

正弦波重畳モデルのパラメータ最適化アルゴリズムと 1ch ブラインド音源分離への応用*

亀岡弘和, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

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、その逆問題として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が、従来、勾配法 [2] やサンプリング法 [3] 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。このような背景のもと、正弦波重畳モデルのパラメータを最尤推定するための、EM アルゴリズムのように安定的かつ効率的な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 周期信号重畳モデルと目的関数

まず、混合信号を次のようなそれぞれ N 個の調波成分からなる K 個の周期信号を重畳した解析信号表現

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

でモデル化する。この中で未知パラメータは各周期信号の基本周波数 μ_k 、第 n 調波成分の複素振幅 $A_{k,n}$ であり、これらをまとめて Θ と表記する。ここで、考える問題は、観測信号 $y(t)$ は区分的に周期信号の重ね合わせであることを仮定し、 $y(t)$ の $t=0$ における最尤の瞬時特徴量 Θ を求めることである。これは

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

を最小化する Θ に相当する。窓関数 $w(t)$ を Gauss 窓 $w(t) \triangleq e^{-dt^2}$ とすると、 $\Phi(\Theta)$ は周波数領域では、

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

と書ける。ただし、 $A_{k,n} = \frac{\tilde{A}_{k,n}}{\sqrt{2d}}$ であり、 d は窓の拡がりを表す、分析条件によって決まる定数、 $Y(\omega)$ は $w(t)y(t)$ の Fourier 変換である。

3 パラメータ最適化アルゴリズム

本章の最適化アルゴリズムは、EM アルゴリズムをヒントにした、補助関数を介したパラメータ反復更

新法の考え方を基礎とする。以降の準備として、補助関数の定義と、補助関数に関する補題を示す。

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

$$\Phi(\Theta) = \min_{\Omega} \Phi^+(\Theta, \Omega) \quad (3)$$

を満たすとき、 $\Phi^+(\Theta, \Omega)$ を目的関数 $\Phi(\Theta)$ の補助関数、 $\Omega = (\Omega_1, \dots, \Omega_L)$ 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 $\Phi^+(\Theta, \Omega)$ 、補助変数を $\Omega = (\Omega_1, \dots, \Omega_L)$ 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 $\Theta_1, \dots, \Theta_J$ に関する最小化

$$\hat{\Omega} = \operatorname{argmin}_{\Omega} \Phi^+(\Theta, \Omega) \quad (4)$$

$$\hat{\Theta}_j = \operatorname{argmin}_{\Theta_j} \Phi^+(\hat{\Theta}^{(j)}, \Omega) \quad (j = 1, \dots, J) \quad (5)$$

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 (証明略)。ただし、 $\hat{\Theta}^{(j)} = (\hat{\Theta}_1, \dots, \hat{\Theta}_{j-1}, \Theta_j, \dots, \Theta_J)$ とする。

このパラメータ更新法をある最適化問題に応用するにあたり、補助関数を最小化する補助変数およびモデルパラメータがいずれも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。ところで有名な 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。

導出は省略するが、 $S_{k,n}(\omega) \triangleq A_{k,n} e^{-\frac{(\omega - n\mu_k)^2}{4d}}$ と置くと、 $\Phi(\Theta)$ の上限を与える次のような不等式

$$\begin{aligned} \Phi(\Theta) &= \int_{-\infty}^{\infty} \left\| Y(\omega) - \sum_{k,n} S_{k,n}(\omega) \right\|^2 d\omega \leq \\ &I \sum_{k,n} \left[\sqrt{2\pi d} \|A_{k,n}\|^2 + \int_{-\infty}^{\infty} \left(m_{k,n}(\omega) \|Y(\omega)\|^2 \right. \right. \\ &\quad \left. \left. - 2e^{-\frac{(\omega - n\mu_k)^2}{4d}} \operatorname{Re} \left[A_{k,n} m_{k,n}^*(\omega) Y^*(\omega) \right] \right) d\omega \right] \quad (6) \end{aligned}$$

が成立し、等号は $m_{k,n}(\omega)$ が

$$\frac{1}{Y(\omega)} \left[S_{k,n}(\omega) + \frac{1}{I} \left(Y(\omega) - \sum_{k,n} S_{k,n}(\omega) \right) \right] \quad (7)$$

* A Parameter Optimization Algorithm for Sinusoidal Model and Its Application to 1ch Blind Source Separation. by KAMEOKA, Hirokazu, ONO, Nobutaka, SAGAYAMA, Shigeki (Graduate School of Information Science and Technology, The University of Tokyo)

のとき成立する。ただし、 $I = KN$ とする。式 (6) 右辺は補助関数の定義を満たすが、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。その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、 $-e^{-x}$ 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 $f(x)$ に対し、任意の点 $\alpha \in \mathbb{R}$ において $f(x) \leq f'(\alpha)(x - \alpha) + f(\alpha)$ が成り立つことに着目すると、実関数 $\alpha_{k,n}(\omega)$ を用い

$$\Phi(\Theta) \leq \sqrt{2\pi d} I \sum_{k,n} \|A_{k,n}\|^2 + I \sum_{k,n} \int_{-\infty}^{\infty} \left(m_{k,n}(\omega) \|Y(\omega)\|^2 + 2\text{Re} \left[A_{k,n} Y_{k,n}^*(\omega) \right] \left(e^{-\alpha_{k,n}(\omega)} \left(\frac{(\omega-n\mu_k)^2}{4d} - \alpha_{k,n}(\omega) \right) - e^{-\alpha_{k,n}(\omega)} \right) \right) d\omega$$

のような不等式が立てられる。定義 1 より、上式右辺 ($\Phi^+(\Theta, m, \alpha)$ とおく。) 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 (7) のときと

$$\alpha_{k,n}(\omega) = \frac{(\omega-n\mu_k)^2}{4d} \quad (8)$$

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り、 $\frac{\partial \Phi^+(\Theta, m, \alpha)}{\partial \mu_k} = 0$ を解くと、

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

が得られる。また、 $\frac{\partial \Phi^+(\Theta, m, \alpha)}{\partial A_{k,n}^*} = 0$ を解くと、複素振幅の更新式も同様に解析的に得られる(省略)。また、ここまでは定数と扱ってきた d も正弦波成分ごとに自由度をもつ変数 $d_{k,n}$ と見なせば、その更新式も解析的に得ることができる。このことの利点は稿を改めて述べることにする。

4 評価実験

シミュレーション実験を通し、勾配法に比べて提案法が初期値依存性や収束速度の面で極めて優れていることが確認されたが、ここでは紙面都合上割愛する。詳細は [4] を参照されたい。

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。また、 $K = 2, N = 30$ とした。

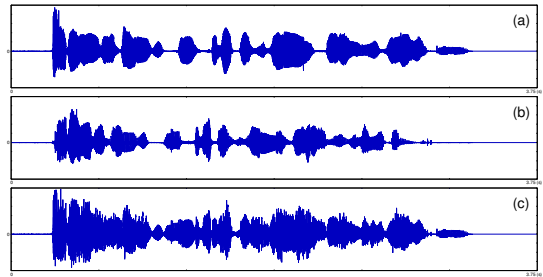


Fig. 1 女性音声 (a) と男性音声 (b) とその混合信号 (c)

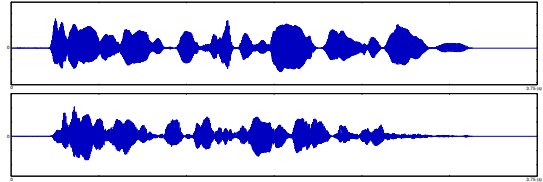


Fig. 2 女性および男性話者音声に対応する分離信号

提案法は、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 1 に示す混合音声を実際に分離した結果の例を図 2 に示す。混合音声の分離後の双方の SN 比の改善は、それぞれ男性 A と女性 B の場合は 7.5dB, 6.1dB、女性 A と女性 B の場合は 4.5dB, 6.3dB、男性 A と男性 B の場合は 5.1dB, 4.0dB であった。提案法では、話者間の基本周波数の違いが、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っくらか下回ってしまうのは予想通りの結果と言える。

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。

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。

参考文献

- [1] R. J. McAulay et al., IEEE Trans. Acoust., Speech, Signal Process., 34(4), pp. 744-754, 1986.
- [2] P. Jinachitra, In Proc. ICASSP'03, 6, 609, 2003.
- [3] S. Godsill et al., In Proc. ICASSP'02, 2, 1769, 2002.
- [4] 亀岡ら, 信学技報, 106(432), pp. 49-54, 2006.