LABORATOIRE

INFORMATIQUE, SIGNAUX ET SYSTÈMES
DE SOPHIA ANTIPOLIS
UMR 6070

A CONTRAST FOR ICA BASED ON THE KNOWLEDGE OF SOURCE KURTOSIS SIGNS

Ronald Phlypo, Vicente Zarzoso, Pierre Comon, Yves D'Asseler, Ignace Lemahieu

Projet BIOMED

Rapport de recherche
ISRN I3S/RR–2007-13–FR

Mars 2007

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 :

Nous proposons un nouveau critère de contraste pour l'Analyse en Composantes Indépendantes (ACI), basé sur la connaissance des signes des kurtosis des sources. Après pré-blanchiment spatial, le contraste peut être maximisé à l'aide d'un traitement par paires dans lequel chaque rotation plane est calculée avec un très faible coût calcul. On prouve que l'indétermination associée à ce contrsate est une matrice de permutation-échelle composée de deux blocs, chacun correspondant à un signe. Par exemple, si on désire extraire une seule source et que le signe de son kurtosis est différent de celui des autres, alors on peut l'extraire sans séparer les autres sources du mélange. Les résultats expérimentaux montrent que les performances d'extraction s'améliorent avec l'écart entre les kurtosis.

MOTS CLÉS :

contraste, séparation aveugle de sources, Analyse en Composantes Indépendantes, déflation rapide

ABSTRACT:

We propose a new contrast criterion for independent component analysis (ICA) based on the prior knowledge of the source kurtosis signs. After prewhitening, the contrast can be optimized by a pairwise processing approach in which plane rotations are found at low computational cost at each iteration. It is proved that the indeterminacy associated with this contrast is a scaled permutation matrix composed of two blocks, each corresponding to a source kurtosis sign. Hence, if the source of interest has a kurtosis sign different from that of the others, it can be extracted without separating the whole mixture. Experimental results show that the source estimation performance improves with the kurtosis gap.

KEY WORDS :

Contrast, Blind Source Separation, Independent Component Analysis, Fast Deflation

A Contrast for ICA Based on the Knowledge of Source Kurtosis Signs

Ronald Phlypo[1,2*], Vicente Zarzoso[1], Pierre Comon[1], Yves D'Asseler[2], Ignace Lemahieu[2]

[1] Laboratoire I3S, CNRS/UNSA Les Algorithmes - Euclide-B, BP 121, 2000 Route des Lucioles, 06903 Sophia Antipolis Cedex, France
{phlypo, zarzoso}@i3s.unice.fr
[2] MEDISIP, ELIS/UGent, IBiTech - Campus Heymans, De Pintelaan 185, B-9000 Ghent, Belgium
{ronald.phlypo, yves.dasseler, ignace.lemahieu}@ugent.be

Abstract. We propose a new contrast for ICA based on the prior knowledge of the source kurtosis signs. It is shown that the contrast can be used for source extraction when the kurtosis sign of the source of interest is different from that of the other sources. Experimental results show that the source estimation reliability increases with the kurtosis gap. In addition, the numerical algorithm presents a very attractive computational complexity.

1 Introduction

Independent Component Analysis (ICA) can be seen as a solution to the blind source separation problem in case the sources underlying the observations are considered statistically independent. In this paper, the datamodel used is based on an instantaneous mixture model of real valued statistical variables $\boldsymbol{x}$ taken from a statistical set of distributions $\mathcal{X} \subset \mathbb{R}^n$, whereupon a mixing matrix $\boldsymbol{M}$ acts to result in the observations $\boldsymbol{y}$, up to some additional noise $\boldsymbol{\eta}$. The model can be represented as:

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{\eta} \ , \tag{1}$$

The mixture matrix $M \in \mathbb{R}^{m \times n}$ is an arbitrary full column rank matrix defining the transformation from $\mathcal{X}$ to $\mathcal{Y}$, the set of statistical distributions of which the observations $\boldsymbol{y}$ are taken. The estimation of $\boldsymbol{M}$ is generally a complex task which can be simplified by prewhitening the observations $\boldsymbol{y}$, yielding $\boldsymbol{z} \in \mathcal{Z} \subset \mathbb{R}^m$. Since the distributions share no second order statistics thanks to the prewhitening step. The mixing matrix to be estimated is hence restricted to the estimation of a matrix $\boldsymbol{Q}$ in the group of orthogonal matrices $\mathcal{Q} \subset \mathbb{R}^{n \times n}$. In the noiseless case and for an equal number of observations and sources, the equivalent mixture model after prewhitening can then be expressed as

$$\boldsymbol{z} = \boldsymbol{Q}\boldsymbol{x} \ , \tag{2}$$

* R. Phlypo would like to thank H. Rix and V. Zarzoso for their kind hospitality at the laboratories of I3S.

where $\boldsymbol{Q}$ is to be estimated, given $\mathcal{Z}$ through its heuristic approximations based on the observations in $\boldsymbol{z}$. Generally, the model given in Eq. (2) is solved by a search algorithm exploring the dataspace, either on a component by component basis (e.g. fastICA [1], robustICA [2] in deflation or regression mode), either through ensemble learning (a.o. JADE [3], most of the bayesian methods). In this paper we propose a source extraction criterion, which can be seen as a component by component decomposition, which will only estimate the source of interest.

Most ICA algorithms consider the set $\mathcal{X}$ as a set of identical distribution, and thus the sources x as independently identically distributed stochastic variables. However, when the assumption of identical distributions does not hold, the solution to the problem is no longer the Maximum Likelihood (ML) estimator. If we suppose our distributions can be characterised by fourth order statistics [4] and we assume symmetric distribution (i.e. they are indistinguishable by their odd order statistics), for identically independently distributed data this translates into $\kappa_{iiii} = \kappa_{jjjj}, \forall i, j$, which is the ML estimator for ICA [5]. Nonetheless, it has been shown that maximising the sum of the kurtosis values (or of their quadratics), is still a contrast for ICA and is thus suitable as an objective to separate independently distributed variables from a mixture [4]. For such a contrast function it is even possible to find an algebraic expression in the two sources, two observations scenario through the Extended ML (EML) estimator [6]. The estimator is not able to separate a mixture when the source kurtosis sum attains zero, though. However, when the kurtosis sum is null, the difference between the source kurtosis becomes significant[3]. Exploiting the latter yields the Approximate EML (AEML) estimator [6].

This paper will extend the 2×2 AEML criterion to AEML based source extraction (AEMLe) from a general $n \times n$ mixture, given that the source of interest differs in kurtosis sign from the other sources in the mixture. This result fits in a more general framework of a contrast for ICA with known source kurtosis signs, the proof of which is given in appendix A.

2 Methods

2.1 Givens Rotations

An orthonormal matrix in $\mathbb{R}^{n \times n}$ can be represented by a series of parameters θ_i, describing subsequent plane rotations in $\mathbb{R}^{n \times n}$. A givens rotation in the plane spanned by the components $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ can thus be represented by an angle θ resulting in the equivalent matrix multiplication:

$$\hat{\boldsymbol{x}}_{ij} = \boldsymbol{Q}(\theta)^T \boldsymbol{z}_{ij} \quad , \quad \text{where } Q(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad . \tag{3}$$

[3] provided we have maximally one gaussian distributed or mesokurtic source in the mixture, which is a general requirement to maintain identifiability in the ICA model

This representation has its major advantage in that one can algebraically calculate an optimal value for the single parameter θ for each signal pair, once an objective is defined for the plane rotations. This reduces the computational burden significantly [4, 7].

2.2 Source Kurtosis Difference as a Contrast

Assuming the signs of the source kurtosis are known, we can put forward a contrast[4] for ICA by defining

$$\Psi(\boldsymbol{Q}) = \sum_{j=1}^{n} \epsilon_j \kappa_{jjjj}^{z} \ , \tag{4}$$

where ϵ_j represents the kurtosis sign of the j^{th} source and the 4^{th} order cumulants are noted as $\kappa_{i_1,i_2,i_3,i_4} = \text{Cum}(z_{i_1}, z_{i_2}, z_{i_3}, z_{i_4})$. For the 4^{th} order standardised marginal cumulants we get:

$$\kappa_{iiii} = \frac{\text{E}\left\{z_i^4\right\} - 3\text{E}\left\{z_i^2\right\}^2}{\text{E}\left\{z_i^2\right\}^2} \ , \tag{5}$$

which is negative for a subgaussian and positive for a supergaussian random variable.

2×2 Source Estimation. In the case of two sources with $(\epsilon_1, \epsilon_2) = (1, -1)$ and two observations, Eq. (4) simplifies to

$$\Psi(\boldsymbol{Q}) = \kappa_{1111} - \kappa_{2222} \ . \tag{6}$$

Maximising the aforementioned criterion by rotating in the variables' plane around an angle $\theta_\star$ using plane givens rotations will yield the source estimates. Expressing the 4^{th} order cumulants of the sources $\lambda_1 = \kappa_{1111}^{\boldsymbol{x}}, \lambda_2 = \kappa_{2222}^{\boldsymbol{x}}$ in function of the cumulants of the observations we obtain the following objective function:

$$\begin{aligned}\Psi(\theta) &= \lambda_2 - \lambda_1 \\ &= \left(\cos^2\theta - \sin^2\theta\right)\alpha + 4\beta\cos\theta\sin\theta \\ &= \alpha\cos 2\theta + 2\beta\sin 2\theta \ ,\end{aligned} \tag{7}$$

where we used the multilinearity of the cumulants and Eqs (3) and (4) and where α and β are given by $(\kappa_{1111}^{z} - \kappa_{2222}^{z})$ and $(\kappa_{1112}^{z} + \kappa_{1222}^{z})$, respectively. The optimal value $\theta_\star$ is then given by the stationary point of Eq. (7), which can be calculated setting its derivative at zero, and thus:

$$2\theta_\star = \arctan\frac{2\beta}{\alpha} \ . \tag{8}$$

This is the result also found in [6] for the AEML estimator.

[4] The proof of Eq. (4) being a contrast can be found in appendix A

Source Extraction Model. The contrast function of Eq. (4) can be rewritten for the extraction of a single source that differs in kurtosis sign from the others in the mixture as:

$$\Psi\left(\boldsymbol{Q}\right) = \epsilon_2 \left(\sum_{j=2}^{n} \kappa^z_{jjjj} \right) - \epsilon_1 \kappa^z_{1111} \ , \tag{9}$$

where we have chosen index 1 as the position for the source of interest. Since the estimation of an orthogonal mixing matrix in $\mathcal{Q}$ can be expressed as a set of successive pairwise givens rotations, it remains to define an appropriate iteration scheme to obtain the source estimate.

Pairwise Optimisation Iteration Scheme. Based on the existing iteration schemes for pairwise optimising and the result in Eq. (8), we can now optimise the contrast function by running several sweeps with a predefined iteration scheme. It has been mentioned in [7, 4, and references therein] that the optimal sweeping for pairwise processing functions can be done by using at each sweep either a fixed sequence (natural cycle), either a sequence based on a ranking according to the value the current source estimates take in the contrast function (jacobian iterations) or by using a threshold based selection at each sweep. For the purpose of source extraction, combination of both the natural cycle method and the jacobian based procedure to update the source estimate seems the most adequate, see 1 for details. By simple adaptation of the updating sequence we can even limit ourselves to the extraction of a single source, avoiding the need to estimate all sources.
Although it has not yet been proven that pairwise optimisation yields a global solution to the contrast, its use is justified since no counterexamples have been found showing misconvergence if the general conditions of the model are met [7, 4, 6]. Its use is further justified by the fact that if the independence criterion holds for each pair, then it holds for the whole set, which can be deduced directly from the information criteria supporting ICA.
The update sequence for single source extraction from n observers, according to the model in (9), then becomes: To ensure the algorithm to terminate, we

Table 1. Sweep algorithm for the proposed contrast with source extraction

Initialise $\hat{\boldsymbol{s}} = \boldsymbol{z}$
Start $\lfloor 1 + \sqrt{m} \rfloor$ sweeps over k
Start: For j from 2 to m
$\hat{\boldsymbol{s}}_{1j}(k+1) = \boldsymbol{Q}\left(\theta_\star\right)^T \hat{\boldsymbol{s}}_{1j}(k)$
Undo possible permutation
End j-loop
End Sweep

need to include the step to undo a possible permutation[5]. The sweeps over the iterations reach termination once no significant rotation has been encountered in a full sweep.

3 Results

Performance Measure. The performance measure is given as the mean of 200 Monte Carlo runs for each simulation point, either using a mixture of binary sources or of FM/AM sources. All observations are obtained through a randomly generated orthogonal $n \times n$ mixture. The performance measure is expressed as the distance between the source of interest and the estimated/extracted source as the source means square error (SMSE):

$$\text{SMSE} = \text{E}\left\{(\hat{x}_1 - \alpha_{opt} x_1)^2\right\} \quad , \tag{10}$$

where $\alpha_{opt} = \text{E}\{\hat{x}_1 x_1\} / \text{E}\{x_1 x_1\}$ which compensates for possible changes in sign and/or amplitude. Recall (from table 1 and Eq. (9)) that the permutation ambiguity has been circumvented by extracting the source of interest at the first position. All simulations on the data have been fed to the JADE [3] algorithm as well, to compare the performance of the extraction algorithm to a full decomposition algorithm. To ensure a fair comparison for the ensemble learning based method and the extraction based AEMLe, we have limited the number of sweeps of the latter to $\max\{\lfloor \#flops_{JADE}(n) \rfloor, \text{convergence}\}$, with a minimum of 1 sweep.

Results on Binary Data. Taking 10^3 samples of binary data series composed of 0's and 1's with probability p and $1 - p$, respectively, we can simulate signals with a kurtosis value ranging from -2 to $\approx 10^3$ by simply altering the distribution parameter p of the Bernouilli distribution. The so obtained data series are used in the following experiments as the source signals $\boldsymbol{x}$, for which $\boldsymbol{x}_1$ has a distinct kurtosis sign with respect to $x_{i \neq 1}$, unless otherwise specified. In figure 1a, the results are shown for a fixed kurtosis value (ref) of the reference source, and a second source ranging from equal kurtosis until a kurtosis difference of 20. The objective to estimate the less kurtotic source is obtained by setting $(\epsilon_1, \epsilon_2) = (-1, 1)$. In figure 1b, the results are shown for the same kurtosis values, but with the objective to extract the most kurtotic source by sign reversion of the ϵ_i in (9). Unfortunately, the aforementioned source scenarios where the separation angle θ can be calculated algebraically are rather unrealistic. A more realistic scenario is given when the reference source has to be distinguished from a mixture with more than two sources and an equal number of observers ($m = n : m, n > 2$). In figure 2a&b the SMSE are presented in function of the number of sources in the mixture, both for a subgaussian and supergaussian source extraction scheme.

[5] Due to limited sample size. This was empirically observed (results not shown).

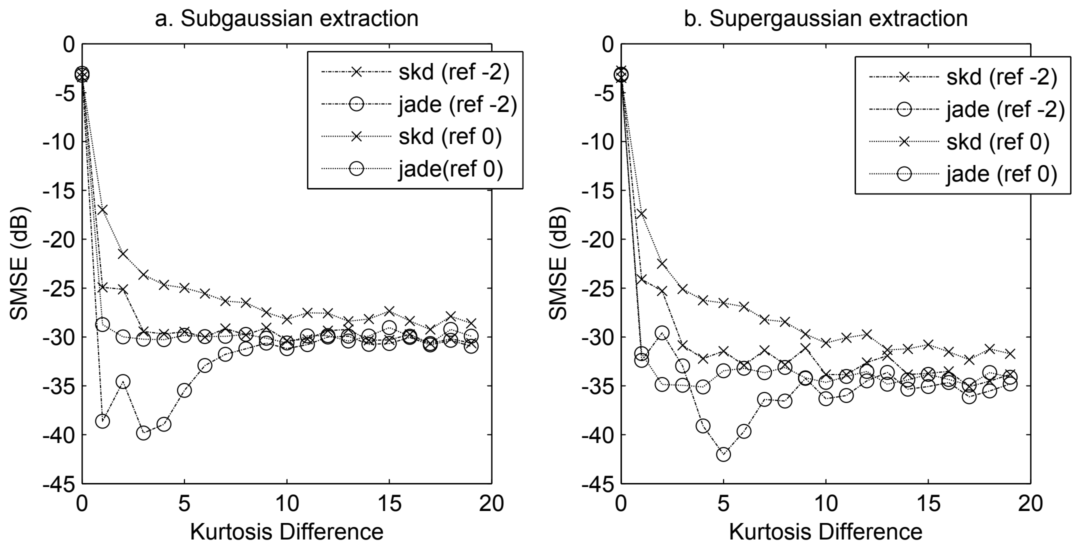

Fig. 1. SMSE for real sources in 2×2 mixtures based on a. gaussian or subgaussian ($\kappa_{1111} = -2$) reference b. gaussian or supergaussian reference

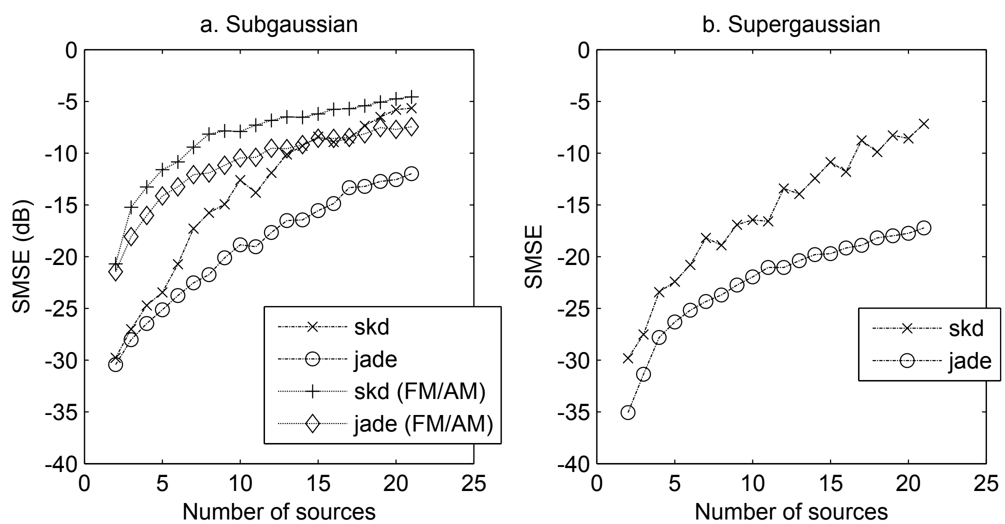

Fig. 2. SMSE for AEMLe in n×n mixtures based on a. subgaussian reference (binary or FM/AM modulated sources, see text) or b. a supergaussian reference

Sources with Nonstationary Statistics. A simulation set containing two simulation models is included, i.e. a model for the supergaussian source based on a tangent function, and a model for the subgaussian source which has been obtained by a sinusoid and its harmonics. By changing the function in the tangent's argument, we can easily alter the kurtosis value of the supergaussian sources giving us a possibility to mimic distinct scenarios. Both source models contain a frequency modulation as well as an amplitude modulation part to avoid stationarity of their statistics. The results of the dataset with extraction of the subgaussian source can be found in figure 2a.

Influence of Noise and Sample Size. Figures 3a and 3b show the effect of noise and sample size on the algorithm's average performance. The SNR was defined as $1/\sigma_\eta^2$ where σ_η^2 stands for the variance of η, temporally and spatially white gaussian noise. The denominator was set to one because the noise has been added to the unitary mixture of normalised (unit-variance) sources.

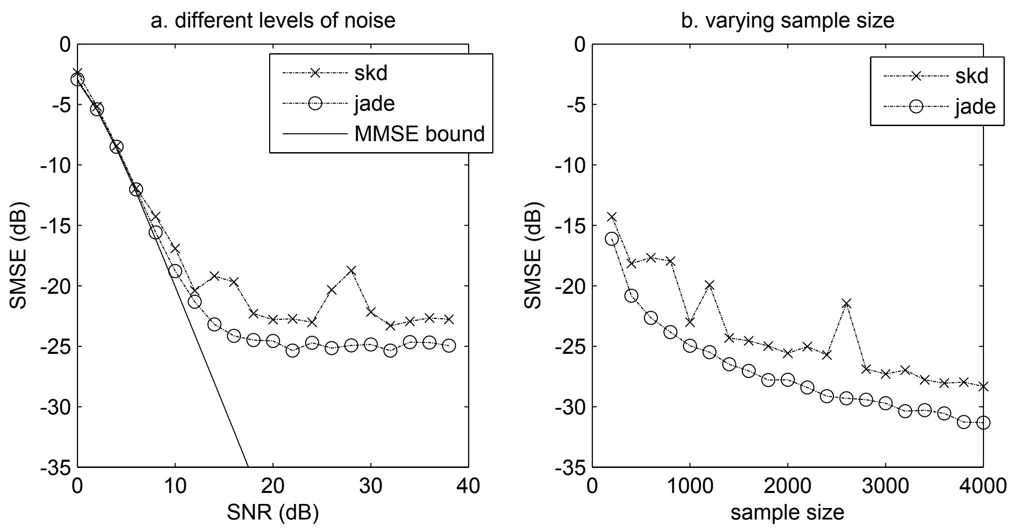

Fig. 3. The SMSE performance measure under influence of a. noise and b. sample size

Violation of the *a priori* Assumption. The results in figures 4a& b and 5a& b show the results for simulated datasets where all sources were sharing the same kurtosis sign, thus violating the basis assumption supporting Eq. (9). Since the case of negative kurtosis sign does not leave a lot of space for kurtotic variation, the kurtosis difference has been increased with steps of one twentieth, whereas the kurtotic difference for the supergaussian case was taken in steps of 1.

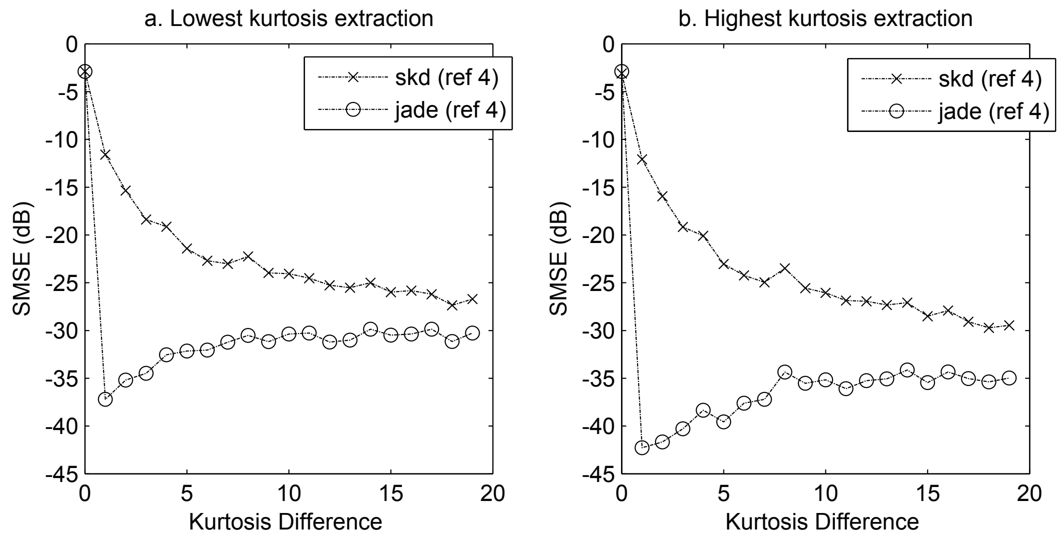

Fig. 4. The SMSE performance measure when all sources are supergaussian having the same kurtosis sign for the scenarios of extraction of a. the lowest kurtotic source and b. the highest kurtotic source.

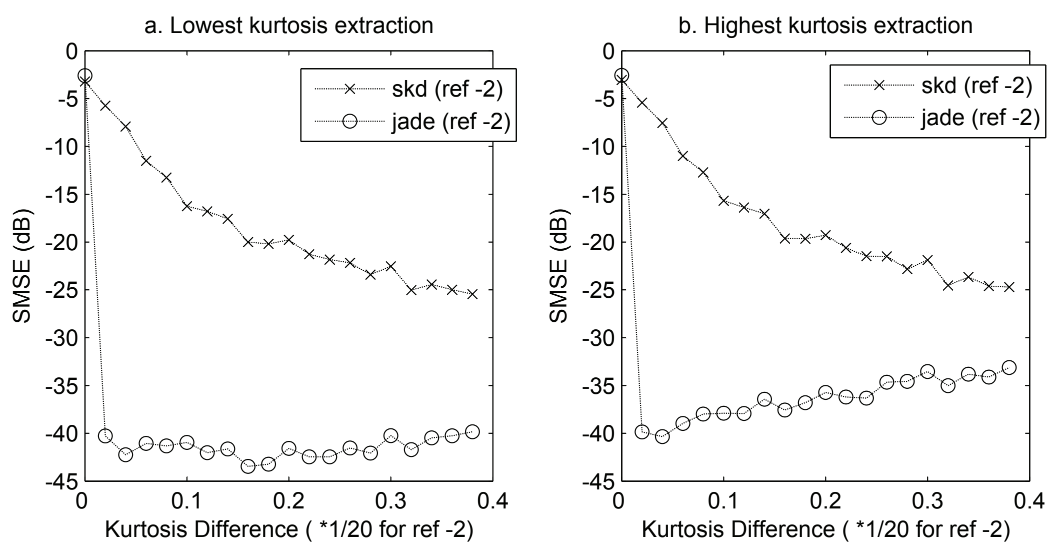

Fig. 5. The SMSE performance measure when all the sources are subgaussian having the same kurtosis sign for the scenarios of extraction of a. the lowest kurtotic source and b. the highest kurtotic source.

4 Discussion

It is clear from the results in figures 1 and 2 that the algorithm is more suited to extract the source with highest kurtosis value than vice versa, probably because more spread is allowed in the positive kurtosis values than in the negative. From the same figures it is clear that the algorithm is most suited for a considerably large kurtotic gap, although overall performance is quite good (-20dB is a mean sample error of $0.01 \times \sigma_{x_1}$).
Comparing the algorithm's performance to a full decomposition algorithm such as JADE, it is obvious that the performance deterioration of the proposed algorithm is negligible for a sufficiently large kurtosis difference (in the subgaussian signal extraction scenario 2dB at a kurtosis difference of 6, whereas 1dB at a difference of 8) and this with a substantially lower calculation effort since there is no need to decompose the signal in its full set of sources.
From figures (4)a& b and (5)a& b it can be deducted that the source extraction model for a 2×2 scenario still holds with comparable performance when the kurtosis sign of all sources is equal (i.e.$\epsilon_1 = \epsilon_2$), the source of interest either being the highest or lowest in kurtosis. These results suggest that the contrast may serve as well as the basis of a full decomposition algorithm with subspace deflation. The proof that Eq. (7) is a contrast no matter the values taken by ϵ_i under the sole condition that $\lambda_1 \neq \lambda_2$ can be found in Appendix B.
The low performance of the algorithms at a kurtotic difference of zero is due to the non identifiability of the sources based on their kurtosis. Extracting the source with lowest or highest kurtosis in the 2 sources/2 observers case will thus with a chance of 0.5 result in extraction of the wrong source, explaining the poor SMSE values for a source kurtosis difference of zero.

5 Conclusion

A new contrast has been proposed for ICA based on the kurtosis signs of the sources. Advantages are that the contrast is able to extract the source of interest, given that it has the lowest or highest kurtosis value in the mixture, with optimal results when the kurtosis sign of the source of interest is different from all other kurtosis signs of the sources in the mixture. Moreover it is presented in the framework of simple pairwise optimisation, a cost-effective closed form estimator in the two signal case. In the multichannel case this framework allows for extraction of the source of interest without having to estimate the whole set of sources.
The extraction of a single source having different kurtosis sign fits in the framework of ICA with known source kurtosis sign (appendix A), a generalisation of the AEML in [6] estimator in the 2 sources/2 observers case and already hinted at in [7] through an extension of the MaxKurt with kurtosis sign inclusion. It's also a generalisation of the contrast proposed in [8] based on the kurtosis sum of the sources. The algorithm's flexibility allows for easy be extensions e.g. by

putting constraints in the updating step. An accompanying paper describing the extraction of atrial activity from electrocardiogram signals shows promising results for addition of a spectral constraint in the updating cycle [9].

References

1. Hyvärinen, A., Oja, E.: A fast fixed-point algorithm for independent component analysis. Neur Comp **9** (1997) 1483–1492
2. Zarzoso, V., Comon, P.: How fast is fastICA? In: Proceedings of the 14th European Signal Processing Conference (EUSIPCO), Firenze, Italy (September 2006)
3. francois Cardoso, J., Souloumiac, A.: Blind beamforming for non-gaussian signals. IEE Proc.-F **140**(6) (1993) 362–370
4. Comon, P.: Independent component analysis, a new concept? Signal Processing **36** (1994) 287–314
5. Harroy, F., Lacoume, J.L.: Maximum likelihood estimators and cramer-rao bounds in source separation. Signal Processing **55**(2) (1996) 167–177
6. Zarzoso, V., Nandi, A.K., Hermann, F., Millet-Roig, J.: Combined estimation scheme for blind source separation with arbitrary source PDFs. IEE Electronics Letters **37**(2) (2001) 132–133
7. Cardoso, J.F.: High-order contrasts for independent component analysis. Neural Computation **11** (1999) 157–192
8. Comon, P., Moreau, E.: Improved contrast dedicated to blind separation in communications. In: Proc. ICASSP-97, 22nd IEEE International Conference on Acoustics, Speech and Signal Processing, Munich, Germany (Apr. 20–24, 1997) 3453–3456
9. Phlypo, R., Zarzoso, V., Comon, P., D'Asseler, Y., Lemahieu, I.: Extraction of atrial activity from the ECG by spectrally constrained kurtosis sign based ICA. In: submitted to ICA 2007. (2007)

Appendix A. Proof of contrast

Consider a set of N independent sources, s_n, $1 \le n \le N$, and denote κ_n their kurtosis. Now assume that the first p sources are known to have a positive kurtosis and the remaining $n-p$ a negative one:

$$\begin{cases} \kappa_i > 0, \forall i : 1 \le i \le p \\ \kappa_i < 0, \forall i : p < i \le N \end{cases} \tag{11}$$

Denote T_i the kurtosis of the separator output y_i. If data measurements have been spatially prewhitened, the output vector $\boldsymbol{y}$ is linked to the source vector $\boldsymbol{s}$ through a real orthogonal transform, $\boldsymbol{Q}$. Denote $\mathcal{S}$ the set of sources satisfying (11), and $\mathcal{Y}$ the set of observations generated by the orthogonal group $\mathcal{Q}$ acting on $\mathcal{S}$. Then with these notations we have the following result:

Proposition 1. *The optimization criterion $\Psi_p(\boldsymbol{Q})$ defined as:*

$$\Psi_p(\boldsymbol{Q}) = \sum_{i=1}^{N} \varepsilon_i T_i \tag{12}$$

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

Proof. By hypothesis, there exists an orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{y} = \boldsymbol{Q}\,\boldsymbol{s}$. Thus, by the multilinearity property of cumulants, we have $T_i = \sum_{j=1}^{N} Q_{ij}\kappa_j$. Hence:

$$\Psi_p(\boldsymbol{Q}) = \sum_{i=1}^{N} \varepsilon_i \sum_{j=1}^{N} Q_{ij}^4 \kappa_j$$

Now, by using the triangular inequality, we have

$$\Psi_p(\boldsymbol{Q}) \le \sum_{j=1}^{N} \sum_{i=1}^{N} |Q_{ij}|^4 \, |\kappa_j| \le \sum_{j=1}^{N} \sum_{i=1}^{N} |Q_{ij}|^2 \, |\kappa_j|$$

the second inequality stemming from the fact that the entries of an orthogonal matrix are of modulus smaller than or equal to one. Now, for any orthogonal matrix, $\sum_i |Q_{ij}|^2 = 1$, so that eventually

$$\Psi_p(\boldsymbol{Q}) \le \sum_{j=1}^{N} |\kappa_j| = \sum_{j=1}^{N} \varepsilon_j \kappa_j = \Psi_p(\boldsymbol{I})$$

This proves the domination. Now if the equality $\Psi_p(\boldsymbol{Q}) = \Psi_p(\boldsymbol{I})$ holds, we must have

$$\sum_{j=1}^{N} \sum_{i=1}^{N} \left[|Q_{ij}|^2 - |Q_{ij}|^4 \right] \, |\kappa_j| = 0$$

Yet, all the terms in the sums are positive, which means that they must all vanish. In other words $|Q_{ij}|^2 - |Q_{ij}|^4 = 0$, $\forall(i,j)$, which can occur only if $|Q_{ij}| \in \{0,1\}$. Because $\boldsymbol{Q}$ is orthogonal, there can be only one nonzero entry in every row and column, which means that $Q_{ij} = \lambda_i\, P_{ij}$, where $\lambda = \pm 1$ and $\boldsymbol{P}$ is a permutation. This proves the discrimination property.

Proposition 2. *Trivial filters associated with the contrast (12) are of the form*

$$\boldsymbol{Q} = \boldsymbol{\Lambda}\, \boldsymbol{P} \tag{13}$$

where $\boldsymbol{P}$ *is a matrix formed of two diagonal blocks of size* $p \times p$ *and* $N{-}p \times N{-}p$ *respectively, containing permutations.*

Proof. We have seen in Proposition 1 that trivial filters are the product of a diagonal matrix containing unit modulus entries and a permutation. So denote them $Q_{ij} = \lambda_i\, P_{ij}$. It remains to prove that matrix $\boldsymbol{P}$ is of the expected block form. Start with the equality (true for all trivial filters, by definition): $\Psi_p(\boldsymbol{Q}) = \Psi_p(\boldsymbol{I})$, and choose the equality below in the chain of equalities proved in Proposition 1:

$$\sum_{i=1}^{N} \varepsilon_i \sum_{j=1}^{N} Q_{ij}^4 \kappa_j = \sum_{j=1}^{N} \varepsilon_j \kappa_j$$

Because $\lambda_j = \pm 1$ and $\boldsymbol{P}$ is a permutation, we have $Q_{ij}^4 = P_{ij}$. Hence, since $\varepsilon_j^2 = 1$ and $\varepsilon_j \kappa_j = |\kappa_j|$, one gets:

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

Yet, every term in this sum is positive, so that they must individually vanish, which yields the relation:

$$\sum_{i=1}^{N} \varepsilon_i\, P_{ij}\, \varepsilon_j = 1,\ \forall j \tag{14}$$

By splitting the sum into two parts, we can replace ε_i by its value: $\sum_{i=1}^{p} P_{ij}\varepsilon_j - \sum_{i=p+1}^{N} P_{ij}\varepsilon_j = 1$. Now distinguish the cases $j \le p$ and $j > p$, and take into account the fact that, for any permutation, $\sum_i P_{ij} = 1$; we get:

$$\begin{cases} 1 - 2\sum_{i=1}^{p} P_{ij} = 1,\ \forall j > p, \\ 2\sum_{i=1}^{p} P_{ij} - 1 = 1,\ \forall j \le p \end{cases}$$

The first equality yields that for any $j > p$, $\sum_{i=1}^{p} P_{ij} = 0$, that is, by positivity, $P_{ij} = 0$. The top right block of $\boldsymbol{P}$ is thus null. Now the second equality yields $\sum_{i=1}^{p} P_{ij} = 1$ for any $j \le p$. As a consequence, every column in the first principal block contains a 1. But in a permutation, there can be only a single 1 in every column. The bottom left block must then be null. We have proved that the permutation matrix is indeed of the form:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_1 & 0 \\ 0 & \boldsymbol{P}_2 \end{pmatrix}$$

This second proposition shows that sources with positive kurtosis may be extracted separately from sources with negative kurtosis, provided that contrast (12) is utilized, and provided that p is known.

Appendix B. AEML: a contrast for $\kappa_{iiii} + \kappa_{jjjj} \neq 0$

From Eq. (7), we know that for an optimal angle of $\theta_\star$, the cross cumulants vanish. This can be deduced from Eq. (8), where we see that for the stationary point it holds that $(\kappa_{1222} + \kappa_{1112}) = 0$. We can now write our stationary point as:

$$\Psi(\theta_\star) = (\kappa_{1111} - \kappa_{2222}) \cos 2\theta_\star \ , \tag{15}$$

$$\dot{\Psi}(\theta_\star) = -2(\kappa_{1111} - \kappa_{2222}) \sin 2\theta_\star = 0 \ , \tag{16}$$

$$\ddot{\Psi}(\theta_\star) = -4(\kappa_{1111} - \kappa_{2222}) \cos 2\theta_\star < 0 \ . \tag{17}$$

From Eq. (16) we get the first restriction on the solution space for θ as $\theta_\star = n\frac{pi}{2}$ or $(\kappa_{1111} - \kappa_{2222}) = 0$. The latter condition does not set any restrictions on our value for θ in the optimum, thus making it impossible to obtain a solution.
For values of θ satisfying Eq. (16), it remains open whether the optimum is a minimum or a maximum. Since maxima are characterised by Eq. (17), we can rewrite this equation to constrain θ's solution space even further by rewriting $\cos 2\theta$ as a function of the solutions to Eq. (16). Eq. (17) then becomes:

$$\ddot{\Psi}(\theta_\star) = 4(\kappa_{1111} - \kappa_{2222})(-1)^n \ . \tag{18}$$

The solutions are now given as a function of n, and a maximum is reached for

$$n = \begin{cases} \text{odd} & \text{if } \kappa_{1111} > \kappa_{2222} \\ \text{even} & \text{if } \kappa_{2222} > \kappa_{1111} \end{cases} \tag{19}$$

Since we know that the signals are prewhitened and thus orthogonal, the solutions to Ψ show that their is a unique solution for independent signals at the maximum of the function, without scaling ambiguity (since all signals are prewhitened, their scale is fixed to 1) and even no permutation ambiguity as can be seen from Eq. (19). This concludes the proof for AEML being a contrast function for any pair $(\kappa_{1111}, \kappa_{2222})$ given $(\kappa_{1111} - \kappa_{2222}) \neq 0$. Remark that there has been made no assumption about the signs of both source kurtosis, making it a general contrast function for any $(\kappa_{1111}, \kappa_{2222})$ in the case of 2 sources/2 observers.