修士論文

NMF による教師あり音声分離のための識別的学習法の改善とその評価尺度の提案

指導教員　峯松 信明　教授

東京大学 大学院工学系研究科 電気系工学専攻

37-196473 紺野瑛介

2021 年 1 月 28 日

概要

本論文ではモノラル音声分離に取り組む。

音声分離の最も基本的な手法として非負値行列因子分解（Nonnegative Matrix Factorization; NMF）によるものがある。NMF により各話者の基底を事前に学習しておき、混合音声が与えられたときは、これらの基底を固定して対応するアクティベーションのみを推定することで、音声分離が実現できる。

本論文の成果は、NMF による基底学習法の 1 つである識別的 NMF と呼ばれる手法に関するものであり、次の 2 つからなる。

第一の成果は識別的 NMF の最適化手法の改善に関するものである。識別的 NMF は解くのが困難な二段階最適化問題として知られ、先行研究では問題の一部を等式制約に置き換えることで、ペナルティ法により一段階化して解く手法が提案されていた。しかしこの等式制約の導出は不自然な仮定に基づくものであり、またペナルティ法は最適化が困難なことが知られている。そこで本論文では最適化問題の最適性条件に基づく、理論的により自然な新しい等式制約を提案し、さらにペナルティ法の代わりに拡張ラグランジュ関数法を用いることで最適化の高速化を行なった。拡張ラグランジュ関数法による識別的 NMF では、ペナルティ法よりも等式制約が高速に収束することが確認された。また、モノラル音声分離実験の結果、提案された等式制約を用いた方がより良い分離性能を示すことが確認された。

第二の成果は、様々な基底学習法により得られる各音源基底間の識別性を定量的に測るオーバーラップ尺度の提案である。識別的 NMF により学習された基底は、他の手法と比べてどのような性質を持っているか、またその「識別性」とは何かということは今まで定量的に議論されてこなかった。もしこれを定量的に評価できる指標が得られれば、識別的 NMF の仕組みや性質を他の手法と比較して議論できる。そこで本研究ではこれを測るオーバーラップ尺度を提案し、その基準として各音源の真の基底を用いることを提案する。真の基底は一般の基底学習法では得られないが、近年、最小体積 NMF という手法により近似的に得られることが明らかになった。そのため、最小体積 NMF により学習される基底を基準として用いる。モノラル音声分離実験の結果、識別的 NMF は単純な NMF よりも、各音源基底間のオーバーラップを小さくするという意味で「識別的」な基底を学習し、これにより分離性能を向上させていると考えられることがわかった。一方で、基準である最小体積 NMF の方が識別的 NMF よりも更にオーバーラップが小さく、かつ分離性能も高いことがわかり、最小体積 NMF の方が実はより「識別的」で、音源分離のための基底学習法として優れている可能性が示唆された。

目次

第 1 章

序論

1.1 本論文の背景：NMF による音源分離

人間は、たくさんの人々が雑談している喫茶店や、車が走行音を立てて行き交う通りの中にいても、自分が注意を向けている対象が発する音を、自然と聞き取ることができる。（ただし、聴覚障害や聴覚過敏などを持つ人を除く。）この能力は**選択的聴取**（selective listening）や**カクテルパーティ効果**（cocktail-party effect）などと呼ばれ、認知科学や心理学の分野において古くから研究され続けている [3]。

このカクテルパーティ効果を、計算機によって実現しようとするのが**音源分離**である。

音源分離は幅広い分野での応用を持つ。例えば、会議の文字起こしシステムや AI スピーカーなどの音声認識器の前処理として、音源分離により目的話者の音声だけを取り出す研究がなされている [53]。また、聴覚障害があり選択的聴取ができない人のために、補聴器で音源分離機能を実現しようとする研究もある [48]。音信号以外でも、例えば脳波解析（ElectroEncephaloGram; EEG）の分野などでもブラインド信号源分離が活用されている [4]。このように音源分離はさまざまな分野で活用され、研究が蓄積され続けているが、依然として満足な性能を持つ決定版と言えるような手法はなく、チャレンジングな問題であり続けている。

音源分離の最も基本的でシンプルな手法として**非負値行列因子分解**（Nonnegative Matrix Factorization; NMF）[10, 49] によるものがある。NMF は音源モデリング手法の一つである。これは、音源信号を、その中で頻出するスペクトルパターンを集めた基底と呼ばれる行列と、それらスペクトルパターンの時間変化する振幅を表すアクティベーションと呼ばれる行列との積に分解する。各音源の学習データに対して NMF を適用し、各音源ごとに基底を事前に学習しておき（これを本論文では**基底学習**と呼ぶ）、混合信号が与えられたときはこれらの基底を固定して対応するアクティベーションのみを推定することで、音源分離が実現できる。これを**教師あり NMF**（supervised NMF）[39] という。

NMF はただの行列分解であり非常にシンプルではあるが、それゆえに音源モデリングの基礎的な手法として非常に広範に用いられている。例えば、NMF を様々な方法で拡張して、発展的な音源分離手法が多数考え出されている。上に述べた教師あり NMF は録音に用いられるマイクが 1 本だけの場合（シングルチャネル）の手法だが、これをマルチチャネルに拡張したマルチチャネル NMF（Multichannel NMF; MNMF）[36] や、マルチチャネルの代表的手法の一つである独立ベクトル分析（Independent Vector Analysis; IVA）[15, 19] と NMF とを組み合わせた独立低ランク行列分析（Independent Low-Rank Matrix Analysis; ILRMA）[20] などはその代表例で

ある［37］。また、行列からテンソルへと拡張した半正定値テンソル分解（Positive SemiDefinite Tensor Factorization; PSDTF）［51, 52］などもある。他にも、NMF の軽量さを活かして、膨大な学習データを必要とする深層学習と組み合わせて音源分離を行う研究もある［2］。音源分離以外にも、例えば音響符号化［34］や声質変換［42］など、音源信号の特徴量抽出を必要とするような場面で NMF はよく用いられる。

以上のように、NMF は音源分離を含む高度な音声音響信号処理技術のための重要なモジュールである。したがって、教師あり NMF による音源分離の性能を上げるために、NMF 自体の性能を向上させたり、その性質を調べたりすることは、音源分離だけにとどまらない正の波及効果を持つ。

1.2　本論文の目的：識別的 NMF の改善とその性質の解明

一般に音源分離は、分離対象の各音源の音響的性質が似通っているほど難しくなる。例えば、人の音声と楽器音の分離は比較的容易だが、音声同士や楽器音同士の分離は難しい。教師あり NMF により分離する場合を考えてみよう。学習データとして各音源のクリーン信号、すなわちその音源の音しか含まないようなデータを用いると、音源群が音響的に似ていた場合、得られる基底も音源間で似通ってしまう。ゆえに混合信号が与えられたときに各音源基底に対応するアクティベーションを推定するのが困難となり、分離性能が落ちてしまうことになる。

このような音源群に対する分離性能を向上させるために、**識別的 NMF**［50］という基底学習法が提案された。識別的 NMF では、学習データである各音源のクリーン信号から仮想的な混合信号を作り出して、これをうまく分離できるように各音源基底を学習する。こうすることで学習時の目的関数がテスト時の目的関数と一致することとなり、「識別的」な基底が学習できて分離性能が向上することが報告されている［29, 32, 40, 50］。

だが、識別的 NMF に関する先行研究には、次の 2 つの不満足な点がある。

第一に、先行研究では識別的 NMF が定義する最適化問題を、直接的かつ正当な方法で解いてはいない。識別的 NMF は、数学的には二段階最適化問題［38］として定式化される。二段階最適化問題とは、最適化問題の制約条件の中にさらに別の最適化問題が含まれるものであり、直接解くことは特殊なケースを除いて非常に困難であることが知られている。そのため、先行研究では様々な仮定を置いてこの問題をより簡単な形に書き換えて間接的に解いたり［29, 40, 50］、あるいはペナルティ法と呼ばれる最適化手法を用いて直接的に解く場合でも［32］、アルゴリズムの導出に理論的に不自然な仮定が含まれるなど、識別的 NMF の最適化の直接的かつ正当な方法は未だ提案されていない。

第二に、先行研究では識別的 NMF がどのような仕組みで「識別的」な基底を学習するか、さらには識別的 NMF が定義する基底間の「識別性」とは何かということについて、他の基底学習法と定量的に比較可能な議論がなされてこなかった。もし、ある基底学習法により得られる各音源基底に対し、それらの間の「識別性」を測る定量的な尺度を設計することができれば、識別的 NMF の仕組みや性質を、他の手法と比較して調べることができる。またそのような尺度は、識別的 NMF に限らず、新しく提案された別の基底学習法に対しても、その性質を調べる際に役立つだろう。

本研究の動機は以上の 2 点にある。我々は以降の章でこれらの問題に取り組んで行く。

1.3　本論文の構成

本論文の構成は以下の通りである。2 章では、以降の章で共通して必要な基礎事項を説明する。3 章は第一の成果である、識別的 NMF の最適化手法の改善について報告する。4 章は第二の成果である、基底間の「識別性」を定量的に測るオーバーラップ尺度の提案と、それを用いた識別的 NMF の評価について報告する。5 章では結論と今後の課題を述べる。

1.4　記号の注意

本論文では $\mathbf{1}$,J をそれぞれ全要素が 1 のベクトルと行列とする。サイズは周囲から明らかに分かる場合省略する。

第 2 章
準備

本章では、以降の 3 章と 4 章とで共通して必要な基礎事項を説明する。2.1 節では音源分離の舞台設定を説明する。2.2 節では音源分離を解くためのアプローチとして空間モデリングと音源モデリングという 2 つのアプローチを説明する。本論文で以降取るのは、この音源モデリングに基づくアプローチである。2.3 節では NMF による音源モデリングの動機付けとして、音源分離の主な対象である音声や楽器音の性質を紹介する。2.4 節では、NMF による音源モデリングの仕方を説明し、また、NMF の最適化問題としての定式化を行う。2.5 節では、NMF による音源モデリングを用いて音源分離をどう実現するかを説明する。2.6 節では NMF の最適化手法として上界最小化アルゴリズムを説明し、これによる NMF の具体的な更新式を述べる。

2.1　音源分離の定義

1.1 節で触れた音源分離について、その舞台設定を説明する。

音源分離のタスクは、複数の音源からの信号を複数の測定器で測定し、これを計算機的な処理によってそれぞれの信号に分離することである。もう少し詳しく述べよう。本論文では特に各音源が人の音声である場合を考えるから、この場合を例に取って説明する。N 人の話者がめいめいに喋っており、これを M 本のマイクで録音する。$M = 1$ の場合を**シングルチャネル**（single channel）、$M \geq 2$ の場合を**マルチチャネル**（multi-channel）という。また、$N > M$、すなわち話者数がマイクの本数より多い系を**劣決定系**（underdetermined system）といい、逆に $N \leq M$ の系を**優決定系**（overdetermined system）という。各音声はマイクに到達するまでに、遅延や干渉、反射などの物理的な過程を経るが、一般にこの過程の詳細は完全には分からない、つまり音の**混合系**（mixing system）は未知（ブラインド）だと仮定する。以上のような状況下での録音から、各々の音声を分離して取り出す技術が**音源分離**である。

混合系が未知であることを強調して**ブラインド音源分離**と呼ぶこともあるが、本論文では以降、ブラインドを省略して単に音源分離と呼び、上で考えたような各音源が音声である場合を特に**音声分離**（speech separation）と呼ぶことにする。

2.2　音源分離へのアプローチ：空間モデリングと音声モデリング

音源分離へのアプローチは大きく分けて 2 つある。**空間モデリング**（spacial modelling）と**音源モデリング**（source modelling）である。ほとんどの音源分離手法はこの 2 つの組み合わせによって作られている。

▶ 空間モデリング

マルチチャネル、すなわち録音するマイクが 2 本以上ある場合には、各マイクへの音の到達時間差を利用して音源群の空間的配置を推定することができ、音源分離に活用できる。これが空間モデリングである。

典型的なのは、音源信号がマイクに到達するまでの伝達系の周波数特性（振幅や位相の変化）を複素ベクトルで表した**ステアリングベクトル**（steering vector）を用いて**ビームフォーマ**（beamformer）を形成し、音源方向の音を強調することで分離する手法である。理想的なビームフォーマとは何かという考え方の違いによっていろいろ変種はあるが、代表的でよく用いられているものに**最小分散無歪応答ビームフォーマ**（Minimum Variance and Distortionless Response beamformer; MVDR beamformer）［14］や**最大信号対雑音比ビームフォーマ**（Maximum Signal-to-Noise Ratio beamformer; Max SNR beamformer）［1］などがある。

近年では、シングルチャネルの場合でもニューラルネットを用いてステアリングベクトルを推定することにより、仮想的にビームフォーマを行える手法が提案されるなど［13]、空間モデリングのアプローチは現在でも盛んに研究されている。

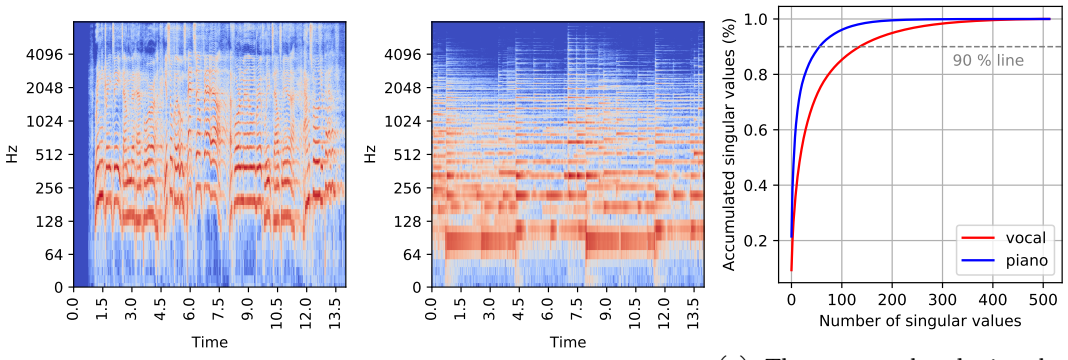
▶ 音源モデリング

もう一方の音源モデリングは、空間ではなく、音源の性質を数学的・統計的にモデリングするアプローチである。これは空間モデリングとは独立して行うことができる。マルチチャネルの場合は空間モデリングと組み合わせて用いられ、音源モデリングのやり方次第によっては分離性能を大きく向上させることができる［37]。本論文ではマルチチャネルではなくシングルチャネルの音源分離を扱うが、この場合には（先述のニューラルビームフォーマのような特例を除いて）音源モデリングのみによって音源分離を行わなければならないから、音源モデリングの良し悪しが分離の良し悪しに直結することになる。

音源分離における音源モデリングとして最も基礎的で、かつ現在でも広く用いられているものとして、2.4 節に述べる NMF によるものがある。

2.3　音源の性質

音源分離の主な対象は人の音声や楽器音である。次節で述べる NMF による音源モデリングのための動機付けとして、ここではこういった音源の性質に関する以下の 3 つの事実を紹介しよう。



(a) Vocal.　(b) Piano.　(c) The accumulated singular value percentages.

Fig. 2.1: Examples of audio spectrograms. (a, b) Amplitude spectrograms of vocal and piano. The audio data were obtained from SiSEC2010 [45]. (c) Accumulated singular value percentages of the amplitude spectrograms of vocal and piano. Less than 150 singular values account for 90 % of the total sum.

▶ 音の加法性

音は空間を伝播する波動であって、異なる音波の間には（音源分離で普通考えているような範囲では）**重ね合わせの理が成り立つ**。

例えばピアノでドとレを同時に鳴らしたときに観測される周波数スペクトルは、それぞれ別々に鳴らしたときの周波数スペクトルの和になって、ドとレに対応する 2 つのピークが現れる。あるいは複数の話者が同時に喋ったとき、その混合音声は個々の音声の和として計測される。ゆえに音信号を数学的モデルで表すならば、それは今述べたような音源内での足し合わせ（ピアノのドとレ）や音源間での足し合わせ（複数話者）を表せる加法的な構造を持つべきである。

▶ 音源の低ランク性

音声や楽器音は**低ランク構造**（low rank structure）**を持つ**。

音は本義的には 1 次元の時間信号である。だが普通その周波数スペクトルは時々刻々と変化する。この点に着目して、音声や楽器音などを処理する際には、**短時間フーリエ変換**（Short-Time Fourier Transform; STFT）を掛けて**スペクトログラム**（spectrogram）と呼ばれる時間周波数信号に変換することがよく行われる。スペクトログラムは複素数値の行列であり、離散周波数と離散時間の 2 つの軸を持つ。本論文では、離散周波数に関するインデックスを**周波数ビン**（frequency bin）と呼び、離散時間に関するインデックスを**時間フレーム**（time frame）と呼ぶことにする。またスペクトログラムの振幅成分（絶対値成分）を**振幅スペクトログラム**（amplitude spectrogram）、位相成分を**位相スペクトログラム**（phase spectrogram）という。

例えば男性音声とピアノの振幅スペクトログラムを見てみると（Fig. 2.1a, 2.1b）、似たようなスペクトルパターンの繰り返しが多いことがわかる。この傾向は特に楽器音であるピアノにおいて顕著である（Fig. 2.1b）。つまり、音声や楽器音の振幅スペクトログラムを行列として見てみると、これは低ランク構造を持ち（Fig. 2.1c）、周波数ビンや時間フレームの数よりも少ない本数のベクトルで表せるはずである。

▶ 位相よりも振幅が重要

人間の音の知覚においては、**音の位相情報よりも振幅情報のほうが重要である**ことが知られている［30, 47］。このことから、音源モデリングにおいては位相よりも振幅のモデリングの方を優先すべきであると言える。（実は、近年では位相情報も利用した音声音響信号処理が活発に議論されているが［31, 56］、本研究では位相情報は取り扱わないこととする。）音信号をスペクトログラムで表した場合を考えると、位相スペクトログラムは無視して振幅スペクトログラムだけモデリングすれば十分だということである。

2.4　NMF による音源モデリング

以上の（1）加法的なモデルであること（2）低ランク性を表せるモデルであること（3）振幅スペクトログラムだけを考慮したモデルでも良いことという 3 つの事情から、**非負値行列因子分解**（Nonnegative Matrix Factorization; NMF）による音源モデリングが音源分離において広く用いられている。本節では NMF を定義し、実際の音信号に対して適用した例を示す。そして NMF を最適化問題として定式化する。

▶ NMF の定義

NMF とは、非負の行列を 2 つの非負の行列の積に分解することである［24, 25, 35］。NMF によってどのように音源モデリングを行うかを説明しよう。音源の振幅スペクトログラムを $V = (\boldsymbol{v}_1 \ldots \boldsymbol{v}_T) \in \mathbb{R}_+^{F \times T}$ とする。F, T はそれぞれ周波数ビンと時間フレームの数であり、列ベクトル $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_T \in \mathbb{R}_+^F$ は各時刻における振幅スペクトルを表す。（今振幅を取っているので V は非負であることに注意。）$\boldsymbol{v}_1, \ldots, \boldsymbol{v}_T$ は、音声や楽器音が低ランクかつ加法的であることから、$K < \min\{F, T\}$ 本の振幅スペクトルパターン $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_K \in \mathbb{R}_+^F$ の非負結合によって

$$\boldsymbol{v}_t \approx \sum_{k=1}^{K} \boldsymbol{w}_k h_{kt} \quad (\forall t)$$

と近似できるはずである。これを行列形式で書けば、

$$V \approx WH$$

となる。このように振幅スペクトログラム V を V より低ランクな 2 つの非負行列 W, H の積でモデリングするのが NMF による音源モデリングである。

（なお厳密には音の加法性は、音が、元々の時間信号表現や、あるいはそれに線形な演算である（短時間）離散フーリエ変換をかけた複素スペクトログラムで表されたときにのみ成り立ち、振幅スペクトログラムに対しては近似的にしか成り立たない。つまり、$\mathcal{V}_1$, $\mathcal{V}_2$ を 2 つの音が別々に鳴ったときの複素スペクトログラムとし、$\mathcal{V}$ を 2 つの音が同時になったときの複素スペクトログラムとすると、

$$\mathcal{V} = \mathcal{V}_1 + \mathcal{V}_2, \quad |\mathcal{V}| \approx |\mathcal{V}_1| + |\mathcal{V}_2|$$

である。音の加法性を正確に取り扱うために、振幅スペクトログラムではなく複素スペクトログラ

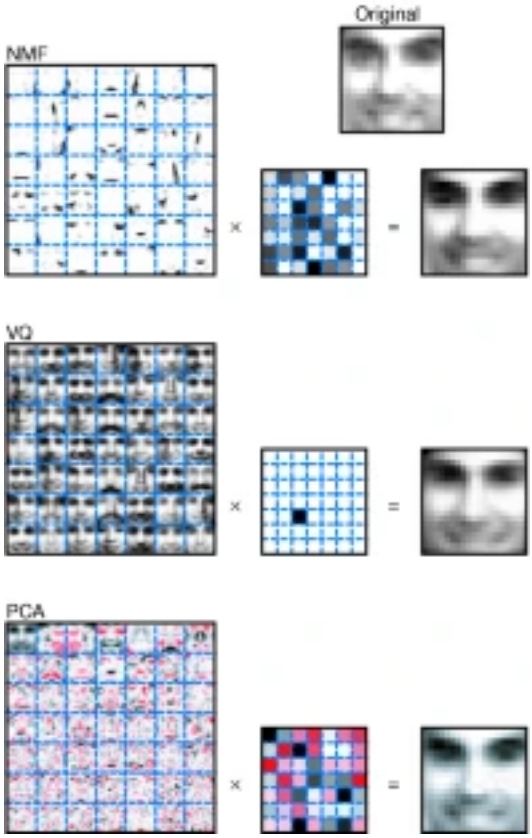
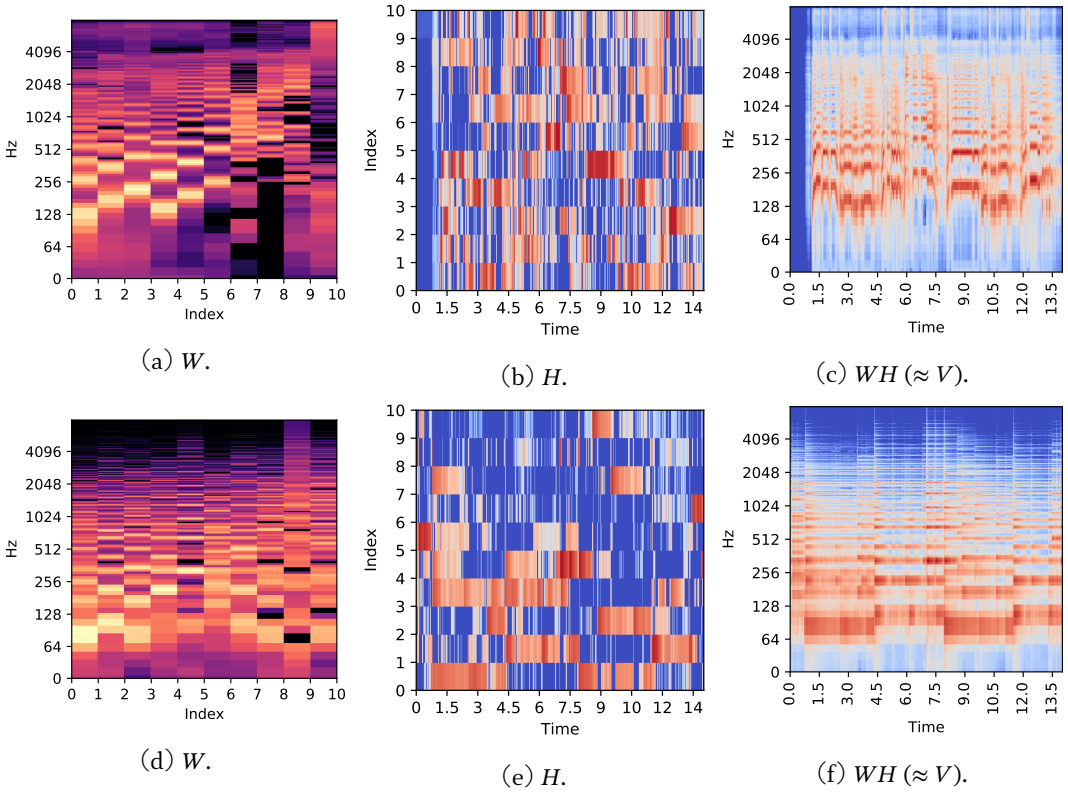


Fig. 2.2: Various matrix factorization methods applied to a face image. NMF learns parts-based and interpretable representation of faces, whereas VQ and PCA do not [24].

ムを直接分解する**複素 NMF**（complex NMF）[18] も提案されているが、これは本研究の対象外とする。）

$W = (\boldsymbol{w}_1 ... \boldsymbol{w}_K) \in \mathbb{R}_+^{F \times K}$ は**基底**（basis）と呼ばれ、各列ベクトル $\boldsymbol{w}_1, ..., \boldsymbol{w}_K \in \mathbb{R}_+^F$ は**基底ベクトル**（basis vector）と呼ばれる。$H = (\boldsymbol{h}_1; ...; \boldsymbol{h}_K) \in \mathbb{R}_+^{K \times T}$ は**アクティベーション**（activation）と呼ばれ（; はベクトルを縦方向に結合することを表す）、各行ベクトル $\boldsymbol{h}_1, ..., \boldsymbol{h}_K \in \mathbb{R}_+^{1 \times T}$ は対応する基底ベクトルの時間変化する重みを表す。厳密ではないが、人間の音声に例えるならば、W はその話者の発声した音素を集めたようなもので、H は各音素が発せられたタイミングのようなものである。楽器音に例えるならば、W はその楽器で弾いたドレミの各音で、H は各音が鳴ったタイミングのようなものである。

W, H の両方に非負性が課されているおかげで、NMF は解釈性の高い行列分解を得ることができる。これは、1999 年の Lee と Seung による画期的な論文 [24] によって初めて指摘された。彼らは、ピクセルを並び替えて 1 次元化した顔画像を並べて行列 V を作り、それに対して NMF を適用した（Fig. 2.2）。ベクトル量子化（Vector Quantization; VQ）や主成分分析（Principal Component Analysis; PCA）などの行列分解手法が人間の顔の全体的な特徴を学習してしまっているのに対して、NMF はパーツ毎の特徴を捉えた、解釈性の高い基底を学習できている。この論文の登場以降、NMF の研究は爆発的に進展することとなった [49]。



(a) W.　(b) H.　(c) WH ($\approx V$).

(d) W.　(e) H.　(f) WH ($\approx V$).

Fig. 2.3: (a, b, c) NMF of the amplitude spectrogram of the vocal. (d, e, f) NMF of the amplitude spectrogram of the piano.

▶ 音信号に対する NMF の適用例

Fig. 2.1 の男性音声とピアノに対して NMF を掛けた例を Fig. 2.3 に示す。ここでは結果が見えやすいように基底ベクトルの本数を $K = 10$ とした。Fig. 2.1c の示す通り、音声や楽器音に対しては、$K = 10$ は NMF で精度良く近似するためには流石に少なすぎるのだが、それでも元の振幅スペクトログラムをそれなりに再現できていることがわかる（Fig. 2.3c, 2.3f)。また、ピアノのアクティベーション（Fig. 2.3e）よりも音声のアクティベーション（Fig. 2.3b）の方が複雑であることなども見て取れる。これはピアノの振幅スペクトログラムよりも音声の振幅スペクトログラムの方がランクが高かった（Fig. 2.1c）ということを反映している。(なおここで用いた NMF は正確には 4 章で後述する最小体積 NMF というものである。)

▶ 最適化問題としての NMF の定式化

ところでこの NMF 型の分解は、

$$V = WH$$

というふうに厳密（exact）に解くことはできない。もし W, H に非負制約がなければこれは特異値分解により解くことができる。しかし非負制約のある NMF の場合、厳密に解くことは NP 困難でありとても難しいことが明らかにされている [43]。

そのため NMF は、通常、厳密に解くことは諦めて

$$\begin{aligned} &\min_{W,H} D(V \mid WH) \\ &\text{s.t. } W, H \geq 0 \end{aligned} \tag{2.1}$$

という最適化問題の形で近似的に解く。ここで D は V と推定 WH との間の誤差を測る距離尺度である。

一般には NMF の目的関数は距離尺度 D のみから成るわけではなく、W や H にどのような性質を求めるかに応じて様々な正則化項が付け加えられる。例えばスパース NMF（sparse NMF）[5, 16, 23] ではアクティベーション H をスパースにするために正則化項 $\|H\|_1$（行列に対する L_1 ノルム）を用いる。当然、目的関数だけでなく制約条件にもバリエーションがある。例えば 4 章では、各基底ベクトル $\boldsymbol{w}_1, ..., \boldsymbol{w}_K$ が確率単体上にある（$\mathbf{1}^\top \boldsymbol{w}_k = \mathbf{1}^\top$ $(\forall k)$）という制約が置かれる。以上の他にも、目的関数や制約条件を改変して無数の NMF の変種が提案されているが、本論文ではそれらはほとんど取り上げない。詳しくはレビュー [10, 49] を参照されたい。

2.5　NMF による音源分離

NMF を二段階で用いることにより、シングルチャネルで音源分離ができる [39]。

▶ plain NMF による基底学習

まず、学習データとして、分離対象である個々の音源 $n = 1, ..., N$ の、その音源の音しか含まれていないような振幅スペクトログラム $V_1, ..., V_N \in \mathbb{R}_+^{F \times T}$ を用意する。このようなデータを**クリーン信号**（clean signal）という。特に音声の場合は**クリーン音声**（clean speech）という。なお $V_1, ..., V_N$ のそれぞれの時間フレーム数は違っていても構わないのだが、ここでは簡単のため全ての音源 $n = 1, ..., N$ で共通で T とした。（必要ならば信号の前後を適当に 0 で埋めることで揃えられる。）

これらの $V_1, ..., V_N$ に対して NMF を行うことにより、各音源の基底 $W_1, ..., W_N$ を分離の事前に学習しておく。前節で述べたように NMF には様々なバリエーションがあるが、ここでは基底学習法として最も単純で、かつよく用いられる次の NMF を導入しておく。

plain NMF

$$\min_{W,H} D_\beta(V \mid WH) \tag{2.2a}$$

$$\text{s.t. } W \in \mathbb{R}_+^{F \times K},\ H \in \mathbb{R}_+^{K \times T} \tag{2.2b}$$

D_β は 2.6.1 節で後述する β ダイバージェンスという距離尺度である。本論文では (2.2) による基底学習を特に **plain NMF** と呼ぶことにする。なお plain NMF の具体的な更新式については 2.6.3 節で述べる。

▶ 教師あり NMF による音源分離

混合信号の振幅スペクトログラム $X \in \mathbb{R}_+^{F \times T}$ が与えられたら（再び T は学習データ $V_1, ..., V_N$ と違っていても構わないのだが簡単のため同じとした）、plain NMF (2.2) により得られた各音源基

底 $W_1, ..., W_N$ を固定して X に対して教師あり NMF

教師あり NMF

$$\min_{H_1,...,H_N} D_\beta\left(X \mid \sum_n W_n H_n\right) \tag{2.3a}$$

$$\text{s.t. } H_1, ..., H_N \in \mathbb{R}_+^{K \times T} \tag{2.3b}$$

を行って、対応するアクティベーション $H_1, ..., H_N$ のみを推定する。教師あり NMF の具体的な更新式については 2.6.4 節で述べる。

推定された $H_1, ..., H_N$ と学習済みの $W_1, ..., W_N$ を掛け合わせることで、X 中の各音源の成分を

$$\hat{V}_n = W_n H_n \quad (\forall n)$$

と推定できる。$\hat{V}_n$ に音源 n の位相スペクトログラムを掛け合わせ逆 STFT を行うことで、音源 n の信号が得られる。

位相の推定は難しいので、与えられた混合信号の位相スペクトログラムをそのまま使うことが多い。本研究でもこの方針を取る。（だが 2.3 節でも述べたように近年では位相情報も積極的に取り扱っていこうと流れがある [31, 56]。直近でも上述の β ダイバージェンスを用いたときに位相情報を推定する研究 [44] が出ているが、本研究ではこの位相スペクトログラム推定の問題は対象外とする。）

2.6　NMF の最適化

本節ではまず、NMF によく用いられる距離尺度である β ダイバージェンスを紹介し、NMF の最適化手法として、上界最小化と呼ばれるアルゴリズムを説明する。そして plain NMF と教師あり NMF に対して上界最小化による更新式を示す。

2.6.1　β ダイバージェンス

NMF による音源モデリング (2.1) で、分離対象の V と推定 WH との間の距離尺度 D として最もよく用いられているのが、plain NMF (2.2) でも使われた **β ダイバージェンス**（β divergence）である [6]。以下にその定義と性質を述べよう。

▶ β ダイバージェンスの定義

まず、2 つのスカラー $x, y \in \mathbb{R}$ の間の β ダイバージェンスは

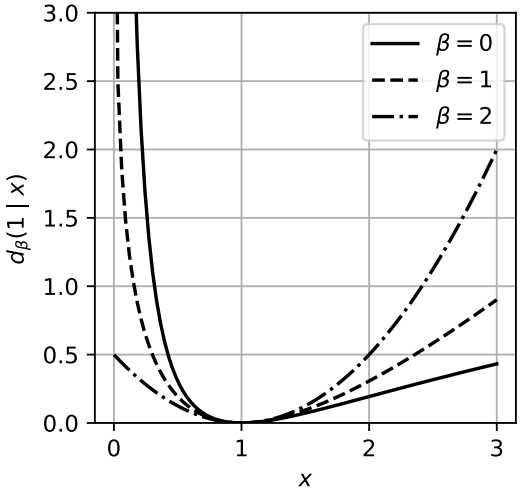


Fig. 2.4: $d_\beta(1 \mid x)$ for $\beta = 0, 1, 2$.

β ダイバージェンス

$$d_\beta(y \mid x) \coloneqq \begin{cases} \dfrac{y^\beta}{\beta(\beta-1)} + \dfrac{x^\beta}{\beta} - \dfrac{yx^{\beta-1}}{\beta-1} & (\beta \in \mathbb{R} \setminus \{0, 1\}) \\ y/x - \log(y/x) - 1 & (\beta = 0) \\ x - y + y\log(y/x) & (\beta = 1) \end{cases} \tag{2.4}$$

で定義される。これは音響信号処理の分野で良く用いられる種々の距離尺度を統合したものであり、$\beta = 0$ のときは**板倉齋藤ダイバージェンス**（Itakura-Saito divergence）、$\beta = 1$ のときは**一般化カルバック・ライブラーダイバージェンス**（generalized Kullback-Leibler divergence）、$\beta = 2$ のときは**ユークリッド距離**（Euclidean distance）となる（Fig. 2.4）。一般に $\forall x,\ y,\ \beta$ に対して

$$d_\beta(y \mid x) \geq 0 \quad (= 0 \text{ if and only if } a = b)$$

が成り立つが、$\beta = 2$ の場合を除いて d_β は距離の 3 条件を満たさないため、数学的には厳密には距離でない。2 つの同じサイズの行列 $X, Y \in \mathbb{R}^{I \times J}$ に対する β ダイバージェンスは、行列要素毎の β ダイバージェンスの和として、

$$D_\beta(Y \mid X) \coloneqq \sum_{ij} d_\beta(y_{ij} \mid x_{ij})$$

で定義される。

▶ ベータダイバージェンスの性質

β の値が変わると、d_β のスケール依存性が変わる。$\sigma > 0$ をスケーリングパラメータとして、

$$d_\beta(\sigma y \mid \sigma x) = \sigma^\beta d_\beta(y \mid x)$$

が成り立つ。例えば $\beta = 0$（板倉齋藤ダイバージェンス）のときは、$d_\beta(\sigma y \mid \sigma x) = d_\beta(y \mid x)$ だから、振幅の小さい音も大きい音も同じ比重で扱うことになる。そのため、例えば楽器のアタック音などの微小な音も重視したい場合に有効である [7]。一方で $\beta = 1, 2$ のときは振幅の大きい音の方が比重が大きくなり、微小なノイズなどを無視してピークに注目したい場合に有効である。

最適化の観点からもいくつか述べておこう。$\beta \leq 1$ では

$$\lim_{x \to 0} d_\beta(y \mid x) = +\infty$$

となるため、最適化の際に自然と x が正の方に誘導される。一方で $\beta > 1$ では $d_\beta(y \mid 0)$ が有限となるため（Fig. 2.4）、$x < 0$ となりうるので、x を非負にするためにステップサイズを調節するなどの工夫が必要となる（付録 A.3）。また、$1 \leq \beta \leq 2$ では $d_\beta(y \mid x)$ は x に関して凸であり最適化がしやすくなる。

2.6.2　上界最小化による NMF の最適化

以降の章では、様々な NMF 手法の最適化問題を解くために、上界最小化アルゴリズムと呼ばれる手法が鍵となる。そこでこの手法の基本的な考え方を説明しておこう。特に、plain NMF（2.2）の場合を例に取る。

▶ separable でない目的関数

plain NMF （2.2）は非凸最適化問題であり、W, H を一気に更新していくことは非常に難しい。そのため、

$$W \leftarrow \arg\min_W D_\beta(V \mid WH)$$
$$H \leftarrow \arg\min_H D_\beta(V \mid WH)$$

というように W に関して最適化、次に H に関して最適化、と交互に最適化していく**ブロック座標降下法**（Block Coordinate Descent; BCD）が基本的な戦略となる。

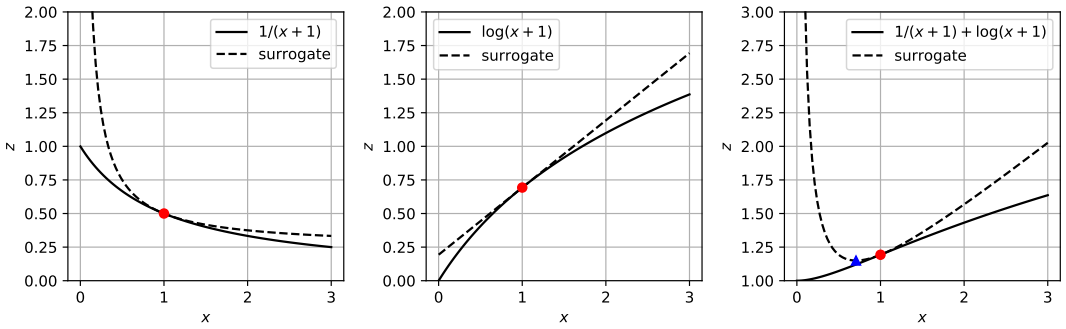
BCD を使って上のように分解すると問題は格段に簡単になるが、依然として、個々の最適化問題を解析的に解くことは難しい。例えば板倉齋藤ダイバージェンスを使ったとき目的関数がどうなるかを考えてみよう。目的関数は

$$\begin{aligned} D_{\mathrm{IS}}(V \mid WH) &= \sum_{ft} d_{\mathrm{IS}}(v_{ft} \mid \boldsymbol{w}_f \boldsymbol{h}_t) \\ &= \sum_{ft} d_{\mathrm{IS}}\left(v_{ft} \mid \sum_k w_{fk} h_{kt}\right) \\ &= \sum_{ft} \left(\frac{v_{ft}}{\sum_k w_{fk} h_{kt}} - \log \frac{v_{ft}}{\sum_k w_{fk} h_{kt}} - 1 \right) \end{aligned}$$

となる。W_n について極値を取る点を探すために例えば w_{fk} で偏微分してみると、最右辺の第 1 項と第 2 項から w_{fk} とそれ以外の成分 $w_{fk'}$ $(k' \neq k)$ が複雑に絡み合った式が出てくる。そのため、w_{fk} による偏微分 $= 0$ という式を $\forall f, k$ に対して同時に解く（つまり連立方程式を解く）のは解析的にはできない。このように変数同士が絡み合っていて、極値を取る点が解析的に求まらないような目的関数を **separable** でない目的関数という。

▶ 上界最小化アルゴリズム

separable でない目的関数を最小化するための常套手段が**上界最小化アルゴリズム**（majorization minimization algorithm）である。今適当な目的関数を考えて、それを $f(x)$ とする。最適化のあるステップ i において $x = x^{(i)}$ だったとしよう。$f(x)$ に対して



(a) Majorization of $1/(x+1)$.　(b) Majorization of $\log(x+1)$.　(c) Majorization of the sum.

Fig. 2.5: Majorization of $1/(x+y)+\log(x+y)$. Surrogate functions were constructed at $(x,y)=(1,1)$ (red circle). Only the xz-plane with $y=1$ is shown. (c) The surrogate function takes its minimum at $x=1/\sqrt{2}$ when $y=1$ (blue triangle).

代理関数

$$g(x \mid x^{(i)}) \geq f(x) + c \quad (\forall x)$$
$$(c \coloneqq g(x^{(i)} \mid x^{(i)}) - f(x^{(i)}))$$

と上から抑えるような関数 g を $x^{(i)}$ における f の**代理関数**（surrogate function）という。最適化の次の解候補点として g を最小化する点

$$x^{(i+1)} = \arg\min_{x} g(x \mid x^{(i)})$$

を選ぶと、代理関数の定義と $x^{(i+1)}$ の選び方から

$$f(x^{(i+1)}) \quad \leq \quad g(x^{(i+1)} \mid x^{(i)}) - c \quad \leq \quad g(x^{(i)} \mid x^{(i)}) - c \quad = \quad f(x^{(i)})$$

が成り立ち、単調非増加な解候補点列 $(x^{(i)})_{i=0,1,2,\dots}$ を生成できることがわかる。g として特に凸で separable なものを設計できれば、解析的に解候補点列を求めることができて、元の目的関数を間接的に最小化していくことができる。

▶ 具体例

上で板倉齋藤ダイバージェンスの例を考えたから、上界最小化アルゴリズムの具体例として

$$f(x,y) = \frac{1}{x+y} + \log(x+y)$$

という関数を最小化してみよう。ブロック座標降下法を用いることを考えて、x に関して最小化してみる。最適化のあるステップにおける解候補点が $(x',y')=(1,1)$ だったとして、この点における代理関数を設計し、次の解候補点を探す。まず第 1 項に対しては、凸関数に対する Jensen の不等式を用いることにより

$$\frac{1}{x+y} \leq \left(\frac{x'}{x'+y'}\right)^2 \frac{1}{x} + \left(\frac{y'}{x'+y'}\right)^2 \frac{1}{y}$$

と抑えられる。$(x',y')=(1,1)$ を代入し、$y=1$ の平面で見てみると

$$\frac{1}{x+1} \leq \frac{1}{4x} + \frac{1}{4}$$

となる（Fig. 2.5a）。次に第 2 項に対しては、正の対数関数は上に凸であるから接線不等式を立てることで

$$\log(x+y) \le \log(x'+y') + \frac{1}{x'+y'}(x+y-x'-y')$$

と抑えられる。$(x', y') = (1, 1)$ を代入し、$y = 1$ の平面で見てみると

$$\log(x+1) \le \log 2 + \frac{x-1}{2}$$

となる（Fig. 2.5b）。第 1 項と第 2 項をまとめると、全体として

$$f(x,y) = \frac{1}{x+y} + \log(x+y) \le \left(\frac{x'}{x'+y'}\right)^2 \frac{1}{x} + \left(\frac{y'}{x'+y'}\right)^2 \frac{1}{y} + \log(x'+y') + \frac{1}{x'+y'}(x+y-x'-y')$$

と抑えられる。$(x', y') = (1, 1)$ を代入し、$y = 1$ の平面で見てみると

$$\frac{1}{x+1} + \log(x+1) \le \frac{1}{4x} + \frac{1}{4} + \log 2 + \frac{x-1}{2} \tag{2.5}$$

となる（Fig. 2.5c）。$y = 1$ を固定して (2.5) を x に関して最小化しよう。x で右辺を偏微分して $= 0$ とおくと

$$0 = \frac{\partial}{\partial x}\left(\frac{1}{4x} + \frac{1}{4} + \log 2 + \frac{x-1}{2}\right) = -\frac{1}{4x^2} + \frac{1}{2} \quad \therefore \quad x = \frac{1}{\sqrt{2}}$$

となる。（NMF なので非負の値を取った。）これが x の次の解候補点となる。

2.6.3　plain NMF の更新式

上界最小化アルゴリズムによって plain NMF (2.2) に対する更新式が導出できる [33]。詳しい導出は省くが、基底 W とアクティベーション H の更新式は

plain NMF の基底とアクティベーションの更新式

$$W \leftarrow W \odot \left(\frac{(V \odot \hat{V}^{\beta-2})H^\top}{\hat{V}^{\beta-1}H^\top}\right)^{\phi(\beta)} \tag{2.6a}$$

$$H \leftarrow H \odot \left(\frac{W^\top(V \odot \hat{V}^{\beta-2})}{W^\top\hat{V}^{\beta-1}}\right)^{\phi(\beta)} \tag{2.6b}$$

となる。ただし $\hat{V} \coloneqq WH$ で、

$$\phi(\beta) \coloneqq \begin{cases} 1/(2-\beta) & (\beta < 1) \\ 1 & (1 \le \beta \le 2) \\ 1/(\beta-1) & (\beta > 2) \end{cases}$$

である。また、$\odot$ や除算、冪乗は行列の要素毎に行う。

2.6.4　教師あり NMF の更新式

plain NMF (2.2) とほぼ同様なやり方で、上界最小化アルゴリズムにより教師あり NMF (2.3) に対する更新式も求まる。各音源の基底 $W_1, ..., W_N$ は固定するからアクティベーション $H_1, ..., H_N$ だけを更新すればよくて、更新式は

教師あり NMF のアクティベーションの更新式

$$H_n \leftarrow H_n \odot \left(\frac{W_n^\top (X \odot \hat{X}^{\beta-2})}{W_n^\top \hat{X}^{\beta-1}} \right)^{\phi(\beta)} \quad (\forall n) \tag{2.7}$$

となる。ただし $\hat{X} \coloneqq \sum_n W_n H_n$ である。

第 3 章

拡張ラグランジュ関数法による識別的 NMF とその音声分離への応用

本章ではペナルティ法による識別的 NMF で用いられる新たな等式制約を提案し、さらに拡張ラグランジュ関数法による高速化を行う。識別的 NMF は NMF の基底を音源間で識別的に学習する手法であり、これを教師あり NMF によるモノラル音源分離のための基底学習法として用いれば、より良い分離性能を達成できる。近年、識別的 NMF に現れる二段階最適化問題を、その一部を等式制約に置き換えペナルティ法により解く手法が提案された。しかし、提案された等式制約は不自然な仮定に基づくことや、ペナルティ法は収束が遅く最適化が難しいなどの問題があった。そこで本章では、最適性条件に基づく理論的により自然な等式制約を提案し、さらに拡張ラグランジュ関数法による高速化を行なった。拡張ラグランジュ関数法による識別的 NMF ではペナルティ法よりも等式制約が高速に収束することが確認された。さらに、モノラル音声分離での実験の結果、提案された等式制約を用いた識別的 NMF がより良い分離性能を示すことが確認された。

3.1 序論

▶ plain NMF による音源分離

教師あり NMF による音源分離において、最も単純な基底学習法は plain NMF を用いることであった。ここにその定式化を再掲する。

plain NMF（再掲）

$$\min_{\mathsf{W},\mathsf{H}} \ \mathcal{R}(\mathsf{W},\mathsf{H}) \qquad \left(\mathcal{R}(\mathsf{W},\mathsf{H}) \coloneqq \sum_n D_\beta(V_n \mid W_n H_n)\right) \tag{3.1a}$$

$$\text{s.t. } W_1,...,W_N \in \mathbb{R}_+^{F\times K},\ H_1,...,H_N \in \mathbb{R}_+^{K\times T} \tag{3.1b}$$

ただし、学習は音源 $n = 1,...,N$ で同時に行うことができるから、それらを 1 つの最適化問題にまとめ、各音源の基底 $W_1,...,W_N$ やアクティベーション $H_1,...,H_N$ をテンソルにまとめてそれぞれ W,H と書いた。これらは三階のテンソルであり、$\mathsf{W} \in \mathbb{R}_+^{N\times F\times K}, \mathsf{H} \in \mathbb{R}_+^{N\times K\times T}$ である。また目的関数を $\mathcal{R}$ と書いた。$\mathcal{R}$ を、基底とアクティベーションの組み合わせによって各音源の振幅スペクトログラムを再構成するという意味で、各音源信号に対する**再構成誤差**（reconstruction error）と呼ぶことにする。

plain NMF による基底学習で利用されるのは各音源のクリーン信号（すなわち、その音源しか含まないデータ）だけである。つまり、学習の力点は基底とアクティベーションによってこのクリーン信号を上手く再構成できるかということだけに置かれる。

ここで、学習された各音源基底を用いて混合信号を分離するときのことを考えてみよう。これは教師あり NMF であり、その最適化問題は

教師あり NMF（再掲）

$$\min_{\mathsf{H}} \ \mathcal{A}(\mathsf{H}) \qquad \left(\mathcal{A}(\mathsf{H}) \coloneqq D_\beta\left(X \mid \sum_n W_n H_n\right)\right) \tag{3.2a}$$

$$\text{s.t. } H_1, ..., H_N \in \mathbb{R}_+^{K\times T} \tag{3.2b}$$

であった。ここで目的関数を $\mathcal{A}$ と書いたが、これは、与えられた各音源の基底を用いて混合信号を分析し、各音源のアクティベーションを取り出すという意味で、混合信号に対する**分析誤差**（analysis error）と呼ぶことにする。この $\mathcal{A}$ の最小化こそが、NMF による音源分離の最終的な目的である。ゆえに、各音源基底は $\mathcal{A}$ がきちんと最小化されやすくなるように学習されるべきであるが、plain NMF による基底学習では各音源信号の再構成誤差 $\mathcal{R}$ の最小化しか考えておらず、$\mathcal{A}$ の最小化は考慮されていない。

▶ 識別的 NMF

さて、分離対象として似通った音響的性質を持つ音源群が与えられたときのことを考えよう。これは非常に良くある設定である。例えば楽曲分析において、似た音色を持つ楽器群のそれぞれの旋律を取り出したい場合や、あるいは会議の文字起こしやスマートスピーカーの音声認識などで、前処理として複数人の混合音声を分離する必要がある場合などである。このとき plain NMF (3.1) で基底学習を行うと、各音源信号の再構成誤差 $\mathcal{R}$ を小さくすることしか考慮されないため、異なる音源間で似通った基底が学習されてしまう可能性がある。するとテスト時の分離 (3.2) の際、与えられた混合信号に対して、どの音源の基底に対応するアクティベーションを割り当てるかが困難になる。結果として、混合信号の分析誤差 $\mathcal{A}$ (3.2a) をきちんと減少させることができず、分離性能が落ちてしまう。

上述の問題に対処するために、識別的 NMF という手法が提案されている [12, 32, 40, 50]。これは混合信号に対する分析誤差 $\mathcal{A}$ も考慮に入れて基底を学習する手法である。これにより基底を識別的に学習することができ、類似した音響的性質を持つ音源群に対して分離性能が向上することが報告されている。

識別的 NMF は、数学的には分析誤差 $\mathcal{A}$ を小さくしつつ再構成誤差 $\mathcal{R}$ を小さくするという二段階最適化問題として定式化される。近年、この問題の一部を等式制約に置き換えて、ペナルティ法により一段階化して解く手法が提案された [32]。しかし、提案された等式制約は不自然な仮定に基づくことや、ペナルティ法は収束が遅く最適化が困難であるなどの問題があった。

そこで本章では、最適性条件に基づく理論的により自然な等式制約を提案し、さらにペナルティ法の代わりに拡張ラグランジュ関数法を用いることにより識別的 NMF の高速化を図る。そして提案手法の性能をモノラル音声分離のタスクにおいて検証する。

本章の構成は以下の通りである。3.2 節では準備として、識別的 NMF の定式化と、先行研究のペナルティ法により解く手法を説明する。3.3 節では新しい等式制約の提案と、拡張ラグランジュ関数法による高速化、およびその更新式の導出を行う。3.4 節ではまず提案手法の収束の振る舞いについて調べ、次に先行研究と提案手法との、モノラル音声分離における性能比較を行う。3.5 節では結論と今後の課題を述べる。

3.2　準備

3.2.1　識別的 NMF

▶ 定式化

3.1 節で述べたように、plain NMF（3.1）による基底学習では音響的性質が類似した音源群に対しては似通った基底が学習されてしまい、分離性能が悪くなるという問題があった。

この問題に対する解決策の一つとして提案されているのが基底の識別的な学習である。識別的学習の定式化の方法は複数あるが［12, 32, 40, 50］、特に文献［50］では各音源信号の再構成誤差 $\mathcal{R}$（3.1a）に加えて、混合信号の分析誤差 $\mathcal{A}$（3.2a）を考慮に入れ、識別的学習を次式のように定式化した。

識別的 NMF

$$\min_{\mathsf{W},\mathsf{H}} \ \mathcal{R}(\mathsf{W},\mathsf{H}) \qquad \left(\mathcal{R}(\mathsf{W},\mathsf{H}) = \sum_n D_\beta(V_n \mid W_n H_n)\right) \tag{3.3a}$$

$$\text{s.t. } \mathsf{H} = \arg\min_{\mathsf{H}'} \ \mathcal{A}(\mathsf{W},\mathsf{H}') \qquad \left(\mathcal{A}(\mathsf{W},\mathsf{H}') = D_\beta\left(X \mid \sum_n W_n H'_n\right)\right) \tag{3.3b}$$

ここで（3.3b）中の X は、学習データである各音源信号のクリーン信号を時間領域で足し合わせて作った仮の混合信号に対し STFT を掛け、振幅スペクトログラムを取ったものである。本論文では（3.3）による識別的学習の定義を特に**識別的 NMF**（Discriminative NMF; DNMF）と呼ぶことにする。（ただし、煩雑さを避けるために省略したが $\mathsf{W},\mathsf{H},\mathsf{H}'$ は非負という制約条件も勿論付く。）

数理最適化の分野では、このように最適化問題の制約条件の中にさらに別の最適化問題が含まれているものを**二段階最適化問題**（bilevel optimization problem）という［38］。上下の最適化問題や目的関数を指して、**上段の最適化問題**（upper level optimization problem）と**下段の最適化問題**（lower level optimization problem）、**上段の目的関数**（upper level objective function）と**下段の目的関数**（lower level objective function）という。

▶ 識別的学習の仕組み

この二段階最適化問題（3.3）の意味するところは、混合信号に対する分析誤差 $\mathcal{A}$（3.3b）を最小化するようなアクティベーション H を用いたときに、各音源信号の再構成誤差 $\mathcal{R}$（3.3a）を最小化するような基底 W を学習せよ、ということである。（$\mathcal{A}$ の最小化において基底 W は未知であるため、$\mathcal{A}$ は（3.2a）とは違って W,H の関数になることに注意。）二段階最適化問題ではあくまで「下段

の目的関数が最小化されている」という条件下で上段の目的関数を最小化しようとするため、識別的 NMF においては混合信号を分離することを最終目標として基底学習を行うこととなり、学習時とテスト時の目的が一致することになる。

識別的 NMF（3.3）により得られる基底がなぜ識別的になるかについては次のような説明ができる。音源数が $N = 2$ の場合を考えよう。ある時間フレーム t において、混合信号 $\boldsymbol{x}_t$ の中で音源 1 は有音（$\boldsymbol{v}_{1t} > 0$）だが音源 2 は無音（$\boldsymbol{v}_{2t} = 0$）だとする。今、各音源の基底 W_1, W_2 が何らかの意味で識別的でない、類似したスペクトルを持っていたとする。そしてそのために混合信号に対する分析誤差 $\mathcal{A}$（3.3b）の最小化の結果、時刻 t において音源 1 ではなく音源 2 にアクティベーションが割り当てられてしまった（$\boldsymbol{h}_{1t} = 0,\ \boldsymbol{h}_{2t} > 0$）としよう。するとこのとき、時刻 t において逆に音源 1 が無音で音源 2 が有音であると推定され、各音源信号に対する再構成誤差 $\mathcal{R}$（3.3a）が上昇してしまうことになる。以上をまとめると、識別的 NMF では、混合信号を分離する際に各音源へのアクティベーションの割り当て誤りが起こりにくいという意味で識別的な基底が学習される。

3.2.2　ペナルティ法による識別的 NMF の二段階最適化

近年、識別的 NMF の二段階最適化問題（3.3）に対して**ペナルティ法**（Penalty Method; PM）を用いた解法が提案された［32］。この手法では、下段の混合信号を分離するという最適化問題を何らかの等式制約

$$C_1 = \ldots = C_N = 0$$

に置き換え、これを次式のように上段の目的関数に組み込むことで、二段階最適化問題を一段階最適化問題に帰着させる。

ペナルティ法による識別的 NMF

$$\min_{\mathrm{W},\mathrm{H}} \ \mathcal{L}(\mathrm{W},\mathrm{H}) \qquad \left(\mathcal{L}(\mathrm{W},\mathrm{H}) \coloneqq \mathcal{R}(\mathrm{W},\mathrm{H}) + \frac{\lambda}{2}\sum_n \|C_n\|_F^2\right) \tag{3.4a}$$

$$\text{s.t. } W_1, \ldots, W_N \in \mathbb{R}_+^{F\times K},\ H_1, \ldots, H_N \in \mathbb{R}_+^{K\times T} \tag{3.4b}$$

$\lambda > 0$ は等式制約項 $C_1, \ldots, C_N$ の考慮の度合いを変えるペナルティパラメータである。ペナルティ法では目的関数 $\mathcal{L}$ を最小化する際に、各ステップにおいて λ を段々と大きくしていく。（λ の具体的な更新式は 3.4.2 節で後述する。）これにより、等式制約を満たしつつ、つまり混合信号に対する分析誤差を最小化しつつ、各音源信号に対する再構成誤差 $\mathcal{R}$ を最小化していくことができる。

先行研究［32］では、識別的 NMF の下段の最適化問題（3.3b）に対して、教師あり NMF の上界最小化アルゴリズムによる更新式（2.7）で最適化するときの更新の停止条件から、ペナルティ法に用いる等式制約を

更新式の停止条件に基づく等式制約

$$0 = C_n' \coloneqq H_n \odot \left(J - \left(\frac{W_n^\top (X \odot \hat{X}^{\beta-2})}{W_n^\top \hat{X}^{\beta-1}}\right)^{\phi(\beta)}\right) \quad (\forall n) \tag{3.5}$$

で定義している。

3.3 提案

3.3.1 最適性条件に基づく新しい等式制約

等式制約（3.5）はあくまでアクティベーションを教師あり NMF の上界最小化アルゴリズムによる更新式に従って更新するときの停止条件であり、識別的 NMF の下段の最適化問題（3.3b）の最適性条件ではない。特に等式制約（3.5）は H_n の行列要素が $= 0$ となるときも満たされてしまう。これをペナルティ項 $\|C'_n\|_F^2$ にすると、H_n に対する L_2 正則化のような効果が現れてしまう。

ところで、そもそも下段の目的関数 $\mathcal{A}$（3.3b）は、β ダイバージェンスが一方の変数を固定した場合には他方の変数について $1 \leq \beta \leq 2$ において凸であることから（2.6.1 節）、H_n に関して凸である。よって、下段の最適化問題（3.3b）について、簡単のため非負値制約 $H'_1, ..., H'_N \in \mathbb{R}_+^{K\times T}$ を無視して考えれば、$1 \leq \beta \leq 2$ においてその大域最適性条件は

最適性条件に基づく等式制約

$$0 = C''_n \coloneqq \frac{\partial \mathcal{A}}{\partial H_n} = W_n^\top \hat{X}^{\beta-1} - W_n^\top (X \odot \hat{X}^{\beta-2}) \quad (\forall n) \tag{3.6}$$

で与えられる。これをペナルティ法に用いる新たな等式制約とすることができる。ただし、本来は非負値制約 $H'_1, ..., H'_N \in \mathbb{R}_+^{K\times T}$ も加味しなければならないため、この等式制約にも一種の近似が入っている。また $\beta < 1$, $\beta > 2$ においては、$\mathcal{A}$ は非凸なのでこの等式制約は H_n が停留点にある条件でしかないことに注意する。

3.3.2 拡張ラグランジュ関数法による高速化

▶ 定式化

ペナルティ法の問題点は、ペナルティパラメータ λ の値が有限のときは一般に元の等式制約下最適化問題の正確な最適解が得られないことである。一方で λ を大きくするにつれて目的関数（3.4a）のヘッセ行列の条件数も大きくなり、収束が遅くなって最適化が困難になることが知られている [54]。

そこで本論文では、**拡張ラグランジュ関数法**（Augmented Lagrangian Method; ALM）による識別的 NMF の二段階最適化を提案する。拡張ラグランジュ関数法では、最適化問題を

拡張ラグランジュ関数法による識別的 NMF

$$\min_{\mathsf{W},\mathsf{H}} \ \mathcal{L}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) \qquad (\mathcal{L}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) \coloneqq \mathcal{R}(\mathsf{W},\mathsf{H}) + \mathcal{C}_{\mathsf{M}}(\mathsf{W},\mathsf{H})) \tag{3.7a}$$

$$\text{s.t. } W_1, ..., W_N \in \mathbb{R}_+^{F\times K},\ H_1, ..., H_N \in \mathbb{R}_+^{K\times T} \tag{3.7b}$$

で定義する。ここで $\mathcal{C}_{\mathsf{M}}$ は、

$$\Gamma_n \coloneqq C_n + \lambda^{-1} M_n \quad (\forall n)$$

とおいて、

$$\mathcal{C}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) \coloneqq \frac{\lambda}{2}\sum_{n}\|\Gamma_n\|_F^2 \tag{3.8}$$

で定義される拡張されたペナルティ項である。ペナルティ法でのペナルティ項（3.4a）に**ラグランジュ乗数**（Lagrange multiplier）$M_1,...,M_N \in \mathbb{R}^{K\times T}$ を導入した点が違いである。M はこれらをまとめたテンソルで、$\mathsf{M} \in \mathbb{R}^{N\times K\times T}$ である。なお等式制約項 C_n としては先行研究の C'_n（3.5）か、提案した C''_n（3.6）を使う。

▶ 拡張ラグランジュ関数法の仕組み

ペナルティ法では基底 W およびアクティベーション H の更新とペナルティパラメータ λ の更新とを交互に繰り返すが、拡張ラグランジュ関数法ではこれらの間に

拡張ラグランジュ関数法におけるラグランジュ乗数の更新式

$$M_n \leftarrow M_n + \lambda C_n \quad (\forall n) \tag{3.9}$$

というラグランジュ乗数の更新を挟む。このようにする理由は次のような力学モデルを考えるとわかりやすい。拡張されたペナルティ項（の音源 n 成分）$(\lambda/2)\|C_n+\lambda^{-1}M_n\|_F^2$ がバネ定数 λ のバネに繋がれた質点 C_n のポテンシャルエネルギーのようなものだと考えると、ペナルティパラメータ λ を大きくしていくことはバネを強くしていくことに相当し、拡張ラグランジュ乗数を（3.9）で更新することは質点が振れている方向とは逆の方向にバネの支点を動かすことに相当する。この支点シフトの操作により、ペナルティ法よりも高速に等式制約項 C_n を 0 に近づけられる。さらにペナルティ法と違って、有限の λ の値でもある適当な条件下で元の等式制約下最適化問題の最適解が得られることが知られている [54]。

3.3.3　更新式の導出

2 種類の等式制約項（3.5）、（3.6）と 2 種類の最適化手法（3.4）、（3.7）で、計 4 種類の場合の基底とアクティベーションの更新式を導出する。ところで、拡張ラグランジュ関数法における更新式さえ得られれば、その更新式において恒等的にラグランジュ乗数 $M_1 = ... = M_N = 0$ とすれば自動的にペナルティ法における更新式が求まる。そこで以下では 2 種類の等式制約を用いた場合の拡張ラグランジュ関数法による更新式を導出しよう。なおペナルティパラメータ λ の更新式については 3.4.2 節で述べる。

さて、NMF の更新式の求め方としては上界最小化アルゴリズム（2.6.2 節）が定番だが、拡張ラグランジュ関数法における目的関数（3.7a）に対して separable な代理関数を設計することは大変難しい。そこで我々は更新式の導出に Févotte rule を採用する [7]。この方法では、例えば音源 n の基底 W_n の更新式は、目的関数を W_n で偏微分したものを正負に分けて $= \Delta^+ - \Delta^-, (\Delta^+, \Delta^- \geq 0)$ とおき、ステップ幅を W_n/Δ^+ とする勾配降下法を考えることで

$$W_n \leftarrow W_n \odot \Delta^-/\Delta^+$$

という乗算型更新式が得られる。他の音源の基底やアクティベーションの場合も全く同様である。

ただし、Févotte rule はあくまで場当たり的な方法に過ぎず、上界最小化アルゴリズムとは違って目的関数の単調非増加性は保証できないことに注意する。

▶ 提案された等式制約項を用いた場合

等式制約項 C_n として提案された C_n''（3.6）を用いた場合の、拡張ラグランジュ関数法による識別的 NMF の更新式を求める。目的関数（3.7a）を再掲しておく：

$$\mathcal{L}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) = \mathcal{R}(\mathsf{W},\mathsf{H}) + \mathcal{C}_{\mathsf{M}}(\mathsf{W},\mathsf{H})$$

まず第一項の各音源に対する再構成誤差 $\mathcal{R}$ は

$$\mathcal{R} = \sum_{nft}\left(\frac{\hat{v}_{nft}^{\beta}}{\beta} - \frac{v_{nft}\hat{v}_{nft}^{\beta-1}}{\beta-1}\right) + \text{const.} \quad (\hat{v}_{nft} \coloneqq \boldsymbol{w}_{nf}\boldsymbol{h}_{nt})$$

と書けた。（ただし $\boldsymbol{W},\boldsymbol{H}$ に依存しない項を省いた。また、$\boldsymbol{w}_{nf} \in \mathbb{R}_{+}^{1\times K}$ は W_n の第 f 行ベクトル、$\boldsymbol{h}_{nt} \in \mathbb{R}_{+}^{K}$ は H_n の第 t 列ベクトルである。）これを w_{nfk} で偏微分すると

$$\frac{\partial \mathcal{R}}{\partial w_{nfk}} = \sum_{t}(\hat{v}_{nft}^{\beta-1} - v_{nft}\hat{v}_{nft}^{\beta-2})h_{nkt}$$

となり、行列形式で書くと

$$\frac{\partial \mathcal{R}}{\partial W_n} = (\hat{V}_n^{\beta-1} - V_n \odot \hat{V}_n^{\beta-2})H_n^{\top} \tag{3.10}$$

となる。一方 $\mathcal{R}$ を h_{nkt} で偏微分すると、

$$\frac{\partial \mathcal{R}}{\partial h_{nkt}} = \sum_{f} w_{nfk}(\hat{v}_{nft}^{\beta-1} - v_{nft}\hat{v}_{nft}^{\beta-2})$$

となり、行列形式で書くと

$$\frac{\partial \mathcal{R}}{\partial H_n} = W_n^{\top}(\hat{V}_n^{\beta-1} - V_n \odot \hat{V}_n^{\beta-2}) \tag{3.11}$$

となる。

次に第二項の拡張されたペナルティ項 $\mathcal{C}_{\mathsf{M}}$ は

$$\mathcal{C}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) = \frac{\lambda}{2}\sum_{n}\|\Gamma_n\|_F^2 = \frac{\lambda}{2}\sum_{nkt}\gamma_{nkt}^2$$

と書けた。これを w_{nfk} で偏微分すると

$$\frac{\partial \mathcal{C}_{\mathsf{M}}}{\partial w_{nfk}} = \lambda \sum_{n'k't}\gamma_{n'k't}\frac{\partial c_{n'k't}}{\partial w_{nfk}}$$

となる。ここで今提案された等式制約項（3.6）を使っているから

$$c_{nkt} = c''_{nkt} = \sum_{f} w_{nfk}(\hat{X}^{\beta-1} - X \odot \hat{X}^{\beta-2})_{ft}$$

であり、

$$\frac{\partial c_{n'k't}}{\partial w_{nfk}} = \delta_{nn'}\delta_{kk'}(\hat{X}^{\beta-1} - X \odot \hat{X}^{\beta-2})_{ft} + \sum_{f'} w_{n'f'k'}[(\beta-1)\hat{X}^{\beta-2} - (\beta-2)X \odot \hat{X}^{\beta-3}]_{f't}\frac{\partial \hat{x}_{f't}}{\partial w_{nfk}}$$

となる。

$$\frac{\partial \hat{x}_{f't}}{\partial w_{nfk}} = \delta_{ff'} h_{nkt}$$

であるから、結局

$$\frac{\partial \mathcal{C}_{\mathrm{M}}}{\partial w_{nfk}} = \lambda \sum_{t} \gamma_{nkt} (\hat{X}^{\beta-1} - X \odot \hat{X}^{\beta-2})_{ft} + \lambda \sum_{n'k't} \gamma_{n'k't} w_{n'fk'} [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}]_{ft} h_{nkt}$$

となる。行列形式で書くと

$$\frac{\partial \mathcal{C}_{\mathrm{M}}}{\partial W_n} = \lambda (\hat{X}^{\beta-1} - X \odot \hat{X}^{\beta-2}) \Gamma_n^{\top} + \lambda \{ Y \odot [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}] \} H_n^{\top} \tag{3.12}$$

となる。ここで

$$Y \coloneqq \sum_{n} W_n \Gamma_n$$

とおいた。一方 $\mathcal{C}_{\mathrm{M}}$ を h_{nkt} で偏微分すると

$$\frac{\partial \mathcal{C}_{\mathrm{M}}}{\partial h_{nkt}} = \lambda \sum_{n'k't'} \gamma_{n'k't'} \frac{\partial c_{n'k't'}}{\partial h_{nkt}}$$

となり、

$$\frac{\partial c_{n'k't'}}{\partial h_{nkt}} = \sum_{f} w_{n'fk'} [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}]_{ft'} \frac{\partial \hat{x}_{ft'}}{\partial h_{nkt}}$$

で、

$$\frac{\partial \hat{x}_{ft'}}{\partial h_{nkt}} = \delta_{tt'} w_{nfk}$$

であるから、結局

$$\frac{\partial \mathcal{C}_{\mathrm{M}}}{\partial h_{nkt}} = \lambda \sum_{n'fk'} \gamma_{n'k't} w_{n'fk'} [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}]_{ft} w_{nfk}$$

となる。行列形式で書くと

$$\frac{\partial \mathcal{C}_{\mathrm{M}}}{\partial H_n} = \lambda W_n^{\top} \{ Y \odot [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}] \} \tag{3.13}$$

となる。

以上をまとめると、等式制約項 C_n として提案した C_n''（3.6）を用いた場合、拡張ラグランジュ関数法における目的関数（3.7a）の、音源 n の基底 W_n による偏微分は、（3.10）、（3.12）から

$$\begin{aligned} \frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial W_n} =& (\hat{V}_n^{\beta-1} - V_n \odot \hat{V}_n^{\beta-2}) H_n^{\top} \\ &+ \lambda (\hat{X}^{\beta-1} - X \odot \hat{X}^{\beta-2}) (\Gamma_n^{+} - \Gamma_n^{-})^{\top} \\ &+ \lambda \{ (Y^{+} - Y^{-}) \odot [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}] \} H_n^{\top} \quad (\forall n) \end{aligned} \tag{3.14}$$

となる。一方、対応するアクティベーション H_n による偏微分は、（3.11）、（3.13）から

$$\begin{aligned}\frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial H_n} =& W_n^\top (\hat{V}_n^{\beta-1} - V_n \odot \hat{V}_n^{\beta-2}) \\ &+ \lambda W_n^\top \{(Y^+ - Y^-) \odot [(\beta-1)\hat{X}^{\beta-2} - (\beta-2) X \odot \hat{X}^{\beta-3}] \quad (\forall n)\end{aligned} \tag{3.15}$$

となる。ただし、更新式を求める際に正負の項に分類する必要があるため、

$$C_n^+ = (C_n'')^+ \coloneqq W_n^\top \hat{X}^{\beta-1}, \quad C_n^- = (C_n'')^- \coloneqq W_n^\top (X \odot \hat{X}^{\beta-2}) \quad (\forall n)$$

$$M_n^+ \coloneqq \max\{M_n, 0\}, \quad M_n^- \coloneqq \max\{-M_n, 0\} \quad (\forall n)$$

$$\Gamma_n^+ \coloneqq C_n^+ + \lambda^{-1} M_n^+, \quad \Gamma_n^- \coloneqq C_n^- + \lambda^{-1} M_n^- \quad (\forall n)$$

$$Y^+ \coloneqq \sum_n W_n \Gamma_n^+, \quad Y^- \coloneqq \sum_n W_n \Gamma_n^-$$

とおいた。これらは全て非負である。また max は行列要素毎に取る。

得られた勾配（3.14）、（3.15）を正負に分けることで更新式が導出できる。ただし β の値に応じて係数の正負が変わることに注意する。$\beta = 0$（板倉齋藤ダイバージェンス）のときは

識別的 NMF の基底とアクティベーションの更新式（$\beta = 0,\ C_n = C_n''$）

$$W_n \leftarrow W_n \odot \frac{\frac{V_n}{\hat{V}_n^2} H_n^\top + \lambda \left(\frac{J}{\hat{X}} (\Gamma_n^-)^\top + \frac{X}{\hat{X}^2} (\Gamma_n^+)^\top \right) + \lambda \left(2 \frac{Y^- \odot X}{\hat{X}^3} + \frac{Y^+}{\hat{X}^2} \right) H_n^\top}{\frac{J}{\hat{V}_n} H_n^\top + \lambda \left(\frac{J}{\hat{X}} (\Gamma_n^+)^\top + \frac{X}{\hat{X}^2} (\Gamma_n^-)^\top \right) + \lambda \left(2 \frac{Y^+ \odot X}{\hat{X}^3} + \frac{Y^-}{\hat{X}^2} \right) H_n^\top} \quad (\forall n) \tag{3.16a}$$

$$H_n \leftarrow H_n \odot \frac{W_n^\top \frac{V_n}{\hat{V}_n^2} + \lambda W_n^\top \left(2 \frac{Y^- \odot X}{\hat{X}^3} + \frac{Y^+}{\hat{X}^2} \right)}{W_n^\top \frac{J}{\hat{V}_n} + \lambda W_n^\top \left(2 \frac{Y^+ \odot X}{\hat{X}^3} + \frac{Y^-}{\hat{X}^2} \right)} \quad (\forall n) \tag{3.16b}$$

となる。$\beta = 1$（一般化 KL ダイバージェンス）のときは

識別的 NMF の基底とアクティベーションの更新式（$\beta = 1,\ C_n = C_n''$）

$$W_n \leftarrow W_n \odot \frac{\frac{V_n}{\hat{V}_n} H_n^\top + \lambda \left(J (\Gamma_n^-)^\top + \frac{X}{\hat{X}} (\Gamma_n^+)^\top \right) + \lambda \frac{Y^- \odot X}{\hat{X}^2} H_n^\top}{J H_n^\top + \lambda \left(J (\Gamma_n^+)^\top + \frac{X}{\hat{X}} (\Gamma_n^-)^\top \right) + \lambda \frac{Y^+ \odot X}{\hat{X}^2} H_n^\top} \quad (\forall n) \tag{3.17a}$$

$$H_n \leftarrow H_n \odot \frac{W_n^\top \frac{V_n}{\hat{V}_n} + \lambda W_n^\top \frac{Y^- \odot X}{\hat{X}^2}}{W_n^\top J + \lambda W_n^\top \frac{Y^+ \odot X}{\hat{X}^2}} \quad (\forall n) \tag{3.17b}$$

となる。$\beta = 2$（ユークリッド距離）のときは

識別的 NMF の基底とアクティベーションの更新式（$\beta = 2,\ C_n = C_n''$）

$$W_n \leftarrow W_n \odot \frac{V_n H_n^\top + \lambda \left(\hat{X} (\Gamma_n^-)^\top + X (\Gamma_n^+)^\top \right) + \lambda Y^- H_n^\top}{\hat{V}_n H_n^\top + \lambda \left(\hat{X} (\Gamma_n^+)^\top + X (\Gamma_n^-)^\top \right) + \lambda Y^+ H_n^\top} \quad (\forall n) \tag{3.18a}$$

$$H_n \leftarrow H_n \odot \frac{W_n^\top V_n + \lambda W_n^\top Y^-}{W_n^\top \hat{V}_n + \lambda W_n^\top Y^+} \quad (\forall n) \tag{3.18b}$$

となる。

▶ 先行研究の等式制約を用いた場合

等式制約項 C_n として先行研究の C'_n（3.5）を用いた場合の、拡張ラグランジュ関数法による識別的 NMF の更新式を求める。前節と全く同様にして更新式を導出できる。目的関数（3.7a）の、音源 n の基底 W_n による偏微分は、

$$\begin{aligned}\frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial W_n} =&(\hat{V}_n^{\beta-1} - V_n \odot \hat{V}_n^{\beta-2})H_n^\top \\ &+ \lambda\phi(\beta)\hat{X}^{\beta-1}(\underline{D}_n^+ - \underline{D}_n^-)^\top \\ &- \lambda\phi(\beta)(X \odot \hat{X}^{\beta-2})(\overline{D}_n^+ - \overline{D}_n^-)^\top \\ &+ \lambda\phi(\beta)(\beta-1)\left[(\underline{Z}^+ - \underline{Z}^-) \odot \hat{X}^{\beta-2}\right]H_n^\top \\ &- \lambda\phi(\beta)(\beta-2)\left[(\overline{Z}^+ - \overline{Z}^-) \odot X \odot \hat{X}^{\beta-3}\right]H_n^\top \quad (\forall n)\end{aligned} \tag{3.19}$$

となる。一方対応するアクティベーション H_n による偏微分は、

$$\begin{aligned}\frac{\partial \mathcal{L}_{\mathrm{M}}}{\partial H_n} =&W_n^\top(\hat{V}_n^{\beta-1} - V_n \odot \hat{V}_n^{\beta-2}) \\ &+ \left(J - \left(\frac{W_n^\top(X \odot \hat{X}^{\beta-2})}{W_n^\top \hat{X}^{\beta-1}}\right)^{\phi(\beta)}\right) \odot (\Gamma_n^+ - \Gamma_n^-) \\ &+ \lambda\phi(\beta)(\beta-1)W_n^\top\left[(\underline{Z}^+ - \underline{Z}^-) \odot \hat{X}^{\beta-2}\right] \\ &- \lambda\phi(\beta)(\beta-2)W_n^\top\left[(\overline{Z}^+ - \overline{Z}^-) \odot X \odot \hat{X}^{\beta-3}\right] \quad (\forall n)\end{aligned} \tag{3.20}$$

となる。ただし、

$$C_n^+ = (C'_n)^+ \coloneqq H_n, \quad C_n^- = (C'_n)^- \coloneqq H_n \odot \left(\frac{W_n^\top(X \odot \hat{X}^{\beta-2})}{W_n^\top \hat{X}^{\beta-1}}\right)^{\phi(\beta)} \quad (\forall n)$$

$$M_n^+ \coloneqq \max\{M_n, 0\}, \quad M_n^- \coloneqq \max\{-M_n, 0\} \quad (\forall n)$$

$$\Gamma_n^+ \coloneqq C_n^+ + \lambda^{-1}M_n^+, \quad \Gamma_n^- \coloneqq C_n^- + \lambda^{-1}M_n^- \quad (\forall n)$$

$$\underline{D}_n^+ = \frac{C_n^- \odot \Gamma_n^+}{W_n^\top \hat{X}^{\beta-1}}, \quad \underline{D}_n^- = \frac{C_n^- \odot \Gamma_n^-}{W_n^\top \hat{X}^{\beta-1}} \quad (\forall n)$$

$$\overline{D}_n^+ = \frac{C_n^- \odot \Gamma_n^+}{W_n^\top (X \odot \hat{X}^{\beta-2})}, \quad \overline{D}_n^- = \frac{C_n^- \odot \Gamma_n^-}{W_n^\top (X \odot \hat{X}^{\beta-2})} \quad (\forall n)$$

$$\underline{Z}^+ \coloneqq \sum_n W_n \underline{D}_n^+, \quad \underline{Z}^- \coloneqq \sum_n W_n \underline{D}_n^-$$

$$\overline{Z}^+ \coloneqq \sum_n W_n \overline{D}_n^+, \quad \overline{Z}^- \coloneqq \sum_n W_n \overline{D}_n^-$$

とおいた。これらは全て非負である。

得られた勾配（3.19）、（3.20）から更新式が導出できる。$\beta = 0$（板倉齋藤ダイバージェンス）のときは

識別的 NMF の基底とアクティベーションの更新式（$\beta = 0,\ C_n = C'_n$）

$$W_n \leftarrow W_n \odot \frac{\frac{V_n}{\hat{V}_n^2} H_n^\top + \frac{\lambda}{2}\left(\frac{J}{\hat{X}}(\underline{D}_n^-)^\top + \frac{X}{\hat{X}^2}(\overline{D}_n^+)^\top + \left(\frac{\underline{Z}^+}{\hat{X}^2} + \frac{2\overline{Z}^- \odot X}{\hat{X}^3}\right) H_n^\top\right)}{\frac{J}{\hat{V}_n} H_n^\top + \frac{\lambda}{2}\left(\frac{J}{\hat{X}}(\underline{D}_n^+)^\top + \frac{X}{\hat{X}^2}(\overline{D}_n^-)^\top + \left(\frac{\underline{Z}^-}{\hat{X}^2} + \frac{2\overline{Z}^+ \odot X}{\hat{X}^3}\right) H_n^\top\right)} \quad (\forall n) \tag{3.21a}$$

$$H_n \leftarrow H_n \odot \frac{W_n^\top \frac{V_n}{\hat{V}_n^2} + \lambda\left(\Gamma_n^- + \left(\frac{W_n^\top \frac{X}{\hat{X}^2}}{W_n^\top \frac{J}{\hat{X}}}\right)^{\frac{1}{2}} \odot \Gamma_n^+ + \frac{1}{2} W_n^\top \left(\frac{\underline{Z}^+}{\hat{X}^2} + \frac{2\overline{Z}^- \odot X}{\hat{X}^3}\right)\right)}{W_n^\top \frac{J}{\hat{V}_n} + \lambda\left(\Gamma_n^+ + \left(\frac{W_n^\top \frac{X}{\hat{X}^2}}{W_n^\top \frac{J}{\hat{X}}}\right)^{\frac{1}{2}} \odot \Gamma_n^- + \frac{1}{2} W_n^\top \left(\frac{\underline{Z}^-}{\hat{X}^2} + \frac{2\overline{Z}^+ \odot X}{\hat{X}^3}\right)\right)} \quad (\forall n) \tag{3.21b}$$

となる。$\beta = 1$（一般化 KL ダイバージェンス）のときは

識別的 NMF の基底とアクティベーションの更新式（$\beta = 1,\ C_n = C'_n$）

$$W_n \leftarrow W_n \odot \frac{\frac{V_n}{\hat{V}_n} H_n^\top + \lambda\left(J(\underline{D}_n^-)^\top + \frac{X}{\hat{X}}(\overline{D}_n^+)^\top + \frac{\overline{Z}^- \odot X}{\hat{X}^2} H_n^\top\right)}{J H_n^\top + \lambda\left(J(\underline{D}_n^+)^\top + \frac{X}{\hat{X}}(\overline{D}_n^-)^\top + \frac{\overline{Z}^+ \odot X}{\hat{X}^2} H_n^\top\right)} \quad (\forall n) \tag{3.22a}$$

$$H_n \leftarrow H_n \odot \frac{W_n^\top \frac{V_n}{\hat{V}_n} + \lambda\left(\Gamma_n^- + \frac{W_n^\top \frac{X}{\hat{X}}}{W_n^\top \hat{X}} \odot \Gamma_n^+ + W_n^\top \frac{\overline{Z}^- \odot X}{\hat{X}^2}\right)}{W_n^\top J + \lambda\left(\Gamma_n^+ + \frac{W_n^\top \frac{X}{\hat{X}}}{W_n^\top \hat{X}} \odot \Gamma_n^- + W_n^\top \frac{\overline{Z}^+ \odot X}{\hat{X}^2}\right)} \quad (\forall n) \tag{3.22b}$$

となる。$\beta = 2$（ユークリッド距離）のときは

識別的 NMF の基底とアクティベーションの更新式（$\beta = 2,\ C_n = C'_n$）

$$W_n \leftarrow W_n \odot \frac{V_n H_n^\top + \lambda\left(\hat{X}(\underline{D}_n^-)^\top + X(\overline{D}_n^+)^\top + \underline{Z}^- H_n^\top\right)}{\hat{V}_n H_n^\top + \lambda\left(\hat{X}(\underline{D}_n^+)^\top + X(\overline{D}_n^-)^\top + \underline{Z}^+ H_n^\top\right)} \quad (\forall n) \tag{3.23a}$$

$$H_n \leftarrow H_n \odot \frac{W_n^\top V_n + \lambda\left(\Gamma_n^- + \frac{W_n^\top X}{W_n^\top \hat{X}} \odot \Gamma_n^+ + W_n^\top \underline{Z}^-\right)}{W_n^\top \hat{V}_n + \lambda\left(\Gamma_n^+ + \frac{W_n^\top X}{W_n^\top \hat{X}} \odot \Gamma_n^- + W_n^\top \underline{Z}^+\right)} \quad (\forall n) \tag{3.23b}$$

となる。

3.4 実験

以降では、ペナルティ法および拡張ラグランジュ関数法による識別的 NMF において先行研究 [32] の等式制約項 (3.5) を用いたものを DNMF（previous）、提案した等式制約項 (3.6) を用い

たものを DNMF (proposed) と呼ぶことにする。本研究では plain NMF を基準として、$\beta = 0, 1, 2$ の場合にペナルティ法および拡張ラグランジュ関数法による 2 種の DNMF の収束の振る舞いを調べ（3.4.2 節）、さらにモノラル音声分離における性能の比較を行った（3.4.3 節）。

3.4.1　実験条件

データセットは ATR 音素バランス 503 文 [22] から作成した。音源数を $N = 2$ とし、話者のペアを FKN/FTK、FKN/MHT、MHT/MSH（FKN と FTK は女性、MHT と MSH は男性）の 3 種類とした。3.4.2 節で述べるハイパーパラメータの決定と収束の確認のための開発データセットとして、FKN/MHT の A、B セットの一人当たり全 100 発話を用いた。また 3.4.3 節の音声分離においては、各ペアに対して、学習データセットとして C、D セットの一人当たり全 100 発話、テストデータセットとして E、F セットの一人当たり全 100 発話を用いた。混合音声は、分離対象のペア中の各話者の、同一の発話内容（つまり、同じ文章）の音声を、SNR を 0 dB として瞬時混合し作成した。標本化周波数は 20 kHz であり、短時間フーリエ変換のフレーム幅は 25.6 ms、フレームシフトは 12.8 ms とした。窓はハニング窓を用いた。各音源の基底ベクトルの本数は $K = 200$ とし、基底 W とアクティベーション H は $[0, 1]$ の区間の乱数により初期化した。またラグランジュ乗数 M の初期値は 0 とした。なお、ゼロ除算を防ぐために、最適化中に $\mathrm{eps} = 10^{-9}$ 以下の行列要素が現れた場合それらは全て eps に置き換えた。最適化の反復回数は 1200 回とした。推定音声の位相には、混合音声の位相をそのまま用いた。音声分離の評価指標としては音源対歪み比（source to distortion ratio; SDR）[46] の改善量（混合音声をそのまま分離音声とした場合からの SDR の改善量）である SDRi を用いた。この値が高いほど分離性能が良い。

3.4.2　提案手法の収束について

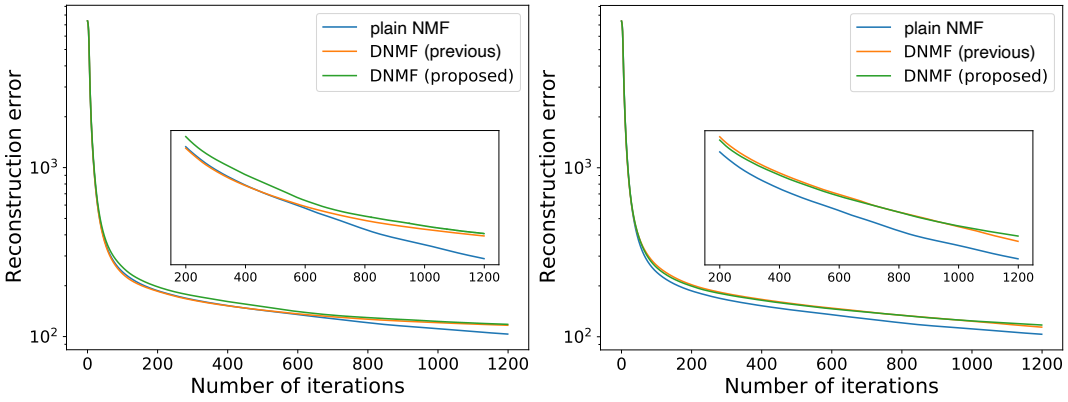
収束について述べる前に、まずペナルティパラメータ λ の更新法について説明する。拡張ラグランジュ関数法の目的関数（3.7a）中の再構成誤差 $\mathcal{R}$ を十分小さくするためには、最適化の反復の初期段階では λ を小さくする必要がある一方で、等式制約 $C_1 = \ldots = C_N = 0$ を満たすためには段々と大きくしていかなければならない。そこで本研究では、先行研究 [32] のやり方にならって、λ を

$$\lambda = \begin{cases} \lambda_{\mathrm{I}} & (1 \leq i \leq 200) \\ \lambda_{\mathrm{I}} \left(\lambda_{\mathrm{L}} / \lambda_{\mathrm{I}}\right)^{(i-201)/999} & (201 \leq i \leq 1200) \end{cases} \tag{3.24}$$

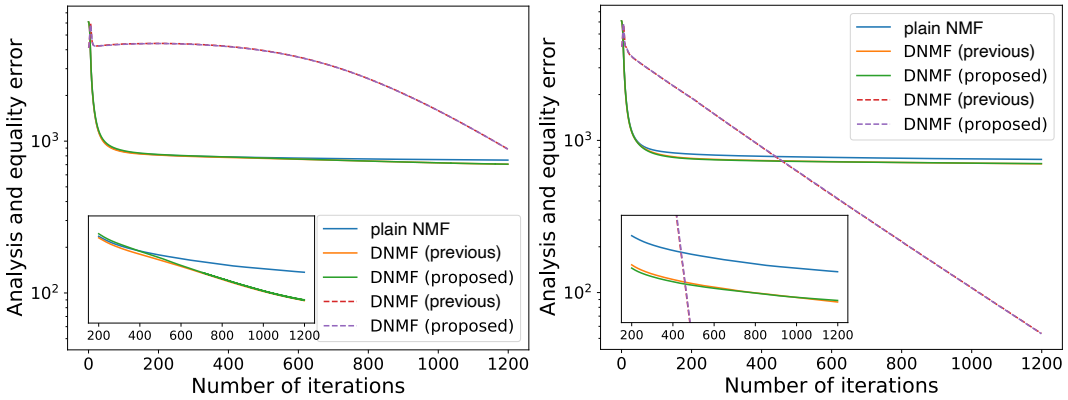
のように更新する。i は反復のインデックスであり、$\lambda_{\mathrm{I}}, \lambda_{\mathrm{L}} > 0$ はそれぞれ λ の初期値と最終値である。開発データセットを用いたチューニングにより、$\beta = 0, 1, 2$ でそれぞれ $(\lambda_{\mathrm{I}}, \lambda_{\mathrm{L}}) = (10^{-4}, 10^{-1})$, $(10^{-5}, 10^{-2})$, $(10^{-7}, 10^{-4})$ と決定した。

これらのハイパーパラメータを用いて、開発データセットにおける提案手法の収束について調べた。Fig. 3.1 に、$\beta = 1$ としてペナルティ法または拡張ラグランジュ関数法により DNMF (previous) と DNMF (proposed) を行った際の、各話者音声に対する再構成誤差 $\mathcal{R}$ (3.1a)、混合音声に対する分析誤差 $\mathcal{A}$ (3.2a) および $\sum_{n=1}^{N} \|C_n\|_1$ で計算された等式誤差（equality error）の収束の振る舞いを示す（ただし、以上の諸量は時間フレーム数 T で正規化してある）。$\mathcal{R}$ と $\mathcal{A}$ については基準として plain NMF の値を載せてある。

Fig. 3.1b, 3.1d を見ると、2 種の DNMF は plain NMF より $\mathcal{R}$ が大きいものの、$\mathcal{A}$ は小さくなっ



(a) Reconstruction errors (PM). (b) Reconstruction errors (ALM).



(c) Analysis errors and equality errors (PM). (d) Analysis errors and equality errors (ALM).

Fig. 3.1: Convergence behavior of the reconstruction errors $\mathcal{R}$, the analysis errors $\mathcal{A}$ and the equality errors $\sum_{n=1}^{N} \|C_n\|_1$ when $\beta = 1$. In (a) and (c), we used PM to execute DNMF. In (b) and (d), we used ALM. In (c) and (d), the solid lines represent the analysis errors and the dotted lines represent the equality errors.

ていることが分かる。これは DNMF が、各話者音声に対する再構成誤差よりも混合音声に対する分析誤差を下げることを重視して、基底の識別的学習を行ったためだと考えられる。加えて Fig. 3.1a, 3.1b を比較すると、ペナルティ法よりも拡張ラグランジュ関数法の方が反復の早い段階で $\mathcal{R}$ に正則化を掛けられている。同様に Fig. 3.1c, 3.1d を比較すると、拡張ラグランジュ関数法の方が $\mathcal{A}$ および等式誤差 $\sum_{n=1}^{N} \|C_n\|_1$ が急速に減少している。この 2 つの差異は、拡張ラグランジュ関数法ではペナルティ法より急速に等式制約が満たされるということの証左だと考えられる。

3.4.3 モノラル音声分離における性能比較

Table 3.1 に $\beta = 0, 1, 2$ での plain NMF、DNMF（previous）および DNMF（proposed）を用いた教師あり NMF による、各話者ペアに対する SDRi を示す。太字の数値が各音源についての最高値である。2 種の DNMF については拡張ラグランジュ関数法を用いた場合に加えてペナルティ法を用いた場合の結果が括弧中に付記されている。

これを見ると $\beta = 1, 2$ のとき DNMF（proposed）が平均的に SDRi が高く、特に $\beta = 2$ で最大である。これは DNMF（proposed）では基底の識別的学習が上手くはたらいたためと考えられる。DNMF（previous）と DNMF（proposed）を比較すると全ての β で後者の方が平均的により

Table 3.1: Separation performance calculated as SDRi [dB]

β	Pair	plain NMF	DNMF (previous) ALM (PM)	DNMF (proposed) ALM (PM)
0	FKN/FTK	**0.164**/**-0.066**	-0.479/-0.570 (-0.353/-0.645)	-0.476/-0.355 (-0.535/-0.490)
	FKN/MHT	1.495/1.516	1.253/1.323 (1.617/1.800)	1.833/**2.327** (**1.873**/2.062)
	MHT/MSH	-0.060/**0.414**	-0.629/0.059 (-0.746/-0.324)	-**0.042**/0.283 (-0.708/0.067)
1	FKN/FTK	1.061/**1.331**	0.971/1.219 (1.083/1.277)	**1.086**/1.284 (0.954/1.100)
	FKN/MHT	2.002/2.495	2.458/2.941 (2.060/2.455)	**2.560**/**3.048** (2.138/2.575)
	MHT/MSH	0.733/0.786	0.670/0.702 (0.510/0.547)	**0.798**/**0.838** (0.790/0.820)
2	FKN/FTK	1.451/1.969	2.223/3.089 (**2.239**/3.161)	2.216/**3.195** (2.180/3.089)
	FKN/MHT	2.353/2.974	2.909/3.721 (2.864/3.499)	**3.004**/**3.784** (2.857/3.528)
	MHT/MSH	0.976/1.040	1.571/**1.649** (1.395/1.477)	**1.590**/1.640 (1.349/1.441)

SDRi が大きい。この理由としては、提案した等式制約項 C_n''（3.6）が先行研究 [32] の等式制約項 C_n'（3.5）より正確であることや、3.3.1 節で指摘したように C_n' を用いた場合は、ペナルティパラメータ λ を大きくしたときに過剰な L_2 正則化が起こってしまっている可能性などが考えられる。また 2 種の DNMF に共通して、ペナルティ法の方が拡張ラグランジュ関数法より高い SDRi を示す場合が見られる。これは、λ が大きすぎたために、拡張ラグランジュ関数法により高速にペナルティ項を減少させていった結果、等式制約が重視されすぎて逆に性能が落ちてしまった可能性を示唆している。

話者ペア毎の違いを見てみると、まず異性ペア（FKN/MHT）では全ての β で DNMF（proposed）が最高の SDRi を達成している。特に、3.3.1 節で述べたように下段の目的関数（3.3b）が非凸である $\beta = 0$ でも DNMF（proposed）が上手く動作していることは注目に値する。一方で同性ペア（FKN/FTK、MHT/MSH）では、$\beta = 0,1$ で plain NMF よりも 2 種の DNMF の SDRi が逆に減少してしまっている場合が見られる。これは、あまりに各音源のスペクトルが似通っている場合には $\beta = 0,1$ での DNMF による識別的学習が逆効果となってしまう可能性を示している。

3.5　結論

3.5.1　まとめ

本章ではペナルティ法による識別的 NMF で利用される等式制約として、識別的 NMF の下段の目的関数の最適性条件から得られる新たな等式制約を提案し、更に拡張ラグランジュ関数法による識別的 NMF の更新式を導出した。モノラル音声分離における性能比較の結果、従来手法に対する提案手法の優位性が示されたが、あまりに似通った音源に対しては識別的 NMF が逆効果となってしまう可能性も確認された。

3.5.2　今後の課題

今後の研究課題として、まず、識別的 NMF がどのように識別的学習を行なっているか、あるいは識別的 NMF が定義する「識別性」とは何かということについて理論的に検討する必要がある。拡張ラグランジュ関数法による識別的 NMF の定式化 (3.7) を見ると、識別的 NMF はある意味で混合音声 X の情報を用いて基底 W に対して正則化を掛けていると考えられる。一方で、教師あり NMF による音声分離のための基底学習法としては他にも、アクティベーションに対してスパース正則化を掛けるスパース NMF [16, 23] などがある。スパース性の度合いが、例えば行列の非ゼロ要素の個数や Hoyer sparsity [16] で定量化できるのと同じように、識別的 NMF の定義する「識別性」も何らかの形で定量化できれば、識別的 NMF の学習する基底の性質を調べ、他の手法と定量的な比較ができる。

また、モノラル音声分離以外の音声音響信号処理タスクへの識別的 NMF の応用も考えられる。例えば NMF による声質変換 [42] では、NMF により学習された話者毎の基底を用いて声質変換を行うが、そのためには各話者に対して識別的な基底を用いることがより望ましいと考えられる。識別的 NMF はそのような基底の学習に役立つだろう。

第 4 章

NMF 基底間の識別性評価のための オーバーラップ尺度

前章では識別的 NMF を取り扱ったが、これは他の基底学習法に比べてどのような性質を持っているか、またその「識別性」とは何かということが課題として残されていた。もし、ある基底学習法により得られる各音源基底に対して、それらの間の「識別性」を定量的に評価できる指標が得られれば、識別的 NMF の仕組みや性質を他の手法と比較して議論できる。そこで本章では、NMF 基底間の識別性を幾何学的に測るオーバーラップ尺度を提案する。そして、それを各手法間で比較するための基準として、各音源の真の基底を用いることを提案する。真の基底を得ることは一般には難しいが、近年、最小体積 NMF を用いることによりこれを近似的に得られることが明らかになった。本章では plain NMF、識別的 NMF、最小体積 NMF の 3 種類の基底学習法によるモノラル音声分離実験において、最小体積 NMF の場合に計算されたオーバーラップ尺度を基準として、各手法のこの尺度や分離性能を比較した。その結果、識別的 NMF は plain NMF よりも、各音源の基底間のオーバーラップを小さくするという意味で「識別的」な基底を学習し、これにより音声分離の性能を向上させていると考えられることが分かった。しかし、基準である最小体積 NMF の方が識別的 NMF よりも更にオーバーラップが小さく、かつ分離性能も高いことがわかり、最小体積 NMF の方が実はより「識別的」であり、音源分離のための基底学習法として優れている可能性が示唆された。

4.1　序論

3 章のあらすじを振り返ろう。NMF による音源分離のための基底学習法として、最も単純なのが plain NMF（2.2）であった。plain NMF は各音源のクリーン信号だけを用いて学習する手法だったが、それに対して、混合信号を分離する際のことも加味して「識別的に」基底を学習する手法が識別的 NMF（3.3）であった。この識別的 NMF を解くために、3 章では拡張ラグランジュ関数法を用いることを提案し 3.7、モノラル音声分離の実験において提案手法が plain NMF や先行研究よりも分離性能が高いことを確認した。

▶ 「識別性」とは何か？：オーバーラップ尺度による定量的評価

拡張ラグランジュ関数法による識別的 NMF の定式化 (3.7) をもう一度見てみよう：

$$\min_{\mathsf{W},\mathsf{H}} \mathcal{L}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) \qquad (\mathcal{L}_{\mathsf{M}}(\mathsf{W},\mathsf{H}) = \mathcal{R}(\mathsf{W},\mathsf{H}) + \mathcal{C}_{\mathsf{M}}(\mathsf{W},\mathsf{H}))$$
$$\text{s.t. } W_1,...,W_N \in \mathbb{R}_+^{F\times K},\ H_1,...,H_N \in \mathbb{R}_+^{K\times T}$$

目的関数 $\mathcal{L}_{\mathsf{M}}$ のうち、$\mathcal{R}$ は plain NMF の目的関数であり、各音源のクリーン信号に対する再構成誤差を表す (3.1a)。一方で $\mathcal{C}_{\mathsf{M}}$ は、混合信号に対する分析誤差 (3.2a) を減少させることを促すペナルティ項であった (3.8)。3.5.2 節で述べたように、このペナルティ項は各音源の基底 $W_1,...,W_N$ が互いに「識別的」になるように何らかの正則化を掛けていると考えられる。

しかし、識別的 NMF が具体的にどのような仕組みで基底を「識別的」に学習するかについては、先行研究 [32, 40, 50] では十分な議論がなされていない。(本論文でも 3.2.1 節でその仕組みを直感的に説明したが、定量的に比較可能な議論ではない。) さらに言えば、識別的 NMF が定義する基底間の「識別性」とは何か、ということも考えられてこなかった。

もし、ある基底学習法により得られる各音源基底に対して、それらの間の「識別性」を定量的に評価する尺度を設計できれば、識別的 NMF の仕組みや性質を他の学習法と比較して議論できるし、あるいは新しく考案した別の学習法に対する評価尺度として利用することもできる。4.2.1 節で述べるように NMF の基底は幾何学的には錐包を生成するから、各音源基底が生成する錐包同士の重なりを測るオーバーラップ尺度を設計できれば、これを「識別性」の評価に使えるはずである。

▶ 識別性評価尺度の基準：最小体積 NMF

さて、そのような「識別性」の評価尺度 $\mathcal{X}$ を設計したとして、それを比較するための基準はどう取るべきだろうか？つまり、どのような基底を使って計算した $\mathcal{X}$ を基準とすれば良いだろうか？それには、各音源の振幅スペクトログラム $V_1,...,V_N$ を生成している、真の基底 $W_1^{\#},...,W_N^{\#}$ を使う必要がある。真の基底を使うことにより、分離対象である音源間の本来的な「識別性」、あるいは幾何学的な重なり具合が測れる。(本研究では各音源信号が実際に NMF 型の生成モデルに従っており、真の基底が実在すると仮定する。) この未知の真の基底は、plain NMF では得ることが難しいが、実は 4.2.2 節で述べるように、最小体積 NMF を使えば近似的に得られる。

そこで本研究ではまず、ある基底学習法によって得られた 2 音源の基底 W_1, W_2 間の「識別性」を測るオーバーラップ尺度 $\mathcal{O}(W_1, W_2)$ を、錐包の代わりに、W_1, W_2 のそれぞれの列ベクトル達が張る平行体の間の重なりを考えて定義する。そして、識別的 NMF を含む様々な基底学習法を使ってモノラル音声分離を行い、最小体積 NMF の場合の $\mathcal{O}$ やその分離性能を基準として、各手法のそれら指標を比較し、各手法の「識別性」とその実効性を定量的に評価することを提案する。

本章の構成は以下の通りである。4.2 節では準備として、NMF の幾何学的解釈と同定可能性の問題に触れ、同定可能性を持つ NMF 手法である最小体積 NMF の説明を行う。4.3 節では基底 W_1, W_2 間のオーバーラップ尺度 $\mathcal{O}(W_1, W_2)$ を定義し、W_1, W_2 が簡単な場合に $\mathcal{O}$ を計算した例を示す。4.4 節では、まずオラクルの学習音声を使って 3 種類の基底学習法 (plain NMF、拡張ラグランジュ関数法による識別的 NMF、最小体積 NMF) でモノラル音声分離を行い、これらの手法のオーバーラップ尺度 $\mathcal{O}$ と分離性能を比較する。さらに、学習音声がオラクルでない一般の場合で

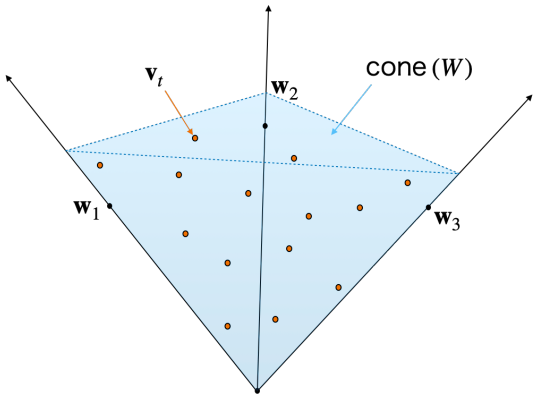


Fig. 4.1: Geometric interpretation of NMF ($K = 3$) [9]. NMF seeks to find basis W such that the cone (W) contains $\boldsymbol{v}_t$s.

もモノラル音声分離を行って、同様の比較を行う。4.5 節では結論と今後の課題を述べる。

4.2　準備

4.2.1　NMF の幾何と同定可能性

▶ NMF の幾何学的解釈

NMF の基底 W は、幾何学的には錐包を生成している。一般に、ベクトル $\boldsymbol{a}_1, ..., \boldsymbol{a}_I$ の非負結合

$$\sum_i \theta_i \boldsymbol{a}_i \quad (\theta_1, ..., \theta_I \geq 0)$$

を**錐結合**（conical combination）と呼ぶ。錐結合全体の集合を**錐包**（conical hull）と呼び、

$$\text{cone}\,(A) = \{A\theta \mid \theta \geq 0\} \quad (A \coloneqq (\boldsymbol{a}_1 \ ... \ \boldsymbol{a}_I),\ \theta \coloneqq (\theta_1 \ ... \ \theta_I))$$

と書く。NMF の式 $\boldsymbol{v}_t = W\boldsymbol{h}_t = \sum_k \boldsymbol{w}_k h_{kt}\,(\forall t)$ を見ると、$H \geq 0$ だから、V の各列ベクトル $\boldsymbol{v}_1, ..., \boldsymbol{v}_T$ は K 本の基底ベクトル $\boldsymbol{w}_1, ..., \boldsymbol{w}_K$ の錐結合となっている。すなわち、

$$\boldsymbol{v}_t \in \text{cone}\,(W) = \{W\theta \mid \theta \geq 0\} \quad (\forall t)$$

であり、$\boldsymbol{v}_1, ..., \boldsymbol{v}_T$ は基底 W の生成する錐包に含まれている（Fig. 4.1）。

▶ NMF の同定可能性

多くの場合において NMF の分解対象 $V = (\boldsymbol{v}_1 \ ... \ \boldsymbol{v}_T)$ を包含する錐包は無数にあり、V を生成した真の基底 $W^\#$ を得ることは難しい。このことを具体的に数式で述べよう。いま仮に真の基底 $W^\#$ のランク K が事前に分かっていたとして、基底の本数を K とおいた NMF により $V = \hat{W}\hat{H}$ という分解が得られたとする。このとき、A を

$$\hat{W}A^{-1} \geq 0,\ \hat{A}H \geq 0$$

を満たす任意の K 次正則行列とすると、$(\hat{W}, \hat{H})$ の変換

$$(\hat{W}, \hat{H}) \leftarrow (\hat{W}A^{-1}, A\hat{H}) \tag{4.1}$$

に対して積 $\hat{W}\hat{H}$ は不変で $= V$ となり、異なる分解が得られてしまう。

変換（4.1）に対する不定性の中でも、特に次の 2 つは本質的なものである。1 つは基底ベクトルの並び替え（permutation）に対する不定性であり、基底のインデックス $k = 1, ..., K$ の順番を決めることはできない。もう 1 つはスケーリングの不定性であり、任意の $c > 0$ に対して $(c\boldsymbol{w}_k)(c^{-1}\boldsymbol{h}_k) = \boldsymbol{w}_k\boldsymbol{h}_k$ となってしまうから、基底のスケールを定めることはできない。（以上の 2 つの不定性はアクティベーションに関するものと読み替えても良い。）まとめれば、NMF により基底 $\hat{W}$ とアクティベーション $\hat{H}$ が得られたとしても、置換とスケーリングからなる変換

$$(\hat{W}, \hat{H}) \leftarrow (\hat{W}B^{-1}, B\hat{H}) \quad (B \coloneqq \Pi\Sigma,\ \Pi\text{: a permutation matrix},\ \Sigma\text{: a positive definite diagonal matrix}) \tag{4.2}$$

に対する不定性を取り除くことは本質的に不可能である。

ある NMF 手法が、置換とスケーリングの不定性を除いて、真の基底と同じ基底を得ることができるとき、言い換えれば得られる基底とアクティベーションが、真の基底およびアクティベーション $W^\#, H^\#$ と（4.2）で結ばれているとき、その手法は**同定可能性**（identifiability）を持つという。（ただし、分解対象のデータが実際に NMF 型で生成されているとしての話である。）同定可能性についてより詳しくはレビュー [9] を参照されたい。

plain NMF（2.2）を含む多くの NMF 手法は同定可能性を持たないが、実は次節で述べるように最小体積 NMF という手法は、緩やかな条件の下で同定可能性を持つ。

4.2.2 最小体積 NMF

▶ 最小体積 NMF の定義

NMF に同定可能性を持たせるためには、真の基底とアクティベーション $W^\#, H^\#$ が何か良い性質を満たしていると仮定したり、NMF の目的関数を変えたり、W, H に非負条件以外の制約条件を付けたりすることによって、可能な変換 A（4.1）の空間を置換とスケーリング（4.2）だけまで狭めるのが常套手段である。

そのために色々な条件が考え出され、さまざまな設定の下で同定可能性が示されてきたが [9]、その中でも特に、文献 [27] は最小体積 NMF という手法が同定可能性を持つことを示した：

最小体積 NMF の同定可能性

$V \in \mathbb{R}_+^{F\times T}$, $W^\# \in \mathbb{R}_+^{F\times K}, H^\# \in \mathbb{R}_+^{K\times T}$ とする。$V = W^\#H^\#$ であり、$\mathrm{rank}(V) = K$ とする。また、$H^\#$ は sufficiently scattered condition（SSC）を満たすとする。このとき、$W \in \mathbb{R}^{F\times K}, H \in \mathbb{R}^{K\times T}$ として、最適化問題

$$\min_{W,H} \det(W^\top W) \tag{4.3a}$$

$$\text{s.t. } V = WH,\ H \geq 0,\ \mathbf{1}^\top W = \mathbf{1}^\top \tag{4.3b}$$

の最適解は置換とスケーリングに関する不定性を除いて $W^\#, H^\#$ と等しい。

この定理の証明は付録 A に書いておくが、ここで直感的な説明をしておこう。SSC（付録 A 参照）とは、直感的には真のアクティベーション $H^\#$ の各列ベクトル $\boldsymbol{h}_1^\#, ..., \boldsymbol{h}_T^\#$ が非負象限

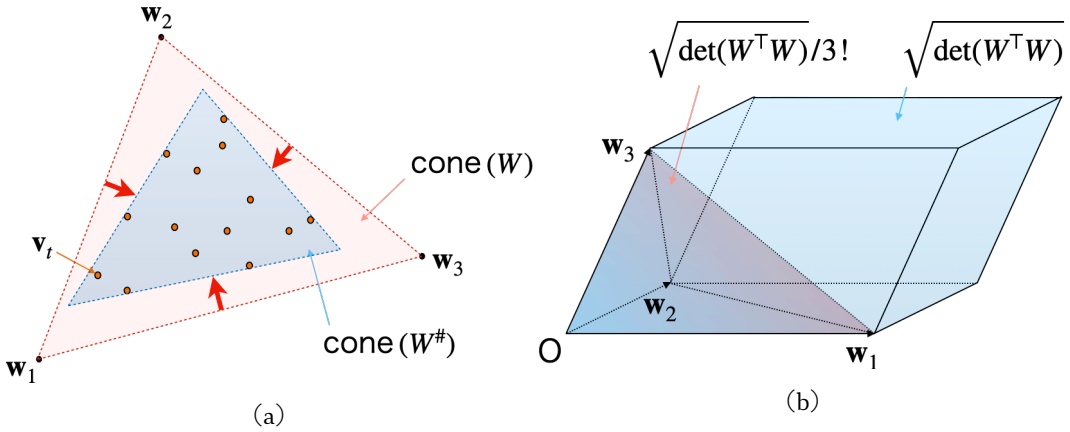


Fig. 4.2: Mechanism of minimum volume NMF ($K = 3$) [9]. (a) Minimum volume NMF seeks to find basis W with the smallest volume such that the cone (W) contains $\boldsymbol{v}_t$s. (b) The parallelpiped (blue) generated by the column vectors of W has a volume of $\sqrt{\det(W^\top W)}$. The simplex (over-painted with red) has a volume of $\sqrt{\det(W^\top W)}/K!$.

(nonnegative orthant) で十分に散らばっているという条件である。$H^\#$ は V と $V = W^\# H^\#$ の関係で結ばれているから、SSC を V に関する言葉で言い換えると、V の各列ベクトル $\boldsymbol{v}_1, ..., \boldsymbol{v}_T$ が真の基底 $W^\#$ の生成する錐包 cone $(W^\#)$ のなかに十分に散らばっているということである。(SSC が成立していることを厳密に確かめるのは NP 困難であることが示されているが [17]、現実にはデータ点の数 T が十分に大きければ満たされていると考えてよい。) このとき、$W^\#$ を知らなくても V の情報から cone $(W^\#)$ の形を推測することができるだろう。データ点群 $\boldsymbol{v}_1, ..., \boldsymbol{v}_T$ を包含するような錐包のなかで、体積が最小となるような基底 W を探せば、これは真の基底 $W^\#$ と (本質的な不定性を除いて) 一致するはずである (Fig. 4.2a)。ここで体積というのは、錐包の体積 (これは無限大) ではなく、W の各列ベクトル $\boldsymbol{w}_1, ..., \boldsymbol{w}_K$ が張る**平行体** (parallelotope) の体積 $\sqrt{\det(W^\top W)}$ (Fig. 4.2b) を使う [11]。目的関数 (4.3a) はまさにこの体積を減少させよということを示している。ただし、同定可能性を持たせるためにはさらに、基底 W の各列ベクトルが確率単体上にある (column-stochastic) という条件 $\mathbf{1}^\top W = \mathbf{1}^\top$ (4.3b) も必要であるが、これも SSC と同様、厳しい条件ではない。以上が、最小体積 NMF の同定可能性の直感的説明である。

最小体積 NMF (4.3) を近似的に解く形に置き換えよう：

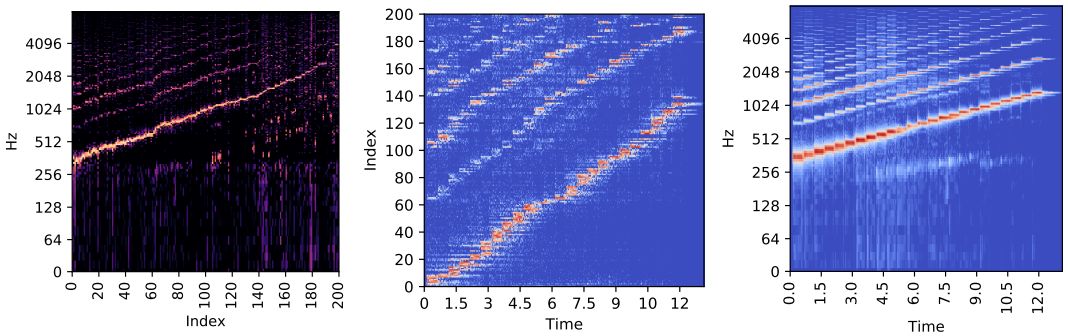
最小体積 NMF

$$\min_{W,H} D_\beta(V \mid WH) + \frac{\lambda}{2} \log\det(W^\top W + \epsilon I) \tag{4.4a}$$

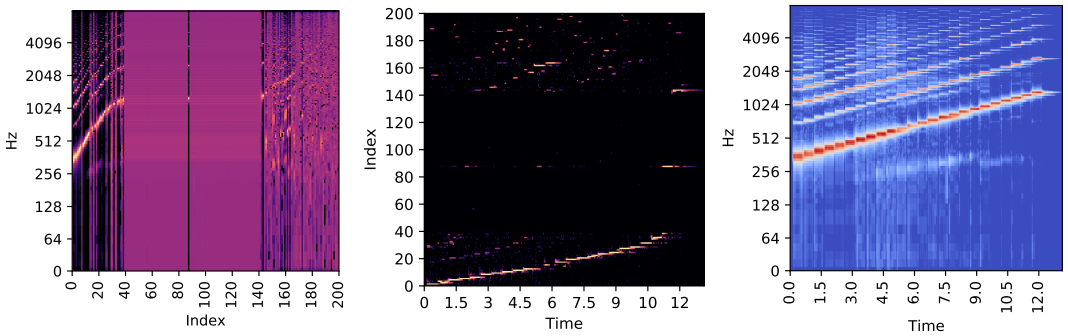
$$\text{s.t. } W \in \mathbb{R}_+^{F \times K},\ H \in \mathbb{R}_+^{K \times T}, \mathbf{1}^\top W = \mathbf{1}^\top \tag{4.4b}$$

第 1 項はデータへのフィットを表し、第 2 項は体積項である。$\lambda > 0$ は体積項の考慮の度合いを変えるペナルティパラメータである。また、W が列フルランクでない場合にも数値計算が安定するように、$W^\top W$ に $\epsilon I\,(\epsilon > 0)$ を足し、さらに log を取っている。こうすることにより、基底の本数 K を真の基底 $W^\#$ のランクより大きくとってしまっても、最小体積 NMF は上手く学習が進むことが明らかにされている [26, 27]。

本章では (4.4) による基底学習法を**最小体積 NMF** (minimum volume NMF; min-vol NMF)



(a) W.　(b) H.　(c) WH.



(d) W.　(e) H.　(f) WH.

Fig. 4.3: NMF of an amplitude spectrogram of clarinet. The audio data were obtained from songKitamura [21]. For visualization, the basis vectors were permuted using the spectral peak. The activations were also permuted accordingly. (a, b, c) Plain NMF. (d, e, f) Minimum volume NMF.

と呼ぶことにし、次節で提案するオーバーラップ尺度の基準として、この最小体積 NMF により得られる基底で計算されたものを使う。

最小体積 NMF の最適化問題（4.4）は上界最小化（2.6.2 節）とラグランジュの未定乗数法（method of Lagrange multiplier）を組み合わせることで解くことができる [28]。具体的な更新とその導出については付録 A を参照されたい。

▶ クラリネットの上昇音階に対する適用例

最小体積 NMF（4.4）によって学習された基底は、他の基底学習法、例えば plain NMF（2.2）などよりも「コンパクト」なものとなる。この 2 つの NMF をクラリネットの振幅スペクトログラムに対して適用した例を Fig. 4.3 に示す。再構成された振幅スペクトログラムはどちらもあまり変わらないが（Fig. 4.3c, 4.3f）、得られる基底とアクティベーションの性質は以下に述べるように全く異なる。

この音データには 2 オクターブ分（24 個）の上昇音階が含まれている。振幅スペクトログラムの特異値の上位 24 個は、全特異値の合計の 84.5 %を占め、ランクも 24 に近いと予想されるが、ここでは 4.4 節での実験の条件と合わせて基底の本数を $K = 200 \gg 24$ として NMF を行なった。（最小体積 NMF は、再構成された振幅スペクトログラム（Fig. 4.3f）を見ればわかるように、ランク落ちの場合でも確かにうまく動作していることがわかる。）plain NMF は、基底の本数が過剰であるためにほぼパルス状の単峰的な基底を学習してしまっており（Fig. 4.3a）、クラリネットのテ

ンプレートを全く得られていない。対応するアクティベーション（Fig. 4.3b）を見ると同時刻に複数の基底が立っており、ハーモニクスを別々の音として学習してしまっていることがわかる。一方で最小体積 NMF は、余剰の基底は体積を増加させてしまうから、これらをフラットで意味のない基底になるよう「押し潰す」ことで、分解対象を精度よく表せる最小限の本数の基底を学習する（Fig. 4.3d）。対応するアクティベーション（Fig. 4.3e）を見ると、各時刻においてはごく少数の基底しか立っておらず、実効的な基底の本数は plain NMF の場合よりも大幅に少ないことがわかる。なお、各手法のアクティベーションのスパース性を Hoyer sparsity [16] の時間平均

$$\frac{1}{T}\sum_t \frac{\sqrt{K} - \|\boldsymbol{h}_t\|_1/\|\boldsymbol{h}_t\|_2}{\sqrt{K}-1}$$

（$[0,1]$ の値をとり、大きいほどスパース性が高い）で測ってみると、plain NMF が 0.91 であるのに対して最小体積 NMF は 0.99 であった。

4.3　提案

4.3.1　NMF 基底間のオーバーラップ尺度

音源数を $N = 2$ とする。簡単のために音源 $n = 1,2$ の基底 W_1, W_2 はどちらも列フルランクだとする。（ランクは同じでなくても良い。）さらにこれらを横方向に連結した $\mathcal{W} \coloneqq (W_1\ W_2)$ も列フルランクだとする。このとき、$\mathcal{W}$ のグラミアンを考えてみよう。まずグラム行列は

$$\mathcal{W}^\top\mathcal{W} = \begin{bmatrix} W_1^\top W_1 & W_1^\top W_2 \\ W_2^\top W_1 & W_2^\top W_2 \end{bmatrix}$$

となる。$\mathcal{W}$ が列フルランクであるからグラム行列は正則で、特に $(1,1)$ 成分 $W_1^\top W_1$ も正則であるため、ブロック分けされた行列式に対する公式

$$\det\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B) \quad \text{when } A \text{ is regular}$$

を使って、グラミアンは

$$\begin{aligned} \det(\mathcal{W}^\top\mathcal{W}) &= \det(W_1^\top W_1)\det(W_2^\top W_2 - W_2^\top W_1(W_1^\top W_1)^{-1}W_1^\top W_2) \\ &= \det(W_1^\top W_1)\det(W_2^\top(I - W_1(W_1^\top W_1)^{-1}W_1^\top)W_2) \\ &= \det(W_1^\top W_1)\det((Q_1W_2)^\top(Q_1W_2)) \end{aligned} \tag{4.5}$$

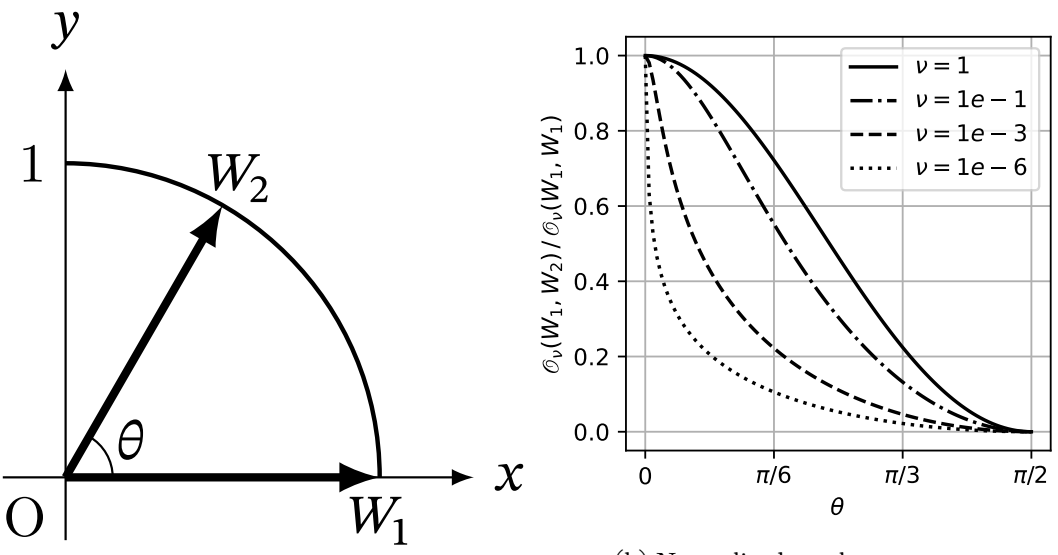
となる。ただし I は単位行列で、

$$Q_1 \coloneqq I - W_1(W_1^\top W_1)^{-1}W_1^\top$$

は W_1 の各列ベクトルが張る空間

$$\mathrm{span}(W_1) \coloneqq \mathrm{span}\{(\text{column vectors of } W_1)\}$$

の直交補空間 $\mathrm{span}(W_1)^\perp$ への直交射影子である。



(a) Problem settings.

(b) Normalized overlap measure.

Fig. 4.4: Calculation of the overlap measure $\mathcal{O}_\nu(W_1, W_2)$ for $W_1 = [1\ 0]^\top$ and $W_2 = [\cos\theta\ \sin\theta]^\top$.

今 W_1, W_2 は列フルランクと考えていたから $\det(W_1^\top W_1)\det(W_2^\top W_2) \neq 0$ なので、これでグラミアン (4.5) を割ってみると

$$\frac{\det(\mathcal{W}^\top \mathcal{W})}{\det(W_1^\top W_1)\det(W_2^\top W_2)} = \frac{\det((Q_1 W_2)^\top (Q_1 W_2))}{\det(W_2^\top W_2)} \tag{4.6}$$

となる。(4.6) の右辺の分母は、音源 2 の基底 W_2 の各列ベクトルが作る平行体の体積の 2 乗である。一方で分子は、直交補空間 $\mathrm{span}(W_1)^\perp$ に射影されたそれらのベクトルが作る平行体の体積の 2 乗である。ゆえに (4.6) は $\mathrm{span}(W_1)$ と $\mathrm{span}(W_2)$ が同じだと最小値 0 を取り、直交していると最大値 1 を取る。(以上と全く同様な議論で、$Q_2 \coloneqq I - W_2(W_2^\top W_2)^{-1} W_2^\top$ とおいて ((4.6) の左辺) $= \det((Q_2 W_1)^\top (Q_2 W_1))/\det(W_1^\top W_1)$ と書くこともできる。)

以上の考察から、(4.6) の平方根は、各音源の基底 W_1, W_2 が作る平行体の間の離れ具合を定量的に表していると考えられる。ただし、今までは $W_1, W_2, \mathcal{W}$ が列フルランクだと仮定してきたが一般にはそうではない。ランク落ちすると行列式は 0 となってしまって (4.6) は計算できない。またランク落ちしなかったとしても、行列式は固有値の掛け算であるから小さな固有値があると正確に計算することが難しい。そこで行列式の計算を安定化させるために $\mathcal{W}^\top \mathcal{W}, W_1^\top W_1, W_2^\top W_2$ に νI $(\nu > 0)$ を足して、さらに log を取って、W_1, W_2 間のオーバーラップ尺度を

オーバーラップ尺度

$$\mathcal{O}_\nu(W_1, W_2) \coloneqq -\frac{1}{2}\log\frac{\det(\mathcal{W}^\top \mathcal{W} + \nu I)}{\det(W_1^\top W_1 + \nu I)\det(W_2^\top W_2 + \nu I)} \tag{4.7}$$

と定義する。ただし、重なり具合が小さいときに小さくなるようにしたいのでマイナスを付けた。

4.3.2　簡単な例

ここでは $\mathcal{O}_\nu$ の計算の簡単な例として

$$W_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ W_2 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \left(0 \le \theta \le \frac{\pi}{2}\right)$$

の場合を考えよう（Fig. 4.4a）。$\theta \to 0$ のとき W_2 は W_1 に近づいていくから、このとき $\mathcal{O}_\nu$ は大きくなっていくことが予想される。

実際に計算してみると

$$\mathcal{O}_\nu(W_1, W_2) = -\frac{1}{2}\log\left(1 - \left(\frac{\cos\theta}{1+\nu}\right)^2\right)$$

となる。これは $\mathrm{span}(W_1) = \mathrm{span}(W_2)$ のとき、特に $W_1 = W_2$ のとき最大値を取るから、$\mathcal{O}_\nu(W_1, W_1)$ で割って

$$\frac{\mathcal{O}_\nu(W_1, W_2)}{\mathcal{O}_\nu(W_1, W_1)} = \log\left(1 - \left(\frac{\cos\theta}{1+\nu}\right)^2\right) \Big/ \log\left(1 - \left(\frac{1}{1+\nu}\right)^2\right) \quad (\in [0,1]) \tag{4.8}$$

と規格化しておく。

Fig. 4.4b に、様々な ν の値に対する (4.8) のグラフを示す。これを見ると、確かに $\theta \to 0$ のとき $\mathcal{O}_\nu(W_1, W_2)$ は最大値を取っている。また $\theta \to \pi/2$ のとき、すなわち W_1 と W_2 が直交しているとき $\mathcal{O}_\nu(W_1, W_2)$ は最小値を取っている。以上のことから、提案したオーバーラップ尺度 (4.7) は確かに基底同士の重なり具合を定量的に評価していると考えられる。

4.4　実験

本研究では 3 種類の基底学習法（plain NMF (2.2)、拡張ラグランジュ関数法による識別的 NMF (3.7)、最小体積 NMF (4.4)）を使って 2 話者（$N = 2$）のモノラル音声分離を行い、提案したオーバーラップ尺度 $\mathcal{O}$ (4.7) と、SDRi（3.4.1 節参照）で測った分離性能とを各手法で比較した。4.4.2 節では、オラクルの学習データを用いた場合、すなわち学習データからテストデータの混合音声を作って、それを分離した場合の結果を示す。4.4.3 節では、学習データとテストデータを分けた一般的な場合の結果を示す。

4.4.1　実験条件

実験データは ATR 音素バランス 503 文 [22] から作成した。このデータセットの話者は男性 6 名、女性 4 名の計 10 名からなり、A セットから J セットまでの計 503 文を各話者が発話したデータが含まれている。1 発話あたりの時間長はおおむね 4–7 s である。標本化周波数は 16 kHz であり、STFT のフレーム幅は 64 ms、フレームシフトは 16 ms とした。窓にはハニング窓を用いた。2 つの実験で共通して、混合音声を作るときは、瞬時混合を仮定してインパルス応答を畳み込まず、開始位置を揃えて SNR を 0 dB として足し合わせて作成した。

4.4.2 節では、10 名の各話者ごとに、A セットから 1 発話ずつ発話内容が互いに異なるものを選び、これら 10 発話から $\binom{10}{2} = 45$ 対の発話ペアを作った。これを学習データとテストデータの両方に用いた。

一方 4.4.3 節では、4 名の話者 FKN、FTK、MHT、MSH（FKN と FTK は女性、MHT と MSH は男性）を選び出し、分離対象の話者ペアを FKN/FTK、FKN/MHT、MHT/MSH の 3 ペアとした。各ペアに対し、学習データとして B セットの一人当たり 10 発話、テストデータとして A セットの一人当たり 50 発話を用いた。4.4.2 節の実験と同様に、混合音声を作るときはペア内の 2 話者で発話内容が異なるようにした。

3 種類の基底学習法で共通して、各音源の基底ベクトルの本数は $K = 200$ とし、基底 W とアクティベーション H は $1 + z$（z は $[0,1]$ の区間の乱数）によって生成した後、W だけ column-stochastic になるよう正規化して初期化した。拡張ラグランジュ関数法による識別的 NMF (3.7) のラグランジュ乗数 M の初期値は 0 とした。またこの識別的 NMF (3.7) と最小体積 NMF (4.4) におけるペナルティパラメータ λ については、W, H の初期値において β ダイバージェンス項とペナルティ項の比が、識別的 NMF では 1 : 1 に、最小体積 NMF では $1 : 10^{-3}$ になるように定めた。識別的 NMF の場合、3 章では最適化の途中で λ を段々と大きくしていったが（3.4.2）、そのために過剰な正則化が起こる可能性があった（3.4.3 節）。そこで本章では λ は初期値のまま固定する。最小体積 NMF (4.4) のパラメータ ϵ は 1 とした。これらのパラメータは以降の実験で使われていないデータから決定した。最適化の反復回数は 1000 回とした。なお、ゼロ除算を防ぐために、最適化中に $\mathrm{eps} = 10^{-12}$ 以下の行列要素が現れた場合それらはすべて eps に置き換えた。

距離尺度としては $\beta = 1$ の一般化 KL ダイバージェンスを用いた。これは、$\beta = 0$ では 3.3.1 節で述べたように識別的 NMF の一段階化が厳密には正当化できないこと、また $\beta = 2$ では最小体積 NMF の性能が低いこと（付録 A.3）を考慮して選んだ。

各基底学習法の間でオーバーラップ尺度を正確に比較するためには、基底のスケールを揃える必要がある。最小体積 NMF (4.4) と同様に、plain NMF でも基底 W に対して column-stochastic となるような制約を付けた。この場合の plain NMF の更新式は [28] と同様のやり方で求められる（付録 B 参照）。一方識別的 NMF (3.7) では W が column-stochastic という制約を満たしながら最適化していくことは難しい。そこで、更新式 (3.17) による最適化の反復が全て終わった後に、column-stochastic となるよう正規化した。

オーバーラップ尺度 $\mathcal{O}_\nu$ (4.7) のパラメータ ν については、Fig. 4.4b を見ると $\nu \to 0$ となるに従ってグラフは急峻になり、$\mathcal{O}_\nu$ は小さい値に偏ってしまうことが分かる。そこで本研究では $\nu = 1$ とする。

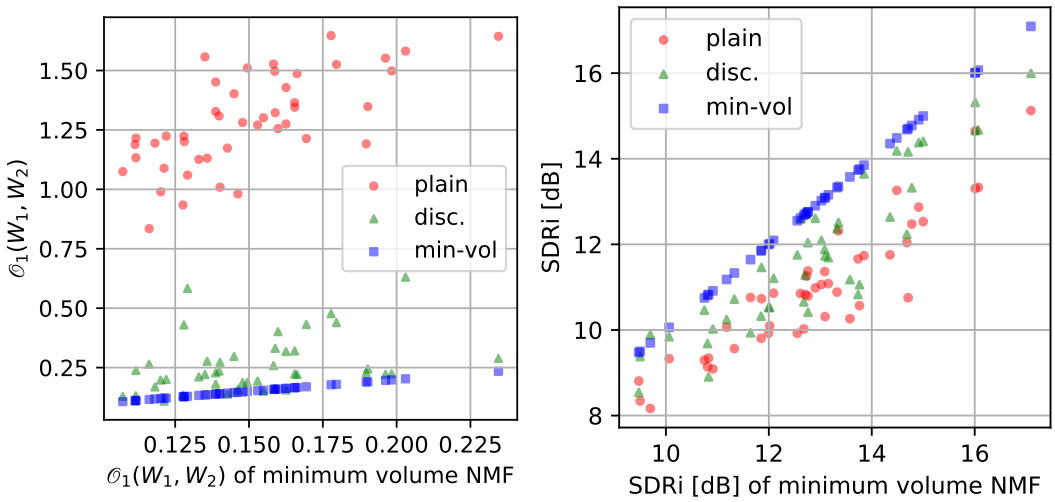
教師あり NMF (2.3) の反復回数は 200 回とした。また、話者 n の振幅スペクトログラムの推定 $\hat{V}_n$ は、混合音声 X に対してウィーナーフィルタを掛けることにより

$$\hat{V}_n = X \odot \frac{W_n H_n}{\sum_n W_n H_n}$$

と求めた。

4.4.2 モノラル音声分離における基底学習の傾向の比較（オラクルの場合）

まず、各基底学習法について、汎化能力ではなく純粋に基底学習の傾向だけを見るために、オラクルの学習データを用いてモノラル音声分離を行った。Fig. 4.5 にその結果を示す。横軸が最小体積 NMF の場合に 4.4.1 節で述べた全 45 発話ペアについて計算されたオーバーラップ尺度 $\mathcal{O}_1$（または SDRi）であり、縦軸が対応するペアに対して各手法の場合に計算された $\mathcal{O}_1$（または SDRi）



(a) Overlap $\mathcal{O}_1(W_1, W_2)$.

(b) SDRi [dB].

Fig. 4.5: Results of the monaural speech separation experiment using the oracle speech data. plain denotes plain NMF, disc. discriminative NMF and min-vol minimum volume NMF.

Table 4.1: Results of the monaural speech separation experiment. As opposed to Fig. 4.5, oracle speech data were not used for training. The abbreviations are the same as Fig. 4.5.

	FKN/FTK		FKN/MHT		MHT/MSH	
	$\mathcal{O}_1(W_1, W_2)$	SDRi [dB]	$\mathcal{O}_1(W_1, W_2)$	SDRi [dB]	$\mathcal{O}_1(W_1, W_2)$	SDRi [dB]
plain	4.78	1.28	5.28	2.32	5.47	1.09
disc.	**0.25**	2.87	0.27	4.77	**0.21**	3.16
min-vol	0.32	**4.21**	**0.26**	**7.76**	0.26	**4.60**

である。同じ発話ペアに対するプロットは同一の垂直線上に並ぶ。

Fig. 4.5a を見ると、識別的 NMF の $\mathcal{O}_1$ は plain NMF よりも有意に小さい。また Fig. 4.5b を見ると、識別的 NMF の方が plain NMF よりも SDRi が平均的に高い。これらのことから、識別的 NMF は各音源基底間のオーバーラップを小さくするという意味で「識別的」な基底を学習し、これにより分離性能を向上させている可能性が考えられる。一方で、基準である最小体積 NMF の $\mathcal{O}_1$ は識別的 NMF よりもさらに小さく（Fig. 4.5a）、SDRi はほとんどのペアで最大である（Fig. 4.5b）。これは、最小体積 NMF によって近似的に得られる真の基底は、実のところ識別的 NMF よりも十分に「識別的」であり、最小体積 NMF の方が分離のための基底学習法として優れているという可能性を示唆する。全体的な傾向として、オラクルの学習データを使った場合、$\mathcal{O}_1$ は plain NMF > 識別的 NMF > 最小体積 NMF となり、SDRi は plain NMF < 識別的 NMF < 最小体積 NMF となった。

4.4.3　モノラル音声分離における性能比較

続いて、オラクルの学習データではなく、テストデータ中に含まれない音声を学習データとして使った一般的な設定で、各基底学習法でモノラル音声分離を行った。その結果を Table 4.1 に示す。FKN/FTK、FKN/MHT、MHT/MSH の 3 ペアに対して 3 手法で音声分離したときの、学

習された各音源基底間のオーバーラップ尺度 $\mathcal{O}_1(W_1, W_2)$ と SDRi を載せている。3 手法の中で、$\mathcal{O}_1(W_1, W_2)$ については最小値を、SDRi については最大値を太字で示してある。

Table 4.1 を見ると、4.4.3 節と同様に plain NMF よりも識別的 NMF の方が $\mathcal{O}_1$ が小さく、SDRi が高いことが確認できる。一方で、基準である最小体積 NMF の $\mathcal{O}_1$ は識別的 NMF と同程度であり、さらに SDRi は識別的 NMF よりも、同性ペア（FKN/FTK、MHT/MSH）の場合は 1.3 dB 以上高く、異性ペア（FKN/MHT）の場合は約 3 dB も高い。4.4.3 節で述べたように、音源分離のための基底学習法として識別的 NMF よりも最小体積 NMF の方が優れていると考えられる。

4.5 結論

4.5.1 まとめ

本章では、ある基底学習法により得られる 2 音源の基底 W_1, W_2 間の「識別性」を幾何学的に測るオーバーラップ尺度 $\mathcal{O}_\nu(W_1, W_2)$ を提案し、それを用いて各手法を比較する際の基準として最小体積 NMF を用いることを提案した。モノラル音声分離における比較の結果、識別的 NMF は plain NMF よりも $\mathcal{O}_\nu$ を小さくするという意味で「識別的」な基底を学習し、これにより分離性能を向上させていると考えられることが分かった。一方で基準である最小体積 NMF は識別的 NMF よりもさらに $\mathcal{O}_\nu$ が小さく、分離性能は最も高いことが分かり、最小体積 NMF の方が実はより「識別的」な基底を学習でき、音源分離のための基底学習法として優れているという可能性が示唆された。

4.5.2 今後の課題

今後の課題として、提案されたオーバーラップ尺度 $\mathcal{O}_\nu(W_1, W_2)$、特に最小体積 NMF の場合に計算された $\mathcal{O}_\nu$ と、音源ペアの分離のしやすさとの間の関係性を調べることが挙げられる。もしこの 2 つの間に明白な相関関係があれば、識別的 NMF が plain NMF よりも $\mathcal{O}_\nu$ を小さくすることと、分離性能を向上させているということとの間に**実際に**因果関係があることが明確に言える。

これを調べるためには、ある音源ペアの分離のしやすさに影響を与える要因を切り分ける必要がある。$\mathcal{O}_\nu(W_1, W_2)$ は基底を使って計算された量であるから、各音源間の周波数領域での重なりを評価していると考えることができるので、分離のしやすさに対して何らかの関係性を持つはずである。一方で、時間領域での重なり、つまりどのタイミングで各音源が有音であるかということも分離のしやすさに影響を与える要因である。本章の実験では、4.4.1 節で述べたように各音源音声の開始位置を揃えて混合音声を作ったため、この時間領域での重なりは排除しなかった。考えられる要因をきちんと切り分けて、$\mathcal{O}_\nu$ と分離のしやすさとの間の関係性が見つけられれば、識別的 NMF に限らない様々な基底学習法を、$\mathcal{O}_\nu$ を使ってより系統的かつ明瞭に評価することができるだろう。

第 5 章
結論

5.1　まとめ

本論文の成果は次の 2 つにまとめられる（3.5.1 節、4.5.1 節）。

1. 識別的 NMF の最適化手法の改善を行なった（3 章）。先行研究のペナルティ法による識別的 NMF（3.4）に用いる等式制約として、最適性条件に基づく理論的により自然なもの（3.6）を提案し、さらにペナルティ法の代わりに拡張ラグランジュ関数法（3.7）を用いることで最適化の高速化を行った。提案手法は先行研究よりも等式制約が高速に収束することが確認され（3.4.2 節）、さらにモノラル音声分離実験の結果、提案手法の方がより良い分離性能を示すことが確認された（3.4.3 節）。
2. 識別的 NMF の基底学習の傾向や性能の評価を行なった（4 章）。一般の基底学習法に対して、それにより学習される各音源基底間の幾何学的な重なり具合を評価するオーバーラップ尺度（4.7）を提案し、それを「識別性」の評価尺度として用いることを提案した。また、オーバーラップ尺度を手法間で比較する際の基準として、各音源の真の基底を用いることを提案し、更にその未知の真の基底を近似的に得る方法として最小体積 NMF を用いることを提案した。モノラル音声分離実験の結果、識別的 NMF は plain NMF よりもオーバーラップを小さくするという意味で「識別的」な基底を学習し、これにより分離性能を向上させていると考えられることがわかった。一方で基準である最小体積 NMF は識別的 NMF よりもさらにオーバーラップが小さく、分離性能もより高いことがわかり、最小体積 NMF の方が実はより「識別的」であり、音源分離のための基底学習法として優れている可能性が示唆された（4.4 節）。

5.2　今後の課題

今後の課題としては次の 2 つが考えられる（3.5.2 節、4.5.2 節）。

1. NMF 声質変換［42］など、モノラル音声分離以外の音声音響信号処理タスクへの、識別的 NMF（3.3）や最小体積 NMF（4.4）の応用。
2. 各音源間の周波数領域や時間領域での重なりと、音源分離のしやすさとの関係性の調査。特に、提案したオーバーラップ尺度（4.7）との相関。

付録 **A**

最小体積 NMF

最小体積 NMF について 4 章で省略したことをここに書いておく。A.1 節では最小体積 NMF の同定可能性の証明を与えておく。A.2 節では距離尺度として一般化 KL ダイバージェンスを用いた場合の更新式を導出する。A.3 節ではユークリッド距離を用いた場合の更新式を導出する。

A.1 同定可能性

最小体積 NMF の同定可能性の定理（4.2.2 節）を再掲する：

最小体積 NMF の同定可能性（再掲）

$V \in \mathbb{R}_+^{F\times T}$, $W^\# \in \mathbb{R}_+^{F\times K}, H^\# \in \mathbb{R}_+^{K\times T}$ とする。$V = W^\# H^\#$ であり、$\mathrm{rank}(V) = K$ とする。また、$H^\#$ は sufficiently scattered condition（SSC）を満たすとする。このとき、$W \in \mathbb{R}^{F\times K}, H \in \mathbb{R}^{K\times T}$ として、最適化問題

$$\min_{W,H} \det(W^\top W) \tag{A.1a}$$

$$\text{s.t. } V = WH, \tag{A.1b}$$

$$H \geq 0, \tag{A.1c}$$

$$\mathbf{1}^\top W = \mathbf{1}^\top \tag{A.1d}$$

の最適解は置換とスケーリングに関する不定性を除いて $W^\#, H^\#$ と等しい。

ここで SSC の定義を与える前に、錐に関する基本事項を確認しておく。以下の用語は基本的に [54] に従っている。空でない集合 $C \in \mathbb{R}^N$ は、

$$\forall x \in C,\ \forall \lambda \geq 0 \qquad \lambda x \in C$$

を満たすとき**錐**（cone）という。例えば 4.2.1 節で述べた、NMF の基底 $W = (\boldsymbol{w}_1 \dots \boldsymbol{w}_K)(\boldsymbol{w}_1, \dots, \boldsymbol{w}_K \in \mathbb{R}_+^F)$ が生成する錐包

$$\mathrm{cone}\,(W) = \{W\boldsymbol{x} \mid \boldsymbol{x} \geq 0\}$$

も錐である。錐 C に対して

$$C^* \coloneqq \{\boldsymbol{y} \in \mathbb{R}^N \mid \boldsymbol{y}^\top \boldsymbol{x} \geq 0\ (\forall \boldsymbol{x} \in C)\}$$

を C の**双対錐**（dual cone）といい、C^* もまた錐となる。一般に、2 つの錐 C_1, C_2 について

$$C_1 \subset C_2 \Rightarrow C_2^* \subset C_1^* \tag{A.2}$$

が成り立つ（証明は例えば [55]）。

さて、SSC の定義は次の通りである：

Sufficiently Scattered Condition（SSC）

$H \in \mathbb{R}_+^{K\times T}$ は次の 2 条件を満たすとき sufficiently scattered であるという。

$$\text{(1) } \mathcal{C} \in \operatorname{cone}(H) \tag{A.3a}$$

$$\text{(2) } \operatorname{cone}(H)^* \cap \operatorname{bd} \mathcal{C}^* = \{\lambda \boldsymbol{e}_k \mid \lambda \geq 0,\ k = 1, \ldots, K\} \tag{A.3b}$$

ただし、$\mathcal{C} \coloneqq \{\boldsymbol{x} \in \mathbb{R}^K \mid \mathbf{1}^\top \boldsymbol{x} \geq \sqrt{K-1}\|\boldsymbol{x}\|_2\}$, $\mathcal{C}^* = \{\boldsymbol{x} \in \mathbb{R}^K \mid \mathbf{1}^\top \boldsymbol{x} \geq \|\boldsymbol{x}\|_2\}$, $\operatorname{cone}(H) = \{H\boldsymbol{x} \mid \boldsymbol{x} \geq 0\}$ であり、bd は集合の境界を、$\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K$ は $\mathbb{R}^K$ の標準基底を表す。

この条件の具体的な意味についてはレビュー [9] を参照されたい。

以上の準備のもとに、定理の証明を与えよう。証明は二段階に分けられる。（以下の証明は [8, 27] に基づく。）

▶ 第一段階

$\hat{W}, \hat{H}$ を最適化問題 (A.1) の実行可能解とする。

このとき (A.1b) より $V = \hat{W}\hat{H}$ であり、$K = \operatorname{rank}(V)$ であることと積のランクに関する不等式 $\operatorname{rank}(V) \leq \operatorname{rank}(\hat{W}), \operatorname{rank}(\hat{H})$ とから $\operatorname{rank}(\hat{W}) = \operatorname{rank}(\hat{H}) = K$。同様に $\operatorname{rank}(W^\#) = \operatorname{rank}(H^\#) = K$。

特に $\hat{W}$ は列フルランクなので左逆 L があって、

$$\hat{H} = L\hat{W}\hat{H} = LW^\# H^\#$$

となる。$\operatorname{rank}(\hat{H}) = \operatorname{rank}(H^\#) = K$ なので $LW^\#$ も可逆で、$A \coloneqq LW^\#$ とおけば、上の式と (A.1b) から

$$\hat{W} = W^\# A^{-1},\ \hat{H} = AH^\# \tag{A.4}$$

と書ける。可逆行列 A が置換とスケーリングに限ることを示せば良い。

基底が column-stochastic という条件 (A.1d) と (A.4) から

$$\mathbf{1}^\top = \mathbf{1}^\top \hat{W} = \mathbf{1}^\top W^\# A^{-1} \quad \therefore \quad \mathbf{1}^\top A = \mathbf{1}^\top \tag{A.5}$$

となり、A も column-stochastic であることが分かる。ただしここで $\mathbf{1}^\top W^\# = \mathbf{1}^\top$ を仮定した。これは真の基底 $W^\#$ の本質的なスケーリング不定性である。

また、(A.1c) と (A.4) から $\hat{H} = AH^\# \geq 0$ だから A の各行ベクトル $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K \in \mathbb{R}^K$ は

$$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K \in \operatorname{cone}(H^\#)^* \tag{A.6}$$

となる。さらに SSC の（A.3a）と双対錐に関する補題（A.2）から $\mathrm{cone}\,(H^{\#})^* \subset \mathcal{C}^*$ だから、

$$\boldsymbol{a}_1, ..., \boldsymbol{a}_K \in \mathcal{C}^* \quad \therefore \quad \|\boldsymbol{a}_k\|_2 \leq \mathbf{1}^\top \boldsymbol{a}_k \ (\forall k) \tag{A.7}$$

となる。

（A.5）と（A.7）より

$$|\det(A)| \leq \prod_k \|\boldsymbol{a}_k\|_2 \quad (\because \text{Hadamard の不等式。等号は } \boldsymbol{a}_1, ..., \boldsymbol{a}_K \text{ が直交}) \tag{A.8}$$

$$\leq \prod_k \mathbf{1}^\top \boldsymbol{a}_k \quad (\because \text{（A.7）。等号は } \boldsymbol{a}_1, ..., \boldsymbol{a}_K \in \mathrm{bd}\,\mathcal{C}^*) \tag{A.9}$$

$$\leq \left(\sum_k \mathbf{1}^\top \boldsymbol{a}_k / K \right)^K \quad (\because \text{相加相乗不等式。等号は } \mathbf{1}^\top \boldsymbol{a}_1 = ... = \mathbf{1}^\top \boldsymbol{a}_K) \tag{A.10}$$

$$= 1 \quad (\because \text{（A.5）}) \tag{A.11}$$

が成り立つ。

等号が全て成り立つとすると、（A.9）と（A.6）と SSC（A.3b）から、

$$\boldsymbol{a}_1, ..., \boldsymbol{a}_K \in \{\lambda \boldsymbol{e}_l \mid \lambda \geq 0,\ l = 1, ..., K\}$$

となる。さらに（A.8）から $\boldsymbol{a}_1, ..., \boldsymbol{a}_K$ は直交し、（A.10）から $\mathbf{1}^\top \boldsymbol{a}_1 = ... = \mathbf{1}^\top \boldsymbol{a}_K$ だから、

$$A = \Pi \quad (\Pi \text{ は } K \text{ 次置換行列})$$

とならなければいけないことが分かる。

▶ 第二段階

逆に等号が 1 つでも成り立たないとすると、（A.11）から

$$|\det(A)| < 1$$

となる。このとき（A.4）より

$$\det(\hat{W}^\top \hat{W}) = |\det(A)|^{-2} \det(W^{\#\top} W^{\#}) > \det(W^{\#\top} W^{\#})$$

となるが、これは $\hat{W}$ が（A.1）の解であることに反する。よって等号は成り立たなければならず、A が置換行列に限ることが示せた（証明終）。

A.2　一般化 KL ダイバージェンスを用いたときの更新式

最小体積 NMF（4.4）において、$\beta = 1$ の一般化 KL ダイバージェンスを用いた場合の更新式を求めよう。この最適化問題は

$$\min_{W,H} D_{\mathrm{KL}}(V \mid WH) + \frac{\lambda}{2} \log\det(W^\top W + \epsilon I) \tag{A.12a}$$

$$\text{s.t. } W \in \mathbb{R}_+^{F \times K},\ H \in \mathbb{R}_+^{K \times T},\ \mathbf{1}^\top W = \mathbf{1}^\top \tag{A.12b}$$

となる。これは以下に示すように、上界最小化（2.6.2 節）とラグランジュの未定乗数法の組み合わせにより解くことができる［28］。なお、アクティベーション H の更新式は plain NMF のもの（2.6b）と同じであるため、基底 W の更新式だけ求める。

まず、目的関数（A.12a）の代理関数を求めよう。第一項はイェンセンの不等式を用いることで

$$\begin{aligned} D_{\mathrm{KL}}(V \mid WH) &= \sum_{ft}\left(\boldsymbol{w}_f\boldsymbol{h}_t - v_{ft} + v_{ft}\log\frac{v_{ft}}{\boldsymbol{w}_f\boldsymbol{h}_t}\right) \\ &\leq \sum_{ft}\left(\boldsymbol{w}_f - \frac{v_{ft}}{\boldsymbol{w}'_f\boldsymbol{h}_t}(\boldsymbol{w}'_f \odot \log \boldsymbol{w}'_f)\right) + \text{const.} \end{aligned} \tag{A.13}$$

と抑えられる。（W に依存しない定数項を省いた。$\boldsymbol{w}_f \in \mathbb{R}^{1\times K}, \boldsymbol{h}_t \in \mathbb{R}^K$ はそれぞれ W の第 f 行ベクトル、H の第 t 列ベクトルである。）等号は $W' = W$ のとき成り立つ。次に体積項は、

$$\frac{\lambda}{2}\log\det(W^\top W + \epsilon I) \leq \lambda\sum_f\left[\boldsymbol{w}_f Y \boldsymbol{w}_f'^{\top} + \frac{1}{2}(\boldsymbol{w}_f - \boldsymbol{w}'_f)\operatorname{diag}(\boldsymbol{\phi}_f)(\boldsymbol{w}_f - \boldsymbol{w}'_f)^\top\right] + \text{const.} \tag{A.14}$$

$$\left(Y \coloneqq (W'^\top W' + \epsilon I)^{-1},\ Y^+ \coloneqq \max\{Y, 0\},\ Y^- \coloneqq \max\{-Y, 0\},\ \boldsymbol{\phi}_f \coloneqq \frac{\boldsymbol{w}'_f(Y^+ + Y^-)}{\boldsymbol{w}'_f}\right)$$

と上から抑えられる（導出は［27, 41］を参照）。等号は $W' = W$ のとき成り立つ。以上の不等式（A.13）と（A.14）から、目的関数（A.12a）の代理関数が

$$\begin{aligned} L(W \mid W') \coloneqq &\sum_{ft}\left(\boldsymbol{w}_f - \frac{v_{ft}}{\boldsymbol{w}'_f\boldsymbol{h}_t}(\boldsymbol{w}'_f \odot \log \boldsymbol{w}'_f)\right) \\ &+ \lambda\sum_f\left[\boldsymbol{w}_f Y \boldsymbol{w}_f'^{\top} + \frac{1}{2}(\boldsymbol{w}_f - \boldsymbol{w}'_f)\operatorname{diag}(\boldsymbol{\phi}_f)(\boldsymbol{w}_f - \boldsymbol{w}'_f)^\top\right] + \text{const.} \end{aligned}$$

と求められる。

この代理関数 L を、基底が column-stochastic という条件（A.12b）の下で最小化する。ラグランジュ乗数 $\boldsymbol{\mu} \in \mathbb{R}^K$ を導入して、ラグランジアンを

$$L_\mu(W \mid W') \coloneqq L + \left(\sum_f \boldsymbol{w}_f - \mathbf{1}^\top\right)\boldsymbol{\mu} + \text{const.}$$

で定義する。L_μ を $\boldsymbol{w}_f$ で偏微分し $= 0$ とおくと、

$$\begin{aligned} 0 = \frac{\partial L_\mu}{\partial \boldsymbol{w}_f} &= \sum_t\left(\mathbf{1}^\top - \frac{v_{ft}}{\boldsymbol{w}'_f\boldsymbol{h}_t}\frac{\boldsymbol{w}'_f}{\boldsymbol{w}_f}\right)\odot \boldsymbol{h}_t^\top + \lambda[\boldsymbol{w}'_f Y + (\boldsymbol{w}_f - \boldsymbol{w}'_f)\odot\boldsymbol{\phi}_f] + \boldsymbol{\mu}^\top \\ \therefore\quad &\lambda\boldsymbol{\phi}_f \odot \boldsymbol{w}_f^2 + \left(\mathbf{1}^\top H^\top - 2\lambda\boldsymbol{w}'_f Y^- + \boldsymbol{\mu}^\top\right)\odot\boldsymbol{w}_f - \boldsymbol{w}'_f \odot \left(\left(\frac{V}{W'H}\right)_f H^\top\right) = 0 \end{aligned}$$

となる。$\forall f$ について上式を解くと、解が

$$W^*(\boldsymbol{\mu}) \coloneqq W' \odot \frac{[(C + \mathbf{1}\boldsymbol{\mu}^\top)^2 + S]^{\frac{1}{2}} - (C + \mathbf{1}\boldsymbol{\mu}^\top)}{D}$$

と求められる。ただし

$$\begin{aligned} C &\coloneqq JH^\top - 2\lambda W'Y^- \\ D &\coloneqq 2\lambda W'(Y^+ + Y^-) \\ S &\coloneqq 2D \odot \left(\frac{V}{W'H}H^\top\right) \end{aligned}$$

である。

等式制約を満たす $\boldsymbol{\mu}$ は**ニュートンラフソン法**（Newton-Raphson method）により求める。解くべき方程式は

$$0 = \gamma(\boldsymbol{\mu}) \coloneqq \mathbf{1}^\top W^*(\boldsymbol{\mu}) - \mathbf{1}^\top$$

である。適当な停止条件（例えば $|\gamma| < 10^{-9}$）が満たされるまで

$$\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} - \frac{\gamma}{\nabla\gamma}$$

という更新を繰り返すことで解 $\boldsymbol{\mu}^*$ を求められる。なお、

$$\nabla\gamma^\top = \mathbf{1}^\top \left\{ \frac{W'}{D} \odot \left[\frac{C + \mathbf{1}\boldsymbol{\mu}^\top}{\sqrt{(C + \mathbf{1}\boldsymbol{\mu}^\top)^2 + S}} - J \right] \right\}$$

である。

得られた $\boldsymbol{\mu}^*$ を用いて基底を

$$W \leftarrow W^*(\boldsymbol{\mu}^*)$$

と更新する。以上が基底 W の更新式である。

A.3　ユークリッド距離を用いたときの更新式

次にユークリッド距離を用いた場合を考えよう。この最適化問題は

$$\min_{W,H} D_{\mathrm{EUC}}(V \mid WH) + \frac{\lambda}{2} \log\det(W^\top W + \epsilon I) \tag{A.15}$$
$$\text{s. t. } W \in \mathbb{R}_+^{F\times K},\ H \in \mathbb{R}_+^{K\times T},\ \mathbf{1}^\top W = \mathbf{1}^\top$$

となる。例のごとく基底の更新式だけ求める。

前節と同様に、上界最小化とラグランジュの未定乗数法の組み合わせにより更新式を導くことができる。ただし、一般化 KL ダイバージェンスのときと違って、等式制約 $\mathbf{1}^\top W^*(\boldsymbol{\mu}) = \mathbf{1}^\top$ を満たすラグランジュ乗数 $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ は解析的に求まり、ニュートン法を用いる必要はない。しかし、この解析解 $W^*(\boldsymbol{\mu}^*)$ は負の成分を含む可能性があり、基底 W の非負性を保証するためにバックトラッキングのような操作を組み込む必要がある。

目的関数 (A.15) の代理関数を求めよう。第一項は、イェンセンの不等式を用いることで

$$D_{\mathrm{EUC}}(V \mid WH) = \sum_{ft} (\boldsymbol{w}_f \boldsymbol{h}_t - v_{ft})^2 \le \sum_{ft} \left[\frac{\boldsymbol{w}_f^2}{\boldsymbol{w}'_f} v'_{ft} - 2\boldsymbol{w}_f v_{ft} \right] \boldsymbol{h}_t + \text{const.} \quad (v'_{ft} \coloneqq \boldsymbol{w}'_f \boldsymbol{h}_t) \tag{A.16}$$

と抑えられる。（W に依存しない定数項を省いた。）等号は $W' = W$ のとき成り立つ。体積項は、一般化 KL ダイバージェンスのときと同様である。不等式（A.16）と（A.14）を合わせて、目的関数 (A.15) の代理関数は

$$L(W \mid W') \coloneqq \sum_{ft} \left[\frac{\boldsymbol{w}_f^2}{\boldsymbol{w}'_f} v'_{ft} - 2\boldsymbol{w}_f v_{ft} \right] \boldsymbol{h}_t + \lambda \sum_f \left[\boldsymbol{w}_f Y \boldsymbol{w}'^\top_f + \frac{1}{2} (\boldsymbol{w}_f - \boldsymbol{w}'_f) \operatorname{diag}(\boldsymbol{\phi}_f) (\boldsymbol{w}_f - \boldsymbol{w}'_f)^\top \right] + \text{const.}$$

と求められる。

この代理関数 L を基底が column-stochastic という条件（A.12b）の下で最小化する。ラグランジュ乗数 $\boldsymbol{\mu} \in \mathbb{R}^K$ を導入して、ラグランジアンを

$$L_\mu(W \mid W') \coloneqq L + \left(\sum_f \boldsymbol{w}_f - \mathbf{1}^\top\right)\boldsymbol{\mu} + \text{const.}$$

で定義する。L_μ を $\boldsymbol{w}_f$ で偏微分し $= 0$ とおくと、

$$\begin{aligned} 0 = \frac{\partial L_\mu}{\partial \boldsymbol{w}_f} &= \sum_t \left(\frac{\boldsymbol{w}_f}{\boldsymbol{w}'_f} \cdot 2v'_{ft} - 2v_{ft}\right) \odot \boldsymbol{h}_t^\top + \lambda[\boldsymbol{w}'_f Y + (\boldsymbol{w}_f - \boldsymbol{w}'_f) \odot \boldsymbol{\phi}_f] + \boldsymbol{\mu}^\top \\ &= \left[\frac{2\boldsymbol{v}'_f H^\top + \lambda \boldsymbol{w}'_f (Y^+ + Y^-)}{\boldsymbol{w}'_f}\right] \odot \boldsymbol{w}_f - 2\boldsymbol{v}_f H^\top - 2\lambda \boldsymbol{w}'_f Y^- + \boldsymbol{\mu}^\top \end{aligned}$$

となる。$\forall f$ について上式を解くと、解が

$$W^*(\mu) = W' \odot \frac{2(VH^\top + \lambda W' Y^-) - \mathbf{1}\boldsymbol{\mu}^\top}{2V'H^\top + \lambda W'(Y^+ + Y^-)}$$

と求められる。

等式制約を満たす $\boldsymbol{\mu}$ を求めよう。

$$A \coloneqq 2(VH^\top + \lambda W' Y^-),\ B \coloneqq 2V'H^\top + \lambda W'(Y^+ + Y^-)$$

とおいて、$W^*(\mu) = W' \odot (A - \mathbf{1}\boldsymbol{\mu}^\top)/B$ と書き直す。すると

$$\begin{aligned} &0 = \mathbf{1}^\top W^*(\boldsymbol{\mu}) - \mathbf{1}^\top = \mathbf{1}^\top \left(W' \odot \frac{A - \mathbf{1}\boldsymbol{\mu}^\top}{B}\right) - \mathbf{1}^\top = \mathbf{1}^\top \left(W' \odot \frac{A}{B}\right) - \boldsymbol{\mu}^\top \odot \left(\mathbf{1}^\top \frac{W'}{B}\right) - \mathbf{1}^\top \\ &\therefore \quad (\boldsymbol{\mu}^*)^\top = \frac{\mathbf{1}^\top (W' \odot A/B) - \mathbf{1}^\top}{\mathbf{1}^\top (W'/B)} \end{aligned}$$

と求まる。

解析解 $W^*(\boldsymbol{\mu}^*)$ は行列要素が負の値を取り得る。そこで、現在の値 W' と $W^*(\boldsymbol{\mu}^*)$ の内分点を取ることで、要素が全て非負になるようにする。ステップ幅を大きくするために、$W^*(\boldsymbol{\mu}^*)$ にできるだけ近い内分点を取りたいので、

$$\min\ \alpha \quad \text{s.t. } 0 \le \alpha \le 1,\ \alpha W' + (1 - \alpha) W^*(\boldsymbol{\mu}^*) \ge 0$$

という最適化問題を解く。これは

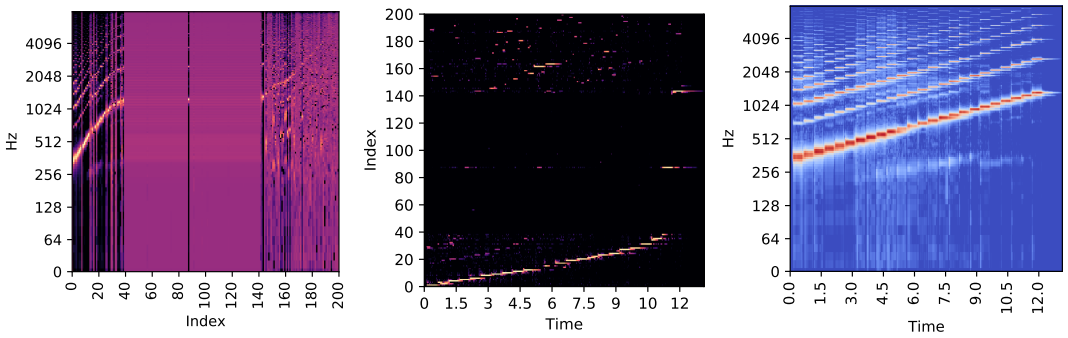
$$\gamma \coloneqq \min_{f,k} \left\{\left(\frac{W^*(\boldsymbol{\mu}^*)}{W'}\right)_{fk}\right\}$$

とおいて、

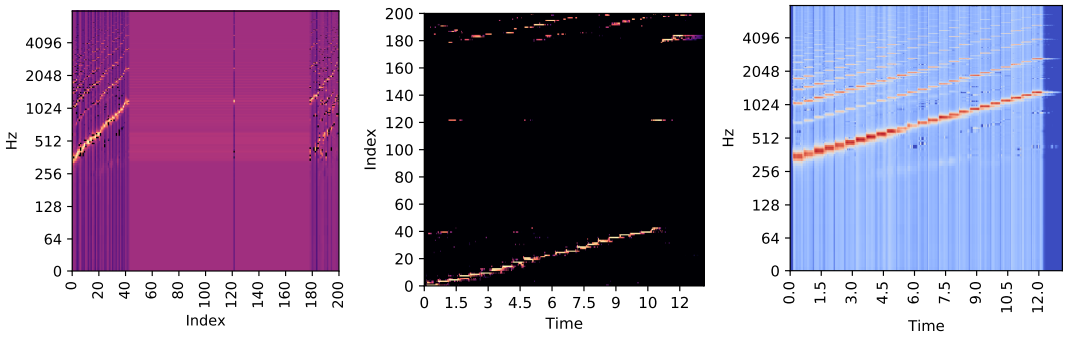
$$\alpha^* = \begin{cases} 0 & (\gamma \ge 0) \\ \dfrac{-\gamma}{1 - \gamma} & (\text{otherwise}) \end{cases}$$

と求められる。これを用いて、基底 W の更新式が

$$W \leftarrow \alpha^* W' + (1 - \alpha^*) W^*(\boldsymbol{\mu}^*) = \begin{cases} W^*(\boldsymbol{\mu}^*) & (\gamma \ge 0) \\ \dfrac{W^*(\boldsymbol{\mu}^*) - \gamma W'}{1 - \gamma} & (\text{otherwise}) \end{cases}$$



(a) W.　(b) H.　(c) WH.



(d) W.　(e) H.　(f) WH.

Fig. A.1: Minimum volume NMF of the amplitude spectrogram of clarinet in section 4.2.2. For visualization, the basis vectors were permuted using the spectral peak. The activations were also permuted accordingly. (a, b, c) Generalized KL divergence (same as (d, e, f) in Fig. 4.3). (d, e, f) Euclidean distance.

と求まる。

▶ クラリネットの上昇音階に対する適用例

Fig. A.1 に、4.2.2 節で触れたクラリネットの上昇音階に対する、ユークリッド距離を用いた最小体積 NMF の適用例を示す。比較のために一般化 KL ダイバージェンスを用いた場合も載せてある。

ユークリッド距離の場合には、2.6.1 節で述べたようにスペクトルのピークにより大きな比重が置かれる。そのためか、基底（Fig. A.1d）や再構成された振幅スペクトログラム（Fig. A.1f）には、振幅の小さい周波数成分にノイズが乗ってしまっており、復元した音を聞いてみてもホワイトノイズのような音が聞こえた。ゆえに 4.4 節の実験では一般化 KL ダイバージェンスを用いた。

付録 B
column-stochastic plain NMF の更新式

最小体積 NMF（付録 A）の場合と同様に、上界最小化（2.6.2 節）とラグランジュの未定乗数法の組み合わせにより最小化できる。

最適化問題は

$$\min_{W,H} D_{\mathrm{KL}}(V \mid WH) \tag{B.1a}$$

$$\text{s.t. } W \in \mathbb{R}_+^{F\times K},\ H \in \mathbb{R}_+^{K\times T},\ \mathbf{1}^\top W = \mathbf{1}^\top \tag{B.1b}$$

となる。付録 A と同様、アクティベーション H の更新式は plain NMF のもの（2.6b）と同じであるので、基底 W の更新式だけ求める。

一般化 KL ダイバージェンスは（A.13）のように上から抑えられたから、目的関数（B.1a）の代理関数は

$$L(W \mid W') \coloneqq \sum_{ft}\left(\boldsymbol{w}_f - \frac{v_{ft}}{\boldsymbol{w}'_f\boldsymbol{h}_t}(\boldsymbol{w}'_f \odot \log \boldsymbol{w}'_f)\right) + \text{const.}$$

となる。

この代理関数 L を、基底が column-stochastic という条件（B.1b）の下で最小化する。ラグランジュ乗数 $\boldsymbol{\mu} \in \mathbb{R}^K$ を導入して、ラグランジアンを

$$L_\mu(W \mid W') \coloneqq L + \left(\sum_f \boldsymbol{w}_f - \mathbf{1}^\top\right)\boldsymbol{\mu} + \text{const.}$$

で定義する。L_μ を $\boldsymbol{w}_f$ で偏微分し $= 0$ とおくと、

$$0 = \frac{\partial L_\mu}{\partial \boldsymbol{w}_f} = \sum_t\left(\mathbf{1}^\top - \frac{v_{ft}}{\boldsymbol{w}'_f\boldsymbol{h}_t}\frac{\boldsymbol{w}'_f}{\boldsymbol{w}_f}\right)\odot \boldsymbol{h}_t^\top + \boldsymbol{\mu}^\top$$

$$\therefore\quad (\mathbf{1}^\top H^\top + \boldsymbol{\mu}^\top)\odot \boldsymbol{w}_f - \boldsymbol{w}'_f \odot \left(\left(\frac{V}{W'H}\right)_f H^\top\right) = 0$$

となる。この解は

$$\boldsymbol{w}_f^*(\boldsymbol{\mu}) = \boldsymbol{w}'_f \odot \frac{\left(\frac{V}{W'H}\right)_f H^\top}{\mathbf{1}^\top H^\top + \boldsymbol{\mu}^\top}$$

である。

等式制約を満たす $\boldsymbol{\mu}$ を求めよう。

$$\boldsymbol{a}_f \coloneqq \left(\frac{V}{W'H}\right)_f H^\top,\ \boldsymbol{b} \coloneqq \mathbf{1}^\top H^\top$$

とおいて、$\boldsymbol{w}_f^*(\mu) = \boldsymbol{w}'_f \odot \boldsymbol{a}_f / (\boldsymbol{b} + \boldsymbol{\mu}^\top)$ と書き直す。すると

$$0 = \sum_f \boldsymbol{w}_f^*(\boldsymbol{\mu}) - \mathbf{1}^\top = \sum_f \boldsymbol{w}'_f \odot \frac{\boldsymbol{a}_f}{\boldsymbol{b} + \boldsymbol{\mu}^\top} - \mathbf{1}^\top$$

$$\therefore \quad \boldsymbol{b} + (\boldsymbol{\mu}^*)^\top = \sum_f \boldsymbol{w}'_f \odot \boldsymbol{a}_f$$

と求まる。

得られた $\boldsymbol{\mu}^*$ を用いて、W の更新式が

$$W \leftarrow W^*(\boldsymbol{\mu}^*) = W' \odot \frac{\frac{V}{W'H} H^\top}{J\left(W' \odot \frac{V}{W'H} H^\top\right)}$$

と求まる。

謝辞

研究は究極に個人的なものでありながら、同時に独りでできるものではありません。

指導教員である峯松信明教授は、この 2 年間自由に研究をする環境を与えて下さいました。テーマを好きに選ばせていただいたお陰で、時間が掛かってはしまいましたが、あまり人気がなくて競争の激しくないテーマを自分で選ぶことができました。これは願ってもないことです。state-of-the-art を目指そうなどと変に気張ることもなく、自分の興味の行くままに研究することができたと感じています。進捗が全くなくても大目に見ていただいていたような気がします。本研究が将来の峯松齋藤研の研究活動や、あるいは広く工学に、少しの足しにでもなればと願います。

齋藤大輔准教授は、私にとって第二の指導教員でした。本研究は、今も過去もこの研究室で他に誰も取り組んでいる人がおらず、独りで全くの手探り状態から始まりました。識別的 NMF のテーマを考えついて齋藤先生にご相談したとき、面白そうだと言っていただけて「この方向性で良いんだ」ととても安心したことを覚えています。五月祭後に一緒に日本酒を飲んでいたときに、至極楽しそうにベイズの話をされていたことも印象深いです。

技術専門員の高橋登様や事務補佐員の池上恵様には研究活動を裏からサポートしていただきました。研究室の費用で何冊も本を購入したので、その度にお手数をお掛けしてしまいました。お陰様で自分では滅多に買えないような高い専門書を読むことができました。

研究室の学生メンバーの皆様には、研究の相談をしたり雑談をしたりと、研究室での普段の生活を充実したものにしていただきました。特に同期の安藤慎太郎氏、後藤駿介氏、白旗悠真氏がいなかったら、私は不安が和らぐことのないまま就職活動を迎えていたと思います。気軽に雑談の出来る環境は、今となっては有難いものです。

齋藤先生にご紹介いただいた、NTT コミュニケーション科学基礎研究所でのインターンはとても貴重な体験となりました。音源分離の分野で世界の第一線にいらっしゃる方々の下で指導を受けるというのは滅多にない機会です。少し変な言い方かも知れませんが、このインターンによって初めて、研究者という存在が私にとって身近なものになった気がしています。インターン指導を担当していただいた荒木章子様や、色々と身辺の世話をしていただいた信田華奈子様をはじめとして、ご指導を賜ったり相談に乗っていただいたりした沢山の所員の方々や、インターンに共に励んだ同期に、この場を借りて感謝します。

最後に。私は結果的に修士課程の半分あまりをコロナ禍で過ごすこととなりました。その期間も含めて、心身ともに支え続けてくれた家族に感謝します。

発表文献

［1］ 紺野瑛介, 齋藤大輔, 峯松信明. 拡張ラグランジュ関数法による識別的非負値行列因子分解と音源分離への応用. 日本音響学会 2020 年秋季研究発表会講演論文集, pp.143–146, 2020.

［2］ 紺野瑛介, 齋藤大輔, 峯松信明. NMF 基底間の識別性に関する定量的尺度. 電子情報通信学会応用音響研究会資料, March 2021（投稿済み）.

［3］ Eisuke Konno, Daisuke Saito and Nobuaki Minematsu. A quantitative measure of discriminability between NMF dictionaries: A comparison with minimum-volume NMF. In *Proceedings of Interspeech*, 2021 (to be submitted).

参考文献

[1] Shoko Araki, Hiroshi Sawada, and Shoji Makino. Blind speech separation in a meeting situation with maximum SNR beamformers. In *Proceedings of ICASSP*, Vol. 1, pp. 41–44, 2007.

[2] Yoshiaki Bando, Masato Mimura, Katsutoshi Itoyama, Kazuyoshi Yoshii, and Tatsuya Kawahara. Statistical speech enhancement based on probabilistic integration of variational autoencoder and non-negative matrix factorization. In *Proceedings of ICASSP*, pp. 716–720, 2018.

[3] E. Colin Cherry. Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, Vol. 25, No. 5, pp. 975–979, 1953.

[4] Andrzej Cichocki, Hyekyoung Lee, Yong-Deok Kim, and Seungjin Choi. Non-negative matrix factorization with α-divergence. *Pattern Recognition Letters*, Vol. 29, No. 9, pp. 1433–1440, 2008.

[5] Julian Eggert and Edgar Körner. Sparse coding and NMF. In *Proceedings of Neural Networks*, Vol. 4, pp. 2529–2533, 2004.

[6] Shinto Eguchi and Yutaka Kano. Robustifying maximum likelihood estimation. *Tokyo Institute of Statistical Mathematics, Tokyo, Japan, Tech. Rep*, 2001.

[7] Cédric Févotte, Nancy Bertin, and Jean-Louis Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, Vol. 21, No. 3, pp. 793–830, 2009.

[8] Xiao Fu, Kejun Huang, and Nicholas D. Sidiropoulos. On identifiability of nonnegative matrix factorization. *IEEE Signal Processing Letters*, Vol. 25, No. 3, pp. 328–332, 2018.

[9] Xiao Fu, Kejun Huang, Nicholas D. Sidiropoulos, and Wing-Kin Ma. Nonnegative matrix factorization for signal and data analytics: Identifiability, algorithms, and applications. *IEEE Signal Processing Magazine*, Vol. 36, No. 2, pp. 59–80, 2019.

[10] Nicolas Gillis. The why and how of nonnegative matrix factorization. In *Regularization, Optimization, Kernels, and Support Vector Machines*, pp. 257–291. Chapman & Hall/CRC, 2014.

[11] Eugene Gover and Nishan Krikorian. Determinants and the volumes of parallelotopes and zonotopes. *Linear Algebra and its Applications*, Vol. 433, No. 1, pp. 28 – 40, 2010.

[12] Emad M. Grais and Hakan Erdogan. Discriminative nonnegative dictionary learning using cross-coherence penalties for single channel source separation. In *Proceedings of Inter-*

speech, pp. 808–812, 2013.

[13] Jahn Heymann, Lukas Drude, and Reinhold Haeb-Umbach. Neural network based spectral mask estimation for acoustic beamforming. In *Proceedings of ICASSP*, pp. 196–200, 2016.

[14] Takuya Higuchi, Nobutaka Ito, Takuya Yoshioka, and Tomohiro Nakatani. Robust MVDR beamforming using time-frequency masks for online/offline ASR in noise. In *Proceedings of ICASSP*, pp. 5210–5214, 2016.

[15] Atsuo Hiroe. Solution of permutation problem in frequency domain ICA, using multivariate probability density functions. In *Proceedings of ICA*, pp. 601–608, 2006.

[16] Patrik O Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, Vol. 5, pp. 1457–1469, 2004.

[17] Kejun Huang, Nicholas D. Sidiropoulos, and Ananthram Swami. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE Transactions on Signal Processing*, Vol. 62, No. 1, pp. 211–224, 2014.

[18] Hirokazu Kameoka, Nobutaka Ono, Kunio Kashino, and Shigeki Sagayama. Complex NMF: A new sparse representation for acoustic signals. In *Proceedings of ICASSP*, pp. 3437–3440, 2009.

[19] Taesu Kim, Hagai T. Attias, Soo-Young Lee, and Te-Won Lee. Blind source separation exploiting higher-order frequency dependencies. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 1, pp. 70–79, 2007.

[20] Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka, and Hiroshi Saruwatari. Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 24, No. 9, pp. 1626–1641, 2016.

[21] Daichi Kitamura, Hiroshi Saruwatari, Hirokazu Kameoka, Yu Takahashi, Kazunobu Kondo, and Satoshi Nakamura. Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 23, No. 4, pp. 654–669, 2015. `http://d-kitamura.net/dataset.html`.

[22] Akira Kurematsu, Kazuya Takeda, Yoshinori Sagisaka, Shigeru Katagiri, Hisao Kuwabara, and Kiyohiro Shikano. ATR Japanese speech database as a tool of speech recognition and synthesis. *Speech Communication*, Vol. 9, No. 4, pp. 357–363, 1990.

[23] Jonathan Le Roux, Felix J. Weninger, and John R. Hershey. Sparse NMF – half-baked or well done? Technical Report TR2015-023, Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA, 2015.

[24] Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, Vol. 401, No. 6755, pp. 788–791, 1999.

[25] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, pp. 556–562. MIT Press, 2001.

[26] Valentin Leplat, Andersen M. S. Ang, and Nicolas Gillis. Minimum-volume rank-deficient

nonnegative matrix factorizations. In *Proceedings of ICASSP*, pp. 3402–3406, 2019.

[27] Valentin Leplat, Nicolas Gillis, and Andersen M. S. Ang. Blind audio source separation with minimum-volume beta-divergence NMF. *IEEE Transactions on Signal Processing*, Vol. 68, pp. 3400–3410, 2020.

[28] Valentin Leplat, Nicolas Gillis, and Jérôme Idier. Multiplicative updates for NMF with β-divergences under disjoint equality constraints. *arXiv e-prints*, p. arXiv:2010.16223, 2020.

[29] Li Li, Hirokazu Kameoka, and Shoji Makino. Discriminative non-negative matrix factorization with majorization-minimization. In *Proceedings of HSCMA*, pp. 141–145, 2017.

[30] Stanley P. Lipshitz, Mark Pocock, and John Vanderkooy. On the audibility of midrange phase distortion in audio systems. *Journal of the Audio Engineering Society*, Vol. 30, No. 9, pp. 580–595, 1982.

[31] Pejman Mowlaee, Rahim Saeidi, and Yannis Stylianou. Advances in phase-aware signal processing in speech communication. *Speech Communication*, Vol. 81, pp. 1–29, 2016.

[32] Hiroaki Nakajima, Daichi Kitamura, Norihiro Takamune, Hiroshi Saruwatari, and Nobutaka Ono. Bilevel optimization using stationary point of lower-level objective function for discriminative basis learning in nonnegative matrix factorization. *IEEE Signal Processing Letters*, Vol. 26, No. 6, pp. 818–822, 2019.

[33] Masahiro Nakano, Hirokazu Kameoka, Jonathan Le Roux, Yu Kitano, Nobutaka Ono, and Shigeki Sagayama. Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with beta-divergence. In *Proceedings of MLSP*, pp. 283–288, 2010.

[34] Joonas Nikunen and Tuomas Virtanen. Object-based audio coding using non-negative matrix factorization for the spectrogram representation. *Journal of the Audio Engineering Society*, 2010.

[35] Pentti Paatero and Unto Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, Vol. 5, No. 2, pp. 111–126, 1994.

[36] Hiroshi Sawada, Hirokazu Kameoka, Shoko Araki, and Naonori Ueda. Multichannel extensions of non-negative matrix factorization with complex-valued data. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 5, pp. 971–982, 2013.

[37] Hiroshi Sawada, Nobutaka Ono, Hirokazu Kameoka, Daichi Kitamura, and Hiroshi Saruwatari. A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF. *APSIPA Transactions on Signal and Information Processing*, Vol. 8, p. e12, 2019.

[38] Ankur Sinha, Pekka Malo, and Kalyanmoy Deb. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, Vol. 22, No. 2, pp. 276–295, 2018.

[39] Paris Smaragdis, Bhiksha Raj, and Madhusudana Shashanka. Supervised and semi-supervised separation of sounds from single-channel mixtures. In *Proceedings of ICA*, pp. 414–421, 2007.

[40] Pablo Sprechmann, Alex M. Bronstein, and Guillermo Sapiro. Supervised non-euclidean sparse NMF via bilevel optimization with applications to speech enhancement. In *Proceedings of HSCMA*, pp. 11–15, 2014.

[41] Ying Sun, Prabhu Babu, and Daniel P. Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, Vol. 65, No. 3, pp. 794–816, 2017.

[42] Ryoichi Takashima, Tetsuya Takiguchi, and Yasuo Ariki. Exemplar-based voice conversion in noisy environment. In *Proceedings of SLT*, pp. 313–317, 2012.

[43] Stephen A. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, Vol. 20, No. 3, pp. 1364–1377, 2010.

[44] Pierre-Hugo Vial, Paul Magron, Thomas Oberlin, and Cédric Févotte. Phase retrieval with bregman divergences and application to audio signal recovery. *arXiv e-prints*, p. arXiv:2010.00392, 2020.

[45] Emmanuel Vincent, Shoko Araki, Fabian Theis, Guido Nolte, Pau Bofill, Hiroshi Sawada, Alexey Ozerov, Vikrham Gowreesunker, Dominik Lutter, and Ngoc Q.K. Duong. The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges. *Signal Processing*, Vol. 92, No. 8, pp. 1928–1936, 2012.

[46] Emmanuel Vincent, Rémi Gribonval, and Cédric Févotte. Performance measurement in blind audio source separation. *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 4, pp. 1462–1469, 2006.

[47] Hermann Von Helmholtz. *On the Sensations of Tone as a Physiological Basis for the Theory of Music*. Longmans, Green, 1912.

[48] DeLiang Wang. Time-frequency masking for speech separation and its potential for hearing aid design. *Trends in Amplification*, Vol. 12, No. 4, pp. 332–353, 2008.

[49] Yu-Xiong Wang and Yu-Jin Zhang. Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 25, No. 6, pp. 1336–1353, 2013.

[50] Felix Weninger, Jonathan Le Roux, John R Hershey, and Shinji Watanabe. Discriminative NMF and its application to single-channel source separation. In *Proceedings of Interspeech*, pp. 865–869, 2014.

[51] Kazuyoshi Yoshii, Kouhei Sekiguchi, Yoshiaki Bando, Mathieu Fontaine, and Aditya Arie Nugraha. Fast multichannel correlated tensor factorization for blind source separation. In *Proceedings of EUSIPCO*, pp. 306–310, 2021.

[52] Kazuyoshi Yoshii, Ryota Tomioka, Daichi Mochihashi, and Masataka Goto. Infinite positive semidefinite tensor factorization for source separation of mixture signals. In *Proceedings of ICML*, Vol. 28, pp. 576–584, 2013.

[53] Takuya Yoshioka, Nobutaka Ito, Marc Delcroix, Atsunori Ogawa, Keisuke Kinoshita, Masakiyo Fujimoto, Chengzhu Yu, Wojciech J. Fabian, Miquel Espi, Takuya Higuchi, Shoko Araki, and Tomohiro Nakatani. The NTT CHiME-3 system: Advances in speech enhance-

ment and recognition for mobile multi-microphone devices. In *Proceedings of ASRU*, pp. 436–443, 2015.

［54］ 寒野善博, 土谷隆. 東京大学工学教程 基礎系 数学 最適化と変分法. 丸善出版株式会社, 2014.

［55］ 福島雅夫. 非線形最適化の基礎. 朝倉書店, 2001.

［56］ 小野順貴. 小特集「位相情報を考慮した音声音響信号処理」にあたって. 日本音響学会誌, Vol. 75, No. 3, pp. 125–129, 2019.