UNIVERSIDADE FEDERAL DO AMAZONAS

INSTITUTO DE CIÊNCIAS EXATAS

DEPARTAMENTO DE MATEMÁTICA

Gleison Stanlley da Silva Marques

Um compêndio do cálculo diferencial em espaços normados com aplicações em otimização contínua

Manaus

16 de novembro de 2023

Gleison Stanlley da Silva Marques

Um compêndio do cálculo diferencial em espaços normados com aplicações em otimização contínua

Trabalho de Conclusão de Curso submetido à Universidade Federal do Amazonas como requisito necessário para obtenção do grau de Licenciado em Matemática.

Universidade Federal do Amazonas – UFAM

Licenciatura em Matemática

Programa de Graduação

Orientadora: Prof.ª Dr.ª Flávia Morgana de Oliveira Jacinto

Manaus

16 de novembro de 2023

Ficha Catalográfica

Elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

M357c Marques, Gleison Stanlley da Silva
Um compêndio do cálculo diferencial em espaços normados com aplicações em otimização contínua / Gleison Stanlley da Silva Marques. - 2023.
195 f. ; 31 cm.

Orientador(a): Flávia Morgana de Oliveira Jacinto.
Trabalho de Conclusão de Curso (graduação) - Universidade Federal do Amazonas, Instituto de Ciências Exatas, Curso de Matemática, Manaus, 2023.

1. Cálculo diferencial. 2. Espaços normados. 3. Otimização contínua. 4. Dimensão infinita. I. Jacinto, Flávia Morgana de Oliveira. II. Universidade Federal do Amazonas. Instituto de Ciências Exatas. Curso de Matemática. III. Título

Agradecimentos

Agradeço à minha avó-mãe, Altamira, por confiar e investir em mim; por me amar e zelar por mim, mesmo à distância.

Agradeço à minha orientadora, Flávia, por todo o carinho, suporte, solicitude, flexibilidade e amizade que me acompanharam todos os dias durante todo o processo de escrita desse trabalho e fora dele.

Agradeço aos membros da banca, professor Thiago e professor Marrocos, por aceitarem o convite de julgar esse trabalho. Ao professor Thiago, meu primeiro orientador pelo Programa de Educação Tutorial (PET) de Matemática, com quem aprendi muito sobre esse mundo que é a Matemática e sobre a carreira acadêmica; ao professor Marrocos, por sempre me ajudar com minhas dúvidas e me incentivar a dar o melhor de mim.

Agradeço ao PET-Matemática pelas bolsas que nunca faltaram durante meus dois anos e meio de permanência no programa e ao professor Roberto Antônio Cordeiro Prata, tutor do programa, por ter sempre me incentivado e confiado no meu trabalho.

Agradeço, em caráter especial, aos meus amigos Diogo Sampaio da Silva, Gustavo Costa de Souza e Vinícius Rosário das Chagas por todos os momentos que passamos juntos estudando e trocando ideias; pelo companheirismo e pela cumplicidade que me permitiram chegar mentalmente são (ou quase) até aqui.

"Sem dúvida já lhes perguntaram muitas vezes para que serve a matemática, e se essas delicadas construções que tiramos inteiras de nosso espírito não são artificiais, concebidas por nosso capricho. Entre os que fazem essa pergunta, devo fazer uma distinção; os práticos reclamam de nós apenas um meio de ganhar dinheiro. Estes não merecem resposta; é a eles, antes, que conviria perguntar para que serve acumular tantas riquezas e se, para ter tempo de adquiri-las, é preciso negligenciar a arte e a ciência, as únicas que podem nos proporcionar espíritos capazes de usufruí-las,

'E por causa da vida perder as razões de viver.'

– Henri Poincaré, O Valor da Ciência

Resumo

Neste trabalho apresentamos um compêndio sobre o cálculo diferencial em espaços normados e abordamos teoremas clássicos e extensões para espaços de dimensão infinita. Além disso, também exploramos aplicações em otimização contínua, incluindo condições de otimalidade de primeira e segunda ordem e generalizações delas. Nosso objetivo é fornecer um material acessível e rigoroso que destaque também a importância do cálculo diferencial na formulação e resolução de problemas de otimização.

Palavras-chave: Cálculo diferencial. Espaços normados. Otimização contínua. Dimensão infinita.

Abstract

In this work, we present a compendium on differential calculus in normed spaces and address classic theorems and their extensions to infinite-dimensional spaces. Furthermore, we also explore applications in continuous optimization, including first and second-order optimality conditions and their generalizations. Our objective is to provide accessible and rigorous material that also highlights the importance of differential calculus in the formulation and resolution of optimization problems.

Keywords: Differential calculus. Normed spaces. Continuous optimization. Infinite-dimensional.

Sumário

Introdução

No capítulo 1, introduzimos o conceito que generaliza a noção de derivada de funções de variável real, a derivada de Fréchet; apresentamos as regras básicas de diferenciação, como a regra da cadeia (ver Teorema 1.3.1), a regra do produto interno (ver Exemplo 1.4.4) e a regra do produto vetorial (ver Exemplo 1.4.5); além de comentarmos também um pouco sobre o vetor e a aplicação gradiente, a matriz jacobiana e a matriz hessiana, tópicos padrões dos cursos de Cálculo Diferencial de várias variáveis.

No capítulo 2, introduzimos algumas das várias versões do Teorema do Valor Médio, como o Teorema do Valor Médio de Lagrange (ver Teorema 2.1.4) e a Desigualdade do Valor Médio (ver Teorema 2.3.3).

No capítulo 3 introduzimos o cálculo de derivadas de ordem superior, construindo uma maneira elegante e mais simples de trabalhar com tais derivadas utilizando a ajuda da Álgebra Multilinear. Mais detalhadamente, introduzimos uma coleção de isometrias lineares entre *espaços ruins* e *espaços bons* e fazemos todas as definições importantes sobre os *espaços bons*. As fórmulas de Taylor encerram esse capítulo.

Em seguida, no capítulo 4, abordamos os teoremas da aplicação inversa e implícita, que constituem ferramentas fundamentais no estudo das variedades diferenciáveis.

Nos capítulos 5, 6, 7 e 8 abordamos um pouco da teoria de otimização com ferramentas de convexidade.

Por fim, os apêndices estão aqui para tornar o texto o mais autocontido possível.

Parte I

Cálculo Diferencial em espaços normados

1 O que é uma aplicação diferenciável?

Reservaremos o termo "função" para aplicações com imagem em $\mathbb{R}$ e "aplicações" para o conceito mais geral. Além disso, lidaremos somente com espaços vetoriais reais.

Sejam E e F espaços normados, $X \subseteq E$, $a \in E$ um ponto de acumulação de X, $f : X \to F$ uma aplicação e $g : X \to \mathbb{R}$ uma função.

Dizemos que $f(x)$ é "o pequeno" de $g(x)$ quando $x \to a$ e escrevemos

$$f(x) = o(g(x)) \quad \text{quando} \quad x \to a$$

se existem uma vizinhança U de a e uma aplicação $\epsilon : X \cap U \to F$ tal que

$$f(x) = \epsilon(x)g(x), \quad \text{para todo} \quad x \in (X \cap U) \setminus \{a\},$$

e

$$\lim_{x\to a} \epsilon(x) = 0_F.$$

Em particular, se $g(x) \neq 0$ numa vizinhança de a, isso é equivalente a

$$\lim_{x\to a} \frac{\|f(x)\|_F}{|g(x)|} = 0.$$

Dizemos que $f(x)$ é "o grande" de $g(x)$ quando $x \to a$ e escrevemos

$$f(x) = O(g(x)) \quad \text{quando} \quad x \to a$$

se existem $C > 0$ e uma vizinhança U de a tal que

$$\|f(x)\|_F \leq C|g(x)|, \quad \text{para todo} \quad x \in (X \cap U) \setminus \{a\}.$$

Observe que

$$f(x) = o(g(x)) \implies f(x) = O(g(x)).$$

O argumento é simples: como $f(x) = \epsilon(x)g(x)$, com $\lim\limits_{x\to a} \epsilon(x) = 0$, então, tomando $\varepsilon := 1$ na definição de limite, existe $\delta > 0$ tal que $\|\epsilon(x)\| < 1$, para todo $x \in B(a,\delta) \setminus \{a\}$. Isso nos dá

$$\|f(x)\| = |g(x)|\|\epsilon(x)\| < |g(x)|, \quad \text{para todo} \quad x \in B(a,\delta) \setminus \{a\}.$$

Logo, $f(x) = O(g(x))$, como queríamos demonstrar.

1.1 Gâteaux vs. Fréchet

Sejam E um espaço vetorial, F um espaço normado e $f : U \subseteq E \to F$ uma aplicação.

A definição abaixo é uma extensão da mesma definição que pode ser encontrada em (COLEMAN, 2012, p. 59).

Definição 1.1.1 (Derivadas direcionais)**.** A **derivada direcional** de f no ponto $a \in U$ na direção do vetor $v \in E$ é, se existe, o limite

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

Se $\dfrac{\partial f}{\partial h}(a)$ existe, para todo $h \in E$, definimos a aplicação

$$\begin{aligned} \frac{\partial f}{\partial (\cdot)}(a) : E &\to F \\ h &\mapsto \frac{\partial f}{\partial h}(a). \end{aligned}$$

Observação 1.1.2. Nos livros de Cálculo, como (CIPOLATTI, 2020, p. 65), é comum denotar a derivada direcional de uma aplicação $f : U \subseteq \mathbb{R}^n \to \mathbb{R}^m$ na direção dos vetores da base canônica, $\dfrac{\partial f}{\partial e_i}(a)$, por $\dfrac{\partial f}{\partial x_i}(a)$, e chamá-los de **derivadas parciais**.

As definições abaixo são versões modificadas, porém equivalentes, de definições que se encontram em (COLEMAN, 2012, pp. 38, 59).

Suponhamos agora que E é um espaço normado e $U \subseteq E$ é um subconjunto aberto.

Definição 1.1.3 (Gâteaux e Fréchet)**.**

- Se $\dfrac{\partial f}{\partial (\cdot)}(a)$ é linear e contínua, dizemos que f é **Gâteaux-diferenciável** em a.
- Dizemos que f é **diferenciável** (ou **Fréchet-diferenciável**) em $a \in U$ se existe uma aplicação $T : E \to F$ linear e contínua tal que

$$f(a+h) - f(a) - Th = o(\|h\|_E) \quad \text{quando} \quad h \to 0_E. \tag{1.1}$$

Proposição 1.1.4 (Definição alternativa de diferenciabilidade)**.** A aplicação f é diferenciável em $a \in U$ se, e somente se, existem aplicações $T, r : E \to F$ tais que T é linear e contínua; e, para todo $h \in E$ tal que $a + h \in U$, temos

$$f(a+h) = f(a) + Th + r(h), \quad \text{com} \quad \lim_{h \to 0_E} \frac{r(h)}{\|h\|_E} = 0_F.$$

Demonstração. Suponhamos f é diferenciável em $a \in U$ e seja $H := \{h \in E : a + h \in U\}$. Observe que H é aberto, já que $H = g^{-1}(U)$, onde $g(h) := a + h$ é uma função contínua. Por definição, existe $T : E \to F$ linear e contínua tal que

$$f(a+h) - f(a) - Th = o(\|h\|_E) \quad \text{quando} \quad h \to 0_E.$$

Por definição, existem uma vizinhança V de 0_E e uma aplicação $\epsilon : V \cap H \to F$ tal que

$$f(a+h) - f(a) - Th = \epsilon(h)\|h\|_E, \quad \text{para todo} \quad h \in (V \cap H) \setminus \{0_E\},$$

e $\lim_{h \to 0_E} \epsilon(h) = 0_F$. Definamos então

$$r(h) := \begin{cases} f(a+h) - f(a) - Th, & \text{se} \quad h \in H, \\ 0_F, & \text{se} \quad h \in E \setminus H. \end{cases}$$

Com essa definição temos

$$r(h) = \epsilon(h)\|h\|_E, \quad \text{para todo} \quad h \in V \cap H.$$

Veja que, como H e V são vizinhanças de 0_E, então $V \cap H$ também é uma vizinhança de 0_E. Logo, existe $\delta > 0$ tal que $B(0_E, \delta) \subseteq V \cap H$, o que nos dá

$$r(h)/\|h\|_E = \epsilon(h), \quad \text{para todo} \quad h \in B(0_E, \delta) \setminus \{0_E\}.$$

Em particular, isso implica que

$$\lim_{h \to 0_E} \frac{r(h)}{\|h\|_E} = \lim_{h \to 0_E} \epsilon(h) = 0_F.$$

Para a recíproca, sejam T e r como no enunciado. Queremos mostrar que

$$r(h) = o(\|h\|_E) \quad \text{quando} \quad h \to 0_E.$$

Defina

$$\epsilon(h) := \begin{cases} r(h)/\|h\|_E, & \text{se} \quad h \in H \setminus \{0_E\}, \\ 0_F, & \text{se} \quad h \in E \setminus (H \setminus \{0_E\}). \end{cases}$$

Então $\lim_{h \to 0_E} \epsilon(h) = 0_F$ e

$$r(h) = \epsilon(h)\|h\|_E,$$

para todo $h \in H \setminus \{0_E\}$, o que termina a prova. □

A demonstração do Teorema abaixo é uma adaptação da ideia exposta em (LIMA, 2002, p. 2).

Teorema 1.1.5 (Unicidade do diferencial)**.** Se f é diferenciável em $a \in U$, então $\dfrac{\partial f}{\partial(\cdot)}(a)$ é a única aplicação linear e contínua satisfazendo (1.1).

Demonstração. Sejam $T, r : E \to F$ como no enunciado da Proposição 1.1.4. Como U é aberto e $a \in U$, então existe $\delta > 0$ tal que $B(a, \delta) \subseteq U$. Daí, fixado $h \in E \setminus \{0\}$, temos

$$Th = \frac{tTh}{t} = \frac{T(th)}{t} = \frac{f(a+th) - f(a)}{t} - \frac{r(th)}{t},$$

para todo $t \in \mathbb{R}$ tal que $0 < |t| < \dfrac{\delta}{\|h\|_E}$. Como

$$\frac{r(th)}{t} = \|h\|_E \frac{r(th)}{t\|h\|_E} = \pm\|h\|_E \frac{r(th)}{\|th\|_E},$$

para todo $t \in \mathbb{R}$ tal que $0 < |t| < \dfrac{\delta}{\|h\|_E}$, e

$$\pm\|h\|_E \frac{r(th)}{\|th\|_E} \xrightarrow{t \to 0} 0_F,$$

então

$$Th = \lim_{t \to 0} \frac{f(a+th) - f(a)}{t} - \lim_{t \to 0} \frac{r(th)}{t} = \frac{\partial f}{\partial h}(a).$$

Isto prova que, para todo $h \in E \setminus \{0\}$, temos

$$Th = \frac{\partial f}{\partial h}(a).$$

O fato de que

$$T0_E = 0_F = \lim_{t \to 0} \frac{f(a + t0_E) - f(a)}{t} = \frac{\partial f}{\partial 0_E}(a)$$

encerra a demonstração. □

Quando f for diferenciável em a, a aplicação linear e contínua $\dfrac{\partial f}{\partial(\cdot)}(a)$ será chamada de **diferencial** (ou **derivada de Fréchet**) de f em a e será denotada por $f'(a)$. Assim, a fim de verificar que f é diferenciável em um ponto $a \in U$ é necessário e suficiente verificar que $\dfrac{\partial f}{\partial(\cdot)}(a)$ é linear e contínua e que

$$\frac{f(a+h) - f(a) - \dfrac{\partial f}{\partial h}(a)}{\|h\|_E} \xrightarrow{\|h\|_E \to 0} 0_F.$$

Proposição 1.1.6. Se f é diferenciável em $a \in U$, então f é contínua em a.

Demonstração. Como

$$\lim_{h \to 0} f(a+h) = \lim_{h \to 0} [f(a) + f'(a)h + r(h)] = \lim_{h \to 0} \left[f(a) + f'(a)h + \|h\| \frac{r(h)}{\|h\|} \right] = f(a),$$

então f é contínua em a. □

O exemplo abaixo foi retirado de um dos exercícios de (LIMA, 2002, p. 4).

Exemplo 1.1.7. A aplicação

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

é diferenciável. Neste exemplo, estamos considerando $\mathbb{R}^2$ munido da norma euclidiana.

Demonstração. Note que

$$\begin{aligned}
\frac{\partial f}{\partial (c,d)}(a,b) &= \lim_{t\to 0} \frac{f((a,b)+t(c,d)) - f(a,b)}{t} \\
&= \lim_{t\to 0} \frac{(2act + c^2t^2 + dt, ct + 2bdt + d^2t^2)}{t} \\
&= \lim_{t\to 0} (2ac + c^2 t + d, c + 2bd + d^2 t) \\
&= (2ac + d, c + 2bd) \\
&= \begin{bmatrix} 2a & 1 \\ 1 & 2b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix}.
\end{aligned}$$

Como $\dfrac{\partial f}{\partial (\cdot)}(a,b)$ é linear e contínua (ver Corolário C.4.12) e

$$(c^2, d^2) = f((a,b) + (c,d)) - f(a,b) - \frac{\partial f}{\partial (c,d)}(a,b),$$

basta verificar que

$$\lim_{(c,d)\to(0,0)} \frac{(c^2,d^2)}{\sqrt{c^2+d^2}} = (0,0).$$

Observe que

$$\frac{(c^2,d^2)}{\sqrt{c^2+d^2}} = \left(\frac{c^2}{\sqrt{c^2+d^2}}, \frac{d^2}{\sqrt{c^2+d^2}} \right).$$

Como

$$0 \leq \frac{c^2}{c^2+d^2} \leq 1$$

e

$$\pm c \xrightarrow{(c,d)\to(0,0)} 0,$$

então

$$\frac{c^2}{\sqrt{c^2+d^2}} = \pm c \cdot \sqrt{\frac{c^2}{c^2+d^2}} \xrightarrow{(c,d)\to(0,0)} 0.$$

Um argumento análogo prova que

$$\frac{d^2}{\sqrt{c^2+d^2}} \xrightarrow{(c,d)\to(0,0)} 0.$$

Por conseguinte, f é diferenciável em (a,b). $\square$

Rigorosamente falando, se E é um espaço vetorial e $\|\cdot\|$ e $\|\cdot\|^{\times}$ são normas distintas em E, então $(E, \|\cdot\|)$ e $(E, \|\cdot\|^{\times})$ são espaços normados distintos, mesmo que $\|\cdot\|$ e $\|\cdot\|^{\times}$ sejam equivalentes. Podemos nos perguntar, então, se a diferenciabilidade de uma aplicação é afetada quando se mudam as normas dos espaços envolvidos por normas equivalentes. O teorema abaixo, que pode ser encontrado em (COLEMAN, 2012, p. 39), garante que não.

Teorema 1.1.8 (Substituição de normas)**.** Se $(E, \|\cdot\|_E)$ e $(F, \|\cdot\|_F)$ são espaços normados, $U \subseteq E$ é um subconjunto aberto e $f : U \to F$ é uma aplicação diferenciável em $a \in U$, então a diferenciabilidade de f em a não é afetada quando substituímos as normas de E e de F por normas que lhes são equivalentes. Mais precisamente, ainda vale

$$\lim_{h \to 0_E} \frac{\|r(h)\|_F^{\times}}{\|h\|_E^{\times}} = 0,$$

onde

$$r(h) := f(a+h) - f(a) - f'(a)h$$

e $\|\cdot\|_E^{\times}$ e $\|\cdot\|_F^{\times}$ são, respectivamente, normas equivalentes às normas $\|\cdot\|_E$ e $\|\cdot\|_F$.

Demonstração. Se $\|\cdot\|_E$ e $\|\cdot\|_F$ são, respectivamente, equivalentes às normas $\|\cdot\|_E^{\times}$ e $\|\cdot\|_F^{\times}$, então existem $C_E, C_F > 0$ tais que

$$\begin{cases} C_E\|x\|_E \leq \|x\|_E^{\times}, & \text{para todo} \quad x \in E, \\ \|y\|_F^{\times} \leq C_F\|y\|_F, & \text{para todo} \quad y \in F. \end{cases}$$

Como

$$0 \leq \frac{\|r(h)\|_F^{\times}}{\|h\|_E^{\times}} \leq \frac{C_F\|r(h)\|_F}{C_E\|h\|_E} = \frac{C_F}{C_E} \cdot \frac{\|r(h)\|_F}{\|h\|_E} \xrightarrow{h \to 0} 0,$$

então, pelo Teorema do Confronto (ver Teorema B.5.10), temos

$$\lim_{h \to 0} \frac{\|r(h)\|_F^{\times}}{\|h\|_E^{\times}} = 0,$$

como queríamos demonstrar. □

Em particular, quando E e F têm dimensão finita, nenhuma norma que escolhamos para eles afeta a diferenciabilidade de uma aplicação em um dado ponto, como consequência do Teorema C.4.9.

Definição 1.1.9 (Derivada)**.** Se E é um espaço normado, $X \subseteq \mathbb{R}$ e $a \in X \cap X'$, dizemos que uma aplicação $f : X \to E$ é **derivável** em a se existe o limite

$$\dot{f}(a) := \frac{d}{dt} f(t)\Big|_{t=a} := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Quando existe, o vetor $\dot{f}(a)$ é chamado a **derivada** de f em a. Dizemos que f é **derivável** se $\dot{f}(x)$ existe, para todo $x \in X$. Nesse caso, a **derivada** de f é a função

$$\frac{d}{dt}f(t) : a \in X \mapsto \frac{d}{dt}f(t)\Big|_{t=a}.$$

O conceito de diferenciabilidade generaliza o conceito de derivada.

Teorema 1.1.10. Se $I \subseteq \mathbb{R}$ é um intervalo aberto, então uma função $f : I \to \mathbb{R}$ é derivável em $a \in I$ se, e somente se, f é diferenciável em a. Nesse caso,

$$\dot{f}(a)h = f'(a)h, \quad \text{para todo} \quad h \in \mathbb{R}\,.$$

Demonstração. Suponhamos que f é derivável em a. Como, para todo $h \in \mathbb{R}$,

$$\frac{\partial f}{\partial h}(a) = \lim_{t\to 0}\frac{f(a+th)-f(a)}{t} = \lim_{t\to 0}\frac{f(a+th)-f(a)}{th}h = \dot{f}(a)h,$$

então $\dfrac{\partial f}{\partial(\cdot)}(a)$ é uma função obviamente linear e contínua. Além disso,

$$\lim_{h\to 0}\frac{f(a+h)-f(a)-\dot{f}(a)h}{h} = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h} - \dot{f}(a) = 0,$$

então f é diferenciável em a e vale

$$f'(a)h = \dot{f}(a)h, \quad \text{para todo} \quad h \in \mathbb{R}\,.$$

Reciprocamente, suponhamos que f é diferenciável em a. Assim,

$$f'(a)1 = \lim_{t\to 0}\frac{f(a+t)-f(a)}{t} = \dot{f}(a),$$

o que termina a prova. □

Observação 1.1.11. Note que se E é um espaço vetorial, F é um espaço normado, $a \in X \subseteq E$ e $f : X \to F$ é uma aplicação, então, dado $h \in E$, a derivada direcional $\dfrac{\partial f}{\partial h}(a)$ existe se, e somente se, $\dot{f}_h(0)$ existe, onde

$$f_h(t) := f(a+th), \text{ para todo } t \in \mathbb{R} \text{ tal que } a+th \in X.$$

Nesse caso, temos

$$\frac{\partial f}{\partial h}(a) = \dot{f}_h(0).$$

Veja que a diferenciabilidade é um conceito pontual: uma aplicação diferenciável em um ponto não é necessariamente diferenciável numa vizinhança desse ponto. Vejamos alguns exemplos disso.

Exemplo 1.1.12. A função

$$f(x) := \begin{cases} x^2, & \text{se} \quad x \in \mathbb{Q}, \\ 0, & \text{se} \quad x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

é derivável apenas em 0.

Demonstração. Por um cálculo direto, $\dot{f}(0) = 0$. Argumentaremos agora que 0 é o único número racional onde f é derivável.

Seja $a \in \mathbb{Q}$ tal que f é derivável em a. Por definição, existe o limite

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

Como

$$\dot{f}(a) = \lim_{h \to 0, h \in \mathbb{Q}} \frac{(a+h)^2 - a^2}{h} = \lim_{h \to 0, h \in \mathbb{Q}} \frac{a^2 + 2ah + h^2 - a^2}{h} = 2a$$

e, por outro lado,

$$\dot{f}(a) = \lim_{h \to 0, h \in \mathbb{R} \setminus \mathbb{Q}} \frac{-a^2}{h} = \lim_{h \to 0, h \in \mathbb{R} \setminus \mathbb{Q}} -\frac{a^2}{h},$$

então devemos ter $a = 0$, porque, do contrário, o último limite não existiria.

Se $a \in \mathbb{R} \setminus \mathbb{Q}$, então, dado um número $h \in \mathbb{R} \setminus \mathbb{Q}$, ainda podemos ter $a + h \in \mathbb{Q}$. Nesse caso, temos

$$\frac{f(a+h) - f(a)}{h} = \frac{a^2 + 2ah + h^2}{h} = \frac{a^2}{h} + (2a + h).$$

O limite da última expressão quando $h \to 0$ existe se, e somente se, existe o limite de $\dfrac{a^2}{h}$ quando $h \to 0$. Como a é não nulo, o limite de $\dfrac{a^2}{h}$ quando $h \to 0$ não existe e, portanto, $\dot{f}(a)$ também não pode existir. $\square$

Exemplo 1.1.13. A função

$$f(x) := \begin{cases} x^2, & \text{se} \quad x \in \mathbb{Q}, \\ -x^2, & \text{se} \quad x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

é derivável apenas em 0.

Demonstração. Por um cálculo direto se verifica que $\dot{f}(0) = 0$. Como f é contínua apenas em 0, então, *a fortiori*, f não é derivável em nenhum outro ponto, pelo Teorema 1.1.6. $\square$

Há, no entanto, alguns exemplos simples (e extremamente importantes) de aplicações diferenciáveis.

Exemplo 1.1.14. Toda aplicação linear contínua $T : E \to F$ é diferenciável, com

$$T'(a) = T, \quad \text{para todo} \quad a \in E.$$

Demonstração. Como

$$\frac{\partial T}{\partial h}(a) = \lim_{t\to 0} \frac{T(a+th) - Ta}{t} = \lim_{t\to 0} Th = Th, \quad \text{para quaisquer} \quad a, h \in E,$$

então

$$\frac{T(a+h) - Ta - \dfrac{\partial T}{\partial h}(a)}{\|h\|_E} = \frac{T(a+h) - Ta - Th}{\|h\|_E} = 0_F, \quad \text{para quaisquer} \quad a, h \in E,$$

donde

$$\frac{T(a+h) - Ta - Th}{\|h\|_E} \xrightarrow{\|h\|_E \to 0} 0_F, \quad \text{para todo} \quad a \in E.$$

Portanto, T é diferenciável e

$$T'(a)h = Th, \quad \text{para quaisquer} \quad a, h \in E,$$

o que nos dá

$$T'(a) = T, \quad \text{para todo} \quad a \in E,$$

como queríamos demonstrar. $\square$

Em particular, se E e F são espaços normados e E tem dimensão finita, então toda aplicação linear $E \to F$ é diferenciável, já que, pelo Corolário C.4.12, toda aplicação linear $E \to F$ é contínua.

O exemplo abaixo, retirado de (COLEMAN, 2012, p. 59), mostra que, embora toda aplicação Fréchet-diferenciável seja Gâteaux-diferenciável, o inverso não é verdadeiro.

Exemplo 1.1.15. A função

$$f(x, y) := \begin{cases} \dfrac{x^6}{x^8 + (y - x^2)^2}, & \text{se} \quad (x, y) \neq (0, 0), \\ 0, \quad \text{se} \quad (x, y) = (0, 0), \end{cases}$$

é Gâteaux-diferenciável em $(0, 0)$, mas não é Fréchet-diferenciável em $(0, 0)$.

Demonstração. Dados $a, b \in \mathbb{R}$, temos

$$\begin{aligned} \frac{\partial f}{\partial(a, b)}(0, 0) = \lim_{t\to 0} \frac{f(t(a, b))}{t} &= \lim_{t\to 0} \frac{\dfrac{t^6 a^6}{t^8 a^8 + (bt - t^2 a^2)^2}}{t} \\ &= \lim_{t\to 0} \frac{t^5 a^6}{t^8 a^8 + (bt - t^2 a^2)^2} \\ &= \lim_{t\to 0} \frac{t^5 a^6}{t^8 a^8 + b^2 t^2 - 2bt^3 a^2 + t^4 a^4} \\ &= \lim_{t\to 0} \frac{t^3 a^6}{a^8 t^6 - 2a^2 bt + a^4 t^2 + b^2}, \end{aligned}$$

donde

$$\frac{\partial f}{\partial(a,b)}(0,0) = \begin{cases} \dfrac{0}{b^2} = 0, \quad \text{se} \quad b \neq 0, \\ \displaystyle\lim_{t\to 0} \frac{t^3a^6}{a^8t^6 + a^4t^2} = \lim_{t\to 0} \frac{ta^6}{a^8t^4 + a^4} = \frac{0}{a^4} = 0, \quad \text{se} \quad b = 0. \end{cases}$$

Logo,

$$\frac{\partial f}{\partial(a,b)}(0,0) = 0, \quad \text{para quaisquer} \quad a, b \in \mathbb{R},$$

e, portanto, $\dfrac{\partial f}{\partial(\cdot)}(0,0)$ é contínua e linear. No entanto, f não é contínua em $(0,0)$, pois,

$$f(t,t^2) = \frac{t^6}{t^8 + (t^2 - t^2)^2} = \frac{1}{t^2}, \quad \text{para todo} \quad t \in \mathbb{R}\setminus\{0\},$$

o que nos dá

$$\lim_{t\to 0} f(t,t^2) = \lim_{t\to 0} \frac{1}{t^2} = +\infty.$$

Por conseguinte, f não pode ser diferenciável em $(0,0)$, pelo Teorema 1.1.6. □

O exemplo abaixo, retirado de (LIMA, 2002, p. 9), mostra que a mera existência de todas as derivadas direcionais não é suficiente para garantir que uma aplicação é Gâteaux-diferenciável.

Exemplo 1.1.16. A função

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

não é Gâteaux-diferenciável em $(0,0)$ (e, consequentemente, não é diferenciável em $(0,0)$).

Demonstração. Se $a, b \in \mathbb{R}$ não são ambos nulos, então

$$\frac{\partial f}{\partial(a,b)}(0,0) = \lim_{t\to 0} \frac{f(ta,tb)}{t} = \lim_{t\to 0} \frac{t^3a^2b}{t^3a^2 + t^3b^2} = \frac{a^2b}{a^2+b^2} = f(a,b).$$

A continuidade de f em $(0,0)$ é assegurada pelo fato de que

$$\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{(x,y)\to(0,0)} \frac{x^2y}{x^2+y^2} = \lim_{(x,y)\to(0,0)} \frac{x^2}{x^2+y^2} \cdot y = 0 = f(0,0).$$

Logo, a aplicação

$$\frac{\partial f}{\partial(\cdot)}(0,0) = f$$

é contínua, mas não é linear e, portanto, f não é Gâteaux-diferenciável em $(0,0)$. □

Para o próximo exemplo, $\mathbb{R}[x]$ denotará o conjunto de todos os polinômios reais de uma variável. Como cada polinômio $a_0 + a_1x + \cdots + a_nx^n \in \mathbb{R}[x]$ pode ser visto como uma sequência de números reais eventualmente nula $(a_0, a_1, \ldots, a_n, 0, \ldots, 0, \ldots)$, é fácil ver que

$$\|a_0 + a_1x + \cdots + a_nx^n\|_\infty := \max_{k=0,\ldots,n} |a_k|$$

define uma norma em $\mathbb{R}[x]$.

Exemplo 1.1.17. (COLEMAN, 2012, p. 20) O operador linear $D : \mathbb{R}[x] \to \mathbb{R}[x]$ dado por

$$D(p) := \frac{d}{dt}p(t)$$

não é diferenciável em nenhum ponto.

Demonstração. Considere a sequência de polinômios

$$p_n(x) := \frac{1}{n}x^n.$$

Como

$$\lim_{n\to\infty} \|p_n\|_\infty = \lim_{n\to\infty} \frac{1}{n} = 0$$

e,

$$\|\dot{p}_n(x)\|_\infty = \|x^{n-1}\|_\infty = 1, \text{ para todo } n \in \mathbb{N},$$

então

$$\lim_{n\to\infty} \|D(p_n)\|_\infty = 1.$$

Como D é uma aplicação linear, então D é contínua se, e somente se,

$$\lim_{p\to 0} D(p) = 0_{\mathbb{R}[x]},$$

pelo Corolário C.4.2. Por conseguinte, D não é contínua em nenhum ponto e, portanto, também não pode ser diferenciável em nenhum ponto, pelo Teorema 1.1.6. □

Bem entendido, o exemplo anterior, em consonância com o Exemplo 1.1.14, deixa claro que aplicações lineares são contínuas se, e somente se, são diferenciáveis e que, em dimensão infinita, a continuidade de uma aplicação linear não decorre de sua linearidade, como é o caso em dimensão finita (ver Teorema C.4.12).

Exemplo 1.1.18. (COLEMAN, 2012, p. 54) Se $E \neq \{0\}$ é um espaço normado, nenhuma norma em E é diferenciável no vetor nulo.

Demonstração. Suponhamos, por redução ao absurdo, que $\|\cdot\|$ é uma norma diferenciável em 0. Desse modo, dado $h \neq 0$, temos

$$-\|h\| = \lim_{t\to 0^-} -\|h\| = \lim_{t\to 0^-} \frac{-t\|h\|}{t} = \lim_{t\to 0^-} \frac{\|th\|}{t} = \frac{\partial \|\cdot\|}{\partial h}(0) = \lim_{t\to 0^+} \frac{t\|h\|}{t} = \lim_{t\to 0} \|h\| = \|h\|,$$

donde $\|h\| = 0$ e, portanto, $h = 0$, o que é absurdo. □

1.2 Regras aritméticas

Para calcular diferenciais de aplicações mais complexas, utilizamos teoremas auxiliares, assim como se faz no cálculo diferencial de uma variável.

Teorema 1.2.1 (Regras aritméticas)**.** Se E e F são espaços normados, $U \subseteq E$ é um subconjunto aberto, $f, g : U \to F$ e $p : U \to \mathbb{R}$ são aplicações diferenciáveis em $a \in U$, então as aplicações $f + g, p \cdot g$ e λg (com $\lambda \in \mathbb{R}$) são diferenciáveis em a, valendo

1. $(f + g)'(a) = f'(a) + g'(a)$;
2. $(p \cdot g)'(a)h = p'(a)h \cdot g(a) + p(a) \cdot g'(a)h$;
3. $(\lambda g)'(a)h = \lambda g'(a)h$.

Além disso, se $p(a) \neq 0$, então

$$\left(\frac{1}{p}\right)'(a)h = -\frac{p'(a)h}{p(a)^2}.$$

Demonstração. Como f e g são diferenciáveis em a, então existem $r, s : E \to F$ tais que

$$\lim_{h \to 0_E} \frac{r(h)}{\|h\|_E} = \lim_{h \to 0_E} \frac{s(h)}{\|h\|_E} = 0_F.$$

e, para todo $h \in E$ tal que $a + h \in U$,

$$f(a + h) = f(a) + f'(a)h + r(h) \quad \text{e} \quad g(a + h) = g(a) + g'(a)h + s(h).$$

Daí,

$$\lim_{h \to 0_E} \frac{(f + g)(a + h) - (f + g)(a) - [f'(a) + g'(a)]h}{\|h\|_E} = \lim_{h \to 0_E} \frac{r(h) + s(h)}{\|h\|_E} = 0_F,$$

donde $f + g$ é diferenciável em a e

$$(f + g)'(a) = f'(a) + g'(a).$$

Com respeito à diferenciabilidade de $p \cdot g$, note que, para todo $h \in E$,

$$\begin{aligned}
\frac{\partial(p \cdot g)}{\partial h}(a) &= \lim_{t \to 0} \frac{(p \cdot g)(a + th) - (p \cdot g)(a)}{t} \\
&= \lim_{t \to 0} \frac{p(a + th)g(a + th) - p(a + th)g(a) + p(a + th)g(a) - p(a)g(a)}{t} \\
&= \lim_{t \to 0} \left[p(a + th)\frac{g(a + th) - g(a)}{t} + \frac{p(a + th) - p(a)}{t}g(a)\right] \\
&= p(a) \cdot g'(a)h + p'(a)h \cdot g(a).
\end{aligned}$$

É fácil ver que $\dfrac{\partial(p\cdot g)}{\partial(\cdot)}(a)$ é linear e contínua. Além disso, como p é diferenciável em a, então existe $q : E \to \mathbb{R}$ tal que

$$\lim_{h\to 0_E} \frac{q(h)}{\|h\|_E} = 0$$

e, para todo $h \in E$ tal que $a + h \in U$, temos

$$p(a+h) = p(a) + p'(a)h + q(h).$$

Desse modo, para todo $h \in E$ tal que $a + h \in U$, temos

$$\begin{aligned}
(p\cdot g)(a+h) - (p\cdot g)(a) - \frac{\partial(p\cdot g)}{\partial h}(a) &= p(a+h)[g(a+h) - g(a) - g'(a)h] \\
&\quad + [p(a+h) - p(a) - p'(a)h]g(a) \\
&\quad + [p(a+h) - p(a) - p'(a)h]g'(a)h \\
&\quad + p'(a)h \cdot g'(a)h \\
&= p(a+h)s(h) + q(h)\cdot g(a) \\
&\quad + q(h)\cdot g'(a)h + p'(a)h\cdot g'(a)h.
\end{aligned}$$

Como $p'(a)$ é linear e contínua, então, pelo Teorema C.4.1, é limitada na esfera, donde

$$\lim_{h\to 0}\left[p(a+h)\frac{s(h)}{\|h\|_E} + \frac{q(h)}{\|h\|_E}\cdot g(a) + \frac{q(h)}{\|h\|_E} g'(a)h + p'(a)\left(\frac{h}{\|h\|_E}\right)\cdot g'(a)h\right] = 0$$

e, portanto, $p\cdot g$ é diferenciável em a, valendo

$$(p\cdot g)'(a)h = p'(a)h\cdot g(a) + p(a)\cdot g'(a)h, \quad \text{para todo} \quad h\in E.$$

Desse modo, se $m(x) := \lambda$, para todo $x \in U$, então m é diferenciável, com

$$m'(a)h = 0, \quad \text{para todo} \quad h \in E,$$

donde $m\cdot g$ é diferenciável e, portanto, para todo $h \in E$

$$(\lambda g)'(a)h = (m\cdot g)'(a)h = m'(a)h\cdot g(a) + m(a)\cdot g'(a)h = \lambda g'(a)h.$$

Por fim, se $p(a) \neq 0$, a aplicação

$$T : h \in E \mapsto -\frac{1}{p(a)^2}\cdot p'(a)h$$

é, evidentemente, linear e contínua. Como, para todo $h \in E$ tal que $a + h \in U$,

$$\begin{aligned}
\frac{1}{p(a+h)} - \frac{1}{p(a)} + \frac{1}{p(a)^2}p'(a)h &= \frac{p(a)^2 - p(a+h)p(a) + p(a+h)p'(a)h}{p(a+h)p(a)^2} \\
&= \frac{-p(a)q(h) + q(h)p'(a)h + p'(a)h\cdot p'(a)h}{p(a+h)p(a)^2},
\end{aligned}$$

então

$$
\begin{aligned}
\lim_{h\to 0_E} \frac{\dfrac{1}{p(a+h)} - \dfrac{1}{p(a)} + \dfrac{1}{p(a)^2}p'(a)h}{\|h\|_E} &= \lim_{h\to 0_E} \frac{-p(a)q(h) + q(h)p'(a)h + p'(a)h \cdot p'(a)h}{\|h\|_E p(a+h)p(a)^2} \\
&= \lim_{h\to 0_E} \frac{1}{p(a+h)p(a)^2} \cdot \lim_{h\to 0_E} l_h \\
&= \frac{1}{p(a)^3} \cdot 0 = 0,
\end{aligned}
$$

onde

$$
l_h := [p'(a)h - p(a)]\frac{q(h)}{\|h\|_E} + p'(a)\left(\frac{h}{\|h\|_E}\right) \cdot p'(a)h
$$

Logo, $\dfrac{1}{p}$ é diferenciável em a, com

$$
\left(\frac{1}{p}\right)'(a)h = -\frac{p'(a)h}{p(a)^2}, \quad \text{para cada} \quad h \in E,
$$

como queríamos demonstrar. □

1.3 A regra da cadeia

Muitas das vezes, o exercício de diferenciar uma aplicação é facilitado quando a enxergamos como uma composição de aplicações cujas derivadas são fáceis de calcular e aplicamos a *Regra da Cadeia.*

Teorema 1.3.1 (Regra da Cadeia)**.** Sejam E, F e V espaços normados, $X \subseteq E, Y \subseteq F$ subconjuntos abertos, $x_0 \in X$ e $f : Y \to V, g : X \to F$ aplicações, com $g(X) \subseteq Y$. Nessas condições, se g é diferenciável em x_0 e f é diferenciável em $g(x_0)$, então a aplicação $f \circ g : X \to V$ também é diferenciável em x_0, valendo

$$
(f \circ g)'(x_0) = f'(g(x_0)) \circ g'(x_0).
$$

Demonstração. Como g é diferenciável em x_0 e f é diferenciável em $g(x_0)$, então

$$
\begin{aligned}
g(x_0 + h) &= g(x_0) + g'(x_0)h + r(h), \quad \text{com} \quad r(h) = o(\|h\|), \\
f(g(x_0) + u) &= f(g(x_0)) + f'(g(x_0))u + s(u), \quad \text{com} \quad s(u) = o(\|u\|).
\end{aligned}
$$

Isso nos dá

$$
\begin{aligned}
(f \circ g)(x_0 + h) - (f \circ g)(x_0) &= f(g(x_0 + h)) - f(g(x_0)) \\
&= f(g(x_0) + \underbrace{g'(x_0)h + r(h)}_{\in F}) - f(g(x_0)) \\
&= f'(g(x_0))(g'(x_0)h + r(h)) + s(g'(x_0)h + r(h)) \\
&= (f'(g(x_0)) \circ g'(x_0))h + \\
&\quad + f'(g(x_0))(r(h)) + s(g'(x_0)h + r(h)),
\end{aligned}
$$

donde vem

$$(f \circ g)(x_0 + h) - (f \circ g)(x_0) - (f'(g(x_0)) \circ g'(x_0))h = f'(g(x_0))(r(h)) + s(g'(x_0)h + r(h)).$$

Precisamos mostrar que

$$R(h) := f'(g(x_0))(r(h)) + s(g'(x_0)h + r(h)) = o(\|h\|).$$

Veja que

$$\|f'(g(x_0))(r(h))\| \le \|f'(g(x_0))\|\|r(h)\| = o(\|h\|).$$

É suficiente então mostrar que

$$S(h) := s(g'(x_0)h + r(h)) = o(\|h\|).$$

Defina então

$$\tilde{s}(u) := \begin{cases} s(u)/\|u\|, & \text{se} \quad u \neq 0, \\ 0_V, & \text{se} \quad u = 0. \end{cases}$$

Isso nos dá

$$s(u) = \tilde{s}(u)\|u\|, \quad \text{para todo} \quad u \in F.$$

Definindo $u(h) := g'(x_0)h + r(h)$, temos

$$S(h)/\|h\| = s(u(h))/\|h\| = \tilde{s}(u(h))\frac{\|u(h)\|}{\|h\|}.$$

Fazendo $h \to 0$, temos $g'(x_0)h \to 0$, pela continuidade de $g'(x_0)$, e

$$r(h) = \|h\| \cdot (r(h)/\|h\|) \to 0,$$

donde $u(h) \to 0$ e, logo $\tilde{s}(u(h)) \to 0$. Além disso,

$$\begin{aligned} \|u(h)\| &= \|g'(x_0)h + r(h)\| \le \|g'(x_0)h\| + \|r(h)\| \\ &\le \|g'(x_0)\|\|h\| + \|r(h)\| = O(\|h\|) + o(\|h\|) = O(\|h\|), \end{aligned}$$

o que termina a prova. □

Dizemos que uma aplicação $f : U \subseteq E \to F$ é C^1 se f é diferenciável e a aplicação $f' : U \to \mathcal{L}(E, F)$ é contínua.

Corolário 1.3.2. Sejam E, F e V espaços normados, $X \subseteq E, Y \subseteq F$ subconjuntos abertos e $f : Y \to V, g : X \to F$ aplicações, com $g(X) \subseteq Y$. Se f e g são C^1, então $h := f \circ g : X \to V$ também é C^1.

Demonstração. Para demonstrar que h é de classe C^1, precisamos mostrar que h é diferenciável em todo ponto de X e que a aplicação derivada $h' : X \to \mathcal{L}(E, V)$ é contínua.

Como g é diferenciável em $x \in X$ e f é diferenciável em $g(x) \in Y$, pela Regra da Cadeia (ver Teorema 1.3.1), a composição $h = f \circ g$ é diferenciável em x e sua derivada é dada por

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

Vamos mostrar agora que h' é contínua. Considere a aplicação bilinear $B : \mathcal{L}(F, V) \times \mathcal{L}(E, F) \to \mathcal{L}(E, V)$ definida por

$$B(T, S) = T \circ S.$$

Temos

$$h'(x) = B(f'(g(x)), g'(x)).$$

Observe que:

- g' é contínua, já que g é C^1 por hipótese.
- A aplicação $f' \circ g$ é a composição da aplicação contínua g (já que g é diferenciável) com a função contínua f' (já que f é C^1). Logo, é contínua.
- A aplicação bilinear B é limitada, pois

  $$\|B(T, S)\| = \|T \circ S\| \leq \|T\|\|S\|,$$

  o que implica que B é contínua.

Então $h' = B \circ \Phi$, onde

$$\Phi(x) := (f'(g(x)), g'(x)),$$

que é uma aplicação contínua. Logo, h' é a composição de duas aplicações contínuas, portanto h' é contínua, provando que h é C^1, como queríamos demonstrar. $\square$

Espaços-produto aparecem de modo natural no Cálculo. O produto interno, o determinante, o produto vetorial e, como veremos em breve, as próprias diferenciais (de ordem superior) são alguns exemplos naturais de aplicações definidas em espaços-produto. Quando estamos diante de uma aplicação cujas imagens caem em algum espaço-produto, podemos analisar a aplicação a partir das suas *aplicações-coordenadas*.

Teorema 1.3.3 (Diferenciação coordenada a coordenada)**.** Sejam $E, E_1, \ldots, E_n$ espaços normados, $U \subseteq E$ um subconjunto aberto e $f : U \to E_1 \times \cdots \times E_n$ uma aplicação. Se $f_i : U \to E_i$ são as aplicações-coordenadas de f, então f é diferenciável em $a \in U$ se, e somente se, $f_1, \ldots, f_n$ são diferenciáveis em a. Nesse caso,

$$f'(a)h = (f_1'(a)h, \ldots, f_n'(a)h), \text{ para todo } h \in E.$$

Demonstração. Se $f, f_1, \ldots, f_n$ são diferenciáveis em a, temos

$$\begin{aligned}
f'(a)h = \lim_{t\to 0} \frac{f(a+th)-f(a)}{t} &= \lim_{t\to 0} \frac{(f_1(a+th), \ldots, f_n(a+th)) - (f_1(a), \ldots, f_n(a))}{t} \\
&= \lim_{t\to 0} \frac{(f_1(a+th)-f_1(a), \ldots, f_n(a+th)-f_n(a))}{t} \\
&= \lim_{t\to 0} \left(\frac{f_1(a+th)-f_1(a)}{t}, \ldots, \frac{f_n(a+th)-f_n(a)}{t} \right) \\
&= \left(\lim_{t\to 0} \frac{f_1(a+th)-f_1(a)}{t}, \ldots, \lim_{t\to 0} \frac{f_n(a+th)-f_n(a)}{t} \right) \\
&= (f_1'(a)h, \ldots, f_n'(a)h)\,.
\end{aligned}$$

Vamos provar agora que f é diferenciável em a se, e somente se, cada f_i é diferenciável em a. Com efeito, $\dfrac{\partial f}{\partial(\cdot)}(a)$ é linear e contínua se, e somente se, cada $\dfrac{\partial f_i}{\partial(\cdot)}(a)$ é linear e contínua. Resta verificar que, pondo

$$r(h) := f(a+h) - f(a) - \frac{\partial f}{\partial h}(a) \quad \text{e} \quad r_i(h) := f_i(a+h) - f_i(a) - \frac{\partial f_i}{\partial h}(a),$$

para todo $h \in E$ tal que $a+h \in U$, e $i = 1, \ldots, n$, temos

$$\lim_{h\to 0} \frac{r(h)}{\|h\|_E} = 0 \iff \lim_{h\to 0} \frac{r_i(h)}{\|h\|_E} = 0, \quad \text{para cada} \quad i = 1, \ldots, n.$$

Por um cálculo direto, temos

$$\frac{r(h)}{\|h\|_E} = \left(\frac{r_1(h)}{\|h\|_E}, \ldots, \frac{r_n(h)}{\|h\|_E} \right),$$

donde se segue o resultado. □

Corolário 1.3.4. Sejam E e F espaços normados, $\{v_i\}_{i=1}^n$ uma base de F, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ uma aplicação. Se $f_1, \ldots, f_n : U \to \mathbb{R}$ são aplicações tais que

$$f(x) = \sum_{i=1}^n f_i(x) v_i, \quad \text{para cada} \quad x \in U,$$

então f é diferenciável em $a \in U$ se, e somente se, $f_1, \ldots, f_n$ são diferenciáveis em a.

Nesse caso,

$$f'(a)h = \sum_{i=1}^n f_i'(a)h \cdot v_i, \quad \text{para cada} \quad h \in E.$$

Demonstração. Suponhamos que f é diferenciável em a e considere $F_i := \operatorname{span}\{v_i\}$, para cada $i = 1, \ldots, n$, e $T : F \to F_1 \times \cdots \times F_n$ o isomorfismo linear dado por

$$T\left(\sum_{i=1}^n \alpha_i v_i \right) = (\alpha_1 v_1, \ldots, \alpha_n v_n), \quad \text{para cada} \quad \{\alpha_i\}_{i=1}^n \subseteq \mathbb{R}\,.$$

Como F possui dimensão finita e T é linear, então T é contínua, pelo Corolário C.4.12, donde T é diferenciável, pelo Exemplo 1.1.14, com $T'(x) = T$, para todo $x \in F$. Pela Regra da Cadeia (ver Teorema 1.3.1), a aplicação

$$T \circ f : U \to F_1 \times \cdots \times F_n$$

é diferenciável em a, com

$$(T \circ f)'(a)h = T'(f(a))(f'(a)h) = T(f'(a)h).$$

Por outro lado, como

$$(T \circ f)(x) = T(f(x)) = T\left(\sum_{i=1}^{n} f_i(x)v_i\right) = (f_1(x)v_1, \ldots, f_n(x)v_n),$$

então, pelo Teorema 1.3.3,

$$(T \circ f)'(a)h = ((f_1'(a)h)v_1, \ldots, (f_n'(a)h)v_n) = T\left(\sum_{i=1}^{n} f_i'(a)h \cdot v_i\right).$$

Pela injetividade de T, temos

$$f'(a)h = \sum_{i=1}^{n} f_i'(a)h \cdot v_i,$$

como queríamos demonstrar.

Reciprocamente, se $f_1, \ldots, f_n$ são diferenciáveis, então cada $g_i(x) := f_i(x)v_i$ é diferenciável, como composição

$$g_i = M_{v_i} \circ f_i$$

de duas aplicações diferenciáveis ($M_{v_i}(\alpha) := \alpha v_i$ é a multiplicação por v_i, que é linear). Logo, pelo Teorema 1.3.3,

$$(T \circ f)(x) = (g_1(x), \ldots, g_n(x))$$

é diferenciável. Como T é isomorfismo linear, T^{-1} está bem definida. Além disso, o domínio de T^{-1} é $F_1 \times \cdots \times F_n$, que tem dimensão finita, logo T^{-1} é contínua e, portanto, diferenciável. Pela Regra da Cadeia,

$$f = T^{-1} \circ (T \circ f)$$

é diferenciável. □

1.4 Derivando transformações multilineares

Nesta seção veremos uma fórmula geral para o cálculo da derivada de transformações k-lineares, qualquer que seja o $k \in \mathbb{N}$.

Teorema 1.4.1. Sejam $E_1, \ldots, E_n$ e F espaços vetoriais, $V := E_1 \times \cdots \times E_n$ e $f : V \to F$ uma aplicação n-linear. Assim, para quaisquer $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in V$, temos

$$f(a_1 + b_1, \ldots, a_n + b_n) = \sum_{0 \le k \le n} \sum_{\substack{c_i \in \{a_i, b_i\},\ \text{para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, c_i = b_i\}| = k}} f(c_1, \ldots, c_n). \tag{1.2}$$

Se, além disso, $E_1, \ldots, E_n$ e F são espaços normados e f é contínua, então, f é diferenciável, valendo, para quaisquer $(a_1, \ldots, a_n), (h_1, \ldots, h_n) \in V$,

$$f'(a_1, \ldots, a_n)(h_1, \ldots, h_n) = \sum_i f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_n).$$

Demonstração. Como f é multilinear, então

$$f(a_1 + b_1, \ldots, a_n + b_n) = \sum_{c_i \in \{a_i, b_i\},\ \text{para todo } i=1,\ldots,n} f(c_1, \ldots, c_n).$$

Por outro lado, se $c_i \in \{a_i, b_i\}$, para cada $i = 1, \ldots, n$, então $c_i = b_i$ para k índices i, com $0 \le k \le n$. Logo,

$$\sum_{c_i \in \{a_i, b_i\},\ \text{para todo } i=1,\ldots,n} f(c_1, \ldots, c_n) = \sum_{0 \le k \le n} \sum_{\substack{c_i \in \{a_i, b_i\},\ \text{para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, c_i = b_i\}| = k}} f(c_1, \ldots, c_n).$$

Isto prova a validade de (1.2).

Suponhamos agora que $E_1, \ldots, E_n$ e F são espaços normados, f é contínua e considere $(a_1, \ldots, a_n), (h_1, \ldots, h_n) \in V$ e $t \in \mathbb{R}$. De (1.2), temos

$$f(a_1 + th_1, \ldots, a_n + th_n) = f(a_1, \ldots, a_n) + \sum_i f(a_1, \ldots, a_{i-1}, th_i, a_{i+1}, \ldots, a_n) + R(t, h),$$

onde

$$\begin{aligned} R(t,h) &:= \sum_{2 \le k \le n} \sum_{\substack{b_i \in \{a_i, th_i\},\ \text{para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = th_i\}| = k}} f(b_1, \ldots, b_n) \\ &= \sum_{2 \le k \le n} t^k \sum_{\substack{b_i \in \{a_i, h_i\},\ \text{para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = h_i\}| = k}} f(b_1, \ldots, b_n). \end{aligned}$$

Assim, como

$$\frac{f(a + th) - f(a)}{t} = \sum_i f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_n) + \frac{R(t,h)}{t},$$

e

$$\lim_{t \to 0} \frac{R(t,h)}{t} = \lim_{t \to 0} \sum_{2 \le k \le n} t^{k-1} \sum_{\substack{b_i \in \{a_i, h_i\},\ \text{para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = h_i\}| = k}} f(b_1, \ldots, b_n) = 0_F,$$

então

$$\frac{\partial f}{\partial(h_1, \ldots, h_n)}(a_1, \ldots, a_n) = \sum_i f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_n).$$

Para simplificar a notação, vamos chamar $\dfrac{\partial f}{\partial(\cdot)}(a)$ de T. Argumentaremos agora que T é linear e contínua.

Com efeito, sejam $u = (u_1, \ldots, u_n), v = (v_1, \ldots, v_n) \in V$ e $\lambda \in \mathbb{R}$. Então,

$$\begin{aligned} T(u + \lambda v) &= \sum_i f(a_1, \ldots, a_{i-1}, u_i + \lambda v_i, a_{i+1}, \ldots, a_n) \\ &= \sum_i f(a_1, \ldots, a_{i-1}, u_i, a_{i+1}, \ldots, a_n) + \lambda \sum_i f(a_1, \ldots, a_{i-1}, v_i, a_{i+1}, \ldots, a_n) \\ &= T(u) + \lambda T(v). \end{aligned}$$

Isto prova que T é linear.

Com respeito à continuidade de T, como f é n-linear, então, pelo Teorema C.7.1, dado $h = (h_1, \ldots, h_n) \in V$, temos

$$\begin{aligned} \|T(h)\|_F &= \left\| \sum_i f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_n) \right\|_F \\ &\leq \sum_i \|f(a_1, \ldots, a_{i-1}, h_i, a_{i+1}, \ldots, a_n)\|_F \\ &\leq \sum_i \left(\|f\|_{\mathcal{L}(E_1, \ldots, E_n; F)} \prod_{\substack{j=1,\ldots,n, \\ j \neq i}} \|a_j\|_{E_j} \|h_i\|_{E_i} \right) \\ &\leq \sum_i \|f\|_{\mathcal{L}(E_1, \ldots, E_n; F)} \|a\|_\infty^{n-1} \|h\|_\infty \\ &\leq n \|f\|_{\mathcal{L}(E_1, \ldots, E_n; F)} \|a\|_\infty^{n-1} \|h\|_\infty. \end{aligned}$$

Logo, pelo Teorema C.4.1, a aplicação T é contínua.

Como

$$R(1, h) = f(a + h) - f(a) - T(h), \text{ para todo } h \in V,$$

resta apenas verificar que

$$\frac{\|R(1, h)\|_F}{\|h\|_\infty} \xrightarrow{\|h\|_\infty \to 0} 0.$$

Da desigualdade triangular,

$$\|R(1, h)\|_F \leq \sum_{2 \leq k \leq n} \sum_{\substack{b_i \in \{a_i, h_i\}, \text{ para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = h_i\}| = k}} \|f(b_1, \ldots, b_n)\|_F \,.$$

Analisemos mais de perto o somatório

$$\sum_{\substack{b_i \in \{a_i, h_i\}, \text{ para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = h_i\}| = k}} \|f(b_1, \ldots, b_n)\|_F.$$

Como há $\binom{n}{k}$ maneiras de escolher a n-upla $(b_1, \ldots, b_n)$, onde cada $b_i \in \{a_i, h_i\}$ e $b_j = h_j$ para k índices j, então há $\binom{n}{k}$ termos $f(b_1, \ldots, b_n)$ a serem somados. Além disso, fixada uma tal n-upla, como f é uma aplicação multilinear contínua, temos

$$\|f(b_1,\dots,b_n)\|_F \leq \|f\|_{\mathcal{L}(E_1,\dots,E_n;F)} \prod_{b_i=h_i} \|b_i\|_{E_i} \prod_{b_j=a_j} \|b_j\|_{E_j} \leq \|f\|_{\mathcal{L}(E_1,\dots,E_n;F)} \|h\|_\infty^k \|a\|_\infty^{n-k}.$$

Nessas condições,

$$\sum_{\substack{b_i\in\{a_i,h_i\},\text{ para todo } i=1,\dots,n,\\ \text{e } |\{i=1,\dots,n\,:\,b_i=h_i\}|=k}} \|f(b_1,\dots,b_n)\|_F \leq \binom{n}{k} \|f\|_{\mathcal{L}(E_1,\dots,E_n;F)} \|h\|_\infty^k \|a\|_\infty^{n-k},$$

donde

$$\|R(1,h)\|_F \leq \sum_{2\leq k\leq n} \binom{n}{k} \|f\|_{\mathcal{L}(E_1,\dots,E_n;F)} \|h\|_\infty^k \|a\|_\infty^{n-k}$$

e, portanto, para $h \neq 0$, temos

$$\frac{\|R(1,h)\|_F}{\|h\|_\infty} \leq \sum_{2\leq k\leq n} \binom{n}{k} \|f\|_{\mathcal{L}(E_1,\dots,E_n;F)} \|h\|_\infty^{k-1} \|a\|_\infty^{n-k}.$$

Se $k \geq 2$, então $k-1 \geq 1$. Assim, no lado direito da última desigualdade destacada, todos os termos são múltiplos de $\|h\|_\infty$, o que faz com que o lado direito convirja para zero quando fazemos $\|h\|_\infty \to 0$. Pelo Teorema do Confronto (ver Teorema B.5.10),

$$\lim_{\|h\|_\infty\to 0} \frac{\|R(1,h)\|_F}{\|h\|_\infty} = 0,$$

como queríamos demonstrar. □

Para o próximo corolário, $\mathcal{M}_n(\mathbb{R})$ denota o conjunto das matrizes quadradas de ordem n com entradas em $\mathbb{R}$ e $\mathrm{GL}_n(\mathbb{R})$ é o conjunto das matrizes $A \in \mathcal{M}_n(\mathbb{R})$ invertíveis.

Corolário 1.4.2 (Regra do determinante)**.** O determinante é diferenciável, valendo

$$\det{}'(A)H = \det A \cdot \operatorname{tr}(A^{-1}H), \quad \text{para quaisquer} \quad A \in \mathrm{GL}_n(\mathbb{R}) \quad \text{e} \quad H \in \mathcal{M}_n(\mathbb{R}).$$

Demonstração. Sejam $H = (h_1,\dots,h_n)$ e $I = (e_1,\dots,e_n)$, onde $\{e_i\}$ é a base canônica de $\mathbb{R}^n$ e $h_j := (h_j^1,\dots,h_j^n) \in \mathbb{R}^n$, para cada $j = 1,\dots,n$. Note que

$$\det{}'(I)H = \sum_{i=1}^n \det(e_1,\dots,e_{i-1},h_i,e_{i+1},\dots,e_n) = \sum_{i=1}^n h_i^i = \operatorname{tr} H.$$

Assim, dados $A \in \mathrm{GL}_n(\mathbb{R})$ e $H \in \mathcal{M}_n(\mathbb{R})$, temos

$$\begin{aligned}\det{}'(A)H = \lim_{t\to 0} \frac{\det(A+tH)-\det(A)}{t} &= \det A \cdot \lim_{t\to 0} \frac{\det(I+tA^{-1}H)-\det(I)}{t}\\ &= \det A \cdot \det{}'(I)(A^{-1}H)\\ &= \det A \cdot \operatorname{tr}(A^{-1}H),\end{aligned}$$

como queríamos demonstrar. □

Corolário 1.4.3 (Regra da transformação bilinear)**.** Sejam E, F_1, F_2 e W espaços normados, $U \subseteq E$ um subconjunto aberto, $B : F_1 \times F_2 \to W$ uma forma bilinear contínua, $a \in U$ e $f : U \to F_1$ e $g : U \to F_2$ aplicações diferenciáveis em a. Nessas condições, a aplicação

$$B(f, g) : U \to W$$

dada por

$$B(f, g)(x) := B(f(x), g(x)), \quad \text{para cada} \quad x \in U,$$

é diferenciável em a, com

$$B'(f, g)(a)h = B(f'(a)h, g(a)) + B(f(a), g'(a)h), \quad \text{para todo} \quad h \in E.$$

Demonstração. Considere a aplicação

$$(f, g) : U \to F_1 \times F_2$$

dada por

$$(f, g)(x) := (f(x), g(x)), \quad \text{para cada} \quad x \in U.$$

Pelo Teorema 1.3.3, essa aplicação é diferenciável, valendo

$$(f, g)'(a)h = (f'(a)h, g'(a)h), \quad \text{para cada} \quad h \in E.$$

Como B é uma aplicação bilinear contínua, então B é diferenciável, pelo Teorema 1.4.1, donde, pela Regra da Cadeia (ver Teorema 1.3.1), a composição $B \circ (f, g)$ é diferenciável em a, valendo, para cada $h \in E$,

$$\begin{aligned}[B \circ (f, g)]'(a)h &= B'(f(a), g(a))(f'(a)h, g'(a)h) \\ &= B(f'(a)h, g(a)) + B(f(a), g'(a)h).\end{aligned}$$

Como

$$B(f, g) = B \circ (f, g),$$

então

$$B(f, g)'(a)h = B(f'(a)h, g(a)) + B(f(a), g'(a)h), \quad \text{para todo} \quad h \in E,$$

como queríamos demonstrar. □

Para o próximo exemplo, relembre a definição de produto interno (ver Definição C.5.1).

Corolário 1.4.4 (Regra do produto interno)**.** Sejam E um espaço normado, F um espaço com produto interno, $U \subseteq E$ um subconjunto aberto e $f, g : U \to F$ aplicações diferenciáveis em $a \in U$. Nessas condições, a função

$$\langle f, g \rangle : U \to \mathbb{R}$$

dada por

$$\langle f, g\rangle(x) := \langle f(x), g(x)\rangle, \quad \text{para cada} \quad x \in U,$$

é diferenciável em a, com

$$\langle f, g\rangle'(a)h = \langle f'(a)h, g(a)\rangle + \langle f(a), g'(a)h\rangle, \quad \text{para todo} \quad h \in E.$$

Demonstração. Pelo Teorema C.5.4, o produto interno $\langle\cdot,\cdot\rangle$ é uma forma bilinear contínua. O resultado então se segue do Corolário 1.4.3. □

Para o próximo corolário, relembre a definição de produto vetorial (ver Exemplo C.7.5).

Corolário 1.4.5 (Regra do produto vetorial)**.** Se E é um espaço normado, $U \subseteq E$ é um subconjunto aberto e $f, g : U \to \mathbb{R}^3$ são aplicações diferenciáveis em $a \in U$, então

$$f \wedge g : U \to \mathbb{R}^3$$

dada por

$$(f \wedge g)(x) := f(x) \wedge g(x), \quad \text{para cada} \quad x \in U,$$

é uma aplicação diferenciável em a, valendo

$$(f \wedge g)'(a)h = f'(a)h \wedge g(a) + f(a) \wedge g'(a)h, \quad \text{para todo} \quad h \in E.$$

Demonstração. Pelo Corolário C.7.5, o produto vetorial é uma forma bilinear contínua. O resultado então se segue do Corolário 1.4.3. □

1.5 Funções diferenciáveis em espaços de Hilbert

Sejam H um espaço de Hilbert, $U \subseteq H$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função diferenciável em $a \in U$. Como $f'(a) : H \to \mathbb{R}$ é um funcional linear contínuo, então, pelo Teorema da Representação de Riesz (ver Teorema C.9.4), existe um único vetor $\nabla f(a) \in H$ tal que

$$\|\nabla f(a)\|_H = \|f'(a)\|_{H^*}$$

e

$$f'(a)h = \langle h, \nabla f(a)\rangle\,, \quad \text{para todo} \quad h \in E.$$

O vetor $\nabla f(a)$ é chamado o **gradiente** de f em a. Quando f for diferenciável em todos os pontos de U, a aplicação

$$\nabla f : x \in U \mapsto \nabla f(x) \in H$$

será chamada a **aplicação gradiente** de f. Quando essa aplicação for diferenciável em a, o operador linear contínuo
$$\nabla^2 f(a) := (\nabla f)'(a)$$
será chamado de **operador hessiano** de f em a. Além disso, se a aplicação grandiente de f for diferenciável em todos os pontos de U, a aplicação

$$\nabla^2 f : x \in U \mapsto \nabla^2 f(x) \in \mathcal{L}(H)$$

será chamada a **aplicação hessiana** de f.

O próximo Teorema apresenta o "rosto" do gradiente em espaços de dimensão finita.

Proposição 1.5.1 (Vetor gradiente em dimensão finita)**.** Se E é um espaço com produto interno, $U \subseteq E$ é um subconjunto aberto, $f : U \to \mathbb{R}$ é uma função diferenciável em $a \in U$ e $\{e_k\}_{k=1}^n \subseteq E$ é uma base ortonormal de E, então

$$\nabla f(a) = \sum_{i=1}^n \frac{\partial f}{\partial e_i}(a) e_i.$$

Demonstração. Como E tem dimensão finita, a norma induzida pelo produto interno necessariamente torna E num espaço de Banach (ver Corolario C.4.11), logo E com o produto interno é um espaço de Hilbert. Como $\nabla f(a) \in E$, existem escalares $\alpha_1, \ldots, \alpha_n$ tais que
$$\nabla f(a) = \sum_{i=1}^n \alpha_i e_i.$$
Daí, dado $k = 1, \ldots, n$,

$$\frac{\partial f}{\partial e_k}(a) = f'(a)e_k = \langle e_k, \nabla f(a)\rangle = \left\langle e_k, \sum_{i=1}^n \alpha_i e_i \right\rangle = \sum_{i=1}^n \alpha_i \langle e_k, e_i\rangle = \sum_{i=1}^n \alpha_i \delta_{ki} = \alpha_k.$$

Logo,
$$\nabla f(a) = \sum_{i=1}^n \frac{\partial f}{\partial e_i}(a) e_i,$$
como queríamos demonstrar. □

Várias fórmulas que vemos nos cursos de Cálculo Diferencial de várias variáveis podem ser facilmente deduzidas com as ferramentas que já desenvolvemos até aqui.

Proposição 1.5.2. (COLEMAN, 2012, Exercise 2.3, p. 45) Se $I \subseteq \mathbb{R}$ é um intervalo aberto, $U \subseteq \mathbb{R}^n$ é um subconjunto aberto, $f : I \to U$ é uma aplicação diferenciável em $a \in I$ e $g : U \to \mathbb{R}$ é uma função diferenciável em $f(a)$, então

$$\frac{d(g \circ f)}{dx}(a) = \sum_{i=1}^n \frac{\partial g}{\partial x_i}(f(a)) \frac{df_i}{dx}(a).$$

Demonstração. Pela Regra da Cadeia (ver Teorema 1.3.1), $g \circ f$ é diferenciável em a, com

$$\begin{aligned}(g \circ f)'(a)h = g'(f(a))(f'(a)h) &= \langle (f_1'(a)h, \ldots, f_n'(a)h), \nabla g(f(a)) \rangle \\ &= \left\langle \sum_{i=1}^{n} f_i'(a)h \cdot e_i, \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(f(a)) \cdot e_i \right\rangle \\ &= \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(f(a)) f_i'(a)h,\end{aligned}$$

donde

$$\frac{d(g \circ f)}{dx}(a) = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(f(a)) \frac{df_i}{dx}(a),$$

como queríamos demonstrar. □

1.6 A matriz jacobiana e a matriz hessiana

Sejam E e F espaços normados, $V := \{u_i\}_{i=1}^{n}$ e $W := \{v_i\}_{i=1}^{m}$ bases de E e F, respectivamente, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ uma aplicação diferenciável em $a \in U$. Para cada $x \in U$, existem únicos $f_1(x), \ldots, f_m(x) \in \mathbb{R}$ tais que

$$f(x) = \sum_{i=1}^{m} f_i(x) v_i.$$

Pelo Corolário 1.3.4,

$$f'(a)h = \sum_{i=1}^{m} f_i'(a)h \cdot v_i, \text{ para todo } h \in E,$$

donde

$$[f'(a)]_W^V = \begin{bmatrix} \dfrac{\partial f_1}{\partial u_1}(a) & \cdots & \dfrac{\partial f_1}{\partial u_n}(a) \\ \vdots & \cdots & \vdots \\ \dfrac{\partial f_m}{\partial u_1}(a) & \cdots & \dfrac{\partial f_m}{\partial u_n}(a) \end{bmatrix}.$$

Em particular, se a norma de E é proveniente de um produto interno e V é um conjunto ortonormal, como

$$\nabla f_i(a) = \sum_{j=1}^{n} \frac{\partial f_i}{\partial u_j}(a) u_j,$$

então

$$[f'(a)]_W^V = \begin{bmatrix} \nabla f_1(a)^t \\ \vdots \\ \nabla f_m(a)^t \end{bmatrix} = \begin{bmatrix} \nabla f_1(a) & \cdots & \nabla f_m(a) \end{bmatrix}^t.$$

Quando $E = \mathbb{R}^n$, $F = \mathbb{R}^m$ e V e W são as bases canônicas de E e F, respectivamente, a matriz

$$[f'(a)] := [f'(a)]_W^V$$

é chamada a **matriz jacobiana** de f no ponto a. Quando $m = n$, a matriz $[f'(a)]$ é uma matriz quadrada cujo determinante é chamado de **jacobiano** de f no ponto a.

Um exemplo extremamente útil de aplicação do conceito de matriz jacobiana é na definição de *matriz hessiana.*

Exemplo 1.6.1 (Matriz hessiana)**.** Sejam $U \subseteq \mathbb{R}^n$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função diferenciável cuja aplicação gradiente $\nabla f : U \to \mathbb{R}^n$ é diferenciável em $a \in U$. Como

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

então cada $\dfrac{\partial f}{\partial x_i}$ é diferenciável em a. A matriz jacobiana de ∇f no ponto a é denotada por $\mathrm{H}\, f(a)$ e chamada a **matriz hessiana** de f em a, isto é,

$$\mathrm{H}\, f(a) := [(\nabla f)'(a)] = \left[\nabla \frac{\partial f}{\partial e_1}(a) \quad \dots \quad \nabla \frac{\partial f}{\partial e_n}(a)\right]^t = \begin{bmatrix} \dfrac{\partial^2 f}{\partial e_1 \partial e_1}(a) & \cdots & \dfrac{\partial^2 f}{\partial e_n \partial e_1}(a) \\ \vdots & \vdots & \vdots \\ \dfrac{\partial^2 f}{\partial e_1 \partial e_n}(a) & \cdots & \dfrac{\partial^2 f}{\partial e_n \partial e_n}(a) \end{bmatrix},$$

onde

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

Note que

$$\mathrm{H}\, f(a) = [\nabla^2 f(a)].$$

O **hessiano** de f em a é o jacobiano do gradiente de f em a, isto é, o determinante da matriz hessiana $\mathrm{H}\, f(a)$. Além disso, o traço da matriz hessiana é denotado por $\Delta f(a)$ e é chamado de **laplaciano** de f em a, isto é,

$$\Delta f(a) := \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2}(a).$$

1.7 Diferenciais parciais

Esta seção é inteiramente baseada em (COLEMAN, 2012, pp. 66-69) será especialmente útil para o Teorema da Aplicação Implícita (ver Teorema 4.2.1).

Sejam $E_1, \dots, E_n, F$ espaços normados, $E := E_1 \times \cdots \times E_n$, $U \subseteq E$ um subconjunto aberto, $f : U \to F$ uma aplicação e $a \in U$. Consideremos a aplicação

$$f_{a,k}(h_k) := f(a_1, \dots, a_{k-1}, h_k, a_{k+1}, \dots, a_n).$$

Então $f_{a,k}$ está definido num subconjunto aberto de E_k:

$$H_{a,k} := \{h_k \in E_k : (a_1, \ldots, a_{k-1}, h_k, a_{k+1}, \ldots, a_n) \in U\}.$$

Vamos argumentar brevemente que todo ponto de $H_{a,k}$ é interior. Dado $h \in H_{a,k}$, como $(a_1, \ldots, a_{k-1}, h, a_{k+1}, \ldots, a_n) \in U$, existe $\varepsilon > 0$ tal que

$$B((a_1, \ldots, a_{k-1}, h, a_{k+1}, \ldots, a_n), \varepsilon) \subseteq U.$$

Dado $y \in B(h, \varepsilon) \subseteq E_k$, temos

$$\|(a_1, \ldots, a_{k-1}, h, a_{k+1}, \ldots, a_n) - (a_1, \ldots, a_{k-1}, y, a_{k+1}, \ldots, a_n)\|_\infty = \|h - y\|_{E_k} < \varepsilon.$$

Logo, $y \in H_{a,k}$. Isso prova que $B(h, \varepsilon) \subseteq H_{a,k}$, como queríamos demonstrar.

Se $f_{a,k}$ é diferenciável em a_k, definimos

$$\partial_k f(a) := f'_{a,k}(a_k) \in \mathcal{L}(E_k, F)$$

e chamamos essa aplicação a k-ésima diferencial parcial de f em a.

Proposição 1.7.1. Se $f : U \to F$ é diferenciável em $a \in U$, então todas as diferenciais parciais $\partial_k f(a)$ existem e

$$f'(a)h = \sum_{k=1}^{n} \partial_k f(a) h_k.$$

Demonstração. Defina $i_k(t) := (a_1, \ldots, a_{k-1}, t, a_{k+1}, \ldots, a_n)$. Pelo Teorema 1.3.3, i_k é diferenciável em a_k, valendo

$$i'_k(a_k)h_k = (0, \ldots, 0, h_k, 0, \ldots, 0).$$

Como f é diferenciável em $a = i_k(a_k)$ e i_k é diferenciável em a_k , então, pela Regra da Cadeia (ver Teorema 1.3.1), $f_{a,k} = f \circ i_k$ é diferenciável em a_k, com

$$\partial_k f(a)h_k = f'_{a,k}(a_k)h_k = f'(i_k(a_k))(i'_k(a_k)h_k) = f'(a)(0, \ldots, 0, h_k, 0, \ldots, 0).$$

Daí, pela linearidade de $f'(a)$, temos

$$f'(a)h = \sum_{k=1}^{n} \partial_k f(a) h_k,$$

como queríamos demonstrar. □

Proposição 1.7.2. Se f é de classe C^1, suas diferenciais parciais existem e são contínuas.

Demonstração. Observe que

$$\|(\partial_k f(a+x) - \partial_k f(a))h_k\| = \|(f'(a+x) - f'(a))(0, \ldots, h_k, \ldots, 0)\| \leq \|f'(a+x) - f'(a)\| \|h_k\|,$$

para todo $h_k \in E_k$. Logo,

$$\|\partial_k f(a + x) - \partial_k f(a)\|_{\mathcal{L}(E_k,F)} \leq \|f'(a + x) - f'(a)\|_{\mathcal{L}(E,F)}.$$

Como f é C^1, fazendo $x \to 0$ no lado direito a expressão inteira do lado direito $\to 0$ e o resultado segue-se. □

2 Teoremas do Valor Médio

Neste capítulo apresentaremos algumas das várias versões do Teorema do Valor Médio, em diferentes contextos, encontrando sua expressão mais geral com o **Teorema da Desigualdade do Valor Médio** (ver Corolário 2.3.3).

2.1 Teorema do Valor Médio de Lagrange

Nesta seção apresentaremos o Teorema do Valor Médio para funções contínuas $f : [a, b] \to \mathbb{R}$ deriváveis em (a, b).

O Teorema abaixo foi retirado de (COLEMAN, 2012, Exercise 2.6., p. 53).

Teorema 2.1.1 (Extensão do Teorema de Rolle)**.** Se E é um espaço normado, $X \subseteq E$ é um conjunto compacto com interior não vazio e $f : X \to \mathbb{R}$ é uma função contínua em X, diferenciável no interior de X e constante na fronteira de X, então f possui pelo menos um ponto estacionário no interior de X.

Demonstração. Pelo Teorema de Weierstrass (ver Corolário B.4.9), existem $x_0, y_0 \in f(X)$ tais que $x_0 = \min f(X)$ e $y_0 = \max f(X)$. Se $x_0 = y_0$, então f é uma função constante e, nesse caso, todos os pontos no interior de X (por hipótese, há pelo menos um ponto no interior de X) são pontos estacionários de f. No entanto, se $x_0 \neq y_0$, então existem $a, b \in X$ distintos tais que $f(a) = x_0$ e $f(b) = y_0$. Como f é constante em ∂X, se fosse $\{a, b\} \subseteq \partial X$, teríamos $x_0 = f(a) = f(b) = y_0$, o que é absurdo. Assim, existe $c \in \{a, b\}$ tal que $c \notin \partial X$. Como $c \in X$ e $c \notin \partial X$, então há uma vizinhança $V \subseteq E$ de c tal que $V \cap (E \setminus X) = \varnothing$, donde $V \subseteq X$ e, portanto, existe $\varepsilon > 0$ tal que $B_\varepsilon(c) \subseteq X$, isto é, $c \in \operatorname{int}(X)$. Logo, pelo Teorema 5.1.2, c é um ponto estacionário de f. $\square$

Como caso particular temos o (LIMA, 2019, Teorema 6, p. 270):

Corolário 2.1.2 (Teorema de Rolle)**.** Se $f : [a, b] \to \mathbb{R}$ é uma função contínua em $[a, b]$, derivável em (a, b) e $f(a) = f(b)$, então existe $c \in (a, b)$ tal que $\dot{f}(c) = 0$.

Uma das aplicações mais importantes do Teorema de Rolle (ver Corolário 2.1.2) é na demonstração do Teorema do Valor Médio (ver Corolário 2.1.3), Teorema que desempenha um papel central no Cálculo Diferencial e Integral.

Corolário 2.1.3 (Teorema do Valor Médio de Cauchy)**.** Se $f, g : [a, b] \to \mathbb{R}$ são contínuas em $[a, b]$ e deriváveis em (a, b), então existe $c \in (a, b)$ tal que

$$[f(a) - f(b)]\dot{g}(c) = [g(a) - g(b)]\dot{f}(c).$$

Demonstração. Defina $\psi : [a,b] \to \mathbb{R}$ pondo

$$\psi(x) := [f(a) - f(b)]g(x) - f(x)[g(a) - g(b)].$$

Como ψ é contínua em $[a,b]$, derivável em (a,b) e

$$\psi(a) = f(a)g(b) - f(b)g(a) = \psi(b),$$

então, pelo Teorema de Rolle (ver Corolário 2.1.2), existe $c \in (a,b)$ tal que $\dot{\psi}(c) = 0$, donde

$$[f(a) - f(b)]\dot{g}(c) = [g(a) - g(b)]\dot{f}(c),$$

como queríamos demonstrar. □

Como caso particular, temos o (LIMA, 2019, Teorema do Valor Médio, de Lagrange, p. 272):

Corolário 2.1.4 (Teorema do Valor Médio de Lagrange)**.** Se $f : [a,b] \to \mathbb{R}$ é contínua em $[a,b]$ e derivável em (a,b), então

$$\dot{f}(c) = \frac{f(b) - f(a)}{b - a}, \text{ para algum } c \in (a,b).$$

Demonstração. Considerando a função $g : [a,b] \to \mathbb{R}$ dada por $g(x) := x$, temos, pelo Teorema do Valor Médio de Cauchy (ver Corolário 2.1.3), que existe $c \in (a,b)$ tal que

$$f(a) - f(b) = [f(a) - f(b)]\dot{g}(c) = [g(a) - g(b)]\dot{f}(c) = (a - b)\dot{f}(c),$$

donde

$$\dot{f}(c) = \frac{f(a) - f(b)}{a - b} = \frac{f(b) - f(a)}{b - a},$$

como queríamos demonstrar. □

2.2 Identidade do Valor Médio

Nesta seção apresentaremos o Teorema do Valor Médio para funções diferenciáveis $f : U \to \mathbb{R}$, onde U é um subconjunto aberto de um espaço normado E contendo um segmento de reta $[a,b]$.

O Teorema abaixo, retirado de (COLEMAN, 2012, Theorem 3.2., p. 62), generaliza o Teorema do Valor Médio de Lagrange (ver Corolário 2.1.4).

Teorema 2.2.1 (Identidade do Valor Médio)**.** Sejam E um espaço normado, $U \subseteq E$ um subconjunto aberto e $[a,b] \subseteq U$. Se $f : U \to \mathbb{R}$ é uma função diferenciável, então

$$f(b) - f(a) = f'(\overline{x})(b - a), \text{ para algum } \overline{x} \in (a,b).$$

Demonstração. Sejam $\varphi : [0,1] \to E$ dada por

$$\varphi(t) := (1-t)a + tb$$

e

$$\psi := f \circ \varphi.$$

Como φ é contínua em $[0,1]$ e diferenciável em $(0,1)$, então ψ é contínua em $[0,1]$ e diferenciável em $(0,1)$, pela Regra da Cadeia (ver Teorema 1.3.1), donde, para $t \in (0,1)$, temos

$$\frac{d\psi}{dt}(t) = \psi'(t)1 = f'(\varphi(t))(\varphi'(t)1) = f'(\varphi(t))(b-a).$$

Pelo Teorema do Valor Médio de Lagrange (ver Corolário 2.1.4), existe $c \in (0,1)$ tal que

$$f'(\overline{x})(b-a) = f'(\varphi(c))\varphi'(c) = \psi'(c) = \frac{\psi(1)-\psi(0)}{1-0} = f(b) - f(a),$$

onde $\overline{x} := \varphi(c)$. □

As definições abaixo, extraídas de (COLEMAN, 2012, p. 63), generalizam a definição de caminhos nos espaços euclidianos $\mathbb{R}^n$.

Definição 2.2.2 (Caminhos)**.** Sejam E um espaço normado e $S \subseteq E$. Um **caminho** em S é uma função contínua $\alpha : I \to S$, onde I é um intervalo compacto de números reais.

Se $x, y \in S$, diremos que um caminho $\alpha : [a,b] \to S$ **liga** x a y quando $\alpha(a) = x$ e $\alpha(b) = y$. O **caminho retilíneo** que liga x a y é o caminho $\beta : [a,b] \to S$ dada por

$$\beta(t) = x + \frac{y-x}{b-a}(t-a).$$

Nesse caso, $\beta([a,b])$ é o segmento de reta $[\beta(a), \beta(b)] = [x,y]$.

Dizemos que um caminho $\alpha : [a,b] \to S$ é um **caminho poligonal** se existe uma partição $P := \{a = t_0 < \ldots < t_p = b\}$ de $[a,b]$ tal que $\alpha|_{[t_{i-1},t_i]}$ é um caminho retilíneo que liga $\alpha(t_{i-1})$ a $\alpha(t_i)$, para cada $i = 1, \ldots, p$.

Veja que, dado um intervalo compacto $[a,b]$, existe um único homeomorfismo não decrescente $[0,1] \to [a,b]$, o que nos permite ver naturalmente qualquer caminho $[a,b] \to S$ como um caminho $[0,1] \to S$.

Caminhos poligonais são particularmente interessantes pelo fato de que são formados por segmentos de reta e, portanto, podemos aplicar a Identidade do Valor Médio (ver Teorema 2.2.1) em cada um desses segmentos de reta. Uma observação trivial é de, dado um espaço normado E e um subconjunto $U \subseteq E$, vale a seguinte cadeia de implicações

$$U \text{ é conexo por caminhos poligonais} \implies U \text{ é conexo por caminhos} \implies U \text{ é conexo},$$

mas uma observação muito mais interessante (e menos trivial) é a de que, se U é aberto, essa cadeia de implicações é reversível.

Teorema 2.2.3. Sejam E um espaço normado e $U \subseteq E$ um subconjunto aberto. São equivalentes:

1. U é conexo;

2. U é conexo por caminhos poligonais;

3. U é conexo por caminhos.

Demonstração. Que 2) $\implies$ 3) é óbvio e 3) $\implies$ 1) constitui o Corolário B.3.21. Resta apenas mostrar 1) $\implies$ 2). Considere em U a relação de equivalência $\sim$ dada por

$$x \sim y \iff \text{existe um caminho poligonal ligando } x \text{ a } y.$$

Para cada $a \in U$, considere $[a]$ a classe de equivalência de a por $\sim$. Fixado $a \in U$, temos $[a] \subseteq U$ e, como U é aberto, se $x \in [a]$, então $x \in U$, donde existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \subseteq U$. Dado $y \in B(x, \varepsilon)$, o caminho

$$\gamma : t \in [0, 1] \mapsto (1 - t)x + ty \in E$$

é tal que $\gamma([0, 1]) \subseteq B(x, \varepsilon)$ e γ liga x a y. Como $x \sim a$ e $y \sim x$, então $y \sim a$, pela transitividade de $\sim$, donde $y \in [a]$ e, portanto, $B(x, \varepsilon) \subseteq [a]$ e, logo, $[a]$ é um conjunto aberto. Como

$$U/\!\sim\, := \{[x] : x \in U\}$$

é uma partição de U, se houvesse outra classe de equivalência em X além de $[a]$, então

$$U = [a] \cup \bigcup_{x \in U \setminus \{a\}} [x]$$

seria união de dois abertos disjuntos não vazios e, portanto, seria desconexo. Logo,

$$U = [a].$$

Isso significa que, para quaisquer $x, y \in U$, temos $x \sim a$ e $y \sim a$, o que nos dá $x \sim y$, pela simetria e transitividade de $\sim$. Logo, existe um caminho poligonal ligando quaisquer dois pontos de U. $\square$

Corolário 2.2.4. (COLEMAN, 2012, Theorem 3.3., p. 63) Se E é um espaço normado, $U \subseteq E$ é um subconjunto aberto conexo e $f : U \to \mathbb{R}$ é uma função diferenciável tal que $f'(a) = 0_{E^*}$, para todo $a \in U$, então f é constante.

Demonstração. Fixe $a \in U$ e considere $x \in U$ outro elemento. Pelo Teorema 2.2.3, existe um caminho poligonal γ conectando a a x, isto é, existem uma partição $P := \{0 = t_0 < t_1 < \ldots < t_p = 1\}$ de $[0,1]$ e um caminho $\gamma : [0,1] \to U$ tal que $\gamma|_{[t_{i-1},t_i]}$ é um caminho retilíneo ligando $\gamma(t_{i-1})$ a $\gamma(t_i)$, para cada $i = 1, \ldots, p$, com $\gamma(0) = a$ e $\gamma(1) = x$. Pela Identidade do Valor Médio (ver Teorema 2.2.1), dado $i = 1, \ldots, p$, existe $c_i \in (\gamma(t_{i-1}), \gamma(t_i))$ tal que

$$f(\gamma(t_i)) - f(\gamma(t_{i-1})) = f'(c_i)(\gamma(t_i) - \gamma(t_{i-1})) = 0.$$

Logo,

$$f(\gamma(t_i)) = f(\gamma(t_{i-1})), \text{ para cada } i = 1, \ldots, p,$$

e, portanto,

$$f(x) = f(\gamma(1)) = f(\gamma(0)) = f(a),$$

como queríamos demonstrar. □

2.3 Desigualdade do Valor Médio

Nesta seção generalizaremos a Identidade do Valor Médio (ver Teorema 2.2.1).

Teorema 2.3.1. (COLEMAN, 2012, Theorem 3.4., p. 64) Sejam $[a,b] \subseteq \mathbb{R}$ um intervalo de números reais e F um espaço normado. Se $f : [a,b] \to F$ e $g : [a,b] \to \mathbb{R}$ são aplicações contínuas em $[a,b]$ e diferenciáveis em (a,b), então

$$\left\|\dot{f}(t)\right\|_F \leq \dot{g}(t), \text{ para todo } t \in (a,b) \implies \|f(b) - f(a)\|_F \leq g(b) - g(a).$$

Demonstração. Sejam $\varepsilon > 0$ e definamos $A(\varepsilon)$ como o conjunto de todos os $x \in [a,b]$ tais que para cada $y \in [a,x]$ se tem

$$\|f(y) - f(a)\|_F \leq g(y) - g(a) + \varepsilon(y - a).$$

Claramente, $a \in A(\varepsilon)$. Além disso, se $x \in A(\varepsilon)$ e $\tilde{x} \in [a,x]$, então $\tilde{x} \in A(\varepsilon)$, donde $A(\varepsilon)$ é um intervalo. Seja $c := \sup A(\varepsilon)$. Afirmamos que $c \in A(\varepsilon)$. Não há o que provar se $c = a$, então suponhamos que este não é o caso. Se $x \in (a,c)$, então

$$\|f(x) - f(a)\|_F \leq g(x) - g(a) + \varepsilon(x - a).$$

Como f e g são contínuas, então, pela Monotonicidade do Limite (ver Teorema A.1.11), a desigualdade se aplica para $x = c$ também, donde $c \in A(\varepsilon)$.

Mostraremos agora que $c = b$. Suponhamos, por redução ao absurdo, que $c < b$. Pela hipótese sobre as derivadas e pelo Teorema B.5.9, existe $\delta > 0$ tal que

$$x \in [(c-\delta, c+\delta) \setminus \{c\}] \cap [a,b] \implies \left\|\frac{f(x) - f(c)}{x - c}\right\|_F \leq \frac{g(x) - g(c)}{x - c} + \varepsilon.$$

Como $b - c > 0$ e $\mathbb{R}$ é um corpo arquimediano, então existe $n_0 \in \mathbb{N}$ tal que $n_0 \geq \dfrac{\delta}{b-c}$. Pondo $\eta := \dfrac{\delta}{n_0}$, temos $\eta \in (0, b-c)$, donde

$$x \in (c, c+\eta) \implies \|f(x) - f(c)\|_F \leq g(x) - g(c) + \varepsilon(x - c).$$

Agora, pela Desigualdade Triangular, temos

$$\begin{aligned}\|f(x) - f(a)\|_F &\leq \|f(x) - f(c)\|_F + \|f(c) - f(a)\|_F \\ &\leq [g(x) - g(c) + \varepsilon(x-c)] + [g(c) - g(a) + \varepsilon(c-a)] \\ &\leq g(x) - g(a) + \varepsilon(x-a),\end{aligned}$$

o que implica que $(c, c+\eta) \subseteq A(\varepsilon)$, contradizendo o fato de que $c = \sup A(\varepsilon)$. Logo, $c = b$ e, portanto,

$$\|f(b) - f(a)\|_F \leq g(b) - g(a) + \varepsilon(b-a) \xrightarrow{\varepsilon \to 0} g(b) - g(a).$$

O resultado se segue da Monotonicidade do Limite (ver Teorema B.5.8). □

Corolário 2.3.2. (COLEMAN, 2012, Corollary 3.1., p. 64) Sejam $[a,b]$ um intervalo não degenerado de números reais, F um espaço normado e $f : [a,b] \to F$ uma aplicação contínua em $[a,b]$ e diferenciável em (a,b). Se existe $K > 0$ tal que

$$\|\dot{f}(t)\|_F \leq K, \text{ para todo } t \in (a,b),$$

então

$$\|f(b) - f(a)\|_F \leq K(b-a).$$

Demonstração. Aplique $g(t) := Kt$ no Teorema anterior. □

Por fim, enunciamos a Desigualdade do Valor Médio, de acordo com (COLEMAN, 2012, Corollary 3.2., p. 65).

Corolário 2.3.3 (Desigualdade do Valor Médio)**.** Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ diferenciável. Se o segmento $[a,b] \subseteq U$, então

$$\|f(b) - f(a)\|_F \leq \sup_{x \in (a,b)} \|f'(x)\|_{\mathcal{L}(E,F)} \|b - a\|_E.$$

Demonstração. Seja $u : [0,1] \to E$ dada por $u(t) := (1-t)a + tb$. Como $g := f \circ u$ é contínua em $[0,1]$ e diferenciável em $(0,1)$, então

$$\dot{g}(t) = g'(t)1 = f'(u(t))(u'(t)1) = f'(u(t))(b-a),$$

donde

$$\|\dot{g}(t)\|_F \leq \sup_{x \in (a,b)} \|f'(x)\|_{\mathcal{L}(E,F)} \|b - a\|_E$$

e, portanto, pelo Corolário 2.3.2,

$$\|g(1) - g(0)\|_F \leq \sup_{x \in (a,b)} \|f'(x)\|_{\mathcal{L}(E,F)} \|b - a\|_E (1 - 0).$$

Isto é,

$$\|f(b) - f(a)\|_F \leq \sup_{x \in (a,b)} \|f'(x)\|_{\mathcal{L}(E,F)} \|b - a\|_E,$$

como queríamos demonstrar. $\square$

3 Diferenciabilidade de ordem superior

Neste capítulo construiremos as derivadas de Fréchet de ordem k, com $k \geq 2$, estabeleceremos uma relação importante entre a derivada de segunda ordem e o operador hessiano (ver Exemplo 3.3.2) e provaremos que, enquanto aplicações multilineares, as derivadas são sempre simétricas (ver Teorema 3.6.3).

3.1 A derivada de segunda ordem

Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ uma aplicação diferenciável. Nessas condições, a **derivada de Fréchet** de f é a aplicação

$$\begin{aligned} f'(\cdot) : U &\to \mathcal{L}(E, F) \\ a &\longmapsto f'(a). \end{aligned}$$

Por definição, dizer que $f'(\cdot)$ é uma aplicação diferenciável em um ponto $a \in U$ significa que existe uma (única) aplicação linear e contínua $E \to \mathcal{L}(E, F)$, que denotaremos por $f^{[2]}(a)$, tal que

$$\frac{f'(a+h) - f'(a) - f^{[2]}(a)h}{\|h\|_E} \xrightarrow{h \to 0_E} 0_{\mathcal{L}(E,F)}.$$

Podemos estabelecer uma isometria linear de modo natural entre os espaços $\mathcal{L}(E, \mathcal{L}(E, F))$ e $\mathcal{L}_2(E, F)$ através da aplicação

$$Q : \mathcal{L}(E, \mathcal{L}(E, F)) \to \mathcal{L}_2(E, F),$$

dada por

$$Q(T)(u, v) := T(v)(u),$$

para quaisquer $u, v \in E$. Com efeito, $Q = J_2$ no Teorema 3.2.2.

A **derivada de Fréchet de segunda ordem** de f em a é a aplicação $Q\left(f^{[2]}(a)\right)$, que denotaremos por $f^{(2)}(a)$. Nessas condições, é imediato que, para quaisquer $u, v \in E$,

$$f^{(2)}(a)(u, v) = f^{[2]}(a)(v)(u) \quad \text{e} \quad \|f^{(2)}(a)\|_{\mathcal{L}_2(E,F)} = \|f^{[2]}(a)\|_{\mathcal{L}(E,\mathcal{L}(E,F))}.$$

3.2 Botando ordem na casa: construindo uma isometria linear

Se E e F são espaços normados e existe uma isometria linear $E \to F$, escreveremos $E \approx F$. Note que essa relação é de equivalência.

Proposição 3.2.1. Sejam E, V e W espaços normados. Então,

$$V \approx W \implies \mathcal{L}_p(E, V) \approx \mathcal{L}_p(E, W).$$

Demonstração. Suponhamos que existe uma isometria linear $\psi : V \to W$.

Note que, para quaisquer $T \in \mathcal{L}_p(E, V)$, $\lambda \in \mathbb{R}$ e $u, v \in E$, temos

$$\begin{aligned}(\psi \circ T)(\ldots, \underbrace{u + \lambda v}_{i\text{-ésima posição}}, \ldots) &= \psi(T(\ldots, u, \ldots) + \lambda T(\ldots, v, \ldots)) \\ &= (\psi \circ T)(\ldots, u, \ldots) + \lambda(\psi \circ T)(\ldots, v, \ldots).\end{aligned}$$

Logo,

$$\psi \circ T \in \mathcal{L}_p(E, W), \quad \text{para cada } T \in \mathcal{L}_p(E, V).$$

Afirmamos que a aplicação

$$\Omega : T \in \mathcal{L}_p(E, V) \mapsto \psi \circ T \in \mathcal{L}_p(E, W)$$

é uma isometria linear. Com efeito, dados $S, T \in \mathcal{L}_p(E, V)$ e $\lambda \in \mathbb{R}$, temos

$$\begin{aligned}\Omega(S + \lambda T)(u) = (\psi \circ (S + \lambda T))(u) &= \psi((S + \lambda T)(u)) \\ &= \psi(S(u) + \lambda T(u)) \\ &= \psi(S(u)) + \lambda\psi(T(u)) \\ &= (\psi \circ S)(u) + \lambda(\psi \circ T)(u) \\ &= (\psi \circ S + \lambda(\psi \circ T))(u) \\ &= (\Omega(S) + \lambda\Omega(T))(u),\end{aligned}$$

para cada $u \in E^p$. Logo, para quaisquer $S, T \in \mathcal{L}_p(E, V)$ e $\lambda \in \mathbb{R}$,

$$\Omega(S + \lambda T) = \Omega(S) + \lambda\Omega(T),$$

isto é, Ω é linear. Além disso, Ω é bijetiva, sendo sua inversa a aplicação

$$S \in \mathcal{L}_p(E, W) \mapsto \psi^{-1} \circ S \in \mathcal{L}_p(E, V)$$

Por fim, como ψ é uma isometria, para quaisquer $u_1, \ldots, u_p \in E$, temos

$$\|(\psi \circ T)(u_1, \ldots, u_p)\|_W = \|\psi(T(u_1, \ldots, u_p))\|_W = \|T(u_1, \ldots, u_p)\|_V,$$

donde

$$\begin{aligned}\|\Omega(T)\|_{\mathcal{L}_p(E,W)} = \|\psi \circ T\|_{\mathcal{L}_p(E,W)} &= \sup_{\substack{\|u_i\|_E = 1, \\ \text{para todo } i \in [p]}} \|(\psi \circ T)(u_1, \ldots, u_p)\|_W \\ &= \sup_{\substack{\|u_i\|_E = 1, \\ \text{para todo } i \in [p]}} \|T(u_1, \ldots, u_p)\|_V \\ &= \|T\|_{\mathcal{L}_p(E,V)}.\end{aligned}$$

Logo, Ω é uma imersão isométrica e, sendo bijetiva, é uma isometria. $\square$

Considere, para cada $n \in \mathbb{N}$,

$$\begin{cases} M_1(E,F) := \mathcal{L}(E,F) \\ M_{n+1}(E,F) := \mathcal{L}(E, M_n(E,F)) \end{cases} \qquad \begin{cases} J_1 := \mathrm{Id}_{\mathcal{L}(E,F)} \\ J_{n+1} : M_{n+1}(E,F) \to \mathcal{L}_{n+1}(E,F) \end{cases}$$

e

$$\begin{cases} H_1 := \mathrm{Id}_{\mathcal{L}(E,F)} \\ H_{n+1} : \mathcal{L}_{n+1}(E,F) \to M_{n+1}(E,F), \end{cases}$$

onde, para cada $T \in M_{n+1}(E,F)$,

$$\begin{aligned} J_{n+1}(T) : E^n \times E &\to F \\ (\overline{x}, y) &\longmapsto J_n(Ty)\overline{x} \end{aligned}$$

e, para cada $S \in \mathcal{L}_{n+1}(E,F)$,

$$\begin{aligned} H_{n+1}(S) : E &\to M_n(E,F) \\ y &\longmapsto J_n^{-1} S(\cdot, y). \end{aligned}$$

Teorema 3.2.2. Cada J_n está bem definido e é uma isometria linear.

Demonstração. Evidentemente, J_1 está bem definido e é uma isometria linear. Suponhamos, por hipótese de indução, que J_k está bem definido e é uma isometria linear.

Dado $T \in M_{k+1}(E,F)$, começamos verificando que $J_{k+1}(T)$ é uma aplicação $(k+1)$-linear contínua. Dados $i \leq j$ naturais e $x_i, x_{i+1}, \ldots, x_j \in E$, defina

$$x_{\geq i}^{\leq j} := (x_i, \ldots, x_j).$$

Para cada $i = 1, \ldots, k$, a aplicação

$$u_i \in E \mapsto J_{k+1}(T)\left(x_{\geq 1}^{\leq i-1}, u_i, x_{\geq i+1}^{\leq k+1}\right) = J_n(Tx_{k+1})\left(x_{\geq 1}^{\leq i-1}, u_i, x_{\geq i+1}^{\leq k}\right)$$

é linear. Isso prova que $J_{k+1}(T)$ é linear em suas k primeiras entradas, resta verificar que é linear na $(k+1)$-ésima. Com efeito, fixe $\overline{x} \in E^k$ e $u, v \in E$ e $\lambda \in \mathbb{R}$. Daí,

$$\begin{aligned} J_{k+1}(T)(\overline{x}, u + \lambda v) = J_k(T(u + \lambda v))\overline{x} &= J_k(Tu + \lambda Tv)\overline{x} \\ &= J_k(Tu)\overline{x} + \lambda J_k(Tv)\overline{x} \\ &= J_{k+1}(T)(\overline{x}, u) + \lambda J_{k+1}(T)(\overline{x}, v). \end{aligned}$$

Note que a hipótese de boa definição de J_k está sendo utilizada em todas as identidades, enquanto a linearidade de J_k é utilizada na penúltima. Isto completa a prova de que $J_{k+1}(T)$ é $(k+1)$-linear.

Vamos verificar agora que $J_{k+1}(T)$ é contínua. Dados $u_1, \ldots, u_{k+1} \in E$, temos

$$\begin{aligned}
\|J_{k+1}(T)(u_1, \ldots, u_{k+1})\|_F &= \|J_k(Tu_{k+1})(u_1, \ldots, u_k)\|_F \\
&\leq \|J_k(Tu_{k+1})\|_{\mathcal{L}_k(E,F)} \prod_{i\in[k]} \|u_i\|_E \\
&= \|Tu_{k+1}\|_{M_k(E,F)} \prod_{i\in[k]} \|u_i\|_E \\
&\leq \|T\|_{M_{k+1}(E,F)} \prod_{i\in[k+1]} \|u_i\|_E.
\end{aligned}$$

Veja que a segunda identidade é válida porque estamos sob a hipótese de que J_k é uma isometria. Do Teorema C.7.1, temos que $J_{k+1}(T)$ é contínua.

Com isso finalizamos a prova da boa definição de J_{k+1}.

Provaremos agora que J_{k+1} é linear. Dados $T_1, T_2 \in M_{k+1}(E, F)$ e $\lambda \in \mathbb{R}$, temos

$$\begin{aligned}
J_{k+1}(T_1 + \lambda T_2)(\overline{x}, y) &= J_k[(T_1 + \lambda T_2)y]\overline{x} \\
&= J_k(T_1 y + \lambda T_2 y)\overline{x} \\
&= [J_k(T_1 y) + \lambda J_k(T_2 y)]\,\overline{x} \\
&= J_k(T_1 y)\overline{x} + \lambda J_k(T_2 y)\overline{x} \\
&= J_{k+1}(T_1)(\overline{x}, y) + \lambda J_{k+1}(T_2)(\overline{x}, y) \\
&= [J_{k+1}(T_1) + \lambda J_{k+1}(T_2)](\overline{x}, y).
\end{aligned}$$

Isto completa a prova da linearidade de J_{k+1}.

Verificaremos agora que J_{k+1} e H_{k+1} são aplicações inversas uma da outra. Para quaisquer $\overline{x} \in E^k, y \in E$ e $T \in \mathcal{L}_{k+1}(E, F)$, temos

$$J_{k+1}(H_{k+1}(T))(\overline{x}, y) = J_k(H_{k+1}(T)y)\overline{x} = J_k(J_k^{-1}T(\cdot, y))\overline{x} = T(\overline{x}, y).$$

Logo, para todo $T \in \mathcal{L}_{k+1}(E, F)$, temos

$$J_{k+1}(H_{k+1}(T)) = T$$

e, portanto,

$$J_{k+1} \circ H_{k+1} = \mathrm{Id}_{\mathcal{L}_{k+1}(E,F)}\,.$$

De modo análogo se verifica que

$$H_{k+1} \circ J_{k+1} = \mathrm{Id}_{M_{k+1}(E,F)}\,.$$

Logo, J_{k+1} é uma aplicação bijetiva, com

$$H_{k+1} = J_{k+1}^{-1}.$$

Resta apenas verificar que J_{k+1} é uma imersão isométrica. Dado $T \in \mathcal{L}_{k+1}(E,F)$, temos, pelo Teorema A.2.8, que

$$\begin{aligned}
\|J_{k+1}(T)\|_{\mathcal{L}_{k+1}(E,F)} &= \sup_{\substack{\|x_i\|_E=1,\\ \text{para todo } i\in[k+1]}} \|J_{k+1}(T)(x_1,\ldots,x_{k+1})\|_F \\
&= \sup_{\|x_{k+1}\|_E=1} \sup_{\substack{\|x_i\|_E=1,\\ \text{para todo } i\in[k]}} \|J_k(Tx_{k+1})(x_1,\ldots,x_k)\|_F \\
&= \sup_{\|x_{k+1}\|_E=1} \|J_k(Tx_{k+1})\|_{\mathcal{L}_k(E,F)} \\
&= \sup_{\|x_{k+1}\|_E=1} \|Tx_{k+1}\|_{M_k(E,F)} \\
&= \|T\|_{M_{k+1}(E,F)}.
\end{aligned}$$

Logo, J_{k+1} é uma isometria.

O resultado segue do Princípio de Indução de Matemática. $\square$

Note que, para quaisquer $m, n \in \mathbb{N}$, temos

$$M_m(E, M_n(E,F)) = M_{m+n}(E,F).$$

Isso pode ser provado como um simples exercício de indução.

Corolário 3.2.3. Dados espaços normados E e F,

$$\mathcal{L}_m(E, \mathcal{L}_n(E,F)) \approx \mathcal{L}_{m+n}(E,F), \text{ para quaisquer } m, n \in \mathbb{N},$$

Demonstração. Fixados m, n naturais, como

$$\mathcal{L}_m(E, \mathcal{L}_n(E,F)) \approx \mathcal{L}_m(E, M_n(E,F)),$$

então

$$\mathcal{L}_m(E, M_n(E,F)) \approx M_m(E, M_n(E,F)) = M_{m+n}(E,F) \approx \mathcal{L}_{m+n}(E,F),$$

como queríamos demonstrar. $\square$

3.3 Classes de diferenciabilidade

Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ uma aplicação tal que $f^{[k]} : U \to M_k(E,F)$ existe e é diferenciável em $a \in U$. Então $f^{[k+1]}(a) := f^{[k]\prime}(a) \in M_{k+1}(E,F)$. A **derivada de Fréchet de ordem** $k+1$ de f em a (ou a $(k+1)$**-ésima derivada de Fréchet** de f em a) é

$$f^{(k+1)}(a) := J_{k+1}\left(f^{[k+1]}(a)\right) \in \mathcal{L}_{k+1}(E,F).$$

Definição 3.3.1 (Classes de diferenciabilidade)**.** Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ uma aplicação.

- Quando f for contínua, diremos que f é de classe C^0.
- Quando f for diferenciável e sua diferencial $f' : U \to \mathcal{L}(E, F)$ for contínua, diremos que f é de **classe** C^1 (ou que f é **continuamente diferenciável**).
- Se $f^{[k]}$ for diferenciável em $a \in U$, diremos que f é $k+1$ **vezes diferenciável** em a. Se $f^{(k+1)} : U \to \mathcal{L}_{k+1}(E, F)$ existe e é contínua, dizemos que f é de **classe** C^{k+1} (ou que f é $k+1$ **vezes continuamente diferenciável**).
- Se $f^{[k]}$ for diferenciável em $a \in U$, para todo $k \in \mathbb{N}$, dizemos que f é de **classe** C^∞ (ou que f é **infinitamente diferenciável**).

Se E e F são espaços normados e $U \subseteq E$ é um subconjunto aberto, o conjunto de todas as aplicações $f^{(k)} : U \to \mathcal{L}_k(E, F)$ será denotado por $C^k(U, F)$, para cada $k \in \mathbb{N} \cup \{0\}$, onde $f^{(0)} := f$ e $\mathcal{L}_0(E, F) := F$.

Para o próximo exemplo, relembre o Exemplo 1.6.1.

Exemplo 3.3.2 (Relação entre o operador hessiano e a diferencial de 2ª ordem)**.** Sejam H um espaço de Hilbert, $U \subseteq H$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função diferenciável cuja aplicação gradiente $\nabla f : U \to H$ é diferenciável em $a \in U$. Nessas condições, f é duas vezes diferenciável em a e

$$f^{(2)}(a)(x, y) = \langle x, \nabla^2 f(a) y\rangle, \quad \text{para quaisquer} \quad x, y \in H.$$

Demonstração. Defina $R_y := \langle \cdot, y\rangle$, para todo $y \in H$. Como o produto interno é uma forma bilinear, então R_y é linear. Pela Desigualdade de Cauchy-Schwarz (ver Teorema C.5.2), temos

$$|R_y(x)| = |\langle x, y\rangle| \leq \|x\|\|y\|, \quad \text{para todo} \quad x \in H.$$

Logo, cada R_y é um funcional linear contínuo. Pelo Teorema da Representação de Riesz (ver Teorema C.9.4), $\|R_y\| = \|y\|$, para todo $y \in H$. Logo a transformação linear $R : H \to H^*$ dada por $R(y) := R_y$ é uma isometria. Em particular, por ser uma isometria, R é contínua. Sendo linear e contínua, R é diferenciável, pelo Exemplo 1.1.14. Como

$$f' = R \circ \nabla f$$

e tanto ∇f quanto R são diferenciáveis em a, então, pela Regra da Cadeia (ver Teorema 1.3.1), f' é diferenciável em a e

$$f^{[2]}(a)y = R'(\nabla f(a))((\nabla f)'(a)y) = R(\nabla^2 f(a)y) = R_{\nabla^2 f(a)y}.$$

Logo,

$$f^{(2)}(a)(x,y) = f^{[2]}(a)y(x) = R_{\nabla^2(f(a))y}(x) = \langle x, \nabla^2 f(a)y\rangle,$$

como queríamos demonstrar. $\square$

Teorema 3.3.3. (COLEMAN, 2012, Proposition 4.5., p. 89) Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto, $f : U \to F$ uma aplicação. Então f é $k+1$ vezes diferenciável em $a \in U$ se, e somente se, $f^{(k)}$ é diferenciável em a. Nesse caso, para quaisquer $h, h_1, \ldots, h_k \in E$,

$$f^{(k)'}(a)h(h_1,\ldots,h_k) = f^{(k+1)}(a)(h_1,\ldots,h_k,h).$$

Demonstração. Suponhamos primeiro que $f^{(k)}$ é diferenciável em a. Como

$$f^{(k)} = J_k \circ f^{[k]},$$

então

$$f^{[k]} = H_k \circ f^{(k)}.$$

Como H_k é uma isometria linear, então H_k é contínuo e, pelo Exemplo 1.1.14, H_k é diferenciável. Pela Regra da Cadeia (ver Teorema 1.3.1), $f^{[k]}$ é diferenciável em a, o que, por definição, quer dizer que f é $k+1$ vezes diferenciável.

Reciprocamente, se f é $k+1$ vezes diferenciável em a, então $f^{[k]}$ é diferenciável em a. Logo, $f^{(k)} = J_k \circ f^{[k]}$ também é diferenciável em a, pela Regra da Cadeia.

Agora, como

$$f^{(k)'}(a)h = J_k'(f^{[k]}(a))(f^{[k+1]}(a)h) = J_k(f^{[k+1]}(a)h),$$

então

$$\begin{aligned} f^{(k)'}(a)h(h_1,\ldots,h_k) &= J_k(f^{[k+1]}(a)h)(h_1,\ldots,h_k) = J_{k+1}(f^{[k+1]}(a))(h_1,\ldots,h_k,h) \\ &= f^{(k+1)}(a)(h_1,\ldots,h_k,h), \end{aligned}$$

como queríamos demonstrar. $\square$

O próximo teorema nos fornece uma técnica para o cálculo das derivadas de ordem superior.

Teorema 3.3.4 (Técnica para calcular diferenciais de ordem superior)**.** Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto, $f : U \to F$ uma aplicação $k+1$ vezes diferenciável em a e considere a aplicação

$$q : x \in U \mapsto f^{(k)}(x)(h_1,\ldots,h_k) \in F,$$

onde $h_1,\ldots,h_k \in E$ são fixados. Então q é diferenciável e, para todo $h_{k+1} \in E$,

$$q'(a)h_{k+1} = f^{(k+1)}(a)(h_1,\ldots,h_{k+1}).$$

Demonstração. Considere a transformação linear

$$\alpha : T \in \mathcal{L}_k(E, F) \mapsto T(h_1, \ldots, h_k) \in F.$$

Como, para cada $T \in \mathcal{L}_k(E, F)$, temos

$$\|\alpha(T)\|_F = \|T(h_1, \ldots, h_k)\|_F \leq \|T\|_{\mathcal{L}_k(E,F)} \prod_{i=1}^{k} \|h_i\|_E,$$

então α é contínua, pelo Teorema C.4.1. Daí, pelo Exemplo 1.1.14, temos

$$\alpha'(a)h = \alpha(h), \quad \text{para todo} \quad h \in E.$$

Como $f^{(k)}$ é diferenciável em a pelo Teorema 3.3.3 e

$$q = \alpha \circ f^{(k)},$$

então, pela Regra da Cadeia (ver Teorema 1.3.1),

$$q'(a)h_{k+1} = \alpha'(f^{(k)}(a))(f^{(k)'}(a)h_{k+1}) = \alpha(f^{(k)'}(a)h_{k+1}) = f^{(k)'}(a)h_{k+1}(h_1, \ldots, h_k).$$

Aplicando novamente o Teorema 3.3.3, temos

$$q'(a)h_{k+1} = f^{(k+1)}(a)(h_1, \ldots, h_{k+1}),$$

o que termina a prova. □

Exemplo 3.3.5. A função $f : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f(x) := \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c,$$

onde A é uma matriz de ordem n simétrica, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$, é de classe C^∞ e vale

$$\begin{aligned} f'(x)h_1 &= \langle Ax + b, h_1\rangle \\ f^{(2)}(x)h_1h_2 &= \langle Ah_2, h_1\rangle \\ f^{(k)}(x)h_1 \ldots h_k &= 0, \quad \text{para todo } k > 2. \end{aligned}$$

Demonstração. Note que

$$f'(x)h_1 = \frac{1}{2}\left(\langle h_1, Ax\rangle + \langle x, Ah_1\rangle\right) + \langle b, h_1\rangle = \langle Ax, h_1\rangle + \langle b, h_1\rangle = \langle Ax + b, h_1\rangle.$$

A fim de calcular $f^{(2)}(x)h_1h_2$, consideremos a função auxiliar $g_1 : \mathbb{R}^n \to \mathbb{R}$ dada por $g_1(x) := f'(x)h_1$, onde $h_1 \in \mathbb{R}^n$ é fixado. Assim,

$$f^{(2)}(x)h_1h_2 = g_1'(x)h_2 = \langle Ah_2, h_1\rangle.$$

Pondo, agora, $g_2(x) := f^{(2)}(x)h_1h_2$, onde $h_2 \in \mathbb{R}^n$ é fixado, temos

$$f^{(3)}(x)h_1h_2h_3 = g_2'(x)h_3 = 0.$$

Logo,

$$f^{(k)}(x)h_1 \ldots h_k = 0,$$

para todo $k > 2$. □

3.4 O polinômio homogêneo

O exemplo abaixo, embora interessante em si mesmo, será extremamente útil para o cálculo de derivadas de ordem superior. Nós o utilizaremos, por exemplo, para demonstrar a **fórmula de Taylor com resto infinitesimal** (ver Teorema 3.7.2).

Sejam E e F espaços vetoriais, $\varphi : E^n \to F$ uma aplicação n-linear simétrica e

$$P : x \in E \mapsto \varphi \cdot x^{(n)} \in F.$$

Nessas condições,

$$P(a+th) = \sum_{k=0}^{n} \binom{n}{k} t^k \varphi\left(a^{(n-k)}, h^{(k)}\right), \quad \text{para quaisquer} \quad a, h \in E, t \in \mathbb{R}, \tag{3.1}$$

e

$$P(x) = 0_F, \quad \text{para todo} \quad x \in E \implies \varphi = 0_{\mathcal{L}_n(E,F)}. \tag{3.2}$$

Além disso, se E e F são espaços normados e φ é contínua, então a aplicação P é de classe C^∞, valendo, para quaisquer $x, h_1, \ldots, h_j \in E$,

$$P^{(j)}(x)h_1 \ldots h_j = \begin{cases} \binom{n}{j} j! \cdot \varphi\left(x^{(n-j)}, h_1, \ldots, h_j\right), & \text{se} \quad j = 1, \ldots, n, \\ 0_F, \quad \text{se} \quad j > n. \end{cases} \tag{3.3}$$

Demonstração. Sejam $a, h \in E$ e $n \in \mathbb{N}$. Do Teorema 1.4.1, temos

$$\varphi \cdot (a+th)^{(n)} = \sum_{0 \le k \le n} \sum_{\substack{b_i \in \{a, th\}, \text{ para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = th\}| = k}} \varphi(b_1, \ldots, b_n).$$

Vamos olhar mais de perto o somatório

$$\sum_{\substack{b_i \in \{a, th\}, \text{ para todo } i \le n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = th\}| = k}} \varphi(b_1, \ldots, b_n).$$

Como há $\binom{n}{k}$ maneiras de escolher a n-upla $(b_1, \ldots, b_n)$, onde cada $b_i \in \{a, th\}$ e $b_j = th$ para k índices j, então há $\binom{n}{k}$ termos $\varphi(b_1, \ldots, b_n)$ a serem somados. Além disso, fixada uma tal n-upla, como φ é uma aplicação multilinear simétrica, temos

$$\varphi(b_1, \ldots, b_n) = \varphi\left((th)^{(k)}, a^{(n-k)}\right) = t^k \varphi\left(h^{(k)}, a^{(n-k)}\right).$$

Nessas condições,

$$\sum_{\substack{b_i \in \{a, th\}, \text{ para todo } i=1,\ldots,n, \\ \text{e } |\{i=1,\ldots,n \,:\, b_i = th\}| = k}} \varphi(b_1, \ldots, b_n) = \binom{n}{k} t^k \varphi\left(h^{(k)}, a^{(n-k)}\right),$$

o que nos permite escrever

$$\varphi \cdot (a+th)^{(n)} = \sum_{0\leq k\leq n} \binom{n}{k} t^k \varphi\left(h^{(k)}, a^{(n-k)}\right).$$

Isto prova a validade de (3.1).

Vamos agora provar a implicação (3.2). Suponhamos que $P(x) = 0_F$, para cada $x \in E$, e $n \geq 1$. Pela identidade (3.1), dados $a, h \in E$ e $t \in \mathbb{R}$, temos

$$0 = \xi\left(\underbrace{P(a+th)}_{=0_F}\right) = \sum_{j=0}^{n} \binom{n}{j} t^j \xi\left(\varphi\left(a^{(n-j)}, h^{(j)}\right)\right), \quad \text{para todo } \xi \in F^*.$$

Fixe $a, h \in E$ e $\xi \in F^*$ e ponha, para cada $t \in \mathbb{R}$,

$$Q(t) := \sum_{j=0}^{n} c_j t^j, \quad \text{onde } c_j := \binom{n}{j} \xi\left(\varphi\left(a^{(n-j)}, h^{(j)}\right)\right).$$

Como Q é um polinômio identicamente nulo, todos os coeficientes c_j's de Q devem ser zero. Por conseguinte,

$$\xi\left(\varphi\left(a^{(n-j)}, h^{(j)}\right)\right) = 0, \quad \text{para quaisquer} \quad a, h \in E, j = 0, \ldots, n \quad \text{e} \quad \xi \in F^*.$$

Pelo Teorema C.2.9,

$$\varphi\left(a^{(n-j)}, h^{(j)}\right) = 0, \quad \text{para quaisquer} \quad a, h \in E \quad \text{e} \quad j = 0, \ldots, n. \tag{3.4}$$

A implicação (3.2) se segue por indução sobre n. Para $n = 1$, temos $P = \varphi$ e a validade da implicação é automática. Suponhamos agora, por hipótese de indução, que o resultado é válido para $n = k$ e consideremos o caso $n = k + 1$.

Para cada $h \in E$, defina $\varphi_h : E^k \to F$ pondo

$$\varphi_h(x_1, \ldots, x_k) := \varphi(h, x_1, \ldots, x_k), \quad \text{para quaisquer} \quad x_1, \ldots, x_k \in E.$$

Note que φ_h é uma aplicação k-linear simétrica. Pela identidade (3.4),

$$\varphi_h \cdot x^{(k)} = \varphi\left(h, x^{(k)}\right) = 0_F, \quad \text{para quaisquer} \quad h, x \in E.$$

Assim, dado $h \in E$, a aplicação

$$P_h : x \in E \mapsto \varphi_h \cdot x^{(k)} \in F$$

é tal que

$$P_h(x) = 0_F, \quad \text{para todo} \quad x \in E,$$

donde, por hipótese de indução,

$$\varphi_h = 0_{\mathcal{L}_k(E,F)}, \quad \text{para todo} \quad h \in E.$$

Em particular, para quaisquer $x_1, \ldots, x_{k+1} \in E$, temos

$$\varphi(x_1, \ldots, x_{k+1}) = \varphi_{x_1}(x_2, \ldots, x_{k+1}) = 0_F.$$

Portanto, a implicação (3.2) é válida para $n = k+1$ e o resultado se segue do Princípio de Indução Matemática. Com isso, finalizamos a prova das expressões (3.1) e (3.2).

Suponhamos agora que E e F são espaços normados e φ é contínua. Provaremos a validade de (3.3).

Começamos com o caso $n = 1$. Como $P = \varphi$, temos

$$P'(a)h = \varphi \cdot h, \quad \text{para quaisquer} \quad a, h \in E,$$

como já vimos. Daí, fixando $a, h_1 \in E$, a aplicação

$$q : x \in E \mapsto P'(x)h_1 = \varphi \cdot h_1$$

é constante, donde

$$P^{(2)}(a)h_1h_2 = q'(a)h_2 = 0_F.$$

Consequentemente,

$$P^{(j)}(a)h_1 \ldots h_j = 0_F, \quad \text{para todo} \quad j \geq 2.$$

Suponhamos agora $n \geq 2$ e considere $a, h \in E$. A partir da expressão (3.1), temos

$$\frac{P(a+th) - P(a)}{t} = n\varphi\left(h, a^{(n-1)}\right) + \sum_{2 \leq k \leq n} \binom{n}{k} t^{k-1} \varphi\left(h^{(k)}, a^{(n-k)}\right),$$

donde

$$\frac{\partial P}{\partial h}(a) = n\varphi\left(h, a^{(n-1)}\right).$$

Como a aplicação

$$h \mapsto \varphi(h, a^{(n-1)})$$

depende linearmente de h, então $\dfrac{\partial P}{\partial(\cdot)}(a)$ é linear. Além disso, como

$$\|n\varphi(h, a^{(n-1)})\|_F \leq n\|\varphi\|_{\mathcal{L}_n(E,F)}\|h\|_E\|a\|_E^{n-1},$$

fazendo $\|h\|_E \to 0$ no lado direito da desigualdade acima, obtemos 0 como limite. Logo, pelo Teorema do Confronto (ver Teorema B.5.10), temos

$$\lim_{h \to 0_E} \|n\varphi(h, a^{(n-1)})\|_F = 0.$$

Portanto, $\dfrac{\partial P}{\partial(\cdot)}(a)$ é contínua. A fim de provar que P é diferenciável, resta provar que

$$\lim_{h \to 0_E} R(a, h) = 0_F, \ \text{para todo} \quad a \in E,$$

onde

$$R(a,h) := \frac{P(a+h) - P(a) - \dfrac{\partial P}{\partial h}(a)}{\|h\|_E} = \frac{\displaystyle\sum_{k=2}^{n} \binom{n}{k} \varphi\left(h^{(k)}, a^{(n-k)}\right)}{\|h\|_E}$$
$$= \sum_{k=2}^{n} \binom{n}{k} \underbrace{\varphi\left(\frac{h}{\|h\|_E}, h^{(k-1)}, a^{(n-k)}\right)}_{=:S(a,h,k)}.$$

Dados $a \in E$ e $k = 2, \ldots, n$, como φ é uma aplicação multilinear contínua, então

$$\|S(a,h,k)\|_F \leq \underbrace{\|\varphi\|_{\mathcal{L}_n(E,F)} \left\|\frac{h}{\|h\|_E}\right\|_E \|h\|_E^{k-1} \|a\|_E^{n-k}}_{=\|\varphi\|_{\mathcal{L}_n(E,F)} \|h\|_E^{k-1} \|a\|_E^{n-k}}.$$

Fazendo $\|h\|_E \to 0$ no lado direito da desigualdade acima, obtemos 0 como limite. Logo, novamente pelo Teorema do Confronto, temos

$$\lim_{h \to 0_E} S(a,h,k) = 0_F, \quad \text{para quaisquer} \quad a \in E \quad \text{e} \quad k = 2, \ldots, n,$$

donde

$$\lim_{h \to 0_E} R(a,h) = \sum_{k=2}^{n} \binom{n}{k} \lim_{h \to 0_E} S(a,h,k) = 0_F, \quad \text{para todo} \quad a \in E,$$

e, portanto, P é diferenciável, com

$$P'(a)h = n\varphi\left(h, a^{(n-1)}\right) = n\varphi\left(a^{(n-1)}, h\right), \quad \text{para todo} \quad a \in E.$$

Suponhamos agora que o resultado é válido para $j = k < n$ e consideremos o caso $j = k+1$. Fixando $h_1, \ldots, h_k \in E$, defina

$$G_k : E^{n-k} \to F$$

pondo, para quaisquer $x_1, \ldots, x_{n-k} \in E$,

$$G_k(x_1, \ldots, x_{n-k}) := \varphi(x_1, \ldots, x_{n-k}, h_1, \ldots, h_k).$$

Da hipótese de indução e da simetria de $G_k \in \mathcal{L}_{n-k}(E,F)$ vem que

$$s_k : x \in E \mapsto P^{(k)}(x)h_1 \ldots h_k = \binom{n}{k} k! \cdot \varphi(x^{(n-k)}, h_1, \ldots, h_k) = \binom{n}{k} k! \cdot G_k x^{(n-k)}$$

é diferenciável, com

$$P^{(k+1)}(a)h_1 \ldots h_{k+1} = s_k'(a)h_{k+1} = (n-k)\binom{n}{k} k! \cdot G_k\left(a^{[n-(k+1)]}, h_{k+1}\right)$$
$$= \binom{n}{k+1}(k+1)! \cdot \varphi\left(a^{[n-(k+1)]}, h_{k+1}, h_1, \ldots, h_k\right)$$
$$= \binom{n}{k+1}(k+1)! \cdot \varphi\left(a^{[n-(k+1)]}, h_1, \ldots, h_{k+1}\right),$$

para quaisquer $a, h_{k+1} \in E$. Isso prova que o resultado é válido para todo $j = 1, \ldots, n$. Em particular, no caso $j = n$, fixados $h_1, \ldots, h_n \in E$, a aplicação

$$x \in E \mapsto P^{(n)}(x)h_1 \ldots h_n = n! \cdot \varphi(h_1, \ldots, h_n)$$

é constante. Logo,

$$P^{(j)}(x) = 0_{\mathcal{L}_j(E,F)},$$

para quaisquer $x \in E$ e $j > n$. □

Aqui vai uma rápida demonstração alternativa da diferenciabilidade de P utilizando a Regra da Cadeia.

Considere a aplicação $g : E \to E^n$ dada por

$$g(x) = x^{(n)}, \quad \text{para cada} \quad x \in E.$$

Pelo Teorema 1.3.3, a aplicação g é diferenciável, valendo

$$g'(x)h = h^{(n)}, \quad \text{para quaisquer} \quad x, h \in E.$$

Além disso, pelo Exemplo 1.4.1, vimos que φ é diferenciável, com

$$\varphi'(x_1, \ldots, x_n)(h_1, \ldots, h_n) = \sum_{i=1}^{n} \varphi(x_1, \ldots, x_{i-1}, h_i, x_{i+1}, \ldots, x_n).$$

Como

$$P = \varphi \circ g,$$

então, pela Regra da Cadeia (ver Teorema 1.3.1), P é diferenciável, com

$$P'(x)h = (\varphi \circ g)'(x)h = \varphi'(g(x))(g'(x)h) = \varphi'(\underbrace{x, \ldots, x}_{n \text{ vezes}})(\underbrace{h, \ldots, h}_{n \text{ vezes}}) = n \cdot \varphi\left(x^{(n-1)}, h\right),$$

para quaisquer $x, h \in E$.

3.5 Compondo aplicações várias vezes diferenciáveis

O Teorema abaixo é uma versão modificada de (COLEMAN, 2012, Theorem 4.9., p. 101). Ele será útil para provar que a composição de aplicações k vezes diferenciáveis também é k vezes diferenciável.

Teorema 3.5.1. Se E, F_1, F_2 e G são espaços normados, $U \subseteq E$ é um subconjunto aberto, $f : U \to F_1$ e $g : U \to F_2$ são aplicações k vezes diferenciáveis em $a \in U$ e $b : F_1 \times F_2 \to G$ é uma aplicação bilinear contínua, então a aplicação $\varphi : U \to G$ dada por

$$\varphi(x) := b(f(x), g(x)), \quad \text{para cada } x \in U,$$

também é k vezes diferenciável em a.

Demonstração. O caso $k = 1$ é consequência do Exemplo 1.4.1.

Suponhamos que o resultado é valido para $k = n$ e consideremos o caso $k = n + 1$. Dados $A \in \mathcal{L}(E, F_1), B \in \mathcal{L}(E, F_2), x \in F_1$ e $y \in F_2$, considere as aplicações

$$l_{A,y}, l_{x,B} : E \to G$$

dadas por

$$l_{A,y}(h) := b(A(h), y) \quad \text{e} \quad l_{x,B}(h) := b(x, B(h)), \quad \text{para cada } h \in E.$$

Claramente, $l_{A,y}, l_{x,B} \in \mathcal{L}(E, G)$. Note que as aplicações

$$\alpha : (A, y) \in \mathcal{L}(E, F_1) \times F_2 \mapsto l_{A,y} \in \mathcal{L}(E, G)$$

e

$$\beta : (x, B) \in F_1 \times \mathcal{L}(E, F_2) \mapsto l_{x,B} \in \mathcal{L}(E, G)$$

são bilineares e contínuas. Como f, f', g e g' são k vezes diferenciáveis em a, então, por hipótese de indução, as aplicações

$$x \in U \mapsto \alpha(f'(x), g(x)) \text{ e } x \in U \mapsto \beta(f(x), g'(x))$$

são k vezes diferenciáveis em a. Como

$$\varphi'(x) = \alpha(f'(x), g(x)) + \beta(f(x), g'(x)),$$

então φ' é k vezes diferenciável em a e, portanto, φ é $k + 1$ vezes diferenciável em a.

O resultado então segue do Princípio de Indução Matemática. □

O corolário abaixo é uma versão modificada do Theorem 4.10. de (COLEMAN, 2012, p. 103).

Corolário 3.5.2. Sejam E, F e V espaços normados e $X \subseteq E$ e $Y \subseteq F$ subconjuntos abertos. Se

$$f : X \to F$$

e

$$g : Y \to V$$

são aplicações, respectivamente, k vezes diferenciável em $a \in X$ e k vezes diferenciável em $f(a)$, com $f(X) \subseteq Y$, então a composição

$$g \circ f : X \to V$$

também é k vezes diferenciável em a.

Demonstração. Note que a aplicação

$$\alpha : (A, B) \in \mathcal{L}(E, F) \times \mathcal{L}(F, G) \mapsto B \circ A \in \mathcal{L}(E, G)$$

é bilinear e contínua. Como f e g são diferenciáveis em a, então, supondo $k \geq 2$, temos

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

Pelo Teorema 3.5.1, a aplicação

$$x \in X \mapsto \alpha(f'(x), g'(f(x))) \in \mathcal{L}(E, G)$$

é $k-1$ vezes diferenciável em a. Logo, $g \circ f$ é k vezes diferenciável em a. □

3.6 A simetria das derivadas de ordem superior

É bem conhecido o fato de que aplicações duas vezes *continuamente* diferenciáveis possuem derivadas de Fréchet de segunda ordem simétricas (ver Definição C.7.6): veja (LIMA, 2002, p. 69) ou (CIPOLATTI, 2020, p. 122). No entanto, o Teorema abaixo, adaptado de (COLEMAN, 2012, p. 91), mostra que a simetria não depende da continuidade da derivada segunda.

Teorema 3.6.1 (Teorema de Schwarz)**.** Se E e F são espaços normados, $U \subseteq E$ é um subconjunto aberto e $f : U \to F$ é duas vezes diferenciável em $a \in U$, então a aplicação bilinear $f^{(2)}(a) \in \mathcal{L}_2(E, F)$ é simétrica.

Demonstração. Como $a \in U$, então existe $\delta_1 > 0$ tal que $B_E(a, \delta_1) \subseteq U$.

Considere $\delta_2 := \dfrac{\delta_1}{2}$, $B := B_E(0, \delta_2)$ e a aplicação

$$\Delta : (h, k) \in B \times B \mapsto f(a + h + k) - f(a + k) - f(a + h) + f(a) \in F.$$

Note que

$$\Delta(h, k) = \Delta(k, h), \text{ para quaisquer } h, k \in B.$$

Além disso, para quaisquer $h, k \in B$, temos

$$\begin{aligned}\Delta(h, k) - f^{(2)}(a)(h, k) = f(a + h + k) - f(a + h) - f'(a + k)h + f'(a)h \\ + [f'(a + k) - f'(a) - f^{(2)}(a)(\cdot, k)]h \\ - [f(a + k) - f(a)].\end{aligned}$$

Para cada $k \in B$, considere a aplicação

$$G_k : h \in B \mapsto f(a + h + k) - f(a + h) - f'(a + k)h + f'(a)h \in F.$$

Da Desigualdade do Valor Médio (ver Teorema 2.3.3), para quaisquer $h, k \in B$, temos

$$\begin{aligned}\|\Delta(h,k) - f^{(2)}(a)(h,k)\|_F &\leq \|G_k(h) - G_k(0)\|_F \\ &\quad + \|f'(a+k) - f'(a) - f^{(2)}(a)(\cdot,k)\|_{\mathcal{L}(E,F)}\|h\|_E \\ &\leq \sup_{0\leq\lambda\leq 1} \|{G_k}'(\lambda h)\|_{\mathcal{L}(E,F)}\|h\|_E \\ &\quad + \|f'(a+k) - f'(a) - f^{(2)}(a)(\cdot,k)\|_{\mathcal{L}(E,F)}\|h\|_E.\end{aligned}$$

Como, para quaisquer $k, u \in B$,

$$\begin{aligned}{G_k}'(u) &= f'(a+u+k) - f'(a+u) - f'(a+k) + f'(a) \\ &= f'(a+u+k) - f'(a) - f^{(2)}(a)(\cdot,u+k) \\ &\quad - (f'(a+u) - f'(a) - f^{(2)}(a)(\cdot,u)) \\ &\quad - (f'(a+k) - f'(a) - f^{(2)}(a)(\cdot,k)),\end{aligned}$$

então, escrevendo $\|\cdot\|$ para $\|\cdot\|_{\mathcal{L}(E,F)}$, para quaisquer $k, u \in B$, temos

$$\begin{aligned}\|{G_k}'(u)\| &\leq \|f'(a+u+k) - f'(a) - f^{(2)}(a)(\cdot,u+k)\| \\ &\quad + \|f'(a+u) - f'(a) - f^{(2)}(a)(\cdot,u)\| \\ &\quad + \|f'(a+k) - f'(a) - f^{(2)}(a)(\cdot,k)\|.\end{aligned}$$

Como f é duas vezes diferenciável em a, então, dado $\varepsilon > 0$, existe $\delta_3 > 0$ tal que, para todo $p \in B$ com $0 < \|p\|_E < \delta_3$, vale

$$\frac{\|f'(a+p) - f'(a) - f^{(2)}(a)(\cdot,p)\|}{\|p\|_E} < \varepsilon$$

e, portanto,

$$\|f'(a+p) - f'(a) - f^{(2)}(a)(\cdot,p)\| < \varepsilon\|p\|_E.$$

Pondo $\delta_4 := \min\{\delta_2, \delta_3\}$, para quaisquer $k, u \in B$ com $\|k\|_E + \|u\|_E < \delta_4$, temos

$$\begin{aligned}\max\{\|u\|_E, \|k\|_E, \|u+k\|_E\} < \delta_4 &\implies \|{G_k}'(u)\| < \varepsilon\|u+k\|_E + \varepsilon\|u\|_E + \varepsilon\|k\|_E \\ &\implies \|{G_k}'(u)\| < 2\varepsilon(\|u\|_E + \|k\|_E),\end{aligned}$$

donde, para quaisquer $\lambda \in [0,1]$ e $h, k \in B$ com $\|h\|_E + \|k\|_E < \delta_4$,

$$\|{G_k}'(\lambda h)\| \leq 2\varepsilon(\|\lambda h\|_E + \|k\|_E) = 2\varepsilon(\lambda\|h\|_E + \|k\|_E) \leq 2\varepsilon\,(\|h\|_E + \|k\|_E)$$

e, portanto, para quaisquer $h, k \in B$ com $\|h\|_E + \|k\|_E < \delta_4$, temos

$$\sup_{0\leq\lambda\leq 1} \|{G_k}'(\lambda h)\| \leq 2\varepsilon(\|h\|_E + \|k\|_E).$$

Logo, dados $h, k \in B$ com $\|h\|_E + \|k\|_E < \delta_4$, temos

$$\begin{aligned}\|\Delta(h,k) - f^{(2)}(a)(h,k)\|_F \leq [2\varepsilon(\|h\|_E + \|k\|_E) + \varepsilon\|k\|_E]\,\|h\|_E &= 2\varepsilon\|h\|_E^2 + 3\varepsilon\|k\|_E\|h\|_E \\ &\leq 2\varepsilon(\|h\|_E + \|k\|_E)^2\end{aligned}$$

que, combinado com a simetria de Δ, nos dá

$$\begin{aligned}\|f^{(2)}(a)(h,k) - f^{(2)}(a)(k,h)\|_F &= \|f^{(2)}(a)(h,k) - \Delta(h,k) + \Delta(k,h) - f^{(2)}(a)(k,h)\|_F \\ &\leq 4\varepsilon(\|h\|_E + \|k\|_E)^2.\end{aligned}$$

Portanto, para quaisquer $h, k \in B$ com $\|h\|_E + \|k\|_E < \delta_4$,

$$\|f^{(2)}(a)(h,k) - f^{(2)}(a)(k,h)\| \leq 4\varepsilon(\|h\|_E + \|k\|_E)^2.$$

Tomando $\alpha := \dfrac{\delta_4}{2}$, temos, para quaisquer $x, y \in E \setminus \{0\}$, vale

$$\left\|\alpha\frac{x}{\|x\|_E}\right\|_E = \left\|\alpha\frac{y}{\|y\|_E}\right\|_E = \alpha < \delta_4 \leq \delta_2,$$

donde $\alpha\dfrac{x}{\|x\|_E}, \alpha\dfrac{y}{\|y\|_E} \in B$. Como, além disso, para quaisquer $x, y \in E \setminus \{0\}$, vale

$$f^{(2)}(a)\left(\alpha\frac{x}{\|x\|_E}, \alpha\frac{y}{\|y\|_E}\right) = \frac{\alpha^2}{\|x\|_E\|y\|_E}f^{(2)}(a)(x,y),$$

então, fixados $x, y \in E \setminus \{0\}$ e tomando $h := \alpha\dfrac{x}{\|x\|_E}$ e $k := \alpha\dfrac{y}{\|y\|_E}$, temos

$$\frac{\alpha^2}{\|x\|_E\|y\|_E}\|f^{(2)}(a)(x,y) - f^{(2)}(a)(y,x)\|_F = \left\|f^{(2)}(a)(h,k) - f^{(2)}(a)(k,h)\right\|_F \leq 16\varepsilon\alpha^2.$$

Por conseguinte, para todo $\varepsilon > 0$, temos

$$\|f^{(2)}(a)(x,y) - f^{(2)}(a)(y,x)\|_F \leq 16\varepsilon\|x\|_E\|y\|_E.$$

De

$$\lim_{\varepsilon\to 0}(16\varepsilon\|x\|_E\|y\|_E) = 0,$$

temos, pelo Teorema do Confronto (ver Teorema B.5.10), que

$$\|f^{(2)}(a)(x,y) - f^{(2)}(a)(y,x)\|_F = \lim_{\varepsilon\to 0}\|f^{(2)}(a)(x,y) - f^{(2)}(a)(y,x)\|_F = 0.$$

Isto prova que, para quaisquer $x, y \in E \setminus \{0\}$,

$$f^{(2)}(a)(x,y) = f^{(2)}(a)(y,x).$$

Por fim, o resultado se segue do fato de que, para quaisquer $x, y \in E$,

$$f^{(2)}(a)(x, 0_E) = f^{(2)}(a)(0_E, y) = 0_F,$$

já que $f^{(2)}(a)$ é uma aplicação bilinear. $\square$

Corolário 3.6.2. O operador hessiano é sempre autoadjunto e, consequentemente, a matriz hessiana (ver Exemplo 1.6.1) é sempre simétrica.

Demonstração. Sejam H um espaço de Hilbert, $U \subseteq H$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função diferenciável cuja aplicação gradiente $\nabla f : U \to H$ é diferenciável em $a \in U$. Nessas condições, f é duas vezes diferenciável em a, pelo Exemplo 3.3.2, e

$$f^{(2)}(a)(x, y) = \langle x, \nabla^2 f(a)y \rangle, \quad \text{para quaisquer } x, y \in H.$$

Pelo Teorema de Schwarz (ver Teorema 3.6.1), para quaisquer $x, y \in H$, temos

$$\langle x, \nabla^2 f(a)y \rangle = f^{(2)}(a)(x, y) = f^{(2)}(a)(y, x) = \langle y, \nabla^2 f(a)x \rangle.$$

Logo, $\nabla^2 f(a)$ é um operador autoadjunto.

Em particular, quando $H = \mathbb{R}^n$, para quaisquer $x, y \in \mathbb{R}^n$, temos

$$x^t \operatorname{H} f(a) y = x^t [\nabla^2 f(a)] y = y^t [\nabla^2 f(a)] x = y^t \operatorname{H} f(a) x,$$

donde a matriz hessiana $\operatorname{H} f(a)$ é simétrica, pelo Corolário C.6.3. □

O corolário abaixo, extraído de (COLEMAN, 2012, p. 92), prova que todas as diferenciais de ordem superior são exemplos de aplicações multilineares simétricas.

Corolário 3.6.3 (Generalização do Teorema de Schwarz)**.** Se E e F são espaços normados, $U \subseteq E$ é um subconjunto aberto e $f : U \to F$ é k vezes diferenciável em $a \in U$, então a aplicação k-linear $f^{(k)}(a) \in \mathcal{L}_k(E, F)$ é simétrica.

Demonstração. Considere $k \geq 2$ um natural para o qual o corolário é verdadeiro e consideremos o caso $k + 1$. Assim, para qualquer permutação $\rho \in S_k$, temos

$$\begin{aligned}
f^{(k)\prime}(a)h(h_1, \ldots, h_k) &= \left(\lim_{t \to 0} \frac{f^{(k)}(a + th) - f^{(k)}(a)}{t} \right) (h_1, \ldots, h_k) \\
&= \lim_{t \to 0} \left[\frac{f^{(k)}(a + th) - f^{(k)}(a)}{t} (h_1, \ldots, h_k) \right] \\
&= \lim_{t \to 0} \frac{f^{(k)}(a + th)(h_1, \ldots, h_k) - f^{(k)}(a)(h_1, \ldots, h_k)}{t} \\
&= \lim_{t \to 0} \frac{f^{(k)}(a + th) \left(h_{\rho(1)}, \ldots, h_{\rho(k)} \right) - f^{(k)}(a) \left(h_{\rho(1)}, \ldots, h_{\rho(k)} \right)}{t} \\
&= \lim_{t \to 0} \left[\frac{f^{(k)}(a + th) - f^{(k)}(a)}{t} \left(h_{\rho(1)}, \ldots, h_{\rho(k)} \right) \right] \\
&= \left(\lim_{t \to 0} \frac{f^{(k)}(a + th) - f^{(k)}(a)}{t} \right) \left(h_{\rho(1)}, \ldots, h_{\rho(k)} \right) \\
&= f^{(k)\prime}(a)h \left(h_{\rho(1)}, \ldots, h_{\rho(k)} \right).
\end{aligned}$$

Desse modo,

$$f^{(k+1)}(a)(h_1, \ldots, h_{k+1}) = f^{(k+1)}(a)\left(h_{\rho(1),}, \ldots, h_{\rho(k)}, h_{k+1}\right). \tag{3.5}$$

Como

$$g^{(2)}(a)(x_k, x_{k+1}) = f^{(k+1)}(a)(x_1, \ldots, x_k, x_{k+1}),$$

pelo Teorema 3.3.4, onde

$$g(x) := f^{(k-1)}(x)(x_1, \ldots, x_{k-1}),$$

então, pelo Teorema 3.6.1, temos, para quaisquer $x, y \in E$,

$$g^{(2)}(a)(x, y) = g^{(2)}(a)(y, x),$$

donde, para quaisquer $x_1, \ldots, x_{k+1} \in E$, temos

$$f^{(k+1)}(a)(x_1, \ldots, x_{k-1}, x_k, x_{k+1}) = f^{(k+1)}(a)(x_1, \ldots, x_{k-1}, x_{k+1}, x_k). \tag{3.6}$$

Seja $\sigma \in S_{k+1}$ uma permutação. Como $\sigma : [k+1] \to [k+1]$ é uma bijeção, existe um único $p \in [k+1]$ tal que $\sigma(p) = k+1$. Se $p = k+1$, então, de (3.5), vem

$$\begin{aligned} f^{(k+1)}(a)\left(x_{\sigma(1)}, \ldots, x_{\sigma(k)}, x_{\sigma(k+1)}\right) &= f^{(k+1)}(a)\left(x_{\sigma(1)}, \ldots, x_{\sigma(k)}, x_{k+1}\right) \\ &= f^{(k+1)}(a)(x_1, \ldots, x_k, x_{k+1}). \end{aligned}$$

Suponhamos agora que $p \neq k+1$. Defina

$$\eta_1, \eta_2 : [k+1] \to [k+1]$$

pondo, para quaisquer $i, j \in [k+1]$, com $i \neq p, k$ e $j \neq p, k+1$,

$$\eta_1(\sigma(i)) = \sigma(i) \text{ e } \eta_2(\sigma(j)) = \sigma(j),$$

e

$$\eta_1(\sigma(p)) = \sigma(k), \eta_1(\sigma(k)) = \sigma(p), \eta_2(\sigma(p)) = \sigma(k+1) \text{ e } \eta_2(\sigma(k+1)) = \sigma(p).$$

Nessas condições,

$$\eta_1, \eta_2 \in S_{k+1}$$

e

$$(\eta_2 \circ \eta_1 \circ \sigma)(k+1) = \eta_2(\eta_1(\sigma(k+1))) = \eta_2(\sigma(k+1)) = \sigma(p) = k+1.$$

Logo,

$$\rho := \eta_2 \circ \eta_1 \circ \sigma \in S_{k+1}$$

é uma permutação tal que

$$\rho(k+1) = k+1.$$

Dados $h := (h_1, \ldots, h_{k+1}) \in E^{k+1}$, ponha

$$\pi h := \left(h_{\pi(1)}, \ldots, h_{\pi(k+1)}\right),$$

para toda permutação $\pi \in S_{k+1}$. Assim, fixado $x = (x_1, \ldots, x_{k+1}) \in E^{k+1}$, de (3.5) temos

$$f^{(k+1)}(a)(\sigma x) = f^{(k+1)}(a)(\eta_1 \sigma x),$$

donde, de (3.6) e do caso anterior, temos

$$f^{(k+1)}(a)(\eta_1 \sigma x) = f^{(k+1)}(a)(\eta_2 \eta_1 \sigma x) = f^{(k+1)}(a)(\rho x) = f^{(k+1)}(a)(x).$$

Logo, o corolário é verdadeiro para $k+1$.

Portanto, o resultado é válido para todo natural $k \geq 2$. □

3.7 Fórmula de Taylor com resto infinitesimal

Lema 3.7.1. (LIMA, 2002, p. 74) Se E e F são espaços normados e $\delta > 0$ tal que $r : B \to F$ é uma aplicação $k+1$ vezes diferenciável em 0_E, onde $B := B_E(0_E, \delta)$, com

$$r^{(j)}(0_E) = 0_{\mathcal{L}_j(E,F)}, \text{ para todo } j = 0, 1, \ldots, k+1,$$

então

$$\lim_{x \to 0} \frac{r(x)}{\|x\|_E^{k+1}} = 0_F.$$

Demonstração. Para $k = 0$, o lema é consequência da definição de função diferenciável. Suponhamos, por hipótese de indução, que ele é válido para $k-1$, onde $k > 0$, e consideremos o caso enunciado. Pela Desigualdade do Valor Médio (ver Corolário 2.3.3),

$$\|r(x)\|_F \leq \|x\|_E M, \quad \text{para todo } x \in E, \text{ onde } M := \sup_{0 \leq t \leq 1} \|r'(tx)\|_{\mathcal{L}(E,F)}.$$

Como r' é $k-1$ vezes diferenciável em B e k vezes diferenciável em 0_E, com

$$r^{(j)}(0_E) = 0_{\mathcal{L}_j(E,F)}, \quad \text{para todo } j = 1, \ldots, k,$$

então, aplicando a hipótese de indução para r', temos

$$\lim_{x \to 0} \frac{r'(x)}{\|x\|_E^k} = 0_F.$$

Assim, dado $\varepsilon > 0$, existe $\delta_1 > 0$ tal que, para todo $x \in B$,

$$\|x\|_E < \delta_1 \implies \frac{\|r'(x)\|_{\mathcal{L}(E,F)}}{\|x\|_E^k} < \varepsilon \implies \|r'(x)\|_{\mathcal{L}(E,F)} < \varepsilon \|x\|_E^k.$$

Dado $t \in [0,1]$, para todo $x \in B$ com $\|tx\|_E = t\|x\|_E < \delta_1$, temos

$$\|r'(tx)\|_{\mathcal{L}(E,F)} < \varepsilon \|tx\|_E^k,$$

donde, para todo $x \in B$,

$$M = \sup_{0 \leq t \leq 1} \|r'(tx)\|_{\mathcal{L}(E,F)} \leq \sup_{0 \leq t \leq 1} \varepsilon \|tx\|_E^k \leq \varepsilon \|x\|_E^k$$

e, portanto,

$$\|r(x)\|_F \leq \varepsilon \|x\|_E^{k+1}.$$

Isto prova que o lema é válido para k.

O resultado se segue do Princípio de Indução Matemática. □

O Teorema abaixo foi retirado de (LIMA, 2002, Teorema 9.1., p. 74).

Teorema 3.7.2 (Fórmula de Taylor com resto infinitesimal)**.** Se E e F são espaços normados, $U \subseteq E$ é um subconjunto aberto e $f : U \to F$ é uma função $k+1$ vezes diferenciável em $a \in U$, então, para todo $h \in E$ tal que $a + h \in U$, temos

$$f(a+h) = f(a) + f'(a)h + \cdots + \frac{1}{(k+1)!} f^{(k+1)}(a)h^{(k+1)} + r(h),$$

com

$$\lim_{h \to 0} \frac{r(h)}{\|h\|_E^{k+1}} = 0_F.$$

Demonstração. Como U é aberto, então existe $\delta > 0$ tal que $B(a,\delta) \subseteq U$. Considere a aplicação $r : B(0_E, \delta) \to F$ dada por

$$r(x) := f(a+x) - f(a) - f'(a)x - \cdots - \frac{1}{(k+1)!} f^{(k+1)}(a)x^{(k+1)}.$$

Veja que $r(0_E) = 0_F$. Faça

$$T_i(x) := 1/(i!) f^{(i)}(a)x^{(i)}, \quad \text{para todo} \quad i = 1, \ldots, k+1.$$

Então, de (3.3), temos

$$T_i^{(j)}(0_E) = \begin{cases} f^{(j)}(a)(h_1, \ldots, h_j), & \text{se} \quad j = i, \\ 1/(i!)\binom{i}{j} j! f^{(i)}(a)(0_E^{(i-j)}, h_1, \ldots, h_j) = 0, & \text{se} \quad j < i, \\ 0_F, \quad \text{se} \quad j > i. \end{cases}$$

Além disso, por indução é fácil ver que se $g(x) := f(a+x)$, então

$$g^{(j)}(x) = f^{(j)}(a+x), \quad \text{para todo} \quad j = 1, \ldots, k, \quad \text{e} \quad g^{(k+1)}(0_E) = f^{(k+1)}(a).$$

Isso nos dá

$$r^{(j)}(0_E) = f^{(j)}(a) - f^{(j)}(a) = 0, \quad \text{para todo} \quad j = 1, \ldots, k+1.$$

Pelo Lema 3.7.1, temos

$$\lim_{x \to 0} \frac{r(x)}{\|x\|_E^{k+1}} = 0_F,$$

como queríamos demonstrar. □

3.8 Fórmula de Taylor com resto de Lagrange

Lema 3.8.1. (COLEMAN, 2012, Lemma 5.2., p. 109) Se E e F são espaços normados, $U \subseteq E$ é um subconjunto aberto, $f : U \to F$ é uma função $(k+1)$-vezes diferenciável, $a \in U$ e $h \in E$ tais que $[a, a+h] \subseteq U$, então $\varphi : [0,1] \to F$ dada por

$$\varphi(t) := f(a+th) + \sum_{i=1}^{k} \frac{(1-t)^i}{i!} f^{(i)}(a+th)h^{(i)},$$

para cada $t \in [0,1]$, é contínua em $[0,1]$ e diferenciável em $(0,1)$ com

$$\frac{d\varphi}{dt}(t) = \frac{(1-t)^k}{k!} f^{(k+1)}(a+th)h^{(k+1)}.$$

Demonstração. A continuidade de φ é evidente. Pondo

$$\varphi_i(x) := f^{(i)}(x)h^{(i)}, \quad \text{para cada } i = 1, \ldots, k,$$

então, pelo Teorema 3.3.4, cada φ_i é diferenciável, com

$$\frac{d}{dt} f^{(i)}(a+th)h^{(i)} = \frac{d}{dt}\varphi_i(a+th) = \varphi_i'(a+th)h = f^{(i+1)}(a+th)h^{(i+1)}.$$

A partir do Teorema 1.2.1 e da Regra da Cadeia (ver Teorema 1.3.1), temos

$$\begin{aligned} \frac{d}{dt}\frac{(1-t)^i}{i!} f^{(i)}(a+th)h^{(i)} = {} & \frac{(1-t)^i}{i!} f^{(i+1)}(a+th)h^{(i+1)} \\ & - \frac{(1-t)^{i-1}}{(i-1)!} f^{(i)}(a+th)h^{(i)}. \end{aligned}$$

Logo,

$$\frac{d\varphi}{dt}(t) = \frac{(1-t)^k}{k!} f^{(k+1)}(a+th)h^{(k+1)}.$$

como queríamos demonstrar. □

Lema 3.8.2. Se $\varphi \in \mathcal{L}_n(E,F)$, então, para todo $x \in E$, temos

$$\|\varphi \cdot x^{(n)}\|_F \le \sup_{\|y\|_E = 1} \|\varphi \cdot y^{(n)}\|_F \cdot \|x\|_E^n \le \|\varphi\|_{\mathcal{L}_n(E,F)} \|x\|_E^n.$$

Demonstração. A segunda desigualdade vem do fato de que, na definição de $\|\cdot\|_{\mathcal{L}_n(E,F)}$, o sup é tomado sobre φ atuando em quaisquer n vetores unitários de E, o que inclui φ atuando em quaisquer n vetores unitários iguais de E.

Provaremos agora a validade da primeira desigualdade. Dado $x \in E$, temos

$$\varphi \cdot x^{(n)} = \|x\|_E^n \cdot \left[\varphi \cdot \left(\frac{x}{\|x\|_E}\right)^{(n)}\right],$$

donde

$$\|\varphi \cdot x^{(n)}\|_F = \left\|\varphi \cdot \left(\frac{x}{\|x\|_E}\right)^{(n)}\right\|_F \|x\|_E^n \leq \sup_{\|y\|_E=1} \|\varphi \cdot y^{(n)}\|_F \cdot \|x\|_E^n,$$

como queríamos demonstrar. □

O Teorema abaixo foi retirado de (COLEMAN, 2012, Theorem 5.2, p. 110).

Teorema 3.8.3 (Fórmula de Taylor com resto de Lagrange)**.** Sejam E e F espaços normados, $U \subseteq E$ um subconjunto aberto e $f : U \to F$ uma aplicação $k+1$ vezes diferenciável. Suponha que $a \in U$ e que $x \in E$ tal que $[a, a+x] \subseteq U$. Assim,

$$f(a+x) = f(a) + f^{(1)}(a)x + \frac{1}{2}f^{(2)}(a)x^{(2)} + \cdots + \frac{1}{k!}f^{(k)}(a)x^{(k)} + \mathcal{R}(a,x),$$

onde

$$\|\mathcal{R}(a,x)\|_F \leq \frac{\|x\|_E^{k+1}}{(k+1)!} \sup_{0\leq\lambda\leq 1} \|f^{(k+1)}(a+\lambda x)\|_{\mathcal{L}_{k+1}(E,F)}.$$

Demonstração. Considere a aplicação $\varphi : [0,1] \to F$ dada por

$$\varphi(t) := f(a+tx) + \sum_{i=1}^{k} \frac{(1-t)^i}{i!} f^{(i)}(a+tx)x^{(i)}.$$

Do Lema 3.8.1 e do Lema 3.8.2, temos

$$\left\|\frac{d\varphi}{dt}(t)\right\|_F \leq \frac{(1-t)^k}{k!}\|f^{(k+1)}(a+tx)\|_{\mathcal{L}_{k+1}(E,F)}\|x\|_E^{k+1} \leq \frac{(1-t)^k}{k!}C,$$

onde

$$C := \sup_{0\leq\lambda\leq 1} \|f^{(k+1)}(a+\lambda x)\|_{\mathcal{L}_{k+1}(E,F)}\|x\|_E^{k+1}.$$

Pondo

$$\psi(t) := -\frac{(1-t)^{k+1}}{(k+1)!}C,$$

temos, pela Regra da Cadeia (ver Teorema 1.3.1) e pelo Teorema 1.2.1,

$$\frac{d\psi}{dt}(t) = \frac{(1-t)^k}{k!}C.$$

Portanto, pelo Teorema 2.3.1,

$$\|\varphi(1) - \varphi(0)\|_F \leq \psi(1) - \psi(0) = \frac{C}{(k+1)!}.$$

Como

$$\varphi(1) - \varphi(0) = f(a+x) - f(a) - \sum_{i=1}^{k} \frac{1}{i!} f^{(i)}(a) x^{(i)},$$

segue-se o resultado. □

4 Os teoremas da aplicação inversa e implícita

Os teoremas da aplicação inversa e da aplicação implícita são especialmente úteis na Geometria, constituindo a base teórica do estudo das variedades diferenciáveis. Essas aplicações no entanto fogem ao escopo deste trabalho: aqui nos preocuparemos apenas em provar os teoremas mencionados.

4.1 O Teorema da Aplicação Inversa

O teorema desta seção estabelece condições sob as quais uma aplicação diferenciável pode ser localmente invertida e, mais importante, que sua aplicação inversa também seja diferenciável.

Para o que se segue, sejam U e V subconjuntos abertos de espaços normados E e F, respectivamente. Dizemos que uma aplicação $f : U \to V$ é um **difeomorfismo** se f for bijetiva e tanto f quanto f^{-1} forem diferenciáveis.

Proposição 4.1.1. (COLEMAN, 2012, Proposition 8.1, p. 161) Sejam E e F espaços de Banach, $U \subseteq E$ e $V \subseteq F$ subconjuntos abertos e $f : U \to V$ um homeomorfismo diferenciável. Se $a \in U$ e $f'(a)$ é invertível, então f^{-1} é diferenciável em $b = f(a)$.

Se, além disso, f é de classe C^1, então existe uma vizinhança aberta U' de a tal que $f|_{U'}$ é um difeomorfismo de classe C^1 sobre sua imagem.

Demonstração. Para simplificar a notação, ponhamos

$$g := f^{-1} \quad \text{e} \quad L := f'(a).$$

Pelo Teorema C.10.4, a aplicação L^{-1} é contínua. Se k é pequeno, $f(a) + k \in V$ e, como f é bijetiva, existe um único $h = h(k) \in U$ tal que

$$f(a + h) = f(a) + k.$$

Primeiro notemos que

$$g(b + k) - g(b) = f^{-1}(f(a) + (f(a + h) - f(a))) - f^{-1}(f(a)) = a + h - a = h.$$

Como g é contínua, $\lim\limits_{k\to 0} h = 0$. Agora,

$$k = f(a + h) - f(a) = L(h) + \|h\|r(h),$$

onde $\lim\limits_{h\to 0} r(h) = 0$. Disso obtemos

$$L^{-1}(k) = h + \|h\|L^{-1}(r(h)) = g(b + k) - g(b) + \|h\|L^{-1}(r(h)).$$

Vamos considerar o termo $\|h\|L^{-1}(r(h))$. Mostraremos que

$$\|h\|L^{-1}(r(h)) = o(\|k\|).$$

Como $L^{-1}(k) = h + \|h\|L^{-1}(r(h))$, temos

$$\|h\| \leq \|h\|L^{-1}(r(h))\| + \|L^{-1}(k)\| \leq \|h\|\|L^{-1}(r(h))\| + \|L^{-1}\|\|k\|.$$

Se k é pequeno, então h também é, donde $L^{-1}(r(h))$ é pequeno. Em particular, existe $\delta > 0$ tal que $\|k\| < \delta \implies \|L^{-1}(r(h))\| < 1/2$. Isso nos permite escrever, para $\|k\| < \delta$,

$$\|h\| \leq \frac{\|L^{-1}\|}{1 - \|L^{-1}(r(h))\|}\|k\| \leq M\|k\|,$$

onde $M := 2\|L^{-1}\| > 0$. Consequentemente,

$$\left\| \frac{\|h\|L^{-1}(r(h))}{\|k\|} \right\| \leq M\|L^{-1}(r(h))\| \xrightarrow{k \to 0} 0.$$

Pelo Teorema do Confronto (ver Teorema B.5.10), temos $\|h\|L^{-1}(r(h)) = o(\|k\|)$, como queríamos demonstrar.

Agora vamos para o segundo item da proposição. Suponhamos que o mapa f também é de classe C^1. O subconjunto $\mathcal{I}(E,F)$ de todos os mapas invertíveis em $\mathcal{L}(E,F)$ é aberto (ver Teorema C.8.11). Seja V' uma vizinhança aberta de $f'(a)$ em $\mathcal{I}(E,F)$ e $U' := (f')^{-1}(V')$. Então U' é um subconjunto aberto de E tal que $f'(x)$ é invertível se $x \in U'$. Agora, aplicando a primeira parte da proposição a cada $x \in U'$, obtemos que f^{-1} é diferenciável em cada $y \in f(U')$. Consequentemente, f restrita a U' é um difeomorfismo. Como f é de classe C^1, f é um difeomorfismo de classe C^1. □

A versão aqui enunciada do Teorema da Aplicação Inversa tal qual sua demonstração se encontram em (COLEMAN, 2012, Theorem 8.1, p. 162).

Teorema 4.1.2 (Teorema da Aplicação Inversa)**.** Sejam E e F espaços de Banach, $U \subseteq E$ aberto e $f : U \to F$ uma aplicação de classe C^1. Se $a \in U$ e $f'(a)$ é invertível, então existe uma vizinhança aberta U' de a tal que $f|_{U'}$ é um difeomorfismo de classe C^1 sobre sua imagem.

Demonstração. Começamos supondo $E = F$, $f(a) = a = 0$ e $f'(a) = \mathrm{Id}_E$.

Como f é de classe C^1, dado $\varepsilon = \dfrac{1}{2}$, existe $r > 0$ tal que, para todo $x \in B[0,r]$, temos

$$\|f'(x) - \mathrm{Id}_E\| \leq \frac{1}{2}.$$

Ponha

$$g(x) := f(x) - x, \quad \text{para cada} \quad x \in U.$$

Então,

$$g'(x) = f'(x) - \mathrm{Id}_E\,.$$

Pela Desigualdade do Valor Médio (ver Corolário 2.3.3), temos

$$\|f(x) - x\| = \|g(x) - g(0)\| \leq \sup_{y\in(0,x)} \|g'(y)\|\|x\| \leq \sup_{y\in(0,x)} \|f'(y) - \mathrm{Id}_E\,\|\|x\| \leq \frac{1}{2}\|x\|.$$

Tome $y \in B(0, \frac{r}{2})$. Então,

$$\|y + x - f(x)\| \leq \|y\| + \|x - f(x)\| \leq \|y\| + \frac{1}{2}\|x\| < \frac{r}{2} + \frac{r}{2} = r,$$

donde temos

$$y + x - f(x) \in B(0, r) \subseteq B[0, r].$$

Para cada $y \in B(0, \frac{r}{2})$, defina

$$\varphi_y : B[0, r] \to B[0, r]$$

pondo

$$\varphi_y(x) := y + x - f(x), \quad \text{para cada} \quad x \in B[0, r].$$

Dados $w, z \in B[0, r]$, temos

$$\begin{aligned}
\|\varphi_y(w) - \varphi_y(z)\| = \|y + w - f(w) - (y + z - f(z))\| &= \|w - f(w) - z + f(z)\| \\
&= \|g(z) - g(w)\| \\
&\leq \sup_{y\in(w,z)} \|g'(y)\|\|z - w\| \\
&\leq \frac{1}{2}\|z - w\|.
\end{aligned}$$

Logo, φ_y é uma contração. Do Teorema do Ponto Fixo de Banach (ver Teorema C.11.2), existe um único $x \in B[0, r]$ tal que $y + x - f(x) = \varphi_y(x) = x$, donde $f(x) = y$. Como $\operatorname{im}\varphi_y \subseteq B(0, r)$, temos $x \in B(0, r)$. Ponha

$$\tilde{U} := f^{-1}\left(B\left(0, \frac{r}{2}\right)\right) \cap B(0, r).$$

Então, $\tilde{U}$ é uma vizinhança de 0 tal que $f(\tilde{U}) = B(0, \frac{r}{2})$. Como

$$|\|w - z\| - \|f(w) - f(z)\|| \leq \|w - z - f(w) + f(z)\| = \|g(z) - g(w)\| \leq \frac{1}{2}\|z - w\|,$$

então

$$\frac{1}{2}\|w - z\| \leq \|f(w) - f(z)\|,$$

donde

$$\|z - w\| \leq 2\|f(w) - f(z)\|$$

e, portanto, $f|_{\tilde{U}}^{-1}$ é 2-Lipschitz, logo contínua. Isto prova que $f : \tilde{U} \to B(0, \frac{r}{2})$ é homeomorfismo. Como f é de classe C^1, a Proposição 4.1.1 garante que existe U' vizinhança aberta de 0 tal que $f|_{U'}$ é difeomorfismo de classe C^1 sobre sua imagem.

Para o caso geral, o mapa ψ definido numa vizinhança de 0_E por

$$\psi(x) := f'(a)^{-1}(f(a+x) - f(a))$$

tem sua imagem contida em E, é de classe C^1 e é tal que $\psi(0) = 0$ e $\psi'(0) = \mathrm{Id}_E$. Do que vimos, existem vizinhanças abertas U e V da origem tais que $\psi : U \to V$ é um difeomorfismo de classe C^1. Para a restrição de f a

$$U' := a + U$$

temos

$$f|_{U'}(x) = f(a) + f'(a)\psi(x-a)$$

e $f|_{U'}$ é uma bijeção sobre

$$W := f(a) + f'(a)V.$$

Como a inversa de $f|_{U'}$ pode ser escrita como

$$f|_{U'}^{-1}(y) = a + \psi^{-1}(f'(a)^{-1}(y - f(a))),$$

então $f|_{U'}^{-1}$ é de classe C^1. Por conseguinte, $f|_{U'}$ é um difeomorfismo de classe C^1. $\square$

4.2 O Teorema da Aplicação Implícita

Muitas vezes as variáveis estão relacionadas por uma equação do tipo $F(x, y) = 0$ e não é possível ou é muito difícil isolar y como uma função explícita de x, na forma $y = f(x)$. O teorema abaixo, extraído de (COLEMAN, 2012, Theorem 8.2, p. 166), estabelece as condições para que, localmente, essa equação defina y como uma função de x.

Relembre a Proposição 1.7.1.

Teorema 4.2.1 (Teorema da Aplicação Implícita)**.** Sejam E_1, E_2 e F espaços de Banach, $U \subseteq E_1 \times E_2$ um subconjunto aberto e $f : U \to F$ um mapa C^1. Suponha que $c \in F$ e que o conjunto S dos pares $(x, y) \in U$ satisfazendo

$$f(x, y) = c$$

é não vazio. Se $(a, b) \in S$ e a diferencial parcial $\partial_2 f(a, b) : E_2 \to F$ é invertível, então existe uma vizinhança aberta U' de (a, b), com $U' \subseteq U$, uma vizinhança V de a em E_1 e um mapa $\varphi : V \to E_2$ de classe C^1 tal que as seguintes afirmações são equivalentes:

1. $(x, y) \in U'$ e $f(x, y) = c$;

2. $x \in V$ e $y = \varphi(x)$.

Demonstração. Defina $g : U \to E_1 \times F$ pondo

$$g(x, y) = (x, f(x, y)).$$

O mapa g é de classe C^1, com

$$g'(x, y)(u, v) = (u, f'(x, y)(u, v)) = (u, \partial_1 f(x, y)u + \partial_2 f(x, y)v).$$

Em particular, em (a, b), temos

$$g'(a, b)(u, v) = (u, f'(a, b)(u, v)) = (u, \partial_1 f(a, b)u + \partial_2 f(a, b)v).$$

Para simplificar a notação, seja $T := g'(a, b)$. Queremos encontrar a inversa T^{-1}. Vamos encontrar h e k tais que

$$T(h, k) = (u, w).$$

Isso nos dá $h = u$ e $\partial_1 f(a, b)h + \partial_2 f(a, b)k = w$. Logo,

$$\partial_2 f(a, b)k = w - \partial_1 f(a, b)h.$$

Por hipótese, $\partial_2 f(a, b)$ é invertível, o que nos dá

$$k = \partial_2 f(a, b)^{-1}(w - \partial_1 f(a, b)h).$$

Logo, $T = g'(a, b)$ é invertível, com

$$g'(a, b)^{-1}(u, w) = T^{-1}(u, w) = (u, \partial_2 f(a, b)^{-1}(w - \partial_1 f(a, b)u)).$$

Aplicando o Teorema da Aplicação Inversa (ver Teorema 4.1.2), obtemos uma vizinhança U' de (a, b), com $U' \subseteq U$, tal que $g|_{U'}$ é um difeomorfismo de classe C^1 sobre sua imagem W. Seja $h : W \to U'$ a sua inversa (também de classe C^1). Então, para $(x, y) \in U'$, temos

$$(x, y) = (h \circ g)(x, y) = h(x, f(x, y)) = (x, h_2(x, z)),$$

onde $h_2 := \pi_2 \circ h$, sendo $\pi_2 : E_1 \times E_2 \to E_2$ a projeção na segunda coordenada, e

$$z := f(x, y).$$

Como h é de classe C^1, então h_2 também é de classe C^1 (a prova disso é semelhante à da Proposição 1.7.2). Ponhamos agora

$$V := \{x \in E_1 : (x, c) \in W\}.$$

V é a imagem inversa de W pela aplicação inclusão $i_c : E_1 \to E_1 \times F$ dado por

$$x \mapsto (x, c)$$

e, portanto, é aberto. Claramente, $a \in V$. O mapa $\varphi = h_2 \circ i_c$ de V em E_2, sendo a composição de mapas de classe C^1, é de classe C^1 (ver Corolário 1.3.2) e satisfaz as sentenças equivalentes do teorema.

De fato, se $(x, y) \in U'$ e $f(x, y) = c$, então $g(x, y) = (x, c)$, o que implica $(x, c) \in W$, ou seja, $x \in V$. Como

$$(x, y) = h(x, c) = (x, \varphi(x)),$$

então $\varphi(x) = y$. Reciprocamente, se $x \in V$ e $\varphi(x) = y$, então

$$(x, y) = (x, h_2(x, c)) = h(x, c) \in U'.$$

Isso nos dá

$$(x, f(x, y)) = g(x, y) = (x, c),$$

donde vem $f(x, y) = c$. $\square$

Parte II

Otimização Contínua

5 Condições de Otimalidade

Sejam X um espaço topológico e $f : X \to \mathbb{R}$ uma função. Dizemos que $a \in X$ é um

1. **minimizador local** de f se existe uma vizinhança $V \subseteq X$ de a tal que $f(a) \leq f(x)$, para todo $x \in V$.
2. **minimizador global** de f se $f(a) \leq f(x)$, para todo $x \in X$.
3. **maximizador local** de f se existe uma vizinhança $V \subseteq X$ de a tal que $f(x) \leq f(a)$, para todo $x \in V$.
4. **maximizador global** de f se $f(x) \leq f(a)$, para todo $x \in X$.
5. **extremo local** de f se a é um minimizador ou um maximizador local de f.
6. **extremo global** de f se a é um minimizador ou um maximizador global de f.
7. **ponto de sela** de f se a não é um extremo local de f.

Podemos dizer então que o problema geral da otimização é o de encontrar extremos locais para uma função $f : X \to \mathbb{R}$, onde X é um espaço topológico. Como o leitor provavelmente já sabe, não existe nenhum método tão geral capaz de nos dizer sempre quando um problema nesses moldes possui ou não solução, nem nenhum método que, mesmo quando se sabe da existência de uma solução, nos permita sempre encontrá-la.

Desse modo, é costumeiro, na literatura de Otimização Contínua (consulte qualquer livro da área), quebrar o problema geral em classes de problemas menores, menos gerais, geralmente munindo o espaço e a função nele definidos de propriedades adicionais, propriedades que costumam aparecer em muitos problemas reais, em modelos reais. Neste capítulo abordaremos um pouco sobre duas classes de problemas extremamente importantes: o problema geral de **Programação não Linear** (PnãoL) e o problema geral de **Otimização Convexa** (OC).

Antes, no entanto, cabe uma pequena observação: não há perda de generalidade em lidar apenas com os problemas de minimização de funções. Com efeito, se X é um espaço topológico e $\overline{x} \in X$ é um maximizador local de uma função $f : X \to \mathbb{R}$, então, por definição, existe $V \subseteq X$ vizinhança de $\overline{x}$ tal que

$$f(x) \leq f(\overline{x}), \quad \text{para todo } x \in V,$$

donde

$$-f(\overline{x}) \leq -f(x), \text{ para todo } x \in V \implies \overline{x} \text{ é minimizador local de } -f.$$

Com base nessa observação, ainda que enunciemos os teoremas aqui abarcando inclusive os casos de maximização, apenas nos preocuparemos em demonstrar os casos envolvendo minimização.

5.1 Condições necessárias

Todas as condições de otimalidade que abordaremos nesse trabalho dependem do fato de que um extremo local de uma função diferenciável sempre é um *ponto estacionário.*

Definição 5.1.1 (Ponto estacionário)**.** Se E é um espaço normado, $U \subseteq E$ é um subconjunto aberto e $f : U \to \mathbb{R}$ é uma função diferenciável, dizemos que $c \in U$ é um **ponto estacionário** (ou **ponto crítico**) de f se $f'(c) = 0_{E^*}$.

Note que se E é um espaço de Hilbert, então

$$f'(a) = 0_{E^*} \iff \nabla f(a) = 0_E.$$

Com efeito, se a é um ponto estacionário de f, temos

$$0 = f'(a)\left(\nabla f(a)\right) = \langle \nabla f(a), \nabla f(a) \rangle = \|\nabla f(a)\|^2 \implies \nabla f(a) = 0_E.$$

Reciprocamente, como, para todo $h \in E$,

$$\nabla f(a) = 0_E \implies f'(a)h = \langle h, \nabla f(a) \rangle = \langle h, 0_E \rangle = 0,$$

então

$$f'(a) = 0_{E^*}.$$

Teorema 5.1.2 (Condição de otimalidade necessária de 1^{a} ordem)**.** Se E é um espaço normado, $U \subseteq E$ é um subconjunto aberto, $f : U \to \mathbb{R}$ é uma função diferenciável em $a \in U$ e a é um extremo local de f, então a é um ponto estacionário de f.

Demonstração. Se a é um minimizador local de f, então existe $\delta > 0$ tal que

$$f(a) \leq f(x), \quad \text{para todo } x \in B(a, \delta) \subseteq U.$$

Dado $h \in E \setminus \{0\}$, como f é diferenciável em a, então

$$f(a + th) = f(a) + tf'(a)h + r(th),$$

para todo $t \in \mathbb{R}$ tal que $a + th \in B(a, \delta)$. Assim,

$$0 < |t| < \frac{\delta}{\|h\|_E} \implies f'(a)h + \frac{r(th)}{t} = \frac{f(a + th) - f(a)}{t}.$$

Como
$$\frac{r(th)}{t} = \|h\|_E \frac{r(th)}{t\|h\|_E} = \pm\|h\|_E \frac{r(th)}{\|th\|_E} \xrightarrow{t\to 0} 0,$$
então, pela Monotonicidade do Limite (ver Teorema B.5.8),
$$0 \geq \lim_{t\to 0^-} \frac{f(a+th)-f(a)}{t} = f'(a)h = \lim_{t\to 0^+} \frac{f(a+th)-f(a)}{t} \geq 0,$$
donde
$$f'(a)h = 0.$$
Portanto,
$$f'(a) = 0_{E^*},$$
isto é, a é um ponto estacionário de f. □

Como se sabe, a recíproca do teorema anterior não é válida em geral. A função $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) := x^3$ é o contraexemplo clássico: f é derivável e possui 0 como seu único ponto estacionário, mas, como se constata facilmente, 0 é um ponto de sela de f.

Definição 5.1.3. Se E é um espaço normado e $T \in \mathcal{L}_n(E, \mathbb{R})$, dizemos que T é

- **definida positiva** se $Tx^{(n)} > 0$, para todo $x \in E \setminus \{0\}$.
- **definida negativa** se $-T$ é definida positiva.
- **definida** se T é definida positiva ou definida negativa.
- **semidefinida positiva** se $Tx^{(n)} \geq 0$, para todo $x \in E$.
- **semidefinida negativa** se $-T$ é semidefinida positiva.
- **semidefinida** se T é semidefinida positiva ou semidefinida negativa.
- **indefinida** se existem $u, v \in E$ tais que $Tu^{(n)} > 0$ e $Tv^{(n)} < 0$.

Note que se $T \in \mathcal{L}_n(E, \mathbb{R})$ é definida, então T é semidefinida.

As mesmas definições se aplicam a uma matriz quadrada A através da forma bilinear
$$T(x, y) := x^t A y,$$
isto é, a matriz A é (semi)definida/indefinida positiva/negativa se a forma bilinear T é (semi)definida/indefinida positiva/negativa.

Observação 5.1.4. Mesmo sendo $T \in \mathcal{L}_n(E, \mathbb{R})$ simétrica e definida positiva, pode-se ter
$$\inf_{\|x\|_E = 1} T \cdot x^{(n)} = 0.$$

O exemplo que daremos utilizará os espaços de sequências $\ell^p(\mathbb{N})$. (Veja os exemplos C.1.4, C.8.6, C.8.7 e Proposição C.3.8). Considere o espaço

$$\ell_1 := \left\{(x_n) \subseteq \mathbb{R} : \sum_{n=1}^{\infty} |x_n| < +\infty\right\}$$

munido da norma

$$\|x\|_{\ell_1} := \sum_{n=1}^{\infty} |x_n|, \quad \text{para todo } x = (x_n) \in \ell_1,$$

e do produto interno

$$\langle x, y\rangle := \sum_{n=1}^{\infty} |x_n, y_n|, \quad \text{para quaisquer } x = (x_n), y = (y_n) \in \ell_1.$$

Ponha, para quaisquer $k, n \in \mathbb{N}$,

$$y_n^k := \begin{cases} \dfrac{1}{k}, & \text{se} \quad n = 1, \ldots, k, \\ 0, & \text{caso contrário,} \end{cases}$$

e

$$y^k := \left(y_m^k\right)_m = \left(\underbrace{\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k}}_{k \text{ vezes}}, 0, \ldots, 0, \ldots\right).$$

Nessa condições, para cada $k \in \mathbb{N}$, temos

$$\sum_{n=1}^{\infty} |y_n^k| = \sum_{n=1}^{\infty} y_n^k = \sum_{n=1}^{k} y_n^k + \sum_{n=k+1}^{\infty} y_n^k = \sum_{n=1}^{k} \frac{1}{k} + \sum_{n=k+1}^{\infty} 0 = k \cdot \frac{1}{k} + 0 = 1 < +\infty,$$

donde

$$y^k \in \ell_1 \quad \text{e} \quad \|y^k\|_{\ell_1} = 1,$$

com

$$\langle y_k, y_k\rangle = \sum_{n=1}^{k} \frac{1}{k^2} = \frac{k}{k^2} = \frac{1}{k}.$$

Logo, $\langle \cdot, \cdot\rangle \in \mathcal{L}_2(\ell_1, \mathbb{R})$ é simétrico, definido positivo e

$$\inf_{\|x\|_{\ell_1}=1} \langle x, x\rangle = 0.$$

Proposição 5.1.5. Sejam E um espaço normado e $T \in \mathcal{L}_k(E, \mathbb{R})$.

- Se k é ímpar e T é semidefinida, então $T = 0_{\mathcal{L}_k(E,\mathbb{R})}$. Consequentemente,

$$k \quad \text{é ímpar} \implies T \quad \text{não é definida.}$$

- Se existe $\alpha > 0$ tal que

$$Th^{(k)} \geq \alpha \|h\|_E^k, \quad \text{para todo} \quad h \in E, \tag{5.1}$$

então T é definida positiva e, por conseguinte, k é par.

Demonstração. Se k é ímpar e T é semidefinida positiva, então, dado $v \in E$,

$$0 \leq T \cdot (-v)^{(k)} = -T \cdot v^{(k)} \leq 0,$$

donde

$$T \cdot v^{(k)} = 0.$$

De (3.2), temos

$$T = 0_{\mathcal{L}_k(E,\mathbb{R})}.$$

Por outro lado, se T é semidefinida negativa, então $-T$ é semidefinida positiva e, portanto, $-T = 0_{\mathcal{L}_k(E,\mathbb{R})}$, o que nos dá

$$T = 0_{\mathcal{L}_k(E,\mathbb{R})}.$$

Suponhamos agora que existe $\alpha > 0$ satisfazendo a desigualdade (5.1). Daí,

$$v \in E \setminus \{0_E\} \implies \|v\|_E > 0 \implies \|v\|_E^k > 0 \implies Tv^{(k)} \geq \alpha \|v\|_E^k > 0.$$

Logo,

$$T \text{ é definida positiva} \implies k \text{ é par},$$

como consequência do item anterior. □

Teorema 5.1.6 (Condição de otimalidade necessária de k-ésima ordem)**.** Sejam E um espaço normado, $U \subseteq E$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função k vezes diferenciável em $a \in U$, com $k \geq 2$, tal que

$$\begin{cases} f^{(j)}(a) = 0_{\mathcal{L}_j(E,\mathbb{R})}, & \text{para} \quad j = 1, \dots, k-1; \\ f^{(k)}(a) \neq 0_{\mathcal{L}_k(E,\mathbb{R})}. \end{cases}$$

Nessas condições,

$$\begin{cases} a \text{ é um minimizador local de } f \implies f^{(k)}(a) \text{ é semidefinida positiva;} \\ a \text{ é um maximizador local de } f \implies f^{(k)}(a) \text{ é semidefinida negativa.} \end{cases}$$

Demonstração. Se a é um minimizador local de f, então existe $\delta_1 > 0$ tal que

$$\|s\|_E < \delta_1 \implies f(a+s) \geq f(a) \implies f(a+s) - f(a) \geq 0.$$

Pela Fórmula de Taylor com resto infinitesimal (ver Teorema 3.7.2), dado $h \in E \setminus \{0_E\}$,

$$\begin{aligned} 0 < t < \delta \implies & \frac{t^k}{k!}\left[f^{(k)}(a)h^{(k)} + k!\frac{r(th)}{t^k}\right] = f(a+th) - f(a) \geq 0 \\ \implies & f^{(k)}(a)h^{(k)} + k!\frac{r(th)}{t^k} = k!\frac{f(a+th) - f(a)}{t^k} \geq 0, \end{aligned}$$

onde $\delta := \dfrac{\delta_1}{\|h\|_E}$. Do fato de que

$$k!\frac{r(th)}{t^k} = k!\|h\|_E^k \frac{r(th)}{t^k\|h\|_E^k} = k! \cdot \left[\pm\|h\|_E^k \frac{r(th)}{\|th\|_E^k}\right] \xrightarrow{t\to 0} 0,$$

temos

$$f^{(k)}(a)h^{(k)} \geq 0,$$

pela Monotonicidade do Limite (ver Teorema B.5.8).

Logo, $f^{(k)}(a)$ é semidefinida positiva. □

Na literatura de Otimização que utilizamos, como (RIBEIRO; KARAS, 2013, pp. 40, 42) e (IZMAILOV; SOLODOV, 2009, pp. 17, 18), são apresentadas apenas condições de otimalidade no espaço $\mathbb{R}^n$ e até 2ª ordem, sendo os critérios de 2ª ordem enunciados com o auxílio da matriz hessiana (ver Teorema 3.3.2).

Corolário 5.1.7 (Condição de otimalidade necessária de 2ª ordem)**.** Se $U \subseteq \mathbb{R}^n$ é um conjunto aberto e $f : U \to \mathbb{R}$ é uma função duas vezes diferenciável em um extremo local $a \in U$ de f, então $\mathrm{H}\, f(a)$ é semidefinida. Mais especificamente,

$$\begin{cases} a \text{ é um minimizador local de } f \implies \mathrm{H}\, f(a) \text{ é semidefinida positiva;} \\ a \text{ é um maximizador local de } f \implies \mathrm{H}\, f(a) \text{ é semidefinida negativa.} \end{cases}$$

Demonstração. Se a é um minimizador local de f, então a é um ponto estacionário de f, pelo Teorema 5.1.2, donde $f^{(2)}(a)$ é semidefinida positiva, pelo Teorema 5.1.6, e, portanto, pelo Exemplo 3.3.2,

$$0 \leq f^{(2)}(a)h^{(2)} = \langle h, \nabla^2 f(a)h\rangle = h^t\, \mathrm{H}\, f(a)h, \quad \text{para todo } h \in \mathbb{R}^n\,.$$

Logo, $\mathrm{H}\, f(a)$ é semidefinida positiva. □

5.2 Condições suficientes

Teorema 5.2.1 (Condição de otimalidade suficiente de k-ésima ordem)**.** Sejam E um espaço normado, $U \subseteq E$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função k vezes diferenciável em $a \in U$, com $k \geq 2$, tal que

$$\begin{cases} f^{(j)}(a) = 0_{\mathcal{L}_j(E,F)}, \text{ para } j = 1, \ldots, k-1; \\ f^{(k)}(a) \neq 0_{\mathcal{L}_k(E,F)}. \end{cases}$$

- Se existe $\alpha > 0$ satisfazendo (5.1) com $T = f^{(k)}(a)$, então k é par, $f^{(k)}(a)$ é definida positiva (ver Corolário 5.1.5) e a é minimizador local estrito de f.

- Se k é ímpar, então a é um ponto de sela de f.

Demonstração. Suponhamos que existe $\alpha > 0$ satisfazendo (5.1) com $T = f^{(k)}(a)$. Pela Fórmula de Taylor com resto infinitesimal (ver Teorema 3.7.2), existe $r : E \to \mathbb{R}$ tal que

$$\lim_{s\to 0} \frac{r(s)}{\|s\|_E^k} = 0 \tag{5.2}$$

e, dado $s \in E$ tal que $a + s \in U$, temos

$$f(a+s) - f(a) = \frac{\|s\|_E^k}{k!}\left[f^{(k)}(a)\left(\frac{s}{\|s\|_E}\right)^{(k)} + k!\frac{r(s)}{\|s\|_E^k}\right].$$

Como

$$f^{(k)}(a)\left(\frac{s}{\|s\|_E}\right)^{(k)} \geq \alpha,$$

para todo $s \in E \setminus \{0_E\}$, então α é uma cota inferior de $\varphi(S_1(0_E))$, onde

$$\varphi : E \to \mathbb{R}$$

é a função dada por

$$\varphi(h) := f^{(k)}(a)h^{(k)},$$

para cada $h \in E$. Portanto,

$$u := \inf \varphi(S_1(0_E)) \geq \alpha > 0.$$

De (5.2), sabemos que existe $\delta > 0$ tal que, para todo $s \in E$ com $0 < \|s\|_E < \delta$, temos

$$-\frac{r(s)}{\|s\|_E^k} \leq \frac{|r(s)|}{\|s\|_E^k} < \frac{u}{2k!},$$

donde

$$k!\frac{r(s)}{\|s\|_E^k} > -\frac{u}{2}$$

e, portanto,

$$f(a+s) - f(a) \geq \frac{\|s\|_E^k}{k!}\left(u - \frac{u}{2}\right) \geq \frac{\|s\|_E^k}{k!}\frac{u}{2} > 0.$$

Logo,

$$f(a+s) > f(a).$$

Isto prova que a é minimizador local estrito de f.

Suponhamos agora, por redução ao absurdo, que k é ímpar e a é um extremo local de f. Nessas condições, $f^{(k)}(a)$ é semidefinida, pelo Teorema 5.1.6, donde

$$f^{(k)}(a) = 0_{\mathcal{L}_k(E,\mathbb{R})},$$

pelo Corolário 5.1.5, o que é absurdo. □

Corolário 5.2.2 (Condição de otimalidade suficiente em dimensão finita)**.** Sejam E um espaço normado de dimensão finita, $U \subseteq E$ um subconjunto aberto e $f : U \to \mathbb{R}$ uma função k vezes diferenciável em $a \in U$ tal que

$$\begin{cases} f^{(j)}(a) = 0_{\mathcal{L}_j(E,\mathbb{R})}, \text{ para } j = 1, \dots, k-1; \\ f^{(k)}(a) \neq 0_{\mathcal{L}_k(E,\mathbb{R})}. \end{cases}$$

- Se $f^{(k)}(a)$ é definida positiva, então a é um minimizador local estrito de f.
- Se $f^{(k)}(a)$ é definida negativa, então a é um maximizador local estrito de f.
- Se $f^{(k)}(a)$ é indefinida, então a é um ponto de sela de f.

Demonstração. Como a função $\varphi : E \to \mathbb{R}$ dada por

$$\varphi(h) := f^{(k)}(a)h^{(k)},$$

para cada $h \in E$, é contínua e $X := S_1(0_E)$ é compacto, então existe $u \in X$ tal que $\varphi(u) = \min \varphi(X)$, pelo Teorema de Weierstrass (ver Corolário B.4.9). Se $f^{(k)}(a)$ é definida positiva, então k é par, pelo Corolário 5.1.5, e

$$\varphi(u) = f^{(k)}(a)u^{(k)} > 0.$$

Daí, dado $h \in E \setminus \{0_E\}$, temos

$$f^{(k)}(a)\left(\frac{h}{\|h\|_E}\right)^{(k)} = \varphi\left(\frac{h}{\|h\|_E}\right) \geq \varphi(u),$$

donde

$$f^{(k)}(a)h^{(k)} \geq \varphi(u)\|h\|_E^k.$$

Logo, a é um minimizador local estrito de f, pelo Teorema 5.2.1.

Se $f^{(k)}(a)$ é indefinida, então $f^{(k)}(a)$ não é semidefinida, e, portanto, a não pode ser nem minimizador local de f, nem maximizador local de f, pelo Teorema 5.1.6. □

A condição suficiente de 2ª ordem também se expressa em termos da matriz hessiana de acordo com o corolário abaixo, retirado de (RIBEIRO; KARAS, 2013, p. 42).

Corolário 5.2.3 (Condição de otimalidade suficiente de 2ª ordem)**.** Se $U \subseteq \mathbb{R}^n$ é um subconjunto aberto, $f : U \to \mathbb{R}$ é uma função duas vezes diferenciável em um ponto estacionário a e $\mathrm{H}\, f(a)$ é definida positiva, então a é minimizador local estrito de f.

Demonstração. Pelo Exemplo 3.3.2, temos

$$f^{(2)}(a)h^{(2)} = \langle h, \nabla^2 f(a)h \rangle = h^t \,\mathrm{H}\, f(a)h > 0, \quad \text{para todo } h \in \mathbb{R}^n \setminus \{0\},$$

donde $f^{(2)}(a)$ é definida positiva. Logo, a é minimizador local de f, pelo Corolário 5.2.2. □

6 Análise Convexa

6.1 Conjuntos convexos

Definição 6.1.1. Seja E um espaço vetorial.

- O **segmento de reta** de extremos em $u, v \in E$ é o conjunto

$$[u, v] := \{(1-t)u + tv : 0 \leq t \leq 1\}.$$

Definimos também o segmento aberto à direita

$$[u, v) := [u, v] \setminus \{v\},$$

o segmento aberto à esquerda

$$(u, v] := [u, v] \setminus \{u\}$$

e o segmento aberto

$$(u, v) := [u, v] \setminus \{u, v\}.$$

- Dizemos que $X \subseteq E$ é **convexo** se $[u, v] \subseteq X$, para quaisquer $u, v \in X$.

Exemplo 6.1.2. Todo espaço vetorial é um conjunto convexo.

Exemplo 6.1.3. Segmentos de reta são conjuntos convexos.

Exemplo 6.1.4. Todas as bolas de um espaço vetorial normado são convexas.

Demonstração. Sejam E um espaço vetorial normado e $a \in E$.

Se $r > 0$ e $x, y \in B[a, r]$, então

$$\begin{aligned}\|(1-t)x + ty - a\| = \|(1-t)x + ty - [(1-t)a + ta]\| &\leq (1-t)\|x - a\| + t\|y - a\| \\ &\leq (1-t)r + tr = r,\end{aligned}$$

donde $[x, y] \subseteq B[a, r]$ e, portanto, $B[a, r]$ é um conjunto convexo.

Logo, toda bola fechada é convexa.

Se $r > 0$ e $x, y \in B(a, r)$, então

$$[x, y] \subseteq B[a, s] \subseteq B(a, r), \quad \text{onde } s := \max\{\|x - a\|, \|y - a\|\}.$$

Portanto, toda bola aberta é convexa. $\square$

Teorema 6.1.5. Sejam V e W espaços vetoriais e seja $f : V \to W$ uma *transformação afim*, isto é, $f = L + w_0$, para alguma aplicação linear $L : V \to W$ e algum $w_0 \in W$. Se $C \subseteq W$ é convexo, então $D := f^{-1}(C) \subseteq V$ é convexo.

Demonstração. Sejam $a, b \in D$ e $t \in [0, 1]$. Então $La + w_0, Lb + w_0 \in C$. Daí,

$$L((1-t)a + tb) + w_0 = (1-t)(La + w_0) + t(Lb + w_0) \in C,$$

o que termina a prova. □

Corolário 6.1.6. Se E é um espaço normado, $C \subseteq E$ é um conjunto convexo e $x, y \in E$, então

$$D := \{t \in \mathbb{R} : x + ty \in C\}$$

é um conjunto convexo. Além disso, se C é aberto, então D também é.

Demonstração. Defina $f(t) := ty + x$. Como $L(t) := ty$ é uma aplicação linear, então f é uma transformação afim. Logo, pelo Exemplo 6.1.5, o conjunto $D = f^{-1}(C)$ é convexo. Além disso, como L é contínua, então f é contínua, logo, se C é aberto, então D também é, como pré-imagem de um aberto por uma função contínua. □

6.2 Funções convexas

Definição 6.2.1 (Função convexa)**.** Se E é um espaço vetorial e $C \subseteq E$ é um conjunto convexo, dizemos que uma função $f : C \to \mathbb{R}$ é uma **função convexa** se

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y), \quad \text{para quaisquer} \quad x, y \in C, \alpha \in [0, 1].$$

Exemplo 6.2.2. A função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) := x^2$ é convexa.

Demonstração. Dados $x, y \in \mathbb{R}$ e $\lambda \in [0, 1]$, temos

$$\begin{aligned} f((1-\lambda)x + \lambda y) = [(1-\lambda)x + \lambda y)]^2 &= [x + \lambda(y-x)]^2 \\ &= x^2 + 2\lambda x(y-x) + \lambda^2(y-x)^2 \\ &\leq x^2 + 2\lambda x(y-x) + \lambda(y-x)^2 \\ &\leq (1-\lambda)x^2 + \lambda y^2 - 2\lambda x^2 \\ &\leq (1-\lambda)f(x) + \lambda f(y). \end{aligned}$$

Logo, f é uma função convexa. □

Exemplo 6.2.3. As normas são funções convexas.

Demonstração. Sejam E um espaço vetorial, $\|\cdot\|$ uma norma em E e $x, y \in E$ e $\lambda \in [0,1]$. Daí,

$$\|(1-\lambda)x + \lambda y)\| \leq (1-\lambda)\|x\| + \lambda\|y\|.$$

Logo, $\|\cdot\|$ é uma função convexa. □

Exemplo 6.2.4. Se E é um espaço com produto interno $\langle \cdot, \cdot \rangle$ e $w \in E$, então a função $f : E \to \mathbb{R}$ dada por $f(x) := \langle w, x \rangle$ é uma função convexa.

Demonstração. Se $x, y \in E$ e $\lambda \in [0,1]$, então

$$f((1-\lambda)x + \lambda y) = \langle w, (1-\lambda)x + \lambda y \rangle = (1-\lambda)\langle w, x \rangle + \lambda \langle w, y \rangle = (1-\lambda)f(x) + \lambda f(y).$$

Logo, f é uma função convexa. □

Definição 6.2.5 (Epígrafo)**.** Se E é um espaço vetorial e $f : X \subseteq E \to \mathbb{R}$ é uma função, o **epígrafo** de f é o conjunto

$$\text{Epi}(f) := \{(x, \gamma) \in X \times \mathbb{R} : f(x) \leq \gamma\}.$$

Teorema 6.2.6. Se E é um espaço vetorial e $C \subseteq E$ é um subconjunto convexo, então $f : C \to \mathbb{R}$ é uma função convexa se, e somente se, $\text{Epi}(f)$ é convexo em $E \times \mathbb{R}$.

Demonstração. Se $x := (a, \lambda), y := (b, \alpha) \in \text{Epi}(f)$ e $(c, \gamma) \in [x, y]$, então existe $t \in [0,1]$ tal que

$$(c, \gamma) = (1-t)x + ty = (1-t)(a, \lambda) + t(b, \alpha) = ((1-t)a + tb, (1-t)\lambda + t\alpha).$$

Como

$$f(c) = f((1-t)a + tb) \leq (1-t)f(a) + tf(b) \leq (1-t)\lambda + t\alpha = \gamma,$$

então $(c, \gamma) \in \text{Epi}(f)$, donde $[x, y] \subseteq \text{Epi}(f)$.

Portanto, $\text{Epi}(f)$ é um conjunto convexo em $E \times \mathbb{R}$.

Reciprocamente, suponhamos que $\text{Epi}(f)$ é convexo em $E \times \mathbb{R}$ e considere $a, b \in C$ e $\alpha \in [0,1]$. Como $(a, f(a)), (b, f(b)) \in \text{Epi}(f)$, então

$$X := [(a, f(a)), (b, f(b))] \subseteq \text{Epi}(f),$$

donde

$$x := (1-\alpha)(a, f(a)) + \alpha(b, f(b)) \in X$$

e, portanto, $x \in \text{Epi}(f)$. Logo,

$$f((1-\alpha)a + \alpha b) \leq (1-\alpha)f(a) + \alpha f(b),$$

já que

$$x = ((1-\alpha)a + \alpha b, (1-\alpha)f(a) + \alpha f(b)).$$

Portanto, f é uma função convexa. □

Teorema 6.2.7 (Definição equivalente de convexidade na reta)**.** Sejam $I \subseteq \mathbb{R}$ um intervalo de números reais e $f : I \to \mathbb{R}$ uma função. Assim, f é convexa se, e somente se, para quaisquer $a, b, x \in I$, com $a < x < b$, vale

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b).$$

Demonstração. Suponhamos válida a desigualdade. Como $\dfrac{b-x}{b-a}$ e $\dfrac{x-a}{b-a}$ são não negativos e

$$\frac{b-x}{b-a} + \frac{x-a}{b-a} = 1,$$

então

$$f(x) \leq \left(1 - \frac{x-a}{b-a}\right) f(a) + \frac{x-a}{b-a} f(b).$$

Tomando

$$\lambda := \frac{x-a}{b-a} \in (0,1),$$

uma vez que $a < x < b$, e

$$x = (1-\lambda)a + \lambda b,$$

donde

$$f((1-\lambda)a + \lambda b) \leq (1-\lambda) f(a) + \lambda f(b).$$

Como $a, b \in I$, então f é convexa. A prova da recíproca é a mesma só refazendo os passos no sentido contrário. $\square$

Teorema 6.2.8 (Inclinações das secantes)**.** Sejam $I \subseteq \mathbb{R}$ um intervalo de números reais e $f : I \to \mathbb{R}$ uma função convexa. Então, para quaisquer $a, b, x \in I$, com $a < x < b$, vale

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x}.$$

Demonstração. Como f é convexa, então, pelo Teorema 6.2.7, temos

$$f(x) \leq f(a) + \frac{f(b) - f(a)}{b-a}(x-a)$$

sempre que $a, b, x \in I$ com $a < x < b$. Como $x - a > 0$, então

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a}.$$

Além disso, como

$$x - b + b - a = x - a,$$

então

$$\begin{aligned} f(x) &\leq f(a) + \frac{f(b) - f(a)}{b-a}(x - b + b - a) \\ &\leq f(a) + \frac{f(b) - f(a)}{b-a}(x-b) + f(b) - f(a), \end{aligned}$$

donde

$$f(x) \leq f(b) + \frac{f(b) - f(a)}{b - a}(x - b).$$

O fato de que $x - b < 0$ nos dá

$$\frac{f(x) - f(b)}{x - b} \geq \frac{f(b) - f(a)}{b - a},$$

como queríamos demonstrar. □

Teorema 6.2.9. Se I é um intervalo limitado de números reais de extremos a e b, com $a < b$, e $f : I \to \mathbb{R}$ é uma função convexa, então f é contínua no interior de I.

Demonstração. Seja $x \in \text{int}(I)$. Como $\text{int}(I) = (a, b)$ é um conjunto aberto, então existe $\delta_1 > 0$ tal que

$$(x - \delta_1, x + \delta_1) \subseteq (a, b),$$

donde

$$[x - \delta, x + \delta] \subseteq (a, b), \text{ onde } \delta := \frac{\delta_1}{2}.$$

Se $y_1, y_2 \in \mathbb{R}$ são tais que

$$x - \delta < y_1 < x < y_2 < x + \delta,$$

então, do Teorema 6.2.8, temos

$$\frac{f(x) - f(x - \delta)}{\delta} \leq \frac{f(x) - f(y_1)}{x - y_1} \leq \frac{f(y_2) - f(x)}{y_2 - x} \leq \frac{f(x + \delta) - f(x)}{\delta}.$$

Tomando

$$\beta := \max\{f(x - \delta) - f(x), f(x + \delta) - f(x)\},$$

então, para todo $y \in (x - \delta, x + \delta)$, temos

$$\frac{-\beta}{\delta} \leq \left| \frac{f(x) - f(y)}{x - y} \right| \leq \frac{\beta}{\delta} \implies |f(x) - f(y)| \leq \frac{\beta}{\delta}|x - y|.$$

Logo, a restrição de f a (a, b) é localmente lipschitziana e, portanto, é contínua. □

Teorema 6.2.10 (Em relação às tangentes)**.** Se I é um intervalo de números reais e $f : I \to \mathbb{R}$ é uma função derivável no interior de I, então as seguintes afirmações são equivalentes:

a) f é convexa;

b) $\dot{f}$ é uma função não decrescente;

c) a reta tangente ao gráfico de f, em qualquer ponto de I, fica abaixo do gráfico de f.

Demonstração. Supondo f convexa, então, do Teorema 6.2.8 e da Monotonicidade do Limite (ver Teorema A.1.11), temos que, para quaisquer $a, b, x \in \text{int}(I)$, com $a < x < b$,

$$\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \implies \dot{f}(a) = \lim_{x\to a^+} \frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a}$$

e

$$\frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x} \implies \frac{f(b)-f(a)}{b-a} \leq \lim_{x\to b^-} \frac{f(b)-f(x)}{b-x} = \dot{f}(b).$$

Logo,

$$a < b \implies \dot{f}(a) \leq \dot{f}(b),$$

Agora, suponhamos b) e provemos c). Dados $a, x \in \text{int}(I)$ distintos, tem-se que $a < x$ ou $x < a$.

Se $a < x$, então, pelo Teorema do Valor Médio de Lagrange (ver Corolário 2.1.4), existe $c \in (a, x)$ tal que

$$\dot{f}(c) = \frac{f(x)-f(a)}{x-a},$$

donde

$$f(x) = f(a) + \dot{f}(c)(x-a) \geq f(a) + \dot{f}(a)(x-a),$$

já que

$$a < c \implies \dot{f}(a) \leq \dot{f}(c).$$

Se $x < a$, então, novamente pelo Teorema do Valor Médio de Lagrange, existe $d \in (x, a)$ tal que

$$\dot{f}(d) = \frac{f(a)-f(x)}{a-x},$$

donde

$$f(x) = f(a) - \dot{f}(d)(a-x) \geq f(a) - \dot{f}(a)(a-x) = f(a) + \dot{f}(a)(x-a),$$

já que

$$d < a \implies -\dot{f}(d) \geq -\dot{f}(a).$$

Por fim, suponhamos c) e provemos a). Sejam $a, b \in \text{int}(I)$, com $a < b$. Se $c \in (a, b)$, então

$$f(a) \geq f(c) + \dot{f}(c)(a-c) \text{ e } f(b) \geq f(c) + \dot{f}(c)(b-c).$$

Além disso, existe $\lambda \in (0, 1)$ tal que

$$c = (1-\lambda)a + \lambda b.$$

Desse modo,

$$(1-\lambda)f(a) \geq (1-\lambda)f(c) + \dot{f}(c)(1-\lambda)(a-c) \text{ e } \lambda f(b) \geq \lambda f(c) + \dot{f}(c)\lambda(b-c),$$

donde

$$(1-\lambda)f(a)+\lambda f(b)\geq f(c)+\dot{f}(c)[\underbrace{(1-\lambda)a+\lambda b}_{=c}-c]=f(c)=f((1-\lambda)a+\lambda b).$$

Portanto, f é convexa. □

Teorema 6.2.11 (Sobre a concavidade)**.** Sejam I um intervalo aberto de números reais e $f: I \to \mathbb{R}$ uma função duas vezes derivável. Então, f é convexa se, e somente se,

$$f^{\{2\}}(x)\geq 0, \quad \text{para todo} \quad x\in I.$$

Demonstração. Se f é convexa, então $\dot{f}$ é não decrescente, donde

$$f^{\{2\}}(x)\geq 0, \text{ para todo } x\in I.$$

Reciprocamente, suponhamos que

$$f^{\{2\}}(x)\geq 0, \text{ para todo } x\in I.$$

Sejam $x, y \in I$, com $x < y$, e $\lambda \in (0,1)$. Então,

$$x<(1-\lambda)x+\lambda y<y$$

e, pelo Teorema do Valor Médio de Lagrange (ver Corolário 2.1.4), pondo

$$z:=(1-\lambda)x+\lambda y,$$

temos

$$\begin{aligned}
\dot{f}(c_1)&=\frac{f(z)-f(x)}{z-x}, \text{ para algum } c_1\in(x,z);\\
\dot{f}(c_2)&=\frac{f(y)-f(z)}{y-z}, \text{ para algum } c_2\in(z,y);\\
f^{\{2\}}(c_3)&=\frac{\dot{f}(c_2)-\dot{f}(c_1)}{c_2-c_1}, \text{ para algum } c_3\in(c_1,c_2).
\end{aligned}$$

Queremos verificar que

$$f((1-\lambda)x+\lambda y)\leq(1-\lambda)f(x)+\lambda f(y),$$

ou, equivalentemente,

$$f((1-\lambda)x+\lambda y)-(1-\lambda)f(x)-\lambda f(y)\leq 0.$$

Calculando a expressão utilizando as identidades que obtemos, encontramos

$$\begin{aligned}
f((1-\lambda)x+\lambda y)-(1-\lambda)f(x)-\lambda f(y) &= (1-\lambda)[f((1-\lambda)x+\lambda y)-f(x)] \\
&\quad + \lambda[f((1-\lambda)x+\lambda y)-f(y)] \\
&= (1-\lambda)\lambda(y-x)\dot{f}(c_1)-\lambda(1-\lambda)(y-x)\dot{f}(c_2) \\
&= \lambda(1-\lambda)(y-x)[\dot{f}(c_1)-\dot{f}(c_2)] \\
&= \underbrace{\lambda}_{>0}\underbrace{(1-\lambda)}_{>0}\underbrace{(y-x)}_{>0}\underbrace{(c_1-c_2)}_{<0}\underbrace{f^{\{2\}}(c_3)}_{\geq 0} \\
&\leq 0,
\end{aligned}$$

como queríamos demonstrar. □

Exemplo 6.2.12. Dados um número real x e um inteiro $k \geq 0$, definimos

$$\binom{x}{k} := \begin{cases} 1, & \text{se} \quad k = 0, \\ 1/(k!) \cdot \prod_{i=0}^{k-1}(x-i), & \text{se} \quad k \geq 1. \end{cases}$$

Fixado $k \geq 0$ inteiro, a função $f : x \in \mathbb{R} \mapsto \binom{x}{k}$ é convexa em $[k-1, +\infty)$.

Demonstração. Se $k = 0$, a função é constante igual a 1; se $k = 1$, temos $f = \mathrm{Id}_{\mathbb{R}}$. Logo, se $k = 0$ ou $k = 1$, f é convexa. Suponhamos então $k \geq 2$ e note que

$$f(x) = \frac{1}{k!} \cdot \prod_{i=0}^{k-1}(x-i)$$

é um produto de k polinômios de grau 1, logo é um polinômio de grau k, sendo suas raízes exatamente $0, 1, \ldots, k-1$. Pelo Teorema de Rolle (ver Corolário 2.1.2), existem $c_1, c_2, \ldots, c_{k-1}$ reais com

$$0 < c_1 < 1 < c_2 < 2 < \ldots < c_{k-1} < k-1$$

tais que

$$\dot{f}(c_i) = 0, \quad \text{para todo} \quad i = 1, \ldots, k-1.$$

Como $\dot{f}(x)$ é um polinômio de grau $k-1$, os c_i's são todas as suas raízes. Aplicando novamente o Teorema de Rolle obtemos $d_1, \ldots, d_{k-2}$ reais com

$$0 < c_1 < d_1 < c_2 < d_2 < \ldots < d_{k-2} < c_{k-1} < k-1$$

tais que

$$f^{\{2\}}(d_i) = 0, \quad \text{para todo } i = 1, \ldots, k-2.$$

Como os d_i's são todas as raízes de $q(x) := f^{\{2\}}(x)$, então $q(x)$ tem sinal constante em $(d_{k-2}, +\infty)$. Observe que $1/(k!) \cdot x^k$ é o termo de maior grau em $f(x)$, logo $k/(k!) \cdot x^{k-1}$ é o termo de maior grau em $\dot{f}(x)$ e, portanto, $(k(k-1))/(k!)\cdot x^{k-2}$ é o termo de maior grau em $f^{\{2\}}(x)$. Observe ainda que

$$(k(k-1))/(k!)\cdot x^{k-2} \xrightarrow{x\to\infty} \begin{cases} 1, & \text{se} \quad k=2, \\ +\infty, & \text{se} \quad k\geq 3. \end{cases}$$

Logo, para todo x suficientemente grande, $q(x) > 0$. A constância do sinal de q em $(d_{k-2}, +\infty)$ garante que $q > 0$ nesse intervalo. Em particular, $q(x) > 0$ para $x \geq k-1$. Pelo Teorema 6.2.11, isso prova que a função f é convexa em $[k-1, +\infty)$, como queríamos demonstrar. □

Teorema 6.2.13. Sejam E um espaço normado e $C \subseteq E$ um subconjunto aberto e convexo. Se $f : C \to \mathbb{R}$ é uma função diferenciável, então as seguintes afirmações são equivalentes:

1. f é convexa;

2. $f(y) \geq f(x) + f'(x)(y-x)$, para quaisquer $x, y \in C$.

Demonstração. Se f é uma função convexa, considere $x, y \in C$ e $\alpha \in (0,1]$. Como

$$f((1-\alpha)x+\alpha y) \leq (1-\alpha)f(x) + \alpha f(y),$$

então

$$\frac{f(x+\alpha(y-x)) - f(x)}{\alpha} \leq f(y) - f(x),$$

donde, da diferenciabilidade de f e da Monotonicidade do Limite (ver Teorema B.5.8), temos

$$f'(x)(y-x) = \frac{\partial f}{\partial(y-x)}(x) = \lim_{\alpha\to 0} \frac{f(x+\alpha(y-x)) - f(x)}{\alpha} \leq f(y) - f(x).$$

Reciprocamente, considere $x, y \in C$ e $\alpha \in (0,1)$. Queremos verificar que

$$f((1-\alpha)x+\alpha y) \leq (1-\alpha)f(x) + \alpha f(y).$$

Como

$$\begin{aligned} f(x) - f((1-\alpha)x+\alpha y) &\geq f'((1-\alpha)x+\alpha y)(\alpha(x-y)) \\ &\geq \alpha f'((1-\alpha)x+\alpha y)(x-y), \end{aligned}$$

então

$$\frac{f(x) - f((1-\alpha)x+\alpha y)}{\alpha} \geq f'((1-\alpha)x+\alpha y)(x-y).$$

Além disso,

$$\begin{aligned} f(y) - f((1-\alpha)x + \alpha y) &\geq f'((1-\alpha)x + \alpha y)((1-\alpha)(y-x)) \\ &\geq (1-\alpha)f'((1-\alpha)x + \alpha y)(y-x), \end{aligned}$$

donde

$$\frac{f(y) - f((1-\alpha)x + \alpha y)}{1-\alpha} \geq -f'((1-\alpha)x + \alpha y)(x-y)$$

e, portanto,

$$\frac{f((1-\alpha)x + \alpha y) - f(y)}{1-\alpha} \leq f'((1-\alpha)x + \alpha y)(x-y) \leq \frac{f(x) - f((1-\alpha)x + \alpha y)}{\alpha}.$$

Desse modo,

$$\alpha[f((1-\alpha)x + \alpha y) - f(y)] \leq (1-\alpha)[f(x) - f((1-\alpha)x + \alpha y)].$$

Logo,

$$f((1-\alpha)x + \alpha y) \leq (1-\alpha)f(x) + \alpha f(y),$$

isto é, f é convexa. □

Lema 6.2.14. Sejam E um espaço vetorial, $C \subseteq E$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função. Então, f é convexa se, e somente se, para quaisquer $x, y \in E$, a função

$$g : t \in \{s \in \mathbb{R} : x + sy \in C\} \mapsto f(x + ty)$$

é uma função convexa.

Demonstração. Sejam $x, y \in E$. Então

$$D := \{s \in \mathbb{R} : x + sy \in C\}$$

é convexo, pelo Corolário 6.1.6. Sejam agora $a, b \in D$ e $t \in [0, 1]$. Temos

$$\begin{aligned} g((1-t)a + tb) &= f(x + ((1-t)a + tb)y) = f((1-t)(x + ay) + t(x + by)) \\ &\leq (1-t)f(x + ay) + tf(x + by) = (1-t)g(a) + tg(b). \end{aligned}$$

Logo, g é uma função convexa.

Reciprocamente, sejam $x, y \in C, t \in [0, 1]$. Nesse caso, $a := x$ e $b := y - x$ estão em E. Definindo

$$D := \{s \in \mathbb{R} : a + sb \in C\},$$

então D é convexo e $0, 1 \in D$. Faça $g(t) := f(a + tb)$. Por hipótese, g é convexa, então

$$\begin{aligned} f((1-t)x + ty) &= f(x + t(y-x)) = f(a + tb) = g(t) = g((1-t)0 + t1) \\ &\leq (1-t)g(0) + tg(1) = (1-t)f(x) + tf(y). \end{aligned}$$

Logo, f é convexa. □

Teorema 6.2.15. Sejam E um espaço normado, $C \subseteq E$ um subconjunto convexo aberto e $f : C \to \mathbb{R}$ uma aplicação duas vezes diferenciável. Então, f é convexa se, e somente se,

$$f^{(2)}(x)h^{(2)} \geq 0, \quad \text{para quaisquer} \quad x \in C \quad \text{e} \quad h \in E.$$

Demonstração. Suponhamos que f é convexa. Dados $x \in C$ e $y \in E$, sabemos, pelo Lema 6.2.14, que a função

$$g : t \in D := \{s \in \mathbb{R} : x + sy \in C\} \mapsto f(x + ty)$$

é convexa. Além disso, D é aberto, pelo Corolário 6.1.6. Como a aplicação

$$m : t \in D \mapsto x + ty \in E$$

é diferenciável, com $m(D) \subseteq C$, então, pela Regra da Cadeia (ver Teorema 1.3.1), para quaisquer $t \in D$ e $h \in E$, temos

$$g'(t)h = f'(m(t))(m'(t)h) = f'(x + ty)(hy) = hf'(x + ty)(y).$$

Logo,

$$\dot{g}(t) = f'(x + ty)(y).$$

Considere a aplicação $\alpha(T) := Ty$, para todo $T \in \mathcal{L}(E, \mathbb{R})$. Então α é linear e

$$|\alpha(T)| = |Ty| \leq \|T\|\|y\|,$$

logo α é contínua, pelo Teorema C.4.1 e, portanto, diferenciável (ver Exemplo 1.1.14). Observe que

$$\dot{g}(t) = (\alpha \circ f' \circ m)(t).$$

Em particular, $n(t) := \dot{g}(t)$ é diferenciável, como composição de três aplicações diferenciáveis, e temos

$$\begin{aligned} n'(t)h &= \alpha'((f' \circ m)(t))((f' \circ m)'(t)h) = \alpha(f^{[2]}(m(t))(m'(t)h)) = f^{[2]}(x + ty)(yh)(y) \\ &= hf^{[2]}(x + ty)(y)(y) = hf^{(2)}(x + ty)(y, y) = hf^{(2)}(x + ty)y^{(2)}. \end{aligned}$$

Logo, como g é convexa, temos, pelo Teorema 6.2.11,

$$0 \leq g^{\{2\}}(t) = \dot{n}(t) = f^{(2)}(x + ty)y^{(2)},$$

donde vem

$$f^{(2)}(x)y^{(2)} \geq 0,$$

já que $0 \in D$.

A prova da recíproca é a mesma apenas refazendo os passos. $\square$

Corolário 6.2.16. Se E é um espaço normado, $C \subseteq E$ é um subconjunto aberto e convexo e $f : C \to \mathbb{R}$ é uma função convexa e diferenciável, então todo ponto estacionário de f é um minimizador global de f.

Demonstração. Se $a \in C$ é um ponto estacionário, então

$$f(x) \geq f(a) + f'(a)(x-a) = f(a) + 0_{E^*}(x-a) = f(a),$$

para todo $x \in C$. $\square$

Finalizamos essa seção com as desigualdades de Young e Jensen (e algumas aplicações).

Teorema 6.2.17 (Desigualdade de Young)**.** Sejam $a, b \geq 0$ e $p, q \in (1, \infty)$ tais que $1/p + 1/q = 1$. Então

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}.$$

Demonstração. Como a função exponencial é convexa, temos

$$\begin{aligned} ab = \exp\left(\frac{1}{p}\log(a^p)\right)\exp\left(\frac{1}{q}\log(b^q)\right) &= \exp\left(\frac{1}{p}\log(a^p) + \frac{1}{q}\log(b^q)\right) \\ &= \exp\left(\left(1 - \frac{1}{q}\right)\log(a^p) + \frac{1}{q}\log(b^q)\right) \\ &\leq \left(1 - \frac{1}{q}\right)a^p + \frac{1}{q}b^q \\ &= \frac{a^p}{p} + \frac{b^q}{q}, \end{aligned}$$

como queríamos demonstrar. $\square$

Corolário 6.2.18. Dados $A, B \geq 0$ e $\theta \in [0, 1]$, temos

$$A^\theta B^{1-\theta} \leq \theta A + (1-\theta)B.$$

Demonstração. A desigualdade vale trivialmente para $\theta = 0$ e $\theta = 1$. Suponhamos então $0 < \theta < 1$. Faça $a = A^\theta, b = B^{1-\theta}, p = 1/\theta$ e tome q tal que $1/p + 1/q = 1$. Então, pela desigualdade de Young (ver Teorema 6.2.17), temos

$$A^\theta B^{1-\theta} = ab \leq \frac{a^p}{p} + \frac{b^q}{q} = \theta A + (1-\theta)B,$$

como queríamos demonstrar. $\square$

Teorema 6.2.19 (Desigualdade de Jensen)**.** Sejam I um intervalo de números reais, $f : I \to \mathbb{R}$ uma função convexa, $x_1, \ldots, x_n \in I$ e $\lambda_1, \ldots, \lambda_n \geq 0$ tais que $\sum \lambda_i = 1$. Então,

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i).$$

Demonstração. Para $n = 2$, a desigualdade é

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2),$$

com $\lambda_1 + \lambda_2 = 1$. Se tomarmos $\lambda_1 = t$, então $\lambda_2 = 1 - t$ para $t \in [0, 1]$. A expressão se torna

$$f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2),$$

que é a própria definição de uma função convexa. Portanto, o caso base é verdadeiro.

Suponhamos agora, por hipótese de indução, que a desigualdade é válida para um inteiro $k \geq 2$. Ou seja, para quaisquer $y_1, \ldots, y_k \in I$ e $\mu_1, \ldots, \mu_k \geq 0$ com $\sum_{i=1}^{k} \mu_i = 1$, temos

$$f\left(\sum_{i=1}^{k} \mu_i y_i\right) \leq \sum_{i=1}^{k} \mu_i f(y_i).$$

Queremos provar que a desigualdade vale para $k+1$ pontos. Sejam $x_1, \ldots, x_{k+1} \in I$ e $\lambda_1, \ldots, \lambda_{k+1} \geq 0$ com $\sum_{i=1}^{k+1} \lambda_i = 1$. Se $\lambda_{k+1} = 1$, a desigualdade é trivial. Assumimos, portanto, $\lambda_{k+1} < 1$. Definimos $C := \sum_{i=1}^{k} \lambda_i = 1 - \lambda_{k+1}$. Como $\lambda_{k+1} < 1$, temos $C > 0$. Então,

$$\begin{aligned}
f\left(\sum_{i=1}^{k+1} \lambda_i x_i\right) &= f\left(\left(\sum_{i=1}^{k} \lambda_i x_i\right) + \lambda_{k+1} x_{k+1}\right) \\
&= f\left(C \cdot \left(\sum_{i=1}^{k} \frac{\lambda_i}{C} x_i\right) + \lambda_{k+1} x_{k+1}\right) \\
\text{(Caso } n = 2\text{)} &\leq C \cdot f\left(\sum_{i=1}^{k} \frac{\lambda_i}{C} x_i\right) + \lambda_{k+1} f(x_{k+1}) \\
\text{(Hipótese de Indução)} &\leq C \cdot \left(\sum_{i=1}^{k} \frac{\lambda_i}{C} f(x_i)\right) + \lambda_{k+1} f(x_{k+1}) \\
&= \left(\sum_{i=1}^{k} C \cdot \frac{\lambda_i}{C} f(x_i)\right) + \lambda_{k+1} f(x_{k+1}) \\
&= \sum_{i=1}^{k} \lambda_i f(x_i) + \lambda_{k+1} f(x_{k+1}) \\
&= \sum_{i=1}^{k+1} \lambda_i f(x_i).
\end{aligned}$$

Isso completa o passo indutivo. Pelo princípio da indução matemática, a desigualdade é válida para todo inteiro $n \geq 2$. □

Corolário 6.2.20. Sejam $n \in \mathbb{N}, t \geq 0$ um inteiro e $x_1, \ldots, x_n \geq t - 1$ números reais. Então,

$$\sum_{i=1}^{n} \binom{x_i}{t} \geq n \cdot \binom{s/n}{t}, \quad \text{onde} \quad s := \sum_{i=1}^{n} x_i.$$

Demonstração. Como a função

$$f : x \in \mathbb{R} \mapsto \binom{x}{t}$$

é convexa para $x \geq t-1$, então, pela desigualdade de Jensen (ver Teorema 6.2.19), temos

$$\frac{1}{n}\sum_{i=1}^{n}\binom{x_i}{t} = \frac{1}{n}\sum_{i=1}^{n} f(x_i) \geq f\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \binom{s/n}{t},$$

como queríamos demonstrar. □

Corolário 6.2.21 (Média aritmética e média geométrica)**.** Dados $x_1, \ldots, x_n \geq 0$, temos

$$\sqrt[n]{x_1 \cdots x_n} \leq \frac{x_1 + \cdots + x_n}{n}.$$

Demonstração. A desigualdade é trivial se algum $x_i = 0$. Suponhamos então que $x_i > 0$ para todo $i = 1, \ldots, n$. Considere a função $f = -\log$. Calculando as derivadas, temos:

$$\dot{f}(x) = -\frac{1}{x} \quad \text{e} \quad f^{\{2\}}(x) = \frac{1}{x^2}.$$

Como $f^{\{2\}} > 0$, pelo Teorema 6.2.11, a função f é convexa. Podemos, então, aplicar a Desigualdade de Jensen (ver Teorema 6.2.19) com os pesos $\lambda_i = 1/n$ para $i = 1, \ldots, n$:

$$f\left(\sum_{i=1}^{n}\frac{1}{n}x_i\right) \leq \sum_{i=1}^{n}\frac{1}{n}f(x_i),$$

o que nos dá

$$-\log\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right) \leq \frac{1}{n}\sum_{i=1}^{n}(-\log x_i)$$

Usando as propriedades do logaritmo,

$$-\log\left(\frac{x_1 + \cdots + x_n}{n}\right) \leq -\frac{1}{n}\log(x_1 \cdots x_n)$$

$$-\log\left(\frac{x_1 + \cdots + x_n}{n}\right) \leq -\log\left((x_1 \cdots x_n)^{1/n}\right)$$

donde vem

$$\log\left(\frac{x_1 + \cdots + x_n}{n}\right) \geq \log\left(\sqrt[n]{x_1 \cdots x_n}\right)$$

Como o log é uma função crescente, temos

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n},$$

como queríamos demonstrar. □

6.3 Otimização Convexa

O problema geral de **Otimização Convexa** (OC) é o de encontrar um minimizador local de uma função $f : C \to \mathbb{R}$, onde E é um espaço normado, $C \subseteq E$ é um conjunto convexo e f é uma função convexa. Esse problema é representado por

$$\text{(OC)} \min_{\text{s.a. } x \in C} f(x).$$

Teorema 6.3.1. Se E é um espaço normado, $C \subseteq E$ é um conjunto convexo e $f : C \to \mathbb{R}$ é uma função convexa, então todo minimizador local de f é um minimizador global.

Demonstração. Seja $x^* \in C$ um minimizador local de f. Por definição, existe um $\delta > 0$ tal que

$$f(x^*) \leq f(y), \quad \text{para todo} \quad y \in C \cap B(x^*, \delta).$$

Afirmamos que x^* é um minimizador global.

Com efeito, seja $z \in C$ um ponto arbitrário. Se $z \in B(x^*, \delta)$, temos $f(x^*) \leq f(z)$. Suponhamos então que $z \notin B(x^*, \delta)$ e consideremos

$$y(t) := (1-t)x^* + tz, \quad \text{para} \quad t \in [0,1].$$

Como $x^*, z \in C$ e C é convexo, então $y([0,1]) = [x^*, z] \subseteq C$. Veja que, para todo $t \in [0,1]$,

$$\|y(t) - x^*\| = \|(1-t)x^* + tz - x^*\| = \|t(z - x^*)\| = t\|z - x^*\|,$$

então, tomando $t_0 \in (0,1)$ tal que $t_0 < \delta/\|z - x^*\|$, temos $y(t_0) \in B(x^*, \delta)$. Isso nos dá, junto com o fato de f ser convexa, que

$$f(x^*) \leq f(y(t_0)) = f((1-t_0)x^* + t_0 z) \leq (1-t_0)f(x^*) + t_0 f(z).$$

Logo,

$$t_0 f(x^*) \leq t_0 f(z)$$

e, portanto,

$$f(x^*) \leq f(z),$$

o que termina a prova. □

Teorema 6.3.2 (Condição de otimalidade de 1ª ordem)**.** Suponhamos que a função objetivo f do problema (OC) é diferenciável. Assim, $x^* \in C$ é minimizador global do problema (OC) se, e somente se, $x^* \in C$ e

$$f'(x^*)(x - x^*) \geq 0, \quad \text{para todo} \quad x \in C. \tag{6.1}$$

Demonstração. Começamos provando a recíproca. Pelo Teorema 6.2.13, dado $x \in C$, temos

$$f(x) \geq f(x^*) + f'(x^*)(x - x^*) \geq f(x^*),$$

donde x^* é minimizador global de f. Reciprocamente, suponhamos que x^* é minimizador global de f, mas não vale (6.1), isto é,

$$f'(x^*)(x - x^*) < 0, \quad \text{para algum} \quad x \in C.$$

Seja $w : [0,1] \to E$ dada por

$$w(t) := (1-t)x^* + tx = x^* + t(x - x^*), \quad \text{para todo} \quad t \in [0,1].$$

Como C é convexo, então

$$w([0,1]) = [x^*, x] \subseteq C.$$

Pela Regra da Cadeia (ver Teorema 1.3.1), a função $f \circ w : [0,1] \to \mathbb{R}$ é diferenciável, com

$$(f \circ w)'(0)1 = f'(w(0))(w'(0)1) = f'(x^*)(x - x^*) < 0,$$

donde

$$\lim_{t \to 0} \frac{f(x^* + t(x - x^*)) - f(x^*)}{t} < 0,$$

e, portanto, pela Permanência de Sinal, existe $\delta > 0$ tal que

$$0 < t < \delta \implies f(w(t)) < f(x^*).$$

Nessas condições,

$$f(w(t_0)) < f(x^*), \quad \text{onde} \quad t_0 := \frac{\delta}{2},$$

o que contraria o fato de que x^* é minimizador global de f. □

Teorema 6.3.3. Sejam E um espaço normado, $C \subseteq E$ um subconjunto convexo aberto e $f_1, f_2 : C \to \mathbb{R}$ funções convexas. Se f_1 é diferenciável e $f := f_1 + f_2$, então as seguintes afirmações são equivalentes:

(a) $\overline{x} \in C$ é solução de (OC), onde f é a função objetivo e C é o conjunto viável;

(b) $f_1'(\overline{x})(y - \overline{x}) + f_2(y) - f_2(\overline{x}) \geq 0$, para todo $y \in C$.

Demonstração. Suponhamos (a) e provemos (b). Supondo que $\overline{x}$ é solução de (OC), então, dado $\lambda \in (0,1)$, temos

$$f(\overline{x}) \leq f((1-\lambda)\overline{x} + \lambda y), \quad \text{para todo} \quad y \in C.$$

Da convexidade de f_2 e da definição de f, temos

$$\begin{aligned} f_1(\overline{x}) + f_2(\overline{x}) &\leq f_1((1-\lambda)\overline{x} + \lambda y) + f_2((1-\lambda)\overline{x} + \lambda y) \\ &\leq f_1((1-\lambda)\overline{x} + \lambda y) + (1-\lambda) f_2(\overline{x}) + \lambda f_2(y), \end{aligned}$$

donde

$$\frac{1}{\lambda}\left[f_1((1-\lambda)\overline{x} + \lambda y) - f_1(\overline{x})\right] + f_2(y) - f_2(\overline{x}) \geq 0.$$

Pela Monotonicidade do Limite (ver Teorema B.5.8), temos

$$f_1'(\overline{x})(y - \overline{x}) + f_2(y) - f_2(\overline{x}) \geq 0, \quad \text{para todo} \quad y \in C. \tag{6.2}$$

Por fim, suponhamos (b) e provemos (a). Pelo Teorema 6.2.13, da convexidade de f_1 temos

$$f_1(y) - f_1(\bar{x}) - f_1'(\overline{x})(y - \overline{x}) \geq 0, \quad \text{para todo} \quad y \in C.$$

Somando essa última desigualdade com a desigualdade (6.2) obtemos

$$f_1(y) + f_2(y) - (f_1(\overline{x}) + f_2(\overline{x})) \geq 0 \implies f(y) \geq f(\overline{x}), \quad \text{para todo} \quad y \in C,$$

o que prova que $\overline{x}$ é solução de (OC). □

7 Método do Gradiente

Nesse capítulo estabeleceremos o **Método do Gradiente** para espaços de Hilbert.

7.1 Algoritmos de descida

A definição abaixo foi retirada de (RIBEIRO; KARAS, 2013, Definição 4.1, p. 66).

Definição 7.1.1 (Direção de descida)**.** Se E é um espaço normado, $f : E \to \mathbb{R}$ é uma função e $\overline{x} \in E$, um ponto $d \in E \setminus \{0_E\}$ é uma **direção de descida** para f, a partir de $\overline{x}$, quando existe $\delta > 0$ tal que

$$f(\overline{x} + td) < f(\overline{x}), \quad \text{para todo} \quad t \in (0, \delta).$$

Teorema 7.1.2. Sejam E um espaço normado, $f : E \to \mathbb{R}$ uma função diferenciável e $\overline{x} \in E$. Se $d \in E$ é tal que

$$f'(\overline{x}) \cdot d < 0$$

então d é uma direção de descida para f, a partir de $\overline{x}$.

Demonstração. Como

$$0 > f'(\overline{x}) \cdot d = \lim_{t \to 0^+} \frac{f(\overline{x} + td) - f(\overline{x})}{t},$$

então, pela Permanência de Sinal, existe $\delta > 0$ tal que

$$0 < t < \delta \implies \frac{f(\overline{x} + td) - f(\overline{x})}{t} < 0 \implies f(\overline{x} + td) < f(\overline{x}),$$

como queríamos demonstrar. □

Corolário 7.1.3. Se H é um espaço de Hilbert, $f : H \to \mathbb{R}$ é uma função diferenciável e $\overline{x} \in H$ não é um ponto estacionário de f, então $-\nabla f(\overline{x})$ é uma direção de descida para f, a partir de $\overline{x}$.

Demonstração. Como $\overline{x}$ não é um ponto estacionário de f, então $\nabla f(\overline{x}) \neq 0_H$, donde

$$f'(\overline{x}) \cdot [-\nabla f(\overline{x})] = \langle \nabla f(\overline{x}), -\nabla f(\overline{x}) \rangle = -\|\nabla f(\overline{x})\|^2 < 0$$

e, portanto, $-\nabla f(\overline{x})$ é uma direção de descida para f, a partir de $\overline{x}$. □

O último corolário é precisamente um dos pilares do **Método do Gradiente**, que introduziremos nessa seção.

Algoritmo 1. Dado $x^0 \in E$

$k = 0$

REPITA enquanto $f'(x^k) \neq 0_{\mathcal{L}(E,\mathbb{R})}$

Calcule d^k tal que $f'(x^k) \cdot d^k < 0$

Escolha $t_k > 0$ tal que $f(x^k + t_k d^k) < f(x^k)$

Faça $x^{k+1} = x^k + t_k d^k$

$k = k + 1$

O Algoritmo 1 ou encontra um ponto estacionário em um número finito de iterações ou gera uma sequência ao longo da qual f decresce. A questão agora é saber se esta sequência tem algum ponto de acumulação e, caso afirmativo, se este ponto é estacionário. Considere $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) := x^2$ e as sequências

$$x^k = 1 + \frac{1}{k+1}$$

e

$$y^k = (-1)^k + \frac{(-1)^k}{k+1}.$$

Ambas podem ser obtidas pelo algoritmo, $x^k \xrightarrow{k\to\infty} 1$ e (y^k) têm dois pontos de acumulação, 1 e -1, mas nenhum desses pontos é estacionário.

Deste modo, se quisermos garantir convergência, a escolha da direção d^k e do tamanho do passo t_k, no Algoritmo 1, não pode ser arbitrária.

7.2 Método de busca exata

Dados um espaço normado E, uma função $f : E \to \mathbb{R}$, um ponto $\overline{x} \in E$ e $d \in \mathbb{R}^n$ uma direção de descida para f, queremos encontrar $\bar{t} > 0$ tal que

$$f(\overline{x} + \bar{t}d) < f(\overline{x}).$$

O **método de busca exata** consiste em

$$\begin{gathered}\text{minimizar } f(\overline{x} + td)\\ \text{sujeito a } t > 0.\end{gathered}$$

Segundo (RIBEIRO; KARAS, 2013, p. 70), “Este problema é, em geral difícil de resolver. Entretanto, para certas funções especiais, existem algoritmos para resolvê-lo.”

Uma pergunta talvez mais fundamental do que a de como calcular essas soluções é a de que se elas existem ou não. Em geral, note que o problema posto não necessariamente possui solução. Considerando a função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(t) := e^{-t}$, note que qualquer $d > 0$ é uma direção de descida para f. Com efeito, dado $t > 0$ e $d > 0$, temos

$$e^{td} > 1 \implies e^{-td} < 1 \implies f(\overline{x} + td) = e^{-\overline{x}-td} < e^{-\overline{x}} = f(\overline{x}).$$

Como a função $g : \mathbb{R}^+ \to \mathbb{R}$ dada por

$$g(t) := f(\overline{x} + td)$$

é derivável, então, se g possuísse qualquer extremo local, isso ocorreria somente em um ponto estacionário, pelo Teorema 5.1.2. No entanto,

$$0 = \dot{g}(t) = d \cdot \dot{f}(\overline{x} + td) \iff -e^{-\overline{x}-td} = \dot{f}(\overline{x} + td) = 0,$$

e, portanto, mesmo sendo g limitada inferiormente, g não possui pontos estacionários.

Sejam H um espaço de Hilbert, $f : H \to \mathbb{R}$ uma função diferenciável possuindo pelo menos um minimizador local, $\overline{x} \in H$ um ponto e $d \in H$ uma direção de descida. Pelo Teorema 5.1.2, sabemos que os candidatos a minimizadores da função $g : \mathbb{R}^+ \to \mathbb{R}$ dada por

$$g(t) := f(\overline{x} + td)$$

são seus pontos estacionários. Como a aplicação $m : \mathbb{R}^+ \to H$ dada por

$$m(t) := \overline{x} + td$$

é diferenciável e f também o é (por hipótese), então g é diferenciável, pela Regra da Cadeia (1.3.1), já que $g = f \circ m$, valendo

$$\dot{g}(t) = g'(t)1 = f'(m(t))(m'(t)1) = f'(m(t))(d) = \langle d, \nabla f(m(t))\rangle = \langle d, \nabla f(\overline{x} + td)\rangle.$$

Em particular, se

$$f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c, \quad \text{para todo} \quad x \in H,$$

onde $A : H \to H$ é um operador autoadjunto e contínuo, $b \in H$ e $c \in \mathbb{R}$, então

$$\langle \nabla f(x), h\rangle = f'(x)h = \langle Ax + b, h\rangle,$$

donde, pela Proposição C.5.5, temos

$$\nabla f(x) = Ax + b$$

e, portanto,

$$\dot{g}(t) = \langle d, A(\overline{x} + td) + b\rangle = \langle d, A\overline{x} + b\rangle + t\langle d, Ad\rangle,$$

isto é,

$$\dot{g}(t) = 0 \iff t = -\frac{\langle d, A\overline{x} + b\rangle}{\langle d, Ad\rangle} = -\frac{\langle d, \nabla f(\overline{x})\rangle}{\langle d, Ad\rangle}.$$

Como d é uma direção de descida para f, então

$$\langle d, \nabla f(\overline{x})\rangle = f'(\overline{x})d < 0,$$

e, portanto, $\langle d, Ad\rangle$ precisa ser positivo a fim de que t também o seja. Como d é uma direção de descida arbitrária, podemos contornar o problema assumindo que

$$\langle x, Ax\rangle > 0, \text{ para todo } x \in H \setminus \{0_H\},$$

ou seja, assumindo que o operador A é *definido positivo*.

Além disso, já vimos que $-\nabla f(\overline{x})$ é uma direção de descida para f. Assim, se $d = -\nabla f(\overline{x})$, obtemos

$$\dot{g}(t) = 0 \iff t = \frac{\langle d, d\rangle}{\langle d, Ad\rangle}.$$

O que acabamos de fazer é, como se verá, calcular o tamanho do passo no *método do gradiente com busca exata* aplicado à generalização do *problema de minimização quadrática*.

7.3 Análise de convergência global

Sejam H um espaço de Hilbert, $T : H \to \mathcal{L}(H)$ uma aplicação contínua tal que, para cada $z \in H$, a forma bilinear $L_z : H \times H \to \mathbb{R}$ dada por

$$L_z(x, y) := \langle x, T(z)y\rangle$$

é definida positiva. Desse modo, $d := -T(x)\nabla f(x)$ é uma direção de descida, já que

$$f'(x) \cdot d = \langle \nabla f(x), -T(x)\nabla f(x)\rangle = -\langle \nabla f(x), T(x)\nabla f(x)\rangle = -L_x\left(\nabla f(x), \nabla f(x)\right) < 0.$$

Temos assim uma maneira de obter direções de descida para o Algoritmo 1. Para facilitar, vamos reescrever o algoritmo com esta escolha da direção de busca. A determinação do tamanho do passo será feita por busca exata.

Algoritmo 2 (Algoritmo de descida)**.** Dado $x^0 \in H$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0_H$

Defina $d^k = -T(x^k)\nabla f(x^k)$

Obtenha $t_k > 0$ tal que $f(x^k + t_k d^k) < f(x^k)$

Faça $x^{k+1} = x^k + t_k d^k$

$k = k + 1$

Definição 7.3.1. (RIBEIRO; KARAS, 2013, Definição 4.11, p. 82) Um algoritmo é dito **globalmente convergente** quando para qualquer sequência (x^k) gerada pelo algoritmo e qualquer ponto de acumulação $\overline{x}$ de (x^k), temos que $\overline{x}$ é um ponto estacionário.

O teorema abaixo, que se encontra em de (RIBEIRO; KARAS, 2013, Teorema 4.12, p. 82), garante a convergência global do Algoritmo 2.

Teorema 7.3.2 (Convergência global do algoritmo de descida)**.** Se f é continuamente diferenciável, o Algoritmo 2, com o tamanho do passo calculado pela busca exata, é globalmente convergente.

Demonstração. Sejam (x^k) uma sequência gerada pelo algoritmo e $\overline{x} \in H$ um ponto de acumulação de (x^k). Considere (x^{n_k}) uma subsequência de (x^k) tal que $\lim\limits_{k\to\infty} x^{n_k} = \overline{x}$ e suponha, por redução ao absurdo, que $\overline{x}$ não é um ponto estacionário de f, isto é, $\nabla f(\overline{x}) \neq 0_H$. Como

$$\overline{d} := -T(\overline{x})\nabla f(\overline{x})$$

é uma direção de descida, então existe $\overline{t} > 0$ tal que

$$\beta := f(\overline{x}) - f(\overline{x} + \overline{t}\overline{d}) > 0.$$

Considere $h : H \to \mathbb{R}$ dada por

$$h(x) = f(x) - f\left(x - \overline{t}T(x)\nabla f(x)\right).$$

Verificaremos agora que h é contínua.

Seja $(y^k) \subseteq H$ tal que $\lim\limits_{k\to\infty} y^k = p$. Da continuidade de T e de ∇f, temos $T(y^k) \xrightarrow[\|\cdot\|_{\mathcal{L}(H)}]{k\to\infty} T(p)$ e $\nabla f(y^k) \xrightarrow[\|\cdot\|_H]{k\to\infty} \nabla f(p)$. Como

$$\begin{aligned}
\|T(y^k)\nabla f(y^k) - T(p)\nabla f(p)\|_H &= \|T(y^k)\nabla f(y^k) - T(y^k)\nabla f(p) \\
&\quad + T(y^k)\nabla f(p) - T(p)\nabla f(y^k) \\
&\qquad + T(p)\nabla f(y^k) - T(p)\nabla f(p)\|_H \\
&\leq \|T(y^k)\left(\nabla f(y^k) - \nabla f(p)\right)\|_H \\
&\quad + \|T(y^k)\nabla f(p) - T(p)\nabla f(y^k)\|_H \\
&\qquad + \|T(p)\left(\nabla f(y^k) - \nabla f(p)\right)\|_H \\
&\leq \|T(y^k)\|_{\mathcal{L}(H)}\|\nabla f(y^k) - \nabla f(p)\|_H \\
&\quad + \|T(y^k)\nabla f(p) - T(p)\nabla f(y^k)\|_H \\
&\qquad + \|T(p)\|_{\mathcal{L}(H)}\|\nabla f(y^k) - \nabla f(p)\|_H,
\end{aligned}$$

então, pela continuidade da norma, de $T(p)$ e pelo Teorema do Confronto (ver Teorema A.1.10), temos $\|T(y^k)\nabla f(y^k) - T(p)\nabla f(p)\|_H \xrightarrow{k\to\infty} 0$, isto é,

$$T(y^k)\nabla f(y^k) \xrightarrow[\|\cdot\|_H]{k\to\infty} T(p)\nabla f(p).$$

Daí, pela continuidade de f, temos

$$h(y^k) = f(y^k) - f(y^k - \overline{t}T(y^k)\nabla f(y^k)) \xrightarrow{k\to\infty} f(p) - f(p - \overline{t}T(p)\nabla f(p)) = h(p).$$

Logo, h é contínua, donde $h(x^{n_k}) \xrightarrow{k\to\infty} h(\overline{x}) = \beta$ e, portanto, existe $k_0 \in \mathbb{N}$ tal que

$$k \geq k_0 \implies |h(x^{n_k}) - \beta| < \frac{\beta}{2} \implies \frac{\beta}{2} < h(x^{n_k}) = f(x^{n_k}) - f(x^{n_k} + \overline{t}d^{n_k}).$$

Como t_{n_k} foi obtido pela busca exata, podemos concluir que

$$k \geq k_0 \implies f(x^{n_k+1}) = f(x^{n_k} + t_{n_k}d^{n_k}) \leq f(x^{n_k} + \overline{t}d^{n_k}) < f(x^{n_k}) - \frac{\beta}{2},$$

donde

$$k \geq k_0 \implies f(x^{n_k}) - f(x^{n_k+1}) > \frac{\beta}{2}. \tag{7.1}$$

Pela continuidade de f, temos

$$\lim_{k\to\infty} f(x^{n_k}) = f(\overline{x}).$$

Como $\left(f(x^k)\right)$ é uma sequência monótona decrescente, então

$$\lim_{k\to\infty} f(x^k) = f(\overline{x}),$$

o que contradiz (7.1). □

Quando tomamos $T(x^k) := \mathrm{Id}_H$ no Algoritmo 2, obtemos o

Algoritmo 3 (Método do Gradiente)**.** Dado $x^0 \in H$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0_H$

Defina $d^k = -\nabla f(x^k)$

Obtenha $t_k > 0$ tal que $f(x^k + t_k d^k) < f(x^k)$

Faça $x^{k+1} = x^k + t_k d^k$

$k = k + 1$

que é globalmente convergente, como consequência imediata do Teorema 7.3.2.

Lema 7.3.3. (RIBEIRO; KARAS, 2013, Lema 5.1, p. 91) No Algoritmo 3, se t_k é obtido por minimização local de $f(x^k + td^k)$, então $(d^{k+1})^t d^k = 0$.

Demonstração. Definindo $\varphi : \mathbb{R} \to \mathbb{R}$ por

$$\varphi(t) := f(x^k + td^k),$$

temos

$$\varphi'(t_k) = \nabla f(x^k + t_k d^k)^t d^k = \nabla f\left(x^{k+1}\right)^t d^k.$$

Portanto, como a busca é feita por minimização local, temos

$$(d^{k+1})^t d^k = -\nabla f\left(x^{k+1}\right)^t d^k = -\varphi'(t_k) = 0,$$

o que prova a afirmação. □

8 O Problema de Minimização Quadrática

O problema geral de minimizar uma função $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ é dito um problema de **Programação não Linear** (PnãoL) e representamos ele por

$$\text{(PnãoL)} \ \min_{\text{s.a. } x \in X} f(x).$$

A função a ser otimizada, nesse caso, a função f, é chamada de **função objetivo**.

Considere H um espaço de Hilbert e $g : H \to \mathbb{R}$ uma função duas vezes diferenciável em $a \in H$. Pela Fórmula de Taylor com resto infinitesimal (ver Teorema 3.7.2) e pelo Exemplo 3.3.2, temos

$$\begin{aligned} g(a+x) &= g(a) + g'(a)x + \frac{1}{2}g^{(2)}(a)x^{(2)} + o(\|x\|^2) \\ &= g(a) + \langle \nabla g(a), x\rangle + \frac{1}{2}\langle x, \nabla^2 g(a)x\rangle + o(\|x\|^2) \\ &= \underbrace{\frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c}_{f(x)} + o(\|x\|^2), \end{aligned}$$

para algum operador autoadjunto A, algum $b \in H$ e algum $c \in \mathbb{R}$. O problema geral de minimizar uma função $f : H \to \mathbb{R}$ nessas condições, onde $H = \mathbb{R}^n$, pode ser encontrado em (RIBEIRO; KARAS, 2013, p. 38) e trataremos dele formalmente agora.

Considere o problema

$$\text{(PnãoL)} \ \min_{\text{s.a. } x \in \mathbb{R}^n} \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c,$$

onde $A \in \mathcal{L}(\mathbb{R}^n)$ é um operador autoadjunto, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

O problema de PnãoL acima é conhecido como o Problema de Minimização Quadrática e neste capítulo iremos resolvê-lo de dois modos: primeiro faremos uso de técnicas de análise e álgebra linear para mostrar que o problema tem uma única solução global; depois, num segundo momento, mostraremos que o mesmo problema pode ser facilmente resolvido usando diretamente algumas das técnicas que obtemos a partir do cálculo diferencial.

Solução sem o cálculo diferencial. Considere a função $f : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f(x) := \frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c.$$

Se $\overline{x} \in \mathbb{R}^n$ é um minimizador local de f, então existe $\delta > 0$ tal que

$$f(\overline{x}) \leq f(x), \quad \text{para todo } x \in B(\overline{x}, \delta).$$

Dado $d \in \mathbb{R}^n$, temos

$$\begin{aligned} f(\overline{x}+td)-f(\overline{x}) &= \frac{1}{2}\left\langle \overline{x}+td, A(\overline{x}+td)\right\rangle + \langle b, \overline{x}+td\rangle - \frac{1}{2}\langle \overline{x}, A\overline{x}\rangle - \langle b, \overline{x}\rangle \\ &= \frac{t}{2}\left[\langle \overline{x}, Ad\rangle + \langle d, A\overline{x}\rangle\right] + \frac{t^2}{2}\langle d, Ad\rangle + t\langle b, d\rangle \\ &= t\langle A\overline{x}, d\rangle + \frac{t^2}{2}\langle d, Ad\rangle + t\langle b, d\rangle = t\langle A\overline{x}+b, d\rangle + \frac{t^2}{2}\langle d, Ad\rangle, \end{aligned}$$

o que nos dá, para $0 < t < \delta/\|d\|$,

$$0 \leq \frac{f(\overline{x}+td)-f(\overline{x})}{t} = \langle A\overline{x}+b, d\rangle + \frac{t}{2}\langle d, Ad\rangle$$

donde, fazendo $t \to 0^+$, temos, pela Monotonicidade do Limite (ver Teorema B.5.8),

$$\langle A\overline{x}+b, d\rangle \geq 0, \quad \text{para todo} \quad d \in \mathbb{R}^n\,.$$

Em particular, para $d := -(A\overline{x}+b)$, temos

$$0 \geq -\|A\overline{x}+b\|^2 = -\langle A\overline{x}+b, A\overline{x}+b\rangle = -\langle A\overline{x}+b, -d\rangle = \langle A\overline{x}+b, d\rangle \geq 0,$$

donde $\|A\overline{x}+b\| = 0$, logo $A\overline{x} = -b$. Ficamos então com

$$f(\overline{x}+d)-f(\overline{x}) = \frac{1}{2}\langle d, Ad\rangle, \quad \text{para todo} \quad d \in \mathbb{R}^n\,.$$

Agora, se $\|d\| < \delta$, vale

$$\langle d, Ad\rangle = 2(f(\overline{x}+d)-f(\overline{x})) \geq 0.$$

Dado $d \in \mathbb{R}^n$, tome $\varepsilon > 0$ tal que $\varepsilon\|d\| < \delta$. Daí,

$$\varepsilon^2\langle d, Ad\rangle = \langle \varepsilon d, A(\varepsilon d)\rangle \geq 0,$$

donde vem $\langle d, Ad\rangle \geq 0$. Logo,

$$f(\overline{x}+d)-f(\overline{x}) = \frac{1}{2}\langle Ad, d\rangle \geq 0, \quad \text{para todo} \quad d \in \mathbb{R}^n,$$

isto é, $\overline{x}$ é um minimizador global de f.

Provaremos agora que, quando se adiciona a hipótese de que A é definida positiva, então um tal minimizador global de fato existe (e é único).

Pelo Teorema Espectral (ver Teorema C.6.6), podemos encontrar uma base ortonormal $\{v_1, \ldots, v_n\}$ de $\mathbb{R}^n$ formada por autovetores de A. Como

$$0 < \langle v_i, Av_i\rangle = \langle v_i, \lambda_i v_i\rangle = \lambda_i\langle v_i, v_i\rangle = \lambda_i,$$

onde λ_i é o autovalor de A associado ao autovetor v_i, então

$$\begin{aligned} \langle x, Ax\rangle &= \left\langle \sum_{i=1}^n x_i v_i, A\left(\sum_{i=1}^n x_i v_i\right)\right\rangle = \left\langle \sum_{i=1}^n x_i v_i, \sum_{i=1}^n x_i\lambda_i v_i\right\rangle = \sum_{i=1}^n \lambda_i\langle x_i v_i, x_i v_i\rangle \\ &\geq \sum_{i=1}^n \lambda\langle x_i v_i, x_i v_i\rangle = \lambda\|x\|^2, \end{aligned}$$

onde $\lambda := \min\{\lambda_i : i = 1, \ldots, n\}$.

Pela Desigualdade de Cauchy-Schwarz (ver Teorema C.5.2),

$$-\langle b, x\rangle \leq |\langle b, x\rangle| \leq \|b\|\|x\|,$$

donde vem $-\|b\|\|x\| \leq \langle b, x\rangle$, logo

$$\frac{1}{2}\langle x, Ax\rangle + \langle b, x\rangle + c \geq \frac{\lambda}{2}\|x\|^2 - \|b\|\|x\| + c = \|x\|^2\left(\frac{\lambda}{2} - \frac{\|b\|}{\|x\|} + \frac{c}{\|x\|^2}\right)$$

e, portanto, $\lim\limits_{\|x\|\to+\infty} f(x) = +\infty$. Daí, existe $r > 0$ tal que

$$\|x\| > r \implies f(x) > |c| + 1.$$

Por conseguinte,

$$L := \{x \in \mathbb{R}^n : f(x) \leq |c| + 1\} \subseteq B[0, r].$$

Como $0 \in L$ e, pela continuidade de f, L é fechado, então L é um conjunto compacto não vazio. Pelo Teorema de Weierstrass (ver Corolário B.4.9), existe $\overline{x} \in L$ tal que

$$f(\overline{x}) \leq f(x), \quad \text{para todo} \quad x \in L.$$

Se $x \notin L$, então $f(\overline{x}) \leq |c| + 1 < f(x)$. Logo, $\overline{x}$ é uma solução global de (PnãoL).

Para provar a unicidade, recordamos que, na primeira parte desta demonstração, estabelecemos que qualquer minimizador local (e, portanto, qualquer minimizador global) deve satisfazer o sistema linear $A\overline{x} = -b$. Sob a hipótese (utilizada para garantir a existência) de que A é definida positiva, sabemos que $\langle x, Ax\rangle > 0$ para todo $x \neq 0$. Logo, o núcleo $\ker A$ de A é $= \{0\}$, o que garante que A é injetiva (veja Proposição C.2.10). Pelo Corolário C.2.8, temos que A é um isomorfismo linear. Isso prova que existe um único $\overline{x} \in \mathbb{R}^n$ tal que $A\overline{x} = -b$. Como o minimizador global (cuja existência foi garantida pelo Teorema de Weierstrass) deve satisfazer esta equação, concluímos que ele é precisamente esta solução única. Portanto, o minimizador global é único. $\square$

Solução alternativa utilizando o cálculo diferencial. Como f é diferenciável, então os candidatos a extremos de f são seus pontos estacionários, pelo Teorema 5.1.2. Se $x \in \mathbb{R}^n$ é um ponto estacionário de f, então $f'(x) = 0_{E^*}$, donde

$$0 = f'(x)h = \langle Ax + b, h\rangle, \quad \text{para todo} \quad h \in E,$$

donde $Ax + b = 0_{\mathbb{R}^n}$, isto é, $Ax = -b$. Como A é definida positiva, então $\ker(A) = \{0_{\mathbb{R}^n}\}$, logo A é injetiva (veja Proposição C.2.10) e, portanto, um isomorfismo linear (ver Corolário C.2.8). Logo,

$$A\overline{x} = -b, \quad \text{para um único} \quad \overline{x} \in \mathbb{R}^n.$$

Como f é duas vezes diferenciável e

$$f^{(2)}(\overline{x})h^{(2)} = \langle Ah, h\rangle > 0, \quad \text{para todo} \quad h \in E \setminus \{0_E\},$$

então $f^{(2)}(\overline{x})$ é definida positiva e, portanto, $\overline{x}$ é um minimizador local de f, pelo Corolário 5.2.2, e, sendo f convexa (pelo Teorema 6.2.13), o Terema 6.3.1 garante que $\overline{x}$ é um minimizador global de f. □

Conclusão

Neste trabalho, fizemos uma apresentação do cálculo diferencial em espaços normados, abordando desde as definições fundamentais até aplicações em otimização contínua. A formalização rigorosa dos principais teoremas e propriedades permitiu compreender como o conceito de derivada pode ser estendido para espaços normados, inclusive de dimensão infinita, mostrando-se uma ferramenta extremamente útil para a teoria de otimização. Esse trabalho reforça a importância do estudo do cálculo diferencial tanto do ponto de vista teórico quanto aplicado.

Como continuação natural deste estudo, a estrutura teórica aqui desenvolvida pode ser utilizada para analisar métodos de otimização mais complexos, como o Método de Newton em espaços de Banach, que dependem fortemente de derivadas de segunda ordem. Adicionalmente, o ferramental explorado abre portas para o estudo de problemas de controle ótimo, equações diferenciais parciais e cálculo variacional, onde a minimização de funcionais em dimensão infinita é central.

Referências

BOTELHO, G.; PELLEGRINO, D.; TEIXEIRA, E. *Fundamentos de Análise Funcional*. 2. ed. Rio de janeiro: SBM, 2015. v. 13. 409 p. (Coleção Textos Universitários, v. 13). ISBN 978-85-83370-68-0. Citado na página 179.

BUENO, H. P. *Álgebra Linear: Um Segundo Curso*. 1. ed. Rio de Janeiro: SBM, 2006. 313 p. (Coleção Textos Universitários). ISBN 85-85818-31-X. Citado 3 vezes nas páginas 149, 177 e 179.

CIPOLATTI, R. *Cálculo Avançado*. 2. ed. Rio de janeiro: SBM, 2020. v. 20. 364 p. (Coleção Textos Universitários, v. 20). ISBN 978-65-990395-6-0. Citado 2 vezes nas páginas 13 e 60.

COLEMAN, R. *Calculus on Normed Vector Spaces*. 1. ed. [S.l.]: Springer, 2012. 249 p. (Universitext). ISBN 978-1-4614-3894-6. Citado 26 vezes nas páginas 13, 17, 20, 22, 35, 37, 39, 40, 41, 42, 43, 44, 52, 58, 59, 60, 63, 67, 68, 70, 71, 73, 178, 190, 193 e 194.

IZMAILOV, A.; SOLODOV, M. *Otimização - volume 1, Condições de otimalidade, elementos de análise convexa e de dualidade*. 2. ed. Rio de janeiro: IMPA, 2009. 253 p. ISBN 978-85-244-0238-8. Citado na página 82.

KREYSZIG, E. *Introductory Functional Analysis with Applications*. [S.l.]: Wiley Classics Library, 1989. 704 p. ISBN 0-471-50731-8. Citado 3 vezes nas páginas 179, 185 e 187.

LIMA, E. L. *Análise no espaço* $\mathbb{R}^n$. Rio de Janeiro: IMPA, 2002. 128 p. (Coleção Matemática Universitária). ISBN 85-244-0189-3. Citado 7 vezes nas páginas 14, 15, 21, 60, 65, 66 e 177.

LIMA, E. L. *Elementos de Topologia Geral*. 3. ed. Rio de janeiro: SBM, 2014. 297 p. (Textos Universitários). ISBN 978-85-85818-43-2. Citado na página 133.

LIMA, E. L. *Espaços Métricos*. 5. ed. Rio de janeiro: IMPA, 2017. 337 p. (Projeto Euclides). ISBN 978-85-244-0158-9. Citado 2 vezes nas páginas 133 e 147.

LIMA, E. L. *Álgebra Linear*. 9. ed. Rio de Janeiro: IMPA, 2018. 357 p. (Coleção Matemática Universitária). ISBN 978-85-244-0420-7. Citado 4 vezes nas páginas 149, 174, 176 e 179.

LIMA, E. L. *Curso de Análise*. 15. ed. Rio de janeiro: IMPA, 2019. v. 1. 320 p. (Projeto Euclides, v. 1). ISBN 978-85-244-0468-9. Citado 5 vezes nas páginas 39, 40, 115, 118 e 148.

MENDELSON, B. *Introduction to Topology*. 3. ed. [S.l.]: Dover Publications, 1990. 206 p. ISBN 978-0-486-66352-4. Citado na página 133.

RIBEIRO, A. A.; KARAS, E. W. *Otimização contínua: Aspectos teóricos e computacionais*. 1. ed. São Paulo: Cengage Learning, 2013. 300 p. ISBN 978-85-221-1501-3. Citado 8 vezes nas páginas 82, 84, 102, 103, 105, 106, 107 e 108.

Apêndices

APÊNDICE A – Números reais

Este capítulo é baseado livremente em (LIMA, 2019).

A.1 Sequências numéricas

Definição A.1.1 (Sequência)**.** Uma **sequência** de elementos em um conjunto X é uma aplicação $x : \mathbb{Z}_{\geq p} \to X$, para algum $p \in \mathbb{Z}$.

O valor de uma sequência x em um ponto n é usualmente denotado por x_n, em vez de $x(n)$, enquanto a sequência x é denotada por $(x_n)_{n=p}$ (ou $(x_n)_{n=p}^{\infty}$, ou $(x_n)_{n\in\mathbb{Z}_{\geq p}}$, ou ainda (x_n), se é claro o ponto p onde a sequência começa). Quando o ponto p não é especificado, subentende-se que $p = 1$.

Quando escrevermos $(x_n) \subseteq X$ queremos dizer que (x_n) é uma sequência tal que

$$x_n \in X, \text{ para todo } n.$$

Definição A.1.2. Dizemos que $x : \mathbb{Z}_{\geq p} \to X$ é uma **subsequência** de $y : \mathbb{Z}_{\geq p} \to X$ se existe uma aplicação crescente $\varphi : \mathbb{Z}_{\geq p} \to \mathbb{Z}_{\geq p}$ tal que $x = y \circ \varphi$.

Por hora, trabalharemos sobre um corpo ordenado $\mathbb{K}$ fixado.

Definição A.1.3 (Convergência)**.** Dizemos que $(x_n) \subseteq \mathbb{K}$ **converge** para $L \in \mathbb{K}$ se, para cada $\varepsilon > 0$, podemos obter $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies |x_n - L| < \varepsilon.$$

Se (x_n) converge para L, escrevemos $x_n \xrightarrow{n\to\infty} L$ ou $\lim\limits_{n\to\infty} x_n = L$. Uma sequência que converge é dita **convergente**, enquanto uma sequência que não converge é dita **divergente**.

Definição A.1.4 (Limites infinitos)**.** Se $(x_n) \subseteq \mathbb{K}$ é tal que, para todo $A > 0$, existe $N \in \mathbb{N}$ tal que

$$n \geq N \implies x_n \geq A,$$

dizemos que "x_n tende para mais infinito" e escrevemos $x_n \xrightarrow{n\to\infty} +\infty$ ou $\lim\limits_{n\to\infty} x_n = +\infty$. Se $(-x_n)$ é tal que $\lim\limits_{n\to\infty} -x_n = +\infty$, dizemos que "$x_n$ tende para menos infinito" e escrevemos $x_n \xrightarrow{n\to\infty} -\infty$ ou $\lim\limits_{n\to\infty} x_n = -\infty$.

Teorema A.1.5 (Unicidade do limite)**.** Sejam $(x_n) \subseteq \mathbb{K}$ e $a, b \in \mathbb{K}$. Se $x_n \xrightarrow{n\to\infty} a$ e $x_n \xrightarrow{n\to\infty} b$, então $a = b$.

Demonstração. Suponhamos, por redução ao absurdo, que $a \neq b$. Assim, tomando

$$\varepsilon := \frac{|a-b|}{2},$$

existem $n_1, n_2 \in \mathbb{N}$ tais que

$$\begin{cases} n \geq n_1 & \Longrightarrow |x_n - a| < \varepsilon, \\ n \geq n_2 & \Longrightarrow |x_n - b| < \varepsilon. \end{cases}$$

Desse modo,

$$n \geq \max\{n_1, n_2\} \implies x_n \in (a-\varepsilon, a+\varepsilon) \cap (b-\varepsilon, b+\varepsilon) = \varnothing.$$

Absurdo! □

Definição A.1.6. Diremos que $(x_n) \subseteq \mathbb{K}$ é uma **sequência de Cauchy** se, para cada $\varepsilon > 0$, podemos obter $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies |x_m - x_n| < \varepsilon.$$

Teorema A.1.7. Toda sequência convergente de $\mathbb{K}$ é de Cauchy.

Demonstração. Sejam $(x_n) \subseteq \mathbb{K}$ e $a \in \mathbb{K}$ tais que $\lim\limits_{n\to\infty} x_n = a$. Daí, dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies |x_n - a| < \frac{\varepsilon}{2}.$$

Por conseguinte,

$$m, n \geq n_0 \implies |x_m - x_n| = |x_m - a + a - x_n| \leq |x_m - a| + |x_n - a| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Logo, (x_n) é de Cauchy. □

Teorema A.1.8. Toda sequência de Cauchy de $\mathbb{K}$ é limitada.

Demonstração. Seja (x_n) uma sequência de Cauchy. Daí, existe $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies |x_m - x_n| < 1.$$

Logo,

$$m \geq n_0 \implies |x_m - x_{n_0}| < 1 \implies x_{n_0} - 1 < x_m < x_{n_0} + 1.$$

Pondo $A := \{x_1, x_2, \ldots, x_{n_0 - 1}\}$, temos

$$\min\{x_{n_0} - 1, \min A\} \leq x_n \leq \max\{x_{n_0} + 1, \max A\},$$

para todo $n \in \mathbb{N}$. □

Teorema A.1.9. Se $(x_n), (y_n) \subseteq \mathbb{K}$ e (x_n) é uma sequência limitada e $\lim\limits_{n\to\infty} y_n = 0$, então $\lim\limits_{n\to\infty} x_n y_n = 0$.

Demonstração. Como (x_n) é limitada, então existe $c > 0$ tal que $|x_n| \leq c$, para todo $n \in \mathbb{N}$. Seja $\varepsilon > 0$. Como $\lim\limits_{n\to\infty} y_n = 0$, então existe $n_0 \in \mathbb{N}$ tal que $n \geq n_0 \implies |y_n| < \dfrac{\varepsilon}{c}$. Desse modo,

$$n \geq n_0 \implies |x_n \cdot y_n| = |x_n| \cdot |y_n| < c \cdot \frac{\varepsilon}{c} = \varepsilon.$$

Portanto, $\lim\limits_{n\to\infty} x_n y_n = 0$. □

Teorema A.1.10 (Teorema do Confronto)**.** Sejam $(x_n), (y_n), (z_n)$ sequências de números reais tais que $x_n \leq y_n \leq z_n$, para todo $n \in \mathbb{N}$, e $a \in \mathbb{R}$ tal que $a = \lim\limits_{n\to\infty} x_n = \lim\limits_{n\to\infty} z_n$. Nessas condições, $\lim\limits_{n\to\infty} y_n = a$.

Demonstração. Se $\varepsilon > 0$, então existem $n_1, n_2 \in \mathbb{N}$ tais que

$$\begin{cases} n \geq n_1 \implies |x_n - a| < \varepsilon \implies a - \varepsilon < x_n < a + \varepsilon \\ n \geq n_2 \implies |z_n - a| < \varepsilon \implies a - \varepsilon < z_n < a + \varepsilon \end{cases}.$$

Tomando $n_0 := \max\{n_1, n_2\}$, temos

$$n \geq n_0 \implies a - \varepsilon < x_n \leq y_n \leq z_n < a + \varepsilon \implies |y_n - a| < \varepsilon.$$

Logo, $\lim\limits_{n\to\infty} y_n = a$. □

Teorema A.1.11 (Monotonicidade do Limite)**.** Se $(x_n), (y_n) \subseteq \mathbb{K}$ são sequências convergentes, então

$$x_n \leq y_n, \text{ para todo } n \in \mathbb{N} \implies \lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n.$$

Demonstração. Suponhamos, por redução ao absurdo, que seja $b < a$ e tome $\varepsilon := \dfrac{a-b}{2}$, onde $a := \lim\limits_{n\to\infty} x_n$ e $b := \lim\limits_{n\to\infty} y_n$. Assim, existem $n_1, n_2 \in \mathbb{N}$ tais que

$$\begin{cases} n \geq n_1 \implies |x_n - a| < \varepsilon, \\ n \geq n_2 \implies |y_n - b| < \varepsilon. \end{cases}$$

Tomando $n_0 := \max\{n_1, n_2\}$, temos

$$n \geq n_0 \implies y_n < b + \varepsilon = a - \varepsilon < x_n,$$

donde $y_{n_0} < x_{n_0}$, o que é absurdo. □

Teorema A.1.12 (Recíproca da Monotonicidade do Limite)**.** Se $(x_n), (y_n) \subseteq \mathbb{K}$ são sequências convergentes, então

$$\lim_{n\to\infty} x_n < \lim_{n\to\infty} y_n \implies \text{existe } n_0 \in \mathbb{N} \text{ tal que } x_n < y_n, \text{ para todo } n \geq n_0.$$

Demonstração. Tomando $\varepsilon := \dfrac{b-a}{2}$, onde $a := \lim\limits_{n\to\infty} x_n$ e $b := \lim\limits_{n\to\infty} y_n =$, existem $n_1, n_2 \in \mathbb{N}$ tais que

$$\begin{cases} n \geq n_1 & \implies |x_n - a| < \varepsilon, \\ n \geq n_2 & \implies |y_n - b| < \varepsilon. \end{cases}$$

Tomando $n_0 := \max\{n_1, n_2\}$, temos $n \geq n_0 \implies x_n < a + \varepsilon = b - \varepsilon < y_n$. □

Teorema A.1.13 (Propriedades aritméticas do limite)**.** Se $\lim\limits_{n\to\infty} x_n = a$ e $\lim\limits_{n\to\infty} y_n = b$, então

1. $\lim\limits_{n\to\infty} (x_n + y_n) = a + b$.
2. $\lim\limits_{n\to\infty} (x_n \cdot y_n) = a \cdot b$.
3. Se $b \neq 0$ e existe $n_1 \in \mathbb{N}$ tal que $n \geq n_1 \implies y_n \neq 0$, então $\lim\limits_{n\to\infty} \dfrac{1}{y_n} = \dfrac{1}{b}$.

Demonstração.

1. Dado $\varepsilon > 0$, existem $n_1, n_2 \in \mathbb{N}$ tais que

$$\begin{cases} n \geq n_1 & \implies |x_n - a| < \varepsilon/2, \\ n \geq n_2 & \implies |y_n - b| < \varepsilon/2. \end{cases}$$

   Tomando $n_0 := \max\{n_1, n_2\}$, temos

$$n \geq n_0 \implies |(x_n + y_n) - (a+b)| = |(x_n - a) + (y_n - b)| \leq |x_n - a| + |y_n - b| < \varepsilon.$$

2. $\lim\limits_{n\to\infty} (x_n y_n - ab) = \lim\limits_{n\to\infty} (x_n y_n - x_n b + x_n b - ab) = \lim\limits_{n\to\infty} [x_n(y_n - b) + b(x_n - a)] = 0.$
3. Como $y_n \xrightarrow{n\to\infty} b$, então $by_n \xrightarrow{n\to\infty} b^2$, donde existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies |by_n - b^2| < \frac{b^2}{2} \implies \frac{b^2}{2} < by_n \implies \frac{1}{by_n} < \frac{2}{b^2}.$$

   Logo,

$$\lim_{n\to\infty} \left(\frac{1}{y_n} - \frac{1}{b}\right) = \lim_{n\to\infty} \frac{b - y_n}{by_n} = \lim_{n\to\infty} (b - y_n) \cdot \frac{1}{by_n} = 0,$$

   já que $\left(\dfrac{1}{by_n}\right)$ é limitada e $\lim\limits_{n\to\infty} (b - y_n) = 0$.

Isso fecha todos os itens. □

Para a prova do teorema abaixo, consulte (LIMA, 2019, Teorema 14, p. 130).

Teorema A.1.14 (Propriedades aritiméticas de limites infinitos)**.**

1. Se $\lim\limits_{n\to\infty} x_n = +\infty$ e (y_n) é uma sequência limitada inferiormente, então $\lim(x_n+y_n) = +\infty$;

2. Se $\lim x_n = +\infty$ e existe $c > 0$ tal que $y_n > c$, para todo $n \in \mathbb{N}$, então $\lim x_n y_n = +\infty$;

3. Seja $x_n > 0$ para todo n. Então
$$\lim x_n = 0 \iff \lim 1/x_n = +\infty;$$

4. Sejam (x_n) e (y_n) sequências de números positivos. Então:

   a) se existe $c > 0$ tal que $x_n > c$ para todo n e se $\lim y_n = 0$, tem-se $\lim \dfrac{x_n}{y_n} = +\infty$;

   b) se (x_n) é limitada e $\lim y_n = +\infty$, então $\lim \dfrac{x_n}{y_n} = 0$.

A.2 Números reais

Continuamos aqui com a hipótese de que $\mathbb{K}$ é um corpo ordenado. Seja $A \subseteq \mathbb{K}$. Dizemos que $c \in \mathbb{K}$ é uma **cota superior** de A se
$$a \leq c, \quad \text{para todo} \quad a \in A.$$

O **supremo** de A é, se existir, o menor $c \in \mathbb{K}$ que é cota superior para A e o denotamos por $\sup A$. Dizemos que A é **limitado superiormente** se existe $c \in \mathbb{K}$ cota superior para A. Se A não é limitado superiormente, escrevemos $\sup A = +\infty$. De modo análogo, A é **limitado inferiormente** se existe $c \in \mathbb{K}$ tal que
$$c \leq a, \quad \text{para todo} \quad a \in A.$$

Todo c nessa condição é chamado **cota inferior** e o maior tal c, se existir, é chamado **ínfimo** de A. Se A não é limitado inferiormente, escrevemos $\inf A = -\infty$. Observe ainda que, com essas definições, o conjunto vazio é limitado inferior e superiormente, mas não tem ínfimo nem supremo. Nesse caso é convencional definir $\inf \varnothing := +\infty$ e $\sup \varnothing := -\infty$.

Já provamos que os limites têm certas propriedades aritméticas, que se estendem até certo ponto, aos limites infinitos (as únicas exceções são os limites que resultam em $+\infty - \infty$ e ∞/∞, que são expressões indeterminadas). Adotaremos a convenção de que
$$-\infty < a < +\infty, \quad \text{para todo} \quad a \in \mathbb{K}\,.$$

Escreveremos $\overline{\mathbb{K}}$ ou $[-\infty, +\infty]$ para denotar $\mathbb{K} \cup \{-\infty, +\infty\}$.

Dados X e Y conjuntos parcialmente ordenados, dizemos que $f : X \to Y$ é

1. **crescente** se $x < y \implies f(x) < f(y)$;

2. **não decrescente** se $x \leq y \implies f(x) \leq f(y)$;

3. **decrescente** se $x < y \implies f(y) < f(x)$; e

4. **não crescente** se $x \leq y \implies f(y) \leq f(x)$.

Uma função f é dita **monótona** se satisfaz qualquer uma dessas condições. Observe que funções crescentes são não decrescentes e funções decrescentes são não crescentes. Assim, a fim de verificar que uma propriedade é válida para funções monótonas, é necessário e suficiente verificar que ela é válida para funções não decrescentes e não crescentes.

Teorema A.2.1. Se $\mathbb{K}$ é um corpo arquimediano, então toda sequência monótona e limitada em $\mathbb{K}$ é de Cauchy.

Demonstração. Seja $(x_n) \subseteq \mathbb{K}$ uma sequência monótona não decrescente e limitada superiormente. Evidentemente, se (x_n) possui um elemento máximo x_{n_0}, então

$$x_{n_0} \leq x_{n_0+1} \leq x_{n_0+2} \leq \cdots \leq x_{n_0+k} \leq x_{n_0},$$

para cada $k \in \mathbb{N}$, donde $\lim\limits_{n\to\infty} x_n = x_{n_0}$ e, portanto, (x_n) é de Cauchy. Suponhamos, assim, que (x_n) não possui elemento máximo.

Afirmação. Dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$x_n \leq x_{n_0} + \varepsilon, \text{ para todo } n \in \mathbb{N}.$$

Demonstração da afirmação. Suponhamos, por redução ao absurdo, que existe $\varepsilon_0 > 0$ tal que $x_n + \varepsilon_0$ não limita superiormente (x_n), para cada $n \in \mathbb{N}$. Como $x_1 + \varepsilon_0$ não limita superiormente (x_n), então existe $n_1 \in \mathbb{N}$ tal que $x_1 + \varepsilon_0 < x_{n_1}$. Como $x_{n_1} + \varepsilon_0$ não limita superiormente (x_n), então existe $n_2 \in \mathbb{N}$ tal que $x_{n_1} + \varepsilon_0 < x_{n_2}$, donde $x_1 + 2\varepsilon_0 < x_{n_2}$. Como $x_{n_2} + \varepsilon_0$ não limita superiormente (x_n), então existe $n_3 \in \mathbb{N}$ tal que $x_{n_2} + \varepsilon_0 < x_{n_3}$. Desse modo, $x_1 + 3\varepsilon_0 < x_{n_3}$. Prosseguindo dessa maneira, obtemos uma subsequência $(x_{n_k})_k$ de (x_n) tal que $x_1 + k\varepsilon_0 < x_{n_k}$. Como $\mathbb{K}$ é arquimediano, então, dado $r > 0$, podemos obter $k_r \in \mathbb{N}$ tal que $k_r > \dfrac{r - x_1}{\varepsilon_0}$, donde $x_{n_{k_r}} > x_1 + k_r\varepsilon_0 > r$ e, portanto, $(x_{n_k})_k$ não é limitada superiormente. Absurdo!

Fixe $\varepsilon > 0$ e tome $n_0 \in \mathbb{N}$ tal que $x_{n_0} + \varepsilon/2$ limita superiormente (x_n). Sejam $m \geq n \geq n_0$. Então

$$x_{n_0} \leq x_n \leq x_m \leq x_{n_0} + \varepsilon/2,$$

donde vem

$$|x_m - x_n| = x_m - x_n \leq (x_{n_0} + \varepsilon/2) - x_n \leq (x_{n_0} + \varepsilon/2) - x_{n_0} = \varepsilon/2 < \varepsilon.$$

Logo, (x_n) é de Cauchy, já que

$$m, n \geq n_0 \implies |x_m - x_n| < \varepsilon.$$

Se $(y_n) \subseteq \mathbb{K}$ é uma sequência monótona não crescente e limitada inferiormente, então $(-y_n)$ é uma sequência monótona não decrescente e limitada superiormente, donde $(-y_n)$ é de Cauchy e, portanto, (y_n) é de Cauchy. □

Teorema A.2.2. Se uma subsequência de uma sequência de Cauchy de $\mathbb{K}$ converge, então a sequência inteira converge.

Demonstração. Sejam (x_n) uma sequência de Cauchy e $(x_{n_k})_k$ uma subsequência convergente. Sejam $a := \lim\limits_{k\to\infty} x_{n_k}$ e $\varepsilon > 0$. Como (x_n) é de Cauchy, então existe $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies |x_m - x_n| < \frac{\varepsilon}{2}.$$

Fixe $n \geq n_0$. Desse modo, $|x_m - x_n| < \varepsilon/2$, para todo $m \geq n_0$. Como $(n_k)_k$ é uma sequência de números naturais ilimitada superiormente, então existe $k_0 \in \mathbb{N}$ tal que $n_{k_0} \geq n_0$. Portanto,

$$k \geq k_0 \implies n_k \geq n_0 \implies |x_{n_k} - x_n| < \frac{\varepsilon}{2}.$$

Pela Monotonicidade do Limite (ver Teorema A.1.11),

$$|a - x_n| = \lim_{k\to\infty} |x_{n_k} - x_n| \leq \lim_{k\to\infty} \frac{\varepsilon}{2} = \frac{\varepsilon}{2} < \varepsilon.$$

Logo, $\lim\limits_{n\to\infty} x_n = a$. □

Definição A.2.3 (Completude no sentido de Cauchy)**.** Um corpo ordenado $\mathbb{K}$ é **Cauchy-completo** se

$$(x_n) \subseteq \mathbb{K} \text{ é de Cauchy} \implies \text{existe } L \in \mathbb{K} \text{ tal que } x_n \xrightarrow{n\to\infty} L.$$

Definição A.2.4 (Completude no sentido de Dedekind)**.** Um corpo ordenado $\mathbb{K}$ é **Dedekind-completo** se todo subconjunto $X \subseteq \mathbb{K}$ não vazio e limitado superiormente tem cota superior mínima em $\mathbb{K}$, isto é, existe o supremo de X em K.

Teorema A.2.5. Se $\mathbb{K}$ é um corpo arquimediano e Cauchy-completo, então $\mathbb{K}$ é Dedekind-completo.

Demonstração. Seja $X \subseteq \mathbb{K}$ não vazio e limitado superiormente em $\mathbb{K}$. Assim, existem $x_1 \in X$ e $y_1 \in \mathbb{K}$ tal que y_1 é cota superior de X. Provaremos que

$$\sup X = \lim_{n\to\infty} x_n = \lim_{n\to\infty} y_n,$$

onde

$$\begin{cases} x_{n+1} & := \begin{cases} \dfrac{x_n + y_n}{2}, & \text{se } \dfrac{x_n + y_n}{2} \text{ não é cota superior de } X, \\ x_n, & \text{caso contrário,} \end{cases} \\ y_{n+1} & := \begin{cases} \dfrac{x_n + y_n}{2}, & \text{se } \dfrac{x_n + y_n}{2} \text{ é cota superior de } X, \\ y_n, & \text{caso contrário.} \end{cases} \end{cases}$$

Facilmente se verifica que, para cada $n \in \mathbb{N}$,

$$x_n \leq x_{n+1} \leq y_{n+1} \leq y_n.$$

Como $\mathbb{K}$ é arquimediano e (x_n) e (y_n) são monótonas e limitadas, então (x_n) e (y_n) são de Cauchy. Como $\mathbb{K}$ é completo, então existem $M, N \in \mathbb{K}$ tais que $\lim\limits_{n\to\infty} x_n = M$ e $\lim\limits_{n\to\infty} y_n = N$.

Sejam

$$A := \{n \in \mathbb{N} : x_{n+1} = x_n\} \text{ e } B := \left\{n \in \mathbb{N} : y_{n+1} = \frac{x_n + y_n}{2}\right\}.$$

Note que $A = B$.

Se A for finito, então existe $n_0 \in \mathbb{N}$ tal que $k \geq n_0 \implies x_{k+1} \neq x_k$. Daí,

$$k \geq n_0 \implies x_{k+1} - x_k = \frac{x_k + y_k}{2} - x_k = \frac{y_k - x_k}{2} \implies \frac{N - M}{2} = 0 \implies M = N.$$

Se A for infinito, então B é infinito, já que $B = A$. Assim,

$$y_{n_k+1} - y_{n_k} = \frac{x_{n_k} + y_{n_k}}{2} - y_{n_k} = \frac{x_{n_k} - y_{n_k}}{2}.$$

Logo,

$$\frac{M - N}{2} = \lim_{k\to\infty} \frac{x_{n_k} - y_{n_k}}{2} = \lim_{k\to\infty} (y_{n_k+1} - y_{n_k}) = 0 \implies M = N.$$

Dado $x \in X$, temos $x \leq y_n$, para todo $n \in \mathbb{N}$, donde $x \leq N$, pela Monotonicidade do Limite (ver Teorema A.1.11) e, portanto, N é cota superior de X. Seja $\varepsilon > 0$. Como $\lim\limits_{n\to\infty} x_n = M = N$, então existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies |x_n - N| < \varepsilon \implies N - \varepsilon < x_n.$$

Como x_{n_0} não é uma cota superior de X, então existe $x \in X$ tal que $N - \varepsilon < x_{n_0} < x$ e, portanto, N é a menor das cotas superiores de X, isto é, $N = \sup X$. $\square$

A partir de agora assumiremos que existe um corpo arquimediano que é Cauchy-completo (e, portanto, Dedekind-completo) que denotaremos por $\mathbb{R}$.

Corolário A.2.6. Se $(x_n) \subseteq \mathbb{R} \cup \{-\infty, +\infty\}$ é uma sequência monótona, então $\lim x_n = L$, para algum $L \in \mathbb{R} \cup \{-\infty, +\infty\}$.

Demonstração. Se (x_n) é limitada, então $(x_n) \subseteq \mathbb{R}$ e é de Cauchy, pelo Teorema A.2.1, logo é convergente.

Suponhamos agora que (x_n) não é limitada. Como ela é monótona, suponhamos que ela é não decrescente. Então, como é ilimitada e é limitada inferiormente, ela é ilimitada superiormente. Então, dado $A > 0$, podemos encontrar $N \in \mathbb{N}$ tal que $x_N \geq A$. A hipótese de que (x_n) é não decrescente nos dá

$$n \geq N \implies x_n \geq A.$$

O outro caso é análogo. $\square$

Observe ainda que, no teorema anterior, se (x_n) é não decrescente, então $L = \sup x_n$; e se é não crescente, então $L = \inf x_n$.

Teorema A.2.7 (Princípio dos Intervalos Encaixados)**.** Se $([a_n, b_n])_{n\in\mathbb{N}}$ é uma sequência não crescente (considerando a relação de ordem como a relação de inclusão) de intervalos de números reais, então

$$\bigcap_{n\in\mathbb{N}} [a_n, b_n] = \left[\sup_{n\in\mathbb{N}} a_n, \inf_{n\in\mathbb{N}} b_n\right].$$

Demonstração. Como $a_1 \leq a_2 \leq \cdots \leq a_n \leq b_n \leq \cdots \leq b_2 \leq b_1$, para todo $n \in \mathbb{N}$, então b_1 é uma cota superior de $A := \{a_n : n \in \mathbb{N}\}$ e a_1 é uma cota inferior de $B := \{b_n : n \in \mathbb{N}\}$, donde $c := \sup A$ e $d := \inf B$ existem enquanto números reais. Além disso, como cada b_n é cota superior de A, então $c \leq b_n$, para todo $n \in \mathbb{N}$, donde $c \leq \inf B = d$. Logo, $a_n \leq c \leq d \leq b_n$, para cada $n \in \mathbb{N}$, e, portanto, $\bigcap_{n\in\mathbb{N}} [a_n, b_n] = [c, d]$. $\square$

Encerramos essa seção com um teorema que será utilizado no capítulo 3.

No que se segue, $[n] := \{1, \ldots, n\}$, para todo $n \in \mathbb{N}$.

Teorema A.2.8. Sejam $X_1, \ldots, X_n$ conjuntos não vazios, $X := \prod_{i\in[n]} X_i$ e $f : X \to \mathbb{R}$ uma função limitada superiormente. Então, para toda permutação $\sigma \in S_n$, temos

$$\sup_{\substack{x_i \in X_i,\\ \text{para todo } i\in[n]}} f(x_1, \ldots, x_n) = \sup_{x_{\sigma(1)}\in X_{\sigma(1)}} \cdots \sup_{x_{\sigma(n)}\in X_{\sigma(n)}} f(x_1, \ldots, x_n).$$

Demonstração. O caso $n = 1$ é óbvio. Suponhamos, por hipótese de indução, que o resultado é válido para $n = k$ e consideremos o caso $n = k + 1$. Fixado $j \in [k + 1]$, para todo $y \in X_j$, temos

$$\alpha := \sup_{\substack{x_i \in X_i,\\ \text{para todo } i\in[k+1]}} f(x_1, \ldots, x_{k+1}) \geq \sup_{\substack{x_i \in X_i,\\ \text{para todo } i\in[k+1]\setminus\{j\}}} f(x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_{k+1}),$$

donde

$$\alpha \geq \sup_{x_j \in X_j} \sup_{\substack{x_i \in X_i, \\ \text{para todo } i \in [k+1] \setminus \{j\}}} f(x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_{k+1}).$$

Dado $\varepsilon > 0$, existe $(y_1, \ldots, y_{k+1}) \in X$ tal que

$$\alpha - \varepsilon < f(y_1, \ldots, y_{k+1}) \leq \sup_{\substack{x_i \in X_i, \\ \text{para todo } i \in [k+1] \setminus \{j\}}} f(x_1, \ldots, x_{j-1}, y_j, x_{j+1}, \ldots, x_{k+1}) \leq \alpha.$$

Portanto,

$$\alpha = \sup_{x_j \in X_j} \sup_{\substack{x_i \in X_i, \\ \text{para todo } i \in [k+1] \setminus \{j\}}} f(x_1, \ldots, x_{k+1}).$$

Ponha

$$Y_p := \begin{cases} X_p, & \text{se } 1 \leq p \leq j-1, \\ X_{p+1}, & \text{se } j \leq p \leq k, \end{cases}$$

e

$$g_{x_j}(z_1, \ldots, z_k) := f(z_1, \ldots, z_{j-1}, x_j, z_j, z_{j+1}, \ldots, z_k),$$

para quaisquer $z_i \in Y_i$ e $i \in [k]$, e $x_j \in X_j$. Então, da hipótese de indução, dada uma permutação $\beta \in S_k$, temos

$$\begin{aligned} \alpha &= \sup_{x_j \in X_j} \sup_{\substack{x_i \in X_i, \\ \text{para todo } i \in [k+1] \setminus \{j\}}} f(x_1, \ldots, x_{k+1}) \\ &= \sup_{x_j \in X_j} \sup_{\substack{z_i \in Y_i, \\ \text{para todo } i \in [k]}} g_{x_j}(z_1, \ldots, z_k) \\ &= \sup_{x_j \in X_j} \sup_{z_{\beta(1)} \in Y_{\beta(1)}} \ldots \sup_{z_{\beta(k)} \in Y_{\beta(k)}} g_{x_j}(z_1, \ldots, z_k) \\ &= \sup_{x_j \in X_j} \sup_{z_{\beta(1)} \in Y_{\beta(1)}} \ldots \sup_{z_{\beta(k)} \in Y_{\beta(k)}} f(z_1, \ldots, z_{j-1}, x_j, z_j, z_{j+1}, \ldots, z_k). \end{aligned}$$

Daí, dada uma permutação $\sigma \in S_{k+1}$, pondo $j := \sigma(1)$, temos

$$\{Y_1, \ldots, Y_k\} = \{X_{\sigma(2)}, \ldots, X_{\sigma(k+1)}\},$$

donde existe pelo menos[1] uma permutação $\beta \in S_k$ tal que

$$Y_{\beta(p)} = X_{\sigma(p+1)},$$

para cada $p \in [k]$. Isso nos dá

$$\alpha = \sup_{x_{\sigma(1)} \in X_{\sigma(1)}} \ldots \sup_{x_{\sigma(k+1)} \in X_{\sigma(k+1)}} f(x_1, \ldots, x_{k+1}).$$

Logo, o resultado é válido para $n = k + 1$. □

[1] Existe exatamente uma se todos os Y_i's são distintos e mais de uma se alguns são iguais.

A.3 Valores de aderência: o $\liminf$ e o $\limsup$

Seja $(x_n) \subseteq \mathbb{R} \cup \{-\infty, +\infty\}$. Então, para cada $k \in \mathbb{N}$,

$$\inf_{n\in\mathbb{N}} x_n \leq \inf_{n\geq k} x_n \leq \inf_{n\geq k+1} x_n \leq \sup_{n\geq k+1} x_n \leq \sup_{n\geq k} x_n \leq \sup_{n\in\mathbb{N}} x_n.$$

Assim, $a_k := \inf\limits_{n\geq k} x_n$ é monótona não decrescente e $b_k := \sup\limits_{n\geq k} x_n$ é monótona não crescente. Por conseguinte,

$$\lim_{k\to\infty} \inf_{n\geq k} x_n = \lim_{k\to\infty} a_k = \sup_{k\in\mathbb{N}} a_k = \sup_{k\in\mathbb{N}} \inf_{n\geq k} x_n$$
$$\lim_{k\to\infty} \sup_{n\geq k} x_n = \lim_{k\to\infty} b_k = \inf_{k\in\mathbb{N}} b_k = \inf_{k\in\mathbb{N}} \sup_{n\geq k} x_n.$$

Definição A.3.1 (Limite inferior e limite superior)**.** O **limite inferior** de $(x_n) \subseteq \mathbb{R} \cup \{-\infty, +\infty\}$ é

$$\liminf_{n\to\infty} x_n := \lim_{k\to\infty} \inf_{n\geq k} x_n,$$

enquanto o **limite superior** de (x_n) é

$$\limsup_{n\to\infty} x_n := \lim_{k\to\infty} \sup_{n\geq k} x_n.$$

Definição A.3.2 (Valor de aderência)**.** Se $(x_n) \subseteq \mathbb{R} \cup \{-\infty, +\infty\}$, dizemos que $a \in \mathbb{R} \cup \{-\infty, +\infty\}$ é um **valor de aderência** de (x_n) se (x_n) possui alguma subsequência (x_{n_k}) tal que $\lim\limits_{k\to\infty} x_{n_k} = a$.

Teorema A.3.3. O menor valor de aderência de uma sequência é o limite inferior e o maior, o limite superior.

Demonstração. Sejam $(x_n) \subseteq \mathbb{R} \cup \{-\infty, +\infty\}$ uma sequência e $L := \liminf x_n$. Vamos provar inicialmente que L é um valor de aderência de (x_n). Seja $a_k := \inf_{n\geq k} x_n$.

Se $L = -\infty$, então $\sup a_k = -\infty$, donde $a_k = -\infty$, para todo $k \in \mathbb{N}$. Em particular, como $a_1 = -\infty$, existe $n_1 \in \mathbb{N}$ tal que $x_{n_1} \leq -1$; como $a_{n_1+1} = -\infty$, existe $n_2 > n_1$ tal que $x_{n_2} \leq -2$; e assim por diante. Isso nos dá uma subsequência (x_{n_k}) tal que

$$-\infty \leq x_{n_k} \leq -k, \quad \text{para todo} \quad k \in \mathbb{N}\,.$$

Logo, pelo Teorema do Confronto, $\lim x_{n_k} = -\infty$.

Agora, se $L = +\infty$, então $a_k \to +\infty$. Dado $M > 0$, escolha $k_0 \in \mathbb{N}$ tal que $a_{k_0} \geq M$. Daí,

$$n \geq k_0 \implies x_n \geq a_{k_0} \geq M,$$

provando que $\lim x_n = +\infty$.

Por fim, suponhamos agora $L \in \mathbb{R}$. Nesse caso, temos $\sup a_k = L \in \mathbb{R}$. Se fosse $a_k = -\infty$, para todo $k \in \mathbb{N}$, então $L = -\infty$. Logo, existe $k_0 \in \mathbb{N}$ tal que $a_{k_0} > -\infty$. Como (a_k) é não decrescente, temos

$$-\infty < a_{k_0} \leq a_k \leq \sup a_n = L < +\infty, \quad \text{para todo} \quad k \geq k_0.$$

Como $L = \lim a_k$, escolha $k_1 \geq k_0$ tal que $L - 1 < a_{k_1} \leq L$. Como $a_{k_1} = \inf_{n \geq k_1} x_n$, escolha $n_1 \geq k_1$ tal que $a_{k_1} \leq x_{n_1} < a_{k_1} + 1$. Então: $L - 1 < x_{n_1} < L + 1$.

Agora escolha $k_2 \geq \max\{k_0, n_1 + 1\}$ tal que $L - 1/2 < a_{k_2} \leq L$. Como $a_{k_2} = \inf_{n \geq k_2} x_n$, escolha $n_2 \geq k_2$ tal que $a_{k_2} \leq x_{n_2} < a_{k_2} + 1/2$. Isso nos dá $L - 1/2 < x_{n_2} < L + 1/2$, com $n_2 \geq k_2 > n_1$.

Continuando dessa forma, obtemos uma subsequência (x_{n_k}) tal que

$$L - 1/k < x_{n_k} < L + 1/k, \quad \text{para todo} \quad k \in \mathbb{N}\,.$$

Pelo Teorema do Confronto (ver Teorema A.1.10), temos $\lim x_{n_k} = L$.

Mostraremos agora que (x_n) não possui nenhum valor de aderência menor do que $\liminf x_n$. Suponhamos, por redução ao absurdo, que há

$$c < \liminf x_n$$

que ainda é um valor de aderência de (x_n). Assim, existe uma subsequência $(x_{n_k})_k$ de (x_n) tal que

$$\lim_{k \to \infty} x_{n_k} = c.$$

Como

$$c < \liminf x_n := \sup \left\{ \inf_{n \geq k} x_n : k \in \mathbb{N} \right\},$$

então existe $k_0 \in \mathbb{N}$ tal que

$$c < \inf_{n \geq k_0} x_n \leq \liminf x_n.$$

Tomando

$$\varepsilon := \inf_{n \geq k_0} x_n - c,$$

existe $k_1 \in \mathbb{N}$ tal que

$$k \geq k_1 \implies |x_{n_k} - c| < \varepsilon \implies x_{n_k} < c + \varepsilon = \inf_{n \geq k_0} x_n.$$

Em particular, tomando $p \geq \max\{k_0 + 1, k_1\}$, temos $n_p \geq n_{k_0+1} \geq k_0 + 1$ e

$$x_{n_p} < \inf_{n \geq k_0} x_n \leq x_{n_p},$$

o que é absurdo.

A demonstração com respeito ao limite superior pode ser realizada de modo completamente análogo. $\square$

Corolário A.3.4. Sejam (x_n) uma sequência e $L \in \mathbb{R} \cup \{-\infty, +\infty\}$. Então $\lim x_n = L$ se, e somente se, $\liminf x_n = \limsup x_n = L$.

Demonstração. Se $\liminf x_n < \limsup x_n$, então (x_n) possui subsequências convergindo para pontos distintos, logo não podemos ter $\lim x_n = L$.

Reciprocamente, suponhamos $\liminf x_n = \limsup x_n = L$. Se $L = +\infty$, então $\lim x_n = +\infty$, como provamos na prova anterior; e se $L = -\infty$, temos $\lim x_n = -\infty$, pelo caso análogo.

Por fim, suponhamos $L \in \mathbb{R}$ e seja $\varepsilon > 0$. Como

$$\liminf x_n - \varepsilon < \liminf x_n \quad \text{e} \quad \limsup x_n < \limsup x_n + \varepsilon,$$

então existem $k_1, k_2 \in \mathbb{N}$ tais que

$$\liminf x_n - \varepsilon < \inf_{n \geq k_1} x_n \leq \liminf x_n = \limsup x_n \leq \sup_{n \geq k_2} x_n < \limsup x_n + \varepsilon = \liminf x_n + \varepsilon.$$

Desse modo,

$$m \geq \max\{k_1, k_2\} \implies \liminf x_n - \varepsilon < \inf_{n \geq k_1} x_n \leq x_m \leq \sup_{n \geq k_2} x_n < \liminf x_n + \varepsilon.$$

Logo, $\lim\limits_{m \to \infty} x_m = \liminf x_n$. □

Corolário A.3.5 (Teorema de Bolzano-Weierstrass)**.** Toda sequência tem pelo menos um valor de aderência em $[-\infty, +\infty]$. Toda sequência limitada admite subsequência convergente.

Demonstração. Toda sequência (x_n) possui uma subsequência (x_{n_k}) tal que $\lim_{k \to \infty} x_{n_k} = \liminf x_n$. Em particular, se (x_n) é limitada, o limite inferior é o limite de uma sequência monótona e limitada, logo $\liminf x_n$ é finito. □

Teorema A.3.6. Se $(x_n) \subseteq \mathbb{R}$ é uma sequência limitada, então, dado $\varepsilon > 0$, podemos obter $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies \liminf x_n - \varepsilon < x_n < \limsup x_n + \varepsilon.$$

Além disso, $\liminf x_n$ e $\limsup x_n$ são, respectivamente, o maior número e o menor número com essa propriedade.

Demonstração. Sejam $a := \liminf x_n$ e $b := \limsup x_n$. Suponhamos, por redução ao absurdo, que existe $\varepsilon_0 > 0$ tal que, para cada $n \in \mathbb{N}$, podemos obter $n_0 \geq n$ tal que $x_{n_0} \leq a - \varepsilon_0$ ou $b + \varepsilon_0 \leq x_{n_0}$. Assim, $x_n \leq a - \varepsilon_0$, para uma infinidade de índices n, ou $b + \varepsilon_0 \leq x_n$, para uma infinidade de índices n.

Se $x_n \leq a - \varepsilon_0$, para uma infinidade de índices n, teríamos uma subsequência (x_{n_k}) de (x_n) tal que $\inf_{n\in\mathbb{N}} x_n \leq x_{n_k} \leq a - \varepsilon_0$. Logo, pelo Teorema de Bolzano-Weierstrass (ver Corolário A.3.5), existe $c \in \mathbb{R}$ tal que alguma subsequência de (x_{n_k}) converge para c. Mas, pela Monotonicidade do Limite (ver Teorema A.1.11), temos $c \leq a - \varepsilon_0 < a$, o que é absurdo. O outro caso pode ser demonstrado de modo completamente análogo a esse.

Suponhamos, por redução ao absurdo, que há $a' > a$ satisfazendo a condição. Como $a = \liminf x_n$, então (x_n) possui uma subsequência (x_{n_k}) convergindo para a. Tomando $\varepsilon := \dfrac{a'-a}{2}$, podemos obter $k_0 \in \mathbb{N}$ tal que

$$k \geq k_0 \implies |x_{n_k} - a| < \varepsilon \implies x_{n_k} < a + \varepsilon = \frac{a'+a}{2}.$$

Como $\varepsilon > 0$, então existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies \frac{a'+a}{2} = a' - \varepsilon < x_n.$$

Tomando $N := \max\{n_{k_0}, n_0\}$, temos

$$n_k \geq N \implies x_{n_k} < \frac{a'+a}{2} < x_{n_k},$$

o que é absurdo.

O outro caso pode ser demonstrado de modo completamente análogo a esse. □

Proposição A.3.7 (Propriedades lim inf e lim sup)**.** Sejam (x_n) uma sequência de números reais e $a, b \in \mathbb{R}$. Então,

- $a \leq \liminf x_n$ se, e somente se, para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

  $$n \geq n_0 \implies x_n \geq a - \varepsilon.$$

  Consequentemente,

  $$\liminf x_n < M \iff \text{existe } \varepsilon > 0 \text{ tal que } x_n < M - \varepsilon \text{ para infinitos } n\text{'s}.$$

  Além disso, se $a < \liminf x_n$, então existem $n_0 \in \mathbb{N}$ e $c < \liminf x_n$ tais que

  $$n \geq n_0 \implies x_n \geq c > a.$$

- $\limsup x_n \leq b$ se, e somente se, para todo $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

  $$n \geq n_0 \implies x_n \leq b + \varepsilon.$$

  Consequentemente,

  $$\limsup x_n > M \iff \text{existe } \varepsilon > 0 \text{ tal que } x_n > M + \varepsilon \text{ para infinitos } n\text{'s}.$$

  Além disso, se $\limsup x_n < b$, então existem $n_0 \in \mathbb{N}$ e $d > \limsup x_n$ tais que

  $$n \geq n_0 \implies x_n \leq d < b.$$

Demonstração. Vamos provar apenas o primeiro item, já que o resultado para o limite superior se segue do primeiro item a partir da observação de que

$$\sup A = -\inf(-A),$$

para todo subconjunto $A \subseteq \mathbb{R}$ não vazio e limitado superiormente. Começaremos provando a validade da equivalência.

Suponhamos inicialmente que $a \leq \liminf x_n$. Assim, fixado $\varepsilon > 0$, como

$$a - \varepsilon < a \leq \liminf x_n,$$

existe no máximo uma finidade de n's tal que

$$x_n < a - \varepsilon.$$

Com efeito, se houvesse (x_{n_k}) tal que $x_{n_k} < a - \varepsilon$, para todo $k \in \mathbb{N}$, tome c um valor de aderência de (x_{n_k}). A monotonicidade do limite força $c \leq a - \varepsilon$. Mas isso nos daria $c < \liminf x_n$, que é o menor valor de aderência possível. Absurdo!

Logo, existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies x_n \geq a - \varepsilon,$$

como queríamos demonstrar.

Reciprocamente, dado $\varepsilon > 0$, a sequência (x_n) tem no máximo finitos termos $< a - \varepsilon$ e, portanto, o mesmo vale para todas as subsequências de (x_n); em particular, isso vale para todas as subsequências convergentes, implicando que elas têm limite $\geq a - \varepsilon$. O resultado se segue da arbitrariedade de ε.

Agora, se $a < \liminf x_n$, fixamos $\varepsilon > 0$ a ser escolhido posteriormente e tomamos qualquer c_0 no intervalo $(a, \liminf x_n)$. Daí, como $\liminf x_n \geq c_0$, então, pelo resultado que acabamos de provar, existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies x_n \geq c_0 - \varepsilon,$$

ficando

$$c_0 - \varepsilon > a \text{ para todo } \varepsilon \in (0, c_0 - a).$$

Então, escolhendo um ε qualquer nesse intervalo e colocando

$$c := c_0 - \varepsilon,$$

temos

$$c < \liminf x_n + 0 = \liminf x_n \text{ e } n \geq n_0 \implies x_n \geq c > a,$$

como queríamos demonstrar. $\square$

Teorema A.3.8 (Extensão da monotonicidade do limite)**.** Se (x_n) e (y_n) são sequências limitadas de números reais tais que

$$x_n \leq y_n, \quad \text{para todo } n \text{ suficientemente grande,}$$

então

$$\liminf x_n \leq \liminf y_n \quad \text{e} \quad \limsup x_n \leq \limsup y_n.$$

Em particular, se (x_n) e (y_n) convergem, então

$$\lim_{n\to\infty} x_n \leq \lim_{n\to\infty} y_n.$$

Demonstração. Vamos fazer a prova por contrapositiva. Se

$$\liminf x_n > \liminf y_n,$$

então, pela Proposição A.3.7, existem $n_0 \in \mathbb{N}$ e $c < \liminf x_n$ tais que

$$n \geq n_0 \implies x_n \geq c > \liminf y_k.$$

Como $\liminf y_k$ é um valor de aderência de (y_k), então existe uma subsequência (y_{k_p}) de (y_k) tal que

$$\lim_{p\to\infty} y_{k_p} = \liminf y_k.$$

De $c > \liminf y_k$, existe $p_1 \in \mathbb{N}$ tal que

$$p \geq p_1 \implies y_{k_p} < c.$$

Como (k_p) é uma sequência crescente de índices naturais, existe $p_2 \in \mathbb{N}$ tal que

$$p \geq p_2 \implies k_p \geq n_0.$$

Daí,

$$p \geq \max\{p_1, p_2\} \implies y_{k_p} < c \leq x_{k_p},$$

o que termina a prova. □

A.4 Séries numéricas

Definição A.4.1 (Séries finitas)**.** Sejam $m \in \mathbb{Z}$ e $(x_k)_{k=m}^{\infty} \subseteq \mathbb{R}$. A **série finita** (ou **soma finita**) $\sum_{i=m}^{n} x_i$ é definida recursivamente por

$$\sum_{i=m}^{n} x_i := 0, \quad \text{para todo} \quad n < m,$$
$$\sum_{i=m}^{n+1} x_i := \sum_{i=m}^{n} x_i + x_{n+1}, \quad \text{para todo} \quad n \geq m - 1.$$

Definição A.4.2 (Série infinita)**.** Dados $m \in \mathbb{Z}$ e $(x_n)_{n=m}^{\infty} \subseteq \mathbb{R}$, a **série** (de termo geral x_n) é a sequência

$$\sum_{n=m}^{\infty} x_n := \left(\sum_{k=m}^{1} x_k, \sum_{k=m}^{2} x_k, \ldots, \sum_{k=m}^{n} x_k, \ldots\right).$$

Dizemos que a série $\sum_{n=m}^{\infty} x_n$ **converge** para $L \in \mathbb{R}$, e escrevemos $\sum_{n=m}^{\infty} x_n = L$, se $\lim_{n\to\infty} \sum_{k=m}^{n} x_k = L$. Dizemos que uma série $\sum_{n=m}^{\infty} x_n$ é **convergente** se existe $L \in \mathbb{R}$ tal que $\sum_{n=m}^{\infty} x_n = L$.

Exemplo A.4.3 (Série geométrica)**.** Se $q \in (0, 1)$, então $\sum q^n$ é uma série convergente.

Demonstração. Como

$$\sum_{k=1}^{n} q^k = \frac{q}{1-q}(1-q^n) \xrightarrow{n\to\infty} \frac{q}{1-q},$$

então

$$\sum q^n = \frac{q}{1-q}.$$

Logo, $\sum q^n$ é convergente. $\square$

Teorema A.4.4. Se $c \in \mathbb{R}$ e $\sum x_n, \sum y_n$ são séries convergentes, então $\sum(x_n+y_n)$ e $\sum(cx_n)$ também são, valendo:

1. $\sum(x_n + y_n) = \sum x_n + \sum y_n$;
2. $\sum(cx_n) = c \sum x_n$.

Demonstração. Como

$$\sum_{k=1}^{n}(x_k + y_k) = \sum_{k=1}^{n} x_k + \sum_{k=1}^{n} y_k \xrightarrow{n\to\infty} \sum x_n + \sum y_n,$$

então $\sum(x_n + y_n) = \sum x_n + \sum y_n$. Além disso, $\sum_{k=1}^{n}(cx_k) = c\sum_{k=1}^{n} x_k \xrightarrow{n\to\infty} c\sum x_n$. $\square$

Teorema A.4.5. Se $\sum x_n$ é uma série de termos não negativos, então $\sum x_n$ converge se, e somente se, $\sum x_n$ é limitada.

Demonstração. Como $\sum x_n$ é não decrescente, então $\sum x_n$ é convergente se, e somente se, $\sum x_n$ é limitada. $\square$

Corolário A.4.6 (Teste da Comparação)**.** Se $\sum x_n$ e $\sum y_n$ são séries de termos não negativos tais que $x_n \leq y_n$, então $\sum y_n$ converge $\implies \sum x_n$ converge.

Definição A.4.7. Dizemos que uma série $\sum x_n$ é **absolutamente convergente** se $\sum |x_n|$ é convergente.

Teorema A.4.8. Toda série absolutamente convergente converge.

Demonstração. Seja $\sum x_n$ uma série absolutamente convergente. Como

$$0 \leq |x_n| - x_n \leq 2|x_n|,$$

para todo $n \in \mathbb{N}$, então, pelo Teste da Comparação (ver Corolário A.4.6), $\sum(|x_n| - x_n)$ é convergente, já que $\sum(2|x_n|)$ é convergente, logo $\sum x_n = \sum |x_n| - \sum(|x_n| - x_n)$. □

Teorema A.4.9 (Lema de Fatou para séries)**.** Seja $(a_{n,k})_{n,k\in\mathbb{N}}$ uma sequência de números reais não negativos. Então

$$\sum_{n=1}^{\infty} \liminf_{k\to\infty} a_{n,k} \leq \liminf_{k\to\infty} \sum_{n=1}^{\infty} a_{n,k}.$$

Demonstração. Como

$$\sum_{n=1}^{N} \liminf_{k\to\infty} a_{n,k} \leq \liminf_{k\to\infty} \sum_{n=1}^{N} a_{n,k} \leq \liminf_{k\to\infty} \sum_{n=1}^{\infty} a_{n,k},$$

para todo $N \in \mathbb{N}$, então

$$\sum_{n=1}^{\infty} \liminf_{k\to\infty} a_{n,k} \leq \liminf_{k\to\infty} \sum_{n=1}^{\infty} a_{n,k},$$

como queríamos demonstrar. □

APÊNDICE B – Topologia Geral

Este capítulo é baseado livremente em (MENDELSON, 1990), (LIMA, 2014) e (LIMA, 2017).

B.1 Espaços topológicos

Uma **topologia** num conjunto X é uma coleção τ de subconjuntos de X, que chamaremos **abertos**, tais que:

1. $X, \varnothing \in \tau$.

2. Se $A, B \in \tau$, então $A \cap B \in \tau$.

3. Se $A_\lambda \in \tau$, para todo $\lambda \in \Lambda$, então $\bigcup\limits_{\lambda \in \Lambda} A_\lambda \in \tau$.

Um **espaço topológico** é um par (X, τ), onde X é um conjunto e τ é uma topologia em X. Como é tradicional na literatura, faremos referência ao espaço topológico X, considerando a topologia subentendida.

Exemplo B.1.1. A topologia usual da reta $\mathbb{R}$ é a coleção τ de todos os subconjuntos $X \subseteq \mathbb{R}$ tais que, para cada $x \in X$, existe um intervalo aberto (a, b) com $x \in (a, b) \subseteq X$.

Proposição B.1.2. Sejam (A, τ) um espaço topológico e $B \subseteq A$. Então, a *topologia induzida*

$$\tau' := \{X \cap B : X \in \tau\}$$

é uma topologia em B.

Demonstração. Como $A, \varnothing \in \tau$, então

$$\varnothing = \varnothing \cap B \in \tau' \text{ e } B = B \cap A \in \tau'.$$

Seja $\{X_\lambda\}_{\lambda \in \Lambda} \subseteq \tau'$. Para cada $\lambda \in \Lambda$, existe $Y_\lambda \in \tau$ tal que $X_\lambda = Y_\lambda \cap B$. Assim,

$$\bigcup_{\lambda \in \Lambda} Y_\lambda \in \tau,$$

donde

$$\bigcup_{\lambda \in \Lambda} X_\lambda = \bigcup_{\lambda \in \Lambda} (Y_\lambda \cap B) = \bigcup_{\lambda \in \Lambda} Y_\lambda \cap B \in \tau'.$$

Por fim, dados $X_1, X_2 \in \tau'$, existem $Y_1, Y_2 \in \tau$ tais que $X_i = B \cap Y_i$, para $i = 1, 2$. Nessas condições,

$$Y_1 \cap Y_2 \in \tau,$$

donde

$$X_1 \cap X_2 = (B \cap Y_1) \cap (B \cap Y_2) = B \cap (Y_1 \cap Y_2) \in \tau'.$$

Logo, τ' é uma topologia em B. $\square$

Definição B.1.3. Sejam A um espaço topológico e $X \subseteq A$.

- O **interior** de X é o conjunto

$$\operatorname{int} X := \bigcup\{Y \subseteq X : Y \text{ é aberto}\}.$$

- X é **fechado** se $A \setminus X$ é aberto.
- O **fecho** de X é o conjunto

$$\overline{X} := \bigcap\{F \subseteq A : X \subseteq F \text{ e } F \text{ é fechado}\}.$$

- A **fronteira** de X é o conjunto

$$\partial X := (X \setminus \operatorname{int} X) \cup (\overline{X} \setminus X).$$

- Um conjunto $V \subseteq A$ é uma **vizinhança** de um ponto $x \in A$ se $x \in \operatorname{int} V$.
- Um ponto $a \in A$ é um **ponto de acumulação** de X se

$$V \text{ é uma vizinhança de } a \implies (V \setminus \{a\}) \cap X \neq \varnothing.$$

- O **derivado** de X é o conjunto X' de todos os pontos de acumulação de X.

B.2 Continuidade

Uma função $f : A \to B$, onde A e B são espaços topológicos, é dita **contínua** se $f^{-1}(Y)$ é aberto de A, para cada Y aberto de B.

Proposição B.2.1. Se A, B e C são espaços topológicos e $f : A \to B$ e $g : B \to C$ são funções contínuas, então $g \circ f : A \to C$ é contínua.

Demonstração. Se Y é aberto de C, então

$$(g \circ f)^{-1}(Y) = f^{-1}(g^{-1}(Y))$$

é aberto, já que $g^{-1}(Y)$ é aberto de B. $\square$

Definição B.2.2 (Continuidade pontual)**.** Dizemos que uma função $f : A \to B$, onde A e B são espaços topológicos, é **contínua** em $a \in A$ se, para toda vizinhança Y de $f(a)$, existe uma vizinhança X de a tal que $f(X) \subseteq Y$.

Teorema B.2.3. Se A e B são espaços topológicos, então uma função $f : A \to B$ é contínua se, e somente se, é contínua em cada $a \in A$.

Demonstração. Suponhamos f uma função contínua e sejam $a \in A$ e Y uma vizinhança aberta de $f(a)$. Como $f^{-1}(Y)$ é uma vizinhança aberta de a e

$$f(f^{-1}(Y)) \subseteq Y,$$

segue-se o resultado.

Reciprocamente, suponhamos f contínua em cada $a \in A$ e considere Y um aberto de B. Dado $x \in f^{-1}(Y)$, como $x \in A$, então f é contínua em x, donde x possui uma vizinhança aberta A_x, com $f(A_x) \subseteq Y$. De

$$f^{-1}(Y) = \bigcup_{f(x) \in Y} A_x,$$

vem que $f^{-1}(Y)$ é aberto. □

Definição B.2.4. Se A e B são espaços topológicos, uma bijeção $f : A \to B$ é um **homeomorfismo** se f e f^{-1} são contínuas. Dizemos que A e B são **homeomorfos** se há um homeomorfismo entre eles.

B.3 Conexidade

Um espaço topológico X é **conexo** se a única forma de escrever

$$X = A \cup B,$$

onde A e B são abertos de X disjuntos, é quando $A = \varnothing$ ou $B = \varnothing$. Um espaço topológico **desconexo** é um espaço que não é conexo.

Um subconjunto D de um espaço topológico X é **conexo** (resp. **desconexo**) se D é um espaço topológico conexo (resp. desconexo), com a topologia induzida de X.

Exemplo B.3.1. Seja X um conjunto qualquer e considere sobre X a *topologia discreta* $\tau := \mathcal{P}(X)$. Então todos os subconjuntos de X com pelo menos dois elementos são desconexos.

Proposição B.3.2. Um subconjunto D de um espaço topológico X é desconexo se, e somente se, existem subconjuntos abertos $A, B \subseteq X$ tais que

$$D \subseteq A \cup B, \quad D \cap A \neq \varnothing \neq D \cap B \quad \text{e} \quad D \cap A \cap B = \varnothing.$$

Demonstração. Seja $D \subseteq X$ desconexo. Então podemos escrever $D = (A \cap D) \cup (B \cap D)$, com A e B abertos de X, e $A \cap D$ e $B \cap D$ disjuntos e não vazios.

Reciprocamente, suponhamos que valha a outra condição do enunciado. Então $D \cap A$ e $D \cap B$ são abertos de D disjuntos e não vazios. □

Proposição B.3.3. Intervalos abertos e limitados de números reais são conexos.

Demonstração. Suponhamos, por redução ao absurdo, que existem $a < b$ em $\mathbb{R}$ tais que o intervalo (a, b) é desconexo. Logo, existem abertos $A, B \subseteq \mathbb{R}$ tais que

$$(a, b) \subseteq A \cup B, A \cap (a, b) \neq \varnothing \neq B \cap (a, b) \text{ e } A \cap B \cap (a, b) = \varnothing.$$

Sejam $x \in A \cap (a, b)$ e $y \in B \cap (a, b)$. Sem perda de generalidade, suponhamos $x < y$ e consideremos o conjunto

$$X := \{z \in A \cap (a, b) : [x, z] \subseteq [x, y]\}.$$

Como X é não vazio (pois $x \in X$) e limitado superiormente (já que y é uma cota superior de X), então $c := \sup X \in \mathbb{R}$. Note que $A \cap (a, b)$ e $B \cap (a, b)$ são abertos. Assim, se $c \in X$, então $c \in A \cap (a, b)$, donde

$$(c - \delta, c + \delta) \subseteq A \cap (a, b), \text{ para algum } \delta > 0,$$

e, portanto,

$$d := c + \frac{\delta}{2} \in A \cap (a, b).$$

Como $\sup X = c < d$, então $d \notin X$, donde $[x, d] \nsubseteq [x, y]$, o que nos dá $y < d$ e, portanto,

$$y \in [c, d] \subseteq (c - \delta, c + \delta) \subseteq A \cap (a, b),$$

já que $c \leq y$. Por conseguinte, $A \cap B \cap (a, b) \neq \varnothing$, o que é absurdo! Logo, $c \notin X$. Como

$$[x, c] \subseteq [x, y] \subseteq (a, b)$$

e $c \notin X$, então $c \notin A \cap (a, b)$, donde $c \notin A$, já que $c \in (a, b)$. De $(a, b) \subseteq A \cup B$, vem $c \in B \cap (a, b)$, o que nos dá

$$(c - \delta, c + \delta) \subseteq B \cap (a, b), \text{ para algum } \delta > 0.$$

Como $c - \delta < c = \sup X$, então existe $h \in X$ tal que $c - \delta < h \leq c < c + \delta$, donde

$$h \in X \cap B \cap (a, b) \subseteq A \cap B \cap (a, b),$$

o que é absurdo.

Consequentemente, todo intervalo aberto limitado é conexo. □

Proposição B.3.4. Em um espaço topológico, a união de uma família de subconjuntos conexos com pelo menos um ponto em comum é um subconjunto conexo.

Demonstração. Sejam X um espaço topológico e $\{C_\lambda\}_{\lambda\in\Lambda}$ uma família de subconjuntos conexos de X com pelo menos um ponto em comum, isto é, com

$$D := \bigcap_{\lambda\in\Lambda} C_\lambda \neq \varnothing.$$

Suponhamos, por redução ao absurdo, que

$$C := \bigcup_{\lambda\in\Lambda} C_\lambda$$

é desconexo. Então, existem subconjuntos abertos A e B de X tais que

$$C \subseteq A \cup B, \quad C \cap A \neq \varnothing \neq C \cap B \quad \text{e} \quad C \cap A \cap B = \varnothing.$$

Tome algum $C_1 \in \{C_\lambda\}$. Como $C_1 \subseteq C \subseteq A \cup B$ e $C_1 \cap A \cap B = \varnothing$, então

$$C_1 \cap A = \varnothing \neq C_1 \cap B \quad \text{ou} \quad C_1 \cap B = \varnothing \neq C_1 \cap A,$$

já que C_1 é conexo. Sem perda de generalidade, suponhamos que seja o primeiro caso. Então $D \subseteq C_1 \subseteq B$. Como $C \cap A \neq \varnothing$, então existe $\lambda \in \Lambda$ tal que $C_\lambda \cap A \neq \varnothing$. E como

$$C_\lambda \cap B \supseteq D \cap B = D \neq \varnothing,$$

temos

$$C_\lambda \subseteq A \cup B, \quad C_\lambda \cap A \neq \varnothing \neq C_\lambda \cap B \quad \text{e} \quad C_\lambda \cap A \cap B = \varnothing,$$

isto é, C_λ não é conexo. Absurdo! □

Corolário B.3.5. Intervalos abertos de números reais são conexos.

Demonstração. Todo intervalo aberto pode ser exprimido como uma união de intervalos abertos limitados com um ponto em comum. □

Note que um subconjunto $X \subseteq \mathbb{R}$ é um intervalo se, e somente se, para quaisquer $u, v \in X$, com $u < v$, tem-se $[u, v] \subseteq X$.

Proposição B.3.6. Sejam X um espaço topológico e $A \subseteq X$. Se

$$A \subseteq B \subseteq \overline{A},$$

então B é conexo. Em particular, o fecho de um conjunto conexo é conexo.

Demonstração. Suponhamos, por redução ao absurdo, que existe $B \subseteq X$ desconexo tal que $A \subseteq B \subseteq \overline{A}$. Então existem subconjuntos abertos $P, Q \subseteq X$ tais que

$$B \subseteq P \cup Q, \quad B \cap P \neq \varnothing \neq B \cap Q \quad \text{e} \quad B \cap P \cap Q = \varnothing.$$

Pelas hipóteses sobre B, temos

$$A \subseteq P \cup Q, \quad \overline{A} \cap P \neq \varnothing \neq \overline{A} \cap Q \quad \text{e} \quad A \cap P \cap Q = \varnothing.$$

Como P e Q são abertos, temos

$$\overline{A} \cap Y \neq \varnothing \iff A \cap Y \neq \varnothing, \quad \text{para todo} \quad Y \in \{P, Q\}.$$

Logo, A é desconexo. Absurdo! □

Corolário B.3.7. Intervalos de números reais são conexos.

Demonstração. É óbvio que intervalos degenerados são conexos. Se $a < b$, já sabemos que (a, b) é conexo, o que nos dá $(a, b], [a, b)$ e $[a, b]$ são conexos, por conterem (a, b) e estarem contidos em $[a, b] = \overline{(a, b)}$. Por um motivo similar, dado $a \in \mathbb{R}$, os intervalos $[a, +\infty)$ e $(-\infty, a]$ também são conexos. □

Corolário B.3.8. Um subconjunto da reta é conexo se, e somente se, é um intervalo.

Demonstração. Já provamos a volta, agora provaremos a ida por contrapositiva. Se $X \subseteq \mathbb{R}$ não é um intervalo, então existem $u, v \in X$ e $w \in \mathbb{R}$ tais que $u < w < v$, mas $w \notin X$, donde

$$X \subseteq (-\infty, w) \cup (w, +\infty),$$

com

$$X \cap (-\infty, w) \neq \varnothing \neq X \cap (w, +\infty),$$

e, portanto, X é desconexo. □

Definição B.3.9. Se X é um espaço topológico, dizemos que $C \subseteq X$ é um **conjunto conexo maximal** de X (ou uma **componente conexa** de X) se C é um conjunto conexo tal que

$$C \subseteq D \implies C = D, \quad \text{para todo} \quad D \subseteq X \quad \text{conexo}.$$

Proposição B.3.10. Sejam X um espaço topológico e C uma componente conexa de X. Então, para qualquer subconjunto conexo U de X,

$$U \cap C \neq \varnothing \iff U \neq \varnothing \quad \text{e} \quad U \subseteq C.$$

Demonstração. A recíproca é óbvia. Suponhamos $U \cap C \neq \varnothing$. Então, $U \cup C$ é um conjunto conexo (veja Proposição B.3.4), donde

$$C \subseteq U \cup C \implies C = U \cup C \implies U \subseteq C,$$

como queríamos demonstrar. □

Proposição B.3.11. Dado um espaço topológico X, são equivalentes:

1. Todo ponto de X possui uma vizinhança conexa.
2. As componentes conexas de X são abertas.
3. X é uma união de abertos conexos.

Demonstração. Vejamos primeiro que 1. $\implies$ 2.

Seja C uma componente conexa de X. Para cada $x \in C$, temos $x \in X$, donde existem um subconjunto aberto U_x e um subconjunto conexo V_x de X com

$$x \in U_x \subseteq V_x.$$

Daí, pela Proposição B.3.10, temos

$$x \in C \implies x \in U_x \subseteq V_x \implies V_x \cap C \neq \varnothing \implies V_x \subseteq C,$$

donde

$$C = \bigcup_{x \in C} V_x = \bigcup_{x \in C} U_x$$

e, portanto, C é aberto.

A implicação 2. $\implies$ 3. se segue do fato de que as componentes conexas de X formam uma partição de X.

A implicação 3. $\implies$ 1. é imediata. □

Definição B.3.12. Um espaço topológico é dito **localmente conexo** se satisfaz qualquer um dos itens da Proposição B.3.11.

Um subconjunto de um espaço topológico é dito **localmente conexo** se é um espaço localmente conexo com a topologia induzida.

Exemplo B.3.13. Todos os abertos da reta são localmente conexos.

Com efeito, dado um aberto $U \subseteq \mathbb{R}$, se $x \in U$, então existe um intervalo aberto (a, b) tal que

$$x \in (a, b) \subseteq U.$$

Logo, (a, b) é uma vizinhança conexa de x na topologia de U.

Proposição B.3.14. A coleção de componentes conexas de um aberto da reta é enumerável e contém somente intervalos abertos.

Demonstração. Seja $A \subseteq \mathbb{R}$ um aberto. As componentes conexas de A são subconjuntos conexos de $\mathbb{R}$, logo todas as componentes de A são intervalos, sendo todos esses intervalos abertos em vista do Exemplo B.3.13.

Sobre a enumerabilidade, considere $\mathcal{C}$ a coleção de todas as componentes conexas de A. Fixe em cada $C \in \mathcal{C}$ um número racional $r_C \in C$. Como a função

$$C \in \mathcal{C} \mapsto r_C \in \mathbb{Q}$$

é injetiva, o resultado se segue da enumerabilidade de $\mathbb{Q}$. □

Teorema B.3.15. Conexidade é *invariante por funções contínuas*, isto é, dados espaços topológicos A e B e uma função contínua $f : A \to B$, se A é conexo, então $f(A)$ é um subconjunto conexo de B.

Demonstração. Vamos fazer a prova por contrapositiva, então começamos supondo $f(A)$ desconexo. Assim, existem abertos Y e Z de B tais que

$$f(A) \subseteq Y \cup Z, \quad f(A) \cap Y \neq \varnothing \neq f(A) \cap Z \quad \text{e} \quad f(A) \cap Y \cap Z = \varnothing.$$

Como

$$A \subseteq f^{-1}(f(A)) \subseteq f^{-1}(Y \cup Z) = f^{-1}(Y) \cup f^{-1}(Z),$$

com $f^{-1}(Y)$ e $f^{-1}(Z)$ abertos (pela continuidade de f),

$$A \cap f^{-1}(Y) \cap f^{-1}(Z) \subseteq f^{-1}(f(A)) \cap f^{-1}(Y) \cap f^{-1}(Z) = f^{-1}(f(A) \cap Y \cap Z) = \varnothing$$

e

$$A \cap f^{-1}(Y) \neq \varnothing \neq A \cap f^{-1}(Z),$$

então A é desconexo. □

Corolário B.3.16 (Teorema do Valor Intermediário)**.** Se $f : [a, b] \to \mathbb{R}$ é contínua e $f(a) < d < f(b)$, então existe $c \in (a, b)$ tal que $f(c) = d$.

Demonstração. Como f é contínua, então $f([a, b])$ é conexo (e, portanto, um intervalo). Desse modo, $d \in f([a, b])$, donde existe $c \in [a, b]$ tal que $f(c) = d$. Como $f(a) < d < f(b)$, então $a < c < b$, isto é, $c \in (a, b)$. □

Corolário B.3.17. Se X e Y são espaços topológicos, $f : X \to Y$ é uma função sobrejetiva contínua e X é conexo, então Y também é conexo.

Demonstração. Sendo f sobrejetiva, $Y = f(X)$ é conexo. □

Corolário B.3.18. Conexidade é um *invariante topológico*, isto é, se X e Y são espaços topológicos homeomorfos, então X é conexo se, e somente se, Y é conexo.

Teorema B.3.19 (Teorema da Alfândega)**.** Sejam M um espaço topológico e $X \subseteq M$. Se $C \subseteq M$ é um conjunto conexo tal que $C \cap X \neq \varnothing \neq C \cap (M \setminus X)$, então $C \cap \partial X \neq \varnothing$.

Demonstração. Se $C \cap \operatorname{int} X = \varnothing$, então todo ponto de $C \cap X$ está na fronteira de X. Analogamente, se $C \cap \operatorname{int}(M \setminus X) = \varnothing$, então todo ponto de $C \cap (M \setminus X)$ está na fronteira de $M \setminus X$, e, portanto, está na fronteira de X.

Suponhamos, portanto, que

$$C \cap \operatorname{int} X \neq \varnothing \neq C \cap \operatorname{int}(M \setminus X).$$

Como C é conexo e

$$C \cap \operatorname{int} X \cap \operatorname{int}(M \setminus X) = \varnothing,$$

então

$$C \not\subseteq \operatorname{int} X \cup \operatorname{int}(M \setminus X).$$

Logo, existe $c \in C$ tal que

$$c \notin \operatorname{int} X \cup \operatorname{int}(M \setminus X),$$

donde

$$c \notin \operatorname{int} X \text{ e } c \notin \operatorname{int}(M \setminus X).$$

Nessas condições, se A é um aberto de M tal que $c \in A$, então, de $c \notin \operatorname{int} X$, temos

$$A \cap (M \setminus X) \neq \varnothing$$

e, do fato de que $c \notin \operatorname{int}(M \setminus X)$, temos

$$A \cap X \neq \varnothing.$$

Consequentemente, $c \in C \cap \partial X$. □

Definição B.3.20. Um espaço topológico X é dito **conexo por caminhos** se, para quaisquer $x, y \in X$, existe um *caminho em X ligando x a y*, isto é, uma função $\varphi : [0,1] \to X$ contínua tal que $\varphi(0) = x$ e $\varphi(1) = y$.

Corolário B.3.21. Sejam M um espaço topológico e $U \subseteq M$ um conjunto conexo por caminhos. Então, U é conexo.

Demonstração. Suponhamos, por redução ao absurdo, que U não é conexo. Então, existem abertos $A, B \subseteq M$ tais que

$$U \cap A \cap B = \varnothing, \quad \tilde{A} := U \cap A \neq \varnothing \neq \tilde{B} := U \cap B \quad \text{e} \quad U \subseteq A \cup B.$$

Como $\tilde{A} \neq \varnothing \neq \tilde{B}$, então existem $a \in \tilde{A}$ e $b \in \tilde{B}$. Por hipótese, existe um caminho $\varphi : [0,1] \to U$ tal que $\varphi(0) = a$ e $\varphi(1) = b$. Como φ é uma aplicação contínua e $[0,1]$ é conexo, então $\operatorname{im}(\varphi) = \varphi([0,1])$ é um conjunto conexo, com

$$a \in \operatorname{im}(\varphi) \cap \tilde{A}$$

e

$$b \in \operatorname{im}(\varphi) \cap \tilde{B} = \operatorname{im}(\varphi) \cap (U \setminus \tilde{A})$$

donde, pelo Teorema da Alfândega (ver Teorema B.3.19), existe $c \in \partial\tilde{A}$ tal que

$$c \in \operatorname{im}(\varphi) \subseteq U \subseteq A \cup B.$$

Do fato de que $\tilde{A}$ é aberto, sabemos que $c \notin \tilde{A}$, donde $c \in \tilde{B}$, que é um conjunto aberto. Como $c \in \partial\tilde{A}$, então

$$U \cap A \cap B = \tilde{A} \cap \tilde{B} \neq \varnothing.$$

Absurdo! $\square$

B.4 Compacidade

Uma **cobertura** de um conjunto X é uma coleção $\mathcal{A}$ de subconjuntos de X tal que $X \subseteq \bigcup \mathcal{A}$. Uma **subcobertura** de $\mathcal{A}$ para X é um subconjunto $\mathcal{B} \subseteq \mathcal{A}$ que ainda é uma cobertura de X. Quando X é um espaço topológico, uma cobertura de X formada apenas por abertos de X é chamada uma **cobertura aberta**.

Um espaço topológico é **compacto** se toda cobertura aberta dele admite uma subcobertura finita.

Um subconjunto D de um espaço topológico X é **compacto** se D é um espaço topológico compacto, com a topologia induzida de X.

Exemplo B.4.1. Todo subconjunto finito de um espaço topológico é compacto.

Proposição B.4.2. O intervalo de números reais da forma $[a,b]$ é compacto.

Demonstração. Suponhamos, por redução ao absurdo, que existe uma cobertura $\{A_\lambda\}_{\lambda\in\Lambda}$ de abertos de $[a,b]$ que não admite subcobertura finita.

Note que $|I_1| = \dfrac{b-a}{2}$, onde

$$I_1 := [a_1, b_1] := \begin{cases} \left[\frac{a+b}{2}, b\right], & \text{se } \left[\frac{a+b}{2}, b\right] \text{ não admite subcobertura finita,} \\ \left[a, \frac{a+b}{2}\right], & \text{caso contrário.} \end{cases}$$

De modo geral, note que $|I_{k+1}| = \dfrac{b-a}{2^{k+1}}$, onde $I_{k+1} := [a_{k+1}, b_{k+1}]$ e

$$[a_{k+1}, b_{k+1}] := \begin{cases} \left[\frac{a_k+b_k}{2}, b_k\right], & \text{se} \quad \left[\frac{a_k+b_k}{2}, b_k\right] \quad \text{não admite subcobertura finita,} \\ \left[a_k, \frac{a_k+b_k}{2}\right], & \text{caso contrário.} \end{cases}$$

Como $I_k \supseteq I_{k+1}$, para cada $k \in \mathbb{N}$, então existe $c \in \bigcap_{k=1}^{\infty} I_k$, pelo Teorema dos Intervalos Encaixados A.2.7. Além disso,

$$c \in [a,b] \subseteq \bigcup_{\lambda \in \Lambda} A_\lambda,$$

donde

$$c \in A_\lambda, \text{ para algum } \lambda \in \Lambda,$$

e, portanto, existe $\delta > 0$ tal que $(c-\delta, c+\delta) \subseteq A_\lambda$, já que A_λ é aberto. Como

$$\lim_{k\to\infty} |I_k| = \lim_{k\to\infty} \frac{b-a}{2^k} = 0,$$

então existe $k_0 \in \mathbb{N}$ tal que $I_{k_0} \subseteq (c-\delta, c+\delta) \subseteq A_\lambda$ e, portanto, I_{k_0} admite $\{A_\lambda\}$ como subcobertura aberta finita. Absurdo! □

Definição B.4.3. Um espaço topológico é **de Hausdorff** se quaisquer dois pontos distintos têm vizinhanças disjuntas.

Proposição B.4.4. A reta é um espaço de Hausdorff.

Demonstração. Sejam $a < b$ em $\mathbb{R}$ e tome $\varepsilon := \dfrac{b-a}{2}$. Então,

$$(a-\varepsilon, a+\varepsilon) \cap (b-\varepsilon, b+\varepsilon) = \varnothing.$$

Logo, a e b têm vizinhanças disjuntas. □

Proposição B.4.5. Seja X um conjunto com pelo menos dois elementos e considere sobre X a *topologia trivial* $\tau := \{\varnothing, X\}$. Então, X não é um espaço de Hausdorff. Note ainda que um subconjunto compacto de X não é necessariamente fechado.

Demonstração. Sejam x e y pontos distintos em X. Como X é o único aberto não vazio de X, então toda vizinhança de x é a mesma vizinhança de y (que é o conjunto X).

Além disso, dado $x \in X$, como $\{x\}$ é finito, então $\{x\}$ é compacto, mas $Y := X \setminus \{x\}$ não é aberto, já que $\varnothing \neq Y \neq X$. Logo, $\{x\}$ não é fechado. □

Proposição B.4.6. Todo subconjunto compacto de um espaço de Hausdorff é fechado.

Demonstração. Sejam X um espaço de Hausdorff, $K \subseteq X$ um conjunto compacto e $x \in X \setminus K$. Como X é Hausdorff, dado $y \in X$, existem subconjuntos abertos disjuntos A_y e B_y de X tais que $x \in A_y$ e $y \in B_y$. Como K é compacto e

$$K \subseteq \bigcup_{y \in K} B_y,$$

então existe $F \subseteq K$ finito tal que $\{B_y\}_{y \in F}$ ainda é uma cobertura de K. Como

$$A := \bigcap_{y \in F} A_y$$

é aberto e $A \cap K = \varnothing$, então $x \in A \subseteq X \setminus K$, donde $x \in \operatorname{int}(X \setminus K)$. Por conseguinte, $X \setminus K$ é aberto, logo K é fechado. $\square$

Teorema B.4.7 (Teorema de Borel-Lebesgue)**.** Um conjunto $X \subseteq \mathbb{R}$ é compacto se, e somente se, é fechado e limitado.

Demonstração. Seja $X \subseteq \mathbb{R}$ um conjunto compacto. Como

$$X \subseteq \mathbb{R} = \bigcup_{n \in \mathbb{N}} (-n, n),$$

então, da compacidade de X, existe $F \subseteq \mathbb{N}$ finito tal que

$$X \subseteq \bigcup_{n \in F} (-n, n)$$

donde X é limitado. O fato de que X é fechado vem do fato de que a reta é Hausdorff.

Reciprocamente, se $X \subseteq \mathbb{R}$ é fechado e limitado, então $X \subseteq [a, b]$, onde $a := \inf X$ e $b := \sup X$. Dada uma cobertura $\{A_\lambda\}_{\lambda \in \Lambda}$ de abertos de X, temos

$$[a, b] \subseteq \bigcup_{\lambda \in \Lambda} A_\lambda \cup (\mathbb{R} \setminus X).$$

Da compacidade de $[a, b]$ (ver Proposição B.4.2), existe $F \subseteq \Lambda$ finito tal que a coleção $\{A_\lambda\}_{\lambda \in F} \cup \{\mathbb{R} \setminus F\}$ ainda é uma cobertura de $[a, b]$, o que nos dá

$$X \subseteq \bigcup_{\lambda \in F} A_\lambda,$$

como queríamos demonstrar. $\square$

Teorema B.4.8. A compacidade é *invariante por funções contínuas*, isto é, se X e Y são espaços topológicos, X é compacto e $f : X \to Y$ é uma função contínua, então $f(X)$ é um subconjunto compacto de Y.

Demonstração. Sejam X e Y espaços topológicos, com X compacto, $f : X \to Y$ uma função contínua e $\{Y_\lambda\}_{\lambda \in \Lambda}$ uma cobertura de abertos de $f(X)$. Desse modo,

$$f(X) \subseteq \bigcup_{\lambda \in \Lambda} Y_\lambda,$$

o que nos dá

$$X \subseteq f^{-1}(f(X)) \subseteq f^{-1}\left(\bigcup_{\lambda\in\Lambda} Y_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(Y_\lambda).$$

Da continuidade de f, temos que $X_\lambda := f^{-1}(Y_\lambda)$ é aberto, para cada $\lambda \in \Lambda$. Logo, $\{X_\lambda\}_{\lambda\in\Lambda}$ é uma cobertura de X, que é um espaço compacto, logo admite uma subcobertura finita $\{X_\lambda\}_{\lambda\in F}$. Por conseguinte,

$$X \subseteq \bigcup_{\lambda\in F} X_\lambda,$$

o que implica

$$f(X) \subseteq f\left(\bigcup_{\lambda\in F} X_\lambda\right) = \bigcup_{\lambda\in F} f(X_\lambda) \subseteq \bigcup_{\lambda\in F} Y_\lambda.$$

Logo, $f(X)$ é compacto. □

Corolário B.4.9 (Teorema de Weierstrass)**.** Se X é um espaço topológico compacto e $f : X \to \mathbb{R}$ é uma função contínua, então existem $\min f(X)$ e $\max f(X)$.

Demonstração. Como $f(X)$ é compacto em $\mathbb{R}$, então, pelo Teorema de Borel-Lebesgue (ver Teorema B.4.7), o conjunto $f(X)$ é fechado e limitado em $\mathbb{R}$. Como $\alpha := \inf f(X)$ e $\beta := \sup f(X)$ são valores de aderência de $f(X)$, então $\alpha, \beta \in f(X)$. □

Corolário B.4.10. Se $f : [a, b] \to \mathbb{R}$ é contínua, então $f([a, b])$ é um intervalo compacto.

Demonstração. Como $[a, b]$ é um intervalo (já que $[a, b]$ é conexo) compacto, então $f([a, b])$ é um intervalo (já que $f([a, b])$ é conexo) compacto. □

Corolário B.4.11. Se X e Y são espaços topológicos, $f : X \to Y$ é uma função sobrejetiva contínua e X é compacto, então Y também é compacto.

Demonstração. Sendo f sobrejetiva, $Y = f(X)$ é compacto. □

Corolário B.4.12. Compacidade é um invariante topológico, isto é, se X e Y são espaços topológicos homeomorfos, então X é compacto se, e somente se, Y é compacto.

B.5 Espaços métricos

Os exemplos mais comuns de espaços topológicos são dados pelos espaços métricos.

Definição B.5.1. Uma **métrica** em um conjunto M é uma função $d : M \times M \to \mathbb{R}$ satisfazendo:

1. $d(x, y) \geq 0$, para quaisquer $x, y \in M$.
2. $d(x, y) = 0 \iff x = y$, para quaisquer $x, y \in M$.

3. $d(x, y) = d(y, x)$, para quaisquer $x, y \in M$.

4. $d(x, z) \leq d(x, y) + d(y, z)$, para quaisquer $x, y, z \in M$.

Um **espaço métrico** é um par (M, d), onde M é um conjunto e d é uma métrica em M.

Definição B.5.2 (Bolas e esferas)**.** Sejam M um espaço métrico e $a \in M$.

- A **bola aberta** de centro a e raio $\delta > 0$ é o conjunto

$$B(a, \delta) := \{x \in M : d(x, a) < \delta\}$$

- A **bola fechada** de centro a e raio $\delta > 0$ é o conjunto

$$B[a, \delta] := \{x \in M : d(x, a) \leq \delta\}$$

- A **esfera** de centro a e raio $\delta > 0$ é o conjunto

$$S(a, \delta) := \{x \in M : d(x, a) = \delta\}.$$

Um fato extremamente importante é que todo espaço métrico M pode ser naturalmente considerado um espaço topológico com a topologia dada pela coleção de todos os subconjuntos $A \subseteq M$ que satisfazem

$$\text{para todo } x \in A, \text{ existe } \delta > 0 \text{ tal que } B(x, \delta) \subseteq A.$$

Essa topologia é chamada a *topologia usual* do espaço métrico M.

Dados espaços métricos M e N, dizemos que $f : M \to N$ é uma imersão isométrica se $d(f(x), f(y)) = d(x, y)$, para quaisquer $x, y \in M$. Se, além disso, f for bijetiva, dizemos que f é uma isometria.

Definição B.5.3. Seja (M, d) um espaço métrico. Dizemos que uma sequência $(x_n) \subseteq M$

1. **converge** para $L \in M$ se $d(x_n, L) \xrightarrow{n\to\infty} 0$.

2. é **de Cauchy** se, para cada $\varepsilon > 0$, podemos obter $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies d(x_m, x_n) < \varepsilon.$$

Escreveremos $x_n \xrightarrow{n\to\infty} L$ ou $\lim\limits_{n\to\infty} x_n = L$ para significar que (x_n) converge para L. Quando acharmos que há risco de confusão, podemos escrever também $x_n \xrightarrow[d]{n\to\infty} L$ para deixar claro que a convergência é relativa à métrica d.

Teorema B.5.4 (Unicidade do limite)**.** Se (M, d) é um espaço métrico e $(x_n) \subseteq M$, então

$$\lim_{n\to\infty} x_n = P \text{ e } \lim_{n\to\infty} x_n = Q \implies P = Q.$$

Demonstração. Se $\lim\limits_{n\to\infty} x_n = P$ e $\lim\limits_{n\to\infty} x_n = Q$, então

$$d(P, Q) = \lim_{n\to\infty} d(P, Q) \leq \lim_{n\to\infty} [d(P, x_n) + d(x_n, Q)] = 0.$$

Por conseguinte,

$$d(P, Q) = 0,$$

donde $P = Q$. □

Definição B.5.5. Um espaço métrico (M, d) é **completo** se

$$(x_n) \subseteq M \text{ é uma sequência de Cauchy} \implies \text{existe } L \in M \text{ tal que } \lim_{n\to\infty} x_n = L.$$

Definição B.5.6 (Limite)**.** Sejam M e N espaços métricos, $f : X \subseteq M \to N$ uma função e $a \in X'$. Escrevemos $\lim\limits_{x\to a} f(x) = L$ se, para qualquer $(x_n) \subseteq X \setminus \{a\}$,

$$x_n \xrightarrow[d_M]{n\to\infty} a \implies f(x_n) \xrightarrow[d_N]{n\to\infty} L.$$

Teorema B.5.7. (LIMA, 2017, Proposição 9, p. 130) Sejam (M, d_M) e (N, d_N) espaços métricos e $a \in M$. Então, $f : M \to N$ é contínua em a se, e somente se, para qualquer $(x_n) \subseteq M$,

$$x_n \xrightarrow[d_M]{n\to\infty} a \implies f(x_n) \xrightarrow[d_N]{n\to\infty} f(a).$$

Teorema B.5.8 (Monotonicidade do Limite)**.** Sejam M um espaço métrico, $V \subseteq M$ uma vizinhança de a e $f, g : M \to \mathbb{R}$ funções tais que $f(x) \leq g(x)$, para todo $x \in V \setminus \{a\}$. Assim,

$$\lim_{x\to a} f(x) = P \text{ e } \lim_{x\to a} g(x) = Q \implies P \leq Q.$$

Demonstração. Suponhamos, por redução ao absurdo, que $Q < P$. Seja $(x_n) \subseteq M \setminus \{a\}$ tal que $\lim\limits_{n\to\infty} x_n = a$. Como V é uma vizinhança de a, então existe $\delta > 0$ tal que $B_\delta(a) \subseteq V$. Do fato de que $\lim\limits_{n\to\infty} d(x_n, a) = 0$, existe $n_1 \in \mathbb{N}$ tal que

$$n \geq n_1 \implies d(x_n, a) < \delta.$$

Sendo $\lim\limits_{n\to\infty} f(x_n) = P$ e $\lim\limits_{n\to\infty} g(x_n) = Q$, então, tomando $\varepsilon := \dfrac{P - Q}{2}$, existem $n_2, n_3 \in \mathbb{N}$ tais que

$$n \geq n_2 \implies |f(x_n) - P| < \varepsilon$$
$$n \geq n_3 \implies |g(x_n) - Q| < \varepsilon.$$

Assim, pondo $n_0 := \max\{n_1, n_2, n_3\}$, temos

$$n \geq n_0 \implies g(x_n) < Q + \varepsilon = P - \varepsilon < f(x_n) \text{ e } d(x_n, a) < \delta$$

donde $g(x_{n_0}) < f(x_{n_0})$ e $x_{n_0} \in B_\delta(a) \subseteq V$, o que é absurdo. □

Teorema B.5.9. Sejam M um espaço métrico, $f, g : X \subseteq M \to \mathbb{R}$ funções e $a \in X'$ tais que $\lim\limits_{x \to a} f(x)$ e $\lim\limits_{x \to a} g(x)$ existem. Se $\lim\limits_{x \to a} f(x) \leq \lim\limits_{x \to a} g(x)$ e $\varepsilon > 0$, então existe $\delta > 0$ tal que

$$f(x) < g(x) + \varepsilon, \text{ para cada } x \in [B(a, \delta) \setminus \{a\}] \cap X.$$

Demonstração. Se, para cada $\delta > 0$, existe $x \in [B_\delta(a) \setminus \{a\}] \cap X$ tal que $f(x) \geq g(x) + \varepsilon$, então, para cada $n \in \mathbb{N}$, o conjunto

$$X_n := \left\{ x \in X : x \in B\left(a, \frac{1}{n}\right) \setminus \{a\} \quad \text{e} \quad f(x) \geq g(x) + \varepsilon \right\}$$

é não vazio. Escolhendo, para cada $n \in \mathbb{N}$, um $x_n \in X$, obtemos uma sequência (x_n) tal que $x_n \in B(a, \frac{1}{n}) \setminus \{a\}$ e $f(x_n) \geq g(x_n) + \varepsilon$. Como $\lim\limits_{n \to \infty} x_n = a$, pelo Teorema do Confronto (ver Teorema A.1.10), então

$$\lim_{x \to a} f(x) = \lim_{n \to \infty} f(x_n) \geq \lim_{n \to \infty} g(x_n) + \varepsilon > \lim_{n \to \infty} g(x_n) = \lim_{x \to a} g(x),$$

pela Monotonicidade do Limite (ver Teorema A.1.11). $\square$

As provas dos dois teoremas a seguir são fáceis adaptações das provas expostas em (LIMA, 2019, Teorema 4, p. 198; Teorema 7, p. 200).

Teorema B.5.10 (Teorema do Confronto)**.** Sejam M um espaço métrico, $f, g, h : X \subseteq M \to \mathbb{R}$ funções e $a \in X'$. Se

$$\begin{cases} f(x) \leq g(x) \leq h(x), \text{ para todo } x \in X \setminus \{a\} \\ \lim\limits_{x \to a} f(x) = \lim\limits_{x \to a} h(x) = L, \end{cases}$$

então $\lim\limits_{x \to a} g(x) = L$.

Teorema B.5.11. Sejam M um espaço métrico, $f, g : X \subseteq M \to \mathbb{R}$ funções e $a \in X'$. Se $\lim\limits_{x \to a} f(x) = P$ e $\lim\limits_{x \to a} g(x) = Q$, então $\lim\limits_{x \to a} [f(x) + g(x)] = P + Q$ e $\lim\limits_{x \to a} [f(x) \cdot g(x)] = P \cdot Q$. Se $Q \neq 0$ e existe uma vizinhança $V \subseteq M$ de a tal que tal que $g(x) \neq 0$, para todo $x \in [V \setminus \{a\}] \cap X$, então $\lim\limits_{x \to a} \dfrac{f(x)}{g(x)} = \dfrac{P}{Q}$. Além disso, se $P = 0$ e g é limitada em $X \setminus \{a\}$, então $\lim\limits_{x \to a} [f(x) g(x)] = 0$.

APÊNDICE C – Álgebra Linear

Neste capítulo enunciaremos as principais definições e teoremas de Álgebra Linear que nos serão úteis. As proposições, os teoremas, etc. sem demonstração ou i) são fatos facilmente verificáveis (e nenhuma referência será citada diretamente) ou ii) são de difícil demonstração, mas são teoremas clássicos. As principais referências desse capítulo serão (LIMA, 2018) e (BUENO, 2006).

C.1 Espaços vetoriais

Definição C.1.1. Um conjunto E munido de operações $+ : E \times E \to E$ e $\cdot : \mathbb{R} \times E \to E$ é um espaço vetorial se

1. $u + (v + w) = (u + v) + w$, para quaisquer $u, v, w \in E$.
2. $u + v = v + u$, para quaisquer $u, v \in E$.
3. Existe $0_E \in E$ tal que $u + 0_E = u$, para todo $u \in E$.
4. Para cada $u \in E$, existe $-u \in E$ tal que $u + (-u) = 0_E$.
5. $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$, para quaisquer $\alpha, \beta \in \mathbb{R}$ e $v \in E$.
6. $\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$, para quaisquer $u, v \in E$ e $\alpha \in \mathbb{R}$.
7. $\alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) \cdot v$, para quaisquer $\alpha, \beta \in \mathbb{R}$ e $v \in E$.
8. $1 \cdot v = v$, para todo $v \in E$.

Exemplo C.1.2. $\mathbb{R}$ é um espaço vetorial com as operações de adição e multiplicação.

Exemplo C.1.3. Dados X um conjunto e E um espaço vetorial, o conjunto $\mathcal{F}(X, E)$ das funções $f : X \to E$ é um espaço vetorial com as operações

$$\begin{aligned} (f + g)(u) &:= f(u) + g(u), \quad \text{para todo} \quad u \in X, \\ (\lambda \cdot f)(u) &:= \lambda \cdot f(u), \quad \text{para quaisquer} \quad u \in X \quad \text{e} \quad \lambda \in \mathbb{R}\,. \end{aligned}$$

Exemplo C.1.4 (Espaço de sequências)**.** Defina, para cada $p > 0$, o conjunto

$$\ell^p(\mathbb{N}) := \left\{ (x_n) \subseteq \mathbb{R} : \sum_n |x_n|^p < +\infty \right\}$$

e seja

$$\ell^\infty(\mathbb{N}) := \left\{(x_n) \subseteq \mathbb{R} : \sup_{n\in\mathbb{N}} |x_n| < +\infty\right\}.$$

Então $\ell^p(\mathbb{N})$, para $p \in (0,\infty]$, é um subconjunto de $\mathcal{F}(\mathbb{N},\mathbb{R})$ que ainda é um espaço vetorial com as operações herdadas.

Demonstração. Seja $0 < p < \infty$. Observe que

$$\left(\frac{|x+y|}{2}\right)^p \leq \left(\frac{|x|+|y|}{2}\right)^p \leq \max\{|x|^p, |y|^p\} \leq |x|^p + |y|^p,$$

o que nos dá

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

Isso garante que a soma de duas sequências em $\ell^p(\mathbb{N})$ continua em $\ell^p(\mathbb{N})$.

Agora sejam $(x_n), (y_n) \in \ell^\infty(\mathbb{N})$. Então existem $C_1, C_2 > 0$ tais que

$$|x_n| \leq C_1 \quad \text{e} \quad |y_n| \leq C_2, \quad \text{para todo} \quad n \in \mathbb{N}.$$

Logo,

$$|x_n + y_n| \leq |x_n| + |y_n| \leq C_1 + C_2, \quad \text{para todo} \quad n \in \mathbb{N}.$$

Isso prova que $(x_n) + (y_n) \in \ell^\infty(\mathbb{N})$. □

Exemplo C.1.5 (Espaço produto)**.** Se $E_1, \dots, E_n$ são espaços vetoriais, então $E := E_1 \times \cdots \times E_n$ é um espaço vetorial com as operações

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) := (a_1 +_{E_1} b_1, \dots, a_n +_{E_n} b_n)$$

e

$$\alpha \cdot (a_1, \dots, a_n) := (\alpha \cdot_{E_1} a_1, \dots, \alpha \cdot_{E_n} a_n).$$

Seja E um espaço vetorial. Dado $A \subseteq E$, o **espaço gerado** por A é o conjunto

$$\operatorname{span} A := \left\{\sum_{i=1}^n \alpha_i v_i : \alpha_1, \dots, \alpha_n \in \mathbb{R}, v_1, \dots, v_n \in A, n \in \mathbb{N}\right\}.$$

Dizemos que $X \subseteq E$ **gera** Y (ou que Y é **gerado** por X) se $Y = \operatorname{span}(X)$.

Um conjunto $X \subseteq E$ é **linearmente independente** (**L.I.**) se

$$\sum_{i=1}^n \alpha_i v_i = 0_E \implies \alpha_1 = \cdots = \alpha_n = 0, \text{ qualquer que seja } \sum_{i=1}^n \alpha_i v_i \in \operatorname{span} X.$$

Um conjunto $B \subseteq E$ é uma **base** de E se B gera E e é linearmente independente.

Proposição C.1.6. Se E é um espaço vetorial, $X := \{v_i\}_{i=1}^n \subseteq E$ é um conjunto L.I. e $v \in \operatorname{span} X$, então existem únicos escalares $\alpha_1, \dots, \alpha_n$ tais que $v = \sum_{i=1}^n \alpha_i v_i$.

Proposição C.1.7. Se E é um espaço vetorial e $B \subseteq E$ é uma base, então, para cada $v \in E \setminus \{0_E\}$, existem únicos $u_1, \ldots, u_n \in B$ e únicos escalares $\alpha_1, \ldots, \alpha_n$ tais que $v = \sum_{i=1}^{n} \alpha_i u_i$ e $\alpha_i \neq 0$, para todo $i = 1, \ldots, n$.

Dizemos que um espaço vetorial E tem **dimensão finita** se existe $X \subseteq E$ finito tal que $\operatorname{span}(X) = E$. Não é difícil ver que se E tem dimensão finita, então E admite uma base finita.

Teorema C.1.8. Se E é um espaço vetorial de dimensão finita e $A, B \subseteq E$ são bases, então $|A| = |B|$.

Definição C.1.9. Se E é um espaço de dimensão finita, sua **dimensão** é a quantidade de elementos de (qualquer) uma de suas bases.

Teorema C.1.10. Se E é um espaço vetorial e $X \subseteq E$, são equivalentes:

1. X é uma base de E.
2. X é um conjunto L.I. maximal.
3. X é um conjunto gerador minimal.

Demonstração. Suponhamos 1) e provemos 2). Se X não fosse um conjunto L.I. maximal, então haveria $Y \subseteq E$ tal que $X \subsetneq Y$ e Y é L.I. Assim, existe $a \in Y$ tal que $a \notin X$. Como $a \in E$ e X é uma base de E, então $a \in \operatorname{span} X$, donde $Z := X \cup \{a\}$ não é L.I., o que é absurdo, já que $Z \subseteq Y$.

Suponhamos 2) e provemos 3). Se X não fosse um conjunto gerador minimal, então haveria $Y \subseteq E$ tal que $Y \subsetneq X$ e $\operatorname{span} Y = E$. Assim, existe $a \in X$ tal que $a \notin Y$. Como $a \in \operatorname{span} Y$, então $Z := Y \cup \{a\}$ não é um conjunto L.I., o que é absurdo, já que $Z \subseteq X$.

Por fim, suponhamos 3) e provemos 1). Se X não fosse uma base de E, então X não é L.I., donde existiria $a \in X$ tal que $a \in \operatorname{span} Y$, onde $Y := X \setminus \{a\}$. Logo, $\operatorname{span} Y = E$, contradizendo a minimalidade de X. $\square$

Teorema C.1.11 (Teorema do Completamento)**.** Se E é um espaço vetorial e $X \subseteq E$ é um conjunto L.I., então existe $\mathcal{B} \subseteq E$ base de E tal que $X \subseteq \mathcal{B}$.

Demonstração. Considere o conjunto

$$\mathcal{A} := \{Z \subseteq E : X \subseteq Z \text{ e } Z \text{ é um conjunto L.I.}\}.$$

Seja $\{H_\lambda\}_{\lambda \in \Lambda} \subseteq \mathcal{A}$ tal que

$$H_\mu \subseteq H_\lambda \text{ ou } H_\lambda \subseteq H_\mu, \text{ para quaisquer } \lambda, \mu \in \Lambda.$$

Note que

$$H_\mu \subseteq \bigcup_{\lambda \in \Lambda} H_\lambda, \text{ para todo } \mu \in \Lambda.$$

Sejam $u_1, \ldots, u_n \in \bigcup_{\lambda \in \Lambda} H_\lambda$ tais que

$$\alpha_1 u_1 + \cdots + \alpha_n u_n = 0_E.$$

Para cada $i = 1, \ldots, n$, existe $\lambda_i \in \Lambda$ tal que $u_i \in H_{\lambda_i}$. Como

$$\bigcup_{i=1}^{n} H_{\lambda_i} = H_{\lambda_0}, \text{ para algum } \lambda_0 \in \{\lambda_1, \ldots, \lambda_n\},$$

então

$$u_1, \ldots, u_n \in H_{\lambda_0},$$

que é um conjunto L.I., donde

$$\alpha_1 = \cdots = \alpha_n = 0.$$

Logo,

$$\bigcup_{\lambda \in \Lambda} H_\lambda \text{ é L.I.} \implies \bigcup_{\lambda \in \Lambda} H_\lambda \in \mathcal{A}.$$

Isto prova que $\bigcup_{\lambda \in \Lambda} H_\lambda$ é, com respeito à relação de ordem de inclusão de conjuntos, uma cota superior de $\{H_\lambda\}_{\lambda \in \Lambda}$.

Pelo Lema de Zorn, o conjunto $\mathcal{A}$ possui pelo menos um elemento maximal, digamos $\mathcal{B}$, com respeito à relação de ordem de inclusão de conjuntos.

Pelo Teorema C.1.10, o conjunto $\mathcal{B}$ é uma base de E. □

Corolário C.1.12 (Existência de base)**.** Se E é um espaço vetorial não trivial, então E admite pelo menos uma base.

Demonstração. Sendo E não trivial, existe $v \in E \setminus \{0_E\}$, donde $\{v\}$ é um conjunto L.I. de E. Pelo Teorema C.1.11, existe pelo menos uma base $\mathcal{B}$ de E tal que $\{v\} \subseteq \mathcal{B}$. □

Corolário C.1.13. Se E é um espaço vetorial de dimensão finita e $X \subseteq E$ é um conjunto L.I. tal que $|X| = \dim E$, então $\operatorname{span} X = E$.

Demonstração. Pelo Teorema do Completamento (ver Teorema C.1.11), o espaço E admite uma base B tal que $X \subseteq B$. Como $|B| = \dim E = |X|$, então $X = B$. □

Teorema C.1.14. Se F é um espaço vetorial de dimensão finita e $E \subseteq F$ é um subespaço vetorial tal que $\dim E = \dim F$, então $E = F$.

C.2 Transformações lineares

Dizemos que uma aplicação $T : E \to F$ é uma **transformação linear** se

$$T(u + \lambda v) = T(u) + \lambda T(v), \quad \text{para quaisquer} \quad u, v \in E \quad \text{e} \quad \lambda \in \mathbb{R}\,.$$

Se $T : E \to F$ é uma transformação linear, $Tu := T(u)$, para todo $u \in E$. Uma transformação linear bijetiva é dita um **isomorfismo linear**.

O conjunto de todas as transformações lineares $T : E \to F$ é denotado por $\mathcal{L}(E, F)$.

Transformações lineares $T : E \to E$ são chamadas **operadores lineares** em E; transformações lineares $T : E \to \mathbb{R}$ são chamadas **funcionais lineares**. Escrevemos $\mathcal{L}(E)$ em vez de $\mathcal{L}(E, E)$ e E^* em vez de $\mathcal{L}(E, \mathbb{R})$. O **espaço dual** de E é o espaço vetorial dado por E^* munido das operações usuais.

Teorema C.2.1. Sejam E um espaço vetorial, $B \subseteq E$ uma base e $f : B \to F$ uma aplicação. Então, existe uma única transformação linear $T : E \to F$ tal que

$$T(u) = f(u), \quad \text{para cada } u \in B.$$

Demonstração. Pelo Corolário C.1.7, para cada $v \in E$, existem únicos $u_1^v, \ldots, u_{n_v}^v \in B$ e escalares $\alpha_1^v, \ldots, \alpha_{n_v}^v$ tais que

$$v = \sum_{i=1}^{n_v} \alpha_i^v u_i^v, \quad \text{com} \quad \alpha_i \neq 0, \quad \text{para todo} \quad i = 1, \ldots, n.$$

Provaremos que a aplicação $T : E \to F$ dada por

$$T(v) := \sum_{i=1}^{n_v} \alpha_i^v f(u_i^v)$$

satisfaz as condições do enunciado. Claramente,

$$T(u) = f(u), \text{ para todo } u \in B.$$

T **é linear.** Dados $v, w \in E$ e um escalar $\lambda \neq 0$, ponha

$$\begin{cases} u_i := u_i^v, \text{ para todo } i = 1, \ldots, n_v; \\ u_{n_v+i} := u_i^w, \text{ para todo } i = 1, \ldots, n_w; \\ \alpha_i := \begin{cases} \alpha_i^v, \text{ se } i = 1, \ldots, n_v; \\ 0, \text{ caso contrário}; \end{cases} \\ \beta_i := \begin{cases} 0, \text{ se } i = 1, \ldots, n_v; \\ \alpha_i^w, \text{ caso contrário}; \end{cases} \\ n := n_v + n_w. \end{cases}$$

Nessas condições,

$$v + \lambda w = \sum_{i=1}^{n} (\alpha_i + \lambda \beta_i) u_i, \text{ com } \alpha_i + \lambda \beta_i \neq 0, \text{ para todo } i = 1, \ldots, n,$$

donde

$$T(v + \lambda w) = \sum_{i=1}^{n} (\alpha_i + \lambda \beta_i) f(u_i) = \sum_{i=1}^{n} \alpha_i f(u_i) + \lambda \sum_{i=1}^{n} \beta_i f(u_i) = T(u) = T(v) + \lambda T(w).$$

Isto prova que T é linear.

T **é único.** Seja $h : E \to F$ uma aplicação linear tal que

$$h(u) = f(u), \quad \text{para cada } u \in B.$$

Dado $v \in E$, temos

$$h(v) = h\left(\sum_{i=1}^{n_v} \alpha_i^v u_i^v\right) = \sum_{i=1}^{n_v} \alpha_i^v h(u_i^v) = \sum_{i=1}^{n_v} \alpha_i^v f(u_i^v) = T(v).$$

Logo, $h = T$, como queríamos demonstrar. $\square$

Teorema C.2.2 (Transformação linear injetiva leva L.I. em L.I.)**.** Se E e F são espaços vetoriais e $T : E \to F$ é uma aplicação linear, então T é injetiva se, e somente se,

$$T(X) \text{ é um conjunto L.I., para cada conjunto L.I. } X \subseteq E.$$

Demonstração. Suponhamos T injetiva e que $X \subseteq E$ é um conjunto linearmente independente. Dado $\{u_i\}_{i=1}^n \subseteq X$, verificaremos que $\{Tu_i\}_{i=1}^n$ é L.I.

Sejam $\beta_1, \ldots, \beta_n \in \mathbb{R}$ tais que

$$T(\beta_1 u_1 + \cdots + \beta_n u_n) = \beta_1 T u_1 + \cdots + \beta_n T u_n = 0_F.$$

Pela injetividade de T, temos

$$\beta_1 u_1 + \cdots + \beta_n u_n = 0_E \implies \beta_1 = \cdots = \beta_n = 0.$$

Logo, $\{Tu_i\}_{i=1}^n$ é L.I., donde $T(X)$ é L.I..

Reciprocamente, suponhamos que $T(X)$ é L.I., para cada conjunto L.I. $X \subseteq E$. Se T não fosse injetiva, então haveria $x \in E \setminus \{0_E\}$ tal que $Tx = 0_F$. Como $\{x\}$ é L.I., então, por hipótese,

$$\{0_F\} = \{Tx\} = T(\{x\})$$

é L.I., o que é absurdo. $\square$

Corolário C.2.3. Se E e F são espaços vetoriais, então uma aplicação linear $T : E \to F$ é bijetiva se, e somente se, para cada $B \subseteq E$ base, $T(B)$ é uma base de F.

Demonstração. Suponhamos que T é bijetiva. Se X é uma base de E, então $T(X)$ é um conjunto L.I., pelo Teorema C.2.2. Dado $v \in F$, existe um único $u \in E$ tal que $Tu = v$, pela bijetividade de T, donde existem $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ e $x_1, \ldots, x_n \in X$ tais que

$$u = \sum_{i=1}^{n} \alpha_i x_i \implies v = Tu = \sum_{i=1}^{n} \alpha_i T x_i \implies v \in \operatorname{span} T(X).$$

Logo, $\operatorname{span} T(X) = F$.

Reciprocamente, suponhamos que $T(B)$ é uma base de F, para cada $B \subseteq E$ base. Se houvesse $u \in E \setminus \{0_E\}$ tal que $Tu = 0_F$, então, pelo Teorema do Completamento (ver Teorema C.1.11), existe $X \subseteq E$ base tal que $u \in X$, donde, por hipótese, $T(X)$ é uma base de F, o que é absurdo, já que $0_F = Tu \in T(X)$. Logo, $\ker T = \{0_E\}$, isto é, T é injetiva.

Provaremos agora que T também é sobrejetiva. Se E é um espaço não trivial, então existe $B \subseteq E$ base, pelo Teorema C.1.12. Daí, por hipótese, $T(B)$ é uma base de F. Dado $v \in F$, existem escalares $\alpha_1, \ldots, \alpha_n$ e $u_1, \ldots, u_n \in B$ tais que

$$v = \sum_{i=1}^{n} \alpha_i T u_i = T\left(\sum_{i=1}^{n} \alpha_i u_i\right) \implies v \in T(E).$$

Logo, $T(E) = F$, isto é, T é sobrejetiva. □

O próximo corolário deixa claro que dois espaços vetoriais são isomorfos se, e somente se, possuem a mesma dimensão.

Corolário C.2.4. Se E e F são espaços vetoriais de dimensão finita, então $\dim E = \dim F$ se, e somente se, existe uma transformação linear bijetiva $T : E \to F$.

Demonstração. Suponhamos que $\dim E = \dim F$. Assim, se A é uma base de E e B é uma base de F, então

$$|A| = \dim E = \dim F = |B|,$$

isto é, existe uma aplicação $g : A \to B$ bijetiva. Pelo Teorema C.2.1, existe uma transformação linear $T : E \to F$ tal que

$$T(x) = g(x), \quad \text{para todo} \quad x \in A.$$

Provaremos agora que T é bijetiva.

Dado $v \in F$, existem escalares $\alpha_1, \ldots, \alpha_n$ e $v_1, \ldots, v_n \in B = g(A) = T(A)$ tais que $v = \sum_{i=1}^{n} \alpha_i v_i$. Para cada $i = 1, \ldots, n$, existe $u_i \in A$ tal que $Tu_i = v_i$, donde

$$v = \sum_{i=1}^{n} \alpha_i T u_i = T\left(\sum_{i=1}^{n} \alpha_i u_i\right)$$

e, portanto, $v \in T(E)$. Logo, T é sobrejetiva.

Se $u \in E$ tal que $Tu = 0_F$, então existem escalares $\alpha_1, \dots, \alpha_n$ e $v_1, \dots, v_n \in A$ tais que $u = \sum_{i=1}^{n} \alpha_i v_i$, donde

$$0_F = Tu = \sum_{i=1}^{n} \alpha_i T v_i = \sum_{i=1}^{n} \alpha_i g(v_i) \implies \alpha_1 = \cdots = \alpha_n = 0$$

e, portanto, $u = 0_E$. Logo, $\ker T = \{0_E\}$, isto é, T é injetiva.

Reciprocamente, suponhamos que existe uma aplicação linear $T : E \to F$ bijetiva.

Se $A \subseteq E$ é uma base, então $T(A)$ é uma base de F, pelo Corolário C.2.3, donde

$$\dim F = |T(A)| = |A| = \dim E,$$

como queríamos demonstrar. □

Sejam E e F espaço vetoriais, com $V := \{v_i\}_{i=1}^{n} \subseteq E$ e $W := \{w_i\}_{i=1}^{m} \subseteq F$ bases de E e de F, respectivamente. Se $T : E \to F$ é uma aplicação linear, então, para cada $j = 1, \dots, n$, existem únicos escalares $\alpha_{1j}, \dots, \alpha_{mj}$ tais que

$$T(v_j) = \sum_{i=1}^{m} \alpha_{ij} w_i.$$

Como, para cada $x \in E$, existem únicos escalares $x_1, \dots, x_n$ tais que

$$x = \sum_{j=1}^{n} x_j v_j,$$

então

$$Tx = \sum_{j=1}^{n} x_j T(v_j) = \sum_{j=1}^{n} x_j \sum_{i=1}^{m} \alpha_{ij} w_i = \sum_{j=1}^{n} \sum_{i=1}^{m} x_j \alpha_{ij} w_i,$$

o que nos permite escrever

$$Tx = \begin{bmatrix} w_1 & \dots & w_m \end{bmatrix} \cdot \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & \vdots & \vdots \\ \alpha_{m1} & \dots & \alpha_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

A matriz de T nas bases V e W é

$$[T]_W^V := \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1n} \\ \vdots & \vdots & \vdots \\ \alpha_{m1} & \dots & \alpha_{mn} \end{bmatrix}.$$

Note que $\dim \mathcal{L}(E, F) = \dim E \cdot \dim F$.

Teorema C.2.5. Sejam E um espaço de dimensão n e F um espaço de dimensão m. Se A e B são, respectivamente, bases de E e de F, então a aplicação

$$T \in \mathcal{L}(E, F) \longmapsto [T]_B^A \in \mathcal{M}(m \times n)$$

é um isomorfismo.

Teorema C.2.6. Se E e F são espaços vetoriais e $T : E \to F$ é uma bijeção linear, então $T^{-1} : F \to E$ também é.

Demonstração. Sejam $u, v \in F$ e $\lambda \in \mathbb{R}$. Como T é bijetiva, então existem únicos $a, b \in E$ tais que $T(a) = u$ e $T(b) = v$, donde

$$u + \lambda v = T(a) + \lambda T(b) = T(a + \lambda b),$$

e, portanto,

$$T^{-1}(u + \lambda v) = a + \lambda b = T^{-1}(u) + \lambda T^{-1}(v).$$

Logo, T é linear. □

Teorema C.2.7 (Teorema do Núcleo e Imagem)**.** Se E e F são espaços vetoriais de dimensão finita e $T : E \to F$ é uma aplicação linear, então

$$\dim E = \dim \ker T + \dim T(E).$$

Demonstração. Seja $B := \{v_i\}_{i=1}^{k}$ uma base de $\ker T$. Se B for uma base de E, então, dado $v \in E$, existem únicos escalares $\alpha_1, \ldots, \alpha_n$ tais que $v = \sum_{i=1}^{n} \alpha_i v_i$, donde

$$Tv = \sum_{i=1}^{n} \alpha_i T v_i = 0_F \implies v \in \ker T.$$

Logo,

$$T(E) = \{0_F\} \implies \dim T(E) = 0 \implies \dim E = \dim \ker T + \dim T(E).$$

Suponhamos agora que B não é uma base de E. Pelo Teorema do Completamento (ver Teorema C.1.11), existem vetores $v_{k+1}, \ldots, v_n \in E$ tais que $B \cup \{v_{k+1}, \ldots, v_n\}$ é uma base de E.

Nosso objetivo é provar que $Y := \{T v_{k+1}, \ldots, T v_n\}$ é uma base de $T(E)$.

Começaremos verificando que Y é um conjunto L.I.. Sejam $\alpha_{k+1}, \ldots \alpha_n$ escalares tais que $\sum_{i=k+1}^{n} \alpha_i T v_i = 0_F$. Assim,

$$T\left(\sum_{i=k+1}^{n} \alpha_i v_i\right) = 0_F \implies \sum_{i=k+1}^{n} \alpha_i v_i \in \ker T.$$

Como B é uma base de $\ker T$, então existem escalares $\alpha_1, \ldots, \alpha_k$ tais que

$$\sum_{i=k+1}^{n} \alpha_i v_i = \sum_{i=1}^{k} \alpha_i v_i \implies \sum_{i=1}^{k} \alpha_i v_i - \sum_{i=k+1}^{n} \alpha_i v_i = 0 \implies \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0.$$

Logo, Y é linearmente independente.

Por fim, verificaremos que $\operatorname{span} Y = T(E)$. Dado $y \in T(E)$, existe $x \in E$ tal que $Tx = y$. Como $x \in E$, então existem escalares $x_1, \ldots, x_n$ tais que

$$x = \sum_{i=1}^{n} x_i v_i \implies y = Tx = \sum_{i=k+1}^{n} x_i T v_i \implies y \in \operatorname{span} Y.$$

Portanto, Y é uma base de $T(E)$, donde $\dim T(E) = n - k$. □

Corolário C.2.8. Se E e F são espaços vetoriais de mesma dimensão finita e $T : E \to F$ é uma aplicação linear, então T é injetiva se, e somente se, T é sobrejetiva.

Demonstração. Suponhamos que T é injetiva. Pelo Teorema do Núcleo e Imagem (ver Teorema C.2.7), temos

$$\dim T(E) = \dim E = \dim F,$$

donde $T(E) = F$, pelo Teorema C.1.14, e, portanto, T é sobrejetiva.

Reciprocamente, suponhamos que T é sobrejetiva. Como $T(E) = F$, então

$$\dim E + \dim \ker T = \dim F + \dim \ker T = \dim T(E) + \dim \ker T = \dim E,$$

donde $\dim \ker T = 0$ e, portanto, $\ker T = \{0_F\}$, isto é, T é injetiva. □

Em dimensão infinita o resultado anterior não vale sempre: o operador linear $T : \mathbb{R}[x] \to \mathbb{R}[x]$ tal que $T(1) = x^2, T(x) = x^4, T(x^2) = x^6$, etc. é injetivo, mas não é sobrejetivo.

Teorema C.2.9. O vetor nulo é o único vetor de um espaço vetorial que zera em todos os funcionais lineares.

Demonstração. Sejam V um espaço vetorial e $v \in V \setminus \{0\}$. Como $\{v\}$ é um conjunto linearmente independente, existe $B \subseteq V$ base tal que $v \in B$. Pelo Teorema C.2.1, existe uma (única) aplicação linear $\xi : V \to \mathbb{R}$ tal que

$$\xi(v) = 1 \quad \text{e} \quad \xi(w) = 0, \quad \text{para todo} \quad w \in B \setminus \{v\}.$$

Isto prova que, dado um vetor não nulo v de um espaço vetorial V, é sempre possível encontrar um funcional linear $\xi \in V^*$ tal que $\xi(v) \neq 0$. □

Terminamos essa seção com um critério em geral de fácil verificação pra determinar se uma transformação linear é injetiva ou não.

Proposição C.2.10. Seja $T : E \to F$ uma transformação linear. Então T é injetiva se, e somente se, $\ker T := \{u \in E : Tu = 0\} = \{0\}$.

C.3 Espaços normados

Definição C.3.1. Uma **norma** em E é uma função $\|\cdot\| : E \to \mathbb{R}$ tal que

1. $\|u\| \geq 0$, para qualquer $u \in E$;
2. $\|u\| = 0 \iff u = 0$, para qualquer $u \in E$;
3. $\|\lambda \cdot u\| = |\lambda| \cdot \|u\|$, para quaisquer $u \in E$ e $\lambda \in \mathbb{R}$;
4. $\|u + v\| \leq \|u\| + \|v\|$, para quaisquer $u, v \in E$.

Se $\|\cdot\|$ é uma norma em E, dizemos que $(E, \|\cdot\|)$ é um **espaço vetorial normado**.

Observe que toda norma $\|\cdot\|$ em um espaço vetorial E induz uma métrica em E dada por

$$d(x, y) := \|x - y\|.$$

Teorema C.3.2 (Continuidade da norma)**.** Se E é um espaço vetorial e $\|\cdot\|$ é uma norma em E, então $\|\cdot\|$ é contínua.

Demonstração. Sejam $a \in E$ e $(x_n) \subseteq E$ tal que $x_n \xrightarrow[\|\cdot\|]{n\to\infty} a$. Como

$$0 \leq \big|\|x_n\| - \|a\|\big| \leq \|x_n - a\| \xrightarrow{n\to\infty} 0,$$

então, pelo Teorema do Confronto (ver Teorema A.1.10),

$$\|x_n\| \xrightarrow{n\to\infty} \|a\|.$$

Logo, $\|\cdot\|$ é uma função contínua. □

Exemplo C.3.3. A função módulo em $\mathbb{R}$ é uma norma.

Exemplo C.3.4. Se E é um espaço vetorial e $\{v_i\}_{i=1}^n$ é uma base de E, então, dado $v \in E$, existem únicos $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tais que $v = \sum_{i=1}^{n} \alpha_i v_i$. As expressões

$$\|v\|_1 := \sum_{i=1}^{n} |\alpha_i| \text{ e } \|v\|_\infty := \max_{i=1,\ldots,n} |\alpha_i|$$

definem normas sobre E.

Exemplo C.3.5. Dado $p \geq 1$, a expressão

$$\|(x_n)\|_p := \begin{cases} \left(\sum_{n=1}^{\infty} |x_n|^p\right)^{1/p}, & \text{se} \quad 1 \leq p < +\infty, \\ \sup_{n\in\mathbb{N}} |x_n|, & \text{se} \quad p = +\infty, \end{cases}$$

define uma norma sobre $\ell^p(\mathbb{N})$ (veja Exemplo C.1.4).

Demonstração. Ver Teorema C.3.7. $\square$

Teorema C.3.6 (Desigualdade de Hölder)**.** Sejam $f \in \ell^p(\mathbb{N})$ e $g \in \ell^q(\mathbb{N})$ com

$$p, q \in [1, +\infty] \quad \text{tais que} \quad \frac{1}{p} + \frac{1}{q} = 1.$$

Nessas condições,

$$fg \in \ell^1(\mathbb{N}) \quad \text{e} \quad \|fg\|_1 \leq \|f\|_p \|g\|_q.$$

Demonstração. Vamos provar primeiro para $p, q \in (1, +\infty)$. Vamos começar mostrando que $fg \in \ell^1(\mathbb{N})$. Pela desigualdade de Young (ver Teorema 6.2.17),

$$|fg| \leq \frac{|f|^p}{p} + \frac{|g|^q}{q},$$

donde vem

$$\|fg\|_1 = \sum_n |(fg)(n)| \leq \sum_n \left(\frac{|f(n)|^p}{p} + \frac{|g(n)|^q}{q} \right) = \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q} < +\infty.$$

Isso prova que $fg \in \ell^1$. Disso também vem que se $\|f\|_p = \|g\|_q = 1$, então

$$\|fg\|_1 \leq \frac{1}{p} + \frac{1}{q} = 1.$$

Aplicando esse resultado às funções $f' := f/\|f\|_p$ e $g' := g/\|g\|_q$, temos

$$\left\| \frac{f}{\|f\|_p} \frac{g}{\|g\|_q} \right\|_1 = \|f'g'\|_1 \leq 1,$$

donde vem

$$\|fg\|_1 \leq \|f\|_p \|g\|_q,$$

como queríamos demonstrar. (Observe que a desigualdade vale trivialmente quando $\|f\|_p = 0$ ou $\|g\|_q = 0$.)

Sejam agora $p = 1$ e $q = +\infty$. Então

$$\|fg\|_1 = \sum_n |f(n)g(n)| \leq \sum_n (|f(n)| \|g\|_\infty) = \|g\|_\infty \sum_n |f_n| = \|g\|_\infty \|f\|_1,$$

como queríamos demonstrar. $\square$

A subaditividade de $\|\cdot\|_p$ é uma desigualdade particularmente especial:

Teorema C.3.7 (Desigualdade de Minkowski)**.** Sejam $f, g \in \ell^p(\mathbb{N})$ com $p \in [1, +\infty]$. Então,

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Demonstração. Para $p = 1$, temos

$$\|f+g\|_1 = \sum_n |(f+g)(n)| \leq \sum_n (|f(n)| + |g(n)|) = \sum_n |f(n)| + \sum_n |g(n)| = \|f\|_1 + \|g\|_1.$$

Para $p = +\infty$, temos

$$|f(n)| \leq \|f\|_\infty \quad \text{e} \quad |g(n)| \leq \|g\|_\infty, \quad \text{para todo} \quad n \in \mathbb{N},$$

donde vem

$$|(f+g)(n)| \leq |f(n)| + |g(n)| \leq \|f\|_\infty + \|g\|_\infty, \quad \text{para todo} \quad n \in \mathbb{N},$$

e, portanto,

$$\|f+g\|_\infty = \sup_n |(f+g)(n)| \leq \|f\|_\infty + \|g\|_\infty.$$

Por fim, suponhamos agora $p > 1$ e seja q tal que $1/p + 1/q = 1$. Veja que a desigualdade é trivialmente verdadeira se $\|f+g\|_p = 0$, então consideremos apenas o outro caso. Observe que

$$|f+g|^p = |f+g||f+g|^{p-1} \leq |f||f+g|^{p-1} + |g||f+g|^{p-1}.$$

Aplicando a desigualdade de Hölder (veja o Teorema C.3.6) para as funções $|f||f+g|^{p-1}$ e $|g||f+g|^{p-1}$, obtemos

$$\||f||f+g|^{p-1}\|_1 \leq \|f\|_p \||f+g|^{p-1}\|_q = \|f\|_p \left(\sum_n |(f+g)(n)|^p \right)^{1/q}$$

e

$$\||g||f+g|^{p-1}\|_1 \leq \|g\|_p \||f+g|^{p-1}\|_q = \|g\|_p \left(\sum_n |(f+g)(n)|^p \right)^{1/q}.$$

Logo,

$$\begin{aligned} \|f+g\|_p^p = \sum_n |(f+g)(n)|^p &= \||f+g|^p\|_1 \leq \||f||f+g|^{p-1}\|_1 + \||g||f+g|^{p-1}\|_1 \\ &\leq \left(\sum_n |(f+g)(n)|^p \right)^{1/q} (\|f\|_p + \|g\|_p) = \|f+g\|_p^{p/q}(\|f\|_p + \|g\|_p) \\ &= \|f+g\|_p^{p-1}(\|f\|_p + \|g\|_p), \end{aligned}$$

já que $1 + p/q = p$. Como $0 < \|f+g\|_p < +\infty$ (veja Exemplo C.1.4), podemos dividir ambos os membros da desigualdade obtida acima por $\|f+g\|_p^{p-1}$, o que nos dá

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p,$$

como queríamos demonstrar. □

Proposição C.3.8. Dados $p < q$ em $[1, +\infty]$, temos

$$\|x\|_q \leq \|x\|_p, \quad \text{para todo} \quad x \in \ell^p(\mathbb{N}).$$

Em particular, $\ell^p(\mathbb{N}) \subseteq \ell^q(\mathbb{N})$.

Demonstração. Suponhamos $q < +\infty$. Seja $x \in \ell^p(\mathbb{N})$. O resultado é óbvio se $\|x\|_p = 0$. Suponhamos então $\|x\|_p > 0$. Defina

$$e := x/\|x\|_p.$$

Como $1 = \|e\|_p^p = \sum_i |e_i|^p$, então cada $|e_i| \leq 1$. Logo, $|e_j|^q \leq |e_j|^p$, para todo $j \in \mathbb{N}$, donde vem

$$\|e\|_q = \left(\sum_i |e_i|^q\right)^{1/q} \leq \left(\sum_i |e_i|^p\right)^{1/q} = 1^{1/q} = 1$$

e, portanto,

$$\|x\|_q \leq \|x\|_p < +\infty.$$

Se $q = +\infty$, dado $i \in \mathbb{N}$, temos $|x_i|^p \leq \sum_j |x_j|^p = \|x\|_p^p$, donde vem $|x_i| \leq \|x\|_p$. Logo,

$$\|x\|_\infty = \sup_i |x_i| \leq \|x\|_p < +\infty,$$

o que termina a prova. □

Exemplo C.3.9. Se $E_1, \dots, E_d$ são espaços normados e $E := E_1 \times \cdots \times E_d$, então a expressão

$$\|x\|_\infty := \max_{i=1,\dots,d} \|x_i\|, \quad \text{para todo} \quad x = (x_1, \dots, x_d) \in E,$$

define uma norma sobre E. (Veja o Exemplo C.1.5.) Além disso, se $\psi : [0, +\infty) \to [0, +\infty)$ é uma função convexa e crescente tal que

$$\psi(0) = 0 \quad \text{e} \quad \lim_{x \to +\infty} \psi(x) = +\infty,$$

então a expressão

$$\|x\|_\psi := \inf\left\{a > 0 : \sum_{i=1}^d \psi\left(\frac{\|x_i\|_{E_i}}{a}\right) \leq 1\right\}, \quad \text{para todo} \quad x = (x_1, \dots, x_d) \in E,$$

define uma norma sobre E.

Resposta. Se $x = (x_1, \dots, x_d) \in E \setminus \{0_E\}$, então

$$\|x_i\|_{E_i} > 0, \text{ para algum } i \in [d].$$

Como

$$\frac{\|x_i\|_{E_i}}{a} \xrightarrow{a \to 0^+} +\infty,$$

então

$$\psi\left(\frac{\|x_i\|_{E_i}}{a}\right) \xrightarrow{a \to 0^+} +\infty$$

donde

$$\sum_{j=1}^d \psi\left(\frac{\|x_j\|_{E_j}}{a}\right) \geq \psi\left(\frac{\|x_i\|_{E_i}}{a}\right) \xrightarrow{a \to 0^+} +\infty.$$

Em particular, para todo $a > 0$ suficientemente pequeno,

$$\sum_{j=1}^{d} \psi\left(\frac{\|x_j\|_{E_j}}{a}\right) > 1.$$

Logo, $\|x\|_\psi \neq 0$.

Vamos verificar agora que

$$\|\lambda x\|_\psi = |\lambda|\|x\|_\psi, \text{ para quaisquer } \lambda \in \mathbb{R} \text{ e } x = (x_1, \ldots, x_d) \in E.$$

Sejam $\lambda \in \mathbb{R}$, com $\lambda \neq 0$, e $x \in E$. (A validade da identidade para $\lambda = 0$ é óbvia.)

Sejam

$$A := \left\{ a > 0 : \sum_{i=1}^{d} \psi\left(\frac{\|\lambda x_i\|_{E_i}}{a}\right) \leq 1 \right\},$$
$$B := \left\{ |\lambda| b > 0 : \sum_{i=1}^{d} \psi\left(\frac{\|x_i\|_{E_i}}{b}\right) \leq 1 \right\}.$$

Note que $A = B$. Com efeito,

$$\begin{aligned} a \in A &\iff a > 0 \text{ e } \sum_{i=1}^{d} \psi\left(\frac{\|\lambda x_i\|_{E_i}}{c}\right) \leq 1 \\ &\iff a > 0 \text{ e } \sum_{i=1}^{d} \psi\left(\frac{\|x_i\|_{E_i}}{|\lambda|^{-1} a}\right) \leq 1 \\ &\iff a = |\lambda| b, \text{ para algum } b > 0 \text{ tal que } \sum_{i=1}^{d} \psi\left(\frac{\|x_i\|_{E_i}}{b}\right) \leq 1 \\ &\iff a \in B. \end{aligned}$$

Logo,

$$\|\lambda x\|_\psi = \inf A = \inf B = |\lambda| \inf \left\{ b > 0 : \sum_{i=1}^{d} \psi\left(\frac{\|x_i\|_{E_i}}{b}\right) \leq 1 \right\} = |\lambda|\|x\|_\psi.$$

Por fim, vamos verificar agora que

$$\|x + y\|_\psi \leq \|x\|_\psi + \|y\|_\psi, \text{ para quaisquer } x, y \in \mathbb{R}^d.$$

Fixe $x, y \in \mathbb{R}^d$ e considere os conjuntos

$$A' := \left\{ a > 0 : \sum_{i=1}^{d} \psi\left(\frac{\|x_i\|_{E_i}}{a}\right) \leq 1 \right\},$$
$$B' := \left\{ b > 0 : \sum_{i=1}^{d} \psi\left(\frac{\|y_i\|_{E_i}}{b}\right) \leq 1 \right\}.$$

Vamos verificar que

$$\sum_{i=1}^{d} \psi\left(\frac{\|x_i + y_i\|_{E_i}}{a + b}\right) \leq 1, \text{ para quaisquer } a \in A' \text{ e } b \in B'.$$

Para isso, fixe $a \in A', b \in B'$ e considere

$$t := \frac{a}{a+b}.$$

Daí,

$$\begin{aligned}
\sum \psi\left(\frac{\|x_i + y_i\|_{E_i}}{a+b}\right) &\leq \sum \psi\left(\frac{\|x_i\|_{E_i}}{a+b} + \frac{\|y_i\|_{E_i}}{a+b}\right) \\
&= \sum \psi\left(t\frac{\|x_i\|_{E_i}}{a} + (1-t)\frac{\|y_i\|_{E_i}}{b}\right) \\
&\leq t\sum \psi\left(\frac{\|x_i\|_{E_i}}{a}\right) + (1-t)\sum \psi\left(\frac{\|y_i\|_{E_i}}{b}\right) \\
&\leq t + (1-t) = 1.
\end{aligned}$$

Disso vem que

$$\|x+y\|_\psi \leq a + b, \text{ para quaisquer } a \in A' \text{ e } b \in B',$$

o que nos permite concluir que

$$\|x+y\|_\psi \leq \|x\|_\psi + \|y\|_\psi,$$

como queríamos demonstrar. □

Para cada $p \geq 1$, a função

$$x \in [0, +\infty) \longmapsto \psi_p(x) := x^p$$

satisfaz as hipóteses do Exemplo C.3.9, fazendo com que $\|\cdot\|_p := \|\cdot\|_{\psi_p}$ seja uma norma sobre E. Note ainda que, dado $x = (x_1, \ldots, x_n) \in E$, temos

$$\begin{aligned}
\|x\|_p = \inf\left\{a > 0 : \sum_{i=1}^n \left(\frac{\|x_i\|_{E_i}}{a}\right)^p \leq 1\right\} &= \inf\left\{a > 0 : \left(\sum_{i=1}^n \|x_i\|_{E_i}^p\right)^{\frac{1}{p}} \leq a\right\} \\
&= \left(\sum_{i=1}^n \|x_i\|_{E_i}^p\right)^{\frac{1}{p}}.
\end{aligned}$$

As normas $\|\cdot\|_p$ sobre o espaço produto $E_1 \times \cdots \times E_n$ constituem para este trabalho o exemplo fundamental de uma família de *normas equivalentes*.

Definição C.3.10. Dizemos que duas normas $\|\cdot\|, \|\cdot\|'$ sobre o mesmo espaço vetorial E são **equivalentes** se $\|x\| = O(\|x\|')$ e $\|x\|' = O(\|x\|)$ quando $x \to 0_E$.

Proposição C.3.11. Duas normas $\|\cdot\|$ e $\|\cdot\|'$ sobre o mesmo espaço vetorial E são equivalentes se, e somente se, existem $a, b > 0$ tais que

$$\begin{cases}
\|x\| \leq a\|x\|', & \text{para todo} \quad x \in E, \\
\|x\|' \leq b\|x\|, & \text{para todo} \quad x \in E.
\end{cases}$$

Demonstração. Suponhamos $\|\cdot\|$ e $\|\cdot\|'$ equivalentes. Então existem $a, b > 0$ e $U \subseteq (E, \|\cdot\|)$ e $V \subseteq (E, \|\cdot\|')$ vizinhanças de 0_E tais que

$$\|x\| \leq a\|x\|' \quad \text{e} \quad \|y\|' \leq b\|y\|, \quad \text{para quaisquer} \quad x \in U, y \in V.$$

Seja $x \in E$ e tome $k_1, k_2 > 0$ tais que $k_1 x \in U$ e $k_2 x \in V$. Temos

$$k_1\|x\| = \|k_1 x\| \leq a\|k_1 x\|' = ak_1\|x\|'$$

e

$$k_2\|x\|' = \|k_2 x\|' \leq b\|k_2 x\| = bk_2\|x\|,$$

donde vem

$$\|x\| \leq a\|x\|' \quad \text{e} \quad \|x\|' \leq b\|x\|,$$

como queríamos demonstrar. A recíproca é imediata. □

Teorema C.3.12. Se $E_1, \ldots, E_n$ são espaços normados e $E := E_1 \times \cdots \times E_n$, então

$$\|(x_1, \ldots, x_n)\|_p = \begin{cases} \left(\sum_i \|x_i\|_{E_i}^p\right)^{\frac{1}{p}}, & \text{se} \quad p \in [1, +\infty), \\ \max_i \|x_i\|_{E_i}, & \text{se} \quad p = \infty, \end{cases}$$

para cada $x = (x_1, \ldots, x_n) \in E$. Além disso, para cada $x \in E$ fixado, a aplicação

$$p \geq 1 \mapsto \|x\|_p$$

é não crescente e todas as normas $\|\cdot\|_p$ sobre E são equivalentes.

Demonstração. Que

$$p \geq 1 \mapsto \|x\|_p$$

é não crescente é a mesma prova de Proposição C.3.8. Além disso, note que

$$\|x\|_\infty = \max \|x_i\|_{E_i} = \left(\max \|x_i\|_{E_i}^p\right)^{\frac{1}{p}} \leq \left(\sum \|x_i\|_{E_i}^p\right)^{\frac{1}{p}} = \|x\|_p.$$

Suponhamos agora $q > p > 1$ e ponhamos $r := \dfrac{q}{p} > 1$. Pela Desigualdade de Hölder (ver Teorema C.3.6), temos

$$\sum \|x_i\|_{E_i}^p \leq \left(\sum \left(\|x_i\|_{E_i}^p\right)^r\right)^{\frac{1}{r}} \left(\sum 1^{\frac{r}{r-1}}\right)^{1-\frac{1}{r}} = \left(\sum \|x_i\|_{E_i}^q\right)^{\frac{p}{q}} n^{1-\frac{p}{q}},$$

donde

$$\|x\|_p \leq n^{\frac{1}{p}-\frac{1}{q}}\|x\|_q.$$

Por fim, supondo agora $\infty \geq q > p = 1$, temos

$$\|x\|_1 \leq n\|x\|_\infty \leq n\|x\|_q.$$

Logo, $\|\cdot\|_p$ e $\|\cdot\|_q$ são equivalentes. □

Teorema C.3.13. Sejam $\|\cdot\|$ e $\|\cdot\|^\times$ duas normas sobre um mesmo espaço vetorial E. Então $\|\cdot\|$ e $\|\cdot\|^\times$ são normas equivalentes se, e somente se, dada uma sequência $(x_n) \subseteq E$, (x_n) converge para $L \in E$ na norma $\|\cdot\| \iff (x_n)$ converge para L na norma $\|\cdot\|^\times$.

Demonstração. Suponhamos primeiro que $\|\cdot\|$ e $\|\cdot\|^\times$ são equivalentes. Então existem $a, b > 0$ tais que

$$\|x\| \leq a\|x\|^\times \quad \text{e} \quad \|x\|^\times \leq b\|x\|, \quad \text{para todo} \quad x \in E.$$

Suponhamos que $x_n \xrightarrow[\|\cdot\|]{n\to\infty} L$. Assim, dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies \|x_n - L\| < \frac{\varepsilon}{b},$$

donde $x_n \xrightarrow[\|\cdot\|^\times]{n\to\infty} L$, já que

$$n \geq n_0 \implies \|x_n - L\|^\times \leq b\|x_n - L\| < b \cdot \frac{\varepsilon}{b} = \varepsilon.$$

Agora suponhamos que $x_n \xrightarrow[\|\cdot\|^\times]{n\to\infty} L$. Assim, dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$n \geq n_0 \implies \|x_n - L\|^\times < \frac{\varepsilon}{a}.$$

Como

$$\|x_n - L\| \leq a\|x_n - L\|^\times < a \cdot \frac{\varepsilon}{a} = \varepsilon,$$

então $x_n \xrightarrow[\|\cdot\|]{n\to\infty} L$.

Reciprocamente, suponhamos que (x_n) converge para $L \in E$ na norma $\|\cdot\| \iff (x_n)$ converge para L na norma $\|\cdot\|^\times$ e consideremos a aplicação identidade $I : (E, \|\cdot\|) \to (E, \|\cdot\|^\times)$. Dada $x_n \to 0$ na norma $\|\cdot\|$, temos, por hipótese, que $I(x_n) = x_n \to 0$ na norma $\|\cdot\|^\times$. Logo, I é contínua em 0. Como I é linear, então, pelo Teorema C.4.1, existe $a > 0$ tal que

$$\|x\|^\times = \|Ix\|^\times \leq a\|x\|, \quad \text{para todo} \quad x \in E.$$

Aplicando o mesmo teorema à aplicação I^{-1} obtemos $b > 0$ tal que

$$\|x\| \leq b\|x\|^\times, \quad \text{para todo} \quad x \in E.$$

Isso prova que as normas são equivalentes. □

C.4 Transformações lineares contínuas

Teorema C.4.1. Se E e F são espaços normados e $T : E \to F$ é linear, então as seguintes afirmações são equivalentes:

1. T é contínua.

2. T é contínua em 0_E.

3. Existe $c > 0$ tal que $\|T(x)\| \leq c$, para todo $x \in E$ com $\|x\| = 1$.

4. Existe $c > 0$ tal que $\|T(x)\| \leq c\|x\|$, para todo $x \in E$.

Demonstração. Que 1) $\implies$ 2) é óbvio.

Suponhamos, por redução ao absurdo, que 2) vale, mas 3) não. Então, para cada $n \in \mathbb{N}$, podemos escolher $x_n \in E$ tal que

$$\|x_n\| = 1 \text{ e } \|Tx_n\| > n.$$

Como T é contínua em 0, tomando $\varepsilon := 1$, existe $\tilde{\delta} > 0$ tal que

$$x \in B(0, \tilde{\delta}) \implies \|Tx\| < 1.$$

Logo, tomando $\delta \in (0, \tilde{\delta})$, temos $\delta x_n \in B[0, \delta] \subseteq B(0, \tilde{\delta})$, donde

$$n < \|Tx_n\| = \frac{1}{\delta}\|T(\delta x_n)\| < \frac{1}{\delta}, \quad \text{para todo } n \in \mathbb{N},$$

o que é absurdo.

Suponhamos 3) e provemos 4). Se $u \neq 0$, então

$$\|T(u)\|_F = \left\| \|u\|_E T\left(\frac{1}{\|u\|_E} u\right) \right\|_F = \|u\|_E \left\| T\left(\frac{1}{\|u\|_E} u\right) \right\|_F \leq c\|u\|_E.$$

Por fim, suponhamos 4) e provemos 1). Sejam $a \in E$ e $(x_n) \subseteq E$ tal que $\lim\limits_{n\to\infty} x_n = a$. Como existe $c > 0$ tal que $\|Tx\|_F \leq c\|x\|_E$, para todo $x \in E$, então

$$0 \leq \|T(x_n - a)\|_F \leq c\|x_n - a\|_E \xrightarrow{n\to\infty} 0,$$

donde, pelo Teorema do Confronto (ver Teorema A.1.10),

$$\lim_{n\to\infty} Tx_n - Ta = \lim_{n\to\infty} (Tx_n - Ta) = \lim_{n\to\infty} T(x_n - a) = 0$$

e, portanto, $\lim\limits_{n\to\infty} Tx_n = Ta$. □

Corolário C.4.2. Se $T : E \to F$ é linear, então T é contínua em $a \in E$ se, e somente se, T é contínua.

Demonstração. Se T é contínua em a, então $\varphi : E \to F$ dada por $\varphi(x) := T(x + a)$ é contínua em 0. O Teorema C.4.1 garante a continuidade de φ e, portanto, a continuidade de T, já que $T(x) = \varphi(x - a)$. □

Exemplo C.4.3. Se E e F são espaços normados, a norma usual do espaço vetorial

$$\mathcal{L}(E,F) := \{T : E \to F : T \text{ é linear e contínua}\}$$

é dada por

$$\|T\| := \sup\{\|Tx\|_F : \|x\|_E = 1\},$$

para cada $T \in \mathcal{L}(E,F)$.

Para o próximo lema, $[k] := \{1, \ldots, k\}$, para todo $k \in \mathbb{N}$.

Lema C.4.4. Se $(x_n^1), \ldots, (x_n^k)$ são sequências de números reais convergentes, então a sequência

$$y_n := \max_{i \in [k]} x_n^i$$

é convergente, com

$$\lim_{n\to\infty} y_n = \max_{i\in[k]} \lim_{n\to\infty} x_n^i.$$

Demonstração. Para $k = 1$ o resultado é obviamente válido.

Suponhamos, por hipótese de indução, que o resultado vale para $k = m$ e consideremos o caso $k = m + 1$. Pondo

$$a_n := \max_{i\in[m]} x_n^i, \text{ para cada } n \in \mathbb{N}, \text{ e } a := \lim a_n$$

e

$$b_n := x_n^{m+1}, \text{ para cada } n \in \mathbb{N}, \text{ e } b := \lim b_n,$$

temos

$$a = \max_{i\in[m]} \lim_{n\to\infty} x_n^i,$$

da hipótese de indução, e

$$\max_{i\in[m+1]} x_n^i = \max\left\{\max_{i\in[m]} x_n^i, x_n^{m+1}\right\} = \frac{a_n + b_n + |a_n - b_n|}{2},$$

donde

$$\max_{i\in[m+1]} x_n^i \xrightarrow{n\to\infty} \frac{a + b + |a - b|}{2} = \max\{a, b\} = \max_{i\in[m+1]} \lim_{n\to\infty} x_n^i,$$

como queríamos demonstrar. □

Teorema C.4.5. Sejam $E_1, \ldots, E_n$ espaços normados e $x^k = (x_1^k, \ldots, x_n^k)$ uma sequência em $E := E_1 \times \cdots \times E_n$. Então,

$$x^k \xrightarrow[\|\cdot\|_\infty]{k\to\infty} y := (y_1, \ldots, y_n) \iff x_i^k \xrightarrow{k\to\infty} y_i, \quad \text{para cada} \quad i = 1, \ldots, n.$$

Demonstração. Começamos supondo que

$$x^k \xrightarrow[\|\cdot\|_\infty]{k\to\infty} y.$$

Como

$$|x_i^k - y_i| \leq \max_{j=1,\ldots,n} |x_j^k - y_j| = \|x^k - y\|_\infty,$$

para cada $k \in \mathbb{N}$, e

$$\|x^k - y\|_\infty \xrightarrow{k\to\infty} 0,$$

então

$$x_i^k \xrightarrow{k\to\infty} y_i,$$

pelo Teorema do Confronto (ver Teorema A.1.10).

A recíproca é consequência direta do Lema C.4.4. □

Teorema C.4.6. Sejam E um espaço normado de dimensão n munido da norma do máximo $\|\cdot\|_\infty$ e $\{v_i\}$ uma base normalizada de E. Então a aplicação

$$T : E \to E_1 \times \cdots \times E_n$$

dada por

$$T\left(\sum_{i=1}^{n} \alpha_i v_i\right) := (\alpha_1 v_1, \ldots, \alpha_n v_n)$$

é uma isometria linear, onde

$$E_i := \operatorname{span}\{v_i\},$$

para cada $i = 1, \ldots, n$.

Demonstração. Vamos mostrar que T é uma imersão isométrica. Dado $x \in E$, existem únicos $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ tais que

$$x = \sum \alpha_i v_i.$$

Nessas condições,

$$\begin{aligned} \|T(x)\|_\infty = \left\|T\left(\sum \alpha_i v_i\right)\right\|_\infty = \|(\alpha_1 v_1, \ldots, \alpha_n v_n)\|_\infty &= \max_i \|\alpha_i v_i\|_\infty \\ &= \max_i |\alpha_i| \\ &= \left\|\sum \alpha_i v_i\right\|_\infty \\ &= \|x\|_\infty. \end{aligned}$$

Logo, T é uma imersão isométrica. Como a aplicação

$$(\alpha_1 v_1, \ldots, \alpha_n v_n) \in E_1 \times \cdots \times E_n \mapsto \sum \alpha_i v_i \in E$$

é obviamente inversa de T, então T é bijetiva e, portanto, T é uma isometria.

Por fim, vamos verificar que que T é linear. Com efeito, dados $x, y \in E$ e $\lambda \in \mathbb{R}$, existem únicos $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_n \in \mathbb{R}$ tais que

$$x = \sum_{i=1}^{n} \alpha_i v_i \text{ e } y = \sum_{i=1}^{n} \beta_i v_i.$$

Daí,

$$\begin{aligned} T(x + \lambda y) = T\left(\sum_{i=1}^{n} (\alpha_i + \lambda\beta_i) v_i\right) &= ((\alpha_1 + \lambda\beta_1)v_1, \ldots, (\alpha_n + \lambda\beta_n)v_n) \\ &= (\alpha_1 v_1, \ldots, \alpha_n v_n) + \lambda(\beta_1 v_1, \ldots, \beta_n v_n) \\ &= T(x) + \lambda T(y). \end{aligned}$$

Logo, T é linear, como queríamos demonstrar. $\square$

Corolário C.4.7. Sejam E um espaço normado e $\{v_i\}$ uma base normalizada de E. Então,

$$x^k := \sum_{i=1}^{n} \alpha_i^k v_i \xrightarrow[\|\cdot\|_\infty]{k\to\infty} \sum_{i=1}^{n} \beta_i v_i \iff (\alpha_i^k)_k \xrightarrow{k\to\infty} \beta_i, \quad \text{para cada} \quad i = 1, \ldots, n.$$

Demonstração. Considere T definida como no Teorema C.4.6 e suponhamos

$$x^k = \sum_{i=1}^{n} \alpha_i^k v_i \xrightarrow[\|\cdot\|_1]{k\to\infty} \sum_{i=1}^{n} \beta_i v_i.$$

Assim,

$$\left(\alpha_1^k v_1, \ldots, \alpha_n^k v_n\right) = T\left(\sum_{i=1}^{n} \alpha_i^k v_i\right) = T\left(x^k\right) \xrightarrow[\|\cdot\|_1]{k\to\infty} T\left(\sum_{i=1}^{n} \beta_i v_i\right) = (\beta_1 v_1, \ldots, \beta_n v_n),$$

donde, pelo Teorema C.4.5, temos

$$\alpha_i^k \xrightarrow{k\to\infty} \beta_i, \quad \text{para cada} \quad i = 1, \ldots, n.$$

Reciprocamente, se

$$\alpha_i^k \xrightarrow{k\to\infty} \beta_i, \quad \text{para cada} \quad i = 1, \ldots, n,$$

então

$$\left(\alpha_1^k v_1, \ldots, \alpha_n^k v_n\right) \xrightarrow[\|\cdot\|_\infty]{k\to\infty} \left(\beta_1 v_1, \ldots, \beta_n v_n\right),$$

donde, pela continuidade de T^{-1}, temos

$$\sum_{i=1}^{n} \alpha_i^k v_i = T^{-1}\left(\alpha_1^k v_1, \ldots, \alpha_n^k v_n\right) \xrightarrow[\|\cdot\|_\infty]{k\to\infty} T^{-1}\left(\beta_1 v_1, \ldots, \beta_n v_n\right) = \sum_{i=1}^{n} \beta_i v_i,$$

como queríamos demonstrar. $\square$

Corolário C.4.8. Sejam E um espaço normado de dimensão n munido da norma do máximo $\|\cdot\|_\infty$ e considere $\{v_i\}$ uma base normalizada de E. Se

$$x^k = \sum_{i=1}^{n} \alpha_i^k v_i$$

é uma sequência limitada, então (x^k) possui uma subsequência convergente.

Demonstração. Considere $j \in \{1, \ldots, n\}$. Como existe $C > 0$ tal que, para todo $k \in \mathbb{N}$,

$$|\alpha_j^k| \leq \max_i |\alpha_i^k| = \left\| \sum_{i=1}^{n} \alpha_i^k v_i \right\|_\infty = \|x^k\|_\infty \leq C,$$

então $(\alpha_j^k)_k$ é uma sequência limitada de números reais, donde, pelo Teorema de Bolzano-Weierstrass (ver Teorema A.3.5), a sequência $(\alpha_j^k)_k$ possui uma subsequência convergente $(\alpha_j^{k_m})_m$. Pelo Corolário C.4.7, $(x^{k_m})_m$ converge. □

Teorema C.4.9 (Equivalência das normas)**.** Se E é um espaço vetorial de dimensão finita, então todas as normas sobre E são equivalentes.

Demonstração. Sejam $\|\cdot\|$ uma norma em E e $\{v_i\}_{i=1}^n$ uma base normalizada, com respeito à norma $\|\cdot\|$, de E. Se

$$\sum_{i=1}^{n} \alpha_i v_i \in E,$$

então

$$\left\| \sum_{i=1}^{n} \alpha_i v_i \right\| \leq \sum_{i=1}^{n} \|\alpha_i v_i\| = \sum_{i=1}^{n} |\alpha_i| \leq \sum_{i=1}^{n} \max_j |\alpha_j| = n \max_i |\alpha_i| = n \left\| \sum_{i=1}^{n} \alpha_i v_i \right\|_\infty .$$

Suponhamos agora, por redução ao absurdo, que $\|\cdot\|$ e $\|\cdot\|_\infty$ não são equivalentes. Daí, para cada $m \in \mathbb{N}$, o conjunto

$$X_m := \{x \in E : m\|x\| < \|x\|_\infty\}$$

é não vazio. Escolhendo um x^m em cada X_m, obtemos uma sequência $(x^m) \subseteq E$ tal que $x^m \in E$ e $m\|x^m\| < \|x^m\|_\infty$. Desse modo,

$$\left\| \frac{x^m}{\|x^m\|_\infty} \right\| = \frac{\|x^m\|}{\|x^m\|_\infty} < \frac{1}{m},$$

para cada $m \in \mathbb{N}$, donde

$$y_m := \frac{x^m}{\|x^m\|_\infty} \xrightarrow{m\to\infty} 0_E,$$

pelo Teorema do Confronto (ver Teorema A.1.10). Como

$$\|y_m\|_\infty = 1, \text{ para cada } m \in \mathbb{N},$$

então (y_m) é limitada na norma $\|\cdot\|_\infty$. Pelo Teorema C.4.8, a sequência (y_m) possui uma subsequência $(y_{m_k})_k$ convergente na norma $\|\cdot\|_\infty$. Daí, existe $y \in E$ tal que

$$y_{m_k} \xrightarrow[\|\cdot\|_\infty]{k\to\infty} y.$$

Como todas as normas são contínuas, temos

$$\|y\|_\infty = \left\|\lim_{k\to\infty} y_{m_k}\right\|_\infty = \lim_{k\to\infty} \|y_{m_k}\|_\infty = \lim_{k\to\infty} 1 = 1.$$

De

$$\|y_{m_k} - y\| \leq n\|y_{m_k} - y\|_\infty \xrightarrow{k\to\infty} 0,$$

vem

$$y_{m_k} \xrightarrow[\|\cdot\|]{k\to\infty} y,$$

pelo Teorema do Confronto, donde $y = 0$, pela unicidade do limite (ver Teorema B.5.4). Por conseguinte,

$$\|y\|_\infty = 0,$$

o que é absurdo. □

Corolário C.4.10 (Teorema de Bolzano-Weierstrass)**.** Toda sequência limitada em um espaço normado de dimensão finita admite uma subsequência convergente.

Corolário C.4.11. Todo espaço normado de dimensão finita é um espaço de Banach.

Corolário C.4.12 (Continuidade de aplicações lineares em dimensão finita)**.** Se E e F são espaços normados, com $\dim E < +\infty$, e $T : E \to F$ é uma aplicação linear, então T é contínua.

Demonstração. Seja $\{v_i\}_{i=1}^n$ uma base de E e considere $(x^k) \subseteq E$ tal que

$$x^k \xrightarrow[\|\cdot\|_E]{k\to\infty} 0_E.$$

Para cada $k \in \mathbb{N}$, existem únicos $\alpha_1^k, \ldots, \alpha_n^k \in \mathbb{R}$ tais que

$$x^k = \sum_{i=1}^n \alpha_i^k v_i.$$

Pelo Teorema C.4.7, temos

$$\alpha_i^k \xrightarrow{k\to\infty} 0, \quad \text{para cada} \quad i = 1, \ldots, n,$$

donde

$$T\left(x^k\right) = T\left(\sum_{i=1}^n \alpha_i^k v_i\right) = \sum_{i=1}^n \alpha_i^k T v_i \xrightarrow[\|\cdot\|_1]{k\to\infty} 0_F.$$

O caso com as outras normas se segue do Teorema C.4.9 e do Teorema C.3.13. □

C.5 Produto interno

Definição C.5.1. Um **produto interno** em E é uma função $\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$ tal que

1. $\langle u + \lambda v, w \rangle = \langle u, w \rangle + \lambda \cdot \langle v, w \rangle$, para quaisquer $u, v, w \in E$;
2. $\langle u, v \rangle = \langle v, u \rangle$, para quaisquer $u, v \in E$;
3. Se $0 \neq u \in E$, então $\langle u, u \rangle > 0$.

Isto é, um produto interno é uma forma bilinear simétrica e definida positiva.

Teorema C.5.2 (Desigualdade de Cauchy-Schwarz)**.** Seja E um espaço munido de um produto interno. Então, para quaisquer $u, v \in E$, temos

$$|\langle u, v \rangle| \leq \sqrt{\langle u, u \rangle \langle v, v \rangle}.$$

Demonstração. Sejam $u, v \in E$. Faça

$$f(\lambda) := \langle u - \lambda v, u - \lambda v \rangle = \langle v, v \rangle \lambda^2 - 2\langle u, v \rangle \lambda + \langle u, u \rangle.$$

Como $f \geq 0$, então

$$\Delta := 4\langle u, v \rangle^2 - 4\langle v, v \rangle \langle u, u \rangle \leq 0,$$

o que nos dá

$$\langle u, v \rangle^2 \leq \langle u, u \rangle \langle v, v \rangle,$$

como queríamos demonstrar. □

Corolário C.5.3. Se E é um espaço vetorial com produto interno $\langle \cdot, \cdot \rangle$, então

$$\|u\| := \sqrt{\langle u, u \rangle}$$

é uma norma em E (chamada a **norma induzida** pelo produto interno).

Teorema C.5.4 (Continuidade do produto interno)**.** Se E é um espaço vetorial e $\langle \cdot, \cdot \rangle$ é um produto interno em E, então $\langle \cdot, \cdot \rangle$ é uma função contínua.

Demonstração. Sejam $a, b \in E$ e $(x_n), (y_n) \subseteq E$ tais que $\lim\limits_{n \to \infty} x_n = a$ e $\lim\limits_{n \to \infty} y_n = b$. Como

$$\begin{aligned} 0 \leq |\langle x_n, y_n \rangle - \langle a, b \rangle| &= |\langle x_n, y_n \rangle - \langle a, y_n \rangle + \langle a, y_n \rangle - \langle a, b \rangle| \\ &= |\langle x_n - a, y_n \rangle + \langle a, y_n - b \rangle| \\ &\leq |\langle x_n - a, y_n \rangle| + |\langle a, y_n - b \rangle| \\ &\leq \|x_n - a\| \|y_n\| + \|a\| \|y_n - b\| \xrightarrow{n \to \infty} 0, \end{aligned}$$

pela Desigualdade de Cauchy-Schwarz (ver Teorema C.5.2) e pela Desigualdade Triangular, donde, pelo Teorema do Confronto (ver Teorema A.1.10),

$$\lim_{n\to\infty} |\langle x_n, y_n\rangle - \langle a, b\rangle| = 0,$$

isto é,

$$\lim_{n\to\infty} \langle x_n, y_n\rangle = \langle a, b\rangle.$$

Portanto, $\langle \cdot, \cdot \rangle$ é uma função contínua. □

A proposição abaixo é um fato útil.

Proposição C.5.5. Se E é um espaço com produto interno e $u, v \in E$ são tais que

$$\langle u, w\rangle = \langle v, w\rangle, \quad \text{para todo} \quad w \in E,$$

então $u = v$.

Demonstração. Como

$$\langle u - v, w\rangle = 0, \quad \text{para todo} \quad w \in E,$$

então, em particular, para $w := u - v$, temos

$$\|u - v\|^2 = \langle u - v, u - v\rangle = 0$$

e, portanto, $u = v$, como queríamos demonstrar. □

C.6 Autovalores, autovetores e operadores autoadjuntos

Começamos as definições dessa seção com (LIMA, 2018, p. 146).

Um vetor $v \neq 0$ em E chama-se um **autovetor** do operador $A : E \to E$ quando existe $\lambda \in \mathbb{R}$ tal que

$$Av = \lambda v.$$

O número $\lambda \in \mathbb{R}$, por sua vez, chama-se **autovalor** do operador A quando existe um vetor não nulo $v \in E$ tal que $Av = \lambda v$. Diz-se então que o autovalor λ corresponde, ou pertence, ao autovetor v e, vice-versa, que o autovetor v também corresponde, ou pertence, ao autovalor λ.

Definição C.6.1. Dizemos que uma matriz quadrada A é **simétrica** se $A^t = A$.

Teorema C.6.2 (Propriedades básicas da transposição)**.**

- Se $x, y, z \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, então $(x + \lambda y)^t z = x^t z + \lambda y^t z$.

- Se $x, y \in \mathbb{R}^n$, então $x^t y = y^t x$.
- Se $x \in \mathbb{R}^n \setminus \{0\}$, então $x^t x > 0$.
- Se $A \in \mathcal{M}(m \times n)$ e $B \in \mathcal{M}(n \times p)$, então $(AB)^t = B^t A^t$.

Todas as propriedades elencadas acima se verificam por um cálculo direto.

Corolário C.6.3. Se A é uma matriz quadrada de ordem n, então

$$A \text{ é uma matriz simétrica} \iff x^t Ay = y^t Ax, \text{ para quaisquer } x, y \in \mathbb{R}^n.$$

Demonstração. Se A é uma matriz simétrica, então

$$x^t Ay = x^t A^t y = (Ax)^t y = y^t Ax.$$

Reciprocamente, seja a_{ij} o j-ésimo termo da i-ésima linha de A. Como

$$a_{ij} = e_i^t A e_j = e_j^t A e_i = a_{ji},$$

então $A = A^t$, isto é, A é uma matriz simétrica. □

Definição C.6.4 (Operador autoadjunto)**.** Se E é um espaço com produto interno, dizemos que um operador linear $T : E \to E$ é **autoadjunto** se a forma bilinear $A : E \times E \to \mathbb{R}$ dada por

$$A(x, y) := \langle x, Ty \rangle$$

é simétrica, isto é,

$$\langle x, Ty \rangle = \langle y, Tx \rangle, \quad \text{para quaisquer} \quad x, y \in E.$$

Teorema C.6.5. Se E é um espaço vetorial de dimensão finita com produto interno, então um operador linear $T : E \to E$ é autoadjunto se, e somente se, a matriz $[T]$ relativa a uma base ortonormal de E é uma matriz simétrica.

Demonstração. Seja $\{e_1, \ldots, e_n\}$ uma base ortonormal de E e suponhamos T autoadjunto. Dados $u = (u_i)_{i=1}^n, v = (v_i)_{i=1}^n \in \mathbb{R}^n$, tome $x = \sum u_i e_i$ e $y = \sum v_j e_j$. Então, $x, y \in E$ e temos

$$u^t[T]v = [x]^t[T][y] = [x]^t[Ty] = \langle x, Ty \rangle = \langle y, Tx \rangle = [y]^t[Tx] = [y]^t[T][x] = v^t[T]u.$$

Logo, $[T]$ é uma matriz simétrica, pelo Corolário C.6.3.

Reciprocamente, suponhamos $[T]$ uma matriz simétrica. Dados $x = \sum x_i e_i, y = \sum y_j e_j \in E$, faça $u = (x_i)_i$ e $v = (y_j)_j$. Então $u, v \in \mathbb{R}^n$ e, novamente pelo Corolário C.6.3, temos

$$\langle x, Ty \rangle = [x]^t[Ty] = [x]^t[T][y] = u^t[T]v = v^t[T]u = [y]^t[T][x] = [y]^t[Tx] = \langle y, Tx \rangle.$$

Logo, T é autoadjunto. □

Teorema C.6.6. (LIMA, 2018, Teorema Espectral, p. 160) Para todo operador autoadjunto $A : E \to E$, num espaço vetorial de dimensão finita munido de um produto interno, existe uma base ortonormal $\{u_1, \ldots, u_n\} \subseteq E$ formada por autovetores de A.

C.7 Transformações multilineares

Sejam $E_1, \ldots, E_n$ e F espaços vetoriais. Uma aplicação

$$T : E_1 \times \cdots \times E_n \to F$$

é n-**linear** se, para quaisquer $i = 1, \ldots, n$ e $v_j \in E_j$, com $j \in \{1, \ldots, n\} \setminus \{i\}$, a aplicação

$$u_i \in E_i \mapsto T(v_1, \ldots, v_{i-1}, u_i, v_{i+1}, \ldots, v_n) \in F$$

é linear.

A aplicação T também é chamada uma **transformação multilinear** e, quando $F = \mathbb{R}$, dizemos que T é uma **forma multilinear**.

Dado $x \in E$, defina

$$x^{(n)} := (\underbrace{x, \ldots, x}_{n \text{ vezes}}).$$

Se E e F são espaços vetoriais e $T : E^n \to F$ é uma aplicação n-linear,

$$Tx^{(n)} := T \cdot x^{(n)} := T\left(x^{(n)}\right), \text{ para todo } x \in E.$$

Além disso, se $T : E_1 \times \cdots \times E_n \to F$ é uma aplicação multilinear,

$$Tu_1 \ldots u_n := T(u_1, \ldots, u_n), \text{ para todo } u_i \in E_i, \text{ com } i = 1, \ldots, n.$$

Quando $E_1, \ldots, E_n$ são espaços normados, ao tratarmos do espaço $E_1 \times \cdots \times E_n$, a não ser quando mencionarmos o contrário, sempre o consideraremos munido da norma $\|\cdot\|_\infty$ (ver Exemplo C.3.9) ou de uma norma que lhe é equivalente (ver Teorema C.3.12).

Teorema C.7.1. Se $E_1, \ldots, E_n$ e F são espaços normados e $T : E_1 \times \cdots \times E_n \to F$ é uma aplicação n-linear, então as seguintes afirmações são equivalentes:

1. T é contínua.

2. T é contínua em $(0_{E_1}, \ldots, 0_{E_n}) \in E_1 \times \cdots \times E_n$.

3. Existe $c > 0$ tal que
$$\|T(u_1, \ldots, u_n)\|_F \leq c,$$
quaisquer que sejam $u_i \in E_i$ com $\|u_i\|_{E_i} = 1$ e $i = 1, \ldots, n$.

4. Existe $c > 0$ tal que
$$\|T(u_1, \ldots, u_n)\|_F \leq c \prod \|u_i\|_{E_i},$$
para todo $u_i \in E_i$, com $i = 1, \ldots, n$.

Demonstração. Idêntica à demonstração do Teorema C.4.1. □

Seguindo (LIMA, 2002, p. 3), se $E_1, \ldots, E_n$ e F são espaços normados, podemos munir naturalmente o espaço vetorial

$$\mathcal{L}(E_1, \ldots, E_n; F) := \{T : E_1 \times \cdots \times E_n \to F : T \text{ é } n\text{-linear e contínua}\}$$

de uma norma pondo, para cada $T \in \mathcal{L}(E_1, \ldots, E_n; F)$,

$$\|T\|_{\mathcal{L}(E_1,\ldots,E_n;F)} := \sup_{\substack{x_i \in E_i \text{ e } \|x_i\|_{E_i}=1, \\ \text{para todo } i=1,\ldots,n}} \|T(x_1, \ldots, x_n)\|_F.$$

Exemplo C.7.2. A aplicação

$$\langle \cdot, \cdot \rangle_2 : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

dada por

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle_2 := \sum_i x_i y_i$$

é uma forma bilinear.

Exemplo C.7.3. O produto interno real é uma forma bilinear.

A existência do determinante, com as propriedades que conhecemos, tem uma construção longa, que não faremos aqui. Os detalhes da construção podem ser consultados na referência indicada.

Teorema C.7.4. (BUENO, 2006, pp. 57-64) Existe uma única função

$$\det : \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ vezes}} \to \mathbb{R}$$

que é n-linear, alternada[1] e tal que

$$\det(e_1, \ldots, e_n) = 1,$$

onde $\{e_i\}$ é a base canônica de $\mathbb{R}^n$.

[1] Dizemos que uma forma n-linear T é **alternada** se $T(c_1, \ldots, c_n) = 0$ sempre que $c_i = c_j$, para $i \neq j$.

Dados $u, v \in \mathbb{R}^3$, como $\det(u, v, \cdot) : \mathbb{R}^3 \to \mathbb{R}$ é um funcional linear contínuo (pois $\mathbb{R}^3$ é um espaço de dimensão finita) e a função

$$\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_2$$

definida no Exemplo C.7.2 é um produto interno em $\mathbb{R}^3$, então, pelo Teorema da Representação de Riesz (ver Teorema C.9.4), existe um único vetor $u \wedge v \in \mathbb{R}^3$ tal que

$$\det(u, v, w) = \langle u \wedge v, w \rangle, \text{ para todo } w \in \mathbb{R}^3 .$$

O vetor $u \wedge v$ é o **produto vetorial** de u e v.

Exemplo C.7.5. O produto vetorial

$$\wedge : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$$

dado por

$$\wedge(u, v) := u \wedge v,$$

para quaisquer $u, v \in \mathbb{R}^3$, é uma aplicação bilinear contínua.

Demonstração. Sejam $w, x, y \in \mathbb{R}^3$ e $\alpha \in \mathbb{R}$. Por um cálculo direto, para cada $z \in \mathbb{R}^3$,

$$\begin{aligned}
\langle w \wedge (x + \alpha y), z \rangle &= \det(w, x + \alpha y, z) \\
&= \det(w, x, z) + \alpha \det(w, y, z) \\
&= \langle w \wedge x, z \rangle + \alpha \langle w \wedge y, z \rangle \\
&= \langle w \wedge x + \alpha (w \wedge y), z \rangle.
\end{aligned}$$

Logo, pela Proposição C.5.5,

$$w \wedge (x + \alpha y) = w \wedge x + \alpha (w \wedge y).$$

Isto prova que o produto vetorial é linear na segunda entrada. Um cálculo análogo mostra que $\wedge$ é linear na primeira entrada também.

A continuidade de $\wedge$ vem do fato de que $\mathbb{R}^3$ é um espaço de dimensão finita. $\square$

Um caso particularmente importante de transformações multilineares é o das transformações *simétricas*.

Definição C.7.6. (COLEMAN, 2012, p. 88) Sejam E e F espaços vetoriais e $T : E^n \to F$ uma aplicação n-linear. Dizemos que T é **simétrica** quando, para quaisquer $h_1, \ldots, h_n \in E$, temos

$$T(h_1, \ldots, h_n) = T\left(h_{\sigma(1)}, \ldots, h_{\sigma(n)}\right),$$

para toda permutação $\sigma \in S_n$.

Exemplo C.7.7. A forma bilinear definida no Exemplo C.7.2 é simétrica.

C.8 Espaços de Banach e espaços de Hilbert

Este capítulo se baseia livrevemente em (BOTELHO; PELLEGRINO; TEIXEIRA, 2015; LIMA, 2018; BUENO, 2006; KREYSZIG, 1989).

Não precisamos, pelo menos para a maior parte da teoria básica do cálculo diferencial, falar sobre espaços de Banach ou espaços de Hilbert. Espaços normados são, em geral, o ambiente suficiente. Esta seção existe, no entanto, porque um ou dois teoremas importantes (ver Teorema 4.1.2 e Teorema 4.2.1) requerem que o espaço normado seja completo (espaços de Banach) e alguns conceitos, como a definição geral de vetor gradiente, requerem a noção de um produto interno completo (espaços de Hilbert), definição essa que é dada via o Teorema da Representação de Riesz (ver Teorema C.9.4).

Definição C.8.1 (Espaços de Banach e espaços de Hilbert)**.** Seja E um espaço vetorial real.

- Dizemos que um espaço normado $(E, \|\cdot\|)$ é um **espaço de Banach** se E é um espaço métrico completo com a métrica induzida pela norma.
- Dizemos que E munido de um produto interno $\langle\cdot,\cdot\rangle$ é um **espaço de Hilbert** se E é um espaço de Banach com a norma induzida pelo produto interno.

Proposição C.8.2. Seja E um espaço de Banach. Um subespaço vetorial de E é um espaço de Banach, com a norma induzida de E, se, e somente se, é fechado em E.

Demonstração. Seja F um subespaço vetorial de E.

$[\Longrightarrow]$ Seja $(x_n) \subseteq F$ uma sequência convergente em E. Assim, existe $x \in E$ tal que $x_n \xrightarrow{n\to\infty} x$. Como (x_n) converge, então (x_n) é de Cauchy em F. Como F é espaço de Banach, então existe $y \in F$ tal que $x_n \xrightarrow{n\to\infty} y$. Pela unicidade do limite, $x = y \implies x \in F$.

$[\Longleftarrow]$ Seja $(x_n) \subseteq F$ uma sequência de Cauchy. Como $F \subseteq E$ e E é espaço de Banach, então existe $x \in E$ tal que $x_n \xrightarrow{n\to\infty} x$. Logo, $x \in \overline{F} = F$. □

Exemplo C.8.3. Todo espaço normado de dimensão finita é um espaço de Banach.

Demonstração. Se E é um espaço normado de dimensão finita, como todas as normas em espaços vetoriais de dimensão finita são equivalentes, a norma de E é equivalente à norma $\|\cdot\|_1$, o que torna E um espaço de Banach. □

Exemplo C.8.4. Se E é um espaço de dimensão finita, qualquer produto interno em E o torna um espaço de Hilbert.

Demonstração. Como todo espaço normado de dimensão finita é Banach, qualquer norma induzida por um produto interno em E o torna um espaço de Banach. □

Exemplo C.8.5. Sejam $V_1, \dots, V_d$ espaços de Banach e $V := \prod_i V_i$. Então, V é um espaço de Banach com a norma

$$\|v\|_p := \left(\sum_i \|v_i\|^p\right)^{1/p},$$

para todo $p \in [1, +\infty)$.

Demonstração. Seja $(v_n) \subseteq V$ uma sequência de Cauchy. Note que

$$\|v_m[i] - v_n[i]\|^p \leq \sum_j \|v_m[j] - v_n[j]\|^p,$$

donde

$$\|v_m[i] - v_n[i]\| \leq \|v_m - v_n\|_p.$$

Dado $\varepsilon > 0$, como (v_n) é de Cauchy, existe $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies \|v_m[i] - v_n[i]\| \leq \|v_m - v_n\|_p < \varepsilon.$$

Logo, $(v_n[i])$ é de Cauchy em V_i e, portanto, convergente (já que V_i é Banach), com

$$v[i] := \lim_n v_n[i] \in V_i.$$

Como

$$\lim_n \sum_i \|v_n[i] - v[i]\|^p = \sum_i \lim_n \|v_n[i] - v[i]\|^p = 0,$$

temos

$$\|v_n - v\|_p = \left(\sum_i \|v_n[i] - v[i]\|^p\right)^{1/p} \xrightarrow{n\to\infty} 0,$$

assegurando a convergência de (v_n). □

Exemplo C.8.6. Dado $p \in [1, +\infty]$, o espaço $\ell^p(\mathbb{N})$ (ver Exemplo C.1.4) com a norma

$$\|(x_n)\|_p := \left(\sum |x_n|^p\right)^{1/p} \quad \text{e} \quad \|(x_n)\|_\infty := \sup_{n\in\mathbb{N}} |x_n|$$

é um espaço de Banach.

Demonstração. Seja $(f_n) \subseteq \ell^p$ uma sequência de Cauchy. Suponhamos primeiro $p = +\infty$. Dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que $\|f_n - f_m\|_\infty < \varepsilon$. Daí,

$$|f_n(j) - f_m(j)| \leq \|f_n - f_m\|_\infty < \varepsilon, \quad \text{para quaisquer} \quad j \in \mathbb{N}, \quad m, n \geq n_0.$$

Logo $(f_n(j))$ é uma sequência de Cauchy em $\mathbb{R}$, para todo j. Definimos

$$f(j) := \lim_{n\to\infty} f_n(j).$$

Fixe $\varepsilon > 0$ e tome $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies \|f_n - f_m\|_\infty < \varepsilon/2.$$

Tome $n \geq n_0$. Temos, para todo $j \in \mathbb{N}$,

$$|f_n(j) - f(j)| = \lim_{m\to\infty} |f_n(j) - f_m(j)| \leq \limsup_{m\to\infty} \|f_n - f_m\|_\infty \leq \varepsilon/2.$$

Logo,

$$\|f_n - f\|_\infty = \sup_{j\in\mathbb{N}} |f_n(j) - f(j)| \leq \varepsilon/2 < \varepsilon.$$

Agora

$$\|f\|_\infty \leq \|f - f_{n_0}\|_\infty + \|f_{n_0}\|_\infty < +\infty,$$

o que termina a prova.

Suponhamos agora $p \in [1, +\infty)$. Então, tomando $\varepsilon = 1$ na definição de sequência de Cauchy, obtemos $n_1 \in \mathbb{N}$ tal que

$$m, n \geq n_1 \implies \|f_m - f_n\|_p < 1/2.$$

Agora, tomando $\varepsilon = 1/2$, obtemos $n_2 \geq n_1 + 1$ tal que

$$m, n \geq n_2 \implies \|f_m - f_n\|_p < 1/4.$$

Tomando $\varepsilon = 1/4$, obtemos $n_3 \geq n_2 + 1$ tal que

$$m, n \geq n_3 \implies \|f_m - f_n\|_p < 1/8.$$

Continuando esse processo obtemos uma subsequência (f_{n_k}) com a propriedade de que

$$\|f_{n_{k+1}} - f_{n_k}\|_p < 1/2^k.$$

Definamos

$$f(j) := f_{n_1}(j) + \sum_{k=1}^{\infty} (f_{n_{k+1}}(j) - f_{n_k}(j)), \quad \text{para todo} \quad j \in \mathbb{N}\,.$$

Para isso estar bem definido, precisamos verificar que a série que aparece na definição de $f(j)$ acima converge. Defina

$$g_N(j) := \sum_{k=1}^{N} |f_{n_{k+1}}(j) - f_{n_k}(j)| \quad \text{e} \quad g(j) := \lim_{N\to\infty} g_N(j).$$

(Veja que esse limite $g(j)$ faz sentido na reta estendida, já que $g_N(j)$ é uma sequência não decrescente; ver Teorema A.2.6.) Como $\|\cdot\|_p$ é uma norma, temos, para cada $N \in \mathbb{N}$,

$$\|g_N\|_p = \left\| \sum_{k=1}^{N} |f_{n_{k+1}} - f_{n_k}| \right\|_p \leq \sum_{k=1}^{N} \|f_{n_{k+1}} - f_{n_k}\|_p < \sum_{k=1}^{N} (1/2^k) < \sum_{k=1}^{\infty} (1/2^k) = 1,$$

pelo Exemplo A.4.3. Pelo Lema de Fatou (ver Teorema A.4.9),

$$\|g\|_p^p = \sum_j g(j)^p = \sum_j \lim_{N\to\infty} g_N(j)^p \leq \liminf_{N\to\infty} \sum_j g_N(j)^p = \liminf_{N\to\infty} \|g_N\|_p^p \leq 1.$$

Em particular, $\|g\|_p \leq 1$ e cada $g(j)$ é finito. Logo, a série

$$\sum_{k=1}^{\infty}(f_{n_{k+1}}(j) - f_{n_k}(j))$$

converge absolutamente para cada $j \in \mathbb{N}$ e, portanto, é convergente (ver Teorema A.4.8). Observe agora que

$$f_{n_N} = f_{n_1} + \sum_{k=1}^{N-1}(f_{n_{k+1}} - f_{n_k}),$$

donde vem

$$\lim_{N\to\infty} f_{n_N}(j) = f(j).$$

Como (f_n) é de Cauchy, dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies \|f_m - f_n\|_p < \varepsilon/2.$$

Daí, tomando $n \geq n_0$ e aplicando Fatou novamente, temos

$$\|f - f_n\|_p^p = \sum_j |f(j) - f_n(j)|^p \leq \liminf_{N\to\infty} \sum_j |f_{n_N}(j) - f_n(j)|^p = \liminf_{N\to\infty} \|f_{n_N} - f_n\|_p^p \leq (\varepsilon/2)^p,$$

o que nos dá

$$\|f - f_n\|_p \leq \varepsilon/2 < \varepsilon.$$

Em particular, isso prova que

$$\|f\|_p = \|f - f_{n_0}\|_p + \|f_{n_0}\|_p < +\infty$$

(isto é, $f \in \ell^p$) e $\|f - f_n\|_p \to 0$, o que termina a prova. □

Exemplo C.8.7. O espaço $\ell^2(\mathbb{N})$ com o produto interno

$$\langle (x_n), (y_n) \rangle := \sum_{n=1}^{\infty} x_n y_n$$

é um espaço de Hilbert.

Demonstração. Veja primeiro que esse produto interno está bem definido: isso é consequência da Desigualdade de Cauchy-Schwarz (ver Teorema C.5.2). Com efeito, dadas $(x_n), (y_n) \in \ell^2$, temos

$$|\langle (x_n), (y_n) \rangle| \leq \sqrt{\sum_n |x_n|^2 \sum_n |y_n|^2} = \|(x_n)\|_2 \|(y_n)\|_2 < +\infty.$$

(As outras propriedades que definem um produto interno são facilmente verificáveis.) Agora, seja $\|\cdot\|$ a norma induzida pelo produto interno. Então

$$\|(x_n)\| := \sqrt{\langle (x_n), (x_n) \rangle} = \sqrt{\sum_{n=1}^{\infty} |x_n|^2} = \|(x_n)\|_2,$$

isto é, $\|\cdot\| = \|\cdot\|_2$. Pelo Exemplo C.8.6, ℓ^2 com a norma $\|\cdot\|_2$ é um espaço de Banach. □

Provaremos agora algumas propriedades importantes de espaços de Banach.

Teorema C.8.8. Seja E um espaço normado. Então, E é um espaço de Banach se, e somente se, toda série $\sum v_n$ absolutamente convergente converge.

Demonstração. Suponhamos E Banach e seja $\sum v_k$ uma série absolutamente convergente. Assim, $\sum \|v_k\| < \infty$, o que equivale a dizer que $s_n := \sum_{k=1}^{n} \|v_k\|$ é uma sequência de Cauchy. Logo, fixado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que, para $m > n \geq n_0$, temos

$$\varepsilon \geq |s_m - s_n| = \left| \sum_{k=1}^{m} \|v_k\| - \sum_{k=1}^{n} \|v_k\| \right| = \sum_{k=n+1}^{m} \|v_k\|.$$

Isso nos dá

$$\left\| \sum_{k=1}^{m} v_k - \sum_{k=1}^{n} v_k \right\| = \left\| \sum_{k=n+1}^{m} v_k \right\| \leq \sum_{k=n+1}^{m} \|v_k\| \leq \varepsilon.$$

Logo, $t_n := \sum_{k=1}^{n} v_k$ é de Cauchy em E. Como E é Banach, então (t_n) converge.

Reciprocamente, suponhamos que toda série absolutamente convergente converge e seja $(x_n) \subseteq E$ uma sequência de Cauchy. Escolha uma subsequência (x_{n_k}) tal que

$$\|x_{n_{k+1}} - x_{n_k}\| \leq \frac{1}{2^k}.$$

(Isso é possível tomando $\varepsilon = 1/2$ e o n_1 correspondente na definição de (x_n) ser de Cauchy; depois tomando $\varepsilon = 1/4$ e escolhendo um n_2 correspondente de tal modo que $n_2 > n_1$; e assim sucessivamente.) Faça agora $v_k := x_{n_{k+1}} - x_{n_k}$. Temos

$$\sum \|v_k\| - \sum \|x_{n_{k+1}} - x_{n_k}\| \leq \sum (1/2^k) = 1.$$

Então $\sum v_k$ é absolutamente convergente, logo converge, digamos, para $L \in E$. Observe que

$$x_{n_{p+1}} = \sum_{k=1}^{p} v_k - x_{n_1} \xrightarrow{p \to \infty} L - x_{n_1}.$$

Assim, (x_{n_p}) é uma subsequência convergente da sequência de Cauchy (x_n), donde, por um argumento análogo ao do Teorema A.2.2, a sequência (x_n) converge. □

Teorema C.8.9. Sejam E um espaço normado e F um espaço de Banach. Então $\mathcal{L}(E, F)$ é um espaço de Banach.

Demonstração. Seja $(T_n) \subseteq \mathcal{L}(E, F)$ uma sequência de Cauchy. Dado $x \in E$, temos

$$\|T_m(x) - T_n(x)\| \leq \|T_m - T_n\| \|x\| \xrightarrow{m,n \to \infty} 0.$$

Logo, $(T_n(x)) \subseteq F$ é uma sequência de Cauchy, para todo $x \in E$. Como F é Banach, para cada $x \in E$, existe

$$T(x) := \lim_{n \to \infty} T_n(x) \in F.$$

Afirmamos que T é (i) linear e (ii) contínua. Vamos começar provando (i). Dados $u, v \in E$ e $\alpha \in \mathbb{R}$, temos

$$T(u+\alpha v) = \lim_{n\to\infty} T_n(u+\alpha v) = \lim_{n\to\infty}(T_n(u)+\alpha T_n(v)) = T(u)+\alpha T(v).$$

Para provar (ii), seja $x \in E$. Como (T_n) é de Cauchy, então

$$|\|T_m\| - \|T_n\|| \leq \|T_m - T_n\| \xrightarrow{m,n\to\infty} 0.$$

Logo, $(\|T_n\|)$ é de Cauchy e, portanto, é limitada, pelo Teorema A.1.8. Isto é, existe $C > 0$ tal que $\|T_n\| \leq C$, para todo $n \in \mathbb{N}$. Isso nos dá

$$\|T(x)\| = \left\|\lim_{n\to\infty} T_n(x)\right\| = \lim_{n\to\infty}\|T_n(x)\| = \limsup_{n\to\infty}\|T_n(x)\| \leq \limsup_{n\to\infty}(\|T_n\|\|x\|) \leq C\|x\|.$$

Pelo Teorema C.4.1, isso prova que T é contínua.

Agora vamos provar que $\|T_n - T\| \xrightarrow{n\to\infty} 0$. Seja $\varepsilon > 0$. Como (T_n) é de Cauchy, existe $n_0 \in \mathbb{N}$ tal que

$$m, n \geq n_0 \implies \|T_m - T_n\| \leq \varepsilon.$$

Fixe $n \geq n_0$. Veja que

$$\|T_n - T\| = \sup_{\|x\|=1}\|T_n(x) - T(x)\|.$$

Agora, se $x \in E$ com $\|x\| = 1$. Temos:

$$\|T_n(x) - T(x)\| = \|T_n(x) - \lim_{m\to\infty} T_m(x)\| = \lim_{m\to\infty}\|T_n(x) - T_m(x)\| \leq \lim_{m\to\infty}\|T_n - T_m\| \leq \varepsilon,$$

o que nos dá

$$\|T_n - T\| \leq \varepsilon,$$

como queríamos demonstrar. □

Lema C.8.10. Se E é um espaço de Banach e $B \in \mathcal{L}(E)$ tem norma $\|B\| < 1$, então $I - B$ é invertível, onde $I : E \to E$ é a identidade.

Demonstração. Observe que

$$\sum \|B^k\| \leq \sum \|B\|^k,$$

onde a série da direita é uma série geométrica, logo convergente. Isso prova que $\sum \|B^k\|$ é convergente, pelo Teste da Comparação. Como E é um espaço de Banach, isso implica que $\sum B^k$ converge para um operador $S \in \mathcal{L}(E)$. (Veja Teorema C.8.8.) Vamos mostrar que $S = (I - B)^{-1}$.

Com efeito,

$$\sum_{k=0}^{n} B^k = I + B + \cdots + B^n$$

e

$$B\sum_{k=0}^{n} B^k = B + B^2 + \cdots + B^{n+1},$$

o que nos dá

$$(I - B)\sum_{k=0}^{n} B^k = I - B^{n+1}. \tag{C.1}$$

Observe que $\|B^{n+1}\| \leq \|B\|^{n+1} \xrightarrow{n\to\infty} 0$, o que implica que $B^{n+1} \to 0$. Logo, como a multiplicação de operadores é contínua, passando o limite com $n \to \infty$ em (C.1), obtemos

$$(I - B)S = I - 0 = I.$$

A prova de que $S(I - B) = I$ é análoga. $\square$

Teorema C.8.11. Sejam E e F espaços de Banach e $\mathcal{I}(E, F)$ o conjunto de todos os $T \in \mathcal{L}(E, F)$ invertíveis. Então $\mathcal{I}(E, F)$ é um subconjunto aberto de $\mathcal{L}(E, F)$.

Demonstração. Sejam $T \in \mathcal{I}(E, F)$ e $r := 1/\|T^{-1}\|$ e tome $S \in B(T, r)$. Então

$$S = T + (S - T) = T(I + B), \quad \text{onde} \quad B := T^{-1}(S - T).$$

Como T é invertível, se $I + B$ for invertível, então S também será. Como

$$\|-B\| = \|T^{-1}(S - T)\| \leq \|T^{-1}\|\|S - T\| < \|T^{-1}\|r = 1,$$

então, pelo Lema C.8.10, $I + B$ é invertível. $\square$

C.9 Teorema da Representação de Riesz

O teorema abaixo, que se encontra em (KREYSZIG, 1989, p. 144), servirá de apoio para a demonstração do Teorema da Representação de Riesz.

Teorema C.9.1 (Vetor minimizante e ortogonalidade)**.** Se X é um espaço vetorial com produto interno, $M \subseteq X$ é um conjunto completo, convexo e não vazio e $x \in X$, então existe um único $y \in M$ tal que

$$\inf_{\tilde{y} \in M} \|x - \tilde{y}\| = \|x - y\|.$$

Se, além disso, M é um subespaço vetorial de X, então $z := x - y$ é ortogonal a M.

Demonstração. **Existência.** Como M é não vazio e 0 é cota inferior de

$$A := \{\|x - \tilde{y}\| : \tilde{y} \in M\},$$

então A é não vazio e limitado inferiormente, donde existe

$$\alpha := \inf A.$$

Daí, para cada $n \in \mathbb{N}$, existe $a_n \in A$ tal que

$$\alpha \leq a_n < \alpha + \frac{1}{n}.$$

Isto é, para cada $n \in \mathbb{N}$, existe $y_n \in M$ tal que

$$\alpha \leq \|x - y_n\| < \alpha + \frac{1}{n}.$$

Pelo Teorema do Confronto (ver Teorema A.1.10),

$$\lim_{n \to \infty} \|x - y_n\| = \alpha.$$

Note que, pela identidade do paralelogramo, temos

$$\|(y_n - x) + (y_m - x)\|^2 + \|y_m - y_n\|^2 = 2(\|y_n - x\|^2 + \|y_m - x\|^2),$$

para quaisquer $m, n \in \mathbb{N}$. Da convexidade de M vem

$$\frac{1}{2}(y_n + y_m) - x \in A \implies \|(y_n - x) + (y_m - x)\|^2 = 4\left\|\frac{1}{2}(y_n + y_m) - x\right\|^2 \geq 4\alpha^2.$$

Como

$$\lim_{k \to \infty} \|y_k - x\|^2 = \alpha^2,$$

então, dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que

$$\begin{aligned} m, n \geq n_0 &\implies |\|y_n - x\|^2 - \alpha^2| < \frac{\varepsilon^2}{4} \text{ e } |\|y_m - x\|^2 - \alpha^2| < \frac{\varepsilon^2}{4} \\ &\implies \|y_n - x\|^2 < \alpha^2 + \frac{\varepsilon^2}{4} \text{ e } \|y_m - x\|^2 < \alpha^2 + \frac{\varepsilon^2}{4} \\ &\implies \|y_m - y_n\|^2 < \varepsilon^2 \implies \|y_m - y_n\| < \varepsilon. \end{aligned}$$

Logo, (y_n) é de Cauchy. Como M é completo, existe $y \in M$ tal que

$$\lim_{n \to \infty} y_n = y,$$

donde, pela continuidade da norma,

$$\alpha = \lim_{n \to \infty} \|x - y_n\| = \|x - y\|.$$

Isto prova a existência.

Unicidade. Sejam $\overline{y} \in M$ tal que

$$\|x - \overline{y}\| = \alpha \quad \text{e} \quad 0 \leq t \leq 1.$$

Daí,

$$(1 - t)\overline{y} + ty \in M$$

e

$$\|x - (1-t)\overline{y} + ty\| = \|(1-t)(x-\overline{y}) + t(x-y)\| \leq (1-t)\|x-\overline{y}\| + t\|x-y\| = \alpha,$$

donde

$$\|x - (1-t)\overline{y} + ty\| = \alpha \implies [y, \overline{y}] \subseteq S_\alpha(x) \implies y = \overline{y},$$

já que a esfera $S(x, \alpha)$ (ver Definição B.5.2) não contém segmentos de reta não degenerados. Isto prova a unicidade.

Ortogonalidade. suponhamos, por redução ao absurdo, que $z \notin M^\perp$. Desse modo, existe $\overline{y} \in M$ tal que $\beta := \langle z, \overline{y} \rangle \neq 0$. Como $\beta \neq 0$, então $\overline{y} \neq 0$. Note que

$$\begin{aligned} \|z - \alpha\overline{y}\|^2 = \langle z - \alpha\overline{y}, z - \alpha\overline{y} \rangle &= \|z\|^2 - \overline{\alpha}\langle z, \overline{y} \rangle - \alpha\langle \overline{y}, z \rangle + \alpha\overline{\alpha}\|\overline{y}\|^2 \\ &= \|z\|^2 - \alpha\langle \overline{y}, z \rangle - \overline{\alpha}(\langle z, \overline{y} \rangle - \alpha\|\overline{y}\|^2), \end{aligned}$$

para todo $\alpha \in \mathbb{R}$. Em particular,

$$\|z - \alpha_0\overline{y}\|^2 = \|z\|^2 - \alpha_0\langle \overline{y}, z \rangle = \|z\|^2 - \frac{|\beta|^2}{\|y\|^2} < \|z\|^2,$$

onde $\alpha_0 := \dfrac{\beta}{\|y\|^2}$. Logo,

$$\|x - h\| = \|x - y - \alpha_0\overline{y}\| = \|z - \alpha_0\overline{y}\| < \|z\| = \|x - y\| = \inf_{\tilde{y} \in M} \|x - \tilde{y}\|,$$

onde $h := y + \alpha_0\overline{y} \in M$, o que é absurdo. Portanto, $z \in M^\perp$. □

O corolário abaixo foi retirado de (KREYSZIG, 1989, p. 146).

Corolário C.9.2 (Soma Direta)**.** Se Y é um subespaço fechado de um espaço de Hilbert H, então

$$H = Y \oplus Y^\perp.$$

Demonstração. Seja $x \in H$. Como Y é convexo (já que é subespaço vetorial de H), completo (pelo Teorema C.8.2) e não vazio (pois $0 \in Y$), então, pelo Teorema C.9.1, existe um único $y \in H$ tal que $\inf_{\overline{y} \in Y} \|x - \overline{y}\| = \|x - y\|$ e $z := x - y \in Y^\perp$, então $x = y + z \in Y + Y^\perp$. Como $Y \cap Y^\perp = \{0\}$, então $H = Y \oplus Y^\perp$. □

Lema C.9.3. Se $f \in \mathcal{L}(E, F)$, então $\ker f$ é um subespaço fechado de E.

Demonstração. Sejam $(x_n) \subseteq \ker f$ uma sequência convergente e $\overline{x} := \lim_{n \to \infty} x_n$. Como $f(x_n) = 0$, para todo $n \in \mathbb{N}$, e f é contínua, então $0 = \lim_{n \to \infty} f(x_n) = f(\overline{x})$, donde $\overline{x} \in \ker f$. Portanto, $\ker f$ é um subespaço fechado de E. □

O teorema abaixo, que se encontra em (KREYSZIG, 1989, p. 188), é utilizado, como já mencionamos, na definição de vetor gradiente (veja a Seção 5 do Capítulo 1).

Teorema C.9.4 (Teorema da Representação de Riesz)**.** Se H é um espaço de Hilbert e $f : H \to \mathbb{R}$ é um funcional linear contínuo, então existe um único vetor $v \in H$ tal que

$$f(h) = \langle h, v \rangle,$$

para todo $h \in H$, e $\|v\| = \|f\|$.

Demonstração. Se $f = 0$, tomamos $v := 0$.

Suponhamos que $f \neq 0$. Assim, existe $w \in H$ tal que $w \notin \ker f$. Como $H = \ker f \oplus \ker f^{\perp}$, como combinação do Corolário C.9.2 e do Lema C.9.3, então existem únicos $a \in \ker f$ e $b \in \ker f^{\perp}$ tais que $w = a + b$. Se fosse $b = 0$, teríamos $w = a \in \ker f$, o que é absurdo. Logo, $b \neq 0$.

Considere a aplicação $g : H \to \mathbb{R}$ dada por

$$g(x) := f(x)b - f(b)x, \quad \text{para cada} \quad x \in H.$$

Como

$$f(g(x)) = f(x)f(b) - f(b)f(x) = 0,$$

então $g(x) \in \ker f$. Como $b \in \ker f^{\perp}$, então $b \perp g(x)$, donde

$$0 = \langle b, g(x) \rangle = \langle b, f(x)b - f(b)x \rangle = f(x)\|b\|^2 - f(b)\langle b, x \rangle \implies f(x) = \langle x, v \rangle,$$

onde $v := \dfrac{f(b)}{\|b\|^2} b$. Isto prova a existência.

Quanto à unicidade, seja $\tilde{v} \in H$ tal que $f(h) = \langle h, \tilde{v} \rangle$, para todo $h \in H$. Assim,

$$\langle h, \tilde{v} \rangle = \langle h, v \rangle \implies \langle h, \tilde{v} - v \rangle = 0,$$

para todo $h \in H$. Em particular,

$$\langle \tilde{v} - v, \tilde{v} - v \rangle = 0 \implies \|\tilde{v} - v\|^2 = 0 \implies \tilde{v} = v.$$

A fim de verificar que $\|v\| = \|f\|$, note que

$$\|f\|\|v\| \geq \left\| f\left(\frac{v}{\|v\|}\right) \right\| \cdot \|v\| = f(v) = \langle v, v \rangle = \|v\|^2 \implies \|f\| \geq \|v\|.$$

Além disso, pela Desigualdade de Cauchy-Schwarz (ver Teorema C.5.2),

$$|f(x)| = |\langle x, v \rangle| \leq \|x\|\|v\| \implies \|f\| = \sup_{\|x\|=1} |\langle x, v \rangle| \leq \|v\|.$$

Logo, $\|v\| = \|f\|$. $\square$

C.10 Teorema da Aplicação Aberta

O nosso objetivo agora é provar o Teorema da Aplicação Aberta (ver Teorema C.10.4) e o Teorema do Ponto Fixo de Banach (ver Teorema C.11.2), que serão utilizados no Capítulo 4.

Lema C.10.1. Sejam E um espaço métrico, $F \subseteq E$ um subconjunto fechado de interior vazio e $J \subseteq E$ um subconjunto aberto não vazio. Nessas condições, E contém uma bola fechada B que não intersecta F e tal que $B \subseteq J$.

Demonstração. Suponhamos, por redução ao absurdo, que esse não seja o caso. Dado $c \in J$, como J é aberto, então

$$B[c, \varepsilon] \subseteq J, \quad \text{para algum } \varepsilon > 0.$$

Note que

$$B_n := B[c, \varepsilon/n] \subseteq B[c, \varepsilon] \subseteq J, \quad \text{para todo } n \in \mathbb{N}\,.$$

Por hipótese, para cada $n \in \mathbb{N}$, existe $c_n \in B_n \cap F$. Como

$$d(c_n, c) \leq \varepsilon/n, \quad \text{para todo } n \in \mathbb{N},$$

então, pelo Teorema do Confronto,

$$\lim_{n\to\infty} c_n = c,$$

donde, do fato de que F é fechado e cada $c_n \in F$, vem $c \in F$.

Logo, $J \subseteq F$ e, portanto, F não tem interior vazio. Absurdo! □

Teorema C.10.2 (Teorema de Baire)**.** Sejam E um espaço métrico completo e $F_1, F_2, \ldots$ subconjuntos fechados de E com interior vazio. Então, $\bigcup_{i\in\mathbb{N}} F_i$ tem interior vazio.

Demonstração. Seja $J \subseteq E$ um subconjunto aberto não vazio. Como F_1 é um subconjunto fechado de E com interior vazio, pelo Lema C.10.1, temos

$$F_1 \cap B_1 = \varnothing, \quad \text{para algum } B_1 := B[y_1, \delta_1] \subseteq J.$$

Como $B(y_1, \delta_1/2)$ é um subconjunto aberto de E e F_2 é um subconjunto fechado de E, então, novamente, temos

$$F_2 \cap B_2 = \varnothing, \quad \text{para algum } B_2 := B[y_2, \delta_2] \subseteq B(y_1, \delta_1/2).$$

Como $B(y_2, \delta_2/2)$ é um subconjunto aberto de E e F_3 é um subconjunto fechado de E, então, novamente, temos

$$F_3 \cap B_3 = \varnothing, \quad \text{para algum } B_3 := B[y_3, \delta_3] \subseteq B(y_2, \delta_2/2).$$

Continuando esse processo, obtemos uma sequência (y_n) de elementos de E, uma sequência (δ_n) de números positivos que decresce para zero e uma sequência encaixada (B_n) de bolas fechadas de E tais que

$$B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots \quad \text{e} \quad B_n := B[y_n, \delta_n].$$

Como

$$d(y_m, y_n) \leq \delta_{\min\{m,n\}} \xrightarrow{m,n\to\infty} 0,$$

então (y_n) é uma sequência de Cauchy em E, donde, da completude de E, existe

$$y := \lim_{n\to\infty} y_n.$$

A mesma desigualdade nos dá, para todo $n \in \mathbb{N}$ fixado,

$$d(y, y_n) = \lim_{m\to\infty} d(y_m, y_n) \leq \lim_{m\to\infty} \delta_{\min\{m,n\}} = \delta_n,$$

o que prova que

$$y \in B_n, \quad \text{para todo } n \in \mathbb{N},$$

donde

$$y \notin F_n, \quad \text{para todo } n \in \mathbb{N},$$

e, portanto,

$$y \notin \bigcup_{n\in\mathbb{N}} F_n.$$

Logo,

$$J \not\subseteq \bigcup_{n\in\mathbb{N}} F_n.$$

Como J era um aberto não vazio arbitrário, isso prova que a união dos F_n não contém nenhum aberto não vazio e, portanto, tem interior vazio. □

Lema C.10.3. (COLEMAN, 2012, Lemma 8.3, p. 185) Sejam E e F espaços de Banach e $L \in \mathcal{L}(E, F)$ sobrejetiva. Então, a imagem de uma bola aberta centrada em 0_E contém uma bola aberta centrada em 0_F.

Demonstração. Defina

$$B_r := B_E(0, r) \quad \text{e} \quad B'_r := B_F(0, r), \quad \text{para todo } r > 0.$$

Da sobrejetividade de L,

$$F = L(E) = L\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} L(B_n) \subseteq \bigcup_{n=1}^{\infty} \overline{L(B_n)} \subseteq F,$$

o que nos permite dizer

$$F = \bigcup_{n=1}^{\infty} \overline{L(B_n)}.$$

Como F tem interior não vazio, o Teorema de Baire (ver Teorema C.10.2) implica que

$$\overline{L(B_m)} \text{ tem interior não vazio, para algum } m \in \mathbb{N}\,.$$

De

$$\varnothing \neq A := \operatorname{int} \overline{L(B_m)} \subseteq \overline{L(B_m)}$$

vem

$$A \cap \overline{L(B_m)} \neq \varnothing.$$

Como A é uma vizinhança de cada um de seus pontos, então

$$C := A \cap L(B_m) \neq \varnothing.$$

Tomando $y_0 \in C$, temos

$$y_0 = L(x_0), \quad \text{para algum } x_0 \in B_m.$$

Note que, se $x \in B_m$, então

$$\|x - x_0\| \leq \|x - 0\| + \|0 - x_0\| \leq m + m = 2m,$$

donde $x - x_0 \in B_{2m}$ e, portanto,

$$L(x) - y_0 = L(x) - L(x_0) = L(x - x_0) \in L(B_{2m}).$$

Logo,

$$L(B_m) - y_0 \subseteq L(B_{2m}),$$

o que também implica

$$\overline{L(B_m)} - y_0 = \overline{L(B_m) - y_0} \subseteq \overline{L(B_{2m})}.$$

Como

$$y_0 \in A \implies 0 \in A - y_0,$$

então

$$0 \in \operatorname{int} \overline{L(B_m)} - y_0 = \operatorname{int}(\overline{L(B_m)} - y_0) \subseteq \operatorname{int} \overline{L(B_{2m})} \subseteq \operatorname{int}(2m\overline{L(B_1)}),$$

donde

$$0 \in \operatorname{int} \overline{L(B_1)}$$

e, portanto, existe $\varepsilon > 0$ tal que

$$B'_\varepsilon \subseteq \overline{L(B_1)}.$$

Afirmamos que

$$B'_\varepsilon \subseteq L(B_3).$$

Tome
$$y \in B'_\varepsilon \quad \text{e} \quad r \in (0, \varepsilon/2)$$
tais que $B(y,r) \subseteq B'_\varepsilon$. Nessas condições,
$$B(y,r) \subseteq \overline{L(B_1)},$$
donde existe $y_1 \in B(y,r) \cap L(B_1)$ e, portanto, existe $x_1 \in B_1$ tal que $L(x_1) = y_1$.

Agora,
$$B'_{\varepsilon/2} = (1/2)B'_\varepsilon \subseteq \overline{(1/2)L(B_1)} = \overline{L(B_{1/2})}.$$
Ponhamos
$$y' := y - y_1$$
e tomamos
$$r' \in (0, \varepsilon/2^2) \text{ suficientemente pequeno tal que } B(y', r') \subseteq B'_{\varepsilon/2}.$$
Como
$$B(y', r') \subseteq \overline{L(B_{1/2})},$$
então existe $y_2 \in B(y', r') \cap L(B_{1/2})$, donde existe $x_2 \in L(B_{1/2})$ tal que $L(x_2) = y_2$. Note que
$$y_2 \in B(y', r') \implies \|y' - y_2\| < \frac{\varepsilon}{2^2} \implies \|y - y_1 - y_2\| < \frac{\varepsilon}{2^2}.$$
Continuando esse processo obtemos uma sequência $(x_n) \subseteq E$ e uma sequência $(y_n) \subseteq F$ tais que $y_n = L(x_n)$, para cada $n \in \mathbb{N}$,
$$\|x_n\| < \frac{1}{2^{n-1}} \quad \text{e} \quad \left\| y - \sum_{i=1}^{n} y_i \right\| < \frac{\varepsilon}{2^n}.$$
Pondo
$$s_n := \sum_{i=1}^{n} x_i,$$
temos
$$\|s_m - s_n\| = \left\| \sum_{i=\min\{m,n\}+1}^{\max\{m,n\}} x_i \right\| \leq \sum_{i=\min\{m,n\}+1}^{\max\{m,n\}} \frac{1}{2^{i-1}} \xrightarrow{m,n\to\infty} 0,$$
garantindo que (s_n) é uma sequência de Cauchy em E. Como E é Banach, existe
$$x := \lim_{n\to\infty} s_n \in E,$$
donde vem
$$\|x\| = \left\| \lim_{n\to\infty} \sum_{i=1}^{n} x_i \right\| = \lim_{n\to\infty} \left\| \sum_{i=1}^{n} x_i \right\| \leq \lim_{n\to\infty} \sum_{i=1}^{n} \|x_i\| \leq \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{2^{i-1}} = 2 < 3.$$
Note ainda que, como L é contínuo, temos
$$L(x) = \lim_{n\to\infty} L(s_n) = \lim_{n\to\infty} \sum_{i=1}^{n} L(x_i) = \lim_{n\to\infty} \sum_{i=1}^{n} y_i = y,$$

o que nos dá $y \in L(B_3)$. Consequentemente, da arbitrariedade de y, temos

$$B'_\varepsilon \subseteq L(B_3) \implies B'_{\varepsilon/3} \subseteq L(B_1),$$

o que termina a prova. $\square$

Finalmente podemos provar o Teorema da Aplicação Aberta.

Teorema C.10.4. (COLEMAN, 2012, Theorem 8.6., p. 187. Open mapping theorem) Sejam E e F espaços de Banach e $L : E \to F$ uma transformação linear contínua e sobrejetiva. Então, L é uma aplicação aplicação aberta, isto é, L mapeia subconjuntos abertos de E em subconjuntos abertos de F. Além disso, se L também for injetiva, então L é um homeomorfismo.

Demonstração. Sejam $U \subseteq E$ um subconjunto aberto e $y \in L(U)$. Então, existe $x \in U$ tal que $L(x) = y$. Como U é aberto, existe $\varepsilon > 0$ tal que $B(x, \varepsilon) \subseteq U \subseteq E$. Pelo Lema C.10.3, existe uma bola aberta $B'_r \subseteq F$ centrada em 0_F tal que $B'_r \subseteq L(B_\varepsilon)$. Então,

$$B'(y, r) = y + B'_r \subseteq y + L(B_\varepsilon) = L(x + B_\varepsilon) = L(B(x, \varepsilon)) \subseteq L(U),$$

mostrando que $L(U)$ é aberto.

Agora, se L também é injetiva, então L é bijetiva e, como L é uma aplicação aberta, $P := L^{-1}$ é uma aplicação tal que pré-imagem por P de subconjuntos abertos de F é um subconjunto aberto de E. Logo, P é contínua. $\square$

C.11 Teorema do Ponto Fixo de Banach

Por fim, vamos provar o Teorema do Ponto Fixo de Banach.

Sejam E um espaço métrico, $S \subseteq E$ e $f : S \to S$ uma aplicação. Dizemos que f é uma **contração** se existe $k \in [0, 1)$ tal que

$$d(f(x), f(y)) \leq k \cdot d(x, y), \quad \text{para quaisquer } x, y \in S.$$

Lema C.11.1. Sejam E um espaço métrico e $(x_n) \subseteq E$. Suponha que existe $\lambda \in [0, 1)$ tal que, para todo $n \in \mathbb{N}$,

$$d(x_{n+2}, x_{n+1}) \leq \lambda \cdot d(x_{n+1}, x_n).$$

Nessas condições, (x_n) é uma sequência de Cauchy.

Demonstração. Se $\lambda = 0$, então (x_n) é identicamente nula. Suponhamos $\lambda > 0$. Assim, dados $m, n \in \mathbb{N}$, com $m > n$, pondo $k := m - n$, temos

$$\begin{aligned} d(x_m, x_n) \leq \sum_{i=1}^{k} d(x_{n+i}, x_{n+i-1}) &\leq (\lambda^{n-1} + \cdots + \lambda^{n+k-2}) \cdot d(x_2, x_1) \\ &= \lambda^{n-1}(1 + \lambda + \cdots + \lambda^{k-1}) \cdot d(x_2, x_1) \\ &= \lambda^{n-1}\frac{1-\lambda^k}{1-\lambda} d(x_2, x_1) \\ &< \frac{\lambda^{n-1}}{1-\lambda} d(x_2, x_1), \end{aligned}$$

já que

$$0 < 1 - \lambda^k < 1.$$

Como

$$\lim_{n\to\infty} \frac{\lambda^{n-1}}{1-\lambda} d(x_2, x_1) = 0,$$

então (x_n) é de Cauchy. □

O Teorema abaixo foi extraído de (COLEMAN, 2012, Theorem 8.7, p. 188).

Teorema C.11.2 (Teorema do Ponto Fixo de Banach)**.** Sejam E um espaço métrico completo e $S \subseteq E$ um subconjunto fechado. Se $f : S \to S$ é uma contração, então f tem um, e somente um, ponto fixo.

Demonstração. **Existência do ponto fixo.** Sejam $x_0 \in S$ e

$$x_n := f(x_{n-1}), \quad \text{para cada } n \in \mathbb{N}\,.$$

Como f é uma contração, então existe $k \in [0, 1)$ tal que

$$d(f(x), f(y)) \leq k \cdot d(x, y), \quad \text{para quaisquer } x, y \in S.$$

Como

$$d(x_{n+2}, x_{n+1}) = d(f(x_{n+1}), f(x_n)) \leq k \cdot d(x_{n+1}, x_n),$$

para todo $n \in \mathbb{N}$, então, pelo Lema C.11.1, a sequência (x_n) é de Cauchy e, portanto, convergente, já que E é um espaço métrico completo. Como $(x_n) \subseteq S$ e S é fechado, então

$$\lim_{n\to\infty} x_n =: x \in S.$$

Como $(f(x_n))$ é uma subsequência de (x_n) e f é contínua, então

$$x = \lim_{n\to\infty} x_n = \lim_{n\to\infty} f(x_n) = f(x).$$

Logo, x é um ponto fixo de f.

Unicidade do ponto fixo. Vamos provar agora que x é o único ponto fixo de f. Com efeito, se $f(y) = y$, então

$$d(x, y) = d(f(x), f(y)) \leq k \cdot d(x, y),$$

o que nos dá

$$\underbrace{(1 - k)}_{>0} d(x, y) \leq 0,$$

donde

$$d(x, y) \leq 0$$

e, portanto,

$$x = y,$$

como queríamos demonstrar. $\square$