QUICK ALGORITHMS FOR INDEPENDENT VECTOR EXTRACTION AND ANALYSIS BASED ON EXACT NEWTON-RAPHSON OPTIMIZATION

Zbyněk Koldovský¹ and Václav Kautský²

¹Acoustic Signal Analysis and Processing Group, Faculty of Mechatronics, Informatics and Interdisciplinary Studies, Technical University of Liberec, Liberec, Czech Republic
²Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic

ABSTRACT

We propose new algorithms for the joint blind extraction and analysis of independent vector/components in linear complex-valued instantaneous mixtures. An efficient parameterization of the mixing and de-mixing matrix is used, which enables us to derive algorithms for extraction based on the exact Newton-Raphson update rule. For the complete independent vector/component analysis, orthogonally constrained algorithms are run in parallel. In experiments, the proposed methods show fast and stable convergence that is competitive to the state-of-the-art algorithms. Moreover, our approach is applicable also to piecewise determined mixtures with constant separating vectors.

Index Terms— Blind Source Separation, Blind Source Extraction, Independent Component Analysis, Independent Vector Analysis, Newton-Raphson Algorithm

1. INTRODUCTION

Independent Component and Vector Analysis (ICA/IVA) are popular methods for Blind Source Separation (BSS) where the goal is to separate K linear instantaneous mixtures of original signals into components that correspond to these signals up to indeterminable scales and order; K = 1, in case of ICA [1]. The main assumption is that the original signals in each mixture are mutually independent, so the signals can be extracted as independent components. IVA separates K > 1 mixtures jointly by using an additional assumption that there are dependencies between independent components from different mixtures which form so-called vector components [2].

Numerous algorithms have been proposed for the real and complex-valued ICA/IVA. Among the most successful methods belong the Natural Gradient (NG) algorithm [3,4] and its generalized version for IVA [4]. FastICA for real-valued [5] and complex-valued signals [6,7] and for IVA [8,9], and the auxiliary-function based algorithms AuxICA [9] and AuxIVA [10]. NG utilizes a gradient-based optimization strategy, which appeared to be useful especially in on-line processing deployments [11]. However, the method is slow and prone to local extremes. The fixed-point optimization-based FastICA and majorization-based AuxICA provide much faster and stable algorithms; for example, AuxIVA was shown successful also in on-line audio source separation [12,13].

There have been several attempts to derive the exact Newton-Raphson (NR) optimization for ICA/IVA in order to obtain fast and stable methods. However, heuristic modifications and approximate Hessians have been used instead in order to avoid high computational burden or bad conditioning issues caused by the Hessian inverse matrix computation; see, e.g., [14, 15]. For example, FastICA has been originally derived as an approximate NR algorithm [5,8,10].

In this paper, we propose new algorithms for the blind extraction of independent vector/components, which, when applied in parallel, can serve also for the complete ICA/IVA. The extraction algorithms are derived based on the recent problem formulation called Independent Component/Vector Extraction (ICE/IVE) [17]. Here, an optimum parameterization of the mixing and de-mixing matrix is used. Based on this, the computations of exact NP update rules are straightforward. Two efficient algorithms are proposed this way, one of which is useful also for recent extensions of the standard instantaneous mixing models: Piecewise determined mixing models [18,19]. Matlab codes of the methods are available online.

The paper is organized as follows. The next section introduces ICA/IVA and the extraction sub-problem and its solution through ICE/IVE. The new algorithms are derived in Section 3. Section 4 is focused on experimental validations and comparisons with the previous ICA/IVA algorithms. Section 5 summarizes the main outcomes of this work.

Notations: Plain, bold, and bold capital letters denote, respectively, scalars, vectors, and matrices. Upper index T, H, or · denotes, respectively, transposition, conjugate transpose, or complex conjugate. The Matlab convention for matrix/vector concatenation will be used, e.g., [1: g] = [1, g] T . E[·] stands for the expectation operator, and E[·] is the average taken over all available samples of the argument. {w k} k is a short notation of w 1, . . . , w K.

2. PROBLEM FORMULATION

Consider K instantaneous mixtures (datasets) of the same dimension d. Let N samples of each mixture be available and be divided into T blocks (for simplicity) of the same length N b, i.e., N = T · N b . A sample of the kth mixture in the tth block is represented by

\[ x_{k,t} = A_{k,t} s_{k,t}, \quad k = 1, \ldots, K, \quad t = 1, \ldots, T, \]  

(1)

where A k,t is the d × d non-singular mixing matrix, and s k,t = [s k,t,1, . . . , s k,t,d] T are independent random variables. The elements of a vector component s i,t = [s i,t,1, . . . , s i,t,d] T , i = 1, . . . , m, are

This work was supported by The Czech Science Foundation through Project No. 17-00902S.

https://asap.ite.tul.cz/downloads/ice/
allowed to feature higher-order dependencies \cite{2}. This is the joint piecewise determined instantaneous mixing model \cite{18}; for \( T = 1 \), it coincides with the standard IVA mixing model and, if also \( K = 1 \), it corresponds to ICA. We will consider complex-valued signals and parameters, nevertheless, the conclusions of this paper are valid for the real-valued case as well.

The problem of separating \( d \) sources from each mixture can be formulated as \( d \) Blind Source Extraction (BSE) tasks, each of which the goal is to extract one independent (vector) source. We now turn to a mixing system parametrization that is suitable for BSE.

Without any loss on generality \cite{1}, let the first source of interest (SOI) to be extracted. To this end, it was shown in \cite{17} that the mixing matrices can be reparametrized as

\[
A_{\text{IVF}}^{k,t} = (a^{k,t} \cdot q^{k,t}) = \left( \gamma^{k,t}_1 \frac{1}{\gamma^{k,t}_1} (h^{k,t}H - I_{d-1}) \right),
\]

where \( a^{k,t} = [\gamma^{k,t}_1; g^{k,t}] \) is the mixing vector corresponding to the first column of \( A^{k,t} \), and \( w^{k,t} = [\beta^{k,t}; h^{k,t}] \) is the separating vector that corresponds to the first row of \( W_{\text{IVF}}^{k,t} = (A_{\text{IVF}}^{k,t})^{-1} \) and also to the first row of \( (A^{k,t})^{-1} \). It holds that

\[
W_{\text{IVF}}^{k,t} = \left( w^{k,t}H \right) = \left( \begin{array}{c} \beta^{k,t} \gamma^{k,t}_1 \\ g^{k,t} \end{array} \right) = \left( \begin{array}{c} h^{k,t}H \\ -\gamma^{k,t}_1 I_{d-1} \end{array} \right),
\]

where \( B^{k,t} = [\gamma^{k,t}_1 - \gamma^{k,t}_1 I_{d-1}] \) is called blocking matrix as it satisfies that \( B^{k,t}a^{k,t} = 0; I_d \) denotes the \( d \times d \) identity matrix. The parameter vectors \( a^{k,t} \) and \( w^{k,t} \) obey the distortionless constraint \( (w^{k,t})H a^{k,t} = 1 \), hence saying that

\[
\gamma^{k,t}_1 (\beta^{k,t})^* = 1 - (h^{k,t})^* g^{k,t}.
\]

The mixed signals are thus modeled as \( x^{k,t} = A_{\text{IVF}}^{k,t} y^{k,t} \), where \( y^{k,t} = [s^{k,t}_1; z^{k,t}] \), and the task is to find \( a^{k,t} \) and \( w^{k,t} \) such that the extracted SOI \( s^{k,t}_1 = (w^{k,t})^* x^{k,t} \) and the background signals \( z^{k,t} = B^{k,t} x^{k,t} \) are independent.

For \( T > 1 \), special variants of \cite{2}-(5) with a reduced number of parameters have been considered in \cite{18,19}. In this paper, we will consider the variant called constant separating vector (CSV) where \( w^{k,t} \) are assumed to be constant over \( t \). CSV appears to be useful, e.g., for extracting a moving SOI \cite{18}. In CSV, the mixing and det-mixing vectors are parametrized as

\[
A_{\text{CSV}}^{k,t} = \left( \begin{array}{c} \gamma^{k,t}_1 \frac{1}{\gamma^{k,t}_1} (h^{k,t})^* \\ g^{k,t} \end{array} \right),
\]

\[
W_{\text{CSV}}^{k,t} = \left( \begin{array}{c} \beta^{k,t} \\ (h^{k,t})^* \end{array} \right) = \left( \begin{array}{c} -\gamma^{k,t}_1 I_{d-1} \end{array} \right).
\]

CSV coincides with the standard mixing models when \( T = 1 \).

Let \( p(s^1_t) \) and \( p(x^{k,t}) \) denote the joint pdf of the vector component \( s^1_t \) and of \( z^{k,t} \), respectively. For simplicity, \( p(s^1_t) \) is assumed constant over \( t \), and the background signals are assumed to be mutually independent across the mixtures. Then, for the CSV model, the joint pdf of one sample of \( x^{k,t} \) reads

\[
p_{x^{k,t}}(\{x^{k,t}\}) = p(s^1_t)^K \prod_{k=1}^K p_{x^{k,t}}(\{x^{k,t}\}) \mid \det \begin{vmatrix} W_{\text{CSV}}^{k,t} \end{vmatrix}^2.
\]

The pdf of the entire batch of data (independently distributed) is \( \prod_{t=1}^T \prod_{k=1}^K p_{x^{k,t}}(\{x^{k,t}\}) \). By using \( \mid \det \begin{vmatrix} W_{\text{CSV}}^{k,t} \end{vmatrix}^2 = |\gamma^{k,t}_1|^{2d-2} \), (Eq. (5) in \cite{17}), the contrast function derived from the log-likelihood function is

\[
\text{C}\left(\{w^k, a^{k,t}\}_{k,t}\right) = \begin{cases} \mathbb{E} \left[ \log f(\{(w^k)^H x^{k,t}\}) \right] \\ + \sum_{k=1}^K \mathbb{E} \left[ \log p_{w^k} (B^{k,t} x^{k,t}) \right] + (d - 2) \sum_{k=1}^K \log |\gamma^{k,t}_1| \end{cases},
\]

where \( f(\cdot) \) is a model pdf that should provide a suitable substitute for the unknown \( p(\cdot) \), and \( < \cdot, \cdot >_t \) denotes the average taken over \( t = 1, \ldots, T \).

3. PROPOSED ALGORITHMS

3.1. Methods for (joint) Blind Source Extraction

Since there is the inherent ambiguity of order in ICA/IVA, it depends on the initialization which source is being extracted by the given BSE method. Let us assume that an initial value for \( w^k \), \( k = 1, \ldots, K \), be given. In \cite{17}, gradient algorithms for \( T = 1 \) were studied where the mixing or separating vectors were sought through optimizing the contrast function. The orthogonal constraint was imposed on pairs \( a^k \) and \( w^k \) in order to avoid saddle points of the contrast function where the vectors correspond to different sources. For example, the gradient w.r.t. \( w^k \) was derived in \cite{17} under the constraint (written with index \( t \) for general use when \( T \geq 1 \))

\[
a^{k,t} = \frac{C^{k,t}_w w^k}{\left( w^k \right)^H C^{k,t}_w w^k},
\]

where \( C^{k,t}_w = \mathbb{E} \left[ (x^{k,t})^H (x^{k,t}) \right] \), leading to the method referred to as OGIVEw. For \( T \geq 1 \), the method was extended to BOGIVEw in \cite{18}.

In the following, we derive algorithms through computing unconstrained gradients and Hessians of \( \text{C} \).[6] Both methods impose the orthogonal constraint after updating the separating vectors in the direction given by the Newton-Raphson rule.

3.1.1. QuickIVE-1

In this algorithm, \( a^{k,t} \) are treated as independent parameters, and \( \beta^{k,t} \) is dependent through \( \mathbb{I} \), i.e., \( \beta^{k,t} = (1 - (h^{k,t})^* g^{k,t})/\gamma^{k,t}_1 \). Since this constraint must be satisfied for every \( t \), the algorithm presented here is suitable only when \( T = 1 \) (standard IVE, not CSV). Therefore, we assume \( T = 1 \) for now and omit index \( t \) in equations.

Now, \( w^k = \left[ 1 - (h^{k,t})^* g^{k,t} \right]/\gamma^{k,t}_1 \), so it holds that

\[
\frac{\partial (w^k)^H x^k}{\partial (h^k)^H} = -\frac{1}{\gamma^k_1} B^k x^k + x^k = -\frac{1}{\gamma^k_1} B^k x^k + \frac{1}{\gamma^k_1} z^k,
\]

where \( x^k = [x^1; x^2] \). Using this identity, straightforward computations give

\[
\nabla h^k = \frac{\partial \text{C}}{\partial (h^k)^H} = \frac{1}{\gamma^k_1} \mathbb{E} \left[ \phi^k z^k \right],
\]

\[
H^k = \frac{\partial^2 \text{C}}{\partial (h^k)^H \partial (h^k)^H} = \frac{1}{\gamma^k_1^2} \mathbb{E} \left[ \phi^k z^k \right]^2 T,
\]
we will assume circular sources, for which the NR update rule is given by
\[
\Delta h^k = - (H_{Wk})^{-1} \nabla h^k,
\] (13)
where \(\Delta h^k = h_{new}^k - h^k\). For general (non-circular) update rule, see Eq. 21 in [20].

By taking the initial values of \(w^k\), the QuickIVE-1 algorithm (QuickICE-1 for \(K = 1\)) proceeds in three main steps. For every \(k = 1, \ldots, K\),
1. update \(h^k\) according to (12).
2. put \(\beta^k = (1 - (g_k^k)^H h^k)/(g_k^k)^2\),
3. normalize updated separating vectors as \(w^k \leftarrow w_{new}^k/\sqrt{(w_{new}^k)^H \hat{C}_k w_{new}^k}\)
and update the mixing vectors using (7).

It is worth noting that, in QuickIVE-1, \(H_{Wk}\) can be well approximated by \(H_{Wk} \approx -\gamma^k \nabla \phi_k\), where \(\gamma_k \approx 1 \frac{\|x_k^k\|}{\|x_k^k\| \hat{C}_k}\), (14)
where \(\hat{C}_k = \hat{E}(z^k x_k^k)^H = B^2 \hat{C}_k (B^H)^H\) and \(\rho_k^H = \hat{E}(\partial \phi_k^H / \partial \hat{C}_k)\).
This is justified by the fact that \((w^k)^H x_k^k\) and \(z^k\) converge to independent variables. Moreover, once \((C_k^k)^{-1}\) is known, \((C_k^k)^{-1}\) can be computed using the matrix inversion lemma, which brings a computationally cheap implementation.

3.1.2. QuickIVE-2
In this variant, \(\gamma^k,t = (1 - (h_k^k)^H g_k^k,t)/(\beta_k^t)^H\) while \(\beta^k\) is treated as independent; \(T\) can be greater than one. Compared to the previous subsection, the second and the third term in (8) is now depending on \(w^k\), and we have to cope with the unknown density \(p_{g_k,t}\). A possible choice is the circular Gaussian pdf with zero mean and covariance given by \(\hat{C}_k^H\); see, e.g., [17] for the justification of this choice. Then, the term \(\hat{E}[\log p_{g_k,t}(B^H x_k^k)]\) in (8) can be replaced by \(-\hat{E}[\partial \phi_k^H (\hat{C}_k^H x_k^H + \text{const.})\partial g_k^t]\), and the gradient of the contrast function (when \(g_k^k\) are treated as constants) reads
\[
\nabla w^k = \frac{\partial \phi_k}{\partial (w^k)^H} = \left< a_k^t - \hat{E}[\phi_k^H x_k^H]\right>_t, \quad (15)
\]
As explained in [17], this gradient needs to be modified so that it is asymptotically equal to zero when \(a_k^t\) and \(w^k\) are the true mixing and separating vectors. This is achieved by dividing the model score function \(\phi_k^H\) by \(\nu_k^t\); \(\nu_k^t\) will be treated as a constant whose value is updated in every iteration. Hence, the modified gradient is
\[
\nabla w^k = \left< \hat{a}_{k,t} - \nu_k^t - \hat{E}[\phi_k^H x_k^H]\right>_t, \quad (16)
\]
and its partial derivative by \(w^k\) reads
\[
H_{Wk} = \frac{\partial \nabla w^k}{\partial w^k} = \left< -\nu_k^t \hat{E}[\phi_k^H x_k^H]\right>_t. \quad (17)
\]
QuickIVE-2 (QuickICE-2 for \(K = 1\)) starts from an initial value and iterates in two main steps: First, the separating vectors are updated in the NR directions given by \(\Delta w^k = -(H_{Wk})^{-1} \nabla w^k\). The second step is the same as step 3 in QuickIVE-1.\footnote{Interestingly, (15) is exactly equal to the “constrained” gradient derived in [17] for \(T = 1\) where \(a_k^t\) is treated as dependent variable on \(w^k\) through the orthogonal constraint (9).}

3.2. Quick IVA
Both previous algorithms can be used as building blocks of methods performing deflation or parallel separation of the observed data into independent vectors/components. Here, we propose the parallel variant, which, similarly to Symmetric FastICA [5], runs \(d\) (differently initialized) QuickIVE-1 or QuickIVE-2 algorithms. Each separating vector is updated by performing one iteration of the BSE method. Then, the updated vectors are orthogonalized through
\[
W^k = W^k ((W^k)^H W^k)^{-1/2}, \quad (18)
\]
where \(W^k\) denotes the \(d \times d\) de-mixing matrix whose rows correspond to the separating vectors. The process is repeated until convergence. The corresponding algorithms will be referred to as QuickIVA-1 and QuickIVA-2 and will be considered only for the standard IVA/ICA, i.e., when \(T = 1\). The problem of performing the complete independent vector/component analysis when \(T > 1\) goes beyond the scope of this paper.

4. SIMULATIONS

4.1. Blind Extraction
The first experiment is focused on the blind extraction problem. We compare QuickIVE-1 and QuickIVE-2 with the gradient-based methods OGICE/OGIVE [17] and with the state-of-the-art algorithms One-unit FastICA [5] and AuxICA/IVA [19]. Each method is initialized in the vicinity of the ideal solution (the true separating vector plus random vector whose norm is 0.1). To compare BSE and BSS methods, in AuxICA/IVA, the initial de-mixing matrix is such that its first row is equal to the initial separating vector and the other rows are orthogonal to the corresponding initial mixing vector. The output signal-to-interference ratio (SIR) of the extracted signal (of the first output channel in case of AuxICA/IVA) is evaluated after the given algorithm is stopped. The stopping criterion used by FastICA is used in all compared methods.

The data for one trial are simulated as follows: Three mixtures \((K = 3)\) of length \(N = 1000\) are generated, where each one consists of six independent signals drawn from the circular Laplacian distribution. The first signals in the \(K\) mixtures are mixed by a random \(K \times K\) unitary matrix, which makes them uncorrelated and dependent. Then, signals in each mixture are mixed by a random non-singular mixing matrix; the whole data thus obey the standard IVA and ICA \((T = 1)\) mixing model (1).

The variant of this experiment is considered also for \(T = 3\) with \(N_k = 1000\), i.e., \(N = 3000\). Here, the mixing matrices on the blocks are randomly generated such that the first three rows of their inverse matrices (true de-mixing matrices) are constant over the blocks. This means that there are three sources in the mixture obeying the CSV mixing model (the first one plays the role of the SOI). OGICE/OGIVE are replaced by their extended variants for CSV called BOGICE/BOGIVE [18], and QuickICA/IVA-2 is used for \(T = 3\) as QuickICA/IVA-1 is not suitable for \(T > 1\).

The algorithms are applied individually to each mixture in case of ICE/ICA and jointly to \(K\) mixtures in case of IVE/IVA. The non-linearities used are \(\phi_k^H(s) = s_k^H / \|s_k\|^2\) in OGICE/OGIVE, \(\phi_k^H(s) = s_k^H / (1 + \|s\|^2)\) in FastICA and QuickICA/IVA, and \(\phi_k^H(s) = 1 / \|s\|^2\) in AuxICA/IVA.

Histograms of the output SIR and the number of iterations resulting from 1000 trials are shown in Fig. 1. Each histogram of SIR has two or three main peaks where the rightmost peak (SIR > 15dB) corresponds to the successful extractions of the SOI. The top of this
peak corresponds to the mean SIR, i.e., to the mean extraction accuracy. This value is known to depend on the distribution of the sources and on the nonlinearity used.

The leftmost peak (SIR < $-15$ dB) corresponds to cases when different independent sources were extracted than the SOI. These cases need not correspond to algorithms’ failures, because they are mostly caused by the ambiguity of order that is inherent to BSE.

The third peak in SIR between $-15$ and $15$ dB points to the cases when the given algorithm got stuck in a saddle point or failed due to mixing model mismatch. The latter occurs in case of AuxICA/IVA and FastICA when $T = 3$ as these methods were designed only for the standard mixing models ($T = 1$).

For $T = 1$, the results show that QuickICE-1 and FastICA achieve comparable performance while FastICA is slightly faster and the both methods yield stable convergence. AuxICE got stuck in saddle points many times, which is cause by the fact that the non-linearity used in it is not suitable for $K = 1$ (see recommended nonlinearities in [9]); AuxIVA is much more stable as it works with $K = 3$ here. It is worth noting that the ICE/ICA algorithms extract the SOI in fewer cases than IVE/IVA. This shows that IVE/IVA can exploit the mutual dependence between the SOIs, which extends the area of convergence and helps in solving the permutation problem [21]. For $T = 3$, BOGICE/IVE and QuickICE/IVE-2 show stable performance.

The histograms of the number of iterations show that the gradient-based methods need a much higher number of iterations (often even more than 1000) to converge compared to the other methods. The proposed algorithms are comparable with the state-of-the-art methods in this respect. Most importantly, QuickIVE-2 shows fast and stable convergence also for the CSV mixing model.

4.2. Blind Separation

In the second experiment, we consider an IVA problem where QuickIVA-1 and QuickIVA-2 are compared with AuxIVA [11] and the corresponding Cramér-Rao-induced lower bound (CRIB) for Interference-to-Signal Ratio (ISR) [22]. In a trial, $K = 3$ random mixtures of $d = 5$ independent vector components are generated, where the components are drawn from the joint pdf given by

$$p(s^1, \ldots, s^K) \propto \exp \left( -\lambda^2 \sum_{i=1}^{K} |s_i|^2 \right)^{\alpha},$$

(19)

where $\lambda > 0$ set to scale data to unit variance, and $\alpha = 0.4$ (for $\alpha < 1$, the pdf is super-Gaussian). The length of data is $N = 5000$.

The methods are initialized by a random mixing matrix; the same nonlinearities are used as in the previous subsection. Fig. 2 shows the ISR averaged over the separated signals and 100 trials as it evolves depending on the computation time (on a PC with 2.6 GHz i7 processor). 50 iterations were done in a trial.

In this experiment, the proposed methods are less computationally demanding and converge faster than AuxIVA. Their final accuracy is slightly worse compared to that of AuxIVA. This limitation is caused by the used nonlinearity [23]; the nonlinearity used within AuxIVA here appears to be more suitable for the selected pdf.

5. CONCLUSIONS

The proposed algorithms provide useful alternatives to the well-known FastIVA/ICA and AuxIVA/ICA for the following reasons:

1. They yield comparably stable and fast performance,
2. no additional constraints on the form of the nonlinear function $\phi_k(\cdot)$ are imposed, unlike in AuxIVA or FastIVA,
3. QuickIVE-2 is efficient also for the CSV mixing model for $T > 1$. Future works will be focused on the stability and performance analysis of these methods in order to optimize their robustness and accuracy.
6. REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Independent Component Analysis and Applications Series. Elsevier Science, 2010.

[2] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Transactions on Audio, Speech, and Language Processing*, pp. 70–79, Jan. 2007.

[3] S. Amari, A. Cichocki, and H. H. Yang, “A new learning algorithm for blind signal separation,” in *Proceedings of Neural Information Processing Systems*, 1996, pp. 757–763.

[4] J. F. Cardoso and B. H. Laheld, “Equivariant adaptive source separation,” *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec 1996.

[5] A. Hyvärinen, “Fast and robust fixed-point algorithm for independent component analysis,” *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[6] E. Bingham and A. Hyvärinen, “A fast fixed-point algorithm for independent component analysis of complex valued signals,” *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, Feb. 2000.

[7] Mike Novey and T. Adali, “On extending the complex fastica algorithm to noncircular sources,” in *IEEE Trans. Signal Processing*, May 2008, pp. 2148–2154.

[8] Intae Lee, Taesu Kim, and Te-Won Lee, “Fast fixed-point independent vector analysis algorithms for convolutive blind source separation,” *Signal Processing*, vol. 87, no. 8, pp. 1859–1871, 2007.

[9] Nobutaka Ono and Shigeaki Miyabe, “Auxiliary-function-based independent component analysis for super-gaussian sources,” in *Latent Variable Analysis and Signal Separation*, Vincent Viognier, Vicente Zarzoso, Eric Moreau, Rémi Gribonval, and Emmanuel Vincent, Eds., Berlin, Heidelberg, 2010, pp. 165–172, Springer Berlin Heidelberg.

[10] Nobutaka Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2011, pp. 189–192.

[11] F. Nesta, P. Svaizer, and M. Omologo, “Convolutive bss of short mixtures by ICA recursively regularized across frequencies,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 3, pp. 624–639, March 2011.

[12] T. Taniguchi, N. Ono, A. Kawamura, and S. Sagayama, “An auxiliary-function approach to online independent vector analysis for real-time blind source separation,” in 2014 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA), May 2014, pp. 107–111.

[13] M. Sunohara, C. Haruta, and N. Ono, “Low-latency real-time blind source separation for hearing aids based on time-domain implementation of online independent vector analysis with truncation of non-causal components,” in 2017 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 216–220.

[14] J. A. Palmer, S. Makeig, K. Kreutz-Delgado, and B. D. Rao, “Newton method for the ica mixture model,” in 2008 *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2008, pp. 1805–1808.

[15] P. Ablin, J. Cardoso, and A. Gramfort, “Faster independent component analysis by preconditioning with hessian approximations,” *IEEE Transactions on Signal Processing*, vol. 66, no. 15, pp. 4040–4049, Aug 2018.

[16] S. Basiri, E. Ollila, and V. Koivunen, “Alternative derivation of fastica with novel power iteration algorithm,” *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1378–1382, Sep. 2017.

[17] Z. Koldovský and P. Tichavský, “Gradient algorithms for complex non-gaussian independent component/vector extraction, question of convergence,” *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1050–1064, Feb 2019.

[18] Z. Koldovský, J. Málek, and J. Janský, “Extraction of independent vector component from underdetermined mixtures through block-wise determined modeling,” in *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, May 2019, vol. 7903–7907.

[19] V. Kautský, Z. Koldovský, P. Tichavský, and V. Zarzoso, “Cramér-Rao bounds for complex-valued independent component extraction: Determined and piecewise determined mixing models,” *arXiv e-prints*, p. arXiv:1907.08790, Jul 2019.

[20] Hualiang Li and Tülay Adali, “Complex-valued adaptive signal processing using nonlinear functions,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, no. 1, pp. 765615, Feb 2008.

[21] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, Sept. 2004.

[22] M. Anderson, G. Fu, R. Phlypo, and T. Adali, “Independent vector analysis: Identification conditions and performance bounds,” *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, Sept 2014.

[23] P. Tichavský, Z. Koldovský, and E. Oja, “Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis,” *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1189–1203, April 2006.