DETERMINED BLIND SOURCE SEPARATION VIA PROXIMAL SPLITTING ALGORITHM

Kohei Yatabe[†] *and Daichi Kitamura*[‡*]

[†]Department of Intermedia Art and Science, Waseda University, Tokyo, Japan
[‡]Department of Information Physics and Computing, The University of Tokyo, Tokyo, Japan

ABSTRACT

The state-of-the-art algorithms of determined blind source separation (BSS) methods based on the independent component analysis (ICA) have gained computational efficiency by the majorization-minimization (MM) principle with a price of losing flexibility. That is, replacing and comparing different source models are not easy in such MM-based framework because it requires efforts to derive a new algorithm each time when one changes the model. In this paper, a general framework for obtaining an ICA-based BSS algorithm is proposed so that a source model can easily be replaced because only a single line of the algorithm must be modified. A sparsity-based extension of the independent vector analysis and a low-rankness-based BSS model using the nuclear norm are also proposed to demonstrate the simplicity and easiness of the proposed framework.

Index Terms— Independence-based separation, frequency domain independent component analysis (FDICA), independent vector analysis (IVA), primal-dual splitting algorithm, proximity operator.

1. INTRODUCTION

Blind source separation (BSS) is methodology for recovering source signals from multiple mixtures without any knowledge about the mixing system. Let a convolutive mixing process be approximated in time-frequency domain as

$$\mathbf{x}[t,f] \approx A[f]\mathbf{s}[t,f], \tag{1}$$

where $\mathbf{x} = [x_1, x_2, \dots x_M]^T$ is an observation obtained by M microphones, $\mathbf{s} = [s_1, s_2, \dots s_N]^T$ is a source signal to be recovered, $A[f]$ is an $M \times N$ mixing matrix, and t and f are indices of time and frequency, respectively. Then, the aim of BSS is to recover N source signals $\mathbf{s}$ from the mixtures $\mathbf{x}$. In a determined or overdetermined situation ($M \geq N$), many of the BSS problems are formulated as an estimation problem of finding an $N \times M$ demixing matrix $W[f]$ which is a left inverse of $A[f]$ (i.e., $W[f]A[f] = I$), and the source signals are recovered by simple multiplication:

$$W[f]\mathbf{x}[t,f] \approx W[f]A[f]\mathbf{s}[t,f] = \mathbf{s}[t,f]. \tag{2}$$

For the sake of simplicity, only a determined situation ($M = N$) is considered in this paper.

For estimating a demixing matrix $W[f]$, statistical independence between source signals is often assumed that leads to a family of independence-based BSS algorithms. Arguably, independent component analysis (ICA) [1] applied in frequency domain (FDICA) [2–6] is one of the most famous methods among them. However, FDICA suffers from the so-called permutation problem [7–10], and thus some recent developments on BSS aim to avoid it by considering more sophisticated models of source signals. For instance, independent vector analysis (IVA) [11–13] assumes co-occurrence among the frequency components in each source, and independent low-rank matrix analysis (ILRMA) [14–16] assumes low-rankness on spectrogram of each source. The key to success of these methods is to incorporate prior knowledge of source signals into their formulations. That is, improvement brought by these methods relies on the preciseness of their source models. Therefore, seeking a better model is the important process for developing a novel and effective BSS method.

However, recent algorithms [16–18] cannot be applied to a different source model directly because they are specialized to each method. These state-of-the-art algorithms are based on the majorization-minimization (MM) principle which requires specially designed upper-bound of the objective functions. That is, one has to derive a new algorithm each time as the source model is modified. Therefore, it may take a lot of time to examine a new source model especially when the model is a complicated one. If a single algorithm can handle a large number of source models without effort, discovering a better source model should become much easier that possibly boosts the development of BSS.

In this paper, a flexible framework for independence-based BSS is proposed based on a proximal splitting algorithm [19–22]. The usefulness of the splitting algorithm comes from its capability of splitting an optimization problem into several easier subproblems which are handled by the proximity operators. We take advantage of this feature to split the ICA-based BSS problem into two parts so that different source models can easily be combined by modifying only a single line of the algorithm. The proposed framework is tested by introducing three new source models (a sparsity-based extension of IVA, a low-rankness-based model using the nuclear norm, and a sparsity-based extension of the low-rank model), and potentiality of the proposed method is indicated by the result.

2. INDEPENDENCE-BASED BSS

As introduced in the previous section, independence-based BSS methods aim to estimate $M \times M$ demixing matrices $\{W[f]\}_{f=1}^F$ which approximately recover the source signals from the observations as $W[f]\mathbf{x}[t,f] \approx \mathbf{s}[t,f]$. Many of them fall into a minimization problem of the following form:

$$\underset{\{W[f]\}_{f=1}^F}{\text{Minimize}} \quad \mathcal{P}(W[f]\mathbf{x}[t,f]) - \sum_{f=1}^{F} \log|\det(W[f])|, \tag{3}$$

where $\mathcal{P}$ is a real-valued penalty function corresponding to the source model. For example, with some constant C,

$$\mathcal{P}(\mathbf{y}[t,f]) = C\,\|\mathbf{y}[t,f]\|_1 = C\sum_{m=1}^{M}\sum_{t=1}^{T}\sum_{f=1}^{F}|y_m[t,f]| \tag{4}$$

*This work was partly supported by JSPS Grant-in-Aid for Research Activity Start-up (17H06572,17H07191).

recovers the traditional FDICA with the Laplace distribution as the source model (which is separable for each f), and

$$\mathcal{P}(\mathbf{y}[t,f]) = C\,\|\mathbf{y}[t,f]\|_{2,1} = C\sum_{m=1}^{M}\sum_{t=1}^{T}\Bigg(\sum_{f=1}^{F}|y_m[t,f]|^2\Bigg)^{\frac{1}{2}} \quad (5)$$

obtains IVA whose source model is the spherical Laplace distribution. ILRMA can also be interpreted as Eq. (3) with

$$\mathcal{P}(\mathbf{y}[t,f]) = C\sum_{m=1}^{M}\mathcal{D}_R(y_m[t,f]), \quad (6)$$

where $\mathcal{D}_R(y_m[t,f])$ is a measure of low-rankness based on the Itakura–Saito non-negative matrix factorization (IS-NMF) [23]:

$$\mathcal{D}_R(y_m[t,f]) = \quad (7)$$
$$\min_{\varphi_{f,r}^{[m]}\geq 0,\psi_{r,t}^{[m]}\geq 0}\sum_{t=1}^{T}\sum_{f=1}^{F}\Bigg(\frac{|y_m[t,f]|^2}{\sum_{r=1}^{R}\varphi_{f,r}^{[m]}\psi_{r,t}^{[m]}} + \log\sum_{r=1}^{R}\varphi_{f,r}^{[m]}\psi_{r,t}^{[m]}\Bigg).$$

From this perspective, it is clear that the difference of performance among these methods is owing to goodness of the penalty function $\mathcal{P}$. Therefore, performance of a BSS method can be improved by finding a better source model and the corresponding penalty function $\mathcal{P}$. For seeking a better model, it is convenient to have a single algorithm that can handle a large number of source models without spending time for its derivation.

3. PROPOSED FRAMEWORK

In this section, a general algorithm for solving the BSS problems in the form of Eq. (3) is proposed by applying a primal-dual splitting algorithm to the reformulated version of Eq. (3).

3.1. Primal-dual splitting algorithm

Let us consider a minimization problem of the form:

$$\underset{\mathbf{w}}{\text{Minimize}}\quad g(\mathbf{w}) + h(L\mathbf{w}), \quad (8)$$

where g and h are proper lower-semicontinuous convex functions, and L is a matrix. A primal-dual splitting algorithm [21] solves this problem by iterating the following procedure[1]:

$$\left\lfloor\begin{array}{l} \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1 g}\big[\,\mathbf{w}^{[k]} - \mu_1\mu_2 L^H\mathbf{y}^{[k]}\,\big], \\ \mathbf{z}\; = \mathbf{y}^{[k]} + L(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]}), \\ \widetilde{\mathbf{y}} = \mathbf{z} - \mathrm{prox}_{h/\mu_2}\big[\,\mathbf{z}\,\big], \\ (\mathbf{w}^{[k+1]},\mathbf{y}^{[k+1]}) = \alpha(\widetilde{\mathbf{w}},\widetilde{\mathbf{y}}) + (1-\alpha)(\mathbf{w}^{[k]},\mathbf{y}^{[k]}), \end{array}\right. \quad (9)$$

where $\mu_1 > 0$ and $\mu_2 > 0$ are step sizes, and $2 > \alpha > 0$ is a relaxation factor which can adjust the speed of convergence ($\alpha = 1$ is the standard speed, and $\alpha > 1$ accelerates and $\alpha < 1$ slows down the algorithm). The important feature of this algorithm is that each function in the problem is handled through the proximity operator [20],

$$\mathrm{prox}_{\mu g}[\mathbf{y}] = \arg\min_{\mathbf{z}}\Big[\,g(\mathbf{z}) + \frac{1}{2\mu}\,\|\mathbf{y}-\mathbf{z}\|_2^2\,\Big], \quad (10)$$

which accepts not only smooth functions but also non-differentiable functions, including sparsity inducing norms in Eqs. (4) and (5), and non-finite functions such as $-\log$ in Eq. (3). That is, difficulty associated with properties of the functions are eliminated within the proximity operator thanks to the proximity term. Therefore, difficulty of an optimization problem of the form Eq. (8) only depends on computational complexity of the corresponding proximity operators. Fortunately, proximity operators of several functions related to ICA-based BSS can be computed quite efficiently as follows [19,20]:

$$\mathrm{prox}_{-\mu\log}[\,y\,] = \frac{y + \sqrt{y^2 + 4\mu}}{2}, \quad (11)$$

$$\big(\mathrm{prox}_{\mu\|\cdot\|_1}[\mathbf{y}]\big)_m[t,f] = \Bigg(1 - \frac{\mu}{|y_m[t,f]|}\Bigg)_+ y_m[t,f], \quad (12)$$

$$\big(\mathrm{prox}_{\mu\|\cdot\|_{2,1}}[\mathbf{y}]\big)_m[t,f] = \Bigg(1 - \frac{\mu}{(\sum_{f=1}^{F}|y_m[t,f]|^2)^{\frac{1}{2}}}\Bigg)_+ y_m[t,f], \quad (13)$$

where $(\cdot)_+ = \max\{0,\cdot\}$.

3.2. Reformulating optimization problem of ICA-based BSS

To apply the primal-dual splitting algorithm, Eq. (3) is reformulated into the form of Eq. (8). Firstly, for considering the proximity operator, the second term is modified. Since determinant of a matrix can be expressed in terms of the singular values as $|\det(W[f])| = \prod_{m=1}^{M}\sigma_m(W[f])$, Eq. (3) can be rewritten as follows[2]:

$$\underset{\{W[f]\}_{f=1}^{F}}{\text{Minimize}}\quad \mathcal{P}(W[f]\mathbf{x}[t,f]) - \sum_{f=1}^{F}\sum_{m=1}^{M}\log\sigma_m(W[f]), \quad (14)$$

where $\sigma_m(W[f])$ is the mth singular value of $W[f]$.

Then, the optimization variables $\{W[f]\}_{f=1}^{F}$ are vectorized to form a single vector. Let $\mathbf{w}$ be an M^2F-dimensional vector corresponding to the demixing filters $\{W[f]\}_{f=1}^{F}$,

$$\mathbf{w} = [\mathbf{w}[1]^T,\mathbf{w}[2]^T,\ldots,\mathbf{w}[F]^T]^T,\quad \big(\,\mathbf{w}[f] = \mathcal{V}(W[f])\,\big) \quad (15)$$

where $\mathcal{V}$ is a linear operator converting a matrix into a vector,

$$\mathcal{V}(W[f]) = [W_{1,1}[f],\ldots,W_{1,M}[f],W_{2,1}[f],\ldots,W_{M,M}[f]]^T, \quad (16)$$

and let $\mathcal{M}$ be a linear operator converting it back into the matrix,

$$\mathcal{M}(\mathbf{w})[f] = W[f], \quad (17)$$

which will be used as $\mathbf{w}[f] = \mathcal{V}(\mathcal{M}(\mathbf{w}[f])) = \mathcal{V}(\mathcal{M}(\mathbf{w})[f])$. With these notations, Eq. (14) can be expressed as follows:

$$\underset{\mathbf{w}}{\text{Minimize}}\quad \mathcal{P}(X\mathbf{w}) - \sum_{f=1}^{F}\sum_{m=1}^{M}\log\sigma_m(\mathcal{M}(\mathbf{w})[f]), \quad (18)$$

where X is a matrix constructed from the observed data $\mathbf{x}[t,f]$ as

$$X = \mathrm{blkdiag}(\boldsymbol{\chi}[1],\boldsymbol{\chi}[2],\ldots,\boldsymbol{\chi}[F]), \quad (19)$$
$$\boldsymbol{\chi}[f] = \mathrm{blkdiag}(\chi[f],\chi[f],\ldots,\chi[f]),\quad (M\text{ times}) \quad (20)$$
$$\chi[f] = [\tau_1[f],\tau_2[f],\ldots,\tau_M[f]], \quad (21)$$
$$\tau_m[f] = [x_m[1,f],x_m[2,f],\ldots,x_m[T,f]]^T, \quad (22)$$

[1]For the sake of simplicity, details of the algorithm are omitted in this paper. The reader may refer to [21] and references therein for detailed explanations and the condition for convergence. Note that, although several primal-dual algorithms can handle more complicated problem than Eq. (8), this paper only focus on that form because it is sufficient for the proposal.

[2]Note that, while Eq. (3) is only defined for square matrices, the extended formulation in Eq. (14) allows rectangular demixing matrices (i.e., $N \neq M$). Therefore, the proposed method does not require the pre-processing method using principle component analysis (PCA) in an over-determined situation, which might improve the performance as discussed in [15].

$\mathrm{blkdiag}(\cdot)$ is an operator constructing a block-diagonal matrix by concatenating inputted matrices diagonally, $\tau_m[f]$ is $T \times 1$, $\chi[f]$ is $T \times M$, $\boldsymbol{\chi}[f]$ is $MT \times M^2$, and X is $FMT \times FM^2$.

Let the second term in Eq. (18) be shortly denoted by $\mathcal{I}$:

$$\mathcal{I}(\mathbf{w}) = -\sum_{f=1}^{F}\sum_{m=1}^{M} \log \sigma_m(\mathcal{M}(\mathbf{w})[f]). \tag{23}$$

Then, Eq. (18) can be rewritten as follows:

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \mathcal{I}(\mathbf{w}) + \mathcal{P}(X\mathbf{w}), \tag{24}$$

which is in the same form of Eq. (8).

3.3. Proposed algorithm

Since the BSS problem is reformulated into the form of Eq. (8), the primal-dual splitting algorithm in Eq. (9) can be applied to it if the proximity operator of each function is efficiently computable.

It is known that a proximity operator of an orthogonally invariant function can be evaluated by applying the corresponding proximity operator to the singular values of the matrix [20]. By regarding $-\log \sigma_m$ in Eq. (23) as $-\log |\sigma_m|$, the proximity operator of $\mathcal{I}(\mathbf{w})$ is obtained as the similar form to the singular value thresholding:

$$(\mathrm{prox}_{\mu\mathcal{I}}[\mathbf{w}])[f] = \mathcal{V}(\, U\, \widetilde{\Sigma}(\mathcal{M}(\mathbf{w})[f])\, V^H), \tag{25}$$

where $W = U\Sigma V^H$ is the singular value decomposition of W,

$$\widetilde{\Sigma}(W) = \mathrm{diag}(\mathrm{prox}_{-\mu\log}[\sigma_1(W)], \dots, \mathrm{prox}_{-\mu\log}[\sigma_M(W)]), \tag{26}$$

$\mathrm{prox}_{-\mu\log}[\cdot]$ is in Eq. (11), and $\mathrm{diag}(\cdot)$ is the operator constructing a diagonal matrix from inputted scalars. In other words, applying the proximity operator of $-\mu \log$ to each singular value of $W[f]$, for each frequency independently, gives $\mathrm{prox}_{\mu\mathcal{I}}[\cdot]$. It is worth mentioning that this operation is stable because it does not magnify $\|\mathbf{w}\|_2$ much [see Eq. (11)] in contrast to MM algorithms which involve inversions of matrices that sometimes lead to instability.

By using Eq. (25) as the main building block, the primal-dual splitting algorithm for BSS problems is obtained as in Algorithm 1. In the 6th line, the proximity operator of $\mathcal{P}$ is required. For FDICA and IVA in the form of Eqs. (4) and (5), the proximity operators are given in Eqs. (12) and (13), respectively. Any other penalty function $\mathcal{P}$ can be incorporated by only changing $\mathrm{prox}_{\mathcal{P}/\mu_2}[\cdot]$ in that line, and therefore this algorithm can be used to test performance of several source models without efforts on modifying the code. Note that an iterative algorithm can be used to evaluate $\mathrm{prox}_{\mathcal{P}/\mu_2}[\cdot]$ since it is defined through the optimization problem in Eq. (10). That is, it is still possible to apply the proposed algorithm even when the corresponding $\mathrm{prox}_{\mathcal{P}/\mu_2}[\cdot]$ does not admit a closed form solution. For example, the penalty function of ILRMA in Eq. (6) yields the optimization problem similar to IS-NMF which could be solved by the existing algorithms. Therefore, by using the proposed algorithm, incorporating existing source models into ICA-based BSS should be easy if the computational complexity is not a concern.

3.4. Proposed algorithm with multiple penalty functions

A source model consists of two or more penalty functions,

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \mathcal{I}(\mathbf{w}) + \sum_{q=1}^{Q} \mathcal{P}_q(X\mathbf{w}), \tag{27}$$

Algorithm 1 PDS-BSS

1: **Input:** X, $\mathbf{w}^{[1]}$, $\mathbf{y}^{[1]}$, μ_1, μ_2, α
2: **Output:** $\mathbf{w}^{[K+1]}$
3: **for** $k = 1, \dots, K$ **do**
4: $\quad \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1\mathcal{I}}[\, \mathbf{w}^{[k]} - \mu_1\mu_2 X^H \mathbf{y}^{[k]}\,]$
5: $\quad \mathbf{z} = \mathbf{y}^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
6: $\quad \widetilde{\mathbf{y}} = \mathbf{z} - \mathrm{prox}_{\frac{1}{\mu_2}\mathcal{P}}[\, \mathbf{z}\,]$
7: $\quad \mathbf{y}^{[k+1]} = \alpha\widetilde{\mathbf{y}} + (1-\alpha)\mathbf{y}^{[k]}$
8: $\quad \mathbf{w}^{[k+1]} = \alpha\widetilde{\mathbf{w}} + (1-\alpha)\mathbf{w}^{[k]}$
9: **end for**

Algorithm 2 PDS-BSS-multiPenalty

Input: X, $\mathbf{w}^{[1]}$, $\mathbf{y}_1^{[1]}, \dots, \mathbf{y}_Q^{[1]}$, μ_1, μ_2, α
Output: $\mathbf{w}^{[K+1]}$
for $k = 1, \dots, K$ **do**
$\quad \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1\mathcal{I}}[\, \mathbf{w}^{[k]} - \mu_1\mu_2 X^H \big(\sum_{q=1}^{Q} \mathbf{y}_q^{[k]}\big)\,]$
$\quad$ **for** $q = 1, \dots, Q$ **do**
$\quad\quad \mathbf{z}_q = \mathbf{y}_q^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
$\quad\quad \widetilde{\mathbf{y}}_q = \mathbf{z}_q - \mathrm{prox}_{\frac{1}{\mu_2}\mathcal{P}_q}[\, \mathbf{z}_q\,]$
$\quad\quad \mathbf{y}_q^{[k+1]} = \alpha\widetilde{\mathbf{y}}_q + (1-\alpha)\mathbf{y}_q^{[k]}$
$\quad$ **end for**
$\quad \mathbf{w}^{[k+1]} = \alpha\widetilde{\mathbf{w}} + (1-\alpha)\mathbf{w}^{[k]}$
end for

can also be handled by the proposed algorithm. Algorithm 1 is easily extended to deal with this problem by vertically concatenating the matrix X as $L = [X^T, \dots, X^T]^T$ (Q times). Then, the primal-dual splitting algorithm for Eq. (27) can be derived by simply applying Eq. (9) to its vertically concatenated version, which is summarized in Algorithm 2. As presented in the algorithm, each penalty function $\mathcal{P}_q$ is independently handled by the corresponding proximity operator. Therefore, this algorithm is applicable to a complicated source model consisting of several simple functions $\{\mathcal{P}_q\}_{q=1}^{Q}$. Such models may include sparse+low-rank model of robust PCA [24, 25] and multi-group sparsity of harmonic/percussive source separation [26].

3.5. Normalization of data matrix

As introduced in Section 2, an ICA-based BSS method usually has a specific constant C which is derived from the corresponding statistical model. Although this constant is important, it may not be easy to derive C for some non-standard models. To circumvent this complication, a normalization method convenient for the proposed algorithm is considered here. Since the step sizes μ_1 and μ_2 in Eq. (9) should be chosen to satisfy $\mu_1\mu_2\|L\|_{\mathrm{s}}^2 \leq 1$ [21][3], they can be set to one if $\|L\|_{\mathrm{s}} = 1$, where $\|\cdot\|_{\mathrm{s}}$ denotes the spectral norm. That is, if the data matrix X is normalized as

$$\widetilde{X} = X/(\sqrt{Q}\,\|X\|_{\mathrm{s}}), \tag{28}$$

then the step sizes can be set to $\mu_1 = 1$ and $\mu_2 = 1$. Remind that α can be arbitrarily chosen from $(0, 2)$ [21] or chosen as 1 which bypasses the last line of Eq. (9). Thus, with this normalization, there is nothing to worry about in regards to the choice of the parameters.

[3]Because of the $-\log\sigma_m$ term, the problem in Eq. (24) is non-convex, which requires an additional analysis for convergence. Such analysis will be considered in the future work, and thus we omitted theoretical details in this paper. Nevertheless, experimental results in the next section show that this choice of step sizes, derived for convex problems, seems work properly.

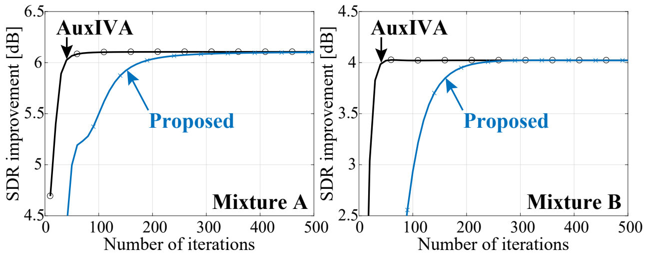

Fig. 1. Comparison between `AuxIVA` [18] and the proposed algorithm. The black lines with circles denotes `AuxIVA`, while the blue lines with cross marks denotes the proposed algorithm.

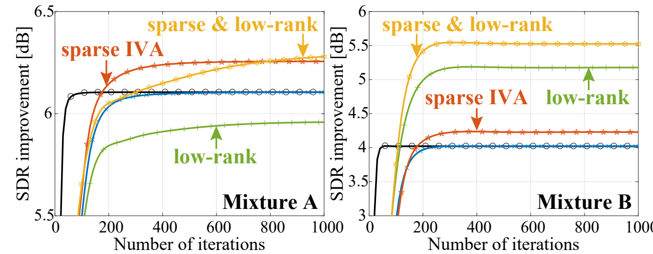

Fig. 2. Performance of several models solved by the proposed algorithm. Each curve represents each model proposed in Section 4.2, while black and blue lines are the ones from Fig. 1 for comparison.

4. EXPERIMENTS

The proposed algorithm was tested by applying it to speech mixtures. The database used in the experiment was a part of SiSEC2011 [27] called `dev1` of the UND task[4]. Live recording (`liverec`) of four female speech sources recorded by two microphones (5 cm spacing) was chosen as the test data. For making the problem determined, two pairs of sources were considered: **Mixture A** consists of two sources arrived from -50° and 45° and **Mixture B** consists of two sources arrived from -10° and 15°, where 0° corresponds to the normal direction to the microphone array. The reverberation time was 130 ms, and 128-ms-long Hann window with 64-ms shift was used. The initial value of demixing matrices $\mathbf{w}^{[1]}$ was set to the identity matrices ($W[f] = I$ for all f). The parameters in the proposed algorithm were $\mu_1 = 1$, $\mu_2 = 1$, and $\alpha = 1.75$, and all initial values of $\mathbf{y}$ in Algorithms 1 and 2 were set to the zero vector.

4.1. Comparison with MM algorithm

In order to compare with the state-of-the-art MM algorithm in [18], the proposed algorithm was applied to IVA whose source model is the spherical Laplace distribution:

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \mathcal{I}(\mathbf{w}) + \|\widetilde{X}\mathbf{w}\|_{2,1}\,, \tag{29}$$

where $\|\cdot\|_{2,1}$ is in Eq. (5). By simply incorporating the proximity operator in Eq. (13) to Algorithm 1, the proposed method is adapted to this problem. The performance at each iteration count was measured by the signal-to-distortion ratio (SDR) [28]. In the MM algorithm [18] which is denoted by `AuxIVA`, the statistical constant C in Eq. (5) was correctly considered.

Figure 1 shows the results of IVA solved by `AuxIVA` and the proposed algorithm. `AuxIVA` reached rapidly to specific values of SDR, where 50 or 100 iteration seems enough for these cases. Although the proposed algorithm required more iterations, it reached to the same values with several hundreds iterations. Running time per iteration of each method was 80.8 ms for `AuxIVA` and 47.4 ms for the proposed algorithm (with Core i5-7200U processor and MATLAB 2017a). This result indicates that the proposed algorithm with the normalization rule in Section 3.5 properly works as the state-of-the-art algorithm with comparable running time.

4.2. Application to several source models

To demonstrate simplicity and easiness of the proposed algorithm, three additional formulations with different source models are considered here. From Eq. (29), it can be clearly seen that IVA attempts to sparsify the spectrograms of separated signals by inducing group sparsity through $\|\cdot\|_{2,1}$, while $\mathcal{I}$ prevents $W[f]$ from being rank deficient. Any other penalty function possibly improves the accuracy if it promotes a property of source signals appropriately [29].

One of such properties is the low-rankness of a spectrogram [14–16]. Then, it may be natural to consider the nuclear norm $\|\cdot\|_*$ because it is well-known as a function inducing low-rankness of a matrix. Therefore, we propose the following model (**low-rank**):

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \mathcal{I}(\mathbf{w}) + \|\mathcal{M}_*(\widetilde{X}\mathbf{w})\|_*\,, \tag{30}$$

where $\mathcal{M}_*$ is the operator converting the vector $\widetilde{X}\mathbf{w}$ into the corresponding matrices (spectrograms of the separated signals), and the nuclear norm is evaluated for each spectrogram. The proximity operator of $\|\cdot\|_*$ is the famous singular-value thresholding [20] which also easily adapts the proposed algorithm to this problem.

Further, it can be presumed that an additional sparsity-inducing term improves separation performance. Thus, to demonstrate the benefit of flexibility of the proposed algorithm, **sparse IVA**,

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \mathcal{I}(\mathbf{w}) + \|\widetilde{X}\mathbf{w}\|_{2,1} + \lambda\,\|\widetilde{X}\mathbf{w}\|_1\,, \tag{31}$$

and a **sparse & low-rank** model,

$$\underset{\mathbf{w}}{\text{Minimize}} \quad \mathcal{I}(\mathbf{w}) + \|\mathcal{M}_*(\widetilde{X}\mathbf{w})\|_* + \lambda\,\|\widetilde{X}\mathbf{w}\|_1\,, \tag{32}$$

are also proposed, where $\lambda > 0$ is a weighting parameter balancing the influence of the norms.

Figure 2 shows the results of these three models solved by the proposed algorithm[5] ($\lambda = 0.002$) together with IVA in Fig. 1 for comparison. For both mixtures, **sparse IVA** and **sparse & low-rank** models resulted in higher SDR than IVA. Although the modification of the proposed algorithm was able to be done within a few minutes, we were able to test the performance of these four BSS models. This fact illustrates the advantage of the proposal in this paper which could contribute to discovery of a new BSS model.

5. CONCLUSIONS

For solving ICA-based BSS problems, this paper proposed a general algorithm which admits a complicated source model consisting of multiple penalty terms. It can handle a large number of models with a slight modification that allows quick investigation of the model's performance. For illustrative examples, three models based on sparsity and low-rankness of the spectrograms were also proposed.

[4] Available at `http://sisec2011.wiki.irisa.fr`

[5] For this experiment, $\sqrt{2}$ in Eq. (28) was omitted for Eqs. (31) and (32) in order to set the trajectories of the four models to the similar ones for easier comparison. It should not be omitted in practice for the sake of reliability.

6. REFERENCES

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

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

[3] 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. Speech Audio Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.

[4] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2004, vol. 3, pp. 885–888.

[5] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[6] 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. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, Mar. 2006.

[7] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of blind signal separation method using directivity pattern under reverberant conditions," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2000, vol. 5, pp. 3140–3143.

[8] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1, pp. 1–24, 2001.

[9] 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 Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[10] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain BSS," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2007, pp. 3247–3250.

[11] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation*, 2006, pp. 601–608.

[12] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ica to multivariate components," in *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation*, 2006, pp. 165–172.

[13] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2007.

[14] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2015, pp. 276–280.

[15] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Relaxation of rank-1 spatial constraint in overdetermined blind source separation," in *Proc. Eur. Signal Process. Conf.*, Aug. 2015, pp. 1261–1265.

[16] 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. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, Sep. 2016.

[17] 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, pp. 165–172.

[18] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2011, pp. 189–192.

[19] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*, pp. 185–212, Springer, New York, 2011.

[20] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, 2014.

[21] N. Komodakis and J. C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," *IEEE Signal Process. Mag.*, vol. 32, no. 6, pp. 31–54, Nov. 2015.

[22] M. Burger, A. Sawatzky, and G. Steidl, *First Order Algorithms in Variational Image Processing*, pp. 345–407, Springer, Cham, 2016.

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

[24] P. S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson, "Singing-voice separation from monaural recordings using robust principal component analysis," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2012, pp. 57–60.

[25] Y. Ikemiya, K. Itoyama, and K. Yoshii, "Singing voice separation and vocal F0 estimation based on mutual combination of robust principal component analysis and subharmonic summation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 11, pp. 2084–2095, Nov. 2016.

[26] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram," in *Proc. Eur. Signal Process. Conf.*, 2008.

[27] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011): - audio source separation -," in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, 2012, pp. 414–422.

[28] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[29] K. Yatabe and Y. Oikawa, "Phase corrected total variation for audio signals," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2018.