EFFICIENT BLIND SOURCE SEPARATION COMBINING CLOSED-FORM SECOND-ORDER ICA AND NONCLOSED-FORM HIGHER-ORDER ICA

Kentaro Tachibana, Hiroshi Saruwatari, Yoshimitsu Mori, Shigeki Miyabe, Kiyohiro Shikano†, Akira Tanaka†
†Nara Institute of Science and Technology, Ikoma, Nara, 630-0192, JAPAN
‡Hokkaido University, Kita-14, Nishi-9, Kita-ku, Sapporo, 060-0814, JAPAN

ABSTRACT
In this paper, first, we propose a computational-cost efficient blind source separation combining closed-form 2nd-order independent component analysis (ICA) and nonclosed-form higher-order ICA. The closed-form solution of the 2nd-order ICA has been recently presented by one of the authors [8]. This mathematical contribution yields an idea of combining the closed-form 2nd-order ICA and the higher-order ICA, where the preceding closed-form ICA can produce a good initial value and the following higher-order ICA can update the separation filters from the advantageous status. Second, we utilize the proposed architecture to address an essential question that which type of statistics is more beneficial to ICA among non-stationarity and non-Gaussianity. This can be conducted owing to the attractive property that the closed-form ICA can provide a good estimate of the theoretical upper limitation of the separation performance among non-stationary ICAs without suffering from poor-convergence problems. Experimental results reveal that the non-Gaussianity-based ICA can outperform the non-stationarity-based ICA.

Index Terms— Separation, speech enhancement, acoustic arrays, acoustic signal processing, adaptive signal processing

1. INTRODUCTION
Blind source separation (BSS) is the approach taken to estimate original source signals using only the information of the mixed signals observed in each input channel. 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 the directions-of-arrival of the sound sources. Owing to the attractive features of BSS, much attention has been paid to BSS in many fields of signal processing such as speech enhancement.

In recent researches of BSS based on independent component analysis (ICA), various methods have been proposed to tackle acoustic-sound separation [1]–[6] which is referred to as convolutive mixing problem. This paper also addresses the BSS problem under reverberant conditions which often arise in many practical audio applications. Generally speaking, almost all the algorithms in ICA, e.g., 2nd-order ICA [2, 3, 5, 6] and higher-order ICA [1, 4] are conducted through nonclosed-form, in other words, iterative, optimization, where the separation filters are improved along with the gradient of an appropriate cost function. However, this property often leads to the difficult problem of the poor and slow convergence [7]. In addition, the latency in the convergence prevents ICA-based BSS from being applicable to real-time processing.

In this paper, first, we newly propose an efficient BSS method combining closed-form 2nd-order ICA and nonclosed-form higher-order ICA. The closed-form solution of the 2nd-order ICA has been recently presented by one of the authors [8]. This mathematical contribution yields an idea of combining the closed-form 2nd-order ICA and the higher-order ICA, where the preceding closed-form ICA can produce a good initial value and the following higher-order ICA can update the separation filters from the advantageous status.

Secondly, based on the above-mentioned structure, we address an essential question that which cost function is better among non-stationarity (on 2nd-order ICA) and non-Gaussianity (on higher-order ICA). This can be conducted using the proposed method’s attractive property that the closed-form ICA approximately shows the theoretical upper limitation of the separation performance among 2nd-order ICAs without suffering from poor-convergence problems. The evaluation of the separation performance in the proposed combination easily indicates the winner of non-stationarity vs. non-Gaussianity in ICA.

2. MIXING PROCESS AND CONVENTIONAL ICA
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.

Multiple mixed signals are observed at the microphone array, and these signals are converted into discrete-time series via an A/D converter. By applying the short-time discrete-time Fourier transform framework, we can express the observed signals, in which multiple source signals are linearly mixed, as follows in the time-frequency domain:

\[ x(f, t) = A(f)x_f(t), \]

where \( x(f, t) = [x_1(f, t), \ldots, x_K(f, t)] \) is the observed signal vector, and \( x_f(t) = [s_1(t), \ldots, s_L(t)] \) is the source signal vector. Also, \( A(f) \) is the mixing matrix which is complex-valued because we introduce a model to deal with the relative time delays among the microphones and room reverberations.

Next, we perform signal separation using the complex-valued unmixing matrix \( W(f) \), so that the \( L \) time-series output \( y(f, t) = y_1(f, t), \ldots, y_L(f, t) \) becomes mutually independent; this procedure can be given as

\[ y(f, t) = W(f)x(f, t). \]

We perform this procedure with respect to all frequency bins.

The optimal \( W(f) \) is obtained by many types of ICAs, where several cost functions are used to measure the independence among sources. The most popular statistics used in the cost functions are non-stationarity and non-Gaussianity. For example, the conventional 2nd-order ICA utilizes non-stationarity of sources. The optimization can be achieved by minimizing, e.g., the following function

\[ \min_{W} \mathbb{E}[\mathbb{E}[\sum_{k=1}^{L} |x_k(f, t) - W(f)s_k(t)|^2|^2]]. \]
In a previous study, closed-form solution of the 2nd-order ICA yielded the joint PDF of \( y(l, t) \).

The principal of the closed-form 2nd-order ICA is derived, especially from the mathematical point of view. This finding has motivated us to combine the proposed method, however, the preceding closed-form 2nd-order ICA has often provided insufficient source-separation performance. To polish up the separation filter matrix and gain the further performance, we propose to combine the nonclosed-form higher-order ICA after the 2nd-order ICA. This strategy regards the separation filter matrix \( W_{SO}(f) \) as an initial value for higher-order ICA's iterative learning.

Next, we apply the singular value decomposition (SVD) to a superposition of \( R_{ti}(f) \), which is represented as

\[
\sum_{i} R_{ti}(f) = U(f) \text{diag}(\lambda_1, \lambda_2, ...) U(f)^H,
\]

where \( \lambda_k \) are the eigenvalues, \( \text{diag}(\lambda_1, ...) \) denotes the diagonal matrix which includes the eigenvalues, and \( U(f) \) is the matrix consisting of the eigenvectors. Then we obtain a full-rank decomposition for pseudo-inverse of \( \sum_{i} R_{ti}(f) \) as follows

\[
\left[ \sum_{i} R_{ti}(f) \right]^+ = L(f) L(f)^H,
\]

\[
L(f) = U(f) \text{diag}(1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, ...).
\]

3. PROPOSED METHOD

3.1. Motivation

In a previous study, closed-form solution of the 2nd-order ICA was proposed by one of the authors [8], who showed that simple algebraic calculations enable the separation of mixed signals without iterative filter updating. This finding has motivated us to combine the closed-form 2nd-order ICA and higher-order ICA, where the computational cost is considerably reduced (see Sect. 3.4). Moreover, our strategy provides a good tool for an insight into the essential question that which cost function is better among non-stationarity and non-Gaussianity (see Sect. 3.5). Hereinafter we describe the detailed algorithm.

3.2. First stage: closed-form 2nd-order ICA

In the original reference [8], the principle of the closed-form 2nd-order ICA was derived, especially from the mathematical point of view. This sub-section briefly describes the overview of signal processing in the closed-form ICA. The strict proofs of the theorem will be omitted due to the limitation of the current manuscript's space.

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

\[
R_{ti}(f) = \langle x(f, t) x(f, t) H \rangle_{t_1 t_i},
\]

where \( \langle \cdot \rangle_{t_1 t_i} \) denotes the time-averaging operator over specific time duration \( t_i \), and \( i = 1, 2, ... \) represent indices of time-averaging block.
3.4. Computational-cost efficiency of proposed method

In the first stage, the closed-form 2nd-order ICA mainly requires the following computations:

Calculation of correlation matrices: The computations for obtaining $R_t(f)$ result in, e.g., more than hundred multiplications-accumulations to deal with the observed signal of several seconds.

Calculation of $L(f)$: To obtain $L(f)$ in Eq. (8), a single SVD should be performed as in Eq. (6), where the computational load is $O(K^3)$ ($K$ corresponds to the dimension of $L(f)$).

Calculation of $T(f)$: The matrix $T(f)$ in Eq. (11) needs one more SVD in Eq. (9) with the computational load of $O(K^3)$.

In summary, overall amount of computations in the closed-form 2nd-order ICA approximately depends on the cost of obtaining $R_t(f)$ because the calculations of $L(f)$ and $T(f)$ are relatively negligible when $K$ is small, e.g., 2 or 3. In addition, it should be mentioned that the whole computations in the closed-form solution are almost the same as those for 1 or 2 iterations in the higher-order ICA, and thus almost all the computational resources can be dedicated to the higher-order ICA part in the second stage. Furthermore, the computational complexities can be totally reduced because the good initialization by the closed-form ICA saves the number of iterations in the following higher-order ICA's updating.

3.5. As a judging tool for non-stationarity vs. non-Gaussianity

Another contribution of the closed-form 2nd-order ICA is concerned with a comparison on non-stationarity and non-Gaussianity. Owing to Eq. (10), $W_{SO}(f)$ given by Eq. (11) can diagonalize each correlation matrix when the covariance of $s(t)$ is negligible. Consequently under such a condition, the cost function defined by Eq. (3) is minimized to be zero, i.e., the following relation holds,

$$J_{so}(W_{SO}(f)) = 0.$$ (14)

This remains us that the closed-form solution Eq. (11) gives a good estimate of the theoretical upper limit of the separation performance among the 2nd-order ICAs based on source non-stationarity. Note that there are no affections from poor-convergence and local-minimum problems which often arise in the conventional nonclosed-form (iterative) method. Therefore, by seeing the results of the first stage and the possible performance increase/decrease by the second stage, we can put a period to the discussion on non-stationarity vs. non-Gaussianity in ICA, i.e., the increase implies the superiority of non-Gaussianity.

4. EXPERIMENTS AND DISCUSSIONS

4.1. Experimental conditions

To evaluate the efficacy of the proposed method, we carried out sound-separation experiments in a real reverberant room illustrated in Fig. 1, where two sources and two directional microphones (stereo-microphone) 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 we prepare three kinds of source direction patterns as follows; $(\theta_1, \theta_2) = (-90^\circ, -10^\circ), (-10^\circ, 0^\circ), \text{or } (30^\circ, 60^\circ)$. We used the speech signals spoken by two male and two female speakers as the source samples, and we generated 6 combinations of speakers. The sampling frequency is 8 kHz and the length of each sound sample is limited to 3 s. The DFT size is 1024, and the frame shift length is 256. The block size for calculation of each $R_t(f)$ is set to 1.5 s in the closed-form 2nd-order ICA part.

4.2. Evaluation of computational-cost efficiency

In order to compare the proposed method with several conventional ICAs, we prepare two higher-order ICAs with different initial filter matrix $W^{(0)}(f)$ as (A) a matrix which has entries of random complex value, and (B) identity matrix. The step-size parameter $\eta$ in the higher-order ICA is fixed to 0.1 throughout the experiments.

Noise reduction rate (NRR) [4], defined as the output signal-to-noise ratio (SNR) in dB minus the input SNR in dB, is used as the objective indication of separation performance. The SNRs are calculated under the assumption that the speech signal of the undesired speaker is regarded as noise. Figures 2–4 show the convergence curves of NRR under different speaker allocations. As for the proposed method, we plot the results only in the higher-order ICA part.

These scores are the averages of 6 speaker combinations.

From the results, we first confirm that the closed-form 2nd-order ICA can score the NRRs of 8–10 dB (see the point of Number of iterations = 0) regardless of the speaker directions. This consistent and tolerable performance is very attractive if we take into account the low computational cost. Application of the higher-order ICA in the second stage can remarkably improve the separation performance, and the proposed BSS outperforms all of the conventional methods, especially on its convergence time.

4.3. Judge of non-stationarity vs. non-Gaussianity in speech

In this subsection, we compare the 2nd-order ICA (non-stationarity) and higher-order ICA (non-Gaussianity). Here we prepare 3 s (max $i = 2$) or 36 s (max $i = 64$) long observed signals to consider the dependence on the data length. The higher-order ICA after the closed-form 2nd-order ICA is conducted as two manners; learning with full-length data, or with 1.5 s data, to equalize the data-size effect of time-block averaging.

As shown in Figures 5–7, the performance of the 2nd-order ICA is slightly improved as the observed data length increases, but it cannot reach the level of the higher-order ICA at all. Although the experimental results are very limited number of evidences, we can speculate that non-Gaussianity is more beneficial than non-stationarity in ICA for speech signals inherently. This result is also consistent with previous all-iteration-type ICAs' results (see, e.g., [6]; unfortunately these methods still suffer from local-minima problem). As far as we know, our new comparison and conclusion are the world's first appearances because we derive it via closed-form method, and this can help the further investigations on this topic.

5. CONCLUSION

First, we proposed a new efficient BSS method combining closed-form 2nd-order ICA and nonclosed-form higher-order ICA, where the preceding closed-form ICA can provide a good initial value and the following higher-order ICA can update the separation filters from the advantageous status. This enables us to reduce the computational
complexities without deteriorating the separation performance. Secondly, using the proposed method, we compare two types of ICAs with non-stationarity and non-Gaussianity. Experimental results reveal that the performances of the 2nd-order ICA are inferior to those of the higher-order ICA.

6. REFERENCES