performance improves to 3% on average (from 2% minimum to 6% maximum) when we first carry out the proposed unsupervised initialization of the first-layer features, again with approximately equal numbers of false alarms and misdetections (97% accuracy, 98.5% precision, and 97% recall). (The decision threshold of the output node is set at zero, i.e., all positive output values are judged as vehicles while negative values are for nonvehicles.)

In contrast with the results above for the multiview inputs, the results of the 3-D CNN trained on the single-view data are worse by 40% on average when tested on the multiview 94-object test set, indicating that the multiview input is indeed important for improved performance. When the multiview data are also made available to the 3-D CNN for training in the single-view fashion (one input $\mathrm{ind} = 0$ in Fig. 6) its test results are still worse but only by 10% on average, suggesting that the data specifics might be responsible for the reduced performance gain.

IV. Conclusion

We presented a useful framework for categorization of segmented 3-D data via convolutional learning systems. The framework is based on extensions and enhancements of the known 2-D convolutional neural network. We implemented the SMD method to train our 3-D CNN in the supervised mode. We demonstrated noticeable performance improvements when the supervised training is preceded by our proposed unsupervised training of the first-layer features, as well as when the multiview representation of the object is used as the input.

Our approach is advantageous over published approaches of template matching [2], [4] and any projection method in general because it is still possible to loose valuable information about 3-D features when projecting onto 2-D.

Though the discussed application utilizes 3-D data obtained by a lidar, the approach itself is clearly not limited to a specific volumetric sensor. Future work may include 1) more extensive experimentation with this system employing larger, multiclass data sets and optimization of our architectural and training (hyper)parameters, and 2) improving computational efficiency of our implementation by combining sparsity of object representation with naturally parallelizable CNN computations.

Acknowledgment

The author would like to thank D. Dolgov for his help in experimental data collection and segmentation, as well as anonymous reviewers for useful suggestions.

References

[1] G. Arnold, T. Klausutis, and K. Sturtz, "3D laser radar recognition approaches," in *Computer Vision Beyond Visible Spectrum*, B. Bhanu and I. Pavlidis, Eds. New York: Springer-Verlag, 2004, pp. 71–114.

[2] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007.

[3] L. A. Feldkamp, D. V. Prokhorov, C. F. Eagen, and F. Yuan, "Enhanced multi-stream Kalman filter training for recurrent networks," in *Nonlinear Modeling: Advanced Black-Box Techniques*, J. Suykens and J. Vewalle, Eds. Norwell, MA: Kluwer, 1998, pp. 29–53.

[4] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.

[5] Y. LeCun, L. Bottou, and Y. Bengio, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[6] Y. LeCun, F.-J. Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting," in *Proc. Comput. Vis. Pattern Recognit.*, 2004, vol. II, pp. 97–104.

[7] P. Lindner and G. Wanielik, "3D LIDAR processing for vehicle safety and environment recognition," in *Proc. SSCI Comput. Intell. Veh. Veh. Syst.*, Nashville, TN, Mar. 30–Apr. 2 2009, pp. 66–71.

[8] M. Ranzato, F.-J. Huang, Y.-L. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2007.

[9] N. N. Schraudolph, "Fast curvature matrix-vector products for second-order gradient descent," *Neural Comput.*, vol. 14, pp. 1723–1738, 2002.

[10] C. K. Toth, A. Barsi, and T. Lovas, "Vehicle recognition from LiDAR data," *Int. Arch. Photogrammetry Remote Sens.*, vol. XXXIV, pt. 3/W13, pp. 162–166, 2003.

[11] J. Weng and N. Zhang, "Optimal in-place learning and the lobe component analysis," in *Proc. World Congr. Comput. Intell.*, Vancouver, BC, Canada, Jul. 16–21, 2006, pp. 3887–3894.

[12] 3D Lidar Specifications [Online]. Available: http://www.velodyne.com/lidar

[13] DARPA Urban Challenge 2007 Homepage [Online]. Available: http://www.darpa.mil/grandchallenge/index.asp

A Contrast Function for Independent Component Analysis Without Permutation Ambiguity

Vicente Zarzoso, Pierre Comon, and Ronald Phlypo

***Abstract*—This brief deals with the problem of blind source separation (BSS) via independent component analysis (ICA). We prove that a linear combination of the separator output fourth-order marginal cumulants (kurtoses) is a valid contrast function for ICA under prewhitening if the weights have the same sign as the source kurtoses. If, in addition, the source kurtoses are different and so are the linear combination weights, 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. If the weights equal the source kurtoses, the contrast is a cumulant matching criterion based on the maximum-likelihood principle. The contrast can be maximized by means of a cost-efficient Jacobi-type pairwise iteration. In the real-valued two-signal case, the asymptotic variance of the resulting Givens angle estimator is determined in closed form, leading to the contrast weights with optimal finite-sample performance. A fully blind solution can be implemented by computing the optimum weights from the initial source estimates obtained by a classical ICA stage. An experimental study validates the features of the proposed technique and shows its superior performance compared to related previous methods.**

***Index Terms*—Blind source separation (BSS), contrast functions, independent component analysis (ICA), Jacobi optimization, kurtosis, performance analysis.**

I. Introduction

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

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

Manuscript received March 24, 2009; revised March 01, 2010; accepted March 01, 2010. Date of publication March 29, 2010; date of current version April 30, 2010.

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

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

Digital Object Identifier 10.1109/TNN.2010.2045128

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 $\mathbf{y} = \mathbf{Fz}$ is equal to the source vector $\mathbf{s}$ up to admissible indeterminacies.

Under the assumption of statistically independent sources, these can be estimated with the tool of independent component analysis (ICA) [4]. 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) [3], [4]. As in the contrast maximization (CoM2) method of [4] based on the sum of the separator-output squared kurtoses, conventional ICA can at best obtain a source-vector estimate of the form $\mathbf{y} = \mathbf{\Lambda P s}$, where $\mathbf{\Lambda}$ is an invertible diagonal matrix and $\mathbf{P}$ is a permutation matrix. While the scale and phase 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 can also result in a poor signal estimation quality caused by error propagation through successive deflation stages [6]. To partially overcome these drawbacks, the contrast recently presented in [10] fixes the permutation ambiguity between sources with different kurtosis signs if these signs are known *a priori*. A computationally efficient Jacobi-type signal-pair sweeping algorithm can be employed to perform source separation or extraction relying on this contrast.

This brief takes a step further in this line of research. A linear combination of the separator-output kurtoses is proven to be a contrast under certain assumptions on the weight coefficients relative to the source kurtosis values (Section III). The contrast of [10] and the cumulant-based approximate ML principle of [3] appear as particular instances of the new criterion. In addition, the new contrast eliminates the permutation ambiguity if the source kurtoses are distinct and so are the weights. As in [4] and [10], the optimization of the new contrast function can be performed by the cost-effective Jacobi-like iterative technique (Section IV). In the real-valued two-signal case, the asymptotic variance of the resulting Givens angle estimator is determined in closed form (Section V). In most practical settings, however, the source statistics are not known in advance. To surmount this difficulty, a fully blind solution with improved performance can still be implemented by using the weight ratios with optimal finite-sample performance computed from the sources estimated by a classical ICA stage (Section VI). Numerical experiments illustrate the comparative performance of the proposed technique (Section VII).

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 fourth-order marginal cumulant or kurtosis, defined as $\mu_i = \mathrm{Cum}\{y_i, y_i, y_i^*, y_i^*\}$, is linked to the whitened observation fourth-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 nondecreasing, i.e., $\kappa_{n+1} \geq \kappa_n, \forall n$.

1) First Assumption (A1): Assume that the first p sources are known to have negative kurtosis $\kappa_n < 0$, $1 \leq n \leq p$, and the remaining $(N-p)$ have 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 [10].

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 $\mathcal{Y} = \mathcal{Q} \cdot \mathcal{S}$.

2) 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 $\{\kappa_i\}$ are nondecreasing, so are $\{\alpha_i\}$. 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 [3], and we are generally bound to ignore the knowledge of ordering and execute a standard ICA algorithm [4]. The contrast proposed in Section III, while incorporating some prior knowledge on the source kurtosis values, is rather robust to inaccuracies in their estimation. This feature will be illustrated in the experiments of Section VII.

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 part A of the Appendix for a proof. Now, it was shown in [3] that, for independent sources and prewhitened observations, the fourth-order cumulant approximation to the ML function results in expression (5) with $\alpha_i = \kappa_i$, $1 \leq i \leq N$ (see [3, eq. (3.9)]). This cumulant-based approximation, however, was never shown to be a contrast. The proof presented in [8] for a similar cumulant-matching approach requires the sources to have the same cumulant sign. Proposition 2 not only proves that the approximate ML criterion is indeed a contrast whatever the source kurtosis signs, but also extends its validity to other values of $\{\alpha_i\}_{i=1}^N$ as long as they fulfill Assumption 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 part B of the Appendix. Proposition 1 (which is also Proposition 1 of [10]) may now be seen as a particular case of Proposition 2, where coefficients α_i are set to ± 1 according to the source kurtosis signs. However, if the source kurtosis values are all distinct, Proposition 3 shows that 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 $\{\alpha_i\}_{i=1}^N$ fulfill the conditions of Assumption A2. By analogy with function (4), called kurtosis sign priors (KSP) contrast [10], we refer to (5) as kurtosis value priors (KVP) contrast.

The KVP contrast is reminiscent of the nonsymmetrical contrasts presented in [8] and the closely related family of blind extraction contrasts later proposed in [5]. However, these nonsymmetrical contrasts are based on the absolute value of higher order cumulants. As a result, the permutation ambiguity cannot be resolved if the source cumulants are different but have the same absolute values. Moreover, the gradient-based algorithms used for the maximization of such contrasts may get trapped in spurious local extrema. The maximization of contrast (5) through the numerical algorithm described in Section IV has yielded permutation-free source separation in all our experiments whenever Assumption A2 is verified, even if two sources have the same absolute kurtoses (see Section VII).

IV. Contrast Optimization

As for function (4), the Jacobi-like pairwise iterative procedure originally proposed for ICA in [4] can also be employed to optimize contrast (5) in the complex as well as the real case. For simplicity, in the sequel we will 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

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

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

$$\Psi_\alpha(\mathbf{y}_{ij}) = \alpha_i \mu_i + \alpha_j \mu_j = \frac{\sum_{k=0}^{4} a_k \xi^k}{(1+\xi^2)^2} \tag{6}$$

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 function are given by the roots of the fourth-degree polynomial

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

which can be obtained algebraically through Ferrari's formula for quartics. The above equation is the same as that found for contrast (4) in [10], 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 (6) is retained; this is the *global* maximizer of $\Psi_\alpha(\mathbf{y}_{ij})$ with respect to ξ 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

$$\alpha_1\kappa_1 + \alpha_2\kappa_2 > 0 \quad \text{and} \quad (\alpha_1 - \alpha_2)(\kappa_1 - \kappa_2) > 0. \tag{8}$$

These are the necessary conditions for the two-signal contrast (6) to have its global maximum at the separation solution without permutation (part C of the Appendix).

V. Asymptotic Performance Analysis

In the real-valued two-signal scenario, the source separation problem reduces to the identification of 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 independent identically distributed (i.i.d.) sources is given by

$$\mathrm{var}(\hat{\theta}) = \frac{\alpha_1^2 \mathrm{E}\{s_1^6\} + \alpha_2^2 \mathrm{E}\{s_2^6\} - 2\alpha_1\alpha_2 \mathrm{E}\{s_1^4\}\mathrm{E}\{s_2^4\}}{T(\alpha_1\kappa_1 + \alpha_2\kappa_2)^2} \tag{9}$$

where T denotes the sample size (part D of the Appendix). The asymptotic variance of (4) is similarly obtained by replacing α_i by ε_i in (9). 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 [4] and the approximate ML estimator of [3]. Equation (9) can be written as a function of a single parameter, the ratio α_2/α_1. The optimal ratio minimizing (9) is readily computed as

$$\left(\frac{\alpha_2}{\alpha_1}\right)_{\mathrm{opt}} = \frac{\kappa_2 \mathrm{E}\{s_1^6\} + \kappa_1 \mathrm{E}\{s_1^4\}\mathrm{E}\{s_2^4\}}{\kappa_1 \mathrm{E}\{s_2^6\} + \kappa_2 \mathrm{E}\{s_1^4\}\mathrm{E}\{s_2^4\}}. \tag{10}$$

To complete the optimal choice of weights (α_1, α_2), it remains to select their signs so that the contrast applicability conditions in the two-signal case [see (8)] are fulfilled. We refer to the KVP contrast with these optimal weights as KVPopt.

The weight ratio with optimal asymptotic performance given by (10) will be different, in general, from the source kurtosis ratio, i.e., $(\alpha_2/\alpha_1)_{\mathrm{opt}} \neq (\kappa_2/\kappa_1)$, even though the latter can be linked to the cumulant-based ML criterion of [3]. This occurs because the ML criterion is only approximate, as it is based on the truncated Gram–Charlier expansion of the source probability density function. Consequently, its finite-sample performance can be improved by fine-tuning the contrast coefficients according to (10).

The above asymptotic performance results are analogous to those in [1], [8], and [9]. As opposed to [8], here we do not require the sources to have the same kurtosis sign. Zarzoso *et al.* [9] find the optimal weight between two estimators of angle θ based on fourth-order cumulants. Abrar and Nandi [1] aim 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 [2]). However, contrary to the present work, the contrasts in [1] and [9] are not designed to reduce the permutation ambiguity of ICA.

VI. If the Source Statistics Are Unknown

In a fully blind problem, the source statistics are unknown and so are the weights $\{\alpha_i\}_{i=1}^N$. 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 [4] is employed to obtain an initial estimation of the sources. Then, the source estimates are ordered according to their

TABLE I
ENSEMBLE STATISTICS OF THE PSEUDORANDOM BINARY SOURCES USED IN THE EXPERIMENTS. $Prob_i$ DENOTES THE PROBABILITY OF ONE OF THE TWO EVENTS IN THE CORRESPONDING BERNOULLI DISTRIBUTION

κ_i	-2	-1.2	-1	1	2	3	5	6
$\mathrm{E}\{s_i^4\}$	1	1.8	2	4	5	6	8	9
$\mathrm{E}\{s_i^6\}$	1	4.04	5	19	29	41	71	89
$Prob_i$	0.5	0.704	0.724	0.827	0.854	0.873	0.899	0.908

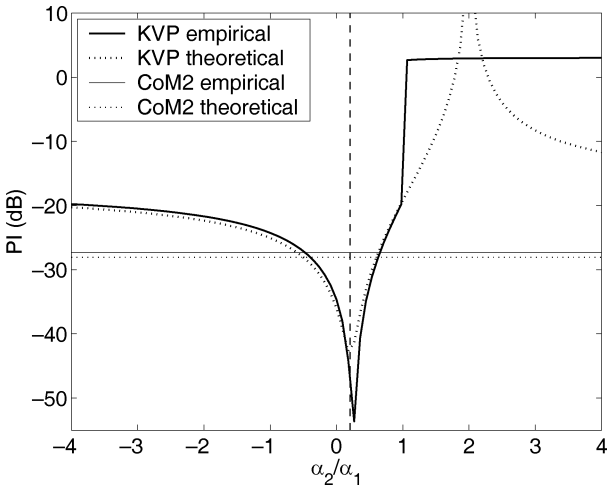

Fig. 1. Fitness of theoretical asymptotic variance. Solid lines represent the average PI values obtained empirically from the separation of random orthogonal mixtures of sources with kurtoses $(-2, 1)$ and $T = 1000$ samples over 100 independent realizations. Dotted lines plot the theoretical asymptotic variance (9) using the source ensemble statistics. The vertical dashed line marks the location of the optimal ratio $(\alpha_2/\alpha_1)_{\mathrm{opt}}$ according to (10).

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. We refer to this fully blind two-stage technique as CoM2-KVPopt.

VII. EXPERIMENTAL PERFORMANCE EVALUATION

A few numerical experiments evaluate the comparative performance of the contrast developed in this brief. In the following, the source signals are N zero-mean unit-variance pseudorandom binary sequences with different kurtosis κ_i, $1 \leq i \leq N$; their ensemble statistics are summarized in Table I. 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

$$\mathrm{PI} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{PI}_i \quad \text{with} \quad \mathrm{PI}_i = (|G_{ii}| - 1)^2 + \sum_{j \neq i}^{N} G_{ij}^2 \qquad (11)$$

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 [4], its permutation ambiguity is resolved by suitably reordering the estimated sources after separation.

Fig. 1 illustrates the fitness of asymptotic variance (9) 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 $(\alpha_2/\alpha_1)_{\mathrm{opt}}$ obtained according to (10) and Table I equals 0.2059 and agrees with the experimental results. The approximation to CoM2's asymptotic variance, obtained from (9) with $\alpha_i = \kappa_i$, $i = 1, 2$, is also very precise. KVP with the optimal weight ratio (KVPopt) achieves a performance gain of up to 20 dB relative to CoM2 for this particular source combination.

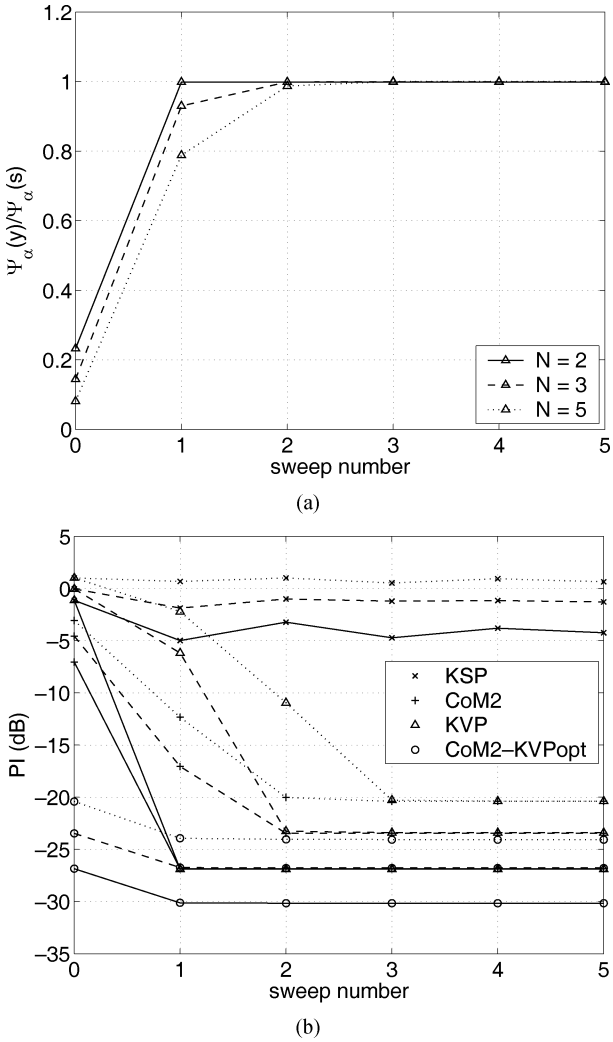

Fig. 2. Source separation performance as a function of the Jacobi sweep number. (a) Normalized KVP contrast. (b) Permutation-sensitive quality index. Mixture sizes: $N = 2$ (solid), $N = 3$ (dashed), $N = 5$ (dotted).

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(a). The trajectories of index (11) are shown in Fig. 2(b). These plots confirm that the Jacobi-like procedure of Section IV is able to maximize contrast (5) and, in turn, this maximization succeeds in separating the sources without permutation. The sources are estimated by KVP as accurately as with the original ICA method of [4] followed by reordering. The KSP method fails to resolve the source permutation as soon as several sources with the same kurtosis sign appear in the mixture, hence the poor average PI values. For all N considered in this experiment, the fully blind CoM2-KVPopt method is able to reduce the PI by half (-3 dB) with just a single sweep of the optimum-weight KVP contrast over the sources estimated by CoM2.

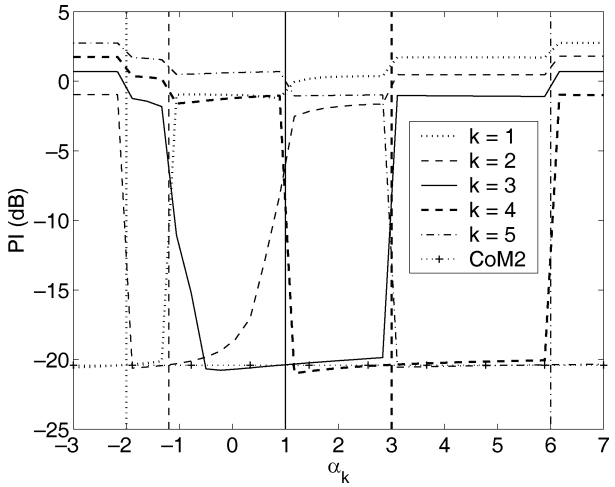

Fig. 3. Source separation performance with varying α_k, for mixtures of $N = 5$ sources and five Jacobi sweeps. Coefficients α_i, $i \neq k$, are matched to their respective theoretical source kurtoses, marked by vertical lines.

KVP's robustness to the choice of weight coefficients in (5) is assessed by Fig. 3, which considers orthogonal mixtures of $N = 5$ sources with kurtosis values $\{-2, -1.2, 1, 3, 6\}$. To obtain each of the curves observed in the plot, 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 using a given set of weight coefficients. A successful permutation-free source separation is achieved for a range of weight values bounded by neighboring source kurtoses, as pointed out by Assumption A2.

VIII. Conclusion

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 [10], proves that the approximate ML criterion of [3] is indeed 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 [4], 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, source separation performance can be further improved by a judicious selection of the weight coefficients in the two-signal case according to the theoretical asymptotic results derived in this work. 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 where 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 a theoretical proof of global convergence, and the extension of the present contrast to single-source extraction.

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 nondecreasing 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 nondecreasing order as well.

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 $v_k > v_{k+p}$, with $p > 0$. By construction, we have $(u_{k+p} - u_k)(v_k - v_{k+p}) > 0$. Expanding the product, we get $u_{k+p}v_k + u_k v_{k+p} > u_k v_k + u_{k+p}v_{k+p}$, 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

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

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

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

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 [7], there exists a set of real positive numbers β_ℓ such that

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

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 vertices, that is, when all β's are null but one, say β_{ℓ_o}. Then, from Lemma 1, $\mathbf{P}(\ell_o)$ precisely relates to j and i so that both $|\alpha_j|$ and $|\kappa_j|$ are sorted in increasing order

$$\Psi_\alpha(\mathbf{y}) \leq \sum_j |\alpha_j \kappa_j| = \Psi_\alpha(\mathbf{s}). \tag{12}$$

If the equality holds, then the same reasoning as in [10] would show that $\mathbf{G} = \mathbf{\Lambda}\mathbf{P}$.

Case 2) Possibly nondistinct 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 will make use of the fact that not only moduli $|\alpha_i|$ are sorted, but also weights α_i themselves. If equality holds in (12), 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 this not be the case, the permutation is not unique: any permutation of indices keeping the

order of κ_n nondecreasing 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 $\mathcal{A}_q$. □

It should be remarked that the above proofs are proper to contrast (5) and not immediate extensions of those in [10].

C. Derivation and Analysis of the Contrast for $N = 2$

The pairwise contrast (6) is actually a function of ξ only, and may be denoted as $\Psi_\alpha(\xi)$ with some abuse of notation. Its first derivative is given by $\Psi'_\alpha(\xi) = \psi_\alpha(\xi)/(1+\xi^2)^3$, where $\psi_\alpha(\xi) = \phi'(\xi)(1+\xi^2) - 4\xi\phi(\xi)$ and $\phi(\xi) = \sum_{k=0}^{4} a_k \xi^k$. 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 finding the roots of polynomial (7). The contrast second derivative is $\Psi''_\alpha(\xi) = \psi'_\alpha(\xi)/(1+\xi^2)^3 - 6\xi\psi_\alpha(\xi)/(1+\xi^2)^4$. 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 horizontal asymptotes with ordinate equal to a_4.

By multilinearity [see (2)], the contrast can be written as a function of the source statistics and the global matrix entries by replacing the whitened observation cumulants by the source cumulants 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

$$a_0 = \alpha_1\kappa_1 + \alpha_2\kappa_2 \qquad a_4 = \alpha_1\kappa_2 + \alpha_2\kappa_1 \tag{13}$$

and $a_k = 0$, $k = 1, 2, 3$. Recall that the stationary points of $\Psi_\alpha(\xi)$ are given by the roots of $\psi_\alpha(\xi)$ and $|\xi| \to +\infty$. Function $\psi_\alpha(\xi) = 4\xi(a_4\xi^2 - a_0)$ cancels at $\xi = 0$ and $\xi = \pm\sqrt{a_0/a_4}$. 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. Since $\psi'_\alpha(\xi) = 4(3a_4\xi^2 - a_0)$, 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 global as well, we also require that $\Psi_\alpha(0) > \Psi_\alpha(\xi)|_{|\xi|\to+\infty}$, that is, $a_0 > a_4$. Taking into account (13), these conditions can be expressed as in (8).

D. Asymptotic Analysis of the Contrast for $N = 2$: Derivation of Variance (9) and Optimal Weight Ratio (10)

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 part C of the Appendix. 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

$$\hat{\xi} \approx -\frac{\hat{\psi}_\alpha(0)}{\hat{\psi}'_\alpha(\hat{\xi})}. \tag{14}$$

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 [see (8)], the numerator of (14) 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

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

Now, for whitened sample cumulants estimated as $T\hat{\kappa}_{iiij} = \sum_{n=0}^{T-1} s_i^3(n)s_j(n)$, with $i \neq j$, some tedious but otherwise straightforward algebraic manipulations show that $T\mathrm{E}\{\hat{\kappa}_{iiij}^2\} = \mathrm{E}\{s_i^6\}$ and $T\mathrm{E}\{\hat{\kappa}_{iiij}\hat{\kappa}_{ijjj}\} = \mathrm{E}\{s_i^4\}\mathrm{E}\{s_j^4\}$. 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 (9). Finally, the weight values minimizing the estimator's asymptotic variance are found by canceling the derivative of (9) with respect to the ratio (α_2/α_1); this readily leads to (10). □

The asymptotic variance of the CoM2 estimator [4] can be worked out similarly.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewer of [10] who conjectured Proposition 2 proven in this brief.

REFERENCES

[1] S. Abrar and A. K. Nandi, "Independent component analysis: Jacobi-like diagonalization of optimized composite-order cumulants," *Proc. R. Soc. A*, vol. 465, no. 2105, pp. 1393–1411, 2009.

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

[3] J.-F. Cardoso, "Higher-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, no. 1, pp. 157–192, 1999.

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

[5] S. Cruces-Alvarez, A. Cichocki, and S. Amari, "From blind signal extraction to blind instantaneous signal separation: Criteria, algorithms, and stability," *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 859–873, Jul. 2004.

[6] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Process.*, vol. 45, no. 1, pp. 59–83, 1995.

[7] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

[8] E. Moreau and N. Thirion-Moreau, "Nonsymmetrical contrasts for sources separation," *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2241–2252, Aug. 1999.

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

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