A Computationally Efficient Online Algorithm for Tracking Single Moving Source using Geometrically Constrained Independent Vector Analysis

Jiahui Hu
Graduate School of Information, Production and Systems
Waseda University
Kitakyushu, Japan
hujiahui22@toki.waseda.jp

Tetsuya Ueda
Graduate School of Information, Production and Systems
Waseda University
Kitakyushu, Japan
t.ueda@akane.waseda.jp

Bo He
Graduate School of Information, Production and Systems
Waseda University
Kitakyushu, Japan
gahaku_bobo@toki.waseda.jp

Shoji Makino
Graduate School of Information, Production and Systems
Waseda University
Kitakyushu, Japan
s.makino@waseda.jp

***Abstract*—This paper discusses Independent Vector Analysis (IVA)-based moving source separation under a source tracking scenario where at most one speaker moves. Previous research proposed a computationally efficient algorithm to update filters for source separation under a situation where only one speaker is moving. To make this algorithm practical, we need to optimize the filters before one speaker starts moving and select which filter corresponds to the moving speaker. Thus, we applied Geometric Constraint (GC) to increase the filter's convergence speed and select the filter automatically. Compared with research that applies GC all the time, our contribution is to apply GC only when all speakers are fixed, which results in lower computational complexity while maintaining its separation performance for our scenario.**

***Index Terms*—Moving source separation, source tracking, geometric constraint, online source steering**

I. Introduction

Blind Source Separation (BSS) is a technique to extract the source signals from their observed mixture. In many consumer electronics applications, such as mobile phones and hearing aids, BSS must operate in real-time. Many popular methods for real-time BSS are based on online Independent Vector Analysis (IVA). Among those, previous research proposed a computationally efficient algorithm, Online Source Steering (OSS) [1], which updates filters for source separation under situations where only one speaker is moving. Although this algorithm has an advantage in updating the filters in a computationally efficient way, it is necessary to handle several issues to make this algorithm practical: we need to finish optimizing the filters and select which filter corresponds to the moving speaker before one speaker starts moving.

For the above issues, it is promising to use spatial prior information such as Direction of Arrival (DOA), which can be roughly given by camera or estimation when considering practical applications. Indeed, researchers have applied Geometric Constraint (GC) derived from the spatial prior information to online source separation [2], which improves the filter's convergence speed and enables us to select the moving speaker's filter during separation.

However, applying GC all the time conflict with original OSS update rules, requiring additional filter update calculations for the original OSS. Moreover, it is hard to obtain DOAs accurately when the speaker is moving. Thus, we propose to apply GC only when all speakers are fixed. Our experiments show that we can achieve computationally efficient filter updates using OSS while achieving high convergence speed and filter selection by applying GC only for a specific period.

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

II. Problem formulation

Let us consider a source tracking scenario where a certain moving source and other $K-1$ spatially stationary sources are mixed and captured by K microphones. We represent the observed signals in the STFT domain as

$$\boldsymbol{x}_{f,t} = \boldsymbol{A}_{f,t}\boldsymbol{s}_{f,t} = \sum_{k=1}^{K} \boldsymbol{a}_{k,f,t} s_{k,f,t} \in \mathbb{C}^K, \tag{1}$$

where $f = 1, \ldots, F$ and $t = 1, \ldots, T$ denote, respectively, the frequency bin and time frame indexes, $\boldsymbol{s}_{f,t} \in \mathbb{C}^K$ represents the K source signals, and the mixing vector $\boldsymbol{a}_{k,f,t} \in \mathbb{C}^K$ is the kth column vector of the mixing matrix $\boldsymbol{A}_{f,t} \in \mathbb{C}^{K\times K}$. We assume that only the ℓth source is moving, i.e., the mixing vectors $\boldsymbol{a}_{k,f,t}$ $(k \neq \ell)$ of stationary sources are time-invariant.

Next, we define source separation and source tracking, which are addressed in this paper. Source separation is defined as the frame-by-frame estimation of separation matrix $\boldsymbol{W}_{f,t} = [\boldsymbol{w}_{1,f,t}, \ldots, \boldsymbol{w}_{K,f,t}]^{\mathsf{H}} = \boldsymbol{A}_{f,t}^{-1} \in \mathbb{C}^{K\times K}$ and separated signals $\boldsymbol{y}_{f,t}$ given by

$$\boldsymbol{y}_{f,t} = \boldsymbol{W}_{f,t}\boldsymbol{x}_{f,t} \in \mathbb{C}^K, \tag{2}$$

where $\boldsymbol{w}_{k,f,t} \in \mathbb{C}^K$ is the kth separation filter. Meanwhile, source tracking is defined as a special case of source separation where only one source is moving. Furthermore, this paper assumes to perform source separation at $t = 1, \ldots, T'$, and source tracking at $t = T'+1, \ldots, T$, where T' $(< T)$ is the start time of source tracking.

We also define filter selection as determining which of the K separation filters corresponds to the moving source. While online IVA-based methods cannot select the filter by minimizing their negative log-likelihood function, previous research realized the filter selection by adding a penalty term derived from GC to the function [2]. This paper uses the following penalty term:

$$\mathcal{L}_{\mathrm{GC}}(\boldsymbol{W}_{f,t}) = \sum_{k=1}^{K} \sum_{\theta \in \Theta_{k,t}} \lambda_{k,\theta} \left| \boldsymbol{w}_{k,f,t}^{\mathsf{H}} \boldsymbol{d}_{\theta,f} \right|^2, \tag{3}$$

where $\lambda_{k,\theta}$ is a weighting coefficient and $\Theta_{k,t} = \{\theta_{1,t}, \ldots, \theta_{K,t}\} \setminus \{\theta_{k,t}\}$ represents DOAs of all speakers at

TABLE I

COMPARISON OF ADVANTAGES BETWEEN PREVIOUS RESEARCH AND OUR PROPOSED METHODS

	Required DOAs	Filter selection	Source separation algorithm and its convergence speed	Source tracking algorithm and its computational complexity
non-GC	None		[3] ... Fine	[1] ... O(K^2)
full-GC	for both separation and tracking	✓	[2] ... Fast	[2] ... O(K^3)
part-GC (Proposed)	for only source separation	✓	[2] ... Fast	[1] ... O(K^2)

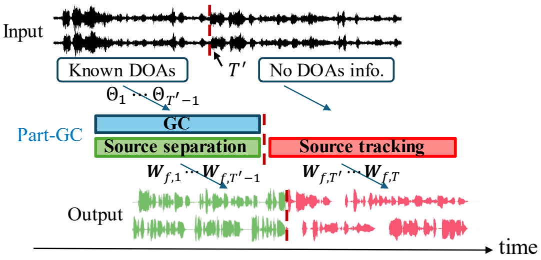

Fig. 1. Schematic of the proposed method.

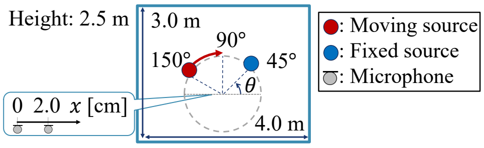

Fig. 2. Experimental sound source and microphone layout. We set the stationary sound source to remain fixed at 45° over 60 seconds. We let the moving sound source be fixed at 150° for 20 seconds, moved on an arc to 90° over the next 20 seconds, and fixed for the last 20 seconds.

time t excluding the direction of the kth source. $\boldsymbol{d}_{\theta,f}$ is the steering vector pointing to the θ direction. By applying and decreasing this penalty term, we can force the kth separation filter $\boldsymbol{w}_{k,f,t}$ to create spatial nulls in the $\theta(\in \Theta_{k,t})$ directions, resulting in fast convergence and effective filter selection.

III. PROPOSED METHOD

Our proposal is to apply the GC term in Eq. (3) to only the source separation, which we denote "part-GC". Specifically, we set $\Theta_{k,t} = \emptyset$ at $t = T' + 1, \ldots, T$. We show a schematic of partGC in Fig. 1. Hereafter, we denote an online source separation method without GC as "non-GC" and a method applying GC all the time as "full-GC" to distinguish conventional and proposed methods. We summarized the algorithms used by each method and their advantages in Table I.

IV. EXPERIMENTAL EVALUATIONS

We conducted an experiment to demonstrate the performance of the following three moving source separation methods: non-GC, full-GC, and the proposed part-GC. The layout of two sound sources and two microphones are shown in Fig. 2. We assume that we know the source in the 150° will move after 20 seconds. For the proposed methods, we provided the DOAs of both sources until T', while for full-GC, we provided the DOAs throughout the entire speech duration. Details of other experimental setups are available in [2].

First, we compared source separation performance and computational efficiency of each method in Table II using Source-to-Distortion Ratios (SDR) and Real-Time Factor (RTF). We denote the RTFs for the source separation and tracking as RTF_{sep} and $\text{RTF}_{\text{track}}$, respectively. In the Table II, non-GC shows the smallest RTF_{sep} and $\text{RTF}_{\text{track}}$, while its SDR was the

TABLE II

COMPARISON OF ALGORITHMS BASED ON RTF AND SDR WITH SOURCE TRACKING START TIME $T' = 20$ [s].

Method	SDR [dB]	RTF_{sep}	$\text{RTF}_{\text{track}}$
non-GC [1], [3]	5.38	0.04	0.01
full-GC [2]	8.67	0.07	0.05
part-GC (Proposed)	8.62	0.07	0.01

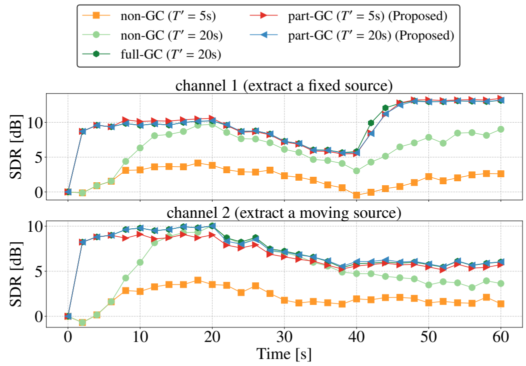

Fig. 3. Average SDR [dB] for each channel in every 2 seconds.

worst because it does not hold filter selection. Full-GC shows higher SDR than non-GC but increases RTF_{sep} and $\text{RTF}_{\text{track}}$. On the other hand, part-GC maintained SDR and decreased $\text{RTF}_{\text{track}}$ to the same extent as non-GC. These results show that it is sufficient to apply GC to only the source separation, and part-GC is useful for effective filter selection while maintaining computational complexity for source tracking.

Next, we evaluated the convergence speed of each method using Fig. 3. Compared with other methods, non-GC ($T' =$ 5s) does not sufficiently converge the filter, and it takes 20s for the convergence. On the other hand, part-GC ($T' =$ 5s) shows almost the same SDR as part-GC ($T' =$ 20s) and full-GC ($T' =$ 20s). This result means that partGC contributes to accelerating the filter convergence and effective filter selection.

V. CONCLUSION

This paper studied the problem of tracking single moving sources. We introduced part-GC, which applied GC only when all sources are fixed. Our experiments showed that part-GC realized computationally efficient filter updates using OSS while achieving high convergence speed and filter selection.

REFERENCES

[1] T. Nakashima, R. Ikeshita, N. Ono, S. Araki, and T. Nakatani, "Fast online source steering algorithm for tracking single moving source using online independent vector analysis," in *Proc. IEEE ICASSP*, 2023, pp. 1–5.

[2] K. Goto, T. Ueda, L. Li, T. Yamada, and S. Makino, "Accelerating online algorithm using geometrically constrained independent vector analysis with iterative source steering," in *Proc. APSIPA*, 2022, pp. 754–759.

[3] T. Nakashima and N. Ono, "Inverse-free online independent vector analysis with flexible iterative source steering," in *Proc. APSIPA*, 2022, pp. 749–753.