GENERALIZATION OF A MAXIMUM-LIKELIHOOD APPROACH TO BLIND SOURCE SEPARATION

Vicente Zarzoso and Asoke K. Nandi

Signal Processing Division, Department of Electronic and Electrical Engineering
University of Strathclyde, 204 George Street, Glasgow G1 1XW, U.K.
Tel: +44 141 548 2205; fax: +44 141 552 2487
e-mail: vicente@spd.eee.strath.ac.uk

ABSTRACT

In the two-source two-sensor blind source separation scenario, only an orthogonal transformation remains to be disclosed once the observations have been whitened. In order to estimate this matrix, a maximum-likelihood (ML) approach has been suggested in the literature, which is only valid for sources with the same symmetric distribution and kurtosis values lying in certain positive range. In the present contribution, the expression for this ML estimator is reviewed and generalized to include almost any source distribution.

1 INTRODUCTION

The present contribution addresses the problem of the blind source separation (BSS). It consists in the reconstruction of a set of independent source signals from another set of measurements which can be regarded as linear mixtures of the sources. Latest research shows the convenience of a two-step strategy to tackle this problem [2], [3]. In a first stage, the set of observations are decorrelated and normalized by means of a pre-whitening processing carried out by second-order analysis. As a result, only an orthogonal transformation remains to be unveiled, which needs recourse to higher-order techniques in a second stage [1], [2], [4]. In the simplified 2-source 2-sensor scenario, the identification of this orthogonal matrix reduces to the estimation of a Givens rotation matrix which can be written as a function of only one parameter: the rotation angle. Basically, a counter-rotation performed using this angle must provide a set of higher-order (usually fourth-order) independent signals: the wanted sources.

Several methods have been proposed in the literature to estimate this angle. In [1], Comon comes up with an estimator that is a function of the 4th-order cumulants of the pre-whitened sensor signals. The same author, in [2], resorts to the maximization of orthogonal contrast functions, giving rise to the independent component analysis (ICA) of the observations. The ML principle is considered in [3]. Taking the Gram-Charlier expansion of the source probability density function (pdf), an optimal angle in the ML sense is found. However, the conditions for the Gram-Charlier expansion to be valid restrict the applicability of this estimator to the case of symmetric sources with the same distribution and kurtosis values lying in certain positive range.

Herein the last procedure is briefly reviewed and, along the same lines, another method is put forward. Its most appealing feature is that the proposed estimator is able to treat almost any combination of source pdfs. Simulations show it effectively valid for many different source distribution combinations.

Section 2 presents a quick summary of the problem and establishes the notation employed throughout the paper. Section 3 reviews the ML approach developed in [3]. Based on this method, section 4 is devoted to introducing a new angle estimator. Simulations carried out to assess its efficacy are reported in section 5. Finally, conclusions are drawn in section 6.

2 PROBLEM, MODEL AND NOTATION

The aim of BSS can be stated as recovering a set of q zero-mean unit-power independent source signals, $\mathbf{x}(k) = [x_1(k), \ldots, x_q(k)]^{\mathrm{T}}$, from a set of p instantaneous linear mixtures, $\mathbf{y}(k) = [y_1(k), \ldots, y_p(k)]^{\mathrm{T}}$. In matrix form and in the noiseless case, this can be modelled as:

$$\mathbf{y}(k) = M\mathbf{x}(k), \qquad k = 1, 2, \ldots, \tag{1}$$

in which $M \in \mathbb{R}^{p\times q}$ denotes the linear transformation from sources to observations, referred to as *transfer* or *mixing matrix*. After pre-whitening, a set of unit-power uncorrelated observations, $\mathbf{z}(k)$, is obtained. They are easily shown to be related to the original sources through an orthogonal transformation:

$$\mathbf{z}(k) = Q\mathbf{x}(k), \quad k = 1, 2, \ldots \tag{2}$$

The estimation of Q requires the use of techniques based on higher-order statistics. Considering the simplified 2-source 2-sensor scenario, the matrix Q becomes a Givens rotation of the form:

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

In *Proceedings EUSIPCO '98*, Vol. IV, pp. 2069-2072, Rhodes, Greece, 8-11 September 1998.

Therefore, only the rotation angle θ remains to be estimated in order to carry out the source extraction.

The independent realizations of the sources and the whitened observations accept a polar as well as a complex form representation [3]:

$$\left.\begin{array}{l}(x_1(k),\, x_2(k)) = x_1(k) + jx_2(k) = \rho_k \mathrm{e}^{j\phi'_k} = \rho_k \angle \phi'_k \\ (z_1(k),\, z_2(k)) = z_1(k) + jz_2(k) = \rho_k \mathrm{e}^{j\phi_k} = \rho_k \angle \phi_k\end{array}\right\} \quad k = 1, 2, \ldots \tag{4}$$

where, in virtue of equation (2), the angles ϕ'_k and ϕ_k are readily related by

$$\phi_k = \phi'_k + \theta. \tag{5}$$

These polar and complex form representations will prove very convenient at certain points of the development. The problem, then, is reduced to determining the angle θ from the whitened observations. Once this has been done, a counter-rotation provides the original source waveforms.

Let us define the following statistical terms [5]. In the first place,

$$\mu^x_{mn} = \mathrm{E}[x_1^m x_2^n] \tag{6}$$

represents the $(m+n)$th-order moment of the bivariate random variable $(x_1,\, x_2)$. Analogously,

$$\kappa^x_{mn} = \mathrm{Cum}_{mn}[x_1,\, x_2] \tag{7}$$

denotes the $(m+n)$th-order cumulant of the same pair of variables. Similar notation can be employed for the whitened observations, just by changing the super-index "x" by "z" in the moment and cumulant expressions. Furthermore, it will be useful to recall the following relationships [5]:

$$\begin{aligned}\kappa^x_{40} &= \mu^x_{40} - 3(\mu^x_{20})^2 = \mu^x_{40} - 3, \\ \kappa^x_{04} &= \mu^x_{04} - 3(\mu^x_{02})^2 = \mu^x_{04} - 3, \\ \mu^x_{22} &= \mu^x_{20}\mu^x_{02} = 1.\end{aligned} \tag{8}$$

The first two equations express the 4th-order marginal cumulants (or kurtosis) of the sources as functions of their respective moments. The assumption that both signals are unit-power has been taken into account. These two equations are also true for the decorrelated measurements, due to (2), which means that they are both unit-power as well. The last identity comes from the source statistical independence assumption.

3 A MAXIMUM LIKELIHOOD ESTIMATOR

By definition, the ML estimator of θ is the value of the rotation angle that maximizes the (log-)likelihood of the given whitened observations, $\mathbf{z}(k)$. Mathematically, this idea can be expressed as:

$$\hat{\theta}_{\mathrm{ML}} = \arg\max_{\theta} \sum_{k=1}^{N} \log p_{\mathbf{z}}(\mathbf{z}(k) \,|\, \theta), \tag{9}$$

for N independent observations $\mathbf{z}(k)$, $k = 1, 2, \ldots, N$, where $p_{\mathbf{z}}(\cdot)$ denotes the pdf of the decorrelated measurements.

In [3], the source joint pdf (jpdf) is approximated by its Gram-Charlier expansion truncated at the fourth-order term. In order for this expansion to be valid, two constraints are introduced. In the first place, the source marginal pdfs must have zero skewness, i.e., zero 3th-order cumulant ($\kappa^x_{30} = \kappa^x_{03} = 0$), which is basically a symmetry condition. In the second place, the source kurtosis must lie in the range $[0, 4]$, which excludes all short-tailed (platykurtic, with negative kurtosis) pdfs and even some long-tailed (leptokurtic, with positive kurtosis) ones, such as the exponential distribution. Then, developing the ML criterion (9) over the Gram-Charlier expansion of the source jpdf and considering sources with the same distribution, the ML estimator of θ is found to be:

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{4}\,\mathrm{arctg}\frac{\sum_k \rho_k^4 \sin 4\phi_k}{\sum_k \rho_k^4 \cos 4\phi_k}. \tag{10}$$

The properties of the estimator are also studied in [3]. For symmetric source pdfs, it is found to be unbiased for any sample size. In addition, the Cramer-Rao lower bound (CRLB) for the estimation of θ under the conditions of this ML development, i.e., when the source pdfs are approximated by their Gram-Charlier expansion, is deduced to be:

$$\mathrm{Var}(\hat{\theta}) \geqslant \frac{6}{N[(\kappa^x_{40})^2 + (\kappa^x_{04})^2]}, \tag{11}$$

where $\hat{\theta}$ is expressed in radians.

4 GENERALIZATION

Dropping the time index k in the sequel for convenience, let us define:

$$\xi \triangleq \mathrm{E}[\rho^4 \mathrm{e}^{j4\phi}]. \tag{12}$$

According to relationships (4) and (5), equation (12) accepts an expansion as a function of the sources and the unknown rotation angle:

$$\xi = \mathrm{e}^{j4\theta}\mathrm{E}[\rho^4 \mathrm{e}^{j4\phi'}] = \mathrm{e}^{j4\theta}\mathrm{E}[(x_1 + jx_2)^4]. \tag{13}$$

But, from the expressions given in (8),

$$\mathrm{E}[(x_1 + jx_2)^4] = \kappa^x_{40} + \kappa^x_{04}. \tag{14}$$

From the last two equations, and providing that $\kappa^x_{40} + \kappa^x_{04} \neq 0$, angle θ is derived as:

$$\theta = \frac{1}{4}\,\mathrm{angle}\left(\xi \cdot \mathrm{sign}(\kappa^x_{40} + \kappa^x_{04})\right). \tag{15}$$

The argument of the sign function is not known, since by definition the sources are not known either. However, from (4) and (8):

$$\beta \triangleq \mathrm{E}[\rho^4] = \mathrm{E}[(x_1^2 + x_2^2)^2] = \kappa^x_{40} + \kappa^x_{04} + 8, \tag{16}$$

which is also available as a function of the whitened data as:

$$\beta = \mathrm{E}[(z_1^2 + z_2^2)^2]. \tag{17}$$

As a result, when ξ and β are estimated from finite sample observations, the following angle estimator arises:

$$\hat{\theta}_{\mathrm{EML}} = \frac{1}{4}\,\mathrm{angle}\left(\xi \cdot \mathrm{sign}(\beta - 8)\right). \tag{18}$$

To arrive at this expression no assumptions on the source pdfs have been made at all, which makes this estimator valid for *any* source distribution combination with *any* type of symmetry, as long as the source kurtosis sum is not zero.

The connection between estimators (10) and (18) can easily be made as follows. Note first that the numerator and denominator of expression (10) are proportional, respectively, to the sample estimates of the imaginary and real parts of ξ defined in (12), that is,

$$\hat{\theta}_{\mathrm{ML}} = \frac{1}{4}\,\mathrm{arctg}\left(\frac{\mathrm{Im}(\xi)}{\mathrm{Re}(\xi)}\right). \tag{19}$$

By comparing the last two formulae and in virtue of the properties and relationship between the "arctg(·)" and "angle(·)" functions[1], it turns out that

$$\hat{\theta}_{\mathrm{ML}} = \hat{\theta}_{\mathrm{EML}} \quad \Leftrightarrow \quad \theta \in \left[-\frac{\pi}{8},\, \frac{\pi}{8}\right]. \tag{20}$$

Moreover, it is simple to show that if the true angle θ is less than $-\pi/8$ or greater than $\pi/8$, then estimator (10)/(19) produces a bias of $+\pi/4$ or $-\pi/4$ radians, respectively, relative to estimator (18). Estimator (10) is hence a particular case of (18) when θ lies in $[-\pi/8,\, \pi/8]$ (or $[-22.5°,\, 22.5°]$). Even with θ in that interval, the conditions of the Gram-Charlier expansion restrict the applicability of (10) to symmetric sources with kurtosis between 0 and 4, although (20) still confirms the validity of (10) outside that kurtosis range and regardless the source symmetry. In conclusion, the ML procedure is extended through (18) to include almost any source distribution. On this account, expression (18) is referred to as *Extended ML (EML)* estimator. Note, however, that the estimator suggested here may not be the ML estimator in all cases, but it is so, at least, under the conditions of [3].

It is worth computing ξ as a function of the statistical properties of the whitened observations. On the one hand, from (12):

$$\begin{aligned}\xi &= \mathrm{E}[(z_1 + jz_2)^4] = \\ &= (\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z) + j4(\kappa_{31}^z - \kappa_{13}^z),\end{aligned} \tag{21}$$

and, on the other hand, from (13), (14), (16) and (17):

$$\xi = \mathrm{e}^{j4\theta}(\beta - 8) = \mathrm{e}^{j4\theta}(\kappa_{40}^z + 2\kappa_{22}^z + \kappa_{04}^z). \tag{22}$$

[1]Remark that $\mathrm{arctg}(\cdot) \in [-\frac{\pi}{2},\, \frac{\pi}{2}]$, whereas $\mathrm{angle}(\cdot) \in [-\pi,\, \pi]$.

In particular, the modulus of (21) and (22) must be equal, which leads to a relationship among the 4th-order cumulants of the whitened sensor outputs:

$$(\kappa_{31}^z - \kappa_{13}^z)^2 - \kappa_{22}^z(\kappa_{40}^z + \kappa_{04}^z) + 2(\kappa_{22}^z)^2 = 0. \tag{23}$$

This relationship was originally deduced by Comon following more algebraic arguments [1]. Finally, from (13) and (14), ξ may be expressed as a function of the source statistics:

$$\xi = \mathrm{e}^{j4\theta}(\kappa_{40}^x + \kappa_{04}^x). \tag{24}$$

Equations (21), (22) and (24) stress the fact that a rotation θ carried out on the source signals manifests itself as an analogous rotation in the 4th-order cumulant space, so that the 4th-order cumulants of the sources and the decorrelated measurements are related by the aforementioned expressions.

5 SIMULATION RESULTS

In order to test the new estimator performance, several Monte Carlo simulations have been run. Three different combinations of source pdfs regarding their tail or kurtosis sign have been considered: both short, long-short and both long, together with three different symmetry combinations: both symmetric, symmetric-asymmetric and both asymmetric. That makes a total of nine distribution pairs for the source signals. The actual source pdfs employed are (in parenthesis the abbreviations used in the results table): uniform ("Uni"), exponential ("Exp"), Laplacian ("Lap"), Rayleigh ("Ray") and a short-skewed distribution ("Shsk"), which is simply an asymmetric triangular pdf.

For each distribution pair, after creating signal realizations made up of 5000 samples, they are made zero-mean, normalized to unit power, and possible remains of statistical dependence up to 4th-order are removed by using the ICA procedure developed in [2]. Then a rotation matrix of fixed angle, $\theta = 15°$, is applied to the set of source signals, giving a hypothetical set of whitened sensor outputs. The estimation of the rotation angle is carried out through expression (18), where ξ and β are obtained from the whitened sensor data by means of the sample estimates of their definition equations (12) and (16). From each realization, two parameters are computed: the bias $(\hat{\theta} - \theta)$ expressed in degrees, and the values of $(\beta - 8)$ in order to test how they approximate the sum of source kurtosis $\kappa_{40}^x + \kappa_{04}^x$. The mean, standard deviation, maximum and minimum value for these two parameters computed over one hundred Monte Carlo runs are summarized in table 1.

From these results, the new extension can be regarded as successful in estimating the rotation angle θ. This is manifested in the low bias and variance of the estimator, which exhibits an unbiased performance. It is also remarkable how $(\beta - 8)$ approximates the sum of source kurtosis very accurately in all cases (e.g., for uniform–exponential sources is approximately equal to $-1.2 + 6 = 4.8$).

It is interesting to compare the variance obtained here with the bound given by (11). For example, for two Laplacian distributions ($\kappa_{40}^x = \kappa_{04}^x = 3$) and $N = 5000$ samples, equation (11) predicts a lower bound for the standard deviation of $0.4678°$. However, from table 1 the empirical standard deviation obtained for this case turns out to be just $0.1952°$, clearly below that limit. This comparison confirms the initial suspicion, already guessed in the previous section and even in [3], that the conditions of the Gram-Charlier expansion are too restrictive and, even more importantly, not necessary to be met if estimator (10) is to be applied. As a conclusion, the CRLB given in (11) can actually be improved, as seen in these simulation results.

6 CONCLUSIONS

The ML approach proposed in [3] has been considered. By using the Gram-Charlier expansion of the source pdf, the angle estimator that maximizes the likelihood of the given observations is obtained. Nevertheless, the validity domain of the Gram-Charlier development restricts the applicability of this ML estimator to symmetric sources with kurtosis lying in the range $[0, 4]$.

Along the same lines, a new expression for the rotation angle estimator is found. This estimator can be seen as an extension of the ML solution of [3], which, under minor alterations, makes it valid to any source pdf combination, since neither pdf expansions nor assumptions on the source pdfs are necessary; only the source kurtosis sum must be different from zero.

Several simulations prove the validity of the theoretical results, offering satisfactory angle estimation in a wide variety of source pdf combinations, with different tails (i.e., kurtosis sign) and symmetries.

A straightforward extension to the general BSS set-up of more than two sources and two sensors has already been accomplished, in the iterative pairwise fashion proposed in [2]. Performance is also very satisfactory for this extension, although results are not shown here due to the lack of space but will be presented in a future paper.

Acknowledgement. The authors, especially V. Zarzoso, would like to thank the University of Strathclyde for financial assistance.

7 REFERENCES

[1] P. Comon, "Separation of Sources using Higher-Order Cumulants", *SPIE Vol. 1152 Advanced Algorithms and Architectures for Signal Processing IV*, 1989, pp. 170–181.

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

[3] 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.

[4] J.-L. Lacoume and P. Ruiz, "Separation of Independent Sources from Correlated Inputs", *IEEE Transactions on Signal Processing*, Vol. 40, No. 12, December 1992, pp. 3074–3078.

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

DISTR.	BOTH SYMMETRIC			SYMMET.–ASYMMET.			BOTH ASYMMETRIC		
TYPE	PAIR	$\beta - 8$	$\hat{\theta} - \theta$	PAIR	$\beta - 8$	$\hat{\theta} - \theta$	PAIR	$\beta - 8$	$\hat{\theta} - \theta$
SHORT-	Uni-	−2.3974	0.0000	Uni-	−1.7904	−0.0511	Shsk-	−1.1989	−0.0043
-SHORT	-Uni	0.0318	0.0048	-Shsk	0.0521	0.2624	-Shsk	0.0658	0.0605
		−2.2832	0.0168		−1.6190	0.6677		−1.0676	0.1575
		−2.4618	−0.0139		−1.9331	−0.8875		−1.4208	−0.2374
SHORT-	Uni-	1.7945	−0.5954	Uni-	4.8091	−0.1862	Shsk-	5.3531	−0.0357
-LONG	-Lap	0.6772	1.7596	-Exp	1.2021	0.6010	-Exp	1.1262	0.3284
		6.5933	3.4871		8.8689	1.0828		10.8068	0.8244
		0.8864	−4.8006		2.6049	−1.8128		3.3907	−0.7187
LONG-	Lap-	5.8996	−0.0060	Exp-	8.9150	−0.1166	Exp-	6.0870	−0.0028
-LONG	-Lap	0.5954	0.1952	-Lap	1.5991	0.3388	-Ray	1.1810	0.4021
		7.0885	0.5475		16.6219	0.6192		11.5700	0.8987
		4.6835	−1.1677		6.0314	−1.2117		4.2050	−1.4049

Table 1: Results of the simulations with the EML estimator. Signals are composed of 5000 samples. Each four-element column displays, from top to bottom, the mean, the standard deviation, the maximum and the minimum value obtained for the corresponding parameter and signal combination over 100 Monte Carlo runs. Angle values $(\hat{\theta} - \theta)$ are expressed in degrees.