Determined Blind Source Separation Using Metric Projection and Proximity Operator of Log-Det Function Under Projection-back Constraint

Kazuki Matsumoto and Kohei Yatabe
Tokyo University of Agriculture and Technology, Tokyo, Japan

***Abstract*—Determined blind source separation (BSS) formulated in the time-frequency domain usually applies pre- and post-processing called *whitening* and *projection back*. They convert the BSS problem into an easier optimization problem, and the existing methods have relied on them. However, these pre- and post-processing steps modify the optimized signals, which prevents us from directly modeling the source signals. In this paper, to directly handle the signals in BSS, we define the *projection-back constraint set* and derive the metric projection and the proximity operator related to it. We also propose two algorithms using the alternating direction method of multipliers (ADMM) and experimentally confirm their correctness.**

***Index Terms*—Audio source separation, time-frequency domain, demixing matrix estimation, affine constraint, alternating direction method of multipliers (ADMM).**

I. Introduction

Determined blind source separation (BSS) aims to recover source signals from a multichannel mixture without information about the mixing system. Many methods formulate it as an estimation problem of demixing matrices in the time-frequency domain and solve it based on some prior knowledge of source signals, such as sparsity [1]–[7] and low-rankness of amplitude spectrograms [8]–[10]. Some of the recent methods use deep neural networks (DNNs) to learn source models from training data [11]–[18]. Prior knowledge of the mixing and demixing systems can also be utilized [19], [20].

Many BSS methods employ pre- and post-processing called *whitening* and *projection back (PB)*, respectively. Whitening helps optimization algorithms, and PB resolves the scale indeterminacy of the separated signals. However, whitening and PB prevent us from directly modeling the output signals of BSS methods. Namely, even if an optimization problem is designed to induce some desirable properties of the demixing system and/or the separated signals, there is no guarantee that these properties are preserved by whitening and PB. This has been noticed by some researchers, but the existing methods used heuristics that only partially resolve this issue [10], [11].

In this paper, we propose a unified optimization framework that integrates the pre- and post-processing stages into the optimization problem so that the overall demixing system and separated signals can be directly modeled. We define the set of demixing matrices that satisfies the PB constraint [21] and derive two optimization algorithms based on the alternating direction methods of multipliers (ADMM) [22]. The metric projection and proximity operator related to the PB constraint set are also derived for computation in the algorithms. Our contributions are summarized as follows: (1) we proposed a unified formulation for determined BSS, which integrates the pre-processing (whitening) and the post-processing (PB) into the optimization problem, allowing direct modeling of the final output of BSS algorithms; (2) we derived analytical formulas of the metric projection onto the PB constraint set and the proximity operator of the log-det function under PB constraint; and (3) we derived and evaluated ADMM-based BSS algorithms incorporating the PB constraint.

This work was partly supported by JST FOREST Program, Grant Number JPMJFR2330.

II. Preliminaries

A. Determined Blind Source Separation

When K source signals are mixed by an unknown mixing system and observed by K microphones, source separation can be formulated as an optimization problem of finding $K \times K$ demixing matrices in the time-frequency domain. Let the source signals at the (f,t)th bin be $\mathbf{s}[f,t] = [s_1[f,t], \ldots, s_K[f,t]]^\mathsf{T} \in \mathbb{C}^K$, and the observed signals be $\mathbf{x}^{(\mathrm{obs})}[f,t] = [x_1^{(\mathrm{obs})}[f,t], \ldots, x_K^{(\mathrm{obs})}[f,t]]^\mathsf{T} \in \mathbb{C}^K$, where $1 \leq t \leq T$ and $1 \leq f \leq F$ are the time and frequency indices, respectively, and $(\cdot)^\mathsf{T}$ denotes the transpose. A convolutive mixing system in the time domain is approximated by the mixing matrix $\mathbf{H}[f] \in \mathbb{C}^{K \times K}$ at each frequency as $\mathbf{x}^{(\mathrm{obs})}[f,t] = \mathbf{H}[f]\,\mathbf{s}[f,t]$. Then, the separated signal $\tilde{\mathbf{y}}[f,t] = \big[\tilde{y}_1[f,t], \ldots, \tilde{y}_K[f,t]\big]^\mathsf{T} \in \mathbb{C}^K$ is obtained by applying the demixing matrix $\widetilde{\mathbf{W}}[f] \in \mathbb{C}^{K \times K}$ for each frequency as [23]

$$\tilde{\mathbf{y}}[f,t] = \widetilde{\mathbf{W}}[f]\,\mathbf{x}^{(\mathrm{obs})}[f,t] \approx \mathbf{s}[f,t]. \tag{1}$$

B. Three-stage Procedure for Demixing Matrix Estimation

The three-stage procedure in Fig. 1 is commonly used for stable estimation of demixing matrices [1]. Each stage can be represented by the application of the $K \times K$ matrices:

1) **Whitening** $\mathbf{Q}[f] \in \mathbb{C}^{K \times K}$: pre-processing matrix for assisting optimization algorithms.
2) **Demixing** $\mathbf{W}[f] \in \mathbb{C}^{K \times K}$: demixing matrix optimized based on some prior knowledge of the source signals and/or the demixing system.
3) **Projection Back (PB)** $\mathbf{D}[f] \in \mathbb{C}^{K \times K}$: post-processing matrix that adjusts the scale of the separated signals.

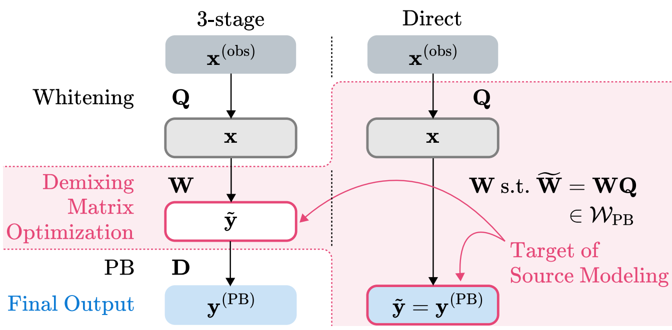

Fig. 1. Comparison between the conventional three-stage source separation (3-stage) and the proposed direct optimization of the final output (Direct). In the three-stage procedure, the scale of the separated signal $\tilde{\mathbf{y}}$ is adjusted by projection back (PB). On the other hand, in the direct approach proposed in this paper, the optimized demixing matrix $\widetilde{\mathbf{W}}$ (which is the composition of the whitening matrix $\mathbf{Q}$ and the demixing matrix $\mathbf{W}$) is constrained to the set $\mathcal{W}_{\mathrm{PB}}$ in Eq. (13). This ensures that the separated signal $\tilde{\mathbf{y}}$ during the optimization stage is the same as the final output after PB $\mathbf{y}^{(\mathrm{PB})}$.

Then, the output signal of a BSS method $\mathbf{y}^{(\mathrm{PB})}[f,t] \in \mathbb{C}^N$ is obtained by the following application of the matrices:

$$\mathbf{y}^{(\mathrm{PB})}[f,t] = \underbrace{\underbrace{\mathbf{D}[f]}_{\text{PB}} \; \underbrace{\underbrace{\mathbf{W}[f]}_{\text{Demixing}} \; \underbrace{\mathbf{Q}[f]}_{\text{Whitening}}}_{\widetilde{\mathbf{W}} \text{ in Eq. (1)}}}_{\mathbf{W}^{(\mathrm{PB})} \text{ (Overall demixing matrix after PB)}} \mathbf{x}^{(\mathrm{obs})}[f,t]. \tag{2}$$

For brevity, we represent all the components altogether by omitting the indices, e.g., $\mathbf{x} = (\mathbf{x}[f,t])_{f=1,t=1}^{F,T}$ and $\mathbf{W} = (\mathbf{W}[f])_{f=1}^{F}$, and shortly write the linear operations as $\mathbf{y} = \mathbf{W}\mathbf{x} = (\mathbf{W}[f]\mathbf{x}[f,t])_{f=1,t=1}^{F,T}$.

The details of each stage are described below.

Stage-1) Whitening matrix $\mathbf{Q}$: First, to convert the optimization problem into an easier one, the whitening matrix $\mathbf{Q}$ is applied to the observed mixture as follows:

$$\mathbf{x} = \mathbf{Q}\mathbf{x}^{(\mathrm{obs})}, \tag{3}$$

where $\mathbf{Q}$ is chosen so that the observed signals after whitening $\mathbf{x}$ satisfy $\sum_{t=1}^{T} \mathbf{x}[f,t](\mathbf{x}[f,t])^{\mathsf{H}} = \mathbf{I}$, see [1].

Stage-2) Demixing matrix $\mathbf{W}$: The next step is the optimization of the demixing matrix $\mathbf{W}$. It can be formulated as the minimization problem as follows [23]:

$$\min_{(\mathbf{W},\tilde{\mathbf{y}})} \;\; \mathcal{P}(\tilde{\mathbf{y}}) - \sum_{f=1}^{F} \log(|\det(\mathbf{W}[f])|) \quad \text{s.t.} \quad \tilde{\mathbf{y}} = \mathbf{W}\mathbf{x}. \tag{4}$$

The objective function consists of the log-det function that ensures the invertibility of the demixing matrix and a penalty function $\mathcal{P} : \mathbb{C}^{K \times T \times F} \to \mathbb{R}$ that corresponds to the prior knowledge of the source signals. When using the local Gaussian model (LGM) [24], [25] as the source prior, the penalty function is reduced to the weighted quadratic function:

$$\mathcal{P}_{\mathrm{LGM}}(\mathbf{y}) = \sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{f=1}^{F} \frac{|y_k[f,t]|^2}{\sigma_k^2[f,t]}, \tag{5}$$

where $\sigma_k^2[f,t] \geq 0$ is the variance parameter for each (f,t)th bin. The variance parameter $\sigma_k^2[f,t]$ is often modeled using nonnegative matrix factorization (NMF) [8]–[10] or estimated using DNNs trained by source signals [11]–[16]. Composition with the whitening matrix $\mathbf{Q}$ gives the demixing matrix $\widetilde{\mathbf{W}}$ and separated signal $\tilde{\mathbf{y}}$ obtained after the optimization stage:

$$\widetilde{\mathbf{W}} = \mathbf{W}\mathbf{Q}, \qquad \tilde{\mathbf{y}} = \mathbf{W}\mathbf{x} \; (= \widetilde{\mathbf{W}}\mathbf{x}^{(\mathrm{obs})}). \tag{6}$$

Stage-3) PB Matrix $\mathbf{D}$: Finally, to address the scale ambiguity of the separated signal, PB is applied as post-processing [1]. PB adjusts the scale of the separated signals by referring to the signal observed at the $k^{(\mathrm{ref})}$th microphone $(1 \leq k^{(\mathrm{ref})} \leq K)$. In this paper, we set $k^{(\mathrm{ref})} = 1$ without loss of generality. Then the PB matrix $\mathbf{D}[f]$ is a diagonal matrix given by

$$\mathbf{D}[f] = \mathrm{diag}(\tilde{h}_{1,1}[f], \ldots, \tilde{h}_{1,K}[f]), \tag{7}$$

where $\mathrm{diag}(\cdot)$ contracts a diagonal matrix from the input, and the elements $(\tilde{h}_{1,k}[f])_{k=1}^{K}$ are the first row of the estimated mixing matrix $\widetilde{\mathbf{H}}[f]$, which is calculated as

$$\widetilde{\mathbf{H}}[f] = (\widetilde{\mathbf{W}}[f])^{-1}. \tag{8}$$

The final output, i.e., the overall demixing matrix $\mathbf{W}^{(\mathrm{PB})}$ (that is applied to the observed mixture $\mathbf{x}^{(\mathrm{obs})}$) and the separated signal $\mathbf{y}^{(\mathrm{PB})}$, is given by

$$\mathbf{W}^{(\mathrm{PB})} = \mathbf{D}\widetilde{\mathbf{W}}, \qquad \mathbf{y}^{(\mathrm{PB})} = \mathbf{D}\tilde{\mathbf{y}} \; (= \mathbf{W}^{(\mathrm{PB})}\mathbf{x}^{(\mathrm{obs})}). \tag{9}$$

This three-stage procedure helps optimization by converting the problem into an easier one. However, the demixing matrix $\mathbf{W}$ and the separated signal $\tilde{\mathbf{y}}$ in the optimization problem in Eq. (4) is different from the final output of BSS methods (i.e., the overall demixing matrix, $\mathbf{W}^{(\mathrm{PB})} = \mathbf{D}\mathbf{W}\mathbf{Q}$, and the separated signals after PB, $\mathbf{y}^{(\mathrm{PB})} = \mathbf{W}^{(\mathrm{PB})}\mathbf{x}^{(\mathrm{obs})}$). Therefore, the source model for the optimization problem in Eq. (4) cannot directly handle the separated signal in the final output.

III. Proposed Method

To directly handle the final output of BSS methods, we propose a unified formulation that imposes PB as a constraint during optimization of the demixing matrix (Fig. 1, Direct). The PB constraint ensures that the signals separated by an algorithm are the same as those after PB. Here, we first formulate the BSS problem under the PB constraint and then derive two ADMM algorithms, along with the metric projection and the proximity operator necessary for implementing the algorithms.

A. Determined BSS Problem Under PB Constraint

When a demixing matrix $\widetilde{\mathbf{W}}[f] \in \mathbb{C}^{K \times K}$ in Eq. (2) is invertible and satisfies the following condition [21],

$$\sum_{k=1}^{K} \tilde{w}_{kj}[f] = \delta_{ij} = \begin{cases} 1 & (j = 1), \\ 0 & (j \neq 1), \end{cases} \tag{10}$$

then the demixing matrix $\widetilde{\mathbf{W}}$ is the same as that after PB[1]:

$$\mathbf{W}^{(\mathrm{PB})}[f] = \widetilde{\mathbf{W}}[f], \tag{11}$$

where δ_{ij} is Kronecker's delta. Therefore, the separated signal after PB, $\mathbf{y}^{(\mathrm{PB})}$, can be directly optimized by introducing the PB constraint on $\widetilde{\mathbf{W}}$ $(=\mathbf{WQ})$ into the problem in Eq. (4) as

$$\begin{aligned} \min_{(\mathbf{W},\mathbf{y}^{(\mathrm{PB})})} \quad & \mathcal{P}(\mathbf{y}^{(\mathrm{PB})}) - \sum_{f=1}^{F} \log(|\det(\mathbf{W}[f])|) \\ \text{s.t.} \quad & \mathbf{y}^{(\mathrm{PB})} = \mathbf{WQx}^{(\mathrm{obs})}, \quad \mathbf{WQ} \in \mathcal{W}_{\mathrm{PB}}^{(K)}, \end{aligned} \tag{12}$$

where we define *PB constraint set* $\mathcal{W}_{\mathrm{PB}}^{(K)}$ as the set of $K \times K$ matrices satisfying condition Eq. (10), i.e.,

$$\mathcal{W}_{\mathrm{PB}}^{(K)} = \left\{ \mathbf{W} \in \mathbb{C}^{K\times K\times F} \,\middle|\, \sum_{k=1}^{K} w_{kj}[f] = \delta_{1j},\ {}^{\forall}(f,j) \right\}. \tag{13}$$

B. Metric Projection to PB Constraint Set $\mathcal{W}_{\mathrm{PB}}$

To handle the PB constraint in Eq. (12) within a proximal splitting algorithm [27], we introduce *metric PB*, defined as the metric projection operator onto the PB constraint set:

$$\mathrm{proj}_{\mathcal{W}_{\mathrm{PB}}^{(K)}}(\widetilde{\mathbf{W}}) = \arg\min_{\mathbf{W}\in\mathcal{W}_{\mathrm{PB}}^{(K)}} \|\mathbf{W} - \widetilde{\mathbf{W}}\|_{\mathrm{Fro}}^2. \tag{14}$$

where $\|\cdot\|_{\mathrm{Fro}}$ denotes the Frobenius norm, and the indices $[f]$ are omitted hereafter for brevity. This projection can be easily computed as follows.

Proposition 1. The metric projection onto the PB constraint set $\mathcal{W}_{\mathrm{PB}}^{(K)}$ is given as follows:

$$\left(\mathrm{proj}_{\mathcal{W}_{\mathrm{PB}}^{(K)}}(\widetilde{\mathbf{W}})\right)_{ij} = \tilde{w}_{ij} + \frac{1}{K}\left(\delta_{1j} - \sum_{k=1}^{K} \tilde{w}_{kj}\right). \tag{15}$$

Proof. The left-hand side in Eq. (10) is the linear operation, and the right-hand side is a constant vector. Therefore, the PB constraint set $\mathcal{W}_{\mathrm{PB}}^{(K)}$ in Eq. (13) is an affine set. The metric projection onto an affine set can be found in, e.g., [27]. ☐

Note that, since $\mathcal{W}_{\mathrm{PB}}^{(K)}$ is a convex set, the metric PB in Eq. (15) is non-expansive (i.e., 1-Lipschitz). This is distinct from the conventional PB in Eqs. (7)–(9) that is not non-expansive due to the matrix inversion. Indeed, as will be shown in the experimental section, heuristic application of the conventional PB causes instability of the algorithms.

[1]Here, the derivation of the condition Eq. (10) is shortly explained. Let $\mathbf{D}[f]$ be the PB matrix corresponding to a demixing matrix $\widetilde{\mathbf{W}}[f]$. Assuming that $\mathbf{W}^{(\mathrm{PB})}[f] = \widetilde{\mathbf{W}}[f]$ in the left part of Eq. (9), Eq. (7) becomes $\mathbf{D}[f] = \mathrm{diag}(\tilde{h}_{1,1}[f], \ldots, \tilde{h}_{1,K}[f]) = \mathbf{I}$. Combining this with Eq. (8), we obtain

$$\widetilde{\mathbf{W}}^{-1}[f] = \widetilde{\mathbf{H}}[f] = \begin{pmatrix} 1 & \cdots & 1 \\ * & \cdots & * \\ \vdots & \ddots & \vdots \\ * & \cdots & * \end{pmatrix},$$

where $*$ can be any scalar. Since $\mathbf{e}_1^{\mathsf{T}}\widetilde{\mathbf{W}}^{-1}[f] = \mathbf{1}^{\mathsf{T}}$, the above equation can be reduced to $\mathbf{1}^{\mathsf{T}}\widetilde{\mathbf{W}}[f] = \mathbf{e}_1^{\mathsf{T}}$ that is equivalent to Eq. (10), where $\mathbf{1} = [1,\ldots,1]^{\mathsf{T}} \in \{1\}^K$, and $\mathbf{e}_1 = [1,0,\ldots,0]^{\mathsf{T}} \in \{0,1\}^K$. A similar discussion can be found in the literature [21], where we independently obtained the same condition at the same time [26].

Algorithm 1 ADMM-BSS using metric PB in Eq. (15)

Input: $\mathbf{x}$, $\mathbf{Q}$, $(\mathbf{y}^{(\mathrm{PB})}, \mathbf{W}^{(\mathrm{PB})}, \check{\mathbf{W}})$, $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, ρ
Output: $\mathbf{y}^{(\mathrm{PB})}$, $\mathbf{W}^{(\mathrm{PB})}$

1: **for** $l = 1, \ldots,$ numIter **do**
2: $\quad \mathbf{y}^{(\mathrm{PB})} \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{P}}(\mathbf{Wx} + \mathbf{u}_1)$
3: $\quad \mathbf{W}^{(\mathrm{PB})} \leftarrow \mathrm{proj}_{\mathcal{W}_{\mathrm{PB}}^{(K)}}(\mathbf{WQ} + \mathbf{u}_2)$
4: $\quad \check{\mathbf{W}} \leftarrow \mathrm{prox}_{(1/\rho)\mathrm{LogDet}}(\mathbf{W} + \mathbf{u}_3)$
5: $\quad (\mathbf{u}_i)_{i=1}^3 \leftarrow (\mathbf{u}_i)_{i=1}^3 + (\mathbf{Wx}, \mathbf{WQ}, \mathbf{W}) - (\mathbf{y}^{(\mathrm{PB})}, \mathbf{W}^{(\mathrm{PB})}, \check{\mathbf{W}})$
6: $\quad \mathbf{W} \leftarrow \big((\mathbf{y}^{(\mathrm{PB})} - \mathbf{u}_1)\,\mathbf{x}^{\mathsf{H}} + (\mathbf{W}^{(\mathrm{PB})} - \mathbf{u}_2)\,\mathbf{Q}^{\mathsf{H}} + (\check{\mathbf{W}} - \mathbf{u}_3)\big)(\mathbf{xx}^{\mathsf{H}} + \mathbf{QQ}^{\mathsf{H}} + \mathbf{I})^{-1}$
7: **end for**

Applying ADMM [22] to Eq. (12) yields Alg. 1, where we used a light notation for operation in the 6th line $\mathbf{ab}^{\mathsf{H}}$ that computes $\sum_{t=1}^{T} \mathbf{a}[f,t](\mathbf{b}[f,t])^{\mathsf{H}}$ for each frequency. The variable $\check{\mathbf{W}} \in \mathbb{C}^{K\times K\times F}$ is for variable splitting, $\mathbf{u}_1 \in \mathbb{C}^{K\times T\times F}$, $\mathbf{u}_2, \mathbf{u}_3 \in \mathbb{C}^{K\times K\times F}$ are dual variables, and $\rho > 0$ is a step size.

C. Proximity Operator of Log-Det Function Under Projection Back Constraint When $K = 2$

Here, we propose a simpler algorithm for the special case $\mathbf{Q} = \mathbf{I}$ and $K = 2$ (i.e., separating two source signals without using whitening). By simultaneously considering the log-det function and the PB constraint in the proximity operator as

$$\begin{aligned} &\mathrm{prox}_{\lambda\,\mathrm{LogDet}+\mathrm{PB}}(\widetilde{\mathbf{W}}) \\ &= \arg\min_{\mathbf{W}\in\mathcal{W}_{\mathrm{PB}}^{(K)}} \left(-\log(|\det(\mathbf{W})|) + (1/2\lambda)\|\mathbf{W} - \widetilde{\mathbf{W}}\|_F^2\right), \end{aligned} \tag{16}$$

its closed-form solution for $K = 2$ can be used for an efficient update of the demixing matrix. We derived the closed-form solution to Eq. (16) based on the following observation.

Proposition 2. Let $\mathbf{W} \in \mathcal{W}_{\mathrm{PB}}^{(K)}$, then it follows that

$$\det(\mathbf{W}) = \det\begin{pmatrix} w_{22} & \cdots & w_{2K} \\ \vdots & \ddots & \vdots \\ w_{K2} & \cdots & w_{KK} \end{pmatrix}. \tag{17}$$

Proof. Using Eq. (10) and the elementary row transformations (adding all the rows to the 1st row), we obtain

$$\det(\mathbf{W}) = \det\begin{pmatrix} \underbrace{\sum_{k=1}^{K} w_{k1}}_{1} & \underbrace{\sum_{k=1}^{K} w_{k2}}_{0} & \cdots & \underbrace{\sum_{k=1}^{K} w_{kK}}_{0} \\ w_{21} & w_{22} & \cdots & w_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ w_{K1} & w_{K2} & \cdots & w_{KK}, \end{pmatrix}.$$

The Laplace expansion along the 1st row yields Eq. (17). ☐

Using this property, we derived a closed-form solution to the proximity operator in Eq. (16) when $K = 2$.

Proposition 3. Let $\widetilde{\mathbf{W}} \in \mathbb{C}^{2\times 2}$, then

$$\begin{aligned} &\mathrm{prox}_{\lambda\,\mathrm{LogDet}+\mathrm{PB}}(\widetilde{\mathbf{W}}) \\ &= \begin{pmatrix} \frac{1+(\tilde{w}_{11}-\tilde{w}_{21})}{2} & -\,\mathrm{prox}_{(-\lambda/2)\log(|\cdot|)}\left(\frac{\tilde{w}_{22}-\tilde{w}_{12}}{2}\right) \\ \frac{1-(\tilde{w}_{11}-\tilde{w}_{21})}{2} & \mathrm{prox}_{(-\lambda/2)\log(|\cdot|)}\left(\frac{\tilde{w}_{22}-\tilde{w}_{12}}{2}\right) \end{pmatrix} \end{aligned} \tag{18}$$

Algorithm 2 ADMM-BSS using the proximity operator in Eq. (18)

Input: $\mathbf{x}$, $(\mathbf{y}^{(\mathrm{PB})}, \mathbf{W}^{(\mathrm{PB})})$, $(\mathbf{u}_1, \mathbf{u}_2)$, ρ
Output: $\mathbf{y}^{(\mathrm{PB})}$, $\mathbf{W}^{(\mathrm{PB})}$

1: **for** $l = 1, \ldots,$ numIter **do**
2: $\quad \mathbf{y}^{(\mathrm{PB})} \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{P}}(\mathbf{W}\mathbf{x} + \mathbf{u}_1)$
3: $\quad \mathbf{W}^{(\mathrm{PB})} \leftarrow \mathrm{prox}_{(1/\rho)\mathrm{LogDet+PB}}(\mathbf{W} + \mathbf{u}_2)$
4: $\quad (\mathbf{u}_i)_{i=1}^2 \leftarrow (\mathbf{u}_i)_{i=1}^2 + (\mathbf{W}\mathbf{x}, \mathbf{W}) - (\mathbf{y}^{(\mathrm{PB})}, \mathbf{W}^{(\mathrm{PB})})$
5: $\quad \mathbf{W} \leftarrow \left((\mathbf{y}^{(\mathrm{PB})} - \mathbf{u}_1)\,\mathbf{x}^{\mathsf{H}} + (\mathbf{W}^{(\mathrm{PB})} - \mathbf{u}_2)\right)(\mathbf{x}\mathbf{x}^{\mathsf{H}} + \mathbf{I})^{-1}$
6: **end for**

holds, where the proximity operator in the matrix is given by

$$\mathrm{prox}_{-\lambda \log(|\cdot|)}(v) = \arg\min_{x \in \mathbb{C}} \left(-\log(|x|) + (1/2\lambda)|x - v|^2\right)$$
$$= (v/|v|)\left(|v| + \sqrt{|v|^2 + 4\lambda}\right)/2. \qquad (19)$$

Proof. Let $\mathbf{W} \in \mathcal{W}_{\mathrm{PB}}^{(2)}$, then $w_{21} = 1 - w_{11}$ and $w_{12} = -w_{22}$ from Eq. (10), and $\det(\mathbf{W}) = w_{22}$ from Prop. 2. Therefore, the objective function in Eq. (16) is reduced to

$$-\log(\underbrace{|w_{22}|}_{|\det(\mathbf{W})|}) + (1/2\lambda)\Big(|\underbrace{-w_{22}}_{w_{12}} - \tilde{w}_{12}|^2 + |w_{22} - \tilde{w}_{22}|^2$$
$$+ |w_{11} - \tilde{w}_{11}|^2 + |\underbrace{1 - w_{11}}_{w_{21}} - \tilde{w}_{21}|^2\Big)$$
$$= \underbrace{-\log(|w_{22}|) + (1/\lambda)\,|w_{22} - (\tilde{w}_{22} - \tilde{w}_{12})/2|^2}_{\text{Minimized at } w_{22}^{\star} = \mathrm{prox}_{(-\lambda/2)\log(|\cdot|)}((\tilde{w}_{22} - \tilde{w}_{12})/2)}$$
$$+ \underbrace{(1/\lambda)\,|w_{11} - (1 + (\tilde{w}_{11} - \tilde{w}_{21}))/2|^2}_{\text{Minimized at } w_{11}^{\star} = (1 + (\tilde{w}_{11} - \tilde{w}_{21}))/2} + C,$$

where C is a constant unrelated to the optimization variable. Using these minimizers $(w_{11}^{\star}, w_{22}^{\star})$ and the above equalities $w_{21} = 1 - w_{11}$ and $w_{12} = -w_{22}$, Eq. (18) is obtained, where $\mathrm{prox}_{-\lambda \log(\cdot)}$ can be found in, e.g., [27]. $\square$

Using this proximity operator, we obtain a simpler ADMM algorithm as in Alg. 2, where the log-det function and the PB constraint are handled simultaneously in the 3rd line.

IV. Experiment

To evaluate the operators and the algorithms we derived, we compared four ADMM-based BSS algorithms:

(a) Standard: ADMM algorithm for solving Eq. (4). It is obtained by replacing $\mathrm{prox}_{(1/\rho)\mathrm{LogDet+PB}}$ in Alg. 2 with $\mathrm{prox}_{(1/\rho)\mathrm{LogDet}}$.
(b) Heuristic: ADMM algorithm for solving Eq. (4), but the conventional PB is applied heuristically in each iteration. We calculated $\mathbf{D}[f]$ using Eqs. (7)–(9) and then applied it to all the variables.
(c) **(Ours) Metric**: Alg. 1 for solving Eq. (12).
(d) **(Ours) Joint**: Alg. 2 for solving Eq. (12). Whitening cannot be used in this case because it assumes $\mathbf{Q} = \mathbf{I}$.

We tested each algorithm with / without whitening in the pre-processing and with / without PB in the post-processing.

A total of 24 mixtures with two sources (i.e., $K = 2$) were generated using the SiSEC dev1 dataset [28]. The window size for STFT was 2048, and the hop size was 1024. For the source model, we used LGM in Eq. (5) with $\sigma_k[f,t] = |h_{1,k}[f]\, s_k[f,t]|^2$. Each algorithm was iterated 10 000 times with $\rho = 100$. The demixing matrix was initialized with $\mathbf{W} = \mathbf{I}$, and we initialized $\mathbf{y}^{(\mathrm{PB})} = \mathbf{W}\mathbf{x}$, $\mathbf{W}^{(\mathrm{PB})} = \mathbf{W}\mathbf{Q}$, $\check{\mathbf{W}} = \mathbf{W}$. The dual variables $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ were initialized with zeros. The separation performance was evaluated by the improvement in signal-to-distortion ratio (SDRi) [29].

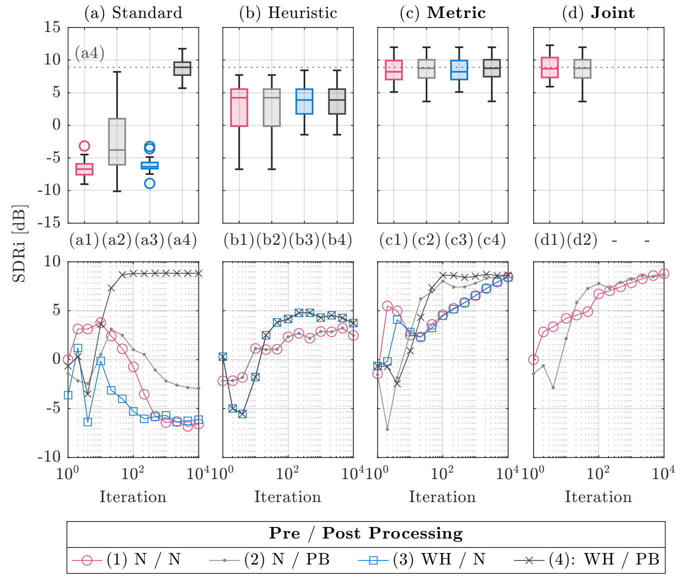

Fig. 2. Separation performance of each algorithm (top) and its transition (bottom) with/without whitening (WH) and projection back (PB) as pre/post-processing, where "N" indicates "without processing". Proposed methods (c) and (d) are emphasized with bold letters. The horizontal line represents the median value of the typical conventional method (a4), i.e., the standard algorithm with whitening and PB.

The experimental results are shown in Fig. 2. The standard method (a) achieved the best separation performance when both whitening and PB were applied (a4), i.e., when the three-stage procedure was used. The heuristic algorithm (b) obtained consistent results irrespective of the application of PB in the post-processing stage. However, they were unstable and the performance was not as good as that in the standard method (a4). For the proposed method, both Metric (c) and Joint (d) achieved a separation performance comparable to the standard method (a4) without requiring PB in the post-processing stage, which confirms that our algorithms can directly obtain the final output of BSS. Moreover, whitening is not required for the proposed algorithms, possibly because the PB constraint restricts the search space of the demixing matrix.

V. Conclusion

This paper proposed a noble formulation of determined BSS that incorporates the pre- and post-processing stages into the optimization problem, along with two types of optimization algorithms based on ADMM. Using the metric PB and the proximity operator of the log-det function under PB constraint, they directly obtained the separated signal after PB. Future work includes exploring advanced prior knowledge to further leveraging the advantages of directly handling the final output.

REFERENCES

[1] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, 2001.

[2] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Independent Component Analysis and Blind Signal Separation*, pp. 601–608. 2006.

[3] T. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: Definition and algorithms," in *Fortieth Asilomar Conf. Signals Syst. Comput.*, 2006, pp. 1393–1396.

[4] I. Lee and G.-J. Jang, "Independent vector analysis based on overlapped cliques of variable width for frequency-domain blind signal separation," *EURASIP J. Adv. Signal Process.*, vol. 2012, pp. 1–12, 2012.

[5] R. Ikeshita, Y. Kawaguchi, M. Togami, Y. Fujita, and K. Nagamatsu, "Independent vector analysis with frequency range division and prior switching," in *Eur. Signal Process. Conf. (EUSIPCO)*, 2017, pp. 2329–2333.

[6] K. Yatabe and D. Kitamura, "Determined BSS based on time-frequency masking and its application to harmonic vector analysis," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 1609–1625, 2021.

[7] K. Yatabe and D. Kitamura, "Time-frequency-masking-based determined BSS with application to sparse IVA," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 715–719.

[8] 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, 2016.

[9] R. Ikeshita and Y. Kawaguchi, "Independent low-rank matrix analysis based on multivariate complex exponential power distribution," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2018, pp. 741–745.

[10] D. Kitamura and K. Yatabe, "Consistent independent low-rank matrix analysis for determined blind source separation," *EURASIP J. Adv. Signal Process.*, vol. 2020, pp. 1–35, 2020.

[11] 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 Trans. Audio Speech Lang. Process.*, vol. 27, no. 10, pp. 1601–1615, 2019.

[12] T. Hasumi, T. Nakamura, N. Takamune, H. Saruwatari, D. Kitamura, Y. Takahashi, and K. Kondo, "Empirical Bayesian independent deeply learned matrix analysis for multichannel audio source separation," in *Eur. Signal Process. Conf. (EUSIPCO)*, 2021, pp. 331–335.

[13] T. Hasumi, T. Nakamura, N. Takamune, H. Saruwatari, D. Kitamura, Y. Takahashi, and K. Kondo, "PoP-IDLMA: Product-of-prior independent deeply learned matrix analysis for multichannel music source separation," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 2680–2694, 2023.

[14] H. Kameoka, L. Li, S. Inoue, and S. Makino, "Supervised determined source separation with multichannel variational autoencoder," *Neural Comput.*, vol. 31, no. 9, pp. 1891–1914, 2019.

[15] L. Li, H. Kameoka, and S. Makino, "Fast MVAE: Joint separation and classification of mixed sources based on multichannel variational autoencoder with auxiliary classifier," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2019, pp. 546–550.

[16] L. Li, H. Kameoka, and S. Makino, "FastMVAE2: On improving and accelerating the fast variational autoencoder-based source separation algorithm for determined mixtures," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 96–110, 2022.

[17] R. Scheibler and M. Togami, "Surrogate source model learning for determined source separation," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2021, pp. 176–180.

[18] K. Matsumoto and K. Yatabe, "Determined BSS by combination of IVA and DNN via proximal average," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2024, pp. 871–875.

[19] Y. Takahashi, D. Kitahara, K. Matsuura, and A. Hirabayashi, "Determined source separation using the sparsity of impulse responses," in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2020, pp. 686–690.

[20] 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 *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, 2018, pp. 746–750.

[21] Y. Kuriki, T. Nakashima, and N. Ono, "Direct update of back-projected demixing matrices in blind source separation," in *Eur. Signal Process. Conf. (EUSIPCO)*, 2024, pp. 922–926.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[23] 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 Inf. Process. (SIP)*, vol. 8, pp. e12, 2019.

[24] A. Ozerov and H. Kameoka, "Gaussian model based multichannel separation," in *Audio Source Separation and Speech Enhancement*, pp. 289–315. 2018.

[25] A. Ozerov, E. Vincent, and F. Bimbot, "A general flexible framework for the handling of prior information in audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 4, pp. 1118–1133, 2011.

[26] K. Matsumoto, K. Yamada, and K. Yatabe, "Constrained determined BSS realizing projection back through metric projection (in Japanese)," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 121–124, Sept. 2024.

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

[28] S. Araki, F. Nesta, E. Vincent, Z. Koldovskỳ, G. Nolte, A. Ziehe, and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011): - audio source separation -," in *Latent Var. Anal. Signal Sep.*, 2012, pp. 414–422.

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