2025 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

Nonnegative Matrix Factorization Using Dirichlet-Distribution-Based Regularization

Haru Ogawa*, Daichi Kitamura*§, and Shoma Ayano*
* National Institute of Technology, Kagawa College, Japan

***Abstract*—In this paper, we propose a novel nonnegative matrix factorization (NMF) using Dirichlet distribution prior. NMF is a low-rank approximation of a nonnegative matrix and is widely used in various tasks in a signal processing field. Many variants of regularized NMF have been proposed, and sparse and smooth regularizations particularly play an important role in the history of NMF. The proposed NMF utilizes Dirichlet-distribution-based regularization, which enables us to impose either sparseness or smoothness to decomposed matrices in a single unified framework. The effectiveness of the proposed method is demonstrated through numerical experiments, including a decomposition accuracy evaluation and a comparison with conventional sparsity-regularized NMF in the task of howling suppression. The results confirm that the proposed approach achieves improved performance in both decomposition quality and practical signal processing applications.**

I. Introduction

Nonnegative matrix factorization (NMF) [1] is an algorithm that approximates a nonnegative observed matrix $\boldsymbol{X}$ by the product of two nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$ under a low-rank constraint. NMF has widely been utilized in various signal processing fields including audio signal processing [2], [3], image and computer vision signal processing [4], [5], and biomedical signal processing [6], [7]. Moreover, many approaches have been proposed to incorporate prior models into the factorized matrices $\boldsymbol{W}$ and/or $\boldsymbol{H}$ in order to guide the optimization toward better solutions. Most of these methods are formulated as a regularized optimization problem. Representative examples include the imposition of sparsity [8], [9], smoothness [10], non-smoothness [11], orthogonality [12], [13], and minimum-volume constraints on the convex hull spanned by the basis vectors [14]. Such methods can be interpreted as maximum a posteriori (MAP) estimation, where specific prior distributions are assumed for the variables.

In this paper, we propose a novel NMF technique that employs a Dirichlet distribution as a prior, unifying both sparsity and smoothness (denseness) regularization within a single framework. By leveraging the properties of the Dirichlet distribution, we enforce each column or row vector of the target variable matrix to sum to unity, thereby imposing a norm constraint. This effectively resolves the scale indeterminacy problem, which often arises in regularized NMF. In addition, we can flexibly induce either sparsity or smoothness by adjusting the hyperparameters (concentration parameters) of the Dirichlet distribution prior.

§Corresponding author: kitamura-d@t.kagawa-nct.ac.jp
This work was supported by JSPS KAKENHI Grant Number 23K24908.

II. Conventional Methods

A. Formulation of NMF

NMF approximates a nonnegative observed matrix $\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{I \times J}$ by the product of a nonnegative basis matrix $\boldsymbol{W} \in \mathbb{R}_{\geq 0}^{I \times K}$ and a nonnegative coefficient matrix $\boldsymbol{H} \in \mathbb{R}_{\geq 0}^{K \times J}$, i.e., $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$, where I and J denote the numbers of rows and columns of $\boldsymbol{X}$, respectively, and K is the number of columns of $\boldsymbol{W}$ (i.e., the number of basis vectors). Typically, $K \ll I, J$ if the goal is low-rank approximation of $\boldsymbol{X}$.

The estimation of $\boldsymbol{W}$ and $\boldsymbol{H}$ is formulated as:

$$\operatorname*{Minimize}_{\boldsymbol{W},\boldsymbol{H}} \ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) \ \text{s.t.} \ w_{ik}, h_{kj} \geq 0 \ \forall i, j, k, \tag{1}$$

where w_{ik} and h_{kj} are the elements of $\boldsymbol{W}$ and $\boldsymbol{H}$, respectively, and $i = 1, 2, \cdots, I$, $j = 1, 2, \cdots, J$, $k = 1, 2, \cdots, K$ are their row/column indices. We denote the kth column of $\boldsymbol{W}$ by $\boldsymbol{w}_k$ (called a basis vector), and the kth row of $\boldsymbol{H}$ by $\boldsymbol{h}_k^{\mathrm{T}}$, where $\cdot^{\mathrm{T}}$ denotes the transpose.

The divergence function $\mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})$ measures the similarity between $\boldsymbol{X}$ and $\boldsymbol{W}\boldsymbol{H}$. In this work, we use the generalized Kullback–Leibler (KL) divergence defined by

$$\mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) = \sum_{i,j} \left(x_{ij} \log \frac{x_{ij}}{\sum_k w_{ik} h_{kj}} - x_{ij} + \sum_k w_{ik} h_{kj} \right), \tag{2}$$

where x_{ij} denotes each element of $\boldsymbol{X}$. A well-known approach to NMF with the generalized KL divergence is the multiplicative update rule [1] based on the so-called majorization-minimization (MM) algorithm [15].

B. Scale Indeterminacy in Regularized NMF

Suppose that the basis matrix $\boldsymbol{W}$ is assumed to follow a prior distribution $p(\boldsymbol{W})$. In that case, the MAP estimation of NMF can be written as:

$$\operatorname*{Minimize}_{\boldsymbol{W},\boldsymbol{H}} \ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) + \mathcal{R}(\boldsymbol{W}) \ \text{s.t.} \ w_{ik}, h_{kj} \geq 0 \ \forall i, j, k, \tag{3}$$

where $\mathcal{R}(\boldsymbol{W}) = -\log p(\boldsymbol{W})$ is a regularization term derived from the prior distribution $p(\boldsymbol{W})$.

If $\mathcal{R}(\boldsymbol{W})$ depends on the scale of $\boldsymbol{W}$, one can make $\mathcal{R}(\boldsymbol{W})$ arbitrarily small by scaling $\boldsymbol{W}$ by some constant factor and multiplying $\boldsymbol{H}$ by the reciprocal of that factor, leaving $\boldsymbol{W}\boldsymbol{H}$ unchanged. Consequently, $\mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})$ remains the same, rendering the regularization term meaningless [9], [13]. This problem occurs when the regularization is imposed on either

$\boldsymbol{W}$ or $\boldsymbol{H}$ alone, originating from the scale indeterminacy between $\boldsymbol{W}$ and $\boldsymbol{H}$. Hence, it is necessary to design the objective function such that the regularization term is scale-invariant (e.g., [13]), or to impose a norm constraint like $\|\boldsymbol{w}_k\| = 1\ \forall k$ (e.g., [9], [14], [16]).

III. Proposed Method

A. Motivation for Dirichlet Priors in NMF

The Dirichlet distribution is defined for vectors on a standard simplex (i.e., nonnegative vectors whose elements sum to 1) and can be interpreted as a probability density function (p.d.f.) on that simplex. Let $\boldsymbol{z} = [z_1, \cdots, z_I]^{\mathrm{T}} \in \mathbb{R}^I_{\geq 0}$ be a random vector following the Dirichlet distribution, such that $\sum_i z_i = 1$. The p.d.f. of the Dirichlet distribution is defined as

$$\boldsymbol{z} \sim p(\boldsymbol{z}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_i z_i^{\alpha_i - 1}, \tag{4}$$

where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_I]^{\mathrm{T}} \in \mathbb{R}^I_{>0}$ is a concentration parameter vector, and $B(\boldsymbol{\alpha}) = [\prod_i \Gamma(\alpha_i)]/\Gamma(\sum_i \alpha_i)$ is the multinomial beta function. Fig. 1 shows examples of the p.d.f. of the Dirichlet distribution when $I = 3$, illustrated on the standard two-dimensional simplex (triangle). The three vertices of the triangle correspond to $\boldsymbol{z} = [1, 0, 0]^{\mathrm{T}}$, $[0, 1, 0]^{\mathrm{T}}$, and $[0, 0, 1]^{\mathrm{T}}$, respectively. The vector $\boldsymbol{\alpha}$ controls how strongly the density concentrates near each vertex. If $\alpha_i < 1$, the density tends to concentrate near the ith vertex (i.e., generating " one-hot " or sparse vectors more frequently). Conversely, if $\alpha_i > 1$, the density tends to generate smoother (dense) vectors whose elements are more evenly distributed. In the special case $\alpha_i = 1\ \forall i$, the distribution is uniform on the simplex.

Using these properties of the Dirichlet distribution, we can design an NMF approach in which each basis vector $\boldsymbol{w}_k$ or coefficient vector $\boldsymbol{h}_k$ follows a Dirichlet prior, thereby inducing either sparsity or smoothness depending on how we set the concentration parameters $\boldsymbol{\alpha}$. In this paper, we call this method "Dirichlet NMF." The proposed method has the following advantages:

- The norm constraint implied by the Dirichlet prior inherently resolves the scale indeterminacy problem discussed in Section 2.2.
- By adjusting the concentration parameters $\boldsymbol{\alpha}$, one can continuously transition between sparse and smooth priors in a unified framework.

In this paper, we focus on Dirichlet priors for the basis vectors $\boldsymbol{w}_k$ as a concrete example. However, the same idea can be applied without loss of generality to $\boldsymbol{h}_k$.

B. Objective Function of NMF with Dirichlet Priors

Assume that each of the K basis vectors $\boldsymbol{w}_1, \cdots, \boldsymbol{w}_K$ follows an independent Dirichlet distribution $p(\boldsymbol{w}; \boldsymbol{\alpha}_k)$:

$$\boldsymbol{w}_k \sim p(\boldsymbol{w}; \boldsymbol{\alpha}_k)\ \forall k, \tag{5}$$

$$\boldsymbol{W} \sim p(\boldsymbol{W}) = \prod_k p(\boldsymbol{w}; \boldsymbol{\alpha}_k)|_{\boldsymbol{w}=\boldsymbol{w}_k}. \tag{6}$$

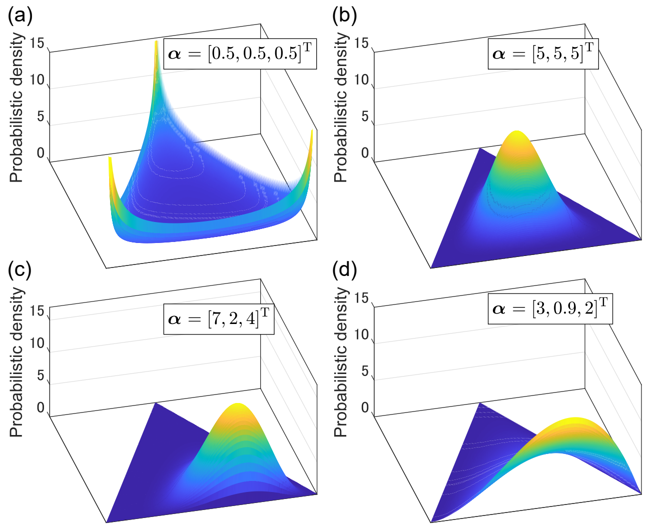

Fig. 1. Dirichlet distributions when $I = 3$, where the support triangle indicates the standard $(I-1)$-dimensional simplex.

By calculating the negative log-likelihood function, we can obtain the regularization term $\mathcal{R}(\boldsymbol{W})$ as

$$\begin{aligned} -\log p(\boldsymbol{W}) &= -\log \prod_k p(\boldsymbol{w}; \boldsymbol{\alpha}_k)|_{\boldsymbol{w}=\boldsymbol{w}_k} \\ &= \sum_k \left[\log B(\boldsymbol{\alpha}) - \sum_i (\alpha_{ik} - 1) \log w_{ik} \right] \\ &\overset{c}{=} \sum_{i,k} (\alpha_{ik} - 1) \log w_{ik}^{-1} \\ &\equiv \mathcal{R}(\boldsymbol{W}), \end{aligned} \tag{7}$$

where α_{ik} denotes the ith element of $\boldsymbol{\alpha}_k$ and $\overset{c}{=}$ denotes equality up to a constant. Thus, the MAP estimation with Dirichlet priors (Dirichlet NMF) is formulated as:

$$\begin{aligned} &\underset{\boldsymbol{W}, \boldsymbol{H}}{\text{Minimize}}\ \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) + \sum_{i,k} (\alpha_{ik} - 1) \log w_{ik}^{-1} \\ &\quad \text{s.t. } w_{ik}, h_{kj} \geq 0\ \forall i, j, k, \quad \sum_i w_{ik} = 1\ \forall k, \end{aligned} \tag{8}$$

where the new constraints impose both nonnegativity of the variables and a norm constraint on each $\boldsymbol{w}_k$. Throughout this paper, we let $\mathcal{J} = \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) + \mathcal{R}(\boldsymbol{W})$ denote the objective function of Dirichlet NMF.

C. Derivation of Optimization Algorithm

In the NMF literature, multiplicative update rules based on the MM algorithm are commonly employed [1]. This approach proceeds as follows: (a) for the current point of the variable, construct a majorization function satisfying certain requirements, (b) solve for the stationary point of this majorization function in closed form, and update the variable for the next iteration. This guarantees the monotonic non-increase of the objective function at each iteration. We also apply this method to the optimization problem in Dirichlet NMF. However, unlike standard NMF, we must handle both nonnegativity and the additional norm constraint on $\boldsymbol{w}_k$. To cope with this problem, we design a majorization function for the objective and then solve a constrained minimization subproblem using the method of Lagrange multipliers and the Karush–Kuhn–Tucker (KKT) conditions[1].

First, we apply Jensen's inequality to the generalized KL divergence:

$$\begin{aligned}
&\mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H})\\
&\quad\stackrel{c}{=}\sum_{i,j}\left(-x_{ij}\log\sum_k w_{ik}h_{kj}+\sum_k w_{ik}h_{kj}\right)\\
&\quad\leq\sum_{i,j}\left(-x_{ij}\sum_k \delta_{ijk}\log\frac{w_{ik}h_{kj}}{\delta_{ijk}}+\sum_k w_{ik}h_{kj}\right), \quad (9)
\end{aligned}$$

where $\delta_{ijk}>0$ is an auxiliary variable that satisfies $\sum_k \delta_{ijk}=1$. Using the inequality (9), we can design a majorization function $\mathcal{J}^+ \geq \mathcal{J}$ as follows:

$$\begin{aligned}
\mathcal{J}^+ \stackrel{c}{=}& \sum_{i,j}\left(-x_{ij}\sum_k \delta_{ijk}\log\frac{w_{ik}h_{kj}}{\delta_{ijk}}+\sum_k w_{ik}h_{kj}\right)\\
&+\sum_{i,k}(\alpha_{ik}-1)\log w_{ik}^{-1}. \quad (10)
\end{aligned}$$

The equality $\mathcal{J}^+=\mathcal{J}$ holds if and only if

$$\delta_{ijk}=\frac{w_{ik}h_{kj}}{\sum_{k'} w_{ik'}h_{k'j}}. \quad (11)$$

We consider the following *constrained* minimization problem of the majorization function:

$$\begin{aligned}
&\underset{\boldsymbol{W},\boldsymbol{H},\boldsymbol{\Delta}}{\text{Minimize}}\ \mathcal{J}^+\\
&\text{s.t. } w_{ik},h_{kj}\geq 0\ \forall i,j,k,\quad \sum_i w_{ik}=1\ \forall k, \quad (12)
\end{aligned}$$

where Δ denotes the set of $\delta_{ijk}\ \forall i,j,k$. The Lagrangian for this optimization is given by

$$\mathcal{L}=\mathcal{J}^+-\sum_k\lambda_k\left(\sum_i w_{ik}-1\right)-\sum_{i,k}\mu_{ik}w_{ik}, \quad (13)$$

where λ_k and μ_{ik} are the Lagrange multipliers for the equality (norm) and inequality (nonnegative) constraints, respectively. The KKT conditions are

$$\frac{\partial\mathcal{L}}{\partial w_{ik}}=0\ \forall i,k, \quad (14)$$

$$\frac{\partial\mathcal{L}}{\partial \lambda_k}=0\ \forall k, \quad (15)$$

$$\mu_{ik}\geq 0\ \forall i,k, \quad (16)$$

$$\frac{\partial\mathcal{L}}{\partial \mu_{ik}}\leq 0\ \forall i,k, \quad (17)$$

$$\mu_{ik}\frac{\partial\mathcal{L}}{\partial \mu_{ik}}=0\ \forall i,k. \quad (18)$$

[1]In [16], a similar constrained optimization for a majorization function is considered. However, the method in [16] treats more generalized formulation and utilizes Newton–Raphson method to calculate the Lagrangian multiplier.

From the derivation of $\mathcal{L}$ w.r.t. each variable, we obtain

$$\frac{\partial\mathcal{L}}{\partial w_{ik}}=\frac{\partial\mathcal{J}^+}{\partial w_{ik}}-\lambda_k-\mu_{ik}, \quad (19)$$

$$\frac{\partial\mathcal{L}}{\partial \lambda_k}=-\left(\sum_i w_{ik}-1\right), \quad (20)$$

$$\frac{\partial\mathcal{L}}{\partial \mu_{ik}}=-w_{ik}. \quad (21)$$

Substituting (21) into (17) and (18) shows that the minimizer must satisfy

$$w_{ik}\geq 0\ \forall i,k, \quad (22)$$

$$\mu_{ik}w_{ik}=0\ \forall i,k. \quad (23)$$

From (23), the minimizer of $\mathcal{L}$ satisfies either $\mu_{ik}=0$ or $w_{ik}=0$. Let $\hat{w}_{ik}$ be the unconstrained minimizer of $\mathcal{L}$ that satisfies (14). If $\hat{w}_{ik}$ satisfies the nonnegativity condition (22), then $\mu_{ik}=0$ and $w_{ik}=\hat{w}_{ik}$ satisfy all of the KKT conditions and coincide with the constrained minimizer of $\mathcal{L}$. In contrast, if $\hat{w}_{ik}$ does not satisfy (22), then $w_{ik}=0$ becomes the constrained minimizer. Thus, the minimizer of w_{ik} can be written as

$$w_{ik}=\begin{cases}\hat{w}_{ik} & (\hat{w}_{ik}\geq 0)\\ 0 & (\text{otherwise})\end{cases}. \quad (24)$$

Therefore, the optimal solution for $\boldsymbol{W}$ in the optimization problem (12) can be obtained by considering only the elements that satisfy $\hat{w}_{ik}\geq 0$. Hereafter, for each k, we define the the set of indices i that satisfy $\hat{w}_{ik}\geq 0$ as $\mathcal{I}_k$.

Next, we solve for $\hat{w}_{ik}\ \forall i\in\mathcal{I}_k$. Setting $\partial\mathcal{L}/\partial w_{ik}=0$, we obtain

$$\sum_j\left(-x_{ij}\frac{\delta_{ijk}}{\hat{w}_{ik}}+h_{kj}\right)-(\alpha_{ik}-1)\frac{1}{\hat{w}_{ik}}-\lambda_k-\mu_{ik}=0. \quad (25)$$

Noting that $\mu_{ik}=0$ for all $i\in\mathcal{I}_k$ due to the condition (23), $\hat{w}_{ik}$ can be obtained by rearranging (25) as follows:

$$\hat{w}_{ik}=\frac{\sum_j x_{ij}\delta_{ijk}+\alpha_{ik}-1}{\sum_j h_{kj}-\lambda_k}\ \forall i\in\mathcal{I}_k,\ \forall k. \quad (26)$$

To eliminate the Lagrange multiplier λ_k in (26), we consider the summation of the minimizer w_{ik} over i, which yields

$$\sum_i w_{ik}=\sum_{i\in\mathcal{I}_k}\frac{\sum_j x_{ij}\delta_{ijk}+\alpha_{ik}-1}{\sum_j h_{kj}-\lambda_k}\ \forall k. \quad (27)$$

By substituting (27) into the norm constraint condition $\sum_i w_{ik}=1\ \forall k$ derived from (15) and (20), we obtain the following equation:

$$\sum_j h_{kj}-\lambda_k=\sum_{i\in\mathcal{I}_k}\left(\sum_j x_{ij}\delta_{ijk}+\alpha_{ik}-1\right)\ \forall k. \quad (28)$$

Since the left-hand side of (28) coincides with the denominator of (26), the Lagrange multiplier λ_k can be eliminated by substituting (28) into (26), and we finally obtain the solution of $\hat{w}_{ik}$ as

$$\hat{w}_{ik} = \frac{\sum_j x_{ij}\delta_{ijk} + \alpha_{ik} - 1}{\sum_{i\in\mathcal{I}_k}\left(\sum_j x_{ij}\delta_{ijk} + \alpha_{ik} - 1\right)} \quad \forall i \in \mathcal{I}_k, \forall k. \quad (29)$$

The constrained minimizer of the majorization function $\mathcal{J}^+$ is given by (24) and (29). Furthermore, by substituting the equality condition of the auxiliary variable δ_{ijk}, (11), into (29), we can rewrite the minimizer (29) as follows:

$$\hat{w}_{ik} = \frac{w_{ik}\sum_j \frac{x_{ij}}{\sum_{k'} w_{ik'}h_{k'j}} h_{kj} + \alpha_{ik} - 1}{\sum_{i\in\mathcal{I}_k}\left(w_{ik}\sum_j \frac{x_{ij}}{\sum_{k'} w_{ik'}h_{k'j}} h_{kj} + \alpha_{ik} - 1\right)} \quad \forall i \in \mathcal{I}_k, \forall k. \quad (30)$$

From (24) and (30), it can be confirmed that the iterative update rule inherently satisfies the nonnegative and norm constraints on $\boldsymbol{w}_k$.

The optimization algorithm for Dirichlet NMF proceeds by:

1) initializing $\boldsymbol{W}$ and $\boldsymbol{H}$ with positive random values with column normalization for $\boldsymbol{W}$,
2) updating $\boldsymbol{W}$ using (24) and (30),
3) updating $\boldsymbol{H}$ by the well-known multiplicative update rule as in [1],

repeating until convergence[2]. Since the update rules for $\boldsymbol{W}$ and $\boldsymbol{H}$ are based on the MM algorithm, the monotonic non-increase of the objective function is theoretically guaranteed at each iteration.

IV. Numerical Experiment

A. Experimental Conditions

In this experiment, we numerically confirmed that the basis vectors in Dirichlet NMF are guided toward either sparse or smooth structures according to the specified concentration parameters. We prepared a 5×3 basis matrix and a 3×10 coefficient matrix, as illustrated in Fig. 2 (a). Two of the basis vectors were set to "one-hot" (sparse) vectors, and the remaining one was set to a smooth (uniform) vector. The elements of coefficient matrix were generated from a uniform distribution in the range $(0, 1)$. An observed 5×10 matrix $\boldsymbol{X}$ was defined as the product of these matrices.

We compared two algorithms: simple NMF without regularization (but with normalization of each column of $\boldsymbol{W}$ at every iteration) and the proposed Dirichlet NMF. In both algorithms, the number of bases was set to $K = 3$, which coincides with the rank of $\boldsymbol{X}$, and the total number of iterations was 100. In each iteration of simple NMF, each column of $\boldsymbol{W}$ was normalized, and the scaling was performed to $\boldsymbol{H}$ so that $\boldsymbol{WH}$ remains unchanged. For Dirichlet NMF, we set the parameters of the Dirichlet priors to $\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_2 = [0.5, 0.5, 0.5, 0.5, 0.5]^\mathrm{T}$ and $\boldsymbol{\alpha}_3 = [1.5, 1.5, 1.5, 1.5, 1.5]^\mathrm{T}$. These parameter settings induce the sparsity into $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ and the smoothness into $\boldsymbol{w}_3$ during the optimization.

[2] A MATLAB implementation of the proposed method is available at https://github.com/d-kitamura/dirichletNmf.

B. Results

The estimated $\boldsymbol{W}$ and $\boldsymbol{H}$ obtained by simple NMF and Dirichlet NMF are shown in Figs. 2 (b) and (c), respectively. Simple NMF fails to recover the oracle basis and coefficient matrices accurately. In contrast, Dirichlet NMF successfully reconstructs both $\boldsymbol{W}$ and $\boldsymbol{H}$ in a nearly perfect manner.

The convergence behavior of the objective function in Dirichlet NMF is illustrated in Fig. 3. The objective function value decreases monotonically. Furthermore, it converges within about 20 iterations in this example. Similar results were obtained in additional experiments with different random seeds for the coefficient and initial matrices.

V. Application to Howling Suppression

A. Principle of Howling Suppression Based on NMF

As a potential application of the proposed method, howling suppression can be considered, which has been studied using adaptive notch filters (e.g., [17], [18]). However, it is difficult for adaptive notch filters to track howling when it occurs at multiple frequencies simultaneously. Therefore, in this section, we conduct experiments on howling suppression using NMF.

Howling sounds appear as stationary signals consisting of a single frequency component and, therefore, exhibit a sparse structure along the frequency axis in a spectrogram. As shown in Fig. 4, the spectral patterns of howling sounds can be estimated by introducing sparsity regularization into a subset of basis vectors. Then, howling suppression can be achieved by reconstructing the spectrogram using only the components not associated with the howling sounds. However, the spectral patterns of howling sounds need to be estimated in the intended basis vectors in advance, and thus appropriate regularization to guide this separation plays a critical role.

We compared two approaches: NMF with L_1-norm-based sparsity regularization [8] (referred to as L_1-sparse NMF) and Dirichlet NMF. Dirichlet NMF is expected not only to apply sparsity regularization to specific basis vectors, but also to introduce smoothness regularization to the remaining ones, thereby encouraging the spectral patterns of howling sounds to be reliably estimated in the intended basis vectors.

We constructed a Wiener filter using the model spectrogram after howling suppression, denoted as $\widehat{\boldsymbol{X}} \in \mathbb{R}_{\geq 0}^{I\times J}$ in Fig. 4. Let $\boldsymbol{X}^{(\mathrm{c})} \in \mathbb{C}^{I\times J}$ and $\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{I\times J}$ denote the complex and amplitude spectrograms of the observed signal, respectively. After estimating a low-rank approximation $\boldsymbol{X} \approx \boldsymbol{WH}$, we define $\widehat{\boldsymbol{X}} \in \mathbb{R}_{\geq 0}^{I\times J}$ as the model spectrogram reconstructed using only the components excluding those corresponding to howling sounds. The howling suppression is then given by

$$\widehat{\boldsymbol{X}}^{(\mathrm{c})} = \frac{\widehat{\boldsymbol{X}}}{\boldsymbol{WH}} \odot \boldsymbol{X}^{(\mathrm{c})}, \quad (31)$$

where $\widehat{\boldsymbol{X}}^{(\mathrm{c})}$ denotes the complex spectrogram after howling suppression. The fraction and the operator $\odot$ represent element-wise division and multiplication, respectively.

(a) Basis mat. (oracle)

0.00e+00	1.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
1.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01

Coef. mat. (oracle)

4.17e−01	3.02e−01	1.86e−01	5.39e−01	2.04e−01	6.70e−01	1.40e−01	9.68e−01	8.76e−01	3.91e−02
7.20e−01	1.47e−01	3.46e−01	4.19e−01	8.78e−01	4.17e−01	1.98e−01	3.13e−01	8.95e−01	1.70e−01
1.14e−04	9.23e−02	3.97e−01	6.85e−01	2.74e−02	5.59e−01	8.01e−01	6.92e−01	8.50e−02	8.78e−01

(b) Basis mat. (simple NMF)

2.53e−02	7.87e−01	3.73e−01
2.10e−02	1.42e−14	1.80e−01
9.35e−01	2.20e−01	1.56e−02
2.10e−02	1.16e−15	1.80e−01
2.10e−02	7.39e−15	1.80e−01

Coef. mat. (simple NMF)

1.91e−01	3.02e−01	1.97e−01	6.26e−01	2.22e−16	7.34e−01	3.02e−01	1.12e+00	6.97e−01	2.14e−01
9.35e−01	1.69e−01	3.37e−01	3.64e−01	1.07e+00	3.98e−01	1.89e−02	2.42e−01	1.13e+00	2.20e−04
2.22e−16	6.60e−02	4.16e−01	6.85e−01	3.10e−02	5.32e−01	8.72e−01	6.32e−01	1.89e−07	9.33e−01

(c) Basis mat. (Dirichlet NMF)

0.00e+00	1.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
1.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01
0.00e+00	0.00e+00	2.00e−01

Coef. mat. (Dirichlet NMF)

4.17e−01	3.02e−01	1.86e−01	5.39e−01	2.04e−01	6.70e−01	1.40e−01	9.68e−01	8.76e−01	3.90e−02
7.20e−01	1.47e−01	3.46e−01	4.19e−01	8.78e−01	4.17e−01	1.98e−01	3.13e−01	8.95e−01	1.70e−01
1.14e−04	9.23e−02	3.97e−01	6.85e−01	2.74e−02	5.59e−01	8.01e−01	6.92e−01	8.50e−02	8.78e−01

Fig. 2. (a) Oracle basis and coefficient matrices that produce the observed matrix $\boldsymbol{X}$, (b) estimates of $\boldsymbol{W}$ and $\boldsymbol{H}$ by simple NMF without regularization, and (c) estimates by the proposed Dirichlet NMF.

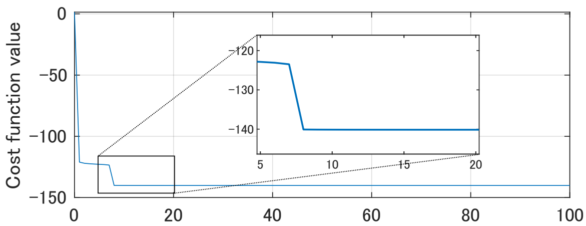

Fig. 3. Convergence behavior of Dirichlet NMF.

B. Conditions

We used three speech signals—`BASIC5000_0024.wav`, `BASIC5000_0035.wav`, and `BASIC5000_0465.wav`—randomly selected from the *JVS corpus* [19]. To simulate multiple howling sounds, we added two sinusoidal waves with frequencies of (1 kHz, 2 kHz), (2 kHz, 4 kHz), and (4 kHz, 8 kHz) to each of the three speech signals, respectively. The first howling component was active during the 2–4 s segment, and the second during the 3–5 s segment. The maximum amplitude of each howling signal was set to 20% of the maximum amplitude of the corresponding clean speech signal. To simulate smooth onsets and offsets, the added sinusoids were multiplied by a Hann window.

In both L_1-sparse NMF and Dirichlet NMF, the first and second basis vectors were regularized to be sparse so that the spectral patterns of the howling sounds would be captured by $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$. For L_1-sparse NMF, the number of basis vectors K and the regularization weight β for the L_1 norm were varied within the ranges $[5, J]$ for K and $[10^{-4}, 50]$ for β. We present results for three representative parameter settings, including the one that achieved the best performance. In the case of Dirichlet NMF, two hyperparameters were introduced: one for the sparse basis vectors, defined as $\boldsymbol{\alpha}_{\mathrm{howl}} = [\alpha_1, \alpha_2]^{\mathrm{T}}$, and another for the remaining basis vectors, defined as $\boldsymbol{\alpha}_{\mathrm{other}} = [\alpha_3, \cdots, \alpha_K]^{\mathrm{T}}$. We tuned K and $\boldsymbol{\alpha}_{\mathrm{howl}}$ under the constraints $\alpha_1 = \alpha_2$ and $\alpha_3 = \cdots = \alpha_K = 1.3$, with the following search ranges: $[5, J]$ for K and $[-10, 1]$ for the elements of $\boldsymbol{\alpha}_{\mathrm{howl}}$. As with L_1-sparse NMF, we report the results for three representative parameter settings in Dirichlet NMF.

Note that, in principle, the concentration parameter $\boldsymbol{\alpha}$ must be strictly positive. However, we empirically found that allowing negative values in $\boldsymbol{\alpha}_{\mathrm{howl}}$ further enhanced the sparsity effect without causing optimization issues. In fact, the update algorithm (30) successfully minimized the cost function while satisfying the constraints for the variables. For this reason, we included negative values in the tuning range of $\boldsymbol{\alpha}_{\mathrm{howl}}$.

As the evaluation metric, we used the source-to-distortion ratio (SDR) [20]. In this experiment, the SDR reflects the overall quality of howling suppression, taking into account both the degree of howling reduction and the absence of artificial distortions in the speech signal.

C. Results

Fig. 5 shows violin plots of the SDR values averaged over the three observed signals. For each method and each observed

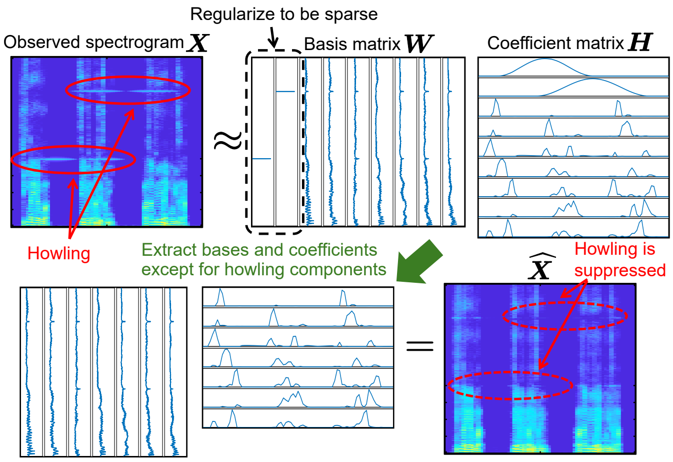

Fig. 4. Howling suppression framework based on NMF with sparsity regularization.

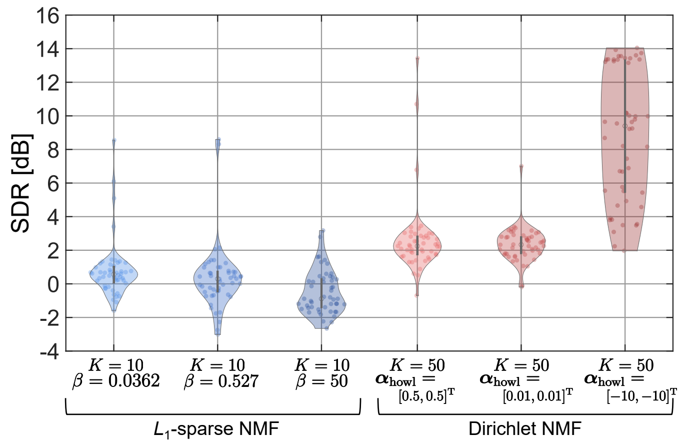

Fig. 5. Violin plots of SDR values for howling suppression. The white circle indicates median value, the gray vertical line shows the 25–27 percentile range, and the violin curve is an estimated distribution.

signal, 50 trials were conducted with different random initializations of $\boldsymbol{W}$ and $\boldsymbol{H}$. Under the optimal conditions for each method, the average SDRs were 0.80 dB for L_1-sparse NMF and 8.83 dB for Dirichlet NMF, confirming the effectiveness of the proposed method in suppressing howling. In L_1-sparse NMF, the howling spectra were not adequately captured by the sparsity-induced basis vectors, resulting in degraded SDR performance. Intriguingly, the best performance in Dirichlet NMF was obtained with a negative hyperparameter setting for $\boldsymbol{\alpha}_{\mathrm{howl}}$. Although such a setting no longer corresponds to the MAP estimation based on the Dirichlet distribution, the enhanced sparsity induced by this generalized Dirichlet NMF framework significantly improves the performance of NMF-based howling suppression.

VI. Conclusion

We presented a new NMF method incorporating a Dirichlet distribution prior. The proposed method allows for inducing either sparse or smooth structures in the factorized matrices by adjusting the concentration parameters, while avoiding the scale indeterminacy issue commonly encountered in regularized NMF. We also derived update rules based on the MM algorithm. Both numerical experiments and a practical application to howling suppression demonstrated the effectiveness of the proposed method.

References

[1] D. Lee and H. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. Adv. Neural Info. Process. Syst.*, 2000.

[2] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.

[3] C. Févotte, E. Vincent, and A. Ozerov, “Single-channel audio source separation with NMF: Divergences, constraints and algorithms,” in *Audio Source Separation*, S. Makino, Ed., pp. 1–24. Springer, Cham, 2018.

[4] Y. Salehani and S. Gazor, “Smooth and sparse regularization for NMF hyperspectral unmixing,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 8, pp. 3677–3692, 2017.

[5] T. Yokota, K. Kawai, M. Sakata, Y. Kimura, and H. Hontani, “Dynamic PET image reconstruction using nonnegative matrix factorization incorporated with deep image prior,” in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, pp. 3126–3135, 2019.

[6] G. Naik and H. Nguyen, “Nonnegative matrix factorization for the identification of EMG finger movements: Evaluation using matrix analysis,” *IEEE J. Biomed. Health Inform.*, vol. 19, no. 2, pp. 478–485, 2015.

[7] D. Gurve and S. Krishnan, “Separation of fetal-ECG from single-channel abdominal ECG using activation scaled non-negative matrix factorization,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 3, pp. 669–680, 2020.

[8] J. L. Roux, F. Weninger, and J. Hershey, “Sparse NMF – half-baked or well done?”, *Mitsubishi Electric Research Laboratories, Tech. Rep.*, vol. TR2015-023, no. 11, 2015.

[9] A. Marmin, J. H. de M. Goulart, and C. Févotte, “Majorization-minimization for sparse nonnegative matrix factorization with the β-Divergence,” *IEEE Trans. Signal Process.*, vol. 71, pp. 1435–1447, 2023.

[10] T. Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 3, pp. 1066–1074, 2007.

[11] A. P.-Montano, J. Carazo, K. Kochi, D. Lehmann, and R. P.-Marqui, “Nonsmooth nonnegative matrix factorization (nsNMF),” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 403–415, 2006.

[12] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, “Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties,” *IEICE Trans. Fundamentals*, vol. E97-A, no. 5, pp. 1113–1118, 2014.

[13] Y. Iwase and D. Kitamura, “Supervised audio source separation based on nonnegative matrix factorization with cosine similarity penalty,” *IEICE Trans. Fundamentals*, vol. E105-A, no. 6, pp. 906–913, 2022.

[14] V. Leplat, N. Gillis, and A. M. S. Ang, “Blind audio source separation with minimum-volume beta-divergence NMF,” *IEEE Trans. Signal Process.*, vol. 68, pp. 3400–3410, 2020.

[15] Y. Sun, P. Babu, and D. P. Palomar, “Majorization-minimization algorithms in signal processing, communications, and machine learning,” *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, 2017.

[16] V. Leplat, N. Gillis, and J. Idier, “Multiplicative updates for NMF with β-divergences under disjoint equality constraints,” *SIAM J. Matrix Anal. Appl.*, vol. 42, no. 2, pp. 730–752, 2021.

[17] P. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, “Regularized adaptive notch filters for acoustic howling suppression,” *European Signal Process. Conf.*, pp. 2574–2578, 2009.

[18] Y. Li, X. Huang, Y. Zheng, Z. Gau, L. Kou, and J. Wan, “Howling detection and suppression based on segmented notch filtering,” *Sensors*, vol. 21, no. 8062, 2021.

[19] S. Takamichi, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji, and H. Saruwatari, “JVS corpus: free Japanese multi-speaker voice corpus,” *arXiv preprint*, 1908.06248, 2019.

[20] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.