

A two-stage frequency-domain blind source separation method for underdetermined instantaneous mixtures

Peng Tianliang Chen Yang

(School of Information Science and Engineering, Southeast University, Nanjing 210096, China)

Abstract: In order to decrease the probability of missing some data points or noises being added in the inverse truncated mixing matrix (ITMM) algorithm, a two-stage frequency-domain method is proposed for blind source separation of underdetermined instantaneous mixtures. The separation process is decomposed into two steps of ITMM and matrix completion in the view that there are many soft-sparse (not very sparse) sources. First, the mixing matrix is estimated and the sources are recovered by the traditional ITMM algorithm in the frequency domain. Then, in order to retrieve the missing data and remove noises, the matrix completion technique is applied to each preliminary estimated source by the traditional ITMM algorithm in the frequency domain. Simulations show that, compared with the traditional ITMM algorithms, the proposed two-stage algorithm has better separation performances. In addition, the time consumption problem is considered. The proposed algorithm outperforms the traditional ITMM algorithm at a cost of no more than one-fourth extra time consumption.

Key words: inverse truncated mixing matrix; underdetermined blind source separation (UBSS); frequency domain; matrix completion

doi: 10.3969/j.issn.1003-7985.2016.02.001

The underdetermined blind source separation (UBSS) technique is the case that the number of sensors is fewer than that of the sources. The sparseness-based approach has been widely employed to solve the UBSS problem. Sparseness-based methods can be mainly classified into two classes. The first class is a two-stage approach: mixing matrix estimation and source recovery. Many researchers have proposed different source recovery approaches, such as the maximum a posteriori (MAP) assumption^[1-2], the MMSE method^[3], and the inverse truncated mixing matrix (ITMM) algorithm^[4]. The second class extracts each source with time-frequency binary masks^[5-6]. The ITMM is a powerful method for underde-

termined blind source separation^[4]. The separation performance is usually affected by the assumptions and parameter sizes of the ITMM algorithm. In order to improve the robustness and quality of the ITMM algorithm, we propose a two-stage method with joint matrix completion and ITMM algorithm in this paper. We apply the ITMM algorithm to estimate the time-frequency (TF) domain source in the first stage. Then, the matrix completion algorithm is implemented to every estimated source's TF matrix. In our simulation, the mixtures of audio and speech signals are considered and the results show that this approach yields a good performance.

1 ITMM Method for UBSS

The block diagram of the proposed underdetermined UBSS algorithm is shown in Fig. 1.

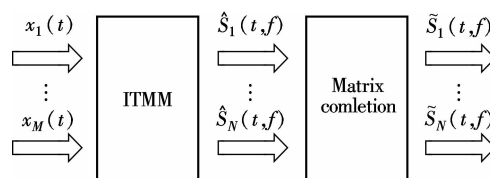


Fig. 1 Block diagram of the proposed underdetermined UBSS algorithm

For the instantaneous UBSS case, let $\mathbf{x}_t = \{x_t(1), x_t(2), \dots, x_t(M)\}^T$ be an M -dimensional column vector corresponding to the output of M sensors at a given discrete time instant t , and let \mathbf{X} be an $M \times T$ matrix corresponding to the sensor data at all times $t = 1, 2, \dots, T$. Let $\mathbf{S} = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_T\}$ be the $N \times T$ matrix of underlying source signals, where $\mathbf{s}_t = \{s_t(1), s_t(2), \dots, s_t(N)\}^T$ and let $\mathbf{A} = \{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N\}$ be the mixing matrix with size $M \times N$, where $\mathbf{a}_i = \{a_{i1}, a_{i2}, \dots, a_{iM}\}^T$ and $M < N$. Then the instantaneous UBSS case can be written as

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (1)$$

or equivalent in vector version as

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t \quad (2)$$

Assumption 1 The column vectors of \mathbf{A} are assumed to be pairwise linearly independent, that is, for any $k, l \in \Lambda$, where $\Lambda = \{1, 2, \dots, N\}$ and $k \neq l$, we can derive the relationship that \mathbf{a}_k and \mathbf{a}_l are linearly independent.

Assumption 1 is usually made in the context of blind source separation (BSS)^[7]. It is also known that BSS

Received 2015-11-23.

Biographies: Peng Tianliang(1984—), male, graduate; Chen Yang(corresponding author), male, doctor, professor, chen@seu.edu.cn.

Foundation item: The National Natural Science Foundation of China (No. 60872074).

Citation: Peng Tianliang, Chen Yang. A two-stage frequency-domain blind source separation method for underdetermined instantaneous mixtures[J]. Journal of Southeast University (English Edition), 2016, 32(2): 135 – 140. doi: 10.3969/j.issn.1003-7985.2016.02.001.

goes only possibly up to an unknown scaling and an unknown permutation of the original sources^[8].

In order to transform signals into the frequency domain, we apply a short-time Fourier transform (STFT) to the sensor observations x_i . Then, the instantaneous mixture at each frequency domain is given by

$$\mathbf{x}(f, t) = \mathbf{A}s(f, t) \quad (3)$$

where $\mathbf{x}(f, t) = \{x_1(f, t), x_2(f, t), \dots, x_M(f, t)\}^T$, $\mathbf{s}(f, t) = \{s_1(f, t), s_2(f, t), \dots, s_N(f, t)\}^T$ are the TF representations of the sensors $\mathbf{x}(t)$ and the sources $\mathbf{s}(t)$, respectively.

Assumption 2 The number of coexisting sources at a single TF point is smaller than that of sensors. In other words, if we assume that the number of sensors is M , then the largest number of sources that exist at any TF point is $M - 1$.

Actually Assumption 2 is a relaxed condition to the W-disjoint orthogonality of the time-frequency masking technique^[6].

The ITMM-UBSS method can be divided into two sections. The first section is the mixing matrix estimation, which is realized by the mask-TFD method^[3] in this paper; the second section is the source recovery in the frequency domain by the traditional ITMM algorithm. As the ITMM algorithm shown in Ref. [4], we can obtain the general version of $s_i(t, f)$ through the union of TF points as follows:

$$s_i(t, f) = \bigcup_{k_1 k_2 \dots k_{M-2} (t, f) \in \Omega_{k_1 k_2 \dots k_{M-2}}} \{s_i(t, f) + s_i(t, f)\} \quad (4)$$

where $s_i(t, f) = x_1^{ik_1 k_2 \dots k_{M-2} l_i}(t, f)$; $s_i(t, f) = 0$; and $x_1^{ik_1 k_2 \dots k_{M-2} l_i}(t, f)$ is derived as

$$\begin{bmatrix} x_1^{ik_1 k_2 \dots k_{M-2} l_i}(t, f) \\ \vdots \\ x_M^{ik_1 k_2 \dots k_{M-2} l_i}(t, f) \end{bmatrix} = \mathbf{A}_{\text{inv}}^{ik_1 k_2 \dots k_{M-2} l_i} \begin{bmatrix} x_1(t, f) \\ \vdots \\ x_M(t, f) \end{bmatrix} = \mathbf{A}_{\text{inv}}^{ik_1 k_2 \dots k_{M-2} l_i} \mathbf{A} \begin{bmatrix} s_1(t, f) \\ \vdots \\ s_N(t, f) \end{bmatrix} \quad (5)$$

where $\mathbf{A}_{\text{inv}}^{ik_1 k_2 \dots k_{M-2} l_i}$ is the inverse matrix of $\mathbf{A}^{ik_1 k_2 \dots k_{M-2} l_i}$, and $\mathbf{A}^{ik_1 k_2 \dots k_{M-2} l_i} = [\mathbf{a}_i, \mathbf{a}_{k_1}, \mathbf{a}_{k_2}, \dots, \mathbf{a}_{k_{M-1}}, \mathbf{a}_{l_i}]$ with $l_j \in \Gamma \setminus i$ and $l_j \notin \Lambda(j = 1, 2, \dots, N - M + 1)$. $x_1^{ik_1 k_2 \dots k_{M-2} l_i}(t, f)$ must satisfy

$$\left. \begin{aligned} & |\text{Im}(x_1^{ik_1 k_2 \dots k_{M-2} l_i}(x_1^{ik_1 k_2 \dots k_{M-2} l_i})^*)| > \delta_{ik_1 k_2 \dots k_{M-2}} \\ & \vdots \\ & |\text{Im}(x_1^{ik_1 k_2 \dots k_{M-2} l_i}(x_1^{ik_1 k_2 \dots k_{M-2} l_i})^*)| > \delta_{ik_1 k_2 \dots k_{M-2}} \end{aligned} \right\} \quad (6)$$

where $\text{Im}(x)$ represents the imaginary part of x ; x^* is the conjugate of x ; $\delta_{ik_1 k_2 \dots k_{M-2}}$ is a very small positive value. In Matlab simulations, we write Eq. (6) as

$$\left. \begin{aligned} & |\text{angle}(x_1^{ik_1 k_2 \dots k_{M-2} l_i}(x_1^{ik_1 k_2 \dots k_{M-2} l_i})^*)| > \delta_{ik_1 k_2 \dots k_{M-2}} \\ & \vdots \\ & |\text{angle}(x_1^{ik_1 k_2 \dots k_{M-2} l_i}(x_1^{ik_1 k_2 \dots k_{M-2} l_i})^*)| > \delta_{ik_1 k_2 \dots k_{M-2}} \end{aligned} \right\} \quad (7)$$

where $\text{angle}(x)$ represents the angle of x .

2 Matrix Completion

Many signals become sparse after they are transformed into the frequency domain, such as speech signals. Suppose that we have a rank- r matrix \mathbf{A} of size $M \times N$, where $r \leq \min(M, N)$. In many engineering problems, the entries of the matrix are often corrupted by errors or noise, and some of the entries can even be missed, or only a set of measurements of the matrix is accessible rather than its entries directly. In general, we model the observed matrix \mathbf{B} to be a set of linear measurements on matrix \mathbf{A} , subject to noise and gross corruptions, i. e. ,

$$\mathbf{B} = \mathbf{L}(\mathbf{A}) + \mathbf{n} \quad (8)$$

where \mathbf{L} is a linear operator, and \mathbf{n} represents the matrix of corruptions. We seek to recover the true matrix \mathbf{A} from \mathbf{B} . We now consider the case where \mathbf{L} is the identity operator and \mathbf{n} is a sparse matrix (i. e. , most of its entries are zero) but whose non-zero entries can be practically unbounded. Since the rank r of \mathbf{A} is unknown, the problem is to find the matrix of the lowest rank that can generate \mathbf{B} when added to an unknown sparse matrix \mathbf{n} . Mathematically, for an appropriate choice of parameter $\eta > 0$, we have the following combinatorial optimization problem to solve:

$$\min_{\mathbf{A}, \mathbf{n}} \text{rank}(\mathbf{A}) + \eta \|\mathbf{n}\|_0 \quad \text{s. t.} \quad \mathbf{B} = \mathbf{A} + \mathbf{n} \quad (9)$$

where η is a positive constant, and $\|\cdot\|_0$ is the counting norm (i. e. , the number of non-zero entries in the matrix). Since this problem cannot be efficiently solved, we consider its convex relaxation instead:

$$\min_{\mathbf{A}, \mathbf{n}} \|\mathbf{A}\|_* + \eta \|\mathbf{n}\|_1 \quad \text{s. t.} \quad \mathbf{B} = \mathbf{A} + \mathbf{n} \quad (10)$$

where $\|\mathbf{A}\|_* = \sum_{k=1}^n \sigma_k(\mathbf{A})$; $\sigma_k(\mathbf{A})$ denotes the k -th largest singular value of \mathbf{A} ; and $\|\cdot\|_1$ represents the matrix 1-norm (i. e. , the sum of absolute values of all entries of the matrix). We call the above convex program the robust principal component analysis (RPCA)^[9-10].

3 Joint Matrix Completion with ITMM

Suppose that we have recovered sources in the frequency domain by the ITMM method. We assume that the estimated source matrix is $\tilde{\mathbf{S}}_i$ with the size of $F \times T$, and F and T are the number of frequency bins and time frames, respectively. The elements of $\tilde{\mathbf{S}}_i$ are $\tilde{s}_i(f, t)$ ($1 \leq f \leq F$, $1 \leq t \leq T$). We assume that $\tilde{\mathbf{S}}_i$ and the true sources \mathbf{S}_i have the following relationship:

$$\tilde{\mathbf{S}}_i = \mathbf{L}(\mathbf{S}_i) + \mathbf{n} \quad (11)$$

Then the problem that we want to obtain \mathbf{S}_i via Eq. (11) is equivalent to the matrix completion problem as in Eq. (8). So, we can use the matrix completion technique for every frequency domain matrix of estimated sources to retrieve the missing data and remove the noise data. We

use the method in Ref. [10] to simulate the matrix completion in this paper.

In summary, the joint matrix completion with the ITMM algorithm is as follows:

1) The traditional ITMM-UBSS algorithm is used in the frequency domain. First, estimate the mixing matrix with the mask-STD method^[3]. Secondly, recover sources with the ITMM algorithm.

2) The matrix completion method is used to estimate each source based on the traditional ITMM-UBSS algorithm in the frequency domain. Then, the time domain sources can be obtained by the inverse STFT.

4 Simulation

To demonstrate the two-stage method, we divide the simulation into two sections: One is the SiSEC 2008 data^[11] simulation, and the other is the simulation with random $M \times 6$ mixing matrix \mathbf{A} for $M = 2, 3, \dots, 5$.

4.1 SiSEC 2008 data

A linear instantaneous stereo mixture (with positive mixing coefficients) of two drum sources and one bass line, and four female speeches mixed with three sensors are used in this paper.

In this paper, the length of STFT is 1 024; the window overlap is 0.5; all the signals are sampled at 16 kHz; and the sample length is 160 000. Then, the size of $\tilde{\mathbf{S}}_i$ is 512×313 . The algorithms are coded with Matlab and run on an Intel Core(TM) 2 Duo (2.20 GHz) processor.

Fig. 2 shows the spectrum of two sensors. The spectrum of three sources, three separated sources with the traditional ITMM, and three separated sources with the proposed method are shown in Fig. 3, Fig. 4 and Fig. 5, respectively. The δ_i of the traditional ITMM and the proposed method are all set to be $\delta_1 = \delta_2 = \delta_3 = 0.3$.

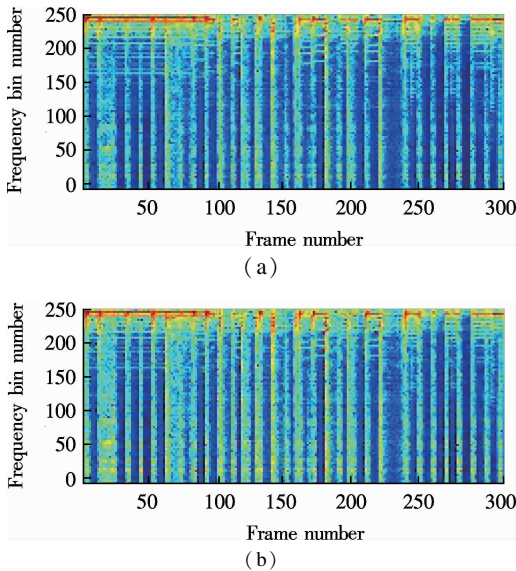


Fig. 2 Spectrum of observations. (a) The first observation; (b) The second observation

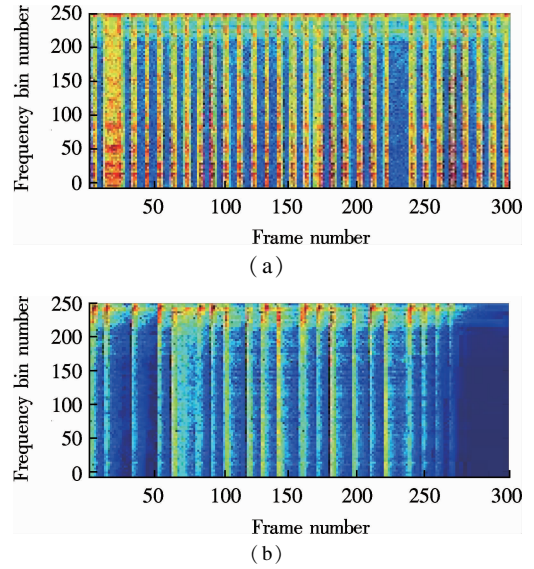


Fig. 3 Spectrum of three sources. (a) The baseline; (b) The first drum; (c) The second drum

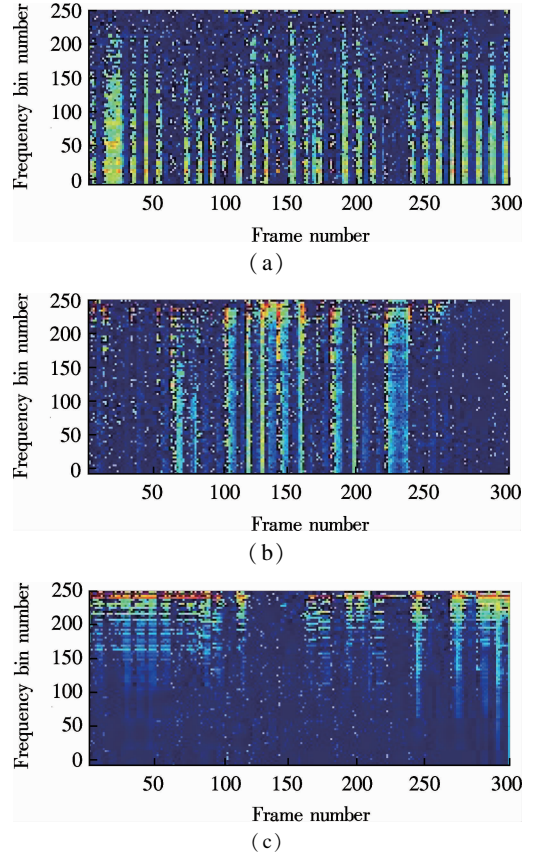


Fig. 4 Spectrum of three separated sources with traditional ITMM. (a) The baseline; (b) The first drum; (c) The second drum

As shown in Fig. 4 and Fig. 5, since the matrix completion technique in the proposed method can retrieve the missing data and reject noise, compared with the traditional ITMM algorithm, the spectra of the proposed method can be seen more clearly. The traditional ITMM algorithm must satisfy Assumption 1 and Assumption 2, which may be the main reason for corruption by errors or noise in the time-frequency domain of estimated sources.

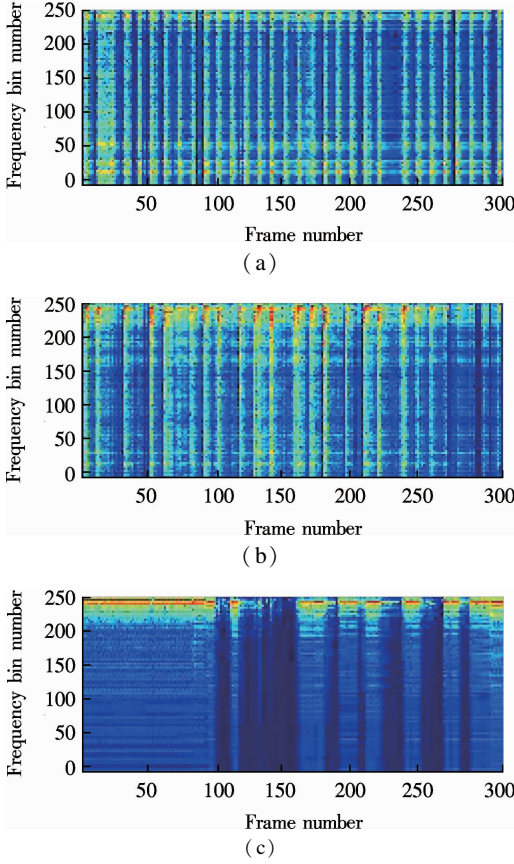


Fig. 5 Spectrum of three separated sources with the proposed method. (a) The baseline; (b) The first drum; (c) The second drum

To measure the performance, we decompose an estimated signal \hat{s} as $\hat{s} = s_{\text{target}} + e_{\text{interf}} + e_{\text{artif}} + e_{\text{noise}}$ [12]. As the measurement of performance, we use the source-to-distortion ratio

$$\mathcal{E}_{\text{SDR}} = 10 \lg \frac{\sum s_{\text{target}}^2}{\sum (e_{\text{interf}} + e_{\text{artif}})^2} \quad (20)$$

the source-to-interference ratio

$$\mathcal{E}_{\text{SIR}} = 10 \lg \frac{\sum s_{\text{target}}^2}{\sum e_{\text{interf}}^2} \quad (21)$$

and the source-to-artifact ratio

$$\mathcal{E}_{\text{SAR}} = 10 \lg \frac{\sum (s_{\text{target}} + e_{\text{interf}})^2}{\sum s_{\text{artif}}^2} \quad (22)$$

The higher values of \mathcal{E}_{SDR} , \mathcal{E}_{SIR} and \mathcal{E}_{SAR} indicate better performance. We must note that the ITMM is a source recovery method, and the mixing matrix estimation problem in UBSS is not considered. We use the method proposed in Ref. [3] to estimate the mixing matrix in this paper.

Tab. 1 is the performance evaluation of the ITMM and the proposed algorithms of mixtures with three audio sources, namely S1, S2 and S3, and two sensors under $\delta_1 = \delta_2 = \delta_3 = 0.3$. As shown in Tab. 1, the proposed

method can obtain a higher performance than the traditional ITMM method except for \mathcal{E}_{SIR} of the third source. Tab. 2 is the performance evaluation of the ITMM and the proposed methods of mixtures with four speeches and three sensors under $\delta_1 = \delta_2 = \delta_3 = \delta_4 = 0.3$. As shown in Tab. 2, the proposed method can obtain a higher performance than the traditional ITMM method except for \mathcal{E}_{SIR} of the fourth source. However, the average performance evaluations of the proposed method are higher than those of the traditional ITMM method as shown in Tab. 1 and Tab. 2. The reason is that the matrix completion of the proposed joint method can complete and extend the frequency bin of estimated sources.

Tab. 1 Performance evaluation of ITMM and proposed methods of mixtures with three audio sources and two sensors dB

Items	Methods	S1	S2	S3
\mathcal{E}_{SDR}	ITMM	7.37	0.79	27.26
	Proposed	8.01	0.92	28.02
\mathcal{E}_{SIR}	ITMM	14.58	13.82	30.14
	Proposed	14.84	14.23	30.00
\mathcal{E}_{SAR}	ITMM	8.36	1.33	30.41
	Proposed	10.00	2.11	32.65

Tab. 2 Quality evaluation of ITMM and proposed methods of mixtures with four speeches and three sensors dB

Items	Methods	S1	S2	S3	S4
\mathcal{E}_{SDR}	ITMM	9.74	7.11	9.32	11.32
	Proposed	11.01	10.16	10.35	12.59
\mathcal{E}_{SIR}	ITMM	16.21	13.52	16.00	18.68
	Proposed	17.08	14.23	16.60	17.57
\mathcal{E}_{SAR}	ITMM	10.95	8.43	10.49	12.26
	Proposed	12.00	9.11	11.86	13.15

Fig. 6 shows the average performances of \mathcal{E}_{SDR} , \mathcal{E}_{SIR} , \mathcal{E}_{SAR} for separating the three speech sources from two sensors under the condition of different values of δ_i . As shown in Fig. 6, \mathcal{E}_{SDR} , \mathcal{E}_{SIR} , and \mathcal{E}_{SAR} can reach their maximum values during different intervals, and obtain high values when $0.2 \leq \delta_i \leq 0.4$, $0.01 \leq \delta_i \leq 0.2$, and $0.8 \leq \delta_i \leq 1.0$, respectively. Our simulation suggests that the speeches and audio waves maintain high quality in the auditory perception when $0.1 \leq \delta_i \leq 0.4$.

4.2 Simulations with random $M \times 6$ mixing matrix A for $M = 2, 3, 4, \dots, 5$

The random simulations are completed in this section. We choose the random $M \times 6$ mixing matrix A for $M = 2, 3, 4, \dots, 5$. Three speech sources and three music sources are chosen from SiSEC 2008 data [11]. We carry out the random simulations 50 times for each mixing model. Fig. 7 shows the simulations of the six sources in the time domain.

Tab. 3 is the average quality evaluation of the ITMM and the proposed methods in random cases. As shown in Tab. 3, \mathcal{E}_{SDR} , \mathcal{E}_{SIR} and \mathcal{E}_{SAR} of the proposed method are all higher than those of the traditional ITMM algorithm ex-

cept for ε_{SIR} of 2×6 and 4×6 when $\delta = 0.05$.

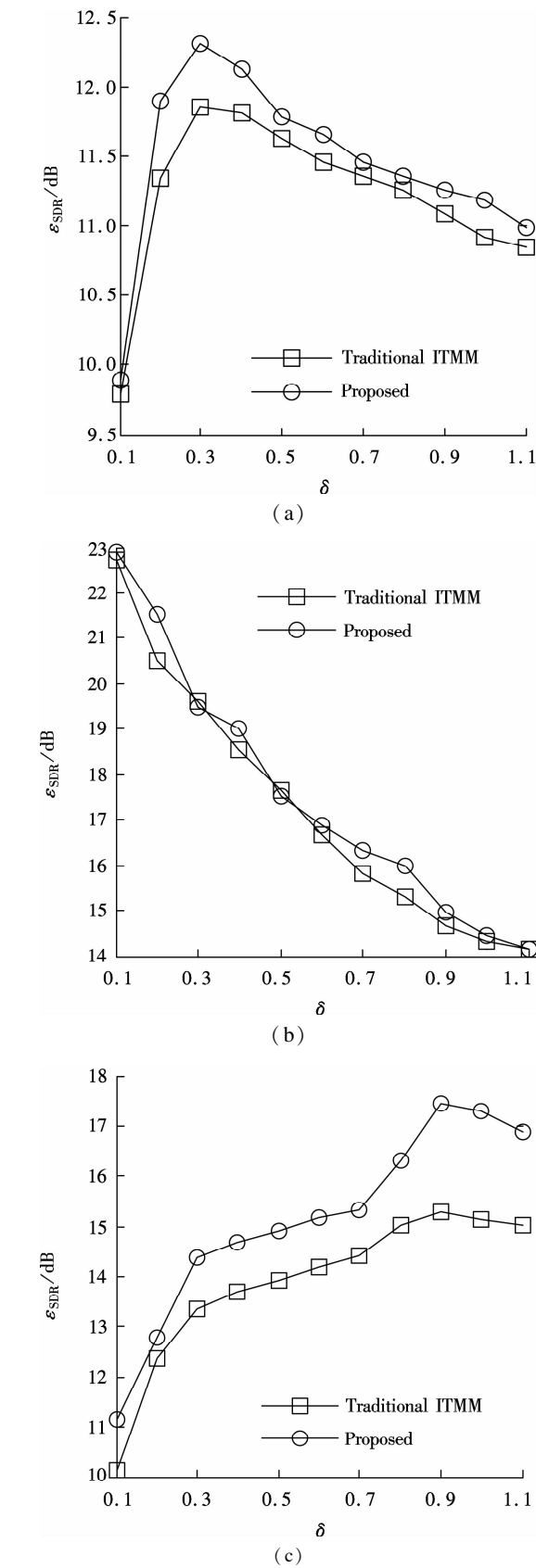


Fig. 6 Average performances of the traditional ITMM and the proposed methods for separating three audio sources from two sensors under different values of δ ($\delta_1 = \delta_2 = \delta_3 = \delta$). (a) ε_{SDR} ; (b) ε_{SIR} ; (c) ε_{SAR}

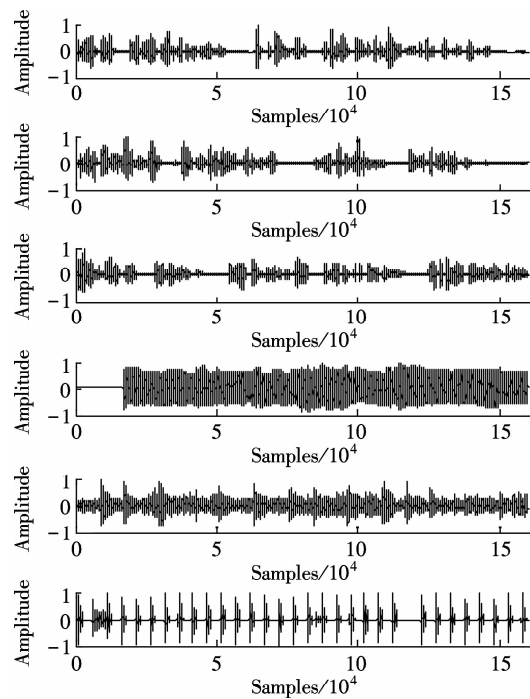


Fig. 7 The six testing sources in the time domain

Tab. 3 Average performance evaluation of ITMM and proposed methods in random cases

Items	2 × 6 matrix		3 × 6 matrix		4 × 6 matrix		5 × 6 matrix	
	ITMM		Proposed		ITMM		Proposed	
	ITMM	Proposed	ITMM	Proposed	ITMM	Proposed	ITMM	Proposed
$\delta = 0.05$								
ε_{SDR}	-4.69	-4.13	1.14	1.19	5.38	5.44	12.92	13.01
ε_{SIR}	9.42	9.38	11.88	12.01	13.98	13.88	21.53	21.66
ε_{SAR}	-3.18	-2.94	2.30	2.33	6.55	6.78	13.88	14.42
$\delta = 0.1$								
ε_{SDR}	-3.09	-2.77	1.92	2.21	4.5	4.76	11.25	11.34
ε_{SIR}	9.17	9.96	10.40	11.95	11.37	13.53	18.53	18.89
ε_{SAR}	-1.46	-1.13	3.53	4.46	6.19	7.01	12.66	12.79
$\delta = 0.2$								
ε_{SDR}	-1.67	-1.34	1.63	1.87	3.26	3.68	9.32	10.93
ε_{SIR}	8.80	10.91	8.74	9.93	9.14	10.31	15.28	16.01
ε_{SAR}	0.19	1.09	4.03	5.72	5.56	6.51	11.36	12.09
$\delta = 0.3$								
ε_{SDR}	-0.89	-0.61	1.21	1.34	1.58	1.78	8.12	9.03
ε_{SIR}	7.94	8.01	6.32	6.89	6.40	7.62	13.65	14.11
ε_{SAR}	1.48	1.54	4.45	4.77	4.96	6.01	10.66	11.24
$\delta = 0.4$								
ε_{SDR}	-0.60	-0.56	1.01	1.06	1.35	1.44	7.87	7.94
ε_{SIR}	7.23	7.45	5.59	5.85	5.78	6.00	12.44	12.67
ε_{SAR}	1.91	3.12	3.67	4.19	4.41	6.13	10.12	12.91

Tab. 4 shows the average computational time (in all the cases shown in Tab. 3) in random simulations. As shown in Tab. 4, the proposed method consumes about 4 s more than the traditional ITMM method in all the cases. The proposed method is a little time-consuming but its average performances are improved.

Tab. 4 Computational time for the proposed and traditional ITMM methods

Cases	ITMM	Proposed
2 × 6 matrix	18.73	22.23
3 × 6 matrix	20.24	23.69
4 × 6 matrix	19.51	23.07
5 × 6 matrix	18.19	21.48

5 Conclusion

In this paper, we propose a two-stage frequency domain algorithm for underdetermined instantaneous blind source separation. In the first stage, we use the ITMM algorithm^[4] to extract the sources in the frequency domain. In the second stage, we use the matrix completion technique to retrieve the missing data and remove the noise data of the estimated source matrix (corrupted by errors or noise) in the time-frequency domain. Simulation results show that the proposed algorithm has a higher performance compared with the conventional ITMM algorithm.

References

[1] Cobos M, Lopez J J. Maximum a posteriori binary mask estimation for underdetermined source separation using smoothed posteriors [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, **20**(7): 2059 – 2064. DOI:10.1109/tasl.2012.2195654.

[2] Leglaive S, Badeau R, Richard G. Multichannel audio source separation with probabilistic reverberation modeling[C]//*IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. New York, USA, 2015; 1 – 5.

[3] Peng Tianliang, Chen Yang, Liu Zengli. A time-frequency domain blind source separation method for underdetermined instantaneous mixtures [J]. *Circuits Systems & Signal Processing*, 2015, **34**(12): 3883 – 3895.

[4] Peng Tianliang, Chen Yang. Inverse truncated mixing matrix (ITMM) algorithm application to underdetermined convolutive blind speech sources separation [C]//*IEEE International Conference on Signal Processing, Commu-*

nications and Computing. Ningbo, China, 2015; 771 – 776.

[5] Yilmaz O, Rickard S. Blind separation of speech mixtures via time-frequency masking [J]. *IEEE Transactions on Signal Processing*, 2004, **52**(7):1830 – 1847.

[6] Sawada H, Araki S, Makino S. Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2011, **19**(3): 516 – 527. DOI:10.1109/tasl.2010.2051355.

[7] Linh-Trung N, Belouchrani A, Abed-Meraim K, et al. Separating more sources than sensors using time-frequency distributions [J]. *EURASIP Journal on Applied Signal Processing*, 2005, **2005**: 2828 – 2847. DOI:10.1155/asp.2005.2828.

[8] Belouchrani A, Abed-Meraim K, Cardoso J F, et al. A blind source separation technique using second-order statistics [J]. *IEEE Transactions on Signal Processing*, 1997, **45**(2): 434 – 444. DOI:10.1109/78.554307.

[9] Cai J F, Candès E J, Shen Z. A singular value thresholding algorithm for matrix completion [J]. *SIAM Journal on Optimization*, 2010, **20**(4): 1956 – 1982. DOI:10.1137/080738970.

[10] Candès E J, Recht B. Exact matrix completion via convex optimization [J]. *Foundations of Computational Mathematics*, 2009, **9**(6): 717 – 772.

[11] Vincent E, Sawada H, Bofill P, et al. First stereo audio source separation evaluation campaign: Data, algorithms and results [M]//*Independent Component Analysis and Signal Separation*. Berlin: Springer, 2007; 552 – 559.

[12] Vincent E, Gribonval R, Févotte C. Performance measurement in blind audio source separation [J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2006, **14**(4): 1462 – 1469. DOI:10.1109/tsa.2005.858005.

一种基于频域的欠定瞬时混合 2 步盲分离法

彭天亮 陈 阳

(东南大学信息科学与工程学院,南京 210096)

摘要:为了减少传统的截断混合矩阵求逆 (ITMM) 算法在个别时频点会丢失数据或者产生噪声信号的概率,提出了一种基于频域的 2 步欠定瞬时盲分离算法. 由于现实中存在大量软稀疏 (稀疏度不是很大) 混合信号,将分离过程分解为 ITMM 和矩阵补偿 2 个步骤. 首先估计出混合矩阵和利用经典的 ITMM 算法对混合信号进行初步恢复,然后对初步估计的信号时频矩阵进行矩阵补偿处理,从而达到修补丢失数据和去除多余数据 (去噪) 的效果. 实验仿真验证了所提出的 2 步分离法相对于传统的 ITMM 算法能够得到更好的分离效果. 此外,对算法的时耗问题进行了研究,相对于传统的 ITMM 算法,所提算法的时耗增加不到四分之一,却能够得到更好的分离效果.

关键词:截断混合矩阵求逆;欠定盲源分离;频域;矩阵填充

中图分类号:TP391