A Contrast for Independent Component Analysis With Priors on the Source Kurtosis Signs

Vicente Zarzoso, *Member, IEEE*, Ronald Phlypo, *Student Member, IEEE*, and Pierre Comon, *Fellow, IEEE*

***Abstract*—A contrast function for independent component analysis (ICA) is presented incorporating the prior knowledge on the sub-Gaussian or super-Gaussian character of the sources as described by their kurtosis signs. The contrast is related to the maximum likelihood principle, reduces the permutation indeterminacy typical of ICA, and proves particularly useful in the direct extraction of a source signal with distinct kurtosis sign. In addition, its numerical maximization can be performed cost-effectively by a Jacobi-like pairwise iteration. Extensions to standardized cumulants of orders other than four are also given.**

***Index Terms*—Blind source separation, contrast functions, higher-order statistics, independent component analysis, kurtosis, performance analysis, standardized cumulants.**

I. Introduction

INDEPENDENT component analysis (ICA) aims at maximizing the statistical independence between the entries of multivariate data. ICA is the fundamental technique for blind source separation (BSS) in linear mixtures when the sources are assumed mutually independent [1]. The plausibility of the assumption in a wide variety of applications has rapidly made of ICA a reference tool in biomedical engineering, communications, and image processing, among many other domains [2]–[4].

In the real-valued noiseless case, ICA assumes the following linear model for the observed data vector $\mathbf{x} \in \mathbb{R}^m$:

$$\mathbf{x} = \mathbf{H}\mathbf{s} \tag{1}$$

where $\mathbf{s} \in \mathbb{R}^n$ contains the independent components or sources and $\mathbf{H} \in \mathbb{R}^{m\times n}$ represents the mixing matrix, with $m \geq n$. The sources are recovered by maximizing a so-called contrast function measuring the statistical independence between the separator output components [1]. Seminal contrasts such as "COM1" and "COM2" originated from cumulant-based approximations (usually at order four) of information-theoretical principles such as maximum likelihood (ML), mutual information, and marginal entropy [1], [5]. The hypothesis that the kurtosis (normalized fourth-order marginal cumulant) of all the sources has the same sign allows the definition of computationally simpler contrasts [5], [6] but is unable to reduce the ambiguity in the ordering of the recovered sources, or permutation indeterminacy, typical in BSS.

The power of the blind approach lies in its robustness to modeling errors, a feature achieved by making as few assumptions about the problem as possible. However, additional information is often available in practice such as the non-Gaussian character of the sources: that of a digital modulation signal depends on the relative probability of its symbols; the atrial activity signal of an atrial fibrillation electrocardiogram is usually sub-Gaussian or quasi-Gaussian; etc. Separation performance can be considerably improved by capitalizing on this information.

The present contribution puts forward a contrast function that takes into account the prior knowledge about the non-Gaussian character of the sources. The new contrast has optimality properties in the ML sense, is efficiently maximized by Jacobi-like iterations, and alleviates (indeed, may totally resolve) the permutation indeterminacy left by blind processing. This latter feature, illustrated in Section IV through simulations, has been successfully put into practice, without mathematical proof, on real signals issued from electrocardiography [7], [8].

Manuscript received September 25, 2007; revised December 20, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yimin Zhang.

V. Zarzoso and P. Comon are with the Laboratoire I3S, Université de Nice-Sophia Antipolis, CNRS, 06903 Sophia Antipolis Cedex, France (e-mail: zarzoso@i3s.unice.fr; pcomon@i3s.unice.fr).

R. Phlypo is with the Department of Electrical and Information Systems (ELIS), Ghent University, Institute for Broadband Technology (IBBT), IBiTech Block Heymans, B-9000 Ghent, Belgium (e-mail: ronald.phlypo@ugent.be).

Digital Object Identifier 10.1109/LSP.2008.919845

II. Contrast Based on Source Kurtosis Signs

Let us first recall the concept of contrast function. The standardization or whitening (second-order processing) of observation (1) yields another vector $\mathbf{z} = \mathbf{Q}\mathbf{s}$, where $\mathbf{Q}$ is a unitary matrix. The sources can then be recovered by applying a unitary transform $\hat{\mathbf{Q}}$, resulting in the separator output $\mathbf{y} = \hat{\mathbf{Q}}^T\mathbf{z} = \mathbf{G}\mathbf{s}$, where $\mathbf{G} = \hat{\mathbf{Q}}^T\mathbf{Q}$. A function $\Psi(\mathbf{y})$ of the separator-output distribution is an orthogonal contrast for ICA if $\Psi(\mathbf{s}) \geq \Psi(\mathbf{G}\mathbf{s})$, for any orthogonal matrix $\mathbf{G}$ (domination), with equality if and only if $\mathbf{G}$ is a trivial filter

$$\mathbf{G} = \mathbf{P}\mathbf{D} \tag{2}$$

where $\mathbf{P}$ is a permutation and $\mathbf{D}$ a non-singular diagonal matrix (discrimination). Consequently, contrast maximization restores the independent sources at the separator output up to a possible permutation and scaling.

Let κ_i denote the ith-source kurtosis and ε_i its sign, $\varepsilon_i = \mathrm{sign}(\kappa_i)$, $1 \leq i \leq n$. We assume in the sequel that p sources have positive kurtosis, $\varepsilon_i = 1$, $1 \leq i \leq p$, and $(n-p)$ sources have negative kurtosis, $\varepsilon_i = -1$, $p < i \leq n$. Symbol μ_i represents the kurtosis of the separator's ith output. Proofs for the mathematical results that follow can be found in the Appendix .

Proposition 1: Criterion

$$\Psi_p(\mathbf{y}) = \sum_{i=1}^{n} \varepsilon_i \mu_i \tag{3}$$

is a contrast function under the above assumptions.

Remark: The maximum likelihood recovery of the source signals under the whitening constraint is achieved by maximizing the following function:

$$\Psi_{\mathrm{ML}}(\mathbf{y}) = \sum_{i=1}^{n} \kappa_i \mu_i. \tag{4}$$

This contrast is obtained from an approximation of the Kullback–Leibler divergence based on the Edgeworth expansion of the separator-output probability density function (pdf) truncated at fourth order [6]. If only the source kurtosis signs are known, contrast (4) naturally reduces to (3). Hence, the latter is expected to inherit the optimality features of the approximate ML estimate while reducing the prior information required. The reduced amount of information helps to keep the desirable features of a blind formulation and is capable of partially solving the permutation ambiguity, as shown by Proposition 2 below.

Remark: Reference [9] addresses the so-called one-bit matching conjecture, whereby the sources can be separated if there exists a one-to-one correspondence between the kurtosis signs of the sources and those resulting from the truncated Gram–Charlier expansion of their pdf's. A function obtained in [9] bears certain resemblance to contrast (3), but the proof of the conjecture is cumbersome and valid only when the source skewness (standardized third-order cumulant) is null. We prove in the Appendix that function (3) is a contrast for all orders $r \geq 3$, of which Proposition 1 is just a particular case for $r = 4$.

Proposition 2: Trivial filters associated with contrast (3) are of the form (2), where

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_2 \end{pmatrix} \tag{5}$$

with $\mathbf{P}_1$ and $\mathbf{P}_2$ being permutation matrices of size $p \times p$ and $(n-p) \times (n-p)$, respectively, and $\mathbf{D}$ made up of unit-norm diagonal entries.

Remark: Sources with positive kurtosis are extracted separately from sources with negative kurtosis by contrast (3), provided that parameter p is known. In particular, a source of interest can be recovered without permutation ambiguity if its kurtosis sign is different from all the others'. The Appendix shows that contrast (3) enjoys this source ordering property for standardized cumulants of even order $r \geq 4$.

III. Contrast Optimization

The Jacobi-like pairwise iteration technique originally proposed in [1] can also be used to optimize contrast function (3). The function is maximized for each signal pair in turn over several sweeps until convergence. Let us assume that we are processing pair $\mathbf{z}_{12} = [z_1, z_2]^T$, the result being readily adapted to other pairs by a simple change of indices. The corresponding two-signal separator output is given by $\mathbf{y}_{12} = \hat{\mathbf{Q}}^T \mathbf{z}_{12}$, where $\hat{\mathbf{Q}}$ is a Givens rotation that can be parameterized as

$$\hat{\mathbf{Q}}(\theta) = \frac{1}{\sqrt{1+t^2}} \begin{pmatrix} 1 & -t \\ t & 1 \end{pmatrix} \tag{6}$$

with $t = \tan\theta$. The associated pairwise contrast is $\Psi(\mathbf{y}_{12}) = \varepsilon_1 \mu_1 + \varepsilon_2 \mu_2$. By virtue of the multilinearity property of cumulants, this function can easily be expressed in terms of the unknown t and the fourth-order cumulants of $\mathbf{z}_{12}$, denoted as $c_{ij} = \mathrm{Cum}_{ij}(z_1, z_2)$, with $(i+j) = 4$ (using Kendall's notation). The stationary points of $\Psi(\mathbf{y}_{12})$ are then found to be the solutions to the quartic equation as follows:

$$a_3 t^4 + 2(a_2 - 2a_4)t^3 + 3(a_1 - a_3)t^2 + 2(2a_0 - a_2)t - a_1 = 0 \tag{7}$$

where $a_0 = (\varepsilon_1 c_{40} + \varepsilon_2 c_{04})$, $a_1 = 4(\varepsilon_1 c_{31} - \varepsilon_2 c_{13})$, $a_2 = 6(\varepsilon_1 + \varepsilon_2) c_{22}$, $a_3 = 4(\varepsilon_1 c_{13} - \varepsilon_2 c_{31})$, and $a_4 = (\varepsilon_1 c_{04} + \varepsilon_2 c_{40})$. The above quartic can be solved by radicals (Ferrari's formula) at a cost that can be considered negligible compared to the cumulant computation. The solutions can also be simply expressed in terms of the extended ML (EML) estimator of [10] if $\varepsilon_1 = \varepsilon_2$ or the alternative EML (AEML) estimator of [11] if $\varepsilon_1 \neq \varepsilon_2$. Typically, about $O(\sqrt{n})$ sweeps over all signal pairs are required for convergence, as suggested in [1]. However, as a by-product of Proposition 2, the extraction of a source of interest with distinct (e.g., positive) kurtosis sign can be carried out by sweeping the contrast over pairs $\mathbf{z}_{1j}$ only, with $\varepsilon_1 = 1, \varepsilon_j = -1$, for $2 \leq j \leq n$. After convergence, the desired source will appear at the first entry of the separator output vector.

IV. Numerical Experiments

The contrast is tested on synthetic random unitary mixtures of $n = 10$ binary signals composed of 1000 samples. Sources kurtosis values of either $\kappa = 2$ (super-Gaussian) or $\kappa = -2$ (sub-Gaussian) are obtained by setting the probability of the two states in the binary distribution accordingly [12]. The error

$$E = \frac{1}{2n(n-1)} \left[\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{|G_{ij}|}{\max\limits_k |G_{ik}|} - 1 \right) + \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \frac{|G_{ij}|}{\max\limits_k |G_{kj}|} - 1 \right) \right] \tag{8}$$

is used as a separation performance criterion [4], [13]. The error is always positive, and zero if and only if matrix $\mathbf{G}$ is a trivial filter of the form (2). Error values are averaged over 250 independent realizations of the sources and the mixing matrix. Three contrasts are considered: "COM2" [1] ($\triangle$ marker); "COM1$+$"; and "COM1$-$"; which correspond to the contrast of [5], assuming that all sources have positive and negative kurtosis, respectively ($+$ and $\times$ markers, resp.); and function (3), which we refer to as "*kurtosis sign priors (KSP)*" contrast ($\circ$ marker). For each tested contrast, we carry out $5(1 + \lfloor \sqrt{n} \rfloor)$ sweeps over all signal pairs.

Fig. 1 shows the performance variation as a function of the number p of sources with positive kurtosis, where p is assumed to be perfectly known *a priori*. As expected, COM1$+$ and COM1$-$ fail to perform the separation, except when all sources have the same kurtosis sign. KSP outperforms the other contrasts.

The robustness of contrast (3) to a mismatch in the prior information is analyzed in Fig. 2, where $\hat{p}$ sources are assumed to

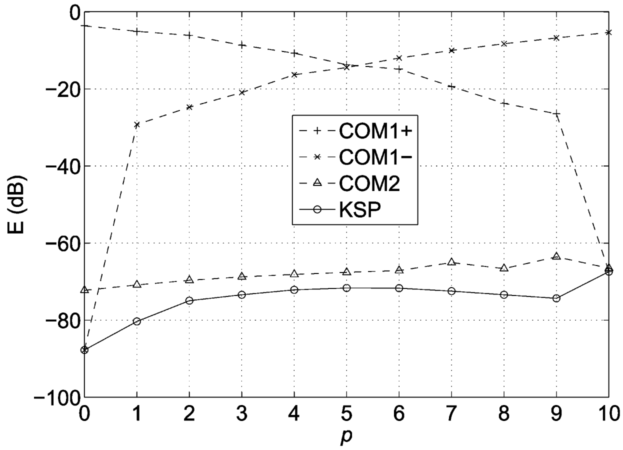

Fig. 1. Source separation performance of ICA contrasts as a function of the number of positive-kurtosis sources p. The KSP method employs the correct value of p.

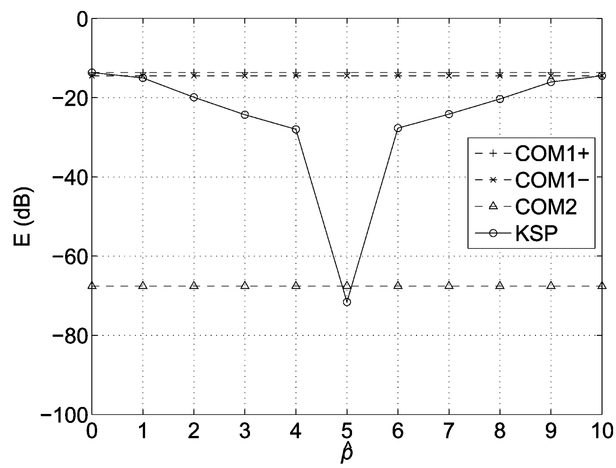

Fig. 2. Source separation performance of ICA contrasts as a function of the estimated number of positive-kurtosis sources $\hat{p}$. The correct value is $p = 5$.

have positive kurtosis while, actually, $p = 5$. KSP's separation performance degrades as the available knowledge becomes less accurate.

Finally, we set $p = 1$ and aim at the single source with positive kurtosis through the extraction procedure described at the end of Section III. Fig. 3 plots the average interference-to-signal ratio (ISR) for the estimation of the first source, defined as

$$ISR = 1 - \frac{|G_{11}|^2}{\sum_{j=1}^{n} |G_{1j}|^2}$$

as a function of the sweep number. This result illustrates the ability of the KSP contrast (3) to extract a source of known kurtosis sign from a mixture where all other sources have the opposite sign, without having to separate the whole mixture and resolve the permutation ambiguity after separation.

V. Conclusions

An orthogonal contrast for ICA has been proposed which takes into account the non-Gaussian character of the source signals as measured by the sign of their fourth-order marginal cumulants (kurtosis). The contrast is linked to an approximate ML principle and is able to separate the independent sources into two groups, depending on their kurtosis sign, thus partially solving the permutation ambiguity usually associated with ICA. The iterative pairwise maximization of the proposed contrast can be carried out at low complexity by closed-form solutions. As opposed to alternative fully blind techniques, the new contrast is particularly suited to the direct extraction of a source with known kurtosis sign distinct from the others'. The principle extends to higher-order cumulants other than kurtosis, as proved in the Appendix.

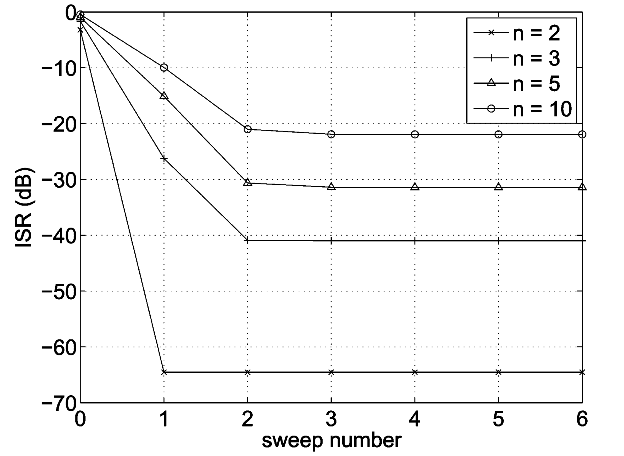

Fig. 3. Source extraction performance of the KSP contrast (3) for different mixture sizes.

Appendix

Proof of Proposition 1: The following proof generalizes the result of Proposition 1 to$r$th-order cumulants, with $r \geq 3$. Accordingly, in the sequel, κ_i and μ_i denote the standardized rth-order cumulant of source s_i and output y_i, respectively, whereas $\varepsilon_i = \text{sign}(\kappa_i)$.

By the multilinearity property of cumulants, we have $\mu_i = \sum_{j=1}^{n} G_{ij}^r \kappa_j$, where $G_{ij} = [\mathbf{G}]_{ij}$. Hence

$$\Psi_p(\mathbf{y}) = \sum_{i=1}^{n} \varepsilon_i \sum_{j=1}^{n} G_{ij}^r \kappa_j.$$

The triangular inequality yields

$$\Psi_p(\mathbf{y}) \leq \sum_{i=1}^{n} \sum_{j=1}^{n} |G_{ij}|^r |\kappa_j| \leq \sum_{i=1}^{n} \sum_{j=1}^{n} |G_{ij}|^2 |\kappa_j|$$

where the right-hand side term stems from the fact that $r \geq 3$ and the orthonormality of matrix $\mathbf{G}$, which can be expressed as $\sum_i |G_{ij}|^2 = 1$. Invoking again this property, we obtain

$$\Psi_p(\mathbf{y}) \leq \sum_{j=1}^{n} |\kappa_j| = \sum_{j=1}^{n} \varepsilon_j \kappa_j = \Psi_p(\mathbf{s}).$$

This proves the domination. Now if the equality $\Psi_p(\mathbf{y}) = \Psi_p(\mathbf{s})$ holds, we must have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left[|G_{ij}|^2 - |G_{ij}|^r \right] |\kappa_j| = 0.$$

Yet all the terms in the sums are positive, and thus, they must all vanish. In other words, $|G_{ij}|^2 - |G_{ij}|^r = 0, \forall\, i,j$, with $r \geq 3$, which can occur only if $|G_{ij}| \in \{0,1\}$. Because $\mathbf{G}$ is orthonormal, it must then have only one nonzero element in every row and column. Hence, $\mathbf{G}$ is of the form (2), with $D_i = [\mathbf{D}]_{ii} = \pm 1$. This proves the discrimination property. Function $\Psi_p(\mathbf{y})$ is thus a contrast for ICA. ∎

Proof of Proposition 2: This proof extends the validity of Proposition 2 to any even order $r \geq 4$. As seen above, equality $\Psi_p(\mathbf{y}) = \Psi_p(\mathbf{s})$ holds if and only if

$$\sum_{i=1}^{n} \varepsilon_i \sum_{j=1}^{n} G_{ij}^r \kappa_j = \sum_{j=1}^{n} \varepsilon_j \kappa_j.$$

Because $D_j = \pm 1$ and $\mathbf{P}$ is a permutation, we have that $G_{ij}^r = P_{ij}$, with $P_{ij} = [\mathbf{P}]_{ij}$, as r is even. Also, $\varepsilon_i^2 = 1$ and $\varepsilon_j \kappa_j = |\kappa_j|$, so that

$$\sum_{j=1}^{n} \left[1 - \sum_{i=1}^{n} \varepsilon_i P_{ij} \varepsilon_j \right] |\kappa_j| = 0.$$

Yet, since all the terms in the sum are positive, they must individually vanish, yielding the relation

$$\sum_{i=1}^{n} \varepsilon_i P_{ij} \epsilon_j = 1, \quad \forall\, j.$$

Now, by splitting the sum into two parts, we are able to replace ε_i by its value, yielding $\sum_{i=1}^{p} P_{ij}\varepsilon_j - \sum_{i=p+1}^{n} P_{ij}\varepsilon_j = 1$. Let us distinguish between the cases $j \leq p$ and $j > p$, and take into account the fact that, for any permutation, $\sum_{i=1}^{n} P_{ij} = 1$. Then

$$\begin{cases} 1 - 2 \sum\limits_{i=p+1}^{n} P_{ij} = 1, & \forall\, j \leq p \\ 1 - 2 \sum\limits_{i=1}^{p} P_{ij} = 1, & \forall\, j > p. \end{cases}$$

The first equality yields, for any $j \leq p$, $\sum_{i=p+1}^{n} P_{ij} = 0$. That is, by positivity, $P_{ij} = 0$. Thus, the $(n-p) \times p$ bottom left block of $\mathbf{P}$ is null. Analogously, we see that for any $j > p$, $\sum_{i=1}^{p} P_{ij} = 0$, and thus, the $p \times (n-p)$ top right block of $\mathbf{P}$ must also be null. Consequently, the permutation matrix takes indeed the form (5). ∎

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[2] S. Haykin, Ed., *Unsupervised Adaptive Filtering*, ser. Adaptive and Learning Systems for Communications, Signal Processing, and Control. New York: Wiley, 2000.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[4] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: Wiley, 2002.

[5] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in *Proc. ICASSP-97, 22nd IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Munich, Germany, Apr. 20–24, 1997, pp. 3453–3456.

[6] J.-F. Cardoso, "Higher-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, pp. 157–192, 1999.

[7] R. Phlypo, Y. D'Asseler, I. Lemahieu, and V. Zarzoso, "Extraction of the atrial activity from the ECG based on independent component analysis with prior knowledge of the source kurtosis signs," in *Proc. EMBC-2007, 29th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society*, Lyon, France, Aug. 23–26, 2007.

[8] R. Phlypo, V. Zarzoso, P. Comon, Y. D'Asseler, and I. Lemahieu, "Extraction of atrial activity from the ECG by spectrally constrained kurtosis sign based ICA," in *Proc. ICA-2007, 7th Int. Conf. Independent Component Analysis and Signal Separation*, London, U.K., Sep. 9–12, 2007.

[9] Z.-Y. Liu, K.-C. Chiu, and L. Xu, "One-bit-matching conjecture for independent component analysis," *Neural Comput.*, vol. 16, no. 2, pp. 383–399, Feb. 2004.

[10] V. Zarzoso and A. K. Nandi, "Blind separation of independent sources for virtually any source probability density function," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2419–2432, Sep. 1999.

[11] V. Zarzoso, A. K. Nandi, F. Herrmann, and J. Millet-Roig, "Combined estimation scheme for blind source separation with arbitrary source PDFs," *Electron. Lett.*, vol. 37, no. 2, pp. 132–133, Jan. 2001.

[12] V. Zarzoso and A. K. Nandi, "Modelling signals of arbitrary kurtosis for testing BSS methods," *Electron. Lett.*, vol. 34, no. 1, pp. 29–30, Jan. 1998.

[13] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proc. ICASSP-94, 19th IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Apr. 1994, vol. 3, pp. 49–52.