

Article

Efficient Overdetermined Independent Vector Analysis Based on Iterative Projection with Adjustment

Ruiming Guo ¹, Zhongqiang Luo ^{1,2,3,*} , Ling Wang ¹ and Li Feng ^{2,4,*} 

- ¹ School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin 644000, China; 321081104113@stu.suse.edu.cn (R.G.); 322081104107@stu.suse.edu.cn (L.W.)
- ² Engineering Research Center of Integration and Application of Digital Learning Technology, Ministry of Education, Beijing 100039, China
- ³ Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin 644000, China
- ⁴ School of Engineering and Technology, The Open University of Sichuan, Chengdu 610073, China
- * Correspondence: luozhongqiang@suse.edu.cn (Z.L.); fengl@scou.cn (L.F.)

Abstract: In this paper, a computationally efficient optimization algorithm for independent vector analysis (IVA) is proposed to accelerate iterative convergence speed and enhance the overdetermined convolutive blind speech separation performance. An iterative projection with adjustment (IPA) is investigated to estimate the unmixing matrix for OverIVA. The IPA algorithm jointly executes the iterative projection (IP) algorithm and the iterative source steering (ISS) algorithm to jointly update one row and one column of the mixing matrix, which can perform computationally-efficient blind source separation. It is achieved by updating one demixing filter and jointly adjusting all the other sources along its current direction. Motivated by its technology superiorities, this paper proposes a modified algorithm for the OverIVA, fully exploiting the computational efficiency of IPA optimization scheme. Experimental results corroborate the proposed OverIVA-IPA algorithm converges faster and performs better than the existing state-of-the-arts algorithms.

Keywords: blind source separation; independent vector analysis; optimization methods; speech separation



Citation: Guo, R.; Luo, Z.; Wang, L.; Feng, L. Efficient Overdetermined Independent Vector Analysis Based on Iterative Projection with Adjustment. *Electronics* **2023**, *12*, 3200. <https://doi.org/10.3390/electronics12143200>

Academic Editor: Javid Taheri

Received: 26 May 2023

Revised: 4 July 2023

Accepted: 20 July 2023

Published: 24 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Blind source separation (BSS) [1] refers to unmixing or extracting the latent sources from the observed mixed signals with minimal prior information. It has become a versatile technology with diverse applications, such as in speech signals [2,3], biomedical signals [4] and digital communication signals [5,6]. Independent component analysis (ICA) [7,8] is one of the most basic means proposed to deal with BSS. ICA is an unsupervised, data-driven blind separation technique for separating linear mixture signals based on non-Gaussian maximization. The frequency domain independent component analysis (FD-ICA) model [9] is proposed for convolutional mixed signals to overcome the high computational complexity of directly implementing ICA in the time domain processing. In FD-ICA, the observed signal is converted from the time domain to the frequency domain representation through the short-time Fourier transform (STFT), and then ICA is applied to estimate the unmixing matrix at each frequency. However, FD-ICA will suffer from the random permutation ambiguity problem.

To solve the permutation ambiguity of ICA, independent vector analysis (IVA) [10,11] has been proposed and gained remarkable attention from scholars. IVA is an extension of ICA for the separation of multiple parallel mixtures. It resolves the random permutation ambiguity of signal separation outputs by exploiting statistical dependencies across datasets to generalize ICA to multiple datasets. IVA preserves the statistical dependency within a frequency source vector and minimizes statistical dependencies between them.

IVA can naturally solve the random permutation problem without any pre-processing or post-processing during the learning process. The traditional IVA algorithm updates the separation matrix through the method based on gradient [10] and fast fixed-point algorithm [12]. The method based on gradient update needs to adjust the parameters such as step size to make the convergence stable, and it needs to balance the convergence speed and stability. To achieve faster convergence, a hyperparameter-free iterative projection (IP) algorithm based on auxiliary-function-based IVA (AuxIVA) was proposed [13]. Recently, fast-converging optimization algorithms have been proposed for AuxIVA, for example, IP2 [14], iterative source steering (ISS) [15], ISS2 [16], and iterative projection with adjustment (IPA) [17]. IPA is the combination of IP and ISS, which solves the problem that IP and ISS can only correct the update in the next iteration when performing the update. This algorithm is superior to other algorithms in terms of convergence speed and performance.

In the case of overdetermined BSS, where the number of non-stationary signals N is smaller than the number of microphones M , i.e., $M > N$. In the multi-source case ($N \geq 2$), the initial approach is to resolve the oversubscription situation by selecting the N best channels [18,19] or reducing the number of channels to N by principal component analysis (PCA) [20,21]. Unfortunately, these methods risk removing the source signal of interest and reducing separation performance. For the case of a single source, several independent vector extraction (IVE) methods [22,23] have been proposed. And on this basis, IVA is extended to the overdetermined situation ($M > N$), and the overdetermined independent vector analysis algorithm (OverIVA) [24] is proposed. The traditional OverIVA relies on the orthogonality constraint (OC), which ignores the sample correlation between the target source signal and the noise signal and makes the limited separation. To solve this problem, an OverIVA [25] is proposed which only utilizes the independence between source signals and the stationarity of Gaussian noise for source separation. Recently, algorithms such as IP and ISS have been combined with OverIVA to achieve efficient overdetermined BSS [24–27].

In this paper, an efficient approach for BSS is proposed, which we call OverIVA-IPA. It combines the technology of OverIVA-IP and OverIVA-ISS with the technology of AuxIVA-IPA. It is an algorithm that can achieve high efficiency and ensure convergence. IP and ISS fix all other sources while doing one of the updates. This means that further correction can only happen at the next iteration. The IPA combines the advantages of the IP and ISS technologies to jointly update the mixing matrix of the source signal. As opposed to IP and ISS, when updating the demixing filter of one source, we simultaneously correct the demixing filters of all other sources accordingly. Therefore, we apply the modified IPA technique to the OverIVA algorithm to update the source part of the demixing matrix as well as the orthogonal noise part. Finally, we validate it in our convolutional speech separation experiments. Experimental results show that the OverIVA-IPA method has faster convergence speed and better performance than the existing OverIVA-IP, OverIVA-IP2, and OverIVA-ISS methods.

The rest of this paper is organized as follows. We describe the background of the overdetermined BSS problem, AuxIVA-IP, OverIVA-IP, AuxIVA-ISS, OverIVA-ISS, and AuxIVA-IPA in Section 2. In Section 3, the proposed algorithm is derived and the time complexity of the algorithm is analyzed. In Section 4, we show the comparative experimental results of different algorithms and conduct related analysis. Section 5 concludes the full text.

2. Background

2.1. Overdetermined Blind Source Separation Model

In general, the model of the BSS algorithm consists of a cost function and an optimization method. The cost function of BSS is constructed according to the characteristics of the limited source and the separation criterion. The purpose of BSS is to find a suitable linear transformation matrix or separation matrix W by optimizing the cost function. The BSS separation process usually estimates the separation matrix and then restores the source signal by estimating the separation matrix. In BSS, usually according to the relationship

between the number of sensors sending and receiving signals, BSS system models can be divided into three categories: determined model ($M = N$), underdetermined model ($M < N$), and overdetermined model ($M > N$). This paper mainly studies the method of optimizing the cost function in the BSS overdetermined model. Assume that under overdetermined conditions ($M > N$), M microphones observe N signal sources. Settings $f = 1, \dots, F$ and $t = 1, \dots, T$ denote the frequency bin index and the time frame index, respectively. After STFT of the observed multi-channel signal, the microphone signal $\mathbf{x}_t^f = [x_{1t}^f, \dots, x_{Mt}^f]^T \in \mathbb{C}^M$ at frequency f and time t is modeled as

$$\mathbf{x}_t^f = \mathbf{A}^f \mathbf{s}_t^f + \boldsymbol{\psi}^f \mathbf{z}_t^f, \quad (1)$$

where $\mathbf{s}_t^f = [s_{1t}^f, \dots, s_{Nt}^f]^T \in \mathbb{C}^N$ is the source signal, $\mathbf{z}_t^f \in \mathbb{C}^{M-N}$ is the noise vector, $\mathbf{A}^f \in \mathbb{C}^{M \times N}$ and $\boldsymbol{\psi}^f \in \mathbb{C}^{M \times (M-N)}$ denote their corresponding mixing matrices, respectively. Our goal is to estimate the unmixing matrix $\tilde{\mathbf{W}}^f \in \mathbb{C}^{M \times M}$ to recover the original vector $\mathbf{s}_t^f$ from the observed signal

$$\begin{bmatrix} \mathbf{s}_t^f \\ \boldsymbol{\Phi}^f \mathbf{z}_t^f \end{bmatrix} = \tilde{\mathbf{W}}^f \mathbf{x}_t^f, \quad (2)$$

where matrix $\boldsymbol{\Phi}^f$ can be an arbitrary reversible linear transformation, indicating that our focus is not to separate the noise vector $\mathbf{z}_t^f$. The parameter $\boldsymbol{\Phi}^f$ is chosen to simplify the processing task such that there is

$$\begin{bmatrix} \mathbf{W}^f \\ \mathbf{U}^f \end{bmatrix} \mathbf{x}_t^f = \tilde{\mathbf{W}}^f \mathbf{x}_t^f, \quad (3)$$

where $\mathbf{W}^f = [w_1^f, \dots, w_N^f]^H \in \mathbb{C}^{M \times N}$ and $\mathbf{U}^f = [\mathbf{J}^f - \mathbf{I}_{M-N}] \in \mathbb{C}^{M \times (M-N)}$ denote the unmixing matrices of the source and noise, respectively, and $\mathbf{J}^f \in \mathbb{C}^{M \times (M-N)}$. In this paper, $(\cdot)^H, (\cdot)^T, \det(\cdot), |\cdot|, (\cdot)^*$ and $(\cdot)^{-1}$ denote the conjugate transpose, transpose, determinant, absolute value, complex conjugate and inverse of $(\cdot)$, respectively.

2.2. Cost Function

In previous studies [24,25], it is usually assumed that the source signal $\mathbf{s}_t^f = [s_{1t}^f, \dots, s_{Nt}^f]^T \in \mathbb{C}^N$ follows a non-Gaussian distribution, such as a circularly symmetric Laplace distribution or a time-varying Gaussian distribution. In this paper, the Laplace source prior model is used, with

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}. \quad (4)$$

Among them, we set the parameter expectation $\mu = 0$, variance $2b^2 = 2$, skewness to 0, and kurtosis to 3. And assuming that the noise signal $\mathbf{z}_t^f \in \mathbb{C}^{M-N}$ follows a stationary Gaussian distribution, we have

$$p(\mathbf{z}_t^f) = \frac{1}{\pi^{M-N} |\det(\mathbf{R}^f)|} e^{-(\mathbf{z}_t^f)^H (\mathbf{R}^f)^{-1} \mathbf{z}_t^f}, \quad (5)$$

where $\mathbf{R}^f$ is the location-space covariance matrix after noise separation. Also, the separated background noise is statistically independent in frequency. By using the distribution of the source and noise signals and ignoring all constants. We can obtain the negative log-likelihood function (objective function) of the observed data, which is given by

$$\begin{aligned} \ell_{OverIVA}^1 = & -2 \sum_{f=1}^F \log |\det(\tilde{\mathbf{W}}^f)| + \frac{1}{T} \sum_{n=1}^N \sum_{t=1}^T G\left(\sqrt{\sum_{f=1}^F |(\tilde{\mathbf{w}}_n^f)^H \mathbf{x}_t^f|}\right) \\ & + \sum_{f=1}^F ((\log \det(\mathbf{R}^f)) + \text{Tr}(\mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{U}^f)^H (\mathbf{R}^f)^{-1})), \end{aligned} \quad (6)$$

where $G(\cdot)$ is a contrast function determined by the distribution of the source signal $\mathbf{s}_t^f$ and $\mathbf{\Gamma}^f = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t^f (\mathbf{x}_t^f)^H$ is the sample covariance matrix of the observed signal, and $\text{Tr}(\cdot)$ is the trace of the matrix. Because the above formula is difficult to directly find the minimum value of $\mathbf{W}^f$, on the contrary, the upper bound of the cost function can be minimized, there is

$$\begin{aligned} \ell_{OverIVA}^2 = & - \sum_{f=1}^F \log |\det(\tilde{\mathbf{W}}^f)| + \sum_{n=1}^N \sum_{f=1}^F ((\tilde{\mathbf{w}}_n^f)^H \mathbf{V}_n^f \tilde{\mathbf{w}}_n^f) \\ & + \sum_{f=1}^F ((\log \det(\mathbf{R}^f)) + \text{Tr}(\mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{U}^f)^H (\mathbf{R}^f)^{-1})), \end{aligned} \quad (7)$$

where

$$\mathbf{V}_n^f = \frac{1}{T} \sum_{t=1}^T \phi(\mathbf{r}_{nt}) \mathbf{x}_t^f (\mathbf{x}_t^f)^H, \quad (8)$$

where $\phi(\mathbf{r}_{nt})$ depends on the definition of the contrast function $G(\cdot)$, $\mathbf{r}_{nt} \in \mathbb{R}$ is a variable with $\mathbf{r}_{nt} \leftarrow \sum_{f=1}^F |(\tilde{\mathbf{w}}_n^f)^H \mathbf{x}_t^f|^2$. Where $\mathbf{x}_t^f = [x_{1t}^f, \dots, x_{Mt}^f]^T \in \mathbb{C}^M$ represents the microphone signal at frequency f and time t , and $\tilde{\mathbf{w}}_n^f$ is the n th unmixing vector at frequency f in the unmixing matrix $\tilde{\mathbf{W}}^f$. It should be noted that the random variables in the cost function are multivariate, and each source signal is also multivariate. During the separation process, the optimization algorithm needs to maintain the statistical dependence within each source vector, and at the same time minimize the statistical dependence between different source vectors, so as to avoid problems such as arrangement ambiguity and separate the source signals.

2.3. ISS and IP of OverIVA and AuxIVA

In the overdetermined case, Ref. [24] conjectures that the strongest source signal has a very non-Gaussian distribution, while noise mixed with other weaker sources will have a distribution that is closer to Gaussian, thus guaranteeing linear independence between vectors (In this paper, the source signal adopts Laplace distribution). There is no doubt that ISS and IP are hybrid matrix updates that can be directly applied to determined situations. However, using these update methods in the overdetermined situation cannot directly extract the target signal from the noise subspace. The lower part $\mathbf{J}^f$ of the matrix needs to be modified so that the noise subspaces remain orthogonal.

To effectively derive the parameter estimation algorithm of OverIVA, it is necessary to refer to the proven Propositions derived from [26,28].

Proposition 1. For any local optimum, a new $\mathbf{U}^f$ can be found without changing the value of the cost function (6).

Proposition 2. $\mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{W}^f)^H = \mathbf{0}_{(M-N) \times N}$, $\mathbf{R}^f = \mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{U}^f)^H$ are the necessary conditions for the optimal solution.

According to the guidance of Proposition 1 above, Formula (3) can be simplified as follows:

$$\begin{bmatrix} \mathbf{W}^f \\ \mathbf{J}^f - \mathbf{I}_{M-N} \end{bmatrix} = \begin{bmatrix} \mathbf{W}^f \\ \mathbf{U}^f \end{bmatrix} = \tilde{\mathbf{W}}^f. \quad (9)$$

Our goal is obvious, to estimate $\mathbf{W}^f$ and $\mathbf{J}^f$ ($M < N$) or only $\mathbf{W}^f$ ($M = N$) that minimized (6).

2.4. Iterative Projection

In OverIVA-IP [25], the rows of the optimized unmixing matrix $\mathbf{W}^f$ are updated in regular order by IP [14]. Its updated rule is

$$\tilde{\mathbf{w}}_k^f \leftarrow (\tilde{\mathbf{W}}^f \mathbf{V}_k^f)^{-1} \mathbf{e}_k (\mathbf{e}_k^T (\tilde{\mathbf{W}}^f \mathbf{V}_k^f \tilde{\mathbf{W}}^H)^{-1} \mathbf{e}_k)^{-1/2}, k = 1, \dots, N. \quad (10)$$

Based on Proposition 2, $\mathbf{R}^f$ is fixed to $\mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{U}^f)^H$, and the update of $\mathbf{J}^f$ must satisfy the orthogonal constraint $\mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{W}^f)^H = \mathbf{0}_{(M-N) \times N}$ between the source subspace and the noise subspace. Due to the equivalence relation of Formula (9), $\mathbf{U}^f \mathbf{\Gamma}^f (\mathbf{W}^f)^H = \mathbf{0}_{(M-N) \times N}$ can be expressed as

$$(\mathbf{J}^f \mathbf{E}_1 - \mathbf{E}_2) \mathbf{\Gamma}^f (\mathbf{W}^f)^H = \mathbf{0}_{(M-N) \times N}, \quad (11)$$

where $\mathbf{E}_1$ and $\mathbf{E}_2$ denote $[\mathbf{I}_N \mathbf{0}_{(M-N) \times N}]$ and $[\mathbf{0}_{(M-N) \times N} \mathbf{I}_{M-N}]$, respectively. The update rule for $\mathbf{J}^f$ is

$$\mathbf{J}^f \leftarrow (\mathbf{E}_2 \mathbf{\Gamma}^f (\mathbf{W}^f)^H) (\mathbf{E}_1 \mathbf{\Gamma}^f (\mathbf{W}^f)^H)^{-1}. \quad (12)$$

2.5. Iterative Source Steering

In OverIVA-ISS [27], the columns of the optimized unmixing matrix $\mathbf{W}^f$ in regular order are updated by ISS [15]. Its updated rule is

$$\mathbf{W}^f \leftarrow \mathbf{W}^f - \mathbf{v}_k^f (\tilde{\mathbf{w}}_k^f)^H, k = 1, \dots, M, \quad (13)$$

$$\mathbf{J}^f \leftarrow (\mathbf{E}_2 \mathbf{\Gamma}^f (\mathbf{W}^f)^H) (\mathbf{E}_1 \mathbf{\Gamma}^f (\mathbf{W}^f)^H)^{-1}, \quad (14)$$

where $\mathbf{v}_k^f = [v_{k1}^f, \dots, v_{kN}^f]^T \in \mathbb{C}^N$ is calculated by

$$v_{kn}^f = \begin{cases} \frac{(\tilde{\mathbf{w}}_n^f)^H \mathbf{V}_n^f \tilde{\mathbf{w}}_k^f}{(\tilde{\mathbf{w}}_k^f)^H \mathbf{V}_n^f \tilde{\mathbf{w}}_k^f} & \text{if } n \neq k, \\ 1 - ((\tilde{\mathbf{w}}_n^f)^H \mathbf{V}_n^f \tilde{\mathbf{w}}_n^f)^{-1/2} & \text{if } n = k. \end{cases} \quad (15)$$

2.6. Iterative Projection with Adjustment

In AuxIVA-IPA [17], the entire mixing matrix $\tilde{\mathbf{W}}^f = \mathbf{W}^f$ jointly performs IP-style and ISS-style updates. It completely re-estimates the k -th unmixing filter and adjusts the values of all other filters by taking steps consistent with the current estimate of source k . Its updated rule is

$$\tilde{\mathbf{W}}^f \leftarrow \mathbf{T}_k(\mathbf{u}, \mathbf{q}) \tilde{\mathbf{W}}^f, \quad k = 1, \dots, M, \quad (16)$$

where $\mathbf{W}^f$ is the estimate of the separation matrix from the previous iteration, while $\mathbf{T}_k(\mathbf{u}, \mathbf{q})$ is the method of each vector update of the mixing matrix, with

$$\mathbf{T}_k(\mathbf{u}, \mathbf{q}) = \mathbf{I} + \mathbf{e}_k (\mathbf{u}^H - \mathbf{e}_k^T) + \tilde{\mathbf{E}}_k \mathbf{q}^* \mathbf{e}_k^T, \quad (17)$$

Update one row and one column of the mixing matrix in each iteration by definition, where $\bar{\mathbf{E}}_k$ is the $M \times (M - 1)$ matrix containing all regular basis vectors except the k th vector.

$$\bar{\mathbf{E}}_k = [e_1 \quad \cdots \quad e_{k-1} \quad e_{k+1} \quad \cdots \quad e_M]. \quad (18)$$

where I denotes the identity matrix and $\mathbf{e}_k$ denotes the k th unit vector. For the update of the column vector $\mathbf{q}$ has

$$\min_{\mathbf{q} \in \mathbb{C}^{M-1}} (\mathbf{q} + \mathbf{A}^{-1}\mathbf{b})^H \mathbf{A} (\mathbf{q} + \mathbf{A}^{-1}\mathbf{b}) - \log((\mathbf{q} - \mathbf{C}^{-1}\mathbf{g})^H \mathbf{C} (\mathbf{q} - \mathbf{C}^{-1}\mathbf{g}) + o) \quad (19)$$

with

$$\mathbf{A} = \text{diag}(\dots, \mathbf{e}_k^T \mathbf{V}_m \mathbf{e}_k, \dots), m = 1, \dots, M, m \neq k, \quad (20)$$

$$\mathbf{b} = [\dots, \mathbf{e}_k^T \mathbf{V}_m \mathbf{e}_k, \dots]^T, m \neq k, \quad (21)$$

$$\mathbf{C} = \bar{\mathbf{E}}_k^T (\mathbf{V}_k^{-1})^* \bar{\mathbf{E}}_k, \quad (22)$$

$$\mathbf{g} = \bar{\mathbf{E}}_k^T (\mathbf{V}_k^{-1})^* \mathbf{e}_k, \quad (23)$$

$$o = \mathbf{e}_k^T (\mathbf{V}_k^{-1})^* \mathbf{e}_k - \mathbf{g}^H \mathbf{C}^{-1} \mathbf{g}. \quad (24)$$

Through the above formula, we can find the optimal solution of the column vector $\mathbf{q}$ in the mixing matrix. Among them, $\mathbf{V}_k$ denotes the k th weighted covariance matrix, with $\mathbf{V}_k \leftarrow ((\mathbf{W}^f \mathbf{V}_k (\mathbf{W}^f)^H)^{-1})^*$. $\mathbf{A}$ and $\mathbf{C}$, respectively, denote the matrix variables of the m th and k th weighted covariance matrices $\mathbf{V}_m$ and $\mathbf{V}_k$ after corresponding transformation. $\mathbf{b}$ and $\mathbf{g}$ denote the vector variables after the corresponding transformation of the m th and k th weighted covariance matrices $\mathbf{V}_m$ and $\mathbf{V}_k$. o denotes the variable value obtained by corresponding transformation of the k th weighted covariance matrix $\mathbf{V}_k$. The update of the row vector $\mathbf{u}$ is

$$\mathbf{u}_k = \frac{\mathbf{V}_k^{-1} \tilde{\mathbf{q}}}{\sqrt{\tilde{\mathbf{q}}_k^H \mathbf{V}_k^{-1} \tilde{\mathbf{q}}_k}} e^{j\theta}, \quad (25)$$

where $\tilde{\mathbf{q}}_k = \mathbf{e}_k - \bar{\mathbf{E}}_k \mathbf{q}^*$, $\theta \in [0, 2\pi]$ is any phase. The optimal solution of $\mathbf{q}$ can be calculated through (19)–(24), and then we can solve the optimal solution of the row vector $\mathbf{u}$ in the mixing matrix by bringing it into (25). The optimal separation matrix can be obtained by updating the mixing matrix through the optimal solution column vector $\mathbf{q}$ and solution row vector $\mathbf{u}$.

3. Proposed Method

3.1. OverIVA-IPA

In the process of blind source separation, the key to the fast separation of signals is to use fewer iterations to reduce the cost function (7) more, thereby improving the separation efficiency. The previously proposed block coordinate descent algorithm IP, IP2 and ISS fix part of the separation matrix and then minimize the cost function on the remaining free variables to separate the signal. Wherein IP and ISS update one row or one column of the unmixing matrix each time, and IP2 updates two rows of the unmixing matrix each time. However, when the separation effect of other source vectors is not good, it may cause poor overall separation performance. We propose an IPA-based OverIVA algorithm. In OverIVA-IP and OverIVA-ISS, updating row-by-row or column-by-column naturally

allows for the separation of the target sources one by one, requiring only further updating of the background noise. In contrast, the IPA algorithm jointly performs IP-style and ISS-style updates to achieve a more efficient BSS. We are thus inspired that the proposed algorithm combines the convergence advantage of the IPA algorithm with the orthogonality constraint of OverIVA. Through the following update method until the cost function can converge to a stable point

$$\mathbf{W}^f \leftarrow \mathbf{T}_k(\mathbf{u}, \mathbf{q})\mathbf{W}^f, \quad k = 1, \dots, M, \quad (26)$$

$$\mathbf{J}^f \leftarrow (\mathbf{E}_2 \mathbf{\Gamma}^f (\mathbf{W}^f)^H) (\mathbf{E}_1 \mathbf{\Gamma}^f (\mathbf{W}^f)^H)^{-1}, \quad (27)$$

where $\mathbf{T}_k(\mathbf{u}, \mathbf{q})$ is given by (16)–(25). Applying the IPA method only to the source part and using the orthogonal constraint to update the remaining noise part to solve the IPA method cannot be directly applied to the entire matrix. In the update process, the IPA method can ensure that each iteration can ensure the proper optimization of the cost function until the final convergence. For the initial value of the matrix $\mathbf{W}^f$, we find that it can be set as the identity matrix to be satisfactory. The final algorithm OverIVA-IPA, which alternately applies updates to $\mathbf{W}^f$ and $\mathbf{J}^f$, is detailed in Algorithm 1.

Algorithm 1 OverIVA by IPA

Input: Microphones signals $\{\mathbf{x}_{tm}^f\}_{f=1, t=1, m=1}^{F, T, M}$

Output: Updated matrix $\mathbf{W}$

for loop $\leftarrow 1$ **to** max.iterations **do**

for $n = 1 : N, t = 1 : T$ **do**

$$\mathbf{r}_{nt} \leftarrow \sum_{f=1}^F |(\mathbf{w}_k^f)^H \mathbf{x}_t^f|^2,$$

for $f = 1 : F$ **do**

$$\mathbf{A} = \text{diag}(\dots, \mathbf{e}_k^T \mathbf{V}_m \mathbf{e}_k, \dots), m = 1, \dots, M \& m \neq k,$$

$$\mathbf{b} = [\dots, \mathbf{e}_k^T \mathbf{V}_m \mathbf{e}_k, \dots]^T, m \neq k,$$

$$\mathbf{C} = \tilde{\mathbf{E}}_k^T (\mathbf{V}_k^{-1})^* \tilde{\mathbf{E}}_k,$$

$$\mathbf{g} = \tilde{\mathbf{E}}_k^T (\mathbf{V}_k^{-1})^* \mathbf{e}_k,$$

$$\mathbf{o} = \mathbf{e}_k^T (\mathbf{V}_k^{-1})^* \mathbf{e}_k - \mathbf{g}^H \mathbf{C}^{-1} \mathbf{g},$$

$$\mathbf{V}_k = ((\mathbf{W}^f \mathbf{V}_k (\mathbf{W}^f)^H)^{-1})^*,$$

 Update $\mathbf{q}$ using (19),

 Update $\mathbf{u}$ using (25),

$$\mathbf{W}^f \leftarrow (\mathbf{I} + \mathbf{e}_k (\mathbf{u}^H - \mathbf{e}_k^T) + \tilde{\mathbf{E}}_k \mathbf{q}^* \mathbf{e}_k^T) \mathbf{W}^f, \quad k = 1, \dots, M,$$

$$\mathbf{J}^f \leftarrow (\mathbf{E}_2 \mathbf{\Gamma}^f (\mathbf{W}^f)^H) (\mathbf{E}_1 \mathbf{\Gamma}^f (\mathbf{W}^f)^H)^{-1}.$$

3.2. Computational Complexity

When the number T of time frames is greater than the number M of microphones, the running time is determined by the computation of the weighted covariance matrix $\mathbf{V}_n^f$. In this case, the weighted covariance matrix $\mathbf{V}_n^f$ of N sources is calculated in each iteration, and the computational complexity of OverIVA-IP and OverIVA-IP2 is $O(FTNM^2)$. The IPA algorithm does not increase the computational complexity in essence and requires a matrix inversion, two matrix multiplications, and an eigenvalue decomposition. The computational complexity of OverIVA-IPA is $O(FTNM^2)$. However, OverIVA with ISS has the particularity that there is an efficient computation of (13), and the complexity is $O(FTNM)$.

4. Numerical Experiment

We compare the performance of our proposed OverIVA-IPA algorithm with existing OverIVA-IP, OverIVA-ISS, and OverIVA-IP2 when applied to convolutional blind source separation in the frequency-domain STFT. We evaluated the performance of the algorithm in terms of Δ SI-SIR, Δ SI-SDR, and separated spectrograms. And it is an experimental comparison carried out under different numbers of sources and different values of signal-to-interference-and-noise ratios (SINR).

4.1. Experimental Environment Settings

To synthesize the mixed signal for evaluation, we simulate the impulse reverberation of 1000 random 3D matrix rooms by using the pyroomacoustics Python package [29]. The three-dimensional matrix room has walls of 6 and 10 m in length and a ceiling height of 2.8 to 4.5 m. The simulated reverb time is sampled uniformly between approximately 60 ms and 450 ms. The source and microphone arrays are randomly placed at least 50 cm away from the wall, and the height is between 1 and 2 m. The array is circular and regular. As shown in Figure 1.

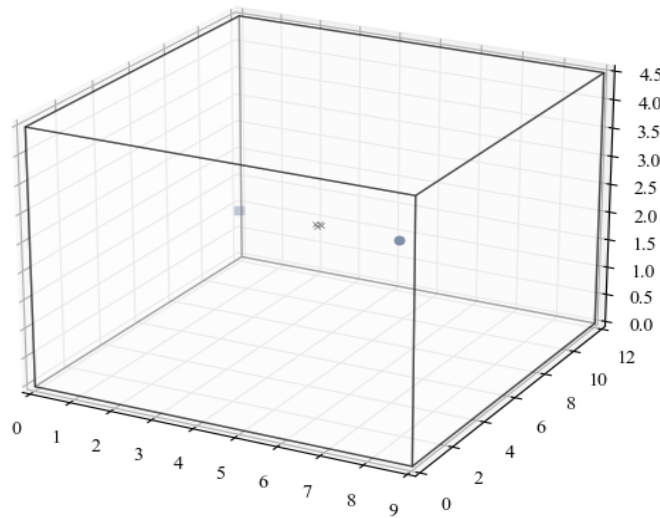


Figure 1. Simulation 3D matrix room.

The three axes of the 3D matrix room in Figure 1 denote the length, width, and height of the room, respectively. Where $\times$ denotes the microphone array, the $\square$ denotes the source signal, and $\circ$ denotes the interferer signal. The number of sources is set to $N = 2$, the source signal is selected from the CNU Arctic CORPUS speech database [30], and 5 additional interference sources are selected to generate diffuse noise. The number of microphones is $M = 4, 6, 8$, and the distance between adjacent microphones is 10 cm. All sound sources are located farther from the array than the critical distance of the room, which is the distance at which direct sound and reverberant energy are equal. This distance can be calculated by

$$d = 0.057 \sqrt{V/T_{60}}, \quad (28)$$

where V is the volume of the room. SINR is defined as

$$\text{SINR} = \frac{\sum_{n=1}^N \sigma_n^2}{Q\sigma_i^2 + \sigma_w^2}, \quad (29)$$

where σ_n^2 , σ_i^2 , and σ_w^2 are the variances of the target source, interferer, and white noise, respectively, for which the specified SINR can be obtained on any reference microphone. After simulating propagation, the variance of the target source is fixed at $\sigma_n^2 = 1$ (at an arbitrary reference microphone). In the comparison experiment, the first microphone is selected as a reference, and its SINR value is fixed. The separation effects at 5 dB, 15 dB,

and 25 dB SINR values are studied. Simulations were performed at 16 kHz, using a 4096 Hamming window with STFT overlapping 3/4.

4.2. Experimental Simulation Results

In the experimental simulation, the performance of various OverIVA algorithms is evaluated using the multivariate Laplacian source prior model. We tested the OverIVA algorithm optimized by IP, IP2, ISS, and IPA methods. We use scale-invariant signal-to-distortion ratio (SI-SDR), scale-invariant signal-to-interference ratio (SI-SIR), and signal spectrogram as our separation performance metrics. SI-SDR measures how much the target signal is degraded, while SI-SIR indicates how much of the other sources remain. High SI-SDR indicates both good separation and high quality. High SI-SIR indicates good separation, but not necessarily preservation of the target source. They are defined as follows. Let $\mathbf{S} \in \mathbb{R}^{T \times M}$ be the matrix containing the M time-domain groundtruth reference signals in its columns. Let $\hat{\mathbf{s}} \in \mathbb{R}^T$ be the estimated signal, and $\mathbf{s}$ one of the columns of $\mathbf{S}$. Then, the definition is as follows.

$$\text{SI-SDR}(\mathbf{s}, \hat{\mathbf{s}}) = \frac{\|\alpha \mathbf{s}\|^2}{\|\alpha \mathbf{s} - \hat{\mathbf{s}}\|^2}, \quad \text{SI-SIR}(\mathbf{s}, \hat{\mathbf{s}}) = \frac{\|\alpha \mathbf{s}\|^2}{\|\mathbf{S} \mathbf{b}\|^2}, \quad (30)$$

where

$$\alpha = \frac{\hat{\mathbf{s}}^T \mathbf{s}}{\|\mathbf{s}\|^2}, \quad \text{and} \quad \mathbf{b} = (\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T (\alpha \mathbf{s} - \hat{\mathbf{s}}). \quad (31)$$

Figures 2 and 3 show the separation performance of OverIVA-IP, OverIVA-IP2, OverIVA-ISS, and OverIVA-IPA. Where, Figure 2 uses $\Delta \text{SI-SDR}$ as the performance index, and Figure 3 uses $\Delta \text{SI-SIR}$ as the performance index. Through the analysis of various performance indicators, it can be seen that the proposed OverIVA-IPA method is superior to other methods in almost all experimental environments. Among them, in the 5 dB and 15 dB environments, the performance of the algorithm is superior. The proposed OverIVA-IPA algorithm is superior to other algorithms in performance, and it is the fastest algorithm to reach higher SI-SDR values and SI-SIR values, where the performance of the six microphones in Figure 2 is comparable to that of OverIVA-IP2 in the 15 dB environment. Table 1 shows that the response speed of OverIVA-IPA is the fastest, and quickly reaches a stable value of SI-SDR with the least number of iterations, and its computational efficiency is several times higher than other algorithms. Table 2 shows the stabilized SI-SDR values of the four methods. The results show that IPA is as effective as other methods in minimizing the cost function, performs better, and converges faster.

Table 1. Number of iterations required for the SI-SDR value to converge within tolerance 0.1 dB.

Method	4 Mics			6 Mics			8 Mics		
	Iterations			Iterations			Iterations		
	5 dB	15 dB	25 dB	5 dB	15 dB	25 dB	5 dB	15 dB	25 dB
OverIVA-IP	35	20	15	27	16	13	26	16	12
OverIVA-IP2	10	8	7	9	8	8	9	9	8
OverIVA-ISS	29	17	13	28	15	13	26	14	12
OverIVA-IPA	7	5	5	6	5	5	6	6	5

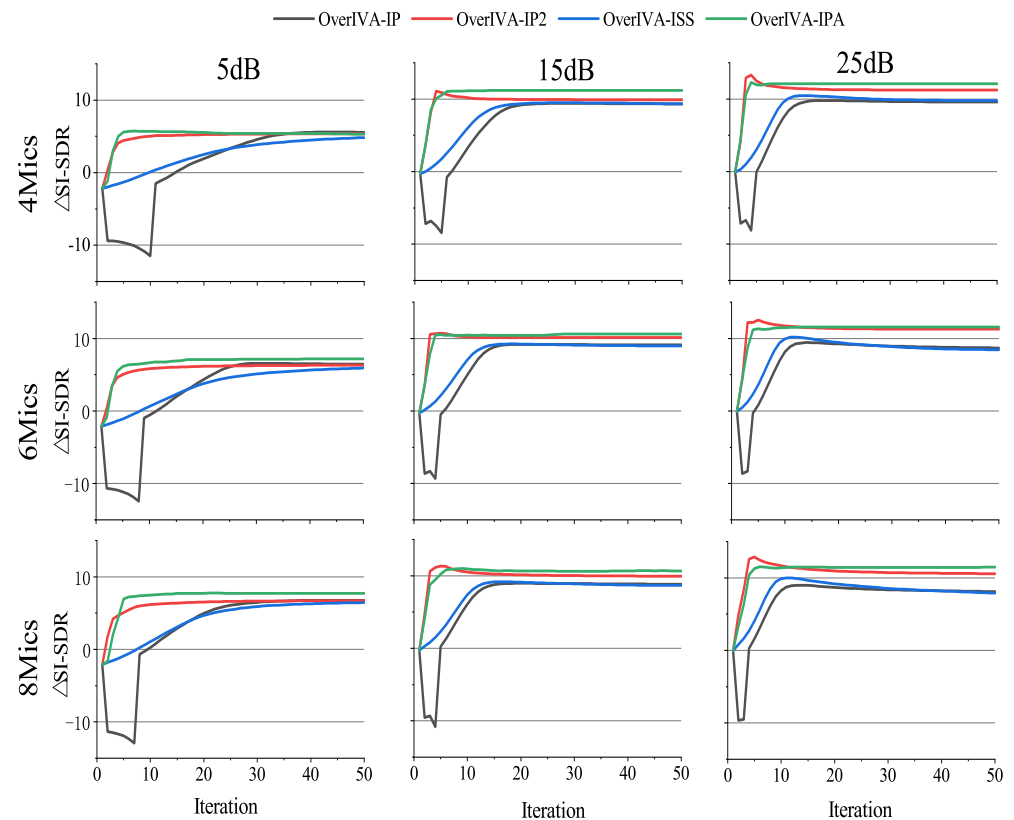


Figure 2. Variation of Δ SI-SDR at 5 dB, 15 dB and 25 dB for OverIVA-IP, OverIVA-IP2, OverIVA-ISS, and OverIVA-IPA for a mixture of two sources. The number of microphones is 4, 6 and 8 from top to bottom.

Table 2. The SI-SDR values of the four algorithms after stabilization.

Method	4 Mics			6 Mics			8 Mics		
	SI-SDR [dB]			SI-SDR [dB]			SI-SDR [dB]		
	5 dB	15 dB	25 dB	5 dB	15 dB	25 dB	5 dB	15 dB	25 dB
OverIVA-IP	5.56	9.32	9.57	6.44	9.11	8.96	6.77	8.83	8.09
OverIVA-IP2	5.35	9.90	11.22	6.33	10.14	11.27	6.71	9.93	10.56
OverIVA-ISS	4.85	9.34	9.77	5.93	8.95	8.45	6.45	8.66	7.86
OverIVA-IPA	5.49	11.18	12.12	7.21	10.61	11.57	7.7	10.86	11.47

At the same time, we can obtain the separated spectrum diagram of the signals obtained by each method in the six mics 25 dB environment as follows (Figure 4):

It can be seen from the separated spectrogram that the proposed OverIVA-IPA algorithm can separate the source signal from the mixed signal better than other algorithms, and the separated signal spectrogram waveform is better than other algorithms in detail.

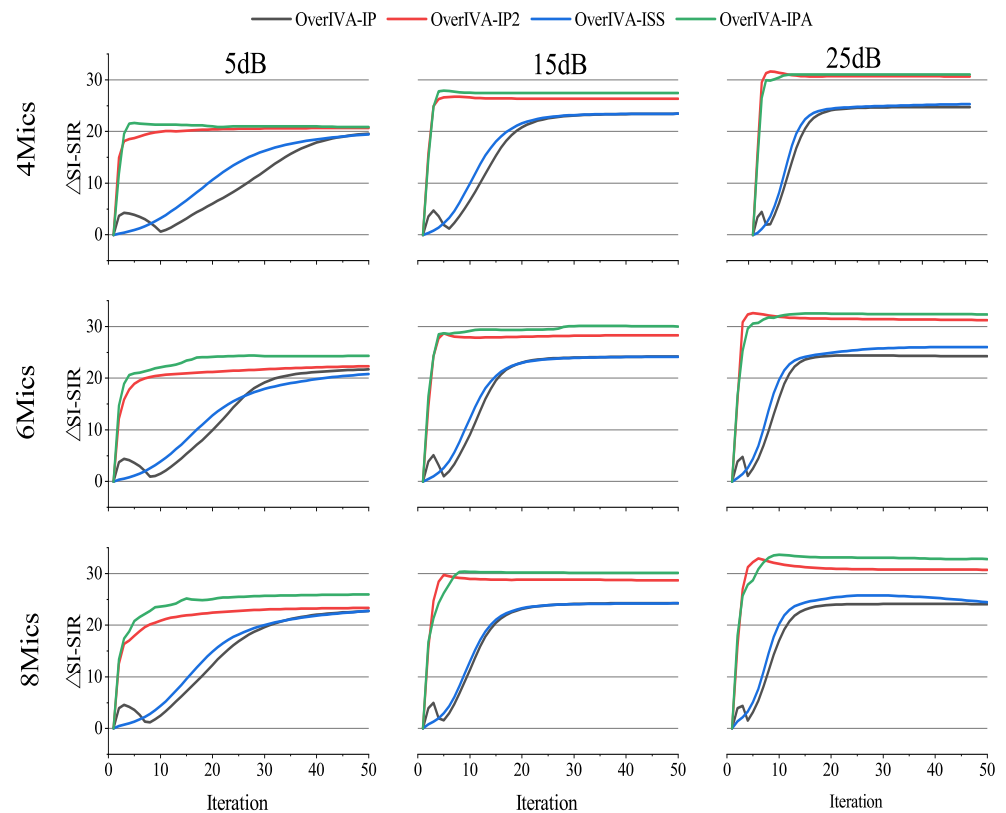


Figure 3. Variation of $\Delta SI-SIR$ at 5 dB, 15 dB and 25 dB for OverIVA-IP, OverIVA-IP2, OverIVA-ISS, and OverIVA-IPA for a mixture of two sources. The number of microphones is 4, 6 and 8 from top to bottom.

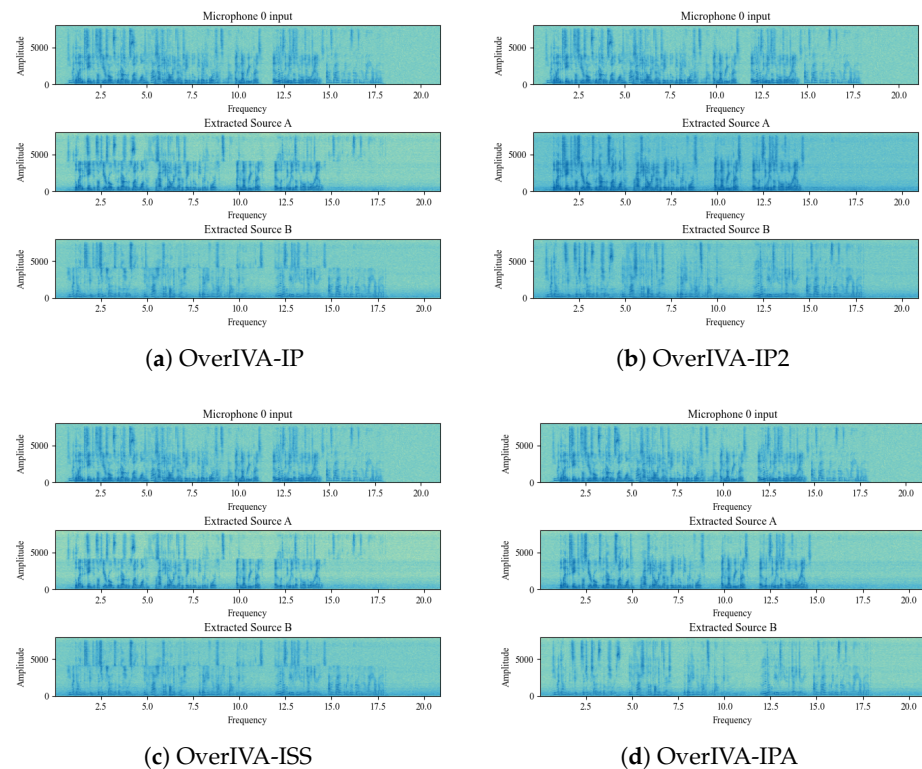


Figure 4. Signal separation spectrum diagram of 4 algorithms under 6 mics 25 dB.

5. Summary and Prospect

We propose an overdetermined independent vector analysis (OverIVA) algorithm optimized using the iterative projection with adjustment (IPA) algorithm. The algorithm applies efficient updates from auxiliary-function-based IVA (AuxIVA). And the complexity of the algorithm is consistent with that of the iterative projection (IP) algorithm. In numerical experiments, we thoroughly investigated the performance of OverIVA using the different update rules for the separation of realistically simulated speech mixtures. Through the analysis of experimental results, the proposed OverIVA-IPA algorithm is superior to other algorithms in all environments. Future work will focus on applying the algorithm to real systems and evaluating its real-time execution performance.

Author Contributions: R.G.: conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft. Z.L.: writing—review, editing, supervision, project administration, resources. L.W.: writing—review, editing, supervision. L.F.: writing—review, editing, supervision, resources, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61801319, in part by Sichuan Science and Technology Program under Grant 2020JDJQ0061, 2021YFG0099, in part by Innovation Fund of Chinese Universities under Grant 2020HYA04001, in part by Innovation Fund of Engineering Research Center of the Ministry of Education of China, Digital Learning Technology Integration and Application (No. 1221009), in part by the 2021 Graduate Innovation Fund of Sichuan University of Science and Engineering under Grant Y2022112.

Data Availability Statement: We are using the concatenated utterances from the the CMU Arctic corpus. <http://doi.org/10.5281/zenodo.3066489>.

Conflicts of Interest: No potential conflict of interest was reported by the authors.

References

1. Cao, X.-R.; Liu, R.-W. General approach to blind source separation. *IEEE Trans. Signal Process.* **1996**, *44*, 562–571. [\[CrossRef\]](#)
2. Cano, E.; FitzGerald, D.; Liutkus, A.; Plumbley, M.D.; Stöter, F.-R. Musical source separation: An introduction. *IEEE Signal Process. Mag.* **2019**, *36*, 31–40. [\[CrossRef\]](#)
3. Guo, R.; Luo, Z.; Li, M. A Survey of Optimization Methods for Independent Vector Analysis in Audio Source Separation. *Sensors* **2023**, *23*, 493. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Cong, F. Blind source separation. In *EEG Signal Processing and Feature Extraction*; Hu, L., Zhang, Z., Eds.; Springer: Singapore, 2019; Chapter 7, pp. 117–140.
5. Luo, Z.; Li, M.; Li, C. Independent vector analysis based blind interference reduction and signal recovery for MIMO IoT green communications. *China Commun.* **2022**, *19*, 79–88. [\[CrossRef\]](#)
6. Luo, Z.; Ruiming, G.; Li, C. Independent Vector Analysis for Blind Deconvolving of Digital Modulated Communication Signals. *Electronics* **2022**, *11*, 1460. [\[CrossRef\]](#)
7. Hyvärinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **2000**, *13*, 411–430. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Stone, J.V. Independent component analysis: An introduction. *Trends Cogn. Sci.* **2002**, *6*, 59–64. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Nesta, F.; Svaizer, P.; Omologo, M. Convolutional BSS of short mixtures by ICA recursively regularized across frequencies. *IEEE Trans. Audio, Speech, Lang. Process.* **2010**, *19*, 624–639. [\[CrossRef\]](#)
10. Kim, T.; Eltoft, T.; Lee, T.-W. Independent vector analysis: An extension of ICA to multivariate components. In *International Conference on Independent Component Analysis and Signal Separation*; Springer: Berlin/Heidelberg, Germany, 2006.
11. Luo, Z. Independent vector analysis: Model, applications, challenges. *Pattern Recognit.* **2023**, *138*, 109376. [\[CrossRef\]](#)
12. Lee, I.; Kim, T.; Lee, T.W. Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. *Signal Process.* **2007**, *87*, 1859–1871. [\[CrossRef\]](#)
13. Ono, N. Stable and fast update rules for independent vector analysis based on auxiliary function technique. In Proceedings of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 16–19 October 2011; pp. 189–192.
14. Nakashima, T.; Scheibler, R.; Wakabayashi, Y.; Ono, N. Faster independent low-rank matrix analysis with pairwise updates of demixing vectors. In Proceedings of the 2020 28th European Signal Processing Conference (EUSIPCO), Amsterdam, The Netherlands, 18–21 January 2021.
15. Scheibler, R.; Ono, N. Fast and stable blind source separation with rank-1 updates. In Proceedings of the ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–9 May 2020.

16. Ikeshita, R.; Nakatani, T. ISS₂: An Extension of Iterative Source Steering Algorithm for Majorization-Minimization-Based Independent Vector Analysis. *arXiv* **2022**, arXiv:2202.00875.
17. Scheibler, R. Independent vector analysis via log-quadratically penalized quadratic minimization. *IEEE Trans. Signal Process.* **2021**, *69*, 2509–2524. [[CrossRef](#)]
18. Nishikawa, T.; Abe, H.; Saruwatari, H.; Shikano, K. Overdetermined blind separation for convolutive mixtures of speech based on multistage ICA using subarray processing. In Proceedings of the 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, QC, Canada, 17–21 May 2004; p. I-225. [[CrossRef](#)]
19. Osterwise, C.; Grant, S.L. On over-determined frequency domain bss. *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **2014**, *22*, 956–966.
20. Joho, M.; Mathis, H.; Lambert, R.H. Overdetermined blind source separation: Using more sensors than source signals in a noisy mixture. In Proceedings of the Independent Component Analysis and Blind Signal Separation ICA 2000, Helsinki, Finland, 19–22 June 2000; pp. 81–86.
21. Lee, I.; Kim, T.; Lee, T.-W. Independent vector analysis for convolutive blind speech separation. In *Blind Speech Separation*; Springer: Dordrecht, The Netherlands, 2007; pp. 169–192.
22. Koldovsk'y, Z.; Tichavsk'y, P. Gradient algorithms for complex non-gaussian independent component/vector extraction, question of convergence. *IEEE Trans. Signal Process.* **2018**, *67*, 1050–1064. [[CrossRef](#)]
23. Scheibler, R.; Ono, N. Fast independent vector extraction by iterative SINR maximization. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 601–605.
24. Scheibler, R.; Ono, N. Independent vector analysis with more microphones than sources. In Proceedings of the 2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 20–23 October 2019.
25. Ikeshita, R.; Nakatani, T.; Araki, S. Overdetermined independent vector analysis. In Proceedings of the ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 591–595.
26. Scheibler, R.; Ono, N. MM algorithms for joint independent subspace analysis with application to blind single and multi-source extraction. *arXiv* **2020**, arXiv:2004.03926.
27. Du, Y.; Scheibler, R.; Togami, M.; Yoshii, K.; Kawahara, T. Computationally-Efficient Overdetermined Blind Source Separation Based on Iterative Source Steering. *IEEE Signal Process. Lett.* **2022**, *29*, 927–931. [[CrossRef](#)]
28. Togami, M.; Scheibler, R. Over-determined speech source separation and dereverberation. In Proceedings of the 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Auckland, New Zealand, 7–10 December 2020; pp. 705–710.
29. Scheibler, R.; Bezzam, E.; Dokmani'c, I. Pyroomacoustics: A Python package for audio room simulation and array processing algorithms. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 351–355.
30. Kominek, J.; Black, A.W. *CMU Arctic Databases for Speech Synthesis*; Technical Report; Carnegie Mellon University: Pittsburgh, PA, USA, 2003.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.