Signal Processing 89 (2009) 2570–2584

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Blind separation of piecewise stationary non-Gaussian sources

Zbyněk Koldovský [a,b,*], Jiří Málek [a], Petr Tichavský [b], Yannick Deville [c], Shahram Hosseini [c]

[a] Faculty of Mechatronics, Informatics and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic
[b] Institute of Information Theory and Automation, Pod vodárenskou věží 4, P.O. Box 18, 182 08 Praha 8, Czech Republic
[c] Laboratoire d'Astrophysique de Toulouse-Tarbes, Université de Toulouse, CNRS, 14 Av. Edouard Belin, 31400 Toulouse, France

ARTICLE INFO

Article history:
Received 15 September 2008
Received in revised form
25 March 2009
Accepted 16 April 2009
Available online 24 April 2009

Keywords:
Independent component analysis
Blind source separation
Cramér–Rao lower bound
FastICA algorithm
Piecewise stationary model

ABSTRACT

We address independent component analysis (ICA) of piecewise stationary and non-Gaussian signals and propose a novel ICA algorithm called Block EFICA that is based on this generalized model of signals. The method is a further extension of the popular non-Gaussianity-based FastICA algorithm and of its recently optimized variant called EFICA. In contrast to these methods, Block EFICA is developed to effectively exploit varying distribution of signals, thus, also their varying variance in time (nonstationarity) or, more precisely, in time-intervals (piecewise stationarity). In theory, the accuracy of the method asymptotically approaches Cramér–Rao lower bound (CRLB) under common assumptions when variance of the signals is constant. On the other hand, the performance is practically close to the CRLB even when variance of the signals is changing. This is demonstrated by comparing our algorithm with various methods that are asymptotically efficient within ICA models based either on the non-Gaussianity or the nonstationarity. The benefit of our algorithm is demonstrated by examples with real-world audio signals.

1. Introduction

The instantaneous linear mixture model is the basic configuration considered in blind source separation (BSS) [1]. The relation between unobserved original signals and observed measured signals is here given by equation

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

where $\mathbf{X}$ and $\mathbf{S}$ are matrices with N columns, each of which represents samples of the measured and the original signals, respectively. We will consider the regular case, thus, the number of rows of $\mathbf{X}$ and $\mathbf{S}$ is the same and equal to d, and the *mixing* matrix $\mathbf{A}$ is a $d \times d$ regular matrix.

Estimating the mixing matrix $\mathbf{A}$ or equivalently the original signals $\mathbf{S}$ from the data $\mathbf{X}$ is the general task of BSS. To solve this problem, a principle giving some assumption about the original signals should be introduced. The most popular one is based on the assumption of their statistical independence, which is used by a certain class of models that fall within a popular BSS discipline called independent component analysis (ICA) [2].

Since Comon's pioneering paper [3], numerous successful algorithms have been proposed using basic models based either on non-Gaussianity [4–6], nonstationarity [7–9] or spectral diversity (coloration) [10,11] of the original signals. Later, various improvements of the earlier methods were developed [12]. The most recent algorithms provide fast and reliable solutions while attaining the best possible accuracy fundamentally limited by the respective Cramér–Rao lower bound (CRLB) [13,15–17].

* Corresponding author at: Faculty of Mechatronics, Informatics and Interdisciplinary Studies, Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic. Tel.: +420 485 353534; fax: +420 485 353112.
E-mail addresses: zbynek.koldovsky@tul.cz (Z. Koldovský), jiri.malek@tul.cz (J. Málek), tichavsk@utia.cas.cz (P. Tichavský), yannick.deville@ast.obs-mip.fr (Y. Deville), shahram.hosseini@ast.obs-mip.fr (S. Hosseini).

0165-1684/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2009.04.021

While methods assuming non-Gaussianity of signals require computation of higher-order statistics (HOS), the methods using nonstationarity or spectral diversity usually need second-order statistics (SOS) only, which provides faster implementations usually through joint approximate diagonalization of a set of matrices; see e.g., [14,17–19]. On the other hand, each approach cannot separate sources if the respective assumptions are not met, which means certain limitations. For instance, the nonstationarity-based methods cannot separate signals having the same dynamics. In this respect, the non-Gaussianity-based methods are popular thanks to their widest application area, e.g., in telecommunications, biomedical signal processing, or speech and audio processing.

Since real signals may often exhibit both non-Gaussianity, nonstationarity or temporal structure, there are attempts to derive methods that combine two or more models to enhance the application area and to improve the performance [20,21]. However, the theoretical background of combined models is much more complex. Most methods therefore rely on various heuristically chosen criteria [22–24] or decision-driven combinations of basic algorithms [20,25], rather than optimizing the performance in a straightforward way through the theory.

This paper focuses on the model combining the non-Gaussianity and the nonstationarity assumptions through the so-called *piecewise stationary* model. The optimum solution of this model was discussed in [26], but few methods were proposed for finding it. The fully general framework was considered by Pham [27]. He proposed an algorithm, from here on named as NSNG, that performs (quasi)-maximum likelihood estimation (MLE), which yields excellent performance in theory. However, in our experimental tests [28], we have observed cases of instability and misconvergence of NSNG. Specifically, the algorithm seems to work well in simple scenarios, e.g., where "few" signals are separated and their properties perfectly fit the model. By contrast, the method failed with nonnegligible probability in more difficult examples or when separating real-world signals such as EEG data or real audio mixtures.

To provide a reliable algorithm with lower computational burden and comparable performance with that of NSNG, we here introduce a further extension of the very popular FastICA algorithm [5], which was originally developed for non-Gaussian signals. The method is called *Block EFICA*[1] as it is an extension of the EFICA algorithm [16] (a theoretically optimized FastICA variant for non-Gaussian signals) for piecewise stationary signals.

The paper is organized as follows. The following section introduces the piecewise stationary model and basic notations used throughout the paper. Section 3 surveys Cramér–Rao bounds that were derived for several levels of generalizations of the basic non-Gaussianity-based ICA model. The proposal of the Block EFICA algorithm is given in Section 4 after short descriptions of its forgoers: FastICA [5] and EFICA [16]. Section 5 provides performance analysis of behavior of FastICA under the assumption of piecewise stationary signals, and introduces optimized selection of important parameters of Block EFICA to achieve the best performance. Finally, experimental results demonstrating performance of the Block EFICA in comparison with other methods are presented by Section 6.

2. Piecewise stationary model

The basic ICA model exploiting non-Gaussianity of the sources is defined by

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{2}$$

where $\mathbf{s} = [s_1, \ldots, s_d]^T$ is a vector of independent random variables (RVs),[2] and each of them represents one of the unknown original signals. In practice, this means that the data matrices $\mathbf{X}$ and $\mathbf{S}$ consist of N i.i.d. realizations of $\mathbf{x}$ and $\mathbf{s}$, respectively, whose relation is given through the transform $\mathbf{A}$. The very assumption of independence of $s_1, \ldots, s_d$ is used for finding the de-mixing transform $\mathbf{A}^{-1}$, which can be achieved only up to an indeterminable order, scales, and signs of its rows.

Compared to the basic ICA model (2), the piecewise stationary model consists in that the samples of the original signals need *not* be identically distributed. The probability density function (pdf) $f_k(x)$ of s_k thus may be different at each time instant/interval.

However, to allow practical estimation of signal statistics on data blocks, we will assume that there are M blocks of $\mathbf{S}$ of the same integer length N/M, where, within each block, the distribution of the signals is unchanging. Therefore, we will use the superscript (ℓ) to denote quantities, RVs or functions that are related to the ℓ-th block. For instance, this means that for each block of data $\mathbf{X}$, say for the ℓ-th block $\mathbf{X}^{(\ell)}$, it holds $\mathbf{X}^{(\ell)} = \mathbf{A}\mathbf{S}^{(\ell)}$, which corresponds to N/M i.i.d. realizations according to model

$$\mathbf{x}^{(\ell)} = \mathbf{A}\mathbf{s}^{(\ell)}, \tag{3}$$

where $\mathbf{x}^{(\ell)}$ and $\mathbf{s}^{(\ell)}$ are vectors of corresponding RVs.

A particular case of the piecewise stationary model, which will be called *Block Gaussian* model [14], is when all the distributions of all RVs in (3) are Gaussian. This means that all signals are white and Gaussian within each block, and the piecewise stationarity consists only in that their variances vary block-by-block.

3. Cramér–Rao lower bounds for independent component analysis

We discuss several bounds that limit the accuracy achievable by blind separation. Such limitation may be given by the Cramér–Rao lower bound that is related to the theoretical model of the original signals. In other words, the bound is different for different models requiring various assumptions about the original signals,

[1] The primary version of Block EFICA introduced in [28] was referred to as *Extended EFICA*.

[2] For simplicity, all RVs considered in the paper are assumed to have zero mean and finite variance.

whereby the separation principles are determined. Although the bounds presented here may all be derived as particular cases of the bound given for the most general piecewise stationary model, for convenience, we start the description with the basic ICA bound and then generalize it gradually.

In general, CRLB is defined for an unbiased estimator of some (multivariate) parameter θ, which is being estimated from a data vector x that has the probability density $f_{x|\theta}(x|\theta)$. CRLB is the lower bound for the covariance matrix of any unbiased estimator $\hat{\theta}$ of θ, i.e.,

$$\mathrm{cov}_\theta\hat{\theta} = \mathrm{E}_\theta[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^T]. \tag{4}$$

If the following Fisher information matrix (FIM) and its inverse exist

$$\mathbf{F} = \mathrm{E}_\theta\left[\frac{1}{f^2_{x|\theta}(x|\theta)}\frac{\partial f_{x|\theta}(x|\theta)}{\partial\theta}\left(\frac{\partial f_{x|\theta}(x|\theta)}{\partial\theta}\right)^T\right], \tag{5}$$

under the regularity conditions [29] it holds that

$$\mathrm{cov}_\theta\,\hat{\theta} \geqslant \mathbf{F}^{-1} = \mathrm{CRLB}[\theta],$$

where the matrix inequality means that the matrix $\mathrm{cov}_\theta\,\hat{\theta} - \mathbf{F}^{-1}$ is positive semidefinite.

In case of the instantaneous linear mixture $\mathbf{X} = \mathbf{AS}$, the parameters intended for the estimation are the elements of $\mathbf{A}^{-1}$. Let $\mathbf{W}$ be an unbiased estimator of $\mathbf{A}^{-1}$. Instead of considering CRLB of $\mathbf{W}$, it is useful to derive the bound for the so-called *gain* matrix $\mathbf{G} = \mathbf{WA}$. Without loss of generality, the indeterminacies of order, signs, and scales of the original signals can be assumed to be resolved. $\mathbf{G}$ should then be close to the identity, and the variances of its nondiagonal elements, $\mathrm{var}[\mathbf{G}_{k\ell}]$, $k \neq \ell$, reflect mean value of residual interference between the separated signals $\mathbf{WX}$. Such a criterion, which is commonly used in signal processing, reflects well the accuracy of the estimator $\mathbf{W}$.

The CRLB for the basic ICA model (2) has been well known since the 1990s [30,31]. We will denote the bound by CRLB_1, and it is given by

$$\mathrm{CRLB}_1[\mathbf{G}_{k\ell}] = \frac{1}{N}\frac{\kappa_\ell}{\kappa_k\kappa_\ell - 1}, \quad k \neq \ell, \tag{6}$$

where $\kappa_k = \mathrm{E}[\psi_k^2(x)]$ with $\psi_k(x) = -f'_k(x)/f_k(x)$ being the score function of the probability density function $f_k(x)$ of the k-th RV s_k. Here s_k is assumed to have unit variance, thus, note that κ's are defined for unit-variance score functions.

The bound for the piecewise stationary model (3) with constant (unit) variance signals, denoted by CRLB_2, is given by [28]

$$\mathrm{CRLB}_2[\mathbf{G}_{k\ell}] = \frac{1}{N}\frac{\overline{\kappa}_\ell}{\overline{\kappa}_k\overline{\kappa}_\ell - 1}, \quad k \neq \ell, \tag{7}$$

where $\overline{\kappa}_k \stackrel{\mathrm{def}}{=} (1/M)\sum_{l=1}^M \kappa_k^{(l)}$.

Now we introduce the most general bound, i.e., for the piecewise stationary model (3) where the variance of the signals is not assumed to be constant. Let $\sigma_k^{2(l)}$ be the variance of $s_k^{(l)}$, $k = 1,\ldots,d$, $l = 1,\ldots,M$, but $\kappa_k^{(l)}$ is still defined for pdf $f_k^{(l)}(\cdot)$ normalized so as to correspond to RV *normalized to unit variance*. Then, the bound could be written in the form

$$\mathrm{CRLB}_3[\mathbf{G}_{k\ell}] = \frac{1}{N}\frac{B_{k\ell}}{A_{k\ell}B_{k\ell} - 1}, \quad k \neq \ell, \tag{8}$$

where

$$A_{k\ell} = \frac{1}{M}\sum_{l=1}^M \frac{\sigma_\ell^{2(l)}}{\sigma_k^{2(l)}}\kappa_k^{(l)}, \tag{9}$$

$$B_{k\ell} = \frac{1}{M}\sum_{l=1}^M \frac{\sigma_k^{2(l)}}{\sigma_\ell^{2(l)}}\kappa_\ell^{(l)}. \tag{10}$$

This result was previously derived, e.g., in [26]. In Appendix A, we provide a simple derivation of the bound using the derivation of FIM from [32].

For the sake of completeness, we introduce the CRLB for the Block Gaussian model, i.e., when all distributions of signals are Gaussian. The bound easily follows from (8) by taking $\kappa_k^{(l)} = 1$ in (9) and (10). We will denote this bound by CRLB_4.

4. Block EFICA algorithm

We here describe our novel algorithm that is an extension of its previous variants FastICA and EFICA. First, the underlying methods are reminded in short as they were proposed for solving the model (2). Second, the building block of the proposed algorithm is given, which is a straightforward extension of the one-unit FastICA algorithm to the piecewise stationary signals. Finally, we introduce the proposed algorithm.

4.1. FastICA and EFICA algorithms

The FastICA algorithm [5] was originally derived as a method for solving the basic ICA problem (2). It is based on optimization of a contrast function

$$c(\mathbf{w}_k) = \mathrm{E}[G(\mathbf{w}_k^T\mathbf{z})], \tag{11}$$

subject to the vector $\mathbf{w}_k^T$, whose optimum value is the k-th row of de-mixing transform. The function $G(\cdot)$, which applies elementwise, is a properly chosen nonlinear function whose derivative will be denoted by $g(\cdot)$. The vector $\mathbf{z}$ is derived by transforming signals $\mathbf{x}$ so that the components of $\mathbf{z}$ are not correlated and have unit variance. After this preprocessing, which is commonly referred to as sphering, it holds that $\mathrm{E}[\mathbf{zz}^T] = \mathbf{I}$.

The optimization of $c(\mathbf{w}_k)$ is based on the iteration

$$\mathbf{w}_k^+ \leftarrow \mathrm{E}[\mathbf{z}g(\mathbf{w}_k^T\mathbf{z})] - \mathbf{w}_k\mathrm{E}[g'(\mathbf{w}_k^T\mathbf{z})]. \tag{12}$$

In practice, i.e., when working with a finite number of signal samples, the theoretical expectations are replaced by respective sample means, thus the resulting de-mixing vectors/matrices are respective estimates thereof.

The original FastICA was developed in two basic versions: the one-unit and the symmetric one. While the one-unit FastICA completes each iteration by normalizing the vector $\mathbf{w}_k^+$, the symmetric FastICA computes d iterations (12) in parallel and does a symmetric orthogonalization[3] of $[\mathbf{w}_1^+,\ldots,\mathbf{w}_d^+]^T$ to yield all rows of the de-mixing matrix, whose practical estimate will be denoted by $\widehat{\mathbf{W}}$.

The theoretical (asymptotic) performance [32] of the one-unit FastICA is characterized by

$$\mathrm{var}[\mathbf{G}^{1U}_{k\ell}] \approx \frac{1}{N} V^{1U}_{k\ell}, \quad k \neq \ell, \tag{13}$$

where $\mathbf{G}^{1U} = \widehat{\mathbf{W}}\mathbf{A}$ is the gain matrix, each of its rows corresponds to the estimation of one de-mixing vector, and $V^{1U}_{k\ell} = \gamma_k/\tau_k^2$ with

$$\begin{aligned} &\mu_k = \mathrm{E}[s_k g(s_k)], \quad \gamma_k = \beta_k - \mu_k^2,\\ &\nu_k = \mathrm{E}[g'(s_k)], \quad \tau_k = \nu_k - \mu_k,\\ &\beta_k = \mathrm{E}[g^2(s_k)]. \end{aligned} \tag{14}$$

In case that the expectations do not exist it may signify either bad choice of the function $G(\cdot)$ or zero leading term in the asymptotic expansion of the variance (13). It is a well-known feature in ICA that the optimum choice of $G(\cdot)$ comes up to $g(\cdot)$ being the score function of s_k, $\psi_k(\cdot)$ [13,31].

Among other things, this knowledge is taken into account by the recently published EFICA algorithm [16], which is designed to attain the best possible performance limited by (6). The method proceeds in three steps: (1) it preestimates all the original signals by means of the symmetric FastICA with the test of saddle points [32], (2) for each $k = 1,\ldots,d$, it adaptively chooses a nonlinearity $g \stackrel{\text{def.}}{=} g_k$ that approximates the score function $\psi_k(\cdot)$, and (3) it does fine-tunings and a refinement.

The fine-tunings consist in further one-unit FastICA iterations for each signal separately, using the nonlinearities found in step 2. The resulting de-mixing vectors from the fine-tunings $\mathbf{w}_1^+,\ldots,\mathbf{w}_d^+$ are then optimally combined by the refinement.

The refinement utilizes optimum weights computed according to

$$c_{k\ell} = \begin{cases} \dfrac{V^{1U}_{k\ell}}{V^{1U}_{\ell k} + 1}, & k \neq \ell, \\ 1, & k = \ell. \end{cases} \tag{15}$$

We remark that we use slightly different definition of the weights from that in [16] since it is handier for forthcoming description of the Block EFICA. The modification simply consists in normalizing the vectors $\mathbf{w}_1^+,\ldots,\mathbf{w}_d^+$, which was not done in [16]; see [33] for details. The weights are used to form matrix

$$\mathbf{W}_k^+ = [c_{k1}\mathbf{w}_1^+/\|\mathbf{w}_1^+\|,\ldots,c_{kd}\mathbf{w}_d^+/\|\mathbf{w}_d^+\|]^T. \tag{16}$$

The k-th row of symmetrically orthogonalized version of $\mathbf{W}_k^+$, i.e., of $(\mathbf{W}_k^+\mathbf{W}_k^{+T})^{-1/2}\mathbf{W}_k^+$, yields the final estimate of $\mathbf{w}_k$. This is done for each $k = 1,\ldots,d$ separately, which relaxes the orthogonality constraint [31] introduced by the symmetric FastICA.

[3] We use the well-established term "symmetric orthogonalization" although "symmetric orthonormalization" would be more accurate.

The asymptotic performance of EFICA is given by

$$\mathrm{var}[\mathbf{G}^{EF}_{k\ell}] \approx \frac{1}{N}\frac{V^{1U}_{k\ell}(V^{1U}_{\ell k}+1)}{V^{1U}_{k\ell}+V^{1U}_{\ell k}+1}, \quad k \neq \ell. \tag{17}$$

The particular case when $g_k = \psi_k$ reveals superior property of EFICA. It holds, then, that $\beta_k = \nu_k = \kappa_k$, $\mu_k = 1$, and $V^{1U}_{k\ell} = 1/(\kappa_k - 1)$. Substituting this into (17) gives

$$\mathrm{var}[\mathbf{G}^{EF}_{k\ell}] \approx \frac{1}{N}\frac{\kappa_\ell}{\kappa_k\kappa_\ell - 1}, \quad k \neq \ell.$$

Compared to the CRLB_1 given by (6), the asymptotic efficiency of EFICA in the framework of the basic ICA model (2) follows.

4.2. One-unit FastICA for piecewise stationary signals

To take into account the piecewise stationary model, we introduce a new definition of the contrast function (11), which is

$$c(\mathbf{w}_k) = \lambda_k^{(1)}\mathrm{E}[G_k^{(1)}(\mathbf{w}_k^T\mathbf{z}^{(1)})] + \cdots\cdots + \lambda_k^{(M)}\mathrm{E}[G_k^{(M)}(\mathbf{w}_k^T\mathbf{z}^{(M)})], \tag{18}$$

where $G_k^{(1)},\ldots,G_k^{(M)}$ are properly chosen nonlinear functions, and $\lambda_k^{(1)},\ldots,\lambda_k^{(M)}$ denote some weights.

It should be noted that this contrast *cannot* be viewed as the contrast (11) with $G(\cdot)$ being a linear combination of $G_k^{(1)},\ldots,G_k^{(M)}$, because each expectation in (18) depends on different distributions from corresponding block of the signals. Also, an important fact is that each term in (18) is a valid contrast function itself. Since the mixing matrix is the same in all blocks, (18) is a valid contrast function as well. In other words, all the contrasts represented by the terms in (18) have the same optimum points.

One-unit FastICA using the contrast function (18), from here on referred to as *block one-unit FastICA*, works in the way that it applies a different nonlinearity $g(\cdot)$ on each block of signals. Thus, the iteration (12) changes to

$$\begin{aligned} \mathbf{w}_k^+ \leftarrow\; & \lambda_k^{(1)}(\mathrm{E}[\mathbf{z}^{(1)}g_k^{(1)}(\mathbf{w}_k^T\mathbf{z}^{(1)})] - \mathbf{w}_k\mathrm{E}[g_k^{(1)\prime}(\mathbf{w}_k^T\mathbf{z}^{(1)})]) \\ & + \cdots + \lambda_k^{(M)}(\mathrm{E}[\mathbf{z}^{(M)}g_k^{(M)}(\mathbf{w}_k^T\mathbf{z}^{(M)})] \\ & - \mathbf{w}_k\mathrm{E}[g_k^{(M)\prime}(\mathbf{w}_k^T\mathbf{z}^{(M)})]), \end{aligned} \tag{19}$$

and the expectations are replaced by sample means in practice.

As can be seen, the original one-unit FastICA is, when setting $\lambda_k^{(l)} = 1/M$ and $g_k^{(l)} = g$, for all $l = 1,\ldots,M$, a particular case of the block version introduced here. Theoretical conclusions derived later in this paper, therefore, yield an insight into the behavior of the original FastICA (and also of other variants of FastICA) when distributions of signals are different from one block to the other.

4.3. Proposed Block EFICA algorithm

The Block EFICA algorithm takes into account the piecewise stationarity of signals. The approach consists of the following three steps that are similar to those in the original EFICA up to the difference that consists in linking the choice of nonlinearities with the fine-tuning into a common step due to higher precision. Also a different approach for the choice of nonlinear functions is used, because variance of signals in blocks need not be equal to one as assumed by the approach used in EFICA.

BEF1 Separation by the symmetric FastICA with the test of saddle points in order to obtain a preestimate of the de-mixing matrix $\widehat{\mathbf{W}}$.

BEF2 Fine-tuning of each row of $\widehat{\mathbf{W}}$ by means of the block one-unit FastICA (Section 4.2). The weights and the nonlinearities in (19) are simultaneously updated as described below. The simplified version of the algorithm, called *Uniform Block EFICA*, selects all the weights equal to an arbitrary nonzero value.

BEF3 The refinement to get the most accurate and final estimate of the whole de-mixing matrix.

A simplified illustration of the flow of Block EFICA is shown in Fig. 1. In the following, we provide more details on the steps of the algorithm.

The pre-estimation of the whole de-mixing matrix in BEF1 could be done by any ICA method, which opens up possible variations of the Block EFICA. Nevertheless, our selection, the symmetric FastICA with the test of saddle points, proves being suitable for wide variety of scenarios [16]. The method allows fast and reliable separation of non-Gaussian signals. Moreover, in practice it generally allows significant separation of piecewise stationary signals as well, which follows from the fact that the symmetric FastICA is a special variant of the block one-unit FastICA introduced in the previous subsection. However, it has limited accuracy due to the nonoptimal choice of the nonlinearity that is fixed for all signals and blocks, and also due to the orthogonality constraint [31] introduced by the symmetric orthogonalization. Therefore, it is a suitable initialization for the fine-tuning done by step BEF2.

In the fine-tunings (BEF2), the estimation of the k-th signal, $k = 1, \ldots, d$, is improved by starting the block one-unit FastICA using appropriately chosen functions $g_k^{(l)}(\cdot)$ and the weights $\lambda_k^{(l)}$, $l = 1, \ldots, M$. Since the best choice of $g_k^{(l)}(\cdot)$ is the score function $\psi_k^{(l)}(\cdot)$, we use the approximation by Pham's estimator from [13]. The details are given below in the extra subsection.

The choice of the weights $\lambda_k^{(l)}$ has an influence on the performance of fine-tunings as well. Since it should be analyzed first, the choice is given afterwards in Section 5 by the expression (27). In that section, we also justify the introduction of the Uniform Block EFICA algorithm, which sets all the weights to a constant.

Finally, the refinement step is done in the similar way as in the original EFICA [16]. The fine-tuned and normalized rows of the separating matrix resulting from BEF2, $\mathbf{w}_1, \ldots, \mathbf{w}_d$, and the weights $c_{k\ell}$ are used to form matrix

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

Then, the k-th row of the matrix $(\mathbf{W}_k^+ \mathbf{W}_k^{+T})^{-1/2}\mathbf{W}_k^+$ yields the final estimate of $\mathbf{w}_k$. The difference compared to EFICA is that the weights $c_{k\ell}$ should be computed accordingly. Namely, (15) is in fact a function of the performance achieved by the fine-tuning in EFICA, i.e., by that of the one-unit FastICA given by (13). However, the fine-tuning in the Block EFICA is done by means of the block one-unit FastICA algorithm, whose performance is different. The performance is analyzed in Section 5, where the analytical expression (28) for the weights follows.

4.4. Parametric estimation of score functions

Parametric estimation of score functions is a well-established problem in statistical theory [34]. The parametric estimator proposed in [13] is suited for the problems tackled by ICA algorithms. It is defined as the minimizer of the mean square distance between a score function $\psi(\cdot)$ and a linear combination of K *basis* functions $h_1(x), \ldots, h_K(x)$, i.e.,

$$\min_{\theta_1,\ldots,\theta_K} \mathrm{E}\left[\left(\psi(x) - \sum_{i=1}^{K} \theta_i h_i(x)\right)^2\right]. \tag{20}$$

The merit consists in the fact that $\mathrm{E}[\psi(x)h(x)] = \mathrm{E}[h'(x)]$ for any function $h(x)$. Thanks to this, the minimization is possible without knowledge of $\psi(\cdot)$ and is fast, because it only requires estimation of $\mathrm{E}[h_i(x)h_j(x)]$ and $\mathrm{E}[h_i'(x)]$,

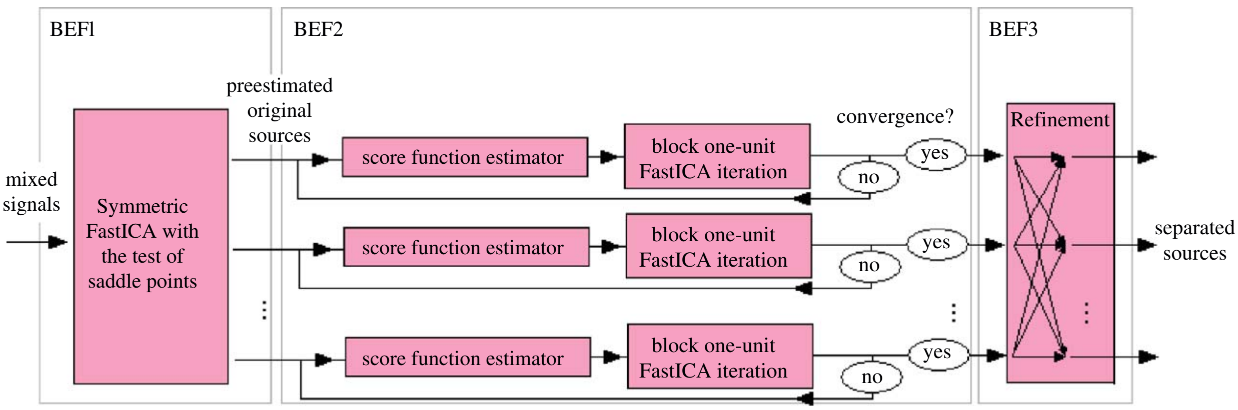

Fig. 1. Flow of the Block EFICA algorithm.

$i,j = 1,\dots,K$. The minimizer is then given by the solution of a set of K linear equations.

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 [37]. This choice turns out to be appropriate for a wide class of distributions and offers a good trade-off between accuracy, speed, and flexibility. For instance, when considering signals with generalized Gaussian distributions, the estimator (20) with our settings used within EFICA yields comparable results with the adaptation originally used thereby [35].

Another advantage of this estimator consists in computational savings: once the moments $\mathrm{E}[h_i(x)h_j(x)]$ and $\mathrm{E}[h_i'(x)]$ are estimated, the results can be used where corresponding moments occur, which is, e.g., in the iteration (19). The burden due to the solution of minimizing equations is, for $K = 2$, negligible, thus, the main slowing-down compared to the adaptation used in EFICA consists in that two nonlinear functions h_1 and h_2 must be evaluated.

Here, we should point out that it is relevant to take into account the identity function $h_i(x) = x$ for the third basis function in (20). Unlike in case of the original FastICA/EFICA, this is meaningful in Block EFICA, because each block $\mathbf{z}^{(l)}$ of the sphered data $\mathbf{z}$ may *not* be sphered. Specifically, when considering $g(x) = \alpha x + h(x)$ in (12) with an arbitrary α and a nonlinearity $h(x)$, the effect of the term αx is zeroed no matter how α is chosen since $\mathrm{E}[\mathbf{z}\mathbf{z}^T] = \mathbf{I}$. It is not so in case of the "block-iteration" (19) due to nonsphered blocks $\mathbf{z}^{(l)}$.

Inclusion of the identity function into the score function estimator, in fact, conveys direct utilization of second-order statistics of signals. The consideration is worthwhile especially when separating signals with changing variance. Therefore, we consider this as an option in the Block EFICA, which is slightly more computationally expensive.

4.5. Choice of the number of blocks

The correct number of blocks M is usually not known in practice. The goal is to choose M such that the distribution of $\mathbf{S}$ may be regarded as constant within each block. On the other hand, M should not be overestimated, because overparametrization may cause higher estimation error. Luckily, Block EFICA is not highly sensitive to this parameter, which is demonstrated by results shown in Fig. 4 in Section 6. It is shown that significant overestimations of M as well as its underestimations do not decrease the performance seriously.

Usually, the choice of optimum M is done by taking into account characteristics of signals to be separated. For example, when separating speech signals, it is worth to select M such that the length of blocks corresponds to 20–25 ms where speech is almost stationary.

Blind selection of M may be based on estimation of residual inter-signal interference (signal-to-interference ratio—SIR) using analytical expressions (29) where corresponding statistics are estimated from separated signals. It is thus possible to see the estimated SIR of separated signals as a function of M. At the beginning, SIR usually improves with growing M, but for larger M the growth is slower and slower. We would select M where the increase of SIR becomes slow; see Fig. 4. A similar approach but more computationally demanding would be when Block EFICA was started with different M's taken from a reasonable range, and the optimum M or its effective value was selected subject to the resulting estimate of SIR. Another possible approach for automated choice of M can be found, e.g., in [36].

5. Performance analysis

In this section, we analyze performance of the proposed Block EFICA algorithm to reveal influence of its parameters on accuracy of separation. Optimization of the theoretical performance subject to the parameters gives their final definition, which also completes the description of the algorithm.

The starting point of the analysis is the derivation of the performance of the block one-unit FastICA, which is achieved by the fine-tunings in BEF2. We generalize the analysis of the original one-unit FastICA from [32] that considers the basic ICA model (2) to the piecewise stationary model with M blocks. The analysis yields the result summarized in the following proposition.

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

(i) *the RVs $s_k^{(l)}$ have zero mean and finite variance $\sigma_k^{2(l)}$ such that it holds that (unit scale)*

$$\frac{1}{M}\sum_{l=1}^{M}\sigma_k^{2(l)} = 1,$$

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

(iii) *the following expectations exist:*

$$\mu_k^{(l)} = \mathrm{E}[s_k^{(l)} g_k^{(l)}(s_k^{(l)})],$$
$$\nu_k^{(l)} = \mathrm{E}[g_k^{(l)\prime}(s_k^{(l)})],$$
$$\beta_k^{(l)} = \mathrm{E}[g_k^{2(l)}(s_k^{(l)})], \tag{21}$$

and

(iv) *the block one-unit FastICA algorithm is started from the correct de-mixing matrix and stops after a single iteration (19).*

Then, the normalized gain matrix elements $N^{1/2}\mathbf{G}_{k\ell}^{\mathrm{B1U}}$ have asymptotically Gaussian distribution $\mathcal{N}(0, V_{k\ell}^{\mathrm{B1U}})$, where

$$V_{k\ell}^{\mathrm{B1U}} = \frac{\overline{\beta}_{k\ell} + \alpha_{k\ell}^2\overline{\sigma}_{k\ell}^2 - 2\alpha_{k\ell}\overline{\mu}_{k\ell}}{\overline{\tau}_k^2} \tag{22}$$

for $k,\ell = 1,\dots,d$, $k \neq \ell$, provided that $\overline{\tau}_k \neq 0$. Here,

$$\overline{\mu}_k = \frac{1}{M}\sum_{l=1}^{M}\lambda_k^{(l)}\mu_k^{(l)},$$

$$\overline{\nu}_k = \frac{1}{M}\sum_{l=1}^{M}\lambda_k^{(l)}\nu_k^{(l)},$$

$$
\begin{aligned}
&\overline{\tau}_k = \overline{\nu}_k - \overline{\mu}_k,\\
&\overline{\beta}_{k\ell} = \frac{1}{M}\sum_{l=1}^{M} (\lambda_k^{(l)})^2 \beta_k^{(l)} \sigma_\ell^{2(l)},\\
&\overline{\mu}_{k\ell} = \frac{1}{M}\sum_{l=1}^{M} \lambda_k^{(l)} \mu_k^{(l)} \sigma_\ell^{2(l)},\\
&\overline{\nu}_{k\ell} = \frac{1}{M}\sum_{l=1}^{M} \lambda_k^{(l)} \nu_k^{(l)} \sigma_\ell^{2(l)},\\
&\overline{\sigma}_{k\ell}^2 = \frac{1}{M}\sum_{l=1}^{M} \sigma_k^{2(l)} \sigma_\ell^{2(l)},\\
&\alpha_{k\ell} = \overline{\mu}_k + (\overline{\nu}_{k\ell} - \overline{\nu}_k)/2.
\end{aligned} \tag{23}
$$

Proof. See Appendix B.

The practical conclusion of this proposition is that the variance of the gain matrix elements obtained from the block one-unit FastICA is approximately

$$
\mathrm{var}[\mathbf{G}_{k\ell}^{\mathrm{B1U}}] \approx \frac{1}{N} V_{k\ell}^{\mathrm{B1U}}, \quad k \neq \ell. \tag{24}
$$

Consequently, the aim is to minimize $V_{k\ell}^{\mathrm{B1U}}$ subject to free parameters (weights) to achieve the best performance in practice. Note that M need not be necessarily equal to a particular value. The proposition is valid if M is such that distributions of signals are constant within each block.

As will be shown later, all the expectations (21), and consequently (23), are important for computing optimum weights (λ's and later the weights for the refinement) needed to achieve the optimum performance. In practice, the expectations are estimated from estimated signals by sample means. Estimation errors are therefore introduced into the weights. Then, the need is that the weights are not much sensitive to the estimation errors so as not to worsen the final performance of the algorithm in practice.

Here we arrive at the problem with the fully general piecewise stationary model. We have found that the resulting formulas for the weights (not shown here to simplify the text) are overparametrized, which causes the higher sensitivity of the weights to the estimation errors of (21) and of the variances $\sigma_k^{2(l)}$. Therefore, to reduce the number of parameters, we introduce an important simplification by assuming the same (unit) variance of signals in all blocks, i.e.,

$$
\sigma_k^{2(l)} = 1, \quad k = 1, \ldots, d, \; l = 1, \ldots, M. \tag{25}
$$

Although the assumption restricts our theoretical conclusions to constant-variance signals, we will show by simulations that the performance of the method is not depressed in practice when the variance of signals is changing. The main reason is that the expectations in (21) depend on the distribution of signals and reflect thus the variance sufficiently, and the parameters $\sigma_k^{2(l)}$ become redundant. This is yet more apparent when the identity function is considered in the score function estimator. The variances are then involved in the moments (21), because the functions $g_k^{(l)}$ have the form $g_k^{(l)}(x) = \alpha x + h_k^{(l)}(x)$, where $h_k^{(l)}(x)$ is a combination of nonlinearities.

By using the constant-variance assumption, (22) simplifies to

$$
V_{k\ell}^{\mathrm{B1U}} = \frac{\overline{\beta}_k - \overline{\mu}_k^2}{\overline{\tau}_k^2}, \quad k \neq \ell, \tag{26}
$$

where $\overline{\beta}_k = (1/M)\sum_{l=1}^{M} (\lambda_k^{(l)})^2 \beta_k^{(l)}$. Now, we derive the optimal choice of $\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}$ by minimizing (26). The result is described by the following proposition.

Proposition 2. *For a fixed* $k \in \{1, \ldots, d\}$, *minimization of* $V_{k\ell}^{\mathrm{B1U}}$ *given by* (26) *subject to* $\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}$ *is achieved for all* $\ell = 1, \ldots, d$, $\ell \neq k$, *when*

$$
\lambda_k^{(J)} = \frac{1}{M}\left(\frac{\tau_k^{(J)}}{\beta_k^{(J)}} + A_k B_k \frac{\mu_k^{(J)}}{\beta_k^{(J)}}\right), \quad J = 1, \ldots, M, \tag{27}
$$

where

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

and

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

Proof. See Appendix C.

After knowing the performance achieved by the fine-tunings stage BEF2, the final performance of the Block EFICA is given after the refinement step BEF3. The refinement, in the original EFICA, utilizes weights $c_{k\ell}$ given by (15), which, in fact, are functions of the performance achieved by the fine-tunings characterized by $V_{k\ell}^{1U}$. Thanks to this relation, the weights that are optimal for the Block EFICA are simply given when inserting $V_{k\ell}^{\mathrm{B1U}}$ into (15) instead of $V_{k\ell}^{1U}$.

Namely, the optimum weights $c_{k\ell}$ for the Block EFICA refinement are given by

$$
c_{k\ell} = \begin{cases} \dfrac{V_{k\ell}^{\mathrm{B1U}}}{V_{\ell k}^{\mathrm{B1U}} + 1}, & k \neq \ell, \\ 1, & k = \ell. \end{cases} \tag{28}
$$

Similarly, the performance of the Block EFICA is analogous to (17), i.e., for $\mathbf{G}^{\mathrm{BEF}}$ being the resulting gain matrix,

$$
\mathrm{var}[\mathbf{G}_{k\ell}^{\mathrm{BEF}}] \approx \frac{1}{N} \frac{V_{k\ell}^{\mathrm{B1U}}(V_{\ell k}^{\mathrm{B1U}} + 1)}{V_{k\ell}^{\mathrm{B1U}} + V_{\ell k}^{\mathrm{B1U}} + 1}, \quad k \neq \ell. \tag{29}
$$

5.1. Optimal performance

Here, we study the special case when the nonlinearities selected by the score function estimator (20) equal the true score functions, i.e., $g_k^{(l)} = \psi_k^{(l)}$, for $k = 1, \ldots, d$, $l = 1, \ldots, M$.

Similarly to the equations above after (17), it holds that $\beta_k^{(l)} = \nu_k^{(l)} = \kappa_k^{(l)}$, $\mu_k^{(l)} = 1$, and $\tau_k^{(l)} = \gamma_k^{(l)} = \kappa_k^{(l)} - 1$. Next, the formula for λ's (27) simplifies to a constant, namely, $\lambda_k^{(J)} = 1/M$, but we may consider all λ's equal to one, because then $\overline{\beta}_k = \overline{\nu}_k = \overline{\kappa}_k$ and $\overline{\mu}_k = 1$. Now, the performance (26) becomes equal to

$$V_{k\ell}^{\rm B1U} = \frac{1}{\overline{\kappa}_k - 1}. \quad (30)$$

Inserting (30) into (29) we get

$$\mathrm{var}[\mathbf{G}_{k\ell}^{\rm BEF}] \approx \frac{1}{N}\frac{\overline{\kappa}_\ell}{\overline{\kappa}_k\overline{\kappa}_\ell - 1}, \quad k \neq \ell. \quad (31)$$

As compared to the CRLB_2 given by (7), it follows that the Block EFICA is asymptotically efficient within the piecewise stationary model with constant variance signals. Although this does not mean the asymptotic efficiency of Block EFICA for the fully general model, we will show by simulations that its performance is usually very close to the CRLB even when variances of signals are not constant.

The uniformity of the weights (27) for the particular case studied here gives rise to the Uniform Block EFICA, as defined in Section 4.3, because $d \cdot M$ parameters $\lambda_k^{(l)}$ need not be estimated when $g_k^{(l)}(\cdot)$ are assumed to be the score functions. This means further reduction of parameters, which may be useful, for instance, when the number of blocks M is unknown and may be overestimated.

6. Experimental results

We have done several experiments simulating various scenarios to demonstrate good performance and versatility of the proposed Block EFICA algorithm [45]. In comparisons, we select algorithms that are supposed to be the most competitive for a given scenario. Thus, the original symmetric FastICA algorithm [5] with the nonlinearity $g(\cdot) = \tanh(\cdot)$ and the original EFICA algorithm [16,37] are considered as competitive methods within non-Gaussianity-based approaches. In several examples, we also consider the BGL algorithm from [14] that is designed for Gaussian nonstationary signals.

The NSNG algorithm [27] stands for a method belonging to the same class of algorithms as Block EFICA. As stated in Section 1, the method performs well in simple examples with "few" signals, but it is considerably unstable in more realistic scenarios. Therefore, we show its performance only in cases where the method yields meaningful results (Figs. 2 and 6).

A common criterion used in experiments is the interference-to-signal ratio (ISR), for the k-th separated signal defined as

$$\mathrm{ISR}_k = \frac{\sum_{\ell=1,\ell\neq k}^{d} \mathbf{G}_{k\ell}^2}{\mathbf{G}_{kk}^2}, \quad (32)$$

where $\mathbf{G} = \mathbf{WA}$ is the gain matrix computed as the product of the separation matrix $\mathbf{W}$ obtained by an algorithm and the known mixing matrix $\mathbf{A}$. Prior to the computation, the rows of $\mathbf{G}$ are permuted to avoid the indeterminacy of their original order. Such permutation is naturally chosen to yield the best value of the criterion.

For each experiment, we show the average computational loads of methods in legends of the corresponding figures. All simulations were running in Matlab$^{\rm TM}$ on a PC with 3 GHz processor and 2 GB of RAM.

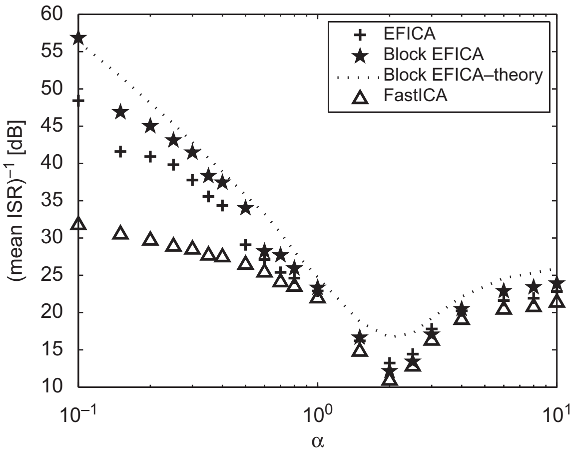

Fig. 2. Mean interference-to-signal ratio of separated signals in the first experiment computed over 100 Monte Carlo trials. Note that CRLB is not defined here, because, in the mixture of 20 signals, some of them are distributed according to RVs with generalized Gaussian distribution with $\alpha \leqslant 0.5$, which have $\kappa = +\infty$; see e.g., Appendix B in [16].

6.1. Validation of the analysis

The examples presented in this subsection aim at validating theoretical conclusions derived in Section 5 and at demonstrating the performance of the Block EFICA in the framework of the piecewise stationary model with constant-variance signals.

To this end, we compare the proposed Block EFICA with the original EFICA algorithm, which performs efficiently when working with signals with generalized Gaussian distributions [16] with parameter α, GGD(α), obeying the basic ICA model (2). However, in the experiments presented here we consider signals with varying distribution from one block to the other. Then, the behavior of EFICA is explained by the analysis of Block EFICA: using the same nonlinear functions in all blocks, the functions cannot match the varying score functions nor the weight for fine-tunings and refinements, thus, the performance of EFICA is *suboptimal*. The same holds for the other FastICA variants.

In the first example, we separate 20 artificial signals of length $N = 10^4$ mixed by a random matrix. Each signal consists of four blocks of the same length $N/4$. The first and the third blocks have Gaussian distribution, which is equivalent with GGD(2), and the second and the fourth blocks have the distribution GGD(α). The parameter α is fixed for each of 20 signals, where its values are uniformly chosen from [0.1, 10]. The variance of all the distributions is one, thus, the signals have constant variance.

Theoretical performance, marked in figures by "theory" in the legend, was estimated from separated signals using (26) and (29). Results of this experiment corroborate validity of the analysis due to proximity of the theoretical results with the empirical ones. They also demonstrate the improved performance of the proposed method compared to EFICA thanks to considering different distributions on the four blocks of signals. We do not demonstrate the

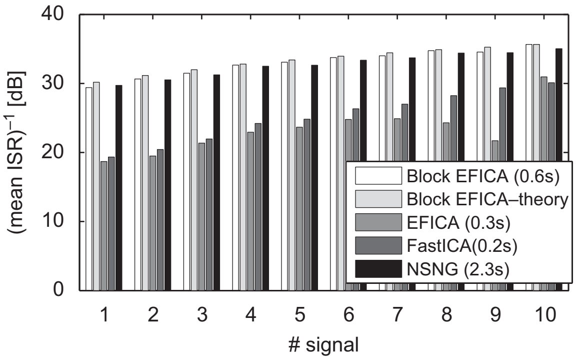

Fig. 3. Mean interference-to-signal ratio of 10 signals of length $N = 10^4$ averaged over 1000 Monte Carlo trials. The first $k \cdot N/10$ samples of the k-th signal are uniformly distributed, and the remainder is Gaussian.

performance of the NSNG algorithm here, because its original implementation[4] is designed for sub-Gaussian signals only, and the method fails to converge in this experiment.

To test a scenario with sub-Gaussian signals, we show in Fig. 3 the performance achieved by separation of 10 signals composed of $M = 10$ blocks. The k-th signal, $k = 1, \ldots, 10$, is uniformly distributed (with variance one) in the first k blocks and Gaussian elsewhere.

Similarly to the previous experiment, this example demonstrates the strongest point of the Block EFICA, which consists in its ability to adapt to varying signal distribution. The same performance was achieved by the NSNG algorithm, and it performed yet better when smaller length of data was considered, which is likely thanks to lower number of parameters compared to Block EFICA. However, also in this scenario, NSNG failed to converge in a few trials. To allow presentation of its performance, the trials where the divergence occurred had to be skipped.

Fig. 4 shows the overall performance averaged over all sources when changing the input parameter M in Block EFICA from 1 to 40. Although performance is optimum for the correct value of $M = 10$, the deterioration of the performance due to overestimation or underestimation of M is not high. For M close to 1 the performance of Block EFICA approaches that of EFICA, which is as expected. Certain local maxima can be observed for M being multiple of 10, which is thanks to fitting the boundaries of blocks exactly to the instants where the distributions of signals are switched. Nevertheless, the negligible improvement demonstrates lower importance of the correct fitting.

The theoretical performance computed using (29) monotonically grows with M. It therefore becomes slightly overoptimistic for higher values of M, because it does not take the practical effect of overparametrization into account. Nevertheless, it may be used in order to choose an effective value of M.

[4] The implementation of the NSNG algorithm was obtained from web-site http://www-lmc.imag.fr/SMS/SASI/bliss.html.

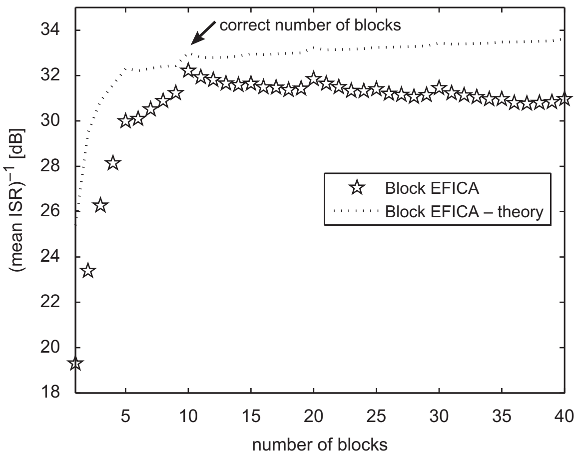

Fig. 4. Average interference-to-signal ratio of 10 sub-Gaussian signals achieved by Block EFICA when changing the number of blocks M considered by the algorithm.

6.2. Signals with changing variance

Since the proposed Block EFICA exploits the piecewise stationary modelling concept, we test its ability to separate nonstationary signals with varying variance. For that purpose, we design a simple experiment where a signal having variable variance is separated from another signal that is stationary. The first (nonstationary) signal has variances, respectively, equal to 1, σ, and σ^2 in the three consecutive blocks of the same length, and the second signal is Gaussian having the constant variance equal to one. An example of the signals for a particular value of the parameter σ, which is considered on interval $(0, 1]$, is shown in Fig. 5.

We consider two situations that differ in selected distribution of the first nonstationary signal. In the first setup, the distribution is Gaussian in all blocks. Then, for σ close to one, where the two signals are almost stationary, the mixture cannot be separated due to Gaussianity of the signals. In the second setup, the distribution is Laplacian, which makes the mixture separable even for σ close to one. The signals can be separated for both cases when σ is close to zero. Then, the first signal is strongly nonstationary and has a different variance-envelope than the second signal, which is the general requirement of the BGL algorithm. Fig. 6 shows results obtained for both settings of the experiment.

The first scenario with Gaussian signals fits the Block Gaussian model. In such a case, the theoretical performance of the BGL algorithm attains corresponding Cramér–Rao bound, here, given by $\text{CRLB}_3 = \text{CRLB}_4$. Therefore, its performance should be optimal, which is confirmed by the results shown by Figs. 6(a) and (b). Similar performance was achieved by the NSNG algorithm without yielding any instability, which reveals its excellent ability to utilize the nonstationarity of signals in simple examples such as the two-dimensional one considered here.

The proposed Block EFICA algorithm achieves comparable results up to $\sigma \in [0.7, 1]$, where the Gaussian signals

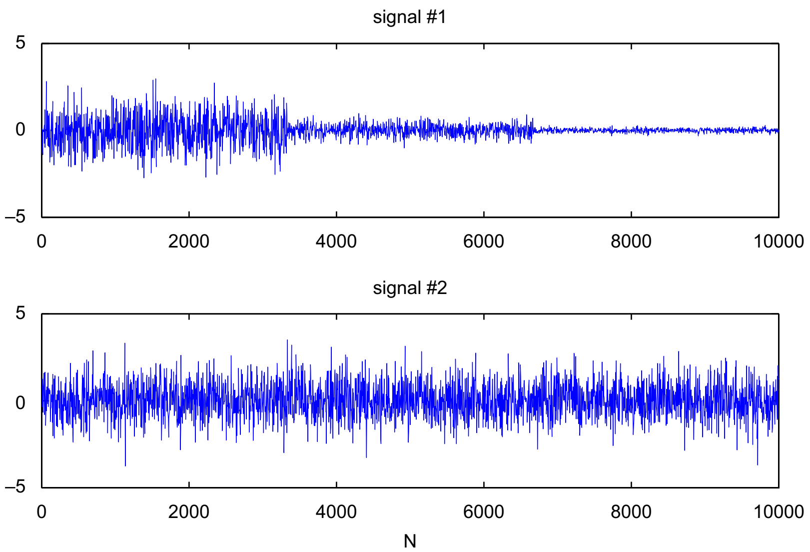

Fig. 5. Illustration of the Gaussian signals of length $N = 10^4$ when the parameter σ that controls the nonstationarity of the first signal equals 0.1.

are almost stationary, which makes them hardly distinguishable for non-Gaussianity-based methods. Hence, the breakdown of the performance is caused by failures of the initialization provided by the symmetric FastICA in the first step BEF1. In our experiments not shown here due to space, we observed that if "good" initialization is guaranteed, the final performance of Block EFICA is comparable with that of the BGL algorithm. Therefore, Block EFICA may be initialized by another method that performs well in this particular case. Nevertheless, our selection, the symmetric FastICA with the test of saddle points, appears to be suitable for most applications as discussed in Section 4.3.

The plots marked by "Block EFICA (identity)" demonstrate further improvement of Block EFICA done via involving the identity function in the score function estimator (see Section 4.4). The better performance shows that the option allows a more effective exploitation of nonstationarity of signals.

The second scenario simulates the case when the original signals exhibit both the non-Gaussianity and the nonstationarity since the distribution of the first signal is Laplacian. Here, the Block EFICA yields performance that is superior to the other methods. The BGL algorithm suffers from stationarity of the signals as σ is approaching one. Conversely, the original EFICA does not utilize effectively their nonstationarity for σ close to zero. The implementation of the NSNG algorithm lacks the ability to accurately estimate the score function of the Laplacian distribution. It has significantly lower performance than EFICA and Block EFICA, nevertheless, its ability to profit both from nonstationarity and non-Gaussianity is confirmed.

6.3. Separation of noisy instantaneous mixtures of speech signals

In this example, we compare performances of algorithms in a noisy scenario. Fig. 7 shows results of separation of 10 speech signals randomly selected from a database, each of length 5000 samples. The signals were mixed by a random matrix, Gaussian noise was added to each mixed channel with the variance corresponding to a given signal-to-noise ratio (input SNR), and the mixture was separated and evaluated in terms of signal-to-interference-plus-noise ratio (SINR). The experiment was designed according to the rules proposed in [39].

Since speech signals often exhibit, beside nonstationarity and non-Gaussianity also spectral diversity, we compare the performance of Block EFICA with the SOBI-RO algorithm from [40] that utilizes the spectral diversity, and ThinICA [11] using also their non-Gaussianity. As can be seen from the results, Block EFICA is not sensitive to the additive noise as inherited from EFICA and FastICA. The achieved SINR decreases smoothly with input SNR. In our example, Block EFICA outperforms the compared algorithms; however, note that the performance strongly depends on properties of the to-be-separated signals.

6.4. Separation of natural convolutive mixture of speech signals

To demonstrate strengths of Block EFICA on real-world data, we present an example where a convolutive mixture of two speech signals recorded by two microphones is separated. The mixture is separated using the procedure from [38] as follows.[5] The first and the most important stage relies on an ICA decomposition of a subspace spanned by delayed signals from microphones, i.e.,

$$x_1(n), x_1(n-1), \ldots, x_1(n-L+1),\\ x_2(n), x_2(n-1), \ldots, x_d(n-L+1), \quad (33)$$

[5] The method from [38] is available at http://itakura.ite.tul.cz/zbynek/tddeconv.htm.

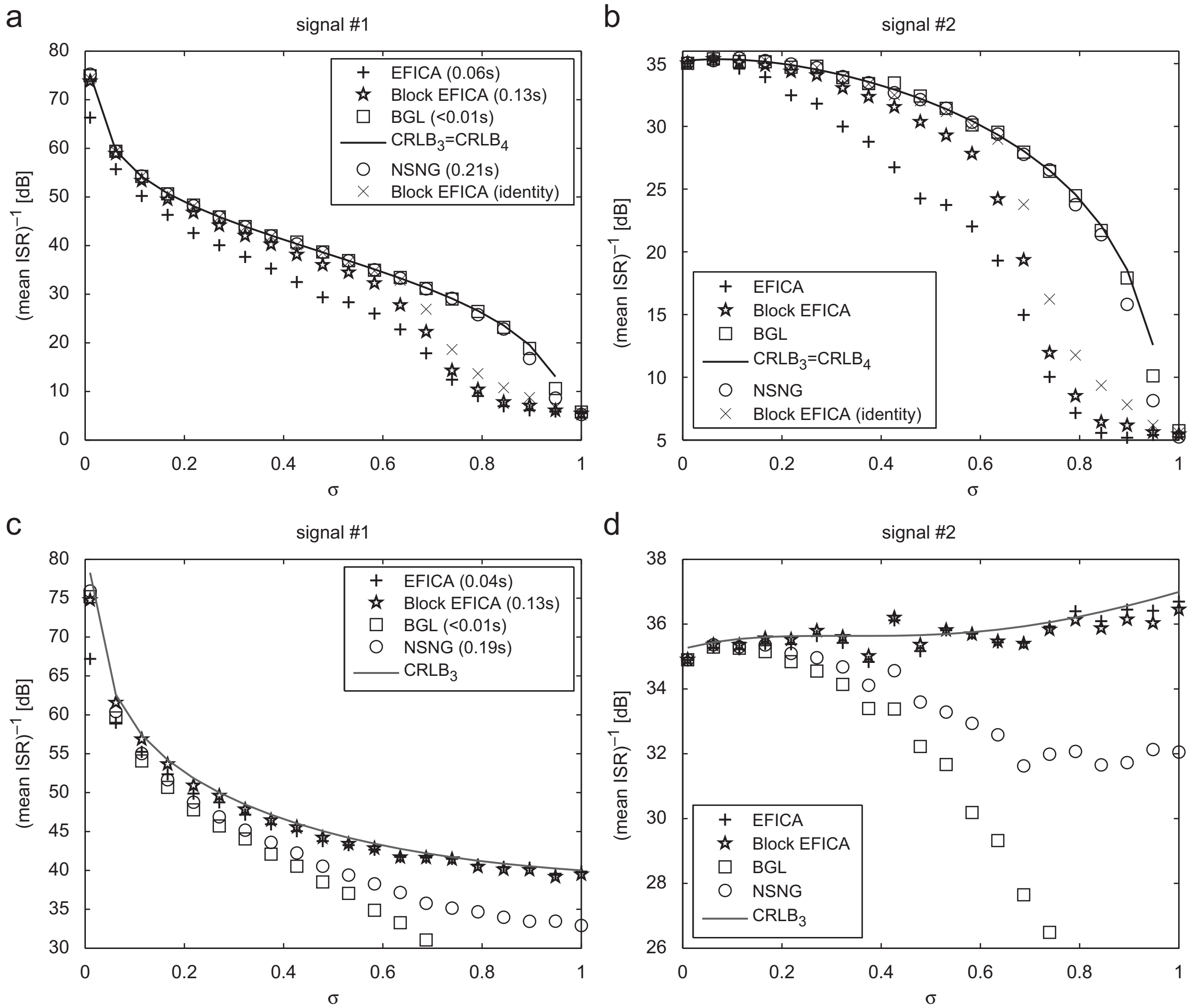

Fig. 6. Results of the experiment with nonstationary signals evaluated by the mean interference-to-signal ratio that was computed from results of 1000 Monte Carlo trials done for each value of the parameter σ. The figures (a) and (b) correspond to the first scenario, where the distribution of signal #1 is Gaussian, while for (c) and (d) the distribution is Laplacian.

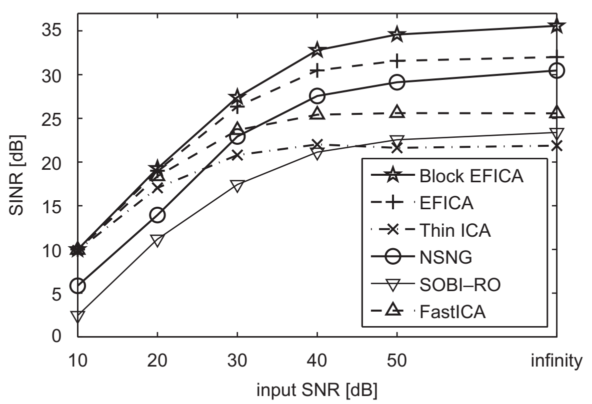

Fig. 7. Results of separation of noisy mixtures of speech signals averaged over 100 independent trials.

where L is the length of separating filters. Note that this way the convolutive mixture problem is transformed into an instantaneous one, thus, we can apply any ICA algorithm that is originally designed for instantaneous mixtures (including Block EFICA). The algorithm thus yields independent components of the subspace (33) that, in fact, correspond to outputs of $d \cdot L$ multiple-input single-output filters of length L. The key objective is that each independent component should contain a contribution of one original source only, which is, in an ideal case, a filtered copy of the source [41–43].

The procedure from [38] continues by grouping the components into clusters that correspond to the same original source. Finally, the clusters (the components in the clusters) are used to reconstruct the original sources; see [38] for further details. Anyway, the idea of this experiment comes from the fact that the final results of

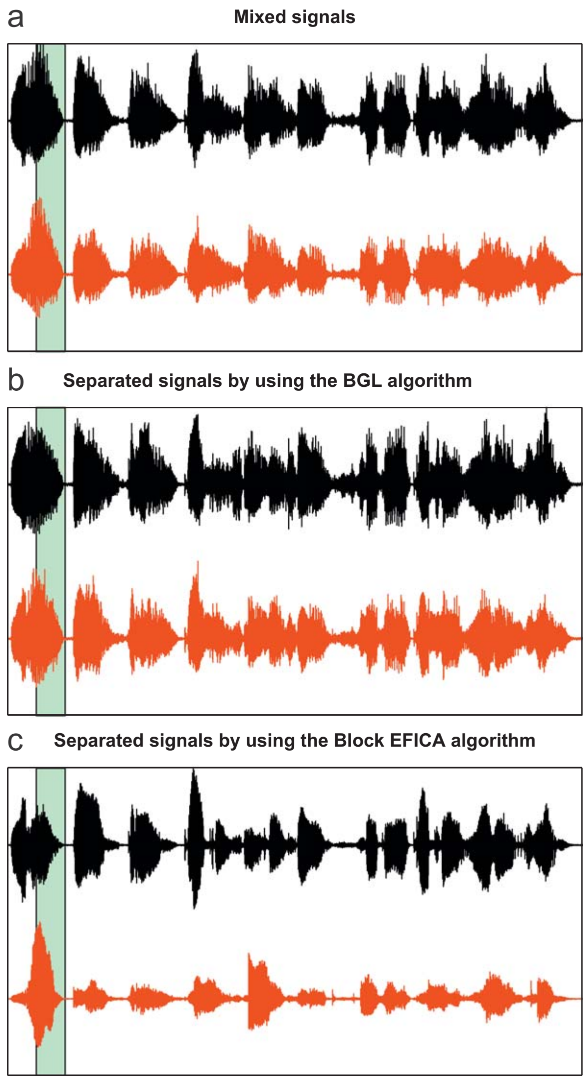

Fig. 8. Results of separation of real-world convolutive mixture of two speech signals recorded by two microphones. Respective ICA methods were applied to the subspace generated by selected data segment of 6000 samples. The segment is delimited by vertical lines in the graphs. (a) Mixed signals. (b) Separated signals by using the BGL algorithm. (c) Separated signals by using Block EFICA algorithm.

the separation provide a benchmark for testing ability of different ICA methods for instantaneous mixtures to separate convolutive audio mixtures, i.e., to yield such independent components that correspond to particular original sources.

Fig. 8(a) shows Lee's data[6] containing real recordings of two speakers (played over loudspeakers) simultaneously saying the digits from one to 10 in English and in Spanish, respectively. The loudspeakers were placed closely to the microphones (60 cm), so direct-path signals and possibly early reflections from the closest objects are much stronger than the other reverberations in the recorded convolutive mixture. Hence, very short separating filters applied through [38] (of the length L) may separate these signals efficiently.

Since the rhythms of the speech signals are similar and synchronized, there occur many short segments (say of length 6000 samples—the sampling frequency is 16 kHz) where the dynamics of the speech signals are very close. Owing to possible changing mixing conditions (e.g., moving sources), the aim is to separate as short segments of signals as possible. However, the similar dynamics of sources in short segments cause malfunctioning of nonstationarity-based methods. From this point of view, the methods that use not only the nonstationarity but also the non-Gaussianity of speech are more flexible, because they do not fail in such situations.

To demonstrate this, Figs. 8(b) and (c) show results of separation with $L = 20$ via BGL[7] and Block EFICA, respectively, when only using a short segment of data for the mixture identification (learning data). Then, the resulting separating filters are applied to the whole signals. Since the mixture is here stationary (the loudspeakers and microphones remain in their positions during the whole recording), the separated signals reveal ability of the ICA methods to separate them using data from the given data segment only.

Since the dynamics of signals are too similar in the chosen segment, the nonstationarity-based BGL algorithm yields poorly separated components of (33), so that average SIR of the finally separated sources is 3.3 dB,[8] while the original SIR of the mixed signals is 3.4 dB. By contrast, the Block EFICA algorithm succeeded to separate the signals yielding average SIR of 12.2 dB, which means "good" result in this convolutive audio source separation task.

7. Conclusions

We have proposed the Block EFICA algorithm that effectively exploits both the non-Gaussianity and the nonstationarity of original signals to separate them. The method efficiently solves the ICA task defined by the piecewise stationary model. It yields comparable performance as methods only intended for marginal cases: the non-Gaussianity-based model or the Block Gaussian model. Namely, it has about the same performance as the EFICA algorithm if the separated signals are stationary and non-Gaussian. In case of Gaussian piecewise stationary signals, Block EFICA is not claimed to be optimum in theory, but in our simulations we have shown that its performance may be close to that of the BGL algorithm that performs optimally in this case.

In so doing, Block EFICA performs best in case of compound scenarios involving non-Gaussian and

[6] Lee's data are available online at http://www.cnl.salk.edu/~tewon/Blind/blind_audio.html.

[7] In fact, the method from [38] utilizes a fast variant of BGL named BG_WEDGE. The algorithm is based on a fast joint diagonalization algorithm with adaptive weights proposed in [17].

[8] The signal-to-interference ratio was evaluated by means of the BSS_EVAL toolbox from [46] that uses projections of signals to avoid indeterminacies due to arbitrary filtering of separated signals. Lee's separated signals were used as the reference "correct" signals.

nonstationary signals. The considered number of blocks M need not be precisely determined, as the method is not highly sensitive to it. Moreover, it yields equivalent performance with that of EFICA when M is equal to one. Finally, Block EFICA provides an appealing alternative to the theoretically optimum NSNG algorithm in terms of better stability and lower computational complexity, especially, when applied to high-dimensional data and, therefore, may be successfully applied to real-world BSS problems.

Acknowledgments

This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and by Grant Agency of the Czech Republic through the projects 102/09/1278 and 102/07/P384.

Appendix A. Derivation of CRLB_3

In this appendix, we provide a simple derivation of CRLB_3 based on results from [32] and the corrections [44].

We start from Eq. (36) of [32] that, for $N=1$, gives the mn-th element of the Fisher information matrix[9] of an independent observation of (2)

$$(\mathbf{F_I})_{mn} = \delta_{ju}\delta_{vi} + \delta_{ji}\delta_{vu}\delta_{vi}(\eta_i - \kappa_i - 2) + \delta_{iu}\delta_{vj}\kappa_i, \tag{34}$$

where $m=(i-1)d+j$ and $n=(u-1)d+v$ with $i,j,u,v=1,\dots,d$, $\eta_i = \mathrm{E}[s_i^2\psi_i^2(s_i)]$, and δ_{ji} is Kronecker's delta. This result can be easily extended for signals with general variance $\sigma_j^2 = \mathrm{E}[s_j^2]$ ([32, p. 1201], the first column, the fourth line of the second item in the enumeration), which gives

$$(\mathbf{F_I})_{mn} = \delta_{ju}\delta_{vi} + \delta_{ji}\delta_{vu}\delta_{vi}\left(\eta_i - \frac{\sigma_v^2}{\sigma_i^2}\kappa_i - 2\right) + \delta_{iu}\delta_{vj}\frac{\sigma_v^2}{\sigma_i^2}\kappa_i, \tag{35}$$

where κ_i and η_i are defined for *normalized* pdfs of the sources in order to be scale-invariant.

Now it follows that the FIM of an observation from the l-th block of the piecewise stationary model (3) should have the block-dependent quantities labelled by the superscript (l), and the FIM of all N independent observations has the mn-th element equal to

$$(\mathbf{F_I})_{mn} = N\left[\delta_{ju}\delta_{vi} + \delta_{ji}\delta_{vu}\delta_{vi}\frac{1}{M}\sum_{l=1}^{M}\left(\eta_i^{(l)} - \frac{\sigma^{2(l)}_v}{\sigma^{2(l)}_i}\kappa_i^{(l)} - 2\right) + \delta_{iu}\delta_{vj}\frac{1}{M}\sum_{l=1}^{M}\frac{\sigma_v^{2(l)}}{\sigma_i^{2(l)}}\kappa_i^{(l)}\right]. \tag{36}$$

The structure of the FIM (36) is the same as in case of the basic ICA model (2), i.e., it can be written in a form $\mathbf{F_I} = \mathbf{P} + \mathbf{\Sigma}$ with $\mathbf{P}$ being a special permutation matrix and $\mathbf{\Sigma}$ being diagonal

$$\mathbf{\Sigma} = \frac{1}{M}\sum_{l=1}^{M}\mathrm{diag}\left[\eta_1^{(l)} - 2, \frac{\sigma_2^{2(l)}}{\sigma_1^{2(l)}}\kappa_1^{(l)}, \dots, \frac{\sigma_d^{2(l)}}{\sigma_1^{2(l)}}\kappa_1^{(l)}, \frac{\sigma_1^{2(l)}}{\sigma_2^{2(l)}}\kappa_2^{(l)}, \eta_2^{(l)} - 2, \frac{\sigma_3^{2(l)}}{\sigma_2^{2(l)}}\kappa_2^{(l)}, \dots, \frac{\sigma_d^{2(l)}}{\sigma_2^{2(l)}}\kappa_2^{(l)}, \dots, \frac{\sigma_1^{2(l)}}{\sigma_d^{2(l)}}\kappa_d^{(l)}, \dots, \frac{\sigma_{d-1}^{2\ (l)}}{\sigma_d^{2(l)}}\kappa_d^{(l)}, \eta_d^{(l)} - 2\right]. \tag{37}$$

Therefore, the inversion of $\mathbf{F_I}$ can be derived using Appendix D of [32]; see the simplification due to the corrections. Using appropriate substitutions according to (90) in [32], the resulting CRLB_3 given by (8) readily follows.

Appendix B. Proof of Proposition 1

We will follow the easiest way by generalizing proof of analogous proposition in [32] (see the Appendix A therefrom). Similar notations will be used, namely, $\mathbf{s}_k$ will be $N \times 1$ vector of samples of the k-th original signal, i.e., the k-th row of $\mathbf{S}$, with the difference that the l-th block of N/M samples is distributed according to RV $s_k^{(l)}$. Owing to the indeterminacy of scale of original signals, the variances of $s_k^{(l)}$ can be assumed to be such that $\mathbf{s}_k$ has unit scale (assumption (i) of the proposition).

Next, the vector $\mathbf{u}_k$ contains normalized elements of $\mathbf{s}_k$ so that $\mathbf{u}_k$ has the second-order sample-moment exactly equal to one. The vectors $\mathbf{z}_k$ 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 elementwise, so that function $\lambda_k^{(l)}g_k^{(l)}(\cdot)$ applies to the l-th block of N/M elements.

Now, using the third assumption of Proposition 1 given by (21), Eqs. (40) and (41) from [32] change, respectively, to

$$N^{-1}\mathbf{s}_k^T g_k(\mathbf{s}_k) \overset{N\to+\infty}{\longrightarrow} \overline{\mu}_k, \tag{38}$$

$$N^{-1}g_k'^T(\mathbf{s}_k)\mathbf{1}_N \overset{N\to+\infty}{\longrightarrow} \overline{\nu}_k. \tag{39}$$

Note that ν denotes the same expectations that are in [32] denoted by ρ. $\mathbf{1}_N$ stands for $N \times 1$ vector of ones.

Using this, all Eqs. (42)–(64) in [32] change according to the substitutions

$$\mu_k \leftarrow \overline{\mu}_k, \tag{40}$$

$$\rho_k \leftarrow \overline{\nu}_k. \tag{41}$$

The only exceptions are Eqs. (42), (45), and (62), which should be revised due to different variance in blocks, and it gives, respectively,

$$N^{-1}g_k'^{\,T}(\mathbf{s}_k)(\mathbf{s}_\ell \odot \mathbf{s}_\ell) \overset{N\to+\infty}{\longrightarrow} \overline{\nu}_{k\ell}, \tag{42}$$

$$\mathbf{g}_k'^{\,T}(\mathbf{u}_\ell \odot \mathbf{u}_\ell) = N\overline{\nu}_{k\ell} + o_p(N), \tag{43}$$

$$\mathrm{E}[(\mathbf{g}_k^T\mathbf{u}_\ell)^2] = N\overline{\beta}_{k\ell}, \tag{44}$$

where $\mathbf{g}_k$ is the simplified notation of $g_k(\mathbf{u}_k)$. Recomputation of (65), (71), and (75) in [32] using the above substitutions readily yields the result of the proposition given by (22).

[9] In the corrections [44], it is shown that the first term in (36) of [32] should be removed. This means that, for $N=1$, the relation is correct.

Appendix C. Proof of Proposition 2

The criterion (26) can be written in the form

$$V_{k\ell}^{\mathrm{B1U}} = \frac{\mathbb{l}_k^T \mathbf{\Gamma}_k \mathbb{l}_k}{\mathbb{l}_k^T \mathbb{t}_k \mathbb{t}_k^T \mathbb{l}_k} \tag{45}$$

with

$$\mathbb{l}_k = [\lambda_k^{(1)}, \ldots, \lambda_k^{(M)}]^T, \tag{46}$$

$$\mathbf{\Gamma}_k = \mathrm{diag}[\beta_k^{(1)}, \ldots, \beta_k^{(M)}] - \frac{1}{M} \mathbb{m}_k \mathbb{m}_k^T, \tag{47}$$

$$\mathbb{m}_k = [\mu_k^{(1)}, \ldots, \mu_k^{(M)}]^T, \tag{48}$$

$$\mathbb{t}_k = [\tau_k^{(1)}, \ldots, \tau_k^{(M)}]^T, \tag{49}$$

$$\tau_k^{(l)} = \nu_k^{(l)} - \mu_k^{(l)}. \tag{50}$$

The goal is to minimize (45) subject to elements of $\mathbb{l}_k$, which is equivalent with maximizing

$$\max_{\mathbb{l}_k} \frac{\mathbb{l}_k^T \mathbb{t}_k \mathbb{t}_k^T \mathbb{l}_k}{\mathbb{l}_k^T \mathbf{\Gamma}_k \mathbb{l}_k}. \tag{51}$$

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 $\mathbf{\Gamma}_k$ ($V_{k\ell}^{\mathrm{B1U}}$ denotes variance, which must be always nonnegative). Since (45) is invariant subject to nonzero multiple of $\mathbb{l}_k$, we can introduce a constraint $\|\mathbb{l}_k\| = \text{const.}$, and (51) can be written in the form of classical eigenvalue problem

$$\max_{\|\mathbb{y}_k\|=1} \frac{\mathbb{y}_k^T \mathbf{\Gamma}_k^{-1/2} \mathbb{t}_k \mathbb{t}_k^T \mathbf{\Gamma}_k^{-1/2} \mathbb{y}_k}{\mathbb{y}_k^T \mathbb{y}_k}. \tag{52}$$

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 (52), is $\mathbb{y}_k = \mathbf{\Gamma}_k^{-1/2} \mathbb{t}_k$. Hence, $\mathbb{l}_k$ that minimizes (45) is

$$\mathbb{l}_k = \mathbf{\Gamma}_k^{-1} \mathbb{t}_k. \tag{53}$$

Using the matrix inversion lemma for computation of $\mathbf{\Gamma}_k^{-1}$, (27) follows.

References

[1] A. Cichocki, S.-I. Amari, Adaptive Signal and Image Processing: Learning Algorithms and Applications, Wiley, New York, 2002.
[2] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley-Interscience, New York, 2001.
[3] P. Comon, Independent Component Analysis, a new concept?, Signal Processing 36 (3) (April 1994) 287–314.
[4] J.-F. Cardoso, A. Souloumiac, Blind beamforming from non-Gaussian Signals, IEE Proceedings—F 140 (6) (December 1993) 362–370.
[5] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Transactions on Neural Networks 10 (3) (1999) 626–634.
[6] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Computation 7 (1995) 1129–1159.
[7] K. Matsuoka, M. Ohya, M. Kawamoto, A neural net for blind separation of nonstationary signals, Neural Networks 8 (3) (May 1995) 411–419.
[8] S. Choi, A. Cichocki, S. Amari, Equivariant nonstationary source separation, Neural Networks 15 (1) (January 2002) 121–130.
[9] J.-I. Hirayama, S.-I. Maeda, S. Ishii, Markov and semi-Markov switching of source appearances for nonstationary independent component analysis, IEEE Transactions on Neural Networks 18 (5) (September 2007).
[10] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, E. Moulines, A blind source separation technique using second-order statistics, IEEE Transactions on Signal Processing 45 (February 1997) 434–444.
[11] S. Cruces, A. Cichocki, Combining blind source extraction with joint approximate diagonalization: thin algorithms for ICA, in: Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), Nara, Japan, April 2003, pp. 463–468.
[12] T.-W. Lee, M. Girolami, T.J. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, Neural Computation 11 (2) (1999) 417–441.
[13] D.T. Pham, P. Garat, Blind separation of mixture of independent sources through a quasi-maximum likelihood approach, IEEE Transactions on Signal Processing 45 (7) (July 1997) 1712–1725.
[14] D.-T. Pham, J.-F. Cardoso, Blind separation of instantaneous mixtures of non stationary sources, IEEE Transactions on Signal Processing 49 (9) (2001) 1837–1848.
[15] R. Boscolo, H. Pan, V.P. Roychowdhury, Independent component analysis based on nonparametric density estimation, IEEE Transactions on Neural Networks 15 (1) (2004) 55–65.
[16] Z. Koldovský, P. Tichavský, E. Oja, Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér–Rao lower bound, IEEE Transactions on Neural Networks 17 (5) (September 2006).
[17] P. Tichavský, A. Yeredor, Fast approximate joint diagonalization incorporating weight matrices, IEEE Transactions on Signal Processing 57 (3) (March 2009) 878–891 codes available at ⟨http://si.utia.cas.cz/downloadPT.htm⟩.
[18] L. Parra, P. Sajda, Blind source separation via generalized eigenvalue decomposition, Journal of Machine Learning Research 4 (2003) 1261–1269.
[19] J.-F. Cardoso, A. Souloumiac, Jacobi angles for simultaneous diagonalization, SIAM Journal on Matrix Analysis and Applications 17 (1996) 161–164.
[20] F. Meinecke, A. Ziehe, M. Kawanabe, K.R. Müller, A resampling approach to estimate the stability of one-dimensional or multidimensional independent components, IEEE Transactions on Biomedical Engineering 49 (12) (December 2002) 1514–1525.
[21] H. Buchner, R. Aichner, W. Kellermann, TRINICON: a versatile framework for multichannel blind signal processing, Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Montreal, Canada 3 (May 2004) 889–892.
[22] K.R. Müller, P. Philips, A. Ziehe, JADETD: combining higher-order statistics and temporal information for blind source separation (with noise), in: Proceedings of ICA '99, Aussois, January 1999.
[23] S. Choi, A. Cichocki, A. Belouchrani, Second order nonstationary source separation, Journal of VLSI Signal Processing 32 (1–2) (August 2002) 93–104.
[24] A. Hyvärinen, A unifying model for blind separation of independent sources, Signal Processing 85 (7) (2005) 1419–1427.
[25] P. Tichavský, Z. Koldovský, A. Yeredor, G.G. Herrero, E. Doron, A hybrid technique for blind non-Gaussian and time-correlated sources using a multicomponent approach, IEEE Transactions on Neural Networks 19 (3) (March 2008) 421–430.
[26] J.-F. Cardoso, D.T. Pham, Separation of non stationary sources. Algorithms and performance, in: S.J. Roberts, R.M. Everson (Eds.), Independent Components Analysis: Principles and Practice, Cambridge University Press, Cambridge, 2001, pp. 158–180.
[27] D.-T. Pham, Blind separation of non stationary non Gaussian sources, in: Proceeding of the EUSIPCO 2002 Conference, Toulouse, France, September 2002.
[28] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, S. Hosseini, Extension of EFICA algorithm for blind separation of piecewise stationary non Gaussian sources, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, Nevada, USA, April 2008, pp. 1913–1916.
[29] R.C. Rao, Linear Statistical Inference and its Applications, second ed., Wiley, New York, 1973.
[30] O. Shalvi, E. Weinstein, Maximum likelihood and lower bounds in system identification with non-Gaussian inputs, IEEE Transactions on Information Theory 40 (2) (March 1994) 328–339.
[31] J.-F. Cardoso, Blind signal separation: statistical principles, Proceedings of the IEEE 90 (8) (October 1998) 2009–2026.
[32] P. Tichavský, Z. Koldovský, E. Oja, Performance analysis of the FastICA algorithm and Cramér–Rao bounds for linear independent

component analysis, IEEE Transactions on Signal Processing 54 (4) (April 2006).
[33] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, S. Hosseini, Performance Analysis of Extended EFICA Algorithm, Technical Report No. 2199, ÚTIA, AV ČR, 2007. Available at ⟨http://itakura.ite.tul.cz/zbynek/publications.htm⟩.
[34] B.W. Silvermann, Density Estimation for Statistics and Data Analysis, Chapman & Hall, London, 1986.
[35] J. Málek, Z. Koldovský, S. Hosseini, Y. Deville, A variant of EFICA algorithm with adaptive parametric density estimator, in: 8th International Workshop on Electronics, Control, Modelling, Measurement, and Signals (ECMS 2007), Liberec, Czech Republic, May 2007, pp. 79–84.
[36] Y. Deville, M. Benali, F. Abrard, Differential source separation for underdetermined instantaneous or convolutive mixtures: concept and algorithms, Signal Processing 84 (10) (October 2004) 1759–1776.
[37] P. Tichavský, Z. Koldovský, E. Oja, Speed and accuracy enhancement of linear ICA techniques using rational nonlinear functions, in: Proceedings of ICA2007, September 2007, pp. 285–292.
[38] Z. Koldovský, P. Tichavský, Time-domain blind audio source separation using advanced component clustering and reconstruction, in: Proceedings of Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA 2008), Trento, Italy, May 2008, pp. 216–219.
[39] Z. Koldovský, P. Tichavský, Methods of fair comparison of performance of linear ICA techniques in presence of additive noise, in: Proceedings of ICASSP 2006, Toulouse, no. V, May 2006, pp. 873–876.
[40] A. Belouchrani, A. Cichocki, Robust whitening procedure in blind source separation context, Electronics Letters 36 (24) (2000) 2050–2053.
[41] H. Buchner, R. Aichner, W. Kellermann, A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics, IEEE Transactions on Speech and Audio Processing 13 (1) (January 2005) 120–134.
[42] S. Douglas, M. Gupta, H. Sawada, S. Makino, Spatio-temporal FastICA algorithms for the blind separation of convolutive mixtures, IEEE Transactions on Audio, Speech and Language Processing 15 (5) (July 2007) 1511–1520.
[43] J. Thomas, Y. Deville, S. Hosseini, Time-domain fast fixed-point algorithms for convolutive ICA, IEEE Signal Processing Letters 13 (4) (2006) 228–231.
[44] P. Tichavský, Z. Koldovský, E. Oja, Corrections of the 'performance analysis of the FastICA algorithm and Cramér–Rao bounds for linear independent component analysis, IEEE Transactions on Signal Processing 56 (4) (April 2008) 1715–1716.
[45] [online] The Matlab^{TM} implementation of the Block EFICA algorithm: ⟨http://itakura.ite.tul.cz/zbynek/eefica.htm⟩.
[46] E. Vincent, R. Gribonval, C. Févotte, Performance measurement in blind audio source separation, IEEE Transactions on Speech and Audio Processing 14 (4) (July 2006) 1462–1469.