日本音響学会誌 75 巻 3 号（2019），pp. 130–138

小特集―位相情報を考慮した音声音響信号処理―

複素生成モデルに基づく非負値行列因子分解と音源分離への応用*

北 村 大 地（香川高等専門学校）**

43.60.Cg, Uv

1. はじめに

非負値行列因子分解（nonnegative matrix factorization: NMF）[1, 2] は，非負要素から成る 2 次元データ，すなわち非負行列から，少数の有意な非負特徴量を抽出する数理アルゴリズムである。非負行列を観測データとするパターン抽出に効果的であり，その解釈の容易さから様々な分野で活用されてきたが，とりわけ音響分野において高度に発展してきた歴史がある。

今，解析対象となる非負観測行列を $\boldsymbol{X} \in \mathbb{R}_{\geq 0}^{I \times J}$ とおく。ここで，I と J はそれぞれ $\boldsymbol{X}$ の行数と列数である。このとき，NMF は次式で示される近似分解を行う。

$$\boldsymbol{X} \approx \hat{\boldsymbol{X}} = \boldsymbol{W}\boldsymbol{H} \tag{1}$$

$$= \sum_k \boldsymbol{w}_k \boldsymbol{h}_k^{\mathrm{T}} \tag{2}$$

ここで，$\boldsymbol{W} = (\boldsymbol{w}_1 \ \cdots \ \boldsymbol{w}_K) \in \mathbb{R}_{\geq 0}^{I \times K}$ は基底行列と呼ばれ，$\boldsymbol{X}$ 中の有意な非負の特徴量ベクトル（基底）$\boldsymbol{w}_k \in \mathbb{R}_{\geq 0}^{I \times 1}$ を列に持つ行列である。また，$\boldsymbol{H} = (\boldsymbol{h}_1 \ \cdots \ \boldsymbol{h}_K)^{\mathrm{T}} \in \mathbb{R}_{\geq 0}^{K \times J}$ は係数行列やアクティベーション行列と呼ばれ，基底 $\boldsymbol{w}_k$ の係数ベクトル $\boldsymbol{h}_k \in \mathbb{R}_{\geq 0}^{J \times 1}$ を行に持つ行列である。式 (1) を要素ごとに記述すると次式となる。

$$x_{ij} \approx \hat{x}_{ij} = \sum_k w_{ik} h_{kj} \tag{3}$$

ここで，x_{ij}，$\hat{x}_{ij}$，w_{ik}，及び h_{kj} はそれぞれ $\boldsymbol{X}$，$\hat{\boldsymbol{X}}$，$\boldsymbol{W}$，及び $\boldsymbol{H}$ の非負要素である。添え字の i, j, 及び k はそれぞれ $\boldsymbol{X}$ の行，列，及び基底のインデクスを表す。一般に基底数 K は I や J よりも十分小さな値に設定されるため，式 (1) は観測行列 $\boldsymbol{X}$ を限られた数（K 個）の基底で表現する低ランク近似と解釈でき，少数の有意な非負特徴量が基底 $\boldsymbol{w}_k$ 及び係数 $\boldsymbol{h}_k$ として得られる。

NMF の変数は次式の最適化問題で推定される。

$$\min_{\boldsymbol{W},\boldsymbol{H}} \mathcal{D}(\boldsymbol{X}|\boldsymbol{W}\boldsymbol{H}) \text{ s.t. } w_{ik}, h_{kj} \geq 0 \ \forall i, j, k \tag{4}$$

ここで $\mathcal{D}(\cdot|\cdot)$ は二つの行列の類似度を測る任意の関数であり，二乗 Euclid 距離 [2]，一般化 Kulback–Leibler（KL）ダイバージェンス [2]，Itakura–Saito（IS）ダイバージェンス [3] 等がよく用いられる。従って式 (4) は，非負観測行列 $\boldsymbol{X}$ を良く近似するような非負低ランクモデル行列 $\boldsymbol{W}\boldsymbol{H}$ を推定する問題となる。式 (4) の解は閉形式では与えられないため，$\boldsymbol{W}$ と $\boldsymbol{H}$ を何等かの方法で初期化（例えば [4] 等）したうえで，反復計算により解を求めるアルゴリズムが提案されている [5, 6]。

音響信号処理においては，信号を短時間フーリエ変換（short-time Fourier transform: STFT）して得られる複素スペクトログラム $\boldsymbol{C} \in \mathbb{C}^{I \times J}$（行列 $\boldsymbol{C}$ の複素要素を c_{ij} と定義する）を非負化したものを観測行列 $\boldsymbol{X}$ とおくことが一般的である。このとき，I は周波数ビン数，J は時間フレーム数に対応する。特に，振幅スペクトログラム（$\boldsymbol{X} = |\boldsymbol{C}|^{\cdot 1}$）やパワースペクトログラム（$\boldsymbol{X} = |\boldsymbol{C}|^{\cdot 2}$）に NMF を適用する例が多い[1]。**Fig. 1** に，パワースペクトログラムを NMF で分解した例を示す。観測行列には音高の異なる二つのピアノ音が含まれているが，基底数を $K = 2$ と設定して分解することで，各音の調波構造（スペクトル）が基底 $\boldsymbol{w}_1$ 及び $\boldsymbol{w}_2$ に，各音の時間強度が係数 $\boldsymbol{h}_1$ 及び $\boldsymbol{h}_2$ にそれぞれ現れている。このように，NMF は音響信号中の

* Nonnegative matrix factorization based on complex generative model and its application to audio source separation.

** Daichi Kitamura (National Institute of Technology, Kagawa Collage, Takamatsu, 761–8058)

[1] 行列に対するドット付き指数と絶対値記号を組み合わせた演算子 $|\cdot|^{\cdot p}$ を，要素ごとの絶対値の p 乗を要素に持つ行列と定義する。すなわち，$[|\boldsymbol{C}|^{\cdot p}]_{ij} = |c_{ij}|^p$ である。

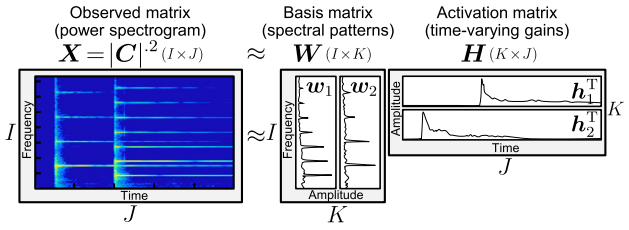

Fig. 1 NMF decomposition for audio signals.

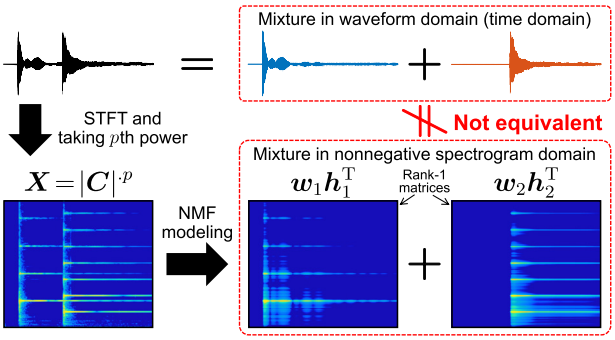

Fig. 2 Wrong mixture model assumed in NMF for audio signals.

頻出スペクトル及びその時間的な強度変化を基底及び係数として抽出できる教師無し学習である。

NMF は音響信号から直感的な特徴量を簡便に抽出できるため，音源分離 [7–11]，自動採譜 [12–14]，音響イベント検出 [15]，音超解像 [16] 等，様々なタスクに応用されてきた。本稿では，非負スペクトログラム $\boldsymbol{X}$ ではなく，複素スペクトログラム $\boldsymbol{C}$ を NMF でモデル化する手法と，複素生成モデルによる統計的解釈及びその拡張理論について紹介し，音源分離問題への応用例を示す。

2. NMF に基づく音響モデリングの問題点

NMF において非負行列 $\boldsymbol{X}$ は，式 (2) に示されるように，ランク 1 の非負行列 $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ の和として近似表現される。この分解モデルは，本来非負であるような観測データ（白黒画像の輝度値や顧客の商品購買数等）に対しては正当性があるが，音響信号の分解に対しては次のような問題が生じる。

Fig. 2 に示すように，本来音響信号の混合は時間領域の信号の和であり，時間周波数領域では各音響信号の複素スペクトログラムの和に対応する。一方で，音響信号に NMF を適用することは，「複素スペクトログラムの和」を「非負のランク 1 スペクトログラム成分 $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ の和」で近似表現することに相当する。一般に二つの複素数値 c_1 及び c_2 に対して非負値の加法性 $|c_1+c_2|^p = |c_1|^p + |c_2|^p$ は $p \neq 0$ で成り立たないため，NMF で仮定される非負スペクトログラム成分の加法性は物理的に間違ったモデルである。より具体的には，二つの音響信号が混合する際の位相ずれによる打ち消し合いが取り扱われないため，NMF によるモデル化では，位相に起因する物理現象が考慮されないまま基底及び係数を推定していることになる。これは NMF を用いた音響信号処理特有の問題であり，NMF が複素行列ではなく非負行列を対象としていることが原因である。人間の聴覚は位相スペクトルに関して鈍感ではあるが，インパルス信号と白色雑音のように，位相スペクトルが人間の知覚に大いに寄与する場合もある。また，NMF で推定した非負スペクトログラム成分 $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ を逆 STFT で時間領域に戻す際にも，$\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ に何等かの位相を付与する必要があり，例えば観測複素スペクトログラム $\boldsymbol{C}$ の位相をそのまま付与する方法，$\boldsymbol{C}$ に対して Wiener フィルタを適用する方法，位相復元法 [17–19] を適用する方法等がある。

この問題を解決するために，これまでに様々な手法が提案されている。本稿では，特に，NMF の分解成分 $w_{ik}h_{kj}$ を複素数に拡張した複素 NMF (complex NMF: CNMF) [20–22] 及び複素 Gauss 分布生成モデルに基づく NMF [3] とその理論拡張手法 [23–25] を取り上げて紹介する。

3. 位相スペクトルを考慮した CNMF

従来の NMF では，位相スペクトルの影響を無視して少数の非負スペクトログラム成分 $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ を推定しており，現実の物理現象との乖離がある。そこで，少数の音響成分を非負値から複素数値に拡張した CNMF が提案されている [20]。CNMF では，複素スペクトル c_{ij} を次式のように分解する。

$$c_{ij} \approx \hat{c}_{ij} = \sum_k \hat{c}_{ij,k} \tag{5}$$

$$= \sum_k w_{ik}h_{kj}e^{\mathrm{j}\phi_{ij,k}} \tag{6}$$

ここで，$\mathrm{j} = \sqrt{-1}$ である。また，$\hat{c}_{ij,k} \in \mathbb{C}$ は複素スペクトル成分であり，$|\hat{c}_{ij,k}| = w_{ik}h_{kj}$ 及び $\arg(\hat{c}_{ij,k}) = \phi_{ij,k}$ である。すなわち，ランク 1 の振幅スペクトログラム成分 $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ と位相スペクトログラム成分 $\boldsymbol{\Phi}_k \in \mathbb{R}_{[0,2\pi]}^{I\times J}$（要素は $\phi_{ij,k}$）から成

る複素スペクトログラム成分 $\hat{\boldsymbol{C}}_k \in \mathbb{C}^{I\times J}$（要素は $\hat{c}_{ij,k}$）を定義し，観測複素スペクトログラム $\boldsymbol{C}$ を

$$\boldsymbol{C} \approx \hat{\boldsymbol{C}} = \sum_k \hat{\boldsymbol{C}}_k \tag{7}$$

と近似するモデルである。

CNMF で推定すべき変数は w_{ik}, h_{kj}, 及び $\phi_{ij,k}$ であり，次式の生成モデルに基づく最尤推定として定式化される。

$$c_{ij} = \hat{c}_{ij} + \varepsilon_{ij} \tag{8}$$

$$\varepsilon_{ij} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2) \tag{9}$$

$$\mathcal{N}_{\mathbb{C}}(c; \mu, \sigma^2) = \frac{1}{\pi\sigma^2} \exp\left(-\frac{|c-\mu|^2}{\sigma^2}\right) \tag{10}$$

ここで，$\mathcal{N}_{\mathbb{C}}(c; \mu, \sigma^2)$ は平均 μ，分散 σ^2 の点対称（等方性）複素 Gauss 分布であり，$\mu = 0$ の場合は **Fig. 3** に示すような複素平面上の原点対称分布である。これは，式 (10) が $\mu = 0$ において複素確率変数のノルム $|c|$ にのみ依存する確率密度関数であることを示している。CNMF は式 (7) のように少数の複素スペクトログラム成分 $\hat{\boldsymbol{C}}_k$ の和で観測複素スペクトログラム $\boldsymbol{C}$ を近似するが，その近似誤差 ε_{ij} が時間周波数ごとに独立な Fig. 3 の分布に従うと仮定すると，変数 w_{ik}, h_{kj}, 及び $\phi_{ij,k}$ の最尤推定として次の最小化問題を得る。

$$\min_{w_{ik}, h_{kj}, \phi_{ij,k}} \sum_{i,j} \left| c_{ij} - \sum_k \hat{c}_{ij,k} \right|^2$$
$$\text{s.t. } w_{ik}, h_{kj} \geq 0 \ \forall i, j, k, \ \sum_i w_{ik} = 1 \ \forall k$$

これは二乗 Euclid 距離に基づく NMF の複素数版と解釈でき，補助関数法を用いることで通常の NMF と同様に最適化が可能である [20]。更に近年では，CNMF の距離規範を一般化 KL ダイバージェンスや β ダイバージェンス（二乗 Euclid 距離，一般化 KL ダイバージェンス，及び IS ダイバージェンスを含むより一般的な類似度関数）へと拡張する手法が提案されている [21, 22]。

以上のように，CNMF は振幅スペクトログラムの低ランク構造を仮定しつつ複素スペクトログラム成分の和で観測信号を近似する物理的に正しいモデルとなっており，NMF の直接的な複素数拡張と解釈される。しかしながら，通常の NMF 変数に加えて，新たに位相を推定しなければならないため，最適化の困難性に起因する極端な初期値依存の問題が指摘されている [26]。

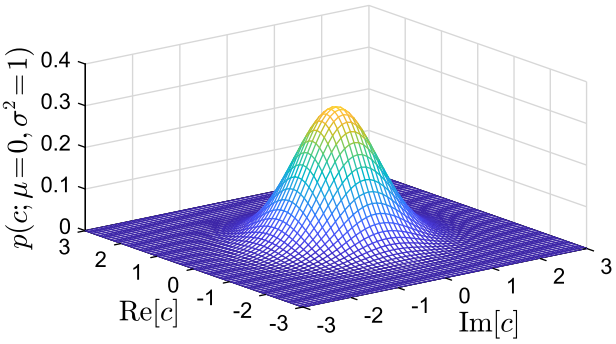

Fig. 3 Isotropic complex Gaussian distribution.

4. 複素生成モデルに基づく NMF

CNMF は，音響信号の混合現象を少数の複素スペクトログラム成分の和でモデル化することで，2 章に述べた問題を解決している。一方，一部の類似度関数に基づく NMF の最適化問題を，複素スペクトログラムの生成モデルに基づく非負パラメータの最尤推定問題として解釈できることが明らかにされており，非負スペクトログラム成分の加法性を統計的な意味で正当化できる方法が発見されている。本章では，NMF の複素生成モデルに基づく解釈について解説し，その理論的な拡張について紹介する。

4.1 IS ダイバージェンスに基づく NMF の統計的解釈

IS ダイバージェンスは次式で定義される。

$$\mathcal{D}(c|\sigma) = \frac{|c|^2}{\sigma^2} - \log\frac{|c|^2}{\sigma^2} - 1 \tag{11}$$

式 (4) における類似度関数 $\mathcal{D}$ を式 (11) で定義し，$|c|^2 = x_{ij}$ 及び $\sigma^2 = \sum_k w_{ik}h_{kj}$ と置いた NMF（ISNMF）の最適化問題は，変数 w_{ik} 及び h_{kj} に無関係な項を省略して記述すると次式のようになる。

$$\min_{\boldsymbol{W},\boldsymbol{H}} \sum_{i,j} \left(\frac{x_{ij}}{\sum_k w_{ik}h_{kj}} + \log \sum_k w_{ik}h_{kj} \right)$$
$$\text{s.t. } w_{ik}, h_{kj} \geq 0 \ \forall i, j, k \tag{12}$$

この最小化問題において，非負観測データの定義を $x_{ij} = |c_{ij}|^2$（パワースペクトル）とした場合，以下に示すような統計的解釈が可能となる。

今，観測された複素スペクトログラム $\boldsymbol{C}$ の各時

間周波数における成分 c_{ij} が $c_{ij} = \sum_k c_{ij,k}$ として少数の複素スペクトル成分 $c_{ij,k}$ の和で構成されていると仮定する。また，各複素スペクトル成分は，次式のように Fig. 3 の原点対称複素 Gauss 分布から生成されたものと仮定する。

$$c_{ij,k} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{ij,k}) \tag{13}$$

ここで，分散 $\sigma^2_{ij,k} > 0$ は周波数 i と時間 j の両方に依存して変化する非負パラメータである。このとき，複素 Gauss 分布の安定性（あるいは再生性）により，$c_{ij,k}$ の和である c_{ij} もまた分散 $\sigma^2_{ij} = \sum_k \sigma^2_{ij,k}$ の複素 Gauss 分布から生成される。

$$\begin{aligned} c_{ij} &\sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{ij}) \\ &= \mathcal{N}_{\mathbb{C}}(0, \textstyle\sum_k \sigma^2_{ij,k}) \end{aligned} \tag{14}$$

式 (14) は観測スペクトルの生成モデルとして，時間周波数ごとに分散の異なる原点対称複素 Gauss 分布 $p(c_{ij}) = \mathcal{N}_{\mathbb{C}}(0, \sigma^2_{ij})$ を仮定している。この生成モデルを **Fig. 4** に示す。パワーの大きな時間周波数グリッドでは大きな分散値を持つため振幅の大きい複素数値 c_{ij} を生成し易く，逆にパワーの小さな時間周波数グリッドではほとんど零付近の複素数値 c_{ij} しか生成しないような分布となっている。また，いずれの時間周波数グリッドも原点対称な分布であることから，位相 $\arg(c_{ij})$ に関しては無情報（一様）な分布である。各時間周波数グリッドにおける分散 σ^2_{ij} は複素スペクトル成分 c_{ij} のパワーの期待値であるため

$$\begin{aligned} \sigma^2_{ij} &= \mathrm{E}[|c_{ij}|^2] \\ &= \sum_k \mathrm{E}[|c_{ij,k}|^2] \\ &= \sum_k \sigma^2_{ij,k} \end{aligned} \tag{15}$$

が成立する。ここで，$\mathrm{E}[\cdot]$ は観測データに対する期待値を表す。式 (14) の生成モデルが周波数 i と時間 j に対して統計的に独立であると仮定すると，観測複素スペクトログラム $\boldsymbol{C}$ の（時間周波数全体における）生成モデルは次式となる。

$$\begin{aligned} \boldsymbol{C} &\sim p(c_{11}, c_{12}, \cdots, c_{IJ}) \\ &= \prod_{i,j} p(c_{ij}) \\ &= \prod_{i,j} \mathcal{N}_{\mathbb{C}}(0, \textstyle\sum_k \sigma^2_{ij,k}) \end{aligned} \tag{16}$$

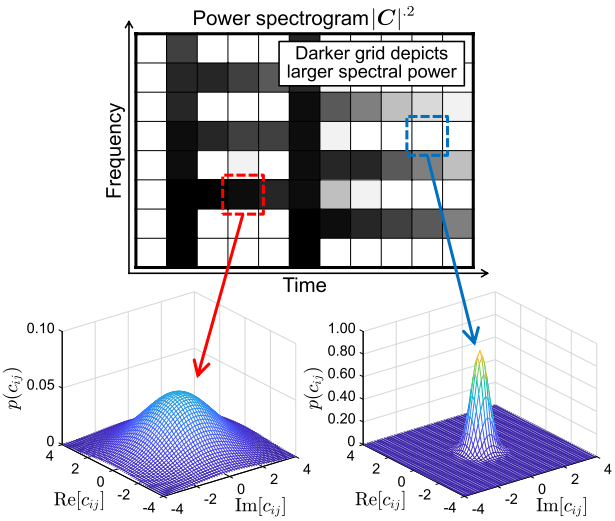

Fig. 4 Local Gaussian model assumed in ISNMF.

この生成モデルは多チャネル信号（多変量分布）にも拡張され，local Gaussian model（LGM）[27] と呼ばれている。

次に，式 (16) の複素生成モデルに基づいて，分散 $\sigma^2_{ij,k}$ の最尤推定問題を考える。観測データ $\boldsymbol{C}$ の尤度関数 $\mathcal{L}$ は次式で表される。

$$\begin{aligned} \mathcal{L} &= \prod_{i,j} \mathcal{N}_{\mathbb{C}}(0, \textstyle\sum_k \sigma^2_{ij,k}) \\ &= \prod_{i,j} \frac{1}{\pi \sum_k \sigma^2_{ij,k}} \exp\left(-\frac{|c_{ij}|^2}{\sum_k \sigma^2_{ij,k}} \right) \end{aligned} \tag{17}$$

更に，負対数尤度関数は

$$\begin{aligned} &-\log \mathcal{L} \\ &= \sum_{i,j} \left(\frac{|c_{ij}|^2}{\sum_k \sigma^2_{ij,k}} + \log \sum_k \sigma^2_{ij,k} + \log \pi \right) \end{aligned} \tag{18}$$

となる。従って，分散の最尤値を得るには，負対数尤度 (18) を最小化する $\sigma^2_{ij,k}$ を求めればよい。ここで，式 (18) と ISNMF の最小化問題 (12) と比較すると，非負観測データ x_{ij} 及び分散 $\sigma^2_{ij,k}$ をそれぞれ $x_{ij} = |c_{ij}|^2$ 及び $\sigma^2_{ij,k} = w_{ik}h_{kj}$ と置くとき，両式は定数項の違いを除いて一致する。

以上より，ISNMF をパワースペクトログラム $\boldsymbol{X} = |\boldsymbol{C}|^{\cdot 2}$ に対して適用することは，式 (16) の複素生成モデル（LGM）を仮定した分散 $\sigma^2_{ij,k}$ の最尤推定問題と等価であることが分かる。また，この複素生成モデルでは，非負成分 $w_{ik}h_{kj}$ の和

が，式 (15) に示す分散 $\sigma_{ij,k}^2 = \mathrm{E}[|c_{ij,k}|^2]$ の和に対応していることが分かる。この事実は，複素スペクトル成分の混合（$c_{ij} = \sum_k c_{ij,k}$）を非負パラメータである分散値の和（$\sigma_{ij}^2 = \sum_k \sigma_{ij,k}^2$）で表現できることを意味し，音響信号に NMF を適用する際の「複素数を非負化して少数の成分に分解」という一見場当たり的にみえる手続きを期待値の意味で正当化できる。従って，パワースペクトログラムを ISNMF で少数の非負スペクトログラム成分 $\boldsymbol{w}_k\boldsymbol{h}_k^{\mathrm{T}}$ の和に近似分解する場合，Fig. 2 に示すような時間領域での信号の混合現象を，時間周波数領域の非負値の和でモデル化することの正当性が保障される。この LGM に基づく定式化は，多チャネル信号の音源分離へと応用されている [28–31]。

4.2 LGM の理論拡張に基づく NMF

ISNMF のように音響信号の非負スペクトログラムの加法性を期待値の意味で正当化するためには，安定性を持つ分布，すなわち安定分布を生成モデルとして仮定する必要がある。安定性とは，同じ分布族から独立に生成された二つの確率変数 v_1 と v_2 に対して，その線形結合 $av_1 + bv_2$ と，別の確率変数 $dv + e$（ここで $a > 0$，$b > 0$，$d > 0$，及び e は定数）がいずれも同一の分布族から生成される性質を指す [32]。このような安定分布を生成モデルとして仮定した場合，確率変数の和をその分布のパラメータの和としてモデル化することが可能となる。事実，原点対称複素 Gauss 分布は安定分布の特殊形であり，次式のように複素数値の和を分散の和で表現できる。

$$\begin{aligned} &c_1 \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_1^2),\ c_2 \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_2^2) \\ &c_1 + c_2 \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_1^2 + \sigma_2^2) \end{aligned} \tag{19}$$

同様に，複素数値の和を 1 次の期待値（尺度母数）γ の和で表現できる分布として，次式がある [32]。

$$\mathcal{C}_{\mathbb{C}}(c; 0, \gamma) = \frac{2^{-1/2}\gamma}{2\pi\left[|c|^2 + \left(2^{-1/2}\gamma\right)^2\right]^{\frac{3}{2}}} \tag{20}$$

ここで，$\mathcal{C}_{\mathbb{C}}(c; 0, \gamma)$ は最頻値 0 かつ尺度母数 γ の原点対称複素 Cauchy 分布である[2]。複素スペクトル成分 $c_{ij,k}$ が式 (20) から生成されるとき，時間周波数ごとに定義される尺度母数は複素スペクトル成分の振幅の期待値 $\gamma_{ij,k} = \mathrm{E}[|c_{ij,k}|]$ に対応する。複素 Cauchy 分布は安定性を持つため，観測複素スペクトル $c_{ij} = \sum_k c_{ij,k}$ の生成モデルもまた

$$c_{ij} \sim \mathcal{C}_{\mathbb{C}}(0, \textstyle\sum_k \gamma_{ij,k}) \tag{21}$$

となり，振幅スペクトル成分 $\gamma_{ij,k} = w_{ik}h_{kj}$ の加法性が期待値の意味で正当化される。

この生成モデルに基づく NMF（Cauchy NMF）[23] が提案され，パワースペクトログラム $|\boldsymbol{C}|^{\cdot 2}$ 及び振幅スペクトログラム $|\boldsymbol{C}|^{\cdot 1}$ の加法性を正当化できる NMF がそれぞれ ISNMF [3] 及び Cauchy NMF [23] として与えられた。また，複素 Gauss 分布と複素 Cauchy 分布の両方を含む一般化である複素 Student's t 分布に基づく NMF（tNMF）[24] も提案されており，自由度パラメータ ν を変化させることで両分布の中間的な生成モデルを仮定することが可能となっている。ただし，安定性は Gauss 分布（$\nu \to \infty$）と Cauchy 分布（$\nu = 1$）の場合にのみ成立するため，非負スペクトログラムの加法性は両分布でのみ正当化される。tNMF も LGM と同様に多チャネル音源分離へと応用され [33, 34]，より高精度な分離を達成している。なお，各 NMF の最適化アルゴリズム（反復更新式）については，それぞれの文献を参照されたい。

5. 複素一般化 Gauss 分布に基づく NMF とスパース雑音除去への応用

本章では，LGM の新たな一般化として，複素一般化 Gauss 分布（generalized Gaussian distribution: GGD）[25] に基づく NMF（GGDNMF）について解説し，その性質を活用したスパース雑音除去の例を紹介する。

5.1 GGDNMF の生成モデル

GGDNMF では，ISNMF で仮定されていた LGM を複素 GGD へと拡張している。すなわち，平均値 0，形状母数 $\rho > 0$，時間周波数ごとに異なる尺度母数 σ_{ij} を持つ原点対称複素 GGD $\mathcal{G}_{\mathbb{C}}(c_{ij}; 0, \rho, \sigma_{ij})$ を，観測複素スペクトログラム $\boldsymbol{C}$ の生成モデルとして仮定する。

$$\boldsymbol{C} \sim \prod_{i,j} \mathcal{G}_{\mathbb{C}}(c_{ij}; 0, \rho, \sigma_{ij})$$

[2] 文献 [23] では，式 (20) の尺度母数 γ の係数 $2^{-1/2}$ はない。これは文献 [23] が内部で参照している文献 [32]（pp. 80–81）の誤植に由来する。係数 $2^{-1/2}$ を付与することで，$|c| = \gamma$ のときに尤度が最大となる。

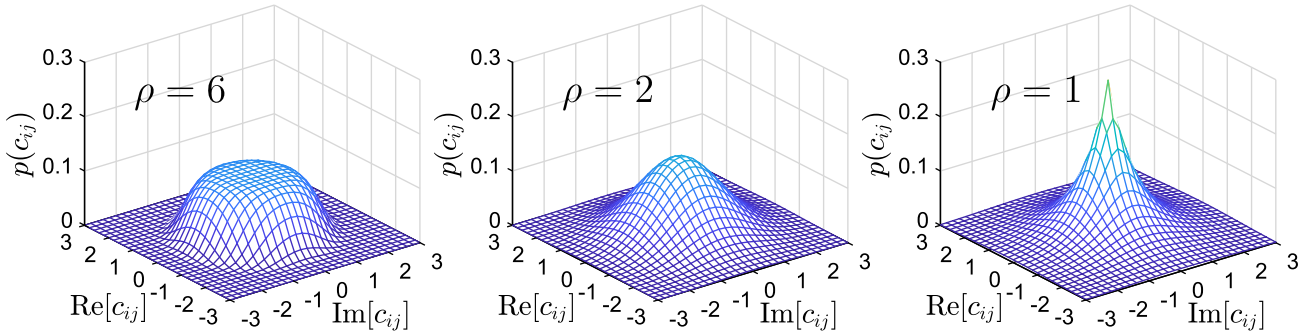

Fig. 5 Isotropic complex GGD with $\sigma_{ij} = 1.5$.

$$= \prod_{i,j} \frac{\rho^{1-\frac{2}{\rho}}}{2^{1-\frac{2}{\rho}}\pi\sigma_{ij}^2\Gamma(2/\rho)} \exp\left[-\frac{2}{\rho}\left(\frac{|c_{ij}|}{\sigma_{ij}}\right)^{\rho}\right] \tag{22}$$

$$\sigma_{ij}^{p} = \sum_{k} w_{ik}h_{kj} \tag{23}$$

ここで，$\Gamma(\cdot)$ はガンマ関数であり，p は低ランク近似の領域に対応するドメインパラメータである。すなわち，$p = 1$ の場合は振幅スペクトログラム $|\boldsymbol{C}|^{\cdot 1}$，$p = 2$ の場合はパワースペクトログラム $|\boldsymbol{C}|^{\cdot 2}$ を $\boldsymbol{W}$ と $\boldsymbol{H}$ で低ランク近似することに対応する。複素 GGD $\mathcal{G}_{\mathbb{C}}(c_{ij}; 0, \rho, \sigma_{ij})$ は **Fig. 5** に示すとおり，$\rho = 2$ 及び $\rho = 1$ のとき，それぞれ原点対称な複素 Gauss 分布及び複素 Laplace 分布に一致する。また，$\rho > 2$ では劣 Gauss (platykurtic)，$\rho < 2$ では優 Gauss (leptokurtic) な分布となる。

5.2 複素 GGD から導かれるダイバージェンス

式 (22) の生成モデルと NMF における類似度関数の関係を明らかにするために，対数尤度差 (deviance) から複素 GGD に基づくダイバージェンスを導出する。複素 GGD の対数尤度は次式となる。

$$\begin{aligned}\log\mathcal{L} &= \log\mathcal{G}_{\mathbb{C}}(c; 0, \rho, \sigma) \\ &= \log\frac{\rho^{1-\frac{2}{\rho}}}{2^{1-\frac{2}{\rho}}\pi\Gamma(2/\rho)} - 2\log\sigma - \frac{2}{\rho}\left(\frac{|c|}{\sigma}\right)^{\rho}\end{aligned}$$

$\partial\log\mathcal{L}/\partial\sigma = 0$ より，σ の最尤値は $\sigma_{\mathrm{ML}} = |c|$ である。対数尤度差 $\mathcal{D} = \log\mathcal{L}(\sigma_{\mathrm{ML}}) - \log\mathcal{L}(\sigma) \geq 0$ は

$$\begin{aligned}\mathcal{D}(c\|\sigma) &= -2\log|c| - \frac{2}{\rho} + 2\log\sigma + \frac{2}{\rho}\left(\frac{|c|}{\sigma}\right)^{\rho} \\ &= \frac{2}{\rho}\left[\left(\frac{|c|}{\sigma}\right)^{\rho} - \log\left(\frac{|c|}{\sigma}\right)^{\rho} - 1\right]\end{aligned} \tag{24}$$

となる。式 (24) は非負かつ $\sigma = |c|$ のとき唯一 0 となるため，ダイバージェンスの公理を満たす。このダイバージェンスは，IS ダイバージェンスの式 (11) と比較すると，形状母数 ρ に関する一般化となっていることが分かる。また，より一般化された類似度関数である α-β ダイバージェンス [35] の一部となっていることも明らかにされている [25]。

5.3 最適化アルゴリズム

GGDNMF の最小化すべき関数は次式である。

$$\begin{aligned}\sum_{i,j}\mathcal{D}(c_{ij}|\sigma_{ij}) = \sum_{i,j}\Bigg[&\frac{|c_{ij}|^{\rho}}{(\sum_k w_{ik}h_{kj})^{\frac{\rho}{p}}} \\ &+ \frac{\rho}{p}\log\sum_k w_{ik}h_{kj}\Bigg]\end{aligned} \tag{25}$$

ただし，定数項を除いて示している。式 (25) を最小化する w_{ik} 及び h_{kj} は補助関数法により導出でき，次式の更新式が得られる [25]。

$$w_{ik} \leftarrow w_{ik}\left[\frac{\sum_j \frac{z_{ij}}{\left(\sum_{k'} w_{ik'}h_{k'j}\right)^2}h_{kj}}{\sum_j \frac{1}{\sum_{k'} w_{ik'}h_{k'j}}h_{kj}}\right]^{\frac{p}{\rho+p}}$$

$$h_{kj} \leftarrow h_{kj}\left[\frac{\sum_i \frac{z_{ij}}{\left(\sum_{k'} w_{ik'}h_{k'j}\right)^2}w_{ik}}{\sum_i \frac{1}{\sum_{k'} w_{ik'}h_{k'j}}w_{ik}}\right]^{\frac{p}{\rho+p}}$$

$$z_{ij} = \left(|c_{ij}|^{\frac{\rho}{p}}\sigma_{ij}^{1-\frac{\rho}{p}}\right)^{p}$$

5.4 スパース雑音除去への応用

複素 GGD における $\rho < 2$ のようにヘビーテイルな分布に基づく NMF の応用例として，信号に足しあわされたスパース雑音（時間周波数領域でスパースに分布する成分）の除去が挙げられる。実験に用いた原信号と雑音を付与した観測信号を **Fig. 6** に示す。このようなスパース雑音は，スペクトル減算法等の非線形な信号処理を適用した際に生じる人工的な成分であり，ミュージカルノイズと呼ばれている。この観測信号 (Fig. 6(b)) に対

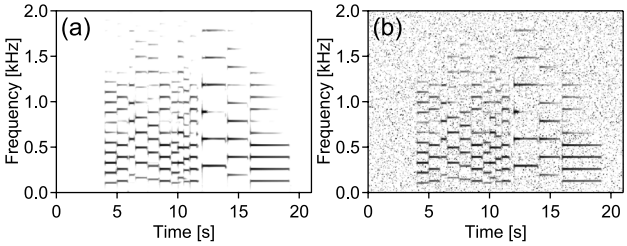

Fig. 6 Power spectrograms of (a) original and (b) observed noisy signals.

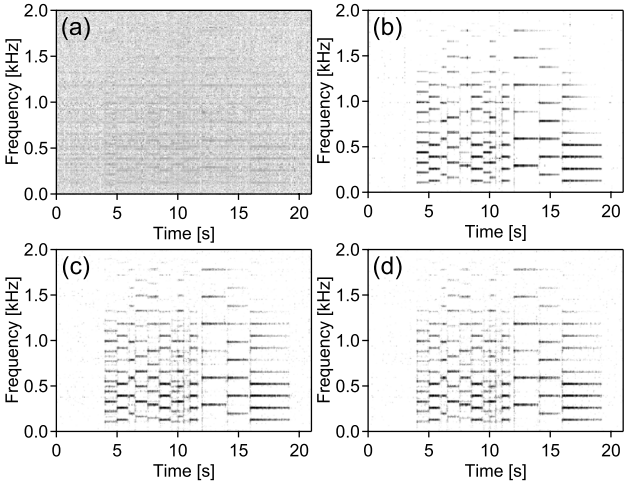

Fig. 7 Examples of estimated power spectrogram by (a) ISNMF (SDR: −13.52 dB), (b) Cauchy NMF (SDR: 3.77 dB), (c) *t*NMF (SDR: 7.26 dB), and (d) GGDNMF (SDR: 7.38 dB).

して，パワースペクトログラムに対する ISNMF，振幅スペクトログラムに対する Cauchy NMF，任意の信号ドメイン（$|\boldsymbol{C}|^{\cdot p}$）に対する *t*NMF 及び GGDNMF を適用した結果を **Fig. 7** に示す。ここで，NMF の基底数はすべて $K = 30$ としている。また，*t*NMF は $\nu = 2$，$p = 0.5$ とし，GGDNMF は $\rho = p = 0.1$ に設定している。評価値には source-to-distortion ratio（SDR）[36] の改善量を用いた。Fig. 7 を見ると，ISNMF 以外の手法はスパース雑音が比較的高精度に除去されていることが確認できる。これは，LGM をよりヘビーテイルな分布へと拡張した Cauchy NMF，*t*NMF，及び GGDNMF が，スパース雑音を外れ値として扱いながら低ランク近似した結果である。すなわち，ヘビーテイルな分布を仮定する NMF では，低ランク性に寄与しないパワーの大きな成分を無視しながら基底 $\boldsymbol{w}_k$ と係数 $\boldsymbol{h}_k$ を推定できるため，Fig. 6(b) のような信号からも頑健に低ランク構造を抽出できる。

Fig. 8 は，Fig. 7 と同様の実験を 6 種類の信号

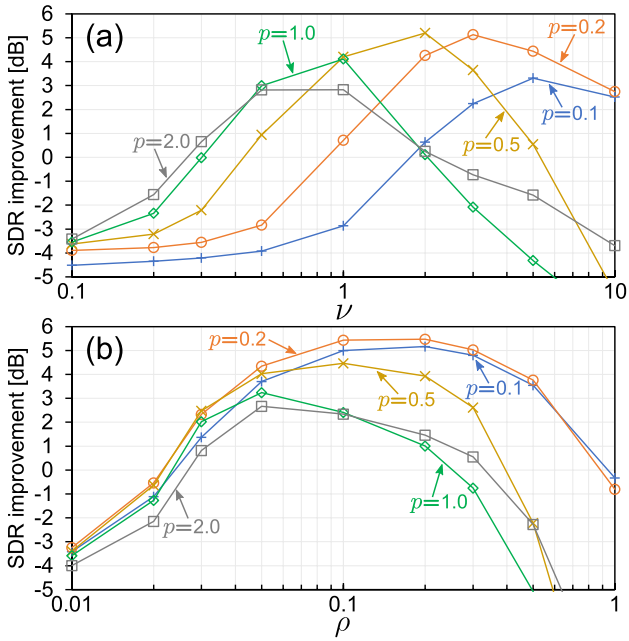

Fig. 8 Average SDR improvements of (a) *t*NMF and (b) GGDNMF for various ν, ρ, and p.

に対して行い，*t*NMF と GGDNMF に関して平均 SDR 改善量を示したグラフである。これをみると，仮定すべき最適な分布の裾の重さ（各線における最高性能の横軸位置）が観測信号のドメイン p に応じて変化していく様子が分かる。これは，p の増加に伴って値の大きなスパース雑音成分が強調され，外れ値としての影響が強くなるためである。SDR 改善量の観点では，*t*NMF と GGDNMF は同程度の雑音除去性能を示している。

6. おわりに

本稿では，音響信号に NMF を適用する際の問題点として，位相スペクトログラムの影響が無視される問題について説明し，その解決法として NMF の複素数拡張である CNMF を紹介した。また，パワースペクトログラムを ISNMF で分解することが LGM に基づく最尤推定と等価である解釈を示し，NMF で仮定される非負スペクトログラムの分解の正当化が可能であることも紹介した。最後に，LGM の理論拡張である Cauchy NMF，*t*NMF，及び GGDNMF について，スパース雑音除去への適用例を示した。

謝　辞

本研究の一部は SECOM 科学技術支援財団，ヤマハ株式会社，及び JSPS 科研費 17H06572 の助成を受けた。

文　献

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, **401**(6755), 788–791 (1999).

[2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization" *Proc. NIPS*, pp. 556–562 (2000).

[3] C. Févotte, N. Bertin and J.-L. Durrieu, "Non-negative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Comput.*, **21**, 793–830 (2009).

[4] D. Kitamura and N. Ono, "Efficient initialization for nonnegative matrix factorization based on nonnegative independent component analysis," *Proc. IWAENC* (2016).

[5] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β-divergence," *Proc. MLSP*, pp. 283–288 (2010).

[6] C. Févotte and J. Idier, "Algorithms for non-negative matrix factorization with the β-divergence," *Neural Comput.*, **23**, 2421–2456 (2011).

[7] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. Audio Speech Lang. Process.*, **15**, 1066–1074 (2007).

[8] P. Smaragdis, B. Raj and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," *Proc. ICA*, pp. 414–421 (2007).

[9] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino and S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," *Proc. ICASSP*, pp. 5365–5368 (2012).

[10] 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. Fundam.*, **E97-A**, 1113–1118 (2014).

[11] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE/ACM Trans. Audio Speech Lang. Process.*, **23**, 654–669 (2015).

[12] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," *Proc. WASPAA*, pp. 177–180 (2003).

[13] S. A. Raczyński, N. Ono and S. Sagayama, "Multipitch analysis with harmonic nonnegative matrix approximation," *Proc. ISMIR*, pp. 381–386 (2007).

[14] N. Bertin, R. Badeau and E. Vincent, "Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription," *IEEE Trans. Audio Speech Lang. Process.*, **18**, 538–549 (2010),

[15] T. Heittola, A. Mesaros, T. Virtanen and A. Eronen, "Sound event detection in multisource environments using source separation" *Proc. CHiME*, pp. 36–40 (2011).

[16] D. Bansal, B. Raj and P. Smaragdis, "Bandwidth expansion of narrowband speech using non-negative matrix factorization," *Proc. Interspeech*, pp. 1505–1508 (2005).

[17] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, **32**, 236–243 (1984).

[18] J. Le Roux, H. Kameoka, N. Ono and S. Sagayama, "Fast signal reconstruction from magnitude STFT spectrogram based on spectrogram consistency," *Proc. DAFx*, pp. 397–403 (2010).

[19] 矢田部浩平, 升山義紀, 草野　翼, 及川靖広, "位相変換による複素スペクトログラムの表現," 音響学会誌, **75**, 147–155 (2019).

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

[21] H. Kameoka, H. Kagami and M. Yukawa, "Complex NMF with the generalized Kullback-Leibler divergence," *Proc. ICASSP*, pp. 56–60 (2017).

[22] P. Magron and T. Virtanen, "Towards complex nonnegative matrix factorization with the beta-divergence," *Proc. IWAENC*, pp. 156–160 (2018).

[23] A. Liutkus, D. Fitzgerald and R. Badeau, "Cauchy nonnegative matrix factorization," *Proc. WASPAA* (2015).

[24] K. Yoshii, K. Itoyama and M. Goto, "Student's t nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation," *Proc. ICASSP*, pp. 51–55 (2016).

[25] 北村大地, 高宗典玄, 最上伸一, 三井祥幹, 猿渡　洋, 高橋　祐, 近藤多伸, "ヘビーテイルな分布に基づく非負値行列因子分解を用いたスパース雑音除去," 音講論集, pp. 441–444 (2018).

[26] P. Magron, R. Badeau and B. David, "Phase recovery in NMF for audio source separation: An insightful benchmark," *Proc. ICASSP*, pp. 81–85 (2015).

[27] A. Ozerov, E. Vincent and F. Bimbot, "A general flexible framework for the handling of prior information in audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, **20**, 1118–1133 (2012).

[28] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, **18**, 550–563 (2010).

[29] H. Sawada, H. Kameoka, S. Araki and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Trans. Audio Speech Lang. Process.*, **21**, 971–982 (2013).

[30] 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.*, **24**, 1626–1641 (2016).

[31] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," *Audio Source Separation*, S. Makino, Ed. (Springer, Cham, 2018), pp. 125–155.

[32] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance* (Chapman & Hall/CRC Press, Boca Raton, 1994).

[33] K. Kitamura, Y. Bando, K. Itoyama and K. Yoshii, "Student's t multichannel nonnegative matrix factorization for blind source separation," *Proc.*

IWAENC (2016).
[34] S. Mogami, D. Kitamura, Y. Mitsui, N. Takamune, H. Saruwatari and N. Ono, "Independent low-rank matrix analysis based on complex Student's t-distribution for blind audio source separation," *Proc. MLSP* (2017).
[35] A. Cichocki, S. Cruces and S. Amari, "Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization," *Entropy*, **13**, 134–170 (2011).
[36] E. Vincent, R. Gribonval and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech Lang. Process.*, **14**, 1462–1469 (2006).

北村　大地

香川高等専門学校専攻科修了 (2012)。奈良先端科学技術大学院大学博士前期課程修了 (2014)。総合研究大学院大学博士後期課程修了 (2017)。博士（情報学）。現在，香川高等専門学校助教。主として統計的信号処理，アレイ信号処理，機械学習，音源分離の研究に従事。日本音響学会学生・若手フォーラム幹事会員（2014–現在）。日本音響学会粟屋潔学術奨励賞 (2015)，日本学術振興会育志賞 (2017)，IEEE SPS Japan Best Paper Award (2017)，電気通信普及財団テレコムシステム技術賞奨励賞 (2018)，日本音響学会独創研究奨励賞板倉記念 (2018) 等受賞。