Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound

Zbyněk Koldovský[1,2], Petr Tichavský[2] and Erkki Oja[3]

[1] Faculty of Nuclear Sciences and Physical Engineering
Czech Technical University, Trojanova 13, 120 00 Prague 2

[2] Institute of Information Theory and Automation, Pod vodárenskou věží 4,
P.O. Box 18, 182 08 Praha 8, Czech Republic
E-mail: {koldovsk,tichavsk}@utia.cas.cz

[3] Neural Networks Research Centre, Helsinki University of Technology,
P.O. Box 5400, 02015 HUT, Finland
E-mail: Erkki.Oja@hut.fi

Abstract— **FastICA is one of the most popular algorithms for Independent Component Analysis, demixing a set of statistically independent sources that have been mixed linearly. A key question is how accurate the method is for finite data samples. We propose an improved version of the FastICA algorithm which is asymptotically efficient, i.e., its accuracy given by the residual error variance attains the Cramér-Rao lower bound. The error is thus as small as possible. This result is rigorously proven under the assumption that the probability distribution of the independent signal components belongs to the class of generalized Gaussian distributions with parameter α, denoted GG(α) for $\alpha > 2$. We name the algorithm EFICA. Computational complexity of a Matlab^{TM} implementation of the algorithm is shown to be only slightly (about three times) higher than that of the standard symmetric FastICA. Simulations corroborate these claims and show superior performance of the algorithm compared with JADE and Non-Parametric ICA on separating sources with distribution GG(α) with arbitrary α, as well as on sources with bi-modal distribution, and a good performance in separating linearly mixed speech signals.**

Keywords— **Independent component analysis, blind source separation, blind deconvolution, Cramér-Rao lower bound, algorithm FastICA**

I. Introduction

Recently, blind techniques such as blind source separation have become popular in the signal processing and machine learning community. One of the central tools for this problem is Independent Component Analysis (ICA) [2], [3]. In this technique, a set of original source signals are retrieved from their mixtures based on the assumption of their mutual statistical independence. The simplest case for ICA is the instantaneous linear noiseless mixing model. In this case, the mixing process can be expressed as

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

where $\mathbf{X}$ is an $d \times N$ data matrix. The rows of $\mathbf{X}$ are the observed mixed signals, thus d is the number of mixed signals and N is their length or the number of samples in each signal. Similarly, the unknown $d \times N$ matrix $\mathbf{S}$ includes samples of the original source signals. $\mathbf{A}$ is an unknown regular $d \times d$ mixing matrix. The mixing matrix is assumed to be square, for simplicity. In the noiseless scenario, this is not a restriction, because if more mixtures than sources are available, they would not improve the signal recovering.

The additive noise is not considered in this paper. Additional experiments that exceed the scope of the paper show, however, that the hereproposed algorithm EFICA outperforms classical FastICA and other algorithms in the noisy scenario, also [28].

A basic assumption in ICA is that the elements of $\mathbf{S}$, denoted s_{ij}, are mutually independent i.i.d. random variables with probability density functions (pdf) $p_i(s_{ij})$ $i = 1, \ldots, d$. The row variables s_{ij} for all $j = 1, \ldots, N$, having the same density, are thus an i.i.d. sample of one of the independent sources denoted by s_i. The key assumptions for the identifiability [1] of the model (1), or solving both $\mathbf{A}$ and $\mathbf{S}$ up to some simple ambiguities, are that all but at most one of the densities $p_i(\cdot)$ are non-Gaussian, and the unknown matrix $\mathbf{A}$ has full rank. In the following, let $\mathbf{W}$ denote the demixing matrix, $\mathbf{W} = \mathbf{A}^{-1}$.

The basic ICA problem and its extensions and applications have been studied widely and many algorithms have been developed. One of the main differences is how the unknown density functions $p_i(\cdot)$ of the original signals are estimated or replaced by suitable nonlinearities in the ICA contrast functions. Non-Gaussianity is the key property. For instance, JADE [5] is based on the estimation of kurtosis via cumulants, NPICA [11] uses a nonparametric model of the density functions, and RADICAL [12] uses an approximation of the entropy of the densities based on order statistics. The FastICA algorithm uses either kurtosis [6] or other measures of

[0]Part of this paper was presented at the IEEE SSP'05 Workshop in Bordeaux. This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and by the Czech Technical University in Prague through the project CTU0508214.

non-Gaussianity in entropy approximations in the form of suitable nonlinear functions $G(\cdot)$ [7].

In spite of the success of ICA in solving even large-scale real world problems, some theoretical questions remain partly open. One of the most central questions is the theoretical accuracy of the developed algorithms. Mostly the methods are compared through empirical studies, which may demonstrate the efficacy in various situations. However, the general validity cannot be proven like this. A natural question is, whether it is possible to reach a Cramér Rao lower bound for separation performance, which is widely accepted as measure of efficiency of estimators.

Many of the algorithms can be shown to converge in theory to the correct solution giving the original sources, under the assumption that the sample size N is infinite. This is unrealistic. For finite data sets, like in the model (1), what typically happens is that the sources are not completely unmixed but some traces of the other sources remain in them. This means that the obtained demixing matrix $\widehat{\mathbf{W}}$ is not exactly the inverse of $\mathbf{A}$, and the matrix of estimated sources $\mathbf{Y} = \widehat{\mathbf{W}}\mathbf{X} = \widehat{\mathbf{W}}\mathbf{A}\mathbf{S}$ is only approximately equal to $\mathbf{S}$. A natural measure of error is the deviation of matrix $\widehat{\mathbf{W}}\mathbf{A}$ from the unit matrix, i.e., the variances of its elements.

The present authors published recently an asymptotic performance analysis of the FastICA algorithm in [8], deriving the exact expression for this error variance. Furthermore, it is compared with the Cramér-Rao lower bound (CRB) for ICA [4], [9], [17], [19], [20], [21], [26], [27] and showed that the accuracy of FastICA is very close, but not equal to, the CRB. The condition for this is that the nonlinearity $G(\cdot)$ in the FastICA contrast function is the integral of the score function $\psi(s)$ of the original signals, or the negative log density, i.e.,

$$G(s) = \int \psi(s)ds = -\int \frac{p_i'(s)}{p_i(s)}ds = -\log p_i(s). \quad (2)$$

The purpose of this paper is to use this analysis to generalize the FastICA algorithm to an improved version so that it becomes asymptotically Fisher efficient, i.e., that the residual error variance becomes equal to the CRB. When the asymptotic performance achieves the CRB we have reached the absolute accuracy that cannot be improved. We call this new variant EFICA (Efficient FastICA).

As far as the authors know, there have been few efforts in developing concrete ICA algorithms based on the concept of asymptotic efficiency. A notable approach in this direction is the method of Estimating functions [23], [24], matrix valued functions $\mathbf{F}(\mathbf{X}, \mathbf{W})$ such that their root $\widehat{\mathbf{W}}(\mathbf{X})$ is an estimator for the true demixing matrix $\mathbf{W}$. Amari and Cardoso [24] derived an optimal class of estimating functions whose roots $\widehat{\mathbf{W}}$ are Fisher-efficient. The theory is general and can be applied to stochastic approximation-type learning algorithms.

Another related paper is [21], which studies asymptotic performance of so-called quasi-maximum likelihood estimate (quasi-MLE). In this method, the true probability density function of sources is replaced by an ad-hoc model density. If the model density coincides with the true density, an asymptotically efficient estimate would be obtained.

The contents of this paper are as follows. In the next Section 2, the results of our previous work are briefly summarized. In Section 3, the improved algorithm is derived and its properties are described. In order to demonstrate the efficiency in practice, Section 4 presents computer simulations. The simulations confirm the excellent performance of the EFICA algorithm and also show that the computational complexity (measured on a Matlab^{TM} implementation) is only about three times that of standard FastICA, which is one of the fastest ICA algorithms. Finally, in Section 5 conclusions are given.

II. Recent results

A. The original FastICA algorithm

The algorithm FastICA was introduced in [6], [7] in two versions: a one-unit approach and a symmetric one. The first preprocessing step, which is common for both versions and for many other ICA algorithms, consists of removing the sample mean and decorrelating the data $\mathbf{X}$, i.e.,

$$\mathbf{Z} = \widehat{\mathbf{C}}^{-1/2}(\mathbf{X} - \overline{\mathbf{X}}) \quad (3)$$

where $\widehat{\mathbf{C}}$ is the sample covariance matrix, $\widehat{\mathbf{C}} = (\mathbf{X} - \overline{\mathbf{X}})(\mathbf{X} - \overline{\mathbf{X}})^T/N$ and $\overline{\mathbf{X}}$ is the sample mean of the mixture data. Now matrix $\mathbf{Z}$ contains the whitened mixtures. The one-unit FastICA algorithm is based on minimization/maximization of the criterion $c(\mathbf{w}) = \mathrm{E}[G(\mathbf{w}^T\mathbf{Z})]$. There $\mathbf{w}$ is the unitary vector of coefficients to be found that separates one of the independent components from the mixture $\mathbf{Z}$ (one row of the separating matrix $\widehat{\mathbf{W}}(\mathbf{Z})$). Function $G(\cdot)$ is a suitable nonlinearity, called contrast function [1], applied elementwise to the row vector $\mathbf{w}^T\mathbf{Z}$; see [2]. The symbol E stands for the sample mean over the elements of the row vector. It is not known in advance which component is being estimated: this mainly depends on the initialization.

To retrieve all the original components, different rows of $\widehat{\mathbf{W}}(\mathbf{Z})$ are estimated under the orthogonality condition, i.e. $\widehat{\mathbf{W}}\widehat{\mathbf{W}}^T = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix. In the one-unit deflation method, the independent components are found one by one, and the weight vector $\mathbf{w}$ is always constrained to be orthogonal to the previously found ones. In the symmetric FastICA, the condition is ensured via a symmetric orthogonalization after parallel one-unit iterations:

$$\mathbf{W}^+ \leftarrow g(\mathbf{W}\mathbf{Z})\mathbf{Z}^T - \mathrm{diag}[g'(\mathbf{W}\mathbf{Z})\mathbf{1}_N]\,\mathbf{W} \quad (4)$$

$$\mathbf{W} \leftarrow (\mathbf{W}^+\mathbf{W}^{+T})^{-1/2}\mathbf{W}^+. \quad (5)$$

There $g(\cdot)$ and $g'(\cdot)$ denote the first and the second derivatives of $G(\cdot)$, respectively, applied elementwise, and $\mathbf{1}_N$ stands for an $N \times 1$ vector of 1's.

The FastICA algorithm is computationally light, robust, and converges very fast. It is available in public-domain software [13]. Recently, it was proposed to complete the symmetric FastICA by a test of saddle points

that eliminates convergence to side minima of the cost function, which may occur for some nonlinearities $G(\cdot)$ [8].

An essential question is the residual error of the algorithm, due to a finite sample of the mixture signals. Let $\widehat{\mathbf{W}}$ be the estimate of the demixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ obtained with the FastICA algorithm. The separation quality is analyzed by means of the so called gain matrix, $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$. Theoretically, $\mathbf{G}$ is the unit matrix, but for finite sample sizes there is residual error. The elements of $\mathbf{G}$ characterize the relative remaining presence of the $j-$th original signal component in the estimated $k-$th component, $j, k = 1, \ldots, d$. In the following, their asymptotic variance is compared to the theoretically optimal one.

B. Analysis of FastICA and the Cramér-Rao bound

Let $\mathbf{G}^{1U}$ and $\mathbf{G}^{SYM}$, respectively, be the gain matrix obtained by the one-unit and the symmetric variant of FastICA using a nonlinear function $g(\cdot)$. The main result shown in [8] was the following: Assume that the original signals s_k in the mixture have zero mean and unit variance, that g is sufficiently smooth, and the following expectations exist:

$$\mu_k \stackrel{\text{def}}{=} \mathrm{E}[s_k g(s_k)] \tag{6}$$

$$\rho_k \stackrel{\text{def}}{=} \mathrm{E}[g'(s_k)] \tag{7}$$

$$\beta_k \stackrel{\text{def}}{=} \mathrm{E}[g^2(s_k)] \tag{8}$$

Then the normalized gain matrix elements $N^{1/2}\mathbf{G}^{1U}_{k\ell}$ and $N^{1/2}\mathbf{G}^{SYM}_{k\ell}$ have asymptotically Gaussian distributions $\mathcal{N}(0, V^{1U}_{k\ell})$ and $\mathcal{N}(0, V^{SYM}_{k\ell})$, with variances

$$V^{1U}_{k\ell} = \frac{\gamma_k}{\tau_k^2} \tag{9}$$

$$V^{SYM}_{k\ell} = \frac{\gamma_k + \gamma_\ell + \tau_\ell^2}{(\tau_k + \tau_\ell)^2}, \tag{10}$$

where

$$\gamma_k = \beta_k - \mu_k^2 \tag{11}$$

$$\tau_k = |\mu_k - \rho_k|. \tag{12}$$

The variances in (9,10) are minimized if the function $g(s)$ equals the score function

$$\psi_k(s) = -\frac{d}{ds}\log p_k(s) = -\frac{p_k'(s)}{p_k(s)} \tag{13}$$

of the corresponding source distribution $p_k(s)$. The minimum variance can be shown to be close, but not to coincide with the Cramér-Rao lower bound (CRB) derived in [4]. It was shown that the CRB is

$$\mathrm{CRB}(\mathbf{G}_{k\ell}) = \frac{1}{N}\frac{\kappa_k}{\kappa_k\kappa_\ell - 1} \tag{14}$$

where $\kappa_k \stackrel{\text{def}}{=} \mathrm{E}\left[\psi_k^2(s_k)\right]$.

III. Efficient FastICA: EFICA

The proposed efficient version of FastICA is based on the following observations: (i) the symmetric FastICA algorithm can be run with different nonlinearity for different sources. (ii) In the symmetrization step of each iteration, it is possible to introduce an auxiliary constants, that can be tuned to minimize mean square estimation error in one (say $k-$th) row of the estimated de-mixing matrix. These estimations can be performed in parallel for all rows - to obtain an estimate of the whole de-mixing matrix, that achieves the corresponding CRB, if the nonlinearities correspond to score functions of the sources. (iii) The algorithm remains to be asymptotically efficient (attaining the CRB) if the theoretically optimum auxiliary constants in the algorithm are replaced by their consistent estimates.

The proposed algorithm EFICA models all independent signals as they have a generalized Gaussian distribution with appropriate parameters α's. The algorithm is summarized in Figure 1. Note that the output is not constrained, unlike symmetric FastICA, in the sense that the separated components need not have exactly zero sample correlations.

In order to explain the proposed algorithm in more details, the notion of "generalized symmetric FastICA" is introduced, and its efficiency is studied in the next subsection. The algorithm EFICA will be presented in detail in subsection B.

A. Generalizing the symmetric FastICA to attain the Cramér-Rao bound

Consider now a version of the symmetric version of FastICA where two changes have been made.

First, as it is not possible to attain the CRB if only one nonlinearity $g(\cdot)$ is used, different nonlinear functions $g_k(\cdot)$, $k = 1, 2, \ldots, d$ will be used for estimation of each row of $\mathbf{W}^+$. Denote $g(\mathbf{WZ}) = [g_1(\mathbf{w}_1^T\mathbf{Z}), g_2(\mathbf{w}_2^T\mathbf{Z}), \ldots, g_d(\mathbf{w}_d^T\mathbf{Z})]^T$ where again each function g_j is applied elementwise. The function $g'(\mathbf{WZ})$ is defined likewise. Eventually, the functions $g_k(\cdot)$ should be the score functions of the sources $\mathbf{s}_k$.

Second, the first step (4) of the iteration will be followed by multiplying each row of $\mathbf{W}^+$ with a suitable positive number c_i $i = 1, \ldots, d$ before the symmetric orthogonalization (5). This will change the length (norm) of each row, which will affect the orientations of the rows after orthonormalization.

One iteration of the new Generalized symmetric FastICA algorithm, with the new definition of $g(\cdot)$, can then be written in three steps:

$$\mathbf{W}^+ \leftarrow g(\mathbf{WZ})\mathbf{Z}^T - \mathrm{diag}[g'(\mathbf{WZ})\mathbf{1}_N]\,\mathbf{W} \tag{15}$$

$$\mathbf{W}^+ \leftarrow \mathrm{diag}[c_1, \ldots, c_d] \cdot \mathbf{W}^+ \tag{16}$$

$$\mathbf{W} \leftarrow (\mathbf{W}^+\mathbf{W}^{+T})^{-1/2}\mathbf{W}^+ \tag{17}$$

This algorithm can be analyzed in the same way as the plain symmetric FastICA in [8] using a simple substitution $\mathbf{W}^+_{k\ell} \leftarrow c_k\mathbf{W}^+_{k\ell}$. The result is that the non-diagonal normalized gain matrix elements for this

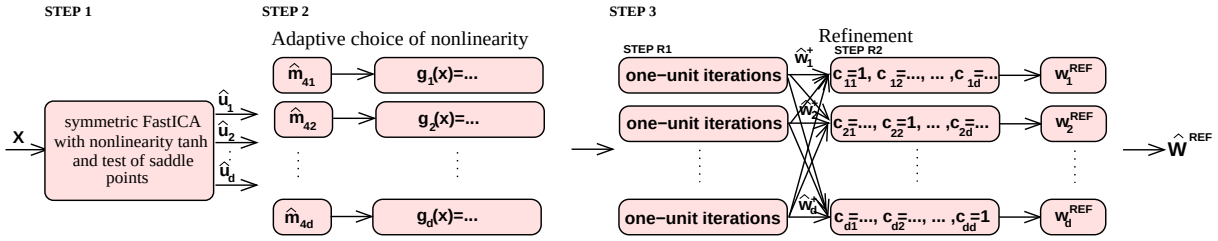


Fig. 1. Flow of the proposed algorithm EFICA.

method, $N^{1/2}\mathbf{G}_{k\ell}^{GS}$, have asymptotically Gaussian distribution $\mathcal{N}(0, V_{k\ell}^{GS})$, where

$$V_{k\ell}^{GS} = \frac{c_k^2\gamma_k + c_\ell^2(\gamma_\ell + \tau_\ell^2)}{(c_k\tau_k + c_\ell\tau_\ell)^2}. \tag{18}$$

Note that the different choice of nonlinearity for each row of $\mathbf{W}$ also changes the definitions (6)-(8) for μ_k, ρ_k, and β_k in the sense that g in those definitions must be replaced by g_k. The definitions for γ_k and τ_k in (11),(12) remain the same but now they, also, depend on their own nonlinearity $g_k(\cdot)$.

The key thing here is that, since $c_1, \ldots, c_d$ are arbitrary positive numbers, the criterion (18) characterizing the asymptotic error variance of the algorithm can be optimized in terms of these free parameters.

The properties of (18) are as follows:

- The variance is invariant with respect to multiplying all parameters c_k by the same factor. Therefore, for a fixed k let c_k be chosen, without any loss of generality, as $c_k = 1$. Let us change the notation so that instead of c_ℓ, we use $c_{k,\ell}$ to denote the other parameters in the case that $c_k = 1$.
- Minimization of (18) with respect to $c_{k,\ell}$, $\ell \neq k$, can be performed analytically in a straightforward way, and it gives

$$c_{k,\ell}^{OPT} = \arg\min_{c_\ell, c_k=1} V_{k\ell}^{GS} = \frac{\tau_\ell\gamma_k}{\tau_k(\gamma_\ell + \tau_\ell^2)}\ . \tag{19}$$

- Combining (18) and (19), the optimum value of the criterion is

$$V_{k\ell}^{OPT} = \min_{c_\ell, c_k=1} V_{k\ell}^{GS} = \frac{\gamma_k(\gamma_\ell + \tau_\ell^2)}{\tau_\ell^2\gamma_k + \tau_k^2(\gamma_\ell + \tau_\ell^2)}. \tag{20}$$

However, note that the optimum choice (19) can be done for a fixed k only, thus, the method implemented via (15)-(17) attains the minimized $V_{k\ell}^{GS}$ (20) for the selected k only, not for all k, ℓ simultaneously. Moreover, optimization of $V_{k\ell}^{GS}$ separately for each k alleviates the orthogonal constraint, and thus, delimits performance of the algorithm [18].

- In the special case that $g_k = \psi_k$, i.e., the k-th nonlinearity equals the score function of the $k-$th signal for each $k = 1, \ldots, d$, it is easy to show that $\beta_k = \rho_k = \kappa_k$, $\mu_k = 1$, and $\tau_k = \gamma_k = \kappa_k - 1$. Then (20) simplifies to

$$V_{k\ell}^{OPT} = \frac{\kappa_k}{\kappa_k\kappa_\ell - 1} = N\,\mathrm{CRB}(\mathbf{G}_{k\ell}). \tag{21}$$

This means that in this special case the estimator attains the CRB for $N \to \infty$.

In practice the true score functions are rarely known in advance, and the generalized symmetric FastICA has only a theoretical meaning. It can be proved, however, that the asymptotic efficiency of the algorithm is maintained if the score functions and the optimum coefficients $c_{k,\ell}^{OPT}$ are replaced by their consistent estimates.

For the consistent estimation, it is necessary to have a consistent initial estimate of the mixing or de-mixing matrix. The ordinary symmetric FastICA is one possible choice. Second, one needs a consistent estimate of the score functions computed from the sample distribution functions of the components. This is a widely studied task, and numerous approaches have been developed either parametric [21] or nonparametric [12], [11], [22]. Note, however, that not every score function can serve a suitable nonlinearity for use in FastICA iteration. Suitable nonlinearity must be continuous and differentiable.

B. Proposed algorithm

In this subsection, an algorithm, called for brevity "EFICA" is proposed, which combines the idea of the generalized symmetric FastICA with an adaptive choice of the function g, which is base on modelling of the distribution of the independent component by generalized Gaussian distribution [15].

The algorithm consists of three steps:

1. Running the original symmetric FastICA until convergence using a standard nonlinearity such as $g(s) = \tanh(s)$.
2. Adaptive choice of different nonlinearities g_k to estimate the score functions of the found sources, based on the outcome of Step 1.
3. A refinement or fine-tuning for each of the found source components by one-unit FastICA, using the nonlinearities found in Step 2, and another fine-tuning using the optimal c_k parameters as in (16),(17).

For easy reference, the outline of the algorithm is summarized in Figure 1. The three steps are now introduced in detail.

B.1 Step 1: Running the symmetric FastICA until convergence

The purpose of Step 1 is to quickly and reliably get preliminary estimates of the original signals. In this step, therefore, the optional nonlinearity in the original symmetric FastICA $g(s) = \tanh(s)$ is used due to its universality, but other possibilities seems to give promising results as well, e.g. $g(s) = \frac{s}{1+s^2}$. Also the test for saddle

points as introduced in [8] is performed to get reliable source estimates.

B.2 Step 2: Adaptive choice of the nonlinearities

Assume that $\widehat{\mathbf{u}}_k$ is the $k-$th estimated independent signal obtained in Step 1.

In many real situations, the distributions of the signals are unimodal and symmetric. in this paper, we focus on a parametric choice of g_k that works well for the class of generalized Gaussian distributions with parameter α, denoted GG(α). This class covers a wide variety of typical distributions including standard Gaussian and Laplacean distributions for $\alpha = 2$ and $\alpha = 1$, respectively, a uniform distribution in the limit as $\alpha \to \infty$, and heavy-tailed distributions for $\alpha \to 0+$ (see Appendix B for the definition).

The score function of the GG(α) distribution is proportional to[1]

$$g(x) = \text{sign}(x) \cdot |x|^{\alpha-1}. \tag{22}$$

A problem with the score function of the GG(α) distribution is that it is not continuous for $\alpha \leq 1$ and thus it is not a valid nonlinearity for FastICA. For these α's the statistical efficiency cannot be achieved by the algorithm using this score function.

Let us look separately at the subgaussian ($\alpha > 2$) and supergaussian ($\alpha < 2$) cases.

Subgaussian case
We propose to use the function $g_k(x) = \text{sign}(x) \cdot |x|^{\alpha-1}$ for subgaussian signals, i.e. for $\alpha > 2$. In this case the parameter α can be well estimated by fitting the theoretical fourth-order moment of the GG(α) distribution

$$m_4(\alpha) = \Gamma\left(\frac{1}{\alpha}\right)\Gamma\left(\frac{5}{\alpha}\right)\Gamma^{-2}\left(\frac{3}{\alpha}\right), \tag{23}$$

with the sample fourth-order moment of the k-th signal

$$\widehat{m}_{4k} = \mathbf{1}_N^T \widehat{\mathbf{u}}_k^{\cdot 4}/N. \tag{24}$$

There $\{\cdot\}^{\cdot 4}$ denotes the elementwise fourth order power. In (23), Γ is the Gamma function. The sample fourth moment indicates well the subgaussianity ($\widehat{m}_{4k} < 3$) or supergaussianity ($\widehat{m}_{4k} > 3$) of the signal and can be used to drive the choice of the nonlinearity g_k.

In Appendix A an asymptotic inversion of (23) at point $\alpha = +\infty$ is performed, giving an estimation

$$\widehat{\alpha}_k = [\nu_1\sqrt{\widehat{m}_{4k} - 1.8} - \nu_2(\widehat{m}_{4k} - 1.8)]^{-1} \tag{25}$$

with $\nu_1 \approx 0.2906$ and $\nu_2 \approx 0.1851$. This is valid for $\widehat{m}_{4k} > \lim_{\alpha\to\infty} m_4(\alpha) = 9/5 = 1.8$. If $\widehat{m}_{4k} \leq 1.8$ or $\widehat{\alpha}_k > 15$, maximum power 14 in function g_k is used to maintain the stability of the algorithm.

Supergaussian case
If the sources have the distribution GG(α) with $\alpha < 2$, the score function (22) appears not to be so useful as the nonlinearity in FastICA. Instead, we suggest to study an ad-hoc choice

$$g_\eta(x) = x\exp\{-\eta|x|\} \tag{26}$$

where η is a free parameter. The intention is to find an optimum choice of η, which would minimize the asymptotic variance in (20). It can be shown (see Appendix C) that for $\alpha < 1/2$ and η going to infinity, the variances $V_{k\ell}^{1U}$ and $V_{k\ell}^{OPT}$ converge to zero. This is in accord with the fact that the CRB in (21) is equal to zero for these α's, because $\kappa_\alpha = +\infty$ (see Appendix B) [15].

The above result suggests, that if $\alpha \leq 1/2$, the parameter η in (26) should be as large as possible. In practice, however, this approach often fails because of poor convergence of the algorithm to the desired solution for large η.

An accurate estimation of parameter α appears to be difficult, if $\alpha < 2$. Fortunately, it is not so important, as it is also shown in Appendix C, and hence we suggest to apply one universal choice of the nonlinearity, that is

$$g_1(x) = x\exp(-\eta_1|x|) \tag{27}$$

where $\eta_1 = 3.348$. With this choice, the parameter α need not be estimated et all. For easy reference, let us call the nonlinearity in (27) "exp1".

In Figure 4 it is shown that quality of separation achievable by this nonlinearity is close to the corresponding CRB for all α's in the interval $[1, 2)$. The difference in terms of SIR is at most 1 dB, if α is not too close to 2. For $\alpha \in (0, 1]$, the performance is no more close to the CRB, which is infinite for $\alpha \in (0, 1/2]$, but at least it outperforms the performance of the classical" nonlinearity "gauss" with $g(x) = x\exp(-x^2/2)$, which was thought to be best for separation of long-tailed signals. The improvement achieved by "exp1" is about 4 dB when $\alpha = 0.5$ and about 8 dB when $\alpha = 0.2$. (We note, however, that "gauss" slightly outperforms "exp1" for $\alpha > 1.4$.)

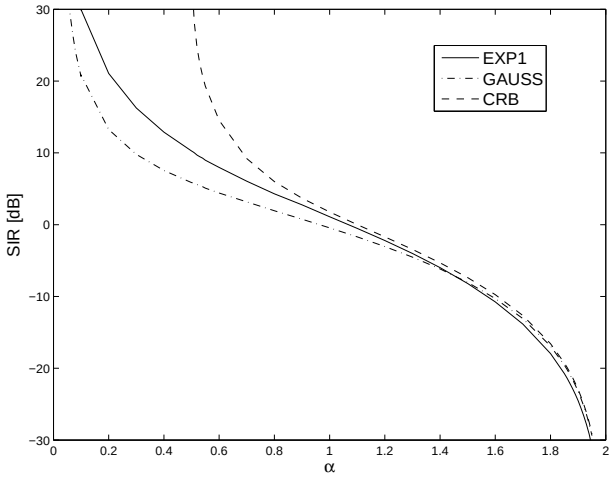


Fig. 2. Theoretically achievable signal-to-interference ratio (SIR) $-10\log_{10} V_{kl}^{OPT}$ for the optimum generalized symmetric FastICA with nonlinearity "exp1" and "gauss" for separation of two sources with the same GG(α) distribution, and the CRB, as a function of α.

[1] It can be easily shown that a nonzero scalar multiplicative factor is irrelevant for usage in the FastICA.

Summary of the Nonlinearity for EFICA
In summary, the nonlinearity of our choice is

$$g_k(x) = \begin{cases} x\exp(-\eta_1|x|) & \text{for} \quad \widehat{m}_{4k} > 3 \\ \text{sign}(x)\cdot|x|^{\min\{\widehat{\alpha}_k - 1, 14\}} & \text{for} \quad 1.8 < \widehat{m}_{4k} \le 3 \\ \text{sign}(x)\cdot|x|^{14} & \text{for} \quad \widehat{m}_{4k} \le 1.8 \end{cases} \tag{28}$$

where $\widehat{m}_{4k}$ is the estimated fourth-order moment of the k-th source signal, given in eq. (24), and $\widehat{\alpha_k}$ is given in (25).

Note that in the vicinity of $\widehat{m}_{4k} = 3$, corresponding to a Gaussian signal, there is a sudden change in nonlinearity, but it does not have any adverse consequences. A signal with nearly Gaussian distribution itself is badly estimable and can be estimated thanks to the other (non-gaussian) signals only. Thus, the effect of selection of the nonlinearity is negligible.

B.3 Step 3: The refinement

The refinement of the initial estimate proceeds in two steps.

The first step, denoted R1, is a more sophisticated implementation of the relation (15). Theoretically, it would suffice to perform (15) once, starting from the initial estimate of $\mathbf{W}$. However, better results are obtained if it is performed separately for each k as series of one unit FastICA iterations, until a convergence is achieved. In the last iteration, however, the normalization step is skipped.

This method works well, if the preliminary estimates of the original signals $\widehat{\mathbf{u}}_k$ from the first step (symmetric FastICA) of the proposed method lie in the right domain of attraction. It might happen, however, that some of the components is difficult to separate from some other component, and the one-unit iterations converge to a wrong component. This patological case can be excluded by checking the condition whether the angle between the component separated by the initial solution and the one unit solution is not too big. If it happens, then the one unit solution should be replaced by the initial estimate.

The step R1 can be summarized as follows.
Step R1: Let $\widehat{\mathbf{W}}^{SYM+} = [\widehat{\mathbf{w}}_1^{SYM+}, \ldots, \widehat{\mathbf{w}}_d^{SYM+}]^T$ and $\widehat{\mathbf{W}}^{SYM} = [\widehat{\mathbf{w}}_1^{SYM}, \ldots, \widehat{\mathbf{w}}_d^{SYM}]^T$, respectively, be the result of (4) and (5) from the last iteration of symmetric FastICA. Assume that Step 2 has been performed, i.e., for each $k = 1, \ldots, d$, $\widehat{m}_{4k}$ has been computed according to (24) with $\widehat{\mathbf{u}}_k = (\widehat{\mathbf{w}}_k^{SYM})^T\mathbf{Z}$, and the optimal nonlinearity g_k has been chosen via (28).

For each $k = 1, \ldots, d$, initialize

$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^{SYM} \tag{29}$$

and under a condition that $|\widehat{\mathbf{w}}_k^T\widehat{\mathbf{w}}_k^{SYM}| \ge 0.95$ iterate the one-unit FastICA

$$\widehat{\mathbf{w}}_k^+ = \mathbf{Z}g_k(\widehat{\mathbf{w}}_k\mathbf{Z}) - \widehat{\mathbf{w}}_k\, g_k'(\widehat{\mathbf{w}}_k\mathbf{Z})\mathbf{1}_N \tag{30}$$
$$\widehat{\mathbf{w}}_k = \widehat{\mathbf{w}}_k^+/\|\widehat{\mathbf{w}}_k^+\| \tag{31}$$

until convergence is achieved. If $|\widehat{\mathbf{w}}_k^T\widehat{\mathbf{w}}_k^{SYM}| < 0.95$, i.e. the new estimate $\widehat{\mathbf{w}}_k$ differs too much from $\widehat{\mathbf{w}}_k^{SYM}$, keep the old (not normalized) result $\widehat{\mathbf{w}}_k^+ = \widehat{\mathbf{w}}_k^{SYM+}$ and put back $g_k(x) = \tanh(x)$.

Now, let $\widehat{\mathbf{W}}^+ = [\widehat{\mathbf{w}}_1^+, \ldots, \widehat{\mathbf{w}}_d^+]^T$ be the result after convergence of the one-unit algorithms. The second refinement step can be summarized as follows.
Step R2: For each $k = 1, \ldots, d$, compute

$$\begin{aligned} \widehat{\mu}_k &= \widehat{\mathbf{u}}_k^T g_k(\widehat{\mathbf{u}}_k)/N, & \widehat{\tau}_k &= |\widehat{\mu}_k - \widehat{\rho}_k| \\ \widehat{\rho}_k &= \widehat{\mathbf{1}}_N^T g_k'(\widehat{\mathbf{u}}_k)/N & \widehat{\gamma}_k &= \widehat{\beta}_k - \widehat{\mu}_k^2 \\ \widehat{\beta}_k &= \widehat{\mathbf{1}}_N^T g_k^2(\widehat{\mathbf{u}}_k)/N \end{aligned} \tag{32}$$

For each $k, \ell = 1, \ldots, d$, compute

$$c_{k\ell} = \begin{cases} \frac{\widehat{\tau}_\ell\widehat{\gamma}_k}{\widehat{\tau}_k(\widehat{\gamma}_\ell + \widehat{\tau}_\ell^2)} & \text{for} \quad \ell \ne k \\ 1 & \text{for} \quad \ell = k \end{cases}$$

Next, for each $k = 1, \ldots, d$, compute

$$\widehat{\mathbf{W}}_k^+ = \text{diag}[c_{k1}, \ldots, c_{kd}]\cdot\widehat{\mathbf{W}}^+ \tag{33}$$
$$\widehat{\mathbf{W}}_k^{aux} = (\widehat{\mathbf{W}}_k^+\widehat{\mathbf{W}}_k^{+T})^{-1/2}\widehat{\mathbf{W}}_k^+ \tag{34}$$
$$\widehat{\mathbf{w}}_k^{REF} = (\widehat{\mathbf{W}}_k^{aux})_{k,\cdot}^T \tag{35}$$

The resulting $k-$th component estimate is $\widehat{\mathbf{u}}_k^{REF} = \mathbf{Z}^T\widehat{\mathbf{w}}_k^{REF}$, and the resulting refined demixing matrix $\widehat{\mathbf{W}}^{REF}$ is

$$\widehat{\mathbf{W}}^{REF} = [\widehat{\mathbf{w}}_1^{REF}, \ldots, \widehat{\mathbf{w}}_d^{REF}]^T \ .$$

This completes the EFICA algorithm.

IV. Simulations

In this section, we provide an experimental comparison of the proposed algorithm with other well-known methods for ICA. For this purpose, we choose the original version of the symmetric FastICA [6] with nonlinearity "tanh", JADE [5], and the nonparametric algorithm NPICA [11], generally believed to give very good results in a variety of separation tasks.

In order to measure the asymptotic errors of the algorithms, we use the achieved signal-to-interference ratio of the k-th signal

$$\text{SIR}_k = \frac{\mathbf{G}_{kk}^2}{\sum_{\ell=1\ell\ne k}^d \mathbf{G}_{k\ell}^2}, \tag{36}$$

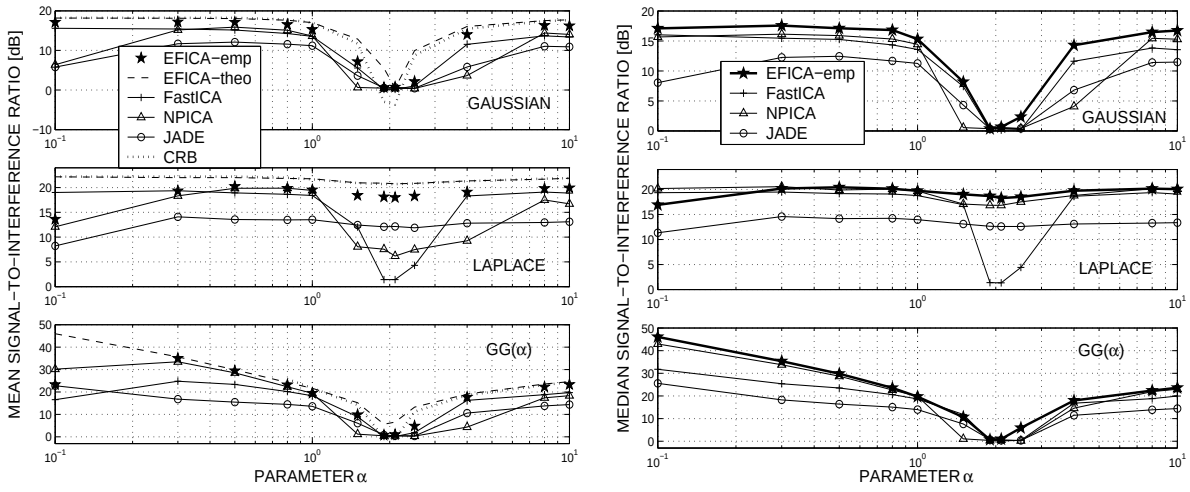
where $\mathbf{G}$ is the acquired gain matrix of the corresponding algorithm. In case of the proposed algorithm, it can be estimated as

$$\overline{\text{SIR}}_k = \frac{N}{\sum_{\ell=1\ell\ne k}^d V_{k\ell}^{OPT}} \tag{37}$$

using (32) and (20).

There occur some random convergence failures in the original FastICA and NPICA, which in our method are fixed by the test of saddle points. To eliminate their effect in the comparisons, both mean and median SIRs from each experiment are shown.

[1]The matlab code for the algorithm EFICA is made available on the Internet at the first author's web page [14].



(a) (b)

Fig. 3. Mean and median SIR of the signal components in Example 1.

A. Experiments with GG(α) distributions

Three simulation experiments with artificial data were carried out to demonstrate the efficiency of the proposed algorithm and its superiority above other methods when all the original signals have generalized Gaussian distributions.

Example 1. One Gaussian signal, 10 Laplacean, and 10 signals with Generalized Gaussian distribution with the same parameter $\alpha \in [0.1, 10]$ and the length $N = 2000$ were generated in 100 independent trials for each α. In each trial, the 21 signals were mixed with a randomly generated matrix, and separated by the proposed method and the algorithms listed above. Averaged results over the 100 trials were computed for each α.

In Figure 3(a), the three diagrams show the dependence of averaged SIRs on α of the Gaussian signal, of the first Laplacean, and of the first signal with Generalized Gaussian distribution. For the proposed EFICA method, both the theoretical SIRs computed via (37) and the empirical ones are presented. Note that the length $N = 2000$ is quite small for separating 21 signals, which also causes overestimation of the attained SIR via (37) since it estimates the asymptotic variance only.

The results can be compared with those in Figure 3(b) where median SIRs from the same experiment are presented. The theoretical SIRs and the CRB are not included there, since they are derived only for mean SIR.

As can be seen from the Figure, the empirical and theoretical SIR (corresponding to the CRB bound) of the EFICA are in good agreement. The empirical SIR is considerably higher than for the comparison methods.

Note that in the vicinity of $\alpha = 2$ there are 11 signals with nearly Gaussian distributions. The CRB on the variance is high (it goes to infinity for $\alpha \to 2$). In this case it may happen that the empirical SIR exceeds the corresponding CRB, as we can see in the upper diagram in Figure 3(a). It happens probably because the estimator is not unbiased here. The MSE of biased estimators can be lower than the CRB.

Example 2. 13 signals of Generalized Gaussian distribution, each with a different value of the 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, were mixed with a random mixing matrix and demixed. The experiment was repeated 100 times with fixed length of data $N = 5000$. The results are plotted in Figure 4. Here, each value of α in the plot correspond to the result for a different original signal. The same conclusions hold as in Example 1.

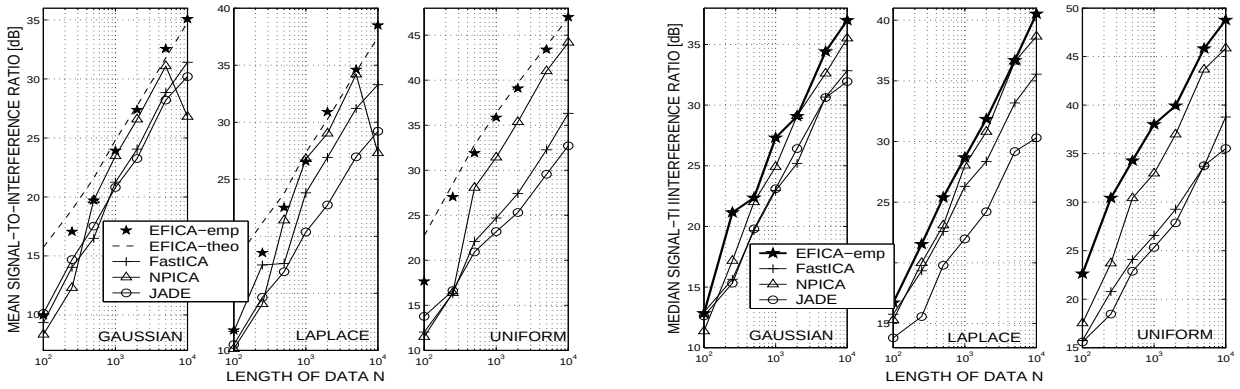
Example 3. To demonstrate the performance of the method for different lengths of data three signals with Gaussian, Laplacean, and uniform distribution were mixed with a random mixing matrix and demixed. The average and median SIRs from 100 independent trials for each length of data are plotted in Figure 5. Again, EFICA outperforms the other methods.

Example 4:. Separation of noisy BPSK signals.
In this example, we consider 10 BPSK signals distorted by Gaussian noise, i.e., i.i.d. data distributed as $\sqrt{1-\varepsilon^2}b + \varepsilon n$, where b is a Bernoulli random variable equal to 1 or -1 with equal probabilities, and n is a standard Gaussian variable. Now the probability density of each signal is (see examples of the densities for different ε in Fig. 6)

$$f_\varepsilon(s) = \frac{1}{2\varepsilon\sqrt{2\pi}} \left(e^{-\frac{\left(s-\sqrt{1-\varepsilon^2}\right)^2}{2\varepsilon^2}} + e^{-\frac{\left(s+\sqrt{1-\varepsilon^2}\right)^2}{2\varepsilon^2}} \right). \quad (38)$$

Data sequences of length $N = 1000$ were generated in 100 independent trials, mixed with a random mixing matrix, and separated. In order to compute the CRB from (14), κ was numerically estimated via a Monte Carlo method. SIR of the first estimated signal as a function of ε is shown in Figure 7.

The results of the experiment need further comments. First, for small ε's the theoretical SIR underestimates the empirical SIR. The reason is that the former SIR is computed from not ideally separated signal components



(a) (b)

Fig. 5. Mean and median SIRs obtained in separating three signal components with varying length.

while its true value approaches infinity. We have used the estimated signal components to mimic the real situation, when the original signals are not available to predict the estimation accuracy. Performance of the symmetric FastICA and of JADE is limited by the orthogonality constraint, which requires that the separated signals must have mutual correlation exactly zero [17]. The NPICA would perform well in separating two or three signals, as it is tailored to separating signals with multimodal distributions. However, in our example with 10 components it fails, even in median, probably because of some implementation or numerical problems.

Example 5: Speech signals separation

To show the performance of the algorithm on real data, although with artificial mixtures, 10 speech signals of length $N = 5000$ were randomly selected from a database of isolated words[2] containing about 200 samples. After centering and normalization, the data were mixed with a random matrix, and consequently separated. Mean and median SIR computed from estimated gain matrices $\mathbf{G} = \widehat{\mathbf{W}}\mathbf{A}$ obtained in 1000 independent trials are summarized in Table I. The proposed EFICA method gives better results than FastICA and JADE. It is outperformed by the algorithm NPICA (in median SIR), though, but note that this is at the expense of much higher computational cost, as shown in the following section.

TABLE I
MEAN AND MEDIAN SIR OF SEPARATED SPEECH SIGNALS

Algorithm	MEAN [dB]	MEDIAN [dB]
EFICA	35.86	27.84
FastICA	27.36	25.03
NPICA	35.80	30.40
JADE	24.21	21.77

B. Complexity of the algorithm

In order to demonstrate the computational complexity, a similar experiment to that in [11] was done. The average CPU time[3] required by the compared methods is shown in Fig. 8(a) for a varying length of data when 6 generalized Gaussian signals with a random parameter $\alpha \in [0.1, 10]$ were separated. The results for a fixed length of data $N = 1000$ and a variable number of signals are in Fig. 8(b).

The complexity of the algorithm is only slightly higher than that of the original symmetric FastICA. The test of saddle points has complexity $\mathcal{O}(d^2N)$; the adaptive choice of nonlinearity (step R1) has complexity $\mathcal{O}(dN)$. Note that the latter two steps have a fixed number of operations, provided that the test of saddle points is negative. Otherwise, only a few additional iterations are needed since the algorithm is initialized almost in the correct solution.

Similarly, only a few additional one-unit iterations in the refinement step R1 are needed.

V. Conclusion

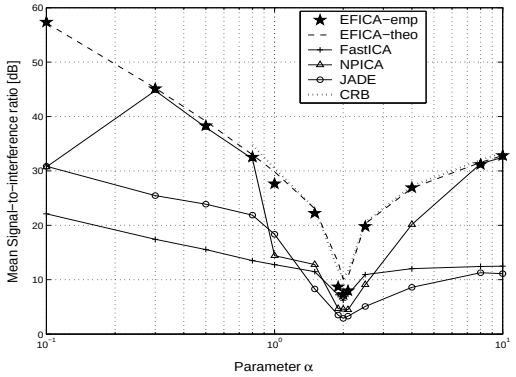
An improved version of the FastICA algorithm was proposed, based on the concept of statistical efficiency. This means that the asymptotic variance of the gain matrix, defined as the product of the estimated unmixing matrix and the original mixing matrix, attains the Cramér-Rao lower bound (CRB) which is the theoretical minimum for the variance. The algorithm was named EFICA.

Two changes have to be made in the standard symmetrical FastICA: first, the nonlinearities must be approximations of the score functions of the true sources. This is achieved by running FastICA in two passes, using the separation results from the first pass to estimate a parametric model of the source densities. Second, a set of extra parameters are added to the algorithm in the form of multiplying the lengths of the weight vectors by some numbers before orthogonalization. By optimizing these numbers, this allows to adjust the asymptotic variance so that it becomes equal to the CRB.

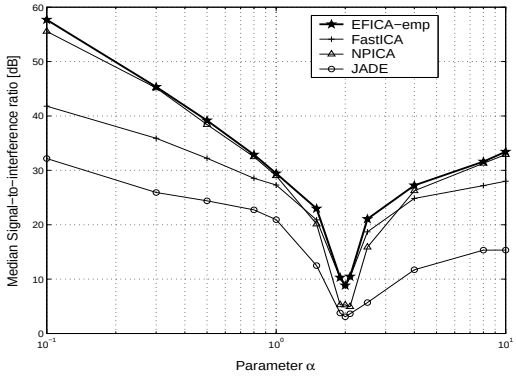
The asymptotic efficiency of EFICA was rigorously proven under the assumption that the source signals

[2] http://noel.feld.cvut.cz/vyu/dzr/cislovky/OBRACENE_BYTY/

[3] The experiment was performed in Matlab^{TM} on a Pentium IV 2.4GHz PC with 512 MB of RAM.



(a) Mean SIR



(b) Median SIR

Fig. 4. Quality of separation of 13 GG(α) components with 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.

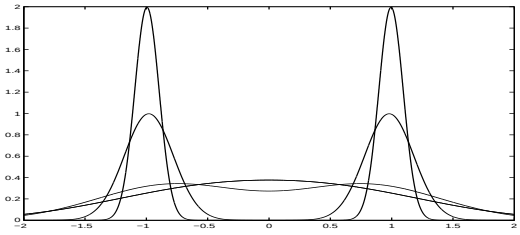


Fig. 6. Probability density (38) for $\varepsilon = 0.1, 0.2, 0.6$ and 0.8, respectively.

follow a Generalized Gaussian density with parameter $\alpha > 0$. In simulations with source signals drawn from this density, the agreement between the theoretical and experimental results was proven. Also, in comparisons to some other ICA algorithms, the EFICA algorithm is superior in this case as predicted by the theory.

It is possible that EFICA will not be optimal (nearly efficient) if the source distributions and their score functions cannot be well modelled by the class of generalized Gaussian distributions - a different than those in Example 4. In that case, the results can be improved by including another method of modelling of the score functions.

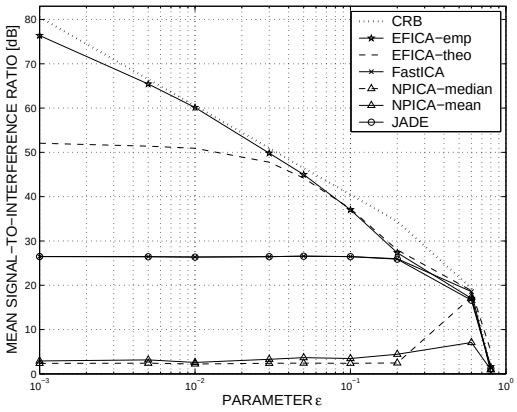


Fig. 7. Separation of noisy BPSK signals.

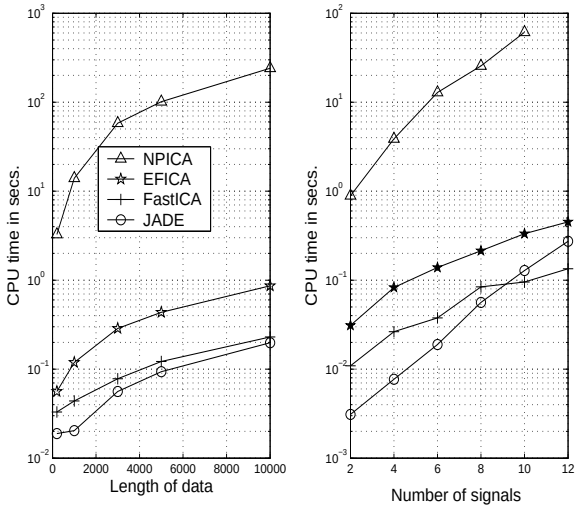


Fig. 8. CPU time needed to separate (a) 6 signals of a various length (b) various number of signals of a fixed length $N = 1000$.

Appendix A

Consider $\alpha \in (2, +\infty)$. The following second order asymptotic expansion of (23) was derived for $\alpha \to +\infty$ in Maple^{TM},

$$z(\alpha) \stackrel{\text{def}}{=} \sqrt{m_4(\alpha) - \frac{9}{5}} = \frac{A}{\alpha} - \frac{B}{\alpha^2} + o(\frac{1}{\alpha^2}), \qquad (39)$$

where $A = \frac{\sqrt{30}\pi}{5}$, $B = \frac{18\sqrt{30}\zeta(3)}{5\pi}$, and $\zeta(\cdot)$ is a Riemann function obeying $\zeta(3) \doteq 1.202$. To invert the above relation, first, substitute $y = \frac{1}{\alpha}$. Then, we can write

$$\begin{aligned} y &= \frac{1}{A}z + \frac{B}{A}y^2 + o(y^2) = \\ &= \frac{1}{A}z + \frac{B}{A}\left(\frac{1}{A}z + \frac{B}{A}y^2 + o(y^2)\right)^2 + o(y^2) = \\ &= \frac{1}{A}z + \frac{B}{A^3}z^2 + o(z^2) \end{aligned} \qquad (40)$$

Using the definition (39) of z gives

$$\alpha \approx \left(\frac{1}{A}\sqrt{m_4(\alpha) - \frac{9}{5}} - \frac{B}{A^3}\left(m_4(\alpha) - \frac{9}{5}\right)\right)^{-1} \qquad (41)$$

See the comparison of inversion of (23) with the derived relation in Fig. 9.

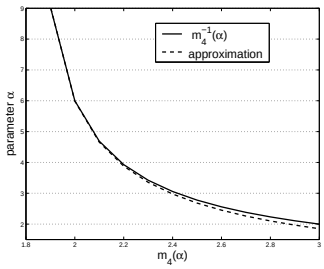


Fig. 9. Comparison of the inversion of (23) with the approximation (41).

Appendix B

Generalized Gaussian distribution family

The generalized Gaussian density function with parameter α, zero mean and variance one, is defined as

$$f_\alpha(x) = c_\alpha \exp\{-(\lambda_\alpha|x|)^\alpha\} \tag{42}$$

where $\alpha > 0$ is a positive parameter that controls the distribution's exponential rate of decay, and $\Gamma(\cdot)$ is the Gamma function, $c_\alpha = \frac{\alpha\lambda_\alpha}{2\Gamma(1/\alpha)}$ and $\lambda_\alpha = \sqrt{\frac{\Gamma(3/\alpha)}{\Gamma(1/\alpha)}}$.

The $k-$th absolute moment for the distribution is

$$\mathrm{E}_\alpha\{|x|^k\} = \int_\infty^\infty |x|^k f_\alpha(x)dx = \frac{1}{\lambda_\alpha^k}\frac{\Gamma\left(\frac{k+1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \tag{43}$$

The score function of the distribution is

$$\psi_\alpha(x) = -\frac{\frac{\partial f_\alpha(x)}{\partial x}}{f_\alpha(x)} = \frac{|x|^{\alpha-1}\mathrm{sign}(x)}{\mathrm{E}_\alpha[|x|^\alpha]} \tag{44}$$

Then, simple computations give

$$\kappa_\alpha = \mathrm{E}_\alpha[\psi_\alpha^2(x)] = \frac{\mathrm{E}_\alpha[|x|^{2\alpha-2}]}{\{\mathrm{E}_\alpha[|x|^\alpha]\}^2} = \frac{\Gamma\left(2-\frac{1}{\alpha}\right)\Gamma\left(\frac{3}{\alpha}\right)}{\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2} \tag{45}$$

Note that $\kappa_\alpha = +\infty$ for $\alpha \le \frac{1}{2}$, $\kappa_\alpha = 1$ for $\alpha = 2$ (the distribution is standard Gaussian), and $\kappa_\alpha \to +\infty$ for $\alpha \to +\infty$.

Appendix C

In this Appendix an asymptotic variance (20) of FastICA with nonlinearity (26) is studied under the assumption that the probability distribution of the independent sources is GG(α) with parameter $\alpha < 2$. Note that (20) can be written in a form

$$V_{k\ell}^{OPT} = \frac{V_{k\ell}^{1U}(V_{\ell k}^{1U}+1)}{V_{k\ell}^{1U}+V_{\ell k}^{1U}+1} \tag{46}$$

where $V_{k\ell}^{1U}$ is given by (9). Since $V_{\ell k}^{1U}$ is independent of the choice of g_k (it is a function of g_ℓ), the nonlinearity g_k which minimizes $V_{k\ell}^{1U}$ for fixed g_ℓ simultaneously minimizes minimizes $V_{k\ell}^{OPT}$.

Let $V_{k\ell}^{1U}(\eta,\alpha)$ be the asymptotic variance computed for parameter η and the parameter of the GG distribution α. We will show that for $\alpha \in (0,\frac{1}{2})$, an optimum η that minimizes (9) is $\eta \to \infty$. In particular,

$$\lim_{\eta\to\infty} V_{k\ell}^{1U}(\eta,\alpha) = 0 \tag{47}$$

for $\alpha \in (0,\frac{1}{2})$.

Proof To prove (47), it suffices to compute asymptotic expansions for quantities in (6)–(8) for $\eta \to \infty$,

$$\begin{aligned}\mu_{\eta,\alpha} &= \mathrm{E}_\alpha[xg_\eta(x)] = \int_{-\infty}^{\infty} xg_\eta(x)f_\alpha(x)\,dx \\ &= 2c_\alpha\int_0^\infty x^2 e^{-\eta x}e^{-(\lambda_\alpha x)^\alpha}dx \\ &= \frac{4c_\alpha}{\eta^3} + O\left(\frac{1}{\eta^{3+\alpha}}\right)\end{aligned} \tag{48}$$

In derivation of (48), the following Taylor series expansion was used,

$$e^{-(\lambda_\alpha x)^\alpha} = 1-(\lambda_\alpha x)^\alpha + \frac{(\lambda_\alpha x)^{2\alpha}}{2!} - \frac{(\lambda_\alpha x)^{3\alpha}}{3!} \pm \ldots \tag{49}$$

together with the definition of the Gamma function $\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}dx$. Similarly, we get

$$\rho_{\eta,\alpha} = \mathrm{E}_\alpha[g'_\eta(x)] = \frac{2\alpha^2 c_\alpha\lambda_\alpha^\alpha\Gamma(\alpha)}{\eta^{1+\alpha}} + O\left(\frac{1}{\eta^{1+2\alpha}}\right) \tag{50}$$

$$\beta_{\eta,\alpha} = \mathrm{E}_\alpha[g_\eta^2(x)] = \frac{c_\alpha}{2\eta^3} + O\left(\frac{1}{\eta^{3+\alpha}}\right) \tag{51}$$

Then, after some algebra, we get

$$\begin{aligned}V_{k\ell}^{1U}(\eta,\alpha) &= \frac{\beta_{\eta,\alpha}-\mu_{\eta,\alpha}^2}{(\mu_{\eta,\alpha}-\rho_{\eta,\alpha})^2} \\ &= \frac{\eta^{2\alpha-1}}{8\alpha^4 c_\alpha\lambda_\alpha^{2\alpha}\Gamma^2(\alpha)} + O(\eta^{\alpha-1})\end{aligned} \tag{52}$$

which proves (47). ■

For any fixed $\frac{1}{2} < \alpha < 2$, $V_{k\ell}^{1U}(\eta,\alpha)$ can be shown to have a global minimum at certain $\eta = \eta_{OPT}(\alpha)$, which was computed numerically and is plotted in Figure 10. We can see that the optimum η grows very fast when α approaches $\frac{1}{2}$.

The variance cannot be minimized simultaneously for all α's. However, the variance does not depend significantly on the value of η, as it can be seen from Figure 11, which shows the ratio $V_{k\ell}^{1U}(\eta,\alpha)/V_{k\ell}^{1U}(\eta_1,\alpha)$ for $\eta = 1, 2, 5$ and 10. As a reference, we have chosen the value η that is optimum for $\alpha = 1$, that is $\eta = \eta_1 = 3.348$.

Acknowledgement

The authors are grateful to the anonymous reviewers for their constructive criticism and valuable comments.

References

[1] P. Comon, "Independent component analysis: a new concept?", *Signal Processing*, 36(3):287-314, Apr. 1994.

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

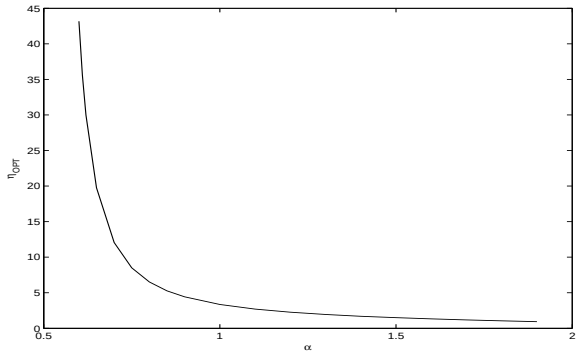


Fig. 10. Optimum parameter η for separation of sources with distribution GG(α) for $\alpha \in (0.5, 2)$.

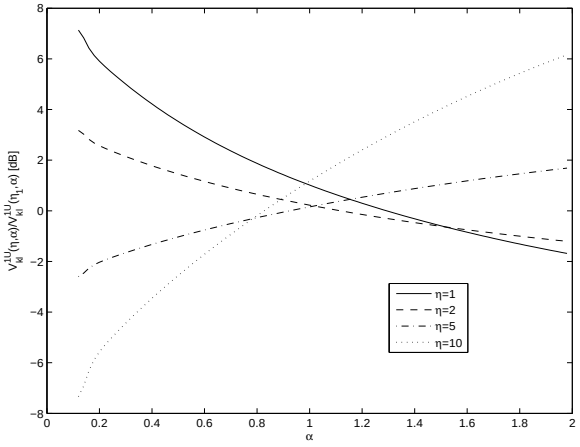


Fig. 11. The relative asymptotic variance $V_{k\ell}^{1U}(\eta, \alpha)/V_{k\ell}^{1U}(\eta_1, \alpha)$ for $\eta = 1, 2, 5, 10$ for shown values of η as a function of parameter α of the GG distribution.

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

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

[5] J.-F. Cardoso and A. Souloumiac, "Blind Beamforming from non-Gaussian Signals", *IEE Proc.-F*, 140(6):362-370, Dec. 1993.

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

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

[8] P. Tichavský, Z. Koldovský, and E. Oja, "Asymptotic Performance Analysis of the Fixed-Point Algorithm (Fast-ICA) for Independent Component Analysis", *IEEE Statistical Signal Processing Workshop*, Bordeaux, July 2005.

[9] 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.

[10] P. Tichavský, Z. Koldovský, and E. Oja, "Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis", accepted for publication in *IEEE Trans. on Signal Processing*, May 2006.

[11] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Independent Component Analysis Based on Nonparametric Density Estimation", *IEEE Trans. on Neural Networks*, vol. 15, no. 1, pp. 55-65, 2004.

[12] E. G. Learned-Miller and J. W. Fisher III, "ICA using spacings estimates of entropy", *Journal of Machine Learning Research*, vol. 4, no. 7-8, pp. 1271-1295, 2004.

[13] `http://www.cis.hut.fi/projects/ica/fastica/`

[14] `http://itakura.kes.tul.cz/zbynek/downloads.htm`

[15] K. Waheed, F. M. Salam, "Blind source recovery using an adaptive generalized Gaussian score function", Circuits and Systems, 2002. MWSCAS-2002. The 2002 45th Midwest Symposium on, 4-7 Aug. 2002 Vol.3, pp. 656-659.

[16] B. Porat, *Digital Signal Processing of Random Signals*, Prentice Hall, New Jersey, 1994.

[17] J.-F. Cardoso, "Blind signal separation: statistical principles", *Proceedings of the IEEE*, vol. 90, n. 8, pp. 2009-2026, October 1998.

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

[19] D. Yellin, B. Friedlander, "Multichannel system identification and deconvolution: performance bounds", *IEEE Trans. on Signal Processing*, Vol. 47, No. 5, May 1999, pp. 1410 - 1414.

[20] O. Shalvi, E. Weinstein, "Maximum likelihood and lower bounds in system identification with non-Gaussian inputs", *IEEE Trans. on Information Theory*, Vol. 40, No. 2, March 1994, pp. 328 - 339.

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

[22] D.T. Pham, "Fast algorithms for mutual information based independent component analysis", *IEEE Trans. on Signal Processing*, Vol. 52, No. 10, October 2004, pp. 2690-2700.

[23] S.-I. Amari, "Superefficiency in blind source separation", *IEEE Trans. on Signal Processing*, vol. 47, no. 4, pp. 936 -944, 1999.

[24] S.-I. Amari and J.- F. Cardoso, "Blind source separation - semiparametric statistical approach", *IEEE Trans. on Signal Processing*, vol. 45, no. 11, pp. 2692-2700, 1997.

[25] P. A. Regalia, E. Kofidis, "Monotonic convergence of fixed-point algorithms for ICA", *IEEE Trans. on Neural Networks*, Vol. 14, No. 4, pp. 943-949, July 2003.

[26] V. Vigneron and C. Jutten, "Fisher information in source separation problems", *Proc. Fifth Intern. Conf. on ICA and BSS*, Granada, Spain, September 2004, pp. 168-176.

[27] Y. Lomnitz and A. Yeredor, "A blind-ML scheme for blind source separation", *Proc. of the 2003 IEEE Workshop on Statistical Signal Processing (SSP2003)*, St. Louis, Sep. 2003.

[28] Z. Koldovský, P. Tichavský "Methods of Fair Comparison of Performance of Linear ICA Techniques in Presence of Additive Noise", to be presented at *ICASSP-2006*.