Capítulo 4

Derivação de ordem superior. Aplicações

4.1 Derivadas parciais e diferenciais de ordem superior

Derivadas parciais

Como vimos, a derivada parcial origina, a partir de uma dada função escalar f, novas funções escalares $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$, então podemos determinar as suas derivadas parciais. Assim

Definição 4.1 *Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$. As derivadas parciais de $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$, nos pontos $\mathbf{a} \in \operatorname{int} D$ onde existem, chamam-se as* derivadas parciais de segunda ordem *(*segundas derivadas*) de f em $\mathbf{a}$:*

$$\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}(\mathbf{a})\right) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x_i}(\mathbf{a} + he_j) - \frac{\partial f}{\partial x_i}(\mathbf{a})}{h}, \quad i, j = 1, \ldots, n,$$

sempre que o limite exista, e denotam-se por

$$f''_{x_j x_i}(\mathbf{a}) \quad ou \quad \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) \quad \left(\frac{\partial^2 f}{\partial x_j^2}(\mathbf{a}) \text{ quando } i = j\right).$$

As derivadas parciais em que os indices i e j são diferentes chamam-se derivadas parciais mistas *(ou* derivadas parciais cruzadas*).*

Observação 4.2 *Nas derivadas parciais mistas o índice à direita indica a componente em relação à qual derivamos em primeiro lugar.*

Para funções de duas variáveis existem quatro derivadas parciais de segunda ordem, nomeadamente,

$$\frac{\partial^2 f}{\partial x^2}, \quad \frac{\partial^2 f}{\partial y \partial x}, \quad \frac{\partial^2 f}{\partial x \partial y}, \quad \frac{\partial^2 f}{\partial y^2}.$$

Sendo $(a, b) \in \mathbb{R}^2$ um ponto onde as derivadas de primeira ordem existem, temos, por exemplo,

$$\frac{\partial^2 f}{\partial x^2}(a, b) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x}(a + h, b) - \frac{\partial f}{\partial x}(a, b)}{h} \quad \text{e} \quad \frac{\partial^2 f}{\partial y \partial x}(a, b) = \lim_{h \to 0} \frac{\frac{\partial f}{\partial x}(a, b + h) - \frac{\partial f}{\partial x}(a, b)}{h}.$$

Exemplo 4.3 *Consideremos a função escalar definida em* $\mathbb{R}^2$ *por*

$$f(x,y) = xy^2 + xe^y$$

e (x,y) *um ponto arbitrário de* $\mathbb{R}^2$*.*

As derivadas parciais de primeira ordem de f *em* (x,y) *são*

$$\frac{\partial f}{\partial x}(x,y) = y^2 + e^y \quad e \quad \frac{\partial f}{\partial y}(x,y) = 2xy + xe^y,$$

e as de segunda ordem são

$$\frac{\partial^2 f}{\partial x^2}(x,y) = 0, \quad \frac{\partial^2 f}{\partial y \partial x}(x,y) = 2y + e^y, \quad \frac{\partial^2 f}{\partial y^2}(x,y) = 2x + xe^y \quad e \quad \frac{\partial^2 f}{\partial x \partial y}(x,y) = 2y + e^y.$$

Repare que, no exemplo anterior, temos igualdade entre as derivadas parciais mistas mas, como mostra o exemplo seguinte, isto nem sempre acontece.

Exemplo 4.4 *Consideremos a função* $f : \mathbb{R}^2 \to \mathbb{R}$ *dada por*

$$f(x,y) = \frac{xy^3}{x^2+y^2} \text{ para } (x,y) \neq (0,0) \quad e \quad f(0,0) = 0.$$

Mostremos que, no ponto $(0,0)$*, a função* f *tem derivadas parciais mistas diferentes.*

Temos

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(0,h) - \frac{\partial f}{\partial x}f(0,0)}{h} \quad e \quad \frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h\to 0} \frac{\frac{\partial f}{\partial y}(h,0) - \frac{\partial f}{\partial y}f(0,0)}{h},$$

sendo

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= \frac{y^3(x^2+y^2) - 2x^2y^3}{(x^2+y^2)^2} = \frac{y^3(-x^2+y^2)}{(x^2+y^2)^2} \\ \frac{\partial f}{\partial y}(x,y) &= \frac{3y^2x(x^2+y^2) - 2xy^4}{(x^2+y^2)^2} = \frac{y^2x(3x^2+y^2)}{(x^2+y^2)^2}, \end{aligned}$$

para $(x,y) \neq (0,0)$*, e*

$$\frac{\partial f}{\partial x}(0,0) = \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} = \lim_{h\to 0} \frac{\frac{0}{h^2}}{h} = 0 = \frac{\partial f}{\partial y}(0,0).$$

Portanto

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{h\to 0} \frac{\frac{\partial f}{\partial x}(0,h) - \frac{\partial f}{\partial x}f(0,0)}{h} = \lim_{h\to 0} \frac{\frac{h^3h^2}{h^4}}{h} = 1$$

e

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{h\to 0} \frac{\frac{\partial f}{\partial y}(h,0) - \frac{\partial f}{\partial y}f(0,0)}{h} = 0.$$

Vejamos sob que condições as derivadas parciais mistas são de certeza iguais.

Teorema 4.5 (Teorema de Schwarz) *Sejam* $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ *e* $(a,b) \in \operatorname{int} D$. *Se existem as derivadas parciais* $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ *e* $\frac{\partial^2 f}{\partial y \partial x}$ *numa vizinhança de* (a,b), *e são contínuas em* (a,b), *então existe* $\frac{\partial^2 f}{\partial x \partial y}(a,b)$ *e tem-se*

$$\frac{\partial^2 f}{\partial y \partial x}(a,b) = \frac{\partial^2 f}{\partial x \partial y}(a,b).$$

Demonstração: Ver [11, p. 202].

Observação 4.6 *O teorema anterior também é válido se trocarmos os papéis desempenhados pelas derivadas* $\frac{\partial^2 f}{\partial y \partial x}$ *e* $\frac{\partial^2 f}{\partial x \partial y}$.

Exemplo 4.7 *A função* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida por*

$$f(x,y) = xy^2 + xe^y$$

verifica o Teorema de Schwarz em qualquer ponto $(x,y) \in \mathbb{R}^2$ *(ver Exemplo 4.3).*

Exemplo 4.8 *A função* $f : \mathbb{R}^2 \to \mathbb{R}$ *dada por*

$$f(x,y) = \frac{xy^3}{x^2+y^2} \text{ para } (x,y) \neq (0,0) \quad \text{e} \quad f(0,0) = 0,$$

já considerada no Exemplo 4.4, satisfaz as condições do Teorema de Schwarz no ponto $(1,1)$, *mas não as satisfaz no ponto* $(0,0)$. *Justifique.*

A partir das derivadas parciais de segunda ordem podemos definir derivadas parciais de terceira ordem e assim sucessivamente. Por exemplo, a derivada de terceira ordem de f num ponto $\mathbf{a}$ relativamente às variáveis x_i, x_j e x_k, define-se como

$$\frac{\partial}{\partial x_k}\left(\frac{\partial^2 f}{\partial x_j \partial x_i}\right)(\mathbf{a})$$

e denota-se por

$$\frac{\partial^3 f}{\partial x_k \partial x_j \partial x_i}(\mathbf{a})$$

ou

$$f'''_{x_k x_j x_i}(\mathbf{a}).$$

No caso $i = j$ usamos o símbolo

$$\frac{\partial^3 f}{\partial x_k \partial x_i^2}(\mathbf{a}) \quad \text{ou} \quad f'''_{x_k x_i x_i}(\mathbf{a})$$

(lê-se "dê três por x_k e x_i quadrado").

Observação 4.9 *O Teorema de Schwarz foi propositadamente enunciado para o caso de funções de duas variáveis e de derivadas parciais de segunda ordem mas pode ser estendido a situações mais gerais. Por exemplo, para uma função de três variáveis é válida a igualdade*

$$\frac{\partial^3 f}{\partial z \partial y \partial x}(\mathbf{a}) = \frac{\partial^3 f}{\partial y \partial x \partial z}(\mathbf{a}) = \frac{\partial^3 f}{\partial y \partial z \partial x}(\mathbf{a}) = \frac{\partial^3 f}{\partial z \partial x \partial y}(\mathbf{a}),$$

desde que todas as derivadas em causa existam e sejam contínuas.

Em geral, tendo em conta o Teorema de Schwarz e juntando a derivação repetida em ordem a uma mesma variável, podemos escrever qualquer derivada de ordem m na forma

$$\frac{\partial^m f}{\partial x_1^{i_1} ... \partial x_n^{i_n}}(\mathbf{a}),$$

onde $i_1, ..., i_n \in \mathbb{N} \cup \{0\}$ são tais que $i_1 + ... + i_n = m$. Se $i_p = 0$ para algum $p = 1, ..., n$, então a função f não se deriva em ordem à variável x_p.

Definição 4.10 *Seja $f : D \to \mathbb{R}$, onde $D \subset \mathbb{R}^n$ é um conjunto aberto.*

(i) Diz-se que f é uma função de classe $\mathcal{C}^k$ *em D, $k \in \mathbb{N}$, (e escreve-se $f \in \mathcal{C}^k(D)$) se f possui todas as derivadas parciais até à ordem k (inclusivé) e estas são contínuas em todos os pontos de D.*

(ii) Diz-se que f é indefinidamente diferenciável *em D ou que f é de* classe $\mathcal{C}^\infty$ *em D (e escreve-se $f \in \mathcal{C}^\infty(D)$) se $f \in \mathcal{C}^k(D)$ para todo $k \in \mathbb{N}$.*

Exemplo 4.11 *A função $f : \mathbb{R}^3 \to \mathbb{R}$ definida por*

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

é de classe $\mathcal{C}^\infty(\mathbb{R})$. (Justifique!)

Diferenciais de ordem superior

Recordemos que o diferencial de uma função $f : D \subset \mathbb{R}^n \to \mathbb{R}$, num ponto $\mathbf{a} \in \operatorname{int} D$, é uma forma linear

$$\mathbf{h} \mapsto \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a}) h_i.$$

Por analogia, o *segundo diferencial* (denotado por $d^2 f(\mathbf{a})$) pode ser definido como uma forma quadrática

$$\mathbf{h} \mapsto \mathbf{h}^T \mathcal{H} f(\mathbf{a}) \mathbf{h} = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) h_i h_j,$$

onde $\mathcal{H}f(\mathbf{a})$ é a *matriz Hessiana* de f em $\mathbf{a}$ definida por

$$\mathcal{H}f(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix}.$$

Exemplo 4.12 *Seja $f : \mathbb{R}^3 \to \mathbb{R}$ a função dada por $f(x,y,z) = xyz$.*

Tem-se

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

e

$$\mathcal{H}f(x,y,z) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial z \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial z \partial y} \\ \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial y \partial z} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} = \begin{bmatrix} 0 & z & y \\ z & 0 & x \\ y & x & 0 \end{bmatrix}.$$

Da mesma maneira se define o diferencial de terceira ordem

$$d^3 f(\mathbf{a})(\mathbf{h}) = \sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{a}) h_i h_j h_k$$

(uma forma homogénea de grau três), etc.

A variável $\mathbf{h}$ significa o acréscimo de $\mathbf{x}$ e costuma ser denotada por $\Delta\mathbf{x}$ ou $d\mathbf{x} = (dx_1, ..., dx_n)$ (e também se chama diferencial da variável independente).

Tendo em conta o Teorema de Schwarz e somando as derivadas mistas em ordem às mesmas variáveis obtemos a seguinte representação do diferencial de ordem m :

$$d^m f(\mathbf{a}) = \sum_{i_1+...+i_n=m} \frac{m!}{i_1!...i_n!} \frac{\partial^m f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a})\, dx_1^{i_1}...\, dx_n^{i_n}. \tag{4.1}$$

Em particular, no caso de duas variáveis x e y temos

$$d^m f(\mathbf{a}) = \sum_{i=0}^{m} C_i^m \frac{\partial^m f}{\partial x^i \partial y^{m-i}}(\mathbf{a})\, dx^i dy^{m-i}, \tag{4.2}$$

onde

$$C_i^m = \frac{m!}{i!\,(m-i)!}$$

são os coeficientes binomiais. Se introduzimos o operador formal

$$\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy$$

a fórmula (4.2) pode ser escrita na forma simplificada

$$d^m f(\mathbf{a}) = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^{\{m\}} f(\mathbf{a}),$$

onde $\{m\}$ significa que, primeiro, aplicamos a fórmula do binómio de Newton, ao operador $\frac{\partial}{\partial x}dx+\frac{\partial}{\partial y}dy$ e, depois, substituímos todos os produtos de $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$ e f pela derivada de f respectiva. Analogamente, ver (4.1), no caso de uma função de n variáveis temos

$$d^m f(\mathbf{a}) = \left(\frac{\partial}{\partial x_1}dx_1 + ... + \frac{\partial}{\partial x_n}dx_n\right)^{\{m\}} f(\mathbf{a}).$$

Podemos chegar ao conceito de diferencial de ordem superior de outra forma. Nomeadamente, lembrando que o primeiro diferencial df é uma função de $2n$ variáveis ($x_1, ..., x_n$ e $dx_1, ..., dx_n$), linear relativamente ao segundo grupo de variáveis, define-se o segundo diferencial $d^2 f$ como o diferencial de df (derivando em ordem a todas as variáveis).

Exemplo 4.13 *Consideremos a função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por*

$$f(x, y) = x^2 y^3.$$

Derivando a primeira vez temos (ver Proposição 2.12)

$$df = y^3 d\left(x^2\right) + x^2 d\left(y^3\right) = 2y^3 x dx + 3x^2 y^2 dy.$$

Então o segundo diferencial de f é

$$\begin{aligned} d^2 f &= d(df) = d\left(2y^3 x dx + 3x^2 y^2 dy\right) \\ &= 2d\left(y^3 x dx\right) + 3d\left(x^2 y^2 dy\right) \\ &= 2\left(d\left(y^3\right) x dx + y^3 dx dx + y^3 x d^2 x\right) + \\ &\quad +3\left(d\left(x^2\right) y^2 dy + x^2 d\left(y^2\right) dy + x^2 y^2 d^2 y\right) \\ &= 2\left(3y^2 x dy dx + y^3 dx^2 + y^3 x d^2 x\right) + \\ &\quad +3\left(2xy^2 dx dy + 2x^2 y dy^2 + x^2 y^2 d^2 y\right) \\ &= 2y^3 dx^2 + 12y^2 x dy dx + 6x^2 y dy^2 + 2y^3 x d^2 x + 3x^2 y^2 d^2 y. \end{aligned} \tag{4.3}$$

Uma vez que x e y são variáveis independentes os diferenciais $d^2 x$ e $d^2 y$ são nulos e, portanto, temos, finalmente,

$$\begin{aligned} d^2 f &= 2y^3 dx^2 + 12y^2 x dy dx + 6x^2 y dy^2 \\ &= (dx, dy)^T \mathcal{H}f\,(dx, dy), \end{aligned}$$

onde a matriz Hessiana

$$\mathcal{H}f = \begin{bmatrix} 2y^3 & 6y^2 x \\ 6y^2 x & 6x^2 y \end{bmatrix}.$$

Observemos que a fórmula (4.3) é válida também quando x e y não são variáveis independentes, isto é, quando dependem de outras variáveis. Por exemplo, se $x = \cos t$, $y = \sin t$, temos

$$dx = -\sin t\ dt; \quad dy = \cos t\ dt$$

e

$$d^2x = -\cos t\ dt^2; \quad d^2y = -\sin t\ dt^2,$$

logo substituindo em (4.3), obtemos

$$\begin{aligned}
d^2f &= 2y^3dx^2 + 12y^2xdydx + 6x^2ydy^2 + 2y^3xd^2x + 3x^2y^2d^2y \\
&= 2\sin^3 t\,(-\sin t\ dt)^2 + 12\sin^2 t\cos t\,(\cos t\ dt)\,(-\sin t\ dt) + 6\cos^2 t\sin t\,(\cos t\ dt)^2 \\
&\quad +2\sin^3 t\cos t\left(-\cos t\ dt^2\right) + 3\cos^2 t\sin^2 t\left(-\sin t\ dt^2\right) \\
&= 2\sin^5 t\ dt^2 - 12\sin^3 t\cos^2 t dt^2 + 6\cos^4 t\sin t\ dt^2 \\
&\quad -2\sin^3 t\cos^2 t\ dt^2 - 3\cos^2 t\sin^3 t\ dt^2 \\
&= \left(2\sin^5 t - 14\sin^3 t\cos^2 t + 6\cos^4 t\sin t - 3\cos^2 t\sin^3 t\right) dt^2.
\end{aligned}$$

Observação 4.14 *Como vimos no exemplo anterior a regra de invariância da forma do primeiro diferencial (ver Teorema 2.31) não é válida para os diferenciais de segunda ordem, porque na representação de d^2f aparecem os segundos diferenciais d^2x e d^2y, que podem não ser nulos se x e y forem variáveis dependentes.*

Como definido antes o segundo diferencial d^2f também depende de $2n$ variáveis $x_1, ..., x_n$ e $dx_1, ..., dx_n$. Então derivando agora d^2f em ordem a todas as variáveis e tendo em conta que $d^2x_1 = ... = d^2x_n = 0$ obtemos o diferencial d^3f de terceira ordem, etc. Em geral, diferencial de ordem m encontra-se pela fórmula iterada

$$d^m f = d\left(d^{m-1}f\right).$$

Observação 4.15 *Temos que ter cuidado com a simbologia. Enquanto que d^kf, d^kx, d^ky, etc denotam os diferenciais de ordem k das respectivas variáveis (ou funções), os símbolos dx^k, dy^k,... representam, a k-ésima potência das variáveis dx, dy, ... (ou dos primeiros diferenciais das variáveis x, y, etc).*

4.2 Fórmula de Taylor

Como sabemos da Análise Matemática I a fórmula de Taylor de uma função real de variável real permite, quando verificadas certas condições, aproximar uma função diferenciável f por um polinómio. Tal aproximação usa-se, por exemplo, para levantar indeterminações, e para aproximar uma função com a maior exactidão possivel. Notemos que a derivada já dá uma boa aproximação da função em torno de um ponto fixo, mas esta aproximação é apenas linear no sentido em que o erro cometido com tal aproximação é um infinitésimo de ordem superior a 1. Usando a segunda derivada podemos aumentar o grau de exactidão, de tal maneira que o erro fica de ordem superior a 2. Considerando derivadas de ordem superior podemos chegar a erros de qualquer ordem de exactidão.

Recordemos o resultado que nos dá a fórmula de Taylor no caso real. Para mais pormenores ver, por exemplo, [2, p. 400].

Teorema 4.16 *Sejam p um número natural, I um intervalo aberto de $\mathbb{R}$, $f : I \to \mathbb{R}$ uma função com derivadas $f', f'', ..., f^{(p)}$ contínuas em I. Se $a \in I$, então para todo o $x \in I$ tem-se*

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + ... + \frac{f^{(p)}(a)}{p!}(x-a)^p + \varepsilon(x), \tag{4.4}$$

onde o resto verifica

$$\lim_{x\to a}\frac{\varepsilon(x)}{(x-a)^p}=0$$

(uma tal função $\varepsilon(x)$ *diz-se um infinitésimo de ordem superior a* p *quando* $x\to a$*, ou um o-pequeno de* $(x-a)^p$ *e denota-se por* $o(x-a)^p$*, ver também Secção 2.2).*

A fórmula (4.4) chama-se *Fórmula de Taylor de ordem* (ou *grau*) p *de* f *no ponto* a.

Observamos que na fórmula (4.4) a diferença $x-a$ é simplesmente o acréscimo do argumento x em torno do ponto a que costuma ser denotado por Δx ou dx (chamado também diferencial da variável independente). Analogamente, a diferença $f(x)-f(a)$ é o acréscimo da função f $(\Delta f(x))$. Além disso, o produto $f^{(i)}(a)\,dx^i$, $i=1,...,p$, é, por definição, o diferencial de ordem i de f. Logo a fórmula (4.4) pode ser escrita na forma

$$\Delta f(x)=df(a)+\frac{1}{2!}d^2f(a)+...+\frac{1}{p!}d^pf(a)+o(\Delta x^p).$$

Notemos ainda que, no caso em que a função f tem derivada de ordem $p+1$ contínua no ponto a, o resto na fórmula acima pode ser escrito na *forma de Lagrange*:

$$\frac{1}{(p+1)!}d^{p+1}f(a+\theta x),$$

onde θ é algúm número entre 0 e 1.

No caso de funções de várias variáveis podemos demonstrar fórmulas parecidas às anteriores.

Teorema 4.17 (Fórmula de Taylor com resto de Lagrange) *Sejam* $f: D\subset\mathbb{R}^n\to\mathbb{R}$ *e* $\mathbf{a}\in\operatorname{int}D$*. Se* f *tem todas as derivadas parciais até à ordem* $p+1$ *(inclusivé) contínuas numa vizinhança de* $\mathbf{a}$*, então para todo* $\mathbf{x}$ *nessa vizinhança tem-se*

$$f(\mathbf{x})-f(\mathbf{a})=df(\mathbf{a})+\frac{1}{2!}d^2f(\mathbf{a})+...+\frac{1}{p!}d^pf(\mathbf{a})+\frac{1}{(p+1)!}d^{p+1}f(\mathbf{a}+\theta\mathbf{x}), \tag{4.5}$$

para algum $\theta\in\,]0,1[$.

Demonstração:

Fixado um ponto $\mathbf{x}$ na vizinhança de $\mathbf{a}$ consideramos a função de uma variável real seguinte

$$\varphi(t)=f(\mathbf{a}+t\Delta\mathbf{x}),$$

onde $\Delta\mathbf{x}=(\Delta x_1,...,\Delta x_n)$; $\Delta x_i=x_i-a_i$, $i=1,..,n$, são as coordenadas do vector $\mathbf{x}-\mathbf{a}$. Como a função f é de classe $\mathcal{C}^{p+1}$ a função real $\varphi:[0,1]\to\mathbb{R}$ tem derivadas contínuas até à ordem $p+1$ (inclusivé). Aplicando a fórmula de Taylor a φ com resto na forma de Lagrange, temos

$$\varphi(1)=\varphi(0)+\frac{1}{2!}\varphi'(0)+...+\frac{1}{p!}\varphi^{(p)}(0)+\frac{1}{(p+1)!}\varphi^{(p+1)}(\theta), \tag{4.6}$$

para algum $\theta \in]0,1[$. Tendo em conta que $\varphi(0) = f(\mathbf{a})$, $\varphi(1) = f(\mathbf{x})$, basta apenas calcular as derivadas $\varphi^{(i)}(0)$, $i = 1, ..., p$, e $\varphi^{(p+1)}(\theta)$ (representando-as em termos das derivadas parciais de f).

Usando a *regra da cadeia* temos

$$\varphi'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a} + t\Delta\mathbf{x})\,\Delta x_i, \tag{4.7}$$

donde $\varphi'(0) = df(\mathbf{a})$. Aplicando novamente a regra da cadeia, de (4.7) sai que

$$\begin{aligned} \varphi''(t) &= \sum_{i=1}^{n} \Delta x_i \left(\frac{\partial f}{\partial x_i}(\mathbf{a} + t\Delta\mathbf{x})\right)' \\ &= \sum_{i=1}^{n} \Delta x_i \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a} + t\Delta\mathbf{x})\,\Delta x_j \\ &= \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a} + t\Delta\mathbf{x})\,\Delta x_i \Delta x_j \\ &= d^2 f(\mathbf{a} + t\Delta\mathbf{x}). \end{aligned}$$

Logo $\varphi''(0) = d^2 f(\mathbf{a})$. Assim, aplicando sucessivamente a regra da cadeia encontramos

$$\varphi^{(k)}(t) = d^k f(\mathbf{a} + t\Delta\mathbf{x}), \quad k = 1, ..., p, p+1,$$

donde $\varphi^{(k)}(0) = d^k f(\mathbf{a})$, $k = 1, ..., p$, e $\varphi^{(p+1)}(\theta) = d^{p+1} f(\mathbf{a} + \theta\Delta\mathbf{x})$. Substituindo todas as derivadas na fórmula (4.6), chegamos a (4.5). ■

Observação 4.18 *O resto na fórmula (4.5), para qualquer* $\mathbf{h} \in \mathbb{R}^n$, *pode ser representado como*

$$\frac{1}{(p+1)!} d^{p+1} f(\mathbf{a} + \theta\mathbf{x})(\mathbf{h}) = \frac{1}{(p+1)!} \|\mathbf{h}\|^{p+1} d^{p+1} f(\mathbf{a} + \theta\mathbf{x}) \left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right),$$

uma vez que $d^{p+1} f(\mathbf{a} + \theta\mathbf{x})$ *é uma forma de grau* $p+1$ *(isto é, representa-se como uma soma de várias parcelas do tipo* $A dx_1^{i_1} ... dx_n^{i_n}$, *onde* A *é um número real e* $i_1 + ... + i_n = p+1$, *ver (4.1)). Como a aplicação* $\mathbf{h} \mapsto d^{p+1} f(\mathbf{a} + \theta\mathbf{x})(\mathbf{h})$ *é contínua logo é limitada na esfera* $\{\mathbf{h} : \|\mathbf{h}\| = 1\}$, *e daqui sai, imediatamente, que*

$$\frac{1}{(p+1)!} \frac{d^{p+1} f(\mathbf{a} + \theta\mathbf{x})(\mathbf{h})}{\|\mathbf{h}\|^p} = \frac{1}{(p+1)!} \|\mathbf{h}\| \, d^{p+1} f(\mathbf{a} + \theta\mathbf{x}) \left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right)$$

tende para zero quando $\mathbf{h} \to \mathbf{0}$, *ou, por outras palavras, que*

$$\frac{1}{(p+1)!} d^{p+1} f(\mathbf{a} + \theta\mathbf{x})(\mathbf{h}) = o(\|\mathbf{h}\|^p).$$

Tendo em conta que $\mathbf{h} = \mathbf{x} - \mathbf{a}$, *chegamos à fórmula de Taylor com resto na forma de Peano:*

$$f(\mathbf{x}) - f(\mathbf{a}) = df(\mathbf{a}) + \frac{1}{2!} d^2 f(\mathbf{a}) + ... + \frac{1}{p!} d^p f(\mathbf{a}) + o(\|\mathbf{x} - \mathbf{a}\|^p). \tag{4.8}$$

Observação 4.19 *Notamos que a fórmula (4.8) se verifica se a função f é apenas continuamente diferenciável até à ordem p (inclusivé). De facto, representamos primeiro o acréscimo $f(\mathbf{x}) - f(\mathbf{a})$ através da fórmula (4.5) mas apenas de ordem p:*

$$f(\mathbf{x}) - f(\mathbf{a}) = df(\mathbf{a}) + \frac{1}{2!}d^2 f(\mathbf{a}) + ... + \frac{1}{p!}d^p f(\mathbf{a} + \theta\mathbf{x}), \tag{4.9}$$

para algum $\theta \in]0,1[$. Pela fórmula (4.1) temos

$$\begin{aligned} &\frac{1}{p!}\left(d^p f(\mathbf{a} + \theta\mathbf{x}) - d^p f(\mathbf{a})\right) \\ = &\frac{1}{p!}\sum_{i_1+...+i_n=p} \frac{p!}{i_1!...i_n!}\left(\frac{\partial^p f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a} + \theta\mathbf{x}) - \frac{\partial^p f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a})\right) h_1^{i_1}...h_n^{i_n}, \end{aligned}$$

onde $h_k = dx_k$, $k = 1, ..., n$. Portanto

$$\begin{aligned} &\frac{1}{p!}\frac{d^p f(\mathbf{a} + \theta\mathbf{x}) - d^p f(\mathbf{a})}{\|\mathbf{h}\|^p} \\ = &\sum_{i_1+...+i_n=p} \frac{1}{i_1!...i_n!}\left(\frac{\partial^p f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a} + \theta\mathbf{x}) - \frac{\partial^p f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a})\right)\left(\frac{h_1}{\|\mathbf{h}\|}\right)^{i_1}...\left(\frac{h_n}{\|\mathbf{h}\|}\right)^{i_n} \end{aligned}$$

e

$$\begin{aligned} &\frac{1}{p!}\frac{\|d^p f(\mathbf{a} + \theta\mathbf{x}) - d^p f(\mathbf{a})\|}{\|\mathbf{h}\|^p} \\ \leq &\sum_{i_1+...+i_n=p} \frac{1}{i_1!...i_n!}\left|\frac{\partial^p f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a} + \theta\mathbf{x}) - \frac{\partial^p f}{\partial x_1^{i_1}...\partial x_n^{i_n}}(\mathbf{a})\right| \end{aligned}$$

tende para zero quando $\mathbf{x} \to \mathbf{a}$ (recordemos que todas as derivadas parciais de ordem p são contínuas). Isto mostra que

$$\frac{1}{p!}d^p f(\mathbf{a} + \theta\mathbf{x}) - d^p f(\mathbf{a}) = o\left(\|\mathbf{h}\|^p\right).$$

Substituindo na fórmula (4.9) obtemos

$$f(\mathbf{x}) - f(\mathbf{a}) = df(\mathbf{a}) + \frac{1}{2!}d^2 f(\mathbf{a}) + ... + \frac{1}{p!}d^p f(\mathbf{a}) + o\left(\|\mathbf{h}\|^p\right).$$

Exemplo 4.20 *Consideremos a função $f : \mathbb{R}^3 \to \mathbb{R}$ definida por $f(x,y,z) = xyz$, e escrevemos a fórmula de Taylor de f de ordem 2 no ponto $(1,0,1)$.*

Pelo Exemplo 4.12 temos

$$f(1,0,1) = 0, \quad \nabla f(1,0,1) = (0,1,0) \quad e \quad \mathcal{H}f(1,0,1) = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Então a fórmula de Taylor é

$$\begin{aligned} xyz &= 0+\langle (0,1,0),(x-1,y,z-1)\rangle + \frac{1}{2}\begin{bmatrix} x-1 & y & z-1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x-1 \\ y \\ z-1 \end{bmatrix}+ \\ & \quad +\varepsilon(x,y,z). \end{aligned}$$

Como

$$\begin{aligned} & \begin{bmatrix} x-1 & y & z-1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} x-1 \\ y \\ z-1 \end{bmatrix} \\ = & \begin{bmatrix} y & x+z-2 & y \end{bmatrix}\begin{bmatrix} x-1 \\ y \\ z-1 \end{bmatrix} \\ = & \; y(x-1)+y(x+z-2)+y(z-1), \end{aligned}$$

então

$$xyz = y+\frac{1}{2}y(x-1+x+z-2+z-1)+\varepsilon(x,y,z).$$

É fácil mostrar que

$$\varepsilon(x,y,z) = y(x-1)(z-1)$$

e que

$$\lim_{(x,y,z)\to(1,0,1)} \frac{\varepsilon(x,y,z)}{\|(x,y,z)-(1,0,1)\|^2} = 0,$$

logo $\varepsilon(x,y,z) = o\left(\|(x-1,y,z-1)\|^2\right)$

Observação 4.21 *Tal como no caso $n=1$, a Fórmula de Taylor de ordem 2 pode ser usada para estudar extremos locais, analisando o sinal de $f(x)-f(a)$, como veremos mais à frente.*

4.3 Extremos locais

Começamos com funções de duas variáveis. Sabemos que o gráfico de uma função $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ diferenciável num ponto $(a,b) \in \operatorname{int} D$ admite um *plano tangente* em $(a,b,f(a,b))$, dado pela equação

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

Se $\frac{\partial f}{\partial x}(a,b) = \frac{\partial f}{\partial y}(a,b) = 0$ este plano é horizontal, isto é paralelo ao plano coordenado xy. Neste caso o ponto (a,b) diz-se um *ponto de estacionaridade* de f. Tais pontos (com excepção de alguns casos específicos) podem ser de três categorias básicas: maximizantes, minimizantes e pontos sela. Se a superfície (gráfico de f) se imaginar como um terreno montanhoso, essas categorias correspondem, respectivamente, aos cumes das montanhas, ao fundo dos vales e às gargantas.

Passemos então às definições rigorosas.

Definição 4.22 *Sejam $f : D \subset \mathbb{R}^n \to \mathbb{R}$ e $\mathbf{a} \in D$.*

(i) Diz-se que $\mathbf{a}$ *é um* maximizante *(resp.* minimizante*)* global *de f em D se*

$$f(\mathbf{x}) \leq f(\mathbf{a}) \quad \textit{(resp. } f(\mathbf{x}) \geq f(\mathbf{a})\textit{)} \tag{4.10}$$

para todo $\mathbf{x} \in D$. Neste caso diz-se que $f(\mathbf{a})$ é um máximo *(resp.* mínimo*)* global *de f.*

(ii) Diz-se que $\mathbf{a}$ *é um* maximizante *(resp.* minimizante*)* local *de f e que $f(\mathbf{a})$ é um* máximo *(resp.* mínimo*)* local *de f se a desigualdade (4.10) é verificada por todo o* $\mathbf{x}$ *pertencente a alguma bola $B_r(\mathbf{a})$ contida em D, ou seja, se*

$$f(\mathbf{x}) \leq f(\mathbf{a}) \quad \textit{(resp. } f(\mathbf{x}) \geq f(\mathbf{a})\textit{)},$$

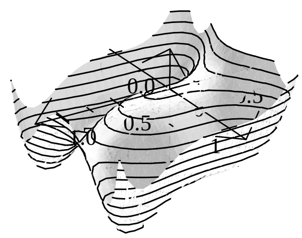
para todo $\mathbf{x} \in D \cap B_r(\mathbf{a})$.

Usamos o termo *extremo* (*extremo local*) para nos referirmos ao máximo ou ao mínimo. Em vez de maximizante (minimizante) também se usa o termo *ponto de máximo* (*ponto de mínimo*, resp.)

Exemplo 4.23 *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ uma função definida por*

$$f(x, y) = x^6 + y^6 - x^3 - y^2.$$

Pela análise do gráfico da função:



vemos que f tem um máximo local em $(0, 0)$, que não é maximizante global porque a função não é majorada. Além disso, e ainda pela análise do gráfico, a função f possui quatro minimizantes, que um estudo mais aprofundado mostra que são globais.

Condição necessária de extremo local

Teorema 4.24 *Supomos que para uma função escalar $f : D \subset \mathbb{R}^n \to \mathbb{R}$ existe uma derivada parcial $\frac{\partial f}{\partial x_i}$ $(i = 1, ..., n)$ num ponto $\mathbf{a} \in \operatorname{int} D$ e que $\mathbf{a}$ é um maximizante ou um minimizante local de f. Então $\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$.*

Demonstração:

Suponhamos que $f(\mathbf{a})$ é máximo local (o caso em que é mínimo local é análogo). Então existe $\delta > 0$ tal que $B_\delta(\mathbf{a}) \subset D$ e $f(\mathbf{x}) \leq f(\mathbf{a})$ para todo $\mathbf{x} \in B_\delta(\mathbf{a}) \subset D$. Em particular, $f(\mathbf{a}) \geq f(\mathbf{a} + h\mathbf{e}_i)$ para todo o $h \in \mathbb{R}$ suficientemente pequeno, onde $\mathbf{e}_i$ é o vector unitário do i-esimo eixo coordenado. Logo, para $h > 0$ temos

$$\frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h} \leq 0, \tag{4.11}$$

enquanto que para $h < 0$ temos

$$\frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h} \geq 0. \tag{4.12}$$

Uma vez que existe a derivada parcial

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h},$$

de (4.11) sai que

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0+} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h} \leq 0.$$

Analogamente, de (4.12) temos

$$\frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{h \to 0-} \frac{f(\mathbf{a} + h\mathbf{e}_i) - f(\mathbf{a})}{h} \geq 0.$$

Assim, concluimos que $\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$. ■

Quando a função f é diferenciável chegamos à seguinte *condição necessária de extremo local*:

Teorema 4.25 *Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ uma função escalar diferenciável num ponto $\mathbf{a} \in \operatorname{int} D$. Se $\mathbf{a}$ é um maximizante ou um minimizante local de f então $df(\mathbf{a}) = 0$.*

Demonstração:

Da representação do diferencial de uma função diferenciável sai, imediatamente, que

$$df(\mathbf{a}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\mathbf{a})\, dx_i = 0,$$

isto é $df(\mathbf{a})$ é uma forma linear nula. ■

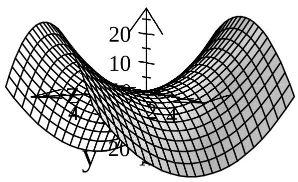
Definição 4.26 *Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável num ponto $\mathbf{a} \in \operatorname{int} D$. Diz-se que $\mathbf{a}$ é um* ponto de estacionaridade *de f se $d\,f(\mathbf{a}) = 0$.*

No caso em que $n = 2$, o Teorema 4.25 afirma que no ponto $(\mathbf{a}, f(\mathbf{a}))$ de máximo ou de mínimo local da função f existe um plano tangente ao gráfico de f paralelo ao plano-xy. Por outro lado, é fácil construir exemplos em que $\nabla f(\mathbf{a}) = \mathbf{0}$ mas em que $f(\mathbf{a})$ não é extremo local de f.

Exemplo 4.27 *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ a função definida por*

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

O gráfico da função é



Temos

$$\frac{\partial f}{\partial x}(x,y) = 2x \quad e \quad \frac{\partial f}{\partial y}(x,y) = -2y.$$

Logo a função tem um único ponto de estacionaridade, o ponto $(0,0)$. No entanto este ponto não dá máximo nem mínimo a f em nenhuma vizinhança da origem, uma vez que $f(0,0) = 0$ é máximo de $f(0,y)$ para todo $y \in \mathbb{R}$, enquanto que $f(0,0)$ é mínimo de $f(x,0)$ para $x \in \mathbb{R}$. Um ponto com esta propriedade (maximizante numa uma direcção e minimizante noutra) chama-se ponto sela.

Exemplo 4.28 *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ a função definida por*

$$f(x,y) = x^3.$$

Temos $\frac{\partial f}{\partial x} = 3x^2$ enquanto $\frac{\partial f}{\partial y} = 0$. Logo qualquer ponto da forma $(0,y_0)$ é ponto de estacionaridade de f. Por outro lado, em qualquer vizinhança de $(0,y_0)$ a função pode admitir valores positivos e valores negativos. Assim $(0,y_0)$ não é ponto de extremo local (nem ponto sela). Isto é um análogo três-dimensional de um ponto de inflexão. *Trata-se, portanto, de um caso degenerado.*

Condição suficiente de extremo local

Para averiguar a natureza dos pontos de estacionaridade (se são, de facto, minimizantes ou maximizantes) utiliza-se a diferenciabilidade de segunda ordem e as propriedades do segundo diferencial. Chegamos assim à condição suficiente que resulta da aplicação da fórmula de Taylor de segunda ordem:

$$f(\mathbf{x}) - f(\mathbf{a}) = df(\mathbf{a}) + \frac{1}{2}d^2 f(\mathbf{a}) + o\left(\|\mathbf{x}-\mathbf{a}\|^2\right). \tag{4.13}$$

Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ uma função de classe $\mathcal{C}^2$ numa vizinhança dum ponto $\mathbf{a} \in \operatorname{int} D$. Se $\mathbf{a}$ é um ponto de estacionaridade de f, logo $df(\mathbf{a}) = 0$ e a igualdade (4.13) pode ser escrita na forma

$$f(\mathbf{x}) - f(\mathbf{a}) = \frac{1}{2} d^2 f(\mathbf{a}) + o\left(\|\mathbf{x} - \mathbf{a}\|^2\right), \tag{4.14}$$

para todo $\mathbf{x}$ próximo de $\mathbf{a}$, ou seja $\|\mathbf{x} - \mathbf{a}\| \leq \varepsilon$. O segundo diferencial $d^2 f(\mathbf{a})$ é a forma quadrática

$$d\mathbf{x}^T \mathcal{H} f(\mathbf{a})\, d\mathbf{x}$$

em relação a dx_i, $i = 1, ..., n$, onde $d\mathbf{x} = \mathbf{x} - \mathbf{a} = (dx_1, ..., dx_n)$ e $\mathcal{H} f(\mathbf{a})$ é a *matriz Hessiana* (simétrica). Por outro lado, podemos escrevemos (4.14) na forma

$$f(\mathbf{x}) - f(\mathbf{a}) = \|d\mathbf{x}\|^2 \left[\frac{1}{2} \left(\frac{d\mathbf{x}}{\|d\mathbf{x}\|} \right)^T \mathcal{H} f(\mathbf{a}) \frac{d\mathbf{x}}{\|d\mathbf{x}\|} + \frac{o\left(\|\mathbf{x} - \mathbf{a}\|^2\right)}{\|\mathbf{x} - \mathbf{a}\|^2} \right]. \tag{4.15}$$

Uma vez que o vector $\frac{d\mathbf{x}}{\|d\mathbf{x}\|}$ tem norma 1 (logo não pode ser nulo) o sinal da diferença $f(\mathbf{x}) - f(\mathbf{a})$ depende da positividade ou da negatividade da forma quadrática. Nomeadamente, temos o seguinte resultado

Teorema 4.29 (Condição suficiente de extremo local) *Seja $f : D \subset \mathbb{R}^n \to \mathbb{R}$ uma função de classe $\mathcal{C}^2$ numa vizinhança dum ponto $\mathbf{a} \in \operatorname{int} D$. Suponhamos que $\mathbf{a}$ é um ponto de estacionaridade de f. Então*

(a) se a forma quadrática $d^2 f(\mathbf{a})$ (ou, equivalentemente, a matriz Hessiana $\mathcal{H} f(\mathbf{a})$) é definida positiva então $\mathbf{a}$ é um minimizante local de f;

(b) se a forma quadrática $d^2 f(\mathbf{a})$ (ou a matriz Hessiana $\mathcal{H} f(\mathbf{a})$) é definida negativa então $\mathbf{a}$ é um maximizante local de f.

Demonstração: Vamos demonstrar apenas a alínea (a), pois a alínea (b) demonstra-se de forma analoga. Consideremos a função contínua

$$\xi \mapsto \Phi(\xi) = \frac{1}{2} \xi^T \mathcal{H} f(\mathbf{a})\, \xi$$

definida na esfera unitária $S = \{\xi \in \mathbb{R}^n : \|\xi\| = 1\}$. Como o conjunto S é compacto em $\mathbb{R}^n$ pelo Teorema de Weierstrasse existe ξ^*, $\|\xi^*\| = 1$, tal que

$$\inf_{\xi \in S} \Phi(\xi) = \Phi(\xi^*) > 0$$

(aplicámos aqui a hipótese da alínea (a) isto é o facto de que a forma quadrática $d^2 f(\mathbf{a})$ é definida positiva). Agora, fixamos $\varepsilon = \frac{1}{2}\Phi(\xi^*) > 0$ e encontramos $\delta > 0$ tal que $\|\mathbf{x} - \mathbf{a}\| \leq \delta$ implica

$$\frac{\left| o\left(\|\mathbf{x} - \mathbf{a}\|^2\right) \right|}{\|\mathbf{x} - \mathbf{a}\|^2} \leq \varepsilon.$$

Então, de (4.15) sai, para todos $\mathbf{x} \in D$, $\|\mathbf{x}-\mathbf{a}\| \leq \delta$, que

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{a}) &\geq \|d\mathbf{x}\|^2 \left(\Phi(\xi^*) - \frac{\left| o\left(\|\mathbf{x}-\mathbf{a}\|^2\right)\right|}{\|\mathbf{x}-\mathbf{a}\|^2} \right) \\ &\geq \|d\mathbf{x}\|^2 \left(\Phi(\xi^*) - \frac{1}{2}\Phi(\xi^*) \right) = \frac{1}{2}\|d\mathbf{x}\|^2 \Phi(\xi^*) > 0. \end{aligned}$$

■

Se a forma quadrática $d^2 f(\mathbf{a})$ é indefinida (isto é, se admite valores positivos e valores próprios negativos) então $\mathbf{a}$ não é um ponto de extremo local. No caso de $\mathbb{R}^2$ esta situação corresponde a um *ponto sela.*

Quando $d^2 f(\mathbf{a})$ é *semi-definida positiva* ou *semi-definida negativa* o Teorema 4.29 não dá qualquer informação sobre a natureza do ponto de estacionaridade. Assim, perante este tipo de situações é necessário examinar caso a caso e aplicar outros métodos. Às vezes, nestes casos, é útil fixar uma parte das coordenadas do ponto $\mathbf{a}$ e fazer variar as outras, e estudar esta nova função que está definida num espaço de dimensão menor.

Exemplo 4.30 *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por*

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

Como $\nabla f = \left(4x^3, 4y - 4y^3\right)$ a função f tem três pontos de estacionaridade, nomeadamente, $(0,0)$ e $(0,\pm 1)$. A matriz Hessiana de f é

$$\mathcal{H}f(x,y) = \begin{bmatrix} 12x^2 & 0 \\ 0 & 4-12y^2 \end{bmatrix}.$$

Logo nos pontos de estacionaridade temos

$$\mathcal{H}f(0,0) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}; \quad \mathcal{H}f(0,\pm 1) = \begin{bmatrix} 0 & 0 \\ 0 & -8 \end{bmatrix},$$

e as respectivas formas quadráticas são

$$d^2 f(0,0) = 4dy^2 \quad e \quad d^2 f(0,\pm 1) = -8dy^2$$

(a primeira é semi-definida positiva; a segunda é semi-definida negativa) não podemos tirar conclusão em nenhum dos casos. No entanto, a natureza do ponto $(0,0)$ pode ser determinada escrevendo

$$f(x,y) = x^4 + y^2\left(2 - y^2\right)$$

e observando que $2-y^2 > 0$ para todos y próximos a zero. Assim, a última igualdade implica que $f(x,y) \geq 0 = f(0,0)$ nalguma vizinhança de $(0,0)$, e portanto o ponto $(0,0)$ é um minimizante local de f.

Para estudar a natureza dos pontos $(0,\pm 1)$ consideramos a função de uma variável $\varphi(y) = f(0,y) = 2y^2 - y^4$. A segunda derivada $\varphi''(y) = 4 - 12y^2$ nos pontos de estacionaridade ± 1 admite valor negativo $\varphi''(\pm 1) = -8$, logo ± 1 são pontos de máximo local de φ. Por outro lado, a função de uma variável $\psi(x) = f(x,\pm 1) = x^4 + 1$ tem um mínimo local em 0. Assim, $(0,\pm 1)$ são pontos sela de f (não são minimizantes nem maximizantes).

Para verificar se a fórma quadrática $d^2 f(\mathbf{a})$ é definida positiva ou negativa podemos encontrar os valores próprios da matriz Hessiana, ou utilizar o critério de Silvestre (ver Secção 1.1).

Exemplo 4.31 *Consideremos a função $f : \mathbb{R}^3 \to \mathbb{R}$ definida por*

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

Temos

$$\nabla f(x, y, z) = \left(3x^2 - 3, 2y - 2, 4z\right) \qquad e \qquad \mathcal{H}f(x, y, z) = \begin{bmatrix} 6x & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

O gradiente é nulo se $x = \pm 1$, $y = 1$ e $z = 0$. Então os pontos de estacionaridade são $(1, 1, 0)$ e $(-1, 1, 0)$. Temos

$$\mathcal{H}f(1, 1, 0) = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Como a matriz é diagonal os valores próprios são os elementos da diaganal principal: $6, 2, 4$. Portanto $f(1, 1, 0)$ é um mínimo local da função. No ponto $(-1, 1, 0)$, a matriz Hessiana é

$$\mathcal{H}f(-1, 1, 0) = \begin{bmatrix} -6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Os valores próprios são $-6, 2, 4$ portanto $(-1, 1, 0)$ não é um ponto de extremo local (ver Proposição 1.32).

Se escrevermos a forma quadrática do segundo diferencial nos pontos de estacionaridade em que a matriz Hessiana é indefinida podemos estudar a sua natureza de tais pontos em mais pormenor. Nomeadamente, podemos ver em relação a que variável (ou em que direcção) é que a função cresce, e em relação a qual decresce, não tendo, portanto, um extremo local nesses pontos. Assim, no exemplo anterior e considerando o ponto de estacionaridade $(-1, 1, 0)$ temos

$$d^2 f(-1, 1, 0) = -6dx^2 + 2dy^2 + 4dz^2 = \left(\sqrt{2}dy - \sqrt{6}dx\right)\left(\sqrt{2}dy + \sqrt{6}dx\right) + 4dz^2.$$

Daqui vê-se que para $z = 0$ a função de duas variáveis $g(x, y) = f(x, y, 0)$ tem sela no ponto $(-1, 1)$: $g(-1 + dx, 1 + dy) \geq g(-1, 1)$ ou $\leq g(-1, 1)$ dependendo se os acréscimos dx e dy satisfazem a desigualdade

$$\left(\sqrt{2}dy - \sqrt{6}dx\right)\left(\sqrt{2}dy + \sqrt{6}dx\right) > 0$$

ou a desigualdade

$$\left(\sqrt{2}dy - \sqrt{6}dx\right)\left(\sqrt{2}dy + \sqrt{6}dx\right) < 0.$$

No caso de $\mathbb{R}^2$ as condições suficientes de mínimo e máximo local podem ser escritas de uma forma muito simples. Pois, seja $f : D \subset \mathbb{R}^2 \to \mathbb{R}$ um função com derivadas de segunda ordem contínuas em torno de um ponto $\mathbf{a} \in \operatorname{int} D$. Supomos que $\mathbf{a}$ é um ponto de estacionaridade de f, isto é,

$$\frac{\partial f}{\partial x}(\mathbf{a}) = \frac{\partial f}{\partial y}(\mathbf{a}) = 0.$$

A matriz Hessiana tem a forma

$$\mathcal{H}f(\mathbf{a}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x \partial y}(\mathbf{a}) & \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) \end{bmatrix}.$$

Pelo critério de Silvestre $\mathcal{H}f(\mathbf{a})$ é definida positiva (negativa) sse $\frac{\partial^2 f}{\partial x^2}(\mathbf{a}) > 0$ (resp. < 0) e

$$\det(\mathcal{H}f(\mathbf{a})) = \frac{\partial^2 f}{\partial x^2}(\mathbf{a}) \frac{\partial^2 f}{\partial y^2}(\mathbf{a}) - \left(\frac{\partial^2 f}{\partial x \partial y}(\mathbf{a})\right)^2 > 0.$$

Exemplo 4.32 *Consideremos a função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por*

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

Temos

$$\nabla f(x,y) = \left(2x - 3 + y^2 - 2y, -2y + 2 + 2xy - 2x\right) \quad e \quad \mathcal{H}f(x,y) = \begin{bmatrix} 2 & 2y-2 \\ 2y-2 & -2+2x \end{bmatrix}.$$

O gradiente é nulo sse

$$\begin{cases} 2x - 3 + y^2 - 2y = 0 \\ -2y + 2 + 2xy - 2x = 0 \end{cases} \Leftrightarrow \begin{cases} 2(x-2) + (y-1)^2 = 0 \\ 2(y-1)(x-1) = 0 \end{cases} \Leftrightarrow \begin{cases} x = 2 \\ y = 1 \end{cases} \vee \begin{cases} y = 1 \pm \sqrt{2} \\ x = 1 \end{cases}.$$

Então os pontos de estacionaridade são $(2,1)$, $\left(1, 1-\sqrt{2}\right)$ e $\left(1, 1+\sqrt{2}\right)$. Como

$$\mathcal{H}f(2,1) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix},$$

então

$$\det \mathcal{H}f(2,1) = 4 > 0 \quad e \quad \frac{\partial^2 f}{\partial x^2}(2,1) = 2 > 0,$$

portanto $(2,1)$ é um minimizante local de f .

Nos pontos $\left(1, 1 \pm \sqrt{2}\right)$ temos

$$\mathcal{H}f\left(1, 1 \pm \sqrt{2}\right) = \begin{bmatrix} 2 & \pm 2\sqrt{2} \\ \pm 2\sqrt{2} & 0 \end{bmatrix}.$$

Como

$$\det \mathcal{H}f\left(1, 1 \pm \sqrt{2}\right) = -8 < 0$$

então $\left(1, 1 \pm \sqrt{2}\right)$ *não são pontos de extremo local. Escrevemos o segundo diferencial*

$$d^2 f\left(1, 1 \pm \sqrt{2}\right) = 2dx^2 \pm 4\sqrt{2}dxdy = 2dx\left(dx \pm 2\sqrt{2}dy\right)$$

e observamos que a função f *satisfaz, por exemplo, a desigualdade*

$$f\left(1 + dx, 1 - \sqrt{2} + dy\right) \geq f\left(1, 1 - \sqrt{2}\right)$$

para todos os acréscimos (suficientemente pequenos) dx *e* dy *tais que* $dx > 0$, $2\sqrt{2}dy < dx$ *ou* $dx < 0$, $2\sqrt{2}dy < dx$, *enquanto que nos casos* $dx > 0$, $2\sqrt{2}dy > dx$ *e* $dx < 0$, $2\sqrt{2}dy > dx$ *verifica-se a desigualdade contrária:*

$$f\left(1 + dx, 1 - \sqrt{2} + dy\right) \leq f\left(1, 1 - \sqrt{2}\right).$$

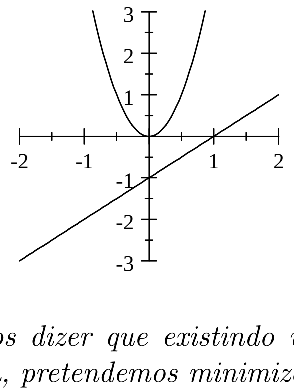
Estes raciocínios justificam que o ponto $\left(1, 1 - \sqrt{2}\right)$ *(analogamente* $\left(1, 1 + \sqrt{2}\right)$*) é um ponto sela.*

4.4 Extremos condicionados. Multiplicadores de Lagrange

Nesta secção vamos estudar os extremos de uma função sujeita a determinadas condições. Um problema deste tipo é, por exemplo, o de determinar um ponto numa curva dada cuja distância à origem seja mínima. Este problema toma a forma: determinar os extremos da função $f(\mathbf{x})$ sujeita à condição $g(\mathbf{x}) = 0$. Em vez de determinar os extremos de f com $\mathbf{x}$ a variar livremente como na secção anterior, aqui a variável $\mathbf{x}$ tem que verificar a condição $g(\mathbf{x}) = 0$. Por isso estes problemas se dizem de *extremos condicionados* (ou *extremos relativos*) e os outros dizem-se problemas de *extremos livres.*

Exemplo 4.33 *Determinemos o ponto da parábola* $y = 4x^2$ *que está mais próximo da recta* $x - y = 1$.

Pretendemos determinar o mínimo da função distância entre dois pontos que devem verificar as condições seguintes: um dos pontos deve pertencer à parábola (e é o ponto pedido) e o outro deve pertencer à recta



A partir da figura anterior podemos dizer que existindo um extremo ele é um mínimo, não existindo qualquer máximo. Ou seja, pretendemos minimizar a função distância

$$\mathrm{d}\left(\left(x, y\right), \left(u, v\right)\right) = \sqrt{\left(x - u\right)^2 + \left(y - v\right)^2}, \tag{4.16}$$

com $(x,y) \in \mathbb{R}^2$ tal que $y = 4x^2$ e $(u,v) \in \mathbb{R}^2$ tal que $u - v = 1$. Escrevendo, por exemplo, y e u em função de x e v, respectivamente, reduzimos o nosso problema à minimização da função

$$f(x,v) = (x - 1 - v)^2 + \left(4x^2 - v\right)^2.$$

Podemos calcular os extremos (locais) livres da função f como fizemos na secção anterior. Comecemos por determinar os pontos de estacionaridade, resolvendo a equação $\nabla f(x,v) = \mathbf{0}$ ou o sistema

$$\begin{cases} \frac{\partial f}{\partial x} = 2(x - 1 - v) + 16x\left(4x^2 - v\right) = 0 \\ \frac{\partial f}{\partial v} = -2(x - 1 - v) - 2\left(4x^2 - v\right) = 0. \end{cases}$$

A única solução deste sistema é o par $(x,v) = \left(\frac{1}{8}, -\frac{13}{32}\right)$, e pelo que dissemos acima trata-se de um minimizante da função em (4.16). Assim, o ponto pedido é $(x,y) = \left(\frac{1}{8}, \frac{1}{16}\right)$ e a distância mínima da parábola à recta é

$$\mathrm{d}\left(\left(\frac{1}{8}, \frac{1}{16}\right), \left(\frac{19}{32}, -\frac{13}{32}\right)\right) = \sqrt{\left(\frac{1}{8} - \frac{19}{32}\right)^2 + \left(\frac{1}{16} + \frac{13}{32}\right)^2} = \frac{15}{32}\sqrt{2}.$$

A necessidade de estudar problemas de *optimização* (isto é, de minimização ou de maximização) com *restrições* aparece também na minimização ou maximização *global.* Lembramos que, para encontrar pontos onde uma função, de uma variável, $f : [a,b] \to \mathbb{R}$ atinge o seu valor extremo determinamos, primeiro, todos os pontos de extremo local (pertencentes ao intervalo aberto $]a,b[$) e, depois, comparamos os valores de f nestes pontos com os valores de f nas extremidades a e b, sendo estes dois últimos pontos de fronteira. Se f é uma função de duas variáveis, definida nalgum domínio $D \subset \mathbb{R}^2$ (limitado ou ilimitado), então o problema de optimização global torna-se mais complexo. Pois, analogamente ao caso de uma função de variável real, podemos encontrar, primeiro, os maximizantes e os minimizantes locais pertencentes ao interior $\operatorname{int} D$, e depois estudamos o comportamento de f na fronteira de D, que, usualmente, é definida através de uma equação do tipo $g(x,y) = 0$, onde g é uma função contínua. Assim chegamos ao problema de extremo condicionado. O mesmo raciocinio pode ser aplicado no caso de funções de mais variáveis $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Neste caso podemos ter mais do que uma restrição, digamos $g_i(\mathbf{x}) = 0$, $\mathbf{x} \in D$, $i = 1, ..., m$. Temos de ter em conta que o número de restrições efectivas m tem de ser estritamente menor do que o número de variáveis n. Caso contrário, o conjunto

$$C = \{\mathbf{x} \in D : g_i(\mathbf{x}) = 0, i = 1, ..., m\} \tag{4.17}$$

pode ser vazio, ou reduzir-se a um número finito de pontos. Por "*restrições efectivas*" entendemos aquelas que não podem ser obtidas a partir das outras. Por exemplo, se $m = 3$ e $g_3(\mathbf{x}) = \mu_1 g_1(\mathbf{x}) + \mu_2 g_2(\mathbf{x})$ para alguns números $\mu_1, \mu_2 \in \mathbb{R}$, então a terceira restrição $g_3(\mathbf{x}) = 0$ não é considerada efectiva e pode ser eliminada. Chegamos, assim, a um caso em que $m = 2$.

O método de substituição usado no Exemplo 4.33 nem sempre pode ser aplicado, pois, em geral, é difícil obter, a partir das condições dadas, algumas das variáveis em função das restantes, mesmo que o Teorema da função implicíta garanta a existência dessas funções.

Vamos agora apresentar o *método dos multiplicadores de Lagrange* que permite obter os pontos onde uma função f, sujeita a determinadas condições $g_1(\mathbf{x}) = 0, ..., g_m(\mathbf{x}) = 0$, pode ter extremo (local). Este método baseia-se no seguinte resultado importante.

Teorema 4.34 *Sejam D um subconjunto aberto de $\mathbb{R}^n$ e $f, g_1, ..., g_m : D \to \mathbb{R}$ funções de classe $\mathcal{C}^1$ em D, com $m < n$. Seja $\mathbf{a} \in C$ (ver (4.17)) um ponto de extremo de f sujeito às restrições $g_i(\mathbf{x}) = 0$, $i = 1, ..., m$. Então, os vectores $\nabla f(\mathbf{a})$, $\nabla g_1(\mathbf{a})$, ..., $\nabla g_m(\mathbf{a})$ são linearmente dependentes, isto é, existem escalares $\lambda_0, \lambda_1, ..., \lambda_m \in \mathbb{R}$ não simultaneamente nulos, tais que*

$$\lambda_0 \nabla f(\mathbf{a}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{a}) = 0. \tag{4.18}$$

Demonstração:

Supondo que os $m+1$ vectores $\nabla f(\mathbf{a})$, $\nabla g_1(\mathbf{a})$, ..., $\nabla g_m(\mathbf{a})$ são linearmente independentes temos que chegar à conclusão de que $\mathbf{a}$ não pode ser minimizante nem maximizante local da função f.

Consideramos a aplicação diferenciável $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^{m+1}$ dada por

$$\mathbf{F}(\mathbf{x}) = (f(\mathbf{x}), g_1(\mathbf{x}), ..., g_m(\mathbf{x})),$$

e a sua matriz de Jacobi Jac $(\mathbf{F})$ de ordem $(m+1) \times n$ (note que $m+1 \leq n$). Como todas as linhas desta matriz são linearmente independentes, a característica de Jac $(\mathbf{F})$ é igual a $m+1$. Isto implica que um menor de Jac $(\mathbf{F})$ de ordem $(m+1) \times (m+1)$ é diferente de zero. Fazendo, se necessário, renumeração das variáveis podemos supôr que este menor consiste nas primeiras $m+1$ colunas, isto é

$$\frac{\partial(f, g_1, ..., g_m)}{\partial(x_1, x_2 ..., x_{m+1})}(\mathbf{a}) \neq 0. \tag{4.19}$$

Igualando as últimas $n-m-1$ coordenadas de $\mathbf{x}$ (se existirem) às respectivas coordenadas do vector $\mathbf{a}$, definimos a aplicação $\mathbf{G} : \mathbb{R} \times \mathbb{R}^{m+1} \to \mathbb{R}^{m+1}$ por

$$\begin{aligned} & \mathbf{G}(y; x_1, ..., x_{m+1}) \\ = \ & \mathbf{F}(x_1, ..., x_{m+1}, a_{m+2}, ..., a_n) - (y, 0, ..., 0). \end{aligned}$$

Sabendo que $\mathbf{G}(f(\mathbf{a}); a_1, a_2, ..., a_{m+1}) = \mathbf{0}$ e que o *Jacobiano* de $\mathbf{G}$ em relação às últimas $m+1$ variáveis é diferente de zero (ver (4.19)) pelo Teorema 3.8 de aplicação implicita existe $\varepsilon > 0$ e uma vizinhança U de $\mathbf{a}^0 = (a_1, a_2, ..., a_{m+1})$ tais que para todo $\mathbf{y} \in \mathbb{R}$, $|y - f(\mathbf{a})| \leq \varepsilon$, encontra-se um único $\hat{\mathbf{x}} = (x_1, x_2, ..., x_{m+1}) \in U$ com

$$\mathbf{G}(y; x_1, ..., x_{m+1}) = 0.$$

Em particular, para $y = f(\mathbf{a}) \pm \varepsilon$ existem pontos $\hat{\mathbf{x}}^{\pm} = \left(x_1^{\pm}, x_2^{\pm}, ..., x_{m+1}^{\pm}\right) \in U$ tais que

$$\mathbf{F}\left(x_1^{\pm}, x_2^{\pm}, ..., x_{m+1}^{\pm}, a_{m+2}, ..., a_n\right) = (f(\mathbf{a}) \pm \varepsilon, 0, ..., 0).$$

Daqui sai que ambos os pontos $\mathbf{x}^{\pm} = \left(x_1^{\pm}, x_2^{\pm}, ..., x_{m+1}^{\pm}, a_{m+2}, ..., a_n\right)$ pertencem a C (isto é, verificam as restrições $g_i(\mathbf{x}^{\pm}) = 0$, $i = 1, ..., m$) e

$$f\left(\mathbf{x}^{-}\right) = f(\mathbf{a}) - \varepsilon < f(\mathbf{a}) < f(\mathbf{a}) + \varepsilon = f\left(\mathbf{x}^{+}\right).$$

Logo, no ponto $\mathbf{a}$ a função f não tem máximo nem mínimo. ■

Observação 4.35 *Considerando extremos condicionados limitaremo-nos apenas ao caso em que os gradientes $\nabla g_i(\mathbf{a})$, $i = 1, .., m$, são linearmente independentes. Então na igualdade (4.17) vamos ter $\lambda_0 \neq 0$ (caso contrário, $\sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{a}) = 0$ donde $\lambda_1 = ... = \lambda_m = 0$, o que é impossível pelo Teorema 4.34). Assim, sem perda de generalidade, pomos $\lambda_0 = 1$ e (4.17) transforma-se em*

$$\nabla f(\mathbf{a}) + \sum_{i=1}^{m} \lambda_i \nabla g_i(\mathbf{a}) = 0.$$

O caso em que os gradientes $\nabla g_i(\mathbf{x})$, $i = 1, .., m$, são linearmente dependentes no ponto $\mathbf{a}$ (por exemplo, em $\mathbb{R}^3$ isto acontece quando as superfícies $g_1(\mathbf{x}) = 0$ e $g_2(\mathbf{x}) = 0$ têm, no ponto $\mathbf{a}$, um plano tangente comum) tem que ser considerado à parte. A independência linear pode ser verificada utilizando a matriz Jacobiana da aplicação $g : D \to \mathbb{R}^m$, $g(\mathbf{x}) = (g_1(\mathbf{x}), ..., g_m(\mathbf{x}))$. Nomeadamente, a matriz $\operatorname{Jac}(g)(\mathbf{a})$ tem que ter característica m. Por outras palavras, tem que existir um menor de ordem $m \times m$ (digamos $\frac{\partial(g_1,...,g_m)}{\partial\left(x_{i_1},...,x_{i_m}\right)}(\mathbf{a})$, com $i_1, ..., i_m$ tomando valores entre 1 e n) diferente de zero.

Tendo em conta o Teorema 4.34 e a Observação 4.35 vamos descrever o *método dos multiplicadores de Lagrange*. Começamos por construir a chamada *função de Lagrange* que depende de $n + m$ variáveis:

$$\begin{aligned} \mathcal{L}(\mathbf{x}; \boldsymbol{\lambda}) &= f(\mathbf{x}) + \langle \boldsymbol{\lambda}, g(\mathbf{x}) \rangle \\ &= f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}), \end{aligned}$$

$\mathbf{x} = (x_1, ..., x_n)$ e $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_m)$. Os escalares $\lambda_1, ..., \lambda_m$ chamam-se *multiplicadores de Lagrange*.

Igualamos a zero o gradiente de $\mathcal{L}(\mathbf{x}; \boldsymbol{\lambda})$ (em ordem a todas as variáveis) e obtemos um sistema de $n + m$ equações (com $n + m$ variáveis):

$$\begin{cases} \frac{\partial f}{\partial x_1}(\mathbf{x}) + \sum_{i=1}^m \lambda_i \frac{\partial g_i}{\partial x_1}(\mathbf{x}) = 0 \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) + \sum_{i=1}^m \lambda_i \frac{g_i}{\partial x_n}(\mathbf{x}) = 0 \\ g_1(\mathbf{x}) = 0 \\ \vdots \\ g_m(\mathbf{x}) = 0. \end{cases}$$

Resolvendo este sistema encontramos um ponto (ou mais) $(\mathbf{a}, \boldsymbol{\lambda})$ (ponto de estacionaridade da função de Lagrange). Um ponto de extremo de f em C é um vector cujas coordenadas são as primeiras n coordenadas do ponto $(\mathbf{a}, \boldsymbol{\lambda})$.

Para averiguar se este ponto é, de facto, mínimizante ou maximizante temos de aplicar as condições de segunda ordem que são parecidas às respectivas condições no caso de extremo local (ver acima) mas com certa particularidade. Nomeadamente, seja $(\mathbf{a}, \boldsymbol{\lambda})$ um ponto de estacionaridade da função de Lagrange $\mathcal{L}$. Fixando λ, consideramos $\mathcal{L}$ como função só de $\mathbf{x}$ (de primeiras n variáveis) e encontramos a forma quadrática do segundo diferencial no ponto $\mathbf{a}$:

$$d^2\mathcal{L}(a;\lambda) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a})\, dx_i dx_j. \tag{4.20}$$

Para $\mathbf{a}$ ser um ponto de mínimo (máximo) local de f no conjunto C esta forma tem que admitir valores positivos (resp., negativos) não para todos os acréscimos $d\mathbf{x} \neq \mathbf{0}$ (como no caso dos extremos livres) mas só para aqueles que estão ligados às condições lineares:

$$\sum_{j=1}^{n} \frac{\partial g_i}{\partial x_j}(\mathbf{a})\, dx_j = 0, \quad i = 1, ..., m. \tag{4.21}$$

Para averiguar isto basta exprimir a partir do sistema (4.21) alguns dos acréscimos, digamos dx_j, $j = m+1, ..., n$, através dos restantes $dx_1, ..., dx_m$ (isto é possível porque os vectores $\nabla g_i(\mathbf{a})$, $i = 1, ..., m$, são linearmente independentes, ver Observação 4.35) e substituir em (4.20). Assim, obtemos uma forma quadrática diferencial no espaço $\mathbb{R}^m$ que pode ser estudada usando os valores próprios ou o critério de Silvestre.

Exemplo 4.36 $(n = 2, m = 1)$ *Consideremos a função escalar* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida por*

$$f(x,y) = ax + by,$$

com $a, b \in \mathbb{R}$ *tais que* $a^2 + b^2 \neq 0$. *Queremos saber quais são os pontos de estacionaridade da restrição de* f *ao círculo unitário* $S^1 = \{(x,y) \in \mathbb{R}^2 : \|(x,y)\| = 1\}$. *Ou seja queremos determinar os pontos de estacionaridade de* f *sujeitos à restrição* $g(x,y) = x^2 + y^2 - 1$. *A função de Lagrange é dada por*

$$\mathcal{L}(x,y;\lambda) = f(x,y) + \lambda g(x,y)$$

Calculando o gradiente $\nabla\mathcal{L}$ *e igualando a zero chegamos ao sistema*

$$\begin{cases} a + 2\lambda x = 0 \\ b + 2\lambda y = 0 \\ x^2 + y^2 - 1 = 0. \end{cases}$$

É fácil verificar que as suas soluções são

$$\left(-\frac{a}{\sqrt{a^2+b^2}}, -\frac{b}{\sqrt{a^2+b^2}}; \frac{\sqrt{a^2+b^2}}{2}\right) \quad e \quad \left(\frac{a}{\sqrt{a^2+b^2}}, \frac{b}{\sqrt{a^2+b^2}}; -\frac{\sqrt{a^2+b^2}}{2}\right). \tag{4.22}$$

Assim, os minimizantes e/ou os maximizantes de f *sujeitos à restrição* $g = 0$ *só podem ser os pontos*

$$(x_1,y_1) := \left(-\frac{a}{\sqrt{a^2+b^2}}, -\frac{b}{\sqrt{a^2+b^2}}\right) \quad e \quad (x_2,y_2) := \left(\frac{a}{\sqrt{a^2+b^2}}, \frac{b}{\sqrt{a^2+b^2}}\right).$$

Pois para eles existem, respectivamente, $\lambda_1 = \frac{\sqrt{a^2+b^2}}{2}$ *e* $\lambda_2 - \frac{\sqrt{a^2+b^2}}{2}$ *tais que os pontos em (4.22) são pontos de estacionaridade de* $\mathcal{L}$.

Averiguemos a natureza do ponto (x_1, y_1). *O segundo diferencial da função* $\mathcal{L}(x, y; \lambda_1)$ *no ponto* (x_1, y_1) *é dado por*

$$d^2\mathcal{L}(x_1, y_1; \lambda_1) = 2\sqrt{a^2+b^2}\left(dx^2 + dy^2\right). \tag{4.23}$$

De $dg(x_1, y_1) = 0$ *sai que*

$$2\left(-\frac{a}{\sqrt{a^2+b^2}}\right)(dx + dy) = 0,$$

ou seja, $dx = -dy$. *Substituindo* dx *por* $-dy$ *em(4.23) obtemos o diferencial* $d^2\mathcal{L}(x_1, y_1; \lambda_1)$ *sujeito à restrição* $dg(x_1, y_1) = 0$, *que é dado por*

$$d^2\mathcal{L}(x_1, y_1; \lambda_1) = 2\sqrt{a^2+b^2}\left(dx^2 + dy^2\right) = 4\sqrt{a^2+b^2}dy^2 > 0, \ \ \forall dy \neq 0.$$

Donde se conclui que a forma quadrática $d^2\mathcal{L}(x_1, y_1; \lambda_1)$, *sujeita à restrição* $dg = 0$ *é definida positiva, e por conseguinte, o ponto* (x_1, y_1) *é um minimizante local de* f *sujeito à restrição* $g = 0$. *Analogamente se mostra que o ponto* (x_2, y_2) *é um maximizante local de* f *sujeito à restrição* $g = 0$.

Como $\nabla g(x_1, y_1) \neq 0$, *o que fizemos está de acordo com a Observação 4.35.*

Observação 4.37 *No exemplo anterior a forma do segundo diferencial* $d^2\mathcal{L}(x_1, y_1; \lambda_1)$ *é, de facto, definida positiva (sem recorrer à restrição), pois* $d^2\mathcal{L}(x_0, y_0; \lambda_1) > 0 \ \forall (dx, dy) \neq (0, 0)$.

Exemplo 4.38 ($n = 3, m = 2$) *Determinemos o ponto na recta de intersecção dos planos de equações* $x - y = 1$ *e* $4x + z = 1$ *que esteja mais próximo de* $(0, 0, 0)$.

Os pontos que estão na recta de intersecção são da forma $(x, 2 - x - z, 1 - 4x)$. *Pretendemos minimizar a função distância entre um ponto* (x, y, z) *na recta de intersecção e o ponto* $(0, 0, 0)$:

$$\mathrm{d}\left((x, y, z), (0, 0, 0)\right) = \sqrt{x^2 + (x-1)^2 + (1-4x)^2},$$

que é equivalente a minimizar a função

$$f(x) = x^2 + (x-1)^2 + (1-4x)^2.$$

Como f *só depende de* x *o problema pode ser resolvido utilizando cálculos de análise real. Mas aqui vamos resolver aplicando o método dos multiplicadores de Lagrange. Pretendemos, então, minimizar a função*

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

sujeita às condições $x - y = 1$ *e* $4x + z = 1$. *Sejam*

$$g_1(x, y, z) = x - y - 1 \quad e \quad g_2(x, y, z) = 4x + z - 1.$$

Temos

$$\mathcal{L}(x, y; \lambda) = f(x, y, z) + \lambda_1 g_1(x, y, z) + \lambda_2 g_2(x, y).$$

Como

$$\nabla f(x, y, z) = (2x, 2y, 2z), \quad \nabla g_1(x, y, z) = (1, -1, 0) \quad e \quad \nabla g_2(x, y, z) = (4, 0, 1).$$

Então $\nabla\mathcal{L}(x,y;\lambda)=\mathbf{0}$ *é equivalente a*

$$\begin{cases} 2x+\lambda_1+4\lambda_2=0\\ 2y-\lambda_1=0\\ 2z+\lambda_2=0\\ x-y-1=0\\ 4x+z-1=0, \end{cases}$$

e a única solução deste sistema é $\left(\frac{5}{18},-\frac{13}{18},-\frac{1}{9};-\frac{13}{9},\frac{2}{9}\right)$. *Como*

$$d^2\mathcal{L}\left(x,y,z;-\frac{13}{9},\frac{2}{9}\right)=2dx^2+2dy^2+2dz^2>0\quad\forall\,(dx,dy,dz)\neq(0,0,0)\,,$$

então (tal como na Observação 4.37) a forma quadrática $d^2\mathcal{L}\left(\frac{5}{18},-\frac{13}{18},-\frac{1}{9};-\frac{13}{9},\frac{2}{9}\right)$ *é definida positiva, e portanto o ponto* $\left(\frac{5}{18},-\frac{13}{18},-\frac{1}{9}\right)$ *é um minimizante local de* f *sujeito às duas restrições dadas.*

Notemos que para $\mathbf{a}=\left(\frac{5}{18},-\frac{13}{18},-\frac{1}{9}\right)$ *temos*

$$\frac{\partial\,(g_1,g_2)}{\partial\,(y,z)}(\mathbf{a})=\det\begin{bmatrix}\frac{\partial g_1}{\partial y}(\mathbf{a}) & \frac{\partial g_1}{\partial z}(\mathbf{a})\\ \frac{\partial g_2}{\partial y}(\mathbf{a}) & \frac{\partial g_2}{\partial z}(\mathbf{a})\end{bmatrix}=\det\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}=-1\neq 0,$$

e portanto o que fizemos está em conformidade com a Observação 4.35.