Introdução à Análise em $\mathbb{R}^n$

J. Campos Ferreira

3 de Junho de 2004

Índice

Introdução

Uma grande parte deste trabalho é o resultado de uma revisão do texto intitulado *Introdução à Análise em* $\mathbb{R}^n$, que redigi há mais de vinte anos para os alunos que então frequentavam no Instituto Superior Técnico a disciplina de Análise Matemática II. Decidi-me a efectuar essa revisão — e até a acrescentar diversos complementos cuja redacção está em curso — porque alguns colegas e amigos me asseguraram que o trabalho poderia ter ainda hoje alguma utilidade, como texto de apoio a uma parte das suas aulas da referida disciplina.

Gostaria de deixar aqui expresso o meu reconhecimento aos Professores Francisco Teixeira, João Palhoto de Matos e Pedro Girão e ao Engenheiro Paulo Abreu pelas suas valiosas contribuições para a concretização deste projecto.

Lisboa, Novembro de 2002

Jaime Campos Ferreira

Capítulo 1

Generalidades e primeiros exemplos

1.1 Introdução

Foi estudada anteriormente a noção geral de função. De forma intuitiva, pode pensar-se que uma função f associa a cada elemento x de um dado conjunto A, chamado *domínio* de f, um e um só elemento $f(x)$ de um conjunto B; o subconjunto de B formado por todos os valores $f(x)$ é, como sabemos, o *contradomínio* de f.

No caso geral, A e B podem ser conjuntos com elementos de natureza qualquer. No entanto, quase todo o nosso trabalho anterior incidiu sobre um caso particular, aliás muito importante: o de tanto A como B serem subconjuntos do conjunto $\mathbb{R}$, dos números reais (dizia-se então, como vimos, que as funções consideradas eram funções reais de (uma) variável real).

Vamos iniciar agora uma generalização desse estudo, de enorme interesse em toda a espécie de aplicações: estudaremos funções reais de m variáveis reais, (com m inteiro positivo), isto é, funções cujo contradomínio é ainda um subconjunto de $\mathbb{R}$ mas cujo domínio é uma parte do conjunto $\mathbb{R}^m = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}$ (produto cartesiano de m factores todos iguais a $\mathbb{R}$). As funções deste tipo são também designadas por funções reais de variável vectorial (expressão relacionada com a designação de vectores, dada correntemente aos elementos de $\mathbb{R}^m$). Mais geralmente ainda, vários aspectos do nosso estudo incidirão sobre funções vectoriais de variável vectorial (funções com domínio $A \subset \mathbb{R}^m$ e contradomínio $B \subset \mathbb{R}^n$, com m e n inteiros positivos).

Neste quadro geral, estudaremos várias noções fundamentais — como as de limite e continuidade — e abordaremos o estudo do cálculo diferencial, bem como algumas das suas aplicações mais importantes.

Recordemos que, antes de iniciarmos o estudo das funções reais de uma variável real, tivemos necessidade de organizar convenientemente os nossos conhecimentos sobre o próprio conjunto $\mathbb{R}$; da mesma forma, teremos agora de começar por estruturar de forma adequada o conjunto $\mathbb{R}^m$, para que possamos assentar numa base sólida o estudo que vamos empreender. Esse trabalho será feito no Capítulo 2, dedicando-se os restantes parágrafos deste capítulo à consideração de exemplos e à exposição de algumas ideias muito simples, que convém abordar nesta fase introdutória do nosso trabalho.

1.2 Alguns exemplos de funções de duas variáveis reais

Uma função real de duas variáveis reais, f, definida numa parte A de $\mathbb{R}^2$, faz corresponder a cada par ordenado de números reais, (x, y), pertencente ao conjunto A, um único número real, $f(x, y)$. Vejamos alguns exemplos concretos.

1. Seja f a função definida pela fórmula:

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

   no conjunto de todos os pontos $(x, y) \in \mathbb{R}^2$.

   Trata-se de uma função de duas variáveis, aqui designadas por x e y; convém lembrar, no entanto, que as letras escolhidas para variáveis são inteiramente secundárias: a mesma função poderia ser definida, por exemplo, pela fórmula:

$$f(u, v) = u^2 + v^2, \qquad (u, v) \in \mathbb{R}^2.$$

   Convém também observar desde já que, se «fixarmos» uma das variáveis num determinado valor, obteremos uma função de uma só variável (a variável «não fixada»); assim, por exemplo, fixando y no valor 2 obter-se-ia a função parcial (que designamos por φ):

$$\varphi(x) = f(x, 2) = x^2 + 4, \qquad x \in \mathbb{R}.$$

   Analogamente, atribuindo a x o valor -1 obter-se-ia uma nova função parcial:

$$\psi(y) = f(-1, y) = y^2 + 1, \qquad x \in \mathbb{R}.$$

   Como é óbvio ter-se-ia, necessariamente:

$$\varphi(-1) = f(-1, 2) = \psi(2).$$

2. Considere-se agora a função g definida pela fórmula:

$$g(x, y) = \sqrt{x^2 + y^2 - 1} - \sqrt{9 - x^2 - y^2},$$

   no conjunto de todos os pontos (x, y) para os quais tem sentido (no conjunto $\mathbb{R}$, onde g toma valores) a expressão que figura no 2º membro.

   Trata-se de uma função de duas variáveis reais cujo domínio pode representar-se no plano xy pela coroa circular determinada pelas circunferências de centro na origem e raios 1 e 3 (incluindo os pontos que pertencem às próprias circunferências).

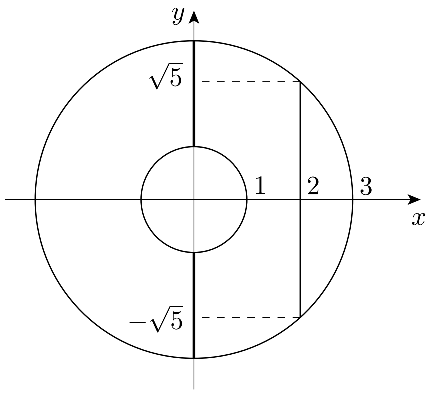

Figura 1.1

Neste caso, se fixarmos por exemplo a variável x no valor 2, obteremos a função parcial:

$$\theta(y) = g(2, y) = \sqrt{y^2 + 3} - \sqrt{5 - y^2}$$

cujo domínio é o intervalo $[-\sqrt{5},\ \sqrt{5}]$. Se, em vez de $x = 2$, pusermos $x = 0$ ou $x = 3$, obteremos respectivamente as funções:

$$g(0, y) = \sqrt{y^2 - 1} - \sqrt{9 - y^2}$$

e

$$g(3, y) = \sqrt{y^2 + 8} - \sqrt{-y^2}$$

O domínio da primeira é o conjunto $[-3,\ -1] \cup [1,\ 3]$ e o da segunda tem apenas um ponto (o ponto 0).

3. Seja h a função definida pela fórmula

$$h(x, y) = \arcsin \frac{x}{y},$$

no conjunto de todos os pontos $(x, y) \in \mathbb{R}^2$ tais que $\arcsin x/y \in \mathbb{R}$.

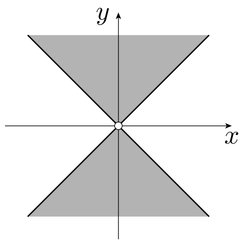

Figura 1.2

É fácil reconhecer que o domínio de h é o conjunto representado geometricamente pelos dois ângulos verticalmente opostos que têm por lados as bissectrizes dos quadrantes pares e dos quadrantes ímpares, e que não contém o eixo das abcissas (os lados dos ângulos referidos pertencem ainda ao domínio, mas não o seu vértice comum).

4. Convém observar que, tal como no caso das funções de uma só variável, para definir uma função de duas variáveis reais não é necessário dar uma expressão analítica. Assim, por exemplo, definir-se-ia também uma função de duas variáveis reais por meio de qualquer dos enunciados seguintes:

   (a) Seja f a função definida em $\mathbb{R}^2$ e tal que

   $$\begin{aligned} f(x,y) &= 0 \quad \text{se } x \text{ e } y \text{ são números inteiros} \\ f(x,y) &= 1 \quad \text{se } x \text{ ou } y \text{ não são inteiros.} \end{aligned}$$

   Ter-se-ia, por exemplo: $f(1,-5) = 0$, $f(2, 1/3) = f(\pi, \sqrt{2}) = 1$, etc.

   (b) Seja g a função cujo domínio é o círculo definido pela desigualdade

   $$x^2 + y^2 \leq 4$$

   e tal que

   $$g(x,y) = \sqrt{1 - x^2 - y^2} \qquad\qquad \text{se } x^2 + y^2 < 1$$

   e

   $$g(x,y) = 0 \qquad\qquad \text{se } 1 \leq x^2 + y^2 \leq 4.$$

   Adiante faremos mais algumas referências às funções mencionadas neste exemplo, a propósito da noção de gráfico de uma função de duas variáveis reais, considerada no parágrafo seguinte.

1.3 Gráficos e linhas de nível

Consideremos no «espaço ordinário» um referencial cartesiano ortonormado. Por um processo bem conhecido, cada ponto P do espaço determina então um terno ordenado de números reais (x, y, z), designados respectivamente por abcissa, ordenada e cota do ponto P; reciprocamente, cada terno ordenado de números reais — isto é, cada elemento de $\mathbb{R}^3$ — determina um ponto do espaço ordinário.

Assim, fixado um referencial, fica estabelecida uma bijecção entre o conjunto $\mathbb{R}^3$ e o espaço ordinário, considerado como conjunto de pontos.

Nestas condições, sendo $z = f(x, y)$ uma função de duas variáveis reais, definida num conjunto $A \subset \mathbb{R}^2$, chama-se gráfico da função f no referencial considerado o conjunto de todos os pontos (x, y, z) cujas coordenadas verificam a condição $z = f(x, y)$ (poderia também dizer-se que o gráfico de f é o conjunto de todos os pontos da forma $\big(x, y, f(x, y)\big)$, com $(x, y) \in A$).

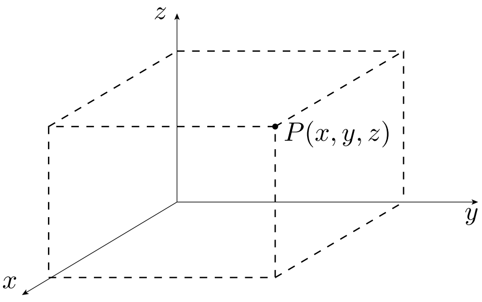

Figura 1.3

Trata-se, como é evidente, de uma generalização natural da noção de gráfico bem conhecida para as funções de uma variável real. No caso destas funções, os gráficos eram geralmente «linhas» (pelo menos se as funções consideradas tivessem «regularidade» suficiente). Para funções de duas variáveis os gráficos serão, na generalidade dos casos que nos interessará considerar, «superfícies», contidas no espaço ordinário.

Na Fig. 1.4 tenta-se dar uma ideia do gráfico da função $z = x^2+y^2$, considerada em 1.2, no exemplo 1. A superfície em causa é um parabolóide de revolução, que pode obter-se fazendo rodar em torno do eixo dos z a parábola situada no plano dos yz e cuja equação neste plano é $z = y^2$.

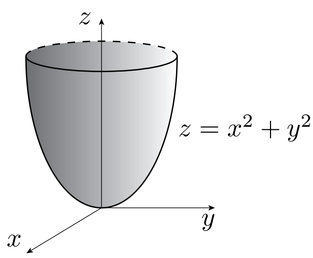

Figura 1.4

Na Fig. 1.5 esboça-se o gráfico da função g, do exemplo 4. b) de 1.2.

Trata-se de uma superfície em forma de «chapéu», cuja «aba» é a coroa circu-

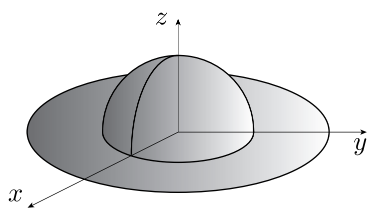

Figura 1.5

lar situada no plano xy e definida pela condição:

$$1 \leq x^2 + y^2 \leq 4,$$

isto é, a coroa limitada pelas circunferências de centro na origem e raios iguais a 1 e 2, (observe-se que nos pontos desta coroa circular, a função $z = g(x, y)$ assume sempre o valor 0, o que significa que toda esta parte do gráfico está situada no plano xy). A parte restante do gráfico é um hemisfério, intersecção da superfície esférica de equação

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

com o «semi-espaço superior», isto é, com o conjunto dos pontos de cota positiva.

Por último, consideremos a função f, definida no exemplo 4. a). O conjunto de pontos que constitui o seu gráfico não é propriamente uma superfície: o gráfico é formado por todos os pontos de um plano paralelo ao plano xy e situado uma unidade acima deste (plano cuja equação é $z = 1$), com excepção dos que têm por abcissa e por ordenada números inteiros, cada um dos quais é «substituído» pela sua projecção ortogonal sobre o plano xy.

Notaremos agora que, embora seja bastante natural, a representação gráfica das funções de duas variáveis reais que temos estado a considerar tem o inconveniente de exigir o recurso a modelos tridimensionais que, quando representados em perspectiva numa folha de papel se tornam bastante menos sugestivos e mais difíceis de interpretar (como o provam algumas das figuras insertas neste parágrafo). Por vezes, pode obter-se uma representação plana mais esclarecedora a respeito do gráfico de uma função de duas variáveis recorrendo às chamadas «linhas de nível», usadas correntemente nas cartas topográficas para indicar a altitude dos terrenos figurados.

A ideia de uma tal representação é muito simples; sugeri-la-emos através de um exemplo, o da função $z = x^2 + y^2$, cujo gráfico esboçámos na Fig. 1.4.

Se intersectarmos esse gráfico com planos paralelos ao plano xy e de cotas positivas (isto é, com planos de equação $z = c$, com c constante positiva), obteremos circunferências de raio tanto maior quanto maior for a cota do plano secante.

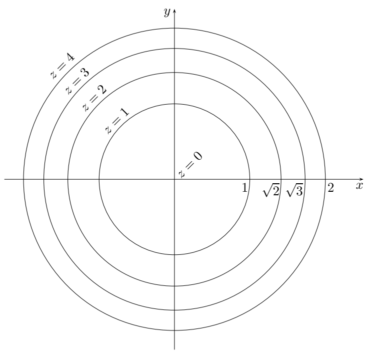

Figura 1.6

As linhas de nível, algumas das quais estão representadas na Fig. 1.6, são precisamente as projecções destas circunferências sobre o plano horizontal (dos xy). Para obter uma equação da linha de nível correspondente à secção do parabolóide com o plano $z = c$ basta substituir z por c na equação $z = x^2 + y^2$, obtendo-se

$$x^2 + y^2 = c,$$

o que mostra que o plano horizontal de cota c intersecta o parabolóide segundo uma circunferência de raio $\sqrt{c}$.

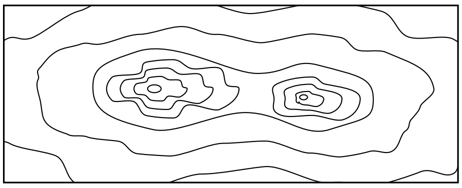

Figura 1.7

Nas cartas topográficas, é frequente as linhas de nível figuradas corresponderem a planos secantes com cotas c em progressão aritmética; em tal caso é fácil, por simples observação, avaliar o declive do terreno representado na carta: assim, nas zonas em que o declive é muito acentuado, como sucede habitualmente junto de picos montanhosos, as linhas de nível apresentam-se muito concentradas, muito próximas umas das outras; nas regiões sensivelmente planas (planícies, etc.) verifica-se uma rarefacção das linhas de nível, que estão então bastante distanciadas (Fig. 1.7).

1.4 Exemplos de funções de mais de duas variáveis reais

A noção de função de três (ou mais) variáveis reais define-se de forma óbvia: função de três variáveis reais é qualquer função cujo domínio seja um subconjunto de $\mathbb{R}^3$ (cubo cartesiano do conjunto $\mathbb{R}$). Uma tal função, f, faz corresponder a cada terno ordenado (x, y, z), pertencente ao seu domínio, um determinado número real designado por $f(x, y, z)$. Mais geralmente, sendo m um número inteiro positivo, qualquer função f cujo domínio A esteja contido em $\mathbb{R}^m$ é uma função de m variáveis reais, cujo valor no ponto $(x_1, x_2, \ldots, x_m) \in A$ se designa por $f(x_1, x_2, \ldots, x_m)$ (por vezes, quando não há risco de confusão, a sequência de m reais $(x_1, \ldots, x_m)$ é representada abreviadamente por $\boldsymbol{x}$ e o valor $f(x_1, \ldots, x_m)$ simplesmente por $f(\boldsymbol{x})$, como se tratasse de uma função de uma só variável real).

Como exemplos consideremos as funções:

$$\begin{aligned} f(x, y, z) &= 1/xyz, \\ g(x, y, z) &= \lim_{n\to\infty} (x^{2n} + y^{2n} + z^{2n}), \\ h(x_1, x_2, \ldots, x_m) &= \log\left(x_1^2 + x_2^2 + \cdots + x_m^2\right), \end{aligned}$$

supostas definidas no conjunto de todos os pontos (do espaço $\mathbb{R}^3$ nos dois primeiros casos e de $\mathbb{R}^m$ no terceiro) nos quais têm sentido as expressões indicadas nos segundos membros das igualdades correspondentes.

Facilmente se reconhece que o domínio de f é o conjunto de todos os pontos de $\mathbb{R}^3$ que não pertencem a qualquer dos planos coordenados (isto é, dos planos dos xy, dos xz e dos yz, cujas equações são, respectivamente, $z = 0$, $y = 0$ e $x = 0$). Com efeito, a expressão $1/xyz$ tem sentido no corpo real sse x, y e z verificam as condições

$$x \neq 0, \quad y \neq 0 \quad \text{e} \quad z \neq 0.$$

Por sua vez, o domínio da função g é o cubo constituído por todos os pontos de $\mathbb{R}^3$ cujas coordenadas verificam as relações: $|x| \leq 1$, $|y| \leq 1$ e $|z| \leq 1$, isto é, o cubo centrado na origem, com faces paralelas aos planos coordenados e arestas de comprimento igual a 2. Para reconhecer que é este o domínio de g basta lembrar que se tem:

$$\lim_{n\to\infty} x^{2n} = \begin{cases} 0 & \text{se } |x| < 1 \\ 1 & \text{se } |x| = 1 \\ +\infty & \text{se } |x| > 1, \end{cases}$$

donde facilmente se deduz também que g toma o valor 0 em todos os pontos (x, y, z) situados no interior do cubo (isto é, que pertencem ao cubo mas não a qualquer das suas faces), o valor 1 em qualquer ponto situado numa face mas não numa aresta, o valor 2 nos pontos das arestas distintos dos vértices e, finalmente, o valor 3 em cada um dos vértices do cubo. Nos pontos (x, y, z) situados fora do cubo tem-se

$$\lim_{n\to\infty} \left(x^{2n} + y^{2n} + z^{2n}\right) = +\infty$$

não estando portanto a função g definida em qualquer desses pontos.

Finalmente, a função h tem por domínio todo o espaço $\mathbb{R}^m$ com excepção da origem; com efeito, $\log(x_1^2 + \cdots + x_m^2)$ converte-se num número real se forem atribuídos a $x_1, \ldots, x_m$ valores tais que

$$x_1^2 + x_2^2 + \cdots + x_m^2 > 0,$$

o que se verifica em qualquer ponto $(x_1, \ldots, x_m) \neq (0, 0, \ldots, 0)$.

Pode observar-se ainda que a função h é constante em cada uma das «hipersuperfícies» definidas (no espaço $\mathbb{R}^m$) por equações da forma

$$x_1^2 + x_2^2 + \cdots + x_m^2 = c,$$

com $c > 0$, as quais podem chamar-se «hipersuperfícies de nível» da função considerada (generalizando a noção de «linha de nível», introduzida anteriormente). No que respeita à representação geométrica de funções de mais de duas variáveis, limitar-nos-emos a observar que, mesmo para uma função de três variáveis reais, $u = f(x, y, z)$, não se pode já visualizar um «gráfico» (no espaço ordinário, tridimensional, não é possível figurar quatro eixos ortogonais dois a dois, para representar as variáveis x, y, z e u). Em tais casos pode ser útil recorrer à representação de «funções parciais», obtidas por fixação de algumas das variáveis.

Assim por exemplo, para estudar a função $z = x^2 + y^2 + t^2$ pode observar-se que, para $t = 0$, a função parcial correspondente, $z = x^2 + y^2$, tem por gráfico o parabolóide representado na Fig. 1.4; para $t = 1$, o gráfico da função parcial $z = x^2 + y^2 + 1$ é também um parabolóide, obtido do anterior por translação de uma unidade na direcção e sentido do eixo dos z, etc..

Deve, no entanto, observar-se que, embora a representação geométrica se torne menos cómoda e também menos útil no caso das funções de m variáveis reais, com $m > 2$, o estudo da teoria destas funções por via analítica se reduz quase sempre a uma generalização simples e directa da teoria correspondente para as funções de duas variáveis; em contrapartida, como teremos oportunidade de ver na sequência, a passagem de uma a duas variáveis «independentes» introduz, de facto, novas situações e algumas dificuldades em vários aspectos da teoria.

Capítulo 2

Estruturação algébrica e topológica de $\mathbb{R}^m$. Sucessões

2.1 O espaço vectorial $\mathbb{R}^m$; produto interno, norma e distância

Sendo m um número inteiro positivo, os elementos do conjunto $\mathbb{R}^m$ são, como sabemos, todas as sequências[1] (ou sucessões finitas) de m números reais, representáveis na forma:

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_m),$$

com $x_1, x_2, \ldots, x_m \in \mathbb{R}$. Como resulta da própria definição de sequência, se $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ e $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)$ são dois elementos de $\mathbb{R}^m$, a igualdade $\boldsymbol{x} = \boldsymbol{y}$ é verificada sse o forem conjuntamente as m igualdades:

$$x_1 = y_1, \quad x_2 = y_2, \quad \ldots, \quad x_m = y_m.$$

Assim, cada elemento $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$ determina, de forma unívoca, cada uma das suas coordenadas, $x_1, x_2, \ldots, x_m$ (designadas, respectivamente, por 1^{a}, 2^{a}, …, m^{a}coordenada de $\boldsymbol{x}$). Nestas condições, fixado $m \in \mathbb{N}_1$, para cada inteiro positivo $j \in \{1, 2, \ldots, m\}$, convencionaremos chamar *projecção de ordem j* e designar por p_j a aplicação de $\mathbb{R}^m$ em $\mathbb{R}$ que faz corresponder a cada $\boldsymbol{x} \in \mathbb{R}^m$ a sua j^{a} coordenada:

$$p_1(\boldsymbol{x}) = x_1, \ldots,\; p_m(\boldsymbol{x}) = x_m,$$

se $\boldsymbol{x} = (x_1, \ldots, x_m)$. Por exemplo, no caso $m = 3$ (e portanto com $\boldsymbol{x} \in \mathbb{R}^3$) os números reais $p_1(\boldsymbol{x})$, $p_2(\boldsymbol{x})$ e $p_3(\boldsymbol{x})$ corresponderiam respectivamente à abcissa, à ordenada e à cota do ponto do «espaço ordinário» identificado com o elemento $\boldsymbol{x}$ (cf. 1.3).

Convém-nos agora introduzir no conjunto $\mathbb{R}^m$ uma operação binária — chamada *adição* — definida pela forma seguinte: sendo $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ e

[1]Recorde-se que, sendo A um conjunto qualquer, uma sequência de m elementos de A é qualquer aplicação do conjunto dos m primeiros inteiros positivos, $\{1, 2, \ldots, m\}$, no conjunto A; a sequência que transforma cada inteiro positivo j (com $1 \leq j \leq m$) no elemento a_j de A é usualmente representada por $(a_1, a_2, \ldots, a_m)$.

$\boldsymbol{y} = (y_1, y_2, \ldots, y_m)$ elementos quaisquer de $\mathbb{R}^m$, a *soma de* $\boldsymbol{x}$ *e* $\boldsymbol{y}$, designada por $\boldsymbol{x} + \boldsymbol{y}$, será por definição o elemento de $\mathbb{R}^m$:

$$\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_m + y_m).$$

> Muitas vezes, para interpretar geometricamente os elementos de $\mathbb{R}^m$ — com $m = 3$ e, de forma análoga, nos casos $m = 1$ e $m = 2$ — é preferível, em lugar da identificação com «pontos» do espaço ordinário referida em 1.3, identificá-los com «vectores» (no sentido atribuído a este termo na Geometria elementar). Nesta interpretação, depois de fixado um referencial $Oxyz$ no espaço ordinário, o elemento (a, b, c) de $\mathbb{R}^3$ fica a corresponder, já não ao ponto P de coordenadas a, b, c no referencial considerado, mas antes ao vector representado por $\vec{OP}$ (onde O é a origem do referencial).
>
> Reconhece-se facilmente que, com esta última interpretação, a adição que acabamos de definir em $\mathbb{R}^m$ corresponde precisamente (se $m \leq 3$) à adição de vectores considerada na Geometria elementar (na qual a soma de dois vectores era geralmente determinada pela «regra do paralelogramo»).

Verifica-se imediatamente que a adição definida em $\mathbb{R}^m$ é comutativa ($\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$) e associativa ($\boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}) = (\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z}$, $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^m$); é também óbvia a existência de um elemento de $\mathbb{R}^m$ — o elemento com todas as coordenadas nulas, que designaremos por $\mathbf{0}$ — tal que, para qualquer $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$; além disso, dado $\boldsymbol{x} \in \mathbb{R}^m$ existe sempre um elemento do mesmo conjunto — chamado simétrico de $\boldsymbol{x}$ e designado por $-\boldsymbol{x}$ — que verifica a condição $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$ (se $\boldsymbol{x} = (x_1, \ldots, x_m)$, então $-\boldsymbol{x} = (-x_1, \ldots, -x_m)$ como é evidente). As quatro propriedades acabadas de mencionar podem sintetizar-se dizendo que $\mathbb{R}^m$, com a operação de adição considerada, é um grupo comutativo.

Aliás, dessas quatro propriedades podem deduzir-se muitas outras (como vimos no estudo dos números reais) de tal forma que, em termos algo imprecisos, poderá afirmar-se que a adição em $\mathbb{R}^m$ — bem como a subtracção, definida de maneira óbvia — gozam da generalidade das propriedades das operações homónimas definidas em $\mathbb{R}$.

Em contrapartida, a multiplicação de números reais não pode generalizar-se ao quadro dos espaços $\mathbb{R}^m$ mantendo todas as suas propriedades essenciais (por exemplo só para valores muito particulares de m é possível definir uma multiplicação de maneira que $\mathbb{R}^m$ fique munido de uma estrutura de corpo). Têm, no entanto, o maior interesse as duas operações de multiplicação que introduziremos na sequência, cada uma das quais constitui, sob certos aspectos, uma generalização da multiplicação usual de números reais.

A primeira dessas operações — chamada *multiplicação por escalares* (ou por números reais) — tem por dados um número real e um elemento de $\mathbb{R}^m$ sendo o resultado um elemento de $\mathbb{R}^m$. Em termos precisos, se $\alpha \in \mathbb{R}$ e $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$, chama-se *produto de* α *por* $\boldsymbol{x}$ e designa-se por $\alpha\boldsymbol{x}$ o elemento de $\mathbb{R}^m$:

$$\alpha\boldsymbol{x} = (\alpha x_1, \alpha x_2, \ldots, \alpha x_m).$$

(Geometricamente, no caso de ser $m \leq 3$, esta operação corresponde à multiplicação usual de um número real por um vector.)

Reconhece-se imediatamente que, se $\alpha, \beta \in \mathbb{R}$ e $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, se tem:

$$\begin{aligned}
&\alpha(\boldsymbol{x}+\boldsymbol{y}) = \alpha\boldsymbol{x} + \alpha\boldsymbol{y}\\
&(\alpha+\beta)\boldsymbol{x} = \alpha\boldsymbol{x} + \beta\boldsymbol{x}\\
&\alpha(\beta\boldsymbol{x}) = (\alpha\beta)\boldsymbol{x}\\
&1\boldsymbol{x} = \boldsymbol{x}
\end{aligned}$$

Todas estas propriedades — e ainda as que resumimos ao afirmar que $\mathbb{R}^m$ é um grupo comutativo relativamente à adição — podem por sua vez sintetizar-se dizendo que $\mathbb{R}^m$, munido com as operações de adição e de multiplicação por escalares, é um espaço vectorial sobre o corpo real, ou apenas um espaço vectorial real[2].

Fixado um inteiro positivo m, convencionaremos designar por $\boldsymbol{e}_j$ (para $j = 1, \ldots, m$) o vector de $\mathbb{R}^m$ com todas as coordenadas nulas excepto a de ordem j, que é igual a 1:

$$\boldsymbol{e}_1 = (1, 0, \ldots, 0), \quad \ldots, \quad \boldsymbol{e}_m = (0, 0, \ldots, 0, 1).$$

Por exemplo, para $m = 2$, haverá só dois vectores a considerar: $\boldsymbol{e}_1 = (1, 0)$ e $\boldsymbol{e}_2 = (0, 1)$.

Nestas condições, sendo $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ um vector qualquer de $\mathbb{R}^m$ ter-se-á, atendendo às definições de adição e multiplicação por escalares:

$$\begin{aligned}
\boldsymbol{x} &= (x_1, \ldots, x_m) = (x_1, 0, \ldots, 0) + \cdots + (0, 0, \ldots, 0, x_m)\\
&= x_1(1, 0, \ldots, 0) + \cdots + x_m(0, 0, \ldots, 0, 1)
\end{aligned}$$

ou

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_m\boldsymbol{e}_m.$$

Sendo $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ vectores de um espaço vectorial real (por exemplo, de um espaço $\mathbb{R}^m$, com $m \in \mathbb{N}_1$), diz-se que um vector $\boldsymbol{u}$ do mesmo espaço é uma *combinação linear* de $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ sse existem k números reais $\alpha_1, \alpha_2, \ldots, \alpha_k$ tais que se verifique a igualdade:

$$\boldsymbol{u} = \alpha_1\boldsymbol{u}_1 + \alpha_2\boldsymbol{u}_2 + \cdots + \alpha_k\boldsymbol{u}_k.$$

Quando qualquer vector $\boldsymbol{u}$ do espaço vectorial considerado pode exprimir-se como combinação linear de $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ de forma única (isto é, quando para todo o $\boldsymbol{u}$ existem escalares $\alpha_1, \alpha_2, \ldots, \alpha_k$ por forma que a igualdade anterior seja verificada e, além disso, a verificação conjunta dessa igualdade e de

$$\boldsymbol{u} = \beta_1\boldsymbol{u}_1 + \beta_2\boldsymbol{u}_2 + \cdots + \beta_k\boldsymbol{u}_k$$

[2]É habitual chamar vectores aos elementos de qualquer espaço vectorial; por isso, de aqui em diante, chamaremos correntemente vectores aos elementos de $\mathbb{R}^m$. Os números reais serão também designados por escalares.

implique $\alpha_1 = \beta_1$, $\alpha_2 = \beta_2$, …, $\alpha_k = \beta_k$) diz-se que os vectores $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ constituem uma *base* do espaço vectorial considerado.

Usando esta terminologia, corrente em Álgebra Linear, poderia dizer-se que os vectores $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_m$ constituem uma base do espaço vectorial $\mathbb{R}^m$ (será útil, como exercício, justificar cuidadosamente esta afirmação). Deve observar-se que o espaço $\mathbb{R}^m$ tem infinitas outras bases; aquela a que nos referimos especialmente costuma ser designada por *base canónica* de $\mathbb{R}^m$.

Introduziremos agora a segunda das operações de «multiplicação» a que anteriormente fizemos referência. Sendo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$ e $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)$, chama-se *produto interno de* $\boldsymbol{x}$ *e* $\boldsymbol{y}$ ao número real

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_m y_m.$$

Assim, por exemplo, ter-se-á, para $i, j \in \{1, 2, \ldots, m\}$.

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j. \end{cases}$$

Verificam-se sem qualquer dificuldade as seguintes propriedades do conceito de produto interno: para qualquer $\alpha \in \mathbb{R}$ e quaisquer $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^m$,

P1) $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x}$

P2) $(\boldsymbol{x} + \boldsymbol{y}) \cdot \boldsymbol{z} = \boldsymbol{x} \cdot \boldsymbol{z} + \boldsymbol{y} \cdot \boldsymbol{z}$ e $\boldsymbol{x} \cdot (\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{x} \cdot \boldsymbol{z}$

P3) $(\alpha \boldsymbol{x}) \cdot \boldsymbol{y} = \alpha(\boldsymbol{x} \cdot \boldsymbol{y}) = \boldsymbol{x} \cdot (\alpha \boldsymbol{y})$

P4) $\boldsymbol{0} \cdot \boldsymbol{0} = 0$ e, para qualquer $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{x} \cdot \boldsymbol{x} > 0$.

A noção de produto interno será usada frequentemente na sequência deste curso; neste momento, convém-nos utilizá-la para introduzir um outro conceito, fundamental em tudo o que segue: a norma de um vector de $\mathbb{R}^m$, generalização da noção de módulo de um número real (ou da de módulo — ou comprimento — de um vector, no sentido considerado na Geometria).

Sendo $\boldsymbol{x} \in \mathbb{R}^m$, chamaremos *norma* de $\boldsymbol{x}$ e designaremos pelo símbolo $\|\boldsymbol{x}\|$, o número real:

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$$

(observe-se que, segundo P4, se tem $\boldsymbol{x} \cdot \boldsymbol{x} \geq 0$, para qualquer $\boldsymbol{x} \in \mathbb{R}^m$).

No caso particular de $\boldsymbol{x}$ ser um vector de $\mathbb{R}^2$ ou $\mathbb{R}^3$ — $\boldsymbol{x} = (x_1, x_2)$ ou $\boldsymbol{x} = (x_1, x_2, x_3)$ — interpretável, no plano ou no espaço ordinário, como um vector $\vec{OP}$ — a norma de $\boldsymbol{x}$, dada por:

$$\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}} = \begin{cases} \sqrt{x_1^2 + x_2^2} & \text{se } m = 2 \\ \sqrt{x_1^2 + x_2^2 + x_3^2} & \text{se } m = 3, \end{cases}$$

coincidirá com o módulo do vector $\vec{OP}$ (ou com a distância do ponto P à origem). Se $m = 1$, o vector $\boldsymbol{x} = (x_1)$ pode identificar-se com o número real x_1 e a sua norma, $\sqrt{x_1^2}$, coincide com o módulo do número real x_1.

Para números reais, sabemos bem que são verificadas as propriedades seguintes:

M1) $|x| \geq 0$ (com $|x| = 0$ sse $x = 0$)

M2) $|xy| = |x||y|$

M3) $|x + y| \leq |x| + |y|$.

Convém-nos agora ver em que medida estas propriedades são generalizáveis (através do conceito de norma) ao caso dos espaços $\mathbb{R}^m$.

A extensão da primeira é trivial; com efeito, se $\boldsymbol{x} = (x_1, \ldots, x_m) \in \mathbb{R}^m$, tem-se, evidentemente:

N1) $\|\boldsymbol{x}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_m^2} > 0$, se $\boldsymbol{x} \neq \mathbf{0}$ e $\|\mathbf{0}\| = 0$.

Para tentar generalizar M2), há que considerar separadamente o produto por escalares e o produto interno. No primeiro caso, deduz-se imediatamente:

$$\begin{aligned} \|\alpha\boldsymbol{x}\| &= \|\alpha(x_1, \ldots, x_m)\| = \|(\alpha x_1, \ldots, \alpha x_m)\| \\ &= \sqrt{\alpha^2 x_1^2 + \cdots + \alpha^2 x_m^2} = |\alpha|\sqrt{x_1^2 + \cdots + x_m^2} \end{aligned}$$

ou

N2) $\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$.

No segundo, obtém-se uma relação de grande utilidade, chamada *desigualdade de Cauchy-Schwarz*:

$$|\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|, \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m.$$

Esta relação pode justificar-se pela forma seguinte: sendo $\boldsymbol{x}$ e $\boldsymbol{y}$ dois vectores quaisquer de $\mathbb{R}^m$, observe-se em primeiro lugar que, para qualquer $\alpha \in \mathbb{R}$, se tem (de acordo com a definição de norma):

$$(\boldsymbol{x} + \alpha\boldsymbol{y}) \cdot (\boldsymbol{x} + \alpha\boldsymbol{y}) = \|\boldsymbol{x} + \alpha\boldsymbol{y}\|^2 \geq 0.$$

Por outro lado, das propriedades indicadas do produto interno logo resulta:

$$\begin{aligned} (\boldsymbol{x} + \alpha\boldsymbol{y}) \cdot (\boldsymbol{x} + \alpha\boldsymbol{y}) &= \boldsymbol{x} \cdot (\boldsymbol{x} + \alpha\boldsymbol{y}) + \alpha\boldsymbol{y} \cdot (\boldsymbol{x} + \alpha\boldsymbol{y}) \\ &= \boldsymbol{x} \cdot \boldsymbol{x} + \alpha\boldsymbol{x} \cdot \boldsymbol{y} + \alpha\boldsymbol{y} \cdot \boldsymbol{x} + \alpha^2\boldsymbol{y} \cdot \boldsymbol{y} \\ &= \|\boldsymbol{x}\|^2 + 2(\boldsymbol{x} \cdot \boldsymbol{y})\alpha + \alpha^2\|\boldsymbol{y}\|^2, \end{aligned}$$

o que permite deduzir que, para qualquer $\alpha \in \mathbb{R}$,

$$\|\boldsymbol{y}\|^2\alpha^2 + 2(\boldsymbol{x} \cdot \boldsymbol{y})\alpha + \|\boldsymbol{x}\|^2 \geq 0.$$

A expressão que figura no 1º membro desta desigualdade é um trinómio do 2º grau em α (com coeficientes dependentes dos vectores dados, $\boldsymbol{x}$ e $\boldsymbol{y}$); como é sabido, para que o trinómio $a\alpha^2 + b\alpha + c$ assuma valores não negativos qualquer que seja o valor real atribuído a α, é *necessário* que o seu discriminante não seja positivo ($b^2 - 4ac \leq 0$).[3] Pode portanto concluir-se que se verifica necessariamente a relação:

$$|\boldsymbol{x} \cdot \boldsymbol{y}|^2 - \|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2 \leq 0,$$

donde imediatamente se deduz a desigualdade de Cauchy-Schwarz.

No caso particular em que $\boldsymbol{x}$ e $\boldsymbol{y}$ são vectores não nulos do espaço $\mathbb{R}^2$ ou $\mathbb{R}^3$, interpretáveis como vectores no plano ou no espaço, pode verificar-se que é válida a relação[4]:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos\theta,$$

onde θ designa o ângulo dos dois vectores (Figura 2.1). Este facto sugere que se procure definir mais geralmente ângulo de dois vectores não nulos do espaço $\mathbb{R}^m$, $\boldsymbol{x}$ e $\boldsymbol{y}$, como sendo o número real:

$$\theta = \arccos \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}.$$

Contudo, não seria legítimo adoptar esta definição se não estivesse assegurado que (para $\boldsymbol{x}, \boldsymbol{y} \neq \mathbf{0}$) a expressão

$$\frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|}$$

assume apenas valores do intervalo $[-1, 1]$; esta garantia resulta imediatamente da desigualdade de Cauchy-Schwarz.

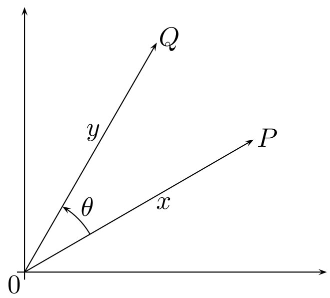

Figura 2.1

É útil mencionar que, quando o produto interno de $\boldsymbol{x}$ e $\boldsymbol{y}$ é nulo (o que se passa se algum desses vectores é igual a $\mathbf{0}$, ou se o seu ângulo é

[3] Convém notar que a afirmação é correcta, mesmo na hipótese de ser nulo o coeficiente de α^2.

[4] Com referencial ortonormado; não sendo o referencial ortonormado, a relação mantém-se, mas o produto interno de $\boldsymbol{x}$ e $\boldsymbol{y}$ não pode ser definido pela fórmula que indicámos.

igual a $\pi/2$) costuma dizer-se que os vectores $\boldsymbol{x}$ e $\boldsymbol{y}$ são *ortogonais*. Pode observar-se ainda que, das igualdades:

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 = (\boldsymbol{x}+\boldsymbol{y})\cdot(\boldsymbol{x}+\boldsymbol{y}) = \|\boldsymbol{x}\|^2 + 2\boldsymbol{x}\cdot\boldsymbol{y} + \|\boldsymbol{y}\|^2,$$

resulta, no caso de $\boldsymbol{x}$ e $\boldsymbol{y}$ serem ortogonais:

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2,$$

relação que generaliza o clássico teorema de Pitágoras.

Podemos agora generalizar, para vectores do espaço $\mathbb{R}^m$, a propriedade M3, relativa ao módulo de uma soma. Sendo $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^m$, tem-se:

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + 2\boldsymbol{x}\cdot\boldsymbol{y} + \|\boldsymbol{y}\|^2\\ &\le \|\boldsymbol{x}\|^2 + 2|\boldsymbol{x}\cdot\boldsymbol{y}| + \|\boldsymbol{y}\|^2,\end{aligned}$$

donde, atendendo à desigualdade de Cauchy-Schwarz,

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2 \le \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 = (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2.$$

Desta relação deduz-se imediatamente a propriedade que pretendíamos obter:

N3) $\|\boldsymbol{x}+\boldsymbol{y}\| \le \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$

Esta propriedade estende-se ao caso de uma soma com qualquer número (finito) de parcelas. Assim (como pode justificar-se por indução, usando a associatividade da adição de vectores e N3), se $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_k\in\mathbb{R}^m$, é válida a relação:

$$\|\boldsymbol{u}_1+\boldsymbol{u}_2+\cdots+\boldsymbol{u}_k\| \le \|\boldsymbol{u}_1\| + \|\boldsymbol{u}_2\| + \cdots + \|\boldsymbol{u}_k\|.$$

Costuma dizer-se que um espaço vectorial real é um *espaço normado* se estiver fixada uma aplicação que faça corresponder a cada vector $\boldsymbol{x}$ do espaço considerado um número real $p(\boldsymbol{x})$, por forma que sejam verificadas as condições seguintes:

1. Para qualquer vector $\boldsymbol{x}$ do espaço, $p(\boldsymbol{x})\ge 0$; $p(\boldsymbol{x}) = 0$ sse $\boldsymbol{x}$ é o vector nulo;
2. $p(\alpha\boldsymbol{x}) = |\alpha|p(\boldsymbol{x})$ (para qualquer real α e qualquer vector $\boldsymbol{x}$);
3. $p(\boldsymbol{x}+\boldsymbol{y}) \le p(\boldsymbol{x}) + p(\boldsymbol{y})$ (quaisquer que sejam os vectores $\boldsymbol{x}$ e $\boldsymbol{y}$).

Em tal caso, o número real $p(\boldsymbol{x})$ chama-se ainda *norma do vector* $\boldsymbol{x}$. Assim, as três propriedades N1, N2 e N3 atrás verificadas permitem-nos afirmar que o espaço $\mathbb{R}^m$, com a aplicação que associa a cada $\boldsymbol{x}$ o real $\|\boldsymbol{x}\|$, é um espaço normado.

Convém observar que podem definir-se no mesmo espaço vectorial $\mathbb{R}^m$ outras «normas», isto é, outras aplicações $p:\mathbb{R}^m\to\mathbb{R}$ verificando as três condições atrás referidas (dir-se-ia então que se tinham introduzido em $\mathbb{R}^m$ outras estruturas de espaço normado). Por exemplo, é fácil verificar que são também normas sobre $\mathbb{R}^m$ as aplicações que fazem corresponder a cada $\boldsymbol{x} = (x_1, \dots, x_m)$ os números reais:

$$p^*(\boldsymbol{x}) = |x_1| + |x_2| + \cdots + |x_m|$$

ou

$$\overline{p}(\boldsymbol{x}) = \max\{|x_1|, |x_2|, \dots, |x_m|\}.$$

É curioso observar — mesmo que de forma necessariamente pouco precisa — que em todas as definições e teoremas subsequentes que farão intervir o conceito de norma, poderíamos utilizar, em lugar da norma inicialmente considerada ($\|\boldsymbol{x}\|$), as duas que acabam de ser mencionadas (ou qualquer das outras infinitas normas que podem definir-se em $\mathbb{R}^m$), obtendo-se sempre resultados essencialmente equivalentes. No entanto, quando na sequência voltarmos a utilizar o termo «norma» deverá sistematicamente entender-se (salvo menção expressa em contrário) que pretendemos referir-nos à única noção de norma considerada antes do início desta nota.

Sendo $\boldsymbol{x} = (x_1, \dots, x_m)$ e $\boldsymbol{y} = (y_1, \dots, y_m)$ dois elementos quaisquer de $\mathbb{R}^m$, chamaremos *distância de* $\boldsymbol{x}$ *a* $\boldsymbol{y}$ — e designaremos por $d(\boldsymbol{x}, \boldsymbol{y})$ — o número real:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_m - y_m)^2}.$$

Para ver como esta definição é natural, basta notar que, no caso de ser m igual a 1, 2 ou 3 — e interpretando agora, de preferência, $\boldsymbol{x}$ e $\boldsymbol{y}$ como pontos da recta, do plano ou do espaço — a igualdade anterior se transforma nas fórmulas para o cálculo da distância de dois pontos, bem conhecidas da Geometria Analítica.

Verificam-se sem qualquer dificuldade as seguintes propriedades da noção de distância: quaisquer que sejam os vectores $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^m$,

D1) $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ e $d(\boldsymbol{x}, \boldsymbol{y}) = 0$ sse $\boldsymbol{x} = \boldsymbol{y}$

D2) $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$

D3) $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$ (*desigualdade triangular*).

Sendo E um conjunto qualquer, costuma chamar-se *distância* sobre E a qualquer função que associe a cada par ordenado (x, y) de elementos de E um número real que poderemos designar ainda por $d(x, y)$, por forma que sejam verificadas, sempre que $x, y, z \in E$, as condições D1, D2 e D3. O conjunto E, com uma determinada função de distância, constitui o que se chama um *espaço métrico*.

Assim o conjunto $\mathbb{R}^m$, com a distância $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$, é um exemplo de um espaço métrico.

A noção de espaço métrico, e também a de espaço normado referida anteriormente, têm, contudo, possibilidades de utilização muito mais vastas do que a correspondente ao caso dos espaços $\mathbb{R}^m$ que estudaremos neste curso. Por exemplo, em muitos espaços funcionais importantes (espaços cujos «vectores» são funções contínuas, funções diferenciáveis, funções integráveis, etc.) podem introduzir-se de maneira natural noções de norma ou de distância, a partir das quais são generalizáveis em grande parte a esses espaços os conceitos e os resultados mais significativos que aqui estudaremos apenas para os espaços $\mathbb{R}^m$. Nomeadamente, no quadro bastante geral dos espaços normados, pode estruturar-se um cálculo diferencial análogo ao que vamos estudar neste curso e que o contém como caso muito particular. E o mais interessante é que, longe de constituirem meras especulações de interesse puramente teórico, essas generalizações da teoria das funções de variável real — que constituem um dos objectos de um ramo da Matemática chamado Análise funcional — são susceptíveis de aplicações de grande alcance na Física, na Engenharia e em diversos outros domínios da Ciência e da Técnica.

Recorrendo à noção de norma (tal como no caso de $\mathbb{R}$ recorremos à de módulo) ou, se preferirmos, à de distância, podemos agora introduzir em $\mathbb{R}^m$ vários conceitos fundamentais, que darão uma base sólida para o nosso estudo do cálculo diferencial em $\mathbb{R}^m$. Esse trabalho será feito, em grande parte, nos parágrafos seguintes deste capítulo, reservando-se a parte restante do presente parágrafo apenas à generalização a $\mathbb{R}^m$ da noção de vizinhança de um número real.

Recordemos que, em $\mathbb{R}$, designámos por vizinhança ϵ de um ponto a ($a \in \mathbb{R}$, $\epsilon > 0$) o conjunto de todos os reais x tais que $|x - a| < \epsilon$. Numa ordem de ideias semelhante, sendo agora $\boldsymbol{a} = (a_1, a_2, \ldots, a_m) \in \mathbb{R}^m$ e ϵ um número real positivo, chamaremos *bola* (ou *bola aberta*) de centro $\boldsymbol{a}$ e raio ϵ ao conjunto de todos os $\boldsymbol{x} \in \mathbb{R}^m$ tais que $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$; para designar este conjunto usaremos o símbolo $B_\epsilon(\boldsymbol{a})$. Eventualmente faremos também referência à *bola fechada* de centro $\boldsymbol{a}$ e raio ϵ, conjunto de todos os pontos $\boldsymbol{x}$ de $\mathbb{R}^m$ tais que $\|\boldsymbol{x} - \boldsymbol{a}\| \leq \epsilon$.

Para $m = 1$ (e $a \in \mathbb{R}$, portanto) $B_\epsilon(a)$ é precisamente a vizinhança ϵ de a já conhecida, representável na recta por um segmento (privado dos extremos) com centro no ponto a e comprimento 2ϵ; para $n = 2$ [ou $n = 3$] e $\boldsymbol{a} = (a_1, a_2)$ [ou $\boldsymbol{a} = (a_1, a_2, a_3)$], a imagem geométrica de $B_\epsilon(\boldsymbol{a})$ é o circulo «aberto» [ou a esfera «aberta»] de centro $\boldsymbol{a}$ e raio ϵ, isto é o conjunto de todos os pontos do plano [ou do espaço] cuja distância ao ponto $\boldsymbol{a}$ é menor do que ϵ.

Com $\boldsymbol{a} \in \mathbb{R}^m$ e sendo ϵ e ϵ' dois reais positivos tais que $\epsilon < \epsilon'$, tem-se, como é evidente, $B_\epsilon(\boldsymbol{a}) \subset B_{\epsilon'}(\boldsymbol{a})$. É também fácil verificar que qualquer bola de $\mathbb{R}^m$ é um conjunto infinito e ainda que a intersecção de todas as bolas centradas no ponto $\boldsymbol{a}$ é o conjunto formado apenas por este ponto, $\{\boldsymbol{a}\}$ (o qual, porém, não é uma bola). Observemos finalmente que, sendo $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^m$ e $\boldsymbol{a} \neq \boldsymbol{b}$, é sempre possível determinar uma bola centrada em $\boldsymbol{a}$, $B_\epsilon(\boldsymbol{a})$, e outra bola centrada em $\boldsymbol{b}$, $B_\delta(\boldsymbol{b})$, que sejam disjuntas (bastará escolher os reais positivos δ e ϵ por forma que $\delta + \epsilon \leq \|\boldsymbol{b} - \boldsymbol{a}\|$).

Na estruturação de alguns conceitos fundamentais da Análise em $\mathbb{R}^m$ — por exemplo, o de limite — as «bolas» acabadas de definir desempenharão naturalmente o papel que coube às «vizinhanças», no caso de $\mathbb{R}$. Vê-lo-emos já no parágrafo seguinte, no que respeita à noção de limite de uma sucessão, e no próximo capítulo, ao estudarmos limites e continuidade para funções mais gerais. Além disso, no parágrafo final deste capítulo, a noção de bola será também utilizada para a definição de diversos conceitos de natureza «topológica», indispensáveis no estudo do cálculo infinitesimal para funções de mais de uma variável real.

> Convém no entanto deixar aqui registado que, no quadro dos espaços métricos, o termo *vizinhança* é usado numa acepção muito mais geral do que aquela a que acabamos de referir-nos. Concretamente, sendo E um espaço métrico e $\boldsymbol{a}$ um elemento de E, chama-se *vizinhança de* $\boldsymbol{a}$ a qualquer conjunto $V \subset E$ que contenha alguma bola de centro $\boldsymbol{a}$; assim, V é vizinhança de $\boldsymbol{a}$ sse existe $\epsilon > 0$ tal que $B_\epsilon(\boldsymbol{a}) \subset V$ ($B_\epsilon(\boldsymbol{a})$ é , por definição o conjunto de todos os elementos $\boldsymbol{x} \in E$ tais que $d(\boldsymbol{x}, \boldsymbol{a}) < \epsilon$, sendo d a função distância considerada no espaço métrico E).
>
> Neste texto, porém, só bastante mais adiante teremos necessidade de utilizar esta noção mais geral de vizinhança, e mesmo assim apenas no quadro do espaço $\mathbb{R}^m$.

2.2 Sucessões em $\mathbb{R}^m$

Comecemos por recordar que, sendo A um conjunto qualquer, se chama *sucessão em* A (ou *sucessão de termos em* A) a qualquer aplicação do conjunto $\mathbb{N}_1$, dos inteiros positivos, no conjunto A.

Se $\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_n\ \ldots$ é uma sucessão em $\mathbb{R}^m$, e se, para cada inteiro positivo $j \leq m$, designarmos por u_{nj} a j^{a} coordenada de $\boldsymbol{u}_n$ (isto é, se pusermos $p_j(\boldsymbol{u}_n) = u_{nj}$), ter-se-á:

$$\begin{aligned}
\boldsymbol{u}_1 &= (u_{11}, u_{12}, \ldots, u_{1m})\\
\boldsymbol{u}_2 &= (u_{21}, u_{22}, \ldots, u_{2m})\\
&\ldots\\
\boldsymbol{u}_n &= (u_{n1}, u_{n2}, \ldots, u_{nm})\\
&\ldots
\end{aligned}$$

Assim, cada sucessão em $\mathbb{R}^m$ determina m sucessões de termos reais, a que chamaremos *sucessões coordenadas* da sucessão dada; mais precisamente, a sucessão numérica:

$$u_{1j}\ u_{2j}\ \ldots\ u_{nj}\ \ldots$$

é a *sucessão coordenada de ordem* j da sucessão $\boldsymbol{u}_n$ considerada ($1 \leq j \leq m$).[5]

Por exemplo, para a sucessão em $\mathbb{R}^2$

$$\boldsymbol{v}_n = \left(\frac{1}{n}, n\right)$$

[5] Já sabemos que só como «abuso de notação» pode aceitar-se o uso do símbolo $\boldsymbol{u}_n$ — que designa o termo de ordem n da sucessão, isto é, o valor por ela assumido no ponto n — para designar a própria sucessão.

as sucessões coordenadas são

$$v_{n1} = \frac{1}{n} \quad \text{e} \quad v_{n2} = n.$$

Estendem-se naturalmente às sucessões em $\mathbb{R}^m$ as operações algébricas definidas no parágrafo 2.1. Assim, sendo $\boldsymbol{u}_n$ e $\boldsymbol{v}_n$ sucessões em $\mathbb{R}^m$ e $\alpha \in \mathbb{R}$, a *soma de* $\boldsymbol{u}_n$ *e* $\boldsymbol{v}_n$ e o *produto de* α *por* $\boldsymbol{u}_n$ são, respectivamente, as sucessões em $\mathbb{R}^m$:

$$\boldsymbol{u}_1 + \boldsymbol{v}_1 \quad \boldsymbol{u}_2 + \boldsymbol{v}_2 \quad \dots \quad \boldsymbol{u}_n + \boldsymbol{v}_n \quad \dots$$

e

$$\alpha\boldsymbol{u}_1 \quad \alpha\boldsymbol{u}_2 \quad \dots \quad \alpha\boldsymbol{u}_n \quad \dots$$

e o *produto interno de* $\boldsymbol{u}_n$ *e* $\boldsymbol{v}_n$ é a sucessão de termos reais:

$$\boldsymbol{u}_1 \cdot \boldsymbol{v}_1 \quad \boldsymbol{u}_2 \cdot \boldsymbol{v}_2 \quad \dots \quad \boldsymbol{u}_n \cdot \boldsymbol{v}_n \quad \dots$$

Introduziremos agora a seguinte definição, que generaliza de forma inteiramente natural uma outra bem conhecida do estudo das sucessões reais:

Seja $\boldsymbol{u}_n$ uma sucessão em $\mathbb{R}^m$ e $\boldsymbol{u}$ um vector de $\mathbb{R}^m$; diz se que $\boldsymbol{u}_n$ *tende* ou *converge para* $\boldsymbol{u}$ — e escreve-se $\boldsymbol{u}_n \to \boldsymbol{u}$ — sse, qualquer que seja a bola centrada em $\boldsymbol{u}$, $B_\epsilon(\boldsymbol{u})$, existe um inteiro positivo p tal que $\boldsymbol{u}_n \in B_\epsilon(\boldsymbol{u})$ para todo o $n > p$.

Reconhece-se sem dificuldade que esta definição poderia também ser formulada, equivalentemente, de qualquer dos modos seguintes:

- $\boldsymbol{u}_n$ converge para $\boldsymbol{u}$ sse, para todo o $\epsilon > 0$ existe p tal que $n > p \Rightarrow \|\boldsymbol{u}_n - \boldsymbol{u}\| < \epsilon$;

ou:

- $\boldsymbol{u}_n$ converge para $\boldsymbol{u}$ sse a sucessão real $\|\boldsymbol{u}_n - \boldsymbol{u}\|$ converge para 0.

Por exemplo, a sucessão em $\mathbb{R}^3$:

$$\boldsymbol{u}_n = \left(\frac{n-1}{n}, 0, \frac{1}{3^n}\right)$$

converge para o vector $\boldsymbol{e}_1 = (1, 0, 0)$. Para o reconhecer, basta notar que:

$$\|\boldsymbol{u}_n - \boldsymbol{e}_1\| = \sqrt{\frac{1}{n^2} + \frac{1}{3^{2n}}}$$

é um infinitésimo.

Naturalmente, diz-se que uma sucessão em $\mathbb{R}^m$, $\boldsymbol{u}_n$, é convergente sse existe $\boldsymbol{u} \in \mathbb{R}^m$ tal que $\boldsymbol{u}_n \to \boldsymbol{u}$.

Antes de prosseguir, convém fazer uma observação simples, que nos facilitará a obtenção de resultados posteriores.

Como vimos, qualquer vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$ pode exprimir-se como combinação linear dos vectores da base canónica, pela forma seguinte:

$$\boldsymbol{x} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_m\boldsymbol{e}_m.$$

Desta igualdade resulta, atendendo a propriedades da norma já estudadas:

$$\begin{aligned}\|\boldsymbol{x}\| &= \|x_1\boldsymbol{e}_1 + \cdots + x_m\boldsymbol{e}_m\| \\ &\le \|x_1\boldsymbol{e}_1\| + \cdots + \|x_m\boldsymbol{e}_m\| \\ &= |x_1|\|\boldsymbol{e}_1\| + \cdots + |x_m|\|\boldsymbol{e}_m\| \\ &= |x_1| + \cdots + |x_m|.\end{aligned}$$

Por outro lado (escolhido arbitrariamente um inteiro positivo $j \le m$), se multiplicarmos internamente por $\boldsymbol{e}_j$ ambos os membros daquela mesma igualdade, obteremos:

$$\begin{aligned}\boldsymbol{x} \cdot \boldsymbol{e}_j &= (x_1\boldsymbol{e}_1 + \cdots + x_m\boldsymbol{e}_m) \cdot \boldsymbol{e}_j \\ &= x_1(\boldsymbol{e}_1 \cdot \boldsymbol{e}_j) + \cdots + x_m(\boldsymbol{e}_m \cdot \boldsymbol{e}_j) \\ &= x_j,\end{aligned}$$

donde, atendendo à desigualdade de Cauchy-Schwarz, se deduz imediatamente:

$$|x_j| = |\boldsymbol{x} \cdot \boldsymbol{e}_j| \le \|\boldsymbol{x}\|\|\boldsymbol{e}_j\| = \|\boldsymbol{x}\|.$$

Assim, para qualquer $j \in \{1, 2, \ldots, m\}$, tem-se:

$$|x_j| \le \|\boldsymbol{x}\| \le |x_1| + |x_2| + \cdots + |x_m|.$$

Consideremos agora uma sucessão em $\mathbb{R}^m$, $\boldsymbol{u}_n = (u_{n1}, u_{n2}, \ldots, u_{nm})$ e um vector $\boldsymbol{a} = (a_1, a_2, \ldots, a_m)$ do mesmo espaço.

De acordo com a observação precedente, ter-se-á (para todo o inteiro positivo $j \le m$ e todo o $n \in \mathbb{N}_1$):

$$|u_{nj} - a_j| \le \|\boldsymbol{u}_n - \boldsymbol{a}\| \le |u_{n1} - a_1| + \cdots + |u_{nm} - a_m|.$$

A primeira destas desigualdades mostra que, se $\boldsymbol{u}_n \to \boldsymbol{a}$ (isto é, se $\|\boldsymbol{u}_n - \boldsymbol{a}\| \to 0$) também $|u_{nj} - a_j| \to 0$ e portanto (qualquer que seja o inteiro positivo $j \le m$) a sucessão coordenada de ordem j, u_{nj}, converge para a_j (em $\mathbb{R}$); a segunda desigualdade permite reconhecer que, reciprocamente, se se tiver $u_{nj} \to a_j$ para $j = 1, \ldots, m$ (o que implica que a soma

$$|u_{n1} - a_1| + |u_{n2} - a_2| + \cdots + |u_{nm} - a_m|$$

tenda para 0), $\|\boldsymbol{u}_n - \boldsymbol{a}\|$ tenderá para 0 e portanto $\boldsymbol{u}_n$ tenderá para $\boldsymbol{a}$ em $\mathbb{R}^m$.

Pode portanto enunciar-se o seguinte:

Teorema 2.1. *Para que uma sucessão em $\mathbb{R}^m$, $\boldsymbol{u}_n = (u_{n1}, \ldots, u_{nm})$ seja convergente é necessário e suficiente que o sejam todas as suas sucessões coordenadas; além disso, na hipótese de convergência de $\boldsymbol{u}_n$ para $\boldsymbol{a} = (a_1, \ldots, a_m)$ tem-se, para $j = 1, \ldots, m$:*

$$a_j = \lim_{n\to\infty} u_{nj}.$$

Daqui decorre imediatamente a unicidade do limite: Se, com $\boldsymbol{b} = (b_1, \ldots, b_m)$, se tiver conjuntamente $\boldsymbol{u}_n \to \boldsymbol{a}$ e $\boldsymbol{u}_n \to \boldsymbol{b}$, ter-se-á também:

$$a_j = \lim_{n\to\infty} u_{nj}, \qquad b_j = \lim_{n\to\infty} u_{nj}$$

e portanto, atendendo à unicidade do limite para sucessões reais, $a_j = b_j$ (para $j = 1, \ldots, m$), isto é, $\boldsymbol{a} = \boldsymbol{b}$.

Naturalmente, quando $\boldsymbol{u}_n$ é convergente, chama-se *limite de* $\boldsymbol{u}_n$ ao (único) vector $\boldsymbol{a}$ tal que $\boldsymbol{u}_n \to \boldsymbol{a}$, podendo então escrever-se $\lim_{n\to\infty} \boldsymbol{u}_n = \boldsymbol{a}$ ou apenas $\lim \boldsymbol{u}_n = \boldsymbol{a}$.

A partir das definições de convergência e limite para sucessões em $\mathbb{R}^m$ ou então (como fizemos na precedente justificação da unicidade do limite) recorrendo ao Teorema 2.1 e a resultados bem conhecidos para as sucessões reais, obtêm-se sem qualquer dificuldade as propriedades seguintes, que nos limitamos a enunciar ($\boldsymbol{u}_n, \boldsymbol{v}_n$ são sucessões em $\mathbb{R}^m$ e $\boldsymbol{a} \in \mathbb{R}^m$; a_n é uma sucessão real).

- *Se para todo o n (a partir de alguma ordem) $\boldsymbol{u}_n = \boldsymbol{a}$, então* $\lim \boldsymbol{u}_n = \boldsymbol{a}$.

- *Se* $\lim \boldsymbol{u}_n = \boldsymbol{a}$ *e $\boldsymbol{u}_{p_n}$ é uma subsucessão de $\boldsymbol{u}_n$,* $\lim \boldsymbol{u}_{p_n} = \boldsymbol{a}$.

- *Se $\boldsymbol{u}_n$ e $\boldsymbol{v}_n$ são sucessões convergentes, $\boldsymbol{u}_n + \boldsymbol{v}_n$, $\boldsymbol{u}_n - \boldsymbol{v}_n$, $\boldsymbol{u}_n \cdot \boldsymbol{v}_n$ e $\|\boldsymbol{u}_n\|$ também o são e:*

$$\begin{aligned}
&\lim(\boldsymbol{u}_n + \boldsymbol{v}_n) = \lim \boldsymbol{u}_n + \lim \boldsymbol{v}_n \\
&\lim(\boldsymbol{u}_n - \boldsymbol{v}_n) = \lim \boldsymbol{u}_n - \lim \boldsymbol{v}_n \\
&\lim(\boldsymbol{u}_n \cdot \boldsymbol{v}_n) = \lim \boldsymbol{u}_n \cdot \lim \boldsymbol{v}_n \\
&\lim \|\boldsymbol{u}_n\| = \|\lim \boldsymbol{u}_n\|.
\end{aligned}$$

- *Se a_n e $\boldsymbol{u}_n$ convergem, $a_n\boldsymbol{u}_n$ também converge e:*

$$\lim(a_n \boldsymbol{u}_n) = (\lim a_n)(\lim \boldsymbol{u}_n).$$

Outro conceito importante que pode generalizar-se naturalmente para sucessões em $\mathbb{R}^m$ é o de sucessão limitada: diz-se que a sucessão $\boldsymbol{u}_n$ é *limitada* sse existe um número real k tal que se tenha $\|\boldsymbol{u}_n\| \leq k$ para todo o $n \in \mathbb{N}$ (ou, o que é equivalente, se existe uma bola centrada na origem[6] que contenha todos os seus termos).

[6] Ou, como é fácil de ver, centrada em qualquer outro ponto $\boldsymbol{a} \in \mathbb{R}^m$.

Supondo $\boldsymbol{u}_n = (u_{n1}, \dots, u_{nm})$ e $j \in \{1, \dots, m\}$, da primeira das desigualdades:

$$|u_{nj}| \leq \|\boldsymbol{u}_n\| \leq |u_{n1}| + \cdots + |u_{nm}|$$

infere-se que, se $\boldsymbol{u}_n$ é limitada, qualquer das suas sucessões coordenadas é uma sucessão limitada (em $\mathbb{R}$); da segunda resulta que, sendo todas as sucessões coordenadas limitadas, $\boldsymbol{u}_n$ é limitada. Portanto:

Teorema 2.2. *Para que uma sucessão em $\mathbb{R}^n$ seja limitada é necessário e suficiente que o seja cada uma das suas sucessões coordenadas.*

Por exemplo, em $\mathbb{R}^3$, é limitada a sucessão:

$$\boldsymbol{u}_n = \left(\frac{\log n}{n}, \left(1 + \frac{1}{n}\right)^n, (-1)^n \right)$$

e não o é

$$\boldsymbol{v}_n = \left(e^{-n}, 2, -n\right).$$

Sabemos bem que, em $\mathbb{R}$, as sucessões convergentes são limitadas. Seja agora $\boldsymbol{u}_n$ uma sucessão convergente em $\mathbb{R}^m$. Pelo Teorema 2.1, todas as sucessões coordenadas de $\boldsymbol{u}_n$ são sucessões (reais) convergentes, e portanto limitadas; daqui, pelo Teorema 2.2, pode concluir-se que $\boldsymbol{u}_n$ é limitada. Assim, também em $\mathbb{R}^m$, *as sucessões convergentes são necessariamente limitadas.*

Um outro resultado importante que nos será necessário na sequência (em particular, no estudo de propriedades fundamentais das funções contínuas) é o que se exprime no seguinte:

Teorema 2.3 (Bolzano--Weierstrass). *Qualquer sucessão limitada (em $\mathbb{R}^m$) tem subsucessões convergentes.*

Demonstração. Para maior simplicidade e clareza, faremos a demonstração para o caso de sucessões em $\mathbb{R}^2$, sendo óbvio que a mesma ideia essencial permite demonstrar a proposição no caso geral (mesmo assim, poderá ser útil ler, antes da demonstração, o exemplo que se lhe segue).

Sendo $\boldsymbol{u}_n = (u_{n1}, u_{n2})$ uma sucessão limitada, serão também limitadas as sucessões reais u_{n1} e u_{n2} (Teorema 2.2). Nestas condições, o teorema de Bolzano--Weierstrass (estudado já para o caso de sucessões reais) permite extrair de u_{n1} uma subsucessão convergente,

$$u_{p_1 1} \; u_{p_2 1} \; \dots \; u_{p_n 1} \; \dots$$

Consideremos a subsucessão de $\boldsymbol{u}_n$:

$$(u_{p_1 1}, u_{p_1 2}) \; (u_{p_2 1}, u_{p_2 2}) \; \dots \; (u_{p_n 1}, u_{p_n 2}) \; \dots,$$

para a qual a 1ª sucessão coordenada é convergente e a 2ª é limitada (por ser subsucessão de u_{n2}). Novo recurso ao teorema de Bolzano--Weierstrass (caso real) permite extrair desta última sucessão numérica limitada uma subsucessão convergente:

$$u_{q_1 2}\ u_{q_2 2}\ \ldots\ u_{q_n 2}\ \ldots$$

Nestas condições, mostra o Teorema 2.1 que a subsucessão de $\boldsymbol{u}_n$:

$$(u_{q_1 1}, u_{q_1 2})\ (u_{q_2 1}, u_{q_2 2})\ \ldots\ (u_{q_n 1}, u_{q_n 2})\ \ldots$$

cujas sucessões coordenadas são ambas convergentes (a 2ª por construção, a 1ª por ser subsucessão de uma sucessão convergente) é necessariamente convergente, o que termina a demonstração. □

Exemplo: Para cada $n \in \mathbb{N}_1$, designemos por r_n o resto da divisão inteira de n por 3 ($r_1 = 1$, $r_2 = 2$, $r_3 = 0$, etc.) e consideremos a sucessão limitada (em $\mathbb{R}^2$):

$$\boldsymbol{u}_n = \left(r_n, (-1)^n + \frac{1}{n}\right).$$

Para obter uma subsucessão convergente de $\boldsymbol{u}_n$, pode começar-se por determinar uma subsucessão convergente de r_n, por exemplo, r_{3n}, que tem todos os termos nulos; ter-se-á então:

$$\boldsymbol{u}_{3n} = \left(0, (-1)^{3n} + \frac{1}{3n}\right).$$

A 2ª sucessão coordenada de $\boldsymbol{u}_{3n}$ não é convergente, mas pode extrair-se dela uma subsucessão convergente, por exemplo considerando apenas os valores de n para os quais o expoente de $(-1)^{3n}$ é par (e portanto múltiplo de 6, visto que já o era de 3). Obtém-se assim a subsucessão $\boldsymbol{u}_{6n}$ de $\boldsymbol{u}_n$:

$$\boldsymbol{u}_{6n} = \left(0, 1 + \frac{1}{6n}\right),$$

que é evidentemente convergente.

Trataremos agora de definir o conceito de sucessão de Cauchy, no quadro das sucessões de termos em $\mathbb{R}^m$. Naturalmente, diremos que uma tal sucessão, $\boldsymbol{u}_n$, é uma *sucessão de Cauchy* (ou uma *sucessão fundamental*) sse, qualquer que seja $\epsilon > 0$ existe p tal que, sempre que os inteiros positivos r e s sejam maiores do que p, se tenha:

$$\|\boldsymbol{u}_r - \boldsymbol{u}_s\| < \epsilon.$$

Supondo $\boldsymbol{u}_n = (u_{n1}, \ldots, u_{nm})$, poderemos deduzir (de modo idêntico ao que usámos já por duas vezes) das desigualdades:

$$|u_{rj} - u_{sj}| \leq \|\boldsymbol{u}_r - \boldsymbol{u}_s\| \leq |u_{r1} - u_{s1}| + \cdots + |u_{rm} - u_{sm}|,$$

que a sucessão $\boldsymbol{u}_n$ é fundamental sse o forem todas as suas sucessões coordenadas. Finalmente, tendo em conta, além deste resultado e do Teorema 2.1, o facto bem conhecido de que uma sucessão de termos reais é convergente sse é fundamental, conclui-se imediatamente que, para que uma sucessão em $\mathbb{R}^m$ seja convergente é necessário e suficiente que seja fundamental.

Finalizaremos este parágrafo com uma breve referência ao conceito de série de termos em $\mathbb{R}^m$. Diremos naturalmente que a série

$$\sum_{n=1}^{\infty} \boldsymbol{u}_n = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_n + \cdots$$

(com $\boldsymbol{u}_n \in \mathbb{R}^m$ qualquer que seja n) é *convergente* sse o for a sucessão $\boldsymbol{s}_n = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \cdots + \boldsymbol{u}_n$; em caso de convergência, chamaremos *soma da série* ao limite de $\boldsymbol{s}_n$. Diremos ainda que a série $\sum \boldsymbol{u}_n$ é *absolutamente convergente* sse for convergente a série de termos reais $\sum \|\boldsymbol{u}_n\|$.

Pondo, para cada $j \in \{1, 2, \ldots, m\}$, $p_j(\boldsymbol{u}_n) = u_{nj}$ (isto é, designando por u_{nj} a sucessão coordenada de ordem j da sucessão $\boldsymbol{u}_n$), reconhece-se imediatamente que $\sum \boldsymbol{u}_n$ é convergente e tem por soma $\boldsymbol{s} = (s_1, s_2, \ldots, s_m)$ sse cada uma das séries $\sum u_{nj}$ converge e tem por soma s_j; e também que $\sum \boldsymbol{u}_n$ é absolutamente convergente sse o forem todas as séries $\sum u_{nj}$.

Assim, o estudo de uma série de termos em $\mathbb{R}^m$ reduz-se trivialmente ao de séries de termos reais, sendo imediata a extensão ao novo quadro de resultados obtidos em estudos anteriores. Eis alguns exemplos, cuja justificação (a partir de resultados conhecidos relativos a séries de termos reais) constituirá um simples exercício:

- Se a série (de termos em $\mathbb{R}^m$) $\sum \boldsymbol{u}_n$ é convergente, $\boldsymbol{u}_n$ converge para $\mathbf{0}$.
- $\sum \boldsymbol{u}_n$ é convergente sse, qualquer que seja $\epsilon > 0$ existe p tal que, sempre que $s > r \geq p$, se tenha

$$\|\boldsymbol{u}_r + \boldsymbol{u}_{r+1} + \cdots + \boldsymbol{u}_s\| < \epsilon.$$

- É convergente qualquer série de termos em $\mathbb{R}^m$ que convirja absolutamente.
- A série $\sum \boldsymbol{u}_n$ é absolutamente convergente se existe uma série convergente de termos reais, $\sum a_n$, tal que (a partir de alguma ordem) se tenha $\|\boldsymbol{u}_n\| \leq a_n$.

2.3 Noções topológicas em $\mathbb{R}^m$

No estudo de diversos temas subsequentes — limites, continuidade, cálculo diferencial para funções de mais de uma variável real — intervirão significativamente certas características dos subconjuntos de $\mathbb{R}^m$ em que as funções consideradas se suporão definidas (recordemos por exemplo, que, para funções contínuas definidas num conjunto limitado $A \subset \mathbb{R}$, podia garantir-se a existência de máximo e mínimo se o conjunto A fosse fechado, não ficando assegurada, fora desta hipótese, a existência de qualquer extremo de f). Torna-se-nos, por isso, necessário estudar para subconjuntos de $\mathbb{R}^m$ algumas noções que costumam ser designadas por noções topológicas e que, como veremos, podem ser todas definidas a partir do conceito de bola.

Para facilitar a compreensão de algumas ideias essenciais começaremos por um exemplo muito simples, no espaço $\mathbb{R}^2$, que suporemos identificado com o plano do modo habitual. Designemos por K o subconjunto de $\mathbb{R}^2$ formado por todos os pares (x, y) tais que:

$$0 \leq x \leq 1 \quad \text{e} \quad 0 \leq y < 1.$$

Geometricamente, K é o conjunto dos pontos situados no quadrado (Figura 2.2), incluindo os pontos de todos os seus lados e vértices, com excepção dos que estão situados na recta de equação $y = 1$ (assim na figura, os pontos P e Q pertencem a K, os pontos R e S não pertencem).

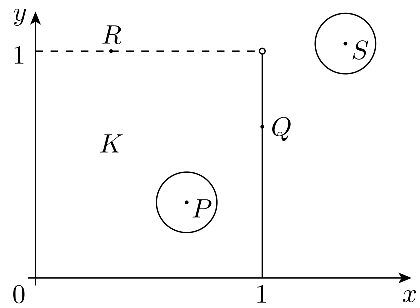

Figura 2.2

No caso do ponto P pode observar-se que, não só o próprio ponto pertence a K, como também pertencem a este conjunto todos os pontos do plano que estejam «suficientemente próximos» de P: mais precisamente, existe uma bola centrada em P tal que todos os pontos desta bola pertencem também ao conjunto K (para obter uma de tais bolas basta escolher para raio um número positivo ϵ inferior ou igual à menor das distâncias de P aos lados do quadrado). De acordo com a definição que introduziremos na sequência, poderemos dizer que o ponto P é interior ao conjunto K.

No caso do ponto S observa-se que, não só S não pertence a K, como também não pertencem ao mesmo conjunto todos os pontos do plano «suficientemente próximos» de S: existem bolas centradas em S que não contêm ponto algum do conjunto K (isto é, que estão contidas no complementar deste conjunto, em relação ao plano); diremos que o ponto S é exterior ao conjunto K.

A situação dos pontos Q e R é diferente de qualquer das anteriores; tanto para Q como para R (e embora o primeiro destes pontos pertença a K e o segundo não, o que não interessa para o efeito em vista) é impossível obter uma bola centrada no ponto considerado — Q ou R — e que esteja, ou contida no conjunto K (como no caso de P) ou contida no seu complementar (como no caso de S); o que se verifica é que qualquer bola centrada em Q ou em R, por menor que seja o seu raio, contém sempre pontos de K e pontos que não pertencem a este conjunto. Os pontos Q e R não são interiores nem exteriores ao conjunto K; diremos que são pontos fronteiros deste conjunto.

Poderemos dar agora, em termos gerais, as definições seguintes:

Seja X um subconjunto qualquer de $\mathbb{R}^m$ e $\boldsymbol{a}$ um elemento de $\mathbb{R}^m$. Diz-se que o ponto $\boldsymbol{a}$ é *interior* ao conjunto X sse existe $\epsilon > 0$ tal que $B_\epsilon(\boldsymbol{a}) \subset X$. Designando por $C(X)$ o complementar de X em $\mathbb{R}^m$ ($C(X) = \mathbb{R}^m \setminus X$), diz-se que $\boldsymbol{a}$ é *exterior* a X sse existe $\epsilon > 0$ tal que $B_\epsilon(\boldsymbol{a}) \subset C(X)$. Assim, dizer que $\boldsymbol{a}$ é exterior a X equivale a dizer que $\boldsymbol{a}$ é interior a $C(X)$.

Diz-se ainda que $\boldsymbol{a}$ é *ponto fronteiro* de X sse, qualquer que seja $\epsilon > 0$, $B_\epsilon(\boldsymbol{a})$ tem pelo menos um ponto de X e pelo menos um ponto de $C(X)$ (ou, o que é o mesmo, se $\boldsymbol{a}$ não é interior nem exterior ao conjunto X).

O conjunto formado por todos os pontos de $\mathbb{R}^m$ que são interiores a X chama-se *interior* do conjunto X, e designa-se por $\operatorname{int} X$ (ou $\overset{\circ}{X}$); definem-se de forma análoga o *exterior de* X ($\operatorname{ext} X$) e a *fronteira* de X ($\operatorname{front} X$ ou ∂X).

Assim, no exemplo do conjunto K há pouco considerado, o interior de K é o conjunto de todos os pontos (x, y) tais que:

$$0 < x < 1 \quad \text{e} \quad 0 < y < 1$$

e a fronteira é formada pelos pontos que pertencem a algum dos lados do quadrado (incluindo os vértices); o exterior de K é constituído por todos os restantes pontos do plano.

Outro exemplo, agora em $\mathbb{R}$ ($m = 1$): sendo $L = [0, 1[\cup \{2\}$ verifica-se imediatamente que o interior de L é o intervalo aberto $]0, 1[$ (para qualquer ponto a deste intervalo existe $\epsilon > 0$ tal que $B_\epsilon(a) =]a - \epsilon, a + \epsilon[\subset L$, e os únicos pontos de $\mathbb{R}$ que possuem esta propriedade são os do intervalo $]0, 1[$); a fronteira de L é o conjunto $\{0, 1, 2\}$ e o exterior é o complementar, em $\mathbb{R}$, do conjunto $[0, 1] \cup \{2\}$.

No caso de $\mathbb{R}^3$, as bolas devem, como sabemos, ser interpretadas como esferas abertas: por exemplo, para o subconjunto de $\mathbb{R}^3$ formado pelos ternos (x, y, z) tais que $z = 0$ — que geometricamente corresponde ao plano dos xy — reconhece-se facilmente que o interior é o conjunto vazio, a fronteira coincide com o próprio conjunto e o exterior é o seu complementar em $\mathbb{R}^3$.

No caso geral de um subconjunto X do espaço $\mathbb{R}^m$, ao qual nos referimos nas definições anteriores, tudo é análogo, salvo a possibilidade de interpretação geométrica, que não subsiste para $m > 3$.

Reconhece-se imediatamente que, qualquer que seja o conjunto $X \subset \mathbb{R}^m$, os três conjuntos $\operatorname{int} X$, $\operatorname{ext} X$ e $\operatorname{front} X$ (dos quais um, ou mesmo dois, podem ser vazios) têm por reunião o conjunto $\mathbb{R}^m$ e são disjuntos dois a dois.

Outra definição importante é a seguinte: chama-se *aderência* ou *fecho* do conjunto $X \subset \mathbb{R}^m$ à reunião do seu interior com a sua fronteira; a aderência de X

é usualmente designada pelo símbolo $\bar{X}$:

$$\bar{X} = \operatorname{int} X \cup \operatorname{front} X,$$

e coincide, portanto, com o complementar do exterior de X.

Aos elementos de $\bar{X}$ chama-se *pontos aderentes* ao conjunto X, sendo fácil reconhecer que, para que $\boldsymbol{a} \in \mathbb{R}^m$ seja aderente ao conjunto X é necessário e suficiente que qualquer bola centrada em $\boldsymbol{a}$ tenha pelo menos um ponto comum com o conjunto X ($B_\epsilon(\boldsymbol{a}) \cap X \neq \emptyset$, para todo o $\epsilon > 0$).

Uma outra caracterização dos pontos aderentes é facultada no seguinte:

Teorema 2.4. *Seja $X \subset \mathbb{R}^m$ e $\boldsymbol{a} \in \mathbb{R}^m$; $\boldsymbol{a}$ é aderente a X sse existe uma sucessão $\boldsymbol{u}_n$ de termos em X que converge para $\boldsymbol{a}$.*

Demonstração. Se existe uma sucessão em X convergente para $\boldsymbol{a}$, é óbvio que qualquer bola centrada em $\boldsymbol{a}$ contém pelo menos um ponto de X, isto é, que $\boldsymbol{a} \in \bar{X}$. Em sentido inverso, se $\boldsymbol{a} \in \bar{X}$, para todo o $\epsilon > 0$ tem-se $B_\epsilon(\boldsymbol{a}) \cap X \neq \emptyset$; escolhendo arbitrariamente um ponto $\boldsymbol{u}_n$ em $B_{\frac{1}{n}}(\boldsymbol{a}) \cap X$, para $n = 1, 2, \ldots$, obtém-se uma sucessão em X que converge para $\boldsymbol{a}$, visto que para todo o n se tem $\|\boldsymbol{u}_n - \boldsymbol{a}\| < 1/n$. □

A aderência de um conjunto X foi definida como reunião de dois conjuntos disjuntos: o interior de X e a fronteira de X. Há uma outra maneira, também significativa, de decompor $\bar{X}$ como reunião de dois conjuntos disjuntos: um deles é o conjunto dos pontos de acumulação do conjunto X — ou derivado de X, designado por X' — o outro o conjunto dos seus pontos isolados. Antes de dar as definições formais, recordemos o exemplo do subconjunto $L = [0, 1[\cup \{2\}$, em $\mathbb{R}$, cuja aderência é o conjunto

$$\bar{L} = [0,\ 1] \cup \{2\},$$

e observemos o seguinte: para o ponto 2, existe uma bola centrada neste ponto na qual ele é o único elemento do conjunto L (é o que se passa em qualquer «bola» $]2-\epsilon,\ 2+\epsilon[$, desde que seja $0 < \epsilon \leq 1$); para qualquer outro ponto $a \in L$, verifica-se que, para todo o $\epsilon > 0$, há elementos do conjunto L distintos de a, em $]a - \epsilon,\ a + \epsilon[$. De acordo com as definições subsequentes, poderemos dizer que o ponto 2 é um ponto isolado de L e que todos os pontos de $[0,\ 1]$ são pontos de acumulação do mesmo conjunto.

Em geral, sendo $X \subset \mathbb{R}^m$ e $\boldsymbol{a} \in \bar{X}$, diremos que $\boldsymbol{a}$ é um *ponto isolado* do conjunto X sse existe $\epsilon > 0$ tal que $B_\epsilon(\boldsymbol{a})$ não contém qualquer elemento de X distinto de $\boldsymbol{a}$ (é fácil ver que, nesta hipótese, se tem necessariamente $\boldsymbol{a} \in X$ pois, de contrário não seria $\boldsymbol{a} \in \bar{X}$); e diremos que $\boldsymbol{a}$ é *ponto de acumulação* de X no caso oposto, isto é, se qualquer bola centrada em $\boldsymbol{a}$ tem pelo menos um ponto de X distinto de $\boldsymbol{a}$ (claro que, neste caso, pode ser $\boldsymbol{a} \in X$ ou $\boldsymbol{a} \notin X$).

O conjunto dos pontos de acumulação de X é, por definição, o *derivado* X', do conjunto X. Reconhece-se facilmente que, para que $\boldsymbol{a} \in X'$, é necessário e suficiente que qualquer bola centrada em $\boldsymbol{a}$ contenha infinitos elementos do conjunto X (se nalguma de tais bolas houvesse apenas um número finito de elementos de X : $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$, designando por ϵ o mínimo das distâncias ao ponto $\boldsymbol{a}$ de cada um desses elementos — com exclusão do próprio ponto $\boldsymbol{a}$, se fosse um deles — logo se vê que $B_\epsilon(\boldsymbol{a})$ não conteria qualquer ponto de X distinto de $\boldsymbol{a}$).

Os pontos de acumulação de um conjunto podem caracterizar-se de modo análogo ao expresso no Teorema 2.4 para os pontos aderentes; enunciaremos essa caracterização no seguinte teorema, cuja demonstração, inteiramente análoga à do Teorema 2.4, poderá ficar como exercício:

Teorema 2.4'. *Seja $X \subset \mathbb{R}^m$ e $\boldsymbol{a} \in \mathbb{R}^m$; $\boldsymbol{a}$ é ponto de acumulação de X sse existe uma sucessão em X, de termos distintos de $\boldsymbol{a}$, que converge para $\boldsymbol{a}$.*

É fácil verificar que, em $\mathbb{R}^m$, qualquer ponto interior de um conjunto é ponto de acumulação do mesmo conjunto (para o reconhecer, basta recordar que qualquer bola é um conjunto infinito): $\operatorname{int} X \subset X'$, para todo o $X \subset \mathbb{R}^m$; daí resulta imediatamente (dado que a reunião de X' com o conjunto dos pontos isolados de X coincide com a reunião do interior com a fronteira do mesmo conjunto) que qualquer ponto isolado de X pertence à fronteira de X.

É óbvio que qualquer ponto interior a um conjunto $X \subset \mathbb{R}^m$ pertence necessariamente ao conjunto X e também que qualquer ponto do conjunto X não pode ser exterior a X, pertencendo, portanto, a $\bar{X}$. Assim, qualquer que seja $X \subset \mathbb{R}^m$, verificam-se necessariamente as relações:

$$\operatorname{int} X \subset X \subset \bar{X}.$$

Pode, em particular, suceder que um conjunto X coincida com o seu interior (isto é, que nenhum dos seus pontos fronteiros lhe pertença: $\operatorname{front} X \subset C(X)$); ou que coincida com a sua aderência (o que se passa se pertencerem a X todos os seus pontos fronteiros: $\operatorname{front} X \subset X$). No primeiro caso, diz-se que X é um *conjunto aberto*, no segundo que é um *conjunto fechado.* Os conjuntos abertos e os conjuntos fechados são, portanto, respectivamente caracterizados pelas igualdades:

$$\operatorname{int} X = X \quad \text{e} \quad \bar{X} = X.$$

As noções de conjunto aberto e conjunto fechado têm grande interesse, como veremos na sequência.

É fácil ver que, em $\mathbb{R}^m$, qualquer bola aberta é um conjunto aberto e qualquer bola fechada é um conjunto fechado.

Para verificar que $B_\epsilon(\boldsymbol{a})$ é um conjunto aberto (qualquer que seja o ponto $\boldsymbol{a} \in \mathbb{R}^m$ e o número positivo ϵ) basta reconhecer que, se $\boldsymbol{b}$ for um ponto arbitrário de $B_\epsilon(\boldsymbol{a})$, existe uma bola centrada em $\boldsymbol{b}$, $B_\delta(\boldsymbol{b})$, contida em $B_\epsilon(\boldsymbol{a})$; ora para que tal se verifique basta escolher δ por forma que se tenha $0 < \delta \leq \epsilon - \|\boldsymbol{b} - \boldsymbol{a}\|$ (o que é possível visto que, por ser $\boldsymbol{b} \in B_\epsilon(\boldsymbol{a})$, se tem $\|\boldsymbol{b}-\boldsymbol{a}\| < \epsilon$). Na realidade, escolhido δ desta forma, ter-se-á, para qualquer $x \in B_\delta(\boldsymbol{b})$,

$$\begin{aligned}
\|\boldsymbol{x}-\boldsymbol{a}\| &= \|(\boldsymbol{x}-\boldsymbol{b})+(\boldsymbol{b}-\boldsymbol{a})\| \\
&\leq \|\boldsymbol{x}-\boldsymbol{b}\| + \|\boldsymbol{b}-\boldsymbol{a}\| \\
&< \delta + \|\boldsymbol{b}-\boldsymbol{a}\| \\
&\leq \epsilon,
\end{aligned}$$

o que mostra que $x \in B_\epsilon(\boldsymbol{a})$ e portanto que $B_\delta(\boldsymbol{b}) \subset B_\epsilon(\boldsymbol{a})$.

Por outro lado, para reconhecer que a bola fechada de centro $\boldsymbol{a}$ e raio ϵ — que, de momento, designaremos por $B^*_\epsilon(\boldsymbol{a})$ — é um conjunto fechado, será suficiente verificar que qualquer ponto $\boldsymbol{c}$ que não pertença a essa bola não lhe pode ser aderente (e ser-lhe-á portanto exterior). Ora se $\boldsymbol{c} \notin B^*_\epsilon(\boldsymbol{a})$, isto é se $\|\boldsymbol{c}-\boldsymbol{a}\| > \epsilon$, escolhido λ tal que $0 < \lambda < \|\boldsymbol{c}-\boldsymbol{a}\| - \epsilon$, ter-se-á $B^*_\epsilon(\boldsymbol{a}) \cap B_\lambda(\boldsymbol{c}) = \emptyset$ visto que, se existisse um ponto $\boldsymbol{x} \in B^*_\epsilon(\boldsymbol{a}) \cap B_\lambda(\boldsymbol{c})$ deveria ter-se:

$$\|\boldsymbol{c}-\boldsymbol{a}\| \leq \|\boldsymbol{c}-\boldsymbol{x}\| + \|\boldsymbol{x}-\boldsymbol{a}\| < \lambda + \epsilon < \|\boldsymbol{c}-\boldsymbol{a}\|$$

o que é absurdo. Pode assim concluir-se que $\boldsymbol{c}$ é exterior a $B^*_\epsilon(\boldsymbol{a})$ e portanto que este conjunto é fechado.

Exprimem-se no teorema seguinte algumas propriedades importantes da noção de conjunto aberto.

Teorema 2.5. *i) A reunião de qualquer família (finita ou infinita) de conjuntos abertos é um conjunto aberto.*

ii) A intersecção de qualquer família finita de conjuntos abertos é um conjunto aberto.

Demonstração. i) Seja $\{A_i\}_{i\in I}$ uma família qualquer de conjuntos abertos, $A = \bigcup_{i\in I} A_i$ a sua reunião. Por definição de reunião, se x é um ponto qualquer de A poderá escolher-se um índice $j \in I$ tal que $\boldsymbol{x} \in A_j$; como A_j é aberto, por hipótese, existirá $\epsilon > 0$ tal que $B_\epsilon(\boldsymbol{x}) \subset A_j$. Segue-se que $B_\epsilon(x) \subset A$ (visto que $A_j \subset A$) o que mostra que $\boldsymbol{x}$ é interior a A, e portanto que A é aberto.

ii) Seja $\{A_1, A_2, \ldots, A_n\}$ uma família finita de conjuntos abertos e seja agora $A = A_1 \cap A_2 \cap \ldots \cap A_n$. Se for $A = \emptyset$ é claro que A será aberto (visto que $\operatorname{int}\emptyset = \emptyset$). De contrário, sendo x um ponto qualquer do conjunto A (que pertencerá portanto a cada um dos conjuntos abertos $A_1, \ldots, A_n$) existirão necessariamente números positivos $\epsilon_1, \ldots, \epsilon_n$ tais que

$$B_{\epsilon_1}(\boldsymbol{x}) \subset A_1, \ldots, B_{\epsilon_n}(\boldsymbol{x}) \subset A_n.$$

Se for então $\epsilon = \min\{\epsilon_1, \ldots, \epsilon_n\}$ ter-se-á também

$$B_\epsilon(\boldsymbol{x}) \subset A_1, \ldots, B_\epsilon(\boldsymbol{x}) \subset A_n$$

e portanto

$$B_\epsilon(\boldsymbol{x}) \subset A_1 \cap \ldots \cap A_n = A,$$

o que prova que A é aberto. □

Convém notar que a intersecção de uma família infinita de conjuntos abertos pode não ser um conjunto aberto; por exemplo, sendo $\boldsymbol{a} \in \mathbb{R}^m$, a intersecção de todas as bolas $B_{\frac{1}{n}}(\boldsymbol{a})$ — com $n = 1, 2, \ldots$ — é o conjunto singular $\{\boldsymbol{a}\}$, que não é aberto.

Pode ainda observar-se que qualquer conjunto aberto do espaço $\mathbb{R}^m$ é reunião de uma família (finita ou infinita) de bolas desse espaço. Com efeito, sendo $A \subset \mathbb{R}^m$ um conjunto aberto, para cada $\boldsymbol{x} \in A$ existirá $\epsilon_x > 0$ tal que $B_{\epsilon_x}(\boldsymbol{x}) \subset A$; e logo se reconhece que se terá então $A = \bigcup_{\boldsymbol{x} \in A} B_{\epsilon_x}(\boldsymbol{x})$.

Como é evidente, um conjunto pode não ser aberto nem fechado (é o que se passa, por exemplo, em $\mathbb{R}$ com um intervalo da forma $[\alpha, \beta[$, com $\alpha, \beta \in \mathbb{R}$ e $\alpha < \beta$; em $\mathbb{R}^2$ com o conjunto K, considerado no primeiro exemplo referido neste parágrafo, etc.). Convém observar, porém, que existem conjuntos que são abertos e fechados; é claro que, para que o conjunto $X \subset \mathbb{R}^m$ seja aberto e fechado é necessário e suficiente que se verifique a igualdade

$$\operatorname{int} X = \bar{X},$$

isto é que não exista qualquer ponto fronteiro do conjunto X (pode provar-se, aliás, que os únicos subconjuntos de $\mathbb{R}^m$ que têm fronteira vazia e que, portanto, são simultaneamente abertos e fechados, são o próprio conjunto $\mathbb{R}^m$ e o conjunto vazio).

Antes de enunciar (no Teorema 2.7) algumas propriedades da noção de conjunto fechado (correspondentes às referidas no Teorema 2.5 para os conjuntos abertos) convém salientar que — como decorre imediatamente da própria definição de ponto fronteiro — para qualquer conjunto $X \subset \mathbb{R}^m$ se verifica a igualdade: $\operatorname{front} X = \operatorname{front} C(X)$.

Daqui decorre trivialmente o:

Teorema 2.6. *Um conjunto é aberto sse o seu complementar é fechado (e portanto é fechado sse o seu complementar é aberto).*

Demonstração. Dizer que X é aberto equivale a dizer que $\operatorname{front} X \subset C(X)$ ou, o que é o mesmo, $\operatorname{front} C(X) \subset C(X)$, o que significa que $C(X)$ é fechado. □

Pode agora enunciar-se o

Teorema 2.7. *i) A intersecção de uma família (finita ou infinita) de conjuntos fechados é um conjunto fechado.*

ii) A reunião de qualquer família finita de conjuntos fechados é um conjunto fechado.

Demonstração. Daremos apenas uma justificação de i), dado que ii) se prova de forma análoga. Sendo $\{F_i\}_{i \in I}$ uma família qualquer de conjuntos fechados, ponha-se, para cada $i \in I$, $A_i = C(F_i)$. Os conjuntos A_i são abertos e portanto é também aberta a sua reunião $A = \bigcup_{i \in I} A_i$. Segue-se que é fechado o conjunto

$$C(A) = C\left(\bigcup_{i \in I} A_i\right) = \bigcap_{i \in I} C(A_i) = \bigcap_{i \in I} F_i,$$

como se pretendia verificar. $\square$

Os conjuntos fechados (e portanto também os conjuntos abertos) podem ser caracterizados recorrendo ao conceito de limite de uma sucessão; com efeito:

Teorema 2.8. *Um conjunto $X \subset \mathbb{R}^m$ é fechado sse, para toda a sucessão convergente $\boldsymbol{u}_n$, de termos em X, se tem $\lim \boldsymbol{u}_n \in X$.*

Demonstração. Sendo X fechado e $\boldsymbol{u}_n$ uma sucessão de termos em X convergente para um ponto $\boldsymbol{a}$ (em $\mathbb{R}^m$), mostra o Teorema 2.4 que $\boldsymbol{a}$ é aderente a X e portanto, por ser $X = \bar{X}$, que $\boldsymbol{a} \in X$.

Se X não é fechado existe pelo menos um ponto aderente a X e não pertencente a este conjunto; e basta observar que (também pelo Teorema 2.4) esse ponto é o limite de alguma sucessão de termos em X para completar a demonstração. $\square$

> É também fácil ver que um conjunto X é fechado sse contém o seu derivado; na realidade, $X' \subset \bar{X}$ e portanto, se X é fechado, $X = \bar{X} \supset X'$; reciprocamente, se se tiver $X' \subset X$ ter-se-á também $\bar{X} \subset X$ (visto que, como já observámos, os pontos isolados de X pertencem necessariamente a este conjunto) e portanto $\bar{X} = X$.

Introduziremos agora um outro conceito importante, o de conjunto limitado[7]: diz-se que um conjunto $X \subset \mathbb{R}^n$ é *limitado* sse existe um real k tal que, para todo o $\boldsymbol{x} \in X$, se tem $\|\boldsymbol{x}\| \leq k$ (pode também dizer-se, equivalentemente, que o conjunto X é limitado sse existe uma bola que o contém).

Em $\mathbb{R}$, os conjuntos limitados nos termos desta definição são precisamente os conjuntos majorados e minorados (aos quais chamávamos já conjuntos limitados). Em $\mathbb{R}^m$, é limitado qualquer conjunto finito, qualquer bola, etc. Não são limitados em $\mathbb{R}^m$ (com $m > 1$), por exemplo, o conjunto dos $\boldsymbol{x} = (x_1, \ldots, x_m)$ tais que $x_1 = 0$ ou, sendo $\boldsymbol{a} \in \mathbb{R}^m$, o conjunto dos $\boldsymbol{x}$ tais que $\boldsymbol{a} \cdot \boldsymbol{x} = 0$; não é também limitado em $\mathbb{R}^m$ o complementar de qualquer conjunto limitado.

Um resultado importante na sequência — por exemplo, para a demonstração de algumas propriedades fundamentais das funções contínuas — é o seguinte:

Teorema 2.9. *Um conjunto $X \subset \mathbb{R}^m$ é limitado e fechado sse qualquer sucessão de termos em X tem uma subsucessão convergente para um ponto de X.*

Demonstração. Suponha-se X limitado e fechado e seja $\boldsymbol{u}_n$ uma sucessão qualquer de termos em X. $\boldsymbol{u}_n$ é limitada (porque $\boldsymbol{u}_n \in X$ para todo o $n \in \mathbb{N}_1$ e X é limitado) e, portanto, pelo Teorema 2.3, pode extrair-se de $\boldsymbol{u}_n$ uma subsucessão convergente, $\boldsymbol{u}_{p_n}$; como X é fechado, decorre do Teorema 2.8 que $\lim \boldsymbol{u}_{p_n} \in X$.

[7]Veremos em estudos mais avançados que a noção de conjunto limitado, aqui definida a partir dos mesmos conceitos (o de bola ou o de norma) utilizados para definir as outras noções introduzidas neste parágrafo, não é, no entanto, propriamente uma «noção topológica», no sentido atribuído a esta expressão em certos contextos mais gerais.

Suponha-se agora que X não é limitado ou não é fechado (podendo evidentemente não ser uma coisa nem outra). Se X não for limitado para todo o $n \in \mathbb{N}_1$ poderá escolher-se $\boldsymbol{u}_n \in X$ tal que $\|\boldsymbol{u}_n\| > n$; obter-se-á assim uma sucessão que não terá qualquer subsucessão limitada nem, portanto, qualquer subsucessão convergente. Se X não for fechado, escolhido um ponto $\boldsymbol{a} \in \bar{X} \setminus X$, existirá (pelo Teorema 2.4) uma sucessão $\boldsymbol{u}_n$ de termos em X convergente para $\boldsymbol{a}$. Tal sucessão não poderá ter qualquer subsucessão convergente para um ponto de X (visto que todas as suas subsucessões convergem para $\boldsymbol{a} \notin X$). $\square$

Na sequência, diremos que um conjunto $X \subset \mathbb{R}^m$ é *compacto* sse for limitado e fechado.

> Diz-se por vezes que um conjunto X é sequencialmente compacto sse é verificada a propriedade seguinte: qualquer sucessão de termos em X tem uma subsucessão que converge para um ponto de X. Assim, poderia exprimir-se o enunciado do precedente Teorema 2.9 dizendo que, em $\mathbb{R}^m$ um conjunto é compacto sse for sequencialmente compacto.
>
> Teremos oportunidade de ver posteriormente que algumas das propriedades mais importantes das funções contínuas num conjunto compacto de $\mathbb{R}$ — tais como a continuidade uniforme (teorema de Heine--Cantor), a existência de máximo e mínimo (teorema de Weierstrass) — se generalizam facilmente ao caso de funções reais, contínuas num conjunto compacto de $\mathbb{R}^m$ (naturalmente, haverá que definir de forma adequada a noção de continuidade para tais funções).
>
> Uma outra propriedade importante — que nos parece útil referir, embora não nos vá ser necessária na sequência — é a seguinte: qualquer conjunto compacto e infinito tem pelo menos um ponto de acumulação (decerto pertencente ao conjunto, por este ser fechado). Mais geralmente, pode provar-se que qualquer conjunto infinito e limitado tem pelo menos um ponto de acumulação (pertencente ou não ao conjunto). Este resultado, que pode deduzir-se sem dificuldade do Teorema 2.3, é também correntemente designado por «teorema de Bolzano--Weierstrass».

Para finalizar este parágrafo, introduziremos outra noção topológica importante (que, em particular, nos permitirá generalizar para funções contínuas de mais de uma variável real o «teorema do valor intermédio»); trata-se da noção de conjunto conexo. A ideia intuitiva de conjunto conexo é a de conjunto formado por «uma só peça» (e não por diversas «peças separadas»). Por exemplo, poderá ver-se que (em $\mathbb{R}$) o intervalo $[0, 1]$ é um conjunto conexo, mas já não o é o seu complementar. No plano, um círculo ou uma circunferência são conjuntos conexos, tal como o complementar de um círculo; não é conexo o complementar de uma circunferência, formado por «duas peças», «separadas» pela própria circunferência.

Antes de darmos uma definição precisa de conjunto conexo, convém introduzir a seguinte: sendo A e B dois subconjuntos *não vazios* de $\mathbb{R}^m$, diremos que A e B são *separados* sse cada um destes conjuntos não contém qualquer ponto que seja aderente ao outro; noutros termos: os conjuntos A e B (tais que $A \neq \emptyset$ e $B \neq \emptyset$) são separados sse forem verificadas as duas igualdades:

$$A \cap \bar{B} = \emptyset, \quad B \cap \bar{A} = \emptyset.$$

É óbvio que dois conjuntos separados são necessariamente disjuntos (de $B \subset \bar{B}$ resulta $A \cap B \subset A \cap \bar{B} = \emptyset$). Mas é fácil ver que a recíproca é falsa. Por exemplo, em $\mathbb{R}$, os conjuntos disjuntos $]-1,\, 0[$ e $[0,\, 1]$ não são separados (o ponto 0, aderente ao primeiro, pertence ao segundo); em $\mathbb{R}^2$, o gráfico da função $\operatorname{sen} 1/x$ e o conjunto formado apenas pelo ponto $(0, a)$ são disjuntos (qualquer que seja $a \in \mathbb{R}$), mas só são separados se for $|a| > 1$.

Seja agora X um subconjunto de $\mathbb{R}^m$. Diz-se que X é um conjunto *desconexo* sse existirem dois conjuntos separados A e B tais que

$$X = A \cup B.$$

Na hipótese contrária, isto é, no caso de não existirem dois conjuntos separados A e B verificando a igualdade precedente, diz-se que X é um conjunto *conexo*.

São exemplos triviais de conjuntos conexos, em $\mathbb{R}^m$, o vazio e qualquer conjunto formado por um só ponto; não é conexo qualquer conjunto finito X, com mais de um ponto (se A for uma parte própria de X — isto é, uma parte de X não vazia e distinta de X — e $B = X \setminus A$ o complementar de A em X, vê-se imediatamente que A e B são conjuntos separados).

No caso de $\mathbb{R}$ $(m = 1)$, o conjunto dos números racionais, $\mathbb{Q}$, é um conjunto desconexo: com efeito, sendo a um irracional qualquer, tem-se:

$$\mathbb{Q} = (\mathbb{Q} \cap \,]-\infty,\, a[) \cup (\mathbb{Q} \cap \,]a,\, +\infty[)$$

e é fácil ver que os conjuntos $\mathbb{Q} \cap \,]-\infty,\, a[$ e $\mathbb{Q} \cap \,]a,\, +\infty[$ são separados.

É útil observar que esta mesma ideia permite reconhecer que, em $\mathbb{R}$, qualquer conjunto conexo X verifica necessariamente a condição seguinte: se pertencerem ao conjunto X dois números reais a e b — com $a < b$ — pertencerão também a esse conjunto todos os reais compreendidos entre a e b, isto é, ter-se-á:

$$[a,\, b] \subset X$$

(tal como no exemplo precedente, basta observar que, se algum ponto c de $]a,\, b[$ não pertencesse a X, este conjunto seria a reunião dos conjuntos separados $X \cap \,]-\infty,\, c[$ e $X \cap \,]c,\, +\infty[$).

Ora é fácil mostrar (e poderá ficar como exercício) que os únicos subconjuntos de $\mathbb{R}$ que verificam a condição indicada são os intervalos.

Pode assim concluir-se que, em $\mathbb{R}$, qualquer conjunto conexo é um intervalo.

Em sentido inverso — e embora não nos seja indispensável na sequência — provaremos agora que qualquer intervalo de $\mathbb{R}$ é um conjunto conexo, o que nos permite enunciar o

Teorema 2.10. *Em $\mathbb{R}$, os conjuntos conexos são precisamente os intervalos.*

Demonstração. Atendendo ao que vimos anteriormente, a demonstração poderá considerar-se terminada se mostrarmos que, sendo I um intervalo qualquer de $\mathbb{R}$, a hipótese de existirem conjuntos separados A e B tais que

$$I = A \cup B$$

conduz necessariamente a uma contradição.

Admitamos então essa hipótese e escolhamos arbitrariamente um ponto $x \in A$ e um ponto $z \in B$; como A e B são disjuntos, ter-se-á necessariamente $x < z$ ou $x > z$. Vamos supor que é $x < z$ (de contrário, bastaria trocar as designações dos conjuntos A e B).

Como I é um intervalo, ter-se-á $[x,\, z] \subset I$, pertencendo então cada ponto do intervalo $[x,\, z]$ a A ou a B (e apenas a um destes conjuntos).

Designemos agora por y o supremo do conjunto $[x,\, z] \cap A$. É óbvio que $y \in [x,\, z]$ (devendo portanto ter-se $y \in A$ ou $y \in B$).

Observando que, como facilmente se reconhece, o supremo de um conjunto é sempre um ponto aderente a esse conjunto, pode inferir-se que y é aderente a $[x,\, z] \cap A$, e portanto também a A (visto que $[x,\, z] \cap A$ é um subconjunto de A). Mas, devendo ter-se $\bar{A} \cap B = \emptyset$, o facto de ser $y \in \bar{A}$ mostra que $y \notin B$ e que, portanto, $y \in A$. Pode então deduzir-se que $y \neq z$ (visto que $z \in B$) e também que o intervalo $]y,\, z]$ não contém qualquer elemento do conjunto A (de contrário não seria y o supremo de $[x,\, z] \cap A$), devendo portanto ter-se $]y,\, z] \subset B$. Nestas condições, porém, y seria aderente ao conjunto B e ter-se-ia $A \cap \bar{B} \neq \emptyset$, em contradição com a hipótese de A e B serem conjuntos separados. $\square$

Capítulo 3

Continuidade e limite

3.1 Continuidade

A definição de continuidade para funções escalares ou vectoriais de variável vectorial é, como vamos ver, uma generalização natural da definição correspondente para funções reais de variável real. A ideia intuitiva essencial continua a ser a seguinte: dizer que f é contínua num ponto a equivale a dizer que todos os valores assumidos por f em pontos «próximos» de a estão «próximos» de $f(a)$ ou, um pouco melhor, que poderá garantir-se que $f(x)$ está «tão próximo quanto se queira» de $f(a)$ desde que se considerem apenas valores de x (pertencentes ao domínio de f e) «suficientemente próximos» de a.

Consideremos em primeiro lugar o caso de uma função real de n variáveis reais ($n \in \mathbb{N}_1$). Sendo $f : D \to \mathbb{R}$, com $D \subset \mathbb{R}^n$ e sendo $\boldsymbol{a}$ um ponto de D, diz-se que f é contínua no ponto $\boldsymbol{a}$ sse, qualquer que seja a vizinhança de $f(\boldsymbol{a})$ — isto é, qualquer que seja o intervalo $]f(\boldsymbol{a}) - \delta,\ f(\boldsymbol{a}) + \delta[$, com $\delta > 0$ — existe uma bola (de $\mathbb{R}^n$) centrada em $\boldsymbol{a}$, $B_\epsilon(\boldsymbol{a})$, tal que para todo o $x \in B_\epsilon(\boldsymbol{a}) \cap D$ se tem $f(\boldsymbol{x}) \in\ ]f(\boldsymbol{a}) - \delta,\ f(\boldsymbol{a}) + \delta[$. Pode também dizer-se, de forma equivalente, que f é contínua em $\boldsymbol{a}$ sse para todo o $\delta > 0$ existe $\epsilon > 0$ tal que, se $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$, então $|f(\boldsymbol{x}) - f(\boldsymbol{a})| < \delta$.

Como primeiro exemplo, consideremos a função $f : \mathbb{R}^2 \to \mathbb{R}$ referida no parágrafo 1.2, ex. 4.a). Mudando as notações, f pode definir-se pela forma seguinte:

$$f(x_1, x_2) = \begin{cases} 0 & \text{se } x_1 \text{ e } x_2 \text{ são inteiros} \\ 1 & \text{se } x_1 \text{ ou } x_2 \text{ não são inteiros.} \end{cases}$$

É fácil ver que f é contínua nos pontos em que toma o valor 1 e não o é naqueles em que toma o valor 0. Para tal, observe-se primeiramente que estes últimos pontos são os vértices de uma «quadrícula» (formada pelas rectas verticais com equações da forma $x = k$, com $k \in \mathbb{Z}$, e pelas horizontais de equação $y = \ell$, $\ell \in \mathbb{Z}$), sendo evidente que, em qualquer bola centrada num desses vértices, há sempre pontos que não são vértices da quadrícula, nos quais f toma o valor 1. Assim, sendo $\boldsymbol{a} = (a_1, a_2)$ um ponto com ambas as coordenadas inteiras, tem-se por um lado $f(\boldsymbol{a}) = 0$, por outro sabe-se que qualquer bola centrada em $\boldsymbol{a}$ contém pontos $\boldsymbol{x}$ tais que $f(\boldsymbol{x}) = 1$. Pode, portanto, concluir-se que, se for δ um número positivo ≤ 1, não existirá $\epsilon > 0$ tal que $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon \Rightarrow |f(\boldsymbol{x}) - f(\boldsymbol{a})| < \delta$, o que mostra que f não é contínua em $\boldsymbol{a}$.

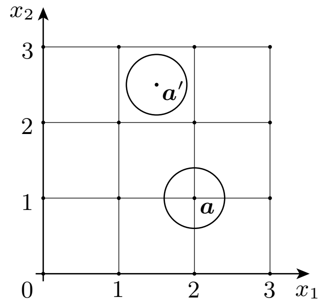

Figura 3.1

Sendo agora $\boldsymbol{a}' = (a'_1, a'_2)$ um ponto cujas coordenadas não sejam ambas números inteiros, vê-se facilmente que pode determinar-se $\epsilon > 0$ por forma que a bola $B_\epsilon(\boldsymbol{a}')$ não contenha qualquer vértice da quadrícula; ter-se-á então $f(\boldsymbol{x}) = 1$ para todo o $\boldsymbol{x} \in B_\epsilon(\boldsymbol{a}')$ e portanto, sendo $\delta > 0$ arbitrário, ter-se-á também $|f(\boldsymbol{x}) - f(\boldsymbol{a}')| < \delta$ sempre que seja $\|\boldsymbol{x} - \boldsymbol{a}'\| < \epsilon$. Pode assim concluir-se que f é contínua em $\boldsymbol{a}'$ (recorde-se no entanto que o facto, verificado neste último caso, de ter sido possível determinar «ϵ» independentemente do valor de «δ» é absolutamente excepcional; em geral, sendo f contínua em $\boldsymbol{a}$, é possível determinar um ϵ para cada δ, mas não um ϵ que convenha simultaneamente para todos os valores positivos de δ).

Como segundo exemplo, provaremos que a função $g : \mathbb{R}^n \to \mathbb{R}$ definida por $g(\boldsymbol{x}) = \|\boldsymbol{x}\|$ é contínua em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$. Para tal comecemos por observar que, das igualdades:

$$\boldsymbol{x} = \boldsymbol{a} + (\boldsymbol{x} - \boldsymbol{a}) \quad \text{e} \quad \boldsymbol{a} = \boldsymbol{x} + (\boldsymbol{a} - \boldsymbol{x})$$

se deduz, por propriedades conhecidas da norma:

$$\|\boldsymbol{x}\| \leq \|\boldsymbol{a}\| + \|\boldsymbol{x} - \boldsymbol{a}\| \quad \text{e} \quad \|\boldsymbol{a}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{a} - \boldsymbol{x}\| = \|\boldsymbol{x}\| + \|\boldsymbol{x} - \boldsymbol{a}\|,$$

e portanto também

$$\|\boldsymbol{x}\| - \|\boldsymbol{a}\| \leq \|\boldsymbol{x} - \boldsymbol{a}\| \quad \text{e} \quad \|\boldsymbol{a}\| - \|\boldsymbol{x}\| \leq \|\boldsymbol{x} - \boldsymbol{a}\|,$$

relações que evidenciam que, para quaisquer vectores $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{R}^n$, se tem:

$$\big|\|\boldsymbol{x}\| - \|\boldsymbol{a}\|\big| \leq \|\boldsymbol{x} - \boldsymbol{a}\|$$

ou

$$|g(\boldsymbol{x}) - g(\boldsymbol{a})| \leq \|\boldsymbol{x} - \boldsymbol{a}\|.$$

Assim, dado $\delta > 0$ bastará tomar $\epsilon = \delta$ para que se tenha $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon \Rightarrow |g(\boldsymbol{x}) - g(\boldsymbol{a})| < \delta$, o que prova o que se pretendia.

Antes de passarmos ao estudo da continuidade no quadro mais geral das funções vectoriais convém fazer algumas observações.

Em primeiro lugar, consideremos um conjunto qualquer D (na sequência ter-se-á quase sempre $D \subset \mathbb{R}^n$ mas por agora não há necessidade de supô-lo) e uma função f definida em D e com valores em $\mathbb{R}^m$. Para cada $x \in D$ o vector $f(x) \in \mathbb{R}^m$ terá m coordenadas (variáveis, em geral, quando x variar em D) que designaremos por $f_1(x), f_2(x), \ldots, f_m(x)$. Assim, a função vectorial f determina m funções escalares definidas em D, $f_1, f_2, \ldots, f_m$, às quais chamaremos naturalmente funções coordenadas de f.

No caso particular de D ser um subconjunto de $\mathbb{R}^n$, cada vector $\boldsymbol{x} \in D$ é, por sua vez, uma sequência $\boldsymbol{x} = (x_1, \ldots, x_n)$, e uma igualdade da forma:

$$\boldsymbol{y} = f(\boldsymbol{x}),$$

com $\boldsymbol{x} \in D$ e $\boldsymbol{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$, poderá ser traduzida por um sistema de m igualdades:

$$\begin{aligned} y_1 &= f_1(x_1, \ldots, x_n) \\ y_2 &= f_2(x_1, \ldots, x_n) \\ &\cdots \\ y_m &= f_m(x_1, \ldots, x_n). \end{aligned}$$

É desta forma (em termos de coordenadas), que muitas vezes são explicitadas as funções vectoriais utilizadas nas aplicações.

Um exemplo particularmente importante neste contexto é o das aplicações lineares de $\mathbb{R}^n$ em $\mathbb{R}^m$. Recorde-se que uma aplicação $f : \mathbb{R}^n \to \mathbb{R}^m$ se diz *linear* sse, quaisquer que sejam os vectores $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ e o escalar α, se tem:

$$f(\boldsymbol{u} + \boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v}) \quad \text{e} \quad f(\alpha \boldsymbol{u}) = \alpha f(\boldsymbol{u}).$$

Convencionemos designar por $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ os vectores da base canónica de $\mathbb{R}^n$, por $\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_m$ os vectores da base canónica de $\mathbb{R}^m$ e ainda — sendo $f : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação linear — por a_{ij} a coordenada de ordem i do vector $f(\boldsymbol{e}_j)$ (para $i \in \{1, 2, \ldots, m\}$ e $j \in \{1, 2, \ldots, n\}$). Dado um vector qualquer $\boldsymbol{x} = (x_1, \ldots, x_n)$ de $\mathbb{R}^n$ e sendo $\boldsymbol{y} = (y_1, \ldots, y_m)$ o valor de f em $\boldsymbol{x}$, deduz-se imediatamente da definição de aplicação linear que deverá ter-se:

$$\boldsymbol{y} = f(\boldsymbol{x}) = f\left(\sum_{j=1}^{n} x_j \boldsymbol{e}_j\right) = \sum_{j=1}^{n} f(x_j \boldsymbol{e}_j) = \sum_{j=1}^{n} x_j f(\boldsymbol{e}_j),$$

donde, atendendo a que

$$f(\boldsymbol{e}_j) = \sum_{i=1}^{m} a_{ij} \boldsymbol{e}'_i$$

resulta:

$$\boldsymbol{y} = \sum_{j=1}^{n} \sum_{i=1}^{m} x_j a_{ij} \boldsymbol{e}'_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \boldsymbol{e}'_i.$$

Como, por outro lado, se verifica também a igualdade:

$$\boldsymbol{y} = \sum_{i=1}^{m} y_i \boldsymbol{e}'_i,$$

a unicidade da expressão de um vector qualquer de $\mathbb{R}^m$ como combinação linear dos vectores de uma base (mencionada em 2.1, quando recordámos a definição de base de um espaço vectorial real) permite deduzir que deverá ter-se, para $i = 1, 2, \dots, m$:

$$y_i = \sum_{j=1}^{n} a_{ij} x_j.$$

Assim, no caso de $f : \mathbb{R}^n \to \mathbb{R}^m$ ser uma aplicação linear, à igualdade $\boldsymbol{y} = f(\boldsymbol{x})$ corresponde (adoptadas as notações acima descritas) o sistema de equações lineares:

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ &\dots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n. \end{aligned}$$

Outra representação possível é, como é sabido, a igualdade matricial:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix},$$

que, como facilmente se reconhece, permite estabelecer uma correspondência bijectiva entre as matrizes do tipo $m \times n$ de elementos reais e as aplicações lineares de $\mathbb{R}^n$ em $\mathbb{R}^m$ (convirá reter que os elementos da coluna de ordem j da matriz correspondente à aplicação f são, ordenadamente, as coordenadas na base canónica de $\mathbb{R}^m$ do vector $f(\boldsymbol{e}_j)$, para $j = 1, \dots, n$).

Outro exemplo com interesse, a que nos referiremos na sequência, este de uma aplicação de $\mathbb{R}^n$ em si mesmo (com $n > 1$), é o da função — que designaremos por μ — determinada pelo sistema:

$$\begin{aligned} y_1 &= x_1 \cos x_2 \cdots \cos x_{n-1} \cos x_n \\ y_2 &= x_1 \cos x_2 \cdots \cos x_{n-1} \operatorname{sen} x_n \\ y_3 &= x_1 \cos x_2 \cdots \operatorname{sen} x_{n-1} \\ &\dots \\ y_{n-1} &= x_1 \cos x_2 \operatorname{sen} x_3 \\ y_n &= x_1 \operatorname{sen} x_2. \end{aligned}$$

Como casos particulares ($n = 2$ e $n = 3$) obtêm-se as fórmulas usuais de mudança de coordenadas cartesianas em coordenadas polares, no plano, ou em coordenadas esféricas, no espaço, as quais, em notações mais correntes, podem escrever-se (ver Figura 3.2):

$$\begin{cases} x = r\cos\theta \\ y = r\,\mathrm{sen}\,\theta \end{cases}$$

ou

$$\begin{cases} x = r\cos\theta\cos\varphi \\ y = r\cos\theta\,\mathrm{sen}\,\varphi \\ z = r\,\mathrm{sen}\,\theta. \end{cases}$$

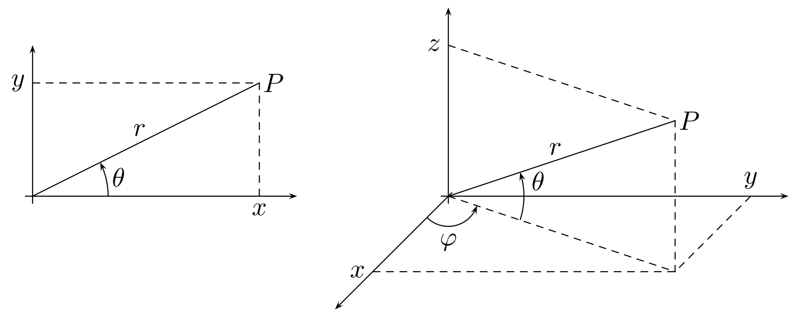

Figura 3.2

Não seria talvez necessário dizer que as operações algébricas introduzidas em $\mathbb{R}^m$ no parágrafo 2.1 se podem estender, de maneira óbvia, às funções vectoriais. Assim, por exemplo, sendo D um conjunto qualquer, f e g duas funções definidas em D e com valores em $\mathbb{R}^m$ e α um número real, a soma de f e g e o produto de α por f são as funções (designadas respectivamente por $f + g$ e αf) definidas em D e tais que, para cada $\boldsymbol{x} \in D$:

$$(f + g)(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x})$$
$$(\alpha f)(\boldsymbol{x}) = \alpha f(\boldsymbol{x}).$$

Verifica-se sem dificuldade que o conjunto de todas as funções definidas em D e com valores em $\mathbb{R}^m$ munido destas duas operações, é um espaço vectorial real.

Pode também definir-se o produto αf no caso mais geral de α ser, não já um escalar, mas uma função escalar definida em D, pondo:

$$(\alpha f)(\boldsymbol{x}) = \alpha(\boldsymbol{x}) f(\boldsymbol{x}), \quad (\boldsymbol{x} \in D).$$

De modo análogo se definem as funções escalares $f \cdot g$ e $\|f\|$.

A definição de continuidade para funções vectoriais é uma extensão imediata da que estudámos no início deste parágrafo. Seja de novo D um subconjunto de

$\mathbb{R}^n$, $f : D \to \mathbb{R}^m$ e $\boldsymbol{a}$ um ponto de D. Diz-se que f é contínua no ponto $\boldsymbol{a}$ sse para toda a bola (de $\mathbb{R}^m$) centrada em $f(\boldsymbol{a})$, $B_\delta(f(\boldsymbol{a}))$, existir uma bola (de $\mathbb{R}^n$) centrada em $\boldsymbol{a}$, $B_\epsilon(\boldsymbol{a})$ tal que se tenha $f(\boldsymbol{x}) \in B_\delta(f(\boldsymbol{a}))$ sempre que $\boldsymbol{x}$ pertença a $B_\epsilon(\boldsymbol{a}) \cap D$. Noutros termos: f é contínua em $\boldsymbol{a}$ sse qualquer que seja $\delta > 0$ existe $\epsilon > 0$ tal que para todo o $\boldsymbol{x}$ que verifique as condições: $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$ se tenha $\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| < \delta$ (como é óbvio, na expressão $\|\boldsymbol{x} - \boldsymbol{a}\|$ a norma considerada é a de $\mathbb{R}^n$, enquanto em $\|f(\boldsymbol{x}) - f(\boldsymbol{a})\|$ é a de $\mathbb{R}^m$; na sequência, cometeremos muitas vezes o «abuso» de usar o mesmo símbolo para designar normas relativas a espaços diferentes, o que não terá inconveniente de maior, porque o contexto sempre tornará evidente qual o espaço que deve ser considerado em cada caso).

Como primeiro exemplo, vejamos que qualquer aplicação linear $f : \mathbb{R}^n \to \mathbb{R}^m$ é contínua em cada ponto $\boldsymbol{a} \in \mathbb{R}^n$. Para tal, recorde-se em primeiro lugar que, sendo $\boldsymbol{x} = (x_1, \ldots, x_n)$ um vector qualquer de $\mathbb{R}^n$, se verificam as desigualdades:

$$|x_j| \le \|\boldsymbol{x}\| \le \sum_{j=1}^n |x_j| \quad (j \in \{1, 2, \ldots, n\})$$

e observe-se que, de $|x_j| \le \|\boldsymbol{x}\|$ para todo o inteiro positivo $j \le n$ se deduz imediatamente a relação:

$$\sum_{j=1}^n |x_j| \le n\|\boldsymbol{x}\|.$$

Sendo agora $f : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação linear, deverá ter-se:

$$\begin{aligned}\|f(\boldsymbol{x})\| &= \left\| f\left(\sum_{j=1}^n x_j \boldsymbol{e}_j\right)\right\| = \left\|\sum_{j=1}^n x_j f(\boldsymbol{e}_j)\right\| \\ &\le \sum_{j=1}^n \|x_j f(\boldsymbol{e}_j)\| = \sum_{j=1}^n |x_j| \|f(\boldsymbol{e}_j)\|.\end{aligned}$$

Designando por M um número positivo maior ou igual a cada um dos n números:

$$\|f(\boldsymbol{e}_1)\|,\ \|f(\boldsymbol{e}_2)\|,\ \ldots,\ \|f(\boldsymbol{e}_n)\|,$$

ter-se-á então também:

$$\|f(\boldsymbol{x})\| \le \sum_{j=1}^n M|x_j| = M\sum_{j=1}^n |x_j| \le Mn\|\boldsymbol{x}\|.$$

Obtida esta relação, válida para qualquer $\boldsymbol{x} \in \mathbb{R}^n$, seja agora $\boldsymbol{a}$ um ponto fixado arbitrariamente em $\mathbb{R}^n$; substituindo na relação referida $\boldsymbol{x}$ por $\boldsymbol{x} - \boldsymbol{a}$ e atendendo a que, por f ser linear, $f(\boldsymbol{x} - \boldsymbol{a}) = f(\boldsymbol{x}) - f(\boldsymbol{a})$, obtém-se:

$$\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| \le Mn\|\boldsymbol{x} - \boldsymbol{a}\|,$$

desigualdade que torna evidente a continuidade de f no ponto $\boldsymbol{a}$: dado $\delta > 0$, bastará tomar ϵ positivo e menor do que δ/Mn para que se tenha $\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| < \delta$ sempre que seja $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$.

Antes de iniciar o estudo de algumas propriedades importantes das funções contínuas, mostraremos ainda que a aplicação de $\mathbb{R}^n$ em si mesmo atrás designada por μ é contínua na origem de $\mathbb{R}^n$ (de posse das propriedades que iremos estudar adiante o resultado obter-se-á mais facilmente e poderá ver-se até que μ é contínua em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$, o que seria difícil neste momento). Com efeito, do sistema de equações que usámos para definir a função μ deduz-se facilmente, por um lado que μ é nula na origem ($\mu(\boldsymbol{O}) = \boldsymbol{O}$), por outro que, sendo $\boldsymbol{x} = (x_1, \ldots, x_n)$ e $\boldsymbol{y} = (y_1, \ldots, y_n) = \mu(\boldsymbol{x})$, se tem:

$$y_1^2 + y_2^2 + \cdots + y_n^2 = x_1^2,$$

isto é:

$$\|\mu(\boldsymbol{x})\|^2 = x_1^2$$

e portanto também:

$$\|\mu(\boldsymbol{x})\| = |x_1| \leq \|\boldsymbol{x}\|,$$

para qualquer $\boldsymbol{x} \in \mathbb{R}^n$. Assim, dado $\delta > 0$ basta tomar $\epsilon = \delta$ para que se verifique a desigualdade $\|\mu(\boldsymbol{x}) - \mu(\boldsymbol{O})\| < \delta$ sempre que seja $\|\boldsymbol{x}\| < \epsilon$.

0 teorema seguinte revela que, tal como no caso das funções reais de variável real, a noção de continuidade pode exprimir-se em termos da noção de limite de sucessões:

Teorema 3.1. *Seja $f : D \to \mathbb{R}^m$ ($D \subset \mathbb{R}^n$) e $\boldsymbol{a} \in D$; para que f seja contínua no ponto $\boldsymbol{a}$ é necessário e suficiente que, sempre que $\boldsymbol{x}_k$ seja uma sucessão[1] em D convergente para $\boldsymbol{a}$, a sucessão $f(\boldsymbol{x}_k)$ convirja para $f(\boldsymbol{a})$.*

Demonstração. Daremos uma demonstração praticamente idêntica à do caso das funções reais de variável real.

Suponha-se em primeiro lugar que f é contínua em $\boldsymbol{a}$ e seja $\boldsymbol{x}_k$ uma sucessão em D convergente para $\boldsymbol{a}$. Dado um número positivo arbitrário δ, existe $\epsilon > 0$ tal que, se $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$, $\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| < \delta$. Como $\boldsymbol{x}_k \to \boldsymbol{a}$, existe um inteiro positivo p tal que $\|\boldsymbol{x}_k - \boldsymbol{a}\| < \epsilon$ para todo o $k > p$; e então, como $\boldsymbol{x}_k \in D$ qualquer que seja $k \in \mathbb{N}_1$, ter-se-á também, para $k > p$, $\|f(\boldsymbol{x}_k) - f(\boldsymbol{a})\| < \delta$, o que prova que $f(\boldsymbol{x}_k) \to f(\boldsymbol{a})$.

Em sentido inverso, se a função f não é contínua no ponto $\boldsymbol{a}$, existe $\delta > 0$ tal que, qualquer que seja $\epsilon > 0$ haverá pelo menos um ponto $\boldsymbol{x}$ pertencente a D e verificando ambas as condições:

$$\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon \quad \text{e} \quad \|f(\boldsymbol{x}) - f(\boldsymbol{a})\| \geq \delta.$$

[1]Evitaremos, naturalmente, o uso (já de si «abusivo») do símbolo $\boldsymbol{x}_n$ para designar a sucessão considerada, dado que a letra n está a ser utilizada para designar a dimensão do espaço que contém o domínio de f.

Pondo $\epsilon = 1/k$ poderá portanto escolher-se (para cada $k \in \mathbb{N}_1$) um ponto $\boldsymbol{x}_k \in D$ por forma que sejam conjuntamente verificadas as desigualdades:

$$\|\boldsymbol{x}_k - \boldsymbol{a}\| < 1/k \quad \text{e} \quad \|f(\boldsymbol{x}_k) - f(\boldsymbol{a})\| \geq \delta.$$

Obter-se-á assim uma sucessão de termos em D, convergente para $\boldsymbol{a}$ (como resulta da primeira dessas desigualdades) e tal que $f(\boldsymbol{x}_k)$ não converge para $f(\boldsymbol{a})$ (como mostra a segunda), o que termina a demonstração. □

De forma sugestiva, embora um pouco imprecisa, pode dizer-se que a continuidade de f no ponto $\boldsymbol{a}$ equivale à possibilidade de permutar os símbolos «f» e «lim»:

$$\lim f(\boldsymbol{x}_k) = f(\lim \boldsymbol{x}_k),$$

quando aplicados sucessivamente a qualquer sucessão em D convergente para $\boldsymbol{a}$.

Tendo em conta o precedente Teorema 3.1 e algumas propriedades da noção de limite de uma sucessão mencionadas em 2.2, obtêm-se sem qualquer dificuldade os resultados seguintes (em cujos enunciados se supõe $\boldsymbol{a} \in D \subset \mathbb{R}^n$; $f, g : D \to \mathbb{R}^m$ e $\alpha : D \to \mathbb{R}$).

- *Se f é constante em D, é contínua em qualquer ponto de D.*

- *Se f e g são contínuas no ponto $\boldsymbol{a}$, também o são $f + g$, $f - g$, $f \cdot g$ e $\|f\|$* (como caso particular — para $m = 1$ — resulta que se as funções reais f e g são contínuas no ponto $\boldsymbol{a} \in D$, são também contínuos no mesmo ponto o seu produto usual, fg, e a função $|f|$).

- *Se α e f são contínuas no ponto $\boldsymbol{a}$, αf também o é; se, além disso, for $\alpha(\boldsymbol{a}) \neq 0$, o cociente $f/\alpha = 1/\alpha f$* — função definida nos pontos $\boldsymbol{x} \in D$ tais que $\alpha(\boldsymbol{x}) \neq 0$ — *é contínuo no ponto $\boldsymbol{a}$* (em particular, o cociente de duas funções reais definidas em D e contínuas no ponto $\boldsymbol{a}$ é uma função contínua no mesmo ponto, desde que nele se não anule a função que figura em denominador).

Sejam agora m, n e p três números inteiros positivos, D um subconjunto de $\mathbb{R}^n$ e E um subconjunto de $\mathbb{R}^p$; sejam ainda g uma aplicação de D em $\mathbb{R}^p$ cujo contradomínio esteja contido em E e f uma aplicação de E em $\mathbb{R}^m$. Nestas condições, a composta $f \circ g$, definida por:

$$(f \circ g)(\boldsymbol{x}) = f\big(g(\boldsymbol{x})\big)$$

é uma aplicação de D em $\mathbb{R}^m$, reconhecendo-se imediatamente (utilizando, por exemplo, o Teorema 3.1) que:

- *Se g é contínua num ponto $\boldsymbol{a} \in D$ e f é contínua no ponto $g(\boldsymbol{a})$, então $f \circ g$ é contínua no ponto $\boldsymbol{a}$.*

Com estes resultados, fica muito facilitado o estudo da continuidade para a generalidade das funções de variável vectorial que surgem mais frequentemente nas aplicações.

Consideremos em primeiro lugar o caso das funções reais ($m = 1$) e, para maior facilidade, suponhamos por agora que são apenas duas as variáveis independentes, que designaremos por x e y, em lugar de x_1 e x_2 (voltamos assim de momento às notações usadas de início, no parágrafo 1.2).

É fácil ver que as funções p_1 e p_2 definidas em $\mathbb{R}^2$ pelas fórmulas:

$$p_1(x, y) = x \quad \text{e} \quad p_2(x, y) = y$$

são contínuas em qualquer ponto $(a, b) \in \mathbb{R}^2$ (para p_1, por exemplo, basta atender a que

$$|p_1(x, y) - p_1(a, b)| = |x - a| \leq \sqrt{(x - a)^2 + (y - b)^2} = \|(x, y) - (a, b)\|,$$

o que mostra que se terá $|p_1(x, y) - p_1(a, b)| < \delta$ sempre que (x, y) pertença à bola de centro (a, b) e raio $\epsilon = \delta$).

Deste facto resulta imediatamente, atendendo a propriedades da continuidade acabadas de referir, que a função f considerada no exemplo 1. de 1.2:

$$f(x, y) = x^2 + y^2 \quad \big((x, y) \in \mathbb{R}^2\big)$$

é contínua em qualquer ponto $(a, b) \in \mathbb{R}^2$ (basta notar que $f = p_1 p_1 + p_2 p_2$ é a soma de produtos de funções contínuas nesse ponto); mais geralmente, pode concluir-se de modo análogo que qualquer função polinomial $P(x, y)$ — isto é, qualquer função que possa representar-se como soma de (um número finito de) «monómios» da forma geral $cx^r y^s$, onde c é uma constante real e r e s inteiros não negativos — é contínua em qualquer ponto de $\mathbb{R}^2$; e também que qualquer função racional de duas variáveis reais, representável como cociente de duas funções polinomiais:

$$\frac{P(x, y)}{Q(x, y)}$$

(não sendo $Q(x, y)$ o polinómio nulo) é contínua em todos os pontos $(x, y) \in \mathbb{R}^2$ tais que $Q(x, y) \neq 0$, isto é, em todos os pontos do seu domínio.

Por sua vez o resultado relativo à continuidade de uma função composta de funções contínuas e alguns dos conhecimentos obtidos no estudo das funções reais de variável real permitem analisar facilmente, do ponto de vista da continuidade, muitas funções não racionais correntes nas aplicações.

A título de exemplo, consideremos a função

$$\varphi(x, y) = \operatorname{arctg} \frac{x^3 + y^3}{1 - x^2}$$

(suposta definida no subconjunto D de $\mathbb{R}^2$ formado por todos os pontos (x, y) que verificam as condições $x \neq 1$ e $x \neq -1$). Como se tem $\varphi = \psi \circ \theta$, com:

$$\psi(u) = \operatorname{arctg} u \quad (u \in \mathbb{R})$$

e

$$\theta(x,y) = \frac{x^3+y^3}{1-x^2} \quad (x \in D),$$

sendo θ contínua em todos os pontos de D (por ser uma função racional) e ψ contínua em cada ponto do contradomínio de θ (visto que é contínua em $\mathbb{R}$) pode concluir-se que φ é contínua em todos os pontos do seu domínio.

Claro que estas ideias se estendem de forma óbvia ao caso de funções reais de n variáveis reais $x_1, x_2, \ldots, x_n$ (com $n > 2$). Por exemplo, a continuidade em qualquer ponto $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ de uma função polinomial $P(\boldsymbol{x}) = P(x_1, \ldots, x_n)$ — isto é, de uma função representável como soma de «monómios» do tipo $cx_1^{r_1}x_2^{r_2}\ldots x_n^{r_n}$ — resulta imediatamente da continuidade (facilmente provada) das «projecções» p_j:

$$p_j(\boldsymbol{x}) = p_j(x_1, \ldots, x_n) = x_j \quad (j \in \{1, \ldots, n\})$$

e dos resultados há pouco enunciados sobre a continuidade das funções constantes e das somas e produtos de funções contínuas. De forma análoga se conclui a continuidade de uma função racional de n variáveis reais:

$$\frac{P(x_1, x_2, \ldots, x_n)}{Q(x_1, x_2, \ldots, x_n)}$$

em todos os pontos $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ tais que $Q(\boldsymbol{x}) \neq 0$; e o teorema que relaciona a continuidade com a composição de funções permite uma vez mais alargar consideravelmente o quadro das funções cujo estudo, deste ponto de vista, pode efectuar-se com extrema simplicidade.

Assim, por exemplo, reconhece-se imediatamente que a função de m variáveis reais mencionada como exemplo em 1.4:

$$y = h(x_1, \ldots, x_m) = \log(x_1^2 + \cdots + x_m^2),$$

que é o resultado da composição de $y = \log u$ (contínua para $u > 0$) com a função polinomial $u = x_1^2 + \cdots + x_m^2$ (contínua em todos os pontos de $\mathbb{R}^m$) é contínua em qualquer ponto de $\mathbb{R}^m$ distinto da origem, isto é, em todos os pontos do seu domínio.

Passemos agora ao caso das funções vectoriais, o qual, como vamos ver, se reduz trivialmente ao das funções reais que acabamos de analisar; neste sentido, o resultado essencial é o que se exprime no seguinte:

Teorema 3.2. *Seja $f : D \to \mathbb{R}^m$, com $D \subset \mathbb{R}^n$ e $\boldsymbol{a} \in D$; para que f seja contínua no ponto $\boldsymbol{a}$ é necessário e suficiente que sejam contínuas no mesmo ponto todas as suas funções coordenadas.*

Demonstração. Consideremos as relações já habituais (verificadas para todo o $\boldsymbol{x} \in D$ e para $i \in \{1, \ldots, m\}$):

$$|f_i(\boldsymbol{x}) - f_i(\boldsymbol{a})| \leq \|f(\boldsymbol{x}) - f(\boldsymbol{a})\| \leq \sum_{i=1}^{m} \big|f_i(\boldsymbol{x}) - f_i(\boldsymbol{a})\big|.$$

A primeira desigualdade mostra que, dado $\delta > 0$, se determinarmos $\epsilon > 0$ por forma que se tenha $\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| < \delta$ sempre que $\boldsymbol{x}$ pertença ao conjunto $B_\epsilon(\boldsymbol{a}) \cap D$ — o que é possível, se f for contínua em $\boldsymbol{a}$ — se terá também (para qualquer inteiro positivo $i \leq m$) $|f_i(\boldsymbol{x}) - f_i(\boldsymbol{a})| < \delta$ para todo o $\boldsymbol{x}$ nesse mesmo conjunto: assim, a continuidade de f implica a de todas as suas funções coordenadas.

Reciprocamente, suponha-se que, para $i \in \{1, \ldots, m\}$, f_i é contínua no ponto $\boldsymbol{a}$ e seja δ um número positivo arbitrário; determine-se para cada inteiro positivo $i \leq m$ um número positivo ϵ_i tal que

$$\boldsymbol{x} \in B_{\epsilon_i}(\boldsymbol{a}) \cap D \Longrightarrow |f_i(\boldsymbol{x}) - f_i(\boldsymbol{a})| < \frac{\delta}{m}$$

e designe-se por ϵ o menor dos números $\epsilon_1, \ldots, \epsilon_m$. Então, sempre que se tenha $x \in B_\epsilon(\boldsymbol{a}) \cap D$ ter-se-á também:

$$\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| \leq \sum_{i=1}^{m} |f_i(\boldsymbol{x}) - f_i(\boldsymbol{a})| < \sum_{i=1}^{m} \frac{\delta}{m} = \delta,$$

o que prova a continuidade de f no ponto $\boldsymbol{a}$. $\square$

Seria agora bastante fácil justificar a continuidade em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$ de uma aplicação linear f de $\mathbb{R}^n$ em $\mathbb{R}^m$: bastaria observar que cada uma das coordenadas de f é uma função polinomial; e seria também quase imediata a prova de uma afirmação anterior, relativa à continuidade em cada ponto do seu domínio da aplicação de $\mathbb{R}^n$ em si mesmo atrás designada por μ: com efeito, facilmente se verifica que cada uma das funções coordenadas de μ é contínua em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$.

Um outro exemplo muito simples: designemos por I a aplicação idêntica de $\mathbb{R}^n$ em si mesmo, $I(\boldsymbol{x}) = \boldsymbol{x}$ para cada $\boldsymbol{x} \in \mathbb{R}^n$; para cada inteiro positivo $j \leq n$, a função coordenada de ordem j da aplicação I é precisamente a projecção p_j,

$$p_j(\boldsymbol{x}) = p_j(x_1, \ldots, x_n) = x_j.$$

E, atendendo a que I é evidentemente contínua em qualquer ponto de $\mathbb{R}^n$, logo se confirma a continuidade, atrás mencionada, de cada projecção p_j em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$.

Antes de passarmos ao estudo da continuidade de um ponto de vista global, enunciaremos ainda dois resultados muito simples — consequências imediatas da definição de continuidade — cujas demonstrações poderão ficar como exercícios para o leitor.

O primeiro pode enunciar-se nos termos seguintes:

Se f é uma função real definida em $D \subset \mathbb{R}^n$ e contínua no ponto $\boldsymbol{a} \in D$ e se $f(\boldsymbol{a}) > 0$, então existe $\epsilon > 0$ tal que, para todo o $\boldsymbol{x} \in B_\epsilon(\boldsymbol{a}) \cap D$ se tem $f(\boldsymbol{x}) > 0$; no caso de uma função vectorial $f : D \to \mathbb{R}^m$, contínua no ponto $\boldsymbol{a}$, poderá por exemplo afirmar-se que, se $f(\boldsymbol{a}) \neq \boldsymbol{b}$ (com $\boldsymbol{b} \in \mathbb{R}^m$), existe $\epsilon > 0$ tal que $f(\boldsymbol{x}) \neq \boldsymbol{b}$ sempre que $\boldsymbol{x} \in B_\epsilon(\boldsymbol{a}) \cap D$.

Antes de enunciar o segundo resultado convém referir que, sendo ainda f uma função definida em D e com valores em $\mathbb{R}^m$, se diz que f é *limitada* sse existe $k \in \mathbb{R}$ tal que $\|f(\boldsymbol{x})\| \leq k$ para todo o $\boldsymbol{x} \in D$, isto é, sse o contradomínio de f, $f(D)$, é um conjunto limitado. Mais geralmente, sendo A um subconjunto de D, diz-se que f é *limitada em A sse a restrição de f a A* (isto é, a função $f_{/A}$ que tem por domínio o conjunto A e verifica a condição $f_{/A}(\boldsymbol{x}) = f(\boldsymbol{x})$ para todo o $\boldsymbol{x} \in A$) é limitada; ou, o que é o mesmo, sse for limitado o conjunto:

$$f(A) = \{f(\boldsymbol{x}) : \boldsymbol{x} \in A\},$$

que é o contradomínio de $f_{/A}$, também designado por *transformado do conjunto A pela função f*.

Após estas definições, podemos enunciar o segundo dos resultados acima referidos:

Se $f : D \to \mathbb{R}^m$, com $D \subset \mathbb{R}^n$ é uma função contínua no ponto $\boldsymbol{a}$, existe $\epsilon > 0$ tal que f é limitada no conjunto $B_\epsilon(A) \cap D$.

Seja agora D um subconjunto de $\mathbb{R}^n$ e f uma função definida em D e com valores em $\mathbb{R}^m$; nestas condições, diz-se que f é *contínua em D* (ou apenas que f é *contínua*) sse f é contínua em cada ponto $\boldsymbol{a} \in D$. Mais geralmente, sendo A um subconjunto de D, diz-se que f é *continua em A* sse $f_{/A}$ é contínua (em A).

Dispomos agora de todos os elementos necessários para a generalização dos teoremas fundamentais — teoremas de Weierstrass e de Heine--Cantor, teorema do valor intermédio, etc. — que estudámos no quadro das funções contínuas de variável real.

Começaremos pelo seguinte:

Teorema 3.3. *Seja D um conjunto compacto de $\mathbb{R}^n$ e $f : D \to \mathbb{R}^m$ uma função contínua (em D). Nestas condições, o contradomínio de f, $f(D)$, é um conjunto compacto.*

Demonstração. De acordo com o Teorema 2.9, bastará provar que qualquer sucessão de termos em $f(D)$ tem uma subsucessão convergente para um ponto de $f(D)$. Seja então $\boldsymbol{y}_k$ uma sucessão qualquer em $f(D)$ e, para cada inteiro positivo k, escolha-se um ponto $\boldsymbol{x}_k \in D$ tal que $f(\boldsymbol{x}_k) = \boldsymbol{y}_k$. Como D é um conjunto compacto, poderá extrair-se da sucessão $\boldsymbol{x}_k$ uma subsucessão $\boldsymbol{x}_{p_k}$ convergente para um ponto $\boldsymbol{x}_0 \in D$; e como f é contínua em D (e portanto em $\boldsymbol{x}_0$), do facto de $\boldsymbol{x}_{p_k}$ convergir para $\boldsymbol{x}_0$ pode deduzir-se que $\boldsymbol{y}_{p_k} = f(\boldsymbol{x}_{p_k})$ — que é uma subsucessão da sucessão $\boldsymbol{y}_k$ — converge para o ponto $\boldsymbol{y}_0 = f(\boldsymbol{x}_0) \in f(D)$, o que termina a demonstração. $\square$

Registaremos agora mais algumas definições importantes, que quase seria desnecessário formular explicitamente dada a sua semelhança com as que conhecemos do estudo das funções de variável real.

Sendo f uma função real definida em $D \subset \mathbb{R}^n$, diz-se que f *tem máximo (em D)* sse existe $\boldsymbol{x}_0 \in D$ tal que $f(\boldsymbol{x}) \leq f(\boldsymbol{x}_0)$ para todo o $\boldsymbol{x} \in D$; qualquer ponto $\boldsymbol{x}_0$ que verifique a condição indicada diz-se um *ponto de máximo* (ou um *maximizante*) de f e o valor $f(\boldsymbol{x}_0)$ é o *máximo* da função (em D), designado por $\max_D f$ ou $\max_{\boldsymbol{x}\in D} f(\boldsymbol{x})$. Definem-se de forma análoga as noções de *mínimo*, *ponto de mínimo*, etc.

Mais geralmente, sendo A um subconjunto qualquer do domínio D da função f, diz-se que f tem *máximo em* A sse $f_{/A}$ tem máximo (em A); e nessa hipótese chama-se *máximo de* f *em* A ($\max_A f$ ou $\max_{\boldsymbol{x}\in A} f(\boldsymbol{x})$) ao máximo da sua restrição, $\max_A f_{/A}$. Como é óbvio, f tem máximo em A sse o conjunto $f(A)$ tiver máximo, verificando-se então a igualdade:

$$\max_{\boldsymbol{x}\in A} f(\boldsymbol{x}) = \max f(A).$$

Definem-se ainda, de forma óbvia, as noções de *supremo* e *ínfimo* de uma função real f num subconjunto A do seu domínio D ($\sup_A f$, $\inf_A f$, etc.) podendo, em particular, ser $A = D$. Para que existam conjuntamente o supremo e o ínfimo de f em A (suposto não vazio) é necessário e suficiente que f seja limitada em A e, em tal hipótese, f terá máximo em A sse existir um ponto $\boldsymbol{x}_0 \in A$ tal que

$$f(\boldsymbol{x}_0) = \sup_{\boldsymbol{x}\in A} f(\boldsymbol{x}),$$

tendo-se então $\max_A f = \sup_A f$; e analogamente para o mínimo e o ínfimo. Como simples consequência do Teorema 3.3, podemos agora enunciar:

Teorema 3.4 (Weierstrass). *Se $D \subset \mathbb{R}^n$ é um conjunto compacto não vazio, qualquer função real f, definida e contínua em D, tem máximo e mínimo nesse conjunto.*

Demonstração. Nas condições da hipótese, decorre do teorema anterior que $f(D)$ é um subconjunto compacto, não vazio, de $\mathbb{R}$; por ser limitado e não vazio, $f(D)$ terá supremo e ínfimo em $\mathbb{R}$, os quais serão necessariamente pontos aderentes a $f(D)$ — é evidente que lhe não podem ser exteriores — e portanto pertencerão a $f(D)$, por este conjunto ser fechado. Conclui-se assim que $f(D)$ tem máximo e mínimo, isto é, que f tem máximo e mínimo no conjunto D. □

Trataremos agora de generalizar uma outra noção de extrema importância, a de continuidade uniforme. Seja $f : D \to \mathbb{R}^m$, com $D \subset \mathbb{R}^n$, e seja A um subconjunto de D (podendo ser, em particular, $A = D$); diz-se que f é *uniformemente contínua no conjunto* A sse para todo o $\delta > 0$ existe $\epsilon > 0$ tal que, quaisquer que sejam os pontos $\boldsymbol{x}, \boldsymbol{x}' \in A$ verificando a condição $\|\boldsymbol{x} - \boldsymbol{x}'\| < \epsilon$, se tiver $\|f(\boldsymbol{x}) - f(\boldsymbol{x}')\| < \delta$.

Como exemplo com interesse, mencionaremos o de uma aplicação linear $f : \mathbb{R}^n \to \mathbb{R}^m$. Para provar que uma tal aplicação é contínua em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$, deduzimos atrás a relação (válida para $\boldsymbol{a}, \boldsymbol{x} \in \mathbb{R}^n$):

$$\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| \leq Mn\|\boldsymbol{x} - \boldsymbol{a}\|,$$

onde M designava um número positivo, independente de $\boldsymbol{a}$ e $\boldsymbol{x}$. É fácil reconhecer agora, a partir desta mesma relação, que f é uniformemente contínua em $\mathbb{R}^n$: com efeito, dado $\delta > 0$, bastará tomar um número positivo $\epsilon < \delta/Mn$ para que se tenha $\|f(\boldsymbol{x}) - f(\boldsymbol{a})\| < \delta$ sempre que $\boldsymbol{a}$ e $\boldsymbol{x}$ sejam dois pontos de $\mathbb{R}^n$ tais que $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$.

Reconhece-se sem dificuldade que uma função uniformemente contínua num conjunto é contínua no mesmo conjunto, sendo a recíproca falsa, em geral, como é sabido do estudo das funções de uma variável real. Verifica-se, no entanto, o seguinte resultado fundamental:

Teorema 3.5 (Heine–Cantor). *Seja D um subconjunto compacto de $\mathbb{R}^n$. Qualquer função $f : D \to \mathbb{R}^m$ contínua em D, é uniformemente contínua no mesmo conjunto.*

Demonstração. Suponha-se que alguma função f, nas condições da hipótese, não era uniformemente contínua em D. Existiria então um número positivo δ tal que, para cada $\epsilon > 0$ seria possível determinar dois pontos $\boldsymbol{x}, \boldsymbol{x}' \in D$ por forma que fossem conjuntamente verificadas as desigualdades:

$$\|\boldsymbol{x} - \boldsymbol{x}'\| < \epsilon \quad \text{e} \quad \|f(\boldsymbol{x}) - f(\boldsymbol{x}')\| \geq \delta.$$

Pondo $\epsilon = 1/k$ (com $k = 1, 2, \ldots$) poderia assim obter-se para cada $k \in \mathbb{N}_1$ um par de pontos $\boldsymbol{x}_k, \boldsymbol{x}'_k \in D$ verificando as condições:

$$\|\boldsymbol{x}_k - \boldsymbol{x}'_k\| < \frac{1}{k} \quad \text{e} \quad \|f(\boldsymbol{x}_k) - f(\boldsymbol{x}'_k)\| \geq \delta.$$

Da sucessão $\boldsymbol{x}_k$, de termos no conjunto compacto D, poderia extrair-se uma subsucessão $\boldsymbol{x}_{p_k}$, convergente para um ponto $\boldsymbol{x}_0 \in D$. E, atendendo às relações:

$$\begin{aligned}
\|\boldsymbol{x}'_{p_k} - \boldsymbol{x}_0\| &= \|(\boldsymbol{x}'_{p_k} - \boldsymbol{x}_{p_k}) + (\boldsymbol{x}_{p_k} - \boldsymbol{x}_0)\| \\
&\leq \|\boldsymbol{x}'_{p_k} - \boldsymbol{x}_{p_k}\| + \|\boldsymbol{x}_{p_k} - \boldsymbol{x}_0\| \\
&< \frac{1}{p_k} + \|\boldsymbol{x}_{p_k} - \boldsymbol{x}_0\|,
\end{aligned}$$

logo se reconhece que também a sucessão $\boldsymbol{x}'_{p_k}$ seria convergente para $\boldsymbol{x}_0$.

Ter-se-ia assim, dada a continuidade de f em D (e portanto em $\boldsymbol{x}_0$)

$$\lim_{k\to\infty} f(\boldsymbol{x}_{p_k}) = f(\boldsymbol{x}_0) = \lim_{k\to\infty} f(\boldsymbol{x}'_{p_k})$$

e portanto também:

$$\lim_{k\to\infty} \big(f(\boldsymbol{x}_{p_k}) - f(\boldsymbol{x}'_{p_k})\big) = \mathbf{0},$$

em contradição com o facto de dever ser verificada, para todo o $k \in \mathbb{N}_1$, a desigualdade:

$$\|f(\boldsymbol{x}_k) - f(\boldsymbol{x}'_k)\| \geq \delta.$$

Esta contradição permite dar por concluída a demonstração do teorema. □

A noção de continuidade uniforme e o precedente teorema de Heine--Cantor ser-nos-ão indispensáveis em diversas fases ulteriores do nosso curso. Um outro resultado com interesse na sequência é o que se exprime no Teorema 3.6, o qual constitui a generalização adequada de um resultado conhecido, relativo à continuidade da função inversa de uma função contínua que aplique injectivamente um intervalo $I \subset \mathbb{R}$ na recta real $\mathbb{R}$.

Teorema 3.6. *Seja $f : D \to \mathbb{R}^m$ uma função contínua no conjunto compacto $D \subset \mathbb{R}^n$ e suponha-se que f aplica injectivamente D em $\mathbb{R}^m$; então a função inversa $g = f^{-1} : f(D) \to \mathbb{R}^n$ é contínua em $f(D)$.*

Demonstração. Tendo em conta o Teorema 3.1, bastará provar que, sendo $\boldsymbol{y}_0$ um ponto arbitrário de $f(D)$ e $\boldsymbol{y}_k$ uma sucessão qualquer de termos em $f(D)$ convergente para $\boldsymbol{y}_0$, se tem necessariamente $g(\boldsymbol{y}_k) \to g(\boldsymbol{y}_0)$.

Ponha-se $\boldsymbol{x}_0 = g(\boldsymbol{y}_0)$ e, para todo o inteiro positivo k, $\boldsymbol{x}_k = g(\boldsymbol{y}_k)$; $\boldsymbol{x}_k$ será uma sucessão de termos em D, $\boldsymbol{x}_0$ um ponto de D e ter-se-á:

$$\boldsymbol{y}_k = f(\boldsymbol{x}_k) \to \boldsymbol{y}_0 = f(\boldsymbol{x}_0),$$

interessando agora provar que $\boldsymbol{x}_k \to \boldsymbol{x}_0$.

Recorrendo directamente à definição de limite de uma sucessão logo se vê que, se $\boldsymbol{x}_k$ não convergisse para $\boldsymbol{x}_0$, existiria $\epsilon > 0$ tal que, para uma infinidade de valores inteiros positivos de k, não seria verificada a condição $\|\boldsymbol{x}_k - \boldsymbol{x}_0\| < \epsilon$; ou, de outra forma: existiria uma subsucessão $\boldsymbol{x}_{p_k}$ de $\boldsymbol{x}_k$ para a qual se teria $\|\boldsymbol{x}_{p_k} - \boldsymbol{x}_0\| \geq \epsilon$ para todo o $k \in \mathbb{N}_1$. Pelo Teorema 2.9, a sucessão $\boldsymbol{x}_{p_k}$, de termos no conjunto compacto D, admitiria por sua vez uma subsucessão $\boldsymbol{x}_{q_k}$ (também subsucessão de $\boldsymbol{x}_k$) convergente para um ponto $\boldsymbol{x}_0' \in D$; mas, verificando-se necessariamente, para todo o inteiro positivo k, a condição:

$$\|\boldsymbol{x}_{q_k} - \boldsymbol{x}_0\| \geq \epsilon,$$

o limite $\boldsymbol{x}_0'$ da sucessão $\boldsymbol{x}_{q_k}$ seria certamente distinto do ponto $\boldsymbol{x}_0$ e portanto — pondo $\boldsymbol{y}_0' = f(\boldsymbol{x}_0')$ — ter-se-ia também, dada a injectividade de f:

$$\boldsymbol{y}_0' = f(\boldsymbol{x}_0') \neq f(\boldsymbol{x}_0) = \boldsymbol{y}_0.$$

Nestas condições, porém, a continuidade de f em $\boldsymbol{x}_0'$ e a convergência de $\boldsymbol{x}_{q_k}$ para $\boldsymbol{x}_0'$ implicariam que a sucessão $\boldsymbol{y}_{q_k} = f(\boldsymbol{x}_{q_k})$ convergisse para $\boldsymbol{y}_0'$, o que é absurdo, porque $\boldsymbol{y}_{q_k}$ é uma subsucessão de $\boldsymbol{y}_k$ e $\boldsymbol{y}_k$, por hipótese, converge para $\boldsymbol{y}_0$. Pode assim considerar-se terminada a demonstração. $\square$

Cada um dos dois teoremas seguintes constitui, de certo ponto de vista, uma generalização natural do clássico teorema do valor intermédio, relativo a funções contínuas num intervalo da recta $\mathbb{R}$; porém, o Teorema 3.8 não é mais do que um simples corolário do:

Teorema 3.7. *Seja $f : D \to \mathbb{R}^m$, com $D \subset \mathbb{R}^n$; se f é contínua (em D) e D é um conjunto conexo, $f(D)$ é também conexo.*

Demonstração. Suponha-se que, sendo f contínua em D, $f(D)$ não era conexo. $f(D)$ seria então a reunião de dois conjuntos separados, isto é, existiriam dois conjuntos não vazios A^* e B^* verificando as condições:

$$A^* \cap \bar{B}^* = \emptyset, \quad \bar{A}^* \cap B^* = \emptyset$$

e

$$f(D) = A^* \cup B^*.$$

Designemos[2] por A o conjunto de todos os pontos $\boldsymbol{x} \in D$ tais que $f(\boldsymbol{x}) \in A^*$ e por B o conjunto dos $\boldsymbol{x} \in D$ tais que $f(\boldsymbol{x}) \in B^*$.

Deduz-se imediatamente que A e B seriam não vazios (porque A^* e B^*, contidos em $f(D)$ são não vazios) e também que $A \cup B = D$ (porque sendo $f(D) = A^* \cup B^*$, para qualquer $\boldsymbol{x} \in D$ se verificaria necessariamente uma das condições $f(\boldsymbol{x}) \in A^*$ ou $f(\boldsymbol{x}) \in B^*$).

Ter-se-ia ainda, como vamos ver,

$$A \cap \bar{B} = \emptyset \quad \text{e} \quad \bar{A} \cap B = \emptyset$$

(provaremos apenas a primeira igualdade, já que a prova da segunda seria idêntica). Com efeito, se algum ponto $\boldsymbol{x}_0 \in A$ fosse aderente ao conjunto B, existiria (pelo Teorema 2.4) uma sucessão $\boldsymbol{x}_k$ de termos em B convergente para $\boldsymbol{x}_0$; mas então, dada a continuidade de f em $\boldsymbol{x}_0$, a sucessão $f(\boldsymbol{x}_k)$, de termos em B^*, convergiria para $f(\boldsymbol{x}_0) \in A^*$ e (pelo mesmo teorema há pouco mencionado) poderia deduzir-se que $f(\boldsymbol{x}_0)$ era aderente ao conjunto B^*, isto é, que $A^* \cap \bar{B}^* \neq \emptyset$.

Assim, na hipótese de f ser contínua e $f(D)$ não ser conexo, concluir-se-ia que o conjunto D era a reunião de dois conjuntos separados, isto é, que o conjunto D era desconexo; esta conclusão é obviamente equivalente ao que pretendíamos provar. □

Uma consequência imediata é o seguinte:

Teorema 3.8 (teorema do valor intermédio). *Seja f uma função real, definida e contínua no conjunto conexo $D \subset \mathbb{R}^n$; se os números reais α e β, com $\alpha < \beta$, pertencem ao contradomínio de f e se γ é um real tal que $\alpha < \gamma < \beta$, então existe pelo menos um ponto $\boldsymbol{x} \in D$ tal que $f(\boldsymbol{x}) = \gamma$.*

Demonstração. Basta observar que, nas condições da hipótese, $f(D)$ é um conjunto conexo de $\mathbb{R}$, isto é, um intervalo (Teorema 2.10). □

[2] Usa-se correntemente em situações como esta a notação $f^{-1}(A^*)$ para designar o conjunto A, chamado *imagem recíproca* ou *imagem inversa por meio de f do conjunto A^**; convém observar que tais notações são usadas mesmo em casos, como o presente, em que se não supõe que f seja injectiva, podendo portanto não existir a função inversa, f^{-1}.

De uma forma geral, os precedentes Teoremas 3.3 a 3.8 estabelecem certas propriedades importantes do contradomínio de uma função (ou da própria função, ou da sua inversa, quando existente) decorrentes da hipótese da função ser contínua e do seu domínio ser um conjunto com certas características especiais (compacto ou conexo). Como é óbvio, qualquer desses resultados é susceptível de uma extensão trivial, resultante de se considerar, em lugar do domínio D da função f, um subconjunto qualquer A de D que possua também as características em causa; a função poderá então ser substituída nos raciocínios pela sua restrição ao conjunto A e só a respeito desta restrição haverá que pôr a hipótese de continuidade. Assim, por exemplo, sendo $f : D \to \mathbb{R}^m$ (com $D \subset \mathbb{R}^n$), $A \subset D$ e f contínua no conjunto A poderá concluir-se que, se A for compacto o mesmo se verificará com $f(A)$ (abreviadamente: as funções contínuas transformam conjuntos compactos em conjuntos compactos); que, na mesma hipótese sobre A, f é uniformemente contínua em A (as funções contínuas em conjuntos compactos são uniformemente contínuas); que $f(A)$ é conexo se A o for (as funções contínuas transformam conjuntos conexos em conjuntos conexos), etc.

No caso $m = 1$, pode ainda concluir-se que, sendo A um compacto não vazio, $f(A)$ tem máximo e mínimo (qualquer função real contínua num conjunto compacto não vazio tem máximo e mínimo nesse conjunto), etc.

Como exercício útil, o leitor poderá procurar exemplos capazes de mostrar que, nos enunciados dos teoremas referidos, não seria possível «enfraquecer» as hipóteses sem prejudicar a generalidade das conclusões; por exemplo: num conjunto que não seja compacto há sempre funções que não são uniformemente contínuas e funções reais contínuas que não têm máximo ou mínimo; num conjunto desconexo há sempre funções contínuas com contradomínio desconexo; uma função não contínua pode transformar conjuntos conexos em conjuntos desconexos e conjuntos compactos em conjuntos que o não sejam, etc.

Outra consequência interessante do Teorema 3.7 é a que se exprime no seguinte:

Teorema 3.9. *Seja X um subconjunto de $\mathbb{R}^n$; se, para qualquer par $(\boldsymbol{a}, \boldsymbol{b})$ de pontos de X existe uma função contínua $\varphi : [0,\, 1] \to \mathbb{R}^n$ verificando as condições: $\varphi(0) = \boldsymbol{a}, \varphi(1) = \boldsymbol{b}$ e $\varphi(t) \in X$ para todo o $t \in [0,\, 1]$, então o conjunto X é conexo.*

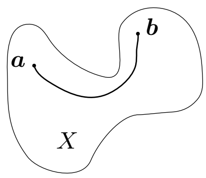

Figura 3.3

A condição referida no enunciado pode ser expressa, de forma mais sugestiva, dizendo que quaisquer pontos $\boldsymbol{a}, \boldsymbol{b} \in X$ podem ser «unidos por uma curva contida em X». Para a demonstração observe-se que, se X não fosse conexo, existiriam conjuntos separados A^* e B^* tais que $A^* \cup B^* = X$. Escolhidos arbitrariamente dois pontos $\boldsymbol{a} \in A^*$ e $\boldsymbol{b} \in B^*$, existiria também, por hipótese, uma função contínua $\varphi : [0,\, 1] \to \mathbb{R}^n$ tal que $\varphi(0) = \boldsymbol{a}, \varphi(1) = \boldsymbol{b}$ e $\varphi([0,\, 1]) \subset X$. Nestas condições, pondo:

$$A = \varphi([0,\, 1]) \cap A^* \quad \text{e} \quad B = \varphi([0,\, 1]) \cap B^*$$

ter-se-ia:

$$\varphi([0,\, 1]) = A \cup B$$

(visto que $\varphi([0,\, 1]) \subset A^* \cup B^* = X$), sendo A e B conjuntos separados (para o reconhecer, basta observar que A e B são não vazios — porque $\boldsymbol{a} \in A$ e $\boldsymbol{b} \in B$ — e ter em conta as relações $A \subset A^*$, $B \subset B^*$ e o facto de A^* e B^* serem conjuntos separados). Assim, concluir-se ia que o conjunto $\varphi([0,\, 1])$ era desconexo o que é absurdo, porque o intervalo $[0,\, 1]$ é conexo e a função φ é contínua.

Costuma-se chamar *conjuntos conexos por arcos* aos conjuntos $X \subset \mathbb{R}^n$ que verificam a condição mencionada no enunciado do Teorema 3.9 (isto é, tais que dois pontos quaisquer $\boldsymbol{a}, \boldsymbol{b} \in X$ podem sempre ser unidos por uma curva contida em X). Nestes termos, o enunciado desse teorema poderia sintetizar-se dizendo que qualquer conjunto conexo por arcos é conexo.

Observe-se, de passagem, que a recíproca é falsa (por exemplo, pode provar-se que o subconjunto X de $\mathbb{R}^2$ formado pelo gráfico da função $\operatorname{sen} 1/x$ ampliado com a origem é conexo, mas que não existe qualquer curva contida em X unindo a origem a outro ponto qualquer do mesmo conjunto).

Observaremos ainda que, na sua generalidade, os resultados obtidos neste parágrafo são válidos em espaços muito mais gerais do que os espaços $\mathbb{R}^n$ — por exemplo, em espaços métricos — sendo as demonstrações praticamente idênticas às que aqui foram feitas (haverá contudo nalguns pontos necessidade de certas «adaptações»: assim, por exemplo, no caso dos Teoremas 3.3, 3.4 e 3.5, haverá que ter em conta que, no quadro geral dos espaços métricos, a noção de conjunto compacto não equivale à de conjunto limitado e fechado).

No entanto, pareceu preferível — mesmo para quem tencione vir a desenvolver bastante os seus estudos no domínio da Análise — que a primeira abordagem destas ideias (para além do estudo das funções reais de variável real) se processasse no quadro particularmente importante e sugestivo facultado pelos espaços $\mathbb{R}^n$. Julga-se assim ter evitado um tratamento demasiado abstracto, cuja profundidade e alcance dificilmente poderiam ser apreendidos neste momento, até por impossibilidade de motivação adequada; e pensa-se também que, ultrapassada esta fase, ficará bastante facilitado o acesso aos pontos de vista mais elevados que alguns leitores decerto desejarão vir a alcançar neste domínio.

3.2 Limite

A noção de limite está muito intimamente relacionada com a de continuidade; em muitos textos, o estudo do conceito de limite precede o das funções contínuas ou é feito a par e passo com o das primeiras propriedades destas funções. Julgou-se contudo preferível estudar em primeiro lugar as propriedades essenciais das funções contínuas, sem qualquer referência à noção de limite, que é talvez um pouco mais elaborada; o estudo dos limites ficará agora muito facilitado e surgirá de modo natural, imediatamente antes do capítulo em que pela primeira vez eles irão ser necessários: a introdução ao cálculo diferencial em $\mathbb{R}^n$.

Para introduzir mais simplesmente a noção de limite, consideraremos em primeiro lugar o caso das funções reais; veremos depois que a extensão às funções vectoriais não oferece a menor dificuldade.

Antes de dar as definições formais, faremos ainda algumas considerações preparatórias. Neste sentido, recorde-se que, sendo f uma função real definida num conjunto $D \subset \mathbb{R}^n$ e $\boldsymbol{a}$ um ponto de D, dizemos que f é contínua no ponto $\boldsymbol{a}$ sse para todo o $\delta > 0$ existe $\epsilon > 0$ tal que

$$\boldsymbol{x} \in B_\epsilon(\boldsymbol{a}) \cap D \Longrightarrow |f(\boldsymbol{x}) - f(\boldsymbol{a})| < \delta.$$

De acordo com as definições que enunciaremos adiante, o facto de esta condição ser verificada poderá também traduzir-se dizendo que «$f(\boldsymbol{x})$ tende para $f(\boldsymbol{a})$ quando $\boldsymbol{x}$ tende para $\boldsymbol{a}$» ou que «$f(\boldsymbol{a})$ é o limite de $f(\boldsymbol{x})$ quando $\boldsymbol{x}$ tende para $\boldsymbol{a}$», e escrevendo[3]:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a}).$$

Suponhamos agora que $\boldsymbol{a}$ é um ponto aderente ao domínio D da função f, não pertencente a esse domínio (neste caso, $\boldsymbol{a}$ será necessariamente ponto de acumulação de D). Não existe então valor da função f no ponto $\boldsymbol{a}$ e f não pode evidentemente ser contínua nesse ponto; mas pode acontecer que exista um número $b \in \mathbb{R}$ com o qual — no lugar de $f(\boldsymbol{a})$ — seja verificada a condição atrás indicada. Se existir de facto $b \in \mathbb{R}$ nessas condições — isto é, tal que, qualquer que seja $\delta > 0$ exista $\epsilon > 0$ por forma que todo o $\boldsymbol{x} \in D$ que satisfaça a condição $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$ verifique também $|f(\boldsymbol{x}) - b| < \delta$ — diremos ainda que $f(\boldsymbol{x})$ tende para b quando $\boldsymbol{x}$ tende para $\boldsymbol{a}$ e escreveremos:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = b.$$

No caso que estamos a considerar ($\boldsymbol{a} \in \bar{D} \backslash D$) é fácil ver que a existência de limite equivale à possibilidade de «prolongar por continuidade a função f ao ponto $\boldsymbol{a}$», isto é, equivale à existência de uma função $\tilde{f}$ — chamada prolongamento por continuidade de f ao ponto $\boldsymbol{a}$ — definida em $D \cup \{\boldsymbol{a}\}$, contínua em $\boldsymbol{a}$ e tal que

[3] Esta notação e também o artigo definido incluido na última das afirmações precedentes só ficarão inteiramente justificados quando se tiver reconhecido a unicidade do limite.

$\tilde{f}_{/D} = f$. Na realidade, se for $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = b$ ver-se-á sem dificuldade que a única função f que satisfaz as condições acabadas de indicar é a função $\tilde{f} : D\cup\{\boldsymbol{a}\} \to \mathbb{R}$ tal que:

$$\tilde{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{se } \boldsymbol{x} \in D \\ b & \text{se } \boldsymbol{x} = \boldsymbol{a}. \end{cases}$$

A título de exemplo, consideremos a função definida pela fórmula:

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

no conjunto $D = \mathbb{R}^2\backslash\{(0, 0)\}$.

Resultados obtidos no parágrafo precedente permitem reconhecer imediatamente que f é contínua em todo o seu domínio; quando (x, y) tender para um ponto qualquer $(a, b) \in D$, a função tenderá portanto para um limite, igual ao seu valor no ponto considerado. Quanto ao ponto $(0, 0)$, é claro que f não é contínua nesse ponto, não existindo sequer o valor $f(0, 0)$. Tem-se, contudo, como vamos ver (em notação de significado evidente):

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

Com efeito, sendo $(x, y) \neq (0, 0)$, tem-se:

$$|f(x, y)| = \frac{|x^2 - y^2|}{\sqrt{x^2 + y^2}} \leq \frac{x^2 + y^2}{\sqrt{x^2 + y^2}} = \sqrt{x^2 + y^2}$$

e portanto a condição $|f(x, y)| < \delta$ será verificada por todo o ponto $(x, y) \in D$ tal que $\|(x, y)\| = \sqrt{x^2 + y^2} < \delta$.

Fica assim provado que $f(x, y)$ tende para 0 quando $(x, y) \to (0, 0)$ e é claro que para prolongar por continuidade a função f à origem bastaria «atribuir-lhe» nesse ponto o valor 0 (esta frase é incorrecta: o prolongamento por continuidade é uma função *distinta* da função f, visto que não tem o mesmo domínio).

Consideremos agora a função φ, definida no mesmo conjunto D do exemplo anterior, pela fórmula:

$$\varphi(x, y) = \frac{x^2 - y^2}{x^2 + y^2}$$

Esta função também é contínua em qualquer ponto do seu domínio e, para $(\alpha, \beta) \neq (0, 0)$, tem-se:

$$\lim_{(x,y)\to(\alpha,\beta)} \varphi(x, y) = \frac{\alpha^2 - \beta^2}{\alpha^2 + \beta^2}.$$

Para averiguar da existência de limite na origem observemos primeiramente que, sendo x e y números reais diferentes de zero, se tem:

$$\varphi(x, 0) = 1 \quad \text{e} \quad \varphi(0, y) = -1$$

o que mostra que a restrição de φ ao «eixo das abcissas privado da origem» (isto é, ao conjunto de todos os pontos $(x, 0)$, com $x \neq 0$) é a função identicamente igual a 1 e que a restrição de φ ao «eixo das ordenadas privado da origem» é a função que toma o valor -1 em qualquer ponto deste conjunto. Esta observação torna evidente que em qualquer bola centrada na origem, por menor que seja o seu raio, haverá sempre pontos em que φ toma o valor 1 e pontos em que φ toma o valor -1 (aliás infinitos, num caso e no outro; na bola de raio ϵ, os pontos $(\epsilon/2, 0)$ e $(0, \epsilon/2)$ podem servir de exemplo de cada um desses casos).

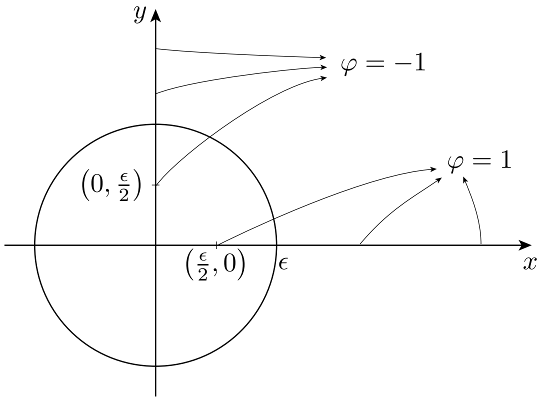

Figura 3.4

Daqui decorre facilmente a impossibilidade da existência de limite. Com efeito, se para algum número real b fosse verdadeira a proposição:

$$\lim_{(x,y)\to(0,0)} \varphi(x, y) = b,$$

fixado δ, por exemplo, no valor 1, deveria existir $\epsilon > 0$ tal que, para qualquer (x, y) pertencente a D e à bola centrada na origem e com raio ϵ, se teria:

$$|\varphi(x, y) - b| < 1.$$

Porém, escolhido um ϵ nessas condições, ter-se-ia:

$$\begin{aligned} 2 &= \left|\varphi\left(\frac{\epsilon}{2}, 0\right) - \varphi\left(0, \frac{\epsilon}{2}\right)\right| = \left|\left[\varphi\left(\frac{\epsilon}{2}, 0\right) - b\right] + \left[b - \varphi\left(0, \frac{\epsilon}{2}\right)\right]\right| \\ &\leq \left|\varphi\left(\frac{\epsilon}{2}, 0\right) - b\right| + \left|\varphi\left(0, \frac{\epsilon}{2}\right) - b\right| < 1 + 1 = 2, \end{aligned}$$

o que é absurdo.

Assim, a função φ não tem limite no ponto $(0, 0)$; e imediatamente se reconhece também que, seja qual for o valor real de b, o «prolongamento» $\tilde{\varphi}$ de φ definido por:

$$\tilde{\varphi}(x, y) = \begin{cases} \varphi(x, y) & \text{se } (x, y) \in D \\ b & \text{se } (x, y) = (0, 0), \end{cases}$$

nunca será contínuo na origem.

Feitas estas considerações, introduziremos agora formalmente a definição de limite, que neste momento já deve ser óbvia:

Seja f uma função real definida num subconjunto D de $\mathbb{R}^n$, e sejam $\boldsymbol{a} = (a_1, \ldots, a_n)$ um ponto aderente a D e b um número real. Diz-se que $f(\boldsymbol{x})$ *tende para b quando $\boldsymbol{x} \to \boldsymbol{a}$* sse para todo o $\delta > 0$ existe $\epsilon > 0$ tal que, sempre que $\boldsymbol{x} \in \mathbb{R}^n$ verifique as condições $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$, se tenha $|f(\boldsymbol{x}) - b| < \delta$.

Convém observar que, em alguns textos, a definição de limite adoptada não é equivalente à que acaba de ser enunciada (tanto no caso, agora considerado, das funções reais de n variáveis reais como no das funções vectoriais que veremos dentro em pouco, e tanto para $n > 1$ como para $n = 1$).

As diferenças entre as duas definições são as seguintes:

1. Em vez de se exigir, como aqui fizemos, que a condição $|f(\boldsymbol{x}) - b| < \delta$ seja verificada sempre que se tenha $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$, impõe-se que essa mesma desigualdade seja satisfeita por todos os pontos $\boldsymbol{x} \in D$ tais que $0 < \|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$. É óbvio que, no caso de $\boldsymbol{a}$ não pertencer a D, as duas definições conduzem exactamente aos mesmos resultados; mas se for $\boldsymbol{a} \in D$, o valor de f em $\boldsymbol{a}$ será «ignorado» na definição aqui não adoptada — sendo o limite, se existir, inteiramente independente de $f(\boldsymbol{a})$ — enquanto pela definição que usaremos neste texto a existência do limite de f num ponto $\boldsymbol{a}$ do seu domínio impõe que se verifique necessariamente a igualdade:
$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a}),$$
e equivale assim à continuidade de f no ponto $\boldsymbol{a}$.

2. Em vez de se exigir que $\boldsymbol{a}$ seja um ponto aderente ao domínio D da função f, como aqui foi feito, impõe-se que $\boldsymbol{a}$ seja ponto de acumulação do mesmo conjunto (deixando-se, assim, de considerar limites em pontos isolados do conjunto D). Esta modificação está inteiramente relacionada com a anterior e, por assim dizer, decorre dela: é fácil ver que o que se impõe ao ponto $\boldsymbol{a}$, em cada caso, é precisamente o que importa para se poder garantir a unicidade do limite.

A diferença entre os dois pontos de vista não é importante, em termos conceptuais; trata-se mais de um «pormenor de ordem técnica», ao qual nos referimos apenas para prevenir o leitor e evitar-lhe eventuais dúvidas e perdas de tempo. Pensamos que a definição que decidimos adoptar — mesmo que obrigue alguns leitores a um pequeno esforço de adaptação que, nesta fase do seu estudo, não poderá já comportar qualquer dificuldade séria — permite organizar de forma mais natural e harmoniosa alguns aspectos da teoria, e por isso a preferimos. De resto, como poderemos ver na sequência, o conceito correspondente à outra definição surgirá aqui também, como caso particular da noção mais geral de limite relativo a um subconjunto A do domínio D da função considerada (precisamente no caso de ser $A = D \setminus \{\boldsymbol{a}\}$).

Consideremos agora o caso mais geral das funções vectoriais. Naturalmente, sendo $f : D \to \mathbb{R}^m$ (com $D \subset \mathbb{R}^n$), $\boldsymbol{a}$ um ponto aderente a D e sendo agora $\boldsymbol{b} = (b_1, \dots, b_m)$ um vector qualquer de $\mathbb{R}^m$, diremos que $f(\boldsymbol{x})$ tende para $\boldsymbol{b}$ quando $\boldsymbol{x}$ tende para $\boldsymbol{a}$ sse qualquer que seja $\delta > 0$ existir $\epsilon > 0$ tal que, para todo o ponto $\boldsymbol{x} \in \mathbb{R}^n$ que verifique as condições $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$, se tenha $\|f(\boldsymbol{x}) - \boldsymbol{b}\| < \delta$.

Prova-se sem qualquer dificuldade (e será consequência imediata de resultados posteriores) que, se $f(\boldsymbol{x})$ tende para $\boldsymbol{b}$ e também para $\boldsymbol{b}'$ quando $\boldsymbol{x}$ tende para $\boldsymbol{a}$, então é necessariamente $\boldsymbol{b} = \boldsymbol{b}'$. Nesta hipótese, o (único) vector $\boldsymbol{b}$ que verifica esta condição é designado por *limite de $f(\boldsymbol{x})$ quando $\boldsymbol{x}$ tende para $\boldsymbol{a}$ ou limite de f no ponto $\boldsymbol{a}$*, podendo escrever-se:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = \boldsymbol{b}$$

ou

$$\lim_{(x_1,\dots,x_n) \to (a_1,\dots,a_n)} f(x_1, \dots, x_n) = (b_1, \dots, b_m).$$

ou ainda, mais simplesmente:

$$\lim_{\boldsymbol{a}} f = \boldsymbol{b}.$$

Vê-se também sem a menor dificuldade (tendo em conta as definições de continuidade e de limite) que a *existência do limite de f no ponto $\boldsymbol{a}$* equivale à existência de um *prolongamento por continuidade de f ao ponto $\boldsymbol{a}$*, isto é, de uma função $\tilde{f} : D \cup \{\boldsymbol{a}\} \to \mathbb{R}^m$ contínua no ponto $\boldsymbol{a}$ e tal que $\tilde{f}_{/D} = f$. No caso em que $\boldsymbol{a}$ (sempre aderente a D) não pertence a D, esse prolongamento é definido por:

$$\tilde{f}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{se } \boldsymbol{x} \in D \\ \lim_{\boldsymbol{a}} f & \text{se } \boldsymbol{x} = \boldsymbol{a}. \end{cases}$$

No caso em que $\boldsymbol{a} \in D$, tem-se evidentemente $D \cup \{\boldsymbol{a}\} = D$ e o prolongamento $\tilde{f}$ coincide com a própria função f (a qual, por existir o limite, é então necessariamente contínua no ponto $\boldsymbol{a}$). Tanto num caso como no outro, é óbvio que (fixado o ponto $\boldsymbol{a} \in D$) o prolongamento $\tilde{f}$ é univocamente determinado pela função f.

Pode ver-se ainda que, se $\boldsymbol{a} \in \mathbb{R}^n$ é um ponto exterior ao domínio D de f (caso excluído na definição de limite) existem sempre infinitas funções $\tilde{f} : D \cup \{\boldsymbol{a}\} \to \mathbb{R}^m$, contínuas em $\boldsymbol{a}$ e tais que $\tilde{f}_{/D} = f$: para obter uma tal função bastaria pôr:

$$\tilde{f}(x) = \begin{cases} f(\boldsymbol{x}) & \text{se } \boldsymbol{x} \in D \\ \boldsymbol{c} & \text{se } \boldsymbol{x} = \boldsymbol{a}, \end{cases}$$

onde $\boldsymbol{c}$ designa um vector arbitrário de $\mathbb{R}^m$.

Assim, para os pontos não aderentes ao domínio de f, haveria sempre possibilidade de «prolongar continuamente» a função, mas o prolongamento, não sendo univocamente determinado, ficaria totalmente desprovido de interesse. É por uma razão semelhante que, na definição de limite, apenas considerámos pontos aderentes ao domínio da função. Com a definição adoptada, o limite (quando existe) é único e a sua existência equivale à de um único prolongamento da função f definido em $D \cup \{\boldsymbol{a}\}$ e contínuo no ponto $\boldsymbol{a}$.

Registaremos agora algumas propriedades da noção de limite, em correspondência com propriedades da continuidade estudadas no parágrafo precedente; as demonstrações, que omitiremos, podem fazer-se de modo análogo ao adoptado no caso da continuidade, ou então reduzir-se a esse caso, como se sugere a propósito do seguinte:

Teorema 3.1'. *Seja $f : D \to \mathbb{R}^m$ ($D \subset \mathbb{R}^n$), $\boldsymbol{a}$ um ponto aderente a D e $\boldsymbol{b}$ um vector de $\mathbb{R}^m$; para que se verifique a igualdade: $\lim_{\boldsymbol{a}} f = \boldsymbol{b}$ é necessário e suficiente que, sempre que $\boldsymbol{x}_k$ seja uma sucessão em D convergente para $\boldsymbol{a}$, a sucessão $f(\boldsymbol{x}_k)$ convirja para $\boldsymbol{b}$.*

A demonstração pode fazer-se de forma quase idêntica à do Teorema 3.1; mas pode também pensar-se que, para que a igualdade $\lim_{\boldsymbol{a}} f = \boldsymbol{b}$ seja verificada, é necessário e suficiente, no caso de ser $\boldsymbol{a} \in D$, que f seja contínua em $\boldsymbol{a}$ e se tenha $f(\boldsymbol{a}) = \boldsymbol{b}$; e no caso de ser $\boldsymbol{a} \in \bar{D} \backslash D$, que seja contínua no ponto $\boldsymbol{a}$ a função $\tilde{f} : D \cup \{\boldsymbol{a}\} \to \mathbb{R}$, que prolonga f e assume no ponto $\boldsymbol{a}$ o valor $\boldsymbol{b}$; assim, a questão do limite fica reduzida à da continuidade e o recurso ao Teorema 3.1 permite completar imediatamente a demonstração.

Nas propriedades seguintes, que nos limitaremos a enunciar, deve supor-se que $D \subset \mathbb{R}^n$, $\boldsymbol{a} \in \bar{D}$; $f, g : D \to \mathbb{R}^m$ e $\alpha : D \to \mathbb{R}$.

- *Se f é constante em D, existe $\lim_{\boldsymbol{a}} f$ e é igual ao valor de f num ponto qualquer de D.*

- *Se f e g têm limite no ponto $\boldsymbol{a}$, também o têm as funções $f+g$, $f-g$, $f \cdot g$ e $\|f\|$, verificando-se as igualdades:*

$$\lim_{\boldsymbol{a}}(f+g) = \lim_{\boldsymbol{a}} f + \lim_{\boldsymbol{a}} g,$$
$$\lim_{\boldsymbol{a}}(f-g) = \lim_{\boldsymbol{a}} f - \lim_{\boldsymbol{a}} g,$$
$$\lim_{\boldsymbol{a}}(f \cdot g) = \lim_{\boldsymbol{a}} f \cdot \lim_{\boldsymbol{a}} g$$

  e

$$\lim_{\boldsymbol{a}} \|f\| = \|\lim_{\boldsymbol{a}} f\|.$$

- *Se α e f têm limite no ponto $\boldsymbol{a}$, αf também e tem-se:*

$$\lim_{\boldsymbol{a}}(\alpha f) = (\lim_{\boldsymbol{a}} \alpha)(\lim_{\boldsymbol{a}} f);$$

se for ainda $\lim_{\boldsymbol{a}} \alpha \neq 0$, *o cociente* f/α *terá limite quando* $\boldsymbol{x} \to \boldsymbol{a}$, *verificando-se a igualdade:*

$$\lim_{\boldsymbol{a}} \frac{f}{\alpha} = \frac{\lim_{\boldsymbol{a}} f}{\lim_{\boldsymbol{a}} \alpha}.$$

Também no caso do limite o estudo das funções vectoriais pode reduzir-se imediatamente ao das funções reais, nos termos do seguinte:

Teorema 3.2'. *Seja* $f : D \to \mathbb{R}^m$ *(com* $D \subset \mathbb{R}^n$*),* $\boldsymbol{a}$ *um ponto aderente a* D, $\boldsymbol{b} = (b_1, \dots, b_m)$ *um vector de* $\mathbb{R}^m$ *e designemos por* f_j *a função coordenada de ordem* j *de* f*; nestas condições, para que se verifique a igualdade* $\lim_{\boldsymbol{a}} f = \boldsymbol{b}$ *é necessário e suficiente que, para cada inteiro positivo* $j \leq m$*, se tenha* $\lim_{\boldsymbol{a}} f_j = b_j$.

Tem também interesse o seguinte resultado, que relaciona da forma desejável a noção de limite com a composição de funções:

Seja $D \subset \mathbb{R}^n$, $E \subset \mathbb{R}^p$, $g : D \to E$ e $f : E \to \mathbb{R}^m$; suponha-se ainda que $\boldsymbol{a}$ é um ponto aderente ao conjunto D. Nestas condições, vê-se imediatamente que, se se tiver $\lim_{\boldsymbol{a}} g = \boldsymbol{b}$, $\boldsymbol{b}$ será necessariamente um ponto aderente ao conjunto E; e também (usando, por exemplo, o Teorema 3.1') que, se existir ainda o limite de f no ponto $\boldsymbol{b}$, existirá também o limite no ponto $\boldsymbol{a}$ da função composta $f \circ g : D \to \mathbb{R}^n$, verificando-se a igualdade:

$$\lim_{\boldsymbol{a}} (f \circ g) = \lim_{\boldsymbol{b}} f.$$

ou, com outra notação:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} (f \circ g)(\boldsymbol{x}) = \lim_{\boldsymbol{y} \to \boldsymbol{b}} f(\boldsymbol{y}).$$

Assim, por exemplo, das igualdades:

$$\lim_{(x,y) \to (0,0)} \frac{x^2 - y^2}{\sqrt{x^2 + y^2}} = 0$$

e

$$\lim_{u \to 0} \cos u = 1,$$

poderá imediatamente deduzir-se que:

$$\lim_{(x,y) \to (0,0)} \cos \frac{x^2 - y^2}{\sqrt{x^2 + y^2}} = 1.$$

Antes de indicarmos algumas outras aplicações do resultado anterior, convém introduzir uma definição:

Nas hipóteses já habituais de D ser um subconjunto de $\mathbb{R}^n$, $\boldsymbol{a}$ um ponto aderente a D e f uma aplicação de D em $\mathbb{R}^m$, consideremos agora um subconjunto A de D ao qual o ponto $\boldsymbol{a}$ seja ainda aderente; $\boldsymbol{a}$ será portanto um ponto aderente ao domínio da função $f_{/A}$, podendo existir ou não o $\lim_{\boldsymbol{a}} f_{/A}$. Quando este limite exista, diremos que a função f tem *limite no ponto* $\boldsymbol{a}$ *relativo ao conjunto* A (o qual será designado por $\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in A}} f(\boldsymbol{x})$) e poremos, por definição:

$$\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in A}} f(\boldsymbol{x}) = \lim_{\boldsymbol{x}\to\boldsymbol{a}} f_{/A}(\boldsymbol{x}).$$

Por exemplo, no caso da função

$$\varphi(x,y) = \frac{x^2-y^2}{x^2+y^2} \qquad (x,y)\in\mathbb{R}^2\backslash\{(0,0)\}$$

ter-se-á, pondo:

$$A = \{(x,0) : x\in\mathbb{R}\backslash\{0\}\} \quad \text{e} \quad B = \{(0,y) : y\in\mathbb{R}\backslash\{0\}\},$$

$$\lim_{\substack{(x,y)\to(0,0)\\ (x,y)\in A}} \varphi(x,y) = 1 \quad \text{e} \quad \lim_{\substack{(x,y)\to(0,0)\\ (x,y)\in B}} \varphi(x,y) = -1.$$

Voltando ao caso geral considerado na definição de limite relativo a um conjunto (e às notações aí adoptadas) designemos agora por g a aplicação de A em D definida por $g(\boldsymbol{x}) = \boldsymbol{x}$, para todo o $\boldsymbol{x}\in A$ (aplicação a que costuma chamar-se *injecção canónica de* A *em* D); ter-se-á então, obviamente:

$$f_{/A} = f\circ g$$

e também $\lim_{\boldsymbol{x}\to\boldsymbol{a}} g(\boldsymbol{x}) = \boldsymbol{a}$. Nestas condições, a relação entre o limite e a composição de funções expressa num resultado precedente permite concluir imediatamente que, se existir o limite de f no ponto $\boldsymbol{a}$, existirá necessariamente — e com o mesmo valor — o limite no ponto $\boldsymbol{a}$ da função $f_{/A}$, isto é, o $\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in A}} f(\boldsymbol{x})$. Assim:

Se existe $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})$*, existe também o limite de* f *no ponto* $\boldsymbol{a}$ *relativo a qualquer conjunto* $A\subset D$ *tal que* $\boldsymbol{a}\in\bar{A}$ *e tem-se necessariamente:*

$$\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in A}} f(\boldsymbol{x}) = \lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}).$$

É este o fundamento de uma técnica corrente para a prova da não existência de determinados limites: sempre que seja possível determinar conjuntos A, $B\subset D$ (com $\boldsymbol{a}\in\bar{A}$ e $\boldsymbol{a}\in\bar{B}$) para os quais se tenha:

$$\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in A}} f(\boldsymbol{x}) \neq \lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in B}} f(\boldsymbol{x})$$

(ou então um só conjunto A, nas mesmas condições, tal que não exista o limite de f no ponto $\boldsymbol{a}$ relativo a A) poderá concluir-se que não existe $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x})$. Assim, a observação feita há pouco sobre os limites relativos aos «eixos coordenados privados da origem» para a função φ permitiria agora concluir com grande facilidade a não existência do limite de φ no ponto $(0,0)$, já atrás reconhecida com mais algum trabalho.

Antes de passar a outros exemplos mencionaremos que, como já foi assinalado, a definição de limite de f num ponto $\boldsymbol{a}$ anteriormente referida como não adoptada neste texto, é um caso particular da de limite relativo a um conjunto; com efeito, vê-se imediatamente que o limite considerado nessa definição se identifica com o que designaríamos agora por

$$\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in D\setminus\{\boldsymbol{a}\}}} f(\boldsymbol{x})$$

e também que, para que este limite possa ser considerado, deverá o ponto $\boldsymbol{a}$ ser aderente ao conjunto $D\setminus\{\boldsymbol{a}\}$, o que equivale a dizer que deverá ser ponto de acumulação do conjunto D.

Vejamos outro exemplo de aplicação da técnica, há pouco referida, utilizável para provar a não existência de limites; seja ψ a função definida (em $D = \mathbb{R}^2\setminus\{(0,0)\}$) pela fórmula:

$$\psi(x,y) = \frac{xy}{x^2+y^2}$$

e designemos por A_m o subconjunto de D formado por todos os pontos da recta de equação $y = mx$ com excepção da origem:

$$A_m = \{(x,mx) : x \neq 0\}.$$

A igualdade (válida para qualquer $x \neq 0$):

$$\psi(x,mx) = \frac{m}{1+m^2}$$

mostra que a função ψ é constante em qualquer dos conjuntos A_m, tendo-se portanto:

$$\lim_{\substack{(x,y)\to(0,0)\\ (x,y)\in A_m}} \psi(x,y) = \frac{m}{1+m^2};$$

do facto deste limite variar com m, deduz-se imediatamente que ψ não tem limite na origem.

A consideração de rectas correntes pela origem — ou, de semirectas com origem nesse ponto — é uma técnica usual, quando se pretende averiguar da eventual não existência do limite de uma função $f(x,y)$ no ponto $(0,0)$. O processo é, aliás, aplicável ao estudo de limites num ponto qualquer (a,b), caso em que podem usar-se rectas passando por este ponto ou semirectas com origem nele (e pode-se também, se se preferir, começar por «transferir o limite para a origem», através da composição de $f(x,y)$ com $x = a+u$, $y = b+v$, reconhecendo-se imediatamente que qualquer dos limites:

$$\lim_{(x,y)\to(a,b)} f(x,y) \quad \text{e} \quad \lim_{(u,v)\to(0,0)} f(a+u,b+v)$$

existe sse o outro existir e que, na hipótese de existência, têm o mesmo valor).

A utilização de técnicas deste tipo no caso de funções de n variáveis reais requer algumas ideias muito simples sobre Geometria Analítica em $\mathbb{R}^n$, a que vamos fazer uma rápida referência.

Sendo $\boldsymbol{a} = (a_1, \ldots, a_n)$ um ponto de $\mathbb{R}^n$ e $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$ um vector não nulo, a recta que passa por $\boldsymbol{a}$ e tem a direcção do vector $\boldsymbol{v}$ é, por definição, o conjunto de todos os pontos $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ representáveis na forma:

$$\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v},$$

com $t \in \mathbb{R}$ (na interpretação geométrica, válida para $n \leq 3$, esta equação representa de facto uma recta, satisfazendo as condições indicadas). A equação $\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v}$ é chamada equação paramétrica da recta considerada, na forma vectorial; e as equações correspondentes, em termos de coordenadas:

$$\begin{aligned} x_1 &= a_1 + tv_1 \\ &\cdots \\ x_n &= a_n + tv_n, \end{aligned}$$

constituem o sistema de equações paramétricas da mesma recta, na forma escalar (Figura 3.5).

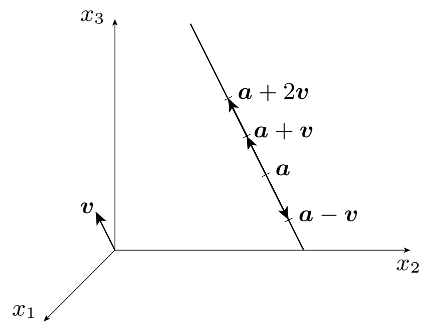

Figura 3.5

Se, em vez de supormos que o parâmetro t assume todos os valores reais, admitirmos que varia num intervalo limitado de $\mathbb{R}$, I, ao conjunto de todos os pontos

$$\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v} \quad (t \in I)$$

chamaremos um segmento de recta (se for $I = [t_1, t_2]$, com $t_1 < t_2$, os pontos $\boldsymbol{a} + t_1\boldsymbol{v}$ e $\boldsymbol{a} + t_2\boldsymbol{v}$ serão os extremos do segmento).

Analogamente, a semirecta (aberta) de origem no ponto $\boldsymbol{a}$ e com a direcção e o sentido do vector $\boldsymbol{v}$ será o conjunto definido pela equação $\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v}$, com $t \in]0, +\infty[$ ($t \in [0, +\infty[$ para a semirecta fechada), etc. Na sequência, designaremos a semirecta definida pela equação:

$$\boldsymbol{x} = \boldsymbol{a} + t\boldsymbol{v},$$

com $t > 0$, pelo símbolo $S_{\boldsymbol{a},\boldsymbol{v}}$ ou, quando o ponto $\boldsymbol{a}$ estiver claramente fixado, apenas por $S_{\boldsymbol{v}}$.

Consideremos agora uma função real f (o caso de uma função vectorial reduzir-se-ia a este por passagem às funções coordenadas), a qual, por razões de comodidade, suporemos definida em todo o conjunto $\mathbb{R}^n$, com eventual excepção de um dado ponto, $\boldsymbol{a}$ (no entanto, tornar-se-á evidente que não haveria qualquer alteração essencial ao que vai seguir-se se admitíssemos, mais geralmente, que f estava definida num conjunto $D \subset \mathbb{R}^n$ tal que, para algum $\epsilon > 0$, se verificasse a relação $B_\epsilon(\boldsymbol{a})\backslash\{\boldsymbol{a}\} \subset D$). Sendo $\boldsymbol{v}$ um vector não nulo, ao limite de f no ponto $\boldsymbol{a}$ relativo ao conjunto $S_{\boldsymbol{v}}$ costuma também chamar-se *limite direccional de f no ponto $\boldsymbol{a}$ segundo o vector $\boldsymbol{v}$* (ou *na direcção e sentido de $\boldsymbol{v}$*). Reconhece-se facilmente que este limite existe sse a função $\varphi_{\boldsymbol{v}} :]0, +\infty[\to \mathbb{R}$ definida pela fórmula:

$$\varphi_{\boldsymbol{v}}(t) = f(\boldsymbol{a} + t\boldsymbol{v})$$

tiver limite quando $t \to 0^+$, verificando-se nessa hipótese a igualdade:

$$\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in S_{\boldsymbol{v}}}} f(\boldsymbol{x}) = \lim_{t\to 0^+} \varphi_{\boldsymbol{v}}(t).$$

Assim, o cálculo de um limite direccional (ou a verificação da sua não existência) reduz-se ao estudo de um problema de limites para uma função de uma só variável real.

É claro que, se existir o limite de f no ponto considerado, existirá também — e com o mesmo valor — o limite direccional segundo qualquer vector $\boldsymbol{v} \neq \boldsymbol{0}$; portanto, se for possível encontrar dois vectores $\boldsymbol{v}_1, \boldsymbol{v}_2$ aos quais correspondam limites direccionais diferentes, poderá concluir-se que a função não tem limite no ponto considerado.

Poderá também concluir-se, em sentido inverso, que se existirem os limites direccionais relativos a todos os vectores (não nulos) $\boldsymbol{v} \in \mathbb{R}^n$ e se todos esses limites direccionais tiverem o mesmo valor, f tem limite no ponto considerado?

Veremos facilmente que a resposta a esta questão deverá ser negativa, se notarmos que, no estudo de cada um dos limites direccionais, os únicos valores de f que se consideram são os que a função assume sobre uma determinada semirecta aberta com origem no ponto $\boldsymbol{a}$; assim, se f estiver definida neste ponto, o valor $f(\boldsymbol{a})$ será «ignorado» na pesquisa de todos os limites direccionais e é óbvio que estes limites poderão existir e ser todos iguais sem que f tenha limite no ponto $\boldsymbol{a}$ (é o que se passa, por exemplo, com a função $f : \mathbb{R}^n \to \mathbb{R}$ que toma o valor 1 em dado ponto $\boldsymbol{a} \in \mathbb{R}^n$ e o valor 0 em todos os outros pontos).

Porém, o que poderá ser um pouco surpreendente é que a existência e igualdade de todos os limites direccionais no ponto $\boldsymbol{a}$ nem sequer garante a existência de limite para a restrição de f a $\mathbb{R}^n\backslash\{\boldsymbol{a}\}$, isto é, do

$$\lim_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\in\mathbb{R}^n\backslash\{\boldsymbol{a}\}}} f(\boldsymbol{x})$$

(o qual, na sequência, designaremos mais simplesmente pelo símbolo $\lim\limits_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\neq\boldsymbol{a}}} f(\boldsymbol{x})$).[4]

Para o mostrar, recorreremos a um exemplo simples, relativo ao caso $n = 2$ (é fácil — e poderá ficar como exercício — a adaptação desse exemplo por forma a provar que, também para $n > 2$, a existência e igualdade de todos os limites direccionais no ponto $\boldsymbol{a}$ não garante a existência de $\lim\limits_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\neq\boldsymbol{a}}} f(\boldsymbol{x})$).

Seja $f : \mathbb{R}^2 \to \mathbb{R}$ a função definida pela forma seguinte:

$$f(x, y) = \begin{cases} 1 & \text{se } x \neq 0 \text{ e } y = x^2 \\ 0 & \text{se } x = 0 \text{ ou } y \neq x^2 \end{cases}$$

Assim, f toma o valor 1 em todos os pontos da parábola de equação $y = x^2$ com excepção da origem e o valor 0 em todos os outros pontos do plano. Vê-se imediatamente que a restrição de f a qualquer semirecta (aberta) com origem em $(0, 0)$ assume o valor 0 em todos os pontos dessa semirecta excepto, quando muito, num ponto (aquele em que a semirecta em causa intersecta a parábola, nos casos em que tal intersecção não é vazia); e daí logo se deduz que todos os limites direccionais de f no ponto $(0, 0)$ são iguais a 0. No entanto, tanto a função como a sua restrição a $\mathbb{R}^2\backslash\{(0, 0)\}$ não podem ter o limite 0 — nem, evidentemente, qualquer outro — quando (x, y) tende para $(0, 0)$: basta notar que, em qualquer bola centrada na origem, há infinitos pontos em que f assume o valor 1 (Figura 3.6).

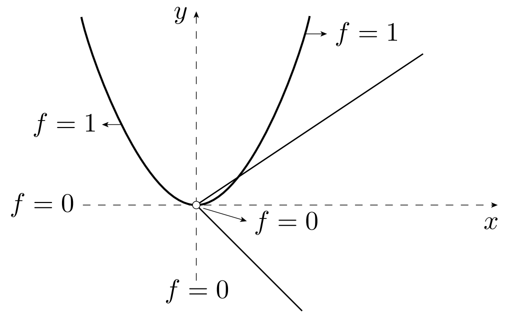

Figura 3.6

O que o exemplo precedente nos permitiu reconhecer pode também observar-se com funções definidas de forma «menos artificial»; para o verificar,

[4]Salvo no caso de ser $n = 1$ (isto é, de f ser uma função de uma variável real); em tal caso, vê-se facilmente que os limites direccionais se identificam com os limites laterais $f(a^+) = \lim_{x\to a^+} f(x)$ e $f(a^-) = \lim_{x\to a^-} f(x)$ (mais precisamente, o limite segundo o vector $k\boldsymbol{e}_1$ coincide com $f(a^+)$ se $k > 0$ e com $f(a^-)$ se $k < 0$) e é sabido que a existência e igualdade dos dois limites laterais assegura a existência de $\lim\limits_{\substack{\boldsymbol{x}\to\boldsymbol{a}\\ \boldsymbol{x}\neq\boldsymbol{a}}} f(x)$.

poderá por exemplo estudar-se, do mesmo ponto de vista, a função racional definida pela expressão:

$$\frac{x^2y}{x^4+y^2},$$

para a qual todos os limites direccionais na origem são nulos, não existindo também limite no mesmo ponto (considere-se, em particular, o limite relativo ao conjunto $\{(x,x^2) : x \neq 0\}$).

Faremos agora uma breve referência a um outro processo, por vezes muito útil para o cálculo de limites de funções de duas variáveis, ou para a verificação da sua não existência. O processo é correntemente designado por «passagem a coordenadas polares» (no caso de funções de três variáveis reais, poderá usar-se a «passagem a coordenadas esféricas» e, mesmo para $n > 3$, poderá recorrer-se de forma análoga à aplicação $\mu : \mathbb{R}^n \to \mathbb{R}^n$ mencionada no parágrafo 3.1).

Designemos por Σ o subconjunto de $\mathbb{R}^2$ formado por todos os pares (r,θ) que verificam a condição $r > 0$ e consideremos a aplicação μ de Σ em $\mathbb{R}^2$ que transforma cada ponto $(r,\theta) \in \Sigma$ no ponto (x,y) tal que:

$$\begin{cases} x = r\cos\theta \\ y = r\,\mathrm{sen}\,\theta. \end{cases}$$

Vê-se imediatamente que, qualquer que seja o número positivo ϵ, a recta de equação $r = \epsilon$ é transformada[5] por μ na circunferência de raio ϵ centrada na origem O do plano xOy; portanto, a faixa plana Σ_ϵ, constituída por todos os pontos (r,θ) tais que $0 < r < \epsilon$ será transformada na bola centrada em $O = (0,0)$, privada do próprio ponto O (Figura 3.7):

$$\mu(\Sigma_\epsilon) = B_\epsilon(O)\backslash\{O\}.$$

Consideremos agora uma função real $f(x,y)$ que, para maior simplicidade, suporemos definida em $\mathbb{R}^2\backslash\{O\}$ (seria imediata a adaptação ao caso de f estar definida num conjunto D tal que $O \in \bar{D}$). Pondo $F = f \circ \mu$, isto é:

$$F(r,\theta) = f(r\cos\theta, r\,\mathrm{sen}\,\theta)$$

para $r > 0$ e $\theta \in \mathbb{R}$, logo se vê que o conjunto dos valores que F assume em todos os pontos da faixa Σ_ϵ coincide com o conjunto dos valores assumidos por f em $B_\epsilon(O)\backslash\{O\}$:

$$F(\Sigma_\epsilon) = f\big(B_\epsilon(O)\backslash\{O\}\big).$$

Sendo assim, para que se verifique a igualdade:

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

[5]É óbvio que a aplicação μ não é injectiva; qualquer que seja o ponto $(r,\theta) \in \Sigma$ e o inteiro k tem-se $\mu(r,\theta+2k\pi) = \mu(r,\theta)$; porém, como é sabido, a restrição de μ ao conjunto formado pelos pontos (r,θ) tais que $r > 0$ e $0 \leq \theta < 2\pi$ aplica bijectivamente este conjunto em $\mathbb{R}^2\backslash\{(0,0)\}$.

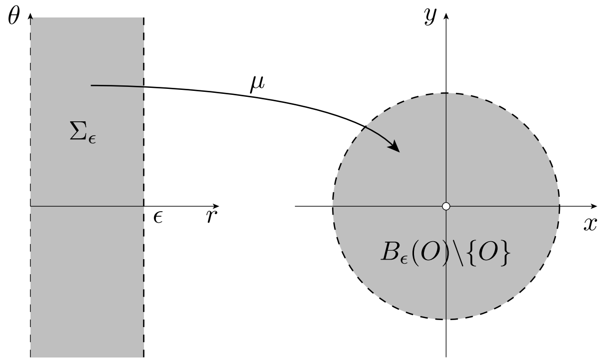

Figura 3.7

a qual — dado que o ponto O não pertence ao domínio de f — significa que, para todo o $\delta > 0$ existe $\epsilon > 0$ tal que $(x, y) \in B_\epsilon(O)\backslash\{O\}$ implica $|f(x, y) - b| < \delta$, é necessário e suficiente que qualquer que seja $\delta > 0$ exista $\epsilon > 0$ por forma que, para todo o ponto (r, θ) pertencente à faixa Σ_ϵ, se tenha $|F(r, \theta) - b| < \delta$.

Noutros termos: a função $f(x, y)$ tende para o limite b quando (x, y) tende para $(0, 0)$ sse F verifica a condição seguinte: para todo o $\delta > 0$ existe $\epsilon > 0$ tal que a desigualdade:

$$|F(r, \theta) - b| < \delta$$

é verificada sempre que seja $0 < r < \epsilon$ (independentemente do valor real atribuído a θ).

A título de exemplo, consideremos a função:

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

para a qual é

$$F(r, \theta) = r(\cos^3\theta + 3\cos^2\theta \operatorname{sen}\theta - \operatorname{sen}^3\theta)$$

e portanto

$$|F(r, \theta)| \le r(|\cos\theta|^3 + 3|\cos\theta|^2|\operatorname{sen}\theta| + |\operatorname{sen}\theta|^3) < 5r;$$

ter-se-á assim $|F(r, \theta)| < \delta$ desde que seja $0 < r < \delta/5$, o que prova que

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

Aproveitaremos esta oportunidade para introduzir, num contexto em que surgem de modo natural, algumas ideias cujo alcance transcende, de longe, a questão particular a que iremos aplicá-las. De qualquer modo essas ideias, aliás estreitamente relacionadas com algumas outras que abordámos anteriormente, no estudo das sucessões e séries de funções de uma variável real, permitir-nos-ão esclarecer melhor alguns aspectos do problema que temos vindo a analisar.

Consideremos uma função $g(u, v)$, definida no conjunto dos pontos $(u, v) \in \mathbb{R}^2$ tais que $u > 0$. Se, para cada $v_0 \in \mathbb{R}$, a função (de uma variável real) $f(u, v_0)$ tem limite (finito) quando $u \to 0^+$ — limite em geral dependente de v_0, que designaremos por $h(v_0)$ — diremos que, quando $u \to 0^+$, a função $g(u, v)$ *converge pontualmente sobre* $\mathbb{R}$ para a função $h(v)$ e escreveremos:

$$\lim_{u \to 0^+} g(u, v) = h(v) \qquad (v \in \mathbb{R}).$$

Por exemplo, para $v \in \mathbb{R}$:

$$\lim_{u \to 0^+} (u \operatorname{sen} v + v \cos u) = v$$

e

$$\lim_{u \to 0^+} e^{\frac{v - |v|}{u}} = H(v),$$

onde H designa a função de Heaviside ($H(v) = 1$ se $v \geq 0$, $H(v) = 0$ se $v < 0$); observe-se que, como mostra o último exemplo, uma função contínua em todo o semiplano $u > 0$ pode convergir pontualmente, quando $u \to 0^+$, para uma função que não é contínua.

De acordo com a definição de convergência pontual, a expressão:

$$\lim_{u \to 0^+} g(u, v) = h(v) \qquad (v \in \mathbb{R})$$

significa que, dado arbitrariamente $\delta > 0$, existe, para cada $v_0 \in \mathbb{R}$, um $\epsilon > 0$ tal que, para $0 < u < \epsilon$, se verifica a desigualdade

$$|g(u, v_0) - h(v_0)| < \delta\,;$$

claro que ϵ depende não só de δ como do ponto v_0 considerado[6], não sendo geralmente possível fixar, para cada $\delta > 0$ um número ϵ — independente de v_0 — por forma que a desigualdade precedente seja verificada sempre que se tenha $0 < u < \epsilon$ (e qualquer que seja $v_0 \in \mathbb{R}$).

Por exemplo, com $g(u, v) = u(1 + v^2)$, função que suporemos definida no semiplano $u > 0$, tem-se

$$\lim_{u \to 0^+} g(u, v) = 0 \qquad (v \in \mathbb{R}).$$

Mas se fixarmos δ, por exemplo, no valor 1, não existirá $\epsilon > 0$ tal que, para $0 < u < \epsilon$ e v real arbitrário, se tenha $|g(u, v)| < 1$; para o reconhecer, basta notar que esta desigualdade equivale a:

$$0 < u < \frac{1}{1 + v^2}$$

e que o conjunto dos números da forma $1/(1 + v^2)$, com $v \in \mathbb{R}$, tem ínfimo nulo (ver Figura 3.8).

[6]Esta frase é pouco precisa: pode talvez sugerir que ϵ ficaria univocamente determinado se se fixassem δ e v_0, o que é obviamente falso.

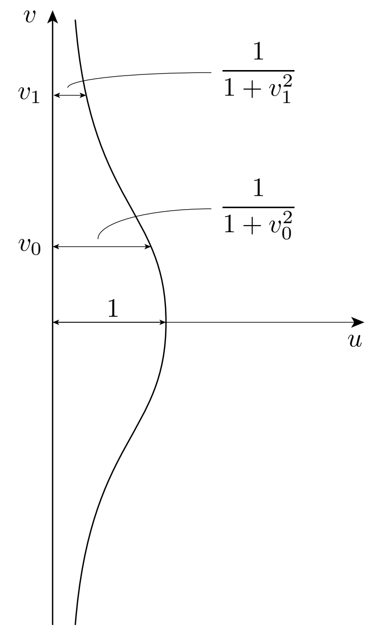

Figura 3.8

Assim, o facto de $g(u,v)$ convergir pontualmente sobre $\mathbb{R}$ para a função $h(v)$, quando $u \to 0^+$, não garante que seja verificada a condição seguinte: qualquer que seja $\delta > 0$ existe uma faixa Σ_ϵ (de largura «uniforme» ϵ, independente de v) tal que, para todo o ponto $(u,v) \in \Sigma_\epsilon$, se tenha $|g(u,v) - h(v)| < \delta$.

Precisamente quando esta última condição se verifica é que dizemos que $g(u,v)$ *converge uniformemente sobre* $\mathbb{R}$ *para a função* $h(v)$, quando $u \to 0^+$ (e de forma análoga se define a convergência uniforme sobre um conjunto qualquer $A \subset \mathbb{R}$).

A noção de convergência uniforme é muito importante em Análise. Em diversas situações, com a convergência pontual (que decerto parece mais natural num primeiro contacto) verificam-se «anomalias» que não são possíveis quando a convergência é uniforme: por exemplo, vimos há pouco que uma função contínua pode convergir pontualmente para uma função não contínua; com convergência uniforme, isso não é possível (prová-lo seria neste momento um bom exercício).

Voltemos agora à questão que nos serviu de pretexto para introduzir estas ideias; seja $f(x,y)$ uma função real definida em $\mathbb{R}^2 \backslash \{(0,0)\}$ e $F(r,\theta) = f(r\cos\theta, r\,\mathrm{sen}\,\theta)$. As conclusões que obtivemos podem agora sintetizar-se do modo seguinte: a condição $\lim_{(x,y)\to(0,0)} f(x,y) = b$ é verificada sse, quando $r \to 0^+$, $F(r,\theta)$ converge uniformemente sobre $\mathbb{R}$ para a (função) constante b.

Por outro lado, é fácil ver que o facto de $F(r,\theta)$ convergir pontualmente sobre $\mathbb{R}$, quando $r \to 0^+$, corresponde precisamente à existência de todos os limites direccionais de $f(x,y)$ no ponto $(0,0)$; no entanto, mesmo que o limite (pontual) seja uma constante, b (caso em que os limites direccionais serão todos iguais a b) a função f só terá limite na origem se a convergência de $F(r,\theta)$ para b for uniforme.

Assim, no caso já atrás considerado de

$$\psi(x,y)=\frac{xy}{x^2+y^2},$$

como a função

$$\psi(r\cos\theta, r\operatorname{sen}\theta)=\frac{1}{2}\operatorname{sen}2\theta,$$

independente de r, converge pontualmente (e até uniformemente) sobre $\mathbb{R}$ para si própria quando $r\to 0^+$, existem todos os limites direccionais de ψ na origem; porém, não sendo estes limites todos iguais (visto que a função $\frac{1}{2}\operatorname{sen}2\theta$ não é constante), ψ não tem limite neste ponto, como já sabíamos.

Como último exemplo, considere-se a função:

$$f(x,y)=H\left(\sqrt{x^2+y^2}-\frac{y}{\sqrt{x^2+y^2}}\right)+H\left(\frac{y}{\sqrt{x^2+y^2}}-\sqrt{x^2+y^2}\right),$$

onde H designa de novo a função de Heaviside. Passando a coordenadas polares obtém-se:

$$F(r,\theta)=H(r-\operatorname{sen}\theta)+H(\operatorname{sen}\theta-r),$$

isto é, a função, definida no semiplano $r>0$ e que assume o valor 1 em todos os pontos (r,θ) deste semiplano, com excepção dos que verificam a condição $r=\operatorname{sen}\theta$, nos quais toma o valor 2.

Vê-se facilmente que:

$$\lim_{r\to 0^+}F(r,\theta)=1 \qquad (\theta\in\mathbb{R})$$

sendo portanto iguais a 1 todos os limites direccionais de f na origem. Porém, como a convergência expressa na fórmula precedente não é uniforme sobre $\mathbb{R}$ (basta observar que, em qualquer faixa Σ_ϵ há pontos (r,θ) com $r=\operatorname{sen}\theta$) pode concluir-se que f não tem limite no ponto $(0,0)$.

Terminaremos este parágrafo com uma breve referência a algumas variantes da noção de limite não enquadradas no estudo anterior (mas tão naturais que quase poderíamos dispensar-nos de mencioná-las explicitamente) e com a introdução de uma notação que nos será útil na sequência.

Sendo f uma função real definida num conjunto $D\subset\mathbb{R}^n$ e $\boldsymbol{a}$ um ponto aderente a D, diz-se que $f(\boldsymbol{x})$ *tende para* $+\infty$ *quando* $\boldsymbol{x}$ *tende para* $\boldsymbol{a}$, e escreve-se:

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}}f(\boldsymbol{x})=+\infty,$$

sse, para todo o $k>0$ existe $\epsilon>0$ tal que, para qualquer ponto $\boldsymbol{x}\in D$ que verifique a condição $\|\boldsymbol{x}-\boldsymbol{a}\|<\epsilon$ se tenha $f(\boldsymbol{x})>k$.[7]

[7]Observe-se que, de acordo com esta definição, uma função real cujo domínio contenha o ponto $\boldsymbol{a}$ não poderá ter limite $+\infty$ nesse ponto.

De maneira análoga se atribui sentido à expressão:

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = -\infty.$$

Assim, ter-se-á, por exemplo:

$$\lim_{(x,y,z)\to(0,0,0)} \frac{1}{\sqrt{x^2+y^2+z^2}} = +\infty.$$

Convém também algumas vezes considerar o limite de uma função real ou vectorial $f(\boldsymbol{x})$ quando $\boldsymbol{x}$ se «afasta indefinidamente (da origem)»; sendo f uma função definida num conjunto não limitado $D \subset \mathbb{R}^n$ e com valores em $\mathbb{R}^m$ e $\boldsymbol{b}$ um vector deste espaço, escreve-se:

$$\lim_{\|\boldsymbol{x}\|\to+\infty} f(\boldsymbol{x}) = \boldsymbol{b}$$

sse para todo o $\delta > 0$ existe k tal que, sempre que um ponto $\boldsymbol{x} \in D$ verifique a condição $\|\boldsymbol{x}\| > k$ se tenha $\|f(\boldsymbol{x}) - \boldsymbol{b}\| < \delta$.

Por exemplo, sendo $\boldsymbol{a}$ um vector qualquer de $\mathbb{R}^n$ e $\boldsymbol{x} \in \mathbb{R}^n$, tem-se:

$$\lim_{\|\boldsymbol{x}\|\to+\infty} \frac{\boldsymbol{a}\cdot\boldsymbol{x}}{\|\boldsymbol{x}\|^2} = 0\,;$$

basta notar que, para $\boldsymbol{x} \neq \boldsymbol{0}$,

$$\left|\frac{\boldsymbol{a}\cdot\boldsymbol{x}}{\|\boldsymbol{x}\|^2}\right| = \frac{|\boldsymbol{a}\cdot\boldsymbol{x}|}{\|\boldsymbol{x}\|^2} \leq \frac{\|\boldsymbol{a}\|\|\boldsymbol{x}\|}{\|\boldsymbol{x}\|^2} = \frac{\|\boldsymbol{a}\|}{\|\boldsymbol{x}\|}$$

e que portanto a desigualdade $\left|\frac{\boldsymbol{a}\cdot\boldsymbol{x}}{\|\boldsymbol{x}\|^2}\right| < \delta$ será verificada desde que seja $\|\boldsymbol{x}\| > \|\boldsymbol{a}\|/\delta$.

De forma óbvia se atribuiria ainda um sentido a expressões tais como:

$$\lim_{\|\boldsymbol{x}\|\to+\infty} f(\boldsymbol{x}) = +\infty, \qquad \lim_{\|\boldsymbol{x}\|\to+\infty} g(\boldsymbol{x}) = -\infty,$$

com f e g funções reais.

Seja D um subconjunto de $\mathbb{R}^n$ e $\boldsymbol{a}$ um ponto interior a D.[8] Sendo f uma função definida em D e com valores em $\mathbb{R}^m$ diremos, naturalmente, que f é um *infinitésimo no ponto* $\boldsymbol{a}$ *(ou quando* $\boldsymbol{x} \to \boldsymbol{a}$*)* sse $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = \boldsymbol{0}$, onde $\boldsymbol{0}$ é o vector nulo de $\mathbb{R}^m$.

Seja agora φ uma função real definida em D, verificando a condição $\varphi(\boldsymbol{x}) \neq 0$ para todo o $\boldsymbol{x} \in D\backslash\{\boldsymbol{a}\}$, e seja ainda $f : D \to \mathbb{R}^m$; diremos que f é *desprezável em relação a* φ *no ponto* $\boldsymbol{a}$ (ou *quando* $\boldsymbol{x} \to \boldsymbol{a}$) e escreveremos:

$$f = o(\varphi) \quad (\text{quando } \boldsymbol{x} \to \boldsymbol{a})$$

[8] Para introduzir as ideias e notações subsequentes bastaria supôr que $\boldsymbol{a}$ era um ponto aderente a D; porém a hipótese $\boldsymbol{a} \in \operatorname{int} D$ é a única que nos vai interessar no cálculo diferencial e, admitindo-a, simplificam-se ligeiramente alguns dos enunciados deste parágrafo.

(ou apenas $f = o(\varphi)$, quando o ponto $\boldsymbol{a}$ estiver claramente fixado) sse existir uma função $f^* : D \to \mathbb{R}^m$, infinitésima no ponto $\boldsymbol{a}$ e tal que:

$$f(\boldsymbol{x}) = \varphi(\boldsymbol{x}) f^*(\boldsymbol{x}),$$

para todo o $\boldsymbol{x} \in D$.

No caso de ser $\varphi(\boldsymbol{a}) \neq 0$ (e portanto $\varphi(\boldsymbol{x}) \neq 0$ para todo o $\boldsymbol{x} \in D$), tem-se $f = o(\varphi)$ sse:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} \frac{f(\boldsymbol{x})}{\varphi(\boldsymbol{x})} = \mathbf{0}.$$

Se for $\varphi(\boldsymbol{a}) = 0$, a relação $f = o(\varphi)$ equivale à conjunção das condições:

$$\lim_{\boldsymbol{x} \to \boldsymbol{a}} \frac{f(\boldsymbol{x})}{\varphi(\boldsymbol{x})} = \mathbf{0} \quad \text{e} \quad f(\boldsymbol{a}) = \mathbf{0}$$

(observe-se que neste caso, não pertencendo o ponto $\boldsymbol{a}$ ao domínio de $f(\boldsymbol{x})/\varphi(\boldsymbol{x})$, a primeira das referidas condições não implica a segunda, que é indispensável para que a relação $f = o(\varphi)$ seja verificada).

Por exemplo, com $m = n = 1$, $a = 0$ e sendo:

$$f(x) = x, \quad g(x) = x^3, \quad h(x) = 1,$$

tem se:

$$g = o(f), \quad f = o(h), \quad g = o(h),$$

relações que se escrevem correntemente na forma:

$$x^3 = o(x), \quad x = o(1), \quad x^3 = o(1).$$

Tem-se também, com $m = 1$, $n = 2$ e $\boldsymbol{a} = (0, 0)$:

$$x^2 - y^2 = o\left(\sqrt{x^2 + y^2}\right),$$

o que pode ainda escrever-se:

$$x^2 = y^2 + o\left(\sqrt{x^2 + y^2}\right).$$

Para utilização posterior convém observar desde já (voltando ao caso geral e supondo fixado um dado ponto $\boldsymbol{a}$) que as relações:

$$f = o(\varphi) \quad \text{e} \quad f = o(|\varphi|)$$

são equivalentes. Verificaremos apenas que a condição $f = o(\varphi)$ implica $f = o(|\varphi|)$, dado que a prova da implicação oposta é idêntica.

Na realidade, existindo uma função $f^* : D \to \mathbb{R}^m$, infinitésima no ponto $\boldsymbol{a}$ e tal que $f(\boldsymbol{x}) = \varphi(\boldsymbol{x}) f^*(\boldsymbol{x})$ para $\boldsymbol{x} \in D$, bastará pôr, por definição:

$$\bar{f}(\boldsymbol{x}) = \begin{cases} \frac{\varphi(\boldsymbol{x})}{|\varphi(\boldsymbol{x})|} f^*(\boldsymbol{x}) & \text{se } \boldsymbol{x} \in D \backslash \{\boldsymbol{a}\} \\ \boldsymbol{0} & \text{se } \boldsymbol{x} = \boldsymbol{a}, \end{cases}$$

para que se tenha $f(\boldsymbol{x}) = |\varphi(\boldsymbol{x})| \bar{f}(\boldsymbol{x})$, sendo $\bar{f} : D \to \mathbb{R}^m$ infinitésima no ponto $\boldsymbol{a}$.

É óbvio que a condição $f = o(1)$, quando $\boldsymbol{x} \to \boldsymbol{a}$, significa precisamente que f é infinitésima no ponto $\boldsymbol{a}$ (convirá talvez notar que, em geral, o facto de se ter $f = o(\varphi)$ não assegura que f seja um infinitésimo: por exemplo, se for $f(x) = 1$ e $\varphi(x) = 1/x$ para $x \in \mathbb{R} \backslash \{0\}$, com $f(0) = \varphi(0) = 0$, a condição $f = o(\varphi)$ quando $x \to 0$ será verificada). E é também evidente que, se φ for um infinitésimo no ponto $\boldsymbol{a}$ e se tiver $f = o(\varphi)$, f será também infinitésima no mesmo ponto.

Neste último caso, costuma-se dizer que f é um *infinitésimo de ordem superior à de φ no ponto $\boldsymbol{a}$*; a ideia intuitiva é a de que, quando $\boldsymbol{x} \to \boldsymbol{a}$, $f(\boldsymbol{x})$ tende para $\boldsymbol{0}$ «mais rapidamente» do que $\varphi(\boldsymbol{x})$ tende para 0.

Um caso de especial importância é o de $\varphi(\boldsymbol{x})$ ser uma função da forma:

$$\varphi(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{a}\|^\alpha,$$

com α real positivo e $\boldsymbol{x} \in D \subset \mathbb{R}^n$. Para se exprimir que a condição:

$$f(\boldsymbol{x}) = o(\|\boldsymbol{x} - \boldsymbol{a}\|^\alpha)$$

é verificada, diz-se que f é um *infinitésimo de ordem superior a α, no ponto $\boldsymbol{a}$*. Interessar-nos-á muito especialmente no próximo capítulo o caso particular dos infinitésimos de ordem superior a 1, que também se dizem infinitésimos de ordem superior à primeira e que são portanto as funções para as quais se tem:

$$f(\boldsymbol{x}) = o(\|\boldsymbol{x} - \boldsymbol{a}\|) \quad (\text{quando } \boldsymbol{x} \to \boldsymbol{a})$$

Como exemplo, mencione-se que a função $\operatorname{sen}^2(x+y)$ é um infinitésimo de ordem superior à primeira (e também de ordem superior a α, para qualquer $\alpha \in]0, 2[$) quando $(x, y) \to (0, 0)$.

Consideremos novamente um conjunto $D \subset \mathbb{R}^n$, um ponto $\boldsymbol{a} \in D$ e duas funções $f : D \to \mathbb{R}^m$ e $\varphi : D \to \mathbb{R}$, supondo ainda que $\varphi(\boldsymbol{x}) \neq 0$ para todo o $\boldsymbol{x} \in D \backslash \{\boldsymbol{a}\}$. No caso de existir $\epsilon > 0$ e uma função $f^* : D \to \mathbb{R}^m$, limitada em $B_\epsilon(\boldsymbol{a})$, por forma que se verifique a igualdade:

$$f(\boldsymbol{x}) = \varphi(\boldsymbol{x}) f^*(\boldsymbol{x})$$

em todo o ponto $\boldsymbol{x} \in D$, diremos que *f é dominada por φ no ponto $\boldsymbol{a}$* (ou *quando $\boldsymbol{x} \to \boldsymbol{a}$*) e escreveremos:

$$f = O(\varphi) \quad (\text{quando } \boldsymbol{x} \to \boldsymbol{a}),$$

ou, se não houver risco de confusão, apenas $f = O(\varphi)$.

É claro que, se $\varphi(\boldsymbol{a}) \neq 0$, dizer que $f = O(\varphi)$ equivale a dizer que o cociente $f(\boldsymbol{x})/\varphi(\boldsymbol{x})$ é limitado nalguma bola centrada no ponto $\boldsymbol{a}$; se for $\varphi(\boldsymbol{a}) = 0$, porém, a condição $f = O(\varphi)$ será verificada sse esse cociente (definido em $D\backslash\{\boldsymbol{a}\}$) for limitado na intersecção do seu domínio com uma bola $B_\epsilon(\boldsymbol{a})$ e se, além disso, for $f(\boldsymbol{a}) = \mathbf{0}$.

Assim, por exemplo, ter-se-á (com $m = 1$, $n = 2$ e sendo $\boldsymbol{a}$ a origem):

$$x^2 = y^2 + O(x^2 + y^2).$$

Outro exemplo, que será útil na sequência: sendo $f : \mathbb{R}^n \to \mathbb{R}^m$ uma aplicação linear, verifica-se a relação $f(\boldsymbol{x}) = O(\|\boldsymbol{x}\|)$; é o que imediatamente se reconhece tendo em conta que (como vimos na pág. 48 ao provar a continuidade das aplicações lineares) pode garantir-se a existência de uma constante C tal que $\|f(\boldsymbol{x})\| \leq C\|\boldsymbol{x}\|$, qualquer que seja $\boldsymbol{x} \in \mathbb{R}^n$.

Indicaremos agora algumas propriedades das relações expressas pelos símbolos «O» e «o», que utilizaremos eventualmente em capítulos seguintes; as demonstrações, com base nas definições dos referidos símbolos e em propriedades bem conhecidas das noções de limite e de função limitada, poderão ficar como exercícios.

Supondo verificadas as condições: $D \subset \mathbb{R}^n$; $\boldsymbol{a} \in D$; $f, g : D \to \mathbb{R}^m$; $\alpha, \varphi, \psi : D \to \mathbb{R}$ e ainda que os símbolos $o(\varphi)$, $O(\varphi)$, $o(\psi)$, etc., se referem todos ao mesmo ponto $\boldsymbol{a}$, tem-se:

- Se $f = o(\varphi)$, então também $f = O(\varphi)$.
- Se $f = o(\varphi)$ e $g = o(\varphi)$, então $f \pm g = o(\varphi)$; se $f = O(\varphi)$ e $g = O(\varphi)$, $f \pm g = O(\varphi)$ (estas proposições exprimem-se por vezes, de forma algo imprecisa, escrevendo: $o(\varphi) \pm o(\varphi) = o(\varphi)$, $O(\varphi) \pm O(\varphi) = O(\varphi)$).
- Se $f = o(\varphi)$ e $\alpha = O(\psi)$ (ou $f = O(\varphi)$ e $\alpha = o(\psi)$), então $\alpha f = o(\varphi\psi)$ (abreviadamente: $O(\psi)o(\varphi) = o(\psi)O(\varphi) = o(\varphi\psi)$).
- Se $f = o(\varphi)$ e $g = O(\psi)$ (ou $f = O(\varphi)$ e $g = o(\psi)$) então $f \cdot g = o(\varphi\psi)$ $\big(o(\varphi) \cdot O(\psi)\big) = o(\varphi\psi)$, etc.)

Vem aqui a propósito transcrever (do livro «Introdução à Análise Matemática», do mesmo autor) os dois parágrafos seguintes (adaptados à situação presente):

As notações «O» e «o», devidas ao matemático alemão Landau, são usadas frequentemente em textos de Matemática e, em determinadas situações, a sua utilidade é manifesta. No entanto, do ponto de vista da coerência lógica, podem merecer algum reparo (por exemplo, contrariamente às regras usuais, das igualdades $f = o(\varphi)$, $g = o(\varphi)$ não pode deduzir-se $f = g$; e é óbvio que de $o(\varphi) + o(\varphi) = o(\varphi)$ não decorre $o(\varphi) = 0$).

Seria na realidade preferível, em lugar de $f = o(\varphi)$, escrever $f \in o(\varphi)$, encarando o símbolo $o(\varphi)$ como representativo de um conjunto de funções definido de maneira conveniente. Não é isto, porém, o que se faz na generalidade dos textos e a verdade é que, do ponto de vista prático, o uso das notações de Landau é muitas vezes cómodo e não conduz a qualquer confusão nos casos habituais.

Capítulo 4

Cálculo diferencial

4.1 Introdução. Alguns aspectos da diferenciabilidade, para funções de uma variável real

Este parágrafo tem carácter introdutório e, salvo por razões de ordem pedagógica, poderia mesmo ser omitido. A noção fundamental do cálculo diferencial — a de derivada — será aqui considerada apenas no quadro das funções reais de variável real, procurando-se destacar alguns aspectos que (embora neste momento possam parecer um pouco rebuscados) virão a constituir a chave para as generalizações a empreender em parágrafos seguintes.

Se pensarmos nas razões do interesse vital do conceito de derivada no estudo das funções de variável real, poderemos detectar dois aspectos distintos, embora estreitamente relacionados.

Em primeiro lugar, a derivada surge como o instrumento natural para medir, localmente, a «taxa de variação» de uma função.

Como é bem sabido, no caso muito particular de uma função polinomial de grau ≤ 1, $\varphi(x) = mx + b$, que tem por gráfico uma recta, é natural adoptar o declive m dessa recta como uma medida da «taxa de variação» da função: atribuído um dado acréscimo positivo à variável x, quanto maior for $|m|$ maior será, em valor absoluto, o acréscimo (positivo ou negativo consoante $m > 0$ ou $m < 0$) sofrido por $\varphi(x)$. Mais geralmente, sendo f uma função real cujo domínio contenha um intervalo aberto $I \subset \mathbb{R}$ e a, $a + h$ dois pontos distintos de I, o cociente:

$$\frac{f(a+h) - f(a)}{h}$$

pode ser encarado como uma medida da «taxa média de variação» de f, por unidade de comprimento, entre os pontos a e $a + h$, e o limite desse cociente quando $h \to 0$ (se existir) será um indicador natural da maior ou menor «rapidez» com que varia $f(x)$ quando x se afasta (pouco) do ponto a.

Veremos no início do parágrafo seguinte como podem estender-se ao caso das funções de n variáveis reais as ideias que acabamos de expor (aliás de forma um tanto imprecisa).

Uma outra ordem de ideias que está na base do interesse fundamental do conceito de derivada insere-se no quadro da «aproximação funcional»; quando, ao pretender estudar uma função sob determinado ponto de vista se depara com dificuldades muito consideráveis, é natural pensar em substituí-la por outra função «mais simples», que permita ainda obter a informação pretendida sem erro excessivo.

Para concretizar as ideias, suponha-se que se conhecia o valor de uma função f no ponto 0 e se pretendia avaliar $f(x)$ num ponto x «próximo» de 0.

Na ausência de qualquer outra informação sobre f, nada poderia fazer-se. Mas se se soubesse que f era contínua na origem, já se tornaria razoável usar, como valor aproximado de $f(x)$, o valor conhecido $f(0)$. Ao fazê-lo cometer-se-ia um erro:

$$r_0(x) = f(x) - f(0),$$

do qual se saberia apenas que era um infinitésimo na origem ($r_0(x) = o(1)$), quando $x \to 0$ e que, portanto, seria decerto «muito pequeno» se x estivesse «suficientemente próximo» de 0.

É claro que este conhecimento seria demasiado escasso para permitir uma majoração do erro em termos quantitativos, que só poderia obter-se se se dispusesse de muito mais informação sobre a função f; no entanto, em determinadas situações concretas, poderia já ter alguma utilidade.

Admitamos agora que a função era, não apenas contínua, mas diferenciável na origem, sendo também conhecido o valor $f'(0)$. Seria então preferível, em princípio, adoptar como valor aproximado de $f(x)$ o número $f(0) + f'(0)x$ (o que corresponderia a usar como aproximação do gráfico de f a sua tangente no ponto $\big(0, f(0)\big)$, em vez da horizontal de equação $y = f(0)$ (Figura 4.1).

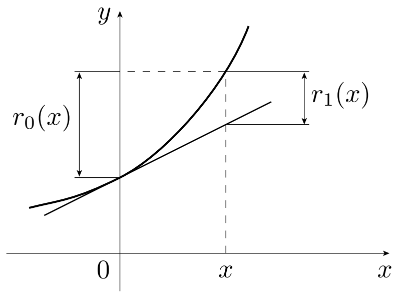

Figura 4.1

0 erro correspondente a esta nova aproximação seria:

$$r_1(x) = f(x) - f(0) - f'(0)x,$$

verificando-se imediatamente que, quando $x \to 0$, se teria, não só $r_1(x) = o(1)$, mas também $r_1(x) = o(x)$; assim, o erro seria agora um infinitésimo de ordem superior à primeira no ponto 0 (de acordo com a definição introduzida no parágrafo precedente, esta última afirmação significa que $r_1(x) = o(|x|)$; mas vimos também que esta condição equivale a $r_1(x) = o(x)$).

Poderia prosseguir-se nesta ordem de ideias,[1] mas não é isso o que nos interessa agora. Convém-nos antes salientar um aspecto, que talvez não pareça muito significativo à primeira vista, mas que acabará por revelar-se essencial. A observação que queremos fazer é a seguinte: no caso de f ser diferenciável na origem, há um e um só número real m que verifica a condição

$$f(x) - f(0) = mx + o(x)$$

(precisamente o número $m = f'(0)$); no caso de f não ser diferenciável na origem, nenhum número verifica a condição referida.

Para verificar estas afirmações basta notar que, se existe (pelo menos) um $m \in \mathbb{R}$ satisfazendo a condição em causa, se tem necessariamente, em qualquer ponto x do domínio de f distinto de 0:

$$\frac{f(x) - f(0)}{x} = m + \frac{o(x)}{x},$$

donde, atendendo a que o segundo membro tende para m quando $x \to 0$, pode concluir-se que f é diferenciável na origem e que $f'(0) = m$; e é óbvio que, reciprocamente, sendo f diferenciável na origem, o número $m = f'(0)$ satisfaz a condição em referência.

Assim, dizer que f é diferenciável no ponto 0 equivale a afirmar a existência de um real m (que aliás será único) tal que o produto mx aproxima o acréscimo $f(x) - f(0)$ com um erro que é um infinitésimo de ordem superior à primeira quando $x \to 0$.

Para podermos dar a esta condição de diferenciabilidade a forma que nos interessará definitivamente, convém recordar (como aliás foi feito no parágrafo 3.1) que uma aplicação linear de $\mathbb{R}^q$ em $\mathbb{R}^p$ pode sempre representar-se por uma matriz de elementos reais do tipo $p \times q$, sendo bijectiva a correspondência entre estas matrizes e aquelas aplicações. Daqui decorre imediatamente que as aplicações lineares de $\mathbb{R}$ em si mesmo (caso $p = q = 1$) se correspondem bijectivamente com as matrizes do tipo $[m]$, com um só elemento real, isto é, com os próprios números reais.

Na realidade, qualquer aplicação linear $L : \mathbb{R} \to \mathbb{R}$ é representável na forma:

$$L(x) = mx \qquad (x \in \mathbb{R}),$$

em que m é um número real determinado univocamente pela aplicação L (m é precisamente o valor de L no ponto 1) e, em sentido inverso, a todo o número real m pode associar-se, por meio da fórmula precedente, uma única aplicação linear de $\mathbb{R}$ em si mesmo.

[1] Como é sabido, poderia em particular reconhecer-se que uma função n vezes diferenciável na origem é aproximável por um polinómio de grau $\leq n$, o seu polinómio de Mac-Laurin, com um erro $r_n(x) = o(x^n)$.

Tendo em conta o resultado expresso na nota anterior — segundo o qual as aplicações lineares de $\mathbb{R}$ em si mesmo podem praticamente «identificar-se» com os próprios números reais — poderíamos então dizer que, para que f seja diferenciável na origem, é necessário e suficiente que exista uma aplicação linear $L_0 : \mathbb{R} \to \mathbb{R}$ tal que, em todo o ponto x do domínio de f, se verifique a igualdade:

$$f(x) - f(0) = L_0(x) + o(x).$$

Considerando, em vez da origem, um ponto qualquer $a \in \mathbb{R}$ (e uma função f cujo domínio contivesse uma vizinhança do ponto a) obteríamos de forma análoga a conclusão seguinte: f é diferenciável no ponto a sse existe uma aplicação linear $L_a : \mathbb{R} \to \mathbb{R}$ tal que, em qualquer ponto x do domínio de f, se tenha:

$$f(x) - f(a) = L_a(x - a) + o(x - a),$$

ou, pondo $x - a = h$:

$$f(a + h) - f(a) = L_a(h) + o(h),$$

(para todo o real h tal que $a + h$ pertença ao domínio de f).

Por exemplo, para $f(x) = x^3$ tem-se, em qualquer ponto $a \in \mathbb{R}$:

$$(a + h)^3 - a^3 = 3a^2h + 3ah^2 + h^3,$$

com h real arbitrário. Neste caso, a aplicação linear L_a é definida por $L_a(h) = 3a^2h$ para todo o $h \in \mathbb{R}$, sendo o termo de erro $3ah^2 + h^3$, que é evidentemente $o(h)$; o número real $3a^2$, que determina a aplicação linear L_a, é precisamente $f'(a)$.

Pode assim dizer-se que as funções diferenciáveis no ponto a são precisamente aquelas cujo acréscimo, $f(a + h) - f(a)$, pode ser aproximado por uma função linear de h, sendo o erro correspondente a essa aproximação um infinitésimo de ordem superior à primeira quando $h \to 0$ (em termos mais intuitivos: a menos de um infinitésimo de ordem superior à primeira, o acréscimo da função é uma função linear do acréscimo da variável independente).

Convirá reter esta conclusão porque, como veremos, ela será a base mais conveniente para a generalização do conceito de derivada ao caso das funções, reais ou vectoriais, de variável vectorial.

4.2 Cálculo diferencial de primeira ordem: derivadas parciais, diferenciabilidade; teorema do valor médio

Seja $f(x, y)$ uma função real definida num conjunto $D \subset \mathbb{R}^2$ e (a, b) um ponto interior a D; procuraremos agora avaliar a «taxa de variação» de $f(x, y)$ quando se atribuam «pequenos acréscimos» ao ponto (x, y), a partir da posição (a, b) (Figura 4.2).

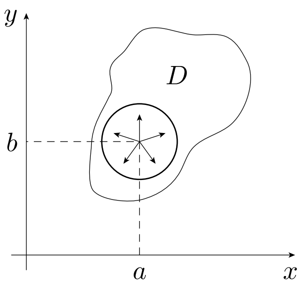

Figura 4.2

Convém observar já que, enquanto no caso das funções de uma variável os «acréscimos» possíveis tinham todos a mesma direcção — a do eixo das abcissas — agora podemos considerar acréscimos (h, k) com qualquer das direcções do plano (deverá naturalmente exigir-se que o ponto $(a + h, b + k)$ pertença ainda ao domínio de f, mas isso decerto se verificará se o módulo do vector (h, k) for suficientemente pequeno, visto que supusemos que o ponto (a, b) era interior a D); e será natural esperar que, em geral, a «taxa de variação» de f dependa da direcção considerada (assim, por exemplo, se $f(x, y)$ designasse a temperatura no ponto (x, y), situado no chão de uma oficina com um forno em funcionamento e uma porta aberta para o exterior, era de esperar que a temperatura aumentasse rapidamente nas direcções que conduziam ao forno e diminuisse nas que levavam à saída).

Para maior simplicidade, consideremos em primeiro lugar duas direcções «privilegiadas»: as dos eixos coordenados. Se (h, k) tiver a direcção do eixo dos x — isto é, se for $k = 0$ e $h \neq 0$ — a «razão incremental» a considerar será:

$$\frac{f(a + h, b) - f(a, b)}{h}$$

Ao limite desta razão quando $h \to 0$, se existir, chama-se *derivada parcial da função f, no ponto (a, b), em ordem à primeira variável*, usando-se para designá-la qualquer dos símbolos: $\mathcal{D}_1 f(a, b)$, $f_1'(a, b)$ ou, se estiver convencionado que a primeira variável é designada por x, $\mathcal{D}_x f(a, b)$, $f_x'(a, b)$, $\partial f/\partial x(a, b)$; quando se tenha escrito $z = f(x, y)$, poderá usar-se ainda o símbolo $\partial z/\partial x(a, b)$ para designar a mesma derivada.

Reconhece-se imediatamente que a derivada parcial $\partial f/\partial x(a, b)$, quando existe, coincide com a derivada (ordinária) no ponto a de uma função de uma única variável real: precisamente a «função parcial» φ que se obtém de f por fixação de y no valor b. Com efeito, pondo $\varphi(x) = f(x, b)$, tem-se (desde que exista uma das derivadas $\varphi'(a)$ ou $\partial f/\partial x(a,b)$):

$$\begin{aligned}\varphi'(a) &= \lim_{h\to 0}\frac{\varphi(a+h)-\varphi(a)}{h}\\ &= \lim_{h\to 0}\frac{f(a+h,b)-f(a,b)}{h}\\ &= \frac{\partial f}{\partial x}(a,b).\end{aligned}$$

De forma análoga se define a derivada parcial de f no ponto (a,b), em ordem à segunda variável (ou em ordem a y), designada por $\mathcal{D}_2 f(a,b)$, $\mathcal{D}_y f(a,b)$, $f'_y(a,b)$, $\partial f/\partial y(a,b)$, etc.:

$$\frac{\partial f}{\partial y}(a,b) = \lim_{k\to 0}\frac{f(a,b+k)-f(a,b)}{k}.$$

Quando existe, esta derivada coincide com a derivada ordinária da função $\psi(y) = f(a,y)$ no ponto b.

Por exemplo, sendo $z = f(x,y) = x^2 + \operatorname{sen} xy$ e $(a,b)\in\mathbb{R}^2$, as funções parciais a considerar serão:

$$\varphi(x) = x^2 + \operatorname{sen} bx, \quad \psi(y) = a^2 + \operatorname{sen} ay,$$

obtendo-se portanto:

$$\frac{\partial z}{\partial x}(a,b) = \varphi'(a) = 2a + b\cos ab$$

e

$$\frac{\partial z}{\partial y}(a,b) = \psi'(b) = a\cos ab.$$

Considerando em vez de (a,b) um ponto qualquer (x,y) — cuja indicação explícita a seguir aos símbolos $\partial z/\partial x$, $\partial z/\partial y$ é muitas vezes omitida — poderia escrever-se:

$$\begin{aligned}\frac{\partial z}{\partial x} &= 2x + y\cos xy\\ \frac{\partial z}{\partial y} &= x\cos xy.\end{aligned}$$

Na prática, para calcular a primeira destas derivadas parciais, derivar-se-ia o segundo membro da igualdade $z = x^2 + \operatorname{sen} xy$ em ordem a x pelas regras usuais da derivação ordinária, considerando y como se fosse uma constante; e de modo análogo para $\partial z/\partial y$. Com efeito, o facto da derivação parcial se reduzir à derivação ordinária da função parcial correspondente torna óbvio que as regras de derivação habituais no caso de uma variável manterão inteira validade para o cálculo de derivadas parciais.

Trataremos agora do problema da derivação em termos mais gerais. Sendo $\boldsymbol{v} = (\alpha,\beta)$ um vector qualquer de $\mathbb{R}^2$, consideremos as equações paramétricas da recta r que passa por (a,b) e tem a direcção do vector $\boldsymbol{v}$ (Figura 4.3):

$$\begin{cases} x = a + t\alpha \\ y = b + t\beta \end{cases} \qquad (t \in \mathbb{R}).$$

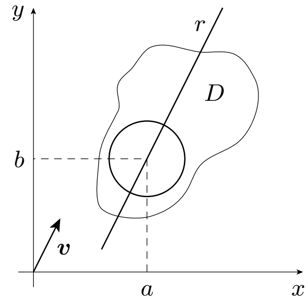

Figura 4.3

Tendo em conta que o ponto (a,b) se supôs interior ao conjunto D, logo se vê que, compondo a função f com a aplicação $t \to (x,y)$ definida pelas equações precedentes, se obterá uma função $\varphi_{\boldsymbol{v}}(t) = f(a+t\alpha, b+t\beta)$, definida num conjunto ao qual o ponto 0 será interior; em qualquer ponto deste conjunto distinto da origem ter-se-á então:

$$\frac{\varphi_{\boldsymbol{v}}(t) - \varphi_{\boldsymbol{v}}(0)}{t} = \frac{f(a+t\alpha, b+t\beta) - f(a,b)}{t}.$$

Ao limite de qualquer destas razões quando $t \to 0$ (se existir) chamaremos *derivada da função f, no ponto (a,b), segundo o vector $\boldsymbol{v}$*; designá-lo-emos por qualquer dos símbolos $\mathcal{D}_{\boldsymbol{v}} f(a,b)$, $\partial f/\partial \boldsymbol{v}(a,b)$, $f'_{\boldsymbol{v}}(a,b)$.

Ter-se-á portanto:

$$\mathcal{D}_{\boldsymbol{v}} f(a,b) = \lim_{t\to 0} \frac{f(a+t\alpha, b+t\beta) - f(a,b)}{t} = \varphi'_{\boldsymbol{v}}(0)$$

sempre que o limite exista; assim, também a derivação segundo um vector arbitrário pode reduzir-se à derivação ordinária.

No caso particular de $\boldsymbol{v}$ ser um vector unitário ($\|\boldsymbol{v}\| = \sqrt{\alpha^2+\beta^2} = 1$), o comprimento do segmento de recta de extremos (a,b) e $(a+t\alpha, b+t\beta)$ é igual a $|t|$ e a razão incremental:

$$\frac{f(a+t\alpha, b+t\beta) - f(a,b)}{t}$$

pode interpretar-se como uma «taxa média de variação» de f, por unidade de comprimento, ao longo do referido segmento; nesse caso é habitual chamar a $\partial f/\partial \boldsymbol{v}(a,b)$ a *derivada direccional de f na direcção (e sentido) de $\boldsymbol{v}$*.

As derivadas parciais são casos particulares do conceito de derivada direccional: $\partial f/\partial x(a,b)$ é evidentemente a derivada de f em (a,b) segundo o vector unitário $\boldsymbol{e}_1 = (1,0)$ e, analogamente, tem-se $\partial f/\partial y(a,b) = \partial f/\partial \boldsymbol{e}_2(a,b)$ (admitida a existência de tais derivadas).

Como exemplo, consideremos a função definida em $\mathbb{R}^2$ pela fórmula: $z = x^2y$, um ponto qualquer (a,b) e um vector $\boldsymbol{v} = (\alpha,\beta)$; ter-se-á:

$$\frac{\partial z}{\partial \boldsymbol{v}}(a,b) = \lim_{t\to 0}\frac{(a+t\alpha)^2(b+t\beta) - a^2b}{t} = 2ab\alpha + a^2\beta.$$

Em particular, para $\boldsymbol{v} = \boldsymbol{e}_1$ e $\boldsymbol{v} = \boldsymbol{e}_2$, obtêm-se as derivadas parciais:

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

Do facto de a derivação segundo um vector se poder reduzir à derivação ordinária decorre facilmente a validade das regras de derivação usuais no novo caso. Assim, por exemplo, sendo f, g funções reais definidas num conjunto $D \subset \mathbb{R}^2$, $(a,b) \in \operatorname{int} D$ e $\boldsymbol{v} \in \mathbb{R}^2$, se existirem (finitas) as derivadas $\partial f/\partial \boldsymbol{v}(a,b)$ e $\partial g/\partial \boldsymbol{v}(a,b)$, existirão também as derivadas das funções $f+g$, $f-g$ e fg segundo o vector $\boldsymbol{v}$, no ponto (a,b) e verificar-se-ão as igualdades:

$$\frac{\partial(f\pm g)}{\partial \boldsymbol{v}}(a,b) = \frac{\partial f}{\partial \boldsymbol{v}}(a,b) \pm \frac{\partial g}{\partial \boldsymbol{v}}(a,b),$$
$$\frac{\partial(fg)}{\partial \boldsymbol{v}}(a,b) = \frac{\partial f}{\partial \boldsymbol{v}}(a,b)g(a,b) + f(a,b)\frac{\partial g}{\partial \boldsymbol{v}}(a,b);$$

se, além disso, for $g(x,y) \neq 0$ em D — ou numa bola centrada em (a,b) — ter-se-á também:

$$\frac{\partial\left(\frac{f}{g}\right)}{\partial \boldsymbol{v}}(a,b) = \frac{\frac{\partial f}{\partial \boldsymbol{v}}(a,b)g(a,b) - f(a,b)\frac{\partial g}{\partial \boldsymbol{v}}(a,b)}{[g(a,b)]^2},$$

etc.

Nas condições anteriormente fixadas sobre D, (a,b) e $\boldsymbol{v}$, consideremos agora o conjunto de todas as funções $f : D \to \mathbb{R}$ que admitem, no ponto (a,b), uma derivada (finita) segundo o vector $\boldsymbol{v}$. É fácil ver que este conjunto, munido das operações usuais de adição de funções e de multiplicação de um número real por uma função, é um espaço vectorial real; além disso, mostram as relações (onde omitimos a indicação do ponto (a,b)):

$$\mathcal{D}_{\boldsymbol{v}}(f+g) = \mathcal{D}_{\boldsymbol{v}}(f) + \mathcal{D}_{\boldsymbol{v}}(g)$$

e

$$\mathcal{D}_{\boldsymbol{v}}(cf) = c\mathcal{D}_{\boldsymbol{v}}(f),$$

válidas para quaisquer funções f, g do referido espaço e qualquer real c, que a aplicação (desse espaço vectorial em $\mathbb{R}$) que faz corresponder a cada função f o número $\mathcal{D}_{\boldsymbol{v}}f(a,b)$ é uma aplicação linear.

Por outro lado, tem também interesse ver como varia a derivada se, fixando a função f e o ponto (a,b), substituirmos o vector $\boldsymbol{v} = (\alpha, \beta)$ por outro vector com a mesma direcção, $c\boldsymbol{v}$ (com $c \in \mathbb{R} \setminus \{0\}$); ter-se-á, sempre que exista algum dos limites considerados:

$$\begin{aligned}
\mathcal{D}_{c\boldsymbol{v}} f(a,b) &= \lim_{t \to 0} \frac{f(a+tc\alpha, b+tc\beta) - f(a,b)}{t} \\
&= c \lim_{t \to 0} \frac{f(a+tc\alpha, b+tc\beta) - f(a,b)}{tc} \\
&= c\mathcal{D}_{\boldsymbol{v}} f(a,b).
\end{aligned}$$

Este resultado poderia talvez sugerir a questão seguinte: será também verdade que, se existirem as derivadas de f segundo dois vectores quaisquer $\boldsymbol{v}_1$, e $\boldsymbol{v}_2$, existirá necessariamente a derivada $\mathcal{D}_{\boldsymbol{v}_1+\boldsymbol{v}_2} f$ (sempre no ponto (a,b), cuja indicação omitimos) verificando-se a relação:

$$\mathcal{D}_{\boldsymbol{v}_1+\boldsymbol{v}_2} f = \mathcal{D}_{\boldsymbol{v}_1} f + \mathcal{D}_{\boldsymbol{v}_2} f?$$

Se tal conjectura fosse verdadeira, esta relação, em conjunto com a igualdade acima provada (apenas no caso $c \neq 0$, mas também obviamente válida se for $c = 0$), $\mathcal{D}_{c\boldsymbol{v}} f = c\mathcal{D}_{\boldsymbol{v}} f$, traduziriam um «comportamento linear» da operação de derivação, já não a respeito das funções sobre as quais actua, mas relativamente aos vectores $\boldsymbol{v} \in \mathbb{R}^2$ segundo os quais essa operação é efectuada.

Veremos oportunamente que a resposta à questão anterior é afirmativa, quando se considerem apenas funções com um certo grau de «regularidade»; em geral, porém, essa resposta é negativa, como vamos ver.

Para esse efeito, consideremos em primeiro lugar a função $f : \mathbb{R}^2 \to \mathbb{R}$ definida pela forma seguinte:

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

Reconhece-se imediatamente que

$$\begin{aligned}
\mathcal{D}_{\boldsymbol{e}_1} f(0,0) &= \frac{\partial f}{\partial x}(0,0) = 0 \\
\mathcal{D}_{\boldsymbol{e}_2} f(0,0) &= \frac{\partial f}{\partial y}(0,0) = 0
\end{aligned}$$

e, portanto, quaisquer que sejam $c_1, c_2 \in \mathbb{R}$,

$$\mathcal{D}_{c_1\boldsymbol{e}_1} f(0,0) = 0 = \mathcal{D}_{c_2\boldsymbol{e}_2} f(0,0).$$

Porém, se for $\boldsymbol{v} = c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2$ um vector com direcção distinta das dos eixos coordenados ($c_1c_2 \neq 0$) não existirá a derivada $\mathcal{D}_{\boldsymbol{v}} f(0,0)$, visto que não existe o limite quando $t \to 0$ da função:

$$\frac{f(c_1t, c_2t) - f(0,0)}{t} = \sqrt{c_1^2 + c_2^2}\frac{|t|}{t}.$$

Mostra este exemplo que podem existir as derivadas de f segundo dois vectores e não existir a derivada segundo a sua soma; mas é fácil ver que, mesmo que esta última derivada também exista, poderá não ser igual à soma das duas primeiras. Para tal, basta considerar a função definida por:

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

e verificar, por exemplo, que $\mathcal{D}_{\boldsymbol{e}_1+\boldsymbol{e}_2} g(0,0) = 2$, enquanto $\mathcal{D}_{\boldsymbol{e}_1} g(0,0) = \mathcal{D}_{\boldsymbol{e}_2} g(0,0) = 0$.

Contrariamente ao que talvez pudesse ser sugerido, a uma primeira vista, por certos resultados válidos no caso das funções de uma variável (no qual, por exemplo, o facto de uma função ter derivada finita num ponto garante a sua continuidade nesse ponto e até a possibilidade de uma «boa aproximação linear», no sentido indicado na parte final do parágrafo 4.1), para funções de duas variáveis reais a existência de derivadas parciais finitas num ponto não assegura sequer que a função nele seja contínua; aliás, não seria difícil prevê-lo se se tivesse em conta que a existência ou não existência das derivadas $\partial f/\partial x(a,b)$ e $\partial f/\partial y(a,b)$, bem como os valores que elas eventualmente assumam, dependem apenas dos valores de f em pontos situados sobre as rectas de equações $y = b$ e $x = a$, não sendo portanto afectados por uma alteração arbitrária da função nos pontos do seu domínio não pertencentes a qualquer dessas rectas (alteração que certamente poderia afectar a continuidade de f em (a,b)).

Mais difícil, porém, seria imaginar que uma função de duas variáveis reais poderia ter derivada (finita), num dado ponto, segundo qualquer vector $\boldsymbol{v} \in \mathbb{R}^2$ e não ser contínua nesse ponto. No entanto, um exemplo a que já nos referimos no parágrafo 3.2, o da função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por:

$$f(x,y) = \begin{cases} 1 & \text{se } x \neq 0 \text{ e } y = x^2 \\ 0 & \text{se } x = 0 \text{ ou } y \neq x^2, \end{cases}$$

permite reconhecê-lo facilmente. Na verdade é óbvio que, para qualquer vector $\boldsymbol{v} \in \mathbb{R}^2$, se tem $\mathcal{D}_{\boldsymbol{v}} f(0,0) = 0$ e já sabemos que f não é contínua na origem.

Recorde-se que, no caso das funções de uma variável, a noção de diferenciabilidade foi definida pela forma seguinte: dizia-se que uma função era diferenciável no ponto a sse tivesse derivada finita nesse ponto; nestas condições, poderia ocorrer que, para funções de duas variáveis, se adoptasse um conceito «análogo», dizendo que $f(x,y)$ era diferenciável no ponto (a,b) sse existissem (finitas) as derivadas parciais $\partial f/\partial x(a,b)$ e $\partial f/\partial y(a,b)$ ou, mais restritivamente, todas as derivadas $\partial f/\partial \boldsymbol{v}(a,b)$, com $\boldsymbol{v}$ vector arbitrário de $\mathbb{R}^2$. As considerações anteriores, porém, revelam que uma tal noção de «diferenciabilidade» não possuiria pelo menos uma das propriedades essenciais verificadas no caso das funções de variável real (a de «diferenciabilidade implicar continuidade»); além disso, será fácil ver posteriormente (o último exemplo indicado poderá servir ainda para esse efeito) que também não ficaria garantida a existência de uma «boa aproximação linear» para as funções diferenciáveis, se esta noção fosse definida por qualquer dos modos há pouco referidos.

A conclusão a tirar é a de que não serão estas as vias convenientes para a generalização do conceito de diferenciabilidade ao caso das funções de duas ou mais variáveis reais. Antes de vermos qual a ordem de ideias que convirá adoptar, vamos estender rapidamente, para funções reais ou vectoriais de n variáveis, as noções definidas na parte inicial deste parágrafo.

Seja D um subconjunto de $\mathbb{R}^n$, $\boldsymbol{a}$ um ponto interior a D e $f : D \to \mathbb{R}$.

Dado um vector $\boldsymbol{v} \in \mathbb{R}^n$, chamaremos *derivada de f no ponto $\boldsymbol{a}$ segundo o vector $\boldsymbol{v}$* ao limite:

$$\lim_{t\to 0} \frac{f(\boldsymbol{a}+t\boldsymbol{v}) - f(\boldsymbol{a})}{t},$$

sempre que exista.

Mais geralmente, com as mesmas hipóteses — excepto a de f ser uma função real definida em D, que deverá ser substituída pela de ser $f : D \to \mathbb{R}^m$ — a *derivada da função vectorial f, no ponto $\boldsymbol{a}$ e segundo $\boldsymbol{v} \in \mathbb{R}^n$*, designada ainda por $\mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a})$, $\partial f/\partial \boldsymbol{v}(\boldsymbol{a})$ ou $f'_{\boldsymbol{v}}(\boldsymbol{a})$ será definida precisamente da mesma maneira:

$$\frac{\partial f}{\partial \boldsymbol{v}}(\boldsymbol{a}) = \lim_{t\to 0} \frac{f(\boldsymbol{a}+t\boldsymbol{v}) - f(\boldsymbol{a})}{t},$$

se existir o limite indicado no segundo membro (o qual será agora um vector do espaço $\mathbb{R}^m$). Designando por f_i a função coordenada de ordem i de f, decorre imediatamente do Teorema 3.2' (parágrafo 3.2) que, para que exista $\partial f/\partial \boldsymbol{v}(\boldsymbol{a})$ é necessário e suficiente que existam (e sejam finitas) as derivadas $\partial f_i/\partial \boldsymbol{v}(\boldsymbol{a})$, para $i = 1, \ldots, m$; em tal hipótese, $\partial f_i/\partial \boldsymbol{v}(\boldsymbol{a})$ será a coordenada de ordem i do vector $\partial f/\partial \boldsymbol{v}(\boldsymbol{a})$.

As propriedades do conceito de derivada segundo um vector, atrás indicadas para o caso das funções de duas variáveis, mantêm-se na situação mais geral agora considerada, com os ajustamentos evidentes.

No caso do vector $\boldsymbol{v} \in \mathbb{R}^n$ ser unitário, costuma ainda dizer-se que $\partial f/\partial \boldsymbol{v}(\boldsymbol{a})$ é a *derivada direccional de f no ponto $\boldsymbol{a}$, na direcção e sentido do vector $\boldsymbol{v}$*. Quando, em particular, $\boldsymbol{v}$ coincide com o vector $\boldsymbol{e}_j$ da base canónica de $\mathbb{R}^n$, obtém-se a derivada parcial de ordem j da função f no ponto $\boldsymbol{a}$, que pode ser designada pelos símbolos $\mathcal{D}_j f(\boldsymbol{a})$, $\partial f/\partial x_j(\boldsymbol{a})$, $f'_{x_j}(\boldsymbol{a})$, etc. ($j = 1, \ldots, n$).

Pondo $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{a} = (a_1, \ldots, a_n)$, ter-se-á evidentemente:

$$\frac{\partial f}{\partial x_j}(a_1, \ldots, a_n) = \lim_{t\to 0} \frac{f(a_1, \ldots, a_{j-1}, a_j + t, a_{j+1}, \ldots, a_n) - f(a_1, \ldots, a_j, \ldots, a_n)}{t},$$

sempre que o limite exista; nos casos mais correntes na prática, o cálculo da derivada parcial $\partial f/\partial x_j$ de uma função vectorial num dado ponto, faz-se separadamente para cada uma das suas funções coordenadas, utilizando as regras usuais na derivação em ordem a x_j e considerando todas as outras variáveis como se fossem constantes.

Por exemplo, para a função que designámos por μ no parágrafo 3.1, definida pelo sistema:

$$\begin{cases} x = r\cos\theta \\ y = r\operatorname{sen}\theta \end{cases}$$

ter-se-á, em qualquer ponto (r, θ) do seu domínio:

$$\begin{cases} \frac{\partial x}{\partial r} = \cos\theta \\ \frac{\partial y}{\partial r} = \operatorname{sen}\theta \end{cases} \qquad \begin{cases} \frac{\partial x}{\partial \theta} = -r\operatorname{sen}\theta \\ \frac{\partial y}{\partial \theta} = r\cos\theta \end{cases}$$

Trataremos agora de generalizar ao caso das funções de n variáveis reais a noção fundamental de função diferenciável. Como teremos oportunidade de ver pelas propriedades que estabeleceremos posteriormente, a «boa definição» é a que ficou claramente sugerida nas considerações finais do parágrafo 4.1.

Consideremos em primeiro lugar o caso de uma função real f, definida num conjunto $D \subset \mathbb{R}^n$: dizer que f é diferenciável no ponto $\boldsymbol{a}$ equivalerá a dizer que o acréscimo $f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})$ pode ser aproximado por uma função linear de $\boldsymbol{h}$, com um erro que será um infinitésimo de ordem superior à primeira quando $\boldsymbol{h} \to \mathbf{0}$.

Em termos precisos: sendo $\boldsymbol{a}$ um ponto interior ao domínio D da função f, diz-se que f é *diferenciável no ponto* $\boldsymbol{a}$ sse existir uma aplicação linear $L_{\boldsymbol{a}} : \mathbb{R}^n \to \mathbb{R}$ tal que se tenha:

$$f(\boldsymbol{a}+\boldsymbol{h}) = f(\boldsymbol{a}) + L_{\boldsymbol{a}}(\boldsymbol{h}) + o(\|\boldsymbol{h}\|)$$

em todo o ponto $\boldsymbol{h}$ tal que $\boldsymbol{a}+\boldsymbol{h} \in D$.

De maneira obviamente equivalente, poderá dizer-se que f é diferenciável em $\boldsymbol{a}$ sse existe uma aplicação linear $L_{\boldsymbol{a}} : \mathbb{R}^n \to \mathbb{R}$ e uma função $\varphi : D \to \mathbb{R}$, infinitésima quando $\boldsymbol{x} \to \boldsymbol{a}$, por forma que se verifique a igualdade:

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + L_{\boldsymbol{a}}(\boldsymbol{x}-\boldsymbol{a}) + \|\boldsymbol{x}-\boldsymbol{a}\|\varphi(\boldsymbol{x}),$$

em todo o ponto $\boldsymbol{x} \in D$.

É sabido que se estabelece uma correspondência bijectiva entre o conjunto das aplicações lineares de $\mathbb{R}^n$ em $\mathbb{R}$ e o conjunto das matrizes (linha) do tipo $1 \times n$ se se associar a cada uma de tais aplicações, L, a matriz $[c_1\ c_2\ \ldots\ c_n]$ tal que $c_1 = L(\boldsymbol{e}_1), \ldots, c_n = L(\boldsymbol{e}_n)$ (sendo $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$, os vectores da base canónica de $\mathbb{R}^n$); e também que, se for $\boldsymbol{x} = (x_1, \ldots, x_n)$ um vector qualquer de $\mathbb{R}^n$, se terá então:

$$L(\boldsymbol{x}) = c_1x_1 + \cdots + c_nx_n.$$

Assim, poderia ainda dizer-se que f é diferenciável no ponto $\boldsymbol{a} \in \operatorname{int} D$ sse existem números reais $\alpha_1, \ldots, \alpha_n$ tais que, sempre que o ponto $\boldsymbol{a}+\boldsymbol{h} = (a_1 + h_1, \cdots, a_n + h_n)$ pertença a D, se tenha:

$$f(a_1+h_1, \ldots, a_n+h_n) = f(a_1, \ldots, a_n) + \alpha_1h_1 + \cdots + \alpha_nh_n + o\left(\sqrt{h_1^2 + \cdots + h_n^2}\right).$$

Antes de passarmos ao caso das funções vectoriais, vejamos um exemplo: a função

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

para $(x,y,z) \in \mathbb{R}^3$, é diferenciável no ponto $(0,0,0)$. Basta observar que se tem, para $x,y,z \in \mathbb{R}$:

$$f(x,y,z) = 3 + 2x - y + o\left(\sqrt{x^2+y^2+z^2}\right)$$

ou

$$f(x,y,z) = f(0,0,0) + L_o[(x,y,z)] + o\left(\sqrt{x^2+y^2+z^2}\right),$$

designando por L_o a aplicação linear de $\mathbb{R}^3$ em $\mathbb{R}$ correspondente à matriz $[2\ -1\ 0]$.

A extensão do conceito de diferenciabilidade às funções vectoriais é agora imediata. Sendo $f : D \to \mathbb{R}^m$, com $D \subset \mathbb{R}^n$ e $\boldsymbol{a} \in \operatorname{int} D$, diremos que f é diferenciável no ponto $\boldsymbol{a}$ sse existir uma aplicação linear $L_{\boldsymbol{a}} : \mathbb{R}^n \to \mathbb{R}^m$ tal que, em todo o ponto $\boldsymbol{h}$ tal que $\boldsymbol{a}+\boldsymbol{h} \in D$, se tenha:

$$f(\boldsymbol{a}+\boldsymbol{h}) = f(\boldsymbol{a}) + L_{\boldsymbol{a}}(\boldsymbol{h}) + o\big(\|\boldsymbol{h}\|\big);$$

ou, de modo equivalente, sse existir uma matriz de elementos reais,

$$M_{\boldsymbol{a}} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \multicolumn{4}{c}{\dotfill} \\ \alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn} \end{bmatrix}$$

e uma função $\varphi : D \to \mathbb{R}^m$, infinitésima no ponto $\boldsymbol{a}$, tais que se verifique, em todo o ponto $\boldsymbol{x} \in D$, a igualdade:

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + M_{\boldsymbol{a}}(\boldsymbol{x}-\boldsymbol{a}) + \|\boldsymbol{x}-\boldsymbol{a}\|\varphi(\boldsymbol{x}),$$

(com a interpretação óbvia dos vectores $f(\boldsymbol{x})$, $f(\boldsymbol{a})$, $\varphi(\boldsymbol{x})$ e $(\boldsymbol{x}-\boldsymbol{a})$ como matrizes coluna).

Em termos de coordenadas, esta igualdade traduzir-se-á pelo sistema de m equações:

$$\begin{aligned} f_i(x_1,\dots,x_n) &= f_i(a_1,\dots,a_n) + \alpha_{i1}(x_1-a_1) + \cdots \\ &\cdots + \alpha_{in}(x_n-a_n) + \|\boldsymbol{x}-\boldsymbol{a}\|\varphi_i(x_1,\dots,x_n) \qquad (i = 1,\dots,m). \end{aligned}$$

Tendo em conta que $\varphi : D \to \mathbb{R}^m$ é infinitésima quando $\boldsymbol{x} \to \boldsymbol{a}$ sse cada uma das suas funções coordenadas φ_i o for, pode então concluir-se imediatamente que:

Teorema 4.1. *Seja $f : D \to \mathbb{R}^m$ (com $D \subset \mathbb{R}^n$) e $\boldsymbol{a} \in \operatorname{int} D$; para que f seja diferenciável no ponto $\boldsymbol{a}$ é necessário e suficiente que cada uma das suas funções coordenadas seja diferenciável no mesmo ponto.*

O esforço feito para obtermos uma boa definição de diferenciabilidade irá agora ser compensado com um série de propriedades enunciadas nos teoremas seguintes, que inclui praticamente todas as que poderíamos considerar desejáveis:

Teorema 4.2. *Se f é diferenciável no ponto $\boldsymbol{a}$:*

1. *f é contínua em $\boldsymbol{a}$,*
2. *para todo o vector $\boldsymbol{v} \in \mathbb{R}^n$, existe a derivada $\mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a})$.*

Demonstração.

1. Sendo f diferenciável no ponto $\boldsymbol{a}$, ter-se-á para todo o $\boldsymbol{x} \in D$:

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + L_{\boldsymbol{a}}(\boldsymbol{x}) - L_{\boldsymbol{a}}(\boldsymbol{a}) + \|\boldsymbol{x} - \boldsymbol{a}\|\varphi(\boldsymbol{x})$$

(sendo $L_{\boldsymbol{a}}$ uma aplicação linear de $\mathbb{R}^n$ em $\mathbb{R}^m$ e $\varphi : D \to \mathbb{R}^m$ um infinitésimo quando $\boldsymbol{x} \to \boldsymbol{a}$); e é claro que cada um dos termos do 2º membro é uma função contínua no ponto $\boldsymbol{a}$ (o primeiro e o terceiro por serem constantes; o segundo, porque, como vimos oportunamente, uma aplicação linear é contínua em qualquer ponto; e o último por ser o produto de uma função escalar contínua em todo o seu domínio pelo infinitésimo φ, obviamente contínuo no ponto $\boldsymbol{a}$).

2. Seja $\boldsymbol{v}$ um vector qualquer de $\mathbb{R}^n$; substituindo, na última igualdade anterior, $\boldsymbol{x}$ por $\boldsymbol{a} + t\boldsymbol{v}$ (o que é legítimo, pelo menos para valores suficientemente pequenos de $|t|$, por $\boldsymbol{a}$ ser interior ao domínio D das funções f e φ) obtém-se:

$$f(\boldsymbol{a} + t\boldsymbol{v}) = f(\boldsymbol{a}) + tL_{\boldsymbol{a}}(\boldsymbol{v}) + |t|\|\boldsymbol{v}\|\varphi(\boldsymbol{a} + t\boldsymbol{v})$$

ou, supondo agora também $t \neq 0$:

$$\frac{f(\boldsymbol{a} + t\boldsymbol{v}) - f(\boldsymbol{a})}{t} = L_{\boldsymbol{a}}(\boldsymbol{v}) + \|\boldsymbol{v}\|\frac{|t|}{t}\varphi(\boldsymbol{a} + t\boldsymbol{v})$$

Quando $t \to 0$, a segunda parcela do 2º membro (produto da função escalar limitada $\|\boldsymbol{v}\|\frac{|t|}{t}$ por $\varphi(\boldsymbol{a}+t\boldsymbol{v})$, que tende evidentemente para $\mathbf{0}$ quando $t \to 0$) é infinitésima e a primeira é constante; existe portanto $\mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a})$ e verifica-se a igualdade:

$$\mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a}) = L_{\boldsymbol{a}}(\boldsymbol{v}).$$

□

Corolário. *Sendo $f : D \to \mathbb{R}^m$ $(D \subset \mathbb{R}^n)$ uma função diferenciável no ponto $\boldsymbol{a}$, existe uma única aplicação linear $L_{\boldsymbol{a}}$ tal que:*

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + L_{\boldsymbol{a}}(\boldsymbol{x} - \boldsymbol{a}) + o\left(\|\boldsymbol{x} - \boldsymbol{a}\|\right),$$

para todo o $\boldsymbol{x}$ no domínio de f, e a matriz correspondente a $L_{\boldsymbol{a}}$ é a matriz:

$$M_{\boldsymbol{a}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{a}) \\ \cdots\cdots & \cdots & \cdots\cdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}$$

(onde f_i é a i^a função coordenada de f).

Demonstração. Atendendo à igualdade final da demonstração do Teorema 4.2:

$$L_{\boldsymbol{a}}(\boldsymbol{v}) = \mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a})$$

e ao facto de a derivada de f segundo um vector $\boldsymbol{v}$ ser única (quando existe), logo se reconhece que, nas condições da hipótese, fica univocamente determinado o valor da aplicação $L_{\boldsymbol{a}}$ em cada vector $\boldsymbol{v} \in \mathbb{R}^n$, o que prova a unicidade de $L_{\boldsymbol{a}}$. Por outro lado (sendo $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ os vectores da base canónica de $\mathbb{R}^n$) os elementos da coluna de ordem j da matriz $M_{\boldsymbol{a}}$ devem ser as coordenadas, na base canónica de $\mathbb{R}^m$, do vector $L_{\boldsymbol{a}}(\boldsymbol{e}_j) = \mathcal{D}_{\boldsymbol{e}_j} f(\boldsymbol{a}) = \partial f/\partial x_j(\boldsymbol{a})$; e já sabemos que essas coordenadas são precisamente as derivadas parciais, $\partial f_i/\partial x_j(\boldsymbol{a})$ $(i = 1, \ldots, m)$ das funções coordenadas de f. □

Registaremos agora algumas definições importantes:

Quando f é diferenciável no ponto $\boldsymbol{a}$, chama-se *derivada de f no ponto $\boldsymbol{a}$*, e designa-se por $f'(\boldsymbol{a})$, a (única) aplicação linear $L_{\boldsymbol{a}}$ que verifica a condição expressa no enunciado do corolário anterior; tem-se, portanto, em qualquer ponto $\boldsymbol{x}$ do domínio de f:

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + f'(\boldsymbol{a})(\boldsymbol{x} - \boldsymbol{a}) + o\left(\|\boldsymbol{x} - \boldsymbol{a}\|\right).$$

À matriz $M_{\boldsymbol{a}}$, correspondente à aplicação $f'(\boldsymbol{a})$, chama-se *matriz jacobiana de f no ponto $\boldsymbol{a}$.*

Sendo $\boldsymbol{h}$ um vector arbitrário de $\mathbb{R}^n$, ao valor da aplicação $f'(\boldsymbol{a})$ no ponto $\boldsymbol{h}$, $f'(\boldsymbol{a})(\boldsymbol{h})$, costuma-se chamar *diferencial da função f no ponto $\boldsymbol{a}$ relativo ao vector $\boldsymbol{h}$*, por vezes designado por $df_{\boldsymbol{a}}(\boldsymbol{h})$; porém, tendo em conta as igualdades:

$$df_{\boldsymbol{a}}(\boldsymbol{h}) = f'(\boldsymbol{a})(\boldsymbol{h}) = L_{\boldsymbol{a}}(\boldsymbol{h}) = \mathcal{D}_{\boldsymbol{h}} f(\boldsymbol{a}),$$

logo se vê que o diferencial de f relativo a um vector qualquer $\boldsymbol{h}$ não é mais do que a derivada da função f segundo esse mesmo vector.

Antes de prosseguirmos na descrição de propriedades importantes da noção de diferenciabilidade, convirá talvez destacar alguns aspectos e casos particulares significativos que decorrem das ideias já expostas e ver alguns exemplos.

Em primeiro lugar recorde-se que, numa observação anterior, vimos que era possível em geral existirem as derivadas $\mathcal{D}_{\boldsymbol{v}_1} f(\boldsymbol{a})$ e $\mathcal{D}_{\boldsymbol{v}_2} f(\boldsymbol{a})$ sem existir, ou existindo com valor diferente da soma daquelas, a derivada $\mathcal{D}_{\boldsymbol{v}_1+\boldsymbol{v}_2} f(\boldsymbol{a})$. Tal não poderá verificar-se, porém, se f for diferenciável no ponto $\boldsymbol{a}$; com efeito, a igualdade:

$$\mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a}) = L_{\boldsymbol{a}}(\boldsymbol{v})$$

mostra precisamente que, fixada a função f e o ponto $\boldsymbol{a}$ (no qual f seja diferenciável) $\mathcal{D}_{\boldsymbol{v}} f(\boldsymbol{a})$ é uma função linear de $\boldsymbol{v}$.

Consideremos agora o caso de f ser uma função real, diferenciável no ponto $\boldsymbol{a} = (a_1, \dots, a_n) \in \operatorname{int} D$; sendo $\boldsymbol{h} = (h_1, \dots, h_n)$ um vector tal que $\boldsymbol{a} + \boldsymbol{h} \in D$ ter-se-á, como vimos:

$$\begin{aligned} f(a_1 + h_1, \dots, a_n + h_n) = f(a_1, \dots, a_n) + \frac{\partial f}{\partial x_1}(a_1, \dots, a_n) h_1 + \cdots \\ \cdots + \frac{\partial f}{\partial x_n}(a_1, \dots, a_n) h_n + o(\|h\|). \end{aligned}$$

Ao vector que tem por coordenadas, na base canónica de $\mathbb{R}^n$, as derivadas parciais $\partial f / \partial x_1(\boldsymbol{a}), \dots, \partial f / \partial x_n(\boldsymbol{a})$ isto é, ao vector:

$$\frac{\partial f}{\partial x_1}(\boldsymbol{a}) \boldsymbol{e}_1 + \cdots + \frac{\partial f}{\partial x_n}(\boldsymbol{a}) \boldsymbol{e}_n,$$

costuma-se chamar *gradiante de f no ponto $\boldsymbol{a}$*; designá-lo-emos por $\nabla f(\boldsymbol{a})$ ou $\operatorname{grad} f(\boldsymbol{a})$. Poderá assim escrever-se:

$$f(\boldsymbol{a} + \boldsymbol{h}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a}) \cdot \boldsymbol{h} + o(\|\boldsymbol{h}\|)$$

onde $\nabla f(\boldsymbol{a}) \cdot \boldsymbol{h}$ designa o produto interno dos vectores $\nabla f(\boldsymbol{a})$ e $\boldsymbol{h}$; claro que este produto interno é precisamente a derivada de f no ponto $\boldsymbol{a}$ segundo o vector $\boldsymbol{h}$:

$$\mathcal{D}_{\boldsymbol{h}} f(\boldsymbol{a}) = \nabla f(\boldsymbol{a}) \cdot \boldsymbol{h}.$$

Como exemplos triviais de funções diferenciáveis surgem, naturalmente, as constantes; em termos um pouco mais gerais, consideremos uma função f cujo domínio $D \subset \mathbb{R}^n$ contenha uma bola $B_\epsilon(\boldsymbol{a})$ e que assuma em todos os pontos desta bola um valor constante $\boldsymbol{c} \in \mathbb{R}^m$. Ter-se-á então, em qualquer ponto $\boldsymbol{x} \in D$:

$$f(\boldsymbol{x}) = \boldsymbol{c} + o(\|\boldsymbol{x} - \boldsymbol{a}\|)$$

ou, designando por $\tilde{0}$ a aplicação (linear) identicamente nula de $\mathbb{R}^n$ em $\mathbb{R}^m$,

$$f(\boldsymbol{x}) = f(\boldsymbol{a}) + \tilde{0}(\boldsymbol{x} - \boldsymbol{a}) + o(\|\boldsymbol{x} - \boldsymbol{a}\|);$$

esta igualdade mostra que f é diferenciável no ponto $\boldsymbol{a}$ e que $f'(\boldsymbol{a})$ é a aplicação nula.

Outros exemplos simples de funções diferenciáveis são facultados pelas próprias aplicações lineares; sendo $g : \mathbb{R}^n \to \mathbb{R}^m$ uma tal aplicação e $\boldsymbol{a}$, $\boldsymbol{x}$ pontos de $\mathbb{R}^n$, decorre da igualdade:

$$g(\boldsymbol{x}) = g(\boldsymbol{a}) + g(\boldsymbol{x} - \boldsymbol{a}) + o(\|\boldsymbol{x} - \boldsymbol{a}\|)$$

(onde, desta vez, o símbolo $o(\|\boldsymbol{x}-\boldsymbol{a}\|)$ está de facto a representar o vector nulo de $\mathbb{R}^m$) que g é diferenciável em $\boldsymbol{a}$ — ponto arbitrário de $\mathbb{R}^n$ — tendo-se precisamente:

$$g'(\boldsymbol{a}) = g.$$

Assim, a derivada de uma aplicação linear coincide com a própria aplicação (e é portanto independente do ponto $\boldsymbol{a}$ considerado).

Em particular, para as projecções p_j, isto é, para as aplicações lineares de $\mathbb{R}^n$ em $\mathbb{R}$ definidas por:

$$p_j(\boldsymbol{x}) = p_j(x_1, \ldots, x_n) = x_j,$$

(para $j = 1, \ldots, n$) tem-se, em qualquer ponto $\boldsymbol{a} \in \mathbb{R}^n$,

$$p_j'(\boldsymbol{a}) = p_j$$

e o diferencial de p_j relativo a um vector $\boldsymbol{h} = (h_1, \ldots, h_n)$ — diferencial que é também independente do ponto $\boldsymbol{a}$, cuja indicação explícita nas notações poderá portanto ser omitida — é dado por:

$$dp_j(\boldsymbol{h}) = p_j(\boldsymbol{h}) = h_j.$$

Convém lembrar agora que, na prática, a projecção p_j é muitas vezes designada de preferência pelo símbolo x_j, correspondente ao valor por ela assumido no ponto $\boldsymbol{x} = (x_1, \ldots, x_n)$ (trata-se uma vez mais do abuso de notação corrente que consiste em usar o símbolo $f(\boldsymbol{x})$ para designar a função f); adoptando este abuso de notação, a fórmula precedente assumiria a forma:

$$dx_j(\boldsymbol{h}) = h_j,$$

a que recorremos dentro em pouco.

Voltemos a considerar uma função real f, diferenciável num ponto $\boldsymbol{a} \in \mathbb{R}^n$ e a expressão já conhecida do seu diferencial:

$$df_{\boldsymbol{a}}(\boldsymbol{h}) = \frac{\partial f}{\partial x_1}(\boldsymbol{a})h_1 + \cdots + \frac{\partial f}{\partial x_n}(\boldsymbol{a})h_n.$$

Como, de acordo com as convenções acabadas de mencionar, se tem $h_j = dx_j(\boldsymbol{h})$ para $j = 1, \ldots, n$, esta expressão poderá também escrever-se:

$$df_{\boldsymbol{a}}(\boldsymbol{h}) = \frac{\partial f}{\partial x_1}(\boldsymbol{a})dx_1(\boldsymbol{h}) + \cdots + \frac{\partial f}{\partial x_n}(\boldsymbol{a})dx_n(\boldsymbol{h}).$$

Na prática, esta fórmula — por vezes chamada «fórmula do diferencial total» — escreve-se habitualmente de modo mais abreviado, omitindo-se a indicação do ponto $\boldsymbol{a}$ (em que f deverá ser diferenciável) e do vector $\boldsymbol{h} \in \mathbb{R}^n$:

$$df = \frac{\partial f}{\partial x_1}dx_1 + \cdots + \frac{\partial f}{\partial x_n}dx_n.$$

Assim, por exemplo, para $f(\boldsymbol{x}) = \log \|\boldsymbol{x}\|$ (com $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n \backslash \{0\}$) ter-se-á, em qualquer ponto $\boldsymbol{x}$ em que f seja diferenciável:[2]

$$\begin{aligned} df &= d\left(\log \sqrt{x_1^2 + \cdots + x_n^2}\right) \\ &= \frac{x_1}{x_1^2 + \cdots + x_n^2} dx_1 + \cdots + \frac{x_n}{x_1^2 + \cdots + x_n^2} dx_n \\ &= \frac{x_1 dx_1 + \cdots + x_n dx_n}{\|\boldsymbol{x}\|^2}, \end{aligned}$$

o que pode também escrever-se, mais simplesmente,

$$df = \frac{\boldsymbol{x} \cdot d\boldsymbol{x}}{\|\boldsymbol{x}\|^2}.$$

Outro exemplo: para a função definida em $\mathbb{R}^2$ pela fórmula:

$$p(x, y) = xy,$$

(que, como veremos, é diferenciável em qualquer ponto de $\mathbb{R}^2$ e cujo valor num ponto (x, y) tal que $x > 0$ e $y > 0$ representa a área do rectângulo de base x e altura y) tem-se:

$$dp = \frac{\partial p}{\partial x}\, dx + \frac{\partial p}{\partial y}\, dy = y\, dx + x\, dy.$$

Se os valores de dx e dy forem «pequenos» em relação a x e y, o diferencial dp dará uma «boa aproximação» do acréscimo da área do rectângulo correspondente à substituição da «base» x por $x + dx$ e da «altura» y por $y + dy$ (Figura 4.4): o erro correspondente a essa aproximação, $dx\, dy$, será um infinitésimo de ordem superior á primeira se o «acréscimo» (dx, dy) tender para $(0, 0)$.

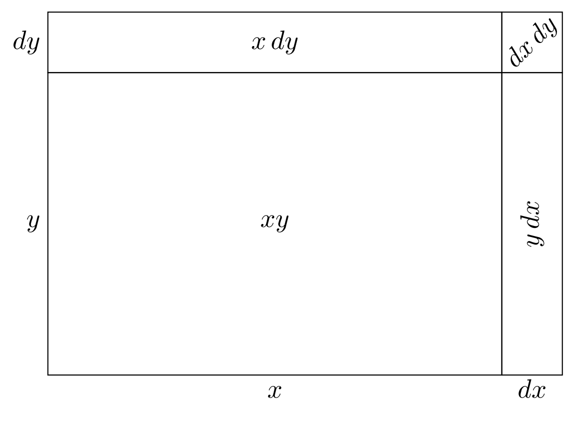

Figura 4.4

[2]Veremos posteriormente que f é diferenciável em qualquer ponto do seu domínio.

Naturalmente, a fórmula do diferencial total estende-se, (de maneira óbvia, para funções vectoriais, reconhecendo-se imediatamente que para uma função $f = (f_1, \ldots, f_m)$ diferenciável em dado ponto $\boldsymbol{x} \in \mathbb{R}^n$ poderá escrever-se, em notação abreviada análoga à que indicámos para o caso das funções reais:

$$\begin{cases} df_1 = \frac{\partial f_1}{\partial x_1} dx_1 + \cdots + \frac{\partial f_1}{\partial x_n} dx_n \\ \quad \cdots \\ df_m = \frac{\partial f_m}{\partial x_1} dx_1 + \cdots + \frac{\partial f_m}{\partial x_n} dx_n \end{cases}$$

Em termos matriciais, este sistema corresponderá à igualdade:

$$\begin{bmatrix} df_1 \\ \ldots \\ df_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \multicolumn{3}{c}{\ldots\ldots\ldots\ldots\ldots} \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ \ldots \\ dx_n \end{bmatrix},$$

que pode representar-se também, de forma mais sintética e designando por $M_{\boldsymbol{x}}$ a matriz jacobiana de f no ponto $\boldsymbol{x}$ considerado:

$$df = M_{\boldsymbol{x}} d\boldsymbol{x}.$$

Outra forma de representar abreviadamente o sistema acima indicado seria escrever:

$$df = f'(\boldsymbol{x})(d\boldsymbol{x})$$

ou, mais simplesmente:

$$df = f'(\boldsymbol{x})d\boldsymbol{x}$$

(nesta última forma, convém notar que o segundo membro não designa propriamente um «produto», mas sim o valor que a aplicação linear $f'(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}^m$ assume em $d\boldsymbol{x} \in \mathbb{R}^n$).

Para evitar qualquer possibilidade de equívoco convirá talvez observar que, nas notações mais precisas que adoptámos de início, quando definimos a noção de diferencial de uma função, as fórmulas precedentes deveriam escrever-se:

$$df_{\boldsymbol{x}}(d\boldsymbol{x}) = f'(\boldsymbol{x})(d\boldsymbol{x}).$$

Do ponto de vista prático, porém, há toda a vantagem em nos habituarmos ao uso das notações simplificadas que são mais correntes, (sem permitir no entanto que daí resulte qualquer prejuízo para a precisão e clareza das ideias).

Como outro exemplo do uso das notações abreviadas usuais, consideremos a função $\mu : \mathbb{R}^3 \to \mathbb{R}^3$ definida pelo sistema:

$$\begin{aligned} x &= \mu_1(r, \theta, \varphi) = r \cos\theta \cos\varphi \\ y &= \mu_2(r, \theta, \varphi) = r \cos\theta \operatorname{sen}\varphi \\ z &= \mu_3(r, \theta, \varphi) = r \operatorname{sen}\theta. \end{aligned}$$

Neste caso o diferencial $d\mu$ (em qualquer ponto em que μ seja diferenciável[3]) ficará determinado pelo sistema:

$$\begin{aligned} dx &= \cos\theta\cos\varphi\, dr - r\,\mathrm{sen}\,\theta\cos\varphi\, d\theta - r\cos\theta\,\mathrm{sen}\,\varphi\, d\varphi \\ dy &= \cos\theta\,\mathrm{sen}\,\varphi\, dr - r\,\mathrm{sen}\,\theta\,\mathrm{sen}\,\varphi\, d\theta + r\cos\theta\cos\varphi\, d\varphi \\ dz &= \mathrm{sen}\,\theta\, dr + r\cos\theta\, d\theta. \end{aligned}$$

Retomaremos agora o estudo das propriedades gerais das funções diferenciáveis. Seja $D \subset \mathbb{R}^n$ e $\boldsymbol{a}$ um ponto interior a D. Se $f, g : D \to \mathbb{R}^m$ são funções diferenciáveis no ponto $\boldsymbol{a}$, ter-se-á, em todo o ponto $\boldsymbol{h}$ tal que $\boldsymbol{a} + \boldsymbol{h} \in D$:

$$\begin{aligned} f(\boldsymbol{a}+\boldsymbol{h}) &= f(\boldsymbol{a}) + f'(\boldsymbol{a})(\boldsymbol{h}) + o\big(\|\boldsymbol{h}\|\big) \\ g(\boldsymbol{a}+\boldsymbol{h}) &= g(\boldsymbol{a}) + g'(\boldsymbol{a})(\boldsymbol{h}) + o\big(\|\boldsymbol{h}\|\big) \end{aligned}$$

Adicionando membro a membro estas igualdades — e tendo em conta que, segundo as notações introduzidas no final do parágrafo 3.2, $o(\|\boldsymbol{h}\|) + o(\|\boldsymbol{h}\|) = o(\|\boldsymbol{h}\|)$ — obtém-se:

$$(f+g)(\boldsymbol{a}+\boldsymbol{h}) = (f+g)(\boldsymbol{a}) + \left[f'(\boldsymbol{a}) + g'(\boldsymbol{a})\right](\boldsymbol{h}) + o\big(\|\boldsymbol{h}\|\big)$$

relação que prova a diferenciabilidade de $f+g$ no ponto $\boldsymbol{a}$ e mostra ainda que $(f+g)'(\boldsymbol{a}) = f'(\boldsymbol{a}) + g'(\boldsymbol{a})$; por outro lado, multiplicando ambos os membros da primeira daquelas igualdades por um escalar arbitrário α, obtém-se imediatamente:

$$(\alpha f)(\boldsymbol{a}+\boldsymbol{h}) = (\alpha f)(\boldsymbol{a}) + \left[\alpha f'(\boldsymbol{a})\right](\boldsymbol{h}) + o\big(\|\boldsymbol{h}\|\big),$$

donde se infere que a função αf é diferenciável em $\boldsymbol{a}$ e que $(\alpha f)'(\boldsymbol{a}) = \alpha f'(\boldsymbol{a})$.

Teorema 4.3. *O conjunto das funções $f : D \to \mathbb{R}^m$ que são diferenciáveis no ponto $\boldsymbol{a}$ (munido das operações usuais de adição de funções e de multiplicação de escalares por funções) é um espaço vectorial real. Sendo f e g duas funções deste espaço e $\alpha \in \mathbb{R}$ tem-se:*

$$\begin{aligned} (f+g)'(\boldsymbol{a}) &= f'(\boldsymbol{a}) + g'(\boldsymbol{a}) \\ (\alpha f)'(\boldsymbol{a}) &= \alpha f'(\boldsymbol{a}). \end{aligned}$$

Sendo E e F dois espaços vectoriais reais, designa-se frequentemente por $L(E, F)$ o conjunto formado por todas as aplicações lineares de E em F; e é bem fácil reconhecer que $L(E, F)$ fica munido de uma estrutura de espaço vectorial real se, como é natural, definirmos a soma de duas aplicações $u, v \in L(E, F)$ como sendo a aplicação (linear) $u + v : E \to F$ tal que

$$(u+v)(\boldsymbol{x}) = u(\boldsymbol{x}) + v(\boldsymbol{x}) \qquad (\forall \boldsymbol{x} \in E)$$

e o produto do número real α pela aplicação u pela fórmula:

$$(\alpha u)(\boldsymbol{x}) = \alpha u(\boldsymbol{x}) \qquad (\forall \boldsymbol{x} \in E).$$

[3] Veremos adiante que μ é diferenciável em qualquer ponto $(r, \theta, \varphi) \in \mathbb{R}^3$.

Nestas condições, as duas igualdades finais do enunciado do Teorema 4.3 poderiam também exprimir-se dizendo que a aplicação que associa a cada função f, do espaço vectorial considerado nesse enunciado, a sua derivada no ponto $\boldsymbol{a}$, $f'(\boldsymbol{a})$, é uma aplicação linear desse espaço em $L(\mathbb{R}^n, \mathbb{R}^m)$.

Para a sequência, convém-nos ainda recordar outro resultado simples de Álgebra Linear, relativo à representação matricial da composta de duas aplicações lineares.

Sejam m, n e p três números inteiros positivos, $u \in L(\mathbb{R}^p, \mathbb{R}^m)$ e $v \in L(\mathbb{R}^n, \mathbb{R}^p)$, as aplicações lineares correspondentes às matrizes:

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1p} \\ \multicolumn{3}{c}{\dotfill} \\ u_{m1} & \cdots & u_{mp} \end{bmatrix} \qquad \text{e} \qquad V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \multicolumn{3}{c}{\dotfill} \\ v_{p1} & \cdots & v_{pn} \end{bmatrix},$$

nas bases canónicas de $\mathbb{R}^n$, $\mathbb{R}^p$ e $\mathbb{R}^m$, que representaremos respectivamente por $(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n)$, $(\bar{\boldsymbol{e}}_1, \dots, \bar{\boldsymbol{e}}_p)$ e $(\boldsymbol{e}_1^*, \dots, \boldsymbol{e}_m^*)$. Reconhece-se sem qualquer dificuldade que $w = u \circ v$ é uma aplicação linear de $\mathbb{R}^n$ em $\mathbb{R}^m$, à qual corresponderá então certa matriz:

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \multicolumn{3}{c}{\dotfill} \\ w_{m1} & \cdots & w_{mn} \end{bmatrix}.$$

Para ver como pode obter-se W a partir de U e V basta ter em conta que (como relembrámos em nota, páginas 45–46) deverá ter-se:

$$\begin{aligned} v(\boldsymbol{e}_j) &= \sum_{k=1}^{p} v_{kj} \bar{\boldsymbol{e}}_k \qquad && (j = 1, \dots, n) \\ u(\bar{\boldsymbol{e}}_k) &= \sum_{i=1}^{m} u_{ik} \boldsymbol{e}_i^* \qquad && (k = 1, \dots, p), \end{aligned}$$

e portanto, para cada inteiro positivo $j \le n$:

$$\begin{aligned} (u \circ v)(\boldsymbol{e}_j) &= \sum_{k=1}^{p} u(v_{kj} \bar{\boldsymbol{e}}_k) \\ &= \sum_{k=1}^{p} \sum_{i=1}^{m} v_{kj} u_{ik} \boldsymbol{e}_i^* \\ &= \sum_{i=1}^{m} \left(\sum_{k=1}^{p} u_{ik} v_{kj} \right) \boldsymbol{e}_i^*. \end{aligned}$$

Como, por outro lado, deve ter-se também:

$$(u \circ v)(\boldsymbol{e}_j) = w(\boldsymbol{e}_j) = \sum_{i=1}^{m} w_{ij} \boldsymbol{e}_i^*,$$

pode concluir-se que será:

$$w_{ij} = \sum_{k=1}^{p} u_{ik} v_{kj} \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n),$$

o que corresponde à usual regra de multiplicarão de matrizes, «linhas por colunas» (segundo a qual o elemento da linha i e coluna j da matriz produto, W, é o «produto interno» da «linha i» da matriz U pela «coluna j» da matriz V).

Assim, à igualdade $w = u \circ v$ entre aplicações lineares corresponde, para as matrizes que as representam, a igualdade $W = UV$ (à composição de aplicações corresponde a multiplicação de matrizes).

Obteremos agora um resultado de interesse fundamental: a generalização, ao caso das funções vectoriais de variável vectorial, da regra de derivação das funções compostas (ou regra da cadeia). Consideremos de novo três números naturais m, n e p, um subconjunto D de $\mathbb{R}^n$ e um subconjunto E de $\mathbb{R}^p$; consideremos ainda uma aplicação $g : D \to \mathbb{R}^p$ tal que $g(D) \subset E$ e uma aplicação $f : E \to \mathbb{R}^m$. Nestas condições:

Teorema 4.4. *Se g é diferenciável no ponto $\boldsymbol{a}$ e f é diferenciável no ponto $\boldsymbol{b} = g(\boldsymbol{a})$, $f \circ g$ é diferenciável em $\boldsymbol{a}$ e verifica-se a igualdade:*

$$(f \circ g)'(\boldsymbol{a}) = f'\big(g(\boldsymbol{a})\big) \circ g'(\boldsymbol{a})$$

Demonstração. Sendo g diferenciável no ponto $\boldsymbol{a}$, ter-se-á, sempre que $\boldsymbol{a}+\boldsymbol{h} \in D$:

$$g(\boldsymbol{a}+\boldsymbol{h}) = g(\boldsymbol{a}) + g'(\boldsymbol{a})(\boldsymbol{h}) + \psi(\boldsymbol{h}),$$

com $\psi(\boldsymbol{h}) = o(\|\boldsymbol{h}\|)$; analogamente, do facto de f ser diferenciável em $\boldsymbol{b} = g(\boldsymbol{a})$ resulta que, para todo o $\boldsymbol{k}$ tal que $\boldsymbol{b}+\boldsymbol{k} \in E$:

$$f(\boldsymbol{b}+\boldsymbol{k}) = f(\boldsymbol{b}) + f'(\boldsymbol{b})(\boldsymbol{k}) + \varphi(\boldsymbol{k}),$$

com $\varphi(\boldsymbol{k}) = o(\|\boldsymbol{k}\|)$.

Substituindo nesta igualdade $\boldsymbol{b}$ por $g(\boldsymbol{a})$ e $\boldsymbol{k}$ por $g(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a})$ (o que é legítimo, visto que $\boldsymbol{b}+\boldsymbol{k} = g(\boldsymbol{a}+\boldsymbol{h}) \in E$) obtém-se:

$$(f \circ g)(\boldsymbol{a}+\boldsymbol{h}) = (f \circ g)(\boldsymbol{a}) + f'\big(g(\boldsymbol{a})\big)\big(g(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a})\big) + \varphi\big(g(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a})\big)$$

ou, atendendo a que $g(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a}) = g'(\boldsymbol{a})(\boldsymbol{h}) + \psi(\boldsymbol{h})$:

$$\begin{aligned}(f \circ g)(\boldsymbol{a}+\boldsymbol{h}) = (f \circ g)(\boldsymbol{a}) &+ \left[f'\big(g(\boldsymbol{a})\big) \circ g'(\boldsymbol{a})\right](\boldsymbol{h}) \\ &+ f'\big(g(\boldsymbol{a})\big)\big(\psi(\boldsymbol{h})\big) + \varphi\big(g(\boldsymbol{a}+\boldsymbol{h}) - g(\boldsymbol{a})\big).\end{aligned}$$

Assim, para terminar a demonstração, bastará provar que se tem:

$$f'\big(g(\boldsymbol{a})\big)\big(\psi(\boldsymbol{h})\big) = o(\|\boldsymbol{h}\|)$$

e
$$\varphi\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)=o(\|\boldsymbol{h}\|).$$

Para este efeito, ponhamos $\psi(\boldsymbol{h})=\|\boldsymbol{h}\|\psi^*(\boldsymbol{h})$, para todo o $\boldsymbol{h}\neq\boldsymbol{0}$ tal que $\boldsymbol{a}+\boldsymbol{h}\in D$ (com a função ψ^* nula na origem de $\mathbb{R}^n$) e $\varphi(\boldsymbol{k})=\|\boldsymbol{k}\|\varphi^*(\boldsymbol{k})$, para qualquer $\boldsymbol{k}\neq\boldsymbol{0}$ tal que $\boldsymbol{b}+\boldsymbol{k}\in E$ (com φ também nula no vector nulo de $\mathbb{R}^p$). Ter-se-á $\lim_{\boldsymbol{h}\to\boldsymbol{0}}\psi^*(\boldsymbol{h})=\boldsymbol{0}$ e $\lim_{\boldsymbol{k}\to\boldsymbol{0}}\varphi^*(\boldsymbol{k})=\boldsymbol{0}$ (aliás, sendo $\boldsymbol{k}=g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})$ e g contínua no ponto $\boldsymbol{a}$, por aí ser diferenciável, ter-se-á também $\lim_{\boldsymbol{h}\to\boldsymbol{0}}\boldsymbol{k}=\boldsymbol{0}$ e portanto $\lim_{\boldsymbol{h}\to\boldsymbol{0}}\varphi^*(\boldsymbol{k})=\lim_{\boldsymbol{h}\to\boldsymbol{0}}\varphi^*\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)=\boldsymbol{0}$).

Observe-se ainda que, como $g'(\boldsymbol{a})$ e $f'\big(g(\boldsymbol{a})\big)$ são aplicações lineares (a primeira de $\mathbb{R}^n$ em $\mathbb{R}^p$, a segunda de $\mathbb{R}^p$ em $\mathbb{R}^m$), existem constantes M e N tais que

$$\|g'(\boldsymbol{a})(\boldsymbol{x})\|\leq M\|\boldsymbol{x}\|\quad\text{e}\quad\|f'\big(g(\boldsymbol{a})\big)(\boldsymbol{y})\|\leq N\|\boldsymbol{y}\|,$$

para qualquer $\boldsymbol{x}\in\mathbb{R}^n$ e qualquer $\boldsymbol{y}\in\mathbb{R}^p$.

Nestas condições pode concluir-se, por um lado que

$$\|f'\big(g(\boldsymbol{a})\big)\big(\psi(\boldsymbol{h})\big)\|\leq N\|\psi(\boldsymbol{h})\|\leq N\|\boldsymbol{h}\|\|\psi^*(\boldsymbol{h})\|,$$

(o que mostra que $f'\big(g(\boldsymbol{a})\big)\big(\psi(\boldsymbol{h})\big)=o(\|\boldsymbol{h}\|)$), e por outro que:

$$\begin{aligned}\big\|\varphi\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)\big\| &= \|g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\|\,\big\|\varphi^*\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)\big\|\\ &= \|g'(\boldsymbol{a})(\boldsymbol{h})+\|\boldsymbol{h}\|\psi^*(\boldsymbol{h})\|\,\big\|\varphi^*\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)\big\|\\ &\leq (\|g'(\boldsymbol{a})(\boldsymbol{h})\|+\|\boldsymbol{h}\|\|\psi^*(\boldsymbol{h})\|)\,\big\|\varphi^*\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)\big\|\\ &\leq \|\boldsymbol{h}\|(M+\|\psi^*(\boldsymbol{h})\|)\,\big\|\varphi^*\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)\big\|.\end{aligned}$$

Assim, observando que a função $M+\|\psi^*(\boldsymbol{h})\|$ é limitada numa vizinhança da origem de $\mathbb{R}^n$ e que $\big\|\varphi^*\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)\big\|$ tende para $\boldsymbol{0}$ quando $\boldsymbol{h}\to\boldsymbol{0}$, imediatamente se obtém a relação $\varphi\big(g(\boldsymbol{a}+\boldsymbol{h})-g(\boldsymbol{a})\big)=o(\|\boldsymbol{h}\|)$, que permite considerar a demonstração terminada. □

Observe-se que, nas condições expressas na hipótese do Teorema 4.4, $g'(\boldsymbol{a})$ é uma aplicação linear de $\mathbb{R}^n$ em $\mathbb{R}^p$, $f'\big(g(\boldsymbol{a})\big)$ uma aplicação linear de $\mathbb{R}^p$ em $\mathbb{R}^m$ e portanto a composta, $f'\big(g(\boldsymbol{a})\big)\circ g'(\boldsymbol{a})$, será uma aplicação linear de $\mathbb{R}^n$ em $\mathbb{R}^m$; acabamos de ver precisamente que esta aplicação coincide com a derivada no ponto $\boldsymbol{a}$ da função composta, $f\circ g$. Tendo em conta o resultado que recordámos na nota anterior ao teorema, é agora muito fácil exprimir, em termos das matrizes jacobianas das funções intervenientes, a regra de derivação das funções compostas e obter, a partir dela, as regras correspondentes para o cálculo de derivadas parciais.

Com efeito, seja $\boldsymbol{x}=(x_1,\ldots,x_n)\in D$, $g(\boldsymbol{x})=\boldsymbol{y}=(y_1,\ldots,y_p)$ e $f(\boldsymbol{y})=\boldsymbol{z}=(z_1,\ldots,z_m)$; designando da forma habitual as funções coordenadas de f e g, as matrizes correspondentes às aplicações lineares $g'(\boldsymbol{a})$ e $f'(\boldsymbol{b})$ — com $\boldsymbol{b}=g(\boldsymbol{a})$ —

serão, respectivamente, as matrizes jacobianas:

$$M_{\boldsymbol{a}}(g) = \begin{bmatrix} \frac{\partial g_1}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial g_1}{\partial x_n}(\boldsymbol{a}) \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial g_p}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial g_p}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}$$

$$M_{\boldsymbol{b}}(f) = \begin{bmatrix} \frac{\partial f_1}{\partial y_1}(\boldsymbol{b}) & \cdots & \frac{\partial f_1}{\partial y_p}(\boldsymbol{b}) \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial f_m}{\partial y_1}(\boldsymbol{b}) & \cdots & \frac{\partial f_m}{\partial y_p}(\boldsymbol{b}) \end{bmatrix}.$$

Pondo, por comodidade de escrita, $h = f \circ g$, entre a matriz correspondente a $h'(\boldsymbol{a})$:

$$M_{\boldsymbol{a}}(h) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial h_1}{\partial x_n}(\boldsymbol{a}) \\ \multicolumn{3}{c}{\dotfill} \\ \frac{\partial h_m}{\partial x_1}(\boldsymbol{a}) & \cdots & \frac{\partial h_m}{\partial x_n}(\boldsymbol{a}) \end{bmatrix}$$

e as matrizes $M_{\boldsymbol{a}}(g)$ e $M_{\boldsymbol{b}}(f)$ verificar-se-á então a relação:

$$M_{\boldsymbol{a}}(h) = M_{\boldsymbol{b}}(f)M_{\boldsymbol{a}}(g),$$

isto é, ter-se-á, para qualquer par (i,j) de inteiros positivos tais que $i \le m$ e $j \le n$:

$$\frac{\partial h_i}{\partial x_j}(\boldsymbol{a}) = \frac{\partial f_i}{\partial y_1}(\boldsymbol{b})\frac{\partial g_1}{\partial x_j}(\boldsymbol{a}) + \frac{\partial f_i}{\partial y_2}(\boldsymbol{b})\frac{\partial g_2}{\partial x_j}(\boldsymbol{a}) + \cdots + \frac{\partial f_i}{\partial y_p}(\boldsymbol{b})\frac{\partial g_p}{\partial x_j}(\boldsymbol{a}).$$

Omitindo a referência expressa aos pontos $\boldsymbol{a}$ e $\boldsymbol{b} = g(\boldsymbol{a})$ (nos quais g e f, respectivamente, devem supor-se diferenciáveis) e cometendo o abuso de notação habitual que consiste em substituir as designações mais precisas $\partial h_i/\partial x_j$, $\partial f_i/\partial y_k$ e $\partial g_k/\partial x_j$ por $\partial z_i/\partial x_j$, $\partial z_i/\partial y_k$ e $\partial y_k/\partial x_j$, a fórmula anterior assume o aspecto muito corrente:[4]

$$\frac{\partial z_i}{\partial x_j} = \frac{\partial z_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial z_i}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \cdots + \frac{\partial z_i}{\partial y_p}\frac{\partial y_p}{\partial x_j},$$

$(i = 1, \dots, m;\ j = 1, \dots, n)$. Em termos pouco precisos a regra para o cálculo de derivadas parciais expressa por esta fórmula — também chamada regra da cadeia — pode traduzir-se dizendo que, para obter a derivada de z_i em ordem a uma dada «variável final» x_j, basta somar os produtos que se obtêm multiplicando a derivada de z_i em ordem a cada uma das «variáveis intermédias», y_k, pela derivada desta em ordem à variável x_j.

[4] Observe-se que, nesta fórmula, o símbolo z_i é usado com dois significadas distintos: no primeiro membro designa a coordenada de ordem i da função composta, $z = h(\boldsymbol{x})$; no segundo, a coordenada da mesma ordem da função $z = f(\boldsymbol{y})$.

Quando, em particular, é $n = 1$ — isto é, quando há uma só variável final, x — a fórmula costuma escrever-se:

$$\frac{dz_i}{dx} = \frac{\partial z_i}{\partial y_1}\frac{dy_1}{dx} + \frac{\partial z_i}{\partial y_2}\frac{dy_2}{dx} + \cdots + \frac{\partial z_i}{\partial y_p}\frac{dy_p}{dx} \qquad (i = 1, \ldots, m)$$

(usando os símbolos dz_i/dx, dy_k/dx, em lugar de $\partial z_i/\partial x$, $\partial y_k/\partial x$, para indicar que as derivadas em causa não são parciais, mas «totais»). Quando há apenas uma variável intermédia, $\boldsymbol{y}$, tem-se analogamente:

$$\frac{\partial z_i}{\partial x_j} = \frac{dz_i}{dy}\frac{\partial y}{\partial x_j} \qquad (i = 1, \ldots, m;\ j = 1, \ldots, n).$$

Evidentemente, no caso particular $m = n = p = 1$, obtém-se a fórmula usual no cálculo diferencial de funções de uma variável:

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$

Antes de aplicarmos a regra da cadeia a alguns exemplos concretos, utilizá-la-emos por diversas vezes na demonstração do seguinte teorema:

Teorema 4.5. *Seja $D \subset \mathbb{R}^n$ e f, g duas funções reais definidas em D e diferenciáveis no ponto $\boldsymbol{a}$. Então a função produto fg é diferenciável em $\boldsymbol{a}$ e tem-se:*

$$(fg)'(\boldsymbol{a}) = g(\boldsymbol{a})f'(\boldsymbol{a}) + f(\boldsymbol{a})g'(\boldsymbol{a}).$$

Se, além disso, for $g(\boldsymbol{a}) \neq 0$, o cociente f/g será também diferenciável no ponto $\boldsymbol{a}$, verificando-se a igualdade:

$$\left(\frac{f}{g}\right)'(\boldsymbol{a}) = \frac{g(\boldsymbol{a})f'(\boldsymbol{a}) - f(\boldsymbol{a})g'(\boldsymbol{a})}{\big(g(\boldsymbol{a})\big)^2}.$$

Demonstração. Em primeiro lugar, observemos que a função $p : \mathbb{R}^2 \to \mathbb{R}$ definida por:

$$p(x, y) = xy$$

é diferenciável em qualquer ponto $(\alpha, \beta) \in \mathbb{R}^2$; para o reconhecer, basta atender à definição de diferenciabilidade e ter em conta a igualdade:

$$(\alpha + h)(\beta + k) = \alpha\beta + (\beta h + \alpha k) + hk$$

a qual, designando de momento por L a aplicação linear de $\mathbb{R}^2$ em $\mathbb{R}$ correspondente à matriz $[\beta\ \alpha]$, pode também escrever-se:

$$p(\alpha + h, \beta + k) = p(\alpha\beta) + L(h, k) + o\left(\sqrt{h^2 + k^2}\right).$$

Portanto, p é diferenciável e a matriz que corresponde a $p'(\alpha, \beta)$ é $[\beta\ \alpha]$.

Em segundo lugar, designando por q a aplicação de D em $\mathbb{R}^2$ definida por

$$q(\boldsymbol{x}) = \big(f(\boldsymbol{x}), g(\boldsymbol{x})\big) \qquad (\boldsymbol{x} \in D),$$

notemos que q é diferenciável no ponto $\boldsymbol{a}$, visto que as suas funções coordenadas f e g, são por hipótese diferenciáveis nesse ponto.

Nestas condições, basta observar que, para todo o $\boldsymbol{x} \in D$, se tem:

$$f(\boldsymbol{x})g(\boldsymbol{x}) = p\big(f(\boldsymbol{x}), g(\boldsymbol{x})\big) = (p \circ q)(\boldsymbol{x}),$$

isto é, $fg = p \circ q$, para que o Teorema 4.4 permita concluir imediatamente que fg é diferenciável no ponto $\boldsymbol{a}$ e também que:

$$(fg)'(\boldsymbol{a}) = p'\big(q(\boldsymbol{a})\big) \circ q'(\boldsymbol{a}) = g(\boldsymbol{a})f'(\boldsymbol{a}) + f(\boldsymbol{a})g'(\boldsymbol{a}).$$

Suponhamos agora $g(\boldsymbol{a}) \neq 0$ e consideremos a função

$$\varphi(\boldsymbol{x}) = \frac{1}{g(\boldsymbol{x})},$$

definida nos pontos $\boldsymbol{x} \in D$ tais que $g(\boldsymbol{x}) \neq 0$ (observe-se que, sendo $g(\boldsymbol{a}) \neq 0$ e g contínua — por ser diferenciável — em $\boldsymbol{a}$, será também $g(\boldsymbol{x}) \neq 0$ em todo o ponto $\boldsymbol{x}$ de alguma bola centrada em $\boldsymbol{a}$, donde resulta que $\boldsymbol{a}$, por hipótese interior a D, será também interior ao domínio de φ).

Pondo, para todo o $t \in \mathbb{R}\backslash\{0\}$, $\psi(t) = 1/t$, ter-se-á evidentemente $\varphi = \psi \circ g$ e portanto, atendendo de novo ao Teorema 4.4, pode concluir-se que φ é diferenciável no ponto $\boldsymbol{a}$, obtendo-se imediatamente a relação:

$$\varphi'(\boldsymbol{a}) = -\frac{1}{\big(g(\boldsymbol{a})\big)^2} g'(\boldsymbol{a}).$$

Finalmente, os resultados já obtidos permitem concluir que, nas condições referidas na hipótese do teorema, o cociente $f/g = (1/g)f$ é diferenciável em $\boldsymbol{a}$ e também que:

$$\begin{aligned} \left(\frac{f}{g}\right)'(\boldsymbol{a}) &= f(\boldsymbol{a})\left(\frac{1}{g}\right)'(\boldsymbol{a}) + \frac{1}{g(\boldsymbol{a})}f'(\boldsymbol{a}) \\ &= \frac{g(\boldsymbol{a})f'(\boldsymbol{a}) - f(\boldsymbol{a})g'(\boldsymbol{a})}{\big(g(\boldsymbol{a})\big)^2}. \end{aligned}$$

□

Os Teoremas 4.3 e 4.5, em conjunto com o facto já verificado de serem diferenciáveis as constantes e as funções coordenadas $p_j(\boldsymbol{x}) = x_j$, permitem reconhecer imediatamente que qualquer função polinominal $P(x_1, \ldots, x_n)$ é diferenciável em todos os pontos de $\mathbb{R}^n$ e que qualquer função racional de n variáveis reais é diferenciável em qualquer ponto do seu domínio. 0 Teorema 4.4, por sua vez, com alguns dos resultados obtidos no estudo da diferenciabilidade das funções de uma variável real, permite concluir a diferenciabilidade de muitas outras funções reais de n variáveis correntes nas aplicações. Finalmente, o estudo das funções vectoriais sob o mesmo ponto de vista pode reduzir-se ao caso das funções reais por meio do Teorema 4.1.

Como primeiro exemplo, consideremos a função $f : \mathbb{R}^n \to \mathbb{R}$ definida pela fórmula:

$$f(\boldsymbol{x}) = e^{\boldsymbol{a}\cdot\boldsymbol{x}} \qquad (\boldsymbol{x} \in \mathbb{R}^n)$$

onde $\boldsymbol{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$. A função f é o resultado da composição de $\varphi(u) = e^u$ com

$$u = \psi(\boldsymbol{x}) = \boldsymbol{a} \cdot \boldsymbol{x} = a_1x_1 + \cdots + a_nx_n$$

e, como φ é diferenciável em $\mathbb{R}$ e ψ (polinómio do 1º grau em $x_1, \dots, x_n$) é diferenciável em $\mathbb{R}^n$, f é diferenciável em $\mathbb{R}^n$.

Num ponto qualquer $\boldsymbol{x} \in \mathbb{R}^n$ tem-se:

$$\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j} e^{a_1x_1+\cdots+a_nx_n} = a_j e^{\boldsymbol{a}\cdot\boldsymbol{x}}$$

e portanto a matriz jacobiana de f é a matriz linha:

$$\begin{bmatrix} a_1e^{\boldsymbol{a}\cdot\boldsymbol{x}} & \cdots & a_ne^{\boldsymbol{a}\cdot\boldsymbol{x}} \end{bmatrix}$$

ou, equivalentemente, o gradiante de f (no ponto $\boldsymbol{x}$) é o vector:

$$\nabla f(\boldsymbol{x}) = e^{\boldsymbol{a}\cdot\boldsymbol{x}}(a_1\boldsymbol{e}_1 + \cdots + a_n\boldsymbol{e}_n) = e^{\boldsymbol{a}\cdot\boldsymbol{x}}\boldsymbol{a}.$$

Consideremos agora a função definida pela fórmula:

$$z = g(x, y) = x^y,$$

no conjunto dos pontos $(x, y) \in \mathbb{R}^2$ tais que $x > 0$. Tem-se, em todo o domínio de g:

$$z = e^{y \log x}$$

e portanto (sendo g a composta de $z = e^u$ com $u = yv$ e $v = \log x$, funções diferenciáveis em todos os pontos dos respectivos domínios) g é também diferenciável em todo o seu domínio.

A matriz jacobiana de g num ponto (x, y) desse domínio é:

$$M_{(x,y)}(g) = \begin{bmatrix} yx^{y-1} & x^y \log x \end{bmatrix}.$$

Se forem agora $x = \alpha(t)$, $y = \beta(t)$ duas funções diferenciáveis em $\mathbb{R}$, a primeira das quais assuma apenas valores positivos (funções que poderemos encarar como as coordenadas de uma aplicação φ de $\mathbb{R}$ em $\mathbb{R}^2$, com contradomínio contido no domínio de g) a função composta:

$$h(t) = g\big(\alpha(t), \beta(t)\big) = \alpha(t)^{\beta(t)}$$

será diferenciável em $\mathbb{R}$ e a sua matriz jacobiana no ponto $t \in \mathbb{R}$ poderá obter-se multiplicando as matrizes:

$$M_{(\alpha(t),\beta(t))}(g) = \begin{bmatrix} \beta(t)\alpha(t)^{\beta(t)-1} & \alpha(t)^{\beta(t)} \log \alpha(t) \end{bmatrix}$$

e

$$M_t(\varphi) = \begin{bmatrix} \alpha'(t) \\ \beta'(t) \end{bmatrix}$$

O resultado (matriz 1×1, que identificamos com o seu único elemento) é uma confirmação da conhecida regra de derivação de uma «potência-exponencial»:

$$h'(t) = \beta(t)\alpha(t)^{\beta(t)-1}\alpha'(t) + \alpha(t)^{\beta(t)}\beta'(t) \log \alpha(t).$$

Como último exemplo, consideremos uma função $f : \mathbb{R}^3 \to \mathbb{R}^2$ de funções coordenadas:

$$\begin{aligned} u &= f_1(x, y, z) \\ v &= f_2(x, y, z) \end{aligned}$$

que suporemos diferenciáveis em qualquer ponto de $\mathbb{R}^3$ e a função $\mu : \mathbb{R}^3 \to \mathbb{R}^3$ definida pelo sistema:

$$\begin{aligned} x &= r \cos\theta \cos\varphi \\ y &= r \cos\theta \operatorname{sen}\varphi \\ z &= r \operatorname{sen}\theta. \end{aligned}$$

Como cada uma das funções coordenadas de μ é diferenciável (por ser um produto de funções diferenciáveis) a função composta $f \circ \mu$ é diferenciável em qualquer ponto $(r, \theta, \varphi) \in \mathbb{R}^3$. Para obter a sua matriz jacobiana basta efectuar o produto:

$$\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \end{bmatrix} \begin{bmatrix} \cos\theta\cos\varphi & -r\operatorname{sen}\theta\cos\varphi & -r\cos\theta\operatorname{sen}\varphi \\ \cos\theta\operatorname{sen}\varphi & -r\operatorname{sen}\theta\operatorname{sen}\varphi & r\cos\theta\cos\varphi \\ \operatorname{sen}\theta & r\cos\theta & 0 \end{bmatrix}.$$

Em particular, se for:

$$\begin{aligned} u &= x^2 + y^2 + z^2 \\ v &= x^2 + y^2 - z^2 \end{aligned}$$

obtém-se como resultado:

$$\begin{bmatrix} 2r & 0 & 0 \\ 2r\cos 2\theta & -2r^2 \operatorname{sen} 2\theta & 0 \end{bmatrix}$$

(observe-se que, neste caso, teria sido menos trabalhoso efectuar previamente a composição, o que conduziria a:

$$\begin{aligned} u &= r^2 \\ v &= r^2 \cos 2\theta, \end{aligned}$$

e obter a partir deste sistema a matriz jacobiana da função composta).

Faremos ainda uma breve referência à questão da diferenciabilidade do produto interno de duas funções vectoriais e do produto de uma função escalar por uma função vectorial.

Sendo $f, g : D \to \mathbb{R}^m$, com $D \subset \mathbb{R}^n$, duas funções diferenciáveis no ponto $\boldsymbol{a}$, ter-se-á, com as notações habituais:

$$f \cdot g = \sum_{i=1}^{m} f_i g_i$$

e portanto dos Teoremas 4.1, 4.5 e 4.3 resulta imediatamente que $f \cdot g$ é diferenciável em $\boldsymbol{a}$ e também que:

$$(f \cdot g)'(\boldsymbol{a}) = \sum_{i=1}^{m} (f_i g_i)'(\boldsymbol{a}) = \sum_{i=1}^{m} \big(g_i(\boldsymbol{a}) f_i'(\boldsymbol{a}) + f_i(\boldsymbol{a}) g_i'(\boldsymbol{a})\big).$$

Assim, se for $\boldsymbol{v} = (v_1, \dots, v_n)$ um vector arbitrário de $\mathbb{R}^n$, ter-se-á:

$$\begin{aligned}(f \cdot g)'(\boldsymbol{a})(\boldsymbol{v}) &= \sum_{i=1}^{m} \left[g_i(\boldsymbol{a})\big(f_i'(\boldsymbol{a})(\boldsymbol{v})\big) + f_i(\boldsymbol{a})\big(g_i'(\boldsymbol{a})(\boldsymbol{v})\big)\right] \\ &= g(\boldsymbol{a}) \cdot \big(f'(\boldsymbol{a})(\boldsymbol{v})\big) + f(\boldsymbol{a}).\big(g'(\boldsymbol{a})(\boldsymbol{v})\big).\end{aligned}$$

Para a derivada parcial em ordem a x_j obtém-se (por exemplo, substituindo $\boldsymbol{v}$ por $\boldsymbol{e}_j$ na igualdade precedente):

$$\frac{\partial(f \cdot g)}{\partial x_j}(\boldsymbol{a}) = g(\boldsymbol{a}) \cdot \frac{\partial f}{\partial x_j}(\boldsymbol{a}) + f(\boldsymbol{a}) \cdot \frac{\partial g}{\partial x_j}(\boldsymbol{a}).$$

Se for agora $\alpha : D \to \mathbb{R}$ uma função diferenciável no ponto $\boldsymbol{a}$ pode reconhecer-se também sem dificuldade (por exemplo, analisando separadamente cada função coordenada do produto αf) que a função vectorial αf é diferenciável em $\boldsymbol{a}$ e que, sendo $\boldsymbol{v} \in \mathbb{R}^n$, é válida a igualdade:

$$(\alpha f)'(\boldsymbol{a})(\boldsymbol{v}) = \big(\alpha'(\boldsymbol{a})(\boldsymbol{v})\big) f(\boldsymbol{a}) + \alpha(\boldsymbol{a})\big(f'(\boldsymbol{a})(\boldsymbol{v})\big);$$

em particular, para $\boldsymbol{v} = \boldsymbol{e}_j$, obtém-se a expressão da derivada parcial em ordem a x_j:

$$\frac{\partial(\alpha f)}{\partial x_j}(\boldsymbol{a}) = \frac{\partial \alpha}{\partial x_j}(\boldsymbol{a}) f(\boldsymbol{a}) + \alpha(\boldsymbol{a}) \frac{\partial f}{\partial x_j}(\boldsymbol{a}).$$

Sendo m e n inteiros positivos e $D \subset \mathbb{R}^n$ um conjunto aberto, convencionemos agora designar[5] por $\mathcal{C}(D, \mathbb{R}^m)$ o conjunto das funções definidas em D, com valores em $\mathbb{R}^m$, contínuas em cada ponto $\boldsymbol{x} \in D$. No caso particular $m = 1$, em lugar de $\mathcal{C}(D, \mathbb{R})$ escrevemos apenas $\mathcal{C}(D)$.

[5] O símbolo $\mathcal{C}(D, \mathbb{R}^m)$ poderá eventualmente ser usado para designar o conjunto das funções definidas e contínuas em D com valores em $\mathbb{R}^m$, mesmo que o subconjunto D de $\mathbb{R}^n$ não seja aberto. Por vezes escreve-se também $\mathcal{C}^0(D, \mathbb{R}^m)$, em lugar de $\mathcal{C}(D, \mathbb{R}^m)$.

Como é óbvio, $\mathcal{C}(D, \mathbb{R}^m)$ é um espaço vectorial real, em relação às operações usuais de adição de funções e de multiplicação de um número real por uma função.

Convencionemos ainda designar por $\mathcal{C}^1(D, \mathbb{R}^m)$ — ou apenas $\mathcal{C}^1(D)$, se $m = 1$ — o subespaço vectorial de $\mathcal{C}(D, \mathbb{R}^m)$ formado pelas funções f que verificam as duas condições seguintes:

1. em cada ponto $\boldsymbol{x} \in D$ e para cada $j \in \{1, \ldots, n\}$ existe a derivada parcial $\mathcal{D}_j f(\boldsymbol{x})$;

2. cada uma das funções $\mathcal{D}_j f : D \to \mathbb{R}^m$ pertence a $\mathcal{C}(D, \mathbb{R}^m)$.

As funções do espaço $\mathcal{C}^1(D, \mathbb{R}^m)$ — ou $\mathcal{C}^1(D)$ — são por vezes designadas por *funções de classe* $\mathcal{C}^1$, definidas em D.

É fácil reconhecer que, sendo f uma função definida em D com valores em $\mathbb{R}^m$ e $f_i = p_i \circ f$ a correspondente função coordenada de ordem i, a condição $f \in \mathcal{C}(D, \mathbb{R}^m)$ é verificada sse $f_i \in \mathcal{C}(D)$, para cada $i \in \{1, \ldots, m\}$. De modo análogo $f \in \mathcal{C}^1(D, \mathbb{R}^m)$ equivale a $f_i \in \mathcal{C}^1(D)$, para $i = 1, \ldots, n$.

Segue-se que em grande parte os resultados enunciados na sequência para funções escalares pertencentes a $\mathcal{C}(D)$ ou $\mathcal{C}^1(D)$ estender-se-iam imediatamente ao caso de funções vectoriais, de $\mathcal{C}(D, \mathbb{R}^m)$ ou $\mathcal{C}^1(D, \mathbb{R}^m)$, respectivamente.

O primeiro destes resultados é o objecto do teorema seguinte, no qual se regista uma condição suficiente de diferenciabilidade de grande utilidade na prática:

Teorema 4.6. *Seja $D \subset \mathbb{R}^n$ um conjunto aberto. Qualquer função $f \in \mathcal{C}^1(D)$ é diferenciável em cada ponto $\boldsymbol{a} \in D$.*

Demonstração. Para maior clareza, faremos a demonstração na hipótese $n = 2$ e indicaremos depois, de modo abreviado, a sua extensão ao caso geral.

Sendo $\boldsymbol{a} = (a_1, a_2)$ um ponto de D e $\boldsymbol{h} = (h_1, h_2)$ um vector de $\mathbb{R}^2$ tal que $\boldsymbol{a} + \boldsymbol{h} \in D$, ponhamos:

$$\theta(h_1, h_2) = f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) - \mathcal{D}_1 f(a_1, a_2) h_1 - \mathcal{D}_2 f(a_1, a_2) h_2.$$

O teorema ficará provado (no caso $n = 2$) se mostrarmos que

$$\lim_{(h_1, h_2) \to (0,0)} \frac{\theta(h_1, h_2)}{\sqrt{h_1^2 + h_2^2}} = 0.$$

Como o ponto $\boldsymbol{a}$ é interior a D, existirá uma bola $B_r(\boldsymbol{a}) \subset D$; nestas condições, se o vector $\boldsymbol{h}$ verificar a condição suplementar $\|\boldsymbol{h}\| = \sqrt{h_1^2 + h_2^2} < r$, todos os pontos da forma $(a_1 + th_1, a_2)$ ou $(a_1 + h_1, a_2 + th_2)$, com $t \in [0,\, 1]$, pertencerão a $B_r(\boldsymbol{a})$.

Considere-se então a igualdade:

$$\begin{aligned} &f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = \\ &\qquad = \big[f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2)\big] + \big[f(a_1 + h_1, a_2) - f(a_1, a_2)\big], \end{aligned}$$

que, se pusermos $\varphi_1(t) = f(a_1 + th_1, a_2)$ e $\varphi_2(t) = f(a_1 + h_1, a_2 + th_2)$, poderá escrever-se:

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = \big[\varphi_2(1) - \varphi_2(0)\big] + \big[\varphi_1(1) - \varphi_1(0)\big].$$

Aplicando o teorema de Lagrange às funções φ_1 e φ_2 em relação ao intervalo [0, 1] — o que é evidentemente legítimo nas condições da hipótese — obtém-se

$$f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = \varphi_2'(c_2) + \varphi_1'(c_1),$$

onde c_1 e c_2 são pontos convenientes do intervalo [0, 1]. Tem-se, porém, para $t \in [0,\ 1]$,

$$\varphi_1'(t) = \mathcal{D}_1 f(a_1 + th_1, a_2)h_1, \quad \varphi_2'(t) = \mathcal{D}_2 f(a_1 + h_1, a_2 + th_2)h_2$$

e portanto

$$\varphi_1'(c_1) = h_1 \mathcal{D}_1 f(a_1 + c_1 h_1, a_2), \quad \varphi_2'(c_2) = h_2 \mathcal{D}_2 f(a_1 + h_1, a_2 + c_2 h_2)$$

donde resulta

$$\begin{aligned}\theta(h_1, h_2) = h_1\big[\mathcal{D}_1 f(a_1 + c_1 h_1, a_2) - \mathcal{D}_1 f(a_1, a_2)\big] \\ + h_2\big[\mathcal{D}_2 f(a_1 + h_1, a_2 + h_2 c_2) - \mathcal{D}_2 f(a_1, a_2)\big].\end{aligned}$$

Segue-se que

$$\begin{aligned}\frac{\theta(h_1, h_2)}{\sqrt{h_1^2 + h_2^2}} = \frac{h_1}{\sqrt{h_1^2 + h_2^2}}\,[\mathcal{D}_1 f(a_1 + c_1 h_1, a_2) - \mathcal{D}_1 f(a_1, a_2)] \\ + \frac{h_2}{\sqrt{h_1^2 + h_2^2}}\,[\mathcal{D}_2 f(a_1 + h_1, a_2 + c_2 h_2) - \mathcal{D}_2 f(a_1, a_2)]\end{aligned}$$

e portanto

$$\begin{aligned}\left|\frac{\theta(h_1, h_2)}{\sqrt{h_1^2 + h_2^2}}\right| \le |\mathcal{D}_1 f(a_1 + c_1 h_1, a_2) - \mathcal{D}_1 f(a_1, a_2)| \\ + |\mathcal{D}_2 f(a_1 + h_1, a_2 + c_2 h_2) - \mathcal{D}_2 f(a_1, a_2)|\,.\end{aligned}$$

Quando $(h_1, h_2) \to (0, 0)$ os pontos $(a_1 + c_1 h_1, a_2)$ e $(a_1 + h_1, a_2 + c_2 h_2)$ tendem ambos para (a_1, a_2) e a continuidade das derivadas $\mathcal{D}_1 f$ e $\mathcal{D}_2 f$ no ponto $\boldsymbol{a}$ — resultante da hipótese de ser $f \in \mathcal{C}^1(D)$ — permite concluir que

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\theta(h_1, h_2)}{\sqrt{h_1^2 + h_2^2}} = 0,$$

o que termina a demonstração (no caso $n = 2$).

No caso geral, sendo $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in D$ e $\boldsymbol{h} = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$ com $\boldsymbol{a} + \boldsymbol{h} \in D$, pôr-se-ia:

$$\theta(\boldsymbol{h}) = f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) - \sum_{j=1}^{n} h_j \mathcal{D}_j f(\boldsymbol{a}).$$

Supondo ainda $\|\boldsymbol{h}\| < r$, com $B_r(\boldsymbol{a}) \subset D$, designando como habitualmente por $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ os vectores da base canónica de $\mathbb{R}^n$ e pondo, por comodidade de notação,

$$\boldsymbol{z_0} = a, \quad \boldsymbol{z}_1 = \boldsymbol{a} + h_1\boldsymbol{e}_1, \quad \ldots, \boldsymbol{z}_j = \boldsymbol{a} + h_1\boldsymbol{e}_1 + \cdots + h_j\boldsymbol{e}_j, \quad \boldsymbol{z}_n = \boldsymbol{a} + \boldsymbol{h},$$

ter-se-ia:

$$f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) = f(\boldsymbol{z}_n) - f(\boldsymbol{z}_0) = \sum_{j=1}^{n} [f(\boldsymbol{z}_j) - f(\boldsymbol{z}_{j-1})]$$

ou, pondo ainda, $\varphi_j(t) = f(\boldsymbol{z}_{j-1} + th_j\boldsymbol{e}_j)$

$$f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a}) = \sum_{j=1}^{n} \varphi_j'(c_j) = \sum_{j=1}^{n} h_j \mathcal{D}_j f(\boldsymbol{z}_{j-1} + c_j h_j \boldsymbol{e}_j),$$

com $c_j \in [0,\, 1]$, para $j = 1, \ldots, n$. Segue-se a igualdade

$$\frac{\theta(\boldsymbol{h})}{\|\boldsymbol{h}\|} = \sum_{j=1}^{n} \frac{h_j}{\|\boldsymbol{h}\|} [\mathcal{D}_j f(\boldsymbol{z}_{j-1} + c_j h_j \boldsymbol{e}_j) - \mathcal{D}_j f(\boldsymbol{a})]$$

da qual, tendo em conta a continuidade das derivadas $\mathcal{D}_j f$ em $\boldsymbol{a}$, se conclui que

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{\theta(\boldsymbol{h})}{\|\boldsymbol{h}\|} = 0.$$

Pode portanto considerar-se a demonstração terminada. $\square$

Exemplo: As funções coordenadas da função $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ tal que

$$\varphi(r, \theta) = (x, y) \quad \text{com} \quad \begin{cases} x = r\cos\theta \\ y = r\,\mathrm{sen}\,\theta \end{cases},$$

pertencem ambas a $\mathcal{C}^1(\mathbb{R}^2)$, o que permite concluir que φ é diferenciável em cada ponto $(r, \theta) \in \mathbb{R}^2$. No ponto (r_0, θ_0) a derivada de φ é a aplicação linear $\varphi'(r_0, \theta_0)$ determinada (na base canónica de $\mathbb{R}^2$) pela matriz

$$\begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{bmatrix}_{(r_0, \theta_0)} = \begin{bmatrix} \cos\theta_0 & -r_0\,\mathrm{sen}\,\theta_0 \\ \mathrm{sen}\,\theta_0 & r_0\cos\theta_0 \end{bmatrix}.$$

É claro que uma função diferenciável pode não ser de classe $\mathcal{C}^1$. É o que se passa, por exemplo, com a função $\psi : \mathbb{R} \to \mathbb{R}$ tal que $\psi(x) = x^2 \operatorname{sen} 1/x$ para $x \neq 0$ (com $\psi(0) = 0$).

O teorema de Lagrange para funções reais de uma variável real tem várias generalizações ao caso de n variáveis. Eis uma das de maior utilidade:

Teorema 4.7. (Lagrange, do valor médio ou dos acréscimos finitos) *Seja D um aberto de $\mathbb{R}^n$, $\boldsymbol{a}$ e $\boldsymbol{b}$ pontos de D tais que o segmento*[6] *$[\boldsymbol{a}, \boldsymbol{b}]$ esteja contido em D; seja ainda $f \in \mathcal{C}^1(D)$; então existe um ponto $\boldsymbol{c} \in \,]\boldsymbol{a}, \boldsymbol{b}[$ tal que*

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

Demonstração. Nas condições da hipótese pode aplicar-se o teorema de Lagrange no intervalo $[0, 1]$ à função (real de variável real) φ definida no mesmo intervalo pela fórmula $\varphi(t) = f\big(\boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a})\big)$, obtendo-se a garantia de existência de um ponto $\theta \in \,]0, 1[$ tal que $\varphi(1) - \varphi(0) = \varphi'(\theta)$. Pondo então $\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}) = \boldsymbol{c}$, basta observar que $\varphi(1) = f(\boldsymbol{b})$, $\varphi(0) = f(\boldsymbol{a})$ e

$$\begin{aligned}
\varphi'(\theta) &= \lim_{\lambda \to 0} \frac{\varphi(\theta + \lambda) - \varphi(\theta)}{\lambda} \\
&= \lim_{\lambda \to 0} \frac{f\big(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a}) + \lambda(\boldsymbol{b} - \boldsymbol{a})\big) - f\big(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a})\big)}{\lambda} \\
&= \mathcal{D}_{\boldsymbol{b}-\boldsymbol{a}} f\big(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a})\big) \\
&= f'\big(\boldsymbol{a} + \theta(\boldsymbol{b} - \boldsymbol{a})\big)(\boldsymbol{b} - \boldsymbol{a}) \\
&= f'(\boldsymbol{c})(\boldsymbol{b} - \boldsymbol{a}),
\end{aligned}$$

para se poder considerar a demonstração terminada. □

1. Observando a demonstração anterior reconhece-se imediatamente que o teorema poderia ter sido enunciado sob forma mais geral: por exemplo, em vez de impôr a condição $f \in \mathcal{C}^1(D)$ bastaria exigir que a função φ fosse contínua no intervalo $[0, 1]$ e diferenciável em $]0, 1[$ para se poder obter da mesma forma a conclusão que figura no enunciado. Uma observação análoga poderia aliás ser feita em relação a alguns outros enunciados de teoremas precedentes e seguintes. Porém, tendo em conta que não nos será necessária maior generalidade nas aplicações que temos em vista, pareceu-nos preferível adoptar em todos os casos enunciados tão simples quanto possível.

2. Convém observar que, se pusermos $\boldsymbol{b} - \boldsymbol{a} = \boldsymbol{h}$, a fórmula indicada no final do teorema 4.7 pode revestir a forma:

$$f(\boldsymbol{a} + \boldsymbol{h}) = f(\boldsymbol{a}) + \mathcal{D}_h f(\boldsymbol{a} + \theta \boldsymbol{h}),$$

[6] Como sabemos, designa-se por $[\boldsymbol{a}, \boldsymbol{b}]$ o conjunto dos pontos da forma $\boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a})$, com $t \in [0, 1]$ e por $]\boldsymbol{a}, \boldsymbol{b}[$ o conjunto dos pontos da mesma forma, agora com $t \in \,]0, 1[$.

ou ainda, com $\|\boldsymbol{h}\| = r$ e $\boldsymbol{h}_* = \frac{1}{r}\boldsymbol{h}$ (onde se supõe agora $\boldsymbol{h} \neq \boldsymbol{0}$, isto é, $\boldsymbol{b} \neq \boldsymbol{a}$):

$$f(\boldsymbol{a}+\boldsymbol{h}) = f(\boldsymbol{a}) + r\mathcal{D}_{\boldsymbol{h}_*} f(\boldsymbol{a}+\theta\boldsymbol{h}).$$

Convém notar também que, na forma indicada, o teorema não subsiste para o caso de funções vectoriais. Por exemplo, sendo $\varphi : \mathbb{R} \to \mathbb{R}^2$ a função definida por $\varphi(t) = (x, y)$, com $x = \cos t$ e $y = \operatorname{sen} t$, para quaisquer reais c e t, $\varphi'(c)(t)$ é a aplicação linear de $\mathbb{R}$ em $\mathbb{R}^2$ determinada pela matriz

$$\begin{bmatrix} -t \operatorname{sen} c \\ t \cos c \end{bmatrix}$$

não podendo portanto existir um ponto $c \in [0,\, 2\pi]$ para o qual se verifique a igualdade:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \varphi(2\pi) - \varphi(0) = \varphi'(c)(2\pi) = \begin{bmatrix} -2\pi \operatorname{sen} c \\ 2\pi \cos c \end{bmatrix}.$$

É no entanto válido o seguinte:

Corolário. *Seja D um aberto de $\mathbb{R}^n$, $\boldsymbol{a}$ e $\boldsymbol{b}$ pontos de D tais que o segmento $[\boldsymbol{a},\, \boldsymbol{b}]$ esteja contido em D. Seja ainda $f \in \mathcal{C}^1(D, \mathbb{R}^m)$. Nestas condições existem pontos $\boldsymbol{c}_1, \boldsymbol{c}_2, \ldots, \boldsymbol{c}_m$ no segmento $[\boldsymbol{a},\, \boldsymbol{b}]$ tais que, designando por $f_1, f_2, \ldots, f_m$ as funções coordenadas de f, se tem:*

$$\begin{bmatrix} f_1(\boldsymbol{b}) - f_1(\boldsymbol{a}) \\ \vdots \\ f_m(\boldsymbol{b}) - f_m(\boldsymbol{a}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{c}_1) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{c}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{c}_m) & \cdots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{c}_m) \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 - \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{b}_n - \boldsymbol{a}_n \end{bmatrix}.$$

Obtém-se imediatamente este corolário, aplicando o teorema anterior, separadamente, a cada uma das funções coordenadas de f; nem sempre é possível, porém, atribuir um valor comum aos pontos $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$, como o exemplo anterior torna evidente.

Para funções reais de variável real, a condição $f'(x) = 0$ em cada ponto x do domínio de f garante que a função f é constante se o domínio em causa é um intervalo de $\mathbb{R}$. Eis a generalização natural deste resultado:

Teorema 4.8. *Seja D um aberto conexo de $\mathbb{R}^n$, $f : D \to \mathbb{R}^m$ uma função diferenciável em cada ponto de D e tal que (designando por $\tilde{O}$ a aplicação nula de $\mathbb{R}^n$ em $\mathbb{R}^m$) se tenha $f'(\boldsymbol{x}) = \tilde{O}$ para cada $\boldsymbol{x} \in D$.*

Então f é constante em D, isto é, existe um vector $\boldsymbol{c} \in \mathbb{R}^m$ tal que $f(\boldsymbol{x}) = \boldsymbol{c}$ para cada $\boldsymbol{x} \in D$.

Demonstração. Sendo $\boldsymbol{x}_0$ um ponto arbitrário de D ponhamos:

$$A = \{\boldsymbol{x} \in D : f(\boldsymbol{x}) = f(\boldsymbol{x}_0)\},\ B = \{\boldsymbol{x} \in D : f(\boldsymbol{x}) \neq f(\boldsymbol{x}_0)\}.$$

Tem-se evidentemente $D = A \cup B$ e A não é vazio ($\boldsymbol{x}_0 \in A$); vamos ver que, se não fosse $B = \emptyset$ (isto é, se f não fosse constante), os conjuntos A e B seriam separados, e portanto D seria desconexo, contrariamente à hipótese.

Se $\boldsymbol{x}_1$ é um ponto arbitrário de B (suposto $B \neq \emptyset$) é $f(\boldsymbol{x}_1) \neq f(\boldsymbol{x}_0)$ e então a continuidade de f em $\boldsymbol{x}_1$ (resultante da hipótese de diferenciabilidade de f), assegura a existência de uma bola $B_\epsilon(\boldsymbol{x}_1) \subset D$ tal que, para qualquer $\boldsymbol{x} \in B_\epsilon(\boldsymbol{x}_1)$,

$$\|f(\boldsymbol{x}) - f(\boldsymbol{x}_1)\| < \|f(\boldsymbol{x}_0) - f(\boldsymbol{x}_1)\|.$$

Segue-se que nenhum ponto de A pertence a $B_\epsilon(\boldsymbol{x}_1)$ (visto que, para $\boldsymbol{x} \in A$, é $f(\boldsymbol{x}) = f(\boldsymbol{x}_0)$), isto é, que $\boldsymbol{x}_1 \notin \bar{A}$. Conclui-se assim que

$$B \cap \bar{A} = \emptyset.$$

Seja agora $\boldsymbol{x}' = (\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n)$ um ponto arbitrário de A, $B_\delta(\boldsymbol{x}')$ uma bola centrada em $\boldsymbol{x}'$ e contida em D; seja ainda $\boldsymbol{x} = (x_1, \ldots, x_n)$ um ponto de $B_\delta(\boldsymbol{x}')$ distinto de $\boldsymbol{x}'$. Designando por $f_1, \ldots, f_m$ as funções coordenadas de f, o corolário anterior assegura a existência de pontos $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m$, pertencentes ao segmento de extremos $\boldsymbol{x}$, $\boldsymbol{x}'$ e portanto a $B_\delta(\boldsymbol{x}') \subset D$ tais que

$$\begin{bmatrix} f_1(\boldsymbol{x}) - f_1(\boldsymbol{x}') \\ \vdots \\ f_m(\boldsymbol{x}) - f_m(\boldsymbol{x}') \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{c}_1) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{c}_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{c}_m) & \cdots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{c}_m) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 - \boldsymbol{x}'_1 \\ \vdots \\ \boldsymbol{x}_n - \boldsymbol{x}'_n \end{bmatrix}.$$

Como se tem $\frac{\partial f_i}{\partial x_j}(\boldsymbol{c}_j) = 0$ para quaisquer valores de i e j, pode concluir-se que $f(\boldsymbol{x}) = f(\boldsymbol{x}')$. Assim, na bola $B_\delta(\boldsymbol{x}')$ não há qualquer ponto do conjunto B, tendo-se portanto

$$A \cap \bar{B} = \emptyset,$$

o que termina a demonstração. □

Não poderia dispensar-se a exigência de o aberto D ser conexo no enunciado do teorema 4.8: por exemplo, a função $g : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ definida pela fórmula $g(x) = \frac{|x|}{x}$ tem derivada nula em todos os pontos do seu domínio sem ser evidentemente constante; e não é difícil reconhecer que poderiam ser dados exemplos análogos nos quais o domínio da função considerada, em vez de $\mathbb{R} \setminus \{0\}$, fosse qualquer conjunto aberto desconexo de $\mathbb{R}^n$ previamente escolhido.

4.3 Cálculo diferencial de ordem superior à primeira; teoremas de Schwarz e Taylor.

Tratemos em primeiro lugar da noção de derivada parcial de ordem superior à primeira; o processo a adoptar para defini-la é praticamente evidente. Considere-se, por exemplo, uma função $f(x, y, z)$, definida num subconjunto D de $\mathbb{R}^3$. Já sabemos então como podem considerar-se definidas em certos subconjuntos de D (eventualmente vazios) as funções (primeiras) derivadas de f : $\frac{\partial f}{\partial x}$, $\frac{\partial f}{\partial y}$ e $\frac{\partial f}{\partial z}$.

A primeira destas funções, por exemplo, admitirá por sua vez em certos pontos do seu domínio (eventualmente em nenhum) derivada parcial em ordem a x, ou a y ou a z. Ficarão assim definidas três novas funções, $\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)$, $\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)$ e $\frac{\partial}{\partial z}\left(\frac{\partial f}{\partial x}\right)$, que designaremos respectivamente por $\frac{\partial^2 f}{\partial x^2}$, $\frac{\partial^2 f}{\partial y \partial x}$ e $\frac{\partial^2 f}{\partial z \partial x}$. De modo análogo se definiriam as derivadas $\frac{\partial^2 f}{\partial x \partial y}$, $\frac{\partial^2 f}{\partial y^2}$ e $\frac{\partial^2 f}{\partial z \partial y}$, etc. Assim para a função $f(x, y, z) = x \operatorname{sen}(yz)$, ter-se-ia, em qualquer ponto $(x, y, z) \in \mathbb{R}^3$:

$$\frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial y^2} = -xz^2 \operatorname{sen}(yz), \quad \frac{\partial^2 f}{\partial z^2} = -xy^2 \operatorname{sen}(yz),$$
$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = z \cos(yz),$$
$$\frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x} = y \cos(yz),$$
$$\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y} = x \cos(yz) - xyz \operatorname{sen}(yz).$$

As derivadas de ordem superior à segunda definem-se de forma análoga. Eis algumas das derivadas de 3ª ordem da função do exemplo anterior (todas definidas em $\mathbb{R}^3$):

$$\frac{\partial^3 f}{\partial x^3} = \frac{\partial^3 f}{\partial y \partial x^2} = \frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^3 f}{\partial x^2 \partial y} = 0,$$
$$\frac{\partial^3 f}{\partial x \partial y^2} = \frac{\partial^3 f}{\partial y \partial x \partial y} = \frac{\partial^3 f}{\partial y^2 \partial x} = -z^2 \operatorname{sen}(yz),$$
$$\frac{\partial^3 f}{\partial y^3} = -xz^3 \cos(yz), \quad \text{etc.}$$

Convém encarar agora a questão em termos mais gerais. Seja f uma função real[7] definida num conjunto $D \subset \mathbb{R}^n$; em certos subconjuntos de D estarão então definidas as derivadas parciais $\mathcal{D}_1 f, \ldots, \mathcal{D}_n f$ (ou $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$). Sendo i e j inteiros positivos menores ou iguais a n, poderá então considerar-se definida, no conjunto formado por todos os pontos em que a função $\mathcal{D}_j f$ admite derivada parcial (finita) em relação à variável x_i, a derivada de 2ª ordem $\mathcal{D}_i(\mathcal{D}_j f) = \mathcal{D}_{i,j} f$ (que poderá também ser designada por $\frac{\partial^2 f}{\partial x_i \partial x_j}$ ou $f''_{x_i x_j}$); naturalmente, o valor da função $\mathcal{D}_{i,j} f$ em cada ponto do seu domínio será precisamente a derivada de $\mathcal{D}_j f$, em ordem a x_i, no ponto considerado.

[7] A extensão ao caso de funções vectoriais é trivial, reconhecendo-se imediatamente que a existência de determinada derivada parcial (de qualquer ordem) de uma função vectorial equivale à existência das derivadas parciais correspondentes para cada uma das suas funções coordenadas, sendo precisamente estas as coordenadas daquelas, na hipótese de existência.

De modo análogo se definiriam as derivadas de 3ª ordem, $\mathcal{D}_{i,j,k}$, de 4ª ordem, $\mathcal{D}_{i,j,k,l}$, etc. Por exemplo, para a função $f : \mathbb{R}^n \to \mathbb{R}$ definida pela fórmula:

$$f(\boldsymbol{x}) = e^{\boldsymbol{a}\cdot\boldsymbol{x}}$$

onde $\boldsymbol{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, tem-se, em qualquer ponto $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ e para qualquer sequência de n inteiros não negativos $(p_1, p_2, \ldots, p_n)$,

$$\frac{\partial^{p_1+p_2+\cdots+p_n} f}{\partial x_1^{p_1} \cdots \partial x_n^{p_n}}(\boldsymbol{x}) = a_1^{p_1} a_2^{p_2} \cdots a_n^{p_n} e^{\boldsymbol{a}.\boldsymbol{x}}.$$

Sendo p um inteiro positivo, convencionaremos dizer que a função real f, definida num aberto $D \subset \mathbb{R}^n$ é uma *função de classe $\mathcal{C}^p$ em D* (e escrever $f \in \mathcal{C}^p(D)$) sse f admitir em cada ponto $\boldsymbol{x} \in D$ derivadas parciais de todas as ordens $\leq p$, sendo cada uma destas derivadas uma função contínua em cada ponto de D; noutros termos, sse para qualquer sequência de n inteiros não negativos $(p_1, p_2, \ldots, p_n)$ verificando a condição $p_1 + p_2 + \cdots + p_n \leq p$, a função $\frac{\partial^{p_1+\cdots+p_n} f}{\partial x_1^{p_1} \cdots \partial x_n^{p_n}}$ pertencer a $\mathcal{C}(D)$. Diremos ainda que f é uma função de classe $\mathcal{C}^\infty$ em D, ou uma função indefinidamente diferenciável em D (e escreveremos $f \in \mathcal{C}^\infty(D)$) sse a condição $f \in \mathcal{C}^p(D)$ for verificada qualquer que seja $p \in \mathbb{N}$.

Define-se de modo semelhante o conceito de função vectorial (definida em D e com valores em $\mathbb{R}^m$) de classe $\mathcal{C}^p$, com p inteiro positivo ou $p = \infty$; designando por $\mathcal{C}^p(D, \mathbb{R}^m)$ o conjunto destas funções, ter-se-á $f \in \mathcal{C}^p(D, \mathbb{R}^m)$ sse cada uma das funções coordenadas de f for um elemento de $\mathcal{C}^p(D)$.

É fácil verificar que uma função de n variáveis tem n^p funções derivadas de ordem p, para qualquer inteiro $p \geq 0$.[8] No entanto em certos casos particulares importantes, identificam-se as derivadas que comportam igual número de derivações em relação a cada uma das variáveis; é o que se verifica com a função $f(x, y, z) = x \operatorname{sen}(yz)$ atrás mencionada, para a qual se tem, por exemplo,

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}, \quad \frac{\partial^3 f}{\partial y \partial x^2} = \frac{\partial^3 f}{\partial x \partial y \partial x} = \frac{\partial^3 f}{\partial x^2 \partial y}, \quad \text{etc.}$$

Um dos resultados mais importantes neste sentido é o que se exprime no seguinte teorema (propositadamente enunciado para o caso de uma função de duas variáveis e de derivadas de 2ª ordem mas que se estenderá depois trivialmente a situações mais gerais).

Teorema 4.9 (Schwarz). *Seja D um aberto de $\mathbb{R}^2$ e $f \in \mathcal{C}^2(D)$; então em qualquer ponto $(a, b) \in D$ verifica-se a igualdade:*

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

[8] Claro que algumas dessas funções derivadas podem ter domínio vazio, isto é, podem reduzir-se à função vazia.

Demonstração. Supondo contida em D a bola centrada em (a, b) e de raio r, $B_r(a, b)$, considere-se a função $\Delta_{(a,b)}f$ (ou mais simplesmente Δf) definida em $B_r(0, 0)$ pela fórmula:

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

(onde (h, k) é um vector arbitrário de norma $< r$).

Pondo $\varphi(t) = f(a + th, b + k) - f(a + th, b)$, ter-se-á

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

donde se obtém, por duas aplicações sucessivas do teorema de Lagrange (legítimas, visto que f é de classe $\mathcal{C}^2$ em D):

$$\begin{aligned}\Delta f(h, k) = \varphi'(c_1) &= h\left[\frac{\partial f}{\partial x}(a + c_1h, b + k) - \frac{\partial f}{\partial x}(a + c_1h, b)\right]\\ &= hk\frac{\partial^2 f}{\partial y\partial x}(a + c_1h, b + c_2k),\end{aligned}$$

com $c_1, c_2 \in [0,\ 1]$.

Pondo agora $\psi(t) = f(a + h, b + tk) - f(a, b + tk)$ ter-se-á, de modo análogo:

$$\begin{aligned}\Delta f(h, k) &= \psi(1) - \psi(0) = \psi'(d_2)\\ &= k[\frac{\partial f}{\partial y}(a + h, b + d_2k) - \frac{\partial f}{\partial y}(a, b + d_2k)]\\ &= hk\frac{\partial^2 f}{\partial x\partial y}(a + d_1h, b + d_2k)\end{aligned}$$

com $d_1, d_2 \in [0,\ 1]$.

Tem-se, portanto, para $\|(h, k)\| < r$ e $h \neq 0,\ k \neq 0$,

$$\frac{\partial^2 f}{\partial y\partial x}(a + c_1h, b + c_2k) = \frac{\Delta f(h, k)}{hk} = \frac{\partial^2 f}{\partial x\partial y}(a + d_1h, b + d_2k).$$

Quando $(h, k) \to (0, 0)$ os pontos $(a + c_1h, b + c_2k)$ e $(a + d_1h, b + d_2k)$ tendem ambos para (a, b); por passagem ao limite, atendendo à hipótese de f ser de classe $\mathcal{C}^2$ em D, obtém-se então imediatamente:

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

□

Deduz-se facilmente do teorema anterior que, no caso de f ser uma função de classe $\mathcal{C}^p$ no aberto $D \subset \mathbb{R}^n$, serão idênticas todas as derivadas que possam obter-se derivando f, por qualquer ordem, p_1 vezes em ordem a x_1, p_2 vezes em ordem a x_2, $\ldots$, p_n vezes em ordem a x_n, desde que seja $p_1 + p_2 + \cdots + p_n \leq$

p. Com efeito, a passagem de uma a outra de tais derivadas poderá sempre ser efectuada por trocas sucessivas da ordem de duas operações de derivação consecutivas, efectuadas sobre funções de classe $\mathcal{C}^2$ e relativas apenas a duas das variáveis consideradas, intervindo nessas operações como se fossem constantes todas as variáveis restantes; e é claro que essas trocas de ordem das derivadas estão legitimadas pelo teorema anterior. Assim, por exemplo, prova-se que

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

(com f de classe $\mathcal{C}^4$), atendendo às sucessivas igualdades:

$$\begin{aligned}\frac{\partial^4 f}{\partial x^2 \partial y \partial z} &= \frac{\partial^2}{\partial x^2}\left(\frac{\partial^2 f}{\partial y \partial z}\right) = \frac{\partial^2}{\partial x^2}\left(\frac{\partial^2 f}{\partial z \partial y}\right)\\
&= \frac{\partial}{\partial x}\left[\frac{\partial^2}{\partial x \partial z}\left(\frac{\partial f}{\partial y}\right)\right] = \frac{\partial}{\partial x}\left[\frac{\partial^2}{\partial z \partial x}\left(\frac{\partial f}{\partial y}\right)\right]\\
&= \frac{\partial^2}{\partial x \partial z}\left(\frac{\partial^2 f}{\partial x \partial y}\right) = \frac{\partial^2}{\partial z \partial x}\left(\frac{\partial^2 f}{\partial y \partial x}\right)\\
&= \frac{\partial}{\partial z}\left[\frac{\partial^2}{\partial x \partial y}\left(\frac{\partial f}{\partial x}\right)\right] = \frac{\partial}{\partial z}\left[\frac{\partial^2}{\partial y \partial x}\left(\frac{\partial f}{\partial x}\right)\right]\\
&= \frac{\partial^4 f}{\partial z \partial y \partial x^2}.\end{aligned}$$

Para justificar estas igualdades basta invocar o teorema 4.4 (além de convenções óbvias relativas à notação das derivadas parciais).

Não seria necessário dizer que a regra de derivação das funções compostas pode aplicar-se, aliás de modo evidente, ao cálculo de derivadas de ordem superior à primeira. Para fixar as ideias num exemplo simples, consideremos a composição de uma função real $y = f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_n)$ de classe $\mathcal{C}^p$ num aberto $D \subset \mathbb{R}^n$, com n funções $x_1 = g_1(t)$, $\ldots$, $x_n = g_n(t)$, de classe $\mathcal{C}^p$, num intervalo aberto $I \subset \mathbb{R}$, tal que $g_1(I) \times \cdots \times g_n(I) \subset D$.

Ter-se-á então, pondo $\varphi(t) = f\big(g_1(t), \ldots, g_n(t)\big)$ (e supondo $p \geq 2$),

$$\frac{d\varphi}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}$$

e portanto,

$$\frac{d^2\varphi}{dt^2} = \sum_{i=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial f}{\partial x_i}\right)\frac{dx_i}{dt} + \frac{\partial f}{\partial x_i}\frac{d^2 x_i}{dt^2}\right].$$

Em geral $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}$ serão ainda funções compostas de t por intermédio de $x_1, \ldots, x_n$, de modo que as suas derivadas (em ordem a t) poderão exprimir-se pelas fórmulas:

$$\frac{d}{dt}\left(\frac{\partial f}{\partial x_i}\right) = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\frac{dx_j}{dt}.$$

Pode assim concluir-se que

$$\frac{d^2\varphi}{dt^2} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{d^2 x_i}{dt^2} + \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\frac{dx_i}{dt}\frac{dx_j}{dt}.$$

Para derivadas de ordem $p > 2$ (e supondo sempre f e $g_1, \ldots, g_n$ funções de classe $\mathcal{C}^p$) tudo seria análogo. Por exemplo, com $p = 3$, ter-se-ia

$$\frac{d^3\varphi}{dt^3} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{d^3 x_i}{dt^3} + 3\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\frac{d^2 x_i}{dt^2}\frac{dx_j}{dt} + \sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}\frac{dx_i}{dt}\frac{dx_j}{dt}\frac{dx_k}{dt}.$$

No caso particular em que as funções $g_1, \ldots, g_n$ são lineares afins, isto é, da forma $g_i(t) = a_i + th_i$ com $a_i, h_i \in \mathbb{R}$ — e com uma ligeira alteração das notações adoptadas — os resultados precedentes conduziriam imediatamente às fórmulas:

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

e em geral, supondo a função f de classe $\mathcal{C}^p$:

$$\varphi^{(p)}(t) = \sum_{i_1, i_2, \ldots, i_p=1}^{n} \frac{\partial^p f}{\partial x_{i_1} \cdots \partial x_{i_p}}(a+th)h_{i_1}h_{i_2}\ldots h_{i_p}.$$

Teremos oportunidade de reencontrar estas fórmulas brevemente, a propósito da demonstração do teorema de Taylor.

Como vimos, no caso de f ser uma função real definida num aberto $D \subset \mathbb{R}^n$ e diferenciável no ponto $\boldsymbol{a} \in D$, a derivada $f'(\boldsymbol{a})$ é uma aplicação linear de $\mathbb{R}^n$ em $\mathbb{R}$ (isto é, um elemento do espaço $L(\mathbb{R}^n, \mathbb{R})$); vimos também que o valor dessa aplicação num vector $\boldsymbol{h}$ (tal que $\boldsymbol{a} + \boldsymbol{h} \in D$), $f'(\boldsymbol{a})(\boldsymbol{h})$, faculta uma aproximação da diferença $f(\boldsymbol{a}+\boldsymbol{h}) - f(\boldsymbol{a})$ que é, em certo sentido, melhor do que a que poderia conseguir-se com qualquer outra aplicação linear de $\mathbb{R}^n$ em $\mathbb{R}$ (dado que, de acordo com o corolário do teorema 4.2, § 4.2, $f'(\boldsymbol{a})$ é o único elemento de $L(\mathbb{R}^n, \mathbb{R})$ cujo valor em $\boldsymbol{h}$ difere de $f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a})$ por um infinitésimo de ordem superior à primeira, quando $\boldsymbol{h} \to \boldsymbol{0}$).

Embora as aproximações desta forma sejam amplamente suficientes para muitos dos objectivos mais correntes (e tenham a grande vantagem de serem particularmente simples) há por vezes necessidade de recorrer a funções de $\boldsymbol{h}$ mais complicadas do que as lineares (funções quadráticas, cúbicas, etc.) para aproximar convenientemente o acréscimo $f(\boldsymbol{a} + \boldsymbol{h}) - f(\boldsymbol{a})$; e é natural pensar que, para esse efeito, convirá começar-se por definir de forma conveniente as derivadas de ordem superior à primeira da função f no ponto $\boldsymbol{a}$, $f''(\boldsymbol{a})$, $f'''(\boldsymbol{a})$, etc.

Começando por $f''(\boldsymbol{a})$, a ideia que ocorre naturalmente para defini-la é a de considerar a derivada, no ponto $\boldsymbol{a}$, da função f'. Mas aqui podem seguir alguns obstáculos, talvez inesperados.

Supondo, para simplificar, que f é diferenciável em todos os pontos de D, f' será uma função com o domínio D (tal como f), mas cujo contradomínio não está já contido em $\mathbb{R}$ (como o da própria função f) mas sim no espaço $L(\mathbb{R}^n, \mathbb{R})$ (visto que, para cada $\boldsymbol{x} \in D$, $f'(\boldsymbol{x})$ é um elemento deste último espaço). Ora para que se pudesse definir o conceito de diferenciabilidade para uma função com valores em $L(\mathbb{R}^n, \mathbb{R})$ seria necessário que este espaço estivesse munido, não apenas da sua estrutura de espaço vectorial (que considerámos na página 90), mas também de algumas noções de carácter topológico (e, para este efeito, o ideal seria dispormos de uma norma sobre o espaço $L(\mathbb{R}^n, \mathbb{R})$, fixada de modo conveniente).

Na realidade, a definição de uma tal norma não se reveste de qualquer dificuldade.[9] No entanto, por esta via, tudo parece ir-se complicando mais do que seria desejável (principalmente se repararmos que, para definir $f'''(\boldsymbol{a})$, $f^{(4)}(\boldsymbol{a})$, etc., deveria ter-se em conta que o contradomínio da função f'' seria um subconjunto de $L\big(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R})\big)$ (isto é, do espaço das aplicações lineares de $\mathbb{R}^n$ em $L(\mathbb{R}^n, \mathbb{R})$), o de f''' um subconjunto de $L\big(\mathbb{R}^n, L(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}))\big)$, etc.

É certo que estas dificuldades são mais aparentes do que reais, podendo ser ultrapassadas directamente com relativa simplicidade. No entanto, numa primeira abordagem do tema, será talvez preferível a via alternativa que seguiremos na sequência. Para torná-la mais natural convirá observar precisamente que, tal como a aplicação linear $f'(\boldsymbol{a})$ assume um valor real quanto aplicada a um vector $\boldsymbol{u} \in \mathbb{R}^n$ (de acordo com a fórmula $f'(\boldsymbol{a})(\boldsymbol{u}) = \mathcal{D}_u f(\boldsymbol{a})$), $f''(\boldsymbol{a})$ deverá assumir um valor real se for sucessivamente aplicada a dois vectores $\boldsymbol{u}, \boldsymbol{v}$ de $\mathbb{R}^n$ (visto que, sendo $f''(\boldsymbol{a}) \in L\big(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R})\big)$, ter-se-á $f''(\boldsymbol{a})(\boldsymbol{u}) \in L(\mathbb{R}^n, \mathbb{R})$ e portanto $\big(f''(\boldsymbol{a})(\boldsymbol{u})\big)(\boldsymbol{v}) \in \mathbb{R}$); e é bem razoável supor que o valor final obtido, que poderemos designar por $f''(\boldsymbol{a})(\boldsymbol{u}, \boldsymbol{v})$, será precisamente $\mathcal{D}_{\boldsymbol{v}}(\mathcal{D}_{\boldsymbol{u}} f)(\boldsymbol{a})$.

Consideremos então uma função f de classe $\mathcal{C}^2$ no aberto $D \subset \mathbb{R}^n$ e, sendo $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ e $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ dois vectores quaisquer de $\mathbb{R}^n$, observemos que se tem, em qualquer ponto $\boldsymbol{x} \in D$:

$$\mathcal{D}_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}) u_i$$

[9] É fácil reconhecer que, como espaço vectorial, $L(\mathbb{R}^n, \mathbb{R})$ é isomorfo ao próprio espaço $\mathbb{R}^n$ (isto é, que existe uma aplicação linear bijectiva $\varphi : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R})$), o que permite «transportar» para $L(\mathbb{R}^n, \mathbb{R})$ a norma que temos vindo a considerar sobre $\mathbb{R}^n$ (ou qualquer das outras infinitas normas que podem considerar-se neste espaço); e pode também provar-se que, qualquer que fosse a aplicação linear bijectiva φ escolhida e qualquer que fosse a norma sobre $\mathbb{R}^n$ que se decidisse transportar para $L(\mathbb{R}^n, \mathbb{R})$ por meio de φ, as noções topológicas resultantes neste último espaço — e a própria noção de diferenciabilidade para funções $f : D \to L(\mathbb{R}^n, \mathbb{R})$ — seriam sempre as mesmas. Assim, qualquer norma fixada sobre $L(\mathbb{R}^n, \mathbb{R})$ serviria para o efeito visado.

ou, omitindo a referência explícita ao ponto $\boldsymbol{x}$:

$$\mathcal{D}_{\boldsymbol{u}} f = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} u_i;$$

ter-se-á também, portanto:

$$\mathcal{D}_{\boldsymbol{v}} \left(\mathcal{D}_{\boldsymbol{u}} f \right) = \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \left(\sum_{i=1}^{n} \frac{\partial f}{\partial x_i} u_i \right) v_j = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j} u_i v_j.$$

Nestas condições, a definição que adoptaremos para a segunda derivada de f no ponto $\boldsymbol{a}$ será a seguinte: supondo f de classe $\mathcal{C}^2(D)$, chamaremos *segunda derivada de f no ponto $\boldsymbol{a} \in D$*, e designaremos pelos símbolos $f''(\boldsymbol{a})$ ou $\mathcal{D}^2 f(\boldsymbol{a})$, a aplicação de $\mathbb{R}^n \times \mathbb{R}^n$ em $\mathbb{R}$ que a cada par $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}^n \times \mathbb{R}^n$ faz corresponder o número real

$$f''(\boldsymbol{a})(\boldsymbol{u}, \boldsymbol{v}) = \mathcal{D}^2 f(\boldsymbol{a})(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) u_i v_j.$$

No caso particular, importante na sequência, de ser $\boldsymbol{u} = \boldsymbol{v}$, ter-se-á, convencionando agora escrever $f''(\boldsymbol{a})\boldsymbol{u}^2$ (ou $\mathcal{D}^2 f(\boldsymbol{a})\boldsymbol{u}^2$) em lugar de $f''(\boldsymbol{a})(\boldsymbol{u}, \boldsymbol{u})$,

$$f''(\boldsymbol{a})\boldsymbol{u}^2 = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) u_i u_j,$$

fórmula a que poderemos dar a forma simbólica:

$$f''(\boldsymbol{a})\boldsymbol{u}^2 = \left[(u_1 \mathcal{D}_1 + u_2 \mathcal{D}_2 + \cdots + u_n \mathcal{D}_n)^2 f \right](\boldsymbol{a}),$$

na qual os «produtos» $\mathcal{D}_i \mathcal{D}_j$, que surgirão no desenvolvimento do «quadrado» que figura no 2º membro, deverão naturalmente ser interpretados como se sugere nas igualdades:

$$\mathcal{D}_i \mathcal{D}_j f(\boldsymbol{a}) = \left[\frac{\partial}{\partial x_i} \left(\frac{\partial}{\partial x_j} f \right) \right](\boldsymbol{a}) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}).$$

De forma análoga, mas supondo agora f de classe $\mathcal{C}^3$, a terceira derivada de f no ponto $\boldsymbol{a}$, $f'''(\boldsymbol{a})$ ou $\mathcal{D}^3 f(\boldsymbol{a})$, será, por definição, a aplicação de $\mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n$ em $\mathbb{R}$ que associa a cada terno $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})$ de vectores de $\mathbb{R}^n$ o número real

$$f'''(\boldsymbol{a})(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\boldsymbol{a}) u_i v_j w_k$$

e, se for $\boldsymbol{u} = \boldsymbol{v} = \boldsymbol{w}$, ter-se-á, escrevendo agora $f'''(\boldsymbol{a})\boldsymbol{u}^3$ em vez de $f'''(\boldsymbol{a})(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{u})$,

$$f'''(\boldsymbol{a})\boldsymbol{u}^3 = \sum_{i,j,k=1}^{n} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\boldsymbol{a}) u_i u_j u_k$$

ou, simbolicamente,

$$f'''(\boldsymbol{a})\boldsymbol{u}^3 = \big[(u_1\mathcal{D}_1 + \cdots + u_n\mathcal{D}_n)^3 f\big](\boldsymbol{a}),$$

com a interpretação óbvia para os «produtos» $\mathcal{D}_i\mathcal{D}_j\mathcal{D}_k$.

Mais geralmente, se p for um inteiro ≥ 1 e f uma função de classe $\mathcal{C}^p$ ter-se-á, com as adaptações de notação já evidentes:[10]

$$f^{(p)}(\boldsymbol{a})\boldsymbol{u}^p = \big[(u_1\mathcal{D}_1 + \cdots + u_n\mathcal{D}_n)^p f\big](\boldsymbol{a}).$$

Por exemplo, no caso de uma função de três variáveis, $f(x, y, z)$, ter-se-á (designando agora por (a, b, c) o ponto considerado e sendo (α, β, γ) o vector $\boldsymbol{u}$):

$$f^{(p)}(a, b, c)(\alpha, \beta, \gamma)^p = \sum_{\substack{i+j+k=p \\ i\geq 0, j\geq 0, k\geq 0}} \frac{p!}{i!\, j!\, k!} \frac{\partial^p f}{\partial x^i \partial y^j \partial z^k}(a, b, c)\alpha^i\beta^j\gamma^k.$$

Estamos agora em condições de provar o

Teorema 4.10 (Taylor). *Seja D um aberto de $\mathbb{R}^n$, $\boldsymbol{a}$ e $\boldsymbol{a}+\boldsymbol{h}$ pontos de D tais que o segmento $[\boldsymbol{a},\, \boldsymbol{a}+\boldsymbol{h}]$ esteja contido em D, p um inteiro positivo e f uma função real de classe $\mathcal{C}^p$ em D; então existe $\theta \in\,]0,\, 1[$ tal que:*

$$f(\boldsymbol{a}+\boldsymbol{h}) = f(\boldsymbol{a}) + f'(\boldsymbol{a})\boldsymbol{h} + \frac{1}{2!}f''(\boldsymbol{a})\boldsymbol{h}^2 + \cdots + \frac{1}{(p-1)!}f^{(p-1)}(\boldsymbol{a})\boldsymbol{h}^{p-1} + r_p(\boldsymbol{h}),$$

onde $r_p(\boldsymbol{h}) = \frac{1}{p!}f^{(p)}(\boldsymbol{a}+\theta\boldsymbol{h})\boldsymbol{h}^p$.

Demonstração. O processo adoptado na demonstração é análogo ao que usámos para provar o teorema do valor médio. Pondo $\varphi(t) = \boldsymbol{a}+t\boldsymbol{h}$, as condições impostas na hipótese são (amplamente) suficientes para que possa aplicar-se a fórmula de Taylor (com resto de Lagrange) à função φ no intervalo $]0,\, 1[$, o que conduz a

$$\varphi(1) = \varphi(0) + \varphi'(0) + \frac{1}{2!}\varphi''(0) + \cdots + \frac{1}{(p-1)!}\varphi^{(p-1)}(0) + \frac{1}{p!}\varphi^{(p)}(\theta)$$

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

[10] Para o desenvolvimento da potência $\left(\sum_{i=1}^n u_i\mathcal{D}_i\right)^p$ poderá ser útil a chamada *fórmula do polinómio de Leibniz*, generalização da fórmula do binómio que se justifica facilmente — a partir desta última — por indução (sobre n):

$$(z_1 + z_2 + \cdots + z_n)^p = \sum_{\substack{p_1+p_2+\cdots+p_n=p \\ p_1\geq 0, \ldots, p_n\geq 0}} \frac{p!}{p_1!\, p_2! \cdots p_n!} z_1^{p_1} z_2^{p_2} \cdots z_n^{p_n}.$$

Convém observar que a função que figura no segundo membro é um polinómio homogéneo (de grau $p = p_1 + p_2 + \cdots + p_n$) em $z_1, z_2, \ldots, z_n$.

Basta agora observar que (com $\boldsymbol{a} = (a_1, \dots, a_n)$, $\boldsymbol{h} = (h_1, \dots, h_n)$ e $\boldsymbol{x} = (x_1, \dots, x_n) = \boldsymbol{a} + t\boldsymbol{h}$) se tem:

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

etc., e portanto

$$\begin{aligned} \varphi'(0) &= f'(\boldsymbol{a})\boldsymbol{h} \\ &\vdots \\ \varphi^{(p-1)}(0) &= f^{(p-1)}(\boldsymbol{a})\boldsymbol{h}^{p-1} \\ \varphi^{(p)}(\theta) &= f^{(p)}(\boldsymbol{a} + \theta\boldsymbol{h})\boldsymbol{h}^p, \end{aligned}$$

para terminar a demonstração. □

1. Supondo $\boldsymbol{h} \neq \boldsymbol{0}$, $r = \|\boldsymbol{h}\|$ e $\boldsymbol{h}_* = 1/r\boldsymbol{h}$, ter-se-á evidentemente:

   $$f'(\boldsymbol{a})\boldsymbol{h} = rf'(\boldsymbol{a})\boldsymbol{h}_*, \quad f''(\boldsymbol{a})\boldsymbol{h}^2 = r^2\boldsymbol{f}''(\boldsymbol{a})\boldsymbol{h}_*^2, \quad \dots$$

   A fórmula precedente pode portanto assumir o aspecto:

   $$\begin{aligned} f(\boldsymbol{a} + \boldsymbol{h}) &= f(\boldsymbol{a} + r\boldsymbol{h}_*) \\ &= f(\boldsymbol{a}) + rf'(\boldsymbol{a})\boldsymbol{h}_* + \frac{r^2}{2!}f''(\boldsymbol{a})\boldsymbol{h}_*^2 + \cdots \\ &\quad \cdots + \frac{r^{p-1}}{(p-1)!}f^{(p-1)}(\boldsymbol{a})\boldsymbol{h}_*^{p-1} + \frac{r^p}{p!}f^{(p)}(\boldsymbol{a} + \theta r\boldsymbol{h}_*)\boldsymbol{h}_*^p. \end{aligned}$$

   No caso particular $n = 1$ e supondo, por exemplo, $h > 0$, obtém-se imediatamente:

   $$\begin{aligned} f(a + r) = f(a) + rf'(a) + \frac{r^2}{2!}f''(a) + \cdots& \\ \cdots + \frac{r^{p-1}}{(p-1)!}f^{(p-1)}(a) + \frac{r^p}{p!}f^{(p)}(a + \theta r),& \end{aligned}$$

   isto é, a clássica fórmula de Taylor com resto de Lagrange conhecida do cálculo diferencial para funções reais de (uma) variável real.

2. Não é difícil reconhecer — atendendo a que, como vimos,$f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p$ é um polinómio homogéneo de grau p em $h_1, \dots, h_n$ e a que a função f se supõe de classe $\mathcal{C}^p$ num aberto contendo o ponto $\boldsymbol{a}$ — que o termo complementar da fórmula de Taylor,

   $$r_p(\boldsymbol{h}) = \frac{1}{p!}f^{(p)}(\boldsymbol{a} + \theta\boldsymbol{h})\boldsymbol{h}^p,$$

(designado ainda por *resto de Lagrange* da mesma fórmula) é um infinitésimo com $\boldsymbol{h}$ de ordem superior a $p-1$:

$$\lim_{\boldsymbol{h}\to\boldsymbol{0}} \frac{r_p(\boldsymbol{h})}{\|\boldsymbol{h}\|^{p-1}} = 0;$$

e é também fácil verificar (atendendo ainda às mesmas razões há pouco invocadas) que esse termo complementar pode assumir a forma:

$$r_p(\boldsymbol{h}) = \frac{1}{p!} f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p + o(\|\boldsymbol{h}\|^p)$$

(resto de Peano).

3. Algumas das convenções de escrita que temos vindo a adoptar permitiram-nos dar à «fórmula de Taylor» inserta no enunciado do teorema 4.10 um aspecto gráfico muito semelhante ao habitual no caso $n = 1$ (isto é, quando se consideram apenas funções de uma variável real). Porém, em muitas situações em que intervêm funções de várias variáveis, pode haver vantagem em dar a essa fórmula uma forma mais explícita, o que aliás não tem qualquer dificuldade se tivermos em conta as referidas convenções de notação. Assim, por exemplo, é fácil reconhecer que, no caso de uma função de três variáveis reais, $f(x, y, z)$, suposta de classe $\mathcal{C}^3$ numa vizinhança do ponto (x_0, y_0, z_0), a fórmula poderia assumir o aspecto:

$$\begin{aligned} f(x,y,z) =& f(x_0,y_0,z_0) \\ &+ \left(\frac{\partial f}{\partial x}\right)_0 (x-x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y-y_0) + \left(\frac{\partial f}{\partial z}\right)_0 (z-z_0) \\ &+ \frac{1}{2}\Bigg[\left(\frac{\partial^2 f}{\partial x^2}\right)_0 (x-x_0)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}\right)_0 (x-x_0)(y-y_0) \\ &\quad + 2\left(\frac{\partial^2 f}{\partial x \partial z}\right)_0 (x-x_0)(z-z_0) + \left(\frac{\partial^2 f}{\partial y^2}\right)_0 (y-y_0)^2 \\ &\quad + 2\left(\frac{\partial^2 f}{\partial y \partial z}\right)_0 (y-y_0)(z-z_0) + \left(\frac{\partial^2 f}{\partial z^2}\right)_0 (z-z_0)^2\Bigg] \\ &+ r_3(x-x_0, y-y_0, z-z_0), \end{aligned}$$

sendo o termo de resto $o\big(\|(x,y,z)-(x_0,y_0,z_0)\|^2\big)$ quando (x, y, z) tende para (x_0, y_0, z_0) (e onde se escreveu $\big(\frac{\partial f}{\partial x}\big)_0, \ldots, \big(\frac{\partial^2 f}{\partial z^2}\big)_0$ em lugar de $\frac{\partial f}{\partial x}(x_0, y_0, z_0), \ldots, \frac{\partial^2 f}{\partial z^2}(x_0, y_0, z_0)$). Feita esta observação, não haverá qualquer inconveniente em regressarmos às notações mais condensadas que temos vindo a utilizar.

No caso de f ser uma função de classe $\mathcal{C}^\infty$ — isto é, de classe $\mathcal{C}^p$ para qualquer $p \in \mathbb{N}$ (caso em que se podem escrever fórmulas de Taylor de ordem p, para todo o inteiro positivo p) — e de se verificar a igualdade $\lim_{p\to\infty} r_p(\boldsymbol{h}) = 0$, qualquer que seja o vector $\boldsymbol{h}$ de norma suficientemente pequena, dir-se-á que a função f é *analítica no ponto* $\boldsymbol{a}$; em alguma bola centrada neste ponto f poderá então ser representada pela sua série de Taylor:

$$f(\boldsymbol{a}) + f'(\boldsymbol{a})h + \frac{1}{2!}f''(\boldsymbol{a})\boldsymbol{h}^2 + \cdots + \frac{1}{p!}f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p + \cdots$$

Para fixar as ideias num exemplo muito simples, considere-se a função $f(x, y) = e^{x-y}$, obviamente de classe $\mathcal{C}^\infty$ em $\mathbb{R}^2$. Para esta função a fórmula de Mac-Laurin — isto é, a fórmula de Taylor relativa ao ponto $(0, 0)$ — poderá escrever-se (designando agora por (x, y) o acréscimo anteriormente designado por $\boldsymbol{h}$):

$$f(x, y) = f(0, 0) + f'(0, 0)(x, y) + \cdots + \frac{1}{(p-1)!}f^{(p-1)}(0, 0)(x, y)^{p-1} + r_p(x, y),$$

com $r_p(x, y) = \frac{1}{p!}f^{(p)}(\theta x, \theta y)(x, y)^p$, para algum $\theta \in]0, 1[$.

Facilmente se verifica que, para qualquer inteiro positivo p e qualquer inteiro i tal que $0 \leq i \leq p$, se tem

$$\frac{\partial^p f}{\partial x^i \partial y^{p-i}}(x, y) = (-1)^{p-i}e^{x-y}$$

e portanto:

$$\begin{aligned} \left|r_p(x, y)\right| &= \frac{1}{p!}\left|\sum_{i=0}^{p} \frac{p!}{i!(p-i)!}(-1)^{p-i}e^{\theta x - \theta y}x^i y^{p-i}\right| \\ &\leq \sum_{i=0}^{p} e^{|x|+|y|}\frac{|x|^i|y|^{p-i}}{i!\,(p-i)!} = e^{|x|+|y|}\frac{(|x|+|y|)^p}{p!}. \end{aligned}$$

Daqui imediatamente decorre que, qualquer que seja $(x, y) \in \mathbb{R}^2$, se terá:

$$\lim_{p\to\infty} r_p(x, y) = 0,$$

o que permite afirmar que, em qualquer ponto do plano, o valor da função $f(x, y)$ coincide com a soma da sua série de Mac-Laurin:

$$f(x, y) = \sum_{p=0}^{\infty} \frac{1}{p!}f^{(p)}(0, 0)(x, y)^p$$

(aceitando a convenção natural: $f^0(0, 0)(x, y)^0 = f(0, 0)$).

Como, para qualquer inteiro positivo p,

$$\begin{aligned} f^p(0, 0)(x, y)^p &= \sum_{i=0}^{p} \frac{p!}{i!\,(p-i)!}\frac{\partial^p f}{\partial x^i \partial y^{p-i}}(0, 0)x^i y^{p-i} \\ &= \sum_{i=0}^{p} \frac{p!}{i!\,(p-i)!}x^i(-y)^{p-i} \\ &= (x-y)^p, \end{aligned}$$

poderá ainda concluir-se que a igualdade $e^{x-y} = \sum_{p=0}^{\infty} \frac{1}{p!}(x-y)^p$ será verificada, qualquer que seja o par (x, y) de números reais.

Não será necessário dizer que, se o objectivo fosse apenas obter este resultado, teria sido bastante mais simples aproveitar os conhecimentos relativos à série de Mac-Laurin da função exponencial e^x e substituir nessa série x por $x - y$.

4.4 Teoremas das funções implícitas e da função inversa

Para dar uma ideia da natureza dos problemas que iremos estudar neste parágrafo (sob a designação tradicional, embora algo imprópria, de «funções implícitas») consideremos em primeiro lugar uma função de duas variáveis, para concretizar $F(x, y) = x^4 - y^2$, e um ponto (a, b) tal que $F(a, b) = 0$. É fácil reconhecer que, se for $a \neq 0$, existirá um rectângulo $I \times J$ centrado no ponto (a, b) no qual a equação $F(x, y) = 0$ poderá ser *univocamente* resolvida em ordem a y, ficando assim determinada uma função $y = f(x)$ tal que, para $(x, y) \in I \times J$ as condições $F(x, y) = 0$ e $y = f(x)$ sejam equivalentes (no nosso caso ter-se-á precisamente $f(x) = x^2$, se for $b > 0$, e $f(x) = -x^2$, se $b < 0$).

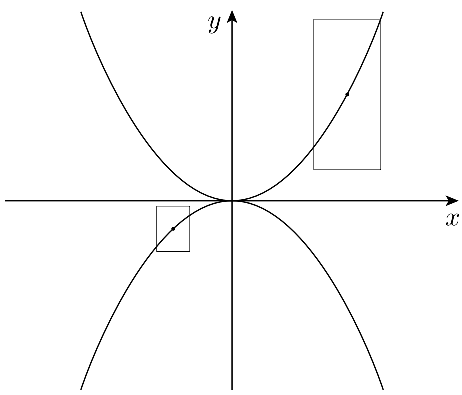

Figura 4.5

Exprimindo a mesma ideia de outro modo: sendo $a \neq 0$, existirão números positivos α e β tais que a cada $x \in \left]a - \alpha, a + \alpha\right[$ corresponda um e um só $y \in \left]b - \beta, b + \beta\right[$ por forma que se verifique a igualdade $F(x, y) = 0$. Pelo contrário, se for $a = 0$ (caso em que terá de ser também $b = 0$ para que se tenha $F(a, b) = 0$) a situação será diferente: quaisquer que sejam os números positivos α e β haverá sempre valores de x no intervalo $\left]-\alpha, \alpha\right[$ para cada um dos quais a equação $F(x, y) = 0$ não determinará univocamente um valor de y em $\left]-\beta, \beta\right[$. Veremos adiante que este facto está relacionado com o anulamento da derivada parcial $\frac{\partial F}{\partial y}$ no ponto $(0, 0)$.

Outro exemplo que poderá ser útil é o da função definida pela expressão $x^2 + y^2 - 1$ (que designaremos de novo por $F(x, y)$). Reconhece-se facilmente que a qualquer ponto (a, b) tal que $F(a, b) = 0$ e $\frac{\partial F}{\partial y}(a, b) \neq 0$ (isto é, a qualquer ponto da circunferência de raio 1 centrada na origem, com excepção de $(-1,0)$ e $(1,0)$) há possibilidade de associar um rectângulo $I \times J$, centrado em (a,b), por forma que a cada $x \in I$ corresponda um e um só $y \in J$ tal que a igualdade $F(x,y) = 0$ seja verificada; para qualquer dos pontos $(-1,0)$ e $(1,0)$ (nos quais a derivada $\frac{\partial F}{\partial y}$ se anula) é claro que essa possibilidade não existe.

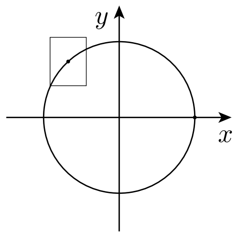

Figura 4.6

No teorema seguinte, que é uma forma ainda bastante particular do chamado teorema das funções implícitas, registam-se condições suficientes para que uma equação da forma $F(x,y) = 0$ permita definir (localmente) uma função $y = f(x)$ e, sob hipóteses convenientes a respeito de F, deduzem-se algumas propriedades da função f e indicam-se processos de cálculo das suas derivadas.

Teorema 4.11. *Seja D um aberto de $\mathbb{R}^2$, $(a,b) \in D$, $F \in \mathcal{C}^1(D)$, $F(a,b) = 0$ e $\frac{\partial F}{\partial y}(a,b) \neq 0$; então:*

1. *existem $\alpha > 0$ e $\beta > 0$ tais que a cada $x \in I =]a-\alpha,\, a+\alpha[$ corresponde um e um só $y_x \in J =]b-\beta,\, b+\beta[$ por forma que se tenha $F(x,y_x) = 0$;*

2. *pondo $f(x) = y_x$ para cada $x \in I$, a função f é de classe $\mathcal{C}^1$ e tem-se, para qualquer $x \in I$,*
$$f'(x) = -\frac{\frac{\partial F}{\partial x}\big(x, f(x)\big)}{\frac{\partial F}{\partial y}\big(x, f(x)\big)}.$$

Demonstração. Pode evidentemente supor-se $\frac{\partial F}{\partial y}(a,b) > 0$ (tudo seria análogo no caso $\frac{\partial F}{\partial y}(a,b) < 0$). Sendo F de classe $\mathcal{C}^1$, existirá $\beta > 0$ tal que $\frac{\partial F}{\partial y}(x,y) > 0$ sempre que se tenha $(x,y) \in I^* \times J^*$, onde $I^* = [a-\beta,\, a+\beta]$ e $J^* = [b-\beta,\, b+\beta]$. Segue-se que, se atribuirmos a x um valor qualquer no intervalo I^*, a função (de y) $F(x,y)$ será estritamente crescente no intervalo J^* (visto que a sua derivada é positiva em todos os pontos desse intervalo); é o que terá de passar-se, em particular, com a função $F(a,y)$, donde — atendendo a que $F(a,b) = 0$ — imediatamente decorrem as desigualdades:

$$F(a, b-\beta) < 0, \qquad F(a, b+\beta) > 0.$$

A continuidade da função F permite agora reconhecer a existência de um número $\alpha > 0$ (que pode evidentemente supor-se $\leq \beta$) tal que, para cada $x \in I =$

$]a-\alpha,\, a+\alpha[$, se tenha

$$F(x, b-\beta) < 0, \qquad F(x, b+\beta) > 0.$$

Destas desigualdades e do facto de a função (de y) $F(x,y)$ ser estritamente crescente e contínua em J^* para qualquer x fixado em I (visto que $I \subset I^*$), segue-se que para cada $x \in I$ existirá um e um só $y_x \in J =]b-\beta,\, b+\beta[$ tal que $F(x, y_x) = 0$ (o que termina a primeira parte da demonstração).

Ponhamos então $f(x) = y_x$, para cada $x \in I$; antes de provar que a função f é da classe $\mathcal{C}^1$ convém ver que é contínua em todos os pontos de I. A continuidade no ponto a é quase evidente: com efeito, o resultado que acabámos de obter (para além de ter possibilitado a definição da própria função f) evidencia que para qualquer x tal que $|x-a| < \alpha$ (isto é, para qualquer $x \in I$) se tem $|f(x)-f(a)| < \beta$ ou seja $f(x) \in J$); assim, se for dado um número positivo δ (que podemos evidentemente supor $\leq \beta$) bastará repetir o raciocínio precedente (agora com δ no lugar de β) para concluir que existe $\epsilon > 0$ (que poderá supor-se $\leq \alpha$) tal que se tenha $|f(x) - f(a)| < \delta$ sempre que seja $|x-a| < \epsilon$ (isto é, para se reconhecer a continuidade de f no ponto a).

Agora, se a' for outro ponto qualquer do intervalo I e $b' = f(a')$ ter-se-á $F(a', b') = 0$ e $\frac{\partial F}{\partial y}(a', b') > 0$ (atendendo à definição da função f e ao facto de se ter $I \times J \subset I^* \times J^*$); poder-se-ia portanto — recomeçando a demonstração da primeira parte, agora com (a', b') no lugar de (a, b) — garantir a existência de números positivos α', β' (podendo evidentemente supor-se $I' =]a'-\alpha',\, a'+\alpha'[\subset I$), tais que a cada $x \in I'$ correspondesse um e um só $y'_x = g(x) \in J' =]b'-\beta',\, b'+\beta'[$ por forma que $F\big(x, g(x)\big) = 0$. Mas então a unicidade da função f anteriormente assegurada permitiria reconhecer que g seria necessariamente a restrição de f ao intervalo I' e, da mesma forma que se provara a continuidade de f no ponto a, provar-se-ia agora a continuidade de g no ponto a', isto é, a continuidade de f neste mesmo ponto.

Trataremos agora de mostrar que, para qualquer $x \in I$, se tem:

$$f'(x) = -\frac{\frac{\partial F}{\partial x}\big(x, f(x)\big)}{\frac{\partial F}{\partial y}\big(x, f(x)\big)};$$

aliás, por um argumento análogo ao que usámos para provar a continuidade de f, também aqui bastará provar que se verifica a igualdade:

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

Para qualquer h tal que $a+h \in I$ ponhamos $f(a+h) - f(a) = k$ (é claro que, sempre que h tender para 0 ter-se-á também $k \to 0$, dada a continuidade de f). Como, por hipótese, F é uma função de classe $\mathcal{C}^1$ e portanto diferenciável, ter-se-á:

$$F(a+h, b+k) - F(a,b) = h\frac{\partial F}{\partial x}(a,b) + k\frac{\partial F}{\partial y}(a,b) + \varphi(h,k)\sqrt{h^2+k^2},$$

onde $\varphi(h,k)$ tende para zero se $\|(h,k)\| = \sqrt{h^2+k^2}$ tender para zero (e portanto também se $h \to 0$, visto que $h \to 0$ implica $\sqrt{h^2+k^2} \to 0$). Tendo em conta que $F(a,b) = 0$ e que $F(a+h, b+k) = F\big(a+h, f(a)+k\big) = F\big(a+h, f(a+h)\big) = 0$, conclui-se facilmente que, para $h \neq 0$, deverá ter-se:

$$\frac{k}{h} = -\frac{\frac{\partial F}{\partial x}(a,b)}{\frac{\partial F}{\partial y}(a,b)} - \frac{\varphi(h,k)}{h\frac{\partial F}{\partial y}(a,b)}\sqrt{h^2+k^2},$$

ou

$$\frac{k}{h} = -\frac{\frac{\partial F}{\partial x}(a,b)}{\frac{\partial F}{\partial y}(a,b)} - \frac{\varphi(h,k)}{\frac{\partial F}{\partial y}(a,b)}\frac{|h|}{h}\sqrt{1+\left(\frac{k}{h}\right)^2}.$$

Se $|h|$ for suficientemente pequeno, será certamente verificada a desigualdade

$$\frac{|\varphi(h,k)|}{\left|\frac{\partial F}{\partial y}(a,b)\right|} < \frac{1}{2}$$

e portanto também

$$\frac{|k|}{|h|} \leq \frac{\left|\frac{\partial F}{\partial x}(a,b)\right|}{\left|\frac{\partial F}{\partial y}(a,b)\right|} + \frac{1}{2}\sqrt{1+\left(\frac{k}{h}\right)^2} \leq \frac{\left|\frac{\partial F}{\partial x}(a,b)\right|}{\left|\frac{\partial F}{\partial y}(a,b)\right|} + \frac{1}{2}\left(1+\frac{|k|}{|h|}\right),$$

o que permite reconhecer que para $|h|$ pequeno (e não nulo), $\frac{|k|}{|h|}$ é limitado. Da última das igualdades precedentes deduz-se então que, quando $h \to 0$, existe o limite de $\frac{k}{h} = \frac{f(a+h)-f(a)}{h}$ e que esse limite é precisamente

$$-\frac{\frac{\partial F}{\partial x}(a,b)}{\frac{\partial F}{\partial y}(a,b)}.$$

Como já referimos, este resultado permite concluir que, para qualquer $x \in I$, se terá:

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

Por sua vez esta igualdade — atendendo a que F é uma função de classe $\mathcal{C}^1$ e a que f é contínua — mostra que a função f' é contínua, isto é, que f é da classe $\mathcal{C}^1$ no intervalo I. □

Antes de registar outras versões mais gerais do teorema das funções implícitas convém fazer algumas observações:

Em primeiro lugar pode notar-se que, depois de assegurada a diferenciabilidade da função $y = f(x)$ definida pela equação $F(x,y) = 0$ nas condições indicadas no Teorema 4.11, a expressão da sua derivada, registada no final do enunciado desse teorema, pode obter-se facilmente por derivação, a partir da igualdade $F\big(x, f(x)\big) = 0$; uma observação análoga poderá ser feita a propósito das fórmulas, relativas a derivadas de funções definidas implicitamente, insertas nos enunciados dos restantes teoremas desta secção.

Deve observar-se também que o não anulamento da derivada $\frac{\partial F}{\partial y}$ no ponto (a, b) não é condição necessária para a existência de uma função $y = f(x)$ univocamente definida, nalguma vizinhança deste ponto[11], pela equação $F(x, y) = 0$. Por exemplo, sendo $F(x, y) = x - y^3$, tem-se $\frac{\partial F}{\partial y}(0, 0) = 0$, embora a equação defina — até globalmente, em todo o conjunto $\mathbb{R}$ — a função $y = \sqrt[3]{x}$ (pode notar-se que esta função não é diferenciável no ponto 0, mas basta considerar o caso da função $F(x, y) = x^3 - y^3$ para se reconhecer que o anulamento de $\frac{\partial F}{\partial y}(0, 0)$ não é incompatível com o facto de a função definida pela equação $F(x, y) = 0$ ser diferenciável no ponto considerado).

Uma outra observação, decerto óbvia para o leitor: se, no enunciado do Teorema 4.11, a hipótese $\frac{\partial F}{\partial y}(a, b) \neq 0$ fosse substituida por $\frac{\partial F}{\partial x}(a, b) \neq 0$, o que poderia concluir-se era a possibilidade de definir univocamente, numa vizinhança conveniente do ponto (a, b), uma função $x = g(y)$, de classe $\mathcal{C}^1$, para a qual se teria (em qualquer ponto y suficientemente próximo de b):

$$g'(y) = -\frac{\frac{\partial F}{\partial y}\big(g(y), y\big)}{\frac{\partial F}{\partial x}\big(g(y), y\big)}.$$

Consideremos agora, a título de exemplo, a função $F(x, y) = x^y - y^x$, definida no 1º quadrante aberto e comecemos por procurar os pontos (a, b) (situados nesse quadrante) que são soluções da equação $F(x, y) = 0$.

É evidente que todos os pontos da forma (a, a) — com $a > 0$ — satisfazem essa condição; mas é fácil ver que há outras soluções. Para tal observemos que a igualdade $F(x, y) = 0$ é equivalente a

$$\frac{\log x}{x} = \frac{\log y}{y}$$

e que, enquanto para qualquer $x \in \,]0, 1] \cup \{e\}$ não há nenhum $y \neq x$ que verifique essa igualdade, já para cada $x \in \,]1, e[\, \cup \,]e, +\infty[$ existe um e um só y distinto de x (que poderemos designar por $h(x)$) tal que

$$\frac{\log h(x)}{h(x)} = \frac{\log x}{x}$$

ou, o que é o mesmo, $x^{h(x)} = \big(h(x)\big)^x$ (cf. Fig. 4.7).

Assim, a igualdade $x^y = y^x$ (com $x, y > 0$) é verificada sse for $y = x$ (com $x > 0$) ou $y = h(x)$ (para $x \in \,]1, +\infty[\setminus\{e\}$).

Na Figura 4.8 esboçam-se os gráficos dessas funções e embora o esboço seja pouco cuidado, chega para sugerir que a equação $F(x, y) = 0$ definirá certamente, numa vizinhança suficientemente pequena de qualquer ponto da

[11]Recorde-se que se chama *vizinhança* de um ponto $\boldsymbol{c} \in \mathbb{R}^m$ a qualquer subconjunto de $\mathbb{R}^m$ que contenha uma bola centrada em $\boldsymbol{c}$.

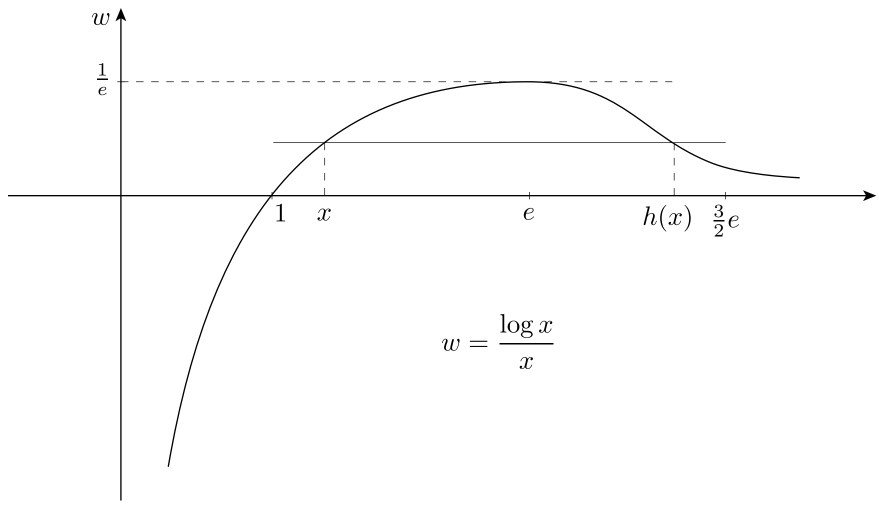

Figura 4.7

forma (a, a) — com $a > 0$ e $a \neq e$ — uma função univocamente determinada (precisamente a função $y = x$) e numa vizinhança suficientemente pequena de qualquer ponto $\big(a, h(a)\big)$ uma outra função (precisamente $y = h(x)$) também determinada de forma única.

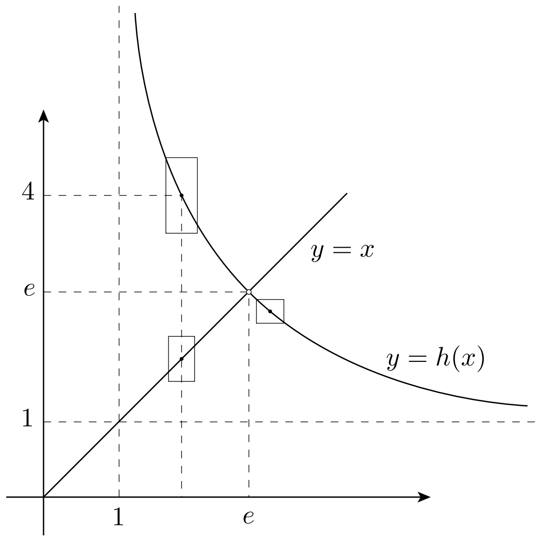

Figura 4.8

Para o ponto (e, e) é óbvio que não será possível determinar uma vizinhança na qual a equação em causa defina univocamente uma função $y = f(x)$ (ou $x = g(y)$), facto que (atendendo ao Teorema 4.11) implica o anulamento nesse ponto da derivada $\frac{\partial F}{\partial y}$ (e da derivada $\frac{\partial F}{\partial x}$). É fácil verificar que, de facto, $\frac{\partial F}{\partial y}(e, e) = 0$ (e $\frac{\partial F}{\partial x}(e, e) = 0$) e também que, em qualquer ponto $(a, b) \neq (e, e)$ e tal que $F(a, b) = 0$, as derivadas parciais da função F

se não anulam; assim, a possibilidade de definir univocamente y como função de x (ou x como função de y) numa vizinhança de tais pontos estaria de facto assegurada, nos termos do teorema das funções implícitas.

Se pretendermos determinar uma equação da tangente ao gráfico da função h num ponto do seu domínio, por exemplo no ponto 2, onde a função assume o valor 4, bastará derivar ambos os membros da igualdade $x^y - y^x = 0$ em ordem a x — supondo $y = h(x)$ — o que conduz a

$$x^y y' \log x + y x^{y-1} - x y^{x-1} y' - y^x \log y = 0,$$

donde decorre imediatamente:

$$h'(2) = \frac{4(1 - \log 2)}{1 - 2\log 2}.$$

Uma equação da tangente será então:

$$y = 4 + \frac{4(1 - \log 2)}{1 - 2\log 2}(x - 2).$$

Por um processo inteiramente análogo ao que usámos na demonstração do Teorema 4.11, obter-se-ia a seguinte versão algo mais geral do teorema das funções implícitas:

Teorema 4.12. *Seja D um aberto de $\mathbb{R}^{n+1}$, $(\boldsymbol{a}, b) = (a_1, \ldots, a_n, b) \in D$, $F \in \mathcal{C}^1(D)$, $F(\boldsymbol{a}, b) = 0$ e $\frac{\partial F}{\partial y}(\boldsymbol{a}, b) \neq 0$; então:*

1. *existem $\alpha > 0$ e $\beta > 0$ tais que a cada $\boldsymbol{x} = (x_1, \ldots, x_n) \in I =]a_1 - \alpha, a_1 + \alpha[\times \cdots \times]a_n - \alpha, a_n + \alpha[$ corresponde um e um só $y_{\boldsymbol{x}} \in J =]b - \beta, b + \beta[$ por forma que se tenha $F(\boldsymbol{x}, y_{\boldsymbol{x}}) = F(x_1, \ldots, x_n, y_{\boldsymbol{x}}) = 0$;*

2. *pondo $f(\boldsymbol{x}) = f(x_1, \ldots, x_n) = y_{\boldsymbol{x}}$ para cada $\boldsymbol{x} \in I$, a função f é de classe $\mathcal{C}^1$ e tem-se, para cada $\boldsymbol{x} \in I$ e cada $i \in \{1, \ldots, n\}$:*

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) = -\frac{\frac{\partial F}{\partial x_i}\big(x_1, \ldots, x_n, f(\boldsymbol{x})\big)}{\frac{\partial F}{\partial y}\big(x_1, \ldots, x_n, f(\boldsymbol{x})\big)}.$$

A demonstração da parte 1. é praticamente idêntica à do Teorema 4.11; para o restante, não há mais que considerar separadamente cada uma das variáveis $x_1, \ldots, x_n$, encarando as restantes variáveis como constantes.

Se, nos enunciados dos Teoremas 4.11 e 4.12, substituíssemos a hipótese $F \in \mathcal{C}^1(D)$ por $F \in \mathcal{C}^p(D)$, com p inteiro maior do que 1 ou $p = \infty$ (conservando todas as restantes hipóteses), poderíamos concluir que seria também $f \in \mathcal{C}^p(I)$ (e não apenas $f \in \mathcal{C}^1(I)$); é o que se verifica sem dificuldade se se tiverem em conta as fórmulas relativas às derivadas da função f que figuram no final dos enunciados dos referidos teoremas.

Como exemplo, determinemos o polinómio de Mac-Laurin de 2ª ordem da função $z = f(x, y)$ definida — numa vizinhança do ponto $(0, 0, 0)$ — pela equação:

$$F(x, y, z) = xz - y + \operatorname{sen} z = 0.$$

Ter-se-á:

$$\begin{aligned}
\frac{\partial F}{\partial x} &= z + x\frac{\partial z}{\partial x} + \cos z\frac{\partial z}{\partial x} = 0,\\
\frac{\partial F}{\partial y} &= x\frac{\partial z}{\partial y} - 1 + \cos z\frac{\partial z}{\partial y} = 0,\\
\frac{\partial^2 F}{\partial x^2} &= 2\frac{\partial z}{\partial x} + x\frac{\partial^2 z}{\partial x^2} - \operatorname{sen} z\left(\frac{\partial z}{\partial x}\right)^2 + \cos z\frac{\partial^2 z}{\partial x^2} = 0,\\
\frac{\partial^2 F}{\partial x\partial y} &= \frac{\partial z}{\partial y} + x\frac{\partial^2 z}{\partial x\partial y} - \operatorname{sen} z\frac{\partial z}{\partial x}\frac{\partial z}{\partial y} + \cos z\frac{\partial^2 z}{\partial x\partial y} = 0,\\
\frac{\partial^2 F}{\partial y^2} &= x\frac{\partial^2 z}{\partial y^2} - \operatorname{sen} z\left(\frac{\partial z}{\partial y}\right)^2 + \cos z\frac{\partial^2 z}{\partial y^2} = 0,
\end{aligned}$$

donde imediatamente decorre que o polinómio de Mac-Laurin em causa é $y - xy$.

Na formulação mais geral do teorema das funções implícitas que estudaremos na sequência tratar-se-á de determinar condições para que um sistema de m equações da forma:

$$\begin{cases} F_1(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0\\ \qquad\ldots\\ F_m(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0, \end{cases}$$

onde $F_1, \ldots, F_m$ são funções definidas num aberto D de $\mathbb{R}^{m+n}$, possa ser resolvido em ordem às m variáveis $y_1, \ldots, y_m$, por forma que cada uma destas fique expressa (localmente) como função das restantes variáveis, $x_1, \ldots, x_n$.

Antes de iniciar o estudo desse problema convirá fazer uma breve referência ao caso particular em que as funções $F_1, \ldots, F_m$ são lineares, assumindo o sistema a forma:

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n + b_{11}y_1 + \cdots + b_{1m}y_m = 0\\ \qquad\ldots\\ a_{m1}x_1 + \cdots + a_{mn}x_n + b_{m1}y_1 + \cdots + b_{mm}y_m = 0. \end{cases}$$

É sabido que, neste caso, as variáveis $y_1, \ldots, y_m$ podem exprimir-se, de forma única, como funções de $x_1, \ldots, x_n$ sse for diferente de zero o determinante[12]

$$\begin{vmatrix} b_{11} & \cdots & b_{1m}\\ \ldots & \ldots & \ldots\\ b_{m1} & \cdots & b_{mm} \end{vmatrix}.$$

[12]Sobre o conceito e propriedades dos determinantes e sobre a resolução de sistemas de equações lineares poderá consultar-se *Álgebra Linear como Introdução à Matemática Aplicada*, Luís T. Magalhães, Texto Editora, ou *Introdução à Álgebra Linear e Geometria Analítica*, F. Dias Agudo, Escolar Editora.

Assim, atendendo a que, na hipótese de as funções F serem lineares:

$$F_i(x_1, \dots, x_n, y_1, \dots, y_m) = a_{i1}x_1 + \cdots + a_{in}x_n + b_{i1}y_1 + \ldots + b_{im}y_m,$$

se tem $b_{ij} = \frac{\partial F_i}{\partial y_j}$ $(i, j \in \{1, \dots, m\})$ (e tendo em conta que, no caso geral, as funções F_i — se forem suficientemente «regulares» — poderão ser localmente aproximadas por funções lineares) é-se naturalmente conduzido a conjecturar que o não anulamento do determinante

$$\begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \cdots & \frac{\partial F_1}{\partial y_m} \\ \cdots & \cdots & \cdots \\ \frac{\partial F_m}{\partial y_1} & \cdots & \frac{\partial F_m}{\partial y_m} \end{vmatrix},$$

será uma hipótese significativa, quando se pretenda garantir, em termos locais, a resolubilidade do sistema em relação às variáveis $y_1, \dots, y_m$.

O determinante em causa, a que usualmente se chama *jacobiano* das funções $F_1, \dots, F_m$ em relação às variáveis $y_1, \dots, y_m$, costuma ser designado pelo símbolo $\frac{\partial(F_1, \dots, F_m)}{\partial(y_1, \dots, y_m)}$. Quanto à conjectura há pouco referida (relativa ao papel desempenhado pela hipótese de não anulamento do jacobiano na resolução do problema que temos vindo a considerar) teremos oportunidade de vê-la confirmada no enunciado do Teorema 4.14.

No entanto, antes de analisar a situação geral considerada nesse teorema,poderá ser conveniente encarar o caso particular a que se refere o

Teorema 4.13. *Seja D um aberto de $\mathbb{R}^{n+2}$, $(\boldsymbol{a}, b, c) = (a_1, a_2, \dots, a_n, b, c) \in D$, $F, G \in \mathcal{C}^1(D)$, $F(\boldsymbol{a}, b, c) = G(\boldsymbol{a}, b, c) = 0$ e $\frac{\partial(F,G)}{\partial(y,z)}(\boldsymbol{a}, b, c) \neq 0$; nestas condições:*

1. *Existe um intervalo aberto I (de $\mathbb{R}^n$, centrado no ponto $\boldsymbol{a}$) e um intervalo aberto J (de $\mathbb{R}^2$, centrado em (b, c)) tais que a cada $\boldsymbol{x} = (x_1, \dots, x_n) \in I$ corresponde um e só um par $(y_{\boldsymbol{x}}, z_{\boldsymbol{x}}) \in J$ por forma que se verifiquem as igualdades:*

$$F(\boldsymbol{x}, y_{\boldsymbol{x}}, z_{\boldsymbol{x}}) = 0, \qquad G(\boldsymbol{x}, y_{\boldsymbol{x}}, z_{\boldsymbol{x}}) = 0.$$

2. *Pondo $f(\boldsymbol{x}) = y_{\boldsymbol{x}}$ e $g(\boldsymbol{x}) = z_{\boldsymbol{x}}$, qualquer que seja $\boldsymbol{x} \in I$, as funções f e g são de classe $\mathcal{C}^1$ e tem-se, para cada $i \in \{1, \dots, n\}$ e em cada ponto $\big(\boldsymbol{x}, f(\boldsymbol{x}), g(\boldsymbol{x})\big) \in I \times J$,*

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) = -\frac{1}{\mathcal{J}}\frac{\partial(F,G)}{\partial(x_i, z)}, \qquad \frac{\partial g}{\partial x_i}(\boldsymbol{x}) = -\frac{1}{\mathcal{J}}\frac{\partial(F,G)}{\partial(y, x_i)},$$

onde

$$\mathcal{J} = \frac{\partial(F,G)}{\partial(y,z)} = \begin{vmatrix} \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z} \\ \frac{\partial G}{\partial y} & \frac{\partial G}{\partial z} \end{vmatrix}.$$

Demonstração. É muito simples a ideia da demonstração que vamos fazer (usando um método que poderia chamar-se «de substituição»): trata-se essencialmente de resolver uma das equações — digamos $F(\boldsymbol{x}, y, z) = 0$ — em ordem a uma das «incógnitas» — digamos z — substituindo depois o resultado obtido, $z = f^*(\boldsymbol{x}, y)$ na outra equação, $G(\boldsymbol{x}, y, z) = 0$, o que conduz a uma nova equação só com $\boldsymbol{x}$ e y, $H(\boldsymbol{x}, y) = G\big(\boldsymbol{x}, y, f^*(\boldsymbol{x}, y)\big) = 0$. Designando por $y = f(\boldsymbol{x})$ a solução desta equação e pondo $z = g(\boldsymbol{x}) = f^*\big(\boldsymbol{x}, f(\boldsymbol{x})\big)$, as funções f e g constituirão a solução do sistema.

Em termos precisos: o não anulamento do jacobiano no ponto $(\boldsymbol{a}, b, c)$ implica que alguma das derivadas $\frac{\partial F}{\partial y}$, $\frac{\partial F}{\partial z}$ será diferente de zero nesse ponto (de contrário seriam nulos os elementos da primeira linha do jacobiano e este não seria diferente de zero). Supondo, por exemplo, $\frac{\partial F}{\partial z}(\boldsymbol{a}, b, c) \neq 0$ poderá deduzir-se, nos termos do Teorema 4.12 e atendendo às restantes hipóteses do Teorema 4.13, que existem dois intervalos abertos, I' (de $\mathbb{R}^{n+1}$, centrado em $(\boldsymbol{a}, b)$) e J' (de $\mathbb{R}$, centrado no ponto c) tais que a cada par $(\boldsymbol{x}, y) \in I'$ corresponda um e um só ponto $z_{\boldsymbol{xy}} \in J'$ por forma que seja verificada a igualdade $F(\boldsymbol{x}, y, z_{\boldsymbol{xy}}) = 0$.

É óbvio que $z_{\boldsymbol{ab}} = c$ e também que, pondo $f^*(\boldsymbol{x}, y) = z_{\boldsymbol{xy}}$ e

$$H(\boldsymbol{x}, y) = G\big(\boldsymbol{x}, y, f^*(\boldsymbol{x}, y)\big),$$

(para cada par $(\boldsymbol{x}, y) \in I'$) se terá $H(\boldsymbol{a}, b) = 0$; e é também evidente que o par (y, z) será solução do sistema $F(\boldsymbol{x}, y, z) = 0, G(\boldsymbol{x}, y, z) = 0$ — com $(\boldsymbol{x}, y) \in I'$ e $z \in J'$ — sse for $z = f^*(\boldsymbol{x}, y)$ e $H(\boldsymbol{x}, y) = 0$.

Decorre ainda do Teorema 4.12 que a função f^* é de classe $\mathcal{C}^1$ e que se verifica a igualdade:

$$\frac{\partial f^*}{\partial y} = -\frac{\frac{\partial F}{\partial y}}{\frac{\partial F}{\partial z}}.$$

Nestas condições, ter-se-á:

$$\frac{\partial H}{\partial y} = \frac{\partial G}{\partial y} + \frac{\partial G}{\partial z}\frac{\partial f^*}{\partial y} = -\frac{1}{\frac{\partial F}{\partial z}}\frac{\partial(F, G)}{\partial(y, z)},$$

o que evidencia que $\frac{\partial H}{\partial y}(\boldsymbol{a}, b) \neq 0$, permitindo-nos portanto, por novo recurso ao Teorema 4.12, concluir que existe um intervalo aberto I (de $\mathbb{R}^n$, centrado em $\boldsymbol{a}$), e um intervalo aberto J'' (de $\mathbb{R}$, centrado em b), tais que a cada $\boldsymbol{x} \in I$ corresponde um e um só $y_{\boldsymbol{x}} \in J''$ por forma que $H(\boldsymbol{x}, y_{\boldsymbol{x}}) = 0$.

Pode naturalmente supor-se que os intervalos I e J'' são tais que $I \times J'' \subset I'$. Nestas condições, pondo $J = J'' \times J'$, $z_{\boldsymbol{x}} = f^*(\boldsymbol{x}, y_{\boldsymbol{x}})$, $f(\boldsymbol{x}) = y_{\boldsymbol{x}}$, $g(\boldsymbol{x}) = z_{\boldsymbol{x}}$, reconhece-se imediatamente que os intervalos I e J e as funções f e g satisfazem as condições referidas no enunciado do teorema, faltando apenas verificar as fórmulas relativas às derivadas parciais dessas funções.

Para esse efeito, derivem-se em ordem a x_i ambos os membros de cada uma das equações

$$F\big(x_1, \ldots, x_n, f(\boldsymbol{x}), g(\boldsymbol{x})\big) = 0 \qquad \text{e} \qquad G\big(x_1, \ldots, x_n, f(\boldsymbol{x}), g(\boldsymbol{x})\big) = 0,$$

o que conduz ao sistema:

$$\begin{cases} \dfrac{\partial F}{\partial x_i} + \dfrac{\partial F}{\partial y}\dfrac{\partial f}{\partial x_i} + \dfrac{\partial F}{\partial z}\dfrac{\partial g}{\partial x_i} = 0 \\ \dfrac{\partial G}{\partial x_i} + \dfrac{\partial G}{\partial y}\dfrac{\partial f}{\partial x_i} + \dfrac{\partial G}{\partial z}\dfrac{\partial g}{\partial x_i} = 0. \end{cases}$$

Bastará resolver este sistema pela regra de Cramer (considerando como incógnitas $\frac{\partial f}{\partial x_i}$ e $\frac{\partial g}{\partial x_i}$) para se obterem as fórmulas referidas no final do enunciado do teorema. $\square$

A título de exemplo, verifiquemos se o sistema

$$\begin{cases} F(x,y,u,v) = e^u + x\cos v = 0 \\ G(x,y,u,v) = e^u + y\,\mathrm{sen}\,v - 1 = 0 \end{cases}$$

define univocamente, nalguma vizinhança do ponto $(-1,1,0,0)$,

1.º x e y como funções de u e v;

2.º u e v como funções de x e y.

No primeiro caso, como o jacobiano $\frac{\partial(F,G)}{\partial(x,y)} = \frac{1}{2}\,\mathrm{sen}\,2v$ se anula no ponto considerado, o Teorema 4.13 não é aplicável. Basta, porém, resolver em ordem a y a equação $G(x,y,u,v) = 0$ para se reconhecer que não poderá existir uma vizinhança do ponto $(0,0)$ tal que a cada par (u,v) pertencente a essa vizinhança corresponda um par (x,y) por forma que essa equação seja verificada.

Para analisar a possiblidade de considerar definidas pelo sistema dado as variáveis u e v como funções de x e y, nalguma vizinhança do ponto $(-1,1,0,0)$, interessa considerar o jacobiano $\frac{\partial(F,G)}{\partial(u,v)}$, que assume nesse ponto o valor 1; desta vez, portanto a conclusão seria afirmativa.

Supondo $u = f(x,y)$, $v = g(x,y)$ — com o par (x,y) «próximo» de $(-1,1)$ e o par (u,v) «próximo» de $(0,0)$ — se pretendêssemos determinar os planos tangentes[13] às superfícies de equações $u = f(x,y)$ e $v = g(x,y)$ no

[13]Supondo $\varphi(x,y)$ diferenciável no ponto (x_0,y_0), a equação do plano tangente à superfície de equação $z = \varphi(x,y)$ no ponto (x_0,y_0,z_0) (onde $z_0 = \varphi(x_0,y_0)$) é:

$$z = z_0 + \frac{\partial \varphi}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial \varphi}{\partial y}(x_0,y_0)(y-y_0).$$

Mais geralmente, sendo $f : D \to \mathbb{R}$ (com $D \subset \mathbb{R}^n$) uma função diferenciável no ponto $\boldsymbol{a} = (a_1,\dots,a_n)$, uma equação do "hiperplano" tangente à "hipersuperfície" $\boldsymbol{y} = f(\boldsymbol{x})$ no ponto $(a_1,\dots,a_n,f(\boldsymbol{a}))$ é:

$$\boldsymbol{y} = f(a_1,\dots,a_n) + \frac{\partial f}{\partial x_1}(\boldsymbol{a})(x_1-a_1) + \cdots + \frac{\partial f}{\partial x_n}(\boldsymbol{a})(x_n-a_n)$$

ou, usando notação mais condensada:

$$\boldsymbol{y} = f(\boldsymbol{a}) + f'(\boldsymbol{a})(\boldsymbol{x}-\boldsymbol{a}).$$

ponto $(-1,1,0)$, bastaria derivar ambos os membros de cada uma das equações $F(x,y,u,v)=0$, $G(x,y,u,v)=0$ em ordem a x e também em ordem a y (considerando u e v como funções de x e y) e resolver os dois sistemas obtidos em relação às «incógnitas» $\frac{\partial u}{\partial x}, \frac{\partial v}{\partial x}$ e $\frac{\partial u}{\partial y}, \frac{\partial v}{\partial y}$ (depois de substituir as variáveis x,y,u,v pelas coordenadas correspondentes do ponto $(-1,1,0,0)$). Obter-se-iam assim os sistemas:

$$\begin{cases} e^u \frac{\partial u}{\partial x} + \cos v - x \operatorname{sen} v \frac{\partial v}{\partial x} = 0 \\ e^u \frac{\partial u}{\partial x} + y \cos v \frac{\partial v}{\partial x} = 0, \end{cases} \qquad \begin{cases} e^u \frac{\partial u}{\partial y} - x \operatorname{sen} v \frac{\partial v}{\partial y} = 0 \\ e^u \frac{\partial u}{\partial y} + \operatorname{sen} v + y \cos v \frac{\partial v}{\partial y} = 0 \end{cases}$$

e portanto, no ponto considerado:

$$\begin{cases} \frac{\partial u}{\partial x} + 1 = 0 \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} = 0, \end{cases} \qquad \begin{cases} \frac{\partial u}{\partial y} = 0 \\ \frac{\partial u}{\partial y} + \frac{\partial v}{\partial y} = 0. \end{cases}$$

Segue-se que as equações dos planos tangentes são, respectivamente,

$$u = -x - 1 \qquad \text{e} \qquad v = x + 1$$

Enunciaremos agora o teorema das funções implícitas, na forma mais geral aqui considerada.

Teorema 4.14. *Seja D um aberto de $\mathbb{R}^{m+n}$, $(\boldsymbol{a},\boldsymbol{b}) = (a_1,\ldots,a_n,b_1,\ldots,b_m)$ um ponto de D e, para cada $j \in \{1,\ldots,m\}$, $F_j(\boldsymbol{x},\boldsymbol{y}) = F_j(x_1,\ldots,x_n,y_1,\ldots,y_m)$ uma função definida e de classe $\mathcal{C}^1$ em D; suponham-se ainda verificadas as condições:*

$$F_j(\boldsymbol{a},\boldsymbol{b}) = 0 \quad (j \in \{1,\ldots,m\})$$

e

$$\frac{\partial(F_1,\ldots,F_m)}{\partial(y_1,\ldots,y_m)}(\boldsymbol{a},\boldsymbol{b}) \neq 0$$

Então:

1. *existe um intervalo aberto I (de $\mathbb{R}^n$, centrado em $\boldsymbol{a}$) e um intervalo aberto J (de $\mathbb{R}^m$, centrado em $\boldsymbol{b}$) tais que a cada $\boldsymbol{x} \in I$ corresponde um e só um $\boldsymbol{y_x} = (y_{1\boldsymbol{x}},\ldots,y_{m\boldsymbol{x}}) \in J$ por forma que $F_j(\boldsymbol{x},\boldsymbol{y_x}) = 0$ (para $j = 1,\ldots,m$);*

2. *pondo $f_1(\boldsymbol{x}) = y_{1\boldsymbol{x}},\ldots,f_m(\boldsymbol{x}) = y_{m\boldsymbol{x}}$, as funções $f_1,\ldots,f_m$ são de classe $\mathcal{C}^1$ em I e tem-se, para $i \in \{1,\ldots,n\}$ e $j \in \{1,\ldots,m\}$:*

$$\frac{\partial f_j}{\partial x_i} = -\frac{\frac{\partial(F_1,F_2,\ldots,F_m)}{\partial(y_1,\ldots,y_{j-1},x_i,y_{j+1},\ldots,y_m)}}{\frac{\partial(F_1,F_2,\ldots,F_m)}{\partial(y_1,y_2,\ldots,y_m)}}.$$

Demonstração. A demonstração pode fazer-se por indução (sobre m): assegurada a veracidade da proposição no caso $m=1$ (pelo Teorema 4.12), admita-se, como hipótese de indução, a sua validade quando se considerem sistemas de $m-1$

equações em $m-1$ «incógnitas» (qualquer que seja o número de variáveis independentes) e comecemos por observar que, sendo diferente de zero no ponto $(\boldsymbol{a}, \boldsymbol{b})$ o jacobiano

$$\mathcal{J} = \begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \dots & \frac{\partial F_1}{\partial y_m} \\ \dots & \dots & \dots \\ \frac{\partial F_m}{\partial y_1} & \dots & \frac{\partial F_m}{\partial y_m} \end{vmatrix}$$

sê-lo-á também pelo menos um dos determinantes de ordem $m-1$ que podem obter-se suprimindo-lhe a última linha e uma das suas colunas[14]. Assim, alterando, se necessário, a ordenação das colunas do determinante $\mathcal{J}$, podemos supor que é diferente de zero em $(\boldsymbol{a}, \boldsymbol{b})$ o determinante

$$\mathcal{J}^* = \begin{vmatrix} \frac{\partial F_1}{\partial y_1} & \dots & \frac{\partial F_1}{\partial y_{m-1}} \\ \dots & \dots & \dots \\ \frac{\partial F_{m-1}}{\partial y_1} & \dots & \frac{\partial F_{m-1}}{\partial y_{m-1}} \end{vmatrix}.$$

Utilizando a hipótese de indução reconhece-se então a existência de intervalos I' (de $\mathbb{R}^{n+1}$, centrado no ponto $(a_1, \dots, a_n, b_m)$) e J' de (de $\mathbb{R}^{m-1}$, centrado em $(b_1, \dots, b_{m-1})$) tais que a cada vector $(\boldsymbol{x}, y_m) = (x_1, \dots, x_n, y_m) \in I'$ corresponda um e um só $\boldsymbol{y}' = (y_1', \dots, y_{m-1}') \in J'$ por forma que se tenha (para cada $j \in \{1, \dots, m-1\}$):

$$F_j(\boldsymbol{x}, y_1', \dots, y_{m-1}', y_m) = 0. \tag{4.1}$$

Ponhamos então:

$$f_j^*(\boldsymbol{x}, y_m) = f_j^*(x_1, \dots, x_n, y_m) = y_j' \qquad (j \in \{1, \dots, m-1\})$$

e ainda

$$H(\boldsymbol{x}, y_m) = F_m\big(\boldsymbol{x}, f_1^*(\boldsymbol{x}, y_m), \dots, f_{m-1}^*(\boldsymbol{x}, y_m), y_m\big). \tag{4.2}$$

Se verificarmos que $\frac{\partial H}{\partial y_m}(\boldsymbol{a}, b_m) \neq 0$, o Teorema 4.12 permitirá reconhecer que a equação $H(\boldsymbol{x}, y_m) = 0$ poderá ser resolvida (localmente) em ordem a y_m, donde decorrerão facilmente os resultados que pretendemos provar. Para tal derivemos em ordem a y_m as $m-1$ equações(4.1) (tendo em conta que $y_j' = f_j^*(\boldsymbol{x}, y_m)$, para $j = 1, \dots, m-1$) e ainda a equação (4.2).

Obteremos o sistema:

$$\begin{cases} \displaystyle\sum_{k=1}^{m-1} \frac{\partial F_j}{\partial y_k} \frac{\partial f_k^*}{\partial y_m} + \frac{\partial F_j}{\partial y_m} = 0 & \qquad (j \in \{1, \dots, m-1\}) \\ \displaystyle\sum_{k=1}^{m-1} \frac{\partial F_m}{\partial y_k} \frac{\partial f_k^*}{\partial y_m} - \frac{\partial H}{\partial y_m} + \frac{\partial F_m}{\partial y_m} = 0. \end{cases}$$

[14]Recorde-se que, de acordo com um Teorema de Laplace, o determinante $\mathcal{J}$ é igual à soma dos produtos que se obtêm multiplicando cada um dos elementos da sua última linha pelos respectivos complementos algébricos; e também que, a menos do sinal, estes complementos algébricos são precisamente os determinantes de ordem $m-1$ acima mencionados.

A resolução deste sistema (considerando como incógnitas $\frac{\partial f_1^*}{\partial y_m}, \ldots, \frac{\partial f_{m-1}^*}{\partial y_m}$ e $\frac{\partial H}{\partial y_m}$), conduz imediatamente à igualdade:

$$\frac{\partial H}{\partial y_m} = \frac{\mathcal{J}}{\mathcal{J}^*},$$

donde decorre o não anulamento da derivada $\frac{\partial H}{\partial y_m}$ que pretendíamos verificar.

Assim, pode garantir-se que existe um intervalo I (de $\mathbb{R}^n$, centrado no ponto $\boldsymbol{a}$) e um intervalo J'' (de $\mathbb{R}$, centrado em $\boldsymbol{b}$) por forma que a cada $\boldsymbol{x} \in I$ corresponda um único $y_{m\boldsymbol{x}} \in J''$ de modo que se verifique a igualdade $H(\boldsymbol{x}, y_{m\boldsymbol{x}}) = 0$.

É claro que podemos supor $I \times J'' \subset I'$; nestas condições, pondo $J = J' \times J''$ e, para cada $\boldsymbol{x} \in I$, $f_m(\boldsymbol{x}) = y_{m\boldsymbol{x}}$ e $f_j(\boldsymbol{x}) = f_j^*\big(\boldsymbol{x}, f_m(\boldsymbol{x})\big)$, para $j = 1, \ldots, m-1$, vê-se imediatamente que os intervalos I e J e as funções $f_1, \ldots, f_m$ satisfazem todas as condições mencionadas no enunciado do teorema, faltando apenas, para terminar a demonstração, verificar as fórmulas relativas às derivadas destas funções. Para este efeito bastará derivar em ordem a x_i ambos os membros de cada uma das equações do sistema dado, o que conduz ao sistema:

$$\frac{\partial F_j}{\partial x_i} + \frac{\partial F_j}{\partial y_1}\frac{\partial f_1}{\partial x_i} + \cdots + \frac{\partial F_j}{\partial y_m}\frac{\partial f_m}{\partial x_i} = 0 \qquad (j \in \{1, \ldots, m\})$$

e resolver este sistema considerando como incógnitas as derivadas $\frac{\partial f_j}{\partial x_i}$.

Aliás, como já foi observado, as fórmulas assim obtidas, para além de evidenciarem que, nas condições da hipótese do teorema, as funções f_j são de classe $\mathcal{C}^1$, permitem também reconhecer que estas funções seriam de classe $\mathcal{C}^p$ (com p inteiro > 1 ou $p = \infty$) se o mesmo se passasse com as funções F_j. $\square$

Uma aplicação simples do teorema das funções implícitas permite obter outro teorema importante, habitualmente designado por teorema da função inversa[15]. Preparando o enunciado desse teorema começaremos por recordar algumas definições e resultados muito correntes, que convém ter presentes no que vai seguir-se.

Como é bem sabido, sendo A e B dois conjuntos quaisquer e f uma aplicação injectiva (ou, como também se diz, invertível) de A em B, a inversa de f é a aplicação $f^{-1} : f(A) \to A$ tal que, para qualquer $x \in A$ e qualquer $y \in f(A)$, $f^{-1}(y) = x$ sse $f(x) = y$.

Segundo um resultado bem conhecido da teoria das funções reais de variável real, uma função contínua $f : I \to \mathbb{R}$ (onde I é um intervalo de $\mathbb{R}$) é injectiva sse for estritamente monótona; em tal caso o seu contradomínio é um intervalo J e a inversa $f^{-1} : J \to I$ é também contínua (e estritamente monótona). Suponhamos agora que o intervalo I é aberto e que a função f é de classe $\mathcal{C}^1$: então, para que f seja invertível e f^{-1} seja também de classe $\mathcal{C}^1$ é necessário e suficiente que seja

[15]Tal possibilidade de aplicação do teorema das funções implícitas é praticamente evidente: basta notar que (em termos pouco precisos) inverter uma função f equivale a resolver em ordem a x a equação $y - f(x) = 0$.

verificada a condição $f'(x) \neq 0$ qualquer que seja $x \in I$; nesta hipótese ter-se-á, como é sabido (designando agora por g a inversa de f):

$$g'\big(f(x)\big) = \big(f'(x)\big)^{-1},$$

para cada $x \in I$.

Convém agora notar que, em certos casos em que a função $f : I \to \mathbb{R}$ não é injectiva (nem, portanto, «globalmente» invertível, isto é, invertível na acepção anteriormente considerada), pode ter interesse analisar a possibilidade de inverter a restrição de f a alguma vizinhança de um ou outro ponto particular do seu domínio; e, quando tal inversão «local» é possível, interessa frequentemente estudar certas propriedades das inversas locais que assim podem obter-se. Por exemplo, se $f : I \to \mathbb{R}$ for uma função de classe $\mathcal{C}^p$ nalguma vizinhança de certo ponto a, é fácil ver que a condição $f'(a) \neq 0$ garante a existência de uma vizinhança U do ponto a tal que a restrição de f a U seja invertível e que f^{-1} seja também uma função de classe $\mathcal{C}^p$.

Exprime-se no enunciado do Teorema 4.15 uma extensão destes resultados ao quadro das funções definidas em abertos de $\mathbb{R}^n$ e com valores neste mesmo espaço.

Como era de esperar, nessa extensão desempenha também um papel fundamental o comportamento da derivada no ponto considerado; porém, no caso $n > 1$, a condição a impor para garantir a invertibilidade local de f (além das propriedades desejáveis de f^{-1}) não será já o não anulamento[16] da derivada $f'(a)$, mas sim que esta aplicação linear de $\mathbb{R}^n$ em si mesmo seja ela própria invertível. Ora para este efeito o que interessa (o que é necessário e suficiente) é que se não anule o jacobiano correspondente.

Antes de enunciar o teorema da aplicação inversa convém introduzir a definição seguinte: sendo p um inteiro ≥ 1 ou $p = \infty$, A e B dois conjuntos abertos de $\mathbb{R}^n$ e $f : A \to B$ uma aplicação bijectiva, diz-se que f é um *difeomorfismo de classe* $\mathcal{C}^p$ sse tanto f como f^{-1} forem funções de classe $\mathcal{C}^p$ (por exemplo, a função $f(x) = x^3$, suposta definida num aberto A de $\mathbb{R}$ é um difeomorfismo de classe $\mathcal{C}^\infty$ sse $0 \notin A$ e não é sequer um difeomorfismo de classe $\mathcal{C}^1$ se $0 \in A$). Convém recordar ainda que, sendo $\boldsymbol{a}$ um ponto qualquer de $\mathbb{R}^n$, é costume chamar *vizinhança de* $\boldsymbol{a}$ a qualquer subconjunto de $\mathbb{R}^n$ que contenha uma bola centrada no ponto $\boldsymbol{a}$.

Teorema 4.15 (Teorema da função inversa). *Seja D um aberto de $\mathbb{R}^n$, $f : D \to \mathbb{R}^n$ uma função de classe $\mathcal{C}^1$ definida pelo sistema:*

$$\begin{aligned} y_1 &= f_1(x_1, \ldots, x_n) \\ &\cdots \\ y_n &= f_n(x_1, \ldots, x_n) \end{aligned}$$

[16] Observe-se que até no caso particular de a função f ser ela própria uma aplicação linear de $\mathbb{R}^n$ em si mesmo (caso em que, qualquer que seja $a \in \mathbb{R}^n$, $f'(a) = f$) a condição $f \neq 0$ é claramente insuficiente para garantir a invertibilidade de f (excepto se for $n = 1$).

(abreviadamente $\boldsymbol{y} = f(\boldsymbol{x})$). Seja ainda $\boldsymbol{a} = (a_1, \ldots, a_n)$ um ponto de D, $\boldsymbol{b} = (b_1, \ldots, b_n) = f(\boldsymbol{a})$ e suponha-se que

$$\frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)}(\boldsymbol{a}) \neq 0$$

(isto é, que a derivada $f'(\boldsymbol{a})$ é uma aplicação bijectiva de $\mathbb{R}^n$ sobre si mesmo).

Nestas condições existe uma vizinhança aberta U do ponto $\boldsymbol{a}$ tal que a restrição de f a U é um difeomorfismo de classe $\mathcal{C}^1$ e, designando por g a inversa de $f_{|U}$, tem-se, em qualquer ponto $\boldsymbol{x} \in U$,

$$g'\big(f(\boldsymbol{x})\big) = \big(f'(\boldsymbol{x})\big)^{-1}.$$

Demonstração. Pondo

$$\begin{aligned} F_1(x_1, \ldots, x_n, y_1, \ldots, y_n) &= f_1(x_1, \ldots, x_n) - y_1 \\ &\cdots \\ F_n(x_1, \ldots, x_n, y_1, \ldots, y_n) &= f_n(x_1, \ldots, x_n) - y_n, \end{aligned}$$

as funções $F_i(\boldsymbol{x}, \boldsymbol{y})$ serão de classe $\mathcal{C}^1$ no aberto (de $\mathbb{R}^{2n}$) $D \times \mathbb{R}^n$ e ter-se-á, para cada $i \in \{1, \ldots, n\}$,

$$F_i(\boldsymbol{a}, \boldsymbol{b}) = 0, \qquad \text{e} \qquad \frac{\partial(F_1, \ldots, F_n)}{\partial(x_1, \ldots, x_n)}(\boldsymbol{a}, \boldsymbol{b}) \neq 0.$$

O Teorema 4.14 garante então a existência de um intervalo aberto J (de $\mathbb{R}^n$, centrado em $\boldsymbol{b}$) e de um intervalo aberto I (de $\mathbb{R}^n$, centrado em $\boldsymbol{a}$) tais que para cada $\boldsymbol{y} \in J$ exista um e só um $\boldsymbol{x}_{\boldsymbol{y}} \in I$ tal que $\boldsymbol{y} = f(\boldsymbol{x}_{\boldsymbol{y}})$; e ainda que, se pusermos $g(\boldsymbol{y}) = \boldsymbol{x}_{\boldsymbol{y}}$ para cada $\boldsymbol{y} \in J$, a função g será de classe $\mathcal{C}^1$ em J. Designando por U a imagem de J por g, $g(J) = U$, reconhece-se imediatamente, não só que $U \subset I$, como também que $\boldsymbol{a} \in U$ (visto que $\boldsymbol{a} = g(\boldsymbol{b})$ e $\boldsymbol{b} \in J$) e ainda que $f_{|U}$ é uma aplicação bijectiva de U sobre J, precisamente a inversa de g (na realidade tem-se:

$$g \circ f_{|U} = I_U, \qquad f_{|U} \circ g = I_J,$$

designando por I_U a aplicação idêntica definida em U e analogamente para I_J).

Para terminar a prova de que $f_{|U}$ é um difeomorfismo de classe $\mathcal{C}^1$ falta apenas verificar que o conjunto U é aberto (porque é óbvio que então a restrição de f a U será, tal como f, uma função de classe $\mathcal{C}^1$). Para tal, comecemos por notar que a imagem por f de um ponto que pertença ao conjunto $I \backslash U$ (se este conjunto não for vazio) não poderá pertencer a J (visto que para cada $\boldsymbol{y} \in J$ existe um e um só $\boldsymbol{x}_{\boldsymbol{y}}$ em I que tem $\boldsymbol{y}$ por imagem e esse ponto $\boldsymbol{x}_{\boldsymbol{y}} = g(\boldsymbol{y})$ pertence necessariamente a U). Agora, se $\boldsymbol{c}$ for um ponto qualquer de U ter-se-á $f(\boldsymbol{c}) \in J$ e, por J ser aberto, existirá $\delta > 0$ tal que $B_\delta\big(f(\boldsymbol{c})\big) \subset J$. Por outro lado, como a função f é contínua em $\boldsymbol{c}$, para algum $\epsilon > 0$ (que pode supor-se suficientemente pequeno para que $B_\epsilon(\boldsymbol{c})$ esteja contida no aberto I) se terá

$$f\big(B_\epsilon(\boldsymbol{c})\big) \subset B_\delta\big(f(\boldsymbol{c})\big) \subset J.$$

Dado, porém, que em $B_\epsilon(\boldsymbol{c})$ não pode existir qualquer ponto de $I \backslash U$ (pois que, como observámos há pouco, as imagens por f de tais pontos não pertencem a J) pode concluir-se que $B_\epsilon(\boldsymbol{c}) \subset U$ e portanto que U é aberto.

Seja agora $\boldsymbol{x}$ um ponto qualquer de U. Da igualdade $g \circ f_{|U} = I_U$ segue-se imediatamente, atendendo à regra de derivação das funções compostas e ao facto de se ter $I'_U(x) = I$, designando por I a aplicação idêntica de $\mathbb{R}^n$ em si mesmo (visto que I_U é a restrição da aplicação linear I a um aberto que contém x),

$$g'\big(f(\boldsymbol{x})\big) \circ f'(\boldsymbol{x}) = I.$$

Finalmente desta relação, tendo em conta o facto de $g'(\boldsymbol{x})$ ser uma aplicação bijectiva[17], decorre a igualdade $g'\big(f(\boldsymbol{x})\big) = \big(f'(\boldsymbol{x})\big)^{-1}$, que pretendíamos provar. □

Convém notar agora que, passando às matrizes jacobianas correspondentes às aplicações lineares que figuram na igualdade $g'\big(f(\boldsymbol{x})\big) \circ f'(\boldsymbol{x}) = I$, se obtem a relação:

$$\begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_j} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_i}{\partial y_1} & \cdots & \frac{\partial x_i}{\partial y_j} & \cdots & \frac{\partial x_i}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_j} & \cdots & \frac{\partial x_n}{\partial y_n} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_i} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial y_j}{\partial x_1} & \cdots & \frac{\partial y_j}{\partial x_i} & \cdots & \frac{\partial y_j}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_i} & \cdots & \frac{\partial y_n}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix},$$

onde as funções $\frac{\partial x_i}{\partial y_j}$ se supõem calculadas no ponto $f(\boldsymbol{x})$, as funções $\frac{\partial y_j}{\partial x_i}$ no ponto $\boldsymbol{x}$ e onde a matriz que figura no 2º membro é evidentemente a matriz identidade de ordem n. Por sua vez desta igualdade decorre imediatamente[18] a relação entre os jacobianos

$$\frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} = \left(\frac{\partial(y_1, \ldots, y_n)}{\partial(x_1, \ldots, x_n)}\right)^{-1},$$

que generaliza a fórmula $\frac{dx}{dy} = 1/\frac{dy}{dx}$, correspondente ao caso $n = 1$.

Por outro lado, efectuando o produto das matrizes do primeiro membro e igualando-o à matriz identidade, obtêm-se as n^2 igualdades:

$$\frac{\partial x_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \frac{\partial x_i}{\partial y_2}\frac{\partial y_2}{\partial x_j} + \cdots + \frac{\partial x_i}{\partial y_n}\frac{\partial y_n}{\partial x_j} = \delta_{ij} \qquad (i, j = 1, \ldots, n),$$

(onde $\delta_{ij} = 1$ se $i = j$ e $\delta_{ij} = 0$ se $i \neq j$).

[17] É muito fácil verificar que uma aplicação linear $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ é injectiva sse for sobrejectiva (e portanto bijectiva); e ainda que, para que seja bijectiva a composta de duas aplicações lineares de $\mathbb{R}^n$ em si mesmo, é necessário e suficiente que ambas o sejam.

[18] Atendendo a que o determinante do produto de duas matrizes (quadradas, da mesma ordem) é igual ao produto dos determinantes dessas matrizes.

Se se pretender determinar as primeiras derivadas parciais da função $x_i = g_i(y_1, \ldots, y_n)$, bastará resolver em ordem a essas derivadas o sistema de n equações que se obtem se, nas igualdades anteriores, fixarmos i e fizermos $j = 1, \ldots, n$.

O resultado obtido,

$$\frac{\partial x_i}{\partial y_j} = (-1)^{i+j} \frac{\frac{\partial(y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_n)}{\partial(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)}}{\frac{\partial(y_1,\ldots,y_n)}{\partial(x_1,\ldots,x_n)}},$$

permite reconhecer uma vez mais que, se no enunciado do teorema da função inversa substituíssemos a hipótese $f \in \mathcal{C}^1(D)$ por $f \in \mathcal{C}^p(D)$ (p inteiro ≥ 1 ou $p = \infty$), poderíamos concluir que g seria uma função de classe $\mathcal{C}^p$ (e portanto um difeomorfismo de classe $\mathcal{C}^p$).

A título de exemplo, consideremos a função f definida pelo sistema:

$$\begin{cases} u = x + \cos \frac{1}{y} \\ v = 1 + \cos \frac{1}{x} \end{cases}$$

no aberto $D \subset \mathbb{R}^2$ formado pelos pontos (x, y) tais que $xy \neq 0$.

Como $f \in \mathcal{C}^\infty(D)$ e o jacobiano

$$\frac{\partial(u, v)}{\partial(x, y)} = -\frac{1}{x^2 y^2} \operatorname{sen} \frac{1}{x} \operatorname{sen} \frac{1}{y}$$

se anula apenas nos pontos $(x, y) \in D$ que verificam pelo menos uma das condições $x = \frac{1}{k\pi}$ ou $y = \frac{1}{\ell\pi}$ (para algum valor de $k, \ell \in \mathbb{Z}\backslash\{0\}$), podemos concluir que qualquer outro ponto de D tem uma vizinhança U tal que $f_{|U}$ é um difeomorfismo de classe $\mathcal{C}^\infty$; e é aliás muito fácil verificar que, em relação aos pontos de D em que o jacobiano se anula, não há de facto possibilidade de inverter a função, mesmo localmente.

Se pretendêssemos determinar a matriz jacobiana da função inversa num dos pontos em que a inversão é possível, bastaria derivar em ordem a u e também em ordem a v o sistema que define a função f, o que conduziria a:

$$\begin{cases} 1 = \frac{\partial x}{\partial u} + \frac{1}{y^2} \operatorname{sen} \frac{1}{y} \frac{\partial y}{\partial u} \\ 0 = \frac{1}{x^2} \operatorname{sen} \frac{1}{x} \frac{\partial x}{\partial u} \end{cases} \qquad \begin{cases} 0 = \frac{\partial x}{\partial v} + \frac{1}{y^2} \operatorname{sen} \frac{1}{y} \frac{\partial y}{\partial v} \\ 1 = \frac{1}{x^2} \operatorname{sen} \frac{1}{x} \frac{\partial x}{\partial v} \end{cases},$$

donde resulta:

$$\frac{\partial x}{\partial u} = 0, \quad \frac{\partial y}{\partial u} = \frac{y^2}{\operatorname{sen} \frac{1}{y}}, \quad \frac{\partial x}{\partial v} = \frac{x^2}{\operatorname{sen} \frac{1}{x}}, \quad \frac{\partial y}{\partial v} = -\frac{x^2 y^2}{\operatorname{sen} \frac{1}{x} \operatorname{sen} \frac{1}{y}}.$$

Assim, por exemplo, no ponto $(x, y) = \left(\frac{2}{\pi}, \frac{2}{\pi}\right)$ a matriz jacobiana da inversa de f (definida numa vizinhança conveniente do ponto $f\left(\frac{2}{\pi}, \frac{2}{\pi}\right) = \left(\frac{2}{\pi}, 1\right)$), seria

$$\begin{bmatrix} 0 & \left(\frac{2}{\pi}\right)^2 \\ \left(\frac{2}{\pi}\right)^2 & -\left(\frac{2}{\pi}\right)^4 \end{bmatrix}.$$

Uma consequência importante do teorema da função inversa é o

Teorema 4.16 (Teorema da aplicação aberta). *Seja D um aberto de $\mathbb{R}^n$, $f : D \to \mathbb{R}^n$ uma função de classe $\mathcal{C}^1$ e suponha-se que, para cada $\boldsymbol{x} \in D$, a aplicação linear $f'(\boldsymbol{x})$ é bijectiva. Então a imagem por f de qualquer subconjunto aberto de D é um conjunto aberto.*

Demonstração. Seja $A \subset D$, A aberto, e seja $\boldsymbol{y}$ um ponto qualquer de $f(A)$. Escolhido um ponto $\boldsymbol{x} \in A$ tal que $\boldsymbol{y} = f(\boldsymbol{x})$, basta aplicar o teorema da função inversa à restrição de f ao conjunto A, $f_{|A}$, em relação ao ponto $\boldsymbol{x}$, para se poder garantir a existência de uma vizinhança aberta U de $\boldsymbol{x}$ contida em A, tal que $f(U)$ é um subconjunto aberto de $f(A)$ e, evidentemente, contém $\boldsymbol{y}$; assim, como cada ponto $\boldsymbol{y} \in f(A)$ tem uma vizinhança contida em $f(A)$, pode concluir-se que este conjunto é aberto. □

4.5 Extremos

Recordámos no parágrafo 3.1 as noções de máximo e mínimo de uma função $f : D \to \mathbb{R}$ num conjunto $A \subset D$, às quais se referem as notações $\max_A f, \min_A f$. Como sabemos é frequente o uso do termo *extremo* para designar indistintamente um máximo ou um mínimo, podendo recorrer-se ao adjectivo *absoluto* (máximo absoluto, mínimo absoluto) para precisar que a noção considerada se refere a todo o domínio da função, isto é, que se trata de $\max_D f$ (também designado apenas por $\max f$) ou $\min_D f$ ($\min f$).

Em muitos casos, porém, interessa considerar os chamados *extremos relativos* (ou *extremos locais*), cujas definições recordaremos agora. Sendo ainda $f : D \to \mathbb{R}$ (com $D \subset \mathbb{R}^n$) e $\boldsymbol{a}$ um ponto de D, diz-se que $\boldsymbol{a}$ é um *ponto de máximo* (ou um *maximizante*) *relativo* da função f, ou ainda que $f(\boldsymbol{a})$ é um *máximo relativo* de f sse existe $\epsilon > 0$ tal que $f(\boldsymbol{x}) \leq f(\boldsymbol{a})$ sempre que $\boldsymbol{x} \in D$ e $\|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$. Se, para algum $\epsilon > 0$, for verificada a condição $f(\boldsymbol{x}) < f(\boldsymbol{a})$ em qualquer ponto x tal que $x \in D$ e $0 < \|\boldsymbol{x} - \boldsymbol{a}\| < \epsilon$, dir-se-á que o máximo relativo $f(\boldsymbol{a})$ é *estrito*. Evidentemente, as definições de *mínimo relativo* e *mínimo relativo estrito* são análogas. É também óbvio que um máximo (ou mínimo) absoluto é também máximo (ou mínimo) relativo.

Assim, por exemplo, a função $\varphi : \mathbb{R}^3 \to \mathbb{R}$ definida pela fórmula $\varphi(x, y, z) = x^2 + y^2 + z^2$ não tem qualquer máximo mas tem um mínimo (absoluto, estrito) assumido na origem do espaço $\mathbb{R}^3$; e é também fácil reconhecer que a função $\psi(x) = x \operatorname{sen} x$ tem um único extremo relativo (estrito), em cada um dos intervalos $[(2k-1)\frac{\pi}{2}, (2k+1)\frac{\pi}{2}]$, com $k \in \mathbb{Z}$ (mínimo se k é par, máximo se k é ímpar), não sendo nenhum deles extremo absoluto.

Um resultado por vezes útil na pesquisa de extremos é o teorema de Weierstrass (teorema 3.4): se o conjunto (não vazio) D for compacto qualquer função definida e contínua em D tem máximo e mínimo absolutos. Outro resultado muito simples e do maior interesse para o mesmo objectivo é o que se exprime no seguinte:

Teorema 4.17. *Seja $f : D \to \mathbb{R}$, com $D \subset \mathbb{R}^n$; então se f é diferenciável no ponto*[19] $\boldsymbol{a}$ *e $f(\boldsymbol{a})$ é um extremo (relativo) de f, tem-se $\frac{\partial f}{\partial x_i}(\boldsymbol{a}) = 0$ para qualquer $i \in \{1, \ldots, n\}$ isto é, a derivada de f no ponto $\boldsymbol{a}$, $f'(\boldsymbol{a})$, é a aplicação nula de $\mathbb{R}^n$ em $\mathbb{R}$.*

Demonstração. Nas condições da hipótese, e supondo $\boldsymbol{a} = (a_1, \ldots, a_n)$, para qualquer $i \in \{1, \ldots, n\}$, a função definida (para todos os valores suficientemente pequenos de $|t|$) pela fórmula $\psi(t) = f(\boldsymbol{a} + t\boldsymbol{e}_i)$ terá um extremo no ponto 0, o que implica o anulamento da derivada $\psi_i'(0) = \frac{\partial f}{\partial x_i}(\boldsymbol{a})$. □

Os pontos (interiores) de D nos quais se anula a derivada de f são chamados *pontos de estacionaridade* ou *pontos críticos* da função; assim, de acordo com o teorema anterior, para que $f(\boldsymbol{a})$ seja extremo (com f diferenciável em $\boldsymbol{a}$) é necessário que $\boldsymbol{a}$ seja ponto de estacionaridade de f.

Sabemos bem que esta condição não é suficiente (por exemplo, $f(x) = x^3$ tem um ponto de estacionaridade na origem sem que tenha qualquer extremo nesse ponto); e sabemos também que pode haver extremos em pontos que não são de estacionaridade: pontos do domínio que não sejam interiores ou então pontos interiores do domínio nos quais a função não seja diferenciável (por exemplo, a restrição de $\sqrt{x^2 + y^2}$ ao círculo $x^2 + y^2 \leq 1$ assume o máximo nos pontos da circunferência que limita esse círculo e o mínimo na origem sem que qualquer destes seja ponto crítico).

Convém referir ainda que os pontos de estacionaridade em que a função não tem extremo são por vezes chamados *pontos de sela.*

Antes de vermos alguns exemplos registaremos o seguinte resultado, que é uma consequência muito simples dos teoremas acabados de mencionar e que pode ser considerado como uma generalização do teorema de Rolle ao quadro das funções reais de mais de uma variável real:

Seja D um aberto limitado não vazio de $\mathbb{R}^n$ e f uma função real contínua na aderência de D, diferenciável em todos os pontos deste conjunto e cuja restrição à fronteira de D é uma função constante; então f' anula-se em algum ponto de D.

A demonstração, praticamente idêntica à do caso $n = 1$ poderá ficar como exercício.

A título de exemplo, consideremos agora a função $f : \mathbb{R}^2 \to \mathbb{R}$,

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

Verifica-se imediatamente que os pontos críticos são, além da origem, os quatro pontos $(1, 1)$, $(1, -1)$, $(-1, 1)$ e $(-1, -1)$. Dado que f assume valores positivos em todos os pontos dos quadrantes ímpares e valores negativos nos pontos dos quadrantes pares (quadrantes abertos), logo se vê que a origem

[19] Recorde-se que, de acordo com a definição de diferenciabilidade que adoptámos, o facto de f ser diferenciável em $\boldsymbol{a}$ exige que este ponto seja interior ao domínio de f.

é um ponto de sela; portanto os únicos extremantes possíveis são os pontos $(1,1)$ e $(-1,-1)$, onde f assume o valor $\frac{1}{e}$, ou $(1,-1)$ e $(-1,1)$ onde o valor de f é $-\frac{1}{e}$. Para ver que estes valores são de facto extremos de f (e até extremos absolutos), comecemos por observar que $f(x,y)$ tende para 0 quando $\|(x,y)\| \to \infty$ (como se reconhece imediatamente, por exemplo se passarmos a coordenadas polares); assim, será possível determinar um número $k > 0$ tal que para $\|(x,y)\| \geq k$ se tenha $|f(x,y)| < \frac{1}{2e}$. Como a restrição de f ao compacto $K = \{(x,y) : \|(x,y)\| \leq k\}$ é contínua, deverá assumir um máximo e um mínimo (absolutos) em pontos de K, pontos decerto interiores a K porque os valores assumidos por f na fronteira e no exterior de K têm módulo menor do que $\frac{1}{2e}$ e $f(1,1) = \frac{1}{e}$, $f(1,-1) = -\frac{1}{e}$; porém, sendo f diferenciável, esses pontos serão necessariamente pontos de estacionaridade e terão portanto de coincidir com alguns dos quatro pontos $(1,1)$, $(1,-1)$, $(-1,1)$ e $(-1,-1)$. Pode então concluir-se que os valores $f(1,1) = f(-1,-1) = \frac{1}{e}$ e $f(-1,1) = f(1,-1) = -\frac{1}{e}$ são extremos relativos (e portanto absolutos) da própria função f, visto que $|f(x,y)| < \frac{1}{2e}$ para $(x,y) \notin K$; daqui decorre também, atendendo aos teoremas 2.10, 3.7 e 3.9 e ao facto evidente de $\mathbb{R}^2$ ser conexo por arcos, que o contradomínio de f é o intervalo $\left[-\frac{1}{e}, \frac{1}{e}\right]$.

Consideremos agora a função definida em $\mathbb{R}^2$ pela fórmula:

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

Como $g(x,y) = (x^2 - y^2)^2 + x^2y^2$ só assume valores não negativos logo se vê que $g(0,0) = 0$ é o mínimo absoluto de função; por outro lado, sendo g contínua e $\lim_{\|(x,y)\|\to\infty} g(x,y) = +\infty$, poderá também concluir-se que o contradomínio de g é o intervalo $[0, +\infty[$. Existirão extremos relativos de g, para além do mínimo absoluto? Se existissem, deveriam ser atingidos em pontos de estacionaridade, visto que g é diferenciável em $\mathbb{R}^2$. Porém, dado que o sistema:

$$\begin{cases} \frac{\partial g}{\partial x} = 2x(\sqrt{2}x - y)(\sqrt{2}x + y) = 0 \\ \frac{\partial g}{\partial y} = 2y(\sqrt{2}y - x)(\sqrt{2}y + x) = 0 \end{cases}$$

tem $(0,0)$ como solução única, logo se conclui que g não tem quaisquer outros pontos de extremo.

Seja ainda

$$h(x,y) = x^2 + 3y^4 - 4y^3 - 12y^2,$$

cujos pontos críticos são $(0,0)$, $(0,2)$ e $(0,-1)$, aos quais correspondem respectivamente os valores da função 0, -32 e -5. Sendo fácil verificar que $h(x,y)$ tende para $+\infty$ quando $\|(x,y)\| \to \infty$, poderá concluir-se, como num dos exemplos precedentes, que $h(0,2)$ é o mínimo absoluto da função e que o seu contradomínio é o intervalo $[-32, +\infty[$; também é fácil reconhecer que o ponto $(0,0)$ é um ponto de sela: basta notar que a função h assume valores positivos em todos os pontos do eixo das abcissas, com excepção da origem, e valores negativos nos pontos do eixo das ordenadas distintos da origem mas suficientemente próximos dela.

Já a determinação da natureza do ponto crítico $(0, -1)$ não será tão simples. Poderemos começar por transferir para esse ponto a origem do sistema de coordenadas (mediante uma translação dos eixos) e, simultaneamente, passar a coordenadas polares; efectuada uma tal mudança de variáveis (que poderá supor-se definida pelo sistema $x = r\cos\theta, y = -1 + r\operatorname{sen}\theta$) obteremos a igualdade:

$$\begin{aligned} h(x,y) - h(0,-1) &= h(r\cos\theta, -1 + r\operatorname{sen}\theta) + 5 \\ &= r^2(18\operatorname{sen}^2\theta + \cos^2\theta + 3r^2\operatorname{sen}^4\theta - 16r\operatorname{sen}^3\theta), \end{aligned}$$

a qual nos vai permitir reconhecer que, para (x, y) "próximo" de $(0, -1)$ mas distinto deste ponto (isto é, para r "próximo" de 0 mas positivo), o valor $h(x, y)$ é sempre maior do que $h(0, -1)$, tendo portanto a função h um mínimo relativo estrito no ponto considerado. Para tal será suficiente mostrar que, se o número positivo r for suficientemente pequeno, todos os valores da função

$$18\operatorname{sen}^2\theta + \cos^2\theta + 3r^2\operatorname{sen}^4\theta - 16r\operatorname{sen}^3\theta$$

serão positivos; para este efeito, porém, basta observar que se tem, para qualquer valor de θ, $18\operatorname{sen}^2\theta + \cos^2\theta \geq 1$ (visto que $18\operatorname{sen}^2\theta \geq \operatorname{sen}^2\theta$) — e portanto também, para qualquer θ e qualquer r, $18\operatorname{sen}^2\theta + \cos^2\theta + 3r^2\operatorname{sen}^4\theta \geq 1$ — e ainda que, se for por exemplo $r < \frac{1}{16}$, será também $|16r\operatorname{sen}^3\theta| \leq 16r < 1$.

A classificação dos pontos de estacionaridade das funções consideradas nos exemplos precedentes foi efectuada por processos mais ou menos casuísticos, que dificilmente parecerão susceptíveis de aplicação em situações de razoável generalidade: um dos objectivos do procedimento adoptado foi precisamente fazer ressaltar o interesse que, para o esclarecimento de questões desta natureza, podem ter alguns dos resultados subsequentes. Como seria fácil prever atendendo ao que se verificou no caso das funções reais de variável real, todos esses resultados decorrem facilmente do teorema de Taylor.

Teorema 4.18. *Seja D um aberto de $\mathbb{R}^n$, $f : D \to \mathbb{R}$ uma função de classe $\mathcal{C}^2$ e $\boldsymbol{a}$ um ponto de estacionaridade de f; nestas condições:*

a) se $f''(\boldsymbol{a})\boldsymbol{h}^2 > 0$ para qualquer vector não nulo $\boldsymbol{h} \in \mathbb{R}^n$, $\boldsymbol{a}$ é um ponto de mínimo relativo estrito da função f;

b) se $\boldsymbol{a}$ é um ponto de mínimo de f, tem-se $f''(\boldsymbol{a})\boldsymbol{h}^2 \geq 0$ para qualquer $\boldsymbol{h} \in \mathbb{R}^n$;

c) se $f''(\boldsymbol{a})\boldsymbol{h}^2 < 0$ para qualquer vector não nulo $\boldsymbol{h} \in \mathbb{R}^n$, $\boldsymbol{a}$ é um ponto de máximo relativo estrito de f;

d) se $\boldsymbol{a}$ é um ponto de máximo de f, $f''(\boldsymbol{a})\boldsymbol{h}^2 \leq 0$ para qualquer $\boldsymbol{h} \in \mathbb{R}^n$;

e) se existem vectores $\boldsymbol{k}, \boldsymbol{l} \in \mathbb{R}^n$ tais que $f''(\boldsymbol{a})\boldsymbol{k}^2 < 0$ e $f''(\boldsymbol{a})\boldsymbol{l}^2 > 0$, $\boldsymbol{a}$ é ponto de sela.

Antes de iniciar a demonstração recordemos que, de acordo com as notações adoptadas em 4.3 e supondo $\boldsymbol{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$, se tem:

$$f''(\boldsymbol{a})\boldsymbol{h}^2 = \sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) h_i h_j.$$

Assim, $f''(\boldsymbol{a})\boldsymbol{h}^2$ é uma forma quadrática, isto é, um polinómio homogéneo do 2º grau em $h_1, \ldots, h_n$ (em geral chama-se *forma de grau p* a qualquer polinómio homogéneo de grau p; para $p = 1, 2, 3, \ldots$, a forma diz-se *linear, quadrática, cúbica*, etc.). Costuma dizer-se que uma forma é *definida positiva* (resp. *definida negativa*) se assume apenas valores positivos (resp. negativos) sempre que seja $\boldsymbol{h} \neq 0$; *semi-definida positiva*[20] (resp. *semi-definida negativa*) se não assumir qualquer valor negativo (resp. positivo); *indefinida* se for susceptível de assumir valores de sinais contrários.

Nestas condições (e continuando a supor que f é uma função de classe $\mathcal{C}^2$ e $\boldsymbol{a}$ um ponto de estacionaridade de f), o teorema 4.18 poderia reenunciar-se nos termos seguintes:

a) se $f''(\boldsymbol{a})\boldsymbol{h}^2$ é definida positiva, $\boldsymbol{a}$ é um ponto de mínimo relativo estrito;

b) se $\boldsymbol{a}$ é um ponto de mínimo, a forma $f''(\boldsymbol{a})\boldsymbol{h}^2$ é semi-definida positiva;

c) se a forma $f''(\boldsymbol{a})\boldsymbol{h}^2$ é definida negativa, $\boldsymbol{a}$ é um ponto de máximo relativo estrito;

d) se $\boldsymbol{a}$ é um ponto de máximo, a forma $f''(\boldsymbol{a})\boldsymbol{h}^2$ é semi-definida negativa;

e) se a forma $f''(\boldsymbol{a})\boldsymbol{h}^2$ é indefinida, $\boldsymbol{a}$ é um ponto de sela.

Demonstração. a) Provaremos que, sendo a forma $f''(\boldsymbol{a})\boldsymbol{h}^2$ definida positiva, existe uma bola centrada em $\boldsymbol{a}$, $B_\delta(\boldsymbol{a})$, tal que $f(x) > f(\boldsymbol{a})$ para qualquer $x \in B_\delta(\boldsymbol{a}) \setminus \{\boldsymbol{a}\}$. Com efeito, como $f''(\boldsymbol{a})\boldsymbol{h}^2$ é uma função contínua de $\boldsymbol{h}$ e assume valores positivos em todos os pontos do compacto $S = \{\boldsymbol{h} \in \mathbb{R}^n : \|\boldsymbol{h}\| = 1\}$, admitirá neste conjunto um mínimo positivo, $m : f''(\boldsymbol{a})\boldsymbol{h}^2 \geq m$, para qualquer $\boldsymbol{h} \in S$. Por outro lado, como f é de classe $\mathcal{C}^2$, é fácil ver que existe $\delta > 0$ tal que, para qualquer $z \in B_\delta(a)$ e qualquer $\boldsymbol{h} \in S, f''(z)\boldsymbol{h}^2 \geq \frac{1}{2}m$.

Seja então $\boldsymbol{x}$ um ponto qualquer de $B_\delta(\boldsymbol{a})$ distinto de $\boldsymbol{a}$ e ponhamos $t = \|\boldsymbol{x} - \boldsymbol{a}\|, \boldsymbol{h} = \frac{1}{t}(\boldsymbol{x} - \boldsymbol{a})$; ter-se-á evidentemente $t \in]0, \delta[$, $\boldsymbol{h} \in S$ e o teorema de Taylor garante a existência de $\theta \in]0, 1[$ tal que

$$f(\boldsymbol{x}) - f(\boldsymbol{a}) = f(\boldsymbol{a} + t\boldsymbol{h}) - f(\boldsymbol{a}) = f'(\boldsymbol{a})(t\boldsymbol{h}) + \frac{1}{2} f''(\boldsymbol{a} + \theta t\boldsymbol{h})(t\boldsymbol{h})^2.$$

[20] Em alguns textos adopta-se uma definição diferente: só se chamam semi-definidas positivas as formas que, não assumindo qualquer valor negativo, se anulam em algum ponto $\boldsymbol{h} \neq \boldsymbol{0}$ (e de modo análogo para as formas semi-definidas negativas).

Como $f'(\boldsymbol{a})(t\boldsymbol{h}) = 0$ por $\boldsymbol{a}$ ser ponto de estacionaridade de f e $\frac{1}{2}f''(\boldsymbol{a} + \theta t\boldsymbol{h})(t\boldsymbol{h})^2 = \frac{1}{2}t^2 f''(\boldsymbol{a} + \theta t\boldsymbol{h})\boldsymbol{h}^2 \geq \frac{1}{4}mt^2 > 0$, (visto que $\boldsymbol{a} + \theta t\boldsymbol{h} \in B_\delta(\boldsymbol{a})$), pode concluir-se que $f(\boldsymbol{x}) > f(\boldsymbol{a})$, isto é, que $f(\boldsymbol{a})$ é um mínimo relativo estrito da função f.

b) Como f é de classe $\mathcal{C}^2$, se existe $\boldsymbol{k} \in \mathbb{R}^n$ tal que $f''(\boldsymbol{a})\boldsymbol{k}^2 < 0$ existirá também $\epsilon > 0$ tal que $f''(\boldsymbol{a} + t\boldsymbol{k})\boldsymbol{k}^2 < 0$ sempre que seja $|t| < \epsilon$; ter-se-á então também, para qualquer t tal que $|t| < \epsilon$ e algum $\theta \in]0, 1[$:

$$f(\boldsymbol{a} + t\boldsymbol{k}) - f(\boldsymbol{a}) = \frac{1}{2}t^2 f(\boldsymbol{a} + \theta t\boldsymbol{k})\boldsymbol{k}^2 < 0.$$

Existirão portanto pontos arbitrariamente próximos de $\boldsymbol{a}$ nos quais f assume valores menores do que $f(\boldsymbol{a})$ e este não poderá ser um mínimo de f.

As proposições c) e d) decorrem de a) e b), respectivamente, por substituição de f por $-f$; e) é consequência também imediata de b) e d). □

Como primeiro exemplo consideremos a função definida em $\mathbb{R}^3 \setminus \{(0, 0, 0)\}$ pela fórmula $f(x, y, z) = z\log(x^2 + y^2 + z^2)$. O sistema que determina os pontos de estacionaridade:

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{2xz}{x^2 + y^2 + z^2} = 0,\\
\frac{\partial f}{\partial y} &= \frac{2yz}{x^2 + y^2 + z^2} = 0,\\
\frac{\partial f}{\partial z} &= \log(x^2 + y^2 + z^2) + \frac{2z^2}{x^2 + y^2 + z^2} = 0
\end{aligned}$$

tem apenas duas soluções situadas fora do plano $z = 0$ — precisamente $(0, 0, \frac{1}{e})$ e $(0, 0 - \frac{1}{e})$ — e uma infinidade de soluções situadas nesse plano: todos os pontos da circunferência em que este plano é intersectado pela superfície cilíndrica $x^2 + y^2 = 1$.

Sendo $(h, k, l) \in \mathbb{R}^3$ tem-se, como é fácil verificar:

$$\begin{aligned}
f''(0, 0, \tfrac{1}{e})(h, k, l)^2 &= 2e(h^2 + k^2 + l^2),\\
f''(0, 0, -\tfrac{1}{e})(h, k, l)^2 &= -2e(h^2 + k^2 + l^2).
\end{aligned}$$

A primeira das formas precedentes é definida positiva, a segunda é definida negativa; segue-se que $f(0, 0, \frac{1}{e}) = -\frac{2}{e}$ é um mínimo estrito e $f(0, 0, -\frac{1}{e}) = \frac{2}{e}$ um máximo estrito da função f; trata-se, evidentemente, de extremos relativos, visto que a função assume todos os valores reais (basta notar que o mesmo se passa com a sua restrição ao eixo dos zz, privado da origem, $f(0, 0, z) = z\log(z^2)$).

Em qualquer outro ponto de estacionaridade — isto é, em qualquer ponto da forma $(a, b, 0)$ com $a^2 + b^2 = 1$ — ter-se-á, como logo se reconhece

$$f''(a, b, 0)(h, k, l)^2 = 4l(ah + bk).$$

Se for $a \neq 0$, esta forma assumirá valores de sinais contrários nos pontos $(1,0,1)$ e $(1,0,-1)$; se $a = 0$ (e portanto $b \neq 0$), os valores da forma em $(0,1,1)$ e $(0,1,-1)$ serão também de sinais contrários. Pode portanto concluir-se que qualquer dos pontos da circunferência determinada pelas equações $x^2 + y^2 = 1$ e $z = 0$ é um ponto de sela da função f.

Sejam agora $\boldsymbol{c}_1 = (c_{11}, \ldots, c_{1n}), \ldots, \boldsymbol{c}_q = (c_{q1}, \ldots, c_{qn})$ q pontos do espaço $\mathbb{R}^n$ e seja $g : \mathbb{R}^n \to \mathbb{R}$ a função definida pela fórmula:

$$g(\boldsymbol{x}) = g(x_1, \ldots, x_n) = \sum_{i=1}^{q} \|\boldsymbol{x} - \boldsymbol{c}_i\|^2 = \sum_{i=1}^{q} \sum_{j=1}^{n} (x_j - c_{ij})^2.$$

Como $\frac{\partial g}{\partial x_j} = 2qx_j - 2\sum_{i=1}^{q} c_{ij}$, para $j = 1, \ldots, n$, a função tem um único ponto de estacionaridade, o ponto $\boldsymbol{a} = \frac{1}{q}\sum_{i=1}^{q} \boldsymbol{c}_i$. Por outro lado, sendo $\frac{\partial^2 g}{\partial x_j^2} = 2q$ $(j = 1, \ldots, n)$ e $\frac{\partial^2 g}{\partial x_i \partial x_j} = 0$ $(i, j = 1, \ldots, n$, com $i \neq j)$ ter-se-á, para qualquer vector $\boldsymbol{h} = (h_1, \ldots, h_n) \in \mathbb{R}^n$,

$$g''(\boldsymbol{a})\boldsymbol{h}^2 = 2q(h_1^2, + \ldots + h_n^2) = 2q\|\boldsymbol{h}\|^2.$$

Pode portanto concluir-se que $g(\boldsymbol{a}) = \sum_{i=1}^{q} \|\boldsymbol{a} - \boldsymbol{c}_i\|^2$ é um mínimo da função g; e se adoptarmos um processo já utilizado num dos exemplos anteriores, tendo em conta que $\lim_{\|x\| \to \infty} g(\boldsymbol{x}) = +\infty$, será fácil reconhecer que esse mínimo é absoluto e que o contradomínio da função é $[g(\boldsymbol{a}), +\infty[$.

No caso particular das funções de duas variáveis reais $(n = 2)$ é muitas vezes útil o seguinte:

Corolário 4.19. *Seja f uma função de classe $\mathcal{C}^2$ no aberto $D \subset \mathbb{R}^2$ e (a, b) um ponto de estacionaridade de f; sendo*

$$A = \frac{\partial^2 f}{\partial x^2}(a, b), \quad B = \frac{\partial^2 f}{\partial x \partial y}(a, b), \quad C = \frac{\partial^2 f}{\partial y^2}(a, b),$$

tem-se:

a) Se $AC - B^2 > 0$, $f(a, b)$ é um mínimo relativo estrito ou um máximo relativo estrito da função f consoante $A > 0$ ou $A < 0$;

b) Se $AC - B^2 < 0$, (a, b) é um ponto de sela.

Demonstração. Adoptando as notações referidas no enunciado do Corolário, ter-se-á, se for (h, k) um vector qualquer de $\mathbb{R}^2$:

$$f''(a, b)(h, k)^2 = Ah^2 + 2Bhk + Ck^2.$$

Assim, no caso $AC - B^2 > 0$ (o que implica $A \neq 0$) será também:

$$f''(a, b)(h, k)^2 = \frac{1}{A}[(Ah + Bk)^2 + (AC - B^2)k^2] \tag{4.3}$$

e portanto — para qualquer vector $(h,k) \neq (0,0)$ — o sinal da forma $f''(a,b)(h,k)^2$ será o mesmo do de A: (a,b) será ponto de mínimo relativo estrito se $A > 0$ e ponto de máximo relativo estrito se $A < 0$.

Na hipótese $AC - B^2 < 0$ poderá ter-se $A \neq 0$ ou $A = 0$. Se for $A \neq 0$ (caso em que a igualdade (4.3) continuará a ser válida) a forma $f''(a,b)(h,k)^2$ terá o sinal de A se for $k = 0$ e $h \neq 0$ e o sinal contrário ao de A se $h = B$ e $k = -A$; finalmente, se $A = 0$ (ainda com $AC - B^2 < 0$, o que implica $B \neq 0$), a forma em referência reduzir-se-á a $k(2Bh + Ck)$ e assumirá valores de sinais contrários se, fixado $k \neq 0$, atribuirmos a h dois valores, um maior e outro menor do que $-\frac{kC}{2B}$; pode portanto concluir-se que, quando $AC - B^2 < 0$, (a,b) é ponto de sela. □

Voltando à situação considerada no enunciado do teorema 4.18, observemos ainda que, para a classificação da forma quadrática $f''(\boldsymbol{a})\boldsymbol{h}^2$ é muitas vezes útil o resultado seguinte[21]: designando por $H(\boldsymbol{a})$ a *matriz hessiana* da função f no ponto $\boldsymbol{a}$, $H(\boldsymbol{a}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a})\right]_{i,j=1}^{n}$, a forma $f''(\boldsymbol{a})\boldsymbol{h}^2$ é definida positiva (resp. definida negativa) se os valores próprios de $H(\boldsymbol{a})$ são todos positivos (resp. negativos) e indefinida se $H(\boldsymbol{a})$ tiver algum valor próprio positivo e algum valor próprio negativo; não havendo valores próprios de sinais contrários, a forma será semi-definida.

A título de exemplo, consideremos a função:

$$\varphi(x,y,z) = x^4 + y^4 + z^4 - 4xyz.$$

O sistema que determina os pontos críticos é $x^3 = yz$, $y^3 = xz$, $z^3 = xy$. Uma solução óbvia — e a única com alguma coordenada nula — é a origem. Para $x \neq 0$, $y \neq 0$ e $z \neq 0$, as duas primeiras equações conduzem imediatamente à relação $\frac{x^3}{y^3} = \frac{y}{x}$ — e portanto a $x = \pm y$ — e as duas últimas a $y = \pm z$. Segue-se que (além da origem) os pontos de estacionaridade são $(1,1,1)$, $(1,-1,-1)$, $(-1,1,-1)$ e $(-1,-1,1)$. Para qualquer destes quatro pontos a equação característica da matriz hessiana:

$$(12 - \lambda)^3 - 48(12 - \lambda) - 128 = 0$$

tem a raíz dupla 16 e a raíz simples 4. Pode portanto concluir-se que cada um desses pontos é um minimizante e é fácil ver que o valor assumido pela função em qualquer deles, -1, é o seu mínimo absoluto.

Esta mesma ordem de ideias não permitiria classificar o ponto de estacionaridade $(0,0,0)$, porque os valores próprios da matriz hessiana correspondente a esse ponto são todos nulos. Mas basta reparar que, sobre a recta de equações $x = y = z$ a função assume, em pontos arbitrariamente próximos da origem, tanto valores positivos como valores negativos, para se poder concluir que se trata de um ponto de sela.

[21] A demonstração deste resultado, bem como o enunciado e demonstração de outros critérios para a classificação de pontos de estacionaridade, podem ser estudados no texto *Álgebra Linear como Introdução à Matemática Aplicada* de Luís Magalhães, já anteriormente citado.

Voltando ao caso $n = 2$ e às condições expressas no enunciado do Corolário 4.19, se designarmos por λ_1 e λ_2 os valores próprios da matriz hessiana da função f no ponto $\boldsymbol{a}$:

$$H(\boldsymbol{a}) = \begin{bmatrix} A & B \\ B & C \end{bmatrix},$$

ter-se-á $\lambda_1\lambda_2 = AC - B^2$, $\lambda_1 + \lambda_2 = A + C$.

Se for $AC - B^2 > 0$ (o que implica $AC > 0$), λ_1 e λ_2 são do mesmo sinal, o sinal de A (ou de C)[22]: a forma $Ah^2 + 2Bhk + Ck^2$ será definida positiva ou definida negativa (e (a, b) será ponto de mínimo ou ponto de máximo) consoante $A > 0$ ou $A < 0$; se for $AC - B^2 < 0, \lambda_1$ e λ_2 terão sinais contrários, a forma será indefinida e (a, b) ponto de sela (evidentemente, estes resultados são apenas uma confirmação do corolário em referência).

Para observar algumas possibilidades de extensão dos resultados precedentes, consideremos novamente um aberto D de $\mathbb{R}^n$, uma função $f : D \to \mathbb{R}$ e suponhamos agora que f é da classe $\mathcal{C}^p$, com $p \geq 2$ e que, em certo ponto $\boldsymbol{a} \in D$ e para qualquer inteiro k tal que $1 \leq k < p$, todas as formas $f^{(k)}(\boldsymbol{a})\boldsymbol{h}^k$ são identicamente nulas. Nestas condições, a fórmula de Taylor:

$$f(\boldsymbol{a} + t\boldsymbol{h}) = f(\boldsymbol{a}) + \frac{1}{p!}t^p f^{(p)}(\boldsymbol{a} + \theta t\boldsymbol{h})\boldsymbol{h}^p$$

permite concluir (de forma inteiramente análoga à que utilizámos ao estabelecer o teorema 4.18) que, consoante a forma $f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p$ seja definida positiva, definida negativa ou indefinida, assim o ponto $\boldsymbol{a}$ será um minimizante estrito, um maximizante estrito ou um ponto de sela da função f; em particular, dado que uma forma de grau ímpar é sempre indefinida (visto que os seus valores em pontos simétricos são simétricos), se p for ímpar (com $f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p$ não identicamente nula) $\boldsymbol{a}$ será um ponto de sela[23]. Também se confirmará agora sem dificuldade que, se $\boldsymbol{a}$ for um ponto de mínimo ou um ponto de máximo, a forma será semidefinida (positiva no primeiro caso, negativa no segundo).

Consideremos agora a hipótese de a forma $f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p$ ser semidefinida (mas não definida nem identicamente nula); neste caso, é usual designar por *direcções singulares* da forma considerada as direcções dos vectores (não nulos) nas quais ela se anula. Para fixar as ideias, suponhamos — até menção expressa em contrário — que a forma é semidefinida positiva; existirão então vectores de $\mathbb{R}^n$ nos quais a forma se anulará (o vector nulo e os que tenham uma direcção singular) e existirão também vectores (todos os restantes) nos quais a forma assumirá um valor positivo.

[22]Tenha-se em conta que os valores próprios da matriz hessiana de uma função de classe $\mathcal{C}^2$ são sempre reais, dado que a matriz é simétrica.

[23]Pode utilizar-se este resultado para verificar que, no caso há pouco mencionado da função $\varphi(x, y, z) = x^4 + y^4 + z^4 - 4xyz$, a origem é um ponto de sela: com efeito tem-se, para qualquer vector $\boldsymbol{h} = (h_1, h_2, h_3) \in \mathbb{R}^3$, $\varphi'(0, 0, 0)\boldsymbol{h} = \varphi''(0, 0, 0)\boldsymbol{h}^2 = 0$ e $\varphi'''(0, 0, 0)\boldsymbol{h}^3 = -4h_1h_2h_3$.

Se $\boldsymbol{h}$ for um destes últimos vectores, mostra a fórmula de Taylor (4.18) que, sempre que o valor absoluto de t for suficientemente pequeno (mas não nulo), se terá $f(\boldsymbol{a} + t\boldsymbol{h}) > f(\boldsymbol{a})$; assim, em tal caso, a função definida (nalguma vizinhança do ponto $t = 0$) pela fórmula $\varphi_{\boldsymbol{h}}(t) = f(\boldsymbol{a} + t\boldsymbol{h})$ terá um mínimo estrito no ponto 0 (ou, o que é o mesmo, terá um mínimo estrito no ponto $\boldsymbol{a}$ a restrição da função f à intersecção do seu domínio com a recta do espaço $\mathbb{R}^n$ que contém esse ponto e tem a direcção do vector $\boldsymbol{h}$). Porém, no caso de $\boldsymbol{h}$ ser um vector com direcção singular, já o ponto 0 poderá ser um minimizante, um ponto de sela ou um maximizante da função $\varphi_{\boldsymbol{h}}$;[24] e quando se tratar de um ponto de sela ou de um ponto de máximo estrito, é claro que a função f não poderá ter um mínimo no ponto $\boldsymbol{a}$ (nem um máximo, pelo que vimos no período precedente): o ponto $\boldsymbol{a}$ será portanto um ponto de sela.

Um caso particular em que esta conclusão será legítima — como decorre de resultados bem conhecidos sobre a existência de extremos para funções reais de uma variável real — é o que se verifica se, para algum vector $\boldsymbol{h}$ com direcção singular, a função $\varphi_{\boldsymbol{h}}(t)$ admitir derivadas de ordem superior a p no ponto 0 e se a primeira de tais derivadas que não seja nula (se alguma existir) for de ordem ímpar ou, se de ordem par, for negativa (visto que $\varphi_{\boldsymbol{h}}$ terá então um ponto de sela ou um máximo estrito no ponto 0). Exemplos muito simples de situações deste tipo verificam-se com as funções β e γ mencionadas na nota 24.

Poderia surgir agora naturalmente a questão seguinte: continuando a supor que a forma $f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p$ é semidefinida positiva, o que poderá concluir-se se, para qualquer vector $\boldsymbol{h}$ com direcção singular (tal como para todos os vectores com outras direcções), a correspondente função $\varphi_{\boldsymbol{h}}$ tiver um mínimo estrito no ponto 0? Ou, equivalentemente, se a restrição de f a uma recta arbitrária de $\mathbb{R}^n$ que contenha o ponto $\boldsymbol{a}$ (intersectada com D) tiver neste ponto um mínimo estrito? Poderá em tal caso afirmar-se que o ponto $\boldsymbol{a}$ é um minimizante da própria função f?

Será natural desconfiar da correcção desta conjectura se se tiver presente que (como oportunamente observámos) para funções de mais de uma variável, o comportamento da função ao longo de todas as rectas que concorrem num ponto não dá informação suficiente sobre algumas características importantes da função, tais como a continuidade ou a existência de limite no ponto considerado.

A conjectura referida é de facto incorrecta. Um exemplo simples (ver figura 4.9) que o evidencia é o da função $f(x, y) = 2x^4 - 3x^2y + y^2$, no seu único ponto de estacionaridade, a origem. A forma quadrática $f''(0, 0)\boldsymbol{h}^2 = 2h_2^2$ (ainda com $\boldsymbol{h} = (h_1, h_2) \in \mathbb{R}^2$) é semidefinida positiva e tem por direcção singular a do eixo das abcissas; a restrição da função a este eixo, como aliás a qualquer outra recta do plano que contenha a

[24] As funções $\alpha(x, y) = x^2 + y^4$, $\beta(x, y) = x^2 + y^3$ e $\gamma(x, y) = x^2 - y^4$, para as quais se tem, com $\boldsymbol{h} = (h_1, h_2) \in \mathbb{R}^2$, $\alpha''(0, 0)\boldsymbol{h}^2 = \beta''(0, 0)\boldsymbol{h}^2 = \gamma''(0, 0)\boldsymbol{h}^2 = 2h_1^2$ (sendo portanto singular, para qualquer das formas consideradas, a direcção do eixo das ordenadas) e ainda $\alpha(0, t) = t^4$, $\beta(0, t) = t^3$, $\gamma(0, t) = -t^4$, exemplificam os três casos.

origem, tem um mínimo estrito neste ponto. No entanto o ponto $(0,0)$ não é um ponto de mínimo, mas sim um ponto de sela da função f.

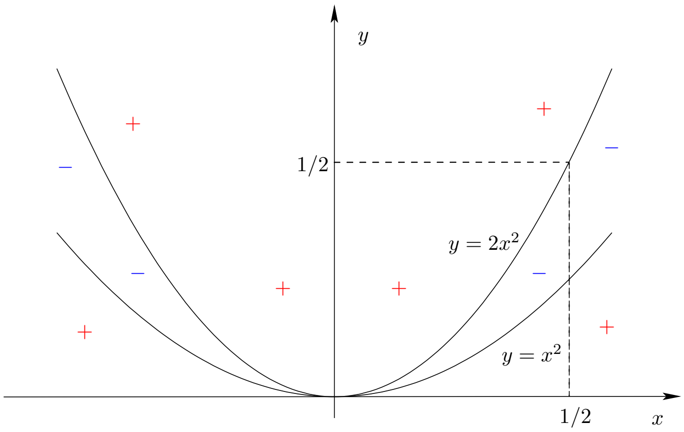

Figura 4.9

Para o reconhecer basta ter em conta a igualdade

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

da qual resulta que se tem $f(x,y) < 0$ se $x^2 < y < 2x^2$ e $f(x,y) > 0$ se $y < x^2$ ou $y > 2x^2$, o que torna evidente que em qualquer vizinhança da origem há pontos onde f assume valores maiores e pontos em que assume valores menores do que $f(0,0) = 0$.

Voltando ao caso geral anteriormente considerado e mantendo todas as restantes hipóteses sobre a função f e o ponto $\boldsymbol{a}$, passemos a supor agora que a forma $f^{(p)}(\boldsymbol{a})\boldsymbol{h}^p$ é semidefinida negativa (sem ser definida nem identicamente nula); é claro que poderemos ainda concluir que $\boldsymbol{a}$ é um ponto de sela da função f se, para algum vector $\boldsymbol{h}$ (necessariamente de direcção singular) o ponto 0 for um ponto de sela ou um minimizante estrito da função $\varphi_{\boldsymbol{h}}(t) = f(\boldsymbol{a} + t\boldsymbol{h})$ (e para que alguma destas circunstâncias se verifique será suficiente que, na hipótese de $\varphi_{\boldsymbol{h}}$ admitir derivadas de ordem superior a p no ponto 0, a primeira destas derivadas que se não anule seja de ordem ímpar ou, se de ordem par, tenha valor positivo). Evidentemente, se para qualquer vector[25] $\boldsymbol{h}$ com direcção singular a função $\varphi_{\boldsymbol{h}}$ for máxima no ponto 0 não será possível, por esta via, tirar qualquer conclusão.

A título de exemplo, consideremos a função

$$f(x,y) = ax^6 - 4x^4 + 4x^2y - y^2$$

(onde a é um parâmetro real), para a qual a origem é o único ponto de estacionaridade. A forma $f''(0,0)(h_1,h_2)^2 = -2h_2^2$ é semidefinida negativa

[25] Como é óbvio, se $\boldsymbol{h}$ e $\boldsymbol{k}$ são dois vectores com a mesma direcção, e se $\varphi_{\boldsymbol{h}}(t)$ tiver um minimizante, um maximizante ou um ponto de sela no ponto 0, o mesmo se passará com $\varphi_{\boldsymbol{k}}(t)$; assim, neste tipo de questões bastará - para cada direcção singular - considerar apenas um vector com essa direcção.

(logo, se houver extremo será um máximo) e a única direcção singular é a do eixo das abcissas. Mas $f(t,0) = -t^4(4 - at^2)$ tem um máximo no ponto $t = 0$, o que não permite tirar qualquer conclusão.

No entanto, se notarmos que $f(x,y) = ax^6 - (y - 2x^2)^2$, tornar-se-á evidente que, se $a > 0$, a restrição da função à parábola $y = 2x^2$ tem um mínimo estrito na origem, donde logo decorre que o ponto $(0,0)$ é, nessa hipótese, um ponto de sela da função f; por observação directa dos valores assumidos pela função é também fácil reconhecer que, se for $a \leq 0$, $f(0,0) = 0$ será o seu máximo absoluto (estrito, excepto se $a = 0$).

Índice Remissivo