Faster independent low-rank matrix analysis with pairwise updates of demixing vectors

Taishi Nakashima, Robin Scheibler, Yukoh Wakabayashi, Nobutaka Ono
Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, JAPAN
{nakashima-taishi@ed.,robin@,wakayuko@,onono@}tmu.ac.jp

***Abstract*—In this paper, we present an algorithm for independent low-rank matrix analysis (ILRMA) of three or more sources that is faster than that for conventional ILRMA. In conventional ILRMA, demixing vectors are updated one by one by the iterative projection (IP) method. The update rules of IP are derived from a system of quadratic equations obtained by differentiating the objective function of ILRMA with respect to demixing vectors. This system of quadratic equations is called hybrid exact-approximate joint diagonalization (HEAD) and no closed-form solution is known yet for three or more sources. Recently, a method that can update two demixing vectors simultaneously has been proposed for independent vector analysis. The method is derived by reducing HEAD for two sources to a generalized eigenvalue problem and solving the problem. Furthermore, the pairwise updates have recently been extended to the case of three or more sources. However, the efficacy of the pairwise updates for ILRMA has not yet been investigated. Therefore, in this work, we apply the pairwise updates of demixing vectors to ILRMA. By replacing the update rules of demixing vectors with the proposed pairwise updates, we accelerate the convergence of ILRMA. The experimental results show that the proposed method yields faster convergence and better performance than conventional ILRMA.**

***Index Terms*—Blind source separation, independent vector analysis, non-negative matrix factorization, independent low-rank matrix analysis**

I. Introduction

Blind source separation (BSS) is a technique of estimating the source signals from a mixture of sources using only the observed signals without any other information. For determined (number of microphones = number of sources) and overdetermined (number of microphones > number of sources) situations, independent component analysis (ICA) [1] is a fundamental technique. In ICA, a demixing system is estimated by assuming a non-Gaussian source distribution and the statistical independence of sources. For a convolutive mixture, frequency-domain ICA (FDICA) [2], [3] has been proposed. In FDICA, a demixing matrix is estimated at each frequency bin by applying ICA to a short-time Fourier transform (STFT) representation of the observed signals. To extend FDICA to a multivariate case, independent vector analysis (IVA) has been proposed [4], [5]. Furthermore, auxiliary-function-based ICA (AuxICA) [6] and auxiliary-function-based IVA (AuxIVA) [7], [8] have also been proposed. Traditionally, in IVA, demixing matrices are updated by the gradient-based method. The gradient-based update rules have tuning parameters such as step size and a trade-off between convergence speed and stability. By contrast, in AuxICA and AuxIVA, there are no tuning parameters, and the monotonic nonincrease in the objective function is theoretically guaranteed. Moreover, the auxiliary-function-based BSS techniques proved to be more stable and faster than the gradient-based methods.

As a state-of-the-art BSS method, independent low-rank matrix analysis (ILRMA) has recently been proposed [9]. In ILRMA, a low-rank matrix model of the spectrogram obtained from each source is assumed, and ILRMA can be interpreted as the method that unifies AuxIVA and multichannel non-negative matrix factorization (MNMF) [10]–[17]. In ILRMA, the demixing matrices are estimated by updating two parameters alternately; the demixing matrix and NMF parameters. ILRMA can achieve better performance than AuxIVA and is more stable than MNMF.

In most studies of ILRMA, the row vectors of the demixing matrix (demixing vectors) are updated one by one by the iterative projection (IP) method first proposed for AuxICA and AuxIVA [8]. This update rule is derived by solving systems of quadratic equations obtained by differentiating the cost function of ILRMA with respect to the demixing vectors. The systems of quadratic equations are called hybrid exact-approximate joint diagonalization (HEAD) [18], and no closed-form solution for three or more sources is yet known. For AuxIVA with two sources and two microphones, a faster algorithm has been proposed for updating the two demixing vectors simultaneously by solving a generalized eigenvalue problem [19], [20]. This pairwise update method is also applicable to pairs of demixing vectors in the case of three and more sources [21]. It has been shown to lead to faster convergence and better performance than methods using one-by-one updates.

On the basis of these studies, we propose a faster algorithm for ILRMA with pairwise updates of demixing vectors for three or more sources. We also investigate the difference in convergence speed between the IP and proposed methods by comparing the achieved values of the objective function. The experimental results showed that the proposed method outperformed conventional ILRMA in terms of convergence speed and the overall performance.

The rest of this paper is organized as follows. In Section II, we formulate the multichannel BSS problem and describe conventional ILRMA. In Section III, we present the proposed method and its algorithm. In Section IV, we show the experimental results and discuss a few implications of the proposed method. Finally, in Section V, we present our conclusions.

II. Background

A. Formulation

Let K and M be the numbers of sources and microphones, respectively. In this paper, henceforth, we consider the determined case, $K = M$. We define the STFT representations of the source, observed, and estimated signals, respectively as follows:

$$\boldsymbol{s}_{f\tau} = \begin{bmatrix} s_{1,f\tau} & \cdots & s_{k,f\tau} & \cdots & s_{K,f\tau} \end{bmatrix}^\top \in \mathbb{C}^{K\times 1}, \quad (1)$$
$$\boldsymbol{x}_{f\tau} = \begin{bmatrix} x_{1,f\tau} & \cdots & x_{k,f\tau} & \cdots & x_{K,f\tau} \end{bmatrix}^\top \in \mathbb{C}^{K\times 1}, \quad (2)$$
$$\boldsymbol{y}_{f\tau} = \begin{bmatrix} y_{1,f\tau} & \cdots & y_{k,f\tau} & \cdots & y_{K,f\tau} \end{bmatrix}^\top \in \mathbb{C}^{K\times 1}, \quad (3)$$

where $f \in \{1, \ldots, F\}$, $\tau \in \{1, \ldots, T\}$, and $k \in \{1, \ldots, K\}$ are the indices of frequency bins, time frames, and channels, respectively, and $^\top$ denotes the vector/matrix transpose. When the STFT window is sufficiently longer than the impulse response, we can represent the observed signal $\boldsymbol{x}_{f\tau}$ as $\boldsymbol{x}_{f\tau} = \boldsymbol{A}_f \boldsymbol{s}_{f\tau}$, where $\boldsymbol{A}_f \in \mathbb{C}^{K\times K}$ is a mixing matrix. If $\boldsymbol{A}_f$ is invertible, we can define the demixing matrix $\boldsymbol{W}_f = \begin{bmatrix} \boldsymbol{w}_{1,f} & \cdots & \boldsymbol{w}_{K,f} \end{bmatrix}^\mathsf{H} = \boldsymbol{A}_f^{-1}$, where $\boldsymbol{w}_{k,f} \in \mathbb{C}^{K\times 1}$ $(k = 1, \ldots, K)$ are the demixing vectors and $^\mathsf{H}$ denotes the Hermitian transpose. Therefore, the estimated signal $\boldsymbol{y}_{f\tau}$ can be represented as

$$\boldsymbol{y}_{f\tau} = \boldsymbol{W}_f \boldsymbol{x}_{f\tau}. \quad (4)$$

B. ILRMA

As introduced earlier, ILRMA is a determined BSS technique unifying IVA and NMF. In ILRMA, the demixing matrices are updated under the assumption that the complex spectrogram from the kth estimated signals is represented as the product of two non-negative matrices, $\boldsymbol{B}_k \in \mathbb{R}_+^{F\times L}$ and $\boldsymbol{H}_k \in \mathbb{R}_+^{L\times T}$ where $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ are the basis and activation matrices, respectively, and $\mathbb{R}_+$ denotes the set of non-negative real numbers. To estimate source model parameters $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$, we assume the source model to be a complex Gaussian distribution. In other words, the generative model of the kth source and its variance $r_{k,f\tau}$ are

$$p(\bar{\boldsymbol{y}}_{1,\tau}, \ldots, \bar{\boldsymbol{y}}_{k,\tau}) = \prod_{k,f} \frac{1}{\pi r_{k,f\tau}} \exp\left(-\frac{|y_{k,f\tau}|^2}{r_{k,f\tau}}\right), \quad (5)$$
$$r_{k,f\tau} = \sum_{l} b_{k,f\ell}\, h_{k,\ell\tau}, \quad (6)$$

where $\bar{\boldsymbol{y}}_{k,\tau}$ is the estimated vector that cosists of all frequency bins defined as $\bar{\boldsymbol{y}}_{k,\tau} = \begin{bmatrix} y_{k,1\tau} & \cdots & y_{k,F\tau} \end{bmatrix}^\top$, $b_{k,f\ell} \in \mathbb{R}_+$ and $h_{k,\ell\tau} \in \mathbb{R}_+$ are the (f,ℓ)th element of $\boldsymbol{B}_k$ and (ℓ,τ)th element of $\boldsymbol{H}_k$, respectively, and $\ell \in \{1, \ldots, L\}$ denotes the index of the bases. Fig. 1 illustrates the overview of ILRMA.

Next, by using the demixing model (4) and the source model (5), and calculating the negative log-likelihood of the observed signals, we obtain the objective function of ILRMA as follows:

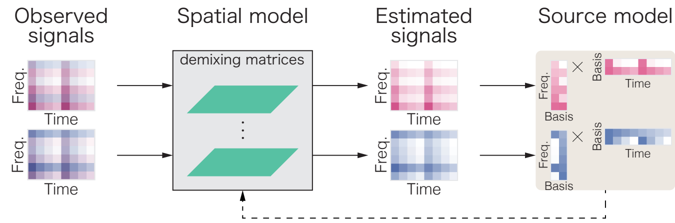


Fig. 1: Overview of source separation in ILRMA (e.g., $K = 2$).

$$\mathcal{J}(\mathsf{W}, \mathsf{B}, \mathsf{H}) = \sum_{k,f,\tau} \left[\frac{|\boldsymbol{w}_{k,f}^\mathsf{H} \boldsymbol{x}_{f\tau}|^2}{r_{k,f\tau}} + \log r_{k,f\tau} \right] - 2T \sum_f \log|\det \boldsymbol{W}_f|, \quad (7)$$

where W, B, and H are the tensors composed of all $\boldsymbol{W}_f$, $\boldsymbol{B}_k$, and $\boldsymbol{H}_k$, respectively. In this paper, the constant terms are omitted.

1) Update of the spatial model: To minimize the objective function (7) with respect to the demixing matrix $\boldsymbol{W}_f$, we obtain the following function $\mathcal{Q}$ by extracting terms that are dependent on W from (7):

$$\mathcal{Q}(\mathsf{W}, \mathsf{U}) = \sum_{k,f} \boldsymbol{w}_{k,f}^\mathsf{H} \boldsymbol{U}_{k,f} \boldsymbol{w}_{k,f} - \sum_f \log|\det \boldsymbol{W}_f|, \quad (8)$$
$$\boldsymbol{U}_{k,f} = \frac{1}{T} \sum_\tau \frac{1}{2 r_{k,f\tau}} \boldsymbol{x}_{f\tau} \boldsymbol{x}_{f\tau}^\mathsf{H}, \quad (9)$$

where $\boldsymbol{U}_{k,f} \in \mathbb{C}^{K\times K}$ and U are the covariance matrix and the tensor composed of all $\boldsymbol{U}_{k,f}$, respectively. Henceforth, we omit the frequency bin index f for simplicity.

By calculating $\partial\mathcal{Q}/\partial\boldsymbol{w}_k = 0$ $(k = 1, \ldots, K)$ and rearranging it, we can obtain the following system of quadratic equations:

$$\boldsymbol{w}_\ell^\mathsf{H} \boldsymbol{U}_k \boldsymbol{w}_k = \delta_{\ell k} \quad (k, \ell = 1, \ldots, K), \quad (10)$$

where $\delta_{\ell k}$ is the Kronecker delta. The problem in (10) is HEAD [18], and no closed-form solution for $K \geq 3$ is known yet [21]. Instead of solving HEAD directly, let us consider minimizing (8) with respect to only one demixing vector $\boldsymbol{w}_m$ while keeping the other $\boldsymbol{w}_k$ $(k \neq m)$ fixed. In this case, the problem can be solved as follows [8]:

$$\boldsymbol{w}_m \leftarrow (\boldsymbol{W}\boldsymbol{U}_m)^{-1} \boldsymbol{e}_m, \quad (11)$$
$$\boldsymbol{w}_m \leftarrow \boldsymbol{w}_m (\boldsymbol{w}_m^\mathsf{H} \boldsymbol{U}_m \boldsymbol{w}_m)^{-\frac{1}{2}}, \quad (12)$$

where $\boldsymbol{e}_m \in \mathbb{R}^K$ denotes the canonical basis vector with the mth element unity. This method is called IP [9]. In this paper, we refer to this IP method as IP-1 to distinguish it from the proposed method described in Section III.

2) Update of the source model: By applying the auxiliary function method [7] to (7), we can obtain multiplicative update rules for the source model parameters $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ [9], [14]:

$$b_{k,f\ell} \leftarrow b_{k,f\ell}\left[\frac{\sum_\tau |y_{k,f\tau}|^2 h_{k,\ell\tau}(\sum_{\ell'} b_{k,f\ell'}h_{k,\ell'\tau})^{-2}}{\sum_\tau h_{k,\ell\tau}(\sum_{\ell'} b_{k,f\ell'}h_{k,\ell'\tau})^{-1}}\right]^{\frac{1}{2}}, \tag{13}$$

$$h_{k,\ell\tau} \leftarrow h_{k,\ell\tau}\left[\frac{\sum_f |y_{k,f\tau}|^2 b_{k,f\ell}(\sum_{\ell'} b_{k,f\ell'}h_{k,\ell'\tau})^{-2}}{\sum_f b_{k,f\ell}(\sum_{\ell'} b_{k,f\ell'}h_{k,\ell'\tau})^{-1}}\right]^{\frac{1}{2}}. \tag{14}$$

III. Pairwise demixing vector updates

A. Pairwise update rules for two sources

In the case of $K = 2$, a closed-form solution of HEAD exists; namely, the two demixing vectors $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are the solutions of the following generalized eigenvalue problem:

$$\boldsymbol{U}_2\boldsymbol{u}_k = \lambda_k \boldsymbol{U}_1\boldsymbol{u}_k \quad (k = 1, 2), \tag{15}$$

where λ_1 and λ_2 are the eigenvalues such that $\lambda_1 \geq \lambda_2$ corresponding to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, respectively. By using these $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ with appropriate normalization as written in (21) below, we can simultaneously update the demixing vectors [19], [22].

Recently, this pairwise update method has been extended to the situation where $K \geq 3$ for AuxIVA [21]. We aim to apply this pairwise update method to ILRMA to accelerate the update of demixing vectors.

B. Pairwise update rules for three or more sources

First, we give an outline of the method proposed in [21]. In the case of $K \geq 3$, let us consider how to update two demixing vectors $\boldsymbol{w}_m$ and $\boldsymbol{w}_n$ $(m \neq n)$ simultaneously while keeping the other $(K-2)$ demixing vectors $\boldsymbol{w}_k$ $(k \neq m, n)$ fixed. By calculating $\partial\mathcal{Q}/\partial\boldsymbol{w}_m = 0$ and $\partial\mathcal{Q}/\partial\boldsymbol{w}_n = 0$ $(m \neq n)$, we can obtain the following systems of $2K$ quadratic equations:

$$\boldsymbol{w}_m^{\mathsf{H}}\boldsymbol{U}_m\boldsymbol{w}_m = 1, \quad \boldsymbol{w}_m^{\mathsf{H}}\boldsymbol{U}_n\boldsymbol{w}_n = 0, \tag{16}$$

$$\boldsymbol{w}_n^{\mathsf{H}}\boldsymbol{U}_m\boldsymbol{w}_m = 0, \quad \boldsymbol{w}_n^{\mathsf{H}}\boldsymbol{U}_n\boldsymbol{w}_n = 1, \tag{17}$$

$$\boldsymbol{w}_{k'}^{\mathsf{H}}\boldsymbol{U}_m\boldsymbol{w}_m = 0, \quad \boldsymbol{w}_{k'}^{\mathsf{H}}\boldsymbol{U}_n\boldsymbol{w}_n = 0 \quad (\forall k' \neq m, n). \tag{18}$$

The pairwise update rule is derived by solving (16)–(18) as follows [21]:

$$\boldsymbol{P}_m \leftarrow (\tilde{\boldsymbol{W}}\boldsymbol{U}_m)^{-1}\boldsymbol{E}_{mn} \qquad \boldsymbol{P}_n \leftarrow (\tilde{\boldsymbol{W}}\boldsymbol{U}_n)^{-1}\boldsymbol{E}_{mn}, \tag{19}$$

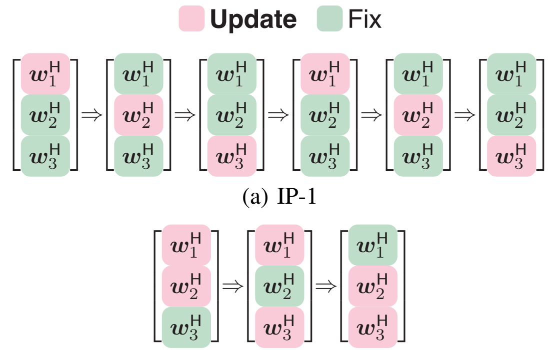
$$\boldsymbol{V}_m \leftarrow \boldsymbol{P}_m^{\mathsf{H}}\boldsymbol{U}_m\boldsymbol{P}_m \qquad \boldsymbol{V}_n \leftarrow \boldsymbol{P}_n^{\mathsf{H}}\boldsymbol{U}_n\boldsymbol{P}_n, \tag{20}$$

$$\boldsymbol{v}_m \leftarrow \boldsymbol{v}_m(\boldsymbol{v}_m^{\mathsf{H}}\boldsymbol{V}_m\boldsymbol{v}_m)^{-\frac{1}{2}} \qquad \boldsymbol{v}_n \leftarrow \boldsymbol{v}_n(\boldsymbol{v}_n^{\mathsf{H}}\boldsymbol{V}_n\boldsymbol{v}_n)^{-\frac{1}{2}}, \tag{21}$$

$$\boldsymbol{w}_m \leftarrow \boldsymbol{P}_m\boldsymbol{v}_m \qquad \boldsymbol{w}_n \leftarrow \boldsymbol{P}_n\boldsymbol{v}_n, \tag{22}$$

where $\tilde{\boldsymbol{W}}$ is the demixing matrix obtained in the previous update, $\boldsymbol{E}_{mn}$ is a $K \times 2$ matrix defined as $\boldsymbol{E}_{mn} = \begin{bmatrix}\boldsymbol{e}_m & \boldsymbol{e}_n\end{bmatrix}$, and $\boldsymbol{v}_m$ and $\boldsymbol{v}_n$ are the eigenvectors of $\boldsymbol{V}_n^{-1}\boldsymbol{V}_m$ corresponding to larger and smaller eigenvalues, respectively. We refer to this update as *IP-2*.

Fig. 2 shows an overview of the difference between IP-1 and IP-2. We can interpret IP-2 as the process by which two steps are updated in IP-1 simultaneously.



(a) IP-1

(b) IP-2

Fig. 2: Illustration of IP-1 and IP-2 when $K = 3$.

Algorithm 1: ILRMA using IP-2

Initialize $\boldsymbol{W}_f$ $\forall f \in \{1, \ldots, F\}$
forall $k = 1, \ldots, K, K+1, \ldots, 2K$ **do**
  $m \leftarrow k \mod K$
  $n \leftarrow (k+1) \mod K$
  Update source model parameters $\boldsymbol{B}_m$, $\boldsymbol{B}_n$ and $\boldsymbol{H}_m$, $\boldsymbol{H}_n$ using (13) and (14).
  Update source models $r_{m,f\tau}$ and $r_{n,f\tau}$ using (6).
  Update covariance matrices $\boldsymbol{U}_m$ and $\boldsymbol{U}_n$ using (9).
  Update two demixing vectors $\boldsymbol{w}_m$ and $\boldsymbol{w}_n$ using (19)–(22).

Therefore, we can obtain a new ILRMA by replacing the conventional update rule of demixing vectors by this method, as shown in Algorithm 1. Henceforth, we refer to the new ILRMA as *FasterILRMA*. Note that the choice of m and n in the proposed method is arbitrary as long as $m \neq n$.

IV. Experiments

To confirm the efficacy of the proposed method, we first compared the convergence of the objective function when using the conventional method with that for the proposed method. Next, we conducted a BSS experiment and evaluated the separation performance.

A. Experimental setup

We conducted experiments for music and speech signals. For the music signals, we used the DSD100 dataset [23], which includes 100 stereo recorded signals that consist of four parts: `bass`, `drums`, `other`, and `vocals`. We excluded `other` because they might include multiple instruments. To obtain monaural sources, we extracted the left channel of `bass`, `drums`, and `vocals`. The original sampling rate was 44.1 kHz, but we downsampled the signals to 16 kHz.

For the speech signals, we used the mixture speech of four speakers obtained from the Japanese Newspaper Article Sentences (JNAS) dataset [24]. The sampling rate was 16 kHz.

We used the `pyroomacoustics` Python package [25] to simulate a rectangular room and create convolutive mixtures. Fig. 3 shows the room used for the simulation and the locations

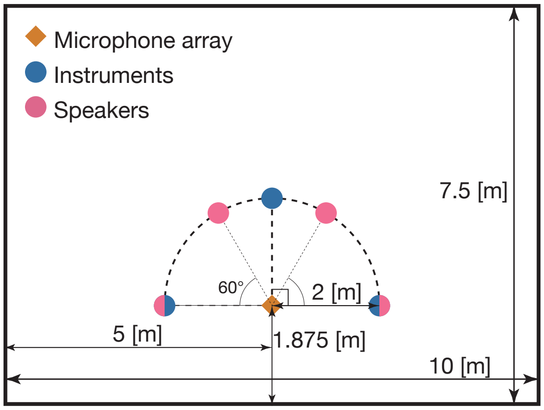


Fig. 3: Setup of experiment.

TABLE I: Parameters of the simulation.

Number of bases L	2, 5, 20
Length of FFT and window function	4096 samples
Frame shift of STFT	2048 samples
Initial demixing matrices $\boldsymbol{W}_f$	Identity matrix
Number of iteration steps	100
STFT window function	Hamming

of sources and the microphone array. The reverberation time was approximately 200 ms. The number of sound sources and microphones was three for the music signals and four for the speech signals, respectively. The microphone array was uniformly linear, with a spacing of 2.83 cm. Moreover, for numerical stability, we set the initial values of the source models $\boldsymbol{B}_k$ and $\boldsymbol{H}_k$ to $0.9\boldsymbol{Z} + 0.1\boldsymbol{I}$ where $\boldsymbol{Z}$ and $\boldsymbol{I}$ are the matrix of values uniformly distributed over $[0, 1)$ and the matrix of ones, respectively. Table I shows the rest of the parameters common in both types of signal.

B. Comparison of convergence of objective function

We compared the convergence of the objective function using the conventional ILRMA with that using the FasterILRMA. Fig. 4 shows an example of the evolution of the objective function for one of the music signals. As shown in Fig. 4, the objective function decreased monotonically in both methods, but the proposed method converged much faster and decreased the objective function to a greater extent than the conventional method. This result implies that the two respective methods may converge to a different local minimum of the objective function of ILRMA. This will be investigated in detail in our future work.

C. Separation performance

To avoid bias from a specific arrangement of sources, we performed experiments for all $K!$ permutations of sources in each mixture. Then, we calculated the scale-invariant signal-to-distortion ratio (SI-SDR) improvements [26] for all sources in all mixtures and averaged them. We compared the two methods: conventional ILRMA [9] and FasterILRMA. The results for the music and speech signals are shown in Figs. 5 and 6, respectively. For music signals, FasterILRMA achieved

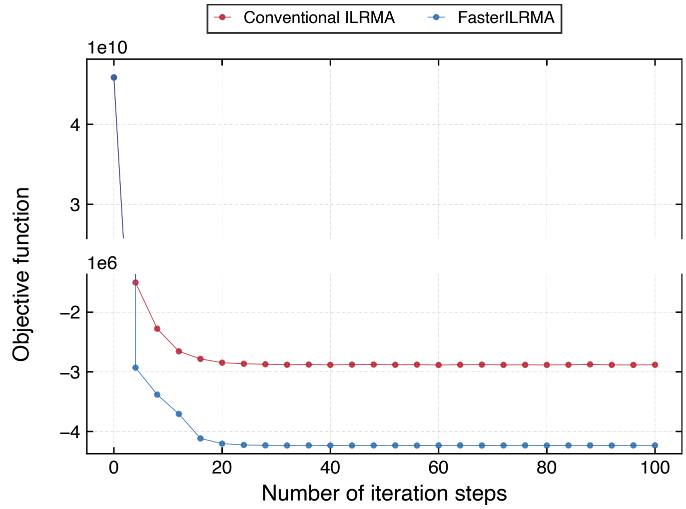


Fig. 4: Values of objective function with conventional and proposed methods.

significantly better performance and faster convergence than conventional ILRMA for all Ls, as shown in Fig. 5. For speech signals, FasterILRMA also outperformed conventional ILRMA for most Ls, as shown in Fig. 6. In particular, FasterILRMA achieved the best performance for $L = 2$, whereas the SI-SDR of conventional ILRMA decreased as L increased. This result is consistent with that described in [9].

An existing work [27] reported that the separation performance of ILRMA tends to be improved by slowing down the update of the source models. We consider that the proposed method improved the separation performance in a similar way because we accelerated the updates of the spatial models in this study, which made the update of the source models slower relatively.

V. Conclusion

We proposed a new algorithm for ILRMA to accelerate the convergence for three or more sources by updating two demixing vectors simultaneously. The experimental evaluation of BSS showed that the proposed FasterILRMA achieved better performance and faster convergence than conventional ILRMA. This result implied that these two methods might converge to a different local minimum of the objective function of ILRMA. In our future work, we will further investigate why the pairwise update of the demixing vectors (IP-2) yields better performance than the sequential update of the demixing vector (IP-1). We will also apply the proposed method to other BSS methods that use IP, such as ILRMA with generalized Kullback–Leibler divergence [28] and independent deeply learned matrix analysis [29].

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP16H01735, JP17H00749, and JST CREST Grant Number JPMJCR19A3 Japan.

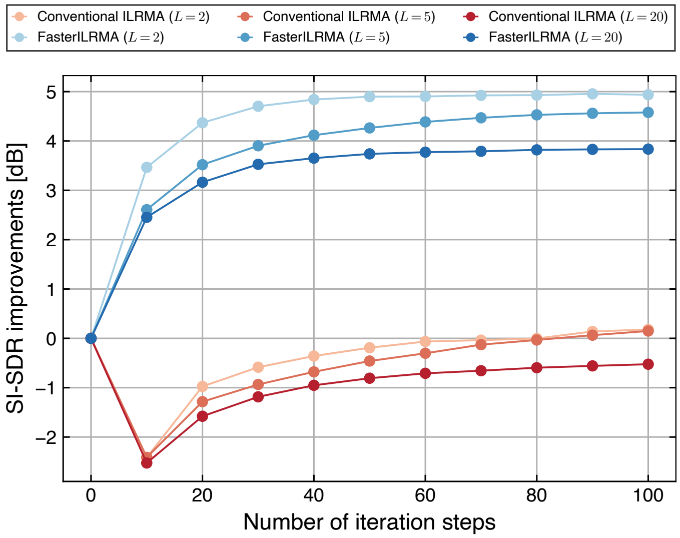


Fig. 5: Average of SI-SDR improvements with DSD100 signals for $L = 2, 5,$ and 20.

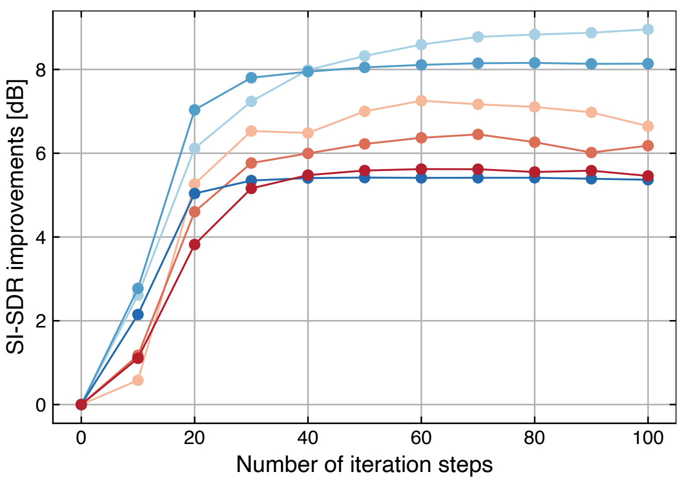


Fig. 6: Average of SI-SDR improvements with speech signals for $L = 2, 5,$ and 20.

References

[1] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1, pp. 21–34, 1998.

[3] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 666–678, 2006.

[4] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. ICA*, 2006, pp. 601–608.

[5] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70–79, 2006.

[6] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," in *Proc. LVA/ICA*, 2010, pp. 165–172.

[7] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[8] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, 2011, pp. 189–192.

[9] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1622–1637, 2016.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[11] ——, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*. MIT Press, 2001, pp. 556–562.

[12] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[13] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proc. ICASSP*, Apr. 2009, pp. 3437–3440.

[14] C. Févotte, N. Bertin, and J. Durrieu, "Nonnegative matrix factorization with the Itakura–Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[15] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, Mar. 2010.

[16] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Under-determined reverberant audio source separation using a full-rank spatial covariance model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1830–1840, Sep. 2010.

[17] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 971–982, May 2013.

[18] A. Yeredor, "On hybrid exact-approximate joint diagonalization," in *Proc. CAMSAP*, 2009, pp. 312–315.

[19] N. Ono, "Fast stereo independent vector analysis and its implementation on mobile phone," in *Proc. IWAENC*, Sep. 2012.

[20] ——, "Blind source separation on iPhone in real environment," in *Proc. EUSIPCO*, Sep. 2013.

[21] ——, "Fast algorithm for independent component/vector/low-rank matrix analysis with three or more sources," in *Proc. 2018 Spring Meeting of Acoustical Society of Japan*, Mar. 2018 (in Japanese), pp. 437–438.

[22] T. Yoshioka, T. Nakatani, and M. Miyoshi, "An integrated method for blind separation and dereverberation of convolutive audio mixtures," in *Proc. EUSIPCO*, Aug. 2008.

[23] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Proc. LVA/ICA*, Aug. 2017, pp. 323–332.

[24] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *Journal of the Acoustical Society of Japan E*, vol. 20, no. 3, pp. 199–206, May 1999.

[25] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulation and array processing algorithms," in *Proc. ICASSP*, Apr. 2018, pp. 351–355.

[26] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "SDR — half-baked or well done?" in *Proc. ICASSP*, May 2019, pp. 626–630.

[27] Y. Mitsui, D. Kitamura, N. Takamune, H. Saruwatari, Y. Takahashi, and K. Kondo, "Independent low-rank matrix analysis based on parametric majorization-equalization algorithm," in *Proc. CAMSAP*, Dec. 2017.

[28] S. Mogami, Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, H. Nakajima, and H. Kameoka, "Independent low-rank matrix analysis based on generalized Kullback–Leibler divergence," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E102-A, no. 2, pp. 458–463, 2019.

[29] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari, and N. Ono, "Independent deeply learned matrix analysis for determined audio source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 10, pp. 1601–1615, Oct. 2019.