BLIND SOURCE SEPARATION WITHOUT OPTIMIZATION CRITERIA?

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 26 63. Fax: +44 141 552 24 87.
e-mail: vicente@spd.eee.strath.ac.uk

ABSTRACT

Blind source separation aims to extract a set of independent signals from a set of observed linear mixtures. After whitening the sensor output, the separation is achieved by estimating an orthogonal transformation, which in the real-mixture two-source two-sensor case is entirely characterized by a single rotation angle. This contribution studies an estimator of such an angle. Even though it is derived from geometric notions based on the scatter-plots of the signals involved, it is found, empirically, to exhibit a performance clearly up to the mark of other methods based on optimality criteria and, theoretically, to improve and generalize one of such procedures. The simplicity of the suggested estimator results in a straightforward adaptive version, which converges regardless of the source distribution, for quite mild conditions, and whose asymptotic analysis is easy to carry out. The applicability of the estimator in a full separation system is also illustrated.

1. INTRODUCTION

In a variety of applications one faces the recovery of a set of unobserved *source* signals from the output of a set of sensors which measure a linear mixture of the sources, where the coefficients of the mixture are also unknown. Instances of this problem, the so-called blind source separation (BSS), are encountered in many different areas, ranging from array processing (direction of arrival estimation, recognition of sources from unknown arrays) to psychology (factor analysis of psychological data), and encompass others such as communications (blind CDMA), speech processing (cocktail party problem) or medical science (antepartum fetal ECG extraction), to name only a few.

The BSS problem has become an emerging research area since the late eighties [7]. The crux of the separation is the source statistical independence assumption. The recovery of the sources is then proven to be equivalent to obtaining a set of independent components at the separator output. To that end, it is convenient to adopt a two-step strategy [5], [9]: in the first step, the measurements are decorrelated and normalized via conventional second-order techniques (principal component analysis); in the second step, higher-order independence is sought. In algebraic terms, it is easily shown that after the first step only an orthogonal matrix needs to be obtained to accomplish the separation. Since any orthogonal matrix can be decomposed as the product of elementary Givens plane rotations, it seems fairly reasonable to take a pairwise approach, processing the whitened output in pairs, until convergence. Even though the convergence of such a strategy has not been theoretically proved yet, it is strongly supported by empirical evidence [4], [5], [6]. In the noiseless two-source two-sensor scenario, an analytical expression for the rotation angle is found in [4], as a function of the cross-cumulants of the whitened observations. Lately, research in the area has witnessed an increasing interest in the use of optimization criteria, through the maximization/minimization of *cost* or *contrast functions*. In [5], the angles are obtained through the maximization of a contrast function made up of the marginal cumulants of the whitened sensor output. Similar cost functions are regarded in [3] to create adaptive separators. The maximum-likelihood (ML) principle is considered in [6], and another closed-form expression of the rotation angle is achieved by optimizing the Gram-Charlier expansion of the observation likelihood function. To arrive at that expression, the sources are assumed to fulfil the validity conditions of the mentioned expansion and to have the same distribution, which diminishes its applicability.

Without direct recourse to any optimization criterion, as opposed to the aforementioned references, another closed-form expression to estimate the relevant angular parameter is derived herein. Interestingly enough, the suggested expression is proven to generalize the approximate ML estimator found in [6], being able to deal with almost any source distribution combination. On the other hand, experiments prove the estimator to perform at the level of quality of other methods based on optimality criteria. In addition, the simplicity of the proposed estimator makes its adaptive version straightforward to obtain, yielding an adaptive algorithm which converges under fairly mild conditions, and does it fast, and whose asymptotic performance analysis is easily analyzed. Other appealing attributes of the angle estimator are highlighted throughout the text.

2. PROBLEM STATEMENT

The objective of the BSS is the reconstruction of 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 $p \geqslant q$ instantaneous linear mixtures measured at the *sensor output*, $\mathbf{y}_k = [y_1(k), \ldots, y_p(k)]^{\mathrm{t}} \in \mathbb{R}^p$, symbol k representing a time index. The noiseless BSS model may be expressed in matrix form as:

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

where matrix $M = (m_{ij}) \in \mathbb{R}^{p\times q}$ contains the mixture coefficients and is hence referred to as *mixing* or *transfer matrix*. Only

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.

two main hypotheses are counted on to achieve the source extraction and mixing matrix identification: the statistical independence of the source components and the linear independence of the columns of M. For the latter matrix, no a priori structure is assumed, and hence the term "blind". The separation is carried out by estimating a *separating matrix* W such as the *global mixing-unmixing system*, given by $G = WM$, exhibits the form PD, P being a permutation matrix and D a regular diagonal matrix. Therefore, an indeterminacy regarding both the ordering and the scale of the recovered sources is allowed, and is indeed inherent to the problem. Consequently, the convention may be adopted, without loss of generality, that the source signals possess an identity covariance matrix: $R_x \triangleq \mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{t}}] = I$. If the sources are to be recovered preserving their original variance, the diagonal entries of D must be unit-norm, the global matrix G then becoming a *quasiidentity* matrix [3].

Usually, the source extraction is accomplished in two steps, associated to the identification of the mixing matrix in two parts, $M = BQ$, with B a regular matrix and Q an orthogonal one (polar decomposition). The first step obtains an estimate of B by means of standard second-order techniques. This process is called (pre-)*whitening*, since transforming the observations according to the pseudoinverse of such a matrix results in the set of whitened components (i.e., decorrelated and unit-variance normalized):

$$\mathbf{z}_k = B^* \mathbf{y}_k = Q\mathbf{x}_k. \tag{2}$$

Hence, B^* is known as *whitening matrix*. Many different ways of performing pre-whitening can be found in the literature [4], [5]. We will concentrate on the estimation of the orthogonal matrix Q.

3. AN OPTIMIZATION-FREE ESTIMATOR

Resorting to the notion of pairwise processing, let us focus on the two-signal case, in which matrix Q reduces to an elementary Givens plane rotation:

$$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{3}$$

The estimation of this matrix, then, reduces to the estimation of a single parameter, angle θ. From equations (2) and (3), the scatter-plot points of the source signals and the whitened observations are rotated versions of one another. In complex form, this relation is readily expressed as:

$$\left.\begin{array}{l} x_1(k) + jx_2(k) = \rho_k \mathrm{e}^{j\phi'_k} \\ z_1(k) + jz_2(k) = \rho_k \mathrm{e}^{j\phi_k} \end{array}\right\} \quad \phi_k = \phi'_k + \theta. \tag{4}$$

The mean point or *centroid* of the whitened-sensor scatter points to the fourth power is

$$\xi \triangleq \mathrm{E}[(z_1 + jz_2)^4] = \mathrm{E}[\rho^4 \mathrm{e}^{j4\phi}], \tag{5}$$

and develops into

$$\xi = \mathrm{E}[(x_1 + jx_2)^4]\, \mathrm{e}^{j4\theta}, \tag{6}$$

by taking into account the relationships in (4). The expectation in the above equation is equal to the source kurtosis sum, $\kappa_{40}^x + \kappa_{04}^x$, where κ_{mn}^x denotes the $(m+n)$th-order cumulant [8] of the source signals. The sources being unknown, that sum does not seem directly available. However, it can be computed from the whitened sensor outputs, as

$$\gamma \triangleq \mathrm{E}[(z_1^2 + z_2^2)^2] - 8 = \mathrm{E}[\rho^4] - 8 = \kappa_{40}^x + \kappa_{04}^x. \tag{7}$$

In conclusion, providing the source kurtosis sum is not null, angle θ can be estimated through:

$$\hat{\theta}_{\mathrm{EML}} = \frac{1}{4}\,\mathrm{angle}(\xi \cdot \mathrm{sign}(\gamma)). \tag{8}$$

Note that no assumptions at all have been made about the source distribution to arrive at (8). It may be applied as long as the source kurtosis sum is distinct from zero. A strikingly similar expression (actually the same when the argument of the centroid (5) lies in $[-\frac{\pi}{2}, \frac{\pi}{2}]$) is obtained in [6] through the ML principle:

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{4}\,\mathrm{arctg}\frac{\sum_k \rho_k^4 \sin 4\phi_k}{\sum_k \rho_k^4 \cos 4\phi_k}. \tag{9}$$

But this latter, due to the restrictions of the Gram-Charlier expansion and other assumptions made during its development, is only valid for symmetric sources with identical distribution and kurtosis value in the range $[0, 4]$. Moreover, estimator (9) is shown [10] to be biased for $\theta \notin [-\frac{\pi}{8}, \frac{\pi}{8}]$, whereas (8) provides the necessary full 90-degree estimation interval. Therefore, expression (8) can be regarded as a generalization of the approximate ML criterion developed in [6], even though no such optimization approach has been taken to determine (8). For these reasons, estimator (8) is referred to as *Extended ML (EML)*. A more detailed link between these two approaches is made in [10] and [11]. A couple of other remarkable features of the above estimator are pointed out next.

Orthogonal invariance. The error introduced in estimator (8) by finite-size sample is merely caused by the source signals, through the sample estimates of the expectations in (6) and (7). From this observation, it is easily proven [12] that estimator (8) is *orthogonal invariant*, thus providing full invariance in the noiseless case and thereby uniform performance characteristics: the source estimation becomes independent of the particular value of the mixing matrix [2].

Asymptotic properties. Both ξ and γ can be expanded as a function of the whitened sensor output cumulants [10]. By consistency of cumulant sample estimates, batch estimator (8) is *consistent* and *asymptotically unbiased*. An exhaustive study of its asymptotic properties is carried out in [11], where a geometrical explanation of the method can also be found.

Extension to more than two signals. In the case we are dealing with more than two mixtures of (possibly) more than two sources, estimator (8) can be applied in turn to each pair of whitened measurements. This pairwise processing idea was already introduced in [4] and mathematically justified in [5]. There is empirical support for this extension with (8), on synthetic [11] as well as on real data [13]. It is found that the number of sweeps over the signal pairs necessary for convergence is about $1 + \sqrt{q}$, which agrees with the value suggested in [5].

How estimator (8) performs when noisy observations are processed is considered in the following section.

4. ASSESSMENT AND COMPARISON

Figures 1 and 2 test the behaviour of the EML method in the presence of noise and compare it with the method described in [5], which is referred to as *HOEVD* (Higher-Order Eigenvalue Decomposition) herein. The HOEVD aims at maximizing the sum of squares of the 4th-order output cumulants, criterion which is proved to ensure maximum independence at that order at the output of the separator. To complete the EML the singular value decomposition (SVD) is used to whiten the measurements, exactly as explained in [5], so that the only difference between the two procedures is the way the pairwise rotations are carried out. The performance index computed (also introduced in [5]) is the gap ε between the true mixing matrix and the estimated one, modulo a post-multiplicative factor of the form PD. The better the separation, the closer ε to zero. All signals are composed of 5000 samples. Mixtures are obtained with a fixed mixing matrix, which, inspired by the experiments in [5], is selected as Toeplitz circulant with the row vector $[1,\,-2,\,3]$. The noise signals are mutually independent as well as independent of the sources. The signal-to-noise ratio (SNR) is defined sensor-wise, that is, as the power due to the sources over the power due to the noise at each sensor, and is chosen to be the same for all sensors. With 3 uniformly-distributed sources and a 3-sensor mixture corrupted by Gaussian noise, the mean ($\mu_{\rm EML}$) and standard deviation ($\sigma_{\rm EML}$) obtained for ε by the EML method over 100 independent Monte Carlo runs at each value of SNR in 1-dB steps is shown in figure 1. The same figure plots the results ($\mu_{\rm HOEVD}$ and $\sigma_{\rm HOEVD}$) obtained over the same mixture and noise realizations by the HOEVD. Both methods provide equal asymptotic behaviour (extreme SNR sides). However, the EML outperforms the HOEVD (lower mean and standard deviation of ε) for low absolute values of SNR, precisely when it becomes more difficult to discern the signals of interest from the noise. The curves obtained under the same simulation conditions as above but with uniformly distributed noise are more discriminant, as seen in figure 2. Remark the anomalous behaviour displayed by the HOEVD in the mentioned SNR range, whereas the EML obeys a more homogeneous performance.

These dissimilar performances seem to be related to the particular choice of the source and noise distributions, as well as the mixing structure. Actually, a number of other experiments with different distributions and mixing matrices indicate that in general the results offered by both methods are very close. This fact is illustrated by figure 3, obtained again under the above conditions but with an arbitrarily chosen non-Toeplitz mixing matrix $M = [1,\,-1,\,1;\;2,\,3,\,4;\;-2,\,1,\,3]$.

5. ADAPTIVE IMPLEMENTATION

The adaptive extension of estimator (8) is equivalent to the adaptive estimation of the centroid location (5) and the source kurtosis sum (7). Calling ν_k the scatter-diagram point of the pre-whitened observations to the fourth power at time instant k:

$$\nu_k = (z_1(k) + jz_2(k))^4 = \rho_k^4 \mathrm{e}^{j4\phi_k}, \qquad (10)$$

then ξ and γ in (5) and (7), respectively, can be adaptively estimated through:

$$\xi_k = (1-\alpha_k)\xi_{k-1} + \alpha_k\nu_k \qquad (11)$$

$$\gamma_k = (1-\alpha_k)\gamma_{k-1} + \alpha_k(|\nu_k| - 8), \qquad (12)$$

where α_k is the adaption coefficient at iteration k, posing the well-known compromise between accuracy and convergence speed. At each iteration k, estimates $\hat{\theta}_k$ of θ are obtained simply by inserting (11) and (12) into (8). From a counter-rotation of this angle, the corresponding source-signal sample at iteration k can be obtained. Such an adaptive procedure is named *Adaptive EML (AEML)*, after its batch equivalent. Equation (11) admits the standard stochastic algorithm form:

$$\xi_k = \xi_{k-1} + \alpha_k H(\xi_{k-1},\,\nu_k), \qquad H(\xi,\,\nu) = \nu - \xi,$$

(and analogously for (12)) and consequently the AEML asymptotic performance (for α_k small enough) can be analyzed with the conventional tools available for the study of such a class of algorithms [1]. Benefiting from those devices, the AEML algorithm described above is proven to converge to the right solution under the same conditions of its batch counterpart, i.e., as long as $\kappa_{40}^x + \kappa_{04}^x \neq 0$. In addition, its stability does not depend on the mixture structure. On the other hand, an analytical expression for the asymptotic probability density function (pdf) of $\hat{\theta}_k$ is determined in [12]. For sufficiently small values of α, the asymptotic variance of $\hat{\theta}_k$ is found to be in the i.i.d. case:

$$\sigma_{\hat{\theta}}^2 \approx \frac{\alpha\sigma_\nu^2}{2|\kappa_{40}^x + \kappa_{04}^x|^2}. \qquad (13)$$

Observe that σ_ν^2 can be written as a function of the source signals only, and hence the performance of the adaptive estimator only depends on them as well. The invariance property of batch estimator (8) (section 3) is thus inherited by its adaptive implementation.

Fast convergence. In order to estimate θ, the pertinent parameter is indeed the orientation of centroid ξ, rather than its precise location. Simulations demonstrate [12] that this orientation is accurately estimated in just a few iterations (nearly independent of the adaption coefficient for a given mixture realization) when the centroid is initialized at the origin of the complex plane.

Figure 4 plots the trajectory of the global matrix coefficients obtained by the AEML (extended to more than two signals in a similar fashion as done for the batch counterpart) for the adaptive separation of 4 pseudorandom binary sources from 4 noiseless instantaneous linear mixtures. The pre-whitening strategy of [3] is employed, λ being its corresponding adaption coefficient. A successful separation is achieved, since G converges to a quasiidentity matrix.

6. CONCLUSIONS

The BSS method studied in this paper is not (explicitly, at least) derived relying on optimization criteria. Despite this, it is proven to possess some very attractive features and to perform at the same (and even higher) level of quality than other methods specially designed to meet certain optimality principles.

At this point, as the question mark in the title evokes, whether such an optimization criterion is hidden behind the method remains to be investigated. Other paths of further research include the extension to complex mixtures, a different generalization to more than two signals, and undertaking the theoretical study of the noise effects.

7. REFERENCES

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

[2] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms", *Proceedings EUSIPCO-94*, Edinburgh, Scotland, U.K., September 1994, pp. 776–779.

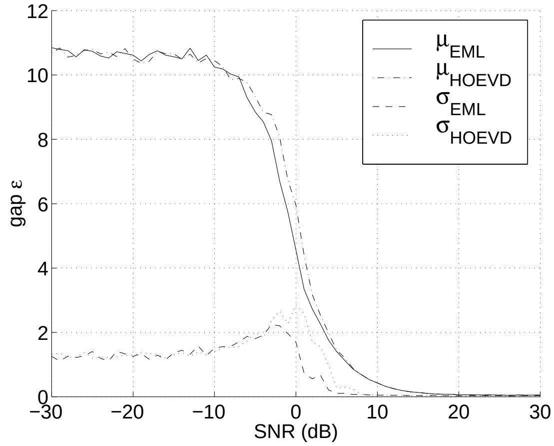

Figure 1: Mean (μ) and standard deviation (σ) of gap ε for a mixture of 3 uniformly distributed sources in additive Gaussian noise.

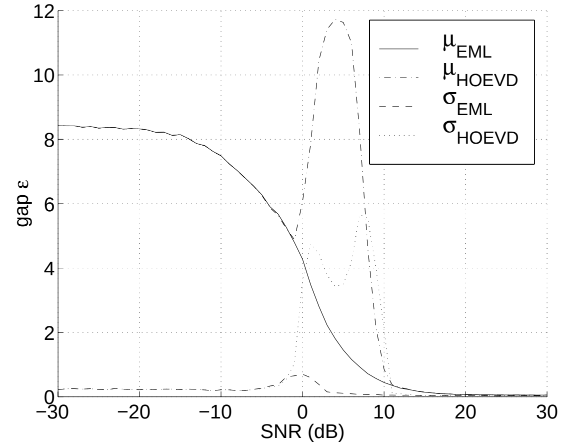

Figure 2: Mean (μ) and standard deviation (σ) of gap ε for a mixture of 3 uniformly distributed sources in additive uniform noise.

[3] 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.

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

[5] P. Comon, "Independent component analysis, A new concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp. 287–314.

[6] 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.

[7] C. Jutten and J. Herault, "Une Solution Neuromimétique au Problème de Séparation de Sources", *Traitement du Signal*, Vol. 5, No. 6, 1989, pp. 389–404.

[8] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. I, Sixth Edition, London: Edward Arnold, 1994.

[9] V. Zarzoso and A. K. Nandi, "The Potential of Decorrelation in Blind Separation of Sources based on Cumulants", *Proceedings ECSAP-97*, Prague, Czech Republic, June 1997, pp. 293–296.

[10] V. Zarzoso and A. K. Nandi, "Generalization of a Maximum-Likelihood Approach to Blind Source Separation", accepted at the forthcoming *EUSIPCO-98*, Rhodes, Greece, 8–11 September 1998.

[11] V. Zarzoso and A. K. Nandi, "Blind Separation of Independent Signals for Virtually any Source Distribution", submitted to *IEEE Transactions on Signal Processing*, 1998.

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

[13] V. Zarzoso and A. K. Nandi, "Blind Separation of Electrocardiographic Sources", submitted to *Medical & Biological Engineering & Computing*, 1998.

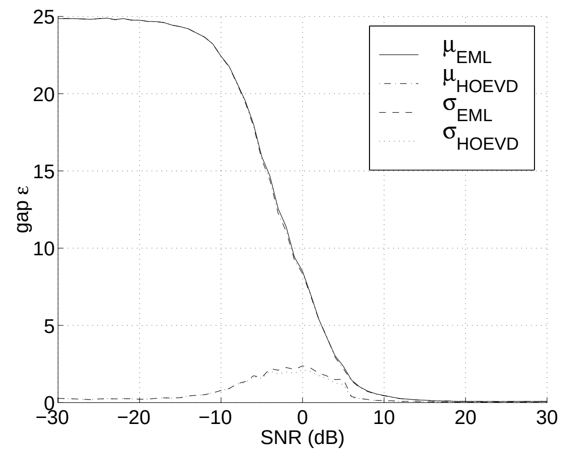

Figure 3: Mean (μ) and standard deviation (σ) of gap ε for another mixture of 3 uniformly distributed sources in additive uniform noise.

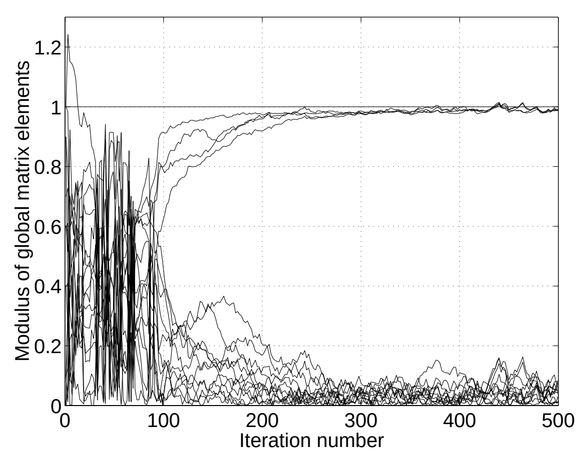

Figure 4: Evolution of the global system in the separation of four binary sources from a four-sensor mixture by the AEML method, with $\lambda = \alpha = 10^{-2}$.