Revista Brasileira de Ensino de Física, v. 28, n. 3, p. 397-401, (2006)
www.sbfisica.org.br

Notas e Discussões

Polinômio dos quadrados mínimos condicionado

(The constrained least squares polynomial)

Oclide J. Dotto[1] e Adalberto A. Dornelles Filho

Departamento de Matemática e Estatística, Universidade de Caxias do Sul, Caxias do Sul, RS, Brasil

Recebido em 20/2/2006; Revisado em 15/5/2006; Aceito em 29/5/2006

O ajustamento polinomial a dados experimentais pelo método de quadrados mínimos é muito usado, e aqui estudamos a introdução de uma simples modificação, que, em certos casos, é mais adequada, isto é, condicionamos o polinômio a conter um dado ponto. Incluímos uma implementação do correspondente algoritmo em Matlab.

Palavras-chave: ajustamento, quadrados mínimos, condicionamento.

Since the least-squares polynomial fit is used very often, we study here a simple modification that is more suitable for some cases, that is, we constrain the polynomial to pass through a given point, besides satisfying the least-squares property. We include an implementation of the corresponding algorithm in Matlab.

Keywords: data fitting, least squares, constraint.

1. Introdução

É popular o ajustamento polinomial discreto dos quadrados mínimos (q. m.). Para evocar o processo desse ajustamento, consideremos os dados da Tabela 1, onde a primeira linha forma uma seqüência finita crescente.

Tabela 1 - Dados a ajustar.

$\mathbf{x}$	x_0	x_1	$\cdots$	x_n
$\mathbf{y}$	y_0	y_1	$\cdots$	y_n

Suponhamos que queiramos ajustar a esses dados o polinômio

$$p(x) = a_k x^k + a_{k-1} x^{k-1} + \cdots + a_1 x + a_0, \qquad (1)$$

de ordem $k \leq n$ dos q. m. A melhor maneira de obter a solução é o uso da álgebra linear como segue. Começamos gerando a matriz de planejamento (*design matrix* [1])

$$\mathbf{X} = \begin{bmatrix} x_0^k & \cdots & x_0^2 & x_0 & 1 \\ x_1^k & \cdots & x_1^2 & x_1 & 1 \\ \vdots & & \vdots & \vdots & \vdots \\ x_n^k & \cdots & x_n^2 & x_n & 1 \end{bmatrix}, \qquad (2)$$

de ordem $(n+1) \times (k+1)$, determinada pelo vetor $\mathbf{x} = [x_0 \;\; x_1 \;\; \ldots \;\; x_n]^t$. A solução única (porque $\mathbf{X}$ é não-singular) da equação normal

$$\mathbf{X}^t \mathbf{X} \mathbf{a} = \mathbf{X}^t \mathbf{y} \qquad (3)$$

é o vetor $\mathbf{a}$ dos coeficientes do polinômio dos q. m. (1).

Do ponto de vista dos métodos de resolução numérica de (3), o uso da matriz de planejamento pode levar a imprecisões uma vez que $\mathbf{X}$ se torna mal-condicionada quando sua ordem é alta. No entanto, isso não é relevante na maioria dos casos práticos, onde dificilmente a ordem do polinômio ajustado é maior que 3.

O desenvolvimento mais completo da teoria e solução numérica dos quadrados mínimos foi feita por Åke Björck em [2], mas pode ser encontrado em muitos bons livros de álgebra linear, como [3, 4, 5]. Uma introdução rápida sobre a solução dos q. m. encontra-se em [6].

Façamos aqui uma pequena digressão útil. Na estimativa da solução $\mathbf{Xa} = \mathbf{y}$ mediante o método dos q. m., é conveniente às vezes ponderar essa estimativa, isto é, usar e *equação normal ponderada*

$$\mathbf{X}^t \mathbf{W} \mathbf{X} \mathbf{a} = \mathbf{X}^t \mathbf{W} \mathbf{y}, \qquad (4)$$

em vez da equação normal ordinária (3), escolhendo, para matriz de ponderação $\mathbf{W}$, uma conveniente matriz simétrica definida positiva, que minimize o vetor residual $\mathbf{r} := \mathbf{Xa} - \mathbf{y}$ no sentido dos q. m. associados ao produto interno definido por $\mathbf{W}$, o que significa simplesmente minimizar $\mathbf{r}^t \mathbf{W} \mathbf{r}$. É conhecido [7] que a escolha *ótima* é a inversa da matriz $\mathbf{V} = [v_{ij}]$ das variâncias-covariâncias dos vetores aleatórios das observações, $\mathbf{y}_i = [y_{1i} \;\; y_{2i} \;\; \cdots \;\; y_{mi}]^t$, $i = 0, 1, \ldots, n$,

[1]E-mail: ojdotto@ucs.br.

isto é, $\mathbf{W} = \mathbf{V}^{-1}$ (cf. também [2, 8]). Ainda, caso os desvios possam ser supostos não-correlacionados, todas as covariâncias são nulas e $\mathbf{V}$ se reduz a uma matriz diagonal, cujos elementos diagonais são as variâncias, que medem a dispersão dos desvios em torno da média, e, nesse caso, para inverter $\mathbf{V}$, basta inverter esses elementos diagonais (que são não-nulos). O efeito do uso de $\mathbf{V}^{-1}$ para ponderação é terem menor peso os desvios maiores.

Voltemos agora ao assunto central do artigo. A curva que representa geometricamente o polinômio dos q. m. não passa pelos pontos definidos pelos dados na Tabela 1, a não ser que coincida com a do polinômio de interpolação, o que ocorre quando $k = n$. No entanto, em certos casos, pela natureza do problema, sabemos de antemão que deve passar por determinado ponto, mesmo que esse ponto não figure entre os dados.

A um polinômio dos q. m., forçado a passar por um ponto P_0 escolhido previamente, chamamos de *polinômio dos q. m. condicionado*, ou *polinômio dos q. m. ancorado no ponto* P_0.

2. Reta dos quadrados mínimos ancorada na origem

A lei de Hooke estabelece que, quando uma força F é aplicada a uma mola homogênea, o alongamento x da mola varia linearmente com essa força, dentro de limites, isto é,

$$F = ax, \tag{5}$$

onde a é dita a constante de Hooke da mola. Para determinar a, executamos experimentos que fornecem uma tabela de duas linhas com os valores de F dados e os valores de x resultantes. Em seguida ajustamos a reta dos q. m. aos dados e a declividade dessa reta é uma estimativa da constante de Hooke da mola. Essa reta quase certamente não passa pela origem, mesmo que a origem se inclua nos dados. No entanto, idealmente, deveria passar pela origem, pois o alongamento é nulo para uma força aplicada nula.

Para forçar a reta dos q. m. a passar pela origem, basta achar a solução dos q. m. do sistema linear

$$\left\{ \begin{array}{l} x_0 a = F_0 \\ x_1 a = F_1 \\ \quad\vdots \\ x_n a = F_n. \end{array} \right. \tag{6}$$

Uma relação de proporcionalidade direta entre duas grandezas, como em (5), ocorre com freqüência. Outros exemplos são: a intensidade i e a tensão v de uma corrente elétrica aplicada a um resistor, a força F e aceleração resultante a aplicada a um móvel, etc.

Exemplo 1

Vejamos um caso específico. A Tabela 2 mostra os valores do alongamento x (em centímetros) de uma mola *vs.* força de tração F (em Newtons) aplicada, ambos em módulo.

Tabela 2 - Alongamentos x (cm) de uma mola *vs.* forças de tração F (N) aplicadas.

F (N)	x (cm)
1,00	1,70
2,00	2,75
3,00	4,08
4,00	5,21
5,00	6,38
6,00	7,88
8,00	11,21
10,00	12,65

Aqui a reta dos q. m. tem por equação

$$F = 0{,}78x - 0{,}16, \tag{7}$$

e a reta dos q. m. ancorada na origem

$$F = 0{,}76x. \tag{8}$$

A Fig. 1 mostra ambas as retas. A primeira não contém a origem. Vemos que as declividades, em unidades coerentes, não coincidem, embora sejam próximas, e é razoável dizer que o valor $a = 0{,}76$ é uma estimativa melhor que $a = 0{,}78$ para a constante da mola.

Para a reta dos q. m. ancorada na origem, a equação matricial (3) pode ser reescrita como

$$\mathbf{x}^t\mathbf{x}\mathbf{a} = \mathbf{x}^t\mathbf{y}. \tag{9}$$

De (9) deduzimos que a declividade a dessa reta é dada por

$$a = \frac{x_0y_0 + x_1y_1 + \cdots + x_ny_n}{x_0^2 + x_1^2 + \cdots + x_n^2}. \tag{10}$$

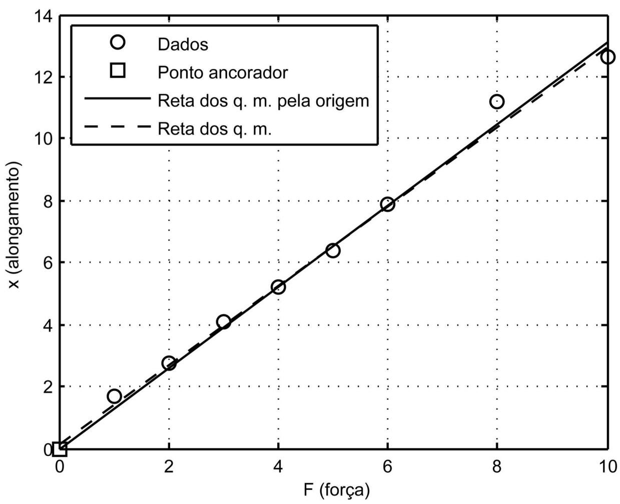

Figura 1 - Retas ancorada e livre dos q. m. ajustadas a dados experimentais.

Em laboratórios de ensino, em vez da fórmula (10), é usada comumente a fórmula

$$a = \frac{1}{n+1}\left(\frac{y_0}{x_0} + \frac{y_1}{x_1} + \cdots + \frac{y_n}{x_n}\right), \qquad (11)$$

que expressa a média aritmética a das declividades y_i/x_i das retas, cada uma das quais contendo a origem e um ponto (x_i, y_i). A expressão (11) é bastante intuitiva e produz um resultado levemente diferente ($a = 0{,}73$, no caso) e, por outro lado, parece não ter nenhuma relação com (10). De fato, elas não são equivalentes. Mas é interessante observar o que segue.

Primeiro, adotando a equação normal ponderada (4) e a inversa $\mathbf{V}^{-1}$ da matriz das variâncias-covariâncias para matriz de ponderação, a fórmula (9) passa a escrever-se

$$\mathbf{x}^t\mathbf{V}^{-1}\mathbf{x}\mathbf{a} = \mathbf{x}^t\mathbf{V}^{-1}\mathbf{y}. \qquad (12)$$

Segundo, suponhamos que, para cada $i \in \{0, 1, 2, \ldots, n\}$, tenhamos feito um conjunto de m observações $y_{1i}, y_{2i}, \ldots, y_{mi}$; então podemos falar da matriz das variâncias-covariâncias dos vetores aleatórios dessas observações, $\mathbf{y}_i = [y_{1i}\ \ y_{2i}\ \ \ldots\ \ y_{mi}]^t$, $i = 0, 1, 2, \ldots, n$. Usando a fórmula normal (12), ponderada pela inversa da matriz $\mathbf{V} = [v_{ij}]$ das variâncias-covariâncias, ou ponderando as declividades y_i/x_i em (11) pelos inversos das correspondentes variâncias dessas declividades (ao invés de tomar a média aritmética), obtemos por ambos os caminhos a fórmula

$$a = \frac{\dfrac{x_0y_0}{v_0} + \dfrac{x_1y_1}{v_1} + \cdots + \dfrac{x_ny_n}{v_n}}{\dfrac{x_0^2}{v_0} + \dfrac{x_1^2}{v_1} + \cdots + \dfrac{x_n^2}{v_n}}, \qquad (13)$$

onde pusemos $v_i := v_{ii}$ (variâncias) e admitimos $v_{ij} = 0$, se $i \neq j$ (covariâncias nulas). Com relação a ponderação em (13), lembramos que, por uma propriedade da variância, temos:

$$\operatorname{var}\left(\frac{\mathbf{y}_i}{x_i}\right) = \left[\frac{v_i}{x_i^2}\right],$$

e o denominador principal em (13) é a soma dos inversos dessas variâncias.

Notamos que a fórmula (10) é recuperada de (13) supondo $v_0 = v_1 = \ldots = v_n$, e, com uma simplificação maior, que a fórmula (11) segue de (13), usando as correspondentes variâncias do vetor das declividades

$$\left[\frac{y_{1i}}{x_i}\ \ \frac{y_{2i}}{x_i}\ \ \cdots\ \ \frac{y_{mi}}{x_i}\right]^t, \qquad (14)$$

todas iguais a $1/(n+1)$, isto é, para $i = 0, 1, 2 \ldots, n$. Isso explica porque as fórmulas (10) e (11) não são equivalentes.

3. Polinômio dos quadrados mínimos condicionado

Em vez de ancorar na origem a reta dos q. m., podemos ancorá-la num ponto qualquer. Mas tratemos do caso geral: mostremos como ancorar o polinômio de ordem k dos q. m. num ponto escolhido (x_0, y_0). O procedimento é feito em três etapas:

1. operamos uma translação de eixos de maneira que a nova origem seja o ponto (x_0, y_0); para isso basta subtrair x_0 das abscissas e y_0 das ordenadas dos dados;

2. ajustamos aos dados transformados o polinômio q dos q. m. ancorado na origem, como no Exemplo 1;

3. fazemos, por último, a transformação inversa da executada na etapa 1, obtendo o polinômio final dos q. m., $p(x) = q(x - x_0)$, ancorado em (x_0, y_0).

Suponhamos que os dados na Tabela 1 já tenham sofrido a alteração descrita na etapa 1. Para operacionalizar a etapa 2, suprimimos a última coluna da matriz de planejamento (2), obtendo uma matriz que indicaremos com $\mathbf{U}$, e achamos a solução a dos q. m. do sistema linear

$$\mathbf{U}^t\mathbf{U}\mathbf{a} = \mathbf{U}^t\mathbf{y}. \qquad (15)$$

O vetor-solução $\mathbf{a}$ será formado pelos coeficientes do polinômio q ancorado na origem, descrito na etapa 2. Para o caso de uma reta, a equação matricial (15) torna-se uma equação escalar, da forma (9), onde $\mathbf{x}$ e $\mathbf{y}$ são os vetores dos dados.

No apêndice implementamos esse algoritmo em MATLAB, com o nome `polqmcond`. O programa produz também a correspondente visualização geométrica.

Exemplo 2

Consideremos a integral elíptica completa

$$E(x) := \int_0^{\pi/2} \sqrt{1 - \operatorname{sen}^2 x\ \ \operatorname{sen}^2\theta}\ \ d\theta. \qquad (16)$$

Mediante integração numérica, obtemos os valores de $E(x)$ (com x dado em graus) e construímos a Tabela 3.

Tabela 3 - Valores de $E(x)$, obtidos com (16) mediante integração numérica.

x	5	10	15	20	25	30
$E(x)$	1,5678	1,5589	1,5442	1,5238	1,4981	1,4675

Se olharmos para a distribuição desses dados, veremos que um ajustamento parabólico é adequado. Além disso, é óbvio que $E(0) = \pi/2$. Então é indicado ancorar a parábola dos q. m. no ponto $(0, \pi/2)$. Utilizando o programa `polqmcond`, obtemos essa parábola

$$P(x) := -0{,}0001118x^2 - 0{,}0001003x + \frac{\pi}{2}. \qquad (17)$$

A parábola "livre" dos q. m. é dada por

$$Q(x) := -0{,}0001089x^2 - 0{,}0002137x + 1{,}5717. \quad (18)$$

A Fig. 2 exibe ambas as parábolas. Visualmente elas se superpõem em quase todo o trajeto.

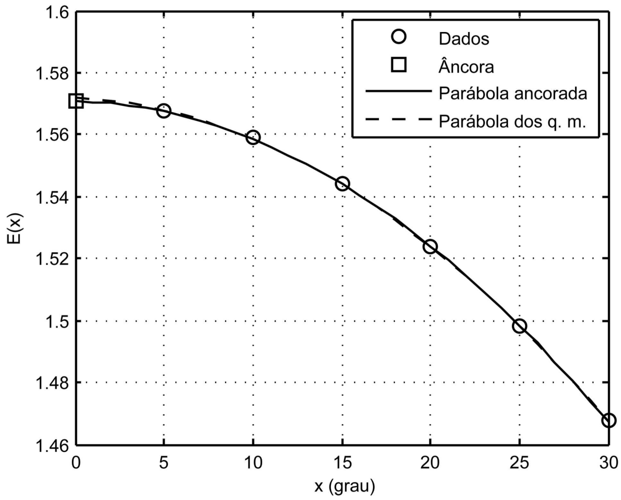

Figura 2 - Parábolas dos q. m. ancorada e livre.

Para compararmos os dois ajustamentos, os testamos nos pontos $x = 2, 12, 17, 27$. Obtivemos a Tabela 4 na qual a segunda linha contém os valores de $E(x)$ obtidos por integração numérica aqui considerados exatos, a terceira linha os valores de $P(x)$ do ajustamento dos q. m. ancorado, e a quarta linha os valores de $Q(x)$ do ajustamento livre.

Tabela 4 - Comparação entre ajustamentos ancorado $P(x)$ e livre $Q(x)$.

x (°)	2	12	17	27
$E(x)$	1,5703	1,5537	1,5367	1,4864
$P(x)$	1,5701	1,5535	1,5368	1,4866
$Q(x)$	1,5709	1,5535	1,5366	1,4866

Vemos que, nesse exemplo, o ajustamento ancorado é melhor para os pontos próximos ao ponto ancorador, como era esperado, e longe dele os ajustamentos tendem a ser equivalentes. Claramente, para um ajustamento linear, por exemplo, a situação longe da âncora é outra, como mostra a Fig. 1.

4. Comentários finais

Nos laboratórios de ensino, é bastante comum realizar experimentos para obter o coeficiente de proporcionalidade entre duas grandezas diretamente proporcionais. Se for usado o recurso do ajuste linear dos q. m., disponível em algumas calculadoras científicas, dificilmente obteremos uma reta que passe exatamente pela origem. Uma elegante solução para essa dificuldade é usar a fórmula (10) que decorre do ajustamento dos q. m. ancorado na origem.

Mais geral, mostramos que é possível e relativamente fácil, ao ajustarmos um polinômio dos quadrados mínimos, condicioná-lo a conter um ponto de interesse. Poder-se-ia tentar desenvolver um algoritmo que o condicione a conter mais de um ponto, contanto que ainda retenha um número de graus de liberdade suficientemente elevado. Por exemplo, um caso extremo oposto é condicionar uma reta dos q. m. a conter dois pontos, o que corresponde a ela ficar com zero graus de liberdade e equivale ao problema trivial de achar a equação da reta que passa por esses dois pontos. Não temos certeza se há interesse prático nessa generalização de ancoragem. De qualquer maneira, parece-nos que o caso interessante, exeqüível e útil, é a ancoragem em um ponto apenas.

Agradecimento

Os autores agradecem ao parecerista a gentil apreciação do artigo, assim como as sugestões.

Apêndice

```
function p = polqmcond(x, y, n, x0, y0)
% ---------------------------------------------------------
% POLQMCOND determina o polinômio dos quadrados mínimos,
% ancorado num ponto fixo. Usa cálculo simbólico.
%
% Entrada:
% x, y, dados (vetores de mesmo numero de componentes);
% n, para obter o polinômio dos q. m. de ordem n-1;
% x0, y0, coordenadas do ponto que o polinômio deve conter.
%
% Saída:
% p, polinômio dos q. m. ancorado em (x0,y0);
% representação geométrica do polinômio, junto com os dados
% e o ponto (x0,y0).
% ---------------------------------------------------------
if n <= 1 | fix(n) ~= n,
   error('O 3.o argumento de entrada deve ser um inteiro >= 2'),
 end, x = x(:)'; a = min([x x0]); b = max([x x0]); x = x - x0;
 m = length(x); y = y(:); y = y - y0;

%As n-1 primeiras colunas da Matriz de Planejamento
for i = 1 : n,
   X(i,:) = x.^(n-i);
end, X = X'; X = X(:, 1:(n-1));

%Polinômio de ordem n-1 dos q. m. ancorado na origem
q = X\y;

%Polinômio de ordem n-1 dos q. m. ancorado no ponto (x0, y0)
syms t, for k = 1 : (n-1),
   r(k) = (t-x0)^k;
end, r = fliplr(r); p = r * q + y0; p = expand(p);

%Representação geométrica dos dados, (x0, y0)
%e o polinômio dos q.m.
 s = vectorize(p); t = a : 0.001 : b; Y = eval(s);
 plot(x + x0, y + y0, 'o', x0, y0, 'sr', t, Y),grid
```

Referências

[1] Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Analysis* (Prentice Hall, Nova York, 2002)

[2] Åke Björck, *Numerical Methods for Least Squares Problems* (SIAM, Philadelphia, 1966).

[3] James W. Demmel, *Applied Numerical Linear Algebra* (SIAM, Philadelphia, 1997).

[4] Gene H. Golub and Charles F. Van Loan, *Matrix Computations* (The John Hopkins University Press, London, 1996).

[5] Lloyd N. Trefethen and David Bau III, *Numerical Linear Algebra* (SIAM, Philadelphia, 1997).

[6] Mário Barone Júnior, *Álgebra Linear* (IME-USP, São Paulo, 2005), 3^a ed.

[7] Gilbert Strang and Kay Borre, *Linear Algebra, Geodesy and GPS* (Wellesley-Cambridge Press, Wellesley, 1997).

[8] José Henrique Vuolo, *Fundamentos da Teoria de Erros* (Ed. Edgard Blücher Ltda., São Paulo, 2005), 2ª ed.