Vol.2017-MUS-116 No.6
2017/8/24

モノラル音響信号に対する音源分離のための無限相関テンソル分解

吉井 和佳[1,2,a)]

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究極の因子分解法について述べる．従来は，非負値行列分解 (NMF) を用いて，混合音の各時刻におけるパワースペクトルを，少数の基底スペクトルの線形和で近似することがよく行われてきた．近年では，複素スペクトルに含まれる位相情報を適切に取り扱うため，半正定値テンソル分解 (PSDTF) が提案されている．PSDTF は，NMF の本質的な拡張となっており，混合音の各時刻における複素スペクトルから計算される局所的な共分散行列を，少数の基底共分散行列の線形和で近似する．この方法により，複素スペクトログラムにおける周波数方向の相関を考慮することができるようになったが，依然として時間方向の独立性が仮定されていた．複素スペクトログラムの周波数軸と時間軸を入れ替えて PSDTF を利用することも可能であるが，時間方向の相関は考慮できても，周波数方向の独立性を仮定せざるを得なかった．この問題を解決するため，本研究では，周波数方向の相関と時間方向との相関を同時に取り扱うことができる相関テンソル分解 (CTF) を提案する．CTF は，PSDTF のさらなる拡張となっており，複素スペクトログラムを構成するすべての時間周波数ビン間の共分散行列を，周波数方向の共分散行列と時間方向の共分散行列とのクロネッカー積の和で近似する．この方法により，すべての時間周波数ビン間の相関を考慮したウィナーフィルタを用いて，音源信号の複素スペクトログラムを一挙に推定することが可能となる．CTF は，NMF や PSDTF と同様に EM アルゴリズムあるいは補助関数法を用いた最尤推定や，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）が可能である．実験の結果，高品質な音源分離ができることを確かめた．

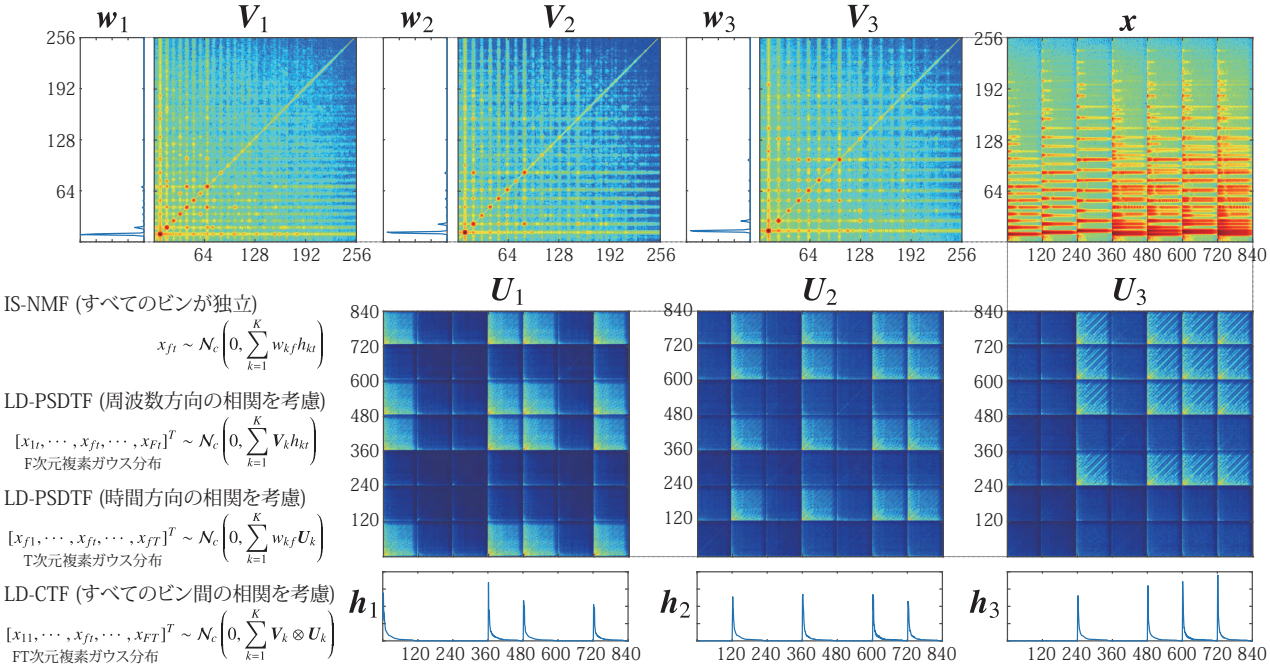
1. はじめに

音楽音響信号に対する音源分離は，音楽情報処理分野における基盤技術のひとつである．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，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] 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めるシステム [1–4] を実現することができる．このとき，分離音の品質の向上は重要な課題である．

モノラル音響信号に対する音源分離を行うには，非負値行列分解 (Nonnegative Matrix Factorization, NMF) [5] を用いるのが主流である．NMF は，入力となる非負値行列（非負値スペクトログラム）を二つの非負値行列（基底スペクトルの集合と対応する音量ベクトルとの集合）の積で近似することができる．NMF には多くの変種が存在するが，混合音の複素スペクトログラム中の時間周波数ビンはすべて独立であり（現実には成立しない），それぞれが異なる複素ガウス分布に従うという仮定のもとでは，混合音のパワースペクトログラムに対して Itakura-Saito (IS) ダイバージェンスに基づく NMF (IS-NMF) を適用することが理論的に妥当である．IS-NMF の結果に基づくウィナーフィルタを用いると，時間周波数ビンごとに独立に，混合音の複素成分を音源信号の複素成分の和に分解することができる．このとき，混合音と音源信号の複素スペクトログラムの位相は同一にならざるを得ず，復元される時間領域の音源信号の品質には限界があった．

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試みはいくつか存在する．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短時間フーリエ変換 (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無矛盾性を評価する関数を提案し，それを最大化するアルゴリズムを導出している [8]．一方，亀岡ら [9] は，混合音の複素スペクトログラムを取り扱うことができる複素 NMF を提案し，無矛盾性に関する評価関数 [10] を組み込む拡張を行っている．ここ

[1] 京都大学 大学院情報学研究科 知能情報学専攻
Yoshida-hoftachi, Sakyo, Kyoto, Kyoto 606-8501, Japan
[2] 理研 革新知能統合研究センター (AIP) 音響情景理解チーム
15F, 1-4-1 Nihonbashi, Chuo, Tokyo 103-0027, Japan
[a)] yoshii@kuis.kyoto-u.ac.jp



*混合音の複素スペクトログラム $\boldsymbol{x} \in \mathbb{C}^{FT}$ に対し，IS-NMF は，周波数方向の非負値ベクトル群 $\boldsymbol{W} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3 \in \mathbb{R}_+^F\}$ および時間方向の非負値ベクトル群 $\boldsymbol{H} = \{\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3 \in \mathbb{R}_+^T\}$ を推定する．LD-PSDTF は，周波数方向の半正定値行列群 $\boldsymbol{V} = \{\boldsymbol{V}_1, \boldsymbol{V}_2, \boldsymbol{V}_3 \in \mathcal{S}_+^F\}$ および時間方向の非負値ベクトル群 $\boldsymbol{H}$，あるいは周波数方向の非負値ベクトル群 $\boldsymbol{W}$ および時間方向の半正定値行列群 $\boldsymbol{U} = \{\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_3 \in \mathcal{S}_+^T\}$ を推定する．LD-CTF は，周波数方向の半正定値行列群 $\boldsymbol{V}$ および時間方向の半正定値行列群 $\boldsymbol{U}$ を推定する（図中では $F = 256, T = 840$）．

図 1 時間周波数領域での音源分離における IS-NMF, LD-PSDTF, LD-CTF の比較．

で，複素スペクトログラムが無矛盾であることが、対応する時間領域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されたい．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．

近年，複素スペクトログラムに含まれる位相情報を取り扱うため，半正定値テンソル分解 (Positive Semidefinite Tensor Factorization, PSDTF) [11, 12] と呼ばれる因子分解法が提案されている．NMF は，**非負値ベクトルを少数の非負値基底ベクトルの線形和で近似**するのに対し，PSDTF は，**半正定値行列を少数の半正定値基底行列の線形和で近似**する．PSDTF にも様々な変種が考えられるが，LogDet (LD) ダイバージェンスに基づく PSDTF (LD-PSDTF) が IS-NMF の自然な拡張となっている．音源分離において，IS-NMF は，各時刻の混合音のパワースペクトルを，各音源の典型的なパワースペクトルパターンの線形和で近似するのに対し，LD-PSDTF は，各時刻の混合音の複素スペクトルから計算される共分散行列を，各音源の複素スペクトルの共分散行列の線形和で近似する．ここで，共分散行列の対角成分がパワースペクトルに対応することに着目されたい．LD-PSDTF では，周波数間の相関を考慮するウィナーフィルタを用いて，時間ごとに独立に，混合音の複素スペクトルを音源信号の複素スペクトルの和に分解することができる．実は，このような周波数領域での分解は，時間領域での分解に対応しており，高品質な時間領域の音源信号を推定できる理由となっている．

実際の音響信号の複素スペクトログラムには，周波数方向だけではなく，時間方向にも相関が存在する．理論上は，定常信号に対してフーリエ変換を用いると，周波数間は独立となることが知られている．そこで，実際の非定常な音響信号に対しては，局所的な定常性を仮定して STFT を行うのが一般的である．しかし，厳密には，窓関数を用いて切り出された短時間信号は定常ではないため，対応する複素スペクトルにおいては，倍音関係や隣接関係にある周波数ビン間には強い相関が発生することが避けられない．また，音響信号は時系列データであるという性質上，時間方向にも相関が存在する．しかし，LD-PSDTF では，周波数方向の相関を考慮することができるが，時間方向の独立性が仮定されていた．混合音の複素スペクトログラムの周波数軸と時間軸を入れ替えて LD-PSDTF を利用することも可能であるが，時間方向の相関は考慮できても，周波数方向の独立性を仮定せざるを得なかった．

本稿では，NMF，PSDTF に続く究極の因子分解技法として，時間周波数上の相関を完全に取り扱える相関テンソル分解 (Correlated Tensor Factorization, CTF) を提案する．CTF も NMF や PSDTF と同様に様々な変種を考えることができ，本稿では，LD ダイバージェンスに基づく CTF (LD-CTF) を取り上げる．**図 1** に IS-NMF，LD-PSDTF，LD-CTF の比較を示す．LD-CTF は，**半正定値行列を，少数の半正定値行列のペアのクロネッカー積の和で近似**する，すなわち，混合音の複素スペクトログラム中の全ての時間

周波数ビン間の超大規模な共分散行列を，各音源に対応する周波数方向の共分散行列と時間方向の共分散行列のクロネッカー積の和で近似する．この結果を用いると，混合音の複素スペクトログラムに対し，全ての時間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，音源の複素スペクトログラムを一挙に得ることができる．

LD ダイバージェンス最小化に基づく LD-CTF は，全ての時間周波数ビン上の超高次元の多変量複素ガウス分布の共分散行列パラメータ（周波数方向の共分散行列と時間方向の共分散行列の少数のペア）の最尤推定を行うことと等価である．本稿では，IS-NMF や LD-PSDTF と同様に，収束性の保証された Expectation-Maximization (EM) アルゴリズムおよび Minorization-Maximization (MM) アルゴリズムを導出する．実際には，適切な音源数を事前に設定することが困難である場合が多い．そこで，観測データに合わせて音源数を自動的に調整するため，ガンマ過程に基づくノンパラメトリックベイズモデル（無限モデル）の定式化を行う．さらに，パラメータの事後分布を推定するため，補助関数を用いた変分ベイズ法を導出する．

理論的には，LD-CTF は究極の分解技法ではあるが，数百万次元を超えるような巨大行列の計算は現実的ではないため，計算量削減のための工夫が必要不可欠である．まず，周波数方向および時間方向の共分散行列をともにブロック対角行列に限定した LD-CTF の近似モデルを提案する．これは，局所的な時間周波数領域に限定して完全な相関を考慮することに相当する．また，これまで時間周波数領域における LD-CTF について議論してきたが，複素スペクトログラムの周波数軸あるいは時間軸に対して，任意の線形変換を施した空間（例：観測音響信号の各フレーム信号を並べただけの時間時間領域や複素スペクトログラムの時間軸に対してさらにフーリエ変換を施した周波数周波数領域など）においても等価な分解が得られる．このことから，周波数方向の共分散行列群を同時にほぼ対角化できるような線形変換と，また，時間方向の共分散行列群を同時にほぼ対角化できるような線形変換をそれぞれ見つけることができれば，変換後の空間における IS-NMF が時間周波数領域での LD-CTF の良い近似になると考えられる．

2. 因子分解と音源分離

本章では，従来の因子分解技法である NMF や PSDTF について説明する．特に，音源分離を行ううえで，IS-NMF [13] や LD-PSDTF [11] の理論的な裏付けについて説明する．

2.1 非負値行列分解 (NMF)

NMF では，非負値行列 $\boldsymbol{X} \in \mathbb{R}_+^{F\times T}$ を二つの非負値行列 $\boldsymbol{W} \in \mathbb{R}_+^{K\times F}$ および $\boldsymbol{H} \in \mathbb{R}_+^{K\times T}$ の積 $\boldsymbol{Y}$ で近似する．

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

ただし，$K \ll \min(F,T)$ は基底数，$\boldsymbol{Y} \in \mathbb{R}_+^{F\times T}$ は低ランクな再構成行列を表す．行列積表現である式 (1) は，各要素ごとに書きくだすことができる．

$$x_{ft} \approx y_{ft} \stackrel{\text{def}}{=} \sum_{k=1}^{K} w_{kf}h_{kt} \tag{2}$$

ここで，$y_{kft} = w_{kf}h_{kt}$ と定義すると，$y_{ft} = \sum_k y_{kft}$ が成立する．観測値 x_{ft} と再構成値 y_{ft} との間の誤差 $\mathcal{C}(x_{ft}|y_{ft})$ を評価する尺度のひとつに，Bregman ダイバージェンス [14] が知られている．

$$\mathcal{C}_\phi(x_{ft}|y_{ft}) = \phi(x_{ft}) - \phi(y_{ft}) - \phi'(y_{ft})(x_{ft} - y_{ft}) \tag{3}$$

ここで，ϕ は厳密に凸な関数である．常に $\mathcal{C}_\phi(x_{ft}|y_{ft}) \geq 0$ が成り立ち，$x_{ft} = y_{ft}$ であるときに限り $\mathcal{C}_\phi(x_{ft}|y_{ft}) = 0$ となる．その特別な場合として，$\phi(x) = x\log x - x$ の場合の Kullback-Leibler (KL) ダイバージェンスや，$\phi(x) = -\log x$ の場合の IS ダイバージェンスがよく知られている．

$$\mathcal{C}_{\text{KL}}(x_{ft}|y_{ft}) = x_{ft}\log\frac{x_{ft}}{y_{ft}} - x_{ft} + y_{ft} \tag{4}$$

$$\mathcal{C}_{\text{IS}}(x_{ft}|y_{ft}) = \frac{x_{ft}}{y_{ft}} - \log\frac{x_{ft}}{y_{ft}} - 1 \tag{5}$$

NMF のコスト関数 $\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y})$ は次式で与えられる．

$$\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y}) = \sum_{f=1}^{F}\sum_{t=1}^{T}\mathcal{C}_\phi(x_{ft}|y_{ft}) \tag{6}$$

このコスト関数を最小化する $\boldsymbol{W}$ および $\boldsymbol{H}$ を求めるため，乗法更新 (Multiplicative Update, MU) 型の MM アルゴリズムが提案されている [15]．これは，ある確率モデルの最尤推定を行うことと等価である．一方，ガンマ事前分布を導入してベイズ推定を行うこともできる [16,17]．

2.2 IS-NMF を用いた音源分離

与えられた混合音の複素スペクトログラム $\boldsymbol{S} \in \mathbb{C}^{F\times T}$（$F$ は周波数ビン数，T はフレーム数）を K 個の音源信号の複素スペクトログラムの和に分解することを考える．音源 k の複素スペクトログラムを $\boldsymbol{Z}_k \in \mathbb{C}^{F\times T}$ とし，周波数領域での瞬時混合過程を仮定すると，以下が成り立つ．

$$\boldsymbol{S} = \sum_{k=1}^{K}\boldsymbol{Z}_k \tag{7}$$

IS-NMF では，潜在変数 z_{kft} が，y_{kft} 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．

$$z_{kft}|y_{kft} \sim \mathcal{N}_c(0, y_{kft}) \tag{8}$$

ここで，$s_{ft} = \sum_k z_{kft}$ かつ $y_{ft} = \sum_k y_{kft}$ であり，複素ガウス分布の再生性から，s_{ft} も複素ガウス分布に従う．

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

これは，s_{ft} の位相に関わらず，時刻 t・周波数 f におけるパワー $x_{ft} \stackrel{\text{def}}{=} |s_{ft}|^2$ が指数分布に従うことと等価である．

$$x_{ft}|y_{ft} \sim \text{Exponential}(y_{ft}) \tag{10}$$

最終的に，観測される複素スペクトログラム $\boldsymbol{S}$ に対する対数尤度関数を導出できる．

$$\begin{aligned}\log p(\boldsymbol{S}|\boldsymbol{Y}) &= \sum_{f=1}^{F}\sum_{t=1}^{T}\log p(s_{ft}|y_{ft}) \\ &= \sum_{f=1}^{F}\sum_{t=1}^{T}\left(-\frac{x_{ft}}{y_{ft}} - \log y_{ft}\right) \\ &\stackrel{c}{=} -\mathcal{C}_{\text{IS}}(\boldsymbol{X}|\boldsymbol{Y})\end{aligned} \tag{11}$$

したがって，対数尤度関数の最大化 $p(\boldsymbol{S}|\boldsymbol{Y})$ は，IS ダイバージェンス $\mathcal{C}_{\text{IS}}(\boldsymbol{X}|\boldsymbol{Y})$ の最小化と等価である．

パラメータ $\boldsymbol{Y}$（$\boldsymbol{W}$ および $\boldsymbol{H}$）が推定できれば，式 (8) および式 (9) から，観測変数 $\boldsymbol{S}$ が与えられたもとで，潜在変数 $\boldsymbol{Z}_k$ の事後分布を求めることができる．

$$p(z_{kft}|s_{ft}) = \mathcal{N}_c\left(z_{kft}\middle|y_{kft}y_{ft}^{-1}s_{ft}, y_{ft} - y_{kft}^2y_{ft}^{-1}\right) \tag{12}$$

この処理はウィナーフィルタと呼ばれ，$\boldsymbol{Z}_k$ と $\boldsymbol{S}$ の位相は同一となる．これは NMF に基づく音源分離の性質であり，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．最後に，重畳加算合成法 [18] を用いれば，$\mathbb{E}[\boldsymbol{Z}_k|\boldsymbol{S}]$ から音源信号を復元することができる．

2.3 半正定値テンソル分解 (PSDTF)

PSDTF では，観測データとして，特別な形式を持つ 3 階のテンソル $\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_t] \in \mathbb{C}^{F\times F\times T}$ が与えられるものとする．ここで，各要素 $\boldsymbol{X}_t \in \mathcal{S}_+^F$ は半正定値行列であるとする．PSDTF の目標は，それぞれの半正定値行列 $\boldsymbol{X}_t$ を K 個の半正定値行列 $\{\boldsymbol{V}_k \in \mathcal{S}_+^F\}_{k=1}^K$（基底行列とよぶ）の線形和で近似することである．

$$\boldsymbol{X}_t \approx \boldsymbol{Y}_t \stackrel{\text{def}}{=} \sum_{k=1}^{K} h_{kt}\boldsymbol{V}_k \tag{13}$$

ここで，$\boldsymbol{Y}_{kt} = h_{kt}\boldsymbol{V}_k$ と定義すると，$\boldsymbol{Y}_t = \sum_k \boldsymbol{Y}_{kt}$ が成立する．$h_{kt} \geq 0$ は n 番目の要素 $\boldsymbol{X}_t$ における基底行列 $\boldsymbol{V}_k$ の重みである．観測行列 $\boldsymbol{X}_t$ と再構成行列 $\boldsymbol{Y}_t$ との間の誤差 $\mathcal{C}_\phi(\boldsymbol{X}_t|\boldsymbol{Y}_t)$ を評価する尺度として，Bregman 行列ダイバージェンス [14] が利用できる．

$$\begin{aligned}\mathcal{C}_\phi(\boldsymbol{X}_t|\boldsymbol{Y}_t) =& \phi(\boldsymbol{X}_t) - \phi(\boldsymbol{Y}_t) \\ &- \text{tr}\big(\nabla\phi(\boldsymbol{Y}_t)^T(\boldsymbol{X}_t - \boldsymbol{Y}_t)\big)\end{aligned} \tag{14}$$

ここで，ϕ は微分可能で厳密に凸な関数である．式 (14) は式 (3) の自然な多次元拡張となっており，常に $\mathcal{C}_\phi(\boldsymbol{X}_t|\boldsymbol{Y}_t) \geq 0$ が成り立ち，$\boldsymbol{X}_t = \boldsymbol{Y}_t$ であるときに限り $\mathcal{C}_\phi(\boldsymbol{X}_t|\boldsymbol{Y}_t) = 0$ となる．その特別な場合として，$\phi(\boldsymbol{Z}) = \text{tr}(\boldsymbol{X}\log\boldsymbol{X} - \boldsymbol{X})$ の場合の von Neumann (vN) ダイバージェンスや，$\phi(\boldsymbol{Z}) = -\log|\boldsymbol{Z}|$ の場合の LogDet (LD) ダイバージェンスなどが知られている [19]．

$$\mathcal{C}_{\text{vN}}(\boldsymbol{X}_t|\boldsymbol{Y}_t) = \text{tr}\left(\boldsymbol{X}\log\boldsymbol{X} - \boldsymbol{X}\log\boldsymbol{Y} - \boldsymbol{X} + \boldsymbol{Y}\right) \tag{15}$$

$$\mathcal{C}_{\text{LD}}(\boldsymbol{X}_t|\boldsymbol{Y}_t) = -\log\left|\boldsymbol{X}_t\boldsymbol{Y}_t^{-1}\right| + \text{tr}\left(\boldsymbol{X}_t\boldsymbol{Y}_t^{-1}\right) - M \tag{16}$$

PSDTF のコスト関数 $\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y})$ は次式で与えられる．

$$\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y}) = \sum_{t=1}^{T}\mathcal{C}_\phi(\boldsymbol{X}_t|\boldsymbol{Y}_t) \tag{17}$$

LD-PSDTF に関しては，このコスト関数を最小化する $\boldsymbol{V}$ および $\boldsymbol{H}$ を求めるため，乗法更新 (Multiplicative Update, MU) 型の MM アルゴリズムが提案されている [11,12]．これは，ある確率モデルの最尤推定を行うことと等価であり，共役事前分布を導入してベイズ推定を行うこともできる．

2.4 LD-PSDTF を用いた音源分離

LD-PSDTF では，音源信号 k の時刻 t における複素スペクトル（潜在変数）を $\boldsymbol{z}_{kt} = [z_{k1t}, \cdots, z_{kFt}]^T \in \mathbb{C}^F$ とすると，$\boldsymbol{z}_{kt}$ は $\boldsymbol{Y}_{kt} \in \mathcal{S}_+^F$ を共分散行列パラメータとする多変量複素ガウス分布に従うことを仮定する．

$$\boldsymbol{z}_{kt}|\boldsymbol{Y}_{kt} \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{Y}_{kt}) \tag{18}$$

ここで，混合音の時刻 t における複素スペクトル（観測変数）を $\boldsymbol{s}_t = [s_{1t}, \cdots, s_{Ft}]^T \in \mathbb{C}^F$ とすると，$\boldsymbol{s}_t = \sum_k \boldsymbol{z}_{kt}$ かつ $\boldsymbol{Y}_t = \sum_k \boldsymbol{Y}_{kt}$ であることと，複素ガウス分布の再生性から，$\boldsymbol{s}_t$ も多変量複素ガウス分布に従う．

$$\boldsymbol{s}_t|\boldsymbol{Y}_t \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{Y}_t) \tag{19}$$

最終的に，混合音の時刻 t における局所的な共分散行列を $\boldsymbol{X}_t \stackrel{\text{def}}{=} \boldsymbol{s}_t\boldsymbol{s}_t^H$ とすると，観測される複素スペクトログラム $\boldsymbol{S}$ に対する対数尤度関数を導出できる．

$$\begin{aligned}\log p(\boldsymbol{S}|\boldsymbol{Y}) &= \sum_{t=1}^{T}\log p(\boldsymbol{s}_t|\boldsymbol{Y}_t) \\ &= \sum_{t=1}^{T} -\log|\boldsymbol{Y}_t| - \text{tr}(\boldsymbol{X}_t\boldsymbol{Y}_t^{-1}) \\ &\stackrel{c}{=} -\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})\end{aligned} \tag{20}$$

したがって，対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ の最大化は，LD ダイバージェンス $\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ の最小化と等価である．

パラメータ $\boldsymbol{Y}$（$\boldsymbol{V}$ および $\boldsymbol{H}$）が推定できれば，式 (18) および式 (19) から，観測変数 $\boldsymbol{S}$ が与えられたもとで，潜在変数 $\boldsymbol{Z}_k$ の事後分布を求めることができる．

$$p(\boldsymbol{z}_{kt}|\boldsymbol{s}_t) = \mathcal{N}_c\left(\boldsymbol{z}_{kt}\middle|\boldsymbol{Y}_{kt}\boldsymbol{Y}_t^{-1}\boldsymbol{s}_t, \boldsymbol{Y}_t - \boldsymbol{Y}_{kt}\boldsymbol{Y}_t^{-1}\boldsymbol{Y}_{kt}\right) \tag{21}$$

このウィナーフィルタを用いると，$\boldsymbol{S}$ から $\boldsymbol{Z}_k$ の位相をある程度正しく復元することができる．

3. 相関テンソル分解

本章では，相関テンソル分解 (Correlated Tensor Factorization, CTF) とよぶ新しい因子分解法について説明する．CTF は NMF や PSDTF の本質的な拡張となっており，最尤推定やベイズ推定が可能である．

3.1 定式化

CTF では，観測データとして，半正定値行列 $\boldsymbol{X} \in \mathcal{S}_+^{FT}$ が与えられるものとする．CTF の目標は，半正定値行列 $\boldsymbol{X}$ を半正定値行列の集合 $\{\boldsymbol{V}_k \in \mathcal{S}_+^F\}_{k=1}^K$ と対応する半正定値行列の集合 $\{\boldsymbol{U}_k \in \mathcal{S}_+^T\}_{k=1}^K$ とのクロネッカー積の和で近似することである．

$$\boldsymbol{X} \approx \boldsymbol{Y} \stackrel{\text{def}}{=} \sum_{k=1}^K \boldsymbol{V}_k \otimes \boldsymbol{U}_k \tag{22}$$

ここで，$\boldsymbol{Y}_k = \boldsymbol{V}_k \otimes \boldsymbol{U}_k$ と定義すると，$\boldsymbol{Y} = \sum_k \boldsymbol{Y}_k$ が成立する．これらすべての半正定値行列が対角行列のとき，式 (22) で定義される CTF は，式 (2) で定義される NMF に帰着する．また，$\{\boldsymbol{V}_k\}_{k=1}^K$ あるいは $\{\boldsymbol{U}_k\}_{k=1}^K$ のいずれかが対角行列の集合であるとき，式 (22) で定義される CTF は，式 (13) で定義される PSDTF に帰着する．CTF では，PSDTF と同様に，観測行列 $\boldsymbol{X}$ と再構成行列 $\boldsymbol{Y}$ との間の誤差 $\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y})$ を評価する尺度として，Bregman 行列ダイバージェンス [14] を利用する．

$$\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y}) = \phi(\boldsymbol{X}) - \phi(\boldsymbol{Y}) - \mathrm{tr}\big(\nabla\phi(\boldsymbol{Y})^T(\boldsymbol{X}-\boldsymbol{Y})\big) \tag{23}$$

CTF では，$\mathcal{C}_\phi(\boldsymbol{X}|\boldsymbol{Y})$ を最小化する $\boldsymbol{V}$ および $\boldsymbol{U}$ を求めることが目標となる．本稿では特に，モノラル音響信号の音源分離を行うのに適した LD ダイバージェンス $\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ に基づく CTF (LD-CTF) について議論する．

3.2 LD-CTF を用いた音源分離

LD-CTF では，音源信号 k の複素スペクトログラム $\boldsymbol{Z}_k$ の全ての時間周波数ビンを並べたベクトル（潜在変数）を $\boldsymbol{z}_k = [z_{k11}, \cdots, z_{k1T}, \cdots, z_{kF1}, \cdots, z_{kFT}]^T \in \mathbb{C}^{FT}$ とし，$\boldsymbol{z}_k$ が $\boldsymbol{Y}_k \in \mathcal{S}_+^{FT}$ を共分散行列パラメータとする多変量複素ガウス分布に従うことを仮定する．

$$\boldsymbol{z}_k|\boldsymbol{Y}_k \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{Y}_k) \tag{24}$$

ここで，混合音の複素スペクトログラム $\boldsymbol{S}$（観測変数）を同様に展開し，$\boldsymbol{s} = [s_{11}, \cdots, s_{1T}, \cdots, s_{F1}, \cdots, s_{FT}]^T \in \mathbb{C}^{FT}$ とすると，$\boldsymbol{s} = \sum_k \boldsymbol{z}_k$ かつ $\boldsymbol{Y} = \sum_k \boldsymbol{Y}_k$ であることと，複素ガウス分布の再生性から，次式が成立する．

$$\boldsymbol{s}|\boldsymbol{Y} \sim \mathcal{N}_c(\boldsymbol{0}, \boldsymbol{Y}) \tag{25}$$

したがって，$\boldsymbol{X} \stackrel{\text{def}}{=} \boldsymbol{s}\boldsymbol{s}^H$ とすると，観測される複素スペクトログラム $\boldsymbol{S}$ に対する対数尤度関数は，次式で与えられる．

$$\begin{aligned}\log p(\boldsymbol{S}|\boldsymbol{Y}) &= \log p(\boldsymbol{s}|\boldsymbol{Y}) \\ &= -\log|\boldsymbol{Y}| - \mathrm{tr}(\boldsymbol{X}\boldsymbol{Y}^{-1}) \\ &\stackrel{c}{=} -\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})\end{aligned} \tag{26}$$

したがって，対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ の最大化は，LD ダイバージェンス $\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ の最小化と等価である．

パラメータ $\boldsymbol{Y}$（$\boldsymbol{V}$ および $\boldsymbol{U}$）が推定できれば，式 (24) および式 (25) から，観測変数 $\boldsymbol{S}$ が与えられたもとで，潜在変数 $\boldsymbol{Z}_k$ の事後分布を求めることができる．

$$p(\boldsymbol{z}_k|\boldsymbol{s}) = \mathcal{N}_c\Big(\boldsymbol{z}_k\Big|\boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{s}, \boldsymbol{Y}-\boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{Y}_k\Big) \tag{27}$$

このウィナーフィルタでは，時間周波数ビン間の完全な相関が考慮されている．また，$\boldsymbol{Z}_k \stackrel{\text{def}}{=} \boldsymbol{z}_k\boldsymbol{z}_k^H$ と定義しておく．

3.3 補助関数法

LD ダイバージェンス $\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ を $\boldsymbol{V}$ および $\boldsymbol{U}$ に関して直接最小化することは困難であるため，補助関数法 [15] を用いて，$\mathcal{C}_{\text{LD}}(\boldsymbol{X}|\boldsymbol{Y})$ を間接的に最小化することを考える．いま，$\boldsymbol{\theta}$ に関して最小化したい目的関数 $\mathcal{F}(\boldsymbol{\theta})$ に対して，

$$\mathcal{F}(\boldsymbol{\theta}) \le \mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi}) \tag{28}$$

を満たす上限関数 $\mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi})$ 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り，任意の値に対して不等式が成立する必要がある．このとき，以下の反復更新則

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

$$\boldsymbol{\theta}^{\text{new}} \leftarrow \mathrm{argmin}_{\boldsymbol{\theta}} \mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi}^{\text{new}}) \tag{30}$$

を用いると，$\mathcal{F}(\boldsymbol{\theta})$ は単調非増加となる（式 (29) により $\mathcal{F}^+(\boldsymbol{\theta}, \boldsymbol{\phi})$ が最大化されて $\mathcal{F}(\boldsymbol{\theta})$ に一致する）．このアルゴリズムの収束性は保証されており，IS-NMF や LD-PSDTF のベイズ学習でも同様の手法が利用されている [11,17]. 対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ を最大化する場合には，下限関数を補助関数として，下限関数を逐次最大化すればよい．

3.4 Expectation-Maximization アルゴリズム

LD-CTF の確率モデル（3.2 節）を考えることで，補助関数法の一種である EM アルゴリズムを用いた最尤推定を行うことができる．対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ を直接最大化することは困難であるので，下限関数を設計する．

$$\begin{aligned}&\log p(\boldsymbol{S}|\boldsymbol{Y}) \\ &= \log \int q(\boldsymbol{Z}) \frac{p(\boldsymbol{S}, \boldsymbol{Z}|\boldsymbol{V}, \boldsymbol{U})}{q(\boldsymbol{Z})} d\boldsymbol{Z} \\ &\ge \int q(\boldsymbol{Z}) \log \frac{p(\boldsymbol{S}, \boldsymbol{Z}|\boldsymbol{V}, \boldsymbol{U})}{q(\boldsymbol{Z})} d\boldsymbol{Z} \\ &= \int q(\boldsymbol{Z}) \log p(\boldsymbol{S}, \boldsymbol{Z}|\boldsymbol{V}, \boldsymbol{U}) d\boldsymbol{Z} - \int q(\boldsymbol{Z}) \log q(\boldsymbol{Z}) d\boldsymbol{Z} \\ &\stackrel{\text{def}}{=} \mathcal{L}\left(q(\boldsymbol{Z}), \boldsymbol{V}, \boldsymbol{U}\right)\end{aligned} \tag{31}$$

ここで，$q(\boldsymbol{Z})$ は潜在変数 $\boldsymbol{Z}$ に関する変分事後分布である．

まず，E ステップにおいて，$\boldsymbol{V}$ および $\boldsymbol{U}$ が既知のもとで，$\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U})$ を最大化する $q(\boldsymbol{Z})$ を求める．具体的には，$q(\boldsymbol{Z})$ が真の事後分布 $p(\boldsymbol{Z}|\boldsymbol{S},\boldsymbol{V},\boldsymbol{U})$ と一致するとき（式 (27) 参照），$\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U})$ は最大化される．

$$\begin{aligned} q(\boldsymbol{Z}) &= p(\boldsymbol{Z}|\boldsymbol{S},\boldsymbol{V},\boldsymbol{U}) = \prod_{k=1}^{K} p(\boldsymbol{z}_k|\boldsymbol{S},\boldsymbol{V}_k,\boldsymbol{U}_k) \\ &= \prod_{k=1}^{K} \mathcal{N}_c\left(\boldsymbol{z}_k \middle| \boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{x}, \boldsymbol{Y}_k - \boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{Y}_k\right) \quad (32) \end{aligned}$$

この結果から潜在変数 $\boldsymbol{Z}$ に関する期待値を計算しておく．

$$\mathbb{E}[\boldsymbol{z}_k|\boldsymbol{X},\boldsymbol{V},\boldsymbol{U}] = \boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{x} \quad (33)$$

$$\begin{aligned} \mathbb{E}[\boldsymbol{Z}_k|\boldsymbol{X},\boldsymbol{V},\boldsymbol{U}] &= \boldsymbol{Y}_k - \boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{Y}_k + \boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{X}\boldsymbol{Y}^{-1}\boldsymbol{Y}_k \\ &= \boldsymbol{Y}_k\left(\boldsymbol{I} + \left(\boldsymbol{Y}^{-1}\boldsymbol{X} - \boldsymbol{I}\right)\boldsymbol{Y}^{-1}\boldsymbol{Y}_k\right) \quad (34) \end{aligned}$$

次に，M ステップにおいて，$q(\boldsymbol{Z})$ が既知のもとで，$\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U})$ を最大化する $\boldsymbol{V}$ および $\boldsymbol{U}$ を求める．まず，$\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U})$ のうち，$\boldsymbol{V}$ および $\boldsymbol{U}$ を含む項を書き下す．

$$\begin{aligned} &\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U}) \\ &\overset{c}{=} \int q(\boldsymbol{Z}) \log p(\boldsymbol{X}|\boldsymbol{Z}) p(\boldsymbol{Z}|\boldsymbol{V},\boldsymbol{U}) d\boldsymbol{Z} \\ &\overset{c}{=} \sum_{k=1}^{K} \mathbb{E}_{q(\boldsymbol{z}_k)}\left[\log \mathcal{N}_c\left(\boldsymbol{z}_k|\boldsymbol{0},\boldsymbol{V}_k\otimes\boldsymbol{U}_k\right)\right] \\ &\overset{c}{=} -T\sum_{k=1}^{K}\log|\boldsymbol{V}_k| - F\sum_{k=1}^{K}\log|\boldsymbol{U}_k| \\ &\quad - \sum_{k=1}^{K}\mathrm{tr}\left((\boldsymbol{V}_k^{-1}\otimes\boldsymbol{U}_k^{-1})\mathbb{E}_{q(\boldsymbol{z}_k)}\left[\boldsymbol{Z}_k\right]\right) \quad (35) \end{aligned}$$

$\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U})$ を $\boldsymbol{V}_k^{-1}$ で偏微分する．

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\boldsymbol{V}_k^{-1}} &= T\boldsymbol{V}_k^T - (\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T) \\ &\quad \left((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k^{-1})\odot\mathbb{E}_{q(\boldsymbol{z}_k)}[\boldsymbol{Z}_k^T]\right)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T) \quad (36) \end{aligned}$$

ここで，$\boldsymbol{1}_*$ は全要素が 1 のベクトルあるいは行列を表す（添え字はデータのサイズ）．これを 0 とおいて $\boldsymbol{V}_k$ について解き，式 (34) を代入すると $\boldsymbol{V}_k$ の更新則を得る．

$$\begin{aligned} \boldsymbol{V}_k &\leftarrow \frac{1}{T}(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\left((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k^{-T})\right. \\ &\left.\odot\left(\boldsymbol{Y}_k\left(\boldsymbol{I} + \left(\boldsymbol{Y}^{-1}\boldsymbol{X} - \boldsymbol{I}\right)\boldsymbol{Y}^{-1}\boldsymbol{Y}_k\right)\right)\right)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T) \quad (37) \end{aligned}$$

同様に，$\boldsymbol{U}_k$ の更新則も求められる．

$$\begin{aligned} \boldsymbol{U}_k &\leftarrow \frac{1}{F}(\boldsymbol{1}_F^T\otimes\boldsymbol{I}_{T,T})\left((\boldsymbol{V}_k^{-T}\otimes\boldsymbol{1}_{T,T})\right. \\ &\left.\odot\left(\boldsymbol{Y}_k\left(\boldsymbol{I} + \left(\boldsymbol{Y}^{-1}\boldsymbol{X} - \boldsymbol{I}\right)\boldsymbol{Y}^{-1}\boldsymbol{Y}_k\right)\right)\right)(\boldsymbol{1}_F\otimes\boldsymbol{I}_{T,T}) \quad (38) \end{aligned}$$

EM アルゴリズムでは，対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ が収束するまで式 (37) および式 (38) を反復する．このとき，$\boldsymbol{V}_k$ と $\boldsymbol{U}_k$ にはスケール不定性があるので，反復のたびに，$\boldsymbol{V}_k$ の対角成分（IS-NMF における基底スペクトル $\boldsymbol{w}_k$ に相当）の和が 1 となるように，両者のスケールの調整を行う．この処理で，$\mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{V},\boldsymbol{U})$ の値は影響を受けない．

3.5 Minorization-Maximization アルゴリズム

対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ に対して EM アルゴリズムとは異なる補助関数（下限関数）を設計することにより，MM アルゴリズムに基づく最尤推定を行うことができる．EM アルゴリズムは，確率モデルの最尤推定にのみ適用できるが，MM アルゴリズムは，補助関数さえ設計できれば，一般的な目的関数の最適化問題に適用できる．

まず，半正定値行列を入力にとる関数の凸性および凹性に基づく不等式を導出しておく．まず，$f(\boldsymbol{Z}) = \log|\boldsymbol{Z}|$ ($\boldsymbol{Z}\in\mathcal{S}_+^M$) 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．

$$\log|\boldsymbol{Z}| \le \log|\boldsymbol{\Omega}| + \mathrm{tr}(\boldsymbol{\Omega}^{-1}\boldsymbol{Z}) - M \quad (39)$$

ここで，$\boldsymbol{\Omega}$ は任意の半正定値行列である．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 $\boldsymbol{A}\in\mathcal{S}_+^M$ 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．

$$\mathrm{tr}\left(\left(\sum_{k=1}^{K}\boldsymbol{Z}_k\right)^{-1}\boldsymbol{A}\right) \le \sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{Z}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{A}\boldsymbol{\Phi}_k^H\right) \quad (40)$$

ここで，$\{\boldsymbol{Z}_k\in\mathcal{S}_+^M\}_{k=1}^K$ は任意の半正定値行列の集合であり，$\{\boldsymbol{\Phi}_k\in\mathbb{C}^{M\times M}\}_{k=1}^K$ は $\sum_k\boldsymbol{\Phi}_k = \boldsymbol{I}_{M,M}$ を満たす補助変数である．ここで，$\boldsymbol{I}_{M,M}$ は $M\times M$ の単位行列である．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 (39) および式 (40) を用いると，対数尤度関数 $p(\boldsymbol{S}|\boldsymbol{Y})$ に対する下限関数を導出できる．

$$\begin{aligned} &\log p(\boldsymbol{S}|\boldsymbol{Y}) \\ &\overset{c}{=} -\log|\boldsymbol{Y}| - \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Y}^{-1}\right) \\ &\overset{c}{\ge} -\log|\boldsymbol{\Omega}| - \sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{Y}_k\boldsymbol{\Omega}^{-1}\right) - \sum_{k=1}^{K}\mathrm{tr}\left(\boldsymbol{Y}_k^{-1}\boldsymbol{\Phi}_k\boldsymbol{X}\boldsymbol{\Phi}_k^H\right) \\ &\overset{\mathrm{def}}{=} \mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{U}) \quad (41) \end{aligned}$$

ここで，$\boldsymbol{\Omega}$ は半正定値行列であり，$\{\boldsymbol{\Phi}_k\}_{k=1}^K$ は $\sum_k\boldsymbol{\Phi}_k = \boldsymbol{I}_{FT,FT}$ を満たす補助変数である．等号が成立する，すなわち，$\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{U})$ を最大化するときの条件（補助パラメータの更新則）は次式で与えられる．

$$\boldsymbol{\Omega} = \boldsymbol{Y} \quad (42)$$

$$\boldsymbol{\Phi}_k = \boldsymbol{Y}_k\boldsymbol{Y}^{-1} \quad (43)$$

次に，$\boldsymbol{\Omega}$ および $\boldsymbol{\Phi}$ が既知のもとで，$\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{U})$ を最大化する $\boldsymbol{V}$ および $\boldsymbol{U}$ を求める．行列変数微分のチェイン則に注意しながら，$\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{V},\boldsymbol{U})$ を $\boldsymbol{V}_k$ で偏微分する．

$$\begin{aligned} &\frac{\partial\mathcal{L}}{\partial\mathrm{vec}^T(\boldsymbol{V}_k)} \\ &= -\mathrm{vec}^T\left((\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\left((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k)\odot\boldsymbol{\Omega}^{-T}\right)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T)\right) \\ &\quad + \mathrm{vec}^T\left((\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\left((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k^{-1})\odot\boldsymbol{\Phi}_k^C\boldsymbol{X}^T\boldsymbol{\Phi}_k^T\right)\right. \\ &\quad \left.(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T)\right)\left(\boldsymbol{V}_k^{-T}\otimes\boldsymbol{V}_k^{-1}\right) \quad (44) \end{aligned}$$

$$= -\mathrm{vec}^T\Big((\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\Big((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k)\odot\boldsymbol{\Omega}^{-T}\Big)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T)\Big) + \mathrm{vec}^T\Big(\boldsymbol{V}_k^{-T}(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\Big((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k^{-1})\odot\boldsymbol{\Phi}_k^C\boldsymbol{X}^T\boldsymbol{\Phi}_k^T\Big)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T)\boldsymbol{V}_k^{-T}\Big) \quad (45)$$

これを 0 とおいて $\boldsymbol{V}_k$ について解き，式 (42) および式 (43) を代入すると $\boldsymbol{V}_k$ の更新則を得る．

$$\boldsymbol{P}_k \stackrel{\mathrm{def}}{=} (\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\Big((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k^T)\odot\boldsymbol{Y}^{-1}\Big)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T) \quad (46)$$

$$\boldsymbol{Q}_k \stackrel{\mathrm{def}}{=} (\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T^T)\Big((\boldsymbol{1}_{F,F}\otimes\boldsymbol{U}_k^T)\odot\boldsymbol{Y}^{-1}\boldsymbol{X}\boldsymbol{Y}^{-1}\Big)(\boldsymbol{I}_{F,F}\otimes\boldsymbol{1}_T) \quad (47)$$

$$\boldsymbol{V}_k \leftarrow \boldsymbol{P}_k^{-\frac{1}{2}}\Big(\boldsymbol{P}_k^{\frac{1}{2}}\boldsymbol{V}_k\boldsymbol{Q}_k\boldsymbol{V}_k\boldsymbol{P}_k^{\frac{1}{2}}\Big)^{\frac{1}{2}}\boldsymbol{P}_k^{-\frac{1}{2}} = \boldsymbol{P}_k^{-1}\circ(\boldsymbol{V}_k\boldsymbol{Q}_k\boldsymbol{V}_k) \quad (48)$$

ここで，$\circ$ は半正定値行列同士の幾何平均を表す [21]．同様に，$\boldsymbol{U}_k$ の更新則も求められる．

$$\boldsymbol{R}_k \stackrel{\mathrm{def}}{=} (\boldsymbol{1}_F^T\otimes\boldsymbol{I}_{T,T})\Big((\boldsymbol{V}_k^T\otimes\boldsymbol{1}_{T,T})\odot\boldsymbol{Y}^{-1}\Big)(\boldsymbol{1}_F\otimes\boldsymbol{I}_{T,T}) \quad (49)$$

$$\boldsymbol{S}_k \stackrel{\mathrm{def}}{=} (\boldsymbol{1}_F^T\otimes\boldsymbol{I}_{T,T})\Big((\boldsymbol{V}_k^T\otimes\boldsymbol{1}_{T,T})\odot\boldsymbol{Y}^{-1}\boldsymbol{X}\boldsymbol{Y}^{-1}\Big)(\boldsymbol{1}_F\otimes\boldsymbol{I}_{T,T}) \quad (50)$$

$$\boldsymbol{U}_k \leftarrow \boldsymbol{R}_k^{-\frac{1}{2}}\Big(\boldsymbol{R}_k^{\frac{1}{2}}\boldsymbol{U}_k\boldsymbol{S}_k\boldsymbol{U}_k\boldsymbol{R}_k^{\frac{1}{2}}\Big)^{\frac{1}{2}}\boldsymbol{R}_k^{-\frac{1}{2}} = \boldsymbol{R}_k^{-1}\circ(\boldsymbol{U}_k\boldsymbol{S}_k\boldsymbol{U}_k) \quad (51)$$

3.6 ノンパラメトリックベイズモデル

本節では，基底数を $K\to\infty$ としたノンパラメトリックベイズモデルについて説明する．無限モデルにおいては，無限個の基底のうちで，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なスパースな学習を行う必要がある．いま，基底の重みを表す無限次元の非負値ベクトル $\boldsymbol{\theta}=[\theta_1,\cdots,\theta_\infty]^T$ を式 (25) に導入する．

$$\boldsymbol{s}|\boldsymbol{Y}\sim\mathcal{N}_c(\boldsymbol{0},\boldsymbol{Y})=\mathcal{N}_c\left(\boldsymbol{0},\sum_{k=1}^K\theta_k(\boldsymbol{V}_k\otimes\boldsymbol{U}_k)\right) \quad (52)$$

ここで，$\boldsymbol{Y}_k=\theta_k(\boldsymbol{V}_k\otimes\boldsymbol{U}_k)$ かつ $\boldsymbol{Y}_k=\sum_k\boldsymbol{Y}_k$ とした．このとき，無限次元の $\boldsymbol{\theta}$ のうちで，一部の要素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な学習を行うには，$\boldsymbol{\theta}$ 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 [17]．最も簡単なガンマ過程の近似方法は，K を十分に大きくとり，各要素 θ_k がガンマ事前分布に従うことを仮定することである．

$$\theta_k\sim\mathcal{G}(\alpha c/K,\alpha) \quad (53)$$

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．

このとき，$K\to\infty$ とすれば，ガンマ過程が得られる．

$$G\sim\mathrm{GaP}(\alpha,G_0) \quad (54)$$

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（確率測度に限らない）．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，$\mathbb{E}[G]=G_0$ 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，$G(\boldsymbol{\Theta}_k)=\theta_k$ を満たす．α は集中度と呼ばれ，α が小さくなるほど $\boldsymbol{\theta}$ はスパースになる．

また，パラメータ $\boldsymbol{U}$ および $\boldsymbol{U}$ に関しては，複素ウィシャート事前分布を用いるのが都合がよい．

$$\boldsymbol{V}_k\sim\mathcal{W}_c(\boldsymbol{R}_0^V,\gamma_0^V) \quad (55)$$

$$\boldsymbol{U}_k\sim\mathcal{W}_c(\boldsymbol{R}_0^U,\gamma_0^U) \quad (56)$$

いま，観測データ $\boldsymbol{S}$ が与えられたもとでの，パラメータの事後分布 $p(\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U}|\boldsymbol{S})=p(\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})/p(\boldsymbol{S})$ を計算したい．しかし，周辺尤度 $p(\boldsymbol{S})=\iiint p(\boldsymbol{S},\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})d\boldsymbol{\theta}d\boldsymbol{V}d\boldsymbol{U}$ の計算は解析的に行えないため，変分ベイズ法を用いて $p(\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U}|\boldsymbol{S})$ を近似的に求める．まず，因子分解が可能な関数形をもつ変分事後分布 $q(\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})$ を考える．

$$q(\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})=\prod_{k=1}^K q(\theta_k)q(\boldsymbol{V}_k)q(\boldsymbol{U}_k) \quad (57)$$

そのうえで，真の事後分布 $p(\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U}|\boldsymbol{S})$ に対する KL ダイバージェンスを最小化するような $q(\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})$ を求めたい．これは，対数周辺尤度 $\log p(\boldsymbol{S})$ の変分下限 $\mathcal{L}$ を最大化することと等価であることが知られている．

$$\begin{aligned}
&\log p(\boldsymbol{S})\\
&\geq \mathbb{E}[\log p(\boldsymbol{S}|\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})]\\
&\quad+\mathbb{E}[\log p(\boldsymbol{\theta})]+\mathbb{E}[\log p(\boldsymbol{V})]+\mathbb{E}[\log p(\boldsymbol{U})]\\
&\quad-\mathbb{E}[\log q(\boldsymbol{\theta})]-\mathbb{E}[\log q(\boldsymbol{V})]-\mathbb{E}[\log q(\boldsymbol{U})]\\
&\stackrel{\mathrm{c}}{\geq}\mathbb{E}[\log\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})]\\
&\quad+\mathbb{E}[\log p(\boldsymbol{\theta})]+\mathbb{E}[\log p(\boldsymbol{V})]+\mathbb{E}[\log p(\boldsymbol{U})]\\
&\quad-\mathbb{E}[\log q(\boldsymbol{\theta})]-\mathbb{E}[\log q(\boldsymbol{V})]-\mathbb{E}[\log q(\boldsymbol{U})]\\
&\stackrel{\mathrm{def}}{=}\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},q(\boldsymbol{\theta}),q(\boldsymbol{V}),q(\boldsymbol{U}))
\end{aligned} \quad (58)$$

ここで，第一項に関して，不等式 (41) では，$\boldsymbol{Y}_k=\boldsymbol{V}_k\otimes\boldsymbol{U}_k$ であったのを，$\boldsymbol{Y}_k=\theta_k(\boldsymbol{V}_k\otimes\boldsymbol{U}_k)$ とした不等式を用いた．このとき，変分下限 $\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},q(\boldsymbol{\theta}),q(\boldsymbol{V}),q(\boldsymbol{U}))$ を最大化するには，$\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},q(\boldsymbol{\theta}),q(\boldsymbol{V}),q(\boldsymbol{U}))$ が収束するまで各変数の最適化を反復する．まず，補助パラメータ $\boldsymbol{\Omega}$ および $\boldsymbol{\Phi}$ に関する最適化を行う．その後，各パラメータに関する変分事後分布を順番に更新すればよい．

$$q(\boldsymbol{\theta})\propto p(\boldsymbol{\theta})\exp(\mathbb{E}_{q(\boldsymbol{V},\boldsymbol{U})}[\log\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})]) \quad (59)$$

$$q(\boldsymbol{V})\propto p(\boldsymbol{V})\exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{U})}[\log\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})]) \quad (60)$$

$$q(\boldsymbol{U})\propto p(\boldsymbol{U})\exp(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{V})}[\log\mathcal{L}(\boldsymbol{\Omega},\boldsymbol{\Phi},\boldsymbol{\theta},\boldsymbol{V},\boldsymbol{U})]) \quad (61)$$

まず，$\mathcal{L}(\boldsymbol{\Omega}, \boldsymbol{\Phi}, q(\boldsymbol{\theta}), q(\boldsymbol{V}), q(\boldsymbol{U}))$ を最大化する補助パラメータ $\boldsymbol{\Omega}$ および $\boldsymbol{\Phi}$ を求める．

$$\boldsymbol{\Omega} = \textstyle\sum_k \mathbb{E}[\boldsymbol{Y}_k] \tag{62}$$

$$\boldsymbol{\Phi}_k = \left(\mathbb{E}\left[\boldsymbol{Y}_k^{-1}\right]\right)^{-1} \left(\textstyle\sum_{k'} \left(\mathbb{E}\left[\boldsymbol{Y}_{k'}^{-1}\right]\right)^{-1}\right)^{-1} \tag{63}$$

次に，$\mathcal{L}(\boldsymbol{\Omega}, \boldsymbol{\Phi}, q(\boldsymbol{\theta}), q(\boldsymbol{V}), q(\boldsymbol{U}))$ を最大化する変分事後分布 $q(\boldsymbol{\theta}), q(\boldsymbol{V}), q(\boldsymbol{U})$ を求める．各分布は一般化逆ガウス (Generalized Inverse Gaussian, GIG) 分布あるいは複素行列 GIG (Matrix GIG, MGIG) 分布で与えられる [11, 22].

$$\begin{aligned} q(\theta_k) &= \mathrm{GIG}(\theta_k | \alpha c / K, \rho_k^\theta, \tau_k^\theta) \\ q(\boldsymbol{V}_k) &= \mathrm{MGIG}(\boldsymbol{V}_k | \gamma_0^V, \boldsymbol{R}_k^V, \boldsymbol{T}_k^V) \\ q(\boldsymbol{U}_k) &= \mathrm{MGIG}(\boldsymbol{U}_k | \gamma_0^U, \boldsymbol{R}_k^U, \boldsymbol{T}_k^U) \end{aligned} \tag{64}$$

ここで，各分布のパラメータは以下で与えられる．

$$\rho_k^\theta = 2\alpha + 2\mathrm{tr}\left(\mathbb{E}[\boldsymbol{V}_k \otimes \boldsymbol{U}_k] \boldsymbol{\Omega}^{-1}\right) \tag{65}$$

$$\tau_k^\theta = 2\mathrm{tr}\left(\mathbb{E}\left[\boldsymbol{V}_k^{-1} \otimes \boldsymbol{U}_k^{-1}\right] \boldsymbol{\Phi}_k \boldsymbol{X} \boldsymbol{\Phi}_k^H\right) \tag{66}$$

$$\begin{aligned} \boldsymbol{R}_k^V &= (\boldsymbol{R}_0^V)^{-1} + \mathbb{E}[\theta_k](\boldsymbol{I}_{F,F} \otimes \boldsymbol{1}_T^T) \\ &\quad \left((\boldsymbol{1}_{F,F} \otimes \mathbb{E}[\boldsymbol{U}_k^T]) \odot \boldsymbol{\Omega}^{-1}\right) (\boldsymbol{I}_{F,F} \otimes \boldsymbol{1}_T) \end{aligned} \tag{67}$$

$$\begin{aligned} \boldsymbol{T}_k^V &= \mathbb{E}[\theta_k^{-1}](\boldsymbol{I}_{F,F} \otimes \boldsymbol{1}_T^T) \\ &\quad \left((\boldsymbol{1}_{F,F} \otimes \mathbb{E}[\boldsymbol{U}_k^{-T}]) \odot \boldsymbol{\Phi}_k \boldsymbol{X} \boldsymbol{\Phi}_k^H\right) (\boldsymbol{I}_{F,F} \otimes \boldsymbol{1}_T) \end{aligned} \tag{68}$$

$$\begin{aligned} \boldsymbol{R}_k^U &= (\boldsymbol{R}_0^U)^{-1} + \mathbb{E}[\theta_k](\boldsymbol{1}_F^T \otimes \boldsymbol{I}_{T,T}) \\ &\quad \left((\mathbb{E}[\boldsymbol{V}_k^T] \otimes \boldsymbol{1}_{T,T}) \odot \boldsymbol{\Omega}^{-1}\right) (\boldsymbol{1}_F \otimes \boldsymbol{I}_{T,T}) \end{aligned} \tag{69}$$

$$\begin{aligned} \boldsymbol{T}_k^U &= \mathbb{E}[\theta_k^{-1}](\boldsymbol{1}_F^T \otimes \boldsymbol{I}_{T,T}) \\ &\quad \left((\mathbb{E}[\boldsymbol{V}_k^{-T}] \otimes \boldsymbol{1}_{T,T}) \odot \boldsymbol{\Phi}_k \boldsymbol{X} \boldsymbol{\Phi}_k^H\right) (\boldsymbol{1}_F \otimes \boldsymbol{I}_{T,T}) \end{aligned} \tag{70}$$

MGIG 分布に従う確率変数の各種期待値を計算するには，ウィシャート分布を用いた重点サンプリングを行う必要がある [23].

3.7 計算量削減

これまで説明した通り，理論上は LD-CTF の最尤推定とベイズ推定が導出できるが，計算量が極めて膨大であり，現実的には実行不可能である．音源分離問題において，F と T はそれぞれ周波数ビン数とフレーム数に対応しており，$FT \times FT$ という巨大な行列を取り扱う必要があることから，LD-CTF の計算量は $O(KF^3T^3)$ である．

計算量を削減する一つの方法は，半正定値行列 $\boldsymbol{V}_k$ や $\boldsymbol{U}_k$ をブロック対角行列に制限することである（**図 2**）．音源分離においては，時間周波数領域を矩形領域に分割し，各ブロックでは完全な相関を考慮するが，ブロック間は独立であるとみなすことを意味する．ここで，$\boldsymbol{V}_k$ と $\boldsymbol{U}_k$ がどちらも対角行列の場合には IS-NMF，いずれかが対角行列の場合には LD-PSDTF となることに注意されたい．ブロック対角 LD-CTF は，制約のない LD-CTF と LD-PSDTF の間の中間的な表現力のモデルになっている．

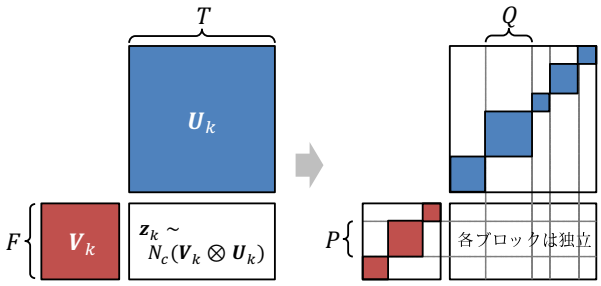


図 2 ブロック対角行列に制限した LD-CTF.

簡単のため，$\boldsymbol{V}_k \in \mathcal{S}_+^F$ は，同一サイズ P の半正定値行列を対角上に I 個に並べた半正定値行列であり，$\boldsymbol{U}_k \in \mathcal{S}_+^T$ は，同一サイズ Q の半正定値行列を対角上に J 個に並べた半正定値行列であるとする．ただし，$F = PI$ および $T = QJ$ とする．実際には，すべての k について共通であれば，各ブロックのサイズは可変でよい．また，必ずしも隣接する行あるいは列を同じブロックにいれる必要はなく，行および列をそれぞれ I 個と J 個のブロックにクラスタリングしておけばよい．このとき，計算量は $O(KIJP^3Q^3)$ となり，大幅な計算量の削減が可能となる．

優れた近似精度と劇的な高速化を同時に達成するための方法として，$\boldsymbol{V}$ および $\boldsymbol{U}$ に対してそれぞれ同時対角化を行うことが考えられる．本稿では，時間周波数領域における複素スペクトログラム $\boldsymbol{S}$ の確率モデルを考えていた．

$$\boldsymbol{S} | \boldsymbol{V}, \boldsymbol{U} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_{k=1}^K \boldsymbol{V}_k \otimes \boldsymbol{U}_k\right) \tag{71}$$

ここで，$\boldsymbol{S}$ は行列であるが，全ての要素を並べたベクトルが多変量複素ガウス分布に従うと考える．$\boldsymbol{S}$ に任意の線形変換 $\boldsymbol{A}$ および $\boldsymbol{B}$ を両側から適用すると，次式が成立する．

$$\boldsymbol{A}\boldsymbol{S}\boldsymbol{B}^H | \boldsymbol{V}, \boldsymbol{U} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_{k=1}^K \boldsymbol{A}\boldsymbol{V}_k\boldsymbol{A}^H \otimes \boldsymbol{B}\boldsymbol{U}_k\boldsymbol{B}^H\right)$$

このとき，すべての k について，$\boldsymbol{A}\boldsymbol{V}_k\boldsymbol{A}^H$ および $\boldsymbol{B}\boldsymbol{U}_k\boldsymbol{B}^H$ が対角行列となれば，変換後の空間では，LD-CTF は IS-NMF に帰着し，高速な実行が可能となる．したがって，同時対角化を行う $\boldsymbol{A}$ および $\boldsymbol{B}$ の推定と，$\boldsymbol{V}$ および $\boldsymbol{U}$ の更新を反復的に繰り返す手法が有望である．**図 3** に，$\boldsymbol{A}$ と $\boldsymbol{B}$ が，ともに離散フーリエ変換 (DFT) 行列である場合の例を示す．もし，時間領域信号が定常であれば，$\boldsymbol{V}_k$ は巡回行列となり，DFT 行列 $\boldsymbol{D}_F$ で対角されるため，$\boldsymbol{A} = \boldsymbol{D}_F$ とすることはよい選択となる．これが，周波数方向には独立であるとして IS-NMF を適用する理由であったが，LD-PSDTF を用いれば分離精度が向上することから，$\boldsymbol{D}_F$ は同時対角化のための最適な変換ではないことが示唆されている．一方，フレーム方向の時間軸の変換については，これまでほとんど議論されてこなかった．ガウス過程の観点から，周波数方向・時間方向の依存性を一挙に解決する試みについてさらなる研究が期待される．

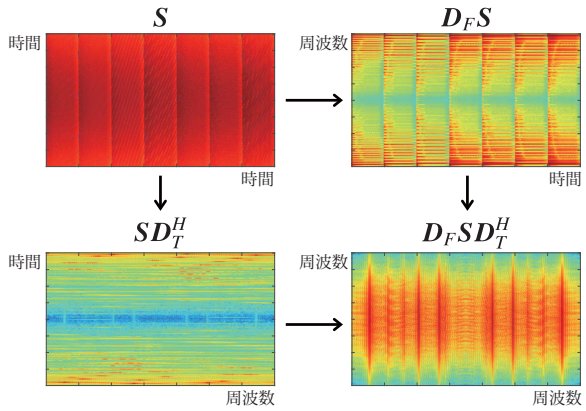


図 3 DFT による空間変換．$\boldsymbol{S} \in \mathbb{C}^{F\times T}$ は混合音の複素スペクトログラム（ここでは F は窓幅）であり，$\boldsymbol{D}_F \in \mathbb{C}^{F\times F}$ および $\boldsymbol{D}_T \in \mathbb{C}^{T\times T}$ はそれぞれ離散フーリエ変換行列である．

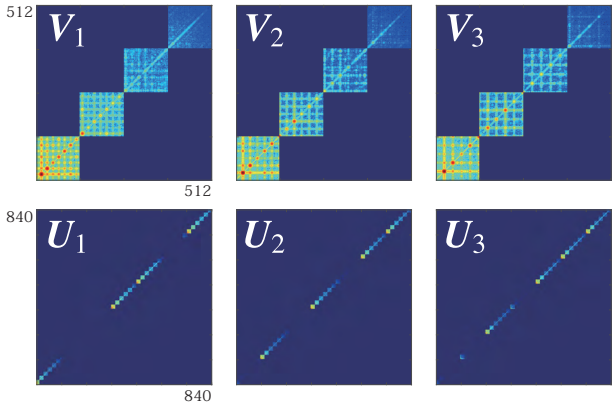


図 4 $(P, Q) = (64, 20)$ のときの LD-CTF の結果．

4. 評価

LD-CTF を用いた音源分離実験について報告する．本稿では，3.5 節で述べた MM アルゴリズムに基づく最尤推定を用いて，LD-CTF の基本的な動作と性能を確認した．

4.1 実験条件

実験には，MIDI のピアノ音を用いた．三つの音高 (C4, E4, G4) をもつ 1.2 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) を連結して 8.4 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．窓幅 512 点ガウス窓を用いて，窓シフト長 160 点の STFT を行った（$F = 256$，$T = 840$）．

合成した混合音を C4, E4, G4 に対応する音源信号に分離することを試みた（$K = 3$）．ブロック対角 LD-CTF の設定は，$(P, Q) = (256, 1), (1, 840), (128, 10), (64, 20), (32, 40)$ とした．比較のため，IS-NMF も評価した．このとき，計算量を削減し，局所解を回避するため，IS-NMF の結果を，LD-CTF の初期値とした．ブロック対角 LD-CTF は，$(P, Q) = (1, 1)$ のとき IS-NMF に，$(P, Q) = (256, 1)$ あるいは $(P, Q) = (1, 840)$ のときそれぞれ周波数方向あるいは時間方向の相関を考慮する LD-PSDTF と等価である．各手法に対して，反復回数は 100 回とした．BSS Eval Toolbox [24] 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．

4.2 実験結果

表 1 に実験結果を示す．$(P, Q) = (256, 1)$ に対応する LD-CTF （周波数方向の相関を考慮した LD-PSDTF）は IS-NMF に対する明確な優位性を示した．LD-PSDTF を時間領域で実行した場合は，SDR は 26.6 [dB] であった [12]．理論的には両者は等価であり，ほぼ同じ分離精度が得られるはずであるが，複素行列計算に関する計算誤差の影響が考えられる．一方，$(P, Q) = (1, 840)$ に対応する LD-CTF（時間方向の相関を考慮した LD-PSDTF）でも優れた分離結果が得られることが本実験で初めて確認された．図 1 を見ると，周波数方向においては，倍音の間には強い相関があり，時間方向においては，音量の大きいフレームの間には強い相関があることがわかる．

表 1 音源分離精度 [dB].

	IS-NMF	LD-PSDTF		LD-CTF		
(P, Q)	$(1, 1)$	$(256, 1)$	$(840, 1)$	$(128, 10)$	$(64, 20)$	$(32, 40)$
SDR	18.88	**21.58**	21.04	19.68	20.60	20.21
SIR	24.14	**27.01**	24.67	25.29	26.17	25.45
SAR	20.45	23.14	**23.50**	21.47	21.47	22.15

ブロック対角 LD-CTF は，IS-NMF より高精度な分離が可能であるはあるものの，LD-PSDTF には及ばなかった．**図 4** に実行結果の例を示す．時間周波数領域を排他的なブロックに区切ることで，ブロックの境界付近で位相が不適切になる可能性が考えられる．また，隣接する周波数を同じブロックにしているが，倍音周波数を同じブロックにいれて，倍音周波数間の相関は完全に取り扱うことで，分離精度は向上することが期待できる．

5. おわりに

本稿では，NMF や PSDTF を特別な場合として含む，相関テンソル分解 (CTF) と呼ぶ究極の因子分解法を提案した．特に，音源分離に適した LD ダイバージェンスに基づく CTF (LD-CTF) を取り上げ，最尤推定のための EM アルゴリズムおよび MM アルゴリズム，さらにガンマ過程に基づくノンパラメトリックベイズモデルと変分ベイズ法を提案した．また，ブロック対角制約や同時対角化などの計算量を削減する方法について議論した．

CTF は一般の N 階のテンソルデータに対して適用でき，よく知られた Canonical Polyadic (CP) 分解のエレガントな数学的拡張でもある．テンソルの各モードにおける完全な相関を取り扱うことができる強力な枠組みであり，音響信号処理分野以外への応用について検討していきたい．

謝辞: 本研究の一部は，JSPS 科研費 26700020, 16H01744 および JST ACCEL No. JPMJAC1602 の支援を受けた.

参考文献

[1] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Query-by-Example Music Information Retrieval by Score-Informed Source Separation and Remixing Technologies, *EURASIP Journal on Advances in Signal Processing*, Vol. 2010 (2010). Article ID 172961.

[2] Goto, M.: Active Music Listening Interfaces based on Signal Processing, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. 4, pp. 1441–1444 (2007).

[3] Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Drumix: An Audio Player with Real-time Drum-Part Rearrangement Functions for Active Music Listening, *IPSJ Digital Courier*, Vol. 3, pp. 134–144 (2007).

[4] Sturmel, N., Liutkus, A., Pinel, J., Girin, L., Marchand, S., Richard, G., Badeau, R. and Daudet, L.: Linear Mixing Models for Active Listening of Music Productions in Realistic Studio Conditions, *The 132nd Audio Engineering Society (AES) Convention* (2012).

[5] Lee, D. and Seung, H.: Algorithms for Non-Negative Matrix Factorization, *Neural Information Processing Systems (NIPS)*, pp. 556–562 (2000).

[6] Griffin, D. W. and Lim, J. S.: Signal Estimation from Modified Short-Time Fourier Transform, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 32, No. 2, pp. 236–243 (1984).

[7] Roux, J. L., Kameoka, H., Ono, N. and Sagayama, S.: Explicit Consistency Constraints for STFT Spectrograms and Their Application to Phase Reconstruction, *Workshop on Statistical and Perceptual Audition (SAPA)*, pp. 23–28 (2008).

[8] Roux, J. L., Kameoka, H., Ono, N. and Sagayama, S.: Fast Signal Reconstruction from Magnitude STFT Spectrogram Based on Spectrogram Consistency, *International Conference on Digital Audio Effects (DAFx)*, pp. 397–403 (2010).

[9] Kameoka, H., Nishimoto, T. and Sagayama, S.: Complex NMF: A New Sparse Representation for Acoustic Signals, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 45–48 (2009).

[10] Roux, J. L., Vincent, E., Mizuno, Y., Kameoka, H., Ono, N. and Sagayama, S.: Consistent Wiener Filtering: Generalized Time-Frequency Masking Respecting Spectrogram Consistency, *International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, pp. 89–96 (2010).

[11] Yoshii, K., Tomioka, R., Mochihashi, D. and Goto, M.: Infinite Positive Semidefinite Tensor Factorization for Source Separation of Mixture Signals, *International Conference on Machine Learning (ICML)*, pp. 576–584 (2013).

[12] Yoshii, K., Tomioka, R., Mochihashi, D. and Goto, M.: Beyond NMF: Time-Domain Audio Source Separation without Phase Reconstruction, *International Society for Music Information Retrieval Conference (ISMIR)*, pp. 369–374 (2013).

[13] Févotte, C., Bertin, N. and Durrieu, J.-L.: Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis, *Neural Computation*, Vol. 21, No. 3, pp. 793–830 (2009).

[14] Bregman, L. M.: 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*, Vol. 7, No. 3, pp. 200–217 (1967).

[15] Nakano, M., Kameoka, H., Roux, J. L., Kitano, Y., Ono, N. and Sagayama, S.: Convergence-Guaranteed Multiplicative Algorithms for Non-Negative Matrix Factorization with Beta Divergence, *International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 283–288 (2010).

[16] Cemgil, A. T.: Bayesian Inference for Nonnegative Matrix Factorisation Models, *Computational Intelligence and Neuroscience*, Vol. 2009, pp. 1–17 (2009).

[17] Hoffman, M., Blei, D. and Cook, P.: Bayesian Nonparametric Matrix Factorization for Recorded Music, *International Conference on Machine Learning (ICML)*, pp. 439–446 (2010).

[18] Allen, J. B. and Rabiner, L. R.: A Unified Approach to Short-Time Fourier Analysis and Synthesis, *IEEE*, Vol. 65, No. 11, pp. 1558–1564 (1977).

[19] Kulis, B., Sustik, M. and Dhillon, I.: Low-rank Kernel Learning with Bregman Matrix Divergences, *Journal of Machine Learning Research (JMLR)*, Vol. 10, pp. 341–376 (2009).

[20] Sawada, H., Kameoka, H., Araki, S. and Ueda, N.: Efficient Algorithms for Multichannel Extensions of Itakura-Saito Nonnegative Matrix Factorization, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 261–264 (2012).

[21] Pusz, W. and Woronowicz, S. L.: Functional Calculus for Sesquilinear Forms and the Purification Map, *Reports on Mathematical Physics*, Vol. 8, No. 2, pp. 159–170 (1975).

[22] Itakura, K., Bando, Y., Nakamura, E., Itoyama, K., Yoshii, K. and Kawahara, T.: Bayesian Multichannel Nonnegative Matrix Factorization for Audio Source Separation and Localization, *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 551–555 (2017).

[23] Yang, M., Li, Y. and Zhang, Z.: Multi-Task Learning with Gaussian Matrix Generalized Inverse Gaussian Model, *International Conference on Machine Learning (ICML)*, pp. 423–431 (2013).

[24] Vincent, E., Gribonval, R. and Févotte, C.: Performance Measurement in Blind Audio Source Separation, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 14, No. 4, pp. 1462–1469 (2006).