Separation Capability of Overcomplete ICA Approaches

Markus Borschbach and Imke Hahn
University of Münster
Dept. of Mathematics and Natural Science
Einsteinstr. 62, D-48149 Münster
GERMANY

Abstract: The Relaxation to have only square matrices in standard ICA leads to an approximation of the inverse of non-quadratic matrices to determine the separation matrix. Synthetic data sets as well as speech data are used to compare the capability of such approaches, called overcomplete ICA on an underdetermined basis. Due to the fact that the mixing matrices are not invertible (because they are not square), the quality of the sources' reconstruction is not excellent. The most extreme case of an undetermined ICA is single channel ICA. But in this paper not the reduction to one sensor is considered but in a maximum case the reduction from eight sensors to two sensor signals. It is shown, which separation quality can still be achieved for the blind separation of the underlying sources. For an improved classification the algorithms are also compared to well-known standard ICA-algorithms.

Key–Words: Overcomplete ICA, Evaluation, Inlier-based ICA, Geometric ICA, Mean Field ICA.

1 Introduction

Independent component analysis (ICA) is a promising blind signal separation technique [4, 7, 1]. There are many different solutions for ICA in the standard case with an equal number of sources and sensors. A more difficult task is to find the unknown mixing matrix and especially the reconstruction of the sources in the case of overcomplete ICA (also referred to as underdetermined ICA) in which there are less sensors than sources available. There are some well-known analyses of algorithms for standard ICA, but so far there is no overview and evaluation of algorithms for overcomplete ICA. This paper deals with the comparison of different approaches in that special case. Points of interest are the analysis of the algorithms for a decreasing number of sensors, a decreasing number of samples and a rising noise level. Every aspect is tested in a Monte Carlo run simulation and graphically illustrated. A selection of the results is presented. The algorithms are compared in the overcomplete and standard ICA case with speech data and synthetic data with subgaussian and supergaussian distributions. Their performance is measured by the signal-to-interference-ratio (SIR) and other performance indices, see section 4.2. In overcomplete ICA it is often possible to estimate the mixing matrix whereas it is very difficult to obtain the estimated sources as the matrix is not quadratic and thereby not invertible. Only in special cases, such as very sparse data, this problem is solvable. Most algorithms for overcomplete ICA are restricted to supergaussian (or even sparse) data and do not achieve satisfying results in other cases. Therefore a Haar-wavelet-transformation is used to produce sparsity in the data.

Section 2. gives a short introduction to the model of overcomplete ICA. Three different approaches to overcomplete ICA are briefly discussed in section 3. In section 4. the test design is described and section 5. shows some of the results.

2 The overcomplete ICA Model

We consider the following model with the unknown, statistical independent original signals $\mathbf{s} = (s_1, \dots, s_m)$, the recorded signals $\mathbf{x} = (x_1, \dots, x_n)$ and the unknown linear mixing matrix $\mathbf{A}$ with linear independent columns (that is full rank),

$$\mathbf{x} = \mathbf{A}\mathbf{s} \text{ with } \mathbf{x} \in \mathbb{R}^{n \times T}, \quad \mathbf{s} \in \mathbb{R}^{m \times T}, \mathbf{A} \in \mathbb{R}^{n \times m}, \tag{1}$$

with T samples and n (recorded) respectively m (original) signals, $m > n$, see [1]. The task of ICA is to estimate the unknown matrix $\mathbf{A}$ and the sources $\mathbf{s}$. In the standard model of ICA, $m = n$, the sources $\mathbf{s}$ can be recovered from the mixing matrix $\mathbf{A}$ as $\mathbf{A}$ is invertible[1]. In overcomplete ICA this is different. The mixing matrix is not invertible and thus it is not enough to find the unmixing-matrix like in standard ICA.

[1] The sources can only be recovered up to a scaling and permutation factor.

Therefore most overcomplete-ICA-algorithms consist of two steps. First the mixing matrix is estimated, second the sources are reconstructed. The second step can be performed by the use of the Moore-Penrose-Pseudo-Inverse or by L1-norm-minimization. We compare the effect of the two approaches to the reconstruction's quality. The quality of each algorithm depends on the algorithm's parameters, the performed steps of preprocessing and the added level of noise. Our testsuite allows a comparison under self-definable testing conditions [2].

3 Algorithms

In this paper three different approaches for overcomplete ICA are examined. They are briefly described below.

3.1 Geometric ICA

Geometric ICA (geoica) by Fabian Theis et al., University of Regensburg, 2003, see [3]. Theis et al. have generalized a geometric algorithm for standard ICA to solve overcomplete ICA as well. For stability of the geometric algorithm it is required that the sources are supergaussian and unimodal. It is assumed that the data is normalized on the unit sphere. Given the recorded data vector $\mathbf{x}$ the aim is to determine the weights of the demixing matrix $\mathbf{W}$ with $\hat{\mathbf{s}} = \mathbf{W}\mathbf{x}$. Therefore $2n$ starting points are picked on the unit sphere $S^{m-1} \in \mathbb{R}^m$. According to a chosen learning rate these elements are moved towards the ICA directions taking into account a zero-neighbourhood-function and an *absolute winner-takes-all learning*. For detailed information see [3]. The sources are recovered in a second step by a maximum likelihood approach.

3.2 Inlier-Based ICA

Inlier-based ICA (ibica) by Harmeling, Meinecke and Müller, Fraunhofer FIRST.IDA, Berlin, 2004, refer to [4], [5]. Ibica also is a geometric algorithm. It is extended by the use of an outlier index to directly find the ICA-directions. After projecting the data on the unit sphere the data points are sorted from very typical points (inlier) to very untypical points (outlier). The idea is, that dense regions of the data, the inlier, directly determine the ICA-directions, which are 1D-subspaces across the origin. By the use of this outlier index a robust algorithm for finding the mixing matrix is achieved.

3.3 Mean Field ICA

Mean Field ICA (mfica) by Ole Winther et al., IMM, Technical University of Denmark, 2002, compare [6]. Mean Field ICA is a probabilistic ICA approach. The mixing matrix is determined by maximum likelihood estimation. To estimate the mean and covariances of the sources, variational mean field theory and linear response theory is used. Various parameters can be chosen, especially to define prior information. The likelihood of the parameters is given by

$$P(\mathbf{x}|\mathbf{A},\mathbf{\Sigma}) = \int P(\mathbf{x}|\mathbf{A},\mathbf{\Sigma},\mathbf{s})P(\mathbf{s})d\mathbf{s}, \qquad (2)$$

with the noise covariance matrix $\mathbf{\Sigma}$. The mixing matrix is estimated by maximum a posteriori (MAP)

$$\mathbf{A}_{MAP} = argmax_{\mathbf{A}} P(\mathbf{A}|\mathbf{x},\mathbf{\Sigma}). \qquad (3)$$

For further information see [6].

4 Set-up

The simulation for thoroughly empirical verfification of the performance comparison were carried out based on a specific ICA-testsuite and well established performance measurements.

4.1 Testsuite

The tests are accomplished by a testsuite[2] that supports the evaluation of BSS algorithms in an automated way. This Matlab-based testsuite was developed to solve the overcomplete ICA problem as well as the standard case. Arbitrary algorithms can be integrated and tested.
In our tests we considered speech data and synthetic data sets. In Matlab the synthetic source signals were generated as follows, refer to [7]:

- Subgaussian: rand(n,5000).
- Supergaussian:

$-\log$(rand(n,T)).*max(0,sign(rand(n,T)-0.5)).

The speech data consist of polish speech data. Each dataset has T=5000 samples. Different linear mixing matrices are considered, e.g. of the dimension 2×3 and 2×4. The values of the mixing matrices are randomly chosen with uniformly distributed elements in the interval $(0, 1)$. The algorithms have been initialized with combinations of their individually required

[2]The testsuite ICYNATOR was developed at the University of Münster in 2005, see [2].

parameters. In the case of overcomplete ICA they all got the number of original sources as input. Sample reduction is done in an automated way. The size of the window used by the algorithms is gradually decreased.

4.2 Performance Measurements

The performance is determined according to several measurements.

- The signal-to-interference-ratio (SIR), see [7].
- The (squared) intersymbol interference (ISI), called *perform2* in the test, see [8].

$$ISI = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{p_{ij}^2}{max_k p_{ik}^2}-1\right) + \sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{p_{ij}^2}{max_k p_{kj}^2}-1\right), \tag{4}$$

for $\mathbf{P} = p_{ij} = \mathbf{AW}$. ISI is zero, if $\mathbf{P}$ is the unit matrix (i.e. the unmixing-matrix $\mathbf{W}$ has been perfectly estimated) and positive otherwise. Thus ISI calculates the difference between the mixing and the separating matrices.

5 Simulation Results

This section contains the achieved performance results of the three tested algorithms. The following problems have been tested, for illustration take a look at the figures.

- Separation of a linear mixing of eight (four) speech signals with a decreasing number of sensors. Figure 1 shows the results.
- Separation of three speech signals under different levels of noise in the case of two sensors. The results are measured by SIR, see left part of figure 2.
- Performance of the three algorithms in separating supergaussian, subgaussian and speech data in the case of two sensors and three sources. Each with the same 2×3 mixing matrices and evaluated by boxplots, see figure 4.
- Separation of four speech signals with a decreasing number of samples. The resulting SIR is presented in the right part of figure 2.
- Comparison to well-known algorithms in the case of standard-ICA with a 3×3 mixing of speech signals. Figure 3 shows boxplots of the resulting SIR.

Altogether the performance results of overcomplete ICA are not very satisfying as in many cases not all of the sources are reconstructed correctly. Often sources are found twice and others are neglected. All three tested algorithms got the number of original sources as input and did not discover the source number alone. In some cases good reconstruction results were achieved, but as we only consider the mean values over 100 runs, the few good results of the algorithms are not preponderating. The tests show that the performance of the algorithms falls with a decreasing number of sensors (figure 1). Geoica performs inferior to the other two algorithms. Mean field ICA remains with the lowest increase of the performance index. Astonishingly the performance results for overcomplete ICA are not affected by the addition of gaussian noise in our test, the moderate performance is not lowered (see left part of figure 2). It is demonstrated that geoica and ibica are restricted to supergaussian data (especially speech data) and perform kind of poor by separating subgaussian data (figure 4). The perform2 value is far from zero. Mfica achieves similar results for subgaussian and supergaussian data. Furthermore the algorithms have been tested in the standard ICA case. In comparison to well-known algorithms like cubica [9], efica [10], jade [11] or flexica [12] their performance was less satisfying, see figure 3. The right part of figure 2 shows the performance of the algorithms in a standard ICA problem with four speech signals and sample reduction. As expected the performance falls with a decreasing number of samples, except in the mean field ICA case. For a specific window size of samples ($T = 1500$), mfica is slightly outperformed by efica only.

6 Conclusion

Performance strength and weaknesses of a selection of overcomplete ICA-approaches have been evaluated. In the case of overcomplete ICA the quality of source reconstruction is not excellent except for individually cases. There are many other approaches to solve overcomplete ICA. So far, the author of these approaches did not provide a unique implementation, e.g. a matlab version. So the impact of individual implementations and its accuracy is of considerable influence [13]. It would be interesting to get an general overview of all overcomplete ICA methods and their performances. A first step towards this idea has been realized. Especially the influence of different preprocessing and sparsification methods is an important issue for further research to review. Furthermore, the question arises if there exists a general boundary of the quality of source reconstruction for over-

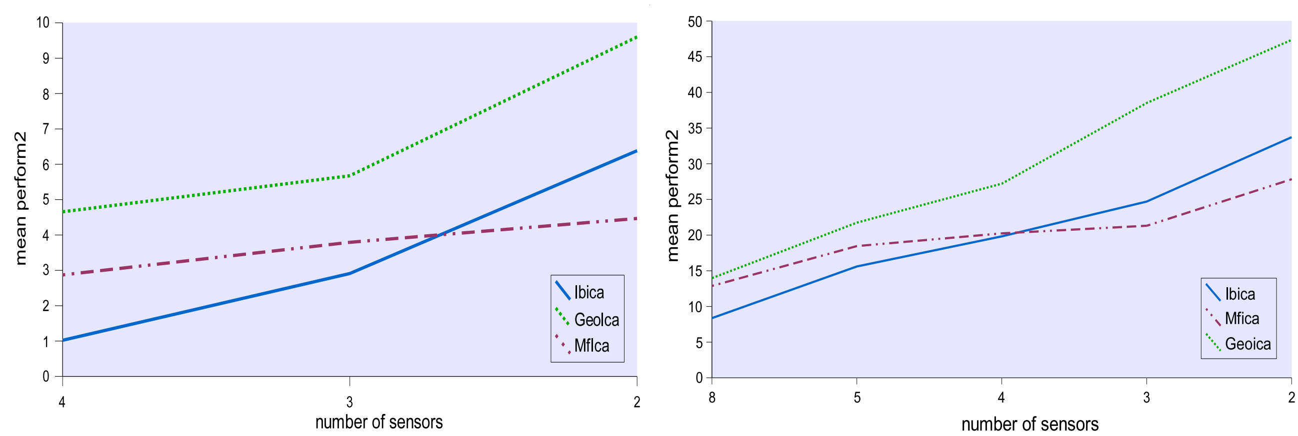


Fig. 1: Mean perform2 of 100 runs: Linear mixture of four (eight) speech signals, each with 5000 samples, with a decreasing number of sensors, from 4 to 2 in the left case and from 8 to 2 in the right case. All three algorithms perform better the more sensors are given. Geoica achieves worse results than ibica and mfica.

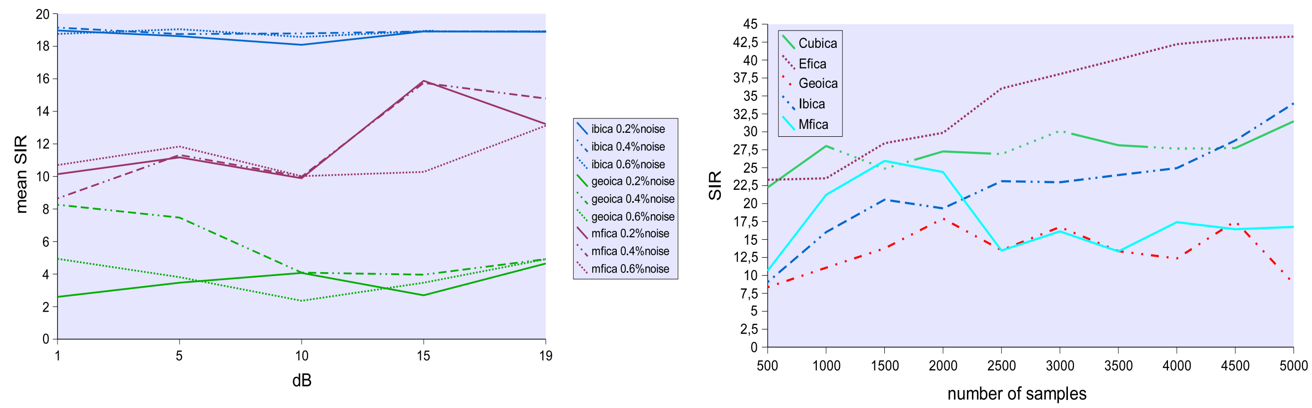


Fig. 2: Each graph: Mean SIR of 100 runs, each with 5000 samples; Left: Mixture of three speech signals, with linear mixing matrices of size 2×3, with different levels of noise. Right: Mixture of four speech signals, with linear mixing matrices of size 4×4, with a decreasing number of samples. For most algorithms the performance lowers with a falling number of samples. Efica outperforms the other algorithms.

complete ICA. The comparison to standard ICA approaches clearly underlines a performance trade-off to standard ICA efficiency, but mean field ICA clearly demonstrates an almost equal performance for a specific window size of samples. From the continuum of ICA related signal separation point of view, an comparison to single channel ICA approaches (e.g. see. [14]) is also strongly recommended.

Acknowledgements: Our thanks to Frank Meinecke et al., Fraunhofer FIRST.IDA Berlin, for providing the algorithm ibica. Thanks as well to Ole Winther et al., IMM, Technical University of Denmark, and to Fabian Theis et al. for publishing their algorithms.

References:

[1] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., 2001.

[2] ICA Matlab Toolbox and Documentation, WWU Münster: Icynator 1.0, http://cs.uni-muenster.de/ICA last date of access: 01.08.2006.

[3] F. J. Theis, E. W. Lang and C. G. Puntonet, A Geometric Algorithm for Overcomplete Linear ICA, *Int. J. of Neurocomputing*, 56, 2004, pp. 381–398.

[4] F. C. Meinecke, S. Harmeling and K.R. Müller, Inlier-based ICA with an Application to Super-

imposed Images, *Int. J. of Imaging Systems and Technology*, 2005, no pp. given.

[5] S. Harmeling, Independent Component Analysis and Beyond, PhD thesis, Universität Potsdam, 2004.

[6] P. A. Højen-Sørensen, O. Winther and L. K. Hansen, Mean Field Approaches to Independent Component Analysis, *Int. J. of Neural Computation*, 14, 2002, pp. 889–918.

[7] A. Cichocki and D. Erdogmus, MLSP Competition: Large scale, Ill-conditioned Blind Source Separation Problem with Limited Number of Samples, 2006.

[8] X. Giannakopoulos, Comparison of Adaptive Independent Component Analysis Algorithms, Master's thesis, Helsinki University, 1998.

[9] T. Blaschke and L. Wiskott, An Improved Cumulant Based Method for Independent Component Analysis, In *Proc. of ICANN*, 2002, pp. 1087–1093.

[10] Z. Koldovsky and P. Tichavsky, Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cram.-Rao Lower Bound, In *Proc. of SSP*, 2005, no pp. given.

[11] J.-F. Cardoso, High-Order Contrasts for Independent Component Analysis, *Int. J. of Neural Computation*, 11(1), 1999, pp. 157–192.

[12] S. Choi, A. Cichocki and S. Amari, Flexible Independent Component analysis, *Int. J. of VLSI Signal Processing*, 26(1/2), 2000, pp. 25–38.

[13] Y. Li, D. Powers and J. Peach, Comparison of Blind Source Separation Algorithms, *Int. J. of Advances in Neural Networks and Applications, WSEAS*, 2000, pp. 18–21.

[14] M. Davies and C. James. Source Separation using Single Channel ICA, In *Proc. of Int. Workshop on ICA*, ICArn, 2006, pp. 13–16.

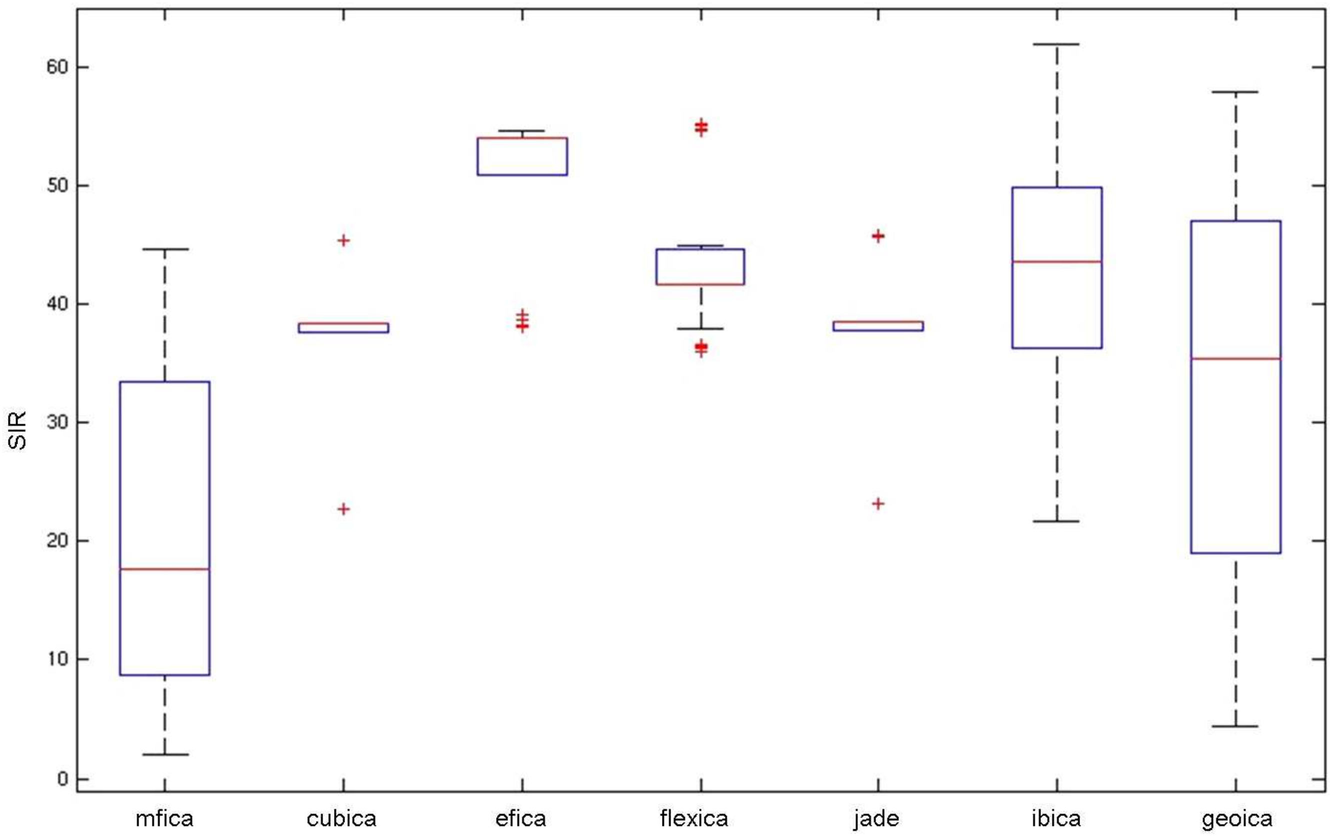
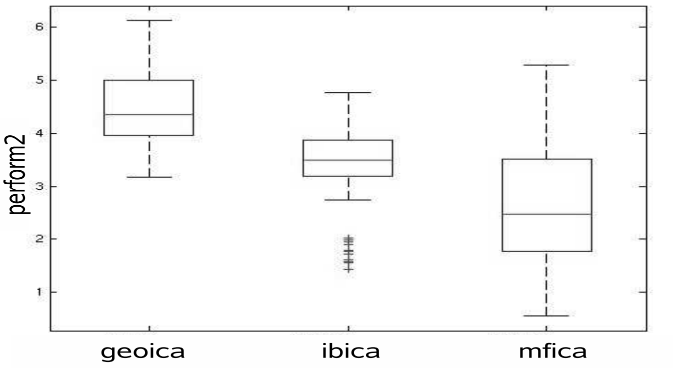
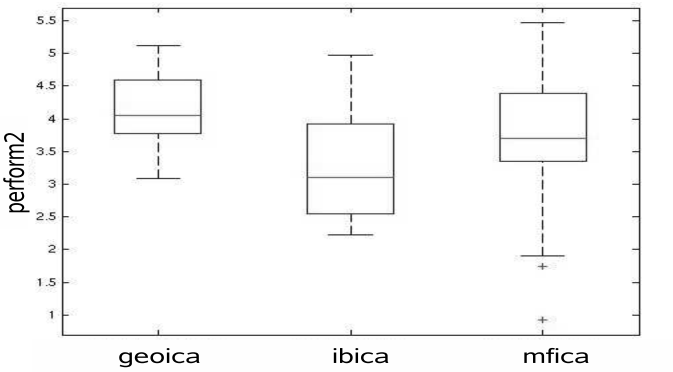


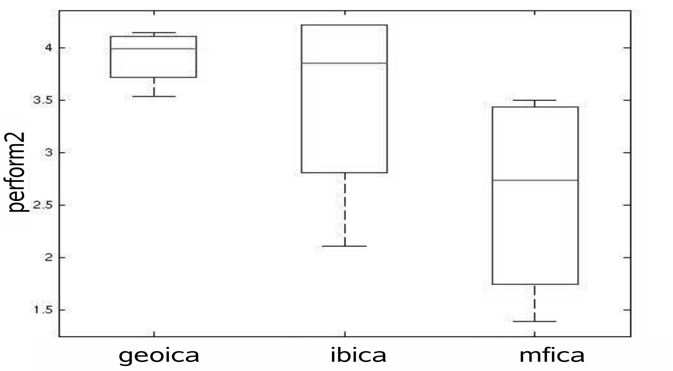
Fig. 3: Mean SIR of 100 runs: Mixture of three speech signals, each with 5000 samples, with linear mixing matrices of size 3×3. In comparison to the standard algorithms, geoica and mfica perform worse, they reach lower SIR values. Efica achieves the best average results.



(a) speech data



(b) subgaussian data



(c) supergaussian data

Fig. 4: Mean perform2 of 100 runs: Mixture of three signals of different data sets, each with 5000 samples, with linear mixing matrices of size 2×3.