

# 凸射影法に基づく Specmurt 分析の共通調波構造推定アルゴリズムとその収束性に関する考察\*

齊藤翔一郎, 亀岡弘和, 小野順貴, 嵯峨山茂樹 (東大情報理工)

## 1 はじめに

我々はこれまで、多重音の音楽音響信号の音高情報を可視化する手法である specmurt 分析 [1] と、その分析に用いる共通調波構造を反復推定するアルゴリズム [2] について提案してきた。このアルゴリズムは、基本周波数分布や調波構造に対する先見情報を用いて共通調波構造を準最適に求めるものであるが、その収束性や安定性に関しては必ずしも保証されていなかった。そこで我々は今回、凸射影法という枠組みを導入することにより、収束性が保証されるような共通調波構造を反復推定を試みた。本稿では、先見情報の凸集合としての表現方法と凸射影反復推定のアルゴリズムを論じ、その収束性と性能について報告する。

## 2 問題の定式化と従来法

### 2.1 Specmurt 分析

一般に楽音は、音高に相当する基本周波数以外に倍音を多く含んでおり、それが音色を構成している。この倍音と基本周波数の位置関係は、対数周波数領域では線形シフトの関係になっていることに注目すると、基本周波数分布  $u(x)$ 、対数周波数軸上の倍音位置に(のみ)倍音強度比の成分を持つ分布(以後、共通調波構造パターンと呼ぶ)を  $h(x)$  とすると、多重音のスペクトル  $v(x)$  を

$$v(x) = u(x) * h(x) \quad (1)$$

とモデル化することができる (Fig. 1,  $x$  は対数周波数)。これに基づき、基本周波数分布  $u(x)$  を多重音スペクトル  $v(x)$  に対する対数周波数での逆畳み込み

$$u(x) = \mathcal{F} \left[ \frac{\mathcal{F}^{-1}[v(x)]}{\mathcal{F}^{-1}[h(x)]} \right] \quad (2)$$

で求める手法を specmurt 分析 [1] と呼ぶ。

### 2.2 共通調波構造の自動推定問題

前節で述べた specmurt 分析では、ユーザーが事前に仮定する共通調波構造パターン  $h(x)$  の形によって結果が大きく異なり、適切なパターンをその都度調べる必要があった。これを解決する1つの方法は、観測スペクトル  $v(x)$  から共通調波構造パターン  $h(x)$  を自動的に推定することであるが、式 (1) のモデルにおいては基本周波数分布  $u(x)$  も未知であるから、これは不良設定問題であり、何らかの先見情報や制約条件が必要となる。

### 2.3 非線形写像を用いた反復推定法

この問題に対して仮定できる先見情報としては、例えば以下のようなものが考えられる。

- A.  $u(x)$  もスペクトルを表すから非負であること、
- B.  $u(x)$  は基本周波数分布であるから、ある大きさ以上の鋭いピークから構成されていること
- C. 共通調波構造  $h(x)$  は各倍音に相当する複数の  $\delta$  関数により、以下のように表わされていること、

$$h(x) = h(x; \mathbf{a}) = \sum_{n=1}^{\infty} a_n \delta(x - \log n) \quad (3)$$

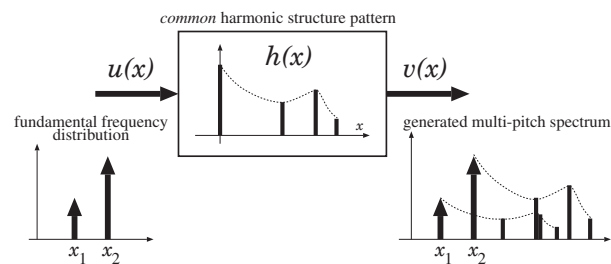
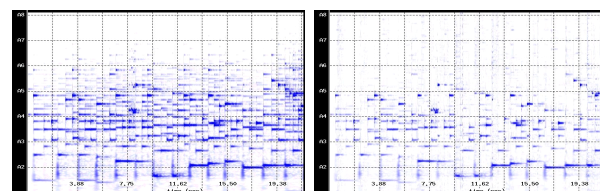


Fig. 1 基本周波数分布と共通調波構造の畳み込み



(a) スペクトログラム (b) specmurt 分析結果

Fig. 2 specmurt 分析による倍音抑圧

我々は一昨年、これらの条件をできるだけ満たすよう、以下の3つのステップを繰り返すことで、 $u_n(x)$ 、 $h_n(x)$  を更新していく反復推定法を提案した [2]。ただし  $n$  はステップ数である。

1.  $v(x)$  と  $h_n(x)$  に対し、式 (2) により  $u_n(x)$  を得る
2.  $u_n(x)$  に対し、非線形写像

$$\bar{u}_n(x) = \frac{u_n(x)}{1 + \exp\{-\alpha(u_n(x) - \beta)\}} \quad (4)$$

を用い、負値をとる部分を除き、不要成分と思われる部分を抑圧する (A, B の先見情報)

3. 式 (3) に基づき、 $\int |v(x) - \bar{u}_n(x) * h(x; \mathbf{a})|^2 dx$  を最小化するパラメータ  $\mathbf{a} = (a_1 \ a_2 \ \dots \ a_N)^T$  を求め、 $h_{n+1}(x) = h(x; \mathbf{a})$  を得る (C の先見情報)。

この反復推定によって出力結果の共通調波構造パターンの初期値に対する依存性が大幅に減少した。RWC 研究用音楽データベースのクラシック No.30 (ノクターン No.2, ショパン) について specmurt 分析を行った結果を Fig. 2 に示す。

ただしこの反復推定法においては、ステップ1の式 (2) が specmurt 領域 (対数周波数  $x$  の逆 Fourier 変換領域) での除算を含んでいるため、まれに推定途中の共通調波構造  $h_n(x)$  が specmurt 領域で零点をもつことがあり、収束しなかったり、解が不安定になる問題があった。

## 3 凸射影法による共通調波構造の推定

### 3.1 specmurt 分析における先見情報の集合表現

解が満たすべき複数の制約条件が全て凸集合であり、それらが共通集合をもつ場合には、ある初期値から各凸集合への射影 (その集合の要素で最も近い点への更新) を繰り返すことで、共通集合に含まれる1点への収束を保証することができる凸射影法 (Projection

\*POCS-based Common Harmonic Structure Estimation Algorithm for Specmurt Analysis and Discussions on Its Convergence. by SAITO, Shoichiro, KAMEOKA, Hirokazu, ONO, Nobutaka, SAGAYAMA, Shigeki (Graduate School of Information Science and Technology, The University of Tokyo)

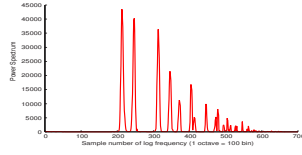
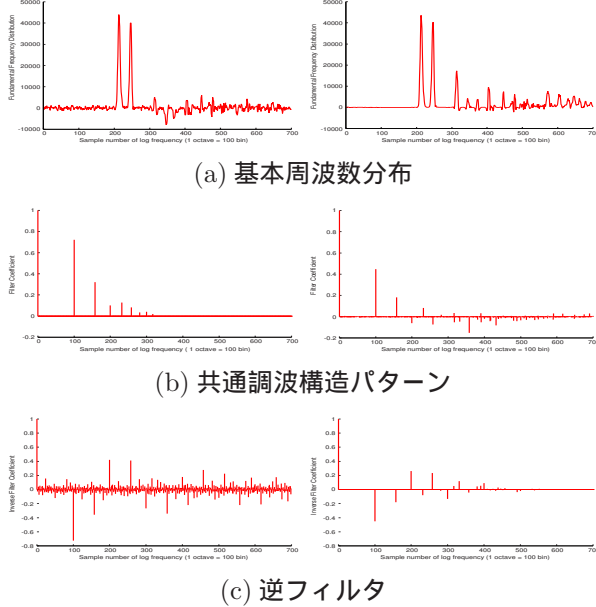


Fig. 3 多重音スペクトル (violin, C4&E4)



(I) 非線型写像による反復推定 (従来法) (II) 凸射影法による逆フィルタ推定 (提案法)

Fig. 4 specmurt 分析による倍音抑圧

Onto Convex Sets; POCS) という枠組みが知られており、画像復元や超解像分野で応用されている [3]. そこでここでは  $u(x)$  の空間を考え、要素間の距離を  $\int |u_n(x) - u_m(x)|^2 dx$  で定義し、凸射影法の specmurt 分析への適用可能性について検討する。

A の条件を満たす  $u(x)$  は明らかに凸集合  $S_A$  をなし、ある  $u(x)$  から  $S_A$  への射影  $\bar{u}(x)$  は簡単に、

$$\bar{u}(x) = \begin{cases} u(x) & (u(x) \geq 0) \\ 0 & (u(x) < 0) \end{cases} \quad (5)$$

という操作で与えられる。

B の条件を満たす  $u(x)$  の集合は、例えばある有意な正の値  $u_0$  を用いて  $S_B = \{u(x) | u(x) = 0 \text{ or } u(x) \geq u_0, \text{ for } \forall x\}$  と表わされるが、これは凸集合ではない。

C の条件を  $u(x)$  の集合としてとらえると、 $S_C = \{u(x) | u(x) = v(x) * h^{-1}(x), h(x) = \sum_{n=1}^{\infty} a_n \delta(x - \log n)\}$  となるが、逆畳み込みが含まれるこの形式のままでは、これが凸集合をなすかどうかは明らかではない。しかし、

$$\left| \sum_{n=2}^{\infty} a_n e^{-jy \log n} \right| < a_1 \quad (6)$$

を仮定すると、無限級数展開により  $h^{-1}(x)$  もまた

$$h^{-1}(x) = h^{-1}(x; \mathbf{b}) = \sum_{n=1}^{\infty} b_n \delta(x - \log n) \quad (7)$$

のように  $h(x)$  と同形の  $\delta$  関数列で表せ、 $S_C$  は凸集合となる。よって我々は調波構造に式 (6) を仮定し、 $u(x)$  から  $S_C$  への射影  $\bar{u}(x)$  を考える。これは、

$$\int |u(x) - v(x) * h^{-1}(x; \mathbf{b})|^2 dx \quad (8)$$

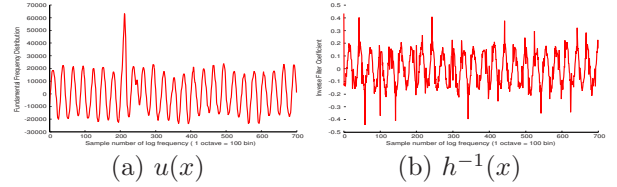


Fig. 5 従来法で推定結果が発散した例

を最小化するパラメータ  $\mathbf{b} = (b_1 \ b_2 \ \dots \ b_N)^T$  を求め、 $\bar{u}(x) = v(x) * h^{-1}(x; \mathbf{b})$  により得られる。これは従来法ステップ 2 での  $a$  の求め方と同様である ([2] 参照)。

以上より、 $A, C$  は凸集合として表現できることがわかった。これらは必ず共通集合 (少なくとも  $u(x) = v(x)$ ) をもつので、凸射影法の原理 [3] より、 $S_A, S_C$  への射影を繰り返すことで、収束性を保証するアルゴリズムを得ることができる。ただし、高い倍音抑圧性能を得るには適切な初期値を与える必要がある。

## 4 比較実験

提案手法について、共通調波構造推定と倍音抑圧効果の双方を従来手法と比較した。使用したのはパイオリンの 2 重奏 (C4 と E4) の実音響信号 (16kHz, モノラル) で、この信号のある解析フレームでのパワースペクトルを実験対象とした (Fig. 3)。このスペクトルに対し、これまで提案してきた非線形写像による反復推定を用いた場合と、今回提案する凸射影法を用いた場合のそれぞれについて、基本周波数分布  $u(x)$ 、共通調波構造パターン  $h(x)$ 、その逆フィルタ  $h^{-1}(x)$  を比較した (Fig. 4)。従来法においては sigmoid 関数のパラメータとして  $\alpha = 15, \beta = 0.5$  を使用し、提案法は共通調波構造に  $f^{-1.0}$  特性を仮定した場合の specmurt 分析結果を初期値として用いている。どちらも反復を 5 回行った。

倍音抑圧効果に関しては、従来法の方が精度よく 2 つの基本周波数を推定できている。ふたつのアルゴリズムで決定的に異なるのは非線形写像 (B の先見情報) を導入しているか否かであり、基本周波数分布において「比較的小さなピークは 0 へと近付ける」という効果が反復推定において大きな役割を果たしていることが予想できる。

一方、従来法は  $h(x)$  を推定するのが目的であるため、実際に観測スペクトルに畳み込まれるフィルタ  $h^{-1}(x)$  の係数は提案手法の方が安定した値が得られている。さらに、従来法ではフーリエ変換領域での零除算の問題から Fig. 5 のように推定結果が激しく振動することが稀に生じていたが、提案手法ではそういった現象は見受けられなかった。このことから、安定性に関しては提案手法の方が優れていることが示された。

## 5 まとめと展望

今回我々は凸射影法を用いて specmurt 分析の共通調波構造パターンの逆フィルタを推定する方法を提案した。比較実験により、従来法に比べ推定精度は低下するが、推定の安定性が向上するという結果を得た。提案手法では解の制約条件がゆるいために局所解へ収束していると考えられ、非線形写像に相当するような別の制約を導入する必要があると考える。

## 参考文献

- [1] 高橋他, 情報処理学会研究報告, 2003-MUS-53, pp.61-66, Dec. 2003.
- [2] 亀岡他, 音講論 (秋), 803-804, 2004.
- [3] D. C. Youla, IEEE Trans. on Medical Imaging, Vol.MI-1, No.2, 81-94, 1982.