EXTENSION OF EFICA ALGORITHM FOR BLIND SEPARATION OF PIECEWISE STATIONARY NON GAUSSIAN SOURCES

Zbyněk Koldovský[1], *Jiří Málek*[1], *Petr Tichavský*[2], *Yannick Deville*[3], *and Shahram Hosseini*[3]

[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
[3]Université Paul Sabatier Toulouse 3 - Observatoire Midi-Pyrénées - CNRS,
Laboratoire d'Astrophysique de Toulouse-Tarbes, 14 Av. Edouard Belin, 31400 Toulouse, France

ABSTRACT

We propose an extension of EFICA algorithm for piecewise stationary and non Gaussian signals. The proposed method is able to profit from varying distribution of the original signals and also from their varying variance, which is demonstrated by simulations with real-world signals. We show that in case of constant-variance signals, the accuracy of the method may achieve the corresponding Cramér-Rao bound, if score functions of the original signals are known in all blocks.

Index Terms— Independent Component Analysis, Piecewise Stationary Signals, Cramér-Rao Lower Bound, Blind Source Separation, FastICA Algorithm

1. INTRODUCTION

The underlying model considered in Independent Component Analysis (ICA) [1, 2] is

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\mathbf{s} = [s_1, \ldots, s_d]^T$ is a vector of independent random variables (RVs), and each of them represents one of unknown original signals. In practice, N i.i.d. realizations of $\mathbf{x}$ are available, that are mixtures of the signals $\mathbf{s}$ via unknown $d \times d$ regular mixing matrix $\mathbf{A}$. Using the assumption of independence of $s_1, \ldots, s_d$, the goal is to estimate the demixing transform $\mathbf{A}^{-1}$ up to an indeterminable order, scales, and signs of its rows.

Numerous methods for separation of i.i.d. signals have been proposed [7, 8, 9]; some recent algorithms [4, 10] were developed to achieve accuracy that approaches the respective Cramér-Rao Lower bound (CRLB) [3]. The bound, for an unbiased estimator $\widehat{\mathbf{W}}$ of $\mathbf{A}^{-1}$, is

$$\mathrm{CRLB}[\mathbf{G}_{k\ell}] = \frac{1}{N}\frac{\kappa_\ell}{\kappa_k\kappa_\ell - 1}, \qquad k \neq \ell, \tag{2}$$

where $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$ is the so-called *gain* matrix, which should be close to the identity[1], and $\kappa_k = \mathrm{E}[\psi_k^2(x)]$, where $\psi_k = -f_k'(x)/f_k(x)$ is the score function of the probability density function (pdf) $f_k(x)$ of the k-th RV s_k. The knowledge of the score functions or their proper estimation is a common necessity of the algorithms to achieve the bound. Another algorithm proposed for the non stationary and non Gaussian scenario is NSGS by Pham [11].

2. FASTICA AND EFICA ALGORITHMS

The FastICA algorithm [8] is based on optimization of a contrast function

$$c(\mathbf{w}_k) = \mathrm{E}[G(\mathbf{w}_k^T\mathbf{z})], \tag{3}$$

where $\mathbf{w}_k^T$ denotes the k-th row of the de-mixing matrix $\widehat{\mathbf{W}}$ to-be estimated, $G(\cdot)$ is a nonlinear function whose derivative will be denoted by $g(\cdot)$, and $\mathbf{z}$ is a vector derived by transforming signals $\mathbf{x}$ so that the elements of $\mathbf{z}$ are not correlated and have unit variance. The optimization of $c(\mathbf{w}_k)$ proceeds via iteration

$$\mathbf{w}_k^+ \quad \leftarrow \quad \mathrm{E}[\mathbf{z}g(\mathbf{w}_k{}^T\mathbf{z})] - \mathbf{w}_k\mathrm{E}[g'(\mathbf{w}_k{}^T\mathbf{z})], \tag{4}$$

where the theoretical expectations are replaced by respective sample means. While the one-unit FastICA completes each iteration by normalizing the vector $\mathbf{w}_k^+$, the symmetric FastICA computes d iterations (4) in parallel and does a symmetric orthogonalization of $[\mathbf{w}_1^+, \ldots, \mathbf{w}_d^+]^T$ to estimate all rows of the demixing matrix $\widehat{\mathbf{W}}$. The theoretical performance [3] of the one-unit FastICA is characterized by

$$\mathrm{var}[\mathbf{G}_{k\ell}^{1\mathrm{U}}] \approx \frac{1}{N}\frac{\gamma_k}{\tau_k^2} \stackrel{\mathrm{def.}}{=} \frac{1}{N}V_{k\ell}^{1\mathrm{U}}, \qquad k \neq \ell, \tag{5}$$

where $\mathbf{G}^{1\mathrm{U}}$ is the gain matrix, each of its rows corresponds to the estimation of one demixing vector, and $\gamma_k = \beta_k - \mu_k^2$,

[0]This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and through the Grant 102/07/P384 of the Grant Agency of the Czech Republic.

[1]Without loss of generality, we assume that the original signals have unit variance.

$\tau_k = \nu_k - \mu_k$, $\mu_k = \mathrm{E}[s_k g(s_k)]$, $\nu_k = \mathrm{E}[g'(s_k)]$, and $\beta_k = \mathrm{E}[g^2(s_k)]$.

EFICA [4] proceeds in three steps: (1) It preestimates all the original signals by means of the symmetric FastICA, (2) for each $k = 1, \ldots, d$, adaptively chooses nonlinearity $g \stackrel{\text{def.}}{=} g_k$ for approximating the score function of the k-th signal, and (3) does the fine-tuning (further one-unit FastICA iterations using the nonlinearities found in step (2)), and a refinement. Using the theoretical performance of the fine-tuning given by (5), in the refinement, optimum weights are computed, for each $k = 1, \ldots, d$, according to[2]

$$c_{k\ell} = \frac{V_{k\ell}^{1\mathrm{U}}}{V_{\ell k}^{1\mathrm{U}} + 1}, \quad k \neq \ell, \qquad c_{kk} = 1, \quad \ell = 1, \ldots, d \tag{6}$$

and used to form matrix

$$\mathbf{W}_k^+ = [c_{k1}\mathbf{w}_1^+/\|\mathbf{w}_1^+\|, \ldots, c_{kd}\mathbf{w}_d^+/\|\mathbf{w}_d^+\|]^T. \tag{7}$$

The k-th row of symmetric orthogonalization of $\mathbf{W}_k^+$ yields the final estimate of $\mathbf{w}_k$. The performance of EFICA is then given by

$$\mathrm{var}[\mathbf{G}_{k\ell}^{\mathrm{EF}}] \approx \frac{1}{N} \frac{V_{k\ell}^{1\mathrm{U}}(V_{\ell k}^{1\mathrm{U}} + 1)}{V_{k\ell}^{1\mathrm{U}} + V_{\ell k}^{1\mathrm{U}} + 1}, \qquad k \neq \ell. \tag{8}$$

In the case when $g_k = \psi_k$, it holds that $\beta_k = \nu_k = \kappa_k$ and $\mu_k = 1$. Then, $V_{k\ell}^{1\mathrm{U}} = 1/(\kappa_k - 1)$, and its substitution into (8) gives (2) and proves the asymptotic efficiency of EFICA.

3. PIECEWISE STATIONARY ICA

The model considered in this paper, called piecewise stationary model, is such that the samples of signals need *not* be identically distributed, specifically, the pdf $f_k(x)$ of s_k may be different at each time instant/interval [11]. To allow practical estimation of signal statistics, we will assume that there are M blocks of the same integer length N/M where the distribution of the signals is unchanging. In that case, the model (1) holds within each block, i.e.,

$$\mathbf{x}^{(I)} = \mathbf{A}\mathbf{s}^{(I)}, \qquad I = 1, \ldots, M. \tag{9}$$

From here on, the superscript (I) denotes quantities, RVs, or functions related to the I-th block.

Prior to the novel algorithm proposal, we suggest that a straightforward adaptation of the FastICA-based algorithms to the piecewise stationary model (9) may consists in redefinition of the contrast (3) by

$$c(\mathbf{w}_k) = \lambda_k^{(1)}\mathrm{E}[G_k^{(1)}(\mathbf{w}_k^T\mathbf{z}^{(1)})] + \cdots + \lambda_k^{(M)}\mathrm{E}[G_k^{(M)}(\mathbf{w}_k^T\mathbf{z}^{(M)})] \tag{10}$$

where $\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}$ are suitable weights, and $G_k^{(1)}, \ldots, G_k^{(M)}$ are properly chosen nonlinear functions. In practice, this means applying different nonlinearity $g(\cdot)$ in (4) on each block of samples of $\mathbf{w}_k^T\mathbf{z}^{(I)}$, i.e. the derivatives $g_k^{(I)}$ of $G_k^{(I)}$.

[2] Note that the definition of the weights in [4] is different due to normalization of vectors $\mathbf{w}_k^+$ in (7), $k = 1, \ldots, d$.

4. EXTENSION OF EFICA ALGORITHM

In this section, we propose an extension of EFICA algorithm, called Extended EFICA, that is tailored to piecewise stationary signals obeying the model (9). The concept consisting of the three following steps is similar to that of the original EFICA described above:

EEF1 Separation by the symmetric FastICA in order to obtain a preestimate of the demixing matrix $\widehat{\mathbf{W}}$.

EEF2 Fine-tuning of each row of $\widehat{\mathbf{W}}$ by means of the one-unit FastICA with the contrast function (10). Selections of the weights and the nonlinearities are simultaneously updated as described below.

EEF3 The refinement to get the most accurate and final estimate of the whole demixing matrix.

As shown in [4], the symmetric FastICA is a well-established way for fast and reliable pre-estimation of $\widehat{\mathbf{W}}$. Here also, performance depends on the nonlinear function $g(\cdot)$, which, in theory, may not be sufficient for various non-Gaussian signals [5]. Luckily, this is not the case of most practical signals such as speech as will be shown in simulations section.

The second and the third steps are due to the accuracy improvement, which can be accomplished if only nonlinear functions $g_k^{(I)}$ are properly chosen. Thanks to the first step, this can be done adaptively using the separated signals therefrom, specifically, the score functions on each block I can be estimated as the optimum choice of the nonlinearities [3, 5]. The weights $\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}$ in (10) could be either set to the same nonzero value, which will be called the *Uniform Extended EFICA*, or according to the expression (13) derived hereafter.

For the score function estimation, we use the flexible and fast parametric estimator proposed by [10], that minimizes mean square distance between a score function and a linear combination of K basis functions $h_1(x), \ldots, h_K(x)$, i.e.,

$$\min_{\theta_1,\ldots,\theta_K} \mathrm{E}\left[\left(\psi(x) - \sum_{i=1}^{K}\theta_i h_i(x)\right)^2\right]. \tag{11}$$

Since $\mathrm{E}[\psi(x)h(x)] = \mathrm{E}[h'(x)]$ for any function $h(x)$, the minimization is fast, because it requires estimation of moments $\mathrm{E}[h_i^2(x)]$ and $\mathrm{E}[h_i'(x)]$, $i = 1, \ldots, K$, and leads to the solution of a set of K linear equations. Moreover, the moments are used in further computations (e.g. in (4)), which yields computational savings.

In our implementation, we have decided for two ($K = 2$) basis functions: $h_1(x) = x^3$, that is good for sub-Gaussian sources, and $h_2(x) = x/(1+6|x|)^2$ working well with super-Gaussian sources [5]. Such choice turns out to be appropriate for a wide class of distributions and offers a good trade-off between accuracy and speed. For instance, when considering signals with Generalized Gaussian distributions, the estimator (11) with our settings yields comparable results with the adaptation proposed in [4], which is tailored to those distributions.

Finally, the refinement step is done in the same way as in [4], i.e. using (6) and (7), with the exception that the weights $c_{k\ell}$ are computed using $V_{k\ell}^{1\text{Ug}}$ instead of $V_{k\ell}^{1\text{U}}$ due to the different performance of the fine-tuning method in EEF2. The expression $V_{k\ell}^{1\text{Ug}}$ is defined below in (12).

5. PERFORMANCE ANALYSIS

Here, we present theoretical performance analysis of the proposed algorithm and derive selections of its parameters therefrom. The analysis assumes constant (unit) variance of the original signals in each block. However, it should be stressed that this is only a working assumption of the analysis, which need not be fulfilled when doing separation in practice.

First, we need to analyze performance of the one-unit FastICA utilizing the contrast function (10), which is the building stone of the fine-tuning in step EEF2. This can be easily done by generalizing results of the analysis in [3], however, we leave out the proof due to lack of space and refer readers to [12]. It is shown that the performance is given by

$$\text{var}[\mathbf{G}_{k\ell}^{1\text{Ug}}] \approx \frac{1}{N} \underbrace{\frac{\frac{1}{M}\sum_{I=1}^{M} {\lambda_k^{(I)}}^2 \beta_k^{(I)} - (\frac{1}{M}\sum_{I=1}^{M}\lambda_k^{(I)}\mu_k^{(I)})^2}{(\frac{1}{M}\sum_{I=1}^{M}\lambda_k^{(I)}\tau_k^{(I)})^2}}_{\stackrel{\text{def.}}{=} V_{k\ell}^{1\text{Ug}}}, \tag{12}$$

where $\mathbf{G}^{1\text{Ug}}$ is the gain matrix resulting from the algorithm, and $\lambda_k^{(I)}, \dots, \lambda_k^{(I)}$ are the weights introduced in (10). The optimal choice of λs is given when minimizing (12) subject to them. In [12] we show that the J-th optimum weight is

$$\lambda_k^{(J)} = \frac{1}{M}\left(\frac{\tau_k^{(J)}}{\beta_k^{(J)}} + A_k B_k \frac{\mu_k^{(J)}}{\beta_k^{(J)}}\right), \tag{13}$$

where $A_k = \left(\sum_{I=1}^{M}\frac{\gamma_k^{(I)}}{\beta_k^{(I)}}\right)^{-1}$ and $B_k = \sum_{I=1}^{M}\frac{\mu_k^{(I)}\tau_k^{(I)}}{\beta_k^{(I)}}$. Hence, the performance achieved in the second step EEF2, is given by inserting (13) into (12).

The final performance of the proposed Extended EFICA is given after analyzing the effect of the refinement step EEF3. The refinement, in the original EFICA, utilizes weights given by (6), which, in fact, are functions of the performance achieved by the fine-tunings characterized by $V_{k\ell}^{1\text{U}}$. Thanks to this relation, the weights that are optimal for the Extended EFICA are simply given when inserting $V_{k\ell}^{1\text{Ug}}$ (defined in (12)) into (6) instead of $V_{k\ell}^{1\text{U}}$. The same holds for the performance of the Extended EFICA, which is analogous to (8), i.e., for $\mathbf{G}^{\text{EEF}}$ being the resulting gain matrix,

$$\text{var}[\mathbf{G}_{k\ell}^{\text{EEF}}] \approx \frac{1}{N}\frac{V_{k\ell}^{1\text{Ug}}(V_{\ell k}^{1\text{Ug}}+1)}{V_{k\ell}^{1\text{Ug}}+V_{\ell k}^{1\text{Ug}}+1} \qquad k \neq \ell. \tag{14}$$

5.1. Cramér-Rao Bound vs. Optimal Performance

The CRLB of the piecewise stationary model under the assumption of constant-variance signals is given by

$$\text{CRLB}[\mathbf{G}_{k\ell}] = \frac{1}{N}\frac{\overline{\kappa_\ell}}{\overline{\kappa_k}\,\overline{\kappa_\ell} - 1}, \qquad k \neq \ell, \tag{15}$$

where $\overline{\kappa_k} \stackrel{\text{def.}}{=} \frac{1}{M}\sum_{I=1}^{M}\kappa_k^{(I)}$. See the proof in the Appendix.

We compare the bound with the performance of Extended EFICA in the special case that the nonlinearities selected in the second step (EEF2) equal the corresponding score functions, i.e., $g_k^{(I)} = \psi_k^{(I)}$, for $k = 1,\dots,d$, $I = 1,\dots,M$. Then, $\beta_k^{(I)} = \nu_k^{(I)} = \kappa_k^{(I)}$, $\mu_k^{(I)} = 1$, $\tau_k^{(I)} = \gamma_k^{(I)} = \kappa_k^{(I)} - 1$, (13) simplifies to a constant $\lambda_J = 1/M$, and

$$V_{k\ell}^{1\text{Ug}} = \frac{1}{\overline{\kappa_k} - 1}. \tag{16}$$

Inserting this into (14) we get the right-hand side of (15). Consequently, Extended EFICA is asymptotically efficient in case of the constant-variance signals when score functions are properly approximated.

The uniformity of the weights (13) for this particular case gives rise to the Uniform Extended EFICA algorithm introduced in Section 4. The weights λs need not be estimated there, since $g_k^{(I)}(\cdot)$ are assumed to be the score functions. This may be useful when the number of blocks M is not known and is overestimated. A possible approach for automated choice of M can be found, e.g., in [13].

6. SIMULATIONS

Our first experiment deals with separation of six randomly mixed signals of length $N = 10^3$ having constant unit variance, whose two blocks of the same length are distributed either differently or alike; see Figure 1. Theoretical performances, marked by "theory" in the acronym, were computed using (8) and (14), respectively, for EFICA and Extended EFICA. Results of this simple experiment shown in Figure 1 corroborate validity of the analysis and demonstrate improved performance of the proposed method compared to the original EFICA.

We also compare the performance of NSNG algorithm [11] that performs (quasi)-MLE estimation without requiring the signal variances to be constant in time. Therefore, it may achieve better performance than Extended EFICA in scenarios where the signal variances are varying. On the other hand, we have observed cases of instability and misconvergence of NSNG, which spoil the average performance of the algorithm, and, therefore, Extended EFICA outperforms it on average.

The second experiment was done with similar setup but with 20 speech signals of length $N = 5000$. The signals were randomly taken from a database of utterances[3], mixed with a random matrix, and separated.

[3]The data-set is available online at [14].

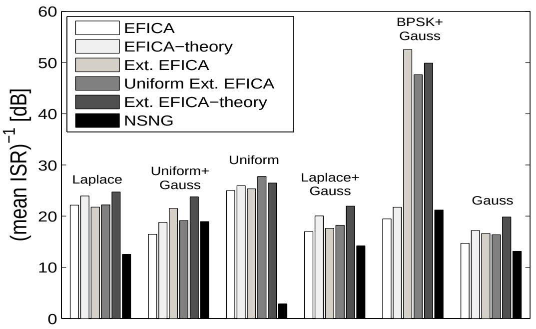

Fig.1 Results of the first experiment averaged over 100 Monte-Carlo trials.

We have compared the performance of Extended EFICA considering $M = 40$ blocks with other known ICA algorithms utilizing non Gaussianity of signals: the symmetric FastICA with the nonlinearity $g = \tanh$ and the Extended Infomax [9]. The results shown in Fig. 2 given by three different criteria averaged over 1000 independent trials demonstrate superior performance of Extended EFICA, where the average improvement against the original EFICA is 1dB in ISR and 2.5dB in SIR.

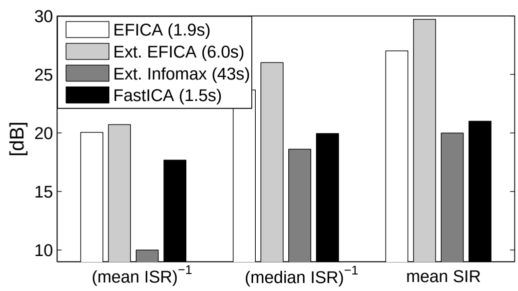

Fig.2 Results of separation of 20 speech signals. In the legend, average computational burdens on PC with 3GHz processor are shown.

7. CONCLUSIONS

An extension of the EFICA algorithm was proposed based on the piecewise stationary model. It is shown that its performance may be optimal, i.e. may achieve Cramér-Rao bound related to the model with constant-variance signals, and it yields significant improvement in separation of real-world signals.

Appendix

In corrections of [3], it is shown that the Fisher information matrix (FIM) of data obeying the model (1) is $\mathbf{F}^{\mathrm{A}} = N(\mathbf{P} + \mathbf{\Sigma})$, where $\mathbf{P}$ is a constant matrix, and $\mathbf{\Sigma}$ depends on $\kappa_1, \ldots, \kappa_d$. Since the observed data are composed of N independent observations of $\mathbf{x}$, $\mathbf{P} + \mathbf{\Sigma}$ is FIM of a single observation.

The independence of the observations holds for the piecewise model (9) as well, therefore, the FIM of data obeying the model (9) is $\mathbf{F}^{\mathrm{B}} = N(\mathbf{P} + \frac{1}{M}\sum_{I=1}^{M}\mathbf{\Sigma}^{(\mathbf{I})})$. Since the structures of $\mathbf{F}^{\mathrm{A}}$ and $\mathbf{F}^{\mathrm{B}}$ are the same, inversion of $\mathbf{F}^{\mathrm{B}}$ giving the CRLB is obtained in the same way as described in the Appendix D of [3], and (15) readily follows.

8. REFERENCES

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

[2] A. Cichocki and S.-I. Amari, *Adaptive signal and image processing: learning algorithms and applications*, Wiley, New York, 2002.

[3] P. Tichavský, Z. Koldovský, and E. Oja, "Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis", *IEEE Trans. on Signal Processing*, Vol. 54, No. 4, April 2006, see also "Corrections" to-be published ibidem, available at `http://si.utia.cas.cz/publiPT.htm`.

[4] Z. Koldovský, P. Tichavský and E. Oja, "Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound", *IEEE Trans. on Neural Networks*, Vol. 17, No. 5, Sept 2006.

[5] P. Tichavský, Z. Koldovský, and E. Oja, "Speed and Accuracy Enhancement of Linear ICA Techniques Using Rational Nonlinear Functions", *Proc. of ICA2007*, pp. 285-292, Sept. 2007.

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

[7] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming from non-Gaussian Signals", *IEE Proc.-F*, vol. 140, no. 6, pp. 362-370, Dec. 1993.

[8] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis", *IEEE Transactions on Neural Networks*, 10(3):626-634, 1999.

[9] T-W. Lee, M. Girolami and T. J. Sejnowski. "Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources", *Neural Computation*, Vol.11(2): 417-441, 1999.

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

[11] D-T. Pham, "Blind Separation of Non Stationary Non Gaussian Sources", *Proceeding of the EUSIPCO 2002 Conference*, Toulouse, France, September 2002.

[12] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, and S. Hosseini, "Performance Analysis of Extended EFICA Algorithm", *Technical report no. 2199*, ÚTIA, AV ČR, 2007. Available at http://itakura.kes.tul.cz/zbynek/publications.htm.

[13] Y. Deville, M. Benali, F. Abrard, "Differential source separation for underdetermined instantaneous or convolutive mixtures: concept and algorithms", Signal Processing, Vol. 84, Issue 10, pp. 1759-1776, Oct. 2004.

[14] [online] The Matlab^{TM} implementation of the Extended EFICA algorithm: `http://itakura.kes.tul.cz/zbynek/eefica.htm`.