

ベイズ高速多チャネル非負値行列因子分解に基づく ブラインド音源分離

市葉 拓馬¹ 須村 允亮^{2,3} Diego Di Carlo³
Aditya Arie Nugraha³ 坂東 宜昭^{4,3} 吉井 和佳^{2,3}

¹京都大学 工学部情報学科 ²京都大学 大学院情報学研究科 ³理化学研究所 AIP ⁴産業技術総合研究所 AIRC

1. はじめに

マイクアレイを用いた多チャネル音源分離は、雑音環境下での人間とロボットとの音声コミュニケーションや音響シーン分類・イベント検出のためのフロントエンドとしての役割が期待されている。近年、音源や環境に関する事前情報を用いないブラインド音源分離 (BSS) において顕著な技術的進歩が見られ、その最新手法として、高速多チャネル非負値行列因子分解 (FastMNMF) [1] が着目されている。本手法は MNMF [2] の特殊系であり、各音源・周波数に対するフルランクの空間共分散行列を同時対角化可能なものに制約することで、空間共分散行列のフルランク性 (非決定条件での適用可能性) を維持しつつモデルの自由度を効果的に制限することにより、高速かつ高精度な分離を実現できる。

ブラインド条件を緩和し、音源方向や伝達特性などの空間事前情報を利用した性能向上の試みも存在する。例えば、FastMNMF における同時対角化行列は決定性 BSS における分離行列と類似した役割を持つことから、所与の分離行列に基づく事前分布を導入して MAP 推定を行う手法が考案されている [3]。分離行列は、独立低ランク行列分析 (ILRMA) [4] などを用いて予め推定しておく。ILRMA は FastMNMF の特殊系であり、空間共分散行列をランク 1 に制約することで、決定条件に限定され、残響への頑健性が低下する一方で、質の悪い局所解に陥りにくい利点がある。ILRMA 自体に関しても、音源方向やマイク形状が既知である場合に、分離行列に対して各音源に対するステアリングベクトルの理論値に基づく事前分布を導入し、MAP 推定を行うことができる [5]。この研究で分離行列の収束保証付き反復最適化アルゴリズムが提案され、あとの研究 [3] でも利用されている。

本稿では、音源・空間事前情報を考慮するためのベイズ FastMNMF を提案する。特に空間事前情報の導入に関して、同時対角化行列 (疑似分離行列) ではなく、その逆行列 (疑似混合行列) に対して事前分布を導入する。本手法は、板倉・齊藤距離に基づくベイズ NMF と、潜在空間から観測空間への変換行列に対して事前分布を導入したベイズ主成分分析 (PCA) との統合とみなせる。ギブスサンプリングを用いて、全パラメータの厳密なベイズ推定が可能になることを示す。また、観測データと事前分布の相対的な重みの制御について考察する。

2. 提案法

M 個のマイクで観測した複素スペクトログラム $\mathbf{X} = \{\mathbf{x}_{ft} \in \mathbb{C}^M\}_{f,t=1}^{F,T}$ から、 N 個の音源イメージ $\mathbf{D} = \{\mathbf{d}_{nft} \in \mathbb{C}^M\}_{n,f,t=1}^{N,F,T}$ を推定する問題を考える。ここで、 F, T は周波数ビン数およびフレーム数を表す。

2.1 ベイズモデルの定式化

FastMNMF では、ある音源 n で発生した信号が、 M 個の伝達経路 (全音源で共通) を通ってマイクに到達し、それらが重畳したものが \mathbf{d}_{nft} であると考えられる。

$$\mathbf{d}_{nft} = \mathbf{A}_f \mathbf{s}_{nft} \quad (1)$$

ここで、 $\mathbf{A}_f \triangleq [\mathbf{a}_{f1}, \dots, \mathbf{a}_{fM}]^H \in \mathbb{C}^{M \times M}$ は周波数 f における混合行列であり、 $\mathbf{a}_{fm} \in \mathbb{C}^M$ は伝達経路 m に対応するステアリングベクトルである。 \mathbf{A}_f の逆行列が同時対角化行列 (分離行列) となることに注意されたい。 $\mathbf{s}_{nft} \triangleq [s_{nft1}, \dots, s_{nftM}]^T \in \mathbb{C}^M$ は M 個の伝達経路に対応する仮想音源であり、非決定条件でも決定条件と同様の定式化が可能になる。ここで、 $\mathbf{d}_{ft} \triangleq \sum_n \mathbf{d}_{nft} \in \mathbb{C}^M$ 、 $\mathbf{s}_{ft} \triangleq \sum_n \mathbf{s}_{nft} \in \mathbb{C}^M$ としておく。

仮想音源 s_{nftm} に対して、複素ガウス分布を仮定する。

$$s_{nftm} \sim \mathcal{N}_{\mathbb{C}}(0, \phi_{nftm}) \quad (2)$$

$$\phi_{nftm} \triangleq g_{nm} \lambda_{nft}, \quad \lambda_{nft} \triangleq \sum_{k=1}^K w_{nkf} h_{nkt} \quad (3)$$

ここで、 $\lambda_{nft} > 0$ は音源 n のパワースペクトル密度であり、 K 個の基底を持つ NMF で表現される。 $w_{nkf}, h_{nkt} > 0$ はそれぞれ、音源 n ・基底 k の周波数 f におけるパワースペクトル密度および時間 t における重みである。 $g_{nm} > 0$ は、音源 n に関する伝達経路 m の重みを表す。

このとき、観測音 \mathbf{x}_{ft} は次式で与えられる。

$$\mathbf{x}_{ft} = \mathbf{d}_{ft} + \mathbf{e}_{ft} \quad (4)$$

ここで、 \mathbf{e}_{ft} は複素ガウス分布に従う白色雑音とする。

$$\mathbf{e}_{ft} \sim \mathcal{N}(\mathbf{0}, \epsilon \mathbf{I}) \quad (5)$$

ここで、 ϵ は雑音のパワースペクトル密度を表す。これらをまとめて、最終的に \mathbf{x}_{ft} の生成モデルを得る。

$$\mathbf{x}_{ft} \sim \mathcal{N}_{\mathbb{C}}\left(\mathbf{0}, \sum_{n=1}^N \lambda_{nft} \mathbf{A}_f \text{Diag}(\mathbf{g}_n) \mathbf{A}_f^H + \epsilon \mathbf{I}\right) \quad (6)$$

ここで、 $\mathbf{g}_n = [g_{n1}, \dots, g_{nM}]^T \in \mathbb{R}_+^M$ とした。

推定すべき全パラメータに対し、事前分布を導入する。

$$w_{nkf} \sim \mathcal{G}(a_0^w, b_0^w), \quad h_{nkt} \sim \mathcal{G}(a_0^h, b_0^h), \quad g_{nm} \sim \mathcal{G}(a_0^g, b_0^g) \\ \epsilon \sim \mathcal{G}(a_0^\epsilon, b_0^\epsilon), \quad \mathbf{a}_{fm} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0^{-1}) \quad (7)$$

2.2 事後分布の推論

事後分布導出の準備として、仮想音源 s_{nftm} における基底 k の寄与分 s_{nkftm} と、それらに対応するパワースペクトル密度 ϕ_{nftm} および ϕ_{nkftm} を定義する。

$$s_{nftm} = \sum_{k=1}^K s_{nkftm}, \quad \phi_{nftm} = \sum_{k=1}^K \phi_{nkftm} \quad (8)$$

さらに、仮想音源の和 \mathbf{s}_{ft} およびそのイメージ \mathbf{d}_{ft} に対応する空間共分散行列 Φ_{ft} および Ψ_{ft} を定義する。

$$\Phi_{ft} \triangleq \sum_{n=1}^N \sum_{k=1}^K \text{Diag}(\phi_{nkft}), \quad \Psi_{ft} \triangleq \mathbf{A}_f \Phi_{ft} \mathbf{A}_f^H \quad (9)$$

ここで、 $\phi_{nkft} \triangleq [\phi_{nkft1}, \dots, \phi_{nkftM}]^T \in \mathbb{R}_+^M$ である。

潜在変数 \mathbf{S}, \mathbf{D} およびパラメータ $\mathbf{W}, \mathbf{H}, \mathbf{G}, \mathbf{A}, \epsilon$ のギブスサンプリングを行う。ただし、各 f, t, m における $\{s_{nkftm}\}_{n,k=1}^{N,K}$ は独立にサンプリングできず、伝承サンプリングを行う。ベイズPCAおよびベイズNMFと同様、各変数に関する条件付き事後分布が導出できる。

$$\mathbf{d}_{ft} \sim \mathcal{N}_C(\mathbf{m}_{ft}, \mathbf{V}_{ft}^{-1}), \quad \mathbf{s}_{ft} \leftarrow \mathbf{A}_f^{-1} \mathbf{d}_{ft} \quad (10)$$

$$\begin{cases} \mathbf{m}_{ft} \triangleq \Psi_{ft} (\Psi_{ft} + \epsilon \mathbf{I})^{-1} \mathbf{x}_{ft} \\ \mathbf{V}_{ft} \triangleq \Psi_{ft}^{-1} + \epsilon^{-1} \mathbf{I} \end{cases} \quad (11)$$

$$\mathbf{a}_{fm} \sim \mathcal{N}_C(\mathbf{u}_{fm}, \mathbf{R}_{fm}^{-1}) \quad (12)$$

$$\begin{cases} \mathbf{u}_{fm} \triangleq \mathbf{R}_{fm}^{-1} \left(\frac{1}{\epsilon} \sum_t \mathbf{s}_{ft} \mathbf{x}_{ftm}^* + \Lambda_0 \boldsymbol{\mu}_0 \right) \\ \mathbf{R}_{fm} \triangleq \frac{1}{\epsilon} \sum_t \mathbf{s}_{ft} \mathbf{s}_{ft}^H + \Lambda_0 \end{cases} \quad (13)$$

$$\epsilon \sim \text{GIG} \left(a_0^\epsilon - FTM, 2b_0^\epsilon, 2 \sum_{ft} |\mathbf{x}_{ft} - \mathbf{d}_{ft}|^2 \right) \quad (14)$$

$$\begin{cases} s \leftarrow s_{ftm} \\ \phi \leftarrow \phi_{ftm} \\ \text{for_each } n \in [1, \dots, N], k \in [1, \dots, K] \\ \quad s_{nkftm} \sim \mathcal{N}_C \left(\frac{\phi_{nkftm}}{\phi} s, \phi_{nkftm} - \frac{\phi_{nkftm}^2}{\phi} \right) \\ \quad s \leftarrow s - s_{nkftm} \\ \quad \phi \leftarrow \phi - \phi_{nkftm} \end{cases} \quad (15)$$

$$w_{nkf} \sim \text{GIG} \left(a_0^w - TM, 2b_0^w, 2 \sum_{tm} \frac{|s_{nkftm}|^2}{h_{nkt} g_{nm}} \right) \quad (16)$$

$$h_{nkt} \sim \text{GIG} \left(a_0^h - FM, 2b_0^h, 2 \sum_{fm} \frac{|s_{nkftm}|^2}{w_{nkf} g_{nm}} \right) \quad (17)$$

$$g_{nm} \sim \text{GIG} \left(a_0^g - KFT, 2b_0^g, 2 \sum_{kft} \frac{|s_{nkftm}|^2}{w_{nkf} h_{nkt}} \right) \quad (18)$$

3. 評価実験

シミュレーションデータを用いてベイズFastMNMFの動作を検証した。一辺10mの立方体形状の部屋の中央に半径1mmの円状マイクアレイを設置し ($M=5$)、 0° および 90° 度方向に1m離れた場所に音源を配置し

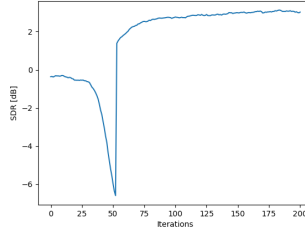


図1: SDR

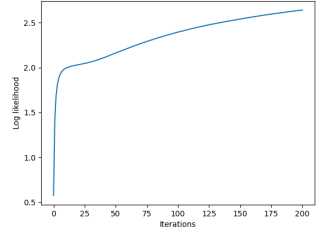


図2: 対数尤度

た ($N=2$). 床面からの高さはいずれも5mとした。音源信号として、CMU ARCTIC データセットから2発話を用いた。初期値は、 $\mathbf{W}, \mathbf{H}, \mathbf{G}$ はすべて1、 $\epsilon = 10^{-6}$ とし、 \mathbf{A} は事前分布からのサンプリングを行った。ハイパーパラメータは $K=8, a_0^* = b_0^* = 1$ ($* \in [w, h, g]$)、 $a_0^s = 10^{-3}, b_0^s = 10, \boldsymbol{\mu}_0 = \mathbf{0}, \Lambda_0 = 10^{-7} \mathbf{I}$ とした。ギブスサンプリングの反復回数は200回とした。

反復に伴う信号対歪み比 (SDR) と対数尤度の変化を、図1と図2にそれぞれ示す。対数尤度は概ね単調増加している一方、現状ではSDRは単調に改善しておらず、通常のFastMNMFと比較して大幅に低いSDRに収束している。音源分離の肝となる混合行列 \mathbf{A} の事後分布 (12) を見ると、観測データと事前分布とが $1/\epsilon$ と Λ_0/T のスケールに比例する重みで考慮されており、両者を適切に設定することが重要であることが示唆される。

4. おわりに

本稿では、音源・空間事前情報を考慮できるベイズFastMNMFを提案した。技術的には、ベイズNMFとベイズPCAに着想を得て、両者を統合したギブスサンプリングアルゴリズムを導出した。これにより、ILRMAを含むFastMNMFの特殊系に関しても全てのパラメータのベイズ的取り扱いが可能になり、残響除去やダイアライゼーションなど複数のタスクを見通しよく統合するための土台ができた。今後、音源分離精度を改善し、初期値依存性を軽減するため、事前分布の影響についてさらに調査を進める予定である。

謝辞: 本研究の一部は、科研費 No. 20H00602 の支援を受けた。

参考文献

- [1] K. Sekiguchi et al. Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation. 28:2610–2625, 2020.
- [2] H. Sawada et al. Multichannel extensions of non-negative matrix factorization with complex-valued data. 21(5):971–982, 2013.
- [3] 加茂 佳吾 et al. 同時対角化行列の事前分布を用いた高速多チャネル非負値行列因子分解によるブラインド音源分離. In 日本音響学会春季研究発表会講演論文集, pages 297–300, 2020.
- [4] D. Kitamura et al. Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. 24(9):1626–1641, 2016.
- [5] 三井 祥幹 et al. 空間事前情報を用いた独立低ランク行列分析. In 信号処理シンポジウム, pages 360–365, 2017.