電気 303 / 電情 303 数値解析 (9)

# 関数近似 (3)
# 最小二乗近似

## 今回の講義の典拠

- 斎藤, 数値解析入門, 東京大学出版会, 2012
- 杉原, 室田, 数値計算法の数理, 岩波書店, 1994
- 森, 数値解析, 第 2 版, 共立出版, 2002.
- J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, 3/e, Springer, 2002.

# 最良近似多項式

- 有界閉区間 $[a, b]$ で定義された連続関数 $f$ を多項式 $p$ によって区間全体でなるべくうまく近似したい場合には，「有限の標本点で $f(x)$ と $p(x)$ を一致させる」のは必ずしも良い方法ではない．

- $\|\cdot\|$ を何らかのノルムとしたとき,

$$\|f - p\|$$

を最小とするような多項式を求めた方が良いこともある.

- $P_n$ を $n$ 次の多項式の全体とする.

$$p^* = \arg\min_{p \in P_n} \|f - p\|$$

を, $f$ の ノルム $\|\cdot\|$ に関する $n$ 次の**最良近似多項式**という.

- 記号 $\arg\min$ や $(\arg\max)$ は, 与えられた評価関数が最小 (最大) となる変数をあらわす記号であり, 転じて「所与の評価関数に関する最小化 (最大化) 問題を解いて, 得られる開を $p^*$ とせよ」という意味で用いられる.

- 最小化問題

$$p^* = \arg\min_{p \in P_n} \|f - p\|$$

では, $\|f - p\|$ が評価関数である. また, この問題では, 「変数」として動くのは多項式 $p$ である (多項式 $p$ が $n$ 次の多項式全体の集合 $P_n$ の中を動く).

- 最良近似多項式を定めるために用いられるノルムは, 典型的には, 無限大ノルム ($\|\cdot\|_\infty$) と $L_2$ ノルムである.

- ノルムを $L_2$ ノルム ($\|\cdot\|_2$) に取ったとき, これに関する $n$ 次の最良近似多項式を $n$ 次の**最小二乗近似**という.

# 最小二乗近似

- $L_2$ ノルムに関する最良近似多項式, すなわち最小二乗近似は, 計算が比較的簡単であり, 応用上もよく用いられる. 以下, これについて解説する.

- 有界閉区間 $[a, b]$(ただし $b > a$) において連続な実数値関数 $f$ を最小二乗近似する $n$ 次の多項式を構成する問題を考える.

- 関数空間自体は無限次元空間であるが, $n$ 次の多項式の全体は, その有限次元 ($n+1$ 次) の部分空間であり, 基底として,

$$\{1, x, \dots, x^n\}$$

が取れる. ただし, ここで言う「基底」は関数空間の基底であり, 関数である.

- 上述の基底を用いると, $f$ を最良近似する 問題は,

$$p(x) = \sum_{k=0}^{n} \alpha_k x^{n-k}$$

に関し, $p(x)$ が $f$ の最良近似となるような

$$\{\alpha_0, \alpha_1, \ldots, \alpha_n\}$$

を求める問題に変換される.

- $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_n)^T$, $\boldsymbol{v}(x) = (1, x, \ldots, x^n)^T$ と定義すると,

$$p(x) = \boldsymbol{\alpha}^T \boldsymbol{v}(x)$$

と書ける.

- 関数 $f$ と $g$ の内積を

$$(f,g) = \int_a^b f(x)g(x)\,dx$$

とする．実数値関数を考えているので複素共役は不要である．

- $g_{kl} = (x^k, x^l)$, $h_k = (f, x^k)$ とする $(0 \leq k, l \leq n)$. $k$ および $l$ の値が零から始まることに注意せよ. 内積の定義より,

$$g_{kl} = \int_a^b x^{k+l}\,dx,$$
$$h_k = \int_a^b f(x)x^k\,dx,$$

であることに注意する.

- $g_{k.l}$ を第 $(k+1, l+1)$ 要素として持つ正方行列を作り, これを $\boldsymbol{G}$ と書く. 定義から, $\boldsymbol{G}$ は対称行列である.

- 任意の零でない多項式 $p$ に対し,

$$\int_a^b (p(x))^2\,dx > 0$$

であるが, これに $p(x) = \boldsymbol{\alpha}^T \boldsymbol{v}(x)$ を代入すると, $\boldsymbol{\alpha} \neq \boldsymbol{0}$ であれば,

$$\int_a^b \boldsymbol{\alpha}^T \boldsymbol{v}(x)(\boldsymbol{v}(x))^T \boldsymbol{\alpha}\,dx = \boldsymbol{\alpha}^T \boldsymbol{G} \boldsymbol{\alpha} > 0$$

となることがわかる.

- よって, $\boldsymbol{G}$ は正定対称行列である.

- $\|f-p\|_2^2 = (f-p, f-p)$ を計算すると,

$$
\begin{aligned}
&\|f-p\|_2^2 \\
&\quad = (f,f) - 2\sum_{k=0}^{n} \alpha_k h_k + \sum_{k=0}^{n}\sum_{l=0}^{n} \alpha_k \alpha_l g_{kl}
\end{aligned}
$$

となる.

- $\boldsymbol{h} = (h_0, \ldots, h_n)^T$ とおくと, 上の式は以下のように書き直される.

$$\|f - p\|_2^2 = (f, f) - 2\boldsymbol{h}^T\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\boldsymbol{G}\boldsymbol{\alpha}.$$

- 評価関数 $\|f-p\|_2^2$ の $\boldsymbol{\alpha}$ による偏微分が零となる点, すなわち

$$\frac{\partial\|f-p\|_2^2}{\partial\boldsymbol{\alpha}} = \mathbf{0}$$

を満たす $\boldsymbol{\alpha}_*$ は評価関数 $\|f-p\|_2^2$ の停留点を与えるが, $\boldsymbol{G}$ が正定対称行列だったから, $\|f-p\|_2^2$ は $\boldsymbol{\alpha}$ に関する正値の二次形式で, したがって $\boldsymbol{\alpha}_*$ によって 評価関数 $\|f-p\|_2^2$ は最小値を達成する.

- したがって, $\boldsymbol{\alpha}_*^T \boldsymbol{v}(x)$ が $f$ の最小二乗の意味で最良近似多項式となる．しばらく, これを最小二乗近似多項式と呼ぶことにする.

- この偏微分を計算すると,

$$\frac{\partial \|f - p\|_2^2}{\partial \boldsymbol{\alpha}} = -\boldsymbol{h}^T + \boldsymbol{\alpha}^T \boldsymbol{G}$$

となる.

- したがって, 連立一次方程式

$$-\boldsymbol{h}^T + \boldsymbol{\alpha}^T \boldsymbol{G} = \boldsymbol{0}^T$$

を解くことにより, $f$ の最小二乗近似多項式が得られる.

- $\boldsymbol{G}$ は対称行列だったから, 先の式全体を転置して整理すると,

$$\boldsymbol{G\alpha} = \boldsymbol{h}$$

が得られる．この連立一次方程式を正規方程式という

- 正規方程式という言葉は上記とは異なる文脈で用いられることがあり，その場合は定義が変わるので注意せよ．これについては後で述べる．

- まとめると，正規方程式を解き，その解を係数として持つ $n$ 次多項式を作ると，その多項式が $f$ の最小二乗近似多項式になる．

- ここまでは, $n$ 次の多項式を, 基底

$$\boldsymbol{v}(x) = \{1, x, \ldots, x^n\}$$

によって表現していた. この基底は直感的にはわかりやすいのだが…

- 数値計算の誤差の影響を考慮すると, この基底は必ずしも好ましくない.

- その理由を説明するために, 条件数という概念を導入しておく.
- 正則行列 $\boldsymbol{A}$ に対し, $\|\boldsymbol{A}\|\|\boldsymbol{A}^{-1}\|$ を $\boldsymbol{A}$ の 条件数という (ノルムの取り方には自由度がある).
- 条件数が大きい行列では, 連立一次方程式の解の誤差が大きくなることがあり得る.

- 基底 $\boldsymbol{v}(x) = \{1, x, \ldots, x^n\}$ を用いた場合, 区間 $[a, b]$ における低次の多項式の積分と高次の多項式を積分では, その絶対値が大きく異なるため, $\boldsymbol{G}$ の条件数が大きくなりがちである.

- たとえば, Mathematica で計算したところ,
  - ▷ $a=0$, $b=1$, $n=10$ とした場合:
    $\boldsymbol{G}$ の条件数は $10^{29}$ のオーダー
  - ▷ $a=0$, $b=10$, $n=10$ とした場合:
    $\boldsymbol{G}$ の条件数は $10^{47}$ のオーダー

  となった.

- この例から，応用上は，多項式の次数が高い場合には，基底

$$\boldsymbol{v}(x) = \{1, x, \ldots, x^n\}$$

は必ずしも好ましくないことが伺い知れる．

- 多項式の集合

$$\{r_0(x), \ldots, r_n(x)\}$$

が内積 $(f, g)$ に関して正規直交基底であれば, $\boldsymbol{G}$ は単位行列となり, 正規方程式は簡単に解ける.

- 正規性までは要求しなくても, 多項式の集合

$$\{r_0(x), \ldots, r_n(x)\}$$

が内積 $(f, g)$ に 関して 直交していて,

$$\forall k, (r_k, r_k) > 0$$

であれば, $\boldsymbol{G}$ が正則な 対角行列となるから, やはり正規方程式は容易に解ける.

- 上述のような条件を (おおむね) 満たす多項式系を, 直交多項式と呼ぶ.
- 直交多項式系を用いることで, 正規方程式は簡単に解けるようになる.

- 今までの議論では, 内積として

$$(f,g) = \int_a^b f(x)g(x)\,dx$$

を考えていたのだが…

- 内積として, もっと一般的なものを考えることがある.

- すなわち, 重み関数と呼ばれる 関数 $w(x)$ を用いて,

$$(f,g)_w = \int_a^b f(x)g(x)w(x)\,dx$$

を内積として用いる. これを重み付き内積と呼ぶ.

- 重み関数 $w(x)$ には,
  - ▷ 非負の連続関数
  - ▷ 有限個の点を除いて零にならない
  - ▷ $\displaystyle\int_a^b w(x)dx$ が有限

という条件が課される.

- $(f, g)_w$ が内積の公理を満たすことは, 定義から直ちに確かめることができる.
- $(f, g)$ は, $(f, g)_w$ において $w(x)$ を恒等的に値が 1 の関数に取った特別な場合である.

- $\|f-p\|_w = \sqrt{(f-p, f-p)_w}$ と定義する.

- $\{r_0(x), \ldots, r_n(x)\}$ を $n$ 次多項式全体がなす部分空間の基底とし,

$$p(x) = \sum_{k=0}^{n} \alpha_k r_k(x)$$

とする.

- $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_n)^T, \boldsymbol{v}(x) = (r_0(x), \ldots, r_n(x))^T$ とする. すると,

$$p(x) = \boldsymbol{\alpha}^T \boldsymbol{v}(x)$$

と書ける.

- 区間 $[a,b]$ で定義された連続関数 $f(x)$ を, $p(x)$ によって,

$$\|f-p\|_w^2$$

に関して最小二乗近似する問題を考える. これを, 重み付き最小二乗近似と呼ぶ.

- $\gamma_{kl} = (r_k, r_l)_w$, $\eta_k = (f, r_k)_w$ とする $(0 \leq k, l \leq n)$. 内積の定義より,

$$\gamma_{kl} = \int_a^b x^{k+l} w(x)\, dx,$$

$$\eta_k = \int_a^b f(x) x^k w(x)\, dx,$$

であることに注意する.

- $\gamma_{k,l}$ を第 $(k+1, l+1)$ 要素として持つ正方行列を作り, これを $\mathbf{\Gamma}$ と書く. 定義から, $\mathbf{\Gamma}$ は対称行列である.

- $p$ を零でない多項式としたとき, 区間 $[a,b]$ において $(p(x))^2w(x)$ が零となる点は高々有限個だから,

$$\int_a^b (p(x))^2 w(x)\,dx > 0$$

である.

- 内積が $(\cdot,\cdot)$ であった場合と同様に, 前ページの式に

$$p(x) = \boldsymbol{\alpha}^T \boldsymbol{v}(x)$$

を代入すると, $\boldsymbol{\alpha} \neq \mathbf{0}$ であれば,

$$\int_a^b \boldsymbol{\alpha}^T \boldsymbol{v}(x)(\boldsymbol{v}(x))^T \boldsymbol{\alpha} w(x)\, dx = \boldsymbol{\alpha}^T \boldsymbol{\Gamma} \boldsymbol{\alpha} > 0$$

となることがわかる.

- よって, $\mathbf{\Gamma}$ は正定対称行列である.

- $\|f-p\|_w^2 = (f-p, f-p)_w$ を計算すると,

$$
\begin{aligned}
&\|f-p\|_w^2 \\
&= (f,f)_w - 2\sum_{k=0}^{n} \alpha_k \eta_k + \sum_{k=0}^{n}\sum_{l=0}^{n} \alpha_k \alpha_l \gamma_{kl}
\end{aligned}
$$

となる.

- $\boldsymbol{\eta} = (\eta_0, \ldots, \eta_n)^T$ とおくと, 上の式は以下のように書き直される.

$$\|f - p\|_w^2 = (f, f)_w - 2\boldsymbol{\eta}^T \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \boldsymbol{\Gamma} \boldsymbol{\alpha}.$$

- 以下は, $(\cdot, \cdot)$ に関する議論の繰り返しになるが…

- 評価関数 $\|f-p\|_w^2$ の $\boldsymbol{\alpha}$ による偏微分が零となる点, すなわち
$$\frac{\partial\|f-p\|_w^2}{\partial\boldsymbol{\alpha}}=\mathbf{0}$$
を満たす $\boldsymbol{\alpha}_*$ は評価関数 $\|f-p\|_w^2$ の停留点を与えるが, $\boldsymbol{\Gamma}$ が正定対称行列だったから, $\|f-p\|_w^2$ は $\boldsymbol{\alpha}$ に関する正値の二次形式で, したがって $\boldsymbol{\alpha}_*$ によって 評価関数 $\|f-p\|_w^2$ は最小値を達成する.

- したがって, $\boldsymbol{\alpha}_*^T \boldsymbol{v}(x)$ が $f$ の重み付き最小二乗近似多項式を与える.

- 上述の偏微分の計算は $(\cdot,\cdot)$ の場合と同一であり, 結果として, この場合も, 正規方程式

$$\mathbf{\Gamma}\boldsymbol{\alpha} = \boldsymbol{\eta}$$

の解 $\boldsymbol{\alpha}_*$ を用い,

$$p(x) = \boldsymbol{\alpha}_*^T \boldsymbol{v}(x)$$

とすることで, $f$ の重み付き最小二乗近似多項式が得られることがわかる.

# 直交多項式系

- 基底が直交していれば正規方程式が簡単に解けることは既に述べた.

- 関数系 $(r_0(x), r_1(x), \ldots)$ において $k \neq l$ なら $(r_k, r_l)_w = 0$ となるとき, これを重み $w$ に関する直交関数系という (岩波数学辞典).
- 上記の直交関数系の定義は $\forall k, (r_k, r_k)_w = 0$ となる場合を許容しており不合理であるが, この用語はすでに定着しているようである.

- Fourier 級数が使いやすい理由のひとつに,

$$\{1, \sin x, \cos x, \sin 2x, \cos 2x, \ldots\}$$

が直交関数系になっていることが挙げられる.

- 多項式の列 $(r_0(x), r_1(x), \ldots)$ が直交関数系になっていて，かつ $\forall k,\ (r_k, r_k)_w > 0$ となるとき，これを**重み $w$ に関する直交多項式系**という．

- ある多項式系が直交するか否かは, 積分する区間や重み関数に依存する. このため, 直交多項式系は, その関数を利用する区間および重み関数と組み合わせて定められる.

- 直交多項式系としては色々なものが知られている．この講義では代表的なものをいくつか紹介するが，網羅的ではないので注意せよ．

# Legendre 多項式

- 以下の式によって定められる多項式を考える:

$$p_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} (x^2 - 1)^n, \quad n = 0, 1, 2, \ldots.$$

- 多項式系

$$\{p_n(x) : n \in \mathbb{N}\}$$

において, $p_n(x)$ は $n$ 次の多項式である．これを, Legendre 多項式と呼ぶ.

- Legendre 多項式の最初の数項は,

$$
\begin{aligned}
p_0(x) &= 1, \\
p_1(x) &= x, \\
p_2(x) &= \frac{3}{2}x^2 - \frac{1}{2}
\end{aligned}
$$

である.

- Legendre 多項式は, 区間 $[-1, 1]$ において, 重みを 1 として利用される. よって, $p_k$ と $p_j$ の内積は, 次のように定義される.

$$(p_k, p_j) = \int_{-1}^{1} p_k(x) p_j(x) dx$$

- この講義では証明を述べないが, 以下が成り立つ:
  - ▷ $k \neq j$ であれば $(p_k, p_j) = 0$
  - ▷ $(p_k, p_k) = \dfrac{2}{2k+1}$

- したがって, Legendre 多項式系は, 区間 $[-1, 1]$ における重み 1 に関する直交多項式系になっている (ただし正規化はされていない).

- 上記の事実の証明は [斎藤], pp. 157–158 を参照せよ．証明にガンマ関数やベータ関数が必要となり，繁雑なので，この講義では証明を省略する．

- $p_k$ が $k$ 次の多項式であることと, 直交化されていることから, Legendre 多項式系は,

$$\{1, x, x^2, \ldots\}$$

を区間 $[-1, 1]$ において重み 1 に関して直交化したものに一致する.

# Laguerre 多項式

- Laguerre 多項式 (系) は, 以下の式によって定められる多項式 $l_n(x)$ を集めたものである.

$$l_n(x) = \frac{e^x}{n!}\frac{d^n}{dx^n}\left(e^{-x}x^n\right), \quad n = 0, 1, 2, \ldots.$$

量子力学の分野で利用される.

- Leguerre 多項式が多項式であることは一見では明らかではないが, 定義式の右側の $e^{-x}$ は何回微分しても消えることはないので, 最終的に定義式の左側の $e^{x}$ と打ち消し合い, 多項式だけが残る.

- Leguerre 多項式は区間 $[0, \infty)$ で, 重みを $e^{-x}$ として利用される. $l_k$ と $l_j$ の 内積は次のように定義される.

$$(l_k, l_j) = \int_0^\infty l_k(x) l_j(x) e^{-x} dx$$

- この講義では証明を述べないが, 以下が成り立つ.
    - ▷ $k \neq j$ であれば $(l_k, l_j) = 0$
    - ▷ $(l_k, l_k) = \dfrac{1}{(k!)^2}$

- 証明は [斎藤], pp. 278–279 を参照.

# Chebychev 多項式

- Chebychev 多項式 (系) は, 以下の式によって定められる多項式 $t_n(x)$ を集めたものである ([杉原, 室田]):

$$t_n(x) = \cos(n\ \arccos x), \quad n = 0, 1, 2, \ldots .$$

- Chebychev「多項式」が実際に多項式になっていることは一見して明らかでないので, これを確認する.

- $n = 0$ とおくと,

$$t_0(x) = \cos 0 = 1$$

となる． したがって, $n = 0$ に対し, Cheby-chev 多項式は多項式 (定数項のみ) である.

- $n = 1$ とおくと,

$$t_1(x) = \cos(\arccos x) = x$$

となる. したがって, $n = 1$ に対し, Chebychev 多項式は多項式である.

- 帰納法により, $t_n(x)$ が $x$ の $n$ 次の多項式であることを示す. ($n = 0, 1$ についてはすでに証明が終わっている).

- $k \leq n$ に対して $t_k$ が $x$ の $k$ 次の多項式になっているものとする．$\arccos x = \theta$ とおき，三角関数の公式

$$\cos(\alpha + \beta) + \cos(\alpha - \beta) = 2 \cos \alpha \cos \beta$$

において $\alpha = n\theta$, $\beta = \theta$ とおくと，

$$\cos((n+1)\theta)) + \cos((n-1)\theta) = 2 \cos n\theta \cos \theta$$

となる．

先の式に $x = \cos\theta$ を代入し, $t_n$ の定義を使うと,

$$t_{n+1}(x) + t_{n-1}(x) = 2xt_n$$

となる. $t_n$ は $n$ 次の多項式, $t_{n-1}$ は $n-1$ 次の多項式だったから, $t_{n+1}$ は $n+1$ 次の多項式である. ■

- Chebychev 多項式は区間 $[-1,1]$ で利用される．重みは $1/\sqrt{1-x^2}$ である．よって，内積は次のように定義される．

$$(t_k, t_j) = \int_{-1}^{1} t_k(x) t_j(x) \frac{dx}{\sqrt{1-x^2}}.$$

- $\theta = \arccos x$ $(x = \cos\theta)$ と変数変換すると, $t_k = \cos k\theta$, $t_l = \cos l\theta$, $\sqrt{1-x^2} = \sin\theta$, $dx = -\sin\theta d\theta$ だから,

$$(t_k, t_j) = \int_{\pi}^{0} \cos k\theta \cos l\theta \frac{-\sin\theta d\theta}{\sin\theta}$$

となる.

- よって, Fourier 級数展開と同じ計算によって,

$$
(t_k, t_j) = \begin{cases} 0, & k \neq j, \\ \pi, & k = j = 0, \\ \frac{\pi}{2}, & k = j > 0 \end{cases}
$$

となる.

- Chebychev 多項式 $t_k(x)$ は,

$$x_j = \cos\frac{j\pi}{k}, 0 \leq j \leq k$$

において極値を取り，すべての $x_j$ で $t_k(x_j)$ の絶対値が 1 で, かつ $t_k(x_j)$ と $t_k(x_{j+1})$ は符号が異なるという特徴がある．この特徴が数値計算で有用であるため, Chebychev 多項式は数値解析でよく用いられる [森].

# 最小二乗法による直線/曲線のあてはめ

- 実験によって 得られた $n$ 個の測定点

$$\{(x_k, y_k) : 1 \leq k \leq n\}$$

に直線あるいは曲線をあてはめる, という問題を考える.

- 補間多項式によってこの目的は達成されるが，測定値に誤差が含まれる場合には，「曲線が測定点を通過する」ことは必ずしも合理的とはいえない．

- 代替案として, パラメータ $\boldsymbol{\alpha}$ を含む非線形関数 $g(x, \boldsymbol{\alpha})$ を準備し,

$$\sum_{k=1}^{n} \left(y_k - g(x_k, \boldsymbol{\alpha})\right)^2$$

が最小となる $\boldsymbol{\alpha}$ を求めるという方法が考えられる. これを (非線形) 最小二乗法と呼ぶことがある.

- なお, 直交多項式に関する議論では, 0 次の多項式 (定数)1 を基底に含めることとの整合性を取るために添字を 0 から始めていたが, 直感的にはわかりにくいと思われるので, 添字を 1 から始めるスタイルに変更している.

- 関数 $g(x, \boldsymbol{\alpha})$ が実験データを説明するための未知パラメータ $\boldsymbol{\alpha}$ を含む数理モデルであるとき, $g(x, \boldsymbol{\alpha})$ のグラフと測定点が適合するパラメータの値は, パラメータの適切な推定値であると考えられる. (非線形) 最小二乗法は, このような状況で用いられる.

- 今まで (非線形) 最小二乗法と書いてきたが, $g(x, \boldsymbol{\alpha})$ が線形関数のときは, 通常の最小二乗法となる.

- 関数 $g(x, \boldsymbol{\alpha})$ をどう選ぶべきかは問題による．線形関数と多項式はふつうに用いられる．物理法則などからこの関数の形が定められることもある．

- $\boldsymbol{\alpha}$ の次元はデータの点数とは無関係だが, 次元がデータ点数より大きいと $\boldsymbol{\alpha}$ を一意的に定められない可能性が高い.

- $g(x, \boldsymbol{\alpha})$ が一般的な非線形関数の場合には,

$$\min_{\boldsymbol{\alpha}} \sum_{k=1}^{n} \left(y_k - g(x_k, \boldsymbol{\alpha})\right)^2$$

は非線形最小化問題となり, 一般には反復解法によって数値的に解く以外の解法はない.

- 一方, $g(x, \boldsymbol{\alpha})$ が既知の関数の線形結合であらわされている場合, すなわち

$$g(x, \boldsymbol{\alpha}) = \sum_{k=1}^{m} \alpha_k r_k(x)$$

となっていて, $r_k(x)$ は既知であるが $\alpha_k$ は未知である場合には, 問題はもう少し簡単になる.

- 以下のように記号を定義する:

$$
\begin{aligned}
\boldsymbol{y} &= (y_1, \ldots, y_n)^T \\
\boldsymbol{r}(x) &= (r_1(x), \ldots, r_m(x)) \\
\boldsymbol{\alpha} &= (\alpha_1, \ldots, \alpha_m)
\end{aligned}
$$

列ベクトルと行ベクトルが混在していること, 一般に $m \neq n$ であることに注意.

- $\boldsymbol{R} = \begin{pmatrix} \boldsymbol{r}(x_1) \\ \vdots \\ \boldsymbol{r}(x_n) \end{pmatrix}$ とすると, 先に述べた曲線のあてはめの問題は,

$$\|\boldsymbol{y} - \boldsymbol{R}\boldsymbol{\alpha}\|_2^2$$

が最小となる $\boldsymbol{\alpha}$ を求める問題に帰着される.

- 先の式をベクトルの転置を使って書き直すと,

$$
\begin{aligned}
\|\boldsymbol{y}-\boldsymbol{R\alpha}\|_2^2 &= (\boldsymbol{y}-\boldsymbol{R\alpha})^T(\boldsymbol{y}-\boldsymbol{R\alpha}) \\
&= \boldsymbol{y}^T\boldsymbol{y}-2\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{R\alpha}+\boldsymbol{\alpha}^T\boldsymbol{R}^T\boldsymbol{R\alpha}
\end{aligned}
$$

となる.

- この時点では, $\boldsymbol{R}^T\boldsymbol{R}$ が正定であることは保証されていないのだが, 当面はこれを仮定する.

- すると, 最小二乗法に関する議論と同じ理由により,

$$\frac{\partial\|\boldsymbol{y}-\boldsymbol{R}\boldsymbol{\alpha}\|_2^2}{\partial\boldsymbol{\alpha}}=\boldsymbol{0}$$

となる $\boldsymbol{\alpha}$ が最適解となる.

- 上記の等式を計算すると

$$\mathbf{0}^T = -\boldsymbol{y}^T \boldsymbol{R} + \boldsymbol{\alpha}^T \boldsymbol{R}^T \boldsymbol{R}$$

となる．これを転置して移項すると, 次式を得る．

$$\boldsymbol{R}^T \boldsymbol{R} \boldsymbol{\alpha} = \boldsymbol{R} \boldsymbol{y}$$

- 方程式

$$\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\alpha} = \boldsymbol{R}\boldsymbol{y}$$

を, (非線形最小二乗法に関する) 正規方程式と呼ぶ.

- 非線形最小二乗法の場合も, 正規方程式は連立一次方程式であり, これを解くことにより, パラメータの推定値が得られる.

- 通常の最小二乗法と異なり, 非線形最小二乗法に関する正規方程式には, 解がないこともあり得る.

- 測定点に直線をあてはめることを**直線回帰**あるいは**線形回帰**, 曲線をあてはめることを**非線形回帰**と呼ぶことがある.

# Scilabによる線形回帰

- Scilabで直線のあてはめをこなう関数は`reglin`である.
- この関数は, スカラーに関する直線回帰だけでなく, $\boldsymbol{x}$と$\boldsymbol{y}$がベクトルで, $\boldsymbol{y}$に写像$\boldsymbol{Ax}+\boldsymbol{b}$をあてはめたい場合にも使える.

- $(\boldsymbol{x}_i, \boldsymbol{y}_i)_{i=1,\ldots,Ns}$ が $Ns$ 個の標本で (列ベクトルとする), $\dim \boldsymbol{x} = n_x$, $\dim \boldsymbol{y} = n_y$ とする.
- 線形回帰をおこなう際に次元を合わせる必要があるので, $\boldsymbol{A} \in \mathbb{R}^{n_y \times n_x}$, $\boldsymbol{b} \in \mathbb{R}^{n_y}$ となる.

- `reglin` を使うときには, $\boldsymbol{x}_i$ を横に $Ns$ 個並べた行列 (仮に `xmat` とする) と, $\boldsymbol{y}_i$ を横に $Ns$ 個並べた行列 (仮に `ymat` とする) を作り, `[A,b,s]=reglin[xmat,ymat]` とする.
- `A`, `b` には回帰の結果得られる行列およびスカラーが, `s` には回帰残差の標準偏差が返される.

- $x$ と $y$ がスカラーの場合の直線回帰: 次に, $x = 1, 2 \ldots, 100$, $y = x+$(正規分布する乱数) として, 線形回帰をおこなった結果を示す. 実行結果から, $\mathtt{a} \simeq 1$, $\mathtt{b} \simeq 0$ となっていることが確認できる.

```
x=1:100;
y=x+rand(1,100,'normal');
[a,b,s]=reglin(x,y);
-->[a,b,s]
 ans  =
   0.9997142 -0.0073729 1.0491074
```

- 次に, `xmat` が 2 行 100 列, `ymat` が 3 行 100 列の乱数行列 (正規分布) の場合の線形回帰の例を次に示す．空白を減らして表示しているので注意.

```
xmat=rand(2,100,'normal');
ymat=rand(3,100,'normal');
[A,b,s]=reglin(xmat,ymat);
```

```
-->A
A  =
-0.0940391 -0.0510797
-0.0298481  0.0738067
-0.0712583  0.0198992
```

```
-->b
b  =
-0.0986652
-0.0063689
-0.0370478
```

```
-->s
s  =
1.0819806 0.0205290 0.0777786
0.0205290 1.0073556 0.1069899
0.0777786 0.1069899 0.9121303
```

# Scilab による曲線のあてはめ

- Scilab でデータに曲線をあてはめるには非線形最小二乗法問題を解く関数 `leastsq` を使うが, 使い方に若干の注意が必要である.

- 例として，曲線 $y = f_p(a,x) = a_0 + a_1x + a_2x^2 + a_3x^3$ にデータをあてはめる問題を考える． $x$ が $-1$ から 1 まで 0.1 刻みで動き，$(a_0,a_1,a_2,a_3) = (1,1,1,1)$ で，$y$ は $f_p(a,x)$ に $[-0.1,0.1]$ の範囲に一様分布する雑音を加えたものとする．

- `leastsq` には引数として最小化したい関数, データ, パラメータの初期値を与える. 初期値次第では適切なあてはめができないことがあるので注意. その場合には乱数等を用いた再試行が必要.

- `leastsq` はパラメータ初期値と同じ次元のパラメータを探索するので, 初期値の次元が実際のパラメータより大きくても, とりあえず動く.

# Scilab のプログラム例:

```
deff('y=fp(a,x)','y=a(1)+a(2).*x+a(3).* (x.^2) + a(4).* (x.^3)');
deff('e=fopt(a,xmat,ymat)','e=(norm(fp(a,xmat)-ymat))^2');

aT=[1 2 3 4]; //パラメータの真値
xmat=-1:.1:1; //標本点の x 座標
ymat=fp(aT,xmat)+0.1*(rand(xmat)-0.5); //y の値を生成
```

# パラメータ推定値の例:

```
--> aEst
aEst  =
1.0050031   1.9672115   2.9836197   4.0484874
```

- 上記のようにすると, `f` には最小化された関数値が, `aEst` にはパラメータ推定値が格納される.

- `leastsq` にはより多くのパラメータを与えることができるが, この講義では略す. 興味がある者はScilab のオンラインマニュアルを参照すること.

- 関数 `datafit` もデータへの曲線をおこなう関数である. 初期値に関する注意は `leastsq` と同様.
- `leastsq` との違いは, パラメータを列ベクトルにしなければならないことと, データとして $x$ 軸と $y$ 軸の値をまとめた行列を与えることである.

# Scilab によるプログラム例:

```
deff('y=fp(a,x)','y=a(1)+a(2).*x+a(3).* (x.^2) + a(4).* (x.^3)');
deff('e=fopt(a,z)','e=(z(2)-fp(a,z(1)))^2');

aT=[1;2;3;4];
xmat=-1:.1:1;
ymat=fp(aT,xmat)+0.1*rand(xmat,'normal');

a0=[1;1;1;1]; //パラメータは列ベクトル
[aEst,err]=datafit(fopt,[xmat;ymat],a0);
```

# パラメータ推定値の例:

```
--> aEst
aEst  =
1.0130090
1.8372007
2.9911182
4.1526024
```

- 誤差を含む実験データのグラフを描くときには, 測定点を完全に通過するわけではない滑らかな曲線を引くことがある
- このような場合, 伝統的には曲線定規などが使われていたが, コンピュータで似たようなことをやることは案外難しい

- Scilab では, `lsq_splin` という関数を使うと, 上記に近いことができる.

- `lsq_splin` は, 標準では, 3 個の引数を取り,
  `[y,d]=lsq_splin(xd,yd,x)`
  のようにして使う.

- `xd` が測定データの $x$ 座標, `yd` が測定データの $y$ 座標で, `x` はスプライン曲線を生成するための $x$ 軸状の分点である.

- この関数は, 与えられた分点から決まるスプライン曲線を使い, 測定データを最小二乗近似する.

- 返却値の `y` はスプライン曲線の $y$ 座標, `d` は導関数で, これらは Scilab で スプライン曲線を生成する関数 `interp` に適合した形になっている.

- `x` の与え方は任意であるが, 刻みを `xd` よりも粗く取らないと, スプライン補間する意味がない.

- もっとも単純には, 以下のように `interp` および `plot` とセットで使うと考えればよい.

```
[y,d]=lsq_splin(xd,yd,x)
yp=interp(xp,x,y,d)
plot(xp,yp)
```

(他にもオプションがあるが, 深入りしない).

- 以下に, 正弦波に $[-0.3, 0, 3]$ の範囲の一様乱数の雑音が乗ったデータから曲線を引くサンプルプログラムを示す.

```
xd=0:.1:2*%pi;
yd=sin(xd)+0.3*(2*rand(1,length(xd))-1);
x=linspace(0,2*%pi,5);
[s,d]=lsq_splin(xd,yd,x);
xp=linspace(0,2*%pi,100);
yp=interp(xp,x,s,d);
scatter(xd,yd);
plot(xp,yp);
```

描画されたグラフは以下の通り.