Capítulo 5
Conjuntos Convexos e Funções Convexas

Sumário

Funções convexas ou côncavas desempenham um papel na Termodinâmica, na Mecânica Estatística, na Mecânica Clássica (notadamente por meio das Transformadas de Legendre, tratada na Seção **??**, página **??**), na Teoria das Probabilidades, na Teoria das Equações Diferenciais, no Cálculo Variacional e em outras áreas. Pretendemos neste breve capítulo apresentar suas definições e suas propriedades básicas para futuro uso e referência. Na Seção 5.3, página 366, obtemos algumas desigualdades úteis envolvendo funções convexas e côncavas, como as desigualdades de Young e de Minkowski, a mais relevante sendo, talvez, a desigualdade de Jensen, apresentada na Proposição 5.19, página 366.

A área da Matemática dedicada ao estudo de propriedades de conjuntos convexos e funções convexas é denominada *Análise Convexa* (vide, *e.g.*, [449]). A Análise Convexa possui também aplicações a problemas de optimização em diversas áreas da Matemática Pura e Aplicada (vide, *e.g.*, [73]). Historicamente, o início da Análise Convexa é creditado a trabalhos de Minkowski[1]. Para uma referência clássica sobre funções convexas e desigualdades associadas às mesmas, vide [225].

Nestas Notas faremos em diversos momentos uso de propriedades de funções convexas ou côncavas, por exemplo, no tratamento da Função Gama de Euler no Capítulo 7, página 408 (vide o Teorema de Bohr-Mollerup, apresentado na Seção 7.5.1, página 425). Propriedades de conjuntos convexos serão evocadas na Teoria dos Espaços de Hilbert (capítulo 41, página 2288), notadamente no Teorema do Melhor Aproximante, Teorema 41.1, página 2292, e em sua generalização a espaços uniformemente convexos, tratada na Seção 5.1.3, página 342, e na Seção 25.6, página 1456.

Para o emprego da teoria das funções convexas na Termodinâmica, vide, *e.g.*, [581], [500], [86].

5.1 Conjuntos Convexos. Noções Básicas

• Combinações lineares convexas em espaços vetoriais reais

Seja $\mathcal{V}$ um espaço vetorial real e sejam v_1, v_2 elementos de $\mathcal{V}$. Um vetor da forma $\lambda v_1 + (1-\lambda)v_2$ com $\lambda \in [0,\ 1]$ é dito ser uma *combinação linear convexa* de v_1 e v_2. O conjunto $\big\{\lambda v_1 + (1-\lambda)v_2,\ \lambda \in [0,\ 1]\big\}$ é dito ser o segmento de reta conectando v_1 a v_2. Note-se que essas definições permanecem mesmo no caso $v_1 = v_2$.

[1]Hermann Minkowski (1864–1909).

Essas definições podem ser generalizadas para mais vetores. Para $m \in \mathbb{N}$, $v_1, \ldots, v_m \in \mathcal{V}$ e $\lambda_1, \ldots, \lambda_m \in [0, 1]$ tais que $\sum_{k=1}^m \lambda_k = 1$, o vetor $\sum_{k=1}^m \lambda_k v_k$ é dito ser uma *combinação linear convexa* de $v_1, \ldots, v_m$. O conjunto

$$\mathrm{Pol}(v_1, \ldots, v_m) := \left\{ \sum_{k=1}^m \lambda_k v_k, \text{ com } \lambda_1, \ldots, \lambda_m \in [0, 1] \text{ e } \sum_{k=1}^m \lambda_k = 1 \right\} \tag{5.1}$$

é denominado *poliedro convexo* gerado por $v_1, \ldots, v_m$.

• Conjuntos afins em espaços vetoriais reais

Seja $\mathcal{V}$ um espaço vetorial real e sejam v_1, v_2 elementos de $\mathcal{V}$. Um vetor da forma $\lambda v_1 + (1-\lambda)v_2$ com $\lambda \in \mathbb{R}$ é dito ser uma *combinação linear afim* de v_1 e v_2. Note que a distinção em relação à noção de combinação linear convexa é que naquele caso λ é restrito ao intervalo $[0, 1]$, enquanto que para combinações lineares afins λ pode ser um elemento de todo $\mathbb{R}$.

Seja $\mathcal{V}$ um espaço vetorial real. Um conjunto não vazio $J \subset \mathcal{V}$ é dito ser um *conjunto afim* se para todos $x, y \in J$ e todo $\lambda \in \mathbb{R}$ valer $\lambda x + (1-\lambda)y \in J$. O conjunto vazio de $\mathcal{V}$ é também, honorificamente, declarado afim.

Exemplo 5.1 Seja $\mathcal{P}_n$ o espaço vetorial de todos os polinômios de grau menor ou igual a $n \in \mathbb{N}_0$ de uma variável complexa z. Assim, cada $p \in \mathcal{P}_n$ é da forma $p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$, com os coeficientes a_k's sendo números complexos. Um polinômio $p \in \mathcal{P}_n$ é dito ser mônico se o coeficiente do termo de maior grau for igual a um, ou seja, se $p(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$. O conjunto $\mathcal{M}_n \subset \mathcal{P}_n$ de todos os polinômios mônicos de grau n é um subconjunto afim de $\mathcal{P}_n$. Verifique! ♦

E. 5.1 *Exercício.* Mostre que um subconjunto afim de um espaço vetorial real é um subespaço se e somente se contiver o vetor nulo. ✠

• Conjuntos convexos em espaços vetoriais reais

Seja $\mathcal{V}$ um espaço vetorial real. Um conjunto não vazio $C \subset \mathcal{V}$ é dito ser um *conjunto convexo* se para todos $x, y \in C$ e todo $\lambda \in [0, 1]$ valer $\lambda x + (1-\lambda)y \in C$, ou seja, se o segmento de reta conectando quaisquer $x, y \in C$ estiver inteiramente contido em C. O conjunto vazio de $\mathcal{V}$ é também, honorificamente, declarado convexo.

É evidente que todo subconjunto afim de $\mathcal{V}$ é também convexo. A recíproca pode não ser verdadeira.

Se $C \subset \mathcal{V}$ é convexo e $z \in C$, dizemos que uma expressão do tipo $z = \lambda x + (1-\lambda)y$ com $x, y \in C$ e $\lambda \in [0, 1]$ é uma *decomposição convexa* de z. Há três situações nas quais uma tal decomposição convexa é trivial: quando $x = y = z$ e $\lambda \in [0, 1]$ é arbitrário, quando $x = z$, $\lambda = 1$ e $y \in C$ é arbitrário ou quando $y = z$, $\lambda = 0$ e $x \in C$ é arbitrário. Nesses casos a decomposição convexa é trivial: $z = z$.

Seja $C \subset \mathcal{V}$ convexo. Dizemos que $z \in C$ é um *ponto interior* de C se existirem $x, y \in C$ com $x \neq y$ e $\lambda \in (0, 1)$ tais que $z = \lambda x + (1-\lambda)y$. Em outras palavras, $z \in C$ é um ponto interior de C se admitir ao menos uma decomposição convexa não trivial.

Dizemos que $z \in C$ é um *ponto extremo* (ou *ponto extremal*) de C se não existirem $x, y \in C$ distintos e $\lambda \in (0, 1)$ tais que $z = \lambda x + (1-\lambda)y$. Em outras palavras, $z \in C$ é um ponto extremo de C se admitir somente decomposições convexas triviais, ou seja, se não for um ponto interior de C.

Exemplo 5.2 No caso em que $\mathcal{V}$ é o espaço $\mathbb{R}^n$, um conjunto $C \subset \mathbb{R}^n$ é convexo se e somente se o segmento de reta conectando dois pontos quaisquer de C estiver inteiramente contido em C. Um triângulo aberto ou fechado em $\mathbb{R}^2$ é um conjunto convexo. Se o triângulo for aberto, todos os seus pontos são interiores e não há pontos extremos. Se o triângulo for fechado todos os seus pontos são interiores, exceto seus três vértices, que são seus únicos pontos extremos. ♦

Exemplo 5.3 Um poliedro convexo como em (5.1) é um conjunto convexo. De fato, se $u = \sum_{k=1}^m \lambda_k v_k$ e $w = \sum_{k=1}^m \lambda'_k v_k$ são elementos de $\mathrm{Pol}(v_1, \ldots, v_m)$, então, para $\mu \in [0, 1]$ temos

$$\mu u + (1-\mu)w = \sum_{k=1}^m \big(\mu\lambda_k + (1-\mu)\lambda'_k\big)v_k ,$$

que é também elemento de $\mathrm{Pol}(v_1, \ldots, v_m)$, pois $\mu\lambda_k + (1-\mu)\lambda'_k \geq 0$ (por ser uma soma de produtos de números não negativos) e

$$\sum_{k=1}^{m}\big(\mu\lambda_k+(1-\mu)\lambda'_k\big) \;=\; \mu\underbrace{\sum_{k=1}^{m}\lambda_k}_{=1}+(1-\mu)\underbrace{\sum_{k=1}^{m}\lambda'_k}_{=1} \;=\; \mu+(1-\mu) \;=\; 1\;,$$

o que, em particular, implica $\mu\lambda_k+(1-\mu)\lambda'_k \leq 1$ para cada k, mostrando que $\mu\lambda_k+(1-\mu)\lambda_k \in [0,\ 1]$. ♦

Um fato elementar mas digno de observação é o seguinte:

Proposição 5.1 *Seja $\mathcal{V}$ um espaço vetorial real. Se $C \subset \mathcal{V}$ é convexo, então C contém todas as combinações lineares convexas finitas de elementos de C.* □

Prova. A prova será feira por indução. Seja $n \in \mathbb{N}$, $n \geq 2$, e suponhamos que para todos $v_1,\ \ldots,\ v_n \in C$ e todos $\lambda_1,\ \ldots,\ \lambda_n \in [0,\ 1]$ com $\sum_{k=1}^{n}\lambda_k = 1$ valha que $\sum_{k=1}^{n}\lambda_k v_k \in C$. Como C é suposto convexo, sabemos de antemão que essa suposição é válida caso $n = 2$.

Sejam $w_1,\ \ldots,\ w_{n+1}$ elementos de C e sejam $\mu_1,\ \ldots,\ \mu_{n+1} \in [0,\ 1]$ com $\sum_{k=1}^{n+1}\mu_k = 1$. Desejamos provar que $W \equiv \sum_{k=1}^{n+1}\mu_k w_k \in C$.

Podemos supor que haja dois elementos não nulos dentre os números $\mu_1,\ \ldots,\ \mu_{n+1} \in [0,\ 1]$, pois se houver apenas um, digamos, μ_1, valerá $\mu_1 = 1$ e $W = w_1 \in C$, não havendo, portanto, o que se provar. Sem perda de generalidade, vamos supor, então, que μ_n e μ_{n+1} sejam não nulos. Podemos escrever

$$W \;=\; \sum_{k=1}^{n-1}\mu_k w_k+(\mu_n+\mu_{n+1})\left[\frac{\mu_n}{\mu_n+\mu_{n+1}}w_n+\frac{\mu_{n+1}}{\mu_n+\mu_{n+1}}w_{n+1}\right]\;. \tag{5.2}$$

Agora, $\frac{\mu_n}{\mu_n+\mu_{n+1}}w_n+\frac{\mu_{n+1}}{\mu_n+\mu_{n+1}}w_{n+1}$ é um elemento de C, por ser uma combinação linear convexa de dois elementos de C. Como $\mu_1+\cdots+\mu_{n-1}+(\mu_n+\mu_{n+1}) = 1$, o lado direito de (5.2) informa que W é uma combinação linear convexa de n elementos de C. Pela hipótese de indução, isso implica $W \in C$. Isso completa a demonstração almejada. ■

5.1.1 Operações que Preservam Convexidade

• **Operações em subconjuntos de um espaço vetorial**

Seja $\mathcal{V}$ um espaço vetorial real e sejam A_1 e A_2 dois subconjuntos não vazios de $\mathcal{V}$. Definimos a *soma* de A_1 e A_2 por $A_1+A_2 := \big\{a_1+a_2,\ a_1 \in A_1,\ a_2 \in A_2\big\}$. Assim, A_1+A_2 é composto por todos os vetores de $\mathcal{V}$ que são a soma de um vetor de A_1 com um vetor de A_2. Se $A \subset \mathcal{V}$ é não vazio, definimos $-A := \{-a,\ a \in \mathcal{V}\}$.

Com mais generalidade, para $\alpha_1,\ \alpha_2 \in \mathbb{R}$ definimos $\alpha_1 A_1+\alpha_2 A_2 := \big\{\alpha_1 a_1+\alpha_2 a_2,\ a_1 \in A_1,\ a_2 \in A_2\big\}$.

Para $A \subset \mathcal{V}$ e $\beta \in \mathcal{V}$, definimos $A+\beta := \{a+\beta,\ a \in A\}$. É claro que $A+\beta = A+\{\beta\}$ com o lado direito dado no sentido de soma de subconjuntos de $\mathcal{V}$.

Se $M : \mathcal{V} \to \mathcal{V}$ é um operador linear e $A \subset \mathcal{V}$, definimos $MA := \big\{Ma,\ a \in A\big\}$.

A transformação $\mathcal{V} \supset A \mapsto MA+\beta \subset \mathcal{V}$ é dita ser uma *transformação afim de subconjuntos* de $\mathcal{V}$.

• **Operações em subconjuntos convexos de um espaço vetorial**

No nosso presente contexto, o seguinte fato é relevante:

Lema 5.1 *Seja $\mathcal{V}$ um espaço vetorial real e sejam C_1 e C_2 dois subconjuntos convexos não vazios de $\mathcal{V}$. Sejam também $\alpha_1,\ \alpha_2 \in \mathbb{R}$. Então, $\alpha_1 C_1+\alpha_2 C_2$ é igualmente convexo.*

Em particular, C_1+C_2 é igualmente convexo e, se $C \subset \mathcal{V}$ é convexo, então $-C$ é igualmente convexo.

Por fim, seja $M : \mathcal{V} \to \mathcal{V}$ um operador linear e $\beta \in \mathcal{V}$. Então, se $C \subset \mathcal{V}$ for convexo, $MC+\beta$ também o é, ou seja, a convexidade é preservada por transformações afins (vide definição na Seção 5.2.2, página 348). □

Prova. Seja $\lambda \in [0, 1]$ e sejam $v_1, w_1 \in C_1$ e $v_2, w_2 \in C_2$. Como C_1 e C_2 são convexos, tem-se $\lambda v_1 + (1-\lambda)w_1 \in C_1$ e $\lambda v_2 + (1-\lambda)w_2 \in C_2$. Com isso, vemos que

$$\lambda(\alpha_1 v_1 + \alpha_2 v_2) + (1-\lambda)(\alpha_1 w_1 + \alpha_2 w_2) = \alpha_1 \big[\underbrace{\lambda v_1 + (1-\lambda)w_1}_{\in C_1}\big] + \alpha_2 \big[\underbrace{\lambda v_2 + (1-\lambda)w_2}_{\in C_2}\big] \in \alpha_1 C_1 + \alpha_2 C_2 ,$$

que é o que desejávamos estabelecer.

Se $C \subset \mathcal{V}$ for convexo, então $C + \beta$ também o é, pois $C + \beta = C + \{\beta\}$ e o conjunto $\{\beta\}$ é convexo, por ser composto de um único elemento. (Uma verificação direta da covexidade de $C + \beta$ é igualmente elementar).

Sejam $M : \mathcal{V} \to \mathcal{V}$ um operador linear, $C \subset \mathcal{V}$ convexo e $\lambda \in [0, 1]$. Sejam dois elementos genréricos de MC da forma Mu e Mv, com $u, v \in C$. Então, pela linearidade de M vale $\lambda Mu + (1-\lambda)Mv = M\big(\lambda u + (1-\lambda)v\big) \in MC$, pois $\lambda u + (1-\lambda)v \in C$, devido à suposta convexidade deste. ■

• **Convexidade e o fecho em espaços normados**

No caso de espaços vetoriais normados, se um conjunto é covexo seu fecho também o é. Esse fato, que possui várias generalizações em outras topologias, é particularmente importante e de demonstração simples.

Proposição 5.2 *Seja $\mathcal{V}$ espaço vetorial real normado, com norma $\|\cdot\|$. Então, se $C \subset \mathcal{V}$ é convexo, seu fecho $\overline{C}$ na topologia métrica induzida pela norma é igualmente convexo.* □

Prova. Sejam $u, v \in \overline{C}$ e considere-se a combinação linear convexa $\lambda u + (1-\lambda)v$ para $\lambda \in [0, 1]$. Como u e v estão no fecho de C, existem sequências $\{u_n \in C, n \in \mathbb{N}\}$ e $\{v_n \in C, n \in \mathbb{N}\}$ convergendo a u e v, respectivamente (pela Proposição 28.12, página 1551). Assim,

$$\Big\|\big(\lambda u + (1-\lambda)v\big) - \big(\lambda u_n + (1-\lambda)v_n\big)\Big\| = \Big\|\lambda(u - u_n) + (1-\lambda)(v - v_n)\Big\| \le \lambda\|u - u_n\| + (1-\lambda)\|v - v_n\| .$$

Como $\|u - u_n\|$ e $\|v - v_n\|$ vão a zero quando n vai a infinito, isso mostrou que a sequência $\lambda u_n + (1-\lambda)v_n$, $n \in \mathbb{N}$, converge a $\lambda u + (1-\lambda)v$ quando $n \to \infty$. Agora, $\lambda u_n + (1-\lambda)v_n$ é um elemento de C, pois C é convexo. Isso provou (novamente pela Proposição 28.12, página 1551) que $\lambda u + (1-\lambda)v$ é elemento de $\overline{C}$, completando a prova. ■

• **Intersecções de convexos são convexos**

A seguinte observação básica sobre conjuntos convexos é muito relevante.

Lema 5.2 *Se $\{C_\mu, \mu \in \Omega\}$ for uma família de conjuntos convexos em um mesmo espaço vetorial real $\mathcal{V}$, então $\bigcap_{\mu\in\Omega} C_\mu$ é igualmente um subconjunto convexo de $\mathcal{V}$.* □

Prova. Se $\bigcap_{\mu\in\Omega} C_\mu$ for vazio, não há o que se provar. Se não for vazio, sejam x e y elementos arbitrários de $\bigcap_{\mu\in\Lambda} C_\mu$. Então, ambos pertencem a cada C_μ, com $\mu \in \Omega$. Como cada C_μ é convexo, concluímos que para cada $\lambda \in [0, 1]$ o vetor $\lambda x + (1-\lambda)y$ é igualmente elemento de cada C_μ, com $\mu \in \Omega$ e, portanto, é um elemento de $\bigcap_{\mu\in\Omega} C_\mu$, provando a afirmação. ■

• **Envoltória convexa e varredura convexa**

Seja $\mathcal{V}$ um espaço vetorial real e seja $A \subset \mathcal{V}$, não vazio. O conjunto A pode não ser convexo, mas é subconjunto de ao menos um conjunto convexo, a saber, do próprio $\mathcal{V}$. Denotamos por $\mathrm{ccl}(A)$ a intersecção de todos os subconjuntos convexos de $\mathcal{V}$ que contêm A. Honorificamente, declaramos também que $\mathrm{ccl}(\emptyset) = \emptyset$. Pelo Lema 5.2, $\mathrm{ccl}(A)$ é convexo para qualquer $A \subset \mathcal{V}$.

O conjunto $\mathrm{ccl}(A)$ é denominado *envoltória convexa*, ou *invólucro convexo*, ou ainda *fecho convexo*[2] de A. Também se diz que $\mathrm{ccl}(A)$ é o *conjunto convexo gerado* por A. Trata-se do "menor" conjunto convexo de $\mathcal{V}$ que contém A.

[2]Em inglês diz-se *convex hull*, *convex envelope* ou *convex closure*.

Definimos a *varredura convexa* de A como o conjunto de todas as combinações lineares convexas finitas de elementos de A, ou seja,

$$\mathrm{cspan}(A) := \left\{\sum_{k=1}^{m}\lambda_k a_k, \text{ para algum } m \in \mathbb{N} \text{ e } a_1, \ldots, a_m \in A, \text{ sendo } \lambda_k \in [0, 1] \text{ e } \sum_{k=1}^{m}\lambda_k = 1\right\}.$$

$\mathrm{cspan}(A)$ é um conjunto convexo, o que pode ser visto da seguinte forma: se $u = \sum_{k=1}^{m}\lambda_k a_k$ e $u' = \sum_{l=1}^{m'}\lambda'_l a'_l$ são dois elementos de $\mathrm{cspan}(A)$ e $\mu \in [0, 1]$, então

$$\mu u + (1-\mu)u' = \sum_{k=1}^{m}\mu\lambda_k a_k + \sum_{l=1}^{m'}(1-\mu)\lambda'_l a'_l ,$$

que é uma combinação linear convexa de finitos elementos de A, pois, claramente,

$$\sum_{k=1}^{m}\mu\lambda_k + \sum_{l=1}^{m'}(1-\mu)\lambda'_l = \mu\underbrace{\sum_{k=1}^{m}\lambda_k}_{=1} + (1-\mu)\underbrace{\sum_{l=1}^{m'}\lambda'_l}_{=1} = \mu + (1-\mu) = 1 ,$$

sendo que, claramente, todos os números $\mu\lambda_k$ e $(1-\mu)\lambda_l$ são elementos de $[0, 1]$.

É um exercício simples provar que $\mathrm{ccl}(A) = \mathrm{cspan}(A)$. De fato, $\mathrm{cspan}(A)$ é convexo e obviamente contém A, estabelecendo que $\mathrm{ccl}(A) \subset \mathrm{cspan}(A)$, pela definição de $\mathrm{ccl}(A)$. Além disso, como $\mathrm{ccl}(A)$ é convexo e contém A, ele contém também, pela Proposição 5.1, página 336, todas as combinações lineares convexas finitas de elementos de A, que é o conjunto $\mathrm{cspan}(A)$, estabelecendo que $\mathrm{ccl}(A) \supset \mathrm{cspan}(A)$ e provando a identidade desejada. Para futura referência capturamos essa afirmação em uma proposição:

Proposição 5.3 *Seja $\mathcal{V}$ um espaço vetorial real e seja A um subconjunto não vazio de $\mathcal{V}$. Então, a envoltória convexa e a varredura convexa de A coincidem:* $\mathrm{ccl}(A) = \mathrm{cspan}(A)$. □

5.1.2 Um Exemplo. Células e Diagramas de Voronoy

O Lema 5.2, página 337, é um excelente pretexto para falarmos sobre alguns objetos matematicamente interessantes e curiosos, e com diversas aplicações atuais: a chamadas *células e diagramas de Voronoy*[3].

Essas noções, que definiremos adiante, encontram usos e aplicações em uma grande variadade de áreas, tais como Aeronáutica (identificação de aeroportos mais próximos a um dado ponto. Vide Fig. 5.2, página 343), Antropologia, Arqueologia (determinação de áreas de influência de civilizações), Arquitetura, Astronomia, Biologia (Fisiologia, Epidemiologia, Zoologia, Ecologia, Forestamento), Cartografia, Computação Gráfica, Cristalografia, Estatística e Análise de Dados, Geofísica, Geografia, Geologia, Marketing, Matemática (Análise de Elementos Finitos, estudo de formas quadráticas, reconhecimento de padrões, modelamento geométrico via triangularizações de superfícies), Metalurgia, Meteorologia, Química, Robótica e Telefonia Celular (identificação de torres de transmissão mais próximas a um dado aparelho).

Células de Voronoy foram definidas e estudadas por esse autor em 1908, em trabalhos sobre propriedades de formas quadráticas, mas surgiram muito antes, tendo sido identificadas em trabalhos de Descartes[4] e Dirichlet[5]. Os trabalhos originais de Voronoy sobre o assunto encontram-se listados em [554].

No que segue, apresentaremos as definições das noções de célula e diagrama de Voronoy no contexto mais geral, em espaços métricos (noção desenvolvida no Capítulo 25, página 1410), para em seguida discutirmos casos mais específicos, e de maior interesse, em espaços vetoriais normados ou eventualmente dotados de um produto escalar.

- **Fatos preliminares**

Seja M um conjunto não vazio dotado de uma métrica d. Para $p,\ q \in M$, distintos, definimos os conjuntos

$$A(p,\ q) := \left\{x \in M \middle|\ d(x,\ p) < d(x,\ q)\right\} \quad \text{e} \quad F(p,\ q) := \left\{x \in M \middle|\ d(x,\ p) \leq d(x,\ q)\right\}. \tag{5.3}$$

[3] Georgy Feodosevich Voronoy (1868–1908). Seu nome é também transcrito como Georgii Feodosevich Voronoi.
[4] René Descartes (1596–1650).
[5] Johann Peter Gustav Lejeune Dirichlet (1805–1859).

Em palavras, $A(p,\ q)$ é a coleção de todos os pontos de M que estão estritamene mais próximos (segundo d) de p do que de q e $F(p,\ q)$ é a coleção de todos os pontos de M que estão mais próximos, ou igualmente próximos, de p do que de q.

É fácil ver que $F(p,\ q) = A(q,\ p)^c \equiv M \setminus A(q,\ p)$ para todos $p \neq q$.

Uma observação que nos será relevante é a seguinte:

Lema 5.3 *Os conjuntos $A(p,\ q)$, para $p \neq q$, são d-abertos. Portanto, os conjuntos $F(p,\ q)$ são d-fechados, por serem complementares de d-abertos.* □

Prova. Seja $x \in A(p,\ q)$. Por hipótese, tem-se que $d(x,\ q) - d(x,\ p) > 0$. Considere-se $r > 0$ tal que

$$r \ < \ \frac{d(x,\ q) - d(x,\ p)}{2} \tag{5.4}$$

e considere-se a bola aberta de raio r centrada em x, que denotamos por $B(x,\ r)$. Afirmamos que $B(x,\ r) \subset A(p,\ q)$. Para tal, tomemos $y \in B(x,\ r)$, arbitrário, e consideremos a diferença $d(y,\ q) - d(y,\ p)$. Sabemos que $d(y,\ p) \leq d(y,\ x) + d(x,\ p)$ e sabemos que $d(x,\ q) \leq d(y,\ q) + d(x,\ y)$, o que implica em $d(y,\ q) \geq d(x,\ q) - d(x,\ y)$. Assim,

$$d(y,\ q) - d(y,\ p) \ \geq \ d(x,\ q) - d(x,\ y) - d(y,\ x) - d(x,\ p) \ = \ d(x,\ q) - d(x,\ p) - 2d(x,\ y) \ > \ d(x,\ q) - d(x,\ p) - 2r \ \stackrel{(5.4)}{>} \ 0 \ .$$

Assim, $d(y,\ p) < d(y,\ q)$, o que significa que $y \in A(p,\ q)$, provando que $B(x,\ r) \subset A(p,\ q)$ e, portanto, que os conjuntos $A(p,\ q)$, para $p \neq q$, são d-abertos. Os conjuntos $F(p,\ q)$ são d-fechados pois $F(p,\ q) = A(q,\ p)^c$. ■

É um exercício fácil mostrar que $A(p,\ q)$ é o interior de $F(p,\ q)$ e que o bordo de $A(p,\ q)$ e de $F(p,\ q)$ é $E(p,\ q) := \big\{x \in M \big| d(x,\ p) = d(x,\ q)\big\}$, o conjunto de pontos equidistantes a p e q.

• Células de Voronoy. Definição e propriedades elementares

Os ingredientes básicos ligados às noções de célula de Voronoy e diagrama de Voronoy são um conjunto M dotado de uma métrica d e um subconjunto $\mathcal{P} \subset M$ de pontos distintos de M. O conjunto $\mathcal{P}$ pode ser finito ou não, contável ou não. Normalmente assume-se que M é completo em relação a d, mas essa hipótese pode ser acrescentada quando necessário.

Definimos a *célula de Voronoy* associada a um particular ponto $q \in \mathcal{P}$, denotada por V_q, como o conjunto de todos os pontos de M cuja distância (segundo d) a q é menor ou igual à distância a todos os demais ponto de $\mathcal{P}$. Formalmente, para $q \in \mathcal{P}$,

$$V_q \ := \ \Big\{x \in M \Big| \, d(x,\ q) \leq d(x,\ p) \text{ para todo } p \in \mathcal{P},\ p \neq q \ \Big\} \ . \tag{5.5}$$

É evidente por essa definição que

$$V_q \ = \ \bigcap_{\substack{p \in \mathcal{P} \\ p \neq q}} F(q,\ p) \ . \tag{5.6}$$

Vide (5.3). Como os conjuntos $F(q,\ p)$ são d-fechados, concluímos imediatamente que as células de Voronoy V_q sempre são subconjuntos d-fechados de M. Mais adiante, faremos uso de (5.6) para demonstrar a convexidade de células de Voronoy em certos casos e ilustraremos o significado de (5.6) no caso em que $M = \mathbb{R}^2$ com a métrica Euclidiana usual, uma situação relevante em aplicações.

O interior de uma célula de Voronoy V_q, denotado por V_q^0, é dado por,

$$V_q^0 \ := \ \Big\{x \in M \Big| \, d(x,\ q) < d(x,\ p) \text{ para todo } p \in \mathcal{P} \ \Big\} \ . \tag{5.7}$$

O interior V_q^0 de uma célula de Voronoy pode ser vazio.

É evidente que cada ponto $q \in \mathcal{P}$ pertence à própria célula V_q.

É também fácil ver que se p e q são elementos distintos de $\mathcal{P}$, então $V_p^0 \cap V_q^0 = \emptyset$, ou seja, que células de Voronoy distintas são *interiormente disjuntas*. Para ver isso, notemos que se $z \in V_p^0 \cap V_q^0$ teríamos, do fato que $z \in V_p^0$ que $d(z,\ p) < d(z,\ q)$ e do fato que $z \in V_q^0$, que $d(z,\ q) < d(z,\ p)$, o que é impossível.

Se p e q são elementos distintos de $\mathcal{P}$ é possível, porém, ter-se $V_p \cap V_q \neq \emptyset$, em cujo caso dizemos que as células V_p e V_q são *células adjacentes*. Caso não vazio, o conjunto $V_p \cap V_q$ é dito ser a *aresta comum* entre as células V_p e V_q.

Note-se que uma aresta comum $V_p \cap V_q$ é sempre um subconjunto fechado de M.

• Diagramas e grafos de Voronoy

Dado um espaço métrico $(M,\ d)$ e um conjunto não vazio $\mathcal{P} \subset M$, a coleção $\{V_p,\ p \in \mathcal{P}\}$ de células de Voronoy é dita ser um *diagrama de Voronoy* associado à coleção de pontos $\mathcal{P}$.

Alguns autores usam a expressão "diagrama de Voronoy" para designar a união disjunta[6]

$$\bigsqcup_{p\in\mathcal{P}} V_p \ := \ \bigcup_{p\in\mathcal{P}} (V_p,\ p) \ .$$

Diagramas de Voronoy são também denominados *tesselações de Voronoy* ou *partições de Voronoy* ou ainda *decomposições de Voronoy*. Algumas áreas específicas têm denominações próprias para esses objetos.

A união de todas as arestas comuns pertencentes a um diagrama de Voronoy fechado, ou seja, o conjunto

$$\bigcup_{\substack{p,\ q\in\mathcal{P}\\ q\neq p}} V_p \cap V_q \ ,$$

é dito ser um *grafo de Voronoy* associado à coleção de pontos $\mathcal{P}$.

• Generalizações

Comentamos de passagem que células, diagramas e grafos de Voronoy podem ser similarmente definidas se $\mathcal{P}$ não for apenas uma coleção de pontos distintos de M, mas uma coleção de subconjuntos disjuntos de M. Não trataremos dessas generalizações, mas apontamos que também há aplicações para as mesmas, como na determinação de bacias hidrográficas na Geografia e Climatologia, onde procura-se determinar os pontos mais próximos a rios levando-se em conta as linhas de cumeada do relevo. Vide *e.g.*, [350] para um estudo na bacia do Alto Sapucaí, na região de Itajubá-MG.

• Diagramas de Voronoy como partições de M

Há diversas questões básicas associadas às noções de diagrama de Voronoy e grafo de Voronoy e iremos no que segue discutir algumas das mais relevantes.

Uma primeira questão importante que se coloca é a seguinte: dado um diagrama de Voronoy $\{V_p,\ p \in \mathcal{P}\}$ associado a uma coleção de pontos $\mathcal{P} \subset M$, quando valerá que $\bigcup_{p\in\mathcal{P}} V_p = M$? Em outras palavras, quando a união de células de Voronoy de um diagrama de Voronoy forma uma partição de M?

Para verificar que essa é uma questão não trivial, notemos que nem sempre ocorre de termos $\bigcup_{p\in\mathcal{P}} V_p$ igual a M.

Exemplo 5.4 No caso geral em que $\mathcal{P}$ não é fechado (na topologia induzida pela métrica d em M) mas não necessariamente finito, nem sempre ocorre de que todo $x \in M$ pertence a alguma célula de Voronoy fechada.

Suponhamos que $\overline{\mathcal{P}} \setminus \mathcal{P}$ seja não vazio e seja $x \in \overline{\mathcal{P}} \setminus \mathcal{P}$. Então, $\inf\big\{d(x,\ p),\ p \in \mathcal{P}\big\} = 0$ e para cada $q \in \mathcal{P}$ existirá ao menos um $p \in \mathcal{P}$, $p \neq q$, tal que $d(x,\ q) > d(x,\ p)$, mostrando que x não pode pertencer a nenhuma célula V_q.

Um caso a se ter em mente. Tome-se $M = \mathbb{R}$ com a métrica usual e $\mathcal{P} = (0,\ \infty)$. Nenhum ponto de $(-\infty,\ 0]$ pertence a alguma célula de Voronoy. Tem-se nesse caso que cada célula de Voronoy contém um único ponto: $V_p = \{p\}$, com $p \in (0,\ \infty)$. ♦

E. 5.2 *Exercício*. Tome-se $M = \mathbb{R}$ com a métrica usual e $\mathcal{P} = [0,\ \infty)$. Mostre que, nesse caso, as células de Voronoy são $V_p = \{p\}$, para $p \in (0,\ \infty)$ e $V_0 = (\infty,\ 0]$. Assim, $\bigcup_{p\in\mathcal{P}} V_p = \mathbb{R}$. ✠

E. 5.3 *Exercício*. Tome-se $M = \mathbb{R}^2$ com a métrica usual e $\mathcal{P} = \{(x,\ 0) \in \mathbb{R}^2,\ \text{com } x \in [0,\ \infty)\}$. Determine as células de Voronoy nesse caso e verifique se $\bigcup_{p\in\mathcal{P}} V_p = \mathbb{R}^2$. Faça o mesmo para o caso em que $\mathcal{P} = \{(x,\ 0) \in \mathbb{R}^2,\ \text{com } x \in (0,\ \infty)\}$. ✠

É importante notar, porém, que no caso geral a condição de $\mathcal{P}$ ser fechado não é suficiente para que o correspondente diagrama de Voronoy seja uma partição de M.

[6] No sentido da definição (1.55), da página 70.

Vamos listar algumas condições sob as quais possamos garantir que $\bigcup_{p\in\mathcal{P}} V_p = M$, ou seja, ou seja, sob as quais um diagrama de Voronoy seja uma partição de M:

1. Caso $\mathcal{P} = M$. Nessa situação trivial, tem-se $V_p = \{p\}$ para cada $p \in M$ e, portanto, o correspondente diagrama de Voronoy é uma partição de M.

2. Caso $\mathcal{P}$ for um conjunto finito. Se $x \in M$, então o conjunto numérico $\{d(x,\ p),\ p \in \mathcal{P}\}$ é finito e, portanto, possui um mínimo, digamos $d(x,\ q)$, para algum $q \in \mathcal{P}$ (não necessariamente único). Logo, para esse q valerá $d(x,\ q) \leq d(x,\ p)$ para todo $p \in \mathcal{P}$, mostrando que $x \in V_q$. Como isso vale para cada $x \in M$, concluímos que todo elemento de M pertence a ao menos uma célula de Voronoy e, portanto, o diagrama de Voronoy associado a $\mathcal{P}$ é uma partição de M.

3. Caso M for um espaço de Hilbert, com a métrica d definida pela norma de M, e $\mathcal{P}$ for um conjunto fechado e convexo de M. Nesse caso, o Teorema do Melhor Aproximante em espaços de Hilbert, Teorema 41.1, página 2292, garante que para cada $x \in M$ existe um único $q \in \mathcal{P}$ tal que $\|x-q\| = \inf\big\{\|x-p\|,\ p \in \mathcal{P}\big\}$. A partir daqui podemos argumentar como no caso em que $\mathcal{P}$ é finito: para esse x valerá $d(x,\ q) \leq d(x,\ p)$ para todo $p \in \mathcal{P}$, mostrando que $x \in V_q$. Como isso vale para cada $x \in M$, concluímos que todo elemento de M pertence a alguma célula de Voronoy e, portanto, o diagrama de Voronoy associado a $\mathcal{P}$ é uma partição de M.

4. A situação anterior pode ser facilmente generalizada para o caso em que M é um espaço vetorial normado uniformemente convexo[7] e completo e $\mathcal{P}$ é um conjunto fechado e convexo de M. Pelo Teorema do Melhor Aproximante, Teorema 25.6, página 1458, vale a mesma conclusão que apresentamos para espaços de Hilbert: para cada $x \in M$ existe um único $q \in \mathcal{P}$ tal que $\|x-q\| = \inf\big\{\|x-p\|,\ p \in \mathcal{P}\big\}$. O restante da argumentação é idêntico.

• Células de Voronoy convexas

Vamos agora particularizar nossa exposição considerando a situação em que M é um espaço vetorial (real ou complexo) cuja métrica é definida por uma norma proveniente de um produto escalar. Essa situação é muito frequente em aplicações e nela as células de Voronoy desfrutam de uma propriedade muito importante: são subconjuntos convexos de M.

Seja V um espaço vetorial (real ou complexo) dotado de um produto escalar, que denotamos por $\langle\cdot,\ \cdot\rangle$. Seja $\|\cdot\|$ a norma associada a esse produto escalar. Sejam p, q pontos distintos de V e seja, como antes, $F(p,\ q) = \big\{x \in V \big|\ \|x-p\| \leq \|x-q\|\big\}$. A condição $\|x-p\| \leq \|x-q\|$ equivale à condição $\|x-q\|^2 - \|x-p\|^2 \geq 0$. Agora,

$$\|x-q\|^2 - \|x-p\|^2 \;=\; \langle x-q,\ x-q\rangle - \langle x-p,\ x-p\rangle \;=\; \|q\|^2 - \|p\|^2 - 2\mathrm{Re}\big(\langle q-p,\ x\rangle\big)\ . \tag{5.8}$$

(Se o espaço V for real a símbolo de parte real, acima, é desnecessário). Assim, $x \in F(p,\ q)$ se e somente se

$$\|q\|^2 - \|p\|^2 - 2\mathrm{Re}\big(\langle q-p,\ x\rangle\big) \;\geq\; 0\ . \tag{5.9}$$

Vamos supor que $x_0,\ x_1 \in F(p,\ q)$, ou seja, satisfazem (5.9). Considere-se $x_\lambda := (1-\lambda)x_0 + \lambda x_1$, com $\lambda \in [0,\ 1]$. Teremos,

$$\begin{aligned}\|q\|^2 - \|p\|^2 - 2\mathrm{Re}\big(\langle q-p,\ x_\lambda\rangle\big) \;&=\; \|q\|^2 - \|p\|^2 - 2(1-\lambda)\mathrm{Re}\big(\langle q-p,\ x_0\rangle\big) - 2\lambda\mathrm{Re}\big(\langle q-p,\ x_1\rangle\big)\\ &=\; (1-\lambda)\Big(\|q\|^2 - \|p\|^2 - 2\mathrm{Re}\big(\langle q-p,\ x_0\rangle\big)\Big) + \lambda\Big(\|q\|^2 - \|p\|^2 - 2\mathrm{Re}\big(\langle q-p,\ x_1\rangle\big)\Big) \;\geq\; 0\ ,\end{aligned}$$

pois, por hipótese, x_0 e x_1 satisfazem (5.9). Isso provou que x_λ também satisfaz (5.9) e, portanto, $x_\lambda \in F(p,\ q)$ para todo $\lambda \in [0,\ 1]$ e quaisquer $x_0,\ x_1 \in F(p,\ q)$, demonstrando que $F(p,\ q)$ é um conjunto convexo.

Do fato de os conjuntos $F(p,\ q)$ serem convexos, segue de (5.6) que as células de Voronoy são também subconjuntos convexos de V, por serem intersecções de conjuntos convexos. Vide Lema 5.2, página 337. Note que essa conclusão independe da cardinalidade de $\mathcal{P}$.

No relevante caso em que $\mathcal{P}$ é um conjunto finito, cada célula V_q será uma intersecção finita de semiespaços $F(q,\ p)$. Nessa situação, se V for um espaço vetorial real de dimensão finita, como $\mathbb{R}^n$ (outro caso relevante em aplicações), as

[7] Para essa noção, vide Seção 5.1.3, página 342 e Seção 25.6, página 1456.

células de Voronoy serão politopos fechados convexos, ou polígonos fechados convexos no caso bidimensional, mas não são necessariamente compactos (*i.e.*, limitados). Por razões históricas, células de Voronoy que sejam polígonos fechados convexos são também denominadas *polígonos de Thiessen*[8] em certas áreas, como Geofísica e Meteorologia.

Chamamos a atenção do leitor para o fato que, no caso de um espaço vetorial normado cuja norma não provenha de um produto escalar, as células de Voronoy **não** são necessariamente conjuntos convexos, mesmo V tendo dimensão finita e mesmo $\mathcal{P}$ sendo finito. Vide Figura 5.1, página 342, para um exemplo ilustrado.

• **Ilustrações**

A Figura 5.1, página 342, exibe o diagrama de Voronoy no plano de um mesmo conjunto de pontos, para duas métricas diferentes, a métrica Euclidiana e a métrica d_1.

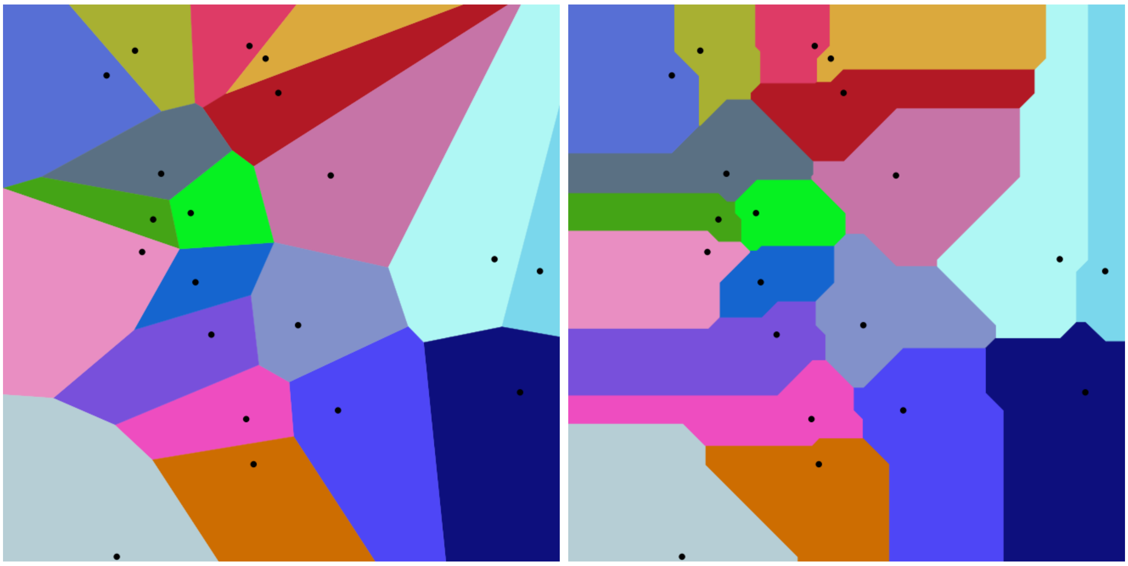

Figura 5.1: À esquerda, o diagrama de Voronoy do conjunto de pontos em preto, segundo a métrica Euclidiana em $\mathbb{R}^2$: $d_2\big((x_1,\, x_2),\, (y_1,\, y_2)\big) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2}$. À direita, o diagrama de Voronoy do mesmo conjunto de pontos, mas segundo a métrica $d_1\big((x_1,\, x_2),\, (y_1,\, y_2)\big) = |y_1 - x_1| + |y_2 - x_2|$. Observe-se que as células de Voronoy do lado esquerdo são convexas, mas as do lado direito não são.
Fonte das figuras: Balu Ertl - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=38534275 e CC BY-SA 1.0, https://commons.wikimedia.org/w/index.php?curid=38535751.

A Figura 5.2, página 343, mostra o diagrama de Voronoy associado a um conjunto de aeroportos dos EUA, segundo a métrica Euclidiana usual. A utilidade desse tipo de diagrama é clara: esse diagrama de Voronoy permite identificar, em caso de emergência, o aeroporto mais próximo a um avião em curso.

• **Algoritmos**

Um dos assuntos mais interessantes e de utilidade prática ligados a diagramas de Voronoy, mas que não discutiremos aqui, é o desenvolvimento de algoritmos para determinar esses diagramas em diferentes circunstâncias. São conhecidos centenas desses algoritmos. São eles que permitem produzir diagramas de Voronoy como os das Figuras 5.1 e 5.2. Remetemos os interessados à vasta lista de referência sobre o assunto, encontrável especialmente na internet.

5.1.3 Espaços Estritamente Convexos e Uniformemente Convexos

Seja $\mathcal{V}$ um espaço vetorial real dotado de uma topologia τ. Dizemos que um conjunto convexo $C \subset \mathcal{V}$ é um *conjunto estritamente convexo* segundo τ se para quaisquer $x,\ y \in C$ o segmento que conecta x a y, excetuando esses dois pontos,

[8] Alfred Henry Thiessen (1872–1956), meteorologista e militar. Artigo original: [536].

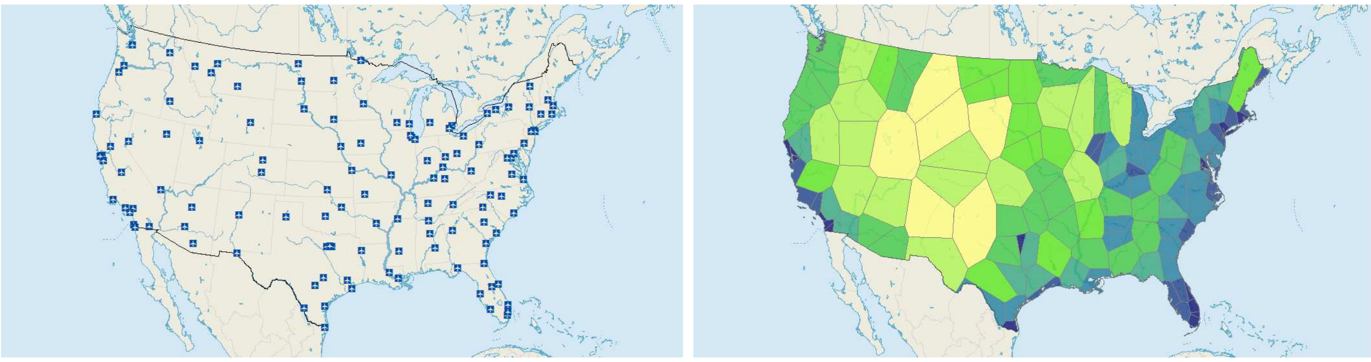

Figura 5.2: À esquerda, uma lista de Aeroportos nos EUA (que seria o conjunto $\mathcal{P}$). À direita, o correspondente diagrama de Voronoy, com a medida usual de distância.
Fonte: GIS Geography. Um mapa interativo desse tipo, em escala mundial, pode ser encontrado em https://www.jasondavies.com/maps/voronoi/airports/

estiver contido no interior de C:

$$\text{para todos } x,\ y \in C \text{ tem-se} \quad \big\{\lambda x + (1-\lambda)y,\ \lambda \in (0,\ 1)\big\} \subset C^0 .$$

Recordar que C^0, o interior de C segundo τ é o "maior" τ-aberto contido em C. Vide Seção 28.3, página 1543.

Para $x,\ y \in \mathcal{V}$ denominaremos $\big\{\lambda x+(1-\lambda)y,\ \lambda \in (0,\ 1)\big\}$ a *corda aberta* conectando x a y e $\big\{\lambda x+(1-\lambda)y,\ \lambda \in [0,\ 1]\big\}$ a *corda fechada* conectando x a y.

Uma situação de interesse particular é aquela em que $\mathcal{V}$ é dotado de uma norma $\|\cdot\|$. De posse de uma norma, $\mathcal{V}$ é um automaticamente um espaço métrico. Há nesse contexto alguns desenvolvimentos da noção acima de conjunto estritamente convexo. Começamos definindo a bola unitária *aberta* centrada em $x \in \mathcal{V}$ por

$$\mathcal{B}_1(x) \ := \ \big\{y \in \mathcal{V} |\ \|y - x\| \ < \ 1\big\} .$$

A bola unitária *fechada* centrada em $x \in \mathcal{V}$ é definida por

$$\overline{\mathcal{B}_1}(x) \ := \ \big\{y \in \mathcal{V} |\ \|y - x\| \ \leq \ 1\big\} .$$

Denotamos também $\mathcal{B}_1 \equiv \mathcal{B}_1(0)$, $\overline{\mathcal{B}_1} \equiv \overline{\mathcal{B}_1}(0)$, as bolas unitárias aberta e fechada, respectivamente, centradas em 0.

É um exercício simples constatar que $\mathcal{B}_1 = \left(\overline{\mathcal{B}_1}\right)^0 = (\mathcal{B}_1)^0$. O fronteira[9], ou bordo, de $\mathcal{B}_1$ é, por definição, conjunto $\partial\mathcal{B}_1 := \overline{\mathcal{B}_1} \setminus (\mathcal{B}_1)^0 = \overline{\mathcal{B}_1} \setminus \mathcal{B}_1$ e, portanto, $\partial\mathcal{B}_1 = \big\{y \in \mathcal{V} |\ \|y\| \ = \ 1\big\}$.

• Espaços normados estritamente convexos

Um espaço vetorial $\mathcal{V}$ dotado de uma norma $\|\cdot\|$ é dito ser um *espaço estritamente convexo* se $\overline{\mathcal{B}_1}$ for um conjunto estritamente convexo, ou seja, se para todo $x,\ y \in \mathcal{V}$ com $\|x\| \leq 1$ e $\|y\| \leq 1$ e todo $\lambda \in (0,\ 1)$ valer $\big\|\lambda x + (1-\lambda)y\big\| < 1$.

Assim, em um espaço estritamente convexo, a corda aberta conectando dois pontos quaisquer $x,\ y \in \overline{\mathcal{B}_1}$ está inteiramente em $\mathcal{B}_1$. Vide Figura 5.3, página 344.

• Espaços normados uniformemente convexos

Uma outra noção aparentada é a de *espaço uniformemente convexo*[10]. Um espaço vetorial $\mathcal{V}$ dotado de uma norma $\|\cdot\|$ é dito ser um *espaço uniformemente convexo* se para todo $0 < \epsilon \leq 2$ existir $\delta \equiv \delta(\epsilon) > 0$ tal que se $x,\ y \in \overline{\mathcal{B}_1}$ e $\|x - y\| \geq \epsilon$, então $\big\|(x+y)/2\big\| \leq 1 - \delta(\epsilon)$.

Vamos esclarecer o significado disso. O ponto $(x+y)/2$ é o ponto intermediário entre x e y. Como $x,\ y \in \overline{\mathcal{B}_1}$, temos sempre $\big\|(x+y)/2\big\| \leq \big(\|x\| + \|y\|\big)/2 \leq 1$ e, portanto, $(x+y)/2 \in \overline{\mathcal{B}_1}$. Assim, $1 - \delta(\epsilon) \leq \big\|(x+y)/2\big\| \leq 1$. Portanto, se

[9] Para a definição, vide (28.6), página 1546.

[10] Essa noção foi introduzida por James Andrew Clarkson (1906-1970) em: James A. Clarkson, "Uniformly convex spaces", Trans. Amer. Math. Soc. **40**, 396–414 (1936). DOI: https://doi.org/10.1090/S0002-9947-1936-1501880-4

movermos os dois pontos x e y dentro de $\overline{\mathcal{B}_1}$ de sorte que o ponto intermediário for se aproximando da borda, ou seja, se $1 - \big\|(x+y)/2\big\|$ for diminuindo, então $\|x-y\|$ não pode permanecer maior que um dado $\epsilon_0 > 0$, pois para esse ϵ_0 teríamos $1 - \big\|(x+y)/2\big\| \geq \delta(\epsilon_0)$, cotrariando a ideia que $1 - \big\|(x+y)/2\big\|$ diminui.

Isso significa que se $x, y \in \overline{\mathcal{B}_1}$ e $\big\|(x+y)/2\big\| \to 1$, então $\|x-y\| \to 0$. Falando de forma intuitiva, se o ponto intermediário de dois pontos x e $y \in \overline{\mathcal{B}_1}$ aproxima-se da borda, os pontos x e y devem aproximar-se um do outro.

Essas ideias nos fazem entender intuitivamente que um espaço uniformemente convexo é um espaço vetorial normado em que a bola unitária fechada é completamente "redonda", ou seja, não possui faces planas. Uma face plana é um subconjunto $\mathcal{F}$ de pontos de $\partial\overline{\mathcal{B}_1}$ tais que para todos $x,\ y \in \mathcal{F}$ vale também $(x+y)/2 \in \mathcal{F}$.

Para exemplificar essas ideias, contemplemos a Figura 5.3, página 344. Vemos na figura da direita que, na presença de faces planas, pode ocorrer de x, y e seu ponto intermediário localizarem-se na borda $\partial\mathcal{B}_1$ sem que a distância entre x e y tenha de ser "pequena". Tal não ocorre, porém, na figura da esquerda.

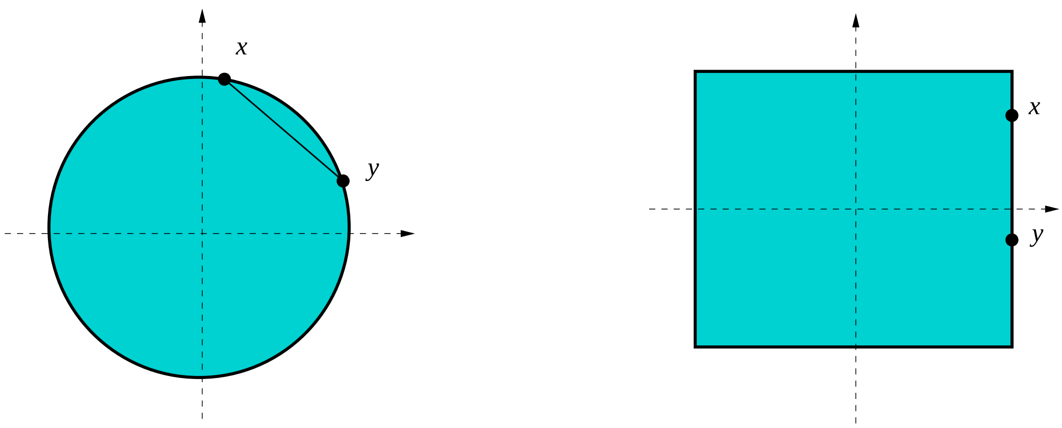

Figura 5.3: Dois conjuntos convexos em $\mathbb{R}^2$. A figura da esquerda (o disco de raio 1 centrado na origem) é a bola de raio 1 na norma euclidiana $\|(x_1,\, x_2)\|_2 = \sqrt{(x_1)^2 + (x_2)^2}$ e a figura da direita (o quadrado de lado 2 centrado na origem) é a bola de raio 1 na norma $\|(x_1,\, x_2)\|_\infty = \max\big\{|x_1|,\ |x_2|\big\}$. A bola da esquerda é uniformemente convexa na norma $\|\cdot\|_2$ e, portanto, é um conjunto estritamente convexo. A bola da direita não é uniformemente convexa na norma $\|\cdot\|_\infty$. Nela a corda aberta conectando os pontos x e y encontra-se no interior do disco. Na figura da direita, a corda aberta conectando os pontos x e y não entra no interior do quadrado. Trata-se, portanto, de um conjunto que não é estritamente convexo.

Na Seção 25.6, página 1456, estabeleceremos um importante resultado sobre espaços uniformemente convexos: um teorema de melhor aproximante, Teorema 25.6, página 1458, o qual tem usos importantes.

Para fechar a presente discussão mostremos que espaços normados uniformemente convexo são estritamente convexos.

Proposição 5.4 *Se um espaço vetorial real $\mathcal{V}$, dotado de uma norma $\|\cdot\|$, é uniformemente convexo nessa norma, então é estritamente convexo na topologia associada a essa norma.* □

Prova. Se $x, y \in \overline{\mathcal{B}_1}$, então para $\lambda \in (0, 1)$ tem-se $\big\|\lambda x + (1-\lambda)y\big\| \leq \lambda\|x\| + (1-\lambda)\|y\|$, que é uma combinação linear comvexa de $\|x\|$ e $\|y\|$. Assim, se $\|x\|$ ou $\|y\|$ são estritamente menores que 1, os vetores $\lambda x + (1-\lambda)y$ também têm norma estritamente menor que 1 e, portanto, estão em $\mathcal{B}_1$, o interior de $\overline{\mathcal{B}_1}$. O caso restante é aquele em que $\|x\| = \|y\| = 1$.

Como $\overline{\mathcal{B}_1}$ é uniformemente convexo, e $\|x-y\| \leq 2$ (pois ambos estão em$\overline{\mathcal{B}_1}$), existe $\delta > 0$ tal que $\big\|(x+y)/2\big\| \leq 1-\delta < 1$, mostrando que o ponto intermediário $z_1 \equiv (x+y)/2$ está em $\mathcal{B}_1$ e, portanto, tem norma estritamente menor que 1.

Pelo argumento anterior, os pontos intermediários entre x e $(x+y)/2$ e entre y e $(x+y)/2$ também estão em $\mathcal{B}_1$. Esses pontos são $z_2 \equiv \frac{3}{4}x + \frac{1}{4}y$ e $z_3 \equiv \frac{1}{4}x + \frac{3}{4}y$. Essa argumento pode ser repetido para o ponto intermediário entre x e z_2, z_2 e z_1, z_1 e z_3 e z_3 e y, resultando em mais pontos intermediários, todos em $\mathcal{B}_1$.

Prosseguindo por indução, esse argumento pode ser repetido um conjunto enumerável de vezes e informa que há um conjunto enumerável denso na corda aberta $\{\lambda x + (1-\lambda)y,\ \lambda \in (0, 1)\}$, especificamente, para todos os λ da forma

$$\lambda_{n,\,k} = \frac{k}{2^n} \quad \text{com} \quad k \in \{0,\ 1\ \ldots,\ 2^n\}\,, \quad n \in \mathbb{N}\,,$$

que está inteiramente dentro de $\mathcal{B}_1$. Pela continuidade da norma concluímos que todos os pontos dessa corda aberta estão em $\mathcal{B}_1$, completando a demonstração. ■

5.2 Funções Convexas e Côncavas em Espaços Vetoriais Reais

• **Funções convexas e côncavas. Definições e Propriedades Básicas**

Seja $\mathcal{V}$ um espaço vetorial real e seja $C \subset \mathcal{V}$ um conjunto convexo. Uma função $f : C \to \mathbb{R}$ é dita ser uma *função convexa* se para todos $x, \ y \in C$ e todo $\lambda \in [0, \ 1]$ valer a desigualdade

$$f\big(\lambda x + (1-\lambda)y\big) \ \leq \ \lambda f(x) + (1-\lambda) f(y) \ . \tag{5.10}$$

Uma função $f : C \to \mathbb{R}$ é dita ser uma *função côncava* se para todos $x, \ y \in C$ e todo $\lambda \in [0, \ 1]$ valer a desigualdade

$$f\big(\lambda x + (1-\lambda)y\big) \ \geq \ \lambda f(x) + (1-\lambda) f(y) \ . \tag{5.11}$$

Uma função $f : C \to \mathbb{R}$ é dita ser uma *função estritamente convexa* se para todos $x, \ y \in C$ com $x \neq y$ e todo $\lambda \in (0, \ 1)$ valer $f\big(\lambda x + (1-\lambda)y\big) < \lambda f(x) + (1-\lambda) f(y)$. Uma função $f : C \to \mathbb{R}$ é dita ser uma *função estritamente côncava* se para todos $x, \ y \in C$ com $x \neq y$ e todo $\lambda \in (0, \ 1)$ valer $f\big(\lambda x + (1-\lambda)y\big) > \lambda f(x) + (1-\lambda) f(y)$.

É elementar constatar que uma função f é côncava se e somente se $-f$ for convexa. Com isso, propriedades de funções côncavas podem ser facilmente derivadas de propriedades correspondentes de funções convexas e, por isso, discutiremos majoritariamente as últimas. O mesmo vale para funções estritamente côncavas e estritamente convexas.

No que segue, salvo menção em contrário, estudaremos funções convexas definidas em conjuntos convexos de $\mathbb{R}$, de $\mathbb{R}^n$ e também de espaços vetoriais reais normados (não necessariamente de dimensão finita).

• **Epigráficos e hipográficos**

Seja C um conjunto convexo não vazio de um espaço vetorial real $\mathcal{V}$ e $F : C \to \mathbb{R}$ uma função convexa. Como ocorre com todas as funções, o conjunto

$$\Gamma(F) \ := \ \big\{\big(x, \ F(x)\big), \ x \in C\big\} \ \subset \ \mathcal{V} \times \mathbb{R} \ ,$$

denota o *gráfico* da função F. O conjunto

$$\mathcal{E}(F) \ := \ \big\{\big(x, \ y\big) \in C \times \mathbb{R} \big| \, y \geq F(x)\big\} \ \subset \ \mathcal{V} \times \mathbb{R} \ ,$$

é denominado *epigráfico* (ou *epigrafo*) de F. Pictoricamente, o epigráfico de uma função convexa F é o conjunto de todos os pontos de $\mathcal{V} \times \mathbb{R}$ que estão sobre ou acima do gráfico de F.

A noção correspondente a epigráfico para o caso de funções côncavas é denominada *hipográfico* (ou *hipografo*):

$$\mathcal{H}(F) \ := \ \big\{\big(x, \ y\big) \in C \times \mathbb{R} \big| \, y \leq F(x)\big\} \ \subset \ \mathcal{V} \times \mathbb{R} \ .$$

Em palavras, o hipógráfico de uma função côncava F é o conjunto de todos os pontos de $\mathcal{V} \times \mathbb{R}$ que estão sobre ou abaixo do gráfico de F.

O conjunto

$$\mathcal{E}^0(F) \ := \ \big\{\big(x, \ y\big) \in C \times \mathbb{R} \big| \, y > F(x)\big\} \ \subset \ \mathcal{V} \times \mathbb{R} \ ,$$

é denominado *epigráfico estrito* de F. Em palavras, o epigráfico estrito de uma função convexa F é o conjunto de todos os pontos de $\mathcal{V} \times \mathbb{R}$ que estão acima do gráfico de F. É claro que $\mathcal{E}(F) = \mathcal{E}^0(F) \cup \Gamma(F)$, uma união disjunta.

O produto cartesiano $\mathcal{V} \times \mathbb{R}$, onde o epigráfico de uma função convexa está definido, possui uma estrutura natural de espaço vetorial real, sendo a combinação linear de dois pares $(v_1, \ x_1)$, $(v_2, \ x_2) \in \mathcal{V} \times \mathbb{R}$, para $\alpha_1, \ \alpha_2 \in \mathbb{R}$, definida por $\alpha_1(v_1, \ x_1) + \alpha_2(v_2, \ x_2) := \big(\alpha_1 v_1 + \alpha_2 v_2, \ \alpha_1 x_1 + \alpha_2 x_2\big)$. O espaço vetorial assim definido é denominado *soma direta* de $\mathcal{V}$ e $\mathbb{R}$ e é denotado por $\mathcal{V} \oplus \mathbb{R}$. Para mais detalhes sobre essa construção – não relevantes para o que segue – vide Seção 2.3.4, página 211.

Proposição 5.5 *Seja C um conjunto convexo não vazio de um espaço vetorial real $\mathcal{V}$ e $F : C \to \mathbb{R}$ uma função. Então, F é convexa se e somente se seu epigráfico for um subconjunto convexo de $\mathcal{V} \oplus \mathbb{R}$.* □

Prova. Seja F convexa e sejam $(v_1,\ y_1)$, $(v_2,\ y_2)$ elementos de $\mathcal{E}(F)$, Isso significa que $y_1 \geq F(v_1)$ e que $y_2 \geq F(v_2)$.

Para $\lambda \in [0,\ 1]$ temos que $\lambda(v_1,\ y_1) + (1-\lambda)(v_2,\ y_2) = \big(\lambda v_1 + (1-\lambda)v_2,\ \lambda y_1 + (1-\lambda)y_2\big)$. Agora, pela convexidade da F,

$$\lambda y_1 + (1-\lambda)y_2 \ \geq\ \lambda F(v_1) + (1-\lambda)F(v_2) \ \geq\ F\big(\lambda v_1 + (1-\lambda)v_2\big) ,$$

estabelecendo que $\lambda(v_1,\ y_1) + (1-\lambda)(v_2,\ y_2)$ é um elemento de $\mathcal{E}(F)$ para todo $\lambda \in [0,\ 1]$, provando a convexidade de $\mathcal{E}(F)$.

Reciprocamente, vamos agora supor que $\mathcal{E}(F)$ seja um conjunto convexo. Isso, em particular, significa que para dois pontos do gráfico de F, $\big(v_1,\ F(v_1)\big)$ e $\big(v_2,\ F(v_2)\big)$, vale que $\lambda\big(v_1,\ F(v_1)\big) + (1-\lambda)\big(v_2,\ F(v_2)\big) \in \mathcal{E}(F)$ para todo $\lambda \in [0,\ 1]$. Assim, $\Big(\lambda v_1 + (1-\lambda)v_2,\ \lambda F(v_1) + (1-\lambda)F(v_2)\Big) \in \mathcal{E}(F)$. Agora, isso significa que $\lambda F(v_1) + (1-\lambda)F(v_2) \geq F\big(\lambda v_1 + (1-\lambda)v_2\big)$, o que afirma que F é convexa. ■

• **Supremo de uma família de funções convexas**

Vamos agora a uma afirmação muito relevante sobre funções convexas definidas em um mesmo conjunto convexo. Ela será evocada, por exemplo, quando discutirmos a noção de Transformada de Legendre na Seção **??**, página **??**.

Proposição 5.6 *Seja $\mathcal{V}$ um espaço vetorial real e C um subconjunto convexo não vazio de $\mathcal{V}$. Seja $\big\{F_\sigma : C \to \mathbb{R},\ \ \sigma \in \Sigma\big\}$ uma família de funções convexas em C tal que para todo $v \in C$ o supremo $\sup\big\{F_\sigma(v),\ \sigma \in \Sigma\big\}$ exista em $\mathbb{R}$. Defina-se a função $F : C \to \mathbb{R}$ por $F(v) := \sup\big\{F_\sigma(v),\ \sigma \in \Sigma\big\}$. Então, $\mathcal{E}(F) = \bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma)$ e F é convexa.* □

Prova. Afirmamos que $\mathcal{E}(F) \subset \bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma)$. De fato, se $(v,\ y) \in \mathcal{E}(F)$, então $y \geq F(v) := \sup\big\{F_\sigma(v),\ \sigma \in \Sigma\big\}$. Logo, $y \geq F_\sigma(v)$ para todo $\sigma \in \Sigma$ e, portanto, $(v,\ y) \in \mathcal{E}(F_\sigma)$ para todo $\sigma \in \Sigma$, o que significa que $\mathcal{E}(F) \subset \mathcal{E}(F_\sigma)$ para todo $\sigma \in \Sigma$, ou seja, $\mathcal{E}(F) \subset \bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma)$.

Afirmamos também que $\mathcal{E}(F) \supset \bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma)$. De fato, se $(v,\ y) \in \bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma)$, então, para todo $\sigma \in \Sigma$ tem-se $(v,\ y) \in \mathcal{E}(F_\sigma)$. Portanto, para todo $\sigma \in \Sigma$ tem-se $y \geq F_\sigma(v)$. Logo, $y \geq \sup\big\{F_\sigma(v),\ \sigma \in \Sigma\big\}$ e, consequentemente, $(v,\ y) \in \mathcal{E}(F)$. Isso estabeleceu que $\bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma) \subset \mathcal{E}(F)$.

Provamos, dessa forma, que $\mathcal{E}(F) = \bigcap_{\sigma\in\Sigma} \mathcal{E}(F_\sigma)$. Pelas hipóteses e pela Proposição 5.5, página 346, cada epigráfico $\mathcal{E}(F_\sigma)$ é convexo. Portanto, pelo Lema 5.2, página 337, $\mathcal{E}(F)$ também é convexo, o que implica, novamente pela Proposição 5.5, que F é uma função convexa. ■

5.2.1 Funções Convexas em Espaços Vetoriais Normados e sua Continuidade

Vamos agora considerar uma importante generalização da nossa discussão para espaços vetoriais reais normados. A leitura do que segue requer noções de topologia de espaços métricos, como espaços vetoriais normados, temas tratados em capítulos posteriores. O principal resultado é:

Proposição 5.7 *Seja $\mathcal{V}$ um espaço vetorial real dotado de uma norma $\|\cdot\|$, seja $C \subset \mathcal{V}$ convexo, com interior C^0 não vazio, e seja $f : C \to \mathbb{R}$ uma função convexa em C. Se f for limitada superiormente em C, ou seja, se possuir um majorante superior finito $\mathsf{S} := \sup\{f(x),\ x \in C\} < \infty$, então f é contínua em C^0, o interior de C.*

Se $f : C \to \mathbb{R}$ côncava em C e limitada inferiormente em C, ou seja, tal que $\mathsf{I} := \inf\{f(x),\ x \in C\} > -\infty$, então, f é contínua em C^0, o interior de C. □

Prova. Provaremos apenas a afirmação para funções convexas, pois a afirmação para funções côncavas decorre imediatamente da mesma (trocando-se f por $-f$).

Vamos supor que existe $z \in C^0$ tal que f não seja contínua em z. Então, existe $\epsilon > 0$ tal que para cada $\delta > 0$ é possível encontrar um $x \in \mathcal{V}$ tal que $\|z - x\| < \delta$ mas com $|f(z) - f(x)| > \epsilon$ (contrariamente, f seria contínua em z).

Como $z \in C^0$ e C^0 é um conjunto aberto, existe $r > 0$ tal que a bola aberta de raio r centrada em z, $B(r,\ z) := \big\{y \in \mathcal{V}\big|\ \|y - z\| < r\big\}$, está inteiramente contida em C^0: $B(r,\ z) \subset C^0$.

Vamos escolher $\lambda \in (0,\ 1]$ tal que

$$\frac{1-\lambda}{\lambda} \ > \ \frac{\mathsf{S} - f(z)}{\epsilon} \, . \tag{5.12}$$

Que uma tal escolha sempre é possível segue do fato que a imagem da função $(0,\ 1] \ni \lambda \mapsto \lambda^{-1} - 1$ é $[0,\ \infty)$, como se vê facilmente. Segue de (5.12) que

$$\frac{\epsilon}{\lambda} \ > \ \mathsf{S} - f(z) + \epsilon \, , \tag{5.13}$$

relação que usaremos adiante.

Tomemos $\delta = r\lambda$. Então, pelo que afirmamos acima, existe $x \in \mathcal{V}$ tal que $\|x - z\| < \delta = r\lambda$ mas $|f(z) - f(x)| > \epsilon$. Note-se que, como $\lambda \in (0,\ 1]$, tem-se que $\delta \leq r$ e, portanto, $x \in B(r,\ z) \subset C^0$. Como $|f(z) - f(x)| > \epsilon$, há duas situações possíveis: *situação a:* $f(x) > f(z) + \epsilon$ e *situação b:* $f(x) < f(z) - \epsilon$.

Situação a: $f(x) > f(z) + \epsilon$. Seja $y \in \mathcal{V}$ dado por

$$y \ := \ \frac{1}{\lambda}x - \frac{1-\lambda}{\lambda}z \, .$$

Teremos $\|y - z\| = \frac{1}{\lambda}\|x - z\| < \frac{\delta}{\lambda} = r$, de modo que $y \in B(r,\ z) \subset C^0$. Porém, temos $x = \lambda y + (1 - \lambda)z$ e como f é convexa, temos $f(x) \leq \lambda f(y) + (1 - \lambda)f(z)$. Logo,

$$f(y) \ \geq \ \frac{1}{\lambda}f(x) - \frac{1-\lambda}{\lambda}f(z) \ > \ \frac{1}{\lambda}(f(z) + \epsilon) - \frac{1-\lambda}{\lambda}f(z) \ = \ f(z) + \frac{\epsilon}{\lambda} \ \overset{(5.13)}{>} \ \mathsf{S} + \epsilon \, .$$

Assim, obtivemos, $f(y) > \mathsf{S} + \epsilon > \mathsf{S}$, o que contraria a definição de S, mostrando que a situação a é impossível.

Situação b: $f(x) < f(z) - \epsilon$. Seja $y \in \mathcal{V}$ dado por

$$y \ := \ \frac{1}{\lambda}z - \frac{1-\lambda}{\lambda}x \, .$$

Teremos $\|y - z\| = \frac{1-\lambda}{\lambda}\|x - z\| < (1 - \lambda)\frac{\delta}{\lambda} = (1 - \lambda)r < r$, de modo que $y \in B(r,\ z) \subset C^0$. Porém, temos $z = \lambda y + (1 - \lambda)x$ e como f é convexa, temos $f(z) \leq \lambda f(y) + (1 - \lambda)f(x)$. Logo,

$$f(y) \ \geq \ \frac{1}{\lambda}f(z) - \frac{1-\lambda}{\lambda}f(x) \ > \ \frac{1}{\lambda}f(z) - \frac{1-\lambda}{\lambda}(f(z) - \epsilon) \ = \ f(z) + \epsilon\frac{1-\lambda}{\lambda} \ \overset{(5.12)}{>} \ \mathsf{S} \, .$$

Assim, obtivemos, $f(y) > \mathsf{S}$, o que novamente contraria a definição de S, mostrando que a situação b também é impossível.

A resolução dessas contradições é que um tal ponto $z \in C^0$ onde f é descontínua não pode existir. ■

• Comentários à Proposição 5.7

Se $\mathcal{V}$ for um espaço vetorial real normado e $f : \mathcal{V} \to \mathbb{R}$ for um funcional linear em $\mathcal{V}$, então f é uma função convexa (e côncava) no conjunto convexo $\mathcal{V}$. Se $\mathcal{V}$ não for um espaço de dimensão finita, linearidade não necessariamente faz de f uma função contínua. Há exemplos bem conhecidos de funcionais lineares descontínuos de um espaço normado em $\mathbb{R}$ ou $\mathbb{C}$, como o funcional delta de Dirac discutido à página 2338. Assim, podem existir funções convexas ou côncavas não continuas em espaços normados de dimensão infinita[11] e aí reside a relevância de resultados como os da Proposição 5.7.

É de se notar também que a condição de limitação superior (inferior) listada na Proposição 5.7 é suficiente, mas não é necessária para que uma função convexa (côncava) seja contínua no interior do seu domínio convexo de definição. Segundo a Proposição 42.1, página 2337, se $\mathcal{V}$ é um espaço vetorial normado, um funcional linear $f : \mathcal{V} \to \mathbb{R}$ é contínuo se e somente se $\sup\{|f(x)|,\ x \in \mathcal{V},\ \|x\| = 1\} < \infty$. Funcionais lineares $f : \mathcal{V} \to \mathbb{R}$ que satisfaçam essa condição de continuidade não são limitados nem superior nem inferiormente em $\mathcal{V}$ (segundo a definição que usamos acima).

[11] Todo o Capítulo 43, página 2531, é dedicado a operadores lineares não contínuos.

• **Continuidade no caso de dimensão finita**

No caso em que $\mathcal{V}$ é um espaço vetorial real normado de dimensão finita (como $\mathbb{R}^n$), a condição de uma função $f : \mathcal{V} \to \mathbb{R}$ convexa (côncava) ser limitada superiormente (inferiormente) é dispensável para a continuidade:

Corolário 5.1 *Seja $C \subset \mathbb{R}^n$, convexo, e seja $f : C \to \mathbb{R}$ uma função convexa ou côncava em C. Então, f é contínua em C^0, o interior de C.* □

Prova. Consideramos apenas o caso em que f é convexa, pois o outro é análogo. Seja $a \in C^0$. Então, é possível encontrar $r > 0$ tal que $B(r,\ a)$, a bola aberta de raio r centrada em a, está inteiramente contida em C^0. Dentro dessa bola é possível encontrar um conjunto finito $v_1, \ \ldots, \ v_{n+1}$ de pontos tais que o conjunto $\mathcal{C} := \{\lambda_1 v_1 + \cdots + \lambda_{n+1} v_{n+1}, \ \lambda_1, \ \ldots \ \lambda_{n+1} \in [0, \ 1]$ com $\lambda_1 + \cdots + \lambda_{n+1} = 1\}$ é uma vizinhança convexa de a. Pela convexidade de f, temos para todo ponto $x = \lambda_1 v_1 + \cdots + \lambda_{n+1} v_{n+1}$ de $\mathcal{C}$ que $f(x) \leq \lambda_1 f(v_1) + \cdots + \lambda_{n+1} f(v_{n+1}) \leq \max\{f(v_1), \ \ldots, \ f(v_{n+1})\}$. Logo, f é limitada superiormente no convexo $\mathcal{C}$ e, pela Proposição 5.7, página 346, f é contínua em $\mathcal{C}$ e, em particular em a. Como a é um ponto arbitrário de C^0 a demonstração está completa. ■

Exemplos 5.5 Considere as funções

$$f(x) = \begin{cases} 1 \, , & \text{se } x = 0 \, , \\ 0 \, , & \text{se } 0 < x < 1 \, , \\ 2 \, , & \text{se } x = 1 \, . \end{cases} \qquad f(x) = \begin{cases} 1 \, , & \text{se } x = 0 \, , \\ x \, , & \text{se } 0 < x < 1 \, , \\ 2 \, , & \text{se } x = 1 \, . \end{cases} \qquad f(x) = \begin{cases} 1/x \, , & \text{se } 0 < x < 1 \, , \\ 2 \, , & \text{se } x = 1 \, . \end{cases}$$

As duas primeiras definidas em $[0, \ 1]$ e a última em $(0, \ 1]$. Todas são convexas em seus respectivos intervalos de definição e são contínuas no intervalo $(0, \ 1)$. ♦

5.2.2 Transformações Afins

Sejam $\mathcal{U}$ e $\mathcal{V}$ dois espaços vetoriais reais. Uma aplicação $T : \mathcal{U} \to \mathcal{V}$ é dita ser uma *transformação afim* se para todo $x, \, y \in \mathcal{U}$ e todo $\lambda \in [0, \, 1]$ valer

$$T\big(\lambda x + (1-\lambda)y\big) = \lambda T(x) + (1-\lambda)T(y) \, . \tag{5.14}$$

A proposição a seguir, sobre transformações afins contínuas entre espaços vetoriais reais normados, é de grande importância, sendo empregada, por exemplo, na demonstração do Teorema de Mazur-Ulam, Teorema 3.8, página 297.

Proposição 5.8 *Se $T : \mathcal{U} \to \mathcal{V}$ é uma transformação afim e contínua entre dois espaços vetoriais reais $\mathcal{U}$ e $\mathcal{V}$, ambos espaços normados*[12]*, então $U : \mathcal{U} \to \mathcal{V}$ definida por $U(x) := T(x) - T(0)$, $x \in \mathcal{U}$, é uma transformação real-linear.* □

Prova. Desejamos provar que para todos $\alpha, \, \beta \in \mathbb{R}$ e todos $x, \, y \in \mathcal{U}$ vale $U(\alpha x + \beta y) = \alpha U(x) + \beta U(y)$. A demonstração é feita em três etapas: **I-III.**.

I. Provemos em primeiro lugar que para todo $x \in \mathcal{U}$ vale

$$U(x) = -U(-x) \, . \tag{5.15}$$

Tomando $\lambda = 1/2$ e $y = -x$ em (5.14), obtemos $T(0) = \frac{1}{2}\big(T(x) + T(-x)\big)$ e disso segue imediatamete que $0 = \frac{1}{2}\big((T(x) - T(0) + (T(-x) - T(0)\big)$, ou seja, $U(x) + U(-x) = 0$, provando (5.15).

II. Vamos agora provar que para todo $\alpha \in \mathbb{R}$ e todo $x \in \mathcal{U}$ vale

$$U(\alpha x) = \alpha U(x) \, . \tag{5.16}$$

Como $U(x) := T(x) - T(0)$ essa relação é evidente para $\alpha = 0$ e, portanto, com (5.15) é suficiente considerarmos $\alpha > 0$.

A relação (5.16) vale, evidentemente, quando $\alpha = 1$. Vemos provar por indução que

$$U(nx) = nU(x) \tag{5.17}$$

[12] A afirmação, e sua demonstração abaixo, é também válida em espaços vetoriais topológicos gerais.

para todo $n \in \mathbb{N}$. Vamos, então supor que para um dado $n \in \mathbb{N}$ valha $U(nx) = nU(x)$. Escrevamos

$$nx \;=\; \lambda x' + (1-\lambda)0 \;,$$

com

$$\lambda \;=\; \frac{n}{n+1} \quad \text{e} \quad x' \;=\; \big(n+1\big)x$$

Como $\lambda \in (0,\, 1)$, podemos usar (5.14) e escrever

$$T(nx) \;=\; T\Big(\lambda x' + (1-\lambda)0\Big) \stackrel{(5.14)}{=} \lambda T(x') + (1-\lambda)T(0) \;=\; \lambda\big(T(x') - T(0)\big) + T(0) \;.$$

Assim, estabelecemos que $U(nx) = \lambda U(x')$, ou seja,

$$U(nx) \;=\; \frac{n}{n+1} U\big((n+1)x\big) \;.$$

Portanto, estabelecemos que $U\big((n+1)x\big) = \frac{n+1}{n} U(nx)$. Pela hipótese de indução, vale $U(nx) = nU(x)$ e, assim, estabelecemos que $U\big((n+1)x\big) = (n+1)U(x)$. Com isso, (5.17) cai demonstrada para <u>todo</u> $n \in \mathbb{N}$.

Seja agora r um número racional positivo da forma $r = p/q$, com $p,\, q \in \mathbb{N}$, e consideremos $U(rx)$. Por (5.17),

$$U(rx) \;=\; U\left(\frac{p}{q}x\right) \stackrel{(5.17)}{=} pU\left(\frac{1}{q}x\right) \;.$$

Porém, (5.17) (que é válida para todo $x \in \mathcal{U}$) também nos diz, substituindo x por $\frac{1}{n}x$ que $U\left(\frac{1}{n}x\right) = \frac{1}{n}U(x)$. Assim, temos

$$U(rx) \;=\; \frac{p}{q}U(x) \;=\; rU(x) \;.$$

Isso estabeleceu (5.16) para todo $\alpha \in \mathbb{Q}$. Pela hipótese de continuidade[13] sobre T, isso prova (5.16) para todo $\alpha \in \mathbb{R}$.

III. Resta considerar o caso em que $\alpha,\, \beta \in \mathbb{R}$, ambos não nulos. Para $x,\, y \in \mathcal{U}$ podemos escrever

$$\alpha x + \beta y \;=\; \lambda x' + (1-\lambda)y' \;,$$

onde

$$\lambda \;=\; \frac{|\alpha|}{|\alpha|+|\beta|} \;, \quad 1-\lambda \;=\; \frac{|\beta|}{|\alpha|+|\beta|} \;, \quad x' \;=\; \frac{\alpha\big(|\alpha|+|\beta|\big)}{|\alpha|}x \quad \text{e} \quad y' \;=\; \frac{\beta\big(|\alpha|+|\beta|\big)}{|\beta|}y \;, \tag{5.18}$$

sendo que $\lambda \in (0,\ 1)$ e $x',\, y' \in \mathcal{U}$. Assim,

$$\begin{aligned}
U(\alpha x + \beta y) \;&=\; U\big(\lambda x' + (1-\lambda)y\big) \;=\; T\big(\lambda x' + (1-\lambda)y\big) - T(0) \\
&\stackrel{(5.14)}{=} \lambda T(x') + (1-\lambda)T(y') - T(0) \;=\; \lambda\Big(T(x') - T(0)\Big) + (1-\lambda)\Big(T(y') - T(0)\Big) \\
&=\; \lambda U(x') + (1-\lambda)U(y') \stackrel{(5.18)}{=} \frac{|\alpha|}{|\alpha|+|\beta|}U\left(\frac{\alpha\big(|\alpha|+|\beta|\big)}{|\alpha|}x\right) + \frac{|\beta|}{|\alpha|+|\beta|}U\left(\frac{\beta\big(|\alpha|+|\beta|\big)}{|\beta|}y\right) \\
&\stackrel{(5.16)}{=} \frac{|\alpha|}{|\alpha|+|\beta|}\frac{\alpha\big(|\alpha|+|\beta|\big)}{|\alpha|}U(x) + \frac{|\beta|}{|\alpha|+|\beta|}\frac{\beta\big(|\alpha|+|\beta|\big)}{|\beta|}U(y) \;=\; \alpha U(x) + \beta U(y) \;,
\end{aligned}$$

como desejávamos estabelecer. ■

• <u>Funções afins de um espaço vetorial real em $\mathbb{R}$</u>

Seja $\mathcal{U}$ um espaço vetorial real e $\mathcal{U}^*$ seu espaço dual. $\mathcal{U}$ deve ser um espaço normado, ou mesmo um espaço vetorial topológico geral, e $\mathcal{U}^*$ seu dual topológico. Uma função $F : \mathcal{U} \to \mathbb{R}$ da forma

$$F(u) \;:=\; \langle w,\ u\rangle + b$$

[13] Este é o único ponto em que a hipótese de continuidade é evocada.

é dita ser uma *função afim* de $\mathcal{U}$ em $\mathbb{R}$. Acima, $w \in \mathcal{U}^*$ e $b \in \mathbb{R}$ são constantes e $\langle \cdot, \cdot \rangle$ denota o emparelhamento (*"pairing"*) entre elementos de $\mathcal{U}^*$ e $\mathcal{U}$.

O leitor há a aperceber-se que a definição acima coincide com a de transformação afim dada em (5.14), página 348 (vide também Proposição 5.8, página 348), para o caso em que $\mathcal{V} = \mathbb{R}$, de fato, aplicações lineares e contínuas entre um espaço normado $\mathcal{U}$ e $\mathbb{R}$ são todas da forma $\langle w, u \rangle$ para algum $w \in \mathcal{U}^*$.

Para $u_1, u_2 \in \mathcal{U}$ e $\lambda \in [0, 1]$ vale

$$F\big(\lambda u_1 + (1-\lambda)u_2\big) = \big\langle w, \lambda u_1 + (1-\lambda)u_2 \big\rangle + b = \lambda\Big(\langle w, u_1\rangle + b\Big) + (1-\lambda)\Big(\langle w, u_2\rangle + b\Big) = \lambda F(u_1) + (1-\lambda)F(u_2) ,$$

o que mostra que toda função afim é convexa e côncava.

5.2.3 Funções Côncavas e Convexas de uma Variável

O estudo de funções convexas de uma variável real é, por si, relevante em aplicações e prepara o terreno para o estudo de funções convexas de várias variáveis reais.

Consideraremos agora funções definidas em um conjunto convexo $I \subset \mathbb{R}$ de interior I^0 não vazio. Podemos ter $I = \mathbb{R}$, ou um intervalo aberto, semiaberto ou fechado, como $[A, B]$, (A, B), $[A, B)$, $(A, B]$, $[A, \infty)$, (A, ∞), $(-\infty, A]$ ou $(-\infty, A)$, com $-\infty < A < B < \infty$. I^2 designa o produto Cartesiano $I \times I$ e I_d designa seu conjunto diagonal: $I_d := \{(x, x), x \in I\} \subset I^2$.

Uma função $f : I \to \mathbb{R}$ é dita ser uma *função convexa* se para todos $x, y \in I$ e todo $\lambda \in [0, 1]$ valer

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y) . \tag{5.19}$$

Uma função $f : I \to \mathbb{R}$ é dita ser uma *função côncava* se para todos $x, y \in I$ e todo $\lambda \in [0, 1]$ valer

$$f\big(\lambda x + (1-\lambda)y\big) \geq \lambda f(x) + (1-\lambda)f(y) . \tag{5.20}$$

E. 5.4 *Exercício.* A noção de convexidade (de concavidade) possui uma interpretação geométrica muito simples para funções de uma variável real. Uma função $f : I \to \mathbb{R}$ é convexa (respectivamente, côncava) se e somente se dados dois pontos quaisquer $x < y$ de seu domínio o gráfico de f no intervalo (x, y) ficar abaixo (respectivamente, acima) da linha reta que conecta o par $(x, f(x))$ ao par $(y, f(y))$, tal como expresso nos gráficos da Figura 5.4, página 350. Justifique essa afirmação com base nas definições. ✠

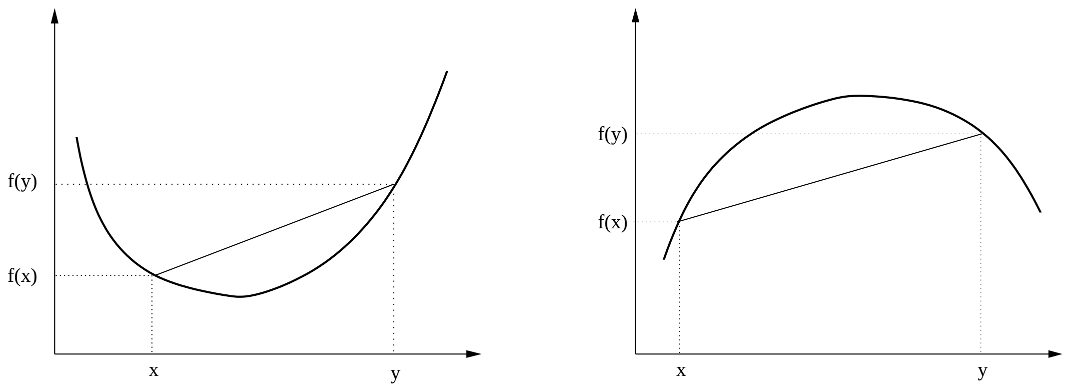

Figura 5.4: O gráfico de uma função convexa (à esquerda) e de uma função côncava (à direita). Em ambos é indicado o segmento de reta conectando par $(x, f(x))$ ao par $(y, f(y))$.

E. 5.5 *Exercício.* Seguindo a definição, mostre que $f(x) = |x|$ e $f(x) = x^2$ são funções convexas em $\mathbb{R}$. ✠

E. 5.6 *Exercício (fácil).* Sejam $f : \mathbb{R} \to \mathbb{R}$ e $g : \mathbb{R} \to \mathbb{R}$ duas funções convexas. Mostre que se f é não decrescente, então $f \circ g$ é convexa. ✠

Se $f : I \to \mathbb{R}$ é convexa, é muito fácil demonstrar, usando indução finita, que para todos $n \in \mathbb{N}$, todos $x_1, \ldots, x_n \in I$ e todos $\lambda_1, \ldots, \lambda_n \in [0, 1]$ tais que $\lambda_1 + \cdots + \lambda_n = 1$ vale

$$f\big(\lambda_1 x_1 + \cdots + \lambda_n x_n\big) \;\leq\; \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n) \,. \tag{5.21}$$

Se f é côncava, temos

$$f\big(\lambda_1 x_1 + \cdots + \lambda_n x_n\big) \;\geq\; \lambda_1 f(x_1) + \cdots + \lambda_n f(x_n) \,. \tag{5.22}$$

Provaremos apenas o caso convexo, pois o outro é análogo. Suponhamos a afirmação válida para $n-1$, com $n \geq 3$. Podemos supor que haja ao menos dois λ's não nulos com $\lambda_1, \ldots, \lambda_n \in [0, 1]$ tais que $\lambda_1 + \cdots + \lambda_n = 1$, pois se houver apenas um, esse deve valer 1 e os demais 0 e não haveria o que se demonstrar. Sem perda de generalidade, suponhamos assim que $\lambda_{n-1} + \lambda_n > 0$. Então, como podemos escrever

$$\lambda_1 x_1 + \cdots + \lambda_n x_n \;=\; \lambda_1 x_1 + \cdots + \lambda_{n-2} x_{n-2} + (\lambda_{n-1} + \lambda_n) \left[\frac{\lambda_{n-1} x_{n-1} + \lambda_n x_n}{\lambda_{n-1} + \lambda_n}\right]$$

temos, pela hipótese de indução,

$$\begin{aligned} f\big(\lambda_1 x_1 + \cdots + \lambda_n x_n\big) \;&=\; f\left(\lambda_1 x_1 + \cdots + \lambda_{n-2} x_{n-2} + (\lambda_{n-1} + \lambda_n) \left[\frac{\lambda_{n-1} x_{n-1} + \lambda_n x_n}{\lambda_{n-1} + \lambda_n}\right]\right) \\ &\leq\; \lambda_1 f(x_1) + \cdots + \lambda_{n-2} f(x_{n-2}) + (\lambda_{n-1} + \lambda_n) f\left(\frac{\lambda_{n-1} x_{n-1} + \lambda_n x_n}{\lambda_{n-1} + \lambda_n}\right) \\ &\leq\; \lambda_1 f(x_1) + \cdots + \lambda_{n-2} f(x_{n-2}) + \lambda_{n-1} f(x_{n-1}) + \lambda_n f(x_n) \,, \end{aligned}$$

sendo que, na última desigualdade, usamos a convexidade de f para obter

$$f\left(\frac{\lambda_{n-1} x_{n-1} + \lambda_n x_n}{\lambda_{n-1} + \lambda_n}\right) \;\leq\; \frac{\lambda_{n-1}}{\lambda_{n-1} + \lambda_n} f(x_{n-1}) + \frac{\lambda_n}{\lambda_{n-1} + \lambda_n} f(x_n) \,.$$

Isso provou (5.21) para todo $n \in \mathbb{N}$.

A desigualdade (5.21) (ou sua forma côncava (5.22)) é por vezes denominada *desigualdade de Jensen*[14]. É importante mencionar que a desigualdade de Jensen pode ser ainda generalizada e (5.21) é apenas sua versão mais simples (discreta). Para uma forma mais geral, vide Proposição 5.19, página 366, em especial, vide (5.50).

• Funções côncavas ou convexas definidas na semirreta $[0, \infty)$

O seguinte resultado sobre funções côncavas ou convexas definidas na semirreta $[0, \infty)$ é útil e será usado nestas Notas.

Lema 5.4 *Seja $f : [0, \infty) \to \mathbb{R}$ com $f(0) = 0$. Então, se f for convexa tem-se*

$$f(x+y) \;\geq\; f(x) + f(y) \,, \qquad \forall x,\ y \in [0, \infty) \tag{5.23}$$

e se f for côncava tem-se

$$f(x+y) \;\leq\; f(x) + f(y) \,, \qquad \forall x,\ y \in [0, \infty) \,. \tag{5.24}$$

□

A propriedade (5.23) é denominada *supaditividade*, enquanto que a propriedade (5.24) é denominada *subaditividade*. Um uso da desigualdade (5.24) será encontrado no estudo de Espaços Métricos, como na Proposição 25.1, página 1416.

Prova do Lema 5.4. Consideremos apenas o caso em que f é convexa, pois se f for côncava valerá que $-f$ é convexa e, portanto, (5.23) implica (5.24). Como $f(0) = 0$, a condição de convexidade implica que

$$f(\lambda x) \;\leq\; \lambda f(x) \tag{5.25}$$

[14] Johan Ludwig William Valdemar Jensen (1859–1925). A desigualdade de Jensen, assim como outros trabalhos do mesmo sobre funções convexas, data de 1906. Para a referência original, vide nota de rodapé 17 à página 360.

para todo $\lambda \in [0,\ 1]$ e todo $x \geq 0$.

Caso $x = y = 0$ a desigualdade (5.23) exige que $f(0) \geq 2f(0)$, o que é trivialmente verdade pois, pelas hipóteses, $f(0) = 0$. Caso x e y não forem simultaneamente nulos, temos

$$f(x+y) \;=\; \frac{x}{x+y}f(x+y) + \frac{y}{x+y}f(x+y) \;\overset{(5.25)}{\geq}\; f\left(\frac{x}{x+y}(x+y)\right) + f\left(\frac{y}{x+y}(x+y)\right) \;=\; f(x) + f(y)\;,$$

demonstrando (5.23). Acima, ao evocarmos (5.25), usamos que $\frac{x}{x+y}$ e $\frac{y}{x+y}$ são ambos, pelas hipóteses, elementos do intervalo $[0,\ 1]$. ■

Observe o leitor que a hipótese de convexidade é empregada na demonstração apenas para justificar (5.25). Nesse contexto, temos o seguinte resultado:

Lema 5.5 *Se $f : \mathbb{R} \to \mathbb{R}$ é subaditiva e satisfaz $f(\lambda x) \leq \lambda f(x)$ para todo $\lambda \in [0,\ 1]$ e todo $x \in \mathbb{R}$, então f é convexa.* □

Prova. Para todos $x,\ y \in \mathbb{R}$ e $\lambda \in [0,\ 1]$ temos

$$f\big(\lambda x + (1-\lambda)y\big) \;\overset{\text{subaditividade}}{\leq}\; f\big(\lambda x\big) + f\big((1-\lambda)y\big) \;\leq\; \lambda f(x) + (1-\lambda)f(y)\;,$$

demonstrando a convexidade. ■

Dos Lemas 5.4 e 5.5 temos, portanto, o seguinte:

Corolário 5.2 *Seja $f : [0,\ \infty) \to \mathbb{R}$ com $f(0) = 0$ e satisfazendo $f(\lambda x) \leq \lambda f(x)$ para todo $\lambda \in [0,\ 1]$ e todo $x \in [0,\ \infty)$. Então, f é convexa se e somente se for subaditiva.*

Analogamente, seja $f : [0,\ \infty) \to \mathbb{R}$ com $f(0) = 0$ e satisfazendo $f(\lambda x) \geq \lambda f(x)$ para todo $\lambda \in [0,\ 1]$ e todo $x \in [0,\ \infty)$. Então, f é côncava se e somente se for supaditiva. □

• Obtendo funções convexas de uma variável a partir de funções convexas em espaços vetoriais reais

O lema a seguir mostra como podemos obter uma função convexa de uma variável real a partir de uma função convexa definida em um espaço vetorial real.

Lema 5.6 *Seja $F : C \to \mathbb{R}$ uma função convexa definida em um subconjunto convexo C de um espaço vetoria real $\mathcal{V}$. Sejam $x_1,\ x_2 \in C$, distintos, e defina-se $f : [0,\ 1] \to \mathbb{R}$ por $f(\alpha) := F\big((1-\alpha)x_1 + \alpha x_2\big)$. Então, f é convexa.* □

Prova. Sejam $\lambda \in [0,\ 1]$ e $\alpha_1,\ \alpha_2 \in [0,\ 1]$. Verifique-se que vale trivialmente a igualdade

$$\big(1 - \lambda\alpha_1 - (1-\lambda)\alpha_2\big)x_1 + \big(\lambda\alpha_1 + (1-\lambda)\alpha_2\big)x_2 \;=\; \lambda\big((1-\alpha_1)x_1 + \alpha_1 x_2\big) + (1-\lambda)\big((1-\alpha_2)x_1 + \alpha_2 x_2\big)\;.$$

Com isso, temos que

$$\begin{aligned}
f\big(\lambda\alpha_1 + (1-\lambda)\alpha_2\big) \quad &= \quad F\Big(\big(1 - \lambda\alpha_1 - (1-\lambda)\alpha_2\big)x_1 + \big(\lambda\alpha_1 + (1-\lambda)\alpha_2\big)x_2\Big) \\
&= \quad F\Big(\lambda\big((1-\alpha_1)x_1 + \alpha_1 x_2\big) + (1-\lambda)\big((1-\alpha_2)x_1 + \alpha_2 x_2\big)\Big) \\
&\overset{\text{convexidade}}{\leq} \quad \lambda F\Big((1-\alpha_1)x_1 + \alpha_1 x_2\Big) + (1-\lambda)F\Big((1-\alpha_2)x_1 + \alpha_2 x_2\Big) \\
&= \quad \lambda f(\alpha_1) + (1-\lambda)f(\alpha_2)\;,
\end{aligned}$$

provando que f é convexa. ■

• **Propriedades do conjunto de funções convexas em I**

Seja $I \subset \mathbb{R}$ convexo. Se $f : I \to \mathbb{R}$ e $g : I \to \mathbb{R}$ são duas funções convexas, então para todos α, $\beta \in [0, \infty)$ a função $\alpha f + \beta g$ é também convexa em I. A prova disso é elementar. Essa propriedade afirma que o conjunto das funções convexas em I é um cone convexo[15]. Essas afirmações valem também para funções côncavas.

Seja $f_n : I \to \mathbb{R}$, $n \in \mathbb{N}$ uma sequência de funções convexas que converge pontualmente a uma função $f : I \to \mathbb{R}$, ou seja, tal que para cada $x \in I$ valha $f(x) = \lim_{n\to\infty} f_n(x)$. Então, f é igualmente convexa. A prova dessa afirmação é elementar e deixada ao estudante. Essa afirmação vale também para funções côncavas.

Seja $\{f_\omega : I \to \mathbb{R},\ \omega \in \Omega\}$ uma família de funções convexas definidas em I tal que para cada $x \in I$ exista $f(x) := \sup\{f_\omega(x),\ \omega \in \Omega\}$. Então, $f : I \to \mathbb{R}$ é também convexa. Vide Proposição 5.6, página 346. Para funções côncavas vale a mesma afirmação, com o supremo substituído pelo ínfimo.

• **Composição de funções convexas em $\mathbb{R}$**

O seguinte resultado simples será por vezes evocado:

Lema 5.7 *Sejam I_1 e I_2 dois subconjuntos convexos de $\mathbb{R}$ e sejam $f : I_2 \to \mathbb{R}$ e $g : I_1 \to I_2$ duas funções convexas, sendo f crescente. Então, a composição $f \circ g : I_1 \to \mathbb{R}$ é também uma função convexa.*

Caso f seja côncava e crescente e g seja côncava, então $f \circ g$ é côncava. □

Prova. É suficiente considerarmos o caso em que ambas f e g são convexas. Seja $\lambda \in [0, 1]$ e sejam u, $v \in I_1$. Então, $g\big(\lambda u + (1-\lambda)v\big) \leq \lambda g(u) + (1-\lambda)g(v)$ e, como f é crescente,

$$
\begin{aligned}
(f\circ g)\big(\lambda u + (1-\lambda)v\big) \; &= \; f\Big(g\big(\lambda u + (1-\lambda)v\big)\Big) \; \leq \; f\big(\lambda g(u) + (1-\lambda)g(v)\big) \\
&\leq \; \lambda f\big(g(u)\big) + (1-\lambda) f\big(g(v)\big) \; = \; \lambda (f\circ g)(u) + (1-\lambda)(f\circ g)(v) \; ,
\end{aligned}
$$

onde, na segunda desigualdade, usamos que f é convexa. Isso provou que $f \circ g$ é convexa.

O caso em que f é côncava e crescente e g é côncava é similar. ■

• **Uma condição equivalente à de convexidade**

Para uma função $f : I \to \mathbb{R}$, considere-se a função simétrica de duas variáveis $R_f \equiv R : I^2 \setminus I_d$ dada por

$$
R_f(x,\ y) \; \equiv \; R(x,\ y) \; := \; \frac{f(x) - f(y)}{x - y} \; , \qquad x \neq y \; .
$$

A proposição que segue mostra que com a função R podemos apresentar uma definição alternativa de convexidade (uma outra caracterização distinta da noção de função convexa será encontrada na Proposição 5.12, página 360).

Proposição 5.9 *Uma função $f : I \to \mathbb{R}$ é convexa se e somente se para todos x, y, $z \in I$, distintos, com $y < z$ valer*

$$
R(x,\ y) \; \leq \; R(x,\ z) \; , \tag{5.26}
$$

ou seja, se e somente se, fixado um dos argumentos, R for monotonamente não decrescente no outro argumento. □

No Exercício E. 5.7, página 354, apresenta-se uma interpretação geométrica da Proposição 5.9.

Prova da Proposição 5.9. **Parte I:** *supondo f convexa provamos (5.26).* Para provarmos (5.26) há três casos a se considerar: $x < y < z$, $y < x < z$ e $y < z < x$.

[15] Para entender essa nomenclatura o estudante deve recordar que se $\mathcal{C}$ é um cone convexo em $\mathbb{R}^3$, então se $\vec{v}$ e $\vec{u}$ são vetores de $\mathcal{C}$, segue que $\alpha\vec{v} + \beta\vec{u}$ é também um vetor de $\mathcal{C}$ para todos α, $\beta \in [0, \infty)$.

Caso 1: $x < y < z$. Como y fica entre x e z, podemos escrever $y = \lambda x + (1-\lambda)z$ adotando para tal $\lambda = \frac{z-y}{z-x}$ (notar que $\lambda \in (0,\ 1)$). Da convexidade de f segue, então, que

$$f(y) \;\leq\; \lambda f(x) + (1-\lambda)f(z) \;=\; \frac{z-y}{x-z}f(x) + \frac{y-x}{z-x}f(z)\;.$$

Subtraindo $f(x)$ de ambos os lados obtemos, após alguns cômputos elementares,

$$\frac{f(y)-f(x)}{y-x} \;\leq\; \frac{f(z)-f(x)}{z-x}\;,$$

ou seja, $R(x,\ y) \leq R(x,\ z)$.

Caso 2: $y < x < z$. Como x fica entre y e z, podemos escrever $x = \lambda y + (1-\lambda)z$ adotando para tal $\lambda = \frac{z-x}{z-y}$ (notar que $\lambda \in (0,\ 1)$). Da convexidade de f segue, então, que

$$f(x) \;\leq\; \lambda f(y) + (1-\lambda)f(z) \;=\; \frac{z-x}{z-y}f(y) + \frac{x-y}{z-y}f(z)\;.$$

Subtraindo $f(x)$ de ambos os lados, podemos escrever

$$0 \;\leq\; \frac{z-x}{z-y}\big(f(y)-f(x)\big) + \frac{x-y}{z-y}\big(f(z)-f(x)\big)$$

do que segue imediatamente que

$$\frac{f(y)-f(x)}{y-x} \;\leq\; \frac{f(z)-f(x)}{z-x}\;,$$

ou seja, $R(x,\ y) \leq R(x,\ z)$.

Caso 3: $y < z < x$. Como z fica entre y e x, podemos escrever $z = \lambda y + (1-\lambda)x$ adotando para tal $\lambda = \frac{x-z}{x-y}$ (notar que $\lambda \in (0,\ 1)$). Da convexidade de f segue, então, que

$$f(z) \;\leq\; \lambda f(y) + (1-\lambda)f(x) \;=\; \frac{x-z}{x-y}f(y) + \frac{z-y}{x-y}f(x)\;.$$

Subtraindo-se $f(x)$ de ambos os lados, obtém-se após cômputos elementares

$$\frac{f(y)-f(x)}{y-x} \;\leq\; \frac{f(z)-f(x)}{z-x}\;,$$

ou seja, novamente $R(x,\ y) \leq R(x,\ z)$. Com isso, (5.26) está estabelecida em todos os casos possíveis em que $y < z$.

Parte II: *supondo (5.26) provamos que f é convexa.* Por (5.26) sabemos que se $y < z$, então

$$\frac{f(x)-f(y)}{x-y} \;\leq\; \frac{f(x)-f(z)}{x-z}$$

para todo x, com $x \neq y$ e $x \neq z$. Tomemos, em particular, $x = \lambda y + (1-\lambda)z$, com $\lambda \in (0,\ 1)$. A última expressão fica

$$\frac{f\big(\lambda y + (1-\lambda)z\big) - f(y)}{(1-\lambda)(z-y)} \;\leq\; -\frac{f\big(\lambda y + (1-\lambda)z\big) - f(z)}{\lambda(z-y)}\;.$$

Cancelando-se o fator $z - y > 0$ de ambos os lados, obtemos $f\big(\lambda y + (1-\lambda)z\big) \leq \lambda f(y) + (1-\lambda)f(z)$, provando a convexidade de f. ■

E. 5.7 *Exercício*. A Proposição 5.9 possui a seguinte interpretação geométrica. Sejam $x,\ y,\ z \in I$ três pontos do domínio de definição de uma função f tais que $y < z$. Considere-se o segmento de linha reta L_1 conectando o ponto $\big(x,\ f(x)\big)$ ao ponto $\big(y,\ f(y)\big)$ (ambos no gráfico gráfico de f) e considere-se o segmento de linha reta L_2 conectando o ponto $\big(x,\ f(x)\big)$ ao ponto $\big(z,\ f(z)\big)$ (ambos também no gráfico de f). Então, o que a Proposição 5.9 afirma é que f é convexa se e somente se a inclinação de L_1 for menor ou igual à inclinação de L_2. Vide o gráfico no lado esquerdo da Figura 5.5, página 355. Justifique essa afirmativa com base na Proposição 5.9. ✠

• **Algumas desigualdades de interesse**

Antes de prosseguirmos apresentemos um resultado que será futuramente evocado nestas Notas.

Lema 5.8 *Seja $f : I \to \mathbb{R}$ convexa. Então, para quaisquer $w, x, y, z \in I$ com $w < x < y < z$ valem as desigualdades*

$$\frac{f(w) - f(x)}{w - x} \leq \frac{f(x) - f(y)}{x - y} \leq \frac{f(y) - f(z)}{y - z} \tag{5.27}$$

e

$$\frac{f(w) - f(x)}{w - x} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(x)}{z - x} . \tag{5.28}$$

□

Prova. Pelas hipóteses e pela Proposição 5.9, página 353, tem-se $R(w, x) \leq R(x, y) \leq R(y, z)$, assim como $R(x, w) \leq R(x, y) \leq R(x, z)$. Escrevendo-se explicitamente o que é a função R, obtemos disso (5.27) e (5.28), respectivamente. ■

O estudante deve atentar para as semelhanças e diferenças entre (5.27) e (5.28). A primeira desigualdade em ambas é a mesma. A diferença está na segunda desigualdade. A desigualdade (5.27) é ilustrada no gráfico à direita da Figura 5.5, página 355. A desigualdade (5.28) é ilustrada no gráfico à esquerda da Figura 5.5.

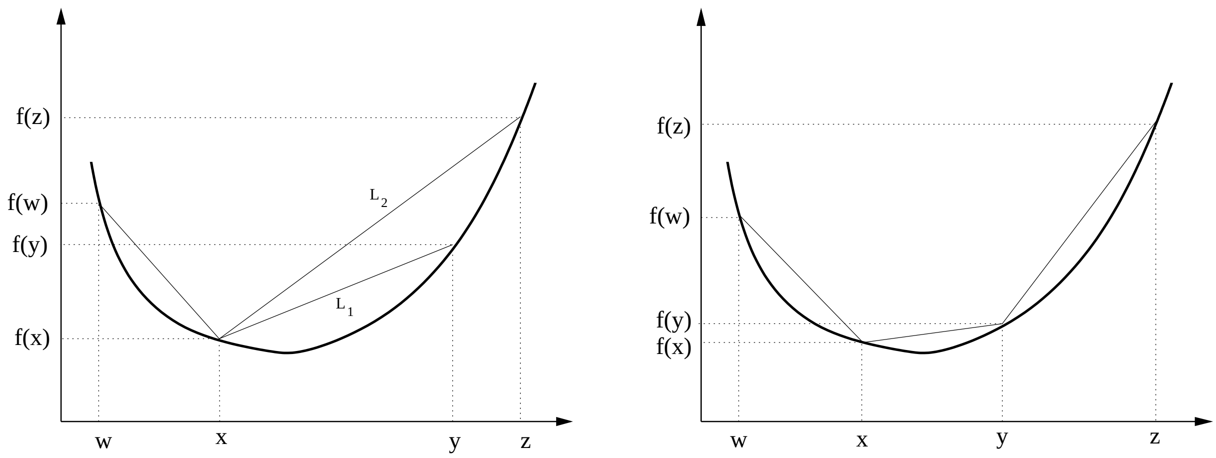

Figura 5.5: À esquerda: gráfico de uma função convexa. As inclinações dos segmentos de reta lá indicados são $R(w, x)$, $R(x, y)$ e $R(x, z)$. A figura deixa claro que $R(w, x) \leq R(x, y) \leq R(x, z)$. À direita: gráfico de uma função convexa. As inclinações dos segmentos de reta lá indicados são $R(w, x)$, $R(x, y)$ e $R(y, z)$. A figura deixa claro que $R(w, x) \leq R(x, y) \leq R(y, z)$.

• **Convexidade e derivadas laterais**

Seja uma função $g : I \to \mathbb{R}$. Para $x \in I^0$ definimos as derivadas laterais que denotamos por $g'_+(x)$ e $g'_-(x)$ por

$$g'_+(x) := \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \frac{g(x + \epsilon) - g(x)}{\epsilon} \qquad \text{e} \qquad g'_-(x) := \lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \frac{g(x) - g(x - \epsilon)}{\epsilon} ,$$

caso esses limites existam. É relevante notar que caso ambos os limites existam, então g é contínua em x, pois teremos $\lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \big(g(x + \epsilon) - g(x)\big) = 0$ e $\lim\limits_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \big(g(x) - g(x - \epsilon)\big) = 0$. É claro que g é diferenciável em x se e somente se $g'_+(x)$ e $g'_+(x)$ existirem em forem iguais.

A proposição que segue revela mais fatos básicos importantes sobre funções convexas. A informação sobre continuidade já fora estabelecida no Corolário 5.1, página 348.

Proposição 5.10 *Se $f : I \to \mathbb{R}$ é convexa, então f é contínua em I^0 e possui em cada ponto de I^0 derivadas laterais à direita e à esquerda, as quais satisfazem a seguinte desigualdade:*

$$f'_-(x) \leq f'_+(x) \leq \frac{f(x') - f(x)}{x' - x} \leq f'_-(x') \leq f'_+(x') \tag{5.29}$$

para todos $x, x' \in I^0$ com $x < x'$. Isso diz que tanto f'_- quanto f'_+ são funções monotonamente não decrescentes em I^0, sendo que $f'_- \leq f'_+$ em todo I^0. Outra afirmação que disso pode ser extraída é que f é não diferenciável em uma coleção no máximo enumerável de pontos. □

Uma demonstração mais geral da continuidade de funções convexas no interior do seu domínio de definição será apresentada na Proposição 5.7, página 346.

Prova da Propossição 5.10. Sejam $w, x, y \in I^0$ com $w < x < y$. Pela Proposição 5.9, página 353, temos $R(x, w) \leq R(x, y)$.

Fixemos w e x. Sabemos, também pela Proposição 5.9, que a função $y \mapsto R(x, y)$ definida para $y > x$ é decrescente quando y diminui para x. Assim, o limite $\lim_{\substack{y \to x \\ y > x}} R(x, y)$ existe, por ser decrescente e limitado inferiormente por $R(x, w)$. Sucede que, pela definição de R, o limite $\lim_{\substack{y \to x \\ y > x}} R(x, y)$ é precisamente a derivada lateral à direita $f'_+(x)$.

Fixemos x e y. Sabemos, também pela Proposição 5.9, que a função $w \mapsto R(x, w)$ definida para $w > x$ é crescente quando w cresce para x. Assim, o limite $\lim_{\substack{w \to x \\ w < x}} R(x, w)$ existe, por ser crescente e limitado superiormente por $R(x, y)$. Sucede que, pela definição de R, o limite $\lim_{\substack{w \to x \\ w < x}} R(x, w)$ é precisamente a derivada lateral à esquerda $f'_-(x)$.

Isso estabeleceu a existência dos limites laterais para todo ponto de I^0 e estabeleceu que f é contínua em todo ponto de I^0.

Sejam agora w, x, y, w', x', y' seis pontos de I^0 tais que $w < x < y < w' < x' < y'$. Fazendo uso da Proposição 5.9, temos

$$R(w, x) \leq R(x, y) \leq R(y, w') \leq R(w', x') \leq R(x', y') ,$$

ou seja,

$$\frac{f(x) - f(w)}{x - w} \leq \frac{f(y) - f(x)}{y - x} \leq \frac{f(w') - f(y)}{w' - y} \leq \frac{f(x') - f(w')}{x' - w'} \leq \frac{f(y') - f(x')}{y' - x'} . \tag{5.30}$$

Tomando-se em (5.30) os limites $\lim_{\substack{w \to x \\ w < x}}$, $\lim_{\substack{y \to x \\ y > x}}$, $\lim_{\substack{w' \to x' \\ w' < x'}}$, $\lim_{\substack{y' \to x' \\ y' > x'}}$ e usando-se a continuidade de f, obtemos

$$f'_-(x) \leq f'_+(x) \leq \frac{f(x') - f(x)}{x' - x} \leq f'_-(x') \leq f'_+(x') ,$$

que é (5.29).

Seja $N \subset I^0$ a coleção de todos os pontos de I^0 nos quais f não seja diferenciável, ou seja, nos quais $f'_-(x) \neq f'_+(x)$. (5.29) informa-nos que se $x, x' \in N$ com $x < x'$, então os intervalos $\big(f'_-(x), f'_+(x)\big)$ e $\big(f'_-(x'), f'_+(x')\big)$ são intervalos disjuntos de $\mathbb{R}$. Agora, $\mathbb{R}$ pode no máximo admitir uma família enumerável de intervalos disjuntos. Logo, N é no máximo enumerável. ■

É instrutivo chamar a atenção do leitor para o fato de a continuidade a que se refere a Proposição 5.10 ser garantida apenas no interior I^0 do domínio de definição I. As funções $f : [0, 1] \to \mathbb{R}$ definidas por

$$f(x) := \begin{cases} 1, & x = 0, \\ x^2, & x \in (0, 1], \end{cases} \qquad f(x) := \begin{cases} 1, & x = 0, \\ 0, & x \in (0, 1], \end{cases}$$

são convexas em todo o intervalo $[0, 1]$, mas não são contínuas em $x = 0$.

- **Retas suportes**

Seja $f : I \to \mathbb{R}$ uma função convexa de uma variável. Uma função afim $r(x) = \alpha x + \beta$, com $\alpha, \beta \in \mathbb{R}$, é dita *suportar* f no ponto $x_0 \in I$ se valer $f(x_0) = r(x_0)$ e $f(x) \geq r(x)$ para todo $x \in I$.

Se r suporta f em x_0 a reta $\{(x,\ r(x)),\ x \in \mathbb{R}\} \subset \mathbb{R}^2$ é dita ser uma *reta suporte* de f nesse ponto. Geometricamente, podemos afirmar que o gráfico de uma função convexa está sempre acima de suas retas suportes e uma reta suporte tangencia o gráfico de f em x_0.

Se a função afim $r(x) = \alpha x + \beta$ suporta f em x_0, a condição $r(x_0) = f(x_0)$ implica que podemos escrever

$$r(x) \;=\; \alpha(x - x_0) + f(x_0) \; .$$

Há uma noção similar para funções côncavas, com a correspondente inversão da desigualdade.

O Corolário 5.3, a seguir, afirma que toda função convexa possui, em cada ponto de seu domínio, ao menos uma reta suporte. Esse corolário é consequência da Proposição 5.10, página 356, e o usaremos quando apresentarmos a demonstração de uma forma geral da desigualdade de Jensen na Proposição 5.19, página 366.

Corolário 5.3 *Se $f : I \to \mathbb{R}$ é convexa, então*

$$f(y) \;\geq\; (y - x)f'_{\pm}(x) + f(x) \tag{5.31}$$

para <u>todos</u> $x,\ y \in I_0$. □

Esse corolário tem a seguinte interpretação geométrica: o gráfico de uma função convexa está sempre acima das retas tangentes ao mesmo (e isso é verdade mesmo em pontos em que a derivada é descontínua, em cujo caso temos duas retas tangentes com inclinações f'_- e f'_+).

Prova do Corolário 5.3. Sejam $x,\ x' \in I_0$. Para $x' > x$ temos de (5.29) as desigualdades $f'_-(x) \leq f'_+(x) \leq \frac{f(x')-f(x)}{x'-x}$, que implicam que

$$f(x') \;\geq\; (x' - x)f'_{\pm}(x) + f(x) \; , \qquad x' \geq x \; . \tag{5.32}$$

Incluímos acima o caso $x' = x$ devido à continuidade de f (também demonstrada na Proposição 5.10).

Também para $x' > x$ temos de (5.29) as desigualdades $\frac{f(x')-f(x)}{x'-x} \leq f'_-(x') \leq f'_+(x')$, que implicam que

$$f(x) \;\geq\; (x - x')f'_{\pm}(x') + f(x') \; , \qquad x' > x \; .$$

Trocando as letras $x \leftrightarrow x'$, isso fica

$$f(x') \;\geq\; (x' - x)f'_{\pm}(x) + f(x) \; , \qquad x > x' \; . \tag{5.33}$$

Contemplando (5.32) e (5.33), vemos que estabelecemos que $f(x') \geq (x' - x)f'_{\pm}(x) + f(x)$ para <u>todos</u> $x,\ x' \in I_0$. ■

• Mais condições para convexidade

A Proposição 5.10, página 356, afirma que se uma função de uma variável é convexa, então suas derivadas laterais são crescentes. Sob hipóteses adequadas é possível garantir a recíproca dessa afirmação. A proposição que segue mostra a forma mais simples dessa recíproca.

Proposição 5.11 *Seja $f : I \to \mathbb{R}$ contínua em I e diferenciável em I^0. Então, uma condição necessária e suficiente para que f seja convexa é que f' seja monotonamente não decrescente, ou seja, que $f'(x) \leq f'(y)$ para todos $x,\ y \in I^0$ com $x \leq y$.* □

Prova. Da Proposição 5.10, página 356, é evidente que convexidade e diferenciabilidade implicam que f' é monotonamente não decrescente, de modo que resta apenas provar a recíproca.

Sejam $x_0,\ x_1 \in I$ com $x_0 < x_1$ e seja $\lambda \in (0,\ 1)$. Definamos $x_\lambda := (1 - \lambda)x_0 + \lambda x_1$. É claro que $x_0 < x_\lambda < x_1$. Com as hipóteses, podemos evocar o Teorema do Valor Médio e afirmar que existem $\xi_0 \in (x_0,\ x_\lambda)$ e $\xi_1 \in (x_\lambda,\ x_1)$ tais que valem

$$\begin{aligned} f(x_\lambda) - f(x_0) \;&=\; f'(\xi_0)\big(x_\lambda - x_0\big) \;=\; \lambda(x_1 - x_0)f'(\xi_0) \; , \\[2mm] f(x_1) - f(x_\lambda) \;&=\; f'(\xi_1)\big(x_1 - x_\lambda\big) \;=\; (1 - \lambda)(x_1 - x_0)f'(\xi_1) \; . \end{aligned}$$

Note-se que $\xi_0 < \xi_1$ e, portanto, $f'(\xi_0) \leq f'(\xi_1)$. Temos, assim,

$$f(x_\lambda) - (1-\lambda)f(x_0) - \lambda f(x_1) \;=\; (1-\lambda)\big(f(x_\lambda) - f(x_0)\big) + \lambda\big(f(x_\lambda) - f(x_1)\big) \;=\; \lambda(1-\lambda)(x_1 - x_0)\big(f'(\xi_0) - f'(\xi_1)\big) \;\leq\; 0 ,$$

pois $f'(\xi_0) \leq f'(\xi_1)$. Isso estabeleceu que $f\big(\lambda x_0 + (1-\lambda)x_1\big) \leq (1-\lambda)f(x_0) + \lambda f(x_1)$ com $\lambda \in (0,\ 1)$. Para $\lambda \in \{0,\ 1\}$ essa relação é trivial e isso demonstra a convexidade de f. ■

O seguinte corolário importante é agora evidente e dispensa demonstração.

Corolário 5.4 *Seja $f : I \to \mathbb{R}$ contínua em I e duas vezes diferenciável em I^0. Então, uma condição necessária e suficiente para que f seja convexa é que $f''(x) \geq 0$ para todo $x \in I^0$.* □

Esse corolário permite confortavelmente determinar se uma função f contínua e duas vezes diferenciável em seu domínio de definição I é convexa (o que se dá caso $f''(x) \geq 0$ para todo $x \in I^0$) ou côncava (o que se dá caso $f''(x) \leq 0$ para todo $x \in I^0$).

Exemplos 5.6 De posse do critério estabelecido no Corolário 5.4, é fácil demonstrar os seguintes fatos. A função $\ln x$ é côncava em $(0,\ \infty)$. As funções $e^{\pm x}$ são convexas em $\mathbb{R}$. As funções x^n com $n \in \mathbb{N}$ par são convexas em $\mathbb{R}$. As funções x^n com $n \in \mathbb{N}$ ímpar são convexas em $[0,\ \infty)$ e côncavas em $(-\infty,\ 0]$. A função $1/(1-x^2)$ é convexa no intervalo $(-1,\ 1)$ e diverge para $x \to \pm 1$. A função $x + 1/x$ é convexa em $(0,\ \infty)$ e côncava em $(-\infty,\ 0)$. As funções $\cos x$ e $\operatorname{sen} x$ são côncavas nos intervalos em que são positivas e convexas nos intervalos em que são negativas. ♦

E. 5.8 *Exercício (fácil).* Considere-se a função $s : [0,\ 1] \to \mathbb{R}$ dada por

$$s(x) \;:=\; \begin{cases} 0 , & \text{se } x = 0 , \\ -x\ln(x) , & \text{se } x \in (0,\ 1] . \end{cases}$$

Constate que $s(x) \geq 0$ para todo $x \in [0,\ 1]$, que s é uma função contínua, duas vezes diferenciável em $(0,\ 1]$ e côncava. ✠

E. 5.9 *Exercício (fácil).* Constate que as funções cos e sen são côncavas (convexas) nos intervalos em que são positivas (negativas). *Sugestão:* use o fato que ambas satisfazem a equação diferencial $y''(x) + y(x) = 0$. ✠

E. 5.10 *Exercício.* Mostre que a função $\ell : [0,\ \infty) \to [0,\ 1)$ definida por $\ell(x) = x/(1+x)$ é crescente e côncava. Sugestão: mostre que $\ell'(x) > 0$ e $\ell''(x) < 0$ para todo $x \in [0,\ \infty)$. Para $n \in \mathbb{N}$, mostre que $\ell^{n\circ} := \underbrace{\ell \circ \cdots \circ \ell}_{n \text{ vezes}}$ é a função $\ell^{n\circ}(x) = x/(1+nx)$ e mostre explicitamente que essa função é também crescente e côncava, estudando suas primeiras derivadas. Alternativamente, evoque o Lema 5.7, página 353. ✠

A Proposição 5.11 tem ainda um outro corolário digno de nota:

Corolário 5.5 *Seja $I \subset \mathbb{R}$ um intervalo convexo do tipo que aqui consideramos e seja $g : I \to \mathbb{R}$ uma função contínua, e não decrescente. Então, a função $f : I \to \mathbb{R}$ definida por*

$$f(x) \;:=\; \int_a^x g(u)\, du$$

com $a \in I$, fixo, é uma função convexa da variável x. □

Prova. É claro que f é contínua e que $f'(x) = g(x)$ para todo $x \in I$. Logo, f' é estritamente crescente e, pela Proposição 5.11, página 357, f é convexa. ■

• **A condição do ponto médio**

Vamos agora apresentar mais uma caracterização de funções convexas e contínuas, a qual possui diversas aplicações. O que mostraremos é que uma função contínua $f : I \to \mathbb{R}$ é convexa se e somente se satisfizer a desigualdade $f\left(\frac{x+y}{2}\right) \leq$

$\frac{f(x)+f(y)}{2}$ para todos $x, y \in I$. Como comentaremos, foi essa caracterização que deu origem histórica à teoria das funções convexas.

Começamos com o seguinte resultado, cuja demonstração é assaz interessante:

Lema 5.9 *Seja $f : I \to \mathbb{R}$ tal que*

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} \tag{5.34}$$

para todos $x, y \in I$. Então, vale

$$f\left(\frac{x_1+\cdots+x_n}{n}\right) \leq \frac{f(x_1)+\cdots+f(x_n)}{n} \tag{5.35}$$

para todo $n \in \mathbb{N}$ e para todos $x_1, \ldots, x_n \in I$. □

Prova. A prova é feita seguindo uma curiosa estratégia de indução[16]: primeiramente mostramos que se a proposição vale para n, então vale para todo número da forma $2^k n$ com $k \in \mathbb{N}$ (indução para a frente!). Em seguida, provamos que se a proposição vale para m ela vale para $m-1$ (indução para trás!). Com isso, todos os naturais são varridos pelo procedimento indutivo.

Vamos assumir (5.35) válida para algum $n \in \mathbb{N}$ (ela vale para $n = 2$, por hipótese). Sejam $x_1, \ldots, x_{2n} \in I$. Então,

$$f\left(\frac{x_1+\cdots+x_{2n}}{2n}\right) = f\left(\frac{\frac{x_1+x_2}{2}+\cdots+\frac{x_{2n-1}+x_{2n}}{2}}{n}\right) \overset{\text{hipótese}}{\leq} \frac{1}{n}\left(f\left(\frac{x_1+x_2}{2}\right)+\cdots+f\left(\frac{x_{2n-1}+x_{2n}}{2}\right)\right)$$

$$\overset{(5.34)}{\leq} \frac{f(x_1)+\cdots+f(x_{2n})}{2n} .$$

Assim, se (5.35) vale para n, vale também para $2n$ e, consequentemente, para todo número da forma $2^k n$ com $k \in \mathbb{N}$. Isso prova os passos indutivos para a frente.

Vamos agora assumir que (5.35) valha para algum $n \geq 4$. Seja $x_1, \ldots, x_{n-1} \in I$ e defina-se $x_n := \frac{1}{n-1}\big(x_1+\cdots+x_{n-1}\big)$. É evidente que $x_n \in I$ e teremos,

$$\frac{x_1+\cdots+x_n}{n} = \frac{x_1+\cdots+x_{n-1}}{n-1} .$$

Verifique! Logo,

$$\begin{aligned} f\left(\frac{x_1+\cdots+x_{n-1}}{n-1}\right) &= f\left(\frac{x_1+\cdots+x_n}{n}\right) \\ &\overset{\text{hipótese}}{\leq} \frac{f(x_1)+\cdots+f(x_n)}{n} \\ &= \frac{f(x_1)+\cdots+f(x_{n-1})}{n} + \frac{1}{n}f\left(\frac{x_1+\cdots+x_{n-1}}{n-1}\right) . \end{aligned}$$

Passando o termo $\frac{1}{n}f\left(\frac{x_1+\cdots+x_{n-1}}{n-1}\right)$ para o lado esquerdo da desigualdade, obtemos imediatamente que

$$f\left(\frac{x_1+\cdots+x_{n-1}}{n-1}\right) = \frac{f(x_1)+\cdots+f(x_{n-1})}{n-1} ,$$

provando que (5.35) vale para $n-1$. Isso prova os passos de indução retrógrada e completa a demonstração. ■

O resultado seguinte apresenta uma caracterização muito útil de funções convexas:

[16]Essa estratégia foi inventada por Cauchy para demonstrar (5.54) e foi empregada no presente contexto por Jensen (para a referência original, vide nota de rodapé 17 à página 360). O texto original de Cauchy é *Cours d'analyse de l'Ecole Royale Polytechnique, premier partie, Analyse algebrique*, Paris 1821.

Proposição 5.12 *Seja $f : I \to \mathbb{R}$, contínua. Então, f é convexa em I se e somente se satisfizer*

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2} \tag{5.36}$$

para todos $x, \ y \in I$. □

Prova. Se f for convexa em I, (5.36) é um caso particular da definição de convexidade (tome-se $\lambda = 1/2$). Vamos provar que se f é contínua em I e lá satisfaz (5.36), então f é convexa em I.

Sejam $a_1, \ a_2 \in \mathbb{N}$ dois números naturais. Sejam também $x_1, \ x_2 \in I$. É claro que $\big(a_1x_1 + a_2x_2\big)/\big(a_1 + a_2\big)$ é um elemento de I. Como

$$a_1x_1 + a_2x_2 = \underbrace{x_1 + \cdots + x_1}_{a_1 \text{vezes}} + \underbrace{x_2 + \cdots + x_2}_{a_2 \text{vezes}}$$

é uma soma de $a_1 + a_2$ elementos de I, podemos evocar o Lema 5.9, página 359, em particular, a relação (5.35) com $n = a_1 + a_2$, e escrever

$$f\left(\frac{a_1x_1 + a_2x_2}{a_1 + a_2}\right) \leq \frac{\overbrace{f(x_1) + \cdots + f(x_1)}^{a_1 \text{vezes}} + \overbrace{f(x_2) + \cdots + f(x_2)}^{a_2 \text{vezes}}}{a_1 + a_2} = \frac{a_1 f(x_2) + a_2 f(x_2)}{a_1 + a_2} . \tag{5.37}$$

Afirmamos que isso implica que se $r \in \mathbb{Q} \cap [0, \ 1]$, então

$$f\big(rx_1 + (1-r)x_2\big) \leq rf(x_1) + (1-r)f(x_2) . \tag{5.38}$$

É suficiente tomarmos $r \in \mathbb{Q} \cap (0, \ 1)$. Como tal, r é da forma $r = p/q$ com $0 < p < q$, sendo $p, \ q \in \mathbb{N}$. Se definirmos $a_1 := p$ e $a_2 := q - p$, teremos $a_1, \ a_2 \in \mathbb{N}$ e poderemos escrever $r = \frac{a_1}{a_1+a_2}$ e $1 - r = \frac{a_2}{a_1+a_2}$. Com essas observações torna-se evidente que a validade de (5.38) segue de (5.37).

Em (5.38), façamos agora o racional r convergir a $\lambda \in [0, \ 1]$, arbitrário. A continuidade da f implica que teremos $f\big(\lambda x_1 + (1-\lambda)x_2\big) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$. Como isso é válido para qualquer $\lambda \in [0, \ 1]$ e quaisquer $x_1, \ x_2 \in I$, a convexidade de f fica estabelecida. ■

Nota histórica. A noção de função convexa foi introduzida por Jensen[17] em 1906. A definição original de Jensen para função convexa era a relação (5.36). Assumindo continuidade para f, Jensen então seguiu os passos que apresentamos na demonstração da Proposição 5.12 e demonstrou (5.19) assim como (5.21) (que ficou conhecida como *desigualdade de Jensen*). ♣

Comentamos, por fim, que a condição de continuidade não pode ser dispensada da Proposição 5.12. Com o uso de bases de Hamel (vide discussão à página 199 e seguintes e, em particular, vide a discussão que sucede a Proposição 2.16, página 201) é possível construir funções não contínuas em $\mathbb{R}$ satisfazendo (5.36) para todos $x, \ y \in \mathbb{R}$ mas que não são convexas (pois seu gráfico é denso em todo $\mathbb{R}^2$ (!)).

• Mais duas caracterizações de funções contínuas convexas

O lema a seguir contém uma informação útil que será evocada adiante.

Lema 5.10 *Seja $f : I \to \mathbb{R}$ uma função convexa. Então, dado qualquer intervalo compacto $[a, \ b] \subset I$ (com $-\infty < a < b < \infty$), o máximo de f nesse intervalo é alcançado em um dos seus pontos extremos, a ou b, ou seja,* $\max\{f(x), \ x \in [a, \ b]\} = \max\{f(a), \ f(b)\}$. □

Prova. Para todo $\lambda \in [0, \ 1]$ tem-se, evidentemente, $\lambda f(a) + (1-\lambda)f(b) \leq \max\{f(a), \ f(b)\}$. Se $x \in [a, \ b]$, então $x = \lambda_0 a + (1-\lambda_0)b$ para algum $\lambda_0 \in [0, \ 1]$. Logo, $f(x_0) = f\big(\lambda_0 a + (1-\lambda_0)b\big) \leq \lambda_0 f(a) + (1-\lambda_0)f(b) \leq \max\{f(a), \ f(b)\}$. Claro é que essa desigualdade $f(x_0) \leq \max\{f(a), \ f(b)\}$ é uma igualdade se x_0 for igual a a ou b. ■

[17] Johan Ludwig William Valdemar Jensen (1859–1925). O trabalho original de Jensen é: J. L. W. V. Jensen, *"Sur les fonctions convexes et les inégalités entre les valeurs moyennes"*, Acta Math. **30**, 175–193 (1906).

O estudante deve notar que a recíproca do Lema 5.10 não é verdadeira: a função cosseno restrita ao intervalo $I = [0,\ 2\pi]$ é contínua e satisfaz $\max\{\cos(x),\ x \in [a,\ b]\} = \max\{\cos(a),\ \cos(b)\}$ para todo compacto $[a,\ b] \subset [0,\ 2\pi]$. Mas a função cosseno não é convexa em $[0,\ 2\pi]$. Em verdade, toda função f definida em um intervalo fechado $[A,\ B]$, contínua, sem nenhum máximo local exceto, eventualmente, os pontos A e B satisfaz $\max\{f(x),\ x \in [a,\ b]\} = \max\{f(a),\ f(b)\}$ para todo $[a,\ b] \subset [A,\ B]$.

A proposição que segue, mencionada em [225], apresenta uma caracterização de convexidade para funções contínuas. Ela mostra o que é necessário supor adicionalmente para que se obtenha uma recíproca à afirmação do Lema 5.10.

Proposição 5.13 *Seja $f : I \to \mathbb{R}$ uma função contínua. Então, f é convexa se e somente se para todo $\alpha \in \mathbb{R}$ o máximo da função $\phi_\alpha(x) := f(x) + \alpha x$ em um intervalo compacto arbitrário $[a,\ b] \subset I$ for sempre assumido em um dos extremos do mesmo. Em outras palavras, f é convexa se e somente se para todo $\alpha \in \mathbb{R}$ valer $\max\{\phi_\alpha(x),\ x \in [a,\ b]\} = \max\{\phi_\alpha(a),\ \phi_\alpha(b)\}$ em todo compacto $[a,\ b] \in I$.* □

Prova. É trivial constatar que $\phi_\alpha(x) := f(x) + \alpha x$ é convexa se e somente se f o for. Logo, pelo Lema 5.10, página 360, os máximos da função ϕ_α em um intervalo compacto arbitrário $[a,\ b] \subset I$ são assumidos em um dos extremos do mesmo.

Vamos agora supor a recíproca, ou seja, que os máximos da função ϕ_α em um intervalo compacto arbitrário $[a,\ b] \subset I$ são assumidos em um dos extremos do mesmo. Assim, temos $f(x) + \alpha x \leq M$ para todo $x \in [a,\ b]$, onde $M := \max\{f(a) + \alpha a,\ f(b) + \alpha b\}$. Escrevamos $x \in [a,\ b]$ na forma $x = \lambda a + (1-\lambda)b$ com $\lambda \in [a,\ b]$. Temos, portanto, $f\big(\lambda a + (1-\lambda)b\big) + \alpha\big(\lambda a + (1-\lambda)b\big) \leq M$, ou seja,

$$f\big(\lambda a + (1-\lambda)b\big) \ \leq \ M - \alpha\big(\lambda a + (1-\lambda)b\big) \ .$$

Subtraindo $\lambda f(a) + (1-\lambda)f(b)$ de ambos os lados dessa desigualdade, teremos

$$f\big(\lambda a + (1-\lambda)b\big) - \lambda f(a) - (1-\lambda)f(b) \ \leq \ M - \lambda\big(f(a) + \alpha a\big) - (1-\lambda)\big(f(b) + \alpha b\big) \ .$$

Observe-se agora que essa desigualdade deve, por hipótese, ser verdadeira para todo $\alpha \in \mathbb{R}$ e que seu lado esquerdo independe de α. Se escolhermos $\alpha = \big(f(a) - f(b)\big)/(b-a)$, teremos $f(a) + \alpha a = f(b) + \alpha b = M$. Nesse caso, o lado direito da desigualdade anula-se e concluímos que $f\big(\lambda a + (1-\lambda)b\big) \leq \lambda f(a) + (1-\lambda)f(b)$. Como a e b são arbitrários em I (com $a < b$) e $\lambda \in [0,\ 1]$ é igualmente arbitrário, isso estabeleceu que f é convexa em todo I. ■

A proposição que segue, também mencionada em [225], apresenta mais uma condição equivalente à convexidade para funções contínuas.

Proposição 5.14 *Seja $f : I \to \mathbb{R}$ uma função contínua. Então, f é convexa se e somente se satisfizer*

$$f(x) \ \leq \ \frac{1}{2h}\int_{x-h}^{x+h} f(t)\, dt \ , \tag{5.39}$$

para todo $x \in I^0$ e todo $h > 0$ tal que $x \pm h \in I$. □

Comentário. A Proposição 5.14 está na raíz da definição das chamadas *funções sub-harmônicas*, tema do qual não trataremos aqui. ♣

Prova da Proposição 5.14. Se f é contínua e convexa então, por (5.34), vale $f(x) \leq \frac{1}{2}\big(f(x+t) + f(x-t)\big)$ para todos $x \in I^0$ e $x \pm t \in I$. Integrando-se ambos os lados t na variável t com t no intervalo $[0,\ h]$, teremos $hf(x) \leq \frac{1}{2}\int_{-h}^{h} f(x+t)\, dt$, como facilmente se vê, e isso equivale a (5.39).

Vamos agora supor que f seja contínua e satisfaça (5.39). Seja $\alpha \in \mathbb{R}$. É elementar constatar que para todo $h > 0$ vale $\alpha x = \frac{1}{2h}\int_{x-h}^{x+h} \alpha t\, dt$. Logo,

$$\phi_\alpha(x) \ \leq \ \frac{1}{2h}\int_{x-h}^{x+h} \phi_\alpha(t)\, dt \ , \tag{5.40}$$

com $\phi_\alpha(x) := f(x) + \alpha x$. Seja $[a,\ b] \subset I$ um intervalo compacto e seja um intervalo $[x_0 - h,\ x_0 + h] \subset [a,\ b]$ com as seguintes propriedades: 1º $\phi_\alpha(y) \leq \phi_\alpha(x_0)$ para todo $y \in [x_0 - h,\ x_0 + h]$ e 2º existe $y_0 \in [x_0 - h,\ x_0 + h]$ tal que $\phi_\alpha(y_0) < \phi_\alpha(x_0)$. Como ϕ_α é contínua, teremos $\frac{1}{2h}\int_{x_0-h}^{x_0+h} \phi_\alpha(t)\, dt < \phi_\alpha(x_0)$. Isso contraria (5.40) e, portanto, devemos ter $\phi_\alpha(y) \geq \phi_\alpha(x_0)$ para todo $y \in [x_0 - h,\ x_0 + h]$. Como isso vale para todo intervalo $[x_0 - h,\ x_0 + h] \subset [a,\ b]$, segue que $\max\{\phi_\alpha(x),\ x \in [a,\ b]\} = \max\{\phi_\alpha(a),\ \phi_\alpha(b)\}$. Pela Proposição 5.13, página 361, isso implica que f é convexa. ■

• A desigualdade de Hermite-Hadamard

Proposição 5.15 *Seja $f : I \to \mathbb{R}$ uma função convexa. Então, f satisfaz*

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a}\int_a^b f(t)\,dt \leq \frac{f(a)+f(b)}{2} \tag{5.41}$$

para todos $a,\ b \in I^0$ com $a < b$. □

A desigualdade (5.41) é conhecida como *desigualdade de Hermite*[18]*-Hadamard*[19], ou simplesmente como *desigualdade de Hadamard*.

Prova da Proposição 5.15. Tomando $h = (b-a)/2$ e $x = (a+b)/2$ em (5.39), obtemos a primeira desigualdade $f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a}\int_a^b f(t)\,dt$. Com a mudança de variáveis $t = \lambda(b-a)+a = \lambda b + (1-\lambda)a$ para $\lambda \in [0,\ 1]$, temos, pela convexidade de f,

$$\frac{1}{b-a}\int_a^b f(t)\,dt = \int_0^1 f\big(\lambda b + (1-\lambda)a\big)\,d\lambda \overset{\text{convex.}}{\leq} \int_0^1 \Big[f(b)\lambda + f(a)(1-\lambda)\Big]\,d\lambda = \frac{f(a)+f(b)}{2},$$

completando a prova. ■

• Convexidade ou concavidade de funções definidas por integrais

Apresentamos, por fim, um resultado útil sobre a convexidade ou concavidade de funções definidas por integrais. Esse resultado pode ser provado a partir da Proposição 5.11, página 357, mas apresentamos aqui uma demonstração que faz uso de menos recursos.

Proposição 5.16 *Para $a < b$, seja $g : [a,\ b] \to \mathbb{R}$ uma função contínua por partes. Seja $f(x) := \int_a^x g(s)\,ds$, $x \in [a,\ b]$. Então, f é convexa em intervalos onde g é não descrescente e côncava em intervalos onde g é não crescente.* □

Prova. Para $a \leq x < x' \leq b$ e $\lambda \in [0,\ 1]$, defina-se $x_\lambda := \lambda x + (1-\lambda)x' \in [x,\ x']$. tem-se

$$f\big(\lambda x + (1-\lambda)x'\big) \equiv f(x_\lambda) = \int_a^{x_\lambda} g(s)\,ds = \int_a^{x'} g(s)\,ds - \int_{x_\lambda}^{x'} g(s)\,ds. \tag{5.42}$$

Tem-se também

$$\lambda f(x) + (1-\lambda)f(x') = f(x') - \lambda\big(f(x') - f(x)\big) = \int_a^{x'} g(s)\,ds - \lambda\int_x^{x'} g(s)\,ds.$$

Portanto,

$$\int_a^{x'} g(s)\,ds = \lambda f(x) + (1-\lambda)f(x') + \lambda\int_x^{x'} g(s)\,ds.$$

Retornando com isso a (5.42), temos

$$f\big(x_\lambda\big) = \lambda f(x) + (1-\lambda)f(x') + D_\lambda, \tag{5.43}$$

onde

$$D_\lambda \equiv \lambda\int_x^{x'} g(s)\,ds - \int_{x_\lambda}^{x'} g(s)\,ds.$$

[18] Charles Hermite (1822–1901).
[19] Jacques Salomon Hadamard (1865–1963).

Vamos agora assumir que g é não decrescente no intervalo $[x,\ x']$, ou seja, se $g(t) \leq g(t')$ sempre que $x \leq t < t' \leq x'$. Desejamos provar que $D_\lambda \leq 0$. Escrevemos,

$$D_\lambda \;=\; \lambda\left[\int_x^{x'} g(s)\,ds - \int_{x_\lambda}^{x'} g(s)\,ds\right] - (1-\lambda)\int_{x_\lambda}^{x'} g(s)\,ds \;=\; \lambda\int_x^{x_\lambda} g(s)\,ds - (1-\lambda)\int_{x_\lambda}^{x'} g(s)\,ds\ .$$

Agora, para g não decrescente,

$$\int_x^{x_\lambda} g(s)\,ds \;\leq\; g(x_\lambda)(x_\lambda - x) \qquad \text{e} \qquad \int_{x_\lambda}^{x'} g(s)\,ds \;\geq\; g(x_\lambda)(x' - x_\lambda)\ . \tag{5.44}$$

Com isso, vemos que

$$D_\lambda \;\leq\; \lambda g(x_\lambda)(x_\lambda - x) - (1-\lambda)g(x_\lambda)(x' - x_\lambda) \;=\; g(x_\lambda)\underbrace{\Big[\lambda(x_\lambda - x) - (1-\lambda)(x' - x_\lambda)\Big]}_{=0} \;=\; 0\ , \tag{5.45}$$

pela definição de x_λ. Assim, por (5.43), fica estabelecido que $f\big(\lambda x + (1-\lambda)x'\big) \leq \lambda f(x) + (1-\lambda)f(x')$, ou seja, que f é convexa. Para provar que f é côncava nos intervalos onde g é não crescente, pode-se proceder de forma análoga (nesse caso as desigualdades (5.44) são revertidas, assim como a de (5.45)) ou pode-se simplesmente considerar o caso acima para g substituída por $-g$ e f por $-f$. ■

5.2.4 Funções Convexas de Várias Variáveis

Vamos agora estender alguns dos resultados anteriores para funções convexas de um número finito de variáveis reais, ou seja, definidas em um subconjunto convexo de algum $\mathbb{R}^n$.

• **Funções convexas em $\mathbb{R}^n$**

No que segue, consideraremos conjuntos convexos $C \subset \mathbb{R}^n$ que sejam também de dimensão n, isto é, que sejam tais que o menor subespaço que contém C seja $\mathbb{R}^n$.

Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \to \mathbb{R}$ é dita ser uma *função convexa* se para todos $x,\ y \in C$ e todo $\lambda \in [0,\ 1]$ valer a desigualdade

$$f\big(\lambda x + (1-\lambda)y\big) \;\leq\; \lambda f(x) + (1-\lambda)f(y)\ . \tag{5.46}$$

Uma função $f : C \to \mathbb{R}$ é dita ser uma *função côncava* se para todos $x,\ y \in C$ e todo $\lambda \in [0,\ 1]$ valer a desigualdade

$$f\big(\lambda x + (1-\lambda)y\big) \;\geq\; \lambda f(x) + (1-\lambda)f(y)\ . \tag{5.47}$$

É elementar constatar que f é côncava se e somente se $-f$ for convexa, sendo, portanto, suficiente estudar propriedades gerais de funções convexas.

Com já vimos no Corolário 5.1, página 348, se $f : C \to \mathbb{R}$ é convexa ou côncava, então f é contínua em C^0, o interior de C.

Conforme já mencionanos no Lema 5.6, página 352, se $f : C \to \mathbb{R}$ é uma função convexa, então ela é uma função convexa de cada uma de suas variáveis individualmente: tomemos, por exemplo, $x,\ y \in C$ na forma $x \equiv (x_1,\ x_2,\ \ldots,\ x_n)$ e $y \equiv (y_1,\ x_2,\ \ldots,\ x_n)$. Então, a convexidade de f implica que $f\big(\lambda x_1 + (1-\lambda)y_1,\ x_2,\ \ldots,\ x_n\big) \leq \lambda f\big(x_1,\ x_2,\ \ldots,\ x_n\big) + (1-\lambda)f\big(y_1,\ x_2,\ \ldots,\ x_n\big)$, que corresponde à afirmação que f é convexa enquanto função de sua primeira variável. Para as demais variáveis individualmente tem-se o mesmo.

A recíproca, porém, não é necessariamente verdadeira: se f é convexa em cada uma de suas variáveis independentemente, ela não é necessariamente convexa enquanto função de suas n variáveis. Como contraexemplo, tome-se a função $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x_1,\ x_2) := x_1 x_2$. É fácil ver que f é convexa como função de x_1 e de x_2 separadamente, mas ela não pode satizfazer (5.46) se, por exemplo, tomarmos $x_1 > 0$, $x_2 = 0$, $y_1 = 0$, $y_2 > 0$ e $\lambda \in (0,\ 1)$. Verifique!

O exercício a seguir mostra uma maneira de se obter uma função convexa em $n+1$ variáveis a partir de uma função convexa de n variáveis.

E. 5.11 *Exercício.* Seja C um conjunto convexo em $\mathbb{R}^n$ e seja $f : C \to \mathbb{R}$ uma função convexa de n variáveis em C. Seja $g(x,\ t) := tf(t^{-1}x)$ uma função de $n+1$ variáves definida em $D := \left\{(x,\ t) \in \mathbb{R}^{n+1} \middle|\ t > 0 \text{ e } t^{-1}x \in C\right\} \subset \mathbb{R}^{n+1}$. Mostre que D é um conjunto convexo em $\mathbb{R}^{n+1}$ (em verdade, um cone convexo em $\mathbb{R}^{n+1}$) e mostre que g é uma função convexa de $n+1$ variáveis em D, ou seja, mostre que

$$g\Big(\lambda(x,\ t) + (1-\lambda)(x',\ t')\Big) \ \equiv\ g\Big(\lambda x + (1-\lambda)x',\ \lambda t + (1-\lambda)t'\Big) \ \leq\ \lambda g\big(x,\ t\big) + (1-\lambda)g\big(x',\ t'\big)$$

para todos $(x,\ t)$, $(x',\ t') \in D$ e $\lambda \in [0,\ 1]$. *Sugestão:* Use o fato que

$$\frac{1}{\lambda t + (1-\lambda)t'}\Big(\lambda x + (1-\lambda)x'\Big) \ = \ \left(\frac{\lambda t}{\lambda t + (1-\lambda)t'}\right)\frac{1}{t}\,x + \left(\frac{(1-\lambda)t'}{\lambda t + (1-\lambda)t'}\right)\frac{1}{t'}\,x' \ ,$$

note que a soma dos termos entre parênteses do lado direito vale 1 e use a convexidade de C e de f. ✠

E. 5.12 *Exercício.* Mostre que toda seminorma em um espaço vetorial real é uma função convexa sobre o mesmo. ✠

• Hiperplanos suportes. Generalização para várias variáveis

Seja $F : C \to \mathbb{R}$ uma função convexa definida em um domínio convexo C de um espaço vetorial $\mathcal{V}$. Para $q \in \mathcal{V}^*$, o espaço dual de $\mathcal{V}$, e $\beta \in \mathbb{R}$. A função afim $R : \mathcal{V} \to \mathbb{R}$ dada por $R(x) = \langle q,\ x\rangle + \beta$ é dita *suportar* F no ponto $x_0 \in C$ se valer $F(x_0) = R(x_0)$ e $F(x) \geq R(x)$ para todo $x \in C$.

Se R suporta F em x_0 o hiperplano $\big\{\big(x,\ R(x)\big),\ x \in \mathcal{V}\big\} \subset \mathcal{V} \times \mathbb{R}$ é dito ser um *hiperplano suporte* de F nesse ponto. Geometricamente, podemos afirmar que um hiperplano suporte de F em x_0 tangencia o gráfico de F em x_0 e o gráfico de F está sempre acima de seus hiperplanos suportes.

Se a função afim $R(x) = \langle q,\ x\rangle + \beta$ suporta F em x_0, então $R(x_0) = F(x_0)$, então $\langle q,\ x_0\rangle + \beta = F(x_0)$ e podemos escrever

$$R(x) \ = \ \langle q,\ (x - x_0)\rangle + F(x_0) \ .$$

De particular importância é o caso de funções convexas $F : C \to \mathbb{R}$ definidas em um domínio convexo C de $\mathbb{R}^n$. Como já observamos, uma tal função $F(x_1,\ x_2,\ \ldots,\ x_n)$, definida em um ponto de coordenadas $(x_1,\ x_2,\ \ldots,\ x_n)$ é uma função convexa de cada uma das variáveis x_k, $k = 1,\ \ldots,\ n$. Portanto, pelas nossas considerações anteriores, F é contínua no interior de C e existem as derivadas partiais laterais

$$\left.\frac{\partial F}{\partial x_k}\right|_{\pm}(x_1,\ x_2,\ \ldots,\ x_n) \ = \ \lim_{\substack{\epsilon\to 0\\ \pm\epsilon>0}} \frac{F(x_1,\ \ldots,\ x_k+\epsilon,\ x_{k+1}\ \ldots,\ x_n) - F(x_1,\ \ldots,\ x_k,\ x_{k+1}\ \ldots,\ x_n)}{\epsilon}$$

para cada $k = 1,\ \ldots,\ n$, sendo que também vale

$$\left.\frac{\partial F}{\partial x_k}\right|_{-}(x_1,\ x_2,\ \ldots,\ x_n) \ \leq \ \left.\frac{\partial F}{\partial x_k}\right|_{+}(x_1,\ x_2,\ \ldots,\ x_n) \ .$$

Vide Proposição 5.10, página 356.

Em cada ponto $x \equiv (x_1,\ x_2,\ \ldots,\ x_n)$ de C temos um máximo de 2^n gradientes laterais

$$\left(\left.\frac{\partial F}{\partial x_1}\right|_{\pm}(x),\ \ldots,\ \left.\frac{\partial F}{\partial x_n}\right|_{\pm}(x)\right) \ ,$$

sendo que, acima, os diversos sinais $\pm$ são independentes. Cada vetor V não nulo do cone positivo gerado[20] por esses vetores define um hiperplano tangente ao gráfico de F em x composto pelos vetores ortogonais ao vetor $\big(V,\ 1\big) \in \mathbb{R}^{n+1}$. Esse hiperplano é um hiperplano suporte a F em x.

A ideia que uma função convexa de n variáveis é também convexa em suas variáveis individuais pode ser estendida, como já mencionanos no Lema 5.6, página 352. Seja $C \subset \mathbb{R}^n$, um conjunto convexo, e $F : C \to \mathbb{R}$, uma função convexa. Tomando $x \in C_0$ e um vetor $u \in \mathbb{R}^n$, defina-se

$$f(t) \ := \ F\big(x + tu\big) \ ,$$

[20] O cone positivo gerado por vetores $v_1, \ldots,\ v_q$ é conjunto de todos os vetores da forma $\lambda_1 v_1 + \cdots + \lambda_q v_q$, para todos $\lambda_j \geq 0$, $j = 1,\ \ldots,\ q$.

com $t \in \mathbb{R}$ definido em algum intervalo I, pequeno o suficiente para garantir que $x + tu \in C_0$. Afirmamos que $f : I \to \mathbb{R}$ é uma função convexa de uma variável. De fato, para $\lambda \in [0,\ 1]$, e $t_1,\ t_2 \in I$,

$$\begin{aligned} f\big(\lambda t_1 + (1-\lambda)t_2\big) \ &= \ F\Big(x + \big(\lambda t_1 + (1-\lambda)t_2\big)u\Big) \\ &= \ F\Big(\lambda\big(x + t_1 u\big) + (1-\lambda)\big(x + t_2 u\big)\Big) \ \leq \ \lambda F\big(x + t_1 u\big) + (1-\lambda)F\big(x + t_2 u\big) \ = \ \lambda f(t_1) + (1-\lambda)f(t_2) \ . \end{aligned}$$

Com isso, aplicam-se à função f os resultados anteriores: f é contínua em I_0 e possui as derivadas laterais $f'_\pm(t)$ para todo $t \in I_0$, sendo que valem também os resultados expressos na 5.10, página 356, e também no Corolário 5.3, página 357:

$$f(t') \ \geq \ (t' - t)f'_\pm(t) + f(t) \ .$$

Agora,

$$f'_\pm(t) \ = \ \sum_{k=1}^{n} u_k \ \frac{\partial F}{\partial x_k}\bigg|_\pm (x + tu) \ ,$$

sendo os sinais $\pm$ de cada termo do lado direito dependentes do sinal de cada u_k.

Assim,

$$F\big(x + t'u\big) \ \geq \ F\big(x + tu\big) + (t' - t)\sum_{k=1}^{n} u_k \ \frac{\partial F}{\partial x_k}\bigg|_\pm (x + tu) \ .$$

Denotando $y = x + tu$, $y' = x + t'u$, isso se lê como

$$F\big(y'\big) \ \geq \ F\big(y\big) + \sum_{k=1}^{n} (y' - y)_k \ \frac{\partial F}{\partial x_k}\bigg|_\pm (y) \ .$$

Tomando $q = \left(\frac{\partial F}{\partial x_1}\Big|_\pm (y),\ \ldots,\ \frac{\partial F}{\partial x_n}\Big|_\pm (y)\right)$ (para quaisquer escolhas dos sinais) a relação acima pode ser lida como $F(y') \geq R(y')$, com

$$R(y') \ := \ \big\langle q,\ (y - y)\big\rangle + F(y) \ ,$$

sendo que $R(y) = F(y)$, o que mostra que as funções R suportam F em y.

Temos, com essas considerações:

Proposição 5.17 *Seja $F : C \to \mathbb{R}$ uma função convexa definida em um domínio convexo C de um espaço vetorial $\mathbb{R}^n$. Então, F possui ao menos um hiperplano suporte para cada $x_0 \in C$. Isso significa que para cada $x_0 \in C$ existem $q \in \mathcal{V}^*$ e $\beta \in \mathbb{R}$ tais que a função afim $R : \mathcal{V} \to \mathbb{R}$ dada por $R(x) = \langle q,\ x\rangle + \beta$ satisfaz $F(x_0) = R(x_0)$ e $F(x) \geq R(x)$ para todo $x \in C$.* □

• **Convexidade e funções duas vezes diferenciáveis**

No caso em que $f : C \to \mathbb{R}$ é duas vezes diferenciável, a proposição que segue fornece condições necessárias e suficientes à convexidade.

Proposição 5.18 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo e seja $f : C \to \mathbb{R}$ uma função contínua e duas vezes diferenciável no interior C^0 de C. Então, f é convexa se e somente se*

$$\sum_{i,\ j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\big(x\big) \ a_i a_j \ \geq \ 0 \tag{5.48}$$

para todo $x \in C^0$ e para todo $(a_1,\ \ldots,\ a_j) \in \mathbb{R}^n$. □

Nota. A condição (5.48) é satisfeita se e somente se todos os autovalores da matriz Hessiana[21] $H_{ij}(x) := \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$, $i, j \in \{1, \ldots, n\}$, forem não negativos. ♣

Prova da Proposição 5.18. Para $x, y \in C^0$ defina-se a função de uma variável

$$g_{x,\,y}(t) \;:=\; f\big(x + t(y-x)\big), \qquad t \in [0,\, 1]\ .$$

Para $t_1, t_2 \in [0, 1]$ temos

$$g_{x,\,y}\big(\lambda t_1 + (1-\lambda)t_2\big) \;=\; f\Big(\lambda\big(x + t_1(y-x)\big) + (1-\lambda)\big(x + t_2(y-x)\big)\Big)\ ,$$

$$g_{x,\,y}(t_1) \;=\; f\big(x + t_1(y-x)\big)\ ,$$

$$g_{x,\,y}(t_2) \;=\; f\big(x + t_2(y-x)\big)\ ,$$

para todo $\lambda \in [0, 1]$. Dessas três relações prova-se facilmente que f é convexa em C se e somente todas as funções $g_{x,\,y}$, $x, y \in C^0$, forem funções convexas de uma variável.

A função g é duas vezes diferenciável, pois f o é. Assim, pelo Corolário 5.4, página 358, f é convexa se e somente se para todos $x, y \in C^0$ valer $g''_{x,\,y} \geq 0$. Agora, pela regra da cadeia,

$$g''_{x,\,y}(t) \;=\; \sum_{i,\,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\big(x + t(y-x)\big)\ (y_i - x_i)(y_j - x_j)\ .$$

Logo, teremos $g''_{x,\,y} \geq 0$ para todos $x, y \in C^0$ se e somente se

$$\sum_{i,\,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x)\ a_i a_j \;\geq\; 0$$

para todo $x \in C^0$ e para todo $(a_1, \ldots, a_j) \in \mathbb{R}^n$. ■

5.3 Algumas Consequências da Convexidade e da Concavidade

Da propriedade de convexidade ou concavidade de funções é possível obter desigualdades muito úteis das quais faremos uso em outros momentos nestas Notas. Nesta Seção apresentaremos e demonstramos algumas delas, como a importate desigualdade de Jensen, a (primeira) desigualdade de Young e algumas outras desigualdades decorrentes da concavidade da função logaritmo ou da convexidade da função exponencial. Todas essas desigualdades são relevantes e possuem aplicações diversas da Mecânica Estatística e Termodinâmica à Análise Funcional.

5.3.1 A Desigualdade de Jensen

Vamos agora apresentar uma importante generalização da desigualdade (5.21). Por simplicidade, consideraremos aqui funções convexas $f : \mathbb{R} \to \mathbb{R}$ definidas em todo $\mathbb{R}$, mas é fácil perceber que é possível também considerar funções convexas definidas em intervalos menores, desde que as devidas restrições sejam feitas às demais funções envolvidas.

Proposição 5.19 (Desigualdade de Jensen) *Seja $f : \mathbb{R} \to \mathbb{R}$ convexa. Seja $p : \mathbb{R} \to [0, \infty)$ uma função positiva e integrável, ou seja, tal que $P := \int_{-\infty}^{\infty} p(t)\,dt < \infty$. Seja $g : \mathbb{R} \to \mathbb{R}$ uma função integrável na medida $p(t)dt$, ou seja, tal que $\int_{-\infty}^{\infty} |g(t)|\,p(t)\,dt < \infty$. Então, vale*

$$P\,f\left(\int_{-\infty}^{\infty} g(t)\,p(t)\,dt\right) \;\leq\; \int_{-\infty}^{\infty} f\big(P\,g(t)\big)\,p(t)\ dt\ . \tag{5.49}$$

[21] Ludwig Otto Hesse (1811–1874).

Essa desigualdade é denominada desigualdade de Jensen. *Um caso particular importante é aquele no qual a função p representa uma distribuição de probabilidades e temos $P = 1$. Nele, a desigualdade de Jensen assume a forma*

$$f\left(\int_{-\infty}^{\infty} g(t)\,p(t)\,dt\right) \;\leq\; \int_{-\infty}^{\infty} f\big(g(t)\big)\,p(t)\;dt\;. \tag{5.50}$$

□

O estudante deve perceber que (5.21) é um caso particular de (5.50). O estudante mais avançado deve também perceber ao longo da demonstração que (5.49) permacece válida se $p(t)dt$ for substituida por qualquer medida positiva e finita $d\mu$ em $\mathbb{R}$ (com $P := \int_{\mathbb{R}} d\mu$) e que (5.50) permanece válida se $p(t)dt$ for substituida por qualquer medida de probabilidade $d\mu$ em $\mathbb{R}$.

Prova da Proposição 5.19. A desigualdade (5.31), página 357, afirma que, para todo $y \in \mathbb{R}$ podemos escrever

$$f(y) \;\geq\; \alpha_x y + \beta_x\;,$$

onde $\alpha_x := f'_\pm(x)$ e $\beta_x := f(x) - xf'_\pm(x)$, sendo $x \in \mathbb{R}$, arbitrário. Tomando-se $y \equiv Pg(t)$, temos

$$f\big(Pg(t)\big) \;\geq\; \alpha_x Pg(t) + \beta_x\;.$$

Logo, como p é não negativa, temos também

$$f\big(Pg(t)\big)p(t) \;\geq\; \alpha_x Pg(t)p(t) + \beta_x p(t)\;.$$

Integrando-se, obtemos

$$\int_{-\infty}^{\infty} f\big(Pg(t)\big)\,p(t)\;dt \;\geq\; \alpha_x P\int_{-\infty}^{\infty} g(t)\,p(t)\;dt + \beta_x P \;=\; f'_\pm(x)P\left(\int_{-\infty}^{\infty} g(t)\,p(t)\;dt\right) + Pf(x) - Pxf'_\pm(x)\;,$$

com $x \in \mathbb{R}$, arbitrário. Tomando-se e $x \equiv \int_{-\infty}^{\infty} g(t)p(t)\,dt$, o primeiro e o terceiro termo do lado direito cancelam-se e obtemos, finalmente,

$$\int_{-\infty}^{\infty} f\big(Pg(t)\big)\,p(t)\;dt \;\geq\; Pf\left(\int_{-\infty}^{\infty} g(t)\,p(t)\,dt\right)\;,$$

que é a desigualdade de Jensen (5.49). ■

5.3.2 A Primeira Desigualdade de Young

Nosso resultado na corrente seção será uma desigualdade útil que denominamos *primeira desigualdade de Young*[22].

Proposição 5.20 (Primeira Desigualdade de Young) *Sejam dois intervalos $[0,\ \alpha)$ e $[0,\ \beta)$, onde $0 < \alpha \leq \infty$ e $0 < \beta \leq \infty$ e seja $F : [0,\ \alpha) \to [0,\ \beta)$ uma função contínua, não negativa, crescente, bijetora e satisfazendo $F(0) = 0$. Então, para todos $x \in [0,\ \alpha)$ e $x \in [0,\ \beta)$ vale*

$$xy \;\leq\; \int_0^x F(s)\;ds + \int_0^y F^{-1}(t)\;dt\;. \tag{5.51}$$

Essa desigualdade é denominada aqui "primeira desigualdade de Young". □

A desigualdade (5.51) tem uma interpretação geométrica muito simples, que descrevemos na Figura 5.6, página 369. Para generalizações da desigualdade (5.51), vide e.g. [225].

Prova da Proposição 5.20. Como $F : [0,\ \alpha) \to [0,\ \beta)$ uma função contínua, não negativa, crescente, bijetora e satisfaz $F^{-1}(0) = 0$, sua inversa $F^{-1} : [0,\ \beta) \to [0,\ \alpha)$ é igualmenete contínua, não negativa, crescente, bijetora e satisfaz $F^{-1}(0) = 0$.

[22] William Henry Young (1863–1942).

Defina-se, para $x \in [0,\ \alpha)$ e $y \in [0,\ \beta)$ as funções

$$A(x) \ := \ \int_0^x F(s)\, ds \qquad \text{e} \qquad B(y) \ := \ \int_0^y F^{-1}(t)\, dt \ .$$

A e B são contínuas e diferenciáveis (com $A'(x) = F(x)$ e $B'(y) = F^{-1}(y)$) e, pelo Corolário 5.5, página 358, ambas são convexas, já que F e F^{-1} são crescentes.

Afirmamos que vale a identidade $B(y) = C(y)$ para todo $y \in [0,\ \beta)$, onde

$$C(y) \ := \ \int_0^{F^{-1}(y)} \Big(y - F(s)\Big)\, ds \ .$$

De fato, C é contínua e diferenciável (pois F^{-1} é contínua) e tem-se[23]

$$C'(y) \ = \ y - F\big(F^{-1}(y)\big) + \int_0^{F^{-1}(y)} \frac{\partial}{\partial y}\Big(y - F(s)\Big)\, ds \ = \ \int_0^{F^{-1}(y)} ds \ = \ F^{-1}(y) \ = \ B'(y) \ ,$$

e como $B(0) = 0$ e $C(0) = 0$ (pois $F^{-1}(0) = 0$), segue que $B(y) = C(y)$ para todo $y \in [0,\ \beta)$.

Agora, a identidade $B(y) = C(y)$ permite escrever

$$B\big(F(x)\big) \ = \ C\big(F(x)\big) \ = \ \int_0^{F^{-1}(F(x))} \Big(F(x) - F(s)\Big)\, ds \ = \ \int_0^x \Big(F(x) - F(s)\Big)\, ds \ = \ xF(x) - A(x) \ .$$

Estabelecemos, portanto, que

$$xF(x) \ = \ A(x) + B\big(F(x)\big) \tag{5.52}$$

para todo $x \in [0,\ \alpha)$, relação que logo usaremos.

Da convexidade estrita de B e do Corolário 5.3, página 357, segue que

$$B(y) \ \geq \ (y - y')F^{-1}(y') + B(y')$$

para todos $y,\ y' \in [0,\ \beta)$. Tomando-se $y' = F(x)$, teremos

$$B(y) \ > \ \big(y - F(x)\big)x + B\big(F(x)\big) \ \stackrel{(5.52)}{=} \ xy - B\big(F(x)\big) - A(x) + B\big(F(x)\big)$$

e, portanto, provamos que

$$xy \ \leq \ A(x) + B(y)$$

para todos $x \in [0,\ \alpha)$ e $y \in [0,\ \beta)$. ■

E. 5.13 *Exercício*. Tomando-se $F(x) = e^x - 1$ em (5.51), obtenha a desigualdade

$$(1 + x)(1 + y) \ \leq \ e^x + (1 + y)\ln(1 + y) \ , \qquad \forall x,\ y \geq 0 \ .$$

Demonstre alternativamente sua validade estudando os mínimos da função $e^x + (1+y)\ln(1+y) - (1+x)(1+y)$ pelo método convencional. ✠

Na desigualdade (5.51), o lado esquerdo é uma função simétrica de x e y, mas o lado direito não é. Isso pode ser remediado simetrizando-se o lado direito no intervalo comum a F e F^{-1}, a saber, $[0,\ \gamma)$, onde $\gamma = \min\{\alpha,\ \beta\}$. Obtem-se, então, nesse intervalo

$$xy \ \leq \ \frac{1}{2}\left(\int_0^x F(s)\, ds + \int_0^y F(s)\, ds\right) + \frac{1}{2}\left(\int_0^x F^{-1}(t)\, dt + \int_0^y F^{-1}(t)\, dt\right) \ .$$

[23] Aqui, usa-se a identidade (prove-a!)

$$\frac{d}{dy}\int_0^{h(y)} J(y,\ s)\, ds \ = \ J\big(y,\ h(y)\big) + \int_0^{h(y)} \frac{\partial J}{\partial y}(y,\ s)\, ds \ ,$$

válida para $h(y)$ contínua e $J(y,\ s)$ contínua e diferenciável, com $\frac{\partial}{\partial y}J(y,\ s)$ contínua.

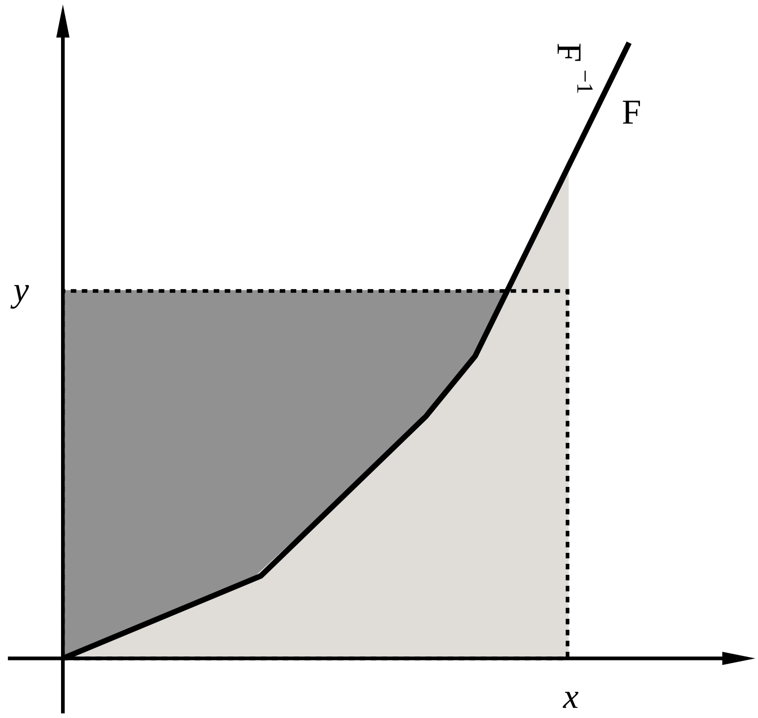

Figura 5.6: O gráfico de uma função contínua, bijetora, positiva e crescente F, com $F(0) = 0$. A área em cinza claro vale $\int_0^x F(s)\, ds$. A área em cinza escuro vale $\int_0^y F^{-1}(t)\, dt$. O retângulo de lados x e y é representado em linhas tracejadas e sua área xy é claramente menor ou igual à soma das duas áreas acinzentadas, que vale $\int_0^x F(s)\, ds + \int_0^y F^{-1}(t)\, dt$. Essa é a interpretação geométrica da Primeira Desigualdade de Young (5.51).

5.3.3 Médias Geométricas, Aritméticas e Desigualdades Correlatas

• **A desigualdade entre média geométrica e média aritmética**

A concavidade da função $\ln x$ tem algumas consequências relevantes. Para $n \in \mathbb{N}$, sejam $x_1, \ldots, x_n$ números positivos e sejam $\lambda_1, \ldots, \lambda_n \in [0, 1]$ tais que $\lambda_1 + \cdots + \lambda_n = 1$. Como $\ln x$ é côncava, (5.22) garante-nos que

$$\ln(\lambda_1 x_1 + \cdots + \lambda_n x_n) \;\geq\; \lambda_1 \ln(x_1) + \cdots + \lambda_n \ln(x_n) \;=\; \ln\left(x_1^{\lambda_1} \cdots x_n^{\lambda_n}\right) .$$

Tomando-se a exponencial dessa desigualdade, obtemos

$$x_1^{\lambda_1} \cdots x_n^{\lambda_n} \;\leq\; \lambda_1 x_1 + \cdots + \lambda_n x_n . \tag{5.53}$$

Note-se que se $\lambda_1, \ldots, \lambda_n \in (0, 1)$ essa desigualdade é válida mesmo que alguns x_k's sejam nulos.

A expressão $\lambda_1 x_1 + \cdots + \lambda_n x_n$ é denominada *média aritmética ponderada* do conjunto de números não negativos $\{x_1, \ldots, x_n\}$ (ponderada pelos fatores $\lambda_1, \ldots, \lambda_n$) e a expressão $x_1^{\lambda_1} \cdots x_n^{\lambda_n}$ é denominada *média geométrica ponderada* do conjunto $\{x_1, \ldots, x_n\}$. A desigualdade (5.53) afirma, portanto, que a média geométrica ponderada de um conjunto finito de números não negativos é sempre menor que a sua média aritmética ponderada. Em [225] o leitor poderá encontrar diversas outras demonstrações da desigualdade (5.53).

Sejam $x_1, \ldots, x_n$ números não negativos. Definimos suas médias geométrica e aritmética (simples) por $\left(x_1 \cdots x_n\right)^{\frac{1}{n}}$ e $\frac{x_1+\cdots+x_n}{n}$, respectivamente. Elas correspondem ao caso em que $\lambda_1 = \cdots = \lambda_n = \frac{1}{n}$. Temos de (5.53), portanto,

$$\left(x_1 \cdots x_n\right)^{\frac{1}{n}} \;\leq\; \frac{x_1 + \cdots + x_n}{n} , \tag{5.54}$$

ou seja, a média geométrica de uma coleção finita de números não negativos é sempre menor ou igual à média aritmética da mesma coleção. Esse resultado é originalmente atribuido a Cauchy[24] e pode ser provado de diversas formas. Cauchy

[24] Augustin Louis Cauchy (1789–1857).

obteve-o, não usando a concavidade do logaritmo, como fizemos, mas por indução (para uma tal prova, vide e.g. [225]).

• A desigualdade de Young

Na demonstração da chamada desigualdade de Hölder em espaços L^p (assunto discutido nas Seções 25.5.1 e 32.4.1, páginas 1443 e 1673, respectivamente) faz-se uso de uma desigualdade elementar conhecida como *desigualdade de Young*[25]. Como a desigualdade de Young tem interesse por si só e algumas outras aplicações, vamos apresentar sua demonstração.

Como já discutimos, a função logaritmo é côncava como função do intervalo $(0, \infty)$ sobre $\mathbb{R}$. Assim, para todos $a, b \in (0, \infty)$ e todo $\lambda \in [0, 1]$ tem-se $\ln\big(\lambda a + (1-\lambda)b\big) \geq \lambda \ln a + (1-\lambda)\ln b = \ln\big(a^\lambda b^{1-\lambda}\big)$. Tomando-se a exponencial de ambos os lados, obtemos

$$a^\lambda b^{1-\lambda} \leq \lambda a + (1-\lambda)b . \tag{5.55}$$

Note-se que se $\lambda \in (0, 1)$, então (5.55) é também válida caso $a = 0$ e/ou $b = 0$. A desigualdade (5.55) é por vezes denominada desigualdade de Young. Como se vê, trata-se meramente de um caso particular de (5.53) (para $n = 2$).

Por vezes a desigualdade de Young é apresentada de forma ligeiramente diferente. Sejam p e q ambos tais que $1 < p < \infty$ e $1 < q < \infty$, mas tais que $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Então, por (5.55), temos para $a, b \in [0, \infty)$.

$$a^{1/p}\, b^{1/q} \quad \leq \quad \frac{a}{p} + \frac{b}{q} , \tag{5.56}$$

sendo que a igualdade só é válida caso $a = b$ (isso segue do fato de o logaritmo ser estritamente côncavo. Vide, porém, nota logo abaixo).

A desigualdade dita de Young (5.56) é bastante elementar, mas ela foi originalmente provada a partir do resultado mais geral e não trivial expresso na *primeira desigualdade de Young*, a desigualdade (5.51) da Proposição 5.20, página 367.

E. 5.14 *Exercício*. Tomando $F(s) = s^{p-1}$ com $1 < p < \infty$, para $s \geq 0$, obtenha da primeira desigualdade de Young (5.51) a desigualdade (5.56). ✠

Há ainda uma terceira desigualdade para produtos de convolução também denominada "desigualdade de Young" e que é vagamente relacionada às desigualdades (5.56) e (5.51). Vide, e.g., [336].

Nota. Por completeza, apresentemos uma segunda demonstração de (5.56) sem uso da concavidade. Notemos, em primeiro lugar, que se $a = 0$ ou $b = 0$ a (5.56) acima é trivialmente satisfeita, pois o lado esquerdo é sempre zero, enquanto que o lado direito é sempre maior ou igual a zero. Vamos, então, supor que a e b sejam ambos não nulos. Tudo o que queremos é provar que $-a^{1/p}b^{1/q} + \dfrac{a}{p} + \dfrac{b}{q}$ é sempre maior ou igual a zero. Podemos escrever a última expressão como $b\left(-t^\alpha + \alpha t + \frac{1}{q}\right)$, onde $\alpha = 1/p$ e $t = a/b$. Como $1 < p < \infty$, temos que $0 < \alpha < 1$ enquanto que $t \geq 0$. Note-se que a função

$$f(x) = -x^\alpha + \alpha x + \frac{1}{q} ,$$

é contínua para $x \in [0, \infty)$ e que, para $x > 0$, tem-se $f'(x) = \alpha\left(1 - x^{\alpha-1}\right)$ e $f''(x) = \alpha(1-\alpha)x^{\alpha-2} > 0$. Assim, $f(x)$ tem um único mínimo local em $x = 1$, onde $f(1) = 0$ (verifique). Fora isso, $f(0) = \frac{1}{q} > 0$ e $\lim_{x\to\infty} f(x) = +\infty$. Desses fatos concluímos facilmente que $f(x) \geq 0$ para todo $x \geq 0$, a igualdade só se dando caso $x = 1$. Isso fecha o que queríamos provar. ♣

E. 5.15 *Exercício*. Mostre que no caso $0 < p < 1$ a desigualdade (5.56) se reverte ($\leq$ deve ser substituído por $\geq$). Nesse caso $q < 0$. ✠

• Desigualdades envolvendo somas de potências

As desigualdades apresentadas na seguinte proposição são muito úteis, especialmente para o propósito de demonstrar que os conjuntos de sequências ℓ_p são espaços vetoriais (vide Seção 25.5.1, página 1443), o mesmo se dando com os conjuntos de funções $L_p(M, d\mu)$ dos quais trataremos no Capítulo 32, página 1634.

Proposição 5.21 *Sejam $a \geq 0$ e $b \geq 0$ dois números reais não negativos.*

I. *Para todo p tal que $0 < p < 1$ tem-se*

$$2^{p-1}\big(a^p + b^p\big) \leq (a+b)^p \leq a^p + b^p . \tag{5.57}$$

[25] William Henry Young (1863–1942).

II. *Para todo p tal que $p \geq 1$ tem-se*

$$a^p + b^p \;\leq\; (a+b)^p \;\leq\; 2^{p-1}\left(a^p + b^p\right) . \tag{5.58}$$

□

Prova. Apresentamos separadamente as demonstrações para os casos I e II.

Caso I. Tomemos $0 < p < 1$ fixo. Vamos primeiramente provar a seguinte desigualdade: para quaisquer $a, \; b \geq 0$ vale

$$(a+b)^p \;\leq\; a^p + b^p . \tag{5.59}$$

Para $a = 0$ isso é óbvio. Seja, então, $a > 0$. Nesse caso, podemos fatorar a^p e a desigualdade acima ficaria,

$$\left(1 + \frac{b}{a}\right)^p \;\leq\; 1 + \left(\frac{b}{a}\right)^p .$$

Para provar isso, tudo o que desejamos é provar que $f(x) := (1+x)^p - 1 - x^p$ satisfaz $f(x) \leq 0$ para todo $x \geq 0$. De fato, tem-se,

$$f'(x) \;=\; -px^{p-1}\left[1 - \frac{1}{\left(1 + \frac{1}{x}\right)^{1-p}}\right] . \tag{5.60}$$

Como $1 + \frac{1}{x} \geq 1$ e $1 - p > 0$, segue que $f'(x) \leq 0$ para todo $x \geq 0$. Com isso, provamos que f é não crescente. Como $f(0) = 0$, segue que $f(x) \leq 0$ para todo $x \geq 0$. Isso provou (5.59).

Vamos agora provar que

$$2^{p-1}\big(a^p + b^p\big) \;\leq\; (a+b)^p . \tag{5.61}$$

Para $x \geq 0$ e $0 < p < 1$ a função $\varphi(x) = x^p$ é côncava. Portanto, para qualquer λ com $0 \leq \lambda \leq 1$, tem-se

$$\lambda\varphi(a) + (1-\lambda)\varphi(b) \;\leq\; \varphi\left(\lambda a + (1-\lambda)b\right) .$$

Para $\lambda = 1/2$, isso fica $\frac{a^p+b^p}{2} \leq \left(\frac{a+b}{2}\right)^p$, que é (5.61), e a prova de (5.57) está completa.

Caso II. Para o caso $p = 1$ a desigualdade (5.58) é evidente. Tomemos, então, $p > 1$ fixo. Vamos primeiramente provar a seguinte desigualdade: para quaisquer $a, \; b \geq 0$ vale

$$a^p + b^p \;\leq\; (a+b)^p . \tag{5.62}$$

Para $a = 0$ isso é óbvio. Seja, então, $a > 0$. Nesse caso, podemos fatorar a^p e a desigualdade acima ficaria,

$$\left(1 + \frac{b}{a}\right)^p \;\geq\; 1 + \left(\frac{b}{a}\right)^p .$$

Para provar isso, tudo o que desejamos é provar que $f(x) := (1+x)^p - 1 - x^p$ satisfaz $f(x) \geq 0$ para todo $x \geq 0$. Agora, por (5.60),

$$f'(x) \;=\; -px^{p-1}\left[1 - \left(1 + \frac{1}{x}\right)^{p-1}\right] .$$

Como $1 + \frac{1}{x} \geq 1$ e $p - 1 > 0$, segue que $f'(x) \geq 0$ para todo $x \geq 0$. Com isso provamos que f é crescente. Como $f(0) = 0$, segue que $f(x) \geq 0$ para todo $x \geq 0$, provando o que queríamos.

Vamos agora provar que

$$(a+b)^p \;\leq\; 2^{p-1}\left(a^p + b^p\right) . \tag{5.63}$$

Para $x \geq 0$ e $p > 1$ a função $\varphi(x) = x^p$ é convexa. Portanto, para qualquer λ com $0 \leq \lambda \leq 1$, tem-se

$$\varphi\left(\lambda a + (1-\lambda)b\right) \;\leq\; \lambda\varphi(a) + (1-\lambda)\varphi(b) .$$

Para $\lambda = 1/2$, isso fica $\left(\frac{a+b}{2}\right)^p \leq \frac{a^p+b^p}{2}$, que é (5.63), e a prova de (5.58) está completa. ■

Um corolário útil é:

Corolário 5.6 *Para $0 < q < 2$ e para todos $z,\ w \in \mathbb{C}$ vale*

$$|z+w|^q + |z-w|^q \ \leq\ 2\big(|z|^q + |w|^q\big) \ . \tag{5.64}$$

Para $q \geq 2$ e para todos $z,\ w \in \mathbb{C}$ vale

$$|z+w|^q + |z-w|^q \ \leq\ 2^{q-1}\big(|z|^q + |w|^q\big) \ . \tag{5.65}$$

□

O Corolário 5.6 pode ser usado para obter-se certas generalizações da identidade do paralelogramo (na forma de desigualdades) em espaços ℓ_p, com $p \geq 1$. Vide relações (25.59) e (25.60), página 1451.

Prova do Corolário 5.6. Para $z,\ w \in \mathbb{C}$ e com $0 < p < 1$, tomemos $a = |z+w|^2$ e $b = |z-w|^2$ na primeira desigualdade em (5.57). Obtemos,

$$\frac{|z+w|^{2p} + |z-w|^{2p}}{2^{1-p}} \ \leq\ \big(|z+w|^2 + |z-w|^2\big)^p \ .$$

Agora, $|z+w|^2 + |z-w|^2 = 2|z|^2 + 2|w|^2$. Assim,

$$\frac{|z+w|^{2p} + |z-w|^{2p}}{2^{1-p}} \ \leq\ 2^p\big(|z|^2 + |w|^2\big)^p \ .$$

Agora, tomemos $a = |z|^2$ e $b = |w|^2$ na segunda desigualdade em (5.57). Teremos $\big(|z|^2 + |w|^2\big)^p \leq |z|^{2p} + |w|^{2p}$. Assim, estabelecemos que

$$\frac{|z+w|^{2p} + |z-w|^{2p}}{2^{1-p}} \ \leq\ 2^p\big(|z|^{2p} + |w|^{2p}\big)$$

que é (5.64) com $q = 2p$.

Provemos agora (5.65). Para $z,\ w \in \mathbb{C}$ e com $p \geq 1$, tomemos $a = |z+w|^2$ e $b = |z-w|^2$ na primeira desigualdade em (5.58). Obtemos,

$$|z+w|^{2p} + |z-w|^{2p} \ \leq\ \big(|z+w|^2 + |z-w|^2\big)^p \ =\ 2^p\big(|z|^2 + |w|^2\big)^p \ .$$

Agora, adotemos $a = |z|^2$ e $b = |w|^2$ na segunda desigualdade em (5.58). Obtemos, $\big(|z|^2 + |w|^2\big)^p \leq 2^{p-1}\left(|z|^{2p} + |w|^{2p}\right)$. Assim, temos

$$|z+w|^{2p} + |z-w|^{2p} \ \leq\ 2^{2p-1}\left(|z|^{2p} + |w|^{2p}\right) \ .$$

Tomando-se $q = 2p$ essa é a desigualdade (5.65). ■

5.3.3.1 A Desigualdade de Minkowski

Uma importante desigualdade empregada do estudo de espaços métricos e na Análise Funcional é a chamada *desigualdade de Minkowski*[26]. Apresentaremos adiante demonstrações dessa degualdade am algumas instâncias, fazendo uso central da convexidade da função $f(x) = x^p$ na região $x > 0$, sendo $p \geq 1$. Outras demonstrações dessa desigualdade podem ser encontradas na Seção 25.5.1, página 1443, e na Seção 32.4.1, página 1673. A estratégia que seguimos na presente seção provém de [453].

Proposição 5.22 (Desigualdade de Minkowski para sequências finitas) *Seja $p \geq 1$. Para $n \in \mathbb{N}$, sejam, $a_k \in \mathbb{C}$ e $b_k \in \mathbb{C}$, $k = 1, \ldots, n$. Então, vale a desigualdade*

$$\left[\sum_{k=1}^{n} \big|a_k + b_k\big|^p\right]^{1/p} \ \leq\ \left[\sum_{k=1}^{n} \big|a_k\big|^p\right]^{1/p} + \left[\sum_{k=1}^{n} \big|b_k\big|^p\right]^{1/p} \ , \tag{5.66}$$

conhecida como desigualdade de Minkowski *para sequências finitas.* □

[26] Hermann Minkowski (1864–1909).

Prova. Se todos os a_k's forem nulos ou se todos os b_k's forem nulos, então (5.66) é evidente. Vamos, então, supor que

$$A := \left[\sum_{k=1}^{n} |a_k|^p\right]^{1/p} \neq 0 \qquad \text{e} \qquad B := \left[\sum_{k=1}^{n} |a_k|^p\right]^{1/p} \neq 0 .$$

É claro que $|a_k + b_k| \leq |a_k| + |b_k|$ e, portanto, $|a_k + b_k|^p \leq \Big(|a_k| + |b_k|\Big)^p$. Assim, podemos escrever que

$$|a_k + b_k|^p \leq \Big(|a_k| + |b_k|\Big)^p = (A+B)^p \left(\frac{A}{A+B}\frac{|a_k|}{A} + \frac{B}{A+B}\frac{|b_k|}{B}\right)^p . \tag{5.67}$$

Agora, como $\frac{A}{A+B} + \frac{B}{A+B} = 1$, com ambos os termos positivos, vemos que $\frac{A}{A+B}\frac{|a_k|}{A} + \frac{B}{A+B}\frac{|b_k|}{B}$ é uma combinação linear convexa dos números $\frac{|a_k|}{A}$ e $\frac{|b_k|}{B}$. Como a função $f(x) = x^p$ é convexa na região $x > 0$ quando $p \geq 1$, segue que

$$\left(\frac{A}{A+B}\frac{|a_k|}{A} + \frac{B}{A+B}\frac{|b_k|}{B}\right)^p \leq \frac{A}{A+B}\left(\frac{|a_k|}{A}\right)^p + \frac{B}{A+B}\left(\frac{|b_k|}{B}\right)^p .$$

Retornando com isso a (5.67), ficamos com

$$|a_k + b_k|^p \leq (A+B)^{p-1}\Big[A^{1-p}|a_k|^p + B^{1-p}|b_k|^p\Big] .$$

Somando-se em k, teremos

$$\sum_{k=1}^{n} |a_k + b_k|^p \leq (A+B)^{p-1}\left[A^{1-p}\underbrace{\sum_{k=1}^{n}|a_k|^p}_{=A^p} + B^{1-p}\underbrace{\sum_{k=1}^{n}|b_k|^p}_{=B^p}\right] = (A+B)^{p-1}\Big[A+B\Big] = (A+B)^p .$$

Logo, tomando-se a p-ésima raíz, obtemos $\left[\sum_{k=1}^{n} |a_k + b_k|^p\right]^{1/p} \leq A + B$, que é a desigualdade (5.66). ■

Um corolário imediato é

Corolário 5.7 (Desigualdade de Minkowski para sequências p-somáveis) *Seja $p \geq 1$ e sejam $\{a_k,\ k \in \mathbb{N}\}$ e $\{b_k,\ k \in \mathbb{N}\}$, duas sequências de números complexos tais que*

$$\sum_{k=1}^{\infty} |a_k|^p < \infty \qquad e \qquad \sum_{k=1}^{\infty} |b_k|^p < \infty . \tag{5.68}$$

Então, vale a desigualdade

$$\left[\sum_{k=1}^{\infty} |a_k + b_k|^p\right]^{1/p} \leq \left[\sum_{k=1}^{\infty} |a_k|^p\right]^{1/p} + \left[\sum_{k=1}^{\infty} |b_k|^p\right]^{1/p} , \tag{5.69}$$

conhecida como desigualdade de Minkowski *para sequências p-somáveis.* □

Prova. Use que (5.66) vale para todo $n \in \mathbb{N}$ e tome o limite $n \to \infty$ levando em conta (5.68). ■

Uma segunda demonstração da desigualdade (5.69) será apresentada na Seção 25.5.1, página 1443.

E. 5.16 *Exercício*. Sejam f e g duas funções contínuas definidas em um intervalo fechado $[a,\ b] \subset \mathbb{R}$ assumindo valores em $\mathbb{C}$. Mostre, imitando a demonstração da Proposição 5.22, que vale a desigualdade

$$\left[\int_a^b |f(x) + g(x)|^p \, dx\right]^{1/p} \leq \left[\int_a^b |f(x)|^p \, dx\right]^{1/p} + \left[\int_a^b |g(x)|^p \, dx\right]^{1/p} , \tag{5.70}$$

também conhecida como *desigualdade de Minkowski*. ✠

Os resultados acima podem ser muitíssimo generalizados (para funções p-integráveis em espaços mensuráveis). Vide Seção 32.4.1, página 1673.

• **O caso** $0 < p < 1$

No caso $0 < p < 1$ a desigualdade de Minkowski não é mais válida na forma que oferecemos acima. O Exercício a seguir ilustra isso.

E. 5.17 *Exercício*. Usando o fato que $f(x) = x^p$ é uma função côncava na região $x > 0$ quando $0 < p < 1$, mostre que para tais valores de p vale

$$\left[\sum_{k=1}^{n} (a_k + b_k)^p\right]^{1/p} \geq \left[\sum_{k=1}^{n} (a_k)^p\right]^{1/p} + \left[\sum_{k=1}^{n} (b_k)^p\right]^{1/p}, \tag{5.71}$$

onde os a_k's e b_k's são números **reais não negativos**. Note a reversão da desigualdade em (5.71) comparada a (5.66). ✣

Para $0 < p < 1$ tem-se, porém, o seguinte resultado:

Corolário 5.8 *Seja* $0 < p < 1$*. Para* $n \in \mathbb{N}$*, sejam,* $a_k \in \mathbb{C}$ *e* $b_k \in \mathbb{C}$*,* $k = 1, \ldots, n$*. Então, vale a desigualdade*

$$\sum_{k=1}^{n} |a_k + b_k|^p \leq \left[\sum_{k=1}^{n} |a_k|^p\right] + \left[\sum_{k=1}^{n} |b_k|^p\right]. \tag{5.72}$$

□

Prova. Temos que $|a_k+b_k| \leq |a_k|+|b_k|$. Como a função $[0, \infty) \ni x \to x^p$ é crescente, segue que $|a_k+b_k|^p \leq \left(|a_k|+|b_k|\right)^p$. Pela desigualdade (5.59), página 371, segue que $\left(|a_k| + |b_k|\right)^p < |a_k|^p + |b_k|^p$. Portanto, $|a_k + b_k|^p \leq |a_k|^p + |b_k|^p$ e disso segue imediatamente (5.72). ■

Nota. O estudante deve observar a principal diferença entre (5.66) e (5.72): a ausência na segunda relação do expoente $1/p$. A relação (5.72) pode ser interpretada como uma versão da Desigualdade de Minkowski para $0 < p < 1$. ♣

5.4 Exercícios Adicionais

E. 5.18 *Exercício dirigido.* Vamos aqui apresentar uma relação útil: uma fórmula fechada para a integral indefinida de uma função inversa. Seja h diferenciável e inversível (em um intervalo adequado, digamos, em todo $\mathbb{R}$). Afirmamos que vale

$$\int h^{-1}(x)\,dx \;=\; xh^{-1}(x) - H\big(h^{-1}(x)\big) + C\;, \tag{5.73}$$

onde H é uma primitiva de h (ou seja, $H'(x) = h(x)$) e C é uma constante de integração. A relação (5.73) pode ser facilmente provada por verificação (faça-o!): derivando ambos os lados e constatando a identidade. Apresentemos uma prova direta. Como $x = h\big(h^{-1}(x)\big)$, temos da regra da cadeia

$$1 \;=\; h'\big(h^{-1}(x)\big)\big(h^{-1}\big)'(x)\;. \tag{5.74}$$

Por outro lado, integração por partes nos diz que

$$\int h^{-1}(x)\,dx \;=\; \int \frac{dx}{dx}h^{-1}(x)\,dx \;=\; xh^{-1}(x) - \int x\big(h^{-1}\big)'(x) \;\stackrel{(5.74)}{=}\; xh^{-1}(x) - \int \frac{x}{h'\big(h^{-1}(x)\big)}\,dx\;.$$

Com a mudança de variáveis $u = h^{-1}(x)$, $x = h(u)$ e $dx = h'(u)du$ na última integral, temos

$$\int h^{-1}(x)\,dx \;=\; xh^{-1}(x) - \int \frac{h(u)}{h'(u)}h'(u)\,du \;=\; xh^{-1}(x) - \int h(u)\,du, \;=\; xh^{-1}(x) - H(u) \;=\; xh^{-1}(x) - H\big(h^{-1}(x)\big)\;,$$

onde H é uma primitiva de h. Isso prova (5.73).

Exemplos elementares de uso de (5.73): Para $h(x) = e^x$, (5.73) fornece (aqui, $x > 0$)

$$\int \ln(x)\,dx \;=\; x\ln(x) - e^{\ln(x)} + C \;=\; x\big(\ln(x) - 1\big) + C\;.$$

Para $h(x) = \cos(x)$ (aqui, $x \in [0,\;\pi]$), temos por (5.73),

$$\int \arccos(x)\,dx \;=\; x\arccos(x) - \;\mathrm{sen}\big(\arccos(x)\big) + C \;=\; x\arccos(x) - \sqrt{1-x^2} + C\;.$$

O estudante pode interessar-se por encontrar outros usos de (5.73). ✠