Geometrically Constrained Independent Vector Analysis with Auxiliary Function Approach and Iterative Source Steering

Kana Goto*, Tetsuya Ueda*, Li Li†, Takeshi Yamada*, Shoji Makino‡
*University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577, Japan
†NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Japan
‡Waseda University, Japan
Email: k.goto@mmlab.cs.tsukuba.ac.jp, t.ueda@mmlab.cs.tsukuba.ac.jp, lili-0805@ieee.org,
takeshi@cs.tsukuba.ac.jp, s.makino@waseda.jp

***Abstract*—In this paper, we propose an alternative algorithm, which is faster and more stable, for geometrically constrained independent vector analysis (GC-IVA) to tackle multichannel speech separation problem. GC-IVA is a method that combines IVA, a blind source separation method, with beamforming-based geometrical constraints, which are defined using the spatial information of the sources, so that it allows us to achieve high separation performance while able to obtain the target speech at the desired output channel. GC-IVA with auxiliary-function approach and vectorwise coordinate descent (GCAV-IVA) is one such method, which has the advantage that no step-size tuning is required, the objective function monotonically decreases, and the algorithm converges fast. However, this method requires matrix inversion, which is computationally expensive and adversely affects numerical stability. To address this problem, we propose an algorithm by using the recently introduced iterative source steering (ISS), which uses a sequence of rank-1 update. ISS does not require matrix inversion and achieves a lower computational complexity per iteration of quadratic in the number of microphones, resulting in the proposed method being faster and more stable than GCAV-IVA. The experimental results revealed that the proposed method had higher source separation performance and shorter execution time than conventional methods.**

***Index Terms*—Multichannel blind source separation, independent vector analysis, geometric constraints, auxiliary function approach, iterative source steering**

I. Introduction

When capturing speeches using a distant microphone, diffuse noise and directional interferences are mixed during recording and they can significantly degrade the performances of many speech processing applications. Blind Source Separation (BSS) methods separate such observed mixtures to provide access to the individual sources of the mixture [1]–[5]. BSS algorithms, including a variety of independent component analysis (ICA) methods, estimate source signals using only the observed signals based on the assumption that source signals are statistically independent with each other.

When BSS is applied to extract specific sources from the observed mixture signal, post-processing is usually required to select the desired sources using additional cues, such as speaker information or spatial information. However, it is preferable to solve source selection jointly with source separation since the clues used for desired source selection can also be helpful for source separation. Geometrically constrained BSS (GC-BSS) [6]–[11] is one of such methods that exploits spatial information to guide the demixing matrices to obtain a signal from a desired direction. Since GC-BSS usually separates signals using a spatial null, which is estimated on the basis of the statistical independence of source signals, it can work with a small number of microphones without any training samples. Geometrically constrained independent vector analysis (GC-IVA) [6], [9], [10] is one of the GC-BSS method, which combines the optimization problem of IVA [2], [3] with beamforming-based geometric constraints derived from a prior spatial information of source signals and the sensor geometry. Many algorithms have been proposed for solving the optimization problem of GC-IVA, including gradient descent method [6]. Among them, GC-IVA with auxiliary-function approach and vectorwise coordinate descent (GCAV-IVA) [10], [11] adopts auxiliary function approach [12] and vectorwise coordinate descent algorithm (VCD) [13], resulting in an algorithm noteworthy in high performance, fast convergence, and no requirement of step-size parameter tuning. These characteristics make GCAV-IVA suitable for practical applications. Furthermore, owing to the well-designed geometric constraints, GCAV-IVA can reduce the negative impact of block permutation between the low- and high-frequency bands in the auxiliary function-based IVA (AuxIVA) [4], [5], and subsequently improve speech separation performance.

Despite all their advantages, the update rules of GCAV-IVA require matrix inversion, which is computational consuming. In addition, the matrix inversion is desired to be avoided since it makes numerical computation unstable. Recently, iterative source steering (ISS) has been introduced for AuxIVA to overcome similar disadvantages [14], which updates the whole demixing matrix with a rank-1 update. This leads to an inverse-free algorithm and reduces the computational complexity. AuxIVA with ISS has been demonstrated to achieve comparable separation performance with the original AuxIVA while significantly reducing computational time when the number of microphones increases.

Towards practical applications, in this paper, we derive an algorithm for GCAV-IVA based on ISS to stabilize the numerical computation and reduce computational cost, which we call "GC-AuxIVA-ISS". It preserves the advantages of fast convergence and non-requirement of pre-training and step-size parameter tuning. Experimental results show that the separation performance of the proposed GC-AuxIVA-ISS is stable and it can perform separation faster than conventional methods.

II. Baseline method: GC-IVA with Auxiliary Function Approach and VCD

Let us consider a determined situation where J sources are observed by I microphones. Let x_{ifn} and y_{jfn} denote the short-time Fourier transform (STFT) coefficients of the signals observed at the i-th microphone and the j-th estimated sources, respectively. Here, $f = 1, \dots, F$ and $n = 1, \dots, N$ are the indices of the frequency and frame, respectively. We denote the frequency-wise vector representation of the observations and the estimated sources by

$$\boldsymbol{x}_{fn} = [x_{1fn}, \dots, x_{Ifn}]^{\mathsf{T}} \in \mathbb{C}^{I}, \tag{1}$$

$$\boldsymbol{y}_{fn} = [y_{1fn}, \dots, y_{Jfn}]^{\mathsf{T}} \in \mathbb{C}^{J}, \tag{2}$$

where $(\cdot)^{\mathsf{T}}$ denotes the transpose. When considering a determined case, where $I = J$, and a time-invariant instantaneous mixture model, where the STFT window length is sufficiently longer than the impulse responses between sources and microphones, the relationship between the observations and the estimated sources can be expressed as

$$\boldsymbol{y}_{fn} = \boldsymbol{W}_f \boldsymbol{x}_{fn}, \tag{3}$$

where $\boldsymbol{W}_f = [\boldsymbol{w}_{1f}, \dots, \boldsymbol{w}_{Jf}]^{\mathsf{H}}$ is an $I \times I$ demixing matrix containing demixing filters $\boldsymbol{w}_{jf} = [w_{1jf}, \dots, w_{Ijf}]^{\mathsf{T}}$, and $(\cdot)^{\mathsf{H}}$ denotes the Hermitian transpose.

IVA assumes that each frame of source follows a multivariate distribution and thus dependencies over frequency components can be exploited to solve frequency-domain permutation alignment. The demixing matrices $\mathcal{W} = \{\boldsymbol{W}_f\}_f$ are estimated by minimizing the following negative log-likelihood function

$$\mathcal{L}_{\text{IVA}}(\mathcal{W}) = \sum_{j=1}^{J} \mathbb{E}[G(\boldsymbol{y}_{jn})] - \sum_{f=1}^{F} \log |\det \boldsymbol{W}_f|, \tag{4}$$

where, $\mathbb{E}[\cdot]$ denotes the expectation operator and $\boldsymbol{y}_{jn} = [y_{j1n}, \dots, y_{jFn}]^{\mathsf{T}} \in \mathbb{C}^{F}$ is the source-wise vector representation. Here, $G(\boldsymbol{y}_{jn})$ is the contrast function having the relationship $G(\boldsymbol{y}_{jn}) = -\log p(\boldsymbol{y}_{jn})$, where $p(\boldsymbol{y}_{jn})$ represents a multivariate probability density function of the j-th source at n-th frame. One typical choice of the contrast function is to use a spherical contract function [2]–[4], which is expressed as

$$G(\boldsymbol{y}_{jn}) = G_R(r_{jn}), \tag{5}$$

$$r_{jn} = ||\boldsymbol{y}_{jn}||_2 = \sqrt{\sum_f |\boldsymbol{y}_{jn}|^2}. \tag{6}$$

Here, $G_R(r)$ is a function of a real-valued scalar variable r, and $|| \cdot ||_2$ denotes the L_2 norm of a vector. By adopting the auxiliary function approach [12], an upper bound is optimized instead of the original objective function, which is expressed as

$$\begin{aligned} \mathcal{L}_{\text{IVA}}(\mathcal{W}) &\leq \mathcal{L}_{\text{AuxIVA}}(\Sigma, \mathcal{W}) \\ &= \frac{1}{2} \sum_{f=1}^{F} \sum_{j=1}^{J} \boldsymbol{w}_{jf}^{\mathsf{H}} \boldsymbol{\Sigma}_{jf} \boldsymbol{w}_{jf} - \sum_{f=1}^{F} \log |\det \boldsymbol{W}_f|. \end{aligned} \tag{7}$$

Here, $\Sigma = \{\boldsymbol{\Sigma}_{jf}\}_{jf}$ and $\boldsymbol{\Sigma}_{jf}$ is the weighted covariance expressed as

$$\boldsymbol{\Sigma}_{jf} = \sum_n \varphi(r_{jn}) \boldsymbol{x}_{fn} \boldsymbol{x}_{fn}^{\mathsf{H}}. \tag{8}$$

Here, $\varphi(r_{jn}) = G_R(r_{jn})^{'}/r_{jn}$ and $(\cdot)'$ denotes the derivative operator.

Now, let us consider geometric constraints [15] that restrict the far-field response of filters estimated by IVA in a set of directions Θ, which is described as

$$\mathcal{L}_{\text{GC}}(\mathcal{W}) = \sum_{j=1}^{J} \sum_{\theta \in \Theta} \lambda_{j\theta} \sum_{f=1}^{F} |\boldsymbol{w}_{jf}^{\mathsf{H}} \boldsymbol{d}_{f\theta} - c_{j\theta}|^2. \tag{9}$$

Here, Θ denotes a set including all directions to be considered, $\boldsymbol{d}_{f\theta}$ is the steering vector pointing to the direction θ, $c_{j\theta}$ is a nonnegative value set for all frequency bins as constraints, and $\lambda_{j\theta} \geq 0$ is a parameter that weighs the importance of the constraint. Note that (9) with $c_{j\theta} = 1$ forces the spatial filter to form a conventional delay-and-sum beamformer steering in the direction θ to preserve the target source whereas a small value of $c_{j\theta}$ essentially creates a spatial null towards the direction θ so that multiple constraints of spatial nulls towards the directions of all interferences can be used to suppress all interferences and preserve the target.

The objective function of GCAV-IVA is summarized as

$$\mathcal{L}(\Sigma, \mathcal{W}) = \mathcal{L}_{\text{AuxIVA}}(\Sigma, \mathcal{W}) + \mathcal{L}_{\text{GC}}(\mathcal{W}). \tag{10}$$

The update rule for Σ is obtained straightforwardly by applying (6) into (8), whereas the update rule for $\mathcal{W}$ is derived by embracing the idea adopted in VCD [13] that arranges the term $\log |\det \boldsymbol{W}|$ with the property of cofactor expansion. The derived update rules are summarized as follows:

$$\boldsymbol{D}_{jf} = \boldsymbol{\Sigma}_{jf} + \sum_{\theta \in \Theta} \lambda_{j\theta} \boldsymbol{d}_{f\theta} \boldsymbol{d}_{f\theta}^{\mathsf{H}} \tag{11}$$

$$\boldsymbol{u}_{jf} = \boldsymbol{D}_{jf}^{-1} \boldsymbol{W}_f^{-1} \boldsymbol{e}_j, \tag{12}$$

$$\hat{\boldsymbol{u}}_{jf} = \boldsymbol{D}_{jf}^{-1} \sum_{\theta \in \Theta} \lambda_{j\theta} c_{j\theta} \boldsymbol{d}_{f\theta}, \tag{13}$$

$$h_{jf} = \boldsymbol{u}_{jf}^{\mathsf{H}} \boldsymbol{D}_{jf} \boldsymbol{u}_{jf}, \tag{14}$$

$$\hat{h}_{jf} = \boldsymbol{u}_{jf}^{\mathsf{H}} \boldsymbol{D}_{jf} \hat{\boldsymbol{u}}_{jf}, \tag{15}$$

$$\boldsymbol{w}_{jf} = \begin{cases} \frac{1}{\sqrt{h_{jf}}} \boldsymbol{u}_{jf} + \hat{\boldsymbol{u}}_{jf} & (\text{if } \hat{h}_{jf} = 0), \\ \frac{\hat{h}_{jf}}{2h_{jf}} \Big[-1 + \sqrt{1 + \frac{4h_{jf}}{|\hat{h}_{jf}|^2}} \Big] \boldsymbol{u}_{jf} + \hat{\boldsymbol{u}}_{jf} & (\text{o.w.}). \end{cases} \tag{16}$$

Here, $\boldsymbol{e}_j$ is the j-th column of the $I \times I$ identity matrix. These update rules are equivalent to those employed in AuxIVA when $\lambda_{j\theta} = 0$ for all j and θ. The details of the derivation are available in [10] and [13].

These update rules have the advantage that no step-size tuning is required, the objective function monotonically deceases, and the algorithm converges fast. However, the matrix inverse required at each iteration is computationally expensive and may adversely affects numerical stability.

III. Proposed method: GC-AuxIVA-ISS

We propose a new update algorithm for GCAV-IVA based on ISS [14], which are lower computational cost and inverse-free. We call it GC-IVA with auxiliary function approach and ISS (GC-AuxIVA-ISS). Instead of updating a single row of the demixing matrix $\boldsymbol{W}_f$ alternately, ISS performs a rank-1 update for the whole demixing matrix as

$$\boldsymbol{W}_f \leftarrow \boldsymbol{W}_f - \boldsymbol{v}_{jf}\boldsymbol{w}_{jf}^{\mathsf{H}} \tag{17}$$

for $j = 1, \ldots, I$. Here, $\boldsymbol{v}_{jf}$ is a vector to be estimated instead of the demixing matrix.

Plugging (17) into (10), we have

$$\begin{aligned}\mathcal{L}(\boldsymbol{v}_{jf}) = &-\sum_{f=1}^{F} \log|\det(\boldsymbol{W}_f - \boldsymbol{v}_{jf}\boldsymbol{w}_{jf}^{\mathsf{H}})| \\ &+ \sum_{f=1}^{F}\sum_{i=1}^{I}\Big\{\frac{1}{2}(\boldsymbol{w}_{if} - v_{ijf}^{*}\boldsymbol{w}_{jf})^{\mathsf{H}}\boldsymbol{\Sigma}_{jf}(\boldsymbol{w}_{if} - v_{ijf}^{*}\boldsymbol{w}_{jf}) \\ &+ \sum_{\theta\in\Theta}\lambda_{i\theta}|(\boldsymbol{w}_{if} - v_{ijf}^{*}\boldsymbol{w}_{jf})^{\mathsf{H}}\boldsymbol{d}_{f\theta} - c_{i\theta}|^2\Big\},\end{aligned} \tag{18}$$

which is the new objective function to be minimized. The index of f is omitted hereafter for the notation simplicity. We derive update rules for cases of $i \neq j$ and $i = j$ separately.

First, when $i \neq j$, we can obtain the partial derivative of $\mathcal{L}(\boldsymbol{v}_j)$ w.r.t. v_{ij}^{*} as

$$\begin{aligned}\frac{\partial}{\partial v_{ij}^{*}}\mathcal{L}(\boldsymbol{v}_j) = &-\frac{1}{2}\sum_n \varphi(r_{in}) y_{in} y_{jn}^{*} + \frac{1}{2} v_{ij}\sum_n \varphi(r_{in})|y_{jn}|^2 \\ &+ \sum_{\theta\in\Theta}\lambda_{i\theta}\{v_{ij}|g_{j\theta}|^2 - g_{j\theta}^{*}(g_{i\theta} - c_{i\theta})\},\end{aligned} \tag{19}$$

where $g_{j\theta} = \boldsymbol{w}_j^{\mathsf{H}}\boldsymbol{d}_\theta$, $\sum_n \varphi(r_{in}) y_{in} y_{jn}^{*} = \boldsymbol{w}_i^{\mathsf{H}}\boldsymbol{\Sigma}_i\boldsymbol{w}_j$, and $\sum_n \varphi(r_{in})|y_{jn}|^2 = \boldsymbol{w}_j^{\mathsf{H}}\boldsymbol{\Sigma}_i\boldsymbol{w}_j$. From $\partial\mathcal{L}(\boldsymbol{v}_j)/\partial v_{ij}^{*} = 0$, we have

$$v_{ij} = \frac{\sum_n \varphi(r_{in}) y_{in} y_{jn}^{*} + 2\sum_{\theta\in\Theta}\lambda_{i\theta} g_{j\theta}^{*}(g_{i\theta} - c_{i\theta})}{\sum_n \varphi(r_{in})|y_{jn}|^2 + 2\sum_{\theta\in\Theta}\lambda_{i\theta}|g_{j\theta}|^2}. \tag{20}$$

Next, when $i = j$, we can obtain the partial derivative of $\mathcal{L}(\boldsymbol{v}_j)$ w.r.t. v_{jj}^{*} as

$$\begin{aligned}\frac{\partial}{\partial v_{jj}^{*}}\mathcal{L}(\boldsymbol{v}_j) = &\frac{1}{2}(1 - v_{jj}^{*})^{-1} - \frac{1}{2}(1 - v_{jj})\sum_n \varphi(r_{jn})|y_{jn}|^2 \\ &+ \sum_{\theta\in\Theta}\lambda_{j\theta}\{v_{jj}|g_{j\theta}|^2 - g_{j\theta}^{*}(g_{j\theta} - c_{j\theta})\},\end{aligned} \tag{21}$$

and equating this expression to zero, we have

$$\begin{aligned}1 - |1 - v_{jj}|^2\Big(\sum_n \varphi(r_{jn})|y_{jn}|^2 + 2\sum_{\theta\in\Theta}\lambda_{j\theta}|g_{j\theta}|^2\Big) \\ + 2(1 - v_{jj})^{*}\sum_{\theta\in\Theta}\lambda_{j\theta}c_{j\theta}g_{j\theta}^{*} = 0.\end{aligned} \tag{22}$$

Because the first and second terms in (22) are real numbers, the third term in (22) must satisfy

$$\mathrm{Im}\Big[(1 - v_{jj})^{*}\sum_{\theta\in\Theta}\lambda_{j\theta}c_{j\theta}g_{j\theta}^{*}\Big] = 0. \tag{23}$$

From $(1 - v_{jj})^{*} \neq 0$ and (23), we have

$$\sum_{\theta\in\Theta}\lambda_{j\theta}c_{j\theta}g_{j\theta} = 0 \tag{24}$$

or

$$(1 - v_{jj})^{*} = \gamma_j\sum_{\theta\in\Theta}\lambda_{j\theta}c_{j\theta}g_{j\theta}, \tag{25}$$

where $\gamma_j \in \mathbb{R}\backslash\{0\}$. When (24) holds, (22) simplifies to

$$v_{jj} = 1 - \Big(\sum_n \varphi(r_{jn})|y_{jn}|^2 + 2\sum_{\theta\in\Theta}\lambda_{j\theta}|g_{j\theta}|^2\Big)^{-1/2}. \tag{26}$$

On the other hand, when (25) holds, we can derive a quadratic equation in γ_j from (22) as follows:

$$1 - \gamma_j^2|\beta_j|^2\alpha_j + 2\gamma_j|\beta_j|^2 = 0, \tag{27}$$

where,

$$\alpha_j = \sum_n \varphi(r_{jn})|y_{jn}|^2 + 2\sum_{\theta\in\Theta}\lambda_{j\theta}|g_{j\theta}|^2, \tag{28}$$

$$\beta_j = \sum_{\theta\in\Theta}\lambda_{j\theta}c_{j\theta}g_{j\theta}. \tag{29}$$

By substituting the solution of (27) into (25), we have

$$\gamma_j = \frac{-|\beta_j| \pm \sqrt{|\beta_j|^2 + \alpha_j}}{-\alpha_j|\beta_j|}, \tag{30}$$

and

$$v_{jj} = 1 - \beta_j^{*}\frac{|\beta_j| \mp \sqrt{|\beta_j|^2 + \alpha_j}}{\alpha_j|\beta_j|}, \tag{31}$$

where the ∓ sign in (31) should be positive (see Appendix A).

In summary, when $i = j$, the minimization of (18) with respect to v_{ij} gives

$$v_{jj} = \begin{cases} 1 - \alpha_j^{-1/2} \ (\beta_j = 0), \\ 1 - \beta_j^{*}\frac{|\beta_j| + \sqrt{|\beta_j|^2 + \alpha_j}}{\alpha_j|\beta_j|} \ (\beta_j \neq 0). \end{cases} \tag{32}$$

After computing $\boldsymbol{v}_j$ by (20) and (32), we need to update the output signal $\boldsymbol{y}_n$ and $\boldsymbol{w}_i^{\mathsf{H}}\boldsymbol{d}_\theta$ by applying (17) as

$$\boldsymbol{y}_n \leftarrow \boldsymbol{y}_n - \boldsymbol{v}_j y_{jn}, \tag{33}$$

$$\boldsymbol{w}_i^{\mathsf{H}}\boldsymbol{d}_\theta \leftarrow \boldsymbol{w}_i^{\mathsf{H}}\boldsymbol{d}_\theta - v_{ij}\boldsymbol{w}_j^{\mathsf{H}}\boldsymbol{d}_\theta. \tag{34}$$

Since $\boldsymbol{w}_i^{\mathsf{H}}\boldsymbol{d}_\theta$ is a scalar, these rules has lower computational cost than that of the conventional method.

IV. Experiment

To evaluate the effectiveness of GC-AuxIVA-ISS, we conducted speech separation experiments. We evaluated each method in terms of separation performance, accuracy of output signal order, and runtime. We compared our proposed method GC-AuxIVA-ISS with GCAV-IVA and AuxIVA-ISS. For clarity, we refer to GCAV-IVA as GC-AuxIVA-VCD hereafter. Since output order of AuxIVA-ISS is arbitrary, we did not calculate the accuracy of output order. We also examined

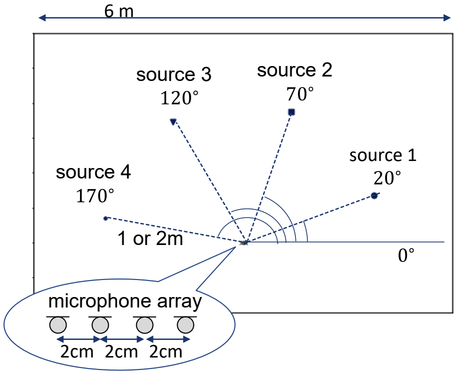


Fig. 1: Layout of sound sources and microphones.

whether the block permutation problem observed in AuxIVA-ISS can be solved by the proposed method.

A. Setup

We used speech signals of 6 speakers (3 males and 3 females) extracted from the ATR Japanese Speech Database [16]. We conducted source separation for 2, 3, and 4 sources. By randomly selecting 2 to 4 different speakers from the database, we generated 48 pair of source signals and mixtures for each case. We used the `pyroomacoustics` Python package [17] to simulate room impulse responses (RIRs), and the layout of sound sources and microphones is shown in Fig. 1. The directions of arrival (DOAs) were set at 20° and 70° for the 2-source case, 20°, 70°, and 120° for the 3-source case, and 20°, 70°, 120°, and 170° for the 4-source case, respectively. The number of microphones was set equal to the number of sources with the interval of microphones at 2 cm. We tested two different reverberant conditions, where the reverberation times (RT_{60}) were about 100 ms and 300 ms. All the speech signals were sampled at 16 kHz. The STFT was computed using a Hanning window, whose length and shift were set at 512 samples (32 ms) and 256 samples (16 ms), respectively. All methods were run for 50 iterations.

In these experiments, we assumed that the correct DOAs of speakers were known and set Θ as follows:

- when $J = 2$, $\Theta = \{20°, 70°\}$,
- when $J = 3$, $\Theta = \{20°, 70°, 120°\}$,
- when $J = 4$, $\Theta = \{20°, 70°, 120°, 170°\}$,

where J is the number of sources. Here, we define $\mathbf{\Lambda} = [\boldsymbol{\lambda}_1 \dots, \boldsymbol{\lambda}_J]^\mathsf{T}$ and $\boldsymbol{C} = [\boldsymbol{c}_1, \dots, \boldsymbol{c}_J]^\mathsf{T}$, where $\boldsymbol{\lambda}_j = [\lambda_{j\theta_1}, \dots, \lambda_{j\theta_T}] \in \mathbb{R}^T$ and $\boldsymbol{c}_j = [c_{j\theta_1}, \dots, c_{j\theta_T}] \in \mathbb{R}^T$. T is the number of elements in Θ, which was equivalent to J. We considered 3 ways to design the constraints.

- *unit response (UR)* constraint: $c_{j\theta} = 1$ when θ is the DOA of the target.
- *Null* constraint: $c_{j\theta} = 0$ when θ is the DOA of the interferences.
- *Double* constraint: constraint using both of the *UR* and *null* constraints.

We achieved *UR* constraint by setting the non-diagonal elements of $\mathbf{\Lambda}$ to zero as $\mathbf{\Lambda} = \Lambda \boldsymbol{I}$, *null* constraint by setting the diagonal elements of $\mathbf{\Lambda} = \Lambda(\boldsymbol{J} - \boldsymbol{I})$ to zero, and *double* constraint by setting as $\mathbf{\Lambda} = \Lambda \boldsymbol{J}$, respectively. Here, Λ is an arbitrary non-negative value, $\boldsymbol{I}$ is a $J \times J$ identity matrix, and $\boldsymbol{J}$ is a $J \times J$ all-ones matrix.

TABLE I: Average SDR [dB], SIR [dB], and accuracy of output signal order over 48 samples in each condition.

method	constraint	Λ	SDR [dB]	SIR [dB]	accuracy of output signal order [%]
2 channel					
AuxIVA-ISS [14]	-	-	10.19	12.20	-
GC-AuxIVA-VCD [10]	*UR*	5	10.09	12.14	100
	null	0.8	11.05	13.26	100
	double	2	10.95	13.16	100
GC-AuxIVA-ISS (proposed)	*UR*	90	10.96	13.18	100
	null	80000	**11.07**	**13.30**	100
	double	80	10.97	13.19	100
3 channel					
AuxIVA-ISS [14]	-	-	9.96	11.98	-
GC-AuxIVA-VCD [10]	*UR*	2	9.69	11.82	100
	null	5	10.62	12.74	100
	double	2	10.58	12.75	100
GC-AuxIVA-ISS (proposed)	*UR*	60	10.14	12.46	100
	null	90000	**10.63**	**12.76**	100
	double	80	10.06	12.11	100
4 channel					
AuxIVA-ISS [14]	-	-	8.64	10.65	-
GC-AuxIVA-VCD [10]	*UR*	15	5.68	7.47	100
	null	40	9.36	11.28	100
	double	2	**9.52**	**11.79**	100
GC-AuxIVA-ISS (proposed)	*UR*	60	8.84	10.98	100
	null	8000	9.17	11.35	100
	double	80	8.73	10.84	100

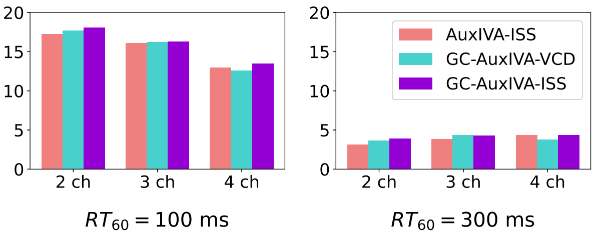


Fig. 2: Average SDR [dB] under reverberant conditions where $RT_{60} = 100$ ms and $RT_{60} = 300$ ms.

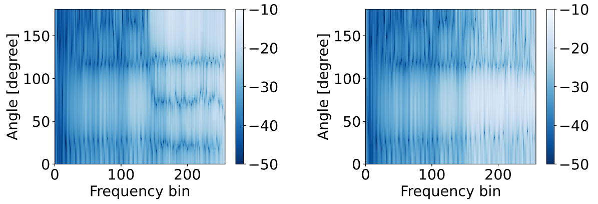
The separation performance was evaluated using the source-to-distortion ratio (SDR) and source-to-interferences ratio (SIR) [18]. The order of the output signals was determined as the one that achieves the highest SIR among all permutations. We investigated several values of $\mathbf{\Lambda}$ and chose the optimal one based on SDR and the accuracy of output signal order.

B. Results

Table I shows the average SDR, SIR, and accuracy of output signal order over 48 samples in each condition, and Fig. 2 shows the average SDR under reverberant conditions, where $RT_{60} = 100$ ms and $RT_{60} = 300$ ms. The proposed GC-AuxIVA-ISS showed the equivalent or higher SDR and SIR scores than the conventional methods. In terms of the accuracy of output signal order, we confirmed that there were no samples with incorrect output order for both GC-AuxIVA-VCD and GC-AuxIVA-ISS. This indicated that the proposed method, as well as GC-AuxIVA-VCD, could correctly guide

TABLE II: Runtime [ms] per iteration.

method	2 ch	3 ch	4 ch
AuxIVA-ISS [14]	33.02	62.31	99.06
GC-AuxIVA-VCD [10]	51.12	120.83	217.84
GC-AuxIVA-ISS (proposed)	33.60	63.16	101.60



(a) AuxIVA-ISS

(b) GC-AuxIVA-ISS

Fig. 3: Examples of beam patterns obtained by AuxIVA-ISS and GC-AuxIVA-ISS with *double* constraint when the number of sources was 4, the target direction was 70°, and $RT_{60} = 300$ ms. Block permutation problem occurred between the low- and high-frequency bands in AuxIVA-ISS, which was avoided in GC-AuxIVA-ISS.

the output order if the weights of the regularization terms were set appropriately.

Figure 3 shows examples of beam pattern obtained by separating the same input sample with AuxIVA-ISS and GC-AuxIVA-ISS. In Fig. 3(a), block permutations occurred between the low- and high-frequency bands, whereas in Fig. 3(b) it did not. This indicated that spatial information was effective for the ISS-based update rules to avoid the block permutation problem.

Table II shows the runtime of each method averaged over iterations for separating a signal with length of 10 seconds. We found that GC-AuxIVA-VCD took longer runtime than AuxIVA-ISS or GC-AuxIVA-ISS. Especially in the case of 4 channels, the ISS-based methods took less than half of runtime of the GC-AuxIVA-VCD.

V. Conclusions

In this paper, we proposed an algorithm for GC-IVA using ISS method, which we call GC-AuxIVA-ISS. GC-AuxIVA-VCD is a method that combines IVA with a set of linear constraints that limit the far-field responses of the demixing filters, whose update rules are derived based on the auxiliary function approach and VCD. The matrix inversion required for each iteration is computationally inefficient and makes numerical computations unstable. On the other hand, the proposed method based on ISS does not require inverse matrix, resulting in a lower computational cost. The experimental results confirmed that the proposed method outperformed the conventional GC-AuxIVA-VCD in terms of separation performance and runtime.

Appendix A. Solution of sign ambiguity in (31)

When $i = j$ and (25) holds, by extracting the terms containing v_{jj} in $\mathcal{L}$ and substituting (28), (29), and (31) into those we obtain:

$$-\log\frac{\left||\beta_j| \mp \sqrt{|\beta_j|^2+\alpha_j}\right|}{\alpha_j} + \frac{1}{2\alpha_j}\Big\{\Big(-|\beta_j| \mp \sqrt{|\beta_j|^2+\alpha_j}\Big)^2 - 4|\beta_j|^2\Big\}. \qquad (35)$$

From (35), $\mathcal{L}$ becomes smaller when we take $+$. Then the $\mp$ sign in (31) should be positive.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 19H04131.

References

[1] A.Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[2] T. Kim, T. Eltoft, and T.-W. Lee, “Independent vector analysis: An extension of ICA to multivariate components,” in *Proc. ICA*, 2006, pp. 165–172.

[3] A. Hiroe, “Solution of permutation problem in frequency domain ICA using multivariate probability density functions,” in *Proc. ICA*, 2006, pp. 601–608.

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

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

[6] A. H. Khan, M. Taseska, and E. A. P. Habets, “A geometrically constrained independent vector analysis algorithm for online source extraction,” in *Proc. LVA/ICA*, 2015, pp. 396–403.

[7] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fastconvergence algorithm combining ICA and beamforming,” *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.

[8] M. Knaak, S. Araki, and S. Makino, “Geometrically constrained independent component analysis,” *IEEE Trans. ASLP*, vol. 15, no. 2, pp. 715–726, 2007.

[9] A. Brendel, T. Haubner, and W. Kellermann, “A unified probabilistic view on spatially informed source separation and extraction based on independent vector analysis,” *IEEE Trans. Signal Processing*, vol. 68, 2020.

[10] L. Li and K. Koishida, “Geometrically constrained independent vector analysis for directional speech enhancement,” in *Proc. ICASSP*, 2020, pp. 846–850.

[11] K. Goto, L. Li, R. Takahashi, S. Makino, and T. Yamada, “Study on geometrically constrained IVA with auxiliary function approach and VCD for in-car communication,” in *Proc. APSIPA*, 2020, pp. 858–862.

[12] D. R Hunter and K. Lange, “A tutorial on MM algorithms,” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

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

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

[15] L. C. Parra and C. V. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” *IEEE Trans. SAP*, vol. 10, no. 6, pp. 352–362, 2002.

[16] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis,” *Speech communication*, vol. 9, no. 4, pp. 357–363, 1990.

[17] R. Scheibler, E. Bezzam, and I. Dokmanic, “Pyroomacoustics: A python package for audio room simulations and array processing algorithms,” in *Proc. ICASSP*, 2018, pp. 351–355.

[18] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE/ACM Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.