モノラル音響信号に対する音源分離のための無限半正定値テンソル分解

吉井 和佳[1,a)]　富岡 亮太[2,b)]　持橋 大地[3,c)]　後藤 真孝[1,d)]

概要：本稿では，モノラルの音響信号に対して音源分離を行うための新しい因子分解法について述べる．この種の音源分離タスクにおいては，非負値行列分解 (NMF) を利用するのが現在の主流である．NMF では，与えられた混合音の振幅あるいはパワースペクトログラムを非負値行列とみなし，それを少数の基底スペクトルとそれらの時間方向のアクティベーションとの積に分解することができる．しかし，NMF の性質上，入力は非負値行列でなくてはならず，混合音の複素スペクトログラムに本来備わっている位相情報は考慮されていなかった．そのため，音源信号を再合成するには，音源信号のスペクトログラムは混合音のスペクトログラムと同一の位相をもつという強い仮定をおく必要があった．本研究では，周波数領域での位相復元を回避するため，混合音を時間領域で直接分解することができる半正定値テンソル分解 (PSDTF) を提案する．PSDTF は NMF の本質的な拡張となっているため，NMF と同様に効率的な乗法更新アルゴリズムに基づく最尤推定が可能であり，ガンマ過程に基づくノンパラメトリックベイズ拡張（基底数の無限化）も可能である．実験の結果，PSDTF は NMF より高品質な音源分離ができることを確かめた．

1. はじめに

音楽音響信号に対する音源分離は，音楽情報検索 (MIR) を支える基礎技術のひとつである．高品質な音源分離が実現できれば，歌唱者の声質や性別，楽器構成などの楽曲の内容に基づいて，より緻密にユーザの好みに合う楽曲を検索できるようになる [1]．また，能動的音楽鑑賞 [2] の一形態として，混合音中に含まれる既存の楽器パートを自分好みに編集しながら音楽鑑賞を楽しむことができるシステム [1–4] を実現するうえでも不可欠な技術である．

モノラル音響信号に対する音源分離タスクでは，非負値行列分解 (Nonnegative Matrix Factorization: NMF) [5] を利用することが主流になっている．NMF は，入力となる非負値行列（振幅あるいはパワースペクトログラム）を二つの非負値行列（基底スペクトルの集合と対応するアクティベーションの集合）の積で近似することができる．その後，ウィナーフィルタを用いて，混合音の複素スペクトログラムを音源信号の複素スペクトログラムの和に分解する処理が行われる．このとき，混合音と音源信号のスペクトログラムの位相は同一であると仮定することが一般的であるため，非適切な位相をもつスペクトログラムから高品質な音源信号を再合成することはできなかった．

これまで，実際の時間領域信号に対応する「無矛盾な」複素スペクトログラムを推定するため，多数の研究が行われてきた．Griffin ら [6] は，与えられた振幅スペクトログラムの位相を復元するため，その振幅スペクトログラムにできるだけ近い振幅スペクトログラムをもつ時間領域信号を推定できる反復 STFT 法を提案している．Le Roux ら [7] は，与えられた複素スペクトログラムの矛盾性を評価するためのコスト関数を導出し，それを最小化するための効率的なアルゴリズムを提案している [8]．一方，亀岡ら [9] は，混合音の複素スペクトログラムを複素成分の加法性に基づいて直接分解することができる複素 NMF を提案している．しかし，得られる音源信号の複素スペクトログラムは必ずしも無矛盾であるとは限らないため，Le Roux らのコスト関数を組み込む方法が提案されている [10]．ここで，複素スペクトログラムが無矛盾性を満たすことは、得られる音源信号が高品質であることを必ずしも意味しないことに注意する必要がある．このことは，周波数領域における位相復元は容易ではないことを示唆している．

本研究では，位相復元を根本的に不要とするため，半正定値テンソル分解 (Positive Semidefinite Tensor Factorization: PSDTF) [11] と呼ぶ新しい因子分解法を提案する．PSDTF では，モノラルの混合音を時間領域において音源信号に直接分解することができる．一般的に，時間領域に

[1] 産業技術総合研究所
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
[2] 東京大学
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
[3] 統計数理研究所
10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
[a)] k.yoshii(at)aist.go.jp
[b)] tomioka(at)mist.i.u-tokyo.ac.jp
[c)] daichi(at)ism.ac.jp
[d)] m.goto(at)aist.go.jp

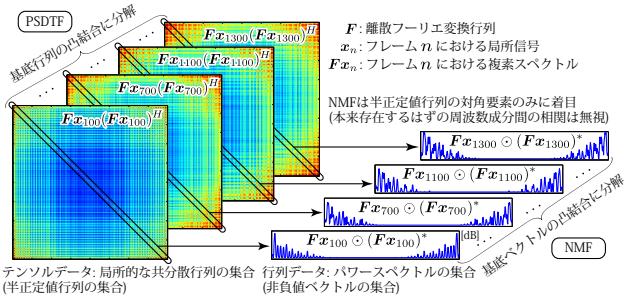

図 1 周波数領域における音源分離：PSDTF は NMF の自然な拡張

おいて，音響信号の位相や波形を何らかの関数形を用いて明示的に表現することは容易ではないため，我々は音響信号の統計的な性質に着目する．まず，少数の定常な基底信号の存在を考え，それぞれが異なるカーネル（基底カーネルと呼ぶ）をもつ定常なガウス過程に従うと仮定する．与えられた混合音は，局所的にみると複数の基底信号の線形和で構成されていると仮定する．ただし，結合係数は各時刻ごとに異なっている．このとき，混合音は局所的にガウス過程に従い，そのカーネルは基底カーネルの凸結合で与えられる．我々の目標は，観測データである混合音の局所的な共分散から，それらを構成する少数の基底カーネルを求めることである．具体的には，乗法更新アルゴリズムに基づく最尤推定あるいは変分ベイズ法に基づくベイズ推定が可能である．さらに，基底数が不明な場合でも，ガンマ過程を用いて理論上無限個の基底の存在を仮定するノンパラメトリックベイズモデルを構成できる．

これまで時間領域における PSDTF について議論してきたが，実は周波数領域において等価な PSDTF を構成することができ，PSDTF は NMF の自然な拡張となっていることが明らかとなる．**図 1** に示すように，PSDTF では，各時刻における複素スペクトルの直積，すなわち**半正定値行列を少数の半正定値基底行列の和に分解**する．一方，NMF では，上記直積行列の対角成分（パワースペクトル），すなわち**非負値ベクトルを少数の非負値基底ベクトルの和に分解**する．ここでの重要な違いは，PSDTF では異なる周波数ビン間の相関が考慮されていることである．このことの妥当性は，スペクトログラムを得るための標準的な方法である短時間フーリエ変換やウェーブレット変換では，周波数ビン間を完全に無相関化することはできないという事実に基づいている．したがって，調波構造などの周波数ビン間の強い相関を考慮しながら音源分離を行うことで，高品質な音源信号を復元することができる．

2. 周波数領域における音源分離

本章では，周波数領域における音源分離を行ううえで，非負値行列分解 (NMF) の理論的な裏付けについて説明する．特に，NMF のよく知られた 2 つの変種である KL-NMF [12] および IS-NMF [13] について議論する．

2.1 非負値行列分解

NMF の目標は，非負値行列 $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{M\times N}$ を低ランク近似する，すなわち二つの非負値行列 $\boldsymbol{W} = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_K] \in \mathbb{R}^{M\times K}$ および $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K]^T \in \mathbb{R}^{K\times N}$ の積で近似することである．

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} \stackrel{\text{def}}{=} \boldsymbol{Y} \tag{1}$$

ここで，$\boldsymbol{w}_k$ および $\boldsymbol{h}_k$ 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．ただし，$K \ll \min(M, N)$ は基底数，$\boldsymbol{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N] \in \mathbb{R}^{M\times N}$ は再構成行列を表す．行列積表現である式 (1) は，各 n に関するベクトル和として書きくだすことができる．

$$\boldsymbol{x}_n \approx \sum_{k=1}^{K} h_{kn}\boldsymbol{w}_k \stackrel{\text{def}}{=} \boldsymbol{y}_n \tag{2}$$

ここで，$\boldsymbol{y}_{kn} = h_{kn}\boldsymbol{w}_k$ と定義すると，$\boldsymbol{y}_n = \sum_k \boldsymbol{y}_{kn}$ が成立する．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 $\mathcal{C}(\boldsymbol{x}_n|\boldsymbol{y}_n)$ を評価する尺度として，Bregman ダイバージェンス [14] が広く利用されている．

$$\mathcal{C}_\phi(\boldsymbol{x}_n|\boldsymbol{y}_n) = \phi(\boldsymbol{x}_n) - \phi(\boldsymbol{y}_n) - \phi'(\boldsymbol{y}_n)^T(\boldsymbol{x}_n - \boldsymbol{y}_n) \tag{3}$$

ここで，ϕ は厳密に凸な関数である．Bregman ダイバージェンスは常に非負であり，$\boldsymbol{x}_n = \boldsymbol{y}_n$ であるときに限りゼロをとる．その特別な形として，$\phi(\boldsymbol{x}) = \sum_m (x_m \log x_m - x_m)$ の場合の Kullback-Leibler (KL) ダイバージェンスや，$\phi(\boldsymbol{x}) = -\sum_m \log x_m$ の場合の Itakura-Saito (IS) ダイバージェンスがよく知られている．コスト関数 $\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y}) = \sum_n \mathcal{C}_\phi(\boldsymbol{x}_n|\boldsymbol{y}_n)$ を最小化する $\boldsymbol{W}$ および $\boldsymbol{H}$ を求めるため，乗法更新アルゴリズム [15] が提案されている．これは，ある確率モデルの最尤推定を行うことに相当する．一方，$\boldsymbol{W}$ および $\boldsymbol{H}$ に適切な事前分布（通常はガンマ分布）を導入してベイズ推定を行う方法も提案されている [16,17].

2.2 音源分離への応用

我々の目的は，与えられた混合音を K 個の音源信号の和に分解することである．NMF を用いる場合，音源分離は周波数領域で行うことになる．いま，混合音の「非負値」スペクトログラムを $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ とし，M および N は周波数ビン数およびフレーム数を表すものとする．このとき，NMF を用いて $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$ という分解を行うと，$\boldsymbol{w}_k \in \mathbb{R}^M$ および $\boldsymbol{h}_k \in \mathbb{R}^N$ はそれぞれ基底スペクトルおよび時間方向の音量変化（アクティベーション）を表す．

潜在変数である音源信号の推定を行うには，NMF の確率モデルを考える必要がある．いま，混合音の複素スペクトログラムを $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_N] \in \mathbb{C}^{M\times N}$ とし，k 番目の音源信号の複素スペクトログラムを $\boldsymbol{S}_k = [\boldsymbol{s}_{k1}, \cdots, \boldsymbol{s}_{kn}] \in \mathbb{C}^{M\times N}$ とする．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．

$$\boldsymbol{S} = \sum_{k=1}^{K} \boldsymbol{S}_k \quad i.e., \quad \boldsymbol{s}_n = \sum_{k=1}^{K} \boldsymbol{s}_{kn} \tag{4}$$

観測変数として $\boldsymbol{S}$ が与えられると，潜在変数である $\boldsymbol{S}_k$ の期待値は次式で計算することができる．

$$\mathbb{E}[s_{knm}|s_{nm}] = \frac{y_{knm}}{y_{nm}} s_{nm} = \frac{w_{km}h_{kn}}{\sum_k w_{km}h_{kn}} s_{nm} \tag{5}$$

この処理はウィナーフィルタリングと呼ばれ，$\boldsymbol{S}_k$ の位相は $\boldsymbol{S}$ の位相と同一であるという仮定がおかれている．最後に，重畳加算合成法 [18] を用いれば，$\mathbb{E}[\boldsymbol{S}_k|\boldsymbol{S}]$ から k 番目の音源信号を復元することができる．しかし，得られた音源信号から複素スペクトログラムを再度計算してみても，$\mathbb{E}[\boldsymbol{S}_k|\boldsymbol{S}]$ とは等しくはならない．なぜなら，$\mathbb{E}[\boldsymbol{S}_k|\boldsymbol{S}]$ は実際の時間領域信号に対応する「無矛盾な」スペクトログラムではないからである．これは NMF に基づく音源分離の原理的な欠陥であり，音源信号の合成品質に限界がある根本的な原因のひとつとなっている．

2.3 KL-NMF に基づく音源分離

KL-NMF は，混合音の「振幅」スペクトログラムを分解するために用いることが一般的である [12]．すなわち，$x_{nm} = |s_{nm}|$ とする．コスト関数は KL ダイバージェンスに基づくものであり，次式で与えられる．

$$\mathcal{C}_{\mathrm{KL}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_{m=1}^{M} \left(x_{nm} \log \frac{x_{nm}}{y_{nm}} - x_{nm} + y_{nm} \right) \tag{6}$$

ただし，式 (6) はスケール不変ではないことに注意する．すなわち，任意の実数 α に対して $\mathcal{C}_{\mathrm{KL}}(\boldsymbol{X}|\boldsymbol{Y}) = \mathcal{C}_{\mathrm{KL}}(\alpha\boldsymbol{X}|\alpha\boldsymbol{Y})$ とはならない．このことは，与えられた混合音全体の音量を変化させると音源分離結果も変化することを意味しており，理論的には妥当であるとはいえない．

確率モデル化のためには，潜在変数 $|s_{knm}|$ が y_{knm} を平均パラメータとするポアソン分布に従うことを仮定する．

$$|s_{knm}|\big|y_{knm} \sim \mathrm{Poisson}(y_{knm}) \tag{7}$$

ここで，振幅の加法性が成立することを仮定すると，式 (4) は $|s_{nm}| = \sum_k |s_{knm}|$ と書きなおせる．すなわち，混合音のスペクトログラムの位相は音源信号のスペクトログラムの位相と同一であると仮定することを意味する．これは通常成立しないが，実用上は都合がよい．いま，$x_{nm} = |s_{nm}|$ かつ $y_{nm} = \sum_k y_{knm}$ であることを思い出すと，ポアソン分布の再生性から次式を得る．

$$x_{nm}\big|y_{nm} \sim \mathrm{Poisson}(y_{nm}) \tag{8}$$

$|s_{nm}| = \sum_k |s_{knm}|$ を満たすポアソン変数 $\{|s_{knm}|\}_{k=1}^K$ に基づく確率モデルを用いると，$x_{nm} = |s_{nm}|$ が与えられたときの $|s_{knm}|$ の期待値は次式で与えられる．

$$\mathbb{E}[|s_{knm}|\big||s_{nm}|] = y_{knm}y_{nm}^{-1}|s_{nm}| \tag{9}$$

最終的に，位相の保存性から式 (5) を得る．

2.4 IS-NMF に基づく音源分離

IS-NMF は，混合音の「パワー」スペクトログラムを分解するために用いることが一般的である [13]．すなわち，$x_{nm} = |s_{nm}|^2$ とする．コスト関数は IS ダイバージェンスに基づくものであり，次式で与えられる．

$$\mathcal{C}_{\mathrm{IS}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_{m=1}^{M} \left(\frac{x_{nm}}{y_{nm}} - \log \frac{x_{nm}}{y_{nm}} - 1 \right) \tag{10}$$

式 (10) はスケール不変であるので，理論的には IS-NMF の方が音源分離に適していることが知られている．

確率モデル化のためには，潜在変数 s_{knm} が y_{knm} を分散パラメータとする複素ガウス分布に従うことを仮定する．

$$s_{knm}|y_{knm} \sim \mathcal{N}_c(0, y_{knm}) \tag{11}$$

ここで，$s_{nm} = \sum_k s_{knm}$ かつ $y_{nm} = \sum_k y_{knm}$ であることを思い出すと，複素ガウス分布の再生性から次式を得る．

$$s_{nm}|y_{nm} \sim \mathcal{N}_c(0, y_{nm}) \tag{12}$$

したがって，$x_{nm} = |s_{nm}|^2$ は指数分布に従うことが分かる．

$$x_{nm}|y_{nm} \sim \mathrm{Exponential}(y_{nm}) \tag{13}$$

$s_{nm} = \sum_k s_{knm}$ を満たすガウス変数 $\{s_{knm}\}_{k=1}^K$ に基づく確率モデルを用いると，s_{nm} が与えられたときの s_{knm} の期待値は式 (5) で，分散は次式で求めることができる．

$$\mathbb{V}[s_{knm}|s_{nm}] = y_{knm} - y_{knm}y_{nm}^{-1}y_{knm} \tag{14}$$

3. 時間領域における音源分離

本章では初心に戻り，音源分離を時間領域における分解問題として考えなおす．具体的には，音源信号の重畳に対する確率モデルを提案し，潜在変数である音源信号を確率的な枠組みで推定する方法について議論する．

3.1 問題設定

我々の目標は，与えられた混合音を K 個の音源信号の和に分解することである．いま，観測データとして N 個の実数値ベクトルの集合 $\boldsymbol{O} = [\boldsymbol{o}_1, \cdots, \boldsymbol{o}_N] \in \mathbb{R}^{M\times N}$ を考える．ここで，$\boldsymbol{o}_n \in \mathbb{R}^M$ は，長さ M の窓を用いて混合音から切り出された n 番目のフレームにおける局所的な信号である．このとき，$\boldsymbol{o}_n$ の分解は次式で与えられる．

$$\boldsymbol{o}_n = \sum_{k=1}^{K} \boldsymbol{o}_{kn} \tag{15}$$

ここで，$\boldsymbol{o}_{kn}$ は k 番目の音源信号から切り出された n 番目のフレームにおける局所信号である．離散フーリエ変換行列を $\boldsymbol{F} \in \mathbb{C}^{M\times M}$ とすると，$\boldsymbol{s}_n = \boldsymbol{F}\boldsymbol{o}_n$ かつ $\boldsymbol{s}_{kn} = \boldsymbol{F}\boldsymbol{o}_{kn}$ であるから，式 (15) で与えられる時間領域表現は式 (4) で与えられる周波数領域表現と等価である．

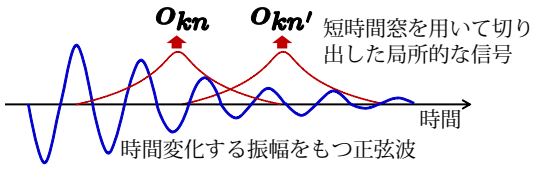

図 2 局所信号 $\boldsymbol{o}_{kn}$ および $\boldsymbol{o}_{kn'}$ は異なる位相や振幅をもっているが，同一の周期に従っていることに着目する．

モノラルの混合音を分離することは不良設定問題であるため，何らかの制約が必要になる．本研究では，k 番目の音源信号における局所的な波形 $\{\boldsymbol{x}_{kn}\}_{n=1}^{N}$ は一見異なっているものの，何らかの統計的な性質（周期性や白色性など）を共有していると仮定する．例えば，**図 2** に示すように，音源信号が時間変化する振幅をもつ正弦波であれば，局所信号 $\boldsymbol{o}_{kn}$ および $\boldsymbol{o}_{kn'}$ $(n \neq n')$ は異なる位相や振幅をもっているものの，定常な周期で振動している．そこで，$\boldsymbol{o}_{kn}$ に対して，非定常な因子 π_{kn} と定常な因子 $\boldsymbol{\phi}_{kn}$ とに基づく分解 $\boldsymbol{o}_{kn} = \pi_{kn}\boldsymbol{\phi}_{kn}$ を考える．ここで，$\boldsymbol{\phi}_{kn}$ は k 番目の定常な「基底信号」から切り出された n 番目のフレームにおける局所信号であり，π_{kn} はその係数である．これらは非負値に限定されていないことに注意する．

結局，観測データ $\boldsymbol{O}$ が与えられたときに，各 k について局所信号 $\{\boldsymbol{o}_{kn}\}_{n=1}^{N}$ を求めることが目標となる．k 番目の音源信号は，重畳加算法合成法 [18] を用いて得ることができる．これらの処理は全て時間領域で実行されるため，短時間フーリエ変換やその逆変換などは必要ない．

3.2 確率モデルの定式化

式 (15) で与えられる分解に対する確率モデルを定式化したい．定常な基底信号に対しては，局所信号 $\{\boldsymbol{\phi}_{kn}\}_{n=1}^{N}$ が同じ共分散をもつことが期待できるので，同一の多次元ガウス分布を事前分布として考えることができる．

$$\boldsymbol{\phi}_{kn}|\boldsymbol{V} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_k), \tag{16}$$

ここで，共分散行列 $\boldsymbol{V}_k \in \mathbb{R}^{M\times M}$ は対角行列に限定されておらず，任意の半正定値行列でよい．音響信号は通常ゼロを中心とする実数として記録されるため，平均ベクトルは $\boldsymbol{0}$ とした．式 (16) は，k 番目の基底信号がカーネル $\boldsymbol{V}_k$ をもつ定常ガウス過程に従うことを意味する．本来，音響信号は離散的な時刻上に存在する不連続な系列ではなく，連続時間上に存在するなめらかな関数とみなすべきであるので，連続関数に対する確率分布を考えることは本質的である．本研究では，任意の M 点における周辺分布が式 (16) で与えられるガウス分布となるため，基底信号が従う確率分布はガウス過程に他ならない．例えば，もし $\boldsymbol{V}_k$ が周期カーネルであれば，局所信号 $\{\boldsymbol{\phi}_{kn}\}_{n=1}^{N}$ は異なる位相をもつ一方，同一の周期に従うことが期待できる．

次に，観測信号 $\boldsymbol{o}_n$ の尤度について考える．$\boldsymbol{o}_{kn}$ と $\boldsymbol{\phi}_{kn}$ の間には線形性 $\boldsymbol{o}_{kn} = \pi_{kn}\boldsymbol{\phi}_{kn}$ が成立しているため，式 (16) から $\boldsymbol{o}_{kn}$ もやはりガウス分布に従うことが分かる．

$$\boldsymbol{o}_{kn}|\boldsymbol{\pi}, \boldsymbol{V} \sim \mathcal{N}(\boldsymbol{0}, \pi_{kn}^2\boldsymbol{V}_k) \tag{17}$$

さらに，式 (15) に着目すると，ガウス分布の再生性から $\boldsymbol{o}_n$ もガウス分布に従うことが分かる．

$$\boldsymbol{o}_n|\boldsymbol{\pi}, \boldsymbol{V} \sim \mathcal{N}\left(\boldsymbol{0}, \sum_{k=1}^{K}\pi_{kn}^2\boldsymbol{V}_k\right) \tag{18}$$

式 (18) には，基底信号の具体的な波形を表す $\{\boldsymbol{\phi}_{kn}\}_{k=1}^{K}$ が含まれておらず，基底信号を周辺化することであらゆる波形の可能性を考慮した表現となっている．したがって，$\{\boldsymbol{\phi}_{kn}\}_{k=1}^{K}$ の位相を明示的に考える必要がなく，より頑健な音源分離ができると期待できる．いま，$h_{kn} = \pi_{kn}^2 \geq 0$, $\boldsymbol{X}_n = \boldsymbol{o}_n\boldsymbol{o}_n^T \succeq \boldsymbol{0}$, $\boldsymbol{Y}_n = \sum_k h_{kn}\boldsymbol{V}_k \succeq \boldsymbol{0}$ とすると[*1]，式 (18) から $\boldsymbol{X}_n$ の対数尤度は次式で与えられる[*2]．

$$\log p(\boldsymbol{X}_n|\boldsymbol{Y}_n) \stackrel{c}{=} -\frac{1}{2}\log|\boldsymbol{Y}_n| - \frac{1}{2}\mathrm{tr}(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}) \tag{19}$$

いま，観測データとしてテンソル $\boldsymbol{\mathcal{X}} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n] \in \mathbb{R}^{M\times M\times N}$ が与えられたとき，対数尤度 $\sum_n \log p(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ が最大となるような $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K]^T \in \mathbb{R}^{K\times N}$ および $\boldsymbol{\mathcal{V}} = [\boldsymbol{V}_1, \cdots, \boldsymbol{V}_K] \in \mathbb{R}^{M\times M\times K}$ を求めたい．この問題は，あとの 4 章に示すように，入力となる $\boldsymbol{X}_n$ がランク 1 の半正定値行列 $(\boldsymbol{X}_n = \boldsymbol{o}_n\boldsymbol{o}_n^T)$ に限定された PSDTF の特別な場合である．したがって，4.3 節で述べる乗法更新アルゴリズムを利用して最尤推定することができる．

3.3 確率モデルに基づく音源分離

$\boldsymbol{H}$ および $\boldsymbol{\mathcal{V}}$ が求まれば，局所信号 $\boldsymbol{o}_{kn} = \pi_{kn}\boldsymbol{\phi}_{kn}$ を確率的な枠組みのもとで推定することができる．ここで，π_{kn} や $\boldsymbol{\phi}_{kn}$ を先に求めておく必要はなく，直接 $\boldsymbol{o}_{kn}$ の事後分布を計算することができる．具体的には，式 (17) および式 (18) から，$\boldsymbol{o}_{kn}$ の事後分布もガウス分布となることが分かり，その平均と分散は次式で与えられる．

$$\mathbb{E}[\boldsymbol{o}_{kn}|\boldsymbol{o}_n] = \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{o}_n \tag{20}$$

$$\mathbb{V}[\boldsymbol{o}_{kn}|\boldsymbol{o}_n] = \boldsymbol{Y}_{kn} - \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{Y}_{kn} \tag{21}$$

ここで，$\boldsymbol{Y}_{kn} = h_{kn}\boldsymbol{V}_k \succeq \boldsymbol{0}$ とした $(\boldsymbol{Y}_n = \sum_k \boldsymbol{Y}_{kn} \succeq \boldsymbol{0})$．式 (20) は，観測信号 $\boldsymbol{o}_n$ のうち，位相や波形を明示的に考えることなく，カーネル $\boldsymbol{V}_k$ で表現できる成分のみを通過させるウィナーフィルタである．時間領域における式 (20) および式 (21) は，周波数領域における式 (5) および式 (14) と形式上は非常に類似している．ただし，我々の提案法ではベクトル $\boldsymbol{o}_{kn}$ ごとに独立な計算になっているのに対し，従来はベクトル $\boldsymbol{s}_{kn}$ の要素 s_{knm} ごとに独立な計算になっている点が異なる．したがって，提案法では，周波数ビン間の相関構造を考慮しながら音源分離を行っていることが示唆される（詳細は次節参照のこと）．

*1 $\boldsymbol{\Psi}$ が半正定値行列であるとき $\boldsymbol{\Psi} \succeq \boldsymbol{0}$ と書く．
*2 $\stackrel{c}{=}$ は定数項を除いて等号が成立することを示す．

3.4 フーリエトリックに基づく近似的高速計算

従来と同様に周波数領域での定式化について議論する（図 1）．離散フーリエ変換行列を $\boldsymbol{F} \in \mathbb{C}^{M\times M}$ とすると，式 (18) から，観測信号 $\boldsymbol{o}_n$ の線形写像である複素スペクトル $\boldsymbol{F}\boldsymbol{o}_n$ は複素ガウス分布に従うことが分かる．

$$\boldsymbol{F}\boldsymbol{o}_n|\boldsymbol{H},\boldsymbol{V} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_{k=1}^{K} h_{kn}\boldsymbol{F}\boldsymbol{V}_k\boldsymbol{F}^H\right) \tag{22}$$

ここで，複素スペクトルの要素間の完全な共分散構造が考慮されていることに注意する．

式 (22) は式 (11) で与えられる IS-NMF の確率モデルの自然な拡張となっていることを示している．数学的には，$\boldsymbol{V}_k$ が巡回行列であれば，$\boldsymbol{F}\boldsymbol{V}_k\boldsymbol{F}^H$ は対角行列となることが知られている（$\boldsymbol{V}_k$ は対角化されるという）．このとき，式 (22) は，複素スペクトルの各要素が独立であると仮定している式 (11) と等価になる．自明な例は，$\boldsymbol{V}_k$ が単位行列である場合，すなわち，$\boldsymbol{\phi}_{kn}$ が定常な白色ガウス性雑音である場合である．一方，$\boldsymbol{V}_k$ が周期カーネルであり，窓幅 M がその周期より十分に大きければ，$\boldsymbol{V}_k$ は巡回行列と類似した斜行型の要素配置をもつ．しかし，厳密には $\boldsymbol{V}_k$ は巡回行列とは異なるため，複素スペクトルの要素間に相関が発生してしまう (**図 4** および**図 5** の実験結果を参照)．ただし，音響信号は白色成分と周期成分とで構成されているとみなせるため，我々の提案法の近似的高速算法として IS-NMF は有用かつ妥当であるといえる．

4. 半正定値テンソル分解

本章では，半正定値テンソル分解 (Positive Semidefinite Tensor Factorization: PSDTF) とよぶ新しい因子分解法について説明する．NMF は N 個の非負値ベクトル（行列データ）を K 個の非負値ベクトルの凸結合で近似するのと同様に，PSDTF は N 個の半正定値行列（テンソルデータ）を K 個の半正定値行列の凸結合で近似する．したがって，PSDTF は NMF のエレガントな拡張になっており，乗法更新アルゴリズムやノンパラメトリックベイズモデルに基づく基底数の無限化などの様々な技術が応用可能である．

4.1 問題設定

はじめに，我々が取り組む問題について定義する．観測データとして，3 階のテンソル $\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n] \in \mathbb{R}^{M\times M\times N}$ が与えられるものとする．ここで，テンソルの各要素 $\boldsymbol{X}_n \in \mathbb{R}^{M\times M}$ は実対称半正定値行列であるとする．$\boldsymbol{X}_n \in \mathbb{C}^{M\times M}$ が複素エルミート半正定値行列である場合でも同様に扱うことができるが，本稿では簡単のため $\boldsymbol{X}_n \in \mathbb{R}^{M\times M}$ の場合について議論する．

PSDTF の目標は，それぞれの半正定値行列 $\boldsymbol{X}_n$ を K 個の半正定値行列 $\{\boldsymbol{V}_k\}_{k=1}^K$（基底行列とよぶ）の凸結合で近似することである．

$$\boldsymbol{X} \approx \sum_{k=1}^{K} \boldsymbol{h}_k \otimes \boldsymbol{V}_k \stackrel{\text{def}}{=} \boldsymbol{Y} \tag{23}$$

ここで，$\otimes$ はクロネッカー積を表す．式 (23) は，各 n に関する行列和として書きくだすことができる．

$$\boldsymbol{X}_n \approx \sum_{k=1}^{K} h_{kn}\boldsymbol{V}_k \stackrel{\text{def}}{=} \boldsymbol{Y}_n \tag{24}$$

ここで，$h_{kn} \geq 0$ は n 番目の要素 $\boldsymbol{X}_n$ における k 番目の基底行列 $\boldsymbol{V}_k$ の重みである．観測行列 $\boldsymbol{X}_n$ と再構成行列 $\boldsymbol{Y}_n$ との間の誤差 $\mathcal{C}_\phi(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を評価する尺度として，本研究では Bregman 行列ダイバージェンス [14] を利用する．

$$\begin{aligned}&\mathcal{C}_\phi(\boldsymbol{X}_n|\boldsymbol{Y}_n)\\ &= \phi(\boldsymbol{X}_n) - \phi(\boldsymbol{Y}_n) - \mathrm{tr}\big(\nabla\phi(\boldsymbol{Y}_n)^T(\boldsymbol{X}_n - \boldsymbol{Y}_n)\big)\end{aligned} \tag{25}$$

ここで，ϕ は微分可能で厳密に凸な関数である．Bregman 行列ダイバージェンスは常に非負であり，$\boldsymbol{X}_n = \boldsymbol{Y}_n$ であるときに限りゼロをとる．本稿では，その特別な形として，$\phi(\boldsymbol{Z}) = -\log|\boldsymbol{Z}|$ である場合の Log-Determinant (LD) ダイバージェンス [19] を用いる場合に着目する．

$$\begin{aligned}&\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)\\ &= -\log\left|\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right| + \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right) - M\end{aligned} \tag{26}$$

非負の実数 x, y 上に定義される IS ダイバージェンス $\mathcal{C}_{\mathrm{IS}}(x|y) = -\log(x/y) + x/y - 1$ は，$M = 1$ とした LD ダイバージェンスの特別な場合である．

我々の目標は, コスト関数 $\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}|\boldsymbol{Y}) = \sum_n \mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を最小化するような $\boldsymbol{H} = [\boldsymbol{h}_1, \cdots, \boldsymbol{h}_K] \in \mathbb{R}^{N\times K}$ および $\boldsymbol{V} = [\boldsymbol{V}_1, \cdots, \boldsymbol{V}_K] \in \mathbb{R}^{M\times M\times K}$ を求めることである．ただし，$\boldsymbol{H}$ に関しては非負値制約が，$\boldsymbol{V}$ に関しては半正定値制約が課されていることに注意する．このような因子分解法を LD-PSDTF と呼ぶことにする．

4.2 補助関数法

本研究では，解析的な計算を可能にするため，補助関数法 [15] を用いて，コスト関数 $\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ を $\boldsymbol{Y}$（$\boldsymbol{H}$ および $\boldsymbol{V}$）に関して間接的に最小化することを考える．いま，$\mathcal{F}(\boldsymbol{\theta})$ を $\boldsymbol{\theta}$ に関して最小化すべき関数であるとすると，

$$\mathcal{F}(\boldsymbol{\theta}) \leq \mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi}) \tag{27}$$

を満たす $\mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi})$ を $\mathcal{F}(\boldsymbol{\theta})$ の補助関数と呼ぶ．ここで，$\boldsymbol{\phi}$ は補助パラメータである．このとき，以下の反復更新則

$$\boldsymbol{\phi}^{\mathrm{new}} \leftarrow \mathrm{argmin}_{\boldsymbol{\phi}}\mathcal{F}^+(\boldsymbol{\theta}^{\mathrm{old}}, \boldsymbol{\phi}) \tag{28}$$

$$\boldsymbol{\theta}^{\mathrm{new}} \leftarrow \mathrm{argmin}_{\boldsymbol{\theta}}\mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi}^{\mathrm{new}}) \tag{29}$$

を用いると，$\mathcal{F}(\boldsymbol{\theta})$ は単調非増加であることが証明できる．このアルゴリズムの収束性は保証されており，IS-NMF のベイズ学習でも同様の手法が利用されている [17].

$\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ に対する補助関数 $\mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ を得るには，半正定値行列を変数とする関数の凸性および凹性に基づく不等式が必要になる．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．

$$\log|\boldsymbol{Z}| \leq \log|\boldsymbol{\Omega}| + \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{Z}) - M \tag{30}$$

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 [20] を適用可能である．

$$\mathrm{tr}\left(\left(\textstyle\sum_{k=1}^{K}\boldsymbol{Z}_k\right)^{-1}\boldsymbol{A}\right) \leq \sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{Z}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{A}\boldsymbol{\Phi}_k^T\right) \tag{31}$$

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．

式 (30) および式 (31) を用いると，式 (26) に対する補助関数 $\mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を導くことができる．

$$\begin{aligned}
&\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) \overset{c}{=} \log|\boldsymbol{Y}_n| + \mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right)\\
&\leq \log|\boldsymbol{\Omega}_n| + \mathrm{tr}\left(\boldsymbol{Y}_n\boldsymbol{\Omega}_n^{-1}\right) - M\\
&\quad + \textstyle\sum_k \mathrm{tr}\left(\boldsymbol{Y}_{kn}^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right)\\
&= \log|\boldsymbol{\Omega}_n| + \textstyle\sum_k \mathrm{tr}\left(h_{kn}\boldsymbol{V}_k\boldsymbol{\Omega}_n^{-1}\right) - M\\
&\quad + \textstyle\sum_k \mathrm{tr}\left(h_{kn}^{-1}\boldsymbol{V}_k^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right) \overset{\mathrm{def}}{=} \mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)
\end{aligned} \tag{32}$$

ここで，$\boldsymbol{\Omega}_n$ は半正定値行列であり，$\{\boldsymbol{\Phi}_{kn}\}_{k=1}^K$ は足すと単位行列となるような補助行列である．等号が成立する，すなわち，$\mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を最小化するときの条件（補助パラメータの反復更新則）は次式で与えられる．

$$\boldsymbol{\Omega}_n = \boldsymbol{Y}_n \qquad \boldsymbol{\Phi}_{kn} = \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1} \tag{33}$$

4.3 乗法更新アルゴリズム

補助関数 $\mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}|\boldsymbol{Y}) = \sum_n \mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を単調減少させることができる乗法更新アルゴリズムを導く．ここでは，スケールの任意性を除くため，$\mathrm{tr}(\boldsymbol{V}_k) = 1$ という制約をおく（あとで説明する LD-PSDTF のベイズ学習においてはこのような制約は必要ない）．$\mathrm{tr}(\boldsymbol{V}_k) = s$ である場合は，$\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ および $\mathcal{U}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ の値を変化させずに，$\boldsymbol{V}_k \leftarrow \frac{1}{s}\boldsymbol{V}_k$ かつ $h_{kn} \leftarrow sh_{kn}$ と更新できる．

まず，式 (32) を h_{kn} について微分してゼロとおき，式 (33) を代入すると，以下の更新則を得る．

$$h_{kn} \leftarrow h_{kn}\sqrt{\frac{\mathrm{tr}\left(\boldsymbol{Y}_n^{-1}\boldsymbol{V}_k\boldsymbol{Y}_n^{-1}\boldsymbol{X}_n\right)}{\mathrm{tr}\left(\boldsymbol{Y}_n^{-1}\boldsymbol{V}_k\right)}} \tag{34}$$

これは，h_{kn} に非負係数を乗ずる乗法更新則となっており，h_{kn} の非負性は自然に保たれている．

次に，式 (32) を $\boldsymbol{V}_k$ について微分してゼロとおき，式 (33) を代入すると，次式を得る．

$$\boldsymbol{V}_k\boldsymbol{P}_k\boldsymbol{V}_k = \boldsymbol{V}_k^{\mathrm{old}}\boldsymbol{Q}_k\boldsymbol{V}_k^{\mathrm{old}} \tag{35}$$

ここで，$\boldsymbol{V}_k^{\mathrm{old}}$ は $\boldsymbol{V}_k$ の現在の値である．$\boldsymbol{P}_k$ および $\boldsymbol{Q}_k$ は半正定値行列であり，次式で与えられる．

$$\boldsymbol{P}_k = \sum_{n=1}^{N} h_{kn}\boldsymbol{Y}_n^{-1} \quad \boldsymbol{Q}_k = \sum_{n=1}^{N} h_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{X}_n\boldsymbol{Y}_n^{-1} \tag{36}$$

ここで，$\boldsymbol{Q}_k$ は半正定値行列であるので，ある下三角行列 $\boldsymbol{L}_k$ についてコレスキー分解 $\boldsymbol{Q}_k = \boldsymbol{L}_k\boldsymbol{L}_k^T$ が可能である．これを用いると，式 (35) を解析的に解くことができ，最終的に以下の更新則を得る．

$$\boldsymbol{V}_k \leftarrow \boldsymbol{V}_k\boldsymbol{L}_k(\boldsymbol{L}_k^T\boldsymbol{V}_k\boldsymbol{P}_k\boldsymbol{V}_k\boldsymbol{L}_k)^{-\frac{1}{2}}\boldsymbol{L}_k^T\boldsymbol{V}_k \tag{37}$$

行列の半正定値性の定義に従えば，ある行列 $\boldsymbol{A}$ が実対称半正定値行列であることは，$\boldsymbol{A} = \boldsymbol{Z}\boldsymbol{Z}^T$ を満たす実行列 $\boldsymbol{Z}$ が存在することと同値である．したがって，式 (37) において $\boldsymbol{V}_k$ の半正定値性は自然に保たれていることが分かる．

4.4 IS-NMF との関連性

LD-PSDTF は IS-NMF の自然な拡張であるが，$\boldsymbol{X}_n$ および $\boldsymbol{V}_k$ が対角行列であれば，LD-PSDTF は IS-NMF と等しくなる．このとき，任意の半正定値行列の対角成分は非負値ベクトルであるので，式 (26) のコスト関数は

$$\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = \mathcal{C}_{\mathrm{IS}}(\mathrm{diag}(\boldsymbol{X}_n)|\mathrm{diag}(\boldsymbol{Y}_n)) \tag{38}$$

とでき，式 (34) および式 (37) で与えられる乗法更新則は IS-NMF のための収束保証付きの乗法更新則 [15] と一致する．IS-NMF に対しては，Expectation-Maximization (EM) アルゴリズムや収束保証がない乗法更新アルゴリズムも利用可能であることが知られている [15]．LD-PSDTF でも同様であるが，経験的な収束性の良さから，本稿では収束保証付きの乗法更新アルゴリズムを提案した．

LD-PSDTF を音源分離に利用するには，$\boldsymbol{X}_n = \boldsymbol{o}_n\boldsymbol{o}_n^T$ とすればよい．式 (19) で与えられる対数尤度の符号を反転させると，式 (26) で与えられるコスト関数と定数項を除いて等しい．したがって，LD-PSDTF を用いて $\boldsymbol{H}$ および $\boldsymbol{V}$ の最尤推定を行うことができる．ただし，本来の LD-PSDTF はランク 1 に限らず，任意のランクの行列 $\boldsymbol{X}_n$ に対して適用可能であることを明記しておく．

5. 無限半正定値テンソル分解

本章では，理論上は可算無限個の基底行列を内包する無限半正定値テンソル分解 (iPSDTF) について述べる．これまで，適切な基底数 K は事前に指定するものであるとして議論してきたが，実際には容易ではない場合が多い．しかし，異なる K で PSDTF を何度も行い，何らかの基準のも

とで最適な K をあとから選ぶ古典的なモデル選択では，計算コストが非常に高くなる問題があった．一方，ノンパラメトリックベイズモデルに基づく iPSDTF では，観測データに合わせて，無限個存在する基底行列のうちのごく一部のみが実質的にアクティベートされる機構を備えている．

5.1 確率モデルの定式化

LD-PSDTF のベイズモデルについて説明する．まず，式 (24) を少し変形した以下の分解を考える．

$$\boldsymbol{X}_n \approx \sum_{k=1}^{K} \theta_k h_{kn} \boldsymbol{V}_k \stackrel{\text{def}}{=} \boldsymbol{Y}_n \tag{39}$$

ここで，$\theta_k \geq 0$ は k 番目の基底行列の重みであり，$\theta_k = 1$ とすれば通常の LD-PSDTF と一致する．$\boldsymbol{\theta} = [\theta_1, \cdots, \theta_K]^T$ の導入は基底数を無限化する際に重要な役割を果たす．

5.1.1 有限モデル

まず，基底数 K が有限である場合のベイズモデルを設計する．具体的には，$\boldsymbol{H}$ および $\boldsymbol{V}$ を確率変数とみなし，なんらかの事前分布を導入する．$\boldsymbol{\theta}$ については $\theta_k = 1$ としておく．非負値 $h_{kn} \geq 0$ および半正定値行列 $\boldsymbol{V}_k \succeq \boldsymbol{0}$ に対する事前分布としては，ガンマ分布およびウィシャート分布を用いることが一般的である．

$$h_{kn} \sim \mathcal{G}(a_0, b_0) \qquad \boldsymbol{V}_k \sim \mathcal{W}(\nu_0, \boldsymbol{V}_0) \tag{40}$$

ここで，a_0 および b_0 は，ガンマ分布の形状およびレートパラメータであり，ν_0 および $\boldsymbol{V}_0$ はウィシャート分布の自由度および尺度行列である．

与えられた半正定値行列 $\{\boldsymbol{X}_n\}_{n=1}^N$ は，それぞれ独立にウィシャート分布に従うと仮定する．

$$\nu \boldsymbol{X}_n | \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V} \sim \mathcal{W}\left(\nu, \sum_{k=1}^{K} \theta_k h_{kn} \boldsymbol{V}_k\right) \tag{41}$$

ここで，ν はウィシャート分布の自由度である．$\boldsymbol{X}_n$ に ν を乗じる理由は，観測行列 $\boldsymbol{X}_n$ の期待値が再構成行列 $\boldsymbol{Y}_n$ と等しくなるようにするためである ($\mathbb{E}[\boldsymbol{X}_n] = \boldsymbol{Y}_n$)．このとき，$\nu \gg M$ であれば，$\mathbb{M}[\boldsymbol{X}_n] = \frac{\nu - M - 1}{\nu} \boldsymbol{Y}_n \approx \boldsymbol{Y}_n$ となる．一方，$\nu < M$ であれば，$\mathbb{M}[\boldsymbol{X}_n]$ は定義されず，$\boldsymbol{X}_n$ はランク落ちになる．また，$M = \nu = 1$ であれば，式 (41) は，式 (13) で与えられる指数分布と等価になる（IS-NMF と等価）．$\boldsymbol{X}_n$ の対数尤度は次式で与えられる．

$$\begin{aligned}\log p(\boldsymbol{X}_n | \boldsymbol{Y}_n) = & C(\nu) + \frac{\nu - M - 1}{2} \log |\boldsymbol{X}_n| \\ & - \frac{\nu}{2} \log |\boldsymbol{Y}_n| - \frac{\nu}{2} \mathrm{tr}\left(\boldsymbol{X}_n \boldsymbol{Y}_n^{-1}\right)\end{aligned} \tag{42}$$

ここで，$C(\nu)$ は ν のみを含む定数項であり，観測行列 $\boldsymbol{X}_n$ のみに依存する第二項も定数項となる．式 (42) と式 (26) を比較すると，対数尤度 $\log p(\boldsymbol{X}|\boldsymbol{Y}) = \sum_n \log p(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を $\boldsymbol{Y}$ に関して最大化する問題は，コスト関数 $\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}|\boldsymbol{Y}) = \sum_n \mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を最小化する問題と等価である．

5.1.2 無限モデル

次に，基底数が $K \to \infty$ である場合のベイズモデルを設計する．重要なことは，無限個の基底行列が存在するとしても，観測データを表現するのに実質的には少数の基底行列しか利用されないという点である．これを実現するため，IS-NMF に対するノンパラメトリックベイズモデル [17] を参考にする．いま，$K \to \infty$ とすると，基底行列の重みを表す $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_\infty]^T$ は無限次元の非負値ベクトルとなり，そのうちの一部のみが有意に大きな値をもち，それ以外はほとんどゼロとなるようにしたい．したがって，$\boldsymbol{\theta}$ に対する事前分布としてガンマ過程を用いるのが自然である．本研究ではまず，以下のガンマ分布を考える．

$$\theta_k \sim \mathcal{G}(\alpha c / K, \alpha) \tag{43}$$

ここで，$\alpha \geq 0$ および $c \geq 0$ は正の超パラメータであり，$\mathbb{E}_{\mathrm{prior}}[\theta_k] = c/K$ および $\mathbb{E}_{\mathrm{prior}}[\sum_k \theta_k] = c$ となっている．$K \to \infty$ とすれば，以下のガンマ過程が得られる．

$$G \sim \mathrm{GaP}(\alpha, G_0) \tag{44}$$

ここで，G_0 はある空間 $\boldsymbol{\Theta}$ 上に定義された基底測度であり，$G(\boldsymbol{\Theta}) = c$ を満たす（確率測度とは限らない）．サンプルされる G は $\boldsymbol{\Theta}$ 上の離散測度となることが知られており，$\mathbb{E}[G] = G_0$ かつ $\mathbb{V}[G] = G_0/\alpha$ となっている．空間 $\boldsymbol{\Theta}$ の微小区間への分割を $\{\boldsymbol{\Theta}_1, \cdots, \boldsymbol{\Theta}_\infty\}$ とすると，$G(\boldsymbol{\Theta}_k) = \theta_k$ となっている．α は集中度と呼ばれ，α が小さくなるほど $\boldsymbol{\theta}$ はよりスパースになる (偏り $\mathbb{V}[G]$ が大きくなる)．

最終的に，GaP-LD-PSDTF のベイズモデルは，式 (40)，式 (41) および式 (43) で与えられる．計算機上では $K \to \infty$ は扱えないが，K を α に比べて十分大きな値に設定すれば，式 (43) はガンマ過程の良い近似となる．このとき，基底測度 G_0 は一様な測度であることを仮定している．

5.2 変分ベイズ法

我々の目標は，観測データ $\boldsymbol{X}$ が与えられたとき，ベイズの公式を用いて確率変数の事後分布 $p(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}|\boldsymbol{X}) = p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V})/p(\boldsymbol{X})$ を計算することである．しかし，周辺尤度 $p(\boldsymbol{X}) = \iiint p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}) d\boldsymbol{\theta} d\boldsymbol{H} d\boldsymbol{V}$ の計算は解析的に行えないため，変分ベイズ法を用いて $p(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}|\boldsymbol{X})$ を近似的に求めることにする．まず，次式のような因子分解が可能な関数形をもつ分布 $q(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V})$ を考える．

$$q(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}) = \prod_{k=1}^{K} \left(q(\theta_k) \left(\prod_{n=1}^{N} q(h_{kn}) \right) q(\boldsymbol{V}_k) \right) \tag{45}$$

真の事後分布 $p(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}|\boldsymbol{X})$ ではこのような変数間の独立性は成立しないが，$q(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V})$ と $p(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}|\boldsymbol{X})$ の間の KL ダイバージェンスを最小化するような $q(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V})$ を求めたい．これは，対数周辺尤度 $\log p(\boldsymbol{X})$ の変分下限 $\mathcal{L}$ を最大化することと同値であることが知られている．

$$\log p(\boldsymbol{X}) \geq \mathbb{E}[\log p(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})] + \mathbb{E}[\log p(\boldsymbol{\theta})] + \mathbb{E}[\log p(\boldsymbol{H})] + \mathbb{E}[\log p(\boldsymbol{V})] - \mathbb{E}[\log q(\boldsymbol{\theta})] - \mathbb{E}[\log q(\boldsymbol{H})] - \mathbb{E}[\log q(\boldsymbol{V})] \equiv \mathcal{L} \quad (46)$$

ここで，第一項では式 (42) に対する期待値計算が必要であるが，解析的な計算は依然困難である．本研究では，$\mathcal{L} \geq \mathcal{L}'$ となるさらなる変分下限 $\mathcal{L}'$ を設計し，$\mathcal{L}'$ の最大化を通して $\mathcal{L}$ を間接的に最大化することを考える．これは，$-\mathcal{L}'$ を $-\mathcal{L}$ の補助関数とみなした場合の補助関数法となっている．具体的には，式 (30) および式 (31) を用いると，第一項の変分下限は次式で与えられる．

$$\begin{aligned}
&\mathbb{E}[\log p(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})] \\
&\overset{c}{=} -\frac{\nu}{2}\sum_{n=1}^{N}\left(\mathbb{E}[\log|\boldsymbol{Y}_n|] + \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right)\right]\right) \quad (47)\\
&\geq -\frac{\nu}{2}\sum_{n=1}^{N}\left(\log|\boldsymbol{\Omega}_n| + \sum_{k=1}^{K}\mathbb{E}\left[\mathrm{tr}\left(\theta_k h_{kn}\boldsymbol{V}_k\boldsymbol{\Omega}_n^{-1}\right)\right] - M \right.\\
&\qquad\left. + \sum_{k=1}^{K}\mathbb{E}\left[\mathrm{tr}\left(\theta_k^{-1}h_{kn}^{-1}\boldsymbol{V}_k^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right)\right]\right)
\end{aligned}$$

ここで，$\boldsymbol{\Omega}_n$ は任意の半正定値行列，$\{\boldsymbol{\Phi}_{kn}\}_{k=1}^K$ は足すと単位行列となるような補助行列である．等号が成立するときの条件は次式で与えられる．

$$\boldsymbol{\Omega}_n = \textstyle\sum_k \mathbb{E}[\boldsymbol{Y}_{kn}] \quad (48)$$

$$\boldsymbol{\Phi}_{kn} = \left(\mathbb{E}[\boldsymbol{Y}_{kn}^{-1}]\right)^{-1}\left(\textstyle\sum_{k'}\left(\mathbb{E}[\boldsymbol{Y}_{k'n}^{-1}]\right)^{-1}\right)^{-1} \quad (49)$$

このとき，$\mathcal{L}'$ を単調増加させるには，各因子に関する変分事後分布を次式に従って順番に更新すればよい．

$$\begin{aligned}
q(\boldsymbol{\theta}) &\propto p(\boldsymbol{\theta})\exp(\mathbb{E}_{q(\boldsymbol{H},\boldsymbol{V})}[\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})])\\
q(\boldsymbol{H}) &\propto p(\boldsymbol{H})\exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{V})}[\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})]) \quad (50)\\
q(\boldsymbol{V}) &\propto p(\boldsymbol{V})\exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{H})}[\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})])
\end{aligned}$$

ここで，$\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})$ は $\log p(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})$ の変分下限であり，式 (47) から期待値計算を省いたものである．

5.3 変分事後分布の計算

式 (40)，式 (41) および式 (43) で定義されたベイズモデルに対して変分ベイズ法を用いて変分事後分布を求めるためには，事前分布と尤度関数との共役性を確立する必要がある．しかし，θ_k に着目すると，ガンマ事前分布の対数は θ_k および $\log\theta_k$ に関する項を含んでいるが，式 (47) で与えられる対数尤度の変分下限は θ_k および θ_k^{-1} に関する項を含んでいる．このような不一致があると，事後分布の形は事前分布と同じガンマ分布にはならない．

本研究では，$q(\theta_k)$ および $q(h_{kn})$ は一般化逆ガウス (Generalized Inverse Gaussian: GIG) 分布となることが分かる．非負の実数上に定義された GIG 分布は次式で与えられる．

$$\mathrm{GIG}(x|\gamma,\rho,\tau) = \frac{(\rho/\tau)^{\frac{\gamma}{2}}}{2K_\gamma(\sqrt{\rho\tau})}x^{\gamma-1}e^{-\frac{1}{2}(\rho x+\tau x^{-1})} \quad (51)$$

ここで，γ および $\rho,\tau \geq 0$ はパラメータであり，K_γ は第二種変形ベッセル関数を表す．式 (51) の対数は $\log x$，x および x^{-1} の項を含んでいる．ガンマ分布は GIG 分布の特別な場合であるので，一種の共役性が成立している．この性質は LD-PSDTF の特別な場合である IS-NMF のベイズ推定 [17] においても利用されている．また，期待値 $\mathbb{E}[x]$ および $\mathbb{E}[x^{-1}]$ は次式で計算可能である．

$$\mathbb{E}[x] = \frac{\sqrt{\tau}K_{\gamma+1}(\sqrt{\rho\tau})}{\sqrt{\rho}K_\gamma(\sqrt{\rho\tau})} \quad \mathbb{E}\left[\frac{1}{x}\right] = \frac{\sqrt{\rho}K_{\gamma-1}(\sqrt{\rho\tau})}{\sqrt{\tau}K_\gamma(\sqrt{\rho\tau})} \quad (52)$$

一方，$q(\boldsymbol{V}_k)$ は行列 GIG (Matrix GIG: MGIG) 分布 [21] となることが分かる．MGIG 分布は半正定値行列上に定義された分布であり，次式で与えられる．

$$\mathrm{MGIG}(\boldsymbol{X}|\gamma,\boldsymbol{R},\boldsymbol{T}) = \frac{2^{\gamma M}}{|\boldsymbol{T}|^{\gamma}B_\gamma(\boldsymbol{R}\boldsymbol{T}/4)}|\boldsymbol{X}|^{\gamma-\frac{M+1}{2}}\exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{R}\boldsymbol{X}+\boldsymbol{T}\boldsymbol{X}^{-1}\right)\right) \quad (53)$$

ここで，γ は実数であり，$\boldsymbol{R},\boldsymbol{T} \succeq \boldsymbol{0}$ は半正定値行列である．M は行列 $\boldsymbol{X}$ のサイズであり，B_γ は第二種行列ベッセル関数 [22] を表す．式 (53) の対数は $\log|\boldsymbol{X}|$，$\boldsymbol{X}$ および $\boldsymbol{X}^{-1}$ の項を含んでいる．ウィシャート分布や逆ウィシャート分布は，MGIG 分布の特別な場合である [23]．ただし，現在のところ期待値 $\mathbb{E}[\boldsymbol{X}]$ および $\mathbb{E}[\boldsymbol{X}^{-1}]$ を解析的に計算することは困難であるため，モンテカルロ法に基づく近似計算法が提案されている [11, 24]．

上記議論を踏まえて，式 (50) に従って変分事後分布を計算すると，最終的に以下を得る．

$$\begin{aligned}
q(\theta_k) &= \mathrm{GIG}(\theta_k|\gamma_k^\theta,\rho_k^\theta,\tau_k^\theta)\\
q(h_{kn}) &= \mathrm{GIG}(h_{kn}|\gamma_{kn}^h,\rho_{kn}^h,\tau_{kn}^h) \quad (54)\\
q(\boldsymbol{V}_k) &= \mathrm{MGIG}(\boldsymbol{V}_k|\gamma_k^V,\boldsymbol{R}_k^V,\boldsymbol{T}_k^V)
\end{aligned}$$

このときの変分パラメータは以下で与えられる．

$$\begin{aligned}
\gamma_k^\theta &= \alpha c/K\\
\rho_k^\theta &= 2\alpha + \nu\sum_{n=1}^{N}\mathrm{tr}\left(\mathbb{E}[h_{kn}\boldsymbol{V}_k]\,\boldsymbol{\Omega}_n^{-1}\right)\\
\tau_k^\theta &= \nu\sum_{n=1}^{N}\mathrm{tr}\left(\mathbb{E}\left[h_{kn}^{-1}\boldsymbol{V}_k^{-1}\right]\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right)\\
\gamma_{kn}^h &= a_0\\
\rho_{kn}^h &= 2b_0 + \nu\mathrm{tr}\left(\mathbb{E}[\theta_k\boldsymbol{V}_k]\,\boldsymbol{\Omega}_n^{-1}\right)\\
\tau_{kn}^h &= \nu\mathrm{tr}\left(\mathbb{E}\left[\theta_k^{-1}\boldsymbol{V}_k^{-1}\right]\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right) \quad (55)\\
\gamma_k^V &= \nu_0/2\\
\boldsymbol{R}_k^V &= \boldsymbol{V}_0^{-1} + \nu\sum_{n=1}^{N}\mathbb{E}[\theta_k h_{kn}]\,\boldsymbol{\Omega}_n^{-1}\\
\boldsymbol{T}_k^V &= \nu\sum_{n=1}^{N}\mathbb{E}\left[\theta_k^{-1}h_{kn}^{-1}\right]\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T
\end{aligned}$$

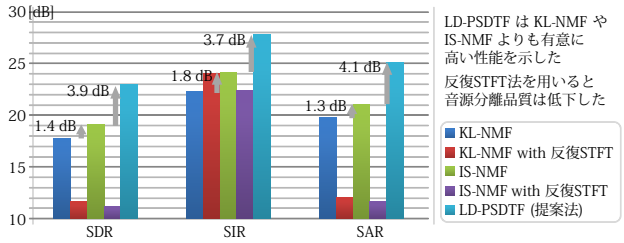

図 3 音源分離実験結果

6. 評価

LD-PSDTF を用いた音源分離実験について報告する．本稿では，4 章で述べた乗法更新アルゴリズムに基づく最尤推定を用いるものとし，5 章で述べた変分ベイズ法に基づくベイズ推定については [11] を参考にされたい．

6.1 実験条件

実験には，RWC 研究用音楽データベース：楽器音 [25] 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，LD-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 ($M = 512$, $N = 1400$)．比較のため，振幅スペクトログラムに対する KL-NMF とパワースペクトログラムに対する IS-NMF も評価した（2.3 節および 2.4 節）．位相情報を復元するため，反復 STFT 法 [6] を用いる場合も評価した．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 [26] 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．

6.2 実験結果

図 3 に示す通り，音源分離タスクにおいて LD-PSDTF は NMF に対する明確な優位性を示した．SDR, SIR, SAR の平均は，KL-NMF では 17.7 dB, 22.2 dB, 19.7 dB，IS-NMF では 19.1 dB, 24.0 dB, 21.0 dB であったのに対し，LD-PSDTF では 23.0 dB, 27.7 dB, 25.1 dB であった[*3]．反復 STFT 法を用いると，音源信号の品質は低下した．このことは，[10] に示唆されている通り，スペクトログラムの無矛盾性の向上が必ずしも音響信号の品質の向上につながらないことを意味する．図 4 および図 5 に示す通り，LD-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 $\boldsymbol{V}$ および音量変化 $\boldsymbol{H}$ を適切に推定することができた．ここで，各基底行列 $\boldsymbol{V}_k$ における斜め縞の間隔は周期を表しており，$\boldsymbol{V}_k$ 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，3.4 節で議論した通り，周波数領域においては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．LD-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，観測行列である $\{\boldsymbol{X}_n\}_{n=1}^N$ は格子状の分布をもっており，斜行型の分布をもつ基底行列 $\{\boldsymbol{V}_k\}_{k=1}^K$ の凸結合では十分な近似となっていないように思える．しかし，LD ダイバージェンスは $\boldsymbol{Y}_n$ が $\boldsymbol{X}_n$ を過大評価しても小さなコストしか課さないことから，妥当な分解となっている．LD-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 IS-NMF で LD-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．

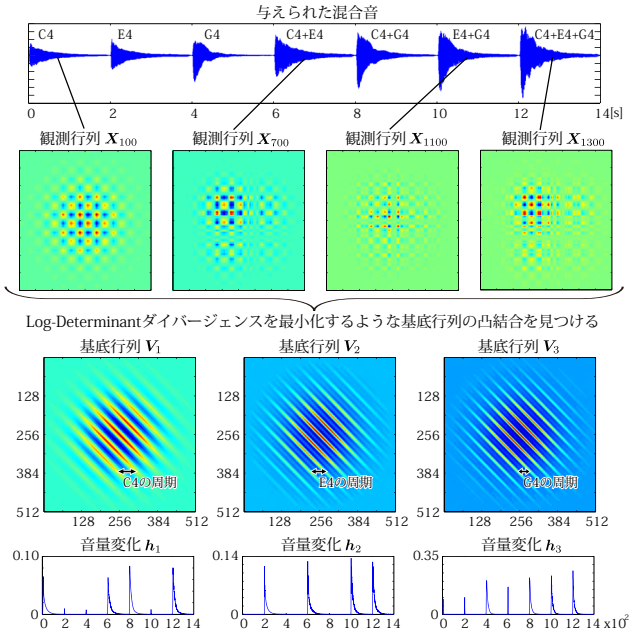

図 4 LD-PSDTF によるピアノ信号の分解結果

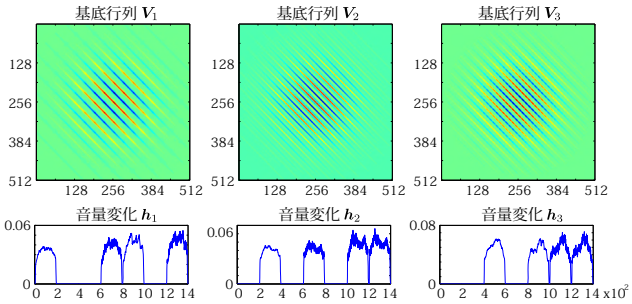

図 5 LD-PSDTF によるクラリネット信号の分解結果

[*3] 著者の WEB サイトにおいてサンプルファイルが試聴可能．

7. おわりに

本稿では，Log-Determinant 半正定値テンソル分解 (LD-PSDTF) と呼ぶ新しい因子分解法を提案した．LD-PSDTF は IS-NMF の自然な拡張となっていることを示し，IS-NMF と同様に収束性が保証された乗法更新則を用いて最尤推定が行えることを示した．また，ガンマ過程に基づくノンパラメトリックベイズモデルを提案し，観測データに合わせて基底数を自動調節する枠組みについて論じた．LD-PSDTF を音源分離に適用すると，KL-NMF や IS-NMF より優れた音源分離結果が得られることを実験で確認した．

今後の研究にはいくつかの興味深い方向性が考えられる．まず，IS-NMF の拡張である複合自己回帰モデル [27,28] と同様に，LD-PSDTF の枠組みにソース・フィルタ理論を組み込むことで，音高ではなく音色に基づく高品質な音源分離の実現に取り組みたい．さらに，PSDTF の興味深いもう一つの変種である von Neumann (vN) ダイバージェンスに基づく PSDTF（式 (25) において $\phi(\boldsymbol{Z}) = \mathrm{tr}(\boldsymbol{Z}\log\boldsymbol{Z} - \boldsymbol{Z})$ となるとき）に対して，最尤推定法およびベイズ推定法を確立したい．vN-PSDTF は KL-NMF の自然な拡張となっているため，幅広い分野への応用が期待できる．

謝辞: 本研究の一部は，JSPS 科研費 23700184, MEXT 科研費 25870192, JST CREST OngaCREST の支援を受けた．

参考文献

[1] K. Itoyama, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. Query-by-example music information retrieval by score-informed source separation and remixing technologies. *EURASIP Journal on Advances in Signal Processing*, Article ID 172961, 2010.

[2] M. Goto. Active music listening interfaces based on signal processing. *ICASSP*, volume 4, pp. 1441–1444, 2007.

[3] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno. Drumix: An audio player with real-time drum-part rearrangement functions for active music listening. *IPSJ Digital Courier*, 3:134–144, 2007.

[4] N. Sturmel, A. Liutkus, J. Pinel, L. Girin, S. Marchand, G. Richard, R. Badeau, and L. Daudet. Linear mixing models for active listening of music productions in realistic studio conditions. *AES Convention*, 2012.

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

[6] D. W. Griffin and J. S. Lim. Signal estimation from modified short-time Fourier transform. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 32(2):236–243, 1984.

[7] J. Le Roux, H. Kameoka, N. Ono, and S. Sagayama. Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction. *SAPA*, pp. 23–28, 2008.

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

[9] H. Kameoka, T. Nishimoto, and S. Sagayama. Complex NMF: A new sparse representation for acoustic signals. *ICASSP*, pp. 45–48, 2009.

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

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

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

[13] 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.

[14] L. M. Bregman. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7(3):200–217, 1967.

[15] 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. *MLSP*, pp. 283–288, 2010.

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

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

[18] J. B. Allen and L. R. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *IEEE*, 65(11):1558–1564, 1977.

[19] B. Kulis, M. Sustik, and I. Dhillon. Low-rank kernel learning with Bregman matrix divergences. *Journal of Machine Learning Research (JMLR)*, 10:341–376, 2009.

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

[21] O. Barndorff-Nielsen, P. Blæsild, J. L. Jensen, and B. Jørgensen. Exponential transformation models. *Royal Society of London*, 379(1776):41–65, 1982.

[22] C. S. Herz. Bessel functions of matrix argument. *Annals of Mathematics*, 61(3):474–523, 1955.

[23] R. W. Butler. Generalized inverse Gaussian distributions and their Wishart connections. *Scandinavian Journal of Statistics*, 25(1):69–75, 1998.

[24] M. Yang, Y. Li, and Z. Zhang. Multi-task learning with Gaussian matrix generalized inverse Gaussian model. *ICML*, pp. 423–431, 2013.

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

[26] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 14(4):1462–1469, 2006.

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

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