Vol.2015-MUS-108 No.2
2015/8/31

非ガウス性モノラル音響信号に対する音源分離のための非負値行列分解と半正定値テンソル分解

吉井 和佳[1,a)]　糸山 克寿[1,b)]　後藤 真孝[2,c)]

概要：本稿では，非ガウス性モノラル音響信号に対して音源分離を行うのに適した，非負値行列分解 (NMF) と半正定値テンソル分解 (PSDTF) の新しい確率モデルを提案する．従来，複素スペクトルは平均を 0 とする複素ガウス分布に従うことを仮定するのが一般的であった．このとき，各フレームにおける複素ガウス分布の分散行列（半正定値行列）を少数の基底分散行列（半正定値行列）の錐結合で表現するのが Log-Determinant ダイバージェンスに基づく PSDTF (LD-PSDTF) であり，半正定値行列を対角行列（対角成分はパワースペクトル密度なので非負ベクトル）に限定すると板倉・斎藤ダイバージェンスに基づく NMF (IS-NMF) に帰着する．しかし，実際の音響信号は非ガウス性（多くの場合優ガウス性）を持つため，ガウス分布より裾の重い確率分布を尤度関数に用いることが望ましい．本研究では，複素ガウス分布を極限形式として含む複素ガウススケール混合 (GSM) 分布を尤度関数とした PSDTF および NMF を提案する．具体的には，複素 GSM 分布のうちで複素 t 分布（特殊形として複素コーシー分布を含む）を尤度関数とした t-PSDTF を定式化し，基底とアクティベーションを同時に最尤推定するための効率的な乗法更新アルゴリズムを導出する．実験の結果，高品質な音源分離ができることを確かめた．

1. はじめに

音楽情報処理分野において，多重音に対する音源分離における有用性から，非負値行列分解 (nonnegative matrix factorization: NMF) は大きな注目を集めている [1–4]. NMF では，多重音の振幅あるいはパワースペクトログラム（非負値行列）を，周波数方向の基底スペクトルの集合（非負値行列）と各基底スペクトルに対応する時間方向の音量変化の集合（非負値行列）との積に分解を行う．このとき，観測スペクトログラムと，得られた 2 つの非負値行列の積である再構成スペクトログラムとの誤差を表すコスト関数を最小化するため，効率的な乗法更新アルゴリズムが導出されている [5]．また，コスト関数の最小化は，ある確率モデルの最尤推定（対数尤度の最大化）としても解釈可能である [6]．NMF の結果を用いれば，多重音の複素スペクトログラムに対して，時間周波数ビンごとに独立にウィナーフィルタを適用することで，音源スペクトログラムを求めることができる（位相は同一となる）.

1　京都大学 大学院情報学研究科 知能情報学専攻
Yoshida-honmachi, Sakyo, Kyoto, Kyoto 606-8501, Japan
2　産業技術総合研究所 情報技術研究部門
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
a)　yoshii(at)i.kyoto-u.ac.jp
b)　itoyama(at)kuis.kyoto-u.ac.jp
c)　m.goto(at)aist.go.jp

我々は以前，NMF の自然な数学的拡張となる，半正定値テンソル分解 (positive semidefinite tensor factorization: PSDTF) [7] と呼ぶ新しい因子分解法を提案した．NMF と同様，PSDTF もコスト関数の違いにより複数の変種が考えられるが，Log-Determinant ダイバージェンスに基づく PSDTF (LD-PSDTF) が音響信号解析に有効である．まず，連続時間上の入力音響信号は局所的に定常なガウス過程に従うことを仮定する．これは，フーリエ変換は線形変換であるから，各フレームの複素スペクトルは連続周波数上の複素ガウス過程に従うことを仮定することと等価である．複素ガウス過程の任意の周辺分布は複素ガウス分布になることから，離散周波数上の複素スペクトルの尤度は複素ガウス分布で与えられる．このとき，各フレームにおける複素ガウス分布の分散行列（半正定値行列）を少数の基底分散行列（半正定値行列）の錐結合で表現し，それら基底行列と非負の重みを同時に最尤推定するのが LD-PSDTF である．ただし，半正定値行列を対角行列（対角成分はパワースペクトル密度なので非負ベクトル）に限定すると板倉・斎藤ダイバージェンスに基づく NMF (IS-NMF) に帰着する．LD-PSDTF の結果を用いれば，多重音の複素スペクトログラムに対して，周波数ビン間の相関を考慮したウィナーフィルタを適用することで，適切な位相をもつ音源スペクトログラムを推定可能となった.

IS-NMF や LD-PSDTF において，多重音である入力音響信号がガウス過程（離散時間上ではガウス分布）に従うという仮定にはある程度妥当性が認められる．各音源信号はガウス過程に従わない場合でも，中心極限定理により，十分に多くの音源信号が重畳することで得られる音響信号はガウス過程に従うとみなせる．実際，単独発話の音声信号はガウス分布よりもラプラス分布などの優ガウス分布でよく近似できることが知られており，独立成分分析では，そのような各音源信号の非ガウス性が，ガウス分布に従う観測信号を分離するときの手がかりとなる [8].

ただし，音源信号の性質によっては，通常の中心極限定理が成立せず，多重音である入力音響信号がガウス分布に従わない場合が存在する．例えば，音源信号がコーシー分布（裾が非常に重く，平均も分散も定義されない）に従う場合は，そのような音源信号が多数重畳しても依然としてコーシー分布に従う [9]．ガウス分布もコーシー分布も安定分布 (stable distribution) の一種であり，それぞれ安定分布の特性指数が 1 および 2 のときの特殊な場合である．安定分布に従う確率変数の和は，同じ特性指数をもつ安定分布に従うという性質があり，音源信号の重畳過程に対する生成モデルを定式化するうえで都合がよい．

本研究では，入力音響信号がガウス分布より裾の重い分布であるガウススケール混合 (Gaussian scale mixture: GSM) 分布に従うと仮定した場合の PSDTF（およびその特殊形としての NMF）の定式化に取り組む．多変量ガウス分布と同様に，多変量 GSM 分布は任意の周辺分布が多変量 GSM 分布となるため，連続時間上の音響信号に対する確率過程として，ガウス過程の代わりに GSM 過程を考えることができる．すなわち，連続周波数上では，複素スペクトルが複素 GSM 過程に従うことを仮定するのと等価である．本稿では，まず，GSM 分布の中でもとりわけ重要性の高い t 分布に着目する．ガウス分布は自由度が無限の t 分布であり，コーシー分布は自由度が 1 の t 分布に相当する．t 分布に基づく確率モデルの最尤推定には，LD-PSDTF に対する乗法更新則を内包する，t-PSDTF のための乗法更新則を導出することができる．

提案手法では，入力音響信号中に含まれる外れ値に対して頑健な低ランク分解が可能になる．NMF や PSDTF の基本的な考え方は，対象となるデータをできるだけ少数の意味のある基底の組み合わせで表現しようとするものであるから，入力データ中に突発的な雑音が含まれていたとしても，それを表現するためにわざわざ新たな基底を準備することは得策ではない．これまでに，回帰問題において，ガウス過程の代わりに t 過程を用いたり [10]，行列分解問題において，雑音がガウス分布ではなく t 分布やラプラス分布に従うと仮定することで [11,12]，外れ値に頑健なモデルが学習できることが報告されている．音楽音響信号には突発的な外れ値が多く，頑健性向上が期待できる．

2. 複素ガウススケール混合分布

本章では，NMF と PSDTF に対する新しい確率モデルを導出するうえで必要なガウススケール混合 (Gaussian scale mixutre: GSM) 分布について説明する．GSM 分布とは，ガウス分布の分散行列のスケールに対してある事前分布を考え，スケールを積分消去することで得られる複合分布 (compound distribution) のことをさす [13]．もとになる分布を多変量複素ガウス分布とすれば，得られる複合分布は多変量複素 GSM 分布となる．一般に，GSM 分布はガウス分布より裾が重く (heavy-tailed)，外れ値に頑健な性質を持ち，具体例として t 分布や K 分布などが知られている．コーシー分布は t 分布の，ラプラス分布は K 分布の特殊な場合である．ガウス分布は，GSM 分布（あるいは t 分布や K 分布）の極限形式としての特殊形である．

複素 GSM 分布は，複素楕円対称 (complex elliptically symmetric: CES) 分布 [13] と呼ばれる，実用上重要なクラスの分布のうちで代表的なものである．そのため，CES 分布がもつ好ましい性質を多く引き継いでおり，周辺分布や条件付き分布が解析的に計算可能である点で都合がよい．GSM 分布を用いると，非ガウス分布およびガウス分布の統一的な取り扱いが可能になる．

2.1 分布の導出

多変量 CGSM 分布の導出について説明する．まず，多変量複素ガウス分布に従う確率変数 $\boldsymbol{x} \in \mathbb{C}^d$ を考える．

$$\boldsymbol{x} \mid \boldsymbol{\Sigma} \sim \mathcal{N}_c(\boldsymbol{x}|\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{1}$$

ここで，$\boldsymbol{\Sigma} \in \mathbb{C}^{d\times d}$ は半正定値行列である．次に，$\boldsymbol{x}$ のアフィン変換で得られる新たな確率変数 $\boldsymbol{z} \in \mathbb{C}^{d\times d}$ を考える．

$$\boldsymbol{z} = \boldsymbol{\mu} + \sqrt{\tau}\boldsymbol{x} \tag{2}$$

ここで，$\boldsymbol{\mu} \in \mathbb{C}^{d\times d}$ は移動パラメータであり，$\tau \geq 0$ はスケールパラメータである．このとき，ガウス分布は確率変数のアフィン変換に対して閉じていることから，$\boldsymbol{z}$ も多変量複素ガウス分布に従う．

$$\boldsymbol{z} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, \tau \sim \mathcal{N}_c(\boldsymbol{z}|\boldsymbol{\mu}, \tau\boldsymbol{\Sigma}) \tag{3}$$

ここで，分散行列のスケールである τ に対して，任意の事前分布（ガンマ分布，逆ガンマ分布，逆ガウス分布，一般化逆ガウス分布など）を考える．

$$\tau \sim p(\tau|\boldsymbol{\theta}) \tag{4}$$

ここで，$\boldsymbol{\theta}$ は事前分布のパラメータである．ただし，スケールの任意性を解消するため，$\mathbb{E}[\tau] = 1$ とする．この事前分布のもとで，式 (3) における τ を積分消去を行う．

$$p(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\theta}) = \int_0^\infty \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}, \tau\boldsymbol{\Sigma}) p(\tau|\boldsymbol{\theta}) d\tau \tag{5}$$

ここで，$p(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\theta})$ が多変量複素 GSM 分布の確率密度関数となる．ただし，スケール τ の事前分布がデルタ分布である（τ の値を一意に定める）場合には，式 (3) で与えられる多変量複素ガウス分布が得られることから，多変量複素 GSM 分布は多変量複素ガウス分布を極限形式として含むことが分かる．

多変量複素 GSM 分布の具体的な確率密度関数を数式で表す際には，量 $Q(\boldsymbol{z})$ を定義しておくと便利である．

$$Q(\boldsymbol{z}) = (\boldsymbol{z}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\boldsymbol{z}-\boldsymbol{\mu}) \tag{6}$$

これは，平均ベクトルが μ で，共分散行列が $\boldsymbol{\Sigma}$ であるようなベクトル $\boldsymbol{z}$ で表される一群の値に対するマハラノビス距離の二乗となっている．

2.2 アフィン変換・周辺分布・条件付き分布

多変量複素 GSM 分布は多変量複素ガウス分布と同じくいくつかの好ましい性質を持つ．まず，多変量複素 GSM 分布はアフィン変換に関して閉じている．すなわち，ある多変量複素 GSM 分布に従う確率変数 $\boldsymbol{z} \in \mathbb{C}^d$ を

$$\boldsymbol{z} \sim \mathrm{GSM}_c(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\theta}) \tag{7}$$

とする．ここで，$\boldsymbol{\mu}$ は位置ベクトル，$\boldsymbol{\Sigma}$ はスケール行列，$\boldsymbol{\theta}$ はその他のパラメータとする．このとき，任意の $\boldsymbol{A} \in \mathbb{C}^{m\times d}$ および $\boldsymbol{b} \in \mathbb{C}^m$ に対して以下が成立する．

$$\boldsymbol{A}\boldsymbol{z}+\boldsymbol{b} \sim \mathrm{GSM}_c(\boldsymbol{A}\boldsymbol{z}+\boldsymbol{b}|\boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b},\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^H,\boldsymbol{\theta}) \tag{8}$$

すなわち，確率分布のパラメータは $\boldsymbol{\mu} \to \boldsymbol{A}\boldsymbol{\mu}+\boldsymbol{b}$ および $\boldsymbol{\Sigma} \to \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^H$ のように変換される．

また，多変量複素 GSM 分布の周辺分布・条件付き分布はいずれも多変量複素 GSM 分布となる．まず，式 (7) において，確率変数 $\boldsymbol{z} \in \mathbb{C}^d$ を 2 つの部分 $\boldsymbol{z}_1 \in \mathbb{C}^{d_1}$ および $\boldsymbol{z}_2 \in \mathbb{C}^{d_2}$ に分割し（ただし $d = d_1 + d_2$），位置ベクトルやスケール行列も対応するように分割しておく．

$$\begin{pmatrix}\boldsymbol{z}_1\\ \boldsymbol{z}_2\end{pmatrix} \sim \mathrm{GSM}_c\left(\begin{pmatrix}\boldsymbol{z}_1\\ \boldsymbol{z}_2\end{pmatrix}\middle|\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix},\begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix},\boldsymbol{\theta}\right) \tag{9}$$

このとき，$\boldsymbol{z}_1$ および $\boldsymbol{z}_2$ の周辺分布は以下で与えられる．

$$\boldsymbol{z}_1 \sim \mathrm{GSM}_c(\boldsymbol{z}_1|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_{11},\boldsymbol{\theta}) \tag{10}$$

$$\boldsymbol{z}_2 \sim \mathrm{GSM}_c(\boldsymbol{z}_2|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_{22},\boldsymbol{\theta}) \tag{11}$$

一方，$\boldsymbol{z}_2$ が与えられた場合の $\boldsymbol{z}_1$ の条件付き分布は以下で与えられる．

$$\boldsymbol{z}_1 \mid \boldsymbol{z}_2 \sim \mathrm{GSM}_c(\boldsymbol{z}_1|\boldsymbol{\mu}_{1|2},\boldsymbol{\Sigma}_{1|2},\boldsymbol{\theta}_{1|2}) \tag{12}$$

ここで，$\boldsymbol{\theta}_{1|2}$ は式 (4) で与えられる事前分布に依存して決まる量であるが，位置ベクトル $\boldsymbol{\mu}_{1|2}$ およびスケール行列 $\boldsymbol{\Sigma}_{1|2}$ は常に以下で与えられる．

$$\boldsymbol{\mu}_{1|2} = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{z}_2-\boldsymbol{\mu}_2) \tag{13}$$

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} \tag{14}$$

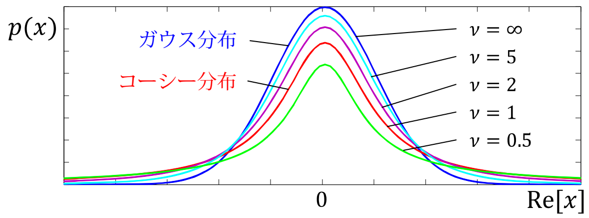

図 1 実数軸でスライスした（虚部を周辺化した）一変量複素 t 分布

2.3 多変量複素 t 分布

式 (4) の事前分布に逆ガンマ分布を与えた場合に，式 (5) で得られる多変量複素 GSM 分布は多変量複素 t 分布となる．参考のため標準 t 分布の確率密度関数を**図 1** に示す．具体的には，$\lambda = \tau^{-1}$ にガンマ事前分布を仮定する．

$$\lambda \sim \mathcal{G}\left(\lambda\middle|\frac{\nu}{2},\frac{\nu}{2}\right) \tag{15}$$

ここで，$\mathcal{G}(a,b)$ は形状パラメータ a，逆スケールパラメータ b をもつガンマ分布を表す．このとき，式 (5) を具体的に計算すると以下を得る．

$$\begin{aligned} p(\boldsymbol{z}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) &= \int \mathcal{N}_c(\boldsymbol{z}|\boldsymbol{\mu},\lambda^{-1}\boldsymbol{\Sigma})\mathcal{G}\left(\lambda\middle|\frac{\nu}{2},\frac{\nu}{2}\right)d\lambda \\ &= \frac{\Gamma(\frac{2d+\nu}{2})}{\Gamma(\frac{\nu}{2})}\frac{2^d}{(\pi\nu)^d}|\boldsymbol{\Sigma}|^{-1}\left(1+\frac{2}{\nu}Q(\boldsymbol{z})\right)^{-\frac{2d+\nu}{2}} \end{aligned} \tag{16}$$

ここで，$\boldsymbol{\mu} \in \mathbb{C}^d$ は位置ベクトル，$\boldsymbol{\Sigma} \in \mathbb{C}^{d\times d}$ はスケール行列，$\nu > 0$ は自由度である．ただし，$\nu < 1$ では平均と分散が，$\nu < 2$ では分散が定義されない（無限大に発散する）ため，$\nu < 2$ の多変量複素 t 分布に従う確率変数に対しては通常の中心極限定理は成立しないことに注意する．

多変量複素 t 分布は，自由度が $\nu = 1$ のとき多変量複素コーシー分布，$\nu \to \infty$ のとき多変量複素ガウス分布となり，いずれも再生性を有する．すなわち，確率変数 $\boldsymbol{z}_1$ および $\boldsymbol{z}_2$ が以下の多変量複素ガウス分布

$$\boldsymbol{z}_1 \sim \mathcal{N}_c(\boldsymbol{z}_1|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) \tag{17}$$

$$\boldsymbol{z}_2 \sim \mathcal{N}_c(\boldsymbol{z}_2|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2) \tag{18}$$

に従うとすると，確率変数の和も多変量複素ガウス分布

$$\boldsymbol{z}_1+\boldsymbol{z}_2 \sim \mathcal{N}_c(\boldsymbol{z}_1+\boldsymbol{z}_2|\boldsymbol{\mu}_1+\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2) \tag{19}$$

に従う．同様のことが多変量複素コーシー分布についても成立する．すなわち，確率変数 $\boldsymbol{z}_1$ および $\boldsymbol{z}_2$ が以下の多変量複素コーシー分布

$$\boldsymbol{z}_1 \sim \mathcal{C}_c(\boldsymbol{z}_1|\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) \tag{20}$$

$$\boldsymbol{z}_2 \sim \mathcal{C}_c(\boldsymbol{z}_2|\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2) \tag{21}$$

に従うとすると，確率変数の和も多変量複素コーシー分布

$$\boldsymbol{z}_1+\boldsymbol{z}_2 \sim \mathcal{C}_c(\boldsymbol{z}_1+\boldsymbol{z}_2|\boldsymbol{\mu}_1+\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_1+\boldsymbol{\Sigma}_2) \tag{22}$$

に従う．t 分布のうちで再生性を有するのはガウス分布とコーシー分布のみであり，一般の自由度 ν に対しては再生性は成立しないことに注意する．

Vol.2015-MUS-108 No.2
2015/8/31

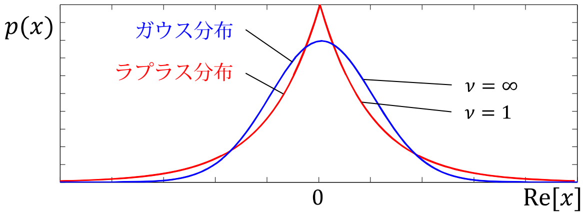

図 2 実数軸でスライスした（虚部を周辺化した）一変量複素 K 分布

2.4 多変量複素 K 分布

式 (4) の事前分布にガンマ分布を与えた場合に，式 (5) で得られる多変量複素 GSM 分布は多変量複素 K 分布となる．参考のため標準 K 分布の確率密度関数を**図 2** に示す．具体的には，τ にガンマ事前分布を仮定する．

$$\tau \sim \mathcal{G}(\tau|\nu,\nu) \tag{23}$$

このとき，式 (5) を具体的に計算すると以下を得る．

$$\mathcal{K}_c(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \int \mathcal{N}_c(\boldsymbol{z}|\boldsymbol{\mu},\tau\boldsymbol{\Sigma})\mathcal{G}(\tau|\nu,\nu)\,d\tau \tag{24}$$

$$= \frac{2\nu^d}{\Gamma(\nu)\pi^d}|\boldsymbol{\Sigma}|^{-1}\left(\sqrt{\nu Q(\boldsymbol{x})}\right)^{\nu-d} K_{\nu-d}\left(2\sqrt{\nu Q(\boldsymbol{x})}\right)$$

ここで，$\boldsymbol{\mu} \in \mathbb{C}^d$ は位置ベクトル，$\boldsymbol{\Sigma} \in \mathbb{C}^{d\times d}$ はスケール行列，$\nu > 0$ は自由度である．多変量複素 t 分布と異なり，任意の ν について平均と分散が定義できる．また，多変量複素 K 分布は，自由度が $\nu = 1$ のとき多変量複素ラプラス分布，$\nu \to \infty$ のとき多変量複素ガウス分布となるが，再生性を持つのは $\nu \to \infty$ のときのみである．

3. 複素ガウス分布に基づく非負値行列分解と半正定値テンソル分解

本章では，モノラル音響信号の音源分離の観点から，半正定値テンソル分解 (PSDTF) [7, 14] のための新しい確率モデルを提案する．まず，コスト関数最小化の立場から，NMF と PSDTF の一般的な定式化を説明し，PSDTF が NMF の自然な数学的拡張になっていることを示す．NMF と同様に PSDTF にも様々なコスト関数が考えうるが，Log-Determinant (LD) ダイバージェンスに基づく PSDTF (LD-PSDTF) は，板倉・斉藤 (IS) ダイバージェンスに基づく NMF (IS-NMF) の数学的拡張となっている．ここで，コスト関数の最小化は，多変量複素ガウス分布で与えられる尤度関数の最大化（確率モデルの最尤推定）と等価であることを示す．

3.1 コスト関数最小化としての IS-NMF

NMF は，非負値行列 $\boldsymbol{X} = [\boldsymbol{x}_1,\cdots,\boldsymbol{x}_N] \in \mathbb{R}_+^{M\times N}$ に対して，二つの非負値行列 $\boldsymbol{W} = [\boldsymbol{w}_1,\cdots,\boldsymbol{w}_K] \in \mathbb{R}_+^{M\times K}$ および $\boldsymbol{H} = [\boldsymbol{h}_1,\cdots,\boldsymbol{h}_K] \in \mathbb{R}_+^{N\times K}$ の積 $\boldsymbol{Y} = [\boldsymbol{y}_1,\cdots,\boldsymbol{y}_N] \in \mathbb{R}_+^{M\times N}$ で近似する（**図 3**）．ただし，$K \ll \min(M,N)$ は予め与えるものとし，$\boldsymbol{X}$ を低ランク行列 $\boldsymbol{Y}$ で近似する．

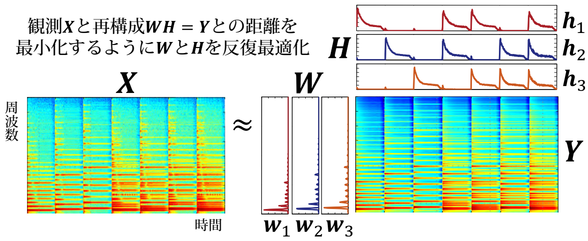

図 3 IS-NMF による混合音スペクトログラムの低ランク分解

$$\boldsymbol{X} \approx \boldsymbol{W} \times_2 \boldsymbol{H} = \boldsymbol{W}\boldsymbol{H}^T \stackrel{\text{def}}{=} \boldsymbol{Y} \tag{25}$$

ここで，$\times_2$ は 2 モード積 [15] であり，$\boldsymbol{w}_k \in \mathbb{R}_+^M$ および $\boldsymbol{h}_k \in \mathbb{R}_+^N$ はそれぞれ基底ベクトルおよび対応するアクティベーションベクトルである．式 (25) は次式で書き直せる．

$$\boldsymbol{x}_n \approx \sum_{k=1}^{K} h_{kn}\boldsymbol{w}_k \stackrel{\text{def}}{=} \boldsymbol{y}_n \tag{26}$$

すなわち，NMF の目標は，**非負ベクトルを少数の非負ベクトルの錐結合で近似すること**である．

観測ベクトル $\boldsymbol{x}_n$ と再構成ベクトル $\boldsymbol{y}_n$ との間の誤差 $\mathcal{D}(\boldsymbol{x}|\boldsymbol{y})$ を評価する尺度として，Kullback-Leibler (KL) ダイバージェンス [5] や板倉・斉藤 (IS) ダイバージェンス [16] がよく用いられる．本稿では，特に後者に着目する．

$$\mathcal{D}_{\text{IS}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_{m=1}^{M}\left(\frac{x_{nm}}{y_{nm}} - \log\frac{x_{nm}}{y_{nm}} - 1\right) \tag{27}$$

この値は常に非負であり，$\boldsymbol{x}_n = \boldsymbol{y}_n$ のときのみ 0 をとる．通常の距離尺度と異なり，対称性が成り立たない，すなわち $\mathcal{D}_{\text{IS}}(\boldsymbol{x}_n|\boldsymbol{y}_n) \neq \mathcal{D}_{\text{IS}}(\boldsymbol{y}_n|\boldsymbol{n}_n)$ であることに注意する．

3.2 コスト関数最小化としての LD-PSDTF

PSDTF は，半正定値行列 [*1] の集合（テンソル）$\boldsymbol{\mathcal{X}} = [\boldsymbol{X}_1,\cdots,\boldsymbol{X}_N] \in \mathbb{C}^{M\times M\times N}$ に対して，少数の半正定値行列の集合（テンソル）$\boldsymbol{\mathcal{W}} = [\boldsymbol{W}_1,\cdots,\boldsymbol{W}_K] \in \mathbb{C}^{M\times M\times K}$ および $\boldsymbol{H} = [\boldsymbol{h}_1,\cdots,\boldsymbol{h}_K]^T \in \mathbb{R}_+^{K\times N}$ の積 $\boldsymbol{\mathcal{Y}} = [\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_N] \in \mathbb{C}^{M\times M\times N}$ で近似する．

$$\boldsymbol{\mathcal{X}} \approx \boldsymbol{\mathcal{W}} \times_3 \boldsymbol{H} \stackrel{\text{def}}{=} \boldsymbol{\mathcal{Y}} \tag{28}$$

ここで，$\times_3$ は 3 モード積 [15] であり，$\boldsymbol{W}_k \in \mathbb{C}^{M\times M} \succ \boldsymbol{0}$ および $\boldsymbol{h}_k \in \mathbb{R}_+^N$ はそれぞれ基底行列およびアクティベーションベクトルである．式 (28) は次式で書き直せる．

$$\boldsymbol{X}_n \approx \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k \stackrel{\text{def}}{=} \boldsymbol{Y}_n \tag{29}$$

すなわち，PSDTF の目標は，**半正定値行列を少数の半正定値ベクトルの錐結合で近似すること**である．

[*1] 半正定値行列と非負値行列とは異なる概念であることに注意．あるエルミート行列 $\boldsymbol{A} = \boldsymbol{A}^H \in \mathbb{R}^{M\times M}$ が半正定値行列であるとは，$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{V}^H$ となる行列 $\boldsymbol{V}$ が存在すること，あるいは，$\boldsymbol{A}$ の M 個の固有値が全て非負であることと同値である．

Vol.2015-MUS-108 No.2
2015/8/31

観測行列 $\boldsymbol{X}_n$ と再構成行列 $\boldsymbol{Y}_n$ との間の誤差 $\mathcal{D}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を評価する尺度として，von-Neumann ダイバージェンス [17] や Log-Determinant (LD) ダイバージェンス [18] が考えられる．本稿では，特に後者に着目する．

$$\mathcal{D}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right) - \log\left|\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right| - M \quad (30)$$

この値は常に非負であり，$\boldsymbol{x}_n = \boldsymbol{y}_n$ のときのみ 0 をとる．

式 (25) と式 (28)，式 (26) と式 (29)，式 (27) と式 (30) を比較すると，PSDTF は NMF の自然な数学的拡張となっていることが分かる（**図 4**）．ここで，半正定値行列を対角行列に限定すると，対角成分には固有値が並び，それらはすべて非負であることから，$\boldsymbol{X}_n = \mathrm{diag}(\boldsymbol{x}_n)$，$\boldsymbol{Y}_n = \mathrm{diag}(\boldsymbol{y}_n)$，$\boldsymbol{W}_k = \mathrm{diag}(\boldsymbol{w}_k)$ とすれば *2，PSDTF は NMF に帰着する．すなわち，NMF では PSDTF のように，M 個の要素間の相関を考慮することができない．

3.3 尤度最大化としての LD-PSDTF

モノラル音響信号の音源分離を題材として，確率モデルの最尤推定の観点から LD-PSDTF を捉え直す．まず，観測される音響信号（混合音）の複素スペクトログラムを $\tilde{\boldsymbol{X}} = [\tilde{\boldsymbol{x}}_1, \cdots, \tilde{\boldsymbol{x}}_N] \in \mathbb{C}^{M\times N}$，$k$ 番目の音源信号の複素スペクトログラムを $\tilde{\boldsymbol{X}}_k = [\tilde{\boldsymbol{x}}_{k1}, \cdots, \tilde{\boldsymbol{x}}_{kN}] \in \mathbb{C}^{M\times N}$ とする．ここで，M は周波数ビン数，N はフレーム数である．観測した混合音が K 個の音源信号の瞬時混合であると仮定すると，以下が成り立つ．

$$\tilde{\boldsymbol{X}} = \sum_{k=1}^{K} \tilde{\boldsymbol{X}}_k \;\Leftrightarrow\; \tilde{\boldsymbol{x}}_n = \sum_{k=1}^{K} \tilde{\boldsymbol{x}}_{kn} \quad (31)$$

我々の目標は，観測変数 $\tilde{\boldsymbol{X}}$ を潜在変数 $\tilde{\boldsymbol{X}}_k$ に分解することである．しかし，これは不良設定問題であるので，何らかの制約を導入して最適解を同定する必要がある．

我々は，各音源信号は局所的に定常なガウス過程に従うと仮定する．すなわち，各フレーム n における局所的な連続信号は定常であり，時不変な共分散関数（カーネル）をもつガウス過程に従うと仮定する．このとき，ガウス過程の任意の周辺分布は多変量ガウス分布になることから，あるサンプリング周波数に従ってサンプルされた離散信号は多変量ガウス分布に従う．したがって，フーリエ変換はアフィン変換であることから，各フレームにおける複素スペクトル $\tilde{\boldsymbol{x}}_{kn}$ は多変量複素ガウス分布に従う．

$$\tilde{\boldsymbol{x}}_{kn} \mid \boldsymbol{Y}_{kn} \sim \mathcal{N}_c(\tilde{\boldsymbol{x}}_{kn}|\boldsymbol{0}, \boldsymbol{Y}_{kn}) \quad (32)$$

ここで，共分散行列を $\boldsymbol{Y}_{kn} = h_{kn}\boldsymbol{W}_k$ とした．すなわち，各音源信号において，$\boldsymbol{W}_k$ で定まる基本的なダイナミクスは定常であるが，そのスケール h_{kn} のみが時変化すると考えることで，音源信号の非定常性を制限する．

*2 $\mathrm{diag}(\boldsymbol{a})$ はベクトル $\boldsymbol{a}$ を対角成分に持つ対角行列を意味する．

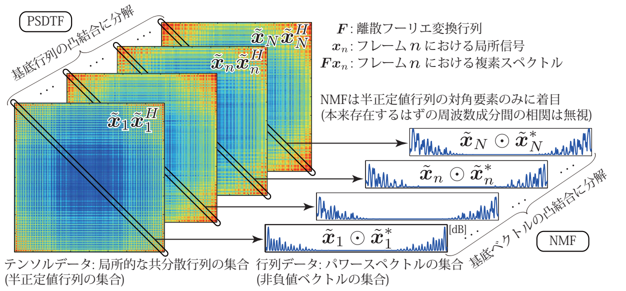

図 4　IS-NMF と LD-PSDTF の比較

このとき，音源信号が重畳して得られる観測信号も局所的に定常なガウス過程に従う．式 (31) に着目すると，複素ガウス分布の再生性から以下を得る．

$$\tilde{\boldsymbol{x}}_n \mid \boldsymbol{Y}_n \sim \mathcal{N}_c(\tilde{\boldsymbol{x}}_n|\boldsymbol{0}, \boldsymbol{Y}_n) \quad (33)$$

ただし，$\boldsymbol{Y}_n = \sum_k \boldsymbol{Y}_{kn}$ である．これが LD-PSDTF の確率モデルであり，対数尤度関数は以下で与えられる．

$$\begin{aligned}\log p(\boldsymbol{x}_n|\boldsymbol{Y}_n) &= -M\log(\pi) - \log|\boldsymbol{Y}_n| - \boldsymbol{x}_n\boldsymbol{Y}_n^{-1}\boldsymbol{x}_n \\ &\overset{c}{=} -\log|\boldsymbol{Y}_n| - \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right) \quad (34)\end{aligned}$$

ここで，$\boldsymbol{X}_n = \boldsymbol{x}_n\boldsymbol{x}_n^H \succ \boldsymbol{0}$ とした．式 (34) を式 (30) と比較すると，$\boldsymbol{X}_n$ は観測データで一定であるから，$\boldsymbol{Y}_n$ に関する対数尤度の最大化は，LD ダイバージェンスの最小化と等価であることが分かる．

3.4 乗法更新アルゴリズム

LD-PSDTF のパラメータ $\boldsymbol{\mathcal{W}}$ および $\boldsymbol{H}$ を求めるため，補助関数法に基づく収束性の保証された乗法更新アルゴリズムが提案されている [7,14]．本稿では結果のみ記すと，乗法更新則は以下で与えられる．

$$\boldsymbol{W}_k \leftarrow \boldsymbol{W}_k\boldsymbol{L}_k(\boldsymbol{L}_k^T\boldsymbol{W}_k\boldsymbol{P}_k\boldsymbol{W}_k\boldsymbol{L}_k)^{-\frac{1}{2}}\boldsymbol{L}_k^T\boldsymbol{W}_k \quad (35)$$

$$h_{kn} \leftarrow h_{kn}\left(\frac{\mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\boldsymbol{W}_k\boldsymbol{Y}_n^{-1}\right)}{\mathrm{tr}\left(\boldsymbol{W}_k\boldsymbol{Y}_n^{-1}\right)}\right)^{\frac{1}{2}} \quad (36)$$

ここで，$\boldsymbol{L}_k$ はコレスキー分解 $\boldsymbol{Q}_k = \boldsymbol{L}_k\boldsymbol{L}_k^T$ で求まる下三角行列であり，$\boldsymbol{P}_k$ および $\boldsymbol{Q}_k$ は次式で求まる．

$$\boldsymbol{P}_k = \sum_{n=1}^{N} h_{kn}\boldsymbol{Y}_n^{-1} \quad \boldsymbol{Q}_k = \sum_{n=1}^{N} h_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{X}_n\boldsymbol{Y}_n^{-1} \quad (37)$$

h_{kn} の非負性と $\boldsymbol{W}_k$ の半正定値性は自然に保たれているが，$\mathrm{tr}(\boldsymbol{W}_k) = 1$ を満たすよう，反復のたびに $\boldsymbol{W}_k$ および $\boldsymbol{h}_k$ のスケールを調整しておく．式 (35) および式 (36) において，半正定値行列を対角行列に限定すると，よく知られた IS-NMF の乗法更新アルゴリズムに帰着する [19,20]．

3.5 LD-PSDTF に基づくウィナーフィルタリング

最終的に，式 (32) および式 (33) から，$\tilde{\boldsymbol{x}}_n$ が与えられたときの $\tilde{\boldsymbol{x}}_{kn}$ の事後分布は多変量複素ガウス分布になること

が分かり，その平均と分散は次式で求めることができる．

$$\mathbb{E}[\tilde{\boldsymbol{x}}_{kn}|\tilde{\boldsymbol{x}}_n] = \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1}\tilde{\boldsymbol{x}}_n \tag{38}$$

$$\mathbb{V}[\tilde{\boldsymbol{x}}_{kn}|\tilde{\boldsymbol{x}}_n] = \boldsymbol{Y}_{kn} - \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{Y}_{kn} \tag{39}$$

このウィナーフィルタリングでは，$\tilde{\boldsymbol{X}}_k$ の位相は $\tilde{\boldsymbol{X}}$ の位相とは異なる点に注意する．各フレーム n ごとに周波数ビン間の相関を考慮しながら一挙に分離を行うことで，高品質な分離が可能となる．一方，IS-NMF では式 (38) や式 (39) における半正定値行列はすべて対角行列に制限されているため，各周波数ビン n, m ごとに独立に行われることになり，$\tilde{\boldsymbol{X}}_k$ の位相は $\tilde{\boldsymbol{X}}$ の位相と同一となる．

4. 複素ガウススケール混合分布に基づく非負値行列分解と半正定値テンソル分解

我々は，多変量複素ガウス分布の代わりに，より裾の重い多変量複素 GSM 分布を尤度関数に用いることで，外れ値に対して頑健な PSDTF（およびその特殊形である NMF）を提案する．本研究では，多変量複素 GSM 分布のなかでも，多変量複素 t 分布（多変量複素コーシー分布を特殊形として含む）を尤度関数に用いる場合に着目する．多変量複素コーシー分布は多変量複素ガウス分布と同様に再生性を持つため，多変量複素コーシー分布に基づく PSDTF は，混合音の生成過程として妥当な解釈を与えることができる．誌面の都合上，多変量複素 K 分布（多変量複素ラプラス分布を特殊形に含む）を尤度関数に用いた PSDTF については項を改めて議論したい．

4.1 複素ガウススケール混合分布の妥当性

一般に，どのような尤度関数が適切であるかは，音源信号やそれらの和である観測信号の性質によって異なる．例えば，音源信号がガウス過程に従うならば，観測信号もガウス過程に従う．また，音源信号がコーシー過程に従うならば，観測信号もコーシー過程に従う．ガウス過程もコーシー過程も t 過程の特別な場合であるが，音源信号が一般の t 過程に従う場合は，観測信号は t 過程に従うとはいえない．一般に，GSM 過程に従う音源信号の和は必ずしも同じ形の GSM 過程に従うとは限らない．中心極限定理により，理論的には，音源数が増加するほど観測信号はガウス過程に漸近する場合も考えられる．

多変量複素ガウス分布と多変量複素コーシー分布に基づく PSDTF のみが「数学的に正しい」モデルであったとしても，より広いクラスの裾の重い分布に基づく PSDTF を定式化しておくことは応用上重要である．これまで，音源信号がガウス過程に従うという仮定のもとでは，音源信号のパワースペクトル密度の加法性が成立するため，混合音のパワースペクトログラムに対して IS-NMF を適用することには一定の妥当性があることが知られていた．一方，混合音の振幅スペクトログラムに対して KL-NMF を適用するうえでは，振幅スペクトルに加法性が成立するという強い仮定を置く必要があった．しかし，実験的には，KL-NMF の方が IS-NMF より優れた音源分離結果を示すことが多い [9]．このことは，音源信号の混合過程を数学的に正しくモデル化することが，必ずしもよい分離結果につながるとは限らないことを示している．

実際の音響信号に対して適切な尤度関数を決定する問題は，モデル選択の問題として定式化できる．将来的に，パラメータに事前分布を導入すれば，エビデンスに基づくベイズ的なモデル選択も可能になると考えられる．

4.2 尤度最大化としての定式化

PSDTF を頑健化するため，式 (33) で与えられる尤度関数において，多変量複素ガウス分布の代わりに，裾の重い多変量複素 GSM 分布を用いることを考える．

$$\tilde{\boldsymbol{x}}_n \mid \boldsymbol{Y}_n \sim \mathrm{GSM}_c(\tilde{\boldsymbol{x}}_n|\boldsymbol{0}, \boldsymbol{Y}_n) \tag{40}$$

ここで，$\boldsymbol{Y}_n = \sum_{k=1}^{K} h_{kn}\boldsymbol{W}_k$ である．このとき，観測データ $\boldsymbol{\mathcal{X}}$ 全体に渡る対数尤度は次式で与えられる．

$$\log p(\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}) = \sum_{n=1}^{N} \log p(\tilde{\boldsymbol{x}}_n|\boldsymbol{Y}_n) \tag{41}$$

ここで，$p(\tilde{\boldsymbol{x}}_n|\boldsymbol{Y}_n)$ は式 (40) で与えられる各フレームごとの尤度である．この対数尤度の一般形を最大化する乗法更新則を導出することは困難であるため，以降，複素 t 分布である場合について議論する．

4.3 多変量複素 t 分布に基づく t-PSDTF

多変量複素 t 分布を尤度関数に用いる場合，最大化すべき対数尤度（式 (41) 参照）は具体的に次式で与えられる．

$$\begin{aligned}
&\log p(\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}) \qquad (42)\\
&\stackrel{c}{=} \sum_{n=1}^{N}\left(-\log|\boldsymbol{Y}_n| - \left(M + \frac{\nu}{2}\right)\log\left(1 + \frac{2}{\nu}\mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right)\right)\right)
\end{aligned}$$

ここで，$\stackrel{c}{=}$ は定数項を除いて等号が成立することを示す．

式 (42) を $\boldsymbol{\mathcal{Y}}$ に関して直接最大化することは困難であるため，補助関数法（A.1 節）を用いて式 (42) の下限関数を設計し，その下限関数を最大化することで，式 (42) を間接的に最大化することを考える．

4.3.1 下限関数導出のための不等式

下限関数導出のために必要な不等式を説明する．まず，$f(\boldsymbol{Z}) = \log|\boldsymbol{Z}|$ が凹関数であることに着目すると，$f(\boldsymbol{Z})$ に対して 1 次のテイラー展開を行うことで，次式を得る．

$$-\log|\boldsymbol{Z}| \geq -\log|\boldsymbol{\Omega}| - \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{Z}) + M \tag{43}$$

ここで，$\boldsymbol{\Omega}$ は任意の半正定値行列（展開点）であり，M は $\boldsymbol{Z}$ のサイズである．等号成立条件は，$\boldsymbol{\Omega} = \boldsymbol{Z}$ で与えられる．

次に，任意の半正定値行列 $\boldsymbol{A}$ に対して $g(\boldsymbol{Z}) = \mathrm{tr}(\boldsymbol{Z}^{-1}\boldsymbol{A})$ は凸関数であることに着目すると，澤田らの提案する不等式 [21] を適用可能である．

$$-\mathrm{tr}\Bigg(\Big(\textstyle\sum_{k=1}^{K}\boldsymbol{Z}_k\Big)^{-1}\boldsymbol{A}\Bigg) \geq -\sum_{k=1}^{K}\mathrm{tr}\Big(\boldsymbol{Z}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{A}\boldsymbol{\Phi}_k^T\Big) \quad (44)$$

ここで，$\{\boldsymbol{Z}_k\}_{k=1}^{K}$ は任意の半正定値行列であり，$\{\boldsymbol{\Phi}_k\}_{k=1}^{K}$ は足すと単位行列になるような補助行列である ($\sum_k \boldsymbol{\Phi}_k = \boldsymbol{I}$)．等号成立条件は，$\boldsymbol{\Phi}_k = \boldsymbol{Z}_k(\sum_{k'}\boldsymbol{Z}_{k'})^{-1}$ で与えられる．

4.3.2 下限関数の導出

これら不等式を用いると，式 (42) に対する補助関数は以下の通り導出できる．

$$\begin{aligned}
&\log p(\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}) \\
&\overset{c}{\geq} \sum_{n=1}^{N}\Bigg(-\log|\boldsymbol{\Omega}_n| - \mathrm{tr}(\boldsymbol{\Omega}_n^{-1}\boldsymbol{Y}_n) + M \\
&\quad - \Big(M+\frac{\nu}{2}\Big)\Big(\psi_n + \psi_n^{-1}\Big(1+\frac{2}{\nu}\mathrm{tr}\big(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\big)\Big) - 1\Big)\Bigg) \\
&\overset{c}{=} \sum_{n=1}^{N}\Bigg(-\mathrm{tr}(\boldsymbol{\Omega}_n^{-1}\boldsymbol{Y}_n) - \psi_n^{-1}\Big(\frac{2M}{\nu}+1\Big)\mathrm{tr}\big(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\big)\Bigg) \\
&\geq \sum_{n=1}^{N}\Bigg(-\mathrm{tr}(\boldsymbol{\Omega}_n^{-1}\boldsymbol{Y}_n) \\
&\quad - \psi_n^{-1}\Big(\frac{2M}{\nu}+1\Big)\sum_{k=1}^{K}\mathrm{tr}\Big(h_{kn}^{-1}\boldsymbol{W}_k^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\Big)\Bigg) \quad (45)
\end{aligned}$$

ここで，$\boldsymbol{\Omega}_n$，ψ_n および $\boldsymbol{\Phi}_{kn}$ は補助変数であり，等号成立条件（下限関数の最大化条件）は以下で与えられる．

$$\boldsymbol{\Omega}_n = \boldsymbol{Y}_n \quad (46)$$

$$\psi_n = 1 + \frac{2}{\nu}\mathrm{tr}\big(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\big) \quad (47)$$

$$\boldsymbol{\Phi}_{kn} = \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1} \quad (48)$$

4.3.3 乗法更新則の導出

この下限関数を $\boldsymbol{W}_k$ および h_{kn} に関してそれぞれ偏微分しゼロとおくことで，以下の更新式を得る．

$$\boldsymbol{W}_k \leftarrow \boldsymbol{W}_k\boldsymbol{L}_k(\boldsymbol{L}_k^T\boldsymbol{W}_k\boldsymbol{P}_k\boldsymbol{W}_k\boldsymbol{L}_k)^{-\frac{1}{2}}\boldsymbol{L}_k^T\boldsymbol{W}_k \quad (49)$$

$$h_{kn} \leftarrow h_{kn}\left(\frac{\mathrm{tr}\big(\pi_n\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\boldsymbol{W}_k\boldsymbol{Y}_n^{-1}\big)}{\mathrm{tr}\big(\boldsymbol{W}_k\boldsymbol{Y}_n^{-1}\big)}\right)^{\frac{1}{2}} \quad (50)$$

ここで，$\boldsymbol{L}_k$ はコレスキー分解 $\boldsymbol{Q}_k = \boldsymbol{L}_k\boldsymbol{L}_k^T$ で求まる下三角行列であり，$\boldsymbol{P}_k$ および $\boldsymbol{Q}_k$ は次式で求まる．

$$\boldsymbol{P}_k = \sum_{n=1}^{N}h_{kn}\boldsymbol{Y}_n^{-1} \quad \boldsymbol{Q}_k = \sum_{n=1}^{N}h_{kn}\boldsymbol{Y}_n^{-1}(\pi_n\boldsymbol{X}_n)\boldsymbol{Y}_n^{-1} \quad (51)$$

また，π_n は次式で求められる．

$$\pi_n = \frac{2M+\nu}{2\mathrm{tr}(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1})+\nu} \quad (52)$$

したがって，h_{kn} の非負性と $\boldsymbol{V}_k$ の半正定値性は自然に保たれているが，$\mathrm{tr}(\boldsymbol{W}_k) = 1$ を満たすよう，反復ごとに $\boldsymbol{W}_k$ および $\boldsymbol{h}_k$ をスケーリングしておく．

式 (49) および式 (50) は，式 (35) および式 (36) の自然な拡張となっていることが分かる．実際，多変量複素 t 分布の自由度を $\nu \to \infty$ とすると，$\pi_n \to 1$ となり，多変量複素ガウス分布に基づく PSDTF (LD-PSDTF) に帰着する．一方，$\nu = 1$ とすると，多変量複素コーシー分布に基づく PSDTF が得られる．すなわち，提案する多変量複素 t 分布に基づく t-PSDTF では，$\pi_n\boldsymbol{X}_n$ を疑似的な観測データであるとして LD-PSDTF を行っていると解釈できる．ただし，π_n は現在の再構成データ $\boldsymbol{Y}_n$ に依存しているため，反復ごとに更新されることに注意する．

4.3.4 *t*-NMF の導出

多変量複素 t 分布に基づく t-PSDTF の特殊形として，複素 t 分布に基づく NMF を導出しておく．3.2 節で説明した通り，半正定値行列を対角行列に限定する場合を考えればよいので，$\boldsymbol{X}_n = \mathrm{diag}(\boldsymbol{x}_n)$，$\boldsymbol{Y}_n = \mathrm{diag}(\boldsymbol{y}_n)$，$\boldsymbol{W}_k = \mathrm{diag}(\boldsymbol{w}_k)$ とすると，乗法更新則は以下で与えられる．

$$w_{km} \leftarrow w_{km}\left(\frac{\sum_n(\pi_{nm}x_{nm})h_{kn}/y_{nm}^2}{\sum_n h_{kn}/y_{nm}}\right)^{\frac{1}{2}} \quad (53)$$

$$h_{kn} \leftarrow h_{kn}\left(\frac{\sum_m(\pi_{nm}x_{nm})w_{km}/y_{nm}^2}{\sum_m w_{km}/y_{nm}}\right)^{\frac{1}{2}} \quad (54)$$

ここで，π_{nm} は以下で与えられる．

$$\pi_{nm} = \frac{2+\nu}{2x_{nm}y_{nm}^{-1}+\nu} \quad (55)$$

このとき，疑似的な観測データ $\pi_{nm}x_{nm}$ は次式となる．

$$\pi_{nm}x_{nm} = \left(\frac{2}{2+\nu}y_{nm}^{-1} + \frac{\nu}{2+\nu}x_{nm}^{-1}\right)^{-1} \quad (56)$$

これは，観測データ x_{nm} と再構成データ y_{nm} に対して，重み $\nu : 2$ で調和平均をとることを意味し，自由度 ν が大きくなるほど，x_{nm} を重視するようになる．反対に，ν が小さくなるほど，y_{nm} を重視するようになる．

t-PSDTF では，疑似的な観測データ $\pi_n\boldsymbol{X}_n$ に対して，低ランクな再構成データ $\boldsymbol{Y}_n$ をフィットさせることが目的であるが，すでに $\pi_n\boldsymbol{X}_n$ には $\boldsymbol{Y}_n$ が取り込まれており，ある種の平滑化が行われていると見ることができる．この結果，観測データに対して過剰にフィットすることが抑制され，外れ値に対して頑健な分解ができると考えられる．

4.4 多変量複素 K 分布に基づく K-PSDTF

多変量複素 K 分布を尤度関数に用いる場合，最大化すべき対数尤度（式 (41) 参照）は，式 (24) を用いて表現できる．4.3 節同様に，補助関数法（A.1 節）を用いて対数尤度の下限関数を設計し，その下限関数を最大化することを考える．このとき，第二種変形ベッセル関数 $K_\nu(x)$ が x に関して対数凸であることを利用する．

5. 評価実験

多変量複素 t 分布に基づく t-PSDTF を用いた予備的な音源分離実験について報告する．

5.1 実験条件

実験には，RWC 研究用音楽データベース：楽器音 [22] に収録されているピアノ (011PFNOM)，エレキギター (131EGLPM) およびクラリネット (311CLNOM) の単独音を用いた．各楽器ごとに，異なる 3 つの音高 (C4, E4, G4) をもつ 2 秒間の音響信号を準備し，それらを 7 つの異なる組み合わせで重畳したもの (C4, E4, G4, C4+E4, C4+G4, E4+G4, C4+E4+G4) を連結することで 14 秒の音響信号を合成した．サンプリング周波数は 16[kHz] とした．

次に，t-PSDTF を用いて，与えられた混合音を C4, E4, G4 に対応する音源信号に分離することを試みた．まず，ガウス窓を用いて局所信号 $\{\boldsymbol{o}_n\}_{n=1}^N$ を切り出し，$\boldsymbol{X}_n = \boldsymbol{o}_n\boldsymbol{o}_n^T$ とすることで観測データ $\{\boldsymbol{X}_n\}_{n=1}^N$ を得た．窓幅は 512 点，シフト長は 160 点とした ($M = 512$, $N = 1400$)．ここで，t 分布などの GSM 分布はアフィン変換に対して閉じている (2.2 節参照）ことから，周波数領域だけではなく，時間領域においても等価な分解が可能である [7]．このとき，ウィナーフィルタリングを行うと，各音源信号の複素スペクトルではなく，時間信号が直接求まる．比較のため，パワースペクトログラムに対する t-NMF も評価した．各手法に対して，基底数は $K = 3$，反復回数は 100 回とし，乗法更新アルゴリズムを用いた．音源分離結果は，BSS Eval Toolbox [23] を用いて，source-to-distortion ratio (SDR)，source-to-interferences ratio (SIR) および sources-to-artifacts ratio (SAR) で評価した．

5.2 実験結果

音源分離タスクにおいて t-PSDTF は t-NMF に対する明確な優位性を示した．SDR, SIR, SAR の平均は，IS-NMF（$\nu \to \infty$ とした t-NMF）では 19.1 dB, 24.0 dB, 21.0 dB であったのに対し，LD-PSDTF（$\nu \to \infty$ とした t-PSDTF）では 23.0 dB, 27.7 dB, 25.1 dB であった．また，コーシー尤度に基づく NMF と PSDTF（$\nu = 1$ とした t-NMF と t-PSDTF）はそれぞれ，IS-NMF と LD-PSDTF とほぼ同等の性能を示した．この予備実験で用いたデータは単純であったが，突発性の雑音が多く含まれるより現実的な音楽音響信号に対する評価を今後実施予定である．

図 5 に示す通り，PSDTF を用いると，減衰音と持続音のいずれに対しても基底行列 $\boldsymbol{\mathcal{W}}$ および音量変化 $\boldsymbol{H}$ を適切に推定できた．ここで，各基底行列 $\boldsymbol{W}_k$ における斜め縞の間隔は周期を表しており，$\boldsymbol{W}_k$ の中心付近は巡回行列に近くなっている．しかし，窓関数の影響で周辺部はそうはならないので，周波数成分間に相関が発生することは原理的に避けられない．PSDTF はこの影響を考慮することで優れた音源分離性能を達成している．また，t-PSDTF に対する乗法更新アルゴリズムが，ν の値に関わらず数値的にも安定して収束することを確認した．

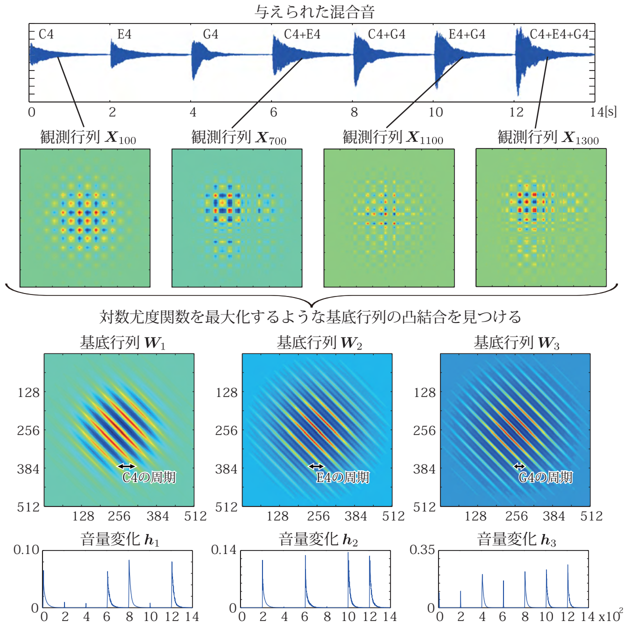

図 5 時間領域 t-PSDTF を用いたピアノ信号の分解結果

PSDTF の主な課題は，計算コストが $O(KNM^3)$ であり，NMF の $O(KNM)$ よりもはるかに大きいことである．計算時間を短縮し，局所解を回避するため，実際には NMF で PSDTF を初期値する，すなわちある程度収束が進むまで基底行列を対角行列に制限して反復更新を行う方法が推奨される．

6. おわりに

本稿では，非ガウス性モノラル音響信号に対して音源分離を行うのに適した，NMF と PSDTF の新しい確率モデルを提案した．実際の音響信号は非ガウス性（多くの場合優ガウス性）を持つため，ガウス分布より裾の重い確率分布である複素ガウススケール混合 (GSM) 分布を尤度関数に用いることが望ましい．そこで、本稿では，複素 GSM 分布のうちで特に重要な分布である複素 t 分布（特殊形として複素コーシー分布を含む）を尤度関数として，基底とアクティベーションを同時に最尤推定するための効率的な乗法更新アルゴリズムを導出した．実験の結果，高品質な音源分離ができることを確かめた．

今後は，多変量複素安定分布を尤度関数とした PSDTF の導出に取り組みたい．安定分布はガウス分布やコーシー分布を特殊な場合として含み，音源信号の重畳をモデル化するうえで数学的に妥当性がある [9]．

謝辞: 本研究の一部は，JSPS 科研費 24220006, 26700020, 26280089，JST CREST OngaCREST，および栢森情報科学振興財団の支援を受けた．

付　　録

A.1　補助関数法

目的関数 $\mathcal{F}(\boldsymbol{\theta})$ を変数 $\boldsymbol{\theta}$ に関して最大化する問題を考える．補助関数法 [4, 25] を用いると，$\mathcal{F}(\boldsymbol{\theta})$ を「間接的に」最大化することができる．まず，$\mathcal{F}(\boldsymbol{\theta})$ に対する補助関数として，以下の下限関数 $\mathcal{F}^{+}(\boldsymbol{\theta}, \boldsymbol{\phi})$ を考える．

$$\mathcal{F}(\boldsymbol{\theta}) \geq \mathcal{F}^{+}(\boldsymbol{\theta}, \boldsymbol{\phi}) \tag{A.1}$$

ここで，$\boldsymbol{\phi}$ は補助変数である．このとき，以下の反復更新則

$$\boldsymbol{\phi}^{\mathrm{new}} \leftarrow \mathrm{argmax}_{\boldsymbol{\phi}} \mathcal{F}^{+}(\boldsymbol{\theta}^{\mathrm{old}}, \boldsymbol{\phi}) \tag{A.2}$$

$$\boldsymbol{\theta}^{\mathrm{new}} \leftarrow \mathrm{argmax}_{\boldsymbol{\theta}} \mathcal{F}^{+}(\boldsymbol{\theta}, \boldsymbol{\phi}^{\mathrm{new}}) \tag{A.3}$$

を用いると，$\mathcal{F}(\boldsymbol{\theta})$ は単調非減少となる．このアルゴリズムの収束性は保証されている．

A.2　確率密度関数

確率分布の確率密度関数を以下に整理しておく．

ガンマ分布

$$\mathcal{G}(x|a,b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}$$

多変量複素ガウス分布

$$\mathcal{N}_c(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi^d} |\boldsymbol{\Sigma}|^{-1} \exp\left(-Q(\boldsymbol{x})\right)$$

多変量複素 t 分布

$$t_c(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{\Gamma(\frac{2d+\nu}{2})}{\Gamma(\frac{\nu}{2})} \frac{2^d}{(\pi\nu)^d} |\boldsymbol{\Sigma}|^{-1} \left(1 + \frac{2}{\nu} Q(\boldsymbol{x})\right)^{-\frac{2d+\nu}{2}}$$

多変量複素 K 分布

$$\mathcal{K}_c(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) = \frac{2\nu^d}{\Gamma(\nu)\pi^d} |\boldsymbol{\Sigma}|^{-1} \left(\sqrt{\nu Q(\boldsymbol{x})}\right)^{\nu-d} K_{\nu-d}\left(2\sqrt{\nu Q(\boldsymbol{x})}\right)$$

参考文献

[1] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman. Dynamic source separation using non-negative factorizations: A unified view. *IEEE Signal Processing Magazine*, 31(3):66–75, 2014.

[2] H. Kameoka and K. Kashino. Composite autoregressive system for sparse source-filter representation of speech. In *ISCAS*, pages 2477–2480, 2009.

[3] K. Yoshii and M. Goto. Infinite composite autoregressive models for music signal analysis. In *ISMIR*, pages 79–84, 2012.

[4] M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *ICML*, pages 439–446, 2010.

[5] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000.

[6] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009:Article ID 785152, 2009.

[7] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *ICML*, pages 576–584, 2013.

[8] A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2004.

[9] A. Liutkus and R. Badeau. Generalized Wiener filtering with fractional power spectrograms. In *ICASSP*, 2015.

[10] A. Shah and A. G. Wilson Zand . Ghahramani. Student-t processes as alternatives to gaussian processes. In *AISTATS*, pages 877–885, 2014.

[11] B. Lakshminarayanan, G. Bouchard, and C. Archambeau. Robust Bayesian matrix factorisation. In *AISTATS*, pages 425–433, 2011.

[12] N. Wang and D.-Y. Yeung. Bayesian robust matrix factorization for image and video processing. In *ICCV*, pages 1785–1792, 2013.

[13] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor. Complex elliptically symmetric distributions: Survey, new results and applications. *IEEE TASLP*, 60(11):5597–5625, 2012.

[14] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Beyond nmf: Time-domain audio source separation without phase reconstruction. In *ISMIR*, pages 369–374, 2013.

[15] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.

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

[17] K. Tsuda, G. Rätsch, and M. K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *JMLR*, 6:995–1018, 2005.

[18] B. Kulis, M. Sustik, and I. Dhillon. Low-rank kernel learning with Bregman matrix divergences. *JMLR*, 10:341–376, 2009.

[19] 亀岡弘和. 非負値行列因子分解. **計測と制御**, 51(9):835–844, 2012.

[20] 亀岡弘和. 非負値行列因子分解の音響信号処理への応用. **日本音響学会誌**, 68(11):559–565, 2012.

[21] H. Sawada, H. Kameoka, S. Araki, and N. Ueda. Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization. In *ICASSP*, pages 261–264, 2012.

[22] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In *ISMIR*, pages 229–230, 2003.

[23] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE TASLP*, 14(4):1462–1469, 2006.

[24] J. Le Roux, E. Vincent, Y. Mizuno, H. Kameoka, N. Ono, and S. Sagayama. Consistent Wiener filtering: Generalized time-frequency masking respecting spectrogram consistency. In *LVA/ICA*, pages 89–96, 2010.

[25] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama. Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta divergence. In *MLSP*, pages 283–288, 2010.