

帯域幅の決定を含む平滑化の一手法

田中佑生子 (指導教員：吉田裕亮)

1 研究背景と目的

科学計測機器からの出力や種々の現象の観測から得られたデータには通常、測定・観測対象からの真の信号以外にも様々なノイズ成分が含まれている。演算処理によるそれらの除去法の一つとして平滑化 (smoothing) があり、代表的なものとして以下に示す移動平均法が挙げられる。

信号を N 個の離散値 $\{g(n)\} (n = 1, 2, \dots, N)$ で表したとき、これらの平滑値を次のように求める。

$$s(m) = \frac{1}{W} \sum_{i=-b}^b w(i)y(m+i)$$

$$\left(\begin{array}{l} m = b+1, b+2, \dots, N-b \\ W = \sum_{i=-b}^b w(i) \quad (w(i) : \text{任意関数}) \end{array} \right)$$

ここで、 $2b+1$ 個の離散点からなる任意関数 $w(i)$ は重み関数 (window)、また $2b+1$ は平滑化の帯域幅 (band width) と呼ばれる。両者は平滑化の結果に大きな影響を及ぼすが、実際の演算ではこれまでの研究で推奨されている既定値もしくは対象となるデータの発生機構や経験に基づき ad hoc に決定されていることが多い。

本研究では、理想的に平滑化されたデータの特徴量を与えられたデータから推定し、平滑化の帯域幅など移動平均モデルのパラメータを決定するひとつの手法を提案する。

2 提案する平滑化法

2.1 モデル

系列 $\{g(n)\} (n = 1, 2, \dots, N)$ の平滑化成分を次のように定める。

$$s(m) = \frac{1}{b} \sum_{i=1}^b R_{m-b+i, m-1+i}(b-i+1) \quad (1)$$

$$\left(\begin{array}{l} m = b, b+1, \dots, N-(b-1), \quad b \geq 4 \\ R_{k,l}(j) : g(k) \sim g(l) \text{ の回帰曲線系列の} \\ \quad \quad \quad j \text{ 番目 } (1 \leq j \leq l-k+1) \end{array} \right)$$

なお、 b 個のデータ点の多項式回帰では十分な次数を用意し、回帰毎に赤池情報量規準等を用いて適切な次数選択を行うものとする。

一つの平滑値が決まるのに必要とされるデータ数を平滑化帯域幅とすると帯域幅は $2b-1$ となるが、このモデルにおいては多項式回帰を施すデータ数、つまり window の意味に近い b を主体に議論を進めていく。

2.2 帯域幅決定の規準式

モデルから得られる平滑化済データには、与えられたデータに対する忠実性とある程度の滑らかさが求められる。ここで、前者を $g(n)$ と $s(m)$ 間の誤差分散 E 、後者を $s(m)$ の第 1 階差二乗和の平均 D で表し、また $s(m)$ の分散を V 、与えられたデータの最も理想的

な平滑化済データを‘真’とにおいて、平滑化の良さを測る指標として次の SG を導入する。

$$SG(b) = \log(b) + T(b) \cdot \log E(b) + \log D(b) \quad (2)$$

$$\left(\begin{array}{l} E(b) = \frac{1}{N-2(b-1)} \sum_{r=b}^{N-(b-1)} \{g(r) - s(r)\}^2 \\ D(b) = \frac{1}{N-2(b-1)-1} \sum_{r=b}^{N-(b-1)-1} \{s(r) - s(r+1)\}^2 \\ T(b) = \frac{V(b)}{E(b)} \cdot \left(\text{真の } \frac{D}{V} \right) \quad (\text{次章参照}) \end{array} \right)$$

このとき、与えられたデータの平滑化に最適な帯域幅 $2b-1$ は、 $SG(b)$ 値を最も小さくする b^* によって決定されるとする。

3 真の D/V の推定と帯域幅の決定

以後、 E, D, V が算出される対象となるデータを次のように表す。

E_*, D_*, V_* について

$$\left(\begin{array}{l} * = all \quad : \text{与えられたデータ } (N \text{ 個}) \\ * = all_S \quad : all \text{ の理想的な平滑化済データ '真' } (N \text{ 個}) \\ * = p(b) \quad : \text{与えられたデータのうち、演算によって} \\ \quad \quad \quad \text{平滑化済される部分 } (N-2(b-1) \text{ 個}) \\ * = p(b)_S \quad : p(b) \text{ の平滑化済データ } (N-2(b-1) \text{ 個}) \end{array} \right)$$

ここで、真の $\frac{D}{V} = \frac{D_{all_S}}{V_{all_S}}$ を推定するにあたり、次の式を導入する。

$$\alpha(x) \cdot \frac{D_p}{V_p}(b) = \frac{D_{p_S}}{V_{p_S}}(b) \left(\doteq \frac{D_{all_S}}{V_{all_S}} \right) \quad (3)$$

$$\left(\begin{array}{l} \alpha(x) = \frac{\frac{D_{all}}{V_{all}} \cdot \frac{x}{K}}{\frac{D_{all}}{V_{all}}} \left(\begin{array}{l} x = 1, 2, \dots, K-1 \\ K : \text{十分に大きい自然数} \end{array} \right) \\ \text{ただし、与えられたデータは} \\ \quad \quad \quad 0 < \frac{D_{all_S}}{V_{all_S}} < \frac{D_{all}}{V_{all}} \\ \text{を満たし、またここでは} \\ \quad \quad \quad \frac{D_p}{V_p}(b) \doteq \frac{D_{all}}{V_{all}}, \quad \frac{D_{p_S}}{V_{p_S}}(b) \doteq \frac{D_{all_S}}{V_{all_S}} \\ \text{であるとみなしている。} \end{array} \right)$$

$p \leq b \leq q$ ($p, q : 4 \leq p \leq q \leq N/2$ を満たす自然数)、 $1 \leq x \leq K$ において (3) の右辺と左辺の値の差が最も小さくなるような組み合わせ p^*, q^*, x^* を求め、

$$x^* \cdot \frac{D_{all}}{V_{all}} \left(= \frac{D_{all_S}}{V_{all_S}} \right) \quad (4)$$

より真の $\frac{D}{V}$ を推定する。

これを (2) に代入し、 $p^* \leq b \leq q^*$ の範囲で最適な平滑化帯域幅を決定すればよい。

4 数値実験

4.1 シミュレーション

適当な箇所にピークを2つ入れた正弦曲線に正規ノイズを加えたデータを平滑化した。ノイズを入れる前のデータを理想的な平滑化済データ‘真’とし、推定された真の $\frac{D}{V}$ と最適な平滑化帯域幅および平滑化済データを以下に示す。また、データ数は600である。

データ

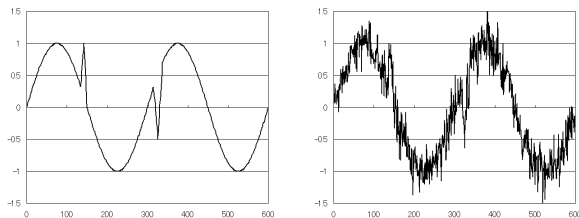


図 1: 真のデータ

図 2: ノイズ入りデータ

実験結果

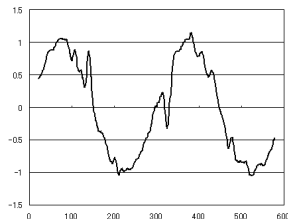


図 3: 平滑化済データ

真の $\frac{D}{V}$ は設定値 0.001669 に対して 0.001039, また最適な平滑化帯域幅は 22 と推定された。

4.2 考察

真の $\frac{D}{V}$ を設定値と大幅にずれることなく推定することができ、また推定された帯域幅を用いた平滑化により、元の形にある程度近いデータを得ることができた。またノイズ量を変えたいくつかのデータに対し同じように平滑化を行ったところ、真の $\frac{D}{V}$ の推定値はいずれも設定値に近い値となり、帯域幅は大きなノイズをもつデータほど大きな値が推定された。

5 実データへの応用

発光しているプラズマの可視光の発光強度 (intensity) の測定データの平滑化を行い、次のような結果を得た。データ数は1000である。

真の $\frac{D}{V}$ は 0.007999, 最適な平滑化帯域幅は 25 と推定された。

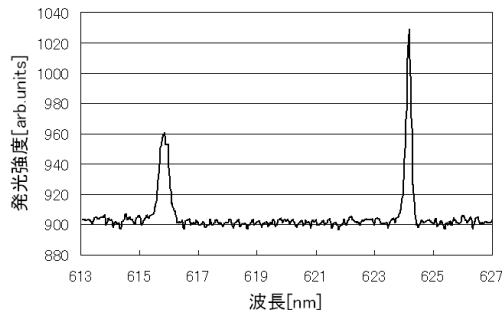


図 4: 元のデータ

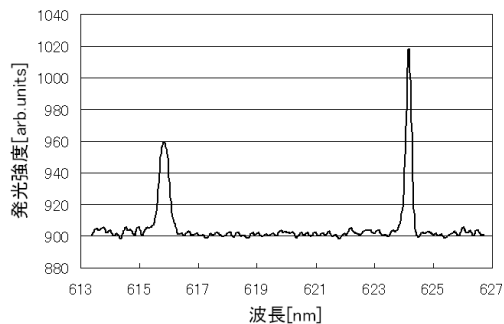


図 5: 平滑化済データ

6 まとめと今後の課題

モデル化に必要なパラメータを客観的な方法で求め、かつ単純な移動平均法では均らされやすいピークを大幅に消すことなくデータを平滑化することができた。しかし、このような部分的な特徴と全体的な形状の両方を十分に捉えるにはさらなる改良が必要であり、また計算時間の短縮も重要な課題である。

謝辞

本研究を進めるにあたり、実データの提供をいただきました株式会社デンソーの棚橋裕基氏に感謝申し上げます。

参考文献

- [1] 南茂夫, 科学計測のための波形データ処理, CQ出版社 (1986).
- [2] Kazutoshi T., Jiro H., “A Comparison Between Kawata-Minami and Savitzky-Golay Smoothing Methods with Raman Spectral Data”, Computer Enhanced Spectroscopy. vol.2, no.1, pp.17-20, Nov.1984.
- [3] Whittaker E. T., “On a New Method of Graduation”, Proc. Edinburgh Math. Soc., vol.41, pp.63-75, 1923.