OVERDETERMINED INDEPENDENT VECTOR ANALYSIS

Rintaro Ikeshita, Tomohiro Nakatani, Shoko Araki

NTT Communication Science Laboratories, NTT Corporation, Kyoto, Japan

ABSTRACT

We address the convolutive blind source separation problem for the (over-)determined case where (i) the number of nonstationary target-sources \( K \) is less than that of microphones \( M \), and (ii) there are up to \( M - K \) stationary Gaussian noises that need not to be extracted. Independent vector analysis (IVA) can solve the problem by separating into \( M \) sources and selecting the top \( K \) highly nonstationary signals among them, but this approach suffers from a waste of computation especially when \( K \ll M \). Channel reductions in preprocessing of IVA by, e.g., principle component analysis have the risk of removing the target signals. We here extend IVA to resolve these issues. One such extension has been attained by assuming the orthogonality constraint (OC) that the sample correlation between the target and noise signals is to be zero. The proposed IVA, on the other hand, does not rely on OC and exploits only the independence between sources and the stationarity of the noises. This enables us to develop several efficient algorithms based on block coordinate descent methods with a problem specific acceleration. We clarify that one such algorithm exactly coincides with the conventional IVA with OC, and also explain that the other newly developed algorithms are faster than it. Experimental results show the improved computational load of the new algorithms compared to the conventional methods. In particular, a new algorithm specialized for \( K = 1 \) outperforms the others.

Index Terms— Blind source separation, over-determined, independent vector analysis, block coordinate descent method, generalized eigenvalue problem

1. INTRODUCTION

Blind source separation (BSS) is a problem of estimating the source signals from their observed mixtures [1,2]. In this paper, we focus on the (over-)determined BSS where the number of nonstationary signals \( K \) is less than that of microphones \( M \), i.e., \( K < M \). There can be up to \( M - K \) stationary Gaussian noises as long as the problem remains (over-)determined.\(^1\) The goal is to extract \( K \) nonstationary signals efficiently. We do not care about separating the noises.

If the mixture is convolutive, \textit{independent vector analysis} (IVA [3,4]) is one of the most fundamental methods to solve BSS. As a straightforward way, we can apply IVA as if there are \( M \) nonstationary sources and select the top \( K (< M) \) highly nonstationary signals among the \( M \) separated signals. This method is, however, computationally intensive and does not run in real time if \( M \) is large.

To improve the computational efficiency, in preprocessing of IVA, we can reduce the number of channels up to \( K \) by using principle component analysis or picking up \( K \) channels with high SNR. These channel reductions, however, have the risk of removing the target signals and often degrade the separation performance [5].

\(^1\)The assumption that there are at most \( M - K \) noises is for the sake of developing efficient algorithms rigorously and can be violated to some extent (see Section 6).

BSS methods for efficiently extracting just one or several target signals with specific properties such as non-Gaussianity has already been studied [6–11]. Among them, we only focus on methods in which the demixing filters can be optimized by using an \textit{iterative projection} (IP) [12–14] technique. IP is a class of \textit{block coordinate descent} (BCD) methods specialized for optimizing the maximum-likelihood-based IVA, with the advantages of high computational efficiency and no hyperparameters. Recently proposed \textit{OverIVA} [5] (IVA for the over-determined case), which is a multi-target-source extension of independent component/vector extraction (ICE/IVE [15, 16]), takes the advantages of IP, resulting in significantly improved computational cost of IVA. In addition to IP, OverIVA as well as ICE/IVE also relies on the orthogonality constraint (OC) that the sample correlation between the separated target and noise signals is to be zero. As OC is a heuristic assumption for developing efficient algorithms, the open problems here are (i) to theoretically clarify the validity of OC, and (ii) to further accelerate OverIVA.

In this paper, we propose a different approach for BSS, which we call \textit{OverIVA} as well. It does not employ OC and exploits only the independence between sources and the stationarity of the Gaussian noises (see Section 3). For this model, we develop several efficient optimization algorithms based on BCD, all of which can be viewed as extensions of IP (see Section 4). We then prove that one such algorithm exactly coincides with OverIVA with OC [5], meaning that the stationarity of the Gaussian noises implicitly implies OC (see Section 5). We also describe that the other newly developed algorithms are expected to be faster than OverIVA with OC (see Section 5), and confirm their effectiveness in the experiment (see Section 6).

2. PROBLEM FORMULATION

Suppose that \( K \) nonstationary source signals are mixed with a stationary noise, or possibly a silent signal, of dimension \( M - K \) and observed with \( M \) microphones, where \( 1 \leq K \leq M - 1 \). In the time-frequency domain, the mixture signal, \( X(f, t) \), is modeled by

\[
x(f, t) = A_s(f) s(f, t) + A_z(f) z(f, t) \in \mathbb{C}^M,\]

\[
A_s(f) = [\alpha_1(f), \ldots, \alpha_K(f)] \in \mathbb{C}^{M \times K}, \quad A_z(f) \in \mathbb{C}^{M \times (M-K)},
\]

\[
s(f, t) = [s_1(f, t), \ldots, s_K(f, t)]^\top \in \mathbb{C}^K, \quad z(f, t) \in \mathbb{C}^{M-K},
\]

where \( f \in \{1, \ldots, F\} \) and \( t \in \{1, \ldots, T\} \) denote the frequency bin and time-frame indexes, respectively. \( ^\top \) is the transpose, \( \alpha_k(f) \) and \( s_k(f, t) \) are the transfer function and signal for the target source \( k \), respectively, and \( A_s(f) \) and \( z(f, t) \) are those for the noise source.

The demixing matrix \( W(f) = [w_1(f), \ldots, w_M(f)] \in \mathbb{C}^{M \times M} \) satisfying \( W(f)^\top A_s(f), A_z(f) = I_M \) translates (1) into

\[
s_k(f, t) = w_k(f)^\top X(f, t), \quad k \in \{1, \ldots, K\},
\]

\[
z(f, t) = W_z(f)^\top X(f, t) \in \mathbb{C}^{M-K},
\]

\[
W_z(f) = [w_{K+1}(f), \ldots, w_M(f)] \in \mathbb{C}^{M \times (M-K)},
\]
where \( I_M \) is the identity matrix and \( H \) is the Hermitian transpose.

The BSS problem dealt with in this paper is to recover the spatial images of the target nonstationary sources, \( \{ a_k(f) s_k(f, t) \}_{k, f, t} \), under the assumption that \( K \) is given and the source signals are independent of each other. Once the demixing matrices \( \{ W_k(f) \} \) are obtained, we can estimate the spatial images using projection back technique [17] as follows (here \( e_k \in \mathbb{C}^M \) denotes the unit vector whose \( k \)th element is equal to one and the others zero):

\[
 a_k(f) s_k(f, t) = (W(f)^{-H} e_k)(u_k(f)^H x(f, t)) \in \mathbb{C}^M. \tag{5}
\]

3. PROBABILISTIC MODEL

We present the probabilistic model of the proposed OverIIVA, which is almost identical to that of the ordinary IVA [3, 4]. The only difference is in the model of the stationary noise that we do not need to estimate. In fact, the proposed model is defined by (2)–(4) and the following (6)–(8): \( s_k(t) := [s_k(1, t), \ldots , s_k(F, t)]^T \in \mathbb{C}^F \), and

\[
 p((s_k(t), z(f, t))_{k,f,t}) = \prod_{k,f,t} p(s_k(t)) \cdot \prod_{f,t} p(z(f, t)), \tag{6}
\]

\[
 s_k(t) \sim \mathbb{C} \mathcal{N}(0, \lambda_k(t) I_F) , \quad k \in \{1, \ldots , K\}, \tag{7}
\]

\[
 z(f, t) \sim \mathbb{C} \mathcal{N}(0, I_{M-K}) . \tag{8}
\]

Note that the noise model (8) with the constant covariance matrix does not sacrifice generality. At first glance, it seems better to employ \( z(f, t) \sim \mathbb{C} \mathcal{N}(0, R(f)) \) with a general \( R(f) \) being a parameter to be optimized. However, by not interested in the noise components, we can freely change the variables to satisfy (8) using the ambiguity between \( A_k(f) \) and \( \{ z(f, t) \}_{f=1}^T \) as follows:

\[
 A_k(f) z(f, t) = (A_k(f) R(f)^H) R(f)^{-1} z(f, t). \tag{9}
\]

4. OPTIMIZATION

We develop an algorithm for the maximum likelihood (ML) estimation of the parameters \( W \) and \( \lambda \). The ML estimation is attained by minimizing the negative log-likelihood \( J \), which is computed as

\[
 J = \sum_{k,f,t} \left[ \frac{\| s_k(t) \|^2}{\lambda_k(t)} + F \log \lambda_k(t) \right] + \sum_{f,t} \| z(f, t) \|^2 - 2T \sum_{f} \log | \det W(f) | + C,
\]

where \( C \) is a constant independent of the parameters.

As is the case with IVA, the proposed algorithm updates \( W \) and \( \lambda \) alternately. When \( W \) is kept fixed, \( \lambda \) can be globally optimized by \( \lambda_k(t) = \frac{1}{T} \| s_k(t) \|^2 \).

In what follows, we will develop several computationally efficient algorithms that optimize \( W \), keeping \( \lambda \) fixed. The objective function \( J \) with respect to \( W = \{ W(f) \}_{f=1}^F \) is additively separable for each frequency bin \( f \). This enables us to split the problem into \( F \) independent problems, each of which is described as (P1) below. We hereafter abbreviate the frequency bin index \( f \) to simplify the notation without mentioning it.

\[
 \text{minimize } J_W, \quad \text{subject to } W \in W, \tag{P1}
\]

\[
 J_W = \sum_{k=1}^K w_k^H G_k w_k + \text{tr}(W_k^H G_k W_k) - 2 \log | \det W_k | , \quad G_k = \frac{1}{T} \sum_{t=1}^T S_k(t) x(t)^H \in \mathbb{C}^{M \times M}, \quad k \in \{1, \ldots , K\},
\]

\[
 G_z = \frac{1}{T} \sum_{t=1}^T f(t) x(t)^H \in \mathbb{C}^{M \times M}. \tag{11}
\]

Table 1. Optimization procedure for each method

| Method | Procedure |
|--------|-----------|
| IP-0 (§ 4.1.1 [13]) | Optimize \( w_1, \ldots , w_K, \ldots , w_M \) cyclically. |
| IP-1 (§ 4.1.2) | Optimize \( w_1, \ldots , w_K, W_k \) cyclically. |
| IP-2 (§ 4.1.3) | Optimize \( w_1, W_k \) simultaneously (when \( K = 1 \)). |
| IP-3 (§ 5 [5]) | Optimize \( w_1, W_z, w_2, W_z, \ldots , w_K, W_z \) cyclically. |

4.1. The algorithms to optimize \( W \)

If \( K = M \) and the noise component does not exist, the problem (P1) is known as the ML based ICA [18–21]. Block coordinate descent (BCD) methods have been proposed to solve it [12–14, 22, 23] and reported to be faster and to give higher separation performance than other algorithms such as the natural gradient method [24] and FastICA [8,9] (see, e.g., [12]). The family of these BCD algorithms specialized to solve the ML based ICA is currently called an iterative projection (IP) method [25]. Even when \( K < M \), IP can directly be applied to the problem (P1) by translating the second term of \( J_W \) as \( \text{tr}(W_k^H G_k W_k) = \sum_{k=1}^M w_k^H G_k w_k \). The problem is, however, a huge wasted computational cost especially in the case of \( K < M \).

We therefore propose two different computationally efficient algorithms based on BCD to solve the problem (P1). We call them IP-1 and IP-2, respectively, because they can be viewed as extensions of the conventional IP, named IP-0 in this paper. All algorithms exactly optimize one or several columns of \( W \) in each iteration while keeping all other columns fixed. The procedures of IP-1 and IP-2 as well as IP-0, are summarized in Table 1. As we will clarify in Section 5, that the conventional OverIIVA [5] can be obtained from the proposed OverIIVA by selecting IP-3 in Table 1 as an optimization procedure of BCD. To make the algorithms work properly, we need to introduce two technical but essential conditions (C1) and (C2):

(C1) \( G_1, \ldots , G_K, G_z \) are positive definite.

(C2) Estimates of \( W \) are always nonsingular during optimization.

As we will see, the algorithms are developed by exploiting the first order necessary optimality (stationary) conditions of the problem (P1) with respect to \( w_k (k = 1, \ldots , K) \) and \( W_z \), which are expressed as follows (see, e.g., [20, 21]):

\[
 \frac{\partial J_W}{\partial w_k} = 0 \Longleftrightarrow W_k^H G_k w_k = e_k \in \mathbb{C}^M, \tag{9}
\]

\[
 \frac{\partial J_W}{\partial W_z} = 0 \Longleftrightarrow W_z^H G_z W_z = E_z \in \mathbb{C}^{M \times (M-K)} , \tag{10}
\]

where \( ^* \) denotes the element-wise conjugate, \( 0_M \in \mathbb{C}^M \) means the zero vector while \( O \in \mathbb{C}^{M \times (M-K)} \) denotes the zero matrix, and we define \( E_z := \{ e_{K+1}, \ldots , e_M \} \in \mathbb{C}^{M \times (M-K)} \).

4.1.1. The review of the conventional IP: IP-0

We review IP-0 for solving the problem (P1) (see [13] for the details). In each iteration, IP-0 updates \( w_k \) for some \( k = 1, \ldots , M > K \) so that it globally minimizes \( J_W \) with respect to \( w_k \) while keeping all other variables fixed. This is achieved under (C1)–(C2) by

\[
 u_k \leftarrow (W_k^H G_k)^{-1} e_k \in \mathbb{C}^M , \tag{11}
\]

\[
 w_k \leftarrow w_k + \mu G_z w_k \left[ u_k^H G_k u_k \right]^{-1} \in \mathbb{C}^M . \tag{12}
\]

2 If (C1) is violated, the problem (P1) has no optimal solutions and the algorithms diverge to infinity (see [26, Proposition 1] for the proof). In practice, we can always guarantee (C1) by adding small \( \epsilon I_M \) heuristically (see Algorithm 1). The condition (C2) is satisfied automatically if we initialize \( W \) as nonsingular. Intuitively, a singular \( W \) implies \( - \log | \det W | = +\infty \), which will never occur during optimization.
Here, $G_k := G_k (k = K + 1, \ldots, M)$, and the $k$th column of $W$ in (11), $w_k$, is set to the current estimate before update. It is guaranteed that updating $w_k$ with (11)-(12) will not increase the value of $J_W$.

4.1.2. The derivation of IP-1

We derive IP-1 that updates $w_1, \ldots, w_K, W_z$ cyclically. The update rules for $w_1, \ldots, w_K$, given by (11)-(12) while that for $W_z$ is defined by (17) below. To begin with, we enumerate all global solutions for the problem of minimizing $J_W$ with respect to $W_z$ when keeping $w_1, \ldots, w_K$ fixed.

**Proposition 1.** Suppose (C1) and (C2). A matrix $W_z$ satisfies the stationary condition (10) if and only if

$$U_z \leftarrow (W^H G_z)^{-1} E_z,$$

(13)

$$W_z \leftarrow U_z \left( U_z^H G_z U_z \right)^{-\frac{1}{2}} Q, \quad Q \in U_{M-K}$$

(14)

where $U_{M-K}$ is the set of all unitary matrices of size $M - K$, and the last $M - K$ columns of $W$ in (13), $W_z$, are set to the current estimate before update.

Moreover, (13)-(14) with an arbitrary $Q \in U_{M-K}$ globally minimizes $J_W$ if $w_1, \ldots, w_K$ are kept fixed.

**Proof.** The first $K$ rows of (10) are linear with respect to $W_z$ and solved as

$$W_z = \begin{pmatrix} U_z \end{pmatrix} \left( U_z^H G_z U_z \right)^{-\frac{1}{2}} Q,$$

(15)

with (11) implying $U_z \in U_{M-K}$, $G_z \in C^{M \times (M-K)}$ is a free parameter. The remaining $M - K$ rows of (10) constrain $B$ and we obtain (13)-(14), which certainly satisfies the stationary condition (10).

We next prove the latter statement. As (13)-(14) satisfies (10), $\text{tr}(W_z^H G_z W_z) = M - K$ holds. It also holds that $|\text{det} W| = |\text{det}[W_z W_z^H Q^{-1}]|$ for any $Q \in U_{M-K}$. Hence, $J_W$ takes the same value on all stationary points. On the other hand, by Proposition 4, $J_W$ attains its minimum at some $W_z$, and this $W_z$ must satisfy the stationary condition (10).

From Proposition 1, as an update formula of $W_z$, we can use (13)-(14) with $Q = I_{M-K}$. Although this formula guarantees the monotonic nonincrease of the cost function, the computation of (14) is not efficient. We therefore propose an acceleration of (13)-(14), resulting in (17) below.

We now show that we can adopt (17) below as an update formula for $W_z$ if we do not need to separate the noise to satisfy (8). Let $U_z$ and $W_z$ be defined by (13) and (14), respectively. Let also $W_z^* \in C^{M \times (M-K)}$ be a matrix satisfying $\text{Im} W_z^* = \text{Im} U_z (= \text{Im} W_z)$, i.e., $W_z^* = W_z R$ for some $R \in C^{(M-K) \times (M-K)}$. Then, concerning (11) and (13),

$$\left(W_z^H G_z \right)^{-1} e_k = \left(W_z^* G_z \right)^{-1} e_k, \quad k = 1, \ldots, K$$

(15)

$$\text{Im} \left( \left(W_z^H G_z \right)^{-1} E_z \right) = \text{Im} \left( \left(W_z^* G_z \right)^{-1} E_z \right),$$

(16)

where $W' := [W_z, W_z^*]$. The equality (15) implies that using $W'$ instead of $W$ does not affect the resultant $w_k$ obtained from (11)-(12) for each $k = 1, \ldots, K$. Also, (16) means that $\text{Im} U_z$ is updated to the same subspace in (13) regardless of whether we use $W$ or $W'$ on the right hand side of (13). Hence, it turns out that we only have to update $W_z$ so as to satisfy $\text{Im} W_z = \text{Im} U_z$, unless we care about the noise components. Such update of $W_z$ can be attained simply by $W_z \leftarrow U_z := \left( W_z^H G_z \right)^{-1} E_z$. In this paper, inspired by [5], we propose a more efficient update:

$$W_z \leftarrow \left( \left( W_z^H G_z E_z \right)^{-1} W_z \right) \left( W_z^* G_z E_z \right) \in C^{M \times (M-K)},$$

(17)

where $E_z := [e_1, \ldots, e_K]$ and $W_z := [w_1, \ldots, w_K] = W E_z$.

We can check using the block matrix inversion that $W_z$ obtained by (17) satisfies $\text{Im} W_z = \text{Im} U_z$, where $U_z$ is given by (13). This concludes the derivation of IP-1 and the procedure of IP-1 is summarized in Algorithm 2.

4.1.3. The derivation of IP-2 only for $K = 1$

When $K = M = 2$ (or $K = 1$ and $M = 2$), it is known that the problem (P1) can be solved directly through a generalized eigenvalue problem [14, 23, 25]. We here extend this direct method to the case where $K = 1$ and $M \geq 2$, which is summarized in Proposition 2.

**Proposition 2.** Suppose (C1) and (C2), and let $K = 1$ and $M \geq 2$. Then, a matrix $W = [u_1, W_z]$ satisfies the stationary condition (9)-(10) if and only if

$$u_1 \in C^M$$ satisfying $G_z u_1 = \lambda G_z u_1,$

(18)

$$W_z = \begin{pmatrix} U_z \end{pmatrix} \left( U_z^H G_z U_z \right)^{-\frac{1}{2}} Q, \quad Q \in U_{M-1},$$

(19)

where $\lambda$ is the generalized eigenvalue problem with $u_1$ and $\lambda \in \mathbb{R}$ being a generalized eigenvector and the corresponding eigenvalue.

Moreover, if $\lambda$ in (18) is chosen as the largest generalized eigenvalue, then any $W$ obtained by (18)-(21) globally minimizes $J_W$.

**Proof.** The “if” part is obvious and we prove the “only if” part. The equations (9)-(10) imply $u_1 G_z u_1 = 1$ and that $G_z u_1$ and $G_z u_1$ are orthogonal to the subspace Im $W_z$ of dimension $M - 1$. Hence, (18) and (20) are necessary. Also, the equations (9)-(10) together with (20) imply $W_z^H G_z W_z = I_{M-1}$ and that $G_z u_1$ are orthogonal to Im $W_z$. Thus, (19) and (21) are necessary.

We next prove the latter statement. As (18)-(21) satisfy (9)-(10), the sum of the first and second terms of $J_W$ becomes $M$. On the other hand, as for the log det term, it holds that

$$|\text{det} W| = \left| \left( u_1 G_z u_1 \right)^{\frac{1}{2}} \cdot \left( U_z^H G_z U_z \right)^{-\frac{1}{2}} \cdot |\text{det} U| \right| = \sqrt{\lambda} |\text{det} (U_z G_z U_z)^{\frac{1}{2}}| |\text{det} U| = \sqrt{\lambda} \text{det}(G_z)^{\frac{1}{2}},$$

where $U := [u_1, U_z]$ and we use $U_z^H G_z u_1 = 0_{M-1}$ in the second equality. Hence, the largest $\lambda$ leads to the smallest $J_W$.

From Proposition 2, we can update $W = [u_1, W_z]$ using (18)-(21) with $\theta = 0$ and $Q = I_{M-1}$, minimizing $J_W$ globally. Note that updating $u_1$ by (18) and (20) is independent of $W_z$ since $G_z$ and $G_z$ are independent of $W_z$. We can thus simplify the procedure of IP-2, which is summarized in Algorithm 3. Interestingly, because $G_z$ and $G_z$ can be viewed as the covariance matrices of the mixture and noise signals, IP-2 turns out to be a MaxSNR beamformer [27, 28]. In other words, OverIVA with IP-2 is a method that alternately updates the target-source power spectrum and MaxSNR beamformer.

5. RELATION TO PRIOR WORKS

The conventional OverIVA [5], denoted as OverIVA-OC, is an acceleration of IVA. It exploits not only the independence of sources but also the orthogonality constraint (OC) [5, 15, 16]): the sample correlation (or inner product) between the target sources and the noise
Algorithm 1 OverIVA

1: Set $\varepsilon_1, \varepsilon_2 \in \mathbb{R}_{>0}$ (we set $\varepsilon_1 = 10^{-5}$ and $\varepsilon_2 = 10^{-4}$ in §6).
2: Initialization: $W(f) = I_M$ for all $f = 1, \ldots, F$.
3: $G_k(f) \leftarrow \frac{1}{N} \sum_{i=1}^{N} |x(f, t)|^2$ for all $f$.
4: repeat
5: $s_k(f, t) \leftarrow u_k(f) G_k(f)^{-1} e_k$
6: $\lambda_k(t) \leftarrow \max \left\{ \frac{1}{2} \sum_k \lambda_k(t) \right\}$ for all $k, t$.
7: $G_k(f) \leftarrow \frac{1}{N} \sum_{i=1}^{N} \frac{s_k(f, t)^2 x(f, t)}{|s_k(f, t)|^2} + \varepsilon_2 I_M$ for all $f, k$.
8: Update $W(f)$ using IP-1, IP-2, or IP-3 for all $f$.
9: Normalization for numerical stability: $c_k := \frac{1}{T} \sum_t \lambda_k(t)$.
10: until convergence
11: In the case of IP-2, update $W_k(f)$ using (17) for all $f$.
12: The separation result is obtained by (5).

Algorithm 2 IP-1

1: for $k = 1, \ldots, K$ do
2: $u_k(f) \leftarrow (W(f)^H G_k(f)^{-1}) e_k$
3: $u_k(f) \leftarrow u_k(f) (W_k(f)^H G_k(f) u_k(f))^{-\frac{1}{2}}$
4: end for
5: $W_k(f) \leftarrow \left( (W_k(f)^H G_k(f) E_k)^{-1} (W_k(f)^H G_k(f) E_k) \right)_{M \times K}$

Algorithm 3 IP-2 (only for $K = 1$)

1: Solve the generalized eigenvalue problem $G_1(f) u = \lambda G_1(f) u$ to obtain the eigenvector $u$ corresponding to the largest eigenvalue.
2: $u_1(f) \leftarrow u (u^H G_1(f) u)^{-\frac{1}{2}}$

Algorithm 4 IP-3 (The conventional OverIVA with OC [5] is equivalent to the proposed OverIVA with IP-3 as shown in §5)

1: for $k = 1, \ldots, K$ do
2: Execute lines 2, 3, and 5 in Algorithm 2 to update $u_k(f)$ and $W_k(f)$.
3: end for

is to be zero, i.e., $W_k^H G_k W_k = O$. This constraint is nothing but the first $K$ rows in (10), meaning that the stationarity of the Gaussian noise implicitly implies OC. In this subsection, we clarify that OverIVA-OC can also be obtained from the proposed OverIVA.

OverIVA-OC restricts $W_k$ to the form $W_k = \left( -B_k \right)_{M \times K}$, where $B_k = (W_k^H G_k E_k)^{-1} (W_k^H G_k E_k)$, this relation between $B_k$ and $W_k$ gives an update rule of $W_k$. The procedure of OverIVA-OC is summarized in Algorithm 4. Note that the update of $W_k$ has to be done immediately after optimizing any other variables $u_1, \ldots, u_K$ so as to always enforce OC in the model.

It is easy to see that Algorithm 4 can also be obtained from the proposed OverIVA by selecting IP-3 as the optimization procedure of BCD. Hence, OverIVA-OC can be viewed as a special case of our OverIVA. The main advantage of our approach is that, by removing OC, we can develop several algorithms including IP-1 and IP-2 that are expected to be more efficient than IP-3. In fact, the computational cost of IP-1 per iteration is slightly less than IP-3. Also, the convergence speed of IP-2 is much faster than IP-3. The advantages of IP-1 and IP-2 are shown experimentally (see Section 6).

6. EXPERIMENTS

We carried out experiments to compare the separation and runtime performances of the following four methods:

AuxIVA [13]: The conventional auxiliary-function-based IVA [13], followed by picking the $K$ signals with largest powers.

OverIVA-OC [5]: It is OverIVA(IP-3). See Algorithms 1 and 4.

OverIVA(IP-1) and OverIVA(IP-2): See Algorithms 1, 2, and 3.

As evaluation data, we generated synthesized convolutive mixtures of target speech and untarget white-noise signals. To this end, we used point-source speech signals from SiSEC2008 [29] and selected a set of room impulse responses (RIR) recorded in the room E2A from RWCP Sound Scene Database [30]. Then, for a given numbers of speakers $K$, white noises $L$, and microphones $M$, (i) we randomly picked $K$ speech signals and $K + L$ RIRs, (ii) generated $L$ white noises, (iii) convolved these $K + L$ point-source signals using the RIRs, and (iv) added the obtained $K + L$ spatial images so that $\text{SNIR} := 10 \log_{10} \frac{\sum_{k=1}^{K} \sigma_k^2}{\sum_{l=1}^{L} \sigma_l^2}$ [dB] becomes a specified value, where $\sigma_k^2$ and $\sigma_l^2$ denote the variances of speech and white noise signals, respectively. We generated 10 mixtures for each condition.

We initialized $W(f) = I_M$, and set the number of optimization iterations to 50 except set it to three in OverIVA(IP-2). The sampling rate was 16 kHz, the reverberation time was 300 ms, the frame length was 4096 (256 ms), and the frame shift was 1/4 of the frame length.

Table 2 shows the BSS performance averaged over 10 samples in terms of SDR [31] and real time factor (RTF) calculated as “the total computation time [sec] divided by the signal length (10 sec).” As expected, the proposed OverIVA is faster than the other methods while providing the comparable SDR. In particular, OverIVA(IP-2) specialized for $K = 1$ significantly outperforms the other methods.

Table 2. The resultant SDR [dB] and real time factor (RTF)

| Speeches & noises | SDR | RTF |
|------------------|-----|-----|
| # channels (#M)  |     |     |
|                  | 3   | 5   | 7   | 3   | 4   | 5   | 6   |
| SDR              | 0.0 | 0.0 | 0.0 | 0.4 | 0.0 | 0.0 | 0.0 |
| RTF              | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 |

Note that OverIVA-OC and OverIVA(IP-1) are the same when $K = 1$. All algorithms were implemented in Python 3.7 and run on a laptop PC with 2.6 GHz Intel Core i7.

7. CONCLUSION

We proposed a computationally efficient IVA for overdetermined BSS, called OverIVA. Unlike the previous OverIVA [5] relying on the orthogonality constraint, our approach exploits only the independence of sources and the stationarity of the Gaussian noise. We verified the effectiveness of the proposed OverIVA in the experiment.

8. APPENDIX

To give a proof of Proposition 1, we need Proposition 4 below, which is a modification of Proposition 3 provided in [20–23]. The proofs of Propositions 3 and 4 are almost the same and so we omit it.

Proposition 3 (see, e.g., [20–23]). Suppose (C1). Then, $J_W$ is lower bounded and attains its minimum.

Proposition 4. Suppose (C1) and let $W := [W_s, W_u] \in \mathbb{C}^{M \times M}$ be a matrix. If the submatrix $W_u$ is full column rank and $W_s$ is a variable, then the function $J(W_u) := \text{tr} (W_s^H G s W_u) - 2 \log |\det W|$ is lower bounded and attains its minimum.
