QUANTIFYING DISCRIMINABILITY BETWEEN NMF BASES

Eisuke Konno, Daisuke Saito, Nobuaki Minematsu

Graduate School of Engineering, The University of Tokyo

ABSTRACT

Discriminative nonnegative matrix factorization (DNMF) has been investigated as a promising basis-learning method for monaural source separation. To the best of our knowledge, however, no good and sound discussion has been made on quantitative definition of discriminability and it is difficult to evaluate how discriminative DNMF is actually. This paper introduces a quantitative measure to calculate how discriminative two NMF bases are. From the viewpoint of our measure, we compare three basis-learning methods of plain NMF, DNMF, and minimum-volume (min-vol) NMF. Experimental results of monaural speech separation reveal that min-vol NMF actually learns as discriminative bases as DNMF and achieves the best separation performance. This is probably because min-vol NMF can learn the most compact basis possible that can cover training data.

***Index Terms*—** Monaural source separation, discriminative NMF, minimum-volume NMF, discriminability

1. INTRODUCTION

Nonnegative matrix factorization (NMF) [1, 2] is a popular approach for monaural source separation. It decomposes the nonnegative amplitude spectrogram of a source signal into a nonnegative basis matrix representing a fixed number of spectral patterns of the source, and an nonnegative activation matrix representing their temporal variations in amplitude. Supervised NMF [3] enables separation utilizing pretrained bases of target sources. Given the amplitude spectrogram of a mixture signal X, with the pretrained bases $\hat{W}_1, \dots, \hat{W}_N$ of N sources obtained from the amplitude spectrograms of clean signals of the sources $V_1, \dots, V_N$, we optimize only their corresponding activations $H_1, \dots, H_N$ to make $\sum_n \hat{W}_n H_n$ approximate X:

$$\hat{H}_1, \dots, \hat{H}_N = \underset{H_1, \dots, H_N}{\arg\min}\, D_\beta\big(X \mid \textstyle\sum_n \hat{W}_n H_n\big) \tag{1}$$

where D_β is the β-divergence, which measures the approximation error [4]. Masking X with the bases and the activations, we reconstruct nth source component of X as

$$X \odot \big(\hat{W}_n \hat{H}_n\big) \oslash \big(\textstyle\sum_{n'} \hat{W}_{n'} \hat{H}_{n'}\big), \tag{2}$$

where $\odot$ and $\oslash$ denote element-wise multiplication and division, respectively.

The performance of supervised NMF (1) mainly depends on a choice of basis-learning methods. The simplest way to train the basis of source n is to use only its clean signal V_n:

$$\min_{W_n, H_n} D_\beta(V_n \mid W_n H_n). \tag{3}$$

We call this method *plain NMF*. Unfortunately, its separation performance is easily degraded by spectral overlap between sources. To solve this problem, *discriminative NMF (DNMF)* [5, 6] has been proposed, which additionally uses the amplitude spectrogram of a mixture signal $\mathcal{X}$ for training:

$$\min_{W_1, \dots, W_N, H_1, \dots, H_N} \sum_n D_\beta(V_n \mid W_n H_n) \tag{4a}$$

$$\text{s.t. } H_1, \dots, H_N = \underset{H'_1, \dots, H'_N}{\arg\min}\, D_\beta\big(\mathcal{X} \mid \textstyle\sum_n W_n H'_n\big), \tag{4b}$$

where $\mathcal{X}$ is obtained artificially by adding the clean signals in the time domain. The constraint (4b) can be regarded as a penalization of inter-bases spectral overlap, because such overlap may induce wrong activations through optimizing the objective function in (4b), which consequently increase the value of the other objective function in (4a). DNMF learns discriminative bases in the sense that they do not represent each other.

However, previous research on DNMF focused mainly on how to optimize the difficult bilevel optimization problem (4) [5, 6, 7, 8, 9, 10]. There has been neither solid enough discussion on the mechanism of the discriminative learning nor a quantitative definition or mathematical formulation of discriminability between NMF bases. With a good and sound measure of discriminability, we can compare various basis-learning methods, not limited to DNMF, from a new viewpoint different from existing ones such as sparsity [11, 12, 13], leading to further refinement of not only supervised NMF but also many widely used separation algorithms stemming from NMF [14, 15, 16, 17].

This paper proposes a quantitative measure of discriminability between two NMF bases by using their geometric overlap, and compare the basis-learning methods above in terms of this measure through experiments on monaural speech separation. In this comparison, we also test another method called *minimum-volume (min-vol) NMF*, described in section 2, because it also seems to learn discriminative bases in a different manner. In fact, the experimental results show that min-vol NMF learns more discriminative bases than DNMF[1].

2. BACKGROUND

Prior to min-vol NMF, let us begin with a geometric interpretation of NMF, on which our definition of discriminability is founded. Suppose that an exact NMF of the amplitude spectrogram of a source signal $V = [\boldsymbol{v}_1 \ \dots \ \boldsymbol{v}_T] \in \mathbb{R}_{\geq 0}^{F \times T}$ has yielded a certain decomposition $V = WH$ where $W \in \mathbb{R}_{\geq 0}^{F \times K}$ and $H \in \mathbb{R}_{\geq 0}^{K \times T}$. The basis W generates a cone (or more precisely, a conic hull) that contains the column vectors of V:

$$\boldsymbol{v}_t \in \mathrm{cone}\,(W) \overset{\text{def}}{=} \big\{W\boldsymbol{y} \mid \boldsymbol{y} \in \mathbb{R}_{\geq 0}^K\big\} \quad (t = 1, \dots, T). \tag{5}$$

This work was conducted as a part of the first author's master study.

[1] This paper is an extended version of our previous technical report [18]. We newly reports experimental results for a non-oracle case.

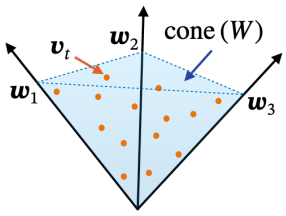
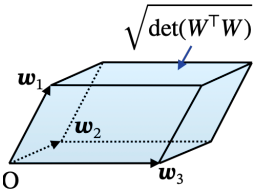
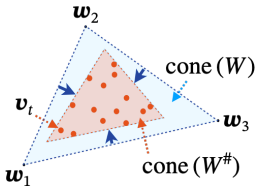


Fig. 1. A geometric interpretation of NMF [19]. cone (W), which is generated by the column vectors $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, and $\boldsymbol{w}_3$ of W, contains all the data points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_t, \ldots, \boldsymbol{v}_T$.



(a) The parallelotope defined by the column vectors of a basis W. Its volume is $\sqrt{\det(W^\top W)}$.



(b) Min-vol NMF seeks the basis W with the smallest volume that can cover the data points [19].

Fig. 2. The mechanism of min-vol NMF.

In other words, NMF is an algorithm that seeks a basis such that the cone it defines can cover all the data points $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_T$ (see Figure 1).

Unfortunately, there generally exist many such solutions; e.g., a set of the standard unit vectors of $\mathbb{R}^F$ generates a cone spreading over the entire nonnegative orthant and can represent any data points, though such a basis may cover extra regions. This makes it difficult to identify the true basis $W^\#$, which generated V. (Here we suppose that source signals certainly follow a generative model of NMF and that they have the true bases.) One simple fact that $WH = \left(WA^{-1}\right)(AH)$ for any nonsingular matrix A shows that calculation of NMF basis and activation always has an ambiguity problem. This intrinsic indeterminacy is inevitable for all NMF variants.

Unlike most of the other NMF variants, min-vol NMF has been proved to provide a solution to the ambiguity problem, to a certain extent under mild conditions [20, Theorem 1].

Theorem 1. *Let $V \in \mathbb{R}_{\geq 0}^{F\times T}$, $W^\# \in \mathbb{R}_{\geq 0}^{F\times K}$, and $H^\# \in \mathbb{R}_{\geq 0}^{K\times T}$. Assume that $V = W^\# H^\#$ with rank $(V) = K$ and $H^\#$ satisfies the sufficiently scattered condition (SSC). Then the optimal solution of the following optimization problem recovers $W^\#$ and $H^\#$ up to permutation and scaling:*

$$\min_{W\in\mathbb{R}^{F\times K}, H\in\mathbb{R}^{K\times T}} \det\left(W^\top W\right) \tag{6a}$$

$$\text{s.t. } V = WH,\ \mathbf{1}^\top W = \mathbf{1}^\top,\ H \geq 0, \tag{6b}$$

where $\mathbf{1}$ denotes a vector of an appropriate size whose entries are all one and $\geq$ denotes element-wise inequalities.

SSC informally means that the column vectors of $H^\#$ is scattered enough in the nonnegative orthant, generating the data points (i.e., the column vectors of V) from which we can estimate the shape of cone $\left(W^\#\right)$. Under SSC and the column-stochastic constraint $\mathbf{1}^\top W = \mathbf{1}^\top$, which are not so restrictive in many problem settings, the estimation of $W^\#$ is realized through minimizing the Gram determinant $\det\left(W^\top W\right)$, which is the squared volume of the parallelotope defined by the column vectors of W (see Figure 2). See [19, 20] for details including the exact definition of SSC.

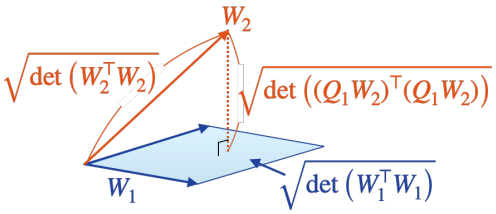


Fig. 3. An illustration of (9) in a three dimensional case. In this case, $\det\left(W_2^\top W_2\right)$ becomes the squared length of W_2 (the solid orange line), whereas $\det\left((Q_1W_2)^\top(Q_1W_2)\right)$ becomes the squared length of the perpendicular line from W_2 to span (W_1) (the dotted orange line).

Let us here return to the subject of discriminability. If the discriminability of DNMF (4) is attributed to its direct penalizing inter-bases spectral overlap through the constraint (4b), then it seems that min-vol NMF also learns discriminative bases in a different and indirect manner. This is because min-vol NMF seeks the most *compact* basis possible so that it may not cover any extra regions, reducing indirectly the spectral overlap concerned. Thus, we also test min-vol NMF in our experiments in section 5.

To train the basis of source n with min-vol NMF, let us rewrite (6) into an approximate form:

$$\min_{W_n, H_n} D_\beta(V_n \mid W_nH_n) + \frac{\lambda}{2}\log\det\left(W_n^\top W_n + \epsilon I\right) \tag{7a}$$

$$\text{s.t. } \mathbf{1}^\top W_n = \mathbf{1}^\top, \tag{7b}$$

where $\lambda > 0$ is a penalty parameter and $\epsilon > 0$. Here we have added the diagonal matrix ϵI and taken logarithm to stabilize computation of the determinant. These modifications allow us to use min-vol NMF even when we use a too large number of the column vectors of W_n, which makes W_n rank deficient [20, 21].

3. GEOMETRIC OVERLAP AND DISCRIMINABILITY

We have shown in section 2 that an NMF basis geometrically generates a cone. For two bases W_1 and W_2, it seems that we can evaluate discriminability between them by measuring the overlap of their cones, or of their parallelotopes. As the volume of the parallelotope defined by a basis is given by the Gram determinant of the basis, let us calculate the Gram determinant of their combined basis $\mathcal{W} \stackrel{\text{def}}{=} [W_1\ W_2]$.

Assuming for simplicity that both W_1 and W_2 are column full rank, we can express the Gram determinant in question as

$$\det\left(\mathcal{W}^\top\mathcal{W}\right) = \det\left(W_1^\top W_1\right)\det\left((Q_1W_2)^\top(Q_1W_2)\right), \tag{8}$$

where $Q_1 \stackrel{\text{def}}{=} I - W_1\left(W_1^\top W_1\right)^{-1}W_1^\top$ is the orthogonal projection onto span $(W_1)^\perp$, which is the orthogonal complement of the span of the column vectors of W_1. Dividing (8) with the Gram determinants of W_1 and W_2, we obtain

$$\frac{\det\left(\mathcal{W}^\top\mathcal{W}\right)}{\det\left(W_1^\top W_1\right)\det\left(W_2^\top W_2\right)} = \frac{\det\left((Q_1W_2)^\top(Q_1W_2)\right)}{\det\left(W_2^\top W_2\right)}. \tag{9}$$

The denominator in the right-hand side of (9) represents the squared volume of the parallelotope defined by the column vectors of W_2, whereas the corresponding numerator represents the squared volume

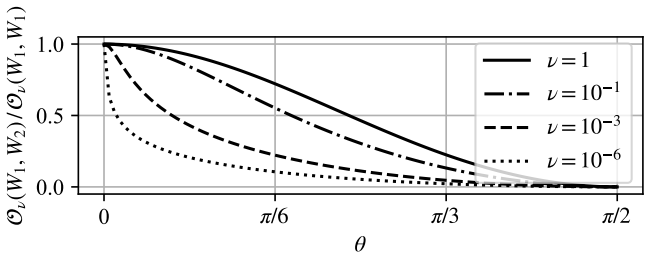


Fig. 4. A calculation example of our proposed discriminability measure $\mathcal{O}_\nu(W_1,W_2)$ for various values of ν, where $W_1=[1,0]^\top$ and $W_2=[\cos\theta,\sin\theta]^\top$ ($\theta\in[0,\pi/2]$). The measure is divided by $\mathcal{O}_\nu(W_1,W_1)$ to be normalized for easy comparison.

of the parallelotope defined by these column vectors projected onto $\mathrm{span}(W_1)^\perp$ (see Figure 3); thus, (9) takes 1 as its maximum when $\mathrm{span}(W_1)$ and $\mathrm{span}(W_2)$ are orthogonal and takes 0 as its minimum when they include the same subspace. The same conclusion holds naturally if we swap the roles of W_1 and W_2.

From the discussion above, this paper proposes the following measure of discriminability between two NMF bases:

$$\mathcal{O}_\nu(W_1,W_2)\overset{\text{def}}{=}\frac{1}{2}\log\frac{\det(W_1^\top W_1+\nu I)\det(W_2^\top W_2+\nu I)}{\det(W^\top W+\nu I)},\quad(10)$$

where $\nu>0$. Here we have modified the determinants as in (7a) because W_1 and W_2 are not necessarily column full rank. We claim that W_1 and W_2 are more *discriminative* when the value of $\mathcal{O}_\nu(W_1,W_2)$ is smaller. Note that min-vol NMF (7) does not directly minimize (10) because it certainly decreases the numerator in the logarithm but does not necessarily increase the denominator. Figure 4 shows a calculation example of $\mathcal{O}_\nu$ for various values of ν, where $W_1=[1,0]^\top$ and $W_2=[\cos\theta,\sin\theta]^\top$ ($\theta\in[0,\pi/2]$) are simple 2×1 matrices. When $\theta=0$, the two bases are completely overlapped and $\mathcal{O}_\nu$ takes its maximum, and when $\theta=\pi/2$, they are orthogonal and $\mathcal{O}_\nu$ takes its minimum. It should be noted that $\mathcal{O}_\nu$ tends to take smaller values when ν is too small as shown in the figure; thus, we use $\nu=1$ in the experiments in section 5.

4. UPDATE RULES

This section briefly summarizes the update rules of supervised NMF, plain NMF, DNMF, and min-vol NMF. Bases of the last three methods will be compared experimentally in section 5 in terms of our discriminability measure proposed in section 3. For divergence, we use the β-divergence with $\beta=1$ (i.e., the generalized Kullback–Leibler (KL) divergence), and add a column-stochastic constraint $\mathbf{1}^\top W_n=\mathbf{1}^\top$ to plain NMF (3) and DNMF (4) so that all the methods yield bases in the same scale.

Supervised NMF. Applying majorization-minimization (MM) algorithm [22] to (1) yields the following update rule [23]:

$$H_n\leftarrow H_n\odot(\hat{W}_n^\top\tilde{X})\oslash(\hat{W}_n^\top J),\quad(11)$$

where $\tilde{X}\overset{\text{def}}{=}X\oslash(\sum_n\hat{W}_nH_n)$ and J denotes a matrix of an appropriate size whose entries are all one.

Plain NMF. A combination of MM algorithm and the method of Lagrange multiplier yields the following update rules of the column-stochastic plain NMF as done in [24]:

$$W_n\leftarrow W_n\odot(\tilde{V}_nH_n^\top)\oslash\{J[W_n\odot(\tilde{V}_nH_n^\top)]\},\quad(12a)$$
$$H_n\leftarrow H_n\odot(W_n^\top\tilde{V}_n)\oslash(W_n^\top J),\quad(12b)$$

where $\tilde{V}_n\overset{\text{def}}{=}V_n\oslash(W_nH_n)$.

DNMF. The update rules of DNMF (4) can be obtained using an augmented Lagrangian method [10]. Replacing the lower-level problem in (4b) with its first order optimality condition

$$0=C_n\overset{\text{def}}{=}\nabla_{H_n}D_{\mathrm{KL}}(\mathcal{X}\mid\textstyle\sum_nW_nH_n)\quad(n=1,\dots,N),\quad(13)$$

we reformulate the difficult bilevel problem (4) into the following single-level one:

$$\min_{W_1,\dots,W_N,H_1,\dots,H_N}\sum_nD_{\mathrm{KL}}(V_n\mid W_nH_n)+\frac{\sigma}{2}\sum_n\left\|C_n+\frac{M_n}{\sigma}\right\|_{\mathrm{F}}^2\quad(14)$$

where $\sigma>0$ is a penalty parameter, $M_1,\dots,M_N$ are Lagrange multipliers, and $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. Optimizing (14) with the column-stochastic constraint is so difficult that we rescale bases and activations after optimization. In [10], we derived the following update rules as done in [25]:

$$W_n\leftarrow W_n\odot\Big[(\tilde{V}_n+\sigma Y^-)H_n^\top+\sigma\Big(J\Gamma_n^{-\top}+\hat{\mathcal{X}}\Gamma_n^{+\top}\Big)\Big]\oslash\Big[(J+\sigma Y^+)H_n^\top+\sigma\Big(J\Gamma_n^{+\top}+\tilde{\mathcal{X}}\Gamma_n^{-\top}\Big)\Big],\quad(15a)$$
$$H_n\leftarrow H_n\odot\big[W_n^\top(\tilde{V}_n+\sigma Y^-)\big]\oslash\big[W_n^\top(J+\sigma Y^+)\big],\quad(15b)$$
$$M_n\leftarrow M_n+\sigma C_n,\quad(15c)$$

where $\hat{\mathcal{X}}\overset{\text{def}}{=}\sum_nW_nH_n$, $\tilde{\mathcal{X}}\overset{\text{def}}{=}\mathcal{X}\oslash\hat{\mathcal{X}}$, $\Gamma_n^+\overset{\text{def}}{=}W_n^\top J+\sigma^{-1}\max\{M_n,0\}$, $\Gamma_n^-\overset{\text{def}}{=}W_n^\top\tilde{\mathcal{X}}+\sigma^{-1}\max\{-M_n,0\}$, $Y^+\overset{\text{def}}{=}(\sum_nW_n\Gamma_n^+)\odot\tilde{\mathcal{X}}\oslash\hat{\mathcal{X}}$, and $Y^-\overset{\text{def}}{=}(\sum_nW_n\Gamma_n^-)\odot\tilde{\mathcal{X}}\oslash\hat{\mathcal{X}}$. The max operation is element-wise.

Min-vol NMF. A combination of MM algorithm and the method of Lagrange multiplier yields the following update rule [24]:

$$W_n\leftarrow W_n^*(\boldsymbol{\mu})\overset{\text{def}}{=}W_n\odot\left\{\sqrt{C(\boldsymbol{\mu})^2+E}-C(\boldsymbol{\mu})\right\}\oslash D,\quad(16)$$

where $Z\overset{\text{def}}{=}\big(W_n^\top W_n+\epsilon I\big)^{-1}$, $Z^+\overset{\text{def}}{=}\max\{Z,0\}$, $Z^-\overset{\text{def}}{=}\max\{-Z,0\}$, $C(\boldsymbol{\mu})\overset{\text{def}}{=}JH_n^\top-2\lambda W_nZ^-+\mathbf{1}\boldsymbol{\mu}^\top$, $D\overset{\text{def}}{=}2\lambda W_n(Z^++Z^-)$, and $E\overset{\text{def}}{=}2D\odot(\tilde{V}_nH_n^\top)$. The square root and the exponentiation are element-wise. The multiplier $\boldsymbol{\mu}$ must satisfy the column-stochastic constraint $0=\gamma(\boldsymbol{\mu})\overset{\text{def}}{=}\mathbf{1}^\top W_n^*(\boldsymbol{\mu})-\mathbf{1}^\top$. The solution $\boldsymbol{\mu}^*$ can be obtained using the Newton–Raphson method: Iterate the update rule $\boldsymbol{\mu}\leftarrow\boldsymbol{\mu}-\gamma\mathbf{1}\oslash\nabla\gamma$ until convergence (e.g., $|\gamma|<10^{-9}$). Note that $(\nabla\gamma)^\top=\mathbf{1}^\top\Big\{W_n\odot\Big[C(\boldsymbol{\mu})\oslash\sqrt{C(\boldsymbol{\mu})^2+E}-J\Big]\oslash D\Big\}$. The update rule of activations is the same as (12b).

5. EXPERIMENTS

Experiments on monaural speech separation were conducted to compare discriminability of different NMF bases among the three basis-learning methods of plain NMF, DNMF, and min-vol NMF. Their discriminabilities were calculated as (10) with $\nu=1$, and their separation performances were further calculated as source-to-distortion ratio improvement (SDRi) [26]. Section 5.3 reports an oracle case where the accurate estimates of source signals in a mixture are available; i.e., test data were created from training data. Here we want to discuss how well the bases of different methods can fit to data in ideal situations. Section 5.4 reports a general non-oracle case where test data were independent of training data.

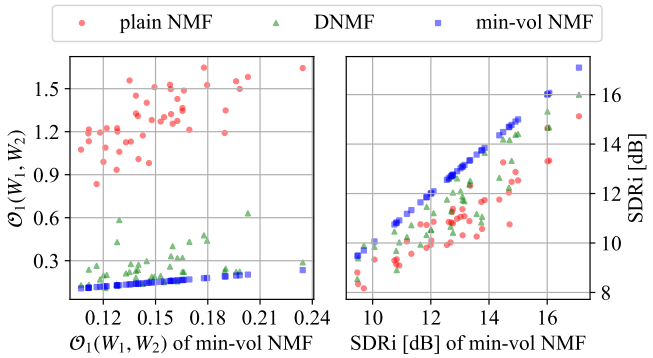


Fig. 5. The results of the monaural speech separation experiment for the oracle case. The points of the three basis-learning methods for the same target pair are on the same vertical line. (Left) Our proposed discriminability measure $\mathcal{O}_\nu(W_1, W_2)$ with $\nu = 1$. (Right) SDRi.

5.1. Data preparation

Speech samples used were 10 sets of the ATR 503 sentences, read aloud by six males and four females [27]. The sentence set is divided into sets A to J. In section 5.3, from each speaker, one utterance was selected from set A, resulting in 10 utterances corresponding to different sentences. Then, $\binom{10}{2} = 45$ utterance pairs were created and used for both training and testing.

In section 5.4, three pairs of speakers FKN-FTK, FKN-MHT, and MHT-MSH, where the initial letter represents sex, were chosen as targets of separation. For each speaker of each pair, 10 utterances were selected from set B as training data. The selected sentences were shared between the pair. For test data of the pair, set A was used to prepare 50 mixtures. In each mixture, two sentences of the two speakers were different as in the oracle case. The mixtures in both cases were prepared simply by adding source signals in the time domain with an SNR of 0 dB. The sampling frequency is 16 kHz, and amplitude spectrograms were obtained using short-time Fourier transform with a 64 ms Hann window and 16 ms frame-shift.

5.2. Optimization settings

The number of basis vectors was set to 200 for both sources. All the matrix elements of bases and activations were initialized from the uniform distribution on the interval $[1, 2]$, and then only the bases were rescaled to be column-stochastic. The Lagrange multipliers in (14) were initialized to zero matrices. The value of σ in (14) was determined so that the ratio of the divergence term to the penalty term at the initial values became $1 : 1$, whereas the value of λ in (7a) was determined so that the ratio became $1 : 10^{-3}$. ϵ in (7a) was set to 1. The number of iterations was set to 1000 for training and 200 for testing. The estimated amplitude spectrogram of a speaker was multiplied by the phase spectrogram of the corresponding mixture to obtain the speaker's time-domain signal.

5.3. Results for oracle case

Figure 5 shows the experimental results for the oracle case. Each point in each scatter plot corresponds to one of the 45 target pairs. The horizontal axes of these scatter plots are the values of $\mathcal{O}_1$ and SDRi of min-vol NMF, whereas the vertical axes are those of all the basis-learning methods. The points of min-vol NMF in each scatter plot thus lie on a straight line. As can be seen from the figure, DNMF achieves a significantly smaller value of $\mathcal{O}_1$ and a larger SDRi on average than plain NMF. These results lead us to the following hypothesis about the mechanism of DNMF: The constraint of DNMF (4b) favors bases W_1 and W_2 such that the cones defined by them have smaller overlap, and thus DNMF decreases our proposed measure, which increases when span (W_1) and span (W_2) include the same subspace. However, min-vol NMF, which does not directly take into account inter-bases spectral overlap in its training step, almost consistently achieves an even smaller value of $\mathcal{O}_1$ and an even larger SDRi than DNMF. This may be because the compact bases that min-vol NMF learns are actually discriminative enough and are better suited to source separation than DNMF.

Table 1. The results of the monaural speech separation experiment for the non-oracle case. The values of $\mathcal{O}_1$ and SDRi are the average over the test data.

	FKN-FTK		FKN-MHT		MHT-MSH	
	$\mathcal{O}_1$	SDRi [dB]	$\mathcal{O}_1$	SDRi [dB]	$\mathcal{O}_1$	SDRi [dB]
plain NMF	4.78	1.28	5.28	2.32	5.47	1.09
DNMF	**0.25**	2.87	0.27	4.77	**0.21**	3.16
min-vol NMF	0.32	**4.21**	**0.26**	**7.76**	0.26	**4.60**

5.4. Results for non-oracle case

Similar trends are confirmed in the non-oracle case as shown in Table 1. DNMF achieves a significantly smaller value of $\mathcal{O}_1$ and a larger SDRi than plain NMF, whereas min-vol NMF achieves a value of $\mathcal{O}_1$ as small as DNMF and an even larger SDRi than DNMF. In the case of the different gender pair FKN-MHT, DNMF and min-vol NMF achieve their biggest increases in SDRi over plain NMF, which are 2.45 dB and 5.44 dB, respectively. Interestingly, plain NMF, unlike DNMF and min-vol NMF, outputs significantly larger values of $\mathcal{O}_1$ in this case than those in the oracle case (see the left of Figure 5). This is probably because plain NMF learned bases like a set of the standard unit vectors to cover all of the training data of the increased size, whereas DNMF and min-vol NMF avoided spreading out basis vectors excessively.

6. CONCLUSIONS

This paper has proposed the measure $\mathcal{O}_\nu(W_1, W_2)$ that quantifies discriminability between two NMF bases W_1 and W_2. To the best of our knowledge, there has been no discussion on min-vol NMF in connection with DNMF or discriminability. The results of the experiments on monaural speech separation may indicate for the first time on a theoretically solid foundation that DNMF is certainly more discriminative than plain NMF, but min-vol NMF can actually learn bases that are compact and discriminative enough at the same time, and is better suited to source separation than DNMF. It suggests a potential use of min-vol NMF not only for source separation but also for other tasks that also need discriminative representations of source characteristics (e.g., NMF-based voice conversion [28]). Our future work will involve studying the relationship between the proposed discriminability measure and separation performance in more detail.

7. REFERENCES

[1] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[2] Nicolas Gillis, *Nonnegative Matrix Factorization*, SIAM, 2020.

[3] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in *Proceedings of ICA*, 2007, pp. 414–421.

[4] Cédric Févotte and Jérôme Idier, "Algorithms for nonnegative matrix factorization with the β-divergence," *Neural Computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[5] Felix Weninger, Jonathan Le Roux, John R. Hershey, and Shinji Watanabe, "Discriminative NMF and its application to single-channel source separation," in *Proceedings of Interspeech*, 2014, pp. 865–869.

[6] Pablo Sprechmann, Alex M. Bronstein, and Guillermo Sapiro, "Supervised non-euclidean sparse NMF via bilevel optimization with applications to speech enhancement," in *Proceedings of HSCMA*, 2014, pp. 11–15.

[7] Li Li, Hirokazu Kameoka, and Shoji Makino, "Discriminative non-negative matrix factorization with majorization-minimization," in *Proceedings of HSCMA*, 2017, pp. 141–145.

[8] Li Li, Hirokazu Kameoka, and Shoji Makino, "Majorization-minimization algorithm for discriminative non-negative matrix factorization," *IEEE Access*, vol. 8, pp. 227399–227408, 2020.

[9] Hiroaki Nakajima, Daichi Kitamura, Norihiro Takamune, Hiroshi Saruwatari, and Nobutaka Ono, "Bilevel optimization using stationary point of lower-level objective function for discriminative basis learning in nonnegative matrix factorization," *IEEE Signal Processing Letters*, vol. 26, no. 6, pp. 818–822, 2019.

[10] Eisuke Konno, Daisuke Saito, and Nobuaki Minematsu, "Discriminative nonnegative matrix factorization using an augmented Lagrangian method and its application to audio source separation," in *Proceedings of 2020 Autumn Meeting of The Acoustical Society of Japan*, 2020, pp. 143–146, (in Japanese).

[11] Patrik O Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[12] Niall Hurley and Scott Rickard, "Comparing measures of sparsity," *IEEE Transactions on Information Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.

[13] Jonathan Le Roux, Felix J. Weninger, and John R. Hershey, "Sparse NMF – half-baked or well done?," Tech. Rep. TR2015-023, Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA, 2015.

[14] Hiroshi Sawada, Hirokazu Kameoka, Shoko Araki, and Naonori Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 971–982, 2013.

[15] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto, "Infinite positive semidefinite tensor factorization for source separation of mixture signals," in *Proceedings of ICML*, 2013, vol. 28, pp. 576–584.

[16] Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka, and Hiroshi Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.

[17] Hiroshi Sawada, Nobutaka Ono, Hirokazu Kameoka, Daichi Kitamura, and Hiroshi Saruwatari, "A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF," *APSIPA Transactions on Signal and Information Processing*, vol. 8, pp. e12, 2019.

[18] Eisuke Konno, Daisuke Saito, and Nobuaki Minematsu, "A quantitative measure of discriminability between NMF dictionaries," *IEICE Technical Report*, vol. 120, no. 397, pp. 134–139, 2021, (in Japanese).

[19] Xiao Fu, Kejun Huang, Nicholas D. Sidiropoulos, and Wing-Kin Ma, "Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 59–80, 2019.

[20] Valentin Leplat, Nicolas Gillis, and Andersen M. S. Ang, "Blind audio source separation with minimum-volume beta-divergence NMF," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3400–3410, 2020.

[21] Valentin Leplat, Andersen M. S. Ang, and Nicolas Gillis, "Minimum-volume rank-deficient nonnegative matrix factorizations," in *Proceedings of ICASSP*, 2019, pp. 3402–3406.

[22] Ying Sun, Prabhu Babu, and Daniel P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2017.

[23] Daichi Kitamura, Hiroshi Saruwatari, Kosuke Yagi, Kiyohiro Shikano, Yu Takahashi, and Kazunobu Kondo, "Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. 97, no. 5, pp. 1113–1118, 2014.

[24] Valentin Leplat, Nicolas Gillis, and Jérôme Idier, "Multiplicative updates for NMF with β-divergences under disjoint equality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 42, no. 2, pp. 730–752, 2021.

[25] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[26] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

[27] Akira Kurematsu, Kazuya Takeda, Yoshinori Sagisaka, Shigeru Katagiri, Hisao Kuwabara, and Kiyohiro Shikano, "ATR Japanese speech database as a tool of speech recognition and synthesis," *Speech Communication*, vol. 9, no. 4, pp. 357–363, 1990.

[28] Ryoichi Takashima, Tetsuya Takiguchi, and Yasuo Ariki, "Exemplar-based voice conversion in noisy environment," in *Proceedings of SLT*, 2012, pp. 313–317.