MLSP'2006 proceedings - 16th Machine Learning for Signal Processing conference
Maynooth (Ireland), 6-8 September 2006, IEEE, pp. 41-46

A LEAST ABSOLUTE BOUND APPROACH TO ICA — APPLICATION TO THE MLSP 2006 COMPETITION

John A. Lee, Frédéric Vrins, Michel Verleysen

Machine Learning Group, Microelectronics Laboratory
Université catholique de Louvain
Place du Levant, 3, B-1348 Louvain-la-Neuve

ABSTRACT

This paper describes a least absolute bound approach as a way to solve the ICA problems proposed in the 2006 MSLP competition. The least absolute bound is an ICA contrast closely related to the support width measure, which has been already studied for the blind extraction of bounded sources. By comparison, the least absolute bound applies to a broader class of sources, including those that are bounded on a single side only. This precisely corresponds to the sources involved in the competition. Practically, the minimization of the least absolute bound relies on a specific deflation algorithm with a loose orthogonality constraint. This allows solving large-scale problems without accumulating errors.

1. INTRODUCTION

The goal of the MLSP 2006 competition proposed by Andrzej Cichocki and Deniz Erdogmus consists in blindly recovering sources $\mathbf{s}(t) = [s_1(t), \dots, s_m(t)]^T$, knowing only mixtures of them. The latter can be written as $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$, where $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ are the observed mixtures and $\mathbf{A}$ is the unknown mixing matrix, which is assumed to be square ($m = n$).

The competition rules bring some a priori information about the sources. For all subproblems, sources are statistically independent and non-negative. Two kinds of sources can be generated: they are either sparse (super-gaussian) or uniformly distributed (sub-gaussian); see Fig. 1. More information regarding the sources and mixing matrices can be found on the competition website (`http://mlsp2006.conwiz.dk/`).

Practically, algorithms implementing Independent Component Analysis (ICA, [1]) can blindly recover the independent sources from the mixtures. In most algorithms, the separation is achieved iteratively by building a matrix $\mathbf{B}$ such that $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ is a 'good estimate' of $\mathbf{s}(t)$. Typically,

J.A.L. is a Scientific Research Worker of the Belgian Fund for Scientific Research (FNRS).

M.V. is a Research Director of the Belgian Fund of Scientific Research (FNRS).

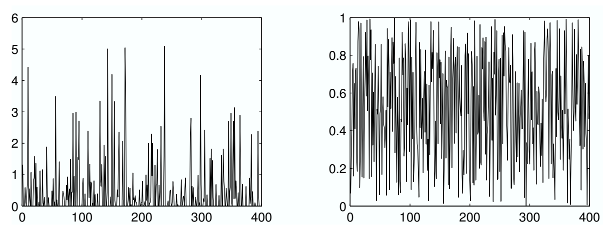

Fig. 1. Sources involved in the MLSP 2006 competition. Sources are statistically independent, either uniformly distributed or sparse. All sources are thus non-negative.

algorithms process prewhitened mixtures $\mathbf{z}(t) = \mathbf{V}\mathbf{x}(t)$, where matrix $\mathbf{V}$ is such that $E[\mathbf{z}(t)\,\mathbf{z}^T(t)] = \mathbf{I}$ and $\mathbf{VA}$ is orthogonal. After prewhitening, ICA reduces to finding an orthogonal matrix $\mathbf{W}$ and $\mathbf{y}(t) = \mathbf{W}\mathbf{z}(t)$ yields an estimate of the sources up to a scaling and permutation. As $\mathbf{B} = \mathbf{WV}$, the quality of both the whitening and ICA procedures are important, especially if the dimension is high.

In order to assess the quality of the source recovery, the competition resorts to the Signal-to-Interference Ratio (SIR), which involves the transfer matrix $\mathbf{C} = \mathbf{BA}$ and can be defined as follows:

$$\mathrm{SIR} = \frac{1}{n}\sum_{i=1}^{n} 10\log_{10}\frac{\max_j c_{ij}^2}{\sum_j c_{ij}^2 - \max_j c_{ij}^2} \quad . \qquad (1)$$

Within the framework of the competition, the SIR is used in a Monte Carlo-process. The SIR higher should be higher than 15dB for at least 90% of the runs, i.e. $P_{90} > 15$dB, where P_{90} is the 90th percentile. Four sub-problems must be solved:

1. Large scale problem: fixed sample size ($N = 5000$), increasing number of mixtures ($n > 50$), random mixing matrix.
2. Small training set problem: fixed number of mixtures ($n = 50$), decreasing sample size ($N < 5000$), random mixing matrix.

3. Highly ill-conditioned problem: $N = 5000$, $n > 1$; for this subproblem, the mixing matrix is a Hilbert matrix multiplied by a random Givens matrix.

4. Noisy mixtures problem: $n = 50$, $N = 1000$, white noise with increasing variance corrupts the mixtures.

The remainder of this paper is organized as follows. Section 2 presents the objective function used to solve the competition problems. Next, Section 3 describes an optimization procedure specifically tailored for this function. Section 4 details the results obtained for the mixtures of the competition. Finally, Section 5 draws the conclusions.

2. PROPOSED ICA CONTRAST FUNCTION

Many general-purpose ICA algorithms offer an appealing tradeoff between performances and speed. However, available a priori knowledge about the sources often incites one to design a specific contrast, fitted to the source properties. For instance, objective functions exploiting sparsity, non-negativity or finite measure support of the sources have been proposed in the literature. Any a priori information about sources may help to improve the separation algorithms, from at least 4 viewpoints:

- Speed and convergence rate.
- Separation quality in terms of a performance index such as the SIR.
- Relaxation of some assumption in the model (such as the source independence or the condition $m \leq n$).
- Derivation of a cost function with more interesting properties (e.g. discriminacy).

For instance, if the sources are bounded, an algorithm based on the Minimum Support Width measure (SWM) can address the three last points, depending on the context [2, 3, 4, 5, 6]. The algorithm described in [7] can extract sources one by one (using a deflation approach [8, 6]) by minimizing $\mathcal{C}_\Omega(y_i) \doteq \sup y_i - \inf y_i$.

2.1. Least absolute bound approach

A variant of the SWM contrast is proposed for extracting the sources involved in the competition problems. About half these sources are uniformly distributed (and thus double-bounded: $a < \inf s_i < \sup s_i < b$, $|a| < \infty, |b| < \infty$), whereas the other half is bounded only on one side ($a < \inf s_i$, $\sup s_i = \infty$; here $a = 0$). In this second case, minimizing the SWM is meaningless and inappropriate. However, a similar reasoning can be derived, based on the milder assumption that the sources are at least bounded on one side. As sources can only be recovered up to a sign change, this

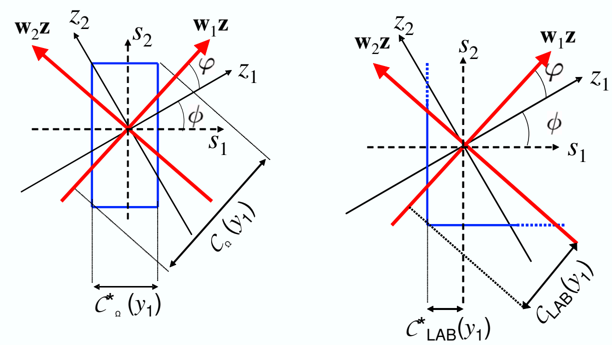

Fig. 2. Graphical interpretation of the SWM (left, $\mathcal{C}_\Omega$) and LAB (right, $\mathcal{C}_{\text{LAB}}$); the minimum values of the contrasts are labelled '$\star$'. The solid lines represents the support boundaries and the bold solid arrows represent the current axes.

leaves two solutions: one can either maximize the infimum of the whitened mixtures or minimize the supremum, what can be related to Erdogan's approach [9] (minimization of the supremum for symmetric bounded signals). These two possibilities can be merged into a single objective function, namely the Least Absolute Bound (LAB):

$$\mathcal{C}_{\text{LAB}}(y_i) \doteq \min\{-\inf y_i, \sup y_i\} \ , \tag{2}$$

which has to be minimized: $\mathbf{w}_i^\star = \min_{\mathbf{w}_i} \mathcal{C}_{\text{LAB}}(\mathbf{w}_i\mathbf{z})$. Both the SWM and LAB can be written as a weighted sum of positive quantities only depending on the sources. In both cases, the weights are the absolute values of the elements of the transfer vector between $\mathbf{s}$ and y_i, i.e. the elements of the i-th row of $\mathbf{C}$; the positive 'source quantities' are the widths of the support convex hulls for the SWM approach, and either the suprema or minus the infima of the sources for the LAB approach. Then, both are equivalent and benefit from similar contrast properties (see Fig. 2).

2.2. Practical estimation

In practice, the output infimum and supremum must be estimated; obviously, an estimator based on the minimum and maximum observed values would be highly sensitive to additive noise and outliers. Therefore averaged order statistics are preferred, like advised in [2] for estimating the SWM. Assuming that y'_i denotes the sorted ith output, the contrast estimator can be written as

$$\hat{\mathcal{C}}_{\text{LAB}}(y_i) \doteq \min\left\{-\frac{1}{q}\sum_{k=1}^{q} y'_i(k), \frac{1}{q}\sum_{k=1}^{q} y'_i(N{+}1{-}k)\right\} \ ,$$

where $y'_i(k)$ is the kth lowest value of $y_i(t)$, $y'_i(N{+}1{-}k)$ the kth highest one and q is an integer between 1 and $\lfloor N/2 \rfloor$. This requires that the sample size N is large enough so that the accuracy of the above estimator is sufficient. Indeed, it can be observed that $\inf y_i = \sum_{j=1}^{m} w_{ij} \inf s_j$, and $\inf y_i$ has thus to be estimated from the sample by $\min_t y_i(t)$, and likewise for the supremum. Clearly, as it holds that $y_i(t) = \sum_{j=1}^{m} w_{ij} s_j(t)$, the last estimation is reliable only if it exists $t^\star$ such that $1 \leq t^\star \leq N$ and $s_j(t^\star) \simeq \inf s_j$ for all $1 \leq j \leq m$, that is when the sample size is large enough, because the sources are assumed to be statistically independent. Under the same condition, $\frac{1}{q}\sum_{k=1}^{q} y_i'(k) \simeq \min_t y_i(t)$ if $y_i'(l) \simeq \min_t y_i(t)$ for all $1 \leq l \leq q$, and likewise for $\frac{1}{q}\sum_{k=1}^{q} y_i'(N+1-k)$.

In the noise-free case, q can be taken close to one; otherwise, q must be slightly increased (see [3] for a discussion of q in the context of the minimum support approach). For all competition problems, q takes on the value $\max\{20, N/50\}$, except for the third one ($q = N/200$).

The estimation of $\hat{\mathcal{C}}_{\text{LAB}}$ involves a sort operation, whose complexity ($\mathcal{O}(N \log N)$) is higher than for more traditional contrasts such as the kurtosis (typically $\mathcal{O}(N)$).

3. OPTIMIZATION SCHEME

As function $\hat{\mathcal{C}}_{\text{LAB}}$ is not everywhere differentiable [6], a specific optimization method is needed. A deflation algorithm that is able to optimize non-differentiable contrasts has been proposed in [7]. Instead of using the derivative of the contrast and a standard gradient ascent or a fixed-point update rule, this algorithm relies on a trial-and-error geodesic search on the manifold of orthogonal matrices. This amounts to updating separation matrix $\mathbf{W}$ by rotating pairs of rows, using Givens matrices; this keeps $\mathbf{W}$ always orthogonal. Unfortunately, although this algorithm shows interesting properties (e.g. monotonic contrast increase), it is only suited for small-sized problem. Indeed, it involves about $2n + 1$ contrast evaluations per iteration and per source, in order to explore thoroughly the contrast landscape. In comparison, FastICA requires only a single evaluation of the contrast gradient. For the MLSP competition, a faster algorithm was clearly required, as proposed hereafter.

Although the LAB and SWM are not differentiable contrasts, a direction that optimizes the contrast can easily be guessed in order to replace the gradient vector. Assume for the sake of simplicity that mixtures involve only bounded sources, e.g. uniform ones, and are whitened. Under this hypothesis, solving the ICA problem amounts to finding an orthogonal matrix $\mathbf{W}$, which can be interpreted geometrically as a rotation. In other words, the joint support of the whitened mixtures looks like a rotated rectangular hyperparallelipiped. Then ICA precisely consists in finding an orthonormal basis such that the hyperparallelipided edges are parallel with the basis axes. Looking at row $\mathbf{w}_i$ of $\mathbf{W}$, computing the infimum and supremum involves the ith ICA output $y_i(t) = \mathbf{w}_i \mathbf{z}(t)$, which is the projection of the

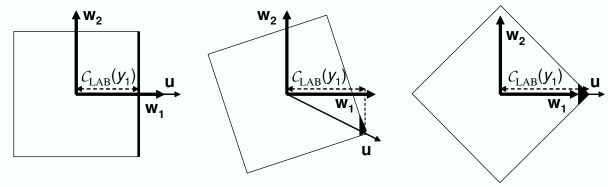

Fig. 3. Principle of the proposed algorithm in the 2D case: (left) LAB minimum; (middle) intermediate situation; (right) LAB maximum. In those three examples, vector $\mathbf{u}$ points towards sample points that determine the value of $\hat{\mathcal{C}}_{\text{LAB}}(\mathbf{w}_1\mathbf{z})$ (bold black line or black triangle). This provides information about how to update $\mathbf{w}_1$: if $\mathbf{u} \neq \mathbf{w}_1$, then rotate $\mathbf{w}_1$ away from $\mathbf{u}$.

whitened sample on the $\mathbf{w}_i$ axis. If the infimum and supremum are estimated using averaged order statistics, only q projected observations are used to approximate both ends of the marginal support. At this stage, let us focus on those q observations at one end of the support and before projection on $\mathbf{w}_i$: averaging them and normalizing the resulting vector gives us a direction $\mathbf{u}$ whose interpretation is as follows. If $\mathbf{w}_i$ is a correct solution of the ICA problem, as on the left of Fig. 3, then the q observations are located on (or near) a face of the hyper-parallelipiped that is perpendicular to $\mathbf{w}_i$; in this setting, $\mathbf{u} \approx \mathbf{w}_i$. On the other hand, if $\mathbf{w}_i$ is not a correct solution of the ICA problem, as in the middle of Fig. 3, then the q observations concentrate in a corner of the hyperparallelipiped; consequently, $\mathbf{u}$ points to that corner. Finally, the right plot of Fig. 3 shows a configuration for which $\mathbf{w}_i \approx \mathbf{u}$, though $\mathbf{w}_i$ is not a good solution. These three configurations correspond respectively to a minimum, an intermediate value and a maximum of the infimum and suggest using $\mathbf{u}$ in a similar way as a gradient vector. As long as the algorithm has not converged, vector $\mathbf{u}$ indicates the direction of the closest corner of the support (w.r.t. direction $\mathbf{w}_i$). Hence $\mathbf{u}$ gives a good estimate of the direction where $\mathbf{w}_i$ may not go. Rotating $\mathbf{w}_i$ 'away' from $\mathbf{u}$ proves to be experimentally a relatively good guess towards the solution: doing so decreases the LAB, at least while no other corner gains the upper hand. Indeed, when approaching the solution, two or more corners struggle for the upper hand from one iteration to the other and the direction indicated by $\mathbf{u}$ might thus change accordingly, in a 'chaotic' way. This justifies the use of an update angle that does not depend on (the angle between $\mathbf{w}_i$ and) $\mathbf{u}$. It is scheduled to start from $\pi/4$ and ulterior increases/decreases depend on the LAB changes (decreases/stagnation). A not too fast decay of the update angle, though time consuming, enables the algorithm to converge slowly but surely on the best solution.

The previous ideas can be implemented through a simple deflation procedure, listed in Fig. 4. As this algorithm is intended to work on 'difficult' problems (large number of mixtures, low number of observations, ill-conditioned mixing matrix), it is expected that whitening results can be inaccurate. In order to be robust against ill-conditioned mixtures, the whitening stage relies on a singular value decomposition (SVD) of the centered mixtures instead of an eigenvalue decomposition (EVD) of the covariance matrix. Additionally, possible badly whitened signals z_i (i.e. those having a variance lower than one and/or nonzero covariances) are discarded from vector $\mathbf{z}$. These signals can perturb the subsequent separation step (the infimum and/or supremum abnormally vanishes in some directions) and jeopardize the recovery of all sources. Hence, assuming without loss of generality that those signals are the p last ones, $\mathbf{z}$ is updated according to $\mathbf{z}(t) \leftarrow [z_1(t), \cdots, z_{n-p}(t)]^T$. Next, the separation algorithm is run on this reduced vector, yielding $n-p$ estimated sources. Finally, those imperfectly whitened mixtures are simply put back as they are in the ICA output: $\mathbf{y}(t) \leftarrow [y_1(t), \cdots, y_{n-p}(t), z_{n-p+1}(t), \cdots, z_n(t)]^T$.

Low sample sizes or large numbers of mixtures can also impair the orthogonality property of $\mathbf{W}$: whitening reduces the *sample* covariance to the identity matrix. In difficult problems, a small discrepancy between the sample covariance and the true covariance can jeopardize the source extraction after whitening. In such cases, small departures from orthogonality allows $\mathbf{W}$ to reach better contrast values and compensate for whitening errors. This justifies the loose orthogonalization of $\mathbf{W}$ in the algorithm. The algorithm merely checks that $\mathbf{w}_i$, the current row of $\mathbf{W}$ is not converging on any previously found solution; if this happens, $\mathbf{w}_i$ is made orthogonal to all previous rows. This approximate orthogonolization also prevents the accumulation of errors, which is typically observed in deflation algorithms.

$[\mathbf{W}, \mathbf{V}] = \text{LABICA}(\mathbf{x}(t))$

1. Whiten the mixtures using a singular value decomposition:
   (a) Center the sample $\mathbf{X} \doteq [\mathbf{x}(t)]$ by removing its mean: $\mathbf{x}(t) \leftarrow \mathbf{x}(t) - \frac{1}{N}\Sigma_t \mathbf{x}(t)$.
   (b) Compute the SVD of the centered sample: $\mathbf{X}^T = \mathbf{U}\mathbf{S}\mathbf{V}^T$.
   (c) Compute $\mathbf{Z} \doteq [\mathbf{z}(t)]$ directly: $\mathbf{Z} = \sqrt{N}\,\mathbf{U}^T$.
   (Depending on the convention, $\mathbf{U}$ is either $N{\times}n$ or $N{\times}N$; in the latter case, keep only the n first columns of $\mathbf{U}$.)
2. Discard the p incorrectly whitened mixtures (i.e. rows $z_i(t)$ having a variance lower than one and/or nonzero covariances).
3. Compute the radial projection of $\mathbf{z}$ on the unit sphere: $\mathbf{z}^\circ(t) = \frac{\mathbf{z}(t)}{\|\mathbf{z}(t)\|}$ for $1 \leq t \leq N$.
4. To extract the ith source, with $1 \leq i \leq n-p$, do:
   (a) Initialize $\mathbf{w}_i$ to any random direction and the update angle α to $\pi/4$.
   (b) Check loose orthogonality: if for some $j < i$ the inequality $|\mathbf{w}_i \mathbf{w}_j^T| < \cos(\pi/12)$ holds then make $\mathbf{w}_i$ orthogonal to all $\mathbf{w}_j$: $\mathbf{w}_i \leftarrow \mathbf{w}_i - \Sigma_j \mathbf{w}_i \mathbf{w}_j^T \mathbf{w}_j$; $\mathbf{w}_i \leftarrow \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}$.
   (c) Compute ith ICA output: $y_i(t) = \mathbf{w}_i \mathbf{z}(t)$ for $1 \leq t \leq N$.
   (d) Estimate the LAB of $y_i(t)$ using mean order statistics:
      - Determine the indexes of the q lowest and q highest values of $y_i(t)$.
      - Average the two corresponding sets of values to obtain the infimum and supremum of $y_i(t)$; keep their minimum absolute value as in (2.2) to obtain $\hat{\mathcal{C}}_{\text{LAB}}(y_i)$.
      - Use the same indexes to compute the direction $\mathbf{u}$ as the average of the corresponding columns of $\mathbf{Z}^\circ$.
      - If $\mathbf{u} \neq \mathbf{w}_i$, make $\mathbf{u}$ orthogonal to $\mathbf{w}_i$ and normalize it: $\mathbf{u} \leftarrow \mathbf{u} - \mathbf{u}\mathbf{w}_i^T \mathbf{w}_i$; $\mathbf{u} \leftarrow \frac{\mathbf{u}}{\|\mathbf{u}\|}$.
   (e) Update $\mathbf{w}_i$ and α:
      - Compute $\mathbf{w}_i' = \cos(\alpha)\mathbf{w}_i - \sin(\alpha)\mathbf{u}$ and $\hat{\mathcal{C}}_{\text{LAB}}(\mathbf{w}_i'\mathbf{z})$ (see step (d) above).
      - If $\hat{\mathcal{C}}_{\text{LAB}}(\mathbf{w}_i'\mathbf{z}) < \hat{\mathcal{C}}_{\text{LAB}}(y_i)$, then let $\alpha \leftarrow 1.01\alpha$ and $\mathbf{w}_i \leftarrow \mathbf{w}_i'$, else $\alpha \leftarrow \alpha/1.2$.
   (f) Go back to step 4(b) if convergence is not attained.
5. Append the p incorrectly whitened mixtures to the extracted sources: $\forall i > n-p, \quad y_i(t) \leftarrow z_i(t)$.

Fig. 4. LABICA: ad hoc deflation procedure to minimize $\hat{\mathcal{C}}_{\text{LAB}}$. After robust SVD-based whitening, sources are extracted one-by-one, with a loose orthogonality constraint preventing error accumulation. The gradient is replaced with contrast-dependent information: the closest support corner direction.

4. RESULTS

4.1. Large-scale problem

In this first subproblem, the sample size is fixed ($N = 5000$) and the number of sources/mixtures is growing ($n > 50$). The proposed algorithm solves it for a quite large number of mixtures. Graphical results in Fig. 5 show that outstanding SIR values are attained for more than 400 mixtures (P_{90} is still higher than 30dB). Processing so many mixtures obviously requires long computation times, even with the fastest algorithms (e.g. FastICA), and justifies the restriction to only 20 Monte-Carlo runs.

4.2. Small training set problem

In this second problem, the number of sources is kept constant ($n = 50$) but the sample size N varies. The results are shown in Fig. 6 for two algorithms: the proposed one and

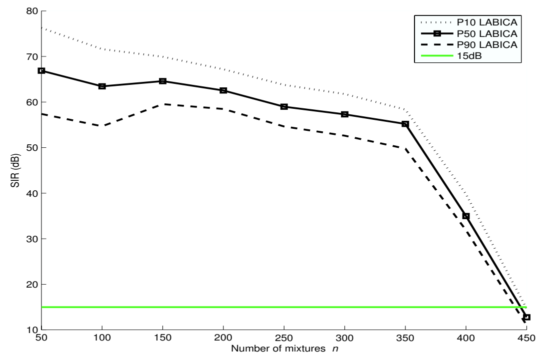

Fig. 5. Results for subproblem 1: SIR performances vs number of sources n for 20 Monte-Carlo runs of LABICA and FastICA with 5000 sample points. $P_{90} > 15$dB holds for more than 300 sources.

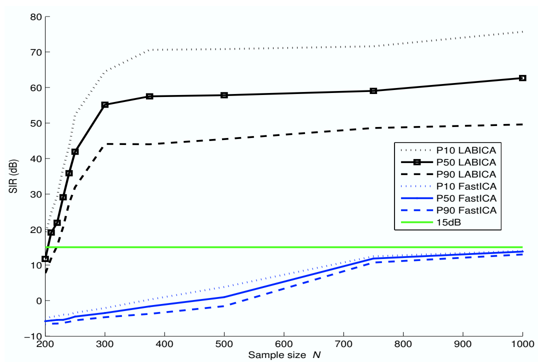

Fig. 6. Results for subproblem 2: SIR performances vs sample size N for 100 Monte-Carlo runs of LABICA and FastICA with 50 sources. Less than 250 observations are needed to achieve $P_{90} > 15$dB.

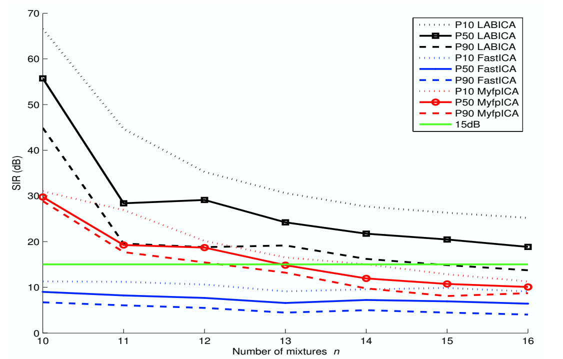

Fig. 7. Results for subproblem 3: SIR performances vs the number of sources n for 100 Monte-Carlo runs of LABICA ($q = N/200$), FastICA and MyfpICA with 5000 sample points. $P_{90} > 15$dB holds up to 14 sources.

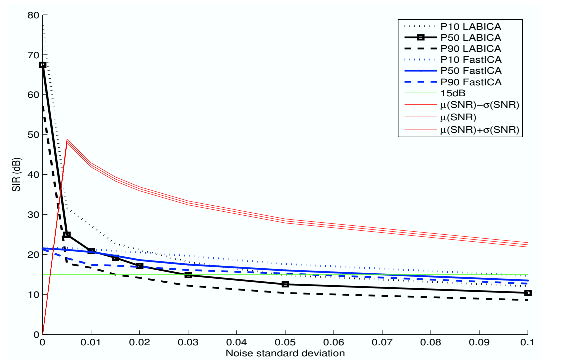

Fig. 8. Results for subproblem 4: SIR performances vs the noise standard deviation, for 100 Monte-Carlo runs of LABICA and FastICA with 5000 sample points and 50 sources; the corresponding SNR curve is plotted alongside.

FastICA (official version 2.5, with 'gaus' nonlinearity and fine tuning enabled). As can be seen, less than 250 sample points are required to achieve a SIR greater than 15dB in 90% of the cases.

4.3. Highly ill-conditioned problem

In this third subproblem, the mixing matrix is the product of a Hilbert matrix with a random Givens matrix. Hence, as the number of mixtures is growing, the separation problem gets more and more ill-conditioned. The results are shown in Fig. 7 for three algorithms: the proposed one, FastICA (as above) and a 'hacked' version of FastICA. The latter, called MyfpICA, works with a SVD-based whitening stage and a kurtosis-driven nonlinearity (either 'kurt' or 'gaus' depending on the kurtosis). In this subproblem, achieving a correct whitening is the main difficulty. The proposed algorithm brings a significant performance gain by using the SVD of the centered sample instead of the EVD of the sample covariance matrix. However, beyond 10 mixtures in this problem, the determinant of the mixing matrix $\mathbf{A}$ is so close to zero that no more than 10 mixtures can be whitened properly, even with the SVD. It has been experimentally observed that additional mixtures after whitening are actually not white; some of them may be correlated and/or have a variance lower than one. In this situation, the trick consists in temporarily discarding these still correlated mixtures after whitening, as proposed in Section 3, so that the separation algorithm can run in good conditions.

4.4. Noisy mixtures problem

The value of estimator $\hat{\mathcal{C}}_{\text{LAB}}(y_i)$ relies on a few sample points only, namely on q sample points with $q \ll N$. Consequently the proposed approach is expected not to be very robust against noise and outliers, especially with low values of q. As can be seen in Fig. 8, the quality of the results is rapidly decreasing as the noise variance is growing.

5. CONCLUSION

This note has described an ICA algorithm able to perform efficiently on at least three subproblems of 2006 MLSP competition proposed by A. Cichocki and D. Erdogmus. The algorithm relies on the support width measure, which has been proved to be an efficient contrast for bounded sources. SWM is not a 'general purpose' contrast, i.e. it can only be used for recovering double-bounded sources. As a matter of fact, not all sources involved in the competition are double-bounded: on average, half of them are. Nevertheless, it is noteworthy that all sources are non-negative. Unfortunately non-negativity does not hold for the estimated sources, because mixing coefficients can be negative. In this case however, non-negativity still reduces to an interesting property, even after centering of the mixtures: the marginal support of the estimated sources has at least either an infimum or a supremum. This property can be used to compute the least absolute bound of the estimated source support, which proves to be a discriminant contrast for the considered sources.

As this contrast is not differentiable, a specifically designed optimization procedure is proposed. It works in a similar way as a gradient descent, except that the contrast gradient is replaced in the update rule with an ad hoc guess. This allows the proposed algorithm to be quite competitive in terms of speed: the computational cost of a single update is low, as for a fixed-point algorithm, though the latter requires less iterations and converges much faster.

Next, the proposed algorithm relaxes the orthogonality constraint on the ICA separating matrix, though mixtures are prewhitened. The algorithm only checks that all source estimates are distinct. The main advantage of this milder constraint is that a deflation approach can be used without accumulating errors in the successive source estimates.

Finally, performance improvement can also stem from optimizations in the prewhitening step. While this operation generally involves an EVD of the sample covariance, better results are obtained with an SVD of the centered sample. Moreover, in the case of an ill-conditioned mixing matrix, it is also advised to discard uncorrectly whitened signals; these can jeopardize the subsequent ICA algorithm and are simply appended as they are to the reduced ICA solution.

6. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley Series on Adaptive Learning Systems for Signal Processing, Communications, and Control. Wiley and Sons, Inc., New York, 2001.

[2] F. Vrins and M. Verleysen, "Minimum support ICA using order statistics. Part I: Quasi-range based support estimation," in *Proc. Intl. Conf. on Independent Component Analysis and Blind Signal Separation (ICA 2006)*, J. Rosca et al., Ed., Charleston, SC, Mar. 2006, Lecture Notes in Computer Science (LNCS 3889), pp. 262–269, Springer.

[3] F. Vrins and M. Verleysen, "Minimum support ICA using order statistics. Part II: Performance analysis," in *Proc. Intl. Conf. Independent Component Analysis and Blind Signal Separation (ICA 2006)*, J. Rosca et al., Ed., Charleston, SC, Mar. 2006, Lecture Notes in Computer Science (LNCS 3889), pp. 270–277, Springer.

[4] F. Vrins, J.A. Lee, and M. Verleysen, "Filtering-free blind separation of correlated images," in *Proc. Intl. Work-conference on Artificial Neural Networks (IWANN): Computational Intelligence and Bioinspired Systems*, A. Prieto J. Cabestany and F. Sandoval, Eds. 2005, Lecture Notes in Computer Science (LNCS 3512), pp. 1091–1099, Springer.

[5] J.A. Lee, F. Vrins, and M. Verleysen, "Non-orthogonal support width ICA," in *Proc. European Symposium on Artificial Neural Networks (ESANN)*, M. Verleysen, Ed., Bruges (Belgium), Apr. 2006, pp. 351–358.

[6] F. Vrins, C. Jutten, and M. Verleysen, "SWM : A Class of Convex Contrasts for Source Separation," in *Proc. Intl. Conf. on Acoustics Speech and Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, pp. V.161–V.164, IEEE.

[7] J.A. Lee, F. Vrins, and M. Verleysen, "A simple ICA algorithm for non-differentiable contrasts," in *Proc. European Signal Processing Conference (EUSIPCO)*, Antalya (Turkey), Sept. 2005, pp. cr1412.1–4.

[8] S. Cruces and I. Duran, "The minimum support criterion for blind source extraction: a limiting case of the strengthened Young's inequality," in *Proc. Intl. Conf. Independent Component Analysis and Blind Signal Separation (ICA 2004)*, C.G. Puntonet and A. Prieto, Eds., Granada (Spain), Sept. 2004, Lecture Notes in Computer Science (LNCS 3195), pp. 57–64, Springer.

[9] A.T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 438–449, 2006.