seção 1 - 2
3 - 4 - 5

Capítulo 7

Aproximação linear e a regra da cadeia

Neste capítulo vamos ver que, sob certas condições, é possível aproximar uma aplicação f: $\mathbb{R}^n \to \mathbb{R}^m$ não-linear "complicada" na vizinhança de um ponto p por uma aplicação bem mais "simples": uma função afim, que é a composição de uma translação com uma transformação linear. Esta transformação linear é o objeto em Cálculo II, equivalente ao conceito de derivada estudada em Cálculo I e desempenhará um papel fundamental em toda teoria que vamos desenvolver. Analisando as propriedades da matriz que define esta transformação linear (algo supostamente mais fácil de se fazer) será possível concluir propriedades muito importantes da função original. Vamos começar com funções de $\mathbb{R}$ para $\mathbb{R}$, seguindo depois para funções de $\mathbb{R}^2$ para $\mathbb{R}$, de $\mathbb{R}^3$ para $\mathbb{R}$, de $\mathbb{R}^n$ para $\mathbb{R}$ e, finalmente, de $\mathbb{R}^n$ para $\mathbb{R}^m$.

7.1 Lembrando Cálculo I: a equação da reta tangente

Você aprendeu em Cálculo I que a reta tangente é a "melhor" reta que aproxima o gráfico de uma função $f\colon \mathbb{R} \to \mathbb{R}$ na vizinhança de um ponto p. Mas em que sentido uma reta é "melhor" do que outra? É preciso estabelecer um critério para fazer uma escolha entre as infinitas retas que existem no plano cartesiano. Se $y = l(x) = \alpha + \beta \cdot x$ é a equação de uma reta, então queremos duas coisas: (1) que $l(p) = f(p)$, isto é, que l e f coincidam no ponto $x = p$ e (2) que $l'(p) = f'(p)$, isto é, que as *derivadas* de l e f também coincidam no ponto $x = p$. Se f é de classe C^1, então existe uma única reta l que atende a estas duas condições, a *reta tangente* ao gráfico de f no ponto p.

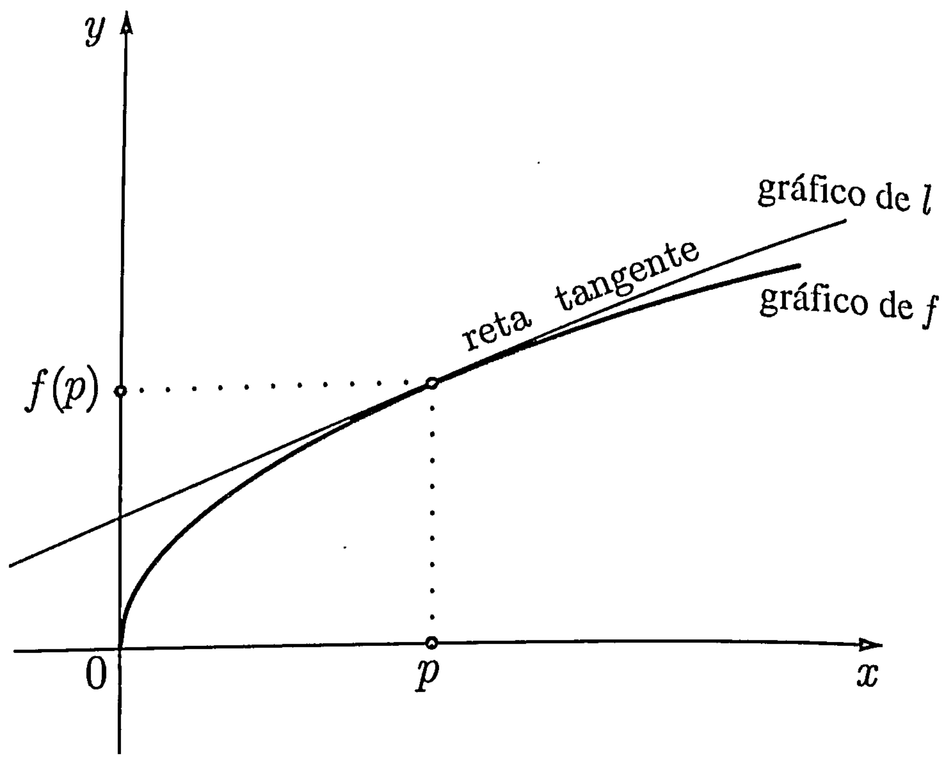

Figura 7.1: A reta tangente ao gráfico de uma função f de uma variável no ponto p.

Teorema 7.1 (A EQUAÇÃO DA RETA TANGENTE) Considere uma função $f\colon D \subset \mathbb{R} \to \mathbb{R}$ de classe C^1 e p um ponto do interior de D. Então existe uma única reta l que satisfaz as condições $l(p) = f(p)$ e $l'(p) = f'(p)$:

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

Mais ainda, vale que

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

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

$$f(x) = f(p) + f'(p) \cdot (x - p) + R(p, x), \qquad \text{com } \lim_{x \to p} \frac{R(p, x)}{x - p} = 0.$$

Demonstração: Se $l(p) = f(p)$ então $\alpha + \beta \cdot p = f(p)$, isto é,

$$\alpha = f(p) - \beta \cdot p.$$

Por outro lado, como $l'(p) = f'(p)$ e $l'(p) = \beta$ segue-se que

$$\beta = f'(p).$$

Desta maneira,

$$l(x) = \alpha + \beta \cdot x = f(p) - \beta \cdot p + \beta \cdot x =$$
$$f(p) + \beta \cdot (x - p) = f(p) + f'(p) \cdot (x - p).$$

Esta é a equação da reta tangente ao gráfico de f no ponto $x = p$. Finalmente, observe que

$$\lim_{x \to p} \frac{f(x) - l(x)}{x - p} = \lim_{x \to p} \frac{f(x) - f(p) - f'(p) \cdot (x - p)}{x - p} =$$
$$\lim_{x \to p} \left(\frac{f(x) - f(p)}{x - p} - f'(p) \right) = f'(p) - f'(p) = 0.$$

□

Em outras palavras, este teorema nos diz que é possível aproximar uma função f de classe C^1 na vizinhança de um ponto p pela equação da reta tangente $l(x) = f(p) + f'(p) \cdot (x - p)$ (uma função afim) e que o erro $R(p, x) = f(x) - l(x)$ cometido nesta aproximação tende a zero mais rapidamente do que $x - p$. Mais formalmente,

$$f(x) = \underbrace{f(p) + f'(p) \cdot (x - p)}_{l(x)} + R(p, x),$$
$$\text{com } \lim_{x \to p} \frac{R(p, x)}{x - p} = \lim_{x \to p} \frac{f(x) - l(x)}{x - p} = 0.$$

Podemos reescrever estas expressões em termos da "pertubação" $h = x - p$ com relação ao ponto p de modo que $x = p + h$ e

$$f(p + h) = f(p) + f'(p) \cdot h + R(p, h), \qquad \text{com } \lim_{h \to 0} \frac{R(p, h)}{h} = 0. \tag{7.1}$$

Exemplo 7.1 Seja $f(x) = e^x$. A equação da reta tangente ao gráfico de f no ponto $p = 0$ é dada por

$$y = l(x) = f(0) + f'(0) \cdot (x - 0) = e^0 + e^0 \cdot x = 1 + x.$$

Desta maneira,

$$e^x = 1 + x + R(0,x), \qquad \text{com } \lim_{x\to 0} \frac{R(0,x)}{x} = 0.$$

Podemos usar $l(x)$ (uma função mais simples) para estimar o valor de $f(x)$, para x próximo de 0. Por exemplo, em $x = 0.2$ temos $l(0.2) = 1 + 0.2 = 1.2$, enquanto que $f(0.2) = e^{0.2} = 1.22140\ldots$. ∎

A equação (7.1) pode ser reescrita como

$$\underbrace{f(p+h) - f(p)}_{V(h)} = \underbrace{f'(p)\cdot h}_{T(h)} + R(p,h)$$

e ela nos diz que a variação $V(h) = f(p+h) - f(p)$ do valor da função f no ponto $x = p$ pode ser aproximada pela *transformação linear*

$$\begin{array}{rccl} T\colon & \mathbb{R} & \to & \mathbb{R} \\ & h & \mapsto & T(h) = f'(p)\cdot h \end{array}$$

e que o erro $R(p,h) = V(h) - T(h)$ cometido nesta aproximação tende a zero mais rapidamente do que h. A matriz associada a esta transformação linear é dada por

$$\left[\frac{df}{dx}(p)\right]_{1\times 1}$$

e ela pode ser identificada com o número real $(df/dx)(p)$.

No caso de funções de uma variável parece não haver ganho substancial em se pensar em transformações lineares e matrizes quando falamos de derivadas. Contudo, como veremos a seguir, esta formulação permitirá a generalização do conceito de aproximação linear para funções vetoriais.

7.2 Aproximação linear em Cálculo II

Funções escalares de 2 variáveis: a equação do plano tangente

Vamos começar com o caso mais simples: sejam $z = f(x,y)$ uma função de duas variáveis de classe C^1 e $\text{p} = (a,b)$ um ponto do interior do domínio de f. Qual é o "melhor" plano que aproxima o gráfico de f na vizinhança do ponto p? Novamente, é preciso estabelecer um critério. Se $z = l(x,y) =$

$\alpha + \beta \cdot x + \gamma \cdot y$ é a equação de um plano, então queremos duas coisas: (1) que $l(\mathrm{p}) = f(\mathrm{p})$, isto é, que l e f coincidam no ponto $\mathrm{p} = (a, b)$ e (2) que

$$\frac{\partial l}{\partial x}(\mathrm{p}) = \frac{\partial f}{\partial x}(\mathrm{p}) \qquad \text{e} \qquad \frac{\partial l}{\partial y}(\mathrm{p}) = \frac{\partial f}{\partial y}(\mathrm{p}),$$

isto é, que as *derivadas parciais* de l e f também coincidam no ponto p. Se f é de classe C^1, então existe um único plano l que atende a estas duas condições, o *plano tangente* ao gráfico de f no ponto p.

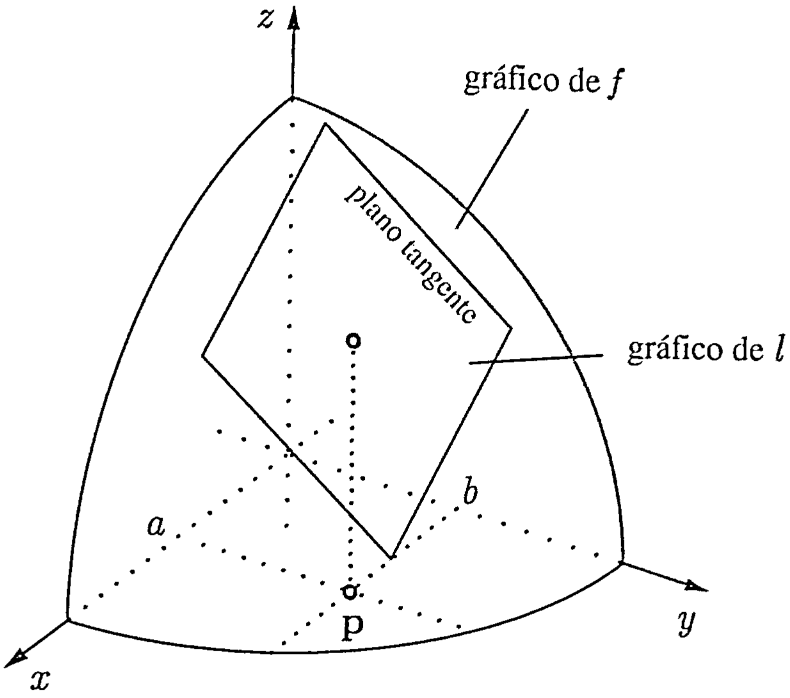

Figura 7.2: O plano tangente ao gráfico de uma função f de duas variáveis no ponto p.

Teorema 7.2 (A EQUAÇÃO DO PLANO TANGENTE) Considere uma função $f\colon D \subset \mathbb{R}^2 \to \mathbb{R}$ de classe C^1 e $\mathrm{p} = (a, b)$ um ponto do interior de D. Então existe um único plano l que satisfaz as condições $l(\mathrm{p}) = f(\mathrm{p})$ e

$$\frac{\partial l}{\partial x}(\mathrm{p}) = \frac{\partial f}{\partial x}(\mathrm{p}) \qquad \text{e} \qquad \frac{\partial l}{\partial y}(\mathrm{p}) = \frac{\partial f}{\partial y}(\mathrm{p}),$$

a saber,

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

Mais ainda, vale que

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

isto é, a diferença entre $f(x,y)$ e $l(x,y)$ vai para zero mais rapidamente do que $||(x-a,y-b)||$. Em outras palavras,

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

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

Demonstração: Se $l(a,b) = f(a,b)$ então $\alpha + \beta\cdot a + \gamma\cdot b = f(a,b)$, isto é,

$$\alpha = f(a,b) - \beta\cdot a - \gamma\cdot b.$$

Por outro lado, como

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

segue-se que

$$\beta = \frac{\partial f}{\partial x}(a,b) \quad \text{e} \quad \gamma = \frac{\partial f}{\partial y}(a,b).$$

Desta maneira, substituindo os valores de α, β e γ na expressão de l vemos que

$$\begin{aligned} l(x,y) &= \alpha + \beta\cdot x + \gamma\cdot y \\ &= f(a,b) - \beta\cdot a - \gamma\cdot b + \beta\cdot x + \gamma\cdot y \\ &= f(a,b) + \beta\cdot(x-a) + \gamma\cdot(y-b) \\ &= f(a,b) + \frac{\partial f}{\partial x}(a,b)\cdot(x-a) + \frac{\partial f}{\partial y}(a,b)\cdot(y-b). \end{aligned}$$

Esta é a equação do plano tangente ao gráfico de f no ponto $\mathrm{p} = (a,b)$. Resta mostrar que esta equação satisfaz a propriedade

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

Desta vez, a demonstração é um pouco mais sofisticada e requer o uso do teorema do valor médio para derivadas. Não vamos fazê-la aqui. Uma demonstração completa deste teorema pode ser encontrada na referência [71], página 833. □

Em outras palavras, este teorema nos diz que é possível aproximar uma função f de classe C^1 de duas variáveis na vizinhança de um ponto $\mathrm{p} = (a,b)$ pela equação do plano tangente $l(x,y) = f(a,b) + (\partial f/\partial x)(a,b)\cdot(x-a) + (\partial f/\partial y)(a,b)\cdot(y-b)$ (uma função afim) e que o erro $R(a,b,x,y) = f(x,y) - l(x,y)$ cometido nesta aproximação tende a zero mais rapidamente do que $||(x-a,y-b)||$. Mais formalmente,

$$f(x,y) = \underbrace{f(a,b) + \frac{\partial f}{\partial x}(a,b)\cdot(x-a) + \frac{\partial f}{\partial y}(a,b)\cdot(y-b)}_{l(x,y)} + R(a,b,x,y),$$

com

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

Podemos reescrever estas expressões em termos da "pertubação" $(h_1,h_2) = (x-a, y-b)$ com relação ao ponto $\mathrm{p} = (a,b)$, de modo que $x = a + h_1$, $y = b + h_2$ e

$$f(a+h_1, b+h_2) = f(a,b) + \frac{\partial f}{\partial x}(a,b)\cdot h_1 + \frac{\partial f}{\partial y}(a,b)\cdot h_2 + R(a,b,h_1,h_2),$$
$$\text{com} \lim_{(h_1,h_2)\to(0,0)} \frac{R(a,b,h_1,h_2)}{||(h_1,h_2)||} = 0. \quad (7.2)$$

Exemplo 7.2 No exemplo (5.1), página 167, estudamos a função de produção Cobb-Douglas

$$z = f(x,y) = 4\cdot x^{3/4}\cdot y^{1/4}.$$

A equação do plano tangente ao gráfico de f no ponto $(a,b) = (10000, 625)$ é dada por

$$z = l(x,y) = f(10000,625) + \frac{\partial f}{\partial x}(10000,625)\cdot(x-10000) + \frac{\partial f}{\partial y}(10000,625)\cdot(y-625).$$

Mas, conforme os cálculos feitos no exemplo (5.1), $f(10000, 625) = 20000$, $(\partial f/\partial x)(10000, 625) = 1.5$ e $(\partial f/\partial y)(10000, 625) = 8$, logo

$$z = l(x, y) = 20000 + 1.5 \cdot (x - 10000) + 8 \cdot (y - 625) = 1.5 \cdot x + 8 \cdot y.$$

Podemos usar $l(x, y)$ (uma função mais simples) para estimar o valor de $f(x, y)$ para (x, y) próximo de $(10000, 625)$. Por exemplo, em $(x, y) = (10010, 623)$ temos $l(10010, 623) = 1.5 \cdot 10010 + 8 \cdot 623 = 19999$, enquanto que $f(10010, 623) = 19998.967\ldots$.

∎

Suponha que um certo fenômeno de interesse seja modelado por uma função $z = f(x, y)$. Em experimentos, os valores de x e y são amostrados com um certo erro percentual. Uma pergunta muito natural é o quanto estes erros afetam o valor z da função. A equação do plano tangente é uma ferramenta bastante útil para se estimar este erro. Vamos ver como esta técnica funciona em um exemplo.

Exemplo 7.3 Considere um balão meteorológico que é liberado ao nível do mar, a uma distância d do olho de um furacão. Sua altitude aumenta à medida que ele se move em direção ao olho. A altitude h que o balão atingirá quando ele estiver no olho do furacão pode ser modelada pela relação

$$h = \pi^2 \cdot g \cdot \frac{d^4}{c^2}, \tag{7.3}$$

onde g é uma constante gravitacional e c é uma medida meteorológica denominada *circulação* da velocidade do vento (a circulação está estreitamente relacionada com a intensidade e a direção dos ventos no interior do furacão).

Supondo que os erros percentuais máximos nas medidas de d e h sejam, respectivamente, $\pm 2\%$ e $\pm 5\%$, vamos utilizar a equação do plano tangente para estimar o erro máximo no cálculo da circulação c. Observe que a relação (7.3) define c como uma função de d e h:

$$c = f(d, h) = \pi \cdot g^{1/2} \cdot \frac{d^2}{h^{1/2}}.$$

A equação do plano tangente ao gráfico de $z = f(x, y)$ no ponto (d^*, h^*) é dada pela expressão

$$l(d, h) = f(d^*, h^*) + \frac{\partial f}{\partial d}(d^*, h^*) \cdot (d - d^*) + \frac{\partial f}{\partial h}(d^*, h^*) \cdot (h - h^*).$$

Se (d,h) está próximo de (d^*,h^*) então $l(d,h)$ fornece uma boa aproximação de $f(d,h)$ e, portanto, a variação $\Delta l(d,h) = l(d,h) - f(d^*,h^*)$ fornece uma boa aproximação para o erro (absoluto) $\Delta f(d,h) = f(d,h) - f(d^*,h^*)$. Sendo assim,

$$\Delta f(d,h) \approx \Delta l(d,h) = \frac{\partial f}{\partial d}(d^*,h^*)\cdot(d-d^*) + \frac{\partial f}{\partial h}(d^*,h^*)\cdot(h-h^*).$$

Agora, derivando a função f com relação a d e h no ponto (d^*,h^*) obtemos, respectivamente,

$$\frac{\partial f}{\partial d}(d^*,h^*) = \pi\cdot g^{1/2}\cdot\frac{2\,d^*}{(h^*)^{1/2}} \quad \text{e} \quad \frac{\partial f}{\partial h}(d^*,h^*) = -\pi\cdot g^{1/2}\cdot\frac{1}{2}\cdot\frac{(d^*)^2}{2\,(h^*)^{3/2}}.$$

Desta maneira, escrevendo $\Delta d = d - d^*$ e $\Delta h = h - h^*$, segue-se que

$$\Delta f(d,h) \approx \Delta l(d,h) = \pi\cdot g^{1/2}\cdot\frac{2\,d^*}{(h^*)^{1/2}}\cdot\Delta d - \pi\cdot g^{1/2}\cdot\frac{1}{2}\cdot\frac{(d^*)^2}{2\,(h^*)^{3/2}}\cdot\Delta h,$$

Como estamos interessados em estimar o erro percentual, vamos dividir esta expressão por $f(d^*,h^*)$ (que é igual a $l(d^*,h^*)$), de modo que

$$\frac{\Delta f(d,h)}{f(d^*,h^*)} \approx \frac{\Delta l(d,h)}{l(d^*,h^*)} = 2\cdot\frac{\Delta d}{d^*} - \frac{1}{2}\cdot\frac{\Delta h}{h^*}.$$

Note que $\Delta d/d^*$ e $\Delta h/h^*$ representam, respectivamente, os erros percentuais nas medidas da distância d e altura h. Não é difícil de ver que os valores máximo e mínimo de $\Delta l(x,y)/l(d^*,h^*)$ são obtidos tomando-se $\Delta d/d^*$ e $\Delta h/h^*$ com sinais contrários:

$$-6.5\% = 2\cdot(-2\%) - \frac{1}{2}\cdot(5\%) \le \frac{\Delta l(d,h)}{l(d^*,h^*)} \le 2\cdot(2\%) - \frac{1}{2}\cdot(-5\%) = 6.5\%.$$

Desta maneira, podemos estimar o erro percentual máximo no cálculo da circulação c em $\pm 6.5\%$. □

A equação (7.2) pode ser reescrita como

$$\underbrace{f(a+h_1,b+h_2) - f(a,b)}_{V(h_1,h_2)} = \underbrace{\frac{\partial f}{\partial x}(a,b)\cdot h_1 + \frac{\partial f}{\partial y}(a,b)\cdot h_2}_{T(h_1,h_2)} + R(a,b,h_1,h_2).$$

Ela nos diz que a variação $V(h_1, h_2) = f(a+h_1, b+h_2) - f(a,b)$ do valor da função f no ponto $\mathrm{p} = (a,b)$ pode ser aproximada pela *transformação linear*

$$\begin{array}{rcl} T: & \mathbb{R}^2 & \to \mathbb{R} \\ & (h_1, h_2) & \mapsto T(h_1,h_2) = \dfrac{\partial f}{\partial x}(a,b)\cdot h_1 + \dfrac{\partial f}{\partial y}(a,b)\cdot h_2 \end{array},$$

e que o erro $R(a,b,h_1,h_2) = V(h_1,h_2) - T(h_1,h_2)$ cometido nesta aproximação tende a zero mais rapidamente do que $||(h_1,h_2)|| = \sqrt{h_1^2+h_2^2}$. A matriz associada a esta transformação linear é dada por

$$\begin{bmatrix} \dfrac{\partial f}{\partial x}(a,b) & \dfrac{\partial f}{\partial y}(a,b) \end{bmatrix}_{1\times 2},$$

de modo que

$$T(h_1,h_2) = \frac{\partial f}{\partial x}(a,b)\cdot h_1 + \frac{\partial f}{\partial y}(a,b)\cdot h_2 = \begin{bmatrix} \dfrac{\partial f}{\partial x}(a,b) & \dfrac{\partial f}{\partial y}(a,b) \end{bmatrix}_{1\times 2} \cdot \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}_{2\times 1}.$$

T é a "melhor" transformação linear que aproxima a variação de f perto do ponto (a,b) ou, equivalentemente, l (que define a equação do plano tangente) é a melhor função afim que aproxima f perto do ponto (a,b).

Funções escalares de 3 variáveis

E para uma função de três variáveis? Se $w = f(x,y,z)$ é uma função de classe C^1, $\mathrm{p} = (a,b,c)$ é um ponto interior do domínio de f e

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

então

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

l é a única função afim que satisfaz as equações. Ou, usando uma notação vetorial mais compacta ($\mathrm{x} = (x, y, z)$ e $\mathrm{p} = (a, b, c)$), temos que se f é de classe C^1, p é um ponto interior do domínio de f e

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

então

$$\lim_{\mathrm{x} \to \mathrm{p}} \frac{f(\mathrm{x}) - l(\mathrm{x})}{||\mathrm{x} - \mathrm{p}||} = 0.$$

Em outras palavras, é possível aproximar uma função f de classe C^1 de três variáveis na vizinhança de um ponto $\mathrm{p} = (a, b, c)$ pela função afim $l(\mathrm{x}) = f(\mathrm{p}) + (\partial f/\partial x)(\mathrm{p}) \cdot (x - a) + (\partial f/\partial y)(\mathrm{p}) \cdot (y - b) + (\partial f/\partial z)(\mathrm{p}) \cdot (z - c)$ e o erro $R(\mathrm{p}, \mathrm{x}) = f(\mathrm{x}) - l(\mathrm{x})$ cometido nesta aproximação tende a zero mais rapidamente do que $||\mathrm{x} - \mathrm{p}||$. Mais formalmente,

$$f(\mathrm{x}) = \underbrace{f(\mathrm{p}) + \frac{\partial f}{\partial x}(\mathrm{p}) \cdot (x - a) + \frac{\partial f}{\partial y}(\mathrm{p}) \cdot (y - b) + \frac{\partial f}{\partial z}(\mathrm{p}) \cdot (z - c)}_{l(\mathrm{x})} + R(\mathrm{p}, \mathrm{x}),$$

com

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

Podemos reescrever estas expressões em termos de $\mathrm{h} = (h_1, h_2, h_3) = \mathrm{x} - \mathrm{p}$ (h descreve a "pertubação" com relação ao ponto p), de modo que $\mathrm{x} = \mathrm{p} + \mathrm{h}$ e

$$f(\mathrm{p} + \mathrm{h}) = f(\mathrm{p}) + \frac{\partial f}{\partial x}(\mathrm{p}) \cdot h_1 + \frac{\partial f}{\partial y}(\mathrm{p}) \cdot h_2 + \frac{\partial f}{\partial z}(\mathrm{p}) \cdot h_3 + R(\mathrm{p}, \mathrm{h}),$$

$$\text{com } \lim_{\mathrm{h} \to 0} \frac{R(\mathrm{p}, \mathrm{h})}{||\mathrm{h}||} = 0. \quad (7.4)$$

A equação (7.4) pode ser reescrita como

$$\underbrace{f(\mathrm{p} + \mathrm{h}) - f(\mathrm{p})}_{V(\mathrm{h})} = \underbrace{\frac{\partial f}{\partial x}(\mathrm{p}) \cdot h_1 + \frac{\partial f}{\partial y}(\mathrm{p}) \cdot h_2 + \frac{\partial f}{\partial z}(\mathrm{p}) \cdot h_3}_{T(\mathrm{h})} + R(\mathrm{p}, \mathrm{h}).$$

Ela nos diz que a variação $V(\mathrm{h}) = f(\mathrm{p} + \mathrm{h}) - f(\mathrm{p})$ do valor da função f no ponto p pode ser aproximada pela *transformação linear* $T\colon \mathbb{R}^3 \to \mathbb{R}$ definida por

$$T(\mathrm{h}) = T(h_1, h_2, h_3) = \frac{\partial f}{\partial x}(\mathrm{p}) \cdot h_1 + \frac{\partial f}{\partial y}(\mathrm{p}) \cdot h_2 + \frac{\partial f}{\partial z}(\mathrm{p}) \cdot h_3,$$

e que o erro $R(\mathrm{p},\mathrm{h}) = V(\mathrm{h}) - T(\mathrm{h})$ cometido nesta aproximação tende a zero mais rapidamente do que $||\mathrm{h}|| = \sqrt{h_1^2 + h_2^2 + h_3^2}$. A matriz associada a esta transformação linear é dada por

$$\begin{bmatrix} \frac{\partial f}{\partial x}(\mathrm{p}) & \frac{\partial f}{\partial y}(\mathrm{p}) & \frac{\partial f}{\partial z}(\mathrm{p}) \end{bmatrix}_{1\times 3},$$

de modo que

$$\begin{aligned} T(\mathrm{h}) &= \frac{\partial f}{\partial x}(\mathrm{p}) \cdot h_1 + \frac{\partial f}{\partial y}(\mathrm{p}) \cdot h_2 + \frac{\partial f}{\partial z}(\mathrm{p}) \cdot h_3 \\ &= \begin{bmatrix} \frac{\partial f}{\partial x}(\mathrm{p}) & \frac{\partial f}{\partial y}(\mathrm{p}) & \frac{\partial f}{\partial z}(\mathrm{p}) \end{bmatrix}_{1\times 3} \cdot \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix}_{3\times 1} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x}(\mathrm{p}) & \frac{\partial f}{\partial y}(\mathrm{p}) & \frac{\partial f}{\partial z}(\mathrm{p}) \end{bmatrix}_{1\times 3} \cdot \mathrm{h}. \end{aligned}$$

T é a "melhor" transformação linear que aproxima a variação de f perto do ponto p ou, equivalentemente, l é a melhor função afim que aproxima f perto do ponto p. Também, l é a única função afim que satisfaz as condições

$$l(\mathrm{p}) = f(\mathrm{p}), \quad \frac{\partial l}{\partial x}(\mathrm{p}) = \frac{\partial f}{\partial x}(\mathrm{p}), \quad \frac{\partial l}{\partial y}(\mathrm{p}) = \frac{\partial f}{\partial y}(\mathrm{p}) \quad \text{e} \quad \frac{\partial l}{\partial z}(\mathrm{p}) = \frac{\partial f}{\partial z}(\mathrm{p}).$$

Funções escalares de n variáveis

Evidentemente, tudo isto pode ser estendido para funções que dependam de n variáveis. Se $w = f(\mathrm{x}) = f(x_1, \ldots, x_n)$ é uma função de classe C^1, $\mathrm{p} = (p_1, \ldots, p_n)$ é um ponto interior do domínio de f e

$$l(\mathrm{x}) = f(\mathrm{p}) + \frac{\partial f}{\partial x_1}(\mathrm{p}) \cdot (x_1 - p_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p}) \cdot (x_n - p_n),$$

então

$$\lim_{\mathrm{x} \to \mathrm{p}} \frac{f(\mathrm{x}) - l(\mathrm{x})}{||\mathrm{x} - \mathrm{p}||} = 0.$$

Em outras palavras, é possível aproximar uma função f de classe C^1 de n variáveis na vizinhança de um ponto p pela função afim l e o erro $R(\mathrm{p},\mathrm{x}) = f(\mathrm{x}) - l(\mathrm{x})$ cometido nesta aproximação tende a zero mais rapidamente do que $||\mathrm{x}-\mathrm{p}||$. Mais formalmente,

$$f(\mathrm{x}) = \underbrace{f(\mathrm{p}) + \frac{\partial f}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n)}_{l(\mathrm{x})} + R(\mathrm{p},\mathrm{x}),$$

com

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

Podemos reescrever estas expressões em termos de $\mathrm{h} = (h_1,\ldots,h_n) = \mathrm{x}-\mathrm{p}$ (h descreve a "pertubação" com relação ao ponto p), de modo que $\mathrm{x} = \mathrm{p}+\mathrm{h}$ e

$$f(\mathrm{p}+\mathrm{h}) = f(\mathrm{p}) + \frac{\partial f}{\partial x_1}(\mathrm{p})\cdot h_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p})\cdot h_n + R(\mathrm{p},\mathrm{h}),$$

$$\text{com } \lim_{\mathrm{h}\to\mathbf{0}} \frac{R(\mathrm{p},\mathrm{h})}{||\mathrm{h}||} = 0. \quad (7.5)$$

A equação (7.5) pode ser reescrita como

$$\underbrace{f(\mathrm{p}+\mathrm{h}) - f(\mathrm{p})}_{V(\mathrm{h})} = \underbrace{\frac{\partial f}{\partial x_1}(\mathrm{p})\cdot h_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p})\cdot h_n}_{T(\mathrm{h})} + R(\mathrm{p},\mathrm{h}).$$

Ela nos diz que a variação $V(\mathrm{h}) = f(\mathrm{p}+\mathrm{h}) - f(\mathrm{p})$ do valor da função f no ponto p pode ser aproximada pela *transformação linear* $T\colon \mathbb{R}^n \to \mathbb{R}$ definida por

$$T(\mathrm{h}) = T(h_1,\ldots,h_n) = \frac{\partial f}{\partial x_1}(\mathrm{p})\cdot h_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p})\cdot h_n,$$

e que o erro $R(\mathrm{p},\mathrm{h}) = V(\mathrm{h}) - T(\mathrm{h})$ cometido nesta aproximação tende a zero mais rapidamente do que $||\mathrm{h}|| = \sqrt{h_1^2 + \cdots + h_n^2}$. A matriz associada a esta transformação linear é dada por

$$\begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f}{\partial x_n}(\mathrm{p}) \end{bmatrix}_{1\times n},$$

de modo que

$$\begin{aligned}
T(\mathrm{h}) &= \frac{\partial f}{\partial x_1}(\mathrm{p}) \cdot h_1 + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p}) \cdot h_n \\
&= \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f}{\partial x_n}(\mathrm{p}) \end{bmatrix}_{1\times n} \cdot \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix}_{n\times 1} \\
&= \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f}{\partial x_n}(\mathrm{p}) \end{bmatrix}_{1\times n} \cdot \mathrm{h}.
\end{aligned}$$

T é a "melhor" transformação linear que aproxima a variação de f perto do ponto p ou, equivalentemente, l é a melhor função afim que aproxima f perto do ponto p. Também, l é a única função afim que satisfaz as condições

$$l(\mathrm{p}) = f(\mathrm{p}), \quad \frac{\partial l}{\partial x_1}(\mathrm{p}) = \frac{\partial f}{\partial x_1}(\mathrm{p}), \quad \ldots, \quad \frac{\partial l}{\partial x_n}(\mathrm{p}) = \frac{\partial f}{\partial x_n}(\mathrm{p}).$$

Vamos resumir os resultados que obtivemos até agora em um teorema.

Teorema 7.3 (APROXIMAÇÃO LINEAR: O CASO ESCALAR) Considere uma função $f\colon D \subset \mathbb{R}^n \to \mathbb{R}$ de classe C^1 e $\mathrm{p} = (p_1, \ldots, p_n)$ um ponto no interior de D. Então existe uma única função afim l que satisfaz as condições $l(\mathrm{p}) = f(\mathrm{p})$ e

$$\frac{\partial l}{\partial x_1}(\mathrm{p}) = \frac{\partial f}{\partial x_1}(\mathrm{p}), \quad \ldots, \quad \frac{\partial l}{\partial x_n}(\mathrm{p}) = \frac{\partial f}{\partial x_n}(\mathrm{p}),$$

a saber,

$$l(\mathrm{x}) = f(\mathrm{p}) + \frac{\partial f}{\partial x_1}(\mathrm{p}) \cdot (x_1 - p_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p}) \cdot (x_n - p_n).$$

Mais ainda, vale que

$$\lim_{\mathrm{x}\to\mathrm{p}} \frac{f(\mathrm{x}) - l(\mathrm{x})}{||\mathrm{x} - \mathrm{p}||} = 0,$$

isto é, a diferença entre $f(\mathrm{x})$ e $l(\mathrm{x})$ tende a zero mais rapidamente do que $||\mathrm{x}-\mathrm{p}||$. Em outras palavras,

$$f(\mathrm{x}) = f(\mathrm{p}) + \frac{\partial f}{\partial x_1}(\mathrm{p}) \cdot (x_1 - p_1) + \cdots + \frac{\partial f}{\partial x_n}(\mathrm{p}) \cdot (x_n - p_n) + R(\mathrm{p},\mathrm{x}),$$

$$\text{com } \lim_{\mathrm{x}\to\mathrm{p}} \frac{R(\mathrm{p},\mathrm{x})}{||\mathrm{x}-\mathrm{p}||} = 0.$$

Funções vetoriais: o caso geral

Também podemos aproximar uma função vetorial f de classe C^1 por uma transformação afim l na vizinhança de um ponto p no interior do domínio de f. O que vamos fazer é uma extensão imediata do teorema (7.3) para funções escalares.

Considere uma função vetorial

$$\begin{array}{rrcl} \mathrm{f}: & D \subset \mathbb{R}^n & \to & \mathbb{R}^m \\ & \mathrm{x} = (x_1, \ldots, x_n) & \mapsto & \mathrm{y} = \mathrm{f}(\mathrm{x}) = (y_1, \ldots, y_m) = (f_1(\mathrm{x}), \ldots, f_m(\mathrm{x})) \end{array}$$

definida em um subconjunto D de $\mathbb{R}^n$. Como vimos, podemos pensar que f: $D \subset \mathbb{R}^n \to \mathbb{R}^m$ é uma maneira de se agrupar ou representar m funções escalares definidas em D:

$$\begin{array}{rrcl} f_1: & D \subset \mathbb{R}^n & \to & \mathbb{R} \\ & \mathrm{x} = (x_1, \ldots, x_n) & \mapsto & y_1 = f_1(\mathrm{x}) = f_1(x_1, \ldots, x_n) \end{array}$$

$$\begin{array}{rrcl} f_2: & D \subset \mathbb{R}^n & \to & \mathbb{R} \\ & \mathrm{x} = (x_1, \ldots, x_n) & \mapsto & y_2 = f_2(\mathrm{x}) = f_2(x_1, \ldots, x_n) \end{array}$$

$$\vdots$$

$$\begin{array}{rrcl} f_m: & D \subset \mathbb{R}^n & \to & \mathbb{R} \\ & \mathrm{x} = (x_1, \ldots, x_n) & \mapsto & y_m = f_m(\mathrm{x}) = f_m(x_1, \ldots, x_n) \end{array}$$

Como f é de classe C^1, então suas funções coordenadas $f_1, f_2, \ldots, f_m$ também são de classe C^1. Logo, podemos aplicar o teorema (7.3) para elas e obter funções afins

$$\begin{cases} l_1(\mathrm{x}) &= f_1(\mathrm{p}) + \dfrac{\partial f_1}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) +\cdots+ \dfrac{\partial f_1}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n),\\ l_2(\mathrm{x}) &= f_2(\mathrm{p}) + \dfrac{\partial f_2}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) +\cdots+ \dfrac{\partial f_2}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n)\\ \quad\vdots & \\ l_m(\mathrm{x}) &= f_m(\mathrm{p}) + \dfrac{\partial f_m}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) +\cdots+ \dfrac{\partial f_m}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n) \end{cases}$$

que aproximam, respectivamente, f_1, f_2, $\ldots$, f_m, nas proximidades do ponto $\mathrm{p} = (p_1, p_2, \ldots, p_n)$. Mais especificamente,

$$\begin{cases} f_1(\mathrm{x}) &= f_1(\mathrm{p}) + \dfrac{\partial f_1}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) +\cdots+ \dfrac{\partial f_1}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n)\\ & \qquad + R_1(\mathrm{p},\mathrm{x}),\\ f_2(\mathrm{x}) &= f_2(\mathrm{p}) + \dfrac{\partial f_2}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) +\cdots+ \dfrac{\partial f_2}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n)\\ & \qquad + R_2(\mathrm{p},\mathrm{x}),\\ \quad\vdots & \\ f_m(\mathrm{x}) &= f_m(\mathrm{p}) + \dfrac{\partial f_m}{\partial x_1}(\mathrm{p})\cdot(x_1-p_1) +\cdots+ \dfrac{\partial f_m}{\partial x_n}(\mathrm{p})\cdot(x_n-p_n)\\ & \qquad + R_m(\mathrm{p},\mathrm{x}), \end{cases}$$

com $R_1(\mathrm{p},\mathrm{x}) = l_1(\mathrm{x}) - f_1(\mathrm{x})$, $R_2(\mathrm{p},\mathrm{x}) = l_2(\mathrm{x}) - f_1(\mathrm{x})$, $\ldots$, $R_m(\mathrm{p},\mathrm{x}) = l_m(\mathrm{x}) - f_m(\mathrm{x})$, e

$$\lim_{\mathrm{x}\to\mathrm{p}} \frac{R_1(\mathrm{p},\mathrm{x})}{||\mathrm{x}-\mathrm{p}||} = 0, \quad \lim_{\mathrm{x}\to\mathrm{p}} \frac{R_2(\mathrm{p},\mathrm{x})}{||\mathrm{x}-\mathrm{p}||} = 0, \quad \ldots, \quad \lim_{\mathrm{x}\to\mathrm{p}} \frac{R_m(\mathrm{p},\mathrm{x})}{||\mathrm{x}-\mathrm{p}||} = 0.$$

Observe que podemos reescrever as m igualdades acima para f_1, f_2, $\ldots$, f_m, usando matrizes:

$$\begin{bmatrix} f_1(\mathrm{x}) \\ f_2(\mathrm{x}) \\ \vdots \\ f_m(\mathrm{x}) \end{bmatrix} = \begin{bmatrix} f_1(\mathrm{p}) \\ f_2(\mathrm{p}) \\ \vdots \\ f_m(\mathrm{p}) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathrm{p}) \\ \frac{\partial f_2}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathrm{p}) \end{bmatrix} \cdot \begin{bmatrix} x_1 - p_1 \\ x_2 - p_2 \\ \vdots \\ x_n - p_n \end{bmatrix}$$

$$+ \begin{bmatrix} R_1(\mathrm{p},\mathrm{x}) \\ R_2(\mathrm{p},\mathrm{x}) \\ \vdots \\ R_m(\mathrm{p},\mathrm{x}) \end{bmatrix}.$$

O objetivo agora é reescrever esta equação matricial com uma notação funcional mais compacta, isto é, *sem o uso de coordenadas*. Para isto, sejam $\mathrm{l}(\mathrm{x}) = (l_1(\mathrm{x}), l_2(\mathrm{x}), \ldots, l_m(\mathrm{x}))$, $\mathrm{R}(\mathrm{p},\mathrm{x}) = (R_1(\mathrm{p},\mathrm{x}), R_2(\mathrm{p},\mathrm{x}), \ldots, R_m(\mathrm{p},\mathrm{x}))$ e

$$D\mathrm{f}(\mathrm{p}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathrm{p}) \\ \frac{\partial f_2}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathrm{p}) \end{bmatrix}_{m\times n}$$

de modo que

$$\mathrm{f}(\mathrm{x}) = \underbrace{\mathrm{f}(\mathrm{p}) + D\mathrm{f}(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p})}_{\mathrm{l}(\mathrm{x})} + \mathrm{R}(\mathrm{p},\mathrm{x}), \tag{7.6}$$

com

$$\lim_{\mathrm{x}\to\mathrm{p}} \frac{\mathrm{R}(\mathrm{p},\mathrm{x})}{||\mathrm{x}-\mathrm{p}||} = \lim_{\mathrm{x}\to\mathrm{p}} \frac{\mathrm{f}(\mathrm{x}) - \mathrm{l}(\mathrm{x})}{||\mathrm{x}-\mathrm{p}||} = \mathbf{0}.$$

Em outras palavras, é possível aproximar uma função vetorial f de classe C^1 na vizinhança de um ponto p pela transformação afim l e o erro $\mathrm{R}(\mathrm{p},\mathrm{x}) =$

$f(x) - l(x)$ cometido nesta aproximação tende a zero mais rapidamente do que $||x - p||$. A equação (7.6) pode ser reescrita como

$$\underbrace{f(p+h) - f(p)}_{V(h)} = \underbrace{Df(p) \cdot h}_{T(h)} + R(p,h).$$

Ela nos diz que a variação $V(h) = f(p+h) - f(p)$ do valor da função f no ponto p pode ser aproximada pela *transformação linear* $T\colon \mathbb{R}^n \to \mathbb{R}^m$ definida por $T(h) = Df(p) \cdot h$, isto é,

$$T(h_1, \ldots, h_n) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(p) & \cdots & \dfrac{\partial f_1}{\partial x_n}(p) \\ \dfrac{\partial f_2}{\partial x_1}(p) & \cdots & \dfrac{\partial f_2}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(p) & \cdots & \dfrac{\partial f_m}{\partial x_n}(p) \end{bmatrix}_{m \times n} \cdot \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix}_{n \times 1}$$

e que o erro $R(p,h) = V(h) - T(h)$ cometido nesta aproximação tende a zero mais rapidamente do que o $||h|| = \sqrt{h_1^2 + \cdots + h_n^2}$. A matriz associada a esta transformação linear é dada por

$$Df(p) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(p) & \cdots & \dfrac{\partial f_1}{\partial x_n}(p) \\ \dfrac{\partial f_2}{\partial x_1}(p) & \cdots & \dfrac{\partial f_2}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(p) & \cdots & \dfrac{\partial f_m}{\partial x_n}(p) \end{bmatrix}_{m \times n}.$$

Ela é denominada a *matriz jacobiana* (ou, mais simplesmente, *jacobiana*) de f no ponto p. A matriz jacobiana $Df(p)$ representa em Cálculo II o mesmo papel que a derivada $f'(p)$ representa em Cálculo I. Muitas idéias, teoremas e algoritmos são elaborados com o seu uso. A transformação linear induzida por $Df(p)$ é denominada a *derivada* de f no ponto p.

Novamente, T é a "melhor" transformação linear que aproxima a variação de f perto do ponto p ou, equivalentemente, l é a melhor transformação afim que aproxima f perto do ponto p. Também, l é a única transformação afim que satisfaz as condições

$$\mathrm{l}(\mathrm{p}) = \mathrm{f}(\mathrm{p}) \quad \text{e} \quad \frac{\partial l_i}{\partial x_j}(\mathrm{p}) = \frac{\partial f_i}{\partial x_j}(\mathrm{p}) \text{ para } i = 1, \ldots, m \text{ e } j = 1, \ldots, n.$$

Vamos resumir os resultados que obtivemos até agora em um teorema.

Teorema 7.4 (APROXIMAÇÃO LINEAR: O CASO VETORIAL) Considere uma função vetorial $\mathrm{f}: D \subset \mathbb{R}^n \to \mathbb{R}^m$ de classe C^1 e $\mathrm{p} = (p_1, \ldots, p_n)$ um ponto do interior de D. Então existe uma única transformação afim l que satisfaz as condições $\mathrm{l}(\mathrm{p}) = \mathrm{f}(\mathrm{p})$ e

$$\frac{\partial l_i}{\partial x_j}(\mathrm{p}) = \frac{\partial f_i}{\partial x_j}(\mathrm{p}) \quad \text{para } i = 1, \ldots, m \text{ e } j = 1, \ldots, m,$$

a saber,

$$\mathrm{l}(\mathrm{x}) = \mathrm{f}(\mathrm{p}) + D\mathrm{f}(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p}),$$

em que

$$D\mathrm{f}(\mathrm{p}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathrm{p}) \\ \frac{\partial f_2}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathrm{p}) \end{bmatrix}_{m \times n}$$

é a matriz jacobiana de f em p. Mais ainda, vale que

$$\lim_{\mathrm{x} \to \mathrm{p}} \frac{\mathrm{f}(\mathrm{x}) - \mathrm{l}(\mathrm{x})}{||\mathrm{x} - \mathrm{p}||} = 0,$$

isto é, a diferença entre f(x) e l(x) tende a zero mais rapidamente do que $||\mathrm{x} - \mathrm{p}||$. Em outras palavras,

$$\mathrm{f}(\mathrm{x}) = \mathrm{f}(\mathrm{p}) + D\mathrm{f}(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p}) + \mathrm{R}(\mathrm{p}, \mathrm{x}),$$

com

$$\lim_{\mathrm{x} \to \mathrm{p}} \frac{\mathrm{R}(\mathrm{p}, \mathrm{x})}{||\mathrm{x} - \mathrm{p}||} = 0.$$

Exemplo 7.4 Considere a função vetorial f: $\mathbb{R}^3 \to \mathbb{R}^2$ definida por

$$(y_1, y_2) = \mathrm{f}(x_1, x_2, x_3) = \left(\frac{x_1 + x_2 + x_3}{3}, \sqrt[3]{x_1 \cdot x_2 \cdot x_3} \right)$$

que foi apresentada no exemplo (6.3). As funções coordenadas de f são

$$y_1 = f_1(x_1, x_2, x_3) = \frac{x_1 + x_2 + x_3}{3} \quad \text{e} \quad y_2 = f_2(x_1, x_2, x_3) = \sqrt[3]{x_1 \cdot x_2 \cdot x_3}.$$

Portanto, a matriz jacobiana de f no ponto (x_1, x_2, x_3) é

$$\begin{aligned} D\mathrm{f}(x_1, x_2, x_3) &= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_1, x_2, x_3) & \frac{\partial f_1}{\partial x_2}(x_1, x_2, x_3) & \frac{\partial f_1}{\partial x_3}(x_1, x_2, x_3) \\ \frac{\partial f_2}{\partial x_1}(x_1, x_2, x_3) & \frac{\partial f_2}{\partial x_2}(x_1, x_2, x_3) & \frac{\partial f_2}{\partial x_3}(x_1, x_2, x_3) \end{bmatrix}_{2\times3} \\ &= \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ \dfrac{x_2 \cdot x_3}{3\sqrt[3]{x_1^2 \cdot x_2^2 \cdot x_3^2}} & \dfrac{x_1 \cdot x_3}{3\sqrt[3]{x_1^2 \cdot x_2^2 \cdot x_3^2}} & \dfrac{x_1 \cdot x_2}{3\sqrt[3]{x_1^2 \cdot x_2^2 \cdot x_3^2}} \end{bmatrix}_{2\times3}. \end{aligned}$$

Uma vez que $\mathrm{f}(1,2,4) = (7/3, 2)$ e

$$D\mathrm{f}(1,2,4) = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 1/3 & 1/6 \end{bmatrix}_{2\times3},$$

segue-se que a equação da transformação afim que melhor aproxima f nas proximidades do ponto $\mathrm{p} = (1,2,4)$ é dada por $\mathrm{l}(\mathrm{x}) = \mathrm{f}(\mathrm{p}) + D\mathrm{f}(\mathrm{p}) \cdot (\mathrm{x} - \mathrm{p})$, isto é,

$$\begin{aligned} \mathrm{l}(x_1, x_2, x_3) &= \begin{bmatrix} 7/3 \\ 2 \end{bmatrix} + \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & 1/3 & 1/6 \end{bmatrix} \cdot \begin{bmatrix} x_1 - 1 \\ x_2 - 2 \\ x_3 - 4 \end{bmatrix} \\ &= \left(x_1/3 + x_2/3 + x_3/3,\ 2\,x_1/3 + x_2/3 + x_3/6 \right). \end{aligned}$$

Podemos usar a transformação afim l para estimar o valor de $\mathrm{f}(x_1, x_2, x_3)$ em pontos (x_1, x_2, x_3) próximos de $(1,2,4)$. Por exemplo,

$$\mathrm{l}(1.1, 1.9, 3.9) = (2.3,\ 2.016\ldots),$$

enquanto que

$$f(1.1, 1.9, 3.9) = (2.3,\ 2.012\ldots).$$

Observe que $l_1(x_1, x_2, x_3) = f_1(x_1, x_2, x_3)$ para todo $(x_1, x_2, x_3) \in \mathbb{R}^3_+$. Isto não é uma coincidência, como veremos mais adiante.

$\square$

Vimos que a matriz jacobiana $Df(\mathrm{p})$ de uma função f em ponto p define a transformação linear que melhor aproxima a variação da função perto deste ponto. Mas o que aconteceria se a própria função já fosse uma transformação linear? Lembre-se que uma transformação linear T nada mais é do que uma função vetorial definida por uma matriz: $T(\mathrm{x}) = A \cdot \mathrm{x}$. Ora, a melhor transformação linear que aproxima a variação de uma transformação linear *é a própria transformação linear*. Como conseqüência, a matriz jacobiana de uma transformação linear é igual à matriz que define a própria transformação linear, isto é, $D\mathrm{T}(\mathrm{p}) = A$. Mais formalmente,

Teorema 7.5 Seja $\mathrm{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ uma transformação linear definida por uma matriz $A_{m\times n}$, isto é, seja

$$\mathrm{T}(\mathrm{x}) = A \cdot \mathrm{x}.$$

Então

$$D\mathrm{T}(\mathrm{p}) = A$$

para todo $\mathrm{p} \in \mathbb{R}^n$.

Demonstração: Para fixar as idéias, vamos fazer o caso $n = 3$ e $m = 2$, isto é,

$$\mathrm{T}(x_1, x_2, x_3) = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = (a\,x_1 + b\,x_2 + c\,x_3,\ d\,x_1 + e\,x_2 + f\,x_3).$$

Como as funções coordenadas de T são

$$\begin{aligned} T_1(x_1, x_2, x_3) &= a\,x_1 + b\,x_2 + c\,x_3, \\ T_2(x_1, x_2, x_3) &= d\,x_1 + e\,x_2 + f\,x_3, \end{aligned}$$

não é difícil de ver que

$$
\begin{aligned}
DT(p_1,p_2,p_3) &= \begin{bmatrix} \dfrac{\partial T_1}{\partial x_1}(p_1,p_2,p_3) & \dfrac{\partial T_1}{\partial x_2}(p_1,p_2,p_3) & \dfrac{\partial T_1}{\partial x_3}(p_1,p_2,p_3) \\ \dfrac{\partial T_2}{\partial x_1}(p_1,p_2,p_3) & \dfrac{\partial T_2}{\partial x_2}(p_1,p_2,p_3) & \dfrac{\partial T_2}{\partial x_3}(p_1,p_2,p_3) \end{bmatrix} \\
&= \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \\
&= A.
\end{aligned}
$$

O caso geral é demonstrado da mesma maneira. Se

$$
T(x_1,x_2,\ldots,x_n) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}
$$

então

$$
T(x_1,x_2,\ldots,x_n) = \left(\sum_{k=1}^{n} a_{1k}\cdot x_k, \sum_{k=1}^{n} a_{2k}\cdot x_k, \ldots, \sum_{k=1}^{n} a_{mk}\cdot x_k \right).
$$

Desta maneira, a i-ésima função coordenada de T é dada por

$$
T_i(x_1,x_2,\ldots,x_n) = \sum_{k=1}^{n} a_{ik}\cdot x_k = a_{i1}\cdot x_1 + a_{i2}\cdot x_2 + \cdots + a_{in}\cdot x_n
$$

de modo que, para calcularmos o elemento na posição ij da matriz jacobiana de T no ponto p, basta derivarmos T_i com relação a x_j no ponto p. Mas note que

$$
\frac{\partial T_i}{\partial x_j}(\mathrm{p}) = \frac{\partial}{\partial x_j}\left(a_{i1}\cdot x_1 + a_{i2}\cdot x_2 + \cdots + a_{in}\cdot x_n \right)\Big|_{\mathrm{x}=\mathrm{p}} = a_{ij},
$$

isto é, o elemento na posição ij da matriz jacobiana $DT(\mathrm{p})$ é igual ao elemento na posição ij da matriz A. Como as matrizes $DT(\mathrm{p})$ e A possuem os mesmos elementos, concluímos que elas são iguais. ∎

Observação: A matriz jacobiana de uma função escalar $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma matriz linha $1 \times n$:

$$Df(\mathrm{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathrm{x}) & \frac{\partial f}{\partial x_2}(\mathrm{x}) & \cdots & \frac{\partial f}{\partial x_n}(\mathrm{x}) \end{bmatrix}_{1\times n},$$

enquanto que a matriz jacobiana de uma curva parametrizada $\alpha\colon \mathbb{R} \to \mathbb{R}^n$, com $\alpha(t) = (\alpha_1(t), \alpha_2(t), \ldots, \alpha_n(t))$, é uma matriz coluna $n \times 1$:

$$D\alpha(t) = \begin{bmatrix} \alpha_1'(t) \\ \alpha_2'(t) \\ \vdots \\ \alpha_n'(t) \end{bmatrix}_{n\times 1}.$$

A matriz jacobiana de uma função compartilha das mesmas propriedades que a derivada de uma função de um variável que você aprendeu em Cálculo I. O próximo teorema resume estas propriedades

Teorema 7.6 (PROPRIEDADES DA MATRIZ JACOBIANA) Considere $\mathrm{f}\colon D_{\mathrm{f}} \subset \mathbb{R}^n \to \mathbb{R}^m$ e $\mathrm{g}\colon D_{\mathrm{g}} \subset \mathbb{R}^n \to \mathbb{R}^m$ duas funções vetoriais de classe C^k, $k \geq 0$. Então

(1) $\mathrm{f} + \mathrm{g}$ é de classe C^1 em $D_{\mathrm{f}} \cap D_{\mathrm{g}}$ e $D(\mathrm{f}+\mathrm{g})(\mathrm{p}) = D\mathrm{f}(\mathrm{p}) + D\mathrm{g}(\mathrm{p})$ para todo $\mathrm{p} \in D_{\mathrm{f}} \cap D_{\mathrm{g}}$.

(2) Se $c \in \mathbb{R}$ é uma constante real então $c \cdot \mathrm{f}$ é de classe C^1 em D_{f} e $D(c \cdot \mathrm{f})(\mathrm{p}) = c \cdot D\mathrm{f}(\mathrm{p})$ para todo $\mathrm{p} \in D_{\mathrm{f}}$.

Se $m = 1$, isto é, se f e g são funções escalares, então as propriedades abaixo também valem.

(3) $f \cdot g$ é de classe C^1 em $D_f \cap D_g$ e $D(f \cdot g)(\mathrm{p}) = g(\mathrm{p}) \cdot Df(\mathrm{p}) + f(\mathrm{p}) \cdot Dg(\mathrm{p})$ para todo $\mathrm{p} \in D_f \cap D_g$.

(4) f/g é de classe C^1 em $D_f \cap D_g - \{\mathrm{q} \in D_g \mid g(\mathrm{q}) = 0\}$ e

$$D\left(\frac{f}{g}\right)(\mathrm{p}) = \frac{g(\mathrm{p}) \cdot Df(\mathrm{p}) - f(\mathrm{p}) \cdot Dg(\mathrm{p})}{[g(\mathrm{p})]^2}$$

para todo $\mathrm{p} \in D_f \cap D_g - \{\mathrm{q} \in D_g \mid g(\mathrm{q}) = 0\}$.

Demonstração: Basta observar que as entradas da matriz jacobiana são formadas por derivadas parciais que, por sua vez, são obtidas através da derivação de uma função de uma variável. Como a operação de derivação em Cálculo I satisfaz todas as propriedades enunciadas acima, estas propriedades estendem-se naturalmente para a matriz jacobiana de funções vetoriais.

Vamos ilustrar esta idéia com a demonstração da primeira propriedade. Da definição de matriz jacobiana, temos que

$$D(\mathrm{f}+\mathrm{g})(\mathrm{p}) = \begin{bmatrix} \dfrac{\partial(f_1+g_1)}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial(f_1+g_1)}{\partial x_n}(\mathrm{p}) \\ \dfrac{\partial(f_2+g_2)}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial(f_2+g_2)}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial(f_m+g_m)}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial(f_m+g_m)}{\partial x_n}(\mathrm{p}) \end{bmatrix}.$$

Por outro lado, vale que $(\partial(f_i+g_i)/\partial x_j)(\mathrm{p}) = (\partial f_i/\partial x_j)(\mathrm{p}) + (\partial g_i/\partial x_j)(\mathrm{p})$ (estamos usando a propriedade de que a derivada parcial da soma é a soma das derivadas parciais de cada parcela) e, portanto,

$$\begin{aligned} D(\mathrm{f}+\mathrm{g})(\mathrm{p}) &= \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathrm{p}) + \dfrac{\partial g_1}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\mathrm{p}) + \dfrac{\partial g_1}{\partial x_n}(\mathrm{p}) \\ \dfrac{\partial f_2}{\partial x_1}(\mathrm{p}) + \dfrac{\partial g_2}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f_2}{\partial x_n}(\mathrm{p}) + \dfrac{\partial g_2}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\mathrm{p}) + \dfrac{\partial g_m}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f_m}{\partial x_n}(\mathrm{p}) + \dfrac{\partial g_m}{\partial x_n}(\mathrm{p}) \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f_1}{\partial x_n}(\mathrm{p}) \\ \dfrac{\partial f_2}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f_2}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial f_m}{\partial x_n}(\mathrm{p}) \end{bmatrix} + \begin{bmatrix} \dfrac{\partial g_1}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial g_1}{\partial x_n}(\mathrm{p}) \\ \dfrac{\partial g_2}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial g_2}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial g_m}{\partial x_1}(\mathrm{p}) & \cdots & \dfrac{\partial g_m}{\partial x_n}(\mathrm{p}) \end{bmatrix} \\ &= D\mathrm{f}(\mathrm{p}) + D\mathrm{g}(\mathrm{p}) \end{aligned}$$

onde, na penúltima igualdade, usamos a definição de soma de duas matrizes.

A demonstração das demais propriedades se faz de maneira análoga e a deixaremos como um exercício para você.

□

Exemplo 7.5 Considere $f(x, y, z) = x^2 + y^2 + z^2$, $g(x, y, z) = x^2 + 1$ e

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

Aplicando-se diretamente a definição de matriz jacobiana para h, vemos que

$$\begin{aligned} Dh(x, y, z) &= \begin{bmatrix} \dfrac{\partial h}{\partial x}(x, y, z) & \dfrac{\partial h}{\partial y}(x, y, z) & \dfrac{\partial h}{\partial z}(x, y, z) \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{2\,x\,(1 - y^2 - z^2)}{(x^2 + 1)^2} & \dfrac{2\,y}{x^2 + 1} & \dfrac{2\,z}{x^2 + 1} \end{bmatrix}. \end{aligned}$$

Por outro lado, aplicando-se a "regra do quociente", concluímos que

$$\begin{aligned} Dh(x, y, z) &= \frac{g(x, y, z) \cdot Df(x, y, z) - f(x, y, z) \cdot Dg(x, y, z)}{[g(x, y, z)]^2} \\ &= \frac{(x^2 + 1) \cdot \begin{bmatrix} 2\,x & 2\,y & 2\,z \end{bmatrix} - (x^2 + y^2 + z^2) \cdot \begin{bmatrix} 2\,x & 0 & 0 \end{bmatrix}}{(x^2 + 1)^2} \\ &= \begin{bmatrix} \dfrac{2\,x\,(1 - y^2 - z^2)}{(x^2 + 1)^2} & \dfrac{2\,y}{x^2 + 1} & \dfrac{2\,z}{x^2 + 1} \end{bmatrix}. \end{aligned}$$

Note que as duas maneiras conduzem ao mesmo resultado. □

Observação: Lembramos que a "regra do quociente" só vale para funções escalares (funções cujo contradomínio é um subconjunto de $\mathbb{R}$) uma vez que a divisão de funções vetoriais não está definida.

7.3 Composição de funções

continua

A partir de duas funções podemos criar novas funções através da soma, diferença, multiplicação ou divisão das funções iniciais. Um outra operação muito importante é a composição de funções.

Definição 7.1 (Composição de funções) Considere f: $A \to B$ e g: $C \to D$ com $f(A) \subset C$. A função h: $A \to D$ obtida aplicando-se primeiro f a x $\in A$ e então aplicando-se g ao resultado f(x) é denominada a *função composta* das funções g e f e denotada por g $\circ$ f:

$$(\mathrm{g} \circ \mathrm{f})(\mathrm{x}) = \mathrm{g}(\mathrm{f}(\mathrm{x})).$$

A figura (7.3) ilustra a composição de g com f.

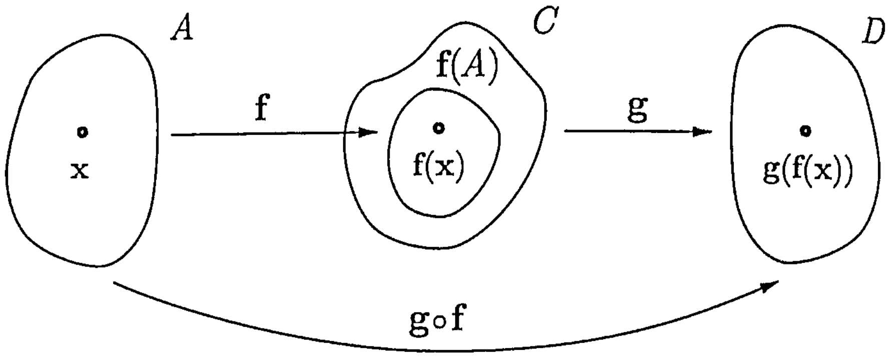

Figura 7.3: A composição de g com f.

Exemplo 7.6 A função $h(x) = \mathrm{sen}^2(x)$ é resultado da composição de $g(x) = x^2$ com $f(x) = \mathrm{sen}(x)$. A função $h(x) = (x+4)^2$ é resultado da composição de $g(x) = x^2$ com $f(x) = x+4$. A função $h(x,y) = \sqrt{x^2+y^2}$ é resultado da composição de $g(x) = \sqrt{x}$ com $f(x,y) = x^2+y^2$. ▫

Exemplo 7.7 Composições de funções aparecem naturalmente em *modelos dinâmicos*, isto é, modelos para os quais as variáveis dependem do tempo. Por exemplo, suponha que a função de produção Cobb-Douglas $z = f(x,y) = 4 \cdot x^{3/4} \cdot y^{1/4}$ represente a produção de uma empresa em função de seu capital x e de seu trabalho y. Suponha também que a curva parametrizada $\alpha(t) = (x(t), y(t)) = (10000 + 100 \cdot t, 625 + 5 \cdot t)$ descreva como o capital $x(t) = 10000 + 100 \cdot t$ e o trabalho $y(t) = 625 + 5 \cdot t$ desta empresa variam no tempo. Então, se quisermos estudar como a produção desta empresa varia no tempo, somos naturalmente levados a considerar a função composta

$$h(t) = (f \circ \alpha)(t) = 4 \cdot (10000 + 100 \cdot t)^{3/4} \cdot (625 + 5 \cdot t)^{1/4}.$$ ▫

É muito importante que você adquira a habilidade de "enxergar" quem está sendo composto com quem, identificando o domínio e o contradomínio de cada função em uma composição, principalmente no caso em que uma ou todas as funções não são dadas explicitamente.

Mais adiante veremos como expressar a matriz jacobiana de uma composição em termos das matrizes jacobianas das funções que fazem parte da composição. Isto resultará na *regra da cadeia*, um teorema que será utilizado com bastante freqüência na teoria que vamos desenvolver. Mais ainda, você certamente usará a regra da cadeia em qualquer disciplina em que uma derivada apareça: Microeconomia, Cálculo III, Cálculo IV, Física, Equações Diferenciais, etc.

O segredo na utilização da regra da cadeia está na identificação correta das funções que fazem parte da composição! Vamos ver alguns exemplos.

Exemplo 7.8 A função $z = f(x(t), y(t))$ é uma função de $\mathbb{R}$ para $\mathbb{R}$ que pode ser descrita como a composição da função escalar $z = f(x, y)$ de $\mathbb{R}^2$ para $\mathbb{R}$ com a curva parametrizada $\alpha(t) = (x(t), y(t))$ de $\mathbb{R}$ para $\mathbb{R}^2$. Mais formalmente, se $z = h(t) = f(x(t), y(t))$, então $h = f \circ \alpha$. $\square$

Exemplo 7.9 A função $z = f(x(u, v), y(u, v))$ é uma função de $\mathbb{R}^2$ para $\mathbb{R}$ que pode ser descrita como a composição da função escalar $z = f(x, y)$ de $\mathbb{R}^2$ para $\mathbb{R}$ com a função vetorial $\mathrm{g}(u, v) = (x(u, v), y(u, v))$ de $\mathbb{R}^2$ para $\mathbb{R}^2$. Mais formalmente, se $z = h(u, v) = f(x(u, v), y(u, v))$, então $h = f \circ \mathrm{g}$. $\square$

Exemplo 7.10 A função $z = f(u + v, u - v + t)$ é uma função de $\mathbb{R}^3$ para $\mathbb{R}$ que pode ser descrita como a composição da função escalar $z = f(x, y)$ de $\mathbb{R}^2$ para $\mathbb{R}$ com a função vetorial $\mathrm{g}(u, v, t) = (u+v, u-v+t)$ de $\mathbb{R}^3$ para $\mathbb{R}^2$. Mais formalmente, se $z = h(u, v, t) = f(u + v, u - v + t)$ então $h = f \circ \mathrm{g}$. Observe que g é uma transformação linear de $\mathbb{R}^3$ para $\mathbb{R}^2$ (por que?). $\square$

Exemplo 7.11 Sejam p um ponto e v um vetor de $\mathbb{R}^n$. A função $z = f(\mathrm{p}+t\cdot\mathrm{v})$ é uma função de $\mathbb{R}$ para $\mathbb{R}$ que pode ser descrita como a composição da função escalar $z = f(\mathrm{x})$ de $\mathbb{R}^n$ para $\mathbb{R}$ com a curva parametrizada $\alpha(t) = \mathrm{p} + t \cdot \mathrm{v}$ de $\mathbb{R}$ para $\mathbb{R}^n$ (α representa a reta que passa pelo p na direção do vetor v). Mais formalmente, se $z = h(t) = f(\mathrm{p} + t \cdot \mathrm{v})$, então $h = f \circ \alpha$. $\square$

7.4 Lembrando Cálculo I: a regra da cadeia

Considere $f\colon D_f \subset \mathbb{R} \to \mathbb{R}$ e $g\colon D_g \subset \mathbb{R} \to \mathbb{R}$ duas funções de classe C^k, com $k \geq 1$, tais que $f(D_f) \subset D_g$. Podemos então construir a composição $h = g \circ f\colon D_h = D_f \subset \mathbb{R} \to \mathbb{R}$.

A *regra da cadeia* afirma que h também é uma função de classe C^k e que sua derivada $h' = (g \circ f)'$ pode ser obtida em termos das derivadas f' e g' de f e g. Mais precisamente,

$$h'(p) = (g \circ f)'(p) = g'(f(p)) \cdot f'(p). \tag{7.7}$$

Exemplo 7.12 A função $h(x) = \operatorname{sen}(x^3 + 4x)$ pode ser escrita como a composição de duas funções: $h(x) = (g \circ f)(x)$, com $f(x) = x^3 + 4x$ e $g(z) = \operatorname{sen}(z)$. Como f e g são funções de classe C^∞ segue-se que h também é uma função de classe C^∞. Vamos usar a regra da cadeia para calcular a derivada de h. Para isto, precisamos calcular as derivadas de f e g. A derivada de g é

$$g'(z) = \cos(z)$$

e a derivada de f é

$$f'(x) = 3\,x^2 + 4.$$

Desta maneira, pela regra da cadeia, a derivada de h é

$$h'(x) = g'(f(x)) \cdot f'(x) = \cos(f(x)) \cdot f'(x) = \cos(x^3 + 4\,x) \cdot \left(3\,x^2 + 4\right). \quad \blacksquare$$

7.5 A regra da cadeia em Cálculo II

Como veremos, a regra da cadeia em Cálculo II é muito parecida com a regra da cadeia de Cálculo I se utilizarmos a notação matricial.

Teorema 7.7 (A REGRA DA CADEIA) Considere $\mathrm{f}\colon D_{\mathrm{f}} \subset \mathbb{R}^n \to \mathbb{R}^m$ e $\mathrm{g}\colon D_{\mathrm{g}} \subset \mathbb{R}^m \to \mathbb{R}^k$ duas funções de classe C^l tais que $\mathrm{f}(D_{\mathrm{f}}) \subset D_{\mathrm{g}}$ de modo que podemos construir a composição $\mathrm{h} = \mathrm{g} \circ \mathrm{f}\colon D_{\mathrm{h}} = D_{\mathrm{f}} \subset \mathbb{R}^n \to \mathbb{R}^k$. Então h também é uma função de classe C^k e a matriz jacobiana de h em p é igual ao produto da matriz jacobiana de g em f(p) pela matriz jacobiana de f em p:

$$D\mathrm{h}(\mathrm{p}) = D(\mathrm{g} \circ \mathrm{f})(\mathrm{p}) = D\mathrm{g}(\mathrm{f}(\mathrm{p})) \cdot D\mathrm{f}(\mathrm{p}). \tag{7.8}$$

Ou, mais explicitamente, se

$$\mathrm{x} = (x_1, \ldots, x_n), \qquad \mathrm{y} = (y_1, \ldots, y_m)$$

e

$$\mathrm{f}(\mathrm{x}) = (f_1(\mathrm{x}), \ldots, f_m(\mathrm{x})), \qquad \mathrm{g}(\mathrm{y}) = (g_1(\mathrm{y}), \ldots, g_l(\mathrm{y})),$$

então a matriz jacobiana de $\mathrm{h}(\mathrm{x}) = (h_1(\mathrm{x}), \ldots, h_l(\mathrm{x}))$ é dada por

$$\underbrace{\begin{bmatrix} \frac{\partial h_1}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial h_1}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_l}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial h_l}{\partial x_n}(\mathrm{p}) \end{bmatrix}}_{D\mathrm{h}(\mathrm{p})} =$$

$$\underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial y_1}(\mathrm{f}(\mathrm{p})) & \cdots & \frac{\partial g_1}{\partial y_m}(\mathrm{f}(\mathrm{p})) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_l}{\partial y_1}(\mathrm{f}(\mathrm{p})) & \cdots & \frac{\partial g_l}{\partial y_m}(\mathrm{f}(\mathrm{p})) \end{bmatrix}}_{D\mathrm{g}(\mathrm{f}(\mathrm{p}))} \cdot \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathrm{p}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathrm{p}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathrm{p}) \end{bmatrix}}_{D\mathrm{f}(\mathrm{p})}.$$

Observe que $D\mathrm{h}(\mathrm{p})$ é uma matriz $l \times n$, $D\mathrm{g}(\mathrm{f}(\mathrm{p}))$ é uma matriz $l \times m$ e $D\mathrm{f}(\mathrm{p})$ é uma matriz $m \times n$, de modo que a multiplicação de matrizes na expressão (7.8) pode ser efetuada.

Note a semelhança entre as expressões (7.7) e (7.8): as derivadas $h'(p)$, $g'(f(p))$ e $f'(p)$ (que em Cálculo I são números reais) são substituídas, respectivamente, pelas matrizes jacobianas $D\mathrm{h}(\mathrm{p})$, $D\mathrm{g}(\mathrm{f}(\mathrm{p}))$ e $D\mathrm{f}(\mathrm{p})$. Evidentemente, o símbolo $\cdot$ na expressão (7.8) indica a multiplicação de matrizes, enquanto que na expressão (7.7) ele indica a multiplicação de números reais.

Uma demonstração formal da regra da cadeia pode ser encontrada nas referências [01, 71] , contudo, vamos tentar fornecer uma justificativa mais informal com base na idéia de aproximação linear desenvolvida no início do capítulo. Pelo teorema (7.4), as variações de f em p e de g em f(p) podem ser aproximadas pelas transformações lineares T e S definidas, respectivamente, pelas matrizes jacobianas $D\mathrm{f}(\mathrm{p})$ e $D\mathrm{g}(\mathrm{f}(\mathrm{p}))$. A composição $\mathrm{S} \circ \mathrm{T}$ destas transformações lineares deve fornecer uma "boa" aproximação da variação da composição $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$ no ponto $\mathbf{p}$. Mas, pelo exercício [22], a composição de duas transformações lineares é também uma transformação linear e a matriz associada à função composta nada mais é do que o produto das matrizes que define cada transformação linear na composição. Desta maneira, é natural esperarmos que a matriz jacobiana $D\mathbf{h}(\mathbf{p})$ seja dada pelo produto das matrizes jacobianas $D\mathbf{g}(\mathbf{f}(\mathbf{p}))$ e $D\mathbf{f}(\mathbf{p})$.

Vamos fazer uma revisão desta justificativa informal usando a notação matricial: já sabemos, pelo teorema (7.4), que

$$\mathbf{g}(\mathbf{y}) \approx \mathbf{g}(\mathbf{q}) + D\mathbf{g}(\mathbf{q}) \cdot (\mathbf{y} - \mathbf{q}),$$

Agora, fazendo $\mathbf{q} = \mathbf{f}(\mathbf{p})$ e $\mathbf{y} = \mathbf{f}(\mathbf{x})$, obtemos

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) \approx \mathbf{g}(\mathbf{f}(\mathbf{p})) + D\mathbf{g}(\mathbf{f}(\mathbf{p})) \cdot (\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{p})). \tag{7.9}$$

Por outro lado, novamente pelo teorema (7.4), temos que

$$\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{p}) + D\mathbf{f}(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}).$$

Substituindo esta expressão no lado direito de (7.9), vemos que

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) \approx \mathbf{g}(\mathbf{f}(\mathbf{p})) + D\mathbf{g}(\mathbf{f}(\mathbf{p})) \cdot (\mathbf{f}(\mathbf{p}) + D\mathbf{f}(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}) - \mathbf{f}(\mathbf{p})),$$

isto é,

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) \approx \mathbf{g}(\mathbf{f}(\mathbf{p})) + D\mathbf{g}(\mathbf{f}(\mathbf{p})) \cdot D\mathbf{f}(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}). \tag{7.10}$$

Aplicando o teorema (7.4) mais uma vez, agora para $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$, segue-se que

$$\mathbf{h}(\mathbf{x}) \approx \mathbf{h}(\mathbf{p}) + D\mathbf{h}(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}),$$

isto é,

$$\mathbf{g}(\mathbf{f}(\mathbf{x})) \approx \mathbf{g}(\mathbf{f}(\mathbf{p})) + D\mathbf{h}(\mathbf{p}) \cdot (\mathbf{x} - \mathbf{p}). \tag{7.11}$$

Comparando-se as expressões (7.10) e (7.11), não é difícil de se suspeitar que

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{p}) = D\mathbf{g}(\mathbf{f}(\mathbf{p})) \cdot D\mathbf{f}(\mathbf{p}).$$

Exercício resolvido 7.1 Considere as funções vetoriais

$$\begin{array}{rccl} \mathbf{f}: & \mathbb{R}^2 & \to & \mathbb{R}^2 \\ & (x, y) & \mapsto & \mathbf{f}(x, y) = \left(x^2 - y^2, x \cdot y\right) \end{array} \quad \text{e}$$

$$\begin{array}{rccl} \mathrm{g}: & \mathbb{R}^2 & \to & \mathbb{R}^3 \\ & (u,v) & \mapsto & \mathrm{g}(u,v) = (\operatorname{sen}(u\cdot v), u+v, u-v) \end{array}$$

Encontre a matriz jacobiana da função composta $\mathrm{h} = \mathrm{g} \circ \mathrm{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$ no ponto $\mathrm{p} = (1,1)$.

SOLUÇÃO: Pela regra da cadeia, para encontrarmos $D\mathrm{h}(1,1)$, basta calcularmos $D\mathrm{g}(\mathrm{f}(1,1))$, $D\mathrm{f}(1,1)$ e, em seguida, efetuarmos a multiplicação de matrizes $D\mathrm{g}(\mathrm{f}(1,1)) \cdot D\mathrm{f}(1,1)$. As funções coordenadas de f são

$$f_1(x,y) = x^2 - y^2 \qquad \text{e} \qquad f_2(x,y) = x \cdot y,$$

de modo que

$$D\mathrm{f}(x,y) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x}(x,y) & \dfrac{\partial f_1}{\partial y}(x,y) \\ \dfrac{\partial f_2}{\partial x}(x,y) & \dfrac{\partial f_2}{\partial y}(x,y) \end{bmatrix} = \begin{bmatrix} 2\cdot x & -2\cdot y \\ y & x \end{bmatrix}.$$

As funções coordenadas de g são

$$g_1(u,v) = \operatorname{sen}(u\cdot v), \qquad g_2(u,v) = u+v \qquad \text{e} \qquad g_3(u,v) = u-v,$$

de modo que

$$D\mathrm{g}(u,v) = \begin{bmatrix} \dfrac{\partial g_1}{\partial u}(u,v) & \dfrac{\partial g_1}{\partial v}(u,v) \\ \dfrac{\partial g_2}{\partial u}(u,v) & \dfrac{\partial g_2}{\partial v}(u,v) \\ \dfrac{\partial g_3}{\partial u}(u,v) & \dfrac{\partial g_3}{\partial v}(u,v) \end{bmatrix} = \begin{bmatrix} v\cdot\cos(u\cdot v) & u\cdot\cos(u\cdot v) \\ 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

No ponto $(x,y) = \mathrm{p} = (1,1)$ temos

$$D\mathrm{f}(1,1) = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}$$

e, no ponto $(u,v) = \mathrm{f}(\mathrm{p}) = \mathrm{f}(1,1) = (1^2 - 1^2, 1\cdot 1) = (0,1)$, temos

$$Dg(f(1,1)) = Dg(0,1) = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Logo, $Dh(1,1) = D(g \circ f)(1,1) = Dg(f(1,1)) \cdot Df(1,1) = Dg(0,1) \cdot Df(1,1)$, isto é,

$$Dh(1,1) = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}}_{Dg(f(1,1))} \cdot \underbrace{\begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}}_{Df(1,1)} = \begin{bmatrix} 2 & -2 \\ 3 & -1 \\ 1 & -3 \end{bmatrix}.$$

Em particular, se h_1, h_2 e h_3 são as funções coordenadas de h, então

$$\frac{\partial h_1}{\partial x}(1,1) = 2, \qquad \frac{\partial h_1}{\partial y}(1,1) = -2,$$

$$\frac{\partial h_2}{\partial x}(1,1) = 3, \qquad \frac{\partial h_2}{\partial y}(1,1) = -1,$$

$$\frac{\partial h_3}{\partial x}(1,1) = 1, \qquad \frac{\partial h_3}{\partial y}(1,1) = -3.$$

Evidentemente, você poderia calcular a matriz jacobiana $Dh(1,1)$ sem a regra da cadeia. Basta efetuar primeiro a composição:

$$\begin{aligned} h(x,y) &= (g \circ f)(x,y) = g(f(x,y)) = g(x^2 - y^2, x \cdot y) \\ &= \left(\operatorname{sen}\left((x^2 - y^2) \cdot x \cdot y\right), x^2 - y^2 + x \cdot y, x^2 - y^2 - x \cdot y\right), \end{aligned}$$

depois calcular a matriz jacobiana $Dh(x,y)$ diretamente:

$$Dh(x,y) = \begin{bmatrix} \frac{\partial h_1}{\partial x}(x,y) & \frac{\partial h_1}{\partial y}(x,y) \\ \frac{\partial h_2}{\partial x}(x,y) & \frac{\partial h_2}{\partial y}(x,y) \\ \frac{\partial h_3}{\partial x}(x,y) & \frac{\partial h_3}{\partial y}(x,y) \end{bmatrix}$$

$$= \begin{bmatrix} \left(3\,x^2\,y - y^3\right)\cos\left(\left(x^2-y^2\right)xy\right) & \left(x^3 - 3\,x\,y^2\right)\cos\left(\left(x^2-y^2\right)xy\right) \\ 2\,x + y & -2\,y + x \\ 2\,x - y & -2\,y - x \end{bmatrix}$$

e, finalmente, substituir os valores $x = 1$ e $y = 1$:

$$D\mathrm{h}(1,1) = \begin{bmatrix} 2 & -2 \\ 3 & -1 \\ 1 & -3 \end{bmatrix}.$$

Você poderia se perguntar: por que calcular a derivada (matriz jacobiana) de uma função composta pela regra da cadeia se é possível efetuar a composição e derivar o resultado diretamente? Neste exercício resolvido, de fato, não vemos a vantagem de um método com relação ao outro. Por outro lado, nem sempre temos explicitamente as funções que fazem parte da composição (veja os exemplos (7.8), (7.9), (7.10) e (7.11)). Por isto, a regra da cadeia é uma ferramenta extremamente útil quando estamos querendo estabelecer resultados teóricos. □

Exercício resolvido 7.2 Sejam $\mathrm{f}\colon \mathbb{R}^2 \to \mathbb{R}^2$ e $\mathrm{g}\colon \mathbb{R}^2 \to \mathbb{R}^3$ as funções vetoriais definidas no exercício resolvido anterior. Encontre a matriz jacobiana da função composta $\mathrm{h} = \mathrm{g} \circ \mathrm{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$ em um ponto $\mathrm{p} = (x, y)$ qualquer.

SOLUÇÃO: Pela regra da cadeia, para encontrarmos $D\mathrm{h}(x,y)$, basta calcularmos $D\mathrm{g}(\mathrm{f}(x,y))$, $D\mathrm{f}(x,y)$ e, em seguida, efetuarmos a multiplicação de matrizes $D\mathrm{g}(\mathrm{f}(x,y)) \cdot D\mathrm{f}(x,y)$. Como vimos no exercício resolvido anterior,

$$D\mathrm{f}(x,y) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x}(x,y) & \dfrac{\partial f_1}{\partial y}(x,y) \\ \dfrac{\partial f_2}{\partial x}(x,y) & \dfrac{\partial f_2}{\partial y}(x,y) \end{bmatrix} = \begin{bmatrix} 2\cdot x & -2\cdot y \\ y & x \end{bmatrix}.$$

e

$$D\mathrm{g}(u,v) = \begin{bmatrix} \dfrac{\partial g_1}{\partial u}(u,v) & \dfrac{\partial g_1}{\partial v}(u,v) \\ \dfrac{\partial g_2}{\partial u}(u,v) & \dfrac{\partial g_2}{\partial v}(u,v) \\ \dfrac{\partial g_3}{\partial u}(u,v) & \dfrac{\partial g_3}{\partial v}(u,v) \end{bmatrix} = \begin{bmatrix} v\cdot\cos(u\cdot v) & u\cdot\cos(u\cdot v) \\ 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

concluímos que

$$\begin{aligned}\frac{\partial h}{\partial u}(u,v) &= \frac{\partial f}{\partial x}(\mathrm{g}(u,v))\cdot\frac{\partial x}{\partial u}(u,v)+\frac{\partial f}{\partial y}(\mathrm{g}(u,v))\cdot\frac{\partial y}{\partial u}(u,v),\\ \frac{\partial h}{\partial v}(u,v) &= \frac{\partial f}{\partial x}(\mathrm{g}(u,v))\cdot\frac{\partial x}{\partial v}(u,v)+\frac{\partial f}{\partial y}(\mathrm{g}(u,v))\cdot\frac{\partial y}{\partial v}(u,v).\end{aligned}$$

▮

Exercício resolvido 7.5 Use a regra da cadeia para calcular as derivadas parciais de primeira ordem da função $z = h(u,v,t) = f(u+v, u-v+t)$ do exemplo (7.10). Assuma que f seja uma função de classe C^1.

SOLUÇÃO: A função $z = h(u,v,t) = f(u+v, u-v+t)$ pode ser descrita como a composição da função escalar $z = f(x,y)$ de $\mathbb{R}^2$ para $\mathbb{R}$ com a função vetorial $\mathrm{g}(u,v,t) = (u+v, u-v+t)$ de $\mathbb{R}^3$ para $\mathbb{R}^2$. Pela regra da cadeia, h também é uma função de classe C^1 e $Dh(u,v,t) = Df(\mathrm{g}(u,v,t))\cdot D\mathrm{g}(u,v,t)$. Como as funções coordenadas de g são dadas por

$$g_1(u,v,t) = u+v \qquad \text{e} \qquad g_2(u,v,t) = u-v+t,$$

segue-se que

$$\begin{aligned}D\mathrm{g}(u,v,t) &= \begin{bmatrix}\frac{\partial g_1}{\partial u}(u,v,t) & \frac{\partial g_1}{\partial v}(u,v,t) & \frac{\partial g_1}{\partial t}(u,v,t)\\ \frac{\partial g_2}{\partial u}(u,v,t) & \frac{\partial g_2}{\partial v}(u,v,t) & \frac{\partial g_2}{\partial t}(u,v,t)\end{bmatrix}\\ &= \begin{bmatrix}1 & 1 & 0\\ 1 & -1 & 1\end{bmatrix}.\end{aligned}$$

Uma vez que

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

temos

$$\begin{aligned}Dh(u,v,t) &= Df(\mathrm{g}(u,v,t))\cdot D\mathrm{g}(u,v,t)\\ &= \begin{bmatrix}\frac{\partial f}{\partial x}(\mathrm{g}(u,v,t)) & \frac{\partial f}{\partial y}(\mathrm{g}(u,v,t))\end{bmatrix}\cdot\begin{bmatrix}1 & 1 & 0\\ 1 & -1 & 1\end{bmatrix}\\ &= \begin{bmatrix}\frac{\partial f}{\partial x}(u+v,u-v+t) & \frac{\partial f}{\partial y}(u+v,u-v+t)\end{bmatrix}\cdot\begin{bmatrix}1 & 1 & 0\\ 1 & -1 & 1\end{bmatrix}.\end{aligned}$$

Efetuando-se a multiplicação de matrizes na expressão anterior e lembrando que, por definição,

$$Dh(u,v,t) = \left[\begin{array}{ccc} \frac{\partial h}{\partial u}(u,v,t) & \frac{\partial h}{\partial v}(u,v,t) & \frac{\partial h}{\partial t}(u,v,t) \end{array} \right],$$

concluímos que

$$\begin{aligned}
\frac{\partial h}{\partial u}(u+v,u-v+t) &= \frac{\partial f}{\partial x}(u+v,u-v+t) + \frac{\partial f}{\partial y}(u+v,u-v+t),\\
\frac{\partial h}{\partial v}(u+v,u-v+t) &= \frac{\partial f}{\partial x}(u+v,u-v+t) - \frac{\partial f}{\partial y}(u+v,u-v+t),\\
\frac{\partial h}{\partial t}(u+v,u-v+t) &= \frac{\partial f}{\partial y}(u+v,u-v+t).
\end{aligned}$$

□

Exercício resolvido 7.6 Use a regra da cadeia para calcular a derivada da função $z = h(t) = f(\mathrm{p} + t\cdot \mathrm{v})$ do exemplo (7.11) em $t = 0$. Assuma que f seja uma função de classe C^1.

SOLUÇÃO: A função $z = h(t) = f(\mathrm{p} + t\cdot \mathrm{v})$ pode ser descrita como a composição da função escalar $z = f(\mathrm{x}) = f(x_1,\ldots,x_n)$ de $\mathbb{R}^n$ para $\mathbb{R}$ com a curva parametrizada de $\mathbb{R}$ para $\mathbb{R}^n$

$$\alpha(t) = \mathrm{p} + t\cdot \mathrm{v} = (p_1,\ldots,p_n) + t\cdot(v_1,\ldots,v_n) = (p_1 + t\cdot v_1,\ldots,p_n + t\cdot v_n).$$

Pela regra da cadeia, h também é uma função de classe C^1 e $Dh(0) = Df(\alpha(0))\cdot D\alpha(0)$. Como as funções coordenadas de α são dadas por $\alpha_1(t) = p_1 + t\cdot v_1$, $\alpha_2(t) = p_2 + t\cdot v_2$, …, $\alpha_n(t) = p_n + t\cdot v_n$, segue-se que

$$D\alpha(t) = \begin{bmatrix} \alpha_1'(t) \\ \vdots \\ \alpha_n'(t) \end{bmatrix}_{n\times 1} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}_{n\times 1}.$$

Uma vez que

$$Df(x_1,\ldots,x_n) = \left[\begin{array}{ccc} \frac{\partial f}{\partial x_1}(x_1,\ldots,x_n) & \cdots & \frac{\partial f}{\partial x_n}(x_1,\ldots,x_n) \end{array} \right]_{1\times n}$$

e $\alpha(0) = p + 0 \cdot v = p$, temos

$$Dh(0) = Df(\alpha(0)) \cdot D\alpha(0) = Df(p) \cdot D\alpha(0)$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) & \cdots & \frac{\partial f}{\partial x_n}(p) \end{bmatrix}_{1\times n} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}_{n\times 1}$$

$$= \begin{bmatrix} \frac{\partial f}{\partial x_1}(p) \cdot v_1 + \cdots + \frac{\partial f}{\partial x_n}(p) \cdot v_n \end{bmatrix}_{1\times 1}.$$

Lembrando que, por definição,

$$Dh(0) = \begin{bmatrix} \frac{dh}{dt}(0) \end{bmatrix}_{1\times 1},$$

concluímos que

$$h'(0) = \frac{dh}{dt}(0) = \frac{\partial f}{\partial x_1}(p) \cdot v_1 + \cdots + \frac{\partial f}{\partial x_n}(p) \cdot v_n = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot v_i.$$

∎

Observação: a maioria dos livros de Cálculo II não enuncia a regra da cadeia com o uso da notação matricial. O que eles fazem é trocar a versão matricial da regra da cadeia

$$\underbrace{\begin{bmatrix} \frac{\partial h_1}{\partial x_1}(p) & \cdots & \frac{\partial h_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial h_k}{\partial x_1}(p) & \cdots & \frac{\partial h_k}{\partial x_n}(p) \end{bmatrix}}_{Dh(p)} =$$

$$\underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial y_1}(f(p)) & \cdots & \frac{\partial g_1}{\partial y_m}(f(p)) \\ \vdots & \ddots & \vdots \\ \frac{\partial g_k}{\partial y_1}(f(p)) & \cdots & \frac{\partial g_k}{\partial y_m}(f(p)) \end{bmatrix}}_{Dg(f(p))} \cdot \underbrace{\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(p) & \cdots & \frac{\partial f_1}{\partial x_n}(p) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(p) & \cdots & \frac{\partial f_m}{\partial x_n}(p) \end{bmatrix}}_{Df(p)}$$

(estabelecida para a função composta $h = g \circ f$) por expressões em termos das entradas das matrizes envolvidas. Mais especificamente, a expressão matricial acima diz, por exemplo, que o elemento da primeira linha e primeira coluna da matriz jacobiana $D\mathrm{h}(\mathrm{p})$ pode ser obtido multiplicando-se a primeira linha da matriz jacobiana $D\mathrm{g}(\mathrm{f}(\mathrm{p}))$ pela primeira coluna da matriz jacobiana $D\mathrm{f}(\mathrm{p})$:

$$\begin{aligned}\frac{\partial h_1}{\partial x_1}(\mathrm{p}) &= \frac{\partial g_1}{\partial y_1}(\mathrm{f}(\mathrm{p}))\cdot\frac{\partial f_1}{\partial x_1}(\mathrm{p})+\cdots+\frac{\partial g_1}{\partial y_m}(\mathrm{f}(\mathrm{p}))\cdot\frac{\partial f_m}{\partial x_1}(\mathrm{p}) \\ &= \sum_{k=1}^{m}\frac{\partial g_1}{\partial y_k}(\mathrm{f}(\mathrm{p}))\cdot\frac{\partial f_k}{\partial x_1}(\mathrm{p}).\end{aligned} \tag{7.12}$$

Podemos obter $k\cdot n$ expressões deste tipo considerando-se os $k\cdot n$ elementos da matriz jacobiana $D\mathrm{h}(\mathrm{p})$, isto é, para cada $i=1,\dots,k$ e $j=1,\dots,n$, o elemento da i-ésima linha e j-ésima coluna da matriz jacobiana $D\mathrm{h}(\mathrm{p})$ pode ser obtido pela multiplicação da i-ésima linha da matriz jacobiana $D\mathrm{g}(\mathrm{f}(\mathrm{p}))$ pela j-ésima coluna da matriz jacobiana $D\mathrm{f}(\mathrm{p})$):

$$\begin{aligned}\frac{\partial h_i}{\partial x_j}(\mathrm{p}) &= \frac{\partial g_i}{\partial y_1}(\mathrm{f}(\mathrm{p}))\cdot\frac{\partial f_1}{\partial x_j}(\mathrm{p})+\cdots+\frac{\partial g_i}{\partial y_m}(\mathrm{f}(\mathrm{p}))\cdot\frac{\partial f_m}{\partial x_j}(\mathrm{p}) \\ &= \sum_{k=1}^{m}\frac{\partial g_i}{\partial y_k}(\mathrm{f}(\mathrm{p}))\cdot\frac{\partial f_k}{\partial x_j}(\mathrm{p}).\end{aligned} \tag{7.13}$$

Assim, em resumo, o que a maioria dos livros de Cálculo II faz é trocar uma única equação matricial (isto é, $D\mathrm{h}(\mathrm{p})=D\mathrm{g}(\mathrm{f}(\mathrm{p}))\cdot D\mathrm{f}(\mathrm{p})$) pelas $k\cdot n$ equações reais definidas em (7.13).

CUIDADO! CUIDADO! CUIDADO!

É preciso tomar muito cuidado com o seguinte abuso de notação que pode ocorrer com a regra da cadeia. Considere funções $z=f(x,y)$ e $\mathrm{g}(u,v)=(g_1(u,v),g_2(u,v))$ de classe C^1. Pela regra da cadeia, a função composta

$$z=h(u,v)=(f\circ g)(u,v)=f(\mathrm{g}(u,v)).$$

é de classe C^1 e

$$\frac{\partial h}{\partial u}(u,v)=\frac{\partial f}{\partial x}(\mathrm{g}(u,v))\cdot\frac{\partial g_1}{\partial u}(u,v)+\frac{\partial f}{\partial y}(\mathrm{g}(u,v))\cdot\frac{\partial g_2}{\partial u}(u,v), \tag{7.14}$$

$$\frac{\partial h}{\partial v}(u,v)=\frac{\partial f}{\partial x}(\mathrm{g}(u,v))\cdot\frac{\partial g_1}{\partial v}(u,v)+\frac{\partial f}{\partial y}(\mathrm{g}(u,v))\cdot\frac{\partial g_2}{\partial v}(u,v). \tag{7.15}$$

Um abuso de notação cometido por vários autores é o de se omitir os pontos onde as derivadas parciais são calculadas, com a justificativa de economia de espaço nas fórmulas. Desta maneira, as expressões (7.14) e (7.15), com este abuso de notação, escrevem-se da seguinte forma:

$$\frac{\partial h}{\partial u} = \frac{\partial f}{\partial x} \cdot \frac{\partial g_1}{\partial u} + \frac{\partial f}{\partial y} \cdot \frac{\partial g_2}{\partial u}, \tag{7.16}$$

$$\frac{\partial h}{\partial v} = \frac{\partial f}{\partial x} \cdot \frac{\partial g_1}{\partial v} + \frac{\partial f}{\partial y} \cdot \frac{\partial g_2}{\partial v}. \tag{7.17}$$

Outro abuso de notação é o se trocar o nome da função pela variável que representa o valor da função em um ponto do domínio. Em nosso exemplo, isto corresponde a trocar f por z, h por w, g_1 por x e g_2 por y:

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial u} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial u}, \tag{7.18}$$

$$\frac{\partial z}{\partial v} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial v} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial v}. \tag{7.19}$$

Contudo, esta notação, apesar de compacta, pode trazer confusão em certos casos. Por exemplo, considere a situação onde $z = f(x,y)$ e $\mathrm{g}(x,y) = (g_1(x,y), g_2(x,y))$. Se escrevermos a regra da cadeia para a função composta $z = h(x,y) = (f \circ \mathrm{g})(x,y) = f(\mathrm{g}(x,y)) = f(g_1(x,y), g_2(x,y))$ com estes abusos de notação, vamos obter as seguintes expressões:

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial x} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial x}, \tag{7.20}$$

$$\frac{\partial z}{\partial y} = \frac{\partial z}{\partial x} \cdot \frac{\partial x}{\partial y} + \frac{\partial z}{\partial y} \cdot \frac{\partial y}{\partial y}. \tag{7.21}$$

Por outro lado, nestas expressões, $\partial x/\partial x$ não significa a derivada parcial de x com relação a x que, no caso, seria igual a 1! Aqui, $\partial x/\partial x$ representa a derivada parcial da primeira função coordenada de g com relação a x, isto é, $\partial x/\partial x = \partial g_1/\partial x$. Também, o símbolo $\partial z/\partial x$ que aparece nos dois lados da equação (7.20) representa duas coisas diferentes. No lado direito, $\partial z/\partial x$ representa a derivada parcial da função f com relação a x, enquanto que, no lado esquerdo, $\partial z/\partial x$ representa a derivada parcial da função composta h com relação a x!

Moral da história: na dúvida, sempre utilize o nome das funções e nunca omita os pontos onde as funções estão calculadas.

7.6 Exercícios

[01] Encontre a equação do plano tangente ao gráfico da função definida por $f(x,y) = x^4 + 2\,x^2y^2 + xy^4 + 10\,y$ em $x = 10$ e $y = 1$.

[02] Use a aproximação linear da função $f(x,y) = x^4 + 2\,x^2y^2 + xy^4 + 10\,y$ no ponto $(10,1)$ para obter uma aproximação de $f(10.36, 1.04)$.

[03] Use a aproximação linear da função de produção $f(x,y) = 6\,x^{2/3}y^{1/2}$ no ponto $(1000,100)$ para obter uma aproximação de $f(998, 101.5)$.

[04] Use a aproximação linear da função $f(x,y,z) = \sqrt{x^{1/2} + y^{1/3} + 5\,z^2}$ no ponto $(4,8,1)$ para obter uma aproximação de $f(4.2, 7.95, 1.02)$.

[05] Use a aproximação linear da função de produção $Q = 3\,K^{2/3}L^{1/3}$ no ponto $(1000,125)$ para obter uma aproximação de $Q(998,128)$.

[06] Por que podemos dizer que os gráficos das funções $f(x,y) = x^2 + y^2$ e $g(x,y) = -x^2 - y^2 + xy^3$ são "tangentes" em $(0,0)$?

[07] Calcule a matriz jacobiana das funções dadas abaixo.

(a) f: $\mathbb{R}^2 \to \mathbb{R}^2$, f$(x,y) = (x,y)$.

(b) f: $\mathbb{R}^2 \to \mathbb{R}^3$, f$(x,y) = (xe^y + \cos(y), x, x + e^y)$.

(c) f: $\mathbb{R}^3 \to \mathbb{R}^2$, f$(x,y,z) = (x + e^z + y, yx^2)$.

(d) f: $\mathbb{R}^2 \to \mathbb{R}^3$, f$(x,y) = (xye^{xy}, x\,\mathrm{sen}(y), 5\,xy^2)$.

[08] Calcule a matriz jacobiana das funções dadas abaixo.

(a) f$(x,y) = (e^x, \mathrm{sen}(xy))$, (b) f$(x,y,z) = (x - y, y + z)$,

(c) f$(x,y) = (x + y, y - x, xy)$, (d) f$(x,y,z) = (x + z, y - 5\,z, x - y)$.

[09] Suponha que um pato esteja nadando segundo a reta definida parametricamente por $x = 3 + 8\,t$ e $y = 3 - 2\,t$. A temperatura da água é dada por $T = x^2\cos(y) - y^2\,\mathrm{sen}(x)$. Encontre a variação da temperatura em função do tempo, dT/dt, de duas maneiras: (a) pela regra da cadeia e (b) expressando T em termos de t e então derivando. Indique explicitamente as funções que fazem parte da composição, não esquecendo de

estabelecer as dimensões do domínio e do contradomínio de cada função envolvida.

[10] Verifique a regra da cadeia para as funções dadas abaixo. Indique explicitamente as funções que fazem parte da composição, não esquecendo de estabelecer as dimensões do domínio e do contradomínio de cada função envolvida.

(a) $f(x,y) = (x^2+y^2)\,\ln\sqrt{x^2+y^2}$, com $x = e^t$ e $y = e^{-t}$.

(b) $f(x,y) = x\,e^{x^2+y^2}$, com $x = t$ e $y = -t$.

(c) $f(x,y,z) = x + y^2 + z^3$, com $x = \cos(t)$, $y = \operatorname{sen}(t)$ e $z = t$.

(d) $f(x,y,z) = e^{x-z}\,(y^2 - x^2)$, com $x = t$, $y = e^t$ e $z = t^2$.

[11] Verifique a regra da cadeia para $u = x/y + y/z + z/x$, onde $x = e^t$, $y = e^{t^2}$ e $z = e^{t^3}$. Indique explicitamente as funções que fazem parte da composição, não esquecendo de estabelecer as dimensões do domínio e do contradomínio de cada função envolvida.

[12] Verifique a regra da cadeia para $u = \operatorname{sen}(xy)$, com $x = t^2 + t$ e $y = t^3$. Indique explicitamente as funções que fazem parte da composição, não esquecendo de estabelecer as dimensões do domínio e do contradomínio de cada função envolvida.

[13] Seja $z = \sqrt{x^2+y^2} + 2\,xy^2$, onde x e y são funções de u. Encontre uma expressão para dz/du usando a regra da cadeia. Indique explicitamente as funções que fazem parte da composição, não esquecendo de estabelecer as dimensões do domínio e do contradomínio de cada função envolvida.

[14] (a) Use a regra da cadeia para encontrar $(d/dx)(x^x)$ utilizando, para isto, a função $f(y,z) = y^z$.

(b) Calcule $(d/dx)(x^x)$ utilizando o cálculo em uma variável.

(c) Qual maneira você prefere?

[15] O que está errado com o seguinte argumento: suponha que $w = f(x,y)$ com $y = x^2$. Pela regra da cadeia,

$$\frac{\partial w}{\partial x} = \frac{\partial w}{\partial x}\cdot\frac{\partial x}{\partial x} + \frac{\partial w}{\partial y}\cdot\frac{\partial y}{\partial x} = \frac{\partial w}{\partial x} + 2\cdot x\cdot\frac{\partial w}{\partial y}.$$

Assim, $0 = 2\cdot x\cdot(\partial w/\partial y)$, de modo que $\partial w/\partial y = 0$.

[16] Considere a função de produção Cobb-Douglas $Q = 4\,K^{3/4}L^{1/4}$. Suponha que as variáveis K e L variem com o tempo t e com a taxa de interesse r segundo as expressões

$$K(t,r) = \frac{10\,t^2}{r} \qquad \text{e} \qquad L(t,r) = 6\,t^2 + 250\,r.$$

Use a regra da cadeia para calcular a taxa de variação de Q com relação a t quando $t = 10$ e $r = 0.1$.

[17] Considere a função de produção Cobb-Douglas $Q = 4\,K^{3/4}L^{1/4}$. Suponha que as variáveis de capital e trabalho, K e L, variem com o tempo t e com a taxa de interesse r, isto é, $K = K(t,r)$ e $L = L(t,r)$. Sabendo que $K(10, 0.1) = 10000$, $L(10, 0.1) = 625$,

$$DK(10,0.1) = \begin{bmatrix} 2000 & -100000 \end{bmatrix} \text{ e } DL(10,0.1) = \begin{bmatrix} 120 & 250 \end{bmatrix},$$

calcule a taxa de variação de Q com relação a t quando $t = 10$ e $r = 0.1$.

[18] Sejam $\mathbf{G}\colon (0,+\infty)\times(0,+\infty) \subset \mathbb{R}^2 \to (0,+\infty)\times(0,+\infty) \subset \mathbb{R}^2$ definida por

$$\mathbf{G}(x,y) = \left(\frac{10\,x^2}{y}, 6\,x^2 + 250\,y\right)$$

e $F\colon (0,+\infty)\times(0,+\infty) \subset \mathbb{R}^2 \to \mathbb{R}$ definida por $F(r,s) = 4\,r^{3/4}s^{1/4}$. Se $H = F \circ \mathbf{G}$, calcule $DH(10, 0.1)$.

[19] Sejam $\mathbf{G}(x,y) = (x^2+1, y^2)$ e $\mathbf{F}(u,v) = (u+v, v^2)$. Calcule a matriz jacobiana de $\mathbf{H}(x,y) = \mathbf{F}(\mathbf{G}(x,y))$ no ponto $(x,y) = (1,1)$.

[20] Sejam $\mathrm{f}(u,v) = (\mathrm{tg}(u-1), u^2 - v^2)$ e $\mathrm{g}(x,y) = (e^{x-y}, x-y)$. Calcule $\mathrm{f}\circ\mathrm{g}$ e $D(\mathrm{f}\circ\mathrm{g})(1,1)$.

[21] Sejam $\mathrm{f}(u,v,w) = (e^{u-w}, \cos(v+u) + \mathrm{sen}(u+v+w))$ e $\mathrm{g}(x,y) = (e^x, \cos(y-x), e^{-y})$. Calcule $\mathrm{f}\circ\mathrm{g}$ e $D(\mathrm{f}\circ\mathrm{g})(0,0)$.

[22] Sejam $\mathbf{T}\colon \mathbb{R}^n \to \mathbb{R}^m$ e $\mathbf{S}\colon \mathbb{R}^m \to \mathbb{R}^k$ transformações lineares definidas, respectivamente, por $\mathbf{T}(\mathrm{x}) = A\cdot\mathrm{x}$ e $\mathbf{S}(\mathrm{y}) = B\cdot\mathrm{y}$, com A uma matriz $m\times n$ e B uma matriz $m \times k$. Mostre que a composição

$$\mathbf{R} = \mathbf{S}\circ\mathbf{T}\colon \mathbb{R}^n \to \mathbb{R}^k$$

também é uma transformação linear e que a matriz associada a $\mathbf{R}$ é dada por $C = B\cdot A$. Moral do exercício: compor transformações lineares corresponde a multiplicar as matrizes que as definem!

[23] Considere w uma função de $\mathbb{R}^2$ para $\mathbb{R}^1$ de classe C^1, $r = y-x$, $s = y+x$ e $F(x,y) = w(r(x,y), s(x,y))$. Calcule $\partial F/\partial x$ e $\partial F/\partial y$ em termos de $\partial w/\partial r$ e $\partial w/\partial s$.

[24] Considere $u = f(x,y)$, $\mathrm{g}(s,t) = (g_1(s,t), g_2(s,t))$ e $h(s,t) = (f \circ \mathrm{g})(s,t)$. Use a regra da cadeia para obter uma expressão para $\partial h/\partial s$ e $\partial h/\partial t$ em termos das derivadas parciais de f, g_1 e g_2. Assuma que f, g_1 e g_2 são funções de classe C^1.

[25] Considere $u = f(x,y,z)$, $\mathrm{g}(r,s,t) = (g_1(r,s,t), g_2(r,s,t), g_3(r,s,t))$ e $h(r,s,t) = (f \circ \mathrm{g})(r,s,t)$. Use a regra da cadeia para obter uma expressão para $\partial h/\partial r$, $\partial h/\partial s$ e $\partial h/\partial t$ em termos das derivadas parciais de f, g_1, g_2 e g_3. Assuma que f, g_1, g_2 e g_3 são funções de classe C^1.

[26] Considere $u = f(x,y)$, $\mathrm{g}(r,s,t) = (g_1(r,s,t), g_2(r,s,t))$ e $h(r,s,t) = (f \circ \mathrm{g})(r,s,t)$. Use a regra da cadeia para obter uma expressão para $\partial h/\partial r$, $\partial h/\partial s$ e $\partial h/\partial t$ em termos das derivadas parciais de f, g_1 e g_2. Assuma que f, g_1 e g_2 são funções de classe C^1.

[27] Seja $z = f(x,y)$ uma função de classe C^1. Sejam também $x = g(u,v) = v\cos(\pi + u) + e^{uv}$ e $y = h(u,v) = u^2 - v^2$. Sabendo que

$$\frac{\partial f}{\partial x}(-1,-4) = 3 \qquad \text{e} \qquad \frac{\partial f}{\partial y}(-1,-4) = 2,$$

determine $(\partial F/\partial u)(0,2)$ e $(\partial F/\partial v)(0,2)$ da função composta $F(u,v) = f(g(u,v), h(u,v))$.

[28] Seja $f\colon \mathbb{R}^3 \to \mathbb{R}$ uma função de classe C^1 tal que $f_x(0,0,0) = 2$, $f_y(0,0,0) = f_z(0,0,0) = 0$. Se $g(u,v) = f(u-v, u^2-1, 3v-3)$, calcule $g_u(1,1)$ e $g_v(1,1)$.

[29] A resistência à retirada de um prego indica sua força de adesão à madeira. Uma fórmula empírica para pregos comuns é

$$P = 2433.5\, S^{5/2} R D,$$

em que P é a resistência à retirada máxima em kg, S é a gravidade específica da madeira a 12% de umidade, R é o raio do prego (em cm) e D é a profundidade (em cm) da penetração do prego na madeira. Um prego Cd comum de 5 cm de comprimento e 0.28 cm de diâmetro é completamente introduzido em pedaço de abeto Douglas, cuja gravidade específica é 0.54. (a) Use a equação da aproximação linear de P para aproximar a resistência máxima à retirada. Nas aplicações, apenas um sexto dessa resistência é considerado seguro, para longos períodos de tempo. (b) Quando os pregos são fabricados, R e D podem variar por $\pm 2\%$, e a gravidade específica de diferentes amostras do abeto Douglas pode variar por $\pm 3\%$. Use a equação da aproximação linear de P para obter uma aproximação do erro percentual máximo no valor calculado de P.

[30] A resistência total R de três resistores R_1, R_2 e R_3, ligados em paralelo, é dada por

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3}.$$

Se as medidas de R_1, R_2 e R_3 são 100, 200 e 400 ohms, respectivamente, com erro máximo de $\pm 1\%$ em cada medida, aproxime o erro máximo no valor calculado de R através da equação da aproximação linear.

[31] A resistência elétrica R de um fio é diretamente proporcional ao seu comprimento e inversamente proporcional ao quadrado do seu diâmetro. Sabendo que o comprimento é medido com um erro possível de $\pm 1\%$ e o diâmetro é medido com um erro possível de $\pm 3\%$, aproxime o erro máximo no valor calculado de R através da equação da aproximação linear.

[32] O fluxo sangüíneo através de uma arteríola é dado por $F = \pi P R^4/(8\,vl)$, em que l é o comprimento da arteríola, R é o raio, P é a diferença de pressão entre as duas extremidades, e v é a viscosidade do sangue. Suponha que v e l sejam constantes. Use a equação da aproximação linear para aproximar a variação percentual no fluxo sangüíneo se o raio decresce 2% e a pressão aumenta 3%.

[33] Se um remédio é ingerido por via oral, o tempo T durante o qual a maior quantidade do remédio permanece na corrente sangüínea pode ser calculado com o auxílio da meia-vida x do remédio no estômago e a meia-vida y do remédio na corrente sangüínea. Para muitos remédios comuns (como a penicilina), T é dado por

$$T = \frac{xy\,(\ln(x) - \ln(y))}{\ln(2)\,(x - y)}.$$

Para um certo remédio, $x = 30$ min e $y = 60$ min. Se o erro máximo na estimativa de cada meia-vida é $\pm 10\%$, use a equação da aproximação linear para estimar o erro máximo no valor calculado de T.

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

(a) Se $f\colon \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 tal que

$$Df(0,0,0) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

para todo $(x, y, z) \in \mathbb{R}^3$, então $f(x, y, z) = 0$ para todo $(x, y, z) \in \mathbb{R}^3$.

(b) Se $f\colon \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 tal que

$$Df(x,y,z) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

para todo $(x, y, z) \in \mathbb{R}^3$, então $f(x, y, z) = 0$ para todo $(x, y, z) \in \mathbb{R}^3$.

* (c) Se $f\colon \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 tal que

$$Df(x,y,z) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

para todo $(x, y, z) \in \mathbb{R}^3$ e $f(0,0,0) = 0$, então $f(x, y, z) = 0$ para todo $(x, y, z) \in \mathbb{R}^3$.

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

(a) Se $f\colon D \subset \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 definida no conjunto aberto

$$D = \{(x,y,z) \in \mathbb{R}^3 \mid x < 1\} \cup \{(x,y,z) \in \mathbb{R}^3 \mid x > 2\}$$

tal que

$$Df(x,y,z) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

para todo (x, y, z) em $\mathbb{R}^3$, então $f(x, y, z) = 0$ para todo $(x, y, z) \in \mathbb{R}^3$.

(b) Se $f\colon D \subset \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 definida no conjunto aberto

$$D = \{(x,y,z) \in \mathbb{R}^3 \mid x < 1\} \cup \{(x,y,z) \in \mathbb{R}^3 \mid x > 2\}$$

tal que

$$Df(x, y, z) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

para todo (x, y, z) em $\mathbb{R}^3$ e $f(0, 0, 0) = 0$, então $f(x, y, z) = 0$ para todo $(x, y, z) \in \mathbb{R}^3$.

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

(a) Se $f\colon \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 satisfazendo

$$Df(0, 0, 0) = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$$

então $f(x, y, z) = 2\,x$ para todo $(x, y, z) \in \mathbb{R}^3$.

(b) Se $f\colon \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 satisfazendo

$$Df(x, y, z) = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$$

para todo $(x, y, z) \in \mathbb{R}^3$, então $f(x, y, z) = 2\,x$ para todo $(x, y, z) \in \mathbb{R}^3$.

* (c) Se $f\colon \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 satisfazendo

$$Df(x, y, z) = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$$

para todo $(x, y, z) \in \mathbb{R}^3$ e $f(0, 0, 0) = 0$, então $f(x, y, z) = 2\,x$ para todo $(x, y, z) \in \mathbb{R}^3$.

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

(a) Se $f\colon D \subset \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 definida no conjunto aberto

$$D = \{(x, y, z) \in \mathbb{R}^3 \mid x < 1\} \cup \{(x, y, z) \in \mathbb{R}^3 \mid x > 2\}$$

satisfazendo

$$Df(x, y, z) = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$$

para todo (x, y, z) no domínio D, então $f(x, y, z) = 2\,x$ para todo $(x, y, z) \in D$.

(b) Se $f\colon D \subset \mathbb{R}^3 \to \mathbb{R}$ é uma função de classe C^1 definida no conjunto aberto

$$D = \{(x, y, z) \in \mathbb{R}^3 \mid x < 1\} \cup \{(x, y, z) \in \mathbb{R}^3 \mid x > 2\}$$

satisfazendo

$$Df(x, y, z) = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$$

para todo (x, y, z) no domínio D e $f(0, 0, 0) = 0$, então $f(x, y, z) = 2x$ para todo $(x, y, z) \in D$.

[38] **(Funções homogêneas)** Dizemos que $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma *função homogênea* de grau k se

$$f(t \cdot \mathrm{x}) = t^k \cdot f(\mathrm{x}), \tag{7.22}$$

para todo $\mathrm{x} \in \mathbb{R}^n$ e pra todo $t > 0$.

(a) Dê exemplos de funções homogêneas de grau 1 e 2 definidas em $\mathbb{R}^2$.

(b) Dê exemplos de funções homogêneas de grau 3 e 4 definidas em $\mathbb{R}^3$.

(c) $f(x_1, x_2) = x_1^2 x_2 + 3\, x_1 x_2^2 + x_2^3$ é uma função homogênea?

(d) Para a função f do item anterior, mostre que

$$x_1 \frac{\partial f}{\partial x_1}(x_1, x_2) + x_2 \frac{\partial f}{\partial x_2}(x_1, x_2) = 3\, f(x_1, x_2),$$

para todo $(x_1, x_2) \in \mathbb{R}^2$.

(e) $f(x_1, x_2) = 4\, x_1^2 x_2^3 - 5\, x_1 x_2^3$ é uma função homogênea?

(f) Para a função f do item anterior, mostre que existe $(x_1, x_2) \in \mathbb{R}^2$ tal que

$$x_1 \frac{\partial f}{\partial x_1}(x_1, x_2) + x_2 \frac{\partial f}{\partial x_2}(x_1, x_2) \neq k\, f(x_1, x_2),$$

para todo $k \in \mathbb{R}$.

(g) **(Teorema de Euler)** Mostre que se $f\colon \mathbb{R}^n \to \mathbb{R}$ é uma função homogênea de grau k, então

$$x_1 \frac{\partial f}{\partial x_1}(\mathrm{x}) + x_2 \frac{\partial f}{\partial x_2}(\mathrm{x}) + \cdots + x_n \frac{\partial f}{\partial x_n}(\mathrm{x}) = k\, f(\mathrm{x}),$$

para todo $\mathrm{x} \in \mathbb{R}^n$.

Funções homogêneas aparecem naturalmente em Economia. Por exemplo, funções de lucro e de custo (que surgem a partir de funções de produção), e funções de demanda (que surgem a partir de funções de utilidade), são automaticamente funções homogêneas nos modelos econômicos tradicionais. O capítulo 20 da referência [71] oferece um excelente estudo de funções homogêneas no contexto econômico.

*[39] Sejam $z = f(u,v)$, $\mathrm{g}(x,y) = (g_1(x,y), g_2(x,y))$ e $h(x,y) = (f \circ \mathrm{g})(x,y)$. Mostre que

$$\begin{aligned}\frac{\partial^2 h}{\partial x^2}(\mathrm{p}) &= \frac{\partial^2 f}{\partial u^2}(\mathrm{g}(\mathrm{p})) \cdot \left(\frac{\partial g_1}{\partial x}(\mathrm{p})\right)^2 + \\ &\quad 2 \cdot \frac{\partial^2 f}{\partial u \partial v}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial g_1}{\partial x}(\mathrm{p}) \cdot \frac{\partial g_2}{\partial x}(\mathrm{p}) + \\ &\quad \frac{\partial^2 f}{\partial v^2}(\mathrm{g}(\mathrm{p})) \cdot \left(\frac{\partial g_2}{\partial x}(\mathrm{p})\right)^2 + \\ &\quad \frac{\partial f}{\partial u}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial^2 g_1}{\partial x^2}(\mathrm{p}) + \frac{\partial f}{\partial v}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial^2 g_2}{\partial x^2}(\mathrm{p}),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 f}{\partial x \partial y}(\mathrm{p}) &= \frac{\partial^2 f}{\partial u^2}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial g_1}{\partial x}(\mathrm{p}) \cdot \frac{\partial g_1}{\partial y}(\mathrm{p}) + \\ &\quad \frac{\partial^2 f}{\partial u \partial v}(\mathrm{g}(\mathrm{p})) \cdot \left(\frac{\partial g_1}{\partial x}(\mathrm{p}) \cdot \frac{\partial g_2}{\partial y}(\mathrm{p}) + \frac{\partial g_1}{\partial y}(\mathrm{p}) \cdot \frac{\partial g_2}{\partial x}(\mathrm{p})\right) + \\ &\quad \frac{\partial^2 f}{\partial v^2}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial g_2}{\partial x}(\mathrm{p}) \cdot \frac{\partial g_2}{\partial y}(\mathrm{p}) + \\ &\quad \frac{\partial f}{\partial u}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial g_1}{\partial x \partial y}(\mathrm{p}) + \frac{\partial f}{\partial v}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial^2 g_2}{\partial x \partial y}(\mathrm{p}),\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 h}{\partial y^2}(\mathrm{p}) &= \frac{\partial^2 f}{\partial u^2}(\mathrm{g}(\mathrm{p})) \cdot \left(\frac{\partial g_1}{\partial y}(\mathrm{p})\right)^2 + \\ &\quad 2 \cdot \frac{\partial^2 f}{\partial u \partial v}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial g_1}{\partial y}(\mathrm{p}) \cdot \frac{\partial g_2}{\partial y}(\mathrm{p}) + \\ &\quad \frac{\partial^2 f}{\partial v^2}(\mathrm{g}(\mathrm{p})) \cdot \left(\frac{\partial g_2}{\partial y}(\mathrm{p})\right)^2 + \\ &\quad \frac{\partial f}{\partial u}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial^2 g_1}{\partial y^2}(\mathrm{p}) + \frac{\partial f}{\partial v}(\mathrm{g}(\mathrm{p})) \cdot \frac{\partial^2 g_2}{\partial y^2}(\mathrm{p}),\end{aligned}$$

onde f, g_1 e g_2 são funções de classe C^2.

WWW

[40] **(O método de Newton)** Em Cálculo I você estudou o método de Newton como um algoritmo com a finalidade de aproximar uma raiz de função real de uma variável, isto é, aproximar um número real p tal que $f(p) = 0$, com $f\colon \mathbb{R} \to \mathbb{R}$. A idéia do método é muito simples. A partir de uma aproximação inicial x_0, calculamos a equação da reta tangente ao gráfico de f no ponto x_0,

$$y = l(x) = f(x_0) + f'(x_0)(x - x_0).$$

Observe que encontrar uma raiz da aproximação l é mais fácil do que encontrar uma raiz da função original f:

$$y = l(x) = f(x_0) + f'(x_0)(x - x_0) \Rightarrow x = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

O número $x_1 = x_0 - f(x_0)/f'(x_0)$ é uma outra aproximação da raiz p de f (estamos supondo, evidentemente, que $f'(x_0) \neq 0$). Basta agora repetir o processo, trocando x_0 por x_1 para calcular x_2, e assim por diante. Espera-se que a seqüência gerada por esta construção convirja para a raiz p de f.

O método de Newton pode ser generalizado para funções vetoriais do tipo

$$\mathrm{F}\colon \mathbb{R}^n \to \mathbb{R}^n,$$

seguindo a mesma construção do caso escalar. A partir de uma aproximação inicial x_0, calculamos a equação da aproximação (afim) de F no ponto x_0,

$$\mathrm{y} = \mathrm{l}(\mathrm{x}) = \mathrm{F}(\mathrm{x}_0) + D\mathrm{F}(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0),$$

onde $D\mathrm{F}(\mathrm{x}_0)$, de tamanho $n \times n$, é a matriz jacobiana de F no ponto x_0. Observe que encontrar uma raiz da aproximação l é mais fácil do que encontrar uma raiz da função original F:

$$\begin{aligned}\mathrm{F}(\mathrm{x}_0) + D\mathrm{F}(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0) = 0 &\Rightarrow D\mathrm{F}(\mathrm{x}_0)(\mathrm{x} - \mathrm{x}_0) = -\mathrm{F}(\mathrm{x}_0) \Rightarrow \\ \mathrm{x} - \mathrm{x}_0 = -[D\mathrm{F}(\mathrm{x}_0)]^{-1}\mathrm{F}(\mathrm{x}_0) &\Rightarrow \mathrm{x} = \mathrm{x}_0 - [D\mathrm{F}(\mathrm{x}_0)]^{-1}\mathrm{F}(\mathrm{x}_0).\end{aligned}$$

O ponto $\mathrm{x}_1 = \mathrm{x}_0 - [D\mathrm{F}(\mathrm{x}_0)]^{-1}\mathrm{F}(\mathrm{x}_0)$ é uma outra aproximação da raiz p de F (estamos supondo, evidentemente, que a matriz $D\mathrm{F}(\mathrm{x}_0)$ seja inversível). Basta agora repetir o processo, trocando x_0 por x_1 para calcular x_2, e assim por diante. Duas observações:

(1) No caso escalar, a derivada $f'(p')$ é um número e, portanto, faz sentido dividir por $f'(p)$. No caso vetorial, a derivada é uma matriz e, portanto, a divisão é substituída pelo uso da inversa. Se usarmos a notação

$$\frac{1}{f(x_0)} = [f(x_0)]^{-1},$$

a fórmula de iteração para o caso vetorial fica muito parecida com o caso escalar.

(2) Numericamente, dado o ponto x_0, não se calcula o ponto x_1 pela fórmula

$$\mathrm{x}_1 = \mathrm{x}_0 - [D\mathrm{F}(\mathrm{x}_0)]^{-1}\mathrm{F}(\mathrm{x}_0),$$

pois calcular a inversa de uma matriz é computacionalmente caro. O que se faz é resolver o sistema linear (algo mais barato):

$$D\mathrm{F}(\mathrm{x}_0)\,\mathrm{z} = -\mathrm{F}(\mathrm{x}_0)$$

e, então, tomar $\mathrm{x}_1 = \mathrm{x}_0 + \mathrm{z}$.

O método de Newton é uma manifestação de uma idéia central na teoria de Cálculo II: para se resolver um problema não-linear (em geral muito difícil), o que se faz é tentar simplificar o problema com uma aproximação linear (o papel da derivada é fundamental nesta parte), resolver o problema no caso simplificado (algo mais fácil de se fazer) e, então, relacionar a solução do problema simplificado com a solução do problema original.

Naturalmente, a exemplo do caso escalar, a seqüência gerada pelo método de Newton vetorial pode não convergir. Não estudaremos, aqui, questões de convergência, testes de parada e detalhes de implementação. Indicamos, ao leitor interessado, as referências [30, 34, 61].

(a) Seja $\mathrm{F}(x_1, x_2) = (x_1^2 + x_2^2 - 1, x_2 - \mathrm{sen}(x_1))$. Calcule os três primeiros elementos da seqüência gerada pelo método de Newton para a função F, usando $\mathrm{x}_0 = (1, 1)$ como tentativa inicial.

(b) Mostre que se $\mathrm{F}\colon \mathbb{R}^n \to \mathbb{R}$ é uma função de classe C^∞ e a seqüência x_k gerada pelo método de Newton para F é convergente, então x_k converge para uma raiz p de F. Dica: mostre que x_k satisfaz

$$\mathrm{x}_{k+1} = \mathrm{x}_k - [D\mathrm{F}(\mathrm{x}_k)]^{-1}\mathrm{F}(\mathrm{x}_k),$$

para todo $k \geq 0$ e, então, tome o limite para $k \to \infty$ dos dois lados desta equação.