FORMULATIONS AND ALGORITHMS FOR MULTICHANNEL COMPLEX NMF

Hiroshi Sawada, Hirokazu Kameoka, Shoko Araki, Naonori Ueda

NTT Communication Science Laboratories, NTT Corporation, Japan

ABSTRACT

This paper studies some formulations and algorithms for the multichannel extension of nonnegative matrix factorization (NMF). We model the inter-channel characteristics of each NMF basis, including both the amplitude ratios and the phase differences on a channel pair. The learned inter-channel characteristics provide useful information for binding each NMF basis to each source component in such a situation that multiple sources are mixed in a convolutive manner and observed at multiple microphones. Effective optimization algorithms based on majorization are derived by using properly designed auxiliary functions. Experimental results show that the algorithms converged favorably regardless of the initialization.

Index Terms— Nonnegative matrix factorization, Complex phase, Inter-channel characteristic, Majorization, Auxiliary function

1. INTRODUCTION

Extracting or identifying frequent sound patterns in recorded sounds is an important task in various kinds of audio signal applications, including sound separation and music transcription. Nonnegative matrix factorization (NMF) [1], which was originally applied to learning parts-based representations of images and to the analysis of documents [2], is also a widely-used technique for such audio applications (e.g., [3]). When we analyze a time-domain single-channel audio signal $x(t)$ with NMF, the signal is typically transformed into the representation x_{ij} in the time-frequency domain by a short-time Fourier transform (STFT), where i and j represent time and frequency, respectively. Then the amplitude $|x_{ij}|$ is used to construct a nonnegative matrix for the NMF analysis.

Some extensions of NMF have been proposed to improve the capabilities for audio applications. Complex NMF [4] takes account of the phase information $x_{ij}/|x_{ij}|$, which is discarded with standard NMF. It improves the analysis accuracy by faithfully obeying the mixing process in the complex domain. Another avenue for extension is to take advantage of multichannel recordings (typically stereo). Research effort has been devoted to considering an instantaneous mixing case [5, 6] and a convolutive mixing case [7]. In this paper, we study a convolutive mixing case, with particular focus on extracting the inter-channel characteristics, including the phase information, of the NMF decomposed components.

2. PRELIMINARIES

2.1. NMF and Complex NMF

Let us start with a single channel case before moving on to a multichannel case in Sect. 3. Let $x(t)$ be a microphone observation for a specific time duration, say $t = 1, \ldots, T$. By applying an STFT to the observation, we have a complex matrix $\mathbf{X}$ whose (i, j)-element is denoted by $x_{ij} \in \mathbb{C}$. Again, i and j represent time and frequency, respectively. Let the matrix size be $I \times J$.

When we employ the standard NMF [1] for the analysis, we only consider the amplitudes $\mathbf{Y} = |\mathbf{X}|$, or equivalently $y_{ij} = |x_{ij}|$, $^\forall i, j$. The generative model can be written as

$$p(\mathbf{Y}|\mathbf{T},\mathbf{V}) = \prod_{i,j} \mathcal{N}(y_{ij}|\sum_{k=1}^K t_{ik}v_{kj}, \tfrac{1}{2}) \tag{1}$$

where $\mathcal{N}$ represents a Gaussian distribution, and K is the number of rank-1 basis matrices. Nonnegative matrices $\mathbf{T}$ and $\mathbf{V}$, whose elements are $t_{ik} \geq 0$ and $v_{kj} \geq 0$, have sizes of $I \times K$ and $K \times J$, respectively.

If we preserve the phase information of $\mathbf{X}$, complex NMF [4] can be applied. The generative model can be written as

$$p(\mathbf{X}|\mathbf{T},\mathbf{V},\mathbf{G}) = \prod_{i,j} \mathcal{N}_c(x_{ij}|\sum_{k=1}^K g_{ijk}t_{ik}v_{kj}, 1) \tag{2}$$

where $\mathcal{N}_c$ is a complex Gaussian distribution, and $\mathbf{G}$ is a third-order tensor whose (i, j, k)-element is a complex $g_{ijk} \in \mathbb{C}$ with a unit amplitude constraint $|g_{ijk}| = 1$.

2.2. Iterative optimization

The negative log likelihood of (1) or (2) can be minimized in an iterative manner by majorization [8] with a properly designed auxiliary function. This section briefly explains the procedure for standard NMF. A similar procedure can be derived for complex NMF as presented in [4].

The negative log-likelihood of (1) is given by

$$\mathcal{L}(\mathbf{T},\mathbf{V}) = \sum_{i,j} |y_{ij} - \sum_{k=1}^K t_{ik}v_{kj}|^2 . \tag{3}$$

Now, we introduce an auxiliary function

$$\mathcal{L}^+(\mathbf{T},\mathbf{V},\mathbf{R}) = \sum_{i,j}\left[y_{ij}^2 - 2y_{ij}(\textstyle\sum_k t_{ik}v_{kj}) + \sum_k \frac{t_{ik}^2 v_{kj}^2}{r_{ijk}}\right] \tag{4}$$

that satisfies

1. $\mathcal{L}(\mathbf{T},\mathbf{V}) \leq \mathcal{L}^+(\mathbf{T},\mathbf{V},\mathbf{R})$,
2. $\mathcal{L}(\mathbf{T},\mathbf{V}) = \min_{\mathbf{R}} \mathcal{L}^+(\mathbf{T},\mathbf{V},\mathbf{R})$.

These are guaranteed according to the following equation, which is derived from Jensen's inequality

$$\left(\sum_{k=1}^K t_{ik}v_{kj}\right)^2 \leq \sum_{k=1}^K \frac{(t_{ik}v_{kj})^2}{r_{ijk}} \tag{5}$$

where $r_{ijk} \geq 0$ and $\sum_{k=1}^K r_{ijk} = 1$.

The negative log likelihood $\mathcal{L}$ is indirectly minimized by repeating the following updates, each of which corresponds to the minimization of the auxiliary function $\mathcal{L}^+$

1. with respect to $\mathbf{R}$: $r_{ijk} \leftarrow \frac{t_{ik}v_{kj}}{\sum_k t_{ik}v_{kj}}$
2. with respect to $\mathbf{T}$: $t_{ik} \leftarrow \frac{\sum_j y_{ij}v_{kj}}{\sum_j v_{kj}^2/r_{ijk}}$
3. with respect to $\mathbf{R}$: $r_{ijk} \leftarrow \frac{t_{ik}v_{kj}}{\sum_k t_{ik}v_{kj}}$
4. with respect to $\mathbf{V}$: $v_{kj} \leftarrow \frac{\sum_i y_{ij}t_{ik}}{\sum_i t_{ik}^2/r_{ijk}}$.

978-1-4577-0539-7/11/$26.00

The $\mathbf{R}$ update is derived from the equality condition of the Jensen's inequality, and the $\mathbf{T}$ and $\mathbf{V}$ updates are derived from the partial derivative of $\mathcal{L}^+$ with respect to t_{ik} or v_{kj}. By substituting r_{ijk} in the $\mathbf{T}$ and $\mathbf{V}$ updates and rearranging them, we obtain the well-known multiplicative updates for NMF [1].

3. MULTICHANNEL COMPLEX NMF

This section proposes two types of multichannel extensions of NMF. One is from complex NMF and the other is from standard NMF. For the sake of simplicity, we consider a stereo case here. Generalization to more channels is straightforward. Let $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ be complex matrices representing time-frequency domain observations of the 1st and 2nd microphone. Both extensions basically model multichannel observations with common nonnegative matrices $\mathbf{T}$ and $\mathbf{V}$, together with a newly introduced complex matrix $\mathbf{H}$ that represents the inter-channel characteristics corresponding to the matrix $\mathbf{T}$.

3.1. Extension from Complex NMF

3.1.1. Model

Let us start with the extension from complex NMF. Without loss of generality, let the 1st channel be the reference channel, where all the inter-channel characteristics are normalized to one. Then, the 1st channel is modeled exactly by complex NMF (2):

$$p(\mathbf{X}^{(1)}|\mathbf{T},\mathbf{V},\mathbf{G}) = \prod_{i,j}\mathcal{N}_c(x_{ij}^{(1)}|\sum_k g_{ijk}t_{ik}v_{kj},1)\,. \quad (6)$$

The 2nd channel is modeled by introducing an $I \times K$ matrix $\mathbf{H}$ whose elements are complex $h_{ik} \in \mathbb{C}$ and by using the same $\mathbf{T},\mathbf{V},\mathbf{G}$ matrices commonly:

$$p(\mathbf{X}^{(2)}|\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H}) = \prod_{i,j}\mathcal{N}_c(x_{ij}^{(2)}|\sum_k g_{ijk}h_{ik}t_{ik}v_{kj},1)\,. \quad (7)$$

It should be emphasized that the inter-channel characteristic h_{ik} is a complex number and thus contains information on the phase difference as well as the amplitude ratio between the 1st and 2nd channels. When we use a microphone array whose microphone inter-distances are not very large, e.g., 4 cm, the phase information is more important than the amplitude information for distinguishing sources coming from different directions. The introduction of h_{ik} in this manner is the main contribution of this paper over the previous work on multichannel NMF [5, 6, 7].

3.1.2. Negative log-likelihood and its auxiliary function

What we want to maximize is the multichannel likelihood

$$p(\mathbf{X}^{(1)},\mathbf{X}^{(2)}|\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H}) = p(\mathbf{X}^{(1)}|\mathbf{T},\mathbf{V},\mathbf{G})\cdot p(\mathbf{X}^{(2)}|\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H}) \quad (8)$$

whose negative log-likelihood is given by

$$\mathcal{L}_{mc}(\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H}) = \sum_{i,j}\left(|x_{ij}^{(1)}-\sum_k g_{ijk}t_{ik}v_{kj}|^2 + |x_{ij}^{(2)}-\sum_k g_{ijk}h_{ik}t_{ik}v_{kj}|^2\right). \quad (9)$$

To follow the iterative optimization procedure explained in Sect. 2.2, we define an auxiliary function

$$\mathcal{L}_{mc}^+(\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H},\mathbf{S}^{(1)},\mathbf{S}^{(2)}) = \sum_{i,j}\left(\sum_k \frac{|s_{ijk}^{(1)} - g_{ijk}t_{ik}v_{kj}|^2}{r_{ijk}} + \sum_k \frac{|s_{ijk}^{(2)} - g_{ijk}h_{ik}t_{ik}v_{kj}|^2}{r_{ijk}}\right) \quad (10)$$

where $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are newly introduced tensors of size $I \times J \times K$ whose elements are complex $s_{ijk}^{(1)}, s_{ijk}^{(2)} \in \mathbb{C}$ satisfying

$$\textstyle\sum_k s_{ijk}^{(1)} = x_{ij}^{(1)}\,, \quad \sum_k s_{ijk}^{(2)} = x_{ij}^{(2)} \quad (11)$$

and $r_{ijk}, {}^\forall i,j,k$ are parameters satisfying $\sum_k r_{ijk} = 1,\ r_{ijk} \geq 0$.

3.1.3. Update rules

The negative log-likelihood $\mathcal{L}_{mc}$ is minimized in the same manner as that explained in Sect. 2.2. The minimization updates for the auxiliary function $\mathcal{L}_{mc}^+$ with respect to $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are given by

$$s_{ijk}^{(1)} = g_{ijk}t_{ik}v_{kj} + r_{ijk}(x_{ij}^{(1)} - \textstyle\sum_k g_{ijk}t_{ik}v_{kj}) \quad (12)$$

$$s_{ijk}^{(2)} = g_{ijk}h_{ik}t_{ik}v_{kj} + r_{ijk}(x_{ij}^{(2)} - \textstyle\sum_k g_{ijk}h_{ik}t_{ik}v_{kj}) \quad (13)$$

It is easily verified that with the above assignment the equality $\mathcal{L}_{mc}(\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H}) = \mathcal{L}_{mc}^+(\mathbf{T},\mathbf{V},\mathbf{G},\mathbf{H},\mathbf{S}^{(1)},\mathbf{S}^{(2)})$ and condition (11) holds. We can interpret this to mean that $s_{ijk}^{(1)}$ is a decomposition of $x_{ij}^{(1)}$ mainly with the k-th component $g_{ijk}t_{ik}v_{kj}$ plus a fraction of the error term $x_{ij}^{(1)} - \sum_k g_{ijk}t_{ik}v_{kj}$.

Regarding the parameter r_{ijk}, it can be specified arbitrarily as long as $\sum_k r_{ijk} = 1$ and $r_{ijk} \geq 0$. However, we specify them using a specific form

$$r_{ijk} = \frac{t_{ik}v_{kj}}{\sum_k t_{ik}v_{kj}} = \frac{t_{ik}v_{kj}}{\hat{x}_{ij}} \quad (14)$$

that was suggested in [4].

The minimization updates for $\mathcal{L}_{mc}^+$ with respect to $\mathbf{T},\mathbf{V},\mathbf{H},\mathbf{G}$ are derived from the partial derivatives of $\mathcal{L}_{mc}^+$ (see appendix), and given by

$$t_{ik} = \frac{\sum_j \hat{x}_{ij}\mathrm{Re}[g_{ijk}^*(s_{ijk}^{(1)} + s_{ijk}^{(2)}h_{ik}^*)]}{(1+|h_{ik}|^2)\sum_j \hat{x}_{ij}v_{kj}} \quad (15)$$

$$v_{kj} = \frac{\sum_i \hat{x}_{ij}\mathrm{Re}[g_{ijk}^*(s_{ijk}^{(1)} + s_{ijk}^{(2)}h_{ik}^*)]}{\sum_i \hat{x}_{ij}t_{ik}(1+|h_{ik}|^2)} \quad (16)$$

$$h_{ik} = \frac{\sum_j \hat{x}_{ij}s_{ijk}^{(2)}g_{ijk}^*}{t_{ik}\sum_j \hat{x}_{ij}v_{kj}} \quad (17)$$

$$g_{ijk} = \frac{s_{ijk}^{(1)} + s_{ijk}^{(2)}h_{ik}^*}{|s_{ijk}^{(1)} + s_{ijk}^{(2)}h_{ik}^*|} \quad (18)$$

where $\hat{x}_{ij} = \sum_k t_{ik}v_{kj}$, and $\mathrm{Re}[\cdot]$ extracts the real part.

In summary, the negative log-likelihood $\mathcal{L}_{mc}$ is iteratively minimized by repeating (12)-(13) and one of the four updates (15)-(18).

3.1.4. Discussion

In our preliminary experiments, the multichannel extension of complex NMF proposed here did not work very well, in the sense that h_{ik} did not successfully reflect the inter-channel characteristics. The reason is that the complex phases have too much freedom to model the observations precisely, as both g_{ijk} and h_{ik} can have any phase. There are two possible approaches to overcoming this problem. The first is to impose some prior or structural constraints to lower the freedom, for instance, a sparsity constraint on the excitation v_{kj} as we did in [4], or a clustering constraint on h_{ik}. However, this line of study will constitute our future work. The second approach is to ignore the phase g_{ijk} of each component $t_{ik}v_{kj}$, and consider only the inter-channel characteristic h_{ik}. This leads to another type of multichannel extension of NMF, which is an extension from standard NMF that we detail in the next subsection.

3.2. Extension from standard NMF

3.2.1. Preprocessing and model

When we apply standard NMF, we only consider the amplitudes of the time-frequency domain observations. This can be performed on the 1st channel (reference channel) by

$$x_{ij}^{(1)} \leftarrow x_{ij}^{(1)} \cdot \left(\frac{x_{ij}^{(1)}}{|x_{ij}^{(1)}|} \right)^* , {}^\forall i, j \, , \tag{19}$$

where $*$ denotes the complex conjugate. To preserve the inter-channel characteristics, we apply the same operation to the 2nd channel as

$$x_{ij}^{(2)} \leftarrow x_{ij}^{(2)} \cdot \left(\frac{x_{ij}^{(1)}}{|x_{ij}^{(1)}|} \right)^* , {}^\forall i, j \, . \tag{20}$$

Please note that with the above preprocessing $\mathbf{X}^{(1)}$ becomes a non-negative matrix but $\mathbf{X}^{(2)}$ remains a complex matrix.

Thus, the 1st channel is modeled exactly by (1):

$$p(\mathbf{X}^{(1)}|\mathbf{T}, \mathbf{V}) = \prod_{i,j} \mathcal{N}(x_{ij}^{(1)} | \sum_k t_{ik} v_{kj}, \tfrac{1}{2}) \tag{21}$$

and the 2nd channel is modeled by using an $I \times K$ matrix $\mathbf{H}$, whose elements are complex $h_{ik} \in \mathbb{C}$, and the same $\mathbf{T}, \mathbf{V}$ matrices as the 1st channel:

$$p(\mathbf{X}^{(2)}|\mathbf{T}, \mathbf{V}, \mathbf{H}) = \prod_{i,j} \mathcal{N}_c(x_{ij}^{(2)} | \sum_k h_{ik} t_{ik} v_{kj}, 1) \, . \tag{22}$$

The difference between the model shown in Sect. 3.1.1 and the above model lies only in the existence of the complex coefficients g_{ijk}.

3.2.2. Negative log-likelihood and update rules

Along with the derivations similar to those described in Sect. 3.1.2 and 3.1.3, we have the following results.

The negative log-likelihood to be minimized is given by

$$\mathcal{L}_m(\mathbf{T}, \mathbf{V}, \mathbf{H}) = \sum_{i,j} \left(|x_{ij}^{(1)} - \sum_k t_{ik} v_{kj}|^2 + |x_{ij}^{(2)} - \sum_k h_{ik} t_{ik} v_{kj}|^2 \right), \tag{23}$$

and its auxiliary function can be defined by

$$\mathcal{L}_m^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{S}^{(1)}, \mathbf{S}^{(2)}) = \sum_{i,j} \left(\sum_k \frac{|s_{ijk}^{(1)} - t_{ik} v_{kj}|^2}{r_{ijk}} + \sum_k \frac{|s_{ijk}^{(2)} - h_{ik} t_{ik} v_{kj}|^2}{r_{ijk}} \right) \tag{24}$$

where $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are tensors of size $I \times J \times K$ whose elements are real $s_{ijk}^{(1)}$ and complex $s_{ijk}^{(2)} \in \mathbb{C}$ and satisfying (11).

The updates with respect to $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are given by

$$s_{ijk}^{(1)} = t_{ik} v_{kj} + r_{ijk}(x_{ij}^{(1)} - \sum_k t_{ik} v_{kj}) \tag{25}$$

$$s_{ijk}^{(2)} = h_{ik} t_{ik} v_{kj} + r_{ijk}(x_{ij}^{(2)} - \sum_k h_{ik} t_{ik} v_{kj}) \tag{26}$$

and those with respect to $\mathbf{T}, \mathbf{V}, \mathbf{H}$ are given by

$$t_{ik} = \frac{\sum_j \hat{x}_{ij} \mathrm{Re}(s_{ijk}^{(1)} + s_{ijk}^{(2)} h_{ik}^*)}{(1 + |h_{ik}|^2) \sum_j \hat{x}_{ij} v_{kj}} \tag{27}$$

$$v_{kj} = \frac{\sum_i \hat{x}_{ij} \mathrm{Re}(s_{ijk}^{(1)} + s_{ijk}^{(2)} h_{ik}^*)}{\sum_i \hat{x}_{ij} t_{ik} (1 + |h_{ik}|^2)} \tag{28}$$

$$h_{ik} = \frac{\sum_j \hat{x}_{ij} s_{ijk}^{(2)}}{t_{ik} \sum_j \hat{x}_{ij} v_{kj}} \tag{29}$$

Table 1. Experimental conditions

Sampling rate	16 kHz
STFT frame size and shift	1024 (64 ms) and 256 (16 ms)
Reverberation time	RT_{60} = 130 ms
Source signals	6 s of simple instrumental music
Direction of sources	70° and 150° (2 sources)
Microphone distance	4 cm

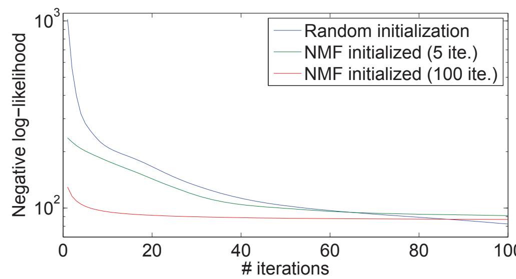

Fig. 1. Convergence behavior

where $\hat{x}_{ij} = \sum_k t_{ik} v_{kj}$, and $\mathrm{Re}(\cdot)$ extracts the real part.

In summary, the negative log-likelihood $\mathcal{L}_m$ is iteratively minimized by repeating (25)-(26) and one of the three updates (27)-(29). The set of these updates (25)-(29) can be considered a special case of the set of updates (12)-(18) shown in Sect. 3.1.3 where g_{ijk} is always fixed at 1 for all i, j, k.

4. EXPERIMENTAL RESULTS

We performed experiments to observe the behavior of the proposed multichannel NMF. As mentioned in Sect. 3.1.4, the multichannel extension from complex NMF did not work well so far. Therefore, we report here the experimental results obtained by the extension from standard NMF shown in Sect. 3.2.

The experimental conditions are summarized in Table 1. We considered a situation where two sounds of instruments coming from different directions were mixed and observed at two microphones. The mixtures were made by convolving measured impulse responses, whose reverberation time was 130 ms, and the sound sources. With the application of STFT, we had two matrices $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ both 513×404 in size. Then, we applied the preprocessing (19) and (20), followed by the multichannel NMF updates (25)-(29). The number K of rank-1 basis matrices was set at 10.

Figure 1 shows the convergence behavior in minimizing the negative log-likelihood (23). We have examined three ways to initialize $\mathbf{T}, \mathbf{V}$ matrices. The first was to initialize them with random matrices. The other two involved initializing them with the result of standard NMF applied to $\mathbf{X}^{(1)}$. The second and third were different as regards the number of standard NMF iterations. The $\mathbf{H}$ matrix was initialized randomly in all cases. We see that the proposed updates (25)-(29) successfully minimized the negative log-likelihood regardless of the initial matrices.

Figure 2 shows the spectrogram reconstructed by the learned $\mathbf{T}$, $\mathbf{V}$ matrices, and Fig. 3 details the $\mathbf{T}$ matrix. Although these two matrices were learned appropriately, there was no explicit information on which basis corresponds to which source. The learned $\mathbf{H}$ matrix provided useful information for solving the basis-source binding problem. Figure 4 shows the phase $h_{ik}/|h_{ik}|$ of each element. We

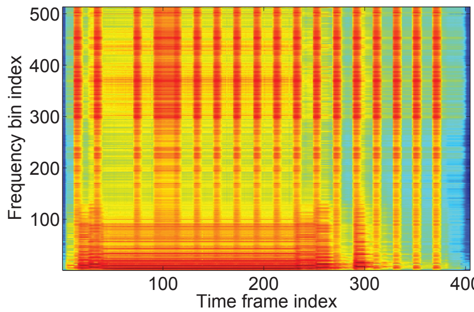

Fig. 2. Spectrogram reconstructed by learned $\mathbf{T}$, $\mathbf{V}$ matrices

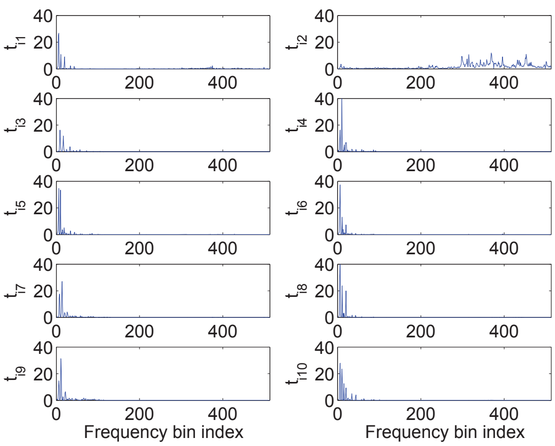

Fig. 3. Each column of $\mathbf{T}$ matrix

can see that the 1st and 2nd column bases have a tendency to negative phases whereas the others tend towards positive phases. By clustering those column bases, it was automatically decided that the 1st and 2nd bases belonged to the 1st source and the others belonged to the 2nd source. As a result of the source separation, the signal-to-distortion ratios [9] were improved from $(-12.04, 12.04)$ dB at the mixture to $(4.22, 16.21)$ dB at the separated sounds.

5. CONCLUSION

We proposed multichannel extensions of complex NMF. We started with a general formulation where the phase g_{ijk} of each component $t_{ik}v_{kj}$ is considered. We then specialized the formulation so that g_{ijk} was always fixed at 1. Experimental results for the specialized version of multichannel NMF were provided to demonstrate the effectiveness. Future work will include the design of a more sophisticated model advanced from the general formulation.

6. APPENDIX

The partial derivatives of $\mathcal{L}^+_{mc}$ (10) with respect to $\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{G}$ are shown below.

$$\frac{\partial \mathcal{L}^+_{mc}}{\partial t_{ik}} = \sum_j \frac{2}{r_{ijk}}[t_{ik}v_{kj}^2 - \mathrm{Re}(s_{ijk}^{(1)}g_{ijk}^*)v_{kj} + |h_{ik}|^2 t_{ik}v_{kj}^2 - \mathrm{Re}(s_{ijk}^{(2)}g_{ijk}^*h_{ik}^*)v_{kj}] \quad (30)$$

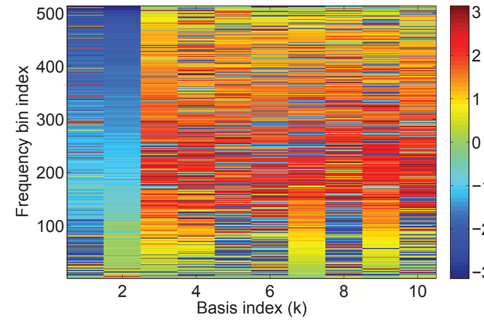

Fig. 4. Phase of each element of $\mathbf{H}$ matrix

$$\frac{\partial \mathcal{L}^+_{mc}}{\partial v_{kj}} = \sum_i \frac{2}{r_{ijk}}[t_{ik}^2 v_{kj} - \mathrm{Re}(s_{ijk}^{(1)}g_{ijk}^*)t_{ik} + |h_{ik}|^2 t_{ik}^2 v_{kj} - \mathrm{Re}(s_{ijk}^{(2)}g_{ijk}^*h_{ik}^*)t_{ik}] \quad (31)$$

$$\frac{\partial \mathcal{L}^+_{mc}}{\partial h_{ik}^*} = \sum_j \frac{t_{ik}v_{kj}}{r_{ijk}}(h_{ik}t_{ik}v_{kj} - s_{ijk}^{(2)}g_{ijk}^*) \quad (32)$$

$$\frac{\partial \mathcal{L}^+_{mc}}{\partial g_{ijk}^*} = \frac{t_{ik}v_{kj}}{r_{ijk}}(s_{ijk}^{(1)} - g_{ijk}t_{ik}v_{kj} + s_{ijk}^{(2)}h_{ik}^* - g_{ijk}|h_{ik}|^2 t_{ik}v_{kj}) \quad (33)$$

Setting these derivatives to zero and substituting r_{ijk} by (14) yields the updates (15)-(17). For the $\mathbf{G}$ update, we use the unit amplitude constraint $|g_{ijk}| = 1$ to derive (18).

7. REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[2] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. ACM SIGIR*, 2003, pp. 267–273.

[3] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2003, pp. 177–180.

[4] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proc. ICASSP 2009*, Apr. 2009, pp. 3437–3440.

[5] D. FitzGerald, M. Cranitch, and E. Coyle, "Non-negative tensor factorisation for sound source separation," in *Proc. Irish Signals Syst. Conf.*, Sept. 2005, pp. 8–12.

[6] R. M. Parry and I. A. Essa, "Estimating the spatial position of spectral components in audio," in *Proc. ICA 2006*. Mar. 2006, pp. 666–673, Springer.

[7] A. Ozerov and C. Fevotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 550–563, Mar. 2010.

[8] J. de Leeuw, "Block-relaxation methods in statistics," in *Information Systems and Data Analysis*, H. H. Bock, W. Lenski, and M. M. Richter, Eds., pp. 308–324. Springer Verlag, 1994.

[9] E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. P. Rosca, "First stereo audio source separation evaluation campaign: data, algorithms and results," in *Proc. ICA 2007*, 2007, pp. 552–559.