LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

A CONTRAST FUNCTION FOR INDEPENDENT COMPONENT ANALYSIS WITHOUT PERMUTATION AMBIGUITY

Vicente Zarzoso, Pierre Comon, Ronald Phlypo

Equipe SIGNAL

Rapport de recherche
ISRN I3S/RR–2009-04–FR

Mars 2009

Laboratoire d'Informatique de Signaux et Systèmes de Sophia Antipolis - UNSA-CNRS
2000, rte.des Lucioles – Les Algorithmes – Bât Euclide B – B.P. 121 – 06903 Sophia-Antipolis Cedex – France
Tél.: 33 (0)4 92 94 27 01 – Fax: 33 (0)4 92 94 28 98 – www.i3s.unice.fr
UMR6070

RÉSUMÉ :

La présente contribution porte sur le problème de la séparation aveugle de sources à travers l'analyse en composantes indépendantes (ACI). Une combinaison linéaire des cumulants marginaux d'ordre quatre (kurtosis) de la sortie du séparateur est un contraste valable pour l'ACI sous l'hypothèse de préblanchiment si les poids ont le même signe que les kurtosis des sources. Si les poids sont égaux aux kurtosis des sources, nous prouvons que le contraste est un critère d'adaptation de cumulants basé sur le principe de maximum de vraisemblance. Si les kurtosis des sources sont différents et les poids aussi (même s'ils ne sont pas adaptés aux premiers), le contraste élimine l'ambiguité de permutation de l'ACI, car les sources estimées sont ordonnées à la sortie du séparateur selon leur kurtosis dans le même ordre que les poids. Dans le cas de deux signaux, la variance asymptotique de l'estimateur de l'angle de Givens résultant est déterminée algébriquement. Le contraste peut être maximisé à faible coût de calcul par un algorithme itératif de type Jacobi opérant sur des paires des signaux. Une étude expérimentale valide les caractéristiques de la technique proposée et la compare à d'autres méthodes basées sur des contrastes.

MOTS CLÉS :

Séparation aveugle de sources, fonctions de contraste, analyse en composantes indépendantes, optimisation de Jacobi, kurtosis, analyse de performances

ABSTRACT:

The present contribution deals with the problem of blind source separation via independent component analysis (ICA). A linear combination of the separator output fourth-order marginal cumulants (kurtoses) is a valid contrast function for ICA under the prewhitening assumption if the weights have the same sign as the source kurtoses. If the weights equal the source kurtoses, we prove that the contrast is a cumulant matching criterion based on the maximum likelihood principle. If the source kurtoses are different and so are the linear combination weights (even if mismatched from the former), the contrast eliminates the permutation ambiguity typical to ICA, as the estimated sources are sorted at the separator output according to their kurtosis values in the same order as the weights. In the two-signal case, the asymptotic variance of the resulting Givens angle estimator is determined in closed form. The contrast can be maximized by means of a cost-efficient Jacobi-type pairwise iteration. An experimental study validates the features of the proposed technique and compares it to related previous methods.

KEY WORDS :

Blind source separation, contrast functions, independent component analysis, Jacobi optimization, kurtosis, performance analysis

A Contrast Function for Independent Component Analysis without Permutation Ambiguity

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

Abstract

The present contribution deals with the problem of blind source separation via independent component analysis (ICA). A linear combination of the separator output fourth-order marginal cumulants (kurtoses) is a valid contrast function for ICA under the prewhitening assumption if the weights have the same sign as the source kurtoses. If the weights equal the source kurtoses, we prove that the contrast is a cumulant matching criterion based on the maximum likelihood principle. If the source kurtoses are different and so are the linear combination weights (even if mismatched from the former), the contrast eliminates the permutation ambiguity typical to ICA, as the estimated sources are sorted at the separator output according to their kurtosis values in the same order as the weights. In the two-signal case, the asymptotic variance of the resulting Givens angle estimator is determined in closed form. The contrast can be maximized by means of a cost-efficient Jacobi-type pairwise iteration. An experimental study validates the features of the proposed technique and compares it to related previous methods.

Index Terms

Blind source separation, contrast functions, independent component analysis, Jacobi optimization, kurtosis, performance analysis.

Manuscript submitted on March 24, 2009.

V. Zarzoso and P. Comon are with the I3S Laboratory, University of Nice - Sophia Antipolis, CNRS, Les Algorithmes, Euclide-B, BP 121, 2000 route des Lucioles, 06903 Sophia Antipolis Cedex, France. {zarzoso, pcomon}@i3s.unice.fr.

R. Phlypo is with the Institute for BroadBand Technology, Department of Electrical and Information Systems, Ghent University, IBiTech Block Heymans, De Pintelaan 185, B-9000 Ghent, Belgium. ronald.phlypo@ugent.be.

I. INTRODUCTION

We consider the problem of blind source separation (BSS) where a set of N possibly complex-valued sources $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T \in \mathbb{C}^N$ are mixed and observed on N sensors. After spatially prewhitening the data, the observation model takes the form:

$$\mathbf{z} = \mathbf{Q}\mathbf{s} \tag{1}$$

where $\mathbf{Q}$ is an unknown $(N \times N)$ unitary matrix, and vector $\mathbf{z} \in \mathbb{C}^N$ represents the whitened observations. The goal is to recover the source realizations from the sole observation of the whitened realizations. To this end, a separating matrix $\mathbf{F}$ is sought so that the separator output vector $\mathbf{y} = \mathbf{F}\mathbf{z}$ is equal to the source vector $\mathbf{s}$ up to admissible indeterminacies. The BSS problem has found a wide range of applications in domains as diverse as telecommunications, geophysics and finance, to name but a few, which helps explain the great interest in this topic witnessed over the last two decades.

Under the assumption of statistically independent sources, these can be estimated with the tool of independent component analysis (ICA) [1]. ICA is typically performed by means of contrast functions quantifying the statistical independence of the separator-output components. Most of these contrasts are functions of cumulant-based approximations of information-theoretical measures such as maximum likelihood (ML) and mutual information (MI) [1], [2]. As in the contrast maximization (CoM2) method of [1], based on the sum of the separator-output squared kurtoses, conventional ICA can at best obtain a source-vector estimate of the form $\mathbf{y} = \mathbf{\Lambda}\mathbf{P}\mathbf{s}$, where $\mathbf{\Lambda}$ is an invertible diagonal matrix and $\mathbf{P}$ is a permutation matrix. While the scale indeterminacy represented by $\mathbf{\Lambda}$ is immaterial in most applications, the permutation ambiguity can lead to an increased computational complexity in situations where only a source, or a small set of sources, is required. In sequential separation schemes, failure to find the source(s) of interest among the first extracted components leads to a poor signal estimation quality caused by error propagation through successive deflation stages.

To overcome these drawbacks, recent works have aimed at reducing the permutation ambiguity of ICA. Reference [3] has proven that the prior knowledge of the source kurtosis signs can fix the permutation ambiguity between sources with different kurtosis signs. As a result, it is possible to directly extract the source of interest if it is the only one to have positive (or negative) kurtosis in the mixture. A computationally efficient Jacobi-type signal-pair sweeping algorithm can be employed to perform source separation or extraction based on this contrast [3].

The present contribution takes a step further in this line of research. A functional based on a weighted linear combination of the separator-output kurtoses is put forward and proven to be a contrast under certain assumptions on the weight coefficients relative to the source kurtosis values. The contrast of [3] and the cumulant-based approximate ML principle of [2] appear as particular instances of the new criterion. In addition, the new contrast eliminates the permutation ambiguity if the source kurtoses are distinct. As in [1], [3], the optimization of the new contrast function can be performed by the cost-effective Jacobi-like iterative technique. Numerical experiments illustrate the comparative performance of the proposed method.

II. PRELIMINARIES AND ASSUMPTIONS

In what follows, all random variables are assumed to have zero mean and unit variance; this conventional standardization is enforced by the prewhitening process and preserved under unitary transformations. The separator-output 4th-order marginal cumulant or kurtosis, defined as $\mu_i = \mathrm{Cum}\{y_i, y_i, y_i^*, y_i^*\}$, is related to the whitened observation 4th-order cumulants, $\gamma_{mnpq} = \mathrm{Cum}\{z_m, z_n, z_p^*, z_q^*\}$, through the multilinear relationship:

$$\mu_i = \sum_{mnpq} F_{im} F_{in} F_{ip}^* F_{iq}^* \gamma_{mnpq} \tag{2}$$

where $F_{ij} = [\mathbf{F}]_{ij}$ and symbol * stands for complex conjugation. If we denote $\mathbf{G}$ the global filter, i.e., $\mathbf{G} = \mathbf{FQ}$ with elements $[\mathbf{G}]_{ij} = G_{ij}$, the separator-output cumulants can also be related to the source kurtoses κ_n as:

$$\mu_i = \sum_{n=1}^{N} |G_{in}|^4 \kappa_n. \tag{3}$$

In the sequel, indices n are assumed, without loss of generality, to be such that κ_n is non decreasing: $\kappa_{n+1} \geq \kappa_n$, $\forall n$.

First assumption (A1). Assume that the first p sources are known to have negative kurtosis, $\kappa_n < 0,\ \ 1 \leq n < p$, and the remaining $(N - p)$ positive kurtosis, $\kappa_n > 0,\ \ p < n \leq N$.

Denote $\mathcal{S}$ the set of sources satisfying this assumption, and $\mathcal{Y}$ the set of observations generated by the orthogonal group $\mathcal{Q}$ acting on $\mathcal{S}$. The following result is proven in [3]:

Proposition 1: The optimization criterion defined as:

$$\Psi_\varepsilon(\mathbf{y}) = \sum_{i=1}^{N} \varepsilon_i \mu_i \tag{4}$$

where $\varepsilon_i = -1$ for $1 \leq i \leq p$, and $\varepsilon_i = 1$ for $p < i \leq N$ is a contrast function over the set of observations $\mathcal{Y} = \mathcal{Q} \cdot \mathcal{S}$.

Second assumption (A2). Assume that the real numbers $\{\alpha_i\}_{i=1}^N$ are related to the unknown source kurtoses $\{\kappa_i\}_{i=1}^N$ via an unknown but strictly increasing function $f(\cdot)$ passing through the origin: $\alpha_i = f(\kappa_i)$.

In other words, we know not only how many positive and negative kurtoses there are (as in assumption A1), but we also know how many are equal and which ones. For instance, if $\alpha_1 < \alpha_2 \leq \alpha_3 < 0 \leq \alpha_4$, then it means that $\kappa_1 < \kappa_2 \leq \kappa_3 < 0 \leq \kappa_4$. Note that because κ_i is non decreasing, so is α_i. In practice, we often have enough information to know such an ordering, but not enough to know the source kurtosis values with good accuracy. This lack of accuracy prevents us from resorting to the ML criterion [2], and we are generally bound to ignore the knowledge of ordering and execute a standard ICA algorithm [1]. The contrast proposed in the next section, while incorporating some prior knowledge on the source kurtosis values, is rather robust to inaccuracies in their estimation. This feature will be illustrated in Sec. V.

III. New Contrast Function

Proposition 2: Under assumption A2, the optimization criterion

$$\Psi_\alpha(\mathbf{y}) = \sum_{i=1}^{N} \alpha_i \mu_i \tag{5}$$

is a contrast function over the set of observations $\mathcal{Y} = \mathcal{Q} \cdot \mathcal{S}$.

See Appendix A for a proof. Now, it was shown in [2] that, for independent sources and prewhitened observations, the 4th-order cumulant approximation to the ML function results in expression (5) with $\alpha_i = \kappa_i$, $1 \leq i \leq N$ [cf. eqn. (3.9) therein]. This cumulant-based approximation, however, was never shown to be a contrast. Proposition 2 not only proves that such a criterion is indeed a contrast, but also extends its validity to other values of $\{\alpha_i\}_{i=1}^N$ as long as they fulfil condition A2. Moreover, we show next that this contrast makes it possible to recover the sources in an order specified beforehand by these coefficients.

Proposition 3: If $\Psi_\alpha(\mathbf{y}) = \Psi_\alpha(\mathbf{s})$, then $\mathbf{y} = \mathbf{\Lambda P s}$, where permutation $\mathbf{P}$ is equal to the identity matrix for every row i (or column i) for which $\alpha_i \neq \alpha_j$, $i \neq j$, and the entries of diagonal matrix $\mathbf{\Lambda}$ are of unit modulus.

This result is proven in Appendix B. Proposition 1 (Proposition 2 of [3]) may now be seen as a particular case of Proposition 3, where coefficients α_i are set to ± 1 according to the source kurtosis signs. However, if the source kurtosis values are all distinct, the maximization of contrast (5) guarantees the recovery of the sources in the order determined by such values relative to weight coefficients α_i. The permutation ambiguity typical to ICA is thus avoided with the use of the new contrast. Again, perfect knowledge of the source kurtoses is not necessary for resolving the permutation ambiguity. Rough guesses of these quantities may suffice, as it is only required that $\{\alpha_i\}_{i=1}^N$ fulfil the conditions of assumption A2.

In [4], a family of cumulant-based contrasts for the blind extraction of $P \leq N$ sources is proposed that resemble (5). Such contrasts are based on the absolute value of rth-order cumulants, and the associated weights must be strictly positive. The natural gradient algorithm used for their maximization does not ensure source recovery free from permutation. If $P = N$ and $r = 4$, contrast (5) can be seen as a generalization of [4] guaranteeing permutation-free source separation.

IV. CONTRAST OPTIMIZATION

As for functional (4), the Jacobi-like pairwise iterative procedure originally proposed in [1] can also be employed to optimize contrast (5). For simplicity, we shall focus on the case of real-valued sources and mixtures. At each iteration, the contrast is maximized for a pair of separator-output signals, $\mathbf{y}_{ij} = [y_i, y_j]^T$, $1 \leq i \neq j, \leq n$, by means of a suitable Givens rotation

$$\mathbf{F}(\xi) = \frac{1}{\sqrt{1+\xi^2}} \begin{bmatrix} 1 & \xi \\ -\xi & 1 \end{bmatrix} \qquad \xi \in \mathbb{R} \tag{6}$$

acting on the corresponding whitened-signal pair $\mathbf{z}_{ij} = [z_i, z_j]^T$. The following claims are proven in Appendix C. Due to the multilinearity relationship of cumulants recalled in eqn. (2), the contrast becomes a rational function of a single parameter ξ with (α_i, α_j) and the 4th-order cumulants of $\mathbf{z}_{ij}$ as coefficients:

$$\Psi_\alpha(\mathbf{y}_{ij}) = \alpha_i \mu_i + \alpha_j \mu_j = \frac{\sum_{k=0}^4 a_k \xi^k}{(1+\xi^2)^2} \tag{7}$$

where $a_0 = \alpha_i \gamma_{iiii} + \alpha_j \gamma_{jjjj}$, $a_1 = 4(\alpha_i \gamma_{iiij} - \alpha_j \gamma_{ijjj})$, $a_2 = 6(\alpha_i + \alpha_j)\gamma_{iijj}$, $a_3 = 4(\alpha_i \gamma_{ijjj} - \alpha_j \gamma_{iiij})$ and $a_4 = \alpha_i \gamma_{jjjj} + \alpha_j \gamma_{iiii}$. The local extrema of this functional are given by the roots of the 4th-degree polynomial

$$a_3 \xi^4 + 2(a_2 - 2a_4)\xi^3 + 3(a_1 - a_3)\xi^2 + 2(2a_0 - a_2)\xi - a_1 \tag{8}$$

which can be obtained algebraically through Ferrari's formula for quartics. The above equation is the same as that found for contrast (4) in [3], but replacing ε_i by α_i in the expressions for coefficients $\{a_k\}_{k=0}^4$. Among the four roots, the one, say ξ_0, maximizing eqn. (7) is retained; this is the *global* maximizer of $\Psi_\alpha(\xi)$ in $\mathbb{R}$. The separator-output signal pair $\mathbf{y}_{ij}$ is then updated by applying matrix $\mathbf{F}(\xi_0)$ onto $\mathbf{z}_{ij}$. The process is repeated for all signal pairs over several sweeps until convergence. Instead of the whitened observations, the most recent update of each separator-output signal is used at each iteration.

Note that in the two-signal case, function $f(\cdot)$ linking coefficients α_i with their respective source kurtoses κ_i (Assumption A2) may not pass through the origin. It actually suffices that:

$$\alpha_1 \kappa_1 + \alpha_2 \kappa_2 > 0 \quad \text{and} \quad (\alpha_1 - \alpha_2)(\kappa_1 - \kappa_2) > 0 \tag{9}$$

These are the necessary conditions for the two-signal contrast (7) to have its global maximum at the separation solution without permutation (Appendix C).

V. ASYMPTOTIC PERFORMANCE ANALYSIS

In the two-signal scenario, the source separation problem reduces to the identification of the rotation angle θ characterizing the Givens rotation $\mathbf{Q}$ in model (1). The asymptotic (large-sample) variance of the estimator of this angle through the maximization of contrast (5) for i.i.d. sources is given by:

$$\mathrm{var}(\hat{\theta}) = \frac{\alpha_1^2 \mathrm{E}\{s_1^6\} + \alpha_2^2 \mathrm{E}\{s_2^6\} - 2\alpha_1\alpha_2 \mathrm{E}\{s_1^4\}\mathrm{E}\{s_2^4\}}{T(\alpha_1\kappa_1 + \alpha_2\kappa_2)^2} \tag{10}$$

where T denotes the sample size (Appendix D). The asymptotic variance of (4) is similarly obtained by replacing α_i by ε_i in eqn. (10). If $\alpha_1 = \kappa_1$ and $\alpha_2 = \kappa_2$, i.e., the weight parameters are adapted to the source kurtoses, the above expression can be shown to be the asymptotic variance of the MI-based CoM2 method of [1] and the approximate ML estimator of [2]. Eqn. (10) can be written as a function of a single parameter, the ratio α_2/α_1. The optimal ratio minimizing (10) is readily computed as:

$$(\alpha_2/\alpha_1)_{\mathrm{o}} = \frac{\kappa_2 \mathrm{E}\{s_1^6\} + \kappa_1 \mathrm{E}\{s_1^4\}\mathrm{E}\{s_2^4\}}{\kappa_1 \mathrm{E}\{s_2^6\} + \kappa_2 \mathrm{E}\{s_1^4\}\mathrm{E}\{s_2^4\}}. \tag{11}$$

To complete the optimal choice of weights (α_1, α_2), it remains to select the sign of α_1 fulfilling the contrast applicability conditions in the two-signal case [eqn. (9)].

The above asymptotic performance results are analogous to those in [5], [6]. Reference [5] finds the optimal weight between two estimators of angle θ based on fourth-order cumulants. Reference [6] aims at the optimal relative weight for a composite contrast made up of squared third- and fourth-order cumulants; CoM2's asymptotic variance can also be obtained as a particular case of the analysis developed therein (a general analysis of CoM2 and related contrasts can be found in [7]). However, contrary to the present work, such contrasts are not designed to reduce the permutation ambiguity of ICA.

VI. IF THE SOURCE KURTOSES ARE UNKNOWN

In a fully blind problem, the source statistics are unknown and so are the weights $\{\alpha_k\}_{k=1}^N$. The optimal weights minimizing the asymptotic variance in the two-signal case cannot be found for the same reason. To surmount this difficulty, a simple two-stage procedure can be proposed as follows. In the first stage, a conventional separation technique such as the CoM2 method [1] is employed to obtain an initial estimation of the sources. Then, the source estimates are ordered according to their kurtosis values. Using the sample statistics of the estimated sources, the optimal weights are computed for each source pair as explained in the previous section. Sweeps are then performed by contrast (5) with the optimal weight coefficients for each signal pair.

VII. Experimental Performance Evaluation

A few numerical experiments test the comparative performance of the contrast developed in this paper. By analogy with function (4), called kurtosis sign priors (KSP) contrast [3], we refer to (5) as *"kurtosis value priors (KVP)"* contrast. In the following, the source signals are N zero-mean unit-variance pseudo-random binary sequences with different kurtosis κ_i, $1 \leq i \leq N$. The source distributions are skewed except for $\kappa_i = -2$. Each source realization, composed of $T = 1000$ samples, is mixed by a random orthogonal matrix with appropriate dimensions, so that no whitening is required. The permutation-sensitive performance index

$$PI = \frac{1}{N}\sum_{i=1}^{N} PI_i, \text{ with } PI_i = (|G_{ii}| - 1)^2 + \sum_{j \neq i}^{N} G_{ij}^2 \tag{12}$$

is averaged over 100 independent realizations of the sources and the mixing matrix. For $N = 2$, this index provides an estimate of $\mathrm{var}(\hat{\theta})$ near permutation-free separation solutions. When considering the CoM2 method of [1], its permutation ambiguity is resolved by suitably re-ordering the estimated sources after separation.

Figure 1 illustrates the fitness of asymptotic variance (10) for the source pair with kurtoses $(-2, 1)$. The theoretical approximation is very accurate in the region where the validity conditions of the KVP contrast hold; the location of the optimal ratio $(\alpha_2/\alpha_1)_\mathrm{o}$ obtained in eqn. (11) agrees with the experimental results. The approximation to CoM2 asymptotic variance, obtained from eqn. (10) with $\alpha_i = \kappa_i$, $i = 1, 2$, is also very precise. KVP with the optimal weight ratio achieves a performance gain of up to 20 dB relative to CoM2 for this particular source combination.

Next, the source kurtoses are randomly chosen without replacement from the set $\{-2, -1, 1, 2, 5\}$. KVP's weight coefficients α_i in (5) are matched to the theoretical source kurtoses, whereas KSP's ε_i in (4) are set to the source kurtosis signs. The average normalized KVP contrast, defined as $\Psi_\alpha(\mathbf{y})/\Psi_\alpha(\mathbf{s})$, is plotted as a function of the pairwise-iteration sweep number in Fig. 2(top). The trajectories of index (12) are shown in Fig. 2(bottom). These plots confirm that the Jacobi-like procedure of Sec. IV is able to maximize contrast (5) and, in turn, this maximization succeeds in separating the sources without permutation. The sources are estimated as accurately as with the original ICA method of [1] followed by re-ordering. The KSP method fails to resolve the source permutation as soon as more than one source with the same kurtosis sign may appear in the mixture, thus the poor average PI values. 'KVP-opt' in Fig. 2(bottom) denotes the procedure described in the previous section that assumes no prior knowledge on the source kurtoses. For all N considered in this experiment, using the optimal weight ratios for each signal pair reduces the PI by half (3 dB) with just a single sweep over the sources estimated by CoM2.

KVP's robustness to the choice of weight coefficients in (5) is assessed by Fig. 3. For the above mixtures realizations, coefficient α_k is left to vary while α_i, $i \neq k$, are kept matched to their respective theoretical source kurtoses. Five sweeps over all signal pairs are performed. A successful permutation-free source separation is achieved for a range of weight values bounded by neighbouring source kurtoses, as pointed out by assumption A2. The impact of the mismatch between the weight parameter and the source sample kurtosis seems to depend on the corresponding source statistics.

VIII. CONCLUSIONS

A contrast function for ICA using fourth-order statistics has been put forward in this paper. The new contrast generalizes a recently proposed function based on the source kurtosis signs [3], proves that the approximate ML criterion of [2] is a contrast and extends it to the case where a mismatch between the weight coefficients and the actual source kurtosis values may appear. In turn, this connection confers the new criterion a certain degree of optimality in the ML sense. As a by-product, our analysis confirms that the CoM2 method of [1], despite arising from the MI principle, presents ML-optimality features, since it achieves, up to permutation, the same asymptotic performance as KVP with weights matched to the source kurtoses. Since these are only approximate ML techniques, asymptotic performance can be further improved by a judicious selection of the weight coefficients in the two-signal case according to theoretical asymptotic analysis results. If the source kurtoses are distinct, only rough guesses on their values suffice for the new contrast to avoid the ICA permutation ambiguity at the separator output. In the case the source statistics are totally unknown a priori, a simple procedure based on the weights with optimal pairwise asymptotic performance can be used to refine a conventional fully blind ICA method. Although the convergence of the pairwise optimization technique used to maximize the contrast is in theory not guaranteed, it has always proven satisfactory in our experiments. Further research should aim at its theoretical proof of global convergence, and the extension of the present contrast to single-source extraction.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewer of [3] who conjectured Proposition 2 proven in this paper.

APPENDIX

A. Proof of Proposition 2

The proposition relies on the following result.

Lemma 1: Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors of $\mathbb{R}^N$, and let the entries of $\mathbf{u}$ be sorted in non decreasing order. Then the permutation $\mathbf{P}$ that maximizes the scalar product $\mathbf{u}^T\mathbf{P}\mathbf{v}$ is the one sorting the elements of $\mathbf{v}$ in non decreasing order.

The proof of this lemma is simple and proceeds by contradiction. Assume that, for the optimal permutation, there exist two entries of $\mathbf{v}$ such that $v_k > v_{k+p}$. By construction, we have $(u_{k+p} - u_k)(v_k - v_{k+p}) > 0$. Expanding the product we get $u_{k+p}v_k + u_kv_{k+p} > u_kv_k + u_{k+p}v_{k+p}$, which shows that transposing the two entries of $\mathbf{v}$ increases the scalar product; hence, the permutation was not optimal.

Now we are ready to prove Proposition 2. Two cases can be distinguished.

- Case 1: Distinct α_i's.

By definition (5) and relationship (3), we can write

$$\Psi_\alpha(\mathbf{y}) \leq \sum_i |\alpha_i| \left| \sum_j G_{ij}^2 G_{ij}^{2*} \kappa_j \right| \leq \sum_{ij} |\alpha_i||G_{ij}|^4|\kappa_j|. \tag{13}$$

Since $\mathbf{G}$ is unitary, we have $|G_{ij}|^4 \leq |G_{ij}|^2$ for any indices, so that:

$$\Psi_\alpha(\mathbf{y}) \leq \sum_{ij} |\alpha_i||G_{ij}|^2|\kappa_j|. \tag{14}$$

Yet, the matrix formed with entries $|G_{ij}|^2$ is itself bistochastic since its rows and columns sum up to one. Hence, from Birkhoff's Theorem [8], there exists a set of real positive numbers β_ℓ such that

$$|G_{ij}|^2 = \sum_\ell \beta_\ell P_{ij}(\ell), \text{ and } \sum_\ell \beta_\ell = 1$$

where $\mathbf{P}(\ell)$ are permutations matrices. This yields the inequality:

$$\Psi_\alpha(\mathbf{y}) \leq \sum_{ij} |\alpha_i||\kappa_j| \sum_\ell \beta_\ell P_{ij}(\ell).$$

The maximum of the right-hand side is reached when the convex linear combination reduces to one of its vertex, that is, when all β's are null but one, say $\beta(\ell_o)$. Then, from Lemma 1, $\mathbf{P}(\ell_o)$ precisely relates j and i, so that both $|\alpha_j|$ and $|\kappa_j|$ are sorted in increasing order:

$$\Psi_\alpha(\mathbf{y}) \leq \sum_j |\alpha_j \kappa_j| = \Psi_\alpha(\mathbf{s}) \tag{15}$$

If the equality holds, then the same reasoning as in [3] would show that $\mathbf{G} = \mathbf{\Lambda}\mathbf{P}$.

- Case 2: Possibly non distinct α_i's.

When α_i's are not distinct, we can group them by packets of equal values. Let $\mathcal{A}_q$ denote the qth such packet. Similarly, values of κ_i can be grouped within the same packets, according to assumption A2. Since permuting indices within a set $\mathcal{A}_q$ does not change the value of the criterion, the proof still holds true. □

B. Proof of Proposition 3

Now we shall make use of the fact that not only moduli $|\alpha_i|$ are sorted, but also weights α_i themselves. If equality holds in (15), it means in particular that there exists a permutation $\mathbf{P}$ such that:

$$\Psi_\alpha(\mathbf{y}) = \sum_{ij} \alpha_i P_{ij} \kappa_j = \sum_j \alpha_j \kappa_j = \Psi_\alpha(\mathbf{s})$$

From Lemma 1, we know that permutation $\mathbf{P}$ is uniquely defined if there is a unique way to sort the κ_n in increasing order. This will be the case if all source kurtoses κ_n are distinct. Should not this be the case, the permutation is not unique: any permutation of indices keeping the order of κ_n non decreasing will still lead to the same maximum of the contrast. The permutation indeterminacy $\mathbf{P}$ is then made up of diagonal blocks $\mathbf{D}(q)$, whose size corresponds to the number of elements in each set $\mathcal{A}_q$. $\square$

It should be remarked that the above proofs are proper to contrast (5) and not immediate extensions of those in [3].

C. Derivation and analysis of the contrast for $N = 2$

In the two-signal case, contrast (5) reduces to $\Psi_\alpha(\mathbf{y}) = \alpha_1\mu_1 + \alpha_2\mu_2$, which is to be maximized under a Givens transformation (6). Using the multilinearity property of cumulants (2), we have:

$$\mu_1 = \frac{\gamma_{1111} + 4\xi\gamma_{1112} + 6\xi^2\gamma_{1122} + 4\xi^3\gamma_{1222} + \xi^4\gamma_{2222}}{(1+\xi^2)^2} \tag{16}$$

$$\mu_2 = \frac{\gamma_{2222} - 4\xi\gamma_{1222} + 6\xi^2\gamma_{1122} - 4\xi^3\gamma_{1112} + \xi^4\gamma_{1111}}{(1+\xi^2)^2} \tag{17}$$

By weighing these expressions by α_1 and α_2, respectively, and adding them together, we readily obtain (7).

Contrast $\Psi_\alpha(\mathbf{y}_{12})$ is actually a function of ξ only, and may be denoted as $\Psi_\alpha(\xi)$ with some abuse of notation. The first derivative of the contrast is given by:

$$\Psi'_\alpha(\xi) = \frac{\psi_\alpha(\xi)}{(1+\xi^2)^3} \tag{18}$$

where $\psi_\alpha(\xi) = P'(\xi)(1+\xi^2) - 4\xi P(\xi)$ and $P(\xi) = \sum_{k=0}^4 a_k \xi^k$. Simple polynomial products lead us to $\psi_\alpha(\xi) = \sum_{k=0}^4 b_k \xi^k$ with $b_0 = a_1$, $b_1 = 2(a_2 - 2a_0)$, $b_2 = 3(a_3 - a_1)$, $b_3 = 2(2a_4 - a_2)$ and $b_4 = -a_3$. The contrast stationary points are the solutions of $\psi_\alpha(\xi) = 0$, which is equivalent to eqn. (8). From (18), the contrast second derivative is given by:

$$\Psi''_\alpha(\xi) = \frac{\psi'_\alpha(\xi)}{(1+\xi^2)^3} - \frac{6\xi\psi_\alpha(\xi)}{(1+\xi^2)^4}. \tag{19}$$

At the stationary points, the second term on the right-hand side cancels out, so that the convexity of the contrast can be studied by analyzing the sign of $\psi'_\alpha(\xi)$ at such points. Other candidate stationary points are $|\xi| \to +\infty$. These are asymptotic horizontal directions with ordinate equal to a_4.

By multilinearity [eqn. (2)], the contrast can be written as a function of the sources and the global matrix entries by replacing the whitened observation cumulants by the source cumulants, defined as $\kappa_{mnpq} = \mathrm{Cum}\{s_m, s_n^*, s_p, s_q^*\}$, in coefficients $\{a_k\}_{k=0}^4$ and redefining $\xi = \tan(\Delta\theta)$, where $\Delta\theta = (\theta - \hat{\theta})$ is the rotation angle parameterizing $\mathbf{G}$. When ensemble statistics are used (i.e., assuming infinite sample size), we have that

$$a_0 = \alpha_1\kappa_1 + \alpha_2\kappa_2 \qquad a_4 = \alpha_1\kappa_2 + \alpha_2\kappa_1 \tag{20}$$

and $a_k = 0$, $k = 1, 2, 3$. Then $\Psi_\alpha(\xi)$ has stationary points if $\psi_\alpha(\xi)$ cancels or when $|\xi| \to +\infty$. Function $\psi_\alpha(\xi) = 4\xi(a_4\xi^2 - a_0)$ is null at $\xi = 0$ and $\xi = \pm\sqrt{\frac{a_0}{a_4}}$. The first root corresponds to the desired permutation-free separation solution. The two other will generally be spurious and can appear only if $\mathrm{sign}(a_0) = \mathrm{sign}(a_4)$. The limit $|\xi| \to +\infty$ achieves source separation with permutation. As explained before, the convexity of the contrast at the stationary points can be ascertained by looking at the sign of $\psi'_\alpha(\xi)$. We have that $\psi'_\alpha(\xi) = 4(3a_4\xi^2 - a_0)$. Accordingly, the desired solution $\xi = 0$ is a local maximum only if $a_0 > 0$. In such a case, the spurious stationary points will be local minima. For the local maximum to be also global, we also require that $\Psi_\alpha(0) > \Psi_\alpha(\xi)|_{|\xi|\to+\infty}$, that is. $a_0 > a_4$. Taking into account eqn. (20), these conditions can be expressed as in eqn. (9).

D. Asymptotic analysis of the contrast for $N = 2$: derivation of variance (10) and optimal weight ratio (11)

If the sample statistics used to compute $\{a_k\}_{k=0}^4$ from finite data length are asymptotically unbiased, so will be the estimator based on the maximization of (5) in the two-signal case, i.e., $\mathrm{E}\{\hat{\theta}\} \to \theta$ as $T \to \infty$. The large-sample variance of the KVP estimator in the (2×2) real-valued scenario can be computed as shown next. First, we denote $\xi = \tan(\Delta\theta)$, with $\Delta\theta = (\theta - \hat{\theta})$, and express the separator output cumulants in terms of the source cumulants, as in Appendix C. For finite sample size, ensemble statistics are approximated by their sample counterparts, giving rise to the sample function $\hat{\Psi}_\alpha(\xi)$. The estimating equation $\hat{\psi}_\alpha(\xi) = 0$ will yield a sample estimate $\hat{\xi}$ of the solution to the contrast optimization. To work out its variance, let us consider the first-order Taylor expansion of $\hat{\psi}_\alpha(\xi)$ around $\hat{\xi}$, which reads: $\hat{\psi}_\alpha(\xi) \approx \hat{\psi}_\alpha(\hat{\xi}) + \hat{\psi}'_\alpha(\hat{\xi})(\xi - \hat{\xi})$. The term $\hat{\psi}_\alpha(\hat{\xi})$ is null since, by hypothesis, $\hat{\xi}$ maximizes the sample contrast. Then, evaluating the above expression at the permutation-free ensemble solution $\xi = 0$ yields:

$$\hat{\xi} \approx -\frac{\hat{\psi}_\alpha(0)}{\hat{\psi}'_\alpha(\hat{\xi})}. \tag{21}$$

For sufficient sample size, we can assume that $\hat{\xi}$ is close enough to the true solution $\xi = 0$ and then $\hat{\psi}'_\alpha(\hat{\xi}) \approx \hat{\psi}'_\alpha(0) = \hat{b}_1 \approx -4\hat{a}_0 = -4(\alpha_1\hat{\kappa}_{1111} + \alpha_2\hat{\kappa}_{2222})$, where we have also considered that the source sample cross-cumulants are negligible relative to the source kurtoses, and so $\hat{a}_2 \ll \hat{a}_0$. Moreover, $\hat{\psi}_\alpha(0) = \hat{b}_0 = \hat{a}_1 = 4(\alpha_1\hat{\kappa}_{1112} - \alpha_2\hat{\kappa}_{1222})$. Under the working assumptions [eqns. (9)], the numerator of (21) will be dominated by the denominator, which can be assumed to be constant and equal to its ensemble average $4a_0 = 4(\alpha_1\kappa_1 + \alpha_2\kappa_2)$. As a result, the variability of $\hat{\xi}$ will mainly stem from the variability of the numerator, so that:

$$\mathrm{E}\{\hat{\xi}^2\} \approx \frac{\mathrm{E}\{(\alpha_1\hat{\kappa}_{1112} - \alpha_2\hat{\kappa}_{1222})^2\}}{(\alpha_1\kappa_1 + \alpha_2\kappa_2)^2}. \tag{22}$$

Now, for whitened sample cumulants estimated as

$$\hat{\kappa}_{iiij} = \frac{1}{T}\sum_{n=0}^{T-1} s_i^3(n)s_j(n) \qquad i \neq j \tag{23}$$

some tedious but otherwise straightforward algebraic derivations show that:

$$\mathrm{E}\{\hat{\kappa}_{iiij}^2\} = \frac{1}{T}\mathrm{E}\{s_i^6\} \qquad \mathrm{E}\{\hat{\kappa}_{iiij}\hat{\kappa}_{ijjj}\} = \frac{1}{T}\mathrm{E}\{s_i^4\}\mathrm{E}\{s_j^4\}.$$

Because $\hat{\xi} \approx 0$, we have $\Delta\theta \approx \hat{\xi}$ and thus $\mathrm{var}(\Delta\theta) \approx \mathrm{var}(\hat{\xi})$. The proof concludes by noticing that $\mathrm{var}(\hat{\theta}) = \mathrm{var}(\Delta\theta)$, hence eqn. (10). Finally, the weight values minimizing the estimator's asymptotic variance is found by cancelling the derivative of (10) with respect to the ratio (α_2/α_1); this readily leads to (11). □

The asymptotic variance of the CoM2 estimator [1] can be worked out similarly.

References

[1] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994, Special Issue on Higher-Order Statistics.

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

[3] V. Zarzoso, R. Phlypo, and P. Comon, "A contrast for independent component analysis with priors on the source kurtosis signs," *IEEE Signal Processing Letters*, vol. 15, pp. 501–504, 2008.

[4] S. Cruces-Alvarez, A. Cichocki, and S. Amari, "From blind signal extraction to blind instantaneous signal separation: criteria, algorithms, and stability," *IEEE Transactions on Neural Networks*, vol. 15, no. 4, pp. 859–873, July 2004.

[5] V. Zarzoso, J. J. Murillo-Fuentes, R. Boloix-Tortosa, and A. K. Nandi, "Optimal pairwise fourth-order independent component analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 3049–3063, Nov. 2006.

[6] S. Abrar and A. K. Nandi, "Independent component analysis: Jacobi-like diagonalization of optimized composite-order cumulants," *Proc. Royal Society A*, Feb. 2009, doi:10.1098/rspa.2008.0381.

[7] L. Albera and P. Comon, "Asymptotic performance of contrast-based blind source separation algorithms," in *Proc. 2nd IEEE Sensor Array and Multichannel Signal Processing Workshop*, Rosslyn, VA, Aug. 4–6, 2002, pp. 244–248.

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

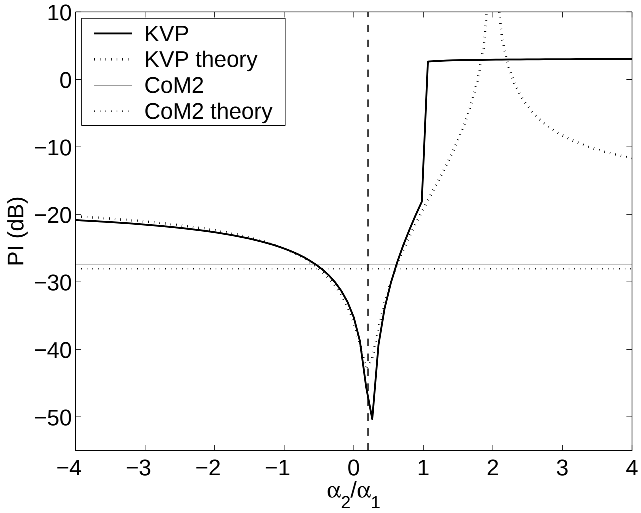

Fig. 1. Fitness of theoretical asymptotic variance. Solid lines represent the average PI values obtained from the separation of random orthogonal mixtures of sources with kurtoses $(-2, 1)$ and $T = 1000$ samples over 100 independent realizations. Dotted lines plot the theoretical asymptotic variance (10) using the source ensemble statistics. The vertical dashed line marks the location of the optimal ratio $(\alpha_2/\alpha_1)_{\rm opt}$ according to (11).

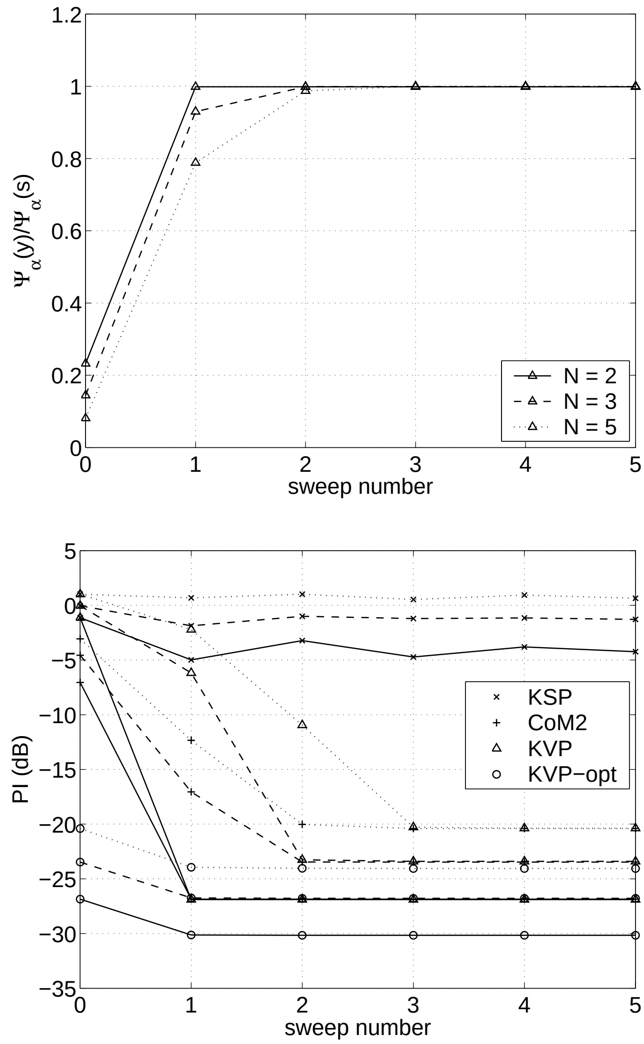

Fig. 2. Source separation performance as a function of the sweep number. (Top) Normalized KVP contrast. (Bottom) Permutation-sensitive quality index. Mixture sizes: $N = 2$ (solid), $N = 3$ (dashed), $N = 5$ (dotted).

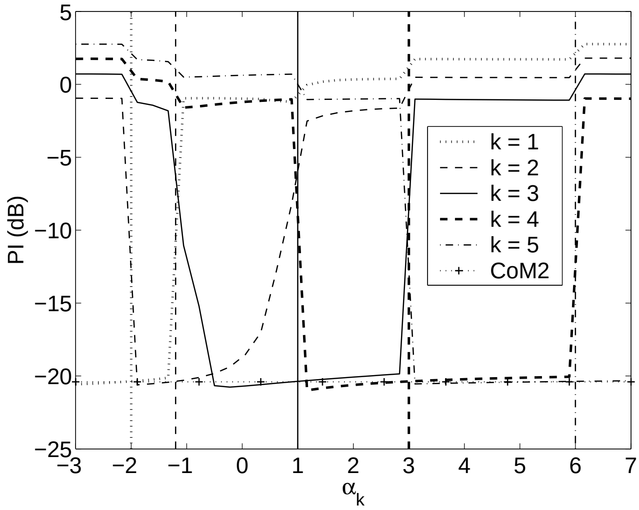

Fig. 3. Source separation performance on the mixture realizations Fig. 2 ($N = 5$, five sweeps) with varying α_k. Coefficients α_i, $i \neq k$, are matched to their respective source kurtoses; these are marked by vertical lines.