Log-likelihood Method to Select Initial Values of Multichannel Non-negative Matrix Factorization

Fuminori Yoshiyama*, Shingo Ueno*ha*, Keisuke Nishijima*
Yusuke Hioka†, Senior Member, IEEE, and Ken’ichi Furuya*, Senior Member, IEEE
* Department of Computer Science and Intelligent Systems, Oita University, Japan
† Department of Mechanical Science and Engineering, University of Auckland, New Zealand

Abstract—A multichannel extension of non-negative matrix factorization (NMF) associates the spatial property of the sources with each of the NMF bases. An initial-value selection method based on log-likelihood for multichannel non-negative matrix factorization (MNMF) is introduced to reduce the variation of the source separation performance. Experimental results showed selecting initial values that provide high log-likelihood would improve the source separation performance of MNMF depending on the sources.

I. INTRODUCTION

Along with the spread of mobile consumer electronics devices with high-speed processors, problems regarding audio such as speech enhancement and noise reduction have been attracting more interests from engineers [1]. Such problems have been known as source separation problem for long time; non-negative matrix factorization (NMF) [2] is a state-of-the-art source separation method which has also been applied to audio signals [3]. The NMF decomposes a signal given in a matrix form into two components (NMF bases and activation matrices) assuming the components take only non-negative values. Nonetheless the standard NMF setting is only suited to a single-channel input, more than one microphones are available on modern consumer electronics devices in these days.

Recently, multichannel extensions of NMF have been receiving attention due to their potential to realize sound source separation with multiple microphones [4]. Multichannel NMF (MNMF) [5] handles complex-valued multichannel observations to maintain the spatial property of the sources, thus Hermitian positive semidefiniteness is used instead of non-negativity. However, it is known that the source separation performance achieved by MNMF is variable due to the selection of the initial values of the iterative method used in NMF.

In this paper, we propose an initial-value selection method based on the log-likelihood of the performance distribution of MNMF. To select appropriate initial values, the proposed method estimates the separation performance for the initial values using the log-likelihood of observations with appropriate statistical models so that the MNMF algorithm can decompose a signal more precisely.

II. MULTICHANNEL NMF

A. Formulation

Let $M$ be the number of microphones, and $x_{ij} = [x_1, \ldots, x_M]^T \in \mathbb{C}$ be a complex valued vector. Observation matrix of MNMF is represented as

$$X_{ij} = x_{ij}x_{ij}^H$$

(1)

where $x_{ij}$ is the observation signal of each microphones. Thus, $X$ is $M \times M$ Hermitian matrix, which can be approximated by $\hat{X}_{ij}$ as follows.

$$X_{ij} \approx \hat{X}_{ij} = \sum_{l} (\sum_{i} H_{il}z_{il})t_{lk}v_{kj}$$

(2)

Fig. 1 shows an example in which MNMF factorizes a hierarchical representation. The parameters $I$, $J$, $K$, and $L$ specify frequency bin, time frame, number of bases, and number of sources, respectively.

B. Cost function

The cost function utilized in MNMF is defined by an arbitrary distance/divergence between $X$ and $\hat{X}$. In this paper, we use the following multichannel Itakura-Saito (IS) divergence:

$$D_{IS}(X_{ij}, \hat{X}_{ij}) = \log N_c(x_{ij}|0, X_{ij}) - \log N_c(\hat{x}_{ij}|0, \hat{X}_{ij})$$

$$= \text{tr}(X_{ij}X_{ij}^{-1}) - \log \det X_{ij}X_{ij} - M.$$

(3)

where $\text{tr}(X) = \sum_{m=1}^{M} x_{mm}$ is the trace of a square matrix $X$. Consequently, the cost function of MNMF is defined:

$$D_{IS}(X, \{T, V, H, Z\}) = \sum_{i=1}^{J} \sum_{j=1}^{J} D_{IS}(X_{ij}, \hat{X}_{ij}).$$

(4)

MNMF is formulated to minimize the cost function, which is achieved by employing multiplicative update rules [5]. Matrices $T$, $V$ are randomly initialized with non-negative entries. The diagonal elements of matrix $H$ are initially all set at $1/M$, and the off-diagonal elements are initially all set at zero. The elements of matrix $Z$ are initialized with random values around $1/L$.

III. INITIAL-VALUE SELECTION METHOD USING LOG-LIKELIHOOD

In the original study of MNMF the initial values $T$, $V$ and $Z$ are randomly selected, which causes the varying source separation performance. The initial-value selection method

Fig. 1. Illustration of MNMF. Non-negative values are shown in gray and complex values are shown in white.
introduced in this study estimates initial values that would deliver high separation performance by MNMF. The method determines the initial values by first running MNMF algorithm with some randomly selected initial values, then it picks up the best initial-value using the log-likelihood method.

The log-likelihood method employs IS divergence as a cost function since IS divergence includes the log-likelihoods of both the observation matrix and the estimation matrix: minimizing the IS divergence is equivalent to maximizing the log-likelihood. Assuming that high source separation performance is achieved when IS divergence is minimized, the log-likelihood method selects the initial values that has provided the minimum value of IS divergence so that MNMF satisfies the following

$$\arg \min_{\mathbf{X}} D_{IS}(\mathbf{X}, \hat{\mathbf{X}})$$

Following selection approaches are employed to specify the initial values:

1) Randomly initialize the initial values,
2) Apply 500 iterations to the MNMF update,
3) Conduct 1) and 2) for some trials,
4) Select the initial values which provide the minimum IS divergence by calculating (4).

IV. EXPERIMENT

To confirm the efficacy of the proposed method, we compared the average source separation performance of the proposed method with that of the random initial-value selection method. Four sets of audio mixture signals were created by music sources listed in Table I. The input signals were generated by convolving the impulse responses measured in a real room, the setup of which is shown in Fig. 2, to the source signals, and then the first 300 frames were extracted.

### TABLE I

<table>
<thead>
<tr>
<th>ID</th>
<th>Author/Song</th>
<th>Snip</th>
<th>Part</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bearlin</td>
<td>85-99</td>
<td>piano</td>
</tr>
<tr>
<td></td>
<td>Roads</td>
<td>(14 sec)</td>
<td>ambient</td>
</tr>
<tr>
<td>2</td>
<td>Fort Minor</td>
<td>69-94</td>
<td>drums</td>
</tr>
<tr>
<td></td>
<td>Remember The Name</td>
<td>(24 sec)</td>
<td>vocals</td>
</tr>
<tr>
<td>3</td>
<td>Another Dreamer</td>
<td>69-94</td>
<td>drums</td>
</tr>
<tr>
<td></td>
<td>The Ones We Love</td>
<td>(25 sec)</td>
<td>vocals</td>
</tr>
<tr>
<td>4</td>
<td>Ultimate Nz Tour</td>
<td>54-78</td>
<td>drums</td>
</tr>
<tr>
<td></td>
<td>Tour</td>
<td>(18 sec)</td>
<td>guitar</td>
</tr>
</tbody>
</table>

The sampling frequency of the signals was 16 kHz, the frame size used for the short-time Fourier transform (STFT) was 1024 samples and the frame was shifted by every 256 samples. The number of NMF bases and the cluster number were set to \( K = 30 \) and \( L = 3 \), respectively. The source separation performance was quantitatively evaluated in terms of the signal-to-distortion ratio (SDR) \[6\].

Fig. 3 shows the source separation performance result obtained by MNMF algorithm with different initial-value selection algorithms, i.e. the log-likelihood (proposed) and random methods. To select 10 initial values using the log-likelihood method, we conducted 100 trials with different initializations for each set of audio mixture signals. The variation of source separation performance was significantly decreased with the signal set ID4 whereas the improvement was less significant with the other signal sets. This implies the effectiveness of the proposed method would depend on the spectral characteristics of source signals.

V. CONCLUSION

An initial-value selection method based on the log-likelihood has been proposed for improving the source separation performance of MNMF algorithm. The method employs IS divergence in the cost function to measure the log-likelihood. Experimental result has shown that the proposed method is effective to significantly increase the separation performance with some audio mixture signals. Further study is needed to improve the method to be robust to the change of source signals.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant number 25350670.

REFERENCES