

# 非負値行列因子分解 NMF の多チャンネル拡張

## Multichannel Extensions of Nonnegative Matrix Factorization

澤田 宏      亀岡 弘和      荒木 章子      上田 修功  
日本電信電話株式会社    NTT コミュニケーション科学基礎研究所

Hiroshi Sawada      Hirokazu Kameoka      Shoko Araki      Naonori Ueda  
NTT Communication Science Laboratories, NTT Corporation

**アブストラクト** 非負値行列因子分解 (NMF: Nonnegative Matrix Factorization) は、音、画像、文書データ等、様々なデータの解析に用いられている。本発表では、特に音響信号を解析するという文脈で、NMF を多チャンネルに拡張する。音響信号を NMF で解析する際には短時間フーリエ変換を施すため、複素数の多次元ベクトルとして与えられるデータに対して、非負性というものを考えることになる。我々は、半正定値エルミート行列を用いてこの非負性をモデル化する。また、ユークリッド距離と板倉齋藤距離のそれぞれの規範に対して、効率的に収束する多チャンネル NMF のための更新式を提案する。楽器音の分離実験を行い、提案法の優位性を示す。

### 1 はじめに

非負値行列因子分解 (NMF: Nonnegative Matrix Factorization) は、与えられた非負値行列  $X$  をより小さい二つの非負値行列  $T$  と  $V$  に因子分解するものである。これは、画像のパーツ分解 [1]、文書クラスタリング [2]、音楽の自動採譜 [3] など、幅広い応用分野を持つ信号処理 / データ解析手法である。音響信号や音楽を対象とする場合は、それらの時間領域波形に短時間フーリエ変換 (STFT: Short-Time Fourier Transform) を施し、さらに各時間周波数スロットの値 (複素数) の絶対値やその二乗を計算する (式 1 参照) ことで、非負値行列  $X$  を作成する。図 1 に示す通り、分解後の行列  $T$  には頻出する音の周波数パターンが基底として抽出され、行列  $V$  には各基底がどのタイミングでアクティブになるかという情報が得られる。

さて近年、図 2 に示すように、音源分離や音源位置推定を目的として、複数のマイクロホンを用いる状況で NMF を多チャンネルに拡張する研究がいくつか行われている。瞬時混合に関わるもの [4], [5] では、各音源から各マイクロホンへのゲインがモデル化されるが、ゲインは非負値であるため、単純に行列からテンソルへ拡張するだけで良い。一方、畳み込み混合では、マイクペア間の位相差が音源分離や方向推定において重要な情報となるため、複素数を扱う必要がある。しかし、NMF の枠組みにおける

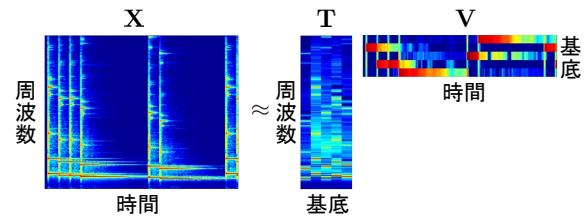


図 1: NMF の音楽信号への適用例

非負性をどのように考えるかは自明ではない。文献 [6], [7] では、音源モデルと混合過程を分け、音源を NMF でモデル化している。また、最適化の方法として EM アルゴリズムを用いているが、収束が非常に遅いという問題点がある。

本稿では、まず、NMF の多チャンネル拡張において、非負性というものをより明示的にモデル化した因子分解の形を定義する。そして、ユークリッド距離と板倉齋藤距離のそれぞれについて多チャンネルでの距離を定義する。その後、これらの距離を効率的に最小化する多チャンネル NMF の更新式を示す。この更新式は、通常の 1 チャンネルで広く用いられている Multiplicative 更新式の形をしており、チャンネル数を 1 にすれば通常の形に帰着することから、NMF の自然な多チャンネル拡張であると言える。また、既存の EM アルゴリズムに基づくものと比較し、提案の更新式が良好に収束することを実験で示す。

本稿では次に、提案する NMF の多チャンネル拡張において、新たに潜在変数  $Z$  を導入する。これにより、空間的特徴に従って NMF の基底がクラスタリングされる仕組みが NMF の枠組みにビルトインされる。この新たな拡張についても同様の Multiplicative 更新式が導かれる。そして、その NMF 結果に基づいてウィナーフィルタを構成することで、音源分離が達成される。楽器やヴォーカルが混合された音楽信号を用いた音源分離の実験では、特に板倉齋藤距離の場合に、以前の筆者らの音源分離手法 [8] に優る結果を得た。

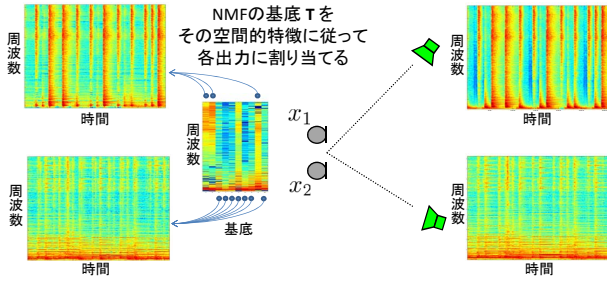


図 2: NMF の多チャンネル拡張により, 音源の位置や到来方向に従って NMF の基底がクラスタリングされ, 音源分離が行える.

## 2 非負値行列因子分解 NMF

本章では, 準備として, 通常の 1 チャンネル NMF の定式化とアルゴリズム [9]–[11] を説明する.

### 2.1 定式化

$\tilde{x}_{ij} \in \mathbb{C}$  を周波数ビン  $i$ , 時間フレーム  $j$  における STFT の値とする. NMF を適用するためには, この複素数を非負値  $x_{ij} \in \mathbb{R}_+$  に変換する必要がある. そのためには, 複素数の絶対値かその二乗

$$x_{ij} = \begin{cases} |\tilde{x}_{ij}| \\ |\tilde{x}_{ij}|^2 = \tilde{x}_{ij} \tilde{x}_{ij}^* \end{cases} \quad (1)$$

が用いられることが多い. ここで  $*$  は複素共役を示す. すべての  $i = 1, \dots, I$  と  $j = 1, \dots, J$  に関する上記変換の結果  $x_{ij}$  により, 行列  $\mathbf{X}$ ,  $[\mathbf{X}]_{ij} = x_{ij}$  を構成する.

NMF は,  $I \times J$  サイズの非負値行列を  $\mathbf{X}$  を,  $I \times K$  の非負値行列  $\mathbf{T}$  と  $K \times J$  の非負値行列  $\mathbf{V}$  の積に因子分解する.  $K$  は NMF の基底の数であり, 一般には解析する人が事前知識や経験に基づいて, あるいは試行錯誤的に決める. 因子分解後の二つの行列の要素  $t_{ik} = [\mathbf{T}]_{ik}$ ,  $v_{kj} = [\mathbf{V}]_{kj}$  は非負値  $t_{ik} \in \mathbb{R}_+$ ,  $v_{kj} \in \mathbb{R}_+$  に限定される.

NMF のアルゴリズムは, 与えられた行列  $\mathbf{X}$  とその因子分解形  $\mathbf{T}\mathbf{V}$  との距離を最小化することを目的として設計される.  $x_{ij}$  と等しくなるべき因子分解形の要素を

$$\hat{x}_{ij} = \sum_{k=1}^K t_{ik} v_{kj} \quad (2)$$

と定義すると, 当該行列間の距離は,

$$D_*(\mathbf{X}, \{\mathbf{T}, \mathbf{V}\}) = \sum_{i=1}^I \sum_{j=1}^J d_*(x_{ij}, \hat{x}_{ij}) \quad (3)$$

と定義される. ここで,  $d_*$  は要素毎の距離であり, 以下の 3 種類が広く用いられている.

ユークリッド距離の 2 乗

$$d_{Eu}(x_{ij}, \hat{x}_{ij}) = |x_{ij} - \hat{x}_{ij}|^2, \quad (4)$$

一般化 Kullback-Leibler (KL) divergence

$$d_{KL}(x_{ij}, \hat{x}_{ij}) = x_{ij} \log \frac{x_{ij}}{\hat{x}_{ij}} - x_{ij} + \hat{x}_{ij}, \quad (5)$$

板倉齋藤距離 (IS divergence)

$$d_{IS}(x_{ij}, \hat{x}_{ij}) = \frac{x_{ij}}{\hat{x}_{ij}} - \log \frac{x_{ij}}{\hat{x}_{ij}} - 1. \quad (6)$$

### 2.2 Multiplicative 更新式

式 (3) と式 (4), (5), (6) のいずれかにより定義される距離は, 以下の手順で最小化できる. まず, 行列  $\mathbf{T}$  と  $\mathbf{V}$  の要素を非負値でランダムに初期化する. その後, 以下の更新式を収束するまで繰り返し適用する [9], [11].

ユークリッド距離の 2 乗

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j x_{ij} v_{kj}}{\sum_j \hat{x}_{ij} v_{kj}}, \quad v_{kj} \leftarrow v_{kj} \frac{\sum_i x_{ij} t_{ik}}{\sum_i \hat{x}_{ij} t_{ik}} \quad (7)$$

一般化 Kullback-Leibler (KL) divergence

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j \frac{x_{ij}}{\hat{x}_{ij}} v_{kj}}{\sum_j v_{kj}}, \quad v_{kj} \leftarrow v_{kj} \frac{\sum_i \frac{x_{ij} t_{ik}}{\hat{x}_{ij}}}{\sum_i t_{ik}} \quad (8)$$

板倉齋藤距離 (IS divergence)

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_j \frac{x_{ij} v_{kj}}{\hat{x}_{ij} \hat{x}_{ij}}}{\sum_j \frac{v_{kj}}{\hat{x}_{ij}}}}, \quad v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_i \frac{x_{ij} t_{ik}}{\hat{x}_{ij} \hat{x}_{ij}}}{\sum_i \frac{t_{ik}}{\hat{x}_{ij}}}} \quad (9)$$

これらの更新式は, 更新前の要素 ( $t_{ik}$  あるいは  $v_{kj}$ ) に別の非負値が掛けられて更新後の値を得る形になっているため, multiplicative と呼ばれる. 行列  $\mathbf{T}$ ,  $\mathbf{V}$  の要素が全て非負値であれば, これらの更新式を何度適用しても非負値のままである. もし距離が完全に 0 になれば, 全ての  $i, j$  に関して  $x_{ij} = \hat{x}_{ij}$  となるため, 掛けられる値が 1 となり更新式では何も変化しないことが観察できる.

## 3 NMF の多チャンネル拡張

本章では, NMF の多チャンネル拡張を考える.

### 3.1 定式化

マイクロホンの数を  $M$  とすれば, ある時間周波数スロットでの STFT の値は  $M$  次元の複素ベクトル  $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_M]^T \in \mathbb{C}^M$  で表現される ( $m$  番目のマイクロホンの STFT の値が  $\tilde{x}_m$  となる). NMF を適用するためには, この複素ベクトルを非負性に関連する何かに変換する必要がある. 本稿では, ベクトルの外積

$$\mathbf{X} = \tilde{\mathbf{x}} \tilde{\mathbf{x}}^H = \begin{bmatrix} |\tilde{x}_1|^2 & \dots & \tilde{x}_1 \tilde{x}_M^* \\ \vdots & \ddots & \vdots \\ \tilde{x}_M \tilde{x}_1^* & \dots & |\tilde{x}_M|^2 \end{bmatrix} \quad (10)$$

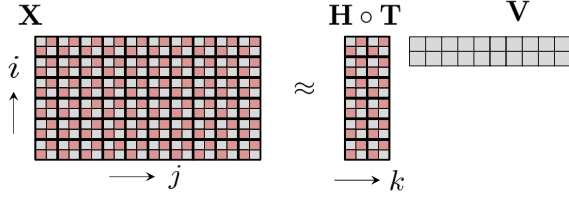


図 3: 多チャンネル NMF の模式図.  $I = 6, J = 10, K = 2, M = 2$  の場合. 非負値はグレー, 複素数は赤で表示.

かその要素の絶対値に平方根を施したもの

$$\mathbf{X} = \begin{bmatrix} |x_1| & \dots & |x_1 x_M|^{\frac{1}{2}} \text{sign}(x_1 x_M^*) \\ \vdots & \ddots & \vdots \\ |x_M x_1|^{\frac{1}{2}} \text{sign}(x_M x_1^*) & \dots & |x_M| \end{bmatrix}, \quad (11)$$

を考える. ここで,  $\text{sign}(x) = \frac{x}{|x|}$  である. いずれにせよ,  $\mathbf{X}$  はエルミート  $\mathbf{X}^H = \mathbf{X}$  かつ半正定値性 (すべての固有値が非負) を持つ行列となる. 次に, 図 3 の左側に示すような, サイズ  $I \times J$  の階層的な行列  $\mathbf{X}$  を考える. その  $i, j$  要素  $[\mathbf{X}]_{ij} = X_{ij}$  は, 周波数ビン  $i$ , 時間フレーム  $j$  に対応し, 上記の通り定義したサイズ  $M \times M$  の行列である.

次に, 因子分解後の形として, 図 3 の右側に示すように, サイズ  $I \times K$  の階層的な行列  $\mathbf{H} \circ \mathbf{T}$  とサイズ  $K \times J$  の非負値行列  $\mathbf{V}$  の積を考える. ここで,  $\mathbf{H}$  はサイズ  $I \times K$  の階層的な行列であり, その  $i, k$  要素  $[\mathbf{H}]_{ik} = H_{ik}$  は, 周波数ビン  $i$ , 基底  $k$  の空間的特徴を表現するサイズ  $M \times M$  のエルミート半正定値行列である. また,  $\circ$  は要素毎の積 (Hadamard product) であり,  $[\mathbf{H} \circ \mathbf{T}]_{ik} = H_{ik} t_{ik}$  となる.  $H_{ik}$  と  $t_{ik}$  の間のスケールの任意性を取り除くため,  $\text{tr}(H_{ik}) = 1, \text{tr}(\mathbf{H}) = \sum_{m=1}^M h_{mm}$  とする.

上記の因子分解後の形から,  $X_{ij}$  と等しくなるべき要素は

$$\hat{X}_{ij} = \sum_{k=1}^K H_{ik} t_{ik} v_{kj}, \quad (12)$$

となるが, これらの要素間の距離  $d_*(X_{ij}, \hat{X}_{ij})$  が定義できれば, 式 (3) と同様に, 全体の距離が

$$D_*(\mathbf{X}, \{\mathbf{T}, \mathbf{V}, \mathbf{H}\}) = \sum_{i=1}^I \sum_{j=1}^J d_*(X_{ij}, \hat{X}_{ij}) \quad (13)$$

と定義できる. 前章で示した 3 種類の距離について多チャンネル拡張が可能かどうかを検討したところ, 以下の 2 種類については複素正規分布から導かれる自然な定義が存在した [12].

ユークリッド距離の 2 乗

$$d_{Eu}(X_{ij}, \hat{X}_{ij}) = \|X_{ij} - \hat{X}_{ij}\|_F^2, \quad (14)$$

多チャンネル板倉齋藤距離 (IS divergence)

$$d_{IS}(X_{ij}, \hat{X}_{ij}) = \text{tr}(X_{ij} \hat{X}_{ij}^{-1}) - \log \det X_{ij} \hat{X}_{ij}^{-1} - M, \quad (15)$$

ここで  $\|\mathbf{B}\|_F^2 = \sum_{m=1}^M \sum_{n=1}^M |b_{mn}|^2$  はフロベニウスノルムの二乗である.

### 3.2 Multiplicative 更新式

式 (13) と式 (14) あるいは (15) により定義される距離は, 以下の更新式を繰り返すことで最小化される [12], [13]. これらの更新式は,  $M = 1, X_{ij} = x_{ij}, H_{ik} = 1$  と仮定すれば, 1チャンネル NMF の更新式, 式 (7) あるいは (9) に帰着される.

EU-NMF (ユークリッド距離の 2 乗)

$$t_{ik} \leftarrow t_{ik} \frac{\sum_j v_{kj} \text{tr}(X_{ij} H_{ik})}{\sum_j v_{kj} \text{tr}(\hat{X}_{ij} H_{ik})} \quad (16)$$

$$v_{kj} \leftarrow v_{kj} \frac{\sum_i t_{ik} \text{tr}(X_{ij} H_{ik})}{\sum_i t_{ik} \text{tr}(\hat{X}_{ij} H_{ik})} \quad (17)$$

$$H_{ik} \leftarrow \left( \sum_j v_{kj} X_{ij} \right) \left( \sum_j v_{kj} \hat{X}_{ij} \right)^{-1} H_{ik}. \quad (18)$$

$H_{ik}$  のエルミート半正定値性を保つために, 以下の後処理が必要となる. まず, エルミート性のために  $H_{ik} \leftarrow \frac{1}{2}(H_{ik} + H_{ik}^H)$  とする. 次に半正定値性のために, 固有値分解  $H_{ik} = \mathbf{U} \mathbf{D} \mathbf{U}^H$  を行い, 対角行列  $\mathbf{D}$  の負の値を 0 にした新たな行列  $\mathbf{D}'$  を用いて  $H_{ik} \leftarrow \mathbf{U} \mathbf{D}' \mathbf{U}^H$  と更新する.

IS-NMF (多チャンネル板倉齋藤距離)

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_j v_{kj} \text{tr}(\hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} H_{ik})}{\sum_j v_{kj} \text{tr}(\hat{X}_{ij}^{-1} H_{ik})}} \quad (19)$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_i t_{ik} \text{tr}(\hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} H_{ik})}{\sum_i t_{ik} \text{tr}(\hat{X}_{ij}^{-1} H_{ik})}} \quad (20)$$

$H_{ik}$  の更新のためには, 代数的 Riccati 方程式 (解き方を次項に示す)

$$\mathbf{H}_{ik}^H \mathbf{A} \mathbf{H}_{ik} = \mathbf{B} \quad (21)$$

を

$$\mathbf{A} = \sum_j v_{kj} \hat{X}_{ij}^{-1}, \quad \mathbf{B} = \mathbf{H}'_{ik} \left( \sum_j v_{kj} \hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} \right) \mathbf{H}'_{ik}$$

として解く. ここで,  $\mathbf{H}'_{ik}$  は更新前の行列を示す.

#### 3.2.1 Riccati 方程式の解き方

式 (21) を解くためには, まず,  $2M \times 2M$  行列

$$\begin{bmatrix} 0 & -\mathbf{A} \\ -\mathbf{B} & 0 \end{bmatrix} \quad (22)$$

の固有値分解を行う．理論的には  $M$  個の固有値が負になり，これらに対応する  $M$  個の固有ベクトルを  $\mathbf{e}_1, \dots, \mathbf{e}_M$  とする．しかし実際には，計算機での有限精度から発生する数値誤差に対応するため， $2M$  の固有値を昇順に従って並べ替え，最初の  $M$  個に対応する固有ベクトルを  $\mathbf{e}_1, \dots, \mathbf{e}_M$  とする．次に，これら  $M$  個の  $2M$  次元ベクトルを

$$\mathbf{e}_m = \begin{bmatrix} \mathbf{f}_m \\ \mathbf{g}_m \end{bmatrix} \quad \text{for } m = 1, \dots, M \quad (23)$$

と分解して表現する．ここで， $\mathbf{f}_m$  と  $\mathbf{g}_m$  は  $M$  次元ベクトルである．そして，行列  $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_M]$  と  $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_M]$  を構成して，

$$\mathbf{H}_{ik} = \mathbf{G}\mathbf{F}^{-1} \quad (24)$$

を計算する．再度数値誤差に対処するために， $\mathbf{H}_{ik} \leftarrow \frac{1}{2}(\mathbf{H}_{ik} + \mathbf{H}_{ik}^H)$  としてエルミート性を担保する．以上で，式 (21) の解  $\mathbf{H}_{ik}$  が得られる．

### 3.3 $\mathbf{H}_{ik}$ の解釈

提案する多チャンネル NMF アルゴリズム，すなわち式 (16)-(18) あるいは (19)-(21) の結果として，行列  $\mathbf{T}$ ,  $\mathbf{V}$ ,  $\mathbf{H}$  が得られる． $\mathbf{T}$  と  $\mathbf{V}$  の解釈については，通常の 1 チャンネル NMF と同様である．本節では，新たに導入した行列  $\mathbf{H}$  の解釈を説明する．

要素  $\mathbf{H}_{ik}$  をランク 1 モデル [14] で

$$\mathbf{H}_{ik} = \mathbf{h}_{ik}\mathbf{h}_{ik}^H + \epsilon_{ik}\mathbf{I}$$

と近似して考える．ここで， $\mathbf{h} = [h_1, \dots, h_M]^T \in \mathbb{C}^M$  は混合ベクトル， $\epsilon$  は小さな正のスカラー値， $\mathbf{I}$  はサイズ  $M \times M$  の単位行列である．この近似形から， $\mathbf{H}_{ik}$  の対角要素は， $k$  番目の NMF 基底の  $i$  番目の周波数ビンにおける各マイクロホンへのゲインの二乗を示し，非対角要素はマイクロホン間の位相差を示していることが分かる．

マイクロホンアレーのサイズ (マイクロホンの間隔) が小さい場合，ゲイン差よりも位相差の方が特徴を示しやすい．図 4 に，多チャンネル NMF で学習した行列  $\mathbf{H}$  から抽出した位相差を示す．基底番号  $k = 1, 2, 5, 6$  は似た空間的特徴を示していることから，これらの基底は同じ方向から到来した音に対応すると推測できる．他に基底番号  $k = 4, 7, 8, 9$  のグループがあるが，これらは別の方向から到来していると推測できる．このような空間的特徴の違いに基づいて NMF の基底をクラスタリングすることで，図 2 に示す通り，音源分離が達成される．クラスタリングの方法は次節で説明する．

### 3.4 基底のクラスタリング

本節では，多チャンネル NMF の定式化を少し変更することで， $K$  個の行列  $\mathbf{H}_{i1}, \dots, \mathbf{H}_{iK}$  を  $L$  個にクラスタリン

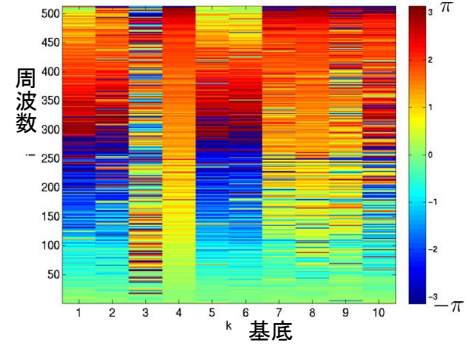


図 4: 学習された空間的特徴の例を  $\arg([\mathbf{H}_{ik}]_{12})$  の値として示す．これらは， $i$  番目の周波数， $k$  番目の NMF 基底に関する 1 つ目と 2 つ目のマイクロホン間の位相差である．

グすることを考える．まず，サイズ  $L \times K$  の新たな行列  $\mathbf{Z}$  を導入する．その要素  $z_{lk}$  は， $k$  番目の行列が  $l$  番目のクラスタに属する ( $z_{lk} = 1$ ) か属さない ( $z_{lk} = 0$ ) かを示す変数とする．そして，式 (12) における  $\mathbf{H}_{ik}$  を  $\sum_{l=1}^L \mathbf{H}_{il}z_{lk}$  で置き換えると

$$\hat{\mathbf{X}}_{ij} = \sum_{k=1}^K \left( \sum_{l=1}^L \mathbf{H}_{il}z_{lk} \right) t_{ik}v_{kj}. \quad (25)$$

となる．この様に  $\hat{\mathbf{X}}_{ij}$  のモデル化を変更しても，式 (14) のユークリッド距離の 2 乗や式 (15) の多チャンネル板倉齋藤距離はそのまま定義できるため，多チャンネル NMF として定式化できる．

さてここで，新たに導入した行列  $\mathbf{Z}$ ,  $[\mathbf{Z}]_{lk} = z_{lk}$  を行列  $\mathbf{T}$  や  $\mathbf{V}$  と同様に最適化するため， $z_{lk} \geq 0$  かつ  $\sum_{l=1}^L z_{lk} = 1$  を満たす連続値として  $z_{lk}$  を定義し直す．この変更により， $z_{lk}$  はその値自体から，事後確率  $p(z_{lk} = 1 | \mathbf{X}, \mathbf{T}, \mathbf{V}, \mathbf{H})$  に基づく  $z_{lk}$  の期待値に置き換わった，と捉えることができる．図 5 は，図 4 に示す  $K = 10$  個の基底を  $L = 2$  個のクラスタにクラスタリングした結果を示す．

さて，式 (12) を式 (25) を置き換えた多チャンネル NMF は，以下の更新式により実行できる [12], [13] ．

EU-NMF (ユークリッド距離の 2 乗)

$$t_{ik} \leftarrow t_{ik} \frac{\sum_l z_{lk} \sum_j v_{kj} \text{tr}(\mathbf{X}_{ij}\mathbf{H}_{il})}{\sum_l z_{lk} \sum_j v_{kj} \text{tr}(\hat{\mathbf{X}}_{ij}\mathbf{H}_{il})} \quad (26)$$

$$v_{kj} \leftarrow v_{kj} \frac{\sum_l z_{lk} \sum_i t_{ik} \text{tr}(\mathbf{X}_{ij}\mathbf{H}_{il})}{\sum_l z_{lk} \sum_i t_{ik} \text{tr}(\hat{\mathbf{X}}_{ij}\mathbf{H}_{il})} \quad (27)$$

$$z_{lk} \leftarrow z_{lk} \frac{\sum_{i,j} t_{ik}v_{kj} \text{tr}(\mathbf{X}_{ij}\mathbf{H}_{il})}{\sum_{i,j} t_{ik}v_{kj} \text{tr}(\hat{\mathbf{X}}_{ij}\mathbf{H}_{il})} \quad (28)$$

$$\mathbf{H}_{il} \leftarrow \left( \sum_k z_{lk}t_{ik} \sum_j v_{kj}\mathbf{X}_{ij} \right) \left( \sum_k z_{lk}t_{ik} \sum_j v_{kj}\hat{\mathbf{X}}_{ij} \right)^{-1} \mathbf{H}_{il} \quad (29)$$

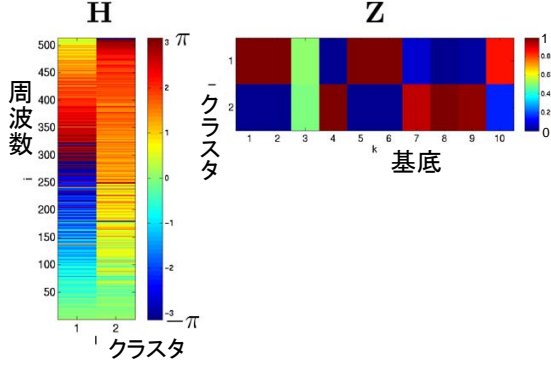


図 5: 図 4 に示した行列  $H$  の要素を 2 つにクラスタリングした様子．これらは  $\arg([H_{il}]_{12})$  の値として示されており,  $l$  はクラスタの番号を示す．行列  $Z$  は  $k$  番目の基底が  $l$  番目のクラスタに属するかどうかを連続値で示している．

### IS-NMF (多チャンネル板倉齋藤距離)

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_l z_{lk} \sum_j v_{kj} \text{tr}(\hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} H_{il})}{\sum_l z_{lk} \sum_j v_{kj} \text{tr}(\hat{X}_{ij}^{-1} H_{il})}}, \quad (30)$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_l z_{lk} \sum_i t_{ik} \text{tr}(\hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} H_{il})}{\sum_l z_{lk} \sum_i t_{ik} \text{tr}(\hat{X}_{ij}^{-1} H_{il})}}, \quad (31)$$

$$z_{lk} \leftarrow z_{lk} \sqrt{\frac{\sum_{i,j} t_{ik} v_{kj} \text{tr}(\hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} H_{il})}{\sum_{i,j} t_{ik} v_{kj} \text{tr}(\hat{X}_{ij}^{-1} H_{il})}}. \quad (32)$$

$H_{il}$  の更新のためには, 代数的 Riccati 方程式

$$H_{il}^H A H_{il} = B \quad (33)$$

を

$$A = \sum_k z_{lk} t_{ik} \sum_j v_{kj} \hat{X}_{ij}^{-1}, \quad (34)$$

$$B = H_{il} \left( \sum_k z_{lk} t_{ik} \sum_j v_{kj} \hat{X}_{ij}^{-1} X_{ij} \hat{X}_{ij}^{-1} \right) H_{il}. \quad (35)$$

として解く, なお,  $z_{lk}$  に関しては, 制約条件  $\sum_{l=1}^L z_{lk} = 1$  を満たすために, 正規化  $z_{lk} \leftarrow z_{lk} / (\sum_l z_{lk})$  を式 (28) あるいは (32) の後処理として行う．

### 3.5 信号の分離

観測信号に STFT を適用した際の, 周波数ビン  $i$ , 時間フレーム  $j$  における STFT の値をチャンネル方向に並べた  $M$  次元複素ベクトルを  $\tilde{x}_{ij}$  とする．また, 同様に  $\tilde{y}_{ij}^{(l)}$  を  $l$  番目の分離信号の STFT 値のベクトルとする．前節で示した基底のクラスタリング機能付き NMF の結果  $T, V, Z, H$  が得られれば, 分離信号は, 各  $m$  番目のチャンネルに適用される通常の 1 チャンネルウィナーフィルタ

$$[\tilde{y}_{ij}^{(l)}]_m = \frac{[H_{il}]_{mm} \sum_{k=1}^K z_{lk} t_{ik} v_{kj}}{\sum_{l=1}^L [H_{il}]_{mm} \sum_{k=1}^K z_{lk} t_{ik} v_{kj}} [\tilde{x}_{ij}]_m, \quad (36)$$

表 1: 18 秒の信号に対して NMF の更新式 1000 回の繰り返しに要した計算時間 (秒)．

	1 チャンネル	提案法	EM
EU-NMF	4.46	271.69	-
IS-NMF	14.16	1082.54	3802.97

か,  $M$  チャンネル全体へのウィナーフィルタ

$$\tilde{y}_{ij}^{(l)} = \left( \sum_{k=1}^K z_{lk} t_{ik} v_{kj} \right) H_{il} \hat{X}_{ij}^{-1} \tilde{x}_{ij}, \quad (37)$$

により得られる．ここで,  $\hat{X}_{ij}$  は (25) で定義されるものである．

### 4 実験

3 音源 2 マイクの状況での混合音を作成し実験を行った．具体的には, 残響時間 200ms の部屋で収録したインパルス応答と, Signal Separation Evaluation Campaign (SiSEC 2011) [15] の Professionally produced music recordings からの楽器音やヴォーカルの音源を畳み込んで, 混合音を作成した．サンプリング周波数は 16 kHz, 短時間フーリエ変換 STFT のフレーム幅は 64 ms, シフトは 16 ms とした．アルゴリズムは Matlab で実装し, Intel Core i7 965 (3.2 GHz) 上で実行した．複素ベクトル  $\tilde{x}$  を半正定値行列  $X$  に変換する際には, EU-NMF では式 (11) を, IS-NMF では式 (10) を用いた．

図 6 に, 3.2 節に示した多チャンネル NMF アルゴリズムの収束の様子を示す．比較のため, 1 チャンネル NMF の様子や, EM アルゴリズムを用いた多チャンネル NMF の様子も示す．図から, 提案の多チャンネル NMF アルゴリズムは, 既存の EM アルゴリズムに基づくものと比べて明らかに良い収束の様子を示し, 1 チャンネル NMF の収束の様子と似ていることが見て取れる．表 1 に 1000 回の繰り返しに要した計算時間を示す．NMF の多チャンネル拡張は実行にかなりの時間を要するが, 既存の EM アルゴリズムよりは軽減されている．また, IS-NMF は EU-NMF より時間を要する．

図 7 に, 3.4 節で提案した基底クラスタリング機能付きの多チャンネル NMF による音源分離の結果を示す．信号の分離には, EU-NMF では式 (36) を, IS-NMF では式 (37) を用いたものがそれぞれ良い結果を出したため, その組み合わせを採用した．IS-NMF は概ね EU-NMF より良い分離性能を示している．これは, 低域と高域を同等の重要度で扱う板倉齋藤距離 (IS-NMF) の良さによるものと考えられる．ユークリッド距離の場合 (EU-NMF) も, 前処理として周波数ビン毎にパワーの正規化を行ったが, 表記の結果となった．また IS-NMF は, 既存の音源分離手法 [8] (UBSS) よりも良い分離性能を達成している．楽

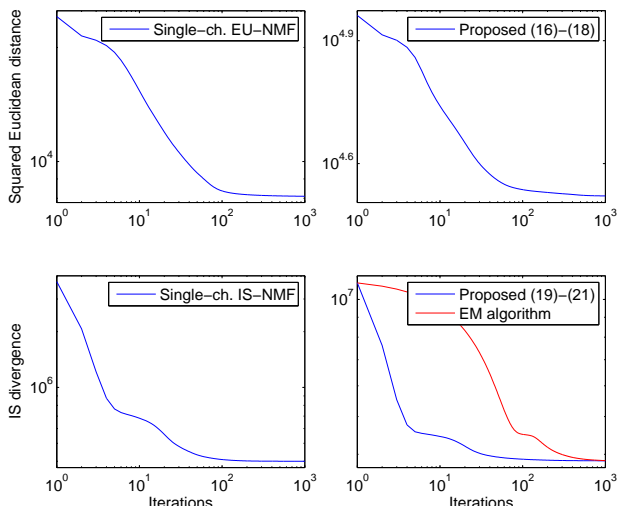


図 6: 18 秒の信号に対する NMF アルゴリズムの収束の様子。両対数プロットで表示。上部は EU-NMF，下部は IS-NMF。左側は 1 チャンネル，右側は多チャンネル。

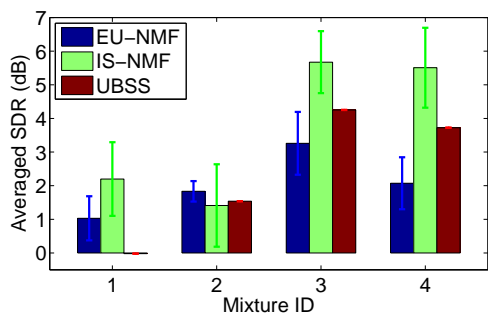


図 7: 音源分離性能。3 つの出力の SDR の平均値。NMF に基づく手法 (EU-NMF と IS-NMF) では、各混合音 ID に対して 10 回の試行を行った。比較のため、既存の音源分離手法 [8] (UBSS) による結果も示す。

器音を多く含む音楽信号に対して、頻出パターンを抽出する NMF の特性が良好に働いたものと言える。表 2 に計算時間を示す。表 1 と同様、多チャンネル NMF は多くの計算時間を要することが、現状の課題として認識される。

## 5 おわりに

多チャンネル NMF の新たな定式化とアルゴリズムを提案した。導出されたアルゴリズムは、これまでの EM アルゴリズムより明らかに良い収束の様子を示し、通常の 1 チャンネル NMF と同様に収束する。また、音源分離の実験を行い、既存の手法より良い分離性能を示した。多くの計算時間を要することが現状の課題であり、今後は、GPGPU などによる効率的な実装を検討していきたい。

表 2: 図 7 に結果を示した音源分離手法について、各混合音に対して要した計算時間 (秒)。

	ID=1	ID=2	ID=3	ID=4
EU-NMF	135.16	207.82	201.96	164.42
IS-NMF	436.42	696.60	675.95	527.63
UBSS [8]	9.98	13.72	12.60	14.42

## 参考文献

- [1] D. D. Lee and H. S. Seung. Learning the parts of objects with nonnegative matrix factorization. *Nature*, Vol. 401, pp. 788–791, 1999.
- [2] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. In *Proc. ACM SIGIR*, pp. 267–273, 2003.
- [3] P. Smaragdakis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. In *Proc. WAS-PAA 2003*, pp. 177–180, October 2003.
- [4] D. FitzGerald, M. Cranitch, and E. Coyle. Non-negative tensor factorisation for sound source separation. In *Proc. Irish Signals Syst. Conf.*, pp. 8–12, September 2005.
- [5] R. M. Parry and I. A. Essa. Estimating the spatial position of spectral components in audio. In *Proc. ICA 2006*, pp. 666–673. Springer, March 2006.
- [6] A. Ozerov and C. Févotte. Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. *IEEE Trans. Audio, Speech and Language Processing*, Vol. 18, No. 3, pp. 550–563, March 2010.
- [7] S. Arberet, A. Ozerov, N.Q.K. Duong, E. Vincent, R. Gribonval, F. Bimbot, and P. Vandergheynst. Nonnegative matrix factorization and spatial covariance model for underdetermined reverberant audio source separation. In *Proc. ISSPA 2010*, pp. 1–4, May 2010.
- [8] H. Sawada, S. Araki, and S. Makino. Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment. *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 19, No. 3, pp. 516–527, 2011.
- [9] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, Vol. 13, pp. 556–562, 2001.
- [10] 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.
- [11] 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 *Proc. MLSP 2010*, pp. 283–288, August 2010.
- [12] H. Sawada, H. Kameoka, S. Araki, and N. Ueda. Multichannel extensions of nonnegative matrix factorization with complex-valued data. Submitted to *IEEE Trans. ASLP*.
- [13] H. Sawada, H. Kameoka, S. Araki, and N. Ueda. Efficient algorithms for multichannel extensions of Itakura-Saito non-negative matrix factorization. In *Proc. ICASSP 2012*, pp. 261–264, March 2012.
- [14] N.Q.K. Duong, E. Vincent, and R. Gribonval. Underdetermined reverberant audio source separation using a full-rank spatial covariance model. *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 18, No. 7, pp. 1830–1840, 2010.
- [15] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe, and A. Benichoux. The 2011 signal separation evaluation campaign (sisec2011)-audio source separation. *Latent Variable Analysis and Signal Separation*, pp. 414–422, 2012.