Capítulo 11

Otimização sem restrições

Neste capítulo vamos desenvolver técnicas que permitam encontrar extremos de uma função escalar f que estejam no *interior* do conjunto admissível D.

11.1 Pontos críticos e a regra de Fermat

Em Cálculo I, um extremo local p de uma função $f\colon I \subset \mathbb{R} \to \mathbb{R}$ derivável, que também é um ponto interior de I, necessariamente é um *ponto crítico* de f, isto é, a derivada de f em p é zero: $f'(p) = 0$. Este teorema é conhecido como a *regra de Fermat* e ele se generaliza facilmente para funções escalares de várias variáveis.

Teorema 11.1 (A regra de Fermat) Sejam $f\colon D_f \subset \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^1 definida em um subconjunto D_f de $\mathbb{R}^n$ e $D \subset D_f$. Se $\mathbf{p} \in D$ é um extremo local de f em D e $\mathbf{p}$ é um *ponto interior* de D, então $\mathbf{p}$ é um *ponto crítico* de f, isto é, a matriz jacobiana de f em $\mathbf{p}$ é a matriz nula,

$$Df(\mathbf{p}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{p}) & \dfrac{\partial f}{\partial x_2}(\mathbf{p}) & \cdots & \dfrac{\partial f}{\partial x_n}(\mathbf{p}) \end{bmatrix}_{1\times n} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}_{1\times n},$$

ou, em notação vetorial,

$$\nabla f(\mathbf{p}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{p}), \frac{\partial f}{\partial x_2}(\mathbf{p}), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{p}) \right) = (0, 0, \ldots, 0),$$

isto é, o vetor gradiente de f em $\mathbf{p}$ é o vetor nulo.

Demonstração: Seja $p \in D$ um extremo local de f em D. Como p é um ponto interior de D, isto significa que podemos caminhar localmente em todas as direções a partir do ponto p sem sairmos do conjunto admissível D. Mais precisamente, dado um vetor v qualquer em $\mathbb{R}^n$, o ponto $p + t \cdot v$ pertence a D para todo t suficientemente pequeno.

Assim se, por absurdo, o gradiente $\nabla f(p)$ de f em p não é o vetor nulo, caminhando-se localmente a partir do ponto p na direção do vetor $v = \nabla f(p)$, estaremos dentro do conjunto admissível D e o valor da função f aumentará pois, pelo teorema (8.2), $\nabla f(p)$ (quando não-nulo) fornece a direção de maior crescimento de f em p. Logo p não pode ser um ponto de máximo local de f em D. Analogamente, caminhando-se localmente a partir do ponto p na direção do vetor $v = -\nabla f(p)$, estaremos dentro do conjunto admissível D e o valor da função f diminuirá pois, pelo teorema (8.2), $-\nabla f(p)$ (quando não-nulo) fornece a direção de maior decrescimento de f em p. Logo p não pode ser um ponto de mínimo local de f em D. Temos assim uma contradição com a hipótese de que p é um extremo local de f em D. Logo, o gradiente de f em p deve ser igual a zero, isto é, p deve ser um ponto crítico de f.

□

A hipótese de que o extremo local p de f seja um ponto interior de D é fundamental na regra de Fermat! Considere, por exemplo, a função $z = f(x, y) = x^2 + y^2$ e o conjunto admissível

$$D = \{(x, y) \in \mathbb{R}^2 \mid x^2/4 + y^2/9 \leq 1\}$$

(uma elipse "cheia"). Os pontos $p = (0, -3)$ e $q = (0, +3)$ são pontos de máximo global (e, portanto, pontos de máximo local) de f em D, mas os vetores gradientes de f nestes pontos, $\nabla f(p) = (0, -6)$ e $\nabla f(q) = (0, +6)$, não se anulam! Isto não contradiz a regra de Fermat pois os pontos p e q não são pontos interiores de D (veja a figura (11.1)). Por outro lado, o ponto $\mathbf{0} = (0, 0)$ é um ponto de mínimo global (e, portanto, um ponto de mínimo local) de f em D que é um ponto interior de D e, de acordo com a regra de Fermat, o vetor gradiente de f em $\mathbf{0}$ é o vetor nulo: $\nabla f(\mathbf{0}) = \mathbf{0}$.

Exercício resolvido 11.1 Diga se cada uma das sentenças a seguir é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

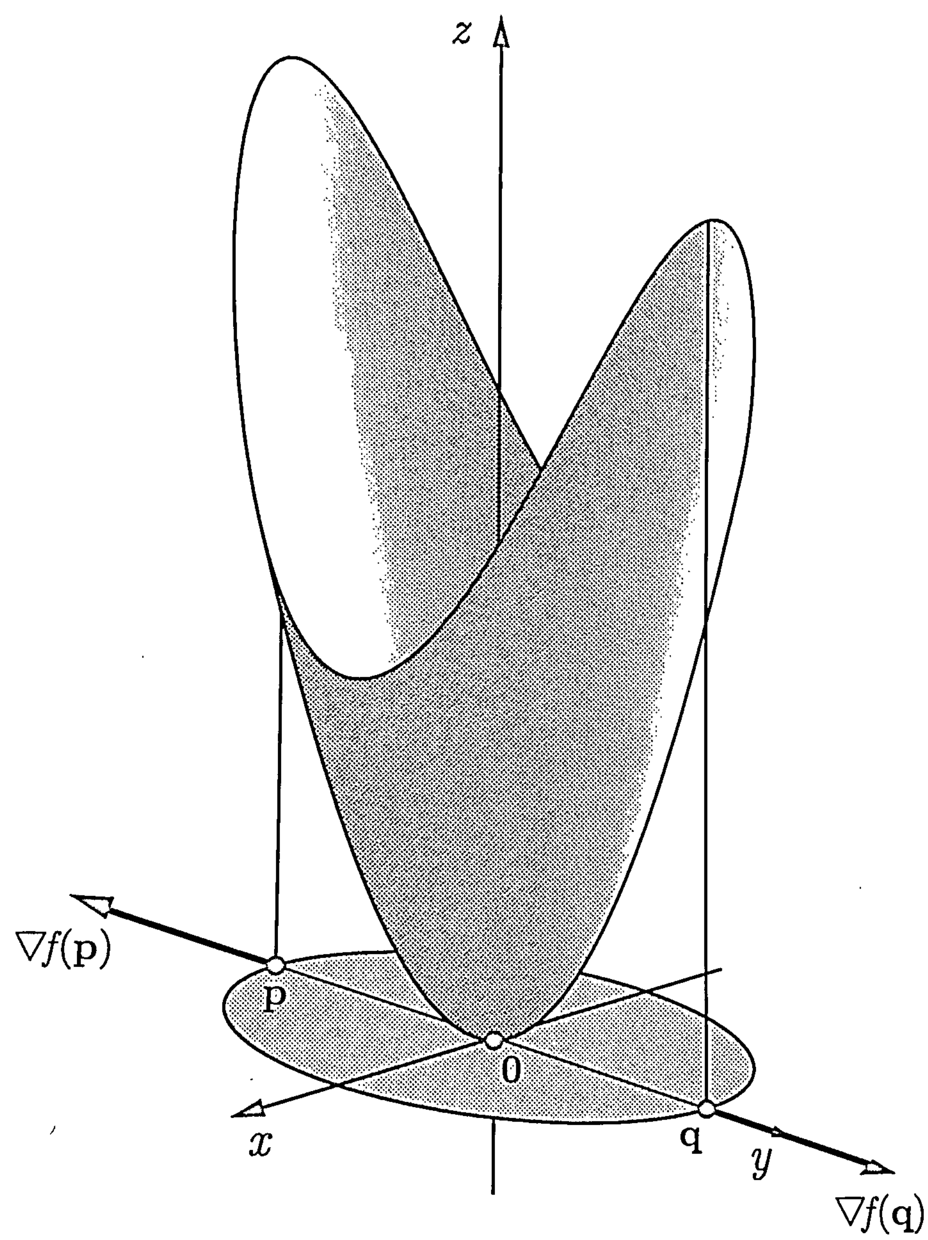

Figura 11.1: Os pontos $\mathbf{p} = (0, -3)$ e $\mathbf{q} = (0, +3)$ são extremos locais de f em $D = \{(x, y) \in \mathbb{R}^2 \mid x^2/4 + y^2/9 \leq 1\}$ mas o vetor gradiente de f em cada um destes pontos é não-nulo. Note que $\mathbf{p}$ e $\mathbf{q}$ são pontos de fronteira de D.

(a) A função $f(x,y,z) = 3\,y + \cos(x^2+z^2) - \operatorname{sen}(x^2-z^2)$ não possui extremos globais em $\mathbb{R}^3$.

(b) A função $f(x,y,z) = 3\,y + \cos(x^2+z^2) - \operatorname{sen}(x^2-z^2)$ não possui extremos globais no conjunto $D = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$.

SOLUÇÃO:

(a) Verdadeira! Como $D = \mathbb{R}^3$ é um conjunto aberto, todo ponto de D é ponto interior. Como todo extremo global é um extremo local, segue-se pela regra de Fermat que um extremo global de f deve ser um ponto crítico de f. Mas f não possui pontos críticos pois

$$\frac{\partial f}{\partial y}(x,y,z) = 3 \neq 0$$

para todo $(x,y,z) \in D = \mathbb{R}^3$. Logo, f não possui extremos globais.

(b) Falsa! O conjunto D é compacto pois D é a esfera de centro em $(0,0,0)$ e raio 1. A função f é contínua pois f é a soma, produto e composição de funções contínuas. Pelo teorema de Weierstrass, f possui pelo menos um mínimo global e pelo menos um máximo global em D. □

CUIDADO! CUIDADO! CUIDADO!

A recíproca da regra de Fermat é falsa!

A regra de Fermat afirma que todo extremo local de uma função f em um conjunto admissível D, que é um ponto interior de D, necessariamente é um ponto crítico de f. Contudo, a recíproca da regra de Fermat é falsa, isto é, nem todo ponto crítico $\mathbf{p} \in D$ de uma função f é um extremo local de f em D. Para um contra-exemplo, considere a função

$$z = f(x,y) = x^2 - y^2$$

e o conjunto admissível $D = \mathbb{R}^2$ (o gráfico de f é a sela de cavalo estudada no exemplo (3.2)). O ponto $\mathbf{p} = (0,0)$ é um ponto crítico de f pois

$$\nabla f(0,0) = \left(\frac{\partial f}{\partial x}(0,0), \frac{\partial f}{\partial y}(0,0)\right) = \left(2\,x, -2\,y\right)\Big|_{(x,y)=(0,0)} = (0,0).$$

Por outro lado, $\mathbf{p} = (0,0)$ não é um extremo local de f em $\mathbb{R}^2$. Para justificar este fato, devemos mostrar que *para toda bola aberta* $B_r(0,0)$ de centro em $(0,0)$ e raio $r > 0$, o ponto $(0,0)$ não é extremo de f em $B_r(0,0) \cap \mathbb{R}^2$, isto é, devemos mostrar que para toda bola aberta $B_r(0,0)$, existem pontos $\mathbf{q}_1 = (x_1, y_1)$ e $\mathbf{q}_2 = (x_2, y_2)$ em $B_r(0,0)$ tais que

$$f(x_1, y_1) < f(0,0) = 0 \qquad \text{e} \qquad f(x_2, y_2) > f(0,0) = 0.$$

Para este fim, considere

$$\mathbf{q}_1 = (x_1, y_1) = (0, r/2) \qquad \text{e} \qquad \mathbf{q}_2 = (x_2, y_2) = (r/2, 0).$$

Estes dois pontos pertencem à bola $B_r(0,0)$ e

$$f(x_1, y_1) = f(0, r/2) = -r^2/4 < f(0,0) = 0$$

e

$$f(x_2, y_2) = f(r/2, 0) = +r^2/4 > f(0,0) = 0.$$

Outra maneira de se ver que $(0,0)$ não é extremo local de f é observar que a função $g(x) = f(x,0) = x^2$ (a restrição de f sobre o eixo x) possui $x = 0$ como único ponto de mínimo local, de forma que $(0,0)$ não pode ser um ponto de máximo local de f. Analogamente, a função $h(y) = f(0,y) = -y^2$ (a restrição de f sobre o eixo y) possui $y = 0$ como único ponto de máximo local, de forma que $(0,0)$ não pode ser um ponto de mínimo local de f (veja a figura (11.2)).

No desenvolvimento acima, o ponto $(0,0)$ é denominado um ponto de sela de $z = f(x,y) = x^2 - y^2$. Um *ponto de sela* é um ponto crítico da função que não é nem máximo e nem mínimo local.

A regra de Fermat funciona como um grande filtro: entre todos os pontos no interior de um conjunto admissível D (freqüentemente em quantidade infinita), os candidatos a extremo local de uma função f em D são os pontos críticos de f em D, isto é, os pontos em D que anulam o vetor gradiente de f. Espera-se que a quantidade de pontos críticos (freqüentemente em quantidade finita) seja consideravelmente menor do que a quantidade de pontos interiores. Vamos ver como isto funciona com um exemplo.

Exemplo 11.1 Considere a função $z = f(x,y) = x^3 - y^3 + 9\,xy$ e o conjunto

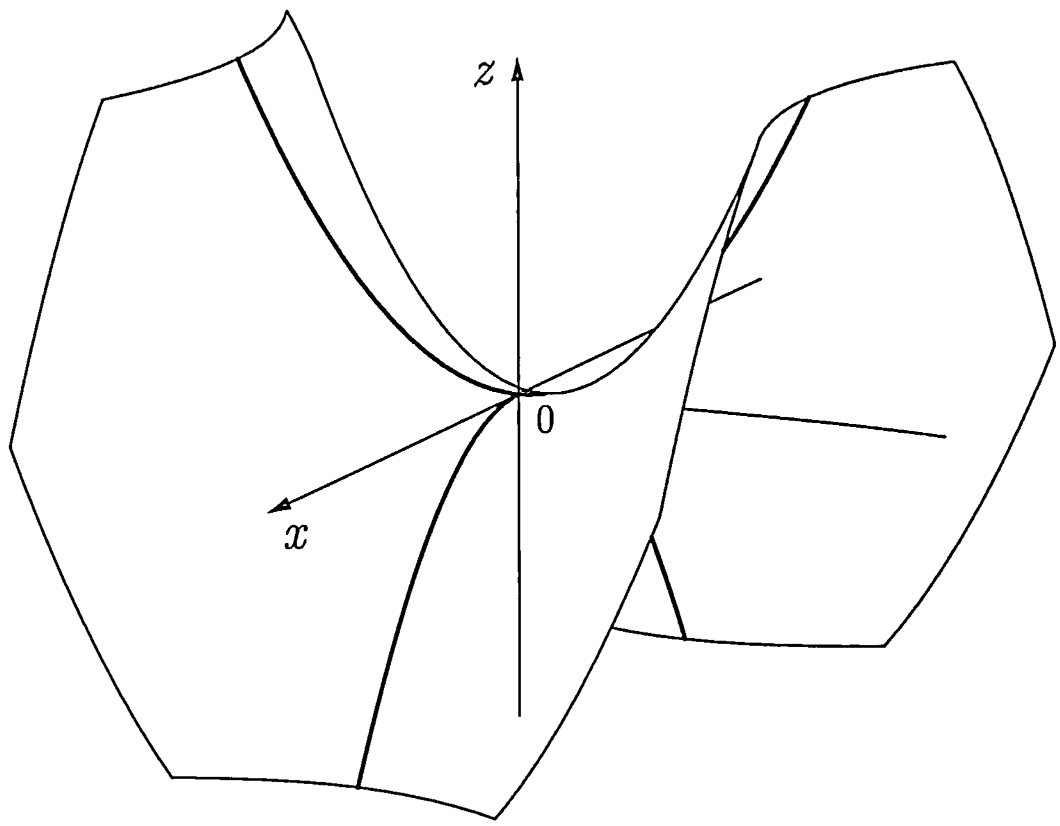

Figura 11.2: O ponto $(0,0)$ é um ponto crítico de $z = f(x,y) = x^2 - y^2$ (a sela de cavalo) mas $(0,0)$ não é um extremo local de f.

admissível $D = \mathbb{R}^2$. Observe que todos os pontos de D são pontos interiores. Se f possui um extremo local em $\mathbb{R}^2$ então, pela regra de Fermat, ele deve ser um ponto crítico de f. Os pontos críticos de f são calculados resolvendo-se a equação

$$\nabla f(x,y) = \left(\frac{\partial f}{\partial x}(x,y), \frac{\partial f}{\partial y}(x,y)\right) = \left(3\,x^2 + 9\,y, -3\,y^2 + 9\,x\right) = (0,0),$$

isto é, resolvendo-se o sistema *não linear*

$$\begin{cases} 3\,x^2 + 9\,y = 0, \\ -3\,y^2 + 9\,x = 0. \end{cases} \tag{11.1}$$

Da primeira equação temos que $y = -x^2/3$. Substituindo na segunda equação e simplificando, obtemos a equação $27\,x - x^4 = x\,(27 - x^3) = 0$, cujas soluções são $x = 0$ e $x = 3$. Como $y = -x^2/3$, concluímos que as soluções do sistema (11.1) são $(0,0)$ e $(3,-3)$. Logo, os pontos críticos de f são $(0,0)$ e $(3,-3)$. Certamente estes não são extremos globais de f em $\mathbb{R}^2$ (exercício!). Contudo, com a teoria que desenvolvemos até aqui, não sabemos se eles são ou não são extremos locais de f. ◻

Nem todo ponto crítico de uma função é um extremo local da função, como mostra o exemplo da sela de cavalo. Precisamos desenvolver uma ferramenta que nos permita "classificar" os pontos críticos de uma função f. Isto será feito com o auxílio do *polinômio de Taylor* de ordem 2 de f, assunto de nossa próxima seção.

Observação. A regra de Fermat também é denominada *condições de primeira ordem* para um extremo local.

11.2 Polinômios de Taylor

Polinômios de Taylor em Cálculo I

No capítulo 7 vimos que a "melhor" reta que aproxima o gráfico de uma função $f\colon \mathbb{R} \to \mathbb{R}$ de classe C^1 em uma vizinhança de um ponto a é dada pela reta tangente:

$$y = l(x) = f(a) + f'(a) \cdot (x - a).$$

Esta é a única reta que satisfaz $l(a) = f(a)$ e $l'(a) = f'(a)$ (l e a derivada de l coincidem com f e a derivada de f no ponto a, respectivamente) ou, equivalentemente, é a única reta que satisfaz a condição

$$\lim_{x \to a} \frac{R(a, x)}{x - a} = \lim_{x \to a} \frac{f(x) - l(x)}{x - a} = 0.$$

A equação da reta tangente define o *polinômio de Taylor* de ordem 1 de f no ponto a: $y = p_1(x) = f(a) + f'(a) \cdot (x - a)$.

Agora, podemos nos perguntar qual é a "melhor" parábola que aproxima o gráfico de uma função $f\colon \mathbb{R} \to \mathbb{R}$ de classe C^2 em uma vizinhança de um ponto a. A resposta é dada pelo próximo teorema.

Teorema 11.2 (O polinômio de Taylor de ordem 2) Considere uma função $f\colon D \subset \mathbb{R} \to \mathbb{R}$ de classe C^2 e a um ponto do interior de D. Então existe um único polinômio p_2 de grau 2 que satisfaz as condições $p_2(a) = f(a)$, $p_2'(a) = f'(a)$ e $p_2''(a) = f''(a)$.

$$y = p_2(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2!} \cdot (x - a)^2.$$

Mais ainda, vale que

$$\lim_{x \to a} \frac{f(x) - p_2(x)}{(x - a)^2} = 0,$$

isto é, a diferença entre $f(x)$ e $p_2(x)$ vai para zero mais rapidamente do que $(x - a)^2$. Em outras palavras,

$$f(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2!} \cdot (x - a)^2 + R_2(a, x),$$

com

$$\lim_{x \to a} \frac{R_2(a, x)}{(x - a)^2} = 0.$$

O polinômio p_2 é denominado o *polinômio de Taylor de ordem* 2 de f no ponto a.

Demonstração: Seja $p_2(x) = \alpha + \beta \cdot x + \gamma \cdot x^2$ um polinômio do segundo grau. Uma vez que $p_2'(x) = \beta + 2 \cdot \gamma \cdot x$ e $p_2''(x) = 2 \cdot 1 \cdot \gamma = 2! \cdot \gamma$, temos

$$p_2''(a) = f''(a) \Rightarrow 2! \cdot \gamma = f''(a) \Rightarrow \gamma = \frac{f''(a)}{2!}.$$

Com o valor de γ podemos encontrar o valor de β:

$$p_2'(a) = f'(a) \Rightarrow \beta + 2 \cdot \gamma \cdot a = f'(a) \Rightarrow$$

$$\beta + f''(a) \cdot a = f'(a) \Rightarrow \beta = f'(a) - f''(a) \cdot a.$$

Com os valores de β e γ podemos encontrar o valor de α:

$$p_2(a) = f(a) \Rightarrow \alpha + \beta \cdot a + \gamma \cdot a^2 = f(a) \Rightarrow$$

$$\alpha + (f'(a) - f''(a) \cdot a) \cdot a + \frac{f''(a)}{2!} \cdot a^2 = f(a) \Rightarrow$$

$$\alpha = f(a) - (f'(a) - f''(a) \cdot a) \cdot a - \frac{f''(a)}{2!} \cdot a^2.$$

Substituindo estes valores de α, β e γ em $p_2(x) = \alpha + \beta \cdot x + \gamma \cdot x^2$, temos

$$p_2(x) = \underbrace{f(a) - (f'(a) - f''(a) \cdot a) \cdot a - \frac{f''(a)}{2!} \cdot a^2}_{\alpha} +$$

$$\underbrace{(f'(a) - f''(a) \cdot a)}_{\beta} \cdot x + \underbrace{\frac{f''(a)}{2!}}_{\gamma} \cdot x^2.$$

Multiplicando-se as expressões acima e colocando-se $f'(a)$ e $f''(a)/2!$ em evidência, concluímos que

$$p_2(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2!} \cdot (x^2 - 2 \cdot a \cdot x + a^2) =$$

$$f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2!} \cdot (x - a)^2.$$

Finalmente, aplicando a regra de L'Hôpital duas vezes, temos

$$\lim_{x \to a} \frac{f(x) - p_2(x)}{(x - a)^2} = \lim_{x \to a} \frac{f'(x) - p_2'(x)}{2 \cdot (x - a)} =$$

$$\lim_{x \to a} \frac{f''(x) - p_2''(x)}{2} = \frac{f''(a) - p_2''(a)}{2} = 0,$$

isto é, o erro $R_2(a, x) = f(x) - p_2(x)$ vai para zero mais rapidamente do que a expressão $(x - a)^2$. □

Exemplo 11.2 Vamos calcular os polinômios de Taylor de ordem 1 e 2 da função exponencial $f(x) = e^x$ em $a = 0$. Como $f(x) = f'(x) = f''(x) = e^x$, segue-se que $f(0) = f'(0) = f''(0) = e^0 = 1$. Assim, o polinômio de Taylor de ordem 1 de f em $a = 0$ é dado por

$$y = p_1(x) = f(0) + f'(0) \cdot (x - 0) = 1 + 1 \cdot (x - 0) = 1 + x$$

e o polinômio de Taylor de ordem 2 de f em $a = 0$ é dado por

$$y = p_2(x) = f(0) + f'(0) \cdot (x - 0) + \frac{f''(0)}{2!} \cdot (x - 0)^2$$

$$= 1 + 1 \cdot (x - 0) + \frac{1}{2} \cdot (x - 0)^2 = 1 + x + \frac{x^2}{2}.$$

Podemos usar p_1 e p_2 para aproximar o valor de f em pontos próximos de $a = 0$. Por exemplo, $p_1(0.2) = 1.2$ e $p_2(0.2) = 1.22$, enquanto que

$f(0.2) = e^{0.2} = 1.22140\ldots$. Outro exemplo: $p_1(1) = 2$ e $p_2(1) = 2.5$ enquanto que $f(1) = e^1 = 2.718\ldots$. A figura (11.3) mostra o gráfico de f, p_1 e p_2. ▫

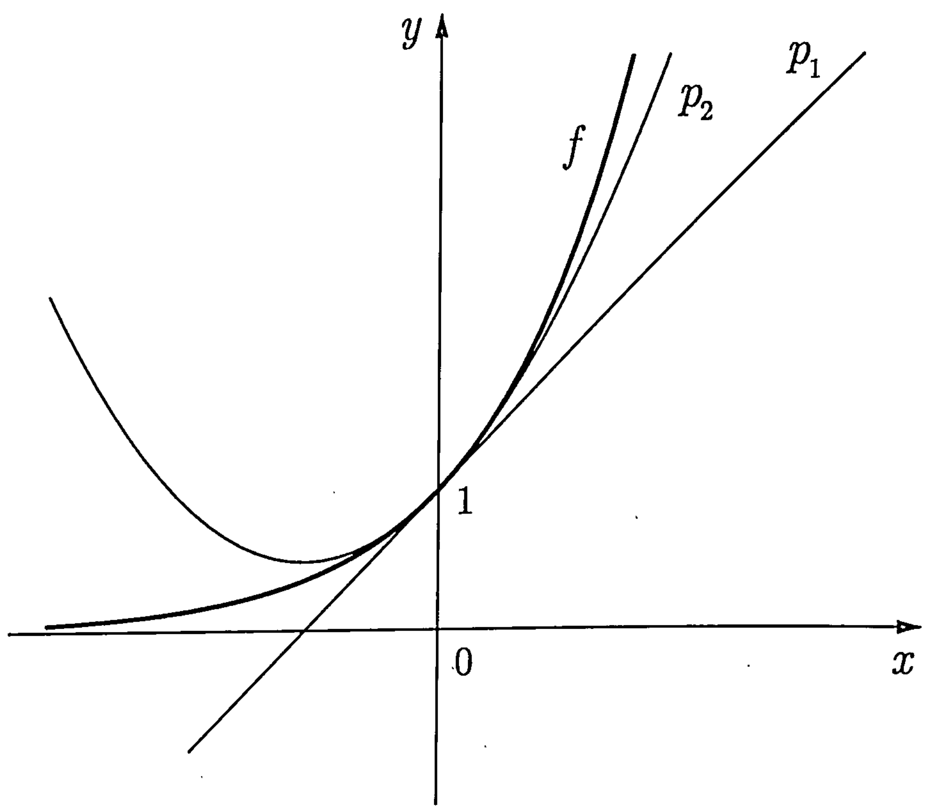

Figura 11.3: Os gráficos de $f(x) = e^x$ e dos polinômios de Taylor $p_1(x) = 1 + x$ de ordem 1 e $p_2(x) = 1 + x + x^2/2$ de ordem 2 de f no ponto $a = 0$.

O polinômio de Taylor de ordem 2 permite construir um classificador de pontos críticos de uma função de uma variável. Sejam $f\colon I \subset \mathbb{R} \to \mathbb{R}$ uma função de classe C^2 e $x = a$ um ponto crítico de f no interior do conjunto I. Do teorema (11.2), sabemos que

$$f(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2!} \cdot (x - a)^2 + R_2(a, x).$$

Uma vez que $x = a$ é ponto crítico de f, isto é, uma vez que $f'(a) = 0$, segue-se que

$$f(x) = f(a) + \frac{f''(a)}{2!} \cdot (x - a)^2 + R_2(a, x).$$

Subtraindo-se $f(a)$ dos dois lados desta equação e dividindo-se por $(x - a)^2$, o que é permitido para pontos $x \neq a$, obtemos que

$$\frac{f(x)-f(a)}{(x-a)^2}=\frac{f''(a)}{2}+\frac{R_2(a,x)}{(x-a)^2}.$$

Suponha que $f''(a)$ seja maior do que zero. Uma vez que $R_2(a,x)/(x-a)^2 \to 0$ quando $x \to a$, existe um $r > 0$ tal que

$$0<|x-a|<r \quad \Rightarrow \quad -\frac{f''(a)}{4}<\frac{R_2(a,x)}{(x-a)^2}<\frac{f''(a)}{4}.$$

Assim, para $0 < |x-a| < r$ temos $R_2(a,x)/(x-a)^2 > -f''(a)/4$ e, conseqüentemente,

$$\frac{f(x)-f(a)}{(x-a)^2}=\frac{f''(a)}{2}+\frac{R_2(a,x)}{(x-a)^2}>\frac{f''(a)}{2}-\frac{f''(a)}{4}=\frac{f''(a)}{4}>0$$

para todo x com $0 < |x-a| < r$. Isto mostra que para todo x diferente de a no intervalo aberto $(a-r, a+r)$, $f(x)-f(a) > 0$, isto é, $f(x) > f(a)$. Logo, $x = a$ é um mínimo local de f. Suponha agora que $f''(a)$ seja menor do que zero. Uma vez que $R_2(a,x)/(x-a)^2 \to 0$ quando $x \to a$, existe um $r > 0$ tal que

$$0<|x-a|<r \quad \Rightarrow \quad \frac{f''(a)}{4}<\frac{R_2(a,x)}{(x-a)^2}<-\frac{f''(a)}{4}.$$

Assim, para $0 < |x-a| < r$ temos $R_2(a,x)/(x-a)^2 < -f''(a)/4$ e, conseqüentemente,

$$\frac{f(x)-f(a)}{(x-a)^2}=\frac{f''(a)}{2}+\frac{R_2(a,x)}{(x-a)^2}<\frac{f''(a)}{2}-\frac{f''(a)}{4}=\frac{f''(a)}{4}<0$$

para todo x com $0 < |x-a| < r$. Isto mostra que para todo x diferente de a no intervalo aberto $(a-r, a+r)$, $f(x)-f(a) < 0$, isto é, $f(x) < f(a)$. Logo, $x = a$ é um máximo local de f. Estes dois resultados constituem as condições de segunda ordem de um extremo local de uma função de uma variável de classe C^2.

Teorema 11.3 (Condições de segunda ordem) Considere uma função $f\colon D \subset \mathbb{R} \to \mathbb{R}$ de classe C^2 e $x = a$ um ponto crítico de f no interior do conjunto admissível D.

(a) Se $f''(a) > 0$, então $x = a$ é um ponto de mínimo local de f.

(b) Se $f''(a) < 0$, então $x = a$ é um ponto de máximo local de f.

(c) Se $f''(a) = 0$, então $x = a$ pode ser um ponto de mínimo local, um ponto de máximo local ou um ponto de sela de f.

O item (c) do teorema acima diz que se a é um ponto crítico de f e $f''(a) = 0$, então as condições de segunda ordem não conseguem classificar o ponto crítico a. Para ver isto, considere as funções $f(x) = x^4$, $g(x) = -x^4$ e $h(x) = x^3$. O ponto $a = 0$ é ponto crítico das três funções e as derivadas de segunda ordem de f, g e h se anulam neste ponto (nesta situação, dizemos que a é um *ponto crítico degenerado* de f). Mas $a = 0$ é um ponto de mínimo local de f, um ponto de máximo local de g e um ponto de sela de h. Veja a figura (11.4).

Evidentemente, podemos considerar polinômios de Taylor de ordem maior do que 2.

Teorema 11.4 (O POLINÔMIO DE TAYLOR DE ORDEM k) Considere uma função $f\colon D \subset \mathbb{R} \to \mathbb{R}$ de classe C^{k+1} e a um ponto do interior de D. Então existe um único polinômio p_k de grau k que satisfaz as condições $p_k(a) = f(a)$, $p_k'(a) = f'(a)$, $\ldots$, $p_k^{(k)}(a) = f^{(k)}(a)$:

$$y = p_k(x) = f(a) + f'(a)\cdot(x-a) + \frac{f''(a)}{2!}\cdot(x-a)^2 + \cdots + \frac{f^{(k)}(a)}{k!}\cdot(x-a)^k.$$

Mais ainda, vale que

$$\lim_{x\to a} \frac{f(x) - p_k(x)}{(x-a)^k} = 0,$$

isto é, a diferença entre $f(x)$ e $p_k(x)$ vai para zero mais rapidamente do que $(x-a)^k$. Em outras palavras,

$$f(x) = p_k(x) + R_k(a, x)$$

com

$$\lim_{x\to a} \frac{R_k(a,x)}{(x-a)^k} = \lim_{x\to a} \frac{f(x) - p_k(x)}{(x-a)^k} = 0.$$

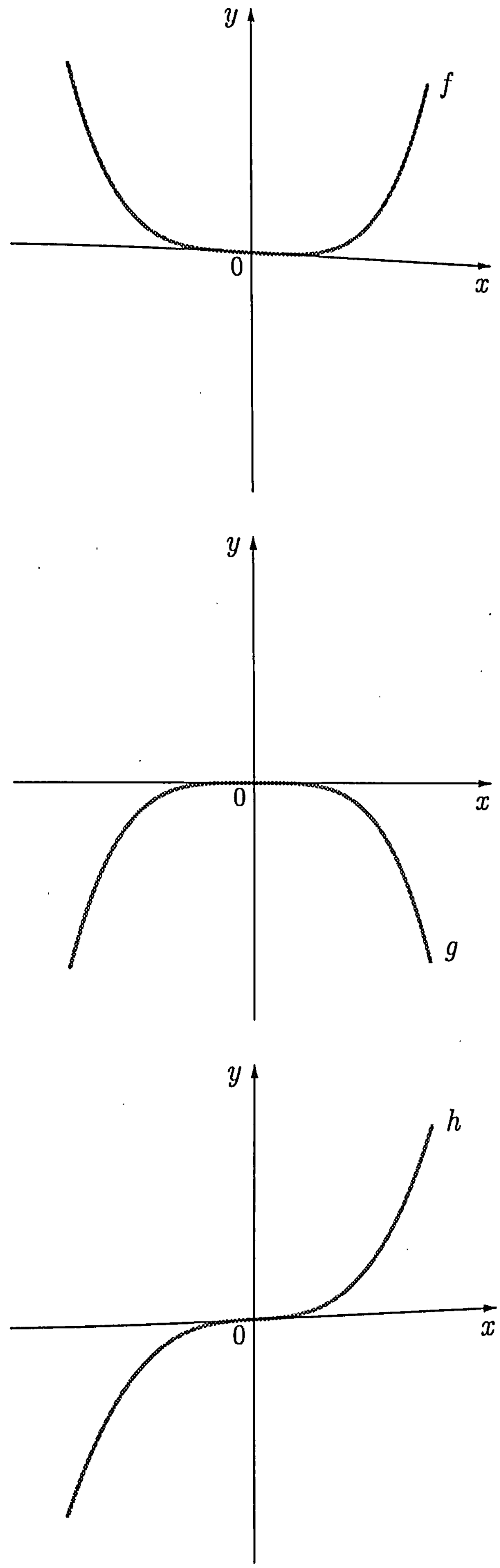

Figura 11.4: Os gráficos de $f(x) = x^4$, $g(x) = -x^4$ e $h(x) = x^3$. Observe que $f'(0) = g'(0) = h'(0) = 0$ e $f''(0) = g''(0) = h''(0) = 0$.

O polinômio p_k é denominado o *polinômio de Taylor de ordem k* de f no ponto a.

Polinômios de Taylor em Cálculo II

No capítulo 7 vimos que o "melhor" plano que aproxima o gráfico de uma função $f: \mathbb{R}^2 \to \mathbb{R}$ de classe C^1 em uma vizinhança de um ponto p é dado pelo plano tangente:

$$z = l(x, y) = f(a, b) + \frac{\partial f}{\partial x}(a, b) \cdot (x - a) + \frac{\partial f}{\partial y}(a, b) \cdot (y - b).$$

Este é o único plano (polinômio de grau um em duas variáveis) que satisfaz as condições $l(a, b) = f(a, b)$, $(\partial l/\partial x)(a, b) = (\partial f/\partial x)(a, b)$ e $(\partial l/\partial y)(a, b) = (\partial f/\partial y)(a, b)$ (l e as derivadas parciais de primeira ordem de l coincidem com f e as derivadas parciais de primeira ordem de f no ponto (a, b), respectivamente) ou, equivalentemente, é o único plano que satisfaz a condição

$$\lim_{(x,y)\to(a,b)} \frac{R(a, b, x, y)}{||(x - a, y - b)||} = \lim_{(x,y)\to(a,b)} \frac{f(x, y) - l(x, y)}{\sqrt{(x - a)^2 + (y - b)^2}} = 0.$$

A equação do plano tangente define o *polinômio de Taylor* de ordem 1 de f no ponto (a, b): $p_1(x, y) = f(a, b) + (\partial f/\partial x)(a, b) \cdot (x - a) + (\partial f/\partial y)(a, b) \cdot (y - b)$. Em notação mais compacta,

$$p_1(\mathrm{x}) = f(\mathrm{p}) + Df(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p}),$$

onde $\mathrm{x} = (x, y)$ e $\mathrm{p} = (a, b)$.

Agora, podemos nos perguntar qual é a "melhor" quádrica (polinômio de grau dois em duas variáveis) que aproxima uma função $f: \mathbb{R}^2 \to \mathbb{R}$ de classe C^2 em uma vizinhança de um ponto (a, b). A resposta é dada pelo próximo teorema.

Teorema 11.5 (O POLINÔMIO DE TAYLOR DE ORDEM 2) Considere uma função $f\colon D \subset \mathbb{R}^2 \to \mathbb{R}$ de classe C^2 e (a,b) um ponto do interior de D. Então existe um único polinômio p_2 de grau 2 que satisfaz as condições

(a) os valores das funções p_2 e f coincidem em (a,b):

$$p_2(a,b) = f(a,b),$$

(b) os valores das derivadas parciais de primeira ordem de p_2 e f coincidem em (a,b):

$$\frac{\partial p_2}{\partial x}(a,b) = \frac{\partial f}{\partial x}(a,b), \quad \frac{\partial p_2}{\partial y}(a,b) = \frac{\partial f}{\partial y}(a,b),$$

(c) os valores das derivadas parciais de segunda ordem de p_2 e f coincidem em (a,b):

$$\begin{aligned} \frac{\partial^2 p_2}{\partial x^2}(a,b) &= \frac{\partial^2 f}{\partial x^2}(a,b), & \frac{\partial^2 p_2}{\partial y^2}(a,b) &= \frac{\partial^2 f}{\partial y^2}(a,b), \\ \frac{\partial^2 p_2}{\partial y \partial x}(a,b) &= \frac{\partial^2 p_2}{\partial x \partial y}(a,b) = \frac{\partial^2 f}{\partial x \partial y}(a,b) &= \frac{\partial^2 f}{\partial y \partial x}(a,b), \end{aligned}$$

a saber, o *polinômio de Taylor de ordem* 2 de f no ponto (a, b):

$$\begin{aligned} p_2(x,y) = A + B\cdot(x-a) + C\cdot(y-b) + \\ D\cdot(x-a)^2 + E\cdot(x-a)\cdot(y-b) + F\cdot(y-b)^2, \end{aligned}$$

com

$$\begin{aligned} A &= f(a,b), & B &= \frac{\partial f}{\partial x}(a,b), & C &= \frac{\partial f}{\partial y}(a,b), \\ D &= \frac{1}{2}\cdot\frac{\partial^2 f}{\partial x^2}(a,b), & E &= \frac{\partial^2 f}{\partial x \partial y}(a,b), & F &= \frac{1}{2}\cdot\frac{\partial^2 f}{\partial y^2}(a,b). \end{aligned}$$

Mais ainda, vale que

$$\lim_{(x,y)\to(a,b)} \frac{R_2(a,b,x,y)}{\|(x-a,y-b)\|^2} = \lim_{(x,y)\to(a,b)} \frac{f(x,y)-p_2(x,y)}{(\sqrt{(x-a)^2+(y-b)^2})^2} = 0,$$

isto é, a diferença entre as funções $f(x,y)$ e $p_2(x,y)$ vai para zero mais rapidamente do que o *quadrado* da distância $\sqrt{(x-a)^2+(y-b)^2}$ entre os pontos (x,y) e (a,b). Em outras palavras, vale que

$$f(x,y) = p_2(x,y) + R_2(a,b,x,y),$$

onde o erro $R_2(a,b,x,y)$ satisfaz a condição

$$\lim_{(x,y)\to(a,b)} \frac{R_2(a,b,x,y)}{||(x-a,y-b)||^2} = \lim_{(x,y)\to(a,b)} \frac{R_2(a,b,x,y)}{(\sqrt{(x-a)^2+(y-b)^2})^2} = 0.$$

O polinômio de Taylor p_2 de f no ponto (a,b) pode ser codificado com o uso de matrizes: escrevendo $\mathrm{x}=(x,y)$ e $\mathrm{p}=(a,b)$, temos

$$p_2(\mathrm{x}) = f(\mathrm{p}) + Df(\mathrm{p})\cdot(\mathrm{x}-\mathrm{p}) + \frac{1}{2}\cdot(\mathrm{x}-\mathrm{p})^T\cdot D^2f(\mathrm{p})\cdot(\mathrm{x}-\mathrm{p}),$$

onde o símbolo ^{T} indica a operação de transposição de matrizes, $Df(\mathrm{p})$ é a matriz jacobiana de f no ponto $\mathrm{p}=(a,b)$ e $D^2f(\mathrm{p})$ é a matriz hessiana de f no ponto $\mathrm{p}=(a,b)$:

$$Df(\mathrm{p}) = \begin{bmatrix} \frac{\partial f}{\partial x}(\mathrm{p}) & \frac{\partial f}{\partial y}(\mathrm{p}) \end{bmatrix}, \quad D^2f(\mathrm{p}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(\mathrm{p}) & \frac{\partial^2 f}{\partial x\partial y}(\mathrm{p}) \\ \frac{\partial^2 f}{\partial y\partial x}(\mathrm{p}) & \frac{\partial^2 f}{\partial y^2}(\mathrm{p}) \end{bmatrix}.$$

O erro $R_2(\mathrm{p},\mathrm{x}) = f(\mathrm{x}) - p_2(\mathrm{x})$ entre as funções f e p_2 satisfaz a condição $\lim_{\mathrm{x}\to\mathrm{p}} R(\mathrm{p},\mathrm{x})/||\mathrm{x}-\mathrm{p}||^2 = 0$.

Exercício resolvido 11.2 Use os polinômios de Taylor de ordem 1 e 2 da função $z=f(x,y)=x^{1/4}y^{3/4}$ no ponto $\mathrm{p}=(1,1)$ para obter uma aproximação de $f(1.1, 0.9)$.

Solução: As derivadas parciais de primeira ordem de f são dadas por

$$\frac{\partial f}{\partial x}(x,y) = \frac{1}{4}\,x^{-3/4}y^{3/4} \quad \text{e} \quad \frac{\partial f}{\partial y}(x,y) = \frac{3}{4}\,x^{1/4}y^{-1/4},$$

e, as derivadas parciais de segunda ordem, por

$$\frac{\partial^2 f}{\partial x^2}(x,y) = -\frac{3}{16}\,x^{-7/4}y^{3/4}, \quad \frac{\partial^2 f}{\partial y^2}(x,y) = -\frac{3}{16}\,x^{1/4}y^{-5/4} \quad \text{e}$$

$$\frac{\partial^2 f}{\partial y \partial x}(x,y) = \frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{3}{16}\,x^{-3/4}y^{-1/4}.$$

Desta maneira,

$$Df(1,1) = \begin{bmatrix} \frac{\partial f}{\partial x}(1,1) & \frac{\partial f}{\partial y}(1,1) \end{bmatrix} = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \end{bmatrix}$$

e

$$D^2 f(1,1) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(1,1) & \frac{\partial^2 f}{\partial x \partial y}(1,1) \\ \frac{\partial^2 f}{\partial y \partial x}(1,1) & \frac{\partial^2 f}{\partial y^2}(1,1) \end{bmatrix} = \begin{bmatrix} -\frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & -\frac{3}{16} \end{bmatrix}.$$

Conseqüentemente,

$$p_1(x,y) = f(1,1) + Df(1,1)\cdot((x,y)-(1,1)) =$$
$$1 + \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \end{bmatrix} \cdot \begin{bmatrix} x-1 \\ y-1 \end{bmatrix} = \frac{1}{4}x + \frac{3}{4}y$$

e

$$p_2(x,y) = p_1(x,y) + \frac{1}{2}((x,y)-(1,1))^T \cdot D^2 f(1,1)\cdot((x,y)-(1,1)) =$$
$$\frac{1}{4}x + \frac{3}{4}y + \frac{1}{2}\begin{bmatrix} x-1 & y-1 \end{bmatrix} \cdot \begin{bmatrix} -\frac{3}{16} & \frac{3}{16} \\ \frac{3}{16} & -\frac{3}{16} \end{bmatrix} \cdot \begin{bmatrix} x-1 \\ y-1 \end{bmatrix} =$$
$$\frac{1}{4}x + \frac{3}{4}y - \frac{3}{32}x^2 + \frac{3}{16}xy - \frac{3}{32}y^2.$$

Sendo assim, $p_1(1.1, 0.9) = (1/4)\,(1.1) + (3/4)\,(0.9) = 0.95$ e $p_2(1.1, 0.9) = (1/4)\,(1.1) + (3/4)\,(0.9) - (3/32)\,(1.1)^2 + (3/16)\,(1.1)\,(0.9) - (3/32)\,(0.9)^2 = 0.94625$, enquanto que $f(1.1, 0.9) = (1.1)^{1/4}(0.9)^{3/4} = 0.9463026\ldots$ □

Evidentemente, podemos construir o polinômio de Taylor de ordem 2 para funções de classe C^2 que dependam de mais do que duas variáveis.

Teorema 11.6 (O POLINÔMIO DE TAYLOR DE ORDEM 2) Considere uma função $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ de classe C^2 e p um ponto do interior de D. Então existe um único polinômio p_2 de grau 2 de n variáveis que satisfaz as condições $p_2(\mathrm{p}) = f(\mathrm{p})$, $Df(\mathrm{p}) = Dp_2(\mathrm{p})$ e $D^2 f(\mathrm{p}) = D^2 p_2(\mathrm{p})$:

$$p_2(\mathrm{x}) = f(\mathrm{p}) + Df(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p}) + \frac{1}{2} \cdot (\mathrm{x} - \mathrm{p})^T \cdot D^2 f(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p}),$$

onde o símbolo ^{T} indica a operação de transposição de matrizes, $Df(\mathrm{p})$ é a matriz jacobiana de f no ponto p e

$$D^2 f(\mathrm{p}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(\mathrm{p}) & \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(\mathrm{p}) & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(\mathrm{p}) \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(\mathrm{p}) & \dfrac{\partial^2 f}{\partial x_2^2}(\mathrm{x}^*) & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n}(\mathrm{p}) \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1}(\mathrm{p}) & \dfrac{\partial^2 f}{\partial x_n \partial x_2}(\mathrm{p}) & \cdots & \dfrac{\partial^2 f}{\partial x_n^2}(\mathrm{p}) \end{bmatrix}_{n \times n}$$

é a matriz hessiana de f no ponto p. Mais ainda, vale que

$$f(\mathrm{x}) = p_2(\mathrm{x}) + R_2(\mathrm{p}, \mathrm{x}),$$

onde o erro $R_2(\mathrm{p}, \mathrm{x})$ satisfaz a propriedade

$$\lim_{\mathrm{x} \to \mathrm{p}} \frac{R_2(\mathrm{p}, \mathrm{x})}{\|\mathrm{x} - \mathrm{p}\|^2} = \lim_{\mathrm{x} \to \mathrm{p}} \frac{f(\mathrm{x}) - p_2(\mathrm{x})}{\|\mathrm{x} - \mathrm{p}\|^2} = 0,$$

isto é, o erro $R_2(\mathrm{p}, \mathrm{x})$ vai para zero mais rapidamente do que o quadrado da distância entre p e x.

Para demonstrar este teorema, ao invés de tentar determinar os coeficientes do polinômio de Taylor resolvendo-se o sistema linear obtido a partir das condições $p_2(\mathrm{p}) = f(\mathrm{p})$, $Df(\mathrm{p}) = Dp_2(\mathrm{p})$ e $D^2 f(\mathrm{p}) = D^2 p_2(\mathrm{p})$ (técnica que usamos no capítulo 7), podemos fazê-lo usando as propriedades do polinômio de Taylor de uma função de uma única variável estabelecidas na subseção anterior. Os detalhes desta demonstração podem ser encontrados nas referências [01, 71].

11.3 Formas quadráticas e matrizes definidas

Como usar o polinômio de Taylor de ordem 2 para classificar pontos críticos de uma função f de n variáveis? Como vimos, se $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ é uma função de classe C^2 e $\mathbf{p}$ é um ponto interior de D, então

$$f(\mathbf{x}) = f(\mathbf{p}) + Df(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) + \frac{1}{2} \cdot (\mathbf{x} - \mathbf{p})^T \cdot D^2 f(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) + R_2(\mathbf{p}, \mathbf{x}).$$

Se $\mathbf{p}$ é um ponto crítico de f, então $Df(\mathbf{p}) = \mathbf{0}$ e, portanto,

$$f(\mathbf{x}) = f(\mathbf{p}) + \frac{1}{2} \cdot (\mathbf{x} - \mathbf{p})^T \cdot D^2 f(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) + R_2(\mathbf{p}, \mathbf{x}).$$

Usando a variável $\mathbf{h} = \mathbf{x} - \mathbf{p}$ (que representa o deslocamento com relação ao ponto $\mathbf{p}$) e negligenciando o erro $R_2(\mathbf{p}, \mathbf{x})$ (que vai para zero rapidamente quando $\mathbf{x} \to \mathbf{p}$), concluímos que

$$f(\mathbf{p} + \mathbf{h}) \approx f(\mathbf{p}) + \frac{1}{2} \cdot \mathbf{h}^T \cdot D^2 f(\mathbf{p}) \cdot \mathbf{h},$$

isto é,

$$f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) \approx \frac{1}{2} \cdot \mathbf{h}^T \cdot D^2 f(\mathbf{p}) \cdot \mathbf{h}.$$

Se o lado direito da expressão acima é maior do que zero para todo $\mathbf{h} \neq \mathbf{0}$ suficientemente pequeno, então o lado esquerdo também deve ser maior do que zero:

$$f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) > 0 \qquad \text{ou} \qquad f(\mathbf{p} + \mathbf{h}) > f(\mathbf{p}),$$

para todo $\mathbf{h} \neq \mathbf{0}$ suficientemente pequeno. Desta maneira, $\mathbf{p}$ é um *mínimo local* de f. Analogamente, se $\mathbf{h}^T \cdot D^2 f(\mathbf{p}) \cdot \mathbf{h}$ é menor do que zero para todo $\mathbf{h} \neq \mathbf{0}$ suficientemente pequeno, então o lado esquerdo também deve ser menor do que zero:

$$f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) < 0 \qquad \text{ou} \qquad f(\mathbf{p} + \mathbf{h}) < f(\mathbf{p}),$$

para todo $\mathbf{h} \neq \mathbf{0}$ suficientemente pequeno. Desta maneira, $\mathbf{p}$ é um *máximo local* de f. *Moral da história*: a aproximação de Taylor indica que o fato de $\mathbf{p}$ ser um extremo local de f está intimamente relacionado com o sinal do polinômio de grau 2 em várias variáveis

$$Q(\mathrm{h}) = \mathrm{h}^T \cdot D^2 f(\mathrm{p}) \cdot \mathrm{h},$$

onde $D^2 f(\mathrm{p})$ é a matriz hessiana de f no ponto crítico p. Funções deste tipo são tão importantes que vamos dar um nome para elas.

Definição 11.1 (FORMAS QUADRÁTICAS) Seja A uma matriz quadrada $n \times n$ formada por números reais. A *forma quadrática* associada à matriz A é a função escalar

$$\begin{array}{rccl} Q\colon & \mathbb{R}^n & \to & \mathbb{R} \\ & \mathrm{h} & \mapsto & Q(\mathrm{h}) = \mathrm{h}^T \cdot A \cdot \mathrm{h} \end{array},$$

isto é,

$$Q(h_1, h_2, \ldots, h_n) = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{1,2} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix},$$

ou ainda,

$$Q(h_1, h_2, \ldots, h_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} h_i h_j$$

onde, como sempre, estamos identificando matrizes 1×1 com números reais.

Como o sinal de $Q(\mathrm{h}) = \mathrm{h}^T \cdot A \cdot \mathrm{h}$ também desempenha um papel importante, somos levados a considerar a seguinte definição.

Definição 11.2 (MATRIZES DEFINIDAS E SEMIDEFINIDAS) Sejam A uma matriz $n \times n$ e $Q(\mathrm{h}) = \mathrm{h}^T \cdot A \cdot \mathrm{h}$ a forma quadrática associada.

(a) A matriz A (ou a forma quadrática Q) é *positiva definida* se

$$Q(\mathrm{h}) = \mathrm{h}^T \cdot A \cdot \mathrm{h} > 0$$

para todo $\mathrm{h} \neq \mathbf{0}$ em $\mathbb{R}^n$.

(b) A matriz A (ou a forma quadrática Q) é *positiva semidefinida* se

$$Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h} \geq 0$$

para todo $\mathbf{h} \neq \mathbf{0}$ em $\mathbb{R}^n$.

(c) A matriz A (ou a forma quadrática Q) é *negativa definida* se

$$Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h} < 0$$

para todo $\mathbf{h} \neq \mathbf{0}$ em $\mathbb{R}^n$.

(d) A matriz A (ou a forma quadrática Q) é *negativa semidefinida* se

$$Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h} \leq 0$$

para todo $\mathbf{h} \neq \mathbf{0}$ em $\mathbb{R}^n$.

(e) A matriz A (ou a forma quadrática Q) é *indefinida* se existem $\mathbf{h}$ e $\mathbf{k}$ em $\mathbb{R}^n$ tais que

$$Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h} > 0 \quad \text{e} \quad Q(\mathbf{k}) = \mathbf{k}^T \cdot A \cdot \mathbf{k} < 0.$$

Observe que toda forma quadrática $Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h}$ satisfaz

$$Q(\mathbf{0}) = \mathbf{0}^T \cdot A \cdot \mathbf{0} = 0.$$

Exemplo 11.3 A matriz

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

é positiva definida pois a forma quadrática associada,

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = h_1^2 + h_2^2,$$

é maior do que 0 para todo $(h_1, h_2) \neq (0, 0)$. Veja a figura (11.5). □

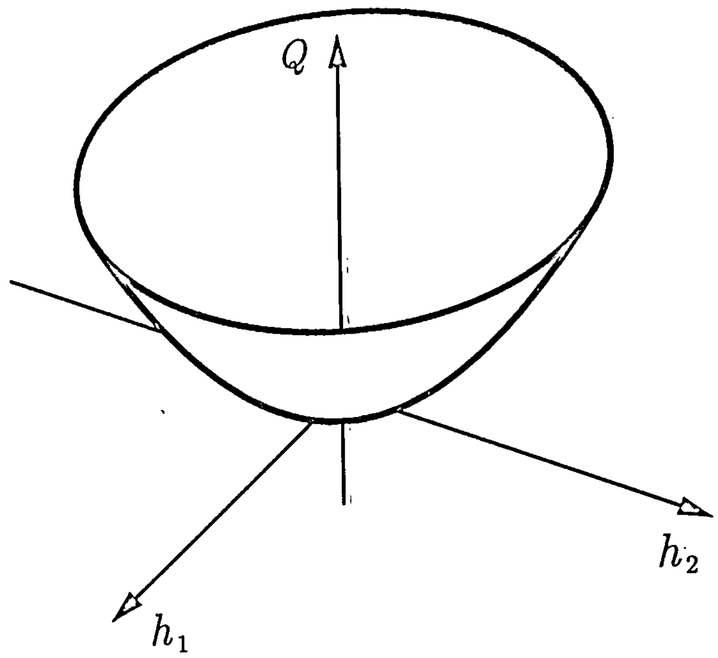

Figura 11.5: Gráfico da forma quadrática positiva definida $Q(h_1, h_2) = h_1^2 + h_2^2$.

Exemplo 11.4 A matriz

$$B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

é negativa definida pois a forma quadrática associada,

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = -h_1^2 - h_2^2,$$

é menor do que 0 para todo $(h_1, h_2) \neq (0,0)$. Veja a figura (11.6). □

Exemplo 11.5 A matriz

$$C = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

é indefinida pois a forma quadrática associada,

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = h_1^2 - h_2^2,$$

é tal que $Q(1,0) = 1 > 0$ e $Q(0,1) = -1 < 0$. □

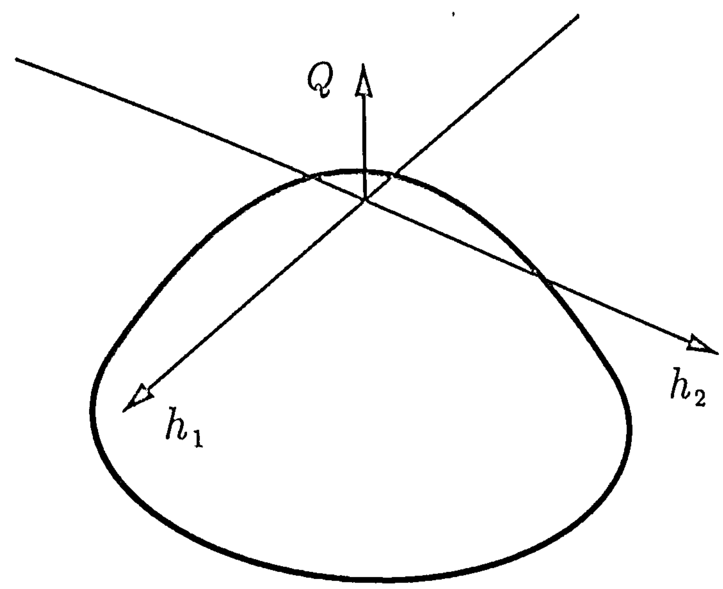

Figura 11.6: Gráfico da forma quadrática negativa definida $Q(h_1, h_2) = -h_1^2 - h_2^2$.

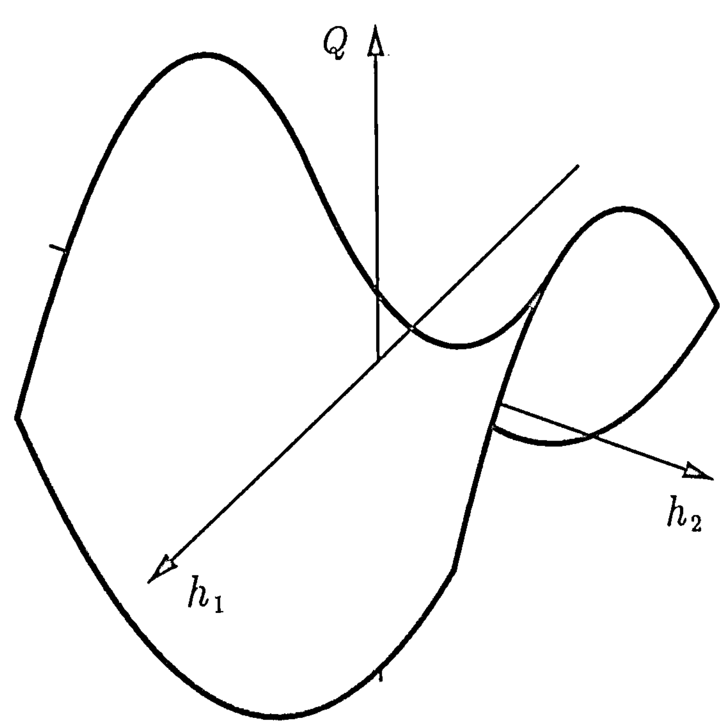

Figura 11.7: Gráfico da forma quadrática indefinida $Q(h_1, h_2) = h_1^2 - h_2^2$.

Exemplo 11.6 A matriz

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

é positiva semidefinida pois a forma quadrática associada,

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = h_1^2,$$

é maior ou igual a 0 para todo $(h_1, h_2) \neq (0,0)$. Veja a figura (11.8). Observe que D não é positiva definida pois $Q(0,1) = 0$. ▫

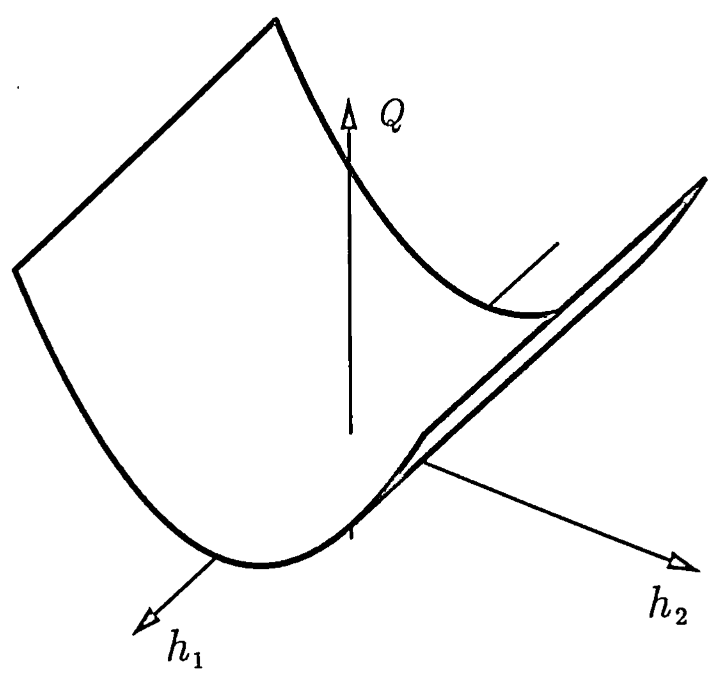

Figura 11.8: Gráfico da forma quadrática positiva semidefinida $Q(h_1, h_2) = h_1^2$.

Exemplo 11.7 A matriz

$$E = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}$$

é negativa semidefinida pois a forma quadrática associada,

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = -h_1^2,$$

é menor ou igual a 0 para todo $(h_1, h_2) \neq (0,0)$. Veja a figura (11.9). Observe que E não é negativa definida pois $Q(0,1) = 0$.

□

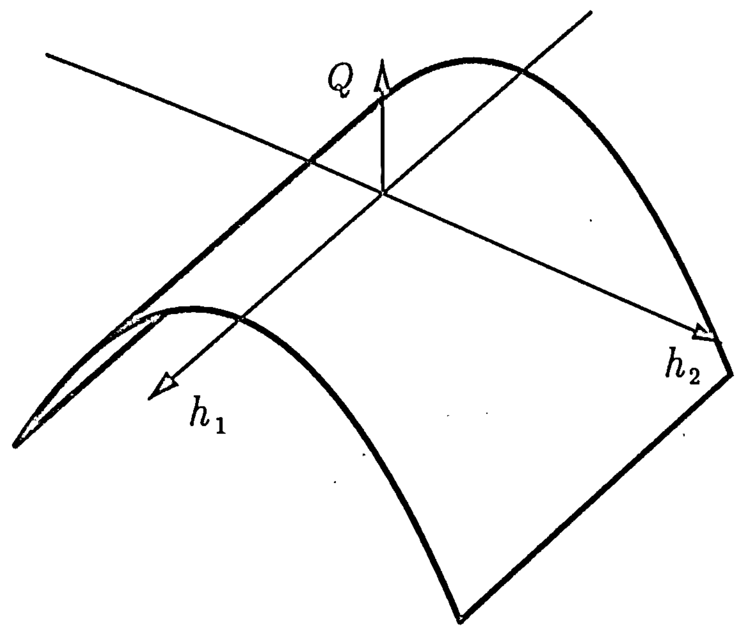

Figura 11.9: Gráfico da forma quadrática negativa semidefinida $Q(h_1, h_2) = -h_1^2$.

Exercício resolvido 11.3 Considere a matriz diagonal

$$D = \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix}.$$

Para quais valores de d_1, d_2, d_3 e d_4 a matriz D é positiva definida, negativa definida, positiva semidefinida, negativa semidefinida e indefinida?

SOLUÇÃO: A forma quadrática associada a matriz D é dada por

$$\begin{aligned} Q(h_1, h_2, h_3, h_4) &= \begin{bmatrix} h_1 & h_2 & h_3 & h_4 \end{bmatrix} \cdot \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} \\ &= d_1\, h_1^2 + d_2\, h_2^2 + d_3\, h_3^2 + d_4\, h_4^2. \end{aligned}$$

Sejam $\mathbf{e}_1 = (1,0,0,0)$, $\mathbf{e}_2 = (0,1,0,0)$, $\mathbf{e}_3 = (0,0,1,0)$ e $\mathbf{e}_4 = (0,0,0,1)$ os vetores da base canônica de $\mathbb{R}^4$.

(a) Se D é uma matriz positiva definida, então devemos ter $Q(\mathbf{e}_1) = d_1 > 0$, $Q(\mathbf{e}_2) = d_2 > 0$, $Q(\mathbf{e}_3) = d_3 > 0$ e $Q(\mathbf{e}_4) = d_4 > 0$. Por outro lado, se d_1, d_2, d_3 e d_4 são todos maiores do que 0, então D é positiva definida pois

$$Q(h_1, h_2, h_3, h_4) = d_1\, h_1^2 + d_2\, h_2^2 + d_3\, h_3^2 + d_4\, h_4^2 \geq 0$$

para todo $(h_1, h_2, h_3, h_4) \in \mathbb{R}^4$ e

$$Q(h_1, h_2, h_3, h_4) = 0 \quad \text{se, e somente se,} \quad h_1 = h_2 = h_3 = h_4 = 0.$$

(b) Se D é uma matriz negativa definida, então devemos ter $Q(\mathbf{e}_1) = d_1 < 0$, $Q(\mathbf{e}_2) = d_2 < 0$, $Q(\mathbf{e}_3) = d_3 < 0$ e $Q(\mathbf{e}_4) = d_4 < 0$. Por outro lado, se d_1, d_2, d_3 e d_4 são todos menores do que 0, então D é negativa definida pois

$$Q(h_1, h_2, h_3, h_4) = d_1\, h_1^2 + d_2\, h_2^2 + d_3\, h_3^2 + d_4\, h_4^2 \leq 0$$

para todo $(h_1, h_2, h_3, h_4) \in \mathbb{R}^4$ e

$$Q(h_1, h_2, h_3, h_4) = 0 \quad \text{se, e somente se,} \quad h_1 = h_2 = h_3 = h_4 = 0.$$

(c) Se D é uma matriz positiva semidefinida, então devemos ter $Q(\mathbf{e}_1) = d_1 \geq 0$, $Q(\mathbf{e}_2) = d_2 \geq 0$, $Q(\mathbf{e}_3) = d_3 \geq 0$ e $Q(\mathbf{e}_4) = d_4 \geq 0$. Por outro lado, se d_1, d_2, d_3 e d_4 são todos maiores ou iguais a 0, então D é positiva semidefinida pois

$$Q(h_1, h_2, h_3, h_4) = d_1\, h_1^2 + d_2\, h_2^2 + d_3\, h_3^2 + d_4\, h_4^2 \geq 0$$

para todo $(h_1, h_2, h_3, h_4) \in \mathbb{R}^4$.

(d) Se D é uma matriz negativa semidefinida, então devemos ter $Q(\mathbf{e}_1) = d_1 \leq 0$, $Q(\mathbf{e}_2) = d_2 \leq 0$, $Q(\mathbf{e}_3) = d_3 \leq 0$ e $Q(\mathbf{e}_4) = d_4 \leq 0$. Por outro lado, se d_1, d_2, d_3 e d_4 são todos menores ou iguais a 0, então D é negativa semidefinida pois

$$Q(h_1, h_2, h_3, h_4) = d_1\, h_1^2 + d_2\, h_2^2 + d_3\, h_3^2 + d_4\, h_4^2 \leq 0$$

para todo $(h_1, h_2, h_3, h_4) \in \mathbb{R}^4$.

(e) Se D é uma matriz indefinida, então algum d_i deve ser menor do que zero pois, se todos os d_i forem maiores ou iguais a zero, então D seria uma matriz positiva semidefinida e não uma matriz indefinida. Analogamente, algum d_j deve ser maior do que zero pois, se todos os d_j forem menores ou iguais a zero, então D seria uma matriz negativa semidefinida e não uma matriz indefinida. Resumindo, se D é uma matriz indefinida, então dois elementos da diagonal de D devem ser diferentes de zero e devem possuir sinais contrários. Por outro lado, se dois elementos da diagonal de D são diferentes de zero e possuem sinais contrários, então D é uma matriz indefinida pois se $d_i < 0$ e $d_j > 0$, então $Q(\mathrm{e}_i) = d_i < 0$ e $Q(\mathrm{e}_j) = d_j > 0$.

□

Com estas definições podemos, finalmente, enunciar o teorema de classificação de pontos críticos, baseado na positividade da matriz hessiana.

Teorema 11.7 (Condições de segunda ordem) Considere uma função $f: D \subset \mathbb{R}^n \to \mathbb{R}$ de classe C^2 e $\mathbf{p}$ um ponto crítico de f no interior do conjunto D.

(a) Se $D^2f(\mathbf{p})$ é uma matriz positiva definida, então $\mathbf{p}$ é um ponto de mínimo local de f.

(b) Se $D^2f(\mathbf{p})$ é uma matriz negativa definida, então $\mathbf{p}$ é um ponto de máximo local de f.

(c) Se $D^2f(\mathbf{p})$ é uma matriz indefinida, então $\mathbf{p}$ é um ponto de sela de f, isto é, $\mathbf{p}$ é um ponto crítico que não é mínimo local e nem máximo local de f.

(d) Se $D^2f(\mathbf{p})$ é uma matriz positiva semidefinida ou uma matriz negativa semidefinida, então $\mathbf{p}$ é um ponto de máximo local, um ponto de mínimo local ou um ponto de sela de f.

O item (d) do teorema acima diz que se $\mathbf{p}$ é um ponto crítico de f e a matriz hessiana $D^2f(\mathbf{p})$ de f em $\mathbf{p}$ é positiva semidefinida ou negativa semidefinida, então as condições de segunda ordem não conseguem classificar o ponto crítico $\mathbf{p}$. Para ver isto, considere as funções $f(x, y) = x^4 + y^4$, $g(x, y) = -x^4 - y^4$ e $h(x) = x^4 - y^4$. O ponto $\mathbf{p} = (0, 0)$ é ponto crítico das três funções e a matriz hessiana destas três funções em $\mathbf{p} = (0, 0)$ é a matriz nula que é uma matriz positiva semidefinida e uma matriz negativa semidefinida ao mesmo tempo (nesta situação, dizemos que $\mathbf{p}$ é um *ponto crítico degenerado* de f). Mas $\mathbf{p} = (0, 0)$ é um ponto de mínimo local de f, um ponto de máximo local de g e um ponto de sela de h (por que?). Uma demonstração do teorema das condições de segunda ordem pode ser encontrada na referência [71].

Exemplo 11.8 No exemplo (11.1) vimos que os únicos pontos críticos da função $z = f(x, y) = x^3 - y^3 + 9\,xy$ no conjunto admissível $D = \mathbb{R}^2$ são $\mathrm{p}_1 = (0, 0)$ e $\mathrm{p}_2 = (3, -3)$. Como todos os pontos do conjunto admissível são pontos interiores, segue-se que p_1 e p_2 são os únicos candidatos a extremo local de f. Vamos aplicar as condições de segunda ordem para tentar classificar estes pontos críticos. Para isto, vamos calcular a matriz hessiana $D^2f(x, y)$ de f no ponto (x, y):

$$D^2f(x,y) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2}(x,y) & \dfrac{\partial^2 f}{\partial x \partial y}(x,y) \\ \dfrac{\partial^2 f}{\partial y \partial x}(x,y) & \dfrac{\partial^2 f}{\partial y^2}(x,y) \end{bmatrix} = \begin{bmatrix} 6\,x & 9 \\ 9 & -6\,y \end{bmatrix}.$$

(a) Classificação do ponto crítico $\mathrm{p}_1 = (0, 0)$.

A matriz hessiana $D^2f(0, 0)$ de f no ponto $(0, 0)$ é dada por

$$D^2f(0,0) = \begin{bmatrix} 0 & 9 \\ 9 & 0 \end{bmatrix}$$

e a correspondente forma quadrática por

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} 0 & 9 \\ 9 & 0 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = 18\,h_1 h_2.$$

Devemos agora estudar o sinal de $Q(h_1, h_2)$. Se conseguirmos dois pontos (h_1, h_2) e (k_1, k_2) para os quais $Q(h_1, h_2) > 0$ e $Q(k_1, k_2) < 0$, então $D^2f(0, 0)$ será uma matriz indefinida e, conseqüentemente, $(0, 0)$ será um ponto de sela de f. Por outro lado, conseguir alguns pontos onde Q é maior do que 0 não garante que $D^2f(0, 0)$ seja uma matriz positiva definida. Analogamente, conseguir alguns pontos onde Q é menor do que 0 não garante que $D^2f(0, 0)$ seja uma matriz negativa definida. Contudo, neste caso, é fácil de se classificar a matriz hessiana $D^2f(0, 0)$ de f no ponto $(0, 0)$. Uma vez que

$$Q(1, 1) = 18 > 0 \quad \text{e} \quad Q(1, -1) = -18 < 0$$

segue-se que $D^2f(0,0)$ é uma matriz indefinida e, portanto, pelas condições de segunda ordem, $(0,0)$ é um ponto de sela de f, isto é, $(0,0)$ é um ponto crítico de f que não é mínimo local e nem máximo local de f.

(b) Classificação do ponto crítico $\mathbf{p}_2 = (3,-3)$.

A matriz hessiana $D^2f(3,-3)$ de f no ponto $(3,-3)$ é dada por

$$D^2f(3,-3) = \begin{bmatrix} 18 & 9 \\ 9 & 18 \end{bmatrix}$$

e a correspondente forma quadrática por

$$Q(h_1,h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} 18 & 9 \\ 9 & 18 \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = 18\,h_1^2 + 18\,h_1h_2 + 18\,h_2^2.$$

Devemos agora estudar o sinal de $Q(h_1,h_2)$. Se conseguirmos dois pontos (h_1,h_2) e (k_1,k_2) para os quais $Q(h_1,h_2) > 0$ e $Q(k_1,k_2) < 0$, então $D^2f(3,-3)$ será uma matriz indefinida e, conseqüentemente, $(3,-3)$ será um ponto de sela de f. Por outro lado, conseguir alguns pontos onde Q é maior do que 0 não garante que $D^2f(3,-3)$ seja uma matriz positiva definida. Analogamente, conseguir alguns pontos onde Q é menor do que 0 não garante que $D^2f(3,-3)$ seja uma matriz negativa definida. Analisar o sinal de $18\,h_1^2 + 18\,h_1h_2 + 18\,h_2^2$ é mais complicado do que analisar o sinal de $18\,h_1h_2$!

Para classificar a matriz hessiana $D^2f(3,-3)$ de f no ponto $(3,-3)$ vamos usar o *método de Lagrange*, que consiste em *completar os quadrados* na expressão para $Q_(h_1,h_2)$:

$$Q(h_1,h_2) = 18\left[h_1^2 + h_1h_2 + h_2^2\right] =$$

$$18\left[\left(h_1 + \frac{1}{2}\,h_2\right)^2 - \frac{1}{4}\,h_2^2 + h_2^2\right] = 18\left[\left(h_1 + \frac{1}{2}\,h_2\right)^2 + \frac{3}{4}\,h_2^2\right].$$

Observe que $Q(h_1,h_2) \geq 0$ para todo $(h_1,h_2) \in \mathbb{R}^2$ e $Q(h_1,h_2) = 0$ se, e somente se, $h_1 = h_2 = 0$. Isto mostra que $D^2f(3,-3)$ é uma matriz positiva definida e, portanto, pelas condições de segunda ordem, $(3,-3)$ é um ponto de mínimo local de f. □

CUIDADO! CUIDADO! CUIDADO!

As condições de segunda ordem estabelecem um critério para decidir se um ponto crítico é um extremo *local*! Elas nada afirmam sobre a *globalidade* do ponto crítico! No exemplo anterior, $(3, -3)$ é um mínimo local de

$$z = f(x, y) = x^3 - y^3 + 9\,xy$$

mas $(3, -3)$ não é um mínimo global de f em $\mathbb{R}^2$. Para vê-lo, basta observar que $f(0, n) = -n^3$ que vai para $-\infty$ quando n vai para $+\infty$.

Existem várias maneiras de se estudar o sinal de uma forma quadrática

$$Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h}$$

associada a uma matriz simétrica A:

(a) com o completamento de quadrados de $Q(\mathbf{h})$ (conhecido como o método de Lagrange),

(b) com o cálculo dos autovalores de A, isto é, com cálculo das raízes do polinômio característico de A,

(c) com o cálculo dos determinantes de certas submatrizes de A (os assim denominados menores principais de A) e

(d) com a decomposição de Cholesky de A (relacionada com a decomposição LU de A).

No exemplo (11.8) usamos o método de Lagrange para classificar o ponto crítico $(3, -3)$ da função $z = f(x, y) = x^3 - y^3 + 9\,xy$.

A decomposição de Cholesky constitui o método computacionalmente mais eficiente para se determinar a positividade de uma matriz. Contudo, gastaríamos muito tempo para estabelecer os pré-requisitos de Álgebra Linear necessários para o entendimento da decomposição de Cholesky. Ao invés disto, vamos descrever como classificar a forma quadrática $Q(\mathbf{h}) = \mathbf{h}^T \cdot A \cdot \mathbf{h}$ em termos de determinantes, o que será satisfatório para nossos propósitos neste curso.

11.4 Menores de uma matriz

Definição 11.3 (Menor de uma matriz) Seja A uma matriz $n \times n$. Um *menor* de ordem k é o determinante de uma submatriz $k \times k$ de A obtida removendo-se $n-k$ linhas e $n-k$ colunas de A.

Exemplo 11.9 Considere a matriz 3×3

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Os menores de ordem 1 da matriz A são dados por

$$|a|, \quad |b|, \quad |c|, \quad |d|, \quad |e|, \quad |f|, \quad |g|, \quad |h| \quad \text{e} \quad |i|.$$

Os menores de ordem 2 da matriz A são dados por

$$\begin{vmatrix} e & f \\ h & i \end{vmatrix}, \quad \begin{vmatrix} d & f \\ g & i \end{vmatrix}, \quad \begin{vmatrix} d & e \\ g & h \end{vmatrix},$$

$$\begin{vmatrix} b & c \\ h & i \end{vmatrix}, \quad \begin{vmatrix} a & c \\ g & i \end{vmatrix}, \quad \begin{vmatrix} a & b \\ g & h \end{vmatrix},$$

$$\begin{vmatrix} b & c \\ e & f \end{vmatrix}, \quad \begin{vmatrix} a & c \\ d & f \end{vmatrix}, \quad \begin{vmatrix} a & b \\ d & e \end{vmatrix}.$$

O menor de ordem 3 da matriz A é dado por

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}.$$

O símbolo $|A|$ denota o determinante da matriz A (não confunda com o módulo de um número real). □

Definição 11.4 (MENOR PRINCIPAL DE UMA MATRIZ) Um *menor principal* de ordem k é um menor de ordem k no qual *as mesmas* linhas e colunas foram removidas. Mais especificamente, se as linhas $i_1, \ldots, i_{n-k}$ foram removidas, então as colunas removidas são as de índices $i_1, \ldots, i_{n-k}$.

Exemplo 11.10 Considere a matriz 3×3

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Os menores principais de ordem 1 da matriz A são dados por

$$|a|, \quad |e| \quad \text{e} \quad |i|.$$

Os menores principais de ordem 2 da matriz A são dados por

$$\begin{vmatrix} e & f \\ h & i \end{vmatrix}, \quad \begin{vmatrix} a & c \\ g & i \end{vmatrix}, \quad \begin{vmatrix} a & b \\ d & e \end{vmatrix}.$$

O menor principal de ordem 3 da matriz A é dado por

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}.$$

O símbolo $|A|$ denota o determinante da matriz A (não confunda com o módulo de um número real). ▫

Definição 11.5 Um *menor principal líder* de ordem k é um menor principal de ordem k onde foram removidas as *últimas* $n-k$ linhas e $n-k$ colunas.

Exemplo 11.11 Considere a matriz 3×3

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}.$$

Os menores principais líderes de ordem 1, 2 e 3 de A são dados, respectivamente, por

$$|a|, \qquad \begin{vmatrix} a & b \\ d & e \end{vmatrix} \qquad \text{e} \qquad \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}.$$

O símbolo $|A|$ denota o determinante da matriz A (não confunda com o módulo de um número real). □

Teorema 11.8 (Classificação de formas quadráticas via menores principais) Seja A uma matriz $n \times n$ simétrica. Denote por $|A_k|$ o *menor principal líder* de ordem k de A.

(a) A é uma matriz positiva definida se, e somente se, *todos os menores principais líderes* $|A_1|, \ldots, |A_n|$ de A são maiores do que zero.

(b) A é uma matriz negativa definida se, e somente se, *todos os menores principais líderes* de ordem *ímpar* são menores do que zero e *todos os menores principais líderes* de ordem *par* são maiores do que zero.

(c) A é uma matriz positiva semidefinida se, e somente se, *todos os menores principais* de A são maiores ou iguais a zero.

(d) A é uma matriz negativa semidefinida se, e somente se, *todos os menores principais* de ordem *ímpar* de A são menores ou iguais a zero e *todos os menores principais* de ordem *par* de A são maiores ou iguais a zero.

(e) Se A não se enquadra em nenhum dos quatro casos anteriores, então A é uma matriz indefinida.

Exemplo 11.12 Suponha que A é uma matriz 4×4 simétrica. Denote por $|A_k|$ o menor principal líder de ordem k.

(a) Se $|A_1| > 0$, $|A_2| > 0$, $|A_3| > 0$ e $|A_4| > 0$, então A é uma matriz positiva definida.

(b) Se $|A_1| < 0$, $|A_2| > 0$, $|A_3| < 0$ e $|A_4| > 0$, então A é uma matriz negativa definida.

(c) Se $|A_1| > 0$, $|A_2| > 0$, $|A_3| = 0$ e $|A_4| < 0$, então A é uma matriz indefinida. De fato: A não é positiva definida por causa de $|A_4|$ (deveria ser > 0 e não < 0), A não é negativa definida por causa de $|A_1|$ (deveria ser < 0 e não > 0), A não é positiva semidefinida por causa de $|A_4|$ (deveria ser ≥ 0 e não < 0), A não é negativa semidefinida por causa de $|A_1|$ (deveria ser ≤ 0 e não > 0). Se A não é definida e nem semidefinida, concluímos que A é indefinida.

(d) Se $|A_1| < 0$, $|A_2| < 0$, $|A_3| < 0$ e $|A_4| < 0$, então A é uma matriz indefinida. De fato, A não é positiva definida por causa de $|A_4|$ (deveria ser > 0 e não < 0), A não é negativa definida por causa de $|A_4|$ (deveria ser > 0 e não < 0), A não é positiva semidefinida por causa de $|A_4|$ (deveria ser ≥ 0 e não < 0), A não é negativa semidefinida por causa de $|A_4|$ (deveria ser ≥ 0 e não < 0). Se A não é definida e nem semidefinida, concluímos que A é indefinida.

(e) Se $|A_1| = 0$, $|A_2| < 0$, $|A_3| > 0$ e $|A_4| = 0$, então A é uma matriz indefinida. De fato, A não é positiva definida por causa de $|A_2|$ (deveria ser > 0 e não < 0), A não é negativa definida por causa de $|A_2|$ (deveria ser > 0 e não < 0), A não é positiva semidefinida por causa de $|A_2|$ (deveria ser ≥ 0 e não < 0), A não é negativa semidefinida por causa de $|A_2|$ (deveria ser ≥ 0 e não < 0). Se A não é definida e nem semidefinida, concluímos que A é indefinida.

(f) Se $|A_1| > 0$, $|A_2| = 0$, $|A_3| > 0$ e $|A_4| > 0$, então A não é uma matriz definida. Certamente A não é negativa semidefinida (pois $|A_3| > 0$). Assim, A pode ser positiva semidefinida ou indefinida. Para decidir se A é positiva semidefinida é preciso calcular todos os menores principais de A e não apenas os quatro menores principais líderes. Se nenhum menor principal é negativo, então A é positiva semidefinida. Se pelo menos um dos menores principais é negativo, então A é indefinida.

(g) Se $|A_1| = 0$, $|A_2| > 0$, $|A_3| = 0$ e $|A_4| > 0$, então A não é uma matriz definida. Assim, A pode ser positiva semidefinida, negativa semidefinida ou indefinida. Para decidir sobre a classificação de A devemos calcular todos os menores principais da matriz A.

□

Exemplo 11.13 No exemplo (11.1) vimos que os únicos pontos críticos da função $z = f(x, y) = x^3 - y^3 + 9xy$ no conjunto admissível $D = \mathbb{R}^2$ são $\mathrm{p}_1 = (0, 0)$ e $\mathrm{p}_2 = (3, -3)$. Como todos os pontos do conjunto admissível são pontos interiores, segue-se que p_1 e p_2 são os únicos candidatos a extremo local de f. No exemplo (11.8) usamos as condições de segunda ordem para classificar estes pontos críticos estudando a positividade das matrizes hessianas

$$A = D^2 f(0,0) = \begin{bmatrix} 0 & 9 \\ 9 & 0 \end{bmatrix} \quad \text{e} \quad B = D^2 f(3,-3) = \begin{bmatrix} 18 & 9 \\ 9 & 18 \end{bmatrix}.$$

Para isto, calculamos explicitamente as formas quadráticas associadas a estas matrizes e empregamos o método de Lagrange (completamento de quadrados) para classificá-las como positiva definida, negativa definida, positiva semidefinida, negativa semidefinida ou indefinida. Vamos usar agora o teorema (11.8) como um método alternativo para classificar as matrizes A e B.

(a) Classificação do ponto crítico $\mathrm{p}_1 = (0, 0)$.

Como $|A_2| = -81 < 0$ segue-se que A não é positiva semidefinida (pois $|A_2|$ deveria ser ≥ 0 e não < 0) e nem negativa semidefinida (pois $|A_2|$ deveria ser ≥ 0 e não < 0). Com isto, A também não é positiva definida e nem negativa definida. Logo, A é indefinida e, conseqüentemente, $(0, 0)$ é um ponto de sela de f.

(b) Classificação do ponto crítico $\mathrm{p}_2 = (3, -3)$.

Como $|B_1| = 18 > 0$ e $|B_2| = 18^2 - 9^2 > 0$ segue-se que todos os menores principais líderes de A são > 0. Logo, A é positiva definida e, conseqüentemente, $(3, -3)$ é um ponto de mínimo local de f. □

A demonstração do teorema (11.8) não é simples e, com a teoria que temos até aqui, ele não parece ser nada óbvio. Na referência [71] você pode encontrar uma demonstração por indução do teorema em sua forma mais geral. Por outro lado, existem situações mais simples onde o teorema (11.8) pode ser facilmente estabelecido. Vamos estudar dois destes casos: a positividade de matrizes diagonais e a positividade de matrizes simétricas 2×2.

Exemplo 11.14 (POSITIVIDADE DE MATRIZES DIAGONAIS) No exercício resolvido (11.3) estudamos a positividade de uma matriz diagonal 4×4. Não é difícil de se ver que as conclusões estabelecidas neste exercício podem facilmente ser generalizadas para matrizes diagonais $n \times n$, isto é, se

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & d_n \end{bmatrix}$$

é uma matriz diagonal $n \times n$, então o estudo do sinal da forma quadrática associada a D,

$$\begin{aligned} Q(h_1, h_2, \ldots, h_n) &= \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix} \cdot \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} \\ &= d_1 h_1^2 + d_2 h_2^2 + \cdots + d_n h_n^2, \end{aligned}$$

nos permite concluir que

(a) D é uma matriz positiva definida se, e somente se, todos os elementos da diagonal principal de D são positivos, isto é, $d_1 > 0$, $d_2 > 0$, $\ldots$, $d_n > 0$,

(b) D é uma matriz negativa definida se, e somente se, todos os elementos da diagonal principal de D são negativos, isto é, $d_1 < 0$, $d_2 < 0$, $\ldots$, $d_n < 0$,

(c) D é uma matriz positiva semidefinida se, e somente se, todos os elementos da diagonal principal de D são maiores ou iguais a zero, isto é, $d_1 \geq 0$, $d_2 \geq 0$, $\ldots$, $d_n \geq 0$,

(d) D é uma matriz negativa semidefinida se, e somente se, todos os elementos da diagonal principal de D são menores ou iguais a zero, isto é, $d_1 \leq 0$, $d_2 \leq 0$, $\ldots$, $d_n \leq 0$ e

(e) D é uma matriz indefinida se, e somente se, existem dois elementos da diagonal principal de D com sinais contrários, isto é, existem índices i e j tais que $d_i > 0$ e $d_j < 0$.

Lembrando que o determinante de uma matriz diagonal nada mais é do que o produto dos elementos da diagonal, é fácil de ver que os menores principais líderes de D são:

$$|D_1| = d_1, \qquad |D_2| = d_1 \cdot d_2, \qquad \ldots, \qquad |D_n| = d_1 \cdot d_2 \cdot \cdots \cdot d_n.$$

Com isto, não é difícil de se ver que

(a) D é uma matriz positiva definida se, e somente se, todos os menores principais líderes de D são positivos, isto é, $|D_1| > 0$, $|D_2| > 0$, $|D_3| > 0$, $|D_4| > 0$, $\ldots$, $|D_n| > 0$ e

(b) D é uma matriz negativa definida se, e somente se, todos os menores principais líderes de ordem ímpar de D são negativos e todos os menores principais líderes de ordem par de D são positivos, isto é $|D_1| < 0$, $|D_2| > 0$, $|D_3| < 0$, $|D_4| > 0$, etc.

Por outro lado, o fato de todos os menores principais líderes de D serem maiores ou iguais a zero não garante que D seja uma matriz positiva semidefinida. Por exemplo, se $d_1 = 0$ e todos os demais elementos da diagonal de D são negativos, então os menores principais líderes de D são todos iguais a zero mas D não é uma matriz positiva semidefinida. Moral da história: para decidir se D é semidefinida ou indefinida não basta estudar apenas os sinais do menores principais líderes de D! É preciso estudar os sinais de *todos* os menores principais! Como as submatrizes principais de D também são matrizes diagonais, segue-se que

(c) D é uma matriz positiva semidefinida se, e somente se, todos os menores principais de D são maiores ou iguais a zero e

(d) D é uma matriz negativa definida se, e somente se, todos os menores principais de ordem ímpar de D são menores ou iguais a zero e todos os menores principais de ordem par de D são maiores ou iguais a zero.

(e) Caso os sinais dos menores principais da matriz D não se enquadrem nos itens (a), (b), (c) e (d), concluímos que D é uma matriz indefinida. □

Exemplo 11.15 (Positividade de matrizes simétricas 2×2) Considere uma matriz simétrica 2×2

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

e a forma quadrática associada

$$Q(h_1, h_2) = \begin{bmatrix} h_1 & h_2 \end{bmatrix} \cdot \begin{bmatrix} a & b \\ b & c \end{bmatrix} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = a\, h_1^2 + 2b\, h_1 h_2 + c\, h_2^2.$$

(a) A é positiva definida se, e somente se, $|A_1| = a > 0$ e $|A_2| = ac - b^2 > 0$.

($\Rightarrow$) Se A é uma matriz positiva definida, então $Q(1,0) = a > 0$ e $Q(-b/a, 1) = -b^2/a + c = (ac - b^2)/a > 0$. Portanto, $|A_1| = a > 0$ e $|A_2| = ac - b^2 > 0$.

($\Leftarrow$) Reciprocamente, se $|A_1| = a > 0$ e $|A_2| = ac - b^2 > 0$, completando-se o quadrado, temos que

$$\begin{aligned} Q(h_1, h_2) &= a\, h_1^2 + 2b\, h_1 h_2 + \frac{b^2}{a}\, h_2^2 - \frac{b^2}{a}\, h_2^2 + c\, h_2^2 \\ &= a \left(h_1^2 + \frac{2b}{a}\, h_1 h_2 + \frac{b^2}{a^2}\, h_2^2 \right) - \frac{b^2}{a}\, h_2^2 + c\, h_2^2 \\ &= a \left(h_1 + \frac{b}{a}\, h_2 \right)^2 + \frac{ac - b^2}{a}\, h_2^2 \end{aligned}$$

é maior ou igual a zero para todo $(h_1, h_2) \in \mathbb{R}^2$. Mas se $Q(h_1, h_2) = 0$, então $h_1 = 0$ e $h_2 = 0$. Portanto, $Q(h_1, h_2) > 0$ para todo $(h_1, h_2) \neq (0,0)$, isto é, A é uma matriz positiva definida.

(b) A é negativa definida se, e somente se, $|A_1| = a < 0$ e $|A_2| = ac - b^2 > 0$.

($\Rightarrow$) Se A é uma matriz negativa definida, então $Q(1,0) = a < 0$ e $Q(-b/a, 1) = -b^2/a + c = (ac - b^2)/a < 0$. Portanto, $|A_1| = a < 0$ e $|A_2| = ac - b^2 > 0$.

($\Leftarrow$) Reciprocamente, se $|A_1| = a < 0$ e $|A_2| = ac - b^2 > 0$, completando-se o quadrado, temos que

$$\begin{aligned} Q(h_1,h_2) &= a\,h_1^2 + 2b\,h_1h_2 + \frac{b^2}{a}\,h_2^2 - \frac{b^2}{a}\,h_2^2 + c\,h_2^2 \\ &= a\left(h_1^2 + \frac{2b}{a}\,h_1h_2 + \frac{b^2}{a^2}\,h_2^2\right) - \frac{b^2}{a}\,h_2^2 + c\,h_2^2 \\ &= a\left(h_1 + \frac{b}{a}\,h_2\right)^2 + \frac{ac-b^2}{a}\,h_2^2 \end{aligned}$$

é menor ou igual a zero para todo $(h_1,h_2) \in \mathbb{R}^2$. Mas se $Q(h_1,h_2) = 0$, então $h_1 = 0$ e $h_2 = 0$. Portanto, $Q(h_1,h_2) < 0$ para todo $(h_1,h_2) \neq (0,0)$, isto é, A é uma matriz negativa definida.

(c) A é positiva semidefinida se, e somente se, $a \geq 0$, $c \geq 0$ e $ac - b^2 \geq 0$.

($\Rightarrow$) Se A é uma matriz positiva semidefinida, então $Q(1,0) = a \geq 0$, $Q(0,1) = c \geq 0$. Temos agora que considerar vários subcasos.

1. Se $a > 0$, então $Q(-b/a,1) = -b^2/a + c = (ac - b^2)/a \geq 0$ e, portanto, $ac - b^2 \geq 0$.
2. Se $c > 0$, então $Q(1,-b/c) = a - b^2/c = (ac-b^2)/c \geq 0$ e, portanto, $ac - b^2 \geq 0$.
3. Se $a = c = 0$, então $Q(1,1) = 2\,b \geq 0$ e $Q(1,-1) = -2\,b \geq 0$, isto é, $b \geq 0$ e $b \leq 0$. Logo, $b = 0$ e, conseqüentemente, $ac - b^2 = 0$.

Em todos os subcasos, $a \geq 0$, $c \geq 0$ e $ac - b^2 \geq 0$, isto é, todos os menores principais de A são maiores ou iguais a zero.

($\Leftarrow$) Reciprocamente, suponha que $a \geq 0$, $c \geq 0$ e $ac - b^2 \geq 0$. Novamente, temos que considerar vários subcasos.

1. Se $a > 0$ então, completando-se o quadrado, temos que
$$Q(h_1,h_2) = a\left(h_1 + \frac{b}{a}\,h_2\right)^2 + \frac{ac-b^2}{a}\,h_2^2$$
é maior ou igual a zero para todo $(h_1,h_2) \in \mathbb{R}^2$.
2. Se $c > 0$ então, completando-se o quadrado, temos que
$$Q(h_1,h_2) = \frac{ac-b^2}{c}\,h_1^2 + c\left(\frac{b}{c}\,h_1 + h_2\right)^2$$

é maior ou igual a zero para todo $(h_1, h_2) \in \mathbb{R}^2$.

3. Se $a = c = 0$, então $ac - b^2 = -b^2 \geq 0$. Portanto, $b = 0$ e, conseqüentemente, $Q(h_1, h_2) = 0$ para todo $(h_1, h_2) \in \mathbb{R}^2$.

Em todos os subcasos, $Q(h_1, h_2) \geq 0$, isto é, A é uma matriz positiva semidefinida.

(d) A é negativa semidefinida se, e somente se, $a \leq 0$, $c \leq 0$ e $ac - b^2 \geq 0$.

($\Rightarrow$) Se A é uma matriz negativa semidefinida, então $Q(1,0) = a \leq 0$, $Q(0,1) = c \leq 0$. Vamos considerar os vários subcasos.

1. Se $a < 0$, então $Q(-b/a, 1) = -b^2/a + c = (ac - b^2)/a \leq 0$ e, portanto, $ac - b^2 \geq 0$.
2. Se $c < 0$, então $Q(1, -b/c) = a - b^2/c = (ac - b^2)/c \leq 0$ e, portanto, $ac - b^2 \geq 0$.
3. Se $a = c = 0$, então $Q(1,1) = 2b \leq 0$ e $Q(1,-1) = -2b \leq 0$, isto é, $b \leq 0$ e $b \geq 0$. Logo, $b = 0$ e, conseqüentemente, $ac - b^2 = 0$.

Em todos os subcasos, $a \leq 0$, $c \leq 0$ e $ac - b^2 \geq 0$, isto é, todos os menores principais de ordem ímpar de A são menores ou iguais a zero e o menor principal de ordem par de A é maior ou igual a zero.

($\Leftarrow$) Reciprocamente, suponha que $a \leq 0$, $c \leq 0$ e $ac - b^2 \geq 0$. Vamos considerar os vários subcasos.

1. Se $a < 0$ então, completando-se o quadrado, temos que
$$Q(h_1, h_2) = a\left(h_1 + \frac{b}{a}\,h_2\right)^2 + \frac{ac - b^2}{a}\,h_2^2$$
é menor ou igual a zero para todo (h_1, h_2).
2. Se $c < 0$ então, completando-se o quadrado, temos que
$$Q(h_1, h_2) = \frac{ac - b^2}{c}\,h_1^2 + c\left(\frac{b}{c}\,h_1 + h_2\right)^2$$
é menor ou igual a zero para todo (h_1, h_2).
3. Se $a = c = 0$, então $ac - b^2 = -b^2 \geq 0$. Portanto, $b = 0$ e, conseqüentemente, $Q(h_1, h_2) = 0$ para todo (h_1, h_2).

Em todos os subcasos, $Q(h_1, h_2) \leq 0$, isto é, A é uma matriz negativa semidefinida.

(e) A é indefinida se, e somente se, $ac - b^2 < 0$.

($\Rightarrow$) Suponha, por absurdo, que A seja uma matriz indefinida e que $ac - b^2 \geq 0$. Então, como A não é positiva semidefinida e nem negativa semidefinida, pelos itens (b) e (c), concluímos que a e c devem possuir sinais contrários, isto é $ac < 0$. Mas então $ac - b^2 < 0$, uma contradição.

($\Leftarrow$) Reciprocamente, suponha que $ac - b^2 < 0$. Então, pelos itens (a), (b), (c) e (d), segue-se que A não é positiva definida, não é negativa definida, não é positiva semidefinida e nem negativa semidefinida. Logo, A é uma matriz indefinida.

□

11.5 Questões de globalidade e convexidade

As condições de segunda ordem (teorema (11.7)) estabelecem um critério para decidir se um ponto crítico é um extremo *local*. Elas nada afirmam sobre a globalidade do ponto crítico! No exemplo (11.8) vimos que $(3, -3)$ é ponto crítico da função $z = f(x, y) = x^3 - y^3 + 9\,xy$ que é um ponto de mínimo local de f, mas não é um ponto de mínimo global de f em $\mathbb{R}^2$.

Como garantir a globalidade de um ponto crítico? Se o conjunto admissível é aberto (o caso que estamos estudando neste capítulo) o teorema de Weierstrass não pode ser utilizado. Nesta seção vamos estudar uma classe específica de funções para as quais é possível garantir a globalidade de pontos críticos: funções convexas e funções côncavas. Mas antes, vamos apresentar um teorema de globalidade de Cálculo I que não pode ser estendido para Cálculo II.

Exercício resolvido 11.4 Diga se cada uma das sentenças a seguir é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Suponha que $f\colon \mathbb{R} \to \mathbb{R}$ seja uma função de classe C^1 que possui um único ponto crítico p que é um ponto de mínimo local de f. Então p é um ponto de mínimo global de f em $\mathbb{R}$.

(b) Suponha que $f\colon \mathbb{R}^2 \to \mathbb{R}$ seja uma função de classe C^1 que possui um único ponto crítico $\mathbf{p}$ que é um ponto de mínimo local de f. Então $\mathbf{p}$ é um ponto de mínimo global de f em $\mathbb{R}^2$.

WWW

SOLUÇÃO:

(a) A sentença é verdadeira! Se p é o único ponto crítico de f, então a derivada f' possui um mesmo sinal em todos os pontos x maiores do que p e um mesmo sinal em todos os pontos x menores do que p. Como, por hipótese, p é um ponto de mínimo local de f, concluímos que f' é positiva em todos os pontos x maiores do que p e negativa em todos os pontos x menores do que p. Desta maneira, f é estritamente crescente no intervalo $[p, +\infty)$ e estritamente decrescente no intervalo $(-\infty, p]$. Portanto, $f(x) > f(p)$ para todo $x \neq p$. Conseqüentemente, p é um ponto de mínimo global de f em $\mathbb{R}$.

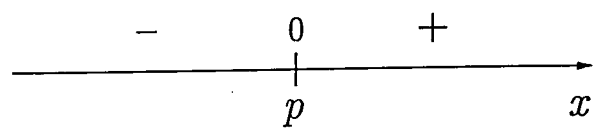

Figura 11.10: Estudo do sinal de f'.

(b) A sentença é falsa! Como um contra-exemplo, considere a função

$$z = f(x, y) = x^2 + (1 - x)^3 y^2$$

definida para todo $(x, y) \in \mathbb{R}^2$. Claramente f é uma função de classe C^1 (de fato, f é uma função C^∞). Os pontos críticos de f são calculados resolvendo-se a equação

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)\right) =$$

$$\left(2\,x - 3\,(1 - x)^2 y^2, 2\,(1 - x)^3 y\right) = (0, 0),$$

isto é, resolvendo-se o sistema *não-linear*

$$\left\{\begin{array}{rcl} 2\,x - 3\,(1 - x)^2 y^2 & = & 0, \\ 2\,(1 - x)^3 y & = & 0. \end{array}\right. \qquad (11.2)$$

Da segunda equação concluímos que $y = 0$ ou $x = 1$. Se $y = 0$, da primeira equação, concluímos que $x = 0$. Por outro lado, se $x = 1$, segue-se da primeira equação que $2 = 0$, um absurdo. Assim, $(0, 0)$ é o *único* ponto crítico de f em $\mathbb{R}^2$. Vamos aplicar as condições de segunda ordem para tentar classificar este ponto crítico e, para isto, vamos calcular a matriz hessiana $D^2 f(x, y)$ de f no ponto (x, y):

$$\begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2}(x,y) & \dfrac{\partial^2 f}{\partial x \partial y}(x,y) \\ \dfrac{\partial^2 f}{\partial y \partial x}(x,y) & \dfrac{\partial^2 f}{\partial y^2}(x,y) \end{bmatrix} = \begin{bmatrix} 2 + 6\,(1-x)y^2 & -6\,(1-x)^2 y \\ -6\,(1-x)^2 y & 2\,(1-x)^3 \end{bmatrix}.$$

Desta maneira, fazendo $(x,y) = (0,0)$, temos que a matriz hessiana $D^2 f(0,0)$ de f no ponto $(0,0)$ é dada por

$$A = D^2 f(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Como $|A_1| = 2 > 0$ e $|A_2| = 4 > 0$, segue-se que A é uma matriz positiva definida e, conseqüentemente, $(0,0)$ é um ponto de mínimo local de f. Contudo, $(0,0)$ não é um ponto de mínimo global de f, uma vez que $f(2,3) = (2)^2 + (1-2)^3(3)^2 = -5 < 0 = f(0,0)$. O gráfico de f no paralelepípedo $[-1.5, 2.5] \times [-3.0, 2.5] \times [-1.0, 2.5]$ está indicado na figura (11.11). ◻

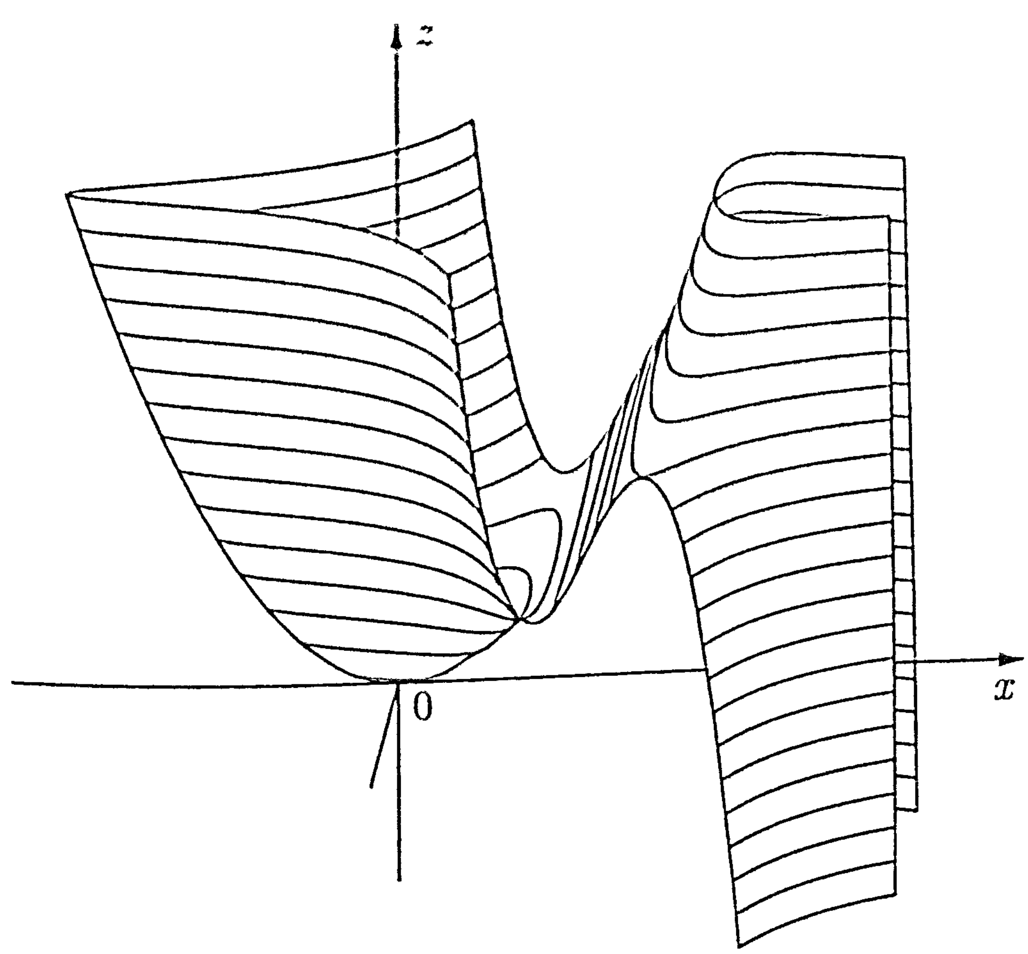

Figura 11.11: O gráfico de $z = f(x,y) = x^2 + (1-x)^3 y^2$ no paralelepípedo $[-1.5, 2.5] \times [-3.0, 2.5] \times [-1.0, 2.5]$.

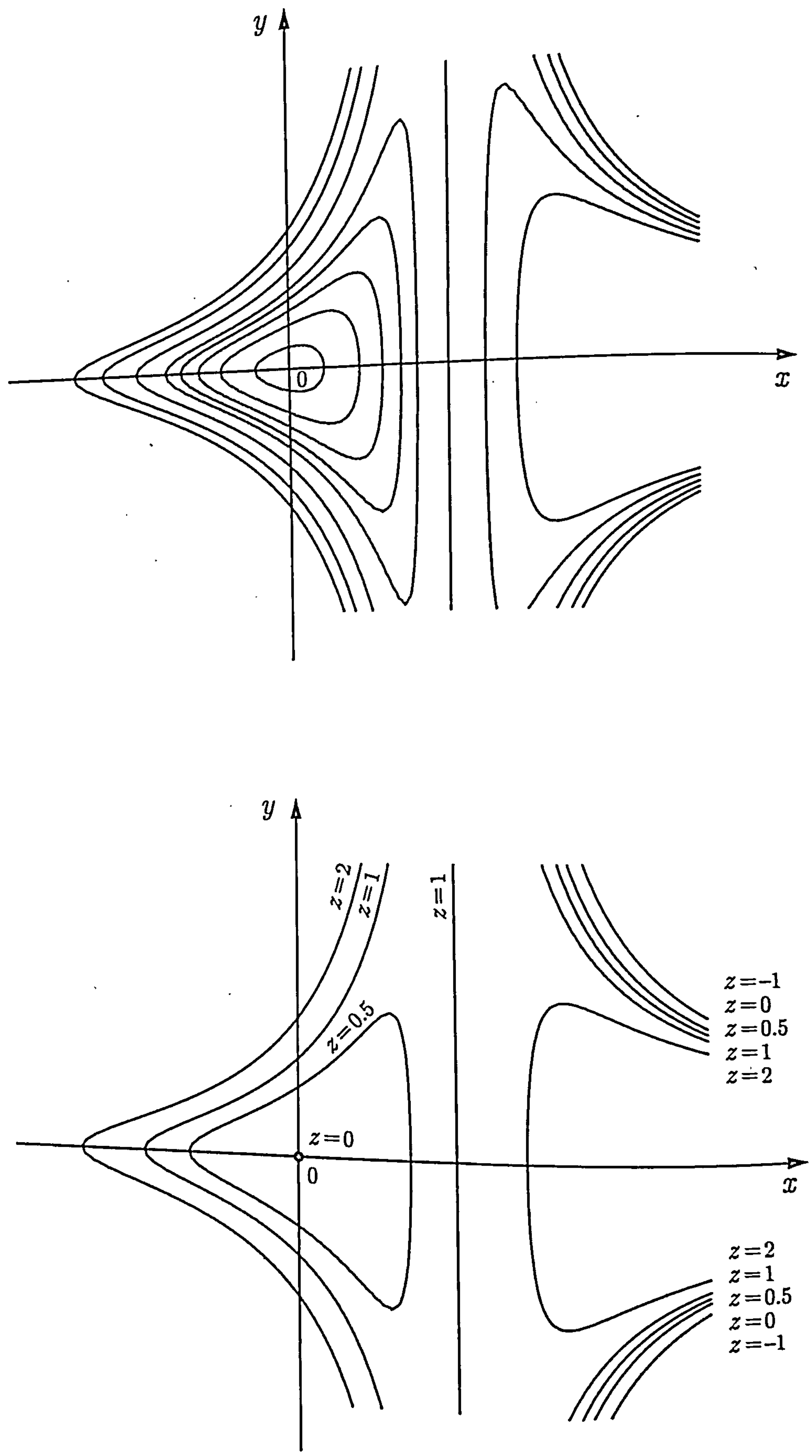

Figura 11.12: As curvas de nível de $z = f(x, y) = x^2 + (1-x)^3 y^2$ no retângulo $[-1.5, 2.5] \times [-3.0, 2.5]$ com $-1.0 \leq z \leq 2.5$.

Convexidade em Cálculo I

Em Cálculo I você aprendeu a definir e identificar funções convexas e côncavas com o uso de derivadas. Vamos dar uma definição alternativa de convexidade que não faz uso de derivadas.

Definição 11.6 (Funções convexas e côncavas)

(a) Dizemos que uma função $f\colon D \subset \mathbb{R} \to \mathbb{R}$ é *convexa* (ou *côncava para cima*) em um intervalo $I \subset D$ se, e somente se,

$$f((1-t)\cdot p + t\cdot q) \leq (1-t)\cdot f(p) + t\cdot f(q), \tag{11.3}$$

para todo $p, q \in I$ e todo $t \in [0,1]$.

(b) Dizemos que uma função $f\colon D \subset \mathbb{R} \to \mathbb{R}$ é *côncava* (ou *côncava para baixo*) em um intervalo $I \subset D$ se, e somente se,

$$f((1-t)\cdot p + t\cdot q) \geq (1-t)\cdot f(p) + t\cdot f(q), \tag{11.4}$$

para todo $p, q \in I$ e todo $t \in [0,1]$.

Vamos entender o que esta definição quer dizer geometricamente. Uma vez escolhidos os pontos p e q no intervalo I, para cada $t \in [0,1]$, a expressão

$$x(t) = (1-t)\cdot p + t\cdot q$$

representa um ponto sobre o segmento de reta que une p e q. Assim, por exemplo, em $t = 0$ temos $x(0) = p$, em $t = 1$ temos $x(1) = q$, em $t = 1/2$ temos $x(1/2) = (p+q)/2$, etc. A expressão $x(t)$ é denominada *combinação linear convexa* dos pontos p e q. Da mesma maneira, para cada $t \in [0,1]$, a expressão

$$s(t) = (1-t)\cdot f(p) + t\cdot f(q)$$

representa um ponto sobre o segmento de reta que une $f(p)$ e $f(q)$. Resumindo, para cada $t \in [0,1]$, $x(t) = (1-t)\cdot p + t\cdot q$ representa um ponto sobre o segmento de reta que une p e q,

$$\Big(x(t), f(x(t))\Big) = \Big((1-t)\cdot p + t\cdot q, f((1-t)\cdot p + t\cdot q)\Big)$$

representa o ponto no *gráfico de f* com abscissa $x(t) = (1-t)\cdot p + t\cdot q$ e

$$\Big(x(t), s(t)\Big) = \Big((1-t)\cdot p + t\cdot q, (1-t)\cdot f(p) + t\cdot f(q)\Big)$$

representa o ponto na *reta secante* que passa pelos pontos $(p, f(p))$ e $(q, f(q))$ com abscissa $x(t) = (1-t)\cdot p + t\cdot q$.

Portanto, dizer que uma função f é convexa em um intervalo I, isto é, dizer que f satisfaz a expressão (11.3) para todo $p, q \in I$ e todo $t \in [0,1]$, significa dizer que o segmento de reta secante que passa pelos pontos $(p, f(p))$ e $(q, f(q))$ sempre está *acima* ou coincide com o gráfico de f (veja a figura (11.13)).

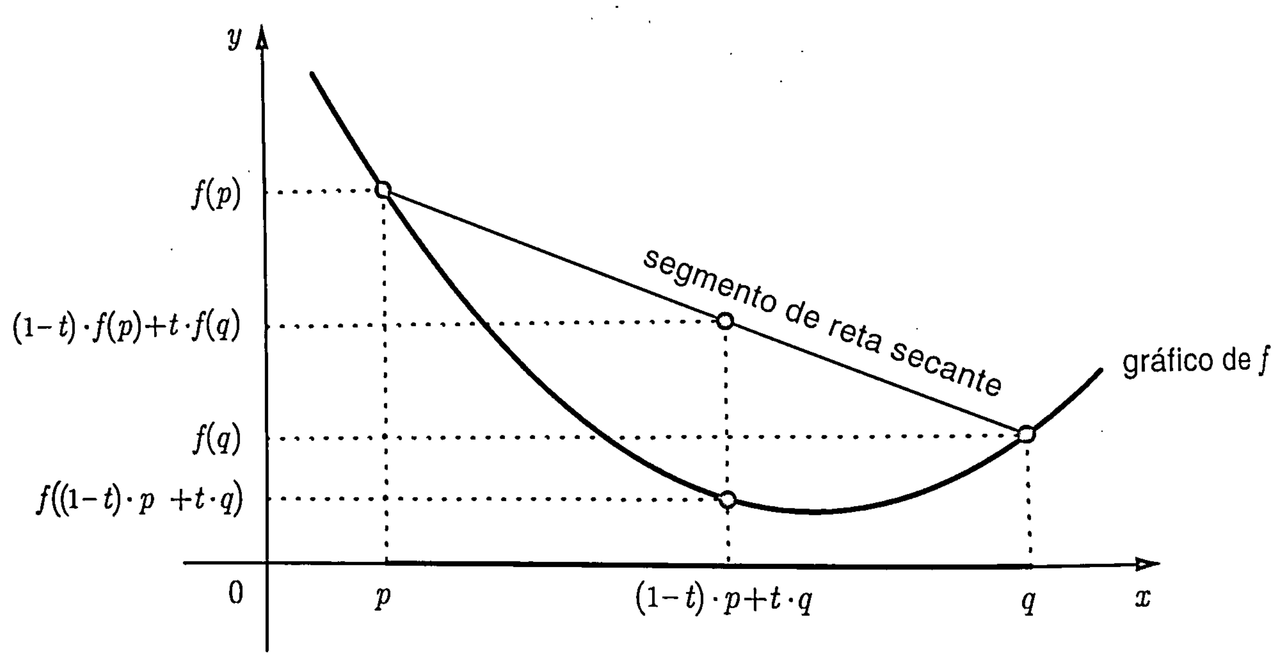

Figura 11.13: Para uma função convexa, o segmento de reta secante fica sempre acima ou coincide com o gráfico da função, para quaisquer escolhas de p e q.

Analogamente, dizer que uma função f é côncava em um intervalo I, isto é, dizer que f satisfaz a expressão (11.4) para todo $p, q \in I$ e todo $t \in [0,1]$, significa dizer que o segmento de reta secante que passa pelos pontos $(p, f(p))$ e $(q, f(q))$ sempre está *abaixo* ou coincide com o gráfico de f (veja a figura (11.14)).

Para funções de classe C^1, podemos usar a primeira derivada para caracterizar convexidade e concavidade.

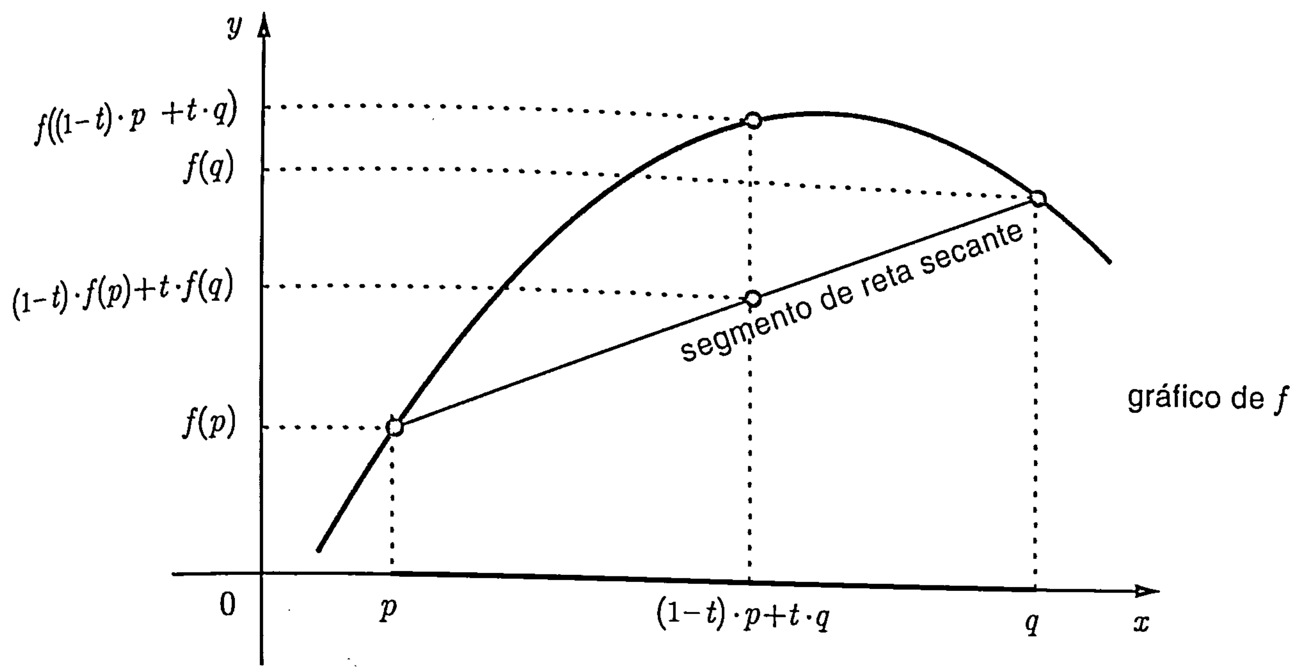

Figura 11.14: Para uma função côncava, o segmento de reta secante fica sempre abaixo ou coincide com o gráfico da função, para quaisquer escolhas de p e q.

Teorema 11.9 Seja $f\colon I \subset \mathbb{R} \to \mathbb{R}$ uma função de classe C^1 definida em um intervalo I.

(a) f é uma função convexa em I se, e somente se,

$$f(q) \geq f(p) + f'(p) \cdot (q-p), \tag{11.5}$$

para todo $p, q \in I$.

(b) f é uma função côncava em I se, e somente se,

$$f(q) \leq f(p) + f'(p) \cdot (q-p), \tag{11.6}$$

para todo $p, q \in I$.

Geometricamente, a expressão (11.5) diz que uma função de classe C^1 definida em um intervalo I é convexa se, e somente se, cada reta tangente ao gráfico de f está sempre *abaixo* ou coincide com o gráfico de f (veja a figura (11.15)).

Analogamente, a expressão (11.6) diz que uma função de classe C^1 defi-

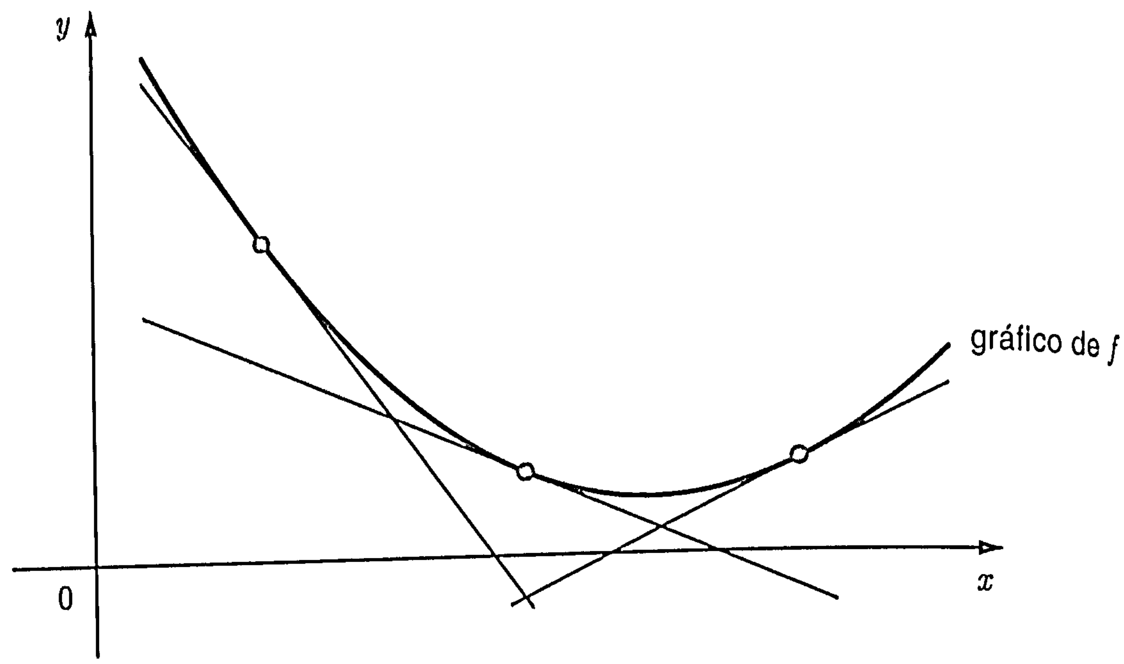

Figura 11.15: Para uma função convexa, cada reta tangente ao gráfico de f está sempre abaixo do gráfico de f.

nida em um intervalo I é côncava se, e somente se, cada reta tangente ao gráfico de f está sempre *acima* do gráfico de f (veja a figura (11.16)).

Você pode encontrar uma demonstração do teorema (11.9) na referência [71]. Como corolário, temos o seguinte teorema de globalidade de pontos críticos.

Teorema 11.10 Sejam $f\colon I \subseteq \mathbb{R} \to \mathbb{R}$ uma função de classe C^1 definida em um intervalo I e $p \in I$ um ponto crítico de f.

(a) Se f é uma função convexa em I, então p é um ponto de mínimo global de f em I.

(b) Se f é uma função côncava em I, então p é um ponto de máximo global de f em I.

Demonstração: Se f é uma função convexa em I, pelo teorema (11.9), vale que $f(q) \geq f(p) + f'(p) \cdot (q-p)$, para todo $q \in I$. Uma vez que p é um ponto crítico de f, segue-se que $f'(p) = 0$ e, portanto, $f(q) \geq f(p)$ para todo $q \in I$, isto é, p é um ponto de mínimo global de f em I. Analogamente, se f é uma função côncava em I, pelo teorema (11.9), vale que $f(q) \leq f(p) + f'(p) \cdot (q-p)$, para todo $q \in I$. Uma vez que p é um ponto crítico de f, segue-se $f'(p) = 0$

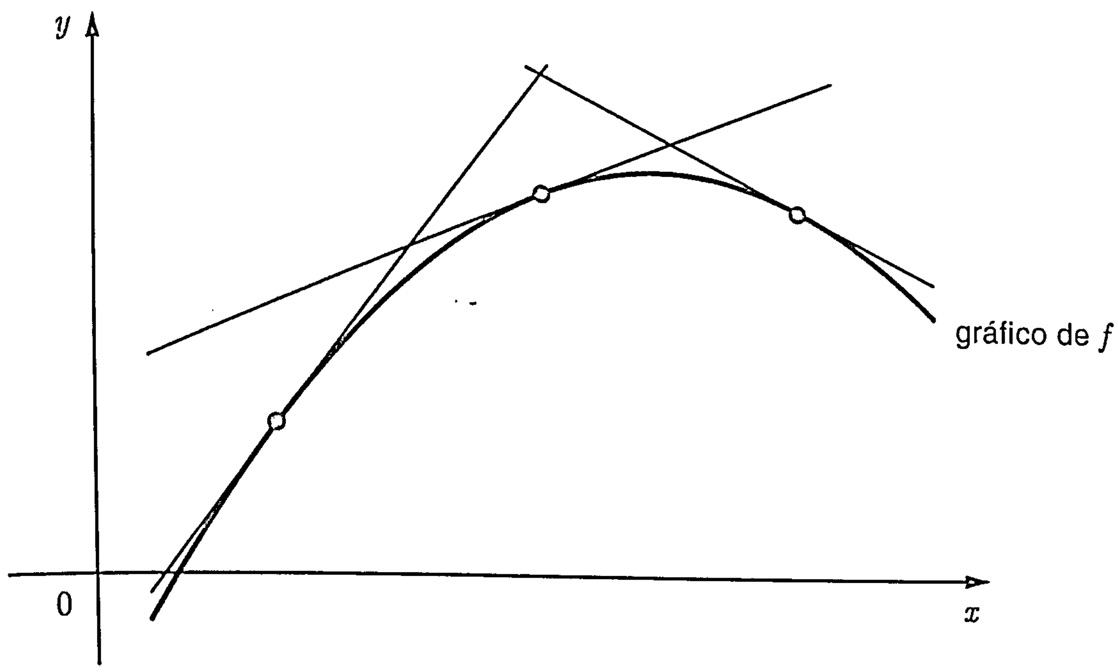

Figura 11.16: Para uma função côncava, cada reta tangente ao gráfico de f está sempre acima do gráfico de f.

e, portanto, $f(q) \leq f(p)$ para todo $q \in I$, isto é, p é um ponto de máximo global de f em I. □

Para funções de classe C^2, podemos usar a segunda derivada para caracterizar convexidade e concavidade, obtendo o famoso critério de Cálculo I.

Teorema 11.11 Seja $f : I \subset \mathbb{R} \to \mathbb{R}$ uma função de classe C^2 definida em um intervalo I.

(a) f é uma função convexa em I se, e somente se, $f''(p) \geq 0$ para todo $p \in I$.

(b) f é uma função côncava em I se, e somente se, $f''(p) \leq 0$ para todo $p \in I$.

Convexidade em Cálculo II

A fim de generalizarmos a definição (11.6) para funções de várias variáveis, é necessário que o domínio destas funções tenha a seguinte propriedade geométrica: o segmento de reta que une dois pontos quaisquer do domínio sempre está contido no domínio. Isto motiva a próxima definição.

Definição 11.7 (Conjuntos convexos) Dizemos que $U \subset \mathbb{R}^n$ é um *conjunto convexo* se, e somente se, para todo $\mathrm{p}, \mathrm{q} \in U$ tem-se

$$(1-t)\cdot \mathrm{p} + t \cdot \mathrm{q} \in U,$$

para todo $t \in [0,1]$, isto é, se o segmento de reta que une dois pontos quaisquer de U está sempre contido em U.

Exemplo 11.16 Na figura (11.17), os conjuntos dos itens (a) e (b) são convexos enquanto que os conjuntos dos itens (c) e (d) não são convexos. ∎

CUIDADO! CUIDADO! CUIDADO!

Existem funções convexas e funções côncavas mas não existem conjuntos côncavos! Se um conjunto não é convexo, ele é denominado *não-convexo*!

Com a noção de conjuntos convexos, a definição de convexidade e concavidade estende-se facilmente para funções escalares de várias variáveis.

Definição 11.8 (Funções convexas e côncavas)

(a) Dizemos que uma função $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um subconjunto convexo U de $\mathbb{R}^n$ é *convexa* se, e somente se,

$$f((1-t)\cdot \mathrm{p} + t\cdot \mathrm{q}) \leq (1-t)\cdot f(\mathrm{p}) + t\cdot f(\mathrm{q}), \tag{11.7}$$

para todo $\mathrm{p}, \mathrm{q} \in U$ e todo $t \in [0,1]$.

(b) Dizemos que uma função $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um subconjunto convexo U de $\mathbb{R}^n$ é *côncava* se, e somente se,

$$f((1-t)\cdot \mathrm{p} + t\cdot \mathrm{q}) \geq (1-t)\cdot f(\mathrm{p}) + t\cdot f(\mathrm{q}), \tag{11.8}$$

para todo $\mathrm{p}, \mathrm{q} \in U$ e todo $t \in [0,1]$.

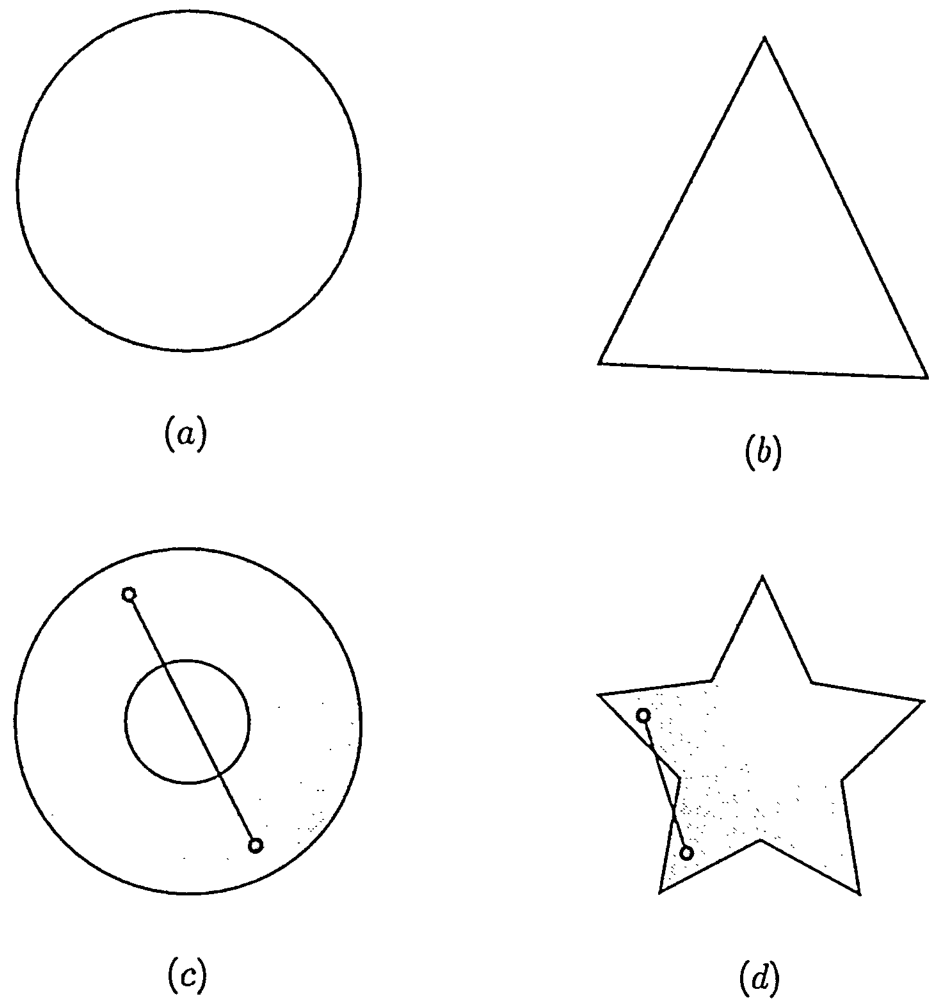

Figura 11.17: Os conjuntos dos itens (a) e (b) são convexos enquanto que os conjuntos dos itens (c) e (d) não são convexos.

A interpretação geométrica é a mesma: dizer que uma função f é convexa em um conjunto convexo U, isto é, dizer que f satisfaz a expressão (11.7) para todo $\mathrm{p}, \mathrm{q} \in U$ e todo $t \in [0,1]$, significa dizer que o segmento de reta secante que passa pelos pontos $(\mathrm{p}, f(\mathrm{p}))$ e $(\mathrm{q}, f(\mathrm{q}))$ sempre está *acima* ou *coincide* com o gráfico de f (veja a figura (11.18)).

Analogamente, dizer que uma função f é côncava em um conjunto convexo U, isto é, dizer que f satisfaz a expressão (11.8) para todo $\mathrm{p}, \mathrm{q} \in U$ e todo $t \in [0,1]$, nada mais quer dizer que o segmento de reta secante que passa pelos pontos $(\mathrm{p}, f(\mathrm{p}))$ e $(\mathrm{q}, f(\mathrm{q}))$ sempre está *abaixo* do gráfico de f.

Os teoremas (11.9), (11.10) e (11.11) generalizam-se facilmente para funções de várias variáveis.

Teorema 11.12 Seja $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^1 definida em um subconjunto convexo U de $\mathbb{R}^n$.

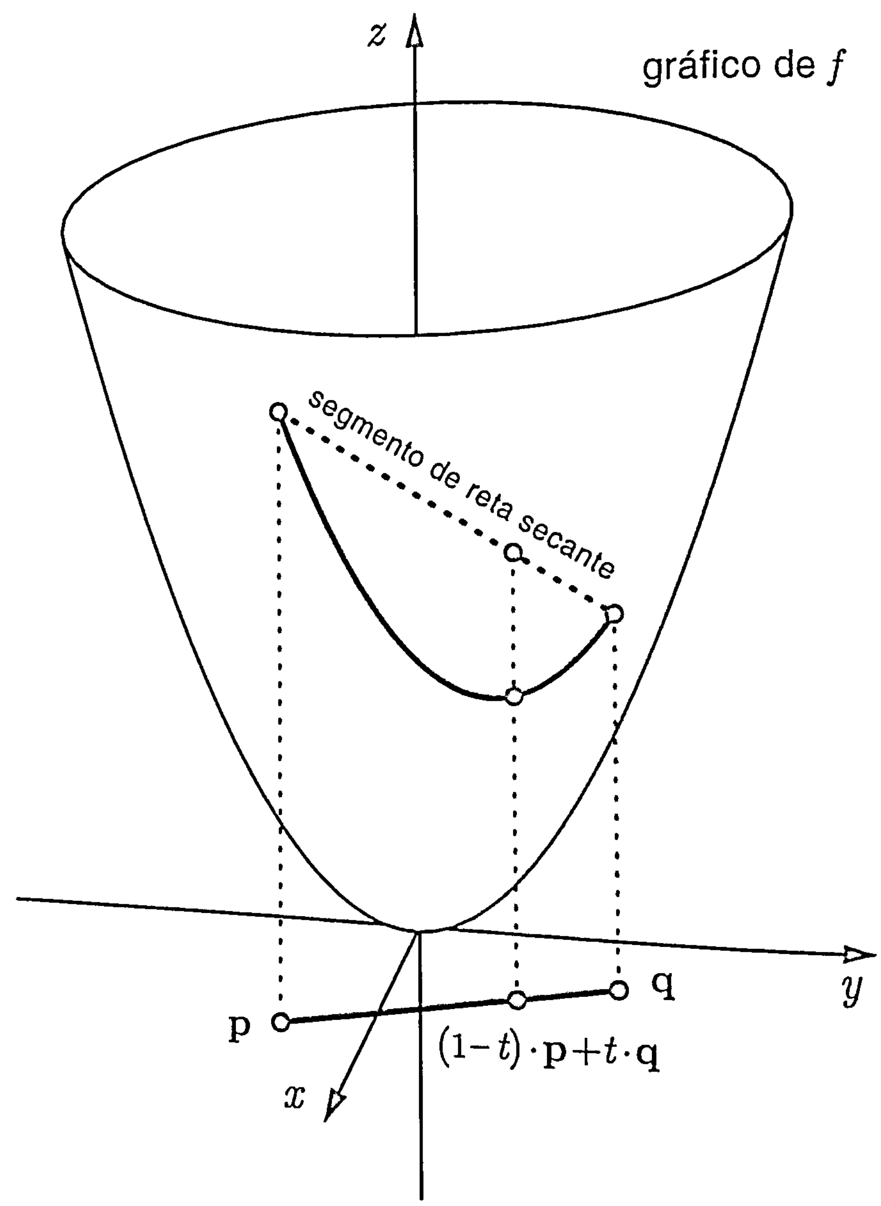

Figura 11.18: Para uma função convexa, o segmento de reta secante fica sempre acima ou coincide com o gráfico da função, para quaisquer escolhas de $\mathbf{p}$ e $\mathbf{q}$.

(a) f é uma função convexa em U se, e somente se,

$$f(\mathbf{q}) \geq f(\mathbf{p}) + Df(\mathbf{p}) \cdot (\mathbf{q} - \mathbf{p}), \tag{11.9}$$

para todo $\mathbf{p}, \mathbf{q} \in U$.

(b) f é uma função côncava em U se, e somente se,

$$f(\mathbf{q}) \leq f(\mathbf{p}) + Df(\mathbf{p}) \cdot (\mathbf{q} - \mathbf{p}), \tag{11.10}$$

para todo $\mathbf{p}, \mathbf{q} \in U$.

Como no caso unidimensional, a expressão (11.9) diz que uma função de classe C^1 definida em um conjunto convexo U é convexa se, e somente se, cada hiperplano tangente ao gráfico de f está sempre *abaixo* ou *coincide* com o gráfico de f. Analogamente, a expressão (11.10) diz que uma função de classe C^1 definida em um conjunto convexo U é côncava se, e somente se, cada hiperplano tangente ao gráfico de f está sempre *acima* ou coincide com o gráfico de f.

Como corolário, temos o seguinte teorema de globalidade de pontos críticos.

Teorema 11.13 Sejam $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^1 definida em um subconjunto convexo U de $\mathbb{R}^n$ e $\mathbf{p} \in U$ um ponto crítico de f.

(a) Se f é uma função convexa em U, então $\mathbf{p}$ é um ponto de mínimo global de f em U.

(b) Se f é uma função côncava em U, então $\mathbf{p}$ é um ponto de máximo global de f em U.

O teorema (11.12) é muito útil para se *demonstrar* propriedades de funções convexas e côncavas. Contudo, ele não é muito prático para se testar a convexidade ou concavidade de uma dada função, isto porque ele exige que verifiquemos uma desigualdade para todo $\mathbf{p}$ e $\mathbf{q}$ no convexo U. De fato, veremos que a generalização do teste da derivada segunda para funções de uma única variável (teorema (11.11)) é mais prática para este fim.

Teorema 11.14 Seja $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^2 definida em um subconjunto convexo e aberto U de $\mathbb{R}^n$.

(a) f é uma função convexa em U se, e somente se, a matriz hessiana $D^2 f(\mathbf{p})$ é positiva semidefinida *para todo* $\mathbf{p} \in U$.

(b) f é uma função côncava em U se, e somente se, a matriz hessiana $D^2 f(\mathbf{p})$ é negativa semidefinida *para todo* $\mathbf{p} \in U$.

Exemplo 11.17 Considere a função $z = f(x,y) = x^4 + x^2y^2 + y^4 - 3x - 8y$ definida em $\mathbb{R}^2$. A matriz hessiana de f é dada por

$$D^2 f(x,y) = \begin{bmatrix} 12x^2 + 2y^2 & 4xy \\ 4xy & 2x^2 + 12y^2 \end{bmatrix}.$$

Uma vez que os dois menores principais de ordem 1,

$$12x^2 + 2y^2 \qquad \text{e} \qquad 2x^2 + 12y^2,$$

e o menor principal de ordem 2,

$$24x^4 + 132x^2y^2 + 24y^4,$$

de $D^2 f(x,y)$ são maiores ou iguais a zero para todo (x,y) em $\mathbb{R}^2$, segue-se pelo teorema (11.14) que f é uma função convexa em $\mathbb{R}^2$. ▫

Exemplo 11.18 Considere a função $z = f(x,y) = xy$ definida em $\mathbb{R}^2$. A matriz hessiana de f é dada por

$$D^2 f(x,y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Uma vez que o menor principal de ordem 2 de $D^2 f(x,y)$ é igual a -1 para todo $(x,y) \in \mathbb{R}^2$, segue-se pelo teorema (11.14) que f não é uma função convexa e nem uma função côncava em $\mathbb{R}^2$. ▫

Exemplo 11.19 Considere a função $z = f(x,y) = x^4 + y^4$ definida em $\mathbb{R}^2$. A matriz hessiana de f é dada por

$$D^2 f(x,y) = \begin{bmatrix} 12\,x^2 & 0 \\ 0 & 12\,y^2 \end{bmatrix}.$$

Uma vez que os menores principais de $D^2 f(x,y)$,

$$12\,x^2, \qquad 12\,y^2 \qquad \text{e} \qquad 144\,x^2 y^2 \qquad .$$

são maiores ou iguais a zero para todo $(x,y) \in \mathbb{R}^2$, segue-se pelo teorema (11.14) que f é uma função convexa em $\mathbb{R}^2$. O único ponto crítico de f é $(0,0)$. Logo, pelo teorema (11.13), $(0,0)$ é ponto de *mínimo global* de f em $\mathbb{R}^2$.

□

CUIDADO! CUIDADO! CUIDADO!

Observe a diferença entre os teoremas (11.7) e (11.14). Se a matriz hessiana de uma função f de classe C^2 em um ponto crítico p é positiva definida, então p é um ponto de mínimo *local* de f. Mas, se a matriz hessiana de f em um ponto crítico p é positiva semidefinida, então p pode ser um extremo local ou um ponto de sela de f, isto é, *as condições de segunda ordem não conseguem classificar o ponto crítico neste caso.* Por outro lado, se a matriz hessiana de f é positiva semidefinida em *todos os pontos* do domínio (convexo) de f, então todo ponto crítico de f será um ponto de mínimo *global* de f. Por exemplo, $(0,0)$ é um ponto crítico da função $z = f(x,y) = x^4 + y^4$. A matriz hessiana de f neste ponto,

$$D^2 f(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

é positiva semidefinida. Com isto, caímos no caso "inconclusivo" do teorema (11.7). Por outro lado, a matriz hessiana de f em um ponto (x,y) qualquer de $\mathbb{R}^2$,

$$D^2 f(x,y) = \begin{bmatrix} 12\,x^2 & 0 \\ 0 & 12\,y^2 \end{bmatrix},$$

é positiva semidefinida para *todo* $(x,y) \in \mathbb{R}^2$. Portanto, pelo teorema (11.14), f é uma função convexa em $\mathbb{R}^2$. Sendo assim, pelo teorema (11.13), o ponto crítico $(0,0)$ é ponto de mínimo *global* de f em $\mathbb{R}^2$.

11.6 O método dos mínimos quadrados

Freqüentemente, quando fazemos experimentos ou observações estatísticas, obtemos dados que podemos dispor em uma tabela:

x	y
x_1	y_1
x_2	y_2
$\vdots$	$\vdots$
x_n	y_n

Gostaríamos de descobrir se os dados obtidos podem estar relacionados de alguma maneira. Em geral, a primeira coisa que tentamos verificar é se os dados satisfazem a equação de uma reta. No caso particular em que a tabela possui apenas dois pontos (com $x_1 \neq x_2$), sabemos que existe uma única reta que passa por eles. Por exemplo, a reta que passa pelos dois pontos da tabela

x	y
1	2
5	7

é dada pela equação

$$y = \frac{3}{4} + \frac{5}{4}x.$$

Contudo, se a tabela possui três ou mais pontos, seria muita sorte que todos estes pontos estivessem alinhados sob uma mesma reta. Neste caso, gostaríamos de encontrar uma reta que aproximasse estes pontos, em algum sentido. Considere a figura (11.19). Para qualquer reta $y = ax + b$, podemos medir a distância vertical de cada um destes n pontos até a reta. Estas distâncias estão representadas pelos comprimentos dos segmentos verticais pontilhados d_1, d_2 e d_3 na figura. É fácil de ver que

$$d_1 = |ax_1 + b - y_1|, \quad d_2 = |ax_2 + b - y_2|, \quad \ldots, \quad d_n = |ax_n + b - y_n|.$$

Desta maneira, poderíamos nos perguntar qual a melhor reta ou, em outras palavras, quais os melhores valores dos coeficientes a e b, a fim de minimizar

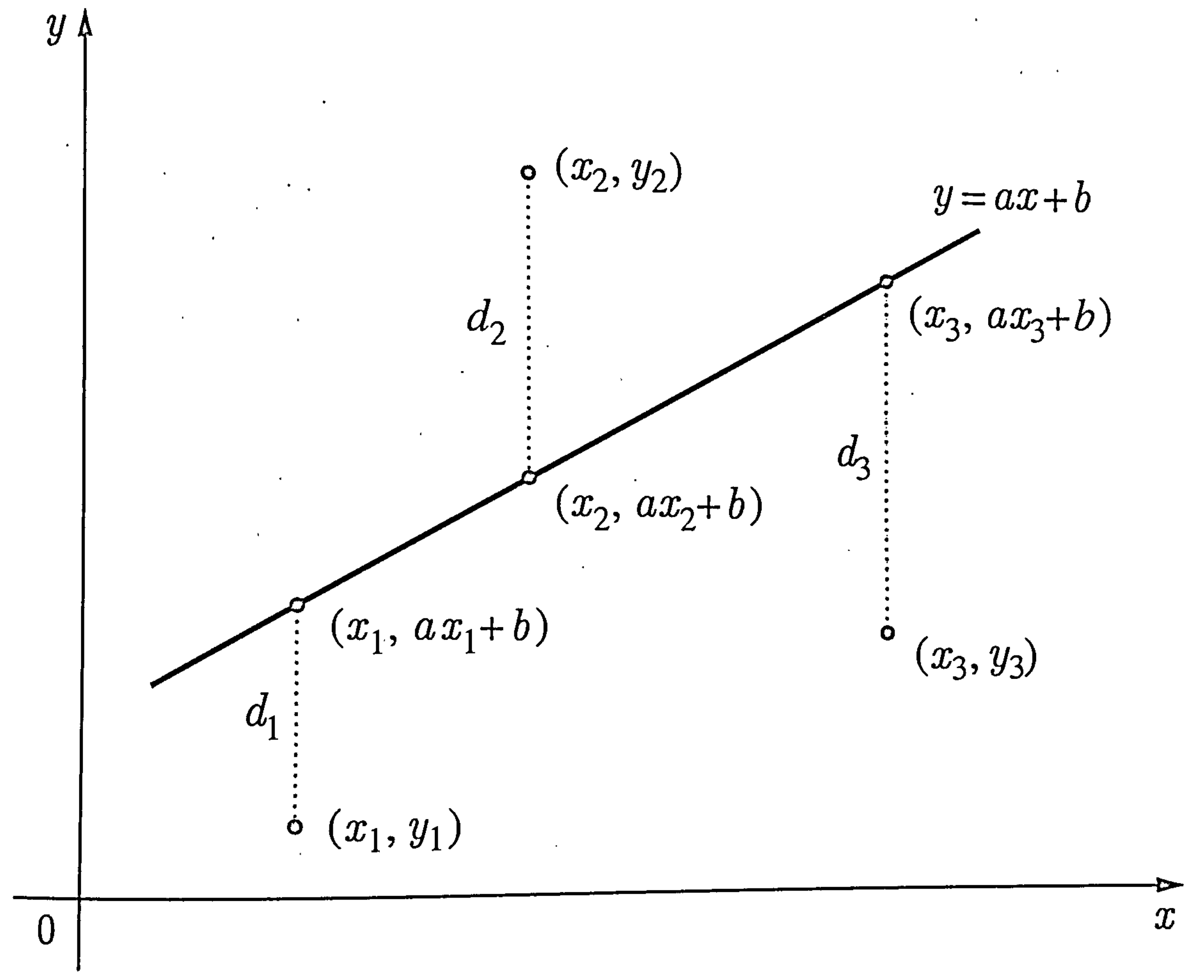

Figura 11.19: A distância vertical agregada entre os três pontos (x_1, y_1), (x_2, y_2), (x_3, y_3) e a reta $y = ax + b$.

a soma dos quadrados dos desvios:

$$\begin{aligned} s(a,b) &= d_1^2 + d_2^2 + \cdots + d_n^2 \\ &= (ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + \cdots + (ax_n + b - y_n)^2 \\ &= \sum_{i=1}^{n} (ax_i + b - y_i)^2. \end{aligned}$$

Você poderia se perguntar por que estamos querendo minimizar a *soma dos quadrados* dos desvios ao invés de, mais simplesmente, minimizar a *soma dos desvios*? Uma boa razão é que a função módulo $|\ \ |$ não é derivável na origem e estamos querendo usar derivadas para resolver este problema! Observe que

$$\begin{aligned} s\colon \quad \mathbb{R}^2 &\to \mathbb{R} \\ (a,b) &\mapsto s(a,b) = \sum_{i=1}^{n} (ax_i + b - y_i)^2 \end{aligned}$$

é uma função de duas variáveis de classe C^∞ definida para todo $(a,b) \in \mathbb{R}^2$. Como $\mathbb{R}^2$ é um conjunto aberto, se s possui um mínimo global (a^*, b^*) em $\mathbb{R}^2$, então, pela regra de Fermat, $\nabla s(a^*, b^*) = (0,0)$, isto é, (a^*, b^*) é um ponto crítico de s. Vamos então calcular os pontos críticos de s. Uma vez que

$$\frac{\partial s}{\partial a}(a,b) = \frac{\partial}{\partial a}\left(\sum_{i=1}^{n}(ax_i + b - y_i)^2\right) = \sum_{i=1}^{n} \frac{\partial}{\partial a}\left((ax_i + b - y_i)^2\right) =$$
$$\sum_{i=1}^{n} 2\,(ax_i + b - y_i)x_i = 2\left(\sum_{i=1}^{n} x_i^2\right) a + 2\left(\sum_{i=1}^{n} x_i\right) b - 2\left(\sum_{i=1}^{n} x_i y_i\right)$$

e, analogamente,

$$\frac{\partial s}{\partial b}(a,b) = \frac{\partial}{\partial b}\left(\sum_{i=1}^{n}(ax_i + b - y_i)^2\right) = \sum_{i=1}^{n} \frac{\partial}{\partial b}\left((ax_i + b - y_i)^2\right) =$$
$$\sum_{i=1}^{n} 2\,(ax_i + b - y_i)1 = 2\left(\sum_{i=1}^{n} x_i\right) a + 2\left(\sum_{i=1}^{n} 1\right) b - 2\left(\sum_{i=1}^{n} y_i\right),$$

segue-se que (a,b) é ponto crítico de s se, somente se, (a,b) é solução do seguinte sistema *linear*

$$\left\{\begin{array}{rcrcl} \left(\sum_{i=1}^{n} x_i^2\right) a & + & \left(\sum_{i=1}^{n} x_i\right) b & = & \sum_{i=1}^{n} x_i y_i, \\ \left(\sum_{i=1}^{n} x_i\right) a & + & n \quad b & = & \sum_{i=1}^{n} y_i, \end{array}\right. \tag{11.11}$$

nas variáveis a e b. Resolver este sistema não é difícil e fica como um exercício mostrar que a única solução deste sistema linear é dada por

$$a^* = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2},$$

$$b^* = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i y_i\right)}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.$$

Resta mostrar que, de fato, este ponto crítico é ponto de mínimo global de s em $\mathbb{R}^2$. Para isto, basta demonstrar que s é uma função convexa em $\mathbb{R}^2$ pois, pelo teorema (11.13), todo ponto crítico de uma função convexa de classe C^1 é ponto de mínimo global da função.

Para mostrar que s é convexa em $\mathbb{R}^2$ vamos demonstrar que a matriz hessiana

$$D^2 s(a,b) = \begin{bmatrix} 2\sum_{i=1}^{n} x_i^2 & 2\sum_{i=1}^{n} x_i \\ 2\sum_{i=1}^{n} x_i & 2n \end{bmatrix}$$

de s é positiva semidefinida para todo $(a,b) \in \mathbb{R}^2$. Os menores principais da matriz hessiana $D^2 s(a, b)$ são dados, respectivamente, por

$$2 \sum_{i=1}^{n} x_i^2, \qquad 2\,n \qquad \text{e} \qquad 4 \left(n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 \right).$$

Claramente, $2 \sum_{i=1}^{n} x_i^2 \geq 0$ e $2\,n \geq 0$ e, se provarmos que a desigualdade

$$n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 = n\,(x_1^2 + \cdots + x_n^2) - (x_1 + \cdots + x_n)^2 \geq 0$$

vale para todo $x_1, x_2, \ldots, x_n \in \mathbb{R}$, então teremos provado que todos os menores principais de $D^2 s(a, b)$ são maiores ou iguais a zero. Com isto, pelo teorema (11.14), s é uma função convexa em $\mathbb{R}^2$. Para provar a desigualdade acima, vamos usar a relação fundamental

$$2\,x_i x_j \leq x_i^2 + x_j^2, \qquad \text{com } x_i, x_j \in \mathbb{R},$$

que pode ser demonstrada da seguinte maneira:

$$(x_i - x_j)^2 \geq 0 \quad \Rightarrow \quad x_i^2 - 2\,x_i x_j + x_j^2 \geq 0 \quad \Rightarrow \quad 2\,x_i x_j \leq x_i^2 + x_j^2.$$

Agora para todo $x_1, x_2, \ldots, x_n \in \mathbb{R}$ vale que

$$\begin{aligned}
(x_1 + \cdots + x_n)^2 = x_1^2 + \cdots + x_n^2 + 2\,x_1 x_2 + 2\,x_1 x_3 + \cdots + {} & 2\,x_1 x_n \\
+ 2\,x_2 x_3 + \cdots + {} & 2\,x_2 x_n \\
& \vdots \\
& + 2\,x_{n-1} x_n.
\end{aligned}$$

Mas, pela relação fundamental, temos que

$$\begin{array}{llll}
2\,x_1 x_2 \leq x_1^2 + x_2^2, & 2\,x_1 x_3 \leq x_1^2 + x_3^2, & \ldots, & 2\,x_1 x_n \leq x_1^2 + x_n^2, \\
 & 2\,x_2 x_3 \leq x_2^2 + x_3^2, & \ldots, & 2\,x_2 x_n \leq x_2^2 + x_n^2, \\
 & & & \vdots \\
 & & & 2\,x_{n-1} x_n \leq x_{n-1}^2 + x_n^2.
\end{array}$$

para todo $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Com estas desigualdades, segue-se que

$$\begin{aligned}(x_1+\cdots+x_n)^2 \le x_1^2+\cdots+x_n^2+(x_1^2+x_2^2)+(x_1^2+x_3^2)+\cdots+ &\ (x_1^2+x_n^2)\\ +(x_2^2+x_3^2)+\cdots+ &\ (x_2^2+x_n^2)\\ &\ \vdots\\ &+(x_{n-1}^2+x_n^2),\end{aligned}$$

isto é

$$(x_1+\cdots+x_n)^2 \le x_1^2+\cdots+x_n^2+(n-1)(x_1^2+\cdots+x_n^2)=n(x_1^2+\cdots+x_n^2).$$

Exemplo 11.20 Para encontrar a reta que "melhor" aproxima os pontos

$$(0,4),\quad (3,3),\quad (4,2),\quad (3,1)\quad \text{e}\quad (5,0),$$

pelo método dos mínimos quadrados, construímos a tabela

i	x_i	y_i	x_iy_i	x_i^2
1	0	4	0	0
2	3	3	9	9
3	4	2	8	16
4	3	1	3	9
5	5	0	0	25
$\sum$	15	10	20	59

Portanto, usando as fórmulas para a^* e b^*,

$$a^*=\frac{5\cdot 20-15\cdot 10}{5\cdot 59-15\cdot 15}=-\frac{5}{7}\qquad \text{e}\qquad b^*=\frac{59\cdot 10-15\cdot 20}{5\cdot 49-15\cdot 15}=\frac{29}{7},$$

concluímos que a equação da reta que "melhor" aproxima estes pontos é dada por

$$y=-\frac{5}{7}\,x+\frac{29}{7}\qquad \text{ou}\qquad 5\,x+7\,=29.$$

A figura (11.20) mostra o gráfico da reta $y=-5\,x/7+29/7$ junto com o desenho dos 5 pontos $(0,4)$, $(3,3)$, $(4,2)$, $(3,1)$ e $(5,0)$ originais. □

O método desenvolvido nesta seção é conhecido como *método dos mínimos quadrados* ou *regressão linear*. A idéia básica pode ser generalizada para equações de curvas mais complicadas do que retas. Por exemplo, poderíamos nos perguntar qual é a melhor parábola que aproxima um conjunto de pontos do plano. Esta idéias também contribuíram para o desenvolvimento da ciência cibernética de Norbert Wiener.

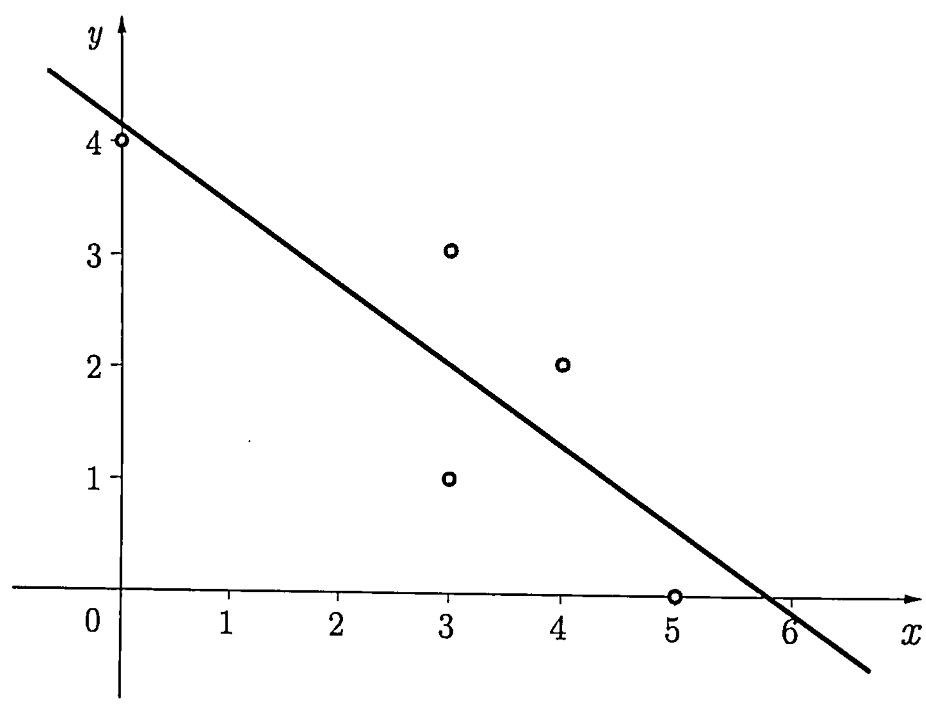

Figura 11.20: Regressão linear dos pontos $(0,4)$, $(3,3)$, $(4,2)$, $(3,1)$ e $(5,0)$.

11.7 Exercícios

[01] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^∞ e D um subconjunto não-vazio de $\mathbb{R}^n$. Se f não possui pontos críticos em D, então f não possui extremos locais em D.

(b) Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^∞ e D um subconjunto aberto e não-vazio de $\mathbb{R}^n$. Se f não possui pontos críticos em D, então f não possui extremos locais em D.

(c) Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^∞ e D um subconjunto não-vazio de $\mathbb{R}^n$. Se f não possui pontos críticos em D, então f não possui extremos globais em D.

(d) Sejam $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função de classe C^∞ e D um subconjunto aberto e não-vazio de $\mathbb{R}^n$. Se f não possui pontos críticos em D, então f não possui extremos globais em D.

[02] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) A função $f(x,y,z) = 3 \cdot y + \cos(x^2+z^2) - \operatorname{sen}(x^2-z^2)$ não possui extremos locais em $\mathbb{R}^3$.

(b) A função $f(x,y,z) = 3 \cdot y + \cos(x^2+z^2) - \operatorname{sen}(x^2-z^2)$ não possui extremos locais no conjunto $\{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$.

(c) A função $f(x,y,z) = 3 \cdot y + \cos(x^2+z^2) - \operatorname{sen}(x^2-z^2)$ não possui extremos globais em $\mathbb{R}^3$.

(d) A função $f(x,y,z) = 3 \cdot y + \cos(x^2+z^2) - \operatorname{sen}(x^2-z^2)$ não possui extremos globais no conjunto $\{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$.

[03] (a) Mostre que a função $f(x,y,z) = \cos(\ln(x^2+y^2+2xy+4)) + 5z$ não possui extremos locais em $\mathbb{R}^3$.

(b) Mostre que a função $f(x,y,z) = \cos(\ln(x^2+y^2+2xy+4)) + 5z$ não possui extremos globais em $\mathbb{R}^3$.

[04] (a) Mostre que $(1,1)$ não é ponto de extremo global da função $f(x,y) = 2\,(x^2+y^2)\,e^{-x^2-y^2}$ definida em $\mathbb{R}^2$.

(b) Mostre que $(1,1)$ não é ponto de extremo local da função $f(x,y) = 2\,(x^2+y^2)\,e^{-x^2-y^2}$ definida em $\mathbb{R}^2$.

[05] Considere o problema da dieta (página 26) onde queremos minimizar

$$\begin{aligned} C(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9) \;=\; & 0.60 \cdot x_1 + 1.00 \cdot x_2 + \\ & 5.00 \cdot x_3 + 1.00 \cdot x_4 + \\ & 0.50 \cdot x_5 + 0.20 \cdot x_6 + \\ & 0.15 \cdot x_7 + 0.40 \cdot x_8 + \\ & 1.00 \cdot x_9 \end{aligned}$$

com $(x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8)$ no conjunto admissível D formado pelos pontos de $\mathbb{R}^9$ que satisfazem as restrições

$$
\begin{cases}
0.140 \cdot x_1 + 0.580 \cdot x_2 + 0.150 \cdot x_3 + \\
\qquad\qquad 26.790 \cdot x_5 + 0.035 \cdot x_6 \;\geq\; 3, \\[2mm]
0.08 \cdot x_1 + 1.30 \cdot x_3 + 0.08 \cdot x_4 + 0.07 \cdot x_5 + \\
\qquad 0.05 \cdot x_6 + 0.04 \cdot x_7 + 0.08 \cdot x_8 + 0.07 \cdot x_9 \;\geq\; 1.1, \\[2mm]
26 \cdot x_3 + 38 \cdot x_4 + 13 \cdot x_5 + x_6 + \\
\qquad\qquad 5 \cdot x_7 + 8 \cdot x_8 + 25 \cdot x_9 \;\geq\; 60, \\[2mm]
1.60 \cdot x_1 + 6.90 \cdot x_3 + 0.20 \cdot x_4 + 0.07 \cdot x_5 + \\
\qquad\qquad 0.02 \cdot x_7 + 0.10 \cdot x_8 + 2.00 \cdot x_9 \;\geq\; 11, \\[2mm]
x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \\
x_5 \geq 0, \quad x_6 \geq 0, \quad x_7 \geq 0, \quad x_8 \geq 0 \quad \text{e} \quad x_9 \geq 0.
\end{cases}
$$

Resolva as questões a seguir, não esquecendo de justificar cuidadosamente sua resposta.

(a) O ponto $(1, 1, 1, 1, 1, 1, 1, 1, 1)$ pertence ao conjunto admissível D?

(b) O ponto $(1, 1, 1, 1, 1, 1, 1, 1, 2)$ pertence ao conjunto admissível D?

(c) O conjunto admissível D é vazio?

(d) O ponto $(1, 1, 1, 1, 1, 1, 1, 1, 2)$ é um ponto interior do conjunto admissível D?

(e) O ponto $(1, 1, 1, 1, 1, 1, 1, 1, 2)$ é um extremo global da função C no conjunto admissível D?

(f) O ponto $(1, 1, 1, 1, 1, 1, 1, 1, 2)$ é um extremo local da função C no conjunto admissível D?

(g) Mostre que C não possui extremos globais no interior do conjunto admissível D.

(h) Mostre que C não possui extremos locais no interior do conjunto admissível D.

(i) Encontre um ponto da fronteira do conjunto admissível D.

[06] Na figura (11.21) encontram-se as curvas de nível de uma função escalar f de classe C^∞ definida em $\mathbb{R}^2$. Os únicos pontos críticos de f são os pontos $\mathbf{p}$, $\mathbf{q}$ e $\mathbf{r}$. Os vetores indicam a direção e sentido do vetor gradiente.

(a) O ponto $\mathbf{p}$ é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

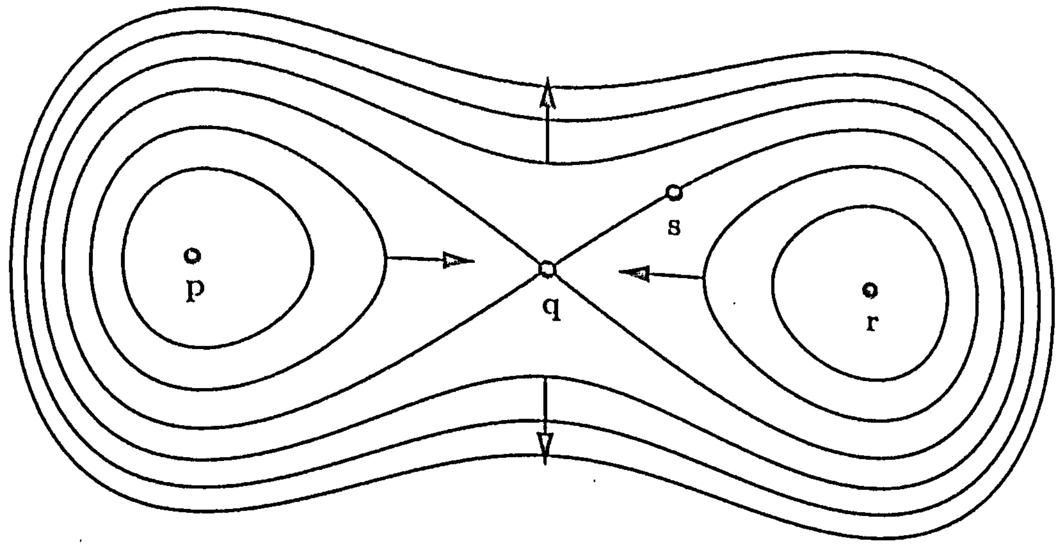

Figura 11.21: As curvas de nível de uma função $f\colon \mathbb{R}^2 \to \mathbb{R}$ de classe C^∞.

(b) O ponto **q** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(c) O ponto **r** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(d) O ponto **s** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua reposta.

[07] Na figura (11.22) encontram-se as curvas de nível de uma função escalar f de classe C^∞ definida em $\mathbb{R}^2$. Os únicos pontos críticos de f são os pontos **p**, **q** e **r**. Os vetores indicam a direção e sentido do vetor gradiente.

(a) O ponto **p** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(b) O ponto **q** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(c) O ponto **r** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(d) O ponto **s** é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua reposta.

[08] Na figura (11.23) encontram-se as curvas de nível de uma função escalar f de classe C^∞ definida em $\mathbb{R}^2$. Os únicos pontos críticos de f são os pontos **p**, **q** e **r**. Os vetores indicam a direção e sentido do vetor gradiente.

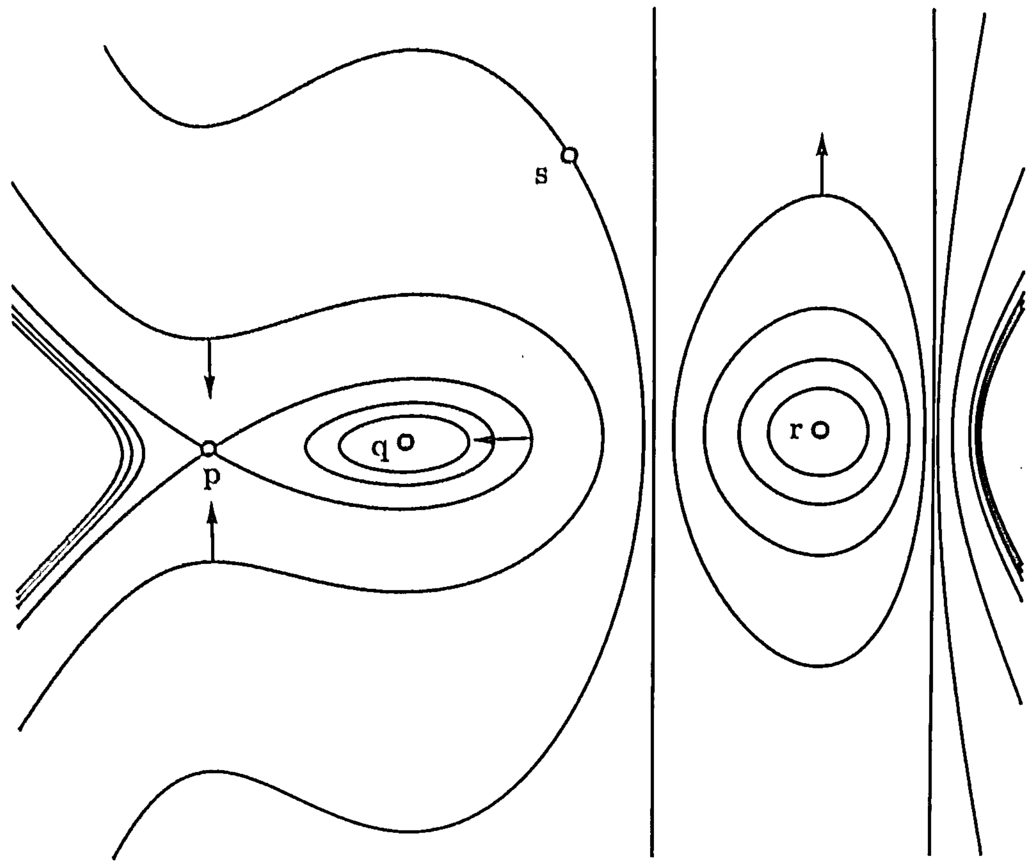

Figura 11.22: As curvas de nível de uma função $f\colon \mathbb{R}^2 \to \mathbb{R}$ de classe C^∞.

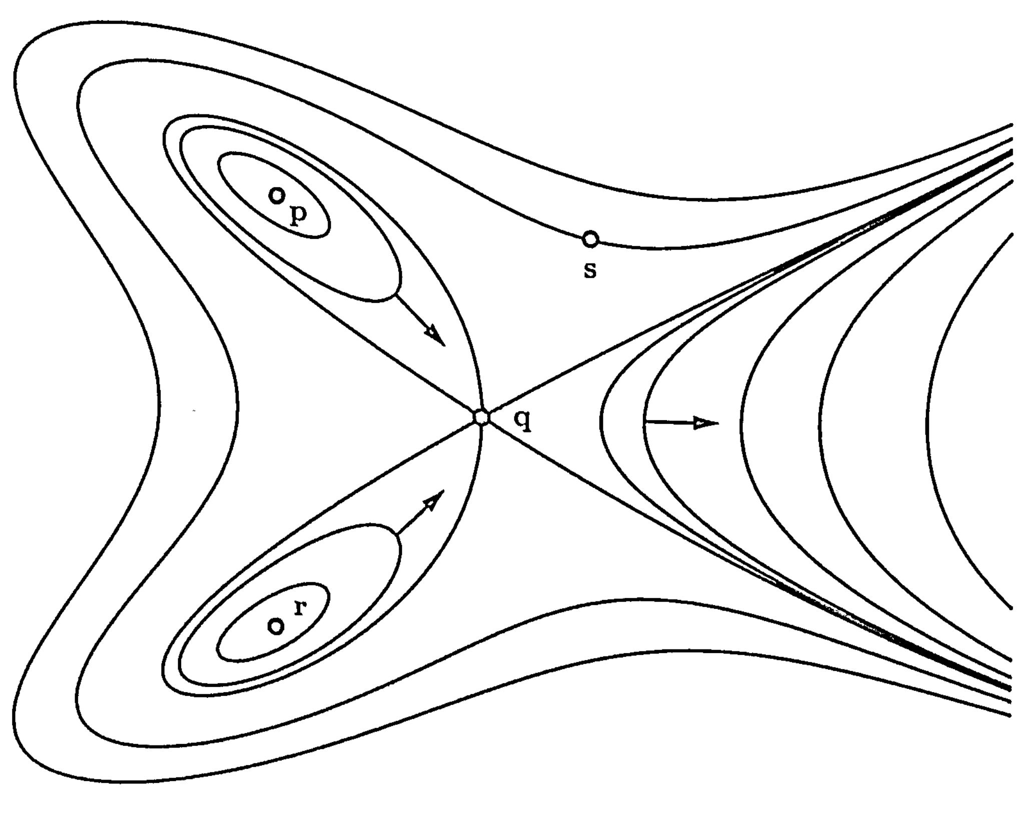

Figura 11.23: As curvas de nível de uma função $f\colon \mathbb{R}^2 \to \mathbb{R}$ de classe C^∞.

(a) O ponto $\mathbf{p}$ é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(b) O ponto $\mathbf{q}$ é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(c) O ponto $\mathbf{r}$ é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua resposta.

(d) O ponto $\mathbf{s}$ é máximo local, mínimo local ou ponto de sela para a função f? Justifique sua reposta.

[09] Aproxime $f(x) = e^x$ em $a = 0$ com um polinômio de Taylor de grau três e com um polinômio de grau quatro. A seguir, calcule os valores destas aproximações em $x = 0.2$ e $x = 1.0$ e compare com os valores corretos.

[10] Use o polinômio de Taylor de ordem dois de $f(x) = x^{3/2}$ no ponto $a = 4$ para obter uma aproximação de $(4.2)^{3/2}$.

[11] Calcule os polinômios de Taylor de ordem um, dois e três da funções $y = f(x) = \sqrt{x+1}$ em $a = 0$ e da função $y = g(x) = \ln(x)$ em $x = 1$. A seguir, calcule os valores destas aproximações em $x = 0.2$ e $x = 1.0$ e compare com os valores corretos.

[12] Calcule os polinômios de Taylor de ordem um e dois das funções abaixo nos pontos indicados.

(a) $f(x, y) = x/(1 + y)$ em $\mathbf{p} = (0, 0)$.

(b) $f(x, y) = e^x\sqrt{1 + y^2}$ em $\mathbf{p} = (0, 0)$.

(c) $f(x, y, z) = x^{1/4}y^{1/2}z^{1/4}$ em $\mathbf{p} = (1, 1, 1)$.

(d) $f(x, y) = k\,x^a y^b$ em $\mathbf{p} = (1, 1)$.

[13] Use o polinômio de Taylor de ordem dois em $\mathbf{p} = (1, 1)$ para aproximar $f(x, y) = x^{1/2}y^{1/2}$ em $\mathbf{q} = (1.2, 0.9)$.

[14] Escreva o polinômio de Taylor de ordem dois da função de duas variáveis $f(x, y) = e^x/(y + 1)$ no ponto $\mathbf{p} = (0, 1)$.

[15] Seja $p(x, y) = a_0 + a_1\,x + a_2\,y + b_1\,x^2 + b_2\,xy + b_3\,y^2$ um polinômio de ordem dois de duas variáveis. Para qualquer (x^*, y^*), mostre que o polinômio de Taylor de f de ordem 2 no ponto (x^*, y^*) é o próprio polinômio p.

[16] Considere a matriz

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos $\mathbf{p}$ e $\mathbf{q}$ em $\mathbb{R}^2$ tais que $Q(\mathbf{p}) > 0$ e $Q(\mathbf{q}) < 0$.

[17] Considere a matriz

$$A = \begin{bmatrix} -3 & 4 \\ 4 & -5 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos $\mathbf{p}$ e $\mathbf{q}$ em $\mathbb{R}^2$ tais que $Q(\mathbf{p}) > 0$ e $Q(\mathbf{q}) < 0$.

[18] Considere a matriz

$$A = \begin{bmatrix} -3 & 4 \\ 4 & -6 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos $\mathbf{p}$ e $\mathbf{q}$ em $\mathbb{R}^2$ tais que $Q(\mathbf{p}) > 0$ e $Q(\mathbf{q}) < 0$.

[19] Considere a matriz

$$A = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos $\mathbf{p}$ e $\mathbf{q}$ em $\mathbb{R}^2$ tais que $Q(\mathbf{p}) > 0$ e $Q(\mathbf{q}) < 0$.

[20] Considere a matriz

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 5 \\ 0 & 5 & 6 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2, h_3)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos $\mathbf{p}$ e $\mathbf{q}$ em $\mathbb{R}^3$ tais que $Q(\mathrm{p}) > 0$ e $Q(\mathrm{q}) < 0$.

[21] Considere a matriz

$$A = \begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2, h_3, h_4)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos p e q em $\mathbb{R}^3$ tais que $Q(\mathrm{p}) > 0$ e $Q(\mathrm{q}) < 0$.

[22] Considere a matriz

$$A = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2, h_3)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos p e q em $\mathbb{R}^3$ tais que $Q(\mathrm{p}) > 0$ e $Q(\mathrm{q}) < 0$.

[23] Considere a matriz

$$A = \begin{bmatrix} 1 & 0 & 3 & 0 \\ 0 & 2 & 0 & 5 \\ 3 & 0 & 4 & 0 \\ 0 & 5 & 0 & 6 \end{bmatrix}.$$

(a) A matriz A é simétrica? Justifique sua resposta.

(b) Escreva explicitamente a forma quadrática $Q = Q(h_1, h_2, h_3, h_4)$ associada à matriz A.

(c) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando o método de Lagrange.

(d) Calcule todos os menores principais de A.

(e) Calcule todos os menores principais líderes de A.

(f) Classifique a forma quadrática (positiva definida, negativa definida, indefinida, positiva semidefinida, negativa semidefinida) usando os menores principais de A.

(g) Encontre, caso existam, pontos $\mathbf{p}$ e $\mathbf{q}$ em $\mathbb{R}^4$ tais que $Q(\mathrm{p}) > 0$ e $Q(\mathrm{q}) < 0$.

[24] Seja A uma matriz simétrica 4×4. Denote o menor principal líder de ordem k de A por $|A_k|$. Classifique a matriz A como positiva definida, negativa definida, indefinida, positiva semidefinida e negativa semidefinida em cada um dos casos abaixo. Justifique cuidadosamente sua resposta.

(a) $|A_1| > 0$, $|A_2| > 0$, $|A_3| > 0$ e $|A_4| > 0$.
(b) $|A_1| < 0$, $|A_2| < 0$, $|A_3| < 0$ e $|A_4| < 0$.
(c) $|A_1| < 0$, $|A_2| > 0$, $|A_3| < 0$ e $|A_4| > 0$.
(d) $|A_1| > 0$, $|A_2| > 0$, $|A_3| > 0$ e $|A_4| < 0$.
(e) $|A_1| = 0$, $|A_2| < 0$, $|A_3| = 0$ e $|A_4| = 0$.

[25] Verdadeira ou falsa? Se $A_{5\times 5}$ é uma matriz simétrica com $|A_1| > 0$, $|A_2| > 0$, $|A_3| > 0$, $|A_4| = 0$ e $|A_5| = 0$, então A é positiva semidefinida. Apresente uma justificativa caso a sentença seja verdadeira ou um contra-exemplo caso ela seja falsa. Aqui $|A_k|$ denota o menor principal líder de ordem k de A.

[26] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se A é uma matriz simétrica $n \times n$ positiva definida, então $B = -A$ é uma matriz simétrica negativa definida.
(b) Se A é uma matriz simétrica $n \times n$ negativa definida, então $B = -A$ é uma matriz simétrica positiva definida.

[27] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se A é uma matriz simétrica $n \times n$ com todos os seus elementos positivos, então A é positiva definida.

(b) Se A é uma matriz simétrica $n \times n$ positiva definida, então todos os seus elementos são positivos.

(c) Se A é uma matriz simétrica $n \times n$ positiva definida, então o maior elemento da matriz aparece na diagonal principal.

(d) Se A é uma matriz simétrica $n \times n$ cujo maior elemento da matriz aparece na diagonal principal, então A é uma matriz positiva definida.

(e) Se A é uma matriz simétrica $n \times n$ positiva definida, então todos os elementos da diagonal principal de A são positivos.

(f) Se A é uma matriz simétrica $n \times n$, cujos elementos da diagonal principal são todos positivos, então A é positiva definida.

[28] Verdadeira ou falsa? A matriz

$$A = \begin{bmatrix} 1 & 1 & 1 & 2 & 9 \\ 1 & 8 & 1 & 1 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 8 & 1 \\ 9 & 1 & 1 & 1 & 1 \end{bmatrix}$$

não é positiva definida. Apresente uma justificativa caso a sentença seja verdadeira ou um contra-exemplo caso ela seja falsa.

[29] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se A é uma matriz simétrica $n \times n$ e o determinante de A é menor do que zero, então A não é uma matriz positiva definida.

(b) Se A é uma matriz simétrica $n \times n$ e o determinante de A é menor do que zero, então A não é uma matriz negativa definida.

[30] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se A é uma matriz simétrica positiva definida, então todos os menores principais de A são maiores ou iguais a zero.

(b) Se A é uma matriz simétrica negativa definida, então todos os menores principais de ordem ímpar de A são menores ou iguais a zero e todos os menores principais de ordem par de A são maiores ou iguais a zero.

*[31] Imite os cálculos feitos no exemplo (11.15) para demonstrar que uma matriz simétrica 3×3 é positiva definida se, e somente se, $|A_1| > 0$, $|A_2| > 0$ e $|A_3| > 0$ e é negativa definida se, e somente se, $|A_1| < 0$, $|A_2| > 0$ e $|A_3| < 0$. Dica: complete o quadrado duas vezes para obter

$$\begin{bmatrix} h_1 & h_2 & h_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = |A_1| r^2 + \frac{|A_2|}{|A_1|} s^2 + \frac{|A_3|}{|A_2|} t^2,$$

onde

$$r = h_1 + \frac{a_{12}}{a_{11}} h_2 + \frac{a_{13}}{a_{11}} h_3, \quad s = h_2 + \frac{a_{11}a_{23} - a_{12}a_{13}}{|A_2|} h_3 \quad \text{e} \quad t = h_3.$$

*[32] **(Forma quadrática de uma matriz quadrada qualquer)** Quase toda função usada neste curso é de classe C^∞. Desta maneira, pelo teorema de Young (página 176), a matriz hessiana desta função é *simétrica* e isto justifica porque temos dado tanta atenção a formas quadráticas definidas por matrizes *simétricas*. Mas o que dizer de formas quadráticas definidas por matrizes não-simétricas? Os exercícios seguintes têm por objetivo esclarecer estes pontos.

(a) Mostre que as formas quadráticas definidas pelas matrizes

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 1 & 5 \\ -5 & -1 \end{bmatrix}$$

são iguais, isto é, mostre que

$$Q_A(\mathrm{h}) = \mathrm{h}^T \cdot A \cdot \mathrm{h} = \mathrm{h}^T \cdot B \cdot \mathrm{h} = Q_B(\mathrm{h})$$

para todo $\mathrm{h} \in \mathbb{R}^2$. Moral da história: duas matrizes diferentes podem induzir à *mesma* forma quadrática.

(b) Por outro lado, mostre que se A e B são matrizes *simétricas* diferentes, então suas formas quadráticas são diferentes.

(c) Dizemos que uma matriz A é *anti-simétrica* se ela satisfaz a propriedade $A^T = -A$. Dê um exemplo de uma matriz anti-simétrica, um exemplo de uma matriz que não é anti-simétrica e um exemplo de uma matriz que não é anti-simétrica e nem simétrica.

(d) Calcule a forma quadrática definida pela matriz anti-simétrica que você forneceu no item anterior.

(e) Mostre que a forma quadrática definida por uma matriz anti-simétrica é a função nula, isto é, mostre que se $A^T = -A$, então $Q(\mathrm{h}) = \mathrm{h}^T \cdot A \cdot \mathrm{h} = 0$ para todo $\mathrm{h} \in \mathbb{R}^n$.

(f) Seja A uma matriz $n \times n$ qualquer. Mostre que as matrizes

$$X = \frac{A + A^T}{2} \qquad \text{e} \qquad Y = \frac{A - A^T}{2}$$

são, respectivamente, simétrica e anti-simétrica. Mostre também que

$$A = X + Y.$$

Moral da história: toda matriz quadrada pode ser escrita como a soma de uma matriz simétrica e uma matriz anti-simétrica. A matriz X é denominada de *parte simétrica* de A e a matriz Y de *parte anti-simétrica* de A.

(g) Mostre que se A é uma matriz anti-simétrica e simétrica ao mesmo tempo, então A é a matriz nula.

(h) Mostre que a decomposição de uma matriz quadrada (como a soma de uma matriz simétrica e uma matriz simétrica) é única, isto é, mostre que se

$$A = X_1 + Y_1 \qquad \text{e} \qquad A = X_2 + Y_2,$$

onde X_1 e X_2 são matrizes simétricas e Y_1 e Y_2 são matrizes anti-simétricas, então

$$X_1 = X_2 \text{ e } Y_1 = Y_2.$$

(i) Conclua que para estudar a positividade de uma forma quadrática definida, por uma matriz quadrada qualquer, basta estudar a positividade da forma quadrática definida por sua parte simétrica.

[33] No teorema de classificação de formas quadráticas via menores principais (teorema (11.8)) é fundamental que a matriz A seja *simétrica*! Para você se convencer deste fato, dê um exemplo de uma matriz quadrada A *indefinida* com todos os menores principais líderes maiores do que zero.

*[34] Encontre uma matriz A com 4 linhas e 4 colunas tal que a forma quadrática definida por A seja igual a

$$Q(h_1, h_2, h_3, h_4) = {h_1}^2 - 2\,h_1h_2 + 2\,h_1h_3 - 2\,h_1h_4 + {h_2}^2 - 2\,h_2h_3 + 2\,h_2h_4 + {h_3}^2 - 2\,h_3h_4 + {h_4}^2.$$

[35] Seja A uma matriz $n \times n$. Quantos menores principais de ordem $k \leq n$ de A existem?

*[36] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se A é uma matriz simétrica positiva definida, então toda submatriz principal de A é positiva definida.

(b) Se A é uma matriz simétrica negativa definida, então toda submatriz principal de A é negativa definida.

(c) Se A é uma matriz simétrica positiva definida, então todos os menores principais de A são maiores do que zero.

(d) Se A é uma matriz simétrica negativa definida, então todos os menores principais de ordem ímpar de A são menores do que zero e todos os menores principais de ordem par de A são maiores do que zero.

[37] Considere as matrizes

$$A = \begin{bmatrix} a & b & 0 & 0 \\ b & c & 0 & 0 \\ 0 & 0 & d & e \\ 0 & 0 & e & f \end{bmatrix}, \quad B = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \quad \text{e} \quad C = \begin{bmatrix} d & e \\ e & f \end{bmatrix}.$$

Diga se cada uma das seguintes sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se A é indefinida, então B é indefinida.

(b) Se B e C são positivas definidas, então A é positiva definida.

(c) Se B é negativa definida e C é positiva definida, então A é negativa definida.

(d) Se B é negativa definida e C é indefinida, então A *não* é negativa definida.

Dica: escreva explicitamente as formas quadráticas associadas às matrizes A, B e C.

[38] Para cada uma das funções abaixo, definidas em $\mathbb{R}^2$, encontre os pontos críticos e classifique-os como mínimo local, máximo local, ponto de sela ou "inconclusivo".

(a) $f(x,y) = xy^2 + x^3y - xy$.
(b) $f(x,y) = x^2 - 6\,xy + 2\,y^2 + 10\,x + 2\,y - 5$.
(c) $f(x,y) = x^4 + x^2 - 6\,xy + 3\,y^2$.
(d) $f(x,y) = 3\,x^4 + 3\,x^2y - y^3$.

[39] Para cada uma das funções abaixo, definidas em $\mathbb{R}^3$, encontre os pontos críticos e classifique-os como mínimo local, máximo local, ponto de sela ou "inconclusivo".

(a) $f(x,y,z) = x^2 + 6\,xy + y^2 - 3\,yz + 4\,z^2 + 6\,x + 17\,y - 2\,z$.
(b) $f(x,y,z) = (x^2 + 2\,y^2 + 3\,z^2)\,e^{(x^2+y^2+z^2)}$.

[40] Mostre que

$$z = f(x,y) = \frac{3\,x^4 - 4\,x^3 - 12\,x^2 + 18}{12\,(1 + 4\,y^2)}$$

possui exatamente um máximo local, exatamente um mínimo local e exatamente um ponto de sela.

[41] Encontre os extremos locais, caso existam, para cada uma das funções abaixo.

(a) $f(x,y) = \ln(x^2 + y^2 + 1)$.
(b) $f(x,y) = x^5y + xy^5 + xy$.
(c) $f(x,y) = (x^2 - y^2 - 1)/(x^2 + y^2 + 1)^2$.
(d) $f(x,y) = xy/(x^2 + y^2 + 1)^2$.
(e) $f(x,y) = 2\,x^3 + y^3 - 3\,x^2 - 3\,y$.
(f) $f(x,y) = 3\,xy^2 + x^3 - 3\,x$.
(g) $f(x,y) = xy\,(1 - x^2 - y^2)$.

[42] Seja $w = f(x, y, z) = x^4 + x^2 z + y^2 + \dfrac{z^2}{2} + xy$.

(a) Encontre todos os pontos críticos de f.

(b) Classifique todos os pontos críticos de f como máximo local, mínimo local ou ponto de sela.

[43] Considere a função

$$f(x, y) = -y^4 - e^{-x^2} + 2\,y^2\sqrt{e^x + e^{-x^2}}$$

definida para todo $(x, y) \in \mathbb{R}^2$.

(a) Mostre que f possui um único ponto crítico que é ponto de mínimo local de f.

(b) Mostre que f não possui mínimos globais.

Este exercício fornece um outro exemplo de uma função de classe C^∞ que possui um único ponto crítico, que é mínimo local mas não é mínimo global. Compare com a função do exercício resolvido (11.4).

[44] Considere a função

$$f(x, y) = 3\,xe^y - x^3 - e^{3\,y}$$

definida para todo $(x, y) \in \mathbb{R}^2$.

(a) Mostre que f possui um único ponto crítico que é ponto de mínimo local de f.

(b) Mostre que f não possui mínimos globais.

Este exercício fornece um outro exemplo de uma função de classe C^∞ que possui um único ponto crítico, que é mínimo local mas não é mínimo global. Compare com a função do exercício resolvido (11.4).

[45] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se $\mathbf{p}$ é um ponto crítico de uma função $f\colon \mathbb{R}^n \to \mathbb{R}$ de classe C^∞ e a matriz hessiana $D^2 f(\mathbf{p})$ no ponto $\mathbf{p}$ é positiva semidefinida, então $\mathbf{p}$ é um mínimo local de f.

(b) Se p é um ponto crítico de uma função $f: \mathbb{R}^n \to \mathbb{R}$ de classe C^∞ e a matriz hessiana $D^2 f(\mathrm{p})$ no ponto p é negativa semidefinida, então p é um máximo local de f.

[46] Suponha que $z = f(x, y)$ possua derivadas parciais de primeira e segunda ordem contínuas para todo $(x, y) \in \mathbb{R}^2$ e que f seja harmônica:

$$\frac{\partial^2 f}{\partial x^2}(x, y) + \frac{\partial^2 f}{\partial y^2}(x, y) = 0, \qquad \text{para todo } (x, y) \in \mathbb{R}^2.$$

Assuma também que $(\partial^2 f/\partial x^2)(x_0, y_0) \neq 0$. Mostre que f não pode possuir um máximo ou um mínimo local em (x_0, y_0).

[47] Mostre que para qualquer escolha de números reais x, y e z, vale que

$$-x^2 + 4\,xy - 2\,xz - 5\,y^2 + 6\,yz - 3\,z^2 \leq 0.$$

Justifique cuidadosamente sua resposta.

[48] Encontre os extremos globais, caso existam, das funções abaixo definidas em $\mathbb{R}^2$. Justifique cuidadosamente sua resposta.

(a) $f(x, y) = xy^2 + x^3 y - xy$.

(b) $f(x, y) = x^2 - 6\,xy + 2\,y^2 + 10\,x + 2\,y - 5$.

(c) $f(x, y) = x^4 + 2\,x^2 y^2 + y^2$.

(d) $f(x, y) = 3\,x^4 + 3\,x^2 y - y^3$.

[49] Encontre os extremos globais, caso existam, das funções abaixo definidas em $\mathbb{R}^2$. Justifique cuidadosamente sua resposta.

(a) $f(x, y, z) = x^2 + 6\,xy + y^2 - 3\,yz + 4\,z^2 + 6\,x + 17\,y - 2\,z$.

(b) $f(x, y, z) = (x^2 + 2\,y^2 + 3\,z^2)\, e^{(x^2+y^2+z^2)}$.

[50] Quais dos conjuntos abaixo é convexo?

(a) $\{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0 \text{ e } x^2 + y^2 \leq 1\}$.

(b) $\{(x, y) \in \mathbb{R}^2 \mid x \geq 0 \text{ e } y \geq 0\}$.

(c) $\{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0 \text{ e } x + y = 1\}$.

(d) $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$.

(e) $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$.

(f) $\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.
(g) $\{(x, y, z) \in \mathbb{R}^3 \mid x \geq 0, y \geq 0, z \geq 0 \text{ e } x + y + z \leq 1\}$.
(h) $\{(x, y, z) \in \mathbb{R}^3 \mid x \geq 0, y \geq 0 \text{ e } z \leq x^2 + y^2\}$.
(i) $\{(x, y, z) \in \mathbb{R}^3 \mid x \geq 0, y \geq 0, z \geq 0, x \leq 1, y \leq 1 \text{ e } z \leq 1\}$.

[51] Mostre, usando a definição (11.7), que o conjunto

$$\{(x, y, z, w) \in \mathbb{R}^4 \mid 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1, 0 \leq w \leq 1\}$$

é convexo.

[52] Diga se cada uma das funções abaixo é convexa, côncava ou nem convexa e nem côncava.

(a) $f(x, y) = \ln(x^2 + y^2 + 1)$, com $x, y \in \mathbb{R}$.
(b) $f(x, y) = x^5 y + xy^5 + xy$, com $x, y \in \mathbb{R}$.
(c) $f(x, y) = (x^2 - y^2 - 1)/(x^2 + y^2 + 1)^2$, com $x, y \in \mathbb{R}$.
(d) $f(x, y) = xy/(x^2 + y^2 + 1)^2$, com $x, y \in \mathbb{R}$.
(e) $f(x, y) = 2\,x^3 + y^3 - 3\,x^2 - 3\,y$, com $x, y \in \mathbb{R}$.
(f) $f(x, y) = 3\,xy^2 + x^3 - 3\,x$, com $x, y \in R$.

[53] Diga se cada uma das funções abaixo é convexa, côncava ou nem convexa e nem côncava.

(a) $f(x) = 3\,e^x + 5\,x^4 - \ln(x)$, com $x > 0$.
(b) $f(x, y) = -3\,x^2 + 2\,xy - y^2 + 3\,x - 4\,y - 1$, com $x, y \in \mathbb{R}$.
(c) $f(x, y, z) = 3\,e^x + 5\,y^4 - \ln(z)$, com $x, y \in \mathbb{R}$ e $z > 0$.
(d) $f(x, y, z) = A\,x^a y^b z^c$, com $a, b, c > 0$ e $x, y, z \geq 0$.

[54] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) A união de dois conjuntos convexos é um conjunto convexo.
(b) A interseção de dois conjuntos convexos é um conjunto convexo.

[55] Diga se cada uma das sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U é convexa, então f possui pelo menos um ponto de mínimo global em U.

(b) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U é convexa, então f possui no máximo um ponto de mínimo global em U.

(c) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U não é convexa, então f não possui pontos de mínimo global em U.

(d) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U não possui pontos de mínimo global em U, então f não é convexa em U.

(e) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U possui pelo menos um ponto de máximo global, então f não é uma função convexa em U.

(f) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U possui pelo menos um ponto de mínimo global, então f não é uma função côncava em U.

(g) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U possui pelo menos um ponto de mínimo global e pelo menos um ponto de máximo global em U, então f não é uma função convexa e nem uma função côncava em U.

(h) Se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U possui pelo menos um ponto de sela em U, isto é, se f possui um ponto crítico em U que não é máximo local e nem mínimo local de f em U, então f não é uma função convexa e nem uma função côncava em U.

[56] Seja $f\colon \mathbb{R}^2 \to \mathbb{R}$ dada por

$$z = f(x,y) = \frac{x^2}{4} + \frac{y^2}{9}.$$

(a) Mostre que f é uma função convexa.

(b) Faça um desenho do *subnível*

$$\{(x,y) \in \mathbb{R}^2 \mid f(x,y) \leq 1\}$$

de f associado ao nível $z = 1$. Mostre que este conjunto é convexo.

(c) Mostre que, de fato, qualquer *subnível*

$$\{(x,y) \in \mathbb{R}^2 \mid f(x,y) \leq c\}$$

de f é um conjunto convexo.

[57] Mostre que se $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ é uma função convexa definida em um subconjunto convexo U de $\mathbb{R}^n$, então o *subnível*

$$\{\mathbf{x} \in U \mid f(\mathbf{x}) \leq c\}$$

de f, associado ao nível $z = c$, é um conjunto convexo para cada $c \in \mathbb{R}$.

[58] Seja $f\colon \mathbb{R}^2 \to \mathbb{R}$ dada por

$$z = f(x, y) = -\frac{1}{1 + x^2 + y^2} + \frac{3}{2}.$$

(a) Mostre que f *não* é uma função convexa.

(b) Faça um desenho do *subnível*

$$\{(x, y) \in \mathbb{R}^2 \mid f(x, y) \leq 1\}$$

de f associado ao nível $z = 1$. Mostre que este conjunto é convexo.

(c) Mostre que, de fato, qualquer *subnível*

$$\{(x, y) \in \mathbb{R}^2 \mid f(x, y) \leq c\}$$

de f é um conjunto convexo.

Este exercício fornece um contra-exemplo para a recíproca da proposição estabelecida no exercício (57) acima.

Moral da história: se f é uma função convexa, então todos os seus subníveis são conjuntos convexos. Por outro lado, existem funções *não-convexas* com todos os seus subníveis convexos.

[59] (a) Mostre que um subconjunto U de $\mathbb{R}^n$ é convexo se, e somente se, para todo $\mathbf{p}, \mathbf{q} \in U$ tem-se

$$\alpha_1 \cdot \mathbf{p} + \alpha_2 \cdot \mathbf{q} \in U,$$

para todo $\alpha_1, \alpha_2 \geq 0$ com $\alpha_1 + \alpha_2 = 1$.

(b) Mostre que um subconjunto U de $\mathbb{R}^n$ é convexo se, e somente se, para todo $\mathbf{p}, \mathbf{q} \in U$ tem-se

$$\alpha_1 \cdot \mathbf{p} + \cdots + \alpha_n \cdot \mathbf{q} \in U,$$

para todo $\alpha_1, \ldots, \alpha_n \geq 0$ com $\alpha_1 + \cdots + \alpha_n = 1$.

[60] Seja $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ uma função definida em um conjunto convexo U de $\mathbb{R}^n$. Mostre que f é convexa em U se, e somente se, $-f$ é côncava em U.

[61] Seja $\mathbf{a}$ um vetor em $\mathbb{R}^n$.

(a) Mostre que $z = f(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ é uma função convexa em $\mathbb{R}^n$.

(b) Mostre que o *semi-espaço*

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a} \cdot \mathbf{x} \leq c\}$$

é um conjunto convexo, para todo $c \in \mathbb{R}$.

[62] (a) Mostre que $z = f(\mathbf{x}) = ||\mathbf{x}|| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ é uma função convexa em $\mathbb{R}^n$.

(b) Mostre que a bola fechada

$$\{\mathbf{x} \in \mathbb{R}^n \mid ||x|| \leq c\}$$

é um conjunto convexo, para todo $c \in \mathbb{R}$.

[63] (**Funções estritamente convexas e questões de unicidade**) Dizemos que uma função $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U de $\mathbb{R}^n$ é *estritamente convexa* se

$$f((1-t)\cdot \mathbf{p} + t \cdot \mathbf{q}) < (1-t)\cdot f(\mathbf{p}) + t \cdot f(\mathbf{q}),$$

para todo $\mathbf{p}, \mathbf{q} \in U$ e todo $t \in (0,1)$. Mostre que uma função estritamente convexa não pode possuir dois pontos de mínimo global diferentes.

[64] (**Funções estritamente côncavas e questões de unicidade**) Dizemos que uma função $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U de $\mathbb{R}^n$ é *estritamente côncava* se

$$f((1-t)\cdot \mathbf{p} + t \cdot \mathbf{q}) > (1-t)\cdot f(\mathbf{p}) + t \cdot f(\mathbf{q}),$$

para todo $\mathbf{p}, \mathbf{q} \in U$ e todo $t \in (0,1)$. Mostre que uma função estritamente côncava não pode possuir dois pontos de máximo global diferentes.

*[65] Verdadeira ou falsa? Seja U um subconjunto de $\mathbb{R}^n$ com a seguinte propriedade: para todo $\mathbf{p}$ e $\mathbf{q}$ em U, o ponto

$$\frac{\mathbf{p}+\mathbf{q}}{2} = \frac{1}{2}\cdot \mathbf{p} + \frac{1}{2}\cdot \mathbf{q}$$

também está em U. Então U é um subconjunto convexo de $\mathbb{R}^n$. Justifique sua resposta.

*[66] Seja U um subconjunto convexo de $\mathbb{R}^n$. Mostre que $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ é convexa se, e somente se,

$$g(t) = f(t \cdot \mathbf{p} + (1 - t) \cdot \mathbf{q})$$

é uma função convexa, onde $\mathbf{p}$ e $\mathbf{q}$ são pontos quaisquer em U e $t \in [0, 1]$. Em outras palavras, mostre que uma função definida em um conjunto convexo é convexa se, e somente se, sua restrição sobre qualquer segmento de reta em U é uma função convexa.

*[67] Seja $f\colon \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Mostre que se f é limitada superiormente, isto é, se existe uma constante $M \in \mathbb{R}$ tal que $f(\mathbf{x}) \leq M$ para todo $\mathbf{x} \in \mathbb{R}^n$, então f é uma função constante em $\mathbb{R}^n$.

*[68] **(A desigualdade de Jensen)**

(a) Mostre que uma função $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U de $\mathbb{R}^n$ é convexa se, e somente se,

$$f(\alpha_1 \cdot \mathbf{p} + \alpha_2 \cdot \mathbf{q}) \leq \alpha_1 \cdot f(\mathbf{p}) + \alpha_2 \cdot f(\mathbf{q}),$$

para todo $\mathbf{p}, \mathbf{q} \in U$ e para todo $\alpha_1, \alpha_2 \geq 0$ com $\alpha_1 + \alpha_2 = 1$.

(b) Mostre que uma função $f\colon U \subset \mathbb{R}^n \to \mathbb{R}$ definida em um conjunto convexo U de $\mathbb{R}^n$ é convexa se, e somente se,

$$f(\alpha_1 \cdot \mathbf{p}_1 + \cdots + \alpha_n \cdot \mathbf{p}_n) \leq \alpha_1 \cdot f(\mathbf{p}_1) + \cdots + \alpha_n \cdot f(\mathbf{p}_n),$$

para todo $\mathbf{p}_1, \ldots, \mathbf{p}_n \in U$ e para todo $\alpha_1, \ldots, \alpha_n \geq 0$ com

$$\alpha_1 + \cdots + \alpha_n = 1.$$

(c) Use o item anterior com a função convexa $y = f(x) = -\ln(x)$ para obter a desigualdade entre as médias aritmética e geométrica:

$$\sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n} \leq \frac{x_1 + x_2 + \cdots x_n}{n},$$

onde $x_1, x_2, \ldots, x_n$ são números reais positivos.

[69] Sejam $u_0 = \ln(x)$ e, para $0 < \alpha \leq 1$, $u_\alpha(x) = (x^\alpha - 1)/\alpha$, definidas no conjunto $U = \{x \in \mathbb{R} \mid x > 0\}$.

(a) Mostre que $u_0(x) = \lim_{\alpha \to 0^+} u_\alpha(x)$, para todo $x \in U$.

(b) Mostre que, para todo $\alpha \in [0, 1]$, a função u_α é côncava, estritamente crescente e satisfaz $u_\alpha(1) = 0$.

Estas funções são usadas em Economia para modelar o benefício ou utilidade em função da quantidade de bens ou dinheiro. A concavidade de u_α significa que a *utilidade marginal* (isto é, o acréscimo na utilidade obtido por um acréscimo fixo na quantidade de bens) decresce quando a quantidade de bens cresce. Em outras palavras, concavidade modela o efeito de *satisfação*.

[70] Diga se cada uma das seguintes sentenças abaixo é verdadeira ou falsa, apresentando uma justificativa caso ela seja verdadeira ou um contra-exemplo caso ela seja falsa.

(a) Seja U um subconjunto convexo e aberto de $\mathbb{R}^n$. Se f e g são funções *convexas* de classe C^∞ definidas em U, então $f + g$ também é uma função convexa em U.

(b) Seja U um subconjunto convexo e aberto de $\mathbb{R}^n$. Se f e g são funções *convexas* de classe C^∞ definidas em U, então $f \cdot g$ também é uma função convexa em U.

[71] Use o método dos mínimos quadrados para encontrar a reta que melhor aproxima os pontos $(0, 1)$, $(1, 3)$, $(2, 2)$, $(3, 4)$ e $(4, 5)$. Faça um desenho dos pontos e da reta.

[72] Use o método dos mínimos quadrados para encontrar a reta que melhor aproxima os pontos $(1, 1)$, $(2, 1)$, $(4, 0)$, $(7, 2)$ e $(9, 9)$. Faça um desenho dos pontos e da reta.

[73] Seja $y = a^*x + b^*$ a reta que "melhor" aproxima os pontos $(x_1, y_1), \ldots,$ (x_n, y_n) de acordo com o método dos mínimos quadrados.

(a) Mostre que

$$\sum_{i=1}^{n}(a^*x + b^* - y_i) = 0,$$

isto é, mostre que os desvios positivos e negativos se cancelam.

(b) Mostre que a reta $y = a^*x + b^*$ sempre passa pelo ponto

$$(\overline{x}, \overline{y}) = \left(\left(\sum_{i=1}^{n} x_i \right) \Big/ n, \left(\sum_{i=1}^{n} y_i \right) \Big/ n \right),$$

formado pelas médias aritméticas dos dados x_i e y_i, com $1 \leq i \leq n$.

*[74] (**Classificação de formas quadráticas via autovalores**) Seja A uma matriz quadrada $n \times n$. Dizemos que um número λ é *autovalor* de A se existe um vetor $\mathbf{x} \in \mathbb{R}^n$ (denominado *autovetor* de A), com $\mathbf{x} \neq 0$, tal que $A \cdot \mathbf{x} = \lambda \cdot \mathbf{x}$.

(a) Considere

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Quais dos vetores abaixo

$$\mathbf{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{z} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

são autovetores da matriz A? Quais são os autovalores associados?

Como encontrar os autovalores de uma matriz A sem ser por tentativa e erro? Observe: se $\mathbf{x}$ é um autovetor associado a um autovalor λ então sabemos que

$$A \cdot \mathbf{x} = \lambda \cdot \mathbf{x}$$

ou, ainda,

$$A \cdot \mathbf{x} - \lambda \cdot I \cdot \mathbf{x} = 0,$$

com I a matriz identidade $n \times n$. Desta maneira,

$$(A - \lambda \cdot I) \cdot \mathbf{x} = 0, \tag{11.12}$$

ou seja, o vetor $\mathbf{x}$ é uma solução não-nula do sistema linear homogêneo $(A - \lambda \cdot I) \cdot \mathbf{x} = 0$. Isto significa que o sistema linear possui infinitas soluções (sistema possível e indeterminado). A fim de que isto ocorra $A - \lambda \cdot I$ não pode ser uma matriz inversível, logo $\det(A - \lambda \cdot I) = 0$. Desta maneira, para se encontrar os autovalores de uma matriz A, basta encontrar as raízes do *polinômio característico* $p(\lambda) = \det(A - \lambda \cdot I)$.

(b) Calcule os autovalores da matriz A definida no item (a).

E como calcular os autovetores sabendo-se os autovalores? Basta substituir o valor do autovalor λ na equação (11.12) e encontrar uma (das infinitas) soluções do sistema formado. Por exemplo, $\lambda = -1$ é um autovalor da matriz A do item (a). Substituindo este valor em (11.12) temos

$$(A - \lambda \cdot I) \cdot \mathbf{x} = \left(\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} - (-1) \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} =$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

É fácil de ver que

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

é uma solução (entre infinitas) do sistema e, portanto, constitui um autovetor de A associado ao autovalor $\lambda = -1$. Para se resolver o sistema em uma situação geral você pode utilizar a técnica de escalonamento ensinada no segundo grau (lembra-se?).

(c) Encontre um autovetor associado ao outro autovalor da matriz A do item (a).

(d) Encontre os autovalores e os correspondentes autovetores da matriz

$$B = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

(e) Mostre que se $\mathbf{v}$ é um autovetor de uma matriz A associado ao autovalor λ, então o vetor $t \cdot \mathbf{v}$ (um múltiplo escalar do vetor $\mathbf{v}$) também é um autovetor de A associado ao mesmo autovalor λ, para todo $t \neq 0$.

Mas qual a relação entre autovalores e autovetores com o processo de se transformar uma matriz simétrica em uma matriz diagonal? Vamos ilustrar o processo através de um exemplo. Considere a matriz

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Os autovalores de A são $\lambda_1 = -1$ e $\lambda_2 = 3$, com os autovetores

$$\mathbf{x} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

respectivamente. Será interessante usar autovetores de tamanho 1. Para consegui-los, basta dividir $\mathbf{x}$ por $||\mathbf{x}||$ e $\mathbf{y}$ por $||\mathbf{y}||$ (note que, pelo item (e), dividir um vetor por um escalar não-nulo não destrói a propriedade do vetor ser autovetor para o mesmo autovalor):

$$\mathbf{u} = \frac{\mathbf{x}}{||\mathbf{x}||} = \begin{bmatrix} -1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix} \quad \text{e} \quad \mathbf{v} = \frac{\mathbf{y}}{||\mathbf{y}||} = \begin{bmatrix} +1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix}.$$

Uma vez que $\mathbf{u}$ e $\mathbf{v}$ são autovetores (de tamanho 1), temos

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix} = (-1) \cdot \begin{bmatrix} -1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix} \quad \text{e}$$

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} +1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix} = (3) \cdot \begin{bmatrix} +1/\sqrt{2} \\ +1/\sqrt{2} \end{bmatrix}.$$

Agora, o "pulo do gato" é escrever estas duas igualdades em uma única igualdade, colecionando os autovetores em uma única matriz:

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} -1/\sqrt{2} & +1/\sqrt{2} \\ +1/\sqrt{2} & +1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} -1/\sqrt{2} & +1/\sqrt{2} \\ +1/\sqrt{2} & +1/\sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix},$$

isto é

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot P = P \cdot \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix},$$

onde P é a matriz formada pelos autovetores. Multiplicando-se os dois lados desta última igualdade obtemos que $P^{-1} \cdot A \cdot P = D$ isto é,

$$P^{-1} \cdot \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot P = \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix},$$

Podemos ainda escrever que $A = P \cdot D \cdot P^{-1}$, isto é,

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = P \cdot \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \cdot P^{-1}.$$

Outro fato surpreendente é que $P^{-1} = P^T$ (^{T} indica transposição):

$$P^{-1} = \begin{bmatrix} -1/\sqrt{2} & +1/\sqrt{2} \\ +1/\sqrt{2} & +1/\sqrt{2} \end{bmatrix} = P^T,$$

isto é, a inversa e a transposta de P são iguais. Para que isto ocorresse, foi fundamental tomar os autovetores de A com tamanho 1. Resumindo: P é uma *matriz ortogonal* (isto é, $P \cdot P^T = I$) e vale que

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = P \cdot \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix} \cdot P^T.$$

(f) Diagonalize a matriz B do item (d), isto é, encontre uma matriz P tal que $B = P \cdot D \cdot P^{-1}$, com D uma matriz diagonal.

(g) Mostre que se P é uma matriz ortogonal (isto é, $P \cdot P^T = I$), então P é uma matriz inversível.

O que fizemos para a matriz A do item (a), pode ser feito para qualquer matriz simétrica, isto é, toda matriz simétrica $n \times n$ pode ser diagonalizada. Este é um resultado importante de Álgebra Linear e ele é conhecido como o Teorema Espectral.

Teorema 11.15 (Teorema Espectral) Seja A uma matriz simétrica $n \times n$. Então existe uma matriz P de tamanho $n \times n$, ortogonal (e, portanto, inversível), tal que

$$A = P \cdot \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \cdot P^T,$$

onde $d_1, d_2, \ldots, d_n$ são os autovalores da matriz A.

O Teorema Espectral permite usar os autovalores de uma matriz para classificar a forma quadrática correspondente. De fato, se A é uma matriz simétrica $n \times n$ então, pelo Teorema Espectral, existe uma matriz P ortogonal tal que $A = P \cdot D \cdot P^T$, onde os elementos da diagonal principal de D são justamente os n autovalores de D. Escrevendo $M = P^T$, de forma que $P = M^T$, temos que assim,

$$Q(\mathbf{h}) = \mathbf{h}^t \cdot A \cdot \mathbf{h} = \mathbf{h}^T \cdot M^T \cdot D \cdot M \cdot \mathbf{h} = \mathbf{x}^T \cdot D \cdot \mathbf{x} = \widetilde{Q}(\mathbf{x}),$$

onde $\mathbf{x} = M \cdot \mathbf{h}$ e $\widetilde{Q}(\mathbf{x}) = \mathbf{x}^T \cdot D \cdot \mathbf{x}$. Uma vez que M é uma matriz inversível, (pois M é ortogonal, veja o item (g)), segue-se que o tipo de positividade de Q é o mesmo tipo de positividade de $\widetilde{Q}$. Mais precisamente, mostre que

(h) Q é positiva definida se, e somente se, $\widetilde{Q}$ é positiva definida; Q é negativa definida se, e somente se, $\widetilde{Q}$ é negativa definida; etc.

Mas estudar a positividade de $\widetilde{Q}$ é muito fácil pois $\widetilde{Q}$ é a forma quadrática de uma matriz diagonal! Você pode então usar os resultados do exemplo (11.14) para concluir o próximo teorema.

Teorema 11.16 Seja A uma matriz simétrica.

(a) A é positiva definida se, e somente se, todos os seus autovalores são positivos.

(b) A é negativa definida se, e somente se, todos os seus autovalores são negativos.

(c) A é positiva semidefinida se, e somente se, todos os seus autovalores são maiores ou iguais a zero.

(d) A é negativa definida se, e somente se, todos os seus autovalores são menores ou iguais a zero.

(e) A é indefinida se, e somente se, existe pelo menos um autovalor positivo e pelo menos um autovalor negativo.

(i) Classifique a positividade das matrizes dos exercícios [16] a [23] usando o teorema (11.16). Qual método você prefere? Lagrange, menores principais ou autovalores?

11.8 Leitura suplementar

Polinômios de Taylor de ordem maior do que 2 podem ser considerados. O polinômio de Taylor p_k de ordem k no ponto $\mathbf{p}$ de uma função f de classe C^k é o único polinômio de grau k cujas derivadas parciais de ordem menor ou igual a k coincidem, respectivamente, com as derivadas parciais de ordem menor ou igual a k da função f no ponto $\mathbf{p}$. Contudo, se $k \geq 3$, não é possível representar p_k com o uso de matrizes, isto porque não existe uma representação matricial para o que seria "a derivada de ordem k" de uma função de n variáveis. Não vamos desenvolver a teoria dos polinômios de Taylor de ordem maior ou igual a 3 aqui. Indicamos ao leitor interessado as referências [01, 48, 71].

Pierre de Fermat (1601–1665) considerou o problema do cálculo de máximos, mínimos e retas tangentes em seus estudos do trabalho de Kepler sobre o paralelepípedo de maior volume que pode ser inscrito em uma esfera (veja a figura (11.24)) e do trabalho de Viete sobre a relação entre os coeficientes e as raízes de um polinômio. Para uma excelente descrição de como Fermat resolvia problemas de otimização (sem o uso de derivadas, que ainda não tinham sido inventadas na época), indicamos a referência [70].

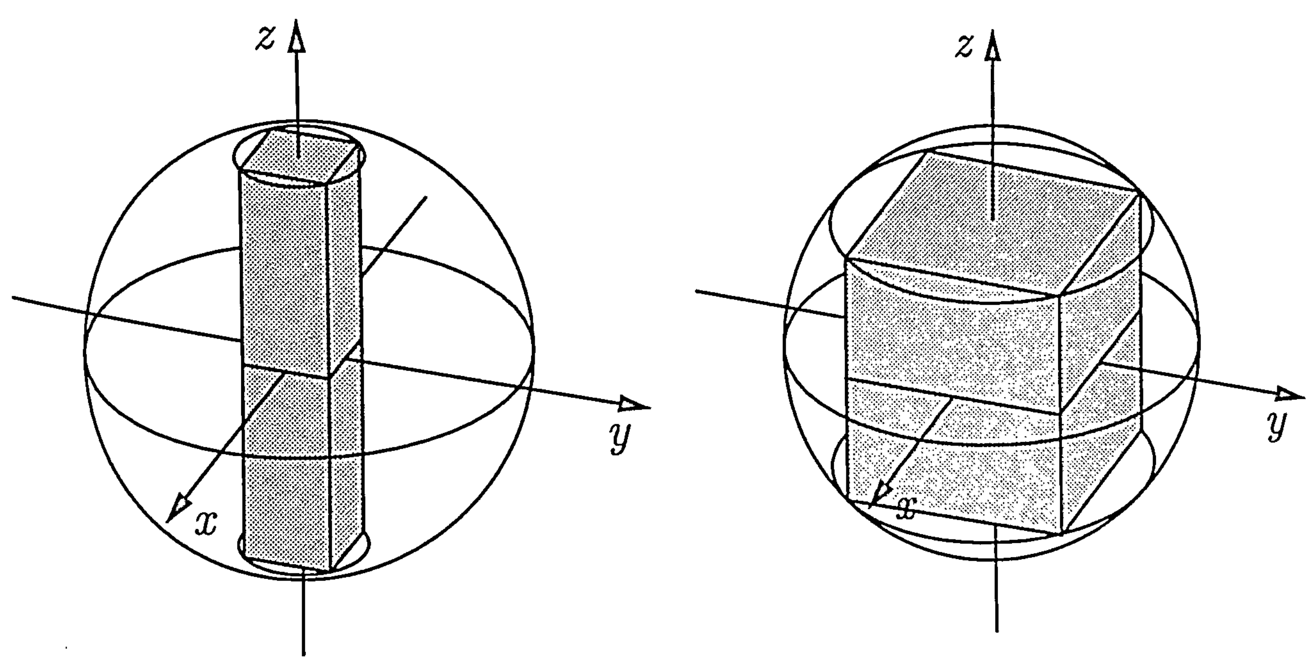

Figura 11.24: O problema de Kepler: encontre o paralelepípedo de maior volume inscrito em uma esfera.