ACCELERATING ALGORITHM OF GEOMETRICALLY CONSTRAINED SOURCE SEPARATION AND DEREVERBERATION USING ITERATIVE SOURCE STEERING

Boyuan Wang, Kaien Mo, Tetsuya Ueda, Shoji Makino

Waseda University, Japan

ABSTRACT

We propose a new algorithm to improve computational efficiency in the geometrically constrained dereverberation and blind source separation method (GC-DR-BSS). Our research GC-ILRMA-T-ISS applies the GC in independent low-rank matrix analysis with time decorrelation and iterative source steering (ILRMA-T-ISS). By utilizing the prior spatial information of desired signals, the output orders will be controllable. Then an efficient optimization algorithm for GC-ILRMA-T-ISS (EGC-ILRMA-T-ISS) is put forward to address the unnecessary calculation of updating joint filters in each iteration in ILRMA-T-ISS. Relying on the merit of the joint filter in ILRMA-T-ISS, EGC-ILRMA-T-ISS's implementation is more simplified and it synchronously possesses the functions of GC-based separation, DR, and high computational efficiency. Simulation results validate that both the GC-ILRMA-T-ISS and EGC-ILRMA-T-ISS can achieve good separation performance in strong reverberation conditions. Besides, the EGC-ILRMA-T-ISS can greatly accelerate the filters' update.

Index Terms— Blind source separation, geometrical constraints, dereverberation, joint optimization, iterative source steering.

1. INTRODUCTION

In the speech signal processing field, a number of the solutions for the blind source separation (BSS) research topic [1–4] have been developed. As an available approach, independent component analysis (ICA) [5] takes advantage of minimizing the inter-channel correlation to separate the mixing signals. As an extension of ICA, independent vector analysis (IVA) [6, 7] can deal with the frequency permutation problem by exploiting the high-order correlations between the frequency bins. To improve the convergence speed and stability, the auxiliary function is introduced into IVA (AuxIVA) [8]. Furthermore, independent low-rank matrix analysis (ILRMA) [9] replaces the IVA source model by the nonnegative matrix factorization (NMF)'s [10] one to build the spectral structures for each frequency bin. Therefore, it can match the sources of harmonic structures.

Based on the above BSS methods, research has been conducted to jointly perform dereverberation, source separation, and permutation alignment. For a strongly reverberated environment, research has applied such a dereverberation (DR) as weighted prediction error (WPE) [11] for preprocessing of BSS. For more accurate source separation, joint majorization algorithms have been proposed that optimize DR and BSS in the same framework (DR-BSS), such as ILRMA with time decorrelation (ILRMA-T) [12, 13], and WPE-IVA [14, 15]. On the other hand, to handle a permutation ambiguity of BSS, researchers have proposed geometrically constrained BSS (GC-BSS) [16–18]. GC-BSS exploits spatial information to guide the permutation of the separated signal according to the specified direction of arrivals (DOAs). These GC and DR have been concurrently integrated into BSS (GC-DR-BSS) recently [19, 20].

This research aims to accelerate the algorithm process for GC-DR-BSS. To realize that, we focus on iterative source steering (ISS) [21], which updates estimation matrices used for BSS and DR with a rank-1 update. It leads to an algorithm without matrix inversion and reduces the computational complexity from cubic to square. Thus, ISS has been applied to several methods [13, 18, 20]. GC-WPE-IVA-ISS [20] is a conventional method that aims to accelerate GC-DR-BSS. It improved the computational efficiency while keeping a comparable separation performance to GC-WPE-IVA. However, it still involves matrix inversion for DR, which requires the cubic order.

In this paper, we propose a computationally efficient GC-DR-BSS method that does not require any matrix inversion for the DR filter's optimization. Recently, ISS update has been applied to ILRMA-T (ILRMA-T-ISS) [13] as an inversion-free algorithm for the DR-BSS field. Thus, we incorporated GC into ILRMA-T-ISS [13] and derived an algorithm, which we call GC-ILRMA-T-ISS. Furthermore, we propose an efficient implementation of GC-ILRMA-T-ISS by inspiring empirical knowledge from the WPE-IVA [14]. We conduct a simulation experiment to confirm the separation performance and computational efficiency of our proposed method.

2. PROBLEM FORMULATION

Assume that there are N sources and M microphones. The source signals and observed signals can be expressed in vectors as

$$\boldsymbol{s}_{f,t} = [s_{1,f,t} \cdots s_{N,f,t}]^{\mathsf{T}} \in \mathbb{C}^N, \tag{1}$$

$$\boldsymbol{x}_{f,t} = [x_{1,f,t} \cdots x_{M,f,t}]^{\mathsf{T}} \in \mathbb{C}^M, \tag{2}$$

where $(\cdot)^{\mathsf{T}}$ denotes transpose, $s_{n,f,t}$ and $x_{m,f,t}$ represent the nth source signal and mth observed signal in the short-time Fourier transformation (STFT) domain. Here, $f = 1 \cdots F$ and $t = 1 \cdots T$ are frequency bin and time frame indices. F and T are the numbers of frequency bins and frames, respectively. The signals observed by microphones can be expressed as

$$\boldsymbol{x}_{f,t} = \sum_{\tau=0}^{L_A-1} \boldsymbol{A}_{f,\tau} \boldsymbol{s}_{f,t-\tau}, \tag{3}$$

where $\boldsymbol{A}_{f,\tau} \in \mathbb{C}^{M \times N}$ is the convolutional mixing matrix at time lag τ, and L_A is the order of time lagged mixing filters. In this paper, we only consider the determined condition, where $N = M$. Based on ILRMA-T framework [22], for the higher quality of speech and efficiency, proposing a united framework to optimize the DR-BSS

topic is reasonable, the output signals can be expressed as

$$\boldsymbol{y}_{f,t} = \tilde{\boldsymbol{W}}_f \tilde{\boldsymbol{x}}_{f,t}, \tag{4}$$

where

$$\tilde{\boldsymbol{W}}_f = [\boldsymbol{W}_f, \overline{\boldsymbol{W}}_f] = \begin{bmatrix} \boldsymbol{w}_{1,f}^{\mathsf{H}}, & \overline{\boldsymbol{w}}_{1,f}^{\mathsf{H}} \\ \vdots & \vdots \\ \boldsymbol{w}_{N,f}^{\mathsf{H}}, & \overline{\boldsymbol{w}}_{N,f}^{\mathsf{H}} \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{\boldsymbol{w}}_{1,f}^{\mathsf{H}} \\ \vdots \\ \tilde{\boldsymbol{w}}_{N,f}^{\mathsf{H}} \end{bmatrix} \in \mathbb{C}^{N \times M(L+1)} \tag{5}$$

$$\tilde{\boldsymbol{x}}_{f,t} = \begin{bmatrix} \boldsymbol{x}_{f,t} \\ \overline{\boldsymbol{x}}_{f,t} \end{bmatrix} \in \mathbb{C}^{M(L+1)}, \tag{6}$$

and $\overline{\boldsymbol{x}}_{f,t} = [\boldsymbol{x}_{f,t-D}^{\mathsf{T}} \cdots \boldsymbol{x}_{f,t-L-D+1}^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{C}^{ML}$ contains past observed signals. Here, $(\cdot)^{\mathsf{H}}$ represents conjugate transpose.

To estimate $\tilde{\boldsymbol{W}}_f$, we use the assumption from conventional algorithms:

- The sources are statistically independent,
- A source signal at each time-frequency bin belongs to a complex Gaussian distribution:
  $p(y_{n,f,t}) = \frac{1}{\pi r_{n,f,t}} \exp(-\frac{|y_{n,f,t}|^2}{r_{n,f,t}})$, where $y_{n,f,t}$ is the nth element of $\boldsymbol{y}_{f,t}$ and $r_{n,f,t}$ is the time-varying variance of the nth source,
- $r_{n,f,t}$ is modeled as $r_{n,f,t} = \sum_{k=1}^{K} b_{n,k,f} a_{n,k,t}$, where K is the number of basis vectors, $b_{n,k,f}$ is the basis coefficient of the nth component, and $a_{n,k,t}$ is the activity of the nth component.

Using the above assumptions, the negative log-likelihood function can be derived as [13]:

$$\mathcal{L}(\Theta) = -2 \sum_{f} \log |\det \boldsymbol{W}_f| + \frac{1}{T} \sum_{n,f,t} \left(\log r_{n,f,t} + \frac{|y_{n,f,t}|^2}{r_{n,f,t}} \right), \tag{7}$$

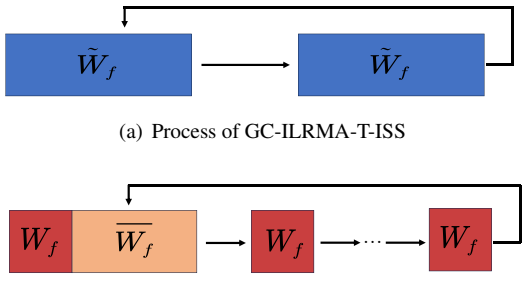
where $\Theta = \{\Theta_{\boldsymbol{W}} \Theta_{\overline{\boldsymbol{W}}}, \Theta_r\}$ is the parameter set. Here, $\Theta_{\boldsymbol{W}} = \{\boldsymbol{W}_f\}$, $\Theta_{\overline{\boldsymbol{W}}} = \{\overline{\boldsymbol{W}}_f\}$, and $\Theta_r = \{r_{n,f,t}\}$. We update both $b_{n,k,f}$ and $a_{n,k,f}$ using non-negative matrix factorization (NMF) [9].

In order to solve the global permutation ambiguity problem in strong reverberation conditions, we introduce the geometrical constraint [16]. The constraint term can be expressed as

$$\mathcal{L}_{\mathrm{GC}} = \sum_{n,\varphi,f} \lambda_{n,\varphi} |\boldsymbol{w}_{n,f}^{\mathsf{H}} \boldsymbol{d}_{\varphi,f} - c_{n,\varphi}|^2, \tag{8}$$

where $\lambda_{n,\varphi}$ is a non-negative weighting coefficient, $\boldsymbol{d}_{\varphi,f} \in \mathbb{C}^M$ is the steering vector points to direction φ and $c_{n,\varphi}$ is the non-negative-valued constraint. Note that when $c_{n,\varphi} = 1$ in (8), it will force the spatial filter to form a conventional delay-and-sum beamforming steering at φ to extract the target signal whereas setting $c_{n,\varphi}$ a small value will create a spatial null to suppress the interference [17]. By involving the geometrical constraints, we can obtain the cost function by adding the constrained term (8) to (7) as

$$\mathcal{L}(\Theta) = -2 \sum_{f} \log|\det \boldsymbol{W}_f| + \frac{1}{T} \sum_{n,f,t} \left(\log r_{n,f,t} + \frac{|y_{n,f,t}|^2}{r_{n,f,t}} \right) + \sum_{n,\varphi,f} \lambda_{n,\varphi} |\boldsymbol{w}_{n,f}^{\mathsf{H}} \boldsymbol{d}_{\varphi,f} - c_{n,\varphi}|^2. \tag{9}$$



(a) Process of GC-ILRMA-T-ISS

(b) Process of EGC-ILRMA-T-ISS

Fig. 1: Relationship between GC-ILRMA-T-ISS and EGC-ILRMA-T-ISS.

The goal of this paper is to devise a computationally efficient algorithm to estimate the optimum $\Theta_{\boldsymbol{W}}$ and $\Theta_{\overline{\boldsymbol{W}}}$.

3. PROPOSED ALGORITHM

3.1. GC-ILRMA-T-ISS

Similarly to ILRMA-T-ISS [13], we update the matrix $\tilde{\boldsymbol{W}}_f$ with one-rank update for $1 \leq n \leq M(L+1)$:

$$\tilde{\boldsymbol{W}}_f \leftarrow \tilde{\boldsymbol{W}}_f - \begin{bmatrix} v_{n,1,f} \\ \vdots \\ v_{n,N,f} \end{bmatrix} \tilde{\boldsymbol{w}}_{n,f}^{\mathsf{H}}, \tag{10}$$

where $\tilde{\boldsymbol{w}}_{n,f} = \boldsymbol{e}_n$ for $n = N+1, \cdots, M(L+1)$ and $\boldsymbol{e}_n$ is the nth column of the identity matrix $\boldsymbol{I}_{M(L+1)} \in \mathbb{R}^{M(L+1) \times M(L+1)}$. We can find the optimum $v_{n,k,f}$ for $k = 1, \cdots, N$ as

$$v_{n,k,f} = \begin{cases} \frac{\tilde{\boldsymbol{w}}_{n,f}^{\mathsf{H}} \tilde{\boldsymbol{U}}_{n,f} \tilde{\boldsymbol{w}}_{k,f} + \sum_{\varphi} \lambda_{n,\varphi} (\tilde{h}_{n,\varphi,f} - c_{n,\varphi}) \tilde{h}_{k,\varphi,f}^*}{\tilde{\boldsymbol{w}}_{k,f}^{\mathsf{H}} \tilde{\boldsymbol{U}}_{n,f} \tilde{\boldsymbol{w}}_{k,f} + \sum_{\varphi} \lambda_{n,\varphi} |\tilde{h}_{k,\varphi,f}|^2} & \text{if } k \neq n \\ 1 - \tilde{\alpha}_{k,f}^{-1/2} & \text{elif } \tilde{\beta}_{k,f} = 0, \\ 1 - \tilde{\beta}_{k,f}^* \frac{|\tilde{\beta}_{k,f}| + \sqrt{|\tilde{\beta}_{k,f}^2| + 4\tilde{\alpha}_{k,f}}}{2\tilde{\alpha}_{k,f} |\tilde{\beta}_{k,f}|} & \text{else,} \end{cases} \tag{11}$$

where

$$\tilde{\boldsymbol{U}}_{n,f} = \frac{1}{T} \sum_{t} \frac{\tilde{\boldsymbol{x}}_{f,t} \tilde{\boldsymbol{x}}_{f,t}^{\mathsf{H}}}{r_{n,f,t}}. \tag{12}$$

$$\tilde{h}_{n,\varphi,f} = \tilde{\boldsymbol{w}}_{n,f}^{\mathsf{H}} \tilde{\boldsymbol{d}}_{\varphi,f}, \tag{13}$$

$$\tilde{\boldsymbol{d}}_{\varphi,f} = [\boldsymbol{d}_{\varphi,f}^{\mathsf{T}}, \boldsymbol{0}_{ML}^{\mathsf{T}}]^{\mathsf{T}} \tag{14}$$

$$\tilde{\alpha}_{k,f} = \tilde{\boldsymbol{w}}_{k,f}^{\mathsf{H}} \tilde{\boldsymbol{U}}_{k,f} \tilde{\boldsymbol{w}}_{k,f} + \sum_{\varphi} \lambda_{k,\varphi} |\tilde{h}_{k,\varphi,f}|^2, \tag{15}$$

$$\tilde{\beta}_{k,f} = \sum_{\varphi} \lambda_{k,\varphi} c_{k,\varphi} \tilde{h}_{k,\varphi,f}. \tag{16}$$

3.2. Efficient optimization algorithm for GC-ILRMA-T-ISS (EGC-ILRMA-T-ISS)

We next propose a new updating rule to realize further computationally efficient optimization, and we will call this method EGC-

ILRMA-T-ISS in the subsequent experiments. Although ILRMA-T-ISS [13] can estimate both $\boldsymbol{W}_f$ and $\overline{\boldsymbol{W}}_f$ efficiently, its update rules are not speculated. We featured a heuristic technique found by previous research, the WPE-IVA [14, 15]. In this research, the nth separated signal can be produced by two steps,

$$\boldsymbol{z}_{n,f,t} = \boldsymbol{x}_{f,t} - \boldsymbol{G}_{n,f}\overline{\boldsymbol{x}}_{f,t}, \tag{17}$$

$$y_{n,f,t} = \boldsymbol{w}_{n,f}^{\mathsf{H}}\boldsymbol{z}_{n,f,t}, \tag{18}$$

where $\boldsymbol{G}_{n,f}$ is the multi-input multi-output (MIMO) WPE filters of the nth source. Within this framework, the WPE filters $\boldsymbol{G}_{n,f}$ converged much faster than $\boldsymbol{W}_f$, so the iteration times of the WPE filters can be much smaller than the separation filters. In the ILRMA-T-ISS framework, the $\tilde{\boldsymbol{W}}_f$ is equal to the multiplication of separation filters $\boldsymbol{W}_f$ and the WPE filters $\boldsymbol{G}_f$ as [13]

$$\tilde{\boldsymbol{W}}_f = \boldsymbol{W}_f[\boldsymbol{I}_M, -\boldsymbol{G}_f], \tag{19}$$

Where $\boldsymbol{I}_M$ represents the identity matrix and $\boldsymbol{I}_M \in \mathbb{R}^{M\times M}$. Featured by the fast convergence of the WPE filters in the WPE-IVA, we intend to update the WPE filters once every several iterations for GC-ILRMA-T-ISS as shown in Fig. 1(b), and EGC-ILRMA-T-ISS is proposed. In this method, the updating of $\tilde{\boldsymbol{W}}_f$ is completely the same as GC-ILRMA-T-ISS. Once $\tilde{\boldsymbol{W}}_f$ is updated, the WPE filters $[\boldsymbol{I}_M, -\boldsymbol{G}_f]$ will be extracted and fixed. When fixing $\boldsymbol{G}_f$, the optimization of $\boldsymbol{W}_f$ is simplified to the GC-IVA-ISS [18] problem whose input is $\boldsymbol{z}_{f,t}$ instead of observed signal $\boldsymbol{x}_{f,t}$. Here, $\boldsymbol{z}_{f,t}$ is the output of MIMO WPE as

$$\boldsymbol{z}_{f,t} = [\boldsymbol{I}_M, -\boldsymbol{G}_f]\tilde{\boldsymbol{x}}_{f,t}. \tag{20}$$

Since the update of the separation matrix $\boldsymbol{W}_f$ is the same as conventional GC-IVA-ISS [18], we will omit the optimization details in this paper because of the space limitation. By multiplying the updated separation filters $\boldsymbol{W}_f$ and the WPE filters $[\boldsymbol{I}_M, -\boldsymbol{G}_f]$ before the next joint optimization, a new $\tilde{\boldsymbol{W}}_f$ is restored, and the problem is turned back to the GC-ILRMA-T-ISS.

3.3. Efficiency analysis of EGC-ILRMA-T-ISS

In this subsection, we will analyze different computation costs of DR filter optimization between conventional GC-WPE-IVA-ISS [20] and our proposed EGC-ILRMA-T-ISS. In the GC-WPE-IVA-ISS, the separation and DR filters are updated respectively while fixing other parameters. Since both of them use the same cost function (9), by fixing other parameters and ignoring the constant variables, the cost function of the WPE filters update can be expressed as

$$\mathcal{L}(\boldsymbol{G}_{n,f}) = \sum_{n,t,f} |\boldsymbol{w}_{n,f}^{\mathsf{H}}(\boldsymbol{x}_{f,t} - \boldsymbol{G}_{n,f}^{\mathsf{H}}\overline{\boldsymbol{x}}_{f,t})|^2 / r_{n,t}. \tag{21}$$

Then we can update $\boldsymbol{G}_{n,f}$ as

$$\boldsymbol{G}_{n,f} \leftarrow \boldsymbol{R}_{n,f}^{-1}\boldsymbol{P}_{n,f}, \tag{22}$$

where

$$\boldsymbol{R}_{n,f} = \sum_t \frac{\overline{\boldsymbol{x}}_{f,t}\overline{\boldsymbol{x}}_{f,t}^{\mathsf{H}}}{r_{n,t}}, \tag{23}$$

$$\boldsymbol{P}_{n,f} = \sum_t \frac{\overline{\boldsymbol{x}}_{f,t}\boldsymbol{x}_{f,t}^{\mathsf{H}}}{r_{n,t}}, \tag{24}$$

are spatio-temporal covariance matrices. Comparing with the joint matrix $\tilde{\boldsymbol{W}}_f$'s optimization in (10), the update process of

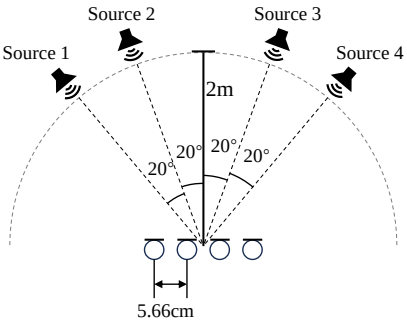


Fig. 2: Simulated experiment environment layout.

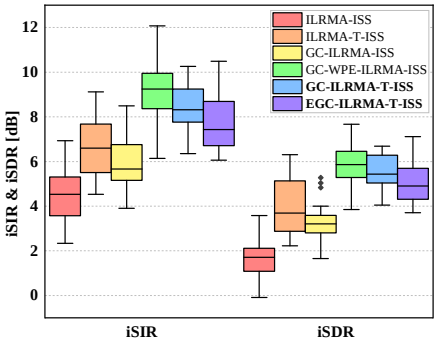


Fig. 3: Improvement SDR and SIR of the related methods in reverberation situation (T_{60} = 600ms).

WPE requires additional calculation of the covariance matrix $\boldsymbol{P}_{n,f}$ as (24) and inverse of the matrix $\boldsymbol{R}_{n,f}$ as (22). More specifically, the total computational complexity of $\tilde{\boldsymbol{W}}_f$ in GC-ILRMA-T-ISS is $\mathcal{O}(FN(ML)T)$ while the computational complexity of only the WPE filters $\boldsymbol{G}_{n,f}$ in GC-WPE-IVA-ISS is $\mathcal{O}(FN(ML)^2\max((ML), T))$. Therefore, our proposed method provides a faster solution for the GC-DR-BSS problem.

4. EXPERIMENT

4.1. Experimental setup

To simulate the real-environmental recordings, the observed signals for our experiment were generated by convolving the speech signals from the TIMIT database [23] with the room impulse responses (RIRs) from the RWCP dataset [24]. For each mixing signal, 4 randomly selected speech fragments from different speakers were used as the clean speech signals, and each of them was concatenated to 10 s long. The room reverberation time (T_{60}) was set to 600 ms. White Gaussian noise was added to control the signal-to-noise ratio (SNR) to 30 dB. The layout of sources and microphones was shown in Fig. 2, where a 4-element uniform linear microphone array (ULA) with an inter-element spacing of 5.66 cm was used and DOAs of four sources are 50°, 70°, 110°, and 130°, respectively. The distance from the center of the ULA to the source signals is 2 m. Twenty-five Monte Carlo simulations were carried out to measure the separation performance. All the signals were sampled at 16 kHz and the STFT transform with a Hann window of 64 ms (1024 samples), and the Hann window sift of 16 ms (256 ms). In order

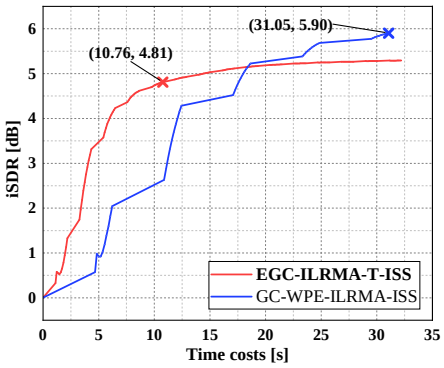


Fig. 4: Convergence speed of EGC-ILRMA-T-ISS and GC-WPE-ILRMA-ISS. Note that in this figure, we marked the 50th iteration's results of GC-WPE-ILRMA-T-ISS and EGC-ILRMA-T-ISS to correspond to the results in Table 1.

to obtain a fair comparison, the method proposed in IVA had been rewritten to ILRMA. We modified IVA-ISS [21], GC-IVA-ISS [18], and GC-WPE-IVA-ISS [20] into ILRMA-ISS, GC-ILRMA-ISS, and GC-WPE-ILRMA-ISS, respectively. Therefore, the comparison methods in our experiments are ILRMA-ISS, ILRMA-T-ISS, GC-ILRMA-ISS, GC-WPE-ILRMA-ISS, GC-ILRMA-T-ISS, and EGC-ILRMA-T-ISS. As for the dereverberation filters in WPE and ILRMA-T-based methods, we set the time delay D to 2 and set the length of taps L to 10. Besides, we set the number of the bases K in the NMF source model by 2. And regarding the parameters of the GC algorithm part, we defined two matrices $\boldsymbol{\Lambda} = [\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_N]^{\mathsf{T}}$ and $\boldsymbol{C} = [\boldsymbol{c}_1, \cdots, \boldsymbol{c}_N]^{\mathsf{T}}$ to denote the weights of importance and beamforming in (8). Here, $\boldsymbol{\lambda}_n = [\lambda_{n,\varphi_1}, \cdots, \lambda_{n,\varphi_n}]^{\mathsf{T}} \in \mathbb{R}^N$ $\boldsymbol{c}_n = [c_{n,\varphi_1}, \cdots, c_{n,\varphi_n}]^{\mathsf{T}} \in \mathbb{R}^N$, and $\Phi = \{\varphi_1, \varphi_2, \cdots, \varphi_N\}$ is the set of DOAs. We set $\boldsymbol{C} = \boldsymbol{I}$ and $\boldsymbol{\Lambda} = \Lambda(\boldsymbol{J} - \boldsymbol{I})$, where Λ is a non-negative value, and $\boldsymbol{J}$ is an all-one matrix. We initialized Λ and decreased it over iterations. According to [25], Λ of ith iteration can be calculated as

$$\Lambda_i = \Lambda_0 \eta^i, \tag{25}$$

where the Λ_0 set by 8000 was the initial value of the Λ, η is the factor of decrease set by 0.8. The results of all the methods in the Fig. 3 were given with 50 iterations. In the GC-WPE-ILRMA-ISS method, we updated the WPE filters every 10 iterations. Similarly, in the EGC-ILRMA-T-ISS, the joint filters $\tilde{\boldsymbol{W}}_f$ were updated every 10 iterations for the balance between computation cost and performance. To evaluate the separated performance, we used improvements of signal-to-distortion ratio (iSDR) and signal-to-interference ratio (iSIR) as the metrics.

4.2. Experimental results

Firstly, we discussed only the separation performance of each method using Fig. 3. In the figure, our proposed GC-ILRMA-T-ISS and EGC-ILRMA-T-ISS had a little drop in iSDR and iSIR, compared with GC-WPE-ILRMA-ISS. However, both proposed methods showed higher iSDR and iSIR than ILRMA-ISS, ILRMA-T-ISS, and GC-ILRMA-ISS. This result meant that our proposed methods could perform source separation more effectively than methods that do not belong to GC-DR-BSS.

Next, we discussed the related source separation methods' effectiveness based on separation performance and convergence speed

Table 1: The improvement of SDR (iSDR) [dB] and the running time of 50 iterations (Time) [s]. Note the above results were averaged across all data sets and we confirmed that all methods shown in this table did not cause output permutation errors.

T_{60} = 600ms		
methods	iSDR	Time
GC-ILRMA-ISS	3.28	8.44
GC-WPE-ILRMA-ISS	5.90	31.05
GC-ILRMA-T-ISS	5.49	53.27
EGC-ILRMA-T-ISS	**4.81**	**10.76**

in our experiments. We showed the comparison of the convergence speed between EGC-ILRMA-T-ISS and GC-WPE-ILRMA-ISS using Fig. 4. As the results were shown in the figure, the converged iSDR of EGC-ILRMA-T-ISS is slightly lower than that of GC-WPE-ILRMA-ISS. However, the iSDR of the proposed method converged quickly and showed a higher value than the conventional method before 18 seconds. This result meant that EGC-ILRMA-T-ISS helps estimate the parameters of GC-DR-BSS efficiently. Finally, we showed iSDR and calculation time of all GC-based methods using Table 1. From the results shown in Table 1, we could find that compared with GC-ILRMA-ISS, although our method added an update operation for the DR matrix, it did not produce too much extra time cost (about 2 seconds). On the other hand, GC-ILRMA-T-ISS, updating the joint matrix in each iteration not only did not bring significant improvement in iSDR but also brought a lot of extra time (about 41 seconds). Despite GC-WPE-ILRMA-ISS showing the highest iSDR, it required about 30 seconds. These results confirmed that the proposed method can accelerate the parameter convergence in the field of GC-DR-BSS.

5. CONCLUSION

In this paper, to achieve a method that jointly optimizes separation and dereverberation filters and addresses the permutation ambiguity problem, we proposed a geometrically constrained ILRMA-T-ISS algorithm. To further exploit the advantages of the ILRMA-T-ISS framework, EGC-ILRMA-T-ISS was proposed to improve efficiency. The experiment results showed that our proposed methods can significantly promote performance in the strong reverberation environment with controllable output order than conventional methods without geometrical constraints or dereverberation methods. Moreover, compared with the existing spatially informed joint separation and dereverberation method, although the final performance is slightly degraded, the computation cost of our proposed method EGC-ILRMA-T-ISS is greatly reduced. In future work, we intend to expand this method to more microphone cases, i.e. the overdetermined situation, and improve its performance while maintaining the low computation cost advantage.

6. ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number 23H03423.

7. REFERENCES

[1] S. Makino, T.-W. Lee, and H. Sawada, *Blind speech separation*. Springer, 2007.

[2] S. Makino, *Audio Source Separation*. Springer, 2018.

[3] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. Wiley-IEEE Press., 2018.

[4] X. Wang, N. Pan, J. Benesty, and J. Chen, "On multiple input/binaural output antiphasic speaker signal extraction," in *Proc. IEEE ICASSP*, 2023, pp. 1–5.

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

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

[7] T. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: Definition and algorithms," in *Proc. ACSSC*, 2006, pp. 1393–1396.

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

[9] 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. on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.

[10] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[11] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Speech dereverberation based on variance-normalized delayed linear prediction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 7, pp. 1717–1731, 2010.

[12] R. Ikeshita, N. Ito, T. Nakatani, and H. Sawada, "A unifying framework for blind source separation based on a joint diagonalizability constraint," in *Proc. EUSIPCO*, 2019, pp. 1–5.

[13] T. Nakashima, R. Scheibler, M. Togami, and N. Ono, "Joint dereverberation and separation with iterative source steering," in *Proc. IEEE ICASSP*, 2021, pp. 216–220.

[14] T. Nakatani, R. Ikeshita, K. Kinoshita, H. Sawada, and S. Araki, "Computationally efficient and versatile framework for joint optimization of blind speech separation and dereverberation," in *Proc. Interspeech*, 2020, pp. 91–95.

[15] T. Nakatani, C. Boeddeker, K. Kinoshita, R. Ikeshita, M. Delcroix, and R. Haeb-Umbach, "Jointly optimal denoising, dereverberation, and source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2267–2282, 2020.

[16] L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 10, no. 6, pp. 352–362, 2002.

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

[18] K. Goto, T. Ueda, L. Li, T. Yamada, and S. Makino, "Geometrically constrained independent vector analysis with auxiliary function approach and iterative source steering," in *Proc. EUSIPCO*, 2022, pp. 757–761.

[19] Y. Yang, X. Wang, A. Brendel, W. Zhang, W. Kellermann, and J. Chen, "Geometrically constrained source extraction and dereverberation based on joint optimization," in *Proc. EUSIPCO*, 2023, pp. 41–45.

[20] K. Mo, X. Wang, Y. Yang, T. Ueda, S. Makino, and J. Chen, "On joint dereverberation and source separation with geometrical constraints and iterative source steering," in *APSIPA ASC*. IEEE, 2023, pp. 1138–1142.

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

[22] R. Ikeshita, N. Ito, T. Nakatani, and H. Sawada, "Independent low-rank matrix analysis with decorrelation learning," in *IEEE WASPAA*, 2019, pp. 288–292.

[23] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," *NASA STI/Recon technical report n*, vol. 93, p. 27403, 1993.

[24] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition." in *LREC*, 2000, pp. 965–968.

[25] 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. IEEE ICASSP*, 2018, pp. 746–750.