2026 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing
February 27 - March 2, 2026, Honolulu, Hawaii, U.S.A.

On Joint Switching Dereverberation and Spatially Regularized Independent Vector Analysis

Guangming Zheng[1], Wei Liu[1,2], Changda Chen[1], Shoji Makino[1]

[1]Waseda University, 2-7 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka 808-0135, Japan
[2]Wuhan University, School of Electronic Information, Wuhan, Hubei 430072, China
E-mail: {guangmingzheng@toki, s.makino@}.waseda.jp

Abstract

Switching Independent Vector Analysis (SwIVA) is a blind source separation algorithm that employs a switching mechanism over multiple separation matrices to enhance separation accuracy. However, SwIVA suffers from the global permutation problem when prior information about the sources is unavailable. Furthermore, it exhibits poor performance under reverberant environments. This paper extends SwIVA to a convolutional beamforming algorithm with a spatially regularized separation method (SR-SwCIVA). We conduct experiments to confirm that SR-SwCIVA achieves excellent separation performance under reverberant conditions while maintaining strong robustness against the global permutation problem.

1. Introduction

Blind source separation (BSS) [1] aims to recover individual source signals from observed mixtures. Independent Component Analysis (ICA) [2] is a fundamental BSS technique that achieves source separation by maximizing the statistical independence of the estimated source signals. However, in the time-frequency (TF) domain, ICA suffers from the frequency permutation problem on frequency bins. Independent Vector Analysis (IVA) [3] resolves this by exploiting higher-order dependency across frequency components. Independent Vector Extraction (IVE) [4] and convolutional beamforming (CBF) algorithm of IVA (CIVA) [5] are extentions to IVA. IVE can extract a specific source of interest, and CIVA has better separation performance than IVA under reverberant environments. To improve performance with limited microphones, a switching mechanism was introduced, allowing dynamic selection of the optimal separation matrix per TF domain, leading to Switching Independent Vector Analysis (SwIVA) [6], SwCIVA [6], and SwIVE [7].

The global permutation problem means that the order of the estimated source signals is inconsistent with the order of the input source signals. Conventional SwIVA requires a careful initialization using the oracle Acoustic Transfer Function (ATF) to ease the above problem. Obtaining the oracle ATF necessitates not only Direction-of-Arrival (DOA) information but also knowledge of other precise acoustic environment properties. Spatially Regularized Switching Independent Vector Analysis (SR-SwIVA) [8] simplifies this requirement significantly by guiding the separation matrix updates solely with the steering vector derived from DOA information. Similarly, in the speech extraction area, the Switching Constant Separating Vector for Moving Source Extraction with Geometric Constraints [9] has been proposed.

Despite these advancements, the performance of SR-SwIVA degrades in reverberant environments. To overcome this, we propose the SR-SwCIVA method, which integrates the spatial regularization and switching mechanism of SR-SwIVA with a CBF structure. This structure enables the system to model and invert the convolution introduced by room acoustics, thereby achieving effective dereverberation and superior separation performance while maintaining strong permutation robustness. Experimental results show that SR-SwCIVA provides a marked improvement in source separation performance compared to SR-SwIVA under reverberant conditions and overcomes the global permutation problem.

2. Signal Model and Problem Formulation

We consider N sources captured by an M-microphone array under a reverberant environment. The observed signal vector $\mathbf{x}_{f,t} \in \mathbb{C}^{M\times 1}$ at the TF domain (f,t) is modeled as:

$$\mathbf{x}_{f,t} = \sum_{n=1}^{N} \mathbf{h}_{n,f} s_{n,f,t} + \mathbf{l} + \mathbf{v}_{f,t}, \tag{1}$$

where $s_{n,f,t}$ and $\mathbf{h}_{n,f} \in \mathbb{C}^{M\times 1}$ are the n-th source signal and the corresponding n-th ATF vector, respectively, $\mathbf{l}$ is the late reverberation, and $\mathbf{v}_{f,t}$ is the additive noise.

The overall task is separated into two parts: dereverberation and separation.

2.1 Dereverberation

CBF-like dereverberation operation estimates the dereverberated signal $\mathbf{z}_{f,t}$:

$$\mathbf{z}_{f,t} = \mathbf{x}_{f,t} - \mathbf{G}_f^{\mathsf{H}} \bar{\mathbf{x}}_{f,t}, \tag{2}$$

where $(\cdot)^{\mathsf{H}}$ denotes the conjugate transpose operator, $\mathbf{G}_f \in \mathbb{C}^{M(L-D)\times M}$ is the dereverberation matrix, and $\overline{\mathbf{x}}_{f,t} = \left[\mathbf{x}_{f,t-D}^{\top}, \ldots, \mathbf{x}_{f,t-L+1}^{\top}\right]^{\top} \in \mathbb{C}^{M(L-D)\times 1}$ is the expanded observation vector of past time frames. The prediction delay $D \geq 1$ can guide dereverberation part to suppress the late reverberation.

2.2 Separation

The separation of $\mathbf{z}_{f,t}$ uses the separation matrix $\mathbf{W}_f \in \mathbb{C}^{M\times M}$:

$$\mathbf{y}_{f,t} = \mathbf{W}_f^{\mathsf{H}} \mathbf{z}_{f,t}, \tag{3}$$

where $\mathbf{y}_{f,t}$ is the estimated signal. To avoid the global permutation problem, a spatial regularization term is imposed on the separation filter $\boldsymbol{w}_{n,f}$ using $\boldsymbol{a}_{n,f}$:

$$\mathcal{L}_{\text{reg}}(\{\mathbf{W}_f\}) = \sum_{n=1}^{N}\sum_{f=1}^{F} ||\boldsymbol{w}_{n,f} - \boldsymbol{a}_{n,f}||_2^2, \tag{4}$$

where $\boldsymbol{w}_{n,f} \in \mathbb{C}^{M\times 1}$ is the n-th column of $\mathbf{W}_f$ and $\boldsymbol{a}_{n,f} \in \mathbb{C}^{M\times 1}$ is the steering vector derived from DOA information [8]. This ℓ_2-norm regularization aligns $\boldsymbol{w}_{n,f}$ with the steering vector $\boldsymbol{a}_{n,f}$.

3. Proposed Method

This section first introduces the switching mechanism, followed by the probabilistic model and objective function, and then concludes with a detailed discussion of the parameter optimization strategy.

3.1 Switching Mechanism

A switching mechanism is incorporated into the dereverberation and separation matrices to select a specific time-invariant matrix for each TF domain via switching values. For dereverberation, we employ a set of I time-invariant matrices $\{\mathbf{G}_f^{(i)}\}_{i=1}^{I}$. The optimal matrix $\mathbf{G}_f^{(i')}$ is selected per TF domain via the binary switching value $\gamma_{f,t}^{(i')} = 1$ ($\sum_i \gamma_{f,t}^{(i)} = 1$). Eq. (2) is thus extended to:

$$\mathbf{z}_{f,t} = \mathbf{x}_{f,t} - \sum_{i=1}^{I} \gamma_{f,t}^{(i)} \left(\mathbf{G}_f^{(i)}\right)^{\mathsf{H}} \overline{\mathbf{x}}_{f,t}. \tag{5}$$

Similarly, for separation, a set of J matrices $\{\mathbf{W}_f^{(j)}\}_{j=1}^{J}$ is applied, with the optimal matrix $\mathbf{W}_f^{(j')}$ chosen by the binary switching value $\delta_{f,t}^{(j')} = 1$ ($\sum_j \delta_{f,t}^{(j)} = 1$). Eq. (3) becomes:

$$\mathbf{y}_{f,t} = \sum_{j=1}^{J} \delta_{f,t}^{(j)} \left(\mathbf{W}_f^{(j)}\right)^{\mathsf{H}} \mathbf{z}_{f,t}. \tag{6}$$

3.2 Probabilistic Model and Objective Function

We assume the estimated source signals $y_{n,f,t}$ are mutually independent and follow a zero-mean, time-varying Gaussian model with variance $\lambda_{n,f,t}$. We define the unified binary switching weight as $\beta_{f,t}^{(i,j)} = \gamma_{f,t}^{(i)}\delta_{f,t}^{(j)}$. The overall optimization is achieved by minimizing the spatially regularized negative log-likelihood function:

$$\begin{aligned}\mathcal{L}_{(\mathcal{G},\mathcal{W},\mathbf{\Lambda},\mathcal{B})} = &\sum_{i,j,f,t}\left(\beta_{f,t}^{(i,j)}\mathcal{L}_{f,t}^{(i,j)}(\mathbf{G}_f^{(i)}, \mathbf{W}_f^{(j)}, \mathbf{\Lambda}_{f,t})\right) \\ &+ \lambda_{\text{reg}} \sum_{j,n,f} \|\boldsymbol{w}_{n,f}^{(j)} - \mathbf{a}_{n,f}\|_2^2,\end{aligned} \tag{7}$$

where λ_{reg} controls the spatial regularization strength,

$$\begin{aligned}&\mathcal{G} = \{\mathbf{G}_f^{(i)}\}_{i,f}, \mathcal{W} = \{\mathbf{W}_f^{(j)}\}_{j,f}, \mathbf{\Lambda} = \{\mathbf{\Lambda}_{f,t}\}_{f,t}, \\ &\mathbf{\Lambda}_{f,t} = \{\lambda_{n,f,t}\}_n, \mathcal{B} = \{\beta_{f,t}^{(i,j)}\}_{i,j,f,t}.\end{aligned} \tag{8}$$

The log-likelihood term $\mathcal{L}_{f,t}^{(i,j)}(\mathbf{G}_f^{(i)}, \mathbf{W}_f^{(j)}, \mathbf{\Lambda}_{f,t})$ is given by:

$$\begin{aligned}\mathcal{L}_{f,t}^{(i,j)}(\mathbf{G}_f^{(i)}, \mathbf{W}_f^{(j)}, \mathbf{\Lambda}_{f,t}) = &\sum_{n=1}^{M}\left(\frac{|y_{n,f,t}^{(i,j)}|^2}{\lambda_{n,f,t}} + \log\lambda_{n,f,t}\right) \\ &- 2\log\left|\det \mathbf{W}_f^{(j)}\right|,\end{aligned} \tag{9}$$

where $y_{n,f,t}^{(i,j)}$ is the n-th separated signal, defined as:

$$y_{n,f,t}^{(i,j)} = \left(\boldsymbol{w}_{n,f}^{(j)}\right)^{\mathsf{H}}\left(\mathbf{x}_{f,t} - \left(\boldsymbol{G}_f^{(i)}\right)^{\mathsf{H}}\bar{\mathbf{x}}_{f,t}\right). \tag{10}$$

3.3 Optimization Algorithm

We employ the coordinate descent algorithm to minimize the objective function (7) by iteratively updating the parameter sets $\mathcal{G}, \mathcal{W}, \mathbf{\Lambda}$, and $\mathcal{B}$.

3.3.1 *Initialization*

The optimization utilizes a two-stage strategy. First is SR-IVA initialization; a single separation matrix $\mathbf{W}_f$ is updated via the SR-IVA algorithm for half iterations, corresponding to $G = 1$ in Figure 1. This result is then replicated to initialize all J separation matrices $\{\mathbf{W}_f^{(j)}\}_{j=1}^{J}$ for the second stage ($G = J$). The remaining iterations are used to update SR-SwCIVA.

3.3.2 *Update of $\mathcal{G}$*

Fixing all other parameters, the cost function for $\mathcal{G}$ is simplified from (7) to:

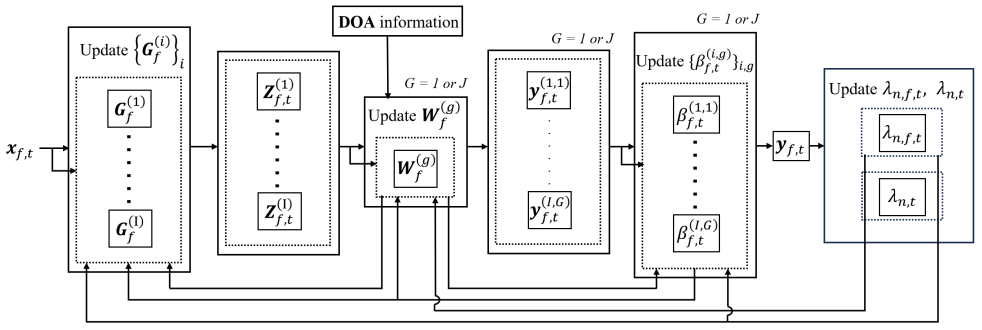

Figure 1: Flowchart of the SR-SwCIVA Optimization Procedure.

$$\mathcal{J}(\mathcal{G}) = \sum_{i,j,f,t} \beta_{f,t}^{(i,j)} \sum_{n=1}^{M} \frac{|(\boldsymbol{w}_{n,f}^{(j)})^{\mathsf{H}}(\mathbf{x}_{f,t} - (\mathbf{G}_f^{(i)})^{\mathsf{H}}\bar{\mathbf{x}}_{f,t})|^2}{\lambda_{n,f,t}}. \tag{11}$$

The update rule for $\mathbf{g}_f^{(i)} = \mathrm{vec}(\mathbf{G}_f^{(i)}) \in \mathbb{C}^{MM(L-D)\times 1}$ is given in a closed-form solution:

$$\mathbf{g}_f^{(i)} = \left(\Psi_f^{(i)}\right)^{-1} \mathrm{vec}\left(\Phi_f^{(i)}\right), \tag{12}$$

where $\Psi_f^{(i)} = \sum_{j,n} \left(\boldsymbol{w}_{n,f}^{(j)}(\boldsymbol{w}_{n,f}^{(j)})^{\mathsf{H}}\right)^* \otimes \mathbf{R}_{n,f}^{(i,j)}$ and $\Phi_f^{(i)} = \sum_{j,n} \mathbf{P}_{n,f}^{(i,j)} \left(\boldsymbol{w}_{n,f}^{(j)}(\boldsymbol{w}_{n,f}^{(j)})^{\mathsf{H}}\right)$, $(\cdot)^*$ is the complex conjugate operator, and $\otimes$ is the Kronecker product operator, with:

$$\mathbf{R}_{n,f}^{(i,j)} = \sum_{t=1}^{T} \frac{\beta_{f,t}^{(i,j)}}{\lambda_{n,f,t}} \bar{\mathbf{x}}_{f,t}\bar{\mathbf{x}}_{f,t}^{\mathsf{H}}, \quad \mathbf{P}_{n,f}^{(i,j)} = \sum_{t=1}^{T} \frac{\beta_{f,t}^{(i,j)}}{\lambda_{n,f,t}} \bar{\mathbf{x}}_{f,t}\mathbf{x}_{f,t}^{\mathsf{H}}. \tag{13}$$

3.3.3 *Update of* $\mathcal{W}$

The cost function for $\mathcal{W}$ is minimized by updating each separation filter $\boldsymbol{w}_{n,f}^{(j)}$ using the Vectorwise Coordinate Descent (VCD) algorithm [10]. The VCD update utilizes the generalized weighted covariance matrix $\mathbf{\Pi}_{n,f}^{(j)}$ and the regularization vector $\mathbf{P}_{n,f}$, defined as:

$$\mathbf{\Pi}_{n,f}^{(j)} = \begin{cases} \mathbf{\Sigma}_{n,f}^{(j)} + \lambda_{\mathrm{reg}}\mathbf{I}_M, & 1 \le n \le N, \\ \mathbf{\Sigma}_{n,f}^{(j)}, & N+1 \le n \le M, \end{cases} \tag{14}$$

$$\mathbf{P}_{n,f} = \begin{cases} \lambda_{\mathrm{reg}}\mathbf{a}_{n,f}, & 1 \le n \le N, \\ \mathbf{0}_M, & N+1 \le n \le M, \end{cases} \tag{15}$$

where $\mathbf{\Sigma}_{n,f}^{(j)} = \left(T_f^{(j)}\right)^{-1} \sum_{i,t} \frac{\beta_{f,t}^{(i,j)}}{\lambda_{n,f,t}} \mathbf{z}_{f,t}^{(i)}(\mathbf{z}_{f,t}^{(i)})^{\mathsf{H}}$ and $T_f^{(j)} = \sum_{i,t} \beta_{f,t}^{(i,j)}$. The VCD procedure then solves the optimization problem for $\boldsymbol{w}_{n,f}^{(j)}$ in a closed-form. The final update rule is:

$$\boldsymbol{w}_{n,f}^{(j)} = \begin{cases} \frac{1}{\sqrt{h_{n,f}^{(j)}}}\mathbf{u}_{n,f}^{(j)} + \tilde{\mathbf{u}}_{n,f}^{(j)}, & \text{if } \hat{h}_{n,f}^{(j)} = 0, \\ \tilde{h}_{n,f}^{(j)}\mathbf{u}_{n,f}^{(j)} + \tilde{\mathbf{u}}_{n,f}^{(j)}, & \text{otherwise,} \end{cases} \tag{16}$$

where $\mathbf{u}_{n,f}^{(j)} = \left(\left(\mathbf{W}_f^{(j)}\right)^{\mathsf{H}} \mathbf{\Pi}_{n,f}^{(j)}\right)^{-1} \mathbf{e}_n$, $\tilde{\mathbf{u}}_{n,f}^{(j)} = \left(\mathbf{\Pi}_{n,f}^{(j)}\right)^{-1} \mathbf{P}_{n,f}$, $h_{n,f}^{(j)} = \left(\mathbf{u}_{n,f}^{(j)}\right)^{\mathsf{H}} \mathbf{\Pi}_{n,f}^{(j)}\mathbf{u}_{n,f}^{(j)}$, $\hat{h}_{n,f}^{(j)} = \left(\mathbf{u}_{n,f}^{(j)}\right)^{\mathsf{H}} \mathbf{\Pi}_{n,f}^{(j)}\tilde{\mathbf{u}}_{n,f}^{(j)}$, and $\tilde{h}_{n,f}^{(j)} = \frac{\hat{h}_{n,f}^{(j)}}{2h_{n,f}^{(j)}}\left[-1 + \sqrt{1 + \frac{4h_{n,f}^{(j)}}{|\hat{h}_{n,f}^{(j)}|^2}}\right]$.

3.3.4 *Update of* $\mathbf{\Lambda}$ *and* $\mathcal{B}$

The variance parameter $\lambda_{n,f,t}$ is updated based on the power of the separated signal $y_{n,f,t}$:

$$\lambda_{n,f,t} = |y_{n,f,t}|^2 + \varepsilon, \tag{17}$$

where ε is a small positive scalar to avoid zero division during the optimization. The binary switching weight $\beta_{f,t}^{(i,j)}$ is updated by selecting the state $\{i,j\}$ that minimizes (9):

$$\beta_{f,t}^{(i,j)} = \begin{cases} 1, & \text{if } \{i,j\} = \arg\min_{\{i',j'\}} \mathcal{L}_{f,t,}^{(i',j')} \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

4. Experiments

In this section, we evaluate the performance of the proposed SR-SwCIVA method against the baseline methods SwIVA, SwCIVA, and SR-SwIVA in a reverberant acoustic environment.

4.1 Experimental Conditions

The setup involved $N = 3$ speech sources and $M = 3$ microphones, with 5 additional point noise sources. Three reverberated speech signals (from CMUARCTIC [11]) were mixed with five cafe background noises [12] at a Signal-to-Noise Ratio (SNR) of 10 dB. Directions of sources were selected from 10° to 170°. The minimum pairwise angular difference between any two speech sources was set to 40°. Room Impulse Responses (RIRs) were generated using the image method for an $8 \times 8 \times 3\,\mathrm{m}^3$-size room, reverberation time $T_{60} = 300$ ms and 600 ms. The Short-Time Fourier Transform (STFT) utilized a 512-sample (32 ms) Hanning window with a 128-sample (8 ms) shift.

All algorithms were initialized with the SR-IVA method for half iterations. They were configured with two dereverberation matrices and two separation matrices ($I = J = 2$) and were executed for $K = 50$ iterations. Performance metrics included the Improvement of Signal-to-Distortion Ratio (SDRi) and Improvement of Signal-to-Interference Ratio (SIRi), along with the Permutation Error (permE) [8].

4.2 Results and Discussion

We summarize the separation and alignment results in Table 1.

SwCIVA significantly outperformed SwIVA in SDRi and SIRi scores due to its integrated dereverberation component. However, both methods suffered from high permutation ambiguity under $T_{60} = 300$ ms and $T_{60} = 600$ ms: SwIVA reached 80% and 100% permE, and SwCIVA reached 65% and 90% permE, highlighting their lack of robust alignment despite using SR-IVA initialization.

By incorporating spatial regularization, both SR-SwIVA and the proposed SR-SwCIVA successfully achieved a 0% permE across both reverberation conditions, confirming the effectiveness of the regularization term in eliminating permutation ambiguity. While SR-SwIVA provides performance gains over SwIVA, its SDRi and SIRi scores significantly degrade at $T_{60} = 600$ ms. The proposed SR-SwCIVA consistently achieved the highest SDRi and SIRi scores under both reverberant conditions. The improvement over SR-SwIVA is particularly pronounced at $T_{60} = 600$ ms, where the CBF model effectively mitigates severe reverberation. This demonstrates the robustness and efficacy of SR-SwCIVA in source separation tasks under highly reverberant conditions.

5. Conclusions

In this paper, we proposed the SR-SwCIVA method, which consistently achieved superior source separation performance than other algorithms under challenging reverberant conditions. Critically, this enhanced performance was attained while retaining the strong robustness against the global permutation problem.

Table 1: Results of compared methods

Method	Reverb Time	SDRi [dB]	SIRi [dB]	permE [%]
SwIVA	300 ms	6.14	14.17	80
SwCIVA	300 ms	7.89	19.30	65
SR-SwIVA	300 ms	8.27	19.29	**0**
SR-SwCIVA (proposed)	300 ms	**9.19**	**24.65**	**0**
SwIVA	600 ms	6.38	15.98	100
SwCIVA	600 ms	7.92	19.12	90
SR-SwIVA	600 ms	6.96	17.03	**0**
SR-SwCIVA (proposed)	600 ms	**8.53**	**21.99**	**0**

References

[1] S. Makino, Audio Source Separation. Springer, 2018.

[2] S. Makino, S. Araki, R. Mukai, and H. Sawada, "Audio source separation based on independent component analysis," in Proc. ISCAS, vol. 5, 2004, pp. V-668–V-671.

[3] 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.

[4] R. Scheibler and N. Ono, "Independent vector analysis with more microphones than sources," in Proc. IEEE WASPAA, 2019, pp. 185–189.

[5] 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.

[6] T. Nakatani, R. Ikeshita, K. Kinoshita, H. Sawada, N. Kamo, and S. Araki, "Switching independent vector analysis and its extension to blind and spatially guided convolutional beamforming algorithms," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 30, pp. 1032-1047, Mar. 2022.

[7] T. Ueda, T. Nakatani, R. Ikeshita, S. Araki, and S. Makino, "DOA informed switching independent vector extraction and beamforming for speech enhancement in underdetermined situations," EURASIP J. Audio, Speech, Music Process., vol. 2024, no. 52, pp. 1–20, Oct. 2024.

[8] T. Ueda, T. Nakatani, R. Ikeshita, S. Araki, and S. Makino, "Spatially-regularized switching independent vector analysis," in Proc. APSIPA ASC, 2023, pp. 2024–2030.

[9] C. Chen, Y. Yang, Y. Zhao, S. Makino, and J. Chen, "Switching constant separating vector for moving source extraction with geometric constraints," in Proc. APSIPA ASC, 2025, pp. 13–18.

[10] 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.

[11] J. Kominek and A. W. Black, "The CMU arctic speech databases," in Proc. ISCA SSW, 2004, pp. 223–224.

[12] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'CHiME' speech separation and recognition challenge: Dataset, task and baselines," in Proc. ASRU, 2015, pp. 504–511.