

特別研究論文

(査読済み)

研 究 題 目

コサイン類似度罰則条件付き非負値行列因子分解に基づく音楽音源分離

提出年月日	2022 年	1 月	31 日
氏 名	岩瀬 佑太		
主 査	北村 大地 講師		
副 査	柿元 健 准教授		
副 査	雛元 洋一 助教		

香川高等専門学校
専攻科
創造工学専攻



Music Source Separation Based on Nonnegative Matrix Factorization with Cosine Similarity Penalty

Yuta Iwase

Advanced Course in Industrial and Systems Engineering
National Institute of Technology, Kagawa College

Abstract

In this thesis, I aim to improve performance of audio source separation for monaural mixture signals that consist of two or more sound sources. In audio source separation for monaural signals, nonnegative matrix factorization (NMF) has been utilized. In particular, semi-supervised NMF (SNMF) can achieve separation by employing few supervised signals such as scale sounds of each instrument. In recent years, penalized SNMF (PSNMF) with orthogonality penalty was proposed. This method forces two basis matrices for target and interference sources to be orthogonal each other. Since basis matrices have spectral patterns of each source and the similarities between these patterns are maximized by orthogonalization, the separation quality is greatly improved by PSNMF. However, in conventional PSNMF, the orthogonality penalty does not affect correctly because of the scale ambiguity between basis and activation matrices in NMF. In this thesis, I propose two new methods that utilize a cosine similarity penalty between the basis matrices of the target and interference sources. Since the cosine similarity penalty does not depend on the lengths of input vector, I can avoid the scale ambiguity problem in NMF decomposition. As the first method, PSNMF with a logarithmic cosine similarity penalty (Log-Cos PSNMF) is proposed. Although Log-Cos PSNMF can improve the separation performance, it encounters another problem of numerical instability because of the logarithm function. To solve this problem, as the second method, I propose PSNMF with a simple cosine similarity penalty that does not include the logarithm function. I compare audio source separation performance of two proposed PSNMF algorithms with conventional PSNMF. The experimental result shows the efficacy of both of the proposed methods. Also, two statistical tests are conducted to objectively conclude the differences between the results obtained by the conventional and proposed methods.

Key Words: audio source separation, nonnegative matrix factorization, orthogonality, cosine similarity

(和訳)

本論文では、2 つ以上の音源からなるモノラル混合信号の音源分離の向上を目指す。モノラル信号の音源分離では、非負値行列因子分解 (NMF) と呼ばれるアルゴリズムに基づく手法が有名である。混合信号の中から抽出したい楽器音信号の教師情報を NMF に活用する手法の 1 つに半教師あり NMF (SNMF) がある。特に、SNMF による音源分離精度を向上させる手法として内積罰則条件付き SNMF (内積型 PSNMF) が定式化されている。この手法は、内積罰則項により、非目的音源の基底行列に含まれる各基底ベクトルが、目的音源の基底行列に含まれるすべての基底ベクトルに対して直交となるように誘導する。基底行列には各音源のスペクトルパターンが含まれているため、互いの基底行列の非類似度を直交化により高めることで分離精度が大きく向上する。しかし、内積型 PSNMF では、基底行列とアクティベーション行列間のスケール不定性により、直交化の罰則条件が正しく作用しない定式化となっている。そこで本論文では、ベクトルの長さに依存せず直交化を促すような罰則項を付与する手法を提案する。この提案手法は、罰則項にコサイン類似度の対数和を与えてるため、対数コサイン類似度罰則条件付き SNMF (Log-Cos 型 PSNMF) と呼ぶ。Log-Cos 型 PSNMF の罰則項では、前述のスケール不定性の問題を回避することができる。しかし、罰則項に対数を用いたことにより、数値的に不安定になるという問題点が新たに生じた。そこで、2 つ目の提案手法として対数を用いない純粋なコサイン類似度を罰則項に与えるコサイン類似度罰則条件付き SNMF (Cos 型 PSNMF) を再定式化する。従来手法である内積型 PSNMF と 2 種類の提案手法の音源分離性能を比較し、2 種類の提案手法がいずれも従来手法より優れた音源分離性能を示すことを実験を通して示す。また実験より得られた結果の差を客観的に結論付けるために統計的検定を 2 種類行う。

目次

第 1 章	緒言	1
1.1	本論文の背景	1
1.2	本論文の目的	2
1.3	本論文の構成	3
第 2 章	NMF を用いた音源分離	5
2.1	まえがき	5
2.2	NMF	5
2.3	STFT	8
2.4	NMF の音源分離への応用	8
2.5	本章のまとめ	10
第 3 章	従来手法	11
3.1	まえがき	11
3.2	SNMF に基づく音楽音源分離	11
3.3	内積型 PSNMF	13
3.4	本章のまとめ	14
第 4 章	対数コサイン類似度罰則条件に基づく提案手法	16
4.1	まえがき	16
4.2	動機	16
4.3	コスト関数	17
4.4	反復更新式の導出	18
4.5	本章のまとめ	21
第 5 章	コサイン類似度罰則条件に基づく提案手法	22
5.1	まえがき	22
5.2	Log-Cos 型 PSNMF の問題点	22
5.3	コスト関数	23
5.4	反復更新式の導出	23
5.5	本章のまとめ	25

第 6 章	音楽信号の分離実験	26
6.1	まえがき	26
6.2	実験条件	26
6.3	実験結果	28
6.4	統計的検定	30
6.5	本章のまとめ	30
第 7 章	結言	32
	謝辞	33
	参考文献	33
付録 A	補助関数法で利用される不等式	37
A.1	接線不等式	37
A.2	Jensen の不等式	37

第 1 章

緒言

1.1 本論文の背景

音源分離とは、複数の音源が混合した状態で観測された信号から、混合する前の個々の音源を分離・抽出する技術である。すなわち、人混みや雑踏の中でも自分に関係のある会話や興味のあるキーワードを自然に聞きとることができるカクテルパーティ効果を、コンピュータを用いて実現することである。この技術は、音声認識へも応用されており、ノイズを抑制し音声の特徴のみを捉えるために音源分離が活用されている。また、音楽に対する音源分離は、Figs. 1.1 及び 1.2 に示すように、コンピュータを用いて音楽信号から楽譜を自動生成する自動採譜や、コンサートのような複数の楽器が同時に演奏されている音楽信号から特定の楽器のメロディのみを抽出する等の活用方法がある。本研究では、特に、モノラル音楽信号に基づく音源分離を対象とする。モノラル信号に対する音源分離技術では非負値行列因子分解 (nonnegative matrix factorization: NMF) [1, 2] と呼ばれる数値アルゴリズムを活用した手法が高精度である。

音源分離は、予め用意しておく手掛かりとなるような情報（これを教師情報と呼ぶ）を用いない「教師なし音源分離」と、教師情報を用いる「教師あり音源分離」の 2 つに大別できる。教師なし音源分離は、教師情報を必要としないことから、あらゆる音響信号への活用が可能である [3]。その反面、高い音源分離精度の実現は困難である。一方、教師あり音源分離では、混合信号中の各音源の種類や楽譜、声色のサンプル等の教師情報を事前に学習し、学習で得られる何らかのモデルを音源分離に積極的に活用する。

教師あり音源分離は、取り扱う教師情報の数によってさらに「全教師あり音源分離」と「半教師あり音源分離」に分けられる [4, 5]。全教師あり音源分離は、全ての音源を分離したい場合に用いられ、教師情報は全ての音源分必要となる。また半教師あり音源分離は、特定の音だけを全体の中から抽出したい場合に用いられ、抽出したい音（目的音）だけの教師情報が必要となる。よって半教師あり音源分離は、全教師あり音源分離よりも予め学習する教師情報の数が少なく済むため、学習にかかるコストが低いという利点がある。

教師情報が大量にある場合、深層ニューラルネットワーク (deep neural network: DNN) と呼ばれる機械学習モデルに基づく音源分離手法が近年盛んに研究され、特に音声の音源分離に

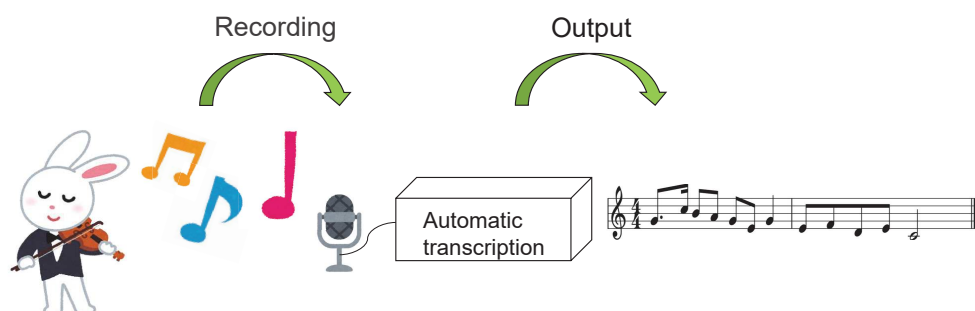


Fig. 1.1: Automatic music transcription.

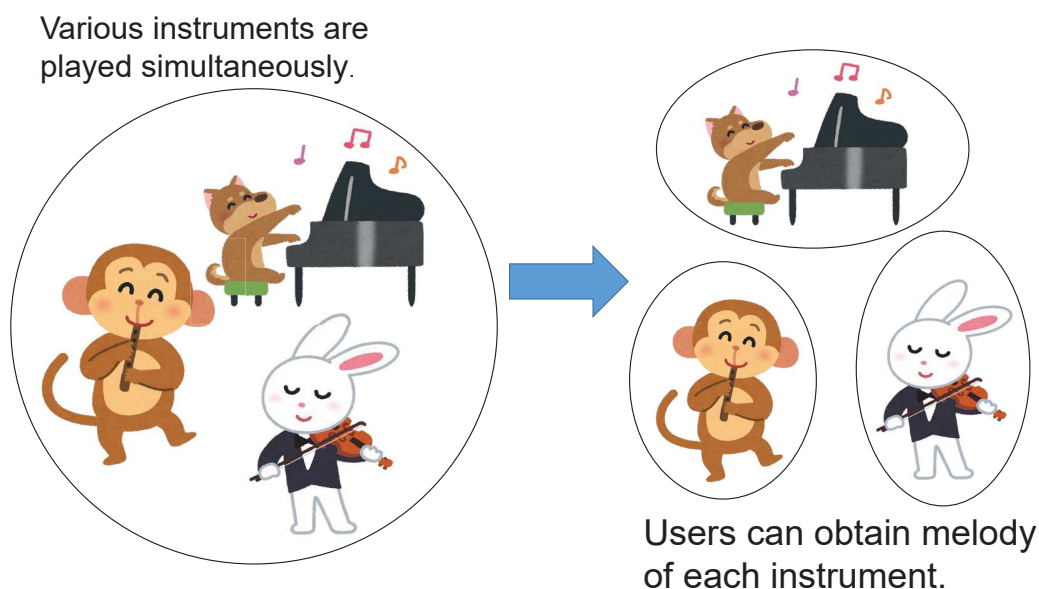


Fig. 1.2: Separation of instrumental sounds.

において成功をおさめている [6, 7]. しかし, そのような DNN を学習するためには, 各音源のソロの録音信号を数百時間分用意する必要がある等の実用上の問題がある. 教師なしの状況や, 分離対象となる音源 (目的音源) の学習用データセットが極僅かしか用意できないような状況においては, NMF に基づく手法がこれまで数多く提案されてきた [3, 4, 5, 8, 9, 10, 11]. 特に, 省サンプルの学習データから特徴量を学習しそのまま活用する教師あり NMF (supervised NMF: SNMF) [4, 5] は, 現実的な学習コストと音源分離性能を両立しており, DNN が盛んに研究される現在においても, 依然として強力な手法である.

1.2 本論文の目的

DNN に基づく音源分離手法は, 前述のとおり, 膨大な教師情報が必要であり, 非現実的である. 本論文では, 現実的な応用可能性の観点から, 目的音に対して少ない (省サンプルの)

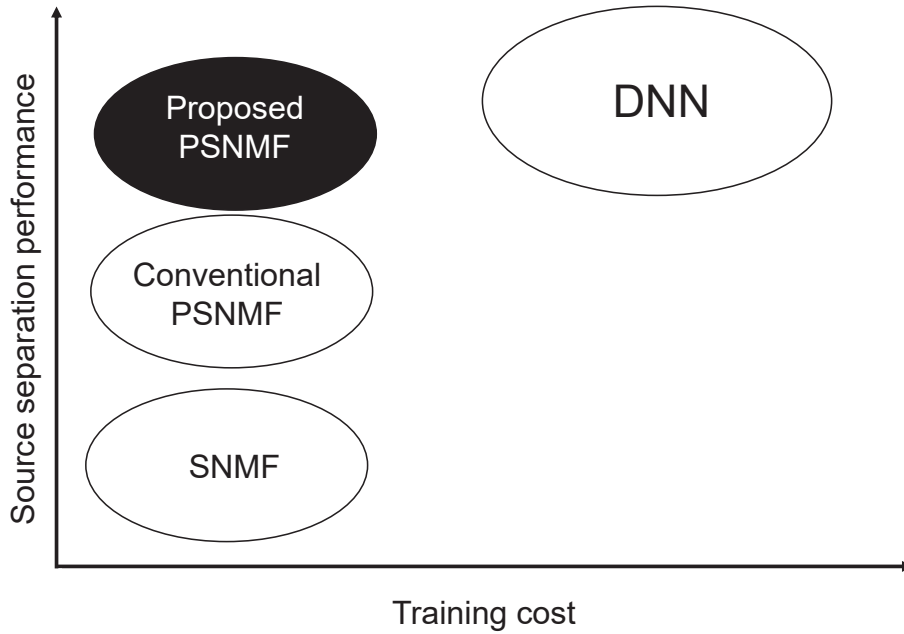


Fig. 1.3: Scope of this thesis.

教師情報が与えられる状況に限定し、この状況に対して今なお強力な手法である（半教師ありの）SNMF に焦点を当て、その改良方法について検討する。

SNMF の性能向上手法には、最適化のコスト関数に直交化罰則と呼ばれる正規化項を付与した内積罰則条件付き SNMF (penalized SNMF: PSNMF)、通称、内積型 PSNMF [5] と呼ばれる手法が提案されており、純粋な（罰則条件のない）SNMF より高い音源分離性能を示すことが実験を通して明らかになっている。しかし、内積型 PSNMF には、NMF が孕むスケール不定性により、罰則条件が最適化に正しく作用している保証がないという問題がある。本論文では、この問題を解決するとともに、更なる分離精度の改善のために、SNMF にコサイン類似度を罰則項として付与する手法を提案する。Fig. 1.3 では SNMF、従来手法である内積型 PSNMF、提案手法、及び DNN に基づく音源分離の位置付けを示す。本論文で提案する手法では、膨大な教師情報を学習データとして用いる DNN ほどの分離精度は見込めない可能性があるが、学習コストは従来手法同様低コストである。さらに、分離精度は従来手法より高くなることを実験で示す。

1.3 本論文の構成

本論文の構成は次の通りである。2 章では、NMF の基礎理論について述べ、音源分離への適用方法を示す。3 章では、SNMF による音源分離と従来手法の内積型 PSNMF の概要と反復更新式を示す。4 章及び 5 章では、本論文で新たに提案するコサイン類似度を用いた罰則条件付き SNMF を 2 種類説明し、それぞれについて定式化する。6 章では、2 種類の提案手法の有効性を確認するために、音楽信号の音楽分離性能を罰則項なしの SNMF、従来手法の内

4 第 1 章 緒言

積型 PSNMF 及び 2 種類の提案手法に関して比較し，その結果を述べる．さらに，比較実験より得られた結果を用いて統計的検定を行い，提案手法の統計的優位性を示す．7 章に本論文の総括と今後の展望を述べる．

第 2 章

NMF を用いた音源分離

2.1 まえがき

本章では、NMF に関する基礎理論と音楽信号を NMF に適用するための手法について説明する。まず、2.2 節では、NMF による行列分解の基礎理論を示す。2.3 節では、NMF に基づく音響信号処理において必須となる短時間フーリエ変換（short-time Fourier transform: STFT）について説明する。2.4 節では、NMF の音源分離への応用を示す。2.5 節で本章のまとめを示す。

2.2 NMF

NMF とは、非負の観測行列を、別の 2 つの非負行列の行列積で近似的に分解する数値アルゴリズムである。このとき、2 つの非負行列の行列積のランクは、観測行列のランクよりも小さくなるように制約しているため、非負制約条件付き低ランク近似とみなせる。この非負制約と低ランク近似の両性質により、観測行列に含まれる少数の有意な特徴量（頻出パターン等）を教師無し学習として抽出することができる。実世界では、音の振幅スペクトログラムや画像等の非負行列で表されるデータが多く存在する。NMF は、そのような非負値のデータを、加法的な非負構成成分に分解することを目的として適用され得られる非負特徴量（構成成分）の解釈の容易さから、商品の購買データパターン抽出 [12] や音源分離 [5] 等のパターン認識を活用する分野で有用である。

NMF による分解は次の式で表される。

$$\mathbf{X} \approx \mathbf{F}\mathbf{Q} \quad (2.1)$$

ここで $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J}$ は非負の観測行列である。また、 $\mathbf{F} \in \mathbb{R}_{\geq 0}^{I \times K}$ 及び $\mathbf{Q} \in \mathbb{R}_{\geq 0}^{K \times J}$ はそれぞれ基底行列及びアクティベーション行列と呼ばれる。 I 及び J はそれぞれ、観測行列の行数と列数である。また、 K は基底ベクトル数（基底行列 $\mathbf{F}$ の列数またはアクティベーション行列 $\mathbf{Q}$ の行数）である。このように、NMF は非負のデータ行列（観測行列） $\mathbf{X}$ を別の二つの非負行列の行列積 $\mathbf{F}\mathbf{Q}$ で近似するアルゴリズムである。基底行列 $\mathbf{F}$ には、観測行列 $\mathbf{X}$ 中の頻出パ

ターンが基底ベクトル（列ベクトル）として含まれ、それらの時間的強度変化がアクティベーション行列 $\mathbf{Q}$ の行ベクトルに含まれる。

式 (2.1) を要素ごとに表すと,

$$x_{ij} \approx \sum_k f_{ik} q_{kj} \quad (2.2)$$

となる。ここで, x_{ij} , f_{ik} , 及び q_{kj} はそれぞれ $\mathbf{X}$, $\mathbf{F}$, 及び $\mathbf{Q}$ の非負要素である。さらに, $i = 1, 2, \dots, I$, $j = 1, 2, \dots, J$, 及び $k = 1, 2, \dots, K$ はそれぞれ $\mathbf{X}$ の行数, $\mathbf{X}$ の列数, 及び基底ベクトル数のインデックスを表す。一般に基底ベクトル数 K は I や J よりも十分に小さい値に設定されるため, NMF は観測行列 $\mathbf{X}$ を K 個の基底ベクトルで表現する低ランク近似と解釈できる。

NMF において $\mathbf{F}$ 及び $\mathbf{Q}$ の推定は, 次式の最適化問題として定式化される。

$$\min_{\mathbf{F}, \mathbf{Q}} \mathcal{D}(\mathbf{X} \| \mathbf{F}\mathbf{Q}) \text{ s.t. } f_{ik}, q_{kj} \geq 0 \quad \forall i, j, k \quad (2.3)$$

ここで, $\mathcal{D}(\cdot \| \cdot)$ は二つの行列の類似度を測る任意の関数であり, NMF の文脈では二乗ユークリッド距離 [2], 一般化 Kulback–Leibler (KL) ダイバージェンス [13], Itakura–Saito (IS) ダイバージェンス [14] 等がよく用いられる。したがって, 式 (2.3) は, 非負観測行列 $\mathbf{X}$ を良く近似するような非負のモデル行列 $\mathbf{F}\mathbf{Q}$ (ただしモデル行列 $\mathbf{F}\mathbf{Q}$ のランクは K) を推定する問題となる。前述の 3 種類の類似度関数はそれぞれ次のように定義されている。

- 二乗ユークリッド距離

$$\mathcal{D}(\mathbf{X} \| \mathbf{F}\mathbf{Q}) = \sum_{i,j} \left(x_{ij} - \sum_k f_{ik} q_{kj} \right)^2 \quad (2.4)$$

- 一般化 KL ダイバージェンス

$$\mathcal{D}(\mathbf{X} \| \mathbf{F}\mathbf{Q}) = \sum_{i,j} \left(x_{ij} \log \frac{x_{ij}}{\sum_k f_{ik} q_{kj}} - x_{ij} + \sum_k f_{ik} q_{kj} \right) \quad (2.5)$$

- IS ダイバージェンス

$$\mathcal{D}(\mathbf{X} \| \mathbf{F}\mathbf{Q}) = \sum_{i,j} \left(\frac{x_{ij}}{\sum_k f_{ik} q_{kj}} - \log \frac{x_{ij}}{\sum_k f_{ik} q_{kj}} - 1 \right) \quad (2.6)$$

式 (2.3) の $\mathbf{F}$ と $\mathbf{Q}$ に関する解は, 上記の 3 種類のいずれの類似度関数を用いても, 閉形式では与えられない。そのため, $\mathbf{F}$ と $\mathbf{Q}$ を何らかの方法で初期化 (例えば [9] 等) したうえで, 反復計算を行うことで式 (2.3) の類似度関数を小さくするような $\mathbf{F}$ と $\mathbf{Q}$ を求める方法が主流である。式 (2.4)–(2.6) の類似度関数を用いた NMF の, majorization-minimization (MM) アルゴリズム [15] に基づく反復更新式を以下に記す。

- 二乗ユークリッド距離

$$f_{ik} \leftarrow f_{ik} \cdot \frac{\sum_j x_{ij} q_{kj}}{\sum_j q_{kj} \sum_{k'} f_{ik'} q_{k'j}} \quad (2.7)$$

$$q_{kj} \leftarrow q_{kj} \cdot \frac{\sum_i x_{ij} f_{ik}}{\sum_i f_{ik} \sum_{k'} f_{ik'} q_{k'j}} \quad (2.8)$$

- 一般化 KL ダイバージェンス

$$f_{ik} \leftarrow f_{ik} \cdot \frac{\sum_j \frac{x_{ij}}{\sum_{k'} f_{ik'} q_{k'j}} q_{kj}}{\sum_j q_{kj}} \quad (2.9)$$

$$q_{kj} \leftarrow q_{kj} \cdot \frac{\sum_i \frac{x_{ij}}{\sum_{k'} f_{ik'} q_{k'j}} f_{ik}}{\sum_i f_{ik}} \quad (2.10)$$

- IS ダイバージェンス

$$f_{ik} \leftarrow f_{ik} \cdot \sqrt{\frac{\sum_j \frac{x_{ij}}{\left(\sum_{k'} f_{ik'} q_{k'j}\right)^2} q_{kj}}{\sum_j \frac{q_{kj}}{\sum_{k'} f_{ik'} q_{k'j}}}} \quad (2.11)$$

$$q_{kj} \leftarrow q_{kj} \cdot \sqrt{\frac{\sum_i \frac{x_{ij}}{\left(\sum_{k'} f_{ik'} q_{k'j}\right)^2} f_{ik}}{\sum_i \frac{f_{ik}}{\sum_{k'} f_{ik'} q_{k'j}}}} \quad (2.12)$$

なお、反復一回目の f_{ik} 及び q_{kj} の初期値には正の乱数を与えることが一般的である。これは、NMF が非負値行列を分離対象としているとともに、初期値に 0 を与えるとその要素は全く更新されなくなることを回避するためである。

2.3 STFT

NMF を音源分離に適用するためには、観測された音響信号を非負行列として表現する必要がある。このために用いられる手法が STFT である。STFT とは Fig. 2.1 に示すように、一次元の時間信号を二次元の時間周波数信号に変換する処理である。STFT の分析窓関数の長さ及びシフト長をそれぞれ Q 及び τ としたとき、時間領域の信号 $z[l]$ の j 番目の短時間区間信号（時間フレーム）は次式で表せられる。

$$\begin{aligned} \mathbf{z}^{[j]} &= [z[(j-1)\tau+1], z[(j-1)\tau+2], \dots, z[(j-1)\tau+Q]]^T \\ &= [z^{[j]}[1], z^{[j]}[2], \dots, z^{[j]}[q], \dots, z^{[j]}[Q]]^T \in \mathbb{R}^Q \end{aligned} \quad (2.13)$$

ここで、 $l = 1, 2, \dots, L$, $j = 1, 2, \dots, J$, 及び $q = 1, 2, \dots, Q$ はそれぞれ離散時間のインデックス、時間フレーム、及び時間フレーム内のサンプルを示す。また、時間フレーム数 J は次式によって与えられる。

$$J = \frac{L}{\tau} \quad (2.14)$$

ただし、信号長 L はセグメント数 J が整数となるように各時間フレームの信号の両端にゼロを挿入する処理（ゼロパディング）が施されている。そして、信号 $\mathbf{z} = [z[1], z[2], \dots, z[L]]^T \in \mathbb{R}^L$ の STFT は次のように表される。

$$\mathbf{Z} = \text{STFT}_{\boldsymbol{\omega}}(\mathbf{z}) \in \mathbb{C}^{I \times J} \quad (2.15)$$

$$z_{ij} = \sum_{q=1}^Q \omega[q] z^{[j]}[q] \exp \left\{ \frac{-\iota 2\pi(q-1)(i-1)}{F} \right\} \quad (2.16)$$

ここで、 $i = 1, 2, \dots, I$ は周波数ビンのインデックス、 F は $\lfloor \frac{F}{2} \rfloor + 1 = I$ を満たす整数（ $\lfloor \cdot \rfloor + 1 = I$ は床関数）、 ι は虚数単位、及び $\boldsymbol{\omega} = [\omega[1], \omega[2], \dots, \omega[Q]] \in \mathbb{R}^Q$ は分析窓関数をそれぞれ示している。このように、時間領域の信号に対して一定幅の短時間ごとに分析窓関数を乗じて離散フーリエ変換を行うことで、横軸が時間、縦軸が周波数のスペクトログラムと呼ばれる複素行列で表すことができる。

2.4 NMF の音源分離への応用

音響信号に NMF を適用する場合、時間信号に STFT を適用して得られるスペクトログラム（複素行列）の振幅値（振幅スペクトログラム）やパワー値（パワースペクトログラム）を非負値観測行列 $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J}$ として扱うことが一般的である。本論文では、常に $\mathbf{X}$ を振幅スペクトログラムとして定義する。人間の聴覚はスペクトログラムの位相の変化をほとんど認識す

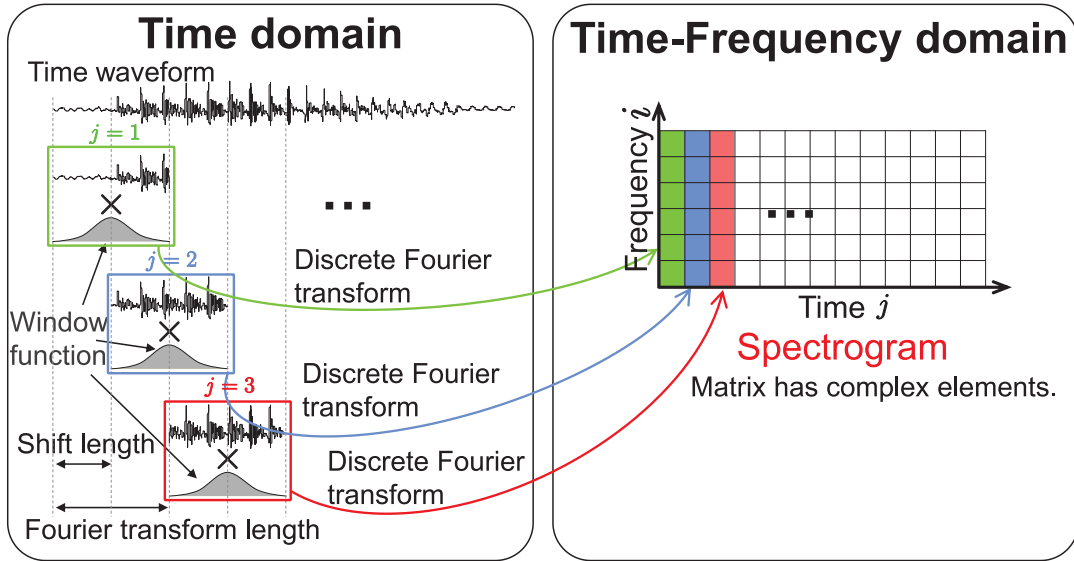


Fig. 2.1: Process of STFT.

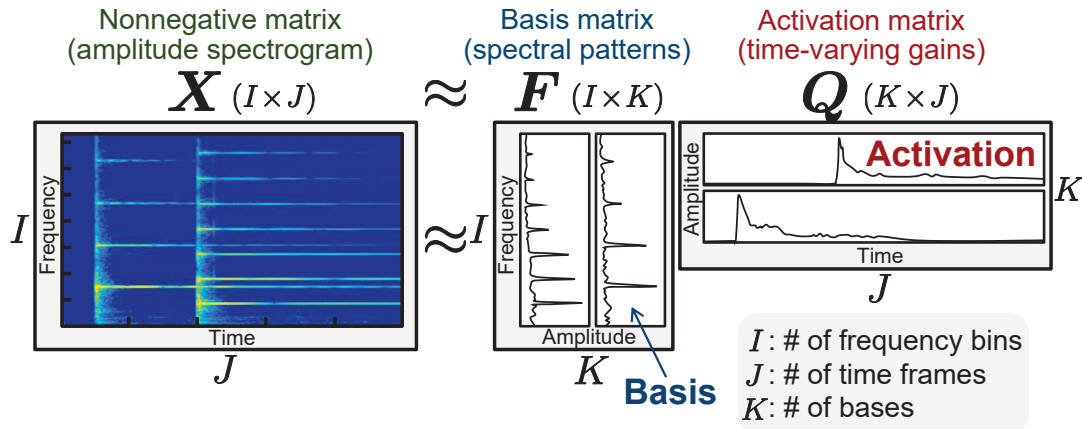


Fig. 2.2: Matrix decomposition by NMF.

ることができたため、本論文でも振幅スペクトログラムのみを音源分離処理の対象とし、位相スペクトログラムは、音源分離アルゴリズムでは取り扱わない。分離信号の振幅スペクトログラムを時間波形に復元する際は、音源分離前の（すなわち観測信号の複素スペクトログラムの）位相をそのまま用いて、分離信号の複素スペクトログラムを構成し、逆 STFT を適用する。

振幅スペクトログラム $\mathbf{X}$ を NMF で分解した様子が Fig. 2.2 である。振幅スペクトログラム $\mathbf{X}$ は NMF を用いることで基底行列 $\mathbf{F}$ とアクティベーション行列 $\mathbf{Q}$ の行列積で近似されている。Fig. 2.2 において、基底行列 $\mathbf{F}$ は 2 つの基底ベクトルから成る ($K = 2$ に設定されている)。すなわち、観測行列 $\mathbf{X}$ のスペクトログラムを 2 本のスペクトルパターン（基底ベクトル）で近似的に表現している。同様に、アクティベーション行列 $\mathbf{Q}$ は、 $\mathbf{F}$ 中の 2 本の基底ベクトルがどの時間にどの程度の強さで生起するかを表す時間的強度変化を行ベクトルとして含んでいる。このように、NMF で抽出される非負構成要素は解釈が容易であり、例えば、抽出

された非負スペクトルパターンを音源ごとに分類することができれば，特定の音源のスペクトルパターンのみを用いて振幅スペクトログラムを再構成するだけで，音源分離が可能となる．しかしながら，このような教師無し音源分類手法は，現実には非常に困難であり，NMF を用いて高音質な音源分離を達成するためには，何らかの教師情報を用いることが効果的である．

2.5 本章のまとめ

本章では，従来手法及び2つの提案手法について述べるにあたり必要となる NMF の基礎理論について述べた．また，NMF を音響信号に適用するために必要不可欠である STFT について説明し，音響信号にどのように用いられるかを説明した．次章では，NMF による音源分離性能を向上させる手法を示す．

第 3 章

従来手法

3.1 まえがき

本章では、NMF による音源分離の既存手法を紹介する。3.2 節では、SNMF による音源分離のアルゴリズムを示す。SNMF は、混合混合信号中の目的音源成分と非目的成分に似たスペクトルを含む場合、分離精度が劣化する問題がある。この問題を回避する手法として、3.3 節では、SNMF の改良手法である内積型 PSNMF について説明する。3.4 節で本章のまとめを示す。

3.2 SNMF に基づく音楽音源分離

NMF を用いた音源分離において、分離対象となる音源（目的音源）の学習用データセットが極僅かしか用意できないような状況においては、省サンプルの学習データから特徴量を学習しそのまま活用する SNMF [4, 5] が比較的高度な音源分離を達成する。SNMF には全教師あり NMF と半教師あり NMF がある。

全教師あり NMF 及び半教師あり NMF はいずれも、学習ステージと分離ステージから構成されている。Fig. 3.1 は全教師あり NMF の音源分離アルゴリズムを示している。学習ステージでは、全ての音源の教師情報（ここでは各音源のサンプル信号）の振幅スペクトログラム $\mathbf{X}_1$ 及び $\mathbf{X}_2$ を NMF でそれぞれ分解し、各音源の頻出スペクトルパターンを含む教師基底行列 $\mathbf{F}_1$ 及び $\mathbf{F}_2$ を学習する。分離ステージでは、事前学習した教師基底行列 $\mathbf{F}_1$ 及び $\mathbf{F}_2$ を用いて、混合信号の振幅スペクトログラム $\mathbf{Y}$ を $\mathbf{F}_1\mathbf{G}_1 + \mathbf{F}_2\mathbf{G}_2$ に近似的に分解することで、 $\mathbf{F}_1\mathbf{G}_1$ 及び $\mathbf{F}_2\mathbf{G}_2$ には、それぞれ事前に学習した各音源の成分が現れ、音源が分離される。Fig. 3.2 は半教師あり NMF の音源分離アルゴリズムを示している。学習ステージでは、目的音源のサンプル信号の振幅スペクトログラム $\mathbf{X}$ を NMF で分解し、頻出スペクトルパターンを含む教師基底行列 $\mathbf{F}$ を学習する。分離ステージでは、事前学習した教師基底行列 $\mathbf{F}$ を用いて混合信号の振幅スペクトログラム $\mathbf{Y}$ を $\mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}$ に近似的に分解することで、目的音源成分 $\mathbf{F}\mathbf{G}$ とその他の音源（非目的音源）成分 $\mathbf{H}\mathbf{U}$ を推定する。本論文では、半教師あり NMF のみを

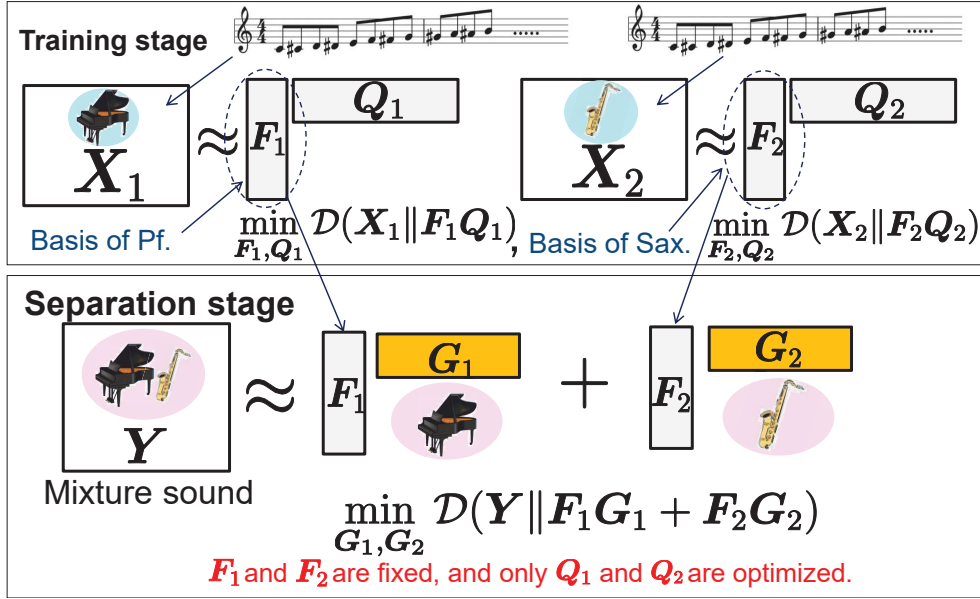


Fig. 3.1: Separation algorithm in full-supervised NMF.

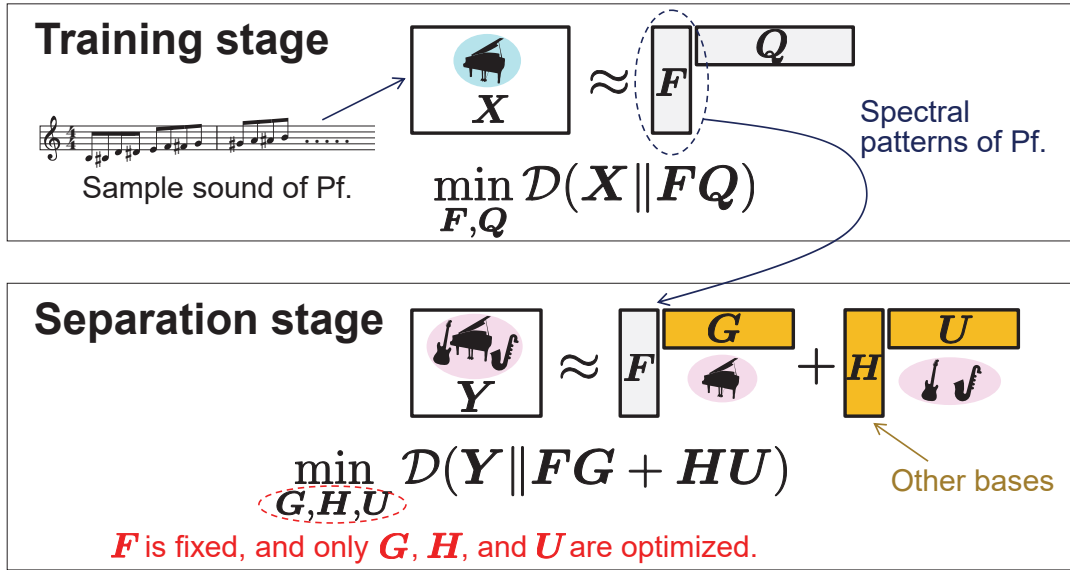


Fig. 3.2: Separation algorithm in semi-supervised NMF.

議論の対象とするが、提案するアイデア等は一般性を失うことなく全教師あり NMF にも応用可能である。以降は半教師あり NMF のことを単に「SNMF」と呼ぶ。

SNMF を用いた音源分離では、振幅スペクトログラム X を一般化 KL ダイバージェンスに基づく NMF で分解する手法が高性能となることが実験的に確認されている [4, 5]。よって、性能の良い手法を改良するのが妥当であるため、本研究では式 (2.5) に示す一般化 KL ダイバージェンスのみを用いる。SNMF では、学習ステージにおいて目的音源の音階を並べた音等のサンプル信号の振幅スペクトログラム X に NMF を適用することで、教師基底行列 F を

事前学習する．このとき， $\mathbf{F}$ は目的音源の頻出スペクトルパターンを K 本の列ベクトル（基底ベクトル）として含んでいる．分離ステージでは，教師基底行列 $\mathbf{F}$ が与えられた（固定された）下で，混合信号の振幅スペクトログラム $\mathbf{Y} \in \mathbb{R}_{\geq 0}^{I \times J}$ を次の最適化問題により分解し， $\mathbf{G}$ ， $\mathbf{H}$ ，及び $\mathbf{U}$ を推定する．

$$\min_{\mathbf{G}, \mathbf{H}, \mathbf{U}} \mathcal{D}(\mathbf{Y} \| \mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}) \quad \text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \quad (3.1)$$

ここで， $\mathbf{H} \in \mathbb{R}_{\geq 0}^{I \times L}$ 及び $\mathbf{U} \in \mathbb{R}_{\geq 0}^{L \times J}$ はそれぞれ非目的音源を表すための基底行列とアクティベーション行列であり， h_{il} 及び u_{lj} はそれぞれ $\mathbf{H}$ 及び $\mathbf{U}$ の非負要素である．また， $l = 1, 2, \dots, L$ は非目的音源の基底行列 $\mathbf{H}$ に含まれる基底ベクトルのインデックスを表す．理想的には，混合信号 $\mathbf{Y}$ 中の目的音源の成分が $\mathbf{F}\mathbf{G}$ ，非目的音源の成分が $\mathbf{H}\mathbf{U}$ として分離され，音源分離が達成される．

MM アルゴリズム [15] に基づく変数行列 $\mathbf{G}$ ， $\mathbf{H}$ ，及び $\mathbf{U}$ の更新則は次式のようになる [5]．

$$g_{kj} \leftarrow g_{kj} \cdot \frac{\sum_i \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} f_{ik}}{\sum_i f_{ik}} \quad (3.2)$$

$$h_{il} \leftarrow h_{il} \cdot \frac{\sum_j \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} u_{lj}}{\sum_j u_{lj}} \quad (3.3)$$

$$u_{lj} \leftarrow u_{lj} \cdot \frac{\sum_i \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} h_{il}}{\sum_i h_{il}} \quad (3.4)$$

ここで， y_{ij} は $\mathbf{Y}$ の要素である．

3.3 内積型 PSNMF

SNMF に基づく音源分離では，混合信号中の目的音源と非目的音源が類似したスペクトルを含む場合，分離精度が劣化する問題がある．これは，類似するスペクトルが教師基底行列 $\mathbf{F}$ または非目的音源の基底行列 $\mathbf{H}$ のいずれを用いても表現出来てしまうことが原因である．すなわち，式 (3.1) で示す最適化問題において，コスト関数が $\mathbf{X}$ と $\mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}$ 間の距離のみで定義されており，このコスト関数には基底行列同士が類似しないような制約が含まれていないことが原因である．この場合，(a) 目的音源の一部が $\mathbf{H}\mathbf{U}$ に混入する，あるいは (b) 非目的音源の一部が $\mathbf{F}\mathbf{G}$ に混入することとなり，音源分離精度が劣化する．SNMF は目的音源のみのサンプル音源を活用するため，前述の現象 (a) 及び (b) の間にはトレードオフが存在する．この両問題のトレードオフに対して，目的音の音質を可能な限り担保する（つまり，(a)

の問題をできるだけ解決する) 手法として, 内積型 PSNMF [5] が提案されている. 内積型 PSNMF は, 分離ステージの最適化において, 直交化罰則項を付与することにより, 非目的音源の基底行列 $\mathbf{H}$ を教師基底行列 $\mathbf{F}$ に強制的に直交するように誘導する. これは, (a) の問題を直接的に解決し, 目的音源成分 $\mathbf{FG}$ の成分が $\mathbf{FG}$ から欠落してしまうことをある程度避けることができる. このとき, 同時に (b) の問題を促進してしまう可能性がある. しかしながら, SNMF では, 混合信号から目的音源を正確に抜き出すことを目的としており, 分離された目的音源の音質を保証することが重要となる. すなわち, (b) の問題よりも (a) の問題の方が重要視されることが多い. そのため, 内積型 PSNMF では, 直交性罰則項の重みを調整することで, (a) と (b) の両問題のトレードオフを制御することができ, 様々な目的への音源分離の応用が可能になる.

内積型 PSNMF では, 分離ステージにおいて次の最適化問題を考える.

$$\min_{\mathbf{G}, \mathbf{H}, \mathbf{U}} \mathcal{D}(\mathbf{Y} \| \mathbf{FG} + \mathbf{HU}) + \mu \mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H}) \quad \text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \quad (3.5)$$

$$\begin{aligned} \mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H}) &= \|\mathbf{F}^T \mathbf{H}\|_{\text{Fr}}^2 \\ &= \sum_{k,l} \left(\sum_i f_{ik} h_{il} \right)^2 \end{aligned} \quad (3.6)$$

ここで, μ は罰則項 $\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ の重み係数である. また, $\|\mathbf{A}\|_{\text{Fr}}$ は任意の行列 $\mathbf{A} \in \mathbb{R}_{\geq 0}^{I \times J}$ のフロベニウスノルムであり, 次式で表される.

$$\|\mathbf{A}\|^2 = \sqrt{\sum_{i,j} a_{ij}^2} \quad (3.7)$$

ここで, a_{ij} は $\mathbf{A}$ の要素である. 罰則項 $\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ は, $\mathbf{F}$ の基底ベクトル (K 本) と $\mathbf{H}$ の基底ベクトル (L 本) の総組み合わせの内積の二乗和に対応する. したがって, $\mathbf{H}$ を推定する際には, 「 $\mathbf{Y}$ と $\mathbf{FG} + \mathbf{HU}$ が近くなること ($\mathcal{D}(\mathbf{Y} \| \mathbf{FG} + \mathbf{HU})$ を最小化すること)」及び「 $\mathbf{H}$ の基底ベクトルが $\mathbf{F}$ の基底ベクトルとなるべく直交すること ($\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ を小さくすること)」の2点が考慮され, 結果として SNMF よりも音源分離が促進されることが実験的に確認されている [5]. MM アルゴリズムに基づく変数行列 $\mathbf{G}$ 及び $\mathbf{U}$ の更新則は式 (3.2) 及び (3.4) に一致し, 変数行列 $\mathbf{H}$ の更新則は次式となる.

$$h_{il} \leftarrow h_{il} \cdot \frac{\sum_j \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} u_{lj}}{\sum_j u_{lj} + \mu f_{ik} \sum_{i'} f_{i'k} h_{i'l}} \quad (3.8)$$

3.4 本章のまとめ

本章では, まず NMF のうち SNMF のアルゴリズムを説明した. その後, SNMF の問題点を述べ, その解決策として既に提案されている内積型 PSNMF を紹介した. 次章では, この

内積型 PSNMF の問題点を述べ，その問題を回避するような別の罰則条件付き SNMF を提案する．

第 4 章

対数コサイン類似度罰則条件に基づく提案手法

4.1 まえがき

前章では、SNMF の改善手法である内積型 PSNMF を紹介したが、実はこの手法は、NMF の持つスケール不定性により、罰則項が正しく寄与しない定式化となっている。そこで本論文では、ベクトルの長さに非依存なコサイン類似度に基づく罰則条件付き SNMF を新たに提案する。本章では、1 つ目の提案手法として、対数コサイン類似度を用いる Log-Cos 型 PSNMF を提案し、定式化及び更新則の導出を示す。

4.2 動機

内積型 PSNMF には、罰則項の効果が直接反映されない本質的な問題がある。これは、NMF が本来持っている基底行列・アクティベーション行列間のスケール不定性が原因である。すなわち、 $\mathbf{X} \approx \mathbf{F}\mathbf{Q}$ という NMF 分解を考えた際に、基底行列 $\mathbf{F}$ に任意の正の係数 α を乗じて $\alpha\mathbf{F}$ と変形しても、同時にアクティベーション行列 $\mathbf{Q}$ を $(1/\alpha)\mathbf{Q}$ とすれば、 $\mathbf{F}\mathbf{Q}$ 自体の値は変わらないという不定性である。

内積型 PSNMF では、罰則項 $\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ の値を小さくすることで $\mathbf{F}$ と $\mathbf{H}$ の非類似性を高めることを考える。しかし、この罰則項は次式のように $\mathbf{H}$ に非常に小さな値 α を乗じるだけで最小化可能である。

$$\lim_{\alpha \rightarrow +0} \mathcal{P}_{\text{inner}}(\mathbf{F}, \alpha\mathbf{H}) = \lim_{\alpha \rightarrow +0} \alpha \|\mathbf{F}^T \mathbf{H}\|_{\text{Fr}}^2 = 0 \quad (4.1)$$

このとき、 $\mathbf{H}\mathbf{U}$ の値が変わらないように、アクティベーション行列 $\mathbf{U}$ は $\mathbf{U} \leftarrow (1/\alpha)\mathbf{U}$ と更新されれば、式 (3.5) のコスト関数の第一項は次式のように不変となる。

$$\mathcal{D}(\mathbf{Y} \|\mathbf{F}\mathbf{G} + \alpha\mathbf{H}[\alpha^{-1}\mathbf{U}]) = \mathcal{D}(\mathbf{Y} \|\mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}) \quad (4.2)$$

したがって、罰則項 $\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ は、実は数理的には直接 $\mathbf{F}$ と $\mathbf{H}$ を直交化しておらず、式

(3.5) の最適化問題は、内積型 PSNMF のコスト関数の式 (3.1) と本質的には等価となる。

この問題を避けるために、文献 [5] では、 $\mathbf{H}$ を更新する度に全基底ベクトルを正規化する（正規化係数の逆数を $\mathbf{U}$ の行ベクトルに乗じて $\mathbf{HU}$ は不変とする）ようにアルゴリズムを実装しており、実験的には性能改善が見られたが、この実装方法は根本的な解決とは言い難い。加えて、この基底正規化により、コスト関数 (3.5) の値が増加するため、MM アルゴリズムの持つ収束性の理論的な保証が失われてしまう。同様の現象は、 L_1 ノルムに基づくスパース正則化をアクティベーション行列に与える NMF でも確認されており、基底ベクトルの長さを制約条件に与える最適化法が提案されている [16]。内積型 PSNMF においても、文献 [16] と同様の制約条件を導入して解くことはできる。しかし、この手法もまた、理論的な収束は保証されていない欠点がある。これは、文献 [16] では、勾配に正負の項を持つ分数を基にした更新則を利用していることに起因している。この更新則は NMF に基づく手法でしばしば用いられることがある [17, 18] が、文献 [16] やそれと同様のノルム制約に基づく最適化に適用した場合の理論的収束性は保証されていない。

そこで、本論文では、スケール不定性の影響を回避しつつ、最適化アルゴリズムの理論的収束性が保証されたアルゴリズムの提案を目指す。提案手法では、基底ベクトル間の大きさには依存せず、角度にのみ依存するような直交性の指標となるコサイン類似度を罰則項に用いる。本章ではまず、コサイン類似度を MM アルゴリズムで解きやすい形に変形するために、対数コサイン類似度を用いた手法である Log-Cos 型 PSNMF について示す。

4.3 コスト関数

Log-cos 型 PSNMF では、次の最適化問題を考える。

$$\min_{\mathbf{G}, \mathbf{H}, \mathbf{U}} \mathcal{J}_1 \quad \text{s.t. } g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \quad (4.3)$$

$$\mathcal{J}_1 = \mathcal{D}(\mathbf{Y} \| \mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}) + \mu \mathcal{P}_{\log\cos}(\mathbf{F}, \mathbf{H}) \quad (4.4)$$

$$\mathcal{P}_{\log\cos}(\mathbf{F}, \mathbf{H}) = \sum_{k,l} \log \frac{\sum_i f_{ik} h_{il}}{\left(\sum_i f_{ik}^2\right)^{\frac{1}{2}} \left(\sum_i h_{il}^2\right)^{\frac{1}{2}}} \quad (4.5)$$

$$= \sum_{k,l} \left[\log \sum_i f_{ik} h_{il} - \frac{1}{2} \log \sum_i f_{ik}^2 - \frac{1}{2} \log \sum_i h_{il}^2 \right] \quad (4.6)$$

ここで、罰則項 $\mathcal{P}_{\log\cos}(\mathbf{F}, \mathbf{H})$ は、 $\mathbf{F}$ の基底ベクトルと $\mathbf{H}$ の基底ベクトルの総組み合わせのコサイン類似度の対数和に対応する。コサイン類似度であるので、基底ベクトルの長さには依らず、角度のみで直交性を測ることができ、NMF におけるスケールの不定性は影響しない。また、基底ベクトルは全て非負であることから、コサイン類似度の値は最小で 0（直交）、最大で 1（平行）となる。純粋なコサイン類似度ではなく、コサイン類似度の対数を罰則項に用いる理由は、式 (4.6) に示すように、分数から成るコサイン類似度を分解するためである。この

変形により、次節で示す MM アルゴリズムに基づく更新式の導出が容易となり、結果として得られる反復更新則も実装のしやすいシンプルなものとなる利点がある。

4.4 反復更新式の導出

式 (4.4) は次のように変形できる。

$$\begin{aligned}\mathcal{J}_1 &= \mathcal{D}(\mathbf{Y} \| \mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}) + \mu \mathcal{P}_{\log\cos}(\mathbf{F}, \mathbf{H}) \\ &= \sum_{i,j} \left[y_{ij} \log y_{ij} - y_{ij} \log \left(\sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right) - y_{ij} + \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right] \\ &\quad + \mu \sum_{k,l} \left[\log \sum_i f_{ik} h_{il} - \frac{1}{2} \log \sum_i f_{ik}^2 - \frac{1}{2} \log \sum_i h_{il}^2 \right] \quad (4.7)\end{aligned}$$

式 (4.7) のコスト関数を最小化する変数 g_{kj} , h_{il} , 及び u_{lj} を求めることが目的であるが、第二項、第六項、及び第八項が対数関数の内部に変数の総和を含んでおり、これの各変数に対する停留点（微分値が 0 となる変数の値）を求めるのは困難である。このようにコスト関数の停留点を解析的に求めることが困難である場合、停留点が解析的に求められる上限関数（コスト関数と接する、または、上回る関数）を設計して、コスト関数の代わりに上限関数の最小点を求めることを繰り返すことで、コスト関数の最小点を求める最適化手法が MM アルゴリズムである。本論文においても、従来の NMF の最適化手法と同様に、MM アルゴリズムに基づく最適化アルゴリズム（反復更新式）を導出する。

式 (4.7) の第二項は変数 (g_{kj} , h_{il} , 及び u_{lj}) の総和を含む負対数関数となっており、上限関数を設計する必要がある。負対数関数 ($-\log$) は、変数に対して凸関数であるため、次式のように Jensen の不等式 (A.1) を適用することで上限関数を設計できる。

$$\begin{aligned}-\log \left(\sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right) &= -\log \left(\sum_k \frac{\alpha_{ijk} f_{ik} g_{kj}}{\alpha_{ijk}} + \sum_l \frac{\beta_{ijl} h_{il} u_{lj}}{\beta_{ijl}} \right) \\ &\leq -\sum_k \alpha_{ijk} \log \frac{f_{ik} g_{kj}}{\alpha_{ijk}} - \sum_l \beta_{ijl} \log \frac{h_{il} u_{lj}}{\beta_{ijl}} \quad (4.8)\end{aligned}$$

ここで $\alpha_{ijk} > 0$ 及び $\beta_{ijl} > 0$ は $\sum_k \alpha_{ijk} = 1$ 及び $\sum_l \beta_{ijl} = 1$ を満たす補助変数である。式 (4.8) の等号成立条件は次式で与えられる。

$$\alpha_{ijk} = \frac{f_{ik} g_{kj}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_l h_{il} u_{lj}} \quad (4.9)$$

$$\beta_{ijl} = \frac{h_{il} u_{lj}}{\sum_k f_{ik} g_{kj} + \sum_{l'} h_{il'} u_{l'j}} \quad (4.10)$$

また、式 (4.7) の第六項も変数 h_{il} の総和を含む正対数関数となっており、上限関数を設計する必要がある。正対数関数 ($+\log$) は、変数に対して凹関数であるため、次式のように接線

不等式 (A.2) を適用することで上限関数を設計できる.

$$\log \sum_i f_{ik} g_{kj} \leq \frac{1}{\delta_{kli}} \left(\sum_i f_{ik} g_{kj} - \delta_{kli} \right) + \log \delta_{kli} \quad (4.11)$$

ここで, $\delta_{kli} > 0$ は $\sum_i \delta_{kli} = 1$ を満たす補助変数である. 式 (4.11) の等号成立条件は次式で与えられる.

$$\delta_{kli} = \sum_i f_{ik} h_{il} \quad (4.12)$$

さらに, 式 (4.7) の第八項も変数 h_{il} の総和を含む負対数関数となっており, 上限関数を設計する必要がある. 次式のように Jensen の不等式を適用することで上限関数を設計できる.

$$\begin{aligned} -\log \sum_i h_{il}^2 &= -\log \sum_i \gamma_{il} \frac{h_{il}^2}{\gamma_{il}} \leq -\sum_i \gamma_{il} \log \frac{h_{il}^2}{\gamma_{il}} \\ &= -2 \sum_i \gamma_{il} \log h_{il} + \sum_i \gamma_{il} \log \gamma_{il} \end{aligned} \quad (4.13)$$

ここで, $\gamma_{il} > 0$ は $\sum_i \gamma_{il} = 1$ を満たす補助変数である. 式 (4.13) の等号成立条件は次式で与えられる.

$$\gamma_{il} = \frac{h_{il}^2}{\sum_{i'} h_{i'l}^2} \quad (4.14)$$

式 (4.7) に式 (4.8), (4.11), 及び (4.13) を適用すると, $\mathcal{J}_1 \leq \mathcal{J}_1^+$ を満たす最終的な上限関数 $\mathcal{J}_1^+$ が次式のように設計できる.

$$\begin{aligned} \mathcal{J}_1 &\leq \mathcal{J}_1^+ \\ &= \sum_{i,j} \left[y_{ij} \log y_{il} - y_{ij} \left(\sum_k \alpha_{ijk} \log \frac{f_{ik} g_{kj}}{\alpha_{ijk}} + \sum_l \beta_{ijl} \log \frac{h_{il} u_{lj}}{\beta_{ijl}} \right) \right. \\ &\quad \left. - y_{ij} + \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right] + \mu \sum_{k,l} \left[\frac{1}{\delta_{kli}} \left(\sum_i f_{ik} h_{il} - \delta_{kli} \right) + \log \delta_{kli} \right. \\ &\quad \left. - \frac{1}{2} \log \sum_i f_{ik}^2 - \sum_i \gamma_{il} \log h_{il} + \frac{1}{2} \sum_i \gamma_{il} \log \gamma_{il} \right] \end{aligned} \quad (4.15)$$

$\mathbf{H}$ の更新式は, 式 (4.15) の右辺を h_{il} で偏微分し 0 とおいた次式 ($\partial \mathcal{J}^+ / \partial h_{il} = 0$) から得られる.

$$\begin{aligned} \sum_j \left[-y_{ij} \beta_{ijl} \cdot \frac{1}{h_{il} u_{lj}} u_{lj} + u_{lj} \right] + \mu \sum_k \left[\frac{1}{\delta_{kli}} f_{ik} - \gamma_{il} \cdot \frac{1}{h_{il}} \right] &= 0 \\ -\sum_j y_{ij} \beta_{ijl} \frac{1}{h_{il}} + \sum_j u_{lj} + \mu \sum_k \frac{1}{\delta_{kli}} f_{ik} - \mu \sum_k \gamma_{il} \frac{1}{h_{il}} &= 0 \end{aligned} \quad (4.16)$$

式 (4.16) を $\frac{1}{h_{il}}$ について整理すると, 次式となる.

$$\frac{1}{h_{il}} \left[\sum_j y_{ij} \beta_{ijl} + \mu \sum_k \gamma_{il} \right] = \sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\delta_{kli}} \quad (4.17)$$

式 (4.17) より, h_{il} について解くと以下の式に変形できる.

$$h_{il} = \frac{\sum_j y_{ij} \beta_{ijl} + \mu \sum_k \gamma_{il}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\delta_{kli}}} \quad (4.18)$$

さらに補助変数の等号成立条件式 (4.9), (4.10), (4.12), 及び (4.14) を上式に代入すると, 反復更新式が得られる.

$$h_{il} \leftarrow \frac{\frac{\sum_j y_{ij} h_{il} u_{lj}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} + \mu \sum_k \frac{h_{il}^2}{\sum_{i'} i'^2}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\sum_{i'} f_{i'k} h_{i'l}}} \quad (4.19)$$

式 (4.19) の分子の第二項において, 基底ベクトル数 k の総和を取っているが第二項の中に k の要素は含まれていない. よって k の総和は第二項全体を K 倍することとなり, 次式で表される.

$$h_{il} \leftarrow \frac{\frac{\sum_j y_{ij} h_{il} u_{lj}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} + \mu K \frac{h_{il}^2}{\sum_{i'} h_{i'l}^2}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\sum_{i'} f_{i'k} h_{i'l}}} \quad (4.20)$$

さらに, 式 (4.20) の分子の全ての項に含まれている共通因数 h_{il} をくくりだすことで最終的な更新式の形は次式となる.

$$h_{il} \leftarrow h_{il} \cdot \frac{\sum_j \frac{y_{ij}}{\sum_{k'} f_{ik'} g_{k'j} + \sum_{l'} h_{il'} u_{l'j}} u_{lj} + \mu K \frac{h_{il}}{\sum_{i'} h_{i'l}^2}}{\sum_j u_{lj} + \mu \sum_k \frac{f_{ik}}{\sum_{i'} f_{i'k} h_{i'l}}} \quad (4.21)$$

なお, Log-Cos 型 PSNMF の罰則項には $\mathbf{G}$ 及び $\mathbf{U}$ の要素が含まれていないため, 変数 $\mathbf{G}$ 及び $\mathbf{U}$ の更新則は罰則項を付与する前の式 (3.2) 及び (3.4) と同様である.

4.5 本章のまとめ

本章では，SNMF の改善手法である内積型 PSNMF の問題点を説明した．そのうえで，内積型 PSNMF のスケール不定性の問題を回避する手法として，SNMF に対数コサイン類似度を罰則項として付与する Log-Cos 型 PSNMF を提案し，反復更新式の導出を行った．次章では，Log-Cos 型 PSNMF の問題点を述べ，代わりとなる新たな手法を提案する．

第 5 章

コサイン類似度罰則条件に基づく提案手法

5.1 まえがき

前章では、内積型 PSNMF の問題を回避する手法として、Log-Cos 型 PSNMF を提案した。しかし、Log-Cos 型 PSNMF では、対数を用いたことにより数値が不安定になるといった問題が新たに生じる。本章では、この問題について詳しく述べたうえで、これを回避する Cos 型 PSNMF を提案し、定式化及び更新則の導出を示す。

5.2 Log-Cos 型 PSNMF の問題点

Log-Cos 型 PSNMF では、反復更新式の導出を容易にするために、コサイン類似度の対数和を罰則項として用いた。この更新式を繰り返すことで、 $\mathbf{F}$ と直交に近づく $\mathbf{H}$ を推定することが期待される。しかし、Log-Cos 型 PSNMF では、 $\mathbf{h}_l$ が $\mathbf{F}$ に含まれる基底ベクトルと完全に直交している場合や、任意の l に対して、 $\mathbf{h}_l = \mathbf{0}$ となる場合に、罰則項が $-\infty$ となってしまう。このような場合、最小化問題の解は $\mathcal{J}_1 \rightarrow -\infty$ を満たす変数 $\mathbf{G}$, $\mathbf{H}$, 及び $\mathbf{U}$ となり、例えば $\mathbf{H} = \mathbf{0}$ ($\mathbf{G}$ 及び $\mathbf{H}$ は任意) という望ましくない解（音源分離を達成しない解）が推定されてしまう。この問題を避けるために、反復更新ごとに次に示すような場当たり的な（しかし NMF の実装においてはしばしばみられる [19]）フロアリング処理を適用する。

$$h_{il} \leftarrow \max(h_{il}, \varepsilon), \quad (5.1)$$

ここで、 ε は計算機イプシロン（倍精度浮動小数点では $\varepsilon \approx 2.22 \times 10^{-16}$ ）を表し、 $\max(\cdot)$ は入力された引数から最大値を返す関数である。Log-Cos 型 PSNMF では、計算機イプシロンを適用して前述の無意味な解に誘導される問題を回避できるが、このような処理は内積型 PSNMF の基底ベクトルの正規化と同様、根本的な解決とは言えない。そこで、本章では、対数を用いないシンプルなコサイン類似度罰則を付与する Cos 型 PSNMF を提案する。Log-Cos 型 PSNMF に比べて導出・更新則が複雑になる欠点があるが、理論的には収束性が

保証された最適化アルゴリズムを得ることができる．また，Log-Cos 型 PSNMF 同様，内積型 PSNMF のスケール不定性の問題を同時に回避することができる．

5.3 コスト関数

Cos 型 PSNMF では次のような最適化問題を考える．

$$\min_{\mathbf{G}, \mathbf{H}, \mathbf{U}} \mathcal{J}_2 \quad \text{s.t.} \quad g_{kj}, h_{il}, u_{lj} \geq 0 \quad \forall i, j, k, l \quad (5.2)$$

$$\mathcal{J}_2 = \mathcal{D}(\mathbf{Y} \| \mathbf{F}\mathbf{G} + \mathbf{H}\mathbf{U}) + \mu \mathcal{P}_{\text{cos}}(\mathbf{F}, \mathbf{H}) \quad (5.3)$$

$$\mathcal{P}_{\text{cos}}(\mathbf{F}, \mathbf{H}) = \sum_{k,l} \frac{\sum_i f_{ik} h_{il}}{(\sum_i f_{ik}^2)^{\frac{1}{2}} (\sum_i h_{il}^2)^{\frac{1}{2}}} \quad (5.4)$$

Cos 型 PSNMF では， $\mathcal{P}_{\text{logcos}}(\mathbf{F}, \mathbf{H})$ の対数関数が省略されており，罰則関数の最小値は 0 であるため，コスト関数 $\mathcal{J}_2$ の最小値も 0 として下限が定義される．これにより，前節で述べたような無意味な解は存在せず，理論的にも正しい定式化となっている．

5.4 反復更新式の導出

コサイン類似度と同様，コスト関数 (5.4) を直接最小化するのは困難である．そこで MM アルゴリズムを適用する．Log-Cos 型 PSNMF と同様に，罰則項には $\mathbf{G}$ 及び $\mathbf{U}$ の要素が含まれないため，変数 $\mathbf{G}$ 及び $\mathbf{U}$ の更新則はそれぞれ式 (3.2) 及び (3.4) と同等である．罰則項 $\mathcal{P}_{\text{cos}}(\mathbf{F}, \mathbf{H})$ は $(\sum_i h_{il}^2)^{-1/2}$ が含まれており， h_{il} の停留点が解析的に求められない原因となっている．この関数は凸関数であるため，Jensen の不等式を用いて次のように上限関数を設計することができる．

$$\begin{aligned} \left(\sum_i h_{il}^2 \right)^{-\frac{1}{2}} &= \left(\sum_i \epsilon_{il} \frac{h_{il}^2}{\epsilon_{il}} \right)^{-\frac{1}{2}} \leq \sum_i \epsilon_{il} \left(\frac{h_{il}^2}{\epsilon_{il}} \right)^{-\frac{1}{2}} \\ &= \sum_i \epsilon_{il}^{\frac{3}{2}} h_{il}^{-1} \end{aligned} \quad (5.5)$$

ここで $\epsilon_{il} > 0$ は $\sum_i \epsilon_{il} = 1$ を満たす補助変数である．式 (5.5) の等号成立条件は次の場合のみ成立する．

$$\epsilon_{il} = \frac{h_{il}^2}{\sum_{i'} h_{i'l}^2} \quad (5.6)$$

一般化 KL ダイバージェンスについて、前章と同様の不等式を適用すると、式 (5.3) 及び (5.5) から $\mathcal{J}_2$ の上限関数 $\mathcal{J}_2^+$ が次のように設計できる。

$$\begin{aligned} \mathcal{J}_2 &\leq \mathcal{J}_2^+ \\ &= \sum_{i,j} \left[y_{ij} \log y_{ij} - y_{ij} \left(\sum_k \alpha_{ijk} \log \frac{f_{ik} g_{kj}}{\alpha_{ijk}} + \sum_l \beta_{ijl} \log \frac{h_{il} u_{lj}}{\beta_{ijl}} \right) \right. \\ &\quad \left. - y_{ij} + \sum_k f_{ik} g_{kj} + \sum_l h_{il} u_{lj} \right] + \mu \sum_{k,l} \left[\left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(\sum_{i',i''} f_{i'k} \epsilon_{i''l}^{\frac{3}{2}} \frac{h_{i'l}}{h_{i''l}} \right) \right] \quad (5.7) \end{aligned}$$

h_{il} の更新則は $\partial \mathcal{J}_2^+ / \partial h_{il} = 0$ を解くことで導出することができる。

$$\sum_j \left[\frac{-y_{ij} \beta_{ijl}}{h_{il}} + u_{lj} \right] + \mu \sum_k \left[\left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(-\frac{1}{h_{il}^2} \epsilon_{il}^{\frac{3}{2}} \sum_{i' \neq i} f_{i'k} h_{i'l} + f_{ik} \sum_{i' \neq i} \epsilon_{i'l}^{\frac{3}{2}} \frac{1}{h_{i'l}} \right) \right] = 0 \quad (5.8)$$

式 (5.8) 全体に h_{il}^2 を乗じることで次式が得られる。

$$h_{il}^2 \sum_j \left[\frac{-y_{ij} \beta_{ijl}}{h_{il}} + u_{lj} \right] + h_{il}^2 \mu \sum_k \left[\left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(-\frac{1}{h_{il}^2} \epsilon_{il}^{\frac{3}{2}} \sum_{i' \neq i} f_{i'k} h_{i'l} + f_{ik} \sum_{i' \neq i} \epsilon_{i'l}^{\frac{3}{2}} \frac{1}{h_{i'l}} \right) \right] = 0 \quad (5.9)$$

式 (5.9) を変形すると次式になる。

$$\begin{aligned} &\sum_j [-y_{ij} \beta_{ijl} h_{il} + u_{lj} h_{il}^2] \\ &+ \mu \sum_k \left[\left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(-\epsilon_{il}^{\frac{3}{2}} \sum_{i' \neq i} f_{i'k} h_{i'l} + f_{ik} h_{il}^2 \sum_{i' \neq i} \epsilon_{i'l}^{\frac{3}{2}} \frac{1}{h_{i'l}} \right) \right] = 0 \quad (5.10) \end{aligned}$$

さらに、式 (5.10) を h_{il} の指数毎に整理すると次式が得られる。

$$\begin{aligned} &h_{il}^2 \left[\sum_j u_{lj} + \mu \sum_k f_{ik} \left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(\sum_{i' \neq i} \epsilon_{i'l}^{\frac{3}{2}} \frac{1}{h_{i'l}} \right) \right] \\ &+ h_{il} \left[-\sum_j y_{ij} \beta_{ijl} \right] + \left[-\mu \epsilon_{il}^{\frac{3}{2}} \sum_k \left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(\sum_{i' \neq i} f_{i'k} h_{i'l} \right) \right] \quad (5.11) \end{aligned}$$

式 (5.11) の h_{il}^2 の係数、 h_{il} の係数、及び第 3 項をそれぞれ次式に置き換える。

$$a_{il} = \sum_j u_{lj} + \mu \sum_k f_{ik} \left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(\sum_{i' \neq i} \epsilon_{i'l}^{\frac{3}{2}} \frac{1}{h_{i'l}} \right) \quad (5.12)$$

$$b_{il} = - \sum_j y_{ij} \beta_{ijl} \quad (5.13)$$

$$c_{il} = -\mu \epsilon_{il}^{\frac{3}{2}} \sum_k \left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(\sum_{i' \neq i} f_{i'k} h_{i'l} \right) \quad (5.14)$$

式 (5.12)–(5.14) を式 (5.11) に適用することで以下のような二次方程式が得られる.

$$a_{il} h_{il}^2 + b_{il} h_{il} + c_{il} = 0 \quad (5.15)$$

したがって, 補助関数 $\mathcal{J}_2^+$ の h_{il} に関する停留点は二次方程式の解の公式で次のように与えられる.

$$h_{il} = \frac{-b_{il} \pm \sqrt{b_{il}^2 - 4a_{il}c_{il}}}{2a_{il}} \quad (5.16)$$

ここで, 式 (5.16) の $\pm$ は h_{il} の解が非負になるように決定する. h_{il} の更新則は, 等号成立条件の式 (4.10), (4.14), 及び (5.6) を式 (5.16) に代入することで得ることができる. しかし, その更新則は非常に複雑になるので, 本論文では代わりに a_{il} , b_{il} , 及び c_{il} の更新則を示す.

$$a_{il} \leftarrow \sum_j u_{lj} + \mu \sum_k f_{ik} \left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left[\sum_{i' \neq i} h_{i'l}^2 \left(\sum_{i''} h_{i''l}^2 \right)^{-\frac{3}{2}} \right] \quad (5.17)$$

$$b_{il} \leftarrow - \sum_j y_{ij} \frac{h_{il} u_{lj}}{\sum_k f_{ik} g_{kj} + \sum_{l'} h_{il'} u_{l'j}} \quad (5.18)$$

$$c_{il} \leftarrow -\mu \left(\frac{h_{il}^2}{\sum_{i'} h_{i'l}^2} \right)^{\frac{3}{2}} \sum_k \left(\sum_{i'} f_{i'k}^2 \right)^{-\frac{1}{2}} \left(\sum_{i' \neq i} f_{i'k} h_{i'l} \right) \quad (5.19)$$

したがって, 更新後の h_{il} は式 (5.17)–(5.19) を更新した後で, 式 (5.16) を計算することで得られる. なお, $\mu = 0$ のとき, 式 (5.16)–(5.19) は式 (3.8) と一致する.

5.5 本章のまとめ

本章では, 提案手法である Cos 型 PSNMF について説明し, 反復更新式の導出を行った. Cos 型 PSNMF とは, SNMF に罰則項としてコサイン類似度を付与する手法である. 次章では, 従来手法の内積型 PSNMF と 4 章及び 5 章で説明した Log-Cos 型 PSNMF 及び Cos 型 PSNMF の比較実験を行い提案手法の有効性を客観的に評価する.

第 6 章

音楽信号の分離実験

6.1 まえがき

本章では SNMF, 内積型 PSNMF, Log-Cos 型 PSNMF 及び Cos 型 PSNMF の混合音源の音源分離性能を比較する. 6.2 節で実験条件を述べ, 6.3 節で実験結果を述べる. 6.4 節では, 4 手法の SDR 値の差が有意なものであるかを結論付けるために, 実験結果を基に統計的検定を行う. 6.5 節で本章のまとめを示す.

6.2 実験条件

提案手法の妥当性を確認するために, SNMF, 内積型 PSNMF [5], Log-Cos 型 PSNMF, 及び Cos 型 PSNMF の 4 手法で音楽信号を対象としたモノラル音源分離の比較実験を行った. 本実験では, 開発データセット及びテストデータセットとして, ヤマハ MU-1000 シンセサイザーが生成した人工音源 songKitamura [20] を使用した. オーボエ (Ob.), トランペット (Tp.), ホルン (Hr.), フルート (Fl.), バイオリン (Vn.), クラリネット (Cl.), ピアノ (Pf.), ハープシコード (Hp.), トロンボーン (Tr.), ファゴット (Fg.), チェロ (Vc.) の 11 種類の楽器から 2 つの楽器メロディーを選び, 同じパワーでミキシングして 2 音源混合モノラル信号 $\mathbf{Y}$ を生成し, 90 個の混合信号を得た. そして, 90 個の混合信号を 45 個の開発データセットと 45 個のテストデータセットにランダムに分割した. 開発データセットは, 各手法における最適なハイパーパラメータ μ を求めるために使用し, そのパラメータ値を用いてテストデータセットにより性能を比較した. Fig. 6.1 に, 4 手法の学習ステージで使用した 2 オクターブの上昇音 (目的音のサンプル音) を示す. したがって学習ステージでは, 目的音源となる楽器の 2 オクターブ上昇音を $\mathbf{X}$ として用い, 教師基底行列 $\mathbf{F}$ を推定した. 実際に音源分離の対象となる混合音 $\mathbf{Y}$ を作成するために使用したメロディーが Fig. 6.2 である. この中から違うメロディーを持つ 2 楽器を選び, そのような混合信号を 90 曲作成した.

評価指標には, 音源対歪み比 (source-to-distortion: SDR) [21] を使用した. これは, 音源分離度合いと分離音の品質 (歪みの少なさ) の両方を加味した総合的な客観指標であり, SDR



Fig. 6.1: Two-octave ascending notes used as training sample sound.



Fig. 6.2: Four melody parts used as mixture observation.

の値が高いほど信号の歪みが小さいことを示す。今、複数の音源が混合しているとき、全楽器の混合音の時間信号を $m[l]$ ，推定したい音源（目的音）の時間信号を $s[l]$ ，推定したい音源以外の音源（非目的音）の時間信号を $n[l]$ すると、次式で表される。

$$m[l] = s[l] + n[l] \quad (6.1)$$

この混合音源 $m[l]$ に対して音源分離を適用して得られる目的音の推定信号を $\hat{m}[l]$ とすると、

$$\hat{m}[l] = e_s[l] + e_n[l] + e_a[l] \quad (6.2)$$

と仮定できる。ここで、 $e_s[l]$ ， $e_n[l]$ ，及び $e_a[l]$ はそれぞれ目的音源成分，推定目的音信号に残留した非目的音源成分，及び音源分離によって生じた人工的な歪み成分（ $s[l]$ 及び $n[l]$ のどちらにも属さない成分）を表す。SDR とは、 $e_s[l]$ ， $e_n[l]$ ，及び $e_a[l]$ を用いて次式で表される。

$$\text{SDR} = 10 \log_{10} \frac{\sum_{l=1}^L |e_s[l]|^2}{\sum_{l=1}^L |e_n[l] + e_a[l]|^2} \quad [\text{dB}] \quad (6.3)$$

したがって、高い SDR 値を達成するには、 $e_n[l]$ と $e_a[l]$ が少なく、 $e_s[l]$ が高精度に推定されている必要がある。これはすなわち、可能な限り目的音源成分を欠かすことなく推定し、可能な限り非目的音源成分を抑圧することに相当する。

各手法における各行列の初期値は、区間 $(0, 1)$ に一様に分布するランダムな値とした。STFT の窓長とシフト長は、それぞれ音楽信号の解析で一般的である 92.9 ms と 46.4 ms に設定した。目的音源の基底ベクトル数 K は、2 オクターブの 24 本と 3 本の計 27 本とした。非目的音源は、どのようなメロディーが与えられるか不明である。そこで、予備実験で基底ベクトル数 L を 30 本、50 本、75 本、及び 100 本の 4 種類に変化させたとき、分離性能にどのように影響を与えるかを調べた。その結果、基底ベクトル数の違いによる分離性能の差はほとんど見られなかったため、本実験では、 $L = 50$ を用いた。

内積型 PSNMF では、4 章で述べたように、罰則項 $\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ を有効にするために、以下の基底ベクトルの正規化を行う必要がある。

$$h_{il} \leftarrow \frac{1}{d_l} h_{il}, \quad (6.4)$$

$$u_{lj} \leftarrow d_l u_{lj}, \quad (6.5)$$

ここで、 $d_l = \sum_i h_{il}$ である。この正規化は、各反復においてすべてのパラメータを更新した後に行った。また、Log-Cos 型 PSNMF では、数値の不安定さを避けるために式 (5.1) を適用後、式 (6.4) 及び (6.5) を適用した。

6.3 実験結果

Fig. 6.3 は内積型 PSNMF, Log-Cos 型 PSNMF, 及び Cos 型 PSNMF における開発データセットの平均 SDR の挙動である。横軸が重み係数であり、縦軸が平均 SDR である。Log-Cos 型 PSNMF と Cos 型 PSNMF の 2 つの提案手法は、最適なハイパーパラメータ設定において内積型 PSNMF を上回ることが確認できる。これは、罰則項 $\mathcal{P}_{\text{logcos}}(\mathbf{F}, \mathbf{H})$ と $\mathcal{P}_{\text{cos}}(\mathbf{F}, \mathbf{H})$ が、 $\mathbf{H}$ の最適化において、教師基底行列 $\mathbf{F}$ のベクトルと非目的音源の基底行列 $\mathbf{H}$ のベクトルの直交化を適切に促しているためである。一方、内積型 PSNMF の罰則項である $\mathcal{P}_{\text{inner}}(\mathbf{F}, \mathbf{H})$ は、NMF におけるスケール不定性により最適な直交化がなされない可能性がある。すなわち、 μ の値を大きくしても $\mathbf{H}$ のスケールが小さくなるだけの更新がなされる可能性がある。Log-Cos 型 PSNMF 及び Cos 型 PSNMF は、スケール不定性の影響を受けないため、SDR 値が高い値となる μ の幅が比較的狭くなっていることも確認できる。

各罰則条件付き SNMF のハイパーパラメータ μ を、開発データセットで最も高い平均 SDR が得られる最適な値に設定し、テストデータセットに対する各手法の SDR 値を比較した。45 個の SDR の結果をバイオリン図に表したものが Fig. 6.4 である。白点は 45 個の SDR の中央値、白点から上下に伸びる灰色の線はそれぞれ第 3 四分位数と第 1 四分位数の範囲を示している。また、バイオリン図の膨らみから、各曲の SDR がどのあたりに集中しているかを知ることができる。結果の平均値と中央値は、Table 6.1 に示すとおりである。Log-Cos 型 PSNMF 及び Cos 型 PSNMF とともに、平均値も中央値も内積型 PSNMF のそれらを上回っていることが確認できる。特に、Cos 型 PSNMF は最良の中央値を示し、内積型 PSNMF の中央値からの改善は 1.42 dB である。この結果は、Log-Cos 型 PSNMF と Cos 型 PSNMF のハイパー

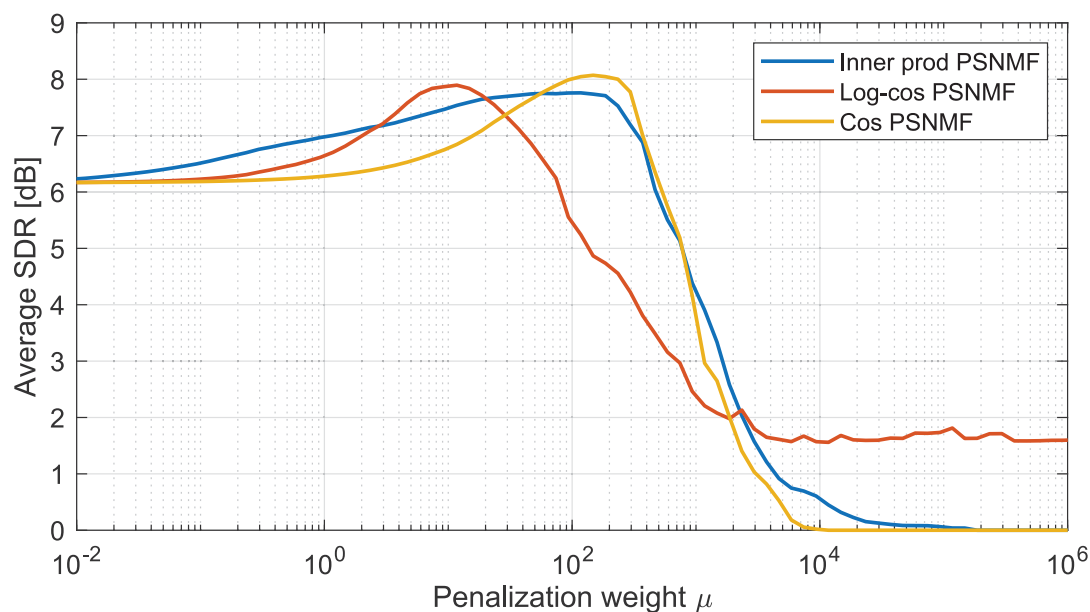


Fig. 6.3: Average SDR behaviors of development dataset. Weight coefficient μ that provides highest average SDR in development dataset is used for performance comparison with test dataset.

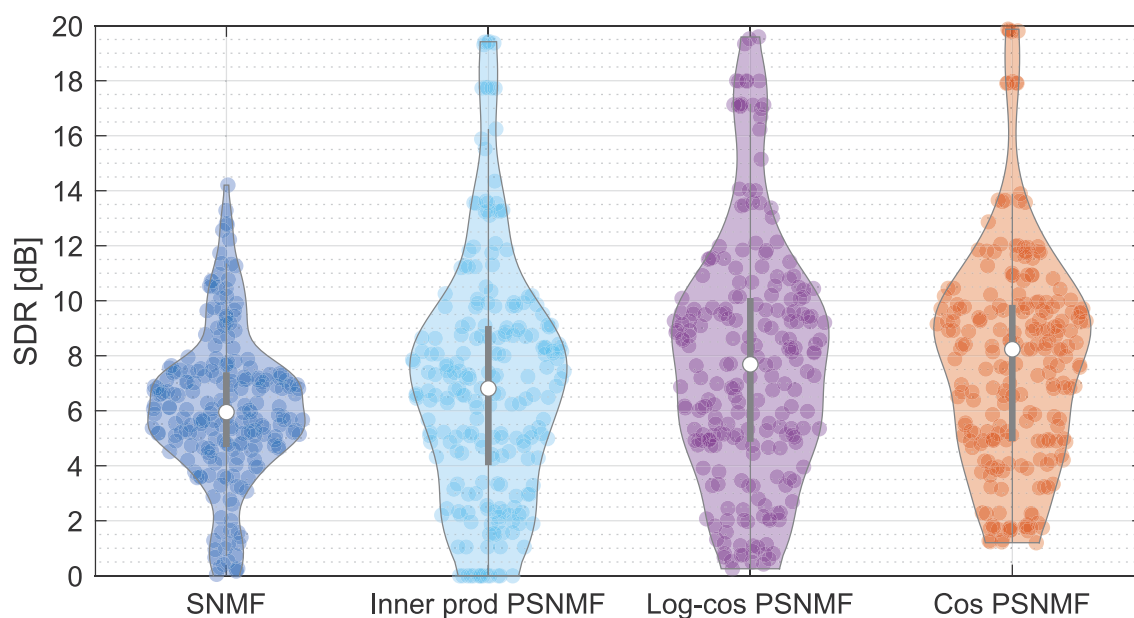


Fig. 6.4: Violin plot of 45 SDR results in test dataset. In each method, white circle indicates median value, gray vertical line shows range of 25–75 percentiles, and violin-shape curve is estimated distributions.

パラメータ μ の最適な設定が、楽器の種類に強く依存しないことを示している。

Table 6.1 Average and median values for test dataset

Method	Average [dB]	Median [dB]
SNMF	6.07	5.95
Inner prod PSNMF	7.01	6.81
Log-Cos PSNMF	7.75	7.69
Cos PSNMF	7.82	8.23

6.4 統計的検定

Table 6.1 の SDR の差の妥当性を評価するために、ペアワイズの Welch の片側 t 検定 [22] とペアワイズの Brunner–Munzel (BM) 検定 [23] の 2 種類の統計的検定を行った。片側 Welch の片側 t 検定と BM 検定の帰無仮説はそれぞれ以下の通りである。

- 分散の異なる正規分布に従う 2 つの母集団 A 及び B を考えるとき、各母集団の標本の平均値 μ_A 及び μ_B は、 $\mu_A \geq \mu_B$ を満たす。
- 2 つの母集団 A 及び B からそれぞれ標本 s_A と s_B を無作為に抽出するとき、 $s_A \geq s_B$ である確率と $s_A < s_B$ である確率が等しい。

Welch の片側 t 検定では平均を考えるため、正規分布を仮定する検定である。一方、BM 検定は観測データの正規分布性を仮定する必要がない検定である。

それぞれの検定で得られた p 値を Tables 6.2 及び 6.3 にそれぞれ示す。 p 値が 0.05 であれば、前述の帰無仮説を 95% の確率で有意に棄却できる。そのため、 p 値が低いほど、帰無仮説は 2 手法で明確な差がある、という結論を導ける。例えば、Log-Cos 型 PSNMF と Cos 型 PSNMF を比較した場合の p 値は、Welch の片側 t 検定で 42.8965 %、BM 検定で 59.7709 % となり、2 つの手法の差が有意ではないという結果になった。一方、内積型 PSNMF と Cos 型 PSNMF を比較した場合の p 値は、Welch の片側 t 検定で 2.06 %、BM 検定で 0.98 % と、十分に低い値となっている。すなわち、両検定において、統計的に有意な確率で帰無仮説を棄却でき、2 手法の間で差が有意であるといえる。

6.5 本章のまとめ

本章では、比較実験を通して提案手法である Cos 型 PSNMF 及び Log-Cos 型 PSNMF が内積型 PSNMF よりも音源分離性能が高いことを示した。さらにその結果を用いて行った 2 種類の統計的検定より、Cos 型 PSNMF の音源分離性能と内積型 PSNMF の音源分離性能との間に明確な差があることを示した。

Table 6.2 p values obtained by pairwise one-sided Welch's t test

Method of group A	Method of group B	p value [%]
SNMF	Inner prod PSNMF	0.3389
SNMF	Log-Cos PSNMF	0.0001
SNMF	Cos PSNMF	0.0000
Inner prod PSNMF	Log-Cos PSNMF	3.9434
Inner prod PSNMF	Cos PSNMF	2.0614
Log-Cos PSNMF	Cos PSNMF	42.8965

Table 6.3 p values obtained by pairwise BM test

Method of group A	Method of group B	p value [%]
SNMF	Inner prod PSNMF	3.1576
SNMF	Log-Cos PSNMF	0.0056
SNMF	Cos PSNMF	0.0000
Inner prod PSNMF	Log-Cos PSNMF	5.9094
Inner prod PSNMF	Cos PSNMF	0.9811
Log-Cos PSNMF	Cos PSNMF	59.7709

第 7 章

結言

本論文では、内積型 PSNMF の持つ問題であるスケール不定性の影響を受けない新たな罰則条件付き SNMF として、対数コサイン類似度を罰則項として SNMF に付与する Log-cos 型 PSNMF 及び純粋なコサイン類似度を SNMF に罰則項として SNMF に付与する Cos 型 PSNMF の 2 手法を新たに提案した。Log-Cos 型 PSNMF では、対数を用いることで計算が簡略化でき実装しやすいといったメリットを含んでいたが、数値的不安定性が生じるといった問題が新たに発生した。Cos 型 PSNMF では、内積型 PSNMF 及び Log-Cos 型 PSNMF と比較すると実装が複雑になるが、両手法にみられる問題を同時に解決できるため理論的に正しいアルゴリズムであるといえる。このことは性能調査実験より、Cos 型 PSNMF を用いた音源分離は既存手法である内積型 PSNMF や Log-Cos 型 PSNMF を用いた音源分離よりも高いという結果を得られたことから裏付けられた。また、Log-Cos 型 PSNMF と内積型 PSNMF の分離性能を比較したとき、Log-Cos 型が少し上回っていることが確認できた。さらに、この差を客観的に結論付けるために、統計的検定を 2 種類行った。提案手法のうち Cos 型 PSNMF は、従来手法の内積型 PSNMF を統計的に十分有意な確率で棄却でき、2 手法の間で差が有意であると結論付けることができた。今後の展望としては、提案手法の両手法とも高い SDR を示すハイパーパラメータの値が狭い範囲に限られているため、最適なハイパーパラメータを観測信号から適応的に算出するアルゴリズムの検討等が必要となることが予想される。本論文では、SNMF のみを対象としたが、教師無しの単純な NMF への活用も可能であることから、本論で提案した手法が広く用いられることが期待される。

謝辞

本論文は、香川高等専門学校電気情報工学科北村研究室にて行われた研究に基づくものです。

まず、本研究を進めるにあたり、ご多忙のところ熱心にご指導くださいました指導教員の北村大地講師に心より感謝申し上げます。北村大地講師には、論文執筆や研究に関する議論など、細部にわたるまで丁寧にご指導いただきました。

本論の副査である柿元健准教授と雛元洋一助教には、論文の構成や記述に関して大変有益な助言を頂き、大変お世話になりました。ここに厚く御礼申し上げます。

また、北村研究室同期の渡辺氏・大藪氏・梶谷氏には、ゼミや日頃のディスカッションのほか、3年に亘る研究室生活を様々な面で支えていただきました。ここに感謝申し上げます。

最後になりますが、現在に至るまで私の学生生活を金銭的に支え、暖かく見守って下さった両親には感謝の念に堪えません。これまで本当にありがとうございました。

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, no. 6755, pp. 788–791, 1999.
- [2] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. NIPS*, pp. 556–562, 2000.
- [3] 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,” in *Proc. ICASSP*, pp. 5365–5368, 2012,
- [4] P. Smaragdis, B. Raj, and M. Shashanka, “Supervised and semi-supervised separation of sounds from single-channel mixtures,” in *Proc. LVA/ICA*, pp. 414–421, 2007.
- [5] 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.
- [6] E. M. Grais, M. U. Sen, and H. Erdogan, “Deep neural networks for signal channel source separation,” in *Proc. ICASSP*, pp. 3734–3738, 2014.
- [7] S. Uhlich, F. Giron, and Y. Mitsufuji, “Deep neural network based instrument extraction from music,” in *Proc. ICASSP*, pp. 2135–2139, 2015.
- [8] T. Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 3, pp. 1066–1074, 2007.
- [9] A. Ozerov, C. Févotte, and M. Charbit, “Factorial scaled hidden Markov model for polyphonic audio representation and source separation,” in *Proc. WASPAA*, pp. 121–124, 2009.
- [10] 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. Audio, Speech, and Language Process.*, vol. 23, no. 4, pp. 654–669, 2015.
- [11] 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. Audio, Speech, and Language Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.
- [12] 蓮本恭介, 後藤正幸, 雲居玄道, “非負値行列因子分解による顧客購買パターン抽出と顧客生涯価値予想,” 情報処理学会第 80 回全国大会講演論文集, pp. 151–152, no. 1, 2018.
- [13] D. FitzGerald, M. Cranitch, and E. Coyle, “On the use of the beta divergence for musical source separation,” in *Proc. Irish Signals Syst. Conf.*, 2009.
- [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, 2009.
- [15] D. R. Hunter and K. Lange, “Quantile regression via an MM algorithm,” *J. Comput. Graph. Stat.*, vol. 9, no. 1, pp. 60–77, 2000.
- [16] J. Le Roux, F. Weninger, and J. R. Hershey, “Sparse NMF - half-baked or well done?,” *Mitsubishi Electric Research Lab. Technical Report*, TR2015-023, 2015.
- [17] 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, 2009.
- [18] A. Liutkus, D. Fitzgerald, and R. Badeau, “Cauchy nonnegative matrix factorization,” in *Proc. WASPAA*, 2015.
- [19] N. Takahashi and M. Seki, “Multiplicative update for a class of constrained optimization problems related to NMF and its global convergence,” in *Proc. EUSIPCO*, pp. 438–442, 2016.
- [20] D. Kitamura, “Open dataset: songKitamura” http://d-kitamura.net/dataset_en.html, accessed Jan. 3. 2022.
- [21] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech, and Language Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.
- [22] B. L. Welch, “The generalization of ‘Student’s’ problem when several different population variances are involved,” *Biometrika*, vol. 34, no. 1/2, pp. 28–35, 1947.
- [23] E. Brunner and U. Munzel, “The nonparametric Behrens-Fisher problem: Asymptotic theory and a small-sample approximation,” *Biometrical Journal*, vol. 42, pp. 17–25, 2000.
- [24] H. Nakajima, D. Kitamura, N. Takamune, H. Saruwatari, and N. Ono, “Bilevel optimization using stationary point of lower-level objective function for discriminative basis learning in nonnegative matrix factorization,” *IEEE Signal Process. Letters*, vol. 26, no. 6, 2019.

発表文献一覧

原著論文

1. Yuta Iwase and Daichi Kitamura, “Supervised audio source separation based on nonnegative matrix factorization with cosine similarity penalty,” IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, vol. E105-A, no. 6, Jun 2022 (in press).

国内学会

1. 岩瀬佑太, 北村大地, “コサイン類似度罰則条件付き半教師あり非負値行列因子分解,” 日本音響学会 2020 年春季研究発表会講演論文集, pp. 425–428, 埼玉, 2020 年 3 月（査読無）.
2. 岩瀬佑太, 北村大地, “コサイン類似度罰則条件付き非負値行列因子分解に基づく音源分離の実験的評価,” 日本音響学会 2021 年秋季研究発表会講演論文集, 2-1P-4, pp. 287–290, オンライン, 2021 年 9 月（査読無）.
3. 岩瀬佑太, 北村大地, “コサイン類似度罰則条件付き非負値行列因子分解に基づく音源分離の仮説検定,” 24 回 日本音響学会関西支部 若手研究者交流研究発表会 p. 33, オンライン, 2021 年 12 月（査読無）.

付録 A

補助関数法で利用される不等式

A.1 接線不等式

補題 A.1. (接線不等式) $f(x)$ が凹関数であるとき, 以下の不等式が成立する.

$$f(x) \leq f'(\bar{x})(x - \bar{x}) + f(\bar{x}) \quad (\text{A.1})$$

不等式中の等号が成立するための条件は $x = \bar{x}$ である.

A.2 Jensen の不等式

補題 A.2. (Jensen の不等式) $\alpha_i > 0$ を, $\sum_i \alpha_i = 1$ を満たす補助変数とする. 関数 $f(x)$ が凸関数であるとき, x_i ($i = 1, \dots, I$) に対して以下の不等式が成立する.

$$f\left(\sum_{i=1}^I \alpha_i x_i\right) \leq \sum_{i=1}^I \alpha_i f(x_i) \quad (\text{A.2})$$

$f(x)$ が狭義凸関数であるとき, 不等式中の等号が成立するための条件は $x_1 = \dots = x_i = \dots = x_I$ である.