

[ポスター講演] 音楽音響信号解析のための 学生 t 分布に基づく非負値行列分解と半正定値テンソル分解

吉井 和佳[†] 糸山 克寿[†] 後藤 真孝^{††}

[†] 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
^{††} 産業技術総合研究所 情報技術研究部門 〒305-8568 茨城県つくば市梅園 1-1-1
E-mail: [†]{yoshii,itoiyama}@kuis.kyoto-u.ac.jp, ^{††}m.goto@aist.go.jp

あらまし 本稿では、裾が重い音響信号の音源分離に適した非負値行列分解 (NMF) と半正定値テンソル分解 (PSDTF) の新しい確率モデルについて述べる。従来、板倉・斎藤ダイバージェンスに基づく NMF (IS-NMF) を用いると、音源信号の複素スペクトル中の各要素が平均 0 の複素ガウス分布に従うという仮定のもとで、混合音のパワースペクトルを少数の音源信号のパワースペクトルの錐結合に分解できることが知られている。しかし、実際には、音源信号は裾が重い確率分布に従う場合が多いため、複素ガウス分布の代わりに複素コーシー分布に基づく Cauchy NMF が提案されている。本研究では、IS-NMF や Cauchy NMF を特殊形に含む、複素 t 分布に基づく NMF (t -NMF) を提案し、統一的な学習アルゴリズムを導出する。 t -NMF は対象となるデータの裾の重さに合わせて複素 t 分布の自由度を調節することで、IS-NMF と Cauchy NMF の中間的な性質をもつ NMF を実現できる。さらに、NMF の拡張である PSDTF に対しても同様の拡張を行い、多変量複素 t 分布に基づく PSDTF (t -PSDTF) を提案する。実験の結果、 t -NMF は IS-NMF より初期値依存性が低く、自由度を小さく抑えることで、安定して高精度な音源分離を行えること、 t -PSDTF は t -NMF より精度の高い分離が行えることを確認した。

キーワード 音楽音響信号解析, 非負値行列分解, 半正定値テンソル分解, 音源分離, t 分布

[Poster Presentation] Nonnegative Matrix Factorization and Positive Semidefinite Tensor Factorization based on Student's t Distributions for Music Signal Analysis

Kazuyoshi YOSHII[†], Katsutoshi ITOYAMA[†], and Masataka GOTO^{††}

[†] Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
^{††} National Institute of Advanced Industrial Science and Technology (AIST)
E-mail: [†]{yoshii,itoiyama}@kuis.kyoto-u.ac.jp, ^{††}m.goto@aist.go.jp

Abstract This paper presents new probabilistic models of nonnegative matrix factorization (NMF) and positive semidefinite tensor factorization (PSDTF) for source separation of heavy-tailed audio signals. It is well known that NMF based on the Itakura-Saito divergence (IS-NMF) is capable of decomposing the power spectrum of a mixture signal into the weighted sum of source spectra. Real audio signals, however, tend to follow heavy-tailed probability distributions such as the Cauchy distribution. In this paper we propose NMF based on the complex t distribution (t -NMF) that includes IS-NMF and Cauchy NMF. In addition, we propose PSDTF based on the multivariate complex t distribution (t -PSDTF). The experimental results showed that t -NMF was less sensitive to initialization than IS-NMF and t -PSDTF significantly outperformed t -NMF in terms of source separation performance.

Key words Music signal analysis, NMF, PSDTF, source separation, t distribution

1. はじめに

モノラル音響信号の音源分離においては、非負値行列分解と

ウィナーフィルタを用いることが標準的になりつつある [1–9]。NMF では、入力となる非負値行列 (非負値ベクトルの集合) を、二つの非負値行列 (基底ベクトルの集合と音量ベクトルの

集合)の積で近似する,すなわち,各非負値ベクトルを少数の非負値基底ベクトルの重み付き和で近似する.NMFを音楽音響信号の非負値スペクトログラム(振幅あるいはパワースペクトログラム)に適用する場合には,得られる基底ベクトルは異なる音高をもつ音源信号のスペクトルに対応することが期待される.最後に,NMFの結果を用いて,時間・周波数ピンごとに独立にウィナーフィルタリングを行うことにより,混合音スペクトルを音源スペクトルの和に分解できる.しかし,音源スペクトルの位相情報は混合音スペクトルのものをそのまま再利用することになり,逆フーリエ変換で得られる時間領域の音源信号の品質に限界があることが指摘されていた[10].

近年,この問題を回避するため,半正定値テンソル分解(PSDTF)が提案されている[11,12].PSDTFでは,入力として半正定値行列の集合が与えられると,各半正定値行列を少数の半正定値基底行列の重み付き和で近似する.図1で示す通り,行列の半正定値性はスカラーやベクトルの非負値性の拡張概念であることから,PSDTFはNMFの自然な拡張となっている.PSDTFを音源分離に用いる場合は,入力となる半正定値行列は,各フレームにおける混合音の複素スペクトルとその共役スペクトルとの直積となる.NMFでは半正定値行列の対角成分,すなわち非負値ベクトルのみを対象としており,周波数ピン間の相関が無視されていることに注意する.一方,PSDTFでは位相情報を取り扱うことができるため,各フレームごとに独立にウィナーフィルタリングを行う(当該フレームの周波数ピンは相関を考慮しながら一挙に処理される)ことにより,音源信号の複素スペクトルを直接推定可能である.

NMFには様々な変種が存在するが,ある仮定のもとで,板倉・斎藤(IS)ダイバージェンスに基づくNMF(IS-NMF)[13](あるいはその拡張であるLogDetダイバージェンスに基づくPSDTF(LD-PSDTF)[11])を用いて音源分離を行うことが理論的に妥当となる.具体的には,音源信号の複素スペクトルが複素ガウス分布に従うと仮定すると,複素ガウス分布は再生性を持つことから,それらが重畳した混合音の複素スペクトルも複素ガウス分布に従う.したがって,パワースペクトルの加法的性が成立し,混合音スペクトルに対する複素ガウス尤度の最大化が,ISダイバージェンスの最小化と等価となる.

実際には,カルバック・ライブラー(KL)ダイバージェンスに基づくNMF(KL-NMF)を用いて音源分離を行うことが一般的である[14].具体的には,音源信号の振幅スペクトルがポアソン分布に従うことを仮定すると,ポアソン分布は再生性を持つことから,混合音の振幅スペクトルもポアソン分布に従う.したがって,振幅スペクトルの加法的性が成立し,混合音スペクトルに対するポアソン尤度の最大化が,KLダイバージェンスの最小化と等価となる.このモデルでは,振幅スペクトルは度数表,すなわち音粒子のヒストグラムであるという物理的に意味を持たない解釈がなされていることに注意する.にもかかわらず,実験的にはKL-NMFがIS-NMFよりも良い結果を与えることが多いことが知られている[15,16].振幅スペクトルの加法的性が成立するように,混合音に対して物理的に意味のある生成モデルを考えられるだろうか?

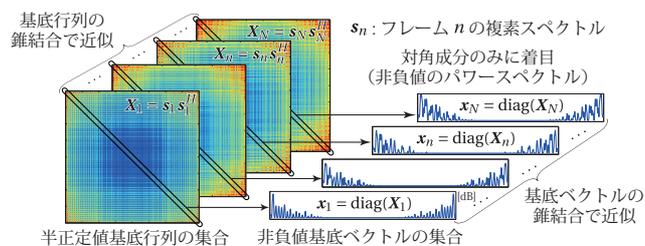


図1 PSDTFとNMFとの比較

この疑問を解消するため,最近,複素コーシー分布に基づくNMF(Cauchy NMF)が提案されている[17].具体的には,音源信号の複素スペクトルが複素コーシー分布に従うと仮定すると,複素コーシー分布も再生性を持つことから,それらが重畳した混合音の複素スペクトルも複素コーシー分布に従う(注1).このとき,近似的に振幅スペクトルの加法的性が成立する.また,複素コーシー分布は裾が重い分布であるため(平均・分散ともに定義されない),外れ値に頑健であるという性質を持っており,実在する音源信号をモデル化するのに都合が良い.

より一般には,音源信号の複素スペクトルが複素対称 α 安定(S α S)分布に従うと仮定すると,混合音の複素スペクトルも複素S α S分布に従う,すなわちFractionalスペクトル(注2)の加法的性が成立する[18].ここで, α は特性指数と呼ばれる分布の裾の重さを調節するパラメータ($0 < \alpha \leq 2$)で, $\alpha = 1, 2$ のときそれぞれ複素コーシー分布および複素ガウス分布に帰着する.この生成モデルのもとで,事後分布推論の観点から,混合音スペクトルを音源スペクトルに分解するための一般化ウィナーフィルタリングが導ける.しかし,複素S α S分布の確率密度関数は $\alpha = 1, 2$ のときを除いて解析的に表現することができないため,一般の α について,複素S α S分布に基づくNMFの最尤推定アルゴリズムは導出されていない.現在のところ,その特殊形であるIS-NMF[13]とCauchy NMF[17]に対する最尤推定アルゴリズムが個別に提案されているのみである.

本稿では,確率密度関数が解析的に表現できる複素 t 分布に基づくNMF(t -NMF)を提案する.複素 t 分布も自由度と呼ばれる裾の重さを調節するパラメータ ν ($0 < \nu$)を持ち, $\nu = 1, \infty$ のときそれぞれ複素コーシー分布および複素ガウス分布に帰着する.ただし,それ以外の一般の ν に対しては,分布の再生性は成立しない.したがって,スペクトルの加法的性も一般に成立しないが,複素S α S分布の軸とは異なるものの,Cauchy NMFとIS-NMFの中間的な性質を持つNMFを実現することは実用面と学術面で有意義である.まず,理論的にスペクトルの加法的性が保証されなくても,実験的に音源分離精度を最大化するよう自由度 ν を最適化できる.また,Cauchy NMFとIS-NMFを含む統一的な最尤推定アルゴリズムを導出することにより,学習時の振る舞いを明らかにすることができる.

(注1): 中心極限定理によると,独立同分布の確率変数を考えると,その確率分布がなんであれ,確率変数の和の確率分布は変数が多くなるにつれて,ガウス分布に収束する.ただし,これが成立するのは確率変数の分散が有限の場合のみであり,確率変数が従う分布の裾が $|x|^{-\alpha-1}$ ($0 < \alpha < 2$)の冪乗で減衰(分散は無限大に発散)する場合は,特性指数 α の安定分布に収束する.

(注2): 複素スペクトルに対して,各要素ごとに絶対値を取り, α 乗したものの, $\alpha = 1$ のとき振幅スペクトル, $\alpha = 2$ のときパワースペクトルとなる.

本稿ではさらに、多変量 t 分布に基づく PSDTF (t -PSDTF) を提案する。 t -PSDTF も t -NMF と同様に、 $\nu = 1, \infty$ のとき Cauchy PSDTF (本稿で提案) と LD-PSDTF [12] に帰着するが、多変量 t 分布の再生性は $\nu = 1$ のときも成立しなくなることに注意する。スペクトルの加法性が成立するのは $\nu = \infty$ である LD-PSDTF のみであるが、実際のデータに対して ν を実験的に最適化できることは実用上重要である。 t -PSDTF に対する最尤推定アルゴリズムは、Cauchy NMF [17], IS-NMF [13], Cauchy PSDTF, LD-PSDTF [12] をすべて特殊形に含む統一的な乗法更新則となっている。Cauchy NMF については、[17] で提案されているものとは異なり、見通しが良く、収束性が保証された新しい学習アルゴリズムを与える。

2. 複素ガウス分布に基づく NMF と PSDTF

本章では、 t -PSDTF を定式化するうえで必要となる数学的な背景について説明する。まず、コスト関数最小化の立場から、NMF [13, 19] と PSDTF [11, 12] の一般的な定式化を説明し、PSDTF が NMF の自然な拡張になっていることを示す。NMF と同様に PSDTF にも様々なコスト関数が考えうるが、LogDet (LD) ダイバージェンスに基づく PSDTF (LD-PSDTF) は、板倉・斉藤 (IS) ダイバージェンスに基づく NMF (IS-NMF) の数学的拡張となっている。IS-NMF や LD-PSDTF を音源分離に適用する例を通して、コスト関数の最小化は (多変量) 複素ガウス分布で与えられる尤度関数の最大化 (確率モデルの最尤推定) と等価であることを示す。

2.1 コスト関数最小化としての IS-NMF

NMF の目標は、与えられた非負値行列 $X = [x_1, \dots, x_N] \in \mathbb{R}_+^{M \times N}$ を、二つの非負値行列 $W = [w_1, \dots, w_K] \in \mathbb{R}_+^{M \times K}$ および $H = [h_1, \dots, h_K] \in \mathbb{R}_+^{N \times K}$ の積 $Y = [y_1, \dots, y_N] \in \mathbb{R}_+^{M \times N}$ で近似することである (図 2)。

$$X \approx W \times_2 H = WH^T \stackrel{\text{def}}{=} Y \quad (1)$$

ここで、 \times_2 は 2 モード積 [20] であり、 $w_k \in \mathbb{R}_+^M$ および $h_k \in \mathbb{R}_+^N$ はそれぞれ基底ベクトルおよび対応するアクティベーションベクトルである。 $K \ll \min(M, N)$ は基底数であり、予め決めておく必要がある。式 (1) は次式で書き直せる。

$$x_n \approx \sum_{k=1}^K h_{kn} w_k \stackrel{\text{def}}{=} y_n \quad (2)$$

すなわち、NMF の目標は、非負値ベクトルを少数の非負値ベクトルの錐結合で近似することである。

観測ベクトル x_n と再構成ベクトル y_n との間の誤差 $\mathcal{D}(x|y)$ を評価する尺度には、KL ダイバージェンス [19] や IS ダイバージェンス [13] などが知られている。本稿では、後者に着目する。

$$\mathcal{D}_{\text{IS}}(x_n|y_n) = \sum_{m=1}^M \left(\frac{x_{nm}}{y_{nm}} - \log \frac{x_{nm}}{y_{nm}} - 1 \right) \quad (3)$$

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

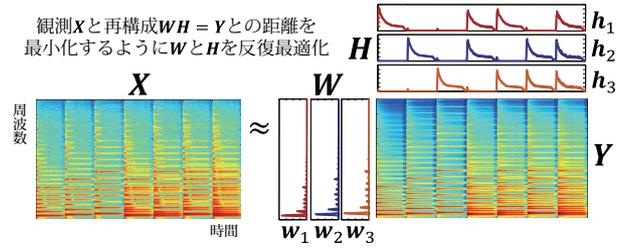


図 2 IS-NMF による混合音スペクトログラムの低ランク分解

2.2 コスト関数最小化としての LD-PSDTF

PSDTF の目標は、半正定値行列^(注3) の集合 (テンソル) $\mathcal{X} = [X_1, \dots, X_N] \in \mathbb{C}^{M \times M \times N}$ に対して、少数の半正定値行列の集合 (テンソル) $\mathcal{W} = [W_1, \dots, W_K] \in \mathbb{C}^{M \times M \times K}$ および $H = [h_1, \dots, h_K] \in \mathbb{R}_+^{N \times K}$ の積 $\mathcal{Y} = [Y_1, \dots, Y_N] \in \mathbb{C}^{M \times M \times N}$ で近似することである。

$$\mathcal{X} \approx \mathcal{W} \times_3 H \stackrel{\text{def}}{=} \mathcal{Y} \quad (4)$$

ここで、 \times_3 は 3 モード積 [20] であり、 $W_k \in \mathbb{C}^{M \times M} \succ 0$ および $h_k \in \mathbb{R}_+^N$ はそれぞれ基底行列およびアクティベーションベクトルである。式 (4) は次式で書き直せる。

$$X_n \approx \sum_{k=1}^K h_{kn} W_k \stackrel{\text{def}}{=} Y_n \quad (5)$$

すなわち、PSDTF の目標は、半正定値行列を少数の半正定値行列の錐結合で近似することである。

観測行列 X_n と再構成行列 Y_n との間の誤差 $\mathcal{D}(X_n|Y_n)$ を評価する尺度として、von-Neumann ダイバージェンス [21] や LogDet ダイバージェンス [22] を利用することが考えられる。本稿では、特に後者に着目する。

$$\mathcal{D}_{\text{LD}}(X_n|Y_n) = \text{tr}(X_n Y_n^{-1}) - \log |X_n Y_n^{-1}| - M \quad (6)$$

この値は常に非負であり、 $x_n = y_n$ のときのみ 0 をとる。通常の距離尺度と異なり、対称性が成り立たない。

式 (1) と式 (4)、式 (2) と式 (5)、式 (3) と式 (6) を比較すると、PSDTF は NMF の自然な拡張であることが明らかである (図 1)。半正定値の対角行列では、対角成分が固有値に対応し、それらはすべて非負値であることから、 $X_n = \text{diag}(x_n)$ 、 $Y_n = \text{diag}(y_n)$ 、 $W_k = \text{diag}(w_k)$ とすれば^(注4)、PSDTF は NMF に帰着する。すなわち、NMF では PSDTF のように、 M 個の要素間の相関を考慮することができない。

2.3 尤度最大化としての IS-NMF と LD-PSDTF

モノラル音響信号の音源分離を例に、確率モデルの最尤推定の観点から IS-NMF と LD-PSDTF を捉え直す。まず、観測される音響信号 (混合音) の複素スペクトログラムを $S = [s_1, \dots, s_N] \in \mathbb{C}^{M \times N}$ 、 k 番目の音源信号の複素スペクトログラムを $S_k = [s_{k1}, \dots, s_{kN}] \in \mathbb{C}^{M \times N}$ とする。ここで、 M

(注3): 半正定値行列と非負値行列とは異なる概念である。ある複素エルミート行列 $A = A^H \in \mathbb{C}^{M \times M}$ が半正定値行列である $\Leftrightarrow A = VV^H$ となる複素行列 V が存在すること $\Leftrightarrow A$ の M 個の固有値が全て非負であること。

(注4): $\text{diag}(a)$ はベクトル a を対角成分に持つ対角行列を意味する。

は周波数ビン数, N はフレーム数である. 観測した混合音が K 個の音源信号の瞬時混合であると仮定すると, 以下が成り立つ.

$$S = \sum_{k=1}^K S_k \Leftrightarrow s_n = \sum_{k=1}^K s_{kn} \quad (7)$$

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

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

$$s_{kn} | Y_{kn} \sim \mathcal{N}^c(s_{kn} | \mathbf{0}, Y_{kn}) \quad (8)$$

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

このとき, 音源信号が重畳して得られる観測信号も局所的に定常なガウス過程に従う. 式 (7) に着目すると, 複素ガウス分布の再生性から以下を得る.

$$s_n | Y_n \sim \mathcal{N}^c(s_n | \mathbf{0}, Y_n) \quad (9)$$

ただし, $Y_n = \sum_k Y_{kn}$ である. したがって, 観測される混合音スペクトル s_n に対する対数尤度関数は以下で与えられる.

$$\begin{aligned} \log p(s_n | Y_n) &= -M \log(\pi) - \log |Y_n| - s_n Y_n^{-1} s_n \\ &\stackrel{c}{=} -\log |Y_n| - \text{tr}(X_n Y_n^{-1}) \end{aligned} \quad (10)$$

ここで, $X_n = s_n s_n^H \succ \mathbf{0}$ とした. 式 (10) を式 (6) と比較すると, X_n 一定であるから, Y_n に関する対数尤度の最大化は, LogDet ダイバージェンスの最小化と等価である. したがって, 式 (10) が LD-PSDTF の確率モデルとなっている.

すべての半正定値行列を対角行列に限定すれば, LD-PSDTF は IS-NMF に帰着する. すなわち, 式 (8) や式 (9) で与えられる多変量複素ガウス分布の分散行列は対角行列となり, 音源スペクトル s_{kn} や混合音スペクトル s_n の各要素が独立であり, それぞれ異なる複素ガウス分布に従うことを意味する.

2.4 乗法更新アルゴリズム

LD-PSDTF のパラメータ W および H を求めるため, 補助関数法に基づく収束性の保証された乗法更新アルゴリズムが提案されている [11, 12]. 本稿では結果のみ記すと, 乗法更新則は以下で与えられる.

$$W_k \leftarrow W_k Q_k^{\frac{1}{2}} (Q_k^{\frac{1}{2}} W_k P_k W_k Q_k^{\frac{1}{2}})^{-\frac{1}{2}} Q_k^{\frac{1}{2}} W_k \quad (11)$$

$$h_{kn} \leftarrow h_{kn} \left(\frac{\text{tr}(X_n Y_n^{-1} W_k Y_n^{-1})}{\text{tr}(W_k Y_n^{-1})} \right)^{\frac{1}{2}} \quad (12)$$

ここで, P_k および Q_k は次式で求まる.

$$P_k = \sum_{n=1}^N h_{kn} Y_n^{-1} \quad Q_k = \sum_{n=1}^N h_{kn} Y_n^{-1} X_n Y_n^{-1} \quad (13)$$

h_{kn} の非負性と W_k の半正定値性は自然に保たれているが, $\text{tr}(W_k) = 1$ を満たすよう, 反復のたびに W_k および h_k のスケールを調整しておく.

すべての半正定値行列を対角行列に限定すれば, よく知られた IS-NMF の乗法更新アルゴリズム [23, 24] に帰着する. 具体的には, 式 (11) および式 (12) において, $X_n = \text{diag}(x_n)$, $Y_n = \text{diag}(y_n)$, $W_k = \text{diag}(w_k)$ とすると以下を得る.

$$w_{km} \leftarrow w_{km} \left(\frac{\sum_n x_{nm} h_{kn} / y_{nm}^2}{\sum_n h_{kn} / y_{nm}} \right)^{\frac{1}{2}} \quad (14)$$

$$h_{kn} \leftarrow h_{kn} \left(\frac{\sum_m x_{nm} w_{km} / y_{nm}^2}{\sum_m w_{km} / y_{nm}} \right)^{\frac{1}{2}} \quad (15)$$

2.5 ウィナーフィルタリング

最終的に, 式 (8) および式 (9) から, 混合音スペクトル s_n が与えられたときの音源スペクトル s_{kn} の事後分布は多変量複素ガウス分布となり, その平均と分散は次式で与えられる.

$$\mathbb{E}[s_{kn} | s_n] = Y_{kn} Y_n^{-1} s_n \quad (16)$$

$$\mathbb{V}[s_{kn} | s_n] = Y_{kn} - Y_{kn} Y_n^{-1} Y_{kn} \quad (17)$$

この処理では, s_k の位相は s の位相とは異なる点に注意する. 各フレーム n ごとに周波数ビン間の相関を考慮しながら一挙に分離を行うことで, 高品質な分離が可能となる. 一方, IS-NMF では式 (16) や式 (17) における半正定値行列はすべて対角行列に制限されているため, 各周波数ビン n, m ごとに独立に分離が行われることになり, s_k の位相は s の位相と同一となる.

3. 複素 t 分布に基づく NMF と PSDTF

本章では, 多変量複素ガウス分布の代わりに, より裾の重い多変量複素 t 分布を尤度関数に用いることで, 外れ値に対して頑健な PSDTF (t -PSDTF) を提案する. 多変量複素 t 分布は, 多変量複素コーシー分布や多変量複素ガウス分布を特殊形として含む. 複素 t 分布に基づく NMF (t -NMF) も同様に定式化することができ, すべての半正定値行列を対角行列に限定すれば, t -PSDTF は t -NMF に帰着する. ただし, 確率分布の再生性が成立するのは, 一変量あるいは多変量複素ガウス分布と一変量複素コーシー分布のみである. この時に限り, すなわち, LD-PSDTF [12], IS-NMF [13], Cauchy NMF [17] に関してのみ, 混合音スペクトルに対する理論的に妥当な生成モデルを考えることができる. 提案する t -PSDTF はこれらをすべて包含する一般的なモデルである.

3.1 複素 α 安定分布と複素 t 分布

1. 章で述べた通り, 本来の目標は複素対称 α 安定 (SaS) 分布に基づく NMF や PSDTF の定式化であることを念頭に置き, 複素 SaS 分布と複素 t 分布の性質を整理しておく. 一変量の場合と多変量の場合で, 確率分布の再生性が成立する条件が異なることに注意する.

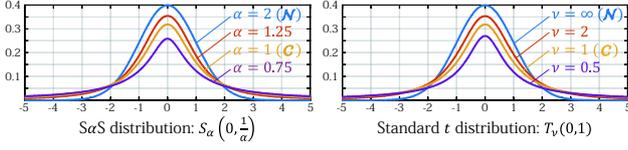


図3 実数の確率変数に対する一変量対称 α 安定分布と t 分布

3.1.1 複素 α 安定分布

一変量複素 α 安定分布の確率密度関数を $S_\alpha^c(\mu, \sigma^\alpha)$ と表す。ここで、 α は特性指数 ($0 < \alpha \leq 2$)、 μ は位置 (location) パラメータ ($\mu \in \mathbb{C}$)、 σ^α は分散 (dispersion) パラメータ ($\sigma > 0$) である。特に $\mu = 0$ のとき複素対称 α 安定 (complex symmetric α -stable: complex SaS) 分布と呼ぶ。一般の α について、 $S_\alpha^c(\mu, \sigma^\alpha)$ は解析的に書けないが、特性関数は簡潔に書ける。

$$\varphi(t) = \mathbb{E}\left[e^{i\text{Re}(t^* X)}\right] = \exp\left(-\left(\frac{1}{2}\sigma\right)^\alpha |t|^\alpha + i\text{Re}(t^* \mu)\right) \quad (18)$$

特性関数と確率密度関数はフーリエ変換の関係にあるので、二つの確率密度関数の畳み込み (二つの確率変数の和の確率分布) は、対応する特性関数の積と等価である。したがって、同じ特性指数の安定分布に従う確率変数の和は、同じ特性指数の安定分布に従う。すなわち、全ての α について再生性を持つ。

$$\begin{cases} X_1 \sim S_\alpha^c(\mu_1, \sigma_1^\alpha) \\ X_2 \sim S_\alpha^c(\mu_2, \sigma_2^\alpha) \end{cases} \Rightarrow X_1 + X_2 \sim S_\alpha^c(\mu_1 + \mu_2, \sigma_1^\alpha + \sigma_2^\alpha) \quad (19)$$

多変量複素 α 安定分布のうち、楕円型 (elliptically contoured) 分布の確率密度関数は $S_\alpha^c(\mu, \Sigma)$ と表す。ここで、 μ は位置ベクトル ($\mu \in \mathbb{C}^d$)、 Σ は分散行列 ($\Sigma \succ \mathbf{0} \in \mathbb{C}^{d \times d}$) である。この分布の特性関数は以下で与えられる。

$$\varphi(\mathbf{t}) = \mathbb{E}\left[e^{i\text{Re}(\mathbf{t}^H X)}\right] = \exp\left(-\left(\frac{1}{2}\mathbf{t}^H \Sigma \mathbf{t}\right)^{\frac{\alpha}{2}} + i\text{Re}(\mathbf{t}^H \mu)\right) \quad (20)$$

このことから、多変量の場合には、 $\alpha = 2$ の多変量複素ガウス分布の場合に限り再生性を持つ。

3.1.2 複素 t 分布

一変量複素 t 分布の確率密度関数を $\mathcal{T}_\nu^c(\mu, \sigma^2)$ と表す。ここで、 ν は自由度 ($\nu > 0$)、 μ は位置パラメータ ($\mu \in \mathbb{C}$)、 σ^2 は分散パラメータ ($\sigma > 0$) である。一般の ν について、 $\mathcal{T}_\nu^c(\mu, \sigma^2)$ は解析的に書ける。

$$\mathcal{T}_\nu^c(x|\mu, \sigma^2) = \frac{2\Gamma(\frac{2+\nu}{2})}{\pi\nu\sigma^2\Gamma(\frac{\nu}{2})} \left(1 + \frac{2}{\nu} \frac{|x - \mu|^2}{\sigma^2}\right)^{-\frac{2+\nu}{2}} \quad (21)$$

したがって、 $\nu \rightarrow \infty$ の複素ガウス分布あるいは $\nu = 1$ の複素コーシー分布のときに限り再生性を持つ。

多変量複素 t 分布の確率密度関数を $\mathcal{T}_\nu^c(\mu, \Sigma)$ と表す。ここで、 μ は位置ベクトル ($\mu \in \mathbb{C}^d$)、 Σ は分散行列 ($\Sigma \succ \mathbf{0} \in \mathbb{C}^{d \times d}$) であり、 $\mathbf{z} = \mathbf{x} - \mu$ とすると、 $\mathcal{T}_\nu^c(\mu, \Sigma)$ は以下で与えられる。

$$\mathcal{T}_\nu^c(\mathbf{x}|\mu, \Sigma) = \frac{2^d \Gamma(\frac{2d+\nu}{2})}{(\pi\nu)^d \Gamma(\frac{\nu}{2}) |\Sigma|} \left(1 + \frac{2}{\nu} \mathbf{z}^H \Sigma^{-1} \mathbf{z}\right)^{-\frac{2d+\nu}{2}} \quad (22)$$

多変量の場合には、 $\nu \rightarrow \infty$ の多変量複素ガウス分布のときに限り再生性を持つ。

3.1.3 複素 α 安定分布と複素 t 分布との比較

複素 α 安定分布では $\alpha = 1, 2$ の場合に、複素 t 分布では $\nu = 1, \infty$ の場合に、それぞれ複素コーシー分布と複素ガウス分布に帰着する。具体的には、一変量の場合は $S_1^c(\mu, \sigma) = \mathcal{T}_1^c(\mu, \sigma^2) = \mathcal{C}^c(\mu, \sigma^2)$ および $S_2^c(\mu, \sigma^2) = \mathcal{T}_\infty^c(\mu, 2\sigma^2) = \mathcal{N}^c(\mu, 2\sigma^2)$ 、多変量の場合は $S_1^c(\mu, \Sigma) = \mathcal{T}_1^c(\mu, \Sigma) = \mathcal{C}^c(\mu, \Sigma)$ および $S_2^c(\mu, \Sigma) = \mathcal{T}_\infty^c(\mu, 2\Sigma) = \mathcal{N}^c(\mu, 2\Sigma)$ が成立する。

参考のため、実数の確率変数に対する一変量の対称 α 安定分布と t 分布の確率密度関数を図3に示す。ここで、 α 安定分布の分散パラメータは $S_2(0, 0.5) = \mathcal{T}_\infty(0, 1)$ かつ $S_1(0, 1) = \mathcal{T}_1(0, 1)$ となるよう適当に調整されている。裾の減衰の仕方は異なるものの、両分布ともに、ガウス分布とコーシー分布の中間的な特徴を持つ裾の重い分布を表現することができる。本研究では、確率密度関数の取り扱いの容易さから複素 t 分布に着目する。

3.2 尤度最大化としての定式化

PSDTF を頑健化するため、式 (9) で与えられる尤度関数において、多変量複素ガウス分布の代わりに、自由度 ν の多変量複素 t 分布を用いることを考える。

$$s_n | \mathbf{Y}_n \sim \mathcal{T}_\nu^c(s_n | \mathbf{0}, \mathbf{Y}_n) \quad (23)$$

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

$$\begin{aligned} \log p(\mathcal{X}|\mathcal{Y}) &= \sum_{n=1}^N \log p(s_n | \mathbf{Y}_n) \\ &\stackrel{c}{=} \sum_{n=1}^N \left(-\log |\mathbf{Y}_n| - \left(M + \frac{\nu}{2}\right) \log \left(1 + \frac{2}{\nu} \text{tr}(\mathbf{X}_n \mathbf{Y}_n^{-1})\right) \right) \end{aligned} \quad (24)$$

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

式 (24) を \mathcal{Y} に関して直接最大化することは困難であるため、補助関数法 (付録) を用いて式 (24) の下限関数を設計し、その下限関数を逐次最大化することで、式 (24) を間接的に最大化することを考える。

3.2.1 下限関数導出のための不等式

下限関数導出のために必要な不等式を説明する。まず、 $f(Z) = \log |Z|$ が凹関数であることに着目すると、 $f(Z)$ に対して一次のテイラー展開を行うことで、次式を得る。

$$-\log |Z| \geq -\log |\Omega| - \text{tr}(\Omega^{-1} Z) + M \quad (25)$$

ここで、 Ω は任意の半正定値行列 (展開点) であり、 M は正方行列 Z の行数あるいは列数である。等号成立条件は、 $\Omega = Z$ で与えられる。次に、任意の半正定値行列 A に対して $g(Z) = \text{tr}(Z^{-1} A)$ 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 [25] を適用可能である。

$$-\text{tr} \left(\left(\sum_{k=1}^K \mathbf{Z}_k \right)^{-1} A \right) \geq -\sum_{k=1}^K \text{tr} \left(\mathbf{Z}_k^{-1} \Phi_k A \Phi_k^H \right) \quad (26)$$

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

3.2.2 下限関数の導出

これら不等式を用いると、式 (24) に対する補助関数は以下の通り導出できる。

$$\begin{aligned}
& \log p(\mathcal{X}|\mathcal{Y}) \\
& \stackrel{c}{\geq} \sum_{n=1}^N \left(-\log |\Omega_n| - \text{tr}(\Omega_n^{-1} \mathbf{Y}_n) + M \right. \\
& \quad \left. - \left(M + \frac{\nu}{2} \right) \left(\psi_n + \psi_n^{-1} \left(1 + \frac{2}{\nu} \text{tr}(\mathbf{X}_n \mathbf{Y}_n^{-1}) \right) - 1 \right) \right) \\
& \stackrel{c}{=} \sum_{n=1}^N \left(-\text{tr}(\Omega_n^{-1} \mathbf{Y}_n) - \psi_n^{-1} \left(\frac{2M}{\nu} + 1 \right) \text{tr}(\mathbf{X}_n \mathbf{Y}_n^{-1}) \right) \\
& \geq \sum_{n=1}^N \left(-\text{tr}(\Omega_n^{-1} \mathbf{Y}_n) \right. \\
& \quad \left. - \psi_n^{-1} \left(\frac{2M}{\nu} + 1 \right) \sum_{k=1}^K \text{tr} \left(h_{kn}^{-1} \mathbf{W}_k^{-1} \Phi_{kn} \mathbf{X}_n \Phi_{kn}^H \right) \right) \quad (27)
\end{aligned}$$

ここで、 Ω_n 、 ψ_n および Φ_{kn} は補助変数であり、等号成立条件（下限関数の最大化条件）は以下で与えられる。

$$\Omega_n = \mathbf{Y}_n \quad (28)$$

$$\psi_n = 1 + \frac{2}{\nu} \text{tr}(\mathbf{X}_n \mathbf{Y}_n^{-1}) \quad (29)$$

$$\Phi_{kn} = \mathbf{Y}_{kn} \mathbf{Y}_n^{-1} \quad (30)$$

3.3 乗法更新アルゴリズム

式 (27) の下限関数を \mathbf{W}_k および h_{kn} に関してそれぞれ偏微分してゼロとおくことで、以下の更新式を得る。

$$\mathbf{W}_k \leftarrow \mathbf{W}_k \mathbf{Q}_k^{\frac{1}{2}} \left(\mathbf{Q}_k^{\frac{1}{2}} \mathbf{W}_k \mathbf{P}_k \mathbf{W}_k \mathbf{Q}_k^{\frac{1}{2}} \right)^{-\frac{1}{2}} \mathbf{Q}_k^{\frac{1}{2}} \mathbf{W}_k \quad (31)$$

$$h_{kn} \leftarrow h_{kn} \left(\frac{\text{tr}(\pi_n \mathbf{X}_n \mathbf{Y}_n^{-1} \mathbf{W}_k \mathbf{Y}_n^{-1})}{\text{tr}(\mathbf{W}_k \mathbf{Y}_n^{-1})} \right)^{\frac{1}{2}} \quad (32)$$

ここで、 \mathbf{P}_k および \mathbf{Q}_k は次式で求まる。

$$\mathbf{P}_k = \sum_{n=1}^N h_{kn} \mathbf{Y}_n^{-1} \quad \mathbf{Q}_k = \sum_{n=1}^N h_{kn} \mathbf{Y}_n^{-1} (\pi_n \mathbf{X}_n) \mathbf{Y}_n^{-1} \quad (33)$$

また、 π_n は次式で求められる。

$$\pi_n = \frac{2M + \nu}{2\text{tr}(\mathbf{X}_n \mathbf{Y}_n^{-1}) + \nu} \quad (34)$$

したがって、 h_{kn} の非負性と \mathbf{V}_k の半正定値性は自然に保たれているが、 $\text{tr}(\mathbf{W}_k) = 1$ を満たすよう、反復ごとに \mathbf{W}_k および h_k をスケールしておく。

式 (31) および式 (32) は、式 (11) および式 (12) の自然な拡張となっている。実際、多変量複素 t 分布の自由度を $\nu \rightarrow \infty$ とすると、 $\pi_n \rightarrow 1$ となり、多変量複素ガウス分布に基づく PSDTF (LD-PSDTF) に帰着する。一方、 $\nu = 1$ とすると、多変量複素コーシー分布に基づく PSDTF が得られる。すなわち、提案する多変量複素 t 分布に基づく t -PSDTF では、 $\pi_n \mathbf{X}_n$ を仮想的な観測行列であるとみなして LD-PSDTF を行っていると解釈できる。ただし、 π_n は再構成行列 \mathbf{Y}_n に依存しているため、 \mathbf{Y}_n が更新されるたびに観測行列 $\pi_n \mathbf{X}_n$ も変化する。

t -PSDTF では、仮想的な観測行列 $\pi_n \mathbf{X}_n$ に対して、低ランクな再構成行列 \mathbf{Y}_n をフィットさせることが目的であるが、すでに $\pi_n \mathbf{X}_n$ には \mathbf{Y}_n の性質が反映されており、ある種の平滑化が行われていると見ることができる。この結果、観測データに対して過剰にフィットすることが抑制され、LD-PSDTF よりも外れ値に対して頑健な分解ができると考えられる。

3.4 t -NMF・IS-NMF・Cauchy NMF の比較

多変量複素 t 分布に基づく t -PSDTF の特殊形として、複素 t 分布に基づく NMF が得られる。2.2 節で説明した通り、対角行列 $\mathbf{X}_n = \text{diag}(\mathbf{x}_n)$ 、 $\mathbf{Y}_n = \text{diag}(\mathbf{y}_n)$ 、 $\mathbf{W}_k = \text{diag}(\mathbf{w}_k)$ を考えると、以下の更新式を得る。

$$w_{km} \leftarrow w_{km} \left(\frac{\sum_n (\pi_{nm} x_{nm}) h_{kn} / y_{nm}^2}{\sum_n h_{kn} / y_{nm}} \right)^{\frac{1}{2}} \quad (35)$$

$$h_{kn} \leftarrow h_{kn} \left(\frac{\sum_m (\pi_{nm} x_{nm}) w_{km} / y_{nm}^2}{\sum_m w_{km} / y_{nm}} \right)^{\frac{1}{2}} \quad (36)$$

ここで、 π_{nm} は次式で求まる。

$$\pi_{nm} = \frac{2 + \nu}{2x_{nm}y_{nm}^{-1} + \nu} \quad (37)$$

このとき、仮想的な観測データ $\pi_{nm}x_{nm}$ は次式となる。

$$\pi_{nm}x_{nm} = \left(\frac{2}{2 + \nu} y_{nm}^{-1} + \frac{\nu}{2 + \nu} x_{nm}^{-1} \right)^{-1} \quad (38)$$

これは、観測データ x_{nm} と再構成データ y_{nm} に対して、重み $\nu:2$ で調和平均をとることを意味し、自由度 ν が大きくなるほど、 x_{nm} を重視するようになり、 ν が小さくなるほど、 y_{nm} を重視するようになる。例えば、IS-NMF は $\nu \rightarrow \infty$ の t -NMF と等価であるため、観測データ x_{nm} のみに着目する。一方、Cauchy NMF は $\nu = 1$ の t -NMF と等価であるため、観測データ x_{nm} と再構成データ y_{nm} を $1:2$ の重みで考慮する。

t -NMF では、IS-NMF と Cauchy NMF に対して、見通しのよい統一的な乗法更新アルゴリズムを与える。特に、Cauchy NMF に対する乗法更新アルゴリズムは、[17] で提案されている二種類のアルゴリズムのいずれとも異なる。

4. 評価実験

t -NMF および t -PSDTF の音源分離精度や初期値依存性を確認するために行った評価実験について報告する。

4.1 実験条件

実験には、RWC 研究用音楽データベース: 楽器音 [26] 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の (C4, E4, G4, C4+E4, C4+G4, E4+G4, C4+E4+G4) を連結することで 14 秒の音響信号を合成した。サンプリング周波数は 16[kHz] とした。まず、ガウス窓を用いて各フレームの複素スペクトル $\{s_n\}_{n=1}^N$ を計算し、半正定値行列 $\mathbf{X}_n = s_n s_n^T$ とすることで観測データ $\{\mathbf{X}_n\}_{n=1}^N$ を得た。窓幅は 512 点、シフト長は 160 点とした ($M = 512$, $N = 1400$)。

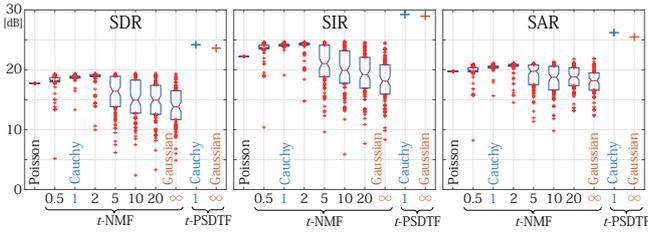


図 4 自由度を変化させた t -NMF と t -PSDTF による音源分離結果

次に、 t -NMF および t -PSDTF を用いて、混合音を 3 つの音高に対応する音源信号に分離することを試みた。 t -NMF では X_n の対角成分 (非負値のパワースペクトル) のみを観測データに用い、 $\nu = 0.5, 1$ (Cauchy NMF [17]), 2, 5, 10, 20, ∞ (IS-NMF [13]) のときを評価した。 t -PSDTF に関しては、 $\nu = 1$ (Cauchy PSDTF), ∞ (LD-PSDTF [12]) のときを評価した。ここで、 t -PSDTF は、多変量複素 t 分布がフーリエ変換 (線形変換) に対して閉じていることから、時間領域における等価な定式化が可能である [12]。このとき、ウィナーフィルタリングを行うと、各音源信号の各フレームの複素スペクトルではなく、局所信号が直接求まる。また、振幅スペクトログラムがポアソン分布に従うと仮定する KL-NMF とも比較した。

各手法はパラメータをランダムに初期化して、100 回試行した。計算量の削減のため、 t -PSDTF は対応する t -NMF の平均的な結果を用いて初期化を行った ($\nu = 1, \infty$ のとき)。音源分離結果は、BSS Eval Toolbox [27] 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

実験結果から、自由度を $\nu = 1, 2$ 程度に設定した t -NMF は、従来の IS-NMF や KL-NMF より優れた分離精度を達成した。どの自由度であっても t -NMF の分離精度の最大値は IS-NMF とほぼ同等であるが、自由度が $\nu = 1, 2$ 程度の t -NMF は初期値依存性が低く、安定して精度の良い音源分離が可能であることが分かった。ここで、自由度が $\nu = 1$ の t -NMF (Cauchy NMF) では、振幅スペクトルの加法性が成立するものの、自由度が $\nu = 2$ の t -NMF の方がわずかに優れている傾向がみられることが興味深い。このことは、スペクトルの加法性が理論的に成立しなくても、実用上は最適な自由度は異なる場合があることを示している。また、 ν の値を逆温度パラメータとみなし、少しずつ上昇させていくことで、よりよい局所解に誘導する一種のアニーリングも可能であると考えられる。

t -PSDTF は t -NMF に対する明確な優位性を示した。図 5 に示す通り、 t -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 W および音量変化 H を適切に推定できた。ここで、各基底行列 W_k における斜め縞の間隔は周期を表しており、 W_k の中心付近は巡回行列に近づいている。しかし、窓関数の影響で周辺部はそうはならないので、周波数成分間に相関が発生することは原理的に避けられない。 t -PSDTF はこの影響を考慮することで優れた音源分離性能を達成している。また、 t -PSDTF に対する乗法更新アルゴリズムが、 ν の値に関わらず、安定して収束することを確認した。

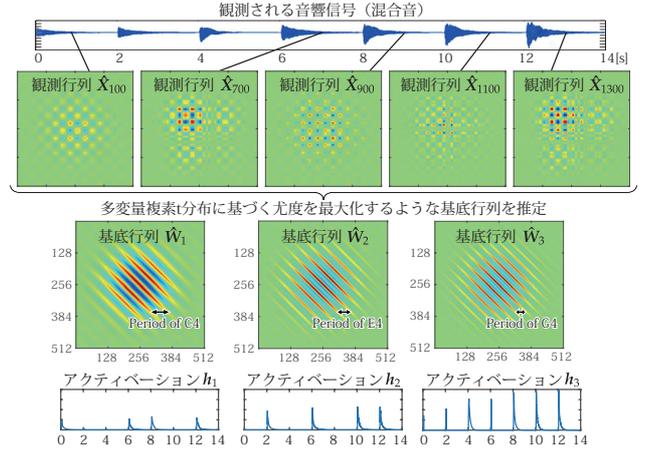


図 5 時間領域 t -PSDTF を用いたピアノ信号の分解結果

5. おわりに

本稿では、モノラル音響信号の音源分離を行うための新しい確率モデルである t -PSDTF およびその特殊形である t -NMF を提案した。実際の音響信号は優ガウス性を持つことが多いため、裾が重い確率分布である多変量複素 t 分布 (特殊形として多変量複素ガウス/コーシー分布を含む) を尤度関数に仮定した。このとき、パラメータを最尤推定するための効率的な乗法更新アルゴリズムを導出した。結果的に、 t -PSDTF は、従来知られていた LD-PSDTF [12], IS-NMF [13], Cauchy NMF [17] をその特殊形として含み、統一的な乗法更新アルゴリズムで学習が可能となった。特に、パワースペクトルの加法性に基づく IS-NMF と振幅スペクトルの加法性に基づく Cauchy NMF の中間的な性質をもつ NMF を実現したことの意義は大きい。実験の結果、 t -NMF は初期値依存性が低く、 t -NMF および t -PSDTF とともに高品質な音源分離ができることを確かめた。

今後は、多変量複素対称 α 安定分布を尤度関数とした SaS-PSDTF およびその特殊形である SaS-NMF の導出に取り組みたい。特に、一変量の複素 SaS 分布は、複素ガウス/コーシー分布を特殊な場合として含み、パワースペクトル ($\alpha = 2$) や振幅スペクトル ($\alpha = 1$) に限らず、あらゆる α に関して Fractional スペクトルの加法性を理論的に保証できる点で画期的である [18]。ただし、確率密度関数は解析的に書くことができず、最尤推定には特別な方法が必要になると考えられる。

付 録

目的関数 $\mathcal{F}(\theta)$ を変数 θ に関して最大化する問題を考える。補助関数法 [28, 29] を用いると、 $\mathcal{F}(\theta)$ を「間接的に」最大化することができる。まず、 $\mathcal{F}(\theta)$ に対する補助関数として、 $\mathcal{F}(\theta) \geq \mathcal{F}^+(\theta, \phi)$ となる下限関数 $\mathcal{F}^+(\theta, \phi)$ を考える。ここで、 ϕ は補助変数である。このとき、以下の反復更新則

$$\phi^{\text{new}} \leftarrow \arg\max_{\phi} \mathcal{F}^+(\theta^{\text{old}}, \phi) \quad (\text{A-1})$$

$$\theta^{\text{new}} \leftarrow \arg\max_{\theta} \mathcal{F}^+(\theta, \phi^{\text{new}}) \quad (\text{A-2})$$

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

多変量複素ガウス分布 ($Q(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$)

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

多変量複素 t 分布

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

多変量複素コーシー分布

$$C^c(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \left(\prod_{i=1}^d \left(i - \frac{1}{2}\right)\right) \frac{2^d}{\pi^d} |\boldsymbol{\Sigma}|^{-1} (1 + 2Q(\mathbf{x}))^{-\frac{2d+1}{2}}$$

謝辞: 本研究の一部は, JSPS 科研費 No. 24220006, 26700020, 26280089, JST CREST, 栢森情報科学振興財団の支援を受けた。

文 献

- [1] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman. Dynamic source separation using nonnegative factorizations: A unified view. *IEEE Signal Processing Magazine*, 31(3):66–75, 2014.
- [2] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009:1–17, 2009.
- [3] A. Ozerov, A. Liutkus, R. Badeau, and G. Richard. Coding-based informed source separation: Nonnegative tensor factorization approach. *IEEE Transactions on Audio, Speech, and Language Processing*, 21(8):1699–1712, 2013.
- [4] D. Liang, M. Hoffman, and D. Ellis. Beta process sparse nonnegative matrix factorization for music. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 375–380, 2013.
- [5] F. J. Canadas-Quesada, P. Vera-Candeas, N. Ruiz-Reyes, J. J. Carabias-Orti, and P. Cabanas-Molero. Percussive/harmonic sound separation by non-negative matrix factorization with smoothness/sparseness constraints. *EURASIP Journal on Audio, Speech, and Music Processing*, 26(1):1–17, 2014.
- [6] N. Souviraà-Labastie, E. Vincent, and F. Bimbot. Music separation guided by cover tracks: Designing the joint NMF model. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 484–488, 2015.
- [7] N. Q. K. Duong D. El Badawy, A. Ozerov. Relative group sparsity for non-negative matrix factorization with application to on-the-fly audio source separation. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 256–260, 2015.
- [8] K. O’Hanlon, M. Sandler, and M. D. Plumbley. Matrix factorisation incorporating greedy Hellinger sparse coding applied to polyphonic music transcription. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3112–3116, 2015.
- [9] C.-W. Wu and A. Lerch. Drum transcription using partially fixed non-negative matrix factorization with template adaptation. In *International Society for Music Information Retrieval Conference (ISMIR)*, 2015.
- [10] J. Le Roux, E. Vincent, Y. Mizuno, H. Kameoka, N. Ono, and S. Sagayama. Consistent Wiener filtering: Generalized time-frequency masking respecting spectrogram consistency. In *International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, pages 89–96, 2010.
- [11] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Beyond NMF: Time-domain audio source separation without phase reconstruction. In *International Society for Music Information Retrieval Conference (ISMIR)*, pages 369–374, 2013.
- [12] K. Yoshii, R. Tomioka, D. Mochihashi, and M. Goto. Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *International Conference on Machine Learning (ICML)*, pages 576–584, 2013.
- [13] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009.
- [14] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 177–180, 2003.
- [15] D. Fitzgerald, M. Cranitch, and E. Coyle. On the use of the beta divergence for musical source separation. In *Irish Signals and Systems Conference (ISSC)*, pages 1–6, 2009.
- [16] B. King, C. Févotte, and P. Smaragdis. Optimal cost function and magnitude power for NMF-based speech separation and music interpolation. In *International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6, 2012.
- [17] A. Liutkus, D. Fitzgerald, and R. Badeau. Cauchy nonnegative matrix factorization. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2015.
- [18] A. Liutkus and R. Badeau. Generalized Wiener filtering with fractional power spectrograms. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 266–270, 2015.
- [19] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *Neural Information Processing Systems (NIPS)*, pages 556–562, 2000.
- [20] T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009.
- [21] K. Tsuda, G. Rätsch, and M. K. Warmuth. Matrix exponentiated gradient updates for on-line learning and Bregman projection. *Journal of Machine Learning Research (JMLR)*, 6:995–1018, 2005.
- [22] B. Kulis, M. Sustik, and I. Dhillon. Low-rank kernel learning with Bregman matrix divergences. *Journal of Machine Learning Research (JMLR)*, 10:341–376, 2009.
- [23] 亀岡弘和. 非負値行列因子分解. 計測と制御, 51(9):835–844, 2012.
- [24] 亀岡弘和. 非負値行列因子分解の音響信号処理への応用. 日本音響学会誌, 68(11):559–565, 2012.
- [25] H. Sawada, H. Kameoka, S. Araki, and N. Ueda. Efficient algorithms for multichannel extensions of Itakura-Saito non-negative matrix factorization. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 261–264, 2012.
- [26] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In *International Conference on Music Information Retrieval (ISMIR)*, pages 229–230, 2003.
- [27] E. Vincent, R. Gribonval, and C. Févotte. Performance measurement in blind audio source separation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 14(4):1462–1469, 2006.
- [28] M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *International Conference on Machine Learning (ICML)*, pages 439–446, 2010.
- [29] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama. Convergence-guaranteed multiplicative algorithms for non-negative matrix factorization with beta divergence. In *International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 283–288, 2010.