Gauss 関数による離散データの fitting

目次

1　導入

任意次元の離散データを Gauss 関数の和で fitting することを考える. N_{dim} 次元, N_{data} 点の離散データは, 表 1 のように一般的に表せる.

表 1　離散データの例

離散座標				データ
x_1^1	x_1^2	$\cdots$	$x_1^{N_{\mathrm{dim}}}$	f_1^{data}
x_2^1	x_2^2	$\cdots$	$x_2^{N_{\mathrm{dim}}}$	f_2^{data}
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
$x_{N_{\mathrm{data}}}^1$	$x_{N_{\mathrm{data}}}^2$	$\cdots$	$x_{N_{\mathrm{data}}}^{N_{\mathrm{dim}}}$	$f_{N_{\mathrm{data}}}^{\mathrm{data}}$

いま, N_{gauss} 個の Gauss 関数の和を次のように表す.

$$f(x^1, x^2, \cdots, x^{N_{\mathrm{dim}}}) = \sum_{j=1}^{N_{\mathrm{gauss}}} a(j) \exp\left[-\frac{1}{2}\sum_{l=1}^{N_{\mathrm{dim}}}\left(\frac{x^l - \mu(l,j)}{\sigma(l,j)}\right)^2\right] \tag{1.1}$$

ここで a, μ, σ はそれぞれ Gauss 関数の高さ, 中心, 広がりを決めるパラメータである. Gauss 関数の重ね合わせによる離散データの fitting とは, 残差を最小化するようなパラメータ a, μ, σ を求める最適化問題である. 最適化するパラメータの総数は $N_{\mathrm{gauss}} \times (1 + 2 \times N_{\mathrm{dim}})$ 個となる.

本稿で解説する手法は任意の次元数 N_{dim}, および任意の Gauss 関数の数 N_{gauss} をサポートする. 離散点が格子状に分布することを前提としないため, データの欠落や疎密も許容される. また以下の説明では, 離散座標 x の各成分や離散点上のデータ f^{data} のとりうる値の範囲がすべて数値的に同程度であることを要請する箇所がある. この要請は fitting 実行前に x, f^{data} を適当に平行移動・スケーリングすることで満足されるため, 実質的には離散座標やデータがとりうる値の範囲に対する要請も存在しない. なお, 実際の fitting では N_{gauss} の代わりに残差 L の threshold を指定する. はじめ $N_{\mathrm{gauss}} = 1$ として fitting を実行し, その後は最小化された L が threshold を満たすまで Gauss 関数を追加してゆく.

2　Newton-Raphson 法

今回は fitting 残差を最小化するために Newton-Raphson 法を用いる. 任意数の Gauss 関数の fitting に対応するため, 任意次元の Newton-Raphson 法の式を与える.

2.1　1 次元の場合

変数 q についての一階微分可能な関数 $f(q)$ に関して, 次の方程式の解を求めたいとする.

$$f(q) = 0 \tag{2.1}$$

表記を簡単にするために, $f(q)$ の任意階数の微分に関して次の記法を用いる.

$$\left(\frac{df}{dq}\right)^0 \equiv \left.\frac{df(q)}{dq}\right|_{q=q^0} \tag{2.2}$$

関数 $f(q)$ の $q=q^0$ における接線 $g(q)$ は,

$$g(q)=\left(\frac{df}{dq}\right)^0(q-q^0)+f(q^0) \tag{2.3}$$

この接線と q 軸との交点を求めるには, $g(q)=0$ を解けばよい. すなわち,

$$\left(\frac{df}{dq}\right)^0(q-q^0)+f(q^0)=0 \tag{2.4}$$

initial guess q^0 を適当に決めて方程式 (2.4) に代入し, 解 q を求める. この q を新しい q^0 として (2.4) に代入すれば, 新たな解 q が得られる. iteration によって q は真の解に向かい速やかに収束する.

2.2　N 次元の場合

変数の組 $\{q_i\}=q_1,q_2,\cdots,q_N$ についての一階微分可能な関数の組 $\{f_i\}=f_1,f_2,\cdots,f_N$ に関して, 次の方程式の解を求めたいとする.

$$\begin{cases} f_1(\{q_i\})=0 \\ \quad\vdots \\ f_N(\{q_i\})=0 \end{cases} \tag{2.5}$$

表記を簡単にするために, $f_i(q)$ の任意階数の微分に関して次の記法を用いる.

$$\left(\frac{\partial f_i}{\partial q_j}\right)^0 \equiv \left.\frac{\partial f_i(\{q_i\})}{\partial q_j}\right|_{\{q_i\}=\{q_i^0\}} \tag{2.6}$$

関数 $f_i(\{q_i\})$ の $q_1=q_1^0,q_2=q_2^0,\cdots,q_N=q_N^0$ における接線 $g_i(\{q_i\})$ は,

$$g_i(\{q_i\})=\sum_{j=1}^{N}\left(\frac{\partial f_i}{\partial q_j}\right)^0(q_j-q_j^0)+f_i(\{q_i^0\}) \tag{2.7}$$

この接線と $\{q_i\}=0$ との交点を求めるには, $g_1(\{q_i\})=0,g_2(\{q_i\})=0,\cdots,g_N(\{q_i\})=0$ を解けばよい. すなわち,

$$\begin{cases} \displaystyle\sum_{j=1}^{N}\left(\frac{\partial f_1}{\partial q_j}\right)^0(q_j-q_j^0)+f_1(\{q_i^0\})=0 \\ \qquad\vdots \\ \displaystyle\sum_{j=1}^{N}\left(\frac{\partial f_N}{\partial q_j}\right)^0(q_j-q_j^0)+f_N(\{q_i^0\})=0 \end{cases} \tag{2.8}$$

initial guess $\{q_i^0\}$ を適当に決めて方程式 (2.8) に代入し, 解 $\{q_i\}$ を求める. この $\{q_i\}$ を新しい $\{q_i^0\}$ として (2.8) に代入すれば, 新たな解 $\{q_i\}$ が得られる. iteration によって $\{q_i\}$ は真の解に向かい速やかに収束する.

連立方程式 (2.8) を変形して,

$$\begin{pmatrix} \left(\frac{\partial f_1}{\partial q_1}\right)^0 & \cdots & \left(\frac{\partial f_1}{\partial q_N}\right)^0 \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial f_N}{\partial q_1}\right)^0 & \cdots & \left(\frac{\partial f_N}{\partial q_N}\right)^0 \end{pmatrix}\begin{pmatrix} q_1-q_1^0 \\ \vdots \\ q_N-q_N^0 \end{pmatrix}+\begin{pmatrix} f_1(\{q_i^0\}) \\ \vdots \\ f_N(\{q_i^0\}) \end{pmatrix}=\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2.9}$$

さらに, 次のように二階テンソル $\boldsymbol{A}$ および一階テンソル $\boldsymbol{b}, \boldsymbol{q}, \boldsymbol{q}^0$ を定義する.

$$\boldsymbol{A} = \begin{pmatrix} \left(\frac{\partial f_1}{\partial q_1}\right)^0 & \cdots & \left(\frac{\partial f_1}{\partial q_N}\right)^0 \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial f_N}{\partial q_1}\right)^0 & \cdots & \left(\frac{\partial f_N}{\partial q_N}\right)^0 \end{pmatrix}, \quad \boldsymbol{b} = -\begin{pmatrix} f_1(\{q_i^0\}) \\ \vdots \\ f_N(\{q_i^0\}) \end{pmatrix}, \quad \boldsymbol{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \quad \boldsymbol{q}^0 = \begin{pmatrix} q_1^0 \\ \vdots \\ q_N^0 \end{pmatrix} \tag{2.10}$$

これらを用いて (2.9) を表すと,

$$\boldsymbol{A} \cdot (\boldsymbol{q} - \boldsymbol{q}^0) = \boldsymbol{b} \tag{2.11}$$

3 基本の連立方程式

3.1 Newton-Raphson 法による残差の最小化

Gauss 関数の重ね合わせ (1.1) による離散データの fitting では, 残差の二乗 L を最小化するような Gauss パラメータの組を求めることが目的となる.

$$L = \frac{1}{\sum_{i=1}^{N_{\mathrm{data}}} w_i} \sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ f(x_i^1, \cdots, x_i^{N_{\mathrm{dim}}}) - f_i^{\mathrm{data}} \right\}^2 \tag{3.1}$$

ただし w_i は残差に対する各離散点の寄与を表す重み因子で, 通常は 1, 特に精度よく fitting したい離散点に対してはより大きな値を設定する. fitting されるパラメータは $\{a(j), \mu(l,j), \sigma(l,j)\}$ $(j = 1, 2, \cdots, N_{\mathrm{gauss}}; l = 1, 2, \cdots, N_{\mathrm{dim}})$ であるが, これをまとめて $\left\{q_j\right\}$ $(j = 1, 2, \cdots, N)$ と表す. ただし,

$$N = (1 + 2 \times N_{\mathrm{dim}}) \times N_{\mathrm{gauss}} \tag{3.2}$$

今回は L をパラメータ $\left\{q_j\right\}$ の関数とみなして次の連立方程式を解く.

$$\begin{cases} \dfrac{\partial L(q_1, q_2, \cdots, q_N)}{\partial q_1} = 0 \\ \quad\vdots \\ \dfrac{\partial L(q_1, q_2, \cdots, q_N)}{\partial q_N} = 0 \end{cases} \tag{3.3}$$

これを Newton-Raphson 法の式に当てはめると, 次の連立方程式を解けば良いことが分かる.

$$\begin{cases} \displaystyle\sum_{j=1}^{N} \left(\frac{\partial^2 L}{\partial q_1 \partial q_j}\right)^0 (q_j - q_j^0) + \left(\frac{\partial L}{\partial q_1}\right)^0 = 0 \\ \quad\vdots \\ \displaystyle\sum_{j=1}^{N} \left(\frac{\partial^2 L}{\partial q_N \partial q_j}\right)^0 (q_j - q_j^0) + \left(\frac{\partial L}{\partial q_N}\right)^0 = 0 \end{cases} \tag{3.4}$$

あるいは,

$$\begin{pmatrix} \left(\frac{\partial^2 L}{\partial q_1 \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_1 \partial q_N}\right)^0 \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial^2 L}{\partial q_N \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_N \partial q_N}\right)^0 \end{pmatrix} \begin{pmatrix} q_1 - q_1^0 \\ \vdots \\ q_N - q_N^0 \end{pmatrix} + \begin{pmatrix} \left(\frac{\partial L}{\partial q_1}\right)^0 \\ \vdots \\ \left(\frac{\partial L}{\partial q_N}\right)^0 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \tag{3.5}$$

ここで, 以下のテンソル $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{q}, \boldsymbol{q}^0$ を用いる.

$$\boldsymbol{A} = \begin{pmatrix} \left(\frac{\partial^2 L}{\partial q_1 \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_1 \partial q_N}\right)^0 \\ \vdots & \ddots & \vdots \\ \left(\frac{\partial^2 L}{\partial q_N \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_N \partial q_N}\right)^0 \end{pmatrix}, \quad \boldsymbol{b} = -\begin{pmatrix} \left(\frac{\partial L}{\partial q_1}\right)^0 \\ \vdots \\ \left(\frac{\partial L}{\partial q_N}\right)^0 \end{pmatrix}, \quad \boldsymbol{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \quad \boldsymbol{q}^0 = \begin{pmatrix} q_1^0 \\ \vdots \\ q_N^0 \end{pmatrix} \tag{3.6}$$

Hessian $\boldsymbol{A}$ は実対称行列である. このとき連立方程式は次のように書ける.

$$\boldsymbol{A} \cdot (\boldsymbol{q} - \boldsymbol{q}^0) = \boldsymbol{b} \tag{3.7}$$

3.2 残差の微分

まず準備として, 以下の記法を定める.

$$\exp[i, j] \equiv \exp\left[-\frac{1}{2} \sum_{l=1}^{N_{\text{dim}}} \left(\frac{x_i^l - \mu(l, j)}{\sigma(l, j)}\right)^2\right] \tag{3.8}$$

$$f_i \equiv f(x_i^1, \cdots, x_i^{N_{\text{dim}}}) = \sum_{j=1}^{N_{\text{gauss}}} a(j) \exp[i, j] \tag{3.9}$$

$$W \equiv \sum_{i=1}^{N_{\text{data}}} w_i \tag{3.10}$$

このとき,

$$L = \frac{1}{W} \sum_{i=1}^{N_{\text{data}}} w_i \left\{f_i - f_i^{\text{data}}\right\}^2 \tag{3.11}$$

また,

$$\frac{\partial \exp[i, p]}{\partial a(p)} = 0 \tag{3.12}$$

$$\begin{aligned} \frac{\partial \exp[i, p]}{\partial \mu(r, p)} &= \exp[i, p] \cdot \left(-\frac{1}{2}\right) \cdot 2 \left(\frac{x_i^r - \mu(r, p)}{\sigma(r, p)}\right) \cdot \left\{-\frac{1}{\sigma(r, p)}\right\} \\ &= \exp[i, p] \frac{x_i^r - \mu(r, p)}{\sigma(r, p)^2} \end{aligned} \tag{3.13}$$

$$\frac{\partial \exp[i, p]}{\partial \mu(r, q)} = 0 \tag{3.14}$$

$$\begin{aligned} \frac{\partial \exp[i, p]}{\partial \sigma(r, p)} &= \exp[i, p] \cdot \left(-\frac{1}{2}\right) \cdot 2 \left(\frac{x_i^r - \mu(r, p)}{\sigma(r, p)}\right) \cdot \left\{-\frac{x_i^r - \mu(r, p)}{\sigma(r, p)^2}\right\} \\ &= \exp[i, p] \frac{\left\{x_i^r - \mu(r, p)\right\}^2}{\sigma(r, p)^3} \end{aligned} \tag{3.15}$$

$$\frac{\partial \exp[i, p]}{\partial \sigma(r, q)} = 0 \tag{3.16}$$

すなわち,

$$\frac{\partial f_i}{\partial a(p)} = \exp[i,p] \tag{3.17}$$

$$\begin{aligned}\frac{\partial f_i}{\partial \mu(r,p)} &= a(p)\frac{\partial \exp[i,p]}{\partial \mu(r,p)}\\ &= a(p)\exp[i,p]\,\frac{x_i^r - \mu(r,p)}{\sigma(r,p)^2}\end{aligned} \tag{3.18}$$

$$\begin{aligned}\frac{\partial f_i}{\partial \sigma(r,p)} &= a(p)\frac{\partial \exp[i,p]}{\partial \sigma(r,p)}\\ &= a(p)\exp[i,p]\,\frac{\left\{x_i^r - \mu(r,p)\right\}^2}{\sigma(r,p)^3}\end{aligned} \tag{3.19}$$

従って, 残差 L の微分は以下のようになる. ただし $\varDelta_i = f_i - f_i^{\mathrm{data}}$ とする.

$$\begin{aligned}\frac{\partial L}{\partial a(p)} &= \frac{1}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \cdot 2\varDelta_i \frac{\partial f_i}{\partial a(p)}\\ &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i\varDelta_i \exp[i,p]\end{aligned} \tag{3.20}$$

$$\begin{aligned}\frac{\partial L}{\partial \mu(r,p)} &= \frac{1}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \cdot 2\varDelta_i \frac{\partial f_i}{\partial \mu(r,p)}\\ &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i\varDelta_i a(p)\exp[i,p]\,\frac{x_i^r - \mu(r,p)}{\sigma(r,p)^2}\end{aligned} \tag{3.21}$$

$$\begin{aligned}\frac{\partial L}{\partial \sigma(r,p)} &= \frac{1}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \cdot 2\varDelta_i \frac{\partial f_i}{\partial \sigma(r,p)}\\ &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i\varDelta_i a(p)\exp[i,p]\,\frac{\left\{x_i^r - \mu(r,p)\right\}^2}{\sigma(r,p)^3}\end{aligned} \tag{3.22}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial a(p)\partial a(q)} &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \frac{\partial f_i}{\partial a(q)}\exp[i,p]\\ &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \exp[i,p]\exp[i,q]\end{aligned} \tag{3.23}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial a(p)\partial \mu(s,q)} &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{\frac{\partial f_i}{\partial \mu(s,q)}\exp[i,p] + \delta_{pq}\varDelta_i \frac{\partial \exp[i,p]}{\partial \mu(s,p)}\right\}\\ &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{a(q)\exp[i,p]\exp[i,q]\,\frac{x_i^s - \mu(s,q)}{\sigma(s,q)^2} + \delta_{pq}\varDelta_i\exp[i,p]\,\frac{x_i^s - \mu(s,p)}{\sigma(s,p)^2}\right\}\end{aligned} \tag{3.24}$$

$$\begin{aligned}\frac{\partial^2 L}{\partial a(p)\partial \sigma(s,q)} &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{\frac{\partial f_i}{\partial \sigma(s,q)}\exp[i,p] + \delta_{pq}\varDelta_i \frac{\partial \exp[i,p]}{\partial \sigma(s,p)}\right\}\\ &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{a(p)\exp[i,p]\exp[i,q]\,\frac{\left\{x_i^s - \mu(s,q)\right\}^2}{\sigma(s,q)^3} + \delta_{pq}\varDelta_i\exp[i,p]\,\frac{\left\{x_i^s - \mu(s,p)\right\}^2}{\sigma(s,p)^3}\right\}\end{aligned} \tag{3.25}$$

$$
\begin{aligned}
\frac{\partial^2 L}{\partial\mu(r,p)\partial\mu(s,q)} &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ \frac{\partial f_i}{\partial\mu(s,q)} a(p)\exp[i,p]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2} \right. \\
&\left. +\delta_{pq}\mathit{\Delta}_i a(p)\frac{\partial \exp[i,p]}{\partial\mu(s,p)}\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2} + \delta_{pq}\delta_{rs}\mathit{\Delta}_i a(p)\exp[i,p]\left(-\frac{1}{\sigma(r,p)^2}\right)\right\} \\
&= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ a(p)a(q)\exp[i,p]\exp[i,q]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2}\frac{x_i^s-\mu(s,q)}{\sigma(s,q)^2} \right. \\
&\left. +\delta_{pq}\mathit{\Delta}_i a(p)\exp[i,p]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2}\frac{x_i^s-\mu(s,p)}{\sigma(s,p)^2} - \delta_{pq}\delta_{rs}\mathit{\Delta}_i a(p)\exp[i,p]\,\frac{1}{\sigma(r,p)^2}\right\}
\end{aligned} \tag{3.26}
$$

$$
\begin{aligned}
\frac{\partial^2 L}{\partial\mu(r,p)\partial\sigma(s,q)} &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ \frac{\partial f_i}{\partial\sigma(s,q)} a(p)\exp[i,p]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2} \right. \\
&\left. +\delta_{pq}\mathit{\Delta}_i a(p)\frac{\partial \exp[i,p]}{\partial\sigma(s,p)}\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2} + \delta_{pq}\delta_{rs}\mathit{\Delta}_i a(p)\exp[i,p]\left(-2\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^3}\right)\right\} \\
&= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ a(p)a(q)\exp[i,p]\exp[i,q]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2}\frac{\left\{x_i^s-\mu(s,q)\right\}^2}{\sigma(s,q)^3} \right. \\
&\left. +\delta_{pq}\mathit{\Delta}_i a(p)\exp[i,p]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^2}\frac{\left\{x_i^s-\mu(s,p)\right\}^2}{\sigma(s,p)^3} - 2\delta_{pq}\delta_{rs}\mathit{\Delta}_i a(p)\exp[i,p]\,\frac{x_i^r-\mu(r,p)}{\sigma(r,p)^3}\right\}
\end{aligned} \tag{3.27}
$$

$$
\begin{aligned}
\frac{\partial^2 L}{\partial\sigma(r,p)\partial\sigma(s,q)} &= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ \frac{\partial f_i}{\partial\sigma(s,q)} a(p)\exp[i,p]\,\frac{\left\{x_i^r-\mu(r,p)\right\}^2}{\sigma(r,p)^3} \right. \\
&\left. +\delta_{pq}\mathit{\Delta}_i a(p)\frac{\partial \exp[i,p]}{\partial\sigma(s,p)}\frac{\left\{x_i^r-\mu(r,p)\right\}^2}{\sigma(r,p)^3} + \delta_{pq}\delta_{rs}\mathit{\Delta}_i a(p)\exp[i,p]\left(-3\frac{\left\{x_i^r-\mu(r,p)\right\}^2}{\sigma(r,p)^4}\right)\right\} \\
&= \frac{2}{W}\sum_{i=1}^{N_{\mathrm{data}}} w_i \left\{ a(p)a(q)\exp[i,p]\exp[i,q]\,\frac{\left\{x_i^r-\mu(r,p)\right\}^2}{\sigma(r,p)^3}\frac{\left\{x_i^s-\mu(s,q)\right\}^2}{\sigma(s,q)^3} \right. \\
&+\delta_{pq}\mathit{\Delta}_i a(p)\exp[i,p]\,\frac{\left\{x_i^r-\mu(r,p)\right\}^2}{\sigma(r,p)^3}\frac{\left\{x_i^s-\mu(s,p)\right\}^2}{\sigma(s,p)^3} \\
&\left. -3\delta_{pq}\delta_{rs}\mathit{\Delta}_i a(p)\exp[i,p]\,\frac{\left\{x_i^r-\mu(r,p)\right\}^2}{\sigma(r,p)^4}\right\}
\end{aligned} \tag{3.28}
$$

4 penalty 関数

Newton-Raphson 法による fitting では, パラメータの発散や, 収束したとしても異常な結果となる可能性がある. こうした問題を防ぐために, penalty 関数の導入が有効であることが知られている. penalty 関数とは, パラメータが正常な値をとっている間は値が小さいが, 異常なパラメータに対しては大きな値を返すような関数である. Gauss 関数による fitting では, 3 種類の penalty 関数 P_1, P_2, P_3 を導入する. 実際に最小化される関数は,

$$
L_{\mathrm{tot}} = L + \lambda_1 P_1 + \lambda_2 P_2 + \lambda_3 P_3 \tag{4.1}
$$

ここで, $\lambda_1, \lambda_2, \lambda_3$ はそれぞれ penalty 関数 P_1, P_2, P_3 によるパラメータの矯正効果の大きさを決める係数である. iteration の過程でいずれかのパラメータが異常な値となった際には, penalty 関数が大きな値を持つことで L_{tot} が大きくなり, Gauss パラメータを正常な範囲に押し戻す.

4.1　極端に小さな σ の禁止

第一の penalty 関数 P_1 は, 極端に小さな σ, すなわち Gauss 関数が極端に鋭い場合に大きな値をとる. 離散点の間隔よりも細く鋭い Gauss 関数を離散点の数だけ用いるような解 (overfitting) は無意味であるため, この penalty 関数によって除外する.

$$P_1 = \sum_{j=1}^{N_{\mathrm{gauss}}} \sum_{l=1}^{N_{\mathrm{dim}}} \left\{ \frac{d(l, I(j))}{\sigma(l, j)} \right\}^2 \tag{4.2}$$

$$\frac{\partial P_1}{\partial \sigma(r, p)} = -2 \frac{d(r, I(p))^2}{\sigma(r, p)^3} \tag{4.3}$$

$$\frac{\partial^2 P_1}{\partial \sigma(r, p) \partial \sigma(s, q)} = 6 \delta_{pq} \delta_{rs} \frac{d(r, I(p))^2}{\sigma(r, p)^4} \tag{4.4}$$

a, μ での微分は 0 である. ただし, $I(j)$ を $\mu(:, j)$ から見て最近接の離散データ点の番号とした. $d(l, i)$ は離散点の密集度合いを表す量で, 離散点の並びが格子点状の場合, 全ての離散点の l 次元軸への射影における離散点 i と他の離散点の間の 0 ではない最小距離を用いればよい. $d(l, i)$ の具体的な計算方法については後述する.

4.2　fitting に寄与しない μ の禁止

第二の penalty 関数 P_2 は, 離散データの存在範囲から大きく離れた μ が存在する場合に大きな値をとる. 残差 L の計算では離散点上の残差のみを考慮しており, このような μ をもつ Gauss 関数は fitting の成否に関与しないが, fitting 結果を実際に運用する際には問題となるため, この penalty 関数により除外する.

$$P_2 = \sum_{j=1}^{N_{\mathrm{gauss}}} \left(\exp \left[\frac{1}{D^2} \sum_{l=1}^{N_{\mathrm{dim}}} \left\{ x^l_{I(j)} - \mu(l, j) \right\}^2 \right] - 1 \right) \tag{4.5}$$

$$\frac{\partial P_2}{\partial \mu(r, p)} = -\frac{2}{D^2} \left\{ x^r_{I(p)} - \mu(r, p) \right\} \exp \left[\frac{1}{D^2} \sum_{l=1}^{N_{\mathrm{dim}}} \left\{ x^l_{I(p)} - \mu(l, p) \right\}^2 \right] \tag{4.6}$$

$$\begin{aligned} \frac{\partial^2 P_2}{\partial \mu(r, p) \partial \mu(s, q)} &= \delta_{pq} \delta_{rs} \frac{2}{D^2} \exp \left[\frac{1}{D^2} \sum_{l=1}^{N_{\mathrm{dim}}} \left\{ x^l_{I(p)} - \mu(l, p) \right\}^2 \right] \\ &\quad - \delta_{pq} \frac{2}{D^2} \left\{ x^r_{I(p)} - \mu(r, p) \right\} \left(-\frac{2}{D^2} \right) \left\{ x^s_{I(p)} - \mu(s, p) \right\} \exp \left[\frac{1}{D^2} \sum_{l=1}^{N_{\mathrm{dim}}} \left\{ x^l_{I(p)} - \mu(l, p) \right\}^2 \right] \\ &= \delta_{pq} \left\{ \frac{4}{D^4} \left\{ x^r_{I(p)} - \mu(r, p) \right\} \left\{ x^s_{I(p)} - \mu(s, p) \right\} + \delta_{rs} \frac{2}{D^2} \right\} \exp \left[\frac{1}{D^2} \sum_{l=1}^{N_{\mathrm{dim}}} \left\{ x^l_{I(p)} - \mu(l, p) \right\}^2 \right] \end{aligned} \tag{4.7}$$

a, σ での微分は 0 である. ただし, $I(j)$ は $\mu(:, j)$ から見て最近接の離散データ点の番号であり, (4.5) の -1 の項は P_2 の最小値を 0 にするために加えている. D は離散点間距離の平均値などとし, その具体的な計算方法については後述する.

4.3　Gauss 関数同士の redundant な重なりの禁止

第三の penalty 関数 P_3 は, Gauss 関数の任意のペアが大きな重なりをもつ, すなわち μ, σ の値が非常に近い場合に大きな値をとる. 例えば 2 つの Gauss 関数の μ, σ が同じで a の絶対値が等しく逆符号の場合, これらの

和は Gauss 関数が 1 つもないことと等価である. Gauss 関数の数を増やすにつれて fitting 精度が向上するためには, redundant な Gauss 関数ペアの生成を防ぐ必要がある.

基底関数 $|j\rangle$ を次のように定義する.

$$|j\rangle \equiv \exp\left[-\frac{1}{2}\sum_{l=1}^{N_{\mathrm{dim}}}\left\{\frac{x^l-\mu(l,j)}{\sigma(l,j)}\right\}^2\right] \tag{4.8}$$

2 つの基底関数 $|j\rangle, |k\rangle$ の重なり積分は,

$$\langle j|k\rangle = \int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\sum_{l=1}^{N_{\mathrm{dim}}}\left\{\frac{x^l-\mu(l,j)}{\sigma(l,j)}\right\}^2\right]\exp\left[-\frac{1}{2}\sum_{l=1}^{N_{\mathrm{dim}}}\left\{\frac{x^l-\mu(l,k)}{\sigma(l,k)}\right\}^2\right]dx^1\cdots dx^{N_{\mathrm{dim}}} \tag{4.9}$$

さらに, $|j\rangle, |k\rangle$ の重なりの二乗 S_{jk} を定義する.

$$S_{jk} \equiv \frac{\langle j|k\rangle^2}{\langle j|j\rangle\langle k|k\rangle} \tag{4.10}$$

$\langle j|k\rangle = \langle k|j\rangle$ より $S_{jk} = S_{kj}$ であるから, $N_{\mathrm{gauss}} \times N_{\mathrm{gauss}}$ 正方行列 S_{jk} は対称である. これを用いて, penalty 関数 P_3 は以下のように定義される.

$$P_3 = \sum_{j=1}^{N_{\mathrm{gauss}}}\sum_{k=j+1}^{N_{\mathrm{gauss}}}\frac{1}{1-S_{jk}} \tag{4.11}$$

次に, S_{jk} を計算するための式を示す. 新たな基底関数 $|l,j\rangle$ と, 2 つの基底関数 $|l,j\rangle, |l,k\rangle$ の重なりの二乗 $s_{l,jk}$ を次のように定義する.

$$|l,j\rangle \equiv \exp\left[-\frac{1}{2}\left\{\frac{x^l-\mu(l,j)}{\sigma(l,j)}\right\}^2\right] \tag{4.12}$$

$$s_{l,jk} \equiv \frac{\langle l,j|l,k\rangle^2}{\langle l,j|l,j\rangle\langle l,k|l,k\rangle} \tag{4.13}$$

このとき,

$$\begin{aligned}
\langle j|k\rangle &= \int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\sum_{l=1}^{N_{\mathrm{dim}}}\left\{\frac{x^l-\mu(l,j)}{\sigma(l,j)}\right\}^2\right]\exp\left[-\frac{1}{2}\sum_{l=1}^{N_{\mathrm{dim}}}\left\{\frac{x^l-\mu(l,k)}{\sigma(l,k)}\right\}^2\right]dx^1\cdots dx^{N_{\mathrm{dim}}} \\
&= \int_{-\infty}^{\infty}\prod_{l=1}^{N_{\mathrm{dim}}}\left(\exp\left[-\frac{1}{2}\left\{\frac{x^l-\mu(l,j)}{\sigma(l,j)}\right\}^2\right]\exp\left[-\frac{1}{2}\left\{\frac{x^l-\mu(l,k)}{\sigma(l,k)}\right\}^2\right]\right)dx^1\cdots dx^{N_{\mathrm{dim}}} \\
&= \prod_{l=1}^{N_{\mathrm{dim}}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left\{\frac{x^l-\mu(l,j)}{\sigma(l,j)}\right\}^2\right]\exp\left[-\frac{1}{2}\left\{\frac{x^l-\mu(l,k)}{\sigma(l,k)}\right\}^2\right]dx^l \\
&= \prod_{l=1}^{N_{\mathrm{dim}}}\langle l,j|l,k\rangle \qquad (4.14) \\
S_{jk} &= \frac{\langle j|k\rangle^2}{\langle j|j\rangle\langle k|k\rangle} \\
&= \frac{\left\{\prod_{l=1}^{N_{\mathrm{dim}}}\langle l,j|l,k\rangle\right\}^2}{\left\{\prod_{l=1}^{N_{\mathrm{dim}}}\langle l,j|l,j\rangle\right\}\left\{\prod_{l=1}^{N_{\mathrm{dim}}}\langle l,k|l,k\rangle\right\}}
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{l=1}^{N_{\mathrm{dim}}} \frac{\langle l, j|l, k\rangle^2}{\langle l, j|l, j\rangle \langle l, k|l, k\rangle} \\
&= \prod_{l=1}^{N_{\mathrm{dim}}} s_{l,jk}
\end{aligned} \tag{4.15}
$$

重なり積分 $\langle l, j|l, k\rangle$ は積分を実行可能である. はじめに, exp の中を平方完成する.

$$
\begin{aligned}
\langle l, j|l, k\rangle &= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left\{\frac{x^l - \mu(l, j)}{\sigma(l, j)}\right\}^2\right] \exp\left[-\frac{1}{2}\left\{\frac{x^l - \mu(l, k)}{\sigma(l, k)}\right\}^2\right] dx^l \\
&= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left\{\frac{\left(x^l\right)^2 - 2\mu(l, j)x^l + \mu(l, j)^2}{\sigma(l, j)^2} + \frac{\left(x^l\right)^2 - 2\mu(l, k)x^l + \mu(l, k)^2}{\sigma(l, k)^2}\right\}\right] dx^l \\
&= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left\{\left(\frac{1}{\sigma(l, j)^2} + \frac{1}{\sigma(l, k)^2}\right)(x^l)^2\right.\right. \\
&\quad \left.\left. -2\left(\frac{\mu(l, j)}{\sigma(l, j)^2} + \frac{\mu(l, k)}{\sigma(l, k)^2}\right)x^l + \left(\frac{\mu(l, j)^2}{\sigma(l, j)^2} + \frac{\mu(l, k)^2}{\sigma(l, k)^2}\right)\right\}\right] dx^l \\
&= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma(l, j)^2} + \frac{1}{\sigma(l, k)^2}\right)\right. \\
&\quad \left.\left\{(x^l)^2 - 2\frac{\mu(l, j)\sigma(l, k)^2 + \mu(l, k)\sigma(l, j)^2}{\sigma(l, j)^2 + \sigma(l, k)^2}x^l + \frac{\mu(l, j)^2\sigma(l, k)^2 + \mu(l, k)^2\sigma(l, j)^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right\}\right] dx^l \\
&= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma(l, j)^2} + \frac{1}{\sigma(l, k)^2}\right)\right. \\
&\quad \left.\left\{\left(x^l - \frac{\mu(l, j)\sigma(l, k)^2 + \mu(l, k)\sigma(l, j)^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right)^2 + \frac{\sigma(l, j)^2\sigma(l, k)^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\frac{(\mu(l, j) - \mu(l, k))^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right\}\right] dx^l \\
&= \exp\left[-\frac{1}{2}\frac{(\mu(l, j) - \mu(l, k))^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right] \\
&\quad \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma(l, j)^2} + \frac{1}{\sigma(l, k)^2}\right)\left(x^l - \frac{\mu(l, j)\sigma(l, k)^2 + \mu(l, k)\sigma(l, j)^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right)^2\right] dx^l
\end{aligned} \tag{4.16}
$$

ここで, 次の変数変換を行う.

$$
t = x^l - \frac{\mu(l, j)\sigma(l, k)^2 + \mu(l, k)\sigma(l, j)^2}{\sigma(l, j)^2 + \sigma(l, k)^2} \tag{4.17}
$$

これにより,

$$
\begin{aligned}
\langle l, j|l, k\rangle &= \exp\left[-\frac{1}{2}\frac{(\mu(l, j) - \mu(l, k))^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right] \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma(l, j)^2} + \frac{1}{\sigma(l, k)^2}\right)t^2\right] dt \\
&= \sqrt{\frac{2\pi}{\frac{1}{\sigma(l, j)^2} + \frac{1}{\sigma(l, k)^2}}} \exp\left[-\frac{1}{2}\frac{(\mu(l, j) - \mu(l, k))^2}{\sigma(l, j)^2 + \sigma(l, k)^2}\right]
\end{aligned} \tag{4.18}
$$

最後の変形ではガウス積分を用いた. 特に $j = k$ のとき,

$$
\langle l, j|l, j\rangle = \sqrt{\frac{2\pi}{\frac{2}{\sigma(l, j)^2}}} \exp[0] = \sqrt{\pi\sigma(l, j)^2} \tag{4.19}
$$

従って,

$$
\begin{aligned}
s_{l,jk} &= \frac{2\pi}{\frac{1}{\sigma(l,j)^2}+\frac{1}{\sigma(l,k)^2}} \frac{1}{\pi\sqrt{\sigma(l,j)^2\sigma(l,k)^2}} \left\{\exp\left[-\frac{1}{2}\frac{(\mu(l,j)-\mu(l,k))^2}{\sigma(l,j)^2+\sigma(l,k)^2}\right]\right\}^2 \\
&= \frac{2\sqrt{\sigma(l,j)^2\sigma(l,k)^2}}{\sigma(l,j)^2+\sigma(l,k)^2} \exp\left[-\frac{(\mu(l,j)-\mu(l,k))^2}{\sigma(l,j)^2+\sigma(l,k)^2}\right] \qquad (4.20)
\end{aligned}
$$

ただし, iteration の過程で $\sigma < 0$ になる場合を考慮して平方根を残しておいた. $\langle l,j|l,k\rangle = \langle l,k|l,j\rangle$ より $s_{l,jk} = s_{l,kj}$ だから, $s_{l,jk}$ も S_{jk} と同様に対称行列である.

さて, 関数 $S_{jk}, s_{l,jk}$ の性質を考える.

$$
\begin{aligned}
&\left\{2\sqrt{\sigma(l,j)^2\sigma(l,k)^2}\right\}^2 - \left\{\sigma(l,j)^2+\sigma(l,k)^2\right\}^2 \\
&= 4\sigma(l,j)^2\sigma(l,k)^2 - \left\{\sigma(l,j)^4 + 2\sigma(l,j)^2\sigma(l,k)^2 + \sigma(l,k)^4\right\} \\
&= -\left\{\sigma(l,j)^4 - 2\sigma(l,j)^2\sigma(l,k)^2 + \sigma(l,k)^4\right\} \\
&= -\left\{\sigma(l,j)^2 - \sigma(l,k)^2\right\}^2 \leq 0 \qquad (4.21)
\end{aligned}
$$

変形して,

$$
\begin{aligned}
\left\{2\sqrt{\sigma(l,j)^2\sigma(l,k)^2}\right\}^2 &\leq \left\{\sigma(l,j)^2+\sigma(l,k)^2\right\}^2 \\
2\sqrt{\sigma(l,j)^2\sigma(l,k)^2} &\leq \sigma(l,j)^2+\sigma(l,k)^2 \qquad (4.22)
\end{aligned}
$$

非負は自明より, 結局,

$$
0 \leq \frac{2\sqrt{\sigma(l,j)^2\sigma(l,k)^2}}{\sigma(l,j)^2+\sigma(l,k)^2} \leq 1 \qquad (4.23)
$$

一方,

$$
-\frac{(\mu(l,j)-\mu(l,k))^2}{\sigma(l,j)^2+\sigma(l,k)^2} \leq 0 \qquad (4.24)
$$

すなわち,

$$
0 < \exp\left[-\frac{(\mu(l,j)-\mu(l,k))^2}{\sigma(l,j)^2+\sigma(l,k)^2}\right] \leq 1 \qquad (4.25)
$$

以上により, $0 \leq s_{l,jk} \leq 1, 0 \leq S_{jk} \leq 1$ が成り立つ. 特に $\mu(l,j) = \mu(l,k)$ かつ $\sigma(l,j) = \sigma(l,k)$ のとき,

$$
\begin{aligned}
s_{l,jk} &= \frac{2\sqrt{\sigma(l,j)^2\sigma(l,j)^2}}{\sigma(l,j)^2+\sigma(l,j)^2} \exp\left[-\frac{(\mu(l,j)-\mu(l,j))^2}{\sigma(l,j)^2+\sigma(l,j)^2}\right] \\
&= \frac{2\sigma(l,j)^2}{2\sigma(l,j)^2} \exp[0] \\
&= 1 \qquad (4.26)
\end{aligned}
$$

つまり, 関数 S_{jk} は Gauss 関数 j,k が (a を除いて) ぴったり重なるときに最大値 1 をとる. (4.11) は $S_{jk} = 1$ で発散し, Gauss 関数同士の redundant な重なりを防ぐ.

以下, penalty 関数 P_3 の微分の式を求める. まず関数 S_{jk} の微分は,

$$\begin{aligned}\frac{\partial S_{jk}}{\partial \mu(l,j)} &= \left\{\prod_{m\neq l} s_{m,jk}\right\} \frac{\partial s_{l,jk}}{\partial \mu(l,j)} \\ &= \frac{S_{jk}}{s_{l,jk}} \frac{\partial s_{l,jk}}{\partial \mu(l,j)} \\ &= S_{jk} \frac{\partial \ln s_{l,jk}}{\partial \mu(l,j)} \end{aligned} \tag{4.27}$$

$$\frac{\partial S_{jk}}{\partial \sigma(l,j)} = S_{jk} \frac{\partial \ln s_{l,jk}}{\partial \sigma(l,j)} \tag{4.28}$$

これを用いて,

$$\begin{aligned}\frac{\partial P_3}{\partial \mu(r,p)} &= \sum_{j\neq p} \frac{\partial}{\partial \mu(r,p)} \left(\frac{1}{1-S_{pj}}\right) \\ &= \sum_{j\neq p} \left\{-\frac{1}{(1-S_{pj})^2}\right\} \left\{-\frac{\partial S_{pj}}{\partial \mu(r,p)}\right\} \\ &= \sum_{j\neq p} \frac{S_{pj}}{(1-S_{pj})^2} \frac{\partial \ln s_{r,pj}}{\partial \mu(r,p)} \end{aligned} \tag{4.29}$$

$$\frac{\partial P_3}{\partial \sigma(r,p)} = \sum_{j\neq p} \frac{S_{pj}}{(1-S_{pj})^2} \frac{\partial \ln s_{r,pj}}{\partial \sigma(r,p)} \tag{4.30}$$

また, $p = q$ のとき,

$$\begin{aligned}\frac{\partial^2 P_3}{\partial \mu(r,p)\partial \mu(s,q)} &= \sum_{j\neq p} \left\{ \frac{\partial S_{pj}}{\partial \mu(s,p)} \frac{1}{(1-S_{pj})^2} \frac{\partial \ln s_{r,pj}}{\partial \mu(r,p)} + S_{pj} \left(-2\frac{1}{(1-S_{pj})^3}\right) \left(-\frac{\partial S_{pj}}{\partial \mu(s,p)}\right) \frac{\partial \ln s_{r,pj}}{\partial \mu(r,p)} \right. \\ &\qquad \left. +\delta_{rs} \frac{S_{pj}}{(1-S_{pj})^2} \frac{\partial^2 \ln s_{r,pj}}{\partial \mu(r,p)\partial \mu(r,p)} \right\} \\ &= \sum_{j\neq p} \left\{ \frac{1+S_{pj}}{(1-S_{pj})^3} S_{pj} \frac{\partial \ln s_{r,pj}}{\partial \mu(r,p)} \frac{\partial \ln s_{s,pj}}{\partial \mu(s,p)} + \delta_{rs} \frac{S_{pj}}{(1-S_{pj})^2} \frac{\partial^2 \ln s_{r,pj}}{\partial \mu(r,p)^2} \right\} \end{aligned} \tag{4.31}$$

$$\frac{\partial^2 P_3}{\partial \mu(r,p)\partial \sigma(s,q)} = \sum_{j\neq p} \left\{ \frac{1+S_{pj}}{(1-S_{pj})^3} S_{pj} \frac{\partial \ln s_{r,pj}}{\partial \mu(r,p)} \frac{\partial \ln s_{s,pj}}{\partial \sigma(s,p)} + \delta_{rs} \frac{S_{pj}}{(1-S_{pj})^2} \frac{\partial^2 \ln s_{r,pj}}{\partial \mu(r,p)\partial \sigma(r,p)} \right\} \tag{4.32}$$

$$\frac{\partial^2 P_3}{\partial \sigma(r,p)\partial \sigma(s,q)} = \sum_{j\neq p} \left\{ \frac{1+S_{pj}}{(1-S_{pj})^3} S_{pj} \frac{\partial \ln s_{r,pj}}{\partial \sigma(r,p)} \frac{\partial \ln s_{s,pj}}{\partial \sigma(s,p)} + \delta_{rs} \frac{S_{pj}}{(1-S_{pj})^2} \frac{\partial^2 \ln s_{r,pj}}{\partial \sigma(r,p)^2} \right\} \tag{4.33}$$

$p \neq q$ のとき,

$$\begin{aligned}\frac{\partial^2 P_3}{\partial \mu(r,p)\partial \mu(s,q)} &= \frac{\partial S_{pq}}{\partial \mu(s,q)} \frac{1}{(1-S_{pq})^2} \frac{\partial \ln s_{r,pq}}{\partial \mu(r,p)} + S_{pq} \left(-2\frac{1}{(1-S_{pq})^3}\right) \left(-\frac{\partial S_{pq}}{\partial \mu(s,q)}\right) \frac{\partial \ln s_{r,pq}}{\partial \mu(r,p)} \\ &\qquad + \delta_{rs} \frac{S_{pq}}{(1-S_{pq})^2} \frac{\partial^2 \ln s_{r,pq}}{\partial \mu(r,p)\partial \mu(r,q)} \\ &= \frac{1+S_{pq}}{(1-S_{pq})^3} S_{pq} \frac{\partial \ln s_{r,pq}}{\partial \mu(r,p)} \frac{\partial \ln s_{s,pq}}{\partial \mu(s,q)} + \delta_{rs} \frac{S_{pq}}{(1-S_{pq})^2} \frac{\partial^2 \ln s_{r,pq}}{\partial \mu(r,p)\partial \mu(r,q)} \end{aligned} \tag{4.34}$$

$$\frac{\partial^2 P_3}{\partial\mu(r,p)\partial\sigma(s,q)} = \frac{1+S_{pq}}{\left(1-S_{pq}\right)^3}S_{pq}\frac{\partial \ln s_{r,pq}}{\partial\mu(r,p)}\frac{\partial \ln s_{s,pq}}{\partial\sigma(s,q)} + \delta_{rs}\frac{S_{pq}}{\left(1-S_{pq}\right)^2}\frac{\partial^2 \ln s_{r,pq}}{\partial\mu(r,p)\partial\sigma(r,q)} \tag{4.35}$$

$$\frac{\partial^2 P_3}{\partial\sigma(r,p)\partial\sigma(s,q)} = \frac{1+S_{pq}}{\left(1-S_{pq}\right)^3}S_{pq}\frac{\partial \ln s_{r,pq}}{\partial\sigma(r,p)}\frac{\partial \ln s_{s,pq}}{\partial\sigma(s,q)} + \delta_{rs}\frac{S_{pq}}{\left(1-S_{pq}\right)^2}\frac{\partial^2 \ln s_{r,pq}}{\partial\sigma(r,p)\partial\sigma(r,q)} \tag{4.36}$$

最後に $\ln s_{l,jk}$ の微分を示すが, ここでは表記を簡単にするために次の記法を用いる.

$$s \equiv s_{l,jk} = s_{l,kj} \tag{4.37}$$

$$\mu_1 \equiv \mu(l,j) \tag{4.38}$$

$$\mu_2 \equiv \mu(l,k) \tag{4.39}$$

$$\sigma_1 \equiv \sigma(l,j) \tag{4.40}$$

$$\sigma_2 \equiv \sigma(l,k) \tag{4.41}$$

このとき (4.20) は,

$$s = \frac{2\sqrt{\sigma_1^2\sigma_2^2}}{\sigma_1^2+\sigma_2^2}\exp\left[-\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right] \tag{4.42}$$

両辺の対数をとると,

$$\begin{aligned}\ln s &= \ln\frac{2\sqrt{\sigma_1^2\sigma_2^2}}{\sigma_1^2+\sigma_2^2} + \ln\left\{\exp\left[-\frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2}\right]\right\} \\ &= \ln 2 + \frac{1}{2}\ln\left(\sigma_1^2\right) + \frac{1}{2}\ln\left(\sigma_2^2\right) - \ln\left(\sigma_1^2+\sigma_2^2\right) - \frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2} \end{aligned}\tag{4.43}$$

従って,

$$\frac{\partial \ln s}{\partial \mu_1} = -\frac{2\left(\mu_1-\mu_2\right)}{\sigma_1^2+\sigma_2^2} \tag{4.44}$$

$$\begin{aligned}\frac{\partial \ln s}{\partial \sigma_1} &= \frac{1}{2}\frac{1}{\sigma_1^2}2\sigma_1 - \frac{1}{\sigma_1^2+\sigma_2^2}2\sigma_1 - \left(-\frac{(\mu_1-\mu_2)^2}{\left(\sigma_1^2+\sigma_2^2\right)^2}2\sigma_1\right) \\ &= \frac{\left(\sigma_1^2+\sigma_2^2\right)^2 - 2\sigma_1^2\left(\sigma_1^2+\sigma_2^2\right) + 2\sigma_1^2\left(\mu_1-\mu_2\right)^2}{\sigma_1\left(\sigma_1^2+\sigma_2^2\right)^2} \\ &= \frac{-\sigma_1^4+\sigma_2^4+2\sigma_1^2\left(\mu_1-\mu_2\right)^2}{\sigma_1\left(\sigma_1^2+\sigma_2^2\right)^2} \end{aligned}\tag{4.45}$$

$$\frac{\partial^2 \ln s}{\partial \mu_1^2} = -\frac{2}{\sigma_1^2+\sigma_2^2} \tag{4.46}$$

$$\frac{\partial^2 \ln s}{\partial \mu_1\partial\mu_2} = \frac{2}{\sigma_1^2+\sigma_2^2} \tag{4.47}$$

$$\frac{\partial^2 \ln s}{\partial \mu_1\partial\sigma_1} = \frac{4\sigma_1\left(\mu_1-\mu_2\right)}{\left(\sigma_1^2+\sigma_2^2\right)^2} \tag{4.48}$$

$$\frac{\partial^2 \ln s}{\partial \mu_1\partial\sigma_2} = \frac{4\sigma_2\left(\mu_1-\mu_2\right)}{\left(\sigma_1^2+\sigma_2^2\right)^2} \tag{4.49}$$

$$\frac{\partial^2 \ln s}{\partial \sigma_1^2} = \frac{-4\sigma_1^3+4\sigma_1\left(\mu_1-\mu_2\right)^2}{\sigma_1\left(\sigma_1^2+\sigma_2^2\right)^2} - \frac{-\sigma_1^4+\sigma_2^4+2\sigma_1^2\left(\mu_1-\mu_2\right)^2}{\left\{\sigma_1\left(\sigma_1^2+\sigma_2^2\right)^2\right\}^2}\left\{\left(\sigma_1^2+\sigma_2^2\right)^2 + \sigma_1\cdot 2\left(\sigma_1^2+\sigma_2^2\right)\cdot 2\sigma_1\right\}$$

$$= \frac{\left\{-4\sigma_1^3 + 4\sigma_1 (\mu_1 - \mu_2)^2\right\} \cdot \sigma_1 \left(\sigma_1^2 + \sigma_2^2\right) + \left\{\sigma_1^4 - \sigma_2^4 - 2\sigma_1^2 (\mu_1 - \mu_2)^2\right\} \left(5\sigma_1^2 + \sigma_2^2\right)}{\sigma_1^2 \left(\sigma_1^2 + \sigma_2^2\right)^3}$$

$$= \frac{\sigma_1^6 - 3\sigma_1^4\sigma_2^2 - 5\sigma_1^2\sigma_2^4 - \sigma_2^6 + \left(-6\sigma_1^4 + 2\sigma_1^2\sigma_2^2\right)(\mu_1 - \mu_2)^2}{\sigma_1^2 \left(\sigma_1^2 + \sigma_2^2\right)^3} \tag{4.50}$$

$$\frac{\partial^2 \ln s}{\partial \sigma_1 \partial \sigma_2} = \frac{4\sigma_2^3}{\sigma_1 \left(\sigma_1^2 + \sigma_2^2\right)^2} - \frac{-\sigma_1^4 + \sigma_2^4 + 2\sigma_1^2 (\mu_1 - \mu_2)^2}{\left\{\sigma_1 \left(\sigma_1^2 + \sigma_2^2\right)^2\right\}^2} 2\sigma_1 \left(\sigma_1^2 + \sigma_2^2\right) \cdot 2\sigma_2$$

$$= \frac{4\sigma_2^3 \left(\sigma_1^2 + \sigma_2^2\right) + \left\{\sigma_1^4 - \sigma_2^4 - 2\sigma_1^2 (\mu_1 - \mu_2)^2\right\} \cdot 4\sigma_2}{\sigma_1 \left(\sigma_1^2 + \sigma_2^2\right)^3}$$

$$= \frac{4\sigma_1\sigma_2 \left(\sigma_1^2 + \sigma_2^2\right) - 8\sigma_1\sigma_2 (\mu_1 - \mu_2)^2}{\left(\sigma_1^2 + \sigma_2^2\right)^3} \tag{4.51}$$

5 残差 Hessian の修正

Newton-Raphson 法を用いた fitting では残差 L の勾配が 0 となる Gauss パラメータを求めているが, これは L の極小値だけでなく, 極大値に行き着く可能性もあることを意味している. fitting が極小に向かうか極大に向かうかを決めているのは L の Hessian である. この Hessian を, ある条件を満たすように修正することで, fitting が極小値に向かうことを保証することができる.

5.1 負の Hessian 固有値による問題

ある N 次元関数 $f(\boldsymbol{q})$ の $\boldsymbol{q}^0$ まわりのテイラー展開は,

$$f(\boldsymbol{q}^0 + \boldsymbol{d}) = \sum_{k=0}^{n} \frac{1}{k!} \left(d_1 \frac{\partial}{\partial q_1} + d_2 \frac{\partial}{\partial q_2} + \cdots + d_N \frac{\partial}{\partial q_N} \right)^k f(\boldsymbol{q}) \bigg|_{\boldsymbol{q}=\boldsymbol{q}^0} \tag{5.1}$$

ただし,

$$\boldsymbol{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_N \end{pmatrix}, \quad \boldsymbol{q}^0 = \begin{pmatrix} q_1^0 \\ \vdots \\ q_N^0 \end{pmatrix}, \quad \boldsymbol{d} = \begin{pmatrix} d_1 \\ \vdots \\ d_N \end{pmatrix} \tag{5.2}$$

$n = 2$ として展開する. $k = 0$ の項は,

$$\frac{1}{0!} f(\boldsymbol{q}^0) = f(\boldsymbol{q}^0) \tag{5.3}$$

$k = 1$ の項は,

$$\frac{1}{1!} \left(d_1 \frac{\partial}{\partial q_1} + \cdots + d_N \frac{\partial}{\partial q_N} \right) f(\boldsymbol{q}) \bigg|_{\boldsymbol{q}=\boldsymbol{q}^0} = d_1 \left. \frac{\partial f(\boldsymbol{q})}{\partial q_1} \right|_{\boldsymbol{q}=\boldsymbol{q}^0} + \cdots + d_N \left. \frac{\partial f(\boldsymbol{q})}{\partial q_N} \right|_{\boldsymbol{q}=\boldsymbol{q}^0}$$

$$= \left(\nabla f(\boldsymbol{q}^0) \right) \cdot \boldsymbol{d} \tag{5.4}$$

$k=2$ の項は,

$$\frac{1}{2!}\left(d_1\frac{\partial}{\partial q_1}+\cdots+d_N\frac{\partial}{\partial q_N}\right)^2 f(\boldsymbol{q})\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0}=\frac{1}{2}\left(d_1^2\ \frac{\partial^2 f(\boldsymbol{q})}{\partial q_1^2}\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0}+\cdots+d_1 d_N\ \frac{\partial^2 f(\boldsymbol{q})}{\partial q_1\partial q_N}\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0}\right.$$
$$\left.+\cdots+d_N d_1\ \frac{\partial^2 f(\boldsymbol{q})}{\partial q_N\partial q_1}\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0}+\cdots+d_N^2\ \frac{\partial^2 f(\boldsymbol{q})}{\partial q_N^2}\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0}\right)$$
$$=\frac{1}{2}\boldsymbol{d}\cdot\left(\nabla^2 f(\boldsymbol{q}^0)\right)\cdot\boldsymbol{d} \tag{5.5}$$

以上により, 関数 $f(\boldsymbol{q})$ の 2 次までのテイラー展開は,

$$f(\boldsymbol{q}^0+\boldsymbol{d})\simeq f(\boldsymbol{q}^0)+\left(\nabla f(\boldsymbol{q}^0)\right)\cdot\boldsymbol{d}+\frac{1}{2}\boldsymbol{d}\cdot\left(\nabla^2 f(\boldsymbol{q}^0)\right)\cdot\boldsymbol{d} \tag{5.6}$$

あるいは,

$$f(\boldsymbol{q}^0+\boldsymbol{d})\simeq f(\boldsymbol{q}^0)+\sum_{i=1}^{N}\frac{\partial f(\boldsymbol{q})}{\partial q_i}\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0} d_i+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} d_i\ \frac{\partial^2 f(\boldsymbol{q})}{\partial q_i\partial q_j}\bigg|_{\boldsymbol{q}=\boldsymbol{q}^0} d_j \tag{5.7}$$

いま関数 f として残差 L を選び $\boldsymbol{d}=\boldsymbol{q}-\boldsymbol{q}^0$ とすれば,

$$L(\boldsymbol{q})\simeq L(\boldsymbol{q}^0)+\sum_{i=1}^{N}\left(\frac{\partial L}{\partial q_i}\right)^0\left(q_i-q_i^0\right)+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(q_i-q_i^0\right)\left(\frac{\partial^2 L}{\partial q_i\partial q_j}\right)^0\left(q_j-q_j^0\right) \tag{5.8}$$

この右辺を新たに関数 $h(\{q_i\})$ とおく. 両辺を q_i で微分すると,

$$\frac{\partial h(\boldsymbol{q})}{\partial q_i}=\left(\frac{\partial L}{\partial q_i}\right)^0+\frac{1}{2}\sum_{j=1}^{N}\left(\frac{\partial^2 L}{\partial q_i\partial q_j}\right)^0\left(q_j-q_j^0\right)+\frac{1}{2}\sum_{j=1}^{N}\left(q_j-q_j^0\right)\left(\frac{\partial^2 L}{\partial q_j\partial q_i}\right)^0$$
$$=\left(\frac{\partial L}{\partial q_i}\right)^0+\sum_{j=1}^{N}\left(\frac{\partial^2 L}{\partial q_i\partial q_j}\right)^0\left(q_j-q_j^0\right) \tag{5.9}$$

Newton-Raphson 法の連立方程式 (3.4) の解において, この式の右辺は 0 となり, 二次関数 h は極値をとる. つまり, Newton-Raphson 法で $\partial L/\partial q_i$ の接線を求めてそれが 0 となるときの $\{q_i\}$ を得ることは, 残差 L の言わば"接放物線"を求めてそれが極値をとるときの $\{q_i\}$ を得ることと等価である.

fitting が残差 L の極小値に向かうためには, この接放物線 $h(\boldsymbol{q})$ が下に凸である必要がある. つまり, あらゆる変位 $\boldsymbol{d}\neq 0$ に関して次式が成り立たなければならない.

$$\boldsymbol{d}\cdot\left(\nabla^2 L(\boldsymbol{q}^0)\right)\cdot\boldsymbol{d}>0 \tag{5.10}$$

ところで, $N\times N$ 行列 $\boldsymbol{A}$ について次の 3 項目は同値である.

- $\boldsymbol{A}$ は正定値である.
- $\boldsymbol{A}$ の固有値はすべて正である.
- あらゆる N 次元ベクトル $\boldsymbol{d}\neq 0$ に対して $\boldsymbol{d}\cdot\boldsymbol{A}\cdot\boldsymbol{d}>0$ が成り立つ.

従って, Hessian$\nabla^2 L(\boldsymbol{q}^0)$ の固有値がすべて正であることは, fitting が残差 L の極小値に収束するための必要十分条件である.

5.2　極端に小さな Hessian 固有値による問題

$N \times N$ 実対称行列 $\boldsymbol{A}$ は, ある直交行列 $\boldsymbol{P}$ により対角化可能である.

$$\boldsymbol{P}^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{P} = \boldsymbol{\Lambda} \tag{5.11}$$

ただし, $\boldsymbol{A}$ の固有値 $\lambda_1, \lambda_2, \cdots, \lambda_N$ に対して,

$$\boldsymbol{\Lambda} \equiv \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \tag{5.12}$$

(5.11) の両辺に左から $\boldsymbol{P}$, 右から $\boldsymbol{P}^{-1}$ を掛けて,

$$\boldsymbol{P} \cdot \boldsymbol{P}^{-1} \cdot \boldsymbol{A} \cdot \boldsymbol{P} \cdot \boldsymbol{P}^{-1} = \boldsymbol{A} = \boldsymbol{P} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{P}^{-1} \tag{5.13}$$

$\boldsymbol{A}$ が残差 L の Hessian のとき, (5.13) を連立方程式 (3.7) に代入して,

$$\boldsymbol{P} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{P}^{-1} \cdot \boldsymbol{q} = \boldsymbol{b} \tag{5.14}$$

ただし, パラメータ変位 $\boldsymbol{q} - \boldsymbol{q}^0$ を新たに $\boldsymbol{q}$ と置いた. 両辺に左から $\boldsymbol{P}^{-1}$ を掛けて,

$$\boldsymbol{P}^{-1} \cdot \boldsymbol{P} \cdot \boldsymbol{\Lambda} \cdot \boldsymbol{P}^{-1} \cdot \boldsymbol{q} = \boldsymbol{\Lambda} \cdot \left(\boldsymbol{P}^{-1} \cdot \boldsymbol{q} \right) = \boldsymbol{P}^{-1} \cdot \boldsymbol{b} \tag{5.15}$$

ここで $\boldsymbol{q}', \boldsymbol{b}'$ を, それぞれ $\boldsymbol{q}, \boldsymbol{b}$ を回転行列 $\boldsymbol{P}^{-1}$ により変換したものとして導入する.

$$\boldsymbol{q}' = \boldsymbol{P}^{-1} \cdot \boldsymbol{q}, \quad \boldsymbol{b}' = \boldsymbol{P}^{-1} \cdot \boldsymbol{b} \tag{5.16}$$

これらを用いると,

$$\boldsymbol{\Lambda} \cdot \boldsymbol{q}' = \boldsymbol{b}' \tag{5.17}$$

成分で書けば,

$$\begin{aligned} \lambda_i {q_i}' &= {b_i}' \\ {q_i}' &= \frac{1}{\lambda_i} {b_i}' \end{aligned} \tag{5.18}$$

また, 次の関係が成り立つ.

$$\boldsymbol{q} = \boldsymbol{P} \cdot \boldsymbol{q}' \tag{5.19}$$

$\boldsymbol{P}$ の列ベクトル $\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_N$ を導入すれば,

$$\boldsymbol{q} = \sum_{i=1}^{N} \boldsymbol{p}_i {q_i}' = \sum_{i=1}^{N} \frac{1}{\lambda_i} \boldsymbol{p}_i {b_i}' \tag{5.20}$$

ここで (5.18) を用いた. Hessian$\nabla^2 L(\boldsymbol{q}^0)$ の固有値 λ_i が 0 に近いとき, 対応する基底 $\boldsymbol{p}_i$ の係数が非常に大きくなることが分かる. これは Newton-Raphson 法における iteration 1 回分の変位 $\boldsymbol{q}$ が過大であることに対応しており, このとき fitting の収束は困難となる.

5.3 Hessian 固有値の修正

ここまでの議論により, fitting が残差 L の極小値に向かって収束するためには, L の Hessian の固有値が常に十分に大きな正の値である必要がある. そこで, Hessian $\boldsymbol{A}$ の全ての固有値を次式により修正することを考える.

$$\lambda_i{}' = \sqrt{\lambda_i^2 + \varepsilon^2} \tag{5.21}$$

$\varepsilon > 0$ は最小固有値であり, 全ての固有値に対して同じ ε を適用する. このとき, (5.20) は,

$$\boldsymbol{q} = \sum_{i=1}^{N} \frac{1}{\lambda_i{}'} \boldsymbol{p}_i b_i{}' \leq \sum_{i=1}^{N} \frac{1}{\varepsilon} \boldsymbol{p}_i b_i{}' = \frac{1}{\varepsilon} \boldsymbol{P} \cdot \boldsymbol{b}' = \frac{1}{\varepsilon} \boldsymbol{b} \tag{5.22}$$

ここで, 次の関係を用いた.

$$\boldsymbol{b} = \boldsymbol{P} \cdot \boldsymbol{b}' \tag{5.23}$$

よって, 一般のベクトル $\boldsymbol{v}$ の成分のうち最大のものを $\{\boldsymbol{v}_i\}_{\max}$ と表すことにすれば,

$$\left\{\left|\boldsymbol{q}_i\right|\right\}_{\max} \leq \frac{1}{\varepsilon} \{|\boldsymbol{b}_i|\}_{\max} \tag{5.24}$$

つまり, 変位 $\boldsymbol{q}$ の各成分の絶対値は次の $q_{\max}$ により上から抑えられる.

$$q_{\max} = \frac{1}{\varepsilon} \{|\boldsymbol{b}_i|\}_{\max} \tag{5.25}$$

これを変形して,

$$\varepsilon = \frac{\{|\boldsymbol{b}_i|\}_{\max}}{q_{\max}} \tag{5.26}$$

これにより, 最大変位 $q_{\max}$ が与えられたときにそれを満たす ε を計算できる. このとき全ての Gauss パラメータの変位を同じ $q_{\max}$ で抑えることになるが, 離散座標の各成分や入力データが数値的に同程度の値をとるようにすれば問題ない. (5.21) により修正された固有値は, (5.13) に代入することで Hessian に反映される.

6 実装上の工夫

6.1 Gauss パラメータの固定

残差 L を最小化するように N 個の Gauss パラメータを fitting するには, Newton-Raphson 法による N 本の連立方程式 (3.4) を iteration によって解く. ここで, k 番目の Gauss パラメータを固定して $N-1$ 個の Gauss パラメータを fitting する場合には, 代わりに以下の $N-1$ 本の連立方程式を解けばよい.

$$
\begin{cases}
\sum_{j\neq k}\left(\frac{\partial^2 L}{\partial q_1 \partial q_j}\right)^0 (q_j - q_j^0) + \left(\frac{\partial L}{\partial q_1}\right)^0 = 0 \\
\vdots \\
\sum_{j\neq k}\left(\frac{\partial^2 L}{\partial q_{k-1} \partial q_j}\right)^0 (q_j - q_j^0) + \left(\frac{\partial L}{\partial q_{k-1}}\right)^0 = 0 \\
\sum_{j\neq k}\left(\frac{\partial^2 L}{\partial q_{k+1} \partial q_j}\right)^0 (q_j - q_j^0) + \left(\frac{\partial L}{\partial q_{k+1}}\right)^0 = 0 \\
\vdots \\
\sum_{j\neq k}\left(\frac{\partial^2 L}{\partial q_N \partial q_j}\right)^0 (q_j - q_j^0) + \left(\frac{\partial L}{\partial q_N}\right)^0 = 0
\end{cases}
\tag{6.1}
$$

これを行列で表すと,

$$
\begin{pmatrix}
\left(\frac{\partial^2 L}{\partial q_1 \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_1 \partial q_{k-1}}\right)^0 & \left(\frac{\partial^2 L}{\partial q_1 \partial q_{k+1}}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_1 \partial q_N}\right)^0 \\
\vdots & & \vdots & \vdots & & \vdots \\
\left(\frac{\partial^2 L}{\partial q_{k-1} \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_{k-1} \partial q_{k-1}}\right)^0 & \left(\frac{\partial^2 L}{\partial q_{k-1} \partial q_{k+1}}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_{k-1} \partial q_N}\right)^0 \\
\left(\frac{\partial^2 L}{\partial q_{k+1} \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_{k+1} \partial q_{k-1}}\right)^0 & \left(\frac{\partial^2 L}{\partial q_{k+1} \partial q_{k+1}}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_{k+1} \partial q_N}\right)^0 \\
\vdots & & \vdots & \vdots & & \vdots \\
\left(\frac{\partial^2 L}{\partial q_N \partial q_1}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_N \partial q_{k-1}}\right)^0 & \left(\frac{\partial^2 L}{\partial q_N \partial q_{k+1}}\right)^0 & \cdots & \left(\frac{\partial^2 L}{\partial q_N \partial q_N}\right)^0
\end{pmatrix}
\begin{pmatrix}
q_1 - q_1^0 \\ \vdots \\ q_{k-1} - q_{k-1}^0 \\ q_{k+1} - q_{k+1}^0 \\ \vdots \\ q_N - q_N^0
\end{pmatrix}
+
\begin{pmatrix}
\left(\frac{\partial L}{\partial q_1}\right)^0 \\ \vdots \\ \left(\frac{\partial L}{\partial q_{k-1}}\right)^0 \\ \left(\frac{\partial L}{\partial q_{k+1}}\right)^0 \\ \vdots \\ \left(\frac{\partial L}{\partial q_N}\right)^0
\end{pmatrix}
=
\begin{pmatrix}
0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0
\end{pmatrix}
\tag{6.2}
$$

(3.5) と比較すると, この式は行列からは k 行目と k 列目を削除し, ベクトルからは k 番目の成分を消去したものである. つまり, 特定の Gauss パラメータを固定した fitting を行いたい場合には, 連立方程式を構成する行列やベクトルから固定したいパラメータに対応する成分を消去すればよい.

残差 L が threshold を満たさず Gauss 関数を追加した後, 全ての Gauss 関数を一度に fitting すると異常なパラメータになる危険が大きいと予想される. そこで, はじめは既存の Gauss 関数のパラメータを固定して新たに追加した関数だけを fitting し, その後は新しい Gauss 関数から順に動かす関数を増やしてゆく方法を採用する.

6.2 指数関数の近似計算による高速化

ガウス関数に含まれる指数関数の計算は比較的遅い. そこで, 離散的な座標について予め正確な指数関数の値を保持しておき, 指数関数の計算が必要になった際には保持しておいたリストから内挿することを考える.

指数関数 e^{-x} $(x \geq 0)$ について, x 軸上に間隔 d で N_{div} 個の離散座標 $\{x_i\}$ $(i = 1, \cdots, N_{\mathrm{div}})$ を置く.

$$x_i = (i-1)d \tag{6.3}$$

各点上の指数関数の値のリスト $\{f_i\}$ $(i = 1, \cdots, N_{\mathrm{div}})$ を以下のように計算する.

$$f_i = \mathrm{e}^{-x_i} - f^0 \tag{6.4}$$

$$f^0 = \mathrm{e}^{-x_{N_{\mathrm{div}}}} \tag{6.5}$$

N_{div} と d は f^0 が十分小さくなるように選ぶ. 任意の $x \geq 0$ における指数関数の値を求めるには, まず $x/d+1$ を整数部分 n と少数部分 a に分割する.

$$\frac{x}{d}+1 = n+a \tag{6.6}$$

このとき, 指数関数の値 $f(x)$ は次のように内挿される.

$$f(x) = \begin{cases} f_n + (f_{n+1}-f_n)\,a & (n < N_{\mathrm{div}}) \\ 0 & (n \geq N_{\mathrm{div}}) \end{cases} \tag{6.7}$$

リストの全体から f^0 を差し引くことで, 関数 $f(x)$ は連続関数となっている.

6.3 グリッドの利用による高速化

penalty 関数 P_1 に含まれる $d(p,i)$ は離散点の局所的な密集度合いを表す量である. P_1 の定義を示す際には, 全ての離散点の p 次元軸への射影における離散点 i と他の離散点の間の 0 ではない最小の距離としていたが, 離散点の並びが格子点状ではない場合には局所的な疎密を反映できないなどうまく機能しない. 一方, i と他すべての離散点との距離を計算し, 最近接のものから順に p 次元座標の差の絶対値を求めて初めてノンゼロになったものを $d(p,i)$ に選ぶこともできる. この方法は局所的な密集度を反映するが, 合計で約 $(N_{\mathrm{data}})^2$ 回の距離計算が必要となり, 例えば $N_{\mathrm{data}} = 10000$ のとき 10000^2 回と, 計算時間の観点から現実的ではない.

そこで, 離散点が分布している各次元軸上の範囲を N_{grid} 等分したグリッドを考える. 離散点の座標は各成分が同程度の値をとるため, $N_{\mathrm{dim}} = 2$ のときは $N_{\mathrm{grid}} \times N_{\mathrm{grid}}$ 個の正方形, $N_{\mathrm{dim}} = 3$ のときは $N_{\mathrm{grid}} \times N_{\mathrm{grid}} \times N_{\mathrm{grid}}$ 個の立方体を考えればよい. 離散点 i が属するグリッドの p 次元方向の密集度を評価するため, 各グリッドの辺を N_{hist} 等分する $\left(N_{\mathrm{grid}}\right)^{N_{\mathrm{dim}}} \times N_{\mathrm{dim}}$ 個のヒストグラムを用意する. 各グリッド内に存在する離散点を各次元軸に射影し, そのグリッドその次元に対応するヒストグラムに割り振る. i 番目の離散点が属するグリッド番号 $G(i)$ の p 次元方向のヒストグラムで 0 ではないビンの数を $n(p, G(i))$ とすると, $d(p,i)$ は次式で定義される. ただし L_p は p 軸方向のグリッドの辺の長さである.

$$d(p,i) = L_p \frac{n(p, G(i))}{N_{\mathrm{hist}}} \tag{6.8}$$

この定義では, $d(p,i)$ は同じグリッドに属する離散点で同じ値となり, グリッドの大きさの精度で局所的な疎密を反映する. その最小値はヒストグラムのビンの幅, 最大値はグリッドの一辺の長さとなる.

また, penalty 関数 P_2 に含まれる D は μ が離散点から離れることが許される距離の目安である. P_2 の定義を示す際に言及したように離散点間の平均距離を採用する場合には次式を用いる. ただし, i 番目の離散点の座標を $\boldsymbol{x}_i$ とした.

$$D = \frac{1}{N_{\mathrm{data}}\left(N_{\mathrm{data}}-1\right)} \sum_i \sum_{j \neq i} \left|\boldsymbol{x}_i - \boldsymbol{x}_j\right| \tag{6.9}$$

この計算にもおよそ $(N_{\mathrm{data}})^2$ 回の距離計算が必要となる. そこで, 代わりにグリッド間距離の重み付き平均で D を定義することを考える. 以下の説明では $N_{\mathrm{dim}} = 2$ を仮定する. p 軸方向のグリッドの辺の長さを L_p, グリッドの位置を整数で表した座標を I_p, J_p などとし, 以下のベクトルを定義する.

$$\boldsymbol{L} = \begin{pmatrix} L_x \\ L_y \end{pmatrix}, \quad \boldsymbol{I} = \begin{pmatrix} I_x \\ I_y \end{pmatrix}, \quad \boldsymbol{J} = \begin{pmatrix} J_x \\ J_y \end{pmatrix} \tag{6.10}$$

このとき, グリッド間の相対座標は次式で表される.

$$\boldsymbol{L}\left(\boldsymbol{I}-\boldsymbol{J}\right) = \begin{pmatrix} L_x\left(I_x - J_x\right) \\ L_y\left(I_y - J_y\right) \end{pmatrix} \tag{6.11}$$

グリッド $\boldsymbol{I}$ に属する離散点の数を $N(\boldsymbol{I})$ とすると, 新たな D は次のように定義される. これはグリッド $\boldsymbol{I}$ と $\boldsymbol{J}$ のペアに含まれる離散点ペアの数 $N(\boldsymbol{I})N(\boldsymbol{J})$ による重みつき平均である.

$$D = \frac{1}{N_{\mathrm{data}}\left(N_{\mathrm{data}} - 1\right)} \sum_{\boldsymbol{I}} \sum_{\boldsymbol{J}} N(\boldsymbol{I})N(\boldsymbol{J}) \left|\boldsymbol{L}(\boldsymbol{I} - \boldsymbol{J})\right| \tag{6.12}$$

このとき, 距離計算の回数はおよそ $\left\{\left(N_{\mathrm{grid}}\right)^{N_{\mathrm{dim}}}\right\}^2$ 回である. 例えば $N_{\mathrm{dim}} = 3, N_{\mathrm{grid}} = 10$ のとき 1000^2 回となる. さらに別の方法として, D をグリッド間距離の二乗の重み付き平均の平方根で定義してもよい. この場合は次式が D の定義式となる.

$$D = \sqrt{\frac{1}{N_{\mathrm{data}}\left(N_{\mathrm{data}} - 1\right)} \sum_{\boldsymbol{I}} \sum_{\boldsymbol{J}} N(\boldsymbol{I})N(\boldsymbol{J}) \left|\boldsymbol{L}(\boldsymbol{I} - \boldsymbol{J})\right|^2} \tag{6.13}$$

余談であるが, (6.13) の二乗と (6.12) の二乗の差はグリッド間距離の分散に等しい. (6.13) に含まれる和は変形できて,

$$\begin{aligned}
&\sum_{\boldsymbol{I}} \sum_{\boldsymbol{J}} N(\boldsymbol{I})N(\boldsymbol{J}) \left|\boldsymbol{L}(\boldsymbol{I} - \boldsymbol{J})\right|^2 \\
&= \sum_{I_x} \sum_{J_x} \sum_{I_y} \sum_{J_y} N(I_x, I_y)N(J_x, J_y) \left\{L_x^2(I_x - J_x)^2 + L_y^2(I_y - J_y)^2\right\} \\
&= \sum_{I_x} \sum_{J_x} \sum_{I_y} \sum_{J_y} N(I_x, I_y)N(J_x, J_y) \left\{L_x^2(I_x - J_x)^2\right\} + \sum_{I_x} \sum_{J_x} \sum_{I_y} \sum_{J_y} N(I_x, I_y)N(J_x, J_y) \left\{L_y^2(I_y - J_y)^2\right\} \\
&= \sum_{I_x} \sum_{J_x} \left\{\sum_{I_y} N(I_x, I_y)\right\} \left\{\sum_{J_y} N(J_x, J_y)\right\} \left\{L_x^2(I_x - J_x)^2\right\} \\
&\quad + \sum_{I_y} \sum_{J_y} \left\{\sum_{I_x} N(I_x, I_y)\right\} \left\{\sum_{J_x} N(J_x, J_y)\right\} \left\{L_y^2(I_y - J_y)^2\right\} \\
&= \sum_{I_x} \sum_{J_x} H_x(I_x)H_x(J_x) \left\{L_x^2(I_x - J_x)^2\right\} + \sum_{I_y} \sum_{J_y} H_y(I_y)H_y(J_y) \left\{L_y^2(I_y - J_y)^2\right\}
\end{aligned} \tag{6.14}$$

ここで $H_p(I_p)$ は, p 軸方向の整数座標が I_p であるグリッドに含まれる離散点の数の総和である. p 軸上にグリッドと同じ刻みのヒストグラムを置き, 全離散点を p 軸に射影してそのヒストグラムに割り振ったと考えてもよい. これにより (6.13) は,

$$D = \sqrt{\frac{1}{N_{\mathrm{data}}\left(N_{\mathrm{data}} - 1\right)} \left\{\sum_{I_x} \sum_{J_x} H_x(I_x)H_x(J_x) \left\{L_x^2(I_x - J_x)^2\right\} + \sum_{I_y} \sum_{J_y} H_y(I_y)H_y(J_y) \left\{L_y^2(I_y - J_y)^2\right\}\right\}} \tag{6.15}$$

この定義によれば, 距離計算の回数はおよそ $N_{\mathrm{dim}} \times \left(N_{\mathrm{grid}}\right)^2$ 回である. 例えば $N_{\mathrm{dim}} = 3, N_{\mathrm{grid}} = 10$ のとき 300 回となる.