Academy of Sciences of the Czech Republic
INSTITUTE OF INFORMATION THEORY AND AUTOMATION
Pod vodárenskou věží 4, 182 08 Prague 8, Czech Republic

# RESEARCH REPORT

Zbyněk Koldovský, Jiří Málek,
Petr Tichavský, Yannick Deville, and
Shahram Hosseini

## Performance Analysis of Extended EFICA Algorithm

No. 2199 October 2007

[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.

**Abstract**

This paper supports the document "Extension of EFICA Algorithm for Blind Separation of Piecewise Stationary Non Gaussian Sources."

# 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 the mixed 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 transformation $\mathbf{A}^{-1}$ up to an indeterminable order, scales, and signs of its rows.

We will assume, without loss of generality, that the original signals have unit scale (variance of RVs $s_1, \ldots, s_d$), and that the original order of signals was retrieved. The accuracy of an ICA algorithm is then characterized by variance of non-diagonal elements of the *gain* matrix $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$, i.e.

$$\text{var}[\mathbf{G}_{k\ell}] \qquad k \neq \ell,$$

with $\widehat{\mathbf{W}}$ being the estimate of $\mathbf{A}^{-1}$ resulting from the algorithm. Note that under the assumption stated above, the gain matrix should be close to the identity.

Numerous methods for separation of i.i.d. signals have been proposed [8, 9, 10] in last two decades. Some recent algorithms [4, 11] were developed to achieve accuracy that approaches the respective Cramér-Rao Lower bound (CRLB) [3]. The bound, for an unbiased estimator [15] $\widehat{\mathbf{W}}$ of $\mathbf{A}^{-1}$, is [3]

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

where $\kappa_k = \text{E}[\psi_k^2(x)]$, 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 $s_k$. An algorithm for which $N \cdot \text{var}[\mathbf{G}_{k\ell}]$ approaches the second fraction of (2) as $N \rightarrow +\infty$ is called *asymptotically efficient*.

# 2 Piecewise Stationary Model for ICA

In this section, we introduce generalization of the i.i.d. model of signals that applies in the basic ICA model (1). It will be called *piecewise stationary* model [13], and it consists in 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.

For practical purposes, we will assume that there are $M$ blocks of the same length where the distribution of the signals is unchanging. Let the length of the blocks $N/M$ be an integer due to simplicity. In that case, the i.i.d. model (1) holds within each block, i.e.,

$$\mathbf{x}^{(I)} = \mathbf{A}\mathbf{s}^{(I)} \qquad I = 1, \ldots, M \tag{3}$$

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

The separability of (1) is limited by the condition that at most one of the signals' distributions is Gaussian. This condition is reduced in case of the piecewise stationary model so that there could be at most one signal that is Gaussian in all blocks.

# 3 Cramér-Rao Lower Bound for ICA of Piecewise Stationary Signals with Unit Variance

When a vector of parameters $\theta$ is estimated from a data vector $x$ that have probability density $f_{x|\theta}(x|\theta)$, the Cramér-Rao lower bound (CRLB) is the lower bound for the variance of any unbiased estimator $\hat{\theta}$. Assume that the following Fisher information matrix (FIM) and its inversion exist:

$$\mathbf{F} = \mathrm{E}_\theta \left[ \frac{1}{f_{x|\theta}^2} \frac{\partial f_{x|\theta}(x|\theta)}{\partial \theta} \left( \frac{\partial f_{x|\theta}(x|\theta)}{\partial \theta} \right)^T \right] \tag{4}$$

Under the regularity conditions [15], it holds

$$\operatorname{cov} \hat{\theta} \geq \mathbf{F}^{-1} = \mathrm{CRLB}_\theta .$$

The derivation of FIM and its inversion is therefore necessary for derivation of the CRLB. This was done in [3] for the ICA model (1), and the results can be easily extended to the piecewise stationary model (3).

It is shown in corrections of [3] that the FIM of data consisting of $N$ samples of $\mathbf{x}$ from (1) is

$$\mathbf{F}^{\mathrm{A}} = N(\mathbf{P} + \mathbf{\Sigma}),$$

where $\mathbf{P}$ is $d^2 \times d^2$ permutation matrix, and the $mn$-th element of $\mathbf{\Sigma}$ is $\delta_{iu}\delta_{vj}[\kappa_i + \delta_{ij}(\eta_i - \kappa_i - 2)]$ for $m = (i-1)d + j$ and $n = (u-1)d + v$, where $\eta_i = \mathrm{E}[s_i^2 \psi_i^2(s_i)]$, and $\delta_{ij}$ is the Kronecker's delta. Since the observations of $\mathbf{x}$ are independent (i.i.d. model), $\mathbf{P} + \mathbf{\Sigma}$ is FIM of each single observation.

The independence of the observations holds for the model (3) as well, and the parameters, which are the elements of $\mathbf{A}$, are the same in each observation. Thus, the FIM of an observation from $I$-th block is $\mathbf{P} + \mathbf{\Sigma}^{(I)}$. From this follows that the FIM of the whole data is

$$\mathbf{F}^{\mathrm{B}} = N \left( \mathbf{P} + \frac{1}{M} \sum_{I=1}^{M} \mathbf{\Sigma}^{(I)} \right).$$

Since the structures of $\mathbf{F}^{\mathrm{A}}$ and $\mathbf{F}^{\mathrm{B}}$ are the same, readily follows that a substitution $\kappa_k \leftarrow \overline{\kappa_k}$ in (2), where

$$\overline{\kappa_k} = \frac{1}{M} \sum_{I=1}^{M} \kappa_k^{(I)},$$

gives the desired CRLB of (3), which is

$$\mathrm{CRLB}[\mathbf{G}_{k\ell}] = \frac{1}{N} \frac{\overline{\kappa_\ell}}{\overline{\kappa_k}\,\overline{\kappa_\ell} - 1} \qquad k \neq \ell. \tag{5}$$

# 4 Performance of the One-unit FastICA Algorithm

The FastICA algorithm [9] is one of the most popular methods originally developed for ICA of i.i.d. signals. A one-unit approach separates one signal from the mixture using a contrast function that measures (non) Gaussianity of the signal. Specifically, the contrast function is defined as a function of elements of a demixing vector $\mathbf{w}_k^T$

$$c(\mathbf{w}_k) = \mathrm{E}[G(\mathbf{w}_k^T \mathbf{z})], \tag{6}$$

where $G(\cdot)$ is a properly chosen nonlinear function. The vector $\mathbf{z}$ is defined as $\mathbf{Cx}$ in the way that elements of $\mathbf{z}$ are not correlated and have unit scale.

The optimization of $c(\mathbf{w}_k)$ proceeds via iteration

$$\mathbf{w}_k^+ \quad \leftarrow \quad \mathrm{E}[\mathbf{z}g({\mathbf{w}_k}^T\mathbf{z})] - \mathbf{w}_k\mathrm{E}[g'({\mathbf{w}_k}^T\mathbf{z})], \tag{7}$$

where $g(\cdot)$ and $g'(\cdot)$ are the first and the second derivative of $G(\cdot)$, respectively. When working with real samples of signals, the theoretical expectations are replaced by respective sample means. Each iteration is completed by normalizing the vector $\mathbf{w}_k^+$, and the whole is repeated until convergence is achieved.

In the end, the demixing vector $\mathbf{w}_k^T$ equals to one row of the whole demixing matrix $\widehat{\mathbf{W}}$ that separates signals $\mathbf{z}$. Which row of $\widehat{\mathbf{W}}$ corresponds to $\mathbf{w}_k^T$ depends strongly on initialization of (7), which has a bearing on indeterminacy of order of the original signals. We may therefore assume, that $\mathbf{w}_k^T$ corresponds to the $k$-th row of $\widehat{\mathbf{W}}$ for $k = 1, \ldots, d$, and we may assume that $d$ properly initialized parallel one-unit FastICA algorithms yield the whole $\widehat{\mathbf{W}}$. The demixing matrix working with signals $\mathbf{x}$ is then $\widehat{\mathbf{W}}^{1\mathrm{U}} = \widehat{\mathbf{W}} \cdot \mathbf{C}$.

Let $\mathbf{G}^{1\mathrm{U}}$ be the gain matrix defined through $\widehat{\mathbf{W}}^{1\mathrm{U}}\mathbf{A}$. Asymptotical analysis of variance of elements of $\mathbf{G}^{1\mathrm{U}}$ in [3] results in

$$\mathrm{var}[\mathbf{G}_{k\ell}^{1\mathrm{U}}] \approx \frac{1}{N}\frac{\gamma_k}{\tau_k^2} \stackrel{\text{def.}}{=} \frac{1}{N}V_{k\ell}^{1\mathrm{U}} \qquad k \neq \ell, \tag{8}$$

where

$$\begin{array}{rclcrcl} \gamma_k & = & \beta_k - \mu_k^2 & \quad & \mu_k & = & \mathrm{E}[s_k\, g(s_k)] \\ \tau_k & = & \mu_k - \nu_k & & \nu_k & = & \mathrm{E}[g'(s_k)] \\ & & & & \beta_k & = & \mathrm{E}[g^2(s_k)] \end{array} \tag{9}$$

The goal of the following subsection is to derive the variance when signals obey the piecewise stationary model, and the contrast function (6) is varied accordingly to the model.

## 4.1 One-unit FastICA for Piecewise Stationary Signals

A new definition of the contrast function (6) that takes into account the piecewise stationary model is

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

where $G_k^{(1)}, \ldots, G_k^{(M)}$ are nonlinear functions, and $\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}$ denote some weights. Note that this contrast is *not* just a linear combination of contrasts (6) with different nonlinearities, unless the distributions are the same, because each expectation value in (10) depends on distributions from corresponding block of (3). Next, each term in (10) is a valid contrast function within each block (each block can be considered for a solitary model (1)). Since the mixing matrix is the same in each block, (10) is a valid contrast function as well.

One-unit FastICA using the contrast function (10) works in the way that it applies different nonlinearity $g(\cdot)$ in the iteration (7) on each block of samples. Thus, the iteration is

$$\mathbf{w}_k^+ \leftarrow \lambda_k^{(1)} \left(\mathrm{E}[\mathbf{z}^{(1)} g_k^{(1)}(\mathbf{w}_k{}^T\mathbf{z}^{(1)})] - \mathbf{w}_k \mathrm{E}[g_k^{(1)\prime}(\mathbf{w}_k{}^T\mathbf{z}^{(1)})]\right) + \dots$$
$$\dots + \lambda_k^{(M)} \left(\mathrm{E}[\mathbf{z}^{(M)} g_k^{(M)}(\mathbf{w}_k{}^T\mathbf{z}^{(M)})] - \mathbf{w}_k \mathrm{E}[g_k^{(M)\prime}(\mathbf{w}_k{}^T\mathbf{z}^{(M)})]\right). \quad (11)$$

When considering real samples of signals, the expectation values are estimated using $N/M$ samples from the corresponding block of data. It should be remarked that (7) and (11) coincide when sample means are considered, and $\lambda_k^{(I)} = 1$ and $g_k^{(I)} = g$, for all $I = 1, \dots, M$.

## 4.2 Performance Analysis

In this section, we derive analysis of the one-unit algorithm for piecewise stationary signals. This means to derive similar result to (8) but for the case when the iteration (11) is used, and the piecewise stationary model of signals is considered. The analysis is done in the same way as in [3], and it requires the assumption that variance of signals is constant in each block, namely, we assume that the variance of RVs $s_1^{(I)}, \dots, s_d^{(I)}$ is one, for all $I = 1, \dots, M$. We stress that this is only a working assumption for the analysis and need *not* be valid in practice. The result of the analysis can be summarized in the following proposition.

**Proposition 1**
For $k = 1, \dots, d$ and $I = 1, \dots, M$, assume that

(i) RVs $s_k^{(I)}$ have zero mean and unit variance,

(ii) the functions $g_k^{(I)}$ are twice continuously differentiable,

(iii) following expectations exist

$$\begin{aligned} \gamma_k^{(I)} &= \beta_k^{(I)} - \mu_k^{(I)2} \\ \tau_k^{(I)} &= \mu_k^{(I)} - \nu_k^{(I)} \end{aligned} \qquad \begin{aligned} \mu_k^{(I)} &= \mathrm{E}[s_k^{(I)} g_k^{(I)}(s_k^{(I)})] \\ \nu_k^{(I)} &= \mathrm{E}[g_k^{(I)\prime}(s_k^{(I)})] \\ \beta_k^{(I)} &= \mathrm{E}[g_k^{(I)2}(s_k^{(I)})], \end{aligned} \quad (12)$$

and

(iv) the one-unit FastICA algorithm is started from the correct demixing matrix and stops after a single iteration (11).

Then, the normalized gain matrix elements $N^{1/2}\mathbf{G}_{k\ell}^{1\mathrm{Ug}}$ have asymptotically Gaussian distribution $\mathcal{N}(0, V_{k\ell}^{1\mathrm{Ug}})$, where

$$V_{k\ell}^{1\mathrm{Ug}} = \frac{\frac{1}{M}\sum_{I=1}^M \lambda_k^{(I)2}\beta_k^{(I)} - (\frac{1}{M}\sum_{I=1}^M \lambda_k^{(I)}\mu_k^{(I)})^2}{(\frac{1}{M}\sum_{I=1}^M \lambda_k^{(I)}\tau_k^{(I)})^2} \quad (13)$$

for $k, \ell = 1, \ldots, d$, $k \neq \ell$, provided that the denominator is nonzero.

**Proof:** See Appendix A.

The practical conclusion of this proposition is that the variance of the gain matrix elements obtained from the one-unit FastICA using iteration (11) is approximately

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

It should be noted that the assumption (iv) is not so restrictive as it seems to be. It reflects the fact that the presented analysis is “local”, and assumes a “good” initialization of the algorithm; see [3]. We close this section by the second proposition that gives optimum choice of the weights in (10) to minimize variance (13).

**Proposition 2**

For a fixed $k \in \{1, \ldots, d\}$, minimization of $V_{k\ell}^{1\mathrm{Ug}}$ subject to $\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}$ is achieved for all $\ell = 1, \ldots, d$ when

$$\lambda_k^{(J)} = \frac{1}{M} \left( \frac{\tau_k^{(J)}}{\beta_k^{(J)}} + A_k B_k \frac{\mu_k^{(J)}}{\beta_k^{(J)}} \right), \qquad J = 1, \ldots, M, \tag{15}$$

where

$$A_k = \left( \sum_{I=1}^{M} \frac{\gamma_k^{(I)}}{\beta_k^{(I)}} \right)^{-1}$$

and

$$B_k = \sum_{I=1}^{M} \frac{\mu_k^{(I)} \tau_k^{(I)}}{\beta_k^{(I)}}.$$

**Proof:** See Appendix B.

# 5 Extended EFICA Algorithm

This section briefly summarizes description of the Extended EFICA algorithm proposed in [18]. The algorithm is tailored to piecewise stationary signals obeying the model (3). It consists of the three following steps, which is similar to the original EFICA [4]:

1. Separation by the symmetric FastICA in order to obtain a preestimate of the demixing matrix $\widehat{\mathbf{W}}$.
2. 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. The simplified version of the algorithm, called *Simplified Extended EFICA*, selects all weights equal to one.
3. The refinement to get the most accurate and final estimation of the whole demixing matrix.

# 6 Notes on the Refinement Procedure

The refinement procedure is the last step both of EFICA [4] and Extended EFICA [18]. Its purpose is to combine demixing vectors resulting from optimized fine-tuning (one-unit FastICA with the criteria (6) and (10), respectively), which is similar to their symmetric orthogonalization that enforces exact orthogonality of the separated signals. However, the refinement relaxes this constraint to avoid performance limitations due to finite-samples processing [7]. This section describes several details of how the refinement is done in the Extended EFICA algorithm.

## 6.1 The Refinement in the Original EFICA

Let $\mathbf{w}_1^+, \ldots, \mathbf{w}_d^+$ be the demixing vectors resulting from the fine-tuning done by the one-unit FastICA. In the original EFICA, the entering vectors are those just after the last iteration (7), i.e., without their normalization. Then, weights $c_{k\ell}^{\mathrm{EF}}$ are computed according to

$$c_{k\ell}^{\mathrm{EF}} = \frac{|\tau_\ell|\gamma_k}{|\tau_k|(\gamma_\ell + \tau_\ell^2)}, \tag{16}$$

and the refinement of the $k$-th demixing vector is given by the $k$-th row of symmetric orthogonalization of matrix

$$\mathbf{W}_k^+ = [c_{k1}^{\mathrm{EF}}\mathbf{w}_1^+, \ldots, c_{kd}^{\mathrm{EF}}\mathbf{w}_d^+]^T, \tag{17}$$

i.e. the $k$-th row of $(\mathbf{W}_k^{+T}\mathbf{W}_k^+)^{-1/2}\mathbf{W}_k^{+T}$.

## 6.2 Alternative Definition of weights (16)

The fine-tuning in the Extended EFICA is done by the one-unit FastICA with the contrast function (10) due to the piecewise stationarity concept. This algorithm yields different performance than the original one-unit algorithm, thus, the optimum weights for the refinement should be changed accordingly. To this end, we provide a handier definition of the weights that is an explicit function of the performance of the fine-tuning.

First, (16) could be written in a form

$$c_{k\ell}^{\mathrm{EF}} = \frac{|\tau_k|}{|\tau_\ell|}\frac{V_{k\ell}^{1\mathrm{U}}}{V_{\ell k}^{1\mathrm{U}} + 1}, \tag{18}$$

where $V_{k\ell}^{1\mathrm{U}}$ defined in (7) is the term characterizing the performance of the fine-tuning. There is left to deal with the first fraction in (18) in order to get a formula that depends on the performance only. This is easily achieved when redefining the weights for normalized vectors $\mathbf{w}_1^+, \ldots, \mathbf{w}_d^+$ in the refinement.

First, we need to show that the norm of $\mathbf{w}_k^+$ equals $|\tau_k|$, $k = 1, \ldots, d$. Let ${\mathbf{w}_k}^T$ be the correct unit-norm demixing vector separating the $k$-th original signal, i.e., the $k$-th row of $\mathbf{W}$ such that $\mathbf{W}\mathbf{z} = \mathbf{s}$. Then, ${\mathbf{w}_k}^T\mathbf{z} = s_k$ and the iteration (7) gives

$$\mathbf{w}_k^+ = \mathrm{E}[\mathbf{z}g(s_k)] - \mathbf{w}_k\mathrm{E}[g'(s_k)].$$

Since the transformation $\mathbf{W}$ must be orthogonal, i.e. $\mathbf{W}^{-1} = \mathbf{W}^T$, it holds that $\mathbf{z} = \mathbf{W}^T\mathbf{s}$ and

$$\mathrm{E}[\mathbf{z}g(s_k)] = \mathbf{W}^T\mathbf{e}_k\mathrm{E}[s_kg(s_k)] = \mathbf{w}_k\mathrm{E}[s_kg(s_k)],$$

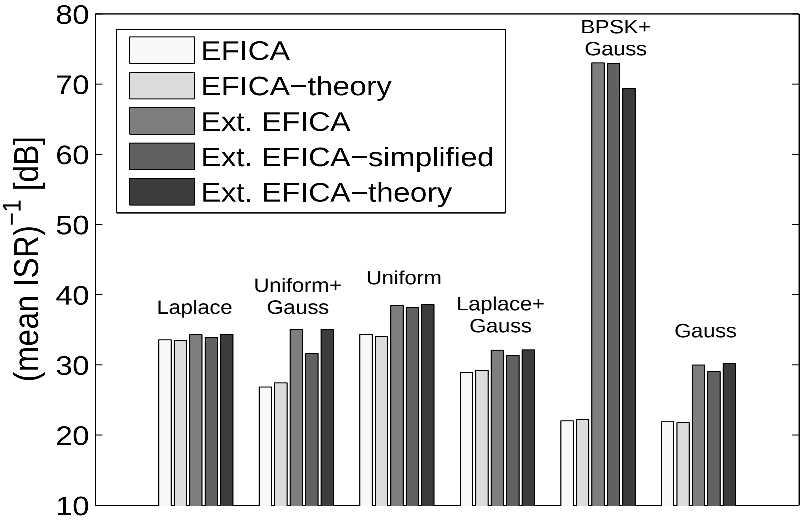


Figure 1: Results of the first experiment averaged over 100 trials.

where $\mathbf{e}_k$ is the $k$-th column of the identity matrix. Hence,

$$\mathbf{w}_k^+ = \mathbf{w}_k \mathrm{E}[s_k g(s_k)] - \mathbf{w}_k \mathrm{E}[g'(s_k)] = (\mu_k - \nu_k)\mathbf{w}_k,$$

thus the norm of $\|\mathbf{w}_k^+\| = |\mu_k - \nu_k| = |\tau_k|$.

Now, the novel definition of the weights for the refinement is

$$c_{k\ell} = \frac{V_{k\ell}^{1\mathrm{U}}}{V_{\ell k}^{1\mathrm{U}} + 1}, \tag{19}$$

with that they are working with normalized vectors $\mathbf{w}_1^+, \ldots, \mathbf{w}_d^+$. The definition of (17) should be therefore changed to

$$\mathbf{W}_k^+ = [c_{k1}\mathbf{w}_1^+/\|\mathbf{w}_1^+\|, \ldots, c_{kd}\mathbf{w}_d^+/\|\mathbf{w}_d^+\|]^T. \tag{20}$$

The equivalence of this definition with (18), which works with non normalized vectors, readily follows.

# 7 Simulation Examples

## 7.1 Demonstration of the Performance Improvements

In the first experiment we demonstrate better performance of Extended EFICA in comparison with the original EFICA when original signals obey the piecewise stationary model. We separate six artificial signals of the length $N = 10^3$ mixed by a random matrix. The signals are composed of two blocks of the same length, each of which has distribution as described in Fig. 1 with variance one. Theoretical performances, marked by "theory", fit well with the real performance, which corroborate validity of the analysis.

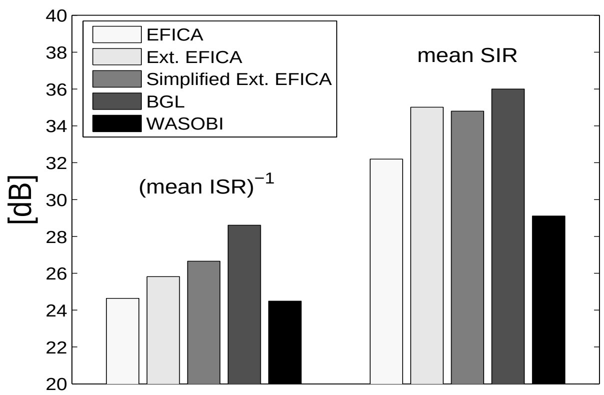


Figure 2: Results of separation of speech signals.

## 7.2 Non Stationarity versus Non Gaussianity

A model of non stationarity of signals considered in the literature for Blind Source Separation (BSS) is one that assumes Gaussian sources with time-varying variances [12]. The Gaussianity is required due to optimality of the proposed algorithm (denoted by BGL - Block Gaussian Likelihood), which jointly diagonalizes correlation matrices of each block of data. However, as the authors claim, the Gaussianity is not necessary for functionality of the method as well as the hypothesis that each observation of signals is independent of the others. On the other hand, the necessary condition for separability is the variance-envelope diversity of all original signals.

By contrast, the theoretical background of ICA algorithms using the non Gaussianity of signals does not take into account the non stationarity, namely, the variance is assumed to be constant. The same holds for Extended EFICA that is based on the piecewise stationarity concept. Luckily, this is not a must in practice analogous to the Gaussianity in the non stationarity model used by BGL.

This is well shown by the following example, where 10 speech signals were randomly taken as a piece of utterance of the length $N = 5000$ from a database[1], mixed, and separated by several algorithms. Averaged results over 1000 independent trials measured by residual Interference-to-Signal Ratio (ISR) and the reciprocal Signal-to-Interference Ratio are shown in Figure 2. For completion, algorithm WASOBI [16] that is based on the third most common model for BSS, i.e. AR Gaussian model of signals, is included.

The algorithm BGL achieves the best results thanks to strong non stationarity of speech. However, the other algorithms achieve comparable results although the non-stationarity is not considered in their theoretical framework. A simplified explanation for this phenomenon is as follows:

> *The ability of EFICA and yet more of the Extended EFICA to profit from the non stationarity of signals consists in that the shape of the distribution (or of its model) varies under hypothesis of unit variance.*

This intuition can be supported by the following example. Figure 3 shows two Gaussian

[1]The database is available at http://itakura.kes.tul.cz/zbynek/downloads.htm.

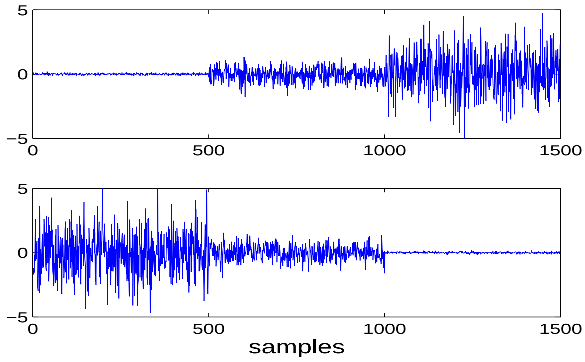


Figure 3: Gaussian signals with different variance in each of three blocks.

signals consisting of three blocks, each of which has variance, respectively, 0.001, 0.1, and 0.899 for the first signal, and in the opposite order for the second signal. Using the concept of piecewise stationarity with constant variance, the distribution of the first "silent" block of the first signal seem to be strongly super-Gaussian, while the last "loud" block seem to be sub-Gaussian. In other words, the signal may seem to be non-Gaussian in two blocks; see the details of histograms in Figure 4.

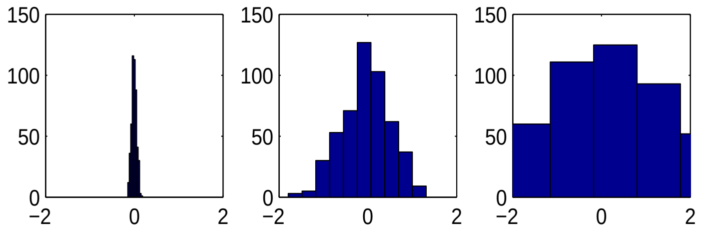


Figure 4: Histograms of three blocks of the first signals from Figure 2 viewed with a fixed range on x-axis.

In the last example, we show results of separation of two signals of the length $N = 1000$. The first signal was Gaussian with variance one. The first and the second half of the second signal was Gaussian with variance one and Laplacean with variance $\sigma^2 \in [0.1, 1]$, respectively. Thus, the second signal is non Gaussian, and the parameter $\sigma$ controls the (non)stationarity of the signal.

Results averaged over 100 trials for each $\sigma$ with a random mixing matrix are shown in Figure 5. For $\sigma$ close to $0.1$, the second signal is strongly non stationary, while for $\sigma$ close to one, both signals are stationary, but the second one is non Gaussian. Therefore, the Extended EFICA achieves superior results thanks to the piecewise stationarity concept.

# 8 Conclusions

This paper provides proofs of two propositions that describe performance of one-unit FastICA algorithm for piecewise stationary signals, and optimum choice of weights in the contrast functions. Further simulation examples are included to demonstrate strengths of the novel Extended

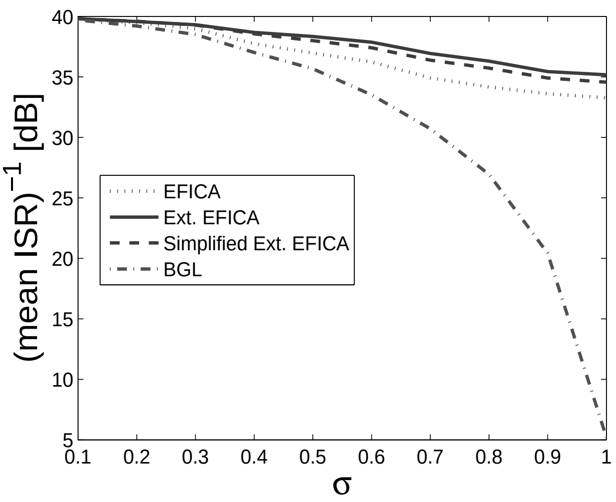


Figure 5: Results of separation of one Gaussian signal and one Gaussian-Laplacean non stationary signal, where the non stationarity is driven by parameter $\sigma$.

EFICA Algorithm.

# Appendix A - Proof of Proposition 1

In this section, we will follow the proof from Appendix A of [3]. From here on we will use different notations than in the rest of the paper, because the vectors of symbols should be replaced by vectors/matrices of samples. Adopted from [3], $\mathbf{s}_k$ will be $N \times 1$ vector of samples of the $k$-th original signal with the difference that the $I$-th block of $M$ samples is distributed according to RV $s_k^{(I)}$. The vector $\mathbf{u}_k$ equals normalized $\mathbf{s}_k$ to the unit scale; similarly $\mathbf{z}_k$ and $\mathbf{x}_k$ denote samples of the respective signals. The nonlinearity $g(\cdot)$ used for the $k$-th signal will be distinguished by the subscript $k$, i.e. $g_k(\cdot)$. It applies to the vectors element-wise, but so that to the $I$-th block of $M$ elements, it applies function $\lambda_k^{(I)} g_k^{(I)}(\cdot)$.

Now, using the third assumption of the proposition 1 (12), the equations (40) and (41) from [3] change, respectively, to

$$N^{-1}\mathbf{s}_k^T g_k(\mathbf{s}_k) \quad \xrightarrow{p} \quad \frac{1}{M}\sum_{I=1}^{M} \lambda_k^{(I)} \mu_k^{(I)} \tag{21}$$

$$N^{-1}g_k'^T(\mathbf{s}_k)\mathbf{1}_N \quad \xrightarrow{p} \quad \frac{1}{M}\sum_{I=1}^{M} \lambda_k^{(I)} \nu_k^{(I)} \tag{22}$$

Note that $\nu$ denotes the same quantities that are in [3] denoted by $\rho$, and $\mathbf{1}_N$ stands for $N \times 1$ vector of ones.

Using this, all the following equations (42)-(74) in [3] change according to the substitutions $\mu_k \leftarrow \frac{1}{M}\sum_{I=1}^{M} \lambda_k^{(I)} \mu_k^{(I)}$ and $\rho_k \leftarrow \frac{1}{M}\sum_{I=1}^{M} \lambda_k^{(I)} \nu_k^{(I)}$. The only exception is the equation (62), which should be recomputed, and it gives

$$\mathrm{E}[(\mathbf{g}_k^T\mathbf{u}_\ell)^2] = \mathrm{E}[\mathbf{g}_k^T\mathbf{g}_k] = N \cdot \frac{1}{M}\sum_{I=1}^{M} \lambda_k^2 \beta_k^{(I)}, \tag{23}$$

where $\mathrm{g}_k$ is the simplified notation of $g_k(\mathbf{u}_k)$. Recomputation of (75) in [3] using the above substitutions readily yields (13). ■

# Appendix B - Proof of Proposition 2

The criterion (13) can be written in the form

$$V_{k\ell}^{1\mathrm{Ug}} = \frac{\mathbb{l}_k^T \mathbf{\Gamma}_k \mathbb{l}_k}{\mathbb{l}_k^T \mathbb{t}_k \mathbb{t}_k^T \mathbb{l}_k} \tag{24}$$

with

$$\mathbb{l}_k = [\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}]^T \tag{25}$$

$$\mathbf{\Gamma}_k = \mathrm{diag}[\beta_k^{(1)}, \ldots, \beta_k^{(M)}] - \frac{1}{M}\mathbb{m}_k \mathbb{m}_k^T \tag{26}$$

$$\mathbb{m}_k = [\mu_k^{(1)}, \ldots, \mu_k^{(M)}]^T \tag{27}$$

$$\mathbb{t}_k = [\tau_k^{(1)}, \ldots, \tau_k^{(M)}]^T \tag{28}$$

The goal is to minimize (24) subject to elements of $\mathbb{l}_k$, which is equivalent with maximizing

$$\max_{\mathbb{l}_k} \frac{\mathbb{l}_k^T \mathbb{t}_k \mathbb{t}_k^T \mathbb{l}_k}{\mathbb{l}_k^T \mathbf{\Gamma}_k \mathbb{l}_k}. \tag{29}$$

Let $\mathbb{y}_k = \mathbf{\Gamma}_k^{1/2}\mathbb{l}_k$, where the matrix $\mathbf{\Gamma}_k^{1/2}$ obeying $\mathbf{\Gamma}_k^{1/2}\mathbf{\Gamma}_k^{1/2} = \mathbf{\Gamma}_k$ exists thanks to positive semidefiniteness of $\mathbf{\Gamma}_k$ ($V_{k\ell}^{1\mathrm{Ug}}$ denotes variance, which must be always nonnegative). Since (24) is invariant subject to nonzero multiple of $\mathbb{l}_k$, we can introduce a constraint $\|\mathbb{l}_k\| = \text{const.}$, and (29) can be written in the form of classical eigenvalue problem

$$\max_{\|\mathbb{y}_k\|=1} \frac{\mathbb{y}_k^T \mathbf{\Gamma}_k^{-1/2} \mathbb{t}_k \mathbb{t}_k^T \mathbf{\Gamma}_k^{-1/2} \mathbb{y}_k}{\mathbb{y}_k^T \mathbb{y}_k}. \tag{30}$$

The rank of the matrix $\mathbf{\Gamma}_k^{-1/2}\mathbb{t}_k\mathbb{t}_k^T\mathbf{\Gamma}_k^{-1/2}$ is one, thus, the eigenvector corresponding to the only nonzero eigenvalue, i.e. the solution of (30), is $\mathbb{y}_k = \mathbf{\Gamma}_k^{-1/2}\mathbb{t}_k$. Hence, $\mathbb{l}_k$ that minimizes (24) is

$$\mathbb{l}_k = \mathbf{\Gamma}_k^{-1}\mathbb{t}_k. \tag{31}$$

Using the matrix inversion lemma for computation of $\mathbf{\Gamma}_k^{-1}$, (15) follows. ■

# 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, "On the performance of orthogonal source separation algorithms", *Proc. EUSIPCO*, pp. 776-779, Edinburgh, September 1994.

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

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

[10] 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.

[11] 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.

[12] D-T. Pham and J-F. Cardoso. "Blind separation of instantaneous mixtures of non stationary sources", *IEEE Trans. Signal Processing*, pp. 1837-1848, vol. 49, no. 9, 2001.

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

[14] J.-F. Cardoso and D. T. Pham, "Separation of non stationary sources. Algorithms and performance.", in *Independent Components Analysis: Principles and Practice*, pp. 158 - 180. S. J. Roberts and R. M. Everson (editors), Cambridge University Press, 2001.

[15] R. C. Rao, *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York, 1973.

[16] A. Yeredor, “Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting,” *IEEE Signal Processing Letters*, vol. 7, pp. 197-200, Jul. 2000.

[17] 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.

[18] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, and S. Hosseini, “Extension of EFICA Algorithm for Blind Separation of Piecewise Stationary Non Gaussian Sources”, submitted to a conference, Oct. 2007.