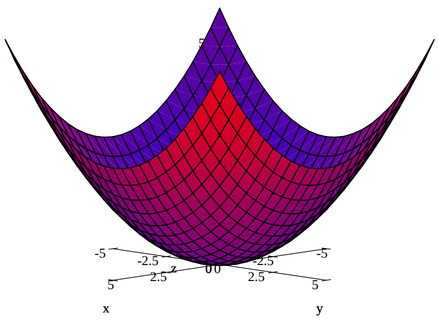
Lições de Cálculo para Economia

Maria do Rosário Grossinho



ISEG
2009

Conteúdo

Introdução

O presente texto foi escrito com base em notas relativas ao estudo de funções de várias variáveis, assunto tradicionalmente tratado na disciplina Matemática II, das licenciaturas em Economia, Finanças e Gestão do ISEG, cuja regência tem sido assegurada pela autora desde 1998, no caso de Economia, e recentemente nas três licenciaturas. Pressupõe que o leitor tem conhecimentos sólidos sobre funções de uma variável.

O conteúdo está concebido de forma clássica. Pretende-se inicialmente que o estudante desenvolva capacidade de visualização de gráficos de funções de $\mathbb{R}^2$ em $\mathbb{R}$. Superfícies de nível e cortes por planos paralelos aos planos da base são alguns dos elementos úteis para esse fim. Avança-se então para o estudo de limites e continuidade de funções, o que pressupõe um ponto prévio para aspectos de estrutura topológica. Em seguida, generaliza-se a várias variáveis o Cálculo Diferencial estudado em disciplina precedente. Este capítulo é pedra de toque do programa, em particular porque constitui uma premissa indispensável para o estudo de problemas de extremo e suas aplicações a problemas de optimização que são abordados no capítulo seguinte. Apresenta-se em apêndice um capítulo sobre alguns aspectos de Álgebra Linear de muita utilidade no estudo de extremos, como por exemplo as formas quadráticas.

A abordagem é rigorosa. Pretende-se ensinar ideias e conceitos, juntando aspectos teóricos à intuição e visualização geométrica. Sendo as definições e teoremas apresentados em dimensão n, de imediato são ilustrados por exemplos em $\mathbb{R}^2$. Ao longo do texto apresentam-se várias aplicações em Economia. Sublinha-se que não há qualquer pretensão de ensinar conceitos e resultados de carácter económico mas tão somente de pôr em evidência através de exemplos a aplicação dos conceitos matemáticos. Espera-se com isso ilustrar a utilidade do investimento feito pelo estudante na aprendizagem matemática e, com isso, favorecer a sua motivação.

Gostaria de agradecer aos vários colegas que ao longo dos anos fizeram parte da equipa docente de Matemática II e que contribuiram para o aperfeiçoamento do texto.

É com muito gosto que exprimo o meu particular reconhecimento à Prof. Lígia Amado e ao Prof. João Janela. Foi o seu incentivo e apoio que levaram a que, a partir de notas de curso dispersas, desse forma coerente a este texto, que constitui o primeiro passo de um projecto didáctico mais abrangente em que estamos envolvidos. Registo ainda a sua inestimável colaboração na correcção de gralhas bem como na introdução de figuras no texto.

ISEG, 2009
Maria do Rosário Grossinho

Capítulo 1

Funções de várias variáveis

Neste capítulo vamos ocupar-nos do estudo de funções reais de variável vectorial e funções vectoriais de variável vectorial. Permitindo traduzir analiticamente o comportamento de certos fenómenos do mundo real, a noção de função real de variável real é contudo ainda insuficiente para responder a situações mais complexas em que a necessidade de várias variáveis se faz sentir. Tal acontece, por exemplo, em situações de natureza económica. Se quisermos representar o lucro anual de uma empresa como dependente dos montantes gastos em investigação e em publicidade, ou do problema da produção, em que o produto final é função do capital e do trabalho, imediatamente reconhecemos a necessidade de considerarmos funções de várias variáveis.

1.1 Função de $\mathbb{R}^m$ em $\mathbb{R}$: noções preliminares e exemplos

De forma intuitiva, poderemos dizer que uma função $f : A \longrightarrow B$ é uma correspondência que associa a cada elemento x pertencente a um conjunto A, um e um só elemento $f(x)$ de um conjunto B; o conjunto A onde a correspondência está definida chama-se **domínio** de f e o subconjunto de B formado pelos valores de $f(x)$ chama-se **contradomínio** de f. Habitualmente representam-se por D_f e CD_f.

1.1.1 Função real de m variáveis reais. Domínio

No que se segue, embora apresentemos as definições na sua forma geral, o especial enfoque da matéria será dado aos casos $\mathbb{R}^2$ e $\mathbb{R}^3$.

Dado $m \in \mathbb{N}$, chama-se **função real de** m **variáveis reais**, ou **de variável vectorial** $\mathbf{x} = (x_1, x_2, ...x_m)$, a toda a função f cujo domínio D_f, isto é, o conjuntos dos elementos em que está definida, é um subconjunto de $\mathbb{R}^m$ e o contradomínio, isto é, o conjunto das imagens $f(\mathbf{x})$ é um subconjunto de $\mathbb{R}$, isto é

$$f : D_f \subset \mathbb{R}^m \longrightarrow \mathbb{R}$$

onde $\mathbb{R}^m = \mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$, é o produto cartesiano de m factores todos iguais a $\mathbb{R}$. Notaremos por $\mathbf{x}$ um elemento de $\mathbb{R}^m$, $m > 1$, e manteremos a notação x para elementos de $\mathbb{R}$.

Exemplo 1 *1.Consideremos a função definida em* $\mathbb{R}^2$ *por*

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

O domínio é $\mathbb{R}^2$ *e o contradomínio é* $\mathbb{R}_0^+$. *Temos, por exemplo:*

$f(-2, 4) = (-2)^2 + 4^2 = 20;$
$f(3, 5) = 3^2 + 5^2 = 34;$
$f(-1, -1) = (-1)^2 + (-1)^2 = 2.$

2. Seja f *definida em* $\mathbb{R}^2$ *por*

$$f(x, y) = \sqrt{x^2 + y^2 - 9.}$$

Neste caso o domínio já não é todo o $\mathbb{R}^2$. *Com efeito,*

$$D_f = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \geq 9\right\},$$

que, do ponto de vista gráfico, não é mais do que o subconjunto do plano, complementar do círculo fechado de centro $(0, 0)$ *e raio* 3.

3. Seja f *definida em* $\mathbb{R}^2$ *por*

$$f(x, y) = \frac{\ln\left(16 - x^2 - y^2\right)}{|x| - |y|}.$$

Neste caso o domínio é dado por

$$D_f = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 16 \ \wedge \ x \neq y \ \wedge \ x \neq -y\right\}.$$

Exercício: faça um esboço da representação gráfica de D_f.

4. Consideremos f *definida em* $\mathbb{R}^3$ *por*

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

O domínio de f é

$$\begin{aligned} D_f &= \left\{(x,y,z) \in \mathbb{R}^3 : e^{x^2+y^2+z^2} > e^9\right\} \\ &= \left\{(x,y,z) \in \mathbb{R}^3 : x^2+y^2+z^2 > 9\right\}. \end{aligned}$$

Trata-se do complementar da esfera de centro $(0,0,0)$ *e raio* 3.

5. *Um estudo sobre produção de leite mostrou que a produção P é função de quatros factores x, y, z e w de acordo com a expressão seguinte*

$$P = 2,5.x^{0,02}y^{0,5}z^{0,25}w^{0,03}.$$

Por exemplo, x, y, z e w poderão ser os parâmetros relativos ao trabalho envolvido, ao consumo de forragem, ao gasto de electricidade e ao gasto de água. O domínio de P enquanto função é

$$D_P = \left\{(x,y,z,w) \in \mathbb{R}^4 : y \geq 0, z \geq 0, w \geq 0\right\} = \mathbb{R} \times \left(\mathbb{R}^3\right)_0^+.$$

Contudo, dado que se trata de um problema de produção, sendo conhecidos os tipos de variáveis diremos que $D_P = \left(\mathbb{R}^4\right)^+$.
Se os parâmetros duplicassem, qual seria o efeito na produção de leite? Para tal, basta atender a que

$$P(2x,2y,2z,2w) = 2,5.(2x)^{0,02}(2y)^{0,5}(2z)^{0,252}(2w)^{0,03}$$

$$= \left(2,5 \times 2^{0,805}\right)x^{0,02}y^{0,5}z^{0,25}w^{0,03} = 2^{0,805}P(x,y,z,w).$$

Nesse caso a produção de leite viria multiplicada por $2^{0,805}$.

1.2 Gráfico e linhas de nível

Seja $f : D_f \subset \mathbb{R}^m \longrightarrow \mathbb{R}$. Chamamos **gráfico de** f ao subconjunto de $\mathbb{R}^{m+1}$ dado por

$$Graf(f) = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in D_f\}.$$

O caso em que $m = 2$ admite uma visualização que torna a noção muito intuitiva. Consideremos, por exemplo, as funções f, g, h :$\mathbb{R}^2 \longrightarrow \mathbb{R}$ definidas por

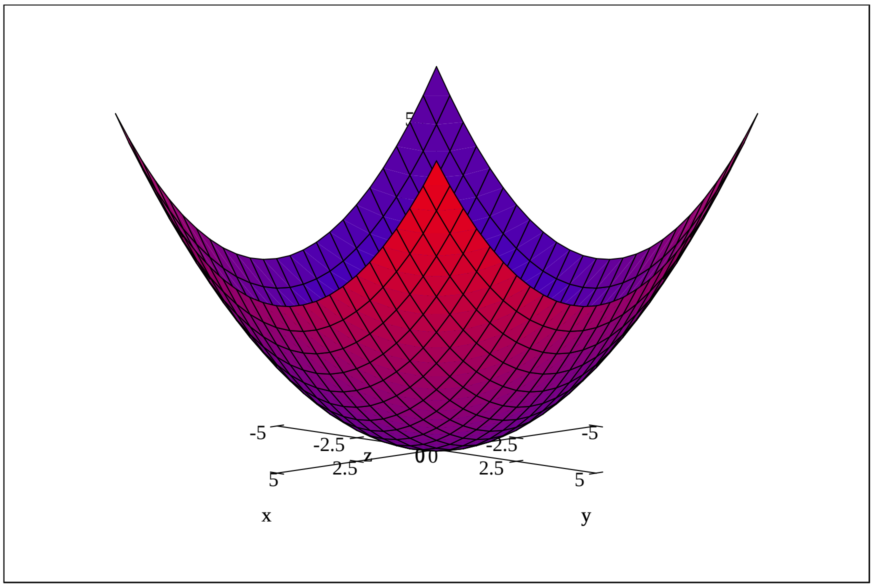
$$f(x,y) = x^2+y^2, \quad g(x,y) = \sqrt{x^2+y^2} \quad \text{e} \quad h(x,y) = x^2-y^2.$$

Os gráficos respectivos são

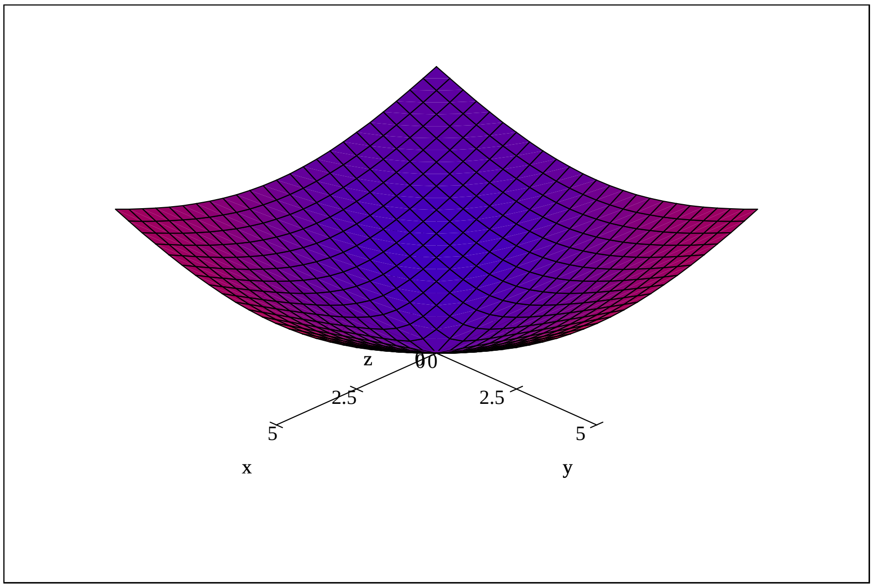
$$
\begin{aligned}
Graf(f) &= \left\{(x,y,z) \in \mathbb{R}^3 : z = x^2 + y^2\right\}, \\
Graf(g) &= \left\{(x,y,z) \in \mathbb{R}^3 : z = \sqrt{x^2 + y^2}\right\}, \\
Graf(h) &= \left\{(x,y,z) \in \mathbb{R}^3 : z = x^2 - y^2\right\}.
\end{aligned}
$$

Geometricamente, temos:

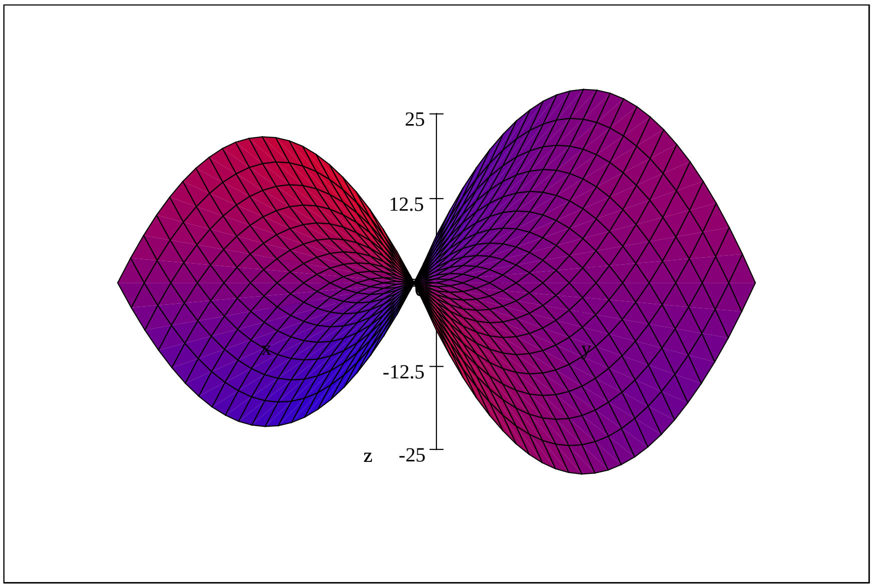
$z = x^2 + y^2$



$z = \sqrt{x^2 + y^2}$



$z = x^2 - y^2$



Como se sabe, o gráfico de uma função real de variável real é uma "linha" em $\mathbb{R}^2$. Como vimos, o gráfico de uma função real definida em $\mathbb{R}^2$ é uma "superfície" em $\mathbb{R}^3$. Trata-se pois de um modelo tridimensional, cujo esboço numa folha de papel apresenta, como é patente, dificuldades na observação das suas propriedades. Nesse sentido, torna-se útil para a compreensão do gráfico e do comportamento da função, o "corte" por planos, nomeadamente verticais ou horizontais. Atentemos no exemplo anterior

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

Corte por planos paralelos ao plano $X0Z$

Corresponde a intersectar o gráfico por planos de equação $y = c$, obtendo assim uma linha plana que é o gráfico de uma função real de uma variável real e não é mais do que uma parábola

$$f(x) = x^2 + c^2.$$

Em particular, se $y = 0$, obtem-se a parábola

$$f(x) = x^2.$$

Corte por planos paralelos ao plano $Y0Z$

Corresponde a intersectar o gráfico por planos de equação $x = c$. Obtêm-se também parábolas mas situadas em planos perpendiculares aos anteriores:

$$f(y) = c^2 + y^2$$

Corte por planos paralelos ao plano $X0Y$ - curvas de nível

Corresponde a intersectar o gráfico por planos horizontais de equação $z = a$, obtendo-se assim curvas de nível cuja projecção no plano $X0Y$ são circunferências centradas em $(0, 0)$ e com raio $\sqrt{a}$. Para uma função arbitrária de $\mathbb{R}^2$ em $\mathbb{R}$, o corte por planos paralelos ao plano $X0Y$ origina as comummente chamadas **curvas de nível** C_a

$$C_a = \left\{(x, y) \in \mathbb{R}^2 : f(x, y) = a\right\}.$$

Na prática, são muitas vezes usadas em mapas de meteorologia para indicar, por exemplo, linhas geográficas de pontos com a mesma pressão atmosférica (isóbaras) ou mapas topográficos para indicar linhas de pontos do terreno que possuem a mesma altitude.

Através dos diferentes cortes que acabámos de descrever, podemos perceber que o gráfico da função $f(x, y) = x^2 + y^2$ é o conjunto de pontos dum parabolóide circular com vértice na origem e com a concavidade "virada para cima".

1.3 Curvas de nível e um modelo económico

Consideremos o modelo de produção $P(K,L)$ onde, para um *input* de capital K e trabalho L, são obtidas P unidades de um determinado produto. Uma curva de nível a para a função P é uma linha no plano KoL dada por

$$C_a = \{(K,L) : P(K,L) = a\}.$$

Esta curva é designada por *isoquântica*. Se se tratar do modelo de Cobb-Douglas

$$P(K,L) = AK^{\alpha}L^{\beta}$$

em que $A > 0$, as *linhas isoquânticas* são

$$C_a = \left\{(K,L) : AK^{\alpha}L^{\beta} = a\right\}$$

ou, resolvendo em ordem a L,

$$C_a = \left\{(K,L) : L = \left(\frac{a}{A}\right)^{\frac{1}{\beta}} K^{-\frac{\alpha}{\beta}}\right\}.$$

Capítulo 2

Limites e Continuidade

2.1 Breves noções topológicas em $\mathbb{R}^m$

No estudo de diversos aspectos das funções de várias variáveis - limites, continuidade, cálculo diferencial, optimização - será importante considerar certas características dos subconjuntos de $\mathbb{R}^m$ onde estas funções estarão definidas. Esta secção sobre noções topológicas pretende introduzir as ideias e definições essenciais à caracterização dos domínios das funções que iremos estudar.

Todos os conceitos que vamos introduzir fazem intervir de algum modo a noção de proximidade. Assim, devemos poder medir a "proximidade" entre pontos de $\mathbb{R}^m$. De entre várias possíveis medidas que poderíamos utilizar para avaliar a distância entre pontos, utilizaremos a de distância euclideana.

Definição 2 *A* ***distância euclideana*** *entre quaisquer dois pontos* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ *é definida por*

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}.$$

É a partir da noção de distância que iremos definir a noção de vizinhança, fundamental em tudo o que se segue. Definimos **bola de centro em** x **e de raio** $\varepsilon > 0$ como sendo o conjunto

$$B_\varepsilon(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^m : d(\mathbf{y}, \mathbf{x}) < \varepsilon\},$$

e definimos os "**vizinhos de raio** ε **de** x" como sendo os pontos de $\mathbb{R}^m$ cuja distância a $\mathbf{x}$ é menor do que ε.

A distância euclideana induz a **norma euclideana**

$$\|\mathbf{x}\| = d(\mathbf{x}, \mathbf{0}),$$

tendo-se

$$\|\mathbf{x} - \mathbf{y}\| = d(\mathbf{x}, \mathbf{y}).$$

A norma euclideana, em $\mathbb{R}$, é precisamente o módulo.

Quando $m = 1$, $\mathbf{x} = x$, tem-se

$$B_\varepsilon(x) = \{y \in \mathbb{R} : \sqrt{(x-y)^2} < \varepsilon\} = \{y \in \mathbb{R} : |x - y| < \varepsilon\},$$

pelo que, em dimensão 1, $B_\varepsilon(x)$ não é mais do que um intervalo aberto centrado em x e de raio ε, isto é,

$$B_\varepsilon(x) =]x - \varepsilon, x + \varepsilon[.$$

Quando $m = 2$,

$$\begin{aligned} B_\varepsilon(\mathbf{x}) &= \{\mathbf{y} \in \mathbb{R}^2 : \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2} < \varepsilon\} \\ &= \{\mathbf{y} \in \mathbb{R}^2 : (y_1 - x_1)^2 + (y_2 - x_2)^2 < \varepsilon^2\}. \end{aligned}$$

Assim, no plano, $B_\varepsilon(\mathbf{x})$ corresponde geometricamente ao interior de um círculo de raio ε e centro em $\mathbf{x}$.

Quando $m = 3$, a noção de bola vai corresponder geometricamente ao interior de uma esfera de raio ε centrada em $\mathbf{x}$.

Para valores da dimensão m superiores a 3, a definição mantém-se, apesar de já não podermos dar-lhe uma interpretação geométrica tão simples.

Para motivar a definição das diversas noções topológicas e simultaneamente facilitar a compreensão das mesmas, vamos começar com um exemplo muito simples de um subconjunto do plano:

$$A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < 1\}.$$

Geometricamente, A é o conjunto dos pontos do plano que está no interior do círculo desenhado na figura 2.1.

Os pontos P, Q e R assinalados na figura estão em posições diferentes relativamente ao conjunto A. O ponto P não só pertence ao conjunto A,

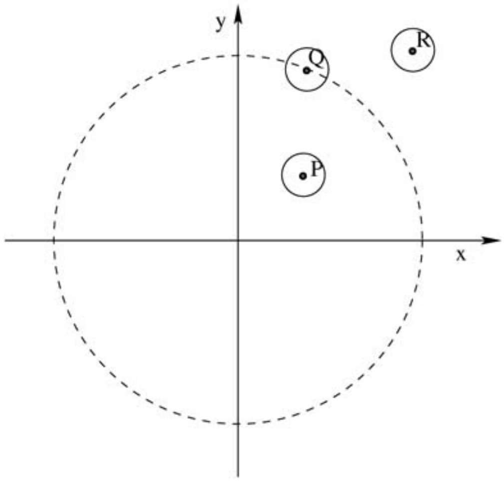


Figura 2.1: Círculo de centro em $(0, 0)$ e raio 1.

como também existe uma vizinhança desse ponto completamente contida no conjunto: diz-se nesse caso que P é um **ponto interior** ao conjunto. O ponto R não só está fora do conjunto (pertence ao complementar[1] de A) como o mesmo se passa com uma sua vizinhança: diz-se neste caso que R é um **ponto exterior** ao conjunto. Relativamente ao ponto Q, uma vez que se situa sobre a circunferência, verifica-se que qualquer vizinhança desse ponto contém pontos quer do conjunto A quer do seu complementar: dizemos nesse caso que Q é um **ponto fronteiro** a A.

Designamos por $int(A)$ o conjunto dos pontos interiores a A, por $ext(A)$ o conjunto dos pontos exteriores a A e por $front(A)$ o conjunto dos pontos fronteiros a A. Resumidamente,

$$\begin{aligned} \mathbf{x} \in int(A) &\Leftrightarrow \exists \varepsilon > 0 : \quad B_\varepsilon(\mathbf{x}) \subseteq A, \\ \mathbf{x} \in ext(A) &\Leftrightarrow \exists \varepsilon > 0 : \quad B_\varepsilon(\mathbf{x}) \subseteq A^C, \\ \mathbf{x} \in front(A) &\Leftrightarrow \forall \varepsilon > 0, \quad B_\varepsilon(\mathbf{x}) \cap A \neq \emptyset, B_\varepsilon(\mathbf{x}) \cap A^C \neq \emptyset. \end{aligned}$$

Da definição de ponto interior, exterior e fronteiro decorre imediatamente que qualquer ponto do espaço satisfaz exactamente uma e só uma das definições, por isso o interior, exterior e fronteira de qualquer conjunto são sempre disjuntos dois a dois e a sua reunião é o espaço inteiro, isto é, estes três conjuntos formam uma partição do espaço.

[1] Designamos por A^C o complementar do conjunto A, isto é $A^C = \mathbb{R}^m \backslash A$

Designa-se por **Aderência** ou **Fecho** de um conjunto a reunião do seu interior com a fronteira,

$$ad(A) = int(A) \cup front(A).$$

Chamamos **conjuntos abertos** aos que coincidem com o seu interior e **conjuntos fechados** aos que coincidem com a sua aderência. É importante referir que um conjunto aberto não é o contrário de conjunto fechado; facilmente podemos encontrar exemplos de conjuntos simultaneamente abertos e fechados ou nem abertos nem fechados.

Um ponto $\mathbf{x}$ diz-se **ponto de acumulação** de um conjunto, se em qualquer vizinhança de $\mathbf{x}$ existir uma infinidade de elementos do conjunto ou, de outro modo, se em qualquer vizinhança do ponto $\mathbf{x}$ existirem elementos do conjunto que não o próprio $\mathbf{x}$. O conjunto dos pontos de acumulação de A é o **conjunto derivado** de A, que se denota por A',

$$\mathbf{x} \in A' \Leftrightarrow \forall \varepsilon > 0, \quad (B_\varepsilon(\mathbf{x}) \backslash \{\mathbf{x}\}) \cap A \neq \emptyset.$$

Prosseguimos esta série de definições referindo o que entendemos por **ponto isolado**, conceito que corresponde à ideia intuitiva de que o ponto será o único do conjunto numa certa vizinhança,

$$\mathbf{x} \text{ é ponto isolado de } A \Leftrightarrow \exists \varepsilon > 0 : \quad (B_\varepsilon(\mathbf{x}) \backslash \{\mathbf{x}\}) \cap A = \emptyset.$$

Note-se que qualquer ponto isolado é ponto fronteiro e que qualquer ponto de acumulação não pode ser exterior ao conjunto, tendo portanto de pertencer à aderência, o que nos mostra que $A' \cup \{\text{pontos isolados de } A\} \subset ad(A)$. Por outro lado, um ponto interior é claramente um ponto de acumulação e um ponto fronteiro que não seja ponto de acumulação será obviamente um ponto isolado provando-se assim a inclusão contrária, $A' \cup \{\text{pontos isolados de } A\} \supset ad(A)$, o que permite concluir que

$$A' \cup \{\text{pontos isolados de } A\} = ad(A).$$

Um conjunto diz-se **limitado** se estiver contido numa bola (de determinado centro e raio), que podemos sem perda de generalidade assumir centrada na origem do referencial. Um subconjunto de $\mathbb{R}^m$ diz-se **compacto** se e só se for limitado e fechado. Como veremos mais adiante a noção de conjunto compacto será especialmente importante em problemas de optimização.

Exemplo 3 *Considerando o exemplo representado na figura 2.1, podemos facilmente ver que todo o ponto do círculo, excepto a circunferência, é interior, pelo que* $int(A) = A$*, sendo por isso* A *um conjunto aberto. O exterior de* A *corresponde neste caso à parte de fora do círculo (excluindo a circunferência) e a fronteira corresponde à circunferência. Assim* $ad(A) = int(A) \cup front(A) \neq A$ *e por isso* A *não é fechado e por isso não pode ser compacto. Finalmente, como não existem pontos isolados* $A' = ad(A)$*.*

2.2 Sucessões em $\mathbb{R}^m$

Faremos neste ponto uma menção breve ao conceito de sucessão em $\mathbb{R}^m$. Partindo da noção de sucessão em $\mathbb{R}$, tópico que habitualmente faz parte do programa de qualquer disciplina de Cálculo em dimensão 1, não é difícil intuir a noção de sucessão em espaços de dimensão superior. É aliás de esperar que propriedades análogas às que se verificam em dimensão 1 tenham aqui enunciados paralelos com a formalização adequada à dimensão do espaço em causa.

Com efeito, por exemplo, dar uma sucessão em $\mathbb{R}^2$ não é mais do que a cada natural $n \in \mathbb{N}$, fazer corresponder um par ordenado (a_n, b_n). Se quisermos ser formais, diremos que uma sucessão em $\mathbb{R}^2$ é uma função S de $\mathbb{N}$ em $\mathbb{R}^2$ tal que $n \xrightarrow{S} (a_n, b_n)$. Os elementos designar-se-ão por

$$(a_1, b_1), (a_2, b_2), (a_3, b_3), \ldots, (a_n, b_n), \ldots$$

ou, de forma abreviada, $((a_n, b_n))_{n\in\mathbb{N}}$, ou ainda, simplificando a notação, $((a_n, b_n))$, ou mesmo (com algum abuso), (a_n, b_n). Uma questão natural é a da existência de limite. É de esperar que:

1) o limite, se existir, seja único,

2) a sucessão $((a_n, b_n))$ convirja para (a, b) se e só se as sucessões coordenadas $((a_n))$ e $((b_n))$ convergirem para a e b, respectivamente, isto é, $a_n \to a$ e $b_n \to b$,

3) se uma sucessão convergir, qualquer sua subsucessão também convirja para o mesmo limite.

Tocaremos apenas estes pontos.

Definição 4 *Uma sucessão em $\mathbb{R}^m$ é uma sequência ordenada de elementos de $\mathbb{R}^m$, $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n, \ldots$ tal que cada elemento, por estar em $\mathbb{R}^m$, possui m coordenadas. Assim, com $n = 1, 2, 3, \ldots$,*

$$\begin{aligned}
\mathbf{u}_1 &= (u_{11}, u_{12}, \ldots, u_{1m})\\
\mathbf{u}_2 &= (u_{21}, u_{22}, \ldots, u_{2m})\\
\mathbf{u}_3 &= (u_{31}, u_{32}, \ldots, u_{3m})\\
&\vdots\\
\mathbf{u}_n &= (u_{n1}, u_{n2}, \ldots, u_{nm}).\\
&\vdots
\end{aligned}$$

*Ao elemento $\mathbf{u}_n$ chama-se **termo geral** pois contém a expressão genérica da sucessão.*

Exemplo 5 *A sucessão $(\mathbf{v}_n)_{n\in\mathbb{N}}$ cujo termo geral é $\mathbf{v}_n = (\frac{1}{n}, \ln(e + \frac{1}{n}))$ é uma sucessão em $\mathbb{R}^2$. Vejamos os primeiros elementos*

$$\begin{aligned}
\mathbf{v}_1 &= (1, \ln(e+1))\\
\mathbf{v}_2 &= (\frac{1}{2}, \ln(e+\frac{1}{2}))\\
\mathbf{v}_3 &= (\frac{1}{3}, \ln(e+\frac{1}{3}))\\
&\vdots\\
\mathbf{v}_n &= (\frac{1}{n}, \ln(e+\frac{1}{n}))\\
&\vdots
\end{aligned}$$

Observemos que tanto as primeiras coordenadas como as segundas constituem sucessões em $\mathbb{R}$. Com efeito, são as duas sucessões que podemos designar por $(v_{n1})_{n\in\mathbb{N}}$ e $(v_{n2})_{n\in\mathbb{N}}$, ou seja,

sucessão $(v_{n1})_{n\in\mathbb{N}}$ $\quad 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$

sucessão $(v_{n2})_{n\in\mathbb{N}}$ $\quad \ln(e+1), \ln(e+\frac{1}{2}), \ln(e+\frac{1}{3}), \ldots, \ln(e+\frac{1}{n}), \ldots$

Tomando agora a sucessão $(\mathbf{z}_n)_{n\in\mathbb{N}}$ em $\mathbb{R}^3$, com $\mathbf{z}_n = (\frac{1}{n}, \ln(e+\frac{1}{n}), \frac{4n^5+5n^3}{8n^5+7n^2})$, *os primeiros elementos são*

$$\begin{aligned}
\mathbf{z}_1 &= (1, \ln(e+1), \frac{3}{5}) \\
\mathbf{z}_2 &= (\frac{1}{2}, \ln(e+\frac{1}{2}), \frac{42}{71}) \\
\mathbf{z}_3 &= (\frac{1}{3}, \ln(e+\frac{1}{3}), \frac{123}{223}) \\
&\vdots \\
\mathbf{z}_n &= (\frac{1}{n}, \ln(e+\frac{1}{n}), \frac{4n^5+5n^3}{8n^5+7n^2}) \\
&\vdots
\end{aligned}$$

Neste caso, para além das duas sucessões em $\mathbb{R}$, $(\mathbf{z}_{n1})$ e $(\mathbf{z}_{n2})$ que são, respectivamente, as sucessões das primeiras e segundas coordenadas de $(\mathbf{z}_n)$ (e coincidem com (v_{n1}) e (v_{n2}), respectivamente), existe ainda a sucessão em $\mathbb{R}$, (z_{n3}), que é precisamente a sucessão das terceiras coordenadas de $(\mathbf{z}_n)$. Neste caso, temos

$$\begin{array}{ll}
\textit{sucessão } (z_{n1})_{n\in\mathbb{N}} & 1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots \\
\textit{sucessão } (z_{n2})_{n\in\mathbb{N}} & \ln(e+1), \ln(e+\frac{1}{2}), \ln(e+\frac{1}{3}), \ldots, \ln(e+\frac{1}{n}), \ldots \\
\textit{sucessão } (z_{n3})_{n\in\mathbb{N}} & \frac{3}{5}, \frac{42}{71}, \frac{123}{223}, \ldots, \frac{4n^5+5n^3}{8n^5+7n^2}, \ldots
\end{array}$$

Os exemplos anteriores ilustram o facto geral de cada sucessão $(\mathbf{u}_n)$ em $\mathbb{R}^m$ comportar precisamente **m** sucessões de termos reais - as **sucessões coordenadas** da sucessão dada. Por exemplo, a sucessão das terceiras coordenadas é

$$\text{sucessão } (u_{n3})_{n\in\mathbb{N}}\ u_{13},\ u_{23},\ u_{33}, \ldots,\ u_{n3}, \ldots$$

e, mais geralmente, a sucessão das coordenadas de ordem j é

$$\text{sucessão } (u_{nj})_{n\in\mathbb{N}}\ u_{1j},\ u_{2j},\ u_{3j}, \ldots,\ u_{nj}, \ldots\ .$$

Definição 6 *Seja $(\mathbf{u}_n)_{n\in\mathbb{N}}$ uma sucessão em $\mathbb{R}^m$ e $\mathbf{u} \in \mathbb{R}^m$. Diz-se que $(\mathbf{u}_n)_{n\in\mathbb{N}}$ **converge para** $\mathbf{u}$ se, qualquer que seja a bola centrada em $\mathbf{u}$, $B_\epsilon(\mathbf{u})$ (com ϵ qualquer), existe um inteiro positivo p tal que $\mathbf{u}_n \in B_\epsilon(\mathbf{u})$, para todo o $n > p$. Ao vector $\mathbf{u}$ chama-se **limite** de $(\mathbf{u}_n)_{n\in\mathbb{N}}$ e usa-se a notação*

$$\mathbf{u}_n \to \mathbf{u} \Leftrightarrow \lim \mathbf{u}_n = \mathbf{u}.$$

Simbolicamente,

$$\mathbf{u}_n \to \mathbf{u} \quad se \quad \forall_{\epsilon>0} : \exists_{p\in\mathbb{N}} : n > p \Rightarrow d(\mathbf{u}_n, \mathbf{u}) < \epsilon.$$

onde $d(\mathbf{u}_n, \mathbf{u})$ *é a distância de* $\mathbf{u}_n$ *a* $\mathbf{u}$*, mais precisamente,*

$$\begin{aligned} d(\mathbf{u}_n, \mathbf{u}) &= \sqrt{(u_{n1}-u_1)^2 + (u_{n2}-u_2)^2 + (u_{n3}-u_3)^2 + \ldots + (u_{nm}-u_m)^2} \\ &= \|\mathbf{u}_n - \mathbf{u}\|. \end{aligned}$$

Analogamente ao que foi dito no início deste ponto, quando não houver perigo de confusão, usaremos a notação simplificada $(\mathbf{u}_n)$ em vez de $(\mathbf{u}_n)_{n\in\mathbb{N}}$.

Proposição 7 *Uma sucessão* $(\mathbf{u}_n)$ *em* $\mathbb{R}^m$ *converge para* $\mathbf{u} \in \mathbb{R}^m$ *se e só se as suas sucessões coordenadas convergirem para as coordenadas de* $\mathbf{u}$*.*

Dem (Condição necessária). Para cada sucessão coordenada (u_{nj}), tem-se

$$|u_{nj} - u_j| = \sqrt{(u_{nj}-u_j)^2} < d(\mathbf{u}_n, \mathbf{u}).$$

Ora, como $(\mathbf{u}_n)$ converge para $\mathbf{u}$ em $\mathbb{R}^m$, sai facilmente da definição que u_{nj} converge para u_j em $\mathbb{R}$, $\forall j = 1, \ldots, m$.

(Condição suficiente) Esta condição deduz-se usando a desigualdade

$$\begin{aligned} d(u_n, u) &= \sqrt{(u_{n1}-u_1)^2 + (u_{n2}-u_2)^2 + (u_{n3}-u_3)^2 + \ldots + (u_{nm}-u_m)^2} \\ &\le |u_{n1}-u_1| + |u_{n2}-u_2| + |u_{n3}-u_3| + \ldots + |u_{nm}-u_m|, \end{aligned}$$

juntamente com a definição de convergência, aplicada a cada uma das sucessões coordenadas. □

Exemplo 8 *Consideremos as sucessões* $(\mathbf{v}_n)$ *e* $(\mathbf{z}_n)$ *do exemplo anterior,*

$$\mathbf{v}_n = \left(\frac{1}{n}, \ln(e + \frac{1}{n})\right) \quad e \quad \mathbf{z}_n = \left(\frac{1}{n}, \ln(e + \frac{1}{n}), \frac{4n^5 + 5n^3}{8n^5 + 7n^2}\right).$$

Pela proposição anterior,

$$\begin{aligned} \lim\ \mathbf{v}_n &= \lim \left(\frac{1}{n}, \ln(e + \frac{1}{n})\right) = \left(\lim \frac{1}{n}, \lim(\ln(e + \frac{1}{n}))\right) = (0,1), \\ \lim\ \mathbf{z}_n &= \lim \left(\frac{1}{n}, \ln(e + \frac{1}{n}), \frac{4n^5 + 5n^3}{8n^5 + 7n^2}\right) \\ &= \left(\lim \frac{1}{n}, lim(\ln(e + \frac{1}{n})), \lim \frac{4n^5 + 5n^3}{8n^5 + 7n^2}\right) \\ &= (0, 1, \frac{1}{2}). \end{aligned}$$

Definição 9 *Dada uma sucessão $(\mathbf{u}_n)$ em $\mathbb{R}^m$, chama-se **subsucessão** de $(\mathbf{u}_n)$ a uma sucessão obtida a partir de $(\mathbf{u}_n)$ por selecção ordenada de alguns termos em número infinito.*

Exemplo 10 *Dada a sucessão de termo geral $\mathbf{w}_n = \left(\frac{1}{n}, \frac{n+1}{n+2}\right)$, isto é,*

$$\left(1, \frac{2}{3}\right), \left(\frac{1}{2}, \frac{3}{4}\right), \left(\frac{1}{3}, \frac{4}{5}\right), \left(\frac{1}{4}, \frac{5}{6}\right), \left(\frac{1}{5}, \frac{6}{7}\right), \ldots, \left(\frac{1}{n}, \frac{n+1}{n+2}\right), \ldots$$

podemos considerar como exemplo de subsucessão

$$\left(\frac{1}{2}, \frac{3}{4}\right), \left(\frac{1}{4}, \frac{5}{5}\right), \left(\frac{1}{6}, \frac{7}{8}\right), \left(\frac{1}{8}, \frac{9}{10}\right), \ldots, \left(\frac{1}{2n}, \frac{2n+1}{2n+2}\right), \ldots$$

que é precisamente a subsucessão dos termos de ordem par de $(\mathbf{w}_n)$.

Proposição 11 *Toda a subsucessão de uma sucessão convergente em $\mathbb{R}^m$ é também convergente e para o mesmo limite.*

Dem O resultado é consequência da proposição anterior e do facto das coordenadas da subsucessão constituirem subsucessões das sucessões coordenadas da sucessão dada, o que reduz o caso à aplicação do resultado análogo de unicidade do limite já conhecido em $\mathbb{R}$. □

Pelo exposto, tentou ilustrar-se o facto de as noções e propriedades das sucessões em $\mathbb{R}^m$ assentarem geralmente nas análogas em $\mathbb{R}$, com as adaptações adequadas de forma natural à dimensão. Espera-se que o leitor, se necessitar, possa utilizar este facto para estabelecer outras propriedades que omitimos aqui. Um exemplo importante e óbvio é o das propriedades operatórias dos limites de sucessões.

2.3 Limites

Introduziremos primeiramente a definição de limite de uma função real de variável vectorial, fazendo especial ênfase nos casos de $\mathbb{R}^2$ e também $\mathbb{R}^3$. Em seguida falaremos do caso de funções vectoriais de variável vectorial.

2.3.1 Definições e propriedades

Seja $\mathcal{D}$ um aberto de $\mathbb{R}^m$, $\mathbf{a}$ um ponto aderente a $\mathcal{D}$ e $b \in \mathbb{R}$. Consideremos uma função $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$.

Definição 12 *(segundo Heine) O limite de $f(\mathbf{x})$ quando $\mathbf{x}$ tende para $\mathbf{a}$ é b se, para toda a sucessão $(\mathbf{a}_n)$ que tende para $\mathbf{a}$ $(\mathbf{a}_n \neq \mathbf{a})$, a sucessão imagem $f(\mathbf{a}_n)$ tender para b,*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b \;\;\Leftrightarrow\;\; [\forall (\mathbf{a}_n) \subset \mathcal{D} \subset \mathbb{R}^m, \mathbf{a}_n \neq \mathbf{a} :\; (\mathbf{a}_n) \to \mathbf{a} \;\Rightarrow\; (f(\mathbf{a}_n)) \to b]\,.$$

Definição 13 *(segundo Cauchy) O limite de $f(\mathbf{x})$ quando $\mathbf{x}$ tende para $\mathbf{a}$ é b se for satisfeita a condição δ, ε,*

$$\forall \delta > 0 : \exists \varepsilon > 0 : \forall \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^m, \mathbf{x} \neq \mathbf{a}, \|\mathbf{x} - \mathbf{a}\| < \varepsilon \Rightarrow |f(\mathbf{x}) - b| < \delta.$$

Como referido anteriormente, $\|\mathbf{x} - \mathbf{a}\| = d(\mathbf{x}, \mathbf{a})$, donde

$$\|\mathbf{x} - \mathbf{a}\| < \epsilon \;\Leftrightarrow\; d(\mathbf{x}, \mathbf{a}) < \epsilon \;\Leftrightarrow\; \mathbf{x} \in B_\epsilon(\mathbf{a}),$$

pelo que usaremos indiferentemente qualquer das expressões.

Teorema 14 *As duas definições anteriores são equivalentes.*

Dem (1) Suponhamos que $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b$ segundo Heine e admitamos, por absurdo, que existe um δ para o qual não existe ε tal que, se $\mathbf{x} \in \mathbb{R}^m$ verifica $d(\mathbf{x}, \mathbf{a}) < \varepsilon$, então $|f(\mathbf{x}) - b| \geq \delta$. Em particular, para cada $\varepsilon = 1/n$, existe um elemento $\mathbf{x}_n$ que verifica simultaneamente

$$d(\mathbf{x}_n, \mathbf{a}) < 1/n \;\text{ e }\; |f(\mathbf{x}_n) - b| \geq \delta.$$

A sucessão $\mathbf{x}_n$ converge para $\mathbf{a}$, porque para $\mu > 0$ dado, ao tomarmos $N > 1/\mu$, obtemos um inteiro tal que $\forall n \geq N$, temos $d(\mathbf{x}_n, \mathbf{a}) \leq 1/N < \mu$.

A sucessão $f(\mathbf{x}_n)$ não pode convergir para b porque, se $n \geq N$, os elementos $f(\mathbf{x}_n)$ não pertencem à bola aberta de centro b e raio δ, o que está em contradição com a definição de limite segundo Heine.

(2) Suponhamos que a condição δ, ε é satisfeita e seja $(\mathbf{x}_n)$ uma sucessão que tende para $\mathbf{a}$. Mostremos que a sucessão $f(\mathbf{x}_n)$ tende para b. Fixemos $\delta > 0$. Precisamos de encontrar um inteiro N tal que, para $n \geq N$, o elemento $f(\mathbf{x}_n)$ pertence à bola aberta de centro b e de raio δ.

Se a condição δ, ε for satisfeita, existe um $\varepsilon > 0$ tal que, se $d(\mathbf{x}_n, \mathbf{a}) < \varepsilon$, então $|f(\mathbf{x}_n) - b| < \delta$. Como a sucessão $(\mathbf{x}_n)$ converge para $\mathbf{a}$, para esse número ε, existe um inteiro N tal que $d(\mathbf{x}_n, \mathbf{a}) < \varepsilon, \forall n \geq N$ e temos, para este mesmo N, $|f(\mathbf{x}_n) - b| < \delta, \forall n \geq N$. $\square$

Nota 15 *O conceito de limite que apresentámos é designado por certos autores por "limite por valores diferentes" .*

Nota 16 *Dado que* $\|\mathbf{x}-\mathbf{a}\| = d(\mathbf{x},\mathbf{a})$, *a condição em* δ, ε *pode também ser enunciada em termos de bolas abertas do seguinte modo: o limite de* $f(\mathbf{x})$ *é* b *quando* $\mathbf{x}$ *tende para* $\mathbf{a}$ *se para toda a bola aberta* $B_\delta(b)$ *(que em* $\mathbb{R}$ *é o intervalo* $]\, b-\delta, b+\delta\, [$ *) existir uma bola aberta* $B_\varepsilon(\mathbf{a})$ *tal que tal que para todo o* $\mathbf{x}$ *em* $B_\varepsilon(\mathbf{a})$, *a sua imagem* $f(\mathbf{x})$ *pertence a* $B_\delta(b)$.

Da definição de limite segundo Cauchy, deduz-se de imediato que

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b \text{ se e só se } \lim_{\mathbf{x}\to\mathbf{a}} |f(\mathbf{x}) - b| = 0.$$

Usando a definição de Heine, é fácil estabelecer as seguintes regras operatórias,

Teorema 17 *Sejam* f *e* g *funções de* $\mathbb{R}^m$ *em* $\mathbb{R}$ *de domínios abertos e* $\mathbf{a}$ *um ponto aderente aos respectivos domínios. Se* $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b$ *e* $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = c$ *então*

1. $\lim_{\mathbf{x}\to\mathbf{a}} (f+g)(\mathbf{x}) = b + c$.
2. $\lim_{\mathbf{x}\to\mathbf{a}} (f\cdot g)(\mathbf{x}) = b\cdot c$.
3. $\lim_{\mathbf{x}\to\mathbf{a}} \dfrac{f(\mathbf{x})}{g(\mathbf{x})} = \dfrac{b}{c}$ *(se* $g(\mathbf{x})$ *e* $c \neq 0$*).*

Teorema 18 *Sejam* $f \leq g \leq h$ *três funções de* $\mathbb{R}^m$ *em* $\mathbb{R}$. *Se* $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b$ *e* $\lim_{\mathbf{x}\to\mathbf{a}} h(\mathbf{x}) = b$, *então* $\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = b$.

Dem Fixemos $\varepsilon > 0$. Pela definição de limite, existem dois números δ_1 e δ_2 tais que

$$\begin{aligned} \|\mathbf{x}-\mathbf{a}\| &< \varepsilon_1 \Rightarrow |f(\mathbf{x}) - b| < \delta, \\ \|\mathbf{x}-\mathbf{a}\| &< \varepsilon_2 \Rightarrow |h(\mathbf{x}) - b| < \delta. \end{aligned}$$

Tomemos $\varepsilon = \min(\varepsilon_1, \varepsilon_2)$. Se $\|\mathbf{x}-\mathbf{a}\| < \varepsilon$, temos

$$b - \delta < f(\mathbf{x}) \leq g(\mathbf{x}) \leq h(\mathbf{x}) < b + \delta,$$

de onde $|g(\mathbf{x}) - b| < \delta$. $\square$

Sendo $B \subset \mathcal{D}$ e $\mathbf{a}$ aderente a B, consideremos o limite de $f(\mathbf{x})$ quando $\mathbf{x} \to \mathbf{a}$, mas ao longo do conjunto B, isto é, com $\mathbf{x} \in B$. Este não é mais do que o limite da restrição de f a B quando $\mathbf{x} \to \mathbf{a}$, pelo que fazem sentido as definições segundo Cauchy e segundo Heine, neste caso aplicadas à função $f\,|_B$. Facilmente se conclui o seguinte:

Proposição 19 *Se* $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b$ *então o limite de* f *em* $\mathbf{a}$ *ao longo de qualquer subconjunto de* $\mathcal{D}$ *a que* $\mathbf{a}$ *seja aderente, existe e tem o valor* b*, isto é*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b \Rightarrow \lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in B}} f(\mathbf{x}) = b, \forall B \subset \mathcal{D} : \mathbf{a} \in ad(B).$$

A propriedade anterior é extremamente importante e torna-se particularmente útil nos casos em que não existe limite.

Teorema 20 *Seja* f *de* $\mathcal{D} \subset \mathbb{R}^m$ *em* $\mathbb{R}$*. Sejam* $B_1, B_2, \ldots, B_s$*, conjuntos disjuntos dois a dois, tais que* $B_1 \cup B_2 \cup \ldots \cup B_s = \mathcal{D}\backslash\{\mathbf{a}\}$ *com* $\mathbf{a}$ *ponto aderente de cada um dos* B_j*. Então a função* f *tem limite* b *quando* $\mathbf{x} \to \mathbf{a}$ *se e só se existirem todos os limites de* f *quando* $\mathbf{x} \to \mathbf{a}$ *ao longo de todos os conjuntos* $B_1, B_2, \ldots, B_s$ *e forem iguais a* b*. Isto é,*

$$\lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in B}} f(\mathbf{x}) = b \Leftrightarrow \lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in B_1}} f(\mathbf{x}) = \lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in B_2}} f(\mathbf{x}) = \ldots = \lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in B_s}} f(\mathbf{x}) = b.$$

Dem Resulta imediatamente da proposição anterior que a condição é necessária. Vejamos o que se passa quanto à condição ser suficiente. Como, para cada j

$$\lim_{\substack{\mathbf{x}\to\mathbf{a}\\ \mathbf{x}\in B_j}} f(\mathbf{x}) = b$$

então, qualquer que seja $\delta > 0$, existe um $\varepsilon_j = \varepsilon_j(\delta)$ tal que

$$\mathbf{x} \in B_{\epsilon_j}(\mathbf{a}) \cap B_j \;\Rightarrow\; f(\mathbf{x}) \in B_\delta(b).$$

Tomando $\varepsilon = \min\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_s\}$, uma vez que cada $\mathbf{x} \in \mathcal{D}\backslash\{\mathbf{a}\}$ pertence a algum B_j, tem-se

$$\mathbf{x} \in B_\epsilon(\mathbf{a}) \;\Rightarrow\; \exists_{1\le j\le s} : \mathbf{x} \in B_{\epsilon_j}(\mathbf{a}) \cap B_j \Rightarrow f(\mathbf{x}) \in B_\delta(b).$$

logo, segundo a definição de Cauchy, $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = b$. $\square$

Nota 21 *Em* $\mathbb{R}$*, o teorema anterior corresponde à propriedade seguinte: o limite de* f *em* a *existe se e só se os limites à esquerda e à direita existirem e forem iguais.*

2.3.2 Cálculo de limites

A caracterização em termos de limites ao longo de subconjuntos permite mostrar que alguns limites não existem.

Exemplo 22 *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ a função definida por*

$$f(x,y) = \begin{cases} \dfrac{xy}{x^2+y^2} & \text{se } (x,y) \neq (0,0), \\ \\ 0 & \text{no caso contrário.} \end{cases}$$

Pretende-se calcular $\lim\limits_{(x,y)\to(0,0)} f(x,y)$ ou concluir que não existe. Sendo $B^* = \{(x,x) \in \mathbb{R}^2\}$ e $B^{**} = \{(0,y) : y \in \mathbb{R}\}$ tem-se

$$\lim_{\substack{(x,y)\to(0,0)\\(x,y)\in B^*}} f(x,y) = \frac{1}{2} \quad e \quad \lim_{\substack{(x,y)\to(0,0)\\(x,y)\in B^{**}}} f(x,y) = 0.$$

Consequentemente, a função $f(x,y)$ não tem limite quando (x,y) tende para $(0,0)$.

Acabámos de calcular dois limites ao longo de duas rectas diferentes, isto é, de duas direcções diferentes, uma de declive 1 e outra vertical.

Poderíamos ter tentado ver o limite ao longo de uma recta genérica passando pelo ponto onde se quer calcular o limite, neste caso o ponto $(0,0)$. Corresponderia a considerar o conjunto $B_m = \{(x,y) \in \mathbb{R}^2 : y = mx\}$, que é o conjunto de todas as rectas que passam no ponto $(0,0)$ e que se distinguem pelo declive m que é característica de cada uma e também da recta vertical B^{**}. Então,

$$\begin{aligned} \lim_{\substack{(x,y)\to(0,0)\\(x,y)\in B_m}} f(x,y) &= \lim_{\substack{(x,y)\to(0,0)\\y=mx}} \frac{xy}{x^2+y^2} = \lim_{x\to 0} \frac{x(mx)}{x^2+(mx)^2} \\ &= \lim_{x\to 0} \frac{mx^2}{x^2(1+m^2)} = \lim_{x\to 0} \frac{m}{1+m^2} = \frac{m}{1+m^2}, \end{aligned}$$

o que mostra que o limite de f quando (x,y) tende para $(0,0)$ ao longo de cada recta $y = mx$ varia com a recta em causa, pelo que não existe limite de $f(x,y)$ quando $(x,y) \to (0,0)$. Dizendo de outro modo, os limites direccionais de $f(x,y)$ quando (x,y) tende para $(0,0)$ variam com a direcção (já não seria preciso ir ver o limite ao longo da recta vertical, a menos que se quisessse determinar especificamente esse valor).

Genericamente, chamam-se **limites direccionais de** f quando (x, y) tende para (a, b) aos limites ao longo das rectas $y = m(x - a) + b$ e também ao longo da recta vertical $x = a$.

O teorema de enquadramento apresentado anteriormente é muito útil para o cálculo explícito de um limite.

Exemplo 23 *Mostremos que* $\lim_{(x,y)\to(0,0)} f(x, y) = 0$, *com*

$$f(x, y) = \frac{xy}{\sqrt{x^2 + y^2}}.$$

Resulta imediatamente da desigualdade $(x - y)^2 \geq 0$ *que* $|xy| \leq x^2 + y^2$*. De onde sai que*

$$0 \leq |f(x, y)| = \left|\frac{xy}{\sqrt{x^2 + y^2}}\right| = \frac{|xy|}{\sqrt{x^2 + y^2}} \leq \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} \leq \sqrt{x^2 + y^2}.$$

Como os membros extremos, da esquerda e da direita, tendem para 0 *quando* (x, y) *tende para* $(0, 0)$*, o resultado deduz-se do teorema referido.*

Vejamos agora uma generalização do exemplo anterior.

Exemplo 24 *Limites em* $\mathbf{0}$ *de fracções racionais. Seja*

$$f(\mathbf{x}) = \frac{P(\mathbf{x})}{\|\mathbf{x}\|^r}$$

uma aplicação de $\mathbb{R}^m$ *em* $\mathbb{R}$ *definida como o quociente de um polinómio* $P(\mathbf{x})$ *de* m *variáveis pela* r*-ésima potência da norma de* $\mathbf{x}$*.*

Se o polinómio for um monómio $x_{i_1} x_{i_2} \dots x_{i_s}$ *de grau maior do que* r*, então* $\lim_{\mathbf{x}\to\mathbf{0}} f(\mathbf{x}) = 0$*. Basta mostrar que*

$$\left|\frac{x_{i_1} x_{i_2} \dots x_{i_s}}{\|\mathbf{x}\|^r}\right| \leq \|\mathbf{x}\|^{s-r},$$

ou mostrar que

$$\frac{(x_{i_1} x_{i_2} \dots x_{i_s})^2}{\|\mathbf{x}\|^{2r}} \leq \|\mathbf{x}\|^{2(s-r)}.$$

Esta condição é evidente porque, para cada i*,* $x_i^2 \leq \|\mathbf{x}\|^2$*.*

O teorema sobre o limite de uma composição é também muito útil. É necessário reconhecer uma composição. Seguem-se dois exemplos de tal situação, que apresentamos de forma formal. Contudo, na prática, o cálculo decorre de forma natural e quase automática, sem que tal reconhecimento seja formalmente feito.

Exemplo 25 *Calcular o limite em* $(0,0,0)$ *da função* $f : \mathbb{R}^3 \to \mathbb{R}$ *definida por*

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

A função f *é a composta de duas funções*

$$f = g \circ h, \quad h : \mathbb{R}^3 \to \mathbb{R}, \quad g : \mathbb{R} \to \mathbb{R},$$

$$h(x,y,z) = x^2+y^2+z^2, \quad g(x) = \frac{\sin x}{x}.$$

Como $\lim\limits_{(x,y,z)\to(0,0,0)} h(x,y,z) = 0$ *e* $\lim\limits_{x\to 0} g(x) = 1$*, tem-se*

$$\lim_{(x,y,z)\to(0,0,0)} f(x,y,z) = 1.$$

Exemplo 26 *Calculemos o limite em* $(0,0)$ *da função* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida por*

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

A função f *escreve-se*

$$f = (x-y).(g \circ h), \quad h : \mathbb{R}^2 \to \mathbb{R}, \quad g : \mathbb{R} \to \mathbb{R},$$

$$h(x,y) = x+y, \quad g(x) = \frac{x}{\sin x}.$$

Como

$$\lim_{(x,y)\to(0,0)} h(x,y) = 0, \ \lim_{x\to 0} g(x) = 1 \ \text{e} \ \lim_{(x,y)\to(0,0)} x-y = 0,$$

tem-se

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0.$$

2.4 Funções de $\mathbb{R}^m$ em $\mathbb{R}^p$

Falámos até aqui de funções reais de variável vectorial. Neste ponto vamos introduzir funções vectoriais de variável vectorial, isto é, cujo domínio está contido em $\mathbb{R}^m$ e o contradomínio está contido em $\mathbb{R}^p$. Consideremos, por exemplo, a função

$$\begin{array}{cccc} \mathbf{f} & : & \mathbb{R}^2 & \to & \mathbb{R}^3 \\ & & (x,y) & \to & \left(x+y, x.y, \sqrt{x^2+y^2}\right). \end{array}$$

A cada elemento (x,y) do domínio em $\mathbb{R}^2$ corresponde um elemento de $\mathbb{R}^3$, $\mathbf{z} = (z_1, z_2, z_3)$, ou, dito de outro modo, três valores ordenados em $\mathbb{R}$: a soma das coordenadas, o produto das coordenadas e a norma de (x,y), isto é, $z_1 = x+y$, $z_2 = x.y$ e $z_3 = \sqrt{x^2+y^2}$. Assim, a função $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^3$ pode ser encarada como tendo três funções coordenadas

$$\begin{aligned} f_1 &: \mathbb{R}^2 \to \mathbb{R} \quad \text{tal que} \quad f_1(x,y) = x+y \\ f_2 &: \mathbb{R}^2 \to \mathbb{R} \quad \text{tal que} \quad f_2(x,y) = x.y \\ f_3 &: \mathbb{R}^2 \to \mathbb{R} \quad \text{tal que} \quad f_3(x,y) = \sqrt{x^2+y^2} \end{aligned}$$

Mais geralmente, uma função $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^p$ tem p funções coordenadas $f_i : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ e pode escrever-se

$$\mathbf{f} = (f_1, f_2, f_3, \ldots, f_p).$$

Logo, $\mathbf{f}$ pode ser considerada como um sistema de p funções reais, cada uma com m variáveis reais e com domínio $\mathcal{D}$,

$$\mathbf{x} = (x_1, \ldots, x_m) \longrightarrow \begin{cases} y_1 = f_1(\mathbf{x}) = f_1(x_1, \ldots, x_m) \\ y_2 = f_2(\mathbf{x}) = f_2(x_1, \ldots, x_m) \\ \ldots \\ y_p = f_p(\mathbf{x}) = f_p(x_1, \ldots, x_m) \end{cases}$$

Por este facto, as noções introduzidas anteriormente para funções de $\mathbb{R}^m$ em $\mathbb{R}$ podem ser alargadas de forma natural a funções de $\mathbb{R}^m$ em $\mathbb{R}^p$ e o estudo de uma função vectorial de variável vectorial $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^p$ está muito ligado ao estudo de p funções reais de variável vectorial $f_i : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}, 1 \leq i \leq p$. Vejamos a noção de limite.

Seja $\mathcal{D}$ um aberto de $\mathbb{R}^m$ e $\mathbf{a}$ um ponto aderente a $\mathcal{D}$. Seja $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^p$ uma função de domínio $\mathcal{D}$ e $\mathbf{b} \in \mathbb{R}^p$.

Definição 27 *(segundo Heine) O limite de* $\mathbf{f}(\mathbf{x})$ *quando* $\mathbf{x}$ *tende para* $\mathbf{a}$ *é* $\mathbf{b}$ *se, para toda a sucessão* $(\mathbf{a}_n)$ $(\mathbf{a}_n \neq \mathbf{a})$ *que tende para* $\mathbf{a}$*, a sucessão imagem* $\mathbf{f}(\mathbf{a}_n)$ *tende para* $\mathbf{b}$*,*

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{b} \;\Leftrightarrow\; \left[\forall (\mathbf{a}_n) \subset \mathbb{R}^m, \mathbf{a}_n \neq \mathbf{a} :\; (\mathbf{a}_n) \to \mathbf{a} \;\Rightarrow\; (\mathbf{f}(\mathbf{a}_n)) \to \mathbf{b}\right].$$

Definição 28 *(segundo Cauchy) O limite de* $\mathbf{f}(\mathbf{x})$ *quando* $\mathbf{x}$ *tende para* $\mathbf{a}$ *é* $\mathbf{b}$ *se for satisfeita a condição* δ, ε*:*

$$\forall \delta > 0\ \exists \varepsilon > 0\ \forall \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^m, \mathbf{x} \neq \mathbf{a}, d(\mathbf{x}, \mathbf{a}) < \varepsilon \Rightarrow d(\mathbf{f}(\mathbf{x}), \mathbf{b}) < \delta.$$

Tal como em $\mathbb{R}$, as duas definições anteriores são equivalentes. Neste caso, apenas há que ter em conta que, sendo o espaço de chegada $\mathbb{R}^p$, a distância não é o módulo mas sim a norma em $\mathbb{R}^p$.

Nota 29 *Utilizando a definição de Heine, por exemplo, é fácil concluir que, tal como no caso dos limites de sucessões, o limite de uma função* $\mathbf{f} : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^p$, $\mathbf{f} = (f_1, f_2, \ldots, f_p)$, *calcula-se coordenada a coordenada,*

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f} = (\lim_{\mathbf{x}\to\mathbf{a}} f_1, \lim_{\mathbf{x}\to\mathbf{a}} f_2, \ldots, \lim_{\mathbf{x}\to\mathbf{a}} f_p).$$

Temos pois que o limite de $\mathbf{f}(\mathbf{x})$ *quando* $\mathbf{x}$ *tende para* $\mathbf{a}$ *é* $\mathbf{b} = (b_1, b_2, \ldots, b_p)$ *se e só se*

$$\lim_{\mathbf{x}\to\mathbf{a}} f_1(\mathbf{x}) = b_1, \quad \lim_{\mathbf{x}\to\mathbf{a}} f_2(\mathbf{x}) = b_2, \quad \ldots, \quad \lim_{\mathbf{x}\to\mathbf{a}} f_p(\mathbf{x}) = b_p.$$

Da definição δ, ε, *deduz-se de imediato que*

$$\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{b} \text{ se e só se } \lim_{\mathbf{x}\to\mathbf{a}} \|\mathbf{f}(\mathbf{x}) - \mathbf{b}\| = 0.$$

2.5 Continuidade

Definição 30 *Uma **função*** $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^p$ *é **contínua** em* $\mathbf{a} \in \mathcal{D}$ *se* $\lim_{\mathbf{x}\to\mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a})$.

A aplicação $\mathbf{f}$ *é contínua num subconjunto* $\mathcal{A}$ *de* $\mathbb{R}^m$ *se for contínua em todos os pontos de* $\mathcal{A}$.

Proposição 31

1. *Uma soma* $\mathbf{f} + \mathbf{g}$ *de funções contínuas em* $\mathbf{a}$ *é contínua em* $\mathbf{a}$.

2. *Se* $f : \mathbb{R}^m \to \mathbb{R}$ *é contínua em* $\mathbf{a}$ *e se* $f(\mathbf{a}) \neq 0$, *então* $1/f$ *é contínua em* $\mathbf{a}$.

3. *Se* $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}^p$ *é contínua em* $\mathbf{a}$ *e* $\mathbf{g} : \mathbb{R}^p \to \mathbb{R}^q$ *é contínua em* $\mathbf{b} = \mathbf{f}(\mathbf{a})$, *então a função composta* $\mathbf{g} \circ \mathbf{f}$ *é contínua em* $\mathbf{a}$.

4. *Toda a função polinomial de* $\mathbb{R}^m$ *em* $\mathbb{R}$ *é contínua.*

Teorema 32 *(Weierstraß) Se* f *for uma função real contínua, definida num subconjunto fechado e limitado* $\mathcal{A}$ *de* $\mathbb{R}^m$, *então* f *tem máximo e mínimo em* $\mathcal{A}$.

Recorde-se que uma função é **limitada** num subconjunto $\mathcal{A}$ se existem constantes m e M tais que, para todo o $\mathbf{x}$ de $\mathcal{A}$, se tem

$$m \leq f(\mathbf{x}) \leq M.$$

Dem Suponhamos que f não é limitada superiormente. Então, existe uma sucessão de pontos $\mathbf{x}_n$ em $\mathcal{A}$ com $f(\mathbf{x}_n) > n$. Como a sucessão $\mathbf{x}_n$ está contida num subconjunto fechado e limitado, admite uma subsucessão $\mathbf{y}_n$ que converge para um ponto $\mathbf{y}$ de $\mathcal{A}$. Como $\mathbf{y}_n = \mathbf{x}_{g(n)}$ com $g(n) \geq n$, tem-se que $f(\mathbf{y}_n) > n$ e $\lim f(\mathbf{y}_n) = +\infty$, sendo $\mathbf{y}_n \to \mathbf{y}$, o que é impossível pois f é uma função real contínua e, portanto $\lim f(\mathbf{y}_n) = f(\mathbf{y})$.

Podemos, então, supor que f é limitada superiormente. Chamemos M ao supremo do conjunto dos $f(\mathbf{x})$, $\mathbf{x} \in \mathcal{A}$. Pretendemos mostrar que existe um ponto $\mathbf{y}_0$ para o qual $M = f(\mathbf{y}_0)$. Pela definição de supremo, para todo o $n \geq 1$, existe um ponto $\mathbf{x}_n$ em $\mathcal{A}$ tal que $f(\mathbf{x}_n) > M - \frac{1}{n}$. A sucessão $\mathbf{x}_n$ admite uma subsucessão $\mathbf{y}_n$ que converge para um ponto $\mathbf{y}_0$ de $\mathcal{A}$. Como $M - \frac{1}{n} < f(\mathbf{y}_n) \leq M$, a sucessão $f(\mathbf{y}_n)$ tende para M. Agora, pela continuidade de f, temos

$$f(\mathbf{y}_0) = \lim f(\mathbf{y}_n) = M.$$

Procede-se da mesma forma para mostrar que f é minorada e que o limite inferior é atingido. $\square$

Capítulo 3

Cálculo Diferencial

O cálculo diferencial constitui uma ferramenta muitíssimo potente no estudo de uma classe muito alargada de funções, nomeadamente no que diz respeito ao seu comportamento e à existência de extremos, aspecto decisivo em questões de optimização. Como tal é de grande utilidade no estudo de modelos em todos os domínios, nomeadamente nos modelos económicos.

Neste capítulo apresentamos conceitos e propriedades relativos à diferenciabilidade de funções definidas em abertos de $\mathbb{R}^m$ e com valores em $\mathbb{R}$, dando ênfase especial às funções de $\mathbb{R}^2$ para $\mathbb{R}$. Estabelecem-se de modo adequado resultados análogos para o caso de funções de $\mathbb{R}^m$ para $\mathbb{R}^p$ uma vez que sabemos que estas não são mais do que sistemas de p funções de $\mathbb{R}^m$ para $\mathbb{R}$.

3.1 Derivadas parciais e derivadas direccionais

3.1.1 Derivadas de primeira ordem

Seja $\mathcal{D} \subset \mathbb{R}^m$ um aberto, $\mathbf{a} = (a_1, \ldots, a_m)$ um ponto de $\mathcal{D}$ e i um índice, $1 \leq i \leq m$. Dada uma função $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$, podemos considerar a função de uma variável que se obtém fixando todas as componentes de $\mathbf{a}$ excepto a de índice i, isto é,

$$x \to f(a_1, \ldots, a_{i-1}, x, a_{i+1,\ldots,}a_m)$$

definida num intervalo $]a_i - \varepsilon, a_i + \varepsilon[$. E podemos considerar a derivada desta função, caso exista.

Definição 33 *Seja $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ uma aplicação, $\mathcal{D}$ aberto e $\mathbf{a} \in \mathcal{D}$. A i-**ésima derivada parcial** de f em $\mathbf{a}$, que se representa por $\dfrac{\partial f}{\partial x_i}(\mathbf{a})$ ou*

$f'_{x_i}(\mathbf{a})$, *é, caso exista,*

$$f'_{x_i}(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}) = \lim_{t \to 0} \frac{f(a_1, \ldots, a_i + t, \ldots, a_m) - f(a_1, \ldots, a_m)}{t},$$

isto é,

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

onde $e_i = (0, 0, ..., 1, ..., 0)$ *é o vector da base canónica de* $\mathbb{R}^m$ *cuja i-ésima componente vale* 1.

Quando $m = 2$*, estas expressões assumem a seguinte forma:*

$$f'_x(a_1, a_2) = \frac{\partial f}{\partial x}(a_1, a_2) = \lim_{t \to 0} \frac{f(a_1 + t, a_2) - f(a_1, a_2)}{t},$$

$$f'_y(a_1, a_2) = \frac{\partial f}{\partial y}(a_1, a_2) = \lim_{t \to 0} \frac{f(a_1, a_2 + t) - f(a_1, a_2)}{t}.$$

Nota 34 *Calcular a i-ésima derivada parcial equivale, portanto, a considerar as variáveis diferentes de* x_i *como constantes e a derivar a função como uma função de uma única variável real* x_i.

Exemplo 35 *Dada a função*

$$f(x, y) = \begin{cases} x^2 \dfrac{x^2 - y^2}{x^2 + y^2}, & se\ x^2 + y^2 \neq 0 \\ \\ 0, & se\ x = y = 0, \end{cases}$$

tem-se

$$\begin{aligned} \frac{\partial f}{\partial x}(0,0) &= \lim_{t \to 0} \frac{f(0 + t, 0) - f(0,0)}{t} \\ &= \lim_{t \to 0} \frac{(0 + t)^2 \frac{(0+t)^2 - 0^2}{(0+t)^2 + 0^2} - 0}{t} \\ &= \lim_{t \to 0} \frac{t^2 \frac{t^2}{t^2}}{t} = \lim_{t \to 0} t = 0, \end{aligned}$$

$$
\begin{aligned}
\frac{\partial f}{\partial y}(0,0) &= \lim_{t\to 0}\frac{f(0,0+t)-f(0,0)}{t}\\
&= \lim_{t\to 0}\frac{0^2\frac{0^2-(0+t)^2}{0^2+(0+t)^2}-0}{t}\\
&= \lim_{t\to 0}\frac{0}{t}=\lim_{t\to 0}0=0,
\end{aligned}
$$

isto é, existem as derivadas parciais da função f no ponto $(0,0)$.

Definição 36 *Seja $f:\mathcal{D}\subset\mathbb{R}^m\to\mathbb{R}$ uma aplicação definida num aberto $\mathcal{D}$ e $\mathbf{a}\in\mathcal{D}$. Se $\mathbf{v}$ é um vector de norma **1** de $\mathbb{R}^m$, a **derivada direccional** de f no ponto $\mathbf{a}$ e na direcção $\mathbf{v}$, $\partial_{\mathbf{v}}f(\mathbf{a})$ é, caso exista, o limite*

$$
\begin{aligned}
\partial_{\mathbf{v}}f(\mathbf{a}) &= \lim_{t\to 0}\frac{f(\mathbf{a}+t\mathbf{v})-f(\mathbf{a})}{t}\\
&= \lim_{t\to 0}\frac{f(a_1+tv_1,\ldots,a_i+tv_i,\ldots,a_m+tv_m)-f(a_1,\ldots,a_m)}{t},
\end{aligned}
$$

onde $\mathbf{a}=(a_1,\ldots,a_m)$ e $\mathbf{v}=(v_1,\ldots,v_m)$.

Quando $m=2$, podemos dar-lhe a interpretação geométrica seguinte. Cortemos o gráfico de f pelo plano vertical π que contém a recta r, passa pelo ponto $\mathbf{a}$ e tem como vector director $\mathbf{v}$.

Obtemos assim uma função de uma variável real definida pela equação $g(t)=f(\mathbf{a}+t\mathbf{v})$.

O número $g'(0)=\partial_{\mathbf{v}}f(\mathbf{a})$ é, portanto, o declive da curva secção do gráfico pelo plano π no ponto $(\mathbf{a},f(\mathbf{a}))$ e na direcção $\mathbf{v}$.

Notas e exemplos:

1. A i-ésima derivada parcial não é mais que a derivada direccional na direcção do i-ésimo vector da base $\mathbf{e}_i$.

2. Calcular a i-ésima derivada parcial equivale, portanto, a considerar as variáveis diferentes de x_i como constantes e a derivar a função como uma função de uma única variável real x_i.

3. A existência de derivadas parciais num ponto não implica a continuidade de f nesse ponto, nem a existência de outras derivadas direccionais. Basta considerar a função $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ definida por

$$f(x,y) = \begin{cases} 1, & \text{se } xy = 0 \\ \\ 0, & \text{se } xy \neq 0. \end{cases}$$

É evidente que $f(0,0) = 1$ e que as derivadas parciais são nulas em $(0,0)$ porque a função é constante nos eixos. No entanto, a função não é contínua em $(0,0)$, porque o limite da função segundo a recta $y = x$ quando x tende para 0 é 0 e não 1. As derivadas direccionais, à excepção das derivadas parciais, não existem.

4. Pode acontecer que todas as derivadas direccionais existam num ponto e que, apesar disso, a função não seja contínua nesse ponto. Consideremos, por exemplo, a função

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4}, & \text{se } (x,y) \neq (0,0) \\ \\ 0, & \text{caso contrário.} \end{cases}$$

Esta função não é contínua na origem (considere-se a curva $x = y^2$), mas todas as derivadas direccionais existem neste ponto:

$$\partial_{(\alpha,\beta)} f(0,0) = \lim_{t\to 0} \frac{f(t\alpha,t\beta) - f(0,0)}{t} = \lim_{t\to 0} \frac{\frac{t^3\alpha\beta^2}{t^2(\alpha^2+t^2\beta^4)}}{t} = \begin{cases} \frac{\beta^2}{\alpha} & \text{se } \alpha \neq 0 \\ 0 & \text{se } \alpha = 0. \end{cases}$$

3.1.2 Derivadas parciais e um modelo económico

Consideremos, tal como no capítulo anterior, o modelo de produção $P(K,L)$ de um certo produto onde, para um *input* de capital K e de trabalho L, são obtidas P unidades desse produto. Então $\dfrac{\partial P}{\partial K} = P'_K$ é a taxa de variação de produção (*output*) P relativamente a K quando L se mantém constante. É designado por **produto marginal do capital.** Analogamente, $\dfrac{\partial P}{\partial L} = P'_L$ é a taxa de variação de produção (*output*) P relativamente a L quando K se mantém constante. Designa-se por **produto marginal do trabalho**.

3.2 Funções diferenciáveis

Seja $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ definida em $\mathcal{D}$ aberto e $\mathbf{a} \in \mathcal{D}$. Como $\mathcal{D}$ é aberto, existe uma $B_\varepsilon(\mathbf{a})$ contida em $\mathcal{D}$. Tomando um $\mathbf{h}$ tal que $\|\mathbf{h}\| < \varepsilon$, tem-se evidentemente $\mathbf{a}+\mathbf{h} \in B_\varepsilon(\mathbf{a})$. A função $f(\mathbf{x})$ diz-se **diferenciável** em $\mathbf{x} = \mathbf{a}$ se e só se existe uma aplicação linear $D_1 : \mathbb{R}^m \to \mathbb{R}$ tal que, para todos os $\mathbf{h}$ com $\|\mathbf{h}\| < \varepsilon$,

$$f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) = D_1(\mathbf{h}) + \varepsilon(\mathbf{h}) \cdot \|\mathbf{h}\|$$

em que $\lim_{\mathbf{h}\to\mathbf{0}} \varepsilon(\mathbf{h}) = 0$.

Notas e comentários:

1. Note-se que uma função diferenciável num ponto deve sempre ser definida numa vizinhança desse ponto.

2. Existem formulações equivalentes da definição. Pondo $\mathbf{x} = \mathbf{a}+\mathbf{h}$, obtém-se, por exemplo:

(Def′). A aplicação f é diferenciável em $\mathbf{a}$ se existe uma aplicação linear $D_1 : \mathbb{R}^m \to \mathbb{R}$ tal que

$$f(\mathbf{x}) - f(\mathbf{a}) = D_1(\mathbf{x}-\mathbf{a}) + \varepsilon(\mathbf{x}-\mathbf{a})\|\mathbf{x}-\mathbf{a}\|$$

onde $\lim_{\mathbf{x}\to\mathbf{a}} \varepsilon(\mathbf{x}-\mathbf{a}) = 0$.

3. Ou ainda: a aplicação f é diferenciável no ponto $\mathbf{a}$ se existe uma aplicação linear $D_1 : \mathbb{R}^m \to \mathbb{R}$ tal que

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - D_1(\mathbf{h})}{\|\mathbf{h}\|} = 0.$$

Teorema 37 *Se f for diferenciável em $\mathbf{a}$, com uma aplicação linear associada L, então todas as derivadas direccionais em $\mathbf{a}$ existem e:*

$$L(\mathbf{u}) = \partial_{\mathbf{u}} f(\mathbf{a}).$$

Dem. Seja $\mathbf{u}$ um vector de norma $\mathbf{1}$; então, por definição,

$$\partial_{\mathbf{u}} f(\mathbf{a}) = \lim_{t\to 0} \frac{f(\mathbf{a}+t\mathbf{u}) - f(\mathbf{a})}{t} = \lim_{t\to 0} \frac{L(t\mathbf{u}) + \varepsilon(t\mathbf{u})\|t\mathbf{u}\|}{t}$$

Como L é linear, $L(t\mathbf{u}) = tL(\mathbf{u})$, de onde resulta

$$\lim_{t\to 0} \frac{L(t\mathbf{u}) + |t|\varepsilon(t\mathbf{u})}{t} = \lim_{t\to 0} \frac{tL(\mathbf{u}) + |t|\varepsilon(t\mathbf{u})}{t}$$

$$= \lim_{t\to 0} L(\mathbf{u}) + \lim_{t\to 0} \frac{|t|\varepsilon(t\mathbf{u})}{t} = L(\mathbf{u}).\square$$

Do teorema anterior deduz-se que dado $\mathbf{x} = \mathbf{a}$, a aplicação linear D_1 introduzida na definição de função diferenciável em $\mathbf{a}$ é única. Tem-se então a definição

Definição 38 *A aplicação linear D_1 acima introduzida designa-se por* ***diferencial de f em*** *a (ou diferencial de primeira ordem de f em $\mathbf{a}$) e representa-se por $Df(\mathbf{a})$ ou $(df)(\mathbf{a})$. Tem a expressão*

$$\boxed{Df(\mathbf{a})(h_1, h_2, \ldots, h_m) = h_1 \frac{\partial f}{\partial x_1}(\mathbf{a}) + h_2 \frac{\partial f}{\partial x_2}(\mathbf{a}) + \ldots + h_m \frac{\partial f}{\partial x_m}(\mathbf{a}).}$$

Nota 39 *Quando $m = 2$, é frequente escrever-se (a, b) em vez de (a_1, a_2) e (h, k) em vez de (h_1, h_2). Assim, tendo em conta o que foi dito acima, $f(x, y)$ é diferenciável em (a, b) se*

$$f(a+h, b+k) - f(a, b) = \frac{\partial f}{\partial x}(a, b)h + \frac{\partial f}{\partial y}(a, b)k + \epsilon(h, k).\, \|(h, k)\|,$$

em que

$$\lim_{(h,k)\to(0,0)} \epsilon(h, k) = 0,$$

ou, o que é equivalente,

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

Exemplo 40 *Retomemos a função*

$$f(x, y) = \begin{cases} x^2 \dfrac{x^2 - y^2}{x^2 + y^2}, & \text{se } x^2 + y^2 \neq 0 \\ \\ 0, & \text{se } x = y = 0, \end{cases}$$

considerada num exemplo anterior. Tendo em conta a última nota, averiguemos se $f(x, y)$ é diferenciável no ponto $(0, 0)$. Ora, já sabemos de cálculos anteriormente feitos, que $\dfrac{\partial f}{\partial x}(0, 0) = 0$ e $\dfrac{\partial f}{\partial y}(0, 0) = 0$. Logo,

$$\begin{aligned}
|\epsilon(h, k)| &= \left| \frac{f(0+h, 0+k) - f(0, 0) - \frac{\partial f}{\partial x}(0, 0)h - \frac{\partial f}{\partial y}(0, 0)k}{\sqrt{h^2 + k^2}} \right| \\
&= \left| \frac{f(h, k) - f(0, 0) - 0.h - 0.k}{\sqrt{h^2 + k^2}} \right| = \left| \frac{h^2 \frac{h^2 - k^2}{h^2 + k^2}}{\sqrt{h^2 + k^2}} \right| \\
&= \left| \frac{h^2 (h^2 - k^2)}{(h^2 + k^2)\sqrt{h^2 + k^2}} \right| \leq \left| \frac{(h^2 + k^2)(h^2 - k^2)}{(h^2 + k^2)\sqrt{h^2 + k^2}} \right| \\
&\leq \left| \frac{h^2 - k^2}{\sqrt{h^2 + k^2}} \right| \leq \frac{h^2 + k^2}{\sqrt{h^2 + k^2}} = \sqrt{h^2 + k^2},
\end{aligned}$$

donde

$$|\epsilon(h,k)| \leq \sqrt{h^2+k^2}.$$

Assim, passando ao limite,

$$0 \leq \lim_{(h,k)\to(0,0)} |\epsilon(h,k)| \leq \lim_{(h,k)\to(0,0)} \sqrt{h^2+k^2} = 0,$$

e portanto, $\lim_{(h,k)\to(0,0)} |\epsilon(h,k)| = 0$, *ou seja* $\lim_{(h,k)\to(0,0)} \epsilon(h,k) = 0$, *o que permite concluir que a função* f *é diferenciável em* $(0,0)$.

3.2.1 Gradiente e matriz jacobiana

Vimos que o diferencial $Df(\mathbf{a})$ é uma aplicação linear de $\mathbb{R}^m$ em $\mathbb{R}$. Para $\mathbf{h} = (h_1, h_2, \ldots, h_m)$, tem-se

$$\begin{aligned} Df(\mathbf{a})(\mathbf{h}) &= \sum_{i=1}^{m} h_i \frac{\partial f}{\partial x_i}(\mathbf{a}) \\ &= \left[\frac{\partial f}{\partial x_1}(\mathbf{a}) \; \frac{\partial f}{\partial x_2}(\mathbf{a}) \; \ldots \; \frac{\partial f}{\partial x_m}(\mathbf{a}) \right] \begin{bmatrix} h_1 \\ h_2 \\ \ldots \\ h_m \end{bmatrix}, \end{aligned}$$

que não é mais do que o produto interno do vector de $\mathbb{R}^m$ chamado **gradiente de f em** a, e representado por $\nabla f(\mathbf{a})$,

$$\nabla f(\mathbf{a}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{a}), \frac{\partial f}{\partial x_2}(\mathbf{a}), \ldots, \frac{\partial f}{\partial x_m}(\mathbf{a}) \right),$$

pelo vector

$$\mathbf{h} = (h_1, h_2, \ldots, h_m).$$

Assim, como notação,

$$Df(\mathbf{a})(\mathbf{h}) = \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle = \nabla f(\mathbf{a}).\mathbf{h} = \mathbf{h}.\nabla f(\mathbf{a}),$$

onde as três últimas expressões têm como significado o produto interno do gradiente de f pelo vector $\mathbf{h}$ e serão usadas indiferentemente.

Nota 41 *Vimos que, se* f *for uma função diferenciável em* $\mathbf{a}$, *então*

(i) existem todas as derivadas direccionais em $\mathbf{a}$, $\partial_{\mathbf{v}} f(\mathbf{a}), \forall \mathbf{v} \in \mathbb{R}^m$,

(ii) a aplicação linear D_1 *associada é o diferencial de primeira ordem, isto é,*

$$D_1(\mathbf{h}) = Df(\mathbf{a})(\mathbf{h}) = \nabla f(\mathbf{a}).\mathbf{h},$$

(iii) $\partial_{\mathbf{v}} f(\mathbf{a}) = D_1(\mathbf{v})$.

Estas condições fornecem-nos um modo alternativo de calcular $\partial_{\mathbf{v}} f(\mathbf{a})$ *(apenas para funções diferenciáveis)*

$$\partial_{\mathbf{v}} f(\mathbf{a}) = \mathbf{v}.\nabla f(\mathbf{a}).$$

Notação 42 *O vector gradiente poderá também ser representado por*

$$gradf(\mathbf{a})$$

ou ainda, quando se pretender reforçar a semelhança com alguma propriedade das funções reais de uma só variável real, por $f'(\mathbf{a})$*, isto é,*

$$\nabla f(\mathbf{a}) = grad\ f(\mathbf{a}) = f'(\mathbf{a}).$$

Consideremos agora uma função $\mathbf{f}$ de $\mathbb{R}^m$ para $\mathbb{R}^p$. O que foi dito para funções de $\mathbb{R}^m$ em $\mathbb{R}$ é válido para funções de $\mathbb{R}^m$ em $\mathbb{R}^p$ e pode ser estabelecido de forma adequada, usando o facto de estas serem sistemas de p funções de $\mathbb{R}^m$ em $\mathbb{R}$. Vejamos o que se passa neste caso com o diferencial. Designemos por $f_1, f_2, \ldots, f_p$ as componentes de $\mathbf{f}$. Para $\mathbf{h} = (h_1, h_2, \ldots, h_m)$, tem-se

$$D\mathbf{f}(\mathbf{a})(\mathbf{h}) = \sum_{i=1}^{m} h_i \frac{\partial \mathbf{f}}{\partial x_i}(\mathbf{a}) = \begin{bmatrix} \displaystyle\sum_{i=1}^{m} h_i \frac{\partial f_1}{\partial x_i}(\mathbf{a}) \\ \displaystyle\sum_{i=1}^{m} h_i \frac{\partial f_2}{\partial x_i}(\mathbf{a}) \\ \cdots \\ \displaystyle\sum_{i=1}^{m} h_i \frac{\partial f_p}{\partial x_i}(\mathbf{a}) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_m}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_m}(\mathbf{a}) \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{a}) & \frac{\partial f_p}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_p}{\partial x_m}(\mathbf{a}) \end{bmatrix} . \begin{bmatrix} h_1 \\ h_2 \\ \dots \\ h_m \end{bmatrix}.$$

À matriz das derivadas parciais chama-se **Matriz Jacobiana** de $\mathbf{f}$ no ponto $\mathbf{a}$ e representa-se por $J_{\mathbf{f}}(\mathbf{a})$,

$$J_{\mathbf{f}}(\mathbf{a}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \frac{\partial f_1}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_m}(\mathbf{a}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{a}) & \frac{\partial f_2}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_2}{\partial x_m}(\mathbf{a}) \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{a}) & \frac{\partial f_p}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f_p}{\partial x_m}(\mathbf{a}) \end{bmatrix}.$$

Assim,

$$D\mathbf{f}(\mathbf{a})(\mathbf{h}) = J_{\mathbf{f}}(\mathbf{a}).\mathbf{h} \tag{3.1}$$

Quando $p = 1$, a matriz Jacobiana reduz-se a uma única linha constituída pelas coordenadas do **gradiente** de f em $\mathbf{a}$, que vimos em cima,

$$\begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{a}) & \frac{\partial f}{\partial x_2}(\mathbf{a}) & \dots & \frac{\partial f}{\partial x_m}(\mathbf{a}) \end{bmatrix}.$$

Notação 43 *Mais uma vez, quando se pretender enfatizar a semelhança entre certas propriedades da matriz jacobiana* $J_{\mathbf{f}}(\mathbf{a})$ *e as correspondentes propriedades da derivada de uma função real de variável real, poder-se-á escrever* $\mathbf{f}'(\mathbf{a})$ *em vez de* $J_{\mathbf{f}}(\mathbf{a})$. *Ter-se-á então a partir de (3.1)*

$$D\mathbf{f}(\mathbf{a})(\mathbf{h}) = J_{\mathbf{f}}(\mathbf{a}).\mathbf{h} = \mathbf{f}'(\mathbf{a})(\mathbf{h}). \tag{3.2}$$

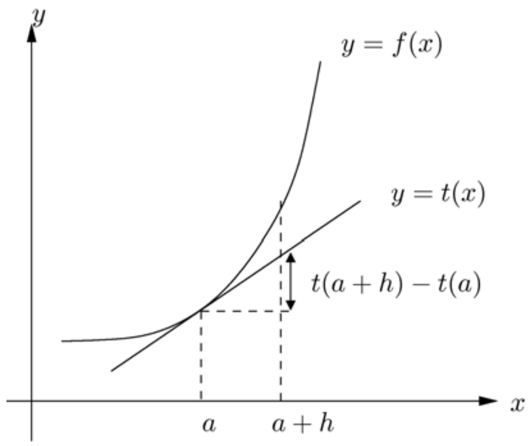


Figura 3.1:

3.3 Interpretações geométricas

3.3.1 Funções de $\mathbb{R}$ em $\mathbb{R}$

Neste caso, $\mathbf{a} = a$ e $\mathbf{h} = h$ e o diferencial é a aplicação linear

$$Df(a)(h) = f'(a).h.$$

O seu valor no ponto h é a diferença de altura entre $t(a)$ e $t(a+h)$, onde $y = t(x)$ designa a recta tangente à curva $y = f(x)$ no ponto de coordenadas $(a, f(a))$ (ver figura 3.1).

3.3.2 Funções de $\mathbb{R}^2$ em $\mathbb{R}$

Consideremos o gráfico da função f,

$$Graf(f) = \left\{(x, y, z) \in \mathbb{R}^3 : z = f(x, y), (x, y) \in D_f\right\}.$$

Se f é diferenciável no ponto (a, b), a equação

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

é a equação de um plano que passa pelo ponto $(a, b, f(a, b))$. Como esse plano, $z = \alpha x + \beta y + \gamma$, é o único plano que passa pelo ponto $(a, b, f(a, b))$ e que verifica a condição

$$\lim_{(x,y)\to(a,b)} \frac{f(x, y) - f(\alpha x + \beta y + \gamma)}{\|(x, y) - (a, b)\|} = 0,$$

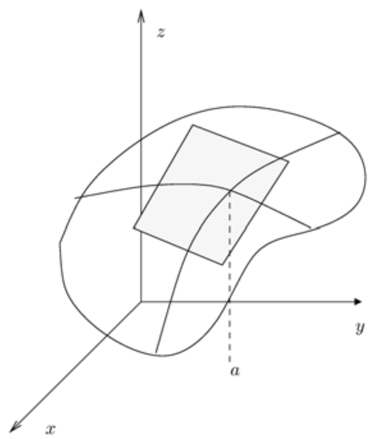


Figura 3.2: Plano tangente à superfície no ponto assinalado.

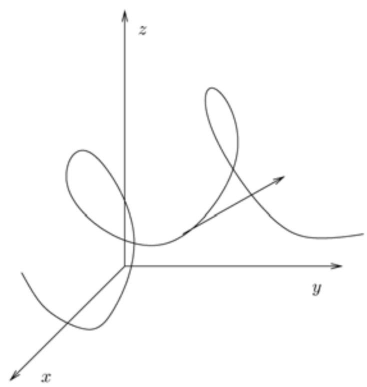


Figura 3.3: Vector tangente à curva no ponto indicado.

trata-se do plano que se “cola” melhor à superfície $Graf(f)$ no ponto $(a, b, f(a, b))$. Chamamos-lhe o **plano tangente** à superfície no ponto $(a, b, f(a, b))$ (figura 3.2).

3.3.3 Funções de $\mathbb{R}$ em $\mathbb{R}^p$

Uma função $\mathbf{f}$ de $\mathbb{R}$ em $\mathbb{R}^p$ representa-se pela sua imagem, que é uma curva em $\mathbb{R}^p$. O vector tangente à curva $\mathbf{f}(t) = (f_1(t), f_2(t), \ldots, f_p(t))$, no ponto $t = t_0$, é o vector $\mathbf{f}'(t_0) = (f_1'(t_0), f_2'(t_0), \ldots, f_p'(t_0))$ (figura 3.3).

O vector tangente a $\mathbf{f}(t)$ em $t = t_0$ é um vector director da recta tangente à curva no ponto $\mathbf{f}(t_0)$.

3.3.4 Funções de $\mathbb{R}^m$ em $\mathbb{R}^p$, $m \geq 2$, $p \geq 2$

Convém visualisar simultaneamente o espaço de partida $\mathbb{R}^m$ e o espaço de chegada $\mathbb{R}^p$. Se fixarmos um ponto $\mathbf{a}$ de $\mathbb{R}^m$, podemos associar à aplicação $\mathbf{f}$ uma aplicação linear $D\mathbf{f}(\mathbf{a})$ entre os mesmos espaços. Essa aplicação linear $D\mathbf{f}(\mathbf{a})$ é a "melhor aproximação" de primeira ordem de $\mathbf{f}$ na vizinhança do ponto $\mathbf{a}$,

$$\mathbf{f}(\mathbf{a}+\mathbf{h}) \simeq \mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{h}).$$

3.4 Derivação da função composta

Como é sabido, o teorema de derivação da função composta, em dimensão 1, garante que, sob condições adequadas,

$$(f \circ g)'(x) = f'(g(x)).g'(x).$$

Apresentamos em seguida a generalização deste resultado para o caso de funções reais de variável vectorial. Usaremos a identificação que mencionámos anteriormente em **??**. Trata-se de um teorema de grande utilidade quer teórica quer prática. Para além dos exemplos calculatórios apresentados nesta secção, veremos que será aplicado ainda neste capítulo na demonstração do Teorema de Euler para funções homogéneas e no Teorema de Lagrange.

Teorema 44 *Sejam $m, n, p \in \mathbb{N}$ e consideremos um subconjunto $\mathcal{D}$ de $\mathbb{R}^m$ e um subconjunto $\mathcal{E}$ de $\mathbb{R}^p$. Consideremos também uma função $\mathbf{g} : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}^p$ tal que $g(\mathcal{D}) \subset \mathcal{E}$ e uma função $\mathbf{f} : \mathcal{E} \subset \mathbb{R}^p \to \mathbb{R}^n$. Se $\mathbf{g}$ for diferenciável num ponto $\mathbf{a} \in \mathbb{R}^m$ e $\mathbf{f}$ for diferenciável no ponto $\mathbf{b} = \mathbf{g}(\mathbf{a})$, $\mathbf{b} \in \mathbb{R}^p$, então $\mathbf{f} \circ \mathbf{g}$ é diferenciável em $\mathbf{a}$ e é satisfeita a igualdade*

$$(\mathbf{f} \circ \mathbf{g})'(\mathbf{a}) = \mathbf{f}'(\mathbf{g}(\mathbf{a})).\mathbf{g}'(\mathbf{a}).$$

Dem Apresentaremos apenas um esboço da demonstração. Sem perda de generalidade, admitiremos que nas expressões seguintes os argumentos das funções pertencem aos domínios respectivos. Sendo $\mathbf{g}$ diferenciável no ponto $\mathbf{a}$,

$$\mathbf{g}(\mathbf{a}+\mathbf{h}) = \mathbf{g}(\mathbf{a}) + \mathbf{g}'(\mathbf{a})(\mathbf{h}) + \mathbf{r_g}(\mathbf{h})$$

em que $\mathbf{r_g}(\mathbf{h})$ satisfaz $\lim\limits_{\mathbf{h}\to\mathbf{0}} \dfrac{\mathbf{r_g}(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0}$. Analogamente, como $\mathbf{f}$ é diferenciável em $\mathbf{b} = \mathbf{g}(\mathbf{a})$ tem-se

$$\mathbf{f}(\mathbf{b}+\mathbf{k}) = \mathbf{f}(\mathbf{b}) + \mathbf{f}'(\mathbf{b})(\mathbf{k}) + \mathbf{r_f}(\mathbf{k}),$$

em que $\mathbf{r_f}(\mathbf{k})$ satisfaz $\lim\limits_{\mathbf{k}\to\mathbf{0}} \dfrac{\mathbf{r_f}(\mathbf{k})}{\|\mathbf{k}\|} = \mathbf{0}$. Fazendo $\mathbf{b} = \mathbf{g}(\mathbf{a})$ e $\mathbf{k} = \mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})$, obtém-se a expressão

$$(\mathbf{f}\circ\mathbf{g})(\mathbf{a}+\mathbf{h}) = (\mathbf{f}\circ\mathbf{g})(\mathbf{a}) + \mathbf{f}'(\mathbf{g}(\mathbf{a}))(\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})) + \mathbf{r_f}(\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a}))$$

e, uma vez que $\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a}) = \mathbf{g}'(\mathbf{a})(\mathbf{h}) + \mathbf{r_g}(\mathbf{h})$, tem-se

$$\begin{aligned}(\mathbf{f}\circ\mathbf{g})(\mathbf{a}+\mathbf{h}) &= (\mathbf{f}\circ\mathbf{g})(\mathbf{a}) + \mathbf{f}'(\mathbf{g}(\mathbf{a})).\mathbf{g}'(\mathbf{a})(\mathbf{h}) + \mathbf{f}'(\mathbf{g}(\mathbf{a}))\mathbf{r_g}(\mathbf{h}) \\ &\quad + \mathbf{r_f}(\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a})).\end{aligned}$$

Prova-se (de forma fácil mas não imediata) que

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{f}'(\mathbf{g}(\mathbf{a}))\mathbf{r_g}(\mathbf{h})}{\|\mathbf{h}\|} = \mathbf{0} \quad \text{e} \quad \lim_{\mathbf{h}\to\mathbf{0}} \frac{\mathbf{r_f}(\mathbf{g}(\mathbf{a}+\mathbf{h}) - \mathbf{g}(\mathbf{a}))}{\|\mathbf{h}\|} = \mathbf{0},$$

o que conclui a demonstração. □

Usando as matrizes jacobianas respectivas, a regra de derivação da função composta pode ser escrita em termos matriciais do seguinte modo: Seja $\mathbf{h} = \mathbf{f}\circ\mathbf{g}$ e sejam $J_\mathbf{f}(\mathbf{b}), J_\mathbf{g}(\mathbf{a})$ e $J_\mathbf{h}(\mathbf{a})$ as matrizes jacobianas de $\mathbf{f}$, $\mathbf{g}$ e $\mathbf{h}$, respectivamente, nos pontos indicados, isto é,

$$J_\mathbf{f}(\mathbf{b}) = \begin{bmatrix} \dfrac{\partial f_1}{\partial y_1}(\mathbf{b}) & \dfrac{\partial f_1}{\partial y_2}(\mathbf{b}) & \dots & \dfrac{\partial f_1}{\partial y_p}(\mathbf{b}) \\ \dfrac{\partial f_2}{\partial y_1}(\mathbf{b}) & \dfrac{\partial f_2}{\partial y_2}(\mathbf{b}) & \dots & \dfrac{\partial f_2}{\partial y_p}(\mathbf{b}) \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial f_n}{\partial y_1}(\mathbf{b}) & \dfrac{\partial f_n}{\partial y_2}(\mathbf{b}) & \dots & \dfrac{\partial f_n}{\partial y_p}(\mathbf{b}) \end{bmatrix}$$

$$J_\mathbf{g}(\mathbf{a}) = \begin{bmatrix} \dfrac{\partial g_1}{\partial x_1}(\mathbf{a}) & \dfrac{\partial g_1}{\partial x_2}(\mathbf{a}) & \dots & \dfrac{\partial g_1}{\partial x_m}(\mathbf{a}) \\ \dfrac{\partial g_2}{\partial x_1}(\mathbf{a}) & \dfrac{\partial g_2}{\partial x_2}(\mathbf{a}) & \dots & \dfrac{\partial g_2}{\partial x_m}(\mathbf{a}) \\ \dots & \dots & \dots & \dots \\ \dfrac{\partial g_p}{\partial x_1}(\mathbf{a}) & \dfrac{\partial g_p}{\partial x_2}(\mathbf{a}) & \dots & \dfrac{\partial g_p}{\partial x_m}(\mathbf{a}) \end{bmatrix}$$

$$J_{\mathbf{f}\circ\mathbf{g}}(\mathbf{a}) = \begin{bmatrix} \frac{\partial(\mathbf{f}\circ\mathbf{g})_1}{\partial x_1}(\mathbf{a}) & \frac{\partial(\mathbf{f}\circ\mathbf{g})_1}{\partial x_2}(\mathbf{a}) & \ldots & \frac{\partial(\mathbf{f}\circ\mathbf{g})_1}{\partial x_m}(\mathbf{a}) \\ \frac{\partial(\mathbf{f}\circ\mathbf{g})_2}{\partial x_1}(\mathbf{a}) & \frac{\partial(\mathbf{f}\circ\mathbf{g})_2}{\partial x_2}(\mathbf{a}) & \ldots & \frac{\partial(\mathbf{f}\circ\mathbf{g})_2}{\partial x_m}(\mathbf{a}) \\ \ldots & \ldots & \ldots & \ldots \\ \frac{\partial(\mathbf{f}\circ\mathbf{g})_n}{\partial x_1}(\mathbf{a}) & \frac{\partial(\mathbf{f}\circ\mathbf{g})_n}{\partial x_2}(\mathbf{a}) & \ldots & \frac{\partial(\mathbf{f}\circ\mathbf{g})_n}{\partial x_m}(\mathbf{a}) \end{bmatrix}$$

Então,

$$J_{\mathbf{f}\circ\mathbf{g}}(\mathbf{a}) = J_{\mathbf{f}}(\mathbf{g}(\mathbf{a})) \times J_{\mathbf{g}}(\mathbf{a}),$$

ou seja, a matriz jacobiana de $\mathbf{f}\circ\mathbf{g}$ no ponto $\mathbf{a}$ é o produto da matriz jacobiana de $\mathbf{f}$ em $\mathbf{b} = \mathbf{g}(\mathbf{a})$ pela matriz jacobiana de $\mathbf{g}$ em $\mathbf{a}$. Assim, cada entrada da matriz jacobiana correspondente à composição, isto é, cada elemento da matriz $J_{\mathbf{f}\circ\mathbf{g}}(\mathbf{a})$, com $i = 1, \ldots, n$ e $j = 1, \ldots, m$, é dado por

$$\frac{\partial(\mathbf{f}\circ\mathbf{g})_i}{\partial x_j}(\mathbf{a}) = \frac{\partial f_i}{\partial y_1}(\mathbf{b})\frac{\partial g_1}{\partial x_j}(\mathbf{a}) + \frac{\partial f_i}{\partial y_2}(\mathbf{b})\frac{\partial g_2}{\partial x_j}(\mathbf{a}) + \ldots + \frac{\partial f_i}{\partial y_p}(\mathbf{b})\frac{\partial g_p}{\partial x_j}(\mathbf{a}),$$

ou, com abuso de notação, mas de forma expressiva,

$$\frac{\partial z_i}{\partial x_j}(\mathbf{a}) = \frac{\partial z_i}{\partial y_1}(\mathbf{b})\frac{\partial y_1}{\partial x_j}(\mathbf{a}) + \frac{\partial z_i}{\partial y_2}(\mathbf{b})\frac{\partial y_2}{\partial x_j}(\mathbf{a}) + \ldots + \frac{\partial z_i}{\partial y_p}(\mathbf{b})\frac{\partial y_p}{\partial x_j}(\mathbf{a}),$$

em que $\dfrac{\partial(\mathbf{f}\circ\mathbf{g})_i}{\partial x_j}(\mathbf{a})$, $\dfrac{\partial f_i}{\partial y_k}(\mathbf{b})$ e $\dfrac{\partial g_k}{\partial x_j}(\mathbf{a})$ foram substituídas por $\dfrac{\partial z_i}{\partial x_j}(\mathbf{a})$, $\dfrac{\partial z_i}{\partial y_k}(\mathbf{b})$ e $\dfrac{\partial y_k}{\partial x_j}(\mathbf{a})$, respectivamente. O abuso de notação consiste em utilizar z_i com dois sentidos diferentes: no primeiro membro z_i diz respeito à componente i da função $f\circ g$, enquanto que no segundo membro diz respeito à componente i da função $f(y)$. Omitindo os pontos $\mathbf{a}$ e $\mathbf{b}$, desde que não haja perigo de confusão, tem-se a expressão sugestiva, também designada por **regra da cadeia**,

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial z_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial z_i}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \ldots + \frac{\partial z_i}{\partial y_p}\frac{\partial y_p}{\partial x_j}.$$

Suponhamos, por exemplo, que

$$\begin{array}{rcccl} \mathbf{g} & : & \mathbb{R} & \to & \mathbb{R}^2 \\ & & x & \to & (y_1, y_2) \end{array} \quad e \quad \begin{array}{rcccl} f & : & \mathbb{R}^2 & \to & \mathbb{R} \\ & & (y_1, y_2) & \to & z. \end{array}$$

Então,

$$\begin{aligned} f \circ g \ &: \ \mathbb{R} \to \mathbb{R} \\ & \quad x \to z \end{aligned}$$

e

$$\frac{dz}{dx} = \frac{\partial z}{\partial y_1}\frac{\partial y_1}{\partial x} + \frac{\partial z}{\partial y_2}\frac{\partial y_2}{\partial x}.$$

Na expressão $\frac{dz}{dx}$, a utilização do símbolo d em vez de ∂ diz respeito ao facto da composição ser apenas função de uma só variável, pelo que, neste caso, a derivada ser total e não apenas parcial.

Suponhamos agora que

$$\begin{array}{rcccl} \mathbf{g} & : & \mathbb{R}^2 & \to & \mathbb{R}^4 \\ & & (x_1, x_2) & \to & (y_1, y_2, y_3, y_4) \end{array} \quad e \quad \begin{array}{rcccl} \mathbf{f} & : & \mathbb{R}^4 & \to & \mathbb{R}^3 \\ & & (y_1, y_2, y_3, y_4) & \to & (z_1, z_2, z_3) \end{array}.$$

Então,

$$\begin{array}{rcccl} \mathbf{f} \circ \mathbf{g} & : & \mathbb{R}^2 & \to & \mathbb{R}^3 \\ & & (x_1, x_2) & \to & (z_1, z_2, z_3) \end{array}$$

e

$$\frac{\partial z_1}{\partial x_1} = \frac{\partial z_1}{\partial y_1}\frac{\partial y_1}{\partial x_1} + \frac{\partial z_1}{\partial y_2}\frac{\partial y_2}{\partial x_1} + \frac{\partial z_1}{\partial y_3}\frac{\partial y_3}{\partial x_1} + \frac{\partial z_1}{\partial y_4}\frac{\partial y_4}{\partial x_1}$$

$$\frac{\partial z_1}{\partial x_2} = \frac{\partial z_1}{\partial y_1}\frac{\partial y_1}{\partial x_2} + \frac{\partial z_1}{\partial y_2}\frac{\partial y_2}{\partial x_2} + \frac{\partial z_1}{\partial y_3}\frac{\partial y_3}{\partial x_2} + \frac{\partial z_1}{\partial y_4}\frac{\partial y_4}{\partial x_2}.$$

Obter-se-iam expressões análogas para $\frac{\partial z_2}{\partial x_j}$, bem como para $\frac{\partial z_3}{\partial x_j}$, substituindo nas expressões acima z_1 por z_2, ou z_3, respectivamente.

Exemplo 45 *Seja $z = x^2 \sin y$ em que $x = v \cos u$ e $y = vu$. Determinemos $\frac{\partial z}{\partial u}$ e $\frac{\partial z}{\partial v}$ num ponto genérico (a, b). Usando a derivada da função composta,*

$$\begin{aligned} \frac{\partial z}{\partial u}(a,b) &= \frac{\partial z}{\partial x}(x(a,b), y(a,b))\frac{\partial x}{\partial u}(a,b) + \frac{\partial z}{\partial y}(x(a,b), y(a,b))\frac{\partial y}{\partial u}(a,b) \\ &= 2x \sin y|_{(x(a,b),y(a,b))} \cdot (-v \sin u)|_{(a,b)} + x^2 \cos y\big|_{(x(a,b),y(a,b))} \cdot v|_{(a,b)} \end{aligned}$$

e como $x(a,b) = a\cos b$ *e* $y(a,b) = ab$*,*

$$\begin{aligned}\frac{\partial z}{\partial u}(a,b) &= 2a\cos b\sin(ab)(-b\sin a) + a^2\cos^2 b\cos(ab)b \\ &= -2ab\cos b\sin a\sin(ab) + a^2 b\cos^2 b\cos(ab).\end{aligned}$$

Em particular, se $(a,b) = (1,\pi)$*, tem-se*

$$\frac{\partial z}{\partial u}(1,\pi) = -2\pi\cos\pi\sin 1\sin\pi + \pi\cos^2\pi\cos\pi = -\pi.$$

3.5 Teoremas da Diferenciabilidade

Comecemos por apresentar o teorema dos acréscimos finitos ou do valor médio de Lagrange. Como se pode observar, a formulação apresentada constitui uma generalização para dimensão superior, do correspondente em dimensão 1. No que se segue, sendo $\mathbf{a},\mathbf{b},\mathbf{c} \in \mathbb{R}^m$, dizemos que $\mathbf{c} \in]\mathbf{a},\mathbf{b}[$ se $c_1 \in]a_1,b_1[,\ldots,c_m \in]a_m,b_m[$.

Teorema 46 *(Teorema do valor médio de Lagrange) Seja* $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ *uma função contínua e com derivadas parciais contínuas em* $\mathcal{D}$*, onde* $\mathcal{D}$ *é um conjunto aberto. Sejam* $\mathbf{a}$ *e* $\mathbf{b}$ *pontos de* $\mathcal{D}$ *tais que o segmento de recta* $[\mathbf{a},\mathbf{b}]$ *está contido em* $\mathcal{D}$*. Então existe um ponto* $\mathbf{c} \in]\mathbf{a},\mathbf{b}[$ *tal que*

$$f(\mathbf{b}) - f(\mathbf{a}) = f'(\mathbf{c}).(\mathbf{b}-\mathbf{a}),$$

(sendo o segundo membro igual ao produto interno do vector gradiente de f *em* $\mathbf{c}$ *pelo vector* $\mathbf{b}-\mathbf{a}$*).*

Dem Consideremos a função real de variável real definida no intervalo $[0,1]$ por

$$\varphi(t) = f(\mathbf{a} + t(\mathbf{b}-\mathbf{a})).$$

Esta função satisfaz as hipóteses do teorema do valor médio de Lagrange em dimensão 1. Logo, existe $\xi \in]0,1[$ tal que

$$\varphi(1) - \varphi(0) = \varphi'(\xi)$$

e como $\varphi'(t) = (\mathbf{b}-\mathbf{a})f'(\mathbf{a}+t(\mathbf{b}-\mathbf{a}))$,

$$f(\mathbf{b}) - f(\mathbf{a}) = \varphi(1) - \varphi(0) = \varphi'(\xi) = (\mathbf{b}-\mathbf{a})f'(\mathbf{a}+\xi(\mathbf{b}-\mathbf{a})).$$

Fazendo $\mathbf{c} = \mathbf{a} + \xi(\mathbf{b} - \mathbf{a})$, sai que

$$f(\mathbf{b}) - f(\mathbf{a}) = (\mathbf{b} - \mathbf{a}) f'(\mathbf{c}),$$

com $\mathbf{c} \in]\mathbf{a}, \mathbf{b}[$, o que termina a demonstração. $\square$

No início deste capítulo vimos que podem existir todas as derivadas direccionais de uma função f num ponto $\mathbf{a}$ e que, apesar disso, f pode não ser contínua em $\mathbf{a}$. Assim, como as derivadas parciais são um caso particular das direccionais, podemos também concluir que a existência de derivadas parciais de uma função f num ponto $\mathbf{a}$ não implica a continuidade de f em $\mathbf{a}$. Contudo, a diferenciabilidade de f implica a continuidade de f. Mas se existirem as derivadas parciais de f numa vizinhança de $\mathbf{a}$ e se elas forem contínuas em $\mathbf{a}$, podemos concluir que a aplicação f é então diferenciável em $\mathbf{a}$ e, portanto, contínua em $\mathbf{a}$. Nos dois teoremas seguintes, são demonstradas estas afirmações.

Teorema 47 *Se f é diferenciável em $\mathbf{a}$ então f é contínua em $\mathbf{a}$.*

Dem Tem-se

$$f(\mathbf{x}) = f(\mathbf{a}) + D_1(\mathbf{x} - \mathbf{a}) + \varepsilon(\mathbf{x} - \mathbf{a})\|\mathbf{x} - \mathbf{a}\|.$$

Como D_1 é linear, D_1 é contínua e portanto $\lim\limits_{\mathbf{x}\to\mathbf{a}} D_1(\mathbf{x} - \mathbf{a}) = D_1(\mathbf{0}) = 0$. Como $\lim\limits_{\mathbf{x}\to\mathbf{a}} \varepsilon(\mathbf{x} - \mathbf{a}) = 0$, sai que $\lim\limits_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = f(\mathbf{a})$ e o resultado fica demonstrado. $\square$

Teorema 48 *Seja f uma função de $\mathbb{R}^m$ em $\mathbb{R}$. Se as derivadas parciais existem numa vizinhança de $\mathbf{a}$ e são contínuas em $\mathbf{a}$, então f é diferenciável em $\mathbf{a}$.*

Dem Precisamos de mostrar que, para $\mathbf{h} = (h_1, h_2, \ldots, h_m)$, se tem

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a}) - \sum_{i=1}^{m} \dfrac{\partial f}{\partial x_i}(\mathbf{a}) h_i}{\|\mathbf{h}\|} = 0.$$

Decomponhamos $f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})$ do seguinte modo

$$\begin{aligned}
f(\mathbf{a}+\mathbf{h})-f(\mathbf{a}) &= f(a_1+h_1,a_2+h_2,\ldots,a_m+h_m)\\
&\quad -f(a_1,a_2,\ldots,a_m)\\
&= f(a_1+h_1,a_2+h_2,\ldots,a_m+h_m)\\
&\quad -f(a_1+h_1,a_2+h_2,\ldots,a_{m-1}+h_{m-1},a_m)\\
&\quad +f(a_1+h_1,a_2+h_2,\ldots,a_{m-1}+h_{m-1},a_m)\\
&\quad -f(a_1+h_1,\ldots,a_{m-2}+h_{m-2},a_{m-1},a_m)\\
&\quad +\ldots\\
&\quad +f(a_1+h_1,a_2,\ldots,a_m)\\
&\quad -f(a_1,a_2,\ldots,a_m).
\end{aligned}$$

Façamos

$$\begin{aligned}
g_i(t) &= f(a_1+h_1,\ldots,a_{i-1}+h_{i-1},a_i+t,a_{i+1},\ldots,a_m)\\
&\quad -f(a_1+h_1,\ldots,a_{i-1}+h_{i-1},a_i,a_{i+1},\ldots,a_m).
\end{aligned}$$

e reescrevamos a expressão de cima na forma

$$f(\mathbf{a}+\mathbf{h})-f(\mathbf{a})=\sum_{i=1}^{m}g_i(h_i).$$

Para demonstrar o resultado, basta mostrar que, para cada i,

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\left|g_i(h_i)-h_i\frac{\partial f}{\partial x_i}(\mathbf{a})\right|}{\|\mathbf{h}\|}=0.$$

Fixemos um $\delta>0$. Como a derivada parcial $\dfrac{\partial f}{\partial x_i}$ existe numa vizinhança de $\mathbf{a}$, podemos escolher $r>0$ tal que $\dfrac{\partial f}{\partial x_i}(\mathbf{z})$ existe para $\mathbf{z}\in B_r(\mathbf{a})$. Como $\dfrac{\partial f}{\partial x_i}$ é contínua em $\mathbf{a}$, existe um ε ($\varepsilon<r$) tal que, se $\|\mathbf{z}-\mathbf{a}\|<\varepsilon$, então

$$\left|\frac{\partial f}{\partial x_i}(\mathbf{z})-\frac{\partial f}{\partial x_i}(\mathbf{a})\right|<\delta.$$

Por hipótese, a função de variável t,

$$f(a_1+h_1,a_2+h_2,\ldots,a_{i-1}+h_{i-1},a_i+t,a_{i+1},\ldots,a_m)$$

é derivável para $\sqrt{t^2+h_1^2+\ldots+h_{i-1}^2}<\varepsilon$. Quando $\|\mathbf{h}\|<\varepsilon$, o teorema dos acréscimos finitos dá

$$g_i(h_i)=h_i\cdot\frac{\partial f}{\partial x_i}(a_1+h_1,\ldots,a_{i-1}+h_{i-1},a_i+\theta h_i,a_{i+1},\ldots,a_m),$$

para um θ, $0 < \theta < 1$.

De onde resulta,

$$
\begin{aligned}
&\frac{\left|g_i(h_i) - h_i\frac{\partial f}{\partial x_i}(\mathbf{a})\right|}{\|\mathbf{h}\|} = \\
= \quad & |h_i| \cdot \frac{\left|\frac{\partial f}{\partial x_i}(a_1 + h_1, \ldots, a_{i-1} + h_{i-1}, a_i + \theta h_i, a_{i+1}, \ldots, a_m) - \frac{\partial f}{\partial x_i}(\mathbf{a})\right|}{\|\mathbf{h}\|} \\
& \leq \frac{|h_i|}{\|\mathbf{h}\|}\delta \leq \delta,
\end{aligned}
$$

como se pretendia demonstrar. $\square$

Podemos resumir a lista de implicações precedentes no diagrama seguinte.

As derivadas parciais de f existem em $B_\varepsilon(\mathbf{a})$
e são contínuas em $\mathbf{a}$

$\Downarrow$

f é diferenciável em $\mathbf{a}$

$\Downarrow$ $\Downarrow$

As derivadas direccionais existem em $\mathbf{a}$ em todas as direcções	f é contínua em $\mathbf{a}$

3.6 Diferenciais de ordem superior

3.6.1 Derivadas parciais de ordem superior

Como vimos anteriormente, dada uma função $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$, $\mathcal{D}$ aberto, a derivada parcial $\dfrac{\partial f}{\partial x_i}$ é novamente uma função de $\mathbb{R}^m$ em $\mathbb{R}$, cujo domínio estará obviamente contido em $\mathcal{D}$. É pois legítimo pensar se existirão as suas derivadas parciais, ou seja, as derivadas de segunda ordem. Como notação, utilizaremos as seguintes expressões

$$f''_{x_i^2}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_i^2}(\mathbf{a}) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right)(\mathbf{a})$$

$$f''_{x_i x_j}(\mathbf{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(\mathbf{a}).$$

Se existirem derivadas parciais de segunda ordem de f, podemos pensar em voltar a derivar, obtendo, caso existam, derivadas de terceira ordem de f. E assim sucessivamente. Vejamos para o caso de uma função definida em $\mathbb{R}^2$ o esquema seguinte, em que admitimos a existência de todas as derivadas parciais pelo menos até à ordem 3:

f
- $f'_x = \frac{\partial f}{\partial x}$
  - $f''_{x^2} = \frac{\partial^2 f}{\partial x^2}$
    - $f'''_{x^3} = \frac{\partial^3 f}{\partial x^3}$
    - $f'''_{x^2 y} = \frac{\partial^3 f}{\partial y \partial x^2}$
  - $f''_{xy} = \frac{\partial^2 f}{\partial y \partial x}$
    - $f'''_{xyx} = \frac{\partial^3 f}{\partial x \partial y \partial x}$
    - $f'''_{xy^2} = \frac{\partial^3 f}{\partial y^2 \partial x}$
- $f'_y = \frac{\partial f}{\partial y}$
  - $f''_{yx} = \frac{\partial^2 f}{\partial x \partial y}$
    - $f'''_{yx^2} = \frac{\partial^3 f}{\partial x^2 \partial y}$
    - $f'''_{yxy} = \frac{\partial^3 f}{\partial y \partial x \partial y}$
  - $f''_{y^2} = \frac{\partial^2 f}{\partial y^2}$
    - $f'''_{y^2 x} = \frac{\partial^3 f}{\partial x \partial y^2}$
    - $f'''_{y^3} = \frac{\partial^3 f}{\partial y^3}$

Apresentamos em seguida um resultado muito importante e de grande utilidade nas aplicações do cálculo diferencial.

Teorema 49 *(Teorema de Schwarz) Seja f uma função definida num aberto $\mathcal{D}$ de $\mathbb{R}^2$ e com valores em $\mathbb{R}$. Se f tiver derivadas parciais em $\mathcal{D}$ até à ordem 2 e estas forem contínuas, isto é, se f for de **classe** $C^2(\mathcal{D})$, então*

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial^2 f}{\partial x_2 \partial x_1}.$$

Dem Tomemos um ponto arbitrário $(a,b) \in \mathcal{D}$. Como $\mathcal{D}$ é aberto, podemos considerar uma bola aberta centrada em (a,b), $B_\epsilon(a,b)$, contida em $\mathcal{D}$. Para (h,k) tal que $(a+h,b+k) \in B_\epsilon(a,b)$, definamos a função seguinte,

$$\psi(h,k) = f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b).$$

Então,

$$\psi(h,k) = \varphi(1) - \varphi(0)$$

onde φ é dada por

$$\varphi(t) = f(a+th,b+k) - f(a+th,b).$$

Aplicando o teorema do valor médio de Lagrange,

$$\begin{aligned}
\psi(h,k) &= \varphi(1) - \varphi(0)\\
&= \varphi'(\xi)\\
&= h\frac{\partial f}{\partial x}(a+\xi h,b+k) - h\frac{\partial f}{\partial x}(a+\xi h,b)\\
&= hk\left[\frac{\partial^2 f}{\partial x \partial y}(a+\xi h,b+\mu k)\right],
\end{aligned}$$

com $\xi,\mu \in]0,1[$. Por outro lado,

$$\psi(h,k) = \omega(1) - \omega(0)$$

onde ω é dada por

$$\omega(t) = f(a+h,b+tk) - f(a,b+tk).$$

Cálculos análogos aos anteriores permitem concluir que

$$\begin{aligned}
\psi(h,k) &= \omega(1) - \omega(0)\\
&= hk\left[\frac{\partial^2 f}{\partial y \partial x}(a+\xi' h,b+\mu' k)\right],
\end{aligned}$$

com $\xi',\mu' \in]0,1[$. Logo

$$\frac{\partial^2 f}{\partial x \partial y}(a+\xi h,b+\mu k) = \frac{\partial^2 f}{\partial y \partial x}(a+\xi' h,b+\mu' k).$$

Portanto, passando ao limite quando $(h,k) \to (0,0)$, conclui-se que,

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

Nota 50 *O teorema anterior permite a generalização correspondente para uma função de classe $C^k(\mathcal{D})$ em $\mathbb{R}^m$, isto é, quaisquer que sejam as derivadas parciais de ordem igual ou inferior a k, é indiferente a ordem das variáveis relativamente às quais é feita a derivação. Como exemplo, suponhamos que temos uma função de classe $C^4(\mathbb{R}^3)$. Então teremos, por exemplo,*

$$\frac{\partial^3 f}{\partial x \partial y \partial z} = \frac{\partial^3 f}{\partial x \partial z \partial y} = \frac{\partial^3 f}{\partial z \partial x \partial y} = \frac{\partial^3 f}{\partial z \partial y \partial x}$$

$$\frac{\partial^4 f}{\partial x \partial y^2 \partial z} = \frac{\partial^4 f}{\partial x \partial y \partial z \partial y} = \frac{\partial^4 f}{\partial y \partial x \partial y \partial z} = \frac{\partial^4 f}{\partial z \partial y \partial x \partial y} = \dots\ .$$

3.6.2 Funções de Classe $\mathbf{C}^k$

Definição 51 *Uma função f definida num aberto $\mathcal{D}$ de $\mathbb{R}^m$ e com valores em $\mathbb{R}$ é de **classe $C^k(\mathcal{D})$**, $k \in \mathbb{N}$, isto é $f \in C^k(\mathcal{D})$, se tiver todas as derivadas parciais em $\mathcal{D}$ até à ordem k e estas forem contínuas.*

*Em particular, f é de **classe $C^\infty(\mathcal{D})$**, ou seja $f \in C^\infty(\mathcal{D})$ se tiver todas as derivadas parciais em $\mathcal{D}$ de qualquer ordem e estas forem contínuas.*

Definimos na secção anterior *diferencial de f em* $\mathbf{a}$, *Df(*$\mathbf{a}$*)*. Com efeito, trata-se do diferencial de primeira ordem de f em $\mathbf{a}$. É uma função $Df(\mathbf{a}) : \mathbb{R}^m \to \mathbb{R}$ dada por

$$Df(\mathbf{a})(h_1, h_2, \dots, h_m) = h_1 \frac{\partial f}{\partial x_1}(\mathbf{a}) + h_2 \frac{\partial f}{\partial x_2}(\mathbf{a}) + \dots + h_m \frac{\partial f}{\partial x_m}(\mathbf{a}).$$

Passemos agora aos diferenciais de ordem superior. Pelos teoremas da secção anterior, poderemos constatar que as funções das definições seguintes satisfazem o teorema de Schwarz.

Para aliviar a notação das funções e quando por comodidade de escrita for adequado, escreveremos para uma função genérica $g(\mathbf{a}) = g|_{(\mathbf{a})} = g_{|(\mathbf{a})}$ ou até, omitindo o ponto $\mathbf{a}$, apenas g. Por exemplo, sendo $g = \dfrac{\partial^2 f}{\partial x_1^2}$, identificaremos

$$\frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) \equiv \left.\frac{\partial^2 f}{\partial x_1^2}\right|_{(\mathbf{a})} \equiv \frac{\partial^2 f}{\partial x_1^2}_{|(\mathbf{a})},$$

e usaremos qualquer destas notações de forma indiferente. Por razões pedagógicas, começaremos por considerar funções definidas em $\mathbb{R}^2$.

Se $f : \mathcal{D} \subset \mathbb{R}^2 \to \mathbb{R}$ for de classe $C^2(\mathcal{D})$, definimos **diferencial de segunda ordem** de f em $\mathbf{a}$ como a função $D^2 f(\mathbf{a}) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ dada por

$$\begin{aligned} D^2 f(\mathbf{a})(\mathbf{u},\mathbf{v}) &= \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a})u_1v_1 + \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a})u_1v_2 + \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a})u_2v_1 + \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a})u_2v_2 \\ &= \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})u_iv_j. \end{aligned}$$

Terá especial interesse nas aplicações do próximo capítulo, o caso em que $\mathbf{u} = \mathbf{v}$. Tem-se nesse caso, com a convenção de escrita $D^2 f(\mathbf{a})(\mathbf{u}^2)$ em vez de $D^2 f(\mathbf{a})(\mathbf{u},\mathbf{u})$,

$$\begin{aligned} D^2 f(\mathbf{a})(\mathbf{u}^2) &= D^2 f(\mathbf{a})(\mathbf{u},\mathbf{u}) = \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a})u_1^2 + 2\frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a})u_1u_2 + \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a})u_2^2 \\ &= \sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})u_iu_j. \end{aligned}$$

Esta notação sugere a seguinte representação formal,

$$D^2 f(\mathbf{a})(\mathbf{u}^2) = \left[\left(u_1\frac{\partial}{\partial x_1} + u_2\frac{\partial}{\partial x_2}\right)^2 f\right](\mathbf{a}).$$

Note-se que o expoente 2 não representa o quadrado. É apenas um modo simbólico de condensar a escrita de diferencial de segunda ordem.

Passemos agora a $\mathbb{R}^m$. Se f for de classe $C^2(\mathcal{D})$, definimos **diferencial de segunda ordem de f em a** como sendo a função $D^2 f(\mathbf{a}) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ dada por

$$D^2 f(\mathbf{a})(\mathbf{u},\mathbf{v}) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})u_iv_j.$$

De forma análoga, se f for de classe $C^3(\mathcal{D})$, definimos **diferencial de terceira ordem de f em a** como sendo a função $D^3 f(\mathbf{a}) : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ dada por

$$D^3 f(\mathbf{a})(\mathbf{u},\mathbf{v},\mathbf{w}) = \sum_{i,j,s=1}^{m} \frac{\partial^3 f}{\partial x_s \partial x_j \partial x_i}(\mathbf{a})u_iv_jw_s.$$

Generalizando, se f for de classe $C^k(\mathcal{D})$, definimos **diferencial de ordem k de f em a** como sendo a função $D^k f(\mathbf{a}) : \mathbb{R}^m \times \mathbb{R}^m \times \ldots \times \mathbb{R}^m \to \mathbb{R}$ dada por

$$D^k f(\mathbf{a})(u^1, u^2, \ldots, u^k) = \sum_{i_1,i_2,\ldots,i_k=1}^{m} \frac{\partial^k f}{\partial x_{i_k} \ldots \partial x_{i_2} \partial x_{i_1}}(\mathbf{a})u^1_{i_1}u^2_{i_2}\ldots u^k_{i_k}.$$

Como já referimos, terá especial interesse nas aplicações do próximo capítulo, o caso em que os vectores são todos iguais. Nesse caso, as expressões anteriores assumem as formas seguintes

$$\begin{aligned}
D^2f(\mathbf{a})(\mathbf{u}^2) &= \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})u_i u_j, \\
D^3f(\mathbf{a})(\mathbf{u}^3) &= \sum_{i,j,s=1}^{m} \frac{\partial^3 f}{\partial x_s \partial x_j \partial x_i}(\mathbf{a})u_i u_j u_s, \\
&\dots \\
D^kf(\mathbf{a})(\mathbf{u}^k) &= \sum_{i_1,i_2,\dots,i_k=1}^{m} \frac{\partial^k f}{\partial x_{i_k} \dots \partial x_{i_2} \partial x_{i_1}}(\mathbf{a})u_{i_1} u_{i_2} \dots u_{i_k}.
\end{aligned}$$

Tal como anteriormente, podemos escrever estas expressões de forma simbólica, em que omitimos o elemento $(\mathbf{a})$,

$$\begin{gathered}
D^2f(\mathbf{u}^2) = \left(u_1\frac{\partial}{\partial x_1} + u_2\frac{\partial}{\partial x_2} + \dots + u_m\frac{\partial}{\partial x_m}\right)^2 f \\
D^3f(\mathbf{u}^3) = \left(u_1\frac{\partial}{\partial x_1} + u_2\frac{\partial}{\partial x_2} + \dots + u_m\frac{\partial}{\partial x_m}\right)^3 f \\
\dots \\
D^kf(\mathbf{u}^k) = \left(u_1\frac{\partial}{\partial x_1} + u_2\frac{\partial}{\partial x_2} + \dots + u_m\frac{\partial}{\partial x_m}\right)^k f.
\end{gathered}$$

Nota 52 *As propriedades operatórias conhecidas para a derivação em dimensão* 1 *permanecem válidas para os diferenciais em espaços de dimensão superior, com as adaptações adequadas. Por exemplo, se f e g forem de classe C^k em $\mathcal{D} \subset \mathbb{R}^m$ então $D^k(f+g)(\mathbf{a})(\mathbf{u}^k) = D^kf(\mathbf{a})(\mathbf{u}^k) + D^kg(\mathbf{a})(\mathbf{u}^k)$.*

3.6.3 Matriz Hessiana

Tomemos novamente o diferencial de segunda ordem de uma função $f \in C^2(\mathcal{D})$ em $\mathbf{a} \in \mathcal{D}$ aplicado a $(\mathbf{u}^2) = (\mathbf{u}, \mathbf{u}) \in \mathbb{R}^m \times \mathbb{R}^m$,

$$D^2f(\mathbf{a})(\mathbf{u}^2) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{a})u_i u_j.$$

A sua expressão não é mais do que um polinómio homogéneo de grau 2, isto é, uma forma quadrática que admite a seguinte formulação matricial,

$$\begin{bmatrix} u_1 & u_2 & \dots & u_m \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m \partial x_1}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m \partial x_2}(\mathbf{a}) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_1 \partial x_m}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_m}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m^2}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_m \end{bmatrix}.$$

À matriz

$$\begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m \partial x_1}(\mathbf{a}) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m \partial x_2}(\mathbf{a}) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_1 \partial x_m}(\mathbf{a}) & \frac{\partial^2 f}{\partial x_2 \partial x_m}(\mathbf{a}) & \dots & \frac{\partial^2 f}{\partial x_m^2}(\mathbf{a}) \end{bmatrix}$$

chama-se **Matriz Hessiana de f no ponto** a e representa-se por $H_f(\mathbf{a})$. Esta matriz determina inteiramente a aplicação $D^2 f(\mathbf{a})$. Uma vez que $f \in C^2(\mathcal{D})$, é válido o teorema de Schwarz e portanto as derivadas de segunda ordem cruzadas são iguais. Logo a matriz hessiana é simétrica. Assim, $D^2 f(\mathbf{a})(\mathbf{u}^2)$ é uma forma quadrática cuja matriz simétrica associada é a matriz Hessiana de f em $\mathbf{a}$,

$$D^2 f(\mathbf{a})(\mathbf{u}^2) = \mathbf{u}^T H_f(\mathbf{a})\mathbf{u}.$$

Em face do exposto, o diferencial de segunda ordem de f em $\mathbf{a}$, $D^2 f(\mathbf{a})(\mathbf{u}^2)$, ou, de forma equivalente, a matriz que lhe está associada, podem ser classificados de acordo com as classificações atribuídas na teoria das formas quadráticas ou das matrizes simétricas (definida positiva, definida negativa, etc). Este facto terá uma aplicação de grande interesse na classificação dos pontos críticos de uma função, como veremos no próximo capítulo.

3.6.4 Funções Homogéneas

Definição 53 *Seja f uma função de $\mathbb{R}^m$ em $\mathbb{R}$. Diz-se que f é homogénea de grau α se para qualquer $\lambda \in \mathbb{R}$ for válida a igualdade*

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) = \lambda^\alpha f(x_1, x_2, \ldots, x_m).$$

Exemplo 54

1. *A função $f(x, y, z) = x^3y^2z$ é homogénea de grau 6. Com efeito, $f(\lambda x, \lambda y,\, \lambda z) = (\lambda x)^3(\lambda y)^2\lambda z = \lambda^6x^3y^2z = \lambda^6 f(x, y, z)$.*

2. *As funções $f(x, y) = \sin(x^3y^2)$ e $f(x, y, z) = x^3y^2z + y^2z$ não são homogéneas.*

Nota 55 *Uma propriedade importante das funções homogéneas consiste no facto de que, se conhecermos o valor da função num ponto P, então conhecemos o valor da função em qualquer outro ponto Q da recta r que passa por P e pela origem. Com efeito, como se sabe por argumentos de geometria analítica, as coordenadas de Q são proporcionais às de P, facto que acontece a todos os pontos de r, e só a esses. Assim, se $P = (a_1, a_2, \ldots, a_m)$ então $Q = (ta_1, ta_2, \ldots, ta_m)$, para algum escalar $t \in \mathbb{R}$. Suponhamos então que f é homogénea de grau α e que $f(P) = B$. Então,*

$$f(Q) = f(ta_1, ta_2, \ldots, ta_m) = t^\alpha f(a_1, a_2, \ldots, a_m) = t^\alpha B.$$

Por exemplo, se f for homogénea de grau 2 e $f(2, 5, 4) = 10$ então $f(6, 15, 12) = 90$. Com efeito,

$$f(6, 15, 12) = f(3 \times 2, 3 \times 5, 3 \times 4) = 3^2 \times 10 = 90.$$

Outra propriedade importante das funções homogéneas é a que constitui o teorema de Euler.

Teorema 56 *(Teorema de Euler) Seja $f : \mathbb{R}^m \to \mathbb{R}$ uma função diferenciável. Se f é homogénea de grau α em $\mathbb{R}^m$, então é válida a igualdade seguinte,*

$$x_1\frac{\partial f}{\partial x_1}(x_1, x_2, \ldots, x_m) + x_2\frac{\partial f}{\partial x_2}(x_1, x_2, \ldots, x_m) + \ldots + x_m\frac{\partial f}{\partial x_m}(x_1, x_2, \ldots, x_m) =$$
$$= \alpha f(x_1, x_2, \ldots, x_m),$$

isto é, pondo $\mathbf{x} = (x_1, x_2, \ldots, x_m)$,

$$\mathbf{x} \cdot \nabla f(\mathbf{x}) = \alpha f(\mathbf{x}).$$

Dem Se f é homogénea de grau α em $\mathbb{R}^m$ então, para qualquer $\lambda \in \mathbb{R}$,

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) = \lambda^\alpha f(x_1, x_2, \ldots, x_m).$$

Derivando em ordem a λ e usando o teorema de derivação da função composta, obtém-se

$$x_1 \frac{\partial f}{\partial x_1}(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) + x_2 \frac{\partial f}{\partial x_2}(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) + \ldots$$

$$\ldots + x_m \frac{\partial f}{\partial x_m}(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) = \alpha \lambda^{\alpha-1} f(x_1, x_2, \ldots, x_m).$$

Como esta igualdade é válida para qualquer $\lambda \in \mathbb{R}$, em particular é válida para $\lambda = 1$, o que dá

$$x_1 \frac{\partial f}{\partial x_1}(x_1, x_2, \ldots, x_m) + x_2 \frac{\partial f}{\partial x_2}(x_1, x_2, \ldots, x_m) + \ldots + x_m \frac{\partial f}{\partial x_m}(x_1, x_2, \ldots, x_m) =$$

$$= \alpha f(x_1, x_2, \ldots, x_m). \square$$

Nota 57 *Consideremos a função* $f(x, y, z) = \sqrt{x^6 + y^6 + z^6}$ *e tomemos* $\lambda \in \mathbb{R}$. *Então,*

$$\begin{aligned} f(\lambda x, \lambda y, \lambda z) &= \sqrt{(\lambda x)^6 + (\lambda y)^6 + (\lambda z)^6} = \sqrt{\lambda^6 (x^6 + y^6 + z^6)} \\ &= |\lambda|^3 \sqrt{x^6 + y^6 + z^6} = |\lambda|^3 f(x, y, z). \end{aligned}$$

Neste caso, a função não é homogénea porque a condição de homogeneidade apenas é satisfeita para valores de λ *positivos. Diz-se então que* f *é positivamente homogénea de grau* 3.

Definição 58 *Uma função* $f : \mathbb{R}^m \to \mathbb{R}$ *é* ***positivamente homogénea de grau*** α *se para qualquer* $\lambda > 0$ *for válida a igualdade*

$$f(\lambda x_1, \lambda x_2, \ldots, \lambda x_m) = \lambda^\alpha f(x_1, x_2, \ldots, x_m).$$

Nota 59 *Tudo o que foi dito sobre homogeneidade para funções* f *de* $\mathbb{R}^m$ *em* $\mathbb{R}$ *permanece válido se o domínio de* f *for um subconjunto* A *de* $\mathbb{R}^m$ *tal que, se* $\mathbf{x} \in A$ *então* $\lambda \mathbf{x} \in A$, *para* $\lambda \in \mathbb{R} \setminus \{0\}$, *no caso de* f *ser homogénea, ou* $\lambda > 0$ *no caso de* f *ser positivamente homogénea. Por exemplo, a função*

$$f(x, y, z) = \frac{1}{z^4} \sin \frac{x^2 + y^2}{z^2}$$

é homogénea de grau -4 *em* $A = \mathbb{R}^3 \setminus \{(x, y, z) \in \mathbb{R}^3 : z \neq 0\}$.

3.6.5 Funções homogéneas e um exemplo económico

Consideremos o modelo de produção já referido anteriormente, mais precisamente o modelo de Cobb-Douglas

$$P(K,L) = AK^{\alpha}L^{\beta}.$$

em que $A > 0$. Ora

$$P(\lambda K, \lambda L) = A\,(\lambda K)^{\alpha}\,(\lambda L)^{\beta} = \lambda^{\alpha+\beta}AK^{\alpha}L^{\beta} = \lambda^{\alpha+\beta}P(K,L),$$

o que mostra que a função de Cobb-Douglas é homogénea de grau $\alpha + \beta$.

Numa função de produção que seja homogénea de grau 1 teremos

$$P(\lambda K, \lambda L) = \lambda P(K,L).$$

Isto traduz-se economicamente dizendo que P tem um **rendimento constante à escala**, isto é, o resultado varia na mesma proporção da variação dos dados. Por exemplo, se o capital e o trabalho duplicarem, a produção também duplicará. Se o grau de homogeneidade for maior do que 1 diz-se que P tem um **rendimento crescente à escala**; analogamente se for inferior a 1, diz-se que P tem **um rendimento decrescente à escala.**

3.7 Fórmula de Taylor

O leitor que tenha feito um curso de Cálculo Infinitesimal em $\mathbb{R}$ tomou certamente contacto com a fórmula de Taylor em dimensão 1 e com os factos seguintes:

Muitas funções podem, sob certas condições, ser aproximadas por polinómios, de tal modo que o erro cometido ao substituir a função pelo polinómio seja “pequeno”;

Por sua vez, quando o erro é pequeno, esses polinómios que aproximam as funções podem ser usados em vez da função original para cálculos (por exemplo, cálculo de integrais, derivadas, limites) sabendo-se, contudo, que o valor obtido não é exacto, mas sim, aproximado.

Estes factos permanecem válidos para funções em $\mathbb{R}^m$. Com efeito, a teoria tem analogias evidentes que poremos em destaque, embora apresente as modificações adequadas ao facto de o espaço ter dimensão superior a 1.

Vamos em seguida ver o teorema de Taylor que, para além de nos fazer olhar para uma função por meio de uma aproximação polinomial, será muito importante no estudo de existência de extremos que faremos no capítulo seguinte.

Simplifiquemos a notação escrevendo $D^k f(\mathbf{a})\mathbf{u}^k$ em vez de $D^k f(\mathbf{a})(\mathbf{u}^k)$, embora sem perdermos o sentido. Então temos

Teorema 60 *(Teorema de Taylor) Seja $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ uma função de classe $C^k(\mathcal{D})$, com $\mathcal{D}$ aberto e $\mathbf{a} \in \mathcal{D}$. Seja $\mathbf{h} \in \mathbb{R}^m$ tal que o segmento $[\mathbf{a}, \mathbf{a}+\mathbf{h}]$ está contido em $\mathcal{D}$. Então,*

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) \quad &= f(\mathbf{a}) + Df(\mathbf{a})\mathbf{h} + \tfrac{1}{2!}D^2 f(\mathbf{a})\mathbf{h}^2 + \tfrac{1}{3!}D^3 f(\mathbf{a})\mathbf{h}^3 + \dots \\ &\dots + \tfrac{1}{(k-1)!}D^{k-1} f(\mathbf{a})\mathbf{h}^{k-1} + r_k(\mathbf{h}), \end{aligned}$$

onde

$$r_k(\mathbf{h}) = \frac{1}{k!}(D^k f)(\mathbf{a}+\theta\mathbf{h})\mathbf{h}^k$$

e, tal como definimos no capítulo anterior,

$$D^j f(\mathbf{a})(u^j) = \sum_{i_1, i_2, \dots, i_j = 1}^{m} \frac{\partial^j f}{\partial x_{i_j} \dots \partial x_{i_2} \partial x_{i_1}}(\mathbf{a}) u_{i_1} u_{i_2} \dots u_{i_j}, \ \textit{com } j = 1, 2, \dots, k.$$

Dem Consideremos a função real de variável real, definida no intervalo $[0, 1]$ por

$$\varphi(t) = f(\mathbf{a}+t\mathbf{h}).$$

Aplicando o teorema de Taylor em dimensao 1 à função φ no intervalo $]0, 1[$ obtemos

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2!}\varphi''(0) + \frac{1}{3!}\varphi'''(0) + \dots + \frac{1}{(k-1)!}\varphi^{(k-1)}(0) + \frac{1}{k!}\varphi^{(k)}(\tau),$$

com $\tau \in]0, 1[$. Ora, pela derivada da função composta,

$$\varphi'(t) = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(\mathbf{a}+t\mathbf{h})h_i = Df(\mathbf{a}+t\mathbf{h})(\mathbf{h})$$

$$\varphi''(t) = \sum_{i,j=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}+t\mathbf{h})h_i h_j = D^2 f(\mathbf{a}+t\mathbf{h})(\mathbf{h}^2)$$

e assim sucessivamente, para $\varphi'''(t), ..., \varphi^{(k)}(t)$. Logo,

$$\begin{aligned}
\varphi'(0) &= Df(\mathbf{a})(\mathbf{h}) \\
\varphi''(0) &= D^2 f(\mathbf{a})(\mathbf{h}^2) \\
&\vdots \\
\varphi^{(k-1)}(0) &= D^{(k-1)} f(\mathbf{a})(\mathbf{h}^{k-1}) \\
\varphi^{(k)}(\tau) &= D^{(k)} f(\mathbf{a}+\tau\mathbf{h})(\mathbf{h}^k).
\end{aligned}$$

e como

$$\varphi(1) = f(\mathbf{a}+\mathbf{h}) \quad \text{e} \quad \varphi(0) = f(\mathbf{a}),$$

substituindo na fórmula de Taylor de φ acima, obtém-se o resultado pretendido, o que termina a demonstração. $\square$

Nota 61 *Ao resto dado pela fórmula anterior, isto é, a*

$$r_k(\mathbf{h}) = \frac{1}{k!}(D^k f)(\mathbf{a}+\theta\mathbf{h})\mathbf{h}^k$$

chama-se ***resto de Lagrange****. Tem-se*

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{r_k(\mathbf{h})}{\|\mathbf{h}\|^{k-1}} = 0,$$

isto é, $r_k(\mathbf{h})$ *é um infinitésimo de ordem superior a* $k-1$*.*

O teorema anterior mostra que o valor da função f num ponto $\mathbf{x}$ pertencente à vizinhança de $\mathbf{a}$, isto é, $\mathbf{x} = \mathbf{a}+\mathbf{h}$, pode ser aproximado pelo valor de um polinómio de grau $k-1$ do seguinte modo,

$$\begin{aligned}
f(\mathbf{a}+\mathbf{h}) \approx \quad & f(\mathbf{a}) + (Df)(\mathbf{a})\mathbf{h} + \tfrac{1}{2!}(D^2 f)(\mathbf{a})\mathbf{h}^2 + \tfrac{1}{3!}(D^3 f)(\mathbf{a})\mathbf{h}^3 + \dots \\
& \dots + \tfrac{1}{(k-1)!}(D^{k-1} f)(\mathbf{a})\mathbf{h}^{k-1},
\end{aligned}$$

cometendo-se o erro

$$\begin{aligned}
r_k(\mathbf{h}) &= \tfrac{1}{k!}(D^k f)(\mathbf{a}+\theta\mathbf{h})\mathbf{h}^k \\
&= f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - (Df)(\mathbf{a})\mathbf{h} - \tfrac{1}{2!}(D^2 f)(\mathbf{a})\mathbf{h}^2 - \tfrac{1}{3!}(D^3 f)(\mathbf{a})\mathbf{h}^3 - \dots \\
&\quad \dots - \tfrac{1}{(k-1)!}(D^{k-1} f)(\mathbf{a})\mathbf{h}^{k-1}.
\end{aligned}$$

Se conhecessemos o valor exacto do erro, poderíamos adicioná-lo ao valor aproximado de f e obteríamos o valor exacto de f. Contudo, não se conhece o valor exacto do erro, o que provém do facto de a expressão de $r_k(\mathbf{h})$ depender do ponto $\mathbf{a}+\theta\mathbf{h}$, que não é conhecido com exactidão. Na prática, este facto é ultrapassado usando uma majoração do erro, o que fornece um intervalo de confiança para o valor de f.

3.8 Função Inversa e Função Implícita

Apresentamos nesta secção dois teoremas de grande relevância no cálculo diferencial: os teorema da função inversa e da função implícita.

Sabendo que é muito comum que uma função não seja invertível, é frequentemente suficiente o conhecimento de que ela é invertível numa vizinhança de um ponto do seu domínio. O Teorema da Função Inversa, que enunciamos em seguida, é um resultado que trata da possibilidade de inverter uma função, ainda que apenas localmente. O teorema também estabelece propriedades de diferenciabilidade da inversa local.

Teorema 62 *(Teorema da Função Inversa) Seja* $A \subset \mathbb{R}^m$ *um conjunto aberto e* $\mathbf{f} : A \longrightarrow \mathbb{R}^m$ *uma aplicação de classe* C^1*. Seja* $\mathbf{a} = (a_1, ..., a_m) \in A$ *e* $\mathbf{b} = (b_1, ..., b_m) = f(\mathbf{a})$ *e suponhamos que o jacobiano é não nulo,* $|J_{\mathbf{f}}(\mathbf{a})| \neq 0$*. Então*
i) existem abertos U *e* V *de* $\mathbb{R}^m$*, tais que* $\mathbf{a} \in U$ *e* $\mathbf{b} \in V$ *e* $\mathbf{f} : U \longrightarrow V$ *é uma bijecção;*
ii) a função (inversa local de $\mathbf{f}$*)* $\mathbf{f}^{-1} : V \longrightarrow U$ *é de classe* C^1 *e*

$$D\mathbf{f}^{-1}(\mathbf{x}) = (D\mathbf{f}(\mathbf{x}))^{-1},$$

isto é, $\mathbf{f}$ *é um difeomorfismo local de classe* C^1*.*

O teorema da função implícita permite afirmar que uma relação entre variáveis pode, sob certas condições, garantir que uma das variáveis se exprime em função das restantes, de forma única. Um exemplo típico de função implicita é o da equação da circunferência que determina uma curva no plano $x0y$. Contudo, à partida não se exprime uma das variáveis em função da restante de forma única. Com efeito, consideremos a circunferência de centro $(0, 0)$ e raio 1

$$x^2 + y^2 = 1.$$

Então, poderemos dizer

$$x = \pm\sqrt{1 - y^2} \text{ ou } y = \pm\sqrt{1 - x^2}\,.$$

Se escolhermos um ponto da curva, já poderemos garantir que na vizinhança desse ponto a expressão é única, constituindo assim uma aplicação bijectiva. Por exemplo, consideremos o ponto $(\frac{1}{2}, \frac{3}{4})$. Este ponto pertence ao quarto de circunferência do primeiro quadrante. Então, considerando

$$\begin{aligned} \varphi &: \ ]0,1[\longrightarrow]0,1[\\ \\ y &= \varphi(x) = \sqrt{1-x^2}\ , \end{aligned}$$

fica evidente que a relação $x^2 + y^2 = 1$ define implicitamente uma função numa vizinhança de $(\frac{1}{2}, \frac{3}{4})$, o que conlcuímos por observação directa explicitando a própria função φ. Ora, nem sempre tal é possível. Mas, mesmo não conseguindo obter a expressão explícita, poderemos, em certos casos, garantir se uma relação de variáveis define implicitamente uma função. Poderemos até dar informações sobre a respectiva diferenciabilidade. É esse o conteúdo do teorema da função implícita. Enunciamos, primeiramente, para o caso de duas variáveis e em seguida para uma situação mais geral.

Teorema 63 *(Teorema da função implícita) Seja D um subconjunto aberto de $\mathbb{R}^2$, f uma função de classe C^1 em D e $(a,b) \in D$. Suponhamos que $f(a,b) = 0$ e $\frac{\partial f}{\partial y}(a,b) \neq 0$. Então*
i) existem intervalos $I_a =]a-\epsilon, a+\epsilon[$ e $I_b =]b-\delta, b+\delta[$ e uma função $\varphi : I_a \longrightarrow I_b$, bijectiva, tal que $f(a, \varphi(a)) = 0$
ii) φ é de classe C^1 e tem-se para $x \in I_a$

$$\varphi'(x) = -\frac{\frac{\partial f}{\partial x}(x,\varphi(x))}{\frac{\partial f}{\partial y}(x,\varphi(x))}.$$

Este teorema admite formulação mais geral:

Teorema 64 *Seja D um subconjunto aberto de $\mathbb{R}^{m+1}$, f uma função de classe C^1 em D. Notemos $(\mathbf{x}, y) = (x_1, ..., x_m, y)$ os elementos de $\mathbb{R}^{m+1}$ e seja $(\mathbf{a}, b) = (a_1, ..., a_m, b) \in D$. Suponhamos que $f(a_1, ..., a_m, b) = 0$ e que $\frac{\partial f}{\partial y}(a_1, ..., a_m, b) \neq 0$. Então*
i) existem intervalos $I_{a_1} =]a_1 - \epsilon_1, a_1 + \epsilon_1[$, ..., $I_{a_m} =]a_m - \epsilon_m, a_m + \epsilon_m[$ e $I_b =]b-\delta, b+\delta[$ e uma função $\varphi : I_{a_1} \times ... \times I_{a_m} \longrightarrow I_b$, bijectiva, tal que $f(a_1, ..., a_m, \varphi(\mathbf{a})) = 0$
ii) φ é de classe C^1 e tem-se para $\mathbf{x} \in I_{a_1} \times ... \times I_{a_m}$

$$\frac{\partial \varphi}{\partial x_i}(\mathbf{x}) = -\frac{\frac{\partial f}{\partial x_i}(\mathbf{x},\varphi(\mathbf{x}))}{\frac{\partial f}{\partial y}(\mathbf{x},\varphi(\mathbf{x}))}.$$

As demonstrações envolvem alguma complexidade que extravaza o enquadramento teórico deste texto e o fim a que se destina. Podem, contudo, ser encontradas, por exemplo em [4], [7].

Capítulo 4

Problemas de Extremo. Optimização

4.1 Introdução

Entre as aplicações mais relevantes do Cálculo encontram-se os problemas de determinação de valores máximos e mínimos. Expressões usadas no quotidiano, tais como, *distância máxima, tamanho óptimo, custo mínimo, lucro máximo,* ilustram precisamente esta questão e correspondem a problemas habitualmente referidos como de optimização. Estes problemas assumem especial importância em Economia e agrupam-se em três categorias: optimização livre, optimização com condições de igualdade e optimização com condições de igualdade e desigualdade. Põe-se também a questão do estudo da existência de óptimos locais ou globais. Os ingredientes matemáticos necessários a esse estudo formam o tópico, genericamente designado por *problemas de extremo,* que constitui o conteúdo deste capítulo, directamente ligado a optimização não linear. Está organizado em três secções que correspondem precisamente aos problemas de optimização referidos. Terminamos o capítulo ilustrando a teoria exposta com aplicações a problemas de Economia.

4.2 Conceitos básicos

Esta secção está directamente relacionada com a optimização livre.

Seja $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ uma função real de m variáveis reais, de domínio $\mathcal{D}$ e $\mathbf{a} \in \mathcal{D}$.

Definição 65 *Diz-se que,*
*- A função f tem um **máximo global** em $\mathbf{a}$ se*

$$f(\mathbf{x}) \leq f(\mathbf{a}), \ \ \forall \mathbf{x} \in \mathcal{D}.$$

*O ponto a é então chamado **ponto de máximo global** ou **maximizante global**.*

*- A função f tem um **máximo local** em $\mathbf{a}$ se existir $\varepsilon > 0$ tal que*

$$f(\mathbf{x}) \leq f(\mathbf{a}), \ \forall \mathbf{x} \in B_\varepsilon(\mathbf{a}) \cap \mathcal{D}.$$

*Neste caso, o ponto $\mathbf{a}$ é chamado um **ponto de máximo local** ou **maximizante local**.*

Analogamente,

*- A função f possui um **mínimo global** em $\mathbf{a}$ se*

$$f(\mathbf{x}) \geq f(\mathbf{a}), \ \ \forall \mathbf{x} \in \mathcal{D}.$$

*Ao ponto a chama-se **ponto de mínimo global** ou **minimizante global**.*

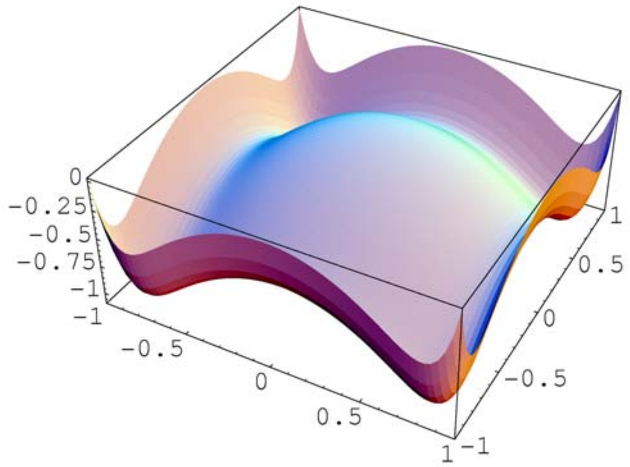
*- A função f tem um **mínimo local** em $\mathbf{a}$ se existir $\varepsilon > 0$ tal que*

$$f(\mathbf{x}) \geq f(\mathbf{a}), \ \forall \mathbf{x} \in B_\varepsilon(\mathbf{a}) \cap \mathcal{D}.$$

*Neste caso, o ponto $\mathbf{a}$ é chamado um **ponto de mínimo local** ou **minimizante local**.*
*Por **extremo** (local ou global) entende-se um máximo ou mínimo (local ou absoluto). Por **extremante** (local ou global) entende-se um ponto que é ponto de máximo ou mínimo (local ou global).*

Muitos autores utilizam as designações "absoluto" e "relativo" em vez de "global" e "local", respectivamente.

Exemplo 66 *A função $f(x, y) = x^{10} + y^{10} - x^2 - y^2$, cujo esboço de gráfico é o seguinte,*



tem um maximizante local em $(0,0)$*, que não é maximizante absoluto, pois* f *não é majorada. O gráfico ilustra ainda o facto de função* f *possuir minimizantes, que um estudo mais cuidado mostra que são absolutos.*

4.3 Extremos livres

Ao longo desta secção assumiremos sempre que $\mathcal{D}$ é um aberto de $\mathbb{R}^m$ e que $f : \mathcal{D} \subset \mathbb{R}^m \to \mathbb{R}$ é uma função diferenciável em $\mathcal{D}$.

O facto de $\mathcal{D}$ ser aberto leva a que a existência de extremantes e extremos esteja apenas relacionada com a expressão da função f, sem que o domínio exerça alguma interferência relevante. Isto é, se $\mathbf{a}$ for extremante local de f em $\mathcal{D}$, essa situação de extremante local não se altera mesmo que faça uma restrição ao domínio de f, desde que contenha o ponto $\mathbf{a}$. Como tal, também não se altera a situação de $f(\mathbf{a})$ ser extremo local, o que nos leva a dizer que **extremos locais num aberto** são **extremos livres**. Esta designação aparece por oposição à de **extremos condicionados**, tópico que será estudado nas secções seguintes e em que haverá uma influência directa do domínio de f, traduzida pela imposição de certas condições.

É claro que sendo $\mathcal{D}$ aberto, nas definições de maximizante local ou minimizante local apresentadas no parágrafo anterior, $B_\varepsilon(\mathbf{a}) \cap \mathcal{D}$ pode ser substituído por $B_\varepsilon(\mathbf{a})$, uma vez que, para todo o ponto $\mathbf{a} \in \mathcal{D}$, existe $\varepsilon > 0$ tal que $B_\varepsilon(\mathbf{a}) \subseteq \mathcal{D}$.

Definição 67 *Se* $\mathbf{a} \in \mathcal{D}$ *for tal que* $Df(\mathbf{a}) \equiv 0$*, então* $\mathbf{a}$ *é chamado* ***ponto***

crítico *ou* ***ponto de estacionaridade*** *de* f.

Notação 68 $Df(\mathbf{a}) \equiv 0$ *significa que a aplicação linear* $Df(\mathbf{a})(\mathbf{h})$ *é a função identicamente nula, isto é,* $Df(\mathbf{a})(\mathbf{h}) = 0, \forall \mathbf{h} \in \mathbb{R}^m$.

Nota 69 ***1.*** *É fácil ver que* $Df(\mathbf{a}) \equiv 0$ *é equivalente a* $\nabla f(\mathbf{a}) = \mathbf{0}$ *ou, ainda, a* $f'_{x_i}(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$ *para todo* $1 \leq i \leq n$. *Com efeito, se* $\frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$, $1 \leq i \leq m$, *ou seja,* $\nabla f(\mathbf{a}) = \mathbf{0}$, *é óbvio que* $Df(\mathbf{a}) \equiv 0$ *uma vez que*

$$\begin{aligned} Df(\mathbf{a}) \equiv 0 \quad &\Longleftrightarrow \quad Df(\mathbf{a})(\mathbf{h}) = 0, \forall \mathbf{h} \in \mathbb{R}^m \\ &\Longleftrightarrow \quad \frac{\partial f}{\partial x_1}(\mathbf{a})h_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})h_2 + ... + \frac{\partial f}{\partial x_m}(\mathbf{a})h_m = 0, \forall \mathbf{h} \in \mathbb{R}^m. \end{aligned}$$

Reciprocamente, se $Df(\mathbf{a}) \equiv 0$ *e se aplicarmos sucessivamente* $Df(\mathbf{a})$ *aos vectores da base canónica de* $\mathbb{R}^m$,

$$\mathbf{e}_1 = (1, 0, ..., 0),\ \mathbf{e}_2 = (0, 1, ..., 0), ...,\ \mathbf{e}_m = (0, 0, ..., 1),$$

obtemos

$$\begin{aligned} Df(\mathbf{a})(\mathbf{e}_1) = 0 \Leftrightarrow \frac{\partial f}{\partial x_1}(\mathbf{a}).1 + \frac{\partial f}{\partial x_2}(\mathbf{a}).0 + ... + \frac{\partial f}{\partial x_m}(\mathbf{a}).0 &= 0 \\ Df(\mathbf{a})(\mathbf{e}_2) = 0 \Leftrightarrow \frac{\partial f}{\partial x_1}(\mathbf{a}).0 + \frac{\partial f}{\partial x_2}(\mathbf{a}).1 + ... + \frac{\partial f}{\partial x_m}(\mathbf{a}).0 &= 0 \\ &\cdots \\ Df(\mathbf{a})(\mathbf{e}_m) = 0 \Leftrightarrow \frac{\partial f}{\partial x_1}(\mathbf{a}).0 + \frac{\partial f}{\partial x_2}(\mathbf{a}).0 + ... + \frac{\partial f}{\partial x_m}(\mathbf{a}).1 &= 0, \end{aligned}$$

isto é,

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) = \frac{\partial f}{\partial x_2}(\mathbf{a}) = \cdots = \frac{\partial f}{\partial x_m}(\mathbf{a}) = 0.$$

Logo, se $Df(\mathbf{a}) \equiv 0$ *então* $f'_{x_i}(\mathbf{a}) = \frac{\partial f}{\partial x_i}(\mathbf{a}) = 0$, $1 \leq i \leq m$, *isto é,* $\nabla f(\mathbf{a}) = \mathbf{0}$.

2. *A observação do ponto 1 é muito utilizada, em termos práticos, na determinação de pontos críticos de uma função* f. *Com efeito, habitualmente procuram-se os pontos críticos* $\mathbf{a} \in \mathcal{D}$ *através da procura dos pontos que* *anulam simultaneamente todas as derivadas parciais de primeira ordem de f, isto é, através da resolução do sistema*

$$\begin{cases} \dfrac{\partial f}{\partial x_1}(\mathbf{a}) = 0 \\ \dfrac{\partial f}{\partial x_2}(\mathbf{a}) = 0 \\ \vdots \\ \vdots \\ \dfrac{\partial f}{\partial x_m}(\mathbf{a}) = 0. \end{cases}$$

Teorema 70 *Se $\mathbf{a} \in \mathcal{D}$ for ponto de máximo ou mínimo local de f então $\mathbf{a}$ é ponto crítico de f.*

Dem Seja $\mathbf{a}$ um ponto de máximo local de f. Considere-se $\mathbf{v} \in \mathbb{R}^m$ um vector arbitrário e $\varepsilon > 0$ tal que $\mathbf{a} + t\mathbf{v} \in \mathcal{D}$ se $|t| < \varepsilon$. Então $t = 0$ é um ponto de máximo local da função de variável real $g(t) = f(\mathbf{a} + t\mathbf{v})$. Logo $g'(0) = 0$, isto é, sendo

$$g'(t) = \frac{\partial f}{\partial x_1}(\mathbf{a} + t\mathbf{v})v_1 + \frac{\partial f}{\partial x_2}(\mathbf{a} + t\mathbf{v})v_2 + ... + \frac{\partial f}{\partial x_m}(\mathbf{a} + t\mathbf{v})v_m, \forall \mathbf{v} \in \mathbb{R}^m,$$

com $|t| < \varepsilon$ então, para $t = 0$,

$$\frac{\partial f}{\partial x_1}(\mathbf{a})v_1 + \frac{\partial f}{\partial x_2}(\mathbf{a})v_2 + ... + \frac{\partial f}{\partial x_m}(\mathbf{a})v_m = 0, \forall \mathbf{v} \in \mathbb{R}^m,$$

isto é, $Df(\mathbf{a}) \equiv 0$, ou seja $\mathbf{a}$ é ponto crítico de f. Se $\mathbf{a}$ for um ponto de mínimo local de f a demonstração é análoga. $\square$

Nota 71 *Sendo f diferenciável, a condição $Df(\mathbf{a}) \equiv 0$ é necessária para que $\mathbf{a}$ seja um ponto extremo de f. Contudo não é suficiente. Por exemplo, consideremos a função definida em $\mathbb{R}^2$ por $f(x, y) = x^2 - y^2$. O ponto $(0, 0)$ é um ponto crítico de f, mas não é um ponto de máximo ou mínimo local de f, conforme ilustra o esboço do gráfico*

[width=6cm]extremosg2.eps

*Neste caso, o ponto $\mathbf{a}$ chama-se **ponto de sela**. Mais precisamente,*

Definição 72 *Diz-se que o ponto $\mathbf{a}$ é um **ponto de sela** de f se*

1. $Df(\mathbf{a}) \equiv 0$,

2. $\forall \varepsilon > 0,\ \exists \mathbf{b}, \mathbf{c} \in B_\varepsilon(\mathbf{a}) :\ f(\mathbf{b}) < f(\mathbf{a}) < f(\mathbf{c})$.

Vamos dar algumas condições suficientes para que um ponto crítico seja ponto de máximo ou mínimo local. Recordemos que se f for uma função real de uma variável real, o ponto a é um maximizante local de f se $f'(a) = 0$ e se $f''(a) < 0$. Da mesma forma, o ponto a é um minimizante local se $f'(a) = 0$ e $f''(a) > 0$. A situação torna-se mais delicada se $f''(a)$ for nula. Observe-se, por exemplo, para as funções $g(x) = x^3$ e $h(x) = x^4$,

$$g'(0) = g''(0) = 0 = h'(0) = h''(0)$$

e, no entanto, o ponto 0 é ponto sela de g e minimizante de h. Para funções de várias variáveis e de classe C^2, a condição $f'(a) = 0$ é substituída pela condição $Df(\mathbf{a}) \equiv 0$ e a condição $f''(a) < 0$ é substituída por uma condição sobre o diferencial de segunda ordem, mais precisamente, pela condição "*$D^2f(\mathbf{a})$ é uma forma quadrática definida negativa*" ou, de forma equivalente, "*$H_f(\mathbf{a})$ é uma matriz simétrica definida negativa*". Da mesma forma, a condição $f''(a) > 0$ é substituída pela condição "*$D^2f(\mathbf{a})$ é uma forma quadrática definida positiva*" ou "*$H_f(\mathbf{a})$ é uma matriz simétrica definida positiva*". A situação "*$D^2f(\mathbf{a})$ é uma forma quadrática indefinida*" corresponderá ao facto de $\mathbf{a}$ ser ponto sela. As situações em que o diferencial de segunda ordem seja uma forma quadrática semidefinida serão mais problemáticas. No entanto, passando à análise de diferenciais de ordem superior, poder-se-á tirar conclusões em bastantes casos. Chama-se também a atenção para o facto de uma análise local, de carácter mais geométrico, numa vizinhança do ponto crítico, permitir muitas vezes tirar conclusões sobre a sua natureza. Os teoremas seguintes tornarão claro o que acabámos de dizer.

Teorema 73 *Seja $f : \mathbb{R}^m \to \mathbb{R}$ uma aplicação de classe $\mathcal{C}^2$ numa vizinhança do ponto $\mathbf{a}$. Supondo que $Df(\mathbf{a}) \equiv 0$, então,*

1. *Se $H_f(\mathbf{a})$ é definida positiva, então $\mathbf{a}$ é um minimizante local;*

2. *Se $H_f(\mathbf{a})$ é definida negativa, então $\mathbf{a}$ é um maximizante local;*

3. *Se $H_f(\mathbf{a})$ é indefinida, então $\mathbf{a}$ é um ponto de sela.*

Dem Consideremos o desenvolvimento de Taylor de ordem 1 de f em torno do ponto $\mathbf{a}$,

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + Df(\mathbf{a})\,(\mathbf{h}) + \frac{1}{2}D^2 f(\mathbf{a})\left(\mathbf{h}^2\right) + o(||\mathbf{h}||^2).$$

Usando linguagem matricial e uma vez que, pelo teorema de Schwartz, $H_f(\mathbf{a})$ é simétrica, a expressão anterior pode ser escrita na forma seguinte

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \mathbf{h}.\nabla f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^T H_f(\mathbf{a})\mathbf{h} + o(||\mathbf{h}||^2).$$

Como o ponto $\mathbf{a}$ é ponto crítico, $\nabla f(\mathbf{a}) = \mathbf{0}$ e esta equação reduz-se a

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^T H_f(\mathbf{a})\mathbf{h} + o(||\mathbf{h}||^2).$$

Se $H_f(\mathbf{a})$ for definida positiva, existe $c > 0$ tal que

$$\mathbf{h}^T H_f(\mathbf{a})\mathbf{h} \geq c||\mathbf{h}||^2, \ \forall h \in \mathbb{R}^m.$$

Por definição do símbolo "$o(...)$", dado c' existe $\varepsilon_{c'} > 0$ tal que se $||\mathbf{h}|| < \varepsilon_{c'}$ então

$$|o(||\mathbf{h}||^2)| < c'||\mathbf{h}||^2.$$

Fixando um $c' < \dfrac{c}{2}$, se $||\mathbf{h}|| < \varepsilon_{c'}$ então

$$f(\mathbf{a}+\mathbf{h}) > f(\mathbf{a}) + \frac{c}{2}||\mathbf{h}||^2 - c'||\mathbf{h}||^2,$$

o que mostra que

$$f(\mathbf{a}+\mathbf{h}) \geq f(\mathbf{a}),$$

com desigualdade estrita se $\mathbf{h} \neq \mathbf{0}$. A demonstração é análoga se $H_f(\mathbf{a})$ for definida negativa. Se $H_f(\mathbf{a})$ for indefinida, existem vectores $\mathbf{v}$ e $\mathbf{w}$ não nulos tais que

$$\mathbf{v}^T H_f(\mathbf{a})\mathbf{v} > 0 \quad \text{e} \quad \mathbf{w}^T H_f(\mathbf{a})\mathbf{w} < 0.$$

Da fórmula de Taylor, se λ for tal que $\mathbf{a}+\lambda\mathbf{v} \in \mathcal{D}$ e $\mathbf{a}+\lambda\mathbf{w} \in \mathcal{D}$, tem-se

$$\begin{aligned} f(\mathbf{a}+\lambda\mathbf{v}) &= f(\mathbf{a}) + \frac{\lambda^2}{2}\mathbf{v}^T H_f(\mathbf{a})\mathbf{v} + o(\lambda^2||\mathbf{v}||^2) \\ f(\mathbf{a}+\lambda\mathbf{w}) &= f(\mathbf{a}) + \frac{\lambda^2}{2}\mathbf{w}^T H_f(\mathbf{a})\mathbf{w} + o(\lambda^2||\mathbf{w}||^2). \end{aligned}$$

Como, para λ suficientemente pequeno,

$$\frac{\lambda^2}{2}\mathbf{v}^T H_f(\mathbf{a})\mathbf{v} > \left|o(\lambda^2||\mathbf{v}||^2)\right| \quad \text{e} \quad \frac{\lambda^2}{2}\mathbf{w}^T H_f(\mathbf{a})\mathbf{w} < -\left|o(\lambda^2||\mathbf{w}||^2)\right|,$$

obtem-se

$$\begin{aligned} f(\mathbf{a}+\lambda\mathbf{v}) &\geq f(\mathbf{a})+\frac{\lambda^2}{2}\mathbf{v}^T H_f(\mathbf{a})\mathbf{v}-\left|o(\lambda^2||\mathbf{v}||^2)\right| > f(\mathbf{a}),\\ f(\mathbf{a}+\lambda\mathbf{w}) &\leq f(\mathbf{a})+\frac{\lambda^2}{2}\mathbf{w}^T H_f(\mathbf{a})\mathbf{w}+\left|o(\lambda^2||\mathbf{w}||^2)\right| < f(\mathbf{a}),\end{aligned}$$

o que mostra que $\mathbf{a}$ é um ponto de sela. □

O teorema não diz nada quando a matriz hessiana é semidefinida positiva ou semidefinida negativa. Neste caso, será preciso examinar caso a caso a situação apresentada.

Combinando o teorema anterior com os resultados sobre formas quadráticas, obtemos o corolário seguinte, que se torna muito cómodo no estudo de extremos locais.

Corolário 74 *Seja $f : \mathbb{R}^m \to \mathbb{R}$ uma aplicação de classe $\mathcal{C}^2$ numa vizinhança do ponto $\mathbf{a}$. Supondo que $Df(\mathbf{a}) \equiv 0$, designando por $H_f(\mathbf{a})$ a matriz hessiana de f no ponto $\mathbf{a}$ e por Δ_i o menor principal primário de ordem i respectivo,*

1. *Se $\Delta_1 > 0$, $\Delta_2 > 0$, $\Delta_3 > 0$,..., $\Delta_m > 0$, então $\mathbf{a}$ é um minimizante local;*
2. *Se $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$,..., $(-1)^m\Delta_m > 0$, então $\mathbf{a}$ é um maximizante local;*
3. *Se $\mathbf{a}$ for um minimizante então $\Delta_1 \geq 0$, $\Delta_2 \geq 0$, $\Delta_3 \geq 0$,..., $\Delta_m \geq 0$;*
4. *Se $\mathbf{a}$ for um maximizante então $\Delta_1 \leq 0$, $\Delta_2 \geq 0$, $\Delta_3 \leq 0$,..., $(-1)^m\Delta_m \geq 0$;*
5. *Se $H_f(\mathbf{a})$ tiver valores próprios de sinais diferentes, então $\mathbf{a}$ é um ponto de sela.*

Exemplo 75 *Determinemos os extremantes da função*

$$f(x,y,z) = -x^3+3xz+4y-y^2-5z^2.$$

1° *Determinação dos pontos críticos de f*

As derivadas parciais são

$$\begin{cases} f'_x(x,y,z) = -3x^2+3z\\ \\ f'_y(x,y,z) = 4-2y\\ \\ f'_z(x,y,z) = 3x-10z.\end{cases}$$

As soluções do sistema $f'_x(x,y,z) = 0$, $f'_y(x,y,z) = 0$, $f'_z(x,y,z) = 0$ *são* $(0,2,0)$ *e* $(\frac{3}{10}, 2, \frac{9}{100})$, *que são, portanto, os pontos críticos de* f.

2° *Classificação dos pontos críticos de f*
A matriz hessiana $H_f(x,y,z)$ *é*

$$\begin{bmatrix} -6x & 0 & 3 \\ 0 & -2 & 0 \\ 3 & 0 & -10 \end{bmatrix}$$

No ponto $(0,2,0)$, *os menores principais são* $\Delta_1 = 0$, $\Delta_2 = 0$ *e* $\Delta_3 > 0$, *pelo que não se aplica nem o caso 1. nem o caso 2. do corolário anterior. Calculando as raízes do polinómio definido por* $\det(H_f(0,2,0) - \lambda I) = 0$, *isto é,*

$$\begin{vmatrix} -\lambda & 0 & 3 \\ 0 & -2-\lambda & 0 \\ 3 & 0 & -10-\lambda \end{vmatrix} = 0,$$

ou ainda, $(-2-\lambda)\left(10\lambda + \lambda^2 - 9\right) = 0$, *obtem-se* $\lambda = -2$ *ou* $\lambda = -5 + \sqrt{34}$ *ou* $\lambda = -5 - \sqrt{34}$, *o que mostra que a matriz hessiana admite dois valores próprios negativos e um positivo. É, portanto, indefinida e* $(0,2,0)$ *é um ponto de sela. No ponto* $(\frac{3}{10}, 2, \frac{9}{100})$, *os menores principais primários são* $\Delta_1 = -\frac{18}{10}$, $\Delta_2 = \frac{18}{5}$ *e* $\Delta_3 = -18$. *A matriz é, portanto, definida negativa e o ponto* $(\frac{3}{10}, 2, \frac{9}{100})$ *é um maximizante de* f.

Quando a matriz hessiana $H_f(\mathbf{a})$ é semidefinida, positiva ou negativa, não podemos tirar conclusões directamente. No entanto, para que o ponto $\mathbf{a}$ seja um maximizante de f, é preciso que $\mathbf{a}$ seja um maximizante de qualquer restrição de f. Esta condição é necessária, como ilustra o exemplo seguinte, não sendo, no entanto, suficiente.

Exemplo 76 *Seja* $f(x,y)$ *a função de duas variáveis reais definida por*

$$f(x,y) = \begin{cases} 0, & \text{se } y \leq x^2 \text{ ou } y \geq 2x^2 \\ x(2x^2-y)(y-x^2), & x^2 \leq y \leq x^2/2 \end{cases}$$

O ponto $(0,0)$ *é um ponto de sela, porque sobre a curva* $y = 3/2x^2$ *a função se torna* $1/4x^5$, *que admite valores positivos e negativos em qualquer vizinhança do ponto* $(0,0)$.

Se f for de classe C^k o teorema anterior tem um correspondente mais geral que referiremos, sem demonstração, em seguida (para mais pormenores consultar [1]).

Notação 77 *No estudo do Cálculo Diferencial em $\mathbb{R}^m$, para representar o "diferencial de ordem k da função f no ponto $\mathbf{a}$" usa-se a expressão*

$$D^k f(\mathbf{a})(\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^k), \ \mathbf{u}^i \in \mathbb{R}^m.$$

No que se segue, quando se pretender $\mathbf{u}^1 = \mathbf{u}^2 = \ldots = \mathbf{u}^k$, escreveremos $D^k f(\mathbf{a})(\mathbf{u})^k$ ou, mais simplesmente, $D^k f(\mathbf{a})(\mathbf{u})$, ou ainda, $D^k f(\mathbf{a})\mathbf{u}$.

Teorema 78 *Seja $\mathbf{a} \in \mathcal{D}$ ponto crítico de f. Se $\underline{n}$ for a ordem do primeiro diferencial de f que não é identicamente nulo em $\mathbf{a}$, isto é, se para algum $n \leq k$, existir $\mathbf{h} : D^n f(\mathbf{a})(\mathbf{h}) \neq 0$, e $D^j f(\mathbf{a}) \equiv 0$, se $1 \leq j < n$, então tem-se:*

1. se $\underline{n}$ for ímpar, $\mathbf{a}$ não é extremante local (ponto sela);

2. se $\underline{n}$ for par e se

a) $D^n f(\mathbf{a})\mathbf{h}$ é positivo para todos os $\mathbf{h} \neq \mathbf{0}$ (forma de grau n definida positiva), então $\mathbf{a}$ é ponto de mínimo local;

b) $D^n f(\mathbf{a})\mathbf{h}$ é negativo para todos os $\mathbf{h} \neq \mathbf{0}$ (forma de grau n definida negativa), então $\mathbf{a}$ é ponto de máximo local;

c) $D^n f(\mathbf{a})\mathbf{h}$ é positivo para certos $\mathbf{h} \neq \mathbf{0}$ e negativo para outros (forma de grau n indefinida), então $\mathbf{a}$ não é extremante local (ponto sela);

d) $D^n f(\mathbf{a})\mathbf{h} \geq 0$ mas existem $\mathbf{h} \neq \mathbf{0}$ constituindo direcções singulares, isto é, tais que $D^n f(\mathbf{a})\mathbf{h} = 0$ (forma de grau n semidefinida positiva), então

(d1) se existir uma direcção singular, $\mathbf{h}_0 \neq \mathbf{0}$, para a qual, a ordem p do primeiro dos diferenciais que não se anulam em $\mathbf{h}_0$ $(p > n)$, ou é ímpar ou, no caso de ser par, $D^p f(\mathbf{a})\mathbf{h}_0 < 0$, então a não é extremante local (ponto sela);

(d2) se, qualquer que seja a direcção singular, $\mathbf{h} \neq \mathbf{0}$, a ordem p do primeiro dos diferenciais que não se anulam em $\mathbf{h}$ $(p > n)$ é par e, além disso, $D^p f(\mathbf{a})\mathbf{h} > 0$ então nada se conclui (podendo o caso ser esclarecido por um estudo directo);

e) $D^n f(\mathbf{a})\mathbf{h} \leq 0$ mas existem $\mathbf{h} \neq \mathbf{0}$ constituindo direcções singulares, isto é, tais que $D^n f(\mathbf{a})\mathbf{h} = 0$ (forma de grau n semidefinida negativa), então

(e1) se existir uma direcção singular, $\mathbf{h}_0 \neq \mathbf{0}$, para a qual, a ordem p do primeiro dos diferenciais que não se anulam em $\mathbf{h}_0$ $(p > n)$, ou é ímpar ou, no caso de ser par, $D^p f(\mathbf{a})\mathbf{h}_0 > 0$, então $\mathbf{a}$ não é extremante local (ponto sela);

(e2) se, qualquer que seja a direcção singular, $\mathbf{h} \neq \mathbf{0}$, a ordem p do primeiro dos diferenciais que não se anulam em $\mathbf{h}$ $(p > n)$ é par e, além disso, $D^p f(\mathbf{a})\mathbf{h} < 0$ então nada se conclui (podendo o caso ser esclarecido por um estudo directo);

Concluindo, podemos resumir o resultado do teorema anterior do seguinte modo:

a não é extremante nos casos *1, 2 c, 2 d1* e *2 e1;*

a é minimizante no caso *2a*;

a é maximizante no caso *2b*;

nada se conclui nos casos *2 d2* e *2 e2.*

Exemplo 79 *Determinemos os extremantes da função*

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

1^o *Determinação dos pontos críticos de f*

$$\begin{cases} f'_x(x,y) = 4x^3 - 2(x-y) = 0 \\ f'_y(x,y) = 4y^3 + 2(x-y) = 0 \end{cases} \Longleftrightarrow \begin{cases} 4x^3 + 4y^3 = 0 \\ -- \end{cases}$$

$$\Longleftrightarrow \begin{cases} x = -y \\ 4y^3 - 4y = 0 \end{cases} \Longleftrightarrow \begin{cases} x = -y \\ y = 0, y = 1, y = -1 \end{cases}$$

Então os pontos críticos são $(0,0), (1,-1)$ *e* $(-1,1)$.

2^0 *Classificação dos pontos críticos*
A matriz hessiana é

$$\begin{bmatrix} 12x^2 - 2 & 2 \\ 2 & 12y^2 - 2 \end{bmatrix}.$$

$\cdot$ *Ponto* $(0,0)$

No ponto $(0,0)$*, a cadeia de menores é* $\Delta_1 = -2$, $\Delta_2 = 0$, *e, portanto, não se pode ainda tirar uma conclusão. Procuremos as direcções singulares. Seja* $\mathbf{h} = (h_1, h_2)$. *A expressão*

$$\begin{aligned} D^2f(0,0)(h_1,h_2) &= \frac{\partial^2 f}{\partial x^2}(0,0)h_1^2 + 2\frac{\partial^2 f}{\partial y \partial x}(0,0)h_1h_2 + \frac{\partial^2 f}{\partial y^2}(0,0)h_2^2 \\ &= -2 \times h_1^2 + 4h_1h_2 - 2 \times h_2^2 = -2(h_1 - h_2)^2 \ \leq \ 0 \end{aligned}$$

mostra que $D^2f(0,0)$ *é semidefinida negativa, e que se anula na direcção singular* $\mathbf{h}_0 = (h_1, h_2)$ *em que* $h_1 = h_2$. *Então estamos no caso* 2.*e*) *do teorema, mas há que averiguar se estamos em* (e_1) *ou* (e_2). *Vejamos então qual o sinal de* $D^3f(0,0)$ *ao longo da direcção singular* $\mathbf{h}_0 = (h_1, h_1)$. *Ora,*

$$D^3f(0,0)(h_1,h_1) = \frac{\partial^3 f}{\partial x^3}(0,0)h_1^3 + 3\frac{\partial^3 f}{\partial x^2 \partial y}(0,0)h_1^3 + 3\frac{\partial^3 f}{\partial x \partial y^2}(0,0)h_1^3 = 0,$$

pelo que não se pode concluir nada. Passemos então ao estudo do sinal de $D^4f(0,0)$ *ao longo da direcção singular* $\mathbf{h}_0 = (h_1, h_1)$. *Tem-se,*

$$D^4f(0,0)(h_1,h_1) = 48h_1^4 > 0, \qquad \text{se } h_1 \neq 0.$$

Logo verifica-se o caso (e_1) *e, portanto,* $(0,0)$ *não é extremante local, é ponto sela.*

· *Pontos* $(1,-1)$ *e* $(-1,1)$

Quer no ponto $(1,-1)$, *quer no ponto* $(-1,1)$ *a cadeia de menores é* $\Delta_1 = 10$, $\Delta_2 = 96$, *e, portanto,* $D^2f(1,-1)$ *e* $D^2f(-1,1)$ *são ambos formas quadráticas definidas positivas. Então, estamos no caso 2.b) do teorema. Logo* $(1,-1)$ *e* $(-1,1)$ *são minimizantes locais de* f.
O mínimo local de f *é* $f(1,-1) = f(-1,1) = -2$.

Exemplo 80 *Mostrar que* $(0,0)$ *é ponto crítico de* f *mas não é extremante*

$$f(x,y) = x^2 - 2xy^2 + y^4 - 2y^5$$

1° *$(0,0)$ é ponto crítico de f.*
As derivadas parciais são

$$\begin{cases} f'_x(x,y) = 2x - 2y^2 \\ f'_y(x,y) = -4xy + 4y^3 - 10y^4 \end{cases}$$

e anulam-se no ponto $(0,0)$ *que é, portanto, ponto crítico de* f.

2° *Classificação do ponto crítico $(0,0)$.*
A matriz hessiana de f *é*

$$\begin{bmatrix} 2 & -4y \\ \\ -4y & -4x + 12y^2 - 40y^3 \end{bmatrix}$$

e, no ponto $(0,0)$, *a cadeia de menores é* $\Delta_1 = 2$, $\Delta_2 = 0$, *e portanto não se pode tirar conclusão. Procuremos as direccões singulares. Seja* $h = (h_1, h_2)$. *A expressão*

$$\begin{aligned} D^2f(0,0)(h_1,h_2) &= \frac{\partial^2 f}{\partial x^2}(0,0)h_1^2 + 2\frac{\partial^2 f}{\partial x \partial y}(0,0)h_1h_2 + \frac{\partial^2 f}{\partial y^2}(0,0)h_2^2 \\ &= 2h_1^2 + 2 \times 0 \times h_1h_2 + 0 \times h_2^2 = 2h_1^2 \end{aligned}$$

confirma que $D^2f(0,0)$ *é semidefinida positiva e indica que a direcção singular é dada por*

$$\mathbf{h}_0 = (0, h_2).$$

Estamos, então, no ponto 2.d) do teorema e vamos determinar se será (d_1) ou (d_2). Vejamos qual é a ordem do primeiro diferencial que não se anula ao longo da direcção singular. Ora,

$$\begin{aligned} D^3f(0,0)(0,h_2) &= \frac{\partial^3 f}{\partial x^3}(0,0)\times 0^2 + 3\frac{\partial^3 f}{\partial x^2\partial y}(0,0)\times 0^2\times h_2 + \\ &\quad +3\frac{\partial^3 f}{\partial x\partial y^2}(0,0)\times 0\times h_2^2 + \frac{\partial^3 f}{\partial y^3}(0,0)h_2^3 \\ &= \frac{\partial^3 f}{\partial y^3}(0,0)h_2^3 = 0\times h_2^3 \;=\; 0 \end{aligned}$$

Então $D^3f(0,0)$ anula-se ao longo da direcção singular $h_0 = (0,h_2)$, mas passando ao cálculo de $D^4f(0,0)(0,h_2)$ tem-se

$$D^4f(0,0)(0,h_2) = 24h_2^4 > 0$$

pelo que se aplica o caso (d_2). Logo nada se pode concluir pelo teorema. Observemos que a função f pode ser escrita como

$$f(x,y) = x^2 - 2xy^2 + y^4 - 2y^5 = (x-y^2)^2 - 2y^5,$$

o que mostra que, ao longo das curvas $y = \pm\sqrt{x}$, $x \geq 0$, se obtem $f(x,\sqrt{x}) = -2\sqrt{x}^5 < 0$ e $f(x,-\sqrt{x}) = 2\sqrt{x}^5 > 0$, se $x > 0$. Logo em qualquer vizinhança do ponto $(0,0)$ a função f toma valores positivos e negativos, pelo que $(0,0)$ é ponto sela.

4.4 Extremos condicionados por igualdades

Neste parágrafo, procuramos maximizar ou minimizar uma função f sujeita a condições adicionais suscitadas por situações específicas traduzidas por igualdades, o que constitui a base da optimização com restrições de igualdade. Isto leva-nos ao estudo da existência de máximos ou mínimo de funções em conjuntos não abertos. Mais precisamente, dada uma função $f:\mathbb{R}^m \to \mathbb{R}$, dita *função objectivo* e p funções, $F_1,\ldots,F_p$ de $\mathbb{R}^m$ em $\mathbb{R}$ $(p < m)$, põe-se a questão de determinar os extremos de f no conjunto M, definido por

$$M = \{\mathbf{x} = (x_1,\ldots,x_m) \in \mathbb{R}^m : F_1(x_1,\ldots,x_m) = 0,\ldots,F_p(x_1,\ldots,x_m) = 0\}$$

isto é, os extremos de $f|_M$. Dito de outro modo, pretende-se determinar o máximos ou mínimos locais ou globais, caso existam, de f sujeita às condições definidas pelas igualdades

$$F_1(x_1,\ldots,x_m) = 0;\; F_2(x_1,\ldots,x_m) = 0;\ldots;F_p(x_1,\ldots,x_m) = 0.$$

Os problemas deste tipo são designados por problemas de **extremos condicionados.**

Exemplo 81 *Determinar o ponto da elipse* $4x^2 + y^2 = 1$ *mais próximo do ponto de coordenadas* $(8, -5)$ *equivale a minimizar a função* $f(x, y) = (x-8)^2 + (y+5)^2$ *sujeita à restrição* $F(x, y) = 4x^2 + y^2 - 1 = 0$.

O teorema seguinte contém uma condição necessária para que um ponto seja um extremante local sob diferentes condições.

Teorema 82 *Sejam* $f, F_1, \ldots, F_p$ *funções definidas num aberto* $\mathcal{D}$ *de* $\mathbb{R}^m$, *com valores em* $\mathbb{R}$ *e de classe* C^1. *Se* $\mathbf{a} \in \mathcal{D}$ *for um extremante local sujeito às condições* $F_i = 0$ *e se a matriz*

$$\begin{bmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{a}) & \frac{\partial F_1}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_1}{\partial x_m}(\mathbf{a}) \\ \frac{\partial F_2}{\partial x_1}(\mathbf{a}) & \frac{\partial F_2}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_2}{\partial x_m}(\mathbf{a}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial F_p}{\partial x_1}(\mathbf{a}) & \frac{\partial F_p}{\partial x_2}(\mathbf{a}) & \cdots & \frac{\partial F_p}{\partial x_m}(\mathbf{a}) \end{bmatrix}$$

tiver característica p, *então existem constantes* $\lambda_1, \lambda_2, \ldots, \lambda_p$ *tais que*

$$\nabla f(\mathbf{a}) + \sum_{i=1}^{p} \lambda_i \nabla F_i(\mathbf{a}) = \mathbf{0}. \tag{4.1}$$

Dem Apresentamos a demonstração no caso particular de uma função de duas variáveis e uma condição apenas, $F_1 = 0$. Seja $\mathbf{a} = (a_1, a_2)$ um extremante local de f tal que $F_1(\mathbf{a}) = 0$, e $\frac{\partial F_1}{\partial x_2}(\mathbf{a}) \neq 0$ (seria análogo se $\frac{\partial F_1}{\partial x_1}(\mathbf{a}) \neq 0$). Então, pelo teorema da função implícita, existe uma única função φ, de classe C^1, definida numa vizinhança de a_1, $B_\epsilon(a_1)$, tal que

$$\begin{aligned} a_2 &= \varphi(a_1) \\ \varphi'(a_1) &= -\frac{\frac{\partial F_1}{\partial x_1}(a_1, a_2)}{\frac{\partial F_1}{\partial x_2}(a_1, a_2)} \end{aligned}$$

e, para os pontos $\mathbf{x} = (x_1, x_2)$ próximos de $\mathbf{a} = (a_1, a_2)$,

$$F(x_1, x_2) = 0 \iff x_2 = \varphi(x_1).$$

Considerando a função de classe C^1 definida em $B_\epsilon(a_1)$ por

$$\Psi(x_1) = f(x_1, x_2) = f(x_1, \varphi(x_1)),$$

temos que a_1 é extremante de Ψ. Logo $\Psi'(a_1) = 0$. Pela derivada da função composta,

$$\begin{aligned}
\Psi'(a_1) = 0 &\iff \frac{\partial f}{\partial x_1}(a_1, \varphi(a_1)) + \frac{\partial f}{\partial x_2}(a_1, \varphi(a_1)).\varphi'(a_1) = 0 \\
&\iff \frac{\partial f}{\partial x_1}(a_1, a_2) - \frac{\partial f}{\partial x_2}(a_1, a_2).\frac{\frac{\partial F_1}{\partial x_1}(a_1, a_2)}{\frac{\partial F_1}{\partial x_2}(a_1, a_2)} = 0 \\
&\iff \frac{\partial f}{\partial x_1}(a_1, a_2) - \frac{\frac{\partial f}{\partial x_2}(a_1, a_2)}{\frac{\partial F_1}{\partial x_2}(a_1, a_2)}.\frac{\partial F_1}{\partial x_1}(a_1, a_2) = 0.
\end{aligned}$$

Então, pondo

$$\lambda = -\frac{\frac{\partial f}{\partial x_2}(a_1, a_2)}{\frac{\partial F_1}{\partial x_2}(a_1, a_2)},$$

obtemos

$$\begin{aligned}
\frac{\partial f}{\partial x_1}(a_1, a_2) + \lambda.\frac{\partial F_1}{\partial x_1}(a_1, a_2) &= 0 \\
\frac{\partial f}{\partial x_2}(a_1, a_2) + \lambda.\frac{\partial F_1}{\partial x_2}(a_1, a_2) &= 0,
\end{aligned}$$

o que mostra que existe λ tal que

$$\nabla f(\mathbf{a}) - \lambda \nabla F(\mathbf{a}) = \mathbf{0}.$$

$\square$

Definição 83 *Os números λ_i dados pelo teorema anterior chamam-se **multiplicadores de Lagrange** e a condição 4.1 exprime o facto de* **a** *ser **ponto crítico de** $f|_M$.*

O teorema anterior indicia, pois, um método de pesquisa de pontos de extremo de $f|_M$, conhecido por: **Método dos multiplicadores de Lagrange.**

Consideremos a função $\mathcal{F}$ definida por

$$\mathcal{F}(x_1, \ldots, x_m; \lambda_1, \ldots, \lambda_p) =$$

$$= f(x_1, \ldots, x_m) + \lambda_1 F_1(x_1, \ldots, x_m) + \lambda_2 F_2(x_1, \ldots, x_m) + \ldots + \lambda_p F_p(x_1, \ldots, x_m).$$

Esta função chama-se **função lagrangeana**. Consideremos então o sistema de $(m+p)$ equações que se obtêm igualando a zero as derivadas da função lagrangeana em ordem às variáveis $x_1, x_2, \ldots, x_m, \lambda_1, \ldots, \lambda_p$,

$$\begin{cases} \dfrac{\partial f}{\partial x_1} + \displaystyle\sum_i \lambda_i \dfrac{\partial F_i}{\partial x_1} = 0 & \text{(derivada em ordem a } x_1\text{)} \\ \dfrac{\partial f}{\partial x_2} + \displaystyle\sum_i \lambda_i \dfrac{\partial F_i}{\partial x_2} = 0 & \text{(derivada em ordem a } x_2\text{)} \\ \vdots & \\ \dfrac{\partial f}{\partial x_m} + \displaystyle\sum_i \lambda_i \dfrac{\partial F_i}{\partial x_m} = 0 & \text{(derivada em ordem a } x_m\text{)} \\ F_1 = 0 & \text{(derivada em ordem a } \lambda_1\text{)} \\ \vdots & \\ F_p = 0 & \text{(derivada em ordem a } \lambda_p\text{)} \end{cases}$$

As soluções deste sistema fornecem os pontos críticos de $f|_M$, bem como os multiplicadores de Lagrange associados.

Nota 84 *Se o conjunto M definido pelas condições $F_1, \ldots F_p$ for um compacto de $\mathbb{R}^m$, isto é, for fechado e limitado, uma vez que a função objectivo f é contínua, então f possui necessariamente um máximo e um mínimo absolutos em M (teorema de Weierstrass). Estes valores são facilmente identificáveis através do cálculo dos valores de f nos diversos pontos críticos de $f|_M$. Nas aplicações, é, em geral, extremos absolutos que se pretende determinar.*

Interpretação dos multiplicadores de Lagrange

Os multiplicadores de Lagrange, parecendo não exibir nenhum papel especial na determinação dos possíveis pontos críticos de $f|_M$ e, consequentemente, nos extremos condicionados, fornecem contudo uma interpretação valiosa nas aplicações sobre a variação do valor dos extremos de f em M, de acordo com a variação das condições. Isto é, por exemplo: se a cada $\mathbf{b} = (b_1, \ldots, b_p) \in \mathbb{R}^p$ definirmos $\theta(\mathbf{b})$ como sendo o máximo da função f em $M_{\mathbf{b}}$ onde

$$M_{\mathbf{b}} = \left\{\mathbf{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m : \gamma_1(x_1, \ldots, x_m) = b_1, \ldots, \gamma_p(x_1, \ldots, x_m) = b_p\right\}$$

o que na notação acima corresponde a

$$F_i\,(x_1, \ldots, x_m) = b_i - \gamma_i(x_1, \ldots, x_m) = 0,$$

pode provar-se que

$$\frac{\partial \theta}{\partial b_j}\,(\mathbf{b}) = \lambda_j.$$

Este facto dá-nos pois informação sobre a variação do óptimo de f em $M_{\mathbf{b}}$, de acordo com a variação de b_j, isto é, de acordo com a variação da restrição γ_j, o que será útil nas aplicações se tivermos de fazer uma análise de tendência.

Condições de segunda ordem: a hessiana orlada

O método dos multiplicadores de Lagrange fornece condições necessárias para que um ponto **a** seja um extremo condicionado local de uma função f sujeita às condições $F_1 = 0, \ldots, F_p = 0$. A utilização da hessiana de função lagrangeana vai permitir dar condições suficientes para decidir se, localmente, um ponto é um ponto de máximo, um ponto de mínimo ou um ponto sela com restrições (para mais pormenores consultar [L]).
Designemos por

- $H_{\mathcal{F}}(\mathbf{a})$ a matriz hessiana da função lagrangeana enquanto função das variáveis $x_1, x_2, \ldots, x_m$,

$$\mathcal{F} = f + \lambda_1 F_1 + \lambda_2 F_2 + \ldots + \lambda_p F_p,$$

- $J_F(\mathbf{a})$ a matriz jacobiana da aplicação $F = (F_1, \ldots, F_p)\,$,
- $J_F^t(\mathbf{a})$ a matriz transposta de $J_F(\mathbf{a})$.

Consideremos a matriz

$$H(\lambda, \mathbf{a}) = \begin{bmatrix} 0 & \cdots & 0 & \dfrac{\partial F_1}{\partial x_1}(\mathbf{a}) & \cdots & \dfrac{\partial F_1}{\partial x_m}(\mathbf{a}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \dfrac{\partial F_p}{\partial x_1}(\mathbf{a}) & \cdots & \dfrac{\partial F_p}{\partial x_m}(\mathbf{a}) \\ & & & & & \\ \dfrac{\partial F_1}{\partial x_1}(\mathbf{a}) & \cdots & \dfrac{\partial F_p}{\partial x_1}(\mathbf{a}) & \dfrac{\partial^2 \mathcal{F}}{\partial x_1^2}(\mathbf{a}) & \cdots & \dfrac{\partial^2 \mathcal{F}}{\partial x_1 \partial x_m}(\mathbf{a}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial F_1}{\partial x_m}(\mathbf{a}) & \cdots & \dfrac{\partial F_p}{\partial x_m}(\mathbf{a}) & \dfrac{\partial^2 \mathcal{F}}{\partial x_m \partial x_1}(\mathbf{a}) & \cdots & \dfrac{\partial^2 \mathcal{F}}{\partial x_m^2}(\mathbf{a}) \end{bmatrix}$$

isto é, abreviadamente,

$$H(\lambda, \mathbf{a}) = \begin{bmatrix} 0 & J_F(\mathbf{a}) \\ J_F^t(\mathbf{a}) & H_{\mathcal{F}}(\mathbf{a}) \end{bmatrix}.$$

Determinemos os menores principais de ordens $2p+1, 2p+2, \ldots, p+m$, mais precisamente,

$$\Delta_{2p+1}, \Delta_{2p+2}, \ldots, \Delta_{p+m}$$

e calculemos a seguinte sucessão de $m-p$ números

$$(-1)^p \Delta_{2p+1}, (-1)^p \Delta_{2p+2}, \ldots, (-1)^p \Delta_{p+m}.$$

A análise dos sinais destes números permitir-nos-á tirar conclusões quanto à natureza do ponto crítico **a**. Assim, temos

1. Se forem todos positivos, então **a** é ponto de mínimo condicionado de f

2. Se $(-1)^p \Delta_{2p+1} < 0$ e daí em diante os sinais alternarem então **a** é ponto de máximo condicionado de f

Este critério é conhecido por **critério da hessiana orlada**. Observe-se, no entanto, que existem muitas situações a que este critério não dá resposta.

Exemplo 85 *Sejam as funções f, g_1 e g_2 definidas por*

$$\begin{aligned} f(x_1, x_2, x_3) &= x_1^2 + x_2^2 + x_3^2 \\ g_1(x_1, x_2, x_3) &= x_1 + x_2 + x_3 - 2 \\ g_2(x_1, x_2, x_3) &= x_1 + 3x_2 + x_3 + 2. \end{aligned}$$

Procuremos encontrar os extremos locais de f sujeita aos constrangimentos $g_1(\mathbf{x}) = g_2(\mathbf{x}) = 0$. A função lagrangeana correspondente é dada por

$$\mathcal{F}(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + \lambda_1(x_1 + x_2 + x_3 - 2) + \lambda_2(x_1 + 3x_2 + x_3 + 2)$$

Se $\mathbf{a} = (a_1, a_2, a_3)$ for um extremo local, então existem λ_1 e λ_2 tais que

$$\begin{aligned} 2a_1 + \lambda_1 + \lambda_2 = 0 \\ 2a_2 + \lambda_1 + 3\lambda_2 = 0 \\ 2a_3 + \lambda_1 + \lambda_2 = 0 \end{aligned}$$

Como, além disso, **a** *deve satisfazer as condições*

$$g_1(a_1, a_2, a_3) = 0 \text{ e } g_2(x_1, x_2, x_3) = 0,$$

devemos também ter

$$a_1 + a_2 + a_3 = 2$$
$$a_1 + 3a_2 + a_3 = -2,$$

donde obtemos facilmente $a_1 = -a_2 = a_3 = 2$, $\lambda_1 = -8$ *e* $\lambda_2 = 4$. *A hessiana orlada em* $(2, -2, 2)$ *é dada por*

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 3 & 1 \\ 1 & 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 & 2 \end{bmatrix}.$$

Aplicando o critério anterior, é apenas preciso conhecer o sinal do menor $\Delta_5(\mathbf{a})$, *que, neste caso, é o determinante da hessiana orlada, pois* $p = 2$. *Ora, como* $\Delta_5(\mathbf{a}) = 16$, *obtem-se* $(-1)^2\Delta_5(\mathbf{a}) > 0$, *donde* $\mathbf{a}$ *é um mínimo local de* f *sujeito às condições dadas por* g_1 *e* g_2. *O ponto* $\mathbf{a}$ *não é mais que o ponto da recta de equações*

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_1 + 3x_2 + x_3 = -2 \end{cases}$$

mais próximo da origem, já que a função f *é o quadrado da distância à origem. Considerações geométricas ter-nos-iam, portanto, permitido saber, sem utilizar este critério, que o único ponto candidato tinha de ser um mínimo.*

Exemplo 86 *Sejam as funções* f *e* g *definidas por*

$$\begin{aligned} f(x_1, x_2, x_3) &= x_1^2 + x_2^2 + x_3^2 \\ g(x_1, x_2, x_3) &= x_1 + 3x_2 + x_3 + 2. \end{aligned}$$

Procuremos encontrar os extremos locais de f *sujeitos à condição* $g(\mathbf{x}) = 0$. *A função lagrangeana correspondente é dada por*

$$\mathcal{F}(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 + \lambda(x_1 + 3x_2 + x_3 + 2).$$

Se $\mathbf{a} = (a_1, a_2, a_3)$ *for um extremo local, então existe* λ *tal que*

$$2a_1 + \lambda = 0$$
$$2a_2 + 3\lambda = 0$$
$$2a_3 + \lambda = 0.$$

Como, além disso, o ponto $\mathbf{a}$ *deve satisfazer a condição* $g(\mathbf{a}) = 0$*, devemos também ter*

$$a_1 + 3a_2 + a_3 = -2.$$

A partir destas quatro equações obtém-se os valores

$$a_1 = -\frac{2}{11}, a_2 = -\frac{6}{11}, a_3 = -\frac{2}{11} \text{ e } \lambda = -\frac{4}{11}.$$

A hessiana orlada no ponto $\mathbf{a} = (-\frac{2}{11}, -\frac{6}{11}, -\frac{2}{11})$ *é dada por*

$$\begin{bmatrix} 0 & 1 & 3 & 1 \\ 1 & 2 & 0 & 0 \\ 3 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix}$$

Como $p = 1$*, e os menores* $\Delta_3(\mathbf{a}) = -20$ *e* $\Delta_4(\mathbf{a}) = -44$*, sai que* $-\Delta_3(\mathbf{a}) > 0$ *e* $-\Delta_4(\mathbf{a}) > 0$*, donde* $\mathbf{a}$ *é ponto de mínimo local sujeito à condição dada por* g*. De facto, encontrámos o ponto do plano de equação* $x_1 + 3x_2 + x_3 = -2$ *mais próximo da origem. Aqui também, considerações geométricas ter-nos-iam permitido saber que o único ponto candidato devia ser um mínimo.*

4.5 Extremos condicionados por desigualdades

Consideremos o conjunto definido em termos de igualdades e desigualdades

$$M = \left\{ \mathbf{x} \in \mathbb{R}^m : \left\langle \begin{array}{l} F_1(x_1, \ldots, x_m) = 0, \ldots, F_p(x_1, \ldots, x_m) = 0, \\ \phi_1(x_1, \ldots, x_m) \leq 0, \ldots, \phi_k(x_1, \ldots, x_m) \leq 0, \end{array} \right. \right\}, \quad (4.2)$$

$\mathbf{x} = (x_1, \ldots, x_m)$. Nesta secção tratar-se-á da determinação de extremos de f no conjunto M, isto é, os extremos de $f|_M$. Dito de outro modo, pretende-se determinar o máximo ou mínimo, caso existam, de f sujeita às condições definidas pelas igualdades

$$F_1(x_1, \ldots, x_m) = 0,\ F_2(x_1, \ldots, x_m) = 0, \ldots, F_p(x_1, \ldots, x_m) = 0$$

e pelas desigualdades

$$\phi_1(x_1, \ldots, x_m) \leq 0, \phi_2(x_1, \ldots, x_m) \leq 0, \ldots, \phi_k(x_1, \ldots, x_m) \leq 0.$$

Seja $\mathbf{a} \in M$. Então podemos escrever, renumerando as funções ϕ_j,

$$\left\{ \begin{array}{l} F_1(\mathbf{a}) = 0, \ldots, F_p(\mathbf{a}) = 0, \\ \phi_1(\mathbf{a}) = 0, \ldots, \phi_s(\mathbf{a}) = 0, \end{array} \right. \quad (4.3)$$

$$\phi_{s+1}(\mathbf{a}) < 0, \ldots, \phi_k(\mathbf{a}) < 0. \quad (4.4)$$

As condições de igualdade, dizem-se **restrições activas** e as de desigualdade estrita, dizem-se **restrições inactivas** (no ponto $\mathbf{a}$). Em tudo o que segue suporemos realizada a hipótese seguinte:

$$\begin{array}{l}\text{os gradientes } \nabla F_i(a), \text{ para } i=1,\ldots,p \text{ e } \nabla\phi_j(\mathbf{a}), \text{ para } j=1,\ldots,s \\ \text{(isto é, das restrições activas) são linearmente independentes.}\end{array} \tag{4.5}$$

Por outras palavras, (4.5) significa que o jacobiano das restrições activas

$$\frac{\partial(F_1,\ldots,F_p,\phi_1,\ldots,\phi_s)}{\partial(x_1,\ldots,x_n)}(\mathbf{a})$$

tem característica $p+s$. Apresentamos sem demonstração o resultados seguinte (para mais pormenores consultar [3],[6]).

Teorema 87 (Kuhn-Tucker) *Seja $f:\ \mathbb{R}^m \to \mathbb{R}$ uma função de classe $\mathcal{C}^1$ e M definido por (4.2). Suponhamos que $f|_M$ tem um mínimo local no ponto $\mathbf{a}\in M$ e que é satisfeita a hipótese (4.5) relativa às condições activas em $\mathbf{a}$. Então, existem números $\lambda_1,\ldots,\lambda_P$, $\mu_1,\ldots,\mu_P$ tais que*

$$\begin{cases} \nabla f(\mathbf{a}) + \sum\limits_{i=1}^{p}\lambda_i\nabla F_i(\mathbf{a}) + \sum\limits_{j=1}^{k}\mu_j\nabla\phi_j(\mathbf{a}) = 0 \\ \mu_j \geq 0 \\ \mu_j\phi_j(\mathbf{a}) = 0,\ j=1,\ldots,P \end{cases} \tag{4.6}$$

Corolário 88 *Se, nas mesmas condições, $f|_M$ tem um máximo local em $\mathbf{a}$, verificam-se condições análogas a (4.6) com "$\mu_j \geq 0$" substituídas por "$\mu_j \leq 0$".*

Exemplo 89 *Determinar o mínimo da função $f(x,y) = x^3+y^3+z^3$ no triângulo $M = \left\{(x,y,z)\in\mathbb{R}^3 : x+y+z=1, x\geq 0, y\geq 0, z\geq 0\right\}$.*
Sendo

$$\begin{aligned} F(x,y,z) &= x+y+z, \\ \phi_1(x,y,z) &= -x,\ \phi_2(x,y,z) = -y,\ \phi_3(x,y,z) = -z \end{aligned}$$

então um mínimo de f em M, (x,y,z), deverá satisfazer,

$$\begin{cases} f'_x(x,y,z) + \lambda F'_x(x,y,z) + \mu_1\phi'_{1_x}(x,y,z) = 0 \\ f'_x(x,y,z) + \lambda F'_x(x,y,z) + \mu_2\phi'_{2_x}(x,y,z) = 0 \\ f'_x(x,y,z) + \lambda F'_x(x,y,z) + \mu_3\phi'_{3_x}(x,y,z) = 0 \\ \mu_1,\mu_2,\mu_3 \geq 0 \\ \mu_1\phi_1(x,y,z)=0,\ \mu_2\phi_2(x,y,z)=0,\ \mu_3\phi_3(x,y,z)=0, \end{cases}$$

com $\lambda, \mu_1, \mu_2, \mu_3 \in \mathbb{R}$. *Isto é,*

$$\begin{cases} 3x^2 + \lambda - \mu_1 = 0 \\ 3y^2 + \lambda - \mu_2 = 0 \\ 3z^2 + \lambda - \mu_3 = 0 \\ \mu_1, \mu_2, \mu_3 \geq 0 \\ \mu_1 x = 0 \\ \mu_2 y = 0 \\ \mu_3 z = 0. \end{cases}$$

Se $\mu_1 = \mu_2 = \mu_3 = 0$, *então, pelas três primeiras equações,*

$$3x^2 = -\lambda = 3y^2 = 3z^2,$$

donde, $x = \pm y = \pm z$. *Como o ponto tem de pertencer ao conjunto* M, *sai que* $x = y = z \geq 0$, *e, portanto,* $3x = 1$, *donde o ponto crítico determinado é* $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. *Se* $\mu_1 = 0, \mu_2 \neq 0$ *e* $\mu_3 \neq 0$, *então* $y = 0 = z$, *donde,* $x = 1$. *Então* $\lambda = -3$ *e, portanto,* $\mu_2 = -3$ *e* $\mu_3 = -3$, *o que não interessa. Se* $\mu_1 \neq 0, \mu_2 = 0$ *e* $\mu_3 \neq 0$ *ou* $\mu_1 \neq 0, \mu_2 \neq 0$ *e* $\mu_3 = 0$, *a situação é análoga. Suponhamos agora que* $\mu_1 = 0, \mu_2 = 0$ *e* $\mu_3 \neq 0$. *Então* $z = 0$ *e* $\lambda = \mu_3$ *e* $3x^2 = -\lambda = 3y^2$, *donde, como o ponto pertence a* M, $x = y \geq 0$ *e, portanto,* $2x = 1$. *Logo* $x = \frac{1}{2}$ *e* $\mu_3 = \lambda = -\frac{3}{4}$, *o que não interessa. Se* $\mu_1 = 0, \mu_2 \neq 0$ *e* $\mu_3 = 0$ *ou* $\mu_1 \neq 0, \mu_2 = 0$ *e* $\mu_3 = 0$, *a situação é análoga. Logo o mínimo de* f *em* M *é atingido no ponto* $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ *e tem o valor* $f(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) = \frac{1}{9}$.

Vimos condições necessárias para que $\mathbf{a}$ seja um mínimo local de $f|_M$. Vamos ver que são também suficientes num caso especial. Para isso, introduzamos primeiramente algumas noções relativas a convexidade.

4.5.1 Convexidade

Um conjunto $A \subseteq \mathbb{R}^m$ diz-se **convexo** se

$$\forall \mathbf{x}, \mathbf{y} \in A, \forall \alpha \in [0,1], \quad \alpha \mathbf{x} + (1-\alpha)\mathbf{y} \in A.$$

Geometricamente, isto significa que, dado qualquer par de pontos de A, o segmento que os une está contido em A (ver 4.1).

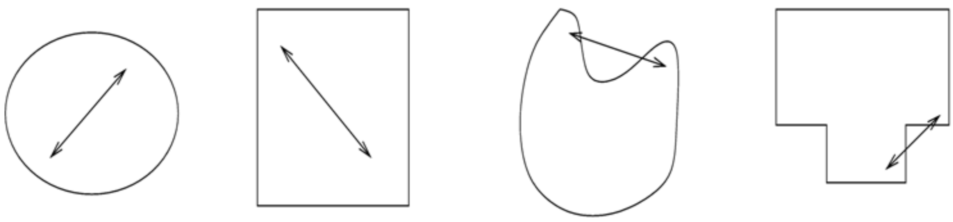


Figura 4.1:

convexo convexo não convexo não convexo

Definição 90 *Seja $A \subseteq \mathbb{R}^m$ um conjunto convexo. Uma função*

$$f : A \subseteq \mathbb{R}^m \to \mathbb{R}$$

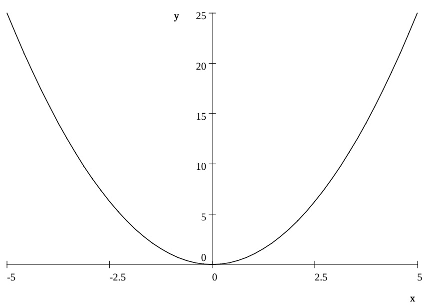
diz-se ***convexa*** *em A se*

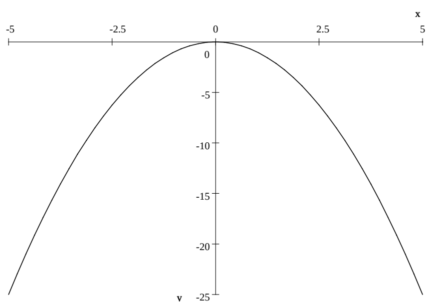
$$\forall \mathbf{x}, \mathbf{y} \in A, \forall \alpha \in [0,1], \quad f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

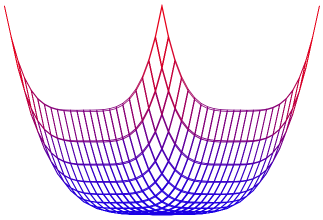
Analogamente, uma função $f : A \subseteq \mathbb{R}^m \to \mathbb{R}$ diz-se ***côncava*** *em A se*

$$\forall \mathbf{x}, \mathbf{y} \in A, \forall \alpha \in [0,1], \quad f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y}) \geq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y})$$

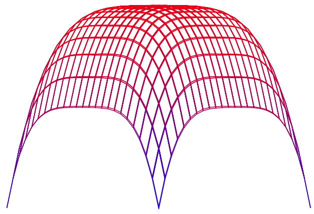
No caso $n = 2$, visualiza-se geometricamente que, se uma função $z = f(x, y)$ for **côncava (convexa)**, então, para todo o par de pontos M e N do gráfico, o segmento que os une nunca se encontra acima (abaixo) do gráfico.
Como ilustram os gráficos seguintes, as funções x^2 e $x^4 + y^4$ são convexas e $-x^2$ e $-x^4 - y^4$ são côncavas, em $\mathbb{R}$ e $\mathbb{R}^2$, respectivamente.

x^2 

$-x^2$ 



$x^4 + y^4$



$-x^4 - y^4$

Admitiremos sem demonstração o resultado seguinte.

Teorema 91 *Seja f uma função de classe $\mathcal{C}^2$ definida num aberto A convexo em $\mathbb{R}^m$. Então*

$$\begin{aligned} &f \text{ é convexa em } A \quad \Leftrightarrow \quad \text{hessiana DP ou SDP em } A,\\ &f \text{ é côncava em } A \quad \Leftrightarrow \quad \text{hessiana DN ou SDN em } A. \end{aligned}$$

Suponhamos que M é da forma

$$M = \{\mathbf{x} \in \mathbb{R}^m : \ \phi_j(\mathbf{x}) \le 0 \text{ para } 1 \le j \le k\} \tag{4.7}$$

onde ϕ_j são funções *convexas* de classe $\mathcal{C}^1$, $j = 1, \ldots, k$. Então M é convexo, como é fácil de verificar, e é válido o seguinte resultado sobre existência de um mínimo local de $f|_M$.

Teorema 92 *Se M é dado por (4.7) e f é uma função convexa de classe $\mathcal{C}^1$ em $\mathbb{R}^m$, então $\mathbf{a} \in M$ é um ponto de mínimo absoluto de $f|_M$ se, e só se, supondo verificada a hipótese (4.5) a respeito das restrições activas em* $\mathbf{a}$, $(\phi_j(\mathbf{a}) = 0)$*, existem escalares $\mu_1, \ldots, \mu_k$, tais que*

$$\nabla f(\mathbf{a}) + \sum_{j=1}^{k} \mu_j \nabla \phi_j(\mathbf{a}) = 0, \ \mu_j \ge 0 \ \text{ e } \mu_j \phi_j(\mathbf{a}) = 0 \tag{4.8}$$

Dem A condição é necessária precisamente pelo teorema de Kuhn-Tucker. Para mostrarmos que, reciprocamente, (4.8) implica que $f(\mathbf{a})$ é o mínimo de f em M, basta mostrar que, se $\phi_j(\mathbf{a}) = 0$

$$b \in M \Rightarrow \nabla \phi_j(\mathbf{a}) \cdot (b - \mathbf{a}) \le 0 \tag{4.9}$$

porque então tem-se $\nabla f(\mathbf{a}) \cdot (b - \mathbf{a}) \ge 0 \ \forall b \in M$. Provemos então (4.9). Consideremos a função $t \mapsto \phi_j(\mathbf{a} + t(b - \mathbf{a}))$, $0 \le t \le 1$. Como $\phi_j(\mathbf{a}) = 0$ e $\mathbf{a} + t(b - \mathbf{a}) \in M \ \forall t \in [0, 1]$, tem-se $\phi_j(\mathbf{a} + t(b - \mathbf{a}))$, e portanto

$$\frac{d}{dt} \, \phi_j(\mathbf{a} + t(b - \mathbf{a}))\Big|_{t=0} \le 0.$$

Ora a derivada que figura no 1.º membro é precisamente $\nabla \phi_j(\mathbf{a}) \cdot (b - \mathbf{a})$, o que termina a demonstração. $\square$

4.6 Problemas de extremo e optimização em Economia

Problema 1 O lucro anual de uma empresa é dado pela expressão seguinte

$$L(x, y) = -x^2 - y^2 + 22x + 18y - 102$$

onde x representa o montante gasto em investigação e y o montante gasto em publicidade (L, x e y expressos em unidades de milhões de euros). Determine o lucro máximo e os valores de x e y que o realizam.

Problema 2 Considere as funções Produção e Custo de uma empresa dadas por

$$\begin{aligned} P &= 3K^{1/3}L^{1/3} \\ C &= K^2 + 2L + 8 \end{aligned}$$

onde K e L. Calcule o custo mínimo para uma produção de 12 unidades.

Problema 3 Designemos por x e y as quantidades de produtos comprados por um consumidor e a respectiva função utilidade representada por

$$U(x, y) = 6xy - 2x.$$

Calcule a utilidade máxima do consumidor, sabendo que a sua restrição orçamental é dada por $8x + 3y = 17$.

Problema 4 (o problema do consumidor) Considere a função utilidade $U(x, y)$ relativa a um consumidor que adquire bens materiais x e y mas cujos gastos não podem exceder o seu rendimento R. Formalize matematicamente o problema da maximização da função utilidade.

Apêndice A

Complementos de Álgebra Linear

A.1 Introdução

Seja $A = [a_{ij}]$ uma matriz quadrada de ordem n e pensemos na transformação linear que a cada cada vector $\mathbf{u} \in \mathbb{R}^n$ faz corresponder um vector imagem em $\mathbb{R}^n$ definido por $A\mathbf{u}$.

Exemplo 93 *Sendo* $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ *e* $\mathbf{u} = \begin{bmatrix} 5 \\ 1 \\ 3 \end{bmatrix}$ *temos que*

$$A\mathbf{u} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 7 \\ 2 \\ 9 \end{bmatrix}.$$

Põe-se a questão de saber se existe algum vector $\mathbf{u} \in \mathbb{R}^n$, cuja imagem pela transformação da matriz A, ou seja $A\mathbf{u}$, tenha a mesma direcção do vector $\mathbf{u}$, isto é, seja um múltiplo de $\mathbf{u}$. Dito de outro modo, será que existe um vector $\mathbf{u} \in \mathbb{R}^n$ e um escalar $\lambda \in \mathbb{R}$ tais que

$$A\mathbf{u} = \lambda\mathbf{u}?$$

Observemos que o vector nulo tem essa propriedade para qualquer escalar pois $A\mathbf{0} = \lambda\mathbf{0}$, $\forall\lambda \in \mathbb{R}$. Essa é, contudo, uma situação trivial. A questão interessante é pois encontrar um vector não nulo em tais condições.

Notação 94 *Quando se refere um vector, digamos um elemento $u \in \mathbb{R}^n$, é muito frequente escrever $\mathbf{u}$, $\bar{u}$ ou $\vec{u}$ para distinguir das suas componentes ou dos escalares ou elementos de $\mathbb{R}$. Usaremos a notação simples $\mathbf{u} \in \mathbb{R}^n$.*

A.2 Valores próprios e vectores próprios

Definição 95 *Seja A uma matriz quadrada de ordem n, $A = [a_{ij}]$. Se existir um escalar $\lambda \in \mathbb{R}$ e um vector $\mathbf{u} \in \mathbb{R}^n$, $\mathbf{u} \neq \mathbf{0}$, tal que*

$$A\mathbf{u} = \lambda\mathbf{u}$$

*diz-se que λ é um **valor próprio** de A e $\mathbf{u}$ é um **vector próprio** de A associado a esse valor próprio λ.*

Exemplo 96 *Tomemos a matriz A (do exemplo anterior) e o vector $\mathbf{u} \in \mathbb{R}^3$ seguintes,*

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad e \quad \mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Observemos que

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

Logo o vector $\mathbf{u} = (2, 1, 0)$ é vector próprio da matriz A associado ao valor próprio $\lambda = 2$.

Poder-se-ão levantar diversas questões, cujas respostas serão dadas ao longo do capítulo. Por exemplo,

1. Um vector próprio corresponde a um único valor próprio?

2. E, reciprocamente, a cada valor próprio de A está associado um único vector próprio de A?

3. Como se determinam os valores próprios? E os vectores próprios?

4. Qual a relevância deste assunto?

Proposição 97 *Dado um vector próprio da matriz A, existe um e um só valor próprio a que ele está associado.*

Proposição 98 *Se* $\mathbf{u}$ *for vector próprio associado a um valor próprio* $\lambda \in \mathbb{R}$ *então qualquer múltiplo de* $\mathbf{u}$ *é também vector próprio de* A *associado ao mesmo valor próprio* λ.

A última proposição responde à segunda questão anteriormente formulada. Contudo, podemos adiantar que a um valor próprio, para além de um vector próprio associado que se determine, bem como dos seus múltiplos, poderão existir outros vectores próprios associados (e respectivos múltiplos) que não sejam múltiplos do primeiro vector determinado. Mais adiante será dada resposta completa a esta questão.

Proposição 99 0 *é valor próprio da matriz* A, *quadrada de ordem* n, *se e só se* A *não for invertível.*

Dem A não é invertível $\Longleftrightarrow r(A) < n \Longleftrightarrow$ o sistema $A\mathbf{u} = \mathbf{0}$ é indeterminado $\Longleftrightarrow \exists \mathbf{u}^* \neq \mathbf{0} : \ A\mathbf{u}^* = 0.\mathbf{u}^* = \mathbf{0} \Longleftrightarrow 0$ é valor próprio de A. $\square$

A.3 Cálculo de valores e vectores próprios

A.3.1 Cálculo dos valores próprios λ

Como vimos, se λ for um valor próprio de A e $\mathbf{u} \neq \mathbf{0}$ um vector próprio associado a esse valor próprio λ, isso significa que

$$A\mathbf{u} = \lambda\mathbf{u},$$

isto é,

$$(A - \lambda I)\,\mathbf{u} = \mathbf{0}.$$

Uma vez que $\mathbf{u} \neq \mathbf{0}$, conclui-se que a matriz $A - \lambda I$ não é invertível, o que, como sabemos, é equivalente a dizer que o seu determinante é nulo. Podemos então resumir este facto na proposição seguinte:

Proposição 100 *λ é um valor próprio de A se e só se* $\det(A-\lambda I)=0$.

Esta proposição fornece um modo de calcular os valores próprios de A que consiste na determinação dos λ que anulam o determinante da matriz $A-\lambda I$.

Exemplo 101 *Tomemos a matriz do exemplo anterior e calculemos todos os seus valores próprios*

$$\begin{aligned}
\det(A-\lambda I) &= \det\left(\begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}\right) \\
&= \det\begin{bmatrix} 1-\lambda & 2 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{bmatrix}.
\end{aligned}$$

Ora, utilizando a proposição anterior,

$$\begin{aligned}
\det(A-\lambda I) &= 0 \\
&\Longleftrightarrow \\
(1-\lambda)(2-\lambda)(3-\lambda) &= 0,
\end{aligned}$$

donde

$$\lambda = 1 \vee \lambda = 2 \vee \lambda = 3.$$

Já tínhamos assinalado que $\lambda = 2$ é valor próprio de A. Pelos cálculos que acabámos de fazer, vemos que A tem ainda mais dois valores próprios: $\lambda = 1$ e $\lambda = 3$.

Nota 102 *Observemos que a determinação dos valores próprios de A consistiu no cálculo das raízes do polinómio de grau* 3,

$$p(\lambda) = (1-\lambda)(2-\lambda)(3-\lambda),$$

ou, dito de outro modo, na resolução da equação

$$(1-\lambda)(2-\lambda)(3-\lambda) = 0.$$

Esta observação antecipa e ilustra o que adiante diremos sobre "polinómio característico" e "equação característica" de uma matriz.

A.3.2 Determinação dos vectores próprios associados a um valor próprio λ

Como já dissemos, dado um valor próprio λ de A, os vectores próprios associados a λ são os vectores $\mathbf{u} \neq \mathbf{0}$ tais que

$$A\mathbf{u} = \lambda\mathbf{u},$$

ou seja,

$$(A - \lambda I)\,\mathbf{u} = \mathbf{0},$$

o que mostra que são as soluções do sistema linear homogéneo

$$\begin{bmatrix} a_{11}-\lambda & a_{12} & \cdots & \cdots & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & & & & a_{2n} \\ & \vdots & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & a_{nn}-\lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ u_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ 0 \end{bmatrix}.$$

Se o sistema for possível e determinado, isso significa que caímos no caso trivial, isto é, apenas o vector nulo satisfaz a equação $(A - \lambda I)\,\mathbf{u} = \mathbf{0}$. Nesse caso não existiriam valores nem vectores próprios (relembra-se que, por definição, o vector próprio não pode ser nulo).

Portanto, a existência de valores e vectores próprios é equivalente ao facto de o sistema anterior ser possível e indeterminado. De acordo com o grau de indeterminação, varia a dimensão do espaço das soluções do sistema, isto é, do espaço dos vectores próprios. Assim,

- se o grau de indeterminação for 1, o espaço dos vectores próprios tem dimensão 1, isto é, dado um vector próprio todos os outros serão os seus múltiplos;

- se o grau de indeterminação for 2, o espaço dos vectores próprios tem dimensão 2, isto é, existem dois vectores próprios linearmente independentes e todos os outros são combinações lineares destes;

- se o grau de indeterminação for n, o espaço dos vectores próprios tem dimensão n, isto é, existem n vectores próprios linearmente independentes e todos os outros são combinações lineares destes.

Definição 103 *Dado um valor próprio λ de A, à união de $\{\mathbf{0}\}$ com o conjunto dos vectores próprios de A associados a λ chama-se **subespaço próprio** de λ e representa-se por E_λ, isto é, $E_\lambda = \{\mathbf{u} \in \mathbb{R}^n : A\mathbf{u} = \lambda\mathbf{u}\}$.*

Exemplo 104 *Vamos determinar os vectores próprios associados ao valor próprio 2 da matriz do exemplo anterior. Temos de resolver o sistema homogéneo $(A - 2I)\,\mathbf{u} = \mathbf{0}$. Temos então, sob a forma matricial,*

$$\left[\begin{array}{ccc|c} 1-2 & 2 & 0 & 0 \\ 0 & 2-2 & 0 & 0 \\ 0 & 0 & 3-2 & 0 \end{array}\right] = \left[\begin{array}{ccc|c} -1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \rightarrow$$

$$\rightarrow \left[\begin{array}{ccc|c} -1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

donde

$$\left\{\begin{array}{c} -u_1 + 2u_2 = 0 \\ u_3 = 0 \end{array}\right. \Longleftrightarrow \left\{\begin{array}{c} u_1 = 2u_2 \\ u_3 = 0 \end{array}\right.$$

ou seja, os vectores próprios são os vectores da forma

$$\mathbf{u} = (2a, a, 0) = a(2, 1, 0), \text{ com } a \in \mathbb{R},$$

isto é, são os múltiplos de $(2, 1, 0)$. Portanto,

$$E_2 =< (2, 1, 0) >,$$

isto é, o espaço gerado pelo vector $(2, 1, 0)$.

De forma análoga veríamos que $E_1 =< (1, 0, 0) >$ e que $E_3 =< (0, 0, 1) >$

Exemplo 105 *Determinemos os valores próprios e respectivos vectores próprios da matriz*

$$B = \begin{bmatrix} 4 & -10 & 10 \\ 0 & 6 & -2 \\ 0 & -2 & 6 \end{bmatrix}.$$

- *Determinação dos valores próprios*

$$\begin{aligned} \det(B - \lambda I) = 0 &\iff \det \begin{bmatrix} 4-\lambda & -10 & 10 \\ 0 & 6-\lambda & -2 \\ 0 & -2 & 6-\lambda \end{bmatrix} = 0 \\ &\iff (4-\lambda)(6-\lambda)^2 - 4(4-\lambda) = 0 \\ &\iff (4-\lambda)\left[(6-\lambda)^2 - 4\right] = 0 \\ &\iff \lambda = 4 \vee \lambda^2 - 12\lambda + 32 = 0 \\ &\iff \lambda = 4 \vee \lambda = 8. \end{aligned}$$

Fazemos notar que $\lambda = 8$ é solução simples e $\lambda = 4$ é solução múltipla, com multiplicidade 2.

- *Determinação dos vectores próprios associados a $\lambda = 4$*

$$\left[B - 4I \vdots \mathbf{0}\right] = \left[\begin{array}{ccc:c} 0 & -10 & 10 & 0 \\ 0 & 2 & -2 & 0 \\ 0 & -2 & 2 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc:c} 0 & -10 & 10 & 0 \\ 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{ccc:c} 0 & -10 & 10 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc:c} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Logo

$$\begin{cases} u_2 - u_3 = 0 \\ u_1 \ \textit{qualquer} \end{cases} \Longleftrightarrow \begin{cases} u_2 = u_3 \\ u_1 \ \textit{qualquer} \end{cases}.$$

Portanto, os vectores próprios associados a $\lambda = 4$ são os vectores não nulos da forma

$$\mathbf{u} = (a, b, b) = a(1, 0, 0) + b(0, 1, 1), \quad \textit{com} \ \ a, b \in \mathbb{R},$$

ou seja, são as combinações lineares não nulas dos vectores $(1, 0, 0)$ e $(0, 1, 1)$. Portanto,

$$E_4 =< (1, 0, 0), (0, 1, 1) >,$$

isto é, o subespaço próprio de $\lambda = 4$ é o espaço gerado pelos vectores $(1, 0, 0)$ e $(0, 1, 1)$. Neste caso tem dimensão 2, o que, como referimos anteriormente, se deve ao facto de o sistema de equações lineares ter grau de indeterminação 2.

De forma análoga veríamos que $E_8 =< (5, -1, 1) >$.

A.3.3 Multiplicidades algébrica e geométrica

Como vimos, dada uma matriz A de ordem n :

- os seus valores próprios são os zeros do polinómio de grau n, $p(\lambda) = \det(A - \lambda I)$. A este polinómio chama-se **polinómio característico**

da matriz A. Sendo um polinómio de grau n tem, no máximo, n raízes reais (de facto, tem exactamente n raízes reais ou complexas mas a justificação deste facto incide no Teorema Fundamental da Álgebra, que sai fora do âmbito do curso);

- à equação $det(A - \lambda I) = 0$ chama-se **equação característica da matriz A** e os valores próprios de A são precisamente as suas soluções;

- os seus valores próprios podem ser distintos ou não, consoante sejam raízes simples ou múltiplas do polinómio característico.

Definição 106 *Dado um valor próprio λ de A, chama-se* ***multiplicidade algébrica de*** λ *à sua multiplicidade enquanto raiz do polinómio característico de A, isto é, à sua multiplicidade enquanto solução da equação característica $p(\lambda) = 0$.*
Por outro lado, chama-se ***multiplicidade geométrica de*** λ *à dimensão do espaço próprio associado E_λ.*

Nota 107 *No exemplo anterior, o valor próprio $\lambda = 4$ tem multiplicidade algébrica 2 e multiplicidade geométrica 2 e o valor próprio $\lambda = 8$ tem multiplicidade algébrica 1 e multiplicidade geométrica 1 (o facto de os valores da multiplicidade algébrica e da geométrica serem o mesmo é pura coincidência).*

A.3.4 Propriedades

Proposição 108 *Seja A uma matriz quadrada de ordem n.*

1. Se A for triangular, os valores próprios de A são os elementos da diagonal principal.

2. Se $\lambda_1, \lambda_2, \lambda_3, \cdots, \lambda_n$ são n valores próprios reais de A (distintos ou não) então $\det(A) = \lambda_1 \lambda_2 \lambda_3 \cdots \lambda_n$.

3. Se A for invertível nenhum dos seus valores próprios pode ser nulo.

Teorema 109 *Seja A uma matriz quadrada de ordem n.*

1. Quaisquer dois vectores próprios associados a dois valores próprios distintos são linearmente independentes.

Mais geralmente:

2. Um sistema de k vectores próprios associados a valores próprios distintos é um sistema de vectores linearmente independente.

Dem 1. Sejam λ_1, λ_2 dois valores próprios distintos de A e $\mathbf{u}_1$, $\mathbf{u}_2$ dois vectores próprios associados. Consideremos uma combinação linear nula desses vectores

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 = \mathbf{0}, \qquad \alpha_1, \alpha_2 \in \mathbb{R}.$$

Multiplicando ambos os membros pela matriz A e usando a definição de valor e vector próprio, obtemos

$$\begin{array}{rcl} \alpha_1 A\mathbf{u}_1 + \alpha_2 A\mathbf{u}_2 & = & \mathbf{0} \\ & \Longleftrightarrow & \\ \alpha_1 \lambda_1 \mathbf{u}_1 + \alpha_2 \lambda_2 \mathbf{u}_2 & = & \mathbf{0} \\ & \Longleftrightarrow & \\ \alpha_1 \lambda_1 \mathbf{u}_1 - \lambda_2 \alpha_1 \mathbf{u}_1 & = & \mathbf{0} \\ & \Longleftrightarrow & \\ \alpha_1 (\lambda_1 - \lambda_2) \mathbf{u}_1 & = & \mathbf{0}. \end{array}$$

Como λ_1, λ_2 são distintos sai que $\alpha_1 = 0$ e, portanto, também $\alpha_2 = 0$, o que prova que $\mathbf{u}_1$, $\mathbf{u}_2$ são linearmente independentes.

2. Tomemos agora três valores próprios de A distintos, λ_1, λ_2 e λ_3 e três vectores próprios associados, $\mathbf{u}_1$, $\mathbf{u}_{2,}$ e $\mathbf{u}_3$, respectivamente. Pelo ponto 1, sabemos que $\mathbf{u}_1$ e $\mathbf{u}_2$ são linearmente independentes. Vejamos agora que $\{\mathbf{u}_1, \mathbf{u}_{2,}\mathbf{u}_3\}$ também constitui um sistema de vectores linearmente independente. Consideremos uma combinação linear nula desses vectores

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 = \mathbf{0} \qquad \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}.$$

Tal como em cima, multiplicando ambos os membros pela matriz A e usando a definição de valor e vector próprio, obtemos

$$\begin{array}{rcl} \alpha_1 A\mathbf{u}_1 + \alpha_2 A\mathbf{u}_2 + \alpha_3 A\mathbf{u}_3 & = & \mathbf{0} \\ & \Longleftrightarrow & \\ \alpha_1 \lambda_1 \mathbf{u}_1 + \alpha_2 \lambda_2 \mathbf{u}_2 + \alpha_3 \lambda_3 \mathbf{u}_3 & = & \mathbf{0} \\ & \Longleftrightarrow & \\ \alpha_1 \lambda_1 \mathbf{u}_1 + \alpha_2 \lambda_2 \mathbf{u}_2 + \lambda_3 (-\alpha_1 \mathbf{u}_1 - \alpha_2 \mathbf{u}_2) & = & \mathbf{0} \\ & \Longleftrightarrow & \\ \alpha_1 (\lambda_1 - \lambda_3) \mathbf{u}_1 + \alpha_2 (\lambda_2 - \lambda_3) \mathbf{u}_2 & = & \mathbf{0}. \end{array}$$

Como $\mathbf{u}_1$ e $\mathbf{u}_2$ são linearmente independentes e $\lambda_1, \lambda_2, \lambda_3$ são distintos sai que $\alpha_1 = 0 = \alpha_2$ e portanto, também $\alpha_3 = 0$, o que prova que $\mathbf{u}_1$, $\mathbf{u}_2$ e $\mathbf{u}_3$ são linearmente independentes.

E assim sucessivamente. Repetindo as vezes necessárias este processo, isto é, juntando sempre mais um vector próprio correspondente a um valor próprio distinto dos anteriores, chegamos à conclusão que os k vectores considerados são linearmente independentes. □

A.4 Diagonalização de uma matriz

A.4.1 Matrizes semelhantes e matriz diagonalizável

Definição 110 *Duas matrizes A e B, quadradas de ordem n, dizem-se **seme-lhantes** se existir uma matriz invertível P, de ordem n, tal que*

$$B = P^{-1}AP.$$

Proposição 111 *Matrizes semelhantes têm os mesmos valores próprios.*

Dem Sendo válidas as igualdades seguintes

$$\begin{aligned} \det(B - \lambda I) &= \det(P^{-1}AP - \lambda I) = \det(P^{-1}AP - \lambda P^{-1}P) \\ &= \det(P^{-1}(AP - \lambda P)) = \det(P^{-1}(A - \lambda I)P) \\ &= \det P^{-1}. \det(A - \lambda I). \det P = \det(A - \lambda I) \end{aligned}$$

fica provado que os polinómios característicos de A e B coincidem, logo têm as mesmas raízes, isto é, A e B têm os mesmos valores próprios. □

Têm particular importância as matrizes semelhantes a uma matriz diagonal. Por exemplo, sabe-se pela proposição anterior que os seus valores próprios serão os mesmos que os valores próprios da matriz diagonal e os desta são precisamente os elementos da sua diagonal principal. Mas poderemos constatar ainda mais propriedades. Sistematizemos estes factos:

Definição 112 *Uma matriz A diz-se **diagonalizável** se for semelhante a uma matriz diagonal, isto é, se existir uma matriz invertível P, de ordem n, tal que*

$$D = P^{-1}AP \qquad \text{em que } D \text{ é matriz diagonal.}$$

*À matriz P chama-se **matriz de diagonalização**.*

Proposição 113 *Se A for diagonalizável então:*

1. Os valores próprios de A são os elementos da diagonal principal da matriz diagonal D a que é semelhante a matriz A;

2. Sendo D a matriz diagonal a que é semelhante a matriz A e P a respectiva matriz de diagonalização, tem-se, qualquer que seja $k \in \mathbb{N}$,

$$A^k = PD^kP^{-1}.$$

Dem 1. Resulta imediatamente da proposição anterior e do facto de os valores próprios de uma matriz diagonal serem precisamente os elementos da diagonal principal.

2. Como $D = P^{-1}AP$, o que é equivalente a $A = PDP^{-1}$, tem-se

$$\begin{aligned} A^k &= \left(PDP^{-1}\right)^k = \underbrace{\left(PDP^{-1}\right)\left(PDP^{-1}\right)\cdots\left(PDP^{-1}\right)}_{k \text{ vezes}} \\ &= PD\left(P^{-1}P\right)D\left(P^{-1}P\right)D\cdots D\left(P^{-1}P\right)DP^{-1} = PD^kP^{-1}, \end{aligned}$$

o que prova o resultado. □

Teorema 114 *Seja A uma matriz quadrada de ordem n, $A = [a_{ij}]$. Então A é diagonalizável sse existem n vectores próprios de A constituindo um sistema de vectores linearmente independentes.*

Dem Condição necessária

Se A for diagonalizável então existe uma matriz P de ordem n invertível tal que $P^{-1}AP = D$ ou, de forma equivalente $AP = PD$.

Sejam

$$P = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \cdots & \mathbf{w}_n \end{bmatrix}$$

onde

$$\mathbf{w}_1 = \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ \vdots \\ \vdots \\ w_{n1} \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} w_{12} \\ w_{22} \\ \vdots \\ \vdots \\ \vdots \\ w_{n2} \end{bmatrix}, \cdots, \mathbf{w}_n = \begin{bmatrix} w_{1n} \\ w_{2n} \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ w_{nn} \end{bmatrix},$$

isto é,

$$P = \begin{bmatrix} w_{11} & w_{12} & \cdots & \cdots & \cdots & w_{1n} \\ w_{21} & w_{22} & & & & w_{2n} \\ & \vdots & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & \cdots & \cdots & w_{nn} \end{bmatrix}.$$

Seja

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & & & & 0 \\ & \vdots & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \lambda_n \end{bmatrix}.$$

Então

$$AP = \begin{bmatrix} A\mathbf{w}_1 & A\mathbf{w}_2 & \cdots & \cdots & A\mathbf{w}_n \end{bmatrix}$$

e

$$PD = \begin{bmatrix} \lambda_1\mathbf{w}_1 & \lambda_2\mathbf{w}_2 & \cdots & \cdots & \lambda_n\mathbf{w}_n \end{bmatrix}.$$

Logo, pela igualdade acima referida $AP = PD$,

$$A\mathbf{w}_1 = \lambda_1\mathbf{w}_1, \quad A\mathbf{w}_2 = \lambda_2\mathbf{w}_2, \cdots, A\mathbf{w}_n = \lambda_n\mathbf{w}_n.$$

Destas igualdades, podemos então observar que

(i) os vectores que constituem as colunas da matriz P, $\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n$, são os vectores próprios de A,

(ii) os elementos da diagonal principal da matriz diagonal D são os valores próprios de A.

Além disso, como P é invertível,

(iii) $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\}$ constitui um sistema de vectores linearmente independentes.

Condição suficiente

Suponhamos que $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\}$ constitui um sistema de vectores próprios de A linearmente independentes correspondendo respectivamente aos valores próprios $\lambda_1, \lambda_2, ..., \lambda_n$. Consideremos as matrizes

$$P = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \cdots & \mathbf{w}_n \end{bmatrix}$$

em que, como anteriormente, $\mathbf{w}_i = [w_{1i}\ w_{2i} \cdots w_{ni}]^T$, para cada $1 \leq i \leq n$ e

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & & & & 0 \\ & \vdots & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \lambda_n \end{bmatrix}.$$

Então

$$\begin{bmatrix} A\mathbf{w}_1 & A\mathbf{w}_2 & \cdots & \cdots & A\mathbf{w}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{w}_1 & \lambda_2\mathbf{w}_2 & \cdots & \cdots & \lambda_n\mathbf{w}_n \end{bmatrix},$$

isto é,

$$AP = PD$$

e, portanto,

$$P^{-1}AP = D,$$

o que mostra que A é diagonizável. □

Como consequência imediata do teorema anterior temos o resultado seguinte

Corolário 115 *Se os n valores próprios da matriz A forem distintos, isto é, cada valor próprio tiver multiplicidade algébrica igual a* 1, *então a matriz A é diagonalizável.*

A.4.2 Teorema espectral para matrizes simétricas

Recordemos que uma matriz quadrada é simétrica se coincidir com a sua transposta, isto é, $A = A^\top$.

Teorema 116 *Seja A uma matriz quadrada de ordem n, simétrica. Então,*
1. Todas as n raízes do polinómio característico de A são reais, isto é, os n valores próprios de A são reais;
2. Os vectores próprios associados a valores próprios distintos são ortogonais;
3. A matriz A é diagonalizável.

Dem 1. A demonstração sai fora do âmbito do curso.

2. Sejam $\mathbf{u}$ e $\mathbf{v}$ dois vectores próprios de A correspondentes a valores próprios distintos λ e μ, respectivamente. Então

$$A\mathbf{u} = \lambda\mathbf{u} \qquad \text{e} \qquad A\mathbf{v} = \mu\mathbf{v}.$$

Considerando os vectores transpostos $\mathbf{u}^\top$ e $\mathbf{v}^\top$ tem-se

$$\mathbf{v}^\top A\mathbf{u} = \lambda\mathbf{v}^\top\mathbf{u} \qquad \text{e} \qquad \mathbf{u}^\top A\mathbf{v} = \mu\mathbf{u}^\top\mathbf{v}.$$

Transpondo ambos os membros da primeira igualdade, obtemos

$$\left(\mathbf{v}^\top A\mathbf{u}\right)^\top = \left(\lambda\mathbf{v}^\top\mathbf{u}\right)^\top$$

e, como A é simétrica,

$$\mathbf{u}^\top A\mathbf{v} = \lambda\mathbf{u}^\top\mathbf{v}.$$

Logo

$$\lambda\mathbf{u}^\top\mathbf{v} = \mu\mathbf{u}^\top\mathbf{v}$$

isto é

$$(\lambda - \mu)\,\mathbf{u}^\top\mathbf{v} = 0.$$

Uma vez que $\lambda \neq \mu$, então $\mathbf{u}^\top\mathbf{v} = 0$, isto é, $\mathbf{u}$ e $\mathbf{v}$ são ortogonais. □

O teorema seguinte antecipa um resultado extremamente importante em Álgebra Linear. Recorda-se que uma matriz quadrada P é ortogonal se for invertível e $P^{-1} = P^\top$.

Teorema 117 *Seja A uma matriz quadrada de ordem n e simétrica. Então* *existe uma matriz ortogonal P tal que*

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & & & & 0 \\ & \vdots & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \lambda_n \end{bmatrix}.$$

Dem Apresentamos a demonstração no caso particular em que os valores próprios de A são todos distintos.

Neste caso, pelo teorema anterior, a matriz A é diagonalizável e possui n vectores próprios $\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n$, correspondentes aos valores próprios $\lambda_1, \lambda_2, \cdots, \lambda_n$, ortogonais dois a dois. Normalizando esses vectores, obtêm-se n vectores próprios ortogonais mas de norma 1, isto é, ortonormais

$$\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}, \mathbf{v}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}, ..., \mathbf{v}_n = \frac{\mathbf{u}_n}{\|\mathbf{u}_n\|}.$$

Considerando a matriz de diagonalização $P = [\mathbf{v}_1\mathbf{v}_2...\mathbf{v}_n]$ basta ver que $P^{-1} = P^\top$ e fica provado o resultado. $\square$

Nota 118 *A demonstração anterior induz que a matriz P pode ser tomada como ortonormal, isto é, ortogonal em que os vectores coluna têm norma 1.*

Tem-se o **Teorema Espectral** para matrizes simétricas:

Teorema 119 *Sejam A uma matriz quadrada de ordem n simétrica e $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ vectores próprios ortonormais correspondentes aos valores próprios $\lambda_1, \lambda_2, \cdots, \lambda_n$. Então a matriz A admite a seguinte decomposição espectral*

$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^\top + \lambda_1 \mathbf{v}_2 \mathbf{v}_2^\top + ... + \lambda_1 \mathbf{v}_n \mathbf{v}_n^\top.$$

Dem Considere-se a matriz diagonal D e a matriz de diagonalização ortonormal P, isto é, a matriz

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_2 & & & & 0 \\ & \vdots & & & & \vdots \\ \vdots & \vdots & \ddots & & & \vdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & \vdots \\ 0 & 0 & \cdots & \cdots & \cdots & \lambda_n \end{bmatrix}$$

e a matriz P, cujas colunas são os vectores próprios ortonormais associados a $\lambda_1, \lambda_2, \cdots, \lambda_n$

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \cdots & \mathbf{v}_n \end{bmatrix}.$$

Aplicando A ao vector coluna $\mathbf{x} \in \mathbb{R}^n$, obtém-se

$$
\begin{aligned}
A\mathbf{x} &= \left(PDP^{-1}\right)\mathbf{x} = \left(PDP^{\top}\right)\mathbf{x} = (PD)\left(P^{\top}\mathbf{x}\right) \\
&= \left(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} D\right)\left(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}^{\top} \mathbf{x}\right) \\
&= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}\left(\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}^{\top} \mathbf{x}\right) \\
&= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}\left(\begin{bmatrix} \mathbf{v}_1^{\top} \\ \mathbf{v}_2^{\top} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{v}_n^{\top} \end{bmatrix}\mathbf{x}\right) \\
&= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \cdots & \lambda_n\mathbf{v}_n \end{bmatrix}\begin{bmatrix} \mathbf{v}_1^{\top}\mathbf{x} \\ \mathbf{v}_2^{\top}\mathbf{x} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{v}_n^{\top}\mathbf{x} \end{bmatrix} \\
&= (\lambda_1\mathbf{v}_1)(\mathbf{v}_1^{\top}\mathbf{x}) + (\lambda_2\mathbf{v}_2)(\mathbf{v}_2^{\top}\mathbf{x}) + ... + (\lambda_n\mathbf{v}_n)(\mathbf{v}_n^{\top}\mathbf{x}) = \\
&= \left(\lambda_1\mathbf{v}_1\mathbf{v}_1^{\top} + \lambda_2\mathbf{v}_2\mathbf{v}_2^{\top} + ... + \lambda_n\mathbf{v}_n\mathbf{v}_n^{\top}\right)\mathbf{x}.
\end{aligned}
$$

Nota 120 *O conjunto dos valores próprios de uma matriz é designado por* ***espectro da matriz****. Daí a designação de "teorema espectral ", que mostra que uma matriz simétrica pode ser definida através do seu espectro.*

A.5 Formas quadráticas em $\mathbb{R}^n$

Definição 121 *Uma **forma quadrática** em n variáveis é uma aplicação*

$$Q : \mathbb{R}^n \longrightarrow \mathbb{R}$$

que se exprime numa base dada através de um polinómio homogéneo de grau 2. Mais precisamente, sendo $\mathbf{x} = (x_1, x_2, ..., x_n)$,

$$\begin{aligned} Q(\mathbf{x}) &= a_{11}x_1^2 + a_{12}x_1x_2 + ... + a_{1n}x_1x_n + \\ &\quad +a_{21}x_2x_1 + a_{22}x_2^2 + ... + a_{2n}x_2x_n + \\ &\quad ... \\ &\quad +a_{n1}x_nx_1 + a_{n2}x_nx_2 + ... + a_{nn}x_n^2, \end{aligned}$$

isto é,

$$Q(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j.$$

Escrevendo $b_{ij} = a_{ij} + a_{ji}$ e tendo em conta que $x_ix_j = x_jx_i$, o que conduz a $(a_{ij}x_ix_j + a_{ji}x_jx_i) = b_{ij}x_ix_j$, obtém-se a expressão equivalente

$$Q(\mathbf{x}) = \sum_{i \leq j} b_{ij}x_ix_j.$$

Exemplo 122 *Vejamos os exemplos seguintes*

1. $Q(x, y) = 5x^2 + 8xy + 7y^2$ é uma forma quadrática em $\mathbb{R}^2$ pois é um polinómio de duas variáveis homogéneo de grau 2.

2. $Q(x, y, z) = 4x^2 + 8xy + 12xz + 8y^2 + 5yz + 20z^2$ é uma forma quadrática em $\mathbb{R}^3$ pois é um polinómio de três variáveis homogéneo de grau 2.

3. $Q(x, y, z, w) = 4x^2 + 4xy + 12xz - 10xw + 8y^2 - 14yw + 20z^2 + 2zw + 2w^2$ é uma forma quadrática em $\mathbb{R}^4$ pois é um polinómio de quatro variáveis homogéneo de grau 2.

Toda a forma quadrática pode ser escrita em forma matricial do seguinte modo

$$\begin{aligned} Q(\mathbf{x}) &= X^\top A X = \\ &= \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \end{aligned}$$

o que mostra que a cada forma quadrática Q em $\mathbb{R}^n$ está sempre associada uma matriz quadrada de ordem n. Reciprocamente, dada uma matriz quadrada A de ordem n, podemos sempre associar-lhe uma forma quadrática Q_A em $\mathbb{R}^n$, isto é, um polinómio homogéneo de grau 2, através do produto de matrizes

$$Q_A(\mathbf{x}) = X^\top A X.$$

Por exemplo, no caso das formas quadráticas apresentadas em cima, temos as formas matriciais:

1. $Q(x,y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

2. $Q(x,y,z) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} 4 & 6 & 6 \\ 2 & 8 & 4 \\ 6 & 1 & 20 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$

3. $Q(x,y,z,w) = \begin{bmatrix} x & y & z & w \end{bmatrix} \begin{bmatrix} 4 & 2 & 6 & -5 \\ 2 & 8 & 0 & -7 \\ 6 & 0 & 20 & 1 \\ -5 & -7 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}.$

É óbvio que, se inicialmente tivessem sido dadas as matrizes quadradas, poderíamos reconstituir as formas quadráticas que lhes estão associadas.

Claramente se observa que, dada uma matriz A, a forma quadrática Q_A que lhe correponde é única. Será que o recíproco é válido? Isto é, dada uma forma quadrática Q ela admite uma forma matricial única? A resposta é negativa. Com efeito, a uma forma quadrática podem ser associadas inúmeras matrizes, o que pode ser enunciado de forma equivalente dizendo que matrizes diferentes podem gerar a mesma forma quadrática.

Tomemos as matrizes

$$A = \begin{bmatrix} 5 & 6 \\ 2 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 9 \\ -1 & 7 \end{bmatrix} \text{ e } C = \begin{bmatrix} 5 & 4 \\ 4 & 7 \end{bmatrix}$$

e determinemos as formas quadráticas que lhes estão associadas, respectivamente

$$\begin{aligned}
Q_A(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5x^2 + 8xy + 7y^2 \\
Q_B(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 9 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5x^2 + 8xy + 7y^2 \\
Q_C(x,y) &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 4 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 5x^2 + 8xy + 7y^2
\end{aligned}$$

Constatamos que

$$Q_A(x,y) = Q_B(x,y) = Q_C(x,y),$$

isto é, as três matrizes dadas estão associadas à mesma forma quadrática (já referida no exemplo anterior). Observamos, contudo, que a matriz C é simétrica e este tipo de acontecimento tem grande relevância como mostra o teorema seguinte:

Teorema 123 *A toda a forma quadrática* $Q : \mathbb{R}^n \longrightarrow \mathbb{R}$

$$Q(\mathbf{x}) = \sum_{i \leq j} b_{ij} x_i x_j$$

está associada uma e uma só matriz simétrica $A = (a_{ij})$. *Essa matriz é definida por* $a_{ii} = b_{ii}$ *e* $a_{ij} = a_{ji} = \frac{b_{ij}}{2}$, *para* $i < j$. *A relação entre a forma quadrática e a matriz é dada pela equação*

$$\sum_{i \leq j} b_{ij} x_i x_j = X^\top A X.$$

Teorema 124 *Toda a forma quadrática $Q : \mathbb{R}^n \longrightarrow \mathbb{R}$ é diagonalizável.*

Definição 125 *Uma forma quadrática $Q : \mathbb{R}^n \longrightarrow \mathbb{R}$ diz-se*

- ***definida positiva (DP)*** se $Q(\mathbf{x}) > 0, \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$,
- ***semidefinida positiva (SDP)*** se $Q(\mathbf{x}) \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$,
- ***definida negativa (DN)*** se $Q(\mathbf{x}) < 0, \forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$,
- ***semidefinida negativa (SDN)*** se $Q(\mathbf{x}) \leq 0, \forall \mathbf{x} \in \mathbb{R}^n$,
- ***indefinida (IND)*** se $\exists \mathbf{x}, \mathbf{y} \in \mathbb{R}^n : Q(\mathbf{x}) > 0 \wedge Q(\mathbf{y}) < 0$.

De forma correspondente, uma matriz simétrica A diz-se

- ***definida positiva (DP)*** se $X^\top AX > 0, \ \forall X \neq \mathbf{0}$;
- ***semidefinida positiva (SDP)*** se $\begin{cases} X^\top AX \geq 0, \ \forall X, \\ \exists Y \neq \mathbf{0}, \ Y^\top AY = 0; \end{cases}$
- ***definida negativa (DN)*** se $X^\top AX < 0, \ \forall X \neq \mathbf{0}$;
- ***semidefinida negativa (SDN)*** se $\begin{cases} X^\top AX \leq 0, \ \forall X, \\ \exists Y \neq \mathbf{0}, \ Y^\top AY = 0; \end{cases}$
- ***indefinida (IND)*** se $\begin{cases} \exists X, \ \text{com} \ X^\top AX < 0, \\ \exists Y, \ \text{com} \ Y^\top AY > 0. \end{cases}$

Proposição 126 *Seja A uma matriz simétrica. Então, usando as abreviaturas introduzidas na definição anterior relativamente à classificação de ma-*

trizes simétricas, tem-se

$$\begin{array}{lcl} A \text{ é } DP & \Leftrightarrow & \text{todos os valores próprios } \lambda_i > 0; \\ A \text{ é } SDP & \Leftrightarrow & \text{um valor próprio } \lambda_i = 0 \text{ e os outros } \geq 0; \\ A \text{ é } DN & \Leftrightarrow & \text{todos os valores próprios } \lambda_i < 0; \\ A \text{ é } SDN & \Leftrightarrow & \text{um valor próprio } \lambda_i = 0 \text{ e os outros } \leq 0; \\ A \text{ é } IND & \Leftrightarrow & \text{um valor próprio } < 0 \text{ e outro } > 0. \end{array}$$

Definição 127 *Uma* ***submatriz principal*** *B de A é uma submatriz obtida suprimindo certas linhas de A bem como as colunas com o mesmo índice.*

Uma ***submatriz principal primária*** *B de A é uma submatriz principal obtida suprimindo as r últimas linhas e as r últimas colunas de A, para um inteiro r compreendido entre 0 e $n-1$.*

Proposição 128 *Se uma submatriz principal B de uma matriz simétrica A possuir pelo menos k valores próprios ≥ 0 (respectivamente > 0, < 0, ≤ 0), então passa-se o mesmo para A.*

Definição 129 *Chamamos* ***menores principais primários*** *aos determinantes das matrizes principais primárias.*

Teorema 130 *Seja A uma matriz simétrica de ordem n. Então*

1. *A é DP se e só se todos os menores principais primários são > 0;*

2. *A é DN se e só se os menores principais primários de ordem ímpar são < 0 e os de ordem par são > 0;*
3. *Se $|A| \neq 0$ e os menores principais não satisfizerem as condições de sinal referidas em 1. ou em 2., então A é indefinida.*

Nota 131 *O teorema anterior não permite classificar uma matriz A quando $|A| = 0$, podendo acontecer que A seja SDP, SDN ou IND. Assim, se $|A| = 0$, a classificação de A deverá ser feita por definição ou através do cálculo dos valores próprios.*

Bibliografia

.

[1] Agudo, F. Dias. *Análise Real* , vol I, Escolar Editora , 1989.

[2] Apostol, T. *Cálculo*, Vol. II, Reverté, 1994.

[3] Chiang, A. Fundamental Methods of Mathematical Economics, McGraw-Hill, 1984.

[4] Ferreira, J. Campos, *Introdução à Análise em* $\mathbb{R}^n$, DMIST, 2003.

[5] Lancaster, K. *Mathematical Economics,* Dover, 1968.

[6] Ribeiro, C. Silva, *Notas de Apoio a Matemática II*, AE-ISEG.

[7] Sanchez, L. *Análise em* $\mathbb{R}^n$, AE - FCUL.