

[ポスター講演] 音楽音響信号解析のための 階層ディリクレ過程に基づく無限潜在的調波配分法

吉井 和佳[†] 後藤 真孝[†]

[†] 産業技術総合研究所 〒305-8568 茨城県つくば市梅園 1-1-1 中央第 2

E-mail: †{k.yoshii,m.goto}@aist.go.jp

あらまし 本稿では、混合音である音楽音響信号に含まれる複数の基本周波数を推定するための無限潜在的調波配分法 (Infinite Latent Harmonic Allocation; iLHA) について述べる。本研究では、音響信号の振幅スペクトルを大量の周波数からなるヒストグラムであると解釈し、観測された周波数それぞれが、ある音源を構成するいずれかの倍音に由来しているとみなす。従来、このような見地のもとで確率モデルを定式化するには、本来未知であるはずの音源数と倍音数を事前に指定しなくてはならないモデル選択問題があった。この問題を回避するため、我々は理論上は可算無限個の音源および倍音を許容する階層ノンパラメトリックベイズモデルを提案する。具体的には、階層ディリクレ過程およびベータ 2 パラメータ過程を利用してネスト型無限混合ガウスモデルを定式化し、周辺化変分ベイズ法を用いてベイズ推論を行う。実験では、iLHA は完全に自動化されているにもかかわらず、利用したデータセットに対して性能が最大化されるように試行錯誤で手動設定を行なった従来手法と同等程度の性能を達成した。

キーワード 音楽音響信号解析, 基本周波数推定, 階層ディリクレ過程, ノンパラメトリックベイズ

[Poster Presentation] Infinite Latent Harmonic Allocation based on Hierarchical Dirichlet Process for Music Signal Analysis

Kazuyoshi YOSHII[†] and Masataka GOTO[†]

[†] National Institute of Advanced Industrial Science and Technology (AIST)

Central 2, 1-1-1 Umezono, Tsukuba, Ibaraki, 305-8568 Japan

E-mail: †{k.yoshii,m.goto}@aist.go.jp

Abstract This paper presents a method called the infinite latent harmonic allocation (iLHA) for detecting multiple fundamental frequencies (F0s) in polyphonic music audio signals. We regard amplitude spectra as histograms of numerous observed frequencies and assume each frequency to be generated from one of the harmonic partials contained in a sound source. From this viewpoint, to formulate conventional probabilistic models, however, we have to specify the numbers of sources and partials, which are unknown in advance. To avoid this problem, we propose a hierarchical nonparametric Bayesian model that allows the infinite numbers of sources and partials in theory. More specifically, we formulated nested infinite Gaussian mixture models using a hierarchical Dirichlet process and Beta two-parameter processes and derived a training method based on collapsed variational Bayes. Our experiments showed that fully automated iLHA performed as well as optimally tuned conventional methods.

Key words Music signals analysis, multiple F0 estimation, hierarchical Dirichlet process, nonparametric Bayes

1. はじめに

近年、ベイズ推論は様々な分野で成功を収めているが、古典的なアプローチでは確率モデルの複雑さを適切に指定する必要があるというモデル選択問題を解決する必要があった。複雑さとは、混合ガウスモデルであれば混合数、隠れマルコフモデル

であれば状態数に相当する。通常、確率モデルのパラメータだけでなくその複雑さも未知であるので、事前および事後の不確実性を適切に扱うことが望ましい。しかし、従来は様々な複雑さをもつ確率モデルを個別に学習し、ある基準のもとで最も優れたモデルを一意に決定していた。このような網羅的な学習は効率的ではなく、自然なベイズ的取り扱いとは言えなかった。

この問題に対する強力な解法として、近年ノンパラメトリックベイズ理論が注目されている。「ノンパラメトリック」とは、パラメータ空間（確率モデルの複雑さ）が固定されておらず、無限個のパラメータが存在する（無限の複雑さを考える）という意味である。もし、観測データが無限であれば、その生成過程を表現するのにすべての、すなわち無限個のパラメータが必要になる。しかし、実際には観測データは有限であるので、本来無限個あるうちのほんの一部のパラメータだけで十分であろう。このように、観測データに合わせて確率モデルの「実的な」複雑さが自動的に調節される。例えば、無限混合ガウスモデルでは、観測データが与えられれば、本来無限個存在するガウス分布のうち実際に利用されたガウス分布の個数の事後分布を推定できる。これは古典的な有限混合モデルとは根本的に異なる。無限混合モデルでは、無限個の異なる混合数のモデルが確率的に重なり合っており、混合数を一意に決定せずに学習や予測ができるので、モデル選択問題とは無縁である。

本研究では、近年研究が盛んな多重音基本周波数解析 [1]～[10] に対する現代的なベイズアプローチのさきがけとして、無限潜在的調波配分法 (Infinite Latent Harmonic Allocation; iLHA) を提案する。本手法は、各時刻に含まれる基本周波数の値（パラメータ）や個数（実的な複雑さ）を含め、どのような未知変数でも一意に決定しない。代わりに、音楽音響信号の振幅スペクトルが観測データとして与えられれば、すべての未知変数の事後分布を求める。我々はディリクレ過程と呼ばれるノンパラメトリックな事前分布を用いることで、振幅スペクトルに対するネスト型無限混合ガウスモデルを定式化する。このモデルは、後藤 [1] や亀岡ら [2] の従来手法に着想を得て、ベイズ的な視点からネスト型有限混合ガウスモデルを定式化し、混合数を正の無限大に発散させたときの極限と等価である。結果的に、各時刻には無限個の音源が含まれていてもよく、各音源には無限個の倍音が含まれていてもよい。ただし、ディリクレ過程の超パラメータは、実的な混合数をどの程度にするかに影響するため、適切なものでなければならない。

本研究では、超パラメータに対しても事前分布（超事前分布と呼ぶ）を設定することで、階層ベイズモデルを構成する。後藤 [1] や亀岡ら [2] の研究では、有限個の倍音の相対強度に対するディリクレ事前分布の超パラメータは手動で設定する必要があった。これは基本周波数推定精度に大きな影響を与えるが、適切な値に定めるのは容易ではない。一方、本研究では、無限個の音源や倍音の事前分布であるディリクレ過程の超パラメータに対して無情報超事前分布を設定し、あらゆる可能性を考慮に入れた解析を行う。多重音基本周波数解析において、我々はパラメータ、複雑さ、重要な超パラメータの自然なベイズ的取り扱いをはじめ可能にし、学習を完全に自動化した。

以降の構成は以下の通りである。第 2 章で、振幅スペクトルの観測データとしての解釈を説明する。第 3 章で、本研究のもとになる従来手法を説明する。第 4 章で、従来手法をベイズ的な観点でとらえ直すことで、ネスト型有限混合ガウスモデルに基づく LHA を導出する。第 5 章で、さらに混合数の極限を考慮することでネスト型無限混合ガウスモデルに基づく iLHA を導

出する。第 6 章で、基本周波数推定精度に関する評価実験について報告する。第 7 章で、本研究のまとめを述べる。

2. 振幅スペクトルの統計的解釈

本研究では、音響信号の振幅スペクトルを、周波数の観測回数に関するヒストグラムであると解釈する。このとき、観測される各周波数は統計的に独立であると仮定する。本来、倍音の関係にある周波数は同じ楽器音から生成されているかもしれないが、このような仮定は適当ではない。しかし、このような便宜的な解釈を行うことで、ベイズモデルを厳密に定式化できる。

2.1 仮定

対象とする音楽音響信号は、調波構造を持つ音のみで構成されているものとする。まず、音響信号に対してウェーブレット解析を行い、振幅スペクトルの系列を得る。フレーム数を D とする。あるフレーム d ($1 \leq d \leq D$) における振幅スペクトルにおいて、周波数 f のパワーが a であったとすると、フレーム d において周波数 f は a 回観測されたとみなす。また、振幅スペクトルは調波構造をもつ K 個の基底から構成されていて、各基底は M 個の倍音で構成されているとする。いま、ウェーブレット解析により対数周波数軸 [cent] を考えているので、 M 個の倍音の周波数は整数倍の関係にはないことに注意する。当初、 K および M は既知の定数として扱うが、第 5 章で正の無限大に発散させた場合を考える。

2.2 観測変数と潜在変数

D フレーム中のすべての観測変数を $\mathbf{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_D\}$ で表わす。ここで、 \mathbf{X}_d はフレーム d において観測した周波数の系列であり、 $\mathbf{X}_d = \{x_{d1}, \dots, x_{dN_d}\}$ で表わす。 N_d は観測した周波数の個数であり、振幅スペクトルの全周波数ビンにわたるパワーの和に等しい。 x_{dn} ($1 \leq n \leq N_d$) は 1 次元のベクトルである。周波数の単位は [cent] であるとする^(注1)。

同様に、 \mathbf{X} に対応する潜在変数を $\mathbf{Z} = \{\mathbf{Z}_1, \dots, \mathbf{Z}_D\}$ で表わす。ここで、 \mathbf{Z}_d はフレーム d における潜在変数系列であり、 $\mathbf{Z}_d = \{z_{d1}, \dots, z_{dN_d}\}$ で表わせる。 x_{dn} が基底 k ($1 \leq k \leq K$) の倍音 m ($1 \leq m \leq M$) から生成されたものであるならば、 z_{dn} は z_{dnkm} のみが 1 で他の要素はすべて 0 の KM 次元のベクトルで表わせる。

3. 関連研究との比較

本章では、ネスト型有限混合ガウス分布を用いた従来の基本周波数推定法を紹介し、我々の提案手法である iLHA との相違点を説明する。

3.1 従来研究での定式化

後藤 [1] や亀岡ら [2] は、対数周波数軸上における調波構造を確率密度関数と解釈し、その近似として有限混合ガウス分布をフィッティングさせる手法を提案した。図 1 に示すように、各基底 k ($1 \leq k \leq K$) は M 個の倍音から構成され、各倍音はガウス分布で表現される。このとき、対数周波数軸上の周波数

(注1): 線形周波数 f_h [Hz] から対数周波数 f_c [cent] へは $f_c = 1200 \log_2(f_h / (440 \cdot 2^{\frac{f_c-440}{12}}))$ で変換した。

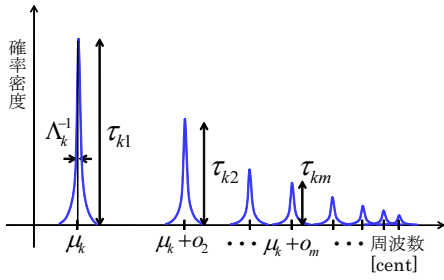


図1 単一基底に対する混合ガウス分布

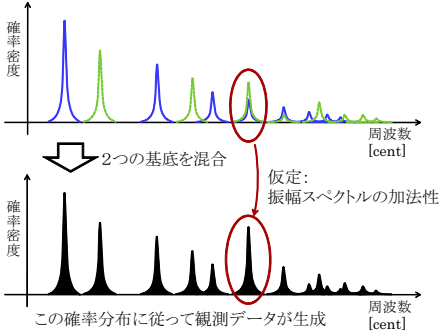


図2 複数基底に対するネスト型混合ガウス分布

$\mathbf{x}[\text{cent}]$ に対する確率密度は、

$$\mathcal{M}_k(\mathbf{x}) = \sum_{m=1}^M \tau_{km} \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_k + \mathbf{o}_m, \boldsymbol{\Lambda}_k^{-1}) \quad (1)$$

で与えられる。ここで、パラメータ $\boldsymbol{\mu}_k$ および $\boldsymbol{\Lambda}_k$ は、ガウス分布の平均および分散であり、基底 k の基本周波数および各倍音まわりでのエネルギーの分散を示す。 τ_{km} ($1 \leq m \leq M$) は倍音 m の相対強度であり、足して1になるよう正規化されている。また、 $\mathbf{o}_m = [1200 \log_2 m]$ であり、 M 個のガウス分布を倍音関係に配置するためのオフセット値である。

図2に示すように、各フレーム d においてこれら K 個の基底を混合することで、ネスト型有限混合ガウス分布を構成できる。このとき、周波数 $\mathbf{x}[\text{cent}]$ に対する確率密度は、

$$\mathcal{M}_d(\mathbf{x}) = \sum_{k=1}^K \pi_{dk} \mathcal{M}_k(\mathbf{x}) \quad (2)$$

で与えられる。ここで、 π_{dk} はフレーム d における基底 k の相対強度である。

3.2 従来研究での学習法

従来は、周波数軸上の振幅スペクトルそのものを確率分布とみなし、ネスト型有限混合ガウス分布とのカルバック・ライブラー情報量 (KLD) を最小化するようなパラメータ $\boldsymbol{\mu}, \boldsymbol{\Lambda}, \boldsymbol{\pi}, \boldsymbol{\tau}$ の点推定を行っていた。表1に基本周波数推定手法の一覧を示す。まず、後藤[1]のPreFEstでは、あらゆる音高に基底を配置することで $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ を固定し、 $\boldsymbol{\pi}, \boldsymbol{\tau}$ のみの推定を行っていた。一方、亀岡ら[2]のHCでは、すべてのパラメータを未知として推定するだけでなく、AICによって適切な K の値を選択していた。ただし、基本的には、これらの手法は異なるフレームごとに独立にパラメータの学習を行っていた。この問題に対処するため、亀岡ら[3]は時間-周波数平面の振幅スペクトログラ

表1 多重音解析手法の比較。

	基底の数	倍音の数	異なるフレームの取り扱い
PreFEst [1]	固定	固定	なし
HC [2]	推定	固定	なし
HTC [3]	固定	固定	時間的連続性を考慮
NMF [9], [10]	固定	—	交換可能
iLHA	無限	無限	交換可能

ムを確率分布とみなし、時間的に強度が変動するように拡張された基底に分解する手法 HTC を提案した。

以上はすべて、事後確率最大化 (MAP) 原理に基づく学習法であった。ベイズ推定と同様に、MAP 推定でもパラメータの事前分布を導入する。しかし、パラメータは常に点推定であり、その不確実性は考慮されていない。また、事前分布の設定に推定結果が大きく影響されるという問題がある。十分な精度が得るためには、特に倍音の相対強度である $\boldsymbol{\tau}$ の事前分布に関して超パラメータを慎重に設定する必要があった。

3.3 提案手法との比較

我々の提案する iLHA は、ノンパラメトリックベイズ学習に基づいており、基底数 K や倍音数 M を事前に指定する必要がない。さらに、無情報超事前分布を導入して階層ベイズモデルを定式化することで、事前分布のチューニングも不要である。また、本来は不十分であるが、非負値行列分解 (NMF) に基づく音源分離手法[9]と同様に、異なるフレーム間の順序は考慮せず、全フレームを同時にモデル化することにした。

4. 潜在的調波配分法

本章では、iLHA を導出するための予備段階として、その有限混合版である LHA について説明する。調波構造を確率的に表現するための基本的なアイデアは3.1節と同じであり、LHA は PreFEst や HC の自然なベイズ的な拡張となっている。しかし、従来は3.2節で述べたとおり、本来観測データであるはずの振幅スペクトルを確率分布として解釈していたため、振幅スペクトルに対する厳密な生成モデルが与えられていなかった。

本研究では、第2章で述べたように、振幅スペクトルは多数観測された周波数のヒストグラムであると解釈する。この解釈のもとで、観測データ \mathbf{X} の生成過程を表現するネスト型有限混合ガウスモデルを厳密に定式化し、LHA におけるベイズ学習法を導出する。

4.1 定式化

図3にLHAにおける確率モデルを示す。完全な同時分布は

$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) p(\mathbf{Z} | \boldsymbol{\pi}, \boldsymbol{\tau}) p(\boldsymbol{\pi}) p(\boldsymbol{\tau}) p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) \quad (3)$$

で与えられる。ここで、前二項はパラメータが与えられたときの観測変数 \mathbf{X} および潜在変数 \mathbf{Z} の尤度であり、後三項はパラメータの事前分布である。具体的には、尤度項は

$$p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{dnkm} \mathcal{N}(x_{dn} | \boldsymbol{\mu}_k + \mathbf{o}_m, \boldsymbol{\Lambda}_k^{-1})^{z_{dnkm}} \quad (4)$$

$$p(\mathbf{Z} | \boldsymbol{\pi}, \boldsymbol{\tau}) = \prod_{dnkm} (\pi_{dk} \tau_{km})^{z_{dnkm}} \quad (5)$$

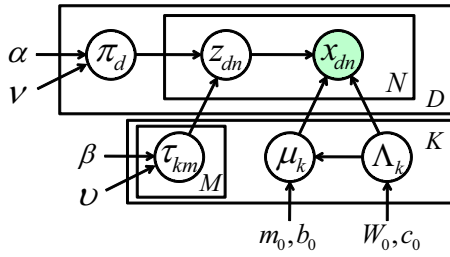


図3 LHAにおけるベイズモデル

で与えられる。また、事前分布として共役事前分布を考える。

$$p(\boldsymbol{\pi}) = \prod_{d=1}^D \text{Dir}(\boldsymbol{\pi}_d | \alpha \boldsymbol{\nu}) \propto \prod_{d=1}^D \prod_{k=1}^K \pi_{dk}^{\alpha \nu_k - 1} \quad (6)$$

$$p(\boldsymbol{\tau}) = \prod_{k=1}^K \text{Dir}(\boldsymbol{\tau}_k | \beta \boldsymbol{\nu}) \propto \prod_{k=1}^K \prod_{m=1}^M \tau_{km}^{\beta \nu_m - 1} \quad (7)$$

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{k=1}^K \mathcal{N}(\boldsymbol{\mu}_k | \mathbf{m}_0, (b_0 \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k | \mathbf{W}_0, c_0) \quad (8)$$

ここで、 $p(\boldsymbol{\pi})$ および $p(\boldsymbol{\tau})$ は K 次元および M 次元ディリクレ分布の積であり、 $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ はガウス・ウィシャート分布の積である。超パラメータに関して、 α および β は集中度と呼ばれ、 ν および ν は正規化されている。また、 \mathbf{m}_0 および b_0 はガウス分布の平均および精度のスケール、 \mathbf{W}_0 および c_0 はウィシャート分布のスケール行列および自由度である。

4.2 変分ベイズ法

ベイズ推定の目的は、観測データ \mathbf{X} が与えられたときに、全ての未知変数の事後分布 $p(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda} | \mathbf{X})$ を求めることである。しかし、この計算を解析的には行うことは困難であるため、変分ベイズ法 (VB) やマルコフ連鎖モンテカルロ法 (MCMC) などの反復解法を利用して事後分布の近似を行う必要がある。一般的に、MCMC の方が精度のよい推定が行えるが、VB の方が収束が高速であるという一長一短がある。

本研究では VB を利用して、効率的に事後分布の推定を行う。いま、解析的に求まるような関数形を仮定した変分事後分布 $q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ を導入し、真の事後分布 $p(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda} | \mathbf{X})$ との KLD を最小化するように $q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ のパラメータ (変分パラメータと呼ぶ) を求めたい。まず、変分事後分布に関して以下の因子分解

$$q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = q(\mathbf{Z})q(\boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \quad (9)$$

が成立することを仮定する。 $q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda})$ を求めるには、VB 版の EM アルゴリズム [11] を利用する。具体的には、以下の VB-E ステップおよび VB-M ステップをエビデンス $p(\mathbf{X})$ の変分下限 (紙面の都合上省略) がほとんど増加しなくなるまで反復すればよい。

$$q^*(\mathbf{Z}) \propto \exp(\mathbb{E}_{\boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}} [\log p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda})]) \quad (10)$$

$$q^*(\boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) \propto \exp(\mathbb{E}_{\mathbf{Z}} [\log p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda})]) \quad (11)$$

4.3 事後分布の反復推定

式 (10) および (11) に従って、変分パラメータの更新則を導出する。

4.3.1 VB-E ステップ

まず、潜在変数 \mathbf{Z} に関する最適な変分事後分布は

$$\begin{aligned} \log q^*(\mathbf{Z}) &= \mathbb{E}_{\boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}} [\log p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + \text{const.} \\ &= \mathbb{E}_{\boldsymbol{\mu}, \boldsymbol{\Lambda}} [\log p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda})] + \mathbb{E}_{\boldsymbol{\pi}, \boldsymbol{\tau}} [\log p(\mathbf{Z} | \boldsymbol{\pi}, \boldsymbol{\tau})] + \text{const.} \\ &= \sum_{dnkm} z_{dnkm} \log \rho_{dnkm} + \text{const.} \end{aligned} \quad (12)$$

で与えられる。ここで、 ρ_{dnkm} を

$$\begin{aligned} \log \rho_{dnkm} &= \mathbb{E}_{\boldsymbol{\pi}_d} [\log \pi_{dk}] + \mathbb{E}_{\boldsymbol{\tau}_k} [\log \tau_{km}] \\ &\quad + \mathbb{E}_{\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k} [\log \mathcal{N}(\mathbf{x}_{dn} | \boldsymbol{\mu}_k + \mathbf{o}_m, \boldsymbol{\Lambda}_k^{-1})] \end{aligned} \quad (13)$$

で定義した。したがって、 $q^*(\mathbf{Z})$ は多項分布

$$q^*(\mathbf{Z}) = \prod_{dnkm} \gamma_{dnkm}^{z_{dnkm}} \quad \text{ただし} \quad \gamma_{dnkm} = \frac{\rho_{dnkm}}{\sum_{km} \rho_{dnkm}} \quad (14)$$

となることが分かる。 γ_{dnkm} は負担率と呼ばれ、観測した周波数 \mathbf{x}_{dn} が基底 k の倍音 m に対応するガウス分布から生成された場合の事後確率である。ここで、 n_{dkm} をフレーム d において基底 k の倍音 m から生成された周波数の個数とする。 n_{dkm} およびその期待値は

$$n_{dkm} = \sum_n z_{dnkm} \quad \mathbb{E}[n_{dkm}] = \sum_n \gamma_{dnkm} \quad (15)$$

で計算できる。あとで必要になるため、十分統計量を

$$\begin{aligned} \mathbb{S}_k[1] &\equiv \sum_{dnm} \gamma_{dnkm} & \mathbb{S}_k[\mathbf{x}] &\equiv \sum_{dnm} \gamma_{dnkm} \mathbf{x}_{dnm} \\ \mathbb{S}_k[\mathbf{x}\mathbf{x}^T] &\equiv \sum_{dnm} \gamma_{dnkm} \mathbf{x}_{dnm} \mathbf{x}_{dnm}^T \end{aligned} \quad (16)$$

と定義しておく。ここで、可読性を向上させるため $\mathbf{x}_{dnm} = \mathbf{x}_{dn} - \mathbf{o}_m$ とした。

4.3.2 VB-M ステップ

次に、パラメータ $\boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}$ に関する最適な変分事後分布は次式で与えられる。

$$q^*(\boldsymbol{\pi}, \boldsymbol{\tau}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{d=1}^D q^*(\boldsymbol{\pi}_d) \prod_{k=1}^K q^*(\boldsymbol{\tau}_k) \prod_{k=1}^K q^*(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) \quad (17)$$

本研究では共役事前分布を利用しているので、事後分布は対応する事前分布と同じ形をとり

$$q^*(\boldsymbol{\pi}_d) = \text{Dir}(\boldsymbol{\pi}_d | \alpha \mathbf{d}) \quad (18)$$

$$q^*(\boldsymbol{\tau}_k) = \text{Dir}(\boldsymbol{\tau}_k | \beta \mathbf{1}) \quad (19)$$

$$q^*(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = \mathcal{N}(\boldsymbol{\mu}_k | \mathbf{m}_k, (b_k \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k | \mathbf{W}_k, c_k) \quad (20)$$

となることが自然に導かれる。このとき、各事後分布の変分パラメータは

$$\alpha_{dk} = \alpha \nu_k + \mathbb{E}[n_{dk}] \quad \beta_{km} = \beta \nu_m + \mathbb{E}[n_{.km}] \quad (21)$$

$$b_k = b_0 + \mathbb{S}_k[1] \quad c_k = c_0 + \mathbb{S}_k[1] \quad (22)$$

$$\mathbf{m}_k = \frac{b_0 \mathbf{m}_0 + \mathbb{S}_k[\mathbf{x}]}{b_0 + \mathbb{S}_k[1]} = \frac{b_0 \mathbf{m}_0 + \mathbb{S}_k[\mathbf{x}]}{b_k} \quad (23)$$

$$\mathbf{W}_k^{-1} = \mathbf{W}_0^{-1} + b_0 \mathbf{m}_0 \mathbf{m}_0^T + \mathbb{S}_k[\mathbf{x}\mathbf{x}^T] - b_k \mathbf{m}_k \mathbf{m}_k^T \quad (24)$$

で与えられる。ここで、ドット (\cdot) はそのインデックスについて和をとることを意味する。

5. 無限潜在的調波配分法

本章では、第4章で説明したネスト型有限混合ガウスモデルをノンパラメトリックベイズ的な視点から拡張することで、ネスト型無限混合ガウスモデルを定式化する。

5.1 基底数の無限化

まず、基底数 K が無限大である場合を考える。具体的には、式 (6) で、各フレーム d ごとに無限次元のディリクレ分布を考えることになる。すると、そのような分布からサンプルされる混合比 π_d は、各基底を選ぶ確率を要素として並べた無限次元のベクトルになる。 π_d は無限次元の離散分布のパラメータであるため、無限個の要素の和が1になるよう正規化されている。現実的には、ごく一部の要素のみが意味のある値をとり、それ以外の無限個の要素はほとんどゼロと等しい。このような確率過程をディリクレ過程 (DP) と呼び、

$$G_d \sim \text{DP}(\alpha, G_0) \quad \text{あるいは} \quad \pi_d \sim \text{DP}(\alpha, \nu) \quad (25)$$

と書く。ここで、 α は集中度、 G_0 は基底測度と呼ばれ、 G_d はサンプルされる測度である。いま、 G_0 および G_d はともに無限次元の離散分布であり、それぞれ混合比 ν および π_d に従って可算無限個の調波構造基底に対して確率を割り当てることができる。具体的には、 G_0 および G_d は

$$G_0(\phi) = \sum_{k=1}^{\infty} \nu_k \delta_{\phi_k} \quad G_d(\phi) = \sum_{k=1}^{\infty} \pi_{dk} \delta_{\phi_k} \quad (26)$$

と書ける。ここで δ_{ϕ_k} は調波構造基底が取りうるパラメータ空間 Φ におけるディラック測度で、 $\phi_k \in \Phi$ において正の無限大をとり、それ以外ではゼロである。 G_0 および G_d は Φ 上で積分すると1になるので正しく確率測度である。

集中度 α は G_0 と G_d の違い、すなわち ν と π_d との違いを制御する重要なパラメータである。 $\alpha \rightarrow \infty$ で両者は完全に一致する。しかし、現実的には $\alpha < 1$ であることが多く、 ν に比べて π_d は一部の基底に確率質量がより偏った分布になることが多い。すなわち、全 D フレームを見渡せば多数の基底が選ばれるが、各フレームに限ってみればより限られた種類の基底が選ばれることになり、我々の直感と合致する。残る問題は、DP のもう一つのパラメータである G_0 すなわち無限次元の ν をどのように設定すればよいかである。

この問題を解決するには、階層ディリクレ過程 (HDP) [12] を利用すればよい。すなわち、 D 個のディリクレ過程の基底測度 G_0 を生成する事前分布として上位階層の DP を考える。具体的には、式 (6) 中で無限次元ディリクレ分布の超パラメータとなっている ν が、上位 DP からのサンプルであると考えられる。このモデルでは、上位 DP で全フレームにわたる基底の選ばれ方 ν がサンプルされ、各フレーム d における DP でフレーム d に特化した基底の選ばれ方 π_d がサンプルされる。このとき、DP からの具体的なサンプル ν は、棒折り過程 (Stick-Breaking Construction) [12] に従うことが知られている。

$$\nu_k = \tilde{\nu}_k \prod_{k'=1}^{k-1} (1 - \tilde{\nu}_{k'}) \quad \tilde{\nu}_k \sim \text{Beta}(1, \gamma) \quad (27)$$

ここで、 γ は上位 DP の集中度である。 ν は GEM(Griffiths-Engen-McCloskey) 分布に従うと表現することもでき、 $\nu \sim \text{GEM}(\gamma)$ と表わす。

HDP の最大の特徴は、 D 個の下位 DP が上位 DP に関して条件付き独立となり、異なるフレーム間で多くの基底が共有できる点にある。現実には観測データは有限であるので、上位 DP からのサンプル ν に従って選ばれる基底の個数は有限である。さらに、各下位 DP からのサンプル π_d に従って基底が絞り込まれる。もし、異なるフレームの DP が完全に独立であれば、それぞれが独立に無限個の基底の中から有限個を選ぶことになるので、異なるフレームで選ばれる基底が完全に排他的なものとなってしまふ。

5.2 倍音数の無限化

次に、基底数 M が無限大である場合を考える。先ほどと違って、異なる基底間での倍音群の共有化を考える必要はないので、通常の DP で十分である。具体的には、式 (7) で、各フレーム k ごとに独立な無限次元のディリクレ分布を考えることになる。本研究では、倍音の相対強度 τ に対してより適切な事前分布が与えられるようにするため、DP の一般化であるベータ2パラメータ過程を利用する。このとき、棒折り過程 $\tau \sim \text{GEM}(\beta\lambda_1, \beta\lambda_2)$ は

$$\tau_{km} = \tilde{\tau}_{km} \prod_{m'=1}^{m-1} (1 - \tilde{\tau}_{km'}) \quad (28)$$

$$\tilde{\tau}_{km} \sim \text{Beta}(\beta\lambda_1, \beta\lambda_2) \quad (29)$$

と与えられる。ここで、 β は正の実数であり、 λ は $\lambda_1 + \lambda_2 = 1$ を満たすものとする。

5.3 超事前分布の導入と定式化

超パラメータである $\alpha, \beta, \gamma, \lambda$ は事後分布の推定結果に大きな影響を与えるので、従来のように手動で設定することはせず、超事前分布を導入して不確実性を適切に扱う。

$$p(\alpha) = \text{Gam}(\alpha|a_\alpha, b_\alpha) \quad p(\gamma) = \text{Gam}(\gamma|a_\gamma, b_\gamma) \quad (30)$$

$$p(\beta) = \text{Gam}(\beta|a_\beta, b_\beta) \quad p(\lambda) = \text{Beta}(\lambda|u_1, u_2) \quad (31)$$

ここで、 $a_{\{\alpha, \beta, \gamma\}}$ および $b_{\{\alpha, \beta, \gamma\}}$ は、ガンマ分布の形状および比率パラメータで、 u_1, u_2 は、ベータ分布のパラメータである。

図4に iLHA における確率モデルを示す。このとき、完全な同時分布は

$$\begin{aligned} p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \tilde{\boldsymbol{\tau}}, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \alpha, \beta, \gamma, \boldsymbol{\lambda}, \tilde{\boldsymbol{\nu}}) \\ = p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\boldsymbol{\mu}, \boldsymbol{\Lambda})p(\mathbf{Z}|\boldsymbol{\pi}, \tilde{\boldsymbol{\tau}}) \\ p(\boldsymbol{\pi}|\alpha, \tilde{\boldsymbol{\nu}})p(\tilde{\boldsymbol{\tau}}|\beta, \boldsymbol{\lambda})p(\alpha)p(\beta)p(\gamma)p(\boldsymbol{\lambda})p(\tilde{\boldsymbol{\nu}}|\gamma) \end{aligned} \quad (32)$$

と与えられる。ここで、 $p(\mathbf{Z}|\boldsymbol{\pi}, \tilde{\boldsymbol{\tau}})$ は式 (29) を式 (5) に代入して得られ、 $p(\boldsymbol{\pi}|\alpha, \tilde{\boldsymbol{\nu}})$ は式 (6) と同じである。 $p(\tilde{\boldsymbol{\nu}}|\gamma)$ および $p(\tilde{\boldsymbol{\tau}}|\beta, \boldsymbol{\lambda})$ は式 (27) および (29) に従って次式で定義できる。

$$p(\tilde{\boldsymbol{\nu}}|\gamma) = \prod_k \text{Beta}(\tilde{\nu}_k|1, \gamma) \quad (33)$$

$$p(\tilde{\boldsymbol{\tau}}|\beta, \boldsymbol{\lambda}) = \prod_{km} \text{Beta}(\tilde{\tau}_{km}|\beta\lambda_1, \beta\lambda_2) \quad (34)$$

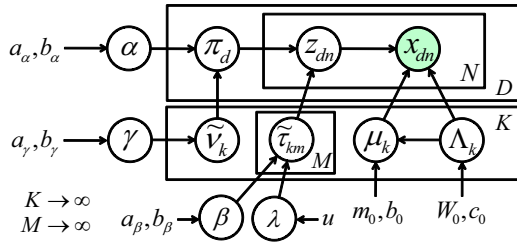


図4 iLHA における階層ベイズモデル

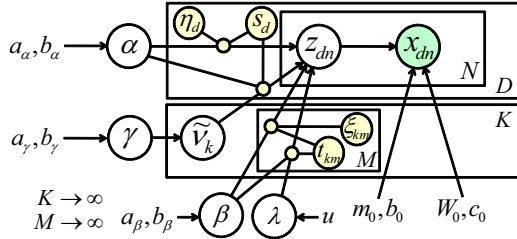


図5 パラメータの積分消去と補助変数の導入

5.4 周辺化変分ベイズ法

HDP に基づく階層ベイズモデルを学習するには、より洗練された学習法である周辺化変分ベイズ法 (CVB) [13] を利用する。通常の VB では、式 (9) の因子分解における潜在変数とパラメータの独立性の仮定が強すぎて、学習後の変分事後分布が真の事後分布と大きく異なったものになる可能性がある。この問題を解決するため、CVB ではパラメータを積分消去して潜在変数のみの空間を考え、観測サンプル間の独立性を仮定する。通常、観測サンプル間には弱い相関しか残っていないため、この仮定は理にかなっている。また、通常の VB は、変分事後分布を解析的に導出するという性質のため、尤度関数-事前分布間に共役性を仮定した確率モデルにしか適用できない。一方、CVB は、階層ベイズモデルのように事前分布-超事前分布間で共役性が崩れる確率モデルに対しても適用できる場合がある。

図5にパラメータ π, τ, μ, Λ を周辺化により積分消去した確率モデルを示す。このとき、残りの確率変数の周辺同時分布は

$$p(\mathbf{X}, \mathbf{Z}, \alpha, \beta, \gamma, \lambda, \tilde{\nu}) = p(\mathbf{X}|\mathbf{Z})p(\mathbf{Z}|\alpha, \beta, \lambda, \tilde{\nu})p(\alpha)p(\beta)p(\gamma)p(\lambda)p(\tilde{\nu}|\gamma) \quad (35)$$

で与えられる。ここで、図4で示したもとの確率モデルにおいては、パラメータ π, τ, μ, Λ に関して共役性が成立しているため、前二項は以下の通り解析的に計算できる。

$$p(\mathbf{X}|\mathbf{Z}) = (2\pi)^{-\frac{n_{\cdot\cdot}}{2}} \prod_k \left(\frac{b_0}{b_{zk}} \right)^{\frac{1}{2}} \frac{B(\mathbf{W}_0, c_0)}{B(\mathbf{W}_{zk}, c_{zk})} \quad (36)$$

$$p(\mathbf{Z}|\alpha, \beta, \lambda, \tilde{\nu}) = \prod_d \frac{\Gamma(\alpha)}{\Gamma(\alpha + n_{d\cdot})} \prod_k \frac{\Gamma(\alpha\nu_k + n_{dk\cdot})}{\Gamma(\alpha\nu_k)} \prod_{km} \frac{\Gamma(\beta)\Gamma(\beta\lambda_1 + n_{\cdot km})\Gamma(\beta\lambda_2 + n_{\cdot k > m})}{\Gamma(\beta\lambda_1)\Gamma(\beta\lambda_2)\Gamma(\beta + n_{\cdot k \geq m})} \quad (37)$$

ここで、 $b_{zk}, \mathbf{W}_{zk}, c_{zk}$ は式 (22) および (24) を計算する際に、式 (16) 中で負担率 γ_{dnkm} の代わりに潜在変数の値 z_{dnkm} を利用して得られる値である。また、 $B(\mathbf{W}_0, c_0)$ および $B(\mathbf{W}_{zk}, c_{zk})$ はウィシャート分布の正規化項である。

ここで、式 (35) に対して CVB を適用することを考えると、全ての超パラメータに関して共役性が成立しているわけではないので、変分事後分布を解析的に導出することができない。この問題を解決するため、データ拡張 [13] と呼ばれる手法を用いて補助変数を導入し、補助変数の変分事後分布の最適化を介して超パラメータの変分事後分布の最適化を行う。いま、 η_d および ξ_{km} をベータ分布に従う確率変数とし、 s_{dk} および $t_{km} = \{t_{km1}, t_{km2}\}$ を $1 \leq s_{dk} \leq n_{dk\cdot}$, $1 \leq t_{km1} \leq n_{\cdot km}$, $1 \leq t_{km2} \leq n_{\cdot k > m}$ を満たす正の整数であるとする。このとき、式 (37) は補助変数を積分消去したあとの \mathbf{Z} の周辺分布であると考えられる。積分消去する前の \mathbf{Z} と補助変数との同時分布は

$$p(\mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}|\alpha, \beta, \lambda, \tilde{\nu}) = \prod_d \frac{\eta_d^{\alpha-1} (1-\eta_d)^{n_{d\cdot}-1}}{\Gamma(n_{d\cdot})} \prod_k \binom{n_{dk\cdot}}{s_{dk}} (\alpha\nu_k)^{s_{dk}} \prod_{km} \frac{\xi_{km}^{\beta-1} (1-\xi_{km})^{n_{\cdot k \geq m}-1}}{\Gamma(n_{\cdot k \geq m})} \binom{n_{\cdot km}}{t_{km1}} (\beta\lambda_1)^{t_{km1}} \binom{n_{\cdot k > m}}{t_{km2}} (\beta\lambda_2)^{t_{km2}} \quad (38)$$

で与えられる。ここで、 $\binom{[]}{[]}$ は組み合わせ数の一種である第一種スターリング数を表す。この分布を補助変数に関して周辺化すると、ベータ分布の正規化項およびガンマ関数の性質に着目することで、式 (37) が得られることは容易に確かめられる。この結果、補助変数を考慮した周辺同時分布は次式で与えられる。

$$p(\mathbf{X}, \mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}, \alpha, \beta, \gamma, \lambda, \tilde{\nu}) = p(\mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}|\alpha, \beta, \lambda, \tilde{\nu})p(\alpha)p(\beta)p(\gamma)p(\lambda)p(\tilde{\nu}|\gamma) \quad (39)$$

CVB では、変分事後分布に関して以下の因子分解

$$q(\mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}, \alpha, \beta, \gamma, \lambda, \tilde{\nu}) = q(\alpha, \beta, \gamma, \lambda)q(\tilde{\nu})q(\eta, \xi, \mathbf{s}, \mathbf{t}|\mathbf{Z}) \prod_{dn} q(z_{dn}) \quad (40)$$

を仮定する。また、実際には無限個の基底や倍音を計算機で扱うことはできないため、混合数に関して打ち切り近似を導入する。すなわち、ある K および M があつたときに、 $k > K, m > M$ を満たす k, m について $q(z_{dnkm}) = 0$ を仮定する。この仮定により、無限混合モデルが有限混合モデルと等価になるわけではないことに注意する。 K や M の値を大きくとるほど事後分布をより正確に近似できるようになるのであって、有限混合モデルにおける混合数を適切に推定する問題とは全く性質が異なる。未知変数の変分事後分布 $q(\mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}, \alpha, \beta, \gamma, \lambda, \tilde{\nu})$ を推定するには、以下の CVB-E ステップと CVB-M ステップをエビデンス $p(\mathbf{X})$ (紙面の都合上省略) がほとんど増加しなくなるまで反復すればよい。

$$q^*(z_{dn}) \propto \exp(\mathbb{E}_{\mathbf{Z}^{-dn}, \alpha, \beta, \gamma, \lambda, \tilde{\nu}} [\log \text{式 (35)}]) \quad (41)$$

$$q^*(\alpha, \beta, \gamma, \lambda) \propto \exp(\mathbb{E}_{\mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}, \tilde{\nu}} [\log \text{式 (39)}]) \quad (42)$$

$$q^*(\tilde{\nu}) \propto \exp(\mathbb{E}_{\mathbf{Z}, \eta, \xi, \mathbf{s}, \mathbf{t}, \alpha, \beta, \gamma, \lambda} [\log \text{式 (39)}]) \quad (43)$$

$$q^*(\eta, \xi, \mathbf{s}, \mathbf{t}|\mathbf{Z}) \propto \exp(\mathbb{E}_{\mathbf{Z}, \alpha, \beta, \gamma, \lambda, \tilde{\nu}} [\log \text{式 (39)}]) \quad (44)$$

ここで、 $-dn$ は d, n 以外の全てのインデックスを表す。

5.5 事後分布の反復推定

式 (41)–(44) に従って、変分パラメータの更新則を導出する。

5.5.1 CVB-E ステップ

$z_{dnkm} = 1$ となる時の変分事後確率は次式で計算できる。

$$\begin{aligned} \log q^*(z_{dnkm} = 1) &= \mathbb{E}_{\mathbf{z}^{-dn}} \left[\log \left(\mathbb{G}[\alpha\nu_k] + n_{dk}^{-dn} \right) \right] \\ &+ \mathbb{E}_{\mathbf{z}^{-dn}} \left[\log \left(\frac{\mathbb{G}[\beta\lambda_1] + n_{.km}^{-dn}}{\mathbb{E}[\beta] + n_{.k \geq m}^{-dn}} \prod_{m'=1}^{m-1} \frac{\mathbb{G}[\beta\lambda_2] + n_{.k > m'}^{-dn}}{\mathbb{E}[\beta] + n_{.k \geq m'}^{-dn}} \right) \right] \\ &+ \mathbb{E}_{\mathbf{z}^{-dn}} \left[\log \mathcal{S}(\mathbf{x}_{dnm} | \mathbf{m}_{zk}^{-dn}, \frac{b_{zk}^{-dn}}{1 + b_{zk}^{-dn}} c_{zk}^{-dn} \mathbf{W}_{zk}^{-dn}, c_{zk}^{-dn}) \right] \\ &+ \text{const.} \end{aligned} \quad (45)$$

ここで、 $\mathbb{G}[x]$ は確率変数 x に関する幾何平均 $\exp(\mathbb{E}[\log x])$ であり、 $\mathcal{S}(x | \boldsymbol{\mu}, \boldsymbol{\Lambda}, \nu)$ は平均 $\boldsymbol{\mu}$ 、精度 $\boldsymbol{\Lambda}$ 、自由度 ν のスチューデント t 分布を表す。 $\mathbf{m}_{zk}^{-dn}, b_{zk}^{-dn}, \mathbf{W}_{zk}^{-dn}, c_{zk}^{-dn}$ は、式 (22), (23) および (24) を計算する際に、式 (16) 中で負担率 γ_{dnkm} の代わりに d, n 以外の潜在変数の値 z_{dnkm} を利用して得られる値である。式 (45) の各項は、一次あるいは二次近似を用いることで精度よく計算可能である [13]–[15]。

5.5.2 CVB-M ステップ

まず、DP の集中度 α, β, γ に関する最適な変分事後分布はガンマ分布となる。

$$q^*(\alpha) \propto \alpha^{a_\alpha + \mathbb{E}[s..] - 1} e^{-\alpha(b_\alpha - \sum_d \mathbb{E}[\log \eta_d])} \quad (46)$$

$$q^*(\beta) \propto \beta^{a_\beta + \mathbb{E}[t..] - 1} e^{-\beta(b_\beta - \sum_{km} \mathbb{E}[\log \xi_{km}])} \quad (47)$$

$$q^*(\gamma) \propto \gamma^{a_\gamma + K - 1} e^{-\gamma(b_\gamma - \sum_k \mathbb{E}[\log(1 - \tilde{\nu}_k)])} \quad (48)$$

λ および $\tilde{\nu}$ に関する最適な変分事後分布はベータ分布となる。

$$q^*(\lambda) \propto \lambda_1^{u_1 + \mathbb{E}[t..] - 1} \lambda_2^{u_2 + \mathbb{E}[t..] - 1} \quad (49)$$

$$q^*(\tilde{\nu}_k) \propto \tilde{\nu}_k^{1 + \mathbb{E}[s..] - 1} (1 - \tilde{\nu}_k)^{\mathbb{E}[\gamma] + \mathbb{E}[s..] - 1} \quad (50)$$

さらに、補助変数 η, ξ, \mathbf{s}, t に関する最適な変分事後分布は

$$q^*(\eta_d) \propto \eta_d^{\mathbb{E}[\alpha] - 1} (1 - \eta_d)^{n_{d..} - 1} \quad (51)$$

$$q^*(\xi_{km} | \mathbf{Z}) \propto \xi_{km}^{\mathbb{E}[\beta] - 1} (1 - \xi_{km})^{n_{k \geq m} - 1} \quad (52)$$

$$q^*(s_{dk} = s | \mathbf{Z}) \propto \begin{bmatrix} n_{dk} \\ s \end{bmatrix} \mathbb{G}[\alpha\nu_k]^s \quad (53)$$

$$q^*(t_{km1} = t | \mathbf{Z}) \propto \begin{bmatrix} n_{.km} \\ t \end{bmatrix} \mathbb{G}[\beta\lambda_1]^t \quad (54)$$

$$q^*(t_{km2} = t | \mathbf{Z}) \propto \begin{bmatrix} n_{.k > m} \\ t \end{bmatrix} \mathbb{G}[\beta\lambda_2]^t \quad (55)$$

で与えられる。ここで、式 (46)–(50) において必要な、事後分布のもとでの補助変数の期待値を計算する際には注意を要する。まず、期待値 $\mathbb{E}[\log \eta_d]$ については、 $n_{d..} = N_d$ が定数であるので、 $\mathbb{E}[\log \eta_d] = \psi(\mathbb{E}[\alpha]) - \psi(\mathbb{E}[\alpha] + n_{d..})$ として求めることができる。ここで、 ψ はディガンマ関数である。一方、期待値 $\mathbb{E}[\log \xi_{km}]$ については、同様に $\mathbb{E}[\log \xi_{km} | \mathbf{Z}] = \psi(\mathbb{E}[\beta]) - \psi(\mathbb{E}[\beta] + n_{k \geq m})$ とできるが、さらに変分事後分布 $q(\mathbf{Z})$ のもとでの期待値を計算する必要がある。しかし、ディガンマ関数は原点付近で対数関数よりはるかに急速に負の無限大に発散するので、単純なテ

表 2 フレームレベルでの F0 推定精度 (F 値 [%])

楽曲番号	手動で最適化		完全に自動化	
	PreFEst [1]	HTC [3]	LHA	iLHA
RWC-MDB-J-2001 No.1	75.8	79.0	70.7	82.2
RWC-MDB-J-2001 No.2	78.5	78.0	69.1	77.9
RWC-MDB-J-2001 No.6	70.4	78.3	49.8	71.2
RWC-MDB-J-2001 No.7	83.0	86.0	70.2	85.5
RWC-MDB-J-2001 No.8	85.7	84.4	55.9	84.6
RWC-MDB-J-2001 No.9	85.9	89.5	68.9	84.7
RWC-MDB-C-2001 No.30	76.0	83.6	81.4	81.6
RWC-MDB-C-2001 No.35	72.8	76.0	58.9	79.6
合計	79.4	82.0	65.8	81.7

イラー展開で近似を行うことはできない。この問題を解決するには、 $n_{.k \geq m} = 0$ を正確に扱い、 $n_{.k \geq m} > 0$ に対してのみテイラー展開に基づく二次近似を適用すればよい (詳細は文献 [13] を参照のこと)。他の補助変数の期待値 $\mathbb{E}[s_{dk}]$, $\mathbb{E}[t_{km1}]$ および $\mathbb{E}[t_{km2}]$ に関しても、同様の特別な処理を適用して計算を行う。

最終的に F0 の値を求めるには、積分消去したパラメータ $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ の変分事後分布を明示的に求め、その MAP 値をとる。具体的には、CVB-E ステップで得られた事後分布 $q(\mathbf{Z})$ を用いて第 4 章における通常の VB-M ステップを一度だけ実行する。

6. 評価実験

本章では、提案する iLHA の性能を検証するため行った比較実験の結果について報告する。

6.1 実験条件

本実験では、RWC 研究用音楽データベースのジャズ音楽およびクラシック音楽 [16] に収録されているピアノソロおよびギターソロ合計 8 曲の冒頭 23 秒に対して、LHA および iLHA の F0 推定精度を評価した。このデータセットは、亀岡らの HTC の研究 [3] で使われたものと同一である。観測データである振幅スペクトルを得るには、時間分解能 16[ms] でガボールウェーブレット解析を行った。音響信号中の正しい F0 の値と時間位置は手作業で付与したものを用いた。F0 推定精度の評価尺度として、フレームレベルでの F 値に着目することにした。LHA および iLHA の事前分布および超事前分布は、すべて無情報事前分布となるように設定し、恣意性を排除した。基底数 K および倍音数 M の値は、これらの楽曲を解析するのに十分に大きいと思われる 60 および 15 とした。各フレーム d ごとに複数の F0 を出力するには、相対強度 π_{dk} と観測周波数の個数 N_d との積がある閾値以上となる基底 k の基本周波数のみを抽出した。このとき、閾値は従来研究と同じく手動で最適化した。

性能比較のため、文献 [3] に報告されている PreFEst および HTC による F0 推定精度を参照した。本研究と正解データは少し異なるが、おおまかに性能を比較するには十分であると考えられる。従来の実験では、PreFEst および HTC における基底数 K 、倍音数 M 、事前分布やその重みなどは、F0 推定精度が最適化されるように正解データを用いてチューニングされていた。これは本来現実的な設定ではないが、各手法の性能の上限を調査するためであったことに注意する。

6.2 実験結果

表 2 に実験結果を示す。iLHA はチューニングやモデル選択を不要としているにもかかわらず、最適化された HTC とほぼ同等の性能を達成することができた。これは、ノンパラメトリックな確率モデルは、最適化されたパラメトリックな確率モデルと同等の性能を達成することができるという経験的な知見とよく一致している。また、HTC は PreFEst と比べてよい精度を達成した。このことは、HTC のように基底の時間的連続性を適切にモデル化することで、iLHA の性能はさらに向上する余地があることを示唆している。

一方、LHA の性能は他の手法よりずっと悪いものになってしまった。この理由は、LHA は階層ベイズモデルに基づいておらず、本来は事前分布を正確に設定する必要があるからである。実際、PreFEst や HTC に対して無情報事前分布を用いると、F0 推定精度は極めて悪化することを確認している。対照的に、iLHA は安定して高い精度を達成できた。

7. おわりに

本稿では、多重音に含まれる未知個数の基本周波数を推定する目的でノンパラメトリックベイズ理論が利用可能であることに着目し、無限潜在的調波配分法 (iLHA) を提案した。本手法は、振幅スペクトル中に上限のない数の基底が存在することを許容し、各基底は上限のない数の倍音で構成されていることを許容する。具体的には、階層ディリクレ過程およびディリクレ過程の一般化であるベータ 2 パラメータ過程を利用することで、振幅スペクトルに対するネスト型無限混合ガウスモデルを定式化した。本モデルは階層ベイズモデルとなっており、事後分布推定に影響を与える事前分布を恣意的に設定する必要がなく、それも含めて観測データのみから推論することができる。モデルの学習には収束の速い周辺化ベイズ法を利用した。実験の結果、無情報事前分布に基づく iLHA が、試行錯誤的に最適化された従来手法と同等以上の性能を達成することを確認した。我々の知る限り、ノンパラメトリックベイズ理論を多重音基本周波数解析に応用したのは本研究が初めてである。

ノンパラメトリックベイズ理論は、音楽情報処理のさまざまな領域で現れるモデル選択問題に対して強力な統一的な解法を提供する。例えば、楽曲の構造を解析するとき、いくつかのセクションに分割すればよいのであろうか。楽曲群をその音楽内容に基づいて、ユーザ群をその嗜好に基づいて分類するとき、いくつかのグループにクラスタリングすればよいのであろうか。このとき、観測データのみを考えるのではなく、無限にデータがあれば無限個のセクションやグループが存在するはずであるが、我々はその一部のみを確率的に観測できているにすぎないという考え方をすることで、原理的にモデル選択問題から解放され、良好な汎化性能を獲得可能である。今後、この強力な枠組みをさまざまな形で応用していきたい。

謝辞 本研究は、JST CREST および科研費 20800084 の支援を受けて行われた。実験を行うにあたり、貴重なデータを提供してくださった亀岡弘和氏 (NTT) に感謝する。

文 献

- [1] M. Goto, "A Real-Time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals," *Speech Communication*, Vol. 43, No. 4, pp. 311–329, 2004.
- [2] H. Kameoka, T. Nishimoto, and S. Sagayama, "Separation of Harmonic Structures based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds," *ICASSP*, Vol. 4, pp. 297–300, 2004.
- [3] H. Kameoka, T. Nishimoto, and S. Sagayama, "A Multipitch Analyzer based on Harmonic Temporal Structured Clustering," *IEEE Trans. on ASLP*, Vol. 15, No. 3, pp. 982–994, 2007.
- [4] A. Klapuri, "Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model," *IEEE Trans. on ASLP*, Vol. 16, No. 2, pp. 255–266, 2008.
- [5] A. Cemgil, H. Kappen, and D. Barber, "A Generative Model for Music Transcription," *IEEE Trans. on ASLP*, Vol. 14, No. 2, pp. 679–694, 2006.
- [6] P. Peeling, A. Cemgil, and S. Godsill, "Generative Spectrogram Factorization Models for Polyphonic Piano Transcription," *IEEE Trans. on ASLP*, Vol. 18, No. 3, pp. 519–527, 2010.
- [7] E. Vincent, N. Bertin, and R. Badeau, "Adaptive Harmonic Spectral Decomposition for Multiple Pitch Estimation," *IEEE Trans. on ASLP*, Vol. 18, No. 3, pp. 528–537, 2010.
- [8] N. Bertin, R. Badeau, and E. Vincent, "Enforcing Harmonicity and Smoothness in Bayesian Non-Negative Matrix Factorization Applied to Polyphonic Music Transcription," *IEEE Trans. on ASLP*, Vol. 18, No. 3, pp. 538–549, 2010.
- [9] P. Smaragdis and J. Brown, "Nonnegative Matrix Factorization for Polyphonic Music Transcription," *WASPAA*, 2003.
- [10] A. Cont, "Realtime Multiple Pitch Observation Using Sparse Nonnegative Constraints," *ISMIR*, pp. 206–211, 2006.
- [11] H. Attias, "A Variational Bayesian Framework for Graphical Models," *NIPS*, pp. 209–215, 2000.
- [12] Y. W. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical Dirichlet Processes," *J. of Am. Stat. Assoc.*, Vol. 101, No. 476, pp. 1566–1581, 2006.
- [13] Y. W. Teh, K. Kurihara, and M. Welling, "Collapsed Variational Inference for HDP," *NIPS*, Vol. 20, 2008.
- [14] J. Sung, Z. Ghahramani, and S.-Y. Bang, "Latent-Space Variational Bayes," *IEEE Trans. on PAMI*, Vol. 30, No. 12, pp. 2236–2242, 2008.
- [15] J. Sung, Z. Ghahramani, and S.-Y. Bang, "Second-Order Latent-Space Variational Bayes for Approximate Bayesian Inference," *IEEE Sig. Proc. Letters*, Vol. 15, pp. 918–921, 2008.
- [16] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース," *情報処理学会論文誌*, Vol. 45, No. 3, pp. 728–738, 2004.