一般社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
EA2017-144,SIP2017-153,SP2017-127 (2018-03)

[ポスター講演] ブラインド音源分離のための高速相関テンソル分解

北村　昂一†　坂東　宜昭†　糸山　克寿†　吉井　和佳†,††　河原　達也†

† 京都大学 大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
†† 理化学研究所 革新的知能統合研究センター (AIP)　〒 103–0027 東京都中央区日本橋
E-mail: †{kitamura,yoshiaki,itoyama,yoshii,kawahara}@sap.ist.i.kyoto-u.ac.jp

あらまし　本稿では，シングルチャネル音源分離のための複素 t 分布に基づく高速相関テンソル分解 (高速 t-CTF) について述べる．相関テンソル分解 (CTF) は，非負値行列分解 (NMF) や半正定値テンソル分解 (PSDTF) の拡張となっており，複素スペクトログラムの周波数方向および時間方向の相関を考慮した音源分離が可能である．しかし，莫大な計算量のため現実的には実行が困難であり，混合音のスペクトログラムが複素ガウス分布に従うという強い仮定が置かれている問題があった．本研究ではまず，相関行列の同時対角化に基づく CTF の高速近似法を提案する．高速 CTF では，複素スペクトログラムの周波数方向および時間方向の相関を無相関化する変換行列の推定と，変換後の空間での非負値行列分解 (NMF) を同時に行う．次に，混合音のスペクトログラムが複素 t 分布に従うことを仮定した高速 t-CTF を導出する．複素 t 分布は，複素対称 α 安定分布と同様に，複素コーシー分布および複素ガウス分布を特殊形として含む裾の重い確率分布であるが，一般に再生性を持たないかわりに，すべての自由度について確率密度関数が陽にかける利点を持ち，最尤推定を行う上で都合がよい．音源分離実験から，高速 t-CTF の特殊形である高速 PSDTF は NMF よりも高い音源分離精度を持つことを示した．

キーワード　相関テンソル分解，半正定値テンソル分解，非負値行列分解，複素 t 分布，同時対角化

1. ま え が き

ブラインド音源分離は，実環境下での音声認識や音環境認識を行う上で必要不可欠な技術である．ブラインド音源分離では，単チャネルまたは多チャネルの信号を音源の位置やマイクロホンに関する事前情報を用いずに，もとの個別の信号に分解することを目的としている．

これまでの音源分離法では，複素スペクトログラムの各時間周波数ビンが独立に複素ガウス分布に従うという仮定が一般的であった．この仮定の下では，複素ガウス分布の再生性から，音源スペクトログラムが重畳してできる混合音スペクトログラムの各時間周波数ビンも複素ガウス分布に従うという望ましい性質を持つ．しかし，実際の複素スペクトログラムの各時間周波数ビンの分布は裾が重い場合が多いので，この仮定は必ずしも適切ではない．近年では，音源スペクトログラムが，裾の重い複素対称 α 安定分布に従うと仮定すると，この分布の再生性から，混合音スペクトログラムも複素対称 α 安定分布に従うことを利用した音源分離手法も提案されている．しかし，一般に，複素対称 α 安定分布の確率密度関数は，$\alpha = 1$ の複素コーシー分布および $\alpha = 2$ の複素ガウス分布以外の場合は，陽に書くことができず，効率的な最尤推定が難しかった．そこで，本研究では，複素 t 分布を用いた音源分離に取り組む．一般に複素 t 分布は再生性を持たないかわりに，すべての自由度について確率密度関数が陽にかけ，複素コーシー分布および複素ガウス分布をその特殊形として含むため，有用であると考えられる．

シングルチャネルの音響信号に対して音源分離を行う場合，NMF が広く用いられている [1–3]．板倉斎藤ダイバージェンスに基づく NMF (IS-NMF) [4] では，各音源の複素スペクトログラムの各時間周波数ビンが複素ガウス分布に従うことが仮定されていた．実際の複素スペクトログラムの各時間周波数ビンの分布は裾の重い場合が多いので，複素コーシー分布に基づく Cauchy NMF [5] が提案されている．複素ガウス分布及び複素コーシー分布を統一的に扱うために，複素 t 分布に基づく NMF が提案され，音源分離精度が向上することが確認されている [6]．このように，NMF では，複素スペクトログラムの各時間周波数ごとに独立に分布をおいていたので，音響信号に本来含まれる周波数方向の相関を考慮して音源分離を行うことができなかった．その問題を解決した半正定値テンソル分解 (Positive Semidefinite Tensor Factorization, PSDTF) [7,8] が提案され，実験では．NMF よりも高い音源分離精度を持つことが確認されている．さらに，実際の音響信号に含まれる周波数方向と時間方向の相関を扱うために，相関テンソル分解 (Correlated Tensor Factorization, CTF) [9] が提案されている．

本研究では，シングルチャネル音源分離のための高速 CTF および複素 t 分布に基づく高速相関テンソル分解 (高速 t-CTF) を提案する．高速 CTF では，周波数方向の共分散行列および時間方向の共分散行列を同時に対角化することで，計算量を削減する．対角化を行うことが可能であれば，変換後の空間では高速 CTF は IS-NMF に帰着するので，高速な実行が可能となる．

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©2018　by　IEICE

2. 関 連 研 究

この章では，モノラルの音響信号に対して音源分離を行う半正定値テンソル分解 (Positive Semidefinite Tensor Factorization, PSDTF) および相関テンソル分解 (Correlated Tensor Factorization, CTF) [9] について述べる．PSDTF は，NMF [10] の拡張となっており，NMF では考慮されていなかった，複素スペクトログラムに備わっている時間方向の相関も考慮し音源分離を行う．

2.1 半正定値テンソル分解

半正定値テンソル分解は各フレームにおける複素スペクトルの直積である半正定値行列を少数の半正定値行列の和に分解する．一方，NMF はこの半正定値行列の対角成分であるパワースペクトルを少数の基底スペクトルの和に分解している．PSDTF では，NMF で考慮されていなかった周波数ビン間の相関を考慮しながら，音源分離を行っている．このため PSDTF では，調波構造などの周波数ビン間の相関を考慮しながら音源分離を行っており，NMF よりも高い精度の音源分離を行うことが可能である．

2.1.1 定 式 化

PSDTF では入力として，半正定値行列の集合 $\boldsymbol{X} = [\boldsymbol{X}_1, \cdots, \boldsymbol{X}_T] \in \mathbb{C}^{M\times M\times T}$ を考える．PSDTF では，各フレーム t における半正定値行列 $\boldsymbol{X}_t$ を K 個の半正定値行列 $\{\boldsymbol{W}_k \in \mathcal{S}_+^F\}_{k=1}^K$ の線形和で近似する．

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

図 1 に PSDTF の概念図を示す．ここで，$h_{kt} \geqq 0$ は t 番目の要素 $\boldsymbol{X}_t$ における基底行列 $\boldsymbol{W}_k$ の重みである．観測行列である $\boldsymbol{X}_t$ と再構成行列 $\boldsymbol{Y}_t$ の誤差を評価するために，LogDet (LD) ダイバージェンス (C_{LD}) [11] が知られている．

$$C_{LD}\left(\boldsymbol{X}_t|\boldsymbol{Y}_t\right) = -\log|\boldsymbol{X}_t\boldsymbol{Y}_t^{-1}| + \mathrm{tr}\left(\boldsymbol{X}_t\boldsymbol{Y}_t^{-1}\right) - M \tag{2}$$

次式で表されるコスト関数 $C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)$ を用いた PSDTF は LD-PSDTF と呼ばれている．

$$C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right) = \sum_{t=1}^{T} C_{LD}\left(\boldsymbol{X}_t|\boldsymbol{Y}_t\right) \tag{3}$$

LD-PSDTF では，コスト関数である $C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)$ を最小化する $\boldsymbol{W}$ および $\boldsymbol{H}$ を求める．$\boldsymbol{W}$ および $\boldsymbol{H}$ を求めるために，乗法更新のアルゴリズムが提案されている [7].

2.1.2 LD-PSDTF を用いた音源分離

LD-PSDTF では，音源信号 k のフレーム t における複素スペクトル $\boldsymbol{z}_{kt} = [z_{kt1}, \cdots, z_{ktF}]^T \in \mathbb{C}^F$ が $\boldsymbol{Y}_{kt} \in \mathcal{S}_+^F$ 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とする．

$$\boldsymbol{z}_{kt}|\boldsymbol{Y}_{kt} \sim \mathcal{N}_c\left(\boldsymbol{0}, \boldsymbol{Y}_{kt}\right) \tag{4}$$

ここで，与えられた複素スペクトログラム $\boldsymbol{S} \in \mathbb{C}^{F\times T}$ のフレーム t の要素 $\boldsymbol{s}_t$ は周波数領域での瞬時混合過程を仮定すると，次

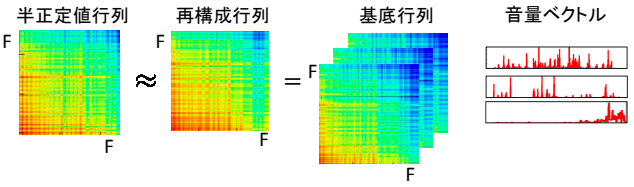

図 1 PSDTF の概念図

式が成り立つ．

$$\boldsymbol{s}_t = \sum_{k}^{K} \boldsymbol{z}_{kt} \tag{5}$$

$\boldsymbol{Y}_t = \sum_k^K \boldsymbol{Y}_{kt}$ および多変量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．

$$\boldsymbol{s}_t|\boldsymbol{Y}_t \sim \mathcal{N}_c\left(\boldsymbol{0}, \boldsymbol{Y}_t\right) \tag{6}$$

ここで，混合音のフレーム t における共分散行列を $\boldsymbol{X}_t = \boldsymbol{s}_t\boldsymbol{s}_t^H$ 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は次式で表される．

$$\begin{aligned}\log p\left(\boldsymbol{S}|\boldsymbol{Y}\right) &= \sum_{t}^{T} \log p\left(\boldsymbol{s}_t|\boldsymbol{Y}_t\right) \\ &= \sum_{t}^{T} \left(-\log|\boldsymbol{Y}_t| - \mathrm{tr}\left(\boldsymbol{X}_t\boldsymbol{Y}_t^{-1}\right)\right) \\ &\stackrel{c}{=} -\mathcal{C}_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)\end{aligned} \tag{7}$$

よって，対数尤度関数 $\log p\left(\boldsymbol{S}|\boldsymbol{Y}\right)$ の最大化は，LD ダイバージェンス $\mathcal{C}_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)$ の最小化と等価であることがわかる．

対数尤度を最大化または LD ダイバージェンスを最小化する最適なパラメータ $\boldsymbol{W}, \boldsymbol{H}$ を推定することができれば，観測行列 $\boldsymbol{S}$ が与えられたときの，潜在変数 $\boldsymbol{Z}_k$ の事後分布を求めることができる．

$$p(\boldsymbol{z}_{kt}|\boldsymbol{s}_t) = \mathcal{N}_c\left(\boldsymbol{z}_{kt}|\boldsymbol{Y}_{kt}\boldsymbol{Y}_t^{-1}\boldsymbol{s}_t, \boldsymbol{Y}_t - \boldsymbol{Y}_{kt}\boldsymbol{Y}_t^{-1}\boldsymbol{Y}_{kt}\right) \tag{8}$$

$\mathbb{E}\left[\boldsymbol{Z}_k|\boldsymbol{S}\right]$ を求めることで，音源 k の信号を復元することができる．

2.2 相関テンソル分解

本節では，モノラル音源信号に対する相関テンソル分解 (CTF) を用いた音源分離について述べる．

2.2.1 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 は与えられた半正定値行列 $\boldsymbol{X}$ を半正定値行列の集合 $\{\boldsymbol{W}_k \in \mathcal{S}_+^F\}_{k=1}^K$ と対応する半正定値行列の集合 $\{\boldsymbol{H}_k \in \mathcal{S}_+^T\}_{k=1}^K$ とのクロネッカー積で近似する．

$$\boldsymbol{X} \approx \boldsymbol{Y} \stackrel{\text{def}}{=} \sum_{k=1}^{K} \boldsymbol{W}_k \otimes \boldsymbol{H}_k \tag{9}$$

ここで，半正定値行列である $\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{H}$ がすべて対角行列であるとき，CTF は NMF に帰着する．また，$\{\boldsymbol{W}_k \in \mathcal{S}_+^F\}_{k=1}^K$ あるいは $\{\boldsymbol{H}_k \in \mathcal{S}_+^T\}_{k=1}^K$ のいずれかが対角行列の集合であるとき，

CTF は PSDTF に帰着する．観測行列 $\boldsymbol{X}$ と再構成行列 $\boldsymbol{Y}$ との間の誤差を評価する尺度として，PSDTF と同様に，LogDet (LD) ダイバージェンス (C_{LD}) [11] が知られている．

$$C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right) = -\log|\boldsymbol{X}\boldsymbol{Y}^{-1}| + \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Y}^{-1}\right) - M \quad (10)$$

LD-CTF では，C_{LD} を最小化する $\boldsymbol{W}$ と $\boldsymbol{H}$ を求める．モノラル音響信号の音源分離に適した ϕ の関数として，LD ダイバージェンスが知られており，$C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)$ に基づく CTF は LD-CTF と呼ばれている．

2.2.2 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 $\boldsymbol{Z}_k$ のすべての時間周波数ビンを並べたベクトルを $\boldsymbol{z}_k = [z_{k11},\cdots,z_{k1T},\cdots,z_{kF1},\cdots,z_{kFT}]^T \in \mathbb{C}^{FT}$ とし，$\boldsymbol{z}_k$ が $\boldsymbol{Y}_k = \boldsymbol{W}_k \otimes \boldsymbol{H}_k$ を共分散パラメータとする多変量複素ガウス分布に従うとする．

$$\boldsymbol{z}_k|\boldsymbol{Y}_k \sim \mathcal{N}_c\left(\boldsymbol{0}, \boldsymbol{Y}_k\right) \quad (11)$$

混合音の観測スペクトログラム $\boldsymbol{S}$ を，音源信号のベクトルと同様に，すべての時間周波数ビンを並べたベクトルにし，$\boldsymbol{s} = [s_{11},\cdots,s_{FT}]^T \in \mathbb{C}^{FT}$ とする．ここで，$\boldsymbol{s} = \sum_k \boldsymbol{z}_k$ と $\boldsymbol{Y} = \sum_k \boldsymbol{Y}_k$ と複素ガウス分布の再生性から，次式が成り立つ．

$$\boldsymbol{s}|\boldsymbol{Y} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{Y}\right) \quad (12)$$

$\boldsymbol{X} = \boldsymbol{s}\boldsymbol{s}^H$ とすると，観測された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．

$$\begin{aligned}\log p\left(\boldsymbol{S}|\boldsymbol{Y}\right) &= \log p\left(\boldsymbol{s}|\boldsymbol{Y}\right)\\ &= -\log|\boldsymbol{Y}| - \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Y}^{-1}\right)\\ &\overset{c}{=} -C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)\end{aligned} \quad (13)$$

よって，対数尤度関数 $\log p\left(\boldsymbol{S}|\boldsymbol{Y}\right)$ の最大化は，LD ダイバージェンス $C_{LD}\left(\boldsymbol{X}|\boldsymbol{Y}\right)$ の最小化と等価であることがわかる．

対数尤度関数 $\log p\left(\boldsymbol{S}|\boldsymbol{Y}\right)$ を最大化または LD ダイバージェンスを最小化する最適なパラメータを推定することができれば，観測行列 $\boldsymbol{S}$ が与えられたときの，潜在変数 $\boldsymbol{Z}_k$ を求めることができる．

$$p(\boldsymbol{z}_k|\boldsymbol{s}) = \mathcal{N}_c\left(\boldsymbol{z}_k|\boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{s}, \boldsymbol{Y} - \boldsymbol{Y}_k\boldsymbol{Y}^{-1}\boldsymbol{Y}_k\right) \quad (14)$$

$\mathbb{E}\left[\boldsymbol{Z}_k|\boldsymbol{S}\right]$ を求めることで，音源 k の信号を復元することができる．

3. 提 案 法

本節では，高速 CTF および高速 t-CTF について述べる．CTF で取り扱う行列は，周波数ビン数を F，フレーム数を T とすると，次元数が $FT \times FT$ の巨大な行列となり，現実的な音源分離は実行不可能である．そこで，現実的に音源分離を実行可能にするために，高速 CTF が望まれる．複素スペクトログラムに対し，周波数方向および時間方向の相関を無相関化する変換行列を求め，変換後の空間で NMF を行うことによって計算量を削減し高速化した，高速相関テンソル分解の手法について述べる．複素ガウス分布や裾の重いコーシー分布を統一的に扱うために，高速 t-CTF の定式化を行い，パラメータ最適化のための乗法更新式を導出した．

3.1 高速相関テンソル分解を用いた音源分離

高速 CTF では，CTF と同様に，音源信号 k の複素スペクトログラム $\boldsymbol{Z}_k$ のすべての時間周波数ビンを並べたベクトルを $\boldsymbol{z}_k = [z_{k11},\cdots,z_{k1T},\cdots,z_{k1F},\cdots,z_{kTF}]^T \in \mathbb{C}^{FT}$ とし，$\boldsymbol{z}_k$ が多変量複素ガウス分布に従うとする．ただし，複素ガウス分布の分散は後に線形変換を行い，対角化を行うため，複素スペクトログラムの周波数方向の線形変換を行う行列 $\boldsymbol{P} = [\boldsymbol{p}_1,\cdots,\boldsymbol{p}_f,\cdots,\boldsymbol{p}_F]^H \in \mathbb{C}^{F\times F}$ および複素スペクトログラムの時間方向の線形変換を行う行列 $\boldsymbol{Q} = [\boldsymbol{q}_1,\cdots,\boldsymbol{q}_t,\cdots,\boldsymbol{q}_T]^H \in \mathbb{C}^{T\times T}$，基底ベクトル $\boldsymbol{w}_k \in \mathbb{R}_+^F$，アクティベーションベクトル $\boldsymbol{h}_k \in \mathbb{R}_+^T$ を用いて，それらのクロネッカー積で表す．

$$\boldsymbol{z}_k \sim \mathcal{N}_c\left(\boldsymbol{0}, \boldsymbol{P}^{-1}\mathrm{Diag}\left(\boldsymbol{w}_k\right)\boldsymbol{P}^{-H} \otimes \boldsymbol{Q}^{-1}\mathrm{Diag}\left(\boldsymbol{h}_k\right)\boldsymbol{Q}^{-H}\right) \quad (15)$$

ただし，Diag($\boldsymbol{a}$) は，$\boldsymbol{a}$ を対角成分に持つ，対角行列とする．混合音の観測スペクトログラム $\boldsymbol{S}$ を，音源信号のベクトルと同様に，すべての時間周波数ビンを並べたベクトルにし，$\boldsymbol{s} = [s_{11},\cdots,s_{FT}]^T \in \mathbb{C}^{FT}$ とする．ここで，混合音の観測スペクトログラムは音源信号の複素スペクトログラムが重畳したものなので $\boldsymbol{s} = \sum_k \boldsymbol{z}_k$ が成り立ち，また，複素ガウス分布の再生性から，次式が成り立つ．

$$\boldsymbol{s} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_{k=1}^{K} \boldsymbol{P}^{-1}\mathrm{Diag}\left(\boldsymbol{w}_k\right)\boldsymbol{P}^{-H} \otimes \boldsymbol{Q}^{-1}\mathrm{Diag}\left(\boldsymbol{h}_k\right)\boldsymbol{Q}^{-H}\right) \quad (16)$$

ここで，$\boldsymbol{s}$ に対して左から $(\boldsymbol{P}\otimes\boldsymbol{Q})$ をかけると次式が成り立ち，分散を対角行列のクロネッカー積で表すことができる．

$$\left(\boldsymbol{P}\otimes\boldsymbol{Q}\right)\boldsymbol{s} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_{k=1}^{K}\mathrm{Diag}\left(\boldsymbol{w}_k\right)\otimes\mathrm{Diag}\left(\boldsymbol{h}_k\right)\right) \quad (17)$$

$\boldsymbol{X} = \boldsymbol{s}\boldsymbol{s}^H$ とすると，観測された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．

$$\begin{aligned}&\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)\\ &\quad\overset{c}{=} -\log|\boldsymbol{Y}| - \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Y}^{-1}\right)\\ &\quad= 2T\log|\det\boldsymbol{P}| + 2F\log|\det\boldsymbol{Q}| - \sum_{t=1}^{T}\sum_{f=1}^{F}\log y_{ft}\\ &\qquad- \sum_{t=1}^{T}\sum_{f=1}^{F}\left(\boldsymbol{p}_f^H\otimes\boldsymbol{q}_t^H\right)\boldsymbol{X}\left(\boldsymbol{p}_f\otimes\boldsymbol{q}_t\right)y_{ft}^{-1}\end{aligned} \quad (18)$$

ただし，$y_{ft} = \sum_{k=1}^{K} h_{kt}w_{kf}$ とする．

対数尤度関数 $\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)$ を最大化するパラメータを推定するために，補助関数法を用い最尤推定を行う．まず，対数尤度を最大化する $\boldsymbol{W}$,$\boldsymbol{H}$,$\boldsymbol{P}$ および $\boldsymbol{Q}$ を求めるために，対数尤度の下限を最大化する補助関数を導出する．

Jensen の不等式を用いて，対数尤度関数 $\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)$ に対する下限関数を導出する．

$$
\begin{aligned}
&\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)\\
&\quad\overset{c}{=}-\log|\boldsymbol{Y}|-\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Y}^{-1}\right)\\
&\quad\geqq 2T\log|\mathrm{det}\boldsymbol{P}|+2F\log|\mathrm{det}\boldsymbol{Q}|-\sum_{t=1}^{T}\sum_{f=1}^{F}\log\alpha_{ft}\\
&\qquad-\sum_{t=1}^{T}\sum_{f=1}^{F}\sum_{k=1}^{K}\frac{h_{kt}w_{kf}}{\alpha_{ft}}-\sum_{t=1}^{T}\sum_{f=1}^{F}\sum_{k=1}^{K}\frac{x_{ft}\lambda_{kft}^{2}}{h_{kt}w_{kf}}\\
&\quad\overset{\mathrm{def}}{=}\mathcal{J}\left(\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q},\alpha_{ft},\lambda_{kft}\right)
\end{aligned}
\tag{19}
$$

ただし，$x_{ft}=\left(\boldsymbol{p}_f^H\otimes\boldsymbol{q}_t^H\right)\boldsymbol{X}\left(\boldsymbol{p}_f\otimes\boldsymbol{q}_t\right)$ とする．ここで，下限関数 $\mathcal{J}\left(\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q},\alpha_{ft},\lambda_{kft}\right)$ を最大化するときの条件，つまり，等号が成立する条件は次式で与えられる．

$$\alpha_{ft}=y_{ft} \tag{20}$$

$$\lambda_{kft}=\frac{h_{kt}w_{kf}}{\sum_{k=1}^{K}h_{kt}w_{kf}} \tag{21}$$

次に，下限関数 $\mathcal{J}\left(\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q},\alpha_{ft},\lambda_{kft}\right)$ 最大化するパラメータ $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{P}$, $\boldsymbol{Q}$ を求める乗法更新アルゴリズムを導出する．

まず，$\boldsymbol{W}$ および $\boldsymbol{H}$ に関する乗法更新式を求める．式 (19) を w_{kf} および h_{kt} についてそれぞれ微分する．得られた式を 0 とおき，w_{kf} および h_{kt} について解き，式 (21) を代入すると以下の乗法更新式を得られる．

$$w_{kf}\leftarrow w_{kf}\sqrt{\frac{\sum_t h_{kt}x_{ft}y_{ft}^{-2}}{\sum_t h_{kt}y_{ft}^{-1}}} \tag{22}$$

$$h_{kt}\leftarrow h_{kt}\sqrt{\frac{\sum_f w_{kf}x_{ft}y_{ft}^{-2}}{\sum_f w_{kf}y_{ft}^{-1}}} \tag{23}$$

次に，$\boldsymbol{P}$ および $\boldsymbol{Q}$ に関する乗法更新式を求める．式 (19) を $\boldsymbol{p}_f^H$ および $\boldsymbol{q}_t^H$ について微分する．それらの式に対して，auxIVA [12] と同様の更新式を適用すると次式の乗法更新式を得られる．

$$\boldsymbol{p}_f\leftarrow\left(\boldsymbol{P}\boldsymbol{U}_f\right)^{-1}\boldsymbol{e}_f \tag{24}$$

$$\boldsymbol{q}_t\leftarrow\left(\boldsymbol{Q}\boldsymbol{R}_t\right)^{-1}\boldsymbol{e}_t \tag{25}$$

ただし，$\boldsymbol{U}_f=\frac{1}{T}\sum_{t=1}^{T}\left(\boldsymbol{E}_{F,F}\otimes\boldsymbol{q}_t^H\right)\boldsymbol{X}\left(\boldsymbol{E}_{F,F}\otimes\boldsymbol{q}_t\right)y_{ft}^{-1}$ と $\boldsymbol{R}_t=\frac{1}{T}\sum_{f=1}^{F}\left(\boldsymbol{p}_f^H\otimes\boldsymbol{E}_{T,T}\right)\boldsymbol{X}\left(\boldsymbol{p}_f\otimes\boldsymbol{E}_{T,T}\right)y_{ft}^{-1}$ とした．

$\boldsymbol{p}_f$ と $\boldsymbol{q}_t$ に対しては，乗法更新式を適用するごとに，次式で正規化を行う．

$$\boldsymbol{p}_f\leftarrow\frac{\boldsymbol{p}_f}{\sqrt{\boldsymbol{p}_f^H\boldsymbol{U}_f\boldsymbol{p}_f}} \tag{26}$$

$$\boldsymbol{q}_t\leftarrow\frac{\boldsymbol{q}_t}{\sqrt{\boldsymbol{q}_t^H\boldsymbol{R}_t\boldsymbol{q}_t}} \tag{27}$$

3.2 複素 t 分布に基づく高速相関テンソル分解

この節では，高速 t-CTF について述べる．高速 t-CTF では，混合音の観測スペクトログラムが多変量複素 t 分布に従うと仮定しているので，複素ガウス分布や複素コーシー分布を含んだ統一的な CTF の音源分離を提案することができる．混合音の観測スペクトログラム $\boldsymbol{S}$ をすべての時間周波数ビンを並べたベクトルにし，$\boldsymbol{s}=\left[s_{11},\cdots,s_{FT}\right]^T\in\mathbb{C}^{FT}$ とする．観測ベクトル $\boldsymbol{s}$ が多変量複素 t 分布に従うと仮定する．

$$\boldsymbol{s}\sim\mathcal{T}_\nu\left(\boldsymbol{0},\sum_{k=1}^{K}\boldsymbol{P}^{-1}\mathrm{Diag}\left(\boldsymbol{w}_k\right)\boldsymbol{P}^{-H}\otimes\boldsymbol{Q}^{-1}\mathrm{Diag}\left(\boldsymbol{h}_k\right)\boldsymbol{Q}^{-H}\right) \tag{28}$$

ただし，自由度 ν の多変量複素 t 分布は次式で表される．自由度 $\nu=0.5,1,2,\infty$ としたときの t 分布を図 2 に示す．

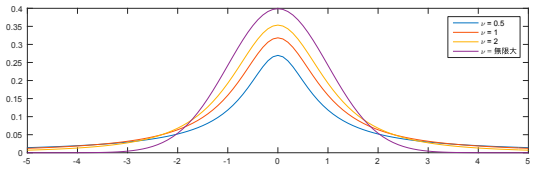

図 2　t　分　布

$$\mathcal{T}_\nu(\boldsymbol{x}|\boldsymbol{0},\boldsymbol{\Sigma})=\frac{\Gamma(\frac{2d+\nu}{2})}{\Gamma(\frac{\nu}{2})}\frac{2^d}{(\upsilon\pi)^d}\frac{1}{|\boldsymbol{\Sigma}|}\left(1+\frac{2}{\nu}\boldsymbol{x}^H\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right)^{-\frac{2M+\nu}{2}} \tag{29}$$

ただし，M は $\boldsymbol{x}$ の次元数を表す．$\boldsymbol{X}=\boldsymbol{s}\boldsymbol{s}^H$ とすると観測された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．

$$
\begin{aligned}
&\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)\\
&\quad\overset{c}{=}-\log|\boldsymbol{Y}|-\frac{2FT+\nu}{2}\log\left(1+\frac{2}{\nu}\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{Y}^{-1}\right)\right)\\
&\quad=2T\log|\mathrm{det}\boldsymbol{P}|+2F\log|\mathrm{det}\boldsymbol{Q}|-\sum_{t=1}^{T}\sum_{f=1}^{F}\log y_{ft}\\
&\qquad-\frac{2FT+\nu}{2}\log\left\{1+\frac{2}{\nu}\sum_{f,t}\left(\boldsymbol{p}_f^H\otimes\boldsymbol{q}_t^H\right)\boldsymbol{X}\left(\boldsymbol{p}_f\otimes\boldsymbol{q}_t\right)y_{ft}^{-1}\right\}
\end{aligned}
\tag{30}
$$

対数尤度関数 $\log(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q})$ に対して，補助関数法を用いることにより，最尤推定を行うことができる．対数尤度を最大化する $\boldsymbol{W},\boldsymbol{H},\boldsymbol{P}$ および $\boldsymbol{Q}$ を 求めるために，対数尤度の下限を最大化する補助関数を導出する．

Jensen の不等式を用いて，対数尤度関数 $\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)$ に対する下限関数を導出する．

$$
\begin{aligned}
&\log\left(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}\right)\\
&\quad\overset{c}{\geqq}2T\log|\mathrm{det}\boldsymbol{P}|+2F\log|\mathrm{det}\boldsymbol{Q}|-\sum_{f,t}\log\alpha_{ft}-\sum_{f,t,k}\frac{h_{kt}w_{kf}}{\alpha_{ft}}\\
&\qquad-\left(\frac{2FT+\nu}{2}\right)\left\{\phi+\phi^{-1}\left(1+\frac{2}{\nu}\sum_{f,t,k}\frac{x_{ft}\lambda_{kft}^{2}}{h_{kt}w_{kf}}\right)\right\}\\
&\quad\overset{\mathrm{def}}{=}\mathcal{M}\left(\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q},\alpha_{ft},\phi,\lambda_{kft},\right)
\end{aligned}
\tag{31}
$$

ここで，下限関数 $\mathcal{M}$ を最大化するときの条件，つまり，等号が成立する条件は次式で与えられる．

$$\alpha_{ft}=y_{ft} \tag{32}$$

$$\phi=\frac{\nu+2\sum_{t,f}\frac{x_{ft}}{y_{ft}}}{\nu} \tag{33}$$

$$\lambda_{kft}=\frac{h_{kt}w_{kf}}{y_{ft}} \tag{34}$$

次に，下限関数 $\mathcal{M}$ を最大化するパラメータ $\boldsymbol{W},\boldsymbol{H},\boldsymbol{P},\boldsymbol{Q}$ を

求める乗法更新式を導出する．

まず，$\boldsymbol{W}$ および $\boldsymbol{H}$ に関する乗法更新式を求める．式 (31) を w_{kf} および h_{kt} について微分する．得られた式を 0 とおき，w_{kf} および h_{kt} について解き，式 (34) を代入すると以下の乗法更新式を得られる．

$$w_{kf} \leftarrow w_{kf} \sqrt{\frac{\frac{2FT+\nu}{\nu+2\sum_{t,f}\frac{x_{ft}}{y_{ft}}} \sum_t h_{kt} x_{ft} y_{ft}^{-2}}{\sum_t h_{kt} y_{ft}^{-1}}} \tag{35}$$

$$h_{kt} \leftarrow h_{kt} \sqrt{\frac{\frac{2FT+\nu}{\nu+2\sum_{t,f}\frac{x_{ft}}{y_{ft}}} \sum_f w_{kf} x_{ft} y_{ft}^{-2}}{\sum_f w_{kf} y_{ft}^{-1}}} \tag{36}$$

次に，$\boldsymbol{P}$ および $\boldsymbol{Q}$ に関する更新則を求める．式 (31) を $\boldsymbol{p}_f^H$ および $\boldsymbol{q}_t^H$ について微分する．それらの式に対して，auxIVA [12] と同様の更新式を適用すると次式の乗法更新式を得られる．

$$\boldsymbol{p}_f \leftarrow (\boldsymbol{P}\boldsymbol{U}_f)^{-1}\boldsymbol{e}_f \tag{37}$$

$$\boldsymbol{q}_t \leftarrow (\boldsymbol{Q}\boldsymbol{R}_t)^{-1}\boldsymbol{e}_t \tag{38}$$

ただし，$\boldsymbol{U}_f = \frac{1}{T}\left(\frac{2FT+\nu}{\nu+2\sum_{t,f}\frac{x_{ft}}{y_{ft}}}\right)\sum_t \left(\boldsymbol{E}_{F,F}\otimes \boldsymbol{q}_t^{\mathrm{H}}\right)\boldsymbol{X}\left(\boldsymbol{E}_{F,F}\otimes \boldsymbol{q}_t^{\mathrm{H}}\right) y_{ft}^{-1}$ と $\boldsymbol{R}_t = \frac{1}{T}\left(\frac{2FT+\nu}{\nu+2\sum_{t,f}\frac{x_{ft}}{y_{ft}}}\right)\sum_{f=1}^{F}\left(\boldsymbol{p}_f^H\otimes \boldsymbol{E}_{T,T}\right)\boldsymbol{X}\left(\boldsymbol{p}_f\otimes \boldsymbol{E}_{T,T}\right) y_{ft}^{-1}$ とした．

$\boldsymbol{p}_f$ と $\boldsymbol{q}_t$ に対しては，更新するたびに，次式で正規化を行う．

$$\boldsymbol{p}_f \leftarrow \frac{\boldsymbol{p}_f}{\sqrt{\boldsymbol{p}_f^H \boldsymbol{U}_f \boldsymbol{p}_f}} \tag{39}$$

$$\boldsymbol{q}_t \leftarrow \frac{\boldsymbol{q}_t}{\sqrt{\boldsymbol{q}_t^H \boldsymbol{R}_t \boldsymbol{q}_t}} \tag{40}$$

ここまで，高速 t-CTF を用いた音源分離の定式化とパラメータの乗法更新式を述べたが，高速 t-CTF は高速 t-PSDTF を特殊形として含む音源分離手法である．基底行列の集合 $\{\boldsymbol{W}_k \in \mathcal{S}_+^F\}_{k=1}^K$ またはアクティベーション行列の集合 $\{\boldsymbol{H}_k \in \mathcal{S}_+^T\}_{k=1}^K$ が対角行列となるとき，高速 t-CTF は高速 t-PSDTF に帰着する．よって，乗法更新式を適用する際に，基底ベクトル $\boldsymbol{w}_k$ と周波数方向の線形変換を行う $\boldsymbol{P}$ のみを更新し，アクティベーションベクトル $\boldsymbol{h}$ と時間方向の線形変換を行う $\boldsymbol{Q}$ の更新を行わなければ，高速 t-CTF は高速 t-PSDTF に帰着する．ただし，時間方向の線形変換を行う $\boldsymbol{Q}$ の初期値は単位行列とする．また，高速 t-PSDTF は複素 t 分布の自由度 $\nu \to \infty$ のとき，高速 PSDTF に帰着する．

4. 評価実験

この章では，高速 PSDTF を用いた音源分離実験について述べる．音源分離精度では，IS-NMF および LD-PSDTF の分離精度との比較を行った．

4.1 実験条件

実験で用いる入力の混合音は，MIDI のピアノ音を用いて作成した．混合音は，三つの音高 (C4, E4, G4) をもつ 1.2 秒間の音響信号を，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) をつなぎ合わせた 8.4 秒の音響信号を用いた．また，音源数 2 での音源分離精度を求めるため，二つの音高 (C4, E4) をもつ 1.2 秒間の音響信号を異なる組み合わせで重畳したもの (C4, E4, C4 + E4) をつなぎ合わせた 3.6 秒の音響信号を混合音として用意した．(C4, G4) および (E4, G4) の音高に対しても，(C4, E4) の音高で作成した音響信号と同様に，1.2 秒の音響信号を異なる組み合わせで重畳し，つなぎ合わせた 3.6 秒の音響信号を入力の混合音として用意した．混合音のサンプリング周波数は 16[kHz] で，窓幅 512 点のガウス窓を用いて，窓シフト長 160 点の STFT を行った．

表 1　音源数 3 の楽曲の音源分離精度 [dB]

	IS-NMF	LD-PSDTF	高速 LD-PSDTF
SDR	18.9	23.0	19.1
SIR	24.2	27.7	23.9
SAR	20.5	25.1	20.8

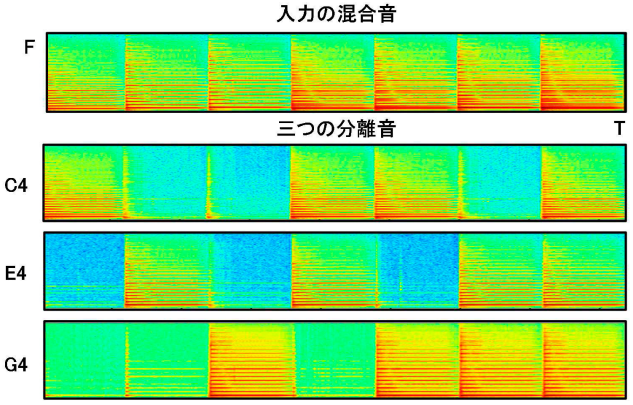

図 3　入力の混合音と分離音のスペクトログラム

実験では，入力の混合音を重畳する前のもとの音響信号 (C4, E4, G4) に分離することを試みた．提案法である高速 PSDTF の音源分離精度を比較するために，IS-NMF および LD-PSDTF の音源分離の精度も評価した．LD-PSDTF および高速 PSDTF の計算量削減とパラメータが局所解に陥ることを回避するために，IS-NMF の実験後のパラメータの値を，LD-PSDTF および高速 PSDTF の初期値とした．高速 PSDTF のパラメータである変換行列 $\boldsymbol{P}$ および $\boldsymbol{Q}$ の初期値は，単位行列とした．各手法の反復回数は，IS-NMF が 300 回，LD-PSDTF が 100 回，高速 PSDTF が 14 回とした．音源分離精度は,BSS Eval Toolbox [13] を用いて, Source-to-Distortion Ratio (SDR), Source-to-Interferences Ratio (SIR), および Source-to-Artifacts Ratio (SAR) で評価をした．

4.2 音源数 3 での実験結果

表 1 に実験結果を示す．提案法である高速 PSDTF は IS-NMF よりも僅かに高く，音源分離精度が少し向上していることがわかった．僅かにしか音源分離精度が向上していないが，これは，高速 PSDTF の反復回数が 14 回と少なく，パラメータの値が収束していないことが原因と考えられる．また，反復回数が少ない理由としては，反復回数が 15 回を超えると，変換行列 $\boldsymbol{P}$ の更新の際に，$\boldsymbol{P}$ が特異値行列となり，逆行列を計算するときに計算誤差の影響がでてくるためである．図 3 に音

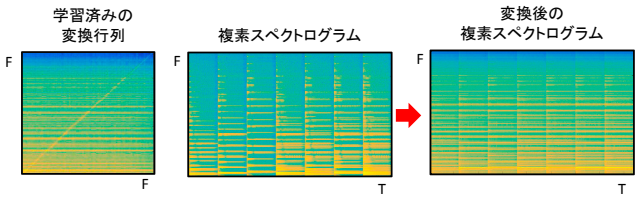

図 4　変換行列を適用したあとのスペクトログラム

表 2　C4 と G4 の足し合わせ音源分離精度 [dB]

	IS-NMF	LD-PSDTF	高速 LD-PSDTF
SDR	19.57	20.99	20.20
SIR	24.33	26.33	25.21
SAR	21.37	22.59	21.87

表 3　C4 と E4 の足し合わせ音源分離精度 [dB]

	IS-NMF	LD-PSDTF	高速 LD-PSDTF
SDR	22.31	25.49	24.83
SIR	27.44	30.54	28.82
SAR	23.94	27.13	27.10

表 4　E4 と G4 の足し合わせ音源分離精度 [dB]

	IS-NMF	LD-PSDTF	高速 LD-PSDTF
SDR	25.27	27.37	27.10
SIR	30.70	33.22	32.65
SAR	26.75	28.68	28.53

源数 3 の混合音とそれを高速 PSDTF で分離した三つの分離音を示す．また，図 4 に入力の複素スペクトログラムに対して変換行列 $\boldsymbol{P}$ をかけたときの複素スペクトログラムを示す．

4.3　音源数 2 での実験結果

表 2-表 4 に実験結果を示す．提案法である高速 PSDTF は IS-NMF に対して，すべての音源数 2 の混合音で音源分離の精度が高いことを示した．音源数 3 の混合音の分離精度と比較して，音源数 2 の混合音の分離では，高速 PSDTF の音源分離精度は LD-PSDTF とほとんど変わらない結果となった．音源数 3 の実験の時と同様に，高速 PSDTF の反復回数は他の二つの手法よりも少なく，14 回であるが，これも変換行列 $\boldsymbol{P}$ の更新の際に，反復回数が 15 回を超えると $\boldsymbol{P}$ が特異値行列となり，逆行列を計算するときに計算誤差の影響が出てくるためである．

5.　おわりに

本稿では，モノラル音響信号に対して CTF の高速化手法 (高速 CTF) および複素 t 分布に基づく高速 CTF (高速 t-CTF) について述べ，高速 CTF の特殊形である高速 LD-PSDTF の音源分離の精度を IS-NMF および LD-PSDTF と比較した．

高速 PSDTF の音源分離の実験では，音源数 3 と音源数 2 の混合音に対して音源分離を行うと，IS-NMF よりも音源分離精度が向上することを確認した．今後の課題として，高速 t-CTF の実装可能なパラメータの更新式を求めることがある．

文　献

[1] Paris Smaragdis, Cedric Fevotte, Gautham J Mysore, Nasser Mohammadiha, and Matthew Hoffman. Static and dynamic source separation using nonnegative factorizations: A unified view. *Signal Processing Magazine*, Vol. 31, No. 3, pp. 66–75, 2014.

[2] Ali Taylan Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational intelligence and neuroscience*, Vol. 2009, , 2009.

[3] Alexey Ozerov, Antoine Liutkus, Roland Badeau, and Gaël Richard. Coding-based informed source separation: Nonnegative tensor factorization approach. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 8, pp. 1699–1712, 2013.

[4] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, pp. 793–830.

[5] Antoine Liutkus, Derry Fitzgerald, and Roland Badeau. Cauchy nonnegative matrix factorization. In *(WASPAA), 2015 IEEE Workshop on*, pp. 1–5.

[6] Kazuyoshi Yoshii, Katsutoshi Itoyama, and Masataka Goto. Student's t nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation. pp. 51–55, 2016.

[7] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto. Infinite positive semidefinite tensor factorization for source separation of mixture signals. *International Conference on Machine Learning (ICML)*, pp. 576–584, 2013.

[8] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto. Beyond nmf: Time-domain audio source separation without phase reconstruction. *International Society for Music Information Retrieval Conference (ISMIR)*, pp. 369–374, 2013.

[9] 吉井和佳, 富岡亮太, 持橋大地, 後藤真孝. 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. 情報処理学会研究報告.[音楽情報科学], pp. 1–10.

[10] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, pp. 556–562, 2001.

[11] Brian Kulis, Mátyás A Sustik, and Inderjit S Dhillon. Low-rank kernel learning with bregman matrix divergences. *Journal of Machine Learning Research*, Vol. 10, No. Feb, pp. 341–376, 2009.

[12] Nobutaka Ono. Stable and fast update rules for independent vector analysis based on auxiliary function technique. *Applications of Signal Processing to Audio and Acoustics*, pp. 189–192, 2011.

[13] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte. Performance measurement in blind audio source separation. *IEEE Transactions on, Audio, Speech, and Language Processing*, Vol. 14, No. 4, pp. 1462–1469, 2006.