

TEA040 Otimização de parâmetros 2026-2

– Notas de aula

Nelson Luís Dias

Prof. Titular, Departamento de Engenharia Ambiental, UFPR

3 de agosto de 2026

Sumário

1	Mínimos quadrados, função afim	3
2	A regressão $y = ax$	7
3	Erros dos parâmetros de $y = ax + b$	10
4	Matrizes e concavidade	15
4.1	Transformações definidas e indefinidas	15
4.2	Máximos e mínimos de funções de mais de uma variável .	16
4.3	Multiplicadores de Lagrange	19

Capítulo 1

Mínimos quadrados, função afim

A equação que equaciona tudo é açafão

Talvez o método de otimização mais conhecido seja a obtenção por mínimos quadrados dos parâmetros a e b da equação da reta,

$$y = ax + b \quad (1.1)$$

através de um conjunto de pontos $(x_i, y_i), i = 1, \dots, n$. Em geral, os pontos observados, ou medidos, apresentam uma certa dispersão, e não se ajustam perfeitamente ao modelo adotado (no caso, a reta (1.1)). A figura 1.1 mostra uma comparação de dados da temperatura média da superfície da água medida em uma bóia (Sentinel, lago Mead, EUA) e a temperatura média da superfície estimada por sensoriamento remoto (MODIS).

Construa uma função objetivo

$$F(a, b; \mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n [(ax_i + b) - y_i]^2, \quad (1.2)$$

e force (a, b) a ser um ponto extremo:

$$\frac{\partial F}{\partial a} = 0, \quad (1.3)$$

$$\frac{\partial F}{\partial b} = 0. \quad (1.4)$$

A 1ª equação acima produz

$$\begin{aligned} \frac{2}{n} \sum_{i=1}^n [(ax_i + b) - y_i] x_i &= 0, \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i &= \sum_{i=1}^n x_i y_i. \end{aligned} \quad (1.5)$$

A 2ª equação produz

$$\begin{aligned} \frac{2}{n} \sum_{i=1}^n [(ax_i + b) - y_i] &= 0, \\ a \sum_{i=1}^n x_i + nb &= \sum_{i=1}^n y_i. \end{aligned} \quad (1.6)$$

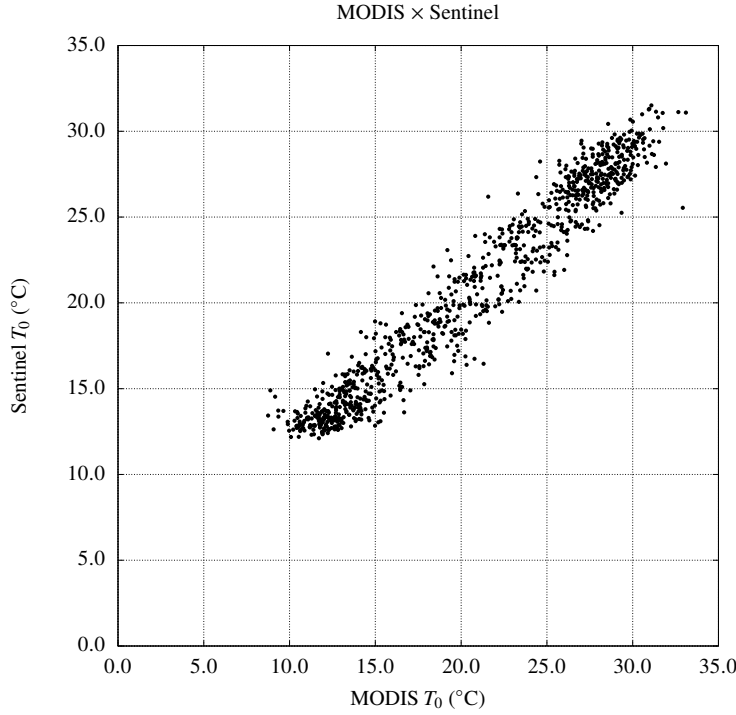


Figura 1.1: Comparação de dados medidos em uma bóia com dados medidos por sensoriamento remoto, lago Mead, EUA.

Temos portanto um sistema 2×2 ,

$$\begin{bmatrix} \sum x_i & n \\ \sum x_i^2 & \sum x_i \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \quad (1.7)$$

cujas soluções são

$$a = \frac{(\sum x_i)(\sum y_i) - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2},$$

$$b = \frac{(\sum x_i)(\sum x_i y_i) - (\sum x_i^2)(\sum y_i)}{(\sum x_i)^2 - n \sum x_i^2}.$$

O denominador das duas equações acima é negativo, e é preferível escrever

$$a = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}, \quad (1.8)$$

$$b = \frac{(\sum x_i^2)(\sum y_i) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}. \quad (1.9)$$

Embora (1.8)–(1.9) sejam frequentemente apresentadas nesta forma, existe uma alternativa preferível tanto analiticamente quanto computacionalmente. Ela consiste em dividir (1.5)–(1.6) por n e reconhecer estatísticas:

$$a \left[\frac{1}{n} \sum_{i=1}^n x_i^2 \right] + b \left[\frac{1}{n} \sum_{i=1}^n x_i \right] = \frac{1}{n} \sum_{i=1}^n x_i y_i,$$

$$\underbrace{a \left[\frac{1}{n} \sum_{i=1}^n x_i \right]}_{\bar{x}} + \underbrace{b}_{\bar{y}} = \left[\frac{1}{n} \sum_{i=1}^n y_i \right];$$

ou

$$\begin{aligned} a \left[\frac{1}{n} \sum_{i=1}^n x_i^2 \right] + b \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i y_i, \\ a \bar{x} + b &= \bar{y}, \end{aligned} \quad (1.10)$$

$$\begin{aligned} a \left[\frac{1}{n} \sum_{i=1}^n x_i^2 \right] + [\bar{y} - a \bar{x}] \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i y_i, \\ a \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right] &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}. \end{aligned} \quad (1.11)$$

É possível reconhecer a variância de $\mathbf{x}$ e a covariância entre $\mathbf{x}$ e $\mathbf{y}$ nas expressões acima:

$$\begin{aligned} \text{Var}\{\mathbf{x}\} = c_{xx} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ &= \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i \bar{x} + \bar{x}^2) \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{2\bar{x}}{n} \sum_{i=1}^n x_i + \bar{x}^2 \\ &= \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2; \end{aligned} \quad (1.12)$$

$$\begin{aligned} \text{Cov}\{\mathbf{x}\} = c_{xy} &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n (x_i y_i - x_i \bar{y} - y_i \bar{x} + \bar{x} \bar{y}) \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \frac{\bar{x}}{n} \sum_{i=1}^n y_i - \frac{\bar{y}}{n} \sum_{i=1}^n x_i + \bar{x} \bar{y} \\ &= \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}. \end{aligned} \quad (1.13)$$

O desvio-padrão s_x é a raiz quadrada da variância:

$$s_x^2 \equiv c_{xx}. \quad (1.14)$$

Segue-se de (1.10)–(1.13) que

$$a = \frac{c_{xy}}{c_{xx}}, \quad (1.15)$$

$$b = \bar{y} - a\bar{x}. \quad (1.16)$$

Vale a pena incluir um descritor da *qualidade do ajuste da reta*, o coeficiente de correlação r_{xy} :

$$r_{xy} = \frac{c_{xy}}{s_x s_y}. \quad (1.17)$$

Isso nos leva a nosso primeiro problema:

1.1 Regressão linear por mínimos quadrados.

- Escreva uma rotina `stat2` para calcular a média e a variância de um array $\mathbf{x}$.
- Escreva uma rotina `covar` para calcular a covariância entre dois arrays $\mathbf{x}$ e $\mathbf{y}$.
- Escreva uma rotina `reglin` para obter a e b a partir de dois arrays $\mathbf{x}$ e $\mathbf{y}$.

1.2 Escreva um programa `reglinT0xy` para ler o arquivo `compareT0.tdat` e calcular a e b .

1.3 Plote a reta $y = ax + b$ obtida.

Capítulo 2

A regressão $y = ax$

Uma regressão mais simples, e muito comum, é:

$$y = ax. \quad (2.1)$$

Vamos precisar de um modelo estocástico para (2.1). Escolhemos

$$Y = ax + \mathcal{E} \quad (2.2)$$

com $E\{\mathcal{E}\} = 0$. O valor esperado de Y é

$$E\{Y\} = ax. \quad (2.3)$$

Neste modelo tanto a quando x são determinísticos. A principal consequência é que Y é *homocedástico* com

$$\text{Var}\{Y\} = \text{Var}\{\mathcal{E}\}. \quad (2.4)$$

Nossa função objetivo agora é

$$F(a; \mathbf{x}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n [ax_i - y_i]^2. \quad (2.5)$$

Derivamos em relação a a , e igualamos a zero:

$$\begin{aligned} \frac{dF}{da} &= 0, \\ \frac{2}{n} \sum_{i=1}^n [ax_i - y_i] x_i &= 0, \\ a \left[\sum_{i=1}^n x_i^2 \right] &= \sum_{i=1}^n x_i y_i, \\ a &= \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}. \end{aligned} \quad (2.6)$$

A regressão “pela origem” envolve momentos em relação à origem! É conveniente definir

$$m_{xx} \equiv \frac{1}{n} \sum_{i=1}^n x_i^2, \quad (2.7)$$

$$m_{yy} \equiv \frac{1}{n} \sum_{i=1}^n y_i^2, \quad (2.8)$$

$$m_{xy} \equiv \frac{1}{n} \sum_{i=1}^n x_i y_i. \quad (2.9)$$

Nessa notação,

$$a = \frac{m_{xy}}{m_{xx}}.$$

O “erro” envolvido na estimativa de a pode ser obtido se nós re-interpretarmos a (e y_i) como variáveis aleatórias, e se supusermos que os Y_i s são independentes:

$$A = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}. \quad (2.10)$$

O valor esperado de A é

$$\begin{aligned} E\{A\} &= \frac{\sum_{i=1}^n x_i E\{Y_i\}}{\sum_{i=1}^n x_i^2}, \\ &= \frac{\sum_{i=1}^n x_i a x_i}{\sum_{i=1}^n x_i^2} = a. \end{aligned} \quad (2.11)$$

Portanto, A é um estimador não-enviesado de a . A variância de A é

$$\begin{aligned} \text{Var}\{A\} &= \frac{\sum_{i=1}^n x_i^2 \text{Var}\{Y\}}{\left[\sum_{i=1}^n x_i^2\right]^2} \\ &= \frac{\sum_{i=1}^n x_i^2}{\left[\sum_{i=1}^n x_i^2\right]^2} \text{Var}\{\mathcal{E}\} \\ &= \frac{\text{Var}\{\mathcal{E}\}}{\sum_{i=1}^n x_i^2}. \end{aligned} \quad (2.12)$$

O estimador para $\text{Var}\{\mathcal{E}\}$ é o quadrado do erro padrão de estimativa s_e^2 :

$$\begin{aligned} s_e^2 &= \frac{1}{n} \sum_{i=1}^n [y_i - ax_i]^2 \\ &= \frac{1}{n} \sum_{i=1}^n [y_i^2 - 2ax_i y_i + a^2 x_i^2] \\ &= \frac{1}{n} \left[\sum_{i=1}^n y_i^2 - 2a \sum_{i=1}^n x_i y_i + a^2 \sum_{i=1}^n x_i^2 \right] \\ &= m_{yy} - 2am_{xy} + a^2 m_{xx}; \\ a &= \frac{m_{xy}}{m_{xx}} \Rightarrow \\ s_e^2 &= m_{yy} - 2 \frac{m_{xy}}{m_{xx}} m_{xy} + \frac{m_{xy}^2}{m_{xx}^2} m_{xx} \\ &= m_{yy} - \frac{m_{xy}^2}{m_{xx}} m_{yy} \end{aligned}$$

$$= [1 - q_{xy}^2] m_{yy}. \quad (2.13)$$

A variável

$$q_{xy} \equiv \frac{m_{xy}}{\sqrt{m_{xx}m_{yy}}} \quad (2.14)$$

funciona como o “coeficiente de correlação” natural para a regressão pela origem. Vamos chamá-la de “coeficiente de quorrelação”. Assim como acontece com r_{xy} definido em (1.17), a desigualdade de Schwarz garante que

$$-1 \leq q_{xy} \leq 1. \quad (2.15)$$

Uma regressão perfeita pela origem possui $q_{xy}^2 = 1$:

$$y_i = ax_i, \quad (2.16)$$

$$\begin{aligned} q_{xy} &= \frac{\frac{1}{n} \sum_{i=1}^n x_i a x_i}{\sqrt{\left[\frac{1}{n} \sum_{i=1}^n x_i^2 \right] \left[\frac{1}{n} \sum_{i=1}^n a^2 x_i^2 \right]}} \\ &= \frac{a m_{xx}}{\sqrt{a^2 m_{xx}^2}} = \frac{a}{\sqrt{a^2}} = \frac{a}{|a|} = \pm 1. \end{aligned} \quad (2.17)$$

Voltando a (2.12), o estimador para o erro de a , s_a , é

$$s_a = \sqrt{\frac{s_e^2}{\sum_{i=1}^n x_i^2}}. \quad (2.18)$$

Finalmente, o coeficiente de determinação, por definição, é

$$c_d = 1 - \frac{s_e^2}{s_y^2}. \quad (2.19)$$

Ao contrário do que acontece com a regressão $y = ax + b$ (à qual retornaremos na próxima aula), é notável que na regressão pela origem é possível que $c_d < 0$. Em outras palavras, é possível que o modelo $y = ax$ produza um erro padrão de estimativa *maior* que o desvio-padrão dos dados observados y_i . À luz de (2.13), fica evidente que o “coeficiente de determinação” (que denominaremos “coeficiente de quodeterminação”) natural para a regressão pela origem é

$$c_q = q_{xy}^2 = 1 - \frac{s_e^2}{m_{yy}}. \quad (2.20)$$

Tarefas:

2.1 Regressão linear através da origem por mínimos quadrados.

- Escreva uma rotina `reglina` para obter a , q_{xy} , s_e e s_a a partir de dois arrays $\mathbf{x}$ e $\mathbf{y}$.

2.2 Escreva um programa `reglinaT0xy` para ler o arquivo `compareT0.tdat` e calcular a , c_d , s_e e s_a .

2.3 Plote a reta $y = ax \pm s_e$ obtida.

Capítulo 3

Erros dos parâmetros de $y = ax + b$

Voltemos ao problema de obter a reta $y = ax + b$, mas agora incluindo os erros envolvidos em a e b .

Vamos seguir [Benjamin e Cornell \(1970, Chapter 4\)](#)

O modelo probabilístico que vamos usar é

$$Y = ax + b + \mathcal{E}, \quad (3.1)$$

onde Y é uma variável aleatória, $\mathcal{E}$ é uma variável aleatória que não depende de x e com média zero, e x por sua vez não é aleatório. Para um valor específico x_i ,

$$Y_i = ax_i + b + \mathcal{E}_i, \quad (3.2)$$

$$E\{Y_i\} = ax_i + b, \quad (3.3)$$

$$\text{Var}\{Y_i\} = \text{Var}\{\mathcal{E}_i\} = \text{Var}\{\mathcal{E}\}. \quad (3.4)$$

Em particular, e muito importante, $\text{Var}\{Y_i\}$ é *diferente* da quantidade

$$s_Y^2 = E\{(Y_i - \bar{Y})^2\}! \quad (3.5)$$

Interessantemente, parece mais fácil calcular os “erros” de a e b a partir de (1.8)–(1.9). Começemos por

$$A = \frac{n \sum x_i Y_i - (\sum x_j)(\sum Y_i)}{d_n}; \quad (3.6)$$

$$d_n = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2. \quad (3.7)$$

A boa notícia é que o denominador d_n *não* é uma variável aleatória. Isso facilita muito as coisas. Além disso,

$$\begin{aligned} d_n &= n^2 \times \frac{1}{n} \sum_{i=1}^n x_i^2 - n^2 \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \\ &= n^2 \left\{ \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right\} \\ &= n^2 s_x^2. \end{aligned} \quad (3.8)$$

Prosseguimos:

$$\begin{aligned}
A &= \frac{1}{n^2 s_x^2} \left\{ n \sum_{i=1}^n x_i Y_i - \sum_{i=1}^n \left(\sum_{j=1}^n x_j \right) Y_i \right\} \\
&= \frac{1}{n^2 s_x^2} \left\{ \sum_{i=1}^n \left[n x_i - \left(\sum_{j=1}^n x_j \right) \right] Y_i \right\}, \\
&= \frac{1}{n^2 s_x^2} \left\{ \sum_{i=1}^n [n x_i - n \bar{x}] Y_i \right\} \\
&= \frac{1}{n s_x^2} \left\{ \sum_{i=1}^n (x_i - \bar{x}) Y_i \right\} \tag{3.9}
\end{aligned}$$

O valor esperado de A é

$$\begin{aligned}
E\{A\} &= \frac{1}{n s_x^2} \sum_{i=1}^n (x_i - \bar{x}) E\{Y_i\} \\
&= \frac{1}{n s_x^2} \sum_{i=1}^n (x_i - \bar{x}) (a x_i + b) \\
&= \frac{a}{n s_x^2} \sum_{i=1}^n (x_i - \bar{x}) x_i + \underbrace{\frac{b}{n s_x^2} \sum_{i=1}^n (x_i - \bar{x})}_{=0} \\
&= \frac{a}{n s_x^2} \left[\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right] \\
&= a. \tag{3.10}
\end{aligned}$$

Portanto, A é um estimador não-enviesado de a . A variância de A é

$$\begin{aligned}
\text{Var}\{A\} &= \frac{1}{(n s_x^2)^2} \left\{ \sum_{i=1}^n (x_i - \bar{x})^2 \text{Var}\{Y_i\} \right\}, \\
&= \frac{n}{n^2 s_x^4} \left\{ \frac{1}{n} \sum_{i=1}^n (\bar{x} - x_i)^2 \text{Var}\{Y_i\} \right\}, \\
&= \frac{n}{n^2 s_x^4} \{s_x^2 \text{Var}\{Y_i\}\}, \\
&= \frac{1}{n s_x^2} \text{Var}\{Y_i\} \tag{3.11}
\end{aligned}$$

$$= \frac{1}{n s_x^2} \text{Var}\{\mathcal{E}_i\}, \tag{3.12}$$

devido a (3.4).

Uma estimativa para $\text{Var}\{\mathcal{E}_i\}$ simples de implementar computacionalmente (embora ineficiente) é o *erro padrão de estimativa de y* ,

$$s_e^2 = \frac{1}{n} \sum_{i=1}^n (y_i - (a x_i + b))^2 \tag{3.13}$$

É possível evitar o cálculo por extenso de (3.13) se já tivermos calculado outras estatísticas. Vejamos:

$$\begin{aligned}
s_e^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - (ax_i + b))^2 \\
&= \frac{1}{n} \sum_{i=1}^n \left([y_i - \bar{y}] - a[x_i - \bar{x}] + \underbrace{\bar{y} - a\bar{x} - b}_{=0} \right)^2 \\
&= \frac{1}{n} \sum_{i=1}^n ([y_i - \bar{y}] - a[x_i - \bar{x}])^2 \\
&= \frac{1}{n} \sum_{i=1}^n ([y_i - \bar{y}]^2 - 2a[x_i - \bar{x}][y_i - \bar{y}] + a^2[x_i - \bar{x}]^2) \\
&= s_y^2 - 2ac_{xy} + a^2s_x^2;
\end{aligned}$$

usando (1.15),

$$\begin{aligned}
&= s_y^2 - 2 \frac{c_{xy}^2}{s_x^2} + \frac{c_{xy}^2}{s_x^4} s_x^2 \\
&= s_y^2 - \frac{c_{xy}^2}{s_x^2 s_y^2} s_y^2 \\
&= [1 - r_{xy}^2] s_y^2.
\end{aligned} \tag{3.14}$$

A estimativa de erro de a (3.12) fica, portanto,

$$\begin{aligned}
s_a^2 &= \frac{s_e^2}{ns_x^2}, \\
s_a &= \frac{s_e}{n^{1/2}s_x}.
\end{aligned} \tag{3.15}$$

Para estimar $\text{Var}\{B\}$, é preferível partir de 1.16; note que agora, entretanto, A e B são estimadores de a e b ; portanto, escrevemos

$$B = \bar{Y} - A\bar{x}. \tag{3.16}$$

O valor esperado de B é

$$\begin{aligned}
E\{B\} &= E\{\bar{Y}\} - E\{A\}\bar{x} \\
&= a\bar{x} + b - a\bar{x} = b.
\end{aligned} \tag{3.17}$$

Portanto, B é um estimador não-enviesado de b .

Observe que $\bar{Y}$ e A não são variáveis aleatórias independentes. Em princípio,

$$\text{Cov}\{A, \bar{Y}\} = E\{A\bar{Y}\} - E\{A\}E\{\bar{Y}\}, \tag{3.18}$$

$$\begin{aligned}
E\{\bar{Y}\} &= E\left\{\frac{1}{n} \sum_{i=1}^n Y_i\right\} \\
&= \frac{1}{n} \sum_{i=1}^n E\{Y_i\}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{i=1}^n ax_i + b \\
&= a\bar{x} + b,
\end{aligned} \tag{3.19}$$

$$\begin{aligned}
E\{A\bar{Y}\} &= E\left\{\left[\frac{1}{ns_x^2} \sum_{i=1}^n (x_i - \bar{x}) Y_i\right] \left[\frac{1}{n} \sum_{j=1}^n Y_j\right]\right\} \\
&= E\left\{\left[\frac{1}{n^2 s_x^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) Y_i Y_j\right]\right\} \\
&= \frac{1}{n^2 s_x^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) E\{Y_i Y_j\} \\
&= \frac{1}{n^2 s_x^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) E\{(ax_i + b + \mathcal{E}_i)(ax_j + b + \mathcal{E}_j)\} \\
&= \frac{1}{n^2 s_x^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) \left[(ax_i + b)(ax_j + b) \right. \\
&\quad \left. + (ax_i + b) E\{\mathcal{E}_j\} + (ax_j + b) E\{\mathcal{E}_i\} + E\{\mathcal{E}_i \mathcal{E}_j\} \right] \\
&= \frac{1}{n^2 s_x^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) \left[(ax_i + b)(ax_j + b) + E\{\mathcal{E}_i \mathcal{E}_j\} \right] \\
&= \frac{1}{n^2 s_x^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}) \left[(ax_i + b)(ax_j + b) + \delta_{ij} \text{Var}\{\mathcal{E}\} \right] \\
&= \frac{1}{s_x^2} \left[\frac{1}{n} \sum_{j=1}^n (ax_j + b) \right] \left[\frac{1}{n} \sum_{i=1}^n (ax_i + b) (x_i - \bar{x}) \right] \\
&\quad + \frac{\text{Var}\{\mathcal{E}\}}{n^2 s_x^2} \sum_{i=1}^n (x_i - \bar{x}) \\
&= \frac{1}{s_x^2} [a\bar{x} + b] \left[\frac{1}{n} \sum_{i=1}^n (ax_i^2 - ax_i \bar{x} + b(x_i - \bar{x})) \right] \\
&= \frac{1}{s_x^2} [a\bar{x} + b] \left\{ a \left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x} \frac{1}{n} \sum_{i=1}^n x_i \right] \right\} \\
&= \frac{1}{s_x^2} [a\bar{x} + b] a \underbrace{\left[\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \right]}_{s_x^2} \\
&= (a\bar{x} + b) a.
\end{aligned} \tag{3.20}$$

Nós agora reunimos (3.10), (3.18), (3.19), e (3.20):

$$\begin{aligned}
\text{Cov}\{A, \bar{Y}\} &= E\{A\bar{Y}\} - E\{A\} E\{\bar{Y}\} \\
&= a(a\bar{x} + b) - a(a\bar{x} + b) = 0.
\end{aligned} \tag{3.21}$$

Portanto, apesar de não serem independentes, A e $\bar{Y}$ são não-correlacionadas.

Prosseguimos para a variância de B :

$$\begin{aligned}
\text{Var}\{B\} &= \text{Var}\{\bar{Y}\} + \bar{x}^2 \text{Var}\{A\} + 2\bar{x} \text{Cov}\{A, \bar{Y}\} \\
&= \text{Var}\{\bar{Y}\} + \bar{x}^2 \text{Var}\{A\} \\
&= \frac{1}{n} \text{Var}\{Y_i\} + \bar{x}^2 \frac{1}{ns_x^2} \text{Var}\{\mathcal{E}\} \\
&= \frac{1}{n} \text{Var}\{\mathcal{E}\} + \bar{x}^2 \frac{1}{ns_x^2} \text{Var}\{\mathcal{E}\} \\
&= \left[s_x^2 + \bar{x}^2\right] \frac{\text{Var}\{\mathcal{E}\}}{ns_x^2} \\
&= \left(\frac{1}{n} \sum_{i=1}^n x_i^2\right) \text{Var}\{A\}.
\end{aligned} \tag{3.22}$$

Tarefas:

3.1 Regressão linear por mínimos quadrados com a estimativa dos erros dos parâmetros.

- Escreva uma rotina `reglinpar` para obter a , b , r , s_e , s_a e s_b a partir de dois arrays $\mathbf{x}$ e $\mathbf{y}$.

3.2 Escreva um programa `reglinparT0xy` para ler o arquivo `compareT0.tdat` e calcular a , b , r , s_e , s_a e s_b .

3.3 Plote a reta $y = ax + b \pm s_e$ obtida. Como podemos visualizar s_a e s_b ?

Capítulo 4

Matrizes e concavidade

4.1 Transformações definidas e indefinidas

Definição 4.1 Uma transformação A é *positiva definida* quando, para qualquer $x \neq 0$,

$$x^T \cdot A \cdot x > 0.$$

Definição 4.2 Uma transformação A é *positiva semidefinida* quando, para qualquer x ,

$$x^T \cdot A \cdot x \geq 0.$$

Definição 4.3 Uma transformação A é *negativa definida* quando, para qualquer $x \neq 0$,

$$x^T \cdot A \cdot x < 0.$$

Definição 4.4 Uma transformação A é *negativa semidefinida* quando, para qualquer x ,

$$x^T \cdot A \cdot x \leq 0.$$

Quando A não é nem positiva semidefinida nem negativa semidefinida, ela é denominada *indefinida*.

O caso de transformações simétricas é particularmente interessante, pois todos os autovalores correspondentes são reais. Então, nós afirmamos, sem prova, o

Teorema 4.1 Se a transformação A é simétrica:

- A é positiva definida se e somente se todos os seus autovalores são positivos.
- A é positiva semidefinida se e somente se todos os seus autovalores são não-negativos.
- A é negativa definida se e somente se todos os seus autovalores são negativos.
- A é negativa semidefinida se e somente se todos os seus autovalores são não-positivos
- A é indefinida se e somente se possui autovalores positivos e negativos.

A título apenas de ilustração, suponha que A é simétrica e que todos os seus n autovalores λ_i sejam positivos. Então, *na base E dos autovetores*, um vetor qualquer $\mathbf{x} \neq \mathbf{0}$ do $\mathbb{R}^n$ pode ser escrito

$$[\mathbf{x}]_E = [x_1 \ x_2 \ \dots \ x_n]^\top,$$

e

$$\mathbf{x}^\top \cdot A \cdot \mathbf{x} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \dots \lambda_n x_n^2 > 0,$$

e A é positiva definida.

4.2 Máximos e mínimos de funções de mais de uma variável

Seja f uma função de n variáveis do $\mathbb{R}^n$ em $\mathbb{R}$:

$$\begin{aligned} f : \mathbb{R}^n &\rightarrow \mathbb{R} \\ (x_1, \dots, x_n) &\mapsto y = f(x_1, \dots, x_n). \end{aligned}$$

Definição 4.5 $f(x_1, \dots, x_n)$ possui um máximo em $\mathbf{x}_0 = (x_{01}, \dots, x_{0n})$ se $f(\mathbf{x}_0) > f(\mathbf{x})$ para todo $\mathbf{x}$ em uma vizinhança de $\mathbf{x}_0$.

Definição 4.6 $f(x_1, \dots, x_n)$ possui um mínimo em $\mathbf{x}_0 = (x_{01}, \dots, x_{0n})$ se $f(\mathbf{x}_0) < f(\mathbf{x})$ para todo $\mathbf{x}$ em uma vizinhança de $\mathbf{x}_0$.

Máximos ou mínimos de uma função são chamados de *extremos*. Cada ponto $\mathbf{x}_0$ correspondente é um *ponto extremo*.

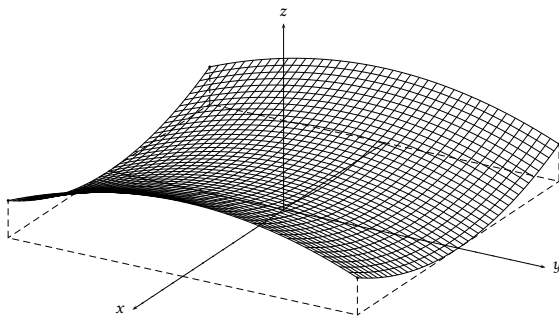
Exemplo 4.1 A função

$$f(x, y) = (200 - (x^2 + y^2))/100$$

possui um máximo em $(x_0, y_0) = (0, 0)$.

Exemplo 4.2 Considere o parabolóide hiperbólico

$$z = (100 + x^2 - y^2)/100,$$



O ponto $(0, 0)$ é um ponto crítico.

Teorema 4.2 (Piskounov, 1969, Cap. VIII, Seção 17, Teorema 1) Se $\mathbf{x}_0$ é um ponto extremo da função f , então ou $\nabla f(\mathbf{x}_0) = \mathbf{0}$, ou o $\nabla f(\mathbf{x}_0)$ não existe.

Definição 4.7 Dizemos que $\mathbf{x}_0$ é um ponto crítico de f se $\nabla f(\mathbf{x}_0) = \mathbf{0}$.

É relativamente fácil intuir “graficamente” o Teorema 4.2: na figura 4.1, note que se em um determinado ponto (x_0, y_0) ∇f existe e é $\neq \mathbf{0}$, então sempre podemos caminhar no sentido de ∇f para valores maiores de f , ou no sentido contrário para valores menores.

Note também que a nulidade ou não-existência do gradiente é uma condição necessária para a ocorrência de um extremo, porém não suficiente; e que portanto um ponto crítico pode não ser um ponto extremo. Por exemplo, a função do exemplo 4.2,

$$f(x, y) = (100 + x^2 - y^2)/100,$$

é tal que $\nabla f(0, 0) = \mathbf{0}$; porém, $(0, 0)$ não é nem um ponto de máximo nem de mínimo. Neste caso, dizemos que $(0, 0)$ é um *ponto de sela*.

Surge portanto a questão: se $\nabla f(\mathbf{x}_0) = \mathbf{0}$, e portanto $\mathbf{x}_0$ é um ponto crítico, como saber se $\mathbf{x}_0$ é um ponto de máximo, de mínimo, ou de sela?

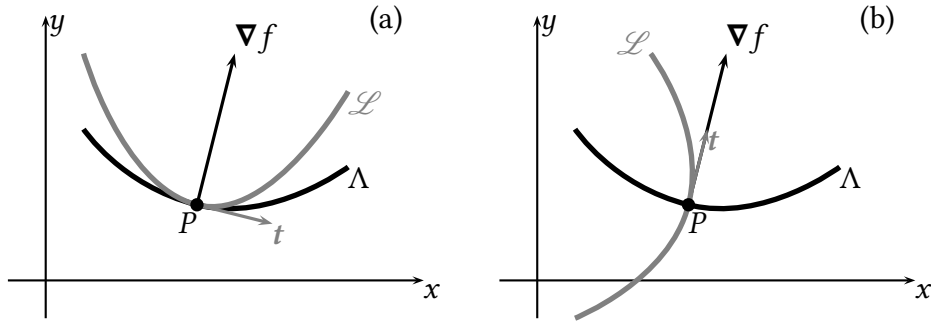


Figura 4.1: (a) o vetor t é tangente à curva de nível Λ e a curva $\mathcal{L}$ é *localmente* paralela a Λ ; (b) o vetor t é perpendicular à curva de nível Λ e a curva $\mathcal{L}$ é *localmente* perpendicular a Λ .

Suponha que f possui derivadas pelo menos até ordem 2 e considere uma expansão em série de Taylor até ordem 2 em torno de x_0 :

$$\begin{aligned} f(x_k) &= f(x_{0k}) + \frac{\partial f(x_{0k})}{\partial x_i} (x_i - x_{0i}) + \frac{1}{2!} \frac{\partial^2 f(x_{0k})}{\partial x_i \partial x_j} (x_i - x_{0i})(x_j - x_{0j}) + \dots \\ &= f(x_{0k}) + \frac{1}{2!} \frac{\partial^2 f(x_{0k})}{\partial x_i \partial x_j} (x_i - x_{0i})(x_j - x_{0j}) + \dots \end{aligned} \quad (4.1)$$

pois $\nabla f(x_0) = \mathbf{0}$.

Note que estamos utilizando sistematicamente notação indicial em (4.1). Note também que a *matriz Hessiana*

$$H_{E,ij} = \frac{\partial^2 f(x_{0k})}{\partial x_i \partial x_j} \quad (4.2)$$

é simétrica: existem n autovetores mutuamente perpendiculares, e n autovalores reais. Inicialmente, nós associamos as coordenadas x_i com a base canônica E , e enfatizamos esse fato em (4.2). Mas é muito mais simples entender as coisas na base (que também é ortogonal) F dos autovetores! Sejam agora, portanto,

$$[x_0]_F = \begin{bmatrix} \xi_{01} \\ \xi_{02} \\ \vdots \\ \xi_{0n} \end{bmatrix} \quad \text{e} \quad [x]_F = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix}$$

as coordenadas de x_0 e x na base F ; nessa base, a matriz hessiana é diagonal:

$$H_{F,ij} = \frac{\partial^2 f(\xi_{0i})}{\partial \xi_i \partial \xi_j} = \lambda_{(i)} \delta_{ij}, \quad (4.3)$$

onde os λ_i s são os autovalores da transformação H . Abusando ligeiramente da notação, utilizemos $f(\xi_1, \dots, \xi_n)$ para indicar a *a mesma função* f porém nas coordenadas da base F . Expandido em série de Taylor temos agora,

$$f(\xi_k) = f(\xi_{0k}) + \frac{1}{2!} \frac{\partial^2 f(\xi_{0k})}{\partial \xi_i \partial \xi_j} (\xi_i - \xi_{0i})(\xi_j - \xi_{0j}) + \dots$$

$$\begin{aligned}
&= f(\xi_{0k}) + \frac{1}{2!} \lambda_{(i)} \delta_{ij} (\xi_i - \xi_{0i}) (\xi_j - \xi_{0j}) + \dots \\
&= f(\xi_{0k}) + \frac{1}{2!} \lambda_{(i)} (\xi_i - \xi_{0i}) (\xi_i - \xi_{0i}) + \dots
\end{aligned}$$

Fazendo $\Delta \xi_i = (\xi_i - \xi_{0i})$ e escrevendo por extenso, portanto,

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{2!} \sum_{i=1}^n \lambda_i \Delta \xi_i^2 + \dots \quad (4.4)$$

A situação agora é bastante clara. Em uma vizinhança de $\mathbf{x}_0$ em que (4.4) seja suficientemente acurada, se todos os λ_i s forem positivos, $f(\mathbf{x}) > f(\mathbf{x}_0)$, e portanto $\mathbf{x}_0$ é um ponto de mínimo. Analogamente, se todos os λ_i s forem negativos, $f(\mathbf{x}) < f(\mathbf{x}_0)$, e portanto ξ_0 é um ponto de mínimo. Caso contrário, $\mathbf{H}$ ou é semidefinida (positiva ou negativa) ou é indefinida. Nos dois primeiros casos, existem um ou mais autovalores nulos, e $\mathbf{H}$ é dita *singular*. No último caso, $\mathbf{H}$ é indefinida. Em resumo, nós agora temos o [Lambers \(2014\)](#):

Teorema 4.3

Se $\mathbf{H}$ é positiva definida em $\mathbf{x}_0$, $\mathbf{x}_0$ é um mínimo local.

Se $\mathbf{H}$ é negativa definida em $\mathbf{x}_0$, $\mathbf{x}_0$ é um máximo local.

Se $\mathbf{H}$ possui autovalores positivos e negativos, $\mathbf{x}_0$ é um ponto de sela.

Se $\mathbf{H}$ é *singular* (isto é, se existem autovalores nulos), nada se pode informar, e $\mathbf{x}_0$ pode ser um ponto de máximo, de mínimo ou de sela, mas isso precisa ser definido por outras informações sobre f .

4.3 Multiplicadores de Lagrange

Nesta seção vamos seguir de perto, inicialmente, [Piskounov \(1969, Cap VIII, seção 18\)](#). Considere o problema de obter o máximo ou mínimo de

$$u = f(x, y) \quad (4.5)$$

sujeito à restrição

$$g(x, y) = 0. \quad (4.6)$$

Um ataque direto envolve resolver a segunda equação, obter $y = y(x)$, e substituir na primeira:

$$u(x) = f(x, y(x)).$$

Note que estamos misturando liberalmente o nome de uma função ($u(x)$) com seu valor (u). A última equação acima leva à condição necessária para a existência de um máximo ou mínimo:

$$\frac{du}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx} = 0.$$

O problema pode ser resolvido, mas requer a obtenção da função (implícita) $y = y(x)$ a partir de $g(x, y) = 0$.

O ponto de extremo pode ser obtido sem se resolver explicitamente para $y(x)$, entretanto. Note que, de (4.6), $dg = 0$. Portanto, temos duas equações

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx} = 0, \quad (4.7)$$

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \frac{dy}{dx} = 0. \quad (4.8)$$

Segue-se que para uma constante arbitrária λ

$$\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \frac{dy}{dx} \right) - \lambda \left(\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y} \frac{dy}{dx} \right) = 0, \quad (4.9)$$

$$\left(\frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} \right) + \left(\frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} \right) \frac{dy}{dx} = 0 \quad (4.10)$$

A última equação acima pode ser satisfeita independentemente de dy/dx se ambos os parênteses forem nulos. Podemos agora resolver um sistema de 3 equações em 3 incógnitas:

$$\frac{\partial f}{\partial x} - \lambda \frac{\partial g}{\partial x} = 0, \quad (4.11)$$

$$\frac{\partial f}{\partial y} - \lambda \frac{\partial g}{\partial y} = 0, \quad (4.12)$$

$$g(x, y) = 0. \quad (4.13)$$

Isso é o mesmo que formar

$$F(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

e impor

$$\frac{\partial F}{\partial x} = 0, \quad (4.14)$$

$$\frac{\partial F}{\partial y} = 0, \quad (4.15)$$

$$\frac{\partial F}{\partial \lambda} = 0. \quad (4.16)$$

λ é denominado o *multiplicador de Lagrange*.

Uma generalização é: ache o máximo/mínimo de

$$u = f(x_1, x_2, \dots, x_n) = f(x_k)$$

sujeito a

$$g_1(x_1, x_2, \dots, x_n) = g_1(x_k) = 0,$$

$$g_2(x_1, x_2, \dots, x_n) = g_2(x_k) = 0,$$

$\vdots$

$$g_m(x_1, x_2, \dots, x_n) = g_m(x_k) = 0.$$

Formamos a função

$$F(x_1, x_2, \dots, x_n, \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_k) - \lambda_l g_l(x_k),$$

e impomos

$$\frac{\partial F}{\partial x_k} = 0, \quad k = 1, 2, \dots, n, \quad (4.17)$$

$$\frac{\partial F}{\partial \lambda_l} = 0, \quad l = 1, 2, \dots, m. \quad (4.18)$$

Exemplo 4.3 Dados $x_1, x_2, \dots, x_n$ com $x_i > 0$, obtenha o máximo de

$$u = [x_1 \times x_2 \dots \times x_n]^{1/n}$$

sujeito a

$$x_1 + x_2 + \dots + x_n - a = 0.$$

SOLUÇÃO

Faça

$$F(x_k, \lambda) = [x_1 \times x_2 \dots \times x_n]^{1/n} - \lambda(x_1 + x_2 + \dots + x_n - a)$$

Agora

$$\begin{aligned} \frac{\partial F}{\partial x_k} &= \frac{1}{n} [x_1 \times x_2 \dots \times x_n]^{1/n-1} \prod_{x_l \neq x_k} x_l - \lambda \\ &= \frac{1}{n} [x_1 \times x_2 \dots \times x_n]^{1/n-1} \frac{[x_1 \times x_2 \dots \times x_n]}{x_k} - \lambda \\ &= \frac{1}{n} \frac{[x_1 \times x_2 \dots \times x_n]^{1/n}}{x_k} - \lambda = \frac{1}{n} \frac{u}{x_k} - \lambda \end{aligned}$$

e

$$\frac{\partial F}{\partial x_k} = 0 \Rightarrow \frac{1}{n} \frac{u}{x_k} = \lambda \Rightarrow x_k = \frac{u}{n\lambda}.$$

Portanto, $x_1 = \dots = x_n$; mas

$$\begin{aligned} x_1 + x_2 + \dots + x_n &= a, \\ nx_k &= a, \\ x_k &= \frac{a}{n}; \\ u &= \left[\left(\frac{a}{n} \right)^n \right]^{1/n} = \frac{a}{n} \end{aligned}$$

O máximo portanto é a/n , donde se segue a desigualdade da média aritmética-geométrica:

$$[x_1 \times x_2 \dots \times x_n]^{1/n} \leq \frac{a}{n} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad \blacksquare$$

Referências Bibliográficas

Benjamin, J. R. e Cornell, C. A. (1970). *Probability, Statistics and Decision for Civil Engineers*. McGraw-Hill, New York.

Lambers, J. V. (2014). MAT280: Multivariable Calculus. Obtido de <https://www.math.usm.edu/lambers/mat280/> em 2023-08-04.

Piskounov, N. (1969). *Differential and Integral Calculus*. MIR, Moscow.