逆畳み込み・畳み込み処理における誤差伝播

井田　隆

名古屋工業大学先進セラミックス研究センター
〒 507-0071 岐阜県多治見市旭ヶ丘 10-6-29

Error Propagation in Deconvolution-Convolution Treatment

Takashi Ida

Advanced Ceramics Research Center, Nagoya Institute of Technology,
10-6-29, Asahigaoka, Tajimi, Gifu 507-0071, JAPAN

This article describes theoretical and practical problems in the error propagation models used in the deconvolution-convolution treatment (DCT) on powder diffraction data, recently proposed by the author. It is shown that more experimental and analytical studies are still required to establish maximum likelihood estimation from the DCT powder diffraction data.

Keywords: error propagation, deconvolution, convolution, powder diffraction

1. はじめに

筆者は 2002 年に非線形の横軸尺度変換とスプライン補間、高速フーリエ変換による畳み込み計算を組み合わせることで、実験で得られる粉末回折強度図形から不要な副ピークや装置収差に由来するピーク位置のシフトおよびピーク形状の変形の効果等を自動的に修正する方法を提案した [1]。この方法の基本的な考え方は 2017 年に解説記事の中で紹介している [2]。また装置の光学収差の一つであり、多くの場合に最も目立つピーク形状の変形及びシフトの原因となる軸発散収差の処理に関して少し改良を加えて再提案を行った [3, 4]。

一方で、2010 年代以降、粉末X線回折測定での一次元検出器利用が急速に拡大し、従来型の検出器と比較して実質的な検出感度が 100 倍程度に向上したことから、従来では目立たなかった微小な副ピークや背景強度のわずかな段構造が目立つようになって来た。また、従来型検出器とスリットの組み合わせでは、集光条件を利用することにより、迷光や散乱光に由来する背景強度を低減させることが可能であったが、一次元検出器ではその方法を用いることができないので、背景強度が相対的に高くなる弱点もある。これらの問題に対応するように従来手法を拡張することは困難な問題ではなく、このことから十数年前に提案した方法が再評価されているようである。

ただし、この方法には、誤差伝播を厳密な形で取り入れることが技術的に困難であるという問題点がある。現時点での粉末回折強度データ解析では、それ以前に自明と思われる実測値の含む統計誤差が考慮に入れられていない状況である。そのような状況から、逆畳み込み・畳み込み処理における誤差伝播推定の不完全さがこの方法の決定的な欠陥となっている印象は受けていないが、問題になりうるとすれば、どのような場合に、どのような形で影響が現れるか、実験的・経験的な手法で検証されるべきであろう。

この解説記事では、理論的な視点と実践的な数値計算技術の両面からこの問題を明確化することを試みる。

2. 畳み込み

２－１　連続変数と離散変数の畳み込みの形式

「畳み込み」の考え方の基本的な部分については 2017 年の解説記事[2]で説明した。連続変数形式の表現では、関数 $f(x)$ と $g(x)$ の「畳み込み」が以下の式で定義される。

$$h(x) = (f * g)(x) = \int_{-\infty}^{\infty} f(x-y)\, g(y) \mathrm{d}y \tag{1}$$

しかし、現実のデータは有界で離散的であり、データ処理もそれを前提として実施される。なお畳み込みモデルに関するフーリエ解析を行う場合、事実上標本点は等間隔であることが前提となる。標本点の間隔を Δx として標本点の位置を

$$x_j = x_0 + j\Delta x \tag{2}$$

$$(j = 0, 1, \cdots, n-1) \tag{3}$$

とする。実験的に観測された離散的な強度データ $\{h_j\}$ を、何か「本来の（理想的な装置で得られるような）強度図形」$\{f_j\}$ と装置関数 $\{g_j\}$ との畳み込みとみなす場合、教科書的には

$$h_i = \sum_j f_{i-j}\, g_j \tag{4}$$

と表現されるだろう。しかし、式（4）のままでは和をとる範囲が明確になっていないので、実用的な定義としては不完全である。離散データ $\{f_j\}$ と $\{g_j\}$ の畳み込み $\{h_j\}$ は

$$h_i = \sum_{j=0}^{n-1} f_{i-j}\, g_j \tag{5}$$

として定義し、さらに $j < 0$ の時 f_j, g_j, h_j は、それぞれ f_{n+j}, g_{n+j}, h_{n+j} と「読み替える」ことにするのが実践的である。

２－２　フーリエ解析における余白領域の処理

前節のように現実のデータ処理を行う結果として、畳み込み前後の強度データに対応する $\{f_j\}$ と $\{h_j\}$ は、いずれも周期 $n\Delta x$ を持つことが前提となる。これは本当に強度データがそのような周期を持つかとは無関係であり、そのように仮定すれば計算が楽になるからという理由のみによると考えて良い。

ここで現実的な問題となるのは、強度データ $\{f_j\}$ と $\{h_j\}$ が左右の両端点でゼロに近い値を持つのでなければ、周期性を仮定して強制的に接続した時に、目に見えて矛盾が生じるということである。

そのために現実のフーリエ解析では左右の端点を接続する際に、一定の幅の「ゼロ」強度データを挿入する「ゼロ詰め」という処理を施す場合が多い。しかし、それでも「畳み込み」の場合には処理結果の両側に本質的には意味のない強度変化が導入されることになり、「逆畳み込み」処理の場合には端点付近に余分な強度振動が現れる。処理後のデータの両端について幾分かのデータを切り落とすことにすることも現実的な対応と思われるが、これは自動処理を目標とする場合に好ましくない。余白領域を設定し、その区間をなんらかの滑らかな曲線で繋げば良いとは想像できるが、現実に測定された強度データを対象とする場合、それはそれほど簡単な処理ではない。余白領域を区分的な一次関数で補間するのは現実的な選択だと思われる。例えば有効なデータが（j=0, 1, ⋯, n_1-1）の範囲だとして、余白領域（j=n_1, n_1+1, ⋯, n-1）は以下のような式で内挿することとした。

$$h_j = \begin{cases} h_{n_1-1} & [n_1 \le j < n_2] \\ \dfrac{h_0(j-n_2) + h_{n_1-1}(n_3-j)}{n_3-n_2} & [n_2 \le j < n_3] \\ h_0 & [n_3 \le j < n] \end{cases} \tag{6}$$

$$n_2 = \frac{2n_1 + n}{3} \tag{7}$$

$$n_3 = \frac{n_1 + 2n}{3} \tag{8}$$

２－３　離散強度データに関する逆畳み込み・畳み込み処理

離散強度 $\{h_j\}$ データのフーリエ変換 $\{H_k\}$ ：

$$H_k = \sum_{j=0}^{n-1} h_j \exp\frac{2\pi i k j}{n} \tag{9}$$

装置関数 $\{g_j\}$ のフーリエ変換 $\{G_k\}$ ：

$$G_k = \sum_{j=0}^{n-1} g_j \exp\frac{2\pi i k j}{n} \tag{10}$$

さらに「畳み込み関数」$\{g'_j\}$ のフーリエ変換 $\{G'_k\}$ ：

$$G'_k = \sum_{j=0}^{n-1} g'_j \exp\frac{2\pi i k j}{n} \tag{11}$$

に対して、逆畳み込み・畳み込み処理後のデータは、

$$f_j = \frac{1}{n} \sum_{k=-n/2}^{n/2-1} \frac{H_k G'_k}{G_k} \exp\left(-\frac{2\pi i k j}{n}\right) \tag{12}$$

で与えられる。

3. 誤差伝播

３－１　逆畳み込み・畳み込み処理に伴う誤差伝播

実測強度データ $\{h_j\}$ の誤差が $\{\Delta h_j\}$ で与えられるとする。逆畳み込み・畳み込みもフーリエ変換処理も線形変換に過ぎないので、処理後のデータ $\{f_j\}$ の含む誤差 $\{\Delta f_j\}$ は、本来ならば正確に予想できるはずである。ここで、便宜的に「装置関数」$\{w_j\}$ と「逆装置関数」$\{w_j^{(-1)}\}$ として、以下の式で定義される関数を導入する。

$$w_j = \frac{1}{n} \sum_{k=-n/2}^{n/2-1} \frac{G_k}{G'_k} \exp\left(-\frac{2\pi i k j}{n}\right) \tag{13}$$

$$w_j^{(-1)} = \frac{1}{n} \sum_{k=-n/2}^{n/2-1} \frac{G'_k}{G_k} \exp\left(-\frac{2\pi i k j}{n}\right) \tag{14}$$

以下の関係が成立する。

$$h_i = \sum_{j=0}^{n-1} f_{i-j}\, w_j \tag{15}$$

$$f_i = \sum_{j=0}^{n-1} h_{i-j}\, w_j^{(-1)} \tag{16}$$

変数 x の期待値を$\langle x \rangle$で表すことにすれば、逆畳み込み・畳み込み処理後のデータの分散･共分散行列（誤差行列）の行列要素 V_{ij} は

$$\begin{aligned} V_{ij} &= \langle (f_i - \langle f_i \rangle)(f_j - \langle f_j \rangle) \rangle \\ &= \sum_{l=0}^{n-1} \sum_{m=0}^{n-1} \langle (h_l - \langle h_l \rangle)(h_m - \langle h_m \rangle) \rangle \, w_{i-l}^{(-1)} w_{j-m}^{(-1)} \\ &= \sum_{l=0}^{n-1} \langle (h_l - \langle h_l \rangle)^2 \rangle \, w_{i-l}^{(-1)} w_{j-l}^{(-1)} \\ &= \sum_{l=0}^{n-1} (\Delta h_l)^2 \, w_{i-l}^{(-1)} w_{j-l}^{(-1)} \end{aligned} \tag{17}$$

という関係を満たす。

実測の粉末回折強度値が常に統計的に独立とみなせるとは限らないが、そのことについての議論はここでは保留する。現実には一般的な粉末回折強度データ解析では独立なものとして扱われる。しかしその扱い方が妥当とみなせる場合であっても、逆畳み込み・畳み込み処理後のデータについては独立性が失われることが明確である。つまり分散・共分散行列の非対角要素（交差項）がゼロでない値をとる。

分散・共分散行列(V_{ij})が既知の場合に、正しく最尤推定をするための方法は知られている。パラメトリックなモデル関数を$f(x)$とする。分散・共分散行列(V_{ij})の逆行列$(W_{ij})=(V_{ij})^{-1}$ を重み行列として定義して

$$S = \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} W_{ij} [f_i - f(x_i)][f_j - f(x_j)] \tag{18}$$

で与えられる重みつき残差二乗和を最小にすれば良い。しかし、粉末回折強度データは、典型的には数千点の強度値を含むデータであるから、式（18）の形式をそのまま用いるのは現実的ではない。重み行列(W_{ij})を対角化するための直交変換を求めておき、強度データとモデルにあらかじめこの変換を施すのが現実的な処理であるように見える。そして、重み行列を対角化するような直交変換は一意に確定するはずだから、その処理を施した後の強度データは、逆畳み込み・畳み込み処理前の実験データ $\{h_j\}$ に一致するはずである。

このような状況は粉末回折強度の波数軸でのフーリエ変換に相当する「原子対相関関数」を用いた解析にも共通することである。いずれも統計学的に正当に取り扱うためには変換前の値に戻して考え直すべきとも解釈できるから、「そもそもこのような変換をすることに意味があったのか？」という疑問が生じるのも自然であろう。別の言い方をすれば、逆畳み込み・畳み込み処理についても原子対相関関数解析にしても、何らかの意味があるとすれば、実験データの統計学的な解釈とは違うところにある。

ただし、単なるフーリエ変換である原子対相関関数と比較すれば、逆畳み込み・畳み込み処理によって得られる強度データの方はデータの統計学的な性質に極端に大きい変更を加えるわけではないので、やや単純化した誤差伝播モデルでも有効である可能性はあると思われる。

３－２　誤差伝播モデル

実測強度データ $\{h_j\}$ と処理後データ $\{f_j\}$ が極端に大きく変わらないのであれば、分散・共分散行列も重み行列も対角項が優勢であることは期待できる。そこで単純化した誤差伝播モデルとして、（I）分散・共分散行列の非対角項を無視するモデルと（II）重み行列の非対角項を無視するモデルについて検討する。

モデル（I）については、式（17）から

$$V_{ii} = \sum_{l=0}^{n-1} (\Delta h_l)^2 \left[w_{i-l}^{(-1)} \right]^2 \tag{19}$$

であることから、「処理後データの分散$\{V_{ii}\}$」が「処理前データの分散$\{(\Delta h_i)^2\}$」と「二乗逆装置関数$\{[w_i^{(-1)}]^2\}$」との畳み込みとして表現されることになる。

一方でモデル（II）については、

$$W_{ii} = \sum_{l=0}^{n-1} (\Delta h_l)^{-2} w_{l-i}^2 \tag{20}$$

となることから、「処理後のデータの逆分散$\{W_{ii}\}$」が「処理前データの逆分散$\{(\Delta h_i)^{-2}\}$」と「反転二乗装置関数$\{w_{-i}^2\}$」との相関として表現されることになる。このことの詳細は文献［1］に記載してある。

現在までの逆畳み込み・畳み込み法の予備的な運用の結果から、誤差伝播モデル（I）では誤差を過小に、モデル（II）では誤差を過大に評価する傾向が認められている。しかし、この傾向の現れ方は、逆畳み込み・畳み込み処理に用いる装置関数モデルにも依存するらしいことがわかってきた。筆者の開発する逆畳み込み・畳み込み処理ソフトウェアの高機能化とともに、状況がやや複雑化している面もある。この問題を解決するためには、さらに体系的な実験と解析が必要になると思われる。

４．おわりに

この記事では、筆者の考案した逆畳み込み・畳み込み処理について、2002 年に提案した当初から最も重要な課題とみなしていた誤差伝播の問題について述べた。

このデータ処理によって、どのように誤差が伝播するかは、理論的には特別に複雑な問題ではない。むしろ粉末回折データは規模が大きいので、正しく誤差伝播を考慮した解析をするためには、計算コストが問題になるという技術的な側面が強い。

誤差伝播モデルとしては当初から分散・共分散行列の非対角項を無視したモデルと重み行列の非対角項を無視したモデルを提案していたが、現時点ではどちらのモデ

ルにもやや不満がある。当面は、実験データに基づいて体系的な解析を実施し、さらに一般性の高い誤差伝播モデルを探索する必要性が高いと思われる。

参考文献

［1］T. Ida and H. Toraya, "Deconvolution of the instrumental functions in powder X-ray diffractometry," *J. Appl. Crystallogr.* 35, 58–68（2002）.

［2］井田隆 " 粉末回折データに対する逆畳み込み・畳み込み処理の考え方，" 名古屋工業大学先進セラミックス研究センター年報，5，37–43（2017）.

［3］T. Ida, S. Ono, D. Hattan, T. Yoshida, Y. Takatsu and K. Nomura, "Deconvolution-convolution treatment on powder diffraction data collected with Cu Kα X-ray and Ni Kβ filter," *Powder Diffr.*（submitted）.

［4］T. Ida, S. Ono, D. Hattan, T. Yoshida, Y. Takatsu and K. Nomura, "Improvement of deconvolution-convolution treatment of axial-divergence aberratioin in Bragg-Brentano geometry" *Powder Diffr.*（submitted）.