社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

［ポスター講演］音楽音響信号解析のためのガンマ過程に基づく無限半正定値テンソル分解

吉井　和佳†　　富岡　亮太††　　持橋　大地†††　後藤　真孝†

† 産業技術総合研究所　〒 305–8568 茨城県つくば市梅園 1–1–1
†† 東京大学　〒 113–8654 東京都文京区本郷 7–3–1
††† 統計数理研究所　〒 190–8518 東京都立川市緑町 10–3
E-mail: †{k.yoshii,m.goto}@aist.go.jp, ††tomioka@mist.i.u-tokyo.ac.jp, †††daichi@ism.ac.jp

あらまし　本稿では，非負値行列分解 (NMF) の自然な拡張となっている半正定値テンソル分解 (PSDTF) を提案し，モノラル信号の音源分離に応用すると優れた性能を発揮することを示す．従来は，与えられた混合音のパワースペクトログラムを非負値行列とみなし，NMF を用いて少数の基底スペクトルと時間方向のアクティベーションとの積に分解する方法が一般的であった．このとき，非負値しか扱えないという NMF の制約上，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扱えず，音源信号を再合成するには分離されたパワースペクトログラムに対して位相復元を行う必要があった．本研究では，PSDTF を用いることで，位相情報を陽に取り扱うことなくモノラル信号を時間領域で直接分離することができる画期的な方法を提案する．また，PSDTF は NMF と同程度に実装が容易で，ガンマ過程に基づいて無限個の基底を許容するノンパラメトリックベイズ拡張も可能であることを示す．
キーワード　音楽音響信号解析，非負値行列分解，音源分離，ガンマ過程，ノンパラメトリックベイズ

[Poster Presentation] Infinite Positive Semidefinite Tensor Factorization based on Gamma Process for Music Signal Analysis

Kazuyoshi YOSHII†, Ryota TOMIOKA††, Daichi MOCHIHASHI†††, and Masataka GOTO†

† National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
†† The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
††† The Institute of Statistical Mathematics (ISM), 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan
E-mail: †{k.yoshii,m.goto}@aist.go.jp, ††tomioka@mist.i.u-tokyo.ac.jp, †††daichi@ism.ac.jp

Abstract　This paper presents positive semidefinite tensor factorization (PSDTF) that is a natural extension of nonnegative matrix factorization (NMF) and shows that PSDTF outperforms NMF in source separation of single-channel audio signals. We propose a PSDTF-based method that can directly separate a given mixture signal into source signals in the time domain without explicitly dealing with phase information. In addition, a nonparametric Bayesian extension of PSDTF is feasible by using the gamma process in a similar manner to NMF.
Key words　Music signal analysis, NMF, source separation, gamma process, nonparametric Bayes

1. はじめに

信号処理分野における重要な研究課題の一つは，モノラル音響信号に対する音源分離である．現在のところ，非負値行列分解 (Nonnegative Matrix Factorization: NMF) [1] に基づく方法が最もよく利用されている．NMF は，入力となる非負値行列（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いう不適切な仮定がおかれることが一般的であった．この問題を解決するため，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しようと多数の研究 [2, 3] が行われてきたが，いまなお周波数領域における位相復元は容易ではない．

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) [4] 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．時間領域に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陽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，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少数の定常な基底信号の存在を考える．与えられた混合音は，局所的にはそれら基底信号の線形和で構成され，結合係数が非定常に時変化すると仮定する．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を行う．ガンマ過程を用いると，実効的な基底数を自動決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ともできる．

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．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) のよく知られた 2 つの変種である KL-NMF [5] および IS-NMF [6] について説明する．

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 $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} \stackrel{\mathrm{def}}{=} \boldsymbol{Y}$ で近似することである．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．各 n に関するベクトル和として書きくだすと次式を得る．

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

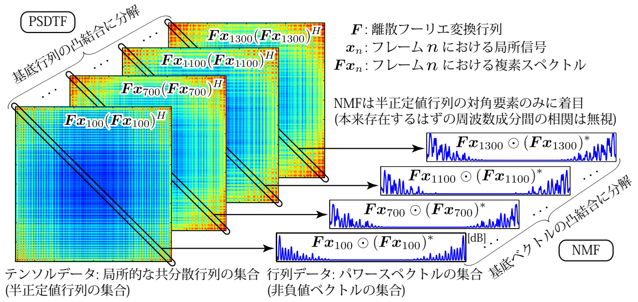

図 1　周波数領域における音源分離：PSDTF は NMF の自然な拡張

ここで，$\boldsymbol{y}_{kn} = h_{kn}\boldsymbol{w}_k$ とすると，$\boldsymbol{y}_n = \sum_k \boldsymbol{y}_{kn}$ 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 [7] が広く利用されている．

$$\mathcal{C}_\phi(\boldsymbol{x}_n|\boldsymbol{y}_n) = \phi(\boldsymbol{x}_n) - \phi(\boldsymbol{y}_n) - \phi'(\boldsymbol{y}_n)^T(\boldsymbol{x}_n - \boldsymbol{y}_n) \tag{2}$$

ここで，ϕ は厳密に凸な関数である．この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 [8] 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 [9, 10]．

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．いま，混合音の「非負」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，混合音の複素スペクトログラム $\boldsymbol{S} = [\boldsymbol{s}_1, \cdots, \boldsymbol{s}_N] \in \mathbb{C}^{M \times N}$，$k$ 番目の音源信号の複素スペクトログラム $\boldsymbol{S}_k = [\boldsymbol{s}_{k1}, \cdots, \boldsymbol{s}_{kn}] \in \mathbb{C}^{M \times N}$ を考える．混合音が K 個の音源信号の瞬時混合であれば

$$\boldsymbol{S} = \sum_{k=1}^{K} \boldsymbol{S}_k \quad i.e., \quad \boldsymbol{s}_n = \sum_{k=1}^{K} \boldsymbol{s}_{kn} \tag{3}$$

が成り立つ．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．

$$\mathbb{E}[s_{knm}|s_{nm}] = \frac{y_{knm}}{y_{nm}} s_{nm} = \frac{w_{km}h_{kn}}{\sum_k w_{km}h_{kn}} s_{nm} \tag{4}$$

この処理はウィナーフィルタと呼ばれ，$\boldsymbol{S}_k$ 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．重畳加算合成法 [11] を用いば，$\mathbb{E}[\boldsymbol{S}_k|\boldsymbol{S}]$ から k 番目の音源信号を復元できる．

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 [5]．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

$$\mathcal{C}_{\mathrm{KL}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_{m=1}^{M} \left(x_{nm} \log \frac{x_{nm}}{y_{nm}} - x_{nm} + y_{nm} \right) \tag{5}$$

で与えられる．ただし，式 (5) はスケール不変ではないことに注意する．このことは，混合音のスペクトル $\boldsymbol{x}_n$ の音量を変化させると音源分離結果 $\boldsymbol{y}_n$ も変化することを意味する．

確率モデルとして解釈するには，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．

$$|s_{knm}|\,|y_{knm} \sim \mathrm{Poisson}(y_{knm}) \tag{6}$$

ここで，式 (3) から本来は $s_{nm} = \sum_k s_{knm}$ であるが，振幅の加法性が成立するという強い仮定のもとでは $|s_{nm}| = \sum_k |s_{knm}|$ が成立する．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．

$$x_{nm}|y_{nm} \sim \mathrm{Poisson}(y_{nm}) \tag{7}$$

$|s_{nm}| = \sum_k |s_{knm}|$ 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 $\mathbb{E}[|s_{knm}|\,||s_{nm}|] = y_{knm}y_{nm}^{-1}|s_{nm}|$ となり，振幅の加法性（位相の保存性）から最終的に式 (4) 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 [6]．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

$$\mathcal{C}_{\mathrm{IS}}(\boldsymbol{x}_n|\boldsymbol{y}_n) = \sum_{m=1}^{M} \left(\frac{x_{nm}}{y_{nm}} - \log \frac{x_{nm}}{y_{nm}} - 1 \right) \tag{8}$$

で与えられる．式 (8) 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．

確率モデルとして解釈するには，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．

$$s_{knm}|y_{knm} \sim \mathcal{N}_c(0, y_{knm}) \tag{9}$$

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 $s_{nm}|y_{nm} \sim \mathcal{N}_c(0, y_{nm})$ が成立する．このことから，$x_{nm} = |s_{nm}|^2$ は指数分布に従うことが分かる．

$$x_{nm}|y_{nm} \sim \mathrm{Exponential}(y_{nm}) \tag{10}$$

$s_{nm} = \sum_k s_{knm}$ 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 (4) で，分散は次式で求めることができる．

$$\mathbb{V}[s_{knm}|s_{nm}] = y_{knm} - y_{knm}y_{nm}^{-1}y_{knm} \tag{11}$$

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，混合音の生成モデルを提案し，潜在変数である音源信号を推定する方法について議論する．

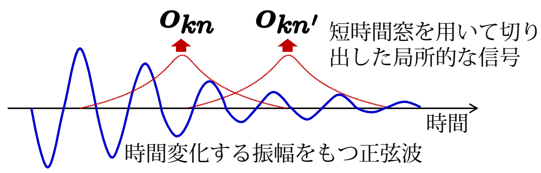

図 2　局所信号 $\boldsymbol{o}_{kn}$ および $\boldsymbol{o}_{kn'}$ は異なる位相や振幅をもっているが，同一の周期に従っていることに着目する．

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フレーム n における局所的な信号である．このとき，$\boldsymbol{o}_n$ の分解

$$\boldsymbol{o}_n = \sum_{k=1}^{K} \boldsymbol{o}_{kn} \tag{12}$$

を考える．$\boldsymbol{o}_{kn}$ は k 番目の音源信号から切り出されたフレーム n 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 (12) の時間領域表現は式 (3) の周波数領域表現と等価である．

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 $\{\boldsymbol{o}_{kn}\}_{n=1}^{N}$ は異なるが，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，音源信号が時変な振幅をもつ正弦波であれば，局所信号 $\boldsymbol{o}_{kn}$ および $\boldsymbol{o}_{kn'}$ $(n \neq n')$ の位相や振幅は異なるが，振動周期は定常である．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」から切り出されたフレーム m における局所信号であり，π_{kn} はその係数である．これらは非負値でなくともよい．

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 [11] を用いて得ることができる．

3.2 確率モデルの定式化

式 (12) に対する確率モデルを定式化したい．基底信号の定常性から局所信号 $\{\boldsymbol{\phi}_{kn}\}_{n=1}^{N}$ 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事前分布を考えることができる．

$$\boldsymbol{\phi}_{kn}|\boldsymbol{V} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_k), \tag{13}$$

ここで，共分散行列 $\boldsymbol{V}_k \in \mathbb{R}^{M \times M}$ は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して振動するため，平均ベクトルは $\boldsymbol{0}$ とした．式 (13) は，任意の連続する M 点における周辺分布がガウス分布となることを意味しており，基底信号が従う確率分布はカーネル $\boldsymbol{V}_k$ をもつ定常ガウス過程に他ならない．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 (13) から $\boldsymbol{o}_{kn}$ もやはりガウス分布に従うことが分かる．

$$\boldsymbol{o}_{kn}|\boldsymbol{\pi},\boldsymbol{V}\sim\mathcal{N}(\boldsymbol{0},\pi_{kn}^2\boldsymbol{V}_k) \tag{14}$$

さらに，式 (12) 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．

$$\boldsymbol{o}_n|\boldsymbol{\pi},\boldsymbol{V}\sim\mathcal{N}\left(\boldsymbol{0},\sum_{k=1}^{K}\pi_{kn}^2\boldsymbol{V}_k\right) \tag{15}$$

式 (15) 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，$h_{kn}=\pi_{kn}^2\geqq 0$，$\boldsymbol{X}_n=\boldsymbol{o}_n\boldsymbol{o}_n^T\succeq\boldsymbol{0}$，$\boldsymbol{Y}_n=\sum_k h_{kn}\boldsymbol{V}_k\succeq\boldsymbol{0}$ とすると(注1)，式 (15) から $\boldsymbol{X}_n$ の対数尤度は次式で与えられる(注2)．

$$\log p(\boldsymbol{X}_n|\boldsymbol{Y}_n)\stackrel{c}{=}-\frac{1}{2}\log|\boldsymbol{Y}_n|-\frac{1}{2}\mathrm{tr}(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}) \tag{16}$$

いま，観測データとしてテンソル $\boldsymbol{X}=[\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n]\in\mathbb{R}^{M\times M\times N}$ 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 $\boldsymbol{V}=[\boldsymbol{V}_1,\cdots,\boldsymbol{V}_K]\in\mathbb{R}^{M\times M\times K}$ を求めたい．これは，4. 章に示すように，入力 $\boldsymbol{X}_n$ 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に基づく最尤推定が行える．

3. 3 確率モデルに基づく音源分離

最後に，$\boldsymbol{H}$ および $\boldsymbol{V}$ を用いて局所信号 $\boldsymbol{o}_{kn}=\pi_{kn}\boldsymbol{\phi}_{kn}$ を確率的な枠組みのもとで推定する．π_{kn} や $\boldsymbol{\phi}_{kn}$ を陽に扱うことなく，式 (14) および式 (15) 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．

$$\mathbb{E}[\boldsymbol{o}_{kn}|\boldsymbol{o}_n]=\boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{o}_n \tag{17}$$

$$\mathbb{V}[\boldsymbol{o}_{kn}|\boldsymbol{o}_n]=\boldsymbol{Y}_{kn}-\boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{Y}_{kn} \tag{18}$$

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 (17) は，観測信号 $\boldsymbol{o}_n$ のうち，位相や波形を陽に考えることなく，カーネル $\boldsymbol{V}_k$ に対応す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 (17) および式 (18) は，周波数領域における式 (4) および式 (11) と非常に類似した形式をもつ．ただし，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，従来法ではベクトル $\boldsymbol{s}_{kn}$ 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）．

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）．式 (15) から，観測信号 $\boldsymbol{o}_n$ の線形写像である複素スペクトル $\boldsymbol{F}\boldsymbol{o}_n$ もまた複素ガウス分布に従うことが分かる．

$$\boldsymbol{F}\boldsymbol{o}_n|\boldsymbol{H},\boldsymbol{V}\sim\mathcal{N}_c\left(\boldsymbol{0},\sum_{k=1}^{K}h_{kn}\boldsymbol{F}\boldsymbol{V}_k\boldsymbol{F}^H\right) \tag{19}$$

(注1)：$\boldsymbol{\Psi}$ が半正定値行列であるとき $\boldsymbol{\Psi}\succeq\boldsymbol{0}$ と書く．

(注2)：$\stackrel{c}{=}$ は定数項を除いて等号が成立することを示す．

式 (19) は式 (9) 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．$\boldsymbol{V}_k$ が巡回行列であれば $\boldsymbol{F}\boldsymbol{V}_k\boldsymbol{F}^H$ は対角行列となり（$\boldsymbol{V}_k$ は対角化されるという），このとき式 (19) 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 (9) 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．PSDTF は NMF の自然な拡張になっており，乗法更新アルゴリズムの導出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が可能である．

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 $\boldsymbol{X}=[\boldsymbol{X}_1,\cdots,\boldsymbol{X}_n]\in\mathbb{R}^{M\times M\times N}$ を考える．本稿では，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が，$\boldsymbol{X}_n\in\mathbb{C}^{M\times M}$ が複素エルミート半正定値行列である場合も同様の取り扱いが可能である．

我々の目標は，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）の凸結合で近似することである．

$$\boldsymbol{X}_n\approx\sum_{k=1}^{K}h_{kn}\boldsymbol{V}_k\stackrel{\mathrm{def}}{=}\boldsymbol{Y}_n \tag{20}$$

ここで，$h_{kn}\geqq 0$ 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 [7] を利用する．

$$\begin{aligned}&\mathcal{C}_\phi(\boldsymbol{X}_n|\boldsymbol{Y}_n)\\&=\phi(\boldsymbol{X}_n)-\phi(\boldsymbol{Y}_n)-\mathrm{tr}\Big(\nabla\phi(\boldsymbol{Y}_n)^T(\boldsymbol{X}_n-\boldsymbol{Y}_n)\Big)\end{aligned} \tag{21}$$

ここで，ϕ は厳密に凸な関数である．この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，$\phi(\boldsymbol{Z})=-\log|\boldsymbol{Z}|$ である場合の Log-Determinant (LD) ダイバージェンス [12] を用いる場合を考える．

$$\mathcal{C}_{\mathrm{LD}}(\boldsymbol{X}_n|\boldsymbol{Y}_n)=-\log\left|\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right|+\mathrm{tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right)-M \tag{22}$$

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．

いま，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 $\boldsymbol{V} = [\boldsymbol{V}_1, \cdots, \boldsymbol{V}_K] \in \mathbb{R}^{M\times M\times K}$ を求めたい．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．

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 [8] を用いて，コスト関数 $\mathcal{C}_{\rm LD}(\boldsymbol{X}|\boldsymbol{Y})$ 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，

$$\mathcal{F}(\boldsymbol{\theta}) \leqq \mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi}) \tag{23}$$

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

$$\boldsymbol{\phi}^{\rm new} \leftarrow {\rm argmin}_{\boldsymbol{\phi}} \mathcal{F}^+(\boldsymbol{\theta}^{\rm old}, \boldsymbol{\phi}) \tag{24}$$

$$\boldsymbol{\theta}^{\rm new} \leftarrow {\rm argmin}_{\boldsymbol{\theta}} \mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi}^{\rm new}) \tag{25}$$

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 [10].

$\mathcal{C}_{\rm LD}(\boldsymbol{X}|\boldsymbol{Y})$ に対する補助関数 $\mathcal{U}_{\rm LD}(\boldsymbol{X}|\boldsymbol{Y})$ を得るには，半正定値行列を変数とする不等式を用いる．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．

$$\log|\boldsymbol{Z}| \leqq \log|\boldsymbol{\Omega}| + {\rm tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{Z}) - M \tag{26}$$

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 $g(\boldsymbol{Z}) = {\rm tr}(\boldsymbol{Z}^{-1}\boldsymbol{A})$ は凸関数であるから，澤田らの不等式 [13] を適用可能である．

$${\rm tr}\left(\left(\sum_{k=1}^K \boldsymbol{Z}_k\right)^{-1} \boldsymbol{A}\right) \leqq \sum_{k=1}^K {\rm tr}\left(\boldsymbol{Z}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{A}\boldsymbol{\Phi}_k^T\right) \tag{27}$$

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 (26) および式 (27) を用いると，式 (22) に対する補助関数 $\mathcal{U}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ を導くことができる．

$$\begin{aligned}
&\mathcal{C}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n) \stackrel{c}{=} \log|\boldsymbol{Y}_n| + {\rm tr}\left(\boldsymbol{X}_n\boldsymbol{Y}_n^{-1}\right)\\
&\leqq \log|\boldsymbol{\Omega}_n| + {\rm tr}\left(\boldsymbol{Y}_n\boldsymbol{\Omega}_n^{-1}\right) - M\\
&\quad + \textstyle\sum_k {\rm tr}\left(\boldsymbol{Y}_{kn}^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right)\\
&= \log|\boldsymbol{\Omega}_n| + \textstyle\sum_k {\rm tr}\left(h_{kn}\boldsymbol{V}_k\boldsymbol{\Omega}_n^{-1}\right) - M\\
&\quad + \textstyle\sum_k {\rm tr}\left(h_{kn}^{-1}\boldsymbol{V}_k^{-1}\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\right) \stackrel{\rm def}{=} \mathcal{U}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n)
\end{aligned} \tag{28}$$

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，$\mathcal{U}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ 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．

$$\boldsymbol{\Omega}_n = \boldsymbol{Y}_n \qquad \boldsymbol{\Phi}_{kn} = \boldsymbol{Y}_{kn}\boldsymbol{Y}_n^{-1} \tag{29}$$

4.3 乗法更新アルゴリズム

補助関数 $\mathcal{U}_{\rm LD}(\boldsymbol{X}|\boldsymbol{Y}) = \sum_n \mathcal{U}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ 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．ただし，スケールの任意性を除くため，${\rm tr}(\boldsymbol{V}_k) = 1$ という制約をおく（LD-PSDTF のベイズ学習においては不要）．${\rm tr}(\boldsymbol{V}_k) = s$ である場合は，$\mathcal{C}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ および $\mathcal{U}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n)$ の値を変化させずに，$\boldsymbol{V}_k \leftarrow \frac{1}{s}\boldsymbol{V}_k$ かつ $h_{kn} \leftarrow s h_{kn}$ と更新できる．

まず，式 (28) を h_{kn} について微分してゼロとおき，式 (29) を代入すると，以下の更新則を得る．

$$h_{kn} \leftarrow h_{kn}\sqrt{\frac{{\rm tr}\left(\boldsymbol{Y}_n^{-1}\boldsymbol{V}_k\boldsymbol{Y}_n^{-1}\boldsymbol{X}_n\right)}{{\rm tr}\left(\boldsymbol{Y}_n^{-1}\boldsymbol{V}_k\right)}} \tag{30}$$

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 (28) を $\boldsymbol{V}_k$ について微分してゼロとおき，式 (29) を代入すると，次式を得る．

$$\boldsymbol{V}_k\boldsymbol{P}_k\boldsymbol{V}_k = \boldsymbol{V}_k^{\rm old}\boldsymbol{Q}_k\boldsymbol{V}_k^{\rm old} \tag{31}$$

ここで，$\boldsymbol{V}_k^{\rm old}$ 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．

$$\boldsymbol{P}_k = \sum_{n=1}^N h_{kn}\boldsymbol{Y}_n^{-1} \quad \boldsymbol{Q}_k = \sum_{n=1}^N h_{kn}\boldsymbol{Y}_n^{-1}\boldsymbol{X}_n\boldsymbol{Y}_n^{-1} \tag{32}$$

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．式 (31) は解析的に解くことができ，最終的に以下の更新則を得る．

$$\boldsymbol{V}_k \leftarrow \boldsymbol{V}_k\boldsymbol{L}_k(\boldsymbol{L}_k^T\boldsymbol{V}_k\boldsymbol{P}_k\boldsymbol{V}_k\boldsymbol{L}_k)^{-\frac{1}{2}}\boldsymbol{L}_k^T\boldsymbol{V}_k \tag{33}$$

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 (33) 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 の自然な拡張であり，$\boldsymbol{X}_n$ 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 (22) のコスト関数は

$$\mathcal{C}_{\rm LD}(\boldsymbol{X}_n|\boldsymbol{Y}_n) = \mathcal{C}_{\rm IS}({\rm diag}(\boldsymbol{X}_n)|{\rm diag}(\boldsymbol{Y}_n)) \tag{34}$$

とでき，式 (30) および式 (33) 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 [8] と一致する．

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．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実質的に必要なだけの基底を学習することができる．

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 (20) を少し変形した以下の分解を考える．

$$\boldsymbol{X}_n \approx \sum_{k=1}^K \theta_k h_{kn}\boldsymbol{V}_k \stackrel{\rm def}{=} \boldsymbol{Y}_n \tag{35}$$

ここで，$\theta_k \geqq 0$ 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．$\boldsymbol{\theta}$ については $\theta_k = 1$ としておく．非負値 $h_{kn} \geqq 0$ 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．

$$h_{kn} \sim \mathcal{G}(a_0, b_0) \qquad \boldsymbol{V}_k \sim \mathcal{W}(\nu_0, \boldsymbol{V}_0) \tag{36}$$

ここで，a_0 および b_0 は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．

$$\nu \boldsymbol{X}_n | \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V} \sim \mathcal{W}\left(\nu, \sum_{k=1}^K \theta_k h_{kn} \boldsymbol{V}_k\right) \tag{37}$$

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 $\boldsymbol{X}_n$ の期待値を再構成行列 $\boldsymbol{Y}_n$ と等しくするためである ($\mathbb{E}[\boldsymbol{X}_n] = \boldsymbol{Y}_n$)．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 (37) は式 (10) で与えられる指数分布と等価になる（IS-NMF）．したがって，$\boldsymbol{X}_n$ の対数尤度は

$$\log p(\boldsymbol{X}_n | \boldsymbol{Y}_n) = C(\nu) + \frac{\nu - M - 1}{2} \log |\boldsymbol{X}_n| - \frac{\nu}{2} \log |\boldsymbol{Y}_n| - \frac{\nu}{2} \mathrm{tr}\left(\boldsymbol{X}_n \boldsymbol{Y}_n^{-1}\right) \tag{38}$$

で与えられる．ここで，$C(\nu)$ は ν のみを含む定数項であり，$\boldsymbol{X}_n$ のみに依存する第二項も定数項となる．式 (38) と式 (22) 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

基底数が $K \to \infty$ である場合のベイズモデルを設計するうえで，無限個の基底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しか利用されないようにしたい．これを実現するため，IS-NMF のノンパラメトリックベイズ拡張 [10] を参考にする．いま，$K \to \infty$ とすると，基底の重みを表す $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_\infty]^T$ は無限次元の非負ベクトルとなり，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．

$$\theta_k \sim \mathcal{G}(\alpha c / K, \alpha) \tag{39}$$

ここで，$\alpha \geqq 0$ および $c \geqq 0$ 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．

$$G \sim \mathrm{GaP}(\alpha, G_0) \tag{40}$$

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．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 (36), 式 (37) および式 (39) 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 (39) 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．

$$q(\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V}) = \prod_{k=1}^K \left(q(\theta_k) \left(\prod_{n=1}^N q(h_{kn}) \right) q(\boldsymbol{V}_k) \right) \tag{41}$$

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．

$$\begin{aligned} \log p(\boldsymbol{X}) &\geqq \mathbb{E}[\log p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V})] \\ &+ \mathbb{E}[\log p(\boldsymbol{\theta})] + \mathbb{E}[\log p(\boldsymbol{H})] + \mathbb{E}[\log p(\boldsymbol{V})] \\ &- \mathbb{E}[\log q(\boldsymbol{\theta})] - \mathbb{E}[\log q(\boldsymbol{H})] - \mathbb{E}[\log q(\boldsymbol{V})] \equiv \mathcal{L} \end{aligned} \tag{42}$$

ここで，第一項は式 (38) に対する期待値計算であるが，解析的な計算は依然困難である．本研究では，$\mathcal{L} \geqq \mathcal{L}'$ 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 (26) および式 (27) を用いると，第一項の変分下限は次式で与えられる．

$$\begin{aligned} &\mathbb{E}[\log p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{V})] \\ &\stackrel{c}{=} -\frac{\nu}{2} \sum_{n=1}^N \left(\mathbb{E}[\log |\boldsymbol{Y}_n|] + \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{X}_n \boldsymbol{Y}_n^{-1}\right)\right] \right) \qquad (43) \\ &\geqq -\frac{\nu}{2} \sum_{n=1}^N \left(\log |\boldsymbol{\Omega}_n| + \sum_{k=1}^K \mathbb{E}\left[\mathrm{tr}\left(\theta_k h_{kn} \boldsymbol{V}_k \boldsymbol{\Omega}_n^{-1}\right)\right] - M \right. \\ &\qquad \left. + \sum_{k=1}^K \mathbb{E}\left[\mathrm{tr}\left(\theta_k^{-1} h_{kn}^{-1} \boldsymbol{V}_k^{-1} \boldsymbol{\Phi}_{kn} \boldsymbol{X}_n \boldsymbol{\Phi}_{kn}^T\right)\right] \right) \end{aligned}$$

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．等号成立条件は次式となる．

$$\boldsymbol{\Omega}_n = \textstyle\sum_k \mathbb{E}[\boldsymbol{Y}_{kn}] \tag{44}$$

$$\boldsymbol{\Phi}_{kn} = \left(\mathbb{E}\left[\boldsymbol{Y}_{kn}^{-1}\right]\right)^{-1} \left(\textstyle\sum_{k'} \left(\mathbb{E}\left[\boldsymbol{Y}_{k'n}^{-1}\right]\right)^{-1} \right)^{-1} \tag{45}$$

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．

$$
\begin{aligned}
q(\boldsymbol{\theta}) &\propto p(\boldsymbol{\theta}) \exp(\mathbb{E}_{q(\boldsymbol{H},\boldsymbol{V})}[\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})]) \\
q(\boldsymbol{H}) &\propto p(\boldsymbol{H}) \exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{V})}[\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})]) \\
q(\boldsymbol{V}) &\propto p(\boldsymbol{V}) \exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{H})}[\log q(\boldsymbol{X}|\boldsymbol{\theta},\boldsymbol{H},\boldsymbol{V})])
\end{aligned}
\tag{46}
$$

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 (43) 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 (36)，式 (37) および式 (39) 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 (43) 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．

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．

$$
\mathrm{GIG}(x|\gamma,\rho,\tau) = \frac{(\rho/\tau)^{\frac{\gamma}{2}}}{2K_\gamma(\sqrt{\rho\tau})} x^{\gamma-1} e^{-\frac{1}{2}(\rho x + \tau x^{-1})} \tag{47}
$$

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 (47) 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 においても利用されている [10]．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．

$$
\mathbb{E}[x] = \frac{\sqrt{\tau}K_{\gamma+1}(\sqrt{\rho\tau})}{\sqrt{\rho}K_\gamma(\sqrt{\rho\tau})} \quad \mathbb{E}\left[\frac{1}{x}\right] = \frac{\sqrt{\rho}K_{\gamma-1}(\sqrt{\rho\tau})}{\sqrt{\tau}K_\gamma(\sqrt{\rho\tau})} \tag{48}
$$

一方，$q(\boldsymbol{V}_k)$ は行列 GIG (Matrix GIG: MGIG) 分布 [14] 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．

$$
\begin{aligned}
\mathrm{MGIG}(\boldsymbol{X}|\gamma,\boldsymbol{R},\boldsymbol{T}) = &\frac{2^{\gamma M}}{|\boldsymbol{T}|^\gamma B_\gamma(\boldsymbol{R}\boldsymbol{T}/4)} |\boldsymbol{X}|^{\gamma-\frac{M+1}{2}} \\
&\exp\left(-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{R}\boldsymbol{X} + \boldsymbol{T}\boldsymbol{X}^{-1}\big)\right)
\end{aligned}
\tag{49}
$$

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 [15] を表す．式 (49) 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 [16]．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 [4,17]．

上記議論を踏まえて，式 (46) 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．

$$
\begin{aligned}
q(\theta_k) &= \mathrm{GIG}(\theta_k|\gamma_k^\theta, \rho_k^\theta, \tau_k^\theta) \\
q(h_{kn}) &= \mathrm{GIG}(h_{kn}|\gamma_{kn}^h, \rho_{kn}^h, \tau_{kn}^h) \\
q(\boldsymbol{V}_k) &= \mathrm{MGIG}(\boldsymbol{V}_k|\gamma_k^V, \boldsymbol{R}_k^V, \boldsymbol{T}_k^V)
\end{aligned}
\tag{50}
$$

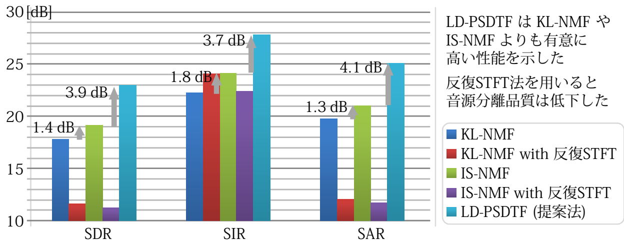

図 3　音源分離実験結果

このときの変分パラメータは以下で与えられる．

$$
\begin{aligned}
\gamma_k^\theta &= \alpha c/K, \quad \rho_k^\theta = 2\alpha + \nu\textstyle\sum_{n=1}^N \mathrm{tr}\big(\mathbb{E}[h_{kn}\boldsymbol{V}_k]\,\boldsymbol{\Omega}_n^{-1}\big) \\
\tau_k^\theta &= \nu\textstyle\sum_{n=1}^N \mathrm{tr}\big(\mathbb{E}\big[h_{kn}^{-1}\boldsymbol{V}_k^{-1}\big]\,\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\big) \\
\gamma_{kn}^h &= a_0, \quad \rho_{kn}^h = 2b_0 + \nu\mathrm{tr}\big(\mathbb{E}[\theta_k\boldsymbol{V}_k]\,\boldsymbol{\Omega}_n^{-1}\big) \\
\tau_{kn}^h &= \nu\mathrm{tr}\Big(\mathbb{E}\big[\theta_k^{-1}\boldsymbol{V}_k^{-1}\big]\,\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T\Big) \\
\gamma_k^V &= \nu_0/2, \quad \boldsymbol{R}_k^V = \boldsymbol{V}_0^{-1} + \nu\textstyle\sum_{n=1}^N \mathbb{E}[\theta_k h_{kn}]\,\boldsymbol{\Omega}_n^{-1} \\
\boldsymbol{T}_k^V &= \nu\textstyle\sum_{n=1}^N \mathbb{E}\big[\theta_k^{-1}h_{kn}^{-1}\big]\,\boldsymbol{\Phi}_{kn}\boldsymbol{X}_n\boldsymbol{\Phi}_{kn}^T
\end{aligned}
\tag{51}
$$

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. 章で述べた最尤推定を用いるものとし，5. 章で述べたベイズ推定については [4] 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 [18] 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（16[kHz]）．

次に，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，$\boldsymbol{X}_n = \boldsymbol{o}_n\boldsymbol{o}_n^T$ とした．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 [2] を用いる場合も評価した．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 [19] 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(注3)．反復 STFT 法を用いるとスペクトログラムの無矛盾性は向上するが，音源信号

(注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．

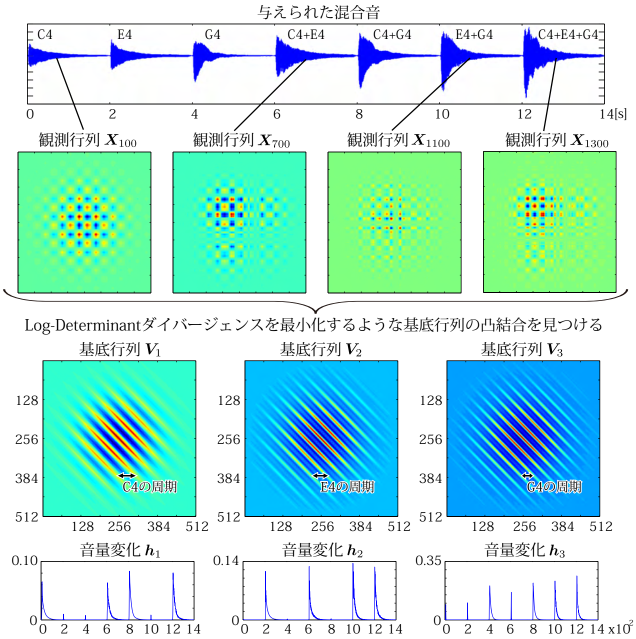

図 4　LD-PSDTF によるピアノ信号の分解結果

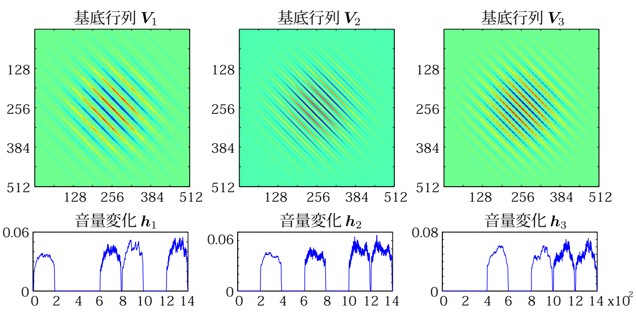

図 5　LD-PSDTF によるクラリネット信号の分解結果

の品質は低下した．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．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．

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おり，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た．

今後の課題として，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 (21) 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，音響信号処理だけではなく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．

文　献

[1] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. *NIPS*, pp. 556–562, 2000.
[2] 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.
[3] J. Le Roux *et al.* Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction. *SAPA*, pp. 23–28, 2008.
[4] K. Yoshii *et al.* Infinite positive semidefinite tensor factorization for source separation of mixture signals. *ICML*, pp. 576–584, 2013.
[5] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. *WASPAA*, pp. 177–180, 2003.
[6] C. Févotte *et al.* Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009.
[7] 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 Comp. Math. and Math. Physics*, 7(3):200–217, 1967.
[8] M. Nakano *et al.* Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta divergence. *MLSP*, pp. 283–288, 2010.
[9] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009:Article ID 785152, 2009.
[10] M. Hoffman *et al.* Bayesian nonparametric matrix factorization for recorded music. *ICML*, pp. 439–446, 2010.
[11] J. Allen and L. Rabiner. A unified approach to short-time Fourier analysis and synthesis. *IEEE*, 65(11):1558–1564, 1977.
[12] B. Kulis *et al.* Low-rank kernel learning with Bregman matrix divergences. *JMLR*, 10:341–376, 2009.
[13] H. Sawada *et al.* Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization. *ICASSP*, pp. 261–264, 2012.
[14] O. Barndorff-Nielsen *et al.* Exponential transformation models. *Royal Society of London*, 379(1776):41–65, 1982.
[15] C. S. Herz. Bessel functions of matrix argument. *Annals of Mathematics*, 61(3):474–523, 1955.
[16] R. W. Butler. Generalized inverse Gaussian distributions and their Wishart connections. *Scandinavian Journal of Statistics*, 25(1):69–75, 1998.
[17] M. Yang *et al.* Multi-task learning with Gaussian matrix generalized inverse Gaussian model. *ICML*, 2013.
[18] M. Goto *et al.* RWC music database: Music genre database and musical instrument sound database. *ISMIR*, 2003.
[19] E. Vincent *et al.* Performance measurement in blind audio source separation. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 14(4):1462–1469, 2006.