

NEW FORMULATIONS AND EFFICIENT ALGORITHMS FOR MULTICHANNEL NMF

Hiroshi Sawada, Hirokazu Kameoka, Shoko Araki, Naonori Ueda

NTT Communication Science Laboratories, NTT Corporation, Japan

ABSTRACT

This paper proposes new formulations and algorithms for a multichannel extension of nonnegative matrix factorization (NMF), intending convolutive sound source separation with multiple microphones. The proposed formulation employs Hermitian positive semidefinite matrices to represent a multichannel version of nonnegative elements. Such matrices are basically estimated for NMF bases, but a source separation task can be performed by introducing variables that relate NMF bases and sources. Efficient optimization algorithms in the form of multiplicative updates are derived by using properly designed auxiliary functions. Experimental results show that two instrumental sounds coming from different directions were successfully separated by the proposed algorithm.

Index Terms— nonnegative matrix factorization, multichannel, positive semidefinite, auxiliary function, source separation

1. INTRODUCTION

Identifying frequent 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, is also a widely-used technique for such audio applications (e.g., [2]). When we analyze a recorded audio signal 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 frequency and time, respectively. Then the amplitudes $|x_{ij}|$ are used to construct a nonnegative matrix for the analysis.

Recently, multichannel extensions of NMF have been receiving attention with a view of realizing sound source localization and separation with multiple microphones, and have been studied for instantaneous mixtures [3, 4] and for convolutive mixtures [5, 6]. In a convolutive case, the phase difference information on time-frequency observations between different microphones is crucial for distinguishing/identifying source directions. Therefore, complex-valued raw STFT observations should be taken into account. In [5, 6], the covariance matrices of multivariate complex Gaussian distributions are modeled with the NMF scheme, and EM algorithms are derived for a source separation task. However, the algorithms are sensitive to initialization, and thus they used the original source information for *perturbed oracle* initializations.

In this paper, we study new types of multichannel NMF formulations for a convolutive mixing case. The new formulations are based on modeling the means of complex Gaussian distributions, and can be seen as natural extensions of standard Euclidean NMF. We study two formulations in which Hermitian positive semidefinite matrices \mathbf{H} representing spatial properties are considered for each NMF basis (the first one) and for each source (the second one). In the second formulation, a source separation task can be per-

formed by automatically clustering the NMF bases to each source. This constitutes a considerable advance on our previous study [7]. The derived algorithms are in the form of multiplicative updates [1]. Experimental results show that the algorithms converge favorably.

2. OPTIMIZATION FRAMEWORK FOR NMF

This section reviews the formulation and optimization framework of standard NMF [1]. This framework can handle complex values [8] and is important for understanding our multichannel extensions of NMF explained in Sect. 3. The generative model of STFT amplitudes \mathbf{A} , $[\mathbf{A}]_{ij} = a_{ij} = |x_{ij}|$, in standard NMF can be written as

$$p(\mathbf{A}|\mathbf{T}, \mathbf{V}) = \prod_{i=1}^I \prod_{j=1}^J \mathcal{N}(a_{ij} | \sum_{k=1}^K t_{ik} v_{kj}, \frac{1}{2}) \quad (1)$$

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.

The negative log likelihood of (1)

$$\mathcal{L}(\mathbf{T}, \mathbf{V}) = -\log p(\mathbf{A}|\mathbf{T}, \mathbf{V}) = \sum_{i,j} |a_{ij} - \sum_k t_{ik} v_{kj}|^2, \quad (2)$$

where constant terms are omitted, can be minimized in an iterative manner by majorization [9] with an auxiliary function

$$\mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{S}) = \sum_{i,j} \sum_{k=1}^K \frac{1}{r_{ijk}} |s_{ijk} - t_{ik} v_{kj}|^2 \quad (3)$$

where s_{ijk} are auxiliary variables that satisfy $\sum_k s_{ijk} = a_{ij}$, and r_{ijk} are parameters that satisfy $\sum_k r_{ijk} = 1$ and $r_{ijk} > 0$. It can be verified that the auxiliary function \mathcal{L}^+ has two properties:

$$\mathcal{L}(\mathbf{T}, \mathbf{V}) \leq \mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{S}), \quad (4)$$

$$\mathcal{L}(\mathbf{T}, \mathbf{V}) = \min_{\mathbf{S}} \mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{S}). \quad (5)$$

Therefore, 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{S} : $s_{ijk} \leftarrow t_{ik} v_{kj} + r_{ijk} (a_{ij} - \sum_k t_{ik} v_{kj})$
2. with respect to \mathbf{T} : $t_{ik} \leftarrow \frac{\sum_j \hat{a}_{ij} s_{ijk}}{\sum_j \hat{a}_{ij} v_{kj}}$
3. with respect to \mathbf{S} : $s_{ijk} \leftarrow t_{ik} v_{kj} + r_{ijk} (a_{ij} - \sum_k t_{ik} v_{kj})$
4. with respect to \mathbf{V} : $v_{kj} \leftarrow \frac{\sum_i \hat{a}_{ij} s_{ijk}}{\sum_i \hat{a}_{ij} t_{ik}}$.

These updates are derived from the partial derivative of \mathcal{L}^+ with respect to the corresponding variables. If we specify the parameters r_{ijk} as

$$r_{ijk} = \frac{t_{ik} v_{kj}}{\hat{a}_{ij}}, \quad \hat{a}_{ij} = \sum_k t_{ik} v_{kj}, \quad (6)$$

then the update for the auxiliary variables is simplified as $s_{ijk} \leftarrow a_{ij} t_{ik} v_{kj} / \hat{a}_{ij}$. By substituting this in the \mathbf{T} and \mathbf{V} updates, we obtain the well-known multiplicative updates for NMF [1]:

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j a_{ij} v_{kj}}{\sum_j \hat{a}_{ij} v_{kj}}, \quad v_{kj} \leftarrow v_{kj} \frac{\sum_i a_{ij} t_{ik}}{\sum_i \hat{a}_{ij} t_{ik}}. \quad (7)$$

3. FORMULATIONS AND ALGORITHMS

3.1. Nonnegativity in Multichannel NMF

In a multichannel case, we have a complex-valued vector $\mathbf{x} = [x_1, \dots, x_M]^T \in \mathbb{C}^M$ for a time and frequency slot, where x_m is the m -th microphone observation. Given that, here we consider how we guarantee the nonnegativity utilized in the NMF scheme.

A trivial way of guaranteeing nonnegativity is to take the absolute value of each element as $[|x_1|, \dots, |x_M|]^T$. However, this ignores the phase information that reflects the directivity of sources from the microphone array. Another way is to select a reference channel (e.g., the 1st channel) and to force the corresponding element to be nonnegative as $\mathbf{x} \leftarrow \mathbf{x} \cdot \text{sign}(x_1^*)$, where $\text{sign}(x) = x/|x|$ and $*$ is the complex conjugate. Although this worked well in our previous study [7], all channels are not treated evenly and the results may be affected by the reference channel selection.

This paper proposes using the outer product

$$\mathbf{X} = \begin{bmatrix} |x_1| & \dots & |x_1 x_M|^{\frac{1}{2}} \text{sign}(x_1 x_M^*) \\ \vdots & \ddots & \vdots \\ |x_M x_1|^{\frac{1}{2}} \text{sign}(x_M x_1^*) & \dots & |x_M| \end{bmatrix} \quad (8)$$

$= \bar{\mathbf{x}} \bar{\mathbf{x}}^H$ of the amplitude square-rooted observation vector $\bar{\mathbf{x}} = [|x_1|^{\frac{1}{2}} \text{sign}(x_1), \dots, |x_M|^{\frac{1}{2}} \text{sign}(x_M)]^T$. Note that $[\mathbf{X}]_{mn}$ contains information regarding the phase difference between the m -th and n -th microphones. We then introduce an $M \times M$ Hermitian positive semidefinite matrix \mathbf{H}_k to model the spatial property of the k -th NMF basis, and describe the outer product as

$$p(\mathbf{X}|\theta) = \prod_{m=1}^M \prod_{n=1}^M \mathcal{N}_c([\mathbf{X}]_{mn} | \sum_{k=1}^K [\mathbf{H}_k]_{mn} t_k v_k, 1) \quad (9)$$

$$\propto \exp(-\|\mathbf{X} - \sum_k \mathbf{H}_k t_k v_k\|_F^2) \quad (10)$$

with $\theta = \{t_1, \dots, t_K, v_1, \dots, v_K, \mathbf{H}_1, \dots, \mathbf{H}_K\}$, where \mathcal{N}_c represents a complex Gaussian distribution, m and n are microphone indexes, and $\|\mathbf{B}\|_F^2 = \sum_{m=1}^M \sum_{n=1}^M |b_{mn}|^2$ is the squared Frobenius norm of a matrix \mathbf{B} . Formulation (10) allows us to naturally extend the Euclidean cost function of the standard NMF to multichannel cases, and to derive effective multiplicative updates with favorable convergence behaviors, as shown below.

The next two subsections present multichannel NMF formulations and their corresponding algorithms. We start with a simple case where Hermitian positive semidefinite matrices are assigned to each NMF basis. We then consider a practical case for source separation by sharing such Hermitian matrices within a source.

3.2. NMF basis-wise spatial property \mathbf{H}_{ik}

This subsection studies a multichannel NMF formulation in which spatial property \mathbf{H}_{ik} depends on frequency i and NMF basis k , and does not depend on time j .

Let \mathbf{x}_{ij} be M -channel observations at frequency i and time j , and let \mathbf{X}_{ij} be the corresponding outer product defined by (8). Following (10) and the NMF scheme, the likelihood of the total observations $\mathbf{X} = \{\mathbf{X}_{11}, \dots, \mathbf{X}_{1J}, \dots, \mathbf{X}_{IJ}\}$ is given as

$$p(\mathbf{X}|\mathbf{T}, \mathbf{V}, \mathbf{H}) \propto \prod_{i,j} \exp(-\|\mathbf{X}_{ij} - \sum_k \mathbf{H}_{ik} t_{ik} v_{kj}\|_F^2), \quad (11)$$

and the negative log-likelihood is given by

$$\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}) = \sum_{i,j} \|\mathbf{X}_{ij} - \sum_k \mathbf{H}_{ik} t_{ik} v_{kj}\|_F^2. \quad (12)$$

Here we introduce a constraint $\|\mathbf{H}_{ik}\|_F = 1$ to fix the scaling ambiguity between \mathbf{H}_{ik} and t_{ik} , and also to simplify the update rules to minimize \mathcal{L} as shown below.

Following the optimization framework explained in Sect. 2, we introduce an auxiliary function

$$\mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{S}) = \sum_{i,j,k} \frac{1}{r_{ijk}} \|\mathbf{S}_{ijk} - \mathbf{H}_{ik} t_{ik} v_{kj}\|_F^2 \quad (13)$$

where \mathbf{S}_{ijk} are auxiliary variables that satisfy $\sum_k \mathbf{S}_{ijk} = \mathbf{X}_{ij}$, and r_{ijk} are parameters that satisfy $\sum_k r_{ijk} = 1$ and $r_{ijk} > 0$. As shown in Appendix 6.1, the auxiliary function \mathcal{L}^+ satisfies two conditions likewise (4) and (5), and the equality $\mathcal{L}^+ = \mathcal{L}$ is satisfied when

$$\mathbf{S}_{ijk} = \mathbf{H}_{ik} t_{ik} v_{kj} + r_{ijk} \mathbf{E}_{ij}, \quad \mathbf{E}_{ij} = \mathbf{X}_{ij} - \sum_k \mathbf{H}_{ik} t_{ik} v_{kj}. \quad (14)$$

The minimization updates with respect to the main variables \mathbf{T} , \mathbf{V} and \mathbf{H} are derived from the partial derivatives of \mathcal{L}^+ (see Appendix 6.2), and given with the definition (6) of \hat{a}_{ij} by

$$t_{ik} \leftarrow \frac{\sum_j \hat{a}_{ij} \text{tr}(\mathbf{S}_{ijk} \mathbf{H}_{ik})}{\sum_j \hat{a}_{ij} v_{kj}}, \quad (15)$$

$$v_{kj} \leftarrow \frac{\sum_i \hat{a}_{ij} \text{tr}(\mathbf{S}_{ijk} \mathbf{H}_{ik})}{\sum_i \hat{a}_{ij} t_{ik}}, \quad (16)$$

$$\mathbf{H}_{ik} \leftarrow \sum_j \hat{a}_{ij} \mathbf{S}_{ijk}, \quad \mathbf{H}_{ik} \leftarrow \mathbf{H}_{ik} / \|\mathbf{H}_{ik}\|, \quad (17)$$

where $\text{tr}(\mathbf{B}) = \sum_{m=1}^M b_{mm}$ is the matrix trace. It is not ensured that the update for \mathbf{H}_{ik} keeps it positive semidefinite, and it sometimes has nonnegative eigenvalues after the update. If so, we apply eigenvalue decomposition as $\mathbf{H}_{ik} = \mathbf{U} \mathbf{D} \mathbf{U}^H$, set all the negative elements of \mathbf{D} to zero, and update $\mathbf{H}_{ik} \leftarrow \mathbf{U} \mathbf{D} \mathbf{U}^H$ with the new \mathbf{D} .

The above \mathbf{T} and \mathbf{V} updates (15) and (16) are not in the form of multiplicative updates [1] and are not very efficient. If we utilize a relation $\hat{a}_{ij} \text{tr}(\mathbf{S}_{ijk} \mathbf{H}_{ik}) = t_{ik} v_{kj} [\hat{a}_{ij} + \text{tr}(\mathbf{E}_{ij} \mathbf{H}_{ik})]$, these updates become multiplicative as shown below. In summary, the negative log-likelihood (12) is iteratively minimized by repeating (18) and one of the three updates (19)-(21).

$$\mathbf{E}_{ij} = \mathbf{X}_{ij} - \sum_k \mathbf{H}_{ik} t_{ik} v_{kj}, \quad \hat{a}_{ij} = \sum_k t_{ik} v_{kj} \quad (18)$$

$$t_{ik} \leftarrow t_{ik} \cdot \left[1 + \frac{\sum_j v_{kj} \text{tr}(\mathbf{E}_{ij} \mathbf{H}_{ik})}{\sum_j v_{kj} \hat{a}_{ij}} \right] \quad (19)$$

$$v_{kj} \leftarrow v_{kj} \cdot \left[1 + \frac{\sum_i t_{ik} \text{tr}(\mathbf{E}_{ij} \mathbf{H}_{ik})}{\sum_i t_{ik} \hat{a}_{ij}} \right] \quad (20)$$

$$\mathbf{H}_{ik} \leftarrow \mathbf{H}_{ik} \sum_j \hat{a}_{ij} v_{kj} + \sum_j \mathbf{E}_{ij} v_{kj}, \quad \mathbf{H}_{ik} \leftarrow \mathbf{H}_{ik} / \|\mathbf{H}_{ik}\| \quad (21)$$

3.3. Source-wise spatial property \mathbf{H}_{io}

When multichannel NMF is applied to a source separation task, it is desired that NMF bases originating from the same source are clustered together. To realize such functionality, this subsection studies another formulation in which spatial property \mathbf{H}_{io} depends on frequency i and source o , where $o = 1, \dots, N$ with N being the number of sources.

The generative model is a modification of (11) by introducing variables z_{ko} , which relates NMF basis k and source o :

$$p(\mathbf{X}|\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z}) \propto \prod_{i,j} \exp(-\|\mathbf{X}_{ij} - \sum_{k,o} z_{ko} \mathbf{H}_{io} t_{ik} v_{kj}\|_F^2) \quad (22)$$

with $z_{ko} \geq 0$ and $\sum_{o=1}^N z_{ko} = 1$. It is interpreted that the k -th NMF basis belongs to the o -th source if z_{ko} is close to 1.

Along with discussions similar to those in the previous subsection, we obtain the following results.

The negative log likelihood to be minimized is given by

$$\mathcal{L}(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z}) = \sum_{i,j} \|\mathbf{X}_{ij} - \sum_{k,o} z_{ko} \mathbf{H}_{io} t_{ik} v_{kj}\|_F^2 \quad (23)$$

and its auxiliary function can be defined by

$$\mathcal{L}^+(\mathbf{T}, \mathbf{V}, \mathbf{H}, \mathbf{Z}, \mathbf{S}) = \sum_{i,j,k,o} \frac{1}{r_{ijk o}} \|\mathbf{S}_{ijk o} - z_{ko} \mathbf{H}_{io} t_{ik} v_{kj}\|_F^2 \quad (24)$$

where $\mathbf{S}_{ijk o}$ are auxiliary variables that satisfy $\sum_{k,o} \mathbf{S}_{ijk o} = \mathbf{X}_{ij}$, and $r_{ijk o}$ are parameters that satisfy $\sum_{k,o} r_{ijk o} = 1$ and $r_{ijk o} > 0$. The equality $\mathcal{L}^+ = \mathcal{L}$ is satisfied when

$$\mathbf{S}_{ijk o} = z_{ko} \mathbf{H}_{io} t_{ik} v_{kj} + r_{ijk o} \mathbf{E}_{ij}, \quad (25)$$

$$\mathbf{E}_{ij} = \mathbf{X}_{ij} - \sum_{k,o} z_{ko} \mathbf{H}_{io} t_{ik} v_{kj}. \quad (26)$$

If we specify the parameters $r_{ijk o}$ as

$$r_{ijk o} = \frac{z_{ko} t_{ik} v_{kj}}{\hat{a}_{ij}}, \quad \hat{a}_{ij} = \sum_{k,o} z_{ko} t_{ik} v_{kj} = \sum_k t_{ik} v_{kj}, \quad (27)$$

multiplicative updates for \mathbf{T} , \mathbf{V} and \mathbf{Z} are obtained as

$$t_{ik} \leftarrow t_{ik} \cdot \left[1 + \frac{\sum_j v_{kj} \sum_o z_{ko} \text{tr}(\mathbf{E}_{ij} \mathbf{H}_{io})}{\sum_j v_{kj} \hat{a}_{ij}} \right], \quad (28)$$

$$v_{kj} \leftarrow v_{kj} \cdot \left[1 + \frac{\sum_i t_{ik} \sum_o z_{ko} \text{tr}(\mathbf{E}_{ij} \mathbf{H}_{io})}{\sum_i t_{ik} \hat{a}_{ij}} \right], \quad (29)$$

$$z_{ko} \leftarrow z_{ko} \sum_{i,j} t_{ik} v_{kj} [\hat{a}_{ij} + \text{tr}(\mathbf{E}_{ij} \mathbf{H}_{io})]. \quad (30)$$

Regarding z_{ko} , a normalization procedure $z_{ko} \leftarrow z_{ko} / \sum_o z_{ko}$ should be conducted to satisfy the constraint $\sum_{o=1}^N z_{ko} = 1$. The update rule for \mathbf{H} is obtained as

$$\mathbf{H}_{io} \leftarrow \mathbf{H}_{io} \sum_k z_{ko} t_{ik} \sum_j \hat{a}_{ij} v_{kj} + \sum_k z_{ko} t_{ik} \sum_j \mathbf{E}_{ij} v_{kj}. \quad (31)$$

After this update, the procedure for ensuring positive semidefiniteness, described just after (17), and $\mathbf{H}_{io} \leftarrow \mathbf{H}_{io} / \|\mathbf{H}_{io}\|$ must be performed. In summary, the negative log-likelihood (23) is iteratively minimized by repeating (26) with $\hat{a}_{ij} = \sum_k t_{ik} v_{kj}$ and one of the four updates (28)-(31).

4. EXPERIMENTS

We considered a situation where two 6-second sounds of instruments coming from different directions (70° and 150°) were mixed and observed at two omnidirectional microphones 4 cm apart. The mixtures were made by convolving measured impulse responses, whose reverberation time was 130 ms, and the sound sources. The sampling rate was 16 kHz. Then, we applied STFT (with a 64 ms frame size and a 16 ms frame shift) to the mixtures and obtained $I \times J = 513 \times 372$ outer products (8).

Figure 1 shows the convergence behavior when minimizing the negative log-likelihoods of NMF basis-wise (12) and source-wise (23). In both cases, the number K of rank-1 basis matrices was set at 10, \mathbf{T} , \mathbf{V} , and \mathbf{Z} matrices were initialized randomly (but with the same random numbers for both runs), and \mathbf{H} elements were all set to $1/M$, with M being the number of microphones. We see that the negative log-likelihoods were minimized efficiently. The program was coded in Matlab and run on an Intel Core i7 965 (3.2-GHz) processor. The computation times for 500 iterations were 83 and 49 seconds, respectively.

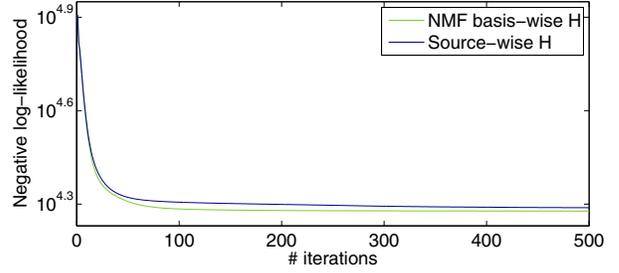


Figure 1: Convergence behavior

In the case (23) of modeling source-wise spatial properties \mathbf{H}_{io} , a source separation task was performed. Figure 2 shows spectrograms of two separated signals. They were constructed with learned matrices \mathbf{T} , \mathbf{V} , and \mathbf{Z} as

$$\mathbf{y}_{ijo} = \mathbf{x}_{ij} \frac{\sum_k z_{ko} t_{ik} v_{kj}}{\sum_k t_{ik} v_{kj}} \quad (32)$$

for each source o and every time-frequency j, i slot. Figure 3 shows the relationship z_{ko} between NMF bases k and sources o . From the figure, it is interpreted that the 2nd basis contributes to the blue output and the remaining bases contribute mainly to the red output. It is an advantage of the proposed method that the source-wise number of NMF bases does not need to be specified but only the total number needs to be specified. As a result of the source separation, the signal-to-distortion ratios [10] were improved from $(-7.37, 7.37)$ dB at the mixture to $(4.83, 12.21)$ dB.

Figure 4 shows the learned spatial properties, or more specifically the arguments of $[\mathbf{H}_{io}]_{12}$, which represent the phase difference between two microphones of each source component o at each frequency i . We observe that the blue and red dots correspond to the sources coming from 70° and 150° , respectively (the anechoic model lines of these two directions are drawn for reference). This way, the multichannel NMF also contributes to source localization.

5. CONCLUSION

We proposed new formulations and algorithms for multichannel NMF. We start with a simple case where each NMF basis k has its own spatial properties \mathbf{H}_{ik} . We then introduced a constraint that such properties \mathbf{H}_{io} should belong to each source o . The constraint forces the learned NMF bases to be clustered into the sources, and contributes to solving a source separation task with the multichannel NMF framework. Experimental results show the effectiveness of the proposed algorithms. Future work will include intensive verification of the algorithms with a variety of sound sources.

6. APPENDIX

6.1. Proof for the auxiliary function condition of (13)

Let us consider the minimization of \mathcal{L}^+ defined in (13) with respect to \mathbf{S} subject to the constraint $\sum_k \mathbf{S}_{ijk} = \mathbf{X}_{ij}$. Introducing Lagrange multipliers λ_{ij} , we have

$$\mathcal{F} = \mathcal{L}^+ + \sum_{i,j} \lambda_{ij} (\sum_k \mathbf{S}_{ijk}^* - \mathbf{X}_{ij}^*). \quad (33)$$

The partial derivatives of \mathcal{F} with respect to \mathbf{S} are given as

$$\frac{\partial \mathcal{F}}{\partial \mathbf{S}_{ijk}^*} = \frac{1}{r_{ijk}} (\mathbf{S}_{ijk} - \mathbf{H}_{ik} t_{ik} v_{kj}) + \lambda_{ij}. \quad (34)$$

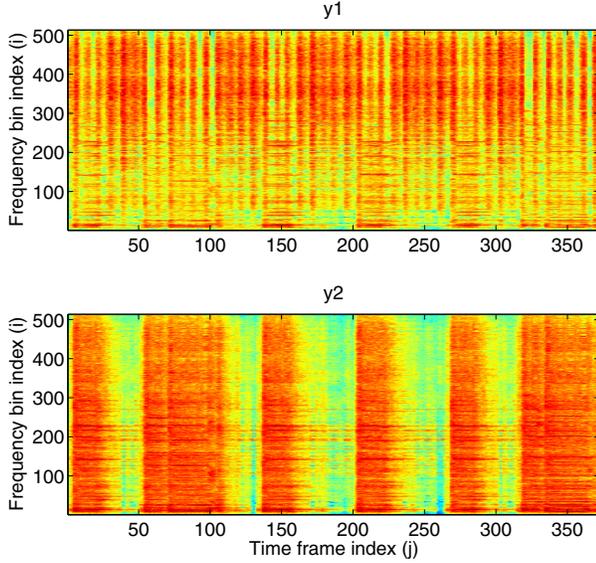


Figure 2: Two separated signals reconstructed by learned \mathbf{T} , \mathbf{V} matrices weighted with \mathbf{Z} (see Eq. 32 for details).

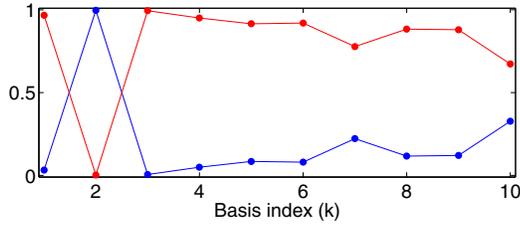


Figure 3: Learned values of z_{ko} .

Setting this to zero gives $S_{ijk} - H_{ik}t_{ik}v_{kj} + \lambda_{ij}r_{ijk} = 0$ and adding these for $k = 1, \dots, K$ gives $-\lambda_{ij} = X_{ij} - \sum_k H_{ik}t_{ik}v_{kj}$. Therefore the minimum \mathcal{L}^+ is obtained when

$$S_{ijk} = H_{ik}t_{ik}v_{kj} + r_{ijk}(X_{ij} - \sum_k H_{ik}t_{ik}v_{kj}), \quad (35)$$

and the minimum value is equal to \mathcal{L} defined in (12).

6.2. Derivation of the update rules

Utilizing $\|\mathbf{B}\|_F^2 = \text{tr}(\mathbf{B}^H \mathbf{B})$, the auxiliary function \mathcal{L}^+ defined in (13) can be simplified as

$$\mathcal{L}^+ = \sum_{i,j,k} \frac{1}{r_{ijk}} \left[\|\mathbf{s}_{ijk}\|_F^2 + t_{ik}^2 v_{kj}^2 - 2t_{ik}v_{kj} \text{tr}(\mathbf{S}_{ijk} \mathbf{H}_{ik}) \right]$$

where the constraint $\|\mathbf{H}_{ik}\|_F = 1$ and Hermitian identities $\mathbf{H}_{ik} = \mathbf{H}_{ik}^H$ and $\mathbf{S}_{ijk} = \mathbf{S}_{ijk}^H$ are utilized. Then, the partial derivatives with respect to \mathbf{T} and \mathbf{V} are

$$\frac{\partial \mathcal{L}^+}{\partial t_{ik}} = \sum_j \frac{2}{r_{ijk}} \left[t_{ik} v_{kj}^2 - v_{kj} \text{tr}(\mathbf{S}_{ijk} \mathbf{H}_{ik}) \right], \quad (36)$$

$$\frac{\partial \mathcal{L}^+}{\partial v_{kj}} = \sum_i \frac{2}{r_{ijk}} \left[t_{ik}^2 v_{kj} - t_{ik} \text{tr}(\mathbf{S}_{ijk} \mathbf{H}_{ik}) \right]. \quad (37)$$

Setting these derivatives to zero and substituting r_{ijk} with (6) yields the updates (15) and (16). The partial derivatives w.r.t. \mathbf{H} are

$$\frac{\partial \mathcal{L}^+}{\partial \mathbf{H}_{ik}^*} = \sum_j \frac{1}{r_{ijk}} (\mathbf{S}_{ijk} - \mathbf{H}_{ik} t_{ik} v_{kj}) (-t_{ik} v_{kj}). \quad (38)$$

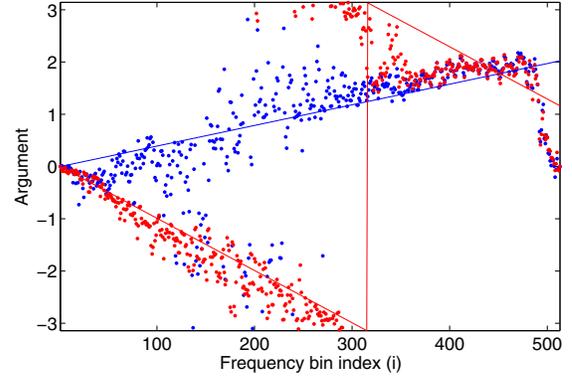


Figure 4: Arguments of learned $[\mathbf{H}_{io}]_{12}$.

Setting the derivative to zero and substituting r_{ijk} with (6), together with the constraint $\|\mathbf{H}_{ik}\|_F = 1$, yields the update (17).

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] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. IEEE Workshop Applcat. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2003, pp. 177–180.
- [3] 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.
- [4] R. M. Parry and I. A. Essa, "Estimating the spatial position of spectral components in audio," in *Proc. ICA 2006*. Springer, Mar. 2006, pp. 666–673.
- [5] 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.
- [6] S. Arberet, A. Ozerov, N. Duong, E. Vincent, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Nonnegative matrix factorization and spatial covariance model for under-determined reverberant audio source separation," in *Proc. ISSPA 2010*, May 2010, pp. 1–4.
- [7] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Formulations and algorithms for multichannel complex NMF," in *Proc. ICASSP 2011*, May 2011, pp. 229–232.
- [8] 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.
- [9] J. de Leeuw, "Block-relaxation methods in statistics," in *Information Systems and Data Analysis*, H. H. Bock, W. Lenski, and M. M. Richter, Eds. Springer Verlag, 1994, pp. 308–324.
- [10] E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. Rosca, "First stereo audio source separation evaluation campaign: Data, algorithms and results," in *Proc. ICA 2007*, 2007, pp. 552–559. [Online]. Available: <http://www.irisa.fr/metiss/SASSECO7/>