FAST NONCLOSED-FORM HIGHER-ORDER ICA WITH FREQUENCY-SUBBAND SELECTION USING CLOSED-FORM SECOND-ORDER ICA

Kentaro Tachibana, Yu Takahashi, Yoshimitsu Mori, Hiroshi Saruwatari, Kiyohiro Shikano†, Akira Tanaka‡

†Nara Institute of Science and Technology
Nara 630-0192, Japan
Phone: +81-743-72-5287
Email: kentaro-t@is.naist.jp

Abstract
In this paper, we proposed a faster convergence algorithm of blind source separation (BSS). We have proposed a computational-cost efficient BSS cascading closed-form 2nd-order independent component analysis (SO-ICA) with nonclosed-form higher-order ICA (HO-ICA). The closed-form solution of SO-ICA has been recently presented by one of the authors. This finding motivates us to combine the closed-form SO-ICA with HO-ICA, where the preceding closed-form ICA produces a good initial value and the following HO-ICA updates the separation filters from the advantageous status. In this paper, we propose a faster convergence method of BSS by selecting optimized frequency-subbands for HO-ICA based on a proper cost function. The experimental result shows the proposed method's efficacy for the convergence speed under the realistic reverberant condition.

1. Introduction
Blind source separation (BSS) is an approach to estimate the original source signals using only information of the observed signal in each microphones. Basically BSS is classified into unsupervised filtering technique, and in that the source-separation procedure requires no training sequences and no a priori information on direct-of-arrivals of the sound sources. Much attention has been paid to BSS in many fields of signal processing such as speech enhancement because of these attractive features.

In recent studies of BSS based on independent component analysis (ICA), various methods have been presented for acoustic-sound separation [1, 2, 3, 4, 5]. This paper also addresses the BSS problem under highly reverberant conditions which often arise in many practical audio applications. Generally speaking, almost all the algorithms in ICA, e.g., 2nd-order ICA (SO-ICA) [3, 5] and higher-order ICA (HO-ICA) [1, 2, 4], are conducted through nonclosed-form, i.e., iterative optimization. However, iterative algorithms are often confronted with local minimum problem or slow convergence. Particularly, the latency due to the slow convergence of ICA-based BSS is the significant problem for real-time processing. Thus, we can not apply ICA-based BSS directly to real-time applications.

On the other hand, the closed-form solution of SO-ICA has been recently proved by one of the authors [6], and we have proposed an efficient BSS method based on the nonclosed-form HO-ICA cascaded with closed-form SO-ICA [7], i.e., the BSS method which performs the nonclosed-form HO-ICA after the closed-form SO-ICA. In this paper, we propose a more faster convergence method of BSS by selecting the optimized frequency-subbands for HO-ICA based on the appropriate cost function. To evaluate the efficacy of the proposed method, we carried out sound-separation experiments in a real reverberant room.

2. Mixing process
In this study, the number of microphones is K and the number of multiple sound sources is L, where we deal with the case of K = L. In the frequency domain, the observed signals, i.e., multiple source mixed signals are given by

\[ X(f, t) = A(f)S(f, t), \]

where \( X(f, t) = [X_1(f, t), \ldots, X_K(f, t)]^T \) is the observed signal vector, and \( S(f, t) = [S_1(f), \ldots, S_L(f)]^T \) is the source signal vector. Also, \( A(f) = [A_{kl}(f)]_{kl} \) is the mixing matrix, where \( [X]_{ij} \) denotes the matrix which includes the element \( X \) in the \( i \)-th row and the \( j \)-th column. The mixing matrix \( A(f) \) is assumed to be complex-valued because we introduce a model to deal with the arrival lags among microphones and room reverberations.

3. Conventional ICA based BSS
We perform signal separation using the complex-valued unmixing matrix, \( W(f) = [W_{kl}(f)]_{kl} \), so that the L time-series output \( Y(f, t) = [Y_1(f, t), \ldots, Y_L(f, t)]^T \) becomes mutually independent. This procedure can be given as

\[ Y(f, t) = W(f)X(f, t). \]
Figure 1. Block diagram of updating separation filter $W_{\text{HO}}(f)$ in conventional method.

Updating $W(f)$ has an inherent disadvantage which is the poor and slow convergence of nonlinear optimization, particularly when we are confronted with very complex convolutive mixtures and unfortunately set a bad initial value. Furthermore, typical ICA-based BSS algorithms require huge computational complexities. The disadvantages reduce the applicability to the general audio applications which often need real-time processing. To solve the problems, we have proposed the efficient BSS combining nonclosed-form HO-ICA and closed-form SO-ICA[7]. Figure 1 shows the block diagram of $W_{\text{HO}}(f)$ in this method. Hereinafter, we describe the detailed algorithm.

3.1. Closed-form SO-ICA

In the original reference [6], the principle of the closed-form SO-ICA was derived, especially from the mathematical point of view. This subsection briefly describes the overview of signal processing in the closed-form SO-ICA.

First, we obtain the correlation matrices with different time points as

$$R_t(f) = \langle X(f,t)X(f,t)^H \rangle_{\text{time}},$$

(3)

where $\langle \cdot \rangle_{\text{time}}$ denotes the time-averaging operator over specific time duration $t_i$, and $i = 1, 2, \ldots$ represents indices of time-averaging block.

Next, we apply the singular value decomposition (SVD) to a superposition of $R_t(f)$, which is represented as

$$\sum_i R_t(f) = U(f)\text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_L)U(f)^H,$$

(4)

where $\lambda_k$ are the singular values, $\text{diag}(\lambda_1, \ldots)$ denotes the diagonal matrix whose diagonal elements consist of the singular values. and $U(f)$ is the matrix consisting of the singular vectors. Then we obtain a full-rank decomposition for pseudo-inverse of $\sum_i R_t(f)$ as follows

$$\left[ \sum_i R_t(f) \right]^+ = L(f)L(f)^H,$$

(5)

$$L(f) = U(f)\text{diag}(1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, \ldots, 1/\sqrt{\lambda_L}).$$

(6)

It can be proved [6] that if the covariance of the sources $S(f, t)$ in $t_i$ is negligible, every $L(f)^HR_t(f)L(f)$ for any $i$ shares the same singular vectors, and this is given via SVD form as

$$L(f)^HR_t(f)L(f) = T(f)\text{diag}(\sigma_1(t_i), \sigma_2(t_i), \ldots, \sigma_L(t_i))T(f)^H,$$

(7)

where $\sigma_k(t_i)$ are the singular values for a specific time block $t_i$, and $T(f)$ denotes the matrix consisting of shared singular vectors which are independent of time-block index $i$. Therefore, for any $i$, the simultaneous diagonalization of $R_t(f)$ can be achieved as follows:

$$T(f)^HL(f)^HR_t(f)L(f)T(f) = \text{diag}(\sigma_1(t_i), \sigma_2(t_i), \ldots, \sigma_L(t_i)),$$

(8)

and this means that the optimal separation filter matrix in the 2nd-order sense is given by

$$W_{\text{SO}}(f) = (L(f)^HT(f))^H.$$ (9)

Note that, for the calculation of $T(f)$ in Eq. (7), it is sufficient for us to only apply a single SVD to an arbitrary single time-block $t_i$ because of the singular vector-sharing property.

It is worth mentioning that Molgedey et al. have shown the closed-form solution only for the case that the number of correlation matrix blocks is up to 2 [8]. In contrast, the algorithm [6] used in the proposed method is the first generalized closed-form solution which can be applicable even to the case of $i > 2$.

3.2. Nonclosed-form HO-ICA

The separation filter matrix $W_{\text{SO}}(f)$ obtained by SO-ICA often provides insufficient source-separation performance. To polish up the separation filter matrix and gain the further performance, we have proposed the BSS method cascading the nonclosed-form HO-ICA after the SO-ICA. This strategy regards the separation filter matrix $W_{\text{SO}}(f)$ as an initial value for HO-ICA’s iterative learning. The HO-ICA is conducted by the following manner:

$$W_{\text{HO}}^{[0]}(f) = W_{\text{SO}}(f),$$

(10)

$$W_{\text{HO}}^{[j+1]}(f) = \eta \left[ I - \langle \Phi(y(f,t))y(f,t)^H \rangle_{\text{time}} \right] W_{\text{HO}}^{[j]}(f) + W_{\text{HO}}^{[j]}(f),$$

(11)

where superscript $[j]$ represents the number of iterations, $I$ is the identity matrix, $\langle \cdot \rangle_{\text{time}}$ denotes the time-averaging operator over whole time indices, and $\Phi(\cdot)$ is the appropriate nonlinear vector function, e.g., [1, 9] (we use [9] in this paper).
4. Proposed method

4.1. Overview

For a real-time application, less computational cost method rather than our previously proposed method is required. Therefore, we propose a more faster convergence method.

In general, the nonclosed-form HO-ICA suffers from poor and slow convergence due to the nonlinear optimization. To mitigate the drawback of the problem, we have proposed to combine HO-ICA with the closed-form SO-ICA. This method shows the superiority of closed-form SO-ICA for the initial value on nonclosed-form HO-ICA[7]. In this paper, we improve our method by introducing more faster convergence method by frequency-subband selection which is aimed to reduce nonclosed-form HO-ICA's computations.

4.2. Algorithm

Figure 2 shows the examples of the signal-to-noise ratio (SNR) of the separation result by closed-form SO-ICA and the nonclosed-form HO-ICA in each frequency, respectively. As shown in Fig. 2, SNRs in some frequency subbands do not improve by nonclosed-form HO-ICA. Therefore, such frequency-subbands can not contribute to SNR improvement even if HO-ICA is conducted.

In the proposed method, we do not utilize such frequency-subbands in HO-ICA. This frequency-subband selection can reduce the computational cost. The proposed method can achieve higher performance than the previously proposed method under the same computational cost. This is due to the fact that the proposed method can increase the number of iterations for optimization. This frequency-subband selection is performed on the basis of the following proper cost function:

\[ J(f) = \| I - (\Phi(Y(f, t))Y^H(f, t))_R \|_F, \]

where \( \| \cdot \|_F \) denotes the frobenius norm operator. Equation (12) is the frobenius norm based on the part of Eq. (11), which is close to 0 when the high separation performance is achieved. In addition, the computational cost of calculating the cost function is almost the same as that of one iteration in the HO-ICA. Also, for only the selected frequency-subband, we update the separation filter \( W_{HO}(f) \) using Eq. (11).

Figure 3 shows the block diagram of \( W_{HO}(f) \) in proposed method.

4.3. Computational cost in proposed method

Computational cost in the closed-form SO-ICA approximately depends on the cost of obtaining \( R_r(f) \) because the calculational cost of \( L(f) \) and \( T(f) \) is relatively little when \( K \) is small, e.g., 2 or 3. The whole computational cost in the closed-form solution is almost the same as those for 1 or 2 iterations in HO-ICA. In addition, the computational load in the frequency selection in Eq.(12) is just the same as that in 1 iteration of HO-ICA. Thus almost all the computational resources can be dedicated to HO-ICA part. Furthermore, the entire computational complexities can be reduced because the good initial value given by the closed-form ICA decreases the number of iterations in the following HO-ICA’s updating.

5. Experiment

5.1. Experimental setup

To evaluate the effectiveness of the proposed method, we carried out sound-separation experiments in a real reverberant room illustrated in Fig. 4, where two sound sources and two omni-directional microphones are set. The reverberation time in this room is 200 ms. Two speech signals are assumed to arrive from different directions, \( \theta_1 \) and \( \theta_2 \), where \( (\theta_1, \theta_2) = (-30^\circ, 20^\circ) \). We used the speech signals spoken by two male and two female speakers as the source samples, and we generated 12 combinations of speakers. The sampling frequency is 8 kHz and the length of each sound sample is limited to 6 s. The DFT size is 1024, and the frame shift length is 256. Noise reduction rate (NRR)[10], i.e., the output SNR minus the input SNR, is used as the objective indication of separation performance.

5.2. Experimental result

Figure 5 shows the computational cost required to achieve NRR of 15 dB. The parameter \( \gamma \) in Fig. 5 is the decimate ratio on frequency-subband for the nonclosed-form HO-ICA, and the case \( \gamma = 0.0 \) corresponds to the conventional method. The axis of ordinate in Fig. 5 shows computational cost. As shown in Fig. 5, in the case \( \gamma = 0.5 \) or \( \gamma = 0.7 \), the computational cost required to achieve NRR of 15 dB is reduced by selecting frequency-subband for the HO-ICA, i.e., the proposed method improves the convergence speed.

6. Conclusion

- 190 -

- 221 -
In this paper, we introduce the previously proposed BSS method cascading closed-form SO-ICA with nonclosed-form HO-ICA. Also, we proposed more fast convergence BSS method than our previously proposed method by selecting frequency-subband for the HO-ICA. The result of signal separation experiments shows that the convergence speed of proposed method overtakes that of the previously proposed method.

References

Figure 3: Block diagram of updating separation filter $W_{HO}(f)$ in proposed method.

Figure 4: Layout of reverberant room used in experiment.

Figure 5: Computational cost required until NRR achieves 15 dB.