

可変基底 NMF に基づく音楽音響信号の解析

中野 允裕^{†1} 北野 佑^{†1} ルルー ジョナトン^{†2}
亀岡 弘和^{†2} 小野 順貴^{†1} 嵯峨山 茂樹^{†1}

本報告では、音楽音響信号から時間に伴って変化する楽器音スペクトルを教師なしに学習することを目的として、新しいスパース表現の枠組みを提案する。スペクトログラムが限られた数の基底スペクトルの重ね合わせにより構成されており、さらに各スペクトルが時間に伴って限られた数の状態を遷移していくと仮定することでスペクトログラムをモデル化し、そのパラメータ最適化アルゴリズムの導出を行う。

Polyphonic music signal analysis using non-negative matrix factorization with deformable bases

MASAHIRO NAKANO,^{†1} YU KITANO,^{†1}
JONATHAN LE ROUX,^{†2} HIROKAZU KAMEOKA,^{†2}
NOBUTAKA ONO^{†1} and SHIGEKI SAGAYAMA^{†1}

This paper presents a new sparse representation for polyphonic music signals. The goal is to learn the time-varying spectral patterns of musical instruments without any prior information. We model the spectrogram of music signals under the assumption that they are composed of a limited number of spectral patterns which have a limited number of states, and we develop an iterative algorithm for the estimation of the model parameters.

^{†1} 東京大学情報理工学系研究科

Graduate School of Information Science and Technology, The University of Tokyo

^{†2} 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

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

1. はじめに

音楽は、多数のメロディーの音響信号が多層に重畳した複雑な音響信号を形成するが、一方で限られたパーツが階層的に組み合わせられることによって構成されていると見なすことも出来る。例えば、多くの楽曲では限られたセクションが繰り返し出現する。また各セクションは限られたメロディーやフレーズによって構成されている。さらに各メロディーは限られた楽器の限られた音高によって演奏される。このように音楽が階層的な限られた数のパーツによって構成されている、という観点から音楽音響信号をモデル化する立場として階層的スパース表現という概念が提唱されている¹⁾。

音楽音響信号に関するスパース表現の代表的な手法として NMF (Nonnegative Matrix Factorization) があげられる²⁾。音響信号に対してよく用いられるのは、スペクトログラムを行列とみなし、スペクトルが限られた数の非負値スペクトルパターンを重ね合わせで表現出来るという仮定の下で NMF を適用しパーツ分解する方法である。NMF は近年、音響信号の強力な解析法として成功を収めており、様々な研究や拡張がなされている。

しかし、同時に NMF は次元削減のための手法という側面も持っており、音楽音響信号が本来持っている多様な姿を表現しにくい場合も少なくない。多くの楽器は時間に伴って変化する非定常なスペクトルを持つことによって音楽に彩りを加えている。例えば、ピアノのような打鍵楽器は、音の立ち上がりにアタックと言われる広帯域なエネルギーを持つ状態からサステイン、ディケイ、リリースといったいくつかの状態を経て消音する。また歌声や楽器のピブラートでは、一音の中でも基本周波数が変化する。NMF によって、このような楽器音の詳細なスペクトルの変化まで表現するためには、単に NMF の分解能を上げるだけでは解決できない問題がある。一つは、NMF の分解能を上げることはモデルの自由度を上げることになり、期待した解を得にくくなることにある。NMF は現在、初期値依存な反復アルゴリズムにより解かれることが一般的であり、モデルの自由度を上げることは意図しない局所解に陥る危険性を高めることに繋がる。もう一点は、分解されたスペクトル同士の関連付けが別の問題として生じることにある。ピアノの立ち上がりと減衰がそれぞれ別のスペクトルパターンとして学習された場合、これらをピアノの 1 音として関連付けるには別の処理が必要となる。

こういった問題意識から、音楽音響信号をより忠実に表現するために、楽器単音のスペクトルが限られた数のスペクトルパターンの状態遷移によって構成されるという階層構造を生かすことを考え、「可変基底 NMF」という新しい階層的スパース表現の枠組みを提案する。

2. 提案法

2.1 可変基底 NMF

本報告で提案するモデルは NMF によるスペクトログラムの分解表現を拡張したものである。NMF によるスペクトログラムの分解表現は、観測スペクトログラムを非負値行列 $\mathbf{Y} = (Y_{\omega,t})_{\Omega \times T} \in \mathbb{R}^{\geq 0, \Omega \times T}$ と見立て

$$Y_{\omega,t} \simeq \sum H_{\omega,d} U_{d,t} \quad (1)$$

となるような基底 $\mathbf{H} = (H_{\omega,d})_{\Omega \times D} \in \mathbb{R}^{\geq 0, \Omega \times D}$ とアクティベーション $\mathbf{U} = (U_{d,t})_{D \times T} \in \mathbb{R}^{\geq 0, D \times T}$ の 2 つの非負値行列を決めることによって得られる。 ω, t はそれぞれ周波数と時刻に対応するインデックスであり、観測スペクトルが D 個のスペクトル (基底) とアクティベーションの積によって表されるというモデルになっている。提案手法は、このスペクトルとアクティベーションのペアが楽器音 1 音高となることを狙っている。実際の楽器音は、時間に伴ってスペクトルが変化することを考慮すると、スペクトルが各時刻 t において一つの状態であると見なし、時刻 t における d 番目の基底の状態を $H_{d,t}^{(\phi_{d,t})}$ で表すことによって、

$$Y_{\omega,t} \simeq \sum_d H_{\omega,d}^{(\phi_{d,t})} U_{d,t} \quad (2)$$

と拡張する。ここで、基底の状態も限られた Φ 個しか存在せず、時間に伴ってそれらを遷移するものと仮定し、 $\mathbf{H} = (H_{\omega,\phi,d})_{\Omega \times \Phi \times D} \in \mathbb{R}^{\geq 0, \Omega \times \Phi \times D}$ を用いてモデルを表現し直す。各 $d = 1, \dots, D$ に対する Φ 個の基底の状態 $\mathbf{h}_\phi^{(d)} = (H_{1,\phi,d}, \dots, H_{\Omega,\phi,d})^T (\phi = 1, \dots, \Phi)$ に関して、許される状態遷移を隣接行列 $\mathbf{A}^{(d)} = (A_{\phi,\phi}^{(d)})_{\Phi \times \Phi}$ を用いて表す。状態 i から状態 j への遷移が許される場合 $A_{i,j}^{(d)} = 1$ とし、遷移が許されない場合は $A_{i,j}^{(d)} = 0$ とする。例えば、

$$\mathbf{A}^{(d)} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad (3)$$

とすると、状態のインデックス $1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1$ の遷移のみが許されることを表している。ここで、 $\mathbf{P} = (P_{\phi,t,d})_{\Phi \times T \times D} \in \mathbb{R}^{\geq 0, \Phi \times T \times D}$ を考え、各 $d = 1, \dots, D$ に対して各時刻 t に基底の状態が $\mathbf{h}_\phi^{(d)}$ のときに $P_{\phi,t,d} = 1$ とし、そうでないときに $P_{\phi,t,d} = 0$ とすることで、提案モデルは

$$Y_{\omega,t} \simeq \sum_{d,\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t} \quad (4)$$

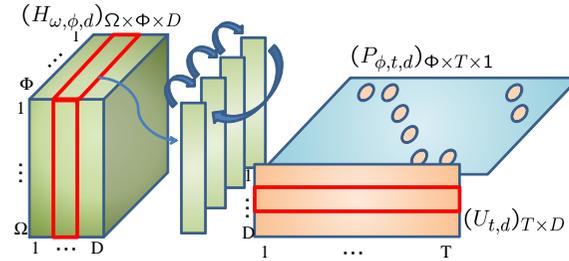


図1 可変基底 NMF の概念図

と表現することができる。今、各 $d = 1, \dots, D$ に対する $\mathbf{A}^{(d)}$ を固定したときにとりうる \mathbf{P} の集合を \mathcal{G} とする。解くべき問題は観測された \mathbf{Y} から \mathbf{H}, \mathbf{U} および $\mathbf{P} \in \mathcal{G}$ を推定することに帰着される。

2.2 最適化アルゴリズム

NMF は一般に観測とモデルの間の何らかの距離尺度を目的関数として最小化する制約付き最適化問題として解かれる。よく用いられる距離尺度としては Euclidean (EUC) distance, Kullback-Leibler (KL) divergence^{2),3)} や Itakura-Saito (IS) divergence⁴⁾ があり、これらはより一般化された β -divergence⁵⁾ の一部として密接な関わりを持っている。

NMF において目的関数を解析的に最適化することは困難であり、主に反復計算によりパラメータを更新していく方法が用いられる。よく知られたアルゴリズムには乗法更新アルゴリズムがあり、EUC distance と KL divergence に対する設計³⁾ から、 β -divergence への適用もされており⁶⁾、最近ではアルゴリズムの安定性に関する研究も行われている⁷⁾。乗法更新アルゴリズム以外にも様々なアルゴリズムが検討されている⁸⁾。

本報告では目的関数を β -divergence とした場合の最適化アルゴリズムを検討する。 β -divergence は様々な定義により用いられる⁹⁾ が、本報告では

$$\mathcal{D}_\beta(y|x) = \begin{cases} \frac{1}{\beta(\beta-1)} (y^\beta + (\beta-1)x^\beta - \beta y x^{\beta-1}) & \beta \in \mathbb{R} \setminus \{0, 1\} \\ y(\log y - \log x) + (x - y) & \beta = 1 \\ \frac{y}{x} - \log \frac{y}{x} - 1 & \beta = 0 \end{cases} \quad (5)$$

という定義を用いる。このとき解くべきは、観測された \mathbf{Y} から

$$\begin{aligned} &\text{minimize} \quad \mathcal{J}(\theta) = \sum_{\omega,t} \mathcal{D}_\beta \left(Y_{\omega,t} \mid \sum_{d,\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t} \right) \\ &\text{subject to} \quad \forall_{\omega,\phi,t} H_{\omega,\phi,t} \geq 0, \forall_{d,t} U_{d,t} \geq 0, \mathbf{P} \in \mathcal{G} \end{aligned} \quad (6)$$

となる $\theta = \{\mathbf{H}, \mathbf{U}, \mathbf{P}\}$ を求める問題に帰着される。この問題を解くにあたって、本報告では次に示す補助関数法を用いる。目的関数 $\mathcal{J}(\theta)$ に対して、

$$\mathcal{J}(\theta) = \min_{\hat{\theta}} \mathcal{J}^{(+)}(\theta, \hat{\theta}) \quad (7)$$

が成り立つ時、 $\mathcal{J}^{(+)}(\theta, \hat{\theta})$ を補助関数、 $\hat{\theta}$ を補助変数と定義する。

定理 1. 次式 (8) の更新を繰り返すことによって目的関数を単調収束させることが出来る。

$$\theta \leftarrow \operatorname{argmin}_{\theta} \mathcal{J}^{(+)}(\theta, \hat{\theta}), \quad \hat{\theta} \leftarrow \operatorname{argmin}_{\hat{\theta}} \mathcal{J}^{(+)}(\theta, \hat{\theta}) \quad (8)$$

証明 1. 第 s ステップ目の $\theta, \hat{\theta}$ をそれぞれ $^{(s)}$ を付けて表すと、

$$\mathcal{J}(\theta^{(s)}) = \mathcal{J}^{(+)}(\theta^{(s)}, \hat{\theta}^{(s+1)}) \geq \mathcal{J}^{(+)}(\theta^{(s+1)}, \hat{\theta}^{(s+1)}) \geq \mathcal{J}(\theta^{(s+1)}) \quad (9)$$

であることから $\mathcal{J}(\theta)$ が下界であるときには目的関数は単調収束する。□

式 (5) において x を変数と見なすと β に応じて各項の凸性や二階導関数の符号が変化することから分かる通り、目的関数は β によって異なる性質を示す。そこで、ここからは各 β ごとに補助関数の設計と最適化アルゴリズムを議論する。

2.2.1 $\beta = 2$: EUC distance

今、 $\sum_d \mu_{\omega,t,d} = 1, \sum_d \hat{Y}_{\omega,t,d} = Y_{\omega,t}$ のもとで $\hat{\theta} = (\hat{Y}_{\omega,t,d})_{\Omega \times T \times D}$ として補助関数 $\mathcal{J}^{(+)}(\theta, \hat{\theta})$ は

$$\mathcal{J}(\theta) \leq \mathcal{J}^{(+)}(\theta, \hat{\theta}) = \frac{1}{2} \sum_{\omega,t,d,\phi} \frac{1}{\mu_{\omega,t,d}} (P_{\phi,t,d} \hat{Y}_{\omega,t,d} - H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t})^2 \quad (10)$$

と設計できる。等号成立条件は

$$\hat{Y}_{\omega,t,d} = \sum_{\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t} + \mu_{\omega,t,d} \left(Y_{\omega,t} - \sum_{d,\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t} \right) \quad (11)$$

である。 $\mu_{\omega,t,d}$ の与え方には自由が許されており、より効率的なアルゴリズムを設計するための検討が行われている¹⁰⁾。ここでは、

$$\mu_{\omega,t,d} = \frac{\sum_{\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t}}{\sum_{d,\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t}} \quad (12)$$

と与える。補助関数が設計できたことによって以下のように更新アルゴリズムを導出することが出来る。更新の第 s ステップにおける各パラメータの値を $^{(s)}$ を付けて表す。

まず \mathbf{P} についてだが、

$$\hat{P}_{\phi,t,d} = \operatorname{argmin}_{\mathbf{P} \in \mathcal{G}} \left\{ \frac{1}{2} \sum_{\omega,t,d,\phi} \frac{1}{\mu_{\omega,t,d}^{(s)}} \left(P_{\phi,t,d} \hat{Y}_{\omega,t,d}^{(s)} - H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t}^{(s)} \right)^2 \right\} \quad (13)$$

は第 s ステップにおいて補助関数 $\mathcal{J}^{(+)}(\theta, \hat{\theta})$ を最小とする \mathbf{P} である。これは動的計画法により効率的に求めることが出来る。しかし、評価実験において検証する通り、 \mathbf{P} の毎反復における最適化は必要以上にパラメータの収束を早めることによって期待していない局所解に

陥りやすくなってしまふ。そこで次のようなスケジューリングを行うことによって、より良い解を得るための誘導を行う。 $\mathbf{k} = (k_1, k_2, \dots, k_{\Phi})$ を $\forall_{\phi} k_{\phi} \geq 0, \sum_{\phi} k_{\phi} = 1$ かつ、

$$\begin{cases} k_1^{(s+1)} \geq k_1^{(s)}, k_1^{(s)} = 1 \\ k_{\phi}^{(s+1)} \leq k_{\phi}^{(s)}, k_{\phi}^{(s)} = 0 \quad (\phi = 2, \dots, \Phi) \end{cases} \quad (14)$$

を満たすものとして与える。なお S は反復回数とする。 $\forall_n \hat{P}_{-(n-1),t,d} = \hat{P}_{\Phi-(n-1),t,d}$ として、

$$P_{\phi,t,d}^{(s+1)} = \sum_{n=1}^{\Phi} k_n^{(s+1)} \hat{P}_{\phi-(n-1),t,d} \quad (15)$$

と更新する。このような更新によって補助関数は各ステップ最小化はされなくなるが、 $\forall_s k_2^{(s)} = \dots = k_{\Phi}^{(s)}$ と与えることによって紙面の都合上詳細は省略するが、毎ステップ単調に減少させることが出来る。したがって、目的関数を単調収束させることの出来るアルゴリズムを設計することが可能である。 H についての更新式は、 $\partial \mathcal{J}^{(+)}(\theta, \hat{\theta}) / \partial H_{\omega,\phi,d} = 0$ を解くことにより、

$$H_{\omega,\phi,d}^{(s+1)} = \frac{\sum_t \hat{Y}_{\omega,t,d}^{(s)} (P_{\phi,t,d}^{(s+1)})^2 U_{d,t}^{(s)} / \mu_{\omega,t,d}^{(s)}}{\sum_t (P_{\phi,t,d}^{(s+1)})^2 U_{d,t}^{(s)} / \mu_{\omega,t,d}^{(s)}} \quad (16)$$

と得られる。同様に U についても $\partial \mathcal{J}^{(+)}(\theta, \hat{\theta}) / \partial U_{d,t} = 0$ を解き、

$$U_{d,t}^{(s+1)} = \frac{\sum_{\omega,\phi} H_{\omega,\phi,d}^{(s+1)} \hat{Y}_{\omega,t,d}^{(s)} (P_{\phi,t,d}^{(s+1)})^2 / \mu_{\omega,t,d}^{(s)}}{\sum_{\omega,\phi} (H_{\omega,\phi,d}^{(s+1)} P_{\phi,t,d}^{(s+1)})^2 / \mu_{\omega,t,d}^{(s)}} \quad (17)$$

を得ることができる。

2.2.2 $\beta = 1$: KL divergence 規準

今、 $\forall_{\omega,t,d} \eta_{\omega,t,d} \geq 0, \sum_d \eta_{\omega,t,d} = 1$ のもとで補助変数を $\hat{\theta} = (\eta_{\omega,t,d})_{\Omega \times T \times D}$ として補助関数 $\mathcal{J}^{(+)}(\theta, \hat{\theta})$ は

$$\mathcal{J}^{(+)}(\theta, \hat{\theta}) = \sum_{\omega,t,d,\phi} \left(\eta_{\omega,t,d} P_{\phi,t,d} Y_{\omega,t} \log \frac{\eta_{\omega,t,d} Y_{\omega,t}}{H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t}} - \frac{Y_{\omega,t}}{D\Phi} + H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t} \right) \quad (18)$$

と設計できる。 $\mathcal{J}(\theta) \leq \mathcal{J}^{(+)}(\theta, \hat{\theta})$ の等号成立条件は

$$\eta_{\omega,t,d} = \frac{\sum_{\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t}}{\sum_{d,\phi} H_{\omega,\phi,d} P_{\phi,t,d} U_{d,t}} \quad (19)$$

である。これより、先ほどと同様の要領で第 s ステップでの更新式は、

$$P_{\phi,t,d} \leftarrow \sum_{n=1}^{\Phi} k_n^{(s)} \hat{P}_{\phi-(n-1),t,d} \quad \left(\text{ただし、} \hat{P}_{\phi,t,d} = \operatorname{argmin}_{\mathbf{P} \in \mathcal{G}} \{ \mathcal{J}^{(+)}(\theta, \hat{\theta}) \} \right) \quad (20)$$

$$H_{\omega,\phi,d} \leftarrow \frac{\sum_t \eta_{\omega,t,d} P_{\phi,t,d} Y_{\omega,t}}{\sum_t P_{\phi,t,d} U_{d,t}}, \quad U_{d,t} \leftarrow \frac{\sum_{\omega,\phi} \eta_{\omega,t,d} P_{\phi,t,d} Y_{\omega,t}}{\sum_{\omega,\phi} H_{\omega,\phi,d} P_{\phi,t,d}} \quad (21)$$

と得られる。

2.2.3 $\beta = 0$: IS divergence 規準

今, $\forall \omega, t, d \eta_{\omega, t, d} \geq 0$, $\sum_d \eta_{\omega, t, d} = 1$, $\forall \omega, t Z_{\omega, t} \geq 0$ のもとで補助変数を $\hat{\theta} = \{(\eta_{\omega, t, d})_{\Omega \times T \times D}, (Z_{\omega, t})_{\Omega \times T}\}$ として補助関数 $\mathcal{J}^{(+)}(\theta, \hat{\theta})$ は

$$\mathcal{J}^{(+)}(\theta, \hat{\theta}) = \sum_{\omega, t, d, \phi} \left\{ \frac{P_{\phi, t, d}(\eta_{\omega, t, d})^2 Y_{\omega, t}}{H_{\omega, \phi, d} U_{d, t}} + \frac{H_{\omega, t, d} P_{\phi, t, d} U_{d, t}}{Z_{\omega, t}} + \frac{\log \frac{Z_{\omega, t}}{Y_{\omega, t}} - 2}{D} \right\} \quad (22)$$

と設計できる. $\mathcal{J}(\theta) \leq \mathcal{J}^{(+)}(\theta, \hat{\theta})$ の等号成立条件は

$$Z_{\omega, t} = \sum_{d, \phi} H_{\omega, t, d} P_{\phi, t, d} U_{d, t}, \quad \eta_{\omega, t, d} = \frac{\sum_{\phi} H_{\omega, t, d} P_{\phi, t, d} U_{d, t}}{\sum_{\phi, d} H_{\omega, t, d} P_{\phi, t, d} U_{d, t}} \quad (23)$$

である. 第 s ステップでの更新式は,

$$P_{\phi, t, d} \leftarrow \sum_{n=1}^{\phi} k_n^{(s)} \hat{P}_{\phi-(n-1), t, d} \quad \left(\text{ただし, } \hat{P}_{\phi, t, d} = \underset{\mathbf{P} \in \mathcal{S}}{\operatorname{argmin}} \{ \mathcal{J}^{(+)}(\theta, \hat{\theta}) \} \right) \quad (24)$$

$$H_{\omega, \phi, t} \leftarrow \sqrt{\frac{\sum_t P_{\phi, t, d}(\eta_{\omega, t, d})^2 Y_{\omega, t} / U_{d, t}}{\sum_t P_{\phi, t, d} U_{d, t} / Z_{\omega, t}}}, \quad U_{d, t} \leftarrow \sqrt{\frac{\sum_{\omega, \phi} P_{\phi, t, d}(\eta_{\omega, t, d})^2 Y_{\omega, t} / H_{\omega, t, d}}{\sum_{\omega, \phi} H_{\omega, t, d} P_{\phi, t, d} / Z_{\omega, t}}} \quad (25)$$

と得られる.

2.2.4 β -divergence 規準 ($\beta \neq 2, 1, 0$)

今, $\forall \omega, t, d \eta_{\omega, t, d} \geq 0$, $\sum_d \eta_{\omega, t, d} = 1$, $\forall \omega, t Z_{\omega, t} \geq 0$ のもとで補助変数を $\hat{\theta} = \{(\eta_{\omega, t, d})_{\Omega \times T \times D}, (Z_{\omega, t})_{\Omega \times T}\}$ として補助関数 $\mathcal{J}^{(+)}(\theta, \hat{\theta})$ は

$$\mathcal{J}^{(+)}(\theta) = \frac{1}{\beta(\beta-1)} \sum_{\omega, t, d, \phi} \begin{cases} \frac{Y_{\omega, t}^{\beta}}{D\Phi} + \eta_{\omega, t, d}(\beta-1) \left(\frac{X_{\omega, t, d, \phi}}{\eta_{\omega, t, d}} \right)^{\beta} + Y_{\omega, t} \mathcal{E}_{\omega, t, d, \phi}^{(2)}, & (\beta > 2, 0 < \beta < 1) \\ \frac{Y_{\omega, t}^{\beta}}{D\Phi} + \eta_{\omega, t, d}(\beta-1) \left(\frac{X_{\omega, t, d, \phi}}{\eta_{\omega, t, d}} \right)^{\beta} - \eta_{\omega, t, d} \beta Y_{\omega, t} \left(\frac{X_{\omega, t, d, \phi}}{\eta_{\omega, t, d}} \right)^{\beta-1}, & (1 < \beta < 2) \\ \frac{Y_{\omega, t}^{\beta}}{D\Phi} + \mathcal{E}_{\omega, t, d, \phi}^{(1)} - \eta_{\omega, t, d} (P_{\phi, t, d})^{2-\beta} \beta Y \left(\frac{X_{\omega, t, d, \phi}}{\eta_{\omega, t, d}} \right)^{\beta-1}, & (\beta < 0) \end{cases} \quad (26)$$

と設計することが出来る. ただし

$$X_{\omega, t, d, \phi} = H_{\omega, t, d} P_{\phi, t, d} U_{d, t} \quad (27)$$

$$\mathcal{E}_{\omega, t, d, \phi}^{(1)} = \beta(\beta-1) Z_{\omega, t}^{\beta-1} X_{\omega, t, d, \phi} - (\beta-1)^2 Z_{\omega, t}^{\beta} \quad (28)$$

$$\mathcal{E}_{\omega, t, d, \phi}^{(2)} = -\beta(\beta-1) Z_{\omega, t}^{\beta-2} X_{\omega, t, d, \phi} + (\beta-1)^2 Z_{\omega, t}^{\beta-1} \quad (29)$$

である. $\mathcal{J}(\theta) \leq \mathcal{J}^{(+)}(\theta, \hat{\theta})$ の等号成立条件は,

$$Z_{\omega, t} = \sum_d X_{\omega, t, d}, \quad \eta_{\omega, t, d} = \frac{X_{\omega, t, d, \phi}}{\sum_d X_{\omega, t, d, \phi}} \quad (30)$$

である. 各パラメータの更新式は同様に導出できるが紙面の都合上省略する.

3. 評価実験

提案法の有効性を検証するために, 3つのシミュレーション実験を行った. また, パラメータの更新アルゴリズムは反復回数 $S = 200$ とし, 第 s ステップの $k_{\phi}^{(s)}$ の値は,

$$k_{\phi}^{(s)} = \begin{cases} \max \left\{ \frac{1}{\Phi}, \frac{2\Phi-2}{S\Phi} s + \frac{2}{\Phi} - 1 \right\} & \phi = 1 \\ \min \left\{ \frac{1}{\Phi}, -\frac{2}{S\Phi} s + \frac{2}{\Phi} \right\} & \phi = 2 \dots \Phi \end{cases} \quad (31)$$

と与えた.

初めに提案手法の基礎的な動作確認として, 楽器単音からスペクトルの状態変化を学習する. 実験に用いた信号は, ピアノ C3 の 4 つの音長の入った音響信号を MIDI によって合成した. 振幅スペクトログラム Y は短時間フーリエ変換 (サンプリング周波数 16kHz, フレーム長 64ms, フレームシフト 32ms, Hanning 窓) により計算した. ピアノにおける基底の状態遷移は left-to-right に起こることを想定し, 隣接行列 $\mathbf{A}^{(d)}$ は,

$$\forall_d \mathbf{A}^{(d)} = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 1 & 0 & \dots & 0 & 1 \end{pmatrix} \quad (32)$$

とした. 基底数 $D = 1$, 状態数 $\Phi = 4$ としてパラメータの初期値を乱数により与え, 目的関数として KL divergence を用いた場合の学習結果を図 2 に示す. 学習結果から, 基底の状態遷移に着目すると, 音の立ち上がりから 2 状態は音長によらず同程度の長さを持ち, 減衰の長さが異なっている様子が確認できる. また基底として学習されたスペクトルパターンに着目すると, 各倍音ごとに異なる減衰をしていることも分かる.

二つ目の実験は, 提案法の音源分離としての性能評価として, 従来の NMF と提案法の SNR の比較を行った. 振幅スペクトログラム上で, 各音源の元信号 $I_{\omega, t}$ と推定した $\hat{I}_{\omega, t}$ の間の SNR (Signal-to-Noise Ratio) は,

$$\text{SNR} = 10 \log_{10} \frac{\sum_{\omega, t} I_{\omega, t}^2}{\sum_{\omega, t} (I_{\omega, t} - \hat{I}_{\omega, t})^2} \quad (33)$$

により計算した. 観測信号としてヴァイオリンによる Db, F, Ab の混合音を用いた. ピブ

ラートによる基底の状態遷移が left-to-right に起こると想定した場合と, ergodic に起こると想定して隣接行列を

$$\forall_d \mathbf{A}^{(d)} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ \vdots & & \ddots & & \vdots \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix} \quad (34)$$

とした場合をそれぞれ実験した. 基底数 $D = 3$, 各基底の状態数 $\Phi = 5$ と与えた. 比較対象とする NMF は基底数 3, 15 として分解を行った. 分離した各信号は元信号のどの音源に属するか分からないため, 元信号全てとの SNR を求め最も大きな値を持つ音源への割り当てを行った. もし割り当ての存在しない音源があった場合は, その音源は SNR の算出には用いず, 代わりに割り当ての存在しなかった音源数の割合をエラー率として計算した. 従来法, 提案法ともに, パラメータの初期値は乱数によって与え, 10 回の試行の SNR とエラー率を表 1 に示した. 目的関数として KL divergence を用いた場合の分離結果の一例を図 3 に示した. 基底の状態変化から, ヴァイオリンのピブラートが状態遷移として学習出来ていることが確認できる. SNR に関しては, IS divergence の場合を除いて, 基底数が 15 の時の従来法と同程度の性能を示している. 従来法では各音源はいくつかの信号に分解され, それらの統合には元信号を利用しているが, 提案法においては各音源ごとに分解された信号が階層的にまとまった形で得ることが出来る. このことから提案法は元信号の事前知識を必要としない音源分離手法としても有用であることが確認できた. しかし, 提案法の基底の状態遷移に関して, ergodic なモデルの方が left-to-right モデルよりも良い分離性能を示しているように, 事前に与える状態遷移の拘束によって性能が変わることも確認できた. 今後は状態遷移も同時に推定するような枠組みへ拡張する予定である.

最後の実験として, 状態遷移を表す \mathbf{P} の更新にスケジューリングを用いた効果を検証した. 観測信号は二つ目の実験と同じヴァイオリンの混合音を用いた. 目的関数として EUC distance を用い, 更新アルゴリズムは反復回数 $S = 100$ とし, スケジューリングを行った場合と行わなかった場合をそれぞれ 3 回ずつ試行した時の目的関数の様子を図 4 に示した. スケジューリングによって局所解が回避できていることが確認できる.

4. おわりに

本報告では時間に伴って変化する楽器音スペクトルを教師なしに学習することを目的として, スペクトルパターン (基底) の状態遷移の拘束を導入した可変基底 NMF という新しい

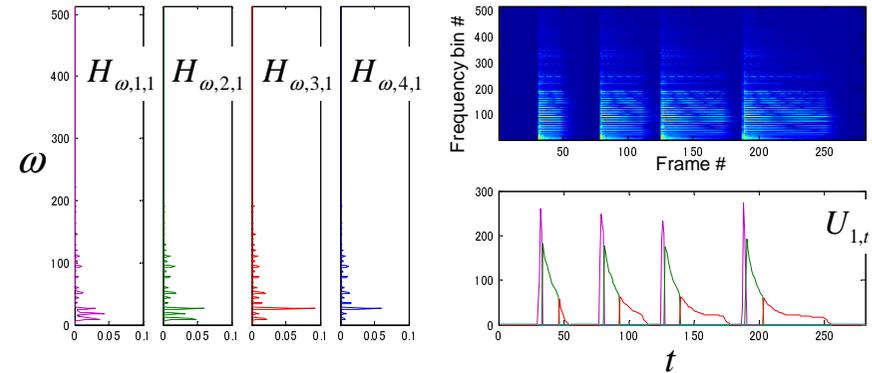


図 2 音長の異なる 4 音を演奏したピアノの観測スペクトログラム (右上) から学習した基底の 4 状態 (左) とそれぞれのアクティベーション (右下)

表 1 各アルゴリズムにおける音源分離性能の SNR(dB) とエラー率 (%)
Table 1 Source Separation Performance (Signal-to-Noise Ratio [dB], error rate [%])

algorithm	SNR(dB)				error rate(%)
	Source1	Source2	Source3	Mean	
EUC-NMF($D = 3$)	13.92	9.46	8.14	10.51	0
EUC-NMF($D = 15$)	12.32	12.98	10.69	11.98	0
EUC-proposed(left-to-right)	15.60	13.70	11.74	13.68	0
EUC-proposed(ergodic)	16.31	14.83	11.72	14.29	0
KL-NMF($D = 3$)	14.74	9.41	8.39	10.85	0
KL-NMF($D = 15$)	14.88	15.99	11.09	13.99	0
KL-proposed(left-to-right)	16.13	13.28	11.52	13.64	0
KL-proposed(ergodic)	16.64	14.64	12.00	14.62	0
IS-NMF($D = 3$)	1.12	2.05	1.66	1.61	0
IS-NMF($D = 15$)	10.77	11.19	10.57	10.84	0
IS-proposed(left-to-right)	1.60	6.11	1.81	3.17	0
IS-proposed(ergodic)	0.44	0.64	0.43	0.47	13

スパース表現の枠組みを提案した. 今後は実音楽に対する実験を行い, 音源分離や自動採譜への有用性を検証していくつもりである. また, 基底の状態遷移を隠れ変数として同時に推定できるような枠組みへの拡張や, 複素 NMF¹¹⁾ と同様に複素スペクトル領域のモデルへの拡張も検討していきたい.

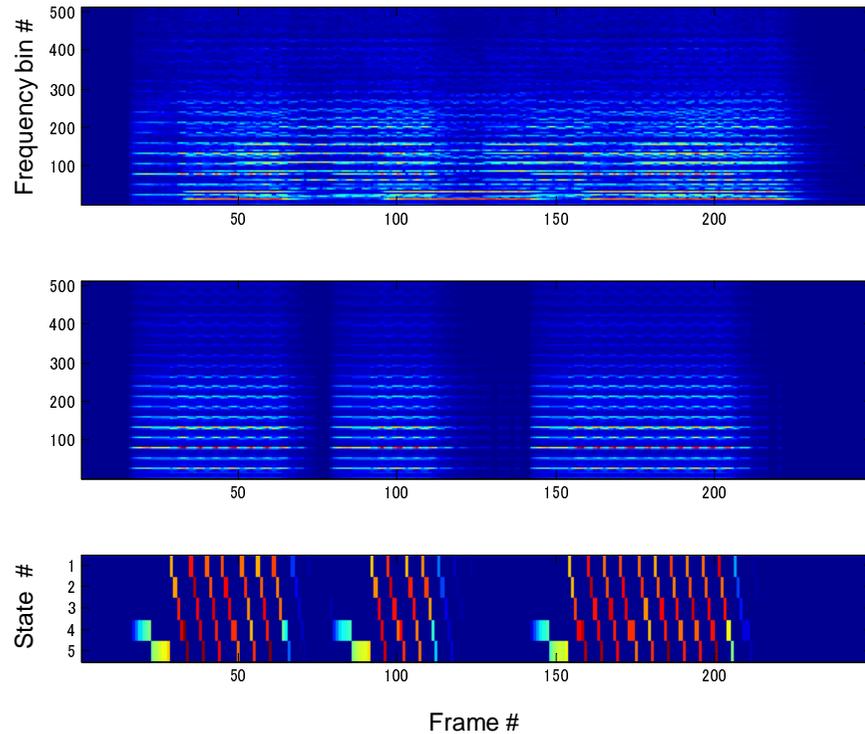


図 3 3 音源を混合した観測スペクトログラム (上段) から分離した 1 音源のスペクトログラム (中段) とその基底の状態遷移 (下段). ビブラートが 5 つの状態の遷移によって表現出来ていることが確認できる.

参 考 文 献

- 1) 亀岡, ルルー, 大石, 柏野, "Music Factorizer: 音楽音響信号をノート単位で編集できるインタフェース," 情報処理学会研究報告, 2009-MUS-81-9, 2009.
- 2) D. D. Lee and H. S. Seung, " Learning the parts of objects by non-negative matrix factorization, " *Nature*, vol. 401, pp. 788-791, Oct. 1999.
- 3) D. D. Lee and H. S. Seung, " Algorithms for non-negative matrix factorization, " in *Proc. of the Conference on Advances in Neural Information Processing Systems*,

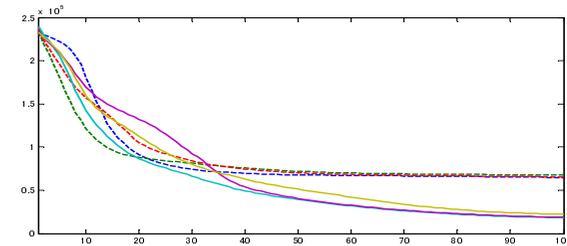


図 4 更新アルゴリズムにおける目的関数の変化. 横軸は反復回数, 縦軸は目的関数の値を表す. 状態遷移を表す P の更新にスケジューリングを行った場合 (実線) に, 行わなかった場合 (破線) よりも目的関数が減少していることが確認できる.

- vol. 13. Vancouver, British Columbia, Canada: MIT Press, Dec. 2001, pp. 556-562.
- 4) F. Itakura and S. Saito, " Analysis synthesis telephony based on the maximum likelihood method, " in *Proc. of 6th International Congress on Acoustics*, Tokyo, Japan, Aug. 1968, pp. C17-C20.
- 5) S. Eguchi and Y. Kano, " Robustifying maximum likelihood estimation, " Tokyo Institute of Statistical Mathematics, Tokyo, Japan, Tech. Rep., 2001. [Online].
- 6) A. Cichocki, R. Zdunek, and S. Amari, " Csiszárs divergences for non-negative matrix factorization : Family of new algorithms, " in *Proc. of 6th International Conference on Independent Component Analysis and Blind Signal Separation*, Charleston, SC, USA, Mar. 2006, pp. 32-39.
- 7) R. Badeau, N. Bertin, and E. Vincent, " On the stability of multiplicative update algorithms. Application to non-negative matrix factorization, " *Telecom ParisTech, Technical report.*, 2009.
- 8) A. Cichocki, R. Zdunek, and S. Amari, " Nonnegative matrix and tensor factorization, " *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 142-145, Jan. 2008.
- 9) 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, Mar. 2009.
- 10) 亀岡, ルルー, " 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, 2009.
- 11) H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A New Sparse Representation for Acoustic Signals," In *Proc. 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3437-3440, 2009.