

# 新しい補助関数を用いた Cauchy NMF の高速最適化

高宗 典玄<sup>1</sup> 北村 大地<sup>1</sup> 小野 順貴<sup>2</sup> 猿渡 洋<sup>1</sup>

概要：本稿では、複素 Cauchy 分布音源モデルに基づく非負値行列因子分解 (Cauchy NMF) の高速な最適化アルゴリズムを提案する。Cauchy NMF は、複素 Cauchy 分布を用いた生成モデルに基づき、振幅スペクトルの加法性を正当化した NMF であり、音響信号への適用が期待されている。Cauchy NMF の既存の最適化アルゴリズムは、収束性を保証しているが多くの反復更新を要する手法と、比較的高速だが収束性が保証されていない手法の二種類が提案されている。そこで、既存のアルゴリズムとは異なる新しい補助関数を設計することで、より高速かつ収束性の保証された反復更新式を導出し、提案手法の有効性を実験的に示す。

## Fast Optimization in Cauchy NMF Based on a New Auxiliary Function

NORIHIRO TAKAMUNE<sup>1</sup> DAICHI KITAMURA<sup>1</sup> NOBUTAKA ONO<sup>2</sup> HIROSHI SARUWATARI<sup>1</sup>

### 1. はじめに

非負値行列因子分解 (nonnegative matrix factorization: NMF) [1], [2] は、非負の観測行列を非負値制約の下で低ランク近似分解するアルゴリズムである。非負の制約と低ランク近似により、観測行列に含まれる潜在的な因子 (頻出パターン等) が基底として選ばれる傾向にある性質から、NMF はこれまで多様なタスクに適用・拡張されている。特に音響信号処理においては、音源分離問題への活用 [3]–[8] が盛んである。

近年 Liutkus et al. によって提案された複素 Cauchy 分布生成モデルに基づく NMF (Cauchy NMF) [9] は、振幅スペクトルの加法性を期待値の意味で正当化しており、今後の音響信号処理への適用が期待されている。しかしながら、文献 [9] で提案されている Cauchy NMF の最適化アルゴリズムは、補助関数法 [10]–[13] (正確には majorization-equalization (ME) algorithm [13]) を用いることで収束性を保証する反面多くの反復更新を要する手法と、比較的高速だがヒューリスティックな反復更新のために収束性が保証されない手法の二種類であり、いずれも一長一短であった。本稿では、

既存の乗算更新則の導出で用いられている補助関数とは異なる新しい補助関数を設計し、majorization-minimization (MM) algorithm [10]–[12] を適用することで、より高速かつ収束性の保証された反復更新式を導出し、提案手法の有効性を実験的に確認する。

### 2. NMF における非負成分の加法性を正当化する生成モデル

#### 2.1 NMF と加法性仮定

NMF は、非負の観測行列  $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{F \times T}$  に対して、下記の最適化問題として定式化される。

$$\min_{\mathbf{W}, \mathbf{H}} \mathcal{D}(\mathbf{Y} | \mathbf{W}\mathbf{H}) \quad \text{s.t. } w_{fk}, h_{kt} \geq 0 \quad \forall f, k, \text{ and } t \quad (1)$$

ここで、 $\mathbf{W} \in \mathbb{R}_{\geq 0}^{F \times K}$  及び  $\mathbf{H} \in \mathbb{R}_{\geq 0}^{K \times T}$  は非負の基底行列及び係数行列と呼ばれ、 $w_{fk}$  及び  $h_{kt}$  はそれらの非負要素である。また、 $\mathcal{D}(\cdot)$  は任意の剥離度 (ダイバージェンス) であり、 $f=1, \dots, F$  及び  $t=1, \dots, T$  はそれぞれ  $\mathbf{X}$  の行及び列のインデックス、 $k=1, \dots, K$  は基底数  $K$  のインデックスを示す。NMF の目的は低ランク近似であるため、基底数は一般的に  $K \ll \min(F, T)$  と設定される。音響信号に NMF を適用する場合は、短時間フーリエ変換 (short-time Fourier transform: STFT) によって得られる複素スペクトログラム  $\mathbf{X} \in \mathbb{C}^{F \times T}$  の振幅  $|\mathbf{X}|^1$  あるいはパワー  $|\mathbf{X}|^2$  を非負の観測

<sup>1</sup> 東京大学

The University of Tokyo, Tokyo 113–8656, Japan

<sup>2</sup> 国立情報学研究所

National Institute of Informatics, Tokyo, 101–8430, Japan

行列  $\mathbf{Y}$  として用いることが一般的である。ここで、行列に対するドット付きの指数は要素毎の指数乗を示し、 $|\cdot|^*$  は要素毎の絶対値の \* 乗を示す。この場合、 $\mathbf{Y}$  に含まれる頻出スペクトルパターンが  $\mathbf{W}$  の列ベクトル  $\mathbf{w}_k$  (基底) として得られ、それらの時間的な強度変化 (アクティベーション) が  $\mathbf{H}$  の行ベクトル  $\mathbf{h}_k^T$  として得られる。

振幅スペクトログラムやパワースペクトログラムの NMF による分解は、次の加法性を仮定している。

$$|x_{ft}|^\alpha = \sum_k |x_{ft,k}|^\alpha \quad (2)$$

ここで、 $x_{ft}$  は複素スペクトログラム  $\mathbf{X}$  の要素 (複素数)、 $x_{ft,k}$  はその分解成分であり、 $x_{ft} = \sum_k x_{ft,k}$  を満たす。また、 $\alpha$  は領域パラメタであり、 $\alpha=1$  が振幅、 $\alpha=2$  がパワーの領域に対応する。一般的には、位相差による打消しのため  $\sum_k |x_{ft,k}|^\alpha \geq \sum_k x_{ft,k} = x_{ft}$  となり、式 (2) の加法性仮定は成立しない。しかしながら、再生性を持つ分布を複素数  $x_{ft,k}$  の生成モデルとして仮定した場合、式 (2) の加法性仮定を期待値の意味で正当化できることが明らかにされている。以下の節においてその具体例を示す。

## 2.2 原点对称複素ガウス分布に基づく NMF

板倉斎藤擬距離 (Itakura-Saito divergence: IS divergence)  $\mathcal{D}_{IS}$  を剥離度に用いる NMF (ISNMF) では、パワースペクトル ( $\alpha=2$ ) の加法性仮定が正当化できる [14]。これは ISNMF の最適化問題が、次式の生成モデルに基づく最尤推定と等価であることに由来する。

$$p(\mathbf{X}_k) = \prod_{ft} p(x_{ft,k}) \quad (3)$$

$$p(x_{ft,k}) = \mathcal{N}_c(0, \sigma_{ft,k}^2) = \frac{1}{\pi \sigma_{ft,k}^2} \exp\left(-\frac{|x_{ft,k}|^2}{\sigma_{ft,k}^2}\right) \quad (4)$$

ここで、 $\mathbf{X}_k \in \mathbb{C}^{F \times T}$  は  $x_{ft,k}$  を要素を持つ複素行列、 $\mathcal{N}_c(0, \sigma^2)$  は零平均かつ分散  $\sigma$  の原点对称 (isotropic) な複素 Gauss 分布である。複素数  $x_{ft,k}$  が確率密度分布 (4) から生成されるとき、その分散  $\sigma_{ft,k}^2$  は  $x_{ft,k}$  のパワーの期待値  $E[|x_{ft,k}|^2]$  である。さらに、複素 Gauss 分布の再生性から、 $\mathbf{X} = \sum_k \mathbf{X}_k$  及び  $x_{ft} = \sum_k x_{ft,k}$  もまた

$$p(\mathbf{X}) = \prod_{ft} p(x_{ft}), \quad p(x_{ft}) = \mathcal{N}_c(0, \sum_k \sigma_{ft,k}^2) \quad (5)$$

に従うため、パワースペクトル  $\sigma_{ft,k}^2$  の加法性が期待値の意味で成立する。 $\mathcal{D}_{IS}(|\mathbf{X}|^2 | \mathbf{W}\mathbf{H})$  の最小化は、 $\sigma_{ft,k}^2 = w_{fk} h_{kt}$  と置くことで式 (4) の生成モデルに基づく  $\mathbf{W}$  及び  $\mathbf{H}$  の最尤推定と等価となるため、パワースペクトルの加法分解の理論的正当性が保証される。

## 2.3 原点对称複素コーシー分布に基づく NMF

振幅領域 ( $\alpha=1$ ) での加法性仮定を正当化する NMF とし

て、近年 Cauchy NMF が提案されている [9]。Cauchy NMF では、次式の生成モデルを仮定して最尤推定を行う。

$$p(\mathbf{X}_k) = \prod_{ft} p(x_{ft,k}) \quad (6)$$

$$p(x_{ft,k}) = C_c(0, \rho_{ft,k}) = \frac{\frac{1}{\sqrt{2}} \rho_{ft,k}}{2\pi \left[ |x_{ft,k}|^2 + \left( \frac{1}{\sqrt{2}} \rho_{ft,k} \right)^2 \right]^{\frac{3}{2}}} \quad (7)$$

ここで、 $C_c(0, \rho)$  は最頻値 0 かつ尺度母数  $\rho$  の原点对称な複素 Cauchy 分布である [15, pp. 80–81]。複素数  $x_{ft,k}$  が確率密度分布 (7) から生成されるとき、その尺度  $\rho_{ft,k}$  は観測信号の振幅  $|x_{ft,k}|$  に対応する\*1。複素 Gauss 分布と同様に複素 Cauchy 分布も再生性を持つため、 $\mathbf{X} = \sum_k \mathbf{X}_k$  及び  $x_{ft} = \sum_k x_{ft,k}$  もまた

$$p(\mathbf{X}) = \prod_{ft} p(x_{ft}), \quad p(x_{ft}) = C_c(0, \sum_k \rho_{ft,k}) \quad (8)$$

に従い、振幅スペクトル  $\rho_{ft,k} = w_{fk} h_{kt}$  の加法性が期待値の意味で成立する。

以上より、振幅スペクトログラム  $|x_{ft}|$  を正当に加法分解する Cauchy NMF が新たに与えられ、今後の応用が期待されている。文献 [9] では、ME algorithm を用いた収束性保証付き乗算更新則 (Cauchy ME) と、収束性が保証されない代わりに高速な乗算更新則 (naive multiplicative update: Cauchy Naive) の二種類の最適化アルゴリズムが提案されている。次章では、既存の Cauchy ME で用いられる補助関数とは異なる新しい補助関数を設計することにより、高速かつ収束性の保証された新しい乗算更新則を提案する。

## 2.4 原点对称複素 Student's $t$ 分布を用いた一般化

ISNMF 及び Cauchy NMF を特殊形に含む一般化 NMF として、生成モデルに原点对称複素 Student's  $t$  分布を用いた Student's  $t$  NMF が Yoshii et al. によって提案されている [16]。本手法では、Student's  $t$  分布の自由度母数  $\nu$  が  $\nu \rightarrow \infty$  のとき ISNMF に対応し、 $\nu=0$  のとき Cauchy NMF に対応する。しかしながら、 $\nu \rightarrow \infty$  及び  $\nu=1$  以外の分布では再生性が無く、加法性は正当化されない。また、導出された乗算更新則は、全ての  $\nu$  に対してモデル  $\mathbf{W}\mathbf{H}$  をパワースペクトログラムと仮定 (即ち  $|\mathbf{X}|^2 \approx \mathbf{W}\mathbf{H}$ ) している。そのため、 $\nu=1$  において振幅スペクトログラムのモデル ( $|\mathbf{X}|^1 \approx \mathbf{W}\mathbf{H}$ ) を取り扱う場合は変数項の凹凸が変化し別の補助関数が必要となるため、本稿では Student's  $t$  NMF の詳細は割愛する。

\*1 文献 [9] では、式 (7) の尺度母数  $\rho$  の係数  $2^{-1/2}$  はない。これは文献 [9] が内部で参照している文献 [15, pp. 80–81] の誤植に由来する。係数  $2^{-1/2}$  を付与することで、 $|x_{ft,k}| = \rho_{ft,k}$  のときに尤度が最大となる。

### 3. 新しい補助関数に基づく Cauchy NMF の高速最適化

#### 3.1 コスト関数

Cauchy NMF の負対数尤度関数は次式で与えられる。

$$\mathcal{L} = -\log p(\mathbf{X}; \mathbf{W}, \mathbf{H}) \stackrel{\text{c}}{=} \sum_{f,t} \left\{ \frac{3}{2} \log \left[ 2|x_{ft}|^2 + \left( \sum_k \rho_{ft,k} \right)^2 \right] - \log \sum_k \rho_{ft,k} \right\} \quad (9)$$

ここで、 $\stackrel{\text{c}}{=}$  は定数項を除いて等しいことを示す。式 (9) に基づく  $\mathbf{W}$  及び  $\mathbf{H}$  の最尤推定問題は、次式で与えられる新しい剥離度の最小化問題と等価である。

$$\mathcal{D}_{\text{Cauchy}}(a|b) = \log \frac{a}{b} + \frac{3}{2} \log \left[ \frac{2a^2 + b^2}{3a^2} \right] \quad (10)$$

剥離度  $\mathcal{D}_{\text{Cauchy}}$  はダイバージェンスの公理 [17] を全て満たすため、以後 Cauchy divergence と呼ぶ。Figure 1 に Cauchy divergence  $\mathcal{D}_{\text{Cauchy}}(y=5|x)$  と、NMF の剥離度としてよく用いられる IS divergence, 一般化 Kullback-Leibler divergence (KL divergence), 二乗 Euclid 距離 (EU distance) の比較を示す。Cauchy divergence は IS divergence と同じくパラメタに関して非凸な関数 (例えば Fig. 1 (b) において  $x \approx 11.6$  に変曲点が存在する) でスケール不変性を持つ。

#### 3.2 補助関数の設計

コスト関数 (9) の最小化において、従来の Cauchy ME [9] の導出で用いられる補助関数を導出する。まず、式 (9) の第一項の  $\log$  関数 ( $\rho_{ft,k}$  に関する凹関数) に接線不等式を適用した後に、 $(\sum_k \rho_{ft,k})^2$  項及び  $-\log \sum_k \rho_{ft,k}$  関数 (いずれも  $\rho_{ft,k}$  に関する凸関数) にそれぞれ Jensen の不等式を適用する。

$$\begin{aligned} \mathcal{L} &\leq \sum_{f,t} \left[ \frac{3}{2\beta_{ft}} \left[ 2|x_{ft}|^2 + \left( \sum_k \rho_{ft,k} \right)^2 - \beta_{ft} \right] + \frac{3}{2} \log \beta_{ft} \right. \\ &\quad \left. - \log \sum_k \rho_{ft,k} \right] \\ &\leq \sum_{f,t} \left[ \frac{3}{2\beta_{ft}} \left( 2|x_{ft}|^2 + \sum_k \frac{\rho_{ft,k}^2}{\gamma_{ft,k}} - \beta_{ft} \right) + \frac{3}{2} \log \beta_{ft} \right. \\ &\quad \left. - \sum_k \gamma_{ft,k} \log \frac{\rho_{ft,k}}{\gamma_{ft,k}} \right] \\ &\equiv \mathcal{L}_{\text{conv}}^{\text{MM}} \quad (11) \end{aligned}$$

ここで、 $\beta_{ft}$  及び  $\gamma_{ft,k}$  は補助変数であり  $\beta_{ft}, \gamma_{ft,k} \geq 0$  及び  $\sum_k \gamma_{ft,k} = 1$  を満たす。式 (11) の等号成立条件はそれぞれ

$$\beta_{ft} = 2|x_{ft}|^2 + \left( \sum_k \rho_{ft,k} \right)^2 \quad (12)$$

$$\gamma_{ft,k} = \frac{\rho_{ft,k}}{\sum_{k'} \rho_{ft,k'}} \quad (13)$$

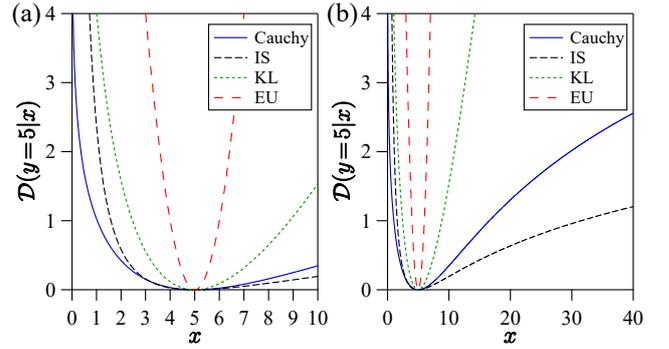


Fig. 1 Comparison of Cauchy divergence, IS divergence, KL divergence, and EU distance: (a) local and (b) global shapes.

で与えられる。式 (11) をパラメタで微分し停留点を求めることで、MM algorithm による乗算更新則が導出可能であるが、Cauchy ME では、ME algorithm を適用するために、 $\mathcal{L}_{\text{conv}}^{\text{MM}}$  に対してさらに  $-\log \rho_{ft,k} \leq \frac{\delta_{ft,k}}{\rho_{ft,k}} - 1 - \log \delta_{ft,k}$  なる不等式を適用し、 $\rho_{ft,k}$  の  $-\log$  項を  $-1$  次項に置き換えている。

$$\begin{aligned} \mathcal{L}_{\text{conv}}^{\text{MM}} &\leq \sum_{f,t} \left[ \frac{3}{2\beta_{ft}} \left( 2|x_{ft}|^2 + \sum_k \frac{\rho_{ft,k}^2}{\gamma_{ft,k}} - \beta_{ft} \right) + \frac{3}{2} \log \beta_{ft} \right. \\ &\quad \left. + \sum_k \gamma_{ft,k} \left( \frac{\delta_{ft,k}}{\rho_{ft,k}} - 1 - \log \frac{\delta_{ft,k}}{\gamma_{ft,k}} \right) \right] \\ &\equiv \mathcal{L}_{\text{conv}}^{\text{ME}} \quad (14) \end{aligned}$$

ここで、 $\delta_{ft,k}$  は補助変数であり  $\delta_{ft,k} \geq 0$  を満たす。式 (14) の等号成立条件は  $\delta_{ft,k} = \rho_{ft,k}$  で与えられる。Cauchy ME では、補助関数 (14) についてパラメタ  $w_{fk}$  及び  $h_{kt}$  の等高点を求めることで乗算更新則が導出される。

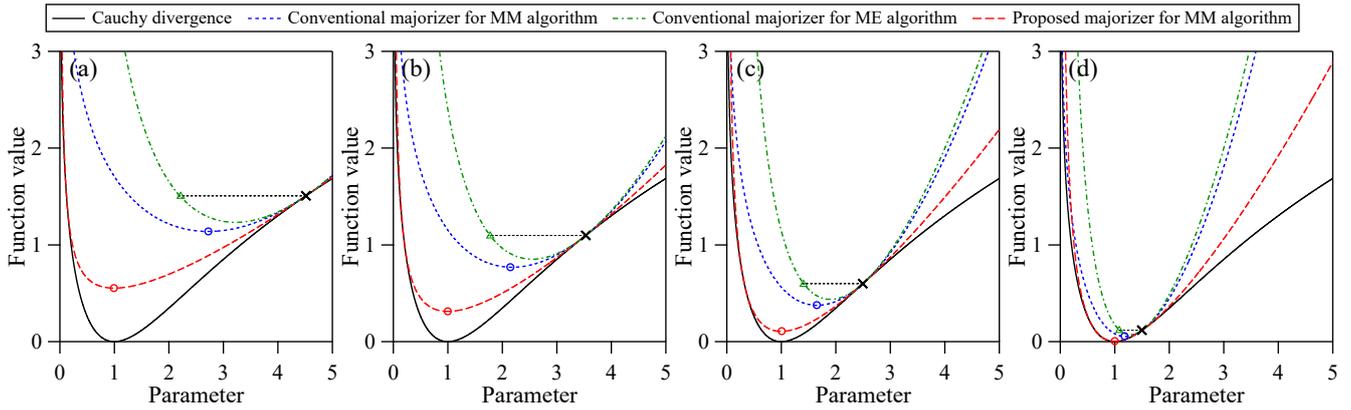
一方で、コスト関数 (9) の複数の対数項をあらかじめ一つにまとめてから接線不等式を適用することも可能である。すなわち、式 (9) を

$$\mathcal{L} = \sum_{f,t} \frac{3}{2} \log \left[ 2|x_{ft}|^2 \left( \sum_k \rho_{ft,k} \right)^{-\frac{2}{3}} + \left( \sum_k \rho_{ft,k} \right)^{\frac{4}{3}} \right] \quad (15)$$

と変形した上で、一つにまとめられた  $\log$  関数に接線不等式を適用し、その後に指数乗項 ( $-\frac{2}{3}$  乗項及び  $\frac{4}{3}$  乗項、いずれも凸関数) のそれぞれに Jensen の不等式を適用することで、新しい補助関数が得られる。

$$\begin{aligned} \mathcal{L} &\leq \sum_{f,t} \left[ \frac{3}{2\epsilon_{ft}} \left( 2|x_{ft}|^2 \left( \sum_k \rho_{ft,k} \right)^{-\frac{2}{3}} + \left( \sum_k \rho_{ft,k} \right)^{\frac{4}{3}} \right) \right. \\ &\quad \left. + \frac{3}{2} \log \epsilon_{ft} \right] \\ &\leq \sum_{f,t} \left[ \frac{3}{2\epsilon_{ft}} \left( 2|x_{ft}|^2 \sum_k \rho_{ft,k}^{-\frac{2}{3}} \gamma_{ft,k}^{\frac{5}{3}} + \sum_k \rho_{ft,k}^{\frac{4}{3}} \gamma_{ft,k}^{-\frac{1}{3}} - \epsilon_{ft} \right) \right. \\ &\quad \left. + \frac{3}{2} \log \epsilon_{ft} \right] \\ &\equiv \mathcal{L}_{\text{prop}}^{\text{MM}} \quad (16) \end{aligned}$$

ここで、 $\epsilon_{ft}$  は補助変数であり  $\epsilon_{ft} \geq 0$  を満たす。式 (16) の



**Fig. 2** Comparison of Cauchy divergence and its majorizers when intersection point is (a) 4.5, (b) 3.5, (c) 2.5, and (d) 1.5, where minimum point of Cauchy divergence is one. Cross marks show intersection points of all functions, circles and triangles are minimum and equalization points of majorizers, respectively.

等号成立条件は式 (13) 及び

$$\epsilon_{ft} = 2|x_{ft}|^2 \left( \sum_k \rho_{ft,k} \right)^{-\frac{2}{3}} + \left( \sum_k \rho_{ft,k} \right)^{\frac{4}{3}} \quad (17)$$

で与えられる。式 (16) は、パラメタ  $w_{fk}$  及び  $h_{kt}$  に関して容易に微分可能であり、MM algorithm が適用できる。

### 3.3 補助関数の性能評価

前節では、コスト関数 (9) の複数の対数項の取り扱い方によって異なる二種類の補助関数  $\mathcal{L}_{\text{conv}}^{\text{MM}}$  (あるいは  $\mathcal{L}_{\text{conv}}^{\text{ME}}$ ) 及び  $\mathcal{L}_{\text{prop}}^{\text{MM}}$  を設計した。従来の補助関数  $\mathcal{L}_{\text{conv}}^{\text{ME}}$  (あるいは  $\mathcal{L}_{\text{conv}}^{\text{MM}}$ ) では、パラメタを含む log 関数と  $-\log$  関数の二項を分離したまま、前者に接線不等式、後者に Jensen の不等式を適用していたのに対し、提案する新しい補助関数  $\mathcal{L}_{\text{prop}}^{\text{MM}}$  では複数の対数項を log 関数にまとめたうえで接線不等式を適用し、その後に各々の凸関数に対して Jensen の不等式を適用している。このような違いに対して設計される補助関数がどのように変化するかを確認するために、Cauchy divergence (10) に対して接線不等式のみを適用した次の三つの補助関数を比較する。

$$\mathcal{D}_{\text{conv}}^{\text{MM}} = \log \frac{a}{b} + \frac{3}{2} \left[ \frac{1}{\beta} (2a^2 + b^2 - \beta) + \log \beta - \log 3a^2 \right] \quad (18)$$

$$\mathcal{D}_{\text{conv}}^{\text{ME}} = \log a + \frac{\delta}{b} - 1 - \log \delta + \frac{3}{2} \left[ \frac{1}{\beta} (2a^2 + b^2 - \beta) + \log \beta - \log 3a^2 \right] \quad (19)$$

$$\mathcal{D}_{\text{prop}}^{\text{MM}} = \frac{3}{2\epsilon} \left[ \frac{2a^2 + b^2}{3a^2} \left( \frac{a}{b} \right)^{\frac{2}{3}} - \epsilon \right] + \frac{3}{2} \log \epsilon \quad (20)$$

ここで、 $\beta$ ,  $\delta$ , 及び  $\epsilon$  は、前節の補助関数設計時に登場した補助変数にそれぞれ対応する。但し、本節では一変数を対象としているため添え字はない。従って、式 (18), 式 (19), 及び式 (20) は、Cauchy NMF のコスト関数 (9) に

対する補助関数  $\mathcal{L}_{\text{conv}}^{\text{MM}}$ ,  $\mathcal{L}_{\text{conv}}^{\text{ME}}$ , 及び  $\mathcal{L}_{\text{prop}}^{\text{MM}}$  の一変数版にそれぞれ対応している (一変数であるため, Jensen の不等式による上限は元の関数に一致する)。

Figure 2 は、式 (10) 及びそれに対する三種類の補助関数 (18), (19), 及び (20) を  $a=1$  としたときのパラメタ  $b$  について示している。Figure 2 では、現在のパラメタ値 (×印) から、補助関数における最小点 (○印, MM algorithm の更新点) 及び等高点 (△印, ME algorithm の更新点) にパラメタが更新される。これらを比較すると、いずれのパラメタ値においても、提案する補助関数の最小点が解 ( $b=a=1$ ) 近傍に更新されることがわかる。

MM algorithm や ME algorithm のような補助関数を用いる最適化手法は、補助関数を設計する際に用いる不等式の種類や適用順序によって多様な更新アルゴリズムが得られる。補助関数の最小点にパラメタを更新する MM algorithm では、本来のコスト関数にできるだけ近い概形の補助関数を設計した方が、原理的により高速な更新アルゴリズムが得られる。一方で、補助関数の等高点にパラメタを更新する ME algorithm は、補助関数の近似精度に対するアルゴリズムの高速性は一概には言えず、等高点がコスト関数の最小点に近くなるような補助関数が望ましい。Cauchy divergence に関しては、log 項を先にまとめて接線不等式を適用する補助関数の方が、より良い更新アルゴリズムを導けることが Fig. 2 より確認できる。但し、Cauchy NMF に関しては、接線不等式の後に適用する Jensen の不等式や、 $f$ ,  $t$ , 及び  $k$  に関する周辺化の影響があるため、本節での議論が有効か否かは明らかではない。従って、提案する補助関数の Cauchy NMF の最適化に対する有効性については、次章で実験的に確認する。

### 3.4 乗算更新則の導出

MM algorithm では、補助関数をパラメタに関して偏微分し停留点を求めることで更新式が導出される。  $\partial \mathcal{L}_{\text{prop}}^{\text{MM}} / \partial w_{fk} = 0$

及び  $\rho_{f_t,k} = w_{f_k} h_{k_t}$  より、次式が得られる。

$$\frac{3}{2} \sum_t \left[ \frac{1}{\epsilon_{f_t}} \left( -\frac{4}{3} |x_{f_t}|^2 w_{f_k}^{-\frac{5}{3}} h_{k_t}^{-\frac{2}{3}} \gamma_{f_t,k}^{\frac{5}{3}} + \frac{4}{3} w_{f_k}^{\frac{1}{3}} h_{k_t}^{\frac{4}{3}} \gamma_{f_t,k}^{-\frac{1}{3}} \right) \right] = 0 \quad (21)$$

上式のうち負の項を移項して  $w_{f_k}$  について解くと

$$w_{f_k}^{\frac{1}{3}} \sum_t \frac{1}{\epsilon_{f_t}} \gamma_{f_t,k}^{-\frac{1}{3}} h_{k_t}^{\frac{4}{3}} = w_{f_k}^{-\frac{5}{3}} \sum_t \frac{1}{\epsilon_{f_t}} |x_{f_t}|^2 \gamma_{f_t,k}^{\frac{5}{3}} h_{k_t}^{-\frac{2}{3}} \quad (22)$$

$$w_{f_k} = \sqrt{\frac{\sum_t \frac{1}{\epsilon_{f_t}} |x_{f_t}|^2 \gamma_{f_t,k}^{\frac{5}{3}} h_{k_t}^{-\frac{2}{3}}}{\sum_t \frac{1}{\epsilon_{f_t}} \gamma_{f_t,k}^{-\frac{1}{3}} h_{k_t}^{\frac{4}{3}}}} \quad (23)$$

となる。等号成立条件 (13) 及び (17) を代入すると、次の乗算型更新式が得られる。

$$w_{f_k}^{(new)} \leftarrow w_{f_k} \sqrt{\frac{\sum_t |x_{f_t}|^2 (\sum_{k'} w_{f_{k'}} h_{k'_t})^{-1} c_{f_t}^{-1} h_{k_t}}{\sum_t (\sum_{k'} w_{f_{k'}} h_{k'_t}) c_{f_t}^{-1} h_{k_t}}} \quad (24)$$

ここで、 $c_{f_t} = 2|x_{f_t}|^2 + (\sum_{k'} w_{f_{k'}} h_{k'_t})^2$  とおいている。同様に、 $h_{k_t}$  の乗算型更新式は次式で得られる。

$$h_{k_t}^{(new)} \leftarrow h_{k_t} \sqrt{\frac{\sum_t |x_{f_t}|^2 (\sum_{k'} w_{f_{k'}} h_{k'_t})^{-1} c_{f_t}^{-1} w_{k_t}}{\sum_t (\sum_{k'} w_{f_{k'}} h_{k'_t}) c_{f_t}^{-1} w_{k_t}}} \quad (25)$$

行列  $\mathbf{W}$  及び  $\mathbf{H}$  で表記すると、次のようになる。

$$\mathbf{W}^{(new)} \leftarrow \mathbf{W} \circ \left( \frac{\{|\mathbf{X}|^2 \circ (\mathbf{W}\mathbf{H})^{-1} \circ \mathbf{C}^{-1}\} \mathbf{H}^T}{\{(\mathbf{W}\mathbf{H}) \circ \mathbf{C}^{-1}\} \mathbf{H}^T} \right)^{\frac{1}{2}} \quad (26)$$

$$\mathbf{H}^{(new)} \leftarrow \mathbf{H} \circ \left( \frac{\mathbf{W}^T \{|\mathbf{X}|^2 \circ (\mathbf{W}\mathbf{H})^{-1} \circ \mathbf{C}^{-1}\}}{\mathbf{W}^T \{(\mathbf{W}\mathbf{H}) \circ \mathbf{C}^{-1}\}} \right)^{\frac{1}{2}} \quad (27)$$

ここで、 $\mathbf{C} = 2|\mathbf{X}|^2 + (\mathbf{W}\mathbf{H})^2$  であり、行列に対する演算子  $\circ$  及び分数は、要素毎の積及び商をそれぞれ表す。

以上より、パラメタ  $\mathbf{W}$  及び  $\mathbf{H}$  を非負の乱数で初期化したうえで、乗算更新則 (26) 及び (27) を反復することで、コスト関数 (9) が減少する。提案手法は MM algorithm に基づいているため、単調非増加性が保証されている。

## 4. 提案手法の実験的評価

### 4.1 実験条件

本実験では、SiSEC2015 [18] の音楽音源分離タスクで公開されているデータセット MSD100 の楽曲を観測信号として用いた。MSD100 では開発データとテストデータでそれぞれ 50 曲のフルトラックが公開されているが、今回はテストデータの楽曲をアルファベット順に並べた際の上位 25 曲について、40 秒から 1 分 40 秒の区間を切り出したものを観測信号とした。25 曲の観測信号それぞれに対して STFT を適用し、振幅スペクトログラム  $|\mathbf{X}|^1$  を非負の観測行列  $\mathbf{Y}$  としたうえで、Cauchy ME, Cauchy Naive, 補助関数  $\mathcal{L}_{\text{conv}}^{\text{MM}}$  (11) に対して MM algorithm を適用して得られる乗算更新則 (Conventional MM), 及び式 (26) と式 (27)

Table 1 Experimental conditions

Window function	Hamming window
Window length	4096 points (92.9 ms)
Window shift length	2048 points (46.4 ms)
Size of observed matrix $\mathbf{Y}$	$F=2049$ and $T=1289$
Number of bases $K$	200
Initialization of $\mathbf{W}$ and $\mathbf{H}$	Uniform random values between (0, 1]

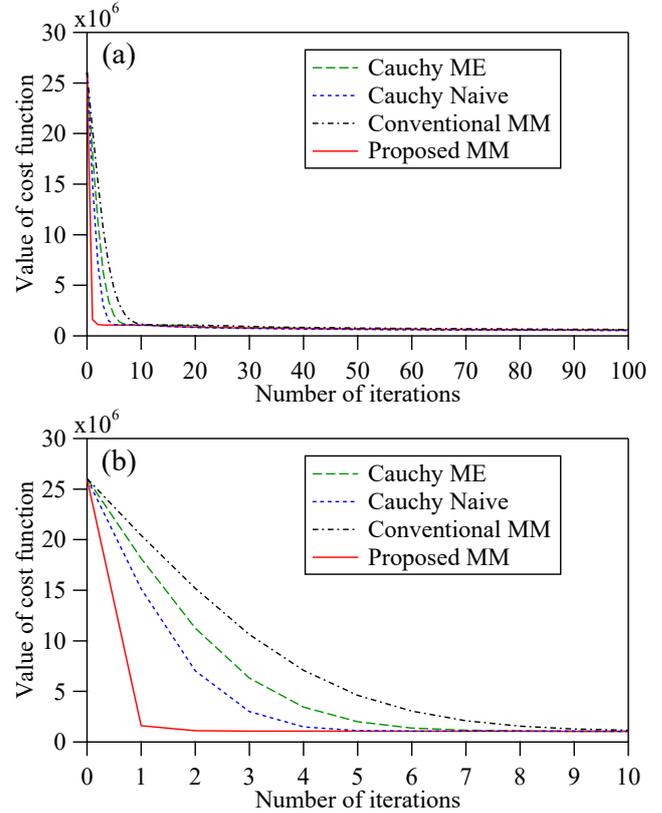


Fig. 3 Averaged convergence behaviors of each update algorithm: (a) between 0–100 iterations and (b) between 0–10 iterations.

の提案手法 (Proposed MM) の 4 種類の反復更新式を、乱数シードを変えて 10 回試行し、平均収束カーブを比較した。Conventional MM の導出については紙面の都合上割愛するが、次式のように得られる。

$$\mathbf{W}^{(new)} \leftarrow \mathbf{W} \circ \left\{ \frac{(\mathbf{W}\mathbf{H})^{-1} \mathbf{H}^T}{[3(\mathbf{W}\mathbf{H}) \circ \mathbf{C}^{-1}] \mathbf{H}^T} \right\}^{\frac{1}{2}} \quad (28)$$

$$\mathbf{H}^{(new)} \leftarrow \mathbf{H} \circ \left\{ \frac{\mathbf{W}^T (\mathbf{W}\mathbf{H})^{-1}}{\mathbf{W}^T [3(\mathbf{W}\mathbf{H}) \circ \mathbf{C}^{-1}]} \right\}^{\frac{1}{2}} \quad (29)$$

なお、これらの乗算更新則は Cauchy Naive の乗算更新則の乗算係数を  $\frac{1}{2}$  乗した形となっている。その他の詳細な実験条件を Table 1 にまとめる。

### 4.2 実験結果

Figure 3 は、25 曲 10 回試行すべての収束カーブの平均を手法毎に示した図である。結果をみると、提案手法の乗算更新則が反復初期 (特に 1 回目の反復) において高速にコス

**Table 2** Example of computational times (sec)

Cauchy ME	18.55
Cauchy Naive	17.06
Conventional MM	17.59
Proposed MM	17.26

ト関数を減少させていることがわかる。また、Conventional MM よりも Cauchy ME の方が高速に収束していることもわかる。これらの更新速度の違いは、Fig. 2 で示した各手法の解近傍への近づき方と整合していることが分かる。Cauchy Naive は理論的な収束性が保証されておらず、事実として観測信号と初期乱数モデルのスケールが大きく外れた場合は更新によってコスト関数値が上昇することもある。しかしながら、実用上は Cauchy ME より高速に収束している。楽曲毎の比較図は割愛するが、Fig. 3 の傾向は、全ての楽曲で同様に確認され、提案手法の優位性が示された。

Table 2 は、各手法の計算時間の一例として、Table 1 に示す実験条件での各手法の 200 回更新時の計算時間を 25 曲に関して平均した結果である。計算には MATLAB 9.1 (64 bit) 環境で Intel Core i7-6950X (3.0 GHz, 10 cores) の CPU を用いている。提案手法の乗算更新則は Cauchy ME よりも高速な反復計算となっている。また、Cauchy Naive と比較しても 200 回更新での差は 0.2 秒程度であり、より少ない回数で収束する提案手法が有効といえる。

## 5. まとめ

本稿では、Cauchy NMF の収束性保証付き高速最適化アルゴリズムを提案した。本手法では、従来の Cauchy NMF で用いられる補助関数とは異なる新しい補助関数を設計しており、MM algorithm を適用することで乗算更新則が導出される。実験的な比較の結果、提案法が従来の ME algorithm に基づく更新式 Cauchy ME 及び収束性の保証されていない Cauchy Naive のいずれよりも高速に収束することが確認された。

謝辞 本研究は、総合科学技術・イノベーション会議により制度設計された革新的研究開発推進プログラム (IMPACT)、JSPS 科研費 16H01735、及び SECOM 科学技術支援財団の助成を受けたものである。

## 参考文献

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2000, vol. 13, pp. 556–562.

[3] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. ASLP*, vol. 15, no. 3, pp. 1066–1074, 2007.

[4] P. Smaragdakis, B. Raj, and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel

mixtures," in *Proc. ICA*, 2007, pp. 414–421.

[5] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino, and S. Sagayama, "Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints," *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.5365–5368, 2012.

[6] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, "Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties," *IEICE Trans. Fundamentals*, vol. E97-A, no. 5, pp. 1113–1118, 2014.

[7] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, "Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration," *IEEE/ACM Trans. ASLP*, vol. 23, no. 4, pp. 654–669, 2015.

[8] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

[9] A. Liutkus, D. FitzGerald, and R. Badeau, "Cauchy nonnegative matrix factorization," in *Proc. WASPAA*, 2015, pp. 1–5.

[10] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of Computational and Graphical Statistics*, vol. 9, no. 1, pp.52–59, 2000.

[11] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, 2004.

[12] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[13] C. Févotte and J. Idier, "Algorithms for nonnegative matrix factorization with the  $\beta$ -divergence," *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.

[14] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura–Saito divergence: with application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[15] G. Samorodnitsky and M. S. Taqqu, *Stable non-Gaussian random processes: stochastic models with infinite variance*, Chapman & Hall/CRC Press, 1994.

[16] K. Yoshii, K. Itoyama, and M. Goto, "Student's  $t$  nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation," in *Proc. ICASSP*, 2016, pp. 51–55.

[17] S. Eguchi, "A differential geometric approach to statistical inference on the basis of contrast functionals," *Hiroshima Mathematical Journal*, vol. 15, no. 2, pp. 341–391, 1985.

[18] N. Ono, Z. Rafii, D. Kitamura, N. Ito, and A. Liutkus, "The 2015 signal separation evaluation campaign," *Proc. LVA/ICA*, 2015.