PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Column-wise update algorithm for independent deeply learned matrix analysis

Naoki MAKISHIMA(1), Norihiro TAKAMUNE(1), Daichi KITAMURA(2), Hiroshi SARUWATARI(1), Yu TAKAHASHI(3), Kazunobu KONDO(3)

(1)The University of Tokyo, Japan
(2)National Institute of Technology, Kagawa College, Japan
(3)Yamaha Corporation, Japan

Abstract

In this paper, we propose a robust demixing filter update algorithm for audio source separation. Audio source separation is a task to recover source signals from multichannel mixtures observed in a microphone array, which can be applied to, e.g., speech recognition and music signal analysis. Recently, independent deeply learned matrix analysis (IDLMA) has been proposed as a state-of-the-art separation method. IDLMA utilizes deep neural network (DNN) inference of source models and blind estimation of demixing filters based on sources' independence. In conventional IDLMA, iterative projection (IP) is exploited to estimate the demixing filters. Although IP is a fast algorithm, when the specific source model is not accurate owing to the bad SNR condition, the successive update of filters will fail hereafter. This is because IP updates the demixing filters in a source-wise manner where only one source model is used for each update. In this paper, we derive a new microphone-wise update rule which exploits all information of the source models simultaneously for each update. Moreover, we propose a method to select the appropriate source- or microphone-wise update rule depending on the source signal's pseudo SNR estimated via the DNNs. Experimental results show the efficacy of the proposed method.

Keywords: Audio source separation, Deep neural network

1 INTRODUCTION

Audio source separation aims to recover source signals from the multichannel mixtures observed by a microphone array. Many types of algorithm have been proposed, e.g., unsupervised (blind) methods [1–17] and supervised (informed) methods [18–22]. *Independent deeply learned matrix analysis* (IDLMA) [23, 24] is a state-of-the-art separation method combining blind estimation of the demixing matrix and deep neural network (DNN) inference of source models. In the conventional source-separation methods including IDLMA, a fast and stable coordinate descent algorithm called iterative projection (IP) [5, 25] is exploited to estimate the demixing matrix. IP updates the demixing matrix in a source-wise (row-wise) manner where only one source model is used for each update. Although IP is a fast algorithm, our preliminary experiments show that it sometimes fails to find a good solution. This is because when the specific source model is not accurate owing to the bad SNR condition, the successive update of filters will fail hereafter. In this paper, we propose a new microphone-wise (column-wise) update rule of the demixing matrix, which exploits all information of the source models simultaneously for each update. Since IP and the proposed column-wise update rule are based on the coordinate-descent-type iterative algorithm, the update order exists and it is arbitrary. In general, different update rule or update order leads to a different solution because of local minima in the cost function, and it is difficult to choose the most appropriate update strategy (here the "strategy" includes a type of row/column-wise update rule, update order, and how to switch them in the iterative optimization). In order to find the appropriate update strategy, we propose an automatic selection method for the update strategy depending on the source signal's pseudo SNR estimated via the DNNs.

2 CONVENTIONAL METHODS

2.1 Formulation

Let M and N be the numbers of microphones and sound sources, respectively. We assume a determined case where $M = N$. The short-time Fourier transform (STFT) of the observed mixtures, estimated signals, source signals are defined as

$$\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijN})^{\mathrm{T}}, \tag{1}$$

$$\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijN})^{\mathrm{T}}, \tag{2}$$

$$\boldsymbol{s}_{ij} = (s_{ij1}, \ldots, s_{ijN})^{\mathrm{T}}, \tag{3}$$

where i and j denote the indexes of frequency bins and time frames, respectively, and $^{\mathrm{T}}$ denotes the transpose. In the determined case, the estimated signals $\boldsymbol{y}_{ij}$ can be represented as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \tag{4}$$

where $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iN})^{\mathrm{H}}$ is the demixing matrix and $\boldsymbol{w}_{in}^{\mathrm{H}}$ is the demixing filter for the nth source, and $^{\mathrm{H}}$ denotes the Hermitian transpose.

2.2 Generative Model and Cost Function

In IDLMA, the following univariate complex Gaussian distribution is assumed as a source generative model:

$$\prod_{i,j} p(y_{ijn}) = \prod_{i,j} \frac{1}{\pi {r_{ijn}}^2} \exp\left(-\frac{|y_{ijn}|^2}{{r_{ijn}}^2} \right), \tag{5}$$

where r_{ijn} is the scale parameter of a Gaussian distribution. Under the model (5), the cost function to be minimized w.r.t. $\boldsymbol{W}_i$ and r_{ijn} (negative log-likelihood of $\boldsymbol{x}_{ij} = \boldsymbol{W}_i^{-1} \boldsymbol{y}_{ij}$) is obtained as

$$\mathscr{L} \stackrel{c}{=} \sum_i \left(\sum_n \boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{Q}_{in} \boldsymbol{w}_{in} - \log |\det \boldsymbol{W}_i|^2 \right) + \frac{1}{J} \sum_{i,j,n} \log r_{ijn}^2, \tag{6}$$

$$\boldsymbol{Q}_{in} = \frac{1}{J} \sum_j \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{H}}}{{r_{ijn}}^2}. \tag{7}$$

We denote the scale parameter matrix as $\boldsymbol{R}_n \in \mathbb{R}_{\geq 0}^{I \times J}$, whose elements are r_{ijn}. The aim of IDLMA is to blindly estimate $\boldsymbol{W}_i$ only from the observed mixtures with the assistance of a DNN. The overview of separation process of IDLMA is shown in Fig. 1.

2.3 Update Rule of Demixing Matrix by IP

In [5, 25], a fast and convergence-guaranteed algorithm called IP was proposed, which can be applied to the sum of a negative log-determinant and a quadratic form. By applying IP to the first and second terms of the right side in (6), the update rule of demixing matrix $\boldsymbol{W}_i$ is obtained as follows:

$$\boldsymbol{w}_{in} \leftarrow (\boldsymbol{W}_i \boldsymbol{Q}_{in})^{-1} \boldsymbol{e}_n, \tag{8}$$

$$\boldsymbol{w}_{in} \leftarrow \frac{\boldsymbol{w}_{in}}{\sqrt{\boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{Q}_{in} \boldsymbol{w}_{in}}}, \tag{9}$$

where $\boldsymbol{e}_n$ denotes the unit vector with the nth element equal to unity. Note that since the update order w.r.t. n is arbitrary, we have to select the appropriate update order from $N!$ options.

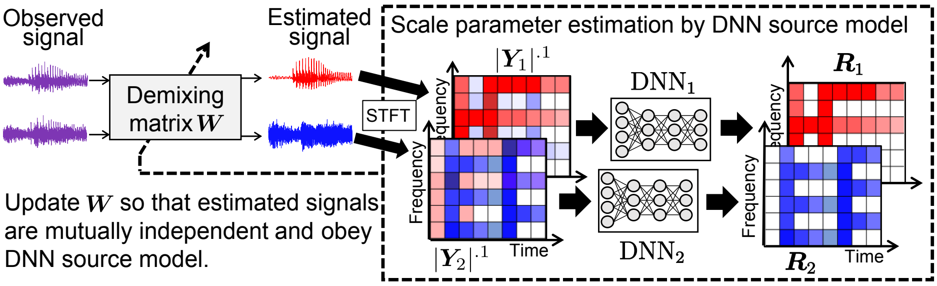

Figure 1. Overview of separation process in IDLMA.

2.4 Update Rule of Scale Parameter Matrix by DNN

DNN$_n$ is pre-trained so that the scale parameter $|\boldsymbol{R}_n|^{\cdot 1}$ is predicted from an input mixture spectrogram $|\tilde{\boldsymbol{X}}|^{\cdot 1}$, where $\tilde{\boldsymbol{X}} \in \mathbb{C}^{I\times J}$ is a mixture spectrogram in the training data and $|\cdot|^{\cdot p}$ for matrices denotes the element-wise absolute and pth-power operations. In inference for open data, the scale parameter matrix is updated by the pre-trained DNN$_n$ as

$$|\boldsymbol{R}_n|^{\cdot 1} \leftarrow \mathrm{DNN}_n(|\boldsymbol{Y}_n|^{\cdot 1}), \tag{10}$$

$$r_{ijn} \leftarrow \max(r_{ijn}, \varepsilon), \tag{11}$$

where $\boldsymbol{Y}_n \in \mathbb{C}^{I\times J}$ is the spectrogram of the estimated signal whose elements are y_{ijn}, temporally obtained through the update of $\boldsymbol{W}_i$ and ε is a small value to increase the numerical stability in IP.

When we define the output scale parameter matrix as $\boldsymbol{D}_n = \mathrm{DNN}_n(|\tilde{\boldsymbol{X}}|^{\cdot 1}) \approx \boldsymbol{R}_n$ and source spectrogram as $\tilde{\boldsymbol{S}}_n \in \mathbb{C}^{I\times J}$, the loss function of DNN$_n$ in (10) can be defined as

$$\mathrm{L}(\boldsymbol{D}_n) = \sum_{i,j} \frac{|\tilde{s}_{ijn}|^2+\delta_1}{d_{ijn}{}^2+\delta_1} - \log\frac{|\tilde{s}_{ijn}|^2+\delta_1}{d_{ijn}{}^2+\delta_1} - 1, \tag{12}$$

where $\tilde{s}_{ijn}$ and d_{ijn} are the elements of $\tilde{\boldsymbol{S}}_n$ and $\boldsymbol{D}_n$, respectively, and δ_1 is a small value to avoid division by zero [20]. Since minimizing (12) corresponds to a simulation for the ML estimation of r_{ijn} in (6) (only limited to the training data), DNN$_n$ can be approximately interpreted as an appropriate source model based on (5). In making the mixture spectrogram $|\tilde{\boldsymbol{X}}|^{\cdot 1}$, source spectrograms are mixed with various SNR as described in [23].

3 PROPOSED METHODS

3.1 Proposed Update Rule of Demixing Matrix

In [23], demixing matrix $\boldsymbol{W}_i$ is updated by (8) and (9). IP updates $\boldsymbol{w}_{in}$ based on given source model $\boldsymbol{R}_n$ and $\boldsymbol{w}_{in'}$ ($n' \neq n$) for each n, which corresponds to a source-wise update of $\boldsymbol{W}_i$. Therefore, updating $\boldsymbol{w}_{in}$ with accurate $\boldsymbol{R}_n$ first leads to a good solution and the update order w.r.t. n is important.

In this paper, we propose a column-wise update rule, which corresponds to a microphone-wise update of $\boldsymbol{W}_i$. Figure 2 illustrates a difference between the conventional IP and the proposed column-wise update algorithm.

We denote the column vector $\tilde{\boldsymbol{w}}_{im}$ of $\boldsymbol{W}_i$ as $\boldsymbol{W}_i = (\tilde{\boldsymbol{w}}_{i1}, \ldots, \tilde{\boldsymbol{w}}_{iN})$, where $\tilde{\boldsymbol{w}}_{im}$ is a microphone-wise vector although $\boldsymbol{w}_{in}$ is a source-wise (row) vector. We pick out the partial cost function $\mathscr{J}$ w.r.t. demixing matrix $\boldsymbol{W}_i$ from (6), which is rewritten as follows w.r.t. $\tilde{\boldsymbol{w}}_{im}$:

$$\begin{aligned}\mathscr{J} &= \sum_i \left(\sum_n \boldsymbol{w}_{in}^{\mathrm{H}}\boldsymbol{Q}_{in}\boldsymbol{w}_{in} - \log|\det\boldsymbol{W}_i|^2\right)\\ &\stackrel{c}{=} \sum_i \left(\tilde{\boldsymbol{w}}_{im}^{\mathrm{H}}\tilde{\boldsymbol{Q}}_{im}\tilde{\boldsymbol{w}}_{im} + \tilde{\boldsymbol{w}}_{im}^{\mathrm{H}}\tilde{\boldsymbol{h}}_{im} + \tilde{\boldsymbol{h}}_{im}^{\mathrm{H}}\tilde{\boldsymbol{w}}_{im} - \log|\boldsymbol{b}_{im}^{\mathrm{H}}\tilde{\boldsymbol{w}}_{im}|^2\right),\end{aligned} \tag{13}$$

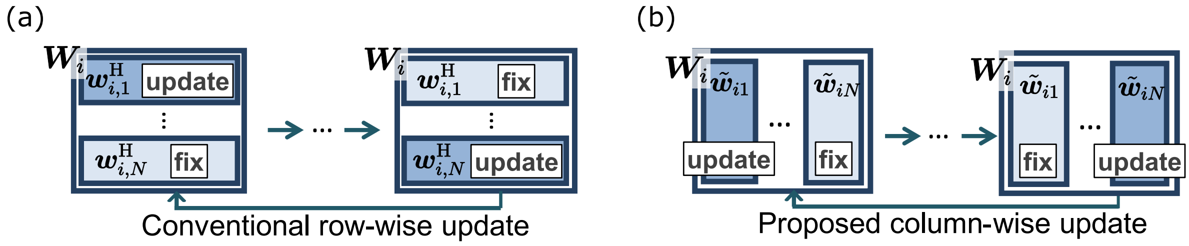

Figure 2. (a) Conventional row-wise update of $\boldsymbol{W}_i$. (b) Proposed column-wise update of $\boldsymbol{W}_i$.

where

$$\tilde{\boldsymbol{Q}}_{im} = \mathrm{diag}(Q_{i1mm}, \ldots, Q_{iNmm}), \tag{14}$$

$$\tilde{\boldsymbol{h}}_{im} = \left(\sum_{m' \neq m} w^*_{i1m'} Q_{i1m'm}, \ldots, \sum_{m' \neq m} w^*_{iNm'} Q_{iNm'm} \right)^{\mathrm{T}}. \tag{15}$$

Here $Q_{inm'm}$ is the (m', m)th element of $\boldsymbol{Q}_{in}$, $w_{inm'}$ is the m'th element of $\boldsymbol{w}_{in}$, $\mathrm{diag}(Q_{i1mm}, \ldots, Q_{iNmm})$ is the diagonal matrix whose elements are $Q_{i1mm}, \ldots, Q_{iNmm}$, $\boldsymbol{b}^{\mathrm{H}}_{im}$ is the mth row vector of the adjugate matrix of $\boldsymbol{W}_i$, and * denotes the complex conjugate. Since IP can not minimize the cost function that includes the linear form, we derive a new coordinate descent algorithm for (13). The derivative of (13) is obtained as

$$\frac{\partial \mathscr{J}}{\partial \tilde{\boldsymbol{w}}^*_{im}} = \tilde{\boldsymbol{Q}}_{im} \tilde{\boldsymbol{w}}_{im} + \tilde{\boldsymbol{h}}_{im} - \frac{\boldsymbol{b}_{im}}{\tilde{\boldsymbol{w}}^{\mathrm{H}}_{im} \boldsymbol{b}_{im}}. \tag{16}$$

From $\partial \mathscr{J} / \partial \tilde{\boldsymbol{w}}^*_{im} = 0$, the stationary point is given in [26] as

$$\boldsymbol{u}_{im} \leftarrow (\tilde{\boldsymbol{Q}}_{im} \boldsymbol{W}^{\mathrm{H}}_i)^{-1} \boldsymbol{e}_m, \tag{17}$$

$$\hat{\boldsymbol{u}}_{im} \leftarrow \tilde{\boldsymbol{Q}}^{-1}_{im} \tilde{\boldsymbol{h}}_{im}, \tag{18}$$

$$a_{im} \leftarrow \boldsymbol{u}^{\mathrm{H}}_{im} \tilde{\boldsymbol{Q}}_{im} \boldsymbol{u}_{im}, \tag{19}$$

$$\hat{a}_{im} \leftarrow \boldsymbol{u}^{\mathrm{H}}_{im} \tilde{\boldsymbol{Q}}_{im} \hat{\boldsymbol{u}}_{im}, \tag{20}$$

$$\tilde{\boldsymbol{w}}_{im} \leftarrow \begin{cases} \frac{\boldsymbol{u}_{im}}{\sqrt{a_{im}}} - \hat{\boldsymbol{u}}_{im} & (\hat{a}_{im} = 0), \\ \frac{\hat{a}_{im}}{2a_{im}} \left[1 - \sqrt{1 + \frac{4a_{im}}{|\hat{a}_{im}|^2}} \right] \boldsymbol{u}_{im} - \hat{\boldsymbol{u}}_{im} & (\hat{a}_{im} \neq 0). \end{cases} \tag{21}$$

Similarly to IP, we have to select the appropriate update order from $N!$ options.

Although we obtain two update rules (row-wise or column-wise), each of which has $N!$ update order, it is difficult to determine the most appropriate update strategy in advance. In order to select the appropriate update strategy that derives high separation performance, we propose the following pseudo-SNR-related criterion ζ:

$$\zeta = \frac{1}{N} \sum_n \frac{\sum_{i,j} \{|\mathrm{DNN}_n(|\boldsymbol{Y}_n|^{\cdot 1})|^{\cdot 2}\}_{ij}}{\sum_l \sum_{i,j} \{|\mathrm{DNN}_l(|\boldsymbol{Y}_n|^{\cdot 1})|^{\cdot 2}\}_{ij}}, \tag{22}$$

where $\{\}_{ij}$ denotes the (i, j)th element of the matrix. Since DNN_n is trained to output the scale parameter of the nth source, the numerator of ζ represents the nth source's power in $\boldsymbol{Y}_n$ and the denominator of ζ represents the sum of all the sources' power, which means that ζ indicates the separation accuracy of $\boldsymbol{Y}_n$. We propose the following procedure to select the update strategy when the demixing matrix is updated.

Step 1: Make $2N!$ copies of $\boldsymbol{W}_i$.

Step 2: Each $\boldsymbol{W}_i$ is updated by each update rule ((8)–(9) or (17)–(21)) and their update order ($N!$ options for n).

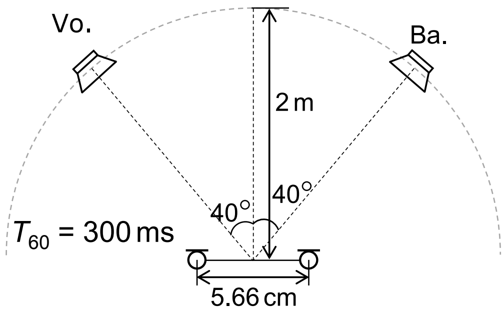

Figure 3. Recording conditions of impulse responses obtained from RWCP database.

Step 3: ζ is calculated via (4) and (22) for each $\boldsymbol{W}_i$.

Step 4: $\boldsymbol{W}_i$ with the highest ζ is taken over to the next iteration.

4 EXPERIMENTAL EVALUATION

We confirmed the validity of the proposed update strategy selection by conducting a music source separation task. We compared six methods: multichannel nonnegative matrix factorization (MNMF) [7], independent low-rank matrix analysis (ILRMA) [9], DNN+WF [27], Duong+DNN [20], conventional row-wise update IDLMA (Row IDLMA), and proposed IDLMA exploiting both row-wise and column-wise updates based on DNN selection (Row+Column IDLMA). MNMF and Duong+DNN estimate the mixing system by the EM algorithm, where MNMF estimates the source model blindly and Duong+DNN estimates it using DNN. ILRMA is a blind method that estimates the demixing system by IP. DNN+WF applies a Wiener filter constructed using all the outputs of the DNN source models to the reference channel signal. Note that MNMF and ILRMA are "blind (unsupervised)" techniques, but we show their performances just for reference to understand to what extent the supervised methods (DNN+WF, Duong+DNN, Row IDLMA, and Row+Column IDLMA) can improve the performance by using pretraining data. For all methods except DNN+WF, we updated the demixing matrix 200 times. In Duong+DNN, Row IDLMA, and Row+Column IDLMA, the scale parameter matrix $\boldsymbol{R}_n$ was updated by DNN$_n$ after every 10 iterations of the spatial parameter optimization. We treated 10 iterations of the spatial parameter optimization as one update strategy and DNN selection of update strategy is done every time after $\boldsymbol{R}_n$ is updated. In Row IDLMA, we averaged the separation performance of all the row-wise update order.

We used the DSD100 dataset of SiSEC2016 [28] as the dry sources and the training dataset of DNN. The 50 songs in the `dev` data were used to train DNN$_n$ and the top 25 songs in alphabetical order in the `test` data were used for performance evaluation. The test songs were trimmed only in the interval of 30 to 60 s. To simulate reverberant mixtures, we produced two-channel observed signals by convoluting the impulse response E2A ($T_{60} = 300\,\mathrm{ms}$) obtained from the RWCP database [29] with each source, and mixtures of bass and vocal were created. The recording conditions of E2A is shown in Fig. 3. All the signals were downsampled to 8 kHz. An STFT was performed using a 512-ms-long Hamming window with a 256-ms-long shift. We used the signal-to-distortion ratio (SDR) [30] to evaluate the total separation performance.

In this paper, the number of hidden layers in the constructed fully connected DNN was set to four. Each layer had 1024 units, and a rectified linear unit was used for the output of each layer. To optimize the DNN, we added the term $(\lambda/2)\sum_q {g_q}^2$ to (12) for regularization, where g_q is the weight coefficient in DNN, and ADADELTA [31] with a 128-size minibatch was performed for 2000 epochs. The parameter ε was experimentally optimized and set to $0.1 \times (IJ)^{-1}\sum_{i,j}\sigma_{ijn}$. The other parameters were set to $\delta_1 = \delta_2 = 10^{-5}$ and $\lambda = 10^{-5}$.

Figure 4 shows the average SDR improvement for bass/vocal separation. The proposed Row+Column IDLMA achieves the best SDR improvement among all the state-of-the-art blind and supervised methods. In particular, the proposed Row+Column IDLMA increases the SDR improvement from Row IDLMA by 1.5 dB, which

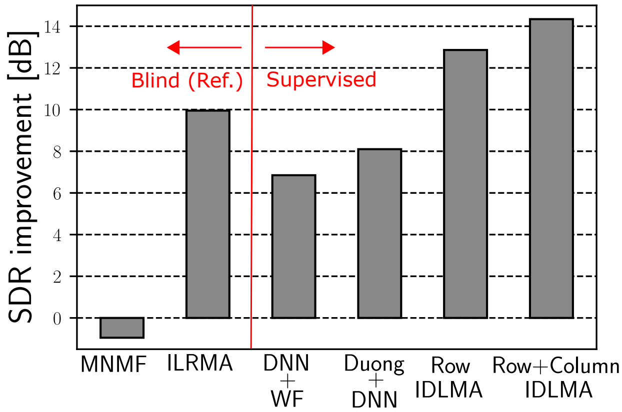

Figure 4. Average SDR improvement of 25 bass/vocal songs.

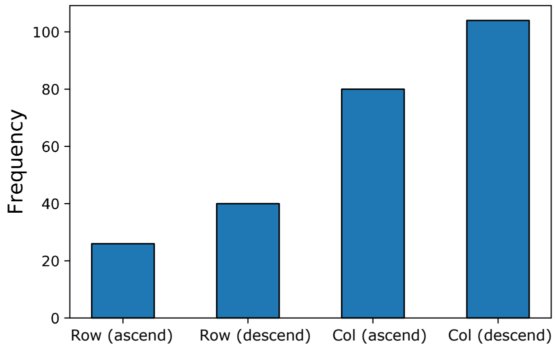

Figure 5. Histogram of selected update rule and update order in Row+Col IDLMA

indicates that DNN selects the appropriate update strategy. Figure 5 shows the histogram of the update rule and order selected in Row+Column IDLMA. In Fig. 5, “Row” or “Col” is the type of update rule and “ascend” or “descend” is the update order. Here “ascend” means that the demixing filter corresponding to Vo. was updated first in “Row” and the vector corresponding to the right microphone in Fig. 3 was updated first in “Col”. It is confirmed that the column-wise update is chosen 2.8 times as much as the row-wise update, which indicates that the column-wise update often leads to a better solution than IP in IDLMA.

5 CONCLUSIONS

In this paper, we proposed a new microphone-wise update algorithm for demixing matrix. Moreover we proposed a method to select the appropriate row- or column-wise update rule depending on the source signal's pseudo SNR estimated via the DNNs. Experimental results showed that the update strategy selection improves the separation performance, which exploits the proposed column-wise update more frequently than the conventional row-wise update.

ACKNOWLEDGEMENTS

This work was partly supported by SECOM Science and Technology Foundation and JSPS KAKENHI Grant Numbers JP17H06101, JP19H01116, and JP19K20306.

REFERENCES

[1] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari. The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech. *IEEE Trans. SAP*, 11(2):109–116, 2003.

[2] 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 Trans. ASLP*, 14(2):666–678, 2006.

[3] A. Hiroe. Solution of permutation problem in frequency domain ica, using multivariate probability density functions. In *Independent Component Analysis and Blind Signal Separation*, pages 601–608, 2006.

[4] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee. Blind source separation exploiting higher-order frequency dependencies. *IEEE Trans. ASLP*, 15(1):70–79, 2007.

[5] N. Ono. Stable and fast update rules for independent vector analysis based on auxiliary function technique. In *Proc. WASPAA*, pages 189–192, 2011.

[6] A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Trans. ASLP*, 18(3):550–563, 2010.

[7] H. Sawada, H. Kameoka, S. Araki, and N. Ueda. Multichannel extensions of non-negative matrix factorization with complex-valued data. *IEEE Trans. ASLP*, 21(5):971–982, 2013.

[8] H. Kameoka, T. Yoshioka, M. Hamamura, J. L. Roux, and K. Kashino. Statistical model of speech signals based on composite autoregressive system with application to blind source separation. In *Proc. LVA/ICA*, pages 245–253, 2010.

[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 Trans. ASLP*, 14(9):1626–1641, 2016.

[10] Y. Mitsui, D. Kitamura, S. Takamichi, N. Ono, and H. Saruwatari. Blind source separation based on independent low-rank matrix analysis with sparse regularization for time-series activity. In *Proc. ICASSP*, pages 21–25, 2017.

[11] 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*, pages 1–5, 2017.

[12] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari. Determined blind source separation with independent low-rank matrix analysis. In Shoji Makino, editor, *Audio Source Separation*, chapter 6, pages 125–155. Springer, Cham, 2018.

[13] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo. Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation. *EURASIP J. Adv. Signal Process.*, 2018(28):1–25, 2018.

[14] R. Ikeshita and Y. Kawaguchi. Independent low-rank matrix analysis based on multivariate complex exponential power distribution. In *Proc. ICASSP*, pages 741–745, 2018.

[15] S. Mogami, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, H. Nakajima, and N. Ono. Independent low-rank matrix analysis based on time-variant sub-Gaussian source model. In *Proc. APSIPA*, pages 1684–1691, 2018.

[16] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari. A review of blind source separation methods: two converging routes to ILRMA originating from ICA and NMF. *APSIPA Trans. Signal. Info. Process*, 8:e12, 2019.

[17] 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. FECCS*, E102-A(2):458–463, 2019.

[18] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura. Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration. *IEEE/ACM Trans. ASLP*, 23(4):654–669, 2015.

[19] S. Araki, T. Hayashi, M. Delcroix, M. Fujimoto, K. Takeda, and T. Nakatani. Exploring multi-channel features for denoising-autoencoder-based speech enhancement. In *Proc. ICASSP*, pages 116–120, 2015.

[20] A. A. Nugraha, A. Liutkus, and E. Vincent. Multichannel audio source separation with deep neural networks. *IEEE/ACM Trans. ASLP*, 24(9):1652–1664, 2016.

[21] Y.-H. Tu, J. Du, L. Sun, and C.-H. Lee. LSTM-based iterative mask estimation and post-processing for multi-channel speech enhancement. In *Proc. APSIPA*, 2017.

[22] K. Qian, Y. Zhang, S. Chang, X. Yang, D. Florencio, and M. Hasegawa-Johnson. Deep learning based speech beamforming. In *Proc. ICASSP*, pages 5389–5393, 2018.

[23] S. Mogami, H. Sumino, D. Kitamura, N. Takamune, S. Takamichi, H. Saruwatari, and N. Ono. Independent deeply learned matrix analysis for multichannel audio source separation. In *Proc. EUSIPCO*, pages 1571–1575, 2018.

[24] N. Makishima, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, and H. Nakajima. Generalized-Gaussian-distribution-based independent deeply learned matrix analysis for multichannel audio source separation. In *Proc. INTERNOISE*, number 1260, 2019 in press.

[25] N. Ono and S. Miyabe. Auxiliary-function-based independent component analysis for super-Gaussian sources. In *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, 2010.

[26] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo. Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis. In *Proc. ICASSP*, pages 746–750, 2018.

[27] S. Uhlich, F. Giron, and Y. Mitsufuji. Deep neural network based instrument extraction from music. In *Proc. ICASSP*, pages 2135–2139, 2015.

[28] 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*, pages 323–332, 2012.

[29] S. Nakamura, K. Hiyane, F. Asano, T. Nishimura, and T. Yamada. Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition. In *Proc. LREC*, pages 965–968, 2000.

[30] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE Trans. ASLP*, 14(4):1462–1469, 2006.

[31] M. D. Zeiler. Adadelta: An adaptive learning rate method. *CoRR*, abs/1212.5701, 2012.