METHODS OF FAIR COMPARISON OF PERFORMANCE OF LINEAR ICA TECHNIQUES IN PRESENCE OF ADDITIVE NOISE

Zbyněk Koldovský[1,2] *and Petr Tichavský*[2]

[1] Faculty of Mechatronic and Interdisciplinary Studies
Technical University of Liberec, Hálkova 6, 461 17 Liberec, Czech Republic
[2]Institute of Information Theory and Automation,
P.O.Box 18, 182 08 Prague 8, Czech Republic

ABSTRACT

Linear ICA model with additive Gaussian noise is frequently considered in many practical applications, because it approaches the reality often much better than the noise-free alternate. In this paper, a number important differences between noisy and noiseless ICA are discussed. It is shown that estimation of the mixing/demixing matrix should *not* be the main goal, in the noisy case. Instead, it is proposed to compare outcome of ICA algorithms with a minimum mean square (MMSE) separation, derived for known mixing model. The signal-to-interference-plus-noise ratio is suggested as the most meaningful performance criterion. A simulation study that compares a few well known ICA algorithms applied to noise data is included.

1. INTRODUCTION

Recently, Independent Component Analysis (ICA) became a very popular method for Blind Source Separation (BSS). Its aim is to transform the mixed random signals into source signals or components using the assumption that the source signals are mutually independent. In this paper, we shall consider an instantaneous linear mixing model. In the noisy case, the mixing process can be expressed as

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}, \tag{1}$$

where $\mathbf{X}$ is a $d \times N$ matrix with the $(k,\ell)-$th element denoted $x_{k\ell}$, d is the number of mixed signals and N is the number of samples. Similarly, $\mathbf{S}$ denotes a matrix of samples of the original signals s_{ij}. $\mathbf{A}$ is an unknown regular $d \times d$ mixing matrix. The signals s_{ij} are modeled as i.i.d. random variables with probability density functions $f_i(s_{ij})$ $i = 1, \ldots, d$ and $j = 1, \ldots, N$ [4] or as independent stationary Gaussian processes [10]. Finally, $\mathbf{N}$ denotes an additive noise, independent of the signals. For simplicity, the noise can be assumed to be a matrix of i.i.d Gaussian random variables, each having zero mean and variance σ^2.

The model (1) approaches reality better then the noise-free one (without $\mathbf{N}$) due to reasons like omnipresent sensor-noise. Moreover, note that the importance of the noise consideration is still greater in case of a more general convolutive model [1] that is closely related to the ICA [2]. The filter-truncation effect, which appears almost in all practical situations, can be modeled as the additive Gaussian noise thanks to the Central Limit Theorem.

There are several special noisy models that are incorporated in (1), e.g., the "source" noise model [3]. Some other authors consider a general noise covariance matrix $\mathbf{\Sigma}$ instead of $\sigma^2\mathbf{I}$. Note, however, that if $\mathbf{\Sigma}$ is known *a priori*, the latter model can be made equivalent to model (1) by pre-multiplying the data with $\mathbf{\Sigma}^{-1/2}$.

While the noise-free model $\mathbf{X} = \mathbf{AS}$ has been deeply investigated, i.e., theoretical limits for separation performance have been derived [4, 5, 6, 7], and several efficient techniques have been developed [8, 9, 10], little work dealing with the noisy model have been conducted [11, 12, 13]. The noise is frequently considered in simulations or in real applications, though.

The paper is organized as follows. Section 2 points out some important differences between the noisy and the noise-free model and shows that the estimation of the mixing (or demixing) matrix does not have such meaning as it is usually thought. In Section 3, minimum mean square error (MMSE) separation is discussed. It is shown that this should be regarded as the general optimal solution of the ICA. Finally, Section 4 provides a comparison of several known ICA methods in a noisy environment.

2. BASIC MODELS DIFFERENCES

In this section, we will point out three important differences between the noisy and noise-free models. The aim is to show that the tasks differ essentially in their solutions. On the other hand, the goals are the same: to retrieve the original sources.

[0]This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and by the Grant Agency of the Czech Republic under Grant 102/05/0278.

2.1. Identifiability

The well-known feature of the noise-free model $\mathbf{X} = \mathbf{AS}$ is that it can be resolved up to the original order, signs and scales of the original sources $\mathbf{S}$. The only assumption is that at most one of the signals $\mathbf{S}$ has a Gaussian distribution.

Note that the indeterminacy of scales of signals $\mathbf{S}$ is equivalent with indeterminacy of norms of columns of the mixing matrix $\mathbf{A}$. Thus, fixing the variance of $\mathbf{S}$, i.e. taking an assumption based on some prior information, enables us to estimate the norms of columns of $\mathbf{A}$ and vice versa. However, in practice, estimation of the original scales is irrelevant.

A necessary condition for estimability of any model parameter is the identifiability [15]. However, it was shown in [16] that the very σ is not identifiable in (1). This is not surprising since σ is indeterminable unless we know either the norms of columns of $\mathbf{A}$ and the original scales of $\mathbf{S}$. Without knowing this, it is impossible to determine how much of the energy of any transformed signal $\mathbf{w}^T\mathbf{X} = \mathbf{w}^T\mathbf{AS} + \mathbf{w}^T\mathbf{N}$ corresponds to the noise.

To conclude: while the scales are irrelevant for signal reconstruction in the noise-free model, it is not so in the noisy model because of the ambiguity of noise variance.

2.2. Optimum solution is ambiguous

There are several criteria for noise-less model separation [17] such as mutual information of the separated signals, marginal entropy of each of the separated signals, likelihood, etc. In spite of leading to different algorithms, their common property is the consistency of the mixing matrix estimation (up to the indeterminacies). Even if an algorithm aims at qualities of the separated signals only [18], the criterion, if suitable, is optimized iff the separated signals equal the original ones, i.e. the separating transformation equals the mixing matrix inversion.

This is no longer truth when assuming model with additive noise, because the original signals cannot be separated from noise. Thus, for instance, the mutual information of signals

$$\mathbf{A}^{-1}\mathbf{X} = \mathbf{S} + \mathbf{A}^{-1}\mathbf{N} \tag{2}$$

is not minimized, in general, due to correlations caused by the second (noise) term $\mathbf{A}^{-1}\mathbf{N}$. Similarly, the entropy of any of signals (2) need not to be minimum since the non-gaussianity can be minimized in different directions $\mathbf{WX}$.

In summary, while optimal solution of the noise-free model is, obviously, achieved whenever the mixing matrix inversion is retrieved, in noisy model (1), the optimality of the solution depends on the criterion used for separation.

2.3. Equivariance

Several ICA algorithms [19, 18] have been shown to have (at least asymptotically) equivariant performance, i.e. their accuracy is independent of the mixing matrix $\mathbf{A}$ (in the noise-free environment). This is caused by equivariance property [20] of the separating criterion used, e.g. mutual information. However, in the noisy environment, many commonly used criteria for separation are no more equivariant.

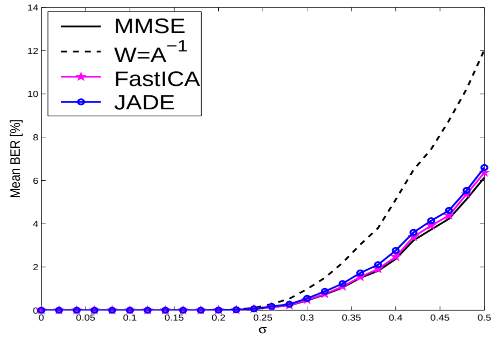

Fig. 1. Mean bit error rate of separated BPSK signals in noisy environment with $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 0.2 \end{pmatrix}$ computed from 100 trials for each $\sigma \in [0, 0.5]$. MMSE solution is much proper than that minimizing SIR (4), i.e. inversion of $\mathbf{A}$.

Let $\mathbf{W}$ be the separating matrix. Then the separated signals are

$$\widehat{\mathbf{S}} = \mathbf{WX} = \mathbf{WAS} + \mathbf{WN}. \tag{3}$$

An example of the equivariant criterion is the signal-to-interference ratio of the k-th separated signal defined by

$$\mathrm{SIR}_k = \frac{(\mathbf{WA})^2_{kk}}{\sum_{i \neq k}^{d} (\mathbf{WA})^2_{ki}} \tag{4}$$

It is obvious that the criterion is optimized (maximized) for $\mathbf{W} = \mathbf{A}^{-1}$. However, note that suitability of the criterion is questionable since it disregards presence of the noise in the separated signals. Moreover, any consistent estimator, that maximizes (4), is not known (if it even exists [16]).

3. MMSE SEPARATION

The fact that the model (1) is not identifiable suggests aiming at qualities of the separated signals rather than the estimation of model parameters (see Fig. 1). Let $\widehat{\mathbf{s}}_k$ be the k-th separated signal. The MMSE solution is defined as one that minimizes

$$\mathrm{E}[(\mathbf{s}_k - \widehat{\mathbf{s}}_k)^2], \tag{5}$$

where $\mathbf{s}_k$ denotes a random variable with the same distribution like the k-th original signal. We assume, for simplicity, that the original signals have zero mean and unit variance.

The MMSE separation can be expressed analytically, if the mixing matrix $\mathbf{A}$ is known. The resultant separating matrix $\mathbf{W}$ is [21]

$$\mathbf{W}^{\mathrm{MMSE}} = \mathbf{A}^T(\mathbf{AA}^T + \sigma^2\mathbf{I})^{-1}. \tag{6}$$

It can be shown that (6) minimizes simultaneously the *signal-to-interference-plus-noise ratio* (SINR) of the separated signals, for the k-th estimated signal, defined by

$$\mathrm{SINR}_k = \frac{(\mathbf{WA})_{kk}^2}{\sum_{i\neq k}^{d}(\mathbf{WA})_{ki}^2 + \sigma^2 \sum_{i=1}^{d} \mathbf{W}_{ki}^2} \tag{7}$$

which is scale-invariant. Note that while the criterion (4) can be arbitrarily large, the criterion (7) is bounded by its value when taking $\mathbf{W} = \mathbf{W}^{\mathrm{MMSE}}$, i.e.

$$\min \mathrm{SINR}_k = \frac{\mathbf{G}_{kk}^2}{\sum_{i\neq k}^{d} \mathbf{G}_{ki}^2 + \sigma^2 \sum_{i=1}^{d}(\mathbf{GA}^{-1})_{ki}^2}, \tag{8}$$

where $\mathbf{G} = (\mathbf{I} + \sigma^2(\mathbf{A}^T\mathbf{A})^{-1})^{-1}$. This bound depends on the mixing matrix $\mathbf{A}$ and the noise variance σ^2 only. A bound that takes signals distributions and number of processed data N into consideration (Cramér-Rao-like bound) has not yet been derived in a closed form in the literature. Nevertheless, as can be seen from simulations (Section 4), (8) provides a more primitive but still useful ultimate bound that limits the attainable one when processing finite amount of data.

In order to study influence of $\mathbf{A}$ and σ on (8) following asymptotic expansion for "small" σ can be derived

$$\min \mathrm{SINR}_k = \frac{1}{\sigma^2\|\mathbf{w}_k\|^2} - B + \mathcal{O}(\sigma^2), \text{ where} \tag{9}$$

$$B = 2 + \frac{1}{\|\mathbf{w}_k\|^4}\Big(\sum_{i\neq k}^{d}(\mathbf{WW}^T)_{ki}^2 - 2\sum_{i=1}^{d}\mathbf{W}_{ki}(\mathbf{WW}^T\mathbf{W})_{ki}\Big),$$

$\mathbf{W} = \mathbf{A}^{-1}$, and $\mathbf{w}_k^T$ denotes k-th row of $\mathbf{W}$. We have following comments:

(1) It is obvious that if $\mathbf{A}$ is unitary, i.e. $\mathbf{AA}^T = \mathbf{I}$, the optimum solution is obtained by taking the separating matrix $\mathbf{W} = \mathbf{A}^{-1} = \mathbf{A}^T$. This corresponds with the fact that the noise $\mathbf{N}$ can be multiplied by an arbitrary unitary matrix since it does not affect its covariance structure and the noise signals remain independent. In that case, the noise can be regarded as the "source" noise [3].

(2) The first term of the expansion (9) shows that if we want to achieve (approximately) uniform bound for all signals, we have to normalize rows of $\mathbf{A}^{-1}$. This is in contrast with an usual erroneous choice of $\mathbf{A}$ in experiments, where mostly rows of $\mathbf{A}$ are normalized.

(3) The second term reflects non-unitarity of $\mathbf{A}$. It vanishes whenever rows of $\mathbf{A}^{-1}$ are orthogonal. Finally, note that accuracy of the asymptotic expansion (9) is violated when $\mathbf{A}$ is nearly singular.

The following example shows, that common ICA techniques (JADE, FASTICA) produce separated signals that are closer to the MMSE solution than to $\mathbf{A}^{-1}\mathbf{X}$, see Fig. 2.

The same fact is demonstrated in the example in Fig. 1, where the goal is to separate BPSK signals from a noisy mixture. Here, the ultimate criterion is the bit error rate. Again,

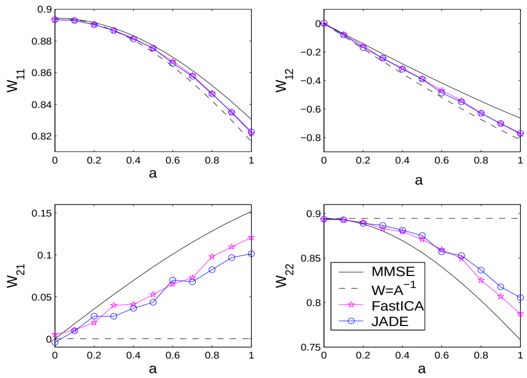

Fig. 2. Elements of the separating matrix $\mathbf{W}$ acquired by several techniques when $\mathbf{A} = \left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right)$, $a \in [0, 1]$, $\sigma = 0.5$.

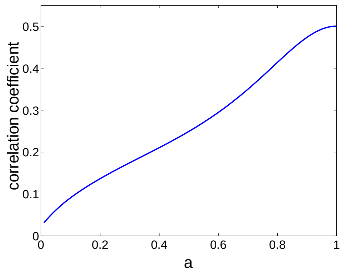

Fig. 3. Correlation coefficient of two MMSE components when $\sigma = 0.3$ for mixing matrix $\mathbf{A} = \left(\begin{smallmatrix} a & a \\ 1 & a \end{smallmatrix}\right)$ with varying parameter $a \in [0, 1]$. Matrix $\mathbf{A}$ is singular for $a = 0, 1$.

FastICA and JADE perform better than separation by the inverse of the true mixing matrix.

The last example (Fig. 3) indicates influence of the mixing matrix on quality of separation. In the presence of the noise, the signals that minimize the MMSE criterion may no longer be uncorrelated. In the example, the correlation goes up to 0.5, when a parameter a approaches 1.

The correlation of the MMSE solution has an adverse effect namely on algorithms which constrain the correlation of the separated signals to be exactly zero [14].

4. CONCLUSIONS

It follows from the previous sections, that a fair comparison of ICA algorithms in presence of additive noise requires (1) a relevant criterion (we recommend to use the SINR and compare it with the MMSE solution) and (2) a well thought-out choice of the mixing matrix.

For example, a common mistake in testing performance of separation techniques is to normalize rows of $\mathbf{A}$ with the intention to get the energy of noise in the ideally separated signals independent of $\mathbf{A}$. Then, SIR is used as the criterion for separation accuracy. This is wrong for several reasons described above and may lead to wrong conclusions about

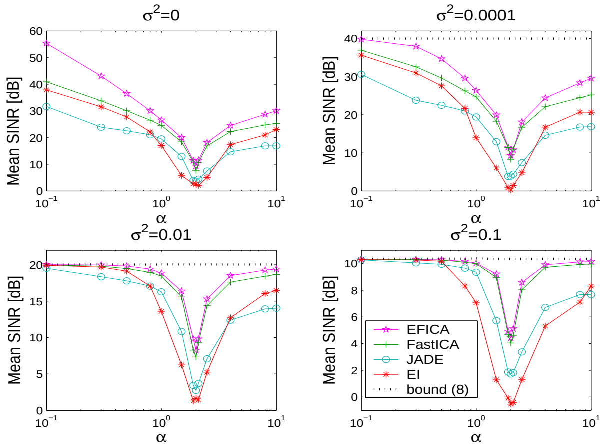

Fig. 4. Quality of separation of 13 signals with Generalized Gaussian distribution measured by mean SINR.

algorithm performance.

As an example of fair comparison using criterion (7) we propose the following experiment: 13 signals of length $N = 5000$ were mixed with a random mixing matrix in each trial. The signals have Generalized Gaussian distribution [9] with parameter α, respectively, equal to 0.1, 0.3, 0.5, 0.8, 1, 1.5, 1.9, 2, 2.1, 2.5, 4, 8, and 10. The mixing matrix is chosen so that its inversion has unitary rows (due to uniform bound (8) for each signal) and is well-conditioned (the ratio of the maximum and the minimum eigenvalue is smaller than 100).

The mixed signals were corrupted by a Gaussian noise with various σs, and separated by well-known ICA algorithms: Extended Infomax (EI) [22], JADE [19], FastICA [18] (the symmetric approach with "tanh" nonlinearity), and EFICA [9]. Mean SINR of each estimated signal from 100 independent trials for each σ is shown in figure 4. It is obvious that the mean SINR declines with growing "gaussianity" ($\alpha = 2$) of the corresponding signal. However, note that the differences in estimability of the signals are less significant when σ becomes large.

5. REFERENCES

[1] A. M. Bronstein, M. M. Bronstein, and M. Zibulevsky, "Relative Optimization for Blind Deconvolution" *IEEE Trans. on Signal Processing*, vol. 53, no. 6, pp. 2018-2026, June 2005.

[2] S.C. Douglas and S. Haykin, "On the Relationship Between Blind Deconvolution and Blind Source Separation", *Proc. 31st Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, pp. 1591-1595, Nov. 1997.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, New York, 2001.

[4] O. Shalvi, E. Weinstein, "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs", *IEEE Tr. Information Theory*, vol. 40, no. 2, pp. 328-339, March 1994.

[5] D. Yellin, B. Friedlander, "Multichannel system identification and deconvolution: performance bounds", *IEEE Tr. Signal Processing*, vol. 47, no. 5, pp. 1410-1414, May 1999.

[6] Z. Koldovský, P. Tichavský and E. Oja, "Cramér-Rao lower bound for linear Independent Component Analysis", *Proc. ICASSP 2005*, Philadelphia, vol. III, pp. 581-584, March 2005.

[7] P. Tichavský, Z. Koldovský, and E. Oja, "Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis", to appear in the April 2006 issue of the *IEEE Trans. on Signal Processing*.

[8] D.T. Pham, Garat, P., "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach", *IEEE Tr. Signal Processing*, vol. 45, no. 7, pp. 1712-1725, July 1997.

[9] Z. Koldovský, P. Tichavský and E. Oja, "Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound". *Proc. of IEEE/SP 13th Workshop on Statistical Signal Processing*, Bordeaux, July 17-20, 2005.

[10] A. Yeredor, "Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting", *IEEE Signal Processing Letters*, vol. 7, no. 7, pp. 197-200, July 2000.

[11] A. Hyvärinen, "Gaussian Moments for Noisy Independent Component Analysis". *IEEE Signal Processing Letters*, vol. 6, pp. 145-147, June 1999.

[12] E. Moulines, J.-F. Cardoso, and E. Gassiat "Maximum Likelihood for Blind Separation and Deconvolution of Noisy Signals Using Mixture Models", *Proc. ICASSP'97 Munich*, vol. 5, pp. 3617-20, 1997.

[13] S.C. Douglas, A. Cichocki, and S. Amari, "A Bias Removal Technique for Blind Source Separation with Noisy Measurements," *Electronics Letters*, vol. 34, no. 14, pp. 1379-1380, July 1998.

[14] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms", *Proc. EUSIPCO*, pp. 776-779, Edinburgh, September 1994.

[15] J. Eriksson and V. Koivunen, "Identifiability and Separability of Linear ICA Models Revisited", *Proc. of ICA2003*, pp. 23-27, April 2003.

[16] M. Davies, "Identifiability Issues in Noisy ICA", *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 470-473, May 2005.

[17] J.-F. Cardoso, "Blind signal separation: statistical principles", *Proc. of the IEEE*, vol. 90, no. 8, pp. 2009-2026, Oct. 1998.

[18] A. Hyvärinen, "A Fast Fixed-Point Algorithm for Independent Component Analysis". *Neural Computation*, 9(7):1483-1492, 1997.

[19] J.-F. Cardoso, "High-order Contrasts for Independent Component Analysis", *Neural Computat.*, vol. 11, no. 1, pp. 157-192, 1999.

[20] J.-F. Cardoso and B. H. Laheld, "Equivariant Adaptive Source Separation", *IEEE Trans. on Signal Processing*, vol. 44, no. 12, pp. 3017-3030, Nov. 1996.

[21] M. Joho, H. Mathis, and R. H. Lambert, " Overdetermined Blind Source Separation: Using More Sensors Than Source Signals in a Noisy Mixture", *Proc. of ICA2000*, Helsinki, Finland, pp. 81-86, June 2000.

[22] S. Makeig, A. J. Bell, T-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data", *Advances in Neural Information Processing Systems* 8:145-151, MIT Press, Cambridge, MA, 1996.