ISS2: An Extension of Iterative Source Steering Algorithm for Majorization-Minimization-Based Independent Vector Analysis

Rintaro Ikeshita and Tomohiro Nakatani
NTT Corporation, Japan

***Abstract*—A majorization-minimization (MM) algorithm for independent vector analysis optimizes a separation matrix $W := [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m]^{\mathsf{H}} \in \mathbb{C}^{m \times m}$ by minimizing a surrogate function of the form $\mathcal{L}(W) := \sum_{i=1}^{m} \boldsymbol{w}_i^{\mathsf{H}} V_i \boldsymbol{w}_i - \log |\det W|^2$, where $m \in \mathbb{N}$ is the number of sensors and positive definite matrices $V_1, \ldots, V_m \in \mathbb{C}^{m \times m}$ are constructed in each MM iteration. For $m \geq 3$, no algorithm has been found to obtain a global minimum of $\mathcal{L}(W)$. Instead, block coordinate descent (BCD) methods with closed-form update formulas have been developed for minimizing $\mathcal{L}(W)$ and shown to be effective. One such BCD is called iterative projection (IP) that updates one or two rows of W in each iteration. Another BCD is called iterative source steering (ISS) that updates one column of the mixing matrix $A := W^{-1}$ in each iteration. Although the time complexity per iteration of ISS is m times smaller than that of IP, the conventional ISS converges slower than the current fastest IP (called IP$_2$) that updates two rows of W in each iteration. We here extend this ISS to ISS$_2$ that can update two columns of A in each iteration while maintaining its small time complexity. To this end, we provide a unified way for developing new ISS type methods from which ISS$_2$ as well as the conventional ISS can be immediately obtained in a systematic manner. Numerical experiments to separate reverberant speech mixtures show that our ISS$_2$ converges in fewer MM iterations than the conventional ISS, and is comparable to IP$_2$.**

***Index Terms*—independent component analysis (ICA), independent vector analysis (IVA), majorization-minimization (MM), block coordinate descent (BCD)**

I. Introduction

INDEPENDENT component analysis (ICA) [1] and its extension, independent vector analysis (IVA) [2], are fundamental blind source separation (BSS) methods that have been applied in numerous fields. Although theoretical properties such as identifiability of ICA [1, Chapter 4] [3] and IVA [4]–[6] have been well studied, the algorithms developed for them still need improvement because fast and stable optimization is indispensable when applied to real-world applications.

Early algorithms for ICA include Infomax [7] and the relative (or natural) gradient method [8], [9]. To accelerate these gradient-based algorithms using curvature information, several second-order algorithms with (relative) Hessian approximation were proposed [10]–[13]. In another research direction, a primal-dual splitting algorithm (e.g., [14]) for IVA [15] and its heuristic extension [16] based on the plug-and-play scheme were recently developed. However, all the above algorithms rely on good policies for determining hyperparameters such as step size, and it is usually difficult to find such policies that are suitable for any kind of signals. Other famous methods, such as FastICA [17] and its improvement [18], assume orthogonal constraint for the separated signals, which is not necessarily optimal, especially for short signals.

To avoid these problems, a majorization-minimization (MM) algorithm [19] for ICA [20], [21] and IVA [22] without such tuning parameters as the step size was proposed about a decade ago (see Section II-B) and has been studied extensively (mainly in the audio source separation community) because it can attain fast and stable optimization. Interestingly, a majorizer (or a surrogate function) constructed in the MM algorithm had already been studied in the ICA literature [23]–[25], not related to the MM approach.

Because the majorizer is non-convex and obtaining a global minimum is difficult, two families of block coordinate descent (BCD) methods [26] with closed-form update formulas were developed. One is called *iterative projection (IP)* [27]–[31] that updates one or two rows of the separation matrix $W \in \mathbb{C}^{m \times m}$ in each BCD iteration, where $m \in \mathbb{N}$ is the number of sensors. The other is called *iterative source steering (ISS)* [32] that updates one column of the mixing matrix $A := W^{-1}$ in each BCD iteration. Although ISS reduces the time complexity of IP by a factor of $1/m$, it requires more MM iterations to converge than the current fastest IP (called IP$_2$) that updates two rows of W in each iteration.

In this paper, we extend the conventional ISS so that it can update two columns of $A := W^{-1}$ in each iteration while keeping its small time complexity. The numerical simulation demonstrates the effectiveness of the proposed approach.

Notation: Let $\mathrm{GL}(m)$ be the set of all $m \times m$ nonsingular matrices over $\mathbb{C}$, and let $\mathcal{S}_+^m \subset \mathbb{C}^{m \times m}$ be the set of all Hermitian positive semidefinite matrices. For a matrix $A \in \mathbb{C}^{m \times n}$, let $A^\top$ and A^{H} denote the transpose and conjugate transpose of A, A_{ij} be the (i,j)th entry of A, $A_{i,\bullet} \in \mathbb{C}^{1 \times n}$ be the ith row of A, $A_{i:i+d,\bullet} := [A_{i,\bullet}^\top, \ldots, A_{i+d,\bullet}^\top]^\top \in \mathbb{C}^{(d+1) \times n}$, and $\mathrm{diag}(A_{i,\bullet}) \in \mathbb{C}^{n \times n}$ be the diagonal matrix whose diagonal entries are $A_{i,\bullet}$. The identity and zero matrices are denoted as $I_d \in \mathbb{C}^{d \times d}$ and $O_{i,j} \in \mathbb{C}^{i \times j}$, respectively.

II. Background

A. Independent Vector Analysis (IVA)

Consider a set of $K \geq 1$ linear mixtures:

$$X^{[k]} = A^{[k]} S^{[k]} \in \mathbb{C}^{m \times n}, \quad k = 1, \ldots, K, \tag{1}$$

where $m \in \mathbb{N}$ is the number of sensors, $n \in \mathbb{N}$ is the number of sample points, $X^{[k]} \in \mathbb{C}^{m \times n}$ is an observation, $S^{[k]} \in \mathbb{C}^{m \times n}$ is the original m source signals, and $A^{[k]} \in \mathrm{GL}(m)$ is called a *mixing matrix*. The goal of IVA is to estimate the set of the *separation matrices* $W^{[k]} \in \mathrm{GL}(m)$, $k = 1, \ldots, K$ satisfying

$$W^{[k]} A^{[k]} = D^{[k]} \Pi, \quad k = 1, \ldots, K, \tag{2}$$

where $D^{[k]}$ and Π are respectively the arbitrary diagonal and permutation matrices of size $m \times m$ that correspond to the *scale and permutation ambiguities* of separated signals $Y^{[k]} \coloneqq W^{[k]} X^{[k]}$. Note that permutation matrix Π must be independent of k to ensure that the orders of the K separated signals $Y^{[1]}, \ldots, Y^{[K]}$ are aligned between different mixtures.

To achieve the above, IVA relies on the assumption that, for each $i = 1, \ldots, m$ and $j = 1, \ldots, n$, the vector

$$\boldsymbol{y}_{ij} \coloneqq [\, Y_{ij}^{[1]}, \ldots, Y_{ij}^{[K]} \,]^\top \in \mathbb{C}^K \tag{3}$$

follows a probability density function with second or higher-order correlation [4]–[6]. Also, it is commonly assumed that the random variables $\{\boldsymbol{y}_{ij}\}_{ij}$ are mutually independent. Under this model, the negative log-likelihood, which yields a cost function of $\mathcal{W} \coloneqq (W^{[k]})_{k=1}^K$, is expressed as:

$$\begin{aligned} \mathcal{L}_0(\mathcal{W}) &\coloneqq -\frac{1}{n} \log p(X^{[1]}, \ldots, X^{[K]}; \mathcal{W}) \\ &= -\frac{1}{n} \sum_{i=1}^m \sum_{j=1}^n \log p(\boldsymbol{y}_{ij}) - \sum_{k=1}^K \log |\det W^{[k]}|^2. \end{aligned} \tag{4}$$

B. Majorization-Minimization Algorithm for IVA

An MM algorithm for ICA was proposed by (Ono and Miyabe, 2010 [20]), rediscovered by (Ablin, Gramfort, Cardoso, and Bach, 2019 [21]), and extended for IVA by (Ono, 2011 [22]). Here we briefly review it.

Let $p(\boldsymbol{y})$ be a circularly-symmetric probability density function of a random variable $\boldsymbol{y}$, and a function $\varphi \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ be given by $\varphi(\|\boldsymbol{y}\|_2) \coloneqq -\log p(\boldsymbol{y})$ with $\|\boldsymbol{y}\|_2 \coloneqq \sqrt{\boldsymbol{y}^\mathsf{H} \boldsymbol{y}}$. We say that $p(\boldsymbol{y})$ is *super-Gaussian* if $\varphi'(r)/r$ is decreasing on $r \in (0, \infty) = \mathbb{R}_{>0}$, where φ' is the first derivative of φ (see, e.g., [20], [33], and [34, pp. 60–61]). For instance, a generalized Gaussian distribution (GGD)

$$\varphi(\|\boldsymbol{y}\|_2) = \|\boldsymbol{y}\|_2^\beta + \text{const.}, \quad 0 < \beta < 2 \tag{5}$$

is super-Gaussian. GGD with $\beta = 1$ is nothing but the Laplace distribution. For a super-Gaussian $\varphi(r)$, we have (see [20])

$$\varphi(r) = \min_{\alpha > 0} \Big[\frac{\varphi'(\alpha)}{2\alpha} \cdot r^2 + \Big(\varphi(\alpha) - \frac{\alpha \varphi'(\alpha)}{2} \Big) \Big] \tag{6}$$

for all $r \in \mathbb{R}_{>0}$ and its minimum is attained at $\alpha = r$. Using (6) for each $\varphi(\|\boldsymbol{y}_{ij}\|_2) \coloneqq -\log p(\boldsymbol{y}_{ij})$ in (4), we can develop an MM algorithm for IVA [22] that alternately updates an auxiliary variable $\Lambda \in \mathbb{R}_{\geq 0}^{m \times n}$ and $\mathcal{W}$ by repeating

$$\Lambda_{ij} \leftarrow \frac{\varphi'(\|\boldsymbol{y}_{ij}\|_2)}{\|\boldsymbol{y}_{ij}\|_2}, \quad i = 1, \ldots, m; \; j = 1, \ldots, n, \tag{7}$$

$$W^{[k]} \in \operatorname*{argmin}_{W^{[k]} \in \mathrm{GL}(m)} \mathcal{L}^{[k]}(W^{[k]}, \Lambda), \quad k = 1, \ldots, K, \tag{8}$$

where we define

$$\mathcal{L}^{[k]}(W^{[k]}, \Lambda) = \sum_{i=1}^m (\boldsymbol{w}_i^{[k]})^\mathsf{H} V_i^{[k]} \boldsymbol{w}_i^{[k]} - \log |\det W^{[k]}|^2, \tag{9}$$

$$V_i^{[k]} = \frac{1}{2n} X^{[k]} \operatorname{diag}(\Lambda_{i,\bullet}) \, (X^{[k]})^\mathsf{H} \in \mathcal{S}_+^m. \tag{10}$$

$$\boldsymbol{w}_i^{[k]} = (W_{i,\bullet}^{[k]})^\mathsf{H} \quad (\Leftrightarrow W^{[k]} = [\boldsymbol{w}_1^{[k]}, \ldots, \boldsymbol{w}_m^{[k]}]^\mathsf{H}), \tag{11}$$

Note that $\Lambda_{i,\bullet} \in \mathbb{R}_{\geq 0}^{1 \times n}$ in (10) is the ith row of Λ.

When $m = 2$ and $W^{[k]} \in \mathbb{C}^{2 \times 2}$, problem (8) has a closed-form solution [24], [35]. However, for $m \geq 3$, no algorithm has been found that obtains a global minimum of (8), and several BCD algorithms were developed. In this paper, we refer to such MM-based IVA approaches as *MM+BCD*.

Hereafter, for ease of notation, we omit the upper right index $\cdot^{[k]}$ when discussing (8)–(11) and simply denote the objective function $\mathcal{L}^{[k]}(W^{[k]}, \Lambda)$ as $\mathcal{L}(W)$.

III. Proposed MM+BCD Algorithm

We generalize the definition of iterative source steering (ISS) to be a family of MM+BCD algorithms that update several columns of $A \coloneqq W^{-1}$ in each iteration based on the minimization of $\mathcal{L}(W)$ with respect to those columns. The conventional ISS [32] (called ISS_1) updates one column of A in each iteration. We extend this ISS_1 to ISS_2 so that it can update two columns of A in each iteration. To this end, we newly provide a unified way to develop ISS_d for any $d \geq 1$.

A. Definition of ISS_d

Let d be a divisor of m and $L \coloneqq m/d$. Consider the partition of A into L submatrices with d columns:

$$A = [\, \underbrace{A^{(1)}}_{d} \mid \cdots \mid \underbrace{A^{(L)}}_{d} \,] \in \mathbb{C}^{m \times m}. \tag{12}$$

ISS_d is an MM+BCD method that cyclically updates

$$\Lambda \to (W, A^{(1)}) \to (W, A^{(2)}) \to \cdots \to (W, A^{(L)}) \tag{13}$$

one by one based on (7) for updating Λ and

$$(W, A^{(\ell)}) \in \operatorname*{argmin}_{(W,\, A^{(\ell)})} \{ \mathcal{L}(W) \mid WA = I_m \} \tag{14}$$

for updating $(W, A^{(\ell)})$ with $\ell = 1, \ldots, L$. When $d = 1$, our definition of ISS_1 coincides with the conventional ISS_1 [32].

B. Multiplicative update (MU) formulation for ISS_d

We show that ISS_d can be written as a multiplicative update (MU) algorithm for W (or equivalently $Y = WX$). To begin with, we provide the following proposition.

Proposition 1. Update rule (14) with $\ell = 1$ is equivalent to the following MU rule for W (and A):

$$T \in \operatorname*{argmin}_{T} \{ \mathcal{L}(TW) \mid T \in \mathcal{D}_{\mathrm{ISS}_d} \}, \tag{15}$$

$$W \leftarrow TW \quad (\text{and } A \leftarrow AT^{-1}), \tag{16}$$

where we define

$$\mathcal{D}_{\mathrm{ISS}_d} \coloneqq \left\{ \begin{bmatrix} P & O_{d,m-d} \\ Q & I_{m-d} \end{bmatrix} \;\middle|\; P \in \mathrm{GL}(d),\, Q \in \mathbb{C}^{(m-d) \times d} \right\}.$$

Proof. By the update of $A^{\text{new}} \leftarrow AT^{-1}$ with $T^{-1} \in \mathcal{D}_{\text{ISS}_d}$, $A^{(1)}$ can take an arbitrary value while $[A^{(2)}, \ldots, A^{(L)}]$ remains unchanged. To keep the constraint $WA = I_m$ in (14), W must be uniquely updated to $W^{\text{new}} \leftarrow TW$:

$$\underbrace{\begin{bmatrix} P^{-1} & O_{d,m-d} \\ -QP^{-1} & I_{m-d} \end{bmatrix} W}_{W^{\text{new}} = TW} \underbrace{A \begin{bmatrix} P & O_{d,m-d} \\ Q & I_{m-d} \end{bmatrix}}_{A^{\text{new}} = AT^{-1}} = I_m.$$

(Note that the set $\mathcal{D}_{\text{ISS}_d}$ is closed under matrix inversion.) This T belongs to and runs over $\mathcal{D}_{\text{ISS}_d}$ when T^{-1} runs over $\mathcal{D}_{\text{ISS}_d}$. Thus, Eq. (14) with $\ell = 1$ is equivalent to (15)–(16). ☐

We next show that Eq. (14) with $\ell \in \{2, \ldots, L\}$ can also be rewritten in the same way as (15)–(16) by properly permuting the rows of W and the columns of A in advance. To see this, let us define a (block) permutation matrix

$$\Pi_d = \left[\begin{array}{c|ccc} & I_d & & \\ & & \ddots & \\ & & & I_d \\ \hline I_d & & & \end{array}\right] \in \mathbb{C}^{m \times m} \tag{17}$$

and permute the rows of (W, Y, Λ) and columns of A by

$$W \leftarrow \Pi_d^{\ell-1} W, \quad Y \leftarrow \Pi_d^{\ell-1} Y, \quad \Lambda \leftarrow \Pi_d^{\ell-1} \Lambda,$$
$$A \leftarrow A(\Pi_d^{\ell-1})^\top = [\, A^{(\ell)}, \ldots, A^{(L)}, A^{(1)}, \ldots, A^{(\ell-1)} \,].$$

This permutation keeps both the objective function and constraint $WA = I_m$ in (14) since $\Pi_d \Pi_d^\top = I_m$. Also, the first d columns of $A(\Pi_d^{\ell-1})^\top$ are $A^{(\ell)}$. Thus, Eq. (14) with $\ell \geq 2$ is also essentially equivalent to (15)–(16). Due to this observation, we only need to address problem (15) below.

C. Derivation of ISS$_2$ (and new derivation of ISS$_1$)

We discuss problem (15) for general $d \geq 1$ and develop a closed-form solution for it when $d = 2$ (proposed ISS$_2$) and $d = 1$ (new derivation of ISS$_1$).

For $T \in \mathcal{D}_{\text{ISS}_d}$, the ith row vector of P (resp. Q) is denoted as $\boldsymbol{p}_i^{\mathsf{H}} \in \mathbb{C}^{1 \times d}$ (resp. $\boldsymbol{q}_{d+i}^{\mathsf{H}} \in \mathbb{C}^{1 \times d}$):

$$P = [\boldsymbol{p}_1, \ldots, \boldsymbol{p}_d]^{\mathsf{H}} \in \mathbb{C}^{d \times d}, \tag{18}$$
$$Q = [\boldsymbol{q}_{d+1}, \ldots, \boldsymbol{q}_m]^{\mathsf{H}} \in \mathbb{C}^{(m-d) \times d}. \tag{19}$$

Then the objective function $\mathcal{L}(TW)$ can be expressed as

$$\begin{aligned} \mathcal{L}(TW) = &\sum_{i=1}^{d} \boldsymbol{p}_i^{\mathsf{H}} G_i \boldsymbol{p}_i - \log |\det P|^2 \\ &+ \sum_{i=d+1}^{m} \begin{bmatrix} \boldsymbol{q}_i \\ 1 \end{bmatrix}^{\mathsf{H}} \begin{bmatrix} G_i & \boldsymbol{g}_i \\ \boldsymbol{g}_i^{\mathsf{H}} & c_i \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_i \\ 1 \end{bmatrix} + \text{const.}, \end{aligned} \tag{20}$$

where for each $i = 1, \ldots, m$,

$$G_i = W_{1:d,\bullet} V_i W_{1:d,\bullet}^{\mathsf{H}} = \frac{1}{2n} Y_{1:d,\bullet} \operatorname{diag}(\Lambda_{i,\bullet}) \, (Y_{1:d,\bullet})^{\mathsf{H}} \in \mathbb{C}^{d \times d},$$
$$\boldsymbol{g}_i = W_{1:d,\bullet} V_i W_{i,\bullet}^{\mathsf{H}} = \frac{1}{2n} Y_{1:d,\bullet} \operatorname{diag}(\Lambda_{i,\bullet}) \, (Y_{i,\bullet})^{\mathsf{H}} \in \mathbb{C}^{d \times 1},$$

and $c_i \in \mathbb{C}$ is constant. Since the variables P and Q are split in (20), we can optimize them separately.

1) Optimization of Q for general $d \geq 1$: Since the objective function $\mathcal{L}(TW)$ is quadratic with respect to Q, and $G_{d+1}, \ldots, G_m$ are positive definite in general (if $d \leq n$), the global optimal solution for Q is obtained as

$$\boldsymbol{q}_i = -G_i^{-1} \boldsymbol{g}_i \in \mathbb{C}^{d \times 1}, \quad i = d+1, \ldots, m. \tag{21}$$

Along with this, the separated signals are updated as

$$Y_{i,\bullet} \leftarrow Y_{i,\bullet} + \boldsymbol{q}_i^{\mathsf{H}} Y_{1:d,\bullet} = Y_{i,\bullet} - \boldsymbol{g}_i^{\mathsf{H}} G_i^{-1} Y_{1:d,\bullet} \in \mathbb{C}^{1 \times n} \tag{22}$$

for each $i = d+1, \ldots, m$. Note that in ISS$_d$ we only need to update (Y, Λ) but not W since the surrogate function $\mathcal{L}(TW)$ given by (20) can be constructed from (Y, Λ) only.

2) Optimization of P for $d \geq 3$: We want to solve

$$P \in \operatorname*{argmin}_{P \in \mathrm{GL}(d)} \sum_{i=1}^{d} \boldsymbol{p}_i^{\mathsf{H}} G_i \boldsymbol{p}_i - \log |\det P|^2. \tag{23}$$

However, as mentioned in Section II-B, obtaining a global minimum of (23) for $d \geq 3$ is a long-standing open problem [24], and we leave this task for future work.

3) Optimization of P for $d = 2$ (ISS$_2$ case): When $d = 2$, problem (23) is known to have a closed-form solution [35]:

$$H = G_1^{-1} G_2 \in \mathbb{C}^{2 \times 2}, \tag{24}$$
$$\theta_1 = \frac{\operatorname{Tr}(H) + \sqrt{(\operatorname{Tr}(H))^2 - 4 \det(H)}}{2}, \quad \theta_2 = \frac{\det H}{\theta_1},$$
$$\boldsymbol{u}_1 = \begin{bmatrix} H_{22} - \theta_1 \\ -H_{21} \end{bmatrix}, \quad \boldsymbol{u}_2 = \begin{bmatrix} -H_{12} \\ H_{11} - \theta_2 \end{bmatrix} \in \mathbb{C}^{2 \times 1}, \tag{25}$$
$$\boldsymbol{p}_i = \frac{\boldsymbol{u}_i}{(\boldsymbol{u}_i^{\mathsf{H}} G_i \boldsymbol{u}_i)^{\frac{1}{2}}} \in \mathbb{C}^{2 \times 1}, \quad i = 1, 2. \tag{26}$$

Along with this, the separated signals are updated by $Y_{1:2,\bullet} \leftarrow P Y_{1:2,\bullet}$. The proposed ISS$_2$ is summarized in Algorithm 1.

4) Optimization of P for $d = 1$ (ISS$_1$ case): When $d = 1$, $p_1 = G_1^{-\frac{1}{2}}$ gives a global minimum of (23). The obtained ISS$_1$ is identical to the conventional ISS$_1$ [32]. Our new derivation has an advantage of providing a systematic way to discuss ISS$_d$, which enabled us to generalize ISS$_1$ to ISS$_2$ as above.

IV. Relation to Prior MM+BCD Algorithms

Iterative projection (IP) is a family of MM+BCD algorithms that optimize several rows of W in each iteration with closed-form update formulas. So far, IP$_1$ [20]–[22] and IP$_2$ [27] (see also [28]–[31]) have been developed as members of IP.

IP$_1$ is an MM+BCD that updates $\Lambda \to \boldsymbol{w}_1 \to \cdots \to \boldsymbol{w}_m$ one by one. The update rule for Λ is given by (7) and that for $\boldsymbol{w}_\ell \coloneqq W_{\ell,\bullet}^{\mathsf{H}}$ can be developed as

$$\boldsymbol{u}_\ell \leftarrow (W V_\ell)^{-1} \boldsymbol{e}_\ell \in \mathbb{C}^{m \times 1}, \quad \boldsymbol{w}_\ell \leftarrow \frac{\boldsymbol{u}_\ell}{(\boldsymbol{u}_\ell^{\mathsf{H}} V_\ell \boldsymbol{u}_\ell)^{\frac{1}{2}}} \in \mathbb{C}^{m \times 1},$$

where V_ℓ is defined by (10) and $\boldsymbol{e}_\ell$ is the ℓ-th column of I_m.

IP$_2$ is an MM+BCD that updates $\Lambda \to [\boldsymbol{w}_1, \boldsymbol{w}_2] \to \cdots \to [\boldsymbol{w}_{m-1}, \boldsymbol{w}_m]$ one by one (when m is even), which improves IP$_1$ (see, e.g., [28]–[31] for details).

Recently, an advanced algorithm called *iterative projection with adjustment (IPA)* was proposed [30]. However, unlike IP

Algorithm 1: IVA by ISS_2

Input: $X^{[k]} \in \mathbb{C}^{m \times n}$ $(k = 1, \ldots, K)$

Initialize $W^{[k]}$ as a whitening matrix for $k = 1, \ldots, K$.

$Y^{[k]} \leftarrow W^{[k]} X^{[k]}$ for each $k = 1, \ldots, K$.

repeat `// outer MM loop`

  $\Lambda_{ij} \leftarrow \varphi'(\|\boldsymbol{y}_{ij}\|_2 + \varepsilon) \,/\, (\|\boldsymbol{y}_{ij}\|_2 + \varepsilon)$, where $\varepsilon = 10^{-10}$ is added to improve numerical stability.

  for $\ell = 1, \ldots, \frac{m}{2}$ **do** `// inner BCD loop`

    for $k = 1, \ldots, K$ **do**

      `/* Update` $Y^{[k]}_{3:m,\bullet} \in \mathbb{C}^{(m-2) \times n}$ `*/`

      for $i = 3, \ldots, m$ **do**

        $G^{[k]}_i = \frac{1}{2n} Y^{[k]}_{1:2,\bullet} \operatorname{diag}(\Lambda_{i,\bullet}) \, (Y^{[k]}_{1:2,\bullet})^{\mathsf{H}}$

        $\boldsymbol{g}^{[k]}_i = \frac{1}{2n} Y^{[k]}_{1:2,\bullet} \operatorname{diag}(\Lambda_{i,\bullet}) \, (Y^{[k]}_{i,\bullet})^{\mathsf{H}}$

        $Y^{[k]}_{i,\bullet} \leftarrow Y^{[k]}_{i,\bullet} - (\boldsymbol{g}^{[k]}_i)^{\mathsf{H}} (G^{[k]}_i)^{-1} Y^{[k]}_{1:2,\bullet}$

      `/* Update` $Y^{[k]}_{1:2,\bullet} \in \mathbb{C}^{2 \times n}$ `*/`

      for $i = 1, 2$ **do**

        $G^{[k]}_i = \frac{1}{2n} Y^{[k]}_{1:2,\bullet} \operatorname{diag}(\Lambda_{i,\bullet}) \, (Y^{[k]}_{1:2,\bullet})^{\mathsf{H}}$

      Update $P^{[k]} \in \mathbb{C}^{2 \times 2}$ using (24)–(26).

      $Y^{[k]}_{1:2,\bullet} \leftarrow P^{[k]} Y^{[k]}_{1:2,\bullet} \in \mathbb{C}^{2 \times n}$

    `/* Permute rows */`

    $\Lambda \leftarrow \Pi_2 \Lambda$, where Π_2 is defined as (17).

    $Y^{[k]} \leftarrow \Pi_2 Y^{[k]}$ for $k = 1, \ldots, K$.

until some convergence criterion is met

Output: $Y^{[k]} \in \mathbb{C}^{m \times n}$ $(k = 1, \ldots, K)$

and ISS, no (fully) closed-form update formula has existed for IPA, because it requires a root-finding algorithm (and for this purpose the Newton-Raphson method is used [30]). Although IPA is important, we will not compare it with IP and ISS in our experiments, since we are focusing on such methods with fully closed-form update formulas.

V. Time Complexity Analysis

The computational time complexity of ISS_2 per MM iteration is dominated by

- the computation of $(G^{[k]}_i, \boldsymbol{g}^{[k]}_i)$ for each $i = 1, \ldots, m$ and loop $\ell = 1, \ldots, \frac{m}{2}$, which costs $\mathrm{O}(Km^2 n)$; and
- the computation of $Y^{[k]}$, which costs $\mathrm{O}(Km^2 n)$.

Thus, ISS_2 has the time complexity of $\mathrm{O}(Km^2 n)$, which is the same as that of ISS_1. For comparison, the time complexity of IP_1 with $d \in \{1, 2\}$ per iteration is dominated by (e.g., [30])

- the computation of covariance matrices $V^{[k]}_1, \ldots, V^{[k]}_m \in \mathcal{S}^m_+$, which costs $\mathrm{O}(Km^3 n)$; and
- the computation of updating $W^{[k]}$, which costs $\mathrm{O}(Km^4)$.

Thus, IP_d ($d \in \{1, 2\}$) has the time complexity of $\mathrm{O}(Km^3 n + Km^4)$, which is m times larger than ISS_d with $d \in \{1, 2\}$.

VI. Experiments

We compared the performance of our proposed ISS_2 and conventional ISS_1, IP_1, and IP_2 when applied to convolutive blind source separation (BSS) in the short-time Fourier transform (STFT) domain [36], where K and n correspond to the numbers of frequency bins and time frames, respectively. This setting is very common in audio source separation [36].

Dataset: We generated synthesized convolutive mixtures of $m \in \{4, 6, 8, 10\}$ speech signals. The signals were captured by a circular array with m microphones and a radius of 5 cm. We obtained speech signals from the TIMIT corpus [37] and concatenated them so that the signal length exceeded 10 seconds. The obtained signals were normalized to have unit power. To obtain acoustic impulse responses (AIR), we used the `pyroomacoustics` Python package [38] and simulated 100 rectangular rooms. The rooms were 5 to 8 m wide and 3 to 5 m high. The arrays were placed in the center of the rooms at a height of 1 m. The speech sources were randomly placed in the room at a height of 1 m, provided that the distances from the array center and the walls were at least 1 m. The reverberation times (T_{60}) ranged from 250 to 400 ms.

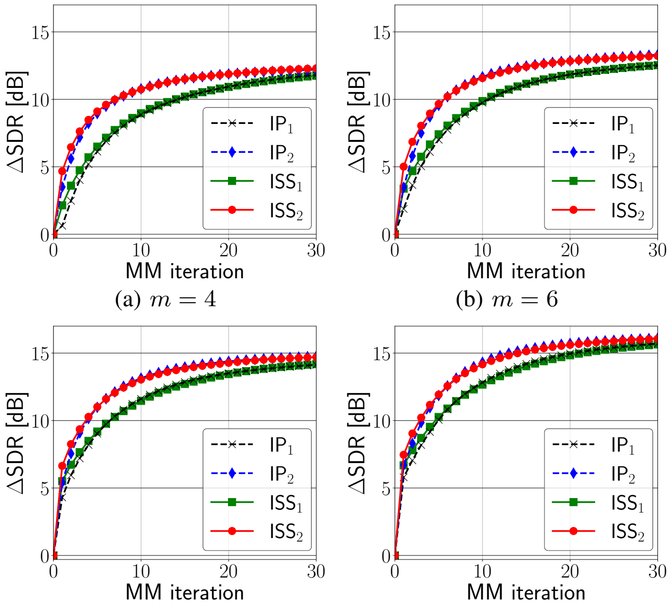
Evaluation criterion: We measured the signal-to-distortion ratio (SDR) [39] between separated signal $\hat{\boldsymbol{s}}$ and oracle reverberant speech signal $\boldsymbol{s}$ at the first microphone. The SDR we used here is sometimes called the scale-invariant SDR [40] and defined as $\mathrm{SDR~[dB]} = 10 \log_{10} \frac{\|\alpha \boldsymbol{s}\|_2^2}{\|\hat{\boldsymbol{s}} - \alpha \boldsymbol{s}\|_2^2}$ with $\alpha = \frac{\hat{\boldsymbol{s}}^\top \boldsymbol{s}}{\|\boldsymbol{s}\|_2^2}$.

Other conditions: The sampling rate was 16 kHz, the STFT frame size was 4096 (256 ms), and the frame shift was 1024 (64 ms). We assumed a Laplace distribution, i.e., (5) with $\beta = 1$, for the separated signals. We initialized $W^{[k]}$ as the whitening matrix $D^{-1/2} U^{\mathsf{H}}$ using the eigenvalue decomposition $UDU^{\mathsf{H}} = \frac{1}{n} X^{[k]} (X^{[k]})^{\mathsf{H}}$ for each $k = 1, \ldots, K$. After separation, the scale ambiguity of IVA, i.e., (2), was restored based on the minimum distortion principle (MDP) [41] (see also [42, Section 2.2] for the details of MDP).

Experimental results: Figure 1 shows the SDR improvement obtained by each method. As we desired, the convergence of the proposed ISS_2 is much faster than ISS_1 and IP_1 and comparable to IP_2 (note that the SDR curves of IP_2 and ISS_2 almost overlap), which clearly shows the effectiveness of our approach. Since the time complexity of ISS_2 is m times smaller than IP_2, one might expect that the runtime of ISS_2 to reach convergence is shorter than that of IP_2; but this was not the case in our experiment with our Python implementation where the runtime of ISS_2 was slightly inferior to that of IP_2. This implementation issue is an important future work.

VII. Conclusion

As BCD algorithms for the MM-based IVA, IP_1, IP_2, and ISS_1 had been developed. We here extended ISS_1 to ISS_2 that updates two columns of the mixing matrix $A \coloneqq W^{-1}$ in each BCD iteration. Our ISS_2 simultaneously achieves both (i) the small time complexity of ISS_1 per MM iteration and (ii) the fast convergence behavior of IP_2, which was confirmed by the numerical experiments.



(a) $m = 4$ (b) $m = 6$ (c) $m = 8$ (d) $m = 10$

Fig. 1: The SDR improvement (ΔSDR) from the initial SDR as a function of the MM iteration. The SDRs were averaged over 100 samples. The average signal length was 13.3 sec.

References

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.

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

[3] B. Afsari, "Sensitivity analysis for the problem of matrix joint diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1148–1171, 2008.

[4] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, 2014.

[5] D. Lahat and C. Jutten, "Joint independent subspace analysis: Uniqueness and identifiability," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 684–699, 2018.

[6] ——, "An alternative proof for the identifiability of independent vector analysis using second order statistics," in *Proc. ICASSP*, 2016, pp. 4363–4367.

[7] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[8] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, 1996.

[9] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Proc. NIPS*, 1996, pp. 757–763.

[10] M. Zibulevsky, "Blind source separation with relative newton method," in *Proc. ICA*, 2003, pp. 897–902.

[11] J. A. Palmer, S. Makeig, K. Kreutz-Delgado, and B. D. Rao, "Newton method for the ICA mixture model," in *Proc. ICASSP*, 2008, pp. 1805–1808.

[12] H. Choi and S. Choi, "A relative trust-region algorithm for independent component analysis," *Neurocomputing*, vol. 70, no. 7-9, pp. 1502–1510, 2007.

[13] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster independent component analysis by preconditioning with Hessian approximations," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4040–4049, 2018.

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

[15] K. Yatabe and D. Kitamura, "Determined blind source separation via proximal splitting algorithm," in *Proc. ICASSP*, 2018, pp. 776–780.

[16] ——, "Determined BSS based on time-frequency masking and its application to harmonic vector analysis," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 1609–1625, 2021.

[17] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.

[18] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster ICA under orthogonal constraint," in *Proc. ICASSP*, 2018, pp. 4464–4468.

[19] K. Lange, *MM optimization algorithms*. SIAM, 2016.

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

[21] P. Ablin, A. Gramfort, J.-F. Cardoso, and F. Bach, "Stochastic algorithms with descent guarantees for ICA," in *Proc. AISTATS*, 2019, pp. 1564–1573.

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

[23] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, 2001.

[24] S. Dégerine and A. Zaïdi, "Determinant maximization of a nonsymmetric matrix with quadratic constraints," *SIAM J. Optim.*, vol. 17, no. 4, pp. 997–1014, 2007.

[25] A. Yeredor, B. Song, F. Roemer, and M. Haardt, "A "sequentially drilled" joint congruence (SeDJoCo) transformation with applications in blind source separation and multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2744–2757, 2012.

[26] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[27] N. Ono, "Fast algorithm for independent component/vector/low-rank matrix analysis with three or more sources," in *Proc. ASJ Spring Meeting*, 2018, (in Japanese).

[28] T. Nakashima, R. Scheibler, Y. Wakabayashi, and N. Ono, "Faster independent low-rank matrix analysis with pairwise updates of demixing vectors," in *Proc. EUSIPCO*, 2021, pp. 301–305.

[29] R. Scheibler and N. Ono, "MM algorithms for joint independent subspace analysis with application to blind single and multi-source extraction," *arXiv:2004.03926v1*, 2020.

[30] R. Scheibler, "Independent vector analysis via log-quadratically penalized quadratic minimization," *IEEE Trans. Signal Process.*, vol. 69, pp. 2509–2524, 2021.

[31] R. Ikeshita, T. Nakatani, and S. Araki, "Block coordinate descent algorithms for auxiliary-function-based independent vector extraction," *IEEE Trans. Signal Process.*, vol. 69, pp. 3252–3267, 2021.

[32] R. Scheibler and N. Ono, "Fast and stable blind source separation with rank-1 updates," in *Proc. ICASSP*, 2020, pp. 236–240.

[33] J. Palmer, D. Wipf, K. Kreutz-Delgado, and B. Rao, "Variational EM algorithms for non-Gaussian latent variable models," in *Proc. NIPS*, vol. 18, 2005, pp. 1059–1066.

[34] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*, 1st ed. Springer Science, 1990, vol. 22.

[35] N. Ono, "Fast stereo independent vector analysis and its implementation on mobile phone," in *Proc. IWAENC*, 2012, pp. 1–4.

[36] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*. John Wiley & Sons, 2018.

[37] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "TIMIT Acoustic-Phonetic Continuous Speech Corpus LDC93S1," Web Download. Philadelphia: Linguistic Data Consortium, Tech. Rep., 1993.

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

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

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

[41] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. ICA*, 2001, pp. 722–727.

[42] R. Scheibler, "Generalized minimal distortion principle for blind source separation," in *Proc. Interspeech*, 2020, pp. 3326–3330.