補間法と最小二乗法

山本昌志[*]

2007 年 11 月 22 日

概 要

実験データのように組になった複数のデータの関係から，値を推定する方法を学ぶ．ラグランジュ補間とスプライン補間，そして最小二乗法について説明する．

1 はじめに

実験やシミュレーションを行っていると $(x_0, y_0), (x_1, y_1), (x_2, y_2), (x_3, y_3)$ のように，離散的にデータが得られるのは普通のことである．この得られた値から，任意の x に対して y の値を推測しなくてはならないことがしばしば生じる．工学実験では曲線定規を使って値を推測したが，ここでは計算機を用いて値を推測することを学習する．

値を計算するためには，$y = f(x)$ を示す関数 f を決めればよい．この関数は，次のような二通りの決め方がある．

補間法 これは，離散的な点，全てを通過する関数を求め，値を推測する方法である (図 1)．データ数が多くなると，問題が生じることがある．

最小二乗法 データをある特定の関数に近似して，値を推測する方法でである (図 2)．通常，関数形を決めるパラメーターよりもデータの数が多く，全てのデータを通る曲線が得られるわけではない．最も誤差が少なくなるように，関数のパラメーターを決める．

今回の講義では，補間法，最小二乗法ともに学習する．ただし，時間の関係で，最小二乗法についてはこのプリントの配布にとどめ，説明は行わない．興味のあるのは，各自学習せよ．補間法にもいろいろ有るが，ここでは最も基本的なラグランジュ補間とスプライン補間について説明する．大抵のデータプロットアプリケーションでは，最小二乗法とスプライン関数が使えるようになっている．これらの大体の内容を理解して，これらの機能を使うとデータの整理が上手になるであろう．

一般的には，補間とは得られたデータの範囲内での値の推定のことを言う．それに対して，範囲外の推定は補外と言う．[1]．例えば，図 3 のように (x_0, y_0)〜(x_3, y_3) のデータがあり，それらを通る曲線が得られたする．データの範囲 $[x_1, x_3]$ の推定を補間，それ以外を補外と言う．ただし，どちらも同じ関数を用いる．

補間に比べて，補外の方が近似の精度が悪くなる場合が多い．このことは，証券価格のグラフを考えると良く分かる．本日までのデータを補間することはそんなに難しくないが，明日以降の価格を推定することは極めて難しくなる．もし，良い精度で明日以降の価格がわかれば，大金持ちになれるであろう．

[*]秋田工業高等専門学校　電気情報工学科

[1]補間のことを内挿，補外のことを外挿ということもある．

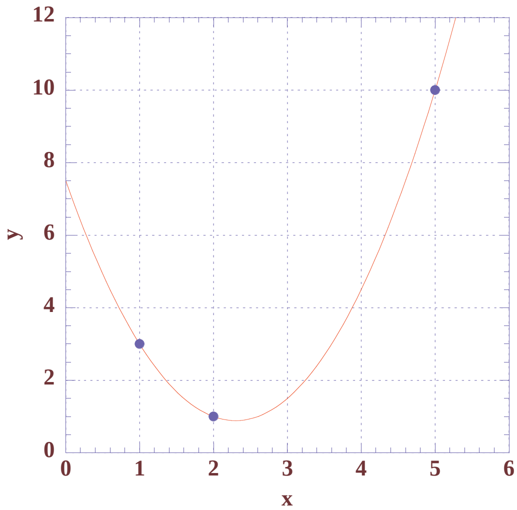

図 1: 補間 (ラグランジュ補間)

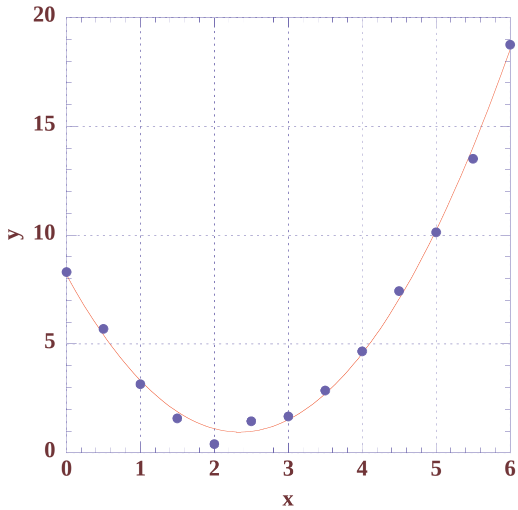

図 2: 最小二乗近似 (2 次関数)

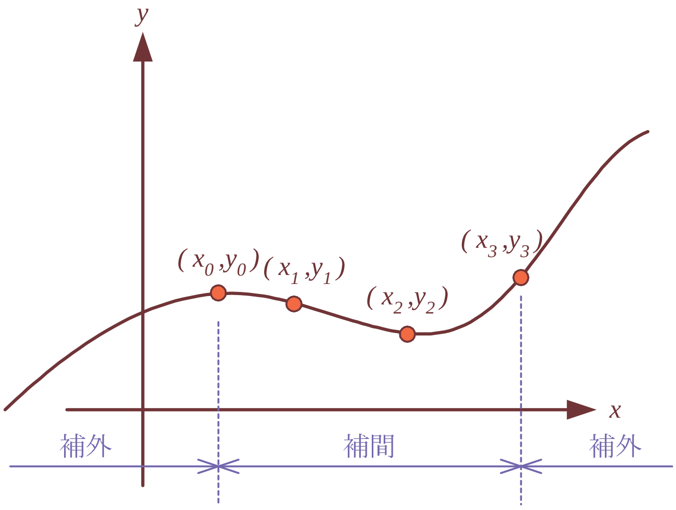

図 3: 補間と補外

2　ラグランジュ補間

2.1　基本的な考え方

ある物理量を測定して $N+1$ 個の値が得られたとする．それらは，(x_0, y_0) , (x_1, y_1), (x_2, y_2), $\cdots$, (x_N, y_N) の 2 次元座標で表すことができる．この全ての点を通る関数を求めることが補間法の課題である．N 次関数を使えばその目的が達成できると容易に分かる．データが 2 個であれば 1 次関数，3 個であれば 2 次関数というようにである．一般的に $N+1$ 個のデータの場合，

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_i x^i + \cdots + a_N x^N \tag{1}$$

と N 次関数を用いて補間するわけである．この係数 a_i を求めれば，補間の関数が求められたことになる．この係数は，N+1 元の連立 1 次方程式を解くことにより求めることができる．

この連立方程式の計算にはかなりの時間が必要であるが，それに代わるもっと良い方法がある．ここでは，N 次関数で表現できれば良いわけで，次のようにする．

$$\begin{aligned}
y = &\frac{(x-x_1)(x-x_2)(x-x_3)\cdots(x-x_N)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)\cdots(x_0-x_N)}y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)\cdots(x-x_N)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)\cdots(x_1-x_N)}y_1 \\
&+ \frac{(x-x_0)(x-x_1)(x-x_3)\cdots(x-x_N)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)\cdots(x_2-x_N)}y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)\cdots(x-x_N)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)\cdots(x_3-x_N)}y_3 \\
&\cdots + \frac{(x-x_0)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_N)}{(x_k-x_0)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_N)}y_k + \cdots \\
&+ \frac{(x-x_0)(x-x_1)(x-x_2)\cdots(x-x_{N-1})}{(x_N-x_0)(x_N-x_1)(x_N-x_2)\cdots(x_N-x_{N-1})}y_N
\end{aligned} \tag{2}$$

この式 (2) を見ると，

- 各項の分母は定数で，分子は N 次関数となっている．全ての項は N 次関数になっているので，この式は N 次関数 (N 次多項式) である．
- x に $x_0, x_1, x_2, \cdots, x_N$ を代入すると，y の値は $y_0, y_1, y_1, \cdots, y_N$ になる．したがって，この N 次関数は全てのデータ点 (x_0, y_0) , (x_1, y_1), (x_2, y_2), $\cdots$, (x_N, y_N) を通過している．

となっている．これは，表現こそ違うものの式 (1) と同じである．式 (1) の a_i を連立方程式を解くことにより補間の関数を求める必要は無く，式 (2) を使えばよいということである．この補間をラグランジュの補間多項式 (Lagrange's interpolating polynomial) という．式 (2) をもうちょっと格好良く書けば，

$$L(x) = \sum_{k=0}^{N} L_k(x) y_k \tag{3}$$

ただし，

$$
\begin{aligned}
L_k(x) &= \frac{(x-x_0)(x-x_1)(x-x_2)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_N)}{(x_k-x_0)(x_k-x_1)(x_k-x_2)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_N)} \\
&= \frac{x-x_0}{x_k-x_0}\times\frac{x-x_1}{x_k-x_1}\times\frac{x-x_2}{x_k-x_2}\times\cdots\times\frac{x-x_{k-1}}{x_k-x_{k-1}}\times\frac{x-x_{k+1}}{x_k-x_{k+1}}\times\cdots\frac{x-x_N}{x_k-x_N} \\
&= \prod_{j=0}^{N(j\neq k)}\frac{x-x_j}{x_k-x_j}
\end{aligned}
\tag{4}
$$

となる．Σ は和の記号であるが，ここで使っている $\prod$ は積の記号である．

2.2　問題点

補間の点数が増えてくると，ラグランジュの補間の近似の精度が悪くなることがある．その具体例を図 4 に示す．これから，補間の関数が振動し，端の方ではかなり精度が悪いことがわかる．ラグランジュの補間では，補間の点数が増えてくると大きな振動が発生して，もはや補間とは言えなくなることがある．ラグランジュの補間には常にこの危険性が付きまとうので，データ点数が多い場合は良い方法ではない．ほかの補間を選択しなくてはならない．

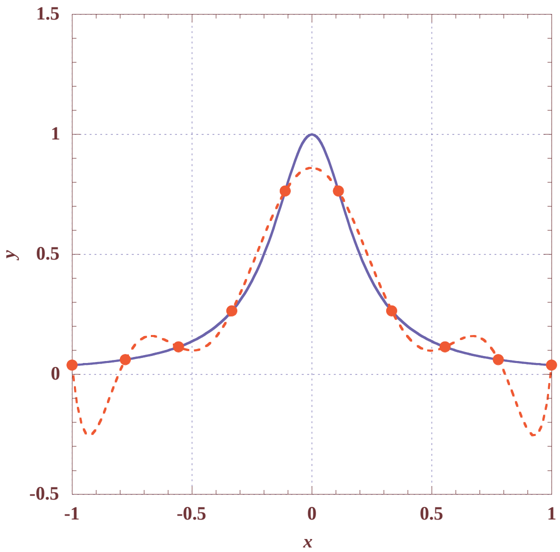

図 4: ラグランジュ補間の問題点．$y=\frac{1}{1+25x^2}$ を 10 点で補間 (点線) したが，両端で振動する．

3　スプライン補間

3.1　区分多項式

ラグランジュの補間はデータ点数が増えてくると関数が振動し，補間の精度が悪くなるのは先に述べたとおりである．そこで，補間する領域をデータ間隔 $[x_i, x_{i+1}]$ に区切り，その近傍の値を使い低次の多項式

で近似することを考える．区分的に近似関数を使うことになるが，上手に近似をしないと境界でその導関数が不連続になる．導関数が連続になるように，上手に近似する方法がスプライン補間 (spline interpolation) である．

ここでは，通常よくつかわれる 3 次のスプライン補間について説明する．3 次関数を補間に使うため，そう呼ばれている．これ以降の説明は，文献 [1] を参考にした．

データは先と同じように $(x_0, y_0)\,, (x_1, y_1),\,(x_2, y_2),\,\cdots\,,(x_N, y_N)$ とする．そして，区間 $[x_j, x_{j+1}]$ で補間に使う関数を $S_j(x)$ とする．この様子を図 5 に示す．

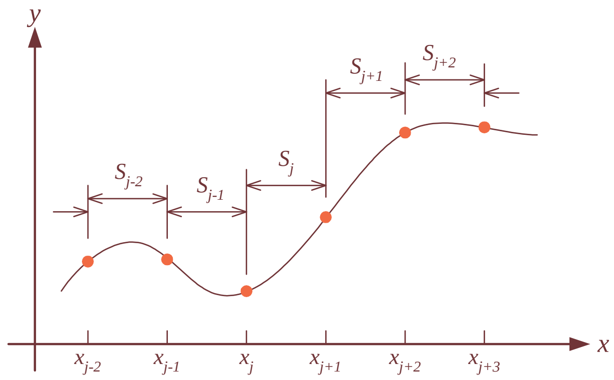

図 5: スプライン補間の区分

3.2 区分多項式の条件と計算方法

3 次のスプライン補間を考えるので，区分多項式は

$$S_j(x) = a_j(x - x_j)^3 + b_j(x - x_j)^2 + c_j(x - x_j) + d_j \qquad (j = 0, 1, 2, 3, \cdots, N-1) \tag{5}$$

となる．この a_j, b_j, c_j, d_j を求めることが，スプライン補間の中心的な問題となる．

$N+1$ 個のデータ数があるため，区分多項式は N 個ある．したがって，区分多項式の係数である未知数は $4N$ 個あることになる．これを求めるためには，$4N$ 個の方程式が必要となる．3 次のスプライン補間に以下の条件を課して，その係数を求めることにする．

[**条件 1**] 全てのデータ点を通る．各々の $S_j(x)$ に対して両端での値が決まるため，$2N$ 個の方程式ができる．

[**条件 2**] 各々の区分補間式は，境界点の 1 次導関数は連続とする．これにより，$N-1$ 個の方程式ができる

[**条件 3**] 各々の区分補間式は，境界点の 2 次導関数も連続とする．これにより，$N-1$ 個の方程式ができる．

以上の 3 つの条件を課すと $4N-2$ 個の方程式で未知数である係数の関係を表現できる．実際，未知数は $4N$ 個なので，2 個方程式が不足している．この不足を補うために，いろいろな条件が考えられるが，通常は両端 x_0 と x_N での 2 次導関数の値を 0 とする．すなわち，$S_0''(x_0) = S_{N-1}''(x_N) = 0$ である．このようにすることにより，全体の関数の形にもっとも影響を少なくすることができる．これを自然スプライン (natural spline) という．自然スプライン以外には，両端の 1 次導関数の値を指定するものもある．

これで全ての条件が決まった．あとは，この条件に満たす連立方程式を求めるだけである．このように，必要な条件が決まった場合，$4N$ 個の未知数 a_j, b_j, c_j, d_j を既知の x_j, y_j を使って連立方程式を作るのが普通である．これも可能であるが，少し手間を省くために，

1. これら $4N$ 個の未知数を x_j と y_j，さらに $x = x_j$ における 2 次導関数の値を u_j で表現する．

$$u_j = S''_k(x_j) \tag{6}$$
$$\text{ただし，}\quad j = 0, 1, 2, \cdots, N$$
$$k = j-1, j$$

2. u_j が満たす連立方程式を作り，u_j を解く．

3. 既知の x_j と y_j と連立方程式により求められた u_j を用いて，区分多項式の係数 a_j, b_j, c_j, d_j を計算する．

というアプローチで問題を解く．ここで，本当に未知数 (a_j, b_j, c_j, d_j) を (x_j, y_j, u_j) で表現することができるのか—という疑問が湧く者もいるだろう．これは，先の連立方程式を作る条件を上手に使うことにより可能なのである．また，このような方法は，問題解決の遠回りをしているように思えるが，以降の説明を見ると実際にはかなり簡潔になることがわかるので我慢して読んで欲しい．

3.2.1 係数の表現

b_j の表現　これは，$u_j = S''_j(x_j)$ から求めることができる．式 (5) から，

$$S''_j(x) = 6a_j(x - x_j) + 2b_j \tag{7}$$

となる．$x = x_j$ の時，これは u_j となるので，

$$b_j = \frac{u_j}{2} \tag{8}$$

が直ちに導かれる．これで，b_j を u_j で表現できたことになる．b_j を表現するためには，x_j と y_j を使っても良かったが，ここではたまたま必要なかったのである．

a_j の表現　これは，2 次導関数が区分多項式の境界で等しいという条件から導くことができる．先ほどは区分多項式の左端 x_j を考えた．次に右端 $x = x_{j+1}$ を考えることにする．右端の導関数は，

$$u_{j+1} = S''_j(x_{j+1}) = 6a_j(x_{j+1} - x_j) + 2b_j \qquad (j = 0, 1, 2, \cdots, N-1) \tag{9}$$

となる．これから，a_j は

$$\begin{aligned} a_j &= \frac{u_{j+1} - 2b_j}{6(x_{j+1} - x_j)} \\ &= \frac{u_{j+1} - u_j}{6(x_{j+1} - x_j)} \qquad (j = 0, 1, 2, \cdots, N-1) \end{aligned} \tag{10}$$

と導くことができる．これで，a_j を x_j と y_j と u_j で表現できたことになる．

d_j の表現　これは，区分多項式は全てのデータ点上を通過するという条件から導くことができる．まずは，多項式 $S_j(x)$ の左端 x_j を考える．ここでは，$S_j(x_j)=y_j$ なので，式 (5) に代入すると

$$d_j = y_j \tag{11}$$

が直ちに導ける．これで，d_j を y_j で表現できたことになる．d_j の表現には，x_j と u_j はたまたま不要であった．

c_j の表現　これもまた，区分多項式は全てのデータ点上を通過するという条件から導くことができる．今度は，多項式 $S_j(x)$ の右端 x_{j+1} である．ここでは，$S_j(x_{j+1})=y_{j+1}$ なので，式 (5) に代入すると

$$a_j(x_{j+1}-x_j)^3 + b_j(x_{j+1}-x_j)^2 + c_j(x_{j+1}-x_j) + d_j = y_{j+1} \tag{12}$$

が導ける．式 (8),(10),(11) を用いると，

$$\begin{aligned} c_j &= \frac{1}{x_{j+1}-x_j}\left[y_{j+1} - a_j(x_{j+1}-x_j)^3 - b_j(x_{j+1}-x_j)^2 - d_j\right] \\ &= \frac{1}{x_{j+1}-x_j}\left[y_{j+1} - \frac{u_{j+1}-u_j}{6(x_{j+1}-x_j)}(x_{j+1}-x_j)^3 - \frac{u_j}{2}(x_{j+1}-x_j)^2 - y_{\mathrm{j}}\right] \\ &= \frac{y_{j+1}-y_j}{x_{j+1}-x_j} - \frac{1}{6}(x_{j+1}-x_j)(2u_j + u_{j+1}) \end{aligned} \tag{13}$$

となる．これで，a_j を x_j と y_j と u_j で表現できたことになる．

以上で，a_j と b_j，c_j，d_j が x_j と y_j，u_j で表せたことになる．x_j と y_j はデータ点なので，既知である．したがって，u_j が分かれば，補間に必要な係数が全て分かるのである．また，連立方程式の [条件 1] と [条件 3] も満たしている．従って，[条件 2] を満たすように u_j を決めれば良いことになる．すると全ての区分多項式の係数が分かるのである．

3.2.2　連立方程式

先に述べたように u_j は，1 次導関数が境界点で等しいという条件から決める．次の式を使うのである．

$$S'(x_{j+1}) = S'_j(x_{j+1}) = S'_{j+1}(x_{j+1}) \qquad (j=0,1,2,\cdots,N-2) \tag{14}$$

これと式 (5) から，

$$3a_j(x_{j+1}-x_j)^2 + 2b_j(x_{j+1}-x_j) + c_j = c_{j+1} \tag{15}$$

となる．あとは，この式の a_j と b_j，c_j を x_j と y_j，u_j で表して，u_j の連立方程式にするだけである．最終的に式は，

$$(x_{j+1}-x_j)u_j + 2(x_{j+2}-x_j)u_{j+1} + (x_{j+2}-x_{j+1})u_{j+2} = 6\left[\frac{y_{j+2}-y_{j+1}}{x_{j+2}-x_{j+1}} - \frac{y_{j+1}-y_j}{x_{j+1}-x_j}\right] \tag{16}$$

となる．この方程式は，$j=0,1,2,\cdots,N-2$ で成り立つ．したがって，式の数は，$N-1$ 個である．u_j の数は $N+1$ 個であるが，$u_0=u_N=0$ としたので，未知の u_j は $N-1$ 個である．式 (16) を解くことにより，全ての u_j が決定できる．これが決まれば，a_j と b_j，そして c_j が計算できる．

$u_0 = u_N = 0$ を代入した連立 1 次方程式は，

$$
\begin{pmatrix}
2(h_0+h_1) & h_1 & & & & & \\
h_1 & 2(h_1+h_2) & h_2 & & & \text{\Large 0} & \\
& h_2 & 2(h_2+h_3) & h_3 & & & \\
& & & \ddots & & & \\
& & & h_{j-1} & 2(h_{j-1}+h_j) & h_j & \\
& \text{\Large 0} & & & & \ddots & \\
& & & & & h_{N-2} & 2(h_{N-2}+h_{N-1})
\end{pmatrix}
\begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_j \\ \vdots \\ u_{N-1} \end{pmatrix}
=
\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_j \\ \vdots \\ v_{N-1} \end{pmatrix}
\tag{17}
$$

となる．ただし，h_j と v_j は以下のとおり．

$$
h_j = x_{j+1} - x_j \qquad (j = 0, 1, 2, \cdots, N-1) \tag{18}
$$

$$
v_j = 6\left[\frac{y_{j+1} - y_j}{h_j} - \frac{y_j - y_{j-1}}{h_{j-1}}\right] \qquad (j = 1, 2, \cdots, N-1) \tag{19}
$$

3.2.3 区分多項式の決定

いうまでもないと思うが，スプライン補間を行う区分多項式 (5) は，以下の手順で求める．

1. 連立方程式 (17) を計算して，各点での二次の導関数の値 u_j を求める．
2. 区分多項式の係数 a_j と b_j，c_j，d_j を式 (10), (8),(13),(11) を計算する．

区分多項式が分かれば，補間の値計算できる．

4 最小二乗法

最小二乗法を感覚的に理解するために，最初に一次関数の最小二乗法を示す．その後，一般的な線形最小二乗法について説明する．非線形最小二乗法は難しいので，範囲外とする．

4.1 簡単な例

最小二乗法というのは，データをある関数で最良近似する方法である．例えば，

$$(1.2,\ 2.2)\quad (2.1,\ 3.8)\quad (3.3,\ 5.6)\quad (4.1,\ 7.1)\quad (5, 8.8) \tag{20}$$

の (x, y) の実験データがあるとする．これを直線で近似 $y = ax + b$ したい．どうすればよいか?—という問題である．誤差の二乗が最小になる直線が最良近似とすることができる．これを最小二乗法 (least squares method) と言う．式で表すと，誤差の二乗の和 $E(a, b)$ は，

$$E(a,b) = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \tag{21}$$

となる．$(x_i,\ y_i)$ は，i 番目のデータで，N はデータの個数である．この誤差が最小になる a と b を捜す．式 (21) は a にも b にも 2 次式でその係数は正の値なので最小値がある．誤差 E の最小値は，それぞれ偏微分した値がゼロとなるときに得ることができる．

$$\frac{\partial E}{\partial a} = -\sum_{i=1}^{N} 2(y_i - ax_i - b)x_i = 0 \qquad \frac{\partial E}{\partial b} = -\sum_{i=1}^{N} 2(y_i - ax_i - b) = 0 \tag{22}$$

これは，a と b の連立方程式である．すなわち，

$$\left\{ \begin{aligned} a\sum_{i=1}^{N} x_i^2 + b\sum_{i=1}^{N} x_i &= \sum_{i=1}^{N} x_i y_i \\ a\sum_{i=1}^{N} x_i + nb &= \sum_{i=1}^{N} y_i \end{aligned} \right. \tag{23}$$

である．これを解くと

$$a = \frac{N\sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2} \qquad b = \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i y_i \sum_{i=1}^{N} x_i}{N\sum_{i=1}^{N} x_i^2 - (\sum_{i=1}^{N} x_i)^2} \tag{24}$$

となる．

最初に示したのデータについて計算してみると，$a = 1.452119$, $b = 0.708006$ となる．ゆえに，最小二乗法による 1 次関数は

$$y = 1.452119x + 0.708006 \tag{25}$$

となる．データをこの間数をプロットすると，図 6 のようになる．

グラフ作成ソフトウェアーは，最小二乗法によるデータのフィッティングをサポートしているものが多い．Excel でも可能なはずである．また，私の WEB ページでは web_gnuplot[2] と称して，最小二乗法が web 上で使えるようになっている．実験データの整理に使うと良い．というか，実験データの整理に最小二乗法を使わないことはあり得ない．

---

[2] http://akita-nct.jp/yamamoto/comp/graph/web_gnuplot/index.php

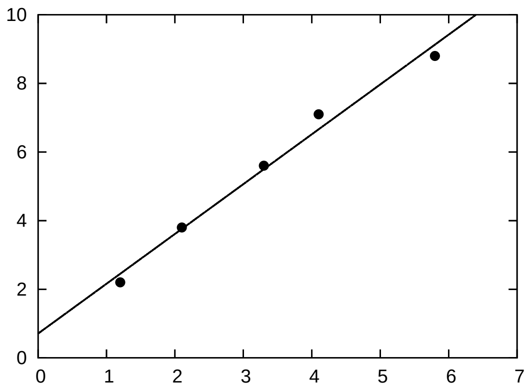

図 6: データを最小二乗法でフィット

ここでは，偏微分により最小二乗法の式を導いたが，線形代数の部分空間への射影を考える方が簡単である．これについては，参考文献 [2] に詳しく書いてある．これは良い教科書なので，一読を勧める．

4.2　線形最小二乗法

一次関数の例を拡張して，線形最小二乗法の説明を行う．ただ，ここの線形最小二乗法の説明は，結構いい加減なところもある．また，線形最小二乗法に適した連立方程式の解き方も述べていない．最小二乗法について詳細に知りたければ，文献 [3] が大いに参考になる．本格的な最小二乗法のプログラムを作成するときには，この文献を見ることを勧める．

データをフィットする線形最小二乗法の関数は，

$$y(x)=\sum_{k=1}^{M}a_kf_k(x) \tag{26}$$

である．ここで，$f_k(x)$ は基底関数 (base function) と呼ばれるもので，M はその個数である．基底関数は，どのような関数でも良い．例えば，

$$y(x)=a_1+a_2x+a_3x^3+a_4\sin(x)+a_4\cos^2(x^2) \tag{27}$$

とすることができる．

ここでは，得られたデータ点 $(x_1,\,y_1),\,(x_2,\,y_2),\,(x_2,\,y_2),\cdots,(x_N,\,y_N)$ に最も近い係数 a_k を求めることが問題となる．近いというのは，一次関数の例と同じように，誤差の二乗和が最小になることを言う．式 (26) の場合，誤差の二乗は，

$$E(a_1,\,a_2,\,a_3,\cdots,a_M)=\sum_{i=1}^{N}\left[y_i-\sum_{k=1}^{M}a_kf_k(x_i)\right]^2 \tag{28}$$

となる．

誤差の二乗和 $E(a_1, a_2, a_3, \cdots, a_M)$ が最小になるように決めれられた係数 $a_1, a_2, a_3, \cdots, a_M$ が最良近似となる．これは，それぞれの係数で偏微分した値がゼロになるようにすれば良い．一次関数の時と全く同じだ! 式で表すと次のようになる．

$$\begin{cases} \dfrac{\partial E}{\partial a_1} = 2\displaystyle\sum_{i=1}^{N} f_1(x_i)\left[y_i - \sum_{k=1}^{M} a_k f_k(x_i)\right] = 0 \\ \dfrac{\partial E}{\partial a_2} = 2\displaystyle\sum_{i=1}^{N} f_2(x_i)\left[y_i - \sum_{k=1}^{M} a_k f_k(x_i)\right] = 0 \\ \dfrac{\partial E}{\partial a_3} = 2\displaystyle\sum_{i=1}^{N} f_3(x_i)\left[y_i - \sum_{k=1}^{M} a_k f_k(x_i)\right] = 0 \\ \qquad\qquad\vdots \\ \dfrac{\partial E}{\partial a_M} = 2\displaystyle\sum_{i=1}^{N} f_M(x_i)\left[y_i - \sum_{k=1}^{M} a_k f_k(x_i)\right] = 0 \end{cases} \tag{29}$$

これをよく見ると，次のような連立方程式に書き直すことができる．

$$\begin{cases} \displaystyle\sum_{i=1}^{N} f_1(x_i)f_1(x_i)a_1 + \sum_{i=1}^{N} f_1(x_i)f_2(x_i)a_2 + \cdots + \sum_{i=1}^{N} f_1(x_i)f_M(x_i)a_M = \sum_{i=1}^{N} f_1(x_i)y_i \\ \displaystyle\sum_{i=1}^{N} f_2(x_i)f_1(x_i)a_1 + \sum_{i=1}^{N} f_2(x_i)f_2(x_i)a_2 + \cdots + \sum_{i=1}^{N} f_2(x_i)f_M(x_i)a_M = \sum_{i=1}^{N} f_2(x_i)y_i \\ \qquad\qquad\vdots \\ \displaystyle\sum_{i=1}^{N} f_M(x_i)f_1(x_i)a_1 + \sum_{i=1}^{N} f_M(x_i)f_2(x_i)a_2 + \cdots + \sum_{i=1}^{N} f_M(x_i)f_M(x_i)a_M = \sum_{i=1}^{N} f_M(x_i)y_i \end{cases} \tag{30}$$

この連立方程式を解いて，$a_1, a_2, a_3, \cdots, a_M$ を求めれば，データをフィットする関数の式 (26) の最適係数が求まったことになる．

連立方程式 (30) の解法については，文献 [3] が詳しい．とりあえずガウス・ジョルダン法で解いてみて，ダメ (特異に近い) ならば，この文献を参考にしてプログラムを作成せよ．特異値分解 (singular value decomposition, SVD) が良いようである．

5 練習問題

[練習 1] 3 点 $(1,1), (3,2), (4,5)$ を通る曲線をラグランジュ補間して，区間 $[-1,5]$ のグラフを書け．

[練習 2] 以下のようなデータが得られた．

x	-1.0	-0.9	-0.8	-0.7	-0.6	-0.5	-0.4	-0.3	-0.2	-0.1	0.0
y	0.038	0.047	0.058	0.075	0.100	0.137	0.200	0.307	0.500	0.800	1.000

x	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
y	0.800	0.500	0.307	0.200	0.137	0.100	0.075	0.058	0.047	0.038	

このデータは，関数 $y = 1/(1 + 25x^2)$ の値である．21 個の点があるが，ここでは補間をするデータの個数と補間のグラフの関係を調べる．

– 等間隔の 6 個のデータ，すなわち

$$x = -1.0, -0.6, -0.2, 0.2, 0.6, 1.0$$

の場合のラグランジュ補間の結果をグラフに示せ．

– 等間隔の 11 個のデータ，すなわち

$$x = -1.0, -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$$

の場合のラグランジュ補間の結果をグラフに示せ．

– 全てのデータを用いた場合のラグランジュ補間の結果をグラフに示せ．

以上の結果から，補間の点数が増加した場合，ラグランジュ補間ではどのようなことが起きるか?–考察せよ．

[**練習 3**] 先の [練習 2] と同じことをスプライン補間で行え．

[**練習 4**] WEB にデータを掲載する．それを二次関数で最小二乗近似せよ．

6 レポート

6.1 内容

少なくとも，以下の問いを行いレポートとして提出すること．これ以上のことを実施しても良い．

[**問 1**] 練習問題の [練習 1] を実施し，プログラムとグラフを提出せよ．

[**問 2**] 練習問題の [練習 3] を実施し，プログラムとグラフを提出せよ．

6.2　提出要領

期限	12 月 20 日 (木)24:00 まで
用紙	A4
提出場所	山本研究室の入口のポスト
表紙	表紙を 1 枚つけて，以下の項目を分かりやすく記述すること．

授業科目名「計算機応用」
課題名「補間法と最小二乗法」
5E　学籍番号　氏名
提出日

参考文献

[1] 高橋大輔. 数値計算. 岩波書店, 1996.

[2] Gilbert Strang. 線形代数とその応用. 産業図書株式会社, 1992.

[3] Willam H. Press et al. NUMERICAL RECIPES in C [日本語版]. 技術評論社, 1996.