

## ディリクレ非負値行列因子分解のハウリング抑圧への応用\*

©小川 遼, 北村 大地 (香川高専), 綾野 翔馬 (科学大)

## 1 はじめに

ハウリングとは、マイクロホン、アンプ、およびスピーカを含む音響システムで形成されるフィードバックループの利得が過大となってしまう生じる発振現象である。一度発生すると狭い周波数帯域に鋭いピークを持つ定常的な音が続く、音質の劣化や聴取者の不快感を引き起こす。会議、講演、および音楽ライブ演奏など多くの場面において、ハウリング抑圧は重要な課題になっている。従来、ハウリング抑圧には、発振周波数周辺を減衰させるノッチフィルタ [1] や、その減衰させるべき周波数を自動的に決定する適応ノッチフィルタ [2] が広く用いられてきた。

ハウリングは、狭帯域に集中した鋭いピークを持つスペクトル構造を持つものに対し、音声や楽器音などは複数の周波数成分から構成される広がりを持ったスペクトル構造を有している。従って、ハウリング抑圧の手法として、観測信号を複数の時間周波数成分に分解し、その中からハウリングに対応する成分のみを抑圧するという手法も考えられる。そこで本稿では、従来のノッチフィルタとは異なるハウリング除去のアプローチを採る目的で、ディリクレ非負値行列因子分解 (nonnegative matrix factorization: NMF) [3] を用いて、観測スペクトログラムを複数の成分に分解し、その中からハウリング成分を識別的に抽出し、ハウリング抑圧を行う手法を提案する。この手法では、NMFの基底ベクトルにディリクレ分布に基づく正則化を導入することで、ハウリング成分の誘導を行う。また、提案手法がスパース NMF [4] と比較して、安定してハウリング抑圧を行えることを実験的に確認する。

## 2 ディリクレ NMF

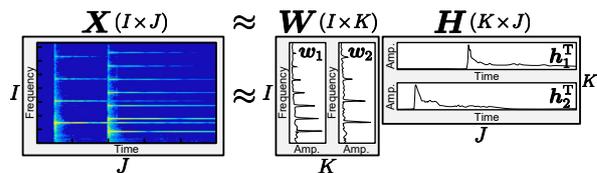
## 2.1 NMFの定式化

NMF [5] では非負観測行列  $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J}$  を非負基底行列  $\mathbf{W} \in \mathbb{R}_{\geq 0}^{I \times K}$  および非負係数行列  $\mathbf{H} \in \mathbb{R}_{\geq 0}^{K \times J}$  の行列積で  $\mathbf{X} \approx \mathbf{W}\mathbf{H}$  と近似する  $\mathbf{W}$  および  $\mathbf{H}$  を推定する。ここで、 $I$  および  $J$  は  $\mathbf{X}$  の行および列数、 $K$  は  $\mathbf{W}$  の列数 (基底数) である。NMF は  $\mathbf{X}$  の低ランク近似を目的とすることが多く、通常  $K \ll \min(I, J)$  である。

$\mathbf{W}$  および  $\mathbf{H}$  の推定は次の最小化問題となる。

$$\underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) \text{ s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k \quad (1)$$

ここで、 $w_{ik}$  および  $h_{kj}$  はそれぞれ  $\mathbf{W}$  および  $\mathbf{H}$  の要素、 $i = 1, \dots, I$  および  $j = 1, \dots, J$  はそれぞれ  $\mathbf{X}$  の行および列のインデックス、 $k = 1, \dots, K$  は  $\mathbf{W}$  の列のインデックスである。また、 $\mathbf{w}_k$  を  $\mathbf{W}$  の  $k$  番目

Fig. 1 Decomposition model of NMF ( $K = 2$ ).

の列ベクトル (基底ベクトル)、 $\mathbf{h}_k^T$  を  $\mathbf{H}$  の  $k$  番目の行ベクトル、 $\cdot^T$  を転置と定義する (Fig. 1 参照)。

最適化問題 (1) 中の  $\mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H})$  は  $\mathbf{X}$  と  $\mathbf{W}\mathbf{H}$  の乖離度を測る関数であり、本稿では次式の一般化 Kullback–Leibler (KL) 擬距離を用いる。

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

ここで、 $x_{ij}$  は  $\mathbf{X}$  の要素を表す。一般化 KL 擬距離を用いた NMF の最適化法は、補助関数法に基づく乗算型反復更新式 [5] が有名である。

## 2.2 ディリクレ分布

ディリクレ分布は、非負制約とノルム制約を満たす確率変数ベクトルに対して定義され、標準単体上の確率密度関数と解釈される。いま、ディリクレ分布に従う  $I$  次元確率変数ベクトルを  $\mathbf{z} = [z_1, \dots, z_I]^T \in \mathbb{R}_{\geq 0}^I$  と定義する。但し、 $\mathbf{z}$  は  $\sum_i z_i = 1$  を満たす。

$$\mathbf{z} \sim p(\mathbf{z}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_i z_i^{\alpha_i - 1} \quad (3)$$

ここで、 $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_I]^T \in \mathbb{R}_{\geq 0}^I$  は母数ベクトル、 $B(\boldsymbol{\alpha}) = [\prod_i \Gamma(\alpha_i)] / \Gamma(\sum_i \alpha_i)$  は多変量ベータ関数である。 $I = 3$  の時のディリクレ分布の確率密度関数を Fig. 2 に示す。ここで、図中の三角状の台は  $(I - 1)$  次元標準単体であり、3つの頂点がそれぞれ  $\mathbf{z} = [1, 0, 0]^T$ ,  $[0, 1, 0]^T$ , および  $[0, 0, 1]^T$  に対応する。

母数  $\boldsymbol{\alpha}$  は各要素が確率密度の各頂点への集中度に対応する。 $\alpha_i < 1$  のときに対応する頂点の確率密度が大きくなり、 $\alpha_i > 1$  は逆に小さくなる。また、 $\alpha_i = 1 \quad \forall i$  のとき、標準単体上で一様な確率密度となる。従って、母数  $\boldsymbol{\alpha}$  を変化させることで、生成される確率変数  $\mathbf{z}$  の傾向を制御できる。Fig. 2 (a) のように  $\alpha_i < 1 \quad \forall i$  とすれば、 $\mathbf{z}$  は one-hot なベクトルが生成されやすくなり、逆に Fig. 2 (b) のように  $\alpha_i > 1 \quad \forall i$  とすれば、 $\mathbf{z}$  は全要素が同じ (dense な) ベクトルが生成されやすくなる。

以上の性質を用いると、NMF の各ベクトル  $\mathbf{w}_k$  を one-hot や dense に誘導する新しい方法として、これらのベクトルの事前分布にディリクレ分布を仮定した最大事後確率推定を考えることができる。

\* Application of Dirichlet Nonnegative Matrix Factorization to Howling Suppression. By Haru OGAWA, Daichi KITAMURA (NIT, Kagawa), and Shoma AYANO (Science Tokyo).

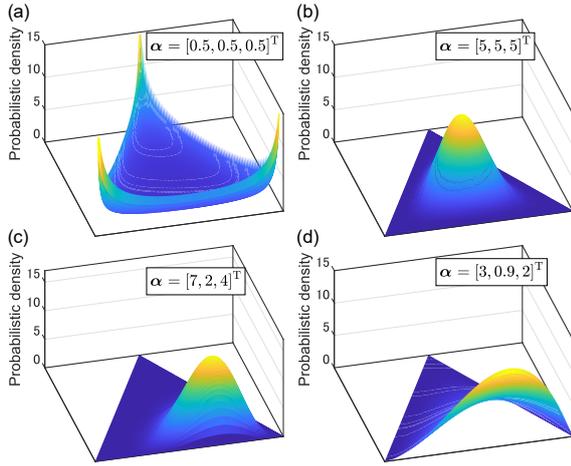


Fig. 2 Dirichlet distributions when  $I = 3$ , where support triangle indicates standard  $(I - 1)$ -simplex.

### 2.3 ディリクレ NMF [3] の目的関数および更新式

ディリクレ NMF は、 $K$  個の基底ベクトル  $\mathbf{w}_1, \dots, \mathbf{w}_K$  が独立にディリクレ分布に従うという事前仮定を導入した NMF である [3]。この事前分布に基づく正則化項を NMF の目的関数に加えることで、基底ベクトルのスペクトル構造を母数によって制御することが可能となる。今、観測行列  $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J}$  に対し、一般化 KL 擬距離を用いたディリクレ NMF を考える。このとき、目的関数は次式で与えられる。

$$\begin{aligned} & \underset{\mathbf{W}, \mathbf{H}}{\text{Minimize}} \mathcal{D}(\mathbf{X} | \mathbf{W}\mathbf{H}) + \sum_{i,k} (\alpha_{ik} - 1) \log w_{ik}^{-1} \\ & \text{s.t. } w_{ik}, h_{kj} \geq 0 \quad \forall i, j, k, \quad \sum_i w_{ik} = 1 \quad \forall k \quad (4) \end{aligned}$$

ここで、 $\alpha_{ik} > 0$  はディリクレ分布の母数を表す。 $\alpha_{ik}$  の値を調整することで、基底ベクトルの要素の分布を one-hot な形状から dense な形状まで制御することが可能である。

最適化には、補助関数法に基づく反復更新式を用いる [3]。各反復において、現在の基底行列  $\mathbf{W}$  から更新候補  $\hat{\mathbf{W}}$  を計算し、次式に従って基底行列を更新する。

$$w_{ik} = \begin{cases} \hat{w}_{ik} & (\text{if } \hat{w}_{ik} \geq 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (5)$$

ここで、 $\hat{w}_{ik}$  は補助関数の制約条件付き最小解として得られる更新候補であり、次式で与えられる。

$$\hat{w}_{ik} = \frac{w_{ik} \sum_j \frac{x_{ij}}{\sum_{k'} w_{ik'} h_{k'j}} h_{kj} + \alpha_{ik} - 1}{\sum_{i \in \mathcal{I}_k} \left( w_{ik} \sum_j \frac{x_{ij}}{\sum_{k'} w_{ik'} h_{k'j}} h_{kj} + \alpha_{ik} - 1 \right)} \quad (6)$$

ここで、 $\mathcal{I}_k$  は  $\hat{w}_{ik} \geq 0$  を満たすインデクス  $i$  の集合である。式 (5) および (6) により、基底行列  $\mathbf{W}$  の非負制約およびノルム制約が常に満たされる。一方、係数行列  $\mathbf{H}$  の反復更新式については、一般化 KL 擬距離を用いた従来の NMF と同一であり、文献 [5] に従う。これらの反復更新式は補助関数法に基づくことから、反復毎に目的関数値の単調非増加性が理論的に

保証される。なお、目的関数および反復更新式の詳細な導出については、文献 [3] を参照されたい。

### 2.4 拡張ディリクレ NMF

前節で述べたディリクレ NMF では、基底ベクトル  $\mathbf{w}_k$  の要素はディリクレ分布の母数  $\boldsymbol{\alpha}_k = [\alpha_{1k}, \alpha_{2k}, \dots, \alpha_{Ik}]^T$  によって制御される。一般に、 $\alpha_{ik} < 1 \quad \forall i$  とすることで基底ベクトル  $\mathbf{w}_k$  は one-hot な構造へと誘導される。しかし、 $\alpha_{ik}$  の値を指数的に 0 に近づけたとしても、式 (4) の正則化項の影響は次第に頭打ちとなり、より強力な one-hot 制約をかけることは困難となる。より強力な one-hot 正則化の実現には、ディリクレ分布の母数  $\alpha_{ik}$  の定義域を拡張する方法が考えられる。そこで本稿では、 $\alpha_{ik}$  の値を 0 以下まで拡張した拡張ディリクレ NMF を提案する。ディリクレ分布は本来  $\alpha_{ik} > 0$  を満たす母数に対してのみ定義される分布であり、 $\alpha_{ik} \leq 0$  では分布ではなくなる。従って、拡張ディリクレ NMF は最大事後確率推定ではなく、単純に式 (4) の正則化付き NMF と解釈される。本稿では、拡張ディリクレ NMF がより強力な one-hot 正則化を実現できるという実用上の利点に着目し、その用途や効果について検討する。

ディリクレ NMF における反復更新式 (6) では、 $\alpha_{ik}$  は基底更新式中に加算項としてのみ現れる。そのため、 $\alpha_{ik} \leq 0$  を許した場合も、更新式の形式自体を変更することなく適用できる。また、非負制約およびノルム制約は式 (5) および (6) によって常に満たされ、更新式そのものは数式上、一貫して定義される。

一方で、 $\alpha_{ik}$  を負の値に設定した場合、各基底ベクトル  $\mathbf{w}_k$  の更新過程において、式 (6) で計算される  $\hat{w}_{ik}$  が全ての  $i$  において負値になってしまう可能性がある。この場合、式 (5) による非負制約の射影が作用し、該当する基底ベクトル  $\mathbf{w}_k$  はゼロベクトルとなり、当該基底ベクトルは以降の反復において更新で変化せず、低ランク近似として有効な成分の推定ができなくなる。

本稿では、このように基底ベクトルがゼロベクトルとなり、NMF として破綻した状態に陥る事象が発生する割合を破綻率と定義する。拡張ディリクレ NMF の有効性を検証するため、従来のディリクレ NMF と比較し、ハウリング抑圧性能に加えて、 $\alpha_{ik}$  に対する破綻率の振る舞いについても実験的に確認する。

## 3 ハウリング抑圧実験

### 3.1 実験条件

本稿では、ディリクレ NMF および拡張ディリクレ NMF をハウリング抑圧のタスクに適用し、 $L_1$  正則化に基づくスパース NMF ( $L_1$ -NMF) [4] との比較を行った。Fig. 3 に、本実験で用いたハウリング抑圧のフレームワークを示す。ハウリング信号が混合している観測信号の振幅スペクトログラムを NMF によって複数の基底成分に分解し、その中からハウリング成分に対応する基底ベクトルおよびアクティベーションを除去する。その後、残りの NMF 成分を用いて時間周波数マスクを構成し、これを観測信号の複素ス

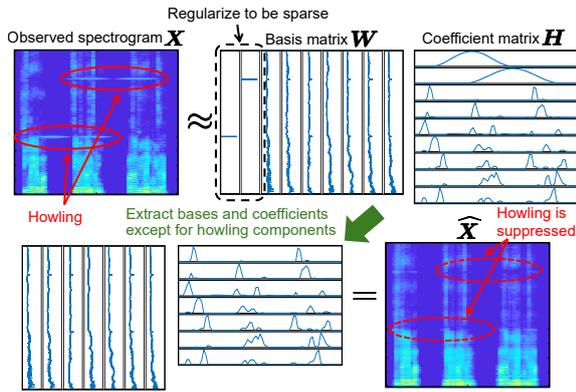


Fig. 3 Howling suppression framework based on NMF with sparsity regularization.

ペクトログラムに適用することで、ハウリング抑圧後の信号を再構成した。

実験には、音声信号および音楽信号の2種類のデータセットを用いた。音声データセットには *JVS corpus* [6] を、音楽データセットには *DSD100* [7] を用いた。いずれのデータセットにおいても、切り出し区間が完全な無音とならない音源のうち、アルファベット順で先頭から10個の信号を選択し、原信号として用いた。各音源信号はモノラル信号とし、サンプリング周波数を16 kHzに統一した。音声信号については信号冒頭から10 s区間を、音楽信号については30 sから40 sの10 s区間をそれぞれ切り出して使用した。

切り出した原信号に対して、異なる振幅および周波数をもつ2個の正弦波信号（ハウリング信号）を人工的に混合することで、ハウリングを含む観測信号を生成した。このハウリング信号の振幅は原信号の最大振幅に対する比率を0.2に設定し、発生区間に対してハン窓を適用することで、時間方向の中央付近で最大振幅となるようにした。周波数は区間 [80, 10000] Hz の範囲で観測信号毎にランダムに設定した。また、2個のハウリング信号はそれぞれ原信号の2 sから5 sおよび4 sから7 sの区間で発生するように生成した。これにより、複数周波数に同時にハウリングが発生する観測信号を模擬した。

$L_1$ -NMF 及びディリクレ NMF は観測信号の振幅スペクトログラムに適用した。短時間フーリエ変換の窓長は、音声信号では64 ms、音楽信号では128 msとし、シフト長はいずれも窓長の1/2とした。窓関数にはハン窓を用いた。各NMFにおいて、基底数  $K$  を複数の条件に設定して実験し、得られた中で最良の結果をもたらした  $K$  についての詳細な結果を示す。但し、いずれの  $K$  においても、ハウリング成分をモデル化する基底ベクトルとして  $w_1$  および  $w_2$  を用意し、原信号をモデル化する基底ベクトルとして  $w_3, \dots, w_K$  を用意した。このとき、ディリクレ NMF および拡張ディリクレ NMF では、ハウリング成分の基底ベクトルのパラメータを  $\alpha_1 = \alpha_2 = [\alpha_{\text{howl}}, \dots, \alpha_{\text{howl}}]^T$  とし、その他の基底ベクトルのパラメータを  $\alpha_{k'} = [\alpha_{\text{other}}, \dots, \alpha_{\text{other}}]^T$  ( $k' = 3, \dots, K$ ) とした。  $\alpha_{\text{howl}}$  および  $\alpha_{\text{other}}$  を広い範囲で変化させることで、母数の組み合わせがハウリング抑圧性能および破綻率に与える影響を検証した。事前に行ったパラメータ探索

に基づき、本稿では代表的なパラメータ設定の組み合わせについての結果を示す。ハウリング抑圧性能の評価指標には、*bss\_eval\_sources* [8] により算出される音源対歪み比 (source-to-distortion ratio: SDR) を用いた。観測信号における SDR と、抑圧後信号における SDR との差を SDR 改善量 (SDRi) として算出し、これを評価指標とした。

### 3.2 実験結果

ディリクレ NMF、拡張ディリクレ NMF、および  $L_1$ -NMF のいずれも初期値に依存する手法であるため、本実験では異なる100個の乱数シードを用いて乱数初期値を与え、各条件について試行を行った。

Fig. 4 (a) および (b) は、ディリクレ NMF および拡張ディリクレ NMF を適用した際の、音声信号および音楽信号に対する結果を示しており、10個の音源それぞれについて100回の試行で得られた SDRi 値の平均をヒートマップとして可視化したものである。各信号における最適条件下での平均 SDRi は、音声信号で7.3 dB、音楽信号で2.1 dBとなった。なお、これらの最適条件に対応するパラメータ領域は、図中に赤枠で示している。このときの基底数  $K$  は、音声信号では  $K = 100$ 、音楽信号では  $K = 5$  である。Fig. 5 は、 $L_1$ -NMF を適用した場合の結果を示しており、音声信号および音楽信号に対して、10個の音源それぞれについて100回の試行で得られた SDRi 値の平均を、正規化重み  $\beta$  毎に示している。最適条件下での平均 SDRi は、音声信号で-6.7 dB、音楽信号で-9.2 dBとなった。なお、本条件では基底数  $K = 50$  を用いている。これらの結果より、ディリクレ NMF および拡張ディリクレ NMF は、 $L_1$ -NMF と比較して、音声信号および音楽信号のいずれにおいても高いハウリング抑圧性能を示すことが確認された。特に音楽において、母数を負の領域まで拡張した拡張ディリクレ NMF では、従来のディリクレ NMF では十分に誘導できなかった基底構造が明確化され、SDRi 値の向上が得られていることが分かる。

Fig. 6 は、音声信号および音楽信号に対して、ディリクレ NMF および拡張ディリクレ NMF を適用した際の2.4節で定義した破綻率の分布を示している。図中の赤枠で囲まれた領域は、すべての試行において破綻が一度も発生しなかったパラメータ領域を示している。図より、 $\alpha_{ik}$  を負の値に設定した場合であっても、すべての条件において破綻が頻発するわけではなく、特定のパラメータ領域では低い破綻率を保ったまま高い SDRi 値が得られていることが確認できる。一方で、母数を負の方向に過度に大きく設定した条件では、破綻率が増加する傾向が見られた。このような破綻率の振る舞いには、基底数および信号の特性が影響していると考えられる。基底数が比較的少ない条件では、1つの基底ベクトルが担う役割が大きくなり、更新過程において基底ベクトルがゼロベクトルになりにくくなるため、破綻が生じにくい傾向がある。一方で、基底数が多い条件では、低ランク近似で用いることのできる基底ベクトルの数に余裕ができるため、一部の基底ベクトルが更新過程で急激に

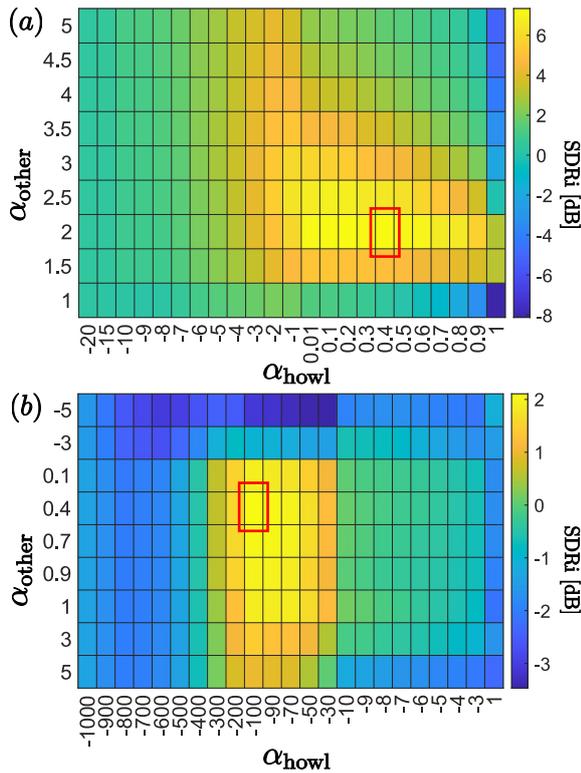


Fig. 4 Average SDRi obtained by Dirichlet NMF ( $\alpha > 0$ ) and extended Dirichlet NMF for (a) speech and (b) music signals. The red boxes indicate the highest performance in each experiment.

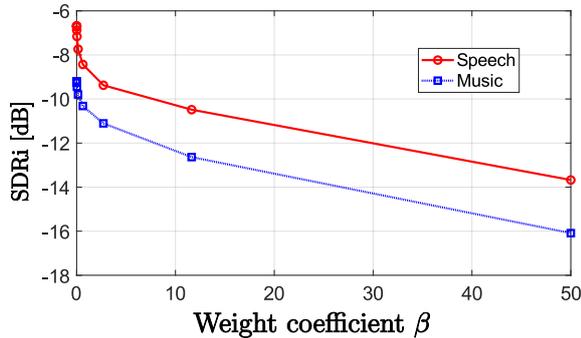


Fig. 5 Average SDRi obtained by  $L_1$ -NMF.  $\beta$  corresponds to the regularization weight.

減衰し、破綻に至る可能性が高くなると考えられる。また、音声信号と音楽信号の間でも破綻率の傾向に差が見られた。音楽信号は音声信号と比較して、振幅スペクトログラムが低ランクになる傾向にあるため、基底構造が明確になりやすく、結果として破綻が生じにくい傾向を示したと考えられる。

以上より、拡張ディリクレ NMF は、破綻率とのトレードオフを伴うものの、適切な基底数および母数の設定を行うことで、従来手法を上回るハウリング抑圧性能を安定して実現できる手法であることが示された。

#### 4 おわりに

本稿では、ディリクレ NMF および拡張ディリクレ NMF をハウリング抑圧に適用し、 $L_1$ -NMF との比較を行った。音声信号および音楽信号を用いた実験によ

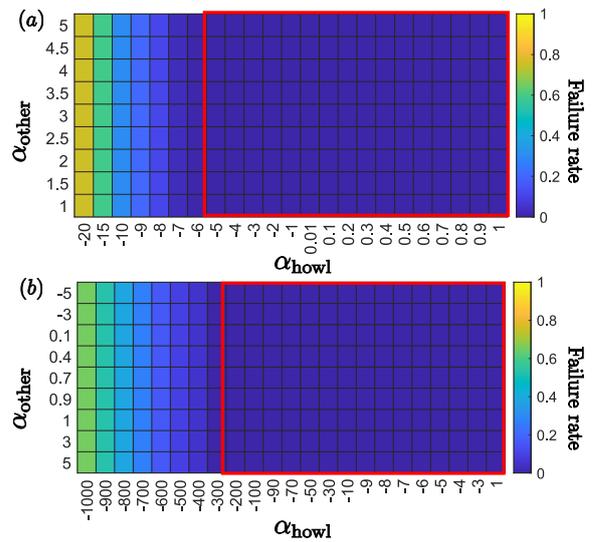


Fig. 6 Distribution of failure rates for extended Dirichlet NMF under different parameter settings. The red boxes indicate parameter regions where the failure rate is zero.

り、提案手法はいずれの信号においても  $L_1$ -NMF と比較して高い SDR 改善量を示し、ハウリング抑圧に有効であることを確認した。特に音楽信号においては、拡張ディリクレ NMF により基底構造の誘導が明確化され、性能向上が得られる条件が存在することを示した。また、拡張ディリクレ NMF における破綻率について検討した結果、パラメータを負の領域まで拡張した場合であっても、適切な条件設定では破綻を生じさせずに高い抑圧性能を実現できることが分かった。

謝辞 本研究の一部は科研費 23K24908 および Google Gift 2024 の助成を受けたものである。

#### 参考文献

- [1] T. van Waterschoot and M. Moonen, "Fifty years of acoustic feedback control: State of the art and future challenges," *Proc. IEEE*, vol. 99, no. 2, pp. 288–327, 2011.
- [2] P. Gil-Cacho, T. van Waterschoot, M. Moonen, and S. H. Jensen, "Regularized adaptive notch filters for acoustic howling suppression," in *Proc. European Signal Process. Conf.*, pp. 2574–2578, 2009.
- [3] 小川遼, 北村大地, 綾野翔馬, "ディリクレ分布に基づく正規化付き非負値行列因子分解," *日本音響学会 2025 年春季研究発表会講演論文集*, pp. 225–228, 2025.
- [4] W. Liu, N. Zheng, and X. Lu, "Non-negative matrix factorization for visual coding," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. III-293, 2003.
- [5] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Advances in Neural Info. Process. Syst.*, 2000.
- [6] 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.
- [7] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," in *Proc. Int. Conf. Latent Variable Anal. Signal Sep.*, pp. 323–332, 2017.
- [8] 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.