

ケプストラム距離正則化に基づく多重音解析

樋口 卓哉^{1,a)} 亀岡 弘和^{1,2,b)}

概要: 本稿では、モノラルの音楽音響信号を楽音ごとのノートに対応した信号に分離する問題を取り扱う。モノラル音源分離の有効なアプローチとして、観測スペクトログラムが限られた種類のスペクトルパーツで構成されていることを仮定して行う、非負値行列因子分解 (Non-negative Matrix Factorization; NMF) が知られているが、スペクトルパーツの数を増やすとオーバーフィッティングが起こりやすくなり、推定した各スペクトルパーツが各楽音に対応したスペクトルにならないという問題点があった。一方で音楽信号の性質に着目してみると、楽曲中には高々何種類かの音色の音しか含まれていないことが多い。そこで、各楽音スペクトルが MFCC 特徴量空間で離散的なクラスタを形成しているという仮定を基に、NMF のフィッティングと、MFCC 特徴量空間における各スペクトルパーツのクラスタ形成を統合的に行うことで、前述したオーバーフィッティングを防ぎながら、音源分離と楽器音認識を同時に行う新しいアプローチを提案する。

1. はじめに

本稿では、モノラルの音楽音響信号を楽音ごとのノートに対応した信号に分離する問題を取り扱う。音楽音響信号の任意の短時間フレームにおけるスペクトルには、通常複数の楽音スペクトルが混在しており、何の仮定も置かずに個々のスペクトルを分離抽出することは難しい。従って、モノラル音源分離の問題においては、いかに一般性のある仮定を置いてうまく解を絞り込めるかが重要となる。モノラル音源分離に対する有効なアプローチとして、近年パートベースモデリングに基づく方法がある。このアプローチは、観測スペクトログラムの全体が限られた種類のパーツだけで構成されていると仮定し、その構成パーツを情報論的な規準に基づいて推定しようというものである。パートベースモデリングに基づく方法の代表格として、非負値行列因子分解 (Non-negative Matrix Factorization; NMF)[1][2] が知られているが、観測スペクトログラムを非負の行列と見なして NMF を適用するアイデアは、観測スペクトログラム全体が限られた種類のスペクトルパー

ツだけで構成されるという仮定がベースとなっている。実際、音楽信号は、限られた種類のスペクトルしか出現しえない、という仮定がある程度成り立つ場合が多く、このことが音楽スペクトログラムの分解方法として NMF が有効である理由の一つである。

しかし、NMF によるスペクトログラム分解法における本質的な問題として、スペクトルパーツの数を大きくした場合に、モデルスペクトルは多様なスペクトログラムを表現できるようになる反面、オーバーフィッティングを起こしやすくなってしまいうというものがある。すなわち、推定された各スペクトルパーツが各楽音に対応したスペクトルにならなくなるという問題である。これは、NMF におけるスペクトルパーツには非負制約以外の制約がなく、その制約の範囲内のいかなる形も取りうるものが主な原因である。

この問題を回避するために、音響信号の持つ独特の規則性に着目し、その規則性に基づいた新たな仮定によって解をさらに絞り込むことを考える。例えば、[3] では、正則化によって NMF を改善する試みがなされている。また音楽信号は、限られた種類のスペクトルパーツだけで構成される、という仮定がある程度成り立つ以外にも、独特の特徴を持っていると言える。例えば、限られた種類の基本周波数の音だけで構成されることや、限られた種類の音長だけで構成されることなどである。そのうちの一つとして、本稿では、限られた種類の楽器音だけで構成されるという特徴に着目する。この特徴に着目し、楽曲は限られた数の音色の音から構成されると仮定して解を限定する手法とし

¹ 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology,
The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

² 日本電信電話株式会社, コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
3-1 Morinosatowakamiya, Atsugi, Kanagawa 243-0198, Japan

a) higuchi@hil.t.u-tokyo.ac.jp

b) kameoka@hil.t.u-tokyo.ac.jp

て, [4], [5] などの手法が提案されており, その有効性が示されている.

楽曲中に高々何種類かの音色の音しか含まれていないと仮定するならば, 各楽音スペクトルは, 何らかの音色特徴量の空間で離散的なクラスタを形成しているはずである. 従ってスペクトルパーツが音色特徴量空間で離散的なクラスタを形成するようにしながら NMF のフィッティングを行えば, 上記のオーバーフィッティングを回避できる可能性があるばかりでなく, 音源分離と楽器音認識が同時に実現できると考えられる. そこで本稿では, 音色特徴量として MFCC を用いて, NMF によるフィッティング規準と MFCC 特徴量空間でのクラスタリング規準を組み合わせた最適化規準による, 新しいモノラル音源分離のアプローチを提案する.

2. 問題設定

2.1 非負値行列因子分解

まず, NMF によるスペクトログラムの分解表現について概説する. NMF によるスペクトログラムの分解表現では, スペクトログラムを非負値行列 $\mathbf{Y} = (Y_{\omega,t})_{\Omega,T} \in \mathbb{R}^{\geq 0, \Omega \times T}$ と見なし,

$$\mathbf{Y} \simeq \mathbf{H}\mathbf{U} \quad (1)$$

となるような非負値行列 $\mathbf{H} = (H_{\omega,k})_{\Omega,K} \in \mathbb{R}^{\geq 0, \Omega \times K}$, $\mathbf{U} = (U_{k,t})_{K,T} \in \mathbb{R}^{\geq 0, K \times T}$ を推定する. ただし $1 \leq \omega \leq \Omega$, $1 \leq t \leq T$ はそれぞれ周波数と時刻に対応するインデックス, $1 \leq k \leq K$ は NMF におけるスペクトルパーツのインデックスである. $\mathbf{y}_t := (Y_{1,t} \dots Y_{\Omega,t})^T$, $\mathbf{h}_k := (H_{1,k} \dots H_{\Omega,k})^T$ とすると式 (1) は

$$\forall t, \mathbf{y}_t \simeq \sum_{k=1}^K U_{k,t} \mathbf{h}_k, \forall k, U_{k,t} \geq 0, \mathbf{h}_k \geq 0 \quad (2)$$

と書き直せることから分かる通り, すべての t における観測データ \mathbf{y}_t を, K 種類のスペクトルパーツ $\mathbf{h}_1 \dots \mathbf{h}_K$ だけで近似しようとすることになる. 以下では $\mathbf{X} = \mathbf{H}\mathbf{U} = (X_{\omega,t})_{\Omega,T}$ をモデルスペクトログラムと呼ぶ.

2.2 音色クラスタリング規準の導入

前述した通り, 音楽信号には限られた種類の音色の音しか含まれていないとするならば, 各楽音のスペクトルは音色特徴量空間で図 1 のような離散的なクラスタを生成しているはずである. そこで音色特徴量空間においてスペクトルパーツをクラスタリングするための規準を導入する. 音色特徴量としてメル周波数ケプストラム係数 (MFCC) を用いると, k 番目のスペクトルパーツ $H_{1,k} \dots H_{\Omega,k}$ の M 次元 MFCC $\mathcal{H}_{1,k}, \dots, \mathcal{H}_{M,k}$ は MFCC の定義により以下の非線形変換

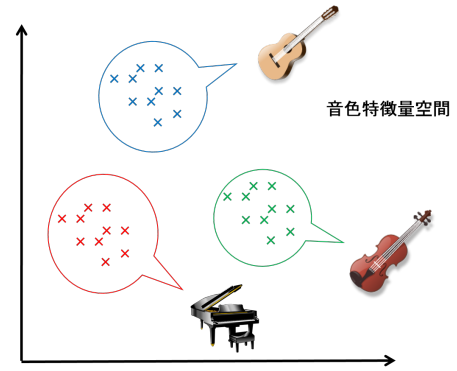


図 1 音色特徴量空間における楽器ごとのクラスタ形成の概念図

$$\mathcal{H}_{m,k} = \sum_{n=1}^N c_{m,n} \log \sum_{\omega} f_{n,\omega} H_{\omega,k} \quad (3)$$

により与えられる. ただし, $f_{n,\omega}$ はメルフィルタバンクにおける n 番目のメルフィルタの係数, $c_{m,n} \in \mathbb{R}$ は離散コサイン変換の係数で, m はケフレンシに対応するインデックスである. ここで $\vec{\mathcal{H}}_k := (\mathcal{H}_{1,k}, \dots, \mathcal{H}_{M,k})^T$ とし, $\vec{\mathcal{H}}_1, \dots, \vec{\mathcal{H}}_K$ を I 個のグループにクラスタリングするための規準を設けたい. そこで, 以下の規準

$$\begin{aligned} J(\mathbf{H}) &= \sum_i \sum_{k \in \mathcal{C}_i} \|\vec{\mathcal{H}}_k - \vec{\mu}_i\|_2^2 \\ &= \sum_i \sum_{k \in \mathcal{C}_i} \sum_m |\mathcal{H}_{m,k} - \mu_{m,i}|^2 \\ &= \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \left| \sum_n c_{m,n} \log \left(\sum_{\omega} f_{n,\omega} H_{\omega,k} \right) - \mu_{m,i} \right|^2 \end{aligned} \quad (4)$$

ただし, $\vec{\mu}_i = (\mu_{1,i} \dots \mu_{M,i})^T$ は i 番目のクラスタセントロイド (i 番目の代表的な音色特徴量値), \mathcal{C}_i は i 番目のクラスタに属するスペクトルパーツのインデックスの集合である. 以上の評価関数を音色クラスタリング規準と呼ぶ.

2.3 最適化問題の定式化

式 (4) の音色クラスタリング規準 $J(\mathbf{H})$ と I-divergence による NMF の規準

$$\begin{aligned} I(\mathbf{H}, \mathbf{U}) &= \sum_{\omega} \sum_t \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \log \sum_k H_{\omega,k} U_{k,t} \right. \\ &\quad \left. - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right) \end{aligned} \quad (5)$$

を統合し,

$$L(\mathbf{H}, \mathbf{U}) = I(\mathbf{H}, \mathbf{U}) + \lambda J(\mathbf{H}) \quad (6)$$

という評価関数を考える. $I(\mathbf{H}, \mathbf{U})$ は観測スペクトル \mathbf{Y} とモデルスペクトル \mathbf{X} の一致度を表す規準, $J(\mathbf{H})$ はスペクトルパーツが MFCC 特徴量空間で I 個の離散的なクラスタを形成しているかどうかを示した規準であり, λ はそれ

ぞれの規準をどれくらいの比率で重く見積もるかを表した定数である。

3. 補助関数法による最適化アルゴリズム

$L(H, U)$ を最小化する H, U は解析的に得ることはできない。そこで、当該最適化問題の局所最適解を探索する反復アルゴリズムを補助関数法に基づき導く。

補助関数法による、目的関数 $F(\Theta)$ の最小化問題の最適化アルゴリズムでは、まず補助変数 Λ を導入し、 $F(\Theta) = \min_{\Lambda} F^+(\Theta, \Lambda)$ を満たす補助関数 $F^+(\Theta, \Lambda)$ を設計する。そして、 $\Theta \leftarrow \operatorname{argmin}_{\Theta} F^+(\Theta, \Lambda)$ と $\Lambda \leftarrow \operatorname{argmin}_{\Lambda} F^+(\Theta, \Lambda)$ を交互に繰り返し、補助関数 $F^+(\Theta, \Lambda)$ の最小化を行うことで、間接的に元の目的関数 $F(\Theta)$ を最小化する。ここで重要なのは、 Θ について最小化しやすい $F^+(\Theta, \Lambda)$ を設計することである。

3.1 補助関数の設計

以下では $J(H)$ に対し上限関数を設計する。まず、

$$\sum_{n=1}^N \alpha_{m,n,k} = 1 \quad (7)$$

を満たす補助変数 $\alpha_{m,n,k}$ を用いて、

$$\left| \sum_n c_{m,n} \log G_{n,k} - \mu_{m,i} \right|^2 \leq \sum_n \frac{1}{\beta_{m,n,k}} (c_{m,n} \log G_{n,k} - \alpha_{m,n,k} \mu_{m,i})^2 \quad (8)$$

なる不等式を立てられる [6]。ただし、 $\beta_{m,n,k}$ は $\sum_{n=1}^N \beta_{m,n,k} = 1$, $\beta_{m,n,k} > 0$ を満たす任意の定数であり、 $G_{n,k} = \sum_{\omega} f_{n,\omega} H_{\omega,k}$ である。等号は、

$$\alpha_{m,n,k} \mu_{m,i} = c_{m,n} \log G_{n,k} + \beta_{m,n,k} \left(\mu_{m,i} - \sum_n c_{m,n} \log G_{n,k} \right) \quad (9)$$

のとき成立する。以上より、

$$\begin{aligned} J(H) &\leq \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{1}{\beta_{m,n,k}} (c_{m,n} \log G_{n,k} - \alpha_{m,n,k} \mu_{m,i})^2 \\ &= \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{1}{\beta_{m,n,k}} c_{m,n}^2 (\log G_{n,k})^2 \\ &\quad + \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{1}{\beta_{m,n,k}} (-2\alpha_{m,n,k} \mu_{m,i} c_{m,n} \log G_{n,k}) \\ &\quad + \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{1}{\beta_{m,n,k}} \alpha_{m,n,k}^2 \mu_{m,i}^2 \quad (10) \end{aligned}$$

$$\begin{aligned} &= \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \left(\sum_m \frac{c_{m,n}^2}{\beta_{m,n,k}} \right) (\log G_{n,k})^2 \\ &\quad + \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \left(\sum_m \frac{-2\alpha_{m,n,k} \mu_{m,i} c_{m,n}}{\beta_{m,n,k}} \right) \log G_{n,k} \\ &\quad + \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{\alpha_{m,n,k}^2 \mu_{m,i}^2}{\beta_{m,n,k}} \quad (11) \end{aligned}$$

が成り立つ。式 (11) 右辺の各項に対し、さらに上限となる関数を設計する。

まず、第一項に関して、以下の不等式を導入する。

補題 1. $x > 0, \xi > 0$ において

$$(\log x)^2 \leq \frac{1}{x} + p(\xi)x + q(\xi) \quad (12)$$

$$p(\xi) = \frac{2 \log \xi}{\xi} + \frac{1}{\xi^2} \quad (13)$$

$$q(\xi) = (\log \xi)^2 - 2 \log \xi - \frac{2}{\xi} \quad (14)$$

が成り立ち、等号は $x = \xi$ のときのみ成立する。

以下、表記の簡単化のため、

$$A_{n,k} := \sum_m \frac{c_{m,n}^2}{\beta_{m,n,k}} \quad (15)$$

と置く。 $A_{n,k}$ は非負なので、第一項の上界関数を設計するには $(\log G_{n,k})^2$ の上界関数を設計できれば十分である。以上の不等式を用いれば、任意の $\xi_{n,k} > 0$ について

$$(\log G_{n,k})^2 \leq \frac{1}{G_{n,k}} + p(\xi_{n,k})G_{n,k} + q(\xi_{n,k}) \quad (16)$$

が言える。さらに、逆数関数が凸関数であることを利用すれば、Jensen の不等式

$$\begin{aligned} \frac{1}{G_{n,k}} &= \frac{1}{\sum_{\omega} f_{n,\omega} H_{\omega,k}} = \frac{1}{\sum_{\omega} \rho_{n,\omega,k} \frac{f_{n,\omega} H_{\omega,k}}{\rho_{n,\omega,k}}} \\ &\leq \sum_{\omega} \rho_{n,\omega,k} \frac{\rho_{n,\omega,k}}{f_{n,\omega} H_{\omega,k}} = \sum_{\omega} \frac{\rho_{n,\omega,k}^2}{f_{n,\omega} H_{\omega,k}} \quad (17) \end{aligned}$$

が立てられる。ただし、 $\rho_{n,\omega,k}$ は

$$\sum_{\omega=1}^{\Omega} \rho_{n,\omega,k} = 1, \quad 0 < \rho_{n,\omega,k} < 1 \quad (18)$$

を満たす任意の変数である。これらをまとめると、

$$(\log G_{n,k})^2 \leq \sum_{\omega} \frac{\rho_{n,\omega,k}^2}{f_{n,\omega} H_{\omega,k}} + p(\xi_{n,k})G_{n,k} + q(\xi_{n,k}) \quad (19)$$

を得る。なお、この不等式の等号は、

$$\xi_{n,k} = G_{n,k} \quad (20)$$

$$\rho_{n,\omega,k} = \frac{f_{n,\omega} H_{\omega,k}}{\sum_{\omega'} f_{n,\omega'} H_{\omega',k}} \quad (21)$$

のとき成立する。

次に、式 (11) の第二項の上界関数を設計する。

$$B_{n,k} := \sum_m \frac{-2\alpha_{m,n,k} \mu_{m,i} c_{m,n}}{\beta_{m,n,k}} \quad (22)$$

と置くと、 $B_{n,k}$ は正と負の符号をとるるので、 $B_{n,k}$ の符号に応じて $\log G_{n,k}$ に対し別の不等式を立てる必要がある。

負の対数関数の上界は Jensen の不等式

$$\begin{aligned} -\log \left(\sum_{\omega} f_{n,\omega} H_{\omega,k} \right) &= -\log \left(\sum_{\omega} \nu_{n,\omega,k} \frac{f_{n,\omega} H_{\omega,k}}{\nu_{n,\omega,k}} \right) \\ &\leq -\sum_{\omega} \nu_{n,\omega,k} \log \frac{f_{n,\omega} H_{\omega,k}}{\nu_{n,\omega,k}} \end{aligned} \quad (23)$$

を, 正の対数関数の上限関数は

$$\log G_{n,k} \leq \frac{1}{\phi_{n,k}} (G_{n,k} - \phi_{n,k}) + \log \phi_{n,k} \quad (24)$$

$$= \frac{G_{n,k}}{\phi_{n,k}} + \log \phi_{n,k} - 1 \quad (25)$$

を用いてそれぞれ別々に設計することができる. これらの不等式は,

$$\nu_{n,\omega,k} = \frac{f_{n,\omega} H_{\omega,k}}{\sum_{\omega'} f_{n,\omega'} H_{\omega',k}} \quad (26)$$

$$\phi_{n,k} = G_{n,k} \quad (27)$$

のとき等号成立する.

以上をまとめると,

$$\begin{aligned} J(H) &\leq \sum_i \sum_{k \in \mathcal{C}_i} \sum_n A_{n,k} (\log G_{n,k})^2 \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_n B_{n,k} \log G_{n,k} \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{\alpha_{m,n,k}^2 \mu_{m,i}^2}{\beta_{m,n,k}} \quad (28) \\ &= \sum_i \sum_{k \in \mathcal{C}_i} \sum_n A_{n,k} (\log G_{n,k})^2 \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \delta_{B_{n,k} \geq 0} |B_{n,k}| \log G_{n,k} \\ &- \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \delta_{B_{n,k} < 0} |B_{n,k}| \log G_{n,k} \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{\alpha_{m,n,k}^2 \mu_{m,i}^2}{\beta_{m,n,k}} \quad (29) \\ &\leq \sum_i \sum_{k \in \mathcal{C}_i} \sum_n A_{n,k} \left(\sum_{\omega} \frac{\rho_{n,\omega,k}^2}{f_{n,\omega} H_{\omega,k}} + p(\xi_{n,k}) G_{n,k} + q(\xi_{n,k}) \right) \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \delta_{B_{n,k} \geq 0} |B_{n,k}| \left(\frac{G_{n,k}}{\phi_{n,k}} + \log \phi_{n,k} - 1 \right) \\ &- \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \delta_{B_{n,k} < 0} |B_{n,k}| \sum_{\omega} \nu_{n,\omega,k} \log \frac{f_{n,\omega} H_{\omega,k}}{\nu_{n,\omega,k}} \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{\alpha_{m,n,k}^2 \mu_{m,i}^2}{\beta_{m,n,k}} \quad (30) \end{aligned}$$

が言える. ただし,

$$\delta_{B_{n,k} \geq 0} = \begin{cases} 1 & (B_{n,k} \geq 0) \\ 0 & (B_{n,k} < 0) \end{cases} \quad (31)$$

$$\delta_{B_{n,k} < 0} = \begin{cases} 1 & (B_{n,k} < 0) \\ 0 & (B_{n,k} \geq 0) \end{cases} \quad (32)$$

である. 以上の上界関数

$$\begin{aligned} J^+(H, \alpha, \xi, \rho, \nu, \phi) &:= \sum_i \sum_{k \in \mathcal{C}_i} \sum_n A_{n,k} \left(\sum_{\omega} \frac{\rho_{n,\omega,k}^2}{f_{n,\omega} H_{\omega,k}} + p(\xi_{n,k}) G_{n,k} + q(\xi_{n,k}) \right) \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \delta_{B_{n,k} \geq 0} |B_{n,k}| \left(\frac{G_{n,k}}{\phi_{n,k}} + \log \phi_{n,k} - 1 \right) \\ &- \sum_i \sum_{k \in \mathcal{C}_i} \sum_n \delta_{B_{n,k} < 0} |B_{n,k}| \sum_{\omega} \nu_{n,\omega,k} \log \frac{f_{n,\omega} H_{\omega,k}}{\nu_{n,\omega,k}} \\ &+ \sum_i \sum_{k \in \mathcal{C}_i} \sum_m \sum_n \frac{\alpha_{m,n,k}^2 \mu_{m,i}^2}{\beta_{m,n,k}} \quad (33) \end{aligned}$$

は $J(H)$ の補助関数としての要件を満たす.

$I(H, U)$ の上限関数については, 通常の NMF と同様 Jensen の不等式を用い,

$$\begin{aligned} I(H, U) &\leq I^+(H, U) \\ &= \sum_{\omega} \sum_t \left(Y_{\omega,t} \log Y_{\omega,t} - Y_{\omega,t} \sum_k \zeta_{\omega,k,t} \log \frac{H_{\omega,k} U_{k,t}}{\zeta_{\omega,k,t}} \right. \\ &\quad \left. - Y_{\omega,t} + \sum_k H_{\omega,k} U_{k,t} \right) \quad (34) \end{aligned}$$

により得られる. 不等式の等号は,

$$\zeta_{\omega,k,t} = \frac{H_{\omega,k} U_{k,t}}{\sum_{k'} H_{\omega,k'} U_{k',t}} \quad (35)$$

のとき成立する.

以上より, $L^+(H, U) = I^+(H, U) + \lambda J^+(H)$ は $L(H, U)$ の補助関数の要件を満たす.

3.2 更新則の導出

以上で設計した補助関数に基づき, H と U の更新則を導く. まず, H の更新則は,

$$\begin{aligned} \frac{\partial L^+}{\partial H_{\omega,k}} &= -\frac{1}{H_{\omega,k}} \sum_t Y_{\omega,t} \zeta_{\omega,k,t} + \sum_t U_{k,t} \\ &+ \lambda \left[\sum_n A_{n,k} \left(-\frac{\rho_{n,\omega,k}^2}{f_{n,\omega} H_{\omega,k}^2} + p(\xi_{n,k}) f_{n,\omega} \right) \right. \\ &\quad \left. + \sum_n \delta_{B_{n,k} \geq 0} \frac{|B_{n,k}| f_{n,\omega}}{\phi_{n,k}} - \sum_n \delta_{B_{n,k} < 0} \frac{|B_{n,k}| \nu_{n,\omega,k}}{H_{\omega,k}} \right] \\ &= -\frac{1}{H_{\omega,k}} \sum_t Y_{\omega,t} \zeta_{\omega,k,t} + \sum_t U_{k,t} \\ &+ \lambda \left[-\frac{1}{H_{\omega,k}^2} \sum_n \frac{A_{n,k} \rho_{n,\omega,k}^2}{f_{n,\omega}} + \sum_n A_{n,k} p(\xi_{n,k}) f_{n,\omega} \right. \\ &\quad \left. + \sum_n \delta_{B_{n,k} \geq 0} \frac{|B_{n,k}| f_{n,\omega}}{\phi_{n,k}} - \frac{1}{H_{\omega,k}} \sum_n \delta_{B_{n,k} < 0} |B_{n,k}| \nu_{n,\omega,k} \right] \\ &= 0 \quad (36) \end{aligned}$$

の解, すなわち, 二次方程式

$$H_{\omega,k}^2 \left[\sum_t U_{k,t} + \lambda \left[\sum_n A_{n,k} p(\xi_{n,k}) f_{n,\omega} + \sum_n \delta_{B_{n,k} \geq 0} \frac{|B_{n,k}| f_{n,\omega}}{\phi_{n,k}} \right] \right] - H_{\omega,k} \left[\sum_t Y_{\omega,t} \zeta_{\omega,k,t} + \lambda \sum_n \delta_{B_{n,k} < 0} |B_{n,k}| \nu_{n,\omega,k} \right] - \lambda \sum_n \frac{A_{n,k} \rho_{n,\omega,k}^2}{f_{n,\omega}} = 0$$

の解となる. U の更新則は

$$\frac{\partial L^+}{\partial U_{k,t}} = -\frac{\sum_{\omega} Y_{\omega,t} \zeta_{\omega,k,t}}{U_{k,t}} + \sum_{\omega} H_{\omega,k} = 0 \quad (37)$$

の解となり, 通常の I-divergence 規準 NMF と同様である. 最適化アルゴリズムは以下のとおりにまとめられる.

(1) 補助変数の更新

$$\alpha_{m,n,k} \mu_{m,i} \leftarrow c_{m,n} \log G_{n,k} + \beta_{m,n,k} \left(\mu_{m,i} - \sum_n c_{m,n} \log G_{n,k} \right) \quad (38)$$

$$\xi_{n,k} \leftarrow G_{n,k} \quad (39)$$

$$\rho_{n,\omega,k} \leftarrow \frac{f_{n,\omega} H_{\omega,k}}{G_{n,k}} \quad (40)$$

$$\nu_{n,\omega,k} \leftarrow \frac{f_{n,\omega} H_{\omega,k}}{G_{n,k}} \quad (41)$$

$$\phi_{n,k} \leftarrow G_{n,k} \quad (42)$$

$$\zeta_{\omega,k,t} \leftarrow \frac{H_{\omega,k} U_{k,t}}{\sum_{k'} H_{\omega,k'} U_{k',t}} \quad (43)$$

(2) パラメータの更新

$$H_{\omega,k} \leftarrow \frac{-b_{\omega,k} + \sqrt{b_{\omega,k}^2 - 4a_{\omega,k} d_{\omega,k}}}{2d_{\omega,k}} \quad (44)$$

$$U_{k,t} \leftarrow \frac{\sum_{\omega} Y_{\omega,t} \zeta_{\omega,k,t}}{\sum_{\omega} H_{\omega,k}} \quad (45)$$

$$G_{n,k} \leftarrow \sum_{\omega} f_{n,\omega} H_{\omega,k} \quad (46)$$

ただし

$$a_{\omega,k} = -\lambda \sum_n \frac{A_{n,k} \rho_{n,\omega,k}^2}{f_{n,\omega}} \quad (47)$$

$$b_{\omega,k} = -\sum_t Y_{\omega,t} \zeta_{\omega,k,t} - \lambda \sum_n \delta_{B_{n,k} < 0} |B_{n,k}| \nu_{n,\omega,k} \quad (48)$$

$$d_{\omega,k} = \sum_t U_{k,t} + \lambda \left[\sum_n A_{n,k} p(\xi_{n,k}) f_{n,\omega} + \sum_n \delta_{B_{n,k} \geq 0} \frac{|B_{n,k}| f_{n,\omega}}{\phi_{n,k}} \right] \quad (49)$$

である. $\bar{\mu}_i$ と C_i の更新については, k-means アルゴリズムを用いれば良い. ここで H , U , $\bar{\mu}_i$, C_i の推定を行うことは, 音源分離と楽器ごとのクラスタリングを統合的に行うことを意味している.

4. 評価実験

提案手法の有効性を評価するため, 音源分離実験を行った. RWC 楽器音データベース [7] の楽器音を 2 つ混合して作成したモノラル音響信号から, 楽器音ごとの音源分離を行った. それぞれの楽器に対応する MFCC クラスタセントロイド $\bar{\mu}_i$ は, あらかじめクリーンな楽器音から学習して固定した. スペクトルパーツ数 K は 50 とした. メルフィルタバンクの数 N は 40 とし, MFCC 次元数 M は 13 とした. 補助関数における任意定数 $\beta_{m,n,k}$ は以下のように反復ごとに更新した.

$$\beta_{m,n,k} \leftarrow \frac{|c_{m,n} \log G_{n,k} - \alpha_{m,n,k} \mu_{m,i}|}{\sum_n |c_{m,n} \log G_{n,k} - \alpha_{m,n,k} \mu_{m,i}|} \quad (50)$$

この値は補助関数 J^+ を $\sum_n \beta_{m,n,k} = 1$, $\beta_{m,n,k} > 0$ の条件化で最小化する $\beta_{m,n,k}$ の値である [8]. H , U の初期値に関しては, まずはじめに I-divergence 規準の NMF を行い, 結果として得られた H と U を用いた. C_i は初期値の H を用いて計算した音色クラスタリング規準で定め, 固定した. 以上を 10 通りの楽器の組み合わせで行った. サンプリング周波数は 10 kHz とした. 短時間フーリエ変換におけるフレーム長は 64 ms, フレームシフト長は 32 ms とした. λ については最初の 5000 回の反復では前の反復結果を用いて計算した I/J として逐次更新を行い, それ以降は更新を行わず固定した (従ってアルゴリズムの収束性は保障される). 全体としてアルゴリズムの反復回数は 10000 回とした. 比較対象として, I-divergence 規準の NMF を行った後に, クラスタセントロイドとの MFCC 特徴量空間での二乗距離によってスペクトルパーツを楽器ごとにクラスタする方法 (以下では多段処理手法と記す) を用いた. 客観評価基準には, 以下の式で表される SNR を用いた.

$$SNR = 10 \log 10 \frac{\sum_i \sum_{\omega} \sum_t s_{i,\omega,t}^2}{\sum_i \sum_{\omega} \sum_t (s_{i,\omega,t} - \hat{s}_{i,\omega,t})^2} \quad (51)$$

ただし $s_{i,\omega,t}$ は i 番目の楽器に対応する元の音源信号, $\hat{s}_{i,\omega,t}$ は

$$\hat{s}_{i,\omega,t} = Y_{\omega,t} \frac{\sum_{k \in C_i} H_{\omega,k} U_{k,t}}{\sum_k H_{\omega,k} U_{k,t}} \quad (52)$$

によって得られる, i 番目の楽器に対応する推定した音源信号である. 高い SNR は高い分離性能を示す.

図 2 に提案法と多段処理法によって得られた分離音の SNR を示す. 10 種類中 9 種類の混合音において, 提案法によって得られた分離音の SNR は多段処理法によるそれを上回った.

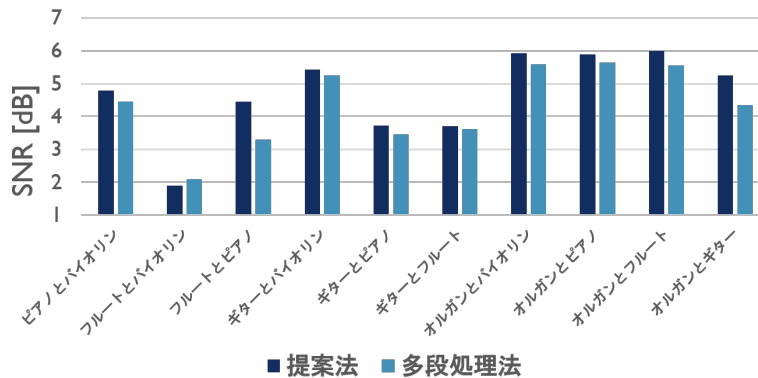


図 2 提案法と多段処理法によって得られた分離音の SNR

5. おわりに

本稿では、NMFによるフィッティング規準とMFCC特徴量空間でのクラスタリング規準を組み合わせた最適化規準を最適化するアルゴリズムを導出し、NMFのオーバーフィッティングを防ぐと共に、音源分離と楽器音認識を同時に実現する新しいアプローチを提案した。楽器ごとの音源分離実験では、NMFを行った後にスペクトルパーツをクラスタリングする手法と比べて、10種類中9種類の混合音において高いSNRを示した。

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] P. Smaragdis, and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," *Proc. WASPAA 2003*, Oct. 2003, pp. 177–180.
- [3] T. Virtanen, "Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criteria," *IEEE Trans. ASLP*, vol. 15, no. 3, pp. 1066–1074, 2007.
- [4] Hirokazu Kameoka, Kunio Kashino, "Composite Autoregressive System for Sparse Source-Filter Representation of Speech," in *Proc. 2009 IEEE International Symposium on Circuits and Systems (ISCAS2009)*, pp. 2477–2480, May 2009.
- [5] 吉井和佳, 後藤真孝, "多重音基本周波数解析のための無限複合自己回帰モデル," *情報処理学会研究報告*, 2012–MUS–96–8, Aug. 2012.
- [6] 亀岡弘和, 小野順貴, 柏野邦夫, 嵯峨山茂樹, "複素 NMF: 新しいスパース信号分解表現と基底系学習アルゴリズム," *日本音響学会 2008 年秋季研究発表会講演論文集*, 2–8–13, pp. 657–660, Sep. 2008.
- [7] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, "RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース," *情報処理学会 音楽情報科学研究会 研究報告 2002–MUS–45–4*, Vol.2002, No.40, pp.19–26, May 2002.
- [8] 亀岡弘和, ルルー・ジョナトン, "Frobenius ノルム規準の非負値行列因子分解における乗法更新式に関する一考察," *日本音響学会 2009 年秋季研究発表会講演論文集*, 1–P–8, pp. 709–712, Sep. 2009.

付 録

A.1 補題 1 の証明

$f(x) = \frac{a}{x} + bx + c$, $g(x) = (\log x)^2$ とし, すべての $x > 0$ で $f(x)$ が $g(x)$ を下回らず, かつ, $f(x)$ と $g(x)$ が 1 点で接する条件を導く. まず, $a = 1$ のとき,

$$f''(x) - g''(x) = \frac{2}{x^3}(1 - x + x \log x) \geq 0, \forall x > 0 \quad (\text{A.1})$$

より, $h(x) = f(x) - g(x)$ は凸関数であることが示される. ここで, $h(x)$ が $x = \xi$ で最小値 $h(\xi) = 0$ をとるには, $f(\xi) = g(\xi)$ かつ $f'(\xi) = g'(\xi)$ を満たす必要がある. これらを連立すると

$$\frac{1}{\xi} + b\xi + c = (\log \xi)^2 \quad (\text{A.2})$$

$$-\frac{1}{\xi^2} + b = \frac{2 \log \xi}{\xi} \quad (\text{A.3})$$

となり,

$$\begin{bmatrix} \xi & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} (\log \xi)^2 - \frac{1}{\xi} \\ \frac{2 \log \xi}{\xi} + \frac{1}{\xi^2} \end{bmatrix} \quad (\text{A.4})$$

$$\begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & -\xi \end{bmatrix} \begin{bmatrix} (\log \xi)^2 - \frac{1}{\xi} \\ \frac{2 \log \xi}{\xi} + \frac{1}{\xi^2} \end{bmatrix} \quad (\text{A.5})$$

$$= \begin{bmatrix} \frac{2 \log \xi}{\xi} + \frac{1}{\xi^2} \\ (\log \xi)^2 - 2 \log \xi - \frac{2}{\xi} \end{bmatrix} \quad (\text{A.6})$$

が得られる.