

多重音基本周波数解析のための無限複合自己回帰モデル

吉井 和佳^{1,a)} 後藤 真孝^{1,b)}

概要: 本稿では、音楽音響信号に含まれる複数の基本周波数 (F0) を推定するための確率的生成モデルについて述べる。従来、ソース・フィルタ理論に基づく非負値行列因子分解 (NMF) では、非負値の観測スペクトログラムを行列とみなし、二種類のスペクトル (ソースのスペクトルおよびフィルタの周波数応答) と、それらの各組み合わせに対する時間方向のアクティベーション (音量変化) との積に分解することが行われていた。このとき、楽器の発音機構は自己回帰 (AR) 系でモデル化できると仮定すれば、F0 をもつ周期信号 (くし型スペクトル) や白色雑音 (平坦スペクトル) などの音源により、楽器の反響特性 (音色) を表現する全極型フィルタが駆動されることで音響信号が生成される過程を考えることができる。しかし、この種の**複合自己回帰モデル**においては、本来未知であるはずのソースやフィルタの個数を事前に指定しておく必要があるという問題があった。この問題を解決するため、ガンマ過程を用いることで、上限のない個数のソースやフィルタを理論上許容可能なノンパラメトリックベイズモデルである**無限複合自己回帰モデル**を提案する。本モデルの学習のため、変分ベイズ法 (VB) と乗法更新則 (MU) を組み合わせた効率的な反復最適化アルゴリズムを導出し、スパースな学習ができることを実験により確認した。

1. はじめに

多重音基本周波数解析は、自動採譜やコード認識などさまざまな音楽解析タスクの基礎をなす重要な技術である。近年、多重音の分離や採譜を行ううえで、非負値行列因子分解 (Nonnegative Matrix Factorization: NMF) に基づく手法 [1–13] が多数提案されている。標準的な NMF [14,15] は、時間・周波数平面上の振幅スペクトログラムあるいはパワースペクトログラムを非負値行列とみなし、少数の基底スペクトルとそれらの各時刻における音量 (アクティベーション) との積に分解する。このような行列の低ランク近似は、一見複雑に見える音楽音響信号であっても、構成音の音高や音色の種類は限られているという事実から、ある程度妥当であると考えられている。また、NMF には乗法更新則 (Multiplicative Update: MU) と呼ばれる収束保証付きの反復解法が存在し、基底と音量とを交互に更新することで局所最適解を効率的に求めることができる利点がある。しかし、通常の NMF を多重音基本周波数解析に利用するうえでの主要な問題として、以下の3点がある。

(1) 基底スペクトル形状は時間変化しないため、楽器音の音色のバリエーションを十分に表現するためには、多数の基底スペクトルが必要になる。例えば、同じ基本周波数 (F0) であっても、スペクトル包絡の形状が異なる楽器音に対しては、それぞれ独立した基底スペクトルが必要になる。このようにモデルの複雑さを単純

に増加させてしまうと、最適化アルゴリズムを用いて性質のよい局所解を求めるのが難しくなる。

(2) NMF を実行したあとで、各基底スペクトルから F0 を推定する処理が別途必要になる。しかし、基底スペクトルの形状に制約を与えずに NMF を実行すると、調波構造をもつ楽器音として不自然なものが基底として推定される可能性がある。このことは、F0 推定を難しくするだけではなく、各基底に対して F0 が存在するかどうかの判定処理も追加で必要になる。

(3) 基底スペクトルの個数 (モデルの複雑さ) は NMF による分解結果に大きな影響を与えるにもかかわらず、事前に手動で指定する必要がある。この問題は、第一の問題にも関連していることに注意されたい。単純には、あらゆる複雑さのモデルを別々に試したあとで最善のものをひとつ選ばばよいが、そのようなモデル選択は計算量の観点で現実的ではない。

上記で議論した通り、制約なしの NMF は柔軟すぎるうえ、音楽的に意味のある分解結果を得られるよう手動で制御することは困難であるといえる。しかし、これら3つの問題はこれまで部分的に扱われていただけであった [1–13]。

統一的な枠組みのもとで上記3つの問題を解決するため、本研究では無限複合自己回帰モデル (Infinite Composite Autoregressive Model: iCARM) と呼ぶ音楽音響信号に対する新しいベイズモデルを提案する (図 1)。本モデルの主要な特長は、以下の3点である。

(1) ソース・フィルタ分解 [1]

基底スペクトルをソーススペクトルと全極型フィルタとの積にさらに分解する。これは、人間の発声機構を

¹ 産業技術総合研究所
Umezono 1-1-1, Tsukuba, Ibaraki 305-8568, Japan

^{a)} k.yoshii(at)aist.go.jp

^{b)} m.goto(at)aist.go.jp

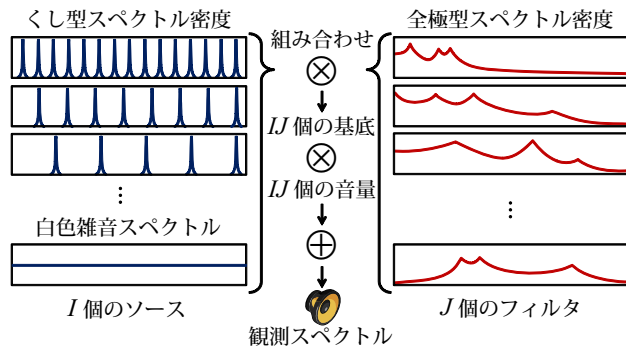


図1 無限複合自己回帰モデル: I 個のソースと J 個のフィルタを組み合わせることで IJ 個の基底スペクトルが得られ、それぞれが各時刻において異なる音量でアクティベートされる。本研究では、 I および J を無限に発散させたときの極限を考える。

自己回帰系とみなすアプローチに着想を得ている。人間は、同じ声の高さ（ソース）を維持しながら、声道の形状（フィルタ）を変化させることで様々な母音を生成できる。楽器の発音機構も同様であるとみなせば、モデルの複雑さの増加を抑えつつ、 F_0 と音色という独立した観点から様々な楽器音を表現できる。

(2) 調波構造モデリング [7,9]

調波構造を持つ楽器音を表現するため、 F_0 をパラメータに持ち、周波数軸上で倍音を等間隔に並べたくし型関数を用いてソーススペクトルを定式化する。このスペクトルが音響学的に妥当性のある全極型フィルタを通過すると、各倍音の相対的な重みは、調波構造を持つ楽器音として適切な値をとるようになる。

さらに、文献 [7,16] における提案にしたがい、全周波数帯域にわたって平坦なスペクトルをもつ白色雑音を表す特別なソーススペクトルを導入する。これは、広い周波数帯域にわたってなだらかなスペクトルをもつ打楽器音（ギターの弦を指ではじいたり、ピアノのハンマーが弦を叩く際の打撃音なども含む）を生成するための駆動音源である。調波構造を持つ楽器音と同様に、音色は全極型フィルタによって特徴づけられる。

(3) ノンパラメトリックベイズ [12]

与えられた観測スペクトログラムを表現するのに必要なソースとフィルタの個数を自動的に調節するため、ノンパラメトリックベイズモデルを構成する。本研究では、従来の複合自己回帰モデル [1] において、ソースとフィルタの個数を事前に指定することはせず、代わりにそれらを無限に発散させたときの極限を考える。このとき、無限次元の事前分布であるガンマ過程を導入することで、本来無限個あるうちで限られた個数のソースとフィルタのみが実質的にアクティベートされるスパースな学習が実現できる。

本研究では、iCARM を学習するため、変分ベイズ法 (Variational Bayes: VB) と乗法更新則 (MU) を組み合わせた新しいタイプの反復最適化アルゴリズムを提案する。

表1 ソース・フィルタ理論に基づく NMF

手法	基準	ソース #	フィルタ #
Kameoka [1]	IS	- I	AR J
Badeau [2]	IS	H I	MA J
Durrieu [3]	IS	(H) I	- J
Virtanen [4]	KL	- I	- J
Carabias-Orti [5]	KL	H I	- J
Heittola [6]	KL	- I^*	- J
安良岡 [7]	KL	H + N I^*	AR J
Hennequin [8]	Beta (0.5)	- I	ARMA J^*
本研究	KL or IS	H + N ∞	AR ∞

(H: 調波構造, N: ノイズ, -: その他, *: 時変化)

2. 関連研究

本章では、近年特に注目されている機械学習技術として、NMF とノンパラメトリックベイズの関連研究を紹介する。

2.1 非負値行列因子分解

NMF は非負値行列に対してスパースな分解を行うための強力なアプローチである [14]。NMF は当初、与えられた顔画像の集合に対して、目や鼻といった少数の局所パーツへの分解を得る目的で利用されている。このような性質は、非負値の基底画像の加算しか許さないという制約から自然に導かれる。調波構造を持つ楽器音は局所的な周波数帯域にエネルギーが集中しているため、NMF は音楽音響信号を局所パーツへ分解する目的に適している。

2.1.1 最適化のための基準

NMF を実行するうえでは、観測データ（スペクトログラム）に対する再構成データ（基底スペクトルの線形結合）からの「乖離度」を測るための基準が必要である。良く利用される基準として、カルバック・ライブラー (KL) ダイバージェンス [14] や板倉・斎藤 (IS) ダイバージェンス [15] が挙げられる（表1）。両者を特別な場合として含むベータダイバージェンスに基づく NMF [17] も提案されている。

音響信号のモデル化においては、IS-NMF の方が理論的には妥当であるが（3.2.1 節参照）、KL-NMF の方がしばしば良い結果をもたらすことが知られている。この理由として、IS ダイバージェンスは凸性を持たないため、勾配降下型の最適化手法では、大域的な最適解を見つけることが困難であることが考えられる。このような傾向は最尤推定において知られているが、ベイズ推定においても同様の傾向が見られるかは自明ではない。本研究では、各基準に基づいて KL-iCARM と IS-iCARM を提案する。

2.1.2 ソース・フィルタ分解

NMF の主要な拡張の一つに、ソース・フィルタ理論に基づくものがある（表1）。例えば、亀岡ら [1] は、観測スペクトログラムを一定個数のソースと AR フィルタ（全極型伝達関数）に分解するための複合自己回帰モデルを提案している。同様のモデルはいくつか提案されているが [3-5]、フィルタは音響学的な視点に立って設計されたものではなかった。一方、Badeau ら [2] は、移動平均 (MA) フィルタ（全零型伝達関数）を用いることを提案している。

いくつかの研究では、モデルの複雑さの増加と引き換えに、ソースやフィルタが時間変化することを許容するモデルを提案している。Heittola ら [6] や安良岡ら [7] は、フィルタは時不変のままであるが、ソースの時間変化を許容するモデルを提案している。一方、Hennequin ら [8] は、ソースを時不変として、自己回帰移動平均 (ARMA) フィルタの時間変化を許容するモデルを提案し、効率的な MU アルゴリズムに基づくパラメータ推定法を導出している。

2.1.3 調波構造モデリング

NMF の別の拡張として、基底スペクトルやソーススペクトルに対し、調波構造を持たせるような制約を与えるものがある。例えば、Vincent ら [10] は、ある周波数帯域内の調波構造スペクトルを表すテンプレートを複数用意し、各基底スペクトルをそれらの重み付き和で表現する手法を提案している。ソース・フィルタ型の NMF においては、Badeau ら [2] は、各ソーススペクトルを、倍音に対応する周波数で 1 を取りやすくなる確率的なバイナリベクトルで表現する手法を提案している。安良岡ら [7] や Hennequin ら [9] は、各ソーススペクトルを、パラメトリックな関数 (例：倍音周波数を中心として配置した急峻なガウス関数の和) で表現している。Carabias-Orti ら [5] は、倍音の重みをいくつかのテンプレートの重み付き和としてさらに分解している。文献 [5,9] では、効率的な MU アルゴリズムに基づくパラメータ推定法を導出している。

安良岡ら [7] の手法の特徴は、全てのソースに調波構造を持たせるのではなく、平坦なスペクトルをもつ特別なソースをひとつ用意しておくことである。これは、人間の発声機構において、音源は周期信号 (くし型スペクトル) か白色雑音 (平坦スペクトル) に大別でき、それら音源信号は、声道 (AR フィルタ) によって調音されるという仮定に基づいている。この仮定は楽器音のモデル化においても、ある程度妥当であると考えられる。本研究では、ベイズ的な枠組みのもとで楽器音の生成過程を記述する。

2.2 ノンパラメトリックベイズ

ノンパラメトリックベイズ理論 [18] は、古典的なベイズ理論を拡張したものである。従来のベイズ推定の枠組みでは、未知のパラメータに対して事前分布を導入することで、観測データが与えられたもとの事後分布を求めることができる。しかし、本来は未知であるはずのモデルの複雑さ (本研究においてはソースやフィルタの個数) は、単なる超パラメータとして事前に指定しておく必要があった。このとき、網羅的に探索を行おうとしても、指定すべき複雑さが複数種類ある場合には計算量が組み合わせ爆発を起こしてしまう。ノンパラメトリックベイズモデルでは、事後分布推定時にそれらを一度に最適化することができる。

近年、ベイズモデルは音楽情報処理においても盛んに利用され、ベイズ拡張された NMF [19] は音源分離タスクにおいて良い結果を与えることが知られている。特に注目すべ

表 2 無限複合自己回帰モデルで利用されるシンボル

M, N	周波数ビン数, フレーム数
I, J	ソース数, フィルタ数 (いずれも無限大に発散)
X_{mn}	ビン m ・フレーム n における振幅またはパワー
Y_{mn}	ビン m ・フレーム n における再構成された値
θ_i	ソース i のグローバルな重み
ϕ_j	フィルタ j のグローバルな重み
W_{im}	ソース i におけるビン m の振幅またはパワー
A_{jm}	フィルタ j におけるビン m のゲイン
H_{ijn}	フレーム n におけるソース i とフィルタ j の音量

きモデルに、Hoffman らによるノンパラメトリックベイズ拡張された NMF であるガンマ過程 NMF (GaP-NMF) [12] が挙げられる。このモデルは、理論上無限個の基底スペクトルが存在することを仮定するが、観測データに合わせて少数の基底が実質的にアクティベートされる機構を有している。この機構は、各基底の時間変化を許容するような複雑なモデル [13] を構成する際にも利用されている。

3. 無限複合自己回帰モデル

本章では、無限複合自己回帰モデル (iCARM) と呼ぶ、多重音基本周波数解析のためのノンパラメトリックベイズモデルについて述べる (図 1 および図 2)。本モデルは、IS ダイバージェンス基準のもとで、パワースペクトログラムを指定された個数のソースと AR フィルタの積に分解する複合自己回帰モデル [1] を拡張したものである。本研究では、KL ダイバージェンス基準のもとで、振幅スペクトログラムを分解する別の定式化も行う (3.2.1 節参照)。ここで、文献 [7] に提案されている通り、各ソースに調波構造を持たせるため、くし型関数を用いることにする。ただし、平坦なスペクトル (ホワイトノイズ) を持ったソースを別の一つだけ用意しておく。さらに、文献 [12] に提案されている通り、ガンマ過程を用いることで、理論上無限個のソースとフィルタが存在することを許容するように拡張する。

3.1 スペクトログラムの分解

最初に、提案モデルの定式化に必要な数学的記号を定義しておく (表 2)。 M を周波数ビン数、 N をフレーム数として、 $M \times N$ の複素スペクトログラムを行列 $\mathbf{X}$ とみなす。また、 I をソース数、 J をフィルタ数とし、これらの値が無限に発散するときを考える。小文字の m, n, i, j はそれぞれインデックスを表すものとする。

本研究では、 $\mathbf{X}$ の非負値表現 (振幅あるいはパワースペクトログラム) を 3 種類の独立した因子 $\mathbf{W}$, $\mathbf{A}$ および $\mathbf{H}$ に分解することを試みる。

$$|X_{mn}| \text{ or } |X_{mn}|^2 \approx \sum_{i,j}^{I,J \rightarrow \infty} \theta_i \phi_j W_{im} A_{jm} H_{ijn} \quad (1)$$

ここで、 W_{im} , A_{jm} および H_{ijn} はそれぞれ、ソース i におけるビン m の振幅またはパワー、フィルタ j におけるビン m のゲイン、フレーム n におけるソース i とフィルタ j の音量を表す。また、2 種類の変数 θ_i および ϕ_j が導入さ

れており、全フレームにわたるソース i あるいはフィルタ j のグローバルな重みを表す。このとき、 I や J が無限大に発散しても、 θ や ϕ のうちで限られた要素のみがある程度以上の大きさを持ち、その他の要素は無視できるくらいに小さい値をとる（厳密にはゼロではない）ようにできる。すなわち、ソースやフィルタの「実質的な」個数 I^+ および J^+ がデータから自動的に決まることを示している。

我々の目的は、観測データ $\mathbf{X}$ が与えられたもとで、未知の確率変数の事後分布 $p(\theta, \phi, \mathbf{H} | \mathbf{X}; \mathbf{W}, \mathbf{A})$ を計算し、 $\mathbf{W}$ および $\mathbf{A}$ を構成するパラメータを推定することである。以降、事前分布 $p(\theta), p(\phi), p(\mathbf{H})$ 、尤度関数 $p(\mathbf{X} | \theta, \phi, \mathbf{H}; \mathbf{W}, \mathbf{A})$ 、パラメトリックな関数 $\mathbf{W}, \mathbf{A}$ の設計について議論する。

3.2 確率モデルの定式化

本節では、KL ダイバージェンスあるいはIS ダイバージェンスに基づく iCARM の設計方法について述べる。

3.2.1 $\mathbf{X}$ に対する尤度関数

本研究では、KL ダイバージェンスあるいはIS ダイバージェンスを乖離度基準として用いる。いま、再構成データ $\mathbf{Y}$ を $Y_{mn} = \sum_{ij} Y_{mn}^{ij}$ および $Y_{mn}^{ij} = \theta_i \phi_j W_{im} A_{jm} H_{ijn}$ と定義すると、式 (1) から X_{mn} と Y_{mn} との乖離度を最小化するような Y_{mn} を求めたい。これは、 X_{mn} が与えられたときに、 Y_{mn} をパラメータにもつポアソン分布あるいは指数分布の最尤推定を行うことと等価である。ここで、 X_{mn} に含まれるソース i とフィルタ j の組み合わせ成分を X_{mn}^{ij} としておく。ただし、 $X_{mn} = \sum_{ij} X_{mn}^{ij}$ が成り立つ。

まず、KL-iCARM では、振幅領域での加算性 $|X_{mn}| = \sum_{ij} |X_{mn}^{ij}|$ が成り立つことを仮定する。これは数学的には正しくないが、実用上はうまく働くことが知られている。いま、 $|X_{mn}^{ij}| \sim \text{Poisson}(Y_{mn}^{ij})$ を仮定すると、ポアソン分布の再生性から $\sum_{ij} |X_{mn}^{ij}| \sim \text{Poisson}(\sum_{ij} Y_{mn}^{ij})$ が成立するので、振幅 $|X_{mn}|$ の尤度は次式で与えられる。

$$|X_{mn}| \sim \text{Poisson}(Y_{mn}) \quad (2)$$

一方、IS-iCARM では、複素領域での加算性 $X_{mn} = \sum_{ij} X_{mn}^{ij}$ に基づいて、数学的には正確な定式化を行う。いま、 $X_{mn}^{ij} \sim \mathcal{N}_c(0, Y_{mn}^{ij})$ を仮定すると、複素ガウス分布の再生性から $\sum_{ij} X_{mn}^{ij} \sim \mathcal{N}_c(0, \sum_{ij} Y_{mn}^{ij})$ が成立する。したがって、パワー $|X_{mn}|^2$ の尤度は次式で与えられる。

$$|X_{mn}|^2 \sim \text{Exponential}(Y_{mn}) \quad (3)$$

3.2.2 θ および ϕ に対するガンマ過程事前分布

無限次元の非負値ベクトルである θ および ϕ に対しては、ガンマ過程事前分布をおくのが都合がよい。具体的には、 θ および ϕ の各要素に対し、以下を仮定する。

$$\theta_i \sim \text{Gamma}\left(\frac{\alpha}{I}, \alpha\right), \quad \phi_j \sim \text{Gamma}\left(\frac{\gamma}{J}, \gamma\right) \quad (4)$$

こうすると、 I が無限大に近づくに連れて、ベクトル θ は、形状パラメータ α を持つガンマ過程から生成される無限個の系列を良く近似するようになる。任意の実数 $\epsilon > 0$ に対

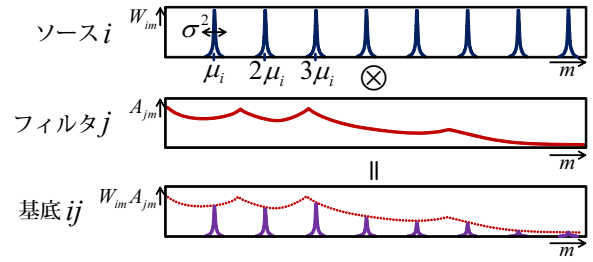


図 2 ソース $\mathbf{W}_i$ とフィルタ $\mathbf{A}_j$ との組み合わせ

して、 $\theta_i > \epsilon$ となる要素の個数 I^+ はほとんど確実に有限であることが証明されている。実用上は、打ち切りレベル I を α より十分に大きくとれば、 θ の I 個の要素のうちいくつかがある程度大きくなるようなスパースな学習が実現できる。同様の議論は、 ϕ に対しても適用できる。

3.2.3 $\mathbf{H}$ に対するガンマチェイン事前分布

各ソース・フィルタの組み合わせの音量が時間方向になめらかに変化するようにするため、 $\mathbf{H}$ に対してガンマチェイン事前分布 [20] をおく。具体的には、以下を仮定する。

$$\begin{aligned} H_{ij1} &\sim \text{Gamma}(\beta, \beta/d) \\ G_{ijn} &\sim \text{Gamma}(\beta, \beta H_{ijn-1}) \\ H_{ijn} &\sim \text{Gamma}(\beta, \beta G_{ijn}) \end{aligned} \quad (5)$$

ここで、 β は、事前分布の強度（音量の時間変化のなめらかさ）を調節する超パラメータである。 G_{ijn} は、隣り合う音量 H_{ijn-1} および H_{ijn} が正の相関を持つようにするための、両者と負の相関をもつ補助変数である ($\mathbb{E}_{\text{prior}}[G_{ijn}] = H_{ijn-1}^{-1}$, $\mathbb{E}_{\text{prior}}[H_{ijn}] = G_{ijn}^{-1}$)。このとき、 G_{ijn} を積分消去すると、正の相関をもつマルコフ連鎖確率 $p(H_{ijn} | H_{ijn-1}) = \frac{\Gamma(2\beta)}{2\Gamma(\beta)} \frac{(H_{ijn-1} H_{ijn})^\beta}{(H_{ijn-1} + H_{ijn})^{2\beta}} H_{ijn}^{-1}$ が得られる。

3.2.4 $\mathbf{W}$ に対するくし型関数

各ソーススペクトル $\mathbf{W}_i$ を H 個のガウス関数の和で構成されるくし型関数で表現する (図 2)。ここで、 H は考慮すべき倍音の個数である。具体的には、以下を考える。

$$W_{im} = \sum_{h=1}^H \exp\left(-\frac{(m - h\mu_i)^2}{2\sigma^2}\right) \quad (6)$$

ここで、 μ_i^{*1} は基本周波数を、 σ は倍音周辺のエネルギーの拡散度合いを表す。また、白色雑音を表す特別なソースに対しては、全周波数帯域にわたって $W_{Im} = 1$ とする。

3.2.5 $\mathbf{A}$ に対する全極型伝達関数

いま、あるフレームにおいて、ソース i とフィルタ j の組み合わせで生成された信号 $x^{ij} \equiv \{x_t^{ij}\}_{t=1}^{2M}$ が P 次の自己回帰過程に従うと仮定すると、以下が成り立つ。

$$x_t^{ij} = -\sum_{p=1}^P a_p^j x_{t-p}^{ij} + s_t^i \quad (7)$$

ここで、 $s^i \equiv \{s_t^i\}_{t=1}^{2M}$ はソース i に対応する信号であり、 $\mathbf{a}_j \equiv \{a_0^j, \dots, a_P^j\}^T$ はフィルタ j の係数である ($a_0^j = 1$)。

*1 基本周波数を $\tilde{\mu}_i$ [Hz] とすると、 $\mu_i = \tilde{\mu}_i / (r/2M)$ [bins] で変換できる。ここで、 r は音響信号のサンプリング周波数である。

*2 $2M$ はフーリエ変換に用いるサンプル数（窓幅）である。

$\mathbf{w}^i \equiv \{w_{ij}^i\}_{j=1}^{2M}$ を $\mathbf{s}^i$ の自己相関関数, $\{W_{im}\}_{m=1}^{2M}$ を $\mathbf{w}^i$ の離散フーリエ変換 (DFT) で得られる複素スペクトル (ウィナー・ヒンチンの定理から $\mathbf{s}^i$ のパワースペクトルと等価) とする. また, $\{X_m^{ij}\}_{m=1}^{2M}$ を $\mathbf{x}^{ij}$ の DFT で得られる複素スペクトルとする. このとき, 音源信号 $\mathbf{s}^i$ が定常なガウス性雑音であると仮定すると, X_m^{ij} は複素ガウス分布 $\mathcal{N}_c(0, \Sigma_m^{ij})$ に従う. ここで, $\Sigma_m^{ij} = W_{im}A_{jm}$ であり, A_{jm} は

$$A_{jm} = \frac{1}{\left| \sum_{p=0}^P a_p^j e^{-2\pi \frac{m}{2M} p} \right|^2} = \frac{1}{\mathbf{a}_j^T \mathbf{U}_m \mathbf{a}_j} \quad (8)$$

である (図 2). $\mathbf{U}_m$ は $(P+1) \times (P+1)$ の Toeplitz 型の行列で, $[\mathbf{U}_m]_{pq} = \cos(2\pi \frac{m}{2M}(p-q))$ で与えられる. したがって, $|X_m^{ij}|^2$ は, $W_{im}A_{jm}$ をパラメータにもつ指数分布に従うことが分かる. すなわち, $\mathbf{x}^{ij}$ が与えられたもとの $\mathbf{a}_j$ の最尤推定は, $\{|X_m^{ij}|^2\}_{m=1}^{2M}$ と $\{W_{im}A_{jm}\}_{m=1}^{2M}$ の間の IS ダイバージェンスを最小化することと等価である*3.

これまでの議論から, KL ダイバージェンスあるいは IS ダイバージェンスに基づく iCARM では次式を仮定する.

$$A_{jm}^{\text{KL}} = \sqrt{\frac{1}{\mathbf{a}_j^T \mathbf{U}_m \mathbf{a}_j}} \quad \text{or} \quad A_{jm}^{\text{IS}} = \frac{1}{\mathbf{a}_j^T \mathbf{U}_m \mathbf{a}_j} \quad (9)$$

ここで, KL-iCARM の場合に平方根を計算する理由は, パワースペクトログラムではなく振幅スペクトログラムを分解するからである (3.2.1 節参照).

3.3 確率モデルの学習

我々の目的は, 対数周辺尤度 $\log p(\mathbf{X}; \mathbf{W}, \mathbf{A})$ が最大化されるように, 事後分布 $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H} | \mathbf{X}; \mathbf{W}, \mathbf{A})$ と関数 $\mathbf{W}$ および $\mathbf{A}$ (パラメータ $\boldsymbol{\mu}, \sigma, \mathbf{a}$) を最適化することである. しかし, 対数周辺尤度や真の事後分布は解析的に計算することができないため, 本研究では変分ベイズ法 (VB) を用いて近似計算を行う. まず, 因子分解が可能なように限定された関数形をもつ事後分布を考える.

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}) = \prod_i q(\theta_i) \prod_j q(\phi_j) \prod_{ijn} q(H_{ijn}) \quad (10)$$

次に, この変分事後分布と真の事後分布との KL ダイバージェンスを最小化するように反復最適化を行う. これは, 対数周辺尤度の変分下限 $\mathcal{L}$ を最大化することと等価である*4. ここで, $\mathcal{L}$ および更新式は次式で与えられる.

$$\begin{aligned} \log p(\mathbf{X}; \mathbf{W}, \mathbf{A}) &\geq \mathbb{E}[\log p(\mathbf{X} | \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})] \\ &+ \mathbb{E}[\log p(\boldsymbol{\theta})] + \mathbb{E}[\log p(\boldsymbol{\phi})] + \mathbb{E}[\log p(\mathbf{H})] \\ &- \mathbb{E}[\log q(\boldsymbol{\theta})] - \mathbb{E}[\log q(\boldsymbol{\phi})] - \mathbb{E}[\log q(\mathbf{H})] \equiv \mathcal{L} \end{aligned} \quad (11)$$

$$\begin{aligned} q(\boldsymbol{\theta}) &\propto \exp(\mathbb{E}_{q(\boldsymbol{\phi}, \mathbf{H})}[\log p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})]) \\ q(\boldsymbol{\phi}) &\propto \exp(\mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{H})}[\log p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})]) \\ q(\mathbf{H}) &\propto \exp(\mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{\phi})}[\log p(\mathbf{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})]) \end{aligned} \quad (12)$$

*3 線形予測分析 (Linear Predictive Coding: LPC) においては, 一般的に, 音源信号 $\mathbf{s}^i$ は白色雑音であることが仮定されている ($W_{im} = 1$). 本研究では, 周期信号の場合を主に考えている.

*4 より正確に言えば, $\mathcal{L}$ は依然として解析的に計算できないので, さらなる変分下限を計算して, それを最大化することになる.

また, $\mathbf{W}$ および $\mathbf{A}$ ($\boldsymbol{\mu}, \sigma$ および $\mathbf{a}$) を最適化するため, VB に乗法更新則 (MU) [8, 9] を組み合わせる手法を提案する. 一般的に, MU における更新式は, 最小化すべきコスト関数 (本研究では $-\mathcal{L}$) を偏微分することで得られる. 例えば, 基本周波数 μ_i の更新式は, コスト関数の偏微分が 2 つの正の項の差 $\frac{\partial \mathcal{L}}{\partial \mu_i} = G_{\mu_i} - F_{\mu_i}$ として記述できるとすると, $\mu_i \leftarrow \mu_i \times \frac{F_{\mu_i}}{G_{\mu_i}}$ で与えられる. したがって, 偏微分がゼロであれば, μ_i は変化せず, そうでなければ勾配とは逆の方向に更新されるため, コスト関数は減少する.

3.3.1 KL-iCARM における変分推論

本節では, KL-iCARM における変分事後分布の更新式を導く. まず, 式 (11) に含まれる対数尤度の期待値 $\mathbb{E}[\log p(\mathbf{X} | \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})]$ は次式で与えられる.

$$\begin{aligned} \mathbb{E}[\log p(\mathbf{X} | \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})] &= \sum_{mn} \log \text{Poisson}(|X_{mn}| | Y_{mn}) \\ &= \sum_{mn} |X_{mn}| \mathbb{E} \left[\log \sum_{ij} \theta_i \phi_j W_{im} A_{jm} H_{ijn} \right] \\ &\quad - \sum_{mn} \sum_{ij} \mathbb{E} [\theta_i \phi_j W_{im} A_{jm} H_{ijn}] - \sum_{mn} \log |X_{mn}| \end{aligned} \quad (13)$$

しかし, 第一項の「和の対数関数の期待値」は解析的に計算できない. そこで, 対数関数 $\log(x)$ は凹関数であることから, $\lambda_k \geq 0$ かつ $\sum_k \lambda_k = 1$ を満たす任意の補助変数 $\boldsymbol{\lambda} = \{\lambda_1, \dots, \lambda_K\}$ に対してイェンセンの不等式

$$\log \left(\sum_k x_k \right) = \log \left(\sum_k \lambda_k \frac{x_k}{\lambda_k} \right) \geq \sum_k \lambda_k \log \left(\frac{x_k}{\lambda_k} \right) \quad (14)$$

が成り立つことを用いると, 式 (13) の変分下限は

$$\begin{aligned} \mathbb{E}[\log p(\mathbf{X} | \boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}; \mathbf{W}, \mathbf{A})] &\geq \sum_{mn} |X_{mn}| \sum_{ij} \lambda_{mnij} \mathbb{E} \left[\log \frac{\theta_i \phi_j W_{im} A_{jm} H_{ijn}}{\lambda_{mnij}} \right] \\ &\quad - \sum_{mn} \sum_{ij} \mathbb{E} [\theta_i \phi_j W_{im} A_{jm} H_{ijn}] - \sum_{mn} \log |X_{mn}| \end{aligned} \quad (15)$$

で与えられる. ここで, $\sum_{ij} \lambda_{mnij} = 1$ となる補助変数 $\boldsymbol{\lambda}$ を導入した. 第一項は「対数関数の和の期待値」に変換され, 解析的に計算可能になっている.

式 (15) の十分統計量は $\log(x)$ および x であり, 式 (4), (5) で与えられる $\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{H}$ に対するガンマ事前分布と共役性が成立していることから, 変分事後分布を

$$\begin{aligned} q(\theta_i) &= \text{Gamma}(a_i^\theta, b_i^\theta), \quad q(\phi_j) = \text{Gamma}(a_j^\phi, b_j^\phi) \\ q(H_{ijn}) &= \text{Gamma}(a_{ijn}^H, b_{ijn}^H) \end{aligned} \quad (16)$$

と仮定すると, 変分パラメータは次式で求められる.

$$\begin{aligned} a_i^\theta &= \frac{\alpha}{I} + \sum_{mnj} |X_{mn}| \lambda_{mnij} \\ b_i^\theta &= \alpha + \sum_{mnj} \mathbb{E} [\phi_j W_{im} A_{jm} H_{ijn}] \\ a_j^\phi &= \frac{\gamma}{J} + \sum_{mni} |X_{mn}| \lambda_{mnij} \\ b_j^\phi &= \gamma + \sum_{mni} \mathbb{E} [\theta_i W_{im} A_{jm} H_{ijn}] \\ a_{ijn}^H &= 2\beta + \sum_m |X_{mn}| \lambda_{mnij} \\ b_{ijn}^H &= \beta \mathbb{E} [G_{ijn} + G_{ijn+1}] + \sum_m \mathbb{E} [\theta_i \phi_j W_{im} A_{jm}] \end{aligned} \quad (17)$$

また、式 (15) を微分してゼロとおくことで、変分下限を最大化する λ_{mnij} は次式で与えられる。

$$\lambda_{mnij} \propto \exp(\mathbb{E}[\log(\theta_i \phi_j W_{im} A_{jm} H_{ijn})]) \quad (18)$$

すなわち、 λ_{mnij} は、ビン m ・フレーム n において、ソース i ・フィルタ j の組み合わせに対応する成分 $\theta_i \phi_j W_{im} A_{jm} H_{ijn}$ が占める割合の期待値を表している。ただし、式 (18) の右辺は $\theta_i \phi_j W_{im} A_{jm} H_{ijn}$ の「幾何平均」であるので、通常の平均 $\mathbb{E}[\theta_i \phi_j W_{im} A_{jm} H_{ijn}]$ を計算する場合とは異なり、小さな成分は消滅に向かう効果が得られる。例えば、 θ に着目すると、 $\mathbb{E}[\theta_i] = \exp(\log(a_i^\theta))/b_i^\theta$ 、 $\exp(\mathbb{E}[\log \theta_i]) = \exp(\psi(a_i^\theta))/b_i^\theta$ であり、後者は前者からほぼ一定のディスカウントがなされ、 a_i^θ が小さいほど影響が大きいことが分かる (図 3 左)。これは NMF のベイズ推定に特有の現象であり、通常の平均を計算する最尤推定や事後確率最大化 (MAP) 推定では、このようなスパース化効果を得ることはできない。

3.3.2 IS-iCARM における変分推論

本節では、IS-iCARM における変分事後分布の更新式を導く。まず、式 (11) に含まれる対数尤度の期待値 $\mathbb{E}[\log p(\mathbf{X}|\theta, \phi, \mathbf{H}; \mathbf{W}, \mathbf{A})]$ は次式で与えられる。

$$\begin{aligned} \mathbb{E}[\log p(\mathbf{X}|\theta, \phi, \mathbf{H}; \mathbf{W}, \mathbf{A})] &= \sum_{mn} \log \text{Exponential}(|X_{mn}|^2 |Y_{mn}|) \\ &= - \sum_{mn} |X_{mn}|^2 \mathbb{E} \left[\frac{1}{\sum_{ij} \theta_i \phi_j W_{im} A_{jm} H_{ijn}} \right] \\ &\quad - \sum_{mn} \mathbb{E} \left[\log \sum_{ij} \theta_i \phi_j W_{im} A_{jm} H_{ijn} \right] \end{aligned} \quad (19)$$

しかし、第一項の「和の逆数の期待値」も第二項の「和の対数関数の期待値」も解析的に計算できないため、文献 [12, 21] で提案されている手法を用いる。まず、第一項の下限関数について考える。関数 $-\frac{1}{x}$ は凹関数であることから、 $\eta_k \geq 0$ かつ $\sum_k \eta_k = 1$ を満たす任意の補助変数 $\boldsymbol{\eta} = \{\eta_1, \dots, \eta_K\}$ に対してイェンセンの不等式

$$\frac{-1}{\sum_k x_k} = \frac{-1}{\sum_k \eta_k \frac{x_k}{\eta_k}} \geq - \sum_k \eta_k \frac{1}{\frac{x_k}{\eta_k}} = - \sum_k \eta_k^2 \frac{1}{x_k} \quad (20)$$

が成り立つことを利用する。次に、第二項の下限関数について考える。式 (13) と異なり、負の対数関数 $-\log(x)$ は凸関数であることから、 ω 周りのテイラー展開を用いて

$$-\log(x) \geq -\log(\omega) + 1 - \frac{x}{\omega} \quad (21)$$

として ω における接線を得る。式 (19) の変分下限は

$$\begin{aligned} \mathbb{E}[\log p(\mathbf{X}|\theta, \phi, \mathbf{H}; \mathbf{W}, \mathbf{A})] &\geq - \sum_{mn} |X_{mn}|^2 \sum_{ij} \eta_{mnij}^2 \mathbb{E} \left[\frac{1}{\theta_i \phi_j W_{im} A_{jm} H_{ijn}} \right] \\ &\quad - \sum_{mn} \left(\log(\xi_{mn}) - 1 + \frac{1}{\xi_{mn}} \sum_{ij} \mathbb{E}[\theta_i \phi_j W_{im} A_{jm} H_{ijn}] \right) \end{aligned} \quad (22)$$

で与えられる。ここで、 $\sum_{ij} \eta_{mnij} = 1$ となる補助変数 $\boldsymbol{\eta}$ および任意の正の実数 ξ_{mnij} を導入した。この結果、全ての項が解析的に計算可能となっている。

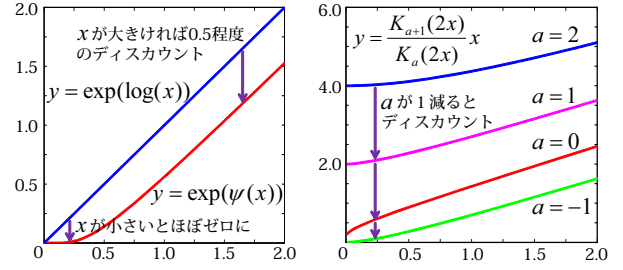


図 3 ディガンマ関数 $\psi(x)$ や第二種変形ベッセル関数 $K_a(x)$ の性質に起因するディスカウント効果

いま、式 (22) の十分統計量は $\log(x)$ および $\frac{1}{x}$ であり、 $\theta, \phi, \mathbf{H}$ に対するガンマ事前分布の十分統計量は $\log(x)$ および x であるので、共役性が満たされていない。しかし、 $\log(x)$ 、 x および $\frac{1}{x}$ を十分統計量にもつ一般化逆ガウス (GIG) 分布 (付録参照) を変分事後分布に用いることで、閉形式の更新式導出が可能になる [12]。具体的には、

$$\begin{aligned} q(\theta_i) &= \text{GIG}(a_i^\theta, b_i^\theta, c_i^\theta), \quad q(\phi_j) = \text{GIG}(a_j^\phi, b_j^\phi, c_j^\phi) \\ q(H_{ijn}) &= \text{GIG}(a_{ijn}^H, b_{ijn}^H, c_{ijn}^H) \end{aligned} \quad (23)$$

と仮定すると、変分パラメータは次式で求められる。

$$\begin{aligned} a_i^\theta &= \frac{\alpha}{I}, \quad b_i^\theta = \alpha + \sum_{mnj} \frac{\mathbb{E}[\phi_j W_{im} A_{jm} H_{ijn}]}{\xi_{mn}} \\ c_i^\theta &= \sum_{mnj} |X_{mn}|^2 \eta_{mnij}^2 \mathbb{E} \left[\frac{1}{\phi_j W_{im} A_{jm} H_{ijn}} \right] \\ a_j^\phi &= \frac{\gamma}{J}, \quad b_j^\phi = \gamma + \sum_{mni} \frac{\mathbb{E}[\theta_i W_{im} A_{jm} H_{ijn}]}{\xi_{mn}} \\ c_j^\phi &= \sum_{mni} |X_{mn}|^2 \eta_{mnij}^2 \mathbb{E} \left[\frac{1}{\theta_i W_{im} A_{jm} H_{ijn}} \right] \\ a_{ijn}^H &= 2\beta, \quad c_{ijn}^H = \sum_m |X_{mn}|^2 \eta_{mnij}^2 \mathbb{E} \left[\frac{1}{\theta_i \phi_j W_{im} A_{jm}} \right] \\ b_{ijn}^H &= \beta \mathbb{E}[G_{ijn} + G_{ijn+1}] + \sum_m \frac{\mathbb{E}[\theta_i \phi_j W_{im} A_{jm}]}{\xi_{mn}} \end{aligned} \quad (24)$$

また、式 (22) を微分してゼロとおくことで、変分下限を最大化する η_{mnij} および ξ_{mn} は次式で与えられる。

$$\eta_{mnij} \propto \mathbb{E} \left[\frac{1}{\theta_i \phi_j W_{im} A_{jm} H_{ijn}} \right]^{-1} \quad (25)$$

$$\xi_{mn} = \sum_{ij} \mathbb{E}[\theta_i \phi_j W_{im} A_{jm} H_{ijn}] \quad (26)$$

式 (22) において、 η_{mnij}^2 は、KL-iCARM における λ_{mnij} と同様に、ビン m ・フレーム n におけるパワー $|X_{mn}|^2$ を各成分に分配する役割を果たしている。ただし、式 (25) の右辺は $\theta_i \phi_j W_{im} A_{jm} H_{ijn}$ の「逆数の平均の逆数」であり、やはり単純な平均 $\mathbb{E}[\theta_i \phi_j W_{im} A_{jm} H_{ijn}]$ に比べてディスカウント効果が発生し、スパースな解へと誘導される。例えば、 θ に着目すると、 $\mathbb{E}[\theta_i] = \frac{K_{a+1}(2\sqrt{bc})\sqrt{c}}{K_a(2\sqrt{bc})\sqrt{b}}$ 、 $\mathbb{E}[\frac{1}{\theta_i}]^{-1} = \frac{K_a(2\sqrt{bc})\sqrt{c}}{K_{a-1}(2\sqrt{bc})\sqrt{b}}$ (添え字 i は省略) であるので、常に $\mathbb{E}[\theta_i]$ より $\mathbb{E}[\frac{1}{\theta_i}]^{-1}$ の方が小さくなる (図 3 右)。特にガンマ過程の場合は $a_i^\theta \rightarrow 0$ であるので、 c_i^θ がゼロに近いほど影響が大きい。

最後に、式 (5) の事前分布で導入された補助変数 $\mathbf{G}$ に関して考える。変分事後分布としてガンマ分布 $q(G_{ijn}) = \text{Gamma}(a_{ijn}^G, b_{ijn}^G)$ を仮定すると、変分パラメータは

$$a_{ijn}^G = 2\beta, \quad b_{ijn}^G = \beta \mathbb{E}[H_{ijn-1} + H_{ijn}] \quad (27)$$

で計算できる。これは KL-iCARM でも同様である。

3.3.3 KL-および IS-iCARM における乗法更新則

ソース $\mathbf{W}$ およびフィルタ $\mathbf{A}$ を構成するパラメータ μ , σ および $\mathbf{a}$ は、乗法更新則を用いて最適化できる。具体的には、式 (15) あるいは式 (22) で与えられる対数尤度の期待値 (または対数周辺尤度 $\log p(\mathbf{X}; \mathbf{W}, \mathbf{A})$) の変分下限の符号を反転したものをコスト関数とみなせば、 $\mu_i \leftarrow G_{\mu_i}^{-1} F_{\mu_i} \mu_i$, $\sigma^2 \leftarrow G_{\sigma^2}^{-1} F_{\sigma^2} \sigma^2$, $\mathbf{a}_j \leftarrow \mathbf{G}_{\mathbf{a}_j}^{-1} \mathbf{F}_{\mathbf{a}_j} \mathbf{a}_j$ とすると、コスト関数を減少させることができる。このとき、各項は

$$\begin{aligned} F_{\mu_i} &= \sum_{mnjh} h(mV_{mnij}^F + h\mu_i V_{mnij}^G) \exp\left(-\frac{(m-h\mu_i)^2}{2\sigma^2}\right) \\ G_{\mu_i} &= \sum_{mnjh} h(mV_{mnij}^G + h\mu_i V_{mnij}^F) \exp\left(-\frac{(m-h\mu_i)^2}{2\sigma^2}\right) \\ F_{\sigma^2} &= \sum_{mnijh} V_{mnij}^F (m - h\mu_i)^2 \exp\left(-\frac{(m-h\mu_i)^2}{2\sigma^2}\right) \\ G_{\sigma^2} &= \sum_{mnijh} V_{mnij}^G (m - h\mu_i)^2 \exp\left(-\frac{(m-h\mu_i)^2}{2\sigma^2}\right) \end{aligned} \quad (28)$$

で与えられる。さらに、KL-iCARM の場合は、

$$\begin{aligned} V_{mnij}^F &= \mathbb{E}[\theta_i \phi_j A_{jm} H_{ijn}], \quad V_{mnij}^G = |X_{mn}| \lambda_{mnij} W_{im}^{-1} \\ \mathbf{F}_{\mathbf{a}_j} &= \sum_{mni} \theta_i \phi_j W_{im} H_{ijn} A_{jm}^3 \mathbf{U}_m \\ \mathbf{G}_{\mathbf{a}_j} &= \sum_{mni} |X_{mn}| \lambda_{mnij} A_{jm}^2 \mathbf{U}_m \end{aligned} \quad (29)$$

となる。一方、IS-iCARM の場合は、

$$\begin{aligned} V_{mnij}^F &= \frac{\mathbb{E}[\theta_i \phi_j A_{jm} H_{ijn}]}{\xi_{mn}} \\ V_{mnij}^G &= |X_{mn}|^2 \eta_{mnij}^2 \mathbb{E}\left[\frac{1}{\theta_i \phi_j W_{im}^2 A_{jm} H_{ijn}}\right] \\ \mathbf{F}_{\mathbf{a}_j} &= \sum_{mni} \frac{\mathbb{E}[\theta_i \phi_j W_{im} H_{ijn}]}{\xi_{mn}} A_{jm}^2 \mathbf{U}_m \\ \mathbf{G}_{\mathbf{a}_j} &= \sum_{mni} |X_{mn}|^2 \eta_{mnij}^2 \mathbb{E}\left[\frac{1}{\theta_i \phi_j W_{im} H_{ijn}}\right] \mathbf{U}_m \end{aligned} \quad (30)$$

となる。ただし、 $\mathbf{a}_j \leftarrow \mathbf{G}_{\mathbf{a}_j}^{-1} \mathbf{F}_{\mathbf{a}_j} \mathbf{a}_j$ で更新すると、 $a_0^j = 1$ は満たされなくなる。そのため、 $a_0^j = 1$ となるように各フィルタ $\mathbf{A}_j$ をスケールリングする処理を導入した。これにより、コスト関数の単調減少性は保証されなくなるが、収束性に実用上問題は無いことを実験により確認している。

4. 評価実験

本章では、KL ダイバージェンスあるいは IS ダイバージェンスに基づく iCARM (KL-iCARM および IS-iCARM) の性能を評価するために行った比較実験について報告する。

4.1 実験条件

実験データは文献 [5, 11] と同様であり、MAPS ピアノデータベース [22] に含まれる「ENSTDkCl」サブセット 30 曲から冒頭 30 秒を抜き出したものを用いた。本研究では、44.1 [kHz]・ステレオのオリジナル音源を 16 [kHz]・モノラルに変換し、窓幅 2048 点、シフト長 160 点の短時間フーリエ変換 (STFT) を行うことで各曲のスペクトログラムを得た。この結果、 $M = 1024$ かつ $N = 3000$ となった。振幅あるいはパワースペクトログラムは、各曲ごとに $\frac{1}{MN} \sum_{mn} |X_{mn}| = 1$ あるいは $\max_{mn} |X_{mn}|^2 = 1$ となるように正規化した。超パラメータは、 $I = 88 + 1$, $J = 10$, $\alpha = 1$, $\beta = \gamma = 0.1$, $H = 20$, $P = 4$, $d = \mathbb{E}_{\text{emp}}[|X_{mn}|]$ or $\mathbb{E}_{\text{emp}}[|X_{mn}|^2]$ とした。最大フィルタ数 $J = 10$ はガンマ過程を正確に近似す

るには小さすぎるが、ピアノの音色のみを扱う本実験においては十分である。基本周波数 $\{\mu_i\}_{i=1}^{88}$ は、標準的なピアノの 88 鍵に対応するように初期化した。

複数の基本周波数は、閾値処理により各フレームごとに検出した。もし、フレーム n においてソース i のゲイン $\sum_j \theta_i \phi_j H_{ijn}$ が閾値よりも大きければ、そのフレームには μ_i が含まれていると判定した。閾値は、30 曲全体に対する F 値の平均が最大化されるように、全体でひとつ定めた。

4.2 実験結果

最初に、提案手法の動作確認のため、楽曲「alb_se2」の冒頭 4.9 秒に対する分解を試みた。この音響信号では、D4, C#4, C4, A3, F#3 の順にピアノ音が一つずつ重畳し、最終的に五重音を構成する。図 4 に示すように、KL-iCARM は適切な個数のソース (5 個の調波構造と 1 個の白色雑音) を見つけることができていた。また、ビン m ・フレーム n における各ソース i の成分 $\mathbb{E}[Y_{mn}^i] = \sum_j \mathbb{E}[\theta_i \phi_j W_{im} A_{jm} H_{ijn}]$ を計算することで、観測データ $\mathbf{X}$ を調波成分と雑音成分に分離することができる。一方、図 5 に示すように、IS-iCARM ではソースやフィルタの個数は過剰に見積もられ、オクターブエラー (正解の F0 の半分を検出) が多く発生する傾向があった。この理由として、IS ダイバージェンスでは、再構成データ Y_{mn} が観測データ X_{mn} を超えてもコスト関数が増加しないことが考えられる。

次に、実験データ全体を評価に用いた。KL-iCARM および IS-iCARM で得られた F 値は、全 30 曲平均でそれぞれ 48.4% および 35.1% であった。この結果は文献 [5, 11] で報告されている最新の成果には及ばないが、無限の複雑さを持つ空間においてスパースな学習を行うことができる性質から、iCARM は依然有望であると考えている。現状のモデルの制約として、適切な個数の基底が推定できたととしても、各時刻におけるオン・オフを判定するために閾値処理が必要なことが挙げられる。今後は、オン・オフに対応するバイナリ変数を導入することを検討したい。

5. おわりに

本研究では、無限複合自己回帰モデル (iCARM) とよぶノンパラメトリックベイズモデルを提案し、観測スペクトログラムを 3 つの因子、すなわち、ソース、フィルタ、それらの組み合わせの音量の積に分解することを試みた。実験の結果、ガンマ過程を用いることで、与えられたデータに対して適切な個数のソースやフィルタを推定することができることが分かった。今後は、多重音基本周波数推定の精度を上げるため、ウェーブレット変換などで得られる対数周波数軸をもつスペクトログラムの利用を検討している。また、フィルタ (スペクトル包絡) は楽器音の音色を表していると考えられるので、さらに、楽器パートの音源分離にも取り組んでいきたい。ソースやフィルタを構成するパラメータのベイズ的な取り扱いも今後の課題である。

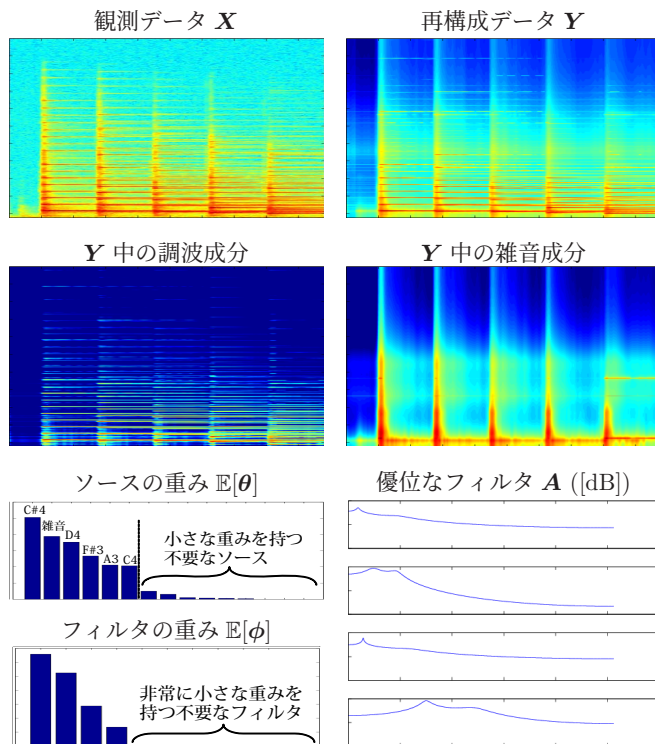


図 4 KL-iCARM による音源分離結果

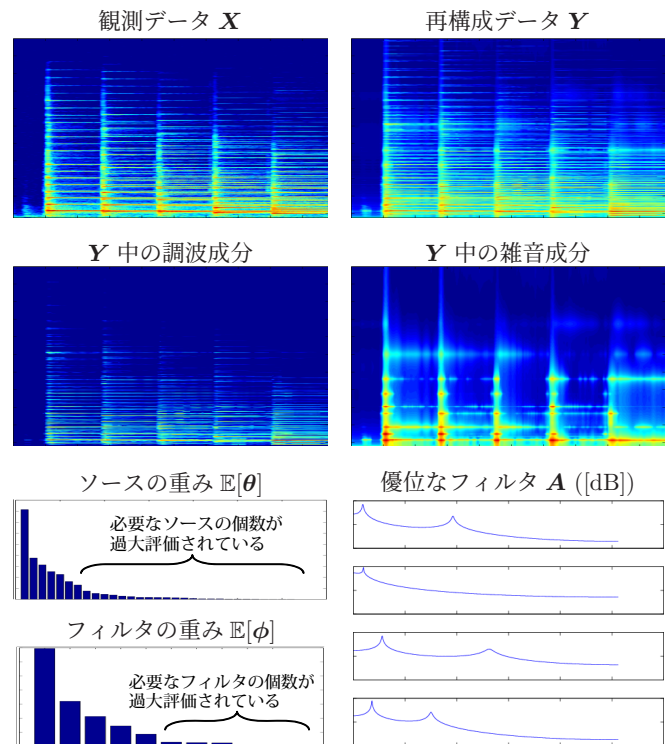


図 5 IS-iCARM による音源分離結果

謝辞: 本研究の一部は, JSPS 科研費 23700184 および JST OngaCREST プロジェクトの支援を受けた。

参考文献

- [1] H. Kameoka and K. Kashino. Composite autoregressive system for sparse source-filter representation of speech. *ICASSP*, pp. 2477–2480, 2009.
- [2] R. Badeau, V. Emiya, and B. David. Expectation-maximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra. *ICASSP*, pp. 3073–3076, 2009.
- [3] J.-L. Durrieu, G. Richard, B. David, and C. Févotte. Source/Filter model for unsupervised main melody extraction from polyphonic audio signals. *IEEE Trans. on ASLP*, 18(3):564–575, 2010.
- [4] T. Virtanen and A. Klapuri. Analysis of polyphonic audio using source-filter model and non-negative matrix factorization. *NIPS Workshop*, 2009.
- [5] J. J. Carabias-Orti, T. Virtanen, P. Vera-Candeas, N. Ruiz-Reyes, and F. J. Cañadas-Quesada. Musical instrument sound multi-excitation model for non-negative spectrogram factorization. *IEEE J. of Sel. Top. in Sig. Proc.*, 5(6):1144–1158, 2011.
- [6] T. Heittola, A. Klapuri, and T. Virtanen. Musical instrument recognition in polyphonic audio using source-filter model for sound separation. *ISMIR*, pp. 327–332, 2009.
- [7] 安良岡直希, 奥乃博. 調波・非調波・音色構造因子分解による音響信号分析と音源分離インターフェースへの応用. 情処研報, 2012-MUS-94, pp. 1–8, 2012.
- [8] R. Hennequin, R. Badeau, and B. David. NMF with time-frequency activations to model nonstationary audio events. *IEEE Trans. on ASLP*, 19(4):744–753, 2011.
- [9] R. Hennequin, R. Badeau, and B. David. Time-dependent parametric and harmonic templates in non-negative matrix factorization. *DAFx*, pp. 1–8, 2010.
- [10] E. Vincent, N. Bertin, and R. Badeau. Adaptive harmonic spectral decomposition for multiple pitch estimation. *IEEE Trans. on ASLP*, 18(3):528–537, 2010.
- [11] N. Bertin, R. Badeau, and E. Vincent. Enforcing har-

- monicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription. *IEEE Trans. on ASLP*, 18(3):538–549, 2010.
- [12] M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. *ICML*, 2010.
- [13] M. Nakano *et al.* Bayesian nonparametric spectrogram modeling based on infinite factorial infinite hidden Markov model. *WASPAA*, pp. 325–328, 2011.
- [14] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. *NIPS*, pp. 556–562, 2000.
- [15] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *NECO*, 21(3):793–830, 2009.
- [16] B. Niedermayer. Improving accuracy of polyphonic music-to-score alignment. *ISMIR*, pp. 585–590, 2009.
- [17] C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the beta-divergence. *NECO*, 23(9):2421–2456, 2011.
- [18] P. Orbanz and Y. W. Teh. Bayesian nonparametric models. *Encyclopedia of Machine Learning*. Springer, 2010.
- [19] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009:Article ID 785152, 2009.
- [20] A. T. Cemgil and O. Dikmen. Conjugate gamma Markov random fields for modelling nonstationary sources. *ICA*, 2007.
- [21] 亀岡弘和, 後藤真孝, 嵯峨山茂樹. スペクトル制御エンベロープによる混合音中の周期および非周期成分の選択的イコライザ. 情処研報, 2006-MUS-66, pp. 77–84, 2006.
- [22] V. Emiya, R. Badeau, and B. David. Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. *IEEE Trans. on ASLP*, 18(6):1643–1654, 2010.

付録

$$\text{Gamma}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad \text{Exponential}(x|\lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$$

$$\text{Poisson}(x|\lambda) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad \text{GIG}(x|a, b, c) = \frac{(b/c)^{\frac{a}{2}} x^{a-1}}{2\mathcal{K}_a(2\sqrt{bc})} e^{-(bx + \frac{c}{x})}$$