

# **Análise Real**

## **Volume 2**

**Elon Lages Lima**

**Rio de Janeiro**

**9 de março de 2004**

# Sumário

<b>1</b>	<b>Topologia do Espaço Euclidiano</b>	<b>1</b>
1	O espaço euclidiano $n$ -dimensional . . . . .	1
2	Bolas e conjuntos limitados . . . . .	5
3	Conjuntos abertos . . . . .	7
4	Seqüências em $\mathbb{R}^n$ . . . . .	9
5	Conjuntos fechados . . . . .	12
6	Conjuntos compactos . . . . .	14
7	Aplicações contínuas . . . . .	18
8	Continuidade uniforme . . . . .	21
9	Homeomorfismos . . . . .	23
10	Conjuntos conexos . . . . .	24
11	Limites . . . . .	30
<b>2</b>	<b>Caminhos em <math>\mathbb{R}^n</math></b>	<b>33</b>
1	Caminhos diferenciáveis . . . . .	33
2	Cálculo diferencial de caminhos . . . . .	35
3	A integral de um caminho . . . . .	37
4	Caminhos retificáveis . . . . .	39
<b>3</b>	<b>Funções Reais de <math>n</math> Variáveis</b>	<b>44</b>
1	Derivadas parciais . . . . .	44
2	Funções de classe $C^1$ . . . . .	46
3	O Teorema de Schwarz . . . . .	52
4	A fórmula de Taylor . . . . .	56
5	Pontos críticos . . . . .	58
6	Funções convexas . . . . .	62
	<b>Apêndice: Continuidade das funções convexas</b> . . . . .	<b>66</b>
<b>4</b>	<b>Funções Implícitas</b>	<b>69</b>
1	Uma função implícita . . . . .	69
2	Hiperfícies . . . . .	72
3	Multiplicador de Lagrange . . . . .	76
<b>5</b>	<b>Aplicações Diferenciáveis</b>	<b>81</b>

1	A derivada como transformação linear . . . . .	81
2	Exemplos de derivadas . . . . .	83
3	Cálculo diferencial de aplicações . . . . .	86
<b>6</b>	<b>Aplicações Inversas e Implícitas</b>	<b>92</b>
1	O Teorema da Aplicação Inversa . . . . .	92
2	Várias Funções Implícitas . . . . .	97
<b>7</b>	<b>Superfícies Diferenciáveis</b>	<b>104</b>
1	Parametrizações . . . . .	104
2	Superfícies diferenciáveis . . . . .	106
3	O espaço vetorial tangente . . . . .	109
4	Superfícies orientáveis . . . . .	111
5	Multiplicadores de Lagrange . . . . .	117
<b>8</b>	<b>Integrais Múltiplas</b>	<b>120</b>
1	A definição de integral . . . . .	120
2	Conjunto de medida nula . . . . .	124
3	Cálculo com integrais . . . . .	129
4	Conjuntos $J$ -mensuráveis . . . . .	132
5	A integral como limite de somas de Riemann . . . . .	134
<b>9</b>	<b>Mudança de Variáveis</b>	<b>139</b>
1	O caso unidimensional . . . . .	140
2	Difeomorfismos primitivos . . . . .	142
3	Todo difeomorfismo $C^1$ é localmente admissível . . . . .	143
4	Conclusão: todo difeomorfismo de classe $C^1$ é admissível . . . . .	144

## Capítulo 1

# Topologia do Espaço Euclidiano

### 1 O espaço euclidiano $n$ -dimensional

Seja  $n$  um número natural. O *espaço euclidiano  $n$ -dimensional*  $\mathbb{R}^n$  é o produto cartesiano de  $n$  fatores iguais a  $\mathbb{R}$ :  $\mathbb{R}^n = \mathbb{R} \times \mathbb{R} \times \dots \times \mathbb{R}$ . Seus elementos, portanto, são as seqüências (ou listas) de  $n$  termos reais  $x = (x_1, \dots, x_n)$ . Para cada  $i = 1, \dots, n$ , o termo  $x_i$  chama-se a  *$i$ -ésima coordenada* de  $x$ . Se  $x = (x_1, \dots, x_n)$  e  $y = (y_1, \dots, y_n)$ , tem-se  $x = y$  se, e somente se,  $x_1 = y_1, \dots, x_n = y_n$ . Assim, toda igualdade entre dois elementos de  $\mathbb{R}^n$  equivale a  $n$  igualdades entre números reais.  $\mathbb{R}^1 = \mathbb{R}$  é o conjunto dos números reais,  $\mathbb{R}^2$  é o modelo numérico do plano e  $\mathbb{R}^3$  é o modelo do espaço euclidiano tridimensional. Por simplicidade, adotaremos o hábito de escrever  $z = (x, y)$  em vez de  $x = (x_1, x_2)$  e  $w = (x, y, z)$  em vez de  $x = (x_1, x_2, x_3)$ .

Os elementos de  $\mathbb{R}^n$  às vezes são chamados *pontos* e às vezes *vetores*. Este segundo nome se aplica principalmente quando se considerarem entre eles as operações que definiremos agora.

A *adição* faz corresponder a cada par de elementos  $x = (x_1, \dots, x_n)$  e  $y = (y_1, \dots, y_n)$  a *soma*

$$x + y = (x_1 + y_1, \dots, x_n + y_n).$$

e a *multiplicação* do número real  $\alpha$  pelo elemento  $x = (x_1, \dots, x_n)$  tem como resultado o *produto*

$$\alpha \cdot x = (\alpha x_1, \dots, \alpha x_n).$$

O vetor  $0 = (0, 0, \dots, 0)$ , cujas coordenadas são todas nulas, chama-se a *origem* de  $\mathbb{R}^n$ . Para todo  $x = (x_1, \dots, x_n)$ , o vetor  $-x = (-x_1, \dots, -x_n)$  chama-se o *oposto*, ou *simétrico* de  $x$ . Dados quaisquer  $x, y, z \in \mathbb{R}^n$  e  $\alpha, \beta \in \mathbb{R}$  valem as

igualdades

$$\begin{aligned}x + y &= y + x, & x + 0 &= x, & -x + x &= 0, \\x + (y + z) &= (x + y) + z, & \alpha(\beta x) &= (\alpha\beta)x, \\(\alpha + \beta)x &= \alpha x + \beta x, & \alpha(x + y) &= \alpha x + \alpha y.\end{aligned}$$

A segunda e a terceira delas dizem que 0 é o *elemento neutro* da adição e  $-x$  é o *inverso aditivo* de  $x$ .

Os vetores  $e_1 = (1, 0, \dots, 0)$ ,  $e_2 = (0, 1, 0, \dots, 0)$ ,  $\dots$ ,  $e_n = (0, \dots, 1)$ , que têm uma única coordenada não-nula, igual a 1, constituem a *base canônica* de  $\mathbb{R}^n$ . A igualdade  $x = (x_1, \dots, x_n)$  significa que  $x = x_1 \cdot e_1 + \dots + x_n \cdot e_n$ .

Existe ainda uma operação que associa a cada par de vetores  $x = (x_1, \dots, x_n)$ ,  $y = (y_1, \dots, y_n)$  o número real

$$\langle x, y \rangle = x_1 y_1 + \dots + x_n y_n,$$

chamado o *produto interno* de  $x$  por  $y$ .

Para  $x, y, z \in \mathbb{R}^n$  e  $\alpha \in \mathbb{R}$  quaisquer, tem-se

$$\begin{aligned}\langle x, y \rangle &= \langle y, x \rangle, & \langle x, y + z \rangle &= \langle x, y \rangle + \langle x, z \rangle, \\ \langle \alpha x, y \rangle &= \alpha \cdot \langle x, y \rangle, & \langle x, x \rangle &> 0 \quad \text{se } x \neq 0.\end{aligned}$$

Segue-se que  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ ,  $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$  e  $\langle x, 0 \rangle = 0$ .

Diz-se que os vetores  $x, y \in \mathbb{R}^n$  são *ortogonais*, e escreve-se  $x \perp y$ , quando  $\langle x, y \rangle = 0$ . Por exemplo,  $\langle e_i, e_j \rangle = 0$  se  $i \neq j$ .

Um exemplo menos trivial de ortogonalidade é o seguinte

**(1.1)** *Seja  $x \in \mathbb{R}^n$  não-nulo. Para todo  $y \in \mathbb{R}^n$ , o vetor  $z = y - \frac{\langle x, y \rangle}{\langle x, x \rangle} \cdot x$  é ortogonal a  $x$ .*

*Demonstração.*  $\langle x, z \rangle = \langle x, y \rangle - \frac{\langle x, y \rangle}{\langle x, x \rangle} \cdot \langle x, x \rangle = 0.$  □

Escrevendo  $y = \frac{\langle x, y \rangle}{\langle x, x \rangle} \cdot x + z$ , vemos assim que, uma vez dado um vetor não-nulo  $x \in \mathbb{R}^n$ , todo vetor  $y \in \mathbb{R}^n$  se escreve como soma de um múltiplo de  $x$  com um vetor ortogonal a  $x$ . Esta decomposição é única pois se  $y = \alpha \cdot x + z$  com  $z \perp x$ , tomando-se o produto interno de ambos os membros por  $x$  obtemos  $\langle x, y \rangle = \alpha \cdot \langle x, x \rangle$ , logo  $\alpha = \langle x, y \rangle / \langle x, x \rangle$ . O vetor  $\alpha x = (\langle x, y \rangle / \langle x, x \rangle) x$  chama-se a *projeção ortogonal* de  $y$  sobre (a reta que contém)  $x$ .

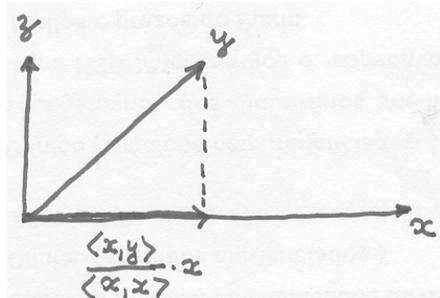


Figura 1.

O número não-negativo  $|x| = \sqrt{\langle x, x \rangle}$  chama-se a *norma* (ou o *comprimento*) do vetor  $x$ . Se  $x = (x_1, \dots, x_n)$  então

$$|x| = \sqrt{x_1^2 + \dots + x_n^2}.$$

Por definição, tem-se  $\langle x, x \rangle = |x|^2$ . Quando  $|x| = 1$ , diz-se que  $x$  é um vetor *unitário*. Para todo  $x \neq 0$ , o vetor  $u = x/|x|$  é unitário.

**(1.2) (Teorema de Pitágoras).** Se  $x \perp y$  então  $|x + y|^2 = |x|^2 + |y|^2$ .

*Demonstração.*  $|x + y|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + 2\langle x, y \rangle + \langle y, y \rangle = \langle x, x \rangle + \langle y, y \rangle = |x|^2 + |y|^2$ .  $\square$

**(1.3) (Desigualdade de Schwarz).** Para quaisquer  $x, y \in \mathbb{R}^n$ , tem-se  $|\langle x, y \rangle| \leq |x| \cdot |y|$ , valendo a igualdade se, e somente se, um dos vetores  $x, y$  é múltiplo do outro.

*Demonstração.* Isto é óbvio se  $x = 0$ . Supondo  $x \neq 0$ , podemos escrever  $y = \alpha x + z$  com  $z \perp x$  e  $\alpha = \langle x, y \rangle / |x|^2$ . Por Pitágoras,  $|y|^2 = \alpha^2 |x|^2 + |z|^2$ , logo  $|y|^2 \geq \alpha^2 |x|^2$ , valendo a igualdade se, e somente se,  $y = \alpha \cdot x$ . Entrando com o valor de  $\alpha$ , vem  $|y|^2 \geq \langle x, y \rangle^2 / |x|^2$ , ou seja,  $\langle x, y \rangle^2 \leq |x|^2 \cdot |y|^2$ , o que nos dá  $|\langle x, y \rangle| \leq |x| \cdot |y|$ , valendo a igualdade se, e somente se,  $y = \alpha \cdot x$ .  $\square$

A norma goza das seguintes propriedades:

1.  $|x| \geq 0$ , valendo  $|x| = 0$  somente quando  $x = 0$ ;
2.  $|\alpha \cdot x| = |\alpha| |x|$ ;
3.  $|x + y| \leq |x| + |y|$ .

A última desigualdade, referindo-se a números não-negativos, equivale a

$$|x + y|^2 \leq (|x| + |y|)^2.$$

Ora,

$$\begin{aligned} |x + y|^2 &= \langle x + y, x + y \rangle = |x|^2 + 2\langle x, y \rangle + |y|^2 \\ &\leq |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2 \end{aligned}$$

pois, em virtude da desigualdade de Schwarz,  $\langle x, y \rangle \leq |x||y|$ .

Mais geralmente, qualquer função  $\mathbb{R}^n \rightarrow \mathbb{R}$ , que associe a cada vetor  $x \in \mathbb{R}^n$  um número  $|x|$  com as três propriedades acima, chama-se uma *norma*. A norma

$$|x| = \sqrt{x_1^2 + \cdots + x_n^2},$$

chama-se *norma euclidiana*.

Há duas outras normas que poderemos utilizar em  $\mathbb{R}^n$  quando houver conveniência. Elas são

1.  $|x|_M = \max \{ |x_1|, \dots, |x_n| \}$  (norma do máximo),
2.  $|x|_S = |x_1| + \cdots + |x_n|$  (norma da soma).

As condições que definem uma norma são fáceis de verificar para estas duas. Também é simples mostrar que, para todo  $x \in \mathbb{R}^n$ , vale

$$|x|_M \leq |x| \leq |x|_S \leq n \cdot |x|_M,$$

onde  $|x|$  é a norma euclidiana.

Quando, num determinado contexto, estivermos usando apenas uma das normas  $|x|_M$  ou  $|x|_S$ , podemos indicá-la com a notação  $|x|$ , por simplicidade.

Para toda norma, vale a desigualdade

$$||x| - |y|| \leq |x - y|.$$

Com efeito, de  $x = (x - y) + y$  resulta que  $|x| \leq |x - y| + |y|$ , logo  $|x| - |y| \leq |x - y|$ .

Trocando os papéis de  $x$  e  $y$ , obtemos  $|y| - |x| \leq |y - x|$ . Mas  $|y - x| = |x - y|$ , logo  $|y| - |x| \leq |x - y|$ . Conclusão:  $||x| - |y|| \leq |x - y|$ .

Uma norma em  $\mathbb{R}^n$  dá origem à noção de *distância*  $d(x, y)$  entre dois pontos  $x, y \in \mathbb{R}^n$ . Para  $x = (x_1, \dots, x_n)$  e  $y = (y_1, \dots, y_n)$ , pomos

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

As três condições que definem uma norma implicam que  $d(x, y)$  tem as propriedades características de uma distância, a saber:

1.  $d(x, y) \geq 0$ , com  $d(x, y) = 0$  se, e somente se,  $x = y$ ;
2.  $d(x, y) = d(y, x)$ ;
3.  $d(x, z) \leq d(x, y) + d(y, z)$  (desigualdade triangular).

Observe que a igualdade  $|\alpha \cdot x| = |\alpha| |x|$  com  $\alpha = -1$  dá  $|-x| = |x|$ , logo  $|x - y| = |y - x|$ . Além disso,  $|x - z| = |x - y + y - z| \leq |x - y| + |y - z|$ , portanto  $d(x, z) \leq d(x, y) + d(y, z)$ .

## 2 Bolas e conjuntos limitados

Dados o ponto  $a \in \mathbb{R}^n$  e o número real  $r > 0$ , a *bola aberta* de centro  $a$  e raio  $r$  é o conjunto  $B(a; r)$  dos pontos  $x \in \mathbb{R}^n$  cuja distância ao ponto  $a$  é menor que  $r$ . Em símbolos:

$$B(a; r) = \{x \in \mathbb{R}^n; |x - a| < r\}.$$

Analogamente, a *bola fechada* de centro  $a$  e raio  $r$  é o conjunto  $B[a; r]$  assim definido:

$$B[a; r] = \{x \in \mathbb{R}^n; |x - a| \leq r\}.$$

Por sua vez, a *esfera* de centro  $a$  e raio  $r$  é o conjunto

$$S[a; r] = \{x \in \mathbb{R}^n; |x - a| = r\}.$$

Evidentemente,  $B[a; r] = B(a; r) \cup S[a; r]$ .

A bola fechada  $B[a; r] \subset \mathbb{R}^n$  também é chamada o *disco*  $n$ -dimensional de centro  $a$  e raio  $r$ . Em particular, o disco  $B[0; 1]$  de centro 0 e raio 1 é chamado o *disco unitário* de  $\mathbb{R}^n$ .

Uma notação especial é reservada para a *esfera unitária* de dimensão  $n - 1$ :

$$S^{n-1} = \{x \in \mathbb{R}^n; |x| = 1\}.$$

Assim,  $S^{n-1}$  é a esfera de centro na origem 0 e raio 1. Quando  $n = 2$ ,  $S^1$  é a circunferência de centro 0 e raio 1.

Acima estamos (pelo menos tacitamente) admitindo que a norma adotada em  $\mathbb{R}^n$  é a euclidiana, já que não foi feita menção em contrário. Convém, entretanto, observar que a forma geométrica das bolas e esferas em  $\mathbb{R}^n$  depende da norma que se considera. Por exemplo, se tomarmos em  $\mathbb{R}^2$  a norma do máximo, a “esfera unitária” é o bordo do quadrado de centro 0 e lados de comprimento 2, paralelos aos eixos. Ainda em  $\mathbb{R}^2$ , com a norma da soma, o “disco unitário” é o quadrado cujos vértices são os pontos  $(1,0)$ ,  $(0,1)$ ,  $(-1, 0)$  e  $(0, -1)$ .

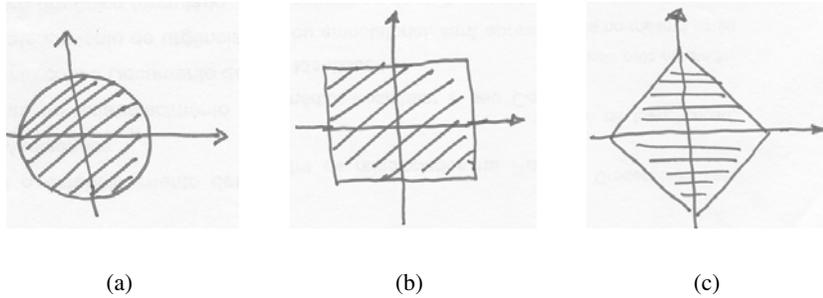


Figura 2. O conjunto dos pontos  $z \in \mathbb{R}^n$  tais que  $|z| \leq 1$ , conforme a norma seja (a) a euclidiana; (b) do máximo, ou (c) da soma.

**Observação.** Indiquemos com as notações  $B$ ,  $B_M$  e  $B_S$  respectivamente as bolas de centro  $a$  e raio  $r$  em  $\mathbb{R}^n$ , relativamente às normas euclidiana, do máximo e da soma. Seja ainda  $B'_M$  a bola de centro  $a$  e raio  $r/n$  na norma do máximo. As desigualdades  $|x|_M \leq |x| \leq |x|_S \leq n|x|_M$  implicam que  $B'_M \subset B_S \subset B \subset B_M$ .

Diz-se que o conjunto  $X \subset \mathbb{R}^n$  é *limitado* quando está contido em alguma bola  $B[a; r]$ . Como  $B[a; r] \subset B[0; k]$ , onde  $k = r + |a|$  (conforme veremos agora), dizer que  $X$  é limitado equivale a dizer que existe  $k > 0$  tal que  $|x| \leq k$  para todo  $x \in X$ .

Para mostrar que  $B[a; r] \subset B[0; r + |a|]$ , note que  $|x - a| \leq r \Rightarrow |x - a + a| \leq |x - a| + |a| \leq r + |a|$ .

Assim,  $x \in B[a; r] \Rightarrow x \in B[0; r + |a|]$ .

Uma aplicação  $f : X \rightarrow \mathbb{R}^n$  diz-se *limitada* quando sua imagem  $f(X) \subset \mathbb{R}^n$  é um conjunto limitado, isto é, quando existe  $c > 0$  tal que  $|f(x)| \leq c$  para todo  $x \in X$ .

Dados  $a \neq b$  em  $\mathbb{R}^n$ , a *reta* que une esses dois pontos é o conjunto  $ab = \{(1 - t)a + tb; t \in \mathbb{R}\}$ . Por sua vez, o *segmento de reta* de extremos  $a, b$  é o conjunto  $[a, b] = \{(1 - t)a + tb; 0 \leq t \leq 1\}$ .

Um conjunto  $X \subset \mathbb{R}^n$  chama-se *convexo* quando o segmento de reta que une dois quaisquer de seus pontos está inteiramente contido em  $X$ . Noutros termos, dizer que  $X$  é convexo equivale a afirmar que

$$a, b \in X, \quad 0 \leq t \leq 1 \quad \Rightarrow \quad (1 - t)a + tb \in X.$$

**Exemplo 1.** Toda bola (aberta ou fechada) é um conjunto convexo. Para fixar as idéias, consideremos a bola fechada  $B = B[x_0; r]$ . Dadas  $a, b \in B$ , temos  $|a - x_0| \leq r$  e  $|b - x_0| \leq r$ . Então, para qualquer  $t \in [0, 1]$  vale  $x_0 = (1 - t)x_0 +$

$tx_0$ , logo

$$\begin{aligned} |(1-t)a + tb - x_0| &= |(1-t)a + tb - (1-t)x_0 - tx_0| \\ &= |(1-t)(a - x_0) + t(b - x_0)| \\ &\leq (1-t)|a - x_0| + t|b - x_0| \\ &\leq (1-t)r + tr = r, \end{aligned}$$

◁

**Exemplo 2.** Seja  $X = \{(x, y) \in \mathbb{R}^2; y \leq x^2\}$ . O conjunto  $X \subset \mathbb{R}^2$  não é convexo. Com efeito os pontos  $a = (-1, 1)$  e  $b = (1, 1)$  pertencem a  $X$  mas  $\frac{1}{2}a + \frac{1}{2}b = (0, 1)$  não pertence a  $X$ . ◁

### 3 Conjuntos abertos

Seja  $a \in X \subset \mathbb{R}^n$ . Diz-se que o ponto  $a$  é *interior* ao conjunto  $X$  quando, para algum  $r > 0$ , tem-se  $B(a; r) \subset X$ . Isto significa que todos os pontos suficientemente próximos de  $a$  também pertencem a  $X$ . O conjunto  $\text{int.}X$  dos pontos interiores a  $X$  chama-se o *interior* do conjunto  $X$ . Evidentemente,  $\text{int.}X \subset X$ . Quando  $a \in \text{int.}X$ , diz-se que  $X$  é uma *vizinhança* de  $a$ .

**Exemplo 3.** Seja  $X = \{(x, y) \in \mathbb{R}^2; y \geq 0\}$  o semi-plano superior fechado. Se  $p = (a, b)$  com  $b > 0$ , então  $p \in \text{int.}X$ . Com efeito, afirmamos que  $B(p; b) \subset X$ . Isto é claro geometricamente.

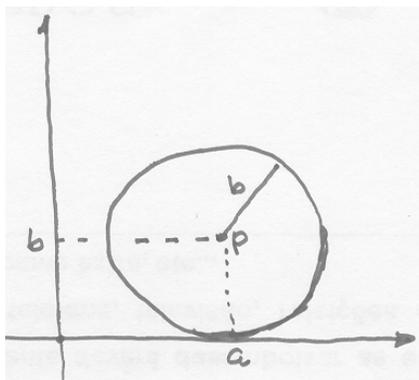


Figura 3.

Em termos mais precisos, argumentamos assim:

$$\begin{aligned}(x, y) \in B &\Rightarrow \sqrt{(x-a)^2 + (y-b)^2} < b \Rightarrow (y-b)^2 < b^2 \\ &\Rightarrow y^2 - 2by + b^2 < b^2 \Rightarrow y^2 < 2by \Rightarrow y > 0 \text{ (pois } b > 0),\end{aligned}$$

e portanto  $(x, y) \in X$ . ◁

**Exemplo 4.** Com a notação do Exemplo 3, os pontos da forma  $q = (a, 0)$ , pertencem a  $X$  porém não são interiores a  $X$ . Com efeito, nenhuma bola  $B(q; r)$  de centro  $q$  pode estar contida em  $X$  pois o ponto  $(a, -r/2)$  pertence a  $B(q; r)$  mas não a  $X$ . Segue-se então que  $\text{int.}X = \{(x, y) \in \mathbb{R}^2; y > 0\}$ .

Um conjunto  $A \subset \mathbb{R}^n$  chama-se *aberto* quando todos os seus pontos são interiores, isto é, quando  $A = \text{int.}A$ . ◁

**Exemplo 5.** Toda bola aberta  $B = B(a; r)$  é um conjunto aberto. Com efeito, seja  $x \in B$ . Então  $|x - a| < r$ , logo  $s = r - |x - a| > 0$ . Afirmamos que,  $B(x; s) \subset B$ . Com efeito,  $y \in B(x; s) \Rightarrow |y - x| < r - |x - a|$ . Logo

$$y \in B(x; s) \Rightarrow |y - a| \leq |y - x| + |x - a| < r - |x - a| + |x - a| = r.$$

Daí concluímos que  $y \in B(a; r)$ . ◁

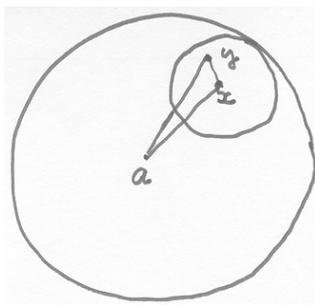


Figura 4.

A *fronteira* de um conjunto  $X \subset \mathbb{R}^n$  é o conjunto  $\text{fr.}X$  formado pelos pontos de  $X$  que não são interiores a  $X$ , juntamente com os pontos de  $\mathbb{R}^n - X$  que não são interiores a  $\mathbb{R}^n - X$ . De forma mais simples: tem-se  $x \in \text{fr.}X$  quando toda bola de centro  $x$  contém pontos de  $X$  e pontos de  $\mathbb{R}^n - X$ .

**Exemplo 6.** Seja  $X = \{(x, y) \in \mathbb{R}^2; y \geq 0\}$ , como no Exemplo 3. De forma análoga ao argumento usado no Exemplo 3, mostra-se que todo ponto de  $\mathbb{R}^2 - X =$

$\{(x, y) \in \mathbb{R}^2; y < 0\}$  é um ponto interior (ou seja, que  $\mathbb{R}^2 - X$  é um conjunto aberto). Logo, nenhum ponto de  $\mathbb{R}^2 - X$  pode estar na fronteira de  $X$ . Segue-se então do Exemplo 4 que  $\text{fr.}X = \{(x, 0); x \in \mathbb{R}\} = \text{eixo dos } xx$ .  $\triangleleft$

**Teorema 1.** (a) Se  $A_1, A_2$  são abertos em  $\mathbb{R}^n$  então  $A_1 \cap A_2$  é aberto.

(b) Se  $(A_\lambda)_{\lambda \in L}$  é uma família arbitrária de conjuntos abertos  $A_\lambda \subset \mathbb{R}^n$  então a reunião  $A \bigcup_{\lambda \in L} A_\lambda$  é um conjunto aberto.

*Demonstração.* Vide vol. 1, pág. 49. Mesma demonstração, substituindo apenas cada intervalo  $(a - \varepsilon, a + \varepsilon)$  pela bola  $B(a; \varepsilon)$ .

Resulta imediatamente do Teorema 1 que a interseção  $A = A_1 \cap \dots \cap A_k$  de um número finito de conjuntos abertos  $A_1, \dots, A_k$  é ainda um conjunto aberto. Entretanto, a interseção de infinitos abertos pode não ser aberta, como mostra o exemplo  $\bigcap_{k=1}^{\infty} B(a; 1/k) = \{a\}$ .

Seja  $X \subset \mathbb{R}^n$ . Diz-se que um subconjunto  $A \subset X$  é *aberto em  $X$*  quando cada ponto  $a \in A$  é centro de uma bola aberta  $B(a; r)$ , tal que  $B(a; r) \cap X \subset A$ . Isto significa que os *pontos de  $X$*  que estão suficientemente próximos de cada  $a \in A$  pertencem a  $A$ . A reunião  $U$  de todas essas bolas é um aberto tal que  $A = U \cap X$ . A recíproca é óbvia, de modo que um conjunto  $A \subset X$  é aberto em  $X$  se, e somente se,  $A = U \cap X$  onde  $U$  é aberto em  $\mathbb{R}^n$ .

Por exemplo, o intervalo  $(0, 1]$  é aberto em  $[0, 1]$  pois  $(0, 1] = (0, 2) \cap [0, 1]$ .  $\square$

#### 4 Seqüências em $\mathbb{R}^n$

Uma *seqüência* em  $\mathbb{R}^n$  é uma função  $f : \mathbb{N} \rightarrow \mathbb{R}^n$ , que associa a cada número natural  $k$  um ponto  $x_k \in \mathbb{R}^n$ . As notações para uma seqüência são  $(x_1, \dots, x_k, \dots)$ ,  $(x_k)_{k \in \mathbb{N}}$  ou simplesmente  $(x_k)$ .

Para cada  $i = 1, \dots, n$ , indicamos com  $x_{ki}$  a  $i$ -ésima coordenada de  $x_k$ . Assim,  $x_k = (x_{k1}, x_{k2}, \dots, x_{kn})$ . Dar uma seqüência em  $\mathbb{R}^n$  equivale a dar as  $n$  seqüências de números reais  $(x_{k1})_{k \in \mathbb{N}}, \dots, (x_{kn})_{k \in \mathbb{N}}$ .

Diz-se que a seqüência  $(x_k)_{k \in \mathbb{N}}$  é *limitada* quando existe uma bola em  $\mathbb{R}^n$  que contém todos os termos  $x_k$ . Isto equivale a dizer que existe  $c > 0$  tal que  $|x_k| \leq c$  para todo  $k \in \mathbb{N}$ . Em virtude das desigualdades que relacionam as três normas que consideramos em  $\mathbb{R}^n$ , ser limitada é uma propriedade da seqüência que independe de qual dessas três normas estamos tratando.

Se a seqüência  $(x_k)$  é limitada então, para todo  $i = 1, \dots, n$ , a seqüência  $(x_{ki})_{k \in \mathbb{N}}$  das  $i$ -ésimas coordenadas de  $x_k$  é também limitada, pois  $|x_{ki}| \leq |x_k|$ . Vale também a recíproca. Para prová-la, adotaremos em  $\mathbb{R}^n$  a norma do máximo.

Então, se  $|x_{k1}| \leq c_1, |x_{k2}| \leq c_2, \dots, |x_{kn}| \leq c_n$  para todo  $k \in \mathbb{N}$ , chamando de  $c$  o maior dos números  $c_1, c_2, \dots, c_n$  teremos  $|x_k| = \max\{|x_{k1}|, \dots, |x_{kn}|\} \leq c$  para todo  $k \in \mathbb{N}$ . Assim, se cada  $(x_{ki})_{k \in \mathbb{N}} (i = 1, \dots, n)$  é limitada, a seqüência  $(x_k)_{k \in \mathbb{N}}$  é limitada.

Uma *subseqüência* de  $(x_k)_{k \in \mathbb{N}}$  é a restrição desta seqüência a um subconjunto infinito  $\mathbb{N}' = \{k_1 < \dots < k_m < \dots\} \subset \mathbb{N}$ . As notações  $(x_k)_{k \in \mathbb{N}'}, (x_{k_m})_{m \in \mathbb{N}}$  ou  $(x_{k_1}, \dots, x_{k_m}, \dots)$  são usadas para indicar uma subseqüência.

Diz-se que o ponto  $a \in \mathbb{R}^n$  é o *limite* da seqüência  $(x_k)$  quando, para todo  $\varepsilon > 0$  dado arbitrariamente, é possível obter  $k_0 \in \mathbb{N}$  tal que  $k > k_0 \Rightarrow |x_k - a| < \varepsilon$ . Noutras palavras:  $k > k_0 \Rightarrow x_k \in B(a; \varepsilon)$ . Escreve-se então  $\lim_{k \rightarrow \infty} x_k = a$ ,

$\lim_{k \in \mathbb{N}} x_k = a$  ou  $\lim x_k = a$ , simplesmente.

De acordo com esta definição, tem-se  $\lim x_k = a$  se, e somente se,  $\lim |x_k - a| = 0$ .

Dizer que  $\lim x_k = a$  significa afirmar que qualquer bola de centro  $a$  contém todos os  $x_k$  com a possível exceção de um número finito de valores de  $k$  (que são  $1, 2, \dots, k_0$ ).

Uma seqüência  $(x_k)$  em  $\mathbb{R}^n$  diz-se *convergente* quando existe  $a = \lim x_k$ . Da observação acima resulta que toda seqüência convergente é limitada. É também óbvio que qualquer subseqüência de uma seqüência convergente é também convergente e tem o mesmo limite.

Observe-se ainda que a definição de limite faz uso de uma norma, porém as desigualdades  $|x|_M \leq |x| \leq |x|_S \leq n \cdot |x|_M$  mostram que a existência e o valor do limite não depende de qual das três normas usuais se está considerando. Este fato será empregado na demonstração do teorema abaixo, onde no final usamos a norma do máximo.

**Teorema 2.** *A seqüência  $(x_k)$  em  $\mathbb{R}^n$  converge para o ponto  $a = (a_1, \dots, a_n)$  se, e somente se, para cada  $i = 1, \dots, n$ , tem-se  $\lim_{k \rightarrow \infty} x_{ki} = a_i$ , isto é, cada coordenada de  $x_k$  converge para a coordenada correspondente de  $a$ .*

*Demonstração.* Para cada  $i = 1, \dots, n$ , tem-se  $|x_{ki} - a_i| \leq |x_k - a|$ , portanto  $\lim x_k = a \Rightarrow \lim_{k \rightarrow \infty} x_{ki} = a_i$ . Reciprocamente, se vale esta última igualdade então, dado  $\varepsilon > 0$ , existem  $k_1, \dots, k_n \in \mathbb{N}$  tais que  $k > k_i \Rightarrow |x_{ki} - a_i| < \varepsilon (i = 1, \dots, n)$ . Tomando  $k_0 = \max\{k_1, \dots, k_n\}$  e adotando em  $\mathbb{R}^n$  a norma do máximo, vemos que  $k > k_0 \Rightarrow |x_k - a| < \varepsilon$ . Logo  $\lim x_k = a$ .  $\square$

**Corolário 1.** *Se  $\lim x_k = a$ ,  $\lim y_k = b$  em  $\mathbb{R}^n$  e  $\lim \alpha_k = \alpha$  em  $\mathbb{R}$  então  $\lim(x_k + y_k) = a + b$  e  $\lim \alpha_k x_k = \alpha a$ .*

Tomando cada seqüência de coordenadas, o corolário resulta da propriedade correspondente em  $\mathbb{R}$ .

Além disso,  $\lim \langle x_k, y_k \rangle = \langle a, b \rangle$ , como se vê facilmente. E a desigualdade  $||x_k| - |a|| \leq |y_k - a|$  mostra ainda que se tem  $\lim |x_k| = |a|$  seja qual for a norma adotada.

**Teorema 3 (Bolzano-Weierstrass).** *Toda seqüência limitada em  $\mathbb{R}^n$  possui uma subseqüência convergente.*

*Demonstração.* Seja  $(x_k)$  uma seqüência limitada em  $\mathbb{R}^n$ . As primeiras coordenadas dos seus termos formam uma seqüência limitada  $(x_{k1})_{k \in \mathbb{N}}$  de números reais, a qual, pelo Teorema de Bolzano-Weierstrass na reta (vol. 1, pág. 25), possui uma subseqüência convergente. Isto é, existem um subconjunto infinito  $\mathbb{N}_1 \subset \mathbb{N}$  e um número real  $a_1$  tais que  $\lim_{k \in \mathbb{N}_1} x_{k1} = a_1$ . Por sua vez, a seqüência limitada  $(x_k)_{k \in \mathbb{N}_1}$  em  $\mathbb{R}$  possui uma subseqüência convergente: existem um subconjunto infinito  $\mathbb{N}_2 \subset \mathbb{N}_1$  e um número real  $a_2$  tais que  $\lim_{k \in \mathbb{N}_2} x_{k2} = a_2$ . E assim por diante, até obtermos  $n$  conjuntos infinitos  $\mathbb{N} \supset \mathbb{N}_1 \supset \mathbb{N}_2 \supset \dots \supset \mathbb{N}_n$  e números reais  $a_1, a_2, \dots, a_n$  tais que  $\lim_{k \in \mathbb{N}_i} x_{ki} = a_i$  para  $i = 1, 2, \dots, n$ . Então pomos  $a = (a_1, \dots, a_n)$  e, pelo Teorema 1, temos  $\lim_{k \in \mathbb{N}_n} x_k = a$ , o que prova o teorema.  $\square$

Uma seqüência de pontos  $x_k \in \mathbb{R}^n$  chama-se uma *seqüência de Cauchy* quando, para todo  $\varepsilon > 0$  dado, existe  $k_0 \in \mathbb{N}$  tal que  $k, r > k_0 \Rightarrow |x_k - x_r| < \varepsilon$ .

Toda seqüência de Cauchy  $(x_k)$  é limitada. Com efeito, tomando  $\varepsilon = 1$  na definição acima, vemos que existe um índice  $k_0$  tal que, salvo possivelmente os pontos  $x_1, \dots, x_{k_0}$  todos os demais termos  $x_k$  pertencem à bola  $B(x_{k_0+1}; 1)$ . Portanto o conjunto dos termos da seqüência é limitado.

A condição para que a seqüência  $(x_k)$  seja de Cauchy pode ser reformulada dizendo-se que  $\lim_{k, r \rightarrow \infty} |x_k - x_r| = 0$ , isto é, que  $\lim_{k, r \in \mathbb{N}} |x_k - x_r| = 0$ . Daí resulta que se  $\mathbb{N}' \subset \mathbb{N}$  é um subconjunto infinito, ou seja, se  $(x_r)_{r \in \mathbb{N}'}$  é uma subseqüência de  $(x_k)$  então  $\lim_{k \in \mathbb{N}, r \in \mathbb{N}'} |x_k - x_r| = 0$ .

**Teorema 4 (Critério de Cauchy).** *Uma seqüência em  $\mathbb{R}^n$  converge se, e somente se, é uma seqüência de Cauchy.*

*Demonstração.* Seja  $(x_k)$  uma seqüência de Cauchy em  $\mathbb{R}^n$ . Sendo limitada, ela possui uma subseqüência convergente  $(x_r)_{r \in \mathbb{N}'}$ . Seja  $a = \lim_{r \in \mathbb{N}'} x_r$ . Temos  $\lim_{r \in \mathbb{N}'} |x_r - a| = 0$  e  $\lim_{k \in \mathbb{N}, r \in \mathbb{N}'} |x_k - x_r| = 0$ , como observamos acima. Então, de  $|x_k - a| \leq |x_k - x_r| + |x_r - a|$  resulta que  $\lim_{k \in \mathbb{N}} |x_k - a| = 0$ , ou seja,  $\lim_{k \rightarrow \infty} x_k = a$ . Reciprocamente, se  $(x_k)$  é convergente, com  $\lim x_k = a$ , então, como  $|x_k - x_r| \leq |x_k - a| + |x_r - a|$ , concluímos que  $\lim_{k, r \rightarrow \infty} |x_k - x_r| = 0$ , ou seja,  $(x_k)$  é de Cauchy.  $\square$

## 5 Conjuntos fechados

Diz-se que o ponto  $a$  é *aderente* ao conjunto  $X \subset \mathbb{R}^n$  quando existe uma seqüência de pontos  $x_k \in X$  tais que  $\lim x_k = a$ .

Chama-se *fecho* do conjunto  $X \subset \mathbb{R}^n$  ao conjunto  $\bar{X}$  formado por todos os pontos aderentes a  $X$ . Portanto  $a \in \bar{X} \Leftrightarrow a = \lim x_k, x_k \in X$ . Dizer que  $a \in \bar{X}$  é o mesmo que afirmar que  $a$  é aderente a  $X$ .

Um conjunto  $F \in \mathbb{R}^n$  chama-se *fechado* quando  $\bar{F} = F$ , isto é, quando o limite de toda seqüência convergente de pontos de  $F$  é ainda um ponto de  $F$ .

Todo ponto  $x \in X$  é aderente a  $X$  pois é limite da seqüência constante  $(x, x, \dots)$ . Assim,  $X \subset \bar{X}$  qualquer que seja  $X \subset \mathbb{R}^n$ . Também é óbvio que  $X \subset Y \Rightarrow \bar{X} \subset \bar{Y}$ .

**Exemplo 7.** Se  $|x| = r$  então  $x$  não pertence à bola aberta  $B = B(0; r)$  porém é aderente a ela. Com efeito, pondo  $x_k = \left(1 - \frac{1}{k}\right)x$  para todo  $k \in \mathbb{N}$ , temos  $x_k \in B(0; r)$  e  $\lim x_k = x$ , logo  $x \in \bar{B}$ .

Reciprocamente, se  $x \in \bar{B}$  então  $x = \lim y_k$  com  $|y_k| < r$  para todo  $k \in \mathbb{N}$ , portanto  $|x| = \lim |y_k| \leq r$ . Conclui-se então que  $x \in \bar{B} \Leftrightarrow |x| \leq r$ , ou seja,  $\bar{B} = B[0; r]$ . O mesmo argumento mostra que o fecho de toda bola aberta  $B(a; r)$  é a bola fechada  $B[a; r]$ .  $\triangleleft$

O teorema abaixo resume as principais propriedades do fecho de um conjunto.

**Teorema 5.** (a) *O ponto  $a$  é aderente ao conjunto  $X \subset \mathbb{R}^n$  se, e somente se, toda bola de centro  $a$  contém algum ponto de  $X$ .*

(b) *Um conjunto  $F \subset \mathbb{R}^n$  é fechado se, e somente se, seu complementar  $\mathbb{R}^n - F$  é aberto. Equivalentemente:  $A \subset \mathbb{R}^n$  é aberto se, e somente se,  $\mathbb{R}^n - A$  é fechado.*

(c) *O fecho de qualquer conjunto  $X \subset \mathbb{R}^n$  é fechado. Noutras palavras: para todo  $X \subset \mathbb{R}^n$  tem-se  $\bar{\bar{X}} = \bar{X}$ .*

*Demonstração.* (a) Se  $a$  é aderente a  $X$  então  $a = \lim x_k$ , com  $x_k \in X$  para todo  $k \in \mathbb{N}$ . Portanto qualquer bola  $B(a; r)$  contém pontos de  $X$ , a saber, todos os  $x_k$  com  $k$  suficientemente grande. Reciprocamente, se toda bola de centro  $a$  contém pontos de  $X$ , podemos escolher, para cada  $k \in \mathbb{N}$ , um ponto  $x_k \in X$  que esteja na bola  $B(a; 1/k)$ , isto é,  $|x_k - a| < 1/k$ . Então  $\lim x_k = a$ , logo  $a$  é aderente a  $X$ .

(b) As seguintes afirmações são equivalentes: (1)  $F$  é fechado. (2) Se  $x \in \mathbb{R}^n - F$  então  $x$  não é aderente a  $F$ . (3) Se  $x \in \mathbb{R}^n - F$  então existe  $r > 0$  tal que  $B(x; r) \subset \mathbb{R}^n - F$  (em virtude da parte (a) acima). (4)  $\mathbb{R}^n - F$  é aberto. Assim,  $F$  fechado  $\Leftrightarrow \mathbb{R}^n - F$  aberto.

Escrevendo  $A = \mathbb{R}^n - F$ , donde  $F = \mathbb{R}^n - A$ , esta última conclusão lê-se assim:  $A$  é aberto se, e somente se,  $\mathbb{R}^n - A$  é fechado.

(c) Se  $x \in \mathbb{R}^n - \bar{X}$  (isto é,  $x$  não é aderente a  $X$ ) então, por (a), existe uma bola  $B = B(x; r)$  que não contém pontos de  $X$ , ou seja,  $X \subset \mathbb{R}^n - B$ . Logo  $\bar{X} \subset \overline{\mathbb{R}^n - B}$ . Mas, pela parte (b) acima,  $\mathbb{R}^n - B$  é fechado; portanto  $\bar{X} \subset \mathbb{R}^n - B$  ou, equivalentemente,  $B \subset \mathbb{R}^n - \bar{X}$ . Assim, todo ponto  $x \in \mathbb{R}^n - \bar{X}$  é um ponto interior e  $\mathbb{R}^n - \bar{X}$  é aberto. Segue-se que  $\bar{X}$  é fechado.  $\square$

Alguns conjuntos  $X \subset \mathbb{R}^n$  não são abertos nem fechados, como  $X = B(a; r) \cup \{b\}$ , onde  $|b - a| = r$ . Ou então  $X =$  conjunto dos pontos de  $\mathbb{R}^n$  com coordenadas racionais ( $X = \mathbb{Q}^n$ ).

Chama-se *distância do ponto  $a \in \mathbb{R}^n$  ao conjunto  $X \subset \mathbb{R}^n$*  ao número

$$d(a; X) = \inf. \{ |x - a| ; x \in X \}.$$

Pela definição de ínfimo, para cada  $k \in \mathbb{N}$  existe um ponto  $x_k \in X$  tal que  $d(a; X) \leq |x_k - a| < d(a; X) + \frac{1}{k}$ , portanto  $\lim_{k \rightarrow \infty} |x_k - a| = d(a; X)$ . A seqüência  $(x_k)$  é certamente limitada, portanto possui uma subsequência convergente. Descartando (por serem desnecessários) os termos  $x_k$  que não estejam nessa subsequência, vemos que existe um ponto  $x_0 = \lim x_k$  tal que  $d(x_0, X) = |x_0 - a|$ . Tem-se  $x_0 \in \bar{X}$ . Se o conjunto  $X$  for fechado então  $x_0 \in X$ . Podemos então enunciar o

**Teorema 6.** *Seja  $F \subset \mathbb{R}^n$  um conjunto fechado. Dado qualquer  $a \in \mathbb{R}^n$  existe (pelo menos um)  $x_0 \in F$  tal que  $|x_0 - a| \leq |x - a|$  para todo  $x \in F$ .*

Noutras palavras: se  $F \subset \mathbb{R}^n$  é fechado então, para  $a \in \mathbb{R}^n$  qualquer, a função  $f : F \rightarrow \mathbb{R}$  dada por  $f(x) = |x - a|$  assume seu valor mínimo em algum ponto  $x_0 \in F$ . Então tem-se  $d(a, F) = |x_0 - a|$ .

Se  $X \subset Y \subset \mathbb{R}^n$ , diz-se que  $X$  é *denso* em  $Y$  quando  $\bar{X} = Y$ . Por exemplo,  $B(a; r)$  é denso em  $B[a; r]$  e  $\mathbb{Q}^n$  é denso em  $\mathbb{R}^n$ .

Dizemos que  $a \in \mathbb{R}^n$  é *ponto de acumulação* do conjunto  $X \subset \mathbb{R}^n$  quando toda bola de centro  $a$  contém algum ponto de  $X$  diferente de  $a$ . (Noutras palavras, quando  $a \in \overline{X} - \{a\}$ .) Um ponto de acumulação de  $X$  pode pertencer a  $X$  ou não. Se  $a \in X$  não é ponto de acumulação de  $X$ , diz-se que  $a$  é um *ponto isolado* de  $X$ . Isto significa que existe  $r > 0$  tal que  $B(a; r) \cap X = \{a\}$ . Quando todos os pontos de  $X$  são isolados, dizemos que  $X$  é um conjunto *discreto*.

**Exemplo 8.** Todos os pontos de uma bola são pontos de acumulação. O conjunto  $\mathbb{Z}^n$  dos pontos de  $\mathbb{R}^n$  com coordenadas inteiras é um conjunto discreto.  $\triangleleft$

As demonstrações dos três teoremas seguintes são omitidas pois são praticamente as mesmas dos seus análogos unidimensionais, provados no volume 1 (págs.

50, 52 e 53). Basta substituir cada intervalo  $(a - r, a + r)$  pela bola  $B(a; r)$  e considerar  $|x|$  como a norma de  $x$ .

**Teorema 7.** *Sejam  $a$  um ponto e  $X$  um subconjunto de  $\mathbb{R}^n$ . As seguintes afirmações são equivalentes:*

- (1)  $a$  é um ponto de acumulação de  $X$ .
- (2)  $a$  é limite de uma seqüência de pontos  $x_k \in X - \{a\}$ .
- (3) Toda bola de centro  $a$  contém uma infinidade de pontos de  $X$ .

**Teorema 8.** *Todo subconjunto infinito limitado  $X \subset \mathbb{R}^n$  admite pelo menos um ponto de acumulação.*

**Teorema 9.** (a) *Se  $F_1$  e  $F_2$  são subconjuntos fechados de  $\mathbb{R}^n$  então  $F_1 \cup F_2$  é também fechado.*

- (b) *Se  $(F_\lambda)_{\lambda \in L}$  é uma família arbitrária de conjuntos fechados então a interseção  $F = \bigcap_{\lambda \in L} F_\lambda$  é um conjunto fechado.*

Cabe aqui a observação de que (a) implica que a reunião  $F_1 \cup \dots \cup F_k$  de um número finito de conjuntos fechados é ainda um conjunto fechado. Entretanto isto não vale para reuniões infinitas. Com efeito, um conjunto qualquer, fechado ou não, é a reunião dos seus pontos, que são conjuntos fechados.

Segue-se do item (2) do Teorema 7 que o fecho do conjunto  $X$  é formado acrescentando-lhe seus pontos de acumulação que por ventura não pertençam a  $X$ .

Seja  $X \subset \mathbb{R}^n$ . Diz-se que um subconjunto  $F \subset X$  é *fechado em  $X$*  quando  $F$  contém todos os seus pontos aderentes *que pertencem a  $X$* . Assim,  $F$  é fechado em  $X$  se, e somente se,  $F = \bar{F} \cap X$ .  $F$  é fechado em  $X$  quando, e somente quando,  $F = G \cap X$  onde  $G \subset \mathbb{R}^n$  é fechado.

Com efeito se  $F = G \cap X$  com  $G$  fechado então  $\bar{F} \subset G$ , logo  $F = F \cap X \subset \bar{F} \cap X \subset G \cap X = F$ , donde  $F = \bar{F} \cap X$  e  $F$  é fechado em  $X$ .

O conjunto  $F \subset X$  é fechado em  $X$  se, e somente se,  $X - F$  (seu complementar relativamente a  $X$ ) é aberto em  $X$ . Com efeito  $F = G \cap X \Leftrightarrow X - F = (\mathbb{R}^n - G) \cap X$ , onde  $G \subset \mathbb{R}^n$  é fechado se, e somente se,  $\mathbb{R}^n - G$  é aberto.

Analogamente,  $A \subset X$  é aberto em  $X$  se, e somente se,  $X - A$  é fechado em  $X$  pois  $A = U \cap X \Leftrightarrow X - A = (\mathbb{R}^n - U) \cap X$  e  $U \subset \mathbb{R}^n$  é aberto se, e somente se,  $\mathbb{R}^n - U$  é fechado.

## 6 Conjuntos compactos

Um conjunto  $X \subset \mathbb{R}^n$  chama-se *compacto* quando é limitado e fechado.

**Exemplo 9.** Toda bola fechada  $B[a; r]$  é compacta e nenhuma bola aberta é. O conjunto  $\mathbb{Z}^n$  é fechado mas não é limitado, logo não é compacto. Toda esfera  $S[a; r]$  é compacta.  $\triangleleft$

**Teorema 10.** As seguintes afirmações sobre o conjunto  $K \subset \mathbb{R}^n$  são equivalentes:

- (1)  $K$  é compacto;
- (2) Toda seqüência de pontos  $x_k \in K$  possui uma subseqüência que converge para um ponto de  $K$ .

*Demonstração.* Se  $K$  é compacto então toda seqüência de pontos  $x_k \in K$  é limitada, pois  $K$  é limitado. Por Bolzano-Weierstrass, uma subseqüência  $(x_k)_{k \in \mathbb{N}}$  converge para um ponto  $a = \lim_{k \in \mathbb{N}} x_k$ . Como  $K$  é fechado, tem-se  $a \in K$ . Logo (1) implica (2). Reciprocamente, se vale (2) então  $K$  é limitado pois do contrário existiria, para cada  $k \in \mathbb{N}$  um ponto  $x_k \in K$  tal que  $|x_k| > k$ . A seqüência  $(x_k)$  assim obtida não possuiria subseqüência limitada logo nenhuma de suas subseqüências seria convergente. Além disso,  $K$  é fechado pois se  $a = \lim x_k$  com  $x_k \in K$  para todo  $k \in \mathbb{N}$  então, por (2) uma subseqüência de  $(x_k)$  convergiria para um ponto de  $K$ . Mas toda subseqüência de  $(x_k)$  converge para  $a$ . Logo  $a \in K$ . Isto mostra que (2)  $\Rightarrow$  (1) e completa a demonstração.  $\square$

Estendendo a discussão da seção 5, dados os conjuntos  $X, Y \subset \mathbb{R}^n$ , podemos definir a *distância* entre eles pondo

$$d(X, Y) = \inf. \{ |x - y|; x \in X, y \in Y \},$$

cabendo-nos agora indagar se, supondo  $X$  e  $Y$  fechados, existem  $x_0 \in X$  e  $y_0 \in Y$  tais que  $d(X, Y) = |x_0 - y_0|$ . Nem sempre. Com efeito, tomando em  $\mathbb{R}^2$  o conjunto  $X$  como sendo o eixo das abcissas, isto é,  $X = \{(x, 0); x \in \mathbb{R}\}$  e  $Y = \{(x, 1/x); x > 0\}$ , ou seja,  $Y =$  ramo positivo da hipérbole  $y = 1/x$ , vemos que  $X$  e  $Y$  são subconjuntos fechados disjuntos em  $\mathbb{R}^2$  tais que  $d(X, Y) = 0$ . Entretanto, vale o seguinte resultado, que contém o Teorema 5 como caso particular:

**Teorema 11.** *Sejam  $K \subset \mathbb{R}^n$  compacto e  $F \subset \mathbb{R}^n$  fechado. Existem  $x_0 \in K$  e  $y_0 \in F$  tais que  $|x_0 - y_0| \leq |x - y|$  para quaisquer  $x \in K$  e  $y \in F$ .*

*Demonstração.* Da definição de ínfimo segue-se que existem seqüências de pontos  $x_k \in K$  e  $y_k \in F$  tais que  $d(K, F) = \lim |x_k - y_k|$ . Passando a uma subseqüência, se necessário, a compacidade de  $K$  nos permite admitir que  $\lim x_k = x_0 \in K$ . Além disso a seqüência  $(y_k)$  é limitada pois  $|y_k| \leq |y_k - x_k| + |x_k|$ , onde  $|y_k - x_k|$  é limitada por ser convergente e  $|x_k|$  é limitada pois  $x_k \in K$ . Logo, passando novamente a uma subseqüência, se necessário, podemos admitir que  $\lim y_k = y_0$ ,

com  $y_0 \in F$  pois  $F$  é fechado. Então  $|x_0 - y_0| = \lim |x_k - y_k| = d(K, F) \leq |x - y|$  para quaisquer  $x \in K$  e  $y \in F$ .  $\square$

**Corolário 2.** *Sejam  $K \subset U \subset \mathbb{R}^n$ , onde  $K$  é compacto e  $U$  é aberto. Existe  $\varepsilon > 0$  tal que toda bola  $B(x; \varepsilon)$ , com raio  $\varepsilon$  e centro num ponto  $x \in K$ , está contida em  $U$ .*

Com efeito, sejam  $x_0 \in K$  e  $y_0 \in F = \mathbb{R}^n - U$  tais que  $|x_0 - y_0| \leq |x - y|$  para quaisquer  $x \in K$  e  $y \in F$ . Ponhamos  $\varepsilon = |x_0 - y_0|$ . Como  $K \subset U$ , vemos que  $K \cap F = \emptyset$ , portanto  $x_0 \neq y_0$  e daí  $\varepsilon > 0$ . Assim, se  $x \in K$  e  $y \notin U$ , tem-se  $|x - y| \geq \varepsilon$ . Noutras palavras, se  $x \in K$  então  $B(x; \varepsilon) \subset U$ .

Se  $F_1 \supset F_2 \supset \dots \supset F_k \supset \dots$  é uma seqüência decrescente de fechados não-vazios em  $\mathbb{R}^n$ , pode ocorrer que  $\bigcap_{k=1}^{\infty} F_k = \emptyset$ . Isto ocorre, por exemplo, quando tomamos  $F_k = [k, +\infty)$  em  $\mathbb{R}$ . O teorema abaixo mostra que isto não acontece quando um dos  $F_k$  é limitado (portanto todos os seguintes são).

**Teorema 12 (Cantor).** *Seja  $K_1 \supset K_2 \supset \dots \supset K_k \supset \dots$  uma seqüência decrescente de compactos não-vazios em  $\mathbb{R}^n$ . Existe pelo menos um ponto  $a \in \mathbb{R}^n$  que pertence a todos os  $K_k$ . Noutros termos:  $\bigcap_{k=1}^{\infty} K_k \neq \emptyset$ .*

*Demonstração.* Para cada  $k \in \mathbb{N}$ , escolhamos um ponto  $x_k \in K_k$ . A seqüência  $(x_k)$  é limitada, logo possui uma subseqüência  $(x_r)_{r \in \mathbb{N}'}$ , que converge para  $a = \lim_{r \in \mathbb{N}'} x_r$ . Mostremos que  $a \in K_k$  para todo  $k \in \mathbb{N}$ . De fato, dado  $k$ , temos  $K_r \subset K_k$  sempre que  $r \in \mathbb{N}'$  e  $r > k$ . Assim,  $r \in \mathbb{N}', r > k \Rightarrow x_r \in K_k$ . Segue-se que  $a = \lim_{r \in \mathbb{N}'} x_r$  pertence ao conjunto fechado  $K_k$ .  $\square$

Uma propriedade fundamental dos conjuntos compactos é o fato de que toda cobertura aberta de um compacto possui uma subcobertura finita. Vejamos isto.

Uma *cobertura* do conjunto  $X \subset \mathbb{R}^n$  é uma família  $(C_\lambda)_{\lambda \in L}$  de subconjuntos  $C_\lambda \subset \mathbb{R}^n$  tais que  $X \subset \bigcup_{\lambda \in L} C_\lambda$ . Isto significa que para cada  $x \in X$  existe um  $\lambda \in L$  tal que  $x \in C_\lambda$ .

Uma *subcobertura* é uma subfamília  $(C_\lambda)_{\lambda \in L'}$ ,  $L' \subset L$ , tal que ainda se tem  $X \subset \bigcup_{\lambda \in L'} C_\lambda$ .

Diz-se que a cobertura  $X \subset \bigcup_{\lambda \in L} C_\lambda$  é *aberta* quando os  $C_\lambda$  forem todos abertos, ou *finita* quando  $L$  é um conjunto finito.

**Teorema 13 (Borel-Lebesgue).** *Toda cobertura aberta  $X \subset \bigcup_{\lambda \in L} A_\lambda$  de um compacto  $X \subset \mathbb{R}^n$  admite uma subcobertura finita  $X \subset A_{\lambda_1} \cup \dots \cup A_{\lambda_k}$ .*

Inicialmente, prepararemos o terreno para estabelecer um lema que torna a demonstração do teorema quase imediata.

Seja  $X \subset \mathbb{R}^n$  um conjunto limitado. O *diâmetro* de  $X$  é o número

$$\text{diam. } X = \sup \{ |x - y|; x, y \in X \}.$$

Segue-se imediatamente desta definição que se  $\text{diam. } X = d$  e  $x \in X$  então  $X \subset B[x; d]$ .

Dado  $\alpha > 0$ , um *cubo* de aresta  $\alpha$  é um produto cartesiano  $C = \prod_{i=1}^n [a_i, a_i + \alpha]$  de  $n$  intervalos de comprimento  $\alpha$ . Se  $x = (x_1, \dots, x_n)$  e  $y = (y_1, \dots, y_n)$  pertencem a  $C$  então, para cada  $i = 1, \dots, n$ , tem-se  $|x_i - y_i| \leq \alpha$  logo  $|x - y| = \sqrt{\sum (x_i - y_i)^2} \leq \alpha \sqrt{n}$ . Tomando  $y_i = a_i + \alpha$  temos  $|x - y| = \alpha \sqrt{n}$ , portanto  $\alpha \sqrt{n}$  é o diâmetro do cubo de aresta  $\alpha$  em  $\mathbb{R}^n$ .

A decomposição  $\mathbb{R} = \bigcup_{m \in \mathbb{Z}} [m\alpha, (m+1)\alpha]$  da reta em intervalos adjacentes de comprimento  $\alpha$  determina uma decomposição de  $\mathbb{R}^n$  como reunião de cubos adjacentes de aresta  $\alpha$ . A saber, para cada  $m = (m_1, \dots, m_n) \in \mathbb{Z}^n$ , pomos  $C_m = \prod_{i=1}^n [m_i\alpha, (m_i+1)\alpha]$  e temos  $\mathbb{R}^n = \bigcup_{m \in \mathbb{Z}^n} C_m$ .

Para todo  $X \subset \mathbb{R}^n$  tem-se  $X = \bigcup_{m \in \mathbb{Z}^n} (X \cap C_m)$ . Se  $X$  é limitado apenas um número finito das interseções  $X \cap C_m$  são não-vazias, logo podemos escrever

$$X = X_1 \cup \dots \cup X_k$$

onde cada  $X_i$  é da forma  $X \cap C_m$ , logo tem diâmetro  $\leq \alpha \sqrt{m}$ . Se  $X$  for compacto então cada  $X_i$  é compacto. Isto prova o

**Lema 1.** *Seja  $K \subset \mathbb{R}^n$  compacto. Para todo  $\varepsilon > 0$  existe uma decomposição  $K = K_1 \cup \dots \cup K_k$  onde cada  $K_i$  é compacto e tem diâmetro  $\leq \varepsilon$ .*

*Demonstração do Teorema de Borel-Lebesgue.* Seja  $K \subset \mathbb{R}^n$  compacto. Suponhamos, por absurdo, que  $K \subset \cup A_\lambda$  seja uma cobertura aberta que não admite subcobertura finita. Exprimamos  $K$  como reunião finita de compactos, todos com diâmetro  $< 1$ . Pelo menos um deles, que chamaremos  $K_1$ , é tal que  $K_1 \subset \cup A_\lambda$  não admite subcobertura finita. Escrevendo  $K_1$  como reunião finita de compactos de diâmetro  $< 1/2$ , vemos que pelo menos um deles, digamos  $K_2$ , não pode ser coberto por um número finito de  $A'_\lambda$ 's. Prosseguindo assim, obtemos uma seqüência decrescente de compactos  $K_1 \supset K_2 \supset \dots \supset K_k \supset \dots$  com  $\text{diam } K_k < 1/k$  e tal que nenhum deles está contido numa reunião finita de  $A'_\lambda$ 's. Em particular, todos os  $K_k$  são não-vazios. Pelo Teorema 12, existe  $a \in \bigcap_{k=1}^{\infty} K_k$ . Para algum  $\lambda$ ,

tem-se  $a \in A_\lambda$ . Como  $A_\lambda$  é aberto, tem-se  $B(a; 1/k) \subset A_\lambda$  para algum  $k$ . Sendo  $a \in K_k$  e  $\text{diam } K_k < 1/k$ , concluímos que  $K_k \subset B(a; 1/k)$ , donde  $K_k \subset A_\lambda$ , o que é uma contradição.

## 7 Aplicações contínuas

Uma aplicação  $f : X \rightarrow \mathbb{R}^n$ , definida no conjunto  $X \subset \mathbb{R}^n$ , associa a cada ponto  $x \in X$  sua imagem  $f(x) = (f_1(x), \dots, f_n(x))$ . As funções reais  $f_1, \dots, f_n : X \rightarrow \mathbb{R}$ , assim definidas, chamam-se as *funções-coordenada* de  $f$ . Escreve-se então  $f = (f_1, \dots, f_n)$ .

Se  $Y \subset \mathbb{R}^n$  é tal que  $f(X) \subset Y$  podemos (com um abuso de notação que é irrelevante em nosso contexto) escrever  $f : X \rightarrow Y$  em vez de  $f : X \rightarrow \mathbb{R}^n$ .

Diz-se que  $f$  é *contínua no ponto*  $a \in X$  quando, para cada  $\varepsilon > 0$  arbitrariamente dado, pode-se obter  $\delta > 0$  tal que

$$x \in X, |x - a| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon.$$

Noutros termos: para cada bola  $B(f(a); \varepsilon)$  dada, existe uma bola  $B(a; \delta)$  tal que  $f(B(a; \delta) \cap X) \subset B(f(a); \varepsilon)$ .

A continuidade de  $f$  no ponto  $a$  independe das normas que se utilizem em  $\mathbb{R}^m$  e  $\mathbb{R}^n$ .

Diremos que  $f : X \rightarrow \mathbb{R}^n$  é uma aplicação *contínua* no conjunto  $X \subset \mathbb{R}^m$  quando  $f$  é contínua em todos os pontos  $a \in X$ .

**Teorema 14.** *Sejam  $X \subset \mathbb{R}^m$ ,  $Y \subset \mathbb{R}^n$ ,  $f : X \rightarrow \mathbb{R}^n$  com  $f(X) \subset Y$  e  $g : Y \rightarrow \mathbb{R}^p$ . Se  $f$  é contínua no ponto  $a \in X$  e  $g$  é contínua no ponto  $f(a)$  então  $g \circ f : X \rightarrow \mathbb{R}^p$  é contínua no ponto  $a$ . Ou seja: a composta de duas aplicações contínuas é contínua.*

*Demonstração.* Seja dado  $\varepsilon > 0$ . A continuidade de  $g$  no ponto  $f(a)$  assegura a existência de  $\lambda > 0$  tal que  $y \in Y, |y - f(a)| < \lambda \Rightarrow |g(y) - g(f(a))| < \varepsilon$ . Por sua vez, dado  $\lambda > 0$ , a continuidade de  $f$  no ponto  $a$  fornece  $\delta > 0$  tal que  $x \in X, |x - a| < \delta \Rightarrow |f(x) - f(a)| < \lambda \Rightarrow |g(f(x)) - g(f(a))| < \varepsilon$ , logo  $g \circ f$  é contínua no ponto  $a$ .  $\square$

**Teorema 15.** (a) *A aplicação  $f : X \rightarrow \mathbb{R}^n$  é contínua no ponto  $a \in X$  se, e somente se, para toda seqüência de pontos  $x_k \in X$  com  $\lim x_k = a$ , tem-se  $\lim f(x_k) = f(a)$ .*

(b) *A aplicação  $f : X \rightarrow \mathbb{R}^n$  é contínua no ponto  $a \in X$  se, e somente se, suas funções-coordenada  $f_1, \dots, f_n : X \rightarrow \mathbb{R}$  são contínuas nesse ponto.*

*Demonstração.* (a) Seja  $f : X \rightarrow \mathbb{R}^n$  contínua no ponto  $a$ . Dada a seqüência de pontos  $x_k \in X$  com  $\lim x_k = a$ , para todo  $\varepsilon > 0$  existe  $\delta > 0$  tal que  $f(B(a; \delta)) \subset B(f(a); \varepsilon)$ . Correspondente a  $\delta$ , existe  $k_0 \in \mathbb{N}$  tal que  $k > k_0 \Rightarrow x_k \in B(a; \delta)$ , logo  $k > k_0 \Rightarrow f(x_k) \in B(f(a); \varepsilon)$ . Isto mostra que  $\lim f(x_k) = f(a)$ . Reciprocamente, suponhamos por absurdo, que  $\lim x_k = a$  implique  $\lim f(x_k) \neq f(a)$ , porém  $f$  seja descontínua no ponto  $a$ . Então existe  $\varepsilon > 0$  com a seguinte propriedade: para todo  $k \in \mathbb{N}$ , podemos encontrar  $x_k \in X$  com  $|x_k - a| < 1/k$  e  $|f(x_k) - f(a)| \geq \varepsilon$ . Assim, temos  $\lim x_k = a$  mas não temos  $\lim f(x_k) = f(a)$ , uma contradição.

(b) Isto decorre imediatamente do Teorema 2 junto com a parte (a), que acabamos de provar.  $\square$

**Teorema 16.** *Seja  $X \subset \mathbb{R}^m$ . Se as aplicações  $f, g : X \rightarrow \mathbb{R}^n$  e  $\alpha : X \rightarrow \mathbb{R}$  são contínuas no ponto  $a \in X$  então são também contínuas nesse ponto as aplicações  $f + g : X \rightarrow \mathbb{R}^n$ ,  $\langle f, g \rangle : X \rightarrow \mathbb{R}^2$ ,  $|f| : X \rightarrow \mathbb{R}$  e  $\alpha f : X \rightarrow \mathbb{R}^n$ , definidas por  $(f + g)(x) = f(x) + g(x)$ ,  $\langle f, g \rangle(x) = \langle f(x), g(x) \rangle$ ,  $|f|(x) = |f(x)|$  e  $(\alpha f)(x) = \alpha(x) \cdot f(x)$ .*

*Demonstração.* Isto resulta do Teorema 15(a) juntamente com o Corolário do Teorema 2.  $\square$

**Teorema 17.** *A imagem  $f(K)$  do conjunto compacto  $K \subset X$  pela aplicação contínua  $f : X \rightarrow \mathbb{R}^n$  é também um conjunto compacto.*

*Demonstração.* Seja  $(y_k)$  uma seqüência de pontos em  $f(K)$ . Para cada  $k \in \mathbb{N}$  existe  $x_k \in K$  tal que  $f(x_k) = y_k$ . Como  $K$  é compacto, uma subseqüência  $(x_k)_{k \in \mathbb{N}'}$  converge para um ponto  $a \in K$ . Sendo  $f$  contínua nesse ponto  $a$ , de  $\lim_{k \in \mathbb{N}'} x_k = a$  resulta, pelo Teorema 15, que  $\lim_{k \in \mathbb{N}'} f(x_k) = f(a)$ . Logo toda seqüência de pontos  $y_k = f(x_k) \in f(K)$  possui uma subseqüência  $(y_k)_{k \in \mathbb{N}'}$  convergente para um ponto  $f(a) \in f(K)$ . Noutras palavras:  $f(K)$  é compacto.  $\square$

**Corolário 3 (Weierstrass).** *Seja  $K \subset \mathbb{R}^m$  compacto. Se  $f : K \rightarrow \mathbb{R}$  é uma função real contínua, então existem  $x_0, x_1 \in K$  tais que  $f(x_0) \leq f(x) \leq f(x_1)$  para todo  $x \in K$ .*

Noutras palavras: toda função real contínua num conjunto compacto  $K$  atinge seus valores mínimo e máximo em pontos de  $K$ .

Para provar o Teorema de Weierstrass basta observar que, sendo  $f(K) \subset \mathbb{R}$  compacto, os números  $y_0 = \inf f(K)$  e  $y_1 = \sup f(K)$  pertencem a  $f(K)$ , isto é,  $y_0 = f(x_0)$  e  $y_1 = f(x_1)$ , com  $x_0, x_1 \in K$ .

**Teorema 18.** *Seja  $X \subset \mathbb{R}^m$ . A aplicação  $f : X \rightarrow \mathbb{R}^n$  é contínua se, e somente se, a imagem inversa  $f^{-1}(A)$  de todo conjunto aberto  $A \subset \mathbb{R}^n$  é um subconjunto aberto em  $X$ .*

*Demonstração.* Seja  $f$  contínua. Se  $A \subset \mathbb{R}^n$  é aberto então, para todo  $x \in f^{-1}(A)$  existe  $\varepsilon > 0$  tal que  $B(f(x); \varepsilon) \subset A$ . Pela continuidade de  $f$ ,  $x$  é centro de uma bola aberta  $B_x$  tal que  $f(B_x \cap X) \subset B(f(x); \varepsilon) \subset A$ , logo  $x \in B_x \cap X \subset f^{-1}(A)$ . Isto valendo para todo  $x \in f^{-1}(A)$ , resulta que  $f^{-1}(A) \subset U \cap X \subset f^{-1}(A)$ , logo  $f^{-1}(A) = U \cap X$ , onde  $U$  é a reunião das bolas abertas  $B_x$ ,  $x \in f^{-1}(A)$ . Reciprocamente, suponhamos que, para todo aberto  $A \subset \mathbb{R}^n$ ,  $f^{-1}(A)$  seja aberto em  $X$ , isto é,  $f^{-1}(A) = U \cap X$  com  $U$  aberto em  $\mathbb{R}^n$ . Então, dado  $x \in X$  e  $\varepsilon > 0$ , tomamos  $A = B(f(x); \varepsilon)$  e obtemos  $U \subset \mathbb{R}^n$  aberto tal que  $U \cap X = f^{-1}(B(f(x); \varepsilon))$ . Certamente  $x \in U$ , logo existe  $\delta > 0$  tal que  $B(x; \delta) \subset U$  e assim  $f(B(x; \delta) \cap X) \subset B(f(x); \varepsilon)$ . Portanto,  $f$  é contínua em todos os pontos  $x \in X$ .  $\square$

**Teorema 19.** *Seja  $X \subset \mathbb{R}^m$ . A aplicação  $f : X \rightarrow \mathbb{R}^n$  é contínua se, e somente se, a imagem inversa de todo conjunto fechado  $F \subset \mathbb{R}^n$  é um subconjunto fechado em  $X$ .*

*Demonstração.* Isto resulta do Teorema 18 se observarmos que, pondo  $A = \mathbb{R}^n - F$  então  $A$  é aberto em  $\mathbb{R}^n$  e que  $f^{-1}(F) = X - f^{-1}(A)$  é fechado em  $X$  se, e somente se,  $f^{-1}(A)$  é aberto em  $X$ .  $\square$

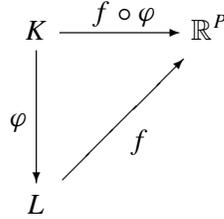
**Observação.** Dada  $f : X \rightarrow \mathbb{R}^n$ , se  $f(X) \subset Y \subset \mathbb{R}^n$  podemos considerar  $f$  como uma aplicação de  $X$  em  $Y$  e escrever  $f : X \rightarrow Y$ . Se  $A$  e  $F$  são subconjuntos de  $\mathbb{R}^n$  então  $f^{-1}(A) = f^{-1}(A \cap Y)$  e  $f^{-1}(F) = f^{-1}(F \cap Y)$ . Logo podemos enunciar os Teoremas 18 e 19 assim: *A aplicação  $f : X \rightarrow Y$  é contínua se, e somente se, a imagem inversa por  $f$  de todo subconjunto aberto (respect. fechado) em  $Y$  é um subconjunto aberto (respect. fechado) em  $X$ .*

**Corolário 4.** *Seja  $X \subset \mathbb{R}^m$  aberto (respect. fechado). A fim de que  $f : X \rightarrow \mathbb{R}^n$  seja contínua é necessário e suficiente que a imagem inversa por  $f$  de todo subconjunto aberto (respect. fechado) em  $\mathbb{R}^n$  seja um conjunto aberto (respect. fechado) em  $\mathbb{R}^m$ .*

**Corolário 5.** *Sejam  $f, g : X \rightarrow \mathbb{R}$  contínuas no conjunto  $X \subset \mathbb{R}^m$ . O conjunto  $A = \{x \in X; f(x) < g(x)\}$  é aberto em  $X$  enquanto os conjuntos  $F = \{x \in X; f(x) \leq g(x)\}$  e  $G = \{x \in X; f(x) = g(x)\}$  são fechados em  $X$ .*

Em particular, tomando  $g$  constante, vemos que o conjunto dos pontos  $x \in X$  tais que  $f(x) < c$  é aberto em  $X$  enquanto as soluções  $x \in X$  da inequação  $f(x) \leq c$  ou da equação  $f(x) = c$  formam conjuntos fechados em  $X$ .

**Teorema 20.** *Sejam  $\varphi : K \rightarrow \mathbb{R}^n$  contínua no compacto  $K \subset \mathbb{R}^m$  e  $L = \varphi(K)$  a imagem (compacta) de  $\varphi$ . A fim de que uma aplicação  $f : L \rightarrow \mathbb{R}^p$  seja contínua, é necessário e suficiente que a composta  $f \circ \varphi : K \rightarrow \mathbb{R}^p$  seja contínua.*



*Demonstração.* Se  $f$  é contínua então  $f \circ \varphi$  é contínua, pelo Teorema 14. Reciprocamente, supondo  $f \circ \varphi$  contínua então, para todo conjunto fechado  $F \subset \mathbb{R}^p$ , a imagem inversa  $(f \circ \varphi)^{-1}(F) = \varphi^{-1}[f^{-1}(F)]$  é um subconjunto fechado de  $K$ , logo é compacto. Então, pelo Teorema 17,  $f^{-1}(F) = \varphi[\varphi^{-1}(f^{-1}(F))]$  é compacto, logo fechado em  $\mathbb{R}^m$ . Segue-se do corolário acima que  $f$  é contínua.  $\square$

**Observação.** Quando se tem uma aplicação arbitrária  $\varphi : K \rightarrow L$  entre dois conjuntos, para todo  $Z \subset L$  vale a inclusão  $\varphi[\varphi^{-1}(Z)] \subset Z$ . Entretanto, quando  $\varphi : K \rightarrow L$  é sobrejetiva, como no caso acima, tem-se  $\varphi[\varphi^{-1}(Z)] = Z$ .

**Exemplo 10.** Tomemos  $K = [0, 2\pi] \subset \mathbb{R}$ ,  $L = S^1 = \{(x, y) \in \mathbb{R}^2; x^2 + y^2 = 1\}$  e  $\varphi : [0, 2\pi] \rightarrow \mathbb{R}^2$  dada por  $\varphi(t) = (\cos t, \sin t)$ . Então  $[0, 2\pi]$  e  $S^1$  são compactos e  $\varphi : [0, 2\pi] \rightarrow S^1$  é contínua e sobrejetiva. Seja agora  $g : [0, 2\pi] \rightarrow \mathbb{R}^n$  uma aplicação contínua tal que  $g(0) = g(2\pi)$ . A partir de  $g$ , podemos definir  $f : S^1 \rightarrow \mathbb{R}^n$ , pondo  $f(\cos t, \sin t) = g(t)$ . Como  $g(0) = g(2\pi)$ ,  $f$  está bem definida. Além disso,  $f \circ \varphi = g$  é contínua. Segue-se do Teorema 20 que  $f$  é contínua. Isto se exprime dizendo que “para definir uma aplicação contínua no círculo  $S^1$  basta defini-la no intervalo  $[0, 2\pi]$  de modo que assuma valores iguais nos extremos 0 e  $2\pi$ .”  $\triangleleft$

## 8 Continuidade uniforme

A adição e a multiplicação de números reais são funções contínuas  $s, p : \mathbb{R}^2 \rightarrow \mathbb{R}$ , definidas por  $s(x, y) = x + y$  e  $p(x, y) = x \cdot y$ . Examinemos a continuidade de cada uma delas no ponto  $(a, b) \in \mathbb{R}^2$ . Para isso, usaremos em  $\mathbb{R}^2$  a norma do máximo, segundo a qual tem-se  $(x, y) \in B((a, b); \delta)$  se, e somente se,  $|x - a| < \delta$  e  $|y - b| < \delta$ .

Começemos com a adição: dado  $\varepsilon > 0$ , tomemos  $\delta = \varepsilon/2$ . Se  $|x - a| < \varepsilon/2$  e  $|y - b| < \varepsilon/2$ , isto é,  $(x, y) \in B((a, b), \delta)$ , então  $|s(x, y) - s(a, b)| = |x + y - (a + b)| \leq |x - a| + |y - b| < \varepsilon$ .

Em seguida, a multiplicação: dado  $\varepsilon > 0$ , temos  $xy - ab = (x - a)(y - b) + (x - a)b + a(y - b)$ , logo, tomando  $\delta > 0$  menor do que cada um dos números  $\frac{1}{3}\sqrt{\varepsilon}$ ,  $\frac{\varepsilon}{3|a|}$  e  $\frac{\varepsilon}{3|b|}$  veremos que se  $|x - a| < \delta$  e  $|y - b| < \delta$  isto é,  $(x, y) \in B((a, b), \delta)$ , então

$$\begin{aligned} |p(x, y) - p(a, b)| &= |xy - ab| \leq |x - a| |y - b| + |x - a| |b| + |a| |y - b| \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned}$$

Note-se a diferença: no caso da adição,  $\delta$  depende apenas de  $\varepsilon$ , mas não do ponto  $(a, b)$  onde a continuidade é testada. Já na multiplicação,  $\delta$  depende não apenas de  $\varepsilon$  mas também de  $(a, b)$ . Se um dos números  $a$  ou  $b$  aumentar, para o mesmo  $\varepsilon$  deve-se tomar  $\delta$  cada vez menor. Isto significa que a adição é uniformemente contínua mas a multiplicação não é. Segue-se a definição pertinente:

Uma aplicação  $f : X \rightarrow \mathbb{R}^n$  diz-se *uniformemente contínua* no conjunto  $X \subset \mathbb{R}^m$  quando, para todo  $\varepsilon > 0$ , for possível obter  $\delta > 0$  tal que  $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon$ , sejam quais forem  $x, y \in X$ .

**Teorema 21.** *A fim de que  $f : X \rightarrow \mathbb{R}^n$  seja uniformemente contínua no conjunto  $X \subset \mathbb{R}^m$  é necessário e suficiente que, para toda seqüência de pontos  $x_k, y_k \in X$  com  $\lim |x_k - y_k| = 0$ , se tenha  $\lim |f(x_k) - f(y_k)| = 0$ .*

**Teorema 22.** *Toda aplicação contínua  $f : X \rightarrow \mathbb{R}^n$ , definida num conjunto compacto  $X \subset \mathbb{R}^m$ , é uniformemente contínua.*

As demonstrações dos Teoremas 21 e 22 são exatamente as mesmas que se encontram nas páginas 83 e 84 do volume 1.

**Exemplo 11.** Uma aplicação  $f : X \rightarrow \mathbb{R}^n$ , definida no conjunto  $X \subset \mathbb{R}^m$ , chama-se *lipschitziana* quando existe  $c > 0$  tal que  $|f(x) - f(y)| \leq c|x - y|$  para quaisquer  $x, y \in X$ . O número  $c$  é chamado uma *constante de Lipschitz* de  $f$ . Toda aplicação lipschitziana é uniformemente contínua: dado  $\varepsilon > 0$ , basta tomar  $\delta = \varepsilon/c$ . A função  $f : [0, 1] \rightarrow \mathbb{R}$ , definida por  $f(x) = \sqrt{x}$ , é uniformemente contínua mas não é lipschitziana. Basta ver que

$$|\sqrt{x} - \sqrt{y}| = \frac{1}{\sqrt{x} + \sqrt{y}} |x - y|$$

e que, com  $x, y \in [0, 1]$  pode-se tornar  $\sqrt{x} + \sqrt{y}$  tão pequeno, (logo  $(\sqrt{x} + \sqrt{y})^{-1}$  tão grande) quanto se queira.  $\triangleleft$

**Exemplo 12.** Toda transformação linear  $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é contínua pois, para cada  $i = 1, 2, \dots, n$ , a  $i$ -ésima função-coordenada de  $A$  é a função contínua  $(x_1, \dots, x_m) \mapsto a_{i1}x_1 + \dots + a_{im}x_m$ , onde  $[a_{ij}]$  é a matriz de  $A$ . A esfera unitária  $S^{m-1} = \{x \in \mathbb{R}^m; |x| = 1\}$  é compacta. Logo  $A$  é limitada em  $S^{m-1}$ . O número

$$|A| = \sup \{ |A \cdot x|; x \in S^{m-1} \}$$

chama-se a *norma* da transformação  $A$ . Para todo vetor  $v \in \mathbb{R}^n$ , tem-se  $|A \cdot v| \leq |A| \cdot |v|$ . Isto é óbvio quando  $v = 0$ . Se  $v \neq 0$  então  $v/|v| \in S^{m-1}$  logo

$$|A \cdot v| = |v| \cdot \left| A \left( \frac{v}{|v|} \right) \right| \leq |A| |v|.$$

Para  $x, y \in \mathbb{R}^n$  quaisquer, tem-se  $|A \cdot x - Ay| = |A(x - y)| \leq |A| \cdot |x - y|$ . Logo a transformação linear  $A$  é uma aplicação lipschitziana, com constante de Lipschitz  $|A|$ .  $\triangleleft$

**Exemplo 13.** Dado  $A \subset \mathbb{R}^n$  não-vazio, seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  definida por  $f(x) = d(x, A)$ . Afirmamos que  $|d(x, A) - d(y, A)| \leq |x - y|$  para quaisquer  $x, y \in \mathbb{R}^n$ . Logo  $f$  é lipschitziana, com constante  $c = 1$ , donde uniformemente contínua. Para provar nossa afirmação, observemos que, dados  $x, y \in \mathbb{R}^n$ , existem  $\bar{a}, \bar{b} \in \bar{A}$  tais que  $d(x, A) = |x - \bar{a}|$  e  $d(y, A) = |y - \bar{b}|$ . (Vide seção 5.)

Temos  $\bar{b} = \lim y_k$ , com  $y_k \in A$ . Como  $|x - \bar{a}| \leq |y - y_k|$  para todo  $k \in \mathbb{N}$ , segue-se que  $|x - \bar{a}| \leq |x - \bar{b}|$ . Conseqüentemente,  $|d(x, A) - d(y, A)| = ||x - \bar{a}| - |y - \bar{b}|| \leq ||x - \bar{b}| - |y - \bar{b}|| \leq |x - y|$ , como queríamos mostrar.

Quando  $|f(x) - f(y)| \leq |x - y|$  para quaisquer  $x, y \in X$ , a aplicação lipschitziana  $f : X \rightarrow \mathbb{R}^n$  chama-se uma *contração fraca*. Se  $|f(x) - f(y)| \leq c|x - y|$  com  $0 < c < 1$ , a aplicação  $f$  chama-se uma *contração*, simplesmente.  $\triangleleft$

## 9 Homeomorfismos

Um *homeomorfismo* do conjunto  $X \subset \mathbb{R}^m$  sobre um conjunto  $Y \subset \mathbb{R}^n$  é uma bijeção contínua  $f : X \rightarrow Y$  cuja inversa  $f^{-1} : Y \rightarrow X$  também é contínua.

**Exemplo 14.** A aplicação  $f : [0, 2\pi) \rightarrow S^1$ , definida por  $f(t) = (\cos t, \sin t)$ , é uma bijeção contínua mas não é um homeomorfismo. Sua inversa  $f^{-1} : S^1 \rightarrow [0, 2\pi)$  aplica o compacto  $S^1$  sobre o intervalo  $[0, 2\pi)$ , que não é compacto, logo é descontínua. Mais precisamente,  $f^{-1}$  é descontínua no ponto  $a = (1, 0) = f(0) \in S^1$ . Com efeito se pusermos, para cada  $k \in \mathbb{N}$ ,  $t_k = (1 - 1/k) \cdot 2\pi$  e  $z_k = (\cos t_k, \sin t_k)$ , teremos  $\lim z_k = a$  mas  $\lim f^{-1}(z_k) = \lim t_k = 2\pi$ , logo não vale  $\lim f^{-1}(z_k) = f^{-1}(a) = 0$ .  $\triangleleft$

**Exemplo 15.** A bola aberta  $B = B(0; 1) \subset \mathbb{R}^n$  é homeomorfa ao espaço  $\mathbb{R}^n$ . De fato, as aplicações  $f : \mathbb{R}^n \rightarrow B$  e  $g : B \rightarrow \mathbb{R}^n$ , definidas por

$$f(x) = \frac{x}{1 + |x|} \quad \text{e} \quad g(y) = \frac{y}{1 - |y|}$$

são contínuas e, como se verifica sem dificuldade, vale  $g(f(x)) = x$ ,  $f(g(y)) = y$ , para quaisquer  $x \in \mathbb{R}^n$  e  $y \in B$ , logo  $g = f^{-1}$ .  $\triangleleft$

**Exemplo 16.** Sejam  $S^n = \{x \in \mathbb{R}^{n+1}; \langle x, x \rangle = 1\}$  a esfera unitária  $n$ -dimensional e  $N = (0, \dots, 0, 1) \in S^n$  seu pólo norte. A *projeção estereográfica*  $\xi : S^n - \{N\} \rightarrow \mathbb{R}^n$  é um importante exemplo de homeomorfismo. Para todo  $x \in S^n - \{N\}$ ,  $\xi(x)$  é o ponto em que a semi-reta  $\overrightarrow{Nx}$  corta o hiperplano  $x_{n+1} = 0$ , o qual identificamos com  $\mathbb{R}^n$ . Os pontos da semi-reta  $\overrightarrow{Nx}$  são da forma  $N + t(x - N)$  com  $t > 0$ . Um tal ponto está no hiperplano  $\mathbb{R}^n$  quando sua última coordenada  $1 + t(x_{n+1} - 1)$  é igual a zero, ou seja, quando  $t = 1/(1 - x_{n+1})$ . Logo  $\xi(x) = x'/(1 - x_{n+1})$ , onde  $x' = (x_1, \dots, x_n)$  para  $x = (x_1, \dots, x_n, x_{n+1})$ . Isto mostra que  $\xi : S^n - \{N\} \rightarrow \mathbb{R}^n$  é contínua. Seja agora  $\varphi : \mathbb{R}^n \rightarrow S^n - \{N\}$  dada por  $\varphi(y) = x$ , onde  $x' = 2y/(|y|^2 + 1)$  e  $x_{n+1} = (|y|^2 - 1)/(|y|^2 + 1)$ . Uma verificação simples mostra que  $\xi(\varphi(y)) = y$  para todo  $y \in \mathbb{R}^n$  e  $\varphi(\xi(x)) = x$  para todo  $x \in S^n$ . Portanto a aplicação contínua  $\varphi : \mathbb{R}^n \rightarrow S^n - \{N\}$  é a inversa de  $\xi$  e, conseqüentemente,  $\xi$  é um homeomorfismo.  $\triangleleft$

**Teorema 23.** Se  $K \subset \mathbb{R}^m$  é compacto então toda aplicação contínua injetiva  $f : K \rightarrow \mathbb{R}^n$  é um homeomorfismo sobre sua imagem (compacta)  $L = f(K)$ .

*Demonstração.* Chamemos de  $g : L \rightarrow K$  a inversa de  $f$ . Como  $L \subset \mathbb{R}^n$  é compacto, portanto fechado, pelo Teorema 19,  $g$  é contínua se, e somente se, para todo conjunto fechado  $F \subset \mathbb{R}^m$ , a imagem inversa  $g^{-1}(F) = \bar{g}^{-1}(F \cap K)$  é um fechado em  $\mathbb{R}^m$ . Mas  $F \cap K$  é compacto, logo  $g^{-1}(F \cap K) = f(F \cap K)$  é compacto (em virtude do Teorema) logo é fechado.  $\square$

O teorema acima mostra por que foi possível dar o Exemplo 14: o intervalo  $[0, 2\pi)$  não é compacto.

## 10 Conjuntos conexos

Uma *cisão* do conjunto  $X \subset \mathbb{R}^n$  é uma decomposição  $X = A \cup B$  onde  $\bar{A} \cap B = A \cap \bar{B} = \emptyset$ , isto é, nenhum ponto de  $A$  é aderente a  $B$  e nenhum ponto de  $B$  é aderente a  $A$ .

Um exemplo óbvio é a *cisão trivial*  $X = X \cup \emptyset$ . Já  $\mathbb{R} - \{0\} = (-\infty, 0) \cup (0, +\infty)$  é uma cisão não-trivial. Por outro lado, pondo  $A = (-\infty, 0]$  e  $B = (0, +\infty)$  a decomposição  $R = A \cup B$  não é uma cisão pois  $0 \in A \cap \bar{B}$ .

Se  $X = A \cup B$  é uma cisão então os pontos de  $X$  que são aderentes a  $A$ , não pertencendo a  $B$ , estão em  $A$ , logo  $A = \bar{A} \cap X$ . Analogamente,  $B = \bar{B} \cap X$ . Assim,  $A$  e  $B$  são ambos fechados em  $X$ . Com  $A = X - B$  e  $B = X - A$ , segue-se que  $A$  e  $B$  são também abertos em  $X$ .

Reciprocamente, se  $A \subset X$  é aberto e fechado em  $X$  então, pondo  $B = X - A$ , a decomposição  $X = A \cup B$  é uma cisão. Com efeito, nenhum ponto de  $X$  aderente a  $A$  pode pertencer a  $B$  pois  $A$  é fechado em  $X$  e, da mesma forma, nenhum ponto de  $X$  aderente a  $B$  pode pertencer a  $A$ .

Em particular, se  $X \subset \mathbb{R}^n$  é aberto, uma cisão  $X = A \cup B$  é uma expressão de  $X$  como reunião de dois abertos disjuntos. E se  $X \subset \mathbb{R}^n$  é fechado, toda cisão  $X = A \cup B$  é a expressão de  $X$  como reunião de dois conjuntos fechados disjuntos. Mais particularmente ainda, se  $X$  é compacto então  $A$  e  $B$  são compactos.

**Exemplo 17.** Escrevendo as linhas de uma matriz, uma após a outra, numa só lista, identificaremos o espaço  $\mathbb{R}^{n^2}$  com o conjunto das matrizes quadradas  $n \times n$ . Sejam  $G_n$ ,  $G_+$  e  $G_-$  respectivamente os conjuntos das matrizes com determinante  $\neq 0$ , das matrizes com determinante  $> 0$  e com determinante  $< 0$ . A igualdade  $G_n = G_+ \cup G_-$  é uma cisão. Com efeito, como o determinante é uma função real contínua  $\det : G_n \rightarrow \mathbb{R}$ , uma seqüência de matrizes com determinantes positivos não pode convergir para uma matriz de determinante negativo. Assim  $\overline{G_+} \cap G_- = \emptyset$ . Analogamente,  $G_+ \cap \overline{G_-} = \emptyset$ .

Um conjunto  $X \subset \mathbb{R}^n$  chama-se *conexo* quando só admite a cisão trivial. Caso contrário, diz-se que  $X$  é *desconexo*.

Como vimos no Exemplo 17 acima, o conjunto das matrizes  $n \times n$  com determinante  $\neq 0$  é desconexo.

Na página 51 do vol. 1 foi provado que todo intervalo da reta  $\mathbb{R}$  (seja ele aberto ou não, limitado ou não) é conexo.

Vale a recíproca: ◁

**Teorema 24.** *Os únicos subconjuntos conexos de  $\mathbb{R}$  são os intervalos.*

*Demonstração.* Suponha que  $X \subset \mathbb{R}$  não seja um intervalo. Então existem  $a < c < b$  tais que  $a, b \in X$  e  $c \notin X$ . Neste caso, pondo  $A = \{x \in X; x < c\}$  e  $B = \{x \in X; x > c\}$ , vemos que  $X = A \cup B$  é uma cisão. Como  $a \in A$  e  $b \in B$ , esta cisão não é trivial. Portanto  $X$  é desconexo. □

**Teorema 25.** (a) *A imagem do conjunto conexo  $X \subset \mathbb{R}^m$  por uma aplicação contínua  $f : X \rightarrow \mathbb{R}^n$  é um conjunto conexo.*

(b) *A reunião  $X = \bigcup_{\lambda \in L} X_\lambda$  de uma família qualquer de conjuntos conexos  $X_\lambda \subset \mathbb{R}^n$  que têm um ponto  $a$  em comum é um conjunto conexo.*

(c) O produto cartesiano  $X \times Y \subset \mathbb{R}^{m+n}$  dos conjuntos  $X \subset \mathbb{R}^m$  e  $Y \subset \mathbb{R}^n$  é um conjunto conexo se, e somente se,  $X$  e  $Y$  são conexos.

(d) O fecho de um conjunto conexo é conexo.

*Demonstração.* (a) Se  $f(X) = A \cup B$  é uma cisão da imagem de  $X$  então  $A$  e  $B$  são ambos abertos e fechados em  $f(X)$ , além de disjuntos. Logo  $f^{-1}(A)$  e  $f^{-1}(B)$  são também disjuntos, abertos e fechados em  $X$ , portanto  $X = f^{-1}(A) \cup f^{-1}(B)$  é uma cisão, a qual é trivial pois  $X$  é conexo. Mas  $A = ff^{-1}(A)$  e  $B = ff^{-1}(B)$  porque  $A$  e  $B$  estão contidos em  $f(X)$ . Assim,  $A$  ou  $B$  é vazio e daí a cisão  $f(X) = A \cup B$  é trivial. Então  $f(X)$  é conexo.

(b) Seja  $a$  tal que  $a \in X_\lambda$  para todo  $\lambda \in L$ . Se  $X = A \cup B$  é uma cisão então o ponto  $a$  pertence a um dos conjuntos,  $A$  ou  $B$ . Digamos que  $a \in A$ . Para todo  $\lambda \in L$ ,  $X_\lambda = (A \cap X_\lambda) \cup (B \cap X_\lambda)$  é uma cisão, qual é trivial pois  $X_\lambda$  é conexo. Como  $a \in A \cap X_\lambda$ , segue-se que  $B \cap X_\lambda$  é vazio. Logo  $B = \bigcup_{\lambda} (B \cap X_\lambda)$  é vazio e a cisão  $X = A \cup B$  é trivial. Portanto  $X$  é conexo.

(c) Se  $X \times Y$  é conexo então  $X$  e  $Y$  são conexos porque são as imagens de  $X \times Y$  pelas projeções  $p : X \times Y \rightarrow X$ ,  $p(x, y) = x$  e  $q : X \times Y \rightarrow Y$ ,  $q(x, y) = y$  as quais são contínuas. Reciprocamente, se  $X$  e  $Y$  são conexos, tomamos um ponto  $c = (a, b) \in X \times Y$ . Para cada  $z = (x, y) \in X \times Y$  o conjunto  $C_z = (X \times \{b\}) \cup (\{x\} \times Y)$  é conexo pois é reunião dos conjuntos conexos  $X \times \{b\}$  e  $\{x\} \times Y$  (homeomorfos respectivamente a  $X$  e  $Y$ ) com o ponto  $(x, b)$  em comum. Além disso, também  $c = (a, b) \in C_z$  para todo  $z \in X \times Y$  e  $X \times Y = \bigcup_z C_z$  logo, pelo item (b),  $X \times Y$  é conexo.

(d) Seja  $\bar{X} = A \cup B$  uma cisão. Então  $X = (A \cap X) \cup (B \cap X)$  também é uma cisão pois  $(A \cap X) \cap (B \cap X) \subset \bar{A} \cap B = \emptyset$  e  $(A \cap X) \cap (\bar{B} \cap X) \subset A \cap \bar{B} = \emptyset$ . Como  $X$  é conexo, tem-se, digamos,  $A \cap X = \emptyset$ . Ora, existe  $U \subset \mathbb{R}^n$  aberto tal que  $A = U \cap \bar{X}$ . Daí  $A \cap X = (U \cap \bar{X}) \cap X = U \cap X$ , logo  $U \cap X = \emptyset$ . Sendo  $U$  aberto, de  $U \cap X = \emptyset$  segue-se que nenhum ponto de  $U$  é aderente a  $X$ . Então  $U \cap \bar{X} = \emptyset$ , ou seja,  $A = \emptyset$ .

Assim, toda cisão  $\bar{X} = A \cup B$  é trivial, portanto  $\bar{X}$  é conexo.  $\square$

**Corolário 6.** Se  $X_1, \dots, X_k$  são conexos então  $X_1 \times \dots \times X_k$  é conexo. Em particular,  $\mathbb{R}^n = \mathbb{R} \times \dots \times \mathbb{R}$  é conexo.

Com efeito,  $X_1 \times X_2 \times X_3 = (X_1 \times X_2) \times X_3$  e assim por diante.

**Corolário 7.** Se  $X \subset \mathbb{R}^n$  é conexo então a imagem de toda função real contínua  $f : X \rightarrow \mathbb{R}$  é um intervalo.

Com efeito, pelo Teorema 24 todo subconjunto conexo de  $\mathbb{R}$  é um intervalo.

Este corolário é conhecido como o Teorema do Valor Intermediário pois pode também ser enunciado assim: “Sejam  $X \subset \mathbb{R}^n$  conexo e  $f : X \rightarrow \mathbb{R}$  contínua. Se  $a, b \in X$  são tais que  $f(a) < f(b)$  então, para cada  $d$  com  $f(a) < d < f(b)$ , existe  $c \in X$  tal que  $f(c) = d$ .”

**Corolário 8 (“Teorema da Alfândega”).** *Seja  $X \subset \mathbb{R}^n$  um conjunto arbitrário. Se um conjunto conexo  $C \subset \mathbb{R}^n$  contém um ponto  $a \in X$  e um ponto  $b \notin X$  então  $C$  contém um ponto  $c \in \text{fr.}X$ .*

Com efeito, a função contínua  $f : C \rightarrow \mathbb{R}$ , definida por  $f(x) = d(x, X) - d(x, \mathbb{R}^n - X)$ , é tal que  $f(a) \leq 0$  e  $f(b) \geq 0$ . Logo, pelo Teorema do Valor Intermediário, deve existir  $c \in C$  tal que  $f(c) = 0$ , isto é,  $d(c, X) = d(c, \mathbb{R}^n - X)$ . Como um desses dois números é zero, ambos o são e daí  $c \in \text{fr.}X$ .

Como  $\mathbb{R}^n$  é conexo, resulta do corolário acima que se o conjunto  $X \subset \mathbb{R}^n$  não é vazio nem coincide com  $\mathbb{R}^n$  então a fronteira de  $X$  não é vazia. De fato, se  $X \neq \emptyset$  e  $X \neq \mathbb{R}^n$  então o conjunto conexo  $\mathbb{R}^n$  contém algum ponto de  $X$  e algum ponto que não pertence a  $X$ , logo contém algum ponto da fronteira de  $X$ .

**Exemplo 18.** Para todo  $n \in \mathbb{N}$ , a esfera  $S^n$  é um conjunto conexo. Com efeito, retirando o pólo norte  $N = (0, \dots, 0, 1)$ , vemos que  $X = S^n - \{N\}$  é conexo por ser homeomorfo a  $\mathbb{R}^n$  (cfr. Exemplo 16). Como  $S^n = \bar{X}$ , segue-se do item (d) que a esfera  $S^n$  é conexa.  $\triangleleft$

**Exemplo 19.** Uma consequência do Teorema do Valor Intermediário é que para toda função real contínua  $f : S^1 \rightarrow \mathbb{R}$  existe (pelo menos) um ponto  $z \in S^1$  tal que  $f(z) = f(-z)$ . Para ver isto, consideremos a função contínua  $\varphi : S^1 \rightarrow \mathbb{R}$ , dada por  $\varphi(z) = f(z) - f(-z)$ . Vale  $\varphi(-z) = -\varphi(z)$ . Assim, ou  $\varphi(z) = 0$  para todo  $z$  (assunto encerrado) ou existe  $a \in S^1$  com  $\varphi(-a) < 0 < \varphi(a)$ , logo  $\varphi(z) = 0$  para algum  $z \in S^1$ , pois  $S^1$  é conexo.  $\triangleleft$

Existe uma noção bem geométrica que fornece uma condição suficiente para a conexidade de um conjunto, que é a conexidade por caminhos.

Um *caminho* num conjunto  $X \subset \mathbb{R}^n$  é uma aplicação contínua  $f : I \rightarrow X$ , definida num intervalo  $I$ .

Por exemplo, dados  $x, y \in \mathbb{R}^n$ , o caminho  $f : [0, 1] \rightarrow \mathbb{R}^n$ , definido por  $f(t) = (1 - t)x + ty$ , chama-se o *caminho retilíneo* que liga  $x$  a  $y$ . Às vezes nos referiremos a ele como o caminho  $[x, y]$ .

Diremos que os pontos  $a, b \in X$  podem ser ligados por um caminho em  $X$  quando existe um caminho  $f : I \rightarrow X$  tal  $a = f(\alpha)$ ,  $b = f(\beta)$  com  $\alpha < \beta \in I$ .

Por exemplo, se  $X \subset \mathbb{R}^n$  é convexo, dois pontos quaisquer  $a, b \in X$  podem ser ligados por um caminho em  $X$ , a saber, o caminho retilíneo  $[a, b]$ .

Se  $a, b \in X$  podem ser ligados por um caminho  $f : I \rightarrow X$ , então existe um caminho  $\varphi : [0, 1] \rightarrow X$  tal que  $\varphi(0) = a$  e  $\varphi(1) = b$ . Basta pôr  $\varphi(t) = f((1-t)\alpha + t\beta)$ , onde  $a = f(\alpha)$  e  $b = f(\beta)$ .

Se  $f, g : [0, 1] \rightarrow X$  são caminhos em  $X$ , com  $f(1) = g(0)$ , então definimos o caminho justaposto  $h = f \vee g : [0, 1] \rightarrow X$  pondo  $h(t) = f(2t)$  se  $0 \leq t \leq 1/2$  e  $h(t) = g(2t - 1)$  se  $1/2 \leq t \leq 1$ . Note que estas duas expressões definem o mesmo valor de  $h(1/2)$ . Como  $h|_{[0, 1/2]}$  e  $h|_{[1/2, 1]}$  são contínuas, segue-se que  $h$  é contínua. Intuitivamente, o caminho  $h$  percorre a trajetória de  $f$  (com velocidade dobrada) até  $t = 1/2$  e depois, para  $t \geq 1/2$ , descreve (ainda com velocidade dobrada) o percurso de  $g$ .

Sejam  $a, b, c$  pontos do conjunto  $X \subset \mathbb{R}^n$ . Se  $a, b$  podem ser ligados por um caminho em  $X$  e  $b, c$  também podem ser ligados por um caminho em  $X$ , então existe um caminho em  $X$  ligando  $a$  a  $c$ . Basta tomar caminhos  $f, g : [0, 1] \rightarrow X$  com  $f(0) = a, f(1) = b, g(0) = b, g(1) = c$  e pôr  $h = f \vee g$ . Então  $h(0) = a, h(1) = c$ .

Um conjunto  $X \subset \mathbb{R}^n$  diz-se *conexo por caminhos* quando dois pontos quaisquer  $a, b \in X$  podem ser ligados por um caminho em  $X$ .

Todo conjunto convexo  $X \subset \mathbb{R}^n$  é conexo por caminhos. Em particular, toda bola (aberta ou fechada) no espaço euclidiano é conexa por caminhos.

A esfera  $S^n = \{x \in \mathbb{R}^{n+1}; \langle x, x \rangle = 1\}$  é conexa por caminhos. Com efeito, dados  $a, b \in S^n$ , se  $a$  e  $b$  não são antípodas, isto é, se  $b \neq -a$ , então  $f : [0, 1] \rightarrow S^n$ , definida por

$$f(t) = \frac{(1-t)a + tb}{|(1-t)a + tb|}.$$

é contínua (pois seu denominador nunca se anula), com  $f(0) = a, f(1) = b$ . Se, porém,  $b = -a$ , tomamos um ponto  $c \in S^n - \{a, b\}$ , ligamos  $a$  com  $c$  e  $c$  com  $b$  pelo processo acima. O caminho justaposto ligará o ponto  $a$  ao seu antípoda  $b$ .

Todo conjunto  $X \subset \mathbb{R}^n$ , conexo por caminhos, é conexo.

Com efeito, fixando  $a \in X$  seja, para cada  $x \in X$ ,  $C_x$  a imagem de um caminho em  $X$  ligando  $a$  até  $x$ . Pelo item (a) do Teorema 25,  $C_x$  é um conjunto conexo que contém  $a$  e  $x$ . Logo, pelo item (b) do mesmo teorema, o conjunto  $X = \bigcup_{x \in X} C_x$  é conexo.

A recíproca é falsa. O conjunto  $X_0 \subset \mathbb{R}^2$ , reunião do gráfico da função  $f(x) = \sin(1/x), 0 < x \leq 1$ , com a origem  $p = (0, 0)$ , é conexo mas não é conexo por caminhos. (Para a demonstração, ver o livro “Espaços Métricos”, do autor, página 103.)

Há, porém, um caso particular importante, no qual a conexidade implica em conexidade por caminhos: quando o conjunto  $X \subset \mathbb{R}^n$  é aberto.

Diremos que  $f : [0, 1] \rightarrow X$  é um *caminho poligonal* em  $X$  quando  $f$  é a justaposição de um número finito de caminhos retilíneos.

**Teorema 26.** *Um aberto  $A \subset \mathbb{R}^n$  é conexo se, e somente se, é conexo por caminhos.*

*Demonstração.* Seja  $A \subset \mathbb{R}^n$  aberto e conexo. Fixemos um ponto  $a \in A$  e consideremos o conjunto  $U$ , formado pelos pontos  $x \in A$  que podem ser ligados ao ponto  $a$  por um caminho poligonal contido em  $A$ . Afirmamos que  $U$  é aberto. Com efeito, seja  $x \in U$ . Sendo  $A$  aberto, existe  $B = B(x; r)$ , com  $x \in B \subset A$ . Como a bola  $B$  é convexa, todo ponto  $y \in B$  pode ser ligado a  $x$  por um segmento de reta contido em  $B$ , logo  $y$  se liga a  $a$  por um caminho poligonal contido em  $A$ . Portanto  $B \subset U$  e  $U \subset A$  é aberto. Também  $V = A - U$  é aberto, pois se  $v \in V$  então  $v$  não pode ser ligado a  $a$  por um caminho poligonal contido em  $A$ . Tomando uma bola aberta  $B_1$ , com  $v \in B_1 \subset A$ , todo  $z \in B_1$  se liga a  $v$  por um segmento de reta contido em  $B_1$ . Se  $z$  pudesse ser ligado a  $a$  por um caminho poligonal contido em  $A$ , justapondo-se  $[v, z]$  a esse caminho, veríamos que  $v \in U$ , um absurdo. Temos então  $A = U \cup V$ , uma cisão. Como  $A$  é conexo e  $a \in U$ , temos  $V = \emptyset$ , donde  $A = U$ , o que prova o teorema.  $\square$

**Corolário 9 (da demonstração).** *Se  $A \subset \mathbb{R}^n$  é aberto e conexo, dois pontos quaisquer de  $A$  podem ser ligados por um caminho poligonal contido em  $A$ .*

Mostraremos a seguir que todo conjunto  $X \subset \mathbb{R}^n$  se exprime como reunião disjunta de subconjuntos conexos máximos, chamados componentes conexas de  $X$ .

Sejam  $x \in X \subset \mathbb{R}^n$ . A *componente conexa* do ponto  $x$  no conjunto  $X$  é a reunião  $C_x$  de todos os subconjuntos conexos de  $X$  que contêm o ponto  $x$ .

Por exemplo, se  $X = \mathbb{Q} \subset \mathbb{R}$  então a componente conexa de qualquer ponto  $x \in X$  é  $\{x\}$ . Por outro lado, se  $X \subset \mathbb{R}^n$  é conexo então, para todo  $x \in X$  temos  $C_x = X$ . Se  $X = \mathbb{R} - \{0\}$  então a componente conexa de 1 em  $X$  é  $(0, +\infty)$  enquanto que a componente conexa de  $-1$  é  $(-\infty, 0)$ .

Dados  $x \in X \subset \mathbb{R}^n$ , a componente conexa  $C_x$  é um conjunto conexo, pelo Teorema 25(b). Na realidade,  $C_x$  é o maior subconjunto conexo de  $X$  contendo o ponto  $x$ . Com efeito, se  $C \subset X$  é conexo e contém  $x$ , então  $C$  é um dos conjuntos cuja reunião é  $C_x$ , logo  $C \subset C_x$ . Mais ainda, se  $C \subset X$  é conexo e tem algum ponto em comum com  $C_x$  então  $C \subset C_x$ , pois  $C \cup C_x$  é conexo contendo  $x$  logo  $C \cup C_x \subset C_x$  e daí  $C \subset C_x$ . Em particular, nenhum subconjunto conexo de  $X$  pode conter  $C_x$  propriamente.

Sejam  $x, y$  dois pontos de  $X$ . Suas componentes conexas  $C_x$  e  $C_y$  ou coincidem ou são disjuntas pois se  $z \in C_x \cap C_y$  então  $C_x \subset C_y$  e  $C_y \subset C_x$ . Assim a relação “ $x$  e  $y$  pertencem à mesma componente conexa em  $X$ ” é uma equivalência no

conjunto  $X$ . As classes de equivalência são as componentes conexas dos pontos de  $X$ .

Toda componente conexa  $C_x$  é um conjunto fechado em  $X$ . Com efeito, sendo  $C_x \subset \bar{C}_x \cap X \subset \bar{C}_x$ , o Teorema 25(d) nos assegura que  $\bar{C}_x \cap X$  é um subconjunto conexo de  $X$ , contendo  $C_x$ . Logo  $\bar{C}_x \cap X = C_x$ , o que mostra que  $C_x$  é fechado em  $X$ .

## 11 Limites

Sejam  $f : X \rightarrow \mathbb{R}^n$  definida no conjunto  $X \subset \mathbb{R}^m$  e  $a \in \mathbb{R}^n$  um ponto de acumulação de  $X$ . Diz-se que  $b \in \mathbb{R}^n$  é o limite de  $f(x)$  quando  $x$  tende para  $a$  e escreve-se  $\lim_{x \rightarrow a} f(x) = b$  quando a seguinte condição é válida:

“para todo  $\varepsilon > 0$  dado, existe  $\delta > 0$  tal que  $x \in X$  e  $0 < |x - a| < \delta$  implicam  $|f(x) - b| < \varepsilon$ .”

O ponto  $a$  pode pertencer ou não a  $X$ . Em muitos dos exemplos mais importantes de limite, na verdade, tem-se  $a \notin X$ . Mas, mesmo que pertença a  $X$ , o ponto  $a$  e o valor  $f(a)$  não desempenham papel algum na definição de limite.

Quando o ponto de acumulação  $a$  pertence a  $X$ , a aplicação  $f : X \rightarrow \mathbb{R}^n$  é contínua no ponto  $a$  se, e somente se,  $\lim_{x \rightarrow a} f(x) = f(a)$ .

A propriedade seguinte decorre imediatamente da definição mas é útil o bastante para ser destacada como um teorema.

**Teorema 27 (Permanência do sinal).** *Sejam  $a$  um ponto de acumulação de  $X \subset \mathbb{R}^n$  e  $f : X \rightarrow \mathbb{R}$  uma função real. Se  $b = \lim_{x \rightarrow a} f(x)$  é um número positivo então existe  $\delta > 0$  tal que  $x \in X$  e  $0 < |x - a| < \delta$  implicam  $f(x) > 0$ .*

*Demonstração.* Como  $b$  é positivo, tomamos  $\varepsilon = b$ . Pela definição de limite, existe  $\delta > 0$  tal que  $x \in X$  e  $0 < |x - a| < \delta$  implicam  $b - \varepsilon < f(x) < b + \varepsilon$ , isto é,  $0 < f(x) < 2b$ , logo  $f(x) > 0$ .  $\square$

Quando  $X$  é um intervalo da reta, tem sentido a noção de *limite lateral* de uma aplicação  $f : I \rightarrow \mathbb{R}^n$ , ou seja, de um caminho, num ponto  $a \in I$ . Por exemplo, se  $a$  não é o extremo superior de  $I$ , diz-se que  $b \in \mathbb{R}^n$  é o *limite à direita* de  $f(t)$  quando  $t$  tende para  $a$ , e escreve-se  $\lim_{t \rightarrow a^+} f(t) = b$ , para significar que

“para todo  $\varepsilon > 0$  dado, existe  $\delta > 0$  tal que  $a < t < a + \delta$  implica  $|f(t) - b| < \varepsilon$ .”

Analogamente se define o limite à esquerda  $\lim_{t \rightarrow a^-} f(t)$ .

Assim como a continuidade de uma aplicação, a existência ao valor do limite se exprimem em termos das funções-coordenada, como veremos agora.

**Teorema 28.** *Seja  $a$  um ponto de acumulação do conjunto  $X \subset \mathbb{R}^m$ . Se as funções-coordenada da aplicação  $f : X \rightarrow \mathbb{R}^n$  são  $f_1, \dots, f_n : X \rightarrow \mathbb{R}$  então tem-se  $\lim_{x \rightarrow a} f(x) = b = (b_1, \dots, b_n)$  se, e somente se,  $\lim_{x \rightarrow a} f_i(x) = b_i$  para cada  $i = 1, \dots, n$ .*

*Demonstração.* Se  $\lim_{x \rightarrow a} f(x) = b$  então, para cada  $i = 1, \dots, n$ , tem-se  $\lim_{x \rightarrow a} f_i(x) = b_i$  porque  $|f_i(x) - b_i| \leq |f(x) - b|$ . Reciprocamente, se  $\lim_{x \rightarrow a} f_i(x) = b_i$  para cada  $i = 1, \dots, n$  então  $\lim_{x \rightarrow a} f(x) = b$  porque  $|f(x) - b| \leq \sum_{i=1}^n |f_i(x) - b_i|$ .  $\square$

A proposição seguinte relaciona o limite de aplicações com o limite de seqüências.

**Teorema 29.** *Seja  $a$  um ponto de acumulação do conjunto  $X \subset \mathbb{R}^m$ . A fim de que se tenha  $\lim_{x \rightarrow a} f(x) = b$  é necessário e suficiente que, para toda seqüência de pontos  $x_k \in X - \{a\}$  com  $\lim x_k = a$ , seja  $\lim f(x_k) = b$ .*

A demonstração é idêntica à feita no vol. 1 (pág. 63).

**Teorema 30.** *Sejam:  $a$  um ponto de acumulação de  $X \subset \mathbb{R}^m$ ,  $b \in Y$  um ponto de acumulação de  $Y \subset \mathbb{R}^n$ ,  $f : X \rightarrow Y$  uma aplicação tal que  $\lim_{x \rightarrow a} f(x) = b$  e  $g : \mathbb{R}^p$  contínua no ponto  $b$ . Então  $\lim_{x \rightarrow a} g(f(x)) = g(b)$ .*

Isto é mais fácil de provar do que enunciar. Basta imitar a demonstração de que a composta de duas aplicações contínuas é contínua (Teorema 14).

**Teorema 31.** *Sejam  $f, g : X \rightarrow \mathbb{R}^n$  e  $\alpha : X \rightarrow \mathbb{R}$  definidas no conjunto  $X \subset \mathbb{R}^m$  e  $a$  um ponto de acumulação de  $X$ . Se existem  $\lim_{x \rightarrow a} f(x) = b$ ,  $\lim_{x \rightarrow a} g(x) = c$  e  $\lim_{x \rightarrow a} \alpha(x) = \alpha_0$ , então existem os limites e valem as igualdades abaixo:*

$$\begin{aligned} \lim_{x \rightarrow a} [f(x) + g(x)] &= b + c, & \lim_{x \rightarrow a} \alpha(x) \cdot f(x) &= \alpha_0 \cdot b \\ \lim_{x \rightarrow a} \langle f(x), g(x) \rangle &= \langle b, c \rangle, & \lim_{x \rightarrow a} |f(x)| &= |b|. \end{aligned}$$

*Demonstração.* A aplicação  $s : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ , definida por  $s(x, y) = x + y$ , é contínua. Observando que  $f(x) + g(x) = s(f(x), g(x))$ , resulta do Teorema 31 que  $\lim_{x \rightarrow a} [f(x) + g(x)] = \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x) = b + c$ . Analogamente para as outras três igualdades.  $\square$

Além disso, é útil saber que se  $\lim_{x \rightarrow a} \alpha(x) = 0$  e  $f : X \rightarrow \mathbb{R}^n$  é limitada na vizinhança de  $a$  (isto é, existem  $\delta > 0$  e  $M > 0$  tais que  $x \in X$  e  $|x - a| < \delta$  implicam  $|f(x)| \leq M$ ) então  $\lim_{x \rightarrow a} \alpha(x)f(x) = 0$ , mesmo que não exista  $\lim_{x \rightarrow a} f(x)$ . (Muito fácil.)

**Exemplo 20.** Seja  $g : \mathbb{R}^2 - \{0\} \rightarrow \mathbb{R}$  definida por  $g(x, y) = x^2y/(x^2 + y^2)$ . Então podemos escrever  $g(x, y) = \alpha(x, y) \cdot f(x, y)$  onde  $\alpha(x, y) = x$  e

$$f(x, y) = \frac{xy}{x^2 + y^2} = \frac{x}{\sqrt{x^2 + y^2}} \cdot \frac{y}{\sqrt{x^2 + y^2}} = \cos \theta \operatorname{sen} \theta,$$

sendo  $\theta$  o ângulo de eixo  $OX$  com o segmento  $Oz$ ,  $z = (x, y)$ . Assim, temos

$$\lim_{(x,y) \rightarrow (0,0)} \alpha(x, y) = 0$$

e  $|f(x, y)| \leq 1$ , logo  $\lim_{(x,y) \rightarrow (0,0)} g(x, y) = 0$ . ◁

Agora que já vimos ser  $\lim_{x \rightarrow a} (f(x) - g(x)) = \lim_{x \rightarrow a} f(x) - \lim_{x \rightarrow a} g(x)$ , podemos demonstrar a seguinte consequência do Teorema 27:

**Teorema 32 (Permanência da desigualdade).** *Sejam  $f, g : X \rightarrow \mathbb{R}$  definidas no conjunto  $X \subset \mathbb{R}^m$  e  $a$  um ponto de acumulação de  $X$ . Se  $f(x) \leq g(x)$  para todo  $x \in X$  e existem  $\lim_{x \rightarrow a} f(x)$  e  $\lim_{x \rightarrow a} g(x)$  então tem-se  $\lim_{x \rightarrow a} f(x) \leq \lim_{x \rightarrow a} g(x)$ .*

*Demonstração.* Se fosse o contrário,  $\lim_{x \rightarrow a} f(x) > \lim_{x \rightarrow a} g(x)$ , teríamos  $\lim_{x \rightarrow a} (f(x) - g(x)) > 0$  e então, pelo Teorema 27, valeria  $f(x) > g(x)$  para todo  $x \in X$  suficientemente próximo de  $a$ , uma contradição. ◻

## Capítulo 2

# Caminhos em $\mathbb{R}^n$

### 1 Caminhos diferenciáveis

Seja  $f : I \rightarrow \mathbb{R}^n$  um caminho, isto é, uma aplicação contínua cujo domínio é um intervalo da reta. Para todo  $t \in I$ , tem-se  $f(t) = (f_1(t), \dots, f_n(t))$ , onde  $f_1, \dots, f_n : I \rightarrow \mathbb{R}$ , as funções-coordenada de  $f$ , são contínuas.

Diz-se que o caminho  $f : I \rightarrow \mathbb{R}^n$  é *diferenciável no ponto*  $t_0 \in I$  quando existe o limite

$$f'(t_0) = \lim_{h \rightarrow 0} \frac{f(t_0 + h) - f(t_0)}{h},$$

chamado a *derivada*, ou o *vetor-velocidade* de  $f$  no ponto  $t_0$ .

Para todo  $h \neq 0$ , as coordenadas do vetor  $[f(t_0 + h) - f(t_0)]/h$  são os números  $[f_i(t_0 + h) - f_i(t_0)]/h$  ( $i = 1, \dots, n$ ). Pelo Teorema 28 do Capítulo 1, o caminho  $f$  é diferenciável no ponto  $t_0$  se, e somente se, suas funções-coordenada o são. No caso afirmativo, tem-se  $f'(t_0) = (f'_1(t_0), \dots, f'_n(t_0))$ . Às vezes se usa também a notação  $\frac{df}{dt}(t_0)$  em vez de  $f'(t_0)$ .

Quando o caminho  $f : I \rightarrow \mathbb{R}^n$  é diferenciável em todos os pontos de  $I$ , diz-se que ele é *diferenciável em*  $I$ . Neste caso, a correspondência  $t \mapsto f'(t)$  define uma aplicação  $f' : I \rightarrow \mathbb{R}^n$ . Quando  $f'$  é contínua, o caminho  $f$  chama-se de *classe*  $C^1$ . Mais geralmente, para todo inteiro  $k > 1$ , diz-se que  $f : I \rightarrow \mathbb{R}^n$  é um caminho *de classe*  $C^k$  quando ele é diferenciável e  $f'$  é de classe  $C^{k-1}$ . Para que  $f$  seja de classe  $C^k$  é necessário e suficiente que cada uma de suas funções-coordenada o seja. Escreve-se então  $f \in C^k$ .

No caso em que  $f'(t_0) \neq 0$ , a definição acima significa que a reta que passa pelo ponto  $f(t_0)$  e tem a direção dada pelo vetor  $f'(t_0)$ , isto é, o conjunto  $\{f(t_0) + \alpha \cdot f'(t_0); \alpha \in \mathbb{R}\}$ , é o limite da secante que passa pelos pontos  $f(t_0)$  e  $f(t_0 + h)$  quando  $h \rightarrow 0$ . Logo é natural chamá-la de *reta tangente* ao caminho  $f$  no ponto

$t$ . Quando  $f'(t_0) = 0$  pode não haver reta alguma que se possa chamar de tangente do ponto  $f(t_0)$ .

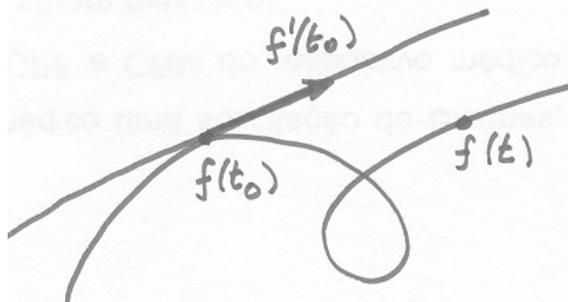


Figura 1.

**Exemplo 1.** Dados  $a \neq b$  em  $\mathbb{R}^n$ , seja  $f : \mathbb{R} \rightarrow \mathbb{R}^n$  o caminho retilíneo que passa pelos pontos  $a$  e  $b$ :  $f(t) = (1-t)a + t \cdot b$ . Para todo  $t \in \mathbb{R}$ ,  $f$  é diferenciável no ponto  $t$ , com  $f'(t) = b - a$ , como se vê diretamente a partir da definição.

Se  $t_0$  não é o extremo superior do intervalo  $I$ , tem sentido considerar a *derivada à direita* do caminho  $f : I \rightarrow \mathbb{R}^n$  no ponto  $t_0$ , a qual é definida por

$$f'_+(t_0) = \lim_{h \rightarrow 0^+} \frac{f(t_0 + h) - f(t_0)}{h},$$

e, de modo análogo, a *derivada à esquerda*  $f'(t_0-)$ , caso  $t_0$  não seja o extremo inferior de  $I$ . Quando  $t_0$  é um ponto interior de  $I$  então  $f$  é diferenciável no ponto  $t_0$  se, e somente se, existem as derivadas  $f'_+(t_0)$  e  $f'_-(t_0)$  sendo elas iguais.  $\triangleleft$

**Exemplo 2.** Seja  $f : \mathbb{R} \rightarrow \mathbb{R}^2$  o caminho definido por  $f(t) = (t, |t|)$ . Para  $t > 0$  tem-se  $f(t) = (t, t)$  e, para  $t < 0$ ,  $f(t) = (t, -t)$ . Logo, para todo  $t \neq 0$  existe  $f'(t)$ , sendo  $f'(t) = (1, 1)$  se  $t > 0$  e  $f'(t) = (1, -1)$  se  $t < 0$ . No ponto  $t = 0$  existem as derivadas laterais  $f'_+(0) = (1, 1)$  e  $f'_-(0) = (1, -1)$ , que são diferentes, logo  $f$  não é diferenciável no ponto  $t = 0$ . Por outro lado, o caminho  $g : \mathbb{R} \rightarrow \mathbb{R}^2$ , definido por  $g(t) = (t|t|, t^2)$ , tem a mesma imagem que  $f$  porém é derivável em todos os pontos, inclusive para  $t = 0$ , valendo  $g'(0) = (0, 0)$ . Com efeito, se  $t \leq 0$  então  $g(t) = (-t^2, t^2)$  e se  $t \geq 0$  vale  $g(t) = (t^2, t^2)$ . Portanto  $g'(t) = (-2t, 2t)$  quando  $t < 0$  e  $g'(t) = (2t, 2t)$  se  $t > 0$ . No ponto  $t = 0$ , temos  $g'_+(0) = g'_-(0) = (0, 0)$ .  $\triangleleft$

**Exemplo 3.** Sejam  $f : \mathbb{R} \rightarrow \mathbb{R}^2$  e  $g : \mathbb{R} \rightarrow \mathbb{R}^3$  os caminhos definidos por  $f(t) = (\cos t, \sin t)$  e  $g(t) = (\cos t, \sin t, t)$ . A imagem de  $f$  é a circunferência unitária

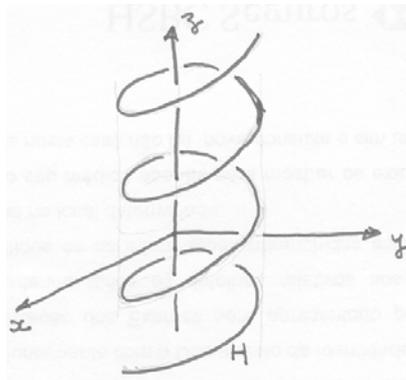


Figura 2.

$S^1$  e a imagem de  $g$  é a hélice  $H$ , cuja projeção sobre o plano  $z = 0$  é  $S^1$ . Ambos,  $f$  e  $g$ , são de classe  $C^k$  para todo  $k \in \mathbb{N}$ , por isso se dizem de classe  $C^\infty$ . Para todo  $t \in \mathbb{R}$  tem-se  $f'(t) = (-\sin t, \cos t)$  e  $g'(t) = (-\sin t, \cos t, 1)$ .  $\triangleleft$

## 2 Cálculo diferencial de caminhos

Sejam  $f, g : I \rightarrow \mathbb{R}^n$  caminhos e  $\alpha : I \rightarrow \mathbb{R}$  uma função real. Se  $f, g$  e  $\alpha$  são diferenciáveis no ponto  $t_0 \in I$  então são também diferenciáveis nesse ponto os caminhos  $f + g$ ,  $\alpha f$  e as funções  $\langle f, g \rangle$  e  $|f| = \sqrt{\langle f, f \rangle}$ , esta última sob a condição de ser  $f(t_0) \neq 0$ .

Valem então as regras abaixo:

1.  $(f + g)'(t_0) = f'(t_0) + g'(t_0)$ ,
2.  $(\alpha f)'(t_0) = \alpha'(t_0) \cdot f(t_0) + \alpha(t_0) \cdot f'(t_0)$ ,
3.  $\langle f, g \rangle'(t_0) = \langle f'(t_0), g(t_0) \rangle + \langle f(t_0), g'(t_0) \rangle$ ,
4.  $|f|'(t_0) = \frac{\langle f(t_0), f'(t_0) \rangle}{|f(t_0)|}$ ,

as quais se provam simplesmente calculando em termos das coordenadas de  $f$  e  $g$ .

Vimos no Exemplo 3 que, em cada ponto, o vetor-velocidade  $f'(t) = (-\sin t, \cos t)$  é perpendicular a  $f(t) = (\cos t, \sin t)$ . A última das regras de derivação acima, segundo a qual  $|f|' = \langle f, f' \rangle / |f|$ , mostra que, mais geralmente, se  $f : I \rightarrow \mathbb{R}^n$  é um caminho diferenciável com  $|f|$  constante (isto é,  $f(t)$  pertence a uma esfera de centro 0) então o vetor-velocidade  $f'(t)$  é perpendicular

a  $f(t)$ , para todo  $t \in I$ . Reciprocamente, se  $\langle f(t), f'(t) \rangle = 0$  para todo  $t \in I$  então  $|f'| = 0$ , logo a função real  $|f| : I \rightarrow \mathbb{R}$  é constante.

Vale também para caminhos diferenciáveis  $f : I \rightarrow \mathbb{R}^n$  o fato de que derivada identicamente nula implica  $f$  constante. Isto pode ser visto diretamente ou a partir do Teorema do Valor Médio, o qual assume, para caminhos, a forma de uma desigualdade.

O Teorema do Valor Médio para funções diferenciáveis  $f : [a, b] \rightarrow \mathbb{R}$  diz que existe  $c$ , com  $a < c < b$ , tal que  $f(b) - f(a) = f'(c)(b - a)$ . Tal igualdade não vale sempre para caminhos  $f : I \rightarrow \mathbb{R}^n$ . Por exemplo, se considerarmos  $f : [0, 2\pi] \rightarrow \mathbb{R}^2$ , dado por  $f(t) = (\cos t, \sin t)$ , temos  $f(2\pi) - f(0) = 0$  mas, como  $|f'(t)| = 1$  para todo  $t \in [0, 2\pi]$  não pode existir  $c \in [0, 2\pi]$  tal que  $f(2\pi) - f(0) = f'(c) \cdot (2\pi - 0)$ .

Tem-se entretanto o seguinte importante resultado:

**Teorema 1 (Desigualdade do Valor Médio).** *Seja  $f : [a, b] \rightarrow \mathbb{R}^n$  um caminho, diferenciável no intervalo aberto  $(a, b)$ , com  $|f'(t)| \leq M$  para todo  $t \in (a, b)$ . Então  $|f(b) - f(a)| \leq M \cdot (b - a)$ .*

*Demonstração.* Definamos  $\varphi : [a, b] \rightarrow \mathbb{R}$  pondo  $\varphi(t) = \langle f(t), f(b) - f(a) \rangle$ . Então, pelo Teorema do Valor Médio (Vol. 1, pág. 96), existe  $c \in (a, b)$  tal que  $\varphi(b) - \varphi(a) = \varphi'(c) \cdot (b - a)$ , pois  $\varphi$  é contínua, derivável em  $(a, b)$ , com  $\varphi'(t) = \langle f'(t), f(b) - f(a) \rangle$ . Mas  $\varphi(b) - \varphi(a) = |f(b) - f(a)|^2$ . Logo, usando a desigualdade de Schwarz, temos:

$$\begin{aligned} |f(b) - f(a)|^2 &= \langle f'(c), f(b) - f(a) \rangle \cdot (b - a) \\ &\leq |f'(c)| |f(b) - f(a)| \cdot (b - a) \\ &\leq M \cdot |f(b) - f(a)| \cdot (b - a). \end{aligned}$$

Cancelando o fator  $|f(b) - f(a)|$ , vem  $|f(b) - f(a)| \leq M \cdot (b - a)$ .  $\square$

**Corolário 1.** *Se o caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  tem derivada nula em todos os pontos de  $(a, b)$  então é constante.*

**Teorema 2 (Regra da Cadeia).** *Sejam  $\varphi : I \rightarrow J$  diferenciável no ponto  $a \in I$  e  $f : J \rightarrow \mathbb{R}^n$  em caminho diferenciável no ponto  $b = f(a)$ . Então o caminho  $f \circ \varphi : I \rightarrow \mathbb{R}^n$  é diferenciável no ponto  $a$ , com  $(f \circ \varphi)'(a) = \varphi'(a) \cdot f'(b)$ .*

*Demonstração.* Aplicar a Regra da Cadeia às funções-coordenada  $f_i \circ \varphi$  do caminho  $f \circ \varphi$ .  $\square$

**Exemplo 4.** Sejam  $f : \mathbb{R} \rightarrow \mathbb{R}^2$  e  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ , com  $f(t) = (\cos t, \sin t)$  e  $\varphi(t) = t^2$ . Então o caminho  $f \circ \varphi : \mathbb{R} \rightarrow \mathbb{R}^2$ , dado por  $(f \circ \varphi)(t) = (\cos t^2, \sin t^2)$ ,

tem vetor-velocidade  $(f \circ \varphi)'(t) = (-2t \operatorname{sen} t^2, 2t \operatorname{cos} t^2) = 2t \cdot (-\operatorname{sen} t^2, \operatorname{cos} t^2)$ , múltiplo escalar do vetor-velocidade de  $f$  no ponto  $\varphi(t)$ .

De um modo geral, a Regra da Cadeia diz que o caminho  $t \mapsto f(\varphi(t))$ , cuja imagem está contida na imagem de  $f$ , tem, para cada  $t \in I$ , vetor-velocidade igual a um múltiplo escalar do vetor-velocidade de  $f$  em  $\varphi(t)$ . ◁

### 3 A integral de um caminho

Lembramos que uma *partição* do intervalo  $[a, b]$  é um conjunto finito  $P = \{t_0 < t_1 < \dots < t_k\}$  com  $t_0 = a$  e  $t_k = b$ . A *norma* de  $P$  é o número  $|P| = \max\{t_i - t_{i-1}; i = 1, \dots, k\}$ . Diz-se que outra partição  $Q$  *refina*  $P$  quando  $P \subset Q$ . Uma *partição pontilhada* é um par  $P^* = (P, \xi)$  onde  $\xi = \{\xi_1, \dots, \xi_k\}$  com  $t_{i-1} \leq \xi_i < t_i$ ,  $1 \leq i \leq k$ .

Dados o caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  e uma partição pontilhada  $P^* = (P, \xi)$  de  $[a, b]$ , a *soma de Riemann* de  $f$  associada a  $P^*$  é definida como

$$\sum(f; P^*) = \sum_{i=1}^k f(\xi_i)(t_i - t_{i-1}).$$

Diz-se que o vetor  $v \in \mathbb{R}^n$  é o *limite* da soma de Riemann  $\sum(f; P^*)$  quando a norma de  $P$  tende a zero, e escreve-se  $v = \lim_{|P| \rightarrow 0} \sum(f; P^*)$ , para significar que, dado arbitrariamente  $\varepsilon > 0$ , existe  $\delta > 0$  tal que  $|P| < \delta \Rightarrow |v - \sum(f; P^*)| < \varepsilon$ , *seja qual for a maneira de pontilhar*  $P$ .

Vimos no Volume 1 (págs. 127 e 137) que se  $f : [a, b] \rightarrow \mathbb{R}$  é contínua então existe  $\lim_{|P| \rightarrow 0} \sum(f; P^*) = \int_a^b f(t)dt$ . Daí resulta que, se  $f : [a, b] \rightarrow \mathbb{R}^n$  é um caminho, existe o limite

$$\lim_{|P| \rightarrow 0} \sum(f; P^*) = \left( \int_a^b f_1(t)dt, \dots, \int_a^b f_n(t)dt \right).$$

Pomos, por definição,

$$\int_a^b f(t)dt = \lim_{|P| \rightarrow 0} \sum(f; P^*).$$

Segue-se da propriedade correspondente para funções reais que

$$\int_a^b [\alpha f(t) + \beta g(t)]dt = \alpha \int_a^b f(t)dt + \beta \int_a^b g(t)dt.$$

Além disso, tem-se a importante desigualdade

$$\left| \int_a^b f(t)dt \right| \leq \int_a^b |f(t)|dt,$$

a qual decorre do fato de que, para toda partição pontilhada  $P^*$  tem-se

$$\left| \sum(f; P^*) \right| \leq \sum(|f|; P^*)$$

pois a norma de uma soma é menor do que ou igual à soma das normas das parcelas.

Em particular, se  $|f(t)| \leq M$  para todo  $t \in [a, b]$  então

$$\left| \int_a^b f(t) dt \right| \leq M \cdot (b - a).$$

**Exemplo 5.** Se  $f : [0, 2\pi] \rightarrow \mathbb{R}^2$  e  $g : [0, 1] \rightarrow \mathbb{R}^2$  são dados por  $f(t) = (\cos t, \sin t)$  e  $g(t) = (t, t^2)$  então  $\int_0^{2\pi} f(t) dt = (0, 0)$  e  $\int_0^1 g(t) dt = (1/2, 1/3)$ .  
◁

Aplicando o Teorema Fundamental do Cálculo a cada uma das coordenadas do caminho de classe  $C^1 f : [a, b] \rightarrow \mathbb{R}^n$ , obtemos o seguinte

**Teorema 3 (Teorema Fundamental do Cálculo para Caminhos).** *Se  $f : [a, b] \rightarrow \mathbb{R}^n$  é um caminho de classe  $C^1$  então*

$$\int_a^b f'(t) dt = f(b) - f(a).$$

Daí resulta outra prova da Desigualdade do Valor Médio (no caso particular de  $f \in C^1$ ), pois se  $|f'(t)| \leq M$  para todo  $t \in [a, b]$  então

$$|f(b) - f(a)| = \left| \int_a^b f'(t) dt \right| \leq M \cdot (b - a).$$

Exprimindo novamente a integral de um caminho em termos das integrais de suas funções-coordenada resulta o Teorema de Mudança de Variável seguinte

Se  $\varphi : [c, d] \rightarrow [a, b]$  é de classe  $C^1$  e  $f : [a, b] \rightarrow \mathbb{R}^n$  é um caminho então

$$\int_{\varphi(c)}^{\varphi(d)} f(t) dt = \int_c^d f(\varphi(t)) \cdot \varphi'(t) dt.$$

Uma simples aplicação desta fórmula nos permite enunciar o Teorema Fundamental do Cálculo assim: se  $f : [a, a+h] \rightarrow \mathbb{R}^n$  é um caminho de classe  $C^1$  então  $f(a+h) - f(a) = h \cdot \int_0^1 f'(a+th) dt$ .

Basta considerar  $\varphi : [0, 1] \rightarrow [a, a+h]$  onde  $\varphi(t) = a + th$  e notar que  $\varphi'(t) = h$ .

Um caminho  $f : I \rightarrow \mathbb{R}^n$  diz-se *uniformemente diferenciável* quando, para todo  $t \in I$  existe um vetor  $f'(t) \in \mathbb{R}^n$  com a seguinte propriedade:

Dado qualquer  $\varepsilon > 0$ , pode-se obter  $\delta > 0$  tal que  $0 < |h| < \delta$  e  $t + h \in I$  implicam  $|f(t + h) - f(t) - f'(t) \cdot h| < \varepsilon|h|$  para qualquer  $t \in I$ .

A diferença entre a diferenciabilidade uniforme e a diferenciabilidade pura e simples situa-se no fato de que o número  $\delta > 0$  depende apenas do  $\varepsilon > 0$  dado, mas não do ponto  $t \in I$  onde se toma a derivada  $f'(t)$ .

**Teorema 4 (Diferenciabilidade Uniforme).** *Todo caminho  $f : [a, b] \rightarrow \mathbb{R}^n$ , de classe  $C^1$  no intervalo compacto  $[a, b]$ , é uniformemente diferenciável.*

*Demonstração.* Pela continuidade uniforme da derivada  $f' : [a, b] \rightarrow \mathbb{R}^n$ , dado  $\varepsilon > 0$  existe  $\delta > 0$  tal que  $|h| < \delta$  e  $t + h \in [a, b]$  implicam  $|f'(t + h) - f'(t)| < \varepsilon$  seja qual for  $t \in [a, b]$ . Observando que, para  $t \in [a, b]$  fixo vale  $\int_t^{t+h} f'(t) ds = f'(t) \cdot h$ , o Teorema Fundamental do Cálculo nos diz que  $0 < |h| < \delta$  e  $t + h \in [a, b]$  implicam

$$|f(t + h) - f(t) - f'(t) \cdot h| = \left| \int_t^{t+h} [f'(s) - f'(t)] ds \right| \leq \varepsilon \cdot |h|$$

para qualquer  $t \in [a, b]$ , o que demonstra o teorema.  $\square$

**Observação.** Vale a recíproca: todo caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  uniformemente diferenciável é de classe  $C^1$ . (Vide “Curso de Análise”, vol. 1, pág. 218 e vol. 2, pág. 88.)

#### 4 Caminhos retificáveis

O comprimento de um caminho  $f : [a, b] \rightarrow \mathbb{R}^n$ , que definiremos a seguir, é a medida da trajetória percorrida pelo ponto  $f(t)$  quando  $t$  varia de  $a$  até  $b$ . Não é o comprimento da curva imagem de  $f$ , pois o ponto  $f(t)$  pode percorrer essa mesma curva de vários modos diferentes, dando origem a caminhos de comprimentos diversos. Por exemplo, o segmento de reta que vai da origem ao ponto  $P = (1, 1)$  do plano tem comprimento  $\sqrt{2}$ . O caminho  $f : [0, 2] \rightarrow \mathbb{R}^2$ , definido por  $f(t) = (2t - t^2, 2t - t^2)$  tem por imagem esse segmento, porém o percorre duas vezes, saindo de  $f(0) = (0, 0)$ , indo até  $f(1) = (1, 1)$  e voltando até  $f(2) = (0, 0)$ . Seu comprimento é, como veremos, igual a  $2\sqrt{2}$ .

Dado um caminho  $f : [a, b] \rightarrow \mathbb{R}^n$ , cada partição  $P = \{a = t_0 < \dots < t = b\}$  de  $[a, b]$  determina uma poligonal inscrita na imagem de  $f$ , cujos vértices são os pontos  $f(a), f(t_1), \dots, f(t_{k-1}), f(b)$ . O comprimento dessa poligonal é o número

$$l(f; P) = \sum_{i=1}^k |f(t_i) - f(t_{i-1})|.$$

Quando não houver perigo de confusão, escreveremos apenas  $l(P)$ , em vez de  $l(f; P)$ .

Diz-se que o caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  é *retificável* quando o conjunto dos números  $l(P)$ , obtidos considerando-se todas as partições  $P$  do intervalo  $[a, b]$ , for limitado. Então o supremo desse conjunto chama-se o *comprimento* do caminho  $f$ , o qual é representado por  $l(f)$ . Assim

$$l(f) = \sup_P l(f; P) = \sup_P l(P).$$

**Exemplo 6.** Seja  $f : [0, 1] \rightarrow \mathbb{R}^n$  o caminho retilíneo  $f(t) = (1 - t)A + tB$ . Para toda partição  $P = \{0 < t_1 < \dots < t_{k-1} < 1\}$  de  $[0, 1]$  tem-se

$$l(P) = \sum |f(t_i) - f(t_{i-1})| = \sum (t_i - t_{i-1})|B - A| = |B - A|.$$

Assim, obviamente vale  $l(f) = |B - A|$ . ◁

**Exemplo 7.** Um caminho não-retificável  $f : [a, b] \rightarrow \mathbb{R}^n$  é aquele em que o ponto  $f(t)$  descreve uma trajetória infinitamente longa no tempo finito  $b - a$ . Um exemplo de tal situação é o caminho  $f : [0, 1] \rightarrow \mathbb{R}^2$ , dado por  $f(t) = (t, \varphi(t))$  o qual percorre o gráfico da função  $\varphi : [0, 1] \rightarrow \mathbb{R}$ . Esta função tem, em cada intervalo

$$\left[ \frac{n}{n+1}, \frac{n+1}{n+2} \right]$$

o gráfico na forma de um triângulo isósceles de altura  $\frac{1}{n+1}$ . Além disso,  $\varphi(1) = 0$ . Se considerarmos, para cada  $n \in \mathbb{N}$ , a partição

$$P_n = \left\{ 0, 1/2, 2/3, \dots, \frac{n+1}{n+2}, 1 \right\}$$

do intervalo  $[0, 1]$ , veremos que  $l(P_n)$  é a soma dos comprimentos dos lados inclinados dos  $n + 1$  primeiros triângulos isósceles que formam o gráfico de  $\varphi$ . Logo  $l(P_n)$  é maior do que a soma das alturas desses triângulos, ou seja,

$$l(P_n) > \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n+1}.$$

Como a série harmônica é divergente, segue-se que o conjunto dos números  $l(P)$  associados ao caminho  $f$  é ilimitado, portanto  $f$  não é retificável. O caminho  $f$  tem comprimento infinito. ◁

Uma observação simples, porém útil, é a seguinte: *se a partição  $Q$  do intervalo  $[a, b]$  refina a partição  $P$  então, dado o caminho  $f : [a, b] \rightarrow \mathbb{R}^n$ , tem-se  $l(P) \leq$*

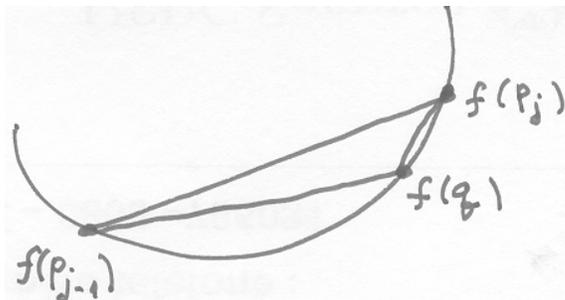


Figura 3.

$l(Q)$ . Para ver isto, basta considerar o caso em que se obtém  $Q$  a partir de  $P$  acrescentando-lhe um só ponto  $q$ , pois cada refinamento de  $P$  pode ser pensado como a repetição de um número finito desses acréscimos. Ora, se  $Q$  difere de  $P$  pela adição do único ponto  $q$ , digamos com  $p_{j-1} < q < p_j$ , então

$$l(Q) - l(P) = |f(q) - f(p_{j-1})| + |f(p_j) - f(q)| - |f(p_j) - f(p_{j-1})| \geq 0$$

pois

$$\begin{aligned} |f(p_j) - f(p_{j-1})| &= |f(p_j) - f(q) + f(q) - f(p_{j-1})| \\ &\leq |f(p_j) - f(q)| + |f(q) - f(p_{j-1})|. \end{aligned}$$

Como no caso da integral, dado um caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  diremos que o número real  $A$  é o *limite de  $l(P)$  quando  $|P|$  tende a zero*, e escreveremos  $\lim_{|P| \rightarrow 0} l(P) = A$ , para significar que, para todo  $\varepsilon > 0$  dado, é possível obter  $\delta > 0$  tal que  $|P| < \delta$  implica  $|l(P) - A| < \varepsilon$ .

**Teorema 5.** Se  $\lim_{|P| \rightarrow 0} l(P) = A$  então  $A = \sup_P l(P)$ , ou seja, o caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  é retificável e  $l(f) = A$ .

*Demonstração.* Se  $\lim_{|P| \rightarrow 0} l(P) = A$ , é claro que  $A \leq \sup_P l(P)$ . Suponhamos, por absurdo, que seja  $A < \sup_P l(P)$ . Então existe uma partição  $Q_0$  tal que  $A < l(Q_0)$ . Seja  $\varepsilon = l(Q_0) - A$ . Pela definição do limite, podemos obter  $\delta > 0$  tal que  $|P| < \delta \Rightarrow A - \varepsilon < l(P) < A + \varepsilon = l(Q_0)$ . Tomemos uma partição qualquer  $P_0$  tal que  $|P_0| < \delta$ . A partição  $P = P_0 \cup Q_0$ , por um lado cumpre  $|P| < \delta$ , logo  $l(P) < l(Q_0)$  e, por outro lado, refina  $Q_0$ , logo  $l(Q_0) \leq l(P)$ . Esta contradição prova o teorema.  $\square$

**Observação.** Vale a recíproca: se  $f$  é retificável, então  $l(f) = \lim_{|P| \rightarrow 0} l(P)$ . (Vide “Curso de Análise”, vol. 2, pág. 99.) Mas somente o teorema acima será usado a seguir.

**Teorema 6.** *Todo caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  de classe  $C^1$  é retificável e  $l(f) = \int_a^b |f'(t)| dt$ .*

*Demonstração.* Para toda partição  $P = \{t_0 < t_1 < \dots < t_k\}$  de  $[a, b]$ , sejam  $\sum(P) = \sum_{i=1}^k |f'(t_{i-1})|(t_i - t_{i-1})$  e  $l(P) = \sum_{i=1}^k |f(t_i) - f(t_{i-1})|$ . Sabemos que  $\lim_{|P| \rightarrow 0} \sum(P) = \int_a^b |f'(t)| dt$ . E, pelo Teorema 4, para todo  $\varepsilon > 0$  dado arbitrariamente, existe  $\delta > 0$  tal que  $|P| < \delta$  implica  $f(t_i) - f(t_{i-1}) = [f'(t_{i-1} + \rho_i)]$  com  $|\rho_i| < \frac{\varepsilon}{b-a}$  para  $i = 1, \dots, k$ . Logo  $l(P) = \sum_{i=1}^k |f(t_i) - f(t_{i-1})| = \sum_{i=1}^k |f'(t_i) + \rho_i|(t_i - t_{i-1})$ , portanto  $|\sum(P) - l(P)| \leq \sum_{i=1}^k |\rho_i|(t_i - t_{i-1}) < \varepsilon$  sempre que  $|P| < \delta$ . Como  $\lim_{|P| \rightarrow 0} \sum(P) = \int_a^b |f'(t)| dt$ , resulta daí que  $\lim_{|P| \rightarrow 0} l(P) = \int_a^b |f'(t)| dt$ . Pelo Teorema 5, concluímos que  $l(f) = \int_a^b |f'(t)| dt$ .  $\square$

Uma *reparametrização* do caminho  $f : [a, b] \rightarrow \mathbb{R}^n$  é um caminho da forma  $f \circ \varphi : [c, d] \rightarrow \mathbb{R}^n$ , onde  $\varphi : [c, d] \rightarrow [a, b]$  é uma função de classe  $C^1$  tal que  $\varphi(c) = a$ ,  $\varphi(d) = b$  e  $\varphi'(u) \geq 0$  para todo  $u \in [c, d]$ . O teorema acima tem, como consequência imediata, o seguinte

**Corolário 2.** *Um caminho de classe  $C^1$ ,  $f : [a, b] \rightarrow \mathbb{R}^n$ , e qualquer sua reparametrização  $f \circ \varphi : [c, d] \rightarrow \mathbb{R}^n$  têm o mesmo comprimento.*

Com efeito, pelo Teorema,

$$\begin{aligned} l(f) &= \int_a^b |f'(t)| dt = \int_c^d \varphi'(u) \cdot |f'(\varphi(u))| du \\ &= \int_c^d |\varphi'(u) \cdot f'(\varphi(u))| du = \int_c^d |(f \circ \varphi)'(u)| du = l(f \circ \varphi). \end{aligned}$$

Para caminhos  $f : [a, b] \rightarrow \mathbb{R}^n$  de classe  $C^1$  com a propriedade adicional de que  $f'(t) \neq 0$  para todo  $t \in [a, b]$  (chamados caminhos *regulares*), existe uma reparametrização especial, “por comprimento de arco”, que apresentamos agora. Dado um tal caminho  $f$ , digamos com  $l(f) = L$ , definimos a função  $\varphi : [a, b] \rightarrow [0, L]$  pondo, para todo  $t \in [a, b]$ ,  $\varphi(t) = \int_0^t |f'(u)| du = l(f | [a, t])$ , comprimento do caminho  $f | [a, t]$ , restrição de  $f$  ao intervalo  $[a, t]$ .

A função  $\varphi : [a, b] \rightarrow [0, L]$ , assim definida, é de classe  $C^1$ , com  $\varphi'(t) = |f'(t)| > 0$  para todo  $t \in [a, b]$ , e  $\varphi(a) = 0$ ,  $\varphi(b) = L$ . Logo é uma bijeção de  $[a, b]$  sobre  $[0, L]$ , cuja inversa  $\varphi^{-1} : [0, L] \rightarrow [a, b]$  é também de classe  $C^1$ , valendo, para todo  $s = \varphi(t) \in [0, L]$ , a fórmula  $(\varphi^{-1})'(s) = \frac{1}{\varphi'(t)} = \frac{1}{|f'(t)|} > 0$ . (cfr. Vol. 1, pág. 92.)

Consideremos a reparametrização  $g = f \circ \varphi^{-1} : [0, L] \rightarrow \mathbb{R}^n$  do caminho  $f$ . Para todo  $s = \varphi(t) \in [0, L]$  temos

$$g'(s) = (\varphi^{-1})'(s) \cdot f'(t) = \frac{f'(t)}{|f'(t)|},$$

portanto  $|g'(s)| = 1$ .

Então, para todo  $s \in [0, L]$ , o comprimento do caminho restrito  $g \mid [0, s]$  tem o valor

$$l(g \mid [0, s]) = \int_0^s |g'(u)| du = \int_0^s du = s.$$

Por este motivo,  $g = f \circ \varphi^{-1}$  chama-se a reparametrização de  $f$  por comprimento de arco.

**Observação.** A fórmula  $l(f) = \int_a^b |f'(t)| dt$  é importante teoricamente mas, em geral, é impraticável procurar calcular essa integral, a não ser numericamente ou então em raros casos especialmente escolhidos, como  $f(t) = (1-t)A + tb$ ,  $f(t) = (\cos t, \sin t)$  e outros.

## Capítulo 3

# Funções Reais de $n$ Variáveis

### 1 Derivadas parciais

Seja  $f: U \rightarrow \mathbb{R}$  uma função definida no aberto  $U \subset \mathbb{R}^n$ . Para cada  $i = 1, \dots, n$ , a  $i$ -ésima derivada parcial de  $f$  no ponto  $a = (a_1, \dots, a_n) \in U$  é o número

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t} = \lim_{t \rightarrow 0} \frac{f(a_1, \dots, a_i + t, \dots, a_n) - f(a)}{t},$$

caso este limite exista. Como  $U$  é aberto, podemos achar  $\delta > 0$  tal que  $a + te_i \in U$  para todo  $t \in (-\delta, \delta)$ . Então está bem definido o caminho retilíneo  $\lambda: (-\delta, \delta) \rightarrow U$ ,  $\lambda(t) = a + te_i$ . A definição acima diz que  $\frac{\partial f}{\partial x_i}(a) = (f \circ \lambda)'(0) =$  derivada, no ponto  $t = 0$ , da função real  $f \circ \lambda: (-\delta, \delta) \rightarrow \mathbb{R}$ .

Observemos que  $\partial f / \partial x_i$  significa a derivada de  $f$  em relação a sua  $i$ -ésima variável, seja qual for o nome que se atribua a ela. Assim

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial y_i} = \frac{\partial f}{\partial z_i}, \text{ etc.}$$

Uma notação alternativa, que evitaria mal-entendidos, seria  $\partial_i f$ . Preferimos a notação tradicional  $\partial f / \partial x_i$  porque ela é conveniente quando se usa a regra da cadeia.

Quando  $n = 2$  ou  $n = 3$ , escrevemos  $(x, y)$  em vez de  $(x_1, x_2)$  e  $(x, y, z)$  em vez de  $(x_1, x_2, x_3)$ . Assim,  $\partial f / \partial x$  é a derivada parcial de  $f$  em relação à primeira variável,  $\partial f / \partial z$  é a terceira derivada parcial de  $f$ , etc.

**Exemplo 1.** Seja  $f: \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x, y) = xy/(x^2 + y^2)$  se  $x^2 + y^2 \neq 0$  e  $f(0, 0) = 0$ . Como  $f(0, y) = 0$  para todo  $y$  e  $f(x, 0) = 0$  para todo  $x$ , segue-se que  $\frac{\partial f}{\partial x}(0, 0) = 0$  e  $\frac{\partial f}{\partial y}(0, 0) = 0$ . Entretanto a função  $f$  é descontínua na origem

$(0, 0)$ . Com efeito, se chamarmos de  $\theta$  o ângulo que o vetor não-nulo  $v = (x, y)$  forma com o eixo das abscissas, veremos que

$$f(x, y) = \frac{x}{\sqrt{x^2 + y^2}} \cdot \frac{y}{\sqrt{x^2 + y^2}} = \cos \theta \cdot \operatorname{sen} \theta.$$

Logo, atribuindo diferentes valores a  $\theta$ , podemos fazer com que  $f(x, y)$  tenha limites diferentes quando  $(x, y)$  tende para  $(0, 0)$  ao longo do segmento  $x = t \cos \theta$ ,  $y = t \operatorname{sen} \theta$ , ou seja, quando  $t \rightarrow 0$ .  $\triangleleft$

O exemplo acima mostra que a existência das  $n$  derivadas parciais no ponto  $a$  não assegura a continuidade da função  $f$  nesse ponto. Para cada  $i = 1, \dots, n$ , a função  $\lambda(t) = f(a + te_i)$  é essencialmente a restrição de  $f$  ao segmento  $(a - \delta e_i, a + \delta e_i)$  da reta que passa pelo ponto  $a$  e é paralela ao  $i$ -ésimo eixo coordenado de  $\mathbb{R}^n$ . A derivada parcial  $\frac{\partial f}{\partial x_i}(a) = (f \circ \lambda)'(0)$  dá informação apenas sobre o comportamento de  $f$  ao longo desse intervalo. Em particular, a existência das  $n$  derivadas parciais de  $f$  no ponto  $a$  implica que a restrição de  $f$  aos  $n$  intervalos paralelos aos eixos, que se cortam no ponto  $a$ , é contínua, embora não garanta a continuidade de  $f : U \rightarrow \mathbb{R}$  em  $a$ .

Se  $\partial f / \partial x_i$  existe e é positiva em todos os pontos do segmento de reta  $[a - \delta e_i, a + \delta e_i]$ , paralelo ao  $i$ -ésimo eixo coordenado, então  $f$  é crescente ao longo desse segmento:  $s < t \Rightarrow f(a + se_i) < f(a + te_i)$ , desde que  $|s - a| \leq \delta$  e  $|t - a| \leq \delta$ . Isto resulta imediatamente do resultado análogo para funções de uma variável.

A noção de derivada parcial também faz sentido para aplicações  $f : U \rightarrow \mathbb{R}^n$ , com  $U \subset \mathbb{R}^m$  aberto. Se  $a \in U$ , põe-se, para cada  $i = 1, \dots, m$ :

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t}.$$

Evidentemente,  $\partial f / \partial x_i$  é um vetor de  $\mathbb{R}^n$ . Se  $f = (f_1, \dots, f_n)$  então

$$\frac{\partial f}{\partial x_i}(a) = \left( \frac{\partial f_1}{\partial x_i}(a), \dots, \frac{\partial f_n}{\partial x_i}(a) \right).$$

Neste capítulo, porém, daremos prioridade às funções com valores numéricos. Para elas tem sentido o vetor gradiente, conceito de forte apelo intuitivo, que contribui para entendermos como cresce (ou decresce)  $f(x)$ .

## 2 Funções de classe $C^1$

Seja  $f : U \rightarrow \mathbb{R}$  uma função que possui as  $n$  derivadas parciais em todos os pontos do aberto  $U \subset \mathbb{R}^n$ . Ficam então definidas  $n$  funções

$$\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} : U \rightarrow \mathbb{R}, \text{ onde } \frac{\partial f}{\partial x_i} : x \mapsto \frac{\partial f}{\partial x_i}(x).$$

Se estas funções forem contínuas em  $U$ , diremos que  $f$  é uma *função de classe  $C^1$*  e escreveremos  $f \in C^1$ .

Uma aplicação  $f : U \rightarrow \mathbb{R}^n$ , definida no aberto  $U \subset \mathbb{R}^n$ , diz-se de *classe  $C^1$*  quando cada uma de suas funções-coordenada  $f_1, \dots, f_n : U \rightarrow \mathbb{R}$  é de classe  $C^1$ .

Muitas propriedades importantes das funções de classe  $C^1$  resultam de serem elas diferenciáveis no sentido seguinte.

Uma função  $f : U \rightarrow \mathbb{R}$ , definida no aberto  $U \subset \mathbb{R}^n$ , diz-se *diferenciável* no ponto  $a \in U$  quando cumpre as seguintes condições:

1. Existem as derivadas parciais  $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$ .
2. Para todo  $v = (\alpha_1, \dots, \alpha_n)$  tal que  $a + v \in U$ , tem-se

$$f(a + v) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \alpha_i + r(v), \quad \text{onde } \lim_{|v| \rightarrow 0} \frac{r(v)}{|v|} = 0.$$

**Observações. 1.** Acima, e sempre que fizermos considerações em torno de um ponto específico  $a$ , escreveremos, por simplicidade,  $\frac{\partial f}{\partial x_i}$  em vez de  $\frac{\partial f}{\partial x_i}(a)$ .

**2.** A essência da definição da diferenciabilidade está na condição  $\lim_{v \rightarrow 0} (r(v)/|v|) = 0$ , pois a igualdade que define o “resto”  $r(v)$  pode ser escrita para qualquer função que possua as  $n$  derivadas parciais.

De  $\lim_{v \rightarrow 0} \frac{r(v)}{|v|} = 0$  resulta que  $\lim_{v \rightarrow 0} r(v) = 0$  pois  $r(v) = \left( \frac{r(v)}{|v|} \right) \cdot |v|$ . Segue-se que  $\lim_{v \rightarrow 0} [f(a + v) - f(a)] = 0$ . Portanto, toda função diferenciável no ponto  $a$  é contínua nesse ponto.

Diremos que  $f : U \rightarrow \mathbb{R}$  é *diferenciável* quando  $f$  for diferenciável em todos os pontos de  $U$ .

Quando  $n = 1$ , a função  $f : U \rightarrow \mathbb{R}$  é diferenciável no ponto  $a$  se, e somente se, possui derivada neste ponto pois, como podemos agora dividir por  $v \in \mathbb{R}$ , de  $f(a + v) - f(a) = df/dx \cdot v + r(v)$  resulta

$$\frac{r(v)}{|v|} = \pm \left[ \frac{f(a + v) - f(a)}{v} - \frac{df}{dx}(a) \right],$$

$$\text{portanto } \lim_{v \rightarrow 0} \frac{r(v)}{|v|} = 0 \iff \lim_{v \rightarrow 0} \frac{f(a+v) - f(a)}{v} = \frac{df}{dx}(a).$$

**Teorema 1.** *Toda função  $f : U \rightarrow \mathbb{R}$  de classe  $C^1$  é diferenciável.*

*Demonstração.* Por simplicidade, suporemos  $U \subset \mathbb{R}^2$ . O caso geral se trata analogamente, apenas com uma notação mais elaborada. Fixemos  $c = (a, b) \in U$  e tomemos  $v = (h, k)$  tal que  $c + v \in U$ . Seja

$$r(v) = r(h, k) = f(a+h, b+k) - f(a, b) - \frac{\partial f}{\partial x} \cdot h - \frac{\partial f}{\partial y} \cdot k,$$

onde as derivadas são calculadas no ponto  $c = (a, b)$ . Podemos escrever

$$\begin{aligned} r(v) &= f(a+h, b+k) - f(a, b+k) + f(a, b+k) - f(a, b) \\ &\quad - \frac{\partial f}{\partial x} \cdot h - \frac{\partial f}{\partial x} \cdot k. \end{aligned}$$

Pelo Teorema do Valor Médio para funções de uma variável real, existem  $\theta_1, \theta_2 \in (0, 1)$  tais que

$$r(v) = \frac{\partial f}{\partial x}(a + \theta_1 h, b+k) \cdot h + \frac{\partial f}{\partial y}(a, b + \theta_2 k) \cdot k - \frac{\partial f}{\partial x} \cdot h - \frac{\partial f}{\partial y} \cdot k,$$

logo

$$\begin{aligned} \frac{r(v)}{|v|} &= \left[ \frac{\partial f}{\partial x}(a + \theta_1 h, b+k) - \frac{\partial f}{\partial x}(a, b) \right] \frac{h}{\sqrt{h^2 + k^2}} \\ &\quad + \left[ \frac{\partial f}{\partial y}(a, b + \theta_2 k) - \frac{\partial f}{\partial y}(a, b) \right] \frac{k}{\sqrt{h^2 + k^2}}. \end{aligned}$$

Quando  $v \rightarrow 0$  os termos dentro dos colchetes acima tendem a zero, pela continuidade das derivadas  $\partial f/\partial x$  e  $\partial f/\partial y$ . Além disso, os termos fora dos colchetes têm valor absoluto  $\leq 1$ . Portanto  $\lim_{v \rightarrow 0} r(v)/|v| = 0$  e então  $f$  é diferenciável.  $\square$

**Corolário 1.** *Toda função de classe  $C^1$  é contínua.*

Às vezes, como na demonstração a seguir, é mais conveniente tomar  $\rho = r(v) = r(v)/|v|$  e escrever  $\rho|v|$  em vez de  $r(v)$ . Então a diferenciabilidade de  $f$  se exprime como

$$f(a+v) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \alpha_i + \rho|v|, \quad \text{com } \lim_{v \rightarrow 0} \rho = 0.$$

**Teorema 2.** *Sejam  $U \subset \mathbb{R}^m$ ,  $V \subset \mathbb{R}^n$  abertos,  $f : U \rightarrow V$  uma aplicação cujas funções-coordenada  $f_1, \dots, f_n$  possuem derivadas parciais no ponto  $a \in U$ , e  $g : V \rightarrow \mathbb{R}$  uma função diferenciável no ponto  $b = f(a)$ . Então  $g \circ f : U \rightarrow \mathbb{R}$  possui derivadas parciais no ponto  $a$  e vale*

$$\frac{\partial(g \circ f)}{\partial x_i} = \sum_{k=1}^n \frac{\partial g}{\partial y_k} \cdot \frac{\partial f_k}{\partial x_i}, \quad i = 1, \dots, m,$$

onde as derivadas parciais relativas aos  $x_i$  são calculadas no ponto  $a$  e as relativas a  $y_k$  são calculadas no ponto  $b = f(a)$ .

Além disso, se  $f$  e  $g$  são ambas de classe  $C^1$  então  $g \circ f \in C^1$ .

**Observação.** No Capítulo 5 provaremos, mais geralmente, que se  $f$  e  $g$  forem diferenciáveis então  $g \circ f$  é diferenciável.

*Demonstração.* Podemos escrever

$$\begin{aligned} g(f(a + te_i)) - g(f(a)) &= \sum_{k=1}^n \frac{\partial g}{\partial y_k} \cdot [f_k(a + te_i) - f_k(a)] \\ &\quad + \rho(t) \cdot |f(a + te_i) - f(a)| \end{aligned}$$

onde, por simplicidade, escrevemos  $\rho(t)$  em vez de  $\rho(v)$  com  $v = f(a + te_i) - f(a)$ . A diferenciabilidade de  $g$  nos dá  $\lim_{t \rightarrow 0} \rho(t) = 0$ . Então

$$\begin{aligned} \frac{g(f(a + te_i)) - g(f(a))}{t} &= \sum_{k=1}^n \frac{\partial g}{\partial y_k} \cdot \frac{f_k(a + te_i) - f_k(a)}{t} \\ &\quad \pm \rho(t) \left| \frac{f(a + te_i) - f(a)}{t} \right|. \end{aligned}$$

Logo

$$\frac{\partial(g \circ f)}{\partial x_i} = \lim_{t \rightarrow 0} \frac{g(f(a + te_i)) - g(f(a))}{t} = \sum_{k=1}^n \frac{\partial g}{\partial y_k} \cdot \frac{\partial f_k}{\partial x_i}$$

pois

$$\lim_{t \rightarrow 0} \rho(t) = 0 \text{ e } \lim_{t \rightarrow 0} \left| \frac{f(a + te_i) - f(a)}{t} \right| = \left| \frac{\partial f}{\partial x_i}(a) \right|.$$

O fato de que  $g \circ f \in C^1$  decorre da expressão de  $\partial(g \circ f)/\partial x_i$  em termos das derivadas parciais de  $g$  e das  $f_k$ , que são contínuas.  $\square$

O *gradiente* de uma função diferenciável  $f : U \rightarrow \mathbb{R}$  no ponto  $a \in U$  é o vetor

$$\text{grad } f(a) = \left( \frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a) \right).$$

Se  $v$  é qualquer vetor de  $\mathbb{R}^n$ , a *derivada direcional* de  $f$  no ponto  $a$ , na direção de  $v$  é, por definição,

$$\frac{\partial f}{\partial v}(a) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t}.$$

Estas definições permitem enunciar os seguintes corolários da Regra da Cadeia. O primeiro deles mostra que, quando  $f$  é diferenciável no ponto  $a$ , a derivada direcional  $\frac{\partial f}{\partial v}(a)$  existe em relação a qualquer vetor  $v$ , dá uma expressão para essa derivada em termos das derivadas parciais de  $f$  e das coordenadas de  $v$  e, finalmente, mostra que, na definição de  $\frac{\partial f}{\partial v}(a)$ , em vez do caminho retilíneo  $t \mapsto a + tv$ , pode-se usar qualquer caminho  $\lambda : (-\delta, \delta) \rightarrow U$  desde que se tenha  $\lambda(0) = a$  e  $\lambda'(0) = v$ . O Corolário 3 é, na realidade, um importante teorema.

**Corolário 2.** *Seja  $f : U \rightarrow \mathbb{R}$  diferenciável no aberto  $U \subset \mathbb{R}^n$ , com  $a \in U$ . Dado o vetor  $v = (\alpha_1, \dots, \alpha_n)$ , se  $\lambda : (-\delta, \delta) \rightarrow U$  é qualquer caminho diferenciável tal que  $\lambda(0) = a$  e  $\lambda'(0) = v$ , tem-se*

$$(f \circ \lambda)'(0) = \langle \text{grad } f(a), v \rangle = \frac{\partial f}{\partial v}(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a) \cdot \alpha_i.$$

Basta aplicar diretamente a fórmula

$$(f \circ \lambda)' = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \frac{d\lambda_i}{dt},$$

observando que, para  $\lambda(t) = (\lambda_1(t), \dots, \lambda_n(t))$ , tem-se  $\alpha_i = \frac{d\lambda_i}{dt}(0)$ . Notar ainda que  $\frac{df}{dv}(a) = (f \circ \lambda)'(0)$  com  $\lambda(t) = a + tv$ , pois  $\lambda'(0) = v$ .

**Corolário 3 (Teorema do Valor Médio).** *Dada  $f : U \rightarrow \mathbb{R}$  diferenciável no aberto  $U \subset \mathbb{R}^n$ , se o segmento de reta  $[a, a + v]$  estiver contido em  $U$  então existe  $\theta \in (0, 1)$  tal que*

$$\begin{aligned} f(a + v) - f(a) &= \frac{\partial f}{\partial v}(a + \theta v) = \langle \text{grad } f(a + \theta v), v \rangle \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(a + \theta v) \cdot \alpha_i \end{aligned}$$

onde  $v = (\alpha_1, \dots, \alpha_n)$ .

Com efeito, considerando o caminho retilíneo  $\lambda : [0, 1] \rightarrow U$ , dado por  $\lambda(t) = a + tv$ , vemos que  $f(a + v) - f(a) = (f \circ \lambda)(1) - (f \circ \lambda)(0)$ . Pelo Teorema do Valor Médio para funções de uma variável real, existe  $\theta \in (0, 1)$  tal que  $(f \circ \lambda)(1) - (f \circ \lambda)(0) = (f \circ \lambda)'(\theta)$ . Pela Regra da Cadeia,

$$(f \circ \lambda)'(\theta) = \sum \frac{\partial f}{\partial x_i}(a + \theta v) \cdot \alpha_i = \frac{\partial f}{\partial v}(a + \theta v) = \langle \text{grad } f(a + \theta v), v \rangle.$$

**Corolário 4.** *Seja  $f : U \rightarrow \mathbb{R}$  diferenciável no aberto  $U \subset \mathbb{R}^n$ . Se o segmento de reta  $[a, a + v]$  estiver contido em  $U$  e existir  $M > 0$  tal que  $|\text{grad } f(a + tv)| \leq M$  para todo  $t \in [0, 1]$  então  $|f(a + v) - f(a)| \leq M \cdot |v|$ .*

Com efeito, pela desigualdade de Schwarz,

$$|f(a + v) - f(a)| = |\langle \text{grad } f(a + \theta v), v \rangle| \leq |\text{grad } f(a + \theta v)| |v| \leq M \cdot |v|.$$

Em particular, se  $U$  é convexo,  $f$  é diferenciável e  $|\text{grad } f(x)| \leq M$  para todo  $x \in U$  então  $|f(y) - f(x)| \leq M|x - y|$  quaisquer que sejam  $x, y \in U$ .

**Corolário 5.** *Seja  $f : U \rightarrow \mathbb{R}$  diferenciável no aberto  $U \subset \mathbb{R}^n$ . Se  $U$  é conexo e  $\text{grad } f(x) = 0$  para todo  $x \in U$  então  $f$  é constante.*

Com efeito, pelo Teorema do Valor Médio (Corolário 3),  $f$  é constante ao longo de todo segmento de reta contido em  $U$ . Ora, sendo o aberto  $U$  conexo, dois quaisquer de seus pontos podem ser ligados por um caminho poligonal (justaposição de segmentos de reta) contido em  $U$ .

Dada  $f : U \rightarrow \mathbb{R}$  de classe  $C^1$ , o conjunto  $f^{-1}(c) = \{x \in U; f(x) = c\}$  é, para todo  $c \in \mathbb{R}$ , chamado o *conjunto de nível  $c$*  da função  $f$ . Quando  $U \subset \mathbb{R}^n$  e  $n = 2$  esse conjunto é geralmente chamado a *curva* ou *linha de nível  $c$*  de  $f$ , a qual é definida pela equação  $f(x, y) = c$ . Analogamente, quando  $n = 3$ , o conjunto  $f^{-1}(c)$ , definido pela equação  $f(x, y, z) = c$  costuma ser chamado a *superfície de nível  $c$*  da função  $f$ . Deve-se observar porém que, para certas funções especialmente escolhidas, tais conjuntos podem ser bem diferentes daquilo que se imagina como uma curva ou uma superfície.

Mencionaremos a seguir algumas propriedades do gradiente. Elas justificam a importância desse vetor, o qual dá interessantes informações sobre o comportamento da função.

Para isto, fixaremos  $a \in U$  e suporemos que  $\text{grad } f(a) \neq 0$ . Então:

- 1) O gradiente aponta para uma direção segundo a qual a função é crescente;

- 2) Dentre todas as direções ao longo das quais a função cresce, a direção do gradiente é a de crescimento mais rápido;
- 3) O gradiente de  $f$  no ponto  $a$  é ortogonal ao conjunto de nível de  $f$  que passa por  $a$ .

Vejam os o que significam estas afirmações.

Em primeiro lugar, pondo  $w = \text{grad } f(a)$  temos

$$\frac{\partial f}{\partial w}(a) = \langle \text{grad } f(a), w \rangle = |\text{grad } f(a)|^2 > 0.$$

Isto quer dizer que se  $\lambda : (-\varepsilon, \varepsilon) \rightarrow U$  é tal que  $\lambda \in C^1$ ,  $\lambda(0) = a$  e  $\lambda'(0) = \text{grad } f(a)$  então a função  $t \mapsto f(\lambda(t))$  tem derivada positiva no ponto  $t = 0$ . Logo, diminuindo  $\varepsilon$  se necessário,  $f \circ \lambda : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$  será uma função crescente. É este o significado de “ $f$  cresce na direção do gradiente.”

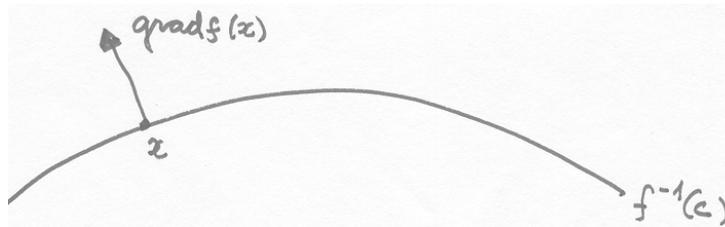


Figura 1.

Como  $\partial f / \partial v = \langle \text{grad } f, v \rangle$ , os vetores  $v$  que apontam para as direções ao longo das quais  $f$  cresce são aqueles para os quais se tem  $\langle \text{grad } f, v \rangle > 0$ , isto é, aqueles que formam um ângulo agudo com  $\text{grad } f(a)$ . Dizer que o crescimento de  $f$  é mais rápido na direção do gradiente significa o seguinte: se  $v \in \mathbb{R}^n$  é tal que  $|v| = |\text{grad } f(a)|$  então

$$\frac{\partial f}{\partial v}(a) \leq \frac{\partial f}{\partial(\text{grad } f(a))}(a).$$

Com efeito, pela desigualdade de Schwarz:

$$\begin{aligned} \frac{\partial f}{\partial v}(a) &= \langle \text{grad } f(a), v \rangle \leq |\text{grad } f(a)| \cdot |v| \\ &= |\text{grad } f(a)|^2 = \frac{\partial f}{\partial(\text{grad } f(a))}(a). \end{aligned}$$

Esclareçamos agora a terceira das afirmações acima.

Dizer que  $w \in \mathbb{R}^n$  é ortogonal ao conjunto de nível  $f^{-1}(c)$  significa que, dado qualquer caminho  $\lambda : (-\varepsilon, \varepsilon) \rightarrow f^{-1}(c)$ , diferenciável no ponto  $t = 0$ , com  $\lambda(0) = a$ , tem-se  $\langle w, \lambda'(0) \rangle = 0$ . Ora,  $\lambda(t) \in f^{-1}(c)$  significa que  $f(\lambda(t)) = c$  para todo  $t \in (-\varepsilon, \varepsilon)$ , portanto  $f \circ \lambda : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$  é constante, igual a  $c$ , logo  $(f \circ \lambda)'(0) = 0$ , ou seja  $\langle \text{grad } f(a), \lambda'(0) \rangle = 0$ .

Assim,  $\text{grad } f(a)$  é ortogonal ao vetor velocidade no ponto  $a = \lambda(0)$  de qualquer caminho diferenciável  $\lambda$  contido no conjunto de nível  $f^{-1}(c)$ .

Ficam portanto constatadas as três propriedades do gradiente acima enunciadas. Vejamos agora alguns exemplos simples.

**Exemplo 2.** Sejam  $f, g, h : \mathbb{R}^2 \rightarrow \mathbb{R}$  definidas por  $f(x, y) = ax + by$  (com  $a^2 + b^2 \neq 0$ ),  $g(x, y) = x^2 + y^2$  e  $h(x, y) = x^2 - y^2$ . A linha de nível  $c$  de  $f$  é a reta definida pela equação  $ax + by = c$ . O vetor  $\text{grad } f(x, y)$  é constante:  $\text{grad } f = (a, b)$  em qualquer ponto  $(x, y) \in \mathbb{R}^2$ . Assim as linhas de nível de  $f$  são retas paralelas umas às outras, todas perpendiculares ao vetor  $v = (a, b)$ .

O conjunto de nível  $c$  da função  $g(x, y) = x^2 + y^2$  é vazio se  $c < 0$  e reduz-se ao ponto  $0 \in \mathbb{R}^2$  quando  $c = 0$ . Para  $c > 0$ , a linha de nível  $c$  é a circunferência de equação  $x^2 + y^2 = c$ , cujo centro é a origem e cujo raio é  $\sqrt{c}$ . O gradiente de  $g$  é  $\text{grad } g(x, y) = (2x, 2y)$ , um vetor colinear com o raio, o que era de esperar pois a tangente da circunferência é perpendicular ao raio no ponto de contacto.

A linha de nível  $0$  da função  $h(x, y) = x^2 - y^2$  é o par de retas perpendiculares definidas pela equação  $x^2 - y^2 = 0$ , que equivale a “ $x + y = 0$  ou  $x - y = 0$ ”. Se  $c > 0$ ,  $x^2 - y^2 = c$  define uma hipérbole cujo eixo é o eixo das abcissas; se  $c < 0$  a hipérbole  $x^2 - y^2 = c$  tem como eixo o eixo das ordenadas. O gradiente de  $h$  é o vetor  $\text{grad } h(x, y) = (2x, -2y)$ . Atribuindo valores particulares a  $x$  e  $y$ , vemos que este vetor é perpendicular à curva de nível que passa em  $(x, y)$  e aponta na direção de crescimento de  $h$ .  $\triangleleft$

Chama-se *ponto crítico* de uma função diferenciável  $f : U \rightarrow \mathbb{R}$  um ponto  $a \in U$  tal que  $\text{grad } f(a) = 0$ .

A função  $f$  do Exemplo 2 não possui ponto crítico. As funções  $g$  e  $h$  do mesmo exemplo têm a origem como ponto crítico. Nota-se em ambos os casos uma quebra de regularidade na disposição das curvas de nível quando se atinge um nível em que há ponto crítico.

### 3 O Teorema de Schwarz

Seja  $f : U \rightarrow \mathbb{R}$  uma função que possui as derivadas parciais  $\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x)$  em todo ponto  $x$  do aberto  $U \subset \mathbb{R}^n$ . A  $j$ -ésima derivada parcial da função

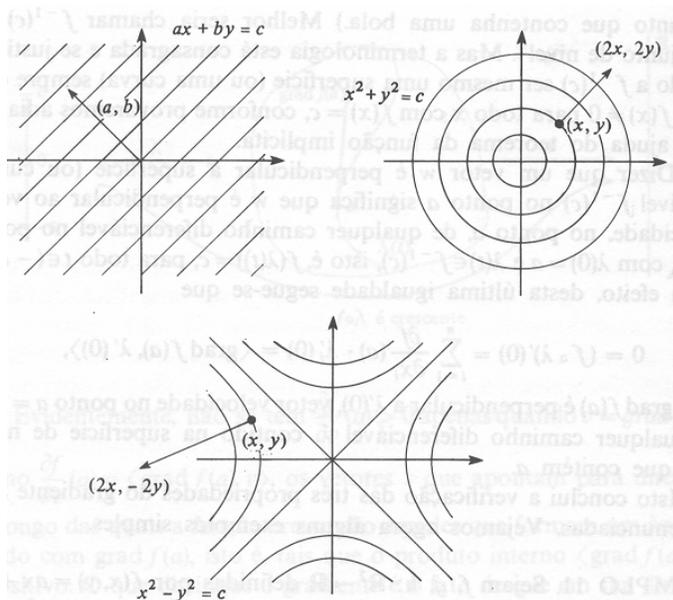


Figura 2.

$\frac{\partial f}{\partial x_i} : U \rightarrow \mathbb{R}$  no ponto  $x \in U$  será indicada por

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \frac{\partial}{\partial x_j} \left( \frac{\partial f}{\partial x_i} \right)(x), \quad i, j = 1, \dots, n.$$

Se essas derivadas parciais de segunda ordem existirem em cada ponto  $x \in U$ , teremos  $n^2$  funções  $\frac{\partial^2 f}{\partial x_j \partial x_i} : U \rightarrow \mathbb{R}$ . Quando tais funções forem contínuas, diremos que  $f$  é de classe  $C^2$  e escreveremos  $f \in C^2$ .

Em geral, a mera existência das derivadas parciais de segunda ordem em todos os pontos onde  $f$  está definida não assegura que se tenha

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j},$$

como se vê no exemplo abaixo.

**Exemplo 3.** Seja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}$  quando  $x^2 +$

$y^2 \neq 0$  e  $f(0, 0) = 0$ . Para todo  $y \neq 0$  tem-se  $f(0, y) = 0$ , logo

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

Portanto

$$\frac{\partial^2 f}{\partial y \partial x}(0, 0) = \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial x}(0, y) \right) = -1.$$

Um cálculo análogo mostra que  $\frac{\partial^2 f}{\partial x \partial y}(0, 0) = 1$ . Logo  $\frac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \frac{\partial^2 f}{\partial y \partial x}(0, 0)$ .

Em todo ponto  $x \in U$  onde existem as derivadas parciais de segunda ordem da função  $f : U \rightarrow \mathbb{R}$ , os números  $h_{ij}(x) = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$  formam uma matriz

$\mathbf{h}(x) = [h_{ij}(x)]$ , chamada a *matriz hessiana* da função  $f$ . O Teorema de Schwarz afirma que se  $f$  é de classe  $C^2$  então a matriz hessiana de  $f$  é simétrica.

A demonstração do Teorema de Schwarz se baseia num resultado, atribuído a Leibniz, segundo o qual é permitido derivar sob o sinal de integral, desde que o resultado da derivação seja uma função contínua. Por sua vez, a demonstração do Teorema de Leibniz utiliza o lema abaixo, que poderia estar no Capítulo 1 mas é colocado aqui para deixar claro como cada proposição depende da anterior.  $\triangleleft$

**Lema 1.** *Sejam  $X \subset \mathbb{R}^m$  um conjunto arbitrário e  $K \subset \mathbb{R}^n$  compacto. Fixemos  $x_0 \in X$ . Se  $f : X \times K \rightarrow \mathbb{R}^p$  é contínua então, para todo  $\varepsilon > 0$  dado, pode-se obter  $\delta > 0$  tal que  $x \in X$  e  $|x - x_0| < \delta$  implicam  $|f(x, t) - f(x_0, t)| < \varepsilon$ , seja qual for  $t \in K$ .*

*Demonstração.* Do contrário existiriam  $\varepsilon > 0$  e seqüências de pontos  $x_k \in X$  e  $t_k \in K$  tais que  $|x_k - x_0| < 1/k$  e  $|f(x_k, t_k) - f(x_0, t_k)| \geq \varepsilon$ . Passando a uma subsequência, se necessário, podemos admitir que  $\lim t_k = t_0 \in K$ . Como, evidentemente,  $\lim x_k = x_0$ , a continuidade de  $f$  nos daria  $\varepsilon \leq \lim |f(x_k, t_k) - f(x_0, t_k)| = |f(x_0, t_0) - f(x_0, t_0)|$ , uma contradição.  $\square$

**Teorema 3 (Derivação sob o sinal de integral).** *Dado  $U \subset \mathbb{R}^n$  aberto, seja  $f : U \times [a, b]$  contínua, tal que a  $i$ -ésima derivada parcial  $\frac{\partial f}{\partial x_i}(x, t)$  existe para todo ponto  $(x, t) \in U \times [a, b]$  e a função  $\partial f / \partial x_i : U \times [a, b] \rightarrow \mathbb{R}$ , assim definida, é contínua. Então a função  $\varphi : U \rightarrow \mathbb{R}$ , dada por  $\varphi(x) = \int_a^b f(x, t) dt$ , possui a  $i$ -ésima derivada parcial em cada ponto  $x \in U$ , sendo  $\frac{\partial \varphi}{\partial x_i}(x) = \int_a^b \frac{\partial f}{\partial x_i}(x, t) dt$ . Em suma: pode-se derivar sob o sinal de integral, desde que o integrando resultante seja uma função contínua.*

*Demonstração.* Pelo Teorema do Valor Médio para funções de uma variável, se  $x$  e  $x + se_i$  pertencem a  $U$  então existe  $\theta \in (0, 1)$  tal que

$$\begin{aligned} & \frac{\varphi(x + se_i) - \varphi(x)}{s} - \int_a^b \frac{\partial f}{\partial x_i}(x, t) dt \\ &= \int_a^b \left[ \frac{f(x + se_i, t) - f(x, t)}{s} - \frac{\partial f}{\partial x_i}(x, t) \right] dt \\ &= \int_a^b \left[ \frac{\partial f}{\partial x_i}(x + \theta se_i, t) - \frac{\partial f}{\partial x_i}(x, t) \right] dt. \end{aligned}$$

Pelo Lema, dado  $\varepsilon > 0$  arbitrariamente, podemos achar  $\delta > 0$  tal que

$$|s| < \delta \implies \left| \frac{\partial f}{\partial x_i}(x + \theta se_i, t) - \frac{\partial f}{\partial x_i}(x, t) \right| < \frac{\varepsilon}{b-a},$$

seja qual for  $t \in [a, b]$ . Então  $|s| < \delta$  implica

$$\left| \frac{\varphi(x + se_i) - \varphi(x)}{s} - \int_a^b \frac{\partial f}{\partial x_i}(x, t) dt \right| < \varepsilon,$$

o que demonstra o teorema.  $\square$

**Teorema 4 (Schwarz).** Se  $f : U \rightarrow \mathbb{R}$  é de classe  $C^2$  então  $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$ .

*Demonstração.* Sem perda de generalidade, podemos supor que  $U = I \times J$  é um retângulo em  $\mathbb{R}^2$ . Fixando  $b \in J$ , o Teorema Fundamental do Cálculo nos diz que, para todo  $(x, y) \in U$ , tem-se

$$f(x, y) = f(x, b) + \int_b^y \frac{\partial f}{\partial y}(x, t) dt.$$

Como  $\partial^2 f / \partial x \partial y$  é contínua, podemos derivar sob o sinal de integral, logo

$$\frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial x}(x, b) + \int_b^y \frac{\partial^2 f}{\partial x \partial y}(x, t) dt.$$

Em seguida, derivamos em relação a  $y$  e obtemos

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

Mais geralmente, para cada inteiro  $k \geq 1$ , podemos considerar as derivadas parciais de ordem  $k$  de uma função  $f : U \rightarrow \mathbb{R}$ , definida no aberto  $U \subset \mathbb{R}^n$ . Por exemplo, para

$$1 \leq i, j, k \leq n, \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(a) \quad \text{significa} \quad \frac{\partial}{\partial x_i} \left( \frac{\partial^2 f}{\partial x_j \partial x_k} \right)(a).$$

Como toda permutação dos índices  $i_1, \dots, i_k$  pode ser obtida por meio de repetidas inversões de índices adjacentes, segue-se do Teorema de Schwarz que a derivada de ordem  $k$

$$\frac{\partial^k f}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_k}}(a)$$

não depende da ordem em que são feitas as derivações, desde que todas as derivadas de ordem  $k$  de  $f$  existam e sejam contínuas.

Uma função  $f : U \rightarrow \mathbb{R}$  que possui, em cada ponto de  $U$ , todas as derivadas parciais de ordem  $k$ , as quais são funções contínuas em  $U$ , chama-se uma *função de classe  $C^k$* . Escreve-se então  $f \in C^k$ . Quando  $f \in C^k$  para todo  $k = 1, 2, 3, \dots$ , diz-se que  $f$  é uma *função de classe  $C^\infty$* .

#### 4 A fórmula de Taylor

A fórmula de Taylor, que estabeleceremos aqui em sua versão restrita aos termos de até segunda ordem, é fundamental para o estudo do comportamento de uma função de classe  $C^2$  na proximidade de um ponto crítico. Ela se baseia no lema abaixo.

**Lema 2.** *Seja  $r : B \rightarrow \mathbb{R}$  de classe  $C^2$  na bola aberta  $B \subset \mathbb{R}^n$ , de centro 0.*

*Se  $r(0) = \frac{\partial r}{\partial x_i}(0) = \frac{\partial^2 r}{\partial x_i \partial x_j}(0) = 0$  para quaisquer  $i, j = 1, \dots, n$ , então*

$$\lim_{v \rightarrow 0} \frac{r(v)}{|v|^2} = 0.$$

*Demonstração.* Sendo  $r : B \rightarrow \mathbb{R}$  uma função de classe  $C^1$  (portanto diferenciável) que se anula, juntamente com todas as suas derivadas  $\partial r / \partial x_i$ , no ponto  $v = 0$ , segue-se da definição de função diferenciável que  $\lim_{v \rightarrow 0} r(v)/|v| = 0$ . Pelo Teorema do Valor Médio (Corolário 3 do Teorema 2), para cada  $v = (\alpha_1, \dots, \alpha_n) \in B$  existe  $\theta$  tal que  $0 < \theta < 1$  e

$$r(v) = \sum_{i=1}^n \frac{\partial r}{\partial x_i}(\theta v) \cdot \alpha_i, \quad \text{logo} \quad \frac{r(v)}{|v|^2} = \sum_{i=1}^n \frac{\frac{\partial r}{\partial x_i}(\theta v)}{|v|} \cdot \frac{\alpha_i}{|v|}.$$

Como cada derivada parcial  $\partial r / \partial x_i$  se anula, juntamente com todas as suas derivadas  $\partial^2 r / \partial x_j \partial x_i$ , no ponto 0, resulta da nossa observação inicial que

$$\lim_{v \rightarrow 0} \left[ \frac{\partial r}{\partial x_i}(\theta v) / |v| \right] = 0 \text{ para todo } i = 1, \dots, n.$$

Além disso, cada quociente  $\alpha_i / |v|$  tem valor absoluto  $\leq 1$ . Por conseguinte  $\lim_{v \rightarrow 0} \frac{r(v)}{|v|^2} = 0$ .  $\square$

**Teorema 5 (Fórmula de Taylor).** *Seja  $f : U \rightarrow \mathbb{R}$  de classe  $C^2$  no aberto  $U \subset \mathbb{R}^n$ . Fixado  $a \in U$ , para todo  $v = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$  tal que  $a + v \in U$ , escrevamos*

$$f(a + v) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \alpha_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j + r(v),$$

as derivadas sendo calculadas no ponto  $a$ . Então  $\lim_{v \rightarrow 0} \frac{r(v)}{|v|^2} = 0$ .

*Demonstração.* De acordo com o Lema, devemos demonstrar que

$$r(v) = f(a + v) - f(a) - \sum_{i=1}^n \frac{\partial f}{\partial x_i} \cdot \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j$$

se anula, juntamente com suas derivadas parciais de primeira e segunda ordem, no ponto  $v = 0$ .

Para fazer o cálculo, começamos lembrando que, na expressão de  $r(v)$ , as variáveis independentes são as coordenadas  $\alpha_1, \dots, \alpha_n$  de  $v$ . É em relação a elas que as derivadas parciais de  $r$  devem ser tomadas, embora continuemos escrevendo  $\partial r / \partial x_i$  e  $\partial^2 r / \partial x_i \partial x_j$ . Observemos também que, no somatório duplo que ocorre na definição de  $r(v)$ , cada par de variáveis  $\alpha_i, \alpha_j$  aparece em duas parcelas iguais, a saber,  $\frac{\partial^2 f}{\partial x_j \partial x_i} \cdot \alpha_j \alpha_i$  e  $\frac{\partial^2 f}{\partial x_i \partial x_j} \alpha_i \alpha_j$ . Levando isto em conta, temos:

$$\frac{\partial r}{\partial x_j}(v) = \frac{\partial f}{\partial x_j}(a + v) - \frac{\partial f}{\partial x_j}(a) - \sum_{i=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \cdot \alpha_i.$$

Derivando outra vez, vem:

$$\frac{\partial^2 r}{\partial x_i \partial x_j}(v) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a + v) - \frac{\partial^2 f}{\partial x_i \partial x_j}(a).$$

Conseqüentemente  $r(0) = 0$ ,  $\frac{\partial r}{\partial x_i}(0) = 0$  e  $\frac{\partial^2 f}{\partial x_i \partial x_j}(0) = 0$  para quaisquer  $i, j = 1, \dots, n$ .  $\square$

**Observação.** Se pusermos  $\rho(v) = \frac{r(v)}{|v|^2}$  quando  $v \neq 0$  e  $\rho(0) = 0$ , a fórmula de Taylor se escreve assim:

$$f(a+v) - f(a) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \alpha_i + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j} \cdot \alpha_i \alpha_j + \rho(v) \cdot |v|^2,$$

onde  $\lim_{v \rightarrow 0} \rho(v) = 0$ .

## 5 Pontos críticos

Uma *forma quadrática*  $H : \mathbb{R}^n \rightarrow \mathbb{R}$  é uma função cujo valor no vetor  $v = (\alpha_1, \dots, \alpha_n)$  é  $\sum_{i,j=1}^n h_{ij} \alpha_i \alpha_j$ , onde  $[h_{ij}]$  é uma matriz simétrica  $n \times n$ . O valor da forma quadrática  $H$  no vetor  $v$  será indicado com a notação  $H \cdot v^2$ . Portanto

$$H \cdot v^2 = \sum_{i,j=1}^n h_{ij} \alpha_i \alpha_j \quad \text{quando } v = (\alpha_1, \dots, \alpha_n).$$

Se  $t \in \mathbb{R}$  então  $H \cdot (tv)^2 = t^2 \cdot H \cdot v^2$ .

A forma quadrática  $H$  chama-se *não-negativa* quando  $H \cdot v^2 \geq 0$  para todo  $v \in \mathbb{R}^n$ , *positiva* quando  $H \cdot v^2 > 0$  para todo  $v \neq 0$  em  $\mathbb{R}^n$  e *indefinida* quando existem  $v, w \in \mathbb{R}^n$  tais que  $H \cdot v^2 > 0$  e  $H \cdot w^2 < 0$ . De modo análogo se definem forma quadrática *negativa* e *não-positiva*. Quando  $H$  é positiva ou negativa, diz-se que ela é *definida*.

**Exemplo 4.** A forma quadrática  $H : \mathbb{R}^n \rightarrow \mathbb{R}$ , onde  $H \cdot v^2 = \langle v, v \rangle$ , é positiva. Como  $\langle v, v \rangle = \alpha_1^2 + \dots + \alpha_n^2$ , a matriz de  $H$  é a identidade. Para todo  $k \in [1, n]$ ,  $H \cdot v^2 = \alpha_1^2 + \dots + \alpha_k^2$  é uma forma quadrática não-negativa em  $\mathbb{R}^n$ . Por outro lado, se pusermos  $H \cdot v^2 = \alpha_1^2 + \dots + \alpha_k^2 - \alpha_{k+1}^2 - \dots - \alpha_n^2$  com  $0 < k < n$ , teremos uma forma quadrática indefinida. Evidentemente, se  $H$  é positiva (respect. não-negativa) então  $-H$  é negativa (respect. não-positiva).  $\triangleleft$

Seja  $H : \mathbb{R}^n \rightarrow \mathbb{R}$  uma forma quadrática cuja matriz é  $[h_{ij}]$ .

Se chamarmos de  $H_0 : \mathbb{R}^n \rightarrow \mathbb{R}^n$  o operador linear cuja matriz na base canônica de  $\mathbb{R}^n$  é também  $[h_{ij}]$ , vemos imediatamente que  $H \cdot v^2 = \langle H_0 \cdot v, v \rangle$  para todo  $v \in \mathbb{R}^n$ . Como a matriz  $[h_{ij}]$  do operador  $H_0$  na base canônica é simétrica,  $H_0$  é

auto-adjunto. Reciprocamente, para qualquer operador auto-adjunto  $H_0 : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , a função  $H : \mathbb{R}^n \rightarrow \mathbb{R}$ , dada por  $H \cdot v^2 = \langle H_0 \cdot v, v \rangle$ , é uma forma quadrática. Quando  $H$  é definida, o operador  $H_0$  é invertível pois  $\langle H_0 \cdot v, v \rangle \neq 0$  para todo  $v \neq 0 \Rightarrow H_0 \cdot v \neq 0$  para todo  $v \neq 0$ .

Dada a função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^2$  no aberto  $U \subset \mathbb{R}^n$ , a *forma quadrática hessiana*  $H(x)$  de  $f$  no ponto  $x \in U$  é aquela cuja matriz é  $[h_{ij}] = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right]$ . Assim, para todo  $v = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ , tem-se

$$H(v) \cdot v^2 = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \cdot \alpha_i \alpha_j.$$

A forma hessiana é usada para determinar a natureza dos pontos críticos da função  $f$ .

Diz-se que  $a \in U$  é um ponto de *máximo local* da função  $f : U \rightarrow \mathbb{R}$  quando existe  $\delta > 0$  tal que  $f(x) \leq f(a)$  para todo  $x \in U \cap B(a; \delta)$ . Analogamente se define um ponto de *mínimo local*. Um ponto  $a$ , de máximo (ou de mínimo) local de uma função diferenciável  $f$ , é um ponto crítico de  $f$ . Com efeito, para todo  $i = 1, \dots, n$ , se  $\delta > 0$  é suficientemente pequeno então a função  $\varphi : (-\delta, \delta) \rightarrow \mathbb{R}$ , dada por  $\varphi(t) = f(a + te_i)$ , está bem definida e possui um máximo (ou mínimo) local no ponto  $t = 0$ . Logo  $0 = \varphi'(0) = \frac{\partial f}{\partial x_i}(a)$ ,  $i = 1, \dots, n$ .

**Exemplo 5.** A origem  $0 \in \mathbb{R}^2$  é ponto crítico das três funções  $f, g, h : \mathbb{R}^2 \rightarrow \mathbb{R}$ , definidas por  $f(x, y) = x^2 + y^2$ ,  $g(x, y) = -x^2 - y^2$  e  $h(x, y) = x^2 - y^2$ . Para  $f$ , a origem é um ponto de máximo, para  $g$  de mínimo e para  $h$  não é máximo nem mínimo pois em qualquer disco de centro 0 a função  $h$  assume valores maiores e menores do que  $0 = h(0, 0)$ .  $\triangleleft$

**Teorema 6.** *Seja  $a \in U$  um ponto crítico da função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^2$ .*

- Se a forma quadrática hessiana  $H(a)$  for positiva então  $a$  é um ponto de mínimo local de  $f$ .*
- Se  $H(a)$  for negativa então  $a$  é um ponto de máximo local.*
- Se  $H(a)$  for indefinida, então  $a$  não é ponto de máximo nem de mínimo local de  $f$ .*

*Demonstração.* a) Por simplicidade, escrevemos  $H$  em vez de  $H(a)$ . Pelo Teorema de Weierstrass, a função contínua positiva  $H$  assume um valor mínimo  $2c > 0$  no conjunto compacto  $S^{n-1}$ . Noutras palavras, existe  $c > 0$  tal que  $H \cdot u^2 \geq 2c$

para todo vetor  $u \in \mathbb{R}^n$  com  $|u| = 1$ . Como  $a$  é um ponto crítico de  $f$ , a fórmula de Taylor se resume a

$$f(a+v) - f(a) = \frac{1}{2}H \cdot v^2 + \rho(v)|v|^2 \quad \text{com} \quad \lim_{v \rightarrow 0} \rho(v) = 0.$$

Como  $v/|v|$  é um vetor unitário (pertencente a  $S^{n-1}$ ), temos

$$\frac{1}{2}H \cdot v^2 = \frac{|v|^2}{2}H \cdot \left(\frac{v}{|v|}\right)^2 \geq \frac{|v|^2}{2} \cdot 2c = |v|^2 \cdot c.$$

Portanto  $f(a+v) - f(a) \geq |v|^2(c + \rho(v))$ .

Pela definição de limite, existe  $\delta > 0$  tal que  $a+v \in U$  e  $0 < |v| < \delta$  implicam  $|\rho(v)| < c$  e conseqüentemente  $c + \rho(v) > 0$ . Logo  $f(a+v) - f(a) > 0$ , isto é,  $f(a) < f(a+v)$  para todo  $v$  tal que  $a+v \in U$  e  $0 < |v| < \delta$ . Assim,  $a$  é um ponto de mínimo local para  $f$ .

b) Segue as mesmas linhas do caso anterior.

c) Dado  $v \in \mathbb{R}^n$ , tem-se  $a+tv \in U$  para todo  $t$  suficientemente pequeno. Então, lembrando que  $H \cdot (tv)^2 = t^2 \cdot H \cdot v^2$ , temos

$$f(a+tv) - f(a) = t^2 \cdot |v|^2 \cdot [H \cdot v^2 + \rho(tv)], \quad \text{com} \quad \lim_{t \rightarrow 0} \rho(tv) = 0.$$

Segue-se, como acima, que para todo  $t$  suficientemente pequeno,  $f(a+tv) - f(a)$  tem o mesmo sinal que  $H \cdot v^2$ . Assim, se  $H$  é indefinida, com  $H \cdot v^2 > 0$  e  $H \cdot w^2 < 0$ , em qualquer bola de centro  $a$  existem pontos  $a+tv$  e  $a+tw$  tais que  $f(a+tv) > f(a)$  e  $f(a+tw) < f(a)$ . Portanto  $f$  não tem máximo nem mínimo local do ponto  $a$ .  $\square$

**Corolário 6.** Se a função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^2$ , possui um mínimo (respect. máximo) local no ponto  $a \in U$  então a forma quadrática hessiana de  $f$  é não-negativa (respect. não-positiva) nesse ponto.

Com efeito, se fosse  $H \cdot v_0^2 < 0$  para algum  $v_0 \in \mathbb{R}^n$ , teríamos  $f(a+tv_0) < f(a)$  para todo  $t$  suficientemente pequeno, e então  $a$  não seria um ponto de mínimo local. Mesmo argumento para máximo local.  $\square$

**Exemplo 6.** Pela demonstração acima, vê-se que quando a forma quadrática hessiana é positiva (respect. negativa) no ponto  $a$  então  $a$  é um ponto de mínimo (respect. máximo) local *estrito*, isto é, numa pequena bola de centro  $a$  não há outros pontos  $x$  com  $f(x) = f(a)$ . Por exemplo, a origem é um ponto de mínimo estrito da função  $f(x, y) = x^2 + y^2$  mas todos os pontos  $(x, 0)$  do eixo das abcissas são pontos de mínimo não-estritos da função  $g(x, y) = y^2$ . (O domínio de ambas

as funções  $f, g \in \mathbb{R}^2$ .) A forma hessiana de  $f$  na origem de  $\mathbb{R}^2$  é  $H \cdot v^2 = 2\alpha^2 + 2\beta^2$  se  $v = (\alpha, \beta)$  enquanto a de  $g$  é  $K \cdot v^2 = 2\alpha^2$ . Vemos que  $H$  é positiva e  $K$  é apenas não-negativa. Já a forma hessiana da função  $h(x, y) = x^2 - y^2$  na origem é  $L \cdot v^2 = 2\alpha^2 - 2\beta^2$ , que é indefinida. Por isso a origem é um ponto crítico que não é máximo nem mínimo local (ponto de sela).  $\triangleleft$

**Exemplo 7.** Poder-se-ia indagar se vale a recíproca do corolário acima. A resposta é negativa. A função  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ , dada por  $f(x, y) = x^2 + y^3$  tem a origem de  $\mathbb{R}^2$  como ponto crítico, no qual a forma hessiana é  $H \cdot v^2 = 2\alpha^2$ , para  $v = (\alpha, \beta)$ . A forma  $H$  é não-negativa porém a origem não é um ponto de mínimo local de  $f$ .  $\triangleleft$

Neste ponto, cabe a pergunta: de que modo podemos determinar se uma dada forma quadrática é positiva, negativa, etc? O método de completar o quadrado, devido a Lagrange, responde a questão. Este método, que se baseia na observação óbvia de que  $a^2 + 2ab = (a + b)^2 - b^2$ , consiste em efetuar sucessivas mudanças de variáveis, visando eliminar, na expressão da forma quadrática  $H$ , os termos como  $xy, xz, yz$ , etc, deixando apenas parcelas do tipo  $x^2, y^2, z^2$  etc.

Os exemplos a seguir ilustram o método de completar os quadrados.

**Exemplo 8.** Seja a forma quadrática  $H(x, y) = x^2 - xy + y^2$  em  $\mathbb{R}^2$ . Completando o quadrado, temos

$$x^2 - xy = x^2 - 2x \cdot \frac{y}{2} = \left(x - \frac{y}{2}\right)^2 - \frac{y^2}{4}.$$

Logo

$$H(x, y) = \left(x - \frac{y}{2}\right)^2 - \frac{y^2}{4} + y^2 = \left(x - \frac{y}{2}\right)^2 + \frac{3}{4}y^2.$$

Portanto  $H(x, y) = s^2 + t^2$  com  $s = x - y/2$  e  $t = \sqrt{3/2} \cdot y$ . Assim, a forma  $H$  é positiva. O mesmo processo, aplicado à forma  $K$ , onde  $K(x, y) = x^2 + 3xy + y^2$  nos dá

$$K(x, y) = \left(x + \frac{3}{2}y\right)^2 - \frac{9}{4}y^2 + y^2,$$

ou seja,

$$K(x, y) = \left(x + \frac{3}{2}y\right)^2 - \frac{5}{4}y^2 = s^2 - t^2 \quad \text{com} \quad s = x + \frac{3}{2}y \quad \text{e} \quad t = \frac{\sqrt{5}}{2}y.$$

Portanto a forma  $K$  é indefinida.  $\triangleleft$

**Exemplo 9.** Seja em  $\mathbb{R}^3$  a forma quadrática  $H(x, y, z) = x^2 + y^2 + z^2 + 3xy + 3xz + 4yz$ . Agrupando os termos que contêm  $x$ , temos:

$$\begin{aligned} x^2 + 3xy + 3xz &= x^2 + 2x \cdot \frac{3}{2}(y+z) = \left[ x + \frac{3}{2}(y+z) \right]^2 - \frac{9}{4}(y+z)^2 \\ &= s^2 - \frac{9}{4}y^2 - \frac{9}{4}z^2 - \frac{9}{2}yz, \quad \text{com } s = x + \frac{3}{2}(y+z). \end{aligned}$$

Logo

$$\begin{aligned} H(x, y, z) &= s^2 + y^2 + z^2 - \frac{9}{4}y^2 - \frac{9}{4}z^2 - \frac{9}{2}yz + 4yz \\ &= s^2 - \frac{5}{4}y^2 - \frac{5}{4}z^2 - \frac{1}{2}yz. \end{aligned}$$

Agrupando os termos que contêm  $y$ :

$$\begin{aligned} -\frac{5}{4}y^2 - \frac{1}{2}yz &= -\frac{5}{4} \left( y^2 + \frac{2}{5}yz \right) = -\frac{5}{4} \left( y + \frac{1}{5}z \right)^2 + \frac{1}{20}z^2 \\ &= -\frac{5}{4}t^2 + \frac{1}{20}z^2 \quad \text{com } t = y + \frac{1}{5}z. \end{aligned}$$

Portanto:

$$H(x, y, z) = s^2 - \frac{5}{4}t^2 + \frac{1}{20}z^2 - \frac{5}{4}z^2 = s^2 - \frac{5}{4}t^2 - \frac{6}{5}z^2.$$

Concluimos então que a forma quadrática  $H$  é indefinida. Com efeito, para  $z = 0$  temos

$$H(x, y, 0) = s^2 - \frac{5}{4}t^2 = \left( x + \frac{3}{2}y \right)^2 - \frac{5}{4}y^2.$$

Logo  $H(x, 0, 0) = x^2$  e, em particular,  $H(1, 0, 0) = 1$ , enquanto  $H(-3/2, 1, 0) = -5/4$ .  $\triangleleft$

## 6 Funções convexas

Seja  $C \subset \mathbb{R}^n$  um conjunto convexo. Uma função  $f : C \rightarrow \mathbb{R}$  chama-se *convexa* quando, para quaisquer  $x, y \in C$  e  $t \in [0, 1]$ , tem-se

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y).$$

Alternativamente:  $f$  é convexa quando, para quaisquer  $x, y \in C$  e  $\alpha, \beta \in [0, 1]$  com  $\alpha + \beta = 1$ , tem-se  $f(\alpha x + \beta y) \leq \alpha \cdot f(x) + \beta \cdot f(y)$ .

Diz-se que  $f: C \rightarrow \mathbb{R}$  é *côncava* quando  $-f$  é convexa. Isto equivale a dizer que, para quaisquer  $x, y \in C$  e  $t \in [0, 1]$  tem-se  $f((1-t)x + ty) \geq (1-t)f(x) + tf(y)$ . Todos os resultados a seguir estabelecidos para funções convexas valem, com as óbvias modificações, para funções côncavas.

A combinação linear  $\alpha_1 v_1 + \dots + \alpha_k v_k$  chama-se uma *combinação convexa* de  $v_1, \dots, v_k \in \mathbb{R}^n$  quando  $\alpha_1 + \dots + \alpha_k = 1$  e  $\alpha_i \geq 0$  para  $i = 1, \dots, k$ .

**Teorema 7.** Se  $C \in \mathbb{R}^n$  é convexo e  $v_1, \dots, v_k \in C$  então toda combinação convexa  $\alpha_1 v_1 + \dots + \alpha_k v_k$  pertence a  $C$ . Além disso, se  $f: C \rightarrow \mathbb{R}$  é uma função convexa, tem-se

$$f\left(\sum_{i=1}^k \alpha_i v_i\right) \leq \sum_{i=1}^k \alpha_i \cdot f(v_i).$$

*Demonstração.* Para  $k = 1$  é óbvio e para  $k = 2$  segue-se da definição de conjunto convexo que a combinação convexa de  $k$  elementos de  $C$  ainda pertence a  $C$ . Supondo este fato verdadeiro para um certo  $k$ , escrevamos uma combinação convexa dos elementos  $v_1, \dots, v_{k+1} \in C$  sob a forma

$$\sum_{i=1}^{k+1} \alpha_i v_i = \sum_{i=1}^k \alpha_i v_i + \alpha_{k+1} v_{k+1}.$$

Sem perda de generalidade, podemos admitir que  $\alpha_{k+1} \neq 1$ . Então, pondo  $\alpha = \sum_{i=1}^k \alpha_i$ , temos  $\alpha_{k+1} = 1 - \alpha$  e  $\alpha \neq 0$ . Pela hipótese de indução, levando em conta que  $\sum_{i=1}^k \frac{\alpha_i}{\alpha} = 1$ , vemos que  $v = \sum_{i=1}^k \frac{\alpha_i}{\alpha} v_i$  pertence a  $C$ . Logo

$$\sum_{i=1}^k \alpha_i v_i = \alpha v + (1 - \alpha) v_{k+1} \in C, \quad \text{pois } C \text{ é convexo.}$$

A segunda parte também se prova por indução, pois

$$\begin{aligned} f\left(\sum_{i=1}^{k+1} \alpha_i v_i\right) &= f\left(\sum_{i=1}^k \alpha_i v_i + \alpha_{k+1} v_{k+1}\right) \\ &= f\left(\alpha \cdot \sum_{i=1}^k \frac{\alpha_i}{\alpha} v_i + (1 - \alpha) v_{k+1}\right) \end{aligned}$$

$$\begin{aligned} &\leq \alpha \cdot f\left(\sum_{i=1}^k \frac{\alpha_i}{\alpha} v_i\right) + (1 - \alpha)f(v_{k+1}) \\ &\leq \alpha \cdot \sum_{i=1}^k \frac{\alpha_i}{\alpha} f(v_i) + (1 - \alpha)f(v_{k+1}) = \sum_{i=1}^{k+1} \alpha_i f(v_i). \quad \square \end{aligned}$$

**Teorema 8.** *Seja  $C \subset \mathbb{R}^n$  convexo. A fim de que a função  $f : C \rightarrow \mathbb{R}$  seja convexa, é necessário e suficiente que, para quaisquer  $a, b \in C$ , a função  $\varphi : [0, 1] \rightarrow \mathbb{R}$ , definida por  $\varphi(t) = f(a + tv)$ ,  $v = b - a$ , seja convexa.*

*Equivalentemente:  $f : C \rightarrow \mathbb{R}$  é convexa se, e somente se, sua restrição a qualquer segmento de reta  $[a, b] \subset C$  é convexa.*

*Demonstração.* Se  $f$  é convexa então, para  $s, t, \alpha \in [0, 1]$  temos

$$\begin{aligned} \varphi((1 - \alpha)s + \alpha t) &= f(a + [(1 - \alpha)s + \alpha t]v) \\ &= f[(1 - \alpha) \cdot (a + sv) + \alpha \cdot (a + tv)] \\ &\leq (1 - \alpha)f(a + sv) + \alpha f(a + tv) \\ &= (1 - \alpha)\varphi(s) + \alpha\varphi(t) \end{aligned}$$

logo  $\varphi$  é convexa.

Reciprocamente se todas as funções  $\varphi$ , definidas do modo acima, são convexas então, dados  $x, y \in C$  e  $\alpha \in [0, 1]$ , pomos  $\varphi(t) = f(x + t(y - x))$  e temos:

$$\begin{aligned} f((1 - \alpha)x + \alpha y) &= f(x + \alpha(y - x)) = \varphi(\alpha) = \varphi((1 - \alpha) \cdot 0 + \alpha \cdot 1) \\ &\leq (1 - \alpha) \cdot \varphi(0) + \alpha \cdot \varphi(1) = (1 - \alpha) \cdot f(x) + \alpha \cdot f(y), \end{aligned}$$

portanto  $f$  é convexa. □

Como aplicação do Teorema 8, mostremos que se  $f : U \rightarrow \mathbb{R}$  é uma função convexa e o conjunto convexo  $U \subset \mathbb{R}^n$  é aberto então, para cada  $a \in U$ , existe a derivada de Gâteaux

$$\frac{\partial f}{\partial v+}(a) = \lim_{t \rightarrow 0+} \frac{f(a + tv) - f(a)}{t}.$$

Com efeito, a função  $\varphi : [0, 1] \rightarrow \mathbb{R}$ , definida por  $\varphi(t) = f(a + tv)$  é convexa, portanto existe a derivada à direita  $\varphi'_+(0)$  (veja Vol. 1, pág. 106). Mas, como se vê facilmente,  $\varphi'_+(0) = \frac{\partial f}{\partial v+}(a)$ .

Daí se conclui, como no Vol. 1, que toda função convexa definida num subconjunto aberto de  $\mathbb{R}$  é contínua. Este resultado continua válido em  $\mathbb{R}^n$  com  $n > 1$  (ver Apêndice deste capítulo) porém não decorre da existência da derivada de Gâteaux, pois uma função em  $\mathbb{R}^n$  pode ser contínua ao longo de cada reta que passa por um ponto  $a$  sem que seja necessariamente contínua nesse ponto.

**Teorema 9.** *Seja  $f : U \rightarrow \mathbb{R}$  definida no aberto convexo  $U \subset \mathbb{R}^n$ . Então:*

- a) *O conjunto  $E(f) = \{(x, y) \in U \times \mathbb{R}; y \geq f(x)\} \subset \mathbb{R}^{n+1}$ , chamado o epigráfico de  $f$ , é convexo se, e somente se,  $f$  é convexa.*
- b) *Supondo-a de classe  $C^1$ , a função  $f$  é convexa se, e somente se, para  $a, a + v \in U$  quaisquer, tem-se*

$$f(a + v) \geq f(a) + \langle \text{grad } f(a), v \rangle.$$

- c) *Quando é de classe  $C^2$ , a função  $f$  é convexa se, e somente se, sua forma quadrática hessiana é não-negativa em todos os pontos de  $U$ .*

*Demonstração.* (a) Seja  $E(f)$  convexo. Para mostrar que  $f$  é convexa, tomamos  $x, x' \in U$  e  $\alpha \in [0, 1]$ . Então  $(x, f(x))$  e  $(x', f(x'))$  pertencem a  $E(f)$ , portanto  $((1 - \alpha)x + \alpha x', (1 - \alpha) \cdot f(x) + \alpha \cdot f(x')) \in E(f)$ . Isto significa que  $(1 - \alpha) \cdot f(x) + \alpha \cdot f(x') \geq f[(1 - \alpha)x + \alpha x']$ , logo  $f$  é convexa. Reciprocamente, supondo  $f$  convexa, sejam  $z = (x, y), z' = (x', y')$  pontos em  $E(f)$  e  $\alpha \in [0, 1]$ . Então  $y \geq f(x)$  e  $y' \geq f(x')$  e daí  $(1 - \alpha)y + \alpha y' \geq (1 - \alpha) \cdot f(x) + \alpha \cdot f(x') \geq f((1 - \alpha)x + \alpha x')$ , a última desigualdade devendo-se à convexidade de  $f$ . Logo  $(1 - \alpha)z + \alpha z' = ((1 - \alpha)x + \alpha x', (1 - \alpha)y + \alpha y')$  pertence a  $E(f)$ , ou seja,  $E(f)$  é um conjunto convexo.

(b) Suponhamos  $f : U \rightarrow \mathbb{R}$  convexa, de classe  $C^1$ . Pelo Teorema 8, se  $a$  e  $a + v$  pertencem a  $U$  então a função  $\varphi : [0, 1] \rightarrow \mathbb{R}$ , definida por  $\varphi(t) = f(a + tv)$ , é convexa. Portanto (v. Teorema 4, pág. 106, vol. 1) tem-se  $\varphi(1) \geq \varphi(0) + \varphi'(0)$ . Mas  $\varphi(1) = f(a + v)$ ,  $\varphi(0) = f(a)$  e  $\varphi'(0) = \langle \text{grad } f(a), v \rangle$ . Logo  $f(a + v) \geq f(a) + \langle \text{grad } f(a), v \rangle$ . Reciprocamente, suponhamos que esta igualdade valha para quaisquer  $a, a + v \in U$ . Então, pondo  $\varphi(t) = f(a + tv)$ , temos uma função  $\varphi : [0, 1] \rightarrow \mathbb{R}$  tal que  $\varphi'(t) = \langle \text{grad } f(a + tv), v \rangle$  para todo  $t \in [0, 1]$ . Ora, pela hipótese admitida sobre  $f$ , para quaisquer  $t, t_0 \in [0, 1]$ ,  $f(a + tv) = f(a + t_0v + (t - t_0)v) = f(a + t_0v + sv)$ , com  $s = t - t_0$ , logo

$$\begin{aligned} f(a + tv) &\geq f(a + t_0v) + \langle \text{grad } f(a + t_0v), sv \rangle \\ &= f(a + t_0v) + \langle \text{grad } f(a + t_0v), v \rangle(t - t_0) \end{aligned}$$

o que pode ser lido como  $\varphi(t) \geq \varphi(t_0) + \varphi'(t_0)(t - t_0)$ . Pelo Teorema 4, pág. 106, Volume 1, a função  $\varphi$  é convexa. O Teorema 8, acima, assegura então que  $f$  é convexa.

(c) Novamente, usamos o Teorema 8 acima, o qual permite reduzir a questão ao caso de uma função de uma variável, e então recaímos outra vez no Teorema 4 da pág. 106 do Volume 1. Com efeito, pondo  $\varphi(t) = f(x + tv)$ , com  $v = (\alpha_1, \dots, \alpha_n)$ , temos

$$\varphi'(t) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x) \alpha_i \quad \text{e} \quad \varphi''(t) = \sum_{i,j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \alpha_i \alpha_j = H(x) \cdot v^2.$$

Temos portanto as seguintes equivalências:  $H(x)$  é não-negativa para todo  $x \in U \iff \varphi''(t) \geq 0$  para quaisquer  $x, x + v \in U$  e  $t \in [0, 1] \iff$  todas as funções  $\varphi$  do tipo  $\varphi(t) = f(x + tv)$  são convexas  $\iff f : U \rightarrow \mathbb{R}$  é convexa.  $\square$

**Corolário 7.** *Todo ponto crítico  $a$  de uma função convexa  $f : U \rightarrow \mathbb{R}$  de classe  $C^1$  é um ponto de mínimo global, isto é,  $f(x) \geq f(a)$  para todo  $x \in U$ .*

### Apêndice: Continuidade das funções convexas

**Teorema 10.** *Seja  $U \subset \mathbb{R}^n$  um aberto convexo. Toda função convexa  $f : U \rightarrow \mathbb{R}$  é contínua.*

A demonstração do Teorema 10 se baseia nos dois lemas abaixo.

**Lema 3.** *Todo ponto de um bloco retangular  $B = \prod_{i=1}^n [a_i, b_i]$  é uma combinação convexa dos vértices desse bloco.*

*Demonstração:* (Por indução). Isto é óbvio para  $n = 1$ . Seja  $n > 1$ . Os vértices do bloco  $B$  são os  $2^n$  elementos do conjunto  $\prod_{i=1}^n \{a_i, b_i\}$ , os quais denotaremos por  $v_j$  ou  $\bar{v}_j$  conforme sua última coordenada seja da forma  $a_k$  ou  $b_k$ . Um ponto arbitrário do bloco  $B$  pode ser escrito como  $p = (x, y)$ , onde  $y \in [a_n, b_n]$  e  $x$  pertence ao bloco  $B' = \prod_{i=1}^{n-1} [a_i, b_i]$ , de dimensão  $n - 1$ . Pela hipótese de indução,  $x = \sum \alpha_j u_j$  é combinação convexa dos vértices  $u_j \in B'$ . Os vértices de  $B$  são  $v_j = (u_j, a_n)$  e  $\bar{v}_j = (u_j, b_n)$ . Pondo  $p_0 = (x, a_n)$  e  $p_1 = (x, b_n)$ , temos  $p_0 = \sum \alpha_j v_j$  e  $p_1 = \sum \alpha_j \bar{v}_j$  (já que  $\sum \alpha_j = 1$ ). Além disso,  $y = (1 - t)a_n + tb_n$ , com

$$t = \frac{y - a_n}{b_n - a_n},$$

logo

$$p = (1-t)p_0 + tp_1 = \sum (1-t)\alpha_j v_j + \sum t\alpha_j \bar{v}_j,$$

o que exprime o ponto arbitrário  $p$  do bloco  $B$  como combinação convexa dos vértices de  $B$ .  $\square$

**Lema 4.** *Toda função convexa  $f : U \rightarrow \mathbb{R}$ , definida num aberto convexo  $U \subset \mathbb{R}^n$ , é localmente majorada por uma constante.*

*Demonstração.* Seja  $A = \prod_{i=1}^n (a_i, b_i)$  o interior de um bloco retangular contido em  $U$ . Se indicarmos com  $w_j, j = 1, \dots, 2^n$ , os vértices de  $A$  teremos, para cada  $x \in A$ ,  $x = \sum \alpha_j w_j$  logo, pela convexidade de  $f$ ,  $f(x) \leq \sum \alpha_j \cdot f(w_j) \leq M$ , onde  $M = \max_j \{f(w_j)\}$ .  $\square$

*Demonstração do Teorema 10.* Para simplificar a escrita, a fim de provar a continuidade de  $f$  no ponto arbitrário  $a \in U$ , podemos admitir que  $a = 0$  e que  $f(0) = 0$ , pois o conjunto  $U_0 = \{x \in \mathbb{R}^n; a - x \in U\}$  é aberto, contém 0 e a função  $g : U_0 \rightarrow \mathbb{R}$ , definida por  $g(x) = f(a - x) - f(a)$ , cumpre  $g(0) = 0$ , é convexa e é contínua no ponto 0 se, e somente se,  $f$  é contínua no ponto  $a$ . Pelo Lema 4, existem  $c > 0$  e  $M > 0$  tais que  $|x| \leq c \Rightarrow f(x) \leq M$ . Seja dado  $\varepsilon > 0$ . Sem perda de generalidade, podemos supor que  $\varepsilon < M$ . A convexidade de  $f$  nos permite afirmar que

$$f\left(\frac{\varepsilon}{M}x\right) = f\left(\left(1 - \frac{\varepsilon}{M}\right) \cdot 0 + \frac{\varepsilon}{M}x\right) \leq \frac{\varepsilon}{M} \cdot f(x)$$

logo

$$f(x) \leq \frac{\varepsilon}{M} \cdot f\left(\frac{M}{\varepsilon}x\right).$$

Tomando  $\delta = \frac{\varepsilon c}{M}$ , vemos que

$$|x| < \frac{\varepsilon c}{M} \implies \left|\frac{M}{\varepsilon}x\right| < c \implies f\left(\frac{M}{\varepsilon}x\right) \leq M \implies f(x) \leq \varepsilon.$$

Além disso,

$$\begin{aligned} 0 &= f(0) = f\left(\frac{M}{M+\varepsilon}x + \frac{\varepsilon}{M+\varepsilon}\left(-\frac{M}{\varepsilon}x\right)\right) \\ &\leq \frac{M}{M+\varepsilon}f(x) + \frac{\varepsilon}{M+\varepsilon} \cdot f\left(-\frac{M}{\varepsilon}x\right). \end{aligned}$$

Simplificando, vem  $M \cdot f(x) + \varepsilon \cdot f(-Mx/\varepsilon) \geq 0$ , donde:

$$f(x) \geq \frac{\varepsilon}{M} \cdot (-f(-Mx/\varepsilon)) \geq \frac{\varepsilon}{M} \cdot (-M) = -\varepsilon.$$

Em resumo:  $|x| < c\varepsilon/M \Rightarrow -\varepsilon \leq f(x) \leq \varepsilon$ , logo  $f$  é contínua no ponto 0.  $\square$

## Capítulo 4

# Funções Implícitas

### 1 Uma função implícita

Os pontos de  $\mathbb{R}^{n+1}$  serão escritos sob a forma  $(x, y)$ , onde  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  e  $y \in \mathbb{R}$ . O teorema abaixo dá significado preciso à afirmação de que “a equação  $f(x, y) = c$  define implicitamente  $y$  como função de  $x$ ” e estabelece uma condição suficiente para que ela seja verdadeira.

**Teorema 1 (Teorema da Função Implícita).** *Dada a função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^k$  ( $k \geq 1$ ) no aberto  $U \subset \mathbb{R}^{n+1}$ , seja  $(x_0, y_0) \in U$  tal que  $f(x_0, y_0) = c$  e  $\frac{\partial f}{\partial y}(x_0, y_0) \neq 0$ . Existem uma bola  $B = B(x_0; \delta)$  e um intervalo  $J = (y_0 - \varepsilon, y_0 + \varepsilon)$  com as seguintes propriedades:*

- 1)  $B \times \bar{J} \subset U$  e  $\frac{\partial f}{\partial y}(x, y) \neq 0$  para todo  $(x, y) \in B \times \bar{J}$ ;
- 2) Para todo  $x \in B$  existe um único  $y = \xi(x) \in J$  tal que  $f(x, y) = f(x, \xi(x)) = c$ .

A função  $\xi : B \rightarrow J$ , assim definida, é de classe  $C^k$  e suas derivadas parciais em cada ponto  $x \in B$  são dadas por

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

*Demonstração.* Para fixar as idéias, admitiremos que  $\frac{\partial f}{\partial y}(x_0, y_0) > 0$ . Pela continuidade de  $\partial f / \partial y$ , existem  $\delta > 0$  e  $\varepsilon > 0$  tais que, pondo  $B = B(x_0, \delta) \subset \mathbb{R}^n$  e  $J = (y_0 - \varepsilon, y_0 + \varepsilon) \subset \mathbb{R}$ , temos  $B \times \bar{J} \subset U$  e  $\frac{\partial f}{\partial y}(x, y) > 0$  para todo

$(x, y) \in B \times \bar{J}$ . Então, para todo  $x \in B$ , a função  $y \mapsto f(x, y)$  é crescente no intervalo  $[y_0 - \varepsilon, y_0 + \varepsilon] = \bar{J}$ . Como  $f(x_0, y_0) = c$ , segue-se que  $f(x_0, y_0 - \varepsilon) < c$  e  $f(x_0, y_0 + \varepsilon) > c$ . Sendo  $f$  contínua, podemos supor  $\delta$  tão pequeno que  $f(x, y_0 - \varepsilon) < c$  e  $f(x, y_0 + \varepsilon) > c$  para todo  $x \in B$ . Pelo Teorema do Valor Intermediário, para cada  $x \in B$ , existe um único  $y = \xi(x) \in \bar{J}$  tal que  $f(x, y) = c$ . Tem-se necessariamente  $y \in J$ . Mostremos que a função  $\xi : B \rightarrow J$  possui derivadas parciais em todo ponto  $x \in B$ .

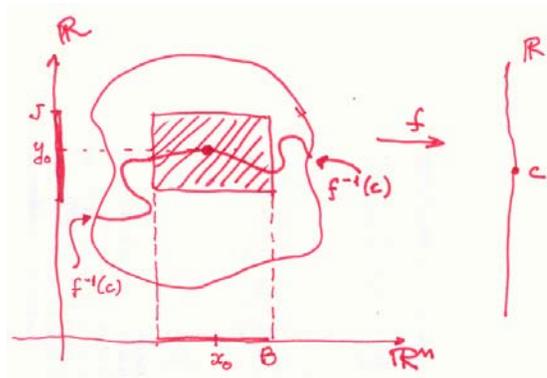


Figura 1.

Com efeito, pondo  $k = k(t) = \xi(x + te_i) - \xi(x)$ , vem  $\xi(x + te_i) = \xi(x) + k$ , logo  $f(x + te_i, \xi(x) + k) = f(x, \xi(x)) = c$ .

Pelo Teorema do Valor Médio, para todo  $t$  existe  $\theta = \theta(t) \in (0, 1)$  tal que

$$\begin{aligned} 0 &= f(x + te_i, \xi(x) + k) - f(x, \xi(x)) \\ &= \frac{\partial f}{\partial x_i}(x + \theta te_i, \xi(x) + \theta k) \cdot t + \frac{\partial f}{\partial y}(x + \theta te_i, \xi(x) + \theta k) \cdot k. \end{aligned}$$

Logo

$$\frac{\xi(x + te_i) - \xi(x)}{t} = \frac{k}{t} = -\frac{\frac{\partial f}{\partial x_i}(x + \theta te_i, \xi(x) + \theta k)}{\frac{\partial f}{\partial y}(x + \theta te_i, \xi(x) + \theta k)}.$$

Neste ponto, admitamos a continuidade de  $\xi$ , que será provada abaixo. Então  $\lim_{t \rightarrow 0} k(t) = 0$ . A continuidade das derivadas parciais de  $f$  nos dá então

$$\frac{\partial \xi}{\partial x_i}(x) = \lim_{t \rightarrow 0} \frac{\xi(x + te_i) - \xi(x)}{t} = -\frac{\frac{\partial f}{\partial x_i}(x, \xi(x))}{\frac{\partial f}{\partial y}(x, \xi(x))}, \quad (1 \leq i \leq n).$$

A expressão de  $\partial\xi/\partial x_i$  mostra que se  $f \in C^k$  então  $\partial\xi/\partial x_i \in C^{k-1}$  para  $i = 1, \dots, n$ , portanto  $\xi \in C^k$ .  $\square$

### Demonstração da continuidade de $\xi$

Pelo Teorema 19 do Capítulo 1 (v. observação que o segue), basta mostrar que, para todo conjunto fechado  $F \subset \bar{J}$ , a imagem inversa  $\xi^{-1}(F)$  é fechada em  $B$ . Ou seja: se a seqüência de pontos  $x_k \in B$  é tal que  $\xi(x_k) \in F$  para todo  $k \in \mathbb{N}$  e  $\lim x_k = \bar{x} \in B$  então,  $\xi(\bar{x}) \in F$ . Ora,  $F$  é compacto, logo uma subsequência de pontos  $x'_k \in B$  é tal que  $\lim \xi(x'_k) = a \in F$ . Logo  $f(\bar{x}, a) = \lim f(x'_k, \xi(x'_k)) = c$ . Mas  $f(\bar{x}, \xi(\bar{x})) = c$ . Pela unicidade de  $\xi(x)$ , segue-se que  $\xi(\bar{x}) = a \in F$ .  $\square$

Considerando o aberto  $V = B \times J \subset \mathbb{R}^{n+1}$ , o teorema acima diz que, nas condições das hipóteses, tem-se

$$f^{-1}(c) \cap V = \{(x, \xi(x)) \in \mathbb{R}^{n+1}; x \in B\}.$$

Noutras palavras,  $f^{-1}(c) \cap V$  é o gráfico da função  $\xi : B \rightarrow \mathbb{R}$ .

**Observação.** Evidentemente, não há nada de especial quanto à última coordenada, exceto simplificar a escrita na demonstração. Se, para algum inteiro  $i \in [1, n+1]$ , tivermos  $\frac{\partial f}{\partial x_i}(z_0) \neq 0$  onde  $z_0 \in U$  e  $f(z_0) = c$ , existirá um aberto  $V \ni z_0$ , tal que, para  $z \in V$ , a equação  $f(z) = c$  definirá  $x_i = \xi(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})$  como função das outras  $n$  coordenadas e  $f^{-1}(c) \cap V$  será o gráfico dessa função  $\xi$ , de classe  $C^k$ . De um modo geral, se  $\text{grad } f(z_0) \neq 0$  e  $f(z_0) = c$  então existe  $V \ni z_0$  aberto tal que  $f^{-1}(c) \cap V$  é o gráfico de uma função real de  $n$  variáveis, de classe  $C^k$ .

**Exemplo 1.** Seja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x, y) = x^2 + y^2$ . Para todo  $(x, y) \in \mathbb{R}^2$ , temos  $\frac{\partial f}{\partial x}(x, y) = 2x$  e  $\frac{\partial f}{\partial y}(x, y) = 2y$ . A equação  $x^2 + y^2 = c$  define o conjunto vazio quando  $c < 0$ . (O Teorema 1 não se aplica, pois não existe o ponto  $(x_0, y_0)$  tal que  $f(x_0, y_0) = c$ .) Quando  $c = 0$ , a equação  $x^2 + y^2 = 0$  é satisfeita apenas quando  $x = y = 0$ . (Agora existe  $(x_0, y_0)$  mas  $\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$ .)

Quando  $c > 0$ , a equação  $x^2 + y^2 = c$  define a circunferência de centro na origem e raio  $\sqrt{c}$ , a qual não é gráfico de função alguma do tipo  $y = \xi(x)$  nem  $x = \zeta(y)$ , pois há retas verticais e horizontais que a cortam em dois pontos. Mas, se considerarmos os abertos

$$\begin{aligned} V^1 &= \{(x, y) \in \mathbb{R}^2; y > 0\}, V^2 = \{(x, y) \in \mathbb{R}^2; y < 0\}, \\ V^3 &= \{(x, y) \in \mathbb{R}^2; x > 0\}, V^4 = \{(x, y) \in \mathbb{R}^2; x < 0\}, \end{aligned}$$

veremos que  $f^{-1}(c) \cap V_1$  e  $f^{-1}(c) \cap V_2$  são gráficos das funções  $\xi_1, \xi_2 : (-1, 1) \rightarrow \mathbb{R}$ , dadas por  $\xi_1(x) = \sqrt{1-x^2}$ ,  $\xi_2(x) = -\sqrt{1-x^2}$ , enquanto  $f^{-1}(c) \cap V_3$  e  $f^{-1}(c) \cap V_4$  são os gráficos de  $\xi_3, \xi_4 : (-1, 1) \rightarrow \mathbb{R}$ , dadas por  $\xi_3(y) = \sqrt{1-y^2}$  e  $\xi_4(y) = -\sqrt{1-y^2}$ . Assim, em  $V_1$  e  $V_2$  a equação  $x^2 + y^2 = c$  (com  $c > 0$ ) define implicitamente  $y$  como função de  $x$  enquanto em  $V_3$  e  $V_4$  define  $x$  como função de  $y$ . Evidentemente, salvo na vizinhança dos 4 pontos  $(\pm\sqrt{c}, 0)$ ,  $(0, \pm\sqrt{c})$ , tem-se a opção de tomar  $y$  como função de  $x$  ou  $x$  como função de  $y$ .  $\triangleleft$

## 2 Hiperfícies

Um conjunto  $M \subset \mathbb{R}^{n+1}$  chama-se uma *hiperfície* de classe  $C^k$  quando é localmente o gráfico de uma função real de  $n$  variáveis de classe  $C^k$ . Mais precisamente, para cada  $p \in M$  deve existir um aberto  $V \subset \mathbb{R}^{n+1}$  e uma função  $\xi : U \rightarrow \mathbb{R}$ , de classe  $C^k$  num aberto  $U \subset \mathbb{R}^n$ , tais que  $p \in V$  e  $V \cap M = \text{gráfico de } \xi$ .

A afirmação “ $V \cap M = \text{gráfico de } \xi$ ” significa que, para um certo inteiro  $i \in [1, n]$ , tem-se

$$V \cap M = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1}; x_i = \xi(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})\}.$$

Evidentemente, dada qualquer função  $f : U \rightarrow \mathbb{R}$  de classe  $C^k$  no aberto  $U \subset \mathbb{R}^n$ , seu gráfico é uma hiperfície  $M = \{(x, f(x)) \in \mathbb{R}^{n+1}; x \in U\}$  de classe  $C^k$ .

Quando  $n = 1$ , uma hiperfície em  $\mathbb{R}^2$  chama-se uma *curva* e, quando  $n = 2$ , tem-se uma *superfície* em  $\mathbb{R}^3$ .

**Exemplo 2.** A esfera  $S^n = \{x \in \mathbb{R}^{n+1}; \langle x, x \rangle = 1\}$  é uma hiperfície  $C^\infty$  em  $\mathbb{R}^{n+1}$ . Com efeito, chamando de  $U$  a bola aberta de centro 0 e raio 1 em  $\mathbb{R}^n$ , pondo, para cada  $i = 1, \dots, n+1$ ,  $V_i = \{x \in \mathbb{R}^{n+1}; x_i > 0\}$ ,  $W_i = \{x \in \mathbb{R}^{n+1}; x_i < 0\}$  e escrevendo  $x^* = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})$ , temos:

$$\begin{aligned} x \in S^n \cap V_i &\iff |x^*| < 1 \quad \text{e} \quad x_i = \sqrt{1 - \langle x^*, x^* \rangle}; \\ x \in S^n \cap W_i &\iff |x^*| < 1 \quad \text{e} \quad x_i = -\sqrt{1 - \langle x^*, x^* \rangle}. \end{aligned}$$

Logo, considerando a função  $\xi : U \rightarrow \mathbb{R}$ , de classe  $C^\infty$ , definida por  $\xi(u) = \sqrt{1 - \langle u, u \rangle}$ , vemos que, para cada  $i = 1, \dots, n+1$ ,  $S^n \cap V_i$  é o gráfico da função  $x_i = \xi(x^*)$  enquanto que  $S^n \cap W_i$  é o gráfico de  $x_i = -\xi(x^*)$ . Como todo ponto  $p \in S^n$  pertence a algum  $V_i$  ou a algum  $W_i$ , concluímos que  $S^n$  é uma hiperfície de classe  $C^\infty$  em  $\mathbb{R}^{n+1}$ .

Seja  $M \subset \mathbb{R}^{n+1}$  uma hiperfície de classe  $C^k$  ( $k \geq 1$ ). A cada ponto  $p \in M$  associaremos o conjunto  $T_p M$ , formado por todos os vetores-velocidade  $v = \lambda'(0)$  dos caminhos  $\lambda : (-\delta, \delta) \rightarrow M$  que são diferenciáveis no ponto 0 e cumprem a

condição  $\lambda(0) = p$ . O conjunto  $T_p M$  é chamado o *espaço vetorial tangente* de  $M$  no ponto  $p$ . Esta denominação se justifica pelo  $\triangleleft$

**Teorema 2.**  $T_p M$  é um subespaço vetorial de dimensão  $n$  em  $\mathbb{R}^{n+1}$ .

*Demonstração.* Seja  $\xi : U \rightarrow \mathbb{R}$  uma função de classe  $C^k$  no aberto  $U \subset \mathbb{R}^n$ , cujo gráfico, formado pelos pontos  $(x, \xi(x)) \in \mathbb{R}^{n+1}$ ,  $x \in U$ , é a interseção  $M \cap V$ , onde  $V \subset \mathbb{R}^{n+1}$  é um aberto que contém  $p = (p_0, \xi(p_0))$ ,  $p_0 \in U$ . Para todo caminho  $\lambda : (-\delta, \delta) \rightarrow M$ , com  $\lambda(0) = p$ , tem-se  $\lambda(t) = (x_1(t), \dots, x_n(t), \xi(x(t)))$ , onde  $x(t) = (x_1(t), \dots, x_n(t))$ . Portanto

$$\lambda'(0) = \left( \frac{dx_1}{dt}, \dots, \frac{dx_n}{dt}, \sum_{i=1}^n \frac{\partial \xi}{\partial x_i} \cdot \frac{dx_i}{dt} \right),$$

as derivadas  $dx_i/dt$  sendo calculadas no ponto  $t = 0$  e  $\partial \xi / \partial x_i$  no ponto  $p_0$ . Isto mostra que todo  $v = \lambda'(0)$  em  $T_p M$  é uma combinação linear dos vetores  $v_1 = (1, 0, \dots, 0, \partial \xi / \partial x_1), \dots, v_n = (0, \dots, 0, 1, \partial \xi / \partial x_n)$ . (Derivadas no ponto  $p_0$ .)

Reciprocamente, toda combinação linear  $v = \sum_{i=1}^n \alpha_i v_i$  é o vetor-velocidade  $\lambda'(0)$  do caminho  $\lambda : (-\delta, \delta) \rightarrow M$  assim definido: tomamos  $v_0 = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$  e pomos  $\lambda(t) = (p_0 + tv_0, \xi(p_0 + tv_0))$ , sendo  $\delta > 0$  escolhido de modo que o segmento de reta  $(p_0, -\delta v_0, p_0 + \delta v_0)$  esteja contido em  $U$ .  $\square$

**Observação.** Como subespaço vetorial de  $\mathbb{R}^{n+1}$ , o espaço vetorial tangente  $T_p M$  contém a origem  $0 \in \mathbb{R}^{n+1}$  e não contém necessariamente o ponto  $p$ , embora nas figuras ele apareça passando por  $p$ . Ocorre que, nas ilustrações, o que se vê é a variedade afim  $p + T_p M$ , paralela a  $T_p M$  por  $p$ .

**Exemplo 3.** O espaço vetorial tangente  $T_p S^n$  é, para todo  $p \in S^n$ , o complemento ortogonal de  $p$ , isto é, o conjunto  $[p]^\perp$  de todos os vetores  $v \in \mathbb{R}^{n+1}$  tais que  $\langle v, p \rangle = 0$ . Com efeito, sendo  $T_p S^n$  e  $[p]^\perp$  ambos subespaços vetoriais de dimensão  $n$  em  $\mathbb{R}^{n+1}$ , para mostrar que eles coincidem, basta provar que  $T_p S^n \subset [p]^\perp$ . Ora, se  $v \in T_p S^n$  então  $v = \lambda'(0)$ , onde  $\lambda : (-\delta, \delta) \rightarrow S^n$  é um caminho diferenciável no ponto 0, com  $\lambda(0) = p$ . Neste caso,  $0 = \frac{d}{dt} \langle \lambda(t), \lambda(t) \rangle = 2 \langle \lambda'(0), \lambda(0) \rangle = 2 \langle v, p \rangle$ .  $\triangleleft$

A seguir, apresentaremos um critério bastante útil para dar exemplos de hiperfícies.

Um número  $c \in \mathbb{R}$  chama-se um *valor regular* de uma função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^1$ , quando não há pontos críticos de  $f$  no nível  $c$ , isto é, quando  $f(x) = c \Rightarrow \text{grad } f(x) \neq 0$ . Diz-se também que  $c$  é um *nível regular* de  $f$ . Quando existe  $x \in U$  tal que  $f(x) = c$  e  $\text{grad } f(x) = 0$ , diz-se que  $c$  é um *nível crítico* de  $f$ .

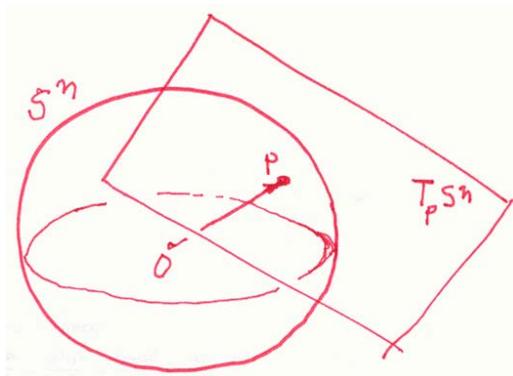


Figura 2.

**Teorema 3.** Se  $c$  é um valor regular da função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{n+1}$ , então  $M = f^{-1}(c)$  é uma hipersuperfície de classe  $C^k$ , cujo espaço vetorial tangente  $T_p M$  é, em cada ponto  $p \in M$ , o complemento ortogonal de  $\text{grad } f(p)$ .

*Demonstração.* O fato de que  $f^{-1}(c)$  é uma hipersuperfície é apenas uma reformulação verbal do Teorema da Função Implícita. (Ver Observação após a prova do Teorema 1.) Quanto ao espaço vetorial tangente  $T_p M$ , como  $M$  é uma superfície de nível da função  $f$ , vemos que todo vetor  $v \in T_p M$  é ortogonal a  $\text{grad } f(p)$ , logo  $T_p M \subset [\text{grad } f(p)]^\perp$ . Sendo ambos subespaços de dimensão  $n$  em  $\mathbb{R}^{n+1}$ , conclui-se que  $T_p M = [\text{grad } f(p)]^\perp$ .  $\square$

**Exemplo 4 (Mais uma vez a esfera).** À luz do Teorema 3, a esfera unitária  $S^n$  é a superfície de nível 1 da função  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ , dada por  $f(x) = \langle x, x \rangle$ . Como  $\text{grad } f(x) = 2x$ , vemos que zero é o único nível crítico de  $f$ . Em particular, 1 é valor regular e  $S^n = f^{-1}(1)$  é uma hipersuperfície  $C^\infty$  e, para todo  $p \in S^n$ , tem-se  $T_p S^n = [\text{grad } f(p)]^\perp = [p]^\perp$ .  $\triangleleft$

**Exemplo 5.** Seja  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  um operador linear auto-adjunto. A função  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , definida por  $f(x) = \langle A \cdot x, x \rangle$  é o que se chama uma *forma quadrática*. Se  $[a_{ij}]$  é a matriz (simétrica) de  $A$  na base canônica de  $\mathbb{R}^n$  então  $f(x) = \sum_{i,j=1}^n a_{ij} x_i x_j$ . Logo  $\partial f / \partial x_i = 2 \sum_{j=1}^n a_{ij} x_j$  e conseqüentemente  $\text{grad } f(x) = 2A \cdot x$ . Supondo agora que o operador  $A$  seja invertível, o único ponto crítico da função  $f$  é a origem 0, onde  $f$  assume o valor zero. Então, para todo  $c \neq 0$  a equação  $f(x) = c$  define uma hipersuperfície. Costuma-se tomar  $c = 1$  e a hipersuperfície definida pela equação  $f(x) = 1$ , ou seja,  $\sum_{i,j=1}^n a_{ij} x_i x_j = 1$ , chama-se

uma *quádrica*. Em particular, se o operador  $A$  é positivo, isto é, se  $f(x) > 0$  para todo  $x \neq 0$ , a quádrica  $f^{-1}(1)$  chama-se um *elipsóide*.  $\triangleleft$

**Exemplo 6.** Seja  $f : \mathbb{R}^{n^2} \rightarrow \mathbb{R}$  a função que associa a cada matriz  $\mathbf{x} = [x_{ij}]$  de  $n$  linhas e  $n$  colunas seu determinante  $f(\mathbf{x}) = \det \mathbf{x}$ . O desenvolvimento de Laplace nos dá

$$f(\mathbf{x}) = \sum_{j=1}^n (-1)^{i+j} x_{ij} \cdot X_{ij},$$

onde o  $ij$ -ésimo menor  $X_{ij}$  é o determinante da matriz  $(n-1) \times (n-1)$  que se obtém de  $\mathbf{x}$  omitindo a  $i$ -ésima linha e a  $j$ -ésima coluna. Segue-se daí que  $\frac{\partial f}{\partial x_{ij}}(\mathbf{x}) = (-1)^{i+j} X_{ij}$ . Em particular, se  $\mathbf{x} = \mathbf{I} =$  matriz identidade  $n \times n$ , temos  $\frac{\partial f}{\partial x_{ij}}(\mathbf{I}) = \delta_{ij}$  (*delta de Kronecker*, igual a 1 quando  $i = j$  e 0 quando  $i \neq j$ ).

Portanto  $\text{grad } f(\mathbf{I}) = \mathbf{I}$ . Seja  $U \subset \mathbb{R}^{n^2}$  o conjunto aberto formado pelas matrizes (invertíveis)  $\mathbf{x}$  tais que  $\det \mathbf{x} \neq 0$ . Para toda  $\mathbf{x} \in U$ , o desenvolvimento de Laplace nos mostra que algum menor  $X_{ij} \neq 0$ , logo  $\text{grad } f(\mathbf{x}) \neq 0$ . Portanto a função  $f : U \rightarrow \mathbb{R}$  não possui pontos críticos: todo número real  $c$  é um valor regular de  $f$ . Logo  $M = f^{-1}(1) =$  conjunto das matrizes reais  $n \times n$  com determinante 1 é uma hipersuperfície  $C^\infty$ .  $M$  é um grupo em relação à multiplicação de matrizes, conhecido como o *grupo unimodular*. O espaço vetorial  $T_{\mathbf{I}}(M)$ , tangente a  $M$  na matriz identidade  $\mathbf{I}$ , é formado pelas matrizes  $\mathbf{x}$  que são perpendiculares (em termos do produto interno de  $\mathbb{R}^{n^2}$ ) ao gradiente  $\text{grad } f(\mathbf{I}) = \mathbf{I}$ . Ora,

$$\langle \mathbf{x}, \mathbf{I} \rangle = \sum_{i,j=1}^n x_{ij} \delta_{ij} = \sum_{i=1}^n x_{ii} = \text{traço de } \mathbf{x}.$$

Assim, o espaço vetorial tangente a  $M$  no ponto  $\mathbf{I}$  é o conjunto das matrizes de traço nulo.  $\triangleleft$

**Observação.** O Teorema 3 é uma boa fonte de exemplos de hipersuperfícies. Mas nem toda hipersuperfície  $M \subset \mathbb{R}^{n+1}$  pode ser obtida como imagem inversa  $M = f^{-1}(c)$  do valor regular  $c$  de uma função  $f : U \rightarrow \mathbb{R}$ . Com efeito, as hipersuperfícies desse tipo admitem um campo contínuo de vetores não-nulos  $v = \text{grad } f : M \rightarrow \mathbb{R}^{n+1}$ , tais que, para todo  $x \in M$ ,  $\langle v(x), w \rangle = 0$  qualquer que seja  $w \in T_x M$ . (Diz-se então que  $v = \text{grad } f$  é um campo de vetores *normais* a  $M$ .) Tais hipersuperfícies são chamadas de *orientáveis*. Um exemplo bem conhecido de superfície não-orientável é a faixa de Moebius. Logo, a faixa de Moebius não é imagem inversa de um valor regular de uma função de classe  $C^1$  definida num aberto de  $\mathbb{R}^3$ .

### 3 Multiplicador de Lagrange

O método do multiplicador de Lagrange se aplica na seguinte situação: tem-se uma função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^1$  no aberto  $U \subset \mathbb{R}^{n+1}$  (função-objetivo), uma hiperfície  $M = \varphi^{-1}(c)$ , imagem inversa do valor regular  $c$  da função  $\varphi : U \rightarrow \mathbb{R}$ , de classe  $C^1$ , e procura-se determinar quais são os pontos críticos da restrição  $f|M$ , ou seja, os pontos críticos  $x$  de  $f$  sujeitos à condição  $\varphi(x) = c$ .

Não se trata de determinar os pontos críticos de  $f : U \rightarrow \mathbb{R}$  que estão localizados sobre a hiperfície  $M$  mas sim os pontos críticos da função  $f|M : M \rightarrow \mathbb{R}$ . É preciso definir o que se entende por isto.

Um ponto  $x \in M$  chama-se um *ponto crítico da restrição  $f|M$*  quando, para todo caminho diferenciável  $\lambda : (-\delta, \delta) \rightarrow M$  com  $\lambda(0) = x$  tem-se  $(f \circ \lambda)'(0) = 0$ . Pondo  $v = \lambda'(0)$ , esta condição significa  $\langle \text{grad } f(x), v \rangle = 0$ . Como  $v$  é um vetor arbitrário pertencente ao espaço vetorial tangente  $T_x M$ , vemos que  $x \in M$  é um ponto crítico de  $f|M$  se, e somente se,  $\text{grad } f(x)$  é ortogonal ao espaço vetorial tangente  $T_x M$ .

Ora,  $\text{grad } \varphi(x)$  é um vetor (não-nulo) ortogonal a  $T_x M$ . Como o complemento ortogonal de  $T_x M$  em  $\mathbb{R}^{n+1}$  tem dimensão 1, segue-se que  $\text{grad } f(x) \perp T_x M$  se, e somente se,  $\text{grad } f(x)$  é um múltiplo de  $\text{grad } \varphi(x)$ . Portanto, podemos enunciar:

O ponto  $x \in U$  é um ponto crítico da restrição  $f|M$  de  $f$  à hiperfície  $M = \varphi^{-1}(c)$  se, e somente se:

- 1)  $\varphi(x) = c$ ;
- 2)  $\text{grad } f(x) = \lambda \cdot \text{grad } \varphi(x)$  para algum  $\lambda \in \mathbb{R}$ .

As condições acima representam um sistema de  $n + 2$  equações (pois a igualdade vetorial 2) acima significa  $n + 1$  igualdades numéricas) nas  $n + 2$  incógnitas  $x_1, \dots, x_{n+1}$  (coordenadas de  $x$ ) e  $\lambda$ . O fator  $\lambda$  é chamado o *multiplicador de Lagrange*. Sua presença torna o número de incógnitas igual ao número de equações, o que viabiliza a solução na prática.

Deve-se notar que se  $x \in M$  é um ponto de mínimo ou de máximo local de  $f|M$  então, para todo caminho diferenciável  $\lambda : (-\delta, \delta) \rightarrow M$  com  $\lambda(0) = x$ , a função  $f \circ \lambda : (-\delta, \delta) \rightarrow \mathbb{R}$  tem um mínimo ou um máximo local no ponto 0, logo  $(f \circ \lambda)'(0) = 0$ . Portanto os mínimos e máximos locais de  $f|M$  estão incluídos na definição de ponto crítico dada acima.

É também evidente que todo ponto crítico  $x$  da função  $f : U \rightarrow \mathbb{R}$  é, com maior razão, ponto crítico da restrição  $f|M$  pois, sendo  $\text{grad } f(x) = 0$ , tem-se  $\langle \text{grad } f(x), v \rangle = 0$  para todo  $v \in \mathbb{R}^{n+1}$ .

Muitas vezes, a condição adicional  $\varphi(x) = c$  é posta sob a forma  $\varphi(x) = 0$ . Isto não representa perda de generalidade. Basta usar, em vez de  $\varphi$ , a função

$\psi(x) = \varphi(x) - c$ . Então  $\psi(x) = 0 \Leftrightarrow \varphi(x) = c$  e  $c$  é valor regular de  $\varphi$  se, e somente se,  $0$  é valor regular de  $\psi$ .

**Exemplo 7.** Seja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  definida por  $f(x, y) = ax + by$ , com  $a^2 + b^2 \neq 0$ . O gradiente de  $f$  é, em todo ponto  $(x, y)$ , o vetor constante não-nulo  $v = (a, b)$ , ortogonal às linhas de nível  $ax + by = c$ , que são retas, duas a duas paralelas. A função  $f$  não tem pontos críticos. Mas

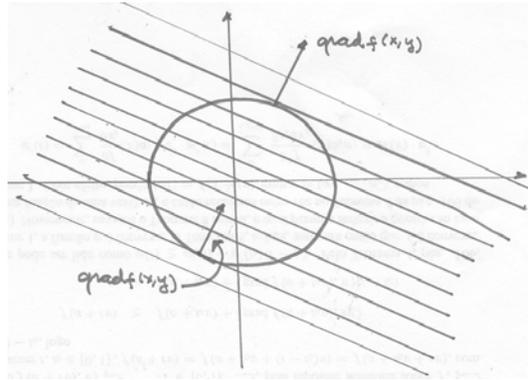


Figura 3.

se  $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$  for dada por  $\varphi(x, y) = x^2 + y^2$  então  $\text{grad } \varphi(x, y) = (2x, 2y)$ ,  $1$  é valor regular de  $\varphi$  e  $M = \varphi^{-1}(1)$  é a circunferência unitária  $x^2 + y^2 = 1$ . Como  $M$  é compacta, a restrição  $f|M$  possui pelo menos dois pontos críticos, nos quais assume seus valores mínimo e máximo. Os pontos críticos de  $f|M$  são as soluções  $(x, y)$  do sistema

$$\text{grad } f(x, y) = \lambda \cdot \text{grad } \varphi(x, y), \quad \varphi(x, y) = 1,$$

ou seja:

$$2\lambda x = a, \quad 2\lambda y = b, \quad x^2 + y^2 = 1.$$

Portanto  $(x, y)$  é um ponto crítico de  $f|M$  se, e somente se, o vetor *unitário*  $z = (x, y)$  é um múltiplo do vetor  $v = (a, b)$ . Isto nos dá

$$(x, y) = \left( \frac{a}{\sqrt{a^2 + b^2}}, \frac{b}{\sqrt{a^2 + b^2}} \right) \quad \text{ou} \quad (x, y) = \left( \frac{-a}{\sqrt{a^2 + b^2}}, \frac{-b}{\sqrt{a^2 + b^2}} \right).$$

Estes são os pontos nos quais  $f(x, y)$  assume seus valores máximo e mínimo em  $M = S^1$ . ◁

**Exemplo 8.** Seja  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  uma forma quadrática. Para todo  $x = (x_1, \dots, x_n)$ , tem-se  $f(x) = \sum_{i,j=1}^n a_{ij}x_i x_j$ , onde  $\mathbf{a} = [a_{ij}]$  é uma matriz simétrica  $n \times n$ .

Alternativamente, tem-se  $f(x) = \langle Ax, x \rangle$ , onde  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  é o operador linear auto-adjunto cuja matriz na base canônica de  $\mathbb{R}^n$  é  $\mathbf{a}$ . Quais são os pontos críticos da restrição  $f|_{S^{n-1}}$ , onde  $S^{n-1}$  é a esfera unitária de  $\mathbb{R}^n$ ? Temos  $S^{n-1} = \varphi^{-1}(1)$ , onde  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$  é definida por  $\varphi(x) = \langle x, x \rangle$  e, como  $\text{grad } \varphi(x) = 2x$ , 1 é valor regular de  $\varphi$ . Por sua vez,  $\frac{\partial f}{\partial x_i}(x) = 2 \cdot \sum_{j=1}^n a_{ij}x_j$ , portanto  $\text{grad } f(x) = 2A \cdot x$ .

Portanto os pontos críticos da restrição  $f|_{S^{n-1}}$  são as soluções  $x$  do sistema  $Ax = 2\lambda x$ ,  $\langle x, x \rangle = 1$ , isto é, são os autovetores do operador  $A$  que têm comprimento 1. Como  $S^{n-1}$  é compacta,  $f$  admite pelo menos 2 pontos críticos em  $S^{n-1}$ , a saber, os pontos em que assume seus valores mínimo e máximo. Isto fornece uma prova de que todo operador auto-adjunto em  $\mathbb{R}^n$  possui autovetores, o que é o passo fundamental para a demonstração do Teorema Espectral.  $\triangleleft$

**Exemplo 9.** Seja  $U \subset \mathbb{R}^n$  o conjunto dos pontos cujas coordenadas são positivas. Consideremos as funções  $f, \varphi : U \rightarrow \mathbb{R}$  definidas, para todo  $x = (x_1, \dots, x_n) \in U$ , como  $f(x) = x_1 \cdot x_2 \dots x_n$  e  $\varphi(x) = x_1 + x_2 + \dots + x_n$ . Fixando  $s > 0$ , procuremos os pontos críticos de  $f|M$  onde  $M = \varphi^{-1}(s)$ . Observemos que  $\text{grad } \varphi(x) = (1, 1, \dots, 1)$  para qualquer  $x \in U$ , de modo que  $M$  é uma hipersuperfície. Por sua vez, temos  $\text{grad } f(x) = (\alpha_1, \dots, \alpha_n)$  com  $\alpha_i = \prod_{j \neq i} x_j$ . Assim,  $x \in M$  é ponto crítico de  $f|M$  se, e somente se, para algum  $\lambda$ , tem-se  $\prod_{j \neq i} x_j = \lambda(i = 1, \dots, n)$ . Dividindo a  $i$ -ésima dessas equações pela  $k$ -ésima, obtemos  $x_k/x_i = 1$ . Assim, o único ponto crítico de  $f|M$  é aquele que tem suas coordenadas iguais, ou seja, é  $p = (s/n, s/n, \dots, s/n)$ . Afirmamos que  $f(p) = (s/n)^n$  é o maior valor de  $f|M$ . Com efeito, a fórmula de  $f$  define uma função contínua no compacto  $\bar{M}$ , onde possui um ponto de máximo, o qual não pode estar em  $\bar{M} - M$  pois  $x_1 \cdot x_2 \dots x_n = 0$  se  $x \in \bar{M} - M$ . Logo esse máximo está em  $M$ , portanto é um ponto crítico, mas  $p$  é o único ponto crítico de  $f|M$ . Conclusão: quando  $n$  números positivos têm soma constante  $s$ , seu produto é máximo, igual a  $(s/n)^n$  quando eles são iguais. Ou ainda, se  $x_1, \dots, x_n$  são positivos então

$$x_1 \cdot x_2 \dots x_n \leq \left( \frac{x_1 + x_2 + \dots + x_n}{n} \right)^n.$$

A desigualdade acima, posta sob a forma

$$\sqrt[n]{x_1 \cdot x_2 \dots x_n} \leq \frac{x_1 + x_2 + \dots + x_n}{n},$$

diz que a média geométrica de números positivos é menor do que ou igual à média aritmética. Além disso, elas coincidem somente quando os números dados são iguais.  $\triangleleft$

**Exemplo 10.** Dadas a função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{n+1}$  e a superfície  $M \subset U$ , os pontos críticos da restrição  $f|_M$  são os pontos  $x \in M$  para os quais  $\text{grad } f(x)$  é ortogonal ao espaço vetorial tangente  $T_x M$ , mesmo quando  $M$  não é obtida como imagem inversa  $\varphi^{-1}(c)$  de um valor regular de uma função  $\varphi : U \rightarrow \mathbb{R}$  de classe  $C^k$ . Isto ficou claro na discussão feita no início desta seção. Como exemplo, consideremos uma superfície  $M \subset \mathbb{R}^{n+1}$ , um ponto  $a \in \mathbb{R}^{n+1}$  não pertencente a  $M$  e indaguemos quais são os pontos  $p \in M$  situados à distância mínima de  $a$ . Trata-se de obter os pontos que tornam mínima a restrição  $f|_M$ , onde

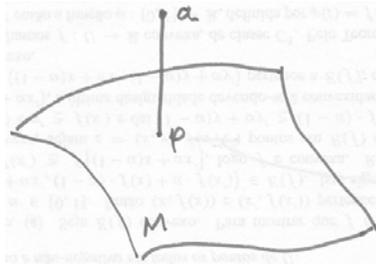


Figura 4.

$f : U \rightarrow \mathbb{R}$ , dada por  $f(x) = |x - a|$ , tem  $U = \mathbb{R}^{n+1} - \{a\}$  por domínio, por isso é de classe  $C^\infty$ . Temos  $f(x) = \sqrt{\sum (x_i - a_i)^2}$ , logo  $\partial f / \partial x_i = (x_i - a_i) / |x - a|$  e daí  $\text{grad } f(x) = x - a / |x - a|$ . Assim, os pontos críticos de  $f$ , entre os quais estão os pontos de  $M$  situados à distância mínima de  $a$ , são os pontos  $x \in M$  tais que  $x - a$  é um vetor normal a  $M$  no ponto  $x$ , isto é,  $\langle x - a, v \rangle = 0$  para todo  $v \in T_x M$ . Em particular, se  $M = S^n$ ,  $x - a \perp T_x S^n$  significa  $x - a = \alpha \cdot x$  isto é,  $x = a / (1 - \alpha)$ . Portanto, neste caso, os únicos pontos críticos de  $f|_{S^n}$  são os pontos  $x \in S^n$  pertencentes à reta  $0a$ , os quais são  $\pm a / |a|$ . Um deles minimiza  $|x - a|$  e o outro maximiza  $f$ .  $\triangleleft$

**Observação.** Os pontos críticos da restrição  $f|_M$  da função  $f : U \rightarrow \mathbb{R}$  à superfície  $\varphi^{-1}(0)$ , onde  $\varphi : U \rightarrow \mathbb{R}$  tem 0 como valor regular, são, como vimos, as soluções  $x$  do sistema de equações  $\text{grad } f(x) = \lambda \cdot \text{grad } \varphi(x)$ ,  $\varphi(x) = 0$ . Se considerarmos a função  $L : U \times \mathbb{R} \rightarrow \mathbb{R}$ , definida por  $L(x, \lambda) = f(x) - \lambda \varphi(x)$ , vemos que as equações acima são satisfeitas se, e somente se,  $\text{grad } L(x, \lambda) = 0$ , ou seja, os pontos críticos de  $f$  sujeitos ao vínculo  $\varphi(x) = 0$  são precisamente os

pontos críticos (livres) da função  $L$ , chamada a *Lagrangiana* do problema.

## Capítulo 5

# Aplicações Diferenciáveis

### 1 A derivada como transformação linear

Uma aplicação  $f : U \rightarrow \mathbb{R}^n$ , definida no aberto  $U \subset \mathbb{R}^m$ , diz-se *diferenciável* no ponto  $a \in U$  quando cada uma das suas funções-coordenada  $f_1, \dots, f_n : U \rightarrow \mathbb{R}$  é diferenciável nesse ponto.

Se este é o caso então, para todo  $v = (\alpha_1, \dots, \alpha_m)$  tal que  $a + v \in U$  e para cada  $i = 1, \dots, m$ , tem-se

$$f_i(a + v) - f_i(a) = \sum_{j=1}^m \frac{\partial f_i}{\partial x_j}(a) \cdot \alpha_j + r_i(v) \quad \text{com} \quad \lim_{v \rightarrow 0} \frac{r_i(v)}{|v|} = 0.$$

A matriz  $Jf(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(a) \\ \vdots \\ \frac{\partial f_n}{\partial x_j}(a) \end{bmatrix} \in M(n \times m)$  chama-se a *matriz jacobiana* de  $f$  no ponto  $a$ .

A transformação linear  $f'(a) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , cuja matriz em relação às bases canônicas de  $\mathbb{R}^m$  e  $\mathbb{R}^n$  é  $Jf(a)$ , chama-se a *derivada da aplicação  $f$*  no ponto  $a$ .

De acordo com a definição de matriz de uma transformação linear, para todo  $v = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$  temos

$$f'(a) \cdot v = (\beta_1, \dots, \beta_n) \quad \text{onde} \quad \beta_i = \sum_{j=1}^m \frac{\partial f_i}{\partial x_j}(a) \cdot \alpha_j = \frac{\partial f_i}{\partial v}(a).$$

Assim, se definirmos, como é natural, a *derivada direcional da aplicação  $f$* , no ponto  $a$ , na direção do vetor  $v$ , como

$$\frac{\partial f}{\partial v}(a) = \lim_{t \rightarrow 0} \frac{f(a + tv) - f(a)}{t},$$

teremos imediatamente

$$\frac{\partial f}{\partial v}(a) = \left( \frac{\partial f_1}{\partial v}(a), \dots, \frac{\partial f_n}{\partial v}(a) \right) = f'(a) \cdot v.$$

Resulta da Regra da Cadeia para funções (Teorema 2 do Capítulo 3), em conformidade com a observação feita logo após sua demonstração, que embora a definição de  $\frac{\partial f}{\partial v}(a)$  tenha sido dada acima como  $\frac{\partial f}{\partial v}(a) = (f \circ \lambda)'(0)$ , onde  $\lambda(t) = a + tv$ , vale, mais geralmente, a igualdade  $\frac{\partial f}{\partial v}(a) = (f \circ \lambda)'(0)$  para qualquer caminho diferenciável  $\lambda : (-\delta, \delta) \rightarrow U$ , com  $\lambda(0) = a$  e  $\lambda'(0) = v$ .

As  $n$  igualdades numéricas que exprimem a diferenciabilidade das funções-coordenada  $f_i$  se resumem na igualdade abaixo, entre vetores de  $\mathbb{R}^n$ :

$$f(a + v) - f(a) = f'(a) \cdot v + r(v), \quad \text{com} \quad \lim_{v \rightarrow 0} \frac{r(v)}{|v|} = 0.$$

Algumas vezes, é mais conveniente escrever esta condição sob a forma

$$f(a + v) - f(a) = f'(a) \cdot v + \rho(v) \cdot |v| \quad \text{com} \quad \lim_{v \rightarrow 0} \rho(v) = 0.$$

Aqui,  $\rho(v) = r(v)/|v|$  para todo  $v \neq 0$  tal que  $a + v \in U$ .

A relação acima caracteriza univocamente a diferenciabilidade da aplicação  $f$  no sentido seguinte: se uma transformação linear  $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é tal que, para  $a, a + v \in U$  tem-se

$$f(a + v) - f(a) = T \cdot v + r(v), \quad \text{com} \quad \lim_{v \rightarrow 0} \frac{r(v)}{|v|} = 0,$$

então  $T = f'(a)$ .

Com efeito, daí resulta, tomando  $tv$  em vez de  $v$ , que:

$$\frac{f(a + tv) - f(a)}{t} = T \cdot v \pm \frac{r(tv)}{|tv|} \cdot |v|,$$

logo

$$T \cdot v = \lim_{t \rightarrow 0} \frac{[f(a + tv) - f(a)]}{t} = \frac{\partial f}{\partial v}(a) = f'(a) \cdot v.$$

Quando  $f : U \rightarrow \mathbb{R}^n$  é diferenciável em todos os pontos de  $U$ , dizemos que  $f$  é *diferenciável em  $U$* . Neste caso, fica definida uma aplicação  $f' : U \rightarrow \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n)$  que faz corresponder a cada  $x \in U$  a transformação linear  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ . Quando for conveniente, identificaremos o conjunto  $\mathcal{L}(\mathbb{R}^m; \mathbb{R}^n)$  das transformações lineares de  $\mathbb{R}^m$  em  $\mathbb{R}^n$  com o conjunto  $M(n \times m)$  das matrizes  $n \times m$  ou com o espaço  $\mathbb{R}^{nm}$ .

Dizer que a aplicação derivada  $f' : U \rightarrow \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n)$ , (ou seja  $f' : U \rightarrow \mathbb{R}^{nm}$ ) é contínua equivale a afirmar a continuidade de cada uma de suas  $nm$  funções-coordenada  $\partial f_i / \partial x_j : U \rightarrow \mathbb{R}$ , isto é, a dizer que  $f$  é uma aplicação de classe

$C^1$  conforme a definição dada no Capítulo 3. Como foi demonstrado no Teorema 1 daquele capítulo, a continuidade das derivadas parciais  $\partial f_i / \partial x_j : U \rightarrow \mathbb{R}$  implica a diferenciabilidade de  $f$ .

Como no caso de funções, aplicações  $f : U \rightarrow \mathbb{R}^n$  de classe  $C^k$  são definidas por indução: diz-se que  $f \in C^k$  quando  $f$  é diferenciável e sua derivada  $f' : U \rightarrow \mathbb{R}^{nm}$  é de classe  $C^{k-1}$ . Se  $f \in C^k$  para todo  $k \in \mathbb{N}$  diz-se que  $f$  é de classe  $C^\infty : f \in C^\infty$ . Então  $f' \in C^\infty$  também.

**Observação.** Na maioria das vezes, a maneira mais simples de verificar que uma aplicação  $f$  é diferenciável consiste em calcular diretamente as derivadas parciais  $\frac{\partial f_i}{\partial x_j}(x)$ , mostrar que elas dependem continuamente de  $x$  e usar o Teorema 1 do Capítulo 3, segundo o qual toda função de classe  $C^1$  é diferenciável. Praticamente todas as aplicações diferenciáveis são de classe  $C^1$ . Ocorre, entretanto, que as propriedades mais relevantes das aplicações  $C^1$  resultam da relação que caracteriza sua diferenciabilidade. Daí a importância deste conceito.

## 2 Exemplos de derivadas

**Exemplo 1.** Sejam  $I \subset \mathbb{R}$  um intervalo aberto e  $f : I \rightarrow \mathbb{R}^n$  um caminho diferenciável no ponto  $a \in I$ . Considerando  $f$  como uma aplicação, sua derivada no ponto  $a$  é a transformação linear  $f'(a) : \mathbb{R} \rightarrow \mathbb{R}^n$  cuja matriz jacobiana tem por única coluna o vetor

$$v = \left( \frac{df_1}{dt}(a), \dots, \frac{df_n}{dt}(a) \right),$$

o qual vem a ser o vetor-velocidade do caminho  $f$  no ponto  $a$ , já indicado com a mesma notação  $f'(a)$  no Capítulo 2. Como transformação linear,  $f'(a) : \mathbb{R} \rightarrow \mathbb{R}^n$  faz corresponder a cada “vetor”  $t \in \mathbb{R}$  o vetor  $t \cdot v \in \mathbb{R}^n$ . Noutros termos:  $f'(a) \cdot t = t \cdot f'(a)$ .  $\triangleleft$

**Exemplo 2.** Seja  $f : U \rightarrow \mathbb{R}$  uma função definida no aberto  $U \subset \mathbb{R}^m$ , diferenciável no ponto  $a \in U$ . Sua derivada é uma transformação linear  $f'(a) : \mathbb{R}^m \rightarrow \mathbb{R}$ , portanto um funcional linear, que associa a cada vetor  $v = (\alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$  o número

$$f'(a) \cdot v = \frac{\partial f}{\partial x_1}(a) \cdot \alpha_1 + \dots + \frac{\partial f}{\partial x_m}(a) \cdot \alpha_m = \frac{\partial f}{\partial v}(a) = \langle \text{grad } f(a), v \rangle.$$

Às vezes se escreve  $df(a)$  e chama-se a *diferencial de  $f$*  à derivada  $f'(a)$ . Em particular, se usarmos a notação tradicional  $x_i : \mathbb{R}^m \rightarrow \mathbb{R}$  para indicar a função

que associa a cada ponto  $x \in \mathbb{R}^m$  sua  $i$ -ésima coordenada  $x_i$ , a diferencial  $dx_i$  desta função é o funcional linear que faz corresponder a cada vetor  $v = (\alpha_1, \dots, \alpha_m)$  sua  $i$ -ésima coordenada  $dx_i \cdot v = \alpha_i$  (mesmo porque, sendo linear, a função  $x_i$  tem derivada constante, igual a si própria). Então

$$df(a) \cdot v = \sum_{i=1}^m \frac{\partial f}{\partial x_i}(a) \cdot \alpha_i = \sum_{i=1}^m \frac{\partial f}{\partial x_i}(a) \cdot dx_i \cdot v.$$

Isto atribui um significado à expressão clássica

$$df = \sum_{i=1}^m \frac{\partial f}{\partial x_i} dx_i.$$

◁

**Exemplo 3.** Se  $f : U \rightarrow \mathbb{R}^n$  é constante então  $f'(x) = 0$  para todo  $x \in U$ . Reciprocamente, se o aberto  $U \subset \mathbb{R}^m$  é conexo e  $f : U \rightarrow \mathbb{R}^n$  possui derivada 0 em todos os pontos  $x \in U$  então  $f$  é constante. (Conforme o Corolário 3 do Teorema 2, Capítulo 3.)

◁

**Exemplo 4.** Se  $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é uma transformação linear então  $T$  é diferenciável e  $T'(x) = T$  para todo  $x \in \mathbb{R}^m$ . Noutras palavras,  $T'(x) \cdot v = T \cdot v$  quaisquer que sejam  $x, v \in \mathbb{R}^m$ . Isto resulta imediatamente da igualdade  $T(x+v) - T \cdot x = T \cdot v + r$ , onde  $r = 0$ , ou então do fato óbvio de que a matriz jacobiana de  $T$  é a própria matriz de  $T$ . Um caso muito particular: a soma  $S : \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ ,  $S \cdot (x, y) = x + y$  é linear, logo  $S'(x, y) \cdot (u, v) = u + v$  quaisquer que sejam  $x, y, u, v \in \mathbb{R}^m$ .

◁

**Exemplo 5.** Seja  $B : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^p$  uma aplicação bilinear, isto é, linear em cada uma de suas duas variáveis. Se escrevermos, para cada par de vetores  $(e_i, e_j)$  das bases canônicas de  $\mathbb{R}^m$  e  $\mathbb{R}^n$  respectivamente,  $B(e_i, e_j) = v_{ij}$ , então, para  $x = (x_1, \dots, x_m)$  e  $y = (y_1, \dots, y_n)$  teremos

$$B(x, y) = \sum_{i,j} x_i y_j v_{ij}.$$

Isto mostra que  $B$  é contínua, logo assume seu valor máximo  $|B|$  no compacto  $S^{m-1} \times S^{n-1}$ . Daí resulta que, para quaisquer  $x \in \mathbb{R}^m$  e  $y \in \mathbb{R}^n$  não-nulos, vale  $|B(x, y)| = |B(x/|x|, y/|y|)| \cdot |x| \cdot |y| \leq |B| \cdot |x| \cdot |y|$ . Para  $x = 0$  ou  $y = 0$ , a desigualdade  $|B(x, y)| \leq |B| \cdot |x| \cdot |y|$  é imediata pois  $B(0, y) = B(x, 0) = 0$ . Mostremos agora que toda aplicação bilinear  $B$  é diferenciável, com

$B'(x, y) \cdot (u, v) = B(u, y) + B(x, v)$ . Com efeito, se  $x, u \in \mathbb{R}^m$  e  $y, v \in \mathbb{R}^n$ , temos pela bilinearidade de  $B$ :

$$B(x + u, y + v) - B(x, y) = B(u, y) + B(x, v) + B(u, v).$$

Observando que  $|(u, v)| = \sqrt{|u|^2 + |v|^2} \geq |v|$ , temos

$$\frac{|B(u, v)|}{|(u, v)|} \leq \frac{|B| \cdot |u| \cdot |v|}{\sqrt{|u|^2 + |v|^2}} \leq |B| |u|,$$

logo  $\lim_{u, v \rightarrow 0} \frac{B(u, v)}{|(u, v)|} = 0$ , comprovando assim a diferenciabilidade de  $B$ .  $\triangleleft$

Nos exemplos 4 e 5 acima (e, obviamente, no Exemplo 3), as aplicações consideradas são de classe  $C^\infty$ . De fato, a derivada  $T' = T : \mathbb{R}^m \rightarrow \mathcal{L}(\mathbb{R}^m; \mathbb{R}^n)$  de uma transformação linear  $T$ , sendo constante, possui derivada nula e todas as derivadas seguintes também serão nulas. Quanto à aplicação bilinear  $B$ , sua derivada  $B' : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^n; \mathbb{R}^p)$  é a transformação linear  $(x, y) \mapsto B(\bullet, y) + B(x, \bullet)$ , recaindo assim no Exemplo 4.

**Exemplo 6 (Derivada complexa).** Uma função de variável complexa  $f : U \rightarrow \mathbb{C}$ , definida no aberto  $U \subset \mathbb{C}$ , pode ser vista como uma aplicação  $f : U \rightarrow \mathbb{R}^2$ , definida no aberto  $U \subset \mathbb{R}^2$ . A derivada da função complexa  $f$  no ponto  $z \in U$  é o número complexo  $f'(z)$ , definido como o limite

$$f'(z) = \lim_{H \rightarrow 0} \frac{f(z + H) - f(z)}{H},$$

quando tal limite existe. Isto equivale a dizer que

$$f(z + H) - f(z) = f'(z) \cdot H + r(H), \quad \text{onde} \quad \lim_{H \rightarrow 0} \frac{r(H)}{|H|} = 0.$$

Acima,  $f'(z) \cdot H$  é uma multiplicação de números complexos. Portanto, a função complexa  $f : U \rightarrow \mathbb{C}$  é derivável no ponto  $z \in U$  se, e somente se, a aplicação  $f : U \rightarrow \mathbb{R}^2$  é diferenciável nesse ponto e, além disso, sua derivada  $f'(z) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  é uma transformação linear do plano que consiste em multiplicar por um número complexo fixo. Ora, se  $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  é uma tal transformação, da forma  $T \cdot z = (a + bi) \cdot z$ , sua matriz na base canônica tem as colunas  $T \cdot 1 = a + bi$  e  $T \cdot i = -b + ai$ , ou seja, sua matriz é do tipo  $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$ . Se, para  $z = x + yi$ ,  $f(z) = u(x, y) + i \cdot v(x, y)$ , a matriz jacobiana de  $f$  é

$$\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}.$$

Segue-se então que a função complexa  $f$  é derivável em  $U$  se, e somente se, valem as relações  $\partial u/\partial x = \partial v/\partial y$  e  $\partial u/\partial y = -\partial v/\partial x$  em todo ponto  $z = x + yi \in U$ . Estas igualdades são conhecidas como as *equações de Cauchy-Riemann*.  $\triangleleft$

A derivada de  $f$ , considerada como função de uma variável complexa é

$$f'(z) = \frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = \frac{\partial v}{\partial y} - i \frac{\partial u}{\partial y}.$$

**Exemplo 7.** Se  $f, g : U \rightarrow \mathbb{R}^n$  são diferenciáveis no ponto  $a \in U \subset \mathbb{R}^m$  então a aplicação  $(f, g) : U \rightarrow \mathbb{R}^n \times \mathbb{R}^n$ , definida por  $(f, g)(x) = (f(x), g(x))$ , é diferenciável no ponto  $a$  e sua derivada é  $(f, g)'(a) \cdot v = (f'(a) \cdot v, g'(a) \cdot v)$ . Se  $f, g \in C^k$  então  $(f, g)$  também é de classe  $C^k$ .  $\triangleleft$

### 3 Cálculo diferencial de aplicações

**Teorema 1 (Regra da Cadeia).** *Sejam  $U \subset \mathbb{R}^m, V \subset \mathbb{R}^n$  abertos e  $f : U \rightarrow \mathbb{R}^n, g : V \rightarrow \mathbb{R}^p$  diferenciáveis nos pontos  $a \in U, b = f(a) \in V$ , com  $f(U) \subset V$ . Então  $g \circ f : U \rightarrow \mathbb{R}^p$  é diferenciável no ponto  $a$  e*

$$(g \circ f)'(a) = g'(b) \cdot f'(a) : \mathbb{R}^m \rightarrow \mathbb{R}^p.$$

Resumidamente: a derivada da aplicação composta é a composta das derivadas.

*Demonstração.* Podemos escrever

$$\begin{aligned} f(a+v) &= f(a) + f'(a) \cdot v + \rho(v) \cdot |v|, & \text{com } \lim_{v \rightarrow 0} \rho(v) &= 0 \quad \text{e} \\ g(b+w) &= g(b) + g'(b) \cdot w + \sigma(w) \cdot |w|, & \text{com } \lim_{w \rightarrow 0} \sigma(w) &= 0. \end{aligned}$$

Então

$$(g \circ f)(a+v) = g(f(a) + f'(a) \cdot v + \rho(v) \cdot |v|).$$

Pondo  $w = f'(a) \cdot v + \rho(v) \cdot |v|$ , obtemos:

$$\begin{aligned} (g \circ f)(a+v) &= g(b+w) = g(b) + g'(b) \cdot f'(a) \cdot v + g'(b) \cdot \rho(v)|v| \\ &\quad + \sigma(w) \cdot |w| = (g \circ f)(a) + [g'(b) \cdot f'(a)] \cdot v + C(v) \cdot |v|, \end{aligned}$$

onde

$$C(v) = g'(b) \cdot \rho(v) + \sigma(w) \cdot \left| f'(a) \cdot \frac{v}{|v|} + \rho(v) \right|.$$

Se  $v \rightarrow 0$  então  $w \rightarrow 0$  e  $f'(a) \cdot \frac{v}{|v|}$  é limitada. Portanto  $\lim_{v \rightarrow 0} C(v) = 0$ , provando o teorema.  $\square$

**Corolário 1.** Se  $f : U \rightarrow \mathbb{R}^n$  e  $g : V \rightarrow \mathbb{R}^p$  (com  $U \subset \mathbb{R}^m$  e  $f(U) \subset V \subset \mathbb{R}^n$ ) são de classe  $C^k$  então  $g \circ f : U \rightarrow \mathbb{R}^p$  é de classe  $C^k$ .

Com efeito a Regra da Cadeia, aplicada num ponto genérico  $x \in U$ , lê-se  $(g \circ f)'(x) = g'(f(x)) \cdot f'(x)$ . Em termos funcionais, temos

$$(g \circ f)' = (g' \circ f) \cdot f' : U \rightarrow \mathcal{L}(\mathbb{R}^m; \mathbb{R}^p),$$

onde  $\circ$  é a composição de aplicações e  $\cdot$  é a multiplicação de transformações lineares, a qual é bilinear, logo  $C^\infty$ . Se  $f$  e  $g$  são de classe  $C^1$ , esta última igualdade mostra que  $(g \circ f)'$  é contínua, logo  $g \circ f \in C^1$ . Por indução, supondo  $f$  e  $g$  de classe  $C^k$ , a mesma igualdade mostra que  $(g \circ f)' \in C^{k-1}$ , logo  $g \circ f \in C^k$ .

**Corolário 2.** Nas condições do Teorema 1, a matriz jacobiana de  $g \circ f$  no ponto  $a$  é o produto da matriz jacobiana de  $g$  no ponto  $f(a)$  pela matriz jacobiana de  $f$  no ponto  $a$ :  $J(g \circ f)(a) = Jg(f(a)) \cdot Jf(a)$ .

Em termos de derivadas parciais, a igualdade acima lê-se

$$\frac{\partial g_i}{\partial x_j} = \sum_{k=1}^n \frac{\partial g_i}{\partial y_k} \cdot \frac{\partial y_k}{\partial x_j}.$$

Nesta fórmula, escrita da maneira tradicional, os  $x_j$  são coordenadas de um ponto em  $U$ , os  $y_k$  em  $V$ ,  $\partial g_i / \partial x_j$  é derivada parcial de  $g_i \circ f$  enquanto  $\partial g_i / \partial y_k$  é derivada de  $g_i$  e, finalmente,  $\partial y_k / \partial x_j$  significa, em nossa notação costumeira,  $\partial f_k / \partial x_j$ . Com tais entendimentos tácitos, essa fórmula tem sobrevivido e sido útil através dos anos.

**Corolário 3 (As regras de derivação).** Sejam  $f, g : U \rightarrow \mathbb{R}^n$  diferenciáveis no ponto  $a \in U \subset \mathbb{R}^m$ ,  $\alpha$  um número real e  $B : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^p$  bilinear. Então:

- 1)  $f + g : U \rightarrow \mathbb{R}^n$  é diferenciável no ponto  $a$ , com  $(f + g)'(a) = f'(a) + g'(a)$ .
- 2)  $\alpha \cdot f : U \rightarrow \mathbb{R}^n$  é diferenciável no ponto  $a$ , com  $(\alpha f)'(a) = \alpha \cdot f'(a)$ .
- 3)  $B(f, g) : U \rightarrow \mathbb{R}^p$ , definida por  $B(f, g)(x) = B(f(x), g(x))$ , é diferenciável no ponto  $a$ , com

$$[B(f, g)]'(a) \cdot v = B(f'(a) \cdot v, g(a)) + B(f(a), g'(a) \cdot v).$$

Os itens 1) e 2) podem ser provados diretamente a partir da definição de aplicação diferenciável ou então considerando as transformações lineares  $S : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\alpha^* : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , definidas por  $S(x, y) = x + y$  e  $\alpha^*(x) = \alpha \cdot x$ . Então é só

observar que  $f + g = S \circ (f, g)$  e  $\alpha \cdot f = \alpha^* \circ f$  e usar a Regra da Cadeia, lembrando que  $S' = S$  e  $(\alpha^*)' = \alpha^*$ , logo

$$(f + g)' = S \circ (f', g') = f' + g' \quad \text{e} \quad (\alpha \cdot f)' = (\alpha^* \circ f)' = \alpha^* \circ f' = \alpha \cdot f'.$$

Quanto ao item 3), basta usar a Regra da Cadeia e os Exemplos 5, 7. Então, como  $B(f, g) = B \circ (f, g)$ , temos, para cada  $v \in \mathbb{R}^m$ :

$$\begin{aligned} [B(f, g)]'(a) \cdot v &= [B \circ (f, g)]'(a) \cdot v = B'(f(a), g(a)) \cdot (f'(a) \cdot v, g'(a) \cdot v) \\ &= B(f'(a) \cdot v, g(a)) + B(f(a), g'(a) \cdot v) \quad \square \end{aligned}$$

**Observação.** Uma aplicação bilinear  $B : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^p$  pode (e deve) ser considerada como uma forma de multiplicar um elemento de  $\mathbb{R}^m$  por outro de  $\mathbb{R}^n$  obtendo um produto em  $\mathbb{R}^p$ . Usando a notação multiplicativa  $x \bullet y$  em vez de  $B(x, y)$ , a regra de derivação do item 3) do Corolário 3 lê-se  $(f \bullet g)' = f' \bullet g + f \bullet g'$  isto é, para todo  $x \in U$  e todo  $v \in \mathbb{R}^p$ ,  $(f \bullet g)'(x) \cdot v = (f'(x) \cdot v) \bullet g(x) + f(x) \bullet g'(x) \cdot v$ . (O ponto maior  $\bullet$  é o produto que substitui  $B$  e o ponto menor  $\cdot$  é a aplicação de uma transformação linear sobre um vetor.)

**Exemplo 8.** Um exemplo frequente de aplicação bilinear é o produto interno de vetores. Se tivermos  $\varphi(x) = \langle f(x), g(x) \rangle$ , com  $f, g : U \rightarrow \mathbb{R}^n$  diferenciáveis em  $U \subset \mathbb{R}^m$  então, para todo  $v \in \mathbb{R}^m$ , vale  $\varphi'(x) \cdot v = \langle f'(x) \cdot v, g(x) \rangle + \langle f(x), g'(x) \cdot v \rangle$ . Em particular, se  $\varphi(x) = \langle f(x), f(x) \rangle = |f(x)|^2$  então  $\psi'(x) \cdot v = 2\langle f(x), f'(x) \cdot v \rangle$ . Levando em conta a fórmula  $(\sqrt{u})' = u'/2\sqrt{u}$  para a derivada da raiz quadrada de uma função real positiva  $u$ , daí resulta que, pondo

$$\xi(x) = \sqrt{\langle f(x), f(x) \rangle} = |f(x)|, \quad \text{tem-se} \quad \xi'(x) \cdot v = \langle f'(x) \cdot v, f(x) \rangle / |f(x)|$$

sempre que  $f(x) \neq 0$ . ◁

**Exemplo 9.** Outro exemplo comum de aplicação bilinear é a multiplicação de matrizes (ou de transformações lineares). Vejamos um caso particular desta situação. Se  $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$  é um operador auto-adjunto então resulta do exemplo anterior que a derivada da forma quadrática  $\varphi(x) = \langle Ax, x \rangle$  atua assim:  $\varphi'(x) \cdot v = \langle Av, x \rangle + \langle Ax, v \rangle = 2\langle Ax, v \rangle$ , levando em conta que  $\langle Av, x \rangle = \langle v, Ax \rangle$ , pela definição de operador auto-adjunto. Algumas pessoas preferem considerar  $A$  como uma matriz  $m \times m$  e  $x \in \mathbb{R}^m$  como uma matriz  $\mathbf{x}$  do tipo  $m \times 1$  (matriz-coluna) cuja transposta  $\mathbf{x}^T$  é uma matriz-linha  $1 \times m$ . Então a forma quadrática  $\varphi$  se escreve como  $\varphi(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$ . Desta maneira, para cada vetor  $v \in \mathbb{R}^m$  (ou seja, para cada matriz  $\mathbf{v}$  do tipo  $m \times 1$ ) tem-se  $\varphi'(\mathbf{x}) \cdot \mathbf{v} = \mathbf{v}^T \mathbf{A} \mathbf{x} + \mathbf{x}^T \mathbf{A} \mathbf{v} = \mathbf{x}^T \mathbf{A} \mathbf{v} + \mathbf{x}^T \mathbf{A} \mathbf{v} = 2\mathbf{x}^T \mathbf{A} \mathbf{v}$ , que corresponde a  $2\langle Ax, v \rangle$  na notação de operadores. ◁

**Exemplo 10.** Seja  $U \subset M(n \times n)$  o conjunto das matrizes invertíveis  $n \times n$ , isto é, das matrizes que têm determinante  $\neq 0$ . Como o determinante é uma função contínua,  $U$  é aberto. Seja  $f : U \rightarrow M(n \times n)$  a aplicação que associa a cada  $\mathbf{x} \in U$  sua inversa  $f(\mathbf{x}) = \mathbf{x}^{-1}$ . Afirmamos que  $f$  é diferenciável e que, em cada ponto  $\mathbf{x} \in U$ , sua derivada  $f'(\mathbf{x}) : M(n \times n) \rightarrow M(n \times n)$  é a transformação linear definida por  $f'(\mathbf{x}) \cdot \mathbf{v} = -\mathbf{x}^{-1} \cdot \mathbf{v} \cdot \mathbf{x}^{-1}$ ,  $\mathbf{v} \in M(n \times n)$ . Para provar isto, atribuiremos a cada matriz  $\mathbf{x} \in M(n \times n)$  a norma  $|\mathbf{x}|$ , igual à norma da transformação linear  $X : \mathbb{R}^n \rightarrow \mathbb{R}^n$  que tem  $\mathbf{x}$  como matriz na base canônica. (Veja Exemplo 12, Capítulo 1.) Mais explicitamente:  $|\mathbf{x}| = \sup\{|X \cdot u|; u \in S^{n-1}\}$ . Como se vê facilmente, se  $\mathbf{x}, \mathbf{y} \in M(n \times n)$  então  $|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}| \cdot |\mathbf{y}|$ . Provemos agora a diferenciabilidade de  $f$ . Escrevemos

$$(\mathbf{x} + \mathbf{v})^{-1} - \mathbf{x}^{-1} = -\mathbf{x}^{-1}\mathbf{v}\mathbf{x}^{-1} + r(\mathbf{v})$$

e mostramos que  $\lim_{\mathbf{v} \rightarrow 0} r(\mathbf{v})/|\mathbf{v}| = 0$ . Com este objetivo, multiplicamos ambos os membros da igualdade acima, à direita, por  $\mathbf{x} + \mathbf{v}$ . Após uma simplificação óbvia, obtemos

$$r(\mathbf{v}) = (\mathbf{x}^{-1} \cdot \mathbf{v})^2 (\mathbf{x} + \mathbf{v})^{-1},$$

donde

$$|r(\mathbf{v})| \leq |\mathbf{x}^{-1}|^2 |\mathbf{v}|^2 |(\mathbf{x} + \mathbf{v})^{-1}|, \quad \frac{|r(\mathbf{v})|}{|\mathbf{v}|} \leq |\mathbf{x}^{-1}|^2 |(\mathbf{x} + \mathbf{v})^{-1}| |\mathbf{v}|$$

e daí  $\lim_{\mathbf{v} \rightarrow 0} r(\mathbf{v})/|\mathbf{v}| = 0$ . (O uso de  $(\mathbf{x} + \mathbf{v})^{-1}$  se justifica pelo fato de que, sendo  $U$  aberto,  $\mathbf{x} \in U \Rightarrow \mathbf{x} + \mathbf{v} \in U$  para toda  $\mathbf{v}$  suficientemente pequena.)  $\triangleleft$

**Observação.** Na verdade a inversão de matrizes  $f : U \rightarrow U$ , considerada no Exemplo 10, é uma aplicação  $C^\infty$ . Isto pode ser verificado diretamente, a partir da fórmula que exprime  $\mathbf{x}^{-1}$  em função de  $\mathbf{x}$ , utilizando a chamada “adjunta clássica” de uma matriz. (Ver também “Análise no Espaço  $\mathbb{R}^n$ ”, página 26.)

Lembremos que a *norma* de uma transformação linear  $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é o número  $|T| = \sup\{|T \cdot u|; u \in S^{m-1}\}$ . Desta definição resulta que, para todo  $v \in \mathbb{R}^m$ , tem-se  $|T \cdot v| \leq |T| \cdot |v|$  e que, se  $S : \mathbb{R}^n \rightarrow \mathbb{R}^p$  é outra transformação linear então  $|S \cdot T| \leq |S| \cdot |T|$ .

**Teorema 2 (Desigualdade do Valor Médio).** *Seja  $f : U \rightarrow \mathbb{R}^n$  diferenciável em todos os pontos do segmento de reta  $[a, a + v] \subset U$ . Se, para todo  $t \in [0, 1]$ , tem-se  $|f'(a + tv)| \leq M$  então  $|f(a + v) - f(a)| \leq M \cdot |v|$ .*

*Demonstração.* O caminho  $\lambda : [0, 1] \rightarrow \mathbb{R}^n$ , definido por  $\lambda(t) = f(a + tv)$ , é diferenciável, com  $\lambda'(t) = f'(a + tv) \cdot v$ , portanto  $|\lambda'(t)| \leq |f'(a + tv)| \cdot |v| \leq M \cdot |v|$  para todo  $t \in [0, 1]$ . Segue-se então da Desigualdade do Valor Médio para caminhos (Teorema 1 do Capítulo 2) que  $|\lambda(1) - \lambda(0)| \leq M \cdot |v| \cdot (1 - 0)$ , isto é,  $|f(a + v) - f(a)| \leq M \cdot |v|$ .  $\square$

**Corolário 4.** *Se o aberto  $U \subset \mathbb{R}^m$  é convexo e  $M > 0$  é tal que a aplicação diferenciável  $f : U \rightarrow \mathbb{R}^n$  cumpre  $|f'(x)| \leq M$  para todo  $x \in U$  então  $f$  satisfaz a condição de Lipschitz  $|f(x) - f(y)| \leq M|x - y|$  para quaisquer  $x, y \in U$ .*

**Teorema 3 (Diferenciabilidade Uniforme).** *Seja  $f : U \rightarrow \mathbb{R}^n$  de classe  $C^1$  no aberto  $U \subset \mathbb{R}^m$ . Se  $K \subset U$  é compacto então  $f$  é uniformemente diferenciável em  $K$ .*

*Demonstração.* Isto significa que, para todo  $\varepsilon > 0$  dado, pode-se obter  $\delta > 0$  tal que  $|v| < \delta$  implica

$$|f(x + v) - f(x) - f'(x) \cdot v| < \varepsilon \cdot |v|$$

qualquer que seja  $x \in K$ . Para estabelecer este resultado, uma vez dado  $\varepsilon > 0$ , devemos inicialmente encontrar  $\delta > 0$  com a seguinte propriedade: para todo  $x \in K$  e todo  $v \in \mathbb{R}^n$  com  $|v| \leq \delta$  tem-se  $x + v \in U$  e  $|f'(x + v) - f'(x)| < \varepsilon$ . Ora, pelo Corolário 2 do Teorema 11, Capítulo 1, existe  $\delta > 0$  tal que toda bola de centro num ponto  $x \in K$  e raio  $2\delta$  está contida em  $U$ . Seja

$$L = \bigcup_{x \in K} B[x; \delta] = \{y \in \mathbb{R}^n; d(y, K) \leq \delta\}.$$

Então  $L$  é um compacto, com  $K \subset L \subset U$ . Se  $x \in K$  e  $|v| \leq \delta$  então  $x + v \in L$ . A aplicação  $f' : L \rightarrow \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$  é uniformemente contínua. Logo, diminuindo  $\delta$  se necessário, podemos admitir que  $|f'(x + v) - f'(x)| < \varepsilon$  para todo  $x \in K$  e todo  $v \in \mathbb{R}^n$  com  $|v| < \delta$ . Evidentemente, isto acarreta que  $|f'(x + tv) - f'(x)| < \varepsilon$  para todo  $t \in [0, 1]$ , pois  $|tv| \leq |v|$  quando  $0 \leq t \leq 1$ . Cumprida esta etapa, consideremos o caminho  $\lambda : [0, 1] \rightarrow \mathbb{R}^n$ , definido por  $\lambda(t) = f(x + tv)$ , com  $x \in K$  e  $|v| < \delta$ . Então  $\lambda'(t) = f'(x + tv) \cdot v$ . Pelo Teorema Fundamental do Cálculo para caminhos,

$$f(x + v) - f(x) = \lambda(1) - \lambda(0) = \int_0^1 \lambda'(t) dt = \int_0^1 f'(x + tv) \cdot v \cdot dt.$$

Logo

$$\begin{aligned} |f(x + v) - f(x) - f'(x) \cdot v| &= \left| \int_0^1 [f'(x + tv) - f'(x)] \cdot v \cdot dt \right| \\ &\leq |f'(x + tv) - f'(x)| |v| \leq \varepsilon \cdot |v|. \end{aligned}$$

provando assim que  $f$  é uniformemente diferenciável em  $K$ .

□

## Capítulo 6

# Aplicações Inversas e Implícitas

### 1 O Teorema da Aplicação Inversa

Na página 97 do Volume 1 foi estabelecido que se  $f : I \rightarrow \mathbb{R}$  é derivável no intervalo  $I \subset \mathbb{R}$ , com  $f'(x) > 0$  para todo  $x \in I$ , então  $f$  é uma bijeção crescente sobre o intervalo  $J = f(I)$  e a função inversa  $g = f^{-1} : J \rightarrow I$  também é derivável, com  $g'(f(x)) = 1/f'(x)$ . Evidentemente, resultado análogo vale com  $f'(x) < 0$ , só que agora  $f$  é decrescente. Na verdade, pelo Teorema de Darboux (pág. 95 do Volume 1), bastaria supor  $f'(x) \neq 0$  para todo  $x \in I$  para garantir que  $f$  é uma bijeção monótona (crescente ou decrescente) de  $I$  sobre  $J = f(I)$ , com  $f^{-1} : J \rightarrow I$  derivável. Nos termos da definição que será dada a seguir, isto significa que a função diferenciável sobrejetiva  $f : I \rightarrow J$ , entre intervalos  $I, J \subset \mathbb{R}$ , é um difeomorfismo se, e somente se,  $f'(x) \neq 0$  para todo  $x \in I$ . Em dimensões superiores esta condição significaria que  $f'(x)$  é um isomorfismo, mas seria apenas necessária para que  $f$  possuísse uma inversa diferenciável.

Sejam  $U \subset \mathbb{R}^m$ ,  $V \subset \mathbb{R}^n$  abertos. Uma aplicação  $f : U \rightarrow V$  chama-se um *difeomorfismo* entre  $U$  e  $V$  quando é uma bijeção diferenciável, cuja inversa  $g = f^{-1} : V \rightarrow U$  também é diferenciável.

Se  $f : U \rightarrow V$  é um difeomorfismo, com  $g = f^{-1} : V \rightarrow U$ , então de  $g \circ f = id_U$  e  $f \circ g = id_V$  resulta, pela Regra da Cadeia, que  $g'(f(x)) \cdot f'(x) = id_{\mathbb{R}^m}$  e  $f'(x) \cdot g'(f(x)) = id_{\mathbb{R}^n}$  para todo  $x \in U$ , portanto  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é um isomorfismo cujo inverso é  $g'(f(x)) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Em particular,  $m = n$ , ou seja, dois abertos em espaços euclidianos de dimensões diferentes não podem ser difeomorfos.

**Exemplo 1.** No Capítulo 1 (Exemplo 15), vimos que a aplicação  $f : \mathbb{R}^m \rightarrow B$ , definida por  $f(x) = x/(1 + |x|)$ , é um homeomorfismo de  $\mathbb{R}^m$  sobre a bola aberta  $B \subset \mathbb{R}^m$ , de centro 0 e raio 1, sendo  $g : B \rightarrow \mathbb{R}^m$ ,  $g(y) = y/(1 - |y|)$  o homeomorfismo inverso de  $f$ . Na verdade, como  $f$  e  $g$  são ambas aplicações

diferenciáveis, as duas são difeomorfismos, um inverso do outro. Deve-se observar, entretanto, que nem todo homeomorfismo diferenciável é um difeomorfismo, isto é, tem inverso diferenciável. O exemplo mais simples disto é  $f : \mathbb{R} \rightarrow \mathbb{R}$ , com  $f(x) = x^3$ . Como  $f'(0) = 0$ , a função inversa de  $f$  (que é  $g(x) = \sqrt[3]{x}$ ) não é diferenciável no ponto  $0 = f(0)$ .

Uma aplicação diferenciável  $f : U \rightarrow \mathbb{R}^m$ , definida no aberto  $U \subset \mathbb{R}^m$ , chama-se um *difeomorfismo local* quando, para cada  $x \in U$  existe uma bola aberta  $B = B(x; \delta) \subset U$  tal que  $f$  aplica  $B$  difeomorficamente sobre um aberto  $V$  contendo  $f(x)$ . Segue-se daí que se  $f : U \rightarrow \mathbb{R}^m$  é um difeomorfismo local então  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$  é um isomorfismo, para todo  $x \in U$ . O Teorema da Aplicação Inversa, que provaremos a seguir, diz que quando  $f \in C^1$  vale a recíproca: se  $f'(x)$  é um isomorfismo para todo  $x \in U$  então  $f$  é um difeomorfismo local.

Decorre da definição acima que um difeomorfismo local  $f : U \rightarrow \mathbb{R}^m$  é uma *aplicação aberta*, isto é, a imagem  $f(A)$  de qualquer aberto  $A \subset U$  é um subconjunto aberto de  $\mathbb{R}^m$ . Com efeito, se tomarmos para cada  $x \in A$  uma bola aberta  $B_x \subset A$ , com centro  $x$ , tal que  $f$  seja um difeomorfismo de  $B_x$  sobre um aberto  $V_x \subset \mathbb{R}^m$ , então  $A = \bigcup_{x \in A} B_x$  e  $f(A) = f(\bigcup_{x \in A} B_x) = \bigcup_{x \in A} f(B_x) = \bigcup_{x \in A} V_x$  é uma reunião de abertos, logo é um aberto.

Observemos ainda que o difeomorfismo local  $f : U \rightarrow \mathbb{R}^m$  é um difeomorfismo (global) de  $U$  sobre o aberto  $V = f(U) \subset \mathbb{R}^m$  se, e somente se, é uma aplicação injetiva.  $\triangleleft$

**Exemplo 2.** Seja  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  definida por  $f(x, y) = (e^x \cos y, e^x \sin y)$ . Evidentemente,  $f \in C^\infty$ . Cada reta vertical  $x = a$  é transformada por  $f$ , com período  $2\pi$  (isto é,  $f(a, y) = f(a, y') \Leftrightarrow y' - y = 2k\pi, k \in \mathbb{Z}$ ), sobre a circunferência de centro 0 e raio  $e^a$ . Cada reta horizontal  $y = b$  é levada por  $f$ , bijectivamente, sobre a semi-reta aberta que parte da origem e passa pelo ponto  $(\cos b, \sin b) \in S^1$ . A imagem de  $f$  é  $\mathbb{R}^2 - \{0\}$ . Em termos da variável complexa  $z = x + iy$ , tem-se  $f(z) = e^z$ . A aplicação  $f$  é um difeomorfismo local (mas não global pois  $f(x, y + 2\pi) = f(x, y)$ ). Isto decorre do Teorema da Aplicação Inversa (Teorema 4, a seguir), pois a matriz jacobiana

$$Jf(x, y) = \begin{bmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{bmatrix}$$

tem determinante  $e^x$ , portanto  $\neq 0$ , logo  $f'(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  é um isomorfismo, para todo  $(x, y) \in \mathbb{R}^2$ . Podemos também chegar à mesma conclusão observando que se  $w_0 = f(z_0)$  então o ramo da função complexa  $\log w$  tal que  $\log w_0 = z_0$  é uma aplicação inversa local de  $f$  no ponto  $w_0 = f(z_0)$ .

Se  $I \subset \mathbb{R}$  é um intervalo aberto então todo difeomorfismo local  $f : I \rightarrow \mathbb{R}$  é um difeomorfismo (global) de  $I$  sobre  $J = f(I)$ .  $\triangleleft$

**Teorema 1.** *Se o difeomorfismo  $f : U \rightarrow V$  é de classe  $C^k$  ( $k \geq 1$ ) então seu inverso  $g = f^{-1} : V \rightarrow U$  também é de classe  $C^k$ .*

*Demonstração.* (Indução em  $k$ ). Para todo  $y = f(x) \in V$ , temos  $g'(y) = [f'(x)]^{-1} = [f'(f^{-1}(y))]^{-1}$ , portanto a aplicação  $g' : V \rightarrow \mathcal{L}(\mathbb{R}^m) = \mathbb{R}^{m^2}$  se exprime como a composta

$$g' = (\text{Inv}) \circ f' \circ f^{-1}$$

onde  $\text{Inv}$  leva todo operador invertível  $X : \mathbb{R}^m \rightarrow \mathbb{R}^m$  no seu inverso  $X^{-1}$ ,  $f' : U \rightarrow \mathcal{L}(\mathbb{R}^m)$ ,  $f'$  leva todo ponto  $x \in U$  na derivada (invertível)  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$  e  $f^{-1} : V \rightarrow U$  é aplicação inversa de  $f$ . Sabemos que  $\text{Inv} \in C^\infty$ . Portanto, se  $f \in C^k$  então  $f' \in C^{k-1}$  e, pela hipótese de indução,  $f^{-1} \in C^{k-1}$ , logo  $g' \in C^{k-1}$ , como composta de três aplicações de classe  $C^{k-1}$ . Por definição, isto significa que  $g \in C^k$ .  $\square$

**Teorema 2.** *Seja  $f : U \rightarrow \mathbb{R}^n$  de classe  $C^1$  no aberto  $U \subset \mathbb{R}^m$ . Se, para algum  $a \in U$ , a derivada  $f'(a) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é injetiva então existem  $\delta > 0$  e  $c > 0$  tais que  $B = B(a; \delta) \subset U$  e, para quaisquer  $x, y \in B$  tem-se  $|f(x) - f(y)| \geq c|x - y|$ . Em particular, a restrição  $f|_B$  é injetiva.*

*Demonstração.* A função  $u \mapsto |f'(a) \cdot u|$  é positiva em todos os pontos  $u$  da esfera unitária  $S^{m-1}$ , a qual é compacta. Pelo Teorema de Weierstrass, existe  $c > 0$  tal que  $|f'(a) \cdot u| \geq 2c$  para todo  $u \in S^{m-1}$ . Por linearidade, segue-se que  $|f'(a) \cdot v| \geq 2c \cdot |v|$  para todo  $v \in \mathbb{R}^m$ . Para todo  $x \in U$ , escrevamos

$$r(x) = f(x) - f(a) - f'(a)(x - a).$$

Então, para  $x, y \in U$  quaisquer, temos

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

Levando em conta que  $|u + v| \geq |u| - |v|$ , segue-se que

$$\begin{aligned} |f(x) - f(y)| &\geq |f'(a) \cdot (x - y)| - |r(x) - r(y)| \\ &\geq 2c \cdot |x - y| - |r(x) - r(y)|. \end{aligned}$$

Observemos que a aplicação  $r$ , acima definida, é de classe  $C^1$ , com  $r(a) = 0$ . Pela continuidade de  $r'$ , existe  $\delta > 0$  tal que  $|x - a| < \delta \Rightarrow x \in U$  e  $|r'(x)| < c$ . A Desigualdade do Valor Médio, aplicada a  $r$  no conjunto convexo  $B = B(x; \delta)$  nos assegura que se  $x, y \in B$  então  $|r(x) - r(y)| \leq c|x - y|$ . Conseqüentemente,  $x, y \in B \Rightarrow |f(x) - f(y)| \geq 2c|x - y| - c|x - y|$ , ou seja,  $|f(x) - f(y)| \geq c|x - y|$ , como queríamos demonstrar.  $\square$

**Teorema 3 (Diferenciabilidade do Homeomorfismo Inverso).** *Seja  $f : U \rightarrow V$  um homeomorfismo diferenciável entre os abertos  $U, V \subset \mathbb{R}^m$ . Se, para algum  $x \in U$ , a derivada  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$  é um operador invertível então o homeomorfismo inverso  $g = f^{-1} : V \rightarrow U$  é diferenciável no ponto  $f(x)$ , com  $g'(f(x)) = [f'(x)]^{-1}$ .*

*Demonstração.* Se  $x, x + v \in U$ , escrevamos  $f(x) = y$  e  $f(x + v) = y + w$ . Então

$$\begin{aligned} w &= f(x + v) - f(x) = f'(x) \cdot v + r(v) \quad \text{onde} \quad \lim_{v \rightarrow 0} \frac{r(v)}{|v|} = 0 \quad \text{e} \\ v &= g(f(x + v)) - g(f(x)) = g(y + w) - g(y). \end{aligned}$$

Para provar que  $f'(x)^{-1}$  é a derivada de  $g$  no ponto  $y$ , escrevamos

$$g(y + w) - g(y) = f'(x)^{-1} \cdot w + s(w) \quad (*)$$

e mostremos que  $\lim_{w \rightarrow 0} \frac{s(w)}{|w|} = 0$ . Entrando na igualdade (\*) com as expressões de  $v$  e  $w$  acima obtidas, vem:

$$v = f'(x)^{-1} [f'(x) \cdot v + r(v)] + s(w),$$

ou seja:

$$v = v + f'(x)^{-1} \cdot r(v) + s(w),$$

donde

$$s(w) = -f'(x)^{-1} \cdot r(v), \quad \text{logo} \quad \frac{s(w)}{|w|} = -f'(x)^{-1} \cdot \frac{r(v)}{|v|} \cdot \frac{|v|}{|w|},$$

isto é:

$$\frac{s(w)}{|w|} = -f'(x)^{-1} \cdot \frac{r(v)}{|v|} \cdot \frac{|v|}{|f(x + v) - f(x)|}.$$

Quando  $w \rightarrow 0$ , tem-se  $v \rightarrow 0$  pela continuidade de  $g$ , logo  $\frac{r(v)}{|v|} \rightarrow 0$ . Além disso, pelo Teorema 2, existem  $\delta > 0$  e  $c > 0$  tais que  $|v| < \delta$  implica

$$|f(x + v) - f(x)| \geq c|v|, \quad \text{portanto} \quad \frac{|v|}{|f(x + v) - f(x)|} \leq \frac{1}{c}.$$

Assim,  $\lim_{w \rightarrow 0} \frac{s(w)}{|w|} = 0$ . □

**Corolário 1.** Se  $f : U \rightarrow V$  é um homeomorfismo de classe  $C^k$  cuja derivada  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^m$  é invertível para todo  $x \in U$  então seu inverso  $g = f^{-1} : V \rightarrow U$  é de classe  $C^k$ .

Com efeito, a derivada  $g' : V \rightarrow \mathcal{L}(\mathbb{R}^m)$ , dada por  $g'(y) = f'(g(y))^{-1}$  para cada  $y \in V$ , pode ser escrita como  $g' = f' \circ \text{Inv} \circ g$ , onde a aplicação  $\text{Inv}$ , de classe  $C^\infty$ , é a inversão de transformações lineares bijetivas e  $f' \in C^{k-1}$ . Admitindo, por indução, que  $g \in C^{k-1}$ , resulta que  $g' \in C^{k-1}$ , logo  $g \in C^k$ .

**Teorema 4 (Teorema da Aplicação Inversa).** Seja  $f : U \rightarrow \mathbb{R}^m$  de classe  $C^k$  ( $k \geq 1$ ) no aberto  $U \subset \mathbb{R}^m$ . Se  $a \in U$  é tal que  $f'(a) : \mathbb{R}^m \rightarrow \mathbb{R}^m$  é invertível então existe uma bola aberta  $B = B(a; \delta) \subset U$  tal que a restrição  $f|_B$  é um difeomorfismo sobre um aberto  $V \ni f(a)$ .

*Demonstração.* Diminuindo  $\delta$ , se necessário, no Teorema 2 podemos admitir que  $\bar{B} = B[a; \delta] \subset U$  e que  $f$  é injetiva no conjunto compacto  $\bar{B}$ , logo é um homeomorfismo de  $B$  sobre  $V = f(B)$ . Pelo Teorema 3, basta então mostrar que  $V \subset \mathbb{R}^m$  é aberto. Seja então  $q = f(p)$ ,  $p \in B$ . Chamando de  $S = S[a, \delta]$  a esfera que é a fronteira de  $\bar{B}$ , a injetividade de  $f|_{\bar{B}}$  assegura que  $q \notin f(S)$ , logo existe  $\varepsilon > 0$  tal que  $|f(x) - q| \geq 2\varepsilon$  para todo  $x \in S$ , pois  $f(S)$  é compacto. Afirmamos que  $B(q; \varepsilon) \subset f(B)$ . Com efeito, se  $y \in B(q; \varepsilon)$ , então o mínimo de  $g(x) = |f(x) - y|$ , quando  $x$  varia no compacto  $\bar{B}$ , não é atingido num ponto  $x \in S$  pois  $x \in S \Rightarrow |f(x) - y| \geq \varepsilon$  enquanto  $|f(p) - y| = |q - y| < \varepsilon$ , com  $p \in B$ . Assim, o mínimo de  $|f(x) - y|$ ,  $x \in \bar{B}$  é atingido num ponto  $x_0 \in B$ . Pelo lema a seguir, isto implica que esse mínimo é zero, portanto  $y = f(x_0)$ , donde  $y \in f(B)$ , ou seja,  $B(q; \varepsilon) \subset f(B)$ .  $\square$

**Lema 1.** Sejam  $U \subset \mathbb{R}^m$  aberto e  $g : U \rightarrow \mathbb{R}^n$  diferenciável no ponto  $a \in U$ , com  $g'(a) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  sobrejetiva. Se  $a$  é um ponto de mínimo local de  $|g(x)|$ ,  $x \in U$ , então  $g(a) = 0$ .

*Demonstração.* Se  $a$  é um ponto de mínimo local para  $|g(x)|$ , será também um ponto de mínimo local para a função  $\varphi : U \rightarrow \mathbb{R}$ , definida por  $\varphi(x) = |g(x)|^2 = \langle g(x), g(x) \rangle$ , logo  $\varphi'(a) = 0$ . Mas, como  $\varphi'(a) \cdot v = 2\langle g'(a) \cdot v, g(a) \rangle$ , isto significa que  $g(a)$  é ortogonal à imagem de  $g'(a)$ , a qual é  $\mathbb{R}^n$ . Logo  $g(a) = 0$ .  $\square$

**Exemplo 3.** Dadas as matrizes  $\mathbf{x}, \mathbf{m} \in M(n \times n)$ , diz-se que  $\mathbf{x}$  é uma raiz quadrada de  $\mathbf{m}$  quando  $\mathbf{x}^2 = \mathbf{m}$ . Nem toda matriz  $\mathbf{m}$  possui raiz quadrada: como  $\det(\mathbf{x}^2) = (\det \mathbf{x})^2$ , uma condição necessária é que  $\det \mathbf{m} \geq 0$ . Mas esta condição não é suficiente pois é fácil ver que, embora a matriz  $\mathbf{m} = \begin{bmatrix} -1 & 0 \\ 1 & -1 \end{bmatrix}$  tenha determinante positivo, não existe  $\mathbf{x} \in M(2 \times 2)$  tal que  $\mathbf{x}^2 = \mathbf{m}$ . O Teorema 4 pode ser usado

para mostrar que toda matriz próxima da identidade  $\mathbf{I}_n$  tem raiz quadrada. Com efeito, consideremos a aplicação  $f : M(n \times n) \rightarrow M(n \times n)$ ,  $f(\mathbf{x}) = \mathbf{x}^2$ , de classe  $C^\infty$ . Sua derivada num ponto  $\mathbf{x} \in M(n \times n)$  é a transformação linear  $f'(\mathbf{x}) : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$ , dada por  $f'(\mathbf{x}) \cdot \mathbf{m} = \mathbf{x} \cdot \mathbf{m} + \mathbf{m} \cdot \mathbf{x}$ . Em particular, para  $\mathbf{x} = \mathbf{I}_n$ , tem-se  $f'(\mathbf{I}_n) \cdot \mathbf{m} = 2\mathbf{m}$ , logo  $f'(\mathbf{I}_n) : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$  é um isomorfismo. Segue-se do Teorema 4 que existe um aberto  $U$  em  $M(n \times n)$ , contendo a matriz identidade, restrita ao qual  $f$  é um difeomorfismo sobre o aberto  $V = f(U)$ . Assim, para toda matriz  $\mathbf{y} \in V$ , existe uma única matriz  $\mathbf{x} = \sqrt{\mathbf{y}} \in U$  tal que  $\mathbf{x}^2 = \mathbf{y}$ . Além disso, a aplicação  $f^{-1} : V \rightarrow U$ ,  $\mathbf{y} \mapsto \sqrt{\mathbf{y}}$ , é de classe  $C^\infty$ .  $\triangleleft$

**Corolário 2 (do Teorema 4).** *Seja  $a \in U$  um ponto crítico da função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^2$  no aberto  $U \subset \mathbb{R}^n$ . Se a matriz hessiana*

$$Hf(a) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right]$$

*é invertível então existe um aberto  $V$ , com  $a \in V \subset U$ , no qual não há outros pontos críticos de  $f$ .*

Com efeito, a matriz hessiana  $Hf(x)$  é, para todo  $x \in U$ , a matriz jacobiana da aplicação

$$F : U \rightarrow \mathbb{R}^n, F(x) = \text{grad } f(x) = \left( \frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right).$$

Como  $Hf(a)$  é invertível,  $F$  é injetiva numa vizinhança  $V \ni a$ , logo  $F(x) \neq F(a)$ , isto é,  $\text{grad } f(x) \neq 0$  para todo  $x \in V$ .

Quando  $\text{grad } f(a) = 0$  e  $Hf(a)$  é invertível,  $a$  chama-se um *ponto crítico não-degenerado* da função  $f$ . O corolário acima diz que os pontos críticos não-degenerados são pontos críticos isolados.

## 2 Várias Funções Implícitas

Os pontos do espaço  $\mathbb{R}^{m+n}$  serão representados sob a forma  $z = (x, y)$ , onde  $x = (x_1, \dots, x_m) \in \mathbb{R}^m$  e  $y = (y_1, \dots, y_n) \in \mathbb{R}^n$ . Um difeomorfismo  $h : U \rightarrow V$ , entre abertos  $U, V \subset \mathbb{R}^{m+n}$ , será chamado de *vertical* quando for do tipo  $h(x, y) = (x, h_2(x, y))$ , ou seja, quando deixar invariante a coordenada  $x$ . O inverso de um difeomorfismo vertical é ainda vertical.

Um difeomorfismo  $\varphi : U \rightarrow V$  é usualmente interpretado como uma transformação geométrica que aplica diferencialmente o conjunto  $U$  sobre o conjunto  $V$ , de forma invertível. Às vezes, porém, é conveniente olhar para  $\varphi$  como uma mudança de coordenadas, em que as coordenadas do ponto  $x \in U$  passam a ser

aquelas da sua imagem  $y = \varphi(x) \in V$ . Sob este ponto de vista, o teorema a seguir diz que se a derivada de uma aplicação  $f$ , de classe  $C^k$ , é sobrejetiva num ponto  $p$  então é possível obter (de modo bastante simples) um sistema de coordenadas, válido numa vizinhança aberta  $Z$  de  $p$ , tal que, em termos dessas novas coordenadas, a aplicação  $f$  assume a expressão

$$(x_1, \dots, x_m, w_1, \dots, w_n) \mapsto (w_1, \dots, w_n).$$

**Teorema 5 (Forma Local das Submersões).** *Seja  $f = (f_1, \dots, f_n)$  uma aplicação de classe  $C^k$  ( $k \geq 1$ ) de um aberto  $U \subset \mathbb{R}^{m+n}$  em  $\mathbb{R}^n$ . Se, num ponto  $p = (a, b) \in U$ , a matriz*

$$\left[ \frac{\partial f_i}{\partial y_j}(p) \right] \quad (i, j = 1, \dots, n)$$

*é invertível então existem abertos  $Z \ni p$  em  $\mathbb{R}^{m+n}$ ,  $V \ni a$  em  $\mathbb{R}^m$ ,  $W \ni c = f(p)$  em  $\mathbb{R}^n$  e um difeomorfismo vertical  $h : V \times W \rightarrow Z$ , de classe  $C^k$ , tal que  $f(h(x, w)) = w$  para todo  $x \in V$  e todo  $w \in W$ .*

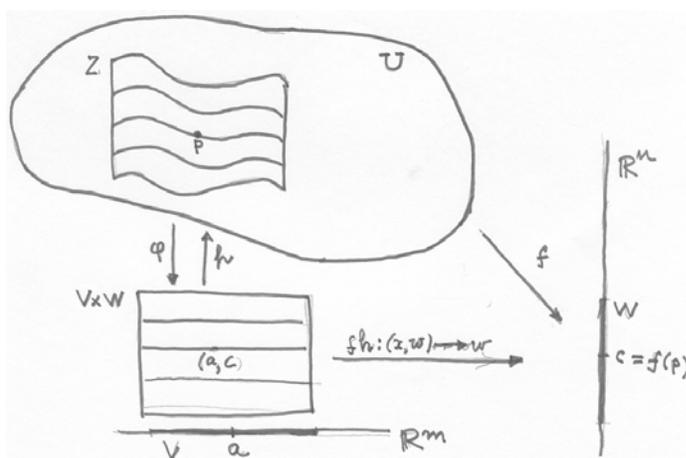


Figura 1.

*Demonstração.* Seja  $\varphi : U \rightarrow \mathbb{R}^m \times \mathbb{R}^n$  a aplicação de classe  $C^k$  definida por  $\varphi(x, y) = (x, f(x, y))$ . A matriz jacobiana de  $\varphi$  tem a forma

$$J\varphi = \begin{bmatrix} \mathbf{I} & \mathbf{a} \\ \mathbf{0} & \mathbf{b} \end{bmatrix},$$

onde  $\mathbf{I}$  é a matriz identidade  $m \times m$  e a matriz  $n \times n$

$$\mathbf{b} = \mathbf{b}(z) = \begin{bmatrix} \frac{\partial f_i}{\partial y_j}(z) \end{bmatrix}$$

é, no ponto  $p = (a, b)$ , invertível.

Pelo Teorema da Aplicação Inversa,  $\varphi$  é um difeomorfismo de um aberto  $Z \ni p$  sobre um aberto de  $\mathbb{R}^m \times \mathbb{R}^n$ , o qual podemos supor da forma  $V \times W$ , onde  $V \subset \mathbb{R}^m$  e  $W \subset \mathbb{R}^n$ , com  $a \in V$  e  $c = f(a, b) \in W$ . O difeomorfismo inverso  $h : V \times W \rightarrow Z$  é da forma  $h(x, w) = (x, h_2(x, w))$ . Então, para qualquer  $(x, w) \in V \times W$ , tem-se

$$\begin{aligned} (x, w) &= \varphi(h(x, w)) = \varphi(x, h_2(x, w)) \\ &= (x, f(x, h_2(x, w))) = (x, f(h(x, w))), \end{aligned}$$

logo  $f(h(x, w)) = w$  para qualquer  $(x, w) \in V \times W$ .  $\square$

Dada  $f : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$ , a matriz de sua derivada  $f'(p) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  tem  $n$  linhas e  $m+n$  colunas. Ela é a matriz jacobiana  $Jf(p)$ . Dizer que a transformação linear  $f'(p)$  é sobrejetiva significa afirmar que é possível escolher  $n$  dessas colunas de modo que a matriz  $n \times n$  resultante seja invertível. No enunciado do teorema acima, as colunas escolhidas são as  $n$  últimas porém isto nada tem de essencial; trata-se apenas de simplificar a notação.

Quando a aplicação  $f : U \rightarrow \mathbb{R}^n$ , com  $U \subset \mathbb{R}^{m+n}$ , possui derivada sobrejetiva  $f'(z) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  em todo ponto  $z \in U$ , diz-se que  $f$  é uma *submersão*. No Teorema 5, a restrição de  $f$  ao aberto  $Z$  é uma submersão.

Com esta terminologia, podemos enunciar o

**Corolário 3.** *Seja  $f : U \rightarrow \mathbb{R}^n$  uma submersão de classe  $C^k$ , definida no aberto  $U \subset \mathbb{R}^{m+n}$ . Para cada ponto  $z \in U$  existem abertos  $Z \subset U$ , contendo  $z$ ,  $W \subset \mathbb{R}^n$  contendo  $c = f(z)$ ,  $V \subset \mathbb{R}^m$  e um difeomorfismo  $h : V \times W \rightarrow Z$  de classe  $C^k$ , tais que  $f(h(x, w)) = w$  para todo  $x \in V$  e todo  $w \in W$ .*

Como  $f'(z) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  é sobrejetiva,  $n$  das  $m+n$  colunas da matriz jacobiana  $Jf(z)$  são linearmente independentes, logo formam uma matriz invertível  $n \times n$ . Se essas forem as últimas colunas, o corolário é meramente o Teorema 5. Se não forem, modificamos ligeiramente a demonstração daquele teorema, permutando inicialmente as coordenadas em  $\mathbb{R}^{m+n}$  de modo que as  $n$  colunas linearmente independentes de  $Jf(z)$  sejam agora as últimas.  $\square$

**Teorema 6 (Teorema das Funções Implícitas).** *Seja  $f = (f_1, \dots, f_n) : U \rightarrow \mathbb{R}^n$  de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$ . Suponhamos que, no ponto  $p = (a, b)$ ,*

com  $f(p) = c$ , a matriz  $n \times n$

$$\left[ \frac{\partial f_i}{\partial y_j}(p) \right] \quad (i, j = 1, \dots, n)$$

seja invertível. Então existem  $Z \subset U$ , aberto contendo  $p$ ,  $V \subset \mathbb{R}^m$ , aberto contendo  $a$ , e  $\xi : V \rightarrow \mathbb{R}^n$  de classe  $C^k$ , com  $\xi(a) = p$ , com a seguinte propriedade:

$$[(x, y) \in Z \text{ e } f(x, y) = c] \iff [x \in V \text{ e } y = \xi(x)].$$

A equivalência acima significa que  $f^{-1}(c) \cap Z$  é o gráfico de  $\xi$ , isto é,

$$f^{-1}(c) \cap Z = \{(x, \xi(x)); x \in V\}.$$

*Demonstração.* Sejam  $Z, V, W$  e  $h$  como no Teorema 5. Definamos  $\xi : V \rightarrow \mathbb{R}^n$  pondo  $\xi(x) = h_2(x, c)$ , onde  $h_2 : V \times W \rightarrow \mathbb{R}^n$  é a segunda coordenada de  $h$ , ou seja,  $h(x, w) = (x, h_2(x, w))$ . Assim,  $(x, y) \in Z \Rightarrow x \in V$  e  $(x, y) = h(x, w)$ ,  $w \in W$ . Se, além disso, tem-se  $f(x, y) = c$  então  $c = f(x, y) = f(x, \xi(x))$  e  $y = \xi(x)$ . Resumindo:  $(x, y) \in Z$  e  $f(x, y) = c$  implicam  $x \in V$  e  $y = \xi(x)$ . Reciprocamente, Se  $x \in V$  e  $y = \xi(x)$  então  $y = h_2(x, c)$  e  $f(x, y) = f(x, h_2(x, c)) = f(h(x, c)) = c$ .  $\square$

O corolário abaixo é uma reformulação mais intrínseca do Teorema 6.

**Corolário 4.** *Seja  $f : U \rightarrow \mathbb{R}^n$  de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$ . Se, no ponto  $p \in U$ , com  $f(p) = c$ , a derivada  $f'(p) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  é sobrejetiva então existe um aberto  $Z \subset U$ , com  $p \in Z$ , tal que  $f^{-1}(c) \cap Z$  é o gráfico de uma aplicação  $\xi : V \rightarrow \mathbb{R}^n$ , de classe  $C^k$  num aberto  $V \subset \mathbb{R}^m$ .*

A abordagem clássica do Teorema das Funções Implícitas era a seguinte: “Se  $f_1, \dots, f_n$  são funções reais de  $m + n$  variáveis,  $k$  vezes continuamente diferenciáveis, e  $p = (a_1, \dots, a_m, b_1, \dots, b_n)$  é uma solução particular do sistema de equações

$$f_1(x_1, \dots, x_m, y_1, \dots, y_n) = c_1$$

$$f_2(x_1, \dots, x_m, y_1, \dots, y_n) = c_2$$

$$\vdots$$

$$f_n(x_1, \dots, x_m, y_1, \dots, y_n) = c_n,$$

sendo a matriz  $n \times n$

$$\left[ \frac{\partial f_i}{\partial x_j}(p) \right]$$

invertível, então as equações acima definem, de modo único, na vizinhança do ponto  $p$  em  $\mathbb{R}^{m+n}$ , as variáveis  $y_1, \dots, y_n$  como funções de classe  $C^k$  das variáveis  $x_1, \dots, x_m$ :  $y_1 = \xi_1(x_1, \dots, x_m), \dots, y_n = \xi_n(x_1, \dots, x_m)$ ."

Escrevendo  $x = (x_1, \dots, x_m)$  e  $\xi(x) = (\xi_1(x), \dots, \xi_n(x))$  tem-se, para cada  $i = 1, \dots, n$ , com  $x$  numa vizinhança de  $a = (a_1, \dots, a_m)$ :

$$f_i(x, \xi_1(x), \dots, \xi_n(x)) = c_i, \quad \text{ou} \quad f_i(x, \xi(x)) = c_i.$$

Derivando cada uma dessas  $n$  identidades em relação a  $x_j$ , vem:

$$\frac{\partial f_i}{\partial x_j} + \sum_{k=1}^n \frac{\partial f_i}{\partial y_k} \cdot \frac{\partial \xi_k}{\partial x_j} = 0, \quad j = 1, \dots, m.$$

Em termos matriciais, isto significa que

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_j} \\ \vdots \\ \frac{\partial f_n}{\partial x_j} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1} & \cdots & \frac{\partial f_n}{\partial y_n} \end{bmatrix} \begin{bmatrix} \frac{\partial \xi_1}{\partial x_j} \\ \vdots \\ \frac{\partial \xi_n}{\partial x_j} \end{bmatrix}$$

ou seja:

$$\begin{bmatrix} \frac{\partial \xi_1}{\partial x_j} \\ \vdots \\ \frac{\partial \xi_n}{\partial x_j} \end{bmatrix} = - \begin{bmatrix} \frac{\partial f_1}{\partial y_1} & \cdots & \frac{\partial f_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_1} & \cdots & \frac{\partial f_n}{\partial y_n} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f_1}{\partial x_j} \\ \vdots \\ \frac{\partial f_n}{\partial x_j} \end{bmatrix}.$$

Isto exhibe as derivadas parciais  $\frac{\partial \xi_i}{\partial x_j}$  a partir de  $f_1, \dots, f_n$ , sem ser necessário conhecer explicitamente as funções  $\xi_1, \dots, \xi_n$ .

Sob o ponto de vista da Álgebra Linear intrínseca, a fim de mostrar como a derivada  $\xi'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  pode ser calculada quando se conhece  $f$  mas não  $\xi$  explicitamente, é preciso estender o conceito de derivada parcial.

As transformações lineares

$$\frac{\partial f}{\partial x}(z) : \mathbb{R}^m \rightarrow \mathbb{R}^n \quad \text{e} \quad \frac{\partial f}{\partial y}(z) : \mathbb{R}^n \rightarrow \mathbb{R}^n,$$

cujas matrizes nas bases canônicas dos espaços euclidianos em questão são

$$\left[ \frac{\partial f_i}{\partial x_j}(z) \right] \in M(n \times m) \quad \text{e} \quad \left[ \frac{\partial f_i}{\partial y_j}(z) \right] \in M(n \times n)$$

são chamadas as *derivadas parciais* de  $f$  no ponto  $z$ , relativamente à decomposição  $\mathbb{R}^{m+n} = \mathbb{R}^m \oplus \mathbb{R}^n$ , obtida ao se escrever cada  $z \in \mathbb{R}^{m+n}$  sob a forma  $z = (x_1, \dots, x_m, y_1, \dots, y_n)$ . Assim,  $\frac{\partial f}{\partial x}(z)$  é a restrição da transformação linear  $f'(z) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  ao subespaço  $\mathbb{R}^m \subset \mathbb{R}^{m+n}$  formado pelos vetores  $(x, 0)$  e  $\frac{\partial f}{\partial y}(z)$  é a restrição de  $f'(z)$  ao subespaço  $\mathbb{R}^n$  que consiste nos vetores da forma  $(0, y)$ . Para todo vetor

$$w = (u, v) \in \mathbb{R}^{m+n}, \quad \text{tem-se} \quad f'(z) \cdot w = \frac{\partial f}{\partial x}(z) \cdot u + \frac{\partial f}{\partial y}(z) \cdot v.$$

Usando estas derivadas parciais, a Regra da Cadeia nos permite concluir, a partir da identidade  $f(x, \xi(x)) = 0$  para todo  $x \in V$ , que

$$\frac{\partial f}{\partial x}(z) + \frac{\partial f}{\partial y}(z) \cdot \xi'(x) = 0, \quad \text{com} \quad z = (x, \xi(x)).$$

Logo

$$\xi'(x) = - \left[ \frac{\partial f}{\partial y}(z) \right]^{-1} \cdot \frac{\partial f}{\partial x}(z),$$

ainda com  $z = (x, \xi(x))$ . Note que a hipótese do Teorema das Funções Implícitas assegura que a transformação linear  $\frac{\partial f}{\partial y}(z) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  é invertível para todo  $z$  na vizinhança de  $p$ .

**Exemplo 4.** Diz-se que o número complexo  $c$  é uma *raiz simples* do polinômio  $p$  quando se tem  $p(z) = (z - c)q(z)$  com  $q(c) \neq 0$ . O Teorema 4 pode ser usado para mostrar que as raízes simples de um polinômio dependem diferenciavelmente dos coeficientes desse polinômio. A fim de provar isto escrevemos, para cada  $a = (a_0, \dots, a_n) \in \mathbb{C}^{n+1} = \mathbb{R}^{2n+2}$  e cada  $z \in \mathbb{C} = \mathbb{R}^2$ ,

$$p_a(z) = p(a_0, \dots, a_n, z) = a_0 + a_1 z + \dots + a_n z^n.$$

Então, de  $p_a(z) = (z - c)q(z)$  resulta  $\frac{\partial p}{\partial z}(c) = p'_a(z) = q(c)$ , logo a matriz jacobiana (real)  $2 \times 2$ ,  $\frac{\partial p}{\partial z}(c)$ , é invertível, por ser a matriz da transformação

linear de  $\mathbb{R}^2$  que consiste na multiplicação pelo número complexo não-nulo  $q(c)$ . Portanto, em virtude do Teorema 4, existem bolas abertas  $B = B(a; \varepsilon)$  em  $\mathbb{C}^{n+1}$  e  $B' = B(c; \delta)$  em  $\mathbb{C}$  tais que, para todo  $b \in B$ , o polinômio  $p_b$  possui uma única raiz  $\xi(b) \in B'$ , a qual é simples, e a aplicação  $\xi : B \rightarrow \mathbb{R}^2$ , assim definida, é de classe  $C^\infty$ .  $\triangleleft$

## Capítulo 7

# Superfícies Diferenciáveis

### 1 Parametrizações

Uma *imersão* do aberto  $U \subset \mathbb{R}^m$  no espaço  $\mathbb{R}^n$  é uma aplicação diferenciável  $f : U \rightarrow \mathbb{R}^n$  tal que, para todo  $x \in U$ , a derivada  $f'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é uma transformação linear injetiva. Isto, naturalmente, só pode ocorrer quando  $m \leq n$ .

Quando  $m = n$ , toda imersão de  $U \subset \mathbb{R}^m$  em  $\mathbb{R}^n$  é um difeomorfismo local. Em geral, para  $m \leq n$  quaisquer, o Teorema 2 do Capítulo 6 assegura que toda imersão é uma aplicação localmente injetiva.

**Exemplo 1.** Se  $I \subset \mathbb{R}$  é um intervalo aberto, as imersões  $f : I \rightarrow \mathbb{R}^n$  são o que chamamos no Capítulo 2 de caminhos regulares. Assim, por exemplo,  $f : \mathbb{R} \rightarrow \mathbb{R}^2$ , definida por  $f(t) = (t^3 - t, t^2)$  é uma imersão de  $\mathbb{R}$  no plano, a qual não é injetiva, pois  $f(-1) = f(1) = (0, 1)$ .

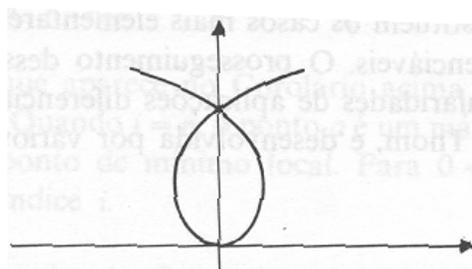


Figura 1.

Uma *parametrização* de classe  $C^k$  e dimensão  $m$  de um conjunto  $V \subset \mathbb{R}^n$  é uma imersão  $\varphi : V_o \rightarrow V$  de classe  $C^k$  que é, ao mesmo tempo, um homeomorfismo do aberto  $V_o \subset \mathbb{R}^m$  sobre  $V$ . ◁

**Exemplo 2.** Dada uma aplicação  $f : V_o \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $V_o \subset \mathbb{R}^m$ , seja  $V = \{(x, f(x)); x \in V_o\} \subset \mathbb{R}^{m+n}$  o gráfico de  $f$ . A aplicação  $\varphi : V_o \rightarrow V$ , dada por  $\varphi(x) = (x, f(x))$ , é uma parametrização de dimensão  $m$  e classe  $C^k$  do conjunto  $V \subset \mathbb{R}^{m+n}$ . Com efeito, se chamarmos de  $\Pi : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^m$  a projeção sobre as  $m$  primeiras coordenadas, a igualdade  $\Pi \circ \varphi = \text{id}_{V_o}$  mostra que  $\varphi$  é um homeomorfismo, cujo inverso é a restrição  $\Pi|_V$  e, em virtude da Regra da Cadeia, que  $\Pi \cdot \varphi'(x) = \text{id}_{\mathbb{R}^m}$ , logo  $\varphi'(x) : \mathbb{R}^m \rightarrow \mathbb{R}^{m+n}$  é injetiva, para todo  $x \in V_o$ , portanto  $\varphi$  é uma imersão.  $\triangleleft$

**Exemplo 3.** Uma imersão  $\varphi : V_o \rightarrow V$  pode muito bem ser bijetiva sem ser um homeomorfismo, logo não é uma parametrização de  $V$ . Um exemplo disso pode ser obtido tomando a restrição do caminho  $f$ , visto no Exemplo 1 acima, ao intervalo  $(-1, +\infty) \subset \mathbb{R}$ . O caminho  $\varphi : (-1, +\infty) \rightarrow \mathbb{R}^2$ , dado por  $\varphi(t) = (t^3 - t, t^2)$ , é uma imersão  $C^\infty$  bijetiva do intervalo  $(-1, +\infty)$  em  $\mathbb{R}^2$  mas não é uma parametrização da sua imagem  $V$  pois a função inversa  $\varphi^{-1} : V \rightarrow (-1, +\infty)$  é descontínua no ponto  $(0, 1) \in V$ . Com efeito, se  $(t_n)$  é uma seqüência decrescente de números reais com  $\lim t_n = -1$ , vemos que  $\lim \varphi(t_n) = (0, 1) = \varphi(1)$  sem que se tenha  $\lim t_n = 1$ .  $\triangleleft$

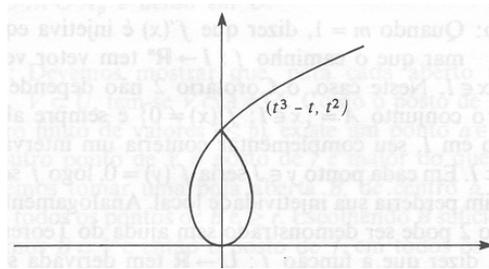


Figura 2.

**Exemplo 4.** Seja  $N = (0, \dots, 0, 1)$  o polo norte da esfera unitária  $S^n = \{x \in \mathbb{R}^{n+1}; \langle x, x \rangle = 1\}$ . Pondo  $V = S^n - \{N\}$  e  $V_o = \mathbb{R}^n$ , o homeomorfismo  $\varphi : V_o \rightarrow V$ , inverso da projeção estereográfica  $\xi$ , (vide Exemplo 16, Capítulo 1), é uma parametrização. Evidentemente,  $\varphi$  é de classe  $C^\infty$  e sua inversa  $\xi : S^n - \{N\} \rightarrow \mathbb{R}^n$  é a restrição de uma aplicação  $C^\infty$  (cujo domínio é o aberto  $U = \{x \in \mathbb{R}^{n+1}; x_{n+1} \neq 1\}$ ). A igualdade  $\xi \circ \varphi = \text{id}_{\mathbb{R}^n}$  mostra, via Regra da Cadeia, que  $\varphi$  é uma imersão, o que completa a verificação.  $\triangleleft$

## 2 Superfícies diferenciáveis

Um conjunto  $M \subset \mathbb{R}^n$  chama-se uma *superfície* de dimensão  $m$  e classe  $C^k$  quando todo ponto  $p \in M$  está contido em algum aberto  $U \subset \mathbb{R}^n$  tal que  $V = U \cap M$  é a imagem de uma parametrização  $\varphi : V_o \rightarrow V$ , de dimensão  $m$  e classe  $C^k$ . O conjunto  $V$  é um aberto em  $M$ , chamado uma *vizinhança parametrizada* do ponto  $p$ . Escreve-se  $m = \dim \cdot M$ .

**Observação.** Na definição acima, supõe-se tacitamente  $k \geq 1$ . Mas teria sentido considerar superfícies de classe  $C^0$ . Bastaria admitir “parametrizações de classe  $C^0$ ”, que são meramente homeomorfismos  $\varphi : V_o \rightarrow V$  de abertos  $V_o \subset \mathbb{R}^m$  sobre abertos  $V \subset M$ . As superfícies de classe  $C^0$  são estudadas na Topologia. Seu interesse em Análise é reduzido, principalmente porque não possuem espaços tangentes.

Quando  $\dim \cdot M = 1$ , a superfície  $M$  chama-se uma *curva*.

**Exemplo 5.** Como  $\mathbb{R}^0 = \{0\}$  reduz-se a um ponto, uma superfície de dimensão 0 em  $\mathbb{R}^n$  é simplesmente um conjunto discreto. No extremo oposto, as superfícies de dimensão  $n$  em  $\mathbb{R}^n$  são os subconjuntos abertos, pois a imagem de uma parametrização de dimensão  $n$  em  $\mathbb{R}^n$  é aberta, em virtude do Teorema da Aplicação Inversa.  $\triangleleft$

**Exemplo 6.** A esfera  $S^n$  é uma superfície de dimensão  $n$  e classe  $C^\infty$  em  $\mathbb{R}^{n+1}$ . Com efeito, a inversa da projeção estereográfica é uma parametrização  $\varphi : \mathbb{R}^n \rightarrow S^n - \{N\}$ . Para obter uma vizinhança parametrizada do polo norte  $N$ , basta considerar  $-\varphi : \mathbb{R}^n \rightarrow S^n - \{N^*\}$ , onde  $N^* = -N$  é o polo sul.  $\triangleleft$

**Exemplo 7.** O produto cartesiano  $M \times N$  de duas superfícies  $M \subset \mathbb{R}^n$  e  $N \subset \mathbb{R}^k$  é uma superfície em  $\mathbb{R}^{n+k}$  pois se  $\varphi : V_o \rightarrow V \subset M$  e  $\psi : W_o \rightarrow W \subset N$  são parametrizações então  $\xi : V_o \times W_o \rightarrow V \times W \subset M \times N$ , dada por  $\xi(x, y) = (\varphi(x), \psi(y))$ , é uma parametrização. Evidentemente,  $\dim(M \times N) = \dim M + \dim N$ . Em particular, o *toro  $m$ -dimensional*  $T^m = S^1 \times \cdots \times S^1$ , produto cartesiano de  $m$  círculos, é uma superfície de dimensão  $m$  e classe  $C^\infty$  em  $\mathbb{R}^{2m}$ .  $\triangleleft$

**Exemplo 8.** O gráfico de uma aplicação  $f : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^m$ , é uma superfície  $M = \{(x, f(x)) \in \mathbb{R}^{m+n}; x \in U\}$ , de dimensão  $m$  e classe  $C^k$  em  $\mathbb{R}^{m+n}$ . Com efeito,  $M$  é a imagem da única parametrização  $\varphi : U \rightarrow M$ ,  $\varphi(x) = (x, f(x))$ .  $\triangleleft$

Ser uma superfície é uma propriedade local: se todo ponto  $p \in M$  está contido num conjunto  $V \subset M$ , aberto em  $M$ , o qual é uma superfície de classe  $C^k$  e dimensão  $m$ , então o conjunto  $M \subset \mathbb{R}^n$  é uma superfície de dimensão  $m$  e classe  $C^k$ .

Em particular, se  $M$  é localmente o gráfico de uma aplicação  $f : V_o \rightarrow \mathbb{R}^n$ , de classe  $C^k$  num aberto  $V_o \subset \mathbb{R}^m$ , então  $M \subset \mathbb{R}^{m+n}$  é uma superfície de classe  $C^k$  e dimensão  $m$ . Assim, por exemplo, as hiperfícies, conforme definidas no Capítulo 4, são superfícies de dimensão  $n - 1$  em  $\mathbb{R}^n$ .

Quando  $M \subset \mathbb{R}^n$  é uma superfície de dimensão  $m$ , costuma-se dizer que  $M$  tem *co-dimensão*  $n - m$ . Portanto, hiperfícies são superfícies de co-dimensão 1.

No teorema abaixo,  $M$  é uma superfície de dimensão  $m$  e classe  $C^k$  em  $\mathbb{R}^n$ . Por “uma projeção  $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ” entendemos a aplicação dada por  $\Pi(x_1, \dots, x_n) = (x_{i_1}, \dots, x_{i_m})$ , definida a partir da escolha de  $m$  índices  $i_1 < \dots < i_m$ , compreendidos entre 1 e  $n$ .

**Teorema 1.** *Seja  $\varphi : V_o \rightarrow V$  uma parametrização em  $M$ . Para cada  $p = \varphi(x_o) \in V$  existe uma projeção  $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$  tal que  $\Pi \circ \varphi$  aplica um aberto  $Z_o$ , com  $x_o \in Z_o \subset V$ , difeomorficamente sobre um aberto  $W_o \subset \mathbb{R}^m$ .*

*Demonstração.* A matriz jacobiana  $\left[ \frac{\partial \varphi_i}{\partial x_j}(x_o) \right] \in M(n \times m)$  tem  $m$  linhas linearmente independentes, de índices  $i_1 < i_2 < \dots < i_m$ . Essas linhas formam a matriz  $m \times m$  invertível  $J = \left[ \frac{\partial \varphi_{i_k}}{\partial x_j}(x_o) \right]$  e os índices  $i_k$  definem uma projeção  $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Observando que  $J$  é a matriz jacobiana da aplicação  $\Pi \circ \varphi : V_o \rightarrow \mathbb{R}^m$ , o Teorema 1 resulta imediatamente do Teorema da Aplicação Inversa.  $\square$

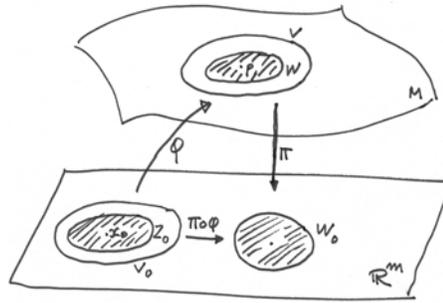


Figura 3.

**Corolário 1.** *Toda superfície de classe  $C^k$  é localmente o gráfico de uma aplicação de classe  $C^k$ .*

Com efeito, usando a notação do Teorema 1, escrevamos os elementos de  $\mathbb{R}^n$  sob a forma  $z = (y, y')$ , onde  $y = \Pi(z)$ . Ponhamos também  $W = \varphi(Z_o)$ . Então

a aplicação  $\psi = \varphi \circ (\Pi \circ \varphi)^{-1} : W_o \rightarrow W$  é uma parametrização. Além disso, para todo  $y \in W_o$ , tem-se

$$\Pi(\psi(y)) = (\Pi \circ \varphi) \circ (\Pi \circ \varphi)^{-1}(y) = y, \quad \text{logo} \quad \psi(y) = (y, y').$$

Assim,  $W$  é o gráfico da aplicação de classe  $C^k$ ,  $f : W_o \rightarrow \mathbb{R}^{n-m}$ , dada por  $f(y) = y'$ .  $\square$

**Corolário 2.** *Seja  $M \subset \mathbb{R}^n$  uma superfície de classe  $C^k$  e dimensão  $m$ . Se uma aplicação  $f : V_o \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $V_o \subset \mathbb{R}^p$ , tiver sua imagem  $f(V_o)$  contida na vizinhança  $W \subset M$ , parametrizada por  $\psi : W_o \rightarrow W$ , então  $\psi^{-1} \circ f : V_o \rightarrow \mathbb{R}^m$  é uma aplicação de classe  $C^k$ .*

Com efeito, para cada ponto  $x_o \in V_o$ , com  $f(x_o) = \psi(y_o)$ , existe, pelo Teorema 1, uma projeção  $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^m$  tal que  $\Pi \circ \psi$  é um difeomorfismo de uma vizinhança de  $y_o$  sobre um aberto de  $\mathbb{R}^m$ . Então, numa vizinhança de  $x_o$ , podemos escrever

$$\psi^{-1} \circ f = (\Pi \circ \psi)^{-1} \circ \Pi \circ f,$$

logo  $\psi^{-1} \circ f$  é de classe  $C^k$ .  $\square$

Sejam  $\varphi : V_o \rightarrow V$  e  $\psi : W_o \rightarrow W$  parametrizações numa superfície  $M$ , de classe  $C^k$  e dimensão  $m$ . Suponhamos que  $V \cap W \neq \emptyset$ . Então todo ponto  $p \in V \cap W$  pode escrever-se como  $p = \varphi(x)$ ,  $x \in V_o$ , ou como  $p = \psi(y)$ ,  $y \in W_o$ , isto é, pode ser representado pelos  $m$  parâmetros que são as coordenadas de  $x$  ou pelas  $m$  coordenadas de  $y$ . A correspondência  $x \mapsto y$ , definida pela relação  $\varphi(x) = \psi(y)$ , é a aplicação

$$\psi^{-1} \circ \varphi : \varphi^{-1}(V \cap W) \longrightarrow \psi^{-1}(V \cap W),$$

chamada *mudança de parametrização*.

**Corolário 3.** *Numa superfície de classe  $C^k$ , toda mudança de parametrização  $\psi^{-1} \circ \varphi$  é um difeomorfismo de classe  $C^k$ .*

Com efeito, pelo Corolário 2,  $\psi^{-1} \circ \varphi$  é uma aplicação de classe  $C^k$ . Pelo mesmo motivo, sua inversa  $\varphi^{-1} \circ \psi$  também é de classe  $C^k$ . Logo  $\psi^{-1} \circ \varphi$  é um difeomorfismo.  $\square$

**Exemplo 9.** O conjunto  $M = \{(x, x^{4/3}); x \in \mathbb{R}\}$ , gráfico da função  $f : \mathbb{R} \rightarrow \mathbb{R}$ ,  $f(x) = x^{4/3}$  é uma curva de classe  $C^1$  em  $\mathbb{R}^2$ : a aplicação  $\varphi : \mathbb{R} \rightarrow \mathbb{R}^2$ , dada por  $\varphi(x) = (x, x^{4/3})$ , é uma parametrização (global) de  $M$ . Cabe observar, porém, que se  $V \subset M$  contém o ponto  $(0, 0)$ , não pode existir uma parametrização

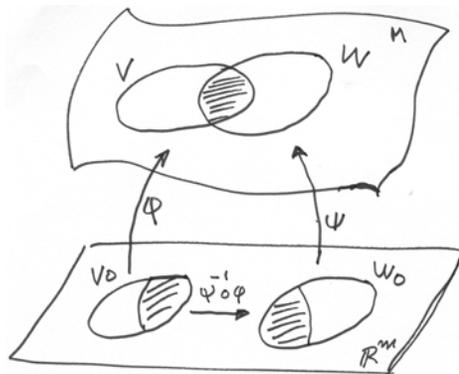


Figura 4.

$\psi : V_o \rightarrow V$  de classe  $C^k$  com  $k > 1$ . Com efeito, se uma tal  $\psi$  existisse então o próprio conjunto  $V$  seria uma curva de classe  $C^k$  logo, pelo Corolário 1, uma vizinhança  $W$  do ponto  $(0, 0)$ , com  $W \subset V$ , seria o gráfico de uma função  $g : W_o \rightarrow \mathbb{R}$ , de classe  $C^k$ . Neste caso, para todo  $x \in W_o$  teríamos  $(x, g(x)) \in W \subset M$ , logo  $g(x) = x^{4/3}$ , mas  $x^{4/3}$  é apenas de classe  $C^1$ . Assim,  $M$  não é uma curva de classe  $C^2$ .  $\triangleleft$

### 3 O espaço vetorial tangente

Seja  $p$  um ponto da superfície  $M$ , de dimensão  $m$  e classe  $C^k$  em  $\mathbb{R}^n$ . O *espaço vetorial tangente* a  $M$  no ponto  $p$  é um subespaço vetorial  $T_p M \subset \mathbb{R}^n$ , que pode ser visto sob dois aspectos:

- 1)  $T_p M$  é o conjunto dos vetores-velocidade  $v = \lambda'(0)$  dos caminhos diferenciáveis  $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$ , tais que  $\lambda(0) = p$ .
- 2)  $T_p M = \varphi'(x_o) \cdot \mathbb{R}^m$  é a imagem da derivada  $\varphi'(x_o) : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , onde  $\varphi : V_o \rightarrow V$  é uma parametrização em  $M$ , com  $\varphi(x_o) = p$ .

A primeira descrição de  $T_p M$  é intrínseca (não depende de escolhas arbitrárias) mas não deixa claro que se trata de um subespaço vetorial de  $\mathbb{R}^n$ . Pela segunda descrição,  $T_p M$  é obviamente um subespaço vetorial de  $\mathbb{R}^n$  mas não é evidente que para outra parametrização  $\psi : W_o \rightarrow W$ , com  $\psi(y_o) = p$ , se tenha  $\psi'(y_o) \cdot \mathbb{R}^m = \varphi'(x_o) \cdot \mathbb{R}^m$ .

As dúvidas ficarão sanadas se mostrarmos que os conjuntos definidos em 1) e 2) são o mesmo. Para ver isto, comecemos com o vetor-velocidade  $v = \lambda'(0)$  de

um caminho diferenciável  $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$ , com  $\lambda(0) = p$ . Restringindo  $\varepsilon$ , se necessário, podemos admitir que a imagem de  $\lambda$  esteja contida na imagem  $V$  de uma parametrização  $\varphi : V_o \rightarrow V \subset M$ , com  $\varphi(x_o) = p$ . Então, pelo Corolário 2,  $\mu = \varphi^{-1} \circ \lambda : (-\varepsilon, \varepsilon) \rightarrow V_o$  é um caminho diferenciável em  $\mathbb{R}^m$ , com  $\mu(0) = x_o$ . Pondo  $u = \mu'(0)$ , temos

$$\varphi'(x_o) \cdot u = \varphi'(x_o) \cdot (\varphi^{-1} \circ \lambda)'(0) = (\varphi \circ \varphi^{-1} \circ \lambda)'(0) = \lambda'(0) = v.$$

Portanto todo vetor  $v = \lambda'(0)$  pertence à imagem  $\varphi'(x_o) \cdot \mathbb{R}^m$  de  $\mathbb{R}^m$  pela derivada de alguma parametrização  $\varphi : V_o \rightarrow V$ , com  $p \in V$ .

Reciprocamente, se  $v = \varphi'(x_o) \cdot u$  então, como  $u = \mu'(0)$ , onde  $\mu : (-\varepsilon, \varepsilon) \rightarrow V_o$  é dado por  $\mu(t) = x_o + t \cdot u$ , temos  $v = \lambda'(0)$  com  $\lambda : (-\varepsilon, \varepsilon) \rightarrow V$ ,  $\lambda(t) = \varphi(\mu(t))$ , logo  $v$  está no conjunto definido em 2).

Como toda parametrização  $\varphi$  é uma imersão, a derivada  $\varphi'(x_o) : \mathbb{R}^m \rightarrow \mathbb{R}^n$  é uma transformação linear injetiva, logo sua imagem  $\varphi'(x_o) \cdot \mathbb{R}^m = T_p M$  é um subespaço vetorial  $m$ -dimensional de  $\mathbb{R}^n$ .

Os vetores

$$\frac{\partial \varphi}{\partial x_1}(x_o) = \varphi'(x_o) \cdot e_1, \dots, \frac{\partial \varphi}{\partial x_m}(x_o) = \varphi'(x_o) \cdot e_m$$

formam uma base de  $T_p M$ , chamada a *base associada à parametrização*  $\varphi$ .

A seguir estenderemos, para superfícies quaisquer, o Teorema 3 do Capítulo 4, provado para o caso de co-dimensão 1.

Seja  $f : U \rightarrow \mathbb{R}^n$  uma aplicação diferenciável, definida no aberto  $U \subset \mathbb{R}^{m+n}$ . Um ponto  $c \in \mathbb{R}^n$  chama-se um *valor regular* de  $f$  quando, para todo  $x \in U$  tal que  $f(x) = c$ , a derivada  $f'(x) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  é uma transformação linear sobrejetiva.

Observe-se que, para  $n = 1$ , a transformação linear  $f'(x) : \mathbb{R}^{m+1} \rightarrow \mathbb{R}$  é sobrejetiva se, e somente se, é diferente de zero, ou seja,  $\text{grad } f(x) \neq 0$ . (Vide Exemplo 2, Capítulo 5.) Portanto esta definição de valor regular estende a que foi dada anteriormente.

**Teorema 2.** *Seja  $c \in \mathbb{R}^n$  um valor regular da aplicação  $f : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$ . A imagem inversa  $M = f^{-1}(c) = \{x \in U; f(x) = c\}$  é uma superfície de classe  $C^k$  e dimensão  $m$  em  $\mathbb{R}^{m+n}$ . O espaço vetorial tangente  $T_p M$ , em cada ponto  $p \in M$ , é o núcleo da derivada  $f'(p) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$ .*

*Demonstração.* Pelo Corolário 4, Capítulo 6,  $M = f^{-1}(c)$  é localmente o gráfico de uma aplicação de classe  $C^k$ , logo é uma superfície. Além disso, para  $p \in M$ , todo vetor  $v \in T_p M$  é da forma  $v = \lambda'(0)$ , onde  $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$  é um caminho diferenciável, com  $\lambda(0) = p$ . Logo  $f'(p) \cdot v = (f \circ \lambda)'(0) = 0$  pois

$f \circ \lambda : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$  é constante, igual a  $c$ . Portanto  $T_p M$  está contido no núcleo de  $f'(p)$ . Como  $f'(p)$  é sobrejetiva, esse núcleo tem dimensão  $m$  e então é igual a  $T_p M$ .  $\square$

**Exemplo 10.** Seja  $O(\mathbb{R}^n)$  o grupo ortogonal, formado pelas matrizes  $\mathbf{x} \in M(n \times n)$ , tais que  $\mathbf{x}\mathbf{x}^T = \mathbf{I}_n$  (matrizes ortogonais). Usaremos o Teorema 2 para mostrar que  $O(\mathbb{R}^n)$  é uma superfície (compacta) de classe  $C^\infty$  e dimensão  $n(n-1)/2$  em  $\mathbb{R}^{n^2}$ . Seja então  $f : M(n \times n) \rightarrow S(\mathbb{R}^n)$  a aplicação definida no conjunto das matrizes  $n \times n$ , com valores no conjunto das matrizes simétricas  $n \times n$ , pela fórmula  $f(\mathbf{x}) = \mathbf{x} \cdot \mathbf{x}^T$ . Já costumamos fazer a identificação  $M(n \times n) = \mathbb{R}^{n^2}$ . Agora identificaremos  $S(\mathbb{R}^n)$  com  $\mathbb{R}^{n(n+1)/2}$  pois uma matriz simétrica  $n \times n$  fica determinada pelos seus elementos da diagonal e acima dela, em número de  $n + (n-1) + \dots + 2 + 1 = n(n+1)/2$ . Assim, escrevemos  $f : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n(n+1)/2}$  e temos  $O(\mathbb{R}^n) = f^{-1}(\mathbf{I}_n)$ . Resta apenas verificar que a matriz identidade  $\mathbf{I}_n$  é um valor regular de  $f$ . Tomando um ponto arbitrário de  $f^{-1}(\mathbf{I}_n)$ , isto é, uma matriz ortogonal  $\mathbf{x}$ , sabemos que a derivada  $f'(\mathbf{x}) : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n(n+1)/2}$  é a transformação linear que toda  $\mathbf{v} \in \mathbb{R}^{n^2}$  faz corresponder  $f'(\mathbf{x}) \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{x}^T + \mathbf{x} \cdot \mathbf{v}^T$ . Para provar que  $f'(\mathbf{x})$  é sobrejetiva, seja dada  $\mathbf{s} \in \mathbb{R}^{n(n+1)/2}$ . Tomando  $\mathbf{v} = \mathbf{s}\mathbf{x}/2$  temos  $f'(\mathbf{x}) \cdot \mathbf{v} = \mathbf{s}\mathbf{x} \cdot \mathbf{x}^T/2 + \mathbf{x}\mathbf{x}^T \cdot \mathbf{s}/2 = \mathbf{s}/2 + \mathbf{s}/2 = \mathbf{s}$ . (Lembre que  $\mathbf{s}^T = \mathbf{s}$ .) Vale  $\dim O(\mathbb{R}^n) = n^2 - n(n+1)/2 = n(n-1)/2$ .  $\triangleleft$

Observemos, em relação ao Exemplo 10, que o espaço vetorial tangente a  $O(\mathbb{R}^n)$  no ponto  $\mathbf{I}_n$  é o conjunto das matrizes anti-simétricas  $n \times n$ , isto é, matrizes  $\mathbf{v}$  tais que  $\mathbf{v} + \mathbf{v}^T = 0$ . Com efeito, sendo a derivada  $f'(\mathbf{I}_n) : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n(n+1)/2}$  dada por  $\mathbf{v} \mapsto \mathbf{v} \cdot \mathbf{I}_n^T + \mathbf{I}_n \cdot \mathbf{v}^T = \mathbf{v} + \mathbf{v}^T$ , vemos que o núcleo de  $f'(\mathbf{I}_n)$ , ou seja, o espaço vetorial tangente a  $O(\mathbb{R}^n)$  no ponto  $\mathbf{I}_n$ , é o conjunto das matrizes anti-simétricas.

**Exemplo 11 (Ainda uma vez a esfera).** Seja  $f : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  definida por  $f(x) = \langle x, x \rangle$ . Como  $f'(x) \cdot v = 2\langle v, x \rangle$ , vemos que, para todo  $x \neq 0$  em  $\mathbb{R}^{n+1}$ , a derivada  $f'(x) : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$  é  $\neq 0$ , logo é sobrejetiva. Como  $f(x) = 0 \Leftrightarrow x = 0$ , conclui-se que todo  $c \neq 0$  em  $\mathbb{R}$  é valor regular de  $f$ . Se  $c < 0$  então  $f^{-1}(c) = \emptyset$ . Se  $c > 0$  então  $f^{-1}(c)$  é a esfera de centro  $O$  e raio  $\sqrt{c}$ .  $\triangleleft$

#### 4 Superfícies orientáveis

Como no caso de hiperfícies (co-dimensão 1), tratado no Capítulo 4, cabe observar que nem toda superfície em  $\mathbb{R}^n$  pode ser obtida como imagem inversa de um valor regular.

Com efeito, se  $M = f^{-1}(c)$  é a imagem inversa do valor regular  $c \in \mathbb{R}^n$  pela aplicação  $f : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$  então, chamando de  $f_1, \dots, f_n : U \rightarrow \mathbb{R}$  as funções-coordenada de  $f$ , vemos que

$\text{grad } f_1, \dots, \text{grad } f_n : U \rightarrow \mathbb{R}^{m+n}$  são campos de vetores de classe  $C^{k-1}$ , com as seguintes propriedades:

- 1) Para todo  $x \in M$ , os vetores  $\text{grad } f_1(x), \dots, \text{grad } f_n(x)$  são ortogonais ao espaço vetorial tangente  $T_x M$ . (Diz-se então que os  $\text{grad } f_i$  são campos de vetores *normais* a  $M$ .)
- 2) Para todo  $x \in M$ , os vetores  $\text{grad } f_1(x), \dots, \text{grad } f_n(x)$  são linearmente independentes.

A afirmação 1) resulta do fato de que, para cada  $i = 1, \dots, n$ , a função  $f_i : U \rightarrow \mathbb{R}$  é constante ao longo de  $M$ . Todo vetor  $v \in T_x M$ , para  $x \in M$  qualquer, é o vetor-velocidade  $v = \lambda'(0)$  de um caminho  $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$ , logo  $f_i \circ \lambda : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$  é constante. Daí,  $\langle \text{grad } f_i(x), v \rangle = (f_i \circ \lambda)'(0) = 0$ .

Por sua vez, a afirmação 2) é equivalente a dizer que  $c$  é um valor regular de  $f$ , pois  $\text{grad } f_1(x), \dots, \text{grad } f_n(x)$  são os vetores-linha da matriz jacobiana  $\left[ \frac{\partial f_i}{\partial x_j}(x) \right] \in M(n \times (m+n))$ . Sua independência linear significa que esta matriz, para todo  $x \in M$ , tem posto  $n$ , logo  $f'(x) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$  é sobrejetiva.

Mas nem toda superfície  $M \subset \mathbb{R}^{m+n}$ , de co-dimensão  $n$ , admite  $n$  campos contínuos linearmente independente de vetores normais. Uma condição necessária para isto é que  $M$  seja orientável, conforme mostraremos agora.

Um *atlas* numa superfície  $M$  é um conjunto de parametrizações  $\varphi : V_o \rightarrow V$  cujas imagens  $V$  cobrem  $M$ . Duas parametrizações  $\varphi : V_o \rightarrow V$  e  $\psi : W_o \rightarrow W$  dizem-se *compatíveis* quando  $V \cap W = \emptyset$  ou quando  $V \cap W \neq \emptyset$  e  $\psi^{-1} \circ \varphi : \varphi^{-1}(V \cap W) \rightarrow \psi^{-1}(V \cap W)$  tem determinante jacobiano positivo em todos os pontos  $x \in \varphi^{-1}(V \cap W)$ . Um atlas  $\mathcal{A}$  na superfície  $M$  chama-se *coerente* quando duas parametrizações quaisquer  $\varphi, \psi \in \mathcal{A}$  são compatíveis. Uma superfície  $M$  chama-se *orientável* quando admite um atlas coerente.

**Teorema 3.** *Se uma superfície  $M \subset \mathbb{R}^{m+n}$ , de co-dimensão  $n$ , admite  $n$  campos contínuos linearmente independentes de vetores normais  $v_1, \dots, v_n : M \rightarrow \mathbb{R}^{m+n}$  então  $M$  é orientável.*

*Demonstração.* Seja  $\mathcal{A}$  o conjunto das parametrizações  $\varphi : V_o \rightarrow V$  em  $M$  tais que  $V_o$  é conexo e, para todo  $x \in V_o$ , a matriz

$$\Phi(x) = \left[ \frac{\partial \varphi}{\partial x_1}(x), \dots, \frac{\partial \varphi}{\partial x_m}(x), v_1(\varphi(x)), \dots, v_n(\varphi(x)) \right],$$

cujas  $m+n$  colunas são os vetores de  $\mathbb{R}^{m+n}$  aí indicados, tem determinante positivo. Como  $V_o$  é conexo e os campos  $v_i$  são contínuos, para ser  $\varphi \in \mathcal{A}$  basta que

det  $\Phi(x) > 0$  para algum  $x \in V_o$ . Se for det  $\Phi(x) < 0$ , escrevemos  $x^* = (-x_1, x_2, \dots, x_m)$  quando  $x = (x_1, x_2, \dots, x_m)$  e pomos  $V_o^* = \{x^*; x \in V_o\}$ . Então  $\varphi^* : V_o^* \rightarrow V$ , dada por  $\varphi^*(x) = \varphi(x^*)$ , é uma parametrização cuja imagem ainda é  $V$  mas det  $\Phi^*(x) > 0$ . Isto mostra que  $\mathcal{A}$  é um atlas em  $M$ . Sejam  $\varphi, \phi \in \mathcal{A}$ , com  $\varphi : V_o \rightarrow V, \phi : W_o \rightarrow W$  e  $V \cap W \neq \emptyset$ . Pondo  $\xi = \psi^{-1} \circ \varphi : \varphi^{-1}(V \cap W) \rightarrow \psi^{-1}(V \cap W)$ , temos  $\phi = \psi \circ \xi$ . A Regra da Cadeia nos dá, para  $x \in \varphi^{-1}(V \cap W)$  e  $y = \xi(x)$ :

$$\frac{\partial \phi}{\partial x_j}(x) = \sum_{i=1}^m a_{ij}(x) \frac{\partial \psi}{\partial y_i}(y), \quad j = 1, \dots, m,$$

onde  $\mathbf{a}(x) = [a_{ij}(x)]$  é a matriz jacobiana de  $\xi$  no ponto  $x$ . Em termos matriciais, estas igualdades significam que

$$\Phi(x) = \Psi(x) \cdot \mathbf{A}(x), \quad \text{onde} \quad \mathbf{A}(x) = \begin{bmatrix} \mathbf{a}(x) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \in M((m+n) \times (m+n))$$

e  $\mathbf{I}$  = matriz identidade  $n \times n$ . Então det  $\Phi(x) = \det \Psi(x) \cdot \det \mathbf{a}(x)$  e daí det  $\mathbf{a}(x) > 0$ . Logo as parametrizações  $\varphi, \psi \in \mathcal{A}$  são compatíveis. O atlas  $\mathcal{A}$  é coerente e a superfície  $M$  é orientável.  $\square$

**Corolário 4.** *Se  $M = f^{-1}(c)$  é a imagem inversa de um valor regular da aplicação  $f : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$ , então  $M$  é uma superfície  $m$ -dimensional orientável.*

Assim, por exemplo, o grupo ortogonal  $O(\mathbb{R}^n)$  é uma superfície orientável.

Para co-dimensão 1, vale a recíproca do Teorema 3. Ela resulta da existência do *produto vetorial*  $w = v_1 \times \dots \times v_n$  de  $n$  vetores em  $\mathbb{R}^{n+1}$ , que descreveremos agora.

O produto  $w = v_1 \times \dots \times v_n$  é igual a zero quando  $v_1, \dots, v_n$  são linearmente dependentes. Caso contrário,  $w$  é ortogonal ao subespaço gerado por esses  $n$  vetores, tem comprimento igual ao volume do paralelepípedo  $n$ -dimensional por eles determinado e seu sentido é dado pela condição det  $[v_1, \dots, v_n, w] > 0$ .

Em termos formais, seja  $\mathbf{m} = [v_1, \dots, v_n]$  a matriz  $n \times (n+1)$  cujas colunas são os vetores dados. Para cada  $i = 1, \dots, n$ , indiquemos com  $\mathbf{m}_i$  a matriz  $n \times n$  obtida de  $\mathbf{m}$  omitindo a  $i$ -ésima coluna. Então o produto vetorial  $w = v_1 \times \dots \times v_n$  é definido por

$$w = v_1 \times \dots \times v_n = \sum_{i=1}^n (-1)^{i+1} \det \mathbf{m}_i \cdot e_i.$$

O desenvolvimento de Laplace de um determinante em relação à sua última coluna mostra que, para todo vetor  $z \in \mathbb{R}^{n+1}$ , tem-se

$$\langle v_1 \times \dots \times v_n, z \rangle = \det [v_1, \dots, v_n, z].$$

Esta última igualdade mostra que, de fato,  $v_1 \times \dots \times v_n = w$  é ortogonal a  $v_1, \dots, v_n$ , que é zero quando esses vetores são linearmente dependentes e que  $\det [v_1, \dots, v_n, w] > 0$  no caso contrário. Além disso, sabe-se que o volume  $(n+1)$ -dimensional do paralelepípedo cujas arestas são  $v_1, \dots, v_n, w$  é o produto do volume  $n$ -dimensional  $V_n$  de sua base (a qual tem  $v_1, \dots, v_n$  como arestas) pelo comprimento de sua altura, que é  $|w|$ , pois  $w$  é ortogonal a essa base. Logo

$$\begin{aligned} |w| \cdot V_n &= \text{vol} [v_1, \dots, v_n, w] = |\det [v_1, \dots, v_n, w]| \\ &= \langle v_1 \times \dots \times v_n, w \rangle = |w|^2. \end{aligned}$$

Simplificando, vem  $|w| = V_n$ , ou seja, o comprimento do produto vetorial  $v_1 \times \dots \times v_n$  é o volume  $n$ -dimensional do paralelepípedo cujas arestas são os vetores  $v_1, \dots, v_n$ .

Concluindo estas considerações sobre o produto vetorial, mostraremos agora que se  $\{u_1, \dots, u_n\}$  e  $\{v_1, \dots, v_n\}$  são bases do subespaço vetorial  $E \subset \mathbb{R}^{n+1}$  e se  $\mathbf{a} = [a_{ij}]$  é a matriz de passagem da primeira para a segunda, isto é,  $v_j = \sum_{i=1}^n a_{ij} u_i$  ( $j = 1, \dots, n$ ), então

$$v_1 \times \dots \times v_n = \det \mathbf{a} \cdot u_1 \times \dots \times u_n.$$

Com efeito, como ambos estes produtos vetoriais são ortogonais ao subespaço  $E \subset \mathbb{R}^{n+1}$ , que tem co-dimensão 1, eles são múltiplos um do outro. Então, fixando os vetores  $u_1, \dots, u_n$ , definimos duas formas  $n$ -lineares alternadas  $f, \Delta$  em  $E$ , pelas condições

$$\begin{aligned} v_1 \times \dots \times v_n &= f(v_1, \dots, v_n) \cdot u_1 \times \dots \times u_n \\ \text{e } \Delta(v_1, \dots, v_n) &= \det [a_{ij}] \quad \text{se } v_j = \sum_{i=1}^n a_{ij} u_i, \quad j = 1, \dots, n. \end{aligned}$$

Sabe-se (v. “Álgebra Linear”, pág. 261) que as formas  $n$ -lineares alternadas num espaço de dimensão  $n$  constituem um espaço vetorial de dimensão 1. Logo existe  $c \in \mathbb{R}$  tal que  $f = c \cdot \Delta$ , ou seja,  $f(v_1, \dots, v_n) = c \cdot \det [a_{ij}]$  para quaisquer  $v_1, \dots, v_n \in E$ . Tomando  $v_1 = u_1, \dots, v_n = u_n$ , temos  $f(u_1, \dots, u_n) = 1$  e  $\Delta(u_1, \dots, u_n) = 1$ , logo  $c = 1$  e daí  $f = \Delta$ , isto é,

$$\begin{aligned} v_1 \times \dots \times v_n &= \det \mathbf{a} \cdot u_1 \times \dots \times u_n, \\ \text{onde } \mathbf{a} &= [a_{ij}] \quad \text{e } v_j = \sum_{i=1}^n a_{ij} u_i, \quad j = 1, \dots, n. \end{aligned}$$

**Teorema 4.** *Toda superfície orientável de co-dimensão 1 admite um campo contínuo de vetores normais não-nulos.*

*Demonstração.* Seja  $M \subset \mathbb{R}^{n+1}$  orientável de dimensão  $n$ . Para toda parametrização  $\varphi : V_o \rightarrow V$  pertencente ao atlas coerente  $\mathcal{A}$ , o qual caracteriza a orientabilidade de  $M$ , definamos o campo contínuo de vetores normais unitários  $u : V \rightarrow \mathbb{R}^{n+1}$  pondo, em cada ponto  $p \in V$ ,

$$w(p) = \frac{\partial \varphi}{\partial x_1}(x) \times \dots \times \frac{\partial \varphi}{\partial x_n}(x), \quad x = \varphi^{-1}(p) \in V_o, \quad \text{e } u(p) = w(p)/|w(p)|.$$

Se  $\psi : W_o \rightarrow W$  for outra parametrização pertencente a  $\mathcal{A}$  então  $\varphi$  e  $\psi$  são compatíveis. Assim, se  $V \cap W \neq \emptyset$ , para todo  $p \in V \cap W$ , com

$$z(p) = \frac{\partial \psi}{\partial y_1}(y) \times \dots \times \frac{\partial \psi}{\partial y_n}(y), \quad y = \psi^{-1}(p) \in W_o,$$

como vimos acima, temos  $w(p) = \det \mathbf{a} \cdot z(p)$  onde  $\mathbf{a}$  é a matriz jacobiana, no ponto  $x$ , da mudança de parametrização  $\psi^{-1} \circ \varphi$ . Logo  $\det \mathbf{a} > 0$  e, conseqüentemente,

$$\frac{w(p)}{|w(p)|} = \frac{z(p)}{|z(p)|} = u(p).$$

Deste modo, o campo unitário normal  $u : M \rightarrow \mathbb{R}^{n+1}$  está bem definido e é, evidentemente, contínuo.  $\square$

O Teorema 4 mostra que a definição de superfície orientável dada no Capítulo 4 é compatível com a definição geral dada aqui.

**Exemplo 12.** Todo subconjunto aberto  $A$  de uma superfície orientável  $M$  é ainda uma superfície orientável. Com efeito, se  $\mathcal{A}$  é um atlas coerente em  $M$  então as restrições  $\varphi|_{(V_o \cap \varphi^{-1}(A))} : V_o \cap \varphi^{-1}(A) \rightarrow V \cap A$  das parametrizações  $\varphi : V_o \rightarrow V$  pertencentes a  $\mathcal{A}$ , com  $V \cap A \neq \emptyset$ , formam um atlas coerente em  $A$ . Portanto se uma superfície bidimensional  $M$  contém uma faixa de Moebius então  $M$  não é orientável.  $\triangleleft$

**Exemplo 13.** O produto  $M \times N$  de duas superfícies  $M$  e  $N$  é uma superfície orientável. Com efeito, se  $\mathcal{A}$  e  $\mathcal{B}$  são atlas coerentes em  $M$  e  $N$  respectivamente então as parametrizações do tipo  $\varphi \times \xi : V_o \times W_o \rightarrow V \times W$ , definidas por  $(\varphi \times \xi)(x, y) = (\varphi(x), \xi(y))$ , onde  $\varphi \in \mathcal{A}$  e  $\xi \in \mathcal{B}$ , formam um atlas em  $M \times N$ , o qual é coerente pois  $(\psi \times \zeta)^{-1} \circ (\varphi \times \xi) = (\psi^{-1} \circ \varphi) \times (\zeta^{-1} \circ \xi)$  e o determinante jacobiano de  $(\psi^{-1} \circ \varphi) \times (\zeta^{-1} \circ \xi)$  é o produto dos determinantes jacobianos de  $\psi^{-1} \circ \varphi$  e  $\zeta^{-1} \circ \xi$ . Reciprocamente, se o produto  $M \times N$  é orientável

então tomamos um atlas coerente  $\mathcal{A}$  em  $M \times N$  e fixamos, de uma vez por todas, uma parametrização  $\xi : Z_o \rightarrow Z$  em  $N$ . O conjunto  $\mathcal{B}$  das parametrizações  $\varphi : V_o \rightarrow V$  em  $M$  tais que  $\varphi \times \xi : V_o \times Z_o \rightarrow V \times Z$  é compatível com todas as parametrizações pertencentes a  $\mathcal{A}$  é certamente um atlas em  $M$ . Além disso, se  $\varphi : V_o \rightarrow V$  e  $\psi : W_o \rightarrow W$  pertencem a  $\mathcal{B}$ , com  $V \cap W \neq \emptyset$ , pondo  $\alpha = (\psi \times \xi)^{-1} \circ (\varphi \times \xi)$ , vemos que  $\alpha = (\psi^{-1} \circ \varphi) \times \text{id}$ , logo o jacobiano de  $\alpha$ , que é positivo, é igual ao de  $\psi^{-1} \circ \varphi$ . Logo  $\mathcal{B}$  é coerente e  $M$  é orientável. Do mesmo modo se mostra que  $N$  também é orientável.  $\triangleleft$

**Exemplo 14.** Em virtude do Teorema 3, a esfera  $S^n$  é uma hiperfície orientável em  $\mathbb{R}^{n+1}$ , pois admite o óbvio campo contínuo de vetores normais unitários  $u : S^n \rightarrow \mathbb{R}^{n+1}$ ,  $u(p) = p$ . Em particular, o círculo  $S^1 \subset \mathbb{R}^2$  é orientável logo, pelo Exemplo 13, o toro  $n$ -dimensional  $T^n = S^1 \times \dots \times S^1$  ( $n$  fatores) é uma superfície orientável em  $\mathbb{R}^{2n}$ .  $\triangleleft$

O lema a seguir serve de fundamento para o Exemplo 15.

**Lema 1.** *Sejam  $\mathcal{A}$  um atlas coerente na superfície orientável  $M$  e  $\varphi : V_o \rightarrow V$  uma parametrização da vizinhança conexa  $V$  em  $M$ . Para qualquer parametrização  $\xi : W_o \rightarrow W$  pertencente a  $\mathcal{A}$ , com  $V \cap W \neq \emptyset$ , o determinante jacobiano  $\det \cdot J(\xi^{-1} \circ \varphi)(x)$  não muda de sinal quando  $x$  varia em  $\varphi^{-1}(V \cap W)$ .*

*Demonstração.* O conjunto  $A$  dos pontos  $p = \varphi(x) \in V$  tais que existe uma parametrização  $\xi : W_o \rightarrow W$ , pertencente ao atlas  $\mathcal{A}$ , com  $p = \xi(w)$ ,  $w \in W_o$  e  $\det J(\xi^{-1} \circ \varphi)(x) > 0$ , é aberto em  $V$ . Analogamente, é aberto o conjunto  $B$  dos pontos  $q \in V$  para os quais existe  $\zeta : Z_o \rightarrow Z$ ,  $\zeta \in \mathcal{A}$ , com  $\zeta(z) = p$  para algum  $z \in Z_o$  e  $\det \cdot J(\zeta^{-1} \circ \varphi)(x) < 0$ . É claro que  $V = A \cup B$ . Além disso,  $A \cap B = \emptyset$  pois se existisse algum ponto  $p = \varphi(x) \in A \cap B$ , teríamos as parametrizações  $\xi : W_o \rightarrow W$ ,  $\zeta : Z_o \rightarrow Z$  pertencentes a  $\mathcal{A}$ , com  $\xi(w) = \zeta(z) = p$  e  $(\xi^{-1} \circ \varphi) \circ (\zeta^{-1} \circ \varphi)^{-1} = \xi^{-1} \circ \zeta$ , logo o determinante jacobiano de  $\xi^{-1} \circ \zeta$  no ponto  $z$  seria o produto dos determinantes  $\det \cdot J(\xi^{-1} \circ \varphi)(x) > 0$  e  $[\det J(\zeta^{-1} \circ \varphi)(x)]^{-1} < 0$ . Então teríamos  $\det \cdot J(\xi^{-1} \circ \zeta)(z) < 0$  e o atlas  $\mathcal{A}$  não seria coerente.  $\square$

**Exemplo 15.** Seja  $M \subset \mathbb{R}^6$  o conjunto das matrizes  $2 \times 3$  de posto 1. Cada elemento  $\mathbf{m} \in M$  será escrito sob a forma  $\mathbf{m} = [u, v]$ , onde os vetores  $u, v \in \mathbb{R}^3$  são as suas linhas. Temos  $M = U \cup V$ , onde  $U$  é o conjunto das matrizes  $\mathbf{m} = [u, v]$  de posto 1 tais que  $u \neq 0$ , enquanto  $V \subset M$  é definido pela condição  $v \neq 0$ . Pondo  $U_o = \mathbb{R} \times (\mathbb{R}^3 - \{0\})$  as aplicações  $\varphi : U_o \rightarrow U$  e  $\psi : U_o \rightarrow V$ , definidas por  $\varphi(t, u) = [u, tu]$  e  $\psi(t, v) = [tv, v]$ , são parametrizações  $C^\infty$ . A interseção  $U \cap V$  é o conjunto das matrizes de posto 1 com ambas as linhas não-nulas, logo  $\varphi^{-1}(U \cap V) = \psi^{-1}(U \cap V) = (\mathbb{R} - \{0\}) \times (\mathbb{R}^3 - \{0\})$  tem duas componentes

conexas:  $\mathbb{R}^+ \times (\mathbb{R}^3 - \{0\})$  e  $\mathbb{R}^- \times (\mathbb{R}^3 - \{0\})$ . A mudança de parametrização  $\xi = \psi^{-1} \circ \varphi : \varphi^{-1}(U \cap V) \rightarrow \psi^{-1}(U \cap V)$  é dada por  $\xi(t, x, y, z) = (1/t, tx, ty, tz)$ . Sua matriz jacobiana em cada ponto  $(t, x, y, z) \in \varphi^{-1}(U \cap V)$  é

$$J\xi(t, x, y, z) = \begin{bmatrix} -\frac{1}{t^2} & 0 & 0 & 0 \\ x & t & 0 & 0 \\ y & 0 & t & 0 \\ z & 0 & 0 & t \end{bmatrix}$$

e seu determinante é igual a  $-t$ . A mudança de parametrização  $\xi = \psi^{-1} \circ \varphi$  tem, portanto, jacobiano negativo em  $\mathbb{R}^+ \times (\mathbb{R}^3 - \{0\})$  e positivo em  $\mathbb{R}^- \times (\mathbb{R}^3 - \{0\})$ . Segue-se do Lema 1 que  $M$  é uma superfície  $C^\infty$ , não-orientável, de dimensão 4 em  $\mathbb{R}^6$ .  $\triangleleft$

## 5 Multiplicadores de Lagrange

Estenderemos agora, para co-dimensão  $n$  qualquer, o método dos multiplicadores de Lagrange, apresentado no Capítulo 4 no caso em que a superfície  $\varphi^{-1}(c)$  tem co-dimensão 1, logo há apenas um multiplicador.

São dadas uma superfície  $M$ , de dimensão  $m$  e classe  $C^k$ , e uma função  $f : U \rightarrow \mathbb{R}$ , de classe  $C^k$  no aberto  $U$ , com  $M \subset U \subset \mathbb{R}^{m+n}$ . Quer-se determinar o conjunto dos pontos críticos da restrição  $f|M$ .

Diz-se que  $p \in M$  é um *ponto crítico* da restrição  $f|M$  quando, para todo caminho diferenciável  $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$ , com  $\lambda(0) = p$ , tem-se  $(f \circ \lambda)'(0) = 0$ .

Como  $\lambda'(0) = v \in T_p M$  e portanto

$$(f \circ \lambda)'(0) = \frac{\partial f}{\partial v}(p) = \langle \text{grad } f(p), v \rangle,$$

concluímos que  $p \in M$  é ponto crítico de  $f|M$  se, e somente se,  $\text{grad } f(p)$  é ortogonal a todos os vetores  $v \in T_p M$ , tangentes a  $M$  no ponto  $p$ , ou seja,  $\text{grad } f(p) \in [T_p M]^\perp$ .

Se  $p \in M$  é um ponto de mínimo (ou máximo) local da restrição  $f|M$  e  $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$  é um caminho diferenciável com  $\lambda(0) = p$  então 0 é um ponto de mínimo (ou máximo) local de  $f \circ \lambda : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$ , logo  $(f \circ \lambda)'(0) = 0$  e então  $p$  é um ponto crítico de  $f|M$ .

**Exemplo 16.** Suponhamos que a superfície  $M \subset \mathbb{R}^{m+n}$  seja um subconjunto fechado. Então, fixado um ponto  $a \in \mathbb{R}^{m+n}$ , existe, entre os pontos de  $M$ , (pelo menos) um ponto  $p$  situado a uma distância mínima de  $a$ . Considerando a função  $f : \mathbb{R}^{m+n} \rightarrow \mathbb{R}$ , dada por  $f(x) = |x - a|^2$ , vemos que  $p$  é um ponto de

mínimo da restrição  $f|M$ . Logo  $\text{grad } f(p)$  é um vetor ortogonal a  $T_pM$ . Mas  $\text{grad } f(p) = 2 \cdot (x - a)$ . Portanto os pontos  $p \in M$  situados à distância mínima do ponto  $a$  são aqueles tais que o vetor  $p - a$  é ortogonal a  $T_pM$ . Evidentemente, vale o mesmo para os pontos de  $M$  mais afastados de  $a$ , caso existam (como ocorre quando  $M$  é compacta).

Suponhamos agora que a superfície  $M = \varphi^{-1}(c)$  seja obtida como imagem inversa do valor regular  $c$  da aplicação  $\varphi : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$ . Se escrevermos  $\varphi(x) = (\varphi_1(x), \dots, \varphi_n(x))$ , a afirmação de que  $c$  é um valor regular de  $\varphi$  significa que os vetores  $\text{grad } \varphi_1(x), \dots, \text{grad } \varphi_n(x)$  são linearmente independentes para todo  $x \in U$  tal que  $\varphi(x) = c$ .

Com efeito, esses  $n$  vetores são as linhas da matriz jacobiana  $J\varphi(x) \in M(n \times (m+n))$ , a qual tem posto  $n$  por ser a matriz da transformação linear sobrejetiva  $\varphi'(x) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$ .

Além disso, conforme já vimos no início da seção 4, em todo ponto  $x \in M = \varphi^{-1}(c)$ , os vetores  $\text{grad } \varphi_1(x), \dots, \text{grad } \varphi_n(x)$  são ortogonais a  $T_xM$ , portanto formam uma base do complemento ortogonal  $[T_xM]^\perp$ .  $\triangleleft$

Podemos então enunciar o

**Método dos multiplicadores de Lagrange.** *Sejam  $f : U \rightarrow \mathbb{R}$  uma função de classe  $C^k$  no aberto  $U \subset \mathbb{R}^{m+n}$  e  $M = \varphi^{-1}(c)$  a imagem inversa do valor regular  $c$  pela aplicação  $\varphi : U \rightarrow \mathbb{R}^n$ , de classe  $C^k$ . A fim de que  $p \in M$  seja um ponto crítico da restrição  $f|M$  é necessário e suficiente que existam números  $\lambda_1, \dots, \lambda_n$  tais que  $\text{grad } f(p) = \lambda_1 \text{grad } \varphi_1(p) + \dots + \lambda_n \text{grad } \varphi_n(p)$ .*

Os números  $\lambda_1, \dots, \lambda_n$  são chamados *multiplicadores de Lagrange*.

De fato  $p$  é ponto crítico de  $f|M$  se, e somente se,  $\text{grad } f(p)$  é ortogonal a  $T_pM$ . Como  $\{\text{grad } \varphi_1(p), \dots, \text{grad } \varphi_n(p)\}$  é uma base do complemento ortogonal de  $T_pM$  em  $\mathbb{R}^{m+n}$ , dizer que  $\text{grad } f(p) \in [T_pM]^\perp$  equivale a afirmar que  $\text{grad } f(p)$  é combinação linear dos gradientes  $\text{grad } \varphi_1(p), \dots, \text{grad } \varphi_n(p)$ .  $\square$

Seja  $c = (c_1, \dots, c_n)$ . Para encontrar os pontos críticos  $p$  da restrição  $f|M$ , devemos resolver o sistema abaixo, de  $m+n$  equações com  $m+n$  incógnitas. (As incógnitas são as  $m+n$  coordenadas de  $p$  mais os  $n$  multiplicadores  $\lambda_i$ ):

$$\begin{cases} \varphi_1(p) = c_1, \dots, \varphi_n(p) = c_n \\ \text{grad } f(p) = \lambda_1 \text{grad } \varphi_1(p) + \dots + \lambda_n \text{grad } \varphi_n(p). \end{cases}$$

A última equação acima é vetorial. Ela equivale às  $m+n$  equações numéricas

$$\frac{\partial f}{\partial x_j}(p) = \lambda_1 \cdot \frac{\partial \varphi_1}{\partial x_j}(p) + \dots + \lambda_n \frac{\partial \varphi_n}{\partial x_j}(p), \quad j = 1, \dots, m+n.$$

**Exemplo 17.** Seja  $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$  uma transformação linear. Defina  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n \rightarrow \mathbb{R}$  pondo  $f(x, y) = \langle A \cdot x, y \rangle = \langle x, A^* \cdot y \rangle = \sum a_{ij} x_i \cdot y_j$  ( $1 \leq i \leq n$ ,  $1 \leq j \leq m$ ). Considerando o valor regular  $(1, 1)$  da aplicação  $\varphi : \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}^2$ , dada por  $\varphi(x, y) = (|x|^2, |y|^2)$ , seja  $M = S^{m-1} \times S^{n-1} = \varphi^{-1}(1, 1)$ . Vejamos quais são os pontos críticos da restrição  $f|M$ .

Para todo  $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$ , temos  $\text{grad } f(x, y) = (A^* \cdot y, A \cdot x) \in \mathbb{R}^m \times \mathbb{R}^n$ . Além disso,  $\varphi = (\varphi_1, \varphi_2)$ , com  $\text{grad } \varphi_1(x, y) = (2x, 0)$  e  $\text{grad } \varphi_2(x, y) = (0, 2y)$ . Por conveniência, tomemos  $\lambda/2$  e  $\mu/2$  como multiplicadores de Lagrange. Um ponto  $p = (x, y) \in M$  é crítico para  $f|M$  se, e somente se,

$$\text{grad } f(x, y) = \frac{\lambda}{2} \cdot \text{grad } \varphi_1(x, y) + \frac{\mu}{2} \text{grad } \varphi_2(x, y),$$

ou seja,  $(A^*y, Ax) = (\lambda x, \mu y)$ . Isto nos dá  $A \cdot x = \mu \cdot y$  e  $A^*y = \lambda \cdot x$ , donde  $\mu = \langle \mu y, y \rangle = \langle Ax, y \rangle = \langle x, A^*y \rangle = \langle x, \lambda \cdot x \rangle = \lambda$ .

Portanto, os pontos críticos de  $f|M$  são os pontos  $(x, y) \in S^{m-1} \times S^{n-1}$  tais que  $Ax = \lambda y$  e  $A^*y = \lambda x$  para um certo  $\lambda \in \mathbb{R}$ . Notemos que então  $\lambda = f(x, y)$  e que  $z \perp x \Rightarrow Az \perp y$ . Assim, se escrevermos  $E = \{z \in \mathbb{R}^m; \langle z, x \rangle = 0\} =$  complemento ortogonal de  $x$  em  $\mathbb{R}^m$ , e  $F =$  complemento ortogonal de  $y$  em  $\mathbb{R}^n$ , a transformação linear  $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$  aplica  $E$  em  $F$ .

Seja então  $p_1 = (u_1, v_1) \in S^{m-1} \times S^{n-1}$  o ponto em que a função  $f$  assume seu valor máximo em  $S^{m-1} \times S^{n-1} : f(u_1, v_1) = \lambda_1$ . Então  $p_1$  é ponto crítico de  $f|M$ . Temos  $Au_1 = \lambda_1 v_1$  e  $A^*v_1 = \lambda_1 u_1$ . Como  $f(x, -y) = -f(x, y)$ , vemos que  $\lambda_1 \geq 0$ . Se  $A \neq 0$  então  $f$  não é identicamente nula em  $M$ , logo  $\lambda_1 > 0$ .  $\triangleleft$

Em seguida consideremos  $A$  como uma transformação linear  $A : E \rightarrow F$ , agora com  $\dim E = m - 1$  e  $\dim F = n - 1$ . Prosseguindo por indução, chegaremos ao seguinte resultado:

**Teorema dos Valores Singulares.** *Seja  $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$  uma transformação linear de posto  $r$ . Existem bases ortonormais  $\{u_1, \dots, u_m\} \subset \mathbb{R}^m$  e  $\{v_1, \dots, v_n\} \subset \mathbb{R}^n$  tais que  $Au_i = \lambda_i v_i$  e  $A^*v_i = \lambda_i u_i$ , onde  $\lambda_i > 0$  para  $i = 1, \dots, r$  e  $\lambda_i = 0$  para  $i \geq r + 1$ .  $\square$*

Os números  $\lambda_1 > 0, \dots, \lambda_r > 0$  são chamados os *valores singulares* de  $\Delta$ .

## Capítulo 8

# Integrais Múltiplas

### 1 A definição de integral

Um *bloco n-dimensional*  $A \subset \mathbb{R}^n$  é um produto cartesiano

$$A = \prod_{i=1}^n [a_i, b_i] = [a_1, b_1] \times \dots \times [a_n, b_n]$$

de  $n$  intervalos compactos  $[a_i, b_i]$ , chamados suas *arestas*. O produto cartesiano  $\prod_{i=1}^n (a_i, b_i)$  dos intervalos abertos  $(a_i, b_i)$  chama-se *bloco n-dimensional aberto*. Quando todas as arestas do bloco  $A$  têm o mesmo comprimento  $a = b_i - a_i$ , diz-se que  $A$  é um *cubo n-dimensional*. Quando  $n = 1$ , os blocos são simplesmente intervalos. Se  $n = 2$ , o bloco reduz-se a um retângulo e o cubo a um quadrado.

Uma *face* do bloco  $A = \prod [a_i, b_i]$  é um conjunto do tipo  $F = \prod_{i=1}^n L_i$  onde, para cada  $i = 1, \dots, n$ , tem-se  $L_i = \{a_i\}$ ,  $L_i = \{b_i\}$  ou  $L_i = \{a_i, b_i\}$ . Diz-se que a face  $F$  tem *dimensão*  $r$  quando exatamente  $r$  dos fatores  $L_i$  são iguais a  $[a_i, b_i]$ . As faces de dimensão zero, isto é, os pontos da forma  $v = (c_1, \dots, c_n)$ , onde  $c_i = a_i$  ou  $c_i = b_i$  para cada  $i = 1, \dots, n$ , são chamadas os *vértices* do bloco.

O *volume n-dimensional* do bloco  $A = \prod_{i=1}^n [a_i, b_i]$  é, por definição, o produto  $\prod_{i=1}^n (b_i - a_i)$  dos comprimentos de suas arestas. Este é também o volume do bloco aberto  $\prod_{i=1}^n (a_i, b_i)$ .

Uma *partição* do bloco  $A = \prod_{i=1}^n [a_i, b_i]$  é um produto cartesiano  $P = P_1 \times \dots \times P_n$ , onde cada  $P_i$  é uma partição do intervalo  $[a_i, b_i]$  (cfr. Vol. 1, Cap. 10, §2). Diz-se que a partição  $Q = Q_1 \times \dots \times Q_n$  *refina* a partição  $P$  quando se tem

$P \subset Q$ . Isto equivale a dizer que  $P_1 \subset Q_1, \dots, P_n \subset Q_n$ .

A partição  $P$  decompõe o bloco  $A$  numa reunião de sub-blocos  $B = I_1 \times \dots \times I_n$ , onde cada  $I_j$  é um intervalo da partição  $P_j$  de  $[a_j, b_j]$ . Estes sub-blocos  $B \subset A$  chamam-se os *blocos da partição*  $P$ . Escreve-se então  $B \in P$ . Se a partição  $Q$  refina  $P$  então cada bloco de  $P$  é a reunião dos blocos de  $Q$  nele contidos.

Se  $B, B'$  são blocos da partição  $P$ , a interseção  $B \cap B'$  é uma face comum a  $B$  e  $B'$  ou é vazia.

Dada a partição  $P = P_1 \times \dots \times P_n$  do bloco  $A$ , como o comprimento  $b_i - a_i$  de cada aresta de  $A$  é a soma dos comprimentos dos intervalos da partição  $P_i$ , segue-se da propriedade distributiva da multiplicação que o volume do bloco  $A$  é a soma dos volumes dos blocos  $B$  da partição  $P$ . Logo, quando  $Q$  refina  $P$ , o volume de cada bloco de  $P$  é a soma dos volumes dos blocos de  $Q$  nele contidos.

Se  $P = P_1 \times \dots \times P_n$  e  $Q = Q_1 \times \dots \times Q_n$  são partições do bloco  $A$ , existem partições de  $A$  que refinam ao mesmo tempo  $P$  e  $Q$ . Uma delas é

$$R = \prod_{i=1}^n (P_i \cup Q_i).$$

Seja  $f : A \rightarrow \mathbb{R}$  uma função real limitada no bloco  $n$ -dimensional  $A$ , digamos com  $m \leq f(x) \leq M$  para todo  $x \in A$ . Dada uma partição  $P$  do bloco  $A$ , para cada bloco  $B \in P$ , indiquemos com  $m_B$  o ínfimo e com  $M_B$  o supremo dos valores  $f(x)$  quando  $x$  varia em  $B$ . Definimos então a *soma inferior*  $s(f; P)$  e a *soma superior*  $S(f; P)$  da função  $f$  relativamente à partição  $P$  pondo

$$s(f; P) = \sum_{B \in P} m_B \cdot \text{vol } B \quad \text{e} \quad S(f; P) = \sum_{B \in P} M_B \cdot \text{vol } B,$$

estas somas estendendo-se a todos os blocos  $B$  da partição  $P$ . Evidentemente  $m_B \leq M_B$  para todo  $B \in P$ , logo  $s(f; P) \leq S(f; P)$ .

Mais do que isto é verdade: para quaisquer partições  $P$  e  $Q$  do bloco  $A$ , tem-se  $s(f; P) \leq S(f; Q)$ .

Para comprovar esta afirmação, observamos primeiro que se uma partição  $P'$  refina a partição  $P$  então  $s(f; P) \leq s(f; P')$  e  $S(f; P') \leq S(f; P)$ .

Com efeito, se o bloco  $B'$  da partição  $P'$  está contido no bloco  $B$  da partição  $P$  então  $m_B \leq m_{B'}$ . Lembrando que cada bloco  $B \in P'$  é a reunião dos blocos  $B' \in P'$  nele contidos, e que  $\text{vol } B = \sum_{B' \subset B} \text{vol } B'$ , segue-se que

$$\begin{aligned} s(f; P) &= \sum_{B \in P} m_B \cdot \text{vol } B = \sum_{B \in P} \left( \sum_{B' \subset B} m_B \cdot \text{vol } B' \right) \\ &\leq \sum_{B' \in P'} m_{B'} \cdot \text{vol } B' = s(f; P'). \end{aligned}$$

Analogamente se vê que  $S(f; P') \leq S(f; P)$  quando  $P'$  refina  $P$ . Assim, quando se refina uma partição, a soma inferior não diminui e a soma superior não aumenta.

Sejam  $P$  e  $Q$  duas partições quaisquer do bloco  $A$ . Tomemos uma partição  $R$  de  $A$  que refine  $P$  e  $Q$  simultaneamente. Temos:

$$s(f; P) \leq s(f; R) \leq S(f; R) \leq S(f; Q),$$

mostrando portanto que  $s(f; P) \leq S(f; Q)$ , ou seja, toda soma inferior de  $f$  é menor do que ou igual a qualquer soma superior.

Definimos, a seguir, a *integral inferior*  $\int_{\underline{A}} f(x)dx$  e a *integral superior*  $\int_{\overline{A}} f(x)dx$  da função limitada  $f : A \rightarrow \mathbb{R}$ , pondo

$$\int_{\underline{A}} f(x)dx = \sup_P s(f; P) \quad \text{e} \quad \int_{\overline{A}} f(x)dx = \inf_P S(f; P),$$

o supremo e o ínfimo acima sendo tomadas em relação a todas as partições  $P$  do bloco  $A$ .

A desigualdade  $s(f; P) \leq S(f; Q)$  implica que

$$m \cdot \text{vol } A \leq \int_{\underline{A}} f(x)dx \leq \int_{\overline{A}} f(x)dx \leq M \cdot \text{vol } A$$

se  $m \leq f(x) \leq M$  para todo  $x \in A$ .

Diz-se que a função limitada  $f : A \rightarrow \mathbb{R}$  é *integrável* no bloco  $n$ -dimensional  $A$  quando suas integrais inferior e superior coincidem. Escreve-se então

$$\int_A f(x)dx = \int_{\underline{A}} f(x)dx = \int_{\overline{A}} f(x)dx$$

e este número é chamado a *integral de  $f$*  no bloco  $A$ .

No caso  $n = 1$ , o bloco  $n$ -dimensional  $A$  reduz-se a um segmento de reta e a definição de integral acima dada coincide com a queda apresentada no Capítulo 10 do Volume 1.

Dada a função  $f : A \rightarrow \mathbb{R}$ , limitada no bloco  $A \subset \mathbb{R}^n$ , o conjunto  $\sigma$  das somas inferiores e o conjunto  $\Sigma$  das somas superiores de  $f$  relativamente às partições  $P$  de  $A$  são subconjuntos do intervalo  $[m \cdot \text{vol } A, M \cdot \text{vol } A]$ , onde  $m \leq f(x) \leq M$  para todo  $x \in A$ . Sabemos que, para quaisquer  $s \in \tau$  e  $S \in \Sigma$ , tem-se  $s \leq S$ . A fim de que seja  $\sup \tau = \inf \Sigma$ , isto é, que  $f$  seja integrável, é necessário e suficiente que, dado arbitrariamente  $\varepsilon > 0$ , existam  $s \in \tau$  e  $S \in \Sigma$  tais que  $S - s < \varepsilon$ . Mais explicitamente:  $f$  é integrável se, e somente se, para todo  $\varepsilon > 0$  dado, existem partições  $R$  e  $Q$  de  $A$  tais que  $S(f; R) - s(f; Q) < \varepsilon$ . Esta condição pode ser aperfeiçoada assim:

**Teorema 1 (Condição imediata de integrabilidade).** *A fim de que a função limitada  $f : A \rightarrow \mathbb{R}$  seja integrável no bloco  $A \subset \mathbb{R}^n$  é necessário e suficiente que, para todo  $\varepsilon > 0$  dado, exista uma partição  $P$  de  $A$  tal que  $S(f; P) - s(f; P) < \varepsilon$ .*

*Demonstração.* A suficiência é óbvia pois a condição acima claramente assegura que  $\sup \tau = \inf \Sigma$ . Quanto à necessidade, supondo  $f$  integrável, ou seja, admitindo que  $\sup \tau = \inf \Sigma$ , dado  $\varepsilon > 0$ , existem partições  $Q$  e  $R$  do bloco  $A$  tais que  $S(f; R) - s(f; Q) < \varepsilon$ . Seja  $P$  uma partição de  $A$  que refine  $Q$  e  $R$  ao mesmo tempo. Então  $s(f; Q) \leq s(f; P) \leq S(f; P) \leq S(f; R)$  portanto  $S(f; P) - s(f; P) \leq S(f; R) - s(f; Q) < \varepsilon$ .

Para todo subconjunto  $X \subset A$ , sejam  $M_X$  o supremo e  $m_X$  o ínfimo dos valores  $f(x)$ , com  $x \in X$ . Escreveremos  $\omega_X = M_X - m_X$  e chamaremos  $\omega_X$  de *oscilação* de  $f$  no conjunto  $X$ .

Às vezes, quando houver necessidade, usaremos a notação mais precisa  $\omega(f; X) = M_X - m_X$  em vez de  $\omega_X$ .  $\square$

Para toda partição  $P$  do bloco  $A$ , temos

$$S(f; P) - s(f; P) = \sum_{B \in P} (M_B - m_B) \text{vol } B = \sum_{B \in P} \omega_B \cdot \text{vol } B$$

Portanto  $f : A \rightarrow \mathbb{R}$  é integrável se, e somente se, para todo  $\varepsilon > 0$  dado, existe uma partição  $P$  de  $A$  tal que

$$\sum_{B \in P} \omega_B \cdot \text{vol } B < \varepsilon.$$

Uma conseqüência imediata desta observação é que toda função contínua  $f : A \rightarrow \mathbb{R}$  é integrável. Com efeito, sendo o bloco  $A$  um conjunto compacto, a função contínua  $f$  é uniformemente contínua. Portanto, dado qualquer  $\varepsilon > 0$ , existe  $\delta > 0$  tal que  $x, y \in A$ ,  $|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon/\text{vol } A$ .

Se tomarmos, em cada aresta  $[a_i, b_i]$  do bloco  $A = \Pi[a_i, b_i]$ , uma partição  $P_i$  cujos intervalos tenham todos comprimento  $< \delta$ , e adotarmos em  $\mathbb{R}^n$  a norma do máximo, então todos os blocos da partição  $P = P_1 \times \cdots \times P_n$  de  $A$  terão diâmetro menor do que  $\delta$ . A função  $f$  sendo contínua, em cada bloco  $B$  da partição  $P$  existem pontos  $a, b$  tais que  $m_B = f(a)$  e  $M_B = f(b)$ , pois  $B$  é compacto. Então  $\omega_B = M_B - m_B = f(b) - f(a) < \varepsilon/\text{vol } A$  e daí

$$\sum_{B \in P} \omega_B \cdot \text{vol } B < \frac{\varepsilon}{\text{vol } A} \cdot \sum_{B \in P} \text{vol } B = \frac{\varepsilon}{\text{vol } A} \cdot \text{vol } A = \varepsilon.$$

Assim  $f$  é integrável.

O fato de que toda função contínua  $f : A \rightarrow \mathbb{R}$  é integrável é muito importante mas não é suficiente para nossos propósitos. A fim de definir a integral de funções cujos domínios são mais gerais do que blocos, precisamos integrar alguns tipos de funções descontínuas. Isto nos leva ao critério de integrabilidade de Lebesgue, o qual se baseia na noção de conjunto de medida nula, que abordaremos no parágrafo seguinte.

## 2 Conjunto de medida nula

Diz-se que o conjunto  $X \subset \mathbb{R}^n$  tem *medida  $n$ -dimensional nula* (segundo Lebesgue), e escreve-se  $\text{med}.X = 0$ , quando, para todo  $\varepsilon > 0$  dado, é possível obter uma cobertura enumerável  $X \subset B_1 \cup \dots \cup B_k \cup \dots$  por meio de blocos abertos  $B_k \subset \mathbb{R}^n$  tais que  $\sum_{k=1}^{\infty} \text{vol } B_k < \varepsilon$ .

Evidentemente, se  $\text{med}.X = 0$  e  $Y \subset X$  então  $\text{med}.Y = 0$ .

**Teorema 2.** *Uma reunião enumerável de conjuntos de medida nula é ainda um conjunto de medida nula.*

*Demonstração.* Sejam  $X_1, \dots, X_k, \dots$  subconjuntos de  $\mathbb{R}^n$  com  $\text{med}.X_k = 0$  para todo  $k \in \mathbb{N}$ . A fim de provar que  $X = \bigcup_{k=1}^{\infty} X_k$  tem medida nula, seja dado  $\varepsilon > 0$ . Para cada  $k \in \mathbb{N}$  podemos obter uma seqüência de blocos  $B_{k1}, B_{k2}, \dots, B_{ki}, \dots$  tais que  $X_k \subset \bigcup_{i=1}^{\infty} B_{ki}$  e  $\sum_{i=1}^{\infty} \text{vol } B_{ki} < \varepsilon/2^k$ . Então  $X$  está contido na reunião (enumerável) de todos os  $B_{ki}$ . Dado qualquer subconjunto finito  $F \subset \mathbb{N} \times \mathbb{N}$ , existe  $j \in \mathbb{N}$  tal que  $(k, i) \in F \Rightarrow k \leq j$  e  $i \leq j$ . Logo

$$\sum_{(k,i) \in F} \text{vol } B_{ki} \leq \sum_{k=1}^j \left[ \sum_{i=1}^j \text{vol } B_{ki} \right] < \sum_{k=1}^j \varepsilon/2^k < \varepsilon.$$

Portanto, seja qual for a maneira de enumerar os  $B_{ki}$  numa seqüência, teremos  $\sum_{k,i} \text{vol } B_{ki} < \varepsilon$ . Assim,  $\text{med}.X = 0$ .  $\square$

**Corolário 1.** *Todo conjunto enumerável tem medida nula.*

Com efeito, todo conjunto enumerável é reunião dos seus pontos, cada um dos quais tem medida nula.

A definição de  $\text{med}.X = 0$ , dada acima com blocos abertos, é conveniente quando se pretende usar o Teorema de Borel-Lebesgue. Noutras ocasiões, pode ser mais adequado empregar blocos fechados. E ainda em certos casos impõe-se o

uso de cubos (abertos ou fechados). Essas alternativas são equivalentes, conforme veremos agora.

**Teorema 3.** *As seguintes afirmações a respeito de um conjunto  $X \subset \mathbb{R}^n$  são equivalentes:*

- (a) *Para todo  $\varepsilon > 0$  dado, pode-se obter uma cobertura enumerável  $X \subset B_1 \cup \dots \cup B_k \cup \dots$  por meio de blocos abertos  $B_k \subset \mathbb{R}^n$  tais que  $\sum_{k=1}^{\infty} \text{vol } B_k < \varepsilon$ .*
- (b) *Vale a afirmação (a), com blocos fechados em vez de abertos.*
- (c) *Vale a afirmação (a), com cubos abertos em vez de blocos.*
- (d) *Vale a afirmação (a), com cubos fechados em vez de blocos abertos.*

*Demonstração.* Mostremos, inicialmente, que (a)  $\Leftrightarrow$  (b). A implicação (a)  $\Rightarrow$  (b) é imediata, pois  $X \subset B_1 \cup \dots \cup B_k \cup \dots$  implica  $X \subset \overline{B}_1 \cup \dots \cup \overline{B}_k \cup \dots$  e  $\text{vol } \overline{B}_k = \text{vol } B_k$ , logo  $\sum \text{vol } B_k < \varepsilon \Rightarrow \sum \text{vol } \overline{B}_k < \varepsilon$ . Quanto a (b)  $\Rightarrow$  (a): dado  $\varepsilon > 0$ , (b) nos autoriza a obter uma cobertura  $X \subset D_1 \cup \dots \cup D_k \cup \dots$  por meio de blocos fechados  $D_k$  com  $\sum \text{vol } D_k < \varepsilon/2$ . Ora, para cada  $k \in \mathbb{N}$ , o bloco  $D_k = \prod_{i=1}^n [a_{ki}, b_{ki}]$  está contido no bloco aberto  $A_k = \prod_{i=1}^n (a_{ki} - \delta, b_{ki} + \delta)$  onde  $\delta > 0$  pode ser escolhido de modo que  $\text{vol } A_k - \text{vol } D_k < \varepsilon/2^{k+2}$ . (Basta notar que  $\text{vol } A_k = \prod_{i=1}^n (b_{ki} - a_{ki} + 2\delta)$  é uma função contínua de  $\delta$ , igual a  $\text{vol } D_k$  quando  $\delta = 0$ .) Então  $X \subset A_1 \cup \dots \cup A_k \cup \dots$ , com

$$\sum_{k=1}^{\infty} \text{vol } A_k < \sum_{k=1}^{\infty} \text{vol } D_k + \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+2}} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

□

A equivalência (c)  $\Leftrightarrow$  (d) se prova exatamente do mesmo modo, bastando observar que, no argumento acima, se  $D_k$  é um cubo então  $A_k$  também é.

Resta portanto, provar que (a)  $\Leftrightarrow$  (c) ou, o que dá no mesmo, que (b)  $\Leftrightarrow$  (d). Ora, é óbvio que (d)  $\Rightarrow$  (b). Para demonstrar a implicação (b)  $\Rightarrow$  (d), começamos provando que, dados um bloco  $B$  e um número  $\delta > 0$ , existe um bloco  $C$  que é uma reunião finita de cubos, contém  $B$  e, além disso,  $\text{vol } C - \text{vol } B < \delta$ . Isto é imediato quando as arestas do bloco  $B = \prod [a_i, b_i]$  têm medidas racionais  $b_k - a_k = p_i/q_i$ . Neste caso, o próprio bloco  $B$  é uma reunião finita de cubos: basta considerar o mínimo múltiplo comum  $m$  dos denominadores  $q_i$  e tomar em cada aresta  $[a_i, b_i]$

de  $B$  a partição  $P_i$  com  $\frac{m p_i}{q^i}$  intervalos, todos de comprimento  $1/m$ . Os blocos da partição  $P = P_1 \times \cdots \times P_n$  do bloco  $B$  são cubos de arestas medindo  $1/m$  e  $B$  é a reunião deles. No caso geral, observamos que, para  $i = 1, \dots, n$ , existem números positivos  $\eta_i$  tão pequenos quanto se queira, tais que  $b_i - a_i + \eta_i$  é racional. Então o bloco  $C = \prod_{i=1}^n [a_i, b_i + \eta_i]$  tem arestas com medidas racionais, logo é reunião finita de cubos. Além disso,  $C$  contém  $B$  e a diferença  $\text{vol } C - \text{vol } B = \prod (b_i - a_i + \eta_i) - \prod (b_i - a_i)$  pode ser tornada tão pequena quanto se deseje, desde que os  $\eta_i$  sejam suficientemente pequenos.

Para completar a prova de que (b)  $\Rightarrow$  (d), seja dado  $\varepsilon > 0$ . Por (b), existe uma cobertura  $X \subset B_1 \cup \dots \cup B_k \cup \dots$  por bloco  $B_k$  tais que  $\sum_{k=1}^{\infty} \text{vol } B_k < \varepsilon/2$ . Como acabamos de ver, cada  $B_k$  está contido numa reunião finita de cubos cuja soma dos volumes é menor do que  $\text{vol } B_k + \varepsilon/2^{k+2}$ . Numerando consecutivamente esses cubos para  $k = 1, 2, \dots$ , chegamos a uma cobertura  $X \subset C_1 \cup \dots \cup C_r \cup \dots$ , onde os cubos  $C_r$  são tais que

$$\sum_{r=1}^{\infty} \text{vol } C_r < \sum_{k=1}^{\infty} \text{vol } B_k + \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+2}} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Isto completa a demonstração do teorema.

**Teorema 4.** *Seja  $f : X \rightarrow \mathbb{R}^n$  uma aplicação lipschitziana no conjunto  $X \subset \mathbb{R}^n$ . Se  $\text{med. } X = 0$  então  $\text{med. } f(X) = 0$ .*

*Demonstração.* Adotemos em  $\mathbb{R}^n$  a norma do máximo. Seja  $c > 0$  tal que  $|f(x) - f(y)| \leq c|x - y|$  para quaisquer  $x, y \in X$ . Dado arbitrariamente  $\varepsilon > 0$ , existe uma cobertura  $X \subset C_1 \cup \dots \cup C_k \cup \dots$  onde cada  $C_k$  é um cubo cuja aresta mede  $a_k$ , com  $\sum_{k=1}^{\infty} \text{vol } C_k = \sum_{k=1}^{\infty} (a_k)^n < \varepsilon/C^n$ . Se  $x, y \in C_k$  então  $|x - y| \leq a_k$ , logo  $|f(x) - f(y)| \leq c \cdot a_k$ . Isto significa que, para todo  $i = 1, \dots, n$ , as  $i$ -ésimas coordenadas de  $f(x)$  e  $f(y)$  pertencem a um intervalo  $J_i$  de comprimento  $c \cdot a_k$ . Portanto  $f(C_k \cap X)$  está contido no cubo  $\prod_{i=1}^n J_i = C'_k$ , de aresta  $c \cdot a_k$ , logo  $\text{vol } C'_k = C^n \cdot (a_k)^n$ . Segue-se que  $f(X) = \bigcup_{k=1}^{\infty} f(C_k \cap X) \subset C'_1 \cup \dots \cup C'_k \cup \dots$ , onde

$$\sum_{k=1}^{\infty} \text{vol } (C'_k) \leq C^n \cdot \sum_{k=1}^{\infty} (a_k)^n < C^n \cdot \frac{\varepsilon}{C^n} = \varepsilon.$$

Logo  $\text{med. } f(X) = 0$ . □

A aplicação mais freqüente do Teorema 4 ocorre quando  $f : U \rightarrow \mathbb{R}^n$  é diferenciável, com derivada limitada no aberto convexo  $U \subset \mathbb{R}^n$ . Se  $|f'(x)| \leq c$  para todo  $x \in U$  então a Desigualdade do Valor Médio nos dá  $|f(x) - f(y)| \leq c \cdot |x - y|$  para quaisquer  $x, y \in U$ , logo  $f$  transforma todo conjunto de medida nula  $X \subset U$  num conjunto de medida nula  $f(X) \subset \mathbb{R}^n$ . A fim de entender este resultado para funções de classe  $C^1$  em abertos não necessariamente convexos, com derivada limitada ou não, será necessário usar o

**Teorema 5 (Lindelöf).** *Toda cobertura aberta  $X \subset \bigcup_{\lambda \in L} U_\lambda$  de um conjunto arbitrário  $X \subset \mathbb{R}^n$  admite uma subcobertura enumerável  $X \subset U_{\lambda_1} \cup \dots \cup U_{\lambda_k} \cup \dots$*

*Demonstração.* Seja  $B$  o conjunto dos blocos abertos em  $\mathbb{R}^n$  cujos vértices têm coordenadas racionais e cada um deles está contido em algum aberto  $U_\lambda$  da cobertura dada. O conjunto  $B$  é enumerável, logo podemos escrever  $B = \{B_1, B_2, \dots, B_k, \dots\}$ . Para cada  $k \in \mathbb{N}$ , escolhamos um índice  $\lambda_k \in L$  tal que  $B_k \subset U_{\lambda_k}$ . Afirmamos que  $\bigcup_{k=1}^{\infty} U_{\lambda_k} = \bigcup_{\lambda \in L} U_\lambda$ . Com efeito, se  $x \in U_\lambda$  então, como  $U_\lambda$  é aberto, existe uma bola aberta de centro  $x$ , contida em  $U_\lambda$ . Se tomarmos em  $\mathbb{R}^n$  a norma do máximo, essa bola é um cubo, cuja aresta podemos supor racional, logo é um  $B_k$ . Assim,  $x \in B_k \subset U_{\lambda_k}$ , portanto todo  $U_\lambda, \lambda \in L$  está contido na reunião dos  $U_{\lambda_k}, k \in \mathbb{N}$ , ou seja  $\bigcup_{\lambda \in L} U_\lambda \subset \bigcup_{k=1}^{\infty} U_{\lambda_k} \subset \bigcup_{\lambda \in L} U_\lambda$ . Segue-se que  $X \subset U_{\lambda_1} \cup \dots \cup U_{\lambda_k} \cup \dots$   $\square$

**Teorema 6.** *Seja  $f : U \rightarrow \mathbb{R}^n$  uma aplicação de classe  $C^1$  no aberto  $U \subset \mathbb{R}^n$ . Se  $X \subset U$  tem medida nula então  $f(X) \subset \mathbb{R}^n$  também tem medida nula.*

*Demonstração.* Para cada  $x \in X$  existe um aberto convexo  $U_x$ , com  $x \in U_x \subset U$ , tal que  $f$  tem derivada limitada em  $U_x$ , logo  $f(X \cap U_x)$  tem medida nula. A cobertura aberta  $X \subset \bigcup_{x \in X} U_x$  admite uma subcobertura enumerável  $X \subset \bigcup_{k=1}^{\infty} U_k$ . Como  $f(X \cap U_k)$  tem medida nula para cada  $k \in \mathbb{N}$ , segue-se que  $f(X) = f\left[\bigcup_{k=1}^{\infty} (X \cap U_k)\right] = \bigcup_{k=1}^{\infty} f(X \cap U_k)$  tem medida nula.  $\square$

**Corolário 2.** *Seja  $f : U \rightarrow \mathbb{R}^m$  uma aplicação de classe  $C^1$  no aberto  $U \subset \mathbb{R}^m$ . Se  $m < n$  então  $f(U) \subset \mathbb{R}^m$  tem medida nula.*

Com efeito, se considerarmos  $\mathbb{R}^m$  como o conjunto dos pontos de  $\mathbb{R}^n$  cujas últimas  $n-m$  coordenadas são nulas, veremos que todo bloco  $m$ -dimensional  $B \subset \mathbb{R}^m \subset \mathbb{R}^n$  tem volume  $n$ -dimensional nulo, pois podemos cobrir  $B$  com um único bloco  $n$ -dimensional  $D = B \times [0, \eta]^{n-m}$ , cujo volume  $n$ -dimensional pode ser

tomado tão pequeno quanto se deseje. Daí resulta que  $\mathbb{R}^m$ , visto como subconjunto de  $\mathbb{R}^n$ , tem medida  $n$ -dimensional nula, pois é reunião enumerável de blocos  $m$ -dimensionais. Em particular, o conjunto  $U \subset \mathbb{R}^m$  tem medida  $n$ -dimensional nula. Isto posto, a partir da aplicação  $f : U \rightarrow \mathbb{R}^n$ , definamos  $F : U \times \mathbb{R}^{n-m} \rightarrow \mathbb{R}^n$  pondo  $F(x, y) = f(x)$ . O conjunto  $U \times 0 \subset U \times \mathbb{R}^{n-m}$  tem medida  $n$ -dimensional nula, logo  $\text{med}.F(U \times 0) = 0$ , pelo Teorema 5. Mas  $F(U \times 0) = f(U)$ , o que prova o corolário.

**Corolário 3.** *Seja  $M \subset \mathbb{R}^n$  uma superfície  $m$ -dimensional de classe  $C^1$ . Se  $m < n$  então  $M$  tem medida  $n$ -dimensional nula.*

Com efeito, para todo  $x \in M$  existe um aberto  $U_x$  em  $\mathbb{R}^n$  tal que  $V_x = U_x \cap M$  é uma vizinhança parametrizada de  $x$ , logo um conjunto de medida nula em  $\mathbb{R}^n$ . A cobertura aberta  $M \subset \bigcup_{x \in M} U_x$  admite, por Lindelöf, uma subcobertura enumerável

$M \subset \bigcup_{k=1}^{\infty} U_k$ , logo  $M = \bigcup_{k=1}^{\infty} (U_k \cap M)$  é reunião enumerável de conjuntos  $V_k = U_k \cap M$ , de medida nula. Assim,  $\text{med}.M = 0$ .

O teorema seguinte, devido a H. Lebesgue, estabelece o critério geral de integrabilidade em termos de noção de conjunto de medida nula. Em sua demonstração, faremos uso do conceito de oscilação de uma função num ponto, que introduziremos agora.

Seja  $f : X \rightarrow \mathbb{R}$  uma função limitada no conjunto  $X \subset \mathbb{R}^n$ . Fixemos  $x \in X$  e, para cada  $\delta > 0$ , ponhamos  $\Omega(\delta) = \omega(f; X \cap B(x; \delta)) =$  oscilação de  $f$  no conjunto dos pontos de  $X$  que distam menos de  $\delta$  do ponto  $x$ . Fica assim definida uma função não-negativa  $\Omega : (0, +\infty) \rightarrow \mathbb{R}$ , a qual é limitada pois  $f$  também é. Além disso,  $\Omega$  é não-decrescente. Logo existe o limite

$$\omega(f; x) = \lim_{\delta \rightarrow 0} \Omega(\delta) = \lim_{\delta \rightarrow 0} \omega(f; X \cap B(x; \delta)) = \inf_{\delta > 0} \omega(f; X \cap B(x; \delta)),$$

que chamaremos a *oscilação da função  $f$  no ponto  $x$* .

Tem-se  $\omega(f; x) = 0$  se, e somente se,  $f$  é contínua no ponto  $x$ . É claro que se  $x \in \text{int}.Y$  e  $Y \subset X$  então  $\omega(f; x) \leq \omega(f; Y)$ .

**Teorema 7 (Lebesgue).** *A função  $f : A \rightarrow \mathbb{R}$ , limitada no bloco  $A \subset \mathbb{R}^n$ , é integrável se, e somente se, o conjunto  $D_f$  dos seus pontos de descontinuidade tem medida nula.*

*Demonstração.* Suponhamos inicialmente que  $\text{med}.D_f = 0$ . Dado arbitrariamente  $\varepsilon > 0$ , seja  $D_f \subset C'_1 \cup \dots \cup C'_k \cup \dots$  uma cobertura enumerável de  $D_f$  por blocos abertos tais que  $\sum \text{vol } C'_k < \varepsilon/2K$ , onde  $K = M - m$  é a diferença entre o sup e o inf de  $f$  em  $A$ . Para cada ponto  $x \in A - D_f$  seja  $C''_x$  um bloco aberto contendo  $x$ , tal que a oscilação de  $f$  no fecho de  $C''_x$  seja inferior a  $\varepsilon/(2 \cdot \text{vol } A)$ . Sendo

A compacto, a cobertura aberta  $A \subset (\cup C'_k) \cup (\cup C''_j)$  admite uma subcobertura finita

$$A \subset C'_1 \cup \dots \cup C'_r \cup C''_1 \cup \dots \cup C''_s.$$

Seja  $P$  uma partição de  $A$  tal que cada bloco aberto  $B \in P$  esteja contido num dos blocos  $C'_k$  ou num  $C''_j$ . Se  $A = \prod_{i=1}^n [a_i, b_i]$  então podemos tomar  $P = P_1 \times \dots \times P_n$  onde, para cada  $i = 1, \dots, n$ ,  $P_i$  é formada pelos pontos  $a_i, b_i$  mais as  $i$ -ésimas coordenadas dos vértices dos blocos  $C'_k$  ou  $C''_j$  que pertençam ao intervalo  $[a_i, b_i]$ . Os blocos de  $P$  contidos em algum  $C'_k$  serão genericamente designados por  $B'$  e os demais blocos de  $P$  (necessariamente contidos em algum  $C''_j$ ) serão chamados  $B''$ . A soma dos volumes dos  $B'$  é menor do que  $\varepsilon/2K$  e, em cada bloco  $B''$ , a oscilação de  $f$  não excede  $\varepsilon/(2 \cdot \text{vol } A)$ . Portanto

$$\begin{aligned} \sum_{B \in P} \omega_B \cdot \text{vol } B &= \sum_{B'} \omega_{B'} \cdot \text{vol } B' + \sum_{B''} \omega_{B''} \cdot \text{vol } B'' \\ &\leq K \cdot \sum \text{vol } B' + \frac{\varepsilon}{2 \cdot \text{vol } A} \sum \text{vol } B'' \\ &< K \cdot \frac{\varepsilon}{2 \cdot K} + \frac{\varepsilon}{2 \cdot \text{vol } A} \cdot \text{vol } A = \varepsilon. \end{aligned}$$

Segue-se que  $f$  é integrável.  $\square$

Reciprocamente, suponhamos  $f$  integrável. Para cada  $k \in \mathbb{N}$ , ponhamos  $D_k = \{x \in A; \omega(f; x) \geq 1/k\}$ , logo  $D_f = D_1 \cup \dots \cup D_k \cup \dots$ . Para mostrar que  $D_f$  tem medida nula, basta provar que  $\text{med}.D_k = 0$  para cada  $k \in \mathbb{N}$ . Seja, então, dado  $\varepsilon > 0$ . Como  $f$  é integrável, existe uma partição  $P$  de  $A$  tal que  $\sum_{B \in P} \omega_B \cdot \text{vol } B < \varepsilon/k$ . Indiquemos genericamente com  $B'$  os blocos da partição  $P$  que contêm algum ponto de  $D_k$  em seu interior. Para cada um desses blocos  $B'$ , vale  $\omega_{B'} \geq 1/k$ . Portanto

$$\frac{1}{k} \sum \text{vol } B' \leq \sum \omega_{B'} \cdot \text{vol } B' \leq \sum_{B \in P} \omega_B \cdot \text{vol } B < \frac{\varepsilon}{k}.$$

Multiplicando por  $k$ , obtemos  $\sum \text{vol } B' < \varepsilon$ . Ora, é claro que  $D_k \subset (\cup B') \cup X$ , onde  $X$  é a reunião das faces próprias dos blocos  $B \in P$  nos quais há algum ponto de  $D_k$ . Sabemos que  $\text{med}.X = 0$ . Segue-se daí que  $\text{med}.D_k = 0$ . Isto completa a demonstração.

### 3 Cálculo com integrais

**Teorema 8.** *Sejam  $f, g : A \rightarrow \mathbb{R}$  funções integráveis no bloco  $A \subset \mathbb{R}^n$  e  $c$  um número real. Então:*

- (1)  $f + g : A \rightarrow \mathbb{R}$  é integrável e  $\int_A [f(x) + g(x)]dx = \int_A f(x)dx + \int_A g(x)dx$ .
- (2)  $c \cdot f : A \rightarrow \mathbb{R}$  é integrável e  $\int_A c \cdot f(x)dx = c \cdot \int_A f(x)dx$ .
- (3) O produto  $f \cdot g : A \rightarrow \mathbb{R}$  é uma função integrável.
- (4) Se  $|g(x)| \geq k > 0$  para todo  $x \in A$  então  $f/g : A \rightarrow \mathbb{R}$  é integrável.
- (5) Se  $f(x) \leq g(x)$  para todo  $x \in A$  então  $\int_A f(x)dx \leq \int_A g(x)dx$ .
- (6)  $|f| : A \rightarrow \mathbb{R}$  é uma função integrável e  $|\int_A f(x)dx| \leq \int_A |f(x)|dx$ .
- (7) Se  $A'$  é um bloco contido em  $A$  e  $f(x) = 0$  para todo  $x \in A - A'$  então  $\int_A f(x)dx = \int_{A'} f(x)dx$ .

*Demonstração.* A integrabilidade das funções  $f + g$ ,  $c \cdot f$ ,  $f \cdot g$ ,  $f/g$  e  $|f|$  resulta do Teorema 7, pois  $D_{f+g} \subset D_f \cup D_g$ ,  $D_{cf} = D_f$  (se  $c \neq 0$ ),  $D_{f \cdot g} \subset D_f \cup D_g$  e  $D_{|f|} \subset D_f$ . Além disso, se  $|g(x)| \geq k > 0$  para todo  $x \in A$  então  $f/g : A \rightarrow \mathbb{R}$  é limitada e, como  $D_{f/g} \subset D_f \cup D_g$ , o quociente  $f/g$  é integrável. As demais afirmações do Teorema 8 se provam exatamente como no caso de funções de uma única variável. (Ver Capítulo 10 do Volume 1.)

O cálculo efetivo da integral de uma função  $f : A \rightarrow \mathbb{R}$ , definida num bloco  $n$ -dimensional, se faz integrando  $f$  sucessivamente em relação a cada uma das suas  $n$  variáveis. Basta aplicar diversas vezes o Teorema 9 abaixo, no qual adotamos as seguintes notações:

Dados os blocos  $A_1 \subset \mathbb{R}^m$  e  $A_2 \subset \mathbb{R}^n$ , os pontos do bloco  $A_1 \times A_2 \subset \mathbb{R}^{m+n}$  escrevem-se como  $(x, y)$ , com  $x \in A_1$  e  $y \in A_2$ . Se  $f : A_1 \times A_2 \rightarrow \mathbb{R}$  é integrável, sua integral é indicada com  $\int_{A_1 \times A_2} f(x, y)dx dy$ . Para cada  $x \in A_1$ , definiremos a função  $f_x : A_2 \rightarrow \mathbb{R}$  pondo  $f_x(y) = f(x, y)$  para todo  $y \in A_2$ , portanto  $f_x$  é essencialmente a restrição de  $f$  ao bloco  $n$ -dimensional  $x \times A_2$ . Mesmo que  $f$  seja integrável, pode ocorrer que, para alguns valores de  $x \in A_1$ , a função  $f_x : A_2 \rightarrow \mathbb{R}$  não o seja. Com efeito, os pontos em que  $f$  é descontínua formam um conjunto  $D$  de medida nula em  $\mathbb{R}^{m+n}$  mas pode existir  $x \in A_1$  tal que  $D \cap (x \times A_2)$  não tenha medida  $n$ -dimensional nula.  $\square$

**Exemplo 1.** Sejam  $A_1 = A_2 = [0, 1]$  e  $f : [0, 1] \times [0, 1] \rightarrow \mathbb{R}$  dada por  $f(x, y) = 0$  se  $x \neq 1/2$ ,  $f(1/2, y) = 0$  se  $y$  é racional,  $f(1/2, y) = 1$  se  $y$  é irracional. O conjunto dos pontos de descontinuidade de  $f$  é  $D_f = 1/2 \times [0, 1]$ , que tem medida nula em  $[0, 1] \times [0, 1]$ , logo  $f$  é integrável. (De fato, sua integral é zero.) Mas  $f_{1/2} : [0, 1] \rightarrow \mathbb{R}$  é a função igual a zero nos pontos racionais e igual a 1 nos irracionais, logo  $f_x : [0, 1] \rightarrow \mathbb{R}$  não é integrável quando  $x = 1/2$ .  $\triangleleft$

**Teorema 9 (Integração repetida).** *Seja  $f : A_1 \times A_2 \rightarrow \mathbb{R}$  integrável no produto dos blocos  $A_1 \subset \mathbb{R}^m$  e  $A_2 \subset \mathbb{R}^n$ . Para todo  $x \in A_1$ , seja  $f_x : A_2 \rightarrow \mathbb{R}$  definida por  $f_x(y) = f(x, y)$ . Ponhamos*

$$\varphi(x) = \int_{\underline{A_2}} f_x(y) dy \quad e \quad \psi(x) = \overline{\int_{A_2} f_x(y) dy}.$$

As funções  $\varphi, \psi : A_1 \rightarrow \mathbb{R}$ , assim definidas, são integráveis, com

$$\int_{A_1} \varphi(x) dx = \int_{A_1} \psi(x) dx = \int_{A_1 \times A_2} f(x, y) dx dy,$$

isto é:

$$\int_{A_1 \times A_2} f(x, y) dx dy = \int_{A_1} dx \left( \int_{\underline{A_2}} f(x, y) dy \right) = \int_{A_1} dx \left( \overline{\int_{A_2} f(x, y) dy} \right).$$

*Demonstração.* As partições do bloco  $A_1 \times A_2$  são da forma  $P = P_1 \times P_2$ , onde  $P_1$  e  $P_2$  são partições dos blocos  $A_1$  e  $A_2$  respectivamente. Os blocos de  $P$  são os produtos  $B_1 \times B_2$  com  $B_1 \in P_1$  e  $B_2 \in P_2$ . Mostraremos que

$$s(f; P) \leq s(\varphi; P_1) \leq S(\varphi; P_1) \leq S(f; P).$$

Daí resultará que  $\varphi$  é integrável e que  $\int_{A_1} \varphi(x) dx = \int_{A_1 \times A_2} f(x, y) dx dy$ . Na verdade, basta provar a primeira das desigualdades acima, pois a segunda é óbvia e a terceira é análoga. Também por analogia, não precisamos provar que  $\int_{A_1} \psi(x) dx = \int_{A_1 \times A_2} f(x, y) dx dy$ .  $\square$

Começamos lembrando que se  $X \subset Y \subset \mathbb{R}$  então  $\inf Y \leq \inf X$ . Segue-se que, para todo bloco  $B_1 \times B_2 \in P_1$  tem-se  $m(f; B_1 \times B_2) \leq m(f_x; B_2)$ , seja qual for  $x \in B_1$ . Portanto

$$\sum_{B_1 \in P_2} m(f; B_1 \times B_2) \cdot \text{vol } B_2 \leq \sum_{B_2 \in P_2} m(f_x; B_2) \cdot \text{vol } B_2 \leq \varphi(x).$$

Como isto vale para todo  $x \in B_1$ , concluímos que:

$$\sum_{B_2 \in P_2} m(f; B_1 \times B_2) \leq m(\varphi; B_1).$$

Portanto

$$\begin{aligned} s(f; P) &= \sum_{B_1 \times B_2 \in P} m(f; B_1 \times B_2) \cdot \text{vol } B_1 \times \text{vol } B_2 \\ &= \sum_{B_1 \in P_1} \left( \sum_{B_2 \in P_2} m(f; B_1 \times B_2) \cdot \text{vol } B_2 \right) \cdot \text{vol } B_1 \\ &\leq \sum_{B_1 \in P_1} m(\varphi; B_1) \cdot \text{vol } B_1 = s(\varphi; P_1). \end{aligned}$$

**Corolário 4.** *Seja  $f : A_1 \times A_2 \times A_2 \rightarrow \mathbb{R}$  integrável no produto dos blocos  $A_1 \subset \mathbb{R}^m$ ,  $A_2 \subset \mathbb{R}^n$  e  $A_3 \subset \mathbb{R}^p$ . Então*

$$\begin{aligned} \int_{A_1 \times A_2 \times A_3} f(x, y, z) dx dy dz &= \int_{A_1} dx \int_{\overline{A_2}} dy \int_{\overline{A_3}} f(x, y, z) dz \\ &= \int_{A_1} dx \int_{\overline{A_2}} dy \int_{\overline{A_3}} f(x, y, z) dz. \end{aligned}$$

Com efeito,

$$\begin{aligned} \int_{A_1 \times A_2 \times A_3} f(x, y, z) dx dy dz &= \int_{A_1 \times A_2} dx dy \left( \int_{\overline{A_3}} f(x, y, z) dz \right) \\ &= \int_{A_1} dx \left( \int_{\overline{A_2}} dy \left( \int_{\overline{A_3}} f(x, y, z) dz \right) \right) \\ &= \int_{A_1} dx \int_{\overline{A_2}} dy \int_{\overline{A_3}} f(x, y, z) dz. \end{aligned}$$

A seguir, vamos estender o conceito de integral para funções definidas em certos subconjuntos  $X \subset \mathbb{R}^n$  que não são necessariamente blocos  $n$ -dimensionais.

#### 4 Conjuntos $J$ -mensuráveis

Dado o conjunto limitado  $X \subset \mathbb{R}^n$ , seja  $A$  um bloco  $n$ -dimensional contendo  $X$ . A *função característica* de  $X$  é a função  $\xi_X : A \rightarrow \mathbb{R}$ , definida por  $\xi_X(x) = 1$  se  $x \in X$  e  $\xi_X(x) = 0$  se  $x \notin X$ .

Se  $X$  e  $Y$  são subconjuntos do bloco  $A$ , as seguintes propriedades da função característica são evidentes:

1.  $\xi_{X \cup Y} = \xi_X + \xi_Y - \xi_{X \cap Y}$ ;
2.  $\xi_{X \cap Y} = \xi_X \cdot \xi_Y$ ;
3. Tem-se  $X \subset Y$  se, e somente se,  $\xi_X \leq \xi_Y$ ; neste caso, vale  $\xi_{Y-X} = \xi_Y - \xi_X$ .

Segue-se de 1. que  $\xi_{X \cup Y} = \xi_X + \xi_Y$  quando  $X$  e  $Y$  são disjuntos.

Se  $X$  estiver contido no interior de  $A$  (o que poderemos supor, sempre que for conveniente) então  $fr.X$  é o conjunto dos pontos de descontinuidade da função  $\xi_X : A \rightarrow \mathbb{R}$ .

O *volume interno* e o *volume externo* do conjunto limitado  $X \subset \mathbb{R}^n$  são definidos, respectivamente, pondo:

$$\text{vol.int. } X = \int_{\overline{A}} \xi_X(x) dx \quad \text{e} \quad \text{vol.ext. } X = \int_A \xi_X(x) dx.$$

Quando a função característica  $\xi_X : A \rightarrow \mathbb{R}$  é integrável, dizemos que  $X$  é  $J$ -mensurável (mensurável segundo Jordan) e que seu *volume  $n$ -dimensional* é

$$\text{vol } X = \int_A \xi_X(x) dx .$$

O item (7) do Teorema 9 assegura que os conceitos acima introduzidos não dependem da escolha do bloco  $A$  contendo  $X$ .

Se  $X \subset A$  e  $P$  é uma partição do bloco  $A$ , as somas inferior e superior da função  $\xi_X : A \rightarrow \mathbb{R}$  relativas à partição  $P$  são

$$\begin{aligned} s(\xi_X; P) &= \text{soma dos volumes dos blocos de } P \text{ contidos em } X; \\ S(\xi_X; P) &= \text{soma dos volumes dos blocos de } P \text{ que intersectam } X. \end{aligned}$$

Portanto, se escrevermos  $v = \text{vol.int. } X$  e  $V = \text{vol.ext. } X$ , veremos que, para todo  $\varepsilon > 0$  dado, existe uma partição  $P$  do bloco  $A$  (o qual contém  $X$ ) tal que a soma dos volumes dos blocos de  $P$  contidos em  $X$  é superior a  $v - \varepsilon$  e a soma dos volumes dos blocos de  $P$  que intersectam  $X$  é inferior a  $V + \varepsilon$ .

**Teorema 10.** (1) *O conjunto limitado  $X \subset \mathbb{R}^n$  é  $J$ -mensurável se, e somente se, sua fronteira tem medida nula.*

(2) *Se  $X, Y \subset \mathbb{R}^n$  são  $J$ -mensuráveis então  $X \cup Y, X \cap Y$  e  $X - Y$  são  $J$ -mensuráveis, com*

$$\begin{aligned} \text{vol}(X \cup Y) &= \text{vol } X + \text{vol } Y - \text{vol}(X \cap Y) \\ \text{e } \text{vol}(X - Y) &= \text{vol } X - \text{vol } Y \quad \text{quando } Y \subset X. \end{aligned}$$

*Demonstração.* (1) Tomando um bloco  $n$ -dimensional  $A$  que contenha  $X$  em seu interior e considerando a função característica  $\xi_X : A \rightarrow \mathbb{R}$ , temos as equivalências:

$$X \text{ é } J\text{-mensurável} \Leftrightarrow \xi_X \text{ é integrável} \Leftrightarrow \text{med. } D_{\xi_X} = 0 \Leftrightarrow \text{med.fr. } X = 0,$$

pois o conjunto  $D_{\xi_X}$  das descontinuidades de  $\xi_X$  coincide com a fronteira de  $X$ .

(2) Basta observar que  $\xi_{X \cup Y} = \xi_X + \xi_Y - \xi_{X \cap Y}$  e que, quando  $Y \subset X$ , vale ainda  $\xi_{X - Y} = \xi_X - \xi_Y$ .  $\square$

**Exemplo 2.** Todo conjunto limitado  $X \subset \mathbb{R}^n$ , cuja fronteira é uma superfície, ou a reunião de um número finito (ou mesmo enumerável) de superfícies, de dimensão  $n - 1$  é  $J$ -mensurável. Isto inclui uma bola fechada e a região compreendida entre duas bolas fechadas concêntricas. Resulta ainda do item (1) acima que um bloco  $n$ -dimensional é  $J$ -mensurável.  $\triangleleft$

**Exemplo 3.** Seja  $X \subset \mathbb{R}$  o conjunto formado pelo intervalo  $[0, 1]$  mais os números racionais do intervalo  $[1, 2]$ . O “volume” interno do conjunto  $X$  é igual a 1 enquanto seu “volume” externo é 2. Portanto  $X$  não é  $J$ -mensurável. Tomando o produto cartesiano de  $n$  cópias de  $X$ , obtém-se um subconjunto de  $\mathbb{R}^n$  que não é  $J$ -mensurável.  $\triangleleft$

**Exemplo 4.** Se  $X \subset \mathbb{R}^n$  é  $J$ -mensurável e  $\text{int.}X = \emptyset$  então  $\text{vol } X = 0$  pois  $s(\xi_X; P) = 0$  para toda partição  $P$  de um bloco que contenha  $X$ . Resulta daí que se  $X$  e  $Y$  são conjuntos  $J$ -mensuráveis sem pontos interiores em comum então  $\text{vol}(X \cup Y) = \text{vol } X + \text{vol } Y$ , pois  $\text{vol}(X \cap Y) = 0$ .  $\triangleleft$

Definiremos agora a integral  $\int_X f(x)dx$  de uma função limitada  $f : X \rightarrow \mathbb{R}$ , cujo domínio é um conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$ . Para isto, consideramos um bloco  $n$ -dimensional  $A$  contendo  $X$  em seu interior e a função

$$\bar{f} : A \rightarrow \mathbb{R},$$

definida por  $\bar{f}(x) = f(x)$  se  $x \in X$  e  $\bar{f}(x) = 0$  se  $x \in A - X$ . Podemos então, por definição

$$\int_{\underline{X}} f(x)dx = \int_{\underline{A}} \bar{f}(x)dx \quad \text{e} \quad \int_{\overline{X}} f(x)dx = \int_{\overline{A}} \bar{f}(x)dx.$$

Diremos que  $f : X \rightarrow \mathbb{R}$  é *integrável* quando tivermos

$$\int_{\underline{X}} f(x)dx = \int_{\overline{X}} f(x)dx$$

ou seja, quando  $\bar{f} : A \rightarrow \mathbb{R}$  for integrável.

Se  $\bar{f} : A \rightarrow \mathbb{R}$  é descontínua num ponto  $x \in A$ , ou  $f$  é descontínua no ponto  $x$  ou  $x$  pertence à fronteira de  $X$ . Noutros termos,  $D_{\bar{f}} \subset D_f \cup \text{fr.}X$ . Como  $\text{fr.}X$  tem medida nula, segue-se que  $f$  é integrável (ou seja,  $\bar{f}$  é integrável) se, e somente se o conjunto  $D_f$  dos seus pontos de descontinuidade tem medida nula.

Valem, evidentemente, para a integral  $\int_X f(x)dx$  as mesmas regras operatórias estabelecidas no Teorema 9 para o caso em que  $X$  é um bloco retangular.

## 5 A integral como limite de somas de Riemann

Mostraremos agora (veja o Teorema 12) que a integral  $\int_X f(x)dx$  é o número real cujos valores aproximados são as “somas de Riemann”  $\sum f(\xi_i)\text{vol } X_i$ , obtidas quando se faz uma decomposição do tipo  $X = X_1 \cup \dots \cup X_k$ , onde os  $X_i$  são

conjuntos  $J$ -mensuráveis, dois a dois sem pontos interiores em comum, tomando-se arbitrariamente  $\xi_i \in X_i$  para cada  $i = 1, \dots, k$ . Esta é a forma mais comu, e a mais intuitiva, de se pensar na integral. Passemos às definições precisas.

Seja  $X \subset \mathbb{R}^n$  um conjunto  $J$ -mensurável. Diz-se que  $D = (X_1, \dots, X_k)$  é uma *decomposição* de  $X$  quando os conjuntos  $X_1, \dots, X_k$  são  $J$ -mensuráveis, sem pontos interiores em comum (isto é,  $X_i \cap X_j \subset \text{fr}.X_i \cap \text{fr}.X_j$  quando  $i \neq j$ ), com  $X = X_1 \cup \dots \cup X_k$ . A *norma* da decomposição  $D$  é o número  $|D| = \max.\text{diam}.X_i =$  maior diâmetro dos conjuntos  $X_1, \dots, X_k$ .

Por exemplo, se  $X \subset \mathbb{R}^n$  é um bloco  $n$ -dimensional, toda partição  $P$  determina uma decomposição  $X = B_1 \cup \dots \cup B_k$ , onde os  $B_k$ , onde os  $B_i$  são os blocos da partição  $P$ .

Seja  $f : X \rightarrow \mathbb{R}$  uma função limitada no conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$ . Dada a decomposição  $D = (X_1, \dots, X_k)$  de  $X$  escreveremos, para cada  $i = 1, \dots, k$ ,  $m_i = \inf.\{f(x); x \in X_i\}$  e  $M_i = \sup.\{f(x); x \in X_i\}$ . Definiremos então a *soma inferior*  $s(f; D)$  e a *soma superior*  $S(f; D)$  pondo

$$s(f; D) = \sum_{i=1}^k m_i \cdot \text{vol } X_i \quad \text{e} \quad S(f; D) = \sum_{i=1}^k M_i \cdot \text{vol } X_i.$$

Diz-se que o número real  $J$  é o limite de  $S(f; D)$  quando  $|D|$  tende a zero, e escreve-se

$$J = \lim_{|D| \rightarrow 0} S(f; D)$$

para significar que, para todo  $\varepsilon > 0$  dado, existe  $\delta > 0$  tal que  $|D| < \delta \Rightarrow |J - S(f; D)| < \varepsilon$ . Analogamente se define o significado da afirmação  $I = \lim_{|D| \rightarrow 0} s(f; D)$ .

**Teorema 11.** Para toda função  $f : X \rightarrow \mathbb{R}$ , limitada no conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$ , tem-se

$$\int_{\underline{X}} f(x) dx = \lim_{|D| \rightarrow 0} s(f; D) \quad \text{e} \quad \int_{\overline{X}} f(x) dx = \lim_{|D| \rightarrow 0} S(f; D).$$

Na demonstração do Teorema 11 usaremos o lema abaixo, cujo enunciado contém a desigualdade  $d(X_i, Y) < \delta$ .

Se  $A$  e  $B$  são subconjuntos não-vazios de  $\mathbb{R}^n$ , costuma-se escrever  $d(A, B) = \inf.\{|x - y|; x \in A, y \in B\}$ . Por conseguinte, a desigualdade  $d(A, B) < \delta$  significa que existem  $x \in A$  e  $y \in B$  com  $|x - y| < \delta$ .

**Lema 1.** Sejam  $Y \subset X \subset \mathbb{R}^n$   $J$ -mensuráveis, com  $\text{vol } Y = 0$ . Para todo  $\varepsilon > 0$  dado, existe  $\delta > 0$  tal que, se  $D$  é uma decomposição de  $X$  com  $|D| < \delta$  então a soma dos volumes dos conjuntos  $X_i \in D$  tais que  $d(X_i, Y) < \delta$  é menor do que  $\varepsilon$ .

*Demonstração.* Dado  $\varepsilon > 0$ , podemos cobrir  $Y$  com uma coleção finita de blocos  $B$  cuja soma dos volumes é  $< \varepsilon$ . Tomando arbitrariamente  $\delta > 0$ , ponhamos cada um desses blocos  $B = \Pi [a_i, b_i]$  dentro do bloco  $B' = \Pi [a_i - 2\delta, b_i + 2\delta]$ . Como  $\lim_{\delta \rightarrow 0} \text{vol } B' = \text{vol } B$ , existe  $\delta > 0$  tal que a soma dos volumes dos blocos  $B'$  é ainda menor do que  $\varepsilon$ . Usando a norma do máximo, podemos assegurar que se  $Z$  é um conjunto de diâmetro  $< \delta$  tal que  $d(Z, B) < \delta$  então  $Z \subset B'$ . Portanto, se  $D = (X_1, \dots, X_k)$  é uma decomposição de  $X$  com  $|D| < \delta$ , vemos que

$$d(X_i, Y) < \delta \Rightarrow d(X_i, B) < \delta \text{ para algum } B \Rightarrow X_i \subset B'.$$

Assim, a soma dos volumes dos conjuntos  $X_i \in D$  tais que  $d(X_i, Y) < \delta$  não excede a soma dos volumes dos blocos  $B'$ , logo é menor do que  $\varepsilon$ .  $\square$

*Demonstração do Teorema 11.* Basta provar a segunda afirmação. Sem perda de generalidade, podemos admitir que  $0 \leq f(x) \leq K$  para todo  $x \in X$ . Com efeito, se somarmos uma constante  $c$  à função  $c$ , tanto a integral superior como o limite acima serão aumentados de  $c \cdot \text{vol } X$ . Seja  $\bar{f} : A \rightarrow \mathbb{R}$  a extensão de  $f$  a um bloco  $n$ -dimensional  $A \supset X$ , com  $\bar{f}(x) = 0$  se  $x \in A - X$ . Dado  $\varepsilon > 0$ , queremos achar  $\delta > 0$  tal que  $|S(f; D) - \int_X f(x) dx| < \varepsilon$  para toda decomposição  $D$  de  $X$  com  $|D| < \delta$ . Ora, dado  $\varepsilon > 0$ , existe uma partição  $P_o$  de  $A$  tal que

$$S(\bar{f}; P_o) < \int_X f(x) dx + \varepsilon/2.$$

Seja  $Y$  a reunião das faces próprias dos blocos de  $P_o$ . Como  $\text{vol } Y = 0$ , o Lema assegura a existência de  $\delta > 0$  tal que, para toda decomposição  $D$  de  $X$  com  $|D| < \delta$ , a soma dos volumes dos conjuntos  $X_i \in D$  com  $d(X_i, Y) < \delta$  é menor do que  $\varepsilon/2K$ .

Seja então  $D$  uma decomposição de  $X$  com norma  $|D| < \delta$ . Chamemos de  $X_\alpha$  os conjuntos de  $D$  tais que  $d(X_\alpha, Y) < \delta$ . Os demais conjuntos de  $D$  serão chamados de  $X_\beta$ . Notemos que cada  $X_\beta$  deve estar contido em algum bloco da partição  $P_o$  pois, do contrário, existiriam  $x, y \in X_\beta$  em blocos distintos de  $P_o$ , logo o segmento de reta  $[x, y]$  conteria algum ponto de  $Y$ . Como  $|x - y| < \delta$ , isto daria  $d(X_\beta, Y) < \delta$  um absurdo. Escrevendo  $M_\alpha = \sup\{f(x); x \in X_\alpha\}$  e  $M_\beta = \sup\{f(x); x \in X_\beta\}$ , vem

$$\begin{aligned} S(f; D) &= \sum M_\alpha \cdot \text{vol } X_\alpha + \sum M_\beta \cdot \text{vol } X_\beta, \quad \text{onde} \\ \sum M_\alpha \cdot \text{vol } X_\alpha &\leq K \cdot \sum \text{vol } X_\alpha < \varepsilon/2 \quad \text{e} \\ \sum M_\beta \cdot \text{vol } X_\beta &= \sum_{B \in P_o} \left( \sum_{X_\beta \subset B} M_\beta \cdot \text{vol } X_\beta \right) \leq \sum_{B \in P_o} M_\beta \cdot \text{vol } B \end{aligned}$$

$$= S(\bar{f}; P_o) < \overline{\int}_X f(x)dx + \varepsilon/2.$$

Assim,  $|D| < \delta \Rightarrow S(f; D) < \overline{\int}_X f(x)dx + \varepsilon/2$ .

Mostraremos agora que  $S(f; D) \geq \overline{\int}_X f(x)dx$  para toda decomposição  $D$  de  $X$ . Com efeito, seja  $Z$  a reunião das fronteiras dos conjuntos  $X_i$  da decomposição  $D$ . Como  $\text{vol } Z = 0$ , o Lema nos dá  $\delta' > 0$  tal que, para toda partição  $P$  do bloco  $A$  com  $|P| < \delta'$ , a soma dos volumes dos blocos de  $P$  que intersectam  $Z$  é menor do que  $\varepsilon/K$ . Tomando  $|P| < \delta'$ , temos

$$S(\bar{f}; P) = \sum M_B \cdot \text{vol } B + \sum M_C \cdot \text{vol } C,$$

onde chamamos de  $B$  os blocos de  $P$  que intersectam  $Z$  e de  $C$  os que estão contidos no interior de algum  $X_i \in D$ . (Observe que, pelo Teorema da Alfândega, um bloco que não esteja contido no interior de algum  $X_i$ , deve intersectar  $Z$ , pois todo bloco é conexo.) Ora, temos

$$\begin{aligned} \sum M_B \cdot \text{vol } B &\leq K \cdot \sum \text{vol } B < \varepsilon \quad \text{e} \\ \sum M_C \cdot \text{vol } C &= \sum_i \left( \sum_{C \subset X_i} M_C \cdot \text{vol } C \right) \leq \sum_i M_i \left( \sum_{C \subset X_i} \text{vol } C \right) \\ &\leq \sum_i M_i \cdot \text{vol } X_i = S(f; D). \end{aligned}$$

Logo  $\overline{\int}_X f(x)dx \leq S(\bar{f}; P) \leq S(f; D) + \varepsilon$ .

Como  $\varepsilon > 0$  é arbitrário, concluímos que  $\overline{\int}_X f(x)dx \leq S(f; D)$  para toda decomposição  $D$  de  $X$ . Isto conclui a demonstração.

**Corolário 5 (da demonstração).** Para toda função limitada  $f : X \rightarrow \mathbb{R}$  no conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$ , tem-se

$$\underline{\int}_X f(x)dx = \sup_D s(f; D) \quad \text{e} \quad \overline{\int}_X f(x)dx = \inf_D S(f; D).$$

Uma *decomposição pontilhada* do conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$  é um par  $D^* = (D, \xi)$ , onde  $D = (X_1, \dots, X_k)$  é uma decomposição de  $X$  e  $\xi = (\xi_1, \dots, \xi_k)$ , com  $\xi_i \in X_i, \dots, \xi_k \in X_k$ . Em termos menos formais, *pontilhar* a decomposição  $D = (X_1, \dots, X_k)$  é escolher um ponto  $\xi_i$  em cada conjunto  $X_i, i = 1, \dots, k$ .

A toda partição pontilhada  $D^*$  fica associada a *soma de Riemann*  $\Sigma (f; D^*)$  definida por

$$\Sigma (f; D^*) = \sum_{i=1}^k f(\xi_i) \cdot \text{vol } X_i .$$

Diz-se que o número  $I$  é o *limite* das somas de Riemann  $\Sigma (f; D^*)$  quando a norma  $|D|$  tende a zero, e escreve-se

$$I = \lim_{|D| \rightarrow 0} \Sigma (f; D^*) ,$$

quando, para todo  $\varepsilon > 0$  dado, pode-se obter  $\delta > 0$  tal que, para toda decomposição  $D$  do conjunto  $X$  com norma  $|D| < \delta$  tem-se  $|\Sigma (f; D^*) - I| < \varepsilon$ , seja qual for a maneira  $D^*$  de se pontilhar  $D$ .

**Teorema 12.** *Se  $f : X \rightarrow \mathbb{R}$  é integrável no conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$  então*

$$\int_X f(x) dx = \lim_{|D| \rightarrow 0} \Sigma (f; D^*)$$

*Demonstração.* Para toda decomposição  $D$  de  $X$  tem-se

$$s(f; D) \leq \Sigma (f; D^*) \leq S(f; D) ,$$

seja qual for o modo  $D^*$  de pontilhar  $D$ . Pelo Teorema 11, temos  $\lim_{|D| \rightarrow 0} S(f; D) = \int_X f(x) dx$ .

Segue-se imediatamente que  $\lim_{|D| \rightarrow 0} \Sigma (f; D^*) = \int_X f(x) dx$ . □

## Capítulo 9

# Mudança de Variáveis

Demonstraremos neste capítulo o importante Teorema da Mudança de Variáveis em integrais múltiplas.

Começaremos estabelecendo as notações.

$U$  e  $V$  são abertos do espaço euclidiano  $\mathbb{R}^n$ ;  $h : U \rightarrow V$  é um difeomorfismo de classe  $C^1$ .

$X$  é um subconjunto compacto  $J$ -mensurável de  $U$ . A fronteira de  $X$ , que tem medida nula, está contida em  $X$  (logo em  $U$ ) e sua imagem por  $h$ , que é a fronteira do compacto  $h(X)$ , tem medida nula (Teorema 6, Capítulo 8). Portanto  $h(X)$  também é um conjunto  $J$ -mensurável.

Finalmente,  $f : h(X) \rightarrow \mathbb{R}$  é uma função integrável.

O Teorema da Mudança de Variáveis diz que a seguinte igualdade é verdadeira:

$$\int_{h(X)} f(y)dy = \int_X f(h(x)) \cdot |\det h'(x)|dx .$$

Ela é análoga para  $n$  variáveis daquela estabelecida no Vol. 1. (Vide Teorema 2, Capítulo 11.) Notam-se, porém, algumas diferenças.

A função que, no caso de uma só variável, desempenhava o papel de  $h$  não precisava ser um difeomorfismo. Para  $n > 1$ , entretanto, pelo menos injetividade de  $h$  (ou algo equivalente) se faz necessário, sem o que a fórmula não vale. (O estudo dessas situações gerais leva à noção de *grau*, que é analisada em detalhe no livro “Curso de Análise”, Vol. 2.)

Outra diferença é o valor absoluto em  $|\det h'(x)|$ . É natural que o determinante substitua a derivada  $h'(x)$  pois, quando  $n > 1$ , esta não é um número; mas o valor absoluto que ocorre na fórmula acima não parece estar presente quando  $n = 1$ . Na verdade, porém, ele está oculto na igualdade

$$\int_{h(a)}^{h(b)} f(y)dy = \int_a^b f(h(x)) \cdot h'(x)dx .$$

De fato, se chamarmos de  $I$  o intervalo  $[a, b]$  e  $J = h(I)$  o intervalo cujos extremos são  $h(a)$  e  $h(b)$ , teremos  $h(a) > h(b)$  quando  $h' < 0$ , logo a fórmula acima significa, em qualquer caso, que

$$\int_J f(y)dy = \int_I f(h(x)) \cdot |h'(x)|dx,$$

pois

$$\int_J f(y)dy = \int_{h(a)}^{h(b)} f(y)dy \quad \text{se } h(a) < h(b),$$

isto é,

$$h'(x) > 0, \quad \text{e} \quad \int_J f(y)dy = \int_{h(b)}^{h(a)} f(y)dy$$

se  $h(b) < h(a)$ , isto é, se  $h'(x) < 0$ .

O Teorema de Mudança de Variáveis será provado por etapas.

## 1 O caso unidimensional

Dado o intervalo  $I = [a, b]$ , escreveremos  $|I| = b - a$ .

**Teorema 1.** *Sejam  $U, V \subset \mathbb{R}$  abertos,  $h : U \rightarrow V$  em difeomorfismo  $C^1$ ,  $I \subset U$  um intervalo compacto,  $J = f(I)$  e  $f : J \rightarrow \mathbb{R}$  uma função limitada. Então*

$$\overline{\int}_J f(y)dy = \overline{\int}_I f(h(x)) \cdot |h'(x)|dx.$$

*Demonstração.* Sem perda de generalidade, podemos admitir que  $f(y) \geq 0$  para todo  $y \in J$  pois, se somarmos a mesma constante positiva  $c$  a ambos os membros da igualdade acima, o lado esquerdo sofrerá o acréscimo de  $c \cdot |J|$  enquanto o acréscimo sofrido pelo lado direito será de  $c \cdot \int_I |h'(x)|dx$ . Como  $h'(x)$  não muda de sinal para  $x \in J$ , o valor desta integral é  $c \cdot |h(b) - h(a)| = c \cdot |J|$  também. Esta observação nos deixa livres para manipular desigualdades.  $\square$

As partições de  $J = h(I)$  são do tipo  $h(P)$ , dadas por intervalos da forma  $J_r = h(I_r)$ , onde os  $I_r$  ( $r = 1, \dots, k$ ) são os intervalos de uma partição  $P$  de  $I$ . Para cada  $r$ , ponhamos  $M_r = \sup_{y \in J_r} f(y) = \sup_{x \in I_r} f(h(x))$  e  $c_r = \sup_{x \in I_r} |h'(x)|$ .

Pelo Teorema do Valor Médio, para cada  $r = 1, \dots, k$  existe  $\xi_r \in I_r$  tal que  $|J_r| = |h'(\xi_r)| \cdot |I_r|$ . Pondo  $\eta_r = c_r - |h'(\xi_r)|$ , temos  $\eta_r \geq 0$ , em virtude da continuidade uniforme de  $h'$  no intervalo  $I$ ,  $\lim_{|P| \rightarrow 0} \eta_r = 0$ . Segue-se que

$$\lim_{|P| \rightarrow 0} \sum_{r=1}^k \eta_r |I_r| = 0 \text{ pois}$$

$$0 \leq \sum_{r=1}^k \eta_r |I_r| \leq (\max_r \eta_r) \cdot \sum_{r=1}^k |I_r| \leq (\max_r \eta_r) \cdot |I|.$$

Então

$$\begin{aligned} S(f; h(P)) &= \sum_{r=1}^k M_r \cdot |J_r| = \sum_{r=1}^k M_r c_r |I_r| - \sum_{r=1}^k \eta_r |I_r| \quad \text{e} \\ S((f \circ h) \cdot |h'|; P) &= \sum_{r=1}^k N_r |I_r|, \end{aligned}$$

pois se  $\varphi, \psi : A \rightarrow \mathbb{R}$  são duas funções não-negativas limitadas quaisquer então  $\sup_{x \in A} (\varphi(x), \psi(x)) \leq \sup_{x \in A} \varphi(x) \cdot \sup_{x \in A} \psi(x)$ . Logo, para toda partição  $P$  do intervalo  $I$ , vale:

$$S((f \circ h) \cdot |h'|; P) \leq S(f; h(P)) + \sum_{r=1}^k \eta_r \cdot |I_r|.$$

Segue-se que

$$\begin{aligned} \overline{\int}_I f(h(x)) \cdot |h'(x)| \cdot dx &= \lim_{|P| \rightarrow 0} S((f \circ h) \cdot |h'|; P) \\ &\leq \lim_{|P| \rightarrow 0} S(f; h(P)) = \overline{\int}_J f(y) dy. \end{aligned}$$

A desigualdade oposta,  $\overline{\int}_J f(y) dy \leq \overline{\int}_I f(h(x)) \cdot |h'(x)| dx$ , resulta da anterior, que vem de ser provada, usando-se  $h^{-1} : J \rightarrow I$  em vez de  $h$ ,  $f \circ h : I \rightarrow \mathbb{R}$  em vez de  $f$  e levando em conta que, para todo  $y = h(x)$ ,  $x \in I$ , tem-se  $(h^{-1})'(y) = 1/h'(x)$ .

Então concluímos que

$$\overline{\int}_J f(y) dy = \overline{\int}_I f(h(x)) \cdot |h'(x)| dx.$$

## 2 Difeomorfismos primitivos

O próximo caso particular que consideraremos é aquele em que  $h$  é um difeomorfismo *primitivo*.

Chamam-se primitivos os difeomorfismos  $h$  de um dos dois tipos seguintes:

*Tipo 1.* São fixados os índices  $i, j$ , com  $1 \leq i < j \leq n$  e  $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$  é dado por

$$h(x) = h(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = (x_1, \dots, x_j, \dots, x_i, \dots, x_n).$$

*Tipo 2.* Tem-se uma função  $\varphi : U \rightarrow \mathbb{R}$ , de classe  $C^1$ , e para todo  $x \in U$  vale

$$h(x) = (\varphi(x), x_2, \dots, x_n).$$

**Teorema 2.** *Seja  $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$  um difeomorfismo primitivo do Tipo 1. Para todo conjunto  $J$ -mensurável  $X \subset \mathbb{R}^n$  e toda função integrável  $f : h(X) \rightarrow \mathbb{R}$  tem-se*

$$\int_{h(X)} f(y) dy = \int_X f(h(x)) \cdot |\det h'(x)| dx.$$

*Demonstração.* O difeomorfismo  $h$  é um operador linear, com  $\det h'(h) = -1$  para todo  $x \in \mathbb{R}^n$ , logo  $|\det h'(x)| = 1$ . Devemos, portanto, mostrar que  $\int_{h(X)} f(y) dy = \int_X f(h(x)) dx$ . Ora, para todo bloco  $B \subset \mathbb{R}^n$ , sua imagem  $h(B)$  é também um bloco, com arestas de mesmo comprimento que as de  $B$ , logo  $\text{vol } h(B) = \text{vol } B$ . Como o volume de um conjunto  $J$ -mensurável  $Z \subset \mathbb{R}^n$  é o ínfimo dos números  $\sum \text{vol } B_i$ , onde os  $B_i = \text{vol } Z$ . Toda decomposição  $h(X) = Y_1 \cup \dots \cup Y_k$  é tal que  $Y_i = h(X_i)$ , onde  $X = X_1 \cup \dots \cup X_k$  é uma decomposição de  $X$ . Todo ponto de  $Y_i$  é da forma  $h(\xi_i)$ , com  $\xi_i \in X_i$ . Logo

$$\begin{aligned} \int_{h(X)} f(y) dy &= \lim_{|D| \rightarrow 0} \sum f(h(\xi_i)) \cdot \text{vol } Y_i \\ &= \lim_{|D| \rightarrow 0} \sum f(h(\xi_i)) \cdot \text{vol } X_i = \int_X f(h(x)) dx. \end{aligned}$$

□

**Teorema 3.** *O Teorema de Mudança de Variáveis em Integrais Múltiplas é válido quando  $X \subset \mathbb{R}^n$  é um bloco retangular e  $h : U \rightarrow V$  é um difeomorfismo primitivo do Tipo 2.*

*Demonstração.* Por conveniência, escreveremos os pontos de  $\mathbb{R}^n$  sob a forma  $x(s, w)$ , com  $s \in \mathbb{R}$  e  $w \in \mathbb{R}^{n-1}$ , e consideraremos o bloco  $X = I \times A$  como produto cartesiano do intervalo  $I = [a, b]$  pelo bloco  $A \subset \mathbb{R}^{n-1}$ . Note-se que,

para todo  $w \in A$  fixado, a função  $\varphi_w : s \mapsto \varphi(s, w) = I$  é um difeomorfismo do intervalo  $I$  sobre o intervalo  $J_w = \varphi_w(I) = \varphi(I \times w)$ . Observemos ainda que a matriz jacobiana de  $h$  tem a primeira linha igual ao gradiente de  $\varphi$  e, a partir da segunda linha, coincide com a matriz identidade  $n \times n$ . Portanto  $\det h'(x) = \frac{\partial \varphi}{\partial s}(s, w) = \varphi'_w(s)$ . Seja  $J \subset \mathbb{R}$  um intervalo compacto contendo  $J_w$  para todo  $w \in A$ . Então  $h(X) \subset J \times A$ . Como de praxe,  $\bar{f} : J \times A \rightarrow \mathbb{R}$  é a função integrável, igual a  $f$  nos pontos de  $h(X)$  e igual a zero nos demais pontos de  $J \times A$ . Então o Teorema 1 nos permite escrever:

$$\begin{aligned} \int_{h(X)} f(y) dy &= \int_{h(X)} f(t, w) dt dw = \int_{J \times A} \bar{f}(t, w) dt dw \\ &= \int_A \left( \int_J \bar{f}_w(t) dt \right) dw = \int_A \left( \int_{J_w} f_w(t) dt \right) dw \\ &= \int_A \left( \int_I f w(\varphi(s, w)) |\varphi'_w(s) ds| dw \right) \\ &= \int_{I \times A} f(\varphi(s, w), w) \cdot |\det h'(s, w)| ds dw \\ &= \int_X f(h(x)) \cdot |\det h'(x)| dx. \end{aligned}$$

□

### 3 Todo difeomorfismo $C^1$ é localmente admissível

Seja  $D$  o conjunto dos difeomorfismos de classe  $C^1$  para os quais é válido o Teorema de Mudança de Variáveis. Os elementos de  $D$  serão chamados *difeomorfismos admissíveis*.

Como sabemos, o objetivo deste capítulo é provar que todo difeomorfismo de classe  $C^1$  é admissível. Acabamos de ver que os difeomorfismos primitivos pertencem a  $D$ . Além disso, como

$$\det((h_1 \circ h_2)'(x)) = \det h'_1(h_2(x)) \cdot \det h'_2(x),$$

vê-se imediatamente que  $h_1 \circ h_2 \in D$  quando  $h_1 \in D$  e  $h_2 \in D$ .

Por exemplo, todo difeomorfismo da forma

$$h(x) = (x_1, \dots, x_{j-1}, \varphi(x), x_{j+1}, \dots, x_n)$$

é admissível pois é composto de dois difeomorfismos primitivos.

Nesta seção, provaremos que todo difeomorfismo de classe  $C^1$  é localmente admissível. Este é o conteúdo do teorema seguinte, o qual, evidentemente, é um resultado provisório.

**Teorema 4.** *Seja  $h : U \rightarrow V$  um difeomorfismo de classe  $C^1$  entre abertos de  $\mathbb{R}^n$ . Todo ponto de  $U$  possui uma vizinhança, restrita à qual  $h$  é admissível.*

*Demonstração.* Basta provar que, dado  $x_0 \in U$ , se  $h$  é definido numa vizinhança de  $x_0$  e tem a forma

$$h(x) = (\varphi_1(x), \dots, \varphi_j(x), x_{j+1}, \dots, x_n)$$

então existe um difeomorfismo  $k$ , de classe  $C^1$ , composto de difeomorfismos primitivos, cuja imagem é uma vizinhança de  $x_0$ , tal que

$$h(k(w)) = (\psi_1(w), \dots, \psi_{j-1}(w), w_j, \dots, w_n).$$

Ora, as  $j$  primeiras linhas da matriz jacobiana de  $h$  são os vetores  $\text{grad } \varphi_1, \dots, \text{grad } \varphi_j$  e as demais linhas coincidem com as da matriz identidade  $n \times n$ . Compondo  $h$ , se necessário, com um difeomorfismo do Tipo 1, podemos admitir que  $\frac{\partial \varphi_j}{\partial x_j}(x_0) \leq 0$ . Pelo Teorema 5 do Capítulo 6 (aplicado à função  $\varphi_j$ ) existe um difeomorfismo admissível

$$k : w \mapsto (w_1, \dots, k_j(w), w_{j+1}, \dots, w_n)$$

cuja imagem é uma vizinhança de  $x_0$  tal que  $\varphi_j(k(w)) = w_j$  para todo  $w$  no domínio de  $k$ . Então

$$h(k(w)) = (\varphi_1(k(w)), \dots, \varphi_{j-1}(k(w)), w_j, \dots, w_n),$$

completando assim a demonstração. □

#### 4 Conclusão: todo difeomorfismo de classe $C^1$ é admissível

Para terminar a demonstração do Teorema de Mudança de Variáveis, vamos usar um resultado topológico elementar que estabeleceremos agora.

Seja  $X \subset \bigcup_{\lambda \in L} C_\lambda$  uma cobertura do conjunto  $X \subset \mathbb{R}^n$ . Diz-se que  $\delta > 0$  é um *número de Lebesgue* dessa cobertura quando todo subconjunto  $Y \subset X$  com diâmetro  $< \delta$  está contido em algum  $C_\lambda$ . Isto equivale a dizer que, para todo  $x \in X$ , o conjunto  $X \cap B(x; \delta/2)$  está contido em algum  $C_\lambda$ .

**Teorema 5.** *Toda cobertura aberta  $X \subset \bigcup_{\lambda \in L} A_\lambda$  de um conjunto compacto  $X \subset \mathbb{R}^n$  possui número de Lebesgue.*

*Demonstração.* Se o teorema fosse falso existiriam, para cada  $n \in \mathbb{N}$ , um ponto  $x_n \in X$  e um número  $\delta_n > 0$  tais que  $\lim \delta_n = 0$  e  $X \cap B(x_n; \delta_n)$  não estaria contido em  $A_\lambda$  qualquer que fosse  $\lambda \in L$ . Passando a uma subsequência, se necessário, podemos admitir que  $\lim x_n = x_0 \in X$ . Seja  $\lambda_0 \in L$  tal que  $x_0 \in A_{\lambda_0}$ . Como  $A_{\lambda_0}$  é aberto, existe  $r > 0$  tal que  $B(x_0; r) \subset A_{\lambda_0}$ . Tomemos  $n \in \mathbb{N}$  tão grande que  $|x_n - x_0| < r/2$  e  $\delta_n < r/2$ . Então  $B(x_n; \delta_n) \subset B(x_0; r)$ , logo  $B(x_n; \delta_n) \subset A_{\lambda_0}$ , um absurdo.  $\square$

**Exemplo 1.** Seja  $X = A_1 \cup A_2 \subset \mathbb{R}^2$ . A cobertura aberta formada por  $A_1 = \{(x, y) \in \mathbb{R}^2; x > 0, y > 1/2\}$  e  $A_2 = \{(x, y) \in \mathbb{R}^2; y < 0\}$  não possui número de Lebesgue.

Com efeito, dado qualquer  $\delta > 0$  seja  $p = (x, -y)$  com  $x > 0, y > 0$  e  $\frac{1}{x} + y < \frac{\delta}{2}$  então  $X \cap B(p; \delta/2)$  é um subconjunto de  $X$  com diâmetro  $< \delta$ , o qual não está contido em  $A_1$  nem  $A_2$ .  $\triangleleft$

**Teorema 6.** *Sejam  $X \subset \mathbb{R}^n$  um conjunto compacto  $J$ -mensurável,  $h : U \rightarrow V$  um difeomorfismo de classe  $C^1$  entre abertos  $U, V \subset \mathbb{R}^n$  e  $f : h(X) \rightarrow \mathbb{R}$  uma função integrável. Então*

$$\int_{h(X)} f(y) dy = \int_X f(h(x)) \cdot |\det h'(x)| dx.$$

*Demonstração.* Existe uma cobertura aberta  $X \subset \bigcup_{x \in X} W_x \subset U$  tal que a restrição de  $h$  a cada  $W_x$  é um difeomorfismo admissível. Seja  $\delta > 0$  um número de Lebesgue dessa cobertura. Tomamos uma decomposição  $D = (X_1, \dots, X_k)$  de  $X$  tal que cada conjunto  $X_i$  tenha diâmetro inferior a  $\delta$ . (Para obter  $D$ , basta tomar uma partição  $P$  de um bloco  $A$  contendo  $X$  de modo que os blocos  $B_i$  de  $P$  tenham arestas  $< \delta$  na norma do máximo, ou  $< \delta/\sqrt{n}$  na norma euclidiana. Em seguida, ponha  $X_i = B_i \cap X$ .) Então

$$\begin{aligned} \int_{h(X)} f(y) dy &= \sum_i \int_{h(X_i)} f(y) dy = \sum_i \int_{X_i} f(h(x)) \cdot |\det h'(x)| dx \\ &= \int_X f(h(x)) \cdot |\det h'(x)| dx. \end{aligned}$$

$\square$

**Corolário 1.** *Seja  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  um operador linear. Para todo conjunto compacto  $J$ -mensurável  $X \subset \mathbb{R}^n$  tem-se  $\text{vol } T(X) = |\det T| \cdot \text{vol } X$ .*

Basta aplicar o Teorema 5 à função característica  $\xi_X$  em lugar de  $f$ , observando que é suficiente considerar o caso em que  $T$  é invertível.