

正弦波重畳モデルのパラメータ最適化アルゴリズムの導出

亀岡 弘和[†] 小野 順貴[†] 嵯峨山茂樹[†]

[†] 東京大学大学院情報理工学系研究科

E-mail: †{kameoka,onono,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 正弦波重畳モデル [1] は、McAulay らが分析合成系に適用し、高品質な合成音声を生成できることを示して以来、分析合成系に留まらず、テキスト音声合成、音響信号符号化圧縮などの用途としても用いられてきた。高品質な合成音声を生成できることは、正弦波重畳モデルが音声や楽音などの音響信号を極めて良く表現するモデルであることの根拠となっているため、音源分離への応用に対する期待も大きい。この枠組においては、正弦波重畳モデルのパラメータの推定精度があらゆる応用における性能に直結する。本章で提案する最適化アルゴリズムは、正弦波重畳モデルのパラメータを効果的に推定するためのものであり、補助関数を介した目的関数の非増加更新法の原理を基礎とする。この考え方は、EM アルゴリズムのヒントにその本質が何であるかを理解した上で明らかになったものである。キーワード 正弦波重畳モデル, EM アルゴリズム, 補助関数, Gabor 変換

Parameter Optimization Algorithm for Sinusoidal Signal Model

Hirokazu KAMEOKA[†], Nobutaka ONO[†], and Shigeki SAGAYAMA[†]

[†] Graduate School of Information Science and Technology, The University of Tokyo

E-mail: †{kameoka,onono,sagayama}@hil.t.u-tokyo.ac.jp

Abstract McAulay *et al.* used the sinusoidal model in a speech analysis and synthesis system and showed that it is able to produce high-quality synthetic speech [1]. This result suggests that the sinusoidal model expresses well the acoustic signals of speech. The performance of any application that uses the sinusoidal model depends critically on how accurate the model parameters can be estimated. This paper aims at presenting an efficient algorithm that optimizes the parameters of the sinusoidal model, inspired by the idea of the EM algorithm.

Key words sinusoidal signal model, EM algorithm, auxiliary function, Gabor transform

1. 序 論

正弦波重畳モデル [1] は、McAulay らが分析合成系に適用し、高品質な合成音声を生成できることを示して以来、分析合成系に留まらず、テキスト音声合成、音声加工、符号化などの用途として広く用いられてきた。特に、高品質な合成音声を生成できることは、正弦波重畳モデルが音声や楽音などの音響信号を極めて良く表現するモデルであることの根拠となっているため、音源分離への応用に対する期待も大きい。一方、音声の自己相関関数を少数の正弦波の重ね合わせで表現しようとする複合正弦波モデル [21] もまた良好な音声符号化方式として知られている。応用場面に依らずこの枠組 (正弦波重畳モデルを用いた信号分析) において共通するのは、いかにして正弦波重畳モデルのパラメータを精度良く推定するかが最も重要な問題となっている点であり、その推定精度があらゆる応用における性能に直結する。

McAulay らが採用した正弦波重畳モデルは、周波数と振幅

がいずれも定常であることを仮定した K 個の周波数成分を重畳した解析信号

$$s(t) \triangleq \sum_{k=1}^K A_k e^{j\mu_k t}, \quad t \in (-\infty, \infty) \quad (1)$$

である [1]。ただし、 μ_k 、 A_k はそれぞれ k 番目の正弦波成分の周波数、複素振幅を表す。短時間分析区間 $t \in [-T, T]$ における対象信号の解析信号を $\tilde{y}(t)$ とするとき、 $\tilde{y}(t)$ が

$$\tilde{y}(t) = s(t) + \epsilon(t), \quad t \in [-T, T] \quad (2)$$

と表され、 $\epsilon(t)$ が Gauss 性白色雑音 $\epsilon(t) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ であるという仮定のもとで $\Theta = \{\mu_k, a_k, \varphi_k\}_{1 \leq k \leq K}$ の最尤パラメータを求めることが具体的な問題設定である。この場合、 $\epsilon(t) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ であるため、 Θ の最尤解は、 L^2 ノルム誤差を最小にする解

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \int_{-\infty}^{\infty} \left\| w(t) (\tilde{y}(t) - s(t)) \right\|^2 dt \quad (3)$$

に相当する。ただし、ここでは $w(t)$ を $t \in [-T, T]$ において 1、それ以外では 0 をとる矩形窓としている。式 (1) が示すように、正弦波重畳モデルは、 A_k に関しては線形に依存するが、 μ_k に関しては非線形に依存するため、 μ_k が固定の場合には A_k の最尤解が解析的に得られることはただちに分かるが、 A_k が固定であったとしても μ_k の最尤解は解析的に得られない。この点が正弦波重畳モデルのパラメータ最適化における難しさの本質であり、そのため、統計的信号処理の分野では μ_k の最尤解を得るための方法がこれまで長年に渡り数多く研究されている [1] ~ [14]。

McAulay らの手法では $\Theta = \{\mu_k, a_k, \varphi_k\}_{1 \leq k \leq K}$ の推定値を得るのに、対象信号の離散パワースペクトル密度 (ピリオドグラム) が最大となるピーク成分の周波数と振幅と位相を決定し、そのピーク成分を信号から減算する操作を K 回繰り返すという簡便な方法が用いられた [1]。単一正弦波のピリオドグラムの最大値を与える周波数が最尤推定量であり、かつその推定量が不偏推定量であるとする Rife らの指摘 [17] ~ [19] が上記の周波数推定法が式 (3) の近似解であることの正当性に対する 1 つの根拠であると言えるが、(1) 離散的なピリオドグラムのピークが真の連続ピリオドグラムの最大値とは必ずしも対応しないこと、(2) 周波数成分が複数ある場合には周波数成分間の干渉のため上記理論が成立しなくなること、(3) 複数の周波数成分が近接する場合にはエネルギー拡散のため各々のピークが正しく検出できないことが起こり得ること、などを考慮するとこの単純な周波数推定法より高精度な推定法の開発が望まれることはごく自然な成り行きと言えるだろう。そうした中で、ピーク成分のパラメータを直接推定値として扱うのではなくピーク近傍の複数を結ぶ補間曲線から極大点を求めてパラメータ推定の高精度化を試みたもの [4] ~ [9] がその簡便さから近年特によく用いられているようだ。しかしながら、これらの手法は、上記の (2) と (3) の問題は依然解決されておらず、McAulay らの方法と同様、式 (3) の近似解を与えるに過ぎないため、その近似精度を向上するための零詰め量や窓関数の設計方法などが中心的な議論の対象である [8], [9]。一方で、式 (3) の解を最急降下法や Newton 法などの勾配法、あるいは Gibbs サンプラーやマルコフ連鎖モンテカルロ (MCMC) 法などの確率的サンプリング手法のような非線形最適化法により数値的に探索する方法も提案されている [3], [10] ~ [14]。

ところで、McAulay らのモデルが K 個の純音信号を混合したものであるのに対して、 K 個の調波信号 (N 個の調波成分で構成され、 n 次調波成分の周波数が基本周波数 μ_k の n 倍であるような信号) を重畳した解析信号

$$s(t) \triangleq \sum_{k=1}^K \sum_{n=1}^N A_{k,n} e^{jn\mu_k t}, \quad t \in (-\infty, \infty) \quad (4)$$

も同様に考えることができる。このモデルは、対象の混合信号が調波信号のみから構成されているような場合の 1ch 音源分離におけるモデルとしてしばしば用いられる [11] ~ [15]。式 (4) において $N = 1$ としたとき、式 (1) と同じモデルを想定したことになるため、McAulay らのモデルは式 (4) の特殊なケースと見ることができる。しかしこのモデルでは、各正弦波の周波数がそれぞれ自由度をもたず $n\mu_k$ のような拘束関係をもつため、

μ_k の最尤解を得るのは上記の McAulay らのモデルの場合よりさらに難しくなる。例えば、上記のうちの前者のアプローチのように中にはピーク抽出後の処理により μ_k を推定する手法も提案されているが、抽出されたピークがどの調波信号の何番目の調波成分に対応するかを決定するための場当たりの閾値判定に頼らざるを得なくなり、得られた μ_k の解の最尤性に関する議論が煩雑化する。そのため、このモデルを用いた従来の音源分離法では、後者のアプローチ (勾配法やサンプリング手法) がとられることが多い [11] ~ [13]。しかしながら、こういった数値計算法には常に局所最適解の問題が付きまとう。勾配法の場合には収束に至るまでの反復計算を無数通りの初期点から行わない限り、また、確率的サンプリング手法の場合には無数回の反復試行を行わない限り、式 (3) の大域的最適解を得られる保証はない。そのため、少ない計算量でいかに解探索を行えるかが問題となる (計算量が少なく済むということはその分だけ多くの異なる初期値条件から解探索が行える) が、勾配法やサンプリング法のようないわば力任せな数値計算しか適用されていないのが現状である。

以上のように、正弦波重畳モデルは音声や楽音を極めて良く表現するモデルであるにもかかわらず、そのパラメータ推定法に関しては議論する余地が残されていた。このような背景のもと、正弦波重畳モデルの最尤パラメータを求めるための新しい最適化アルゴリズムを導出することが本稿の目的である。

2. 正弦波重畳モデルと目的関数

まず、混合信号を次のような正弦波重畳モデル

$$s(t) = \sum_{k=1}^K \sum_{n=1}^N \tilde{A}_{k,n} e^{jn\mu_k t}, \quad t \in (-\infty, \infty) \quad (5)$$

でモデル化する。この中で未知パラメータは K 個の調波信号の基本周波数 μ_k 、 n 次調波成分の複素振幅 $A_{k,n}$ (すなわち、 $a_{k,n} = |A_{k,n}|$ は振幅、 $\varphi_{k,n} = \arg(A_{k,n})$ は初期位相を表す。) であり、これらをまとめて

$$\Theta = \{\mu_k, \{\tilde{A}_{k,n}\}_{1 \leq n \leq N}\}_{1 \leq k \leq K} \quad (6)$$

と表記する。ここで、考える問題は、観測混合信号 $y(t)$ の $t = 0$ における最尤の瞬時特徴量 Θ を求めることであり、これは

$$\Phi(\Theta) \triangleq \int_{-\infty}^{\infty} \left\| w(t) (y(t) - s(t)) \right\|^2 dt \quad (7)$$

を最小化する Θ に相当する。さて、窓関数 $w(t)$ は、先に用いた矩形窓の代わりに、時刻 $t = 0$ を中心とした Gauss 窓

$$w(t) \triangleq e^{-dt^2}, \quad t \in (-\infty, \infty) \quad (8)$$

とすると、 $\Phi(\Theta)$ は周波数領域では Parseval の等式により、

$$\Phi(\Theta) = \int_{-\infty}^{\infty} \left\| Y(\omega) - \sum_{k=1}^K \sum_{n=1}^N A_{k,n} e^{-\frac{(\omega - n\mu_k)^2}{4d}} \right\|^2 d\omega \quad (9)$$

と書ける。ただし、 $A_{k,n} = \frac{\tilde{A}_{k,n}}{\sqrt{2d}}$ であり、 d は窓の広がりを表す、分析条件によって決まる定数、 $Y(\omega)$ は $w(t)y(t)$ の Fourier 変換であり、すなわち、Gabor 変換による $t = 0$ における時間周波数分解成分を意味する。次章より、式 (9) を最小化するパラメータ最適化アルゴリズムを導出していく。

3. 最適パラメータ推定アルゴリズム

3.1 補助関数を用いた目的関数の非増加更新法

本章が提案するパラメータ最適化アルゴリズムは、EM アルゴリズムからヒントを得て導かれた、補助関数を用いた目的関数の非増加更新法の考え方を基礎とする。次節以降の定式化に向けた準備として、まず補助関数の定義を行ったのちに、補助関数に関する補題を示す。

定義 1. パラメータ $\Theta = (\Theta_1, \dots, \Theta_I)$ に関して最小化したい目的関数を $\Phi(\Theta)$ とすると、

$$\Phi(\Theta) \leq \Phi^+(\Theta, m) \quad (10)$$

を満たすとき、 $\Phi^+(\Theta, m)$ を目的関数 $\Phi(\Theta)$ の補助関数、 $m = (m_1, \dots, m_J)$ を補助変数と定義する。

補題 1. 目的関数を $\Phi(\Theta)$ 、 $\Phi(\Theta)$ の補助関数を $\Phi^+(\Theta, m)$ 、補助変数を $m = (m_1, \dots, m_J)$ とすると、補助関数の、補助変数に関する最小化と、パラメータ $\Theta_1, \dots, \Theta_I$ に関する最小化

$$\hat{m} = \operatorname{argmin}_m \Phi^+(\Theta, m) \quad (11)$$

$$\begin{aligned} \hat{\Theta}_1 &= \operatorname{argmin}_{\Theta_1} \Phi^+(\Theta_1, \dots, \Theta_I, m) \\ &\vdots \\ \hat{\Theta}_I &= \operatorname{argmin}_{\Theta_I} \Phi^+(\hat{\Theta}_1, \dots, \hat{\Theta}_{I-1}, \Theta_I, m) \end{aligned} \quad (12)$$

を行うと、目的関数 $\Phi(\Theta)$ は単調減少する。 $\Phi(\Theta)$ が下に有界ならば、上記の一連の更新の反復計算は収束性が保証される。

証明： 任意のパラメータを $\Theta^{(0)}$ とし、式 (11) と式 (12) により $\Phi(\Theta)$ が単調減少することを示す。まず式 (11) により、

$$\Phi(\Theta^{(0)}) = \Phi^+(\Theta^{(0)}, \hat{m}) \quad (13)$$

であり、式 (12) の更新により明らかに

$$\Phi^+(\Theta^{(0)}, \hat{m}) \geq \Phi^+(\hat{\Theta}, \hat{m}) \quad (14)$$

である。補助関数の定義 (10) より、

$$\Phi^+(\hat{\Theta}, \hat{m}) \geq \Phi(\hat{\Theta}) \quad (15)$$

なので、

$$\Phi(\Theta^{(0)}) = \Phi^+(\Theta^{(0)}, \hat{m}) \geq \Phi^+(\hat{\Theta}, \hat{m}) \geq \Phi(\hat{\Theta}) \quad (16)$$

である。以上より、各更新による目的関数の非増加が示された。□

この目的関数更新法をある最適化問題に応用するにあたり、不等式 (10) の等号が成立するときの補助変数が解析的に求められるかどうか、補助関数を最小化するモデルパラメータが解析的に求められるかどうかは鍵であり、そのようになるようにうまく補助関数を設計することが重要である。ところで、これまでの章で議論してきた EM アルゴリズムは、この反復最適化法の原理の 1 つの例に相当する。

3.2 Feder の補題 [20]

式 (9) の補助関数は、Feder らが示唆した次のような補題 [20] によって作ることができる。

補題 2. $t \in (\infty, \infty)$ においてある複素数値関数 $m_i(x)$ が

$$\sum_{i=1}^I m_i(x) = \sum_{i=1}^I m_i^*(x) = 1 \quad (17)$$

を満たすとき、

$$\int \left\| y(x) - \sum_{i=1}^I s_i(x) \right\|^2 dt \leq \sum_{i=1}^I \int \frac{\left\| m_i(x)y(x) - s_i(x) \right\|^2}{\beta_{k,n}} dt$$

が成り立ち、

$$m_i(x) = \frac{1}{y(x)} \left[s_i(x) + \beta_i \left(y(x) - \sum_{i=1}^I s_i(x) \right) \right] \quad (18)$$

のとき等号が成立する。ただし、 β_i は $\beta_i \in (0, 1)$ 、 $\sum_i \beta_i = 1$ を満たす任意の係数である。

証明： Lagrange 未定乗数法を用いた変分法により、不等式の右辺の $m_i(x)$ に関する最小値が左辺と等しくなることを示す。また、そのときの $m_i(x)$ を求める。そこでまず、不等式の右辺に条件 (17) のための Lagrange 未定乗数項を付加した汎関数

$$\begin{aligned} J[m] \triangleq & \sum_{i=1}^I \frac{1}{\beta_i} \int_{-\infty}^{\infty} \left\| m_i(x)y(x) - s_i(x) \right\|^2 dt \\ & - \int_{-\infty}^{\infty} \lambda(x) \left(\sum_{i=1}^I m_i^*(x) - 1 \right) dt \end{aligned} \quad (19)$$

の極値を求めることとする。 $J[m]$ の $m_i^*(x)$ に関する変分は、

$$\delta J[m] = \sum_{i=1}^I \int_{-\infty}^{\infty} \left(\frac{\partial J[m]}{\partial m_i^*} \right) \delta m_i^* dt \quad (20)$$

と書かれるが、これが恒等的に 0 であるためには、 $\frac{\partial J[m]}{\partial m_i^*} = 0$ である必要がある。よって、

$$\frac{\partial J[m]}{\partial m_i^*} = \frac{1}{\beta_i} y^*(x) \left(m_i(x)y(x) - s_i(x) \right) - \lambda(x) \quad (21)$$

を 0 とおくと、

$$m_i(x) = \frac{1}{\|y(x)\|^2} \left(\beta_i \lambda(x) + y^*(x) s_i(x) \right) \quad (22)$$

を得る。条件 (17) より、

$$\lambda(x) = \|y(x)\|^2 - y^*(x) \sum_{i=1}^I s_i(x) \quad (23)$$

である。これを式 (22) に代入すると、最終的に式 (18) を得る。これを上不等式の右辺に代入すると、たしかに左辺と等しくなる。次に、この極値が最小解であるかどうかは、 $J[m]$ の $m(x)$ に関する Hesse 行列 $\operatorname{diag}\left(\frac{\|y(x)\|^2}{\beta_1}, \dots, \frac{\|y(x)\|^2}{\beta_I}\right)$ が正定値であることから示すことができる。□

3.3 補助関数の設計

補題 2 より、 $\beta_{k,n} \in (0, 1)$, $\sum_{k,n} \beta_{k,n} = 1$ とし、 $S_{k,n}(\omega) \triangleq A_{k,n} e^{-\frac{(\omega-n\mu_k)^2}{4d}}$ と置くと式 (9) に関して次のような不等式

$$\begin{aligned} \Phi(\Theta) &= \int_{-\infty}^{\infty} \left\| Y(\omega) - \sum_{k=1}^K \sum_{n=1}^N S_{k,n}(\omega) \right\|^2 d\omega \\ &\leq \sum_{k=1}^K \sum_{n=1}^N \frac{1}{\beta_{k,n}} \int_{-\infty}^{\infty} \left\| m_{k,n}(\omega) Y(\omega) - S_{k,n}(\omega) \right\|^2 d\omega \quad (24) \end{aligned}$$

が成立し、等号は $m_{k,n}(\omega)Y(\omega)$ が

$$S_{k,n}(\omega) + \beta_{k,n} \left(Y(\omega) - \sum_{k=1}^K \sum_{n=1}^N S_{k,n}(\omega) \right) \quad (25)$$

のとき成立する。不等式 (24) の右辺を $\Phi^+(\Theta, m)$ と置く。定義 1 より、 $\Phi^+(\Theta, m)$ は目的関数 $\Phi(\Theta)$ の補助関数であり、 $m_{k,n}(\omega)$ は補助変数である。そして式 (25) は、補題 1 の式 (11)(補助変数の更新式) に相当する。

ところで、この不等式が意味するところは、混合信号と正弦波重畳モデルの複素スペクトル領域の L^2 ノルム誤差が、混合信号を分配フィルタ $m_{k,n}(\omega)$ によって任意に分解した各成分 $m_{k,n}(\omega)Y(\omega)$ と各正弦波成分との L^2 ノルム誤差の総和をとったものの最小値となっているということである。そして、式 (25) は、混合信号と正弦波重畳モデルとの間にある誤差を配分率 $\beta_{k,n}$ に応じて $\{k, n\}$ 番目の正弦波成分に付加したものを $m_{k,n}(\omega)Y(\omega)$ とすることが、 $Y(\omega)$ の最適な分解の仕方であることを表している。以後、表記の簡略化のため、 $Y_{k,n}(\omega) \triangleq m_{k,n}(\omega)Y(\omega)$ とする。

補題 1 より、次に、 $\Phi^+(\Theta, m)$ を Θ に関して最小化することを考える。 $\Phi^+(\Theta, m)$ は

$$\begin{aligned} \Phi^+(\Theta, m) &= \sum_{k=1}^K \sum_{n=1}^N \frac{1}{\beta_{k,n}} \int_{-\infty}^{\infty} \left(\|Y_{k,n}(\omega)\|^2 + \|S_{k,n}(\omega)\|^2 \right. \\ &\quad \left. - 2e^{-\frac{(\omega-n\mu_k)^2}{4d}} \operatorname{Re} \left[A_{k,n} Y_{k,n}^*(\omega) \right] \right) d\omega \quad (26) \end{aligned}$$

と書けるが、上式の第二項の積分は Gauss 積分により解析的に計算可能であることに気づく。すなわち、

$$\int_{-\infty}^{\infty} \|S_{k,n}(\omega)\|^2 d\omega = \sqrt{2\pi d} |A_{k,n}|^2 \quad (27)$$

であり、この項は μ_k に依らない。以上より、式 (26) は次のように書ける。

$$\begin{aligned} \Phi^+(\Theta, m) &= \sqrt{2\pi d} \sum_{k=1}^K \sum_{n=1}^N \frac{\|A_{k,n}\|^2}{\beta_{k,n}} \\ &\quad + \sum_{k=1}^K \sum_{n=1}^N \frac{1}{\beta_{k,n}} \int_{-\infty}^{\infty} \left(\|Y_{k,n}(\omega)\|^2 \right. \\ &\quad \left. - 2e^{-\frac{(\omega-n\mu_k)^2}{4d}} \operatorname{Re} \left[A_{k,n} Y_{k,n}^*(\omega) \right] \right) d\omega \quad (28) \end{aligned}$$

式 (28) の補助関数は依然として μ_k に関して非線形であり、このままではまだ μ_k の更新式に関しては解析的に求められな

い。そこで、 μ_k の更新式に関する議論の前に、 μ_k を固定した状態での振幅 $a_{k,n}$ と位相 $\varphi_{k,n}$ の更新式に関しては解析的に得られることをひとまず確認しておく。 $\frac{\partial \Phi^+(\Theta, m)}{\partial A_{k,n}^*} = 0$ を解くと、

$$A_{k,n} = \frac{1}{\sqrt{2\pi d}} \int_{-\infty}^{\infty} e^{-\frac{(\omega-n\mu_k)^2}{4d}} Y_{k,n}(\omega) d\omega \quad (29)$$

を得る。振幅と位相は、それぞれ上で求めた $A_{k,n}$ を用いて $a_{k,n} = |A_{k,n}|$, $\varphi_{k,n} = \arg(A_{k,n})$ と書ける。なお、この極値が最小解かどうかは 2 階偏導関数が正であることから明らかである。次節では、指数関数の性質に着目し、それに基づき、 μ_k の更新式を解析的に得るためのさらなる補助関数が作れることを示す。

3.4 凹関数の性質に基づく補助関数の設計

式 (28) において μ_k の更新式が解析的に得られない原因となっているのは、非線形項 $\exp(-\frac{(\omega-n\mu_k)^2}{4d})$ である。ここで、我々が着目したのは、 $-e^{-x}$ のような型の関数が微分可能な凹関数である点と、微分可能な凹関数に対して次のような定理が成り立つということである。

補題 3. 実関数 $f(x)$ が微分可能な凹関数のとき、任意の点 $\alpha \in \mathbb{R}$ において次の不等式が成り立つ。

$$f(x) \leq f(\alpha) + (x - \alpha)f'(\alpha) \quad (30)$$

証明：凹関数の定義より、任意の異なる 2 点 $x, \alpha \in \mathbb{R}$ 、任意の実数 $\gamma \in (0, 1)$ に対して、

$$f(\gamma x + (1 - \gamma)\alpha) \geq \gamma f(x) + (1 - \gamma)f(\alpha) \quad (31)$$

が成り立つとき、 $f(x)$ を凹関数という。上式は、

$$(x - \alpha) \frac{f(\alpha + \gamma(x - \alpha)) - f(\alpha)}{\gamma(x - \alpha)} \geq f(x) - f(\alpha) \quad (32)$$

と書き換えることができるが、 $\gamma \rightarrow 0$ のもとでは、 $f(x)$ は微分可能であるので、

$$\lim_{\gamma \rightarrow 0} \frac{f(\alpha + \gamma(x - \alpha)) - f(\alpha)}{\gamma(x - \alpha)} = f'(\alpha) \quad (33)$$

より、これを式 (32) に代入すると、

$$(x - \alpha)f'(\alpha) \geq f(x) - f(\alpha) \quad (34)$$

を得る。□

$-e^{-x}$ は微分可能な凹関数であるから、補題 3 より、任意の点 $\alpha \in \mathbb{R}$ において

$$-e^{-x} \leq -e^{-\alpha} + (x - \alpha)e^{-\alpha} \quad (35)$$

が成り立つ。従って、式 (35) の x を $\frac{(\omega-n\mu_k)^2}{4d}$ 、 α を実関数 $\alpha_{k,n}(\omega)$ と対応させると、 $\forall \omega \in \mathbb{R}$ において

$$\begin{aligned} &-e^{-\frac{(\omega-n\mu_k)^2}{4d}} \\ &\leq -e^{-\alpha_{k,n}(\omega)} + \left(\frac{(\omega-n\mu_k)^2}{4d} - \alpha_{k,n}(\omega) \right) e^{-\alpha_{k,n}(\omega)} \quad (36) \end{aligned}$$

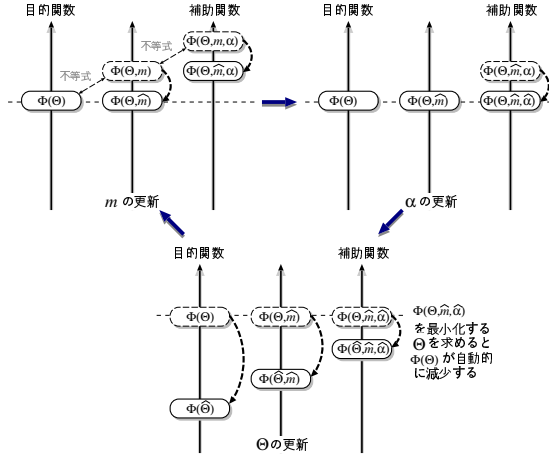


図1 補助関数を介したパラメータ最適化アルゴリズムの図解。

のような不等式が立てられる。式 (36) より、

$$\begin{aligned} \Phi^+(\Theta, m) &\leq \sqrt{2\pi d} \sum_{k=1}^K \sum_{n=1}^N \frac{\|A_{k,n}\|^2}{\beta_{k,n}} + \\ &\sum_{k=1}^K \sum_{n=1}^N \frac{1}{\beta_{k,n}} \int_{-\infty}^{\infty} \left[\|Y_{k,n}(\omega)\|^2 + 2\text{Re} \left[A_{k,n} Y_{k,n}^*(\omega) \right] \right. \\ &\left. \left\{ -e^{-\alpha_{k,n}(\omega)} + e^{-\alpha_{k,n}(\omega)} \left(\frac{(\omega - n\mu_k)^2}{4d} - \alpha_{k,n}(\omega) \right) \right\} \right] d\omega \end{aligned}$$

のように、式 (28) の上限を得る。上不等式右边を $\tilde{\Phi}^+(\Theta, m, \alpha)$ と置くと、

$$\Phi(\Theta) \leq \tilde{\Phi}^+(\Theta, m) \leq \tilde{\Phi}^+(\Theta, m, \alpha) \quad (37)$$

であるので、定義1より、これもまた $\Phi(\Theta)$ の補助関数として扱える。また、 $m_{k,n}(\omega)$ および $\alpha_{k,n}(\omega)$ が、この場合、いずれも補助変数となる。そして等号成立は、 m が式 (25) のときと

$$\alpha_{k,n}(\omega) = \frac{(\omega - n\mu_k)^2}{4d} \quad (38)$$

のときである。

以上の補助関数を導いたことの最大の意義は、 $\Phi(\Theta, m, \alpha)$ を最小化する μ_k を解析的に求めることができる点にある。ここでは実際にピッチ周波数 μ_k の更新式が解析的に得られることを確認しておく。 $\frac{\partial \Phi^+(\Theta, m, \alpha)}{\partial \mu_k} = 0$ を解くと、 $\sum_n \frac{n^2}{\beta_{k,n}} \int_{-\infty}^{\infty} e^{-\alpha_{k,n}(\omega)} \text{Re}[A_{k,n} Y_{k,n}^*(\omega)] d\omega \neq 0$ であるならば、

$$\mu_k = \frac{\sum_n \frac{n}{\beta_{k,n}} \int_{-\infty}^{\infty} e^{-\alpha_{k,n}(\omega)} \text{Re}[A_{k,n} Y_{k,n}^*(\omega)] \omega d\omega}{\sum_n \frac{n^2}{\beta_{k,n}} \int_{-\infty}^{\infty} e^{-\alpha_{k,n}(\omega)} \text{Re}[A_{k,n} Y_{k,n}^*(\omega)] d\omega} \quad (39)$$

が得られる。

3.5 最適パラメータ推定アルゴリズムのまとめ

以上の正弦波重畳モデルのパラメータ最適化アルゴリズムは次のようにまとめられる。また、この反復計算における各ステップでの目的関数値の変遷を図解したものを図1に示す。

Step 0 $\{\mu_k, \{A_{k,n}\}_{1 \leq n \leq N}\}_{1 \leq k \leq K}$ の初期設定。

Step 1 式 (25) による $m_{k,n}(\omega) Y(\omega)$ の更新。

Step 2 式 (38) による $e^{-\alpha_{k,n}(\omega)}$ の更新。

Step 3 式 (29) による $A_{k,n}$ の更新。

Step 4 式 (39) による μ_k の更新後、Step 1 に戻る。

4. 評価実験

4.1 シミュレーションによる収束性の確認

ここでの目的は、提案法と勾配法によるパラメータ推定の初期値依存性や収束速度を比べることである。ここで比較対象の勾配法を用いたパラメータ推定法 (以後、単に勾配法と呼ぶ。) は、Jinachitra の手法 [12] を参考にしたもので、式 (11) により $m_{k,n}(\omega)$ を更新するステップ、式 (29) により $A_{k,n}$ を更新するステップ、式 (28) を μ_k に関する最急降下更新により減少させるステップによって構成される。これと提案法を比べることにより、本稿で提案したピッチ周波数推定方式が局所解回避能力などの面で有効かどうかを示すことができる。

この比較実験では、パラメータの真値があらかじめ分かっている信号 (合成信号) を分析対象とした。具体的には、2つの周期信号 (ピッチ周波数が 270Hz、200Hz) は 10 個の調波成分からなり、各成分の振幅と位相は乱数生成により決め、これらを加算して混合信号を作成した。 n 次調波成分の振幅と位相の乱数値の範囲はそれぞれ $[\frac{1}{n}, \frac{3}{n}]$ および $[0, 2\pi)$ とした。また、正弦波重畳モデルにおいて $K = 2, N = 10$ とした。この合成信号 (16KHz サンプリング) から Gabor 変換により短時間複素スペクトル $Y(\omega)$ を得た。ただし、Gabor 変換における拡散パラメータ d は 0.067 とした。

さまざまな初期値条件のもとで、ピッチ周波数 μ_1, μ_2 が、提案法および勾配法によって各ステップで更新されていく様子をそれぞれ図2、3に示す (ピッチ周波数以外のパラメータの更新模様については省略する)。各図の上下のグラフにおいて、同色および同線種の折れ線は同一の反復計算による μ_1 と μ_2 の更新値の変遷を表す。なお、振幅 $A_{k,n}$ の初期値は 0 とした。

図2、3を比べてみると、勾配法では、 μ_1, μ_2 の初期値が真値 (270Hz、270Hz) にある程度近い値でないと、真値以外の停留点に陥ってしまうことが多いが、提案法では、広範囲の初期値から真値に素早く収束していく様子が確認できる。このシミュレーション実験の結果は、勾配法を用いる多数の従来法よりも提案法が局所解回避能力と収束速度の面で優れていることを示す1つの例証である。

4.2 混合音声の1ch ブラインド音源分離

次に、ここでは提案法の1ch ブラインド音源分離としての基本性能を確認する。対象とする混合信号は、ATR 研究用日本語音声データベースのセット B 中の音声データを用い、男性同士、女性同士、男性と女性の音声波形を加算したものとした。

すべての音声データは 16kHz にリサンプリングし、フレーム間隔 10ms の Gabor 変換により周波数解析した。前節同様、Gabor 変換における拡散パラメータ d は 0.067 とした。また、正弦波重畳モデルの各調波信号の調波成分数 N は 30 とした。

ここで採用したアルゴリズムの全容は次のとおりである。正弦波重畳モデルの初期調波信号数 K を 10 から開始し、パラメータの反復推定の過程で、複数の調波信号モデルのピッチ周波数パラメータが (1) ある一定値より近くなった場合、(2) ほ

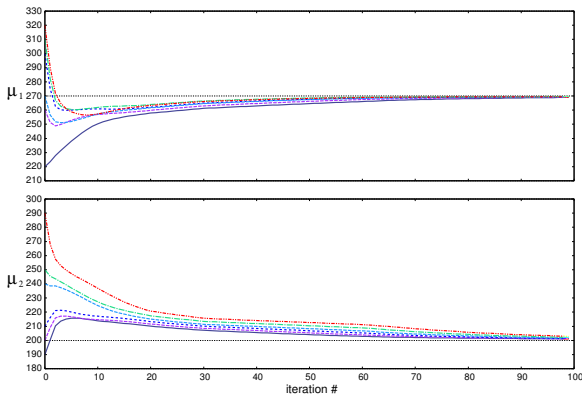


図2 提案法によるピッチ周波数パラメータ更新の様子

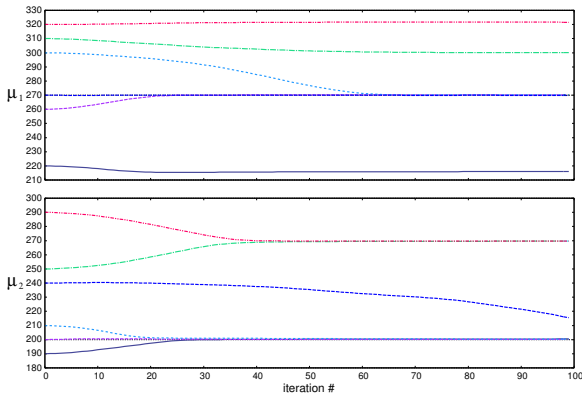


図3 最急降下法によるピッチ周波数パラメータ更新の様子

ば整数倍関係になった場合には低い方の調波信号だけを残し、残りは除外する。収束後、エネルギー全パワーの大きい2つの調波信号だけを残し、再びパラメータ推定を行う。このようにして、最終的に得られた2つの調波信号を分離推定信号とする。また、 μ_k の初期値は、まず観測混合信号の複素スペクトルの実部または虚部において極大/極小点を与える周波数をすべて見つけ、その中でパワーの大きいものから順に選んだ10個の周波数とした。上記は、各短時間窓(フレーム)で分離信号を推定する手続きであるが、各々のフレームで分離した信号がどの音源に対応したものであるかは本来は定かではない。今回の実験では、この対応づけができていない状況での仮の音源分離性能を調べるために、分離信号がどの音源に対応するかを混合する前の各信号との近さを調べて判定することにした。

以上の条件のもとで、図4に示す混合音声を実際に分離した結果の例を図5に示す。男性話者Aと女性話者Bの混合音声(男性話者Aから見てSN比が-0.3dB)の分離後の双方のSN比は、7.2dB, 6.4dB(7.5dB, 6.1dBの改善)、女性話者Aと女性話者Bの混合音声(女性話者Aから見てSN比が1.5dB)の分離後の双方のSN比は6.0dB, 4.8dB(4.5dB, 6.3dBの改善)、男性話者Aと男性話者Bの混合音声(男性話者Aから見てSN比が-0.3dB)の分離後の双方のSN比は4.8dB, 4.3dB(5.1dB, 4.0dBの改善)であった。提案法では、話者間のピッチ周波数の違いが、音源分離の手がかりとなっているため、同性同士の混合音声の分離精度が異性同士のそれよりいくらか下回ってしまうのは予想通りの結果と言える。

本稿で提案した手法はフレームごとに独立にパラメータ推定

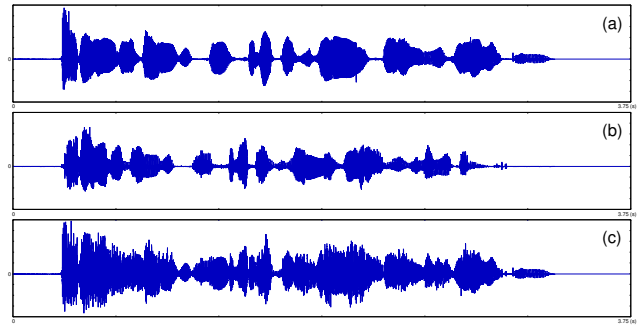


図4 女性話者音声(a)と男性話者音声(b)とその混合信号(c)。

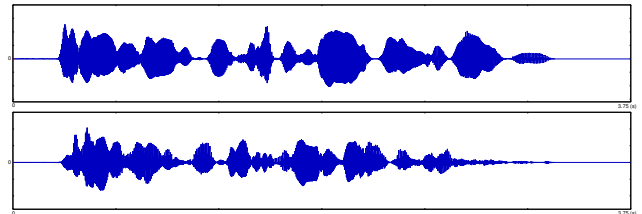


図5 女性および男性話者音声に対応する分離信号。

するため、分離信号の位相変化が不連続になったり、振幅が急激に変化することが多々あるため、今後、隣接する複数フレームに渡った協調的なパラメータ推定を行うことができれば、大幅なミュージカルノイズ軽減やSN比改善が見込まれる。

5. まとめ

本章では、正弦波重畳モデルのパラメータ推定問題の核である周波数推定ないし基本周波数推定の難しさの本質が、正弦波重畳モデルが周波数パラメータに関して非線形である点にあることに注目し、EMアルゴリズムをヒントにした補助関数を用いた新しい反復推定アルゴリズムを導いた。

文献

- [1] R. J. McAulay et al., *IEEE Trans. Acoust., Speech, Signal Process.*, 34(4), 744, 1986.
- [2] D. C. Rife et al., *Bell System Technical Journal*, 55(9), 1389, 1976.
- [3] T. J. Abatzoglou, *IEEE Trans. Acoust., Speech, Signal Process.*, 33(1), 77, 1985.
- [4] J. O. Smith III et al., In *Proc. ICMC'87*, 290, 1987.
- [5] Ph. Depalle et al., In *Proc. IEEE WASPAA'97*, 1997.
- [6] Y. V. Zakharov et al., *IEEE Electronic Letters*, 35(19), 1608, 1999.
- [7] S. S. Abeysekera, In *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, 417, 2001.
- [8] J. C. Brown et al., *J. Acoust. Soc. Am.*, 94(2), 662, 1993.
- [9] 安部ら, *信学技報*, EA2004-61, 7, 2004.
- [10] D. Starer et al., *IEEE Trans. Acoust. Speech, Signal Process.*, 40(6), 1528, 1992.
- [11] D. Chazan et al., In *Proc. ICASSP'93*, 2, 728, 1993.
- [12] P. Jinachitra, In *Proc. ICASSP'03*, 6, 609, 2003.
- [13] S. Godsill et al., In *Proc. ICASSP'02*, 2, 1769, 2002.
- [14] T. Tolonen, In *Proc. AES 106th Convention*, 1999.
- [15] T. Virtanen et al., In *Proc. ICASSP'00*, 2, 765, 2000.
- [16] D. J. Thomson, In *Proc. IEEE*, 70(9), 1055, 1982.
- [17] D. C. Rife et al., *IEEE Trans. Info. Theory*, 20(5), 591, 1974.
- [18] B. G. Quinn, *IEEE Trans. Signal Process.*, 42, 1264, 1994.
- [19] B. G. Quinn, *IEEE Trans. Signal Process.*, 45(3), 814, 1997.
- [20] M. Feder et al., *IEEE Trans., Acoust., Speech, Signal Process.*, 36(4), 477, 1988.
- [21] 嵯峨山ら, *通信学会誌*, J64-A(2), 105, 1981.