Matemática para o exame da ANPEC [1]

Alexandre L. Madureira

Laboratório Nacional de Computação Científica—LNCC, Brasil
URL: http://www.lncc.br/∼alm

[1] 04 de setembro de 2020

RESUMO. Estas notas de aula são relativas ao curso de preparação promovido pela FGV para a parte de matemática do exame da ANPEC. Estas notas devem servir de apoio, e certamente não eliminam a necessidade de se usar os já clássicos, aprimorados e vários livros didáticos. Mencionamos alguns deles na biliografia.

Neste curso apresento alguns tópicos de álgebra linear, cálculo e análise que estão presentes no exame da ANPEC, e que são importantes para uma formação mais sólida de futuros pós-graduandos em economia. Espero apresentar algum rigor matemático aos alunos, e mostrar como este deve ser utilizado em conjunto com a intuição matemática, nunca esquecendo o objetivo que é aprimorar a arte de resolver questões.

Uma particularidade das notas é que, ao fim destas há soluções de questões das provas da ANPEC de matemática dos últimos anos. Isto não seria possível sem a ajuda de Gustavo Lopo Andrade, Gustavo Pereira, e Lucas Alves, que gentilmente concordaram em apresentar suas soluções em TEX. Acho que estas soluções serão úteis para a grande comunidade de alunos que se prepara para os exames da ANPEC. Meus agradecimentos mais sinceros a estes três jovens!

Eu tomei a liberdade de modificar minimamente a notação usada em algumas das questões, a fim de torná-la homogênea e coincidir com as notações usadas nestas notas. Editei também minimamente as questões submetidas pelos alunos, a fim de tornar suas (deles) soluções mais próximas do estilo, linguagem e notações usadas no restante das notas.

São usadas estas notas várias ideias e notações de outros livros, como [**4**, **15**, **22**] em álgebra linear. A bibliografia básica sugerida pela ANPEC é dada por [**4**, **6**, **25**], e a complementar é [**1**, **10**, **11**, **15**, **30**].

Sumário

Os tópicos destas notas seguem a orientação da própria ANPEC. São eles:

(1) Noção de Conjunto

Relação de pertinência. Relação de inclusão, operações de interseção, união, diferença. Produto cartesiano. Relações.

(2) Noções de Geometria Analítica

Coordenadas no plano e no espaço. Fórmulas de distância. Vetores livres no plano e no espaço. Produto escalar, produto vetorial, perpendicularidade. Equações da reta no plano e no espaço, equações de planos. Inequações lineares. Parábola e hipérbole.

(3) Funções

Funções injetoras, sobrejetoras e bijetoras. Representação gráfica. Soma, diferença, produto, quociente e composição de funções.

(4) Álgebra Linear

Operações com matrizes. Matriz inversa, transposta e adjunta. Resolução de sistemas lineares. Determinantes. Regra de Cramer. Espaços vetoriais. Subespaços. Base e dimensão. Produto interno, ortogonalidade. Projeções. Transformações lineares. Núcleo e imagem. Matriz de uma transformação linear. Autovalores e autovetores. Polinômios característicos operadores diagonalizáveis. Operadores auto-adjuntos, operadores ortogonais. Formas bilineares.

(5) Funções de uma variável real -

Limites. Funções contínuas. Funções deriváveis. Reta tangente e reta normal. Regras de derivação: derivada da soma, do produto, do quociente, regra da cadeia, derivada da inversa. Elasticidade. Derivadas sucessivas. Funções trigonométricas. Função exponencial e logarítmica. Regra de L'Hôpital. Intervalos de concavidade e convexidade. Ponto de inflexão. Polinômio de Taylor.

(6) Integrais

Teorema fundamental do cálculo, primitivação por partes e por substituição. Áreas planas. Integrais impróprias.

(7) Sequências e séries

Convergência e divergência de seqüências e séries. Série geométrica, teste da comparação, da razão, da raiz, teste da integral. Séries alternadas.

(8) Matemática financeira

Juros simples. Juros compostos. Desconto e taxa de desconto. Séries de pagamento. Fluxo de caixa. Sistema de amortização.

(9) Funções de várias variáveis reais

Derivadas parciais. Diferencial total. Gradiente. Regra da cadeia. Funções implícitas. Teorema do envelope. Funções homogêneas. Teorema de Euler. Condições de 1ª e 2ª ordens para máximos e mínimos de funções de várias variáveis reais. Condições de 1ª e 2ª ordens para otimização condicionada com restrições de igualdade e desigualdade. Integrais duplas. Mudança de variáveis em integrais duplas.

(10) Equações diferenciais e em diferenças

Equações lineares de 1ª ordem e equações lineares de 2ª ordem com coeficientes constantes. Sistema de duas equações lineares de 1ª ordem homogêneo com coeficientes constantes.

CAPíTULO 1

Conjuntos e Funções

[1]

Esta parte do texto pretende apenas expor algumas dificuldades básicas, da parte talvez mais fundamental da matemática (excluindo-se a lógica). Duas referências também introdutórias, mas muito mais completas, são os livros do Terence Tao [**28**], e do Paul Halmos [**19**].

A primeira dificuldade encontrada é definir o que é um conjunto. Uma saída (questionável) é simplesmente dizer que um conjunto é uma "coleção" ou família de objetos (ou elementos ou membros). Se um objeto x faz parte de um conjunto A, dizemos que ele pertence à A e escrevemos $x \in A$ (o símbolo $\notin$ indica que quando um elemento não pertence a um conjunto).

Espera-se que o uso da palavra "coleção"acima não traga confusões. O termo coleção será a seguir utilizado para conjuntos cujos elementos são também conjuntos.

1.1. Definições básicas

Repetimos aqui o Axioma A.1.1, que garante a existência do conjunto vazio.

Axioma 1.1.1 (do conjunto vazio). Existe um conjunto que não contém nenhum elemento, i.e,

$$(\forall x)(x \notin \emptyset).$$

Considere agora dois conjuntos A e B.

- Dizemos que A está contido em B e escrevemos $A \subseteq B$ se todo elemento de A é elemento de B. Pode-se também escrever $B \supseteq A$ (lê-se B contém A) para indicar $A \subseteq B$.
- Se A não está contido em B escrevemos $A \nsubseteq B$.
- Dizemos que dois conjuntos A e B são iguais, e escrevemos $A = B$ se $A \subseteq B$ e $B \subseteq A$.
- Se não forem iguais, dizemos que são diferentes e escrevemos $A \neq B$.
- Também escrevemos $A \subsetneq B$ se $A \subseteq B$ mas $A \neq B$. Dizemos neste caso que A está *propriamente* contido em B.

O seguinte axioma é importante, nos garante que a "forma usual" de definir conjuntos é "segura," ou seja, quando definimos um conjunto obtemos um e apenas um conjunto (mesmo que seja vazio).

Axioma 1.1.2 (da especificação). Seja A um conjunto, e para cada $x \in A$, seja $P(x)$ uma proposição. Então existe um único conjunto B composto de todos os elementos x de A tais que $P(x)$ seja verdade.

[1]Última Atualização: 25/11/2019

O conjunto acima é denotado por $\{x \in A : P(x)\}$. Quando o conjunto A é claro pelo contexto, podemos escrever simplesmente $\{x : P(x)\}$. Este conjunto é formado por *todos os elementos* x que estejam em A e tais que a propiedade $P(x)$ seja verdadeira. Uma última forma de denotar os conjuntos é simplesmente descrever seus elementos entre as chaves.

Por exemplo, o conjunto dos números pares pode ser denotado por

$$\{x \in \mathbb{Z} : x \text{ é divisível por } 2\}.$$

Pode-se ainda usar a *definição construtiva* $\{2x : x \in \mathbb{Z}\}$, ou, sendo um pouco menos formal, enumerar todos os elementos do conjunto: $\{\dots, -4, -2, 0, 2, 4, 6, \dots\}$.

Vale aqui descrever uma situação interessante dada pelo *Paradoxo de Russel*. É natural perguntar-se o quão grande podem ser conjuntos. Por exemplo, existe um conjunto U tal que todos os conjuntos existentes sejam *elementos* de U? Se U existe, então, pelo Axioma da especificação (Axioma 1.1.2) podemos formar

$$R = \{x \in U : x \text{ é conjunto e } x \notin x\}.$$

Então $R \notin U$. De fato, se $R \in U$, então $R \in R$ ou $R \notin R$. Vamos dividir em dois casos:

(1) Se $R \in R$, então $R \notin R$ pois por definição, R é formado pelos conjuntos que *não* se autocontém.
(2) Se $R \notin R$, então R não satisfaz as propriedades que definem R. No caso de *não* se autoconter. Logo $R \in R$.

Em ambas possibilidades (1) e (2) obtemos absurdos. Logo $R \notin U$. Mas U é exatamente o conjunto que contém *todos* os outros.... Somos levados a concluir que tal conjunto U não pode existir.

O próximo passo é definir as operações usuais. Por incrível que possa parecer, o mais difícil é definir a união entre dois conjuntos, e para isto é necessário um axioma.

Axioma 1.1.3 (da união). Para qualquer coleção de conjuntos, existe um conjunto que contém todos os elementos pertencentes a pelo menos um conjunto da coleção.

Podemos agora definir a união entre dois conjuntos A e B. Para tanto, note que pelo Axioma da união, existe um conjunto U que contém todos os elementos de A e de B. Definimos então $A \cup B = \{x \in U : x \in A \text{ ou } x \in B\}$.

Observe entretanto a seguinte armadilha. O Axioma da união não garante que o tal conjunto contendo A e B é único, somente garante que existe. Podemos ter por exemplo um outro conjunto $\hat{U}$ contendo A e de B. Seja agora $C = \{x \in \hat{U} : x \in A \text{ ou } x \in B\}$. Para a união ser definida de forma única, temos que garantir que $C = A \cup B$. Isto é verdade, e para provar basta argumentar que $C \subseteq A \cup B$ e $C \supseteq A \cup B$.

Com o Axioma da especificação, podemos definir as seguintes operações.

- O conjunto interseção entre A e B é $A \cap B = \{x \in A : x \in B\}$. Dizemos que dois conjuntos A e B são *disjuntos* se $A \cap B = \emptyset$.
- O conjunto diferença A menos B é $A \backslash B = \{x \in A : x \notin B\}$. O conjunto resultante também denotado por $A - B$ e chamado de complemento de B em relação à A.
- Quando é claro quem é o conjunto A, denotamos $A \backslash B$ por $\mathcal{C}(B)$, e o chamamos de complemento de B.

Observação. É fácil generalizar os conceitos acima para uniões e interseções arbitrárias de conjuntos. Por exemplo, dado $n \in \mathbb{N}$ e conjuntos $A_1, \dots, A_n$, definimos

$$\cup_{i=1}^{n} A_i = \{x : x \in A_i \text{ para algum } i \in \{1, \dots, n\}\}.$$

Outra forma é definir $I = \{1, \dots, n\}$ e escrever

$$\cup_{i=1}^{n} A_i = \cup_{i \in I} A_i = \{x : \text{ existe } i \in I \text{ tal que } x \in A_i\}.$$

É simples generalizar o conceito acima para conjuntos $A_1, A_2, \dots,$ bastando para tal considerar $I = \mathbb{N}$:

$$\cup_{i=1}^{\infty} A_i = \cup_{i \in \mathbb{N}} A_i = \{x : \text{ existe } i \in \mathbb{N} \text{ tal que } x \in A_i\}.$$

Em termos de operações entre conjuntos, é útil a regra de *De Morgam*, que diz que para conjuntos E_n, onde $n \in \mathbb{N}$, temos que

$$\mathcal{C}(\cup_{i \in \mathbb{N}} E_n) = \cap_{i \in \mathbb{N}} \mathcal{C}(E_n), \qquad \mathcal{C}(\cap_{i \in \mathbb{N}} E_n) = \cup_{i \in \mathbb{N}} \mathcal{C}(E_n). \tag{1.1.1}$$

Outro conceito útil é o de *par ordenado*. Dados dois elementos, ou objetos a e b, formamos o par (a, b), e chamamos a e b de (primeiro e segundo) componentes de (a, b). Dizemos (definimos) que um par ordenado é *igual* a outro se os respectivos componentes forem iguais, i.e., $(a, b) = (a', b')$ se $a = a'$ e $b = b'$.

Do ponto de vista axiomático, não é claro que dados dois elementos, exista o par ordenado formado por eles. Uma forma de se definir o par ordenado (a, b) é por $\{a, \{a, b\}\}$.

Lema 1.1.4. Dados elementos a, b, defina o par ordenado (a, b) como o conjunto $\{a, \{a, b\}\}$. Mostre então que $(w, x) = (y, z)$ se e somente se $w = y$ e $y = z$.

O mais importante na verdade, é como pares ordenados são formados (por elementos de dois conjuntos) e quando são iguais (quando os componentes são iguais).

Definimos agora *produtos cartesianos*. Dados dois conjuntos A e B, definimos o conjunto $A \times B = \{(a, b) : a \in A, \, b \in B\}$ como sendo o composto pelos pares ordenados.

Observação. A extensão destes conceitos para *n-úplas* ordenadas e produtos cartesianos com n conjuntos é natural.

1.2. Relações e partições

Chamamos R de *relação entre A e B* se R é subconjunto de $A \times B$. Similarmente, dizemos que $a \in A$ e $b \in B$ são *relacionados* se $(a, b) \in \mathrm{R}$. Uma *relação binária* num conjunto A é um subconjunto $\mathrm{R} \subseteq A \times A$. Dado $a, \, b \in A$, denotamos $(a, b) \in \mathrm{R}$ por $a \, \mathrm{R} \, b$, e $(a, b) \notin \mathrm{R}$ por $a \not\mathrm{R} \, b$.

Exemplo 1.1. Nos reais, $=$, $\geq$, $<$, etc definem relações binárias. Considere por exemplo $A = \{1, 2, 3\}$, e defina $< \subseteq A \times A$ por $< = \{(1, 2), (1, 3), (2, 3)\}$. Então $1 < 2$, $1 < 3$ e $2 < 3$.

Definição 1.2.1. *Dizemos que uma relação R em A é:*

i) completa: para todo $a, \, b \in A$ tem-se $a \, R \, b$ ou $b \, R \, a$
ii) transitiva: para todo $a, \, b, \, c \in A$ tais que $a \, R \, b$ e $b \, R \, c$ tem-se $a \, R \, c$
iii) reflexiva: para todo $a \in A$ tem-se $a \, R \, a$
iv) simétrica: para todo $a, \, b \in A$ tais que $a \, R \, b$ tem-se $b \, R \, a$

v) assimétrica: para todo $a,\, b \in A$ *tais que* $a\,R\,b$ *tem-se* $b \not R\, a$
vi) antissimétrica: para todo $a,\, b \in A$ *tais que* $a\,R\,b$ *e* $b\,R\,a$ *tem-se* $a = b$

Uma *relação de equivalência* $\sim$ num conjunto A é uma relação binária reflexiva, simétrica e transitiva. Um exemplo trivial de relação de equivalência é a relação de igualdade $=$.

EXEMPLO 1.2. Seja $\mathbb{Z}^* = \mathbb{Z}\backslash\{0\}$. Então a relação

$$(a,b) \sim (c,d) \iff ad = bc$$

é de equivalência. De fato, note que $\sim$ é

(1) reflexiva: $(a,b) \sim (a,b)$ pois $ab = ba$.
(2) simétrica: seja $(a,b) \sim (c,d)$. Então, por definição, $ad = bc$. Então $(c,d) \sim (a,b)$ pois $bc = ad$.
(3) transitiva: seja $(a,b) \sim (c,d)$ e $(c,d) \sim (m,n)$. Segue-se por definição que $ad = bc$ e $cn = dm$. Quero mostrar que $an = bm$. Mas $adn = bcn = bdm$. Como $d \neq 0$, temos que $adn = bdm$. Portanto, $(a,b) \sim (m,n)$.

Seja agora um conjunto não vazio X e $\mathcal{P}(X)$ o *conjunto das partes de* X, i.e., é a coleção contendo todos os subconjuntos de X:

$$\mathcal{P}(X) = \{A :\ A \subseteq X\}.$$

Dada uma relação de equivalência em X e $x \in X$, podemos definir a classe de equivalência de x como sendo

$$[x] = \{\hat{x} \in X :\ \hat{x} \sim x\}.$$

Denotamos o conjunto de todas as classes de equivalência de X por $X/\sim\subseteq \mathcal{P}(X)$, onde

$$X/\sim = \{[x] :\ x \in X\}.$$

TEOREMA 1.2.2 ([**13**]). *Seja* $\sim$ *uma relação de equivalência definida num conjunto não-vazio* X*. Então*

(1) para todo $x \in X$, $x \in [x]$
(2) para todo $x,\, y \in X$, $x \sim y$ *se e somente se* $[x] = [y]$
(3) para todo $x,\, y \in X$, $x \not\sim y$ *se e somente se* $[x] \cap [y] = \emptyset$

O teorema acima nos diz que as classes de equivalência "dividem" ou "particionam" um conjunto X. Considere a seguinte definição formal. Uma coleção $\mathcal{T} \subset \mathcal{P}(X)$ é uma *partição* de X se

(1) $\emptyset \notin \mathcal{T}$
(2) para todo $A,\, B \in \mathcal{T}$, temos $A = B$ ou $A \cap B = \emptyset$
(3) para todo $x \in X$ existe conjunto $A \in \mathcal{T}$ tal que $x \in A$

Por exemplo, $\{\mathbb{R}_{<0}, \{0\}, \mathbb{R}_{>0}\}$ define uma partição de $\mathbb{R}$.

COROLÁRIO 1.2.3 (do Teorema 1.2.2). Seja $\sim$ uma relação de equivalência definida num conjunto não-vazio X. Então $X/\sim$ é uma partição de X.

A volta do corolário acima também vale, ou seja, uma partição define uma relação de equivalência, como nos mostra o teorema a seguir.

TEOREMA 1.2.4 ([**13**]). *Seja X conjunto não vazio, e $\mathcal{T}$ uma partição de X. Seja a relação de equivalência $\sim$ dada por*

$$x \sim y \iff (\exists A \in \mathcal{T})(x \in A \wedge y \in A),$$

para quaisquer x, $y \in X$. Além disto, as classes de equivalência definidas por $\sim$ são exatamente os elementos de $\mathcal{T}$, i.e., $X/\sim = \mathcal{T}$.

1.3. Funções

Da definição de relação vem o importante conceito de função. Uma *função entre A e B* nada mais é que uma relação entre A e B, e sendo assim $f \subseteq A \times B$. Esta relação entretanto satisfaz a seguinte restrição: para todo $a \in A$ existe um único $b \in B$ tal que $(a, b) \in f$. Denotamos esta relação especial por $f : A \to B$. Dado $a \in A$, $b \in B$, dizemos que $f(a) = b$ se $(a, b) \in f$.

Comumente nos "esquecemos"desta definição e tratamos funções de forma mais informal e direta. Este pecadilho matemático não chega a atrapalhar nossos objetivos, mas é importante ter em mente a definição formal. Na "prática", uma função é uma regra que associa a cada elemento $x \in A$, um elemento $f(x) \in B$. Chamamos o conjunto A de *domínio* da função f e o denotamos por $D(f)$. Chamamos o conjunto B de *contradomínio* da função f. Escrevemos $f : A \to B$, ou ainda

$$\begin{aligned} f : A &\to B \\ x &\mapsto f(x). \end{aligned}$$

Se $E \subseteq A$, chamamos de *imagem de E* ao conjunto

$$f(E) = \{f(x) : x \in E\}.$$

Similarmente, dado um conjunto H, chamamos de *imagem inversa de H* o conjunto

$$f^{-1}(H) = \{x : f(x) \in H\}.$$

Se $f(A) = B$ dizemos que f é *sobrejetiva* (ou simplesmente *sobre*). Dizemos que f é *injetiva* (ou *biunívoca* ou *um a um* ou 1-1) quando, dados a, $a' \in D(f)$, se $f(a) = f(a')$ então $a = a'$. Numa forma mais compacta, escrevemos que para todo a, $a' \in D(f)$ temos

$$f(a) = f(a') \implies a = a'.$$

Se f é injetiva e sobre, a chamamos de *bijetiva* ou de uma *bijeção*.

Dado $f : A \to B$ e um subconjunto $A' \subseteq A$, podemos definir a *função restrição* $g = f|_{A'}$ onde $g : A' \to B$ é dada por $g(a') = f(a')$ para todo $a' \in A'$. Da forma análoga, dizemos que $h : A'' \to B$ é uma *extensão* de f se $A \subseteq A''$ e $h(a) = f(a)$ para todo $a \in A$.

Dados conjuntos A, B e C e funções $f : A \to B$ e $g : B \to C$, definimos a *função composta* $h : A \to C$ por $h(a) = g(f(a))$ para todo $a \in A$. Denotamos a função composta por $h = f \circ g$.

Dizemos que $g : B \to A$ é *função inversa* de f se

$$g(f(x)) = x \quad \text{para todo } x \in A, \qquad f(g(y)) = y \quad \text{para todo } y \in B.$$

Quando esta existir, denotamos a inversa de f por f^{-1}.

Observação. Note que a definição de *imagem inversa* independe de existir ou não a função inversa. Por exemplo, a função $f:\mathbb{R}\to\mathbb{R}$ dada por $f(x)=x^2$ não tem inversa. Entretanto $f^{-1}(\mathbb{R})=\mathbb{R}$.

Exemplo 1.3. Seja

$$\begin{aligned} f :&(0,4)\to\mathbb{R}\\ x&\mapsto\sqrt{x}. \end{aligned}$$

Então o domínio é $(0,4)$ e a imagem é $(0,2)$. Note que f não é invertível pois f não é sobrejetiva. Entretanto as imagens inversas

$$f^{-1}\big((1,2)\big)=(1,4),\qquad f^{-1}(\{2\})=\{4\},\qquad f^{-1}\big([-2,0)\big)=\emptyset,\qquad f^{-1}(\emptyset)=\emptyset.$$

são bem-definidas.

A seguir nos concentramos sobre a importante questão da *existência de uma função inversa*. Considerando $f:A\to B$, note que se f não for sobrejetiva, não existirá uma inversa f^{-1}. De fato, se existe $b\in B$ tal que $f(a)\neq b$ para todo $a\in A$, então $f(g(b))\neq b$ para qualquer função $g:B\to A$. De forma análoga, se f não for injetiva, i.e., se existirem $a\neq a'$ em A tais que $f(a)=f(a')=b\in B$, entao não é possível definir $g(b)$ tal que $g(f(a))=a$ e $g(f(a'))=a'$. Estes argumentos mostram que sobrejetividade e injetividade são condições emphnecessárias para a exitência de inversa. Na verdade, estas condições são também equivalentes, como mostra o lema abaixp.

Lema 1.3.1. Sejam A e B conjuntos e considere a função $f:A\to B$. Então f é invertível se e somente se é sobrejetiva e injetiva.

Demonstração. ($\Longrightarrow$) Suponha que f seja invertível, e denote $g=f^{-1}:B\to A$. Sejam $a,a'\in A$ tais que $f(a)=f(a')$. Então $a=g(f(a))=g(f(a'))=a'$. Logo f é injetiva.

Seja agora $b\in B$, e seja $a=g(b)$. Enão $f(a)=f(g(b))=b$, e portanto f é sobre.

($\Longleftarrow$) Suponha agora f uma bijeção, i.e., sobrejetiva e injetiva. Logo, dado $b\in B$, existe um único $a\in A$ tal que $f(a)=b$. Defina então $g(b)=a$. Note que $f(g(b))=f(a)=b$ e $g(f(a))=g(b)=a$. Como b é arbitrário, definimos a função $g:B\to A$. $\square$

1.4. Exercícios

Exercício 1.1. Mostre que

(1) $\{x\in\mathbb{R}: x^2\geq 0\}=\mathbb{R}$.
(2) $\{x\in\mathbb{R}: x>0\}\subsetneq\{x\in\mathbb{R}: x^2\geq 0\}$.
(3) $\mathbb{R}\not\subseteq\{x\in\mathbb{R}: x^2\geq 0\}$.

Exercício 1.2. Mostre a regra de *De Morgam* dada em (1.1.1).

Exercício 1.3. Mostre que $\{a,a\}=\{a\}$.

Exercício 1.4. Sejam A e B dois conjuntos disjuntos, i.e., $A\cap B=\emptyset$. Seja $X=A\cup B$. Mostre que $A=X\backslash B$ e $B=X\backslash A$.

Exercício 1.5. Sejam A e B dois conjuntos, e $C=(A\backslash B)\cup(B\backslash A)$. Mostre que $C=(A\cup B)\backslash(A\cap B)$ e que $C\cap A\cap B=\emptyset$.

Exercício 1.6. Sejam $X_1, X_2, \ldots,$ conjuntos tais que $X_i \subseteq X_{i+1}$ para todo $i \in \mathbb{N}$. Mostre que $\cap_{i\in\mathbb{N}} X_i = X_1$.

Exercício 1.7. Dado um conjunto A, mostre que $A \times \emptyset = \emptyset$.

Exercício 1.8. Sejam A, B e C conjuntos. Mostre que $A\times(B\cup C) = (A\times B)\cup(A\times C)$ e $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

Exercício 1.9. Sejam A, B, C e D conjuntos. Mostre que $(A \times B) \cup (C \times D) \subseteq (A \cup C) \times (B \cup D)$. Mostre que, em geral, a inclusão não é própria.

Exercício 1.10. Seja a relação $\sim$ em $\mathbb{R}^2$ dada por $(x_1, y_1) \sim (x_2, y_2)$ se $x_1^2+y_1^2 = x_2^2+y_2^2$. Mostre que $\sim$ é uma relação de equivalência. Interprete geometricamente a relação $\sim$.

Exercício 1.11. Tentando generalizar a relação de equivalência do exercício 1.10, dada uma função $f : \mathbb{R}^2 \to \mathbb{R}$, defina a relação $(x_1, y_1) \sim (x_2, y_2)$ se $f(x_1, y_1) = f(x_2, y_2)$. Determine se é verdadeiro ou falso que $\sim$ seja de equivalência. Prove suas afirmativas.

Exercício 1.12. Interprete $\sim$ do exercício 1.11 considerando as curvas de nível $c(t) = \{(x, y) : f(x, y) = t\}$. Mostre que $\{c(t) : t \in \mathbb{R}\}$ determina uma partição de $\mathbb{R}^2$.

Exercício 1.13. Seja X conjunto não vazio e $\sim$ uma relação de equivalência em X. Mostre que $X/\sim$ define uma partição de X.

Exercício 1.14. Seja X um conjunto e $\succcurlyeq$ uma *preordenação* em X, i.e., é uma relação reflexiva e transitiva. Defina a relação $\sim$ tal que $a \sim b$ se $a \succcurlyeq b$ e $b \succcurlyeq a$. Mostre que $\sim$ é relação de equivalência.

Exercício 1.15. Sejam A, B conjuntos e $f : A \to B$ bijeção. Se $f^{-1} : B \to A$ for a função inversa de f, mostre que f^{-1} é bijeção.

Exercício 1.16. Sejam A, B e C conjuntos e $f : A \to B$, $g : B \to C$ bijeções. Mostre que a função composta $g \circ f : A \to C$ dada por $g \circ f(x) = g(f(x))$ é bijeção. Se f^{-1} e g^{-1} forem as funções inversas de f e g, quem é $(g \circ f)^{-1}$? Justifique suas conclusões.

Exercício 1.17. Seja A um conjunto e $f : A \to B$ injetiva. Mostre que a função $f : A \to f(A)$ é bijeção.

Exercício 1.18. Seja $g : X \to Y$ sobrejetiva. Mostre que existe $f : Y \to X$ tal que $g \circ f(y) = y$ para todo $y \in Y$.

CAPíTULO 2

Noções de geometria analítica

[1]

Neste capítulo falaremos sobre noções como coordenadas, distância, vetores, produtos escalar e vetorial, perpendicularidade, equações da reta no plano e espaço, equações de planos, inequações lineares, parábolas, hipérboles. Uma boa referência é [**26**]

Consideraremos o $\mathbb{R}^n$ o como o conjunto das n-úplas ordenadas de números reais, como definido abaixo.

Definição 2.0.1. *Seja $\mathbb{R}^n$ o conjunto das n-úplas ordenadas de números reais, i.e,*

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, \dots, x_n) : x_i \in \mathbb{R} \text{ para } i = 1, \dots, n\}.$$

Definimos então as operações produto por escalar e soma da seguinte forma:

$$\alpha\mathbf{x} = (\alpha x_1, \dots, \alpha x_n), \qquad \mathbf{x} + \mathbf{y} = (x_1 + y_1, \dots, x_n + y_n),$$

onde $\mathbf{x} = (x_1, \dots, x_n)$ e $\mathbf{y} = (y_1, \dots, y_n)$ estão em $\mathbb{R}^n$, e $\alpha \in \mathbb{R}$. Pode-se checar que $\mathbb{R}^n$ é espaço vetorial com as operações acima descritas.

2.1. Coordenadas

Seja $\mathcal{B}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ base do $\mathbb{R}^n$. Então, segundo o Teorema 3.5.11, todo vetor do $\mathbb{R}^n$ pode ser escrito de forma única como combininação linear dos vetores de $\mathcal{B}$, i.e., dado um vetor $\mathbf{w} \in \mathbb{R}$ qualquer, existem números reais $\alpha_1, \dots, \alpha_n$ que são os únicos tais que

$$\mathbf{w} = \alpha_1\mathbf{v}_1 + \cdots \alpha_n\mathbf{v}_n.$$

Dizemos então que $\alpha_1, \dots, \alpha_n$ são as *coordenadas* de $\mathbf{w}$ na base $\mathcal{B}$.

A base mais simples que existe é a *base canônica*, dada por $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, onde, para $i \in \{1, \dots, n\}$, o vetor $\mathbf{e}_i$ é definido tal que a iésima coordenada vale um e as demais coordenadas valem zero, i.e.,

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \dots, 0), \quad \dots, \quad \mathbf{e}_n = (0, 0, \dots, 0, 1).$$

Chamamos este vetores de *vetores da base canônica*. Note que podemos escrever um ponto $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ como $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$. Neste caso, $x_1, \dots, x_n$ são as *coordenadas de* $\mathbf{x}$ *na base canônica*.

[1]Última Atualização: 11/04/2020

Existe uma identificação natural dos pontos em $\mathbb{R}^n$ com suas coordenadas na base canônica. Usaremos neste texto a seguinte notação. Para cada $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, indicaremos por $\vec{\mathbf{x}} \in \mathbb{R}^{n\times 1}$ a matriz coluna das coordenadas na base canônica dada por

$$\vec{\mathbf{x}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{2.1.1}$$

Na verdade não seremos tão preciosistas e escreveremos que $\vec{\mathbf{x}} \in \mathbb{R}^n$ também.

Exemplo 2.1. Nem sempre as bases são tão simples. Por exemplo, $\{(1,1);(0,1)\}$ determina uma base em $\mathbb{R}^2$. Para determinar as coordenadas de um vetor (a,b) qualquer em $\mathbb{R}^2$, temos que achar $\alpha_1, \alpha_2 \in \mathbb{R}$ tais que

$$(a,b) = \alpha_1(1,1) + \alpha_2(0,1).$$

Nesta base, o vetor $(1,1)$ tem 1 e 0 como coordenadas, e o vetor $(0,1)$ tem 0 e 1 como coordenadas, pois

$$(1,1) = 1\cdot(1,1) + 0\cdot(0,1),$$
$$(0,1) = 0\cdot(1,1) + 1\cdot(0,1).$$

Já o vetor $(-2,-1)$ tem -2 e 1 como coordenadas, pois se

$$(-2,-1) = \alpha_1(1,1) + \alpha_2(0,1),$$

então $\alpha_1 = -2$ e $\alpha_1 + \alpha_2 = -1$. Logo $\alpha_2 = 1$.

2.2. Distância, norma, produtos escalar e vetorial

Dados dois vetores $\mathbf{x} = (x_1, x_2)$ e $\mathbf{y} = (y_1, y_2)$ do $\mathbb{R}^2$, a distância $d(\mathbf{x}, \mathbf{y})$ entre eles é dada pelo "tamanho" do vetor $\mathbf{x} - \mathbf{y} = (x_1 - y_1, x_2 - y_2)$. Para medir tamanho de vetores, usamos a noção de *norma*. No $\mathbb{R}^2$, definimos a norma de um vetor $\mathbf{x} = (x_1, x_2)$ por

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}.$$

Esta é a *norma euclidiana*, que no caso mais geral, em $\mathbb{R}^n$, é dada por

$$\|(x_1, \dots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Voltando ao conceito de distância, temos que a distância entre dois pontos do $\mathbb{R}^2$ dados por $\mathbf{x}$ e $\mathbf{y}$, nas várias normas, é dada por

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}, \qquad \|\mathbf{x} - \mathbf{y}\|_\infty = \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

Exemplo 2.2. Considere o vetor $(3,4)$. Então $\|(3,4)\| = \sqrt{9+16} = 5$ e $\|(3,4)\|_\infty = 4$.

Os dois exemplos de norma acima são casos particulares da noção mais geral de norma, que se aplica em espaços vetoriais em geral, ver a Definição 3.6.2.

Uma outra importante *ferramenta* matemáticas quando se trabalha em espaços vetoriais é o conceito de produto interno.

Em $\mathbb{R}^2$, se $\mathbf{x} = (x_1, x_2)$, e $\mathbf{y} = (y_1, y_2)$, o produto interno canônico é dado por

$$\mathbf{x} \cdot \mathbf{y} = \vec{\mathbf{x}}^T \vec{\mathbf{y}} = x_1 y_1 + x_2 y_2.$$

Em $\mathbb{R}^n$, para $\mathbf{x} = (x_1, \dots, x_n)$, e $\mathbf{y} = (y_1, \dots, y_n)$, definimos

$$\mathbf{x} \cdot \mathbf{y} = \vec{\mathbf{x}}^T \vec{\mathbf{y}} = x_1 y_1 + \cdots + x_n y_n.$$

Note que podemos definir as normas euclidianas usando o produto interno

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \sqrt{x_1^2 + \cdots + x_n^2} \tag{2.2.1}$$

Assim como no caso de norma, um produto interno não precisa ser o canônico, basta obedecer algumas “regras”. Veja a Definição 3.6.1. O que é interessante é que a relação entre norma e produto interno vista em (2.2.1) é somente um exemplo do caso mais geral. Sempre que temos um produto interno, podemos definir uma norma. Isto será visto no Capítulo 3. Abaixo temos a desigualdade de *Cauchy–Schwartz* no $\mathbb{R}^n$. Deixaremos a demonstração para o caso geral visto no Teorema 3.6.3.

TEOREMA 2.2.1. *Considere a norma e o produto interno canônicos do $\mathbb{R}^n$. Então vale a desigualdade de Cauchy–Schwartz*

$$|\mathbf{x} \cdot \mathbf{y}| \le \|\mathbf{x}\| \|\mathbf{y}\| \qquad \textit{para todo } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{2.2.2}$$

Além disto, a igualdade $|\mathbf{x} \cdot \mathbf{y}| = \|\mathbf{x}\| \|\mathbf{y}\|$ vale se e somente se $\mathbf{x} = \alpha \mathbf{y}$ para algum $\alpha \in \mathbb{R}$.

Note que da desigualdade de Cauchy–Schwartz obtemos a desigualdade triangular:

$$\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|, \tag{2.2.3}$$

sempre que $\| \cdot \|$ for dada por (2.2.1).

Finalmente, dados dois vetores $\mathbf{x}$, $\mathbf{y}$ do $\mathbb{R}^n$, definimos o cosseno do ângulo formado por eles por **[26]**

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}. \tag{2.2.4}$$

Dizemos então que dois vetores $\mathbf{x}$, $\mathbf{y}$ são ortogonais, ou perpendiculares, se $\mathbf{x} \cdot \mathbf{y} = 0$. Note que devido à desigualdade de Cauchy-Schwartz (2.2.2) que o cosseno toma sempre valores entre -1 e 1.

2.3. Projeção de vetores

Considere os vetores dados por $\mathbf{x}$ e $\mathbf{y}$. Definimos a projeção ortogonal de $\mathbf{x}$ em $\mathbf{y}$ como sendo o vetor $\mathbf{w} = \alpha \mathbf{y}$ tal que $\mathbf{x} - \mathbf{w}$ é ortogonal a $\mathbf{y}$. Logo

$$\mathbf{x} \cdot \mathbf{y} = \alpha \mathbf{y} \cdot \mathbf{y} \implies \alpha = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{y}\|^2} = \frac{\|\mathbf{x}\| \cos\theta}{\|\mathbf{y}\|}.$$

2.4. A reta no plano e espaço

Uma reta r do $\mathbb{R}^n$ é um conjunto de pontos que pode ser definido por um ponto $\mathbf{x}_0$ a ela pertencente, e a uma direção $\mathbf{v}$ dada, e portanto r pode ser parametrizada por

$$r = \{\mathbf{x}_0 + t\mathbf{v} : t \in \mathbb{R}\}.$$

Analogamente, se $\mathbf{x}$, $\mathbf{y}$ são dois pontos de r, então $\mathbf{v} = \mathbf{x} - \mathbf{y}$ determina a direção da reta.

Em duas dimensões, é possível obter as *equações cartesianas* da reta, i.e., dado um ponto (x, y) da reta, é possível determinar uma relação entre as coordenadas. Suponha que $(x_0, y_0) + t(v_1, v_2$ seja uma parametrização de uma reta r. Então qualquer ponto (x, y) pode ser dado por

$$x = x_0 + tv_1, \qquad y = y_0 + tv_2$$

para algum $t \in \mathbb{R}$. Multiplicando a primeira equação por v_2, a segunda por v_1 e tirando a diferença obtemos

$$v_2 x - v_1 y = c,$$

onde $c = v_2 x_0 - v_1 y_0$.

De forma análoga, se uma reta r no plano é determinada por $ax + by + d = 0$, onde a e b não são simultaneamente nulos, então temos a seguinte dituação:

(1) caso $b = 0$, temos que a reta é simplesmente dada por $x - d/a = 0$, ou seja, é a reta vertical dada por x constante, i.e.,

$$r = \{(d/a, 0) + t\mathbf{e}_2 : t \in \mathbb{R}\}.$$

(2) suponha agora $b \neq 0$. Para determinarmos sua direção, vamos achar dois pontos pertencentes à r. Para tal, basta determinar o valos de y quando x for igual a zero e a um, por exemplo. No caso temos que $(0, -d/b)$ e $\big(1, (-d-a)/b\big)$ pertencem à r. Logo, $\big(1, (-d-a)/b\big) - (0, -d/b) = (1, -a/b)$ é paralelo à r e então $\mathbf{v} = b(1, -a/b) = (b, -a)$ também o é. Logo

$$r = \{(0, -d/b) + t\mathbf{v} : t \in \mathbb{R}\}.$$

Note que $-a/b$ é exatamente o número que indica a *inclinação* de r.

É interessante e útil notar que, em ambos os casos, (a, b) é vetor perpendicular à r e que $\mathbf{v} = (b, -a)$ é direção da reta.

Observação. Vemos portanto que em duas dimensões, podemos escrever a reta via parametrização ou usando as equações cartesianas, e que de uma forma obtemos a outra.

Exemplo 2.3. Seja r_1 reta dada por

$$r_1 = \{(1, 2) + t(3, -1) : t \in \mathbb{R}\}.$$

Ache r_2 passando por $(1, 1)$ e paralela à r_1.

Solução. A solução é simples pois como r_2 é paralela à r_1, ambas tem a mesma direção, que no caso é $(3, -1)$. Como $(1, 1) \in r_2$, então $r_2 = \{(1, 1) + t(3, -1) : t \in \mathbb{R}\}$.

O exemplo abaixo lida com interseção de retas.

Exemplo 2.4 (interseção de retas). Determine se as retas $r_1 = \{(0,0,1) + t(1,-1,3) : t \in \mathbb{R}\}$ e $r_2 = \{(1,2,0) + t(0,3,-1) : t \in \mathbb{R}\}$ se interseptam, e em qual ponto.

Solução. Note que as retas se interseptam se e somente se elas tiverem um ponto em comum, ou seja se existirem t, s tais que $(1,-2,1) + t(1,-1,3) = (1,2,0) + s(0,3,-1)$. Isto equivale a resolver o sistema

$$1 + t = 1; \qquad -2 - t = 2 + 3s; \qquad 1 + 3t = -s.$$

Este sistema de equações pode ter uma, zero ou infinitas soluções.

Exemplo 2.5 (ângulo entre retas). Determine o ângulo entre as retas dadas por $y = ax+b$ e $y = cx + d$.

Solução. Vimos acima que $(1,a)$ e $(1,c)$ determinam as direções destas retas. Portanto, se θ é o ângulo entre as duas retas, temos que

$$\cos\theta = \frac{(1,a)\cdot(1,c)}{\|(1,a)\|\|(1,c)\|} = \frac{1+ac}{\sqrt{(1+a^2)(1+c^2)}}$$

Note que $(1,a)\cdot(1,c)$ indica que este cálculo resultará no cosseno do *maior ângulo*, e que portanto pode ser conveniente utilizar como resposta $\pi - \theta$.

Exemplo 2.6 (distância de ponto a reta). Ache a distância d entre um ponto $\mathbf{p} \in \mathbb{R}^2$ e a reta r determinada por $by + ax + c = 0$.

Solução. Sem perda de generalidade, suponha $b = 1$. Sabemos que r tem direção $\mathbf{v} = (1,-a)$. Seja $\mathbf{q} \in r$ tal que $\mathbf{p} - \mathbf{q}$ seja perpendicular a $\mathbf{v}$. Portanto $\mathbf{p} - \mathbf{q} = \alpha(a,1)$ para algum $\alpha \in \mathbb{R}$, e a distância é dada por $d = \|\mathbf{p} - \mathbf{q}\| = |\alpha|(1+a^2)^{1/2}$. Como $\mathbf{q} \in r$, então

$$q_2 = -aq_1 - c, \qquad p_1 - q_1 = a\alpha, \qquad p_2 - q_2 = \alpha.$$

e obtemos da segunda equação:

$$q_1 = p_1 - a(p_2 - q_2) = p_1 - ap_2 - a^2q_1 - ac$$

Portanto $q_1 = (p_1 - ap_2 - ac)/(1+a^2)$. Logo

$$-a\alpha = -p_1 + q_1 = -p_1 + \frac{p_1 - ap_2 - ac}{1+a^2} = -a\frac{p_1a + p_2 + c}{1+a^2}$$

Se $a \neq 0$, temos $\alpha = p_1a + p_2 + c/(1+a^2)$, e portanto a distância é dada por

$$d = \frac{|p_1a + p_2 + c|}{\sqrt{1+a^2}}.$$

Note que pelas contas acima, se $a = 0$, então $q_1 = p_1$ e $q_2 = -c$. Então $\alpha = p_2 + c$, e a fórmula acima se aplica mesmo neste caso.

Em geral, se a reta é dada por $ax + by + c = 0$, então a distância é dada por

$$\frac{|ap_1 + bp_2 + c|}{\sqrt{a^2+b^2}}.$$

2.5. Cônicas no plano

Uma cônica no plano é o conjunto de pontos $\{(x,y) \in \mathbb{R}^2 : ax^2+bxy+cy^2+dx+ey+f = 0\}$, onde $a,\dots,f \in \mathbb{R}$. Pedimos ainda que a, b ou c seja diferente de zero. Uma outra forma de exigir isto é impor que $|a|+|b|+|c| \neq 0$. Se definirmos a forma quadrática $Q(x,y) = ax^2+bxy+cy^2$, e a forma linear $F(x,y) = dx+ey$, temos que $Q(x,y)+F(x,y)+f = 0$. A seguir mostramos exemplos de cônicas em sua *forma reduzida*.

EXEMPLO 2.7.
$$\frac{x^2}{a^2}+\frac{y^2}{b^2} = 1 \text{ (elipse)}, \qquad \frac{x^2}{a^2}-\frac{y^2}{b^2} = 1 \text{ (hipérbole)}, \qquad x^2-ey = 0 \text{ (parábola)}.$$
No caso da elipse e da hipérbole, impomos que a e b sejam não nulos. No caso da parábola, a imposição é que e seja não zero.

Podemos ainda ter casos degenerados, como o exemplo a seguir nos mostra.

EXEMPLO 2.8. O caso da hipérbole degenerada é dado, para a e b não nulos, por $(x^2/a^2)-(y^2/b^2) = 0$, o que implica em $y = \pm bx/a$.

No caso da parábola degenerada, para $a \neq 0$, temos $ax^2-f = 0$, e portanto $x = \pm\sqrt{f/a}$.

Elipses degeneradas são dadas por $ax^2+by^2 = 0$, com a e b positivos. Logo $x = y = 0$ é o único ponto da cônica.

Finalmente, temos cônicas dadas por conjuntos vazios (elipses e parábolas degeneradas), se $ax^2+by^2+r^2 = 0$ e $r \neq 0$, $a \geq 0$ e $b \geq 0$, e a ou b são não nulos.

Como dissemos, todos os exemplo acima estão em sua forma reduzida, mas este não é a forma mais geral possível. Porém, todas as cônicas podem ser reescritas em forma reduzida após mudanças de coordenadas. Veja o exemplo abaixo.

EXEMPLO 2.9. ([**4**]) Seja a cônica dada por $2x^2+2y^2+4xy+4\sqrt{2}x+12\sqrt{2}y-8 = 0$. Para reescrevê-la na forma reduzida, seguimos os passos abaixo.

Passo i: reescrever a cônica em forma matricial:
$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 4\sqrt{2} & 12\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} - 8 = 0, \text{ onde } A = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}.$$

Passo ii: diagonalizar a matriz A. Primeiro vemos que A tem como autovalores $\lambda_1 = 0$ e $\lambda_2 = 4$ e os correspondentes autovetores
$$\vec{\mathbf{v}}^1 = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 \\ 1 \end{bmatrix}, \qquad \vec{\mathbf{v}}^2 = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$
Note que se definirmos a matriz $M = \begin{bmatrix} \vec{\mathbf{v}}^1 & \vec{\mathbf{v}}^2 \end{bmatrix}$, então $M^{-1} = M$ e
$$\begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} = M^{-1}AM = MAM.$$
Se (x_1,y_1) são as coordenadas de (x,y) na base $\{\mathbf{v}^1,\mathbf{v}^2\}$, i.e., se
$$\vec{\mathbf{x}} = x_1\vec{\mathbf{v}}^1 + y_1\vec{\mathbf{v}}^2 = M\begin{bmatrix} x_1 \\ y_1 \end{bmatrix},$$
então
$$Q(x,y) = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \end{bmatrix} M^T AM \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Passo iii: reescrever a parte linear em termos de (x_1, x_2). Note que temos

$$\begin{bmatrix} 4\sqrt{2} & 12\sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 4\sqrt{2} & 12\sqrt{2} \end{bmatrix} M \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

Passo iv: eliminar as constantes. Note que em termos de (x_1, x_2) a cônica é dada por

$$\begin{bmatrix} x_1 & y_1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} 4\sqrt{2} & 12\sqrt{2} \end{bmatrix} M \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - 8 = 0.$$

Reescrevendo a expressão acima em sua forma não matricial, temos que

$$y_1^2 + 2x_1 + 4y_1 - 2 = 0.$$

Completando quadrados temos que $(y_1^2 + 4y_1 + 4) + 2x_1 - 6 = 0$, e portanto

$$(y_1 + 2)^2 + 2(x_1 - 3) = 0.$$

Introduzindo novas coordenadas $y_2 = y_1 + 2$ e $x_2 = x_1 - 3$ obtemos que

$$y_2^2 + 2x_2 = 0,$$

e a cônica é uma parábola.

As contas acima podem ser feitas em geral, como mostra o resultado abaixo.

Teorema 2.5.1. *Seja a cônica definida por $ax^2 + bxy + cy^2 + dx + ey + f = 0$, i.e.,*

$$\begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + f = 0, \text{ onde } A = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix},$$

e sejam λ_1 e λ_2 os autovalores de A. Então

(1) se $\lambda_1\lambda_2 > 0$, então a cônica é uma elipse
(2) se $\lambda_1\lambda_2 < 0$, então a cônica é uma hipérbole
(3) se $\lambda_1\lambda_2 = 0$, então a cônica é uma parábola

Corolário 2.5.2. Como o sinal de $\lambda_1\lambda_2$ é o mesmo de $-(b^2 - 4ac)$, podemos concluir que

(1) se $b^2 - 4ac < 0$, então a cônica é uma elipse
(2) se $b^2 - 4ac > 0$, então a cônica é uma hipérbole
(3) se $b^2 - 4ac = 0$, então a cônica é uma parábola

Olhemos com mais cuidado as propriedades de cada cônica.

2.5.1. Elipses. Dados dois pontos $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ (os *focos* da elipse) e $r > d(\boldsymbol{f}_1, \boldsymbol{f}_2)$, a elipse será dada pelos pontos $\mathbf{x} \in \mathbb{R}^2$ tais que

$$d(\mathbf{x}, \boldsymbol{f}_1) + d(\mathbf{x}, \boldsymbol{f}_2) = r.$$

Considere por simplicidade os focos situados no eixo x e equidistantes da origem, i.e., $\boldsymbol{f}_1 = (-c, 0)$, $\boldsymbol{f}_2 = (c, 0)$, para algum $a > 0$. Então tomando $a = r/2$, obtemos os pontos $\boldsymbol{A}_1 = (-a, 0)$ e $\boldsymbol{A}_2 = (a, 0)$ da elipse e formamos seu *eixo maior*. Similarmente, tomando $b > 0$ tal que $2(b^2 + c^2)^{1/2} = r$, obtemos os pontos $\boldsymbol{B}_1 = (0, -b)$ e $\boldsymbol{B}_2 = (0, b)$ da elipse e formamos seu *eixo menor*. Neste caso temos que $r = d(\boldsymbol{A}_1, \boldsymbol{A}_2) = 2a$ e um ponto $\boldsymbol{P} = (x, y)$ qualquer da elipse satisfaz

$$d(P, \boldsymbol{f}_1) + d(P, \boldsymbol{f}_2) = d(\boldsymbol{A}_1, \boldsymbol{A}_2).$$

Logo $\sqrt{(x-c)^2+y^2}+\sqrt{(x+c)^2+y^2}=2a$, e então

$$\sqrt{(x-c)^2+y^2}=2a-\sqrt{(x-c)^2+y^2}$$

Elevando ao quadrado e rearrumando a expressão, obtemos $a^2-cx=a\sqrt{(x-c)^2+y^2}$. Elevando novamente ao quadrado, obtemos

$$a^2(a^2-c^2)=(a^2-c^2)x^2+a^2y^2$$

Usando agora que $a^2=b^2+c^2$, obtemos $a^2b^2=b^2x^2+a^2y^2$, i.e.,

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1.$$

2.5.2. Hipérboles. Dados dois pontos $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ (os *focos* da hipérbole) e $r<d(\boldsymbol{f}_1,\boldsymbol{f}_2)$, a hipérbole será dada pelos pontos $\mathbf{x}\in\mathbb{R}^2$ tais que

$$|d(\mathbf{x},\boldsymbol{f}_1)-d(\mathbf{x},\boldsymbol{f}_2)|=r.$$

Caso os focos da hipérbole estejam situados em $(-c,0)$, $(c,0)$, sua equação será dada por

$$\frac{x^2}{a^2}-\frac{y^2}{b^2}=1.$$

2.5.3. Parábolas. Dado um ponto $\boldsymbol{f}$ e uma reta r, a parábola é definida como o conjunto de pontos $\mathbf{x}$ tais que

$$d(\mathbf{x},\boldsymbol{f})=d(\mathbf{x},r),$$

onde $d(\mathbf{x},r)$ indica a distância de $\mathbf{x}$ a reta r. O ponto $\boldsymbol{f}$ é denominado *foco* da parábola, e r sua *diretriz*.

2.6. Produto vetorial

Uma outra operação com vetores é o *produto vetorial*. Sejam $\mathbf{x}$, $\mathbf{y}$ vetores em R^3. Então definimos

$$\mathbf{x}\times\mathbf{y}=(x_2y_3-x_3y_2,x_3y_1-x_1y_3,x_1y_2-x_2y_1).$$

Uma outra forma de escrever é

$$\mathbf{x}\times\mathbf{y}=\left(\det\begin{pmatrix}x_2 & x_3\\ y_2 & y_3\end{pmatrix},-\det\begin{pmatrix}x_1 & x_3\\ y_1 & y_3\end{pmatrix},\det\begin{pmatrix}x_1 & x_2\\ y_1 & y_2\end{pmatrix}\right),$$

onde $\det(A)$ denota o determinante da matriz A. Algumas propriedades do produto vetorial são dadas abaixo:

(1) $\mathbf{x}\times\mathbf{y}=-\mathbf{y}\times\mathbf{x}$
(2) $\mathbf{x}\times\mathbf{y}$ é ortogonal a $\mathbf{x}$ e $\mathbf{y}$
(3) $(\alpha\mathbf{x})\times\mathbf{y}=\alpha(\mathbf{x}\times\mathbf{y})=\mathbf{x}\times(\alpha\mathbf{y})$, para todo $\alpha\in\mathbb{R}$
(4) $(\mathbf{x}+\mathbf{y})\times\mathbf{z}=\mathbf{x}\times\mathbf{z}+\mathbf{y}\times\mathbf{z}$
(5) $\|\mathbf{x}\times\mathbf{y}\|^2=\|\mathbf{x}\|^2\|\mathbf{y}\|^2-(\mathbf{x}\cdot\mathbf{y})^2$
(6) $\|\mathbf{x}\times\mathbf{y}\|=\|\mathbf{x}\|\|\mathbf{y}\|\sin\theta$
(7) $\mathbf{x}\times\mathbf{x}=0$
(8) (produto misto) $\mathbf{x}\cdot(\mathbf{y}\times\mathbf{z})=\det\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\\ \mathbf{z}\end{pmatrix}=\det\begin{pmatrix}x_1 & x_2 & x_3\\ y_1 & y_2 & y_3\\ z_1 & z_2 & z_3\end{pmatrix}$
(9) $\mathbf{x}\times\mathbf{y}=0$ se e somente se $\mathbf{x}=\alpha\mathbf{y}$ para algum $\alpha\in\mathbb{R}$

(10) $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$, $\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1$, $\mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$

Uma interessante aplicação das propriedades do produto vetorial é na determinação de áreas e volumes. Considere o paralelograma gerado pelos vetores não colineares $\mathbf{u}$, $\mathbf{v}$. Então o paralelograma gerado temárea $\|\mathbf{u}\|h$, onde h é a altura do parelograma. Mas $h = \|\mathbf{v}\| \sin\theta$, sendo θ o ângulo gerado pelos dois vetores. Portanto a área é dada por $\|\mathbf{u}\times\mathbf{v}\|$, onde usamos a propriedade (6) acima.

Considere agora o paralelepípedo formado por veotres não linearmente dependentes $\mathbf{u}$, $\mathbf{v}$ e $\mathbf{w}$. Para calcular o volume, usaremos o produto misto (8) acima. De fato, o volume é dada pela área $\|\mathbf{u}\times\mathbf{v}\|$ do palelograma formado por $\mathbf{u}$ e $\mathbf{v}$, vezes a altura do paralelepípedo vezes $\|\mathbf{w}\||\cos\alpha|$, onde α é o ângulo formado por $\mathbf{u}\times\mathbf{v}$ e $\mathbf{w}$. Logo o volume é dado por

$$\|\mathbf{u}\times\mathbf{v}\|\|\mathbf{w}\|\cos\alpha = \|\mathbf{u}\times\mathbf{v}\|\|\mathbf{w}\|\frac{|(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}|}{\|\mathbf{u}\times\mathbf{v}\|\|\mathbf{w}\|} = |(\mathbf{u}\times\mathbf{v})\cdot\mathbf{w}| = \left|\det\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix}\right|.$$

2.7. Planos no espaço

Um plano no espaço é definido por um ponto a ele pertencente e a um vetor ortogonal ao plano. Seja Σ um plano, $\hat{\mathbf{x}} \in \Sigma$ e $\mathbf{n}$ vetor perpendicular a Σ. Então

$$\Sigma = \{\mathbf{x} \in \mathbb{R}^3 : (\mathbf{x} - \hat{\mathbf{x}}) \cdot \mathbf{n} = 0\}.$$

Expandindo nas coordenadas, temos que para $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2, \hat{x}_3)$, e $\mathbf{n} = (n_1, n_2, n_3)$, que um ponto qualquer de Σ satisfaz $(x_1 - \hat{x}_1)n_1 + (x_2 - \hat{x}_2)n_2 + (x_3 - \hat{x}_3)n_3 = 0$. Reescrevemos esta equação como $n_1 x + n_2 x + n_3 x = \hat{x}_1 n_1 + \hat{x}_2 n_2 + \hat{x}_3 n_3 = \hat{\mathbf{x}} \cdot \mathbf{n}$, que é da forma geral

$$ax_1 + bx_2 + cx_3 = d,$$

com $a = n_1$, $b = n_2$, $c = n_3$, $d = \hat{\mathbf{x}} \cdot \mathbf{n}$.

EXEMPLO 2.10. Ache a menor distância do ponto $\hat{\mathbf{p}} = (1, 0, 1)$ ao plano dado por $x + 2y - z = 2$.

Solução.

Passo i: precisamos primeiro achar algum ponto $\mathbf{p}_0$ pertencente ao plano. Por exemplo $(1, 1, 1)$.

Passo ii: Seja $\mathbf{v} = (1, 1, 1) - (1, 0, 1) = (0, 1, 0)$. Então a projeção de $\mathbf{v}$ em $\mathbf{n} = (1, 2, -1)$ é dada por

$$\mathbf{w} = \frac{\mathbf{n}\cdot\mathbf{v}}{\|\mathbf{n}\|^2}\mathbf{n} = \frac{1}{3}\mathbf{n}.$$

Então a distância de $\hat{\mathbf{p}}$ ao plano é dada simplesmente pela norma de $\mathbf{w}$, o seja, a distância é de $\|\mathbf{w}\| = \|\mathbf{n}\|/3 = \sqrt{6}/3$.

Passo iii: para achar o ponto $\mathbf{p}_M$ do plano que tem distância mínima até $\hat{\mathbf{p}}$, basta notar que $\mathbf{p}_M + \mathbf{w} = \hat{\mathbf{p}}$, e portanto

$$\mathbf{p}_M = \hat{\mathbf{p}} - \mathbf{w}.$$

EXEMPLO 2.11. Ache a menor distância do ponto $\hat{\mathbf{p}} = (x, y, z)$ ao plano dado por $ax + by + cz + d = 0$.

Solução.

Passo i: considere um ponto $\mathbf{p}_0 = (x_0, y_0, z_0)$ pertencente ao plano.

Passo ii: Seja $\mathbf{v} = (x - x_0, y - y_0, z - z_0)$. Então a projeção de $\mathbf{v}$ em $\mathbf{n} = (a, b, c)$ é dada por

$$\mathbf{w} = \frac{\mathbf{n} \cdot \mathbf{v}}{\|\mathbf{n}\|^2}\mathbf{n}.$$

Então a distância de $\hat{\mathbf{p}}$ ao plano é dada simplesmente pela norma de $\mathbf{w}$, o seja, a distância é de

$$\|\mathbf{w}\| = \frac{|a(x - x_0) + b(y - y_0) + c(z - z_0)|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|ax + by + cz + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

já que $d = -(ax_0 + by_0 + cz_0)$.

Portanto, a distância mínima entre o ponto (x_0, y_0, z_0) e o plano definido por $ax + by + cz + d = 0$ é dada pela fórmula acima.

Outra forma de se definir um plano é, dados três pontos não colineares a ele pertencentes, definir um vetor normal ao plano via produto vetorial. De fato, se $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ pertencem a um plano, então $\mathbf{n} = (\mathbf{y} - \mathbf{x}) \times (\mathbf{z} - \mathbf{x})$ é perpendicular a este mesmo plano.

Exemplo 2.12. Dadas duas retas $r_1 = \{\mathbf{p}^1 + t\mathbf{d}^1 : t \in \mathbb{R}\}$ e $r_2 = \{\mathbf{p}^2 + t\mathbf{d}^2 : t \in \mathbb{R}\}$, onde $\mathbf{d}^1$ e $\mathbf{d}^2$ não são colineares, ache pontos $\mathbf{x}^1 \in r_1$ e $\mathbf{x}^2 \in r_2$ que têm distância mínima.

Solução. Antes de mais nada, observe que como $\mathbf{d}^1$ e $\mathbf{d}^2$ não são colineares, o problema tem solução única. Note também que os pontos de distância minima são tais que $\mathbf{x}^1 - \mathbf{x}^2$ são ortogonais às retas r_1 e r_2. Portanto, escrevendo $\mathbf{x}^1 = \mathbf{p}^1 + t\mathbf{d}^1$ e $\mathbf{x}^2 = \mathbf{p}^2 + s\mathbf{d}^2$, temos que achar t e s tais que

$$[(\mathbf{p}^1 + t\mathbf{d}^1) - (\mathbf{p}^2 + t\mathbf{d}^2)] \cdot \mathbf{d}^1 = 0, \qquad [(\mathbf{p}^1 + s\mathbf{d}^1) - (\mathbf{p}^2 + s\mathbf{d}^2)] \cdot \mathbf{d}^2 = 0.$$

Como temos duas equações e duas incógnitas, podemos resolver o sistema acima. A distância mínima é dada por $\|\mathbf{x}^1 - \mathbf{x}^2\|$.

2.8. Desigualdade lineares

Às vezes precisamos otimizar uma certa função definida no $\mathbb{R}^n$ em domínios que satisfazem alguma restrição, por exemplo que as coordenadas sejam todas não negativas, i.e. $x_i \geq 0$ para $i = 1, \ldots, n$. Estes tipos de restrição são dadas por *desigualdade lineares.* Por simplicidade, ficaremos apenas no caso do plano, quando $n = 2$, mas o caso geral é análogo.

Em geral as desigualdades lineares são dadas na forma $ax_1 + bx_2 + c \leq 0$, onde a, b, c são números reais (para evitar trivialidades, suporemos sempre que a ou b são não nulos). Estas desigualdades determinam a região do plano $\{(x, y) \in \mathbb{R}^2 : ax + by + c \leq 0\}$. Tendo várias desigualdades, podemos considerar a interseção entre os domínios por elas determinados, a chamada *região admissível.* Esta interseção pode ser nula, não limitado, ou limitada. Neste último caso, a região será dada por um polígono.

Exemplo 2.13. Ache os pontos de $\mathbb{R}^2$ tais que

$$\begin{aligned} 3x + 4y - 5 &\leq 0, \\ y &\leq 1, \\ x &\geq 0, \\ y &\geq 0. \end{aligned}$$

É um problema típico tentar agora minimizar uma função linear nalguma região como a dada no exemplo 2.13.

EXEMPLO 2.14. Ache o mínimo de $p(x,y) = 2x + 3 - 5$ na região determinada no exemplo 2.13.

Solução. Note que as curvas de nível da função p são dadas por retas no plano. Neste caso, para achar os pontos de máximo e mínimo de p, basta procurar entre os vértices. Este é apenas um exemplo do caso geral, como enunciado no resultado a seguir.

TEOREMA 2.8.1. *Se uma região admissível D definida por desigualdades lineares é limitada, então máximos e mínimos de $p(x,y) = ax + by + c$ em D ocorrem nos vértices de D.*

2.9. Exercícios

EXERCÍCIO 2.1. Ache as coordenadas de um vetor (u_1, u_2) qualquer na base exemplo 2.1. Mostre que são unicamente determinados por u_1 e u_2.

EXERCÍCIO 2.2. No $\mathbb{R}^n$, seja $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Mostre que os vetores de $\mathcal{B}$ são *linearmente independentes* se e somente se as coordenadas de todo vetor do $\mathbb{R}^n$ são unicamente determinadas.

EXERCÍCIO 2.3. Mostre que $\big|\|\mathbf{x}\| - \|\mathbf{y}\|\big| \le \|\mathbf{x} - \mathbf{y}\|$ (isto vale para qualquer norma) e que $\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 - 2\mathbf{x} \cdot \mathbf{y} + \|\mathbf{y}\|^2$ (isto vale para qualquer norma que venha de produto interno) para todo $\mathbf{x}$, $\mathbf{y}$ do $\mathbb{R}^n$.

EXERCÍCIO 2.4. Considere uma norma vinda de produto interno. Prove o Teorema de Pitágoras.

EXERCÍCIO 2.5. Mostre que a norma euclidiana, e a norma do exemplo 3.13 são de fato normas, segundo a definição 3.6.2.

EXERCÍCIO 2.6. Mostre que existe uma constante $c \in \mathbb{R}$ tal que

$$\|\mathbf{x}\|_\infty \le \|\mathbf{x}\| \le c\|\mathbf{x}\|_\infty$$

para todo $\mathbf{x} \in \mathbb{R}^2$. Mostre que o mesmo vale para vetores do $\mathbb{R}^n$. Como é que esta constante depende de n?

EXERCÍCIO 2.7. Mostre que o ângulo θ entre a diagonal de um cubo e as suas arestas é tal que $\cos\theta = 1/\sqrt{3}$.

EXERCÍCIO 2.8. Verifique que o trângulo com vértices em $(0, -1,)$, $(-1, 1)$, $(2, 0)$ é retângulo.

EXERCÍCIO 2.9. Mostre a lei dos cossenos, que diz que um triângulo com lados de tamanho a, b e c, e com os lados de tamanho a e b determinando um ângulo θ, obedecem à relação:

$$c^2 = a^2 + b^2 - 2ab\cos\theta$$

EXERCÍCIO 2.10. Seja $\mathbf{y}$ vetor não nulo. Mostre que se $\mathbf{z}$ é a projeção de $\mathbf{x}$ em $\mathbf{y}$, i.e., $\mathbf{z} = \alpha\mathbf{y}$ e $(\mathbf{x} - \mathbf{z}) \cdot \mathbf{y} = 0$, então $\alpha = \mathbf{x} \cdot \mathbf{y}/\|\mathbf{y}\|^2$ e $\|\mathbf{z}\| = \|\mathbf{x}\|\cos\theta$.

Exercício 2.11. Mostre que a área do paralelogramo determinado pelos vetores $\mathbf{x}$ e $\mathbf{y}$ é dada por $\|\mathbf{x} \times \mathbf{y}\|$.

Exercício 2.12. Usando a notação do Exemplo 2.12, ache a distância entre as duas retas, com $\mathbf{p}^1 = (0, 0, 0)$, $\mathbf{d}^1 = (1, 0, 1)$, $\mathbf{p}^2 = (-1, 1, 2)$ e $\mathbf{d}^2 = (-1, -1, 0)$.

CAPíTULO 3

Álgebra Linear

[1]

Neste capítulo trataremos resumidamente de várias noções de álgebra linear, como operações com matrizes, matriz inversa, transposta e adjunta, resolução de sistemas lineares, determinantes, regra de Cramer, espaços vetoriais e subespaços, base e dimensão, produto interno, ortogonalidade, projeções, transformações lineares, núcleo e imagem, matriz de uma transformação linear. Autovalores e autovetores, polinômios característicos, operadores diagonalizáveis, operadores auto-adjuntos, operadores ortogonais, e formas bilineares.

3.1. Operações com matrizes

Denotaremos por $\mathbb{R}^{m\times n}$ o espaço das matrizes reais com m linhas e n colunas. Se $A \in \mathbb{R}^{m\times n}$, Denotaremos por $A_{i,j}$ o elemento da i-ésima linha e j-ésima coluna de A. A soma e multiplicação de matrizes é definida da forma usual, isto é, se A, $B \in \mathbb{R}^{m\times n}$, então $C = A + B \in \mathbb{R}^{m\times n}$ é dada por $C_{i,j} = A_{i,j} + B_{i,j}$. A multiplicação para matrizes $D \in \mathbb{R}^{m\times n}$, $E \in \mathbb{R}^{n\times o}$, é definida tal que $C = AB \in \mathbb{R}^{m\times o}$ é dada por $C_{i,j} = \sum_{k=1}^{n} A_{i,k}B_{k,j}$.

Chamaremos de matriz identidade, e denotaremos por I, à matriz tal que $I_{i,i} = 1$ e $I_{i,j} = 0$ se $i \neq j$, para $i, j = 1, \ldots, n$.

3.2. Matriz inversa, transposta e adjunta

Dada $A \in \mathbb{R}^{n\times n}$, se existir $B \in \mathbb{R}^{n\times n}$ tal que $AB = I$ e $BA = I$, então dizemos que A é invertível e que B é a inversa de A. Escrevemos ainda $B = A^{-1}$.

Uma forma de se computar a matriz inversa de A, quando esta existir, é via *matriz dos cofatores*. Seja $\hat{A}^{i,j} \in \mathbb{R}^{n-1\times n-1}$ obtida de A "retirando" de A sua i-ésima linha e j-ésima coluna. Por exemplo, dada

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

temos que

$$\hat{A}^{1,1} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}, \quad \hat{A}^{1,2} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad \hat{A}^{1,3} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}, \quad \hat{A}^{2,1} = \begin{bmatrix} 2 & 3 \\ 8 & 9 \end{bmatrix}, \quad \hat{A}^{2,2} = \begin{bmatrix} 1 & 3 \\ 7 & 9 \end{bmatrix},$$
$$\hat{A}^{2,3} = \begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}, \quad \hat{A}^{3,1} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}, \quad \hat{A}^{3,2} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}, \quad \hat{A}^{3,3} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}.$$

[1]Última Atualização: 18/06/2012

Definimos $\Delta \in \mathbb{R}^{n\times n}$ como sendo a *matriz cofator* de A, onde $\Delta_{i,j} = (-1)^{i+j}\det \hat{A}^{i,j}$. No caso do exemplo acima, temos

$$\Delta = \begin{bmatrix} \det \hat{A}^{1,1} & -\det \hat{A}^{1,2} & \det \hat{A}^{1,3} \\ -\det \hat{A}^{2,1} & \det \hat{A}^{2,2} & -\det \hat{A}^{2,3} \\ \det \hat{A}^{3,1} & -\det \hat{A}^{3,2} & \det \hat{A}^{3,3} \end{bmatrix} = \begin{bmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{bmatrix}.$$

Observação. A matriz Δ^T é também chamada de *adjunta*. É um péssimo nome, que provavelmente deriva de uma tradução infeliz do inglês *adjugate*. O termo matriz adjunta é utilizado mais comumente como sendo simplesmente a transposta de uma matriz (no caso real). Entretanto, na ANPEC, pode aparecer o termo "matrix adjunta" para denominar Δ^T.

Após o cômputo de Δ temos que se A for invertível, então

$$A^{-1} = \frac{\Delta^T}{\det A}.$$

No exemplo acima temos que $\det A = 0$, e portanto a matriz não é invertível. Na verdade temos o importante resultado que afirma que A é invertível se e somente se seu determinante é não nulo.

Note que para conferir se uma matriz é ou não inversa de outra, basta executar a multiplicação matricial e checar se resulta na matriz identidade. Por exemplo, se A e B são invertíveis, então $(AB)^{-1} = B^{-1}A^{-1}$ pois

$$(AB)(B^{-1}A^{-1}) = ABB^{-1}A^{-1} = AA^{-1} = I, \quad (B^{-1}A^{-1})(AB) = BAA^{-1}B^{-1} = BB^{-1} = I.$$

Voltando ao caso geral, dada $A \in \mathbb{R}^{m\times n}$ definimos a matriz transposta de A denotada por A' (ou A^T), onde $A'_{i,j} = A_{j,i}$. Neste caso, as linhas se tornam colunas, e as colunas se tornam linhas. Note que se $A \in \mathbb{R}^{m\times n}$ e $B \in \mathbb{R}^{n\times o}$, e além disto, $C = AB$, então $C' = B'A'$ pois

$$C'_{i,j} = C_{j,i} = \sum_{k=1}^{n} A_{j,k}B_{k,i} = \sum_{k=1}^{n} B'_{i,k}A'_{k,j}.$$

No caso mais geral, considere dois espaços vetoriais V e W, que tenham produtos internos $\langle\cdot,\cdot\rangle_V$ e $\langle\cdot,\cdot\rangle_W$, e seja $T : V \to W$ operador linear. Definimos então a transposta de T como sendo $T' : W \to V$ tal que

$$\langle \mathbf{v}, T'\mathbf{w}\rangle_V = \langle T\mathbf{v}, \mathbf{w}\rangle_W \quad \text{para todo } \mathbf{v} \in V, \mathbf{w} \in W. \tag{3.2.1}$$

Na verdade, na definição acima estamos considerando *Espaços de Hilbert*, mas isto é outra conversa. No caso $V = W = \mathbb{R}^n$ com o produto interno usual, se tomarmos $\mathbf{v} = \mathbf{e}_i$ e $\mathbf{w} = \mathbf{e}_j$ em (3.2.1), temos $[T]'_{i,j} = [T]_{j,i}$ (onde $[T]$ é a representação matricial de T na base canônica).

3.3. Resolução de sistemas lineares

Seja $A \in \mathbb{R}^{m\times n}$ e $\vec{\mathbf{b}} \in \mathbb{R}^m$. Queremos descobrir se existe, e neste caso, quem é, $\vec{\mathbf{x}} \in \mathbb{R}^n$ tal que $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ (chamado de *sistema linear*). Este problema pode não ter solução ($0x = 1$), ter solução única ($2x = 1$), ou ter infinitas soluções ($x + y = 1$). Note entretanto que se A for invertível, então o sistema tem solução única dada por $\vec{\mathbf{x}} = A^{-1}\vec{\mathbf{b}}$.

Em cálculos manuais, a melhor forma de se descobrir se um sistema tem solução é reduzindo-o a uma forma triangular superior, usando a *matriz ampliada*, como nos mostra o exemplo abaixo [**4**, pag.33].

Seja

$$\begin{cases} x_1 + 4x_2 + 3x_3 = 1, \\ 2x_1 + 5x_2 + 4x_3 = 4, \\ x_1 - 3x_2 - 2x_3 = 5. \end{cases}$$

Obtemos então a matriz ampliada, que reduzimos a uma forma triangular superior:

$$\begin{bmatrix} 1 & 4 & 3 & 1 \\ 2 & 5 & 4 & 4 \\ 1 & -3 & -2 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 4 & 3 & 1 \\ 0 & -3 & -2 & 2 \\ 0 & -7 & -5 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 4 & 3 & 1 \\ 0 & -3 & -2 & 2 \\ 0 & 0 & -1/3 & -2/3 \end{bmatrix}$$

Voltando a forma de equações, temos da última linha que $x_3 = 2$. Usando a segunda linha obtemos $x_2 = -2$. Finalmente, da primeira linha temos $x_1 = 3$.

3.4. Determinantes e a regra de Cramer

O determinante é uma função $\mathbb{R}^{n\times n} \to \mathbb{R}$ tal que se uma matriz $A \in \mathbb{R}^{n\times n}$ é dada por $A = [\vec{\mathbf{v}}_1 \, \dots \, \vec{\mathbf{v}}_n]$, então $\det(\cdot)$ é a (única) forma "multilinear alternada" definida em $\vec{\mathbf{v}}_1 \times \cdots \times \vec{\mathbf{v}}_n \mapsto \det(A)$ e tal que $\det(I) = 1$. Por multilinear quer-se dizer que é uma função linear em cada uma das colunas de A. Por alternada quer-se dizer que trocando-se duas linhas de lugar, o determinante é multiplicado por -1, ver [**17**]. Denotaremos o determinante de A por $\det(A)$ ou $|A|$.

Uma outra forma de se definir determinantes é usando-se *permutações*. Seja $I_n = \{1, \dots, n\}$ e $\sigma : I_n \to I_n$ uma bijeção tal que $\sigma(1, \dots, n) = (\sigma_1, \dots, \sigma_n) \in I_n$. Considere S_n o conjunto de todas as permutações de I_n, e denote por $\operatorname{sgn}(\sigma)$ o *sinal* ou *assinatura* de $\sigma \in S_n$, i.e., $\operatorname{sgn}(\sigma) = 1$ se é necessário um número par de inversões para se obter $(\sigma_1, \dots, \sigma_n)$ de $(1, \dots, n)$. Analogamente $\operatorname{sgn}(\sigma) = -1$ se é necessário um número ímpar de inversões. Então

$$\det(A) = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma_1} \dots a_{n,\sigma_n}.$$

Como exemplos, note que se $n = 2$, há duas permutaçoes possíveis:

$$\sigma(1, 2) = (1, 2), \quad \sigma(1, 2) = (2, 1).$$

Portanto o determinante de uma matriz A que seja 2×2 é dado por $\det(A) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$. Já uma matriz 3×3 tem como permutações

$$\sigma(1, 2, 3) = (1, 2, 3), \quad \sigma(1, 2, 3) = (2, 1, 3), \quad \sigma(1, 2, 3) = (1, 3, 2), \quad \sigma(1, 2, 3) = (3, 2, 1),$$
$$\sigma(1, 2, 3) = (3, 1, 2), \quad \sigma(1, 2, 3) = (2, 3, 1).$$

Finalmente, note que são sempre $n!$ permutações possíveis, no caso de matrizes $n \times n$.

Algumas propriedades fundamentais de determinantes são dadas abaixo, supondo-se que $A = [\vec{\mathbf{v}}_1 \, \dots \, \vec{\mathbf{v}}_n]$:

(1) Se existir alguma linha ou coluna zero, então o determinante se anula.
(2) $\det A = \det A^T$, portanto propriedades que valem para linhas, valem para colunas.
(3) $|\vec{\mathbf{v}}_1 \dots \alpha\vec{\mathbf{v}}_j \, \dots \, \vec{\mathbf{v}}_n| = \alpha|\vec{\mathbf{v}}_1 \, \dots \, \vec{\mathbf{v}}_n|$.

(4) $|\vec{\mathbf{v}}_1 \dots \vec{\mathbf{v}}_i \dots \vec{\mathbf{v}}_j \dots \vec{\mathbf{v}}_n| = -|\vec{\mathbf{v}}_1 \dots \vec{\mathbf{v}}_j \dots \vec{\mathbf{v}}_i \dots \vec{\mathbf{v}}_n|$.
(5) $|\vec{\mathbf{v}}_1 \dots \vec{\mathbf{v}}_i + \vec{\mathbf{w}} \dots \vec{\mathbf{v}}_n| = |\vec{\mathbf{v}}_1 \dots \vec{\mathbf{v}}_i \dots \vec{\mathbf{v}}_n| + |\vec{\mathbf{v}}_1 \dots \vec{\mathbf{w}} \dots \vec{\mathbf{v}}_n|$
(6) $\det(AB) = \det(A)\det(B)$

É importante notar que $\det(A+B) \neq \det(A) + \det(B)$. O contraexemplo mais simples é dado por

$$0 = \det \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \det\left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \right) \neq \det \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \det \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = 2.$$

Uma interessante propriedade de determinantes é dada pelo desenvolvimento de Laplace (ver a notação para $\hat{A}_{i,j}$ na página 21):

$$\begin{aligned} (3.4.1) \qquad \det A &= a_{1,1} \det \hat{A}_{1,1} - a_{1,2} \det \hat{A}_{1,2} + \cdots \pm a_{1,N} \det \hat{A}_{N,1} \\ &= a_{1,1}\Delta_{1,1} + a_{1,2}\Delta_{1,2} + \cdots + a_{1,N}\Delta_{N,1} = a_{i,1}\Delta_{i,1} + a_{i,2}\Delta_{i,2} + \cdots + a_{i,N}\Delta_{N,i}. \end{aligned}$$

Considere agora o sistema linear $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$, onde $A \in \mathbb{R}^{n\times n}$ é invertível. Então

$$\vec{\mathbf{x}} = A^{-1}\vec{\mathbf{b}} = \frac{1}{\det A} \begin{bmatrix} \Delta_{1,1} & \dots & \Delta_{n,1} \\ \vdots & \dots & \vdots \\ \Delta_{1,n} & \dots & \Delta_{n,n} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Usando (3.4.1) temos que

$$x_1 = \frac{1}{\det A}(b_1\Delta_{1,1} + \cdots + b_n\Delta_{n,1}) = \frac{1}{\det A} \begin{vmatrix} b_1 & a_{1,2} & \cdots & a_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ b_n & a_{n,2} & \cdots & a_{n,n} \end{vmatrix}$$

Analogamente,

$$x_i = \frac{1}{\det A} \begin{vmatrix} a_{1,1} & \cdots & b_1 & \cdots & a_{1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & \cdots & b_n & \cdots & a_{n,n} \end{vmatrix}.$$

para $i = 1, \dots, n$. Esta identidade é conhecida como *Regra de Cramer*.

Exemplo 3.1 (Tirado de [**4**]). Seja o sistema linear dado por $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$, onde

$$A = \begin{bmatrix} 2 & -3 & 7 \\ 1 & 0 & 3 \\ 0 & 2 & -1 \end{bmatrix}, \quad \vec{\mathbf{b}} = \begin{bmatrix} 1 \\ 5 \\ 0 \end{bmatrix}.$$

Como $\det A = -1$, então, pela Regra de Cramer,

$$x_1 = \frac{1}{\det A} \begin{vmatrix} 1 & -3 & 7 \\ 5 & 0 & 3 \\ 0 & 2 & -1 \end{vmatrix} = -49, \quad x_2 = \begin{vmatrix} 2 & 1 & 7 \\ 1 & 5 & 3 \\ 0 & 0 & -1 \end{vmatrix} = 9, \quad x_3 = \begin{vmatrix} 2 & -3 & 1 \\ 1 & 0 & 5 \\ 0 & 2 & 0 \end{vmatrix} = 18.$$

3.5. Espaços vetoriais, subespaços, base e dimensão

O exemplo mais comum e intuitivo de espaço vetorial é o $\mathbb{R}^n$, ver Definição 2.0.1. Entretanto, uma definição mais geral é de grande utilidade. A menos que explicitamente mencionado, neste texto nos restringiremos a espaços vetoriais sobre o corpo dos reais.

DEFINIÇÃO 3.5.1. *Um espaço vetorial V sobre os reais é um conjunto cujos elementos chamamos de vetores, com duas operações binárias, soma vetorial e multiplicação por escalar tais que*

(1) $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$*, para todo* $\mathbf{x},\mathbf{y}\in V$
(2) $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{y}+(\mathbf{x}+\mathbf{z})$*, para todo* $\mathbf{x},\mathbf{y},\mathbf{z}\in V$
(3) Existe um elemento $\mathbf{0}\in V$ *tal que* $\mathbf{0}+\mathbf{x}=\mathbf{x}$*, para todo* $\mathbf{x}\in V$
(4) Para todo $\mathbf{x}\in V$*, existe um elemento* $\mathbf{y}\in V$ *tal que* $\mathbf{y}+\mathbf{x}=$
(5) $1\mathbf{x}=\mathbf{x}$*, para todo* $\mathbf{x}\in V$
(6) $(\alpha+\beta)\mathbf{x}=\alpha\mathbf{x}+\beta\mathbf{x}$*, para todo* α*,* $\beta\in\mathbb{R}$ *e para todo* $\mathbf{x}\in V$
(7) $\alpha(\beta\mathbf{x})=(\alpha\beta)\mathbf{x}$*, para todo* α*,* $\beta\in\mathbb{R}$ *e para todo* $\mathbf{x}\in V$
(8) $\alpha(\mathbf{x}+\mathbf{y})=\alpha\mathbf{x}+\alpha\mathbf{y}$*, para todo* $\alpha\in\mathbb{R}$ *e para todo* $\mathbf{x},\mathbf{y}\in V$

Alguns resultados podem ser obtidos imediatamente:

LEMA 3.5.2. Seja V um espaço vetorial sobre os reais. Então temos que

(1) O vetor zero é único
(2) Todo elemento de $\mathbf{x}\in V$ tem um único negativo dado por $(-1)\mathbf{x}$
(3) $0\mathbf{x}=\mathbf{0}$ para todo $\mathbf{x}\in V$
(4) $\alpha\mathbf{0}=\mathbf{0}$ para todo $\alpha\in\mathbb{R}$

DEMONSTRAÇÃO. Demonstraremos apenas a primeira afirmativa. As demais ficam como exercícios. Para demonstrar (1), suponha que $\mathbf{0}_1$ e $\mathbf{0}_2$ sejam dois zeros de V. Logo

$$\mathbf{0}_1=\mathbf{0}_2+\mathbf{0}_1=\mathbf{0}_1+\mathbf{0}_2=\mathbf{0}_2,$$

onde usamos que a hipótese de que $\mathbf{0}_1$ é zero e a propriedade (3) da Definição 3.5.1, seguida da propriedade (1). Na última igualdade usamos a hipótese de que $\mathbf{0}_1$ é zero e novamente a propriedade (3) da Definição de 3.5.1. $\square$

EXEMPLO 3.2. O espaço das matrizes $m\times n$ reais denotado por $\mathbb{R}^{m\times n}$ é um espaço vetorial com a definição usual de soma de matrizes e multiplicação por escalar.

EXEMPLO 3.3. O espaço F das funções de $\mathbb{R}$ em $\mathbb{R}$, com as operações

$$(u+v)(x)=u(x)+v(x) \text{ para todo } x\in\mathbb{R} \text{ e todas } u,v\in F,$$
$$(\alpha u)(x)=\alpha u(x) \text{ para todo } x\in\mathbb{R}, \text{ toda } u\in F \text{ e todo } \alpha\in\mathbb{R},$$

é espaço vetorial, ver Exercício 3.1.

3.5.1. Subespaço vetorial. Seja V um espaço vetorial e $W\subseteq V$. Então dizemos que W é *subespaço vetorial de* V se W for também um espaço vetorial. Para que isto aconteça, basta que

(1) $W\neq\emptyset$
(2) se $\mathbf{u}$, $\mathbf{v}\in W$, então $\mathbf{u}+\mathbf{v}\in W$
(3) se $\alpha\in\mathbb{R}$ e $\mathbf{u}\in W$, então $\alpha\mathbf{u}\in W$

Note se W é subespaço de V, então o vetor nulo $\mathbf{0} \in W$ pois como W é não vazio, então existe algum $\mathbf{u} \in W$. Mas então $\mathbf{0} = 0\mathbf{u} \in W$, por causa de (3).

**** por figura mostrando subespaço vetorial e não subespaço vetorial *****

Exemplo 3.4. Note que $\{(0, y, z) : y, z \in \mathbb{R}\}$ é subespaço vetorial do $\mathbb{R}^3$, mas que $\{(0, y, z) \in \mathbb{R}^3 : y \geq 0,\ z \geq 0\}$ não o é.

Exemplo 3.5. O espaço das matrizes diagonais $n \times n$ é subespaço do espaço vetorial das matrizes $n \times n$.

Exemplo 3.6. O espaço dos polinômios quadráticos é subespaço vetorial do espaço das funções.

Exemplo 3.7. Se V é espaço vetorial e $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$, então

$$\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \stackrel{\text{def}}{=} \{\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k : \alpha_1, \dots \alpha_k \in \mathbb{R}\}$$

é subespaço vetorial de V. Chamamos o termo $\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k$ de *combinação linear de* $\mathbf{v}_1, \dots, \mathbf{v}_k$.

Os próximos resultados respondem à pergunta natural: interseções e uniões de subespaços são ainda subespaços?

Lema 3.5.3. Sejam W_1 e W_2 subespaços vetoriais de um espaço vetorial V. Então $W_1 \cap W_2$ é subespaço vetorial de V.

Demonstração. Como W_1 e W_2 são ambos subespaços, então $\mathbf{0} \in W_1$ e $\mathbf{0} \in W_2$. Logo $\mathbf{0} \in W_1 \cap W_2$ e então $W_1 \cap W_2 \neq \emptyset$.

Além disto, se $\mathbf{u}$, $\mathbf{v} \in W_1 \cap W_2$, então $\mathbf{u}$, $\mathbf{v} \in W_1$ e $\mathbf{u}$, $\mathbf{v} \in W_2$. Logo, pela propriedade (2) de subespaços vetoriais, $\mathbf{u} + \mathbf{v} \in W_1$ e $\mathbf{u} + \mathbf{v} \in W_2$. Mas então $\mathbf{u} + \mathbf{v} \in W_1 \cap W_2$.

Da mesma forma, se $\alpha \in \mathbb{R}$ e $\mathbf{u} \in W_1 \cap W_2$, então, pela propriedade (3) de subespaços vetoriais, $\alpha\mathbf{u} \in W_1$ e $\alpha\mathbf{u} \in W_2$. Logo $\alpha\mathbf{u} \in W_1 \cap W_2$. ☐

Como podemos ver no exemplo a seguir, a união de subespaços vetoriais *não é* subespaço vetorial.

Exemplo 3.8. Sejam $A_1 = \{(x, 0) : x \in \mathbb{R}\}$ e $A_2 = \{(0, x) : x \in \mathbb{R}\}$ subespaços de $\mathbb{R}^2$. Seja $A = A_1 \cup A_2 = \{(x, y) : x = 0 \text{ ou } y = 0\}$. Logo $(1, 0) \in A$ e $(0, 1) \in A$, mas $(1, 0) + (0, 1) = (1, 1) \notin A$.

Apesar da união não ser necessariamente subespaços, há uma forma de se "juntar" subespaços vetoriais e obter outro subespaço, como vemos no resultado a seguir.

Lema 3.5.4. Sejam W_1 e W_2 subespaços vetoriais de um espaço vetorial V. Seja o conjunto

$$W_1 + W_2 \stackrel{\text{def}}{=} \{\mathbf{w}_1 + \mathbf{w}_2 : \mathbf{w}_1 \in W_1 \text{ e } \mathbf{w}_2 \in W_2\}.$$

Então $W_1 + W_2$ é subespaço vetorial de V.

DEMONSTRAÇÃO. Note que $\mathbf{0} \in W_1 + W_2$, logo $W_1 + W_2 \neq \emptyset$. Sejam agora $\mathbf{u}$, $\mathbf{v} \in W_1 + W_2$. Logo existem $\mathbf{u}_1 \in W_1$ e $\mathbf{u}_2 \in W_2$ tais que $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$, pela definição de $W_1 + W_2$. Da mesma forma, existem $\mathbf{v}_1 \in W_1$ e $\mathbf{v}_2 \in W_2$ tais que $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$. Mas então

$$\mathbf{u} + \mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{v}_1 + \mathbf{v}_2 = (\mathbf{u}_1 + \mathbf{v}_1) + (\mathbf{u}_2 + \mathbf{v}_2)$$

é a soma de um vetor de W_1, dado por $\mathbf{u}_1 + \mathbf{v}_1$, com outro de W_2, dado por $\mathbf{u}_2 + \mathbf{v}_2$. Logo $\mathbf{u} + \mathbf{v} \in W_1 + W_2$.

Analogamente, se $\alpha \in \mathbb{R}$ e $\mathbf{u} \in W_1 + W_2$, então existem $\mathbf{u}_1 \in W_1$ e $\mathbf{u}_2 \in W_2$ tais que $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$. Mas então

$$\alpha\mathbf{u} = \alpha(\mathbf{u}_1 + \mathbf{u}_2) = \alpha\mathbf{u}_1 + \alpha\mathbf{u}_2$$

é a soma de um vetor de W_1 com outro de W_2. Logo $\alpha\mathbf{u} \in W_1 + W_2$. □

OBSERVAÇÃO. Algumas observações quanto a soma de espaços. A primeira é que W_1+W_2 é apenas uma notação, afinal soma de conjuntos não é uma operação que faça sentido em geral. A segunda observação diz respeito ao importante caso $W_1 \cap W_2 = \emptyset$. Neste caso dizemos que a soma é *direta* e a representamos por $W_1 \oplus W_2$. Note que podemos estender a noção de soma direta para mais que dois espaços, como em $W_1 \oplus W_2 \oplus \cdots \oplus W_N$. Note que $\mathbb{R}^n = \mathbb{R} \oplus \mathbb{R} \oplus \cdots \oplus \mathbb{R}$ é a soma direta do $\mathbb{R}$ repetida n vezes.

3.5.2. Base e dimensão. Sejam $\mathbf{v}_1, \ldots, \mathbf{v}_k$ vetores de um espaço vetorial V. Se

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = 0 \implies \alpha_1 = \cdots = \alpha_k = 0$$

então dizemos que $\mathbf{v}_1, \ldots, \mathbf{v}_k$ são *linearmente independentes*, ou *L.I.*. Vetores que não são L.I. são chamados de L.D., ou *linearmente dependentes*. Outra forma de dizer que $\mathbf{v}_1, \ldots, \mathbf{v}_k$ são L.D. é quando existirem escalares $\alpha_1, \ldots, \alpha_k$, nem todos nulos e tais que

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_k\mathbf{v}_k = 0.$$

Com o conceito de independência linear, podemos definir o que é uma base de um espaço vetorial. Dado um espaço V, dizemos que $\mathbf{v}_1, \ldots, \mathbf{v}_n \in V$ é base de V se

(1) $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = V$
(2) $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ é L.I.

OBSERVAÇÃO. Trataremos aqui sempre de espaços vetoriais de dimensão finita, e isto quer dizer que existe uma base finita para os espaços.

A seguir enunciamos alguns resultados sobre bases.

TEOREMA 3.5.5. *Se* $V = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$*, então é sempre possível extrair uma base de* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

TEOREMA 3.5.6. *Se* $V = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *então o conjunto* $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ *é L.D. sempre que* $m > n$.

COROLÁRIO 3.5.7. Qualquer base de V tem sempre o mesmo número de elementos.

O corolário acima tem grande importância pois nos diz que existe um número inerente a V, que não depende da escolha da base. A este número chamamos de *dimensão de* V, ou $\dim V$.

TEOREMA 3.5.8. *Qualquer conjunto L.I. de vetores pode ser completado a fim de formar uma base.*

COROLÁRIO 3.5.9. Se $\dim V = n$, qualquer conjunto L.I. com n vetores é base.

TEOREMA 3.5.10. *Sejam W_1 e W_2 subespaços de V. Então*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Pelo resultado acima, se a soma é direta, então $\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2$.

TEOREMA 3.5.11. *Seja $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ base de V. Então todo vetor de V pode ser escrito de forma única como combinação linear de $\mathbf{v}_1, \dots, \mathbf{v}_n$.*

OBSERVAÇÃO. **Caros alunos: rever mudança de bases!**

3.6. Produto interno, ortogonalidade e projeções

Duas importantes *ferramentas* matemáticas quando se trabalha em espaços vetoriais são produtos internos e normas.

DEFINIÇÃO 3.6.1. *Seja V espaço vetorial sobre os reais. Um* produto interno *é uma função de $V \times V \to \mathbb{R}$, denotado por $\mathbf{x}, \mathbf{y} \mapsto \mathbf{x} \cdot \mathbf{y}$ e tal que*

(1) $\mathbf{x} \cdot \mathbf{x} > 0$ para todo $\mathbf{x} \in V$ com $\mathbf{x} \neq \mathbf{0}$
(2) $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ para todo $\mathbf{x}$, $\mathbf{y} \in V$
(3) $(\alpha\mathbf{x}) \cdot \mathbf{y} = \alpha(\mathbf{x} \cdot \mathbf{y})$ para todo $\alpha \in \mathbb{R}$ e todo $\mathbf{x}$, $\mathbf{y} \in V$
(4) $(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$ para todo $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in V$

Outra notação usual para produtos internos é $\langle \cdot, \cdot \rangle$.

Note que da definição acima concluímos imediatamente que para todo $\mathbf{x} \in V$,

$$\mathbf{0} \cdot \mathbf{x} = (0\mathbf{0}) \cdot \mathbf{x} = 0(\mathbf{0} \cdot \mathbf{x}) = 0.$$

EXEMPLO 3.9. Em $\mathbb{R}^2$, se $\mathbf{x} = (x_1, x_2)$, e $\mathbf{y} = (y_1, y_2)$, o produto interno canônico é dado por

$$\mathbf{x} \cdot \mathbf{y} = \vec{\mathbf{x}}^T \vec{\mathbf{y}} = x_1 y_1 + x_2 y_2.$$

Em $\mathbb{R}^n$, para $\mathbf{x} = (x_1, \dots, x_n)$, e $\mathbf{y} = (y_1, \dots, y_n)$, definimos

$$\mathbf{x} \cdot \mathbf{y} = \vec{\mathbf{x}}^T \vec{\mathbf{y}} = x_1 y_1 + \cdots + x_n y_n.$$

EXEMPLO 3.10. Em $\mathbb{R}^2$, a operação

$$(x_1, x_2) \cdot (y_1, y_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 2x_1y_1 - x_1y_2 - x_2y_1 + 4x_2y_2$$

define um produto interno. De fato, a primeira propriedade (positividade) é verdadeira pois

$$(x_1, x_2) \cdot (x_1, x_2) = 2x_1^2 - 2x_1x_2 + 4x_2^2 = 2[(x_1 - x_2/2)^2 + 7x_2^2/4] > 0,$$

se $(x_1, x_2) \neq (0, 0)$. As outras propriedades do produto interno são mais fáceis de serem checadas.

EXEMPLO 3.11. Considere o espaço vetorial das funções contínuas em $[0, 1]$, com as operações de multiplicação por escalar e soma como no Exemplo 3.3. Então a operação dada pela integral de Riemann

$$f \cdot g = \int_0^1 f(x)g(x)\,dx$$

define um produto interno deste espaço.

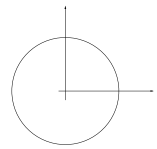

FIGURA 1. Conjunto $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\| = 1\}$.

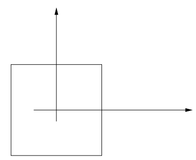

FIGURA 2. Conjunto $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_\infty = 1\}$

Introduzimos agora a noção de *norma*. Num espaço vetorial, uma boa forma de se medir distâncias entre vetores é através de normas. Em particular, o conceito normas ajuda na definição canônica de conjuntos abertos e fechados, como veremos a seguir.

DEFINIÇÃO 3.6.2. *Dado um espaço vetorial V, uma* norma *é uma função de V em $\mathbb{R}$, denotada por $\mathbf{x} \mapsto \|\mathbf{x}\|$, e tal que*

(1) $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$ *para todo* $\mathbf{x}$, $\mathbf{y} \in V$ *(desigualdade triangular)*
(2) $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$ *para todo* $\mathbf{x} \in V$, *e para todo* $\alpha \in \mathbb{R}$
(3) $\|\mathbf{x}\| > 0$ *para todo* $\mathbf{x} \in V$ *tal que* $\mathbf{x} \neq \mathbf{0}$

Quando um espaço vetorial V tem uma norma associada, dizemos que é um *espaço normado*.

EXEMPLO 3.12. Em $\mathbb{R}^2$,

$$\|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2}$$

define uma norma. Na Figura 1 temos que o conjunto de pontos $\mathbf{x}$ tais que $\|\mathbf{x}\| = 1$ é dado por um círculo. No caso mais geral, em $\mathbb{R}^n$,

$$\|(x_1, \dots, x_n)\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

também define uma norma.

EXEMPLO 3.13. Outra norma em $\mathbb{R}^n$ é dada por

$$\|(x_1, \dots, x_n)\|_\infty = \max_{1 \le j \le n} |x_j|.$$

Na Figura 2 vemos que o conjunto de pontos $\mathbf{x}$ tais que $\|\mathbf{x}\|_\infty = 1$ é dado por um quadrado. Compare com a Figura 1.

O resultado abaixo é importante pois mostra que *todo* produto interno induz uma norma.

TEOREMA 3.6.3. *Seja V um espaço vetorial com produto interno. Então*

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

define uma norma em V. Além disto, vale a desigualdade de Cauchy-Schwartz

$$|\mathbf{x} \cdot \mathbf{y}| \le \|\mathbf{x}\|\|\mathbf{y}\| \qquad \text{para todo } \mathbf{x}, \mathbf{y} \in V. \tag{3.6.1}$$

DEMONSTRAÇÃO. Como o produto interno garante que sempre teremos $\mathbf{x}\cdot\mathbf{x} \geq 0$, então a operação acima está bem definida. Mostraremos primeiro (3.6.1). Seja $\mathbf{z} = \mathbf{x} - (\mathbf{x}\cdot\mathbf{y})\mathbf{y}/\|\mathbf{y}\|^2$. Então

$$\mathbf{z}\cdot\mathbf{y} = \mathbf{x}\cdot\mathbf{y} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}\cdot\mathbf{y} = 0,$$

e

$$0 \leq \|\mathbf{z}\|^2 = \mathbf{z}\cdot\mathbf{z} = \mathbf{z}\cdot\mathbf{x} = \mathbf{x}\cdot\mathbf{x} - \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{x}\cdot\mathbf{y}.$$

Logo

$$(\mathbf{x}\cdot\mathbf{y})^2 \leq \|\mathbf{x}\|^2\|\mathbf{y}\|^2,$$

e (3.6.1) vale.

Para mostrar a propriedade (1) da definição de norma, note que

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

e assim temos (1). As propriedade (2) e (3) seguem-se imediatamente da definição e das propriedades do produto interno. □

OBSERVAÇÃO. Note pela demonstração acima que a igualdade $|\mathbf{x}\cdot\mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$ vale se e somente se $\mathbf{x} = \alpha\mathbf{y}$ para algum $\alpha \in \mathbb{R}$. Ver exercício 3.3.

Bem como no caso do $\mathbb{R}^n$, ver (2.2.4), podemos definir cossenos de "ângulos entre dois vetores" não nulos $\mathbf{x}$ e $\mathbf{y} \in V$ por

$$\cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, \tag{3.6.2}$$

que toma valores entre -1 e 1 devido à desigualdade de Cauchy-Schwartz (3.6.1). Dizemos também que $\mathbf{x}$, $\mathbf{y}$ são ortogonais, ou perpendiculares, se $\mathbf{x}\cdot\mathbf{y} = 0$.

Outra generalização interessante é dada por projeções. Dados $\mathbf{u}$ e $\mathbf{v} \in V$ não nulos, chamamos $\mathbf{w}$ de *projeção ortogonal* (ou simplesmente de *projeção*) de $\mathbf{u}$ em $\mathbf{v}$ se

(1) $\mathbf{w} = \alpha\mathbf{v}$ para algum $\alpha \in \mathbb{R}$
(2) $(\mathbf{u}-\mathbf{w})\cdot\mathbf{v} = 0$

Note que para $\mathbf{w}$ ficar bem definido, basta calcular α. Mas note que de (1) e (2), temos que $(\mathbf{u}-\alpha\mathbf{v})\cdot\mathbf{v} = 0$, e portanto $\alpha\mathbf{v}\cdot\mathbf{v} = \mathbf{u}\cdot\mathbf{v}$. Como $\mathbf{v}$ é não nulo,

$$\alpha = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{v}\|^2} = \frac{\|\mathbf{u}\|\cos\theta}{\|\mathbf{v}\|}, \qquad \mathbf{w} = \frac{\|\mathbf{u}\|\cos\theta}{\|\mathbf{v}\|}\mathbf{v}.$$

onde $\cos\theta$ é como em (3.6.2). A norma de $\mathbf{w}$ é dada por $\|\mathbf{w}\| = \|\mathbf{u}\|\cos\theta$.

3.7. Transformações lineares, núcleo, imagem e representações matriciais

Dados dois espaços vetoriais V_1 e V_2, dizemos que uma função $T : V_1 \to V_2$ é uma *função, transformação* ou *aplicação linear* se

$$T(\mathbf{x}+\alpha\mathbf{y}) = T(\mathbf{x}) + \alpha T(\mathbf{y}) \qquad \text{para todo } \mathbf{x},\, \mathbf{y} \in V_1 \text{ e todo } \alpha \in \mathbb{R}.$$

Note que em particular, para toda aplicação linear linear temos $T(\mathbf{0}) = \mathbf{0}$, pois

$$T(\mathbf{0}) = T(0\mathbf{0}) = 0T(\mathbf{0}) = \mathbf{0}.$$

Seja $\mathcal{L}(V, W)$ o espaço das aplicações lineares $T : V \to W$ para as quais existe $M \in \mathbb{R}$ tal que

$$\|T\mathbf{x}\|_W \leq M\|\mathbf{x}\|_V,$$

Neste caso dizemos que T é limitada. Se V e W forem de dimensão finita, então toda transformação linear é limitada, ver exercício 3.5. É possível definir operações canônicas de multiplicação por escalar e soma em $\mathcal{L}(V, W)$ de tal forma que este seja um espaço vetorial, ver exercício 3.2.

O exemplo principal de transformação linear em espaços de dimensões finitas é dado por multiplicação de matrizes. De fato, seja $T : \mathbb{R}^m \to \mathbb{R}^n$ definida por $T(\vec{\mathbf{u}}) = A\vec{\mathbf{u}}$, onde $A \in \mathbb{R}^{n\times m}$. Então T é linear pois $T(\vec{\mathbf{u}} + \alpha\vec{\mathbf{v}}) = A(\vec{\mathbf{u}} + \alpha\vec{\mathbf{v}}) = A\vec{\mathbf{u}} + \alpha A\vec{\mathbf{v}} = T(\vec{\mathbf{u}}) + \alpha T(\vec{\mathbf{v}})$.

Observe que para definir uma aplicação linear qualquer $T : \mathbb{R}^m \to \mathbb{R}^n$, basta defini-la numa base $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ do $\mathbb{R}^m$, i.e., basta conhecer $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$. De fato, se $\mathbf{x} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n$, então

$$T(\mathbf{x}) = T(\alpha_1\mathbf{v}_1 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_n T(\mathbf{v}_n).$$

Num certo sentido, todas as transformações lineares em espaçoes de dimensões finitas são dadas por matrizes. De forma mais precisa, seja $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ base de V e $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ base de W. Então, se $\mathbf{x} \in V$ é dado por $\mathbf{x} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_m\mathbf{v}_m$, então

$$T(\mathbf{x}) = \alpha_1 T(\mathbf{v}_1) + \cdots + \alpha_m T(\mathbf{v}_m).$$

Seja $A_{j,i}$ a j-ésima coordenada de $T(\mathbf{v}_i)$ na base $\{\mathbf{w}_1, \dots, \mathbf{w}_n\}$, i.e.,

$$T(\mathbf{v}_i) = A_{1,i}\mathbf{w}_1 + \cdots + A_{n,i}\mathbf{w}_n. \tag{3.7.1}$$

Logo

$$\begin{aligned} T(\mathbf{x}) &= \alpha_1(A_{1,1}\mathbf{w}_1 + \cdots + A_{n,1}\mathbf{w}_n) + \cdots + \alpha_m(A_{1,m}\mathbf{w}_1 + \cdots + A_{n,m}\mathbf{w}_n) \\ &= (\alpha_1 A_{1,1} + \cdots + \alpha_m A_{1,m})\mathbf{w}_1 + \cdots + (\alpha_1 A_{n,1} + \cdots + \alpha_m A_{n,m})\mathbf{w}_n = \beta_1\mathbf{w}_1 + \cdots + \beta_n\mathbf{w}_n, \end{aligned}$$

onde

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} A_{1,1} & \cdots & A_{1,m} \\ \vdots & \dots & \vdots \\ A_{n,1} & \cdots & A_{n,m} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}. \tag{3.7.2}$$

Se $\mathbf{w} = T(\mathbf{u})$, então a matriz $A \in \mathbb{R}^{n\times m}$ com coeficientes $A_{i,j}$ mapeia as coordenadas de $\mathbf{u}$ nas coordenadas de $\mathbf{v}$. Note que a matriz A depende fortemente das bases de V e W. Dizemos que A é a representação de ou matriz associada a T nas bases $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ e $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$. Por vezes, esta representação é também escrita como $[T]_{\mathcal{W}}^{\mathcal{V}}$.

Um exemplo importante é quando as bases são canônicas. Neste caso, basta ver de (3.7.1) que $A_{j,i}$ é dada pela j-ésima coordenada de $T(\mathbf{e}_i)$.

Dois importantes conjuntos relacionados a uma aplicação linear $T : V \to W$ são seu núcleo e sua imagem, dados por

$$N(T) = \{\mathbf{v} \in V : T(\mathbf{v}) = 0\} \subseteq V, \qquad \operatorname{Im}(T) = \{T(\mathbf{v}) : \mathbf{v} \in V\} \subseteq W.$$

O núcleo recebe também a notação $\ker(T)$, do inglês *kernel*, e a imagem de V por T também recebe a notação $T(V)$ ou $R(T)$ (do inglês *range*). É importante notar que tanto o núcleo como a imagem de uma transformação linear são espaços vetoriais. Para tal, basta checar que estes são subespaços de V e W respectivamente. Ver exercício 3.6.

Se $\operatorname{Im}(V) = W$ dizemos que T é sobrejetora. Se

$$T(\mathbf{u}) = T(\mathbf{v}) \implies \mathbf{u} = \mathbf{v},$$

então dizemos que T é injetora, ou $1-1$. Temos também o seguinte resultado.

TEOREMA 3.7.1. *$N(T) = \mathbf{0}$ se e somente se T é injetiva.*

DEMONSTRAÇÃO. ($\Longrightarrow$) Suponha que $N(T) = \mathbf{0}$. Sejam $\mathbf{u}$, $\mathbf{v} \in V$ tais que $T(\mathbf{u}) = T(\mathbf{v})$. Então $T(\mathbf{u} - \mathbf{v}) = \mathbf{0}$ e portanto $\mathbf{u} - \mathbf{v} = \mathbf{0}$. Logo T é injetiva.

($\Longleftarrow$) Suponha T injetora e $T(\mathbf{u}) = \mathbf{0}$. Então $T(\mathbf{u}) = \mathbf{0} = T(\mathbf{0})$. Como T é injetora, então $\mathbf{u} = \mathbf{0}$. □

Temos a seguir um importante resultado.

TEOREMA 3.7.2 (Teorema do núcleo e da imagem). *Seja $T : V \to W$ aplicação linear, onde V e W são espaços vetoriais. Então* $\dim N(T) + \dim \operatorname{Im}(V) = \dim V$.

COROLÁRIO 3.7.3. Considere as hipóteses do Teorema acima. Temos então que

(i) Se $\dim V = \dim W$, então T é injetora se e somente se T é sobrejetora.
(ii) Se T for injetora, então T leva vetores LI em vetores LI. E se além disto $\dim W = \dim V$, então T leva base em base.

Note que há relação entre as dimensões do núcleo e imagem de uma transformação linear T e o posto e nulidade da matriz que representa esta transformação (em quaisquer bases):

$$\dim \operatorname{Im}(T) = \operatorname{posto}([T]_{\mathcal{W}}^{\mathcal{V}}), \qquad \dim N(T) = \text{nulidade de } ([T]_{\mathcal{W}}^{\mathcal{V}}).$$

Note que pelo Teorema 3.7.2 que

$$\dim N(T) = \text{nulidade de } ([T]_{\mathcal{W}}^{\mathcal{V}}) = \text{número de colunas de } [T]_{\mathcal{W}}^{\mathcal{V}} - \operatorname{posto}([T]_{\mathcal{W}}^{\mathcal{V}}).$$

3.8. Autovalores, polinômios característicos e operadores diagonalizáveis

Nesta seção falaremos sobre autovalores, autovetores, e suas propriedades. Por absoluta falta de espaço/tempo, não faremos contas, mas a forte recomendação para quem tem dificuldades ou não se lembra direito como se calcula autovalores, e autovetores é que olhe, por exemplo, o livro [**4**].

3.8.1. Autovalores, autovetores e polinômios característicos. Seja $T : V \to V$ transformação linear. Dizemos que $\lambda \in \mathbb{C}$ é um *autovalor* de T se existe vetor não nulo, chamado de *autovetor*, $\mathbf{v} \in V$ tal que $T\mathbf{v} = \lambda\mathbf{v}$. Chamamos ainda $(\lambda, \mathbf{v})$ de *autopar*.

Seja λ autovalor de T. Então o conjunto

$$E_\lambda = \{\mathbf{v} \in V : T\mathbf{v} = \lambda\mathbf{v}\}$$

é um subespaço vetorial de V, chamado *autoespaço* de λ. Ver Exercício 3.7.

Suponha agora um operador linear dado por uma matriz $A \in \mathbb{R}^{n\times n}$, definindo a aplicação linear $\vec{\mathbf{x}} \to A\vec{\mathbf{x}}$. Para achar autovalores e autovalores de A, basta achar soluções não triviais, i.e., não nulas, para o sistema $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$. Isto só será possível de $\det(A - \lambda I) = 0$. Note que $\det(A - \lambda I)$ é um polinômio em termos de λ, ao qual damos o nome de *polinômio característico*, e denotamos por $P(\lambda)$. O problema de achar autovalores resume-se então ao problema de encontrar as raízes de $P(\lambda)$. Isto é *sempre possível*, segundo o teorema fundamental da Álgebra, desde que admita-se autovalores complexos.

Depois de encontrado um autovalor λ, pode-se encontrar os autovetores correspondentes resolvendo-se $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$. Note que este sistema *sempre* tem soluções não triviais, já que $A - \lambda I$ não é invertível, ver o exercício 3.8.

Vamos agora, nos exemplos abaixo, ver o conceito de multiplicidade *algébrica* e *geométrica*.

EXEMPLO 3.14. Seja $A = 2I$. Então, se λ for autovalor, temos que $\det(2I - \lambda I) = 0$, i.e., $(\lambda - 2)^2 = 0$. Logo $\lambda = 2$ é o único autovalor. Como autovetores temos que $(2I - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$, ou seja, $0\vec{\mathbf{x}} = \vec{\mathbf{0}}$. Portanto todo $\vec{\mathbf{x}} \in \mathbb{R}^2$ é autovetor, e neste caso, o autoespaço tem dimensão dois. Dizemos que λ tem multiplicidade algébrica dois, e multiplicidade geométrica dois.

EXEMPLO 3.15. Seja agora a matriz

$$B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}.$$

O polinômio característico é dado por $P(\lambda) = (2 - \lambda)^2$, o mesmo do exemplo 3.14, e portanto $\lambda = 2$ é o único autovalor. Entretanto ao calcular os autovetores vemos que se $(B - 2I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$, então

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Logo $x_2 = 0$, e os autovetores são múltiplos de $[1, 0]^T$. Dizemos então que λ tem multiplicidade algébrica dois, e multiplicidade geométrica um.

3.8.2. Operadores diagonalizáveis. Seja $T : V \to V$ operador linear e V espaço vetorial de dimensão finita. Uma característica interessante de autovalores é que, quando estes formam uma base de V, a matriz que representa T é diagonal. De fato, observe em (3.7.1) que se $\mathbf{v}_i$ é autovetor, então

$$T(\mathbf{v}_i) = \lambda \mathbf{v}_i,$$

onde tomamos $\mathbf{w}_j = \mathbf{v}_j$ para todo j. Conclua então que a matriz A em (3.7.2) é diagonal.

Como é bastante conveniente representar um operador por uma matriz diagonal, é natural perguntar se, dado um operador linear, ele é *diagonalizável*, i.e., se existe uma base tal que sua representação nesta base é uma matriz diagonal. De forma mais simples, dizemos que um operador $T : V \to V$ é *diagonalizável* se existe uma base de V formada por autovetores. Isto não será sempre possível, como pode ser visto no exemplo 3.15.

O resultado abaixo é importante para garantir tal base. Ele garante que autovetores correspondentes a autovetores distintos são LI.

TEOREMA 3.8.1. *Sejam $\lambda_1, \dots, \lambda_k$ autovalores distintos de T. Então os correspondentes autovetores $\mathbf{v}_1, \dots, \mathbf{v}_k$ são LI.*

DEMONSTRAÇÃO. Suponha que $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_k \mathbf{v}_k = 0$. Para mostrar que estes vetores são LI, temos que mostrar que $\alpha_k = \cdots = \alpha_1 = 0$. Aplicando $T - \lambda_1 I$ à combinação linear acima, obtemos

$$\alpha_1(\lambda_1 - \lambda_1)\mathbf{v}_1 + \alpha_2(\lambda_2 - \lambda_1)\mathbf{v}_2 + \cdots + \alpha_k(\lambda_k - \lambda_1)\mathbf{v}_k = 0,$$

e portanto $\alpha_2(\lambda_2-\lambda_1)\mathbf{v}_2+\cdots+\alpha_k(\lambda_k-\lambda_1)\mathbf{v}_k=0$. Aplicamos agora $T-\lambda_2 I$ e obtemos

$$\alpha_2(\lambda_2-\lambda_1)(\lambda_2-\lambda_2)\mathbf{v}_2+\alpha_3(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)\mathbf{v}_3+\cdots+\alpha_k(\lambda_k-\lambda_1)(\lambda_k-\lambda_2)\mathbf{v}_k=0,$$

e obtemos que $\alpha_3(\lambda_3-\lambda_1)(\lambda_3-\lambda_2)\mathbf{v}_3+\cdots+\alpha_k(\lambda_k-\lambda_1)(\lambda_k-\lambda_2)\mathbf{v}_k=0$. Procedendo desta forma, aplicando $T-\lambda_3 I,\dots,T-\lambda_{k-1}I$ temos que

$$\alpha_k(\lambda_k-\lambda_1)(\lambda_k-\lambda_2)\dots(\lambda_k-\lambda_{k-1})\mathbf{v}_k=0.$$

Como os autovalores $\lambda_1,\dots,\lambda_k$ são distintos entre si, obtemos que $\alpha_k=0$. Voltando os passos anteriores, é possível ver que $\alpha_{k-1}=\cdots=\alpha_1=0$. □

Corolário 3.8.2. Se uma transformação linear em espaços de dimensão n tiver n autovalores distintos, então existe uma base formada por autovetores.

Outra forma de se definir matrizes diagonalizáveis, é exigir que sejam similares a uma matriz diagonal. Dizemos que duas matrizes A e B são similares se existe uma matriz P invertível tal que $B=P^{-1}AP$. O resultado abaixo trata de propriedades de matrizes similares.

Teorema 3.8.3. *Matrizes similares têm o mesmo determinante, mesmo traço, mesmo polinômio característico, e mesmos autovalores.*

Outro resultado interessante é que similaridade forma uma relação de equivalência, ver exercício 3.10.

Note que se a matriz A é diagonalizável e $D=P^{-1}AP$ é diagonal, então as colunas de P são exatamente os autovetores de A. Para ver isto, suponha que

$$D=\begin{bmatrix}\lambda_1 & 0 & 0 & \cdots & 0\\ 0 & \lambda_2 & 0 & \cdots & 0\\ \vdots & \vdots & & & \vdots\\ 0 & 0 & \cdots & 0 & \lambda_n\end{bmatrix}=P^{-1}AP,$$

onde $P=[\vec{\mathbf{v}}_1\dots\vec{\mathbf{v}}_n]$. Então $PD=AP$ e $\lambda_i\vec{\mathbf{v}}_i=A\vec{\mathbf{v}}_i$.

3.9. Operadores auto-adjuntos, operadores ortogonais

Na seção anterior, pouco pudemos dizer a respeito de que matrizes são diagonalizáveis ou não. Em casos especiais é possível conseguir resultados mais interessantes.

Dizemos que uma transformação linear T é *auto-adjunta* ou *simétrica* se $T'=T$. Uma definição equivalente em espaços com produto interno V é dizer que $\langle T\mathbf{u},\mathbf{v}\rangle=\langle\mathbf{u},T\mathbf{v}\rangle$ para todo $\mathbf{u}$, $\mathbf{v}\in V$.

Caso $T^{-1}=T'$, dizemos que T é ortogonal. A mesma definição e terminologia é empregada no caso de matrizes.

Um exemplo de matriz ortogonal e simétrica é a identidade, e de ortogonal e não simétrica é

$$\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix},$$

se $\theta\neq k\pi$.

Note que se A é ortogonal, então

$$(\det A)^2=\det A\det A'=\det(AA')=\det I=1,$$

e portanto $\det A = \pm 1$.

Segue-se imediatamente da definição de matrizes ortogonais, que os seus vetores colunas e vetores linhas são ortonormais (ou seja, são ortogonais entre si e todos têm norma um).

Outro resultado importante para operadores auto-adjuntos vem abaixo.

TEOREMA 3.9.1. *Seja $T : V \to V$ auto-adjunto, e λ_1, $\mathbf{v}_1$ e λ_2, $\mathbf{v}_2$ autovalores e autovetores de T. Se $\lambda_1 \neq \lambda_2$ então $\mathbf{v}_1$ e $\mathbf{v}_2$ são ortogonais: $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.*

DEMONSTRAÇÃO. Note que

$$\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = (T\mathbf{v}_1) \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot (T\mathbf{v}_2) = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2.$$

Logo $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Como $\lambda_1 \neq \lambda_2$ então $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. □

Uma importante propriedade de operadores auto-adjuntos é dada pelo resultado a seguir.

TEOREMA 3.9.2. *Seja $T : V \to V$ auto-adjunto. Então existe uma base ortonormal de V formada por autovalores de T.*

Note que o resultado acima *não* diz que toda base de autovalores de operadores auto-adjuntos é ortonormal, apenas que *existe* uma base ortonormal. De fato, por exemplo, para a matriz identidade n-dimensional, todo vetor de $\mathbb{R}^n$ é autovetor. Entretanto nem toda base do é $\mathbb{R}^n$ ortonormal.

Finalmente, o resultado abaixo serve para caracterizar quais matrizes são ortogonais.

TEOREMA 3.9.3. *Seja $T : V \to V$. Então as afirmativas abaixo são equivalentes.*
(1) T é ortogonal
(2) T leva bases ortogonais em bases ortogonais
(3) T preserva produtos internos, i.e., $\langle T\mathbf{u}, T\mathbf{v}\rangle = \langle \mathbf{u}, \mathbf{v}\rangle$, para todo $\mathbf{u}$, $\mathbf{v} \in V$.
(4) T preserva normas, i.e., $\|T\mathbf{u}\| = \|\mathbf{u}\|$ para todo $\mathbf{u} \in V$.

3.10. Formas lineares e bilineares

Uma forma linear definida um espaço vetorial V é simplesmente um operador linear $F : V \to \mathbb{R}$. Note que se $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ é base de V, então

$$F(\mathbf{v}) = F(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n) = \alpha_1 F(\mathbf{v}_1) + \cdots + \alpha_n F(\mathbf{v}_n) = \begin{bmatrix} F(\mathbf{v}_1) & \ldots & F(\mathbf{v}_n) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix},$$

ou seja, dada uma base, a forma F num vetor $\mathbf{v}$ pode ser caracterizada simplesmente como o produto interno do vetor formado pelos valores de F nos vetores da base com as coordenadas de $\mathbf{v}$ na base.

Uma forma bilinear B é um operador definido em $V \times V$ e tomando valores em $\mathbb{R}$, e que seja linear em cada um dos argumentos, i.e.,

$$V : V \times V \to \mathbb{R}$$
$$(\mathbf{x}, \mathbf{y}) \mapsto \mathbb{R},$$

onde $B(\mathbf{x} + \alpha\mathbf{y}, \mathbf{z}) = B(\mathbf{x}, \mathbf{z}) + \alpha B(\mathbf{y}, \mathbf{z})$ e $B(\mathbf{x}, \mathbf{y} + \alpha\mathbf{z}) = B(\mathbf{x}, \mathbf{y}) + \alpha B(\mathbf{x}, \mathbf{z})$, para todo $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in V$ e $\alpha \in \mathbb{R}$. Se $B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x})$ para todo $\mathbf{x}$, $\mathbf{y} \in V$, dizemos que a forma B é simétrica.

Dada uma matriz $A \in \mathbb{R}^{n\times n}$, o exemplo usual de forma bilinear é defnido por $(\vec{\mathbf{x}}, \vec{\mathbf{y}}) \to \vec{\mathbf{x}} \cdot A\vec{\mathbf{y}}$. Se A for simétrica, esta forma assim definida será simétrica.

Finalmente, assim como no caso de formas bilineares, se $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ é base de V, então

$$B(\mathbf{u}, \mathbf{v}) = \begin{bmatrix} u_1 & \dots & u_n \end{bmatrix} \begin{bmatrix} B(\mathbf{v}_1, \mathbf{v}_1) & \cdots & B(\mathbf{v}_1, \mathbf{v}_n) \\ \vdots & \dots & \vdots \\ B(\mathbf{v}_n, \mathbf{v}_1) & \cdots & B(\mathbf{v}_n, \mathbf{v}_n) \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix},$$

onde $\mathbf{u} = u_1\mathbf{v}_1 + \cdots + u_n\mathbf{v}_n$ e $\mathbf{v} = v_1\mathbf{v}_1 + \cdots + v_n\mathbf{v}_n$.

3.11. Exercícios

EXERCÍCIO 3.1. Mostre que o espaço das funções, definido no Exemplo 3.3 é de fato um espaço vetorial.

EXERCÍCIO 3.2. Defina operações de multiplicação por escalar e soma em $\mathcal{L}(V_1, V_2)$, tais que este seja um espaço vetorial com estas operações.

EXERCÍCIO 3.3. Dado um espaço vetorial com produto interno $\mathbf{x} \cdot \mathbf{y}$ e norma $\|\mathbf{x}\| = (\mathbf{x} \cdot \mathbf{x})^{1/2}$, mostre que $|\mathbf{x} \cdot \mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$ se e somente se $\mathbf{x} = \alpha\mathbf{y}$ para algum $\alpha \in \mathbb{R}$.

EXERCÍCIO 3.4. Seja V espaço vetorial normado e de dimensão finita. Seja $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ base de V, e $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1, \dots, \alpha_n\end{bmatrix}^T$ o vetor formado pelas coordenadas de $\mathbf{v} \in V$ nesta base. Mostre que existem constantes c_0, c_1, que dependem da base mas não de $\mathbf{v}$ tais que

$$c_0\|\boldsymbol{\alpha}\|_{\mathbb{R}^n} \le \|\mathbf{v}\|_V \le c_1\|\boldsymbol{\alpha}\|_{\mathbb{R}^n}.$$

Acima, $\|\cdot\|_{\mathbb{R}^n}$ denota a norma canônica do $\mathbb{R}^n$.

EXERCÍCIO 3.5. Mostre que se V e W forem espaços vetoriais de dimensão finita, então toda transformação linear $T : V \to W$ é limitada.

EXERCÍCIO 3.6. Sejam V e W espaços vetoriais, e $T : V \to W$ aplicação linear. Mostre que $N(T)$ e $\mathrm{Im}(V)$ são subespaços vetoriais de V e W respectivamente.

EXERCÍCIO 3.7. Seja $T : V \to W$ operador linear e λ autovalor de T. Mostre que $\{\mathbf{v} \in V : T\mathbf{v} = \lambda\mathbf{v}\}$ é um subespaço vetorial de V.

EXERCÍCIO 3.8. Seja $A \in \mathbb{R}^{n\times n}$, e λ autovalor de A. Mostre que o sistema $(A - \lambda I)\vec{\mathbf{x}} = \vec{\mathbf{0}}$ *sempre* tem soluções não triviais.

EXERCÍCIO 3.9. Faça os detalhes da demonstração do teorema 3.8.1 no caso $k = 3$.

EXERCÍCIO 3.10. Mostre que similaridade forma uma relação de equivalência.

CAPíTULO 4

Limites de funções

[1] A fim de discutirmos a noção de *continuidade de funções*, precisamos entender *limites de funções*. Este conceito será importante quando falarmos em derivação. Antes entretanto de definirmos limites de funções, daremos algumas definições básicas relacionadas aos números reais.

4.1. Definições básicas envolvendo funções

Um dos conceitos mais importantes em Matemática é o de funções. Apesar de termos apresentado na página 5 uma definição de função como um caso particular de relação entre conjuntos, a descrição mais usual é dizer que uma função é uma regra que associa elementos enter dois conjuntos de uma forma específica. Para nossos propósitos entretanto, esta "definição" bastará.

Sendo mais específico, considere A e B dois conjuntos, e uma função denominada f entre A e B. Então f é uma regra que associa a cada elemento $x \in A$, um elemento $f(x) \in B$. Chamamos o conjunto A de *domínio* da função f e o denotamos por $D(f)$. Chamamos o conjunto B de *contradomínio* da função f. Escrevemos $f : A \to B$, ou ainda

$$\begin{aligned} f : A &\to B \\ x &\mapsto f(x). \end{aligned}$$

Se $E \subseteq A$, chamamos de *imagem de E* ao conjunto

$$f(E) = \{f(x) : x \in E\}.$$

Similarmente, se $H \subseteq B$, chamamos de *imagem inversa de H* o conjunto

$$f^{-1}(H) = \{x : f(x) \in H\}.$$

dizemos que f é *sobrejetiva* (ou simplesmente, *sobre*) se $f(A) = B$, i.e., se para todo $y \in B$ existe $x \in A$ tal que $f(x) = y$.

Dizemos que f é *injetiva* (ou *biunívoca* ou *um a um* ou 1-1) quando, para a, a' no domínio da f,

$$f(a) = f(a') \implies a = a'.$$

Outra forma de se definir injetividade é quando

$$a \neq a' \implies f(a) \neq f(a').$$

Se f é injetiva e sobre, a chamamos de *bijetiva* ou de uma *bijeção*.

Dizemos que $g : B \to A$ é *função inversa* de f se

$$g(f(x)) = x \quad \text{para todo } x \in A, \qquad f(g(y)) = y \quad \text{para todo } y \in B.$$

[1] Última Atualização: 02/07/2012

Quando esta existir, denotamos a inversa de f por f^{-1}. Note que uma função tem inversa se e somente se é sobrejetiva e injetiva, ver problema 4.1. Portanto, se $f : A \to B$ é injetiva, então restringindo o contradomínio B à imagem de A por f, teremos f sobrejetiva. Em outras palavras, seja $f : A \to f(A)$. Então f é claramente sobre. Se for também injetiva, haverá a inversa $f^{-1} : f(A) \to A$.

Exemplo 4.1. Seja $f : (0, 1) \to \mathbb{R}$ dada por $f(x) = x^2 + 1$. Então f não tem inversa pois, apesar de ser injetiva, não é sobrejetiva. Mas como $f\big((0, 1)\big) = (1, 2)$, então se definirmos

$$f : (0, 1) \to (1, 2)$$
$$x \mapsto x^2,$$

então teremos f sobrejetiva. Logo existe a inversa $f^{-1} : (1, 2) \to (0, 1)$.

Observação. Note que a definição de *imagem inversa* independe de existir ou não a função inversa. Por exemplo, a função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$ não tem inversa. Entretanto $f^{-1}(\mathbb{R}) = \mathbb{R}$.

Existem várias operações com funções, entre elas a composição. Sejam A, B e C conjuntos, e $f : A \to B$ e $g : B \to C$ funções. Então podemos definir uma função $h : A \to C$ dada pela *composição* de f e g, i.e., $h(x) \stackrel{\text{def}}{=} g(f(x))$, onde $x \in A$. Neste caso usamos a notação $h = g \circ f$, e dizemos que h é a composta da f com a g.

Outra operação que pode ser muitas vezes executada é soma, diferença, produto, divisão, de funções. Por exemplo, sejam A conjunto e $f : A \to \mathbb{R}$ e $g : A \to \mathbb{R}$ funções. Então podemos definir a função $h = f + g$ tal que

$$h(x) = f(x) + g(x) \text{ para todo } x \in A.$$

Da mesma forma podemos definir fg por $(fg)(x) = f(x)g(x)$. Outras operações são definidas analogamente.

Temos que tomar cuidado entretanto se a definição faz sentido. Por exemplo, se $f : A \to \mathbb{R}$ se anula em algum ponto de A, então não podemos definir $\phi : A \to \mathbb{R}$ por $\phi(x) = 1/f(x)$. Igualmente, não faz sentido definir $1/g$, se $g : A \to \mathbb{R}^n$, para $n > 1$, pois não podemos dividir por vetores.

Observação. A vizualiação de funções é possível, via gráficos. Por exemplo, o gráfico de uma equação que depende de (x, y) é simplesmente o conjunto de pontos do plano que satisfazem esta equação. Por exemplo, o gráfico de uma função $f : \mathbb{R} \to \mathbb{R}$ é dado pelo conjunto $\big\{(x, y) \in \mathbb{R}^2 : y = f(x)\big\}$. Já o gráfico de $x^2 + y^2 = 3$ é formado por todos os pontos de $\mathbb{R}^2$ que têm norma igual a $\sqrt{3}$.

4.2. Intervalos na reta

Neste capítulo, falaremos sobre intervalos na reta. Falaremos também sobre vizinhanças, cuja noção é baseada na função valor absoluto (que nada mais é que uma norma nos reais).

Intervalos na reta serão conjuntos como os abaixo:

(1) Intervalo aberto: $(a, b) = \{x \in \mathbb{R} : a < x < b\}$
(2) Intervalo fechado: $[a, b] = \{x \in \mathbb{R} : a \le x \le b\}$
(3) $[a, b) = \{x \in \mathbb{R} : a \le x < b\}$

(4) $(a, b] = \{x \in \mathbb{R} : a < x \leq b\}$
(5) $[a, +\infty) = \{x \in \mathbb{R} : a \leq x\}$
(6) $(a, +\infty) = \{x \in \mathbb{R} : a < x\}$
(7) $(-\infty, b] = \{x \in \mathbb{R} : x \leq b\}$
(8) $(-\infty, b) = \{x \in \mathbb{R} : x < b\}$
(9) $(-\infty, +\infty) = \mathbb{R}$

Prosseguimos no sentido de definirmos vizinhanças. Para tal precisamos da noção de distância entre dois pontos x e y da reta, que é dada pelo valor absoluto de $x - y$, i.e., por $|x - y|$. Para um número real a, o valor absoluto (ou módulo) de a é dado por

$$|a| = \begin{cases} a & \text{se } a \geq 0, \\ -a & \text{se } a < 0. \end{cases}$$

Exemplo 4.2. Por definição $|5| = 5$, e $|-5| = -(-5) = 5$.

Lema 4.2.1. Algumas propriedades dos números reais:
(1) $|-a| = |a|$ para todo $a \in \mathbb{R}$.
(2) $|ab| = |a||b|$ para todo a, $b \in \mathbb{R}$.
(3) Dados a, $k \in \mathbb{R}$ temos que $|a| \leq k$ se e somente se $-k \leq a \leq k$.
(4) $-|a| \leq a \leq |a|$ para todo $a \in \mathbb{R}$.

Demonstração. (1) Se $a = 0$, então $|0| = 0 = |-0|$. Se $a > 0$, então $-a < 0$ e logo $|-a| = -(-a) = a = |a|$. Se $a < 0$, então $-a > 0$ e $|-a| = -a = |a|$.
(2) Exercício.
(3) Exercício.
(4) Tome $k = |a|$ no ítem (3) do lema. Então $|a| \leq |a| \implies -|a| \leq a \leq |a|$.

□

Lema 4.2.2 (Desigualdade Triangular). Para todo a, $b \in \mathbb{R}$ temos

$$|a + b| \leq |a| + |b|.$$

Demonstração. Sabemos que $-|a| \leq a \leq |a|$ e $-|b| \leq b \leq |b|$. Logo, $-|a| - |b| \leq a + b \leq |a| + |b|$. Pelo ítem (3) do Lema 4.2.1 temos que $|a + b| \leq |a| + |b|$, como queríamos demonstrar. □

Seja $a \in \mathbb{R}$ e considere o intervalo

$$B_\epsilon(a) = \{x \in \mathbb{R} : |x - a| < \epsilon\} = (a - \epsilon, a + \epsilon).$$

Uma *vizinhança* de a é qualquer conjunto contendo $B_\epsilon(a)$ para algum $\epsilon > 0$.

4.3. Funções inversas

Investigaremos mais de perto agora quando uma função é invertível. Em particular nos concentraremos nas funções monótonos, i.e., funções que são crescentes ou decrescentes, que definiremos a seguir.

Seja $I \subset \mathbb{R}$. Diremos que uma função $f : I \to \mathbb{R}$ é crescente se dados dois pontos x, y de I tais que $x < y$ tem-se que $f(x) < f(y)$. Se $x < y$ implica em $f(x) \leq f(y)$, diremos que f é *não decrescente*.

Analogamente, $g : I \to \mathbb{R}$ é decrescente se $x < y$ implica em $f(x) > f(y)$, e não crescente se $x < y$ implica em $f(x) \geq f(y)$. Se uma função é crescente ou decrescente, a chamamos de *monótona*.

OBSERVAÇÃO. A terminologia acima não é unânime. Alguns autores preferem chamar funções crescentes como definidas acima como estritamente crescentes, funções não-decrescentes como crescentes.

Note que toda função monótona é injetiva. A volta vale também, mas sob a condição da função ser contínua. Isto é, toda função contínua definida num intervalo e injetiva é monótona, ver Exercício 4.3.

TEOREMA 4.3.1. *Seja I um intervalo e $f : I \to \mathbb{R}$ contínua. Então*

1. *se f for monótona, então $f(I)$ é intervalo, e a função inversa $f^{-1} : f(I) \to \mathbb{R}$ é contínua*
2. *se f for injetiva então é monótona.*

4.4. Limites de funções

Seja $I = (a, b)$ um intervalo não vazio e $f : I \to \mathbb{R}$, e seja $c \in [a, b]$. Dizemos que L é o limite de f em c se para todo $\epsilon > 0$ existir $\delta > 0$ tal que

$$x \in (c - \delta, c + \delta) \cap I, \quad x \neq c \implies f(x) \in (L - \epsilon, L + \epsilon).$$

Neste caso, escrevemos $L = \lim_{x \to c} f(x)$, e dizemos que f converge para L no ponto c. Outra forma de escrever a definição acima é dizendo que para todo $\epsilon > 0$, existe $\delta > 0$ tal que

$$x \in I, \quad 0 < |x - c| < \delta \implies |f(x) - L| < \epsilon.$$

Uma observação a respeito da definição acima é que o valor do limite em c independe do valor que f assume em c. Na verdade, f não precisa nem estar definida neste ponto. Somente quando discutirmos continuidade é que o valor em c será importante, mas isto fica para o próximo capítulo.

Antes de começarmos a calcular limites, é interessante também ver que as seguintes definições são equivalentes, e qualquer uma delas pode ser usada no estudo de limites.

LEMA 4.4.1 (Critérios equivalentes para limites). Seja $I = (a, b)$ e $f : I \to \mathbb{R}$, e seja $c \in [a, b]$. Então as afirmativas são equivalentes:

1. $\lim_{x \to c} f(x) = L$.
2. Seja (x_n) sequência em I com $x_n \neq c$ para todo n e $\lim_{n \to \infty} x_n = c$. Então $\big(f(x_n)\big)$ converge e $\lim_{n \to \infty}\big(f(x_n)\big) = L$.

DEMONSTRAÇÃO. (1) $\implies$ (2) Seja $\epsilon > 0$, e (x_n) em $I \backslash \{c\}$ tal que $\lim_{n \to \infty} x_n = c$. Por hipótese existe δ tal que

$$x \in I, \quad 0 < |x - c| < \delta \implies |f(x) - L| < \epsilon. \tag{4.4.1}$$

Seja $N \in \mathbb{N}$ tal que $|x_n - c| < \delta$ se $n \geq N$. Então, por (4.4.1) tem-se $|f(x_n) - L| < \epsilon$ e conclui-se que a sequência $\big(f(x_n)\big)$ converge para L.

(2) $\implies$ (1)(*por contradição*) Assuma que (2) valha, e que (1) seja falso. Logo existe vizinhança $\delta > 0$ tal que para todo $n \in \mathbb{N}$ existe $x_n \in (c - 1/n, c + 1/n) \cap I$, com $x_n \neq c$ e $f(x_n) \notin (L - \delta, L + \delta)$. Isto é uma contradição pois por (2) teríamos que ter $\lim_{n \to \infty}\big(f(x_n)\big) = L$. $\square$

EXEMPLO 4.3. Seja $f(x) = x$. Então $\lim_{x\to c} f(x) = c$. De fato, dado $\epsilon > 0$, tome $\delta = \epsilon$, pois

$$0 < |x - c| < \delta \implies |f(x) - f(c)| = |x - c| < \delta = \epsilon.$$

EXEMPLO 4.4. Seja $f : \mathbb{R} \to \mathbb{R}$ definida por

$$f(x) = \begin{cases} x, & \text{se } x \in \mathbb{Q} \\ 0, & \text{se } x \in \mathbb{R}\backslash\mathbb{Q} \end{cases}$$

Então f tem limite bem definido em $c = 0$, mas não nos demais pontos. Dado $\epsilon > 0$, seja $\delta = \epsilon$. Se $|x| < \delta$, então $|f(x)| = 0 \le |x| < \delta = \epsilon$ caso $x \in \mathbb{R}\backslash\mathbb{Q}$, e $|f(x)| = |x| < \delta = \epsilon$ caso $x \in \mathbb{Q}$. Logo $|x - 0| < \delta$ implica em $|f(x) - f(0)| = |f(x)| < \epsilon$. Portanto f tem limite no zero.

Nos demais pontos tal limite não existe pela densidade dos racionais nos irracionais e vice-versa. De fato, dado $x \in \mathbb{R}$, existe (x_n) sequência em $\mathbb{R}\backslash\mathbb{Q}$ e (y_n) sequência em $\mathbb{Q}$, ambas convergentes para x com $x_n \neq x$ e $y_n \neq x$ para todo $n \in \mathbb{N}$. Mas $\lim_{n\to\infty} f(x_n) = 0$ e $\lim_{n\to\infty} f(y_n) = \lim_{n\to\infty}(y_n) = x \neq 0$. Portanto f não tem limite para $x \neq 0$.

EXEMPLO 4.5. Ache $\lim_{x\to 0} f(x)$ de $f : \mathbb{R}^+ \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} x \sin \frac{1}{x} & \text{se } x \in \mathbb{Q} \cap \mathbb{R}^+, \\ 0 & \text{se } x \in \mathbb{R}^+\backslash\mathbb{Q}. \end{cases}$$

Primeiro note para $x \in \mathbb{R}^+$ que se $x \in \mathbb{Q}$, então $|f(x)| \le |x|$ pois $|\sin 1/x| \le 1$. Se $x \notin \mathbb{Q}$, então $|f(x)| = 0 \le |x|$. Em ambos os casos temos $|f(x)| \le |x|$. Então, dado $\epsilon > 0$, seja $\delta = \epsilon$. Então se $x \in \mathbb{R}^+$ e

$$0 < |x| < \delta = \epsilon \implies |f(x) - 0| \le |x| < \epsilon.$$

Logo f tem limite no zero e o limite é zero, i.e., $\lim_{x\to 0} f(x) = 0$

LEMA 4.4.2 (Unicidade do limite). Seja $I = (a, b)$ e $f : I \to \mathbb{R}$, e seja $c \in [a, b]$. Então f pode ter, no máximo, um limite em c.

DEMONSTRAÇÃO. Sejam L_1 e L_2 dois limites de f em c. Portanto, dado $\epsilon > 0$ existem δ_1 e δ_2 tais que

$$x \in A, \quad 0 < |x - c| < \delta_1 \implies |f(x) - L_1| < \frac{\epsilon}{2},$$
$$x \in A, \quad 0 < |x - c| < \delta_2 \implies |f(x) - L_2| < \frac{\epsilon}{2}.$$

Tome $\delta = \min\{\delta_1, \delta_2\}$. Então tomando-se $0 < |x - c| < \delta$ implica em

$$|L_1 - L_2| \le |L_1 - f(x)| + |f(x) - L_2| < \epsilon.$$

Como ϵ é arbitrário, temos $L_1 = L_2$. □

EXEMPLO 4.6. Seja

$$\text{sgn(x)} = \begin{cases} 1 & \text{se } x > 0, \\ 0 & \text{se } x = 0, \\ -1 & \text{se } x < 0. \end{cases}$$

Tomando-se as sequências $(-1/n)$ e $(1/n)$, ambas convergindo para $c = 0$ mas nunca atingindo este valor, tem-se $\big(f(-1/n)\big) = -1$ e $\big(f(1/n)\big) = 1$. Então esta função não tem limite em $c = 0$, pois se o limite existe, este tem que ser único.

Assim como no caso de sequências, podemos definir operações com funções, como soma, subtração, etc. Se $f : I \to \mathbb{R}$ e $g : I \to \mathbb{R}$, então definimos $(f + g) : I \to \mathbb{R}$ por $(f + g)(x) = f(x) + g(x)$. De forma análoga definimos $(f - g)(x) = f(x) - g(x)$ e $(fg)(x) = f(x)g(x)$. Se g é tal que $g(x) \neq 0$ para todo $x \in I$, definimos também $(f/g)(x) = f(x)/g(x)$. Valem então resultados como o limite da soma é a soma do limite, etc.

Lema 4.4.3. Seja $I = (a, b)$. Sejam $f : I \to \mathbb{R}$ e $g : I \to \mathbb{R}$, e seja $c \in [a, b]$. Suponha que existam os limites $\lim_{x\to c} f(x)$ e $\lim_{x\to c} g(x)$. Então

(1) $\lim_{x\to c}(f + g)(x) = \lim_{x\to c} f(x) + \lim_{x\to c} g(x)$
(2) $\lim_{x\to c}(f - g)(x) = \lim_{x\to c} f(x) - \lim_{x\to c} g(x)$
(3) $\lim_{x\to c}(fg)(x) = \lim_{x\to c} f(x) \lim_{x\to c} g(x)$
(4) $\lim_{x\to c}(f/g)(x) = \lim_{x\to c} f(x)/ \lim_{x\to c} g(x)$, se g for tal que $g(x) \neq 0$ para todo $x \in I$, e $\lim_{x\to c} g(x) \neq 0$.

Os resultados acima podem ser estendidos para um número finito de operações.

Demonstração. Demonstraremos apenas (1). As demais demonstrações são similares.

Seja (x_n) sequência em I com $x_n \neq c$ para todo $n \in \mathbb{N}$ e $\lim_{n\to\infty} x_n = c$. Então $(f + g)(x_n) = f(x_n) + g(x_n)$ converge pois é soma de sequências convergentes e $\lim_{n\to\infty}\big((f + g)(x_n)\big) = \lim_{n\to\infty}\big(f(x_n)\big) + \lim_{n\to\infty}\big(g(x_n)\big)$. ☐

Exemplo 4.7. Seja $n \in \mathbb{N}$. Então $\lim_{x\to c} x^n = (\lim_{x\to c} x)^n = c^n$.

Exemplo 4.8. Se $c > 0$, então $\lim_{x\to c} 1/x = 1/(\lim_{x\to c} x) = 1/c$.

Uma condição extra tem que ser imposta quando lidamos com composição de funções. É natural perguntar, supondo-se que $\lim_{x\to c} g(x) = L$, quando

$$\lim_{x\to c} f(g(x)) = f(\lim_{x\to c} g(x)) = f(L)$$

ocorre. E a resposta é que a igualdade acima é verdadeira se $\lim_{y\to L} f(y) = f(L)$. Em outras palavras, basta que f seja contínua em L.

Uma outra propriedade de funções que têm limite em um ponto é a de limitação local, i.e., a função é limitada numa vizinhança do ponto. Observe que uma função localmente limitada não necessariamente é globalmente limitada, como veremos mais a seguir.

Definição 4.4.4. *Seja $I = (a, b)$ e $f : I \to \mathbb{R}$ e $c \in [a, b]$. Dizemos que f é* limitada numa vizinhança de c *se existem $\delta > 0$ e constante M tais que*

$$x \in (c - \delta, c + \delta) \cap I \implies |f(x)| \leq M.$$

Dizemos também que f é localmente limitada em c.

Para mostrar que se $f : I \to \mathbb{R}$ tem limite em c de I então f é localmente limitada em c, basta primeiro tomar $\epsilon = 1$. Dado $L = \lim_{x\to c} f(x)$, existe $\delta > 0$ tal que

$$x \in I, \quad 0 < |x - c| < \delta \implies |f(x) - L| < 1.$$

Neste caso, temos $|f(x)| \leq |f(x) - L| + |L| < 1 + L$. Se $c \notin I$ defina $M = 1 + |L|$. Se $c \notin I$ defina $M = \max\{|f(c)|, 1 + |L|\}$. Em qualquer dos casos temos que

$$x \in I, \quad |x - c| < \delta \implies |f(x)| < M.$$

Da discussão acima concluímos imediatamente que $f(x) = 1/x$ é localmente limitada em todo ponto $c \neq 0$. Além disso concluímos que f não tem limite em $c = 0$ pois não é limitada localmente em torno deste ponto.

Alguns resultados que valem para sequências podem ser estendidos para limites de funções. Por exemplo, do Lema 4.4.5 tiramos o seguinte resultado. Sua demonstração é um exercício.

Lema 4.4.5 (limite de sanduíches). Sejam $I = (a, b)$ e f, g e h funções de I em $\mathbb{R}$, e seja $c \in [a, b]$. Suponha que para todo $x \in I$ com $x \neq c$ tivermos $f(x) \leq g(x) \leq h(x)$, e que $\lim_{x\to c} f(x) = \lim_{x\to c} h(x) = L$. Então $\lim_{x\to c} g(x) = L$.

Lema 4.4.6. Sejam $I = (a, b)$ e $f : I \to \mathbb{R}$, e seja $c \in [a, b]$. Suponha que para todo $x \in I$ com $x \neq c$ tivermos $a \leq g(x) \leq b$, e que existe o limite de f em c. Então $a \leq \lim_{x\to c} f(x) \leq b$.

4.5. Limites laterais, infinitos e no infinito

Assim como na seção anterior, assumimos que $I = (a, b) \subset \mathbb{R}$ e $f : I \to \mathbb{R}$. Seja agora $c \in [a, b]$. Dizemos que L é limite à direita de f em c se para todo $\epsilon > 0$ existe $\delta > 0$ tal que

$$x \in I, \quad 0 < x - c < \delta \implies |f(x) - L| < \epsilon.$$

Neste caso escrevemos que $\lim_{x\to c^+} f(x) = L$.

Definição similar vale para limite à esquerda (e escrevemos $\lim_{x\to c^-} f(x) = L$).

É possível mostrar que se c ṕonto de acumulação tanto de $I \cap (c, +\infty)$ como de $I \cap (-\infty, c)$, então

$$\lim_{x\to c} f(x) = L \iff \lim_{x\to c^+} f(x) = \lim_{x\to c^-} f(x) = L. \tag{4.5.1}$$

Exemplo 4.9. Seja $f(x) = \mathrm{sgn(x)}$, como no exemplo 4.6. Como $\lim_{x\to 0^+} f(x) = 1$ e $\lim_{x\to 0^-} f(x) = -1$, então não existe limite de f no zero.

Outra definição importante é a de *limite infinito*. Dizemos que f tende a $+\infty$ em c se para todo $\alpha \in \mathbb{R}$ existe $\delta > 0$ tal que

$$x \in I, \quad 0 < |x - c| < \delta \implies f(x) > \alpha.$$

Escrevemos então que $\lim_{x\to c} = +\infty$.

Definição similar vale para f tende a $-\infty$ em c.

Exemplo 4.10. $\lim_{x\to 0} 1/x^2 = +\infty$. De fato, dado $\alpha > 0$, tomando $\delta = 1/\sqrt{\alpha}$ temos

$$0 < |x| < \delta \implies x^2 < \delta^2 = \frac{1}{\alpha} \implies \frac{1}{x^2} > \alpha.$$

Exemplo 4.11. Seja $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$. Então g não tende a $-\infty$ ou a $+\infty$ no zero pois $g(x) < 0$ se $x < 0$ e $g(x) > 0$ se $x > 0$.

Finalmente definimos limites "no infinito". Seja $a \in \mathbb{R}$ e $f : (a, +\infty) \to \mathbb{R}$. Dizemos que $L \in \mathbb{R}$ é limite de f quando $x \to +\infty$ se para todo ϵ existe $k > a$ tal que

$$x > k \implies |f(x) - L| < \epsilon.$$

Analogamente podemos definir limite de f quando $x \to -\infty$.

Exemplo 4.12. $\lim_{x\to-\infty} 1/x = \lim_{x\to+\infty} 1/x = 0$.

Exemplo 4.13. Nem sempre existe limite "no infinito. Tome por exemplo $\sin(x)$.

4.6. Exercícios

Exercício 4.1. Mostre que uma função tem inversa se e somente se é sobrejetiva e injetiva.

Exercício 4.2. Demonstre os ítens (2) e (3) no Lema 4.2.1.

Exercício 4.3. Construa $f : (0, 1) \to \mathbb{R}$ injetiva e não monótona.

Exercício 4.4. Mostre que se $x \neq y$ são números reais, então existem vizinhanças U de x e V de y tais que $U \cap V = \emptyset$.

Exercício 4.5. Demonstre o Lema 4.4.5.

Exercício 4.6. Demonstre o Lema 4.4.6.

Exercício 4.7. Demonstre a equivalência 4.5.1.

CAPíTULO 5

Continuidade e Funções Contínuas

[1] A partir das definições de limites de funções do capítulo anterior, fica mais fácil definir continuidade e estudar suas propriedades.

5.1. Introdução e exemplos

Seja $A \subset \mathbb{R}$ e $f : A \to \mathbb{R}$. Dizemos que f é *contínua em* $c \in A$ se para todo $\epsilon > 0$ existe $\delta > 0$ tal que

$$x \in (c - \epsilon, c + \epsilon) \cap A \implies f(x) \in (c - \delta, c + \delta).$$

Finalmente, dizemos que f é contínua em $B \subset A$ se f for contínua em todos os pontos de B.

Observação. Note que, quando $\lim_{x\to c} f(x)$ está bem definido,

$$f \text{ é contínua em } c \iff f(c) = \lim_{x\to c} f(x).$$

Observação. Note uma diferença na definição de limite de função e continuidade num ponto c. Para definir limite, a função não precisava nem estar definida em c, e se estivesse, o valor de $f(c)$ não tinha importância.

Lema 5.1.1. Seja $A \subset \mathbb{R}$ e $f : \to \mathbb{R}$. Seja $c \in A$. Então as afirmativas abaixo são equivalentes.

(1) f é contínua em c.
(2) Para todo $\epsilon > 0$ existe $\delta > 0$ tal que

$$x \in A, \quad |x - c| < \delta \implies |f(x) - f(c)| < \epsilon.$$

(3) Se (x_n) é tal que $x_n \in A$ para todo $n \in \mathbb{N}$ e $\lim_{n\to\infty} x_n = c$, então $\lim_{n\to\infty} f(x_n) = f(c)$.

Outro resultado útil é o seguinte *critério de descontinuidade:* assumindo as hipóteses do Lema 5.1.1, temos que f não é contínua em c se e somente se existe sequência (x_n) em A convergindo para c mas $\big(f(x_n)\big)$ não convergindo para $f(c)$.

Exemplo 5.1. $g(x) = x$ é contínua em $\mathbb{R}$. De fato, para todo $c \in \mathbb{R}$, temos $\lim_{x\to c} g(x) = c = g(c)$.

Exemplo 5.2. A função sgn(x) (ver exemplo 4.6) não é contínua no zero, já que não existe $\lim_{x\to 0}$ sgn(x).

[1] Última Atualização: 23/06/2012

Exemplo 5.3. Seja $f : \mathbb{R} \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q}, \\ 0 & \text{caso contrário,} \end{cases}$$

é descontínua para todo $x \in \mathbb{R}$. Para mostrar isto, assuma $x \in \mathbb{Q}$, e uma sequência (x_n) em $\mathbb{R}\backslash\mathbb{Q}$ convergindo para x. Neste caso, $\lim_{n\to\infty}\big(f(x_n)\big) = 0 \neq 1 = f(x)$. Da mesma forma, se $x \notin \mathbb{Q}$, tomamos uma sequência (x_n) em $\mathbb{Q}$ convergindo para x, e temos $\lim_{n\to\infty}\big(f(x_n)\big) = 1 \neq 0 = f(x)$.

As vezes, é possível estender uma função de forma contínua para pontos "fora" do domínio original. Por exemplo, seja $I = (a, c)$ e $f : I \to \mathbb{R}$. Se existir $\lim_{x\to c} f(x)$, então definimos $f(c)$ como sendo este limite, e f será contínua em c.

Exemplo 5.4. Considere a função similar ao problema 4.4, mas desta vez definida apenas para reais positivos:

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

Então $\lim_{x\to 0} f(x) = 0$ e podemos estender f continuamente no zero definindo

$$g : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}, \quad g(x) = \begin{cases} f(x), & \text{se } x \in \mathbb{R}^+, \\ 0, & \text{se } x = 0. \end{cases}$$

Então temos g contínua no zero (e somente no zero).

Exemplo 5.5. É claro que nem sempre tal extensão contínua é possível. Por exemplo no caso de $f : \mathbb{R}^+ \to \mathbb{R}$ dada por $f(x) = 1/x$, não se pode definir $f(0)$ tal que $f : \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ seja contínua.

5.1.1. Composição de funções. Em geral, se f e g são contínuas, então $f + g$, $f - g$, fg também o são. Da mesma forma, se $h(x) \neq 0$ para todo x do domínio, então f/h é contínua. O próximo resultado garante que a *composição* de funções contínuas também é contínua.

Teorema 5.1.2. *Sejam A, $B \subset \mathbb{R}$, e $f : A \to B$ e $g : B \to \mathbb{R}$. Assuma f contínua em $c \in A$ e g contínua em $f(c) \in B$. Enão a composição $g \circ f : A \to \mathbb{R}$ é contínua em c.*

Exemplo 5.6. A função $g(x) = |x|$ é contínua em $\mathbb{R}$. Realmente, como

$$|g(x) - g(y)| = ||x| - |y|| \leq |x - y|,$$

se (x_n) converge para x então

$$|g(x_n) - g(x)| \leq |x_n - x| \implies \lim_{n\to\infty}\big(g(x_n)\big) = g(x).$$

Portanto, se $f : A \to \mathbb{R}$ é contínua em $c \in A$, entao $h(x) = |f(x)|$ também o é, pois $h = g \circ f$ é composição de funções contínuas.

5.2. Funções Contínuas em intervalos fechados e limitados

Um resultado com várias aplicações diz que funções contínuas definidas em conjuntos fechados e limitados são limitadas e atingem seus pontos extremos. Chamamos um intervalo de fechado limitado quando é da forma $[a,b]$, para $a<b$. Na verdade, todos os resultados abaixo, excetuando-se o Teorema 5.2.5, valem em conjuntos mais gerais, por exemplo em uniões finitas de intervalos fechados e limitados.

DEFINIÇÃO 5.2.1. *Dizemos que $f: A \to \mathbb{R}$ é limitada em A se existe $M \in \mathbb{R}$ tal que $|f(x)| \le M$ para todo $x \in A$.*

EXEMPLO 5.7. $\sin x$ é limitada em $\mathbb{R}$ pois $|\sin x| \le 1$ para todo $x \in \mathbb{R}$.

EXEMPLO 5.8. $1/x$ não é limitada em $\mathbb{R}^+$. Entretanto $1/x$ é limitada em $(1/2, +\infty)$ pois $|1/x| \le 2$ para todo x neste intervalo.

TEOREMA 5.2.2. *Seja $I = [a,b]$, e $f: I \to \mathbb{R}$ contínua em I. Então f é limitada em I.*

Outra noção importante é o de máximos e mínimos. Dizemos que $f: A \to \mathbb{R}$ tem valor máximo em A se existe $x^* \in A$ tal que $f(x^*)$ é cota superior de $f(A)$. De forma análoga dizemos que f tem valor mínimo em A se existe $x_* \in A$ tal que $f(x_*)$ é cota inferior de $f(A)$. Chamamos x^* de ponto de valor máximo e x_* de ponto de valor mínimo.

OBSERVAÇÃO. Se uma função f como acima definida assume seus valores máximo e mínimo em A, então f é limitada em A.

EXEMPLO 5.9. $f: (-1,1) \to \mathbb{R}$ dada por $f(x) = 1/(1-x^2)$ não é limitada em $(-1,1)$, mas é limitada em $[-1/2, 1/2]$ por exemplo.

EXEMPLO 5.10. $f(x) = x$ é contínua e limitada em $(-1,1)$, mas não assume valor máximo nem mínimo em $(-1,1)$. Entretanto f assume seus valores máximo e mínimo em $[-1,1]$.

EXEMPLO 5.11. $h(x) = 1/(1+x^2)$ é limitada em $\mathbb{R}$, assume seu valor máximo em $x^* = 0$, mas não assume seu valor mínimo. Isto porque $\inf h(\mathbb{R}) = 0 \ne h(x)$ para todo $x \in \mathbb{R}$.

OBSERVAÇÃO. Note que pontos de máximo e mínimo não são únicos em geral. Por exemplo, $f(x) = x^2$ tem -1 e 1 como seus dois pontos de máximo em $[-1,1]$.

TEOREMA 5.2.3 (Pontos Extremos). *Seja $I = [a,b]$, e $f: I \to \mathbb{R}$ contínua em I. Então f tem pelo menos um ponto de máximo e um de mínimo em I.*

Outro resultado de grande importância é o Teorema do Valor Intermediário que garante a preservação de intervalos por funções contínua.

TEOREMA 5.2.4 (Teorema do Valor Intermediário). *Sejam $a<b$ e $f: [a,b] \to \mathbb{R}$ contínua. Se existe $d \in \mathbb{R}$ tal que $f(a) < d < f(b)$, então existe $c \in (a,b)$ tal que $f(c) = d$.*

Concluímos esta parte com uma importante consequência dos resultados anteriores.

TEOREMA 5.2.5. *Seja I intervalo fechado limitado e $f: I \to \mathbb{R}$ função contínua. Então $f(I)$ é intervalo fechado limitado.*

5.3. Exercícios

EXERCÍCIO 5.1. Determine os pontos de continuidade da função $[x]$, que retorna para cada $x \in \mathbb{R}$ o maior inteiro menor ou igual a x. Por exemplo, $[2] = 2$, $[2.5] = 2$, $[-2.5] = -3$.

CAPíTULO 6

Diferenciação

[1] Neste capítulo vemos a noção de diferenciabilidade e suas aplicações.

6.1. Definições e Exemplos

Seja $f : I \to \mathbb{R}$, onde I é um intervalo em $\mathbb{R}$. Dizemos que f é diferenciável em $c \in I$ se existe um número real L onde dado $\epsilon > 0$ existe $\delta > 0$ tal que

$$x \in I, \quad 0 < |x - c| < \delta \implies \left| \frac{f(x) - f(c)}{x - c} - L \right| < \epsilon.$$

Chamamos L de derivada de f em c, e escrevemos $L = f'(c)$.

Note que se f é diferenciável em c, então

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}.$$

Se f é diferenciável em todo ponto de I dizemos que f é diferenciável em I. Neste caso note que a derivada f' é uma função de I em $\mathbb{R}$.

Exemplo 6.1. Se $f(x) = x^2$, então para $c \in \mathbb{R}$ tem-se

$$f'(c) = \lim_{x \to c} \frac{x^2 - c^2}{x - c} = \lim_{x \to c} \frac{(x + c)(x - c)}{x - c} = \lim_{x \to c} (x + c) = 2c.$$

Teorema 6.1.1. *Se $f : I \to \mathbb{R}$, onde I é um intervalo em $\mathbb{R}$ é diferenciável em $c \in I$, então f é continua em c.*

Demonstração. Seja $L = f'(c)$. Dado $\epsilon > 0$, existe $\delta > 0$ tal que

$$x \in I, \quad 0 < |x - c| < \delta \implies L - \epsilon < \left| \frac{f(x) - f(c)}{x - c} \right| < L + \epsilon.$$

Seja $\bar{\delta} = \min\{\delta, \epsilon/(L + \epsilon)\}$. Então

$$x \in I, \quad 0 < |x - c| < \bar{\delta} \implies |f(x) - f(c)| = \left| \frac{f(x) - f(c)}{x - c} \right| |x - c| \leq (L + \epsilon)\delta \leq \epsilon.$$

Logo f é continua em c. □

Observação. Pelo teorema acima, diferenciabilidade implica em continuidade. O inverso entretanto não é verdade em geral. Seja por exemplo $f : \mathbb{R} \to \mathbb{R}$ onde $f(x) = |x|$. Então f é continua em $\mathbb{R}$ mas não é diferenciável em zero pois para $x \neq 0$ temos

$$\left| \frac{f(x) - f(0)}{x - 0} \right| = \frac{|x|}{x} = \begin{cases} 1 & \text{se } x > 0, \\ -1 & \text{se } x < 0. \end{cases}$$

[1] Última Atualização: 26/06/2012

Logo o limite quando $x \to 0$ não existe.

6.2. Propriedades da Derivada

Seja f e g funções de $I \to \mathbb{R}$, onde I é um intervalo em $\mathbb{R}$, ambas diferenciáveis em $c \in I$. Então

(1) $(\alpha f)'(c) = \alpha f'(c)$, onde $\alpha \in \mathbb{R}$. De fato, se $x \neq c$, então

$$\frac{(\alpha f)(x) - (\alpha f)(c)}{x - c} = \alpha \frac{f(x) - f(c)}{x - c}.$$

(2) $(f + g)'(c) = f'(c) + g'(c)$.

(3) Se $p = fg$, então se $x \neq c$,

$$\frac{p(x) - p(c)}{x - c} = \frac{f(x)g(x) - f(c)g(c)}{x - c} = \frac{f(x)g(x) - f(c)g(x) + f(c)g(x) - f(c)g(c)}{x - c}$$
$$= \frac{f(x) - f(c)}{x - c} g(x) + f(c) \frac{g(x) - g(c)}{x - c}.$$

Logo existe $\lim_{x \to c}(p(x) - p(c))/(x - c)$ e

$$p'(c) = \lim_{x \to c} \frac{p(x) - p(c)}{x - c} = \lim_{x \to c} \left[\frac{f(x) - f(c)}{x - c} g(x)\right] + \lim_{x \to c} \left[f(c) \frac{g(x) - g(c)}{x - c}\right]$$
$$= f'(c)g(c) + f(c)g'(c).$$

(4) Se $g(x) \neq 0$ para todo $x \in I$, então seja $h(x) = f(x)/g(x)$. Logo se $x \neq c$,

$$\frac{h(x) - h(c)}{x - c} = \frac{\frac{f(x)}{g(x)} - \frac{f(c)}{g(c)}}{x - c} = \frac{f(x)g(c) - f(c)g(x)}{(x - c)g(x)g(c)}$$
$$= \frac{f(x)g(c) - f(c)g(c)}{(x - c)g(x)g(c)} + \frac{f(c)g(c) - f(c)g(x)}{(x - c)g(x)g(c)} = \frac{f(x) - f(c)}{(x - c)} \frac{1}{g(x)} - \frac{f(c)}{g(x)g(c)} \frac{g(x) - g(c)}{x - c}.$$

Logo existe $\lim_{x \to c}(h(x) - h(c))/(x - c)$ e

$$h'(c) = \lim_{x \to c} \frac{h(x) - h(c)}{x - c} = f'(c) \frac{1}{g'(c)} - \frac{f(c)}{g^2(x)} g'(c).$$

Exemplo 6.2. Pela regra acima temos que se $f(x) = x^n$, para $n \in \mathbb{N}$, então f é diferenciável e $f'(c) = nx^{n-1}$.

Observe que $f : I \to \mathbb{R}$ é diferenciável em $c \in I$ com $f'(c) = L$ se e somente se existir uma função r tal que

$$f(x) = f(c) + (x - c)L + r(x - c), \quad \text{com } \lim_{h \to 0} \frac{r(h)}{h} = 0.$$

De forma equivalente escrevemos $h = x - c$ e

$$f(c + h) = f(c) + hL + r(h) \quad \text{com } \lim_{h \to 0} \frac{r(h)}{h} = 0.$$

TEOREMA 6.2.1 (Regra da Cadeia). *Sejam I e J intervalos em $\mathbb{R}$ e $g : I \to \mathbb{R}$ e $f : J \to \mathbb{R}$, onde $f(J) \subset I$. Se f é diferenciável em $c \in J$ e g é diferenciável em $f(c)$, ent ao $g \circ f$ é diferencável em c e*

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

DEMONSTRAÇÃO. Seja $d = f(c)$. Note que para h tal que $c+h \in J$ e k tal que $d+k \in I$, temos

$$f(c+h) = f(c) + hf'(c) + r(h) \quad \text{com } \lim_{h\to 0} \frac{r(h)}{h} = 0.$$
$$g(d+k) = g(d) + kg'(d) + p(k) \quad \text{com } \lim_{k\to 0} \frac{p(k)}{k} = 0.$$

Definindo $k = f(c+h) - f(c) = hf'(c) + r(h)$, temos

$$\begin{aligned} g \circ f(c+h) &= g(f(c+h)) = g(d+k) = g(d) + kg'(d) + p(k) \\ &= g(d) + (hf'(c) + r(h))g'(d) + p(f(c+h) - f(c)) = g(d) + hf'(c)g'(d) + q(h) \end{aligned}$$

onde $q(h) = r(h)g'(d) + p(f(c+h) - f(c))$. Finalmente,

$$\lim_{h\to 0} \frac{q(h)}{h} = g'(d) \lim_{h\to 0} \frac{r(h)}{h} + \lim_{h\to 0} \frac{p(f(c+h) - f(c))}{h}.$$

Se $f(c+h) = f(c)$ numa vizinhança de c, então $p(f(c+h) - f(c)) = 0$. Caso contrário,

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

De qualquer forma concluímos que

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

□

EXEMPLO 6.3. Seja

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x}, & \text{se } x \neq 0 \\ 0, & \text{se } x = 0. \end{cases}$$

Logo, para $x \neq 0$ temos $f'(x) = 2x \sin 1/x - \cos 1/x$. Em $x = 0$ usamos a definição:

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0} x \sin \frac{1}{x} = 0.$$

Logo f é diferenciável em $\mathbb{R}$ mas f' não é contínua no zero.

TEOREMA 6.2.2 (Derivada da Função Inversa). *Seja $I \subset \mathbb{R}$ intervalo, $f : I \to \mathbb{R}$ contínua e invertível com inversa $g : J \to \mathbb{R}$ contínua, e $J = f(I)$. Se f é diferenciável em $c \in I$, então g é diferencável em $d = f(c)$ se e somente se $f'(c) \neq 0$. Neste caso,*

$$g'(d) = \frac{1}{f'(c)} = \frac{1}{f'(g(d))}$$

DEMONSTRAÇÃO. Tendo que g é continua. Al'em disso, se $y \in J\backslash\{d\}$, então $g(y) \neq c$. Logo, se $f'(c) \neq 0$,

$$\lim_{y\to d}\frac{g(y)-g(d)}{y-d}=\lim_{y\to d}\frac{g(y)-c}{f(g(y))-f(c)}=\lim_{y\to d}\left(\frac{f(g(y))-f(c)}{g(y)-c}\right)^{-1}=\frac{1}{f'(c)}.$$

Logo g é diferenciável em d e $g'(d)=1/f'(c)$. Analogamente, se g é diferenciável em d, então usando a regra da cadeia e que $g(f(x))=x$, temos

$$g'(f(c))f'(c)=1,$$

e então $f'(c)\neq 0$. □

EXEMPLO 6.4. Seja $f:\mathbb{R}^+\to\mathbb{R}^+$ dada por $f(x)=x^n$, onde $n\in\mathbb{N}$. Então f tem inversa $g:\mathbb{R}^+\to\mathbb{R}^+$, e $g(y)=\sqrt[n]{y}$. Para $y>0$ temos então

$$g'(y)=\frac{1}{ny^{\frac{n-1}{n}}}.$$

Note que g não é diferenciável no zero pois $f'(0)=0$.

6.3. Aplicações

Uma primeira e importante aplicação diz respeito a pontos extremos locais. Dizemos que uma função$f:I\to\mathbb{R}$, onde $I\subset\mathbb{R}$ é um intervalo, tem um *máximo local* em $c\in I$ se existe $\delta>0$ tal que

$$x\in(c-\delta,c+\delta)\cap I \implies f(x)\leq f(c).$$

Definição análoga serve para mínimo local. Chamamos um ponto de máximo ou mínimo local de ponto extremo local.

O resultado a seguir descreve condição necessária para um ponto ser extremo local.

TEOREMA 6.3.1 (Ponto extremo interior). *Seja $f:I\to\mathbb{R}$, onde $I\subset\mathbb{R}$ é um intervalo, e $c\in I$ ponto extremo local. Se f é diferenciável em c, então $f'(c)=0$.*

DEMONSTRAÇÃO. Sem perda de generalidade, assuma c ponto de máximo local. Então, se $f'(c)>0$ temos

$$0<f'(c)=\lim_{x\to c}\frac{f(x)-f(c)}{x-c}\implies\frac{f(x)-f(c)}{x-c}>0$$

numa vizinhança de c. Logo, para $x>c$ tem-se $f(x)>f(c)$, contradição pois c é ponto de máximo local. De forma semelhante não podemos ter $f'(c)<0$. Logo $f'(c)=0$. □

A seguir apresentamos um resultado com importantes por si e por suas consequências. É o *Teorema do Valor Médio*, que vemos a seguir na sua versão mais simples, o *Teorema de Rolle*.

TEOREMA 6.3.2 (Teorema de Rolle). *Seja $a<b\in\mathbb{R}$ e $f:[a,b]\to\mathbb{R}$ continua e diferenciável em $[a,b]$. Assuma ainda que $f(a)=f(b)=0$. Então existe $c\in(a,b)$ tal que $f'(c)=0$.*

Demonstração. Se f é identicamente nula em $[a, b]$, então o resultado é verdadeiro. Caso contrário, então f assume algum valor positivo ou negativo em (a, b). Sem perda de generalidade, suponha que f assuma algum valor positivo. Como $[a, b]$ é intervalo fechado e limitado, então f atinge seu máximo em algum $c \in (a, b)$. Mas pelo Teorema do Ponto extremo interior 6.3.1, $f'(c) = 0$, como queríamos demonstrar. □

Teorema 6.3.3 (Teorema do Valor Médio). *Seja $a < b \in \mathbb{R}$ e $f : [a, b] \to \mathbb{R}$ continua e diferenciável em $[a, b]$. Então existe $c \in (a, b)$ tal que*

$$f(b) - f(a) = f'(c)(b - a).$$

Demonstração. Seja

$$\phi(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a).$$

Então $\phi(a) = \phi(b) = 0$. Como f é diferenciável em $[a, b]$, então ϕ também o é no mesmo intervalo. Logo, pelo Teorem de Rolle 6.3.2 existe $c \in (a, b)$ tal que $\phi'(c) = 0$. Portanto

$$f'(x) = \frac{f(b) - f(a)}{b - a}.$$

□

Uma primeira aplicação do Teorema do Valor Médio garante que se uma função definida num intervalo tem derivada identicamente igual a zero, então a função é constante.

Lema 6.3.4. Assuma que $f : [a, b] \to \mathbb{R}$ seja contínua em $[a, b]$, onde $a < b$, e diferenciável em (a, b). Se $f'(x) = 0$ para todo x, então f é constante em $[a, b]$.

Demonstração. Seja $a < x < b$. Pelo Teorema do Valor Médio 6.3.3, existe $c \in (a, x)$ tal que $f(x) - f(a) = f'(c)(x - a)$. Como $f'(c) = 0$, temos $f(x) = f(a)$. Como x é arbitrário, temos f constante em (a, b). Mas continuidade temos f constante em $[a, b]$. □

Observe que pelo resultado acima, se f, g são funções diferenciáveis que tem a mesma derivada, então f e g diferem por uma constante.

A aplicação seguinte do Teorema do Valor Médio garante condições necessárias e suficientes para uma função ser crescente num intervalo.

Lema 6.3.5. Seja $I \subset \mathbb{R}$ intervalo e $f : I \to \mathbb{R}$ diferenciável em I. Então

1. f é crescente em I se e somente se $f'(x) \geq 0$ para todo $x \in I$.
2. f é decrescente em I se e somente se $f'(x) \leq 0$ para todo $x \in I$.

Demonstração. Assuma f crescente.

($\Longrightarrow$) Para x, $c \in I$,

$$x < c \text{ ou } x > c \implies \frac{f(x) - f(c)}{x - c} \geq 0.$$

Portanto

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c} \geq 0.$$

($\Longrightarrow$) Assuma $f'(x) \geq 0$ para todo $x \in I$. Sejam $x_1 < x_2$ com $x_1 < x_2 \in I$. Usando o teorema do valor médio 6.3.3, existe $c \in (x_1, x_2)$. □

Observação. É possível modificar a demonstração acima e mostrar que $f'(x) > 0$ implica em f estritamente crescente. Entretanto, mesmo funções que tem derivada nula em alguns pontos podem ser estritamente crescentes, como por exemplo $f(x) = x^3$.

Observação. Não é verdade que se $f'(c) > 0$ para algum ponto c no domínio da f implique em f crescente numa vizinhança de c. Como exemplo considere

$$g(x) = \begin{cases} x + 2x^2 \sin \frac{1}{x} & \text{se } x \neq 0, \\ 0 & \text{se } x = 0, \end{cases}$$

é diferenciável em zero com $g'(0) = 1$, mas não é crescente e, nenhuma vizinhança do zero.

Outra aplicação do Teorema do Valor Médio segue no exemplo abaixo.

Exemplo 6.5. Seja $f(x) = \exp(x)$. Então $f'(x) = \exp(x)$. Queremos mostrar que

$$\exp(x) > 1 + x \text{ para todo} x \neq 0. \tag{6.3.1}$$

Seja $x > 0$. Então aplicando o Teorema do Valor Médio em $[0, x]$ temos que existe $c \in (0, x)$ tal que

$$\exp(x) - \exp(0) = \exp(c)(x - 0).$$

Como $c > 0$, então $\exp(c) > \exp(0) = 1$. Logo

$$\exp(x) > 1 + x.$$

Para $x < 0$, os argumentos são semelhantes e portanto a desigualdade (6.3.1) vale.

6.4. Teorema de Taylor e Aplicações

Uma ferramenta poderosa em análise com várias consequências é o Teorema de Taylor, que é na verdade também uma aplicação do Teorema do Valor Médio.

A expansão de Taylor aproxima localmente uma função que pode ser complicada por um polinômio. Suponha que $f : I \to \mathbb{R}$ onde $I \subset \mathbb{R}$ tenha $n \geq 0$ derivadas num ponto $x_0 \in I$. Defina

$$P_n(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x - x_0)^2}{2} + \cdots + f^{(n)}(x_0)\frac{(x - x_0)^n}{n!},$$

onde usamos a notaçãoque $g^{(k)}(c)$ indica a k-ésima deriva de g num ponto c.

Note que com a definição acima, temos $f^{(k)}(x_0) = P_n^{(k)}(x_0)$ para $k = 1, \dots, n$. Chamamos P_n de polinômio de Taylor de ordem n para f em x_0, e o resultado abaixo diz o quão boa é a aproximação de uma função por seu polinômio de Taylor.

Teorema 6.4.1 (Fórmula de Taylor com resto de Lagrange). *Seja $n \geq 0$ e $I = [a, b]$, com $a < b$. Seja $f : I \to \mathbb{R}$ n vezes diferenciável em I com $f^{(n)}$ contínua em I e tal $f^{(n+1)}$ exista em (a, b). Se x_0, $x \in I$ então existe $\xi \in (x_0, x) \cap (x, x_0)$ tal que*

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x - x_0)^2}{2} + \cdots + f^{(n)}(x_0)\frac{(x - x_0)^n}{n!} + f^{(n+1)}(\xi)\frac{(x - x_0)^{n+1}}{(n+1)!}. \tag{6.4.1}$$

Demonstração. Sejam x_0, $x \in I$. Sem perda de generalidade, assuma $x > x_0$. Defina $J = [x_0, x]$ e seja $F : J \to \mathbb{R}$ dada por

$$F(t) = f(x) - f(t) - (x-t)f'(t) - \cdots - \frac{(x-t)^n}{n!}f^{(n)}(t).$$

Logo

$$F'(t) = -\frac{(x-t)^n}{n!}f^{(n+1)}(t)$$

Definindo $G : J \to \mathbb{R}$ por

$$G(t) = F(t) - \left(\frac{x-t}{x-x_0}\right)^{n+1} F(x_0),$$

temos $G(x_0) = G(x) = 0$. Pelo Teorema de Rolle 6.3.2 existe $\xi \in (x_0, x)$ tal que

$$0 = G'(\xi) = F'(\xi) + (n+1)\frac{(x-\xi)^n}{(x-x_0)^{n+1}}F(x_0).$$

Portanto

$$\begin{aligned} F(x_0) = -\frac{1}{n+1}\frac{(x-x_0)^{n+1}}{(x-\xi)^n}F'(\xi) = \frac{1}{n+1}\frac{(x-x_0)^{n+1}}{(x-\xi)^n}\frac{(x-\xi)^n}{n!}f^{(n+1)}(\xi) \\ = \frac{(x-x_0)^{n+1}}{(n+1)!}f^{(n+1)}(\xi). \end{aligned}$$

□

Exemplo 6.6. Seja $f : I \to \mathbb{R}$, onde $I = [a, b] \subset \mathbb{R}$, com $a < b$. Assuma que f e suas derivadas f', f'',..., $f^{(n+1)}$ existam e sejam contínuas em I. Se $f^{(n+1)}(x) = 0$ para todo $x \in I$ e $f(x_0) = f'(x_0) = \cdots = f^{(n)}(x_0) = 0$ para algum $x_0 \in I$, então $f(x) = 0$ para todo $x \in I$. De fato, pelo Teorema de Taylor 6.4.1, dado $x \in I$, existe ξ entre x e x_0 tal que

$$\begin{aligned} f(x) = f(x_0) + f'(x_0)(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2} + \cdots + f^{(n)}(x_0)\frac{(x-x_0)^n}{n!} \\ + f^{(n+1)}(\xi)\frac{(x-x_0)^{n+1}}{(n+1)!}. \end{aligned}$$

Mas por hipótese, $f^{(i)}(x_0)$ para $i = 0, \ldots, n$, e $f^{(n+1)} \equiv 0$ em I. Em particular, como $\xi \in I$, temos $f^{(n+1)}(\xi) = 0$. Portanto, $f(x) = 0$ para todo $x \in I$.

Uma primeira aplicação refere-se à caracterização de extremos locais.

Teorema 6.4.2. *Seja $a < b \in \mathbb{R}$ e $I = [a, b]$. Sejam $x_0 \in (a, b)$ e $k \geq 2$ número inteiro. Supondo que f',...,$f^{(k)}$ existam, que sejam contínuas em I, e que $f'(x_0) = \cdots = f^{(k-1)}(x_0) = 0$ mas $f^{(k)}(x_0) \neq 0$, temos que*

(1) Se k é par e $f^{(k)}(x_0) > 0$, então f tem mínimo local em x_0.
(2) Se k é par e $f^{(k)}(x_0) < 0$, então f tem máximo local em x_0.
(3) Se k é ímpar, então x_0 não é máximo nem mínimo local.

DEMONSTRAÇÃO. Pelo Teorema de Taylor, para $x \in I$ existe ξ entre x_0 e x tal que

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(x_0)\frac{(x-x_0)^2}{2} + \cdots + f^{(k-1)}(x_0)\frac{(x-x_0)^{(k-1)}}{(k-1)!} + f^k(\xi)\frac{(x-x_0)^k}{k!} = f(x_0) + f^k(\xi)\frac{(x-x_0)^k}{k!}.$$

Assumindo agora que $f^{(k)}(x_0) > 0$, como $f^{(k)}$ é contínua então existe $\delta > 0$ tal que $f^{(k)}(x) > 0$ para todo $x \in U = (x_0 - \delta, x_0 + \delta)$. Se $x \in U$, então $\xi \in U$ e então $f^{(\xi)}(x) > 0$. Se n é par, então para $x \neq x_0$ temos

$$f^k(\xi)\frac{(x-x_0)^k}{k!} > 0.$$

Logo

$$x \in U\backslash\{x_0\} \implies f(x) - f(x_0) > 0 \implies x_0 \text{ é mínimo local},$$

e portanto (1) está demonstrado.

Para demonstrar (2) o argumento é semelhante.

Finalmente, se k é ímpar, então $(x-x_0)/k!$ é positivo para $x > x_0$ e negativo para $x < x_0$. Logo $f(x) > f(x_0)$ ou $f(x) < f(x_0)$ dependendo do sinal de $x - x_0$. Logo a proposição (3) é verdadeira. $\square$

Uma segunda aplicação diz respeito à funções convexas. Seja $I \subset \mathbb{R}$ um intervalo. Dizemos que $f : I \to \mathbb{R}$ é convexa em I se para todo $t \in [0,1]$ e x_1, $x_2 \in I$ temos

$$f\big((1-t)x_1 + tx_2\big) \leq (1-t)f(x_1) + tf(x_2).$$

Graficamente, uma função é convexa se o gráfico de f entre x_1 e x_2 está abaixo da reta que une os pontos $(x_1, f(x_1))$ e $(x_2, f(x_2))$.

TEOREMA 6.4.3. *Seja I intervalo aberto e $f : I \to \mathbb{R}$. Então f é convexa se e somente se $f''(x) \geq 0$ para todo $x \in I$.*

DEMONSTRAÇÃO. ($\Longleftarrow$)Assuma que $f''(x) \geq 0$ para todo $x \in I$. Sejam $x_1 < x_2 \in I$ e $0 < t < 1$. Definindo $x_0 = (1-t)x_1 + tx_2$, pelo Teorema de Taylor existe $\xi_1 \in (x_1, x_0)$ e $\xi_2 \in (x_0, x_2)$ tais que

$$f(x_1) = f(x_0) + f'(x_0)(x_1 - x_0) + \frac{1}{2}f''(\xi_1)(x_1 - x_0)^2,$$
$$f(x_2) = f(x_0) + f'(x_0)(x_2 - x_0) + \frac{1}{2}f''(\xi_2)(x_2 - x_0)^2.$$

Como $f''(\xi_1) \geq 0$ e $f''(\xi_2) \geq 0$, então

$$\begin{aligned}
&(1-t)f(x_1) + tf(x_2) \\
&\quad = f(x_0) + [(1-t)x_1 + tx_2 - x_0]f'(x_0) + \frac{(1-t)}{2}f''(\xi_1)(x_1 - x_0)^2 + \frac{t}{2}f''(\xi_2)(x_2 - x_0)^2 \\
&\quad = f(x_0) + \frac{(1-t)}{2}f''(\xi_1)(x_1 - x_0)^2 + \frac{t}{2}f''(\xi_2)(x_2 - x_0)^2 \geq f(x_0).
\end{aligned}$$

Logo f é convexa.

($\Longrightarrow$) Sejam $x_1 < x < x_2 \in I$. Então $x = (1-t)x_1 + tx_2$ para $t = (f(x_2) - f(x_1))/(x_2 - x_1)$. Logo, se f é convexa,

$$\frac{f(x) - f(x_1)}{x - x_1} \leq \frac{(1-t)f(x_1) + tf(x_2) - f(x_1)}{t(x_2 - x_1)} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

e

$$\frac{f(x_2) - f(x)}{x_2 - x} \geq \frac{f(x_2) - [(1-t)f(x_1) + tf(x_2)]}{(1-t)(x_2 - x_1)} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}.$$

Portanto,

$$x_1 < x < x_2 \Longrightarrow \frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_2) - f(x)}{x_2 - x},$$

e

$$x_1 < x_2 \Longrightarrow f'(x_1) = \lim_{x \to x_1} \frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \lim_{x \to x_2} \frac{f(x_2) - f(x)}{x_2 - x} = f'(x_2).$$

Logo f' é função crescente em I e então $f''(x) \geq 0$ para todo $x \in I$. □

6.4.1. pontos de inflexão e concavidades. Seja f uma função real duas vezes diferenciável. Dizemos que um ponto é de inflexão se este "separa" curvas de concavidades contrárias. Se c é ponto de inflexão então $f''(c) = 0$. Para descobrir se um ponto c onde a segunda derivada se anula é de inflexão, basta checar se f'' muda de sinal no intervalo $(c - \epsilon, c + \epsilon)$, para todo $\epsilon > 0$.

Quanto a concavidades, dizemos que f é *côncava* (para baixo) em (a, b) se $f'' < 0$ em (a, b). Dizemos que f é *convexa* (côncava para cima) em (a, b) se $f'' > 0$ em (a, b).

6.5. Regra de L'Hôpital

Considere o problema de achar

$$\lim_{x \to 0} \frac{\sin x}{x},$$

se este limite existir. Surpreendentemente, vale a regra de que, nestes casos, o limite da razão das funções é igual ao limite da razão das derivadas das funções.

TEOREMA 6.5.1. *Sejam f e g duas funções reais diferenciáveis definiddas na intervalo (a, b). Suponha também que g e g' seja não nula e que*

$$\lim_{x \to b} f(x) = \lim_{x \to b} g(x) = 0, \text{ ou que } \lim_{x \to b} f(x) = \lim_{x \to b} g(x) = \pm\infty.$$

Temos então que

$$\text{se } \lim_{x \to b} \frac{f'(x)}{g'(x)} = \alpha, \text{ então } \lim_{x \to b} \frac{f(x)}{g(x)} = \alpha.$$

Mesmo se α for $-\infty$ ou ∞, o resultado continua valendo. Vale também se $x \to a$, ou mesmo para pontos interiores, onde basta tomar os dois limites laterais.

EXEMPLO 6.7.

$$\lim_{x \to 0} \frac{\sin x}{x} = \lim_{x \to 0} \frac{\cos x}{1} = 1.$$

Exemplo 6.8. Nunca usar l'Hôspital cegamente! Aplicando l'Hôspital:

$$\lim_{x\to\pi^-} \frac{\sin x}{1-\cos x} = \lim_{x\to\pi^-} \frac{\cos x}{\sin x} = -\infty.$$

Isto está errado! Note que o numerador $\lim_{x\to\pi^-} \sin x = 0$. Entretanto, o denominador $\lim_{x\to\pi^-}(1-\cos x) = 2$. Logo

$$\lim_{x\to\pi^-} \frac{\sin x}{1-\cos x} = \frac{\lim_{x\to\pi^-} \sin x}{\lim_{x\to\pi^-}(1-\cos x)} = 0.$$

Considere o limite $\lim_{x\to a} f(x)g(x)$. Se $\lim_{x\to a} f(x) = 0$ e $\lim_{x\to a} g(x) = \infty$, temos limites de produtos indeterminados do tipo "$\infty \cdot 0$". Escrevemos o produto como uma das opções abaixo

$$fg = \frac{f}{1/g} = \frac{g}{1/f},$$

e depois aplicamos l'Hôspital.

Exemplo 6.9.

$$\lim_{x\to 0} x\ln x = \lim_{x\to 0} \frac{\ln x}{1/x} = \lim_{x\to 0} \frac{1/x}{-1/x^2} = -\lim_{x\to 0} x = 0.$$

O caso de *diferença indeterminada* "$\infty - \infty$" pode ser calculado da seguinte forma. Suponha que $\lim_{x\to a} f(x) = \infty$ e $\lim_{x\to a} g(x) = \infty$. Então $\lim_{x\to a}[f(x) - g(x)]$ pode ser calculado manipulando-se algebricamente $f - g$ para obter um limite da forma ∞/∞ ou $0/0$.

Exemplo 6.10.

$$\lim_{x\to(\pi/2)^-}(\sec x - \tan x) = \lim_{x\to(\pi/2)^-}\left(\frac{1}{\cos x} - \frac{\sin x}{\cos x}\right) = \lim_{x\to(\pi/2)^-} \frac{1-\sin x}{\cos x} = \lim_{x\to(\pi/2)^-} \frac{-\cos x}{-\sin x} = 0.$$

Potências indeterminadas 0^0, ∞^0, 1^∞, podem surgir quando calculamos limites do tipo $f(x)^{g(x)}$. Nestes casos tomamos logarítimos:

$$y = f(x)^{g(x)} \implies \ln y = g(x)\ln f(x),$$

ou escrever $f(x)^{g(x)} = e^{g(x)\ln f(x)}$.

Exemplo 6.11. O $\lim_{x\to 0^+}(1+\sin 4x)^{\cot x}$ pode ser calculado definindo-se $y = (1 + \sin 4x)^{\cot x}$ e portanto

$$\ln y = \cot x \ln(1+\sin 4x)$$

e então

$$\lim_{x\to 0^+} \ln y = \lim_{x\to 0^+} \cot x \ln(1+\sin 4x) = \lim_{x\to 0^+} \frac{\ln(1+\sin 4x)}{\tan x} = \lim_{x\to 0^+} \frac{\frac{4\cos 4x}{1+\sin 4x}}{\sec^2 x} = 4.$$

Note que $\lim_{x\to 0^+} \ln y = 4$, então

$$\lim_{x\to 0^+}(1+\sin 4x)^{\cot x} = \lim_{x\to 0^+} y = \lim_{x\to 0^+} e^{\ln y} = e^4.$$

Exemplo 6.12. Para calcular $\lim_{x\to 0^+} x^x$ escrevemos $x^x = e^{x\ln x}$, e então

$$\lim_{x\to 0^+} x^x = \lim_{x\to 0^+} e^{x\ln x} = e^0 = 1.$$

Note que usamos $\lim_{x\to 0^+} x\ln x = 0$.

6.6. Exercícios

EXERCÍCIO 6.1. Assuma $f : \mathbb{R} \to \mathbb{R}$ diferenciável em $c \in \mathbb{R}$ e $f(c) = 0$. Mostre então que $g(x) = |f(x)|$ é diferenciável em c se e somente se $f'(c) = 0$.

EXERCÍCIO 6.2. Seja $f : \mathbb{R} \to \mathbb{R}$ dada por

$$f(x) = \sum_{i=1}^{n} (x - c_i)^2,$$

onde $c_i \in \mathbb{R}$ para $i = 1, \ldots, n$, e $n \in \mathbb{N}$. Ache um ponto de mínimo relativo de f. Mostre que é único.

EXERCÍCIO 6.3. Seja $I \subset \mathbb{R}$ um intervalo e $f : I \to \mathbb{R}$ diferenciável. Mostre que se f' é positiva em I, i.e., $f'(x) > 0$ para todo $x \in I$, então f é estritamente crescente.

CAPíTULO 7

Funções trigonométricas, logarítmicas e exponenciais

[1]

Neste capítulo descrevemos algumas funções especiais, como as funções trigonométricas, o logarítmo e a exponencial.

7.1. Funções trigonométricas

Definimos aqui algumas funções trigonométricas, começando pelas funções seno e cosseno. Nossas definições diferem das definições geométricas "usuais", pois usamos séries de potências. Entretanto são as mesmas funções, como pode ser visto em [Djairo Figueiredo].

7.1.1. Senos e cossenos. Definimos as funções sin e cos de $\mathbb{R} \to \mathbb{R}$ através das séries de potências

$$\sin x = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n+1}}{(2n+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots, \qquad \cos x = \sum_{n=0}^{\infty}(-1)^n \frac{x^{2n}}{(2n)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \dots.$$

A derivação termo a termo nas séries de potências acima é válida, e portanto

$$\sin' x = \sum_{n=0}^{\infty}(-1)^n(2n+1)\frac{x^{2n}}{(2n+1)!} = \cos x, \qquad \cos' x = \sum_{n=0}^{\infty}(-1)^n 2n \frac{x^{2n-1}}{(2n)!} = -\sin x.$$

Note que das definições, a função sen é ímpar ($\sin(-x) = -\sin x$), e a função cosseno é par($\cos(-x) = \cos x$). Temos ainda uma igualdade fundamental, dada pelo resultado abaixo.

Lema 7.1.1. Para todo $x \in \mathbb{R}$, tem-se que $\sin^2 x + \cos^2 x = 1$.

Demonstração. Seja $f(x) = \sin^2 x + \cos^2$. Então $f'(x) = 2\sin x \cos x - 2\cos x \sin x = 0$. Logo f é constante. Basta agora ver que $f(0) = 1$, e então

$$1 = f(0) = f(x) = \sin^2 x + \cos^2 .$$

□

Corolário 7.1.2. Para todo $x \in \mathbb{R}$, temos que $|\sin x| \le 1$ e $|\cos x| \le 1$.

Valem também as identidades abaixo.

Lema 7.1.3. Para todo $x \in \mathbb{R}$, tem-se que

$$\sin(a+b) = \sin a \cos b + \sin b \cos a, \qquad \cos(a+b) = \cos a \cos b - \sin a \sin b.$$

Finalmente, uma propriedade importante destas funções são suas periodicidades. Dizemos que uma função $f : \mathbb{R} \to \mathbb{R}$ é *periódica*, com período T se $f(x+T) = f(x)$ para todo $x \in \mathbb{R}$.

[1]Última Atualização: 22/06/2020

LEMA 7.1.4. As funções sin e cos são periódicas com período 2π.

7.1.2. Outras funções trigonométricas. Seja $\tan x = \sin x/\cos x$ definida em $\mathbb{R}$ excetuando-se $\{\pm\pi/2, \pm 3\pi/2, \dots\}$. Note que $\tan x$ é periódica com período π, pois

$$\tan(x+\pi) = \frac{\sin(x+\pi)}{\cos(x+\pi)} = \frac{-\sin x}{-\cos x} = \tan x.$$

Outras funções trigonométricas são:

$$\cot x = \frac{\cos x}{\sin x}, \qquad \sec x = \frac{1}{\cos x}, \qquad \csc x = \frac{1}{\sin x}.$$

7.1.3. Funções trigonométricas inversas. Note que como a função $\sin x$ é crescente em $[-\pi/2, \pi/2]$, então ela possui uma função inversa, que é denominada de $\arcsin x$. Mais especificamente,

$$\begin{aligned} \arcsin : [-1,1] &\to [-\pi/2, \pi/2] \\ y &\mapsto \arcsin y. \end{aligned}$$

Esta função é diferenciável e

$$\frac{d}{dy}\arcsin y = \frac{1}{\sin'(\arcsin y)} = \frac{1}{\cos(\arcsin y)}.$$

Mas usando que $\cos^2(\arcsin y) = 1 - \sin^2(\arcsin y) = 1 - y^2$, concluímos que

$$\frac{d}{dy}\arcsin y = \frac{1}{\sqrt{1-y^2}}.$$

Analogamente temos as funções $\arccos : [-1,1] \to [0,\pi]$ com

$$\frac{d}{dy}\arccos y = \frac{-1}{\sqrt{1-y^2}},$$

e $\arctan : [-\infty, \infty] \to [-\pi/2, \pi/2]$, com

$$\frac{d}{dy}\arctan y = \frac{1}{\tan'(\arctan y)} = \frac{1}{\sec^2(\arctan y)} = \frac{1}{1+y^2},$$

pois $\sec^2 x = 1 + \tan^2 x$. Existem ainda as funções trigonométricas inversas para a $\cot x$, a sec x e a $\csc x$, sobre as quais não nos extenderemos.

7.2. Funções log e exponencial

Duas funções que têm importância fundamenal na matemática são dadas pelo logaritmo e sua inversa, a função exponencial. Há formas diversas de definirmos estas funções, e escolhemos aquela que nos parece mais direta.

7.2.1. O logaritmo. Definimos, para $x > 0$,

$$\ln x = \int_1^x \frac{1}{s}\, ds.$$

Observamos diretamente da definição que se $x > 1$, então $\ln x > 0$ e que se $x \in (0, 1)$, então

$$\ln x = \int_1^x \frac{1}{s}\, ds = -\int_x^1 \frac{1}{s}\, ds < 0.$$

Segue ainda de definição que $\ln 1 = 0$. Temos ainda os seguintes resultados:

(1) $\ln x$ é crescente
(2) $\ln x$ é contínua
(3) $\ln' x = 1/x$
(4) $\ln(xy) = \ln x + \ln y$
(5) $\ln x^r = r \ln x$, para $r \in \mathbb{Q}$
(6) $\lim_{x\to\infty} \ln x = \infty$
(7) $\lim_{x\to 0} \ln x = -\infty$

As demonstrações dos resultados acima não são complicadas. O resultado (1) vem do fato que se $x > y{>}0$, então $\ln x - \ln y = \int_y^x 1/s\, ds > 0$. Os fatos dados por (2) e (3) são resultados diretos da definição do ln e as propriedades das integrais. Para demonstrar (4), definimos $f(x) = \ln(xy)$, e portanto $f'(x) = 1/x = \ln' x$. Logo $f(x) - \ln x$ é constante. Como $f(1) - \ln 1 = \ln y$, obtemos o resultado. A identidade (5) é verdadeira para $r = 0$ pois $\ln 1 = 0$. Para r natural, aplicamos (4) $r - 1$ vezes, pois $x^r = x \cdots x$. Se $r = 1/n$, então usamos (4) novamente com $x = x^{1/n} \cdots x^{1/n}$. O caso geral para racionais positivos vem de $x^{m/n} = x^m x^{1/n}$. Para expoentes negativos, note que se $r > 0$ por exemplo, $0 = \ln(x^r x^{-r}) = \ln x^r + \ln x^{-r}$. Portanto, $\ln x^{-r} = -\ln x^r = -r \ln x$.

Observação. Vale a pena ressaltar que (5) vale também para qualquer r real, mas esta afirmativa esbarra no fato de que ainda não temos uma definição para x^r, quando r não é racional. Este lapso será resolvido somente em (7.2.1).

Finalmente, para (6) (7), basta usar que o logarímo é função crescente e considerar as sequências $\ln 2^n$ e $\ln 2^{-n}$.

Podemo definir a logarítmo "na base b" via

$$\log_b(x) = \frac{\ln(x)}{\ln(b)},$$

para $b \neq 1$ positivo.

7.2.2. A exponencial. Como a função ln é estritamente crescente, ela é invertível. Denominando esta inversa por $\exp x$, onde $\exp : \mathbb{R} \to (0, +\infty)$, note que

(1) $\exp(0) = 1$
(2) $x > 0 \implies \exp x > 1$
(3) $x < 0 \implies \exp x < 1$
(4) $\exp x$ é contínua
(5) $\exp x$ é diferenciável e $\exp' x = \exp x$
(6) $\exp(x + y) = \exp x \cdot \exp y$
(7) $\exp(\alpha x) = (\exp x)^\alpha$, para $\alpha \in \mathbb{R}$

A demonstração da fórmula em (5) é dada por

$$\exp' x = \frac{1}{\ln'(\exp x)} = \exp x.$$

Com a ajuda das funções acima descritas, podemos definir

$$a^b = \exp(b \ln a) \text{ para } a > 0 \text{ e } b \in \mathbb{R}. \tag{7.2.1}$$

Definimos o número especial $e = \exp 1$. Note então de (7) que

$$\exp \alpha = \exp(\alpha 1) = (\exp 1)^\alpha = e^\alpha \text{ para } \alpha \in \mathbb{R}.$$

CAPíTULO 8

Integração

[1]

Sem entrar em detalhes a respeito da definição de integral (Riemann), enunciamos algumas properiedades importantes. Consideraremos inicialmente (Seção 8.1) somente funções limitadas em intervalos limitados, mas não necessariam,ente contínuas e aplicações para determinar áreas Seção 8.2. Funções não limitadar e/ou intervalos limitados serão considerados na seção seguinte (Seção 8.3)

8.1. Propriedade básicas de integrais de funções limitadas

Primeiros veremos algumas propriedades fundamentais de funções diferenciáveis, e a seguir abordaremos a importante relação entre integrabilidade ew diferenciabilidade. Na última seção falaremos um pouco sobre técnicas que podem ajudar nos cálculos de algumas ingtegrais.

8.1.1. Algumas propriedades fundamentais. Considere abaixo $f : [a, b] \to \mathbb{R}$ e $g : [a, b] \to \mathbb{R}$ limitadas, $a < b$ números reais . As integrais serão sempre no domínio $[a, b]$. Temos então os seguintes resultados.

(1) Se f for contínua, então é diferenciável
(2) Se f for monótona (i.e., for função crescente ou decrescente), ela é integrável
(3) A integral da soma de funções é a soma das integrais. O mesmo vale para diferença e produto por escalar $\alpha \in \mathbb{R}$:

$$\int_a^b \alpha f(x) + g(x)\, dx = \alpha \int_a^b f(x)\, dx + \int_a^b g(x)\, dx,$$

(4) $\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx$ e $\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx$.
(5) Se f e g são integráveis, então o produto fg é integrável
(6) Existem funções não integráveis. Por exemplo:

$$f(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ -1 & \text{se } x \in \mathbb{R} \backslash \mathbb{Q} \end{cases}$$

(7) O exemplo do ítem 5 mostra que f^2 pode ser integrável, mesmo que f não o seja
(8) Se f é integrável, e $g(x) = f(x)$ a menos de um número finito de pontos, então g é integrável e

$$\int_a^b f(x)\, dx = \int_a^b g(x)\, dx$$

[1] Última Atualização: 06/07/2020

(9) Podemos sempre decompor uma função como a soma de suas partes positivas e negativas. Sejam

$$f^+(x) = \begin{cases} f(x) & \text{se } f(x) \geq 0 \\ 0 & \text{caso contrário} \end{cases}, \qquad f^-(x) = \begin{cases} 0 & \text{se } f(x) \geq 0 \\ -f(x) & \text{caso contrário} \end{cases},$$

Note que f^+ e f^- assumem somente valores positivos, e que, por construção, $f(x) = f^+(x) - f^-(x)$ e $|f(x)| = f^+(x) + f^-(x)$. Quanto a integrabilidade, se f for integrável, então f^+, f^- e $|f|$ são integráveis. Note que $|f|$ ser integrável não implica em f integrável, como nos mostra o exemplo apresentado no ítem 5.

(10) Se $f(x) \geq g(x)$ para todo $x \in (a, b)$ (ou a menos de um número finito de pontos), então

$$\int_a^b f(x)\,dx \geq \int_a^b g(x)\,dx$$

Se $f(x) > g(x)$ então a integral de f é estritamentente maior que a integral de g. Uma consequência imediata é que se $\alpha_0 \leq f(x) \leq \alpha_1$, onde α_0 e α_1 são números reais, então

$$\alpha_0(b - a) \leq \int_a^b f(x)\,dx \leq \alpha_1(b - a)$$

(11) O módulo da integral é menor ou igual a integral de módulo:

$$|\int_a^b f(x)\,dx| \leq \int_a^b |f(x)|\,dx$$

(12) Se f é função para e g é funçãoímpar, então

$$\int_{-a}^a f(x)\,dx = 2\int_0^a f(x)\,dx, \qquad \int_{-a}^a g(x)\,dx = 0.$$

Exemplo 8.1. Diga se é verdadeiro ou falso:

$$\int_{-2}^2 (ax^2 + bx + c)\,dx = 2\int_{-2}^2 (ax^2 + c)\,dx$$

Resposta: verdadeiro, pois

$$\int_{-2}^2 (ax^2 + bx + c)\,dx = \int_{-2}^2 ax^2\,dx + \int_{-2}^2 bx\,dx + \int_{-2}^2 c\,dx = 2\int_0^2 ax^2\,dx + 2\int_0^2 c\,dx$$

$$= 2\int_{-2}^2 (ax^2 + c)\,dx$$

8.1.2. Primitivas e o Teorema Fundamental do Cálculo. Provavelmente o resultado mais importante em se tratando de integrais é o *Teorema Fundamental do cálculo*. Suponha que $f : [a, b] \to \mathbb{R}$ seja integrável, e defina $F : [a, b] \to \mathbb{R}$ por

$$F(x) = \int_a^b f(s)\,ds \quad \text{para } x \in [a, b].$$

Então F é contínua em $[a, b]$. Além disto, se f for contínua em $c \in [a, b]$, então F é diferenciável em c e $F'(c) = f(c)$. Se f for contínua em todos os pontos de seu domínio, então F é chamada de *primitiva* da f.

A continuidade da f é essencial para a diferenciabilidade de F. Por exemplo, considere $f : [-1, 1] \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} 0 & \text{se } x < 0 \\ 1 & \text{se } x \geq 0 \end{cases}.$$

Então

$$F(x) = \int_{-1}^{x} f(x)\, dx = \begin{cases} 0 & \text{se } x < 0 \\ x & \text{se } x \geq 0 \end{cases}.$$

Note que F é contínua em $[-1, 1]$, mas não diferenciável em $x = 0$, ponto em que f é descontínua.

Veja também que duas primitivas de uma função diferem por uma constante. Seja $f : [a, b] \to \mathbb{R}$ e F e $\hat{F}$ suas primitivas. Então $F' = \hat{F}' = f$ em todos os pontos do domínio. Logo, $(F' - \hat{F}') = 0$ e portanto $F' = \hat{F}' + C$ para alguma constante C.

Denotamos a primitiva de f por

$$\int f(x)\, dx,$$

também chamada de *integral indefinida* da f.

TEOREMA 8.1.1 (Fundamental do Cálculo (parte 1)). *Seja f contínua. Então*

$$F(x) = \int_{a}^{x} f(s)\, ds$$

é diferenciável e $F'(x) = f(x)$.

TEOREMA 8.1.2 (Fundamental do Cálculo (parte 2)). *Seja f integrável. Se F é primitiva de f, então*

$$\int_{a}^{b} f(x)\, dx = \int_{a}^{b} F'(x)\, dx = F(b) - F(a).$$

DEMONSTRAÇÃO. Seja $g(x) = \int_a^x f(s)\, ds + F(a)$ pelo Teorema Fundamental do Cálculo (parte 1), temos que $g'(x) = f(x) = F'(x)$. Logo, $g(x) = F(x) + C$ para alguma constante C. Como $g(a) = F(a)$, então $C = 0$ e $g(x) = F(x)$. Logo,

$$F(b) = g(b) = \int_{a}^{b} f(s)\, ds + F(a).$$

□

É muito comum a notação

$$F\Big|_a^b = F\Big]_a^b = F(b) - F(a).$$

Há que se tomar cuidado com o uso de TFC. Por exemplo, se

$$g(x) = \int_{0}^{x^2} f(x)\, dx,$$

então falso que $g'(x) = f(x^2)$. A forma correta de se calcular g' é a seguinte. Seja $h(x) = x^2$ e $I(s) = \int_0^s f(x)\, dx$. Então $I'(s) = f(s)$ e como $g(x) = I(h(x))$, temos pela regra da cadeia

$$g'(x) = I'(h(x))h'(x) = f(h(x))2x = 2xf(x^2).$$

Considere agora o caso geral.

LEMA 8.1.3. Seja p, q funções diferenciáveis em toda reta, e

$$g(x) = \int_{p(x)}^{q(x)} f(t)\,dt.$$

Então $g'(x) = f(q(x))q'(x) - f(p(x))p'(x)$.

DEMONSTRAÇÃO. Note primeiro que

$$g(x) = \int_{p(x)}^{q(x)} f(t)\,dt = \int_{p(x)}^{0} f(t)\,dt + \int_{0}^{q(x)} f(t)\,dt = -\int_{0}^{p(x)} f(t)\,dt + \int_{0}^{q(x)} f(t)\,dt.$$

Seja $I(s) = \int_0^s f(t)\,dt$. Então $g(x) = -I(p(x)) + I(q(x))$, e portanto

$$g'(x) = -I'(p(x))p'(x) + I'(q(x))q'(x) = -f(p(x))p'(x) + f(q(x))q'(x).$$

□

Um resultado importante é dado pela fórmula do valor médio. O valor médio da integral de $f[a,b] \to \mathbb{R}$ é dado por

$$f_{\text{med}} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

TEOREMA 8.1.4 (Fórmula do valor médio para integrais). *Seja $f[a,b] \to \mathbb{R}$ contínua em $[a,b]$. Então existe $c \in [a,b]$ tal que*

$$f(c) = f_{med}.$$

DEMONSTRAÇÃO. Como f é contínua, então existe m (valor mínimo de f em $[a,b]$) e M (valor máximo de f em $[a,b]$), $m \le f(x) \le M$ para todo $x \in [a,b]$. Logo,

$$m \le d \le M, \qquad \text{onde } d = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Mas pelo Teorema do valor intermediário, existe $c \in (a,b)$ tal que $f(c) = d$. □

TEOREMA 8.1.5 (Fórmula de Taylor com resto integral). *Seja $n \ge 0$ e $I = [a,b]$, com $a < b$. Seja $f : I \to \mathbb{R}$ n vezes diferenciável em I com $f^{(n)}$ contínua em I e tal $f^{(n+1)}$ exista em (a,b). Se x_0, $x \in I$ então*

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2} + \cdots + f^{(n)}(x_0)\frac{(x-x_0)^n}{n!} + \frac{(x-x_0)^{n+1}}{(n+1)!}\int_0^1 (1-t)^n f^{(n+1)}(x_0 + t(x-x_0))\,dt.$$

Note que do Teorema acima obtemos a Fórmula de Taylor com resto de Lagrange se $f^{(n+1)}$ for contínua. De fato, usando o Fórmula do valor médio para integrais, existe $s \in (0,1)$ tal que

$$f^{(n+1)}(x_0 + s(x-x_0)) = \int_0^1 (1-t)^n f^{(n+1)}(x_0 + t(x-x_0))\,dt.$$

Denotando $\xi = x_0 + s(x-x_0)$, obtemos a fórmula (6.4.1).

8.1.3. Cálculo das integrais. Pode-se notar depois de algumas tentativas que achar primitivas ou calcular o valor de integrais não é tarefa fácil. Aqui falaremos de duas técnicas comumente usadas nesta tarefa. Consideraremos aqui somente funções integráveis.

8.1.3.1. *Integração por partes.* A primeira técnica, bem simples, é dada executando-se *integração por partes.* Sejam f e g diferenciáveis. Então

$$(fg)\Big|_a^b = \int_a^b \big(f(x)g(x)\big)'\,dx = \int_a^b f'(x)g(x)\,dx + \int_a^b f(x)g'(x)\,dx$$

Logo

$$\int_a^b f'(x)g(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f(x)g'(x)\,dx.$$

Um exemplo onde este truque pode ser usado é no cálculo de $\int_0^{2\pi} \sin^2 x\,dx$. Note que

$$\begin{aligned}\int_0^{2\pi} \sin^2 x\,dx &= \int_0^{2\pi} \sin x \sin x\,dx = -\int_0^{2\pi} \cos' x \sin x\,dx \\ &= -\cos x \sin x\Big|_0^{2\pi} + \int_0^{2\pi} \cos x \sin' x\,dx = \int_0^{2\pi} \cos^2 x\,dx = \int_0^{2\pi} 1 - \sin^2 x\,dx \\ &= 2\pi - \int_0^{2\pi} \sin^2 x\,dx.\end{aligned}$$

Passando $\int_0^{2\pi} \sin^2 x\,dx$ para o lado esquerdo temos que $\int_0^{2\pi} \sin^2 x\,dx = \pi$.

8.1.3.2. *Regra da substituição.* Outro resultado que é bastante útil é a mudança de variáveis no domínio de integração. Seja $f : [a,b] \to \mathbb{R}$ integrável, e $\phi : [c,d] \to [a,b]$ derivável, com $\phi(c) = a$ e $\phi(d) = b$, e tal que ϕ' seja integrável. Então

$$\int_c^d f(\phi(s))\phi'(s)\,ds = \int_{\phi(c)}^{\phi(b)} f(x)\,dx = \int_a^b f(x)\,dx. \tag{8.1.1}$$

A fórmula é prática pois pode ser mais simples integrar $f(x)$ do que $(\phi(s))\phi'(s)$. Para lembrar a fórmula basta definir $u(s) = \phi(s)$, e então $du/ds = \phi'(s)$, i.e., "$du = \phi'(s)ds$", e

$$\int_c^d f(\phi(s))\phi'(s)\,ds = \int_{u(c)}^{u(d)} f(u)\,du.$$

Em boa parte das aplicações, usamos ϕ invertível. Se denotarmos $u = \phi^{-1}$, temos que $\phi'(s) = 1/u'(\phi(s))$ e portanto, de (8.1.1), temos

$$\int_a^b f(x)\,dx = \int_c^d \frac{f(u^{-1}(s))}{u'(\phi(s))}\,ds = \int_{u(a)}^{u(b)} \frac{f(u^{-1}(s))}{u'(u^{-1}(s))}\,ds.$$

A forma de lembrar é usando $ds/dx = u'$ implica formalmente em $dx = ds/u'$.

Nos exemplos abaixo, vemos como usar estas identidades.

Exemplo 8.2. Para calcular $\int_0^{\pi} \frac{\sin x}{\cos^3 x}\,dx$, notamos que $\sin x = -\cos' x$ e usamos $u(x) = \cos x$. Então $du = -\sin x\,dx$ e

$$\int_0^{\pi} \frac{\sin x}{\cos^3 x}\,dx = -\int_{\cos 0}^{\cos \pi} \frac{1}{u^3}\,du = \frac{1}{2u^2}\Big|_1^{-1} = 0.$$

Na verdade, a integral acima poderia ser calculada diretamente observando-se por simetria que

$$\int_0^{\pi/2} \frac{\sin x}{\cos^3 x}\, dx = -\int_{\pi/2}^{\pi} \frac{\sin x}{\cos^3 x}\, dx$$

EXEMPLO 8.3. Para calcular $\int_0^{\pi/6} \cos^2(x)\sin^3(x)\, dx$, primeiro notamos que

$$\cos^2(x)\sin^3(x) = \cos^2(x)\sin^2(x)\sin(x) = \cos^2(x)(1-\cos^2(x))\sin(x).$$

Então definimos $u = \cos x$, e então $du = -\sin x dx$. Como $u(0) = \cos(0) = 1$ e $u(\pi/6) = \cos(\pi/6) = \sqrt{3}/2$, obtemos

$$\int_0^{\pi/6} \cos^2(x)\sin^3(x)\, dx = -\int_1^{\sqrt{3}/2} u^2(1-u^2)\, du = -\left(\frac{u^3}{3}-\frac{u^5}{5}\right)\Bigg]_1^{\sqrt{3}/2} = -\frac{3\sqrt{3}}{24}+\frac{9\sqrt{3}}{160}+\left(\frac{1}{3}-\frac{1}{5}\right)$$
$$= \frac{-11\sqrt{3}}{160} + \frac{2}{15}.$$

Para integrais indefinidas, não se esquecer de substituir novamente "o valor de u", como no exemplo abaixo.

EXEMPLO 8.4. Para calcular $\int \frac{\sin(\ln(x))}{x}\, dx$, usamos a substituição $u = \ln(x)$, e como $du = (1/x)dx$, obtemos

$$\int \frac{\sin(\ln(x))}{x}\, dx = \int \sin(u)\, du = -\cos(u) + C = -\cos(\ln(x)) + C.$$

8.2. Áreas planas

Como as integrais definidas dão a área (com sinal) sob determinadas curvas, nada mais natural que usar integrais para cálculo de áreas [**9**].

EXEMPLO 8.5. Por exemplo, para calcular a área entre a curva $f(x) = 2x$, os pontos $x = 0$ e $x = 3$, e o eixo dado por $y = 0$, há duas maneiras. Podemos usar a fórmula da área do triângulo (base $\times$ altura/2) e ver que $A = 3 \times 6/2 = 9$. Usando integrais,

$$A = \int_0^3 2x\, dx = x^2\Big|_0^3 = 9,$$

como era de se esperar.

É claro que nem todos cálculos de áreas são tão simples como o do exemplo acima. Podemos por exemplo, calcular áreas determinadas por curvas mais sofisticadas.

EXEMPLO 8.6. Para calcular a área A determinada pela curva $f(x) = \sin x$ e os pontos $x = \pi$ e $x = 2\pi$ e o eixo dado por $y = 0$, basta ver que

$$\int_\pi^{2\pi} \sin x\, dx = -\cos x\Big|_\pi^{2\pi} = \cos\pi - \cos 2\pi = -2.$$

É claro que uma área não pode ser negativa. O que dá "errado" neste exemplo é que a função sin é sempre negativa entre π e 2π. A área determinada então é simplesmente o negativo da integral, i.e., $A = 2$.

Um cuidado extra tem que ser tomado se a função tomar valores positivos e negativos no intervalo de interesse. Por exmplo, no exemplo acima, para achar a área de sin entre $x = 0$ e $x = 2\pi$, não se pode simplesmente calcular

$$\int_{\pi}^{2\pi} \sin x\, dx = 0.$$

Tem que se dividir o domínio que se quer integrar nas partes onde a função é positiva e onde é negativa. A área é dada na verdade por

$$A = \int_0^{\pi} \sin x\, dx - \int_{\pi}^{2\pi} \sin x\, dx = 4.$$

Observação. A área dada pela integral sem considerar os sinais é chamada às vezes de *área algébrica*, que pode ser negativa ou nula. A área que é sempre postiva, como as determinadas acima, é por vezes chamada *área geométrica*.

Outro problema mais interessante relacionado a áreas é o de *áreas entre curvas* (na verdade, os exemplos acima já são deste tipo, mas uma das curvas é dada por $y = 0$). Considere as funções reais f e g, definidas em $\mathbb{R}$. Pode-se perguntar qual é a área entre as curvas $f(x)$, $g(x)$, $x = a$ e $x = b$. Neste caso, se $f(x) \geq g(x)$ entre a e b, então a área é dada por

$$A = \int_a^b f(x) - g(x)\, dx.$$

Se f for maior que g apenas em parte do domínio, a integral tem que ser "quebrada" em partes para que não surjam "áreas negativas". Uma forma geral de se escrever a área é

$$A = \int_a^b |f(x) - g(x)|\, dx.$$

Neste caso, não precisamos nos preocupar se $f(x) \geq g(x)$ ou não.

Exemplo 8.7. Seja $f(x) = 2x + 3$, e $g(x) = x^2$. Determine a área compreendida entre f e g e entre $x = 1$ e $x = 3$. Note que em $[1, 3]$, temos $f(x) \geq g(x)$, e portanto podemos integrar $f - g$ para calcular a área:

$$A = \int_1^3 (2x + 3 - x^2)\, dx = \left(x^2 + 3x - \frac{x^3}{3} \right) \Bigg|_1^3 = 9 + 9 - 9 - 1 - 3 + \frac{1}{3} = \frac{16}{3}.$$

Exemplo 8.8. Considere a seguinte questão Anpec 2016, que envolvia calcular a área entre as curvas $y = -x^2 + 6$ e $y = x$. Para resolver temos que achar os pontos de interseção entre as curvas, i.e., resolver $-x^2 + 6 = x$, i.e., $x \in \{2, -3\}$. Note que no intervalo $(-3, 2)$ temos $-x^2 + 6 > x$ (para ver isto, basta calcular o valor das duas funçõesnum ponto; $x = 0$ por exemplo). A área é então dada por

$$\begin{aligned} \int_{-3}^{2} -x^2 + 6 - x\, dx = (-\frac{x^3}{3} - \frac{x^2}{2} + 6x)]_{-3}^{2} &= (-\frac{8}{3} - \frac{8}{2} + 12) - (9 - \frac{9}{2} - 18) \\ &= -\frac{8}{3} + 8 + 9 + \frac{9}{2} = -\frac{16}{6} + \frac{102}{6} + \frac{27}{6} = \frac{103}{6}. \end{aligned}$$

8.3. Integrais impróprias

Integrais impróprias são integrais de funções ilimitadas, ou em domínios ilimitados, e seus valores são dados através de limites, se estes existirem. Nestes casos, dizemos que as integrais *existem*, ou *convergem*.

Em domínios ilimitados, as integrais podem ser

$$\int_a^\infty f\,dx = \lim_{b\to\infty}\int_a^b f\,dx, \qquad \int_{-\infty}^b f\,dx = \lim_{a\to-\infty}\int_a^b f\,dx,$$

$$\int_{-\infty}^\infty f\,dx = \int_{-\infty}^0 f\,dx + \int_0^\infty f\,dx.$$

É importante atentar para um detalhe nas integrais em $(-\infty,\infty)$. Para esta existir, tem que existir o limite $\lim_{b\to\infty}\int_0^b f\,dx$ e o limite $\lim_{a\to-\infty}\int_a^0 f\,dx$, *separadamente*. Note que isto é diferente de escever

$$\int_{-\infty}^\infty f\,dx = \lim_{a\to\infty}\int_{-a}^a f\,dx.$$

Observação (Isto não cai na ANPEC). O valor acima é conhecido como *valor principal de Cauchy*, usado em alguns ramos da matemática. Para ver a diferença entre as duas definições, considere a função sinal de x dada por $f(x) = \operatorname{sgn}(x)$. Isto é, $f(x) = 1$ para números positivos e $f(x) = -1$ para números negativos. Então f não é integrável em $(-\infty,\infty)$ pois não existe $\lim_{b\to\infty}\int_0^b f\,dx = \lim_{b\to\infty} b$ nem $\lim_{a\to-\infty}\int_a^0 f\,dx = -\lim_{a\to-\infty} a$. Mas o valor principal de Cauchy está bem definido pois $\int_{-a}^a f\,dx = 0$ para todo a e então

$$\int_{-\infty}^\infty f\,dx = \lim_{a\to\infty}\int_{-a}^a f\,dx = 0.$$

¤

Integrais de funções com assíntotas verticais são definidas de forma análoga. Seja $f : (a,b) \to \mathbb{R}$ e suponha f integrável em $[a+\delta, b]$ para todo $\delta > 0$. Definimos então

$$\int_a^b f\,dx = \lim_{\delta\to 0}\int_{a+\delta}^b f\,dx$$

quando este limite existir. Quando a função é ilimitada numa vizinhança de b, a definição é análoga:

$$\int_a^b f\,dx = \lim_{\delta\to 0}\int_a^{b-\delta} f\,dx$$

O último caso é quando a singularidade fica no interior do intervalo. Por exemplo, seja $c \in (a,b)$ tal que f seja integrável em $(a, c-\delta)$ e em $(c+\delta, b)$. Definimos então

$$\int_a^b f\,dx = \int_a^c f\,dx + \int_c^b f\,dx = \lim_{\delta\to 0^+}\int_a^{c-\delta} f\,dx + \lim_{\delta\to 0^+}\int_{c+\delta}^b f\,dx.$$

Novamente, no caso acima, os dois limites têm que existir. Por exemplo, a função dada em $[-1,1]$ por $f(0) = 0$ e, se $x \neq 0$ por $f(x) = \operatorname{sgn}(x)/x$ não é integrável, apesar de termos

$$\lim_{\delta\to 0}\left(\int_{-1}^{-\delta} \frac{-1}{x}\,dx + \int_\delta^1 \frac{1}{x}\,dx\right) = 0.$$

Exemplo 8.9. A integral imprópria $\int_0^1 1/\sqrt{x}\,dx$ está bem definida pois

$$\lim_{\delta\to 0}\int_\delta^1 \frac{1}{\sqrt{x}}\,dx = \lim_{\delta\to 0} 2\sqrt{x}\Big|_\delta^1 = 2.$$

Exemplo 8.10. As integrais impróprias $\int_0^1 1/x\,dx$ e $\int_1^\infty 1/x\,dx$ não existem pois

$$\lim_{\delta\to 0}\int_\delta^1 \frac{1}{x}\,dx = \lim_{\delta\to 0}\log x\Big|_\delta^1 = +\infty, \qquad \lim_{b\to\infty}\int_1^b \frac{1}{x}\,dx = \lim_{b\to\infty}\log x\Big|_1^b = +\infty.$$

Vemos neste exemplo que $\lim_{x\to\infty} f(x) = 0$ não garante que f seja integrável.

Exemplo 8.11. Considere agora a função $f(x) = 1/x^p$. Veremos que o caso $p = 1$ é o "caso limite". Vamos calcular

$$\text{a)} \int_0^1 x^p\,dx, \qquad \text{b)} \int_1^\infty x^p\,dx.$$

Note que a primitiva de x^{-p} é dada por $\ln x$ se $p = 1$, e $x^{1-p}/(1-p)$ se $p \neq 1$, e portanto

$$\int_a^b \frac{1}{x^p}\,dx = \begin{cases} \ln(b/a) & \text{se } p = 1, \\ \frac{b^{1-p}-a^{1-p}}{1-p} & \text{se } p \neq 1. \end{cases}$$

Primeiro, considerando integrais entre 0 e 1, temos que

$$\int_0^1 \frac{1}{x^p}\,dx = \lim_{a\to 0}\int_a^1 \frac{1}{x^p}\,dx = \lim_{a\to 0}\begin{cases} \ln(b/a) & \text{se } p = 1 \\ \frac{1-a^{1-p}}{1-p} & \text{se } p \neq 1 \end{cases} = \begin{cases} +\infty & \text{se } p \geq 1, \\ \frac{1}{1-p} & \text{se } p < 1. \end{cases}$$

Para integrais no intervalo $(1, +\infty)$ temos

$$\int_1^{+\infty} \frac{1}{x^p}\,dx = \lim_{b\to\infty}\int_1^b \frac{1}{x^p}\,dx = \lim_{b\to\infty}\begin{cases} \ln(b) & \text{se } p = 1 \\ \frac{b^{1-p}-1}{1-p} & \text{se } p \neq 1 \end{cases} = \begin{cases} +\infty & \text{se } p \leq 1, \\ \frac{1}{p-1} & \text{se } p > 1. \end{cases}$$

Exemplo 8.12. Calcule $\int_1^3 \frac{1}{\sqrt{x-1}}\,dx$.

Solução: observe que a integral é imprópria pois o integrando tem uma assíntota vertical em $x = 1$. Usando a definição, temos

$$\int_1^3 \frac{1}{\sqrt{x-1}}\,dx = \lim_{a\to 1^+}\int_a^3 \frac{1}{\sqrt{x-1}}\,dx.$$

como

$$\int_a^3 \frac{1}{\sqrt{x-1}}\,dx = 2(x-1)^{1/2}|_a^3 = 2(\sqrt{2} - \sqrt{a-1}),$$

temos

$$\int_1^3 \frac{1}{\sqrt{x-1}}\,dx = \lim_{a\to 1^+}\int_a^3 \frac{1}{\sqrt{x-1}}\,dx = \lim_{a\to 1^+} 2(\sqrt{2} - \sqrt{a-1}) = 2\sqrt{2}.$$

Integrais impróprias envolvendo assíntotas verticas *têm* que ser calculadas tomando limites, *mesmo que pareça desnecessário*. Considere o exemplo abaixo.

Exemplo 8.13. Calcule $\int_0^2 \frac{1}{x-1}\,dx$, se possível.

Solução: Note que comoo existe uma assíntota vertical em $x = 1$, temos que calcular

$$\int_0^2 \frac{1}{x-1}\,dx = \int_0^1 \frac{1}{x-1}\,dx + \int_1^2 \frac{1}{x-1}\,dx,$$

e a integral à esquerda só existirá se *ambas* integrais à direita forem convergentes. Note entretanto que

$$\int_0^1 \frac{1}{x-1}\,dx = \lim_{a\to 1^-}\int_0^a \frac{1}{x-1}\,dx = \lim_{a\to 1^-}(\ln|x-1|)\Big|_0^a = \lim_{a\to 1^-}(\ln|a-1| - \ln 1) = -\infty.$$

Logo, a integral não converge, e não é nem necessário calcular $\int_1^2 \frac{1}{x-1}\,dx$.

Note entretanto que se tentássemos calcular $\int_0^2 \frac{1}{x-1}\,dx$ diretamente, o resultado seria errado:

$$\int_0^2 \frac{1}{x-1}\,dx = \ln|x-1|)\Big|_0^2 = \ln 1 - \ln 1 = 0.$$

A exemplo de séries, dizemos que uma integral imprópria de uma função f *converge absolutamente* se $|f|$ for integrável. Temos então o seguinte resultado.

Teorema 8.3.1. *Suponha f contínua. Se a integral de f converge absolutamente, então a integral de f converge. Em outras palavras, $|f|$ integrável implica em f integrável.*

Um resultado importante é dado pelo Teorema da Comparação abaixo. O resultado pode auxiliar na decisão de quando uma integral imprópria converge ou diverge, mesmo sem o cálculo de seu valor.

Teorema 8.3.2 (Teorema da Comparação). *Sejam $f \ge 0$ e $g \ge 0$, com $f \le g$ no domínio de integração. Então, $\int g$ convergente implica em $\int f$ convergente. e $\int f$ divergente implica em $\int g$ divergente.*

Exemplo 8.14. Decida se $\int_0^\infty e^{-x^2}\,dx$ é ou não integrável.

Solução: Note que $\int_0^\infty e^{-x^2}\,dx = \int_0^1 e^{-x^2}\,dx + \int_1^\infty e^{-x^2}\,dx$. Como e^{-x^2} é contínua, basta analisar a integral entre 1 e ∞. Neste caso, como $x \ge 1$ implica em $x^2 \ge x$, temos $e^{-x^2} \le e^{-x}$, e

$$0 \le \int_1^\infty e^{-x^2}\,dx \le \int_1^\infty e^{-x}\,dx$$

Mas $\int_1^\infty e^{-x}\,dx$ é convergente pois

$$\lim_{a\to+\infty}\int_1^a e^{-x}\,dx = -\lim_{a\to+\infty} e^{-x}|_1^a = -\lim_{a\to+\infty}(e^{-a} - \frac{1}{e}) = \frac{1}{e}.$$

Então, como $\int_1^\infty e^{-x}\,dx$ é convergente, então $\int_0^\infty e^{-x^2}\,dx$ também o é.

Exemplo 8.15. A integral $\int_0^\infty (1+\sin^2 x)x^{-1/2}\,dx$ é divergente pelo Teorema da Comparação, pois $\int_0^\infty x^{-1/2}\,dx$ diverge.

CAPíTULO 9

Sequências e Séries

[1] Neste capítulo veremos sequências e séries. Como série nada mais é que um caso particular de sequências, veremos os dois tópicos de forma unificada, chamando atenção para as possíveis diferenças.

9.1. Definição e resultados preliminares

Uma sequência em $\mathbb{R}$ é simplesmente uma função de $\mathbb{N}$ em $\mathbb{R}$. Portanto $X : \mathbb{N} \to \mathbb{R}$ indica uma sequência de números reais, que escrevemos também como (x_n), ou ainda $(x_1, x_2, x_3, \dots)$. Para indicar o n-ésimo valor da sequência escrevemos simplesmente x_n.

Sejam $c_1, c_2 \dots$ números reais. Uma série

$$\sum_{i=1}^{\infty} c_i \tag{9.1.1}$$

pode ser compreendida através da sequência s_n definida por

$$s_n = \sum_{i=1}^{n} c_i. \tag{9.1.2}$$

Em geral, a expressão (9.1.1) nem sempre faz sentido, enquanto (9.1.2) está sempre bem-definida.

Exemplo 9.1. $x_n = (-1)^n$ define a sequência $(-1, 1 - 1, 1, -1, 1, -1, \dots)$.

Exemplo 9.2. A sequência de Fibonacci é definida recursivamente por $x_1 = 1$, $x_2 = 1$, e $x_{n+1} = x_n + x_{n-1}$ para $n \geq 2$. Portanto temos $(x_n) = (1, 1, 2, 3, 5, 8, \dots)$.

Podemos realizar com sequências várias das operações que realizamos com números reais, como por exemplo somar, subtrair, etc. Sejam por exemplo (x_n) e (y_n) duas sequências em $\mathbb{R}$, e $c \in \mathbb{R}$. Então definimos

$$(x_n)+(y_n) = (x_n+y_n), \qquad (x_n)-(y_n) = (x_n-y_n), \qquad (x_n)\cdot(y_n) = (x_n y_n), \qquad c(x_n) = (cx_n).$$

Exemplo 9.3. Se $x_n = (2, 4, 6, 8, \dots)$ e $(y_n) = (1, 1/2, 1/3, 1/4, \dots)$, então $x_n \cdot y_n = (2, 2, 2, \cdots)$.

A primeira pergunta que surge quando tratamos de sequências é quanto à convergência destas, isto é, se quando n aumenta, os termos x_n se aproximam de algum valor real. Note que para isto, não importa o que acontece com finitos termos da sequência, mas sim seu comportamento assintótico com respeito a n. Em outras palavras queremos determinar o comportamento das sequências no "limite".

[1] Última Atualização: 27/07/2020

Definição 9.1.1. *Dizemos que* $x \in \mathbb{R}$ *é* limite de uma sequência (x_n), *se para todo* $\epsilon > 0$, *existe* $N \in \mathbb{N}$ *tal que* $|x - x_n| < \epsilon$ *para todo* $n \geq N$. *Escrevemos neste caso que* $x_n \to x$, *ou que* $x = \lim x_n$, *ou ainda*

$$x = \lim_{n \to \infty} x_n.$$

De forma resumida, $x_n \to x$ *se para todo* ϵ *existir* $N \in \mathbb{N}$ *tal que*

$$n \geq N \implies |x - x_n| < \epsilon.$$

Se uma sequência não tem limite, dizemos que ela diverge ou é divergente.

A definição para séries é análoga.

Definição 9.1.2. *Dizemos que* $s \in \mathbb{R}$ *é* limite de uma série $\sum_{i=1}^{\infty} c_n$, *se para todo* $\epsilon > 0$, *existe* $N \in \mathbb{N}$ *tal que*

$$\left| s - \sum_{i=1}^{n} c_n \right| < \epsilon \quad \textit{para todo } n \geq N.$$

Escrevemos neste caso que que $x = \sum_{i=1}^{\infty} c_n$. *Se uma série não tem limite, dizemos que ela diverge ou é divergente.*

Exemplo 9.4. Se $x_n = 1$, então $\lim x_n = 1$. De fato, dado $\epsilon > 0$, para todo $n \geq 1$ temos $|x_n - 1| = 0 < \epsilon$.

Exemplo 9.5. $\lim(1/n) = 0$. De fato, dado $\epsilon > 0$, seja N tal que $1/N < \epsilon$. Logo, para todo $n > N$ temos $|1/n - 0| = 1/n < 1/N < \epsilon$.

Exemplo 9.6. $(0, 2, 0, 2, 0, 2, 0, 2, \dots)$ não converge para 0. De fato, tome $\epsilon = 1$. Então para todo $N \in \mathbb{N}$ temos $2N > N$ e $x_{2N} = 2$. Portanto $|x_{2N} - 0| = 2 > \epsilon$.

Observe que diferentes situações ocorrem nos exemplos acima. No primeiro, a sequência é constante, e a escolha de N independe de ϵ. Já no exemplo seguinte, N claramente depende de ϵ.

A seguir, no exemplo 9.6 o objetivo é mostar que um certo valor x *não* é o limite da sequência (x_n). Mostramos então que existe pelo menos um certo $\epsilon > 0$ tal que para todo N, conseguimos achar $n > N$ tal que $|x_n - x| > \epsilon$. Note que o que fizemos foi *negar* a convergência.

Talvez a segunda pergunta mais natural em relação aos limites de sequências é quanto a unicidade destes, quando existirem. A resposta é afirmativa, como mostra o resultado abaixo.

Teorema 9.1.3 (Unicidade de limite). *Uma sequência (uma série) pode ter no máximo um limite.*

Demonstração. Considere que (x_n) é uma sequência de reais tal que $x_n \to x$ e $x_n \to x'$, com $x \neq x'$. Sejam $\epsilon = |x - x'|/2 > 0$, e sejam N e $N' \in \mathbb{N}$ tais que $|x_n - x| < \epsilon$ para todo $n > N$ e $|x_n - x'| < \epsilon$ para todo $n > N'$. Logo, se $n > \max\{N, N'\}$, então

$$|x - x'| \leq |x - x_n| + |x_n - x'| < 2\epsilon = |x - x'|.$$

Como um número não pode ser estritamente menor que ele mesmo, temos uma contradição. Portanto $x = x'$ e o limite é único. ∎

Para mostrar convergência, podemos usar o resultado seguinte.

TEOREMA 9.1.4. *Seja* (x_n) *uma sequência em* $\mathbb{R}$. *Então as afirmativas são equivalentes.*

(1) (x_n) *converge para* x.
(2) Para toda vizinhança V *de* x *existe* $N \in \mathbb{N}$ *tal que*

$$n \geq N \implies x_N \in V.$$

DEMONSTRAÇÃO. Fica como exercício. □

As vezes, uma sequência se aproxima de algum valor de forma mais lenta que alguma outra sequência que converge para 0. É possível assim garantir convêrgencia, como o resultado a seguir nos mostra.

LEMA 9.1.5. Seja (a_n) sequência em $\mathbb{R}$ convergente para 0. Se para (x_n) sequência em $\mathbb{R}$ existir $c > 0$ tal que

$$|x_n - x| \leq c|a_n| \quad \text{para todo } n \in \mathbb{N},$$

então $x_n \to x$.

DEMONSTRAÇÃO. Como (a_n) converge, dado $\epsilon > 0$, seja $N \in \mathbb{N}$ tal que $|a_n| < \epsilon/c$ para todo $n > N$. Logo

$$|x_n - x| \leq c|a_n| < \epsilon \quad \text{para todo } n > N,$$

e $\lim x_n = x$. □

COROLÁRIO 9.1.6. Seja (a_n) sequência em $\mathbb{R}$ convergente para 0. Se para (x_n) sequência em $\mathbb{R}$ existir $c > 0$ e $N \in \mathbb{N}$ tal que

$$|x_n - x| \leq c|a_n| \quad \text{para todo } n \geq N,$$

então $x_n \to x$.

EXEMPLO 9.7. Seja $x_n = (2/n)\sin(1/n)$. Enão

$$|x_n - 0| \leq 2\frac{1}{n}.$$

Como $1/n \to 0$, podemos usar o lema acima para garantir que $\lim[(2/n)\sin(1/n)] = 0$.

Uma outra noção importante é o de limitação de uma sequência. Neste caso, mesmo quando a sequência não converge, podemos conseguir alguns resultados parciais, como veremos mais a seguir.

DEFINIÇÃO 9.1.7. *Dizemos que uma sequência* (x_n) *é limitada quando existe um número real* M *tal que* $|x_n| \leq M$ *para todo* $n \in \mathbb{N}$.

Um primeiro resultado intuitivo é que toda sequência convergente é limitada. De fato, é razoável pensar que se a sequência converge, ela não pode ter elementos arbitrariamente grandes em valor absoluto.

TEOREMA 9.1.8. *Toda sequência convergente é limitada*

DEMONSTRAÇÃO. Seja (x_n) sequência convergente e seja x seu limite. Seja $\epsilon = 1$. Como (x_n) converge, existe N tal que $|x - x_n| < 1$ para todo $n > N$. Logo, usando a desigualdade triangular temos

$$|x_n| \leq |x_n - x| + |x| < 1 + |x| \quad \text{para todo } n > N.$$

Falta agora limitar os N primeiros termos da sequência. Seja então

$$M = \max\{|x_1|, |x_2|, |x_3|, \dots, |x_N|, 1 + |x|\}.$$

Portanto $|x_n| \leq M$ para todo $n \in \mathbb{N}$. □

Existem situações em que a sequência não converge, mas *tendem* para $+\infty$ ou $-\infty$. Por exemplo, as sequencias $(1, 0, 9, 0, 25, 0, 36, \dots)$ e (n^2) não convergem pois não são limitadas. Entretanto, para a sequência dada por $x_n = n^2$, para todo número $C > 0$ existe $N \in \mathbb{N}$ tal que

$$n > N \implies x_n > M.$$

quando isto acontece, dizemos que $x_n \to +\infty$ ou que $\lim x_n = +\infty$. Definição análoga vale para $-\infty$.

Outro resultado importante trata de limites de sequências que são resultados de operações entre sequências. Por exemplo, dadas duas sequências convergente, o limite da soma das sequências é a soma dos limites. E assim por diante.

LEMA 9.1.9. Seja (x_n) e (y_n) tais que $\lim x_n = x$ e $\lim y_n = y$. Então

(1) $\lim(x_n + y_n) = x + y$.
(2) $\lim(x_n - y_n) = x - y$.
(3) $\lim(x_n y_n) = xy$.
(4) $\lim(cx_n) = cx$, para $c \in \mathbb{R}$.
(5) se $y_n \neq 0$ para todo n e $y \neq 0$, então $\lim(x_n/y_n) = x/y$.

DEMONSTRAÇÃO. (1) Dado $\epsilon > 0$, seja $N \in \mathbb{N}$ tal que $|x_n - x| < \epsilon/2$ e $|y_n - y| < \epsilon/2$ para todo $n \geq N$. Logo

$$|x_n + y_n - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon \quad \text{para todo } n \geq N.$$

(2) A demonstração é basicamente a mesma de (1), tomando-se o devido cuidado com os sinais.

(3) Para todo $n \in \mathbb{N}$ temos

$$|x_n y_n - xy| \leq |x_n y_n - x_n y| + |x_n y - xy| = |x_n||y_n - y| + |y||x_n - x|.$$

Seja $M \in \mathbb{R}$ tal que $|x_n| < M$ e $|y| < M$. Tal constante M existe pois como (x_n) converge, ela é limitada. Agora, dado $\epsilon > 0$, seja N tal que $|y_n - y| < \epsilon/(2M)$ e $|x_n - x| < \epsilon/(2M)$ para todo $n \geq N$. Logo,

$$|x_n y_n - xy| \leq M[|y_n - y| + |x_n - x|] < \epsilon,$$

para todo $n \geq N$.

Deixamos (4) e (5) como exercícios para o leitor. □

OBSERVAÇÃO. Os resultados do lema acima continuam válidos para um número finito de somas, produtos, etc.

Outro resultado interessante nos diz que se a_n é sequência de números positivos, então $x_n \to 0$ se e somente se $1/x_n \to \infty$ (resultado análogo vale trocando-se $+\infty$ por $-\infty$). Ver exercício 9.3.

Outros resultados importantes para tentar achar um "candidato" limite vêm a seguir. O primeiro nos diz que se temos uma sequência de números positivos, então o limite, se existir, tem que ser não negativo, podendo ser zero. A seguir, aprendemos que se temos uma sequência "sanduichadas" entre outras duas sequências convergentes que tem o mesmo limite, então a sequência do meio converge e tem também o mesmo limite.

Lema 9.1.10. Seja (x_n) convergente com $\lim x_n = x$. Se existe $N \in \mathbb{N}$ tal que $x_n \geq 0$ para todo $n > N$, então $x \geq 0$.

Demonstração. (por contradição) Assuma que $x < 0$. Seja então $\epsilon = -x/2 > 0$. Como (x_n) converge para x, seja $N \in \mathbb{N}$ tal que $|x_n - x| < \epsilon$ para todo $n > N$. Logo, $x_{N+1} \in (x - \epsilon, x + \epsilon)$, isto é, $x_{N+1} < x + \epsilon = x/2 < 0$. Obtivemos então uma contradição pois x_{N+1} não é negativo. $\square$

Corolário 9.1.11. Se (x_n) e (x_n) são convergentes com $\lim x_n = x$ e $\lim y_n = y$, e se existe $N \in \mathbb{N}$ tal que $x_n \geq y_n$ para todo $n > N$, então $x \geq y$.

Demonstração. Se $z_n = x_n - y_n$, então $\lim z_n = \lim x_n - \lim y_n = x - y$. O presente resultado segue então do Lema 9.1.10. $\square$

Lema 9.1.12 (sanduíche de sequências). Sejam (x_n), (y_n) e (z_n) sequências tais que $x_n \leq y_n \leq z_n$ para todo $n > N$, para algum $N \in \mathbb{N}$. Assuma ainda que (x_n) e (z_n) convergem com $\lim x_n = \lim z_n$. Então (y_n) converge e $\lim y_n = \lim x_n = \lim z_n$.

Demonstração. Seja $a = \lim x_n = \lim z_n$. Dado $\epsilon > 0$, existe N tal que $|x_n - a| < \epsilon$ e $|z_n - a| < \epsilon$ para todo $n > N$. Logo

$$-\epsilon < x_n - a \leq y_n - a \leq z_n - a < \epsilon \implies |x_n - a| < \epsilon$$

para todo $n > N$, como queríamos demonstrar. $\square$

Exemplo 9.8. (n) diverge pois não é limitada.

Exemplo 9.9. Seja $S_n = 1 + 1/2 + 1/3 + 1/4 + \cdots + 1/n$. Mostraremos que (S_n) não é limitada, e portanto divergente. Note que

$$\begin{aligned} x_{2^n} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^{n-1}+1} + \frac{1}{2^n}\right) \\ &= 1 + \frac{1}{2} + \sum_{i=3}^{4} \frac{1}{n} + \sum_{i=5}^{8} \frac{1}{n} + \cdots + \sum_{i=2^{n-1}+1}^{2^n} \frac{1}{n} > 1 + \frac{1}{2} + \sum_{i=3}^{4} \frac{1}{4} + \sum_{i=5}^{8} \frac{1}{8} + \cdots + \sum_{i=2^{n-1}+1}^{2^n} \frac{1}{2^n} \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots + \frac{1}{2} = 1 + \frac{n}{2}. \end{aligned}$$

Logo (S_n) não é limitada, e portanto diverge.

Outra forma de ver que a sequência acima diverge é por indução. Quero mostrar que $S_{2^n} \geq 1 + n/2$. Note que $S_2 = 1 + 1/2$. Assumindo que $S_{2^{n-1}} \geq 1 + (n-1)/2$ temos

$$S_{2^n} = S_{2^{n-1}} + \frac{1}{2^{n-1}+1} + \cdots + \frac{1}{2^n} > 1 + \frac{(n-1)}{2} + \frac{1}{2} > 1 + \frac{n}{2},$$

como queríamos demonstrar. Mais uma vez a conclusão é que (S_n) não é limitada, logo diverge.

EXEMPLO 9.10. $\lim_{n\to\infty}\big((2n+1)/n\big) = 2$. De fato,
$$\frac{2n+1}{n} = (2) + \big(\frac{1}{n}\big).$$
Como $\lim_{n\to\infty}(2) = 2$ e $\lim_{n\to\infty}(1/n) = 0$, nós obtemos o resultado.

EXEMPLO 9.11. $\lim_{n\to\infty}\big(2n/(n^2+1)\big) = 0$, pois
$$\frac{2n}{n^2+1} = \frac{2/n}{1+1/n^2}.$$
Como $\lim_{n\to\infty}(2/n) = 0$ e $\lim_{n\to\infty}(1+1/n^2) = 1 \neq 0$, podemos aplicar o resultado sobre quociente de sequências.

EXEMPLO 9.12. A sequência
$$x_n = \frac{1}{n^2}\sum_{i=1}^{n} i$$
converge. Primeiro note que
$$\sum_{i=1}^{n} i = \frac{n^2+n}{2}. \tag{9.1.3}$$
Para $n = 1$ o resultado (9.1.3) é trivial. Assuma (9.1.3) vedadeiro para $n = k$. Temos então que
$$\sum_{i=1}^{k+1} i = \frac{k^2+k}{2} + k + 1 = \frac{k^2+3k+2}{2} = \frac{(k+1)^2+(k+1)}{2},$$
e portanto fórmula (9.1.3) é verdadeira. Temos então que
$$x_n = \frac{n^2+n}{2n^2} = \frac{1}{2}\left(1+\frac{1}{n}\right) = \frac{1}{2} + \left(\frac{1}{2n}\right).$$
Logo (x_n) é soma de duas sequências convergentes, $(1/2)$ e $(1/2)(1/n)$ e
$$\lim_{n\to\infty} x_n = \lim_{n\to\infty}\frac{1}{2} + \lim_{n\to\infty}\frac{1}{2n} = \frac{1}{2}.$$

EXEMPLO 9.13. Seja (x_n) sequência convergente em $\mathbb{R}$,e seja $x \in \mathbb{R}$ seu limite. Então a sequência definida por
$$\frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$
converge e tem x como seu limite.

Sem perda de generalidade, supomos que (x_n) converge para zero. Para o caso geral quando (x_n) converge para x basta tratar a sequência $(x_n - x)$.

Seja $S_n = (x_1 + x_2 + \cdots + x_n)/n$. Como (x_n) converge, então é limitada. Seja M tal que $|x_n| < M$ para todo $n \in \mathbb{N}$. Dado $\epsilon > 0$, seja N^* tal que $M/N^* < \epsilon$ e $\sup\{|x_n| : n \geq N^*\} < \epsilon$. Então, temos $S_n = \check{S}_n + \hat{S}_n$, onde
$$\check{S}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_{N^*}), \qquad \hat{S}_n = \frac{1}{n}(x_{N^*} + x_{N^*+1} + \cdots + x_n).$$

Então (S_n) é a soma de duas sequências convergentes. De fato para $n > (N^*)^2$, temos $|\check{S}_n| \leq N^*M/n \leq M/N^* < \epsilon$. Além disso, $|\hat{S}_n| < \epsilon(n - N^*)/n < \epsilon$. Portanto (S_n) converge.

Exemplo 9.14. $\lim_{n\to\infty}\big((\sin n)/n\big) = 0$ pois como $-1 \leq \sin n \leq 1$, então

$$-1/n \leq (\sin n)/n \leq 1/n,$$

e o resultado segue do lema 9.1.12.

Outro resultado importante refere-se à convergência do valor absoluto de sequências: se uma sequência converge, entao a sequência de valores absolutos também converge. A reciproca *não* é verdadeira. Basta considerar como contra-exemplo a sequência $\big((-1)^n\big)$. Neste caso a sequência diverge mas a sequência de seus valores absolutos converge.

Lema 9.1.13. Seja (x_n) convergente. Então $(|x_n|)$ também o é.

Demonstração. Exercício. □

Lema 9.1.14 (teste da razão). Seja (x_n) sequência de números positivos tal que (x_{n+1}/x_n) converge e $\lim_{n\to\infty}(x_{n+1}/x_n) < 1$. Então (x_n) converge e $\lim_{n\to\infty}(x_n) = 0$.

Demonstração. Seja $L = \lim_{n\to\infty}(x_{n+1}/x_n)$. Então, por hipótese, $L < 1$. Seja r tal que $L < r < 1$. Portanto dado $\epsilon = r - L > 0$, existe N tal que $x_{n+1}/x_n < L + \epsilon = r$ para todo $n \geq N$. Logo,

$$0 < x_{n+1} < x_n r < x_{n-1}r^2 < x_{n-2}r^3 < \cdots < x_N r^{n-N+1} \quad \text{para todo } n \geq N.$$

Se $c = x_N r^{-N}$. , então $0 < x_{n+1} < cr^{n+1}$. O resultado segue do Corolário 9.1.6, pois como $r < 1$, então $\lim_{n\to\infty} r^n = 0$. □

Corolário 9.1.15. Seja (x_n) tal que $x_n \neq 0$ para todo $n \in \mathbb{N}$ e

$$L = \lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|}$$

existe e $L > 1$. Então para todo $C \in \mathbb{R}$ existe $N^* \in \mathbb{N}$ tal que

$$n \geq N^* \implies |x_n| > C.$$

Demonstração. basta considerar o teste da razão para $y_n = 1/x_n$. Neste caso,

$$\lim_{n\to\infty} \frac{|y_{n+1}|}{|y_n|} = \lim_{n\to\infty} \frac{|x_n|}{|x_{n+1}|} = \lim_{n\to\infty} \frac{1}{\frac{|x_{n+1}|}{|x_n|}} = \frac{1}{\lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|}} = \frac{1}{L} < 1.$$

Logo (y_n) converge para zero, e para todo $C \in \mathbb{R}^+$ existe N^* tal que

$$n \geq N^* \implies |y_n| < \frac{1}{C}.$$

Portanto para $n \geq N^*$ temos $|x_n| > C$ e (x_n) não é limitada e não converge. □

Exemplo 9.15. Seja $(x_n) = n/2^n$. Então

$$\lim_{n\to\infty}\Big(\frac{x_{n+1}}{x_n}\Big) = \lim_{n\to\infty}\Big(\frac{n+1}{2^{n+1}}\frac{2^n}{n}\Big) = \frac{1}{2}\lim_{n\to\infty}\Big(\frac{n+1}{n}\Big) = \frac{1}{2}.$$

Pelo teste da razão temos $\lim_{n\to\infty}(x_n) = 0$

Exemplo 9.16. Note que para $x_n = 1/n$, temos $\lim_{n\to\infty} x_{n+1}/x_n = 1$ e (x_n) converge. Entretanto, para $x_n = n$, temos $\lim_{n\to\infty} x_{n+1}/x_n = 1$ mas (x_n) *não convergente*. Portanto o teste não é conclusivo quando o limite da razão entre os termos é um.

9.2. Subsequências e Teorema de Bolzano–Weierstrass

Seja (x_k) sequência em $\mathbb{R}$ e

$$k_1 < k_2 < k_3 < \cdots < k_j < \ldots$$

sequência de números naturais. Então dizemos que (x_{k_j}) é uma *subsequência* de (x_k).

Exemplo 9.17. Se $(x_k) = (1, 1/2, 1/3, 1/4, \ldots)$, então $(1, 1/2, 1/4, 1/6, 1/8, \ldots)$ e (x_{2^k}) são subsequências de (x_k).

Um primeiro resultado relacionado com subsequências nos diz que se uma sequência converge para um determinado limite, então todas as subsequências convergem e têm o mesmo limite.

Lema 9.2.1. Se uma sequência (x_k) converge para x, então todas as subsequências de (x_k) são convergentes e têm o mesmo limite x.

Exemplo 9.18. $\big((-1)^n\big)$ diverge pois se convergisse para algum $x \in \mathbb{R}$, suas subsequências convirgiriam este mesmo valor. Mas

$$\lim_{n\to\infty} \big((-1)^{2n}\big) = 1, \qquad \lim_{n\to\infty} \big((-1)^{2n+1}\big) = -1.$$

Exemplo 9.19. Seja (x_k) sequência convergente para l e tal que $x_{2k} = x_k^2$. Então

$$l = \lim_{n\to\infty} (x_{2k}) = \lim_{k\to\infty} x_k \lim_{k\to\infty} x_k = l^2.$$

Logo $l = 0$ ou $l = 1$. Para concluirmos qual dos dois candidatos a limite é o correto, precisaríamos de mais informações sobre a sequência. Por exemplo, se $x_k = a^k$ para $a < 1$, temos que $l = 0$ pois a sequência é limitada superiormente por $a < 1$. Então $l = 1$ não pode ser limite, e $\lim_{k\to\infty}(a^k) = 0$. Por outro lado, se $a = 1$ então $l = 1$.

Finalmente mostramos um importante resultado que nos garante convergência de alguma subsequência mesmo quando a sequência original não converge.

Teorema 9.2.2 (Bolzano–Weierstrass para sequências). *Toda sequência limitada em $\mathbb{R}$ tem pelo menos uma subsequência convergente.*

9.3. Sequências de Cauchy

Um conceito importante tratando-se de sequências é o de sequências de Cauchy. Formalmente, dizemos que uma sequência (x_k) é *de Cauchy* se para todo $\epsilon > 0$ existe $K^* \in \mathbb{N}$ tal que

$$|x_k - x_m| < \epsilon \quad \text{para todo } k, m \geq K^*.$$

Enunciar a seguir (sem demonstrar) a equivalência entre convergência e o critério de Cauchy.

Teorema 9.3.1 (Critério de convergência de Cauchy). *Uma sequência é convergente se e somente se é de Cauchy.*

Um corolário do resultado acima é que toda sequência de Cauchy é limitada (afina, toda sequência de Cauchy converge, e toda sequência convergente é limitada).

COROLÁRIO 9.3.2. Toda sequência de Cauchy é limitada.

EXEMPLO 9.20. Em geral, se (x_i) é tal que $|x_{i+1} - x_i| < c_i$, onde $S_i = \sum_{k=1}^{i} c_k$ é convergente, então (x_i) é convergente. De fato, mostramos abaixo que a sequência é de Cauchy, e portanto converge. Note que para $i > j$, temos
(9.3.1)
$$|x_i - x_j| \leq |x_i - x_{i-1}| + |x_{i-1} - x_{i-2}| + \cdots + |x_{j+1} - x_j| \leq c_{i-1} + c_{i-2} + \cdots + c_j = S_{i-1} - S_{j-1}.$$
Como S_i converge, então é de Cauchy. Logo dado $\epsilon > 0$, existe $K^* \in \mathbb{N}$ tal que $i > j > K^*$ implica que $|S_{i-1} - S_{j-1}| < \epsilon$. Logo, por (9.3.1) temos que $i > j > K^*$ implica que $|x_i - x_j| < \epsilon$ e (x_i) é de Cauchy.

9.4. Sequências Monótonas

Um classe muito especial de sequências é a de sequências monótonas. Uma sequência monótona é tal que seus valores não "oscilam", i.e., eles ou nunca diminuem ou nunca aumentam. Pode-se ver que a definição de sequência monótona é restritas a uma dimensão.

DEFINIÇÃO 9.4.1. *Dizemos que uma sequência* (x_n) *é* crescente *se* $x_1 < x_2 < \cdots < x_n < \ldots$ *e* não decrescente *se* $x_1 \leq x_2 \leq \cdots \leq x_n \leq \ldots$ *Similarmente, uma sequência* (x_n) *é* decrescente *se* $x_1 > x_2 > \cdots > x_n > \ldots$ *e* não crescente *se* $x_1 \geq x_2 \geq \cdots \geq x_n \geq \ldots$. *Finalmente, uma sequência é* monótona *se for crescente ou decrescente.*

EXEMPLO 9.21. $(1, 2, 3, 4, \ldots)$ é crescente, e $(1, 2, 3, 3, 3, 3, \ldots)$ não decrescentes.

EXEMPLO 9.22. $(1/n)$ é decrescente.

EXEMPLO 9.23. $(-1, 1, -1, 1, -1, \ldots)$ não é monótona.

TEOREMA 9.4.2. *Uma sequência não crescente ou não decrescente é convergente se e somente se é limitada.*
Além disso, se (x_n) *é não decrescente, então* $\lim_{n\to\infty}(x_n) = \sup\{x_n : n \in \mathbb{N}\}$. *Da mesma forma, se* (x_n) *é não crescente, então* $\lim_{n\to\infty}(x_n) = \inf\{x_n : n \in \mathbb{N}\}$.

DEMONSTRAÇÃO. ($\Longrightarrow$) Já vimos que toda sequência convergente é limitada.
($\Longleftarrow$) Assuma (x_n) não decrescente e limitada. Seja $x = \sup x_n : n \in \mathbb{N}$. Então dado $\epsilon > 0$, existe N tal que $x - \epsilon < x_N \leq x < x + \epsilon$, pois x é o supremo. Logo, para todo $n > N$ temos $x - \epsilon < x_N \leq x_n \leq x < x + \epsilon$, portanto x_n converge para x. Se a sequência for não-crescente, a demonstração é análoga. □

TEOREMA 9.4.3. *Uma sequência de números reais* (x_n), *monótona não decrescente e limitada converge para "seu supremo", i.e., converge para* $\sup\{x_n : n \in \mathbb{N}\}$.

DEMONSTRAÇÃO. Seja $x = \sup\{x_n : n \in \mathbb{N}\}$ (que existe pois a sequência é limitada). Então pela definição de supremo, para todo $\epsilon > 0$, existe $N^* \in \mathbb{N}$ tal que $x_{N^*} \in (x - \epsilon, x)$. Logo como a sequência é monótona não decrescente, temos
$$n > N^* \implies x_n > x_{N^*} > x - \epsilon.$$

Mas para todo $n \in \mathbb{N}$ temos $x_n \le x$ por definição de supremo. Logo

$$n > N^* \implies x_n > x_{N^*} > x - \epsilon \quad \text{e } x_n < x + \epsilon \implies x_n \in (x - \epsilon, x + \epsilon).$$

□

EXEMPLO 9.24. (a^n) diverge se $a > 1$ pois é ilimitada.

EXEMPLO 9.25. (a^n) converge se $0 < a \le 1$ pois é monótona decrescente e limitada. Além disso, $\lim_{n\to\infty}(a^n) = 0$, pois $\inf\{a_n : n \in \mathbb{N}\} = 0$.

EXEMPLO 9.26. Seja $y_1 = 1$ e $y_{n+1} = (1 + y_n)/3$. Mostraremos que (y_n) é convergente e achamos seu limite. Note que $y_2 = 2/3 < 1 = y_1$. Vamos mostrar por indução que $0 < y_{n+1} < y_n$. Esta afirmativa vale para $n = 1$. Assuma verdadeira para $n = k - 1$, isto é $0 < y_k < y_{k-1}$. Então para $n = k$ temos

$$y_{k+1} = (1 + y_k)/3 < (1 + y_{k-1})/3 = y_k,$$

e como $y_k > 0$, então $y_{k+1} > 0$, como queríamos. Portanto a sequência é monótona não crescente e limitada inferiormente por zero. Portanto converge. Seja y seu limite. Então

$$y = \lim_{n\to\infty} y_{n+1} = \lim_{n\to\infty} (1 + y_n)/3 = (1 + y)/3.$$

Logo $y = 1/2$.

EXEMPLO 9.27. Seja $y_1 = 1$, e $y_{n+1} = (2y_n + 3)/4$. Note que $y_2 = 5/4 > y_1$. Para mostrar que $y_{n+1} > y_n$ em geral, usamos indução. Note que para $n = 1$ o resultado vale. Assuma agora que valha também para $n = k$ para algum k, i.e., $y_{k+1} > y_k$. Então

$$y_{k+2} = \frac{1}{4}(2y_{k+1} + 3) > \frac{1}{4}(2y_k + 3) = y_{k+1}.$$

Logo, por indução, $y_{n+1} > y_n$ para todo $n \in \mathbb{N}$, e (y_n) é não decrescente. Para mostrar que é limitada, note que $|y_1| < 2$. Mais uma vez usamos indução a fim de provar que em geral $|y_n| < 2$. Assuma que $|y_k| < 2$. Logo,

$$|y_{k+1}| = |\frac{1}{4}(2y_{k+1} + 3)| < \frac{1}{4}(2|y_{k+1}| + 3) < \frac{7}{4} < 2.$$

Por indução, segue-se que $|y_n| < 2$ para todo $n \in \mathbb{N}$. Como (y_n) é monótona e limitada, então é convergente. Seja $y = \lim_{n\to\infty}(y_n)$. Então

$$y = \lim_{n\to\infty}(y_n) = \lim_{n\to\infty}((2y_n + 3)/4) = ((2y + 3)/4).$$

resolvendo a equação algébrica acima, temos $y = 3/2$.

9.5. Séries

Devido às suas peculiaridades, as séries possuem algumas propriedades próprias. Um resultado importante em se tratando de séries versa sobre o comportamento assintótico dos termos que a compõem. Temos o seguinte resultado.

LEMA 9.5.1. Seja a série dada por $\sum_{i=1}^{\infty} c_i$ convergente. Então $\lim_{i\to\infty} c_i = 0$.

Demonstração. Note que se a série converge para um valor S, as sequência parciais $S_n = \sum_{i=1}^{n} c_i$ convergem para o mesmo valor S. Logo

$$\lim_{i\to\infty} c_i = \lim_{i\to\infty}(S_i - S_{i-1}) = \lim_{i\to\infty} S_i - \lim_{i\to\infty} S_{i-1} = S - S = 0.$$

□

Observação. Do resultado acima, concluimos que se c_i não tem limite, ou se seu limite não é zero, a série não converge.

Exemplo 9.28. Considere as séries de termos alternados $\sum_{i=1}^{\infty}(-1)^i$. Como a sequência $(-1)^i$ não converge, a série não é convergente. O mesmo vale para $\sum_{n=1}^{\infty}\sin(n\pi)$.

Uma classe de séries especial é dada pelas *séries geométricas*. Para $r \in \mathbb{R}$, esta série é dada por

$$\sum_{i=0}^{\infty} r^i.$$

Note que as somas parciais $S_n = \sum_{i=0}^{n} r^i$ divergem se $|r| \geq 1$, e convergem se $|r| < 1$. De fato, se $|r| \geq 1$, então r^i não converge a zero, e portanto a série não converge. Para $|r| < 1$, temos

$$S_n = \frac{1 - r^{n+1}}{1 - r},$$

e portanto a série converge para $\lim_{n\to\infty} S_n = 1/(1-r)$.

Detalharemos abaixo a questão de convergência absoluta e condicional, e testes de convergência próprios para séries. Dizemos que a série $\sum_{i=1}^{\infty} c_i$ converge *absolutamente* se a série $\sum_{i=1}^{\infty} |c_i|$ converge. E dizemos que $\sum_{i=1}^{\infty} c_i$ converge *condicionalmente* se $\sum_{i=1}^{\infty} c_i$ converge, mas não converge absolutamente, i.e., $\sum_{i=1}^{\infty} |c_i|$ diverge. O resultado abaixo nos garante que a convcergência absoluta é suficiente para uma série convergir. A primeira demonstração usa a noção de série de Cauchy, e apresentamos uma segunda demonstração que dispensa este conceito.

Teorema 9.5.2. *Toda série que converge absolutamente é convergente.*

Demonstração usando sequências de Cauchy. Seja $\sum_{i=1}^{\infty} c_i$ tal que $\sum_{i=1}^{\infty} |c_i|$ convirja, e denote por $s_n = \sum_{i=1}^{n} c_i$ as somas parciais. Note que, para $n > m$,

$$|s_n - s_m| \leq \sum_{m}^{n} |c_i| = \hat{s}_n - \hat{s}_m,$$

onde $\hat{s}_n = \sum_{i=1}^{n} |c_i|$. Como $\hat{s}_n$ é de Cauchy, então s_n é também de Cauchy, e portanto convergente. □

Demonstração alternativa [**14**]. Seja $p_i = \max\{a_i, 0\}$, e $q_i = \min\{a_i, 0\}$. Então $|a_i| = p_i - q_i$ e $a_i = p_i + q_i$. Então $p_i \leq |a_i|$ e $q_i \leq |a_i|$. Logo $\sum_{i=1}^{\infty} p_i$ e $\sum_{i=1}^{\infty} q_i$ convergem pois são limitadas superiormente por $\sum_{i=1}^{\infty} |a_i|$. Logo

$$s_n = \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} p_i + \sum_{i=1}^{n} q_i$$

converge. □

Observação. Note que na demonstração acima, que se $\sum_{i=1}^{\infty} a_i$ é condicionalmente convergente, então $\sum_{i=1}^{\infty} p_i$ e $\sum_{i=1}^{\infty} q_i$ divergem. De fato, se apenas uma destas séries convergisse, então

$$\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} p_i + \sum_{i=1}^{\infty} q_i$$

divergiria. Se as duas convergissem, então

$$\sum_{i=1}^{\infty} |a_i| = \sum_{i=1}^{\infty} p_i - \sum_{i=1}^{\infty} q_i$$

convergiria.

Observação. Uma pergunta natural é o que aconteceria se trocássemos a ordem em que se soma a série. Se a série for absolutamente convergente, então a ordem da soma pode ser trocada e a série ainda converge para o mesmo valor. Entretanto, se a série for condicionalmente convergente, o seguinte surpreendente resultado vale: é possível trocar a ordem da soma tal que a série convirja para *qualquer* número real.

Observação. Note que se $\sum_{n=1}^{\infty} a_n$ converge absolutamente, então $\sum_{n=1}^{\infty} a_n^2$ converge absolutamente.

Prosseguimos agora descrevendo alguns testes que podem ser usados para definir convergências de séries.

Lema 9.5.3 (Teste da razão). Suponha que $\lim_{i\to\infty} |c_{i+1}/c_i| = L$. Então se $L < 1$, a série $\sum_{i=1}^{\infty} c_i$ converge absolutamente, e se $L > 1$, a série diverge.

Demonstração. Faremos aqui a prova no caso em que $|c_{i+1}/c_i| < L < 1$ para todo inteiro i. Note que neste caso,

$$|c_i| < L|c_{i-1}| < L^2|c_{i-2}| < \cdots < L^{i-1}|c_1|.$$

Então $\sum_{i=1}^{\infty} |c_i| < |c_1| \sum_{i=1}^{\infty} L^{i-1}$ converge absolutamente pois $|L| < 1$. $\square$

Lema 9.5.4 (Teste da raiz). Suponha que $\lim_{i\to\infty} \sqrt[i]{|c_i|} = L$. Então se $L < 1$, a série $\sum_{i=1}^{\infty} c_i$ converge absolutamente, e se $L > 1$, a série diverge.

Demonstração. Consideraremos somente o caso $\sqrt[i]{|c_i|} < L < 1$ para todo i. Então $\sum_{i=1}^{\infty} |c_i| < \sum_{i=1}^{\infty} L^i$ que converge absolutamente pois $|L| < 1$. $\square$

Observação. Tanto o teste da razão como a da raiz são inconclusivo caso $L = 1$, como os exemplos $\sum_{i=1}^{\infty} 1/i$ e $\sum_{i=1}^{\infty} 1/i^2$ ilustram.

Lema 9.5.5 (Teste da integral). Seja $f : [1, +\infty) \to \mathbb{R}$ não negativa e monótona decrescente. Então $\sum_{i=1}^{\infty} f(i)$ converge se e somente se $\int_0^{\infty} f(x)\,dx < +\infty$.

Lema 9.5.6 (Série de termos positivos). Seja $\sum_{n=1}^{\infty} a_n$, onde (a_n) é sequência de números positivos. Então a série converge se e somente se a sequência das somas parciais é limitada.

Demonstração. Para demonstrar este resultado, basta ver que as somas parciais da série determinam uma sequência monótona, e esta converge se e somente se é limitada. $\square$

Lema 9.5.7 (Teste da comparação). Sejam $\sum_{i=1}^{\infty} a_n$ e $\sum_{i=1}^{\infty} b_n$ duas séries de termos não negativos tais que $a_n \leq cb_n$ para todo $n \in \mathbb{N}$. Então

- Se $\sum_{n=1}^{\infty} b_n$ converge então $\sum_{n=1}^{\infty} a_n$ converge
- Se $\sum_{n=1}^{\infty} a_n$ diverge então $\sum_{n=1}^{\infty} b_n$ diverge

Uma aplicação importante do Lema 9.5.6 é na determinação da convergência da série harmônica alternada, como mostramos a seguir.

Exemplo 9.29 ([**3**]). Considere a *série harmônica alternada* $\sum_{i=1}^{\infty}(-1)^{i+1}/i$. Então a série converge pelo seguinte argumento. Considere a série "par" das somas parciais

$$S_{2n} = (1 - 1/2) + (1/3 - 1/4) + \cdots + (1/(2n-1) - 1/(2n)),$$

e "ímpar"

$$S_{2n+1} = 1 - (1/2 - 1/3) - (1/4 - 1/5) - \cdots - (1/(2n) - 1/(2n+1)).$$

Como cada termo entre parênteses nas séries acima são positivos, temos que a série par é monótona crescente e a ímpar monótona decrescente. Note a seguir que

$$0 < S_{2n} < S_{2n} + 1/(2n+1) = S_{2n+1} < 1, \tag{9.5.1}$$

e portanto ambas são limitadas e convergem. Por (9.5.1), concluimos que convergem para o mesmo valor, e que toda a série converge.

Muitas vezes é possível calcular o valor exato da série, como nos exemplos abaixo

Exemplo 9.30. Considere a série

$$\sum_{i=1}^{\infty} \frac{1}{n(n+1)}$$

Note que $1/[n(n+1)] = 1/n - 1/(n+1)$, e portanto a n-ésima soma parcial s_n é dada por

$$s_n = (1 - 1/2) + (1/2 - 1/3) + \cdots + (1/n - 1/(n+1)) = 1 - 1/(n+1).$$

Logo a série converge para $\lim_{n\to\infty} s_n = 1$.

Da mesma forma, podemos calcular o valor de séries envolvendo termos como

$$\frac{1}{(n+1)(n+2)} = \frac{1}{(n+1)} - \frac{1}{(n+2)}.$$

9.6. Exercícios

Exercício 9.1. Demonstrar o Teorema 9.1.13.

Exercício 9.2. Demonstrar o Teorema 9.1.4.

Exercício 9.3. Mostre que se x_n é sequência de números positivos, então $x_n \to 0$ se e somente se $1/x_n \to \infty$.

Exercício 9.4. Seja (x_n) tal que $x_n \neq 0$ para todo $n \in \mathbb{N}$ e

$$L = \lim_{n\to\infty} \frac{|x_{n+1}|}{|x_n|}$$

existe e $L > 1$. Mostre que para todo $C \in \mathbb{R}$ existe $N^* \in \mathbb{N}$ tal que

$$n \geq N^* \implies |x_n| > C.$$

EXERCÍCIO 9.5. Mostre o resultado da Observação 9.5

CAPíTULO 10

Funções de várias variáveis

[1] Neste capítulo estudamos as propriedades de funções de várias variáveis. Nos concentramos principalmente nas questões relativas a diferenciabilidade destas funções e aplicações.

10.1. Introdução

Consideamos aqui funções $f : \mathbb{R}^m \to \mathbb{R}^n$, onde, tipicamente, $m = 2, 3$ e $n = 1$. Assim como no caso unidimensional, dizemos que uma função f é contínua em $\mathbf{x}_0$ se dado $\epsilon > 0$ existe $\delta > 0$ tal que

$$\|\mathbf{x} - \mathbf{x}_0\| < \delta \implies \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \epsilon.$$

Analogamente ao caso unidimensional, podemos definir o conceito de limite de funções num determinado ponto, ver Exercício 10.1.

Outro conceito importante é o de curvas (em duas dimensões) ou superfícies (em três dimensões) de nível, que chamaremos sempre de curvas de nível. Uma curva de nível é dada pelo conjunto de pontos que têm imagem constante, i.e., dado $c \in \mathbb{R}$, definimos

$$\{\mathbf{x} \in \mathbb{R}^m : f(\mathbf{x}) = c\}.$$

10.2. Derivadas parciais e planos tangentes

Seja $f : \mathbb{R}^m \to \mathbb{R}$. Então definimos a *i-ésima derivada parcial* de f num ponto $\mathbf{x} = (x_1, \dots, x_m)$ por

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{h \to 0} \frac{f(x_1, \dots, x_{i-1}, x_i + h, x_{i+1}, \dots, x_m) - f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_m)}{h}$$

quando o limite acima existir. Se cada uma das derivadas parciais existirem, definimos o *vetor gradiente* (também chamado de *gradiente* pelos íntimos) dada por

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_m}(\mathbf{x})\right).$$

Note-se que depois de obtida uma derivada parcial, pode-se derivar novamente para se obter segundas derivadas parciais. Por exemplo, derivando-se na direção i e depois na direção j obtemos a função $\partial^2 f/\partial x_j \partial x_i$. Para funções "suaves", a ordem em que se deriva não importa. Em particular, se $\partial^2 f/\partial x_j \partial x_i(\mathbf{x}_0)$ e $\partial^2 f/\partial x_i \partial x_j(\mathbf{x}_0)$ existem e são contínuas numa vizinhança de $\mathbf{x}_0$, então

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0).$$

[1] Última Atualização: 31/08/2020

Considere agora o caso bidimesional e $f:\mathbb{R}^2\to\mathbb{R}$. Com o conceito de derivadas parciais, é possível, quando a função é suave o suficiente, definir o plano tangente ao gráfico de f num determinado ponto $\mathbf{p}_0=(x_0,y_0)\in\mathbb{R}^2$. Começamos por lembrar que a reta tangente no caso $m=1$ é definida pelos pontos (x,z) tais que

$$z-f(x_0)-f'(x_0)(x-x_0)=0.$$

Fixando primeiramente y_0 temos que $f(x,y_0)$ é uma função de uma dimensão, e a reta tangente na direção x então é dada pelos pontos (x,y_0,z) tais que

$$z-f(x_0,y_0)-\frac{\partial f}{\partial x}(x_0,y_0)(x-x_0)=0. \tag{10.2.1}$$

De forma análoga, fixando-se x_0, obtém-se a reta tangente à $f(x,y_0)$ na direção y, e esta é dada pelos pontos (x_0,y,z) tais que

$$z-f(x_0,y_0)-\frac{\partial f}{\partial y}(x_0,y_0)(y-x_0)=0. \tag{10.2.2}$$

Dadas as duas retas acima, tangentes ao gráfico de f, podemos determinar o pano tangente P a este mesmo gráfico. Tomando $x=x_0+1$ em (10.2.1) e $y=y_0+1$ em (10.2.2), temos que

$$\left(x_0+1,y_0,f(x_0,y_0)+\frac{\partial f}{\partial x}(x_0,y_0)\right),\qquad \left(x_0,y_0+1,f(x_0,y_0)+\frac{\partial f}{\partial y}(x_0,y_0)\right)$$

pertencem ao plano P. Como $\big(x_0,y_0,f(x_0,y_0)\big)$ também pertence a P, então os vetores

$$\mathbf{v}_x=\left(1,0,\frac{\partial f}{\partial x}(x_0,y_0)\right),\qquad \mathbf{v}_y=\left(0,1,\frac{\partial f}{\partial x}(x_0,y_0)\right)$$

são tangentes a P. Portanto

$$\mathbf{N}=\mathbf{v}_x\times\mathbf{v}_y=\left(\frac{\partial f}{\partial x}(x_0,y_0),\frac{\partial f}{\partial y}(x_0,y_0),-1\right)$$

é normal a P. Como temos que

$$P=\{(x,y,z):\ \big(x-x_0,y-y_0,z-f(x_0,y_0)\big)\cdot\mathbf{N}=0\}$$

então P é definido pela equação

$$\frac{\partial f}{\partial x}(x_0,y_0)(x-x_0)+\frac{\partial f}{\partial y}(x_0,y_0)(y-y_0)-z+f(x_0,y_0)=0.$$

10.3. Diferenciabilidade

A noção de diferenciabilidade e de derivada em dimensões maiores simplesmente generaliza de forma natural a derivada unidimensional. Seja $\boldsymbol{f}:\mathbb{R}^m\to\mathbb{R}^n$ e $\mathbf{x}\in\mathbb{R}^m$. Dizemos que $\boldsymbol{f}$ é diferenciável em $\mathbf{x}$ se

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|\boldsymbol{f}(\mathbf{x}+\mathbf{h})-\boldsymbol{f}(\mathbf{x})-\boldsymbol{f}'(\mathbf{x})\mathbf{h}\|}{\|\mathbf{h}\|}=0,$$

onde $\boldsymbol{f}'(\mathbf{x})$ tem a representação matricial $[\boldsymbol{f}'(\mathbf{x})] \in \mathbb{R}^{n\times m}$ dada por

$$[\boldsymbol{f}'(\mathbf{x})] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_m} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}.$$

A matriz $[\boldsymbol{f}'(\mathbf{x})]$ também é chamada de *matriz jacobiana* de f no ponto $\mathbf{x}$. . Chamamos $\boldsymbol{f}'(\mathbf{x})$ de derivada de $\boldsymbol{f}$ em $\mathbf{x}$, e que também denotamos por $\boldsymbol{D}f(\mathbf{x})$.

Assim como em uma dimensão, $\boldsymbol{f}$ é diferenciável em $\mathbf{x}$ se e somente se existir uma função $\boldsymbol{r} : \mathbb{R}^m \to \mathbb{R}^n$ tal que

$$\boldsymbol{f}(\mathbf{x}+\mathbf{h}) = \boldsymbol{f}(\mathbf{x}) + \boldsymbol{f}'(\mathbf{x})(\mathbf{h}) + \mathbf{r}(\mathbf{h}) \quad \text{com } \lim_{\mathbf{h}\to\mathbf{0}} \frac{\|\mathbf{r}(\mathbf{h})\|}{\|\mathbf{h}\|} = 0. \tag{10.3.1}$$

Note que pela identidade acima, temos imediatamente que diferenciabilidade implica em continuidade.

Exemplo 10.1. Considere $\boldsymbol{A} : \mathbb{R}^m \to \mathbb{R}^n$ aplicação linear e $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$ dada por $\boldsymbol{f}(\mathbf{x}) = \boldsymbol{A}(\mathbf{x}) + \boldsymbol{c}$, onde $\boldsymbol{c} \in \mathbb{R}^n$ é vetor constante. Então $\boldsymbol{f}'(\mathbf{x})(\mathbf{h}) = \boldsymbol{A}(\mathbf{h})$ para todo $\mathbf{x}, \mathbf{h} \in \mathbb{R}^m$. Note que neste caso, a derivada $\boldsymbol{f}'(\mathbf{x})$ é na verdade *independente* de $\mathbf{x}$.

Exemplo 10.2. Seja a matriz $A \in \mathbb{R}^{m\times n}$ e $\mathbf{c} = (c_1, \dots, c_m)$ vetor constante. Considere ainda $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$ tal que para $\mathbf{x} = (x_1, \dots, x_m)$ tem-se $\boldsymbol{f}(\mathbf{x}) = (f_1(\mathbf{x}), \dots, f_n(\mathbf{x}))$ e

$$\vec{\boldsymbol{f}}(\mathbf{x}) = A\vec{\mathbf{x}} + \vec{\boldsymbol{c}} \qquad \text{i.e., } f_i(\mathbf{x}) = \sum_{j=1}^m A_{ij}x_j + c_j \quad \text{para } i = 1, \dots, n.$$

Então, para $\mathbf{h} = (h_1, \dots, h_m)$ tem-se $\boldsymbol{f}'(\mathbf{x})(\mathbf{h}) = \mathbf{y}$ onde $\vec{\mathbf{y}} = A\vec{\mathbf{h}}$, i.e.,

$$y_i = \sum_{j=1}^m A_{ij}h_j.$$

Compare com o exemplo 10.1.

Uma interessante forma de analisarmos uma função em várias variáveis é restringindo esta função numa direção e usando propriedades de funções de apenas uma variável. Para tanto, sejam $\mathbf{u} \in \mathbb{R}^m$ com $\|\mathbf{u}\| = 1$, e $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$. Dado $\mathbf{x} \in \mathbb{R}^m$, definimos a *derivada direcional de* $\boldsymbol{f}$ *em* $\mathbf{x}$ *na direção* $\mathbf{u}$ por $\boldsymbol{D}_{\mathbf{u}}\boldsymbol{f}(\mathbf{x}) \in \mathbb{R}^n$, onde

$$\boldsymbol{D}_{\mathbf{u}}\boldsymbol{f}(\mathbf{x}) = \lim_{t\to 0} \frac{\boldsymbol{f}(\mathbf{x}+t\mathbf{u}) - \boldsymbol{f}(\mathbf{x})}{t} = 0.$$

quando o limite acima existir.

No caso em que $\mathbf{u} = \mathbf{e}_i$, então temos a *derivada parcial* em relação à iésima coordenada e escrevemos

$$\boldsymbol{D}_{\mathbf{e}_i}\boldsymbol{f}(\mathbf{x}) = \frac{\partial \boldsymbol{f}}{\partial x_i}(\mathbf{x}).$$

É importante ressaltar que a *existência de derivadas parciais em relação às coordenadas não implica na existência de derivadas direcionais em geral.* Considere o simples exemplo abaixo.

EXEMPLO 10.3. Seja $f : \mathbb{R}^2 \to \mathbb{R}$ dada por

$$f(x,y) = \begin{cases} \frac{x}{y} & \text{se } y \neq 0, \\ 0 & \text{se } y = 0. \end{cases}$$

Então

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

mas a derivada direcional na direção (a,b) não existe se a e b são não nulos, pois não existe o limite quando $t \to 0$ de

$$\frac{f(ta,tb) - f(0,0)}{t} = \frac{1}{t}\frac{a}{b}.$$

A situação muda se supusermos diferenciabilidade, como mostra o resultado a seguir.

TEOREMA 10.3.1. *Seja* $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$ *diferenciável em* $\mathbf{x} \in \mathbb{R}^m$. *Seja* $\mathbf{u} \in \mathbb{R}^m$ *com* $\|\mathbf{u}\| = 1$. *Então existe a derivada direcional* $\boldsymbol{D}_{\mathbf{u}}f(\mathbf{x})$, *e esta é dada por*

$$\boldsymbol{D}_{\mathbf{u}}\boldsymbol{f}(\mathbf{x}) = \boldsymbol{f}'(\mathbf{x})(\mathbf{u}).$$

DEMONSTRAÇÃO. Como $\boldsymbol{f}$ é diferenciável em $\mathbf{x}$, então para todo $\epsilon > 0$ existe $\delta > 0$ tal que

$$\mathbf{h} \in \mathbb{R}^m,\ 0 < \|\mathbf{h}\| < \delta \implies \frac{\|\boldsymbol{f}(\mathbf{x}+\mathbf{h}) - \boldsymbol{f}(\mathbf{x}) - \boldsymbol{f}'(\mathbf{x})(\mathbf{h})\|}{\|\mathbf{h}\|} < \epsilon.$$

Tomando $\mathbf{h} = t\mathbf{u}$, com $|t| \in (0,\delta)$, temos

$$\left\| \frac{\boldsymbol{f}(\mathbf{x}+t\mathbf{u}) - \boldsymbol{f}(\mathbf{x})}{t} - \boldsymbol{f}'(\mathbf{x})(\mathbf{u}) \right\| < \epsilon.$$

Logo

$$\lim_{t\to 0} \frac{\boldsymbol{f}(\mathbf{x}+t\mathbf{u}) - \boldsymbol{f}(\mathbf{x})}{t} = \boldsymbol{f}'(\mathbf{x})(\mathbf{u}),$$

e portanto a derivada direcional existe e é dada por $\boldsymbol{f}'(\mathbf{x})(\mathbf{u})$. $\square$

OBSERVAÇÃO. Considerando $n = 1$, temos que

$$\boldsymbol{D}_{\mathbf{u}}f(\mathbf{x}) = f'(\mathbf{x})(\mathbf{u}) = \nabla f(\mathbf{x}) \cdot \mathbf{u} = \|\nabla f(\mathbf{x})\|\|\mathbf{u}\|\cos\theta \leq \|\nabla f(\mathbf{x})\|\|\mathbf{u}\|, \tag{10.3.2}$$

onde θ é o ângulo formado por $\nabla f(\mathbf{x})$ e $\mathbf{u}$. Portanto, a derivada direcional atinge seu maior valor se $\theta = 0$, i.e., quando $\mathbf{u}$ aponta na direção do gradiente. Outra forma de se ler a equação (10.3.2) é observar que a direção do gradiente é a direção de crescimento máximo da função, e que a direção contrária ao gradiente é a direção de menor crescimento (menor derivada).

A existência de derivadas direcionais *não implica em diferenciabilidade*. Para ilustrar tal fato, considere a função

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

Então

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

mas dado (a,b) com $\|(a,b)\|^2 = a^2 + b^2 = 1$ e $b \neq 0$, temos

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

e a derivada direcional é dada por

$$D_{(a,b)} f(0,0) = \frac{a^2}{b}. \tag{10.3.3}$$

Entretanto, se f fosse diferenciável, teríamos

$$D_{(a,b)} f(0,0) = \boldsymbol{f}'(0,0)(a,b) = \frac{\partial f}{\partial x}(0,0)a + \frac{\partial f}{\partial y}(0,0)b = 0,$$

uma contradição com (10.3.3). Logo f não é diferenciável em $(0,0)$ apesar de ter todas as derivadas direcionais neste ponto. Note que $f(x,x^2) = 1$ para $x \neq 0$, e portanto f é descontínua em $(0,0)$.

Apesar da existência de derivadas direcionais num determinado ponto não garantir a diferenciabilidade neste ponto, a existência e continuidade das derivadas parciais numa *vizinhança* dum ponto garante a diferenciabilidade, como podemos ver no resultado a seguir.

TEOREMA 10.3.2. *Seja* $\mathbf{f} : \mathbb{R}^m \to \mathbb{R}$*. Se* $\partial\mathbf{f}/\partial x_i$ *existir e for contínua numa vizinhança de* $\mathbf{x}$ *para* $i = 1, \dots, m$*, então* $\mathbf{f}$ *é diferenciável em* $\mathbf{x}$*.*

Outro resultado de grande importância diz respeito à diferenciabilidade de composições de funções, garantindo que se duas funções são diferenciáveis, então a composição também o é.

TEOREMA 10.3.3 (Regra da Cadeia). *Sejam* $\boldsymbol{f} : \mathbb{R}^l \to \mathbb{R}^m$ *e* $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^n$*. Se* $\boldsymbol{f}$ *é diferenciável em* $\mathbf{x} \in \mathbb{R}^l$ *e* $\boldsymbol{g}$ *é diferenciável em* $\boldsymbol{f}(\mathbf{x})$*, então* $\boldsymbol{g} \circ \boldsymbol{f}$ *é diferenciável em* $\mathbf{x}$ *e*

$$(\boldsymbol{g} \circ \boldsymbol{f})'(\mathbf{x}) = \mathbf{g}'(\boldsymbol{f}(\mathbf{x})) \circ \boldsymbol{f}'(\mathbf{x}).$$

Veremos agora várias aplicações da regra da cadeia.

- **Aplicação 1:**

  EXEMPLO 10.4. Seja $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^n$, e seja a função $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^n$ inversa de $\boldsymbol{f}$, isto é,

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

  para todo $\mathbf{x}$, $\mathbf{y}$ em $\mathbb{R}^n$. Se $\boldsymbol{f}$ é diferenciável em $\mathbf{x} \in \mathbb{R}^n$, e $\boldsymbol{g}$ é diferenciável em $\mathbf{y} = \boldsymbol{f}(\mathbf{x})$, então $\boldsymbol{f}'(\mathbf{x})$ e $\mathbf{g}'(\mathbf{y})$ são inversas uma da outra, isto é,

  $$\boldsymbol{f}'(\mathbf{x}) \circ \mathbf{g}'(\mathbf{y}) = \mathbf{g}'(\mathbf{y}) \circ \boldsymbol{f}'(\mathbf{x}) = \boldsymbol{I},$$

  onde $\boldsymbol{I}$ é o operador identidade $\boldsymbol{I}(\mathbf{x}) = \mathbf{x}$.

  De fato, seja $\boldsymbol{h}(\mathbf{x}) = \boldsymbol{g}(\boldsymbol{f}(\mathbf{x})) = \mathbf{x}$. Derivando $\boldsymbol{h}(\mathbf{x}) = \mathbf{x}$, temos $\mathbf{h}'(\mathbf{x}) = \boldsymbol{I}$. Usando a regra da cadeia para $\boldsymbol{h}(\mathbf{x}) = \boldsymbol{g}(\boldsymbol{f}(\mathbf{x}))$, temos $\mathbf{h}'(\mathbf{x}) = \mathbf{g}'(\mathbf{y})\boldsymbol{f}'(\mathbf{x})$. Logo, $\mathbf{g}'(\mathbf{y})\boldsymbol{f}'(\mathbf{x}) = \boldsymbol{I}$. De forma análoga segue-se que $\boldsymbol{f}'(\mathbf{x})\mathbf{g}'(\mathbf{y}) = \boldsymbol{I}$.

- **Aplicação 2:** Uma aplicação imediata da regra da cadeia é dada no seguinte teorema do valor médio para funções de várias variáveis. Na verdade, esta é uma aplicação imediata do teorema do valor médio unidimensional (Teorema 6.3.3) quando restringimos uma função de várias variáveis a um segmento de reta.

TEOREMA 10.3.4. *Seja* $f : \mathbb{R}^m \to \mathbb{R}$, *diferenciável. Sejam* $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^m$ *e* $S = \{\mathbf{x} + t(\mathbf{y} - \mathbf{x}) : t \in (0,1)\}$. *Então existe* $\boldsymbol{\xi} \in S$ *tal que*

$$f(\mathbf{y}) - f(\mathbf{x}) = \boldsymbol{f}'(\boldsymbol{\xi})(\mathbf{y} - \mathbf{x}).$$

DEMONSTRAÇÃO. Este resultado segue-se de uma aplicação do teorema do valor médio unidimensional (Teorema 6.3.3) para a função $\phi : [0,1] \to \mathbb{R}$ dada por $\phi(t) = f\big(\mathbf{x} + t(\mathbf{y} - \mathbf{x})\big)$. Note ainda que pela regra da cadeia temos que

$$\phi'(t) = \boldsymbol{f}'\big(\mathbf{x} + t(\mathbf{y} - \mathbf{x})\big)(\mathbf{y} - \mathbf{x}).$$

□

É interessante notar que não vale uma "generalização trivial" para o teorema do valor médio quando a imagem de uma função está no $\mathbb{R}^n$, para $n \geq 2$. Como exemplo, considere a função $\phi : \mathbb{R} \to \mathbb{R}^2$ dada por $\phi(t) = (\sin t, \cos t)$. Tomando-se os pontos $t = 0$ e $t = 2\pi$, vemos que não existe $\xi \in [0, 2\pi]$ tal que

$$\mathbf{0} = \phi(0) - \phi(2\pi) = \phi'(\xi)(2\pi - 0) = 2\pi\phi'(\xi).$$

pois $\phi'(\xi) \neq \mathbf{0}$ para todo ξ.

Existe entretanto o seguinte resultado para funções em $\mathbb{R}^n$.

TEOREMA 10.3.5. *Seja* $f : \mathbb{R}^m \to \mathbb{R}^n$, *diferenciável. Sejam* $\mathbf{x}_0$, $\mathbf{x}_1 \in \mathbb{R}^m$ *e seja* S *o segmento de reta unindo estes pontos. Então existe* $\boldsymbol{\xi} \in S$ *tal que*

$$\|\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)\| \leq \|\boldsymbol{f}'(\boldsymbol{\xi})(\mathbf{x}_1 - \mathbf{x}_0)\|.$$

DEMONSTRAÇÃO. Seja $\mathbf{v} = \boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)$, e $\boldsymbol{\phi}(\mathbf{x}) = \boldsymbol{f}(\mathbf{x}) \cdot \mathbf{v}$. Então

$$\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_0) = (\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)) \cdot \mathbf{v} = \|\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)\|^2,$$

e $\boldsymbol{\phi}'(\mathbf{x})(\mathbf{h}) = [\boldsymbol{f}'(\mathbf{x})(\mathbf{h})] \cdot \mathbf{v}$. Pelo Teorema do valor médio dado pelo Teorema 10.3.4, existe $\boldsymbol{\xi} \in S$ tal que $\boldsymbol{\phi}(\mathbf{x}_1) - \boldsymbol{\phi}(\mathbf{x}_0) = \boldsymbol{\phi}'(\boldsymbol{\xi})(\mathbf{x}_1 - \mathbf{x}_0)$, i.e.,

$$\|\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)\|^2 = [\boldsymbol{f}'(\boldsymbol{\xi})(\mathbf{x}_1 - \mathbf{x}_0)] \cdot (\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)) \leq \|\boldsymbol{f}'(\boldsymbol{\xi})(\mathbf{x}_1 - \mathbf{x}_0)\|\|\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)\|.$$

Finalmente, se $\boldsymbol{f}(\mathbf{x}_1) = \boldsymbol{f}(\mathbf{x}_0)$, o resultado é trivial. Caso contrário dividimos ambos os lados da desigualdade acima para concluir a demonstração. □

OBSERVAÇÃO. Podemos continuar a desigualdade acima mais um passo e obter

$$\|\boldsymbol{f}(\mathbf{x}_1) - \boldsymbol{f}(\mathbf{x}_0)\| \leq \|\boldsymbol{f}'(\boldsymbol{\xi})\|\|(\mathbf{x}_1 - \mathbf{x}_0)\|.$$

Aplicação 3: Gradientes são ortogonais a curvas de níveis. Seja $\boldsymbol{\phi} : \mathbb{R} \to \mathbb{R}^2$ parametrização "suave" de uma curva de nível de uma função $f(x,y)$. Então $f(\boldsymbol{\phi}(t))$ é constante, e

$$0 = \frac{d}{dt} f(\boldsymbol{\phi}(t)) = \nabla f(\boldsymbol{\phi}(t)) \cdot \boldsymbol{\phi}'(t).$$

Portanto, como $\boldsymbol{\phi}'$ é tangente a curva de nível da f em $\mathbf{x}$, temos que $\nabla f(\mathbf{x})$ é ortogonal a curvas de nível de f no ponto $\mathbf{x}$.

Aplicação 4: Uma outra aplicação da regra da cadeia tem a ver com funções homogêneas. Dizemos que $f : \mathbb{R}^m \to \mathbb{R}$ é *homogênea de grau* k se $f(t\mathbf{x}) = t^k f(\mathbf{x})$ para todo $t \in \mathbb{R}$. Por exemplo, $x^2/(yz)$ é homogênea de grau zero, e $\sqrt{x^5}$ é homogênea de grau $5/2$. Um Teorema devido à Euler afirma que f é homogênea de grau k se e somente se

$$\nabla f(\mathbf{x}) \cdot \mathbf{x} = k f(\mathbf{x}). \tag{10.3.4}$$

Para obter (10.3.4) a partir de uma função homogênea f, basta derivar $f(t\mathbf{x}) = t^k f(\mathbf{x})$ em relação a t usando a regra da cadeia:

$$\mathbf{x} \cdot \nabla f(t\mathbf{x}) = k t^{k-1} f(\mathbf{x}),$$

e depois tomar $t = 1$.

Aplicação 5: É possível obter a regra de Leibniz através da regra da cadeia. Seja

$$F(t_1, t_2, t_3) = \int_{t_1}^{t_2} f(t_3, y)\, dy.$$

Então

$$\frac{d}{dx} F(u(x), v(x), x) = \frac{\partial F}{\partial t_1}\frac{du}{dx} + \frac{\partial F}{\partial t_2}\frac{dv}{dx} + \frac{\partial F}{\partial t_3},$$

implica em

$$\frac{d}{dx}\int_{u(x)}^{v(x)} f(x,y)\, dy = -f(x, u(x))\frac{du}{dx} + f(x, v(x))\frac{dv}{dx} + \int_{u(x)}^{v(x)} \frac{\partial f(x,y)}{\partial x}\, dy.$$

10.4. Matriz Hessiana, Fórmula de Taylor e pontos críticos

Note que a derivada de uma função de uma função de $f : \mathbb{R}^m \to \mathbb{R}$ num determinado ponto $\mathbf{x}$ foi definida como uma aplicação linear de $\mathbb{R}^m$ em $\mathbb{R}$. No caso, para $\mathbf{x}$ fixo, teríamos $\boldsymbol{f}'(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}$ dada por

$$\boldsymbol{f}'(\mathbf{x})(\mathbf{y}) = \frac{\partial f}{\partial x_1}(\mathbf{x}) y_1 + \frac{\partial f}{\partial x_2}(\mathbf{x}) y_2 + \cdots + \frac{\partial f}{\partial x_m}(\mathbf{x}) y_m,$$

onde $\mathbf{y} \in \mathbb{R}^m$.

De forma análoga, definimos a segunda derivada de f num ponto $\mathbf{x}$ fixado como sendo a função bilinear $f''(\mathbf{x}) : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ tal que

$$f''(\mathbf{x})(\mathbf{y}, \mathbf{z}) = \sum_{i,j=1}^{m} \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j} y_i z_j, \quad \text{onde } \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right),$$

e $\mathbf{y}, \mathbf{z} \in \mathbb{R}^m$. Uma forma mais compacta de escrever a definição acima é usando-se a *matriz hessiana* H dada por $H_{ij}(\mathbf{x}) = \partial^2 f(\mathbf{x})/\partial x_i \partial x_j$. Logo

$$f''(\mathbf{x})(\mathbf{y}, \mathbf{z}) = (\vec{\mathbf{y}})^t H(\mathbf{x}) \vec{\mathbf{z}}.$$

Observação. Um interessante resultado garante que se f for suficientemente suave num determinado ponto $\mathbf{x}_0$ (é suficiente que as segundas derivadas existam e sejam contínuas numa vizinhança aberta de $\mathbf{x}_0$) teremos que *não importa a ordem em que se toma as derivadas*, i.e., $\partial^2 f/\partial x_i \partial x_j = \partial^2 f/\partial x_j \partial x_i$, e portanto a matriz hessiana é simétrica. Este tipo de resultado, com diferentes hipóteses, é atribuido à *Clairaut* em **[29]**, e à *Schwarz* em **[2, 17]**.

Definições para derivadas de ordem mais alta seguem o mesmo formato, sendo estas *aplicações multilineares*. Entretanto para os nossos propósitos, a matriz hessiana basta.

Apresentamos no teorema a seguir a fórmula de Taylor, e nos restringimos ao caso particular de polinômios quadráticos. Este teorema será de fundamental importância para caracterizarmos pontos extremos.

TEOREMA 10.4.1 (Taylor). *Seja $f : \mathbb{R}^m \to \mathbb{R}$ duas vezes diferenciável em Ω, com derivadas contínuas. Para $\mathbf{x} \in \mathbb{R}^m$ e $\mathbf{h} \in \mathbb{R}^m$, existe $\hat{t} \in (0,1)$ tal que para $\boldsymbol{\xi} = \mathbf{x} + \hat{t}\mathbf{h}$ tem-se*

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \boldsymbol{f}'(\mathbf{x})(\mathbf{h}) + \frac{1}{2}f''(\boldsymbol{\xi})(\mathbf{h},\mathbf{h}). \tag{10.4.1}$$

DEMONSTRAÇÃO. Seja $\phi : [0,1] \to \mathbb{R}$ dada por $\phi(t) = f\big(\mathbf{x} + t\mathbf{h}\big)$. Aplicando o Teorema de Taylor em uma dimensão (Teorema 6.4.1), obtemos que existe $\hat{t} \in (0,1)$ tal que

$$\phi(1) = \phi(0) + \phi'(0) + \frac{1}{2}\phi''(\hat{t}).$$

Note que

$$\phi'(t) = \mathbf{f}'(\mathbf{x}+t\mathbf{h})(\mathbf{h}) = \sum_{i=1}^{m} \frac{\partial f}{\partial x_i}(\mathbf{x}+t\mathbf{h})h_i, \qquad \phi''(t) = \sum_{j=1}^{m}\sum_{i=1}^{m} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}+t\mathbf{h})h_i h_j,$$

e usando a definição de ϕ obtemos o resultado diretamente. $\square$

OBSERVAÇÃO. Note que exigindo que as segundas derivadas sejam contínuas, podemos usar o fato de que a "ordem" das segundas derivadas não importam.

Assim como em uma dimensão, usaremos o Teorema de Taylor para estudarmos pontos extremos de uma função. Dizemos que $f : \mathbb{R}^m \to \mathbb{R}$ tem um *máximo local* em $\mathbf{x} \in \mathbb{R}^m$ se existe $\delta > 0$ tal que

$$\|\mathbf{y} - \mathbf{x}\| < \delta \implies f(\mathbf{y}) \le f(\mathbf{x}). \tag{10.4.2}$$

Dizemos que $\mathbf{x}$ é *máximo estrito local* se valer a desigualdade estrita em (10.4.2). Definição análoga serve para *mínimo local* e *mínimo estrito local*. Chamamos um ponto de máximo ou mínimo local de *ponto extremo local*, e um ponto de máximo ou mínimo estrito local de *ponto extremo estrito local*.

O resultado que obtemos a seguir, relativo a pontos extremos interiores, é análogo ao caso unidimensional, ver o Teorema 6.3.1, e diz primeiro que pontos extremos interiores são *pontos críticos*, i.e., pontos em que a derivada se anula. O resultado mostra também que se um ponto $\mathbf{x}$ é de mínimo local, então a forma bilinear $f''(\mathbf{x})$ é *semi-definida positiva*, i.e, $f''(\mathbf{x})(\mathbf{h},\mathbf{h}) \ge 0$ para todo $\mathbf{h} \in \mathbb{R}^m$. De forma análoga se um ponto é de máximo local, então $f''(\mathbf{x})$ é *semi-definida negativa*, i.e, $f''(\mathbf{x})(\mathbf{h},\mathbf{h}) \le 0$ para todo $\mathbf{h} \in \mathbb{R}^m$.

Em termos matriciais, $f''(\mathbf{x})$ é semi-definida positiva se a matriz hessiana $H(\mathbf{x})$ o for, i.e., se $(\vec{\mathbf{h}})^t H(\mathbf{x})\vec{\mathbf{h}} \ge 0$ para todo $\mathbf{h} \in \mathbb{R}^m$, e semi-definida negativa se $(\vec{\mathbf{h}})^t H(\mathbf{x})\vec{\mathbf{h}} \le 0$ para todo $\mathbf{h} \in \mathbb{R}^m$.

TEOREMA 10.4.2 (Ponto extremo interior). *Seja $f : \mathbb{R}^m \to \mathbb{R}$ e $\mathbf{x} \in \mathbb{R}^m$ ponto extremo local. Se f é diferenciável em $\mathbf{x}$, então $\mathbf{x}$ é ponto crítico, i.e., $\boldsymbol{f}'(\mathbf{x}) = 0$. Se além disto, f for duas vezes diferenciável com derivadas segundas contínuas, então temos que*

(1) se $\mathbf{x}$ *for ponto de mínimo local, então* $f''(\mathbf{x})(\mathbf{h},\mathbf{h}) \geq 0$ *para todo* $\mathbf{h} \in \mathbb{R}^m$*,*
(2) se $\mathbf{x}$ *for ponto de máximo local, então* $f''(\mathbf{x})(\mathbf{h},\mathbf{h}) \leq 0$ *para todo* $\mathbf{h} \in \mathbb{R}^m$*.*

DEMONSTRAÇÃO. Para mostrar que $\mathbf{x}$ é ponto crítico, basta usar o Teorema 10.3.1 e mostrar que as derivadas parciais se anulam, pois dado o vetor $\mathbf{e}_i$ temos que a função $\phi(t) = f(\mathbf{x}+t\mathbf{e}_i)$ tem ponto extremo local em $t = 0$. Usando o Teorema 6.3.1 vemos que $\phi'(0) = 0$. Mas então

$$0 = \phi'(0) = f'(\mathbf{x})(\mathbf{e}_i) = \frac{\partial f}{\partial x_i}(\mathbf{x})$$

e concluímos que $\boldsymbol{f}'(\mathbf{x}) = 0$.

Suponha agora que f seja duas vezes diferenciável com derivadas segundas contínuas, e que $\mathbf{x}$ seja ponto de mínimo local. Então $\mathbf{x}$ é ponto crítico, como acabamos de mostrar, e pelo Teorema de Taylor em várias dimensões (Teorema 10.4.1), temos que

$$f(\mathbf{x}+s\mathbf{u}) - f(\mathbf{x}) = \frac{s^2}{2} f''(\boldsymbol{\xi}_s)(\mathbf{u},\mathbf{u}),$$

para todo s suficientemente pequeno e $\mathbf{u} \in \mathbb{R}^m$, onde $\boldsymbol{\xi}_s$ é ponto do segmento unindo $\mathbf{x}$ e $\mathbf{x}+s\mathbf{u}$. Quando $s \to 0$, temos que $\boldsymbol{\xi}_s \to \mathbf{x}$, e usando a continuidade de f'' concluímos que

$$f''(\mathbf{x})(\mathbf{u},\mathbf{u}) = \lim_{s\to 0} f''(\boldsymbol{\xi}_s)(\mathbf{u},\mathbf{u}) = 2\lim_{s\to 0} \frac{f(\mathbf{x}+s\mathbf{u}) - f(\mathbf{x})}{s^2} \geq 0,$$

pois como $\mathbf{x}$ é mínimo local, então $f(\mathbf{x}+s\mathbf{u}) - f(\mathbf{x}) \geq 0$ para todo s suficientemente pequeno. Portanto $f''(\mathbf{x})(\mathbf{u},\mathbf{u}) \geq 0$, como queríamos demonstrar. □

Os resultados acima nos dão condições necessárias para um ponto interior ser extremo local, porém estas não são suficientes (vide exemplo $f(x) = x^3$). Dizemos que um ponto é *de sela* quando a derivada se anula mas este não é extremo local. Um caso interessante é quando a função é localmente crescente na direção de uma coordenada e decrescente na direção de outra. Por exemplo, $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x,y) = x^2 - y^2$, ver Figura 1. Ver também a sela de macaco dada por $f(x,y) = x^3 - 3xy^2$, Figura 2 (tirada de [**24**]).

O resultado a seguir nos fornece algumas condições suficientes para um ponto ser de máximo, mínimo ou de sela. Mais precisamente, temos que se um ponto crítico $\mathbf{x}$ de uma função suave tem $f''(\mathbf{x})$ *positiva definida*, i.e, $f''(\mathbf{x})(\mathbf{h},\mathbf{h}) > 0$ para todo $\mathbf{h} \in \mathbb{R}^m \backslash \{\mathbf{0}\}$, então ele é mínimo estrito local. De forma análoga, se $f''(\mathbf{x})$ é *negativa definida*, i.e, $f''(\mathbf{x})(\mathbf{h},\mathbf{h}) < 0$ para todo $\mathbf{h} \in \mathbb{R}^m \backslash \{\mathbf{0}\}$, então ele é máximo estrito local. O último caso é quando $f''(\mathbf{x})$ é *indefinida* i.e, existem $\mathbf{h}$, $\boldsymbol{\xi}$ em $\mathbb{R}^m$ tais que $[f''(\mathbf{x})(\mathbf{h},\mathbf{h})][f''(\mathbf{x})(\boldsymbol{\xi},\boldsymbol{\xi})] < 0$. Aí então $\mathbf{x}$ é ponto de sela.

TEOREMA 10.4.3. *Seja* $f : \mathbb{R}^m \to \mathbb{R}$ *duas vezes diferenciável, com derivadas contínuas, e* $\mathbf{x} \in \mathbb{R}^m$ *ponto crítico. Temos então que*
(1) se $f''(\mathbf{x})$ *for positiva definida então* $\mathbf{x}$ *é mínimo estrito local,*
(2) se $f''(\mathbf{x})$ *for negativa definida então* $\mathbf{x}$ *é máximo estrito local,*
(3) se $f''(\mathbf{x})$ *for indefinida então* $\mathbf{x}$ *é ponto de sela.*

DEMONSTRAÇÃO. Mostraremos apenas o caso em que $f''(\mathbf{x})$ é positiva definida. Neste caso, devido à continuidade das segundas derivadas, $f''(\cdot)$ é positiva definida numa vizinhança aberta de $\mathbf{x}$. Para $\mathbf{y} \neq \mathbf{x}$ satisfazendo as condições do Teorema de Taylor no $\mathbb{R}^m$

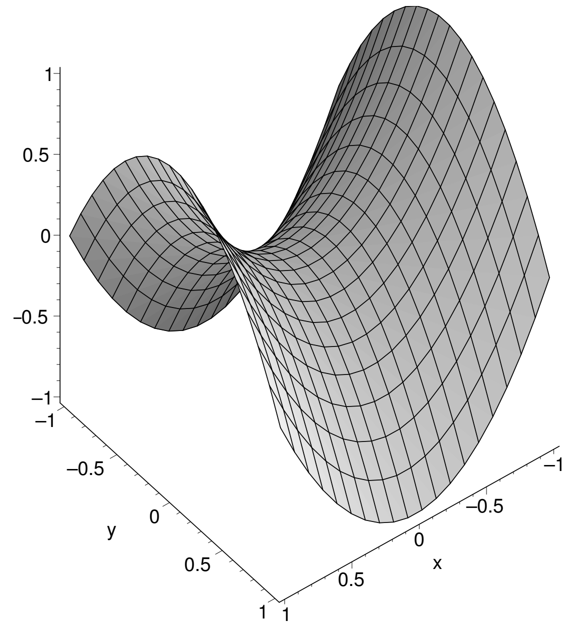

FIGURA 1. Gráfico de $x^2 - y^2$, que tem ponto de sela em $(0, 0)$.

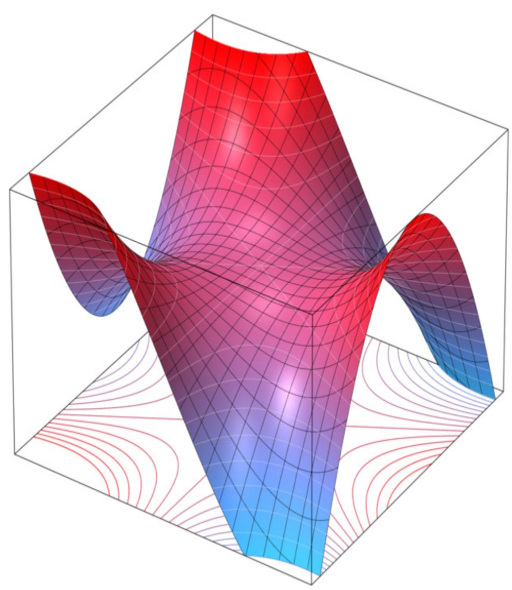

FIGURA 2. Gráfico da sela de macaco dada por $x^3 - 3xy^2$, com ponto de sela em $(0, 0)$.

(Teorema 10.4.1), e suficientemente próximo de $\mathbf{x}$, temos que existe $\boldsymbol{\xi}$ pertencente ao segmento de reta entre $\mathbf{y}$ e $\mathbf{x}$ e tal que

$$f(\mathbf{y}) - f(\mathbf{x}) = \frac{1}{2} f''(\boldsymbol{\xi})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}). \tag{10.4.3}$$

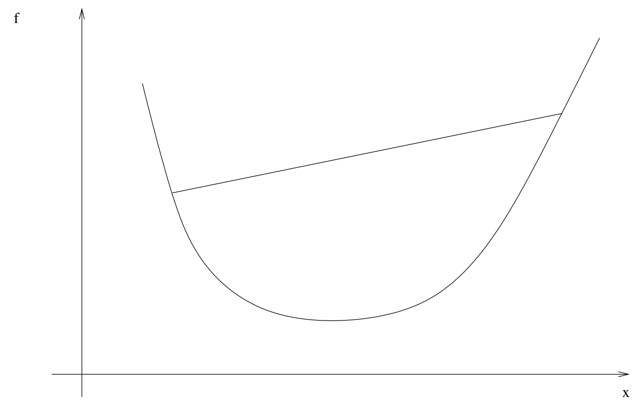

FIGURA 3. Função convexa.

Portanto $\mathbf{x}$ é mínimo estrito local pois a espressão do lado direito de (10.4.3) é estritamente positiva. $\square$

Note que apesar do teorema anterior dar condições suficientes para determinar se um ponto crítico é ou não extremo local, ainda é preciso descobrir se a f'' é positiva ou negativa definida ou indeterminada. Esta dificuldade é contornável, pois existem vários resultados de álgebra linear que dizem, por exemplo, quando uma matriz é ou não positiva definida. Por exemplo, uma matriz simétrica é positiva definida se e somente se seus autovalores são positivos. A referência [**12**] apresenta este e vários outros resultados relacionados ao tema.

EXEMPLO 10.5. Seja $F : \mathbb{R}^m \to \mathbb{R}$ dada por

$$F(\mathbf{x}) = c + \vec{\mathbf{b}}^t\vec{\mathbf{x}} + \frac{1}{2}\vec{\mathbf{x}}^t A\vec{\mathbf{x}},$$

onde $A \in \mathbb{R}^{m\times m}$ é simétrica positiva definida, $\vec{\mathbf{b}} \in \mathbb{R}^{m\times 1}$, e $c \in \mathbb{R}$. Então $\mathbf{x}^*$ é ponto de mínimo estrito de F se e somente se $A\vec{\mathbf{x}}^* = -\vec{\mathbf{b}}$. De fato, se $\mathbf{x}^*$ é ponto de mínimo estrito de F, então $F'(\mathbf{x}^*) = 0$. Mas a matriz jacobiana $[F'(\mathbf{x}^*)] \in \mathbb{R}^{m\times 1}$ é dada por

$$[F'(\mathbf{x}^*)] = (\vec{\mathbf{x}}^*)^t A + \vec{\mathbf{b}}^t,$$

e portanto $A\vec{\mathbf{x}}^* = -\vec{\mathbf{b}}$. Por outro lado, se $A\vec{\mathbf{x}}^* = -\vec{\mathbf{b}}$, então $F'(\mathbf{x}^*) = 0$. Como a matriz hessiana de F, dada por A, é positiva definida, então $\mathbf{x}^*$ é ponto de mínimo estrito de F.

Uma segunda aplicação do Teorema 10.4.1 diz respeito à funções convexas definidas em convexos. Dizemos que $\Omega \subseteq \mathbb{R}^m$ é convexo se $\mathbf{x}$, $\mathbf{y} \in \Omega$ implica em $(1-t)\mathbf{x} + t\mathbf{y} \in \Omega$ para todo $t \in [0,1]$. Dizemos que $f : \Omega \to \mathbb{R}$ é convexa em Ω se

$$f\big((1-t)\mathbf{x} + t\mathbf{y}\big) \leq (1-t)f(\mathbf{x}) + tf(\mathbf{y}).$$

para todo $t \in [0,1]$. Graficamente, uma função é convexa se o gráfico de f entre $\mathbf{x}$ e $\mathbf{y}$ está abaixo da reta que une os pontos $(\mathbf{x}, f(\mathbf{x}))$ e $(\mathbf{y}, f(\mathbf{y}))$, como ilustra a Figura 3.

Existem inúmeros resultados relacionados a convexidade. Em particular, um mínimo local é também global, e se o mínimo local é estrito, segue-se a unicidade de mínimo global [**21**].

Teorema 10.4.4. *Seja $f : \mathbb{R}^m \to \mathbb{R}$ duas vezes diferenciável, com derivadas contínuas. Então as afirmativas abaixo são equivalentes:*

(1) f é convexa
(2) $f''(\mathbf{x})$ é semi-definida positiva para todo $\mathbf{x} \in \mathbb{R}^m$.

Demonstração. ($\Leftarrow$) Suponha que $f''(\mathbf{x})$ seja semi-definida positiva em $\mathbb{R}^m$. Seja S o segmento de reta unindo $\mathbf{x}$ e $\mathbf{y} \in \mathbb{R}^m$, e seja $0 < t < 1$. Definindo $\mathbf{x}_0 = (1-t)\mathbf{x} + t\mathbf{y}$, pelo Teorema de Taylor existe $\boldsymbol{\xi}_1 \in S$ entre $\mathbf{x}$ e $\mathbf{x}_0$, e $\boldsymbol{\xi}_2 \in S$ entre $\mathbf{x}_0$ e $\mathbf{y}$ tais que

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \boldsymbol{f}'(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2} f''(\boldsymbol{\xi}_1)(\mathbf{x} - \mathbf{x}_0, \mathbf{x} - \mathbf{x}_0),$$
$$f(\mathbf{y}) = f(\mathbf{x}_0) + \boldsymbol{f}'(\mathbf{x}_0)(\mathbf{y} - \mathbf{x}_0) + \frac{1}{2} f''(\boldsymbol{\xi}_2)(\mathbf{y} - \mathbf{x}_0, \mathbf{y} - \mathbf{x}_0).$$

Como $f''(\boldsymbol{\xi}_1)$ e $f''(\boldsymbol{\xi}_2)$ são ambas semi-definidas positivas, então

$$\begin{aligned}
&(1-t)f(\mathbf{x}) + tf(\mathbf{y}) \\
&= f(\mathbf{x}_0) + \boldsymbol{f}'(\mathbf{x}_0)[(1-t)\mathbf{x} + t\mathbf{y} - \mathbf{x}_0] + \frac{(1-t)}{2} f''(\boldsymbol{\xi}_1)(\mathbf{x} - \mathbf{x}_0, \mathbf{x} - \mathbf{x}_0) + \frac{t}{2} f''(\boldsymbol{\xi}_2)(\mathbf{y} - \mathbf{x}_0, \mathbf{y} - \mathbf{x}_0) \\
&= f(\mathbf{x}_0) + \frac{(1-t)}{2} f''(\boldsymbol{\xi}_1)(\mathbf{x} - \mathbf{x}_0, \mathbf{x} - \mathbf{x}_0) + \frac{t}{2} f''(\boldsymbol{\xi}_2)(\mathbf{y} - \mathbf{x}_0, \mathbf{y} - \mathbf{x}_0) \geq f(\mathbf{x}_0).
\end{aligned}$$

Logo f é convexa.

($\Rightarrow$) Se f é convexa,

$$f((1-t)\mathbf{x} + t\mathbf{y}) \leq (1-t)f(\mathbf{x}) + tf(\mathbf{y})$$

e para $t \in (0, 1]$ temos que

$$\frac{f((1-t)\mathbf{x} + t\mathbf{y}) - f(\mathbf{x})}{t} \leq f(\mathbf{y}) - f(\mathbf{x}).$$

Tomando o limite $t \to 0$ obtemos $\boldsymbol{f}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$. Seja $s = \|\mathbf{x} - \mathbf{y}\|$ e $\mathbf{h} = (\mathbf{y} - \mathbf{x})/s$. Usando agora a fórmula de Taylor obtemos que existe $\hat{s} \in (0, s)$ tal que

$$\frac{1}{2} f''(\mathbf{x} + \hat{s}\mathbf{h})(s\mathbf{h}, s\mathbf{h}) = \frac{1}{2} f''(\mathbf{x} + \hat{s}\mathbf{h})(\mathbf{y} - \mathbf{x}, \mathbf{y} - \mathbf{x}) = f(\mathbf{y}) - f(\mathbf{x}) - \boldsymbol{f}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0.$$

Usando a bilinearidade da aplicação $f''(\mathbf{x} + \hat{s}\mathbf{h})$, temos

$$f''(\mathbf{x} + \hat{s}\mathbf{h})(\mathbf{h}, \mathbf{h}) \geq 0.$$

para todo $\mathbf{h} \in B_1(\mathbf{0})$. Tomando $\mathbf{y} \to \mathbf{x}$ temos $s \to 0$ e portanto $\hat{s} \to 0$. Usando a continuidade de f'' concluímos a demonstração. □

Observação. Note que no processo de demonstração do Teorema 10.4.4, mostramos também que uma função f ser convexa implica em $\boldsymbol{f}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) \leq f(\mathbf{y}) - f(\mathbf{x})$ para todo $\mathbf{x}$, $\mathbf{y}$.

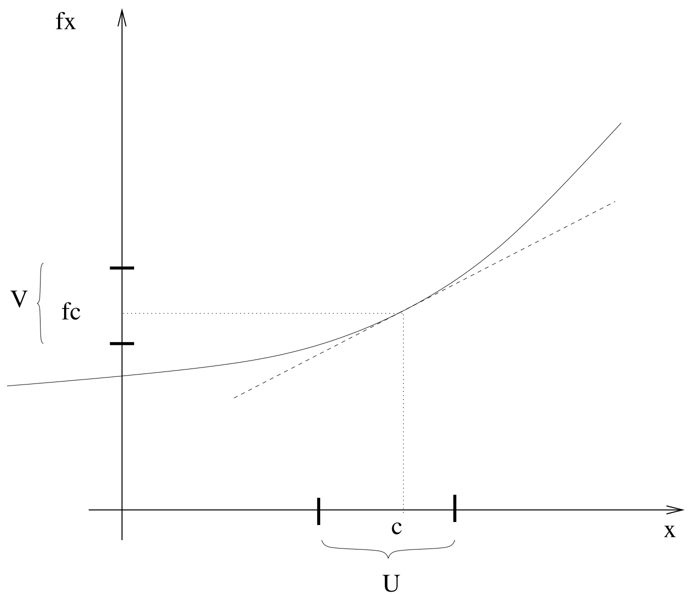

Figura 4. Teorema da função inversa.

10.5. Teorema da Função Inversa e da Função Implícita

10.5.1. Teorema da Função Inversa. Como motivação considere primeiro o caso unidimensional, e seja $f : \mathbb{R} \to \mathbb{R}$ "suave". Se $f'(x) \neq 0$ para algum $x \in \mathbb{R}$, então f é localmente invertível, i.e, f é injetiva numa vizinhança aberta U de x e existe $g = f^{-1} : V \to U$, onde $U = f(V)$, tal que

$$g(f(x)) = x, \qquad \text{para todo } x \in U.$$

No caso a "suavidade" necessária é que a função tenha derivadas contínuas. Dado $\Omega \subseteq \mathbb{R}^m$, dizemos que uma função $f : \Omega \to \mathbb{R}^n$ é de classe $\mathcal{C}^1(\Omega)$ se é diferenciável com derivadas contínuas em Ω.

Teorema 10.5.1 (Função Inversa). *Seja $\Omega \subseteq \mathbb{R}^m$ e $\boldsymbol{f} : \Omega \to \mathbb{R}^m$ de classe $\mathcal{C}^1(\Omega)$. Seja $\hat{\mathbf{x}} \in \Omega$ tal que $D = \boldsymbol{f}'(\hat{\mathbf{x}})$ é invertível. Então existe uma vizinhança aberta U de $\hat{\mathbf{x}}$ tal que*

(1) $\boldsymbol{f} : U \to V = \boldsymbol{f}(U)$ é injetiva, e V é aberto.
(2) Seja $\boldsymbol{g} : V \to U$ a função inversa de $\boldsymbol{f}$ definida por

$$\boldsymbol{g}(\boldsymbol{f}(\mathbf{x})) = \mathbf{x} \qquad \textit{para todo } \mathbf{x} \in U.$$

Então $\boldsymbol{g} \in \mathcal{C}^1(V)$ e para $\hat{\mathbf{y}} = \boldsymbol{f}(\hat{\mathbf{x}})$ tem-se $\boldsymbol{g}'(\hat{\mathbf{y}}) = [\boldsymbol{f}'(\hat{\mathbf{x}})]^{-1}$.

Observação. Note que o teorema acima tem caráter *local*. Em particular, é possível construir funções não injetivas em seu domínios que possuem matrizes jacobianas invertíveis em todos os pontos. Entretanto em *uma dimensão*, se a derivada de uma função sobrejetiva de classe $\mathcal{C}^1$ não se anular em nenhum ponto de um intervalo aberto, a função é globalmente invertível.

10.5.2. Teorema da função implícita. O teorema de função implícita trata da importante questão de solvabilidade de equações dadas de forma implícita. A pergunta é simples:

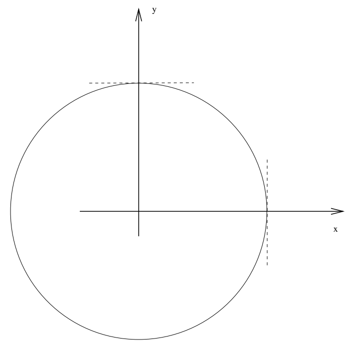

FIGURA 5. Conjunto $\{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

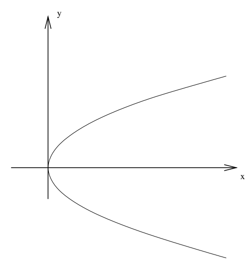

FIGURA 6. Conjunto $\{(x,y) \in \mathbb{R}^2 : x = y^2\}$.

dados os pontos $(\mathbf{x}, \mathbf{y})$, soluções de uma equação $F(\mathbf{x}, \mathbf{y}) = 0$, será que é possível escrever $\mathbf{y}$ em função de $\mathbf{x}$, numa vizinhança de $\mathbf{x}$?

Como uma primeira motivação, considere $F(x,y) = x^2 + y^2 - 1$. Então a curva de nível determinada por $F(x,y) = 0$ é dada pelo círculo de raio unitário, como nos mostra a Figura 5. Seja $(a,b) \in \mathbb{R}^2$ tal que $F(a,b) = 0$. Por exemplo $(0,1)$ e $(0,-1)$ satisfazem esta condição, ou seja, numa vizinhaça de $(0,1)$, podemos escrever y como função de x, na caso $y = \sqrt{1-x^2}$. E em torno de $(0,-1)$ podemos escrever $y = -\sqrt{1-x^2}$. Uma pergunta natural é se existe uma função ϕ tal que $F(x, \phi(x)) = 0$, e $\phi(a) = b$. A resposta é *globalmente, não*. Mas *localmente sim*, se $\partial F / \partial y(a,b) \neq 0$.

Um segundo exemplo é dado por $F(x,y) = x - y^2$, ver Figura 6. Para se ter $F(x, \phi(x)) = 0$, pode-se escolher $\phi(x) = \sqrt{x}$ ou $\phi(x) = -\sqrt{x}$. Entretanto nenhuma das duas funções está definida na vizinhança de $x = 0$. Note que $\partial F / \partial y(0,0) = 0$.

Um exemplo final, agora em dimensões maiores. Sejam $T_1 : \mathbb{R}^m \to \mathbb{R}^n$ e $T_2 : \mathbb{R}^n \to \mathbb{R}^n$ transformações lineares, e $F : \mathbb{R}^{m+n} \to \mathbb{R}^n$ dada por $F(\mathbf{x}, \mathbf{y}) = T_1\mathbf{x} + T_2\mathbf{y}$. Então podemos escrever a equação $F(\mathbf{x}, \mathbf{y}) = 0$ somente em função de $\mathbf{x}$ se T_2 for invertível. Neste caso temos $F(\mathbf{x}, -T_2^{-1}T_1\mathbf{x}) = 0$. Note que dados $\mathbf{u} \in \mathbb{R}^m$ e $\mathbf{v}^n$, tem-se

$$\boldsymbol{F}'(\boldsymbol{a}, \boldsymbol{b})(\mathbf{u}, \mathbf{v}) = T_1\mathbf{u} + T_2\mathbf{v} \tag{10.5.1}$$

Se definirmos a aplicação linear $L : \mathbb{R}^n \to \mathbb{R}^n$ dada por $L : \mathbf{v} \to \boldsymbol{F}'(\boldsymbol{a}, \boldsymbol{b})(0, \mathbf{v})$, teremos $L = T_2$. Então a condição de solvabilidade é de L seja invertível.

Observação. No exemplo acima, podemos usar um abuso de notação e escrever (10.5.1) como

$$\boldsymbol{F}'(\boldsymbol{a},\boldsymbol{b})(\mathbf{u},\mathbf{v}) = \frac{\partial F}{\partial \mathbf{x}}(\boldsymbol{a},\boldsymbol{b})\cdot\mathbf{u} + \frac{\partial F}{\partial \mathbf{y}}(\boldsymbol{a},\boldsymbol{b})\mathbf{v}.$$

Portanto a condição de solvabilidade é que $\partial F/\partial \mathbf{y}(\boldsymbol{a},\boldsymbol{b})$ seja invertível.

Teorema 10.5.2 (Função implícita). *Seja $\Omega \subseteq \mathbb{R}^{m+n}$ um aberto, e $(\mathbf{x}_0,\mathbf{y}_0)\in\Omega$. Seja $\boldsymbol{F}:\Omega\to\mathbb{R}^n$ de classe $\mathcal{C}^1(\Omega)$, e tal que $\boldsymbol{F}(\mathbf{x}_0,\mathbf{y}_0)=0$. Se a transformação linear de $\mathbb{R}^n$ em $\mathbb{R}^n$ definida por $\mathbf{v}\mapsto\boldsymbol{F}'(\mathbf{x}_0,\mathbf{y}_0)(\mathbf{0},\mathbf{v})$ for invertível, então existe uma vizinhança aberta W de $\mathbf{x}_0$, e uma única função $\boldsymbol{\phi}:W\to\mathbb{R}^n$, que é $\mathcal{C}^1(W)$ e tal que $\mathbf{y}_0=\boldsymbol{\phi}(\mathbf{x}_0)$ e $F(\mathbf{x},\boldsymbol{\phi}(\mathbf{x}))=0$ para todo $\mathbf{x}\in W$.*

Demonstração. Sem perda de generalidade, suponha $\mathbf{x}_0=\mathbf{0}$ e $\mathbf{y}_0=\mathbf{0}$. Seja $H:\Omega\to\mathbb{R}^{m+n}$ dada por $H(\mathbf{x},\mathbf{y})=(\mathbf{x},\boldsymbol{F}(\mathbf{x},\mathbf{y}))$. Então $H'(\mathbf{0},\mathbf{0})$ é invertível. Pelo teorema da função inversa (Teorema 10.5.1), existe vizinhança aberta U de $(\mathbf{0},\mathbf{0})$ em $\mathbb{R}^{m+n}$ tal que $V=H(U)$ é vizinhança aberta em $\mathbb{R}^{m+n}$. Além disto existe $\Phi:V\to U$ inversa de H de classe $\mathcal{C}^1$. Escrevendo $\Phi=(\phi_1,\phi_2)$, onde $\phi_1:V\to\mathbb{R}^m$ e $\phi_1:V\to\mathbb{R}^n$, temos

$$(\mathbf{x},\mathbf{y})=H\circ\Phi(\mathbf{x},\mathbf{y})=H(\phi_1(\mathbf{x},\mathbf{y}),\phi_2(\mathbf{x},\mathbf{y}))=(\phi_1(\mathbf{x},\mathbf{y}),F(\phi_1(\mathbf{x},\mathbf{y}),\phi_2(\mathbf{x},\mathbf{y}))).$$

Logo,

$$\mathbf{x}=\phi_1(\mathbf{x},\mathbf{y}),\qquad \mathbf{y}=F(\mathbf{x},\phi_2(\mathbf{x},\mathbf{y})), \tag{10.5.2}$$

para todo $(\mathbf{x},\mathbf{y})\in V$. Então $W=\{\mathbf{x}\in\mathbb{R}^m:\ (\mathbf{x},\mathbf{0})\in\mathbf{V}\}$ é vizinhança aberta de $\mathbf{x}=\mathbf{0}$ em $\mathbb{R}^m$. Definindo $\phi(\mathbf{x})=\phi_2(\mathbf{x},\mathbf{0})$, temos $\phi(\mathbf{0})=\mathbf{0}$, e segue-se de (10.5.2) que $F(\mathbf{x},\phi(\mathbf{x}))=\mathbf{0}$. Como Φ é de classe $\mathcal{C}^1$, então ϕ_2, e portanto ϕ também é de classe $\mathcal{C}^1$. □

Observação. Se escrevermos a matriz jacobiana em formato de blocos, como

$$\big(\boldsymbol{F}'(\mathbf{x}_0,\mathbf{y}_0)\big)=\big(\tfrac{\partial \boldsymbol{F}}{\partial \mathbf{x}}(\mathbf{x}_0,\mathbf{y}_0)\quad \tfrac{\partial \boldsymbol{F}}{\partial \mathbf{y}}(\mathbf{x}_0,\mathbf{y}_0)\big)$$

temos que a condição do teorema da função implícita é dada por $\partial\boldsymbol{F}/\partial\mathbf{y}(\mathbf{x}_0,\mathbf{y}_0)$ ser invertível.

Note que o Teorema da função implícita *não* garante existência de $(\mathbf{x}_0,\mathbf{y}_0)$ tal que $\boldsymbol{F}(\mathbf{x}_0,\mathbf{y}_0)=0$. Por exemplo, considere $F:\mathbb{R}^2\to\mathbb{R}$ dada por $F(x,y)=x^2+y^2+1$. Então F não tem raiz, apesar de $\partial F/\partial y(x_0,y_0)=2y\neq 0$ sempre que $y\neq 0$.

Exemplo 10.6. Considere a caso onde $F:\mathbb{R}^2\to\mathbb{R}$ é suave e seja $(x_0,y_0)\in\mathbb{R}^2$ tal que x_0, $y_0\in\mathbb{R}$, $F(x_0,y_0)=c$ para alguma constante $c\in\mathbb{R}$, e $\partial F/\partial y(x_0,y_0)\neq 0$. Então, aplicando o teorema de função implícita para $F(x_0,y_0)-c$, existe função ϕ definida numa vizinhança de x_0 e tal que $\phi(x_0)=y_0$ e $F(x,\phi(x))=c$. Note que

$$0=\frac{d}{dx}F(\mathbf{x},\phi(\mathbf{x}))=\frac{\partial F}{\partial x}(\mathbf{x},\phi(\mathbf{x}))+\frac{\partial F}{\partial y}(\mathbf{x},\phi(\mathbf{x}))\phi'(\mathbf{x}),$$

e portanto $\phi'(\mathbf{x})=-\frac{\partial F/\partial x}{\partial F/\partial y}$ calculada no ponto $(\mathbf{x},\phi(\mathbf{x}))$.

10.6. Minimização com restrições

Para problemas de minimização com restrições, dois importantes resultados nos dão condições suficientes para que um ponto seja extremo. São os teoremas de Lagrange e de Kuhn–Tucker, que apresentamos abaixo.

10.6.1. Condições de primeira ordem.

10.6.1.1. *Restrições de igualdade.* Dadas funções reais $f, g_1, \dots, g_k$ definidas num aberto Ω de $\mathbb{R}^m$, consideramos o problema de minimizar f restrita ao conjunto de raízes de $g_1, \dots, g_k$ em Ω. O Teorema de Lagrange nos dá condições necessárias que um candidato a mínimo de tal problema tem que satisfazer.

Começamos com um caso mais simples, de funções definidas no plano e onde há apenas uma restrição. Sejam f e g funções do $\mathbb{R}^2$ em $\mathbb{R}$, e suponha que $\mathbf{x}^* \in \mathbb{R}^2$ é tal que

$$f(\mathbf{x}^*) = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^2,\ g(\mathbf{x}) = 0\}.$$

Apesar de termos imposto $\mathbf{x}^*$ como mínimo global, isto é somente uma simplificação. A teoria não se modifica em nada se $\mathbf{x}^*$ for somente mínimo local.

Suponha agora que $\nabla g(\mathbf{x}^*) \neq 0$ e considere $\mathbf{r} : \mathbb{R} \to \mathbb{R}^2$ uma parametrização para a curva de nível determinada por $g(x, y) \equiv 0$, e tal que $\mathbf{r}(0) = \mathbf{x}^*$. A existência de tal parametrização (local) pode ser justificada através do teorema da função implícita. Pela regra da cadeia,

$$0 = \frac{d}{dt} g(\mathbf{r}(t)) = \nabla g(\mathbf{r}(t)) \cdot \mathbf{r}'(t)$$

concluímos que $\nabla g(\mathbf{r}(t))$ é ortogonal a $\mathbf{r}'(t)$. De forma análoga, temos que $f(\mathbf{r}(t))$ atinge seu máximo em $t = 0$. Logo, $df(\mathbf{r}(0))/dt = 0$, e como

$$\frac{d}{dt} f(\mathbf{r}(t)) = \nabla f(\mathbf{r}(t)) \cdot \mathbf{r}'(t),$$

temos que $\nabla f(\mathbf{x}^*)$ é também ortogonal a $\mathbf{r}'(0)$. Em duas dimensões, se $\nabla g(\mathbf{x}^*)$ e $\nabla f(\mathbf{x}^*)$ são ortogonais a $\mathbf{r}'(0)$ isto significa que $\nabla g(\mathbf{x}^*)$ e $\nabla f(\mathbf{x}^*)$ são colineares, i.e., existe $\lambda^* \in \mathbb{R}$ tal que

$$\nabla g(\mathbf{x}^*) = \lambda^* \nabla f(\mathbf{x}^*).$$

O caso mais geral, em dimensões maiores e com número arbitrário de restrições, é enunciado no teorema abaixo.

TEOREMA 10.6.1 (Lagrange). *Sejam $f, g_1, \dots, g_k$ funções reais definidas em $\mathbb{R}^m$ com derivadas contínuas. Suponha que exista $\delta > 0$ e $\mathbf{x}^* \in \mathbb{R}^m$ e tal que*

$$f(\mathbf{x}^*) = \min\{f(\mathbf{x}) :\ \mathbf{x} \in B_\delta(\mathbf{x}^*)\ \text{ e } g_1(\mathbf{x}) = \cdots = g_k(\mathbf{x}) = 0\}.$$

Então existem números $\mu, \lambda_1^*, \dots, \lambda_k^*$ não todos nulos e tais que

$$\mu \nabla f(\mathbf{x}^*) = \lambda_1^* \nabla g_1(\mathbf{x}^*) + \cdots + \lambda_k^* \nabla g_k(\mathbf{x}^*). \tag{10.6.1}$$

Além disto, se $\{\nabla g_1(\mathbf{x}^*), \dots, \nabla g_k(\mathbf{x}^*)\}$ é linearmente independente, então pode-se tomar $\mu = 1$.

DEMONSTRAÇÃO. Supondo que $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{n-k} \times \mathbb{R}^k$ e $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^k$ seja tal que $\nabla_{\mathbf{x}_2} \mathbf{g}(\mathbf{x}^*)$ seja invertível, então existe $\mathbf{h} : \mathbb{R}^{n-k} \to \mathbb{R}^k$ tal que $\mathbf{g}\big(\mathbf{x}_1, \mathbf{h}(\mathbf{x}_1)\big) = \mathbf{0}$ para todo $\mathbf{x}_1 \in \mathbb{R}^{n-k}$. Seja então $F : \mathbb{R}^{n-k} \to \mathbb{R}$ dada por

$$F(\mathbf{x}_1) = f\big(\mathbf{x}_1, \mathbf{h}(x_1)\big). \tag{10.6.2}$$

Tem-se imediatamente que

$$f(\mathbf{x}_1^*, \mathbf{x}_2^*) = F(\mathbf{x}_1^*) = \max_{\mathbf{x}_1 \in \mathbb{R}^{n-k}} F(\mathbf{x}_1).$$

Obtemos então pela regra da cadeia que

$$\mathbf{0} = \nabla F(\mathbf{x}_1^*) = \nabla_{\mathbf{x}_1} f(\mathbf{x}_1^*, \mathbf{x}_2^*) + \nabla_{\mathbf{x}_2} f\big(\mathbf{x}_1^*, \mathbf{x}_2^*\big) \cdot \nabla_{\mathbf{x}_1} \mathbf{h}(\mathbf{x}_1^*)$$

Como $\mathbf{g}\big(\mathbf{x}_1, \mathbf{h}(\mathbf{x}_1)\big) = \mathbf{0}$, então

$$\nabla_{\mathbf{x}_1} \mathbf{g}(\mathbf{x}_1^*, \mathbf{x}_2^*) + \nabla_{\mathbf{x}_2} \mathbf{g}\big(\mathbf{x}_1^*, \mathbf{x}_2^*\big) \nabla_{\mathbf{x}_1} \mathbf{h}(\mathbf{x}_1^*) = \mathbf{0}$$

e portanto $\nabla_{\mathbf{x}_1} \mathbf{h}(\mathbf{x}_1^*) = -\big[\nabla_{\mathbf{x}_2} \mathbf{g}\big(\mathbf{x}_1^*, \mathbf{x}_2^*\big)\big]^{-1} \nabla_{\mathbf{x}_1} \mathbf{g}(\mathbf{x}_1^*, \mathbf{x}_2^*)$. Obtemos então

$$\nabla_{\mathbf{x}_1} f(\mathbf{x}_1^*, \mathbf{x}_2^*) = \boldsymbol{\lambda} \cdot \nabla_{\mathbf{x}_1} \mathbf{g}(\mathbf{x}_1^*, \mathbf{x}_2^*)$$

onde $\boldsymbol{\lambda} = \nabla_{\mathbf{x}_2} f(\mathbf{x}_1^*, \mathbf{x}_2^*) \cdot [\nabla_{\mathbf{x}_2} \mathbf{g}(\mathbf{x}_1^*, \mathbf{x}_2^*)]^{-1}$. Note entretanto que

$$\nabla_{\mathbf{x}_2} f(\mathbf{x}_1^*, \mathbf{x}_2^*) = \boldsymbol{\lambda} \cdot [\nabla_{\mathbf{x}_2} \mathbf{g}\big(\mathbf{x}_1^*, \mathbf{x}_2^*\big)]^{-1},$$

e então

$$\nabla_{\mathbf{x}} f\big(\mathbf{x}^*\big) = \boldsymbol{\lambda} \cdot \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}^*). \tag{10.6.3}$$

□

Observação. Note que uma forma compacta de descrever as condições de Lagrange vem através da definição do Lagrangeano $\mathcal{L} : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ dado por

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda} \cdot \mathbf{g}(\mathbf{x}), \tag{10.6.4}$$

onde $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k) \in \mathbb{R}^k$ e $\mathbf{g}(\mathbf{x}) = \big(g_1(\mathbf{x}), \dots, g_k(\mathbf{x})\big) \in \mathbb{R}^k$. Então as condições necessárias (10.6.3) para $\mathbf{x}^*$ ser ponto de mínimo são dadas por $\nabla \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = 0$, i.e.,

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \nabla_{\mathbf{x}} f\big(\mathbf{x}^*\big) - \boldsymbol{\lambda}^* \cdot \nabla_{\mathbf{x}} \mathbf{g}(\mathbf{x}^*) = \mathbf{0}, \quad \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*) = -\mathbf{g}(\mathbf{x}^*) = \mathbf{0}. \tag{10.6.5}$$

Os números $\lambda_1^*, \dots, \lambda_k^*$ acima são conhecidos por *multiplicadores de Lagrange*, e em muitas aplicações têm significado próprio.

Apresentamos, seguindo **[8]**, alguns argumentos que indicam o porquê do resultado valer. A argumentação é toda baseada na aproximação dada pelo Teorema do Valor Médio:

$$g_i(\mathbf{x}^* + \mathbf{s}) \approx g_i(\mathbf{x}^*) + \nabla g_i(\mathbf{x}^*) \cdot \mathbf{s}, \tag{10.6.6}$$

Diremos que $\mathbf{s} \in \mathbb{R}^m$ é uma *direção factível* se $\mathbf{x}^* + \mathbf{s}$ satisfizer todas as restrições, i.e., $g_i(\mathbf{x}^* + \mathbf{s}) = 0$ para todo $i = 1, \dots, k$. Logo, para termos $\mathbf{s}$ factível, temos $g_i(\mathbf{x}^* + \mathbf{s}) = g_i(\mathbf{x}^*) = 0$. Logo, por (10.6.6), vemos que $\mathbf{s}$ é factível se e somente se

$$\nabla g_i(\mathbf{x}^*) \cdot \mathbf{s} = 0 \text{ para } i = 1, \dots, k. \tag{10.6.7}$$

Note que a derivada direcional da f em $\mathbf{x}^*$ na direção factível $\mathbf{s}$ não pode ser negativa, pois isto indicaria que $f(\mathbf{x}^* + \mathbf{s}) < f(\mathbf{x}^*)$, uma contradição com $\mathbf{x}^*$ ser mínimo local. Logo

$$\nabla f(\mathbf{x}^*) \cdot \mathbf{s} < 0 \tag{10.6.8}$$

não pode ocorrer junto com (10.6.7).

A condição necessária e suficiente para que (10.6.7) e (10.6.8) nunca ocorram simultaneamente é que existam $\lambda_1^*, \dots, \lambda_k^*$ tais que

$$\nabla f(\mathbf{x}^*) = \lambda_1^* \nabla g_1(\mathbf{x}^*) + \cdots + \lambda_k^* \nabla g_k(\mathbf{x}^*).$$

É imediato checar que a condição é suficiente. Para ver que é também necessária, suponha que (10.6.7) e (10.6.8) n ao ocorram simultaneamente para nenhuma direçã factível $\mathbf{s}$, e considere P o subespaço vetorial formado por $\nabla g_1(\mathbf{x}^*), \dots, \nabla g_k(\mathbf{x}^*)$. Se $\nabla f(\mathbf{x}^*) \notin P$ então existe $\mathbf{u}$ não nulo e ortogonal a P. Por (10.6.7), $\mathbf{u}$ é factível. Seja $\mathbf{v} \in P$ tal que $\mathbf{u} + \mathbf{v} = \nabla f(\mathbf{x}^*)$. Então $\mathbf{s} = -\mathbf{u}$ satisfaz (10.6.7) e (10.6.8), uma contradição.

10.6.1.2. *Restrições de desigualdade.* Uma outra situação de minimização com restrições ocorre quando as restrições são dadas por desigualdades, e não mais como acima. neste caso temos o Teorema de Kuhn–Tucker, dado abaixo.

TEOREMA 10.6.2 (Kuhn–Tucker). *Sejam $f, h_1, \dots, h_k$ funções reais definidas em $\mathbb{R}^m$ com derivadas contínuas. Suponha que existam $\delta > 0$ e $\mathbf{x}^* \in \mathbb{R}^m$ tais que*

$$f(\mathbf{x}^*) = min\{f(x) : \mathbf{x} \in B_\delta(\mathbf{x}^*) \text{ e } h_1(\mathbf{x}) \geq 0, \cdots, h_k(\mathbf{x}) \geq 0\}.$$

Então as seguintes afirmativas são verdadeiras:

(1) existem números $\mu, \lambda_1^*, \dots, \lambda_k^*$ não todos nulos e tais que

$$\mu \nabla f(\mathbf{x}^*) = \lambda_1^* \nabla h_1(\mathbf{x}^*) + \cdots + \lambda_k^* \nabla h_k(\mathbf{x}^*).$$

(2) seja $i \in \{1, \dots, k\}$ tal que $h_i(\mathbf{x}^*) > 0$. Então pode-se impor $\lambda_i^* = 0$.

(3) se o conjunto $V = \{\nabla h_i(\mathbf{x}^*) : h_i(\mathbf{x}^*) = 0, \text{ onde } 1 \leq i \leq k\}$ é linearmente independente, então pode-se tomar $\mu = 1$ e $\lambda_1^* \geq 0, \dots, \lambda_k^* \geq 0$.

Seguiremos aqui apresentar os argumentos que apresentamos justificando o Teorema de Lagrange, seguindo novamente [**8**]. Podemos supor, sem perda de generalidade, que $h_i(x^*) = 0$ para todo $i = 1, \dots, k$. De fato, se $h_i(x^*) > 0$, então podemos tomar uma vizinhança de $\mathbf{x}^*$ tal que h_i seja estritamente positiva nesta vizinhança (pois h_i é positiva). É este o motivo podemos tomar $\lambda_i^* = 0$ nestes casos.

Se $\mathbf{s}$ é direção factível, i.e., satisfaz as desigualdades, então, $h_i(\mathbf{x}^* + \mathbf{s}) \geq 0$. Logo, como $h_i(\mathbf{x}^*) = 0$, por (10.6.6), temos

$$\nabla h_i(\mathbf{x}^*) \cdot \mathbf{s} \geq 0 \text{ para } i = 1, \dots, k. \tag{10.6.9}$$

Desde que a derivada direcional da f em $\mathbf{x}^*$ na direção factível $\mathbf{s}$ não pode ser negativa, temos que

$$\nabla f(\mathbf{x}^*) \cdot \mathbf{s} < 0 \tag{10.6.10}$$

não pode ocorrer junto com (10.6.9).

A condição necessária e suficiente para que (10.6.9) e (10.6.10) nunca ocorram simultaneamente é que existam $\lambda_1^* \geq 0, \dots, \lambda_k^* \geq 0$ tais que

$$\nabla f(\mathbf{x}^*) = \lambda_1^* \nabla h_1(\mathbf{x}^*) + \cdots + \lambda_k^* \nabla h_k(\mathbf{x}^*).$$

Novamente, é fácil checar que a condição é suficiente. Para ver que $\nabla f(\mathbf{x}^*)$ é combinação linear de $\nabla h_1(\mathbf{x}^*), \dots, \nabla h_k(\mathbf{x}^*)$, basta argumentar como no caso de restrição estrita. Sem perda de generalidade suponha agora que $\lambda_1^* < 0$. Seja P o espaço vetorial formado pelas combinações lineares de $\{\nabla h_2(\mathbf{x}^*), \dots, \nabla h_k(\mathbf{x}^*)\}$. Seja agora $\mathbf{s}$ vetor ortogonal ao subespaço P, e "apontando na mesma direção que $\nabla h_1(\mathbf{x}^*)$", i.e., tal que $\nabla h_1(\mathbf{x}^*) \cdot \mathbf{s} \geq 0$. Por construção temos então que (10.6.9) e (10.6.10) ocorrem com tal $\mathbf{s}$, uma contradição. Logo temos sempre $\lambda_i^* \geq 0$. Note que no argumento acima usamos que $\nabla h_1(\mathbf{x}^*)$ é linearmente independente de $\nabla h_2(\mathbf{x}^*), \dots, \nabla h_k(\mathbf{x}^*)$ (caso contrário $\nabla h_1(\mathbf{x}^*) \in P$). Esta hipótese de independência linear é um artifício desta demonstração, e pode ser eliminada [**8**].

10.6.1.3. *Restrições de positividade.* Seja $f : \mathbb{R}^n \to \mathbb{R}$ "suave" e considere o seguinte problema de otimização com restrições:

$$\min_{\mathbf{x}\in\mathbb{R}^m} f(\mathbf{x}) \quad \text{tal que } \mathbf{x} \geq \mathbf{0}. \tag{10.6.11}$$

e suponha que $\mathbf{x}^*$ resolva (10.6.11). Seja $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ base canônica de $\mathbb{R}^m$ e $i \in \{1, \dots, m\}$. Seja $f_i : [0, \infty) \to \mathbb{R}$ dada por

$$f_i(\tau) = f(x_1^*, x_2^*, \dots, x_{i-1}^*, \tau, x_{i+1}^*, \dots, x_m^*).$$

Note então que f_i atinge mínimo em $\tau = x_i^*$. Se $x_i = 0$, então $f_i'(x_i^*) \geq 0$, caso contrário, $f_i'(x_i^*) \geq 0$. Podemos escrever estas condições de forma resumida por

$$f_i'(x_i^*) \geq 0, \qquad x_i^* f_i'(x_i^*) = 0, \qquad x_i^* \geq 0.$$

Como $f_i'(x_i^*) = \partial f/\partial x_i(x_i^*)$, reescrevemos as condições acima por

$$\nabla f(\mathbf{x}^*) \geq \mathbf{0}, \qquad \nabla f(\mathbf{x}^*) \cdot \mathbf{x}^* = 0, \qquad \mathbf{x}^* \geq \mathbf{0}.$$

10.6.1.4. *Restrições de positividade com desigualdade.* Com a introdução de variáveis de folga, pode-se reduzir o problema com restrições de desigualdade a problemas com restrições de positividade. Seja $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^k$ "suave" e $\boldsymbol{b} \in \mathbb{R}^k$, com $k \geq 1$, e

$$\min_{\mathbf{x}\in\mathbb{R}^m} f(\mathbf{x}) \quad \text{tal que } \mathbf{g}(\mathbf{x}) \geq \boldsymbol{b},\ \mathbf{x} \geq \mathbf{0}. \tag{10.6.12}$$

Introduzindo a *variável de folga* $\boldsymbol{s} = \mathbf{g}(\mathbf{x}) - \boldsymbol{b}$, obtemos um problema na forma:

$$\min_{\mathbf{x}\in\mathbb{R}^m} f(\mathbf{x}) \quad \text{tal que } \boldsymbol{s} + \boldsymbol{b} - \mathbf{g}(\mathbf{x}) = \mathbf{0}, \quad \boldsymbol{s} \geq \mathbf{0}, \quad \mathbf{x} \geq \mathbf{0}. \tag{10.6.13}$$

Para obtermos condições suficientes para pontos de ótimo, basta combinar a demonstração do Teorema 10.6.1 com as ideias da Seção 10.6.1.3. O lagrangeano será agora dado por

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda} \cdot \big(\boldsymbol{s} + \boldsymbol{b} - \mathbf{g}(\mathbf{x})\big),$$

e as condições necessárias de otimalidade são

$$\begin{gathered}
\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \nabla_{\mathbf{x}} f - \boldsymbol{\lambda} \cdot \nabla_{\mathbf{x}} \mathbf{g} \geq \mathbf{0}, \qquad \nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \cdot \mathbf{x} = (\nabla_{\mathbf{x}} f - \boldsymbol{\lambda} \cdot \nabla_{\mathbf{x}} \mathbf{g}) \cdot \mathbf{x} = 0, \\
\nabla_{\boldsymbol{s}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = \boldsymbol{\lambda} \geq \mathbf{0}, \qquad \nabla_{\boldsymbol{s}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) \cdot \boldsymbol{s} = \boldsymbol{\lambda} \cdot \boldsymbol{s} = 0, \\
\nabla_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda}) = \boldsymbol{s} + \boldsymbol{b} - \mathbf{g} = \mathbf{0}, \qquad \boldsymbol{s} \geq \mathbf{0}, \qquad \mathbf{x} \geq \mathbf{0}.
\end{gathered}$$

computed at $\mathbf{x} = \mathbf{x}^*$, $\boldsymbol{s} = \boldsymbol{s}^*$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$. Eliminando $\boldsymbol{s} = \mathbf{g} - \boldsymbol{b}$, obtém-se finalmente

$$\begin{gathered}
\nabla_{\mathbf{x}} f - \boldsymbol{\lambda} \cdot \nabla_{\mathbf{x}} \mathbf{g} \geq \mathbf{0}, \qquad (\nabla_{\mathbf{x}} f - \boldsymbol{\lambda} \cdot \nabla_{\mathbf{x}} \mathbf{g}) \cdot \mathbf{x} = 0, \\
\boldsymbol{\lambda} \geq \mathbf{0}, \qquad \boldsymbol{\lambda} \cdot (\mathbf{g} - \boldsymbol{b}) = 0, \qquad \mathbf{g} - \boldsymbol{b} \geq \mathbf{0}, \qquad \mathbf{x} \geq \mathbf{0}.
\end{gathered}$$

10.6.2. Condições de segunda ordem. Considere o caso de restrições de igualdade somente, como no Teorema 10.6.1. Considere $\mathbf{x}^*$ e $\boldsymbol{\lambda}^*$ tal que (10.6.5) valha. Temos por Taylor, para $\mathbf{s}$ admissível, que

$$f(\mathbf{x}^*+\mathbf{s}) = \mathcal{L}(\mathbf{x}^*+\mathbf{s},\boldsymbol{\lambda}^*) \approx \mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)+\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)\cdot\mathbf{s}+\frac{1}{2}\mathbf{s}\cdot W^*\mathbf{s} = f(\mathbf{x}^*)+\frac{1}{2}\mathbf{s}\cdot W^*\mathbf{s} \tag{10.6.14}$$

onde $\mathcal{L}$ foi definida em (10.6.1), e W^* é a Hessiana de $\mathcal{L}$ em relação a $\mathbf{x}$, i.e,

$$\mathbf{s}\cdot W^*\mathbf{s} = \sum_{i,j=1}^{m}\frac{\partial^2\mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)}{\partial x_i\partial x_j}\mathbf{s}_i\mathbf{s}_j.$$

Como $\mathbf{x}^*$ é mínimo, obtemos de (10.6.14) a condição necessária $\mathbf{s}\cdot W^*\mathbf{s}\ge 0$ para todo $\mathbf{s}$ admissível, i.e., todo $\mathbf{s}$ tal que satisfaça (10.6.7). Similarmente, de (10.6.14) obtemos a condição suficiente para que $\mathbf{x}^*$ seja mínimo estrito, que é

$$\mathbf{s}\cdot W^*\mathbf{s} > 0$$

para todo $\mathbf{s}$ admissível.

Para o caso de restrições com desigualdade, seja $\mathbf{s}$ admissível, i.e., tal que $h_i(\mathbf{x}^*+\mathbf{s})\ge 0$. Então definimos agora

$$\mathcal{L}(\mathbf{x},\boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda}\cdot\mathbf{h}(\mathbf{x}),$$

onde $\mathbf{h}(\mathbf{x}) = \big(h_1(\mathbf{x}),\dots,h_k(\mathbf{x})\big)\in\mathbb{R}^m$. Portanto

(10.6.15)

$$\begin{aligned} f(\mathbf{x}^*+\mathbf{s}) = \mathcal{L}(\mathbf{x}^*+\mathbf{s},\lambda^*)+\boldsymbol{\lambda}^*\cdot\mathbf{h}(\mathbf{x}^*+\mathbf{s}) &\approx \mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)+\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}^*,\boldsymbol{\lambda}^*)\cdot\mathbf{s}+\frac{1}{2}\mathbf{s}\cdot W^*\mathbf{s}+\boldsymbol{\lambda}^*\cdot\mathbf{h}(\mathbf{x}^*+\mathbf{s}) \\ &= f(\mathbf{x}^*) - \boldsymbol{\lambda}^*\cdot\mathbf{h}(\mathbf{x}^*) + \frac{1}{2}\mathbf{s}\cdot W^*\mathbf{s} + \sum_{i=1}^{k}\lambda_i^* h_i(\mathbf{x}^*+\mathbf{s}). \end{aligned}$$

Como $\lambda_i^*\ge 0$, podemos definir $\mathcal{A}_+ = \{i \text{ inteiro}: \lambda_i^* > 0,\ 1\le i\le k\}$ e $\mathcal{A}_0^* = \{i \text{ inteiro}: \lambda_i^* = 0,\ 1\le i\le k\}$. Logo, usando que $h_i(\mathbf{x}^*) = 0$,

$$f(\mathbf{x}^*+\mathbf{s}) - f(\mathbf{x}^*) \approx \frac{1}{2}\mathbf{s}\cdot W^*\mathbf{s} + \sum_{i\in\mathcal{A}_+}\lambda_i^*\nabla h_i(\mathbf{x}^*)\cdot\mathbf{s},$$

Buscamos então condições suficientes para que o lado direito da equação acima seja positivo. Se $\mathbf{s}$ é factível, então vale (10.6.9). Se $\nabla h_i(\mathbf{x}^*)\cdot\mathbf{s} > 0$, para $\mathbf{s}$ "pequeno o suficiente" temos que

$$\frac{1}{2}\mathbf{s}\cdot W^*\mathbf{s} + \sum_{i\in\mathcal{A}_+}\lambda_i^*\nabla h_i(\mathbf{x}^*)\cdot\mathbf{s} > 0$$

e $f(\mathbf{x}^*+\mathbf{s}) > f(\mathbf{x}^*)$. Entretanto, se $\nabla h_i(\mathbf{x}^*)\cdot\mathbf{s} = 0$, uma condição suficiente para que $f(\mathbf{x}^*+\mathbf{s})\ge f(\mathbf{x}^*)$ é exigir que $\mathbf{s}\cdot W^*\mathbf{s} > 0$ para todo $\mathbf{s}$ admissível e ortogonal a $\nabla h_i(\mathbf{x}^*)$ para $i\in\mathcal{A}_+$.

10.7. Exercícios

EXERCÍCIO 10.1. Extenda para $f : \mathbb{R}^m \to \mathbb{R}^n$ o conceito de limite de funções num determinado ponto $\mathbf{x}_0 \in \mathbb{R}^m$.

EXERCÍCIO 10.2. Seja f e g funções de $\mathbb{R}^m \to \mathbb{R}^n$, diferenciáveis em $\mathbf{x} \in \mathbb{R}^m$. Mostre usando a definição de derivadas que $(f+g)'(\mathbf{x}) = f'(\mathbf{x}) + g'(\mathbf{x})$. Seja $h : \mathbb{R}^m \to \mathbb{R}$ também diferenciável em $\mathbf{x} \in \mathbb{R}^m$. Mostre usando a definição de derivadas que $(hg)'(\mathbf{x}) = h'(\mathbf{x})g(\mathbf{x}) + h(\mathbf{x})g'(\mathbf{x})$.

EXERCÍCIO 10.3. Sejam $a < b$ números reais, e $f : [a,b] \to \mathbb{R}$ contínua em $[a,b]$ e diferenciável em (a,b). Mostre que entre duas raízes consecutivas de f' existe no máximo uma raiz de f.

EXERCÍCIO 10.4. Seja $f : \mathbb{R}^2 \to \mathbb{R}$ dada por

$$f(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & \text{para } (x,y) \neq (0,0), \\ 0 & \text{para } (x,y) = (0,0). \end{cases}$$

Mostre que a derivada direcional de f em $(0,0)$ com respeito a $\mathbf{u} = (a,b)$ existe e que

$$D_{\mathbf{u}} f(0,0) = \frac{b^2}{a}, \qquad \text{se } a \neq 0.$$

Mostre que f não é contínua e portanto não é diferenciável no $(0,0)$.

EXERCÍCIO 10.5 (Kevmasan [**20**], Example 1.1.1). Mostre que $f : \mathbb{R}^2 \to \mathbb{R}$ dada por

$$f(x,y) = \begin{cases} \dfrac{x^5}{(y-x^2)+x^4} & \text{para } (x,y) \neq (0,0), \\ 0 & \text{para } (x,y) = (0,0). \end{cases}$$

tem todas as derivadas direcionais em $(0,0)$ iguais a zero, mas que f não é diferenciável no $(0,0)$. (Dica: considere $\mathbf{h} = (h,h^2)$ em (10.3.1)).

EXERCÍCIO 10.6. Seja $Q = (0,1) \times (0,1)$. Suponha que $f : Q \to \mathbb{R}$, e $g : Q \to \mathbb{R}$ sejam diferenciáveis em Q. Mostre que se $\mathbf{f}'(x) = \mathbf{g}'(x)$ para todo $\mathbf{x} \in Q$, então existe constante c tal que $f(\mathbf{x}) = g(\mathbf{x}) + c$ para todo $\mathbf{x} \in Q$.

EXERCÍCIO 10.7. Seja $\Omega \subset \mathbb{R}^n$, aberto e com a seguinte propriedade: existe $\mathbf{x}^* \in \Omega$ tal que para todo $\mathbf{x} \in \Omega$, a reta $S_{\mathbf{x}} = \{t\mathbf{x} + (1-t)\mathbf{x}^* : t \in [0,1]\}$ está contida em Ω, i.e., $S_{\mathbf{x}} \subset \Omega$. Seja $f : \Omega \to \mathbb{R}$ função diferenciável em Ω e tal que todas as derivadas parciais de $f(\mathbf{x})$ são nulas, para todo $\mathbf{x} \in \Omega$. Mostre que f é constante.

EXERCÍCIO 10.8. Seja $B = \{\mathbf{x} \in \mathbb{R}^m : \|\mathbf{x}\| \leq 1\}$ e $f : B \to \mathbb{R}$ função contínua em B, diferenciável no interior de B e tal que $f \equiv 0$ na fronteira de B. Mostre que f tem ponto crítico no interior de B.

EXERCÍCIO 10.9 (Mínimos Quadrados). Considere para $i = 1, \dots, n$ os pontos $(x_i, y_i) \in \mathbb{R}^2$, e seja $p : \mathbb{R} \to \mathbb{R}$ dada por $p(x) = ax^2 + bx + c$ tal que a, b e c minimizam o erro

$\sum_{i=1}^n |p(x_i) - y_i|^2$. Mostre que a, b e c satisfazem as equações

$$a \sum_{i=1}^n x_i^4 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i^2 y_i,$$
$$a \sum_{i=1}^n x_i^3 + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i = \sum_{i=1}^n x_i y_i,$$
$$a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i + cn = \sum_{i=1}^n y_i.$$

Exercício 10.10. Seja $A \subset \mathbb{R}^m$ compacto, e A° o conjunto dos pontos interiores de A. Seja $f : A \to \mathbb{R}$ duas vezes diferenciável, com derivadas contínuas, em A°. Suponha ainda que f se anule em toda a fronteira de A, e que f'' seja negativa definida para todo ponto em A°. Mostre que $f(\mathbf{x}) > 0$ para todo $\mathbf{x} \in A^\circ$.

Exercício 10.11. Mostre, usando o Teorema 10.4.3, que $(0,0)$ é ponto de sela de $f(x,y) = x^2 - y^2$, e ponto de mínimo estrito local de $f(x,y) = x^2 + y^2$.

Exercício 10.12. Sejam as funções $f : \mathbb{R} \to \mathbb{R}$ e $g : \mathbb{R} \to \mathbb{R}$ duas vezes diferenciáveis, com as segundas derivadas contínuas. Suponha que o zero seja ponto de mínimo estrito de f e g, e que $f(0) = g(0) = 1$. Seja $\varphi : \mathbb{R}^2 \to \mathbb{R}$ dada por $\varphi(x,y) = f(x)g(y)$. O que podemos afirmar sobre a Hessiana de φ em $(0,0)$ (nada pode ser afirmado, ela é indefinida, positiva definida, positiva semi-definida, negativa definida, etc)? O que podemos afirmar sobre o ponto $(0,0)$ em relação à φ (nada pode ser afirmado, é ponto de máximo, de máximo estrito, de mínimo, de mínimo estrito, de sela, etc)? Justifique suas respostas.

Exercício 10.13. Seja $\mathbf{f} : \mathbb{R} \to \mathbb{R}^3$ diferenciável e tal que $\|\mathbf{f}(t)\| = 1$ para todo $t \in \mathbb{R}$. Mostre então que $\mathbf{f}'(t) \cdot \mathbf{f}(t) = 0$. O vetor $\mathbf{f}'(t)$ é o *vetor tangente da curva* $\mathbf{f}$ *em* t.

Exercício 10.14. Seja $\Omega \subseteq \mathbb{R}^m$ aberto e $f : \Omega \to \mathbb{R}$ diferenciável em $\mathbf{x} \in \Omega$. Seja $\nabla f(\mathbf{x}) = (\partial f/\partial x_1, \dots, \partial f/\partial x_m)(\mathbf{x}) \in \mathbb{R}^m$. Supondo que $\mathbf{x}$ não é ponto crítico de f, mostre que a derivada direcional $D_{\mathbf{u}} f(\mathbf{x})$ atinge seu máximo quando $\mathbf{u} = c\nabla f(\mathbf{x})$ para algum $c > 0$. O vetor ∇f é chamado de *vetor gradiente de* f, e dá a direção de "maior crescimento" da função f no ponto $\mathbf{x}$.

CAPíTULO 11

Equações Diferenciais Ordinárias

[1] Neste capítulo discutiremos aspectos relacionados a soluções de equações diferenciais ordinárias (EDOs).

11.1. Equações lineares de primeira ordem

Considere a equação diferencial ordinária de primeira ordem

$$y' + p(x)y = g(x), \tag{11.1.1}$$

$$y(x_0) = y_0, . \tag{11.1.2}$$

Se f e g forem continuas, então existe um única solução para o problema acima num intervalo aberto contendo x_0. Esta solução é dada como $y = y_h + y_p$, soma de uma solução homogenea y_h com uma solucao particular Y_p. Uma solução homogênea é dada por qualquer função que satisfaça $y_h' + p(x)y_h = 0$. Por outro lado, uma solução particular tem que satisfazer (11.1.1).

Para encontrar uma solução particular, usamos *fatores de integração.* Seja $\mu(x)$ tal que

$$\mu'(x) = p(x)\mu(x). \tag{11.1.3}$$

Suponha por um momento que saibamos resolver a equação acima. Neste caso, multiplicando (11.1.1) por μ, obtemos

$$y'\mu + p(x)y\mu = g(x)\mu.$$

Porém, $(y\mu)' = y'\mu + y\mu'$, e portanto

$$(y\mu)' = -p(x)y\mu + g(x)\mu + y\mu' = [-p(x)y + g(x) + yp(x)]\mu = g(x)\mu.$$

Then

$$y(x) = \frac{1}{\mu(x)} \int_{x_0}^{x} g(s)\mu(s)\, ds + y_0.$$

Falta ainda resolver (11.1.3). Suponha por um momento que μ seja estritamente positiva. Então, de

$$\big(\ln(\mu(x))\big)' = \frac{\mu'(x)}{\mu(x)} = p(x)$$

temos

$$\mu(x) = \exp\left(\int p(s)\, ds\right).$$

[1] Última Atualização: 21/09/2020

EXEMPLO 11.1 ([**5**]). Considere o problema de achar $y(t)$ tal que

$$ty' + 2y = 4t^2, y(1) = 2. \tag{11.1.4}$$

Reescrevendo a equação acima na sua forma usual, vemos que $y' + 2y/t = 4t$. Seja $\mu(t)$ tal que

$$\mu(x) = \exp\left(\int \frac{2}{s}\, ds\right) = \exp\big(2\ln t\big) = t^2.$$

Multiplicando a equação (11.1.4) por μ, obtemos

$$y't^2 + 2yt = 4t^3,$$

e então $(yt^2)' = 4t^3$. Integrando, vemos que

$$yt^2 = \int 4s^3\, ds = t^4 + c.$$

Da condição inicial $y(1) = 1$, concluímos que $c = 1$ e

$$y(t) = t^2 + \frac{1}{t}.$$

11.2. Equações de segunda ordem com coeficientes constantes

Considere agora uma equação da forma

$$ay'' + by' + cy(t) = g(t). \tag{11.2.1}$$

Novamente note que se y_h resolve a equação homogênea $ay_h'' + by_h' + cy_h(t) = 0$, e se y_p é uma solução particular, i.e., resolve (11.2.1), então $y_h + y_p$ também é solução de (11.2.1).

Consideraremos primeiro o problema homogêneo

$$ay'' + by' + cy(t) = g(t). \tag{11.2.2}$$

supondo $g = 0$ e buscando soluções *linearmente independentes* y_1^h e y_2^h (i.e., se existirem constantes c_1 e c_2 tais que $c_1y_1^h(t) + c_2y_2^h(t)$ para todo x, então $c_1 = c_2 = 0$). Por substituição, veja que se y_1^h e y_2^h são soluções de (11.2.2), então $c_1y_1^h(t) + c_2y_2^h(t)$ também o são.

Substituindo $y(t) = e^{rt}$ na equação, onde r é constante, obtemos

$$(ar^2 + br + c)e^{rt} = 0$$

para todo x. Como e^{rx} não se anula, temos que r é raiz de

$$ar^2 + br + c = 0, \tag{11.2.3}$$

são a chamada *equação característica* de equação diferencial (11.2.2). Note que se (11.2.3) tiver raízes r_1, r_2 diferentes (independente de serem reais ou não), temos que e^{r_1t} e e^{r_2t} serão soluções linearmente independentes da equação homogênea (11.2.2). A solução geral para (11.2.2) é dada por $y(t) = c_1e^{r_1t} + c_2e^{r_2t}$, e para determinar as constantes são necessárias duas condições, normalmente dadas por

$$y(0) = y_0, \qquad y'(0) = y_0'.$$

Substituindo, temos

$$y(0) = c_1 + c_2 = y_0, \qquad y(0) = c_1r_1 + c_2r_2 = y_0'.$$

As duas condições acima determinam de forma única c_1 e c_2, quaisquer que sejam os valores de y_0 e y_0'.

No caso em que as raízes da equação característica(11.2.3) sejam complexas conjugadas, i.e., $r_1 = \lambda + i\mu$ e $r_2 = \lambda - i\mu$, com $\mu \neq 0$ onde $i^2 = -1$, temos pela fórmula de Euler que

$$e^{r_1 t} = e^{\lambda t}\big(\cos(\mu t) + i\sin(\mu t)\big), \qquad e^{r_2 t} = e^{\lambda t}\big(\cos(\mu t) - i\sin(\mu t)\big).$$

Somando e subtraindo as soluções acima obtemos novas soluções linearmente independentes, i.e.,

$$e^{r_1 t} + e^{r_2 t} = 2e^{\lambda t}\cos(\mu t), \qquad e^{r_1 t} - e^{r_2 t} = 2ie^{\lambda t}\sin(\mu t).$$

Como 2 e $2i$ são constantes, podemos buscar a solução geral como combinação linear das funções reais

$$u(t) = e^{\lambda t}\cos(\mu t), \qquad v(t) = e^{\lambda t}\sin(\mu t).$$

Estas serão também linearmente independentes, desde que $\mu \neq 0$.

Então, dados y_0 e y_0', buscamos $y(t) = c_1 u(t) + c_2 v(t)$ onde

$$y(0) = c_1 u(0) + c_2 v(0) = c_1 \mu = y_0, \qquad y'(0) = c_1 u'(0) + c_2 v'(0) = y_0'.$$

Note que $c_1 = y_0/\mu$. Para determinar c_2, é preciso determinar o valor de $u'(0)$ e $v'(0)$.

O último caso de interesse é quando $r = r_1 = r_2$, neste caso tentamos soluções da forma e^{rt} e te^{rt}, i.e., $y(t) = c_1 e^{rt} + c_2 e^{rt}$, onde c_1 e c_2 são determinadas pelas condições iniciais.

11.3. Sistemas homogêneos lineares de duas equações com coeficientes constantes

Considere o sistema de equações

$$x_1' = ax_1 + bx_2,$$
$$x_2' = cx_1 + dx_2,$$

com condições iniciais $x_1(0) = x_1^0$ e $x_2(0) = x_2^0$. O sistema pode ser escrito na forma matricial

$$\mathbf{x}' = A\mathbf{x}, \qquad \mathbf{x}(0) = \mathbf{x}_0, \tag{11.3.1}$$

onde

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad \mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}, \qquad \mathbf{x}_0 = \begin{pmatrix} x_1^0 \\ x_2^0 \end{pmatrix}.$$

Há duas formas possíveis de se resolver o sistema acima. A primeira é por substituição. Se $b = c = 0$, então na verdade temos duas equações desacopladas, simples de serem resolvidas. Se $b \neq 0$ (o caso $c \neq 0$ é análogo), então

$$x_2 = \frac{1}{b}x_1' - \frac{a}{b}x_1,$$

e então

$$x_2' = \frac{1}{b}x_1'' - \frac{a}{b}x_1'.$$

Portanto

$$\frac{1}{b}x_1'' - \frac{a}{b}x_1' = cx_1 + dx_2$$

e obtemos o sistema de segunda ordem para x_1:

$$\frac{1}{b}x_1'' - \frac{a}{b}x_1' - cx_1 - d(\frac{1}{b}x_1' - \frac{a}{b}x_1) = 0.$$

A segunda maneira de se estudar sistemas é usando álgebra linear. Considere $\mathbf{x} = e^{\alpha t}\mathbf{v}$. Então

$$A\mathbf{x} = \mathbf{x}'(t) \iff A(e^{\alpha t}\mathbf{v}) = \alpha e^{\alpha t}\mathbf{v}.$$

Dividindo ambos os lados por $e^{\alpha t}$ obtemos que $\mathbf{x}(t) = e^{\alpha t}\mathbf{v}$ é uma solução para (11.3.1) se e somente se $A\mathbf{v} = \alpha\mathbf{v}$, i.e., $\mathbf{v}$ é autovetor de A com autovalor α. Considere as seguintes situações:

(1) Autovalores reais e diferentes. Sejam $\alpha_1 \neq \alpha_2$ autovalores reais, e sejam $\mathbf{v}_1$ e $\mathbf{v}_2$ autovetores correspondentes. Então estes são linearmente independentes, e podemos buscar soluções da forma $c_1e^{\alpha_1 t}\mathbf{v}_1 + c_2e^{\alpha_2 t}\mathbf{v}_2$. As constantes c_1, c_2 são calculadas tomando-se $t = 0$ e obtendo-se

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{x}_0. \tag{11.3.2}$$

(2) Autovalores complexos. Sejam $\alpha_1 \neq \alpha_2$ autovalores complexos de A. Como A é real, temos que os autovalores são conjugados, i.e., podem ser escritos como $\lambda \pm i\mu$. Além disto, os autovetores também são conjugados. De fato, se $(A - \alpha I)\mathbf{v} = 0$, então $(A - \bar{\alpha}I)\bar{\mathbf{v}} = 0$.

Suponha $\mu \neq 0$, considere o autovalor $\lambda+i\mu$ e o autovetor correspondente $\mathbf{v}_r+i\mathbf{v}_i$, onde $\mathbf{v}_r$ e $\mathbf{v}_i$ são vetores reais. Então

$$e^{(\lambda+i\mu)t}(\mathbf{v}_r + i\mathbf{v}_i) = e^{\lambda t}(\cos\mu t + i\sin\mu t)(\mathbf{v}_r + i\mathbf{v}_i) = e^{\lambda t}(\mathbf{v}_1 + i\mathbf{v}_2),$$

onde $\mathbf{v}_1 = \cos\mu t\mathbf{v}_r - \sin\mu t\mathbf{v}_i$ e $\mathbf{v}_2 = \sin\mu t\mathbf{v}_r + \cos\mu t\mathbf{v}_i$. Pode-se mostrar que $\mathbf{v}_1$ e $\mathbf{v}_2$ são linearmente independentes e podemos tomar $e^{\lambda t}\mathbf{v}_1$ e $e^{\lambda t}\mathbf{v}_2$ como soluções independentes e buscar uma solução geral da forma $c_1e^{\alpha_1 t}\mathbf{v}_1 + c_2e^{\alpha_2 t}\mathbf{v}_2$ resolvendo (11.3.2).

(3) Autovalores repetidos. Se λ é autovalor duplo com dois autovalores linearmente independentes $\mathbf{v}_1$ e $\mathbf{v}_2$, então a solução geral é dada por $e^{\alpha t}(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)$ e (11.3.2). Caso o autoespaço associado a λ seja unidimensional (com autovetor $\mathbf{v}$), então a solução geral é dada por

$$c_1e^{\alpha t}\mathbf{v} + c_2e^{\alpha t}(t\mathbf{v} + \mathbf{u}),$$

onde $(A - \lambda I)\mathbf{u} = \mathbf{v}$.

11.4. Equações de diferenças

A equação de recorrência de ordem n linear homogênea discreta com coeficientes constantes consiste em achar $y : \mathbb{N} \to \mathbb{R}$ tal que

$$y_t = c_1y_{t-1} + c_2y_{t-2} + \cdots + c_dy_{t-d}, \tag{11.4.1}$$

onde $c_d \neq 0$, $t \in \mathbb{N}$, e denotamos $y_t = y(t)$. Para determinar de forma única soluções para equações deste tipo, as condições iniciais $y_0, \ldots, y_{d-1}$ têm que ser dadas.

Note que as somas de soluções são também soluções. Buscamos então soluções do tipo $y_t = r^t$. Substituindo em 11.4.1, obtemos que

$$r^t = c_1r^{t-1} + c_2r^{t-2} + \cdots + c_{d-1}r^{t-d+1} + c_dr^{t-d} = r^{t-d}(c_1r^{d-1} + c_2r^{d-2} + \cdots + c_{d-1}r + c_d).$$

Concluímos então que $y_t = r^t$ é solução se e somente se r é raiz do polinômio de ordem d:

$$r^d - c_1 r^{d-1} - c_2 r^{d-2} + \cdots - c_{d-1} r - c_d = 0, \tag{11.4.2}$$

chamado de polinômio característico. Se todas as d raízes $r_1, \ldots, r_n$ forem distintas, então a solução geral é dada por

$$y_t = a_1 r_1^t + \cdots + a_d r_d^d.$$

onde as constantes $a_1, \ldots, a_d$ são determinadas pelas condições iniciais. Suponha agora que haja uma raíz repetida, por exemplo, suponha que r_1 tenha multiplicidade 3. Então a expressão $ar_1^t + btr_1^t + ct^2 r_1^t$ também será solução.

11.4.1. Relações recursivas de segunda ordem. Considere uma relação da forma

$$y_t = Ay_{t-1} + By_{t-2}.$$

Então a equação característica é dada por

$$r^2 = Ar + B.$$

No caso de raízes $r_1 \neq r_2$ distintas, então a solução geral é da forma $y_t = c_1 r_1^t + c_2 r_2^t$, onde c_1 e c_2 são constantes a serem determinadas pelas condições iniciais. Note que se as raízes forem complexas, nada muda (só temos que trabalhar com números complexos).

No caso de duas raízes iguais $r_1 = r_2$, então a solução é dada por $y_t = c_1 r_1^t + c_2 t r_1^t$. Para ver que tr_1^t é também solução do problema, note que

$$r^2 - Ar - B = (r - r_1)^2 = r^2 - 2rr_1 + r_1^2.$$

i.e., $A = 2r_1$ e $B = -r_1^2$. Então

$$tr_1^t = t(Ar_1^{t-1} + Br^{t-2}) = Atr_1^{t-1} + Btr^{t-2} = A(t-1)r_1^{t-1} + B(t-2)r^{t-2} + Ar_1^{t-1} + 2Br_1^{t-2}$$

Entretanto, $Ar_1 + 2B = 2r_1 r_1 + 2(-r_1^2) = 0$, e portanto,

$$tr_1^t = A(t-1)r_1^{t-1} + B(t-2)r^{t-2}.$$

e então tr_1^t é solução.

APÊNDICE A

Uma introdução não tão formal aos fundamentos da matemática

[1]

A matemática se baseia na argumentação lógica. Outras áreas do conhecimento, talvez todas, podem também reclamar para si tal propriedade, Entretanto a matemática é o *próprio* desenvolvimento da argumentação formal, é a "lógica aplicada."

Este aspecto da matemática tem consequências interessantes; seus resultados independem da época, cultura e região em que foram gerados. O Teorema de Pitágoras, demonstrado por fanáticos matemáticos (os pitagóricos), cerca de 500 A.C., será válido em qualquer lugar e época (http://mathworld.wolfram.com/PythagoreanTheorem.html).

Outras áreas têm teorias "exatas" que são na verdade aproximações da realidade, com "validade" somente sob determinadas condições (por exemplo, teoria da relatividade versus física quântica). Mesmo certas definições podem mudar. Como exemplo, em 1997 a unidade de tempo *segundo* foi definida mais uma vez (http://en.wikipedia.org/wiki/Second). Quanto ao pobre quilograma, bem, este ainda busca uma definição adequada aos nossos tempos (http://en.wikipedia.org/wiki/Kilogram).

Parece-me desnecessário comentar sobre a volatilidade de várias teorias econômicas...

Nestes rápidos comentários que seguem, pretendo passear por alguns aspectos de como a matemática funciona. Uma ótima referência é o livro do Terence Tao [**28**].

A.1. Argumentação formal

A.1.1. Afirmativas. Como funciona a argumentação formal na prática? Objetos fundamentais são as afirmativas (ou afirmações ou expressões lógicas), que sempre são verdadeiras ou falsas, mas nunca verdadeiras *e* falsas simultaneamente. Por exemplo[2]

$$1 + 1 = 2, \tag{A.1.1}$$

$$1 = 2. \tag{A.1.2}$$

Vou me adiantar afirmando que (A.1.1) é verdadeira e (A.1.2) é falsa. Esperando que o leitor já tenha se recuperado da surpresa, cabe aqui comentar que frases sem sentido como

$$= 1 + 3-$$

não são afirmativas. Expressões do tipo 3+1 também não. Uma regra usual é que afirmativas têm verbos.

Afirmativas podem ser combinadas com "ou" e "e" gerando outras. Por exemplo, se a é um número real qualquer, então a afirmativa ($a > 0$ ou $a \le 0$) é verdadeira, mas ($a > 0$ e $a \le 0$) não o é. A regra geral é que se X e Y são afirmativas, então (X *e* Y) só é verdadeira

[1]Última Atualização: 09/01/2008

[2]Suponho, por enquanto, que as propriedades de conjuntos e dos números reais são conhecidas

se X e Y forem *ambas* verdadeiras. Similarmente, (X *ou* Y) só é falsa se X e Y forem *ambas* falsas. Note que se apenas uma das afirmativas for verdadeira, (X *ou* Y) é verdadeira. Note que esta noção pode diferir de um possível uso corriqueiro do *ou*, como na frase *ou eu, ou ele ficamos.* Neste caso quer-se dizer que ou eu fico, ou ele fica, mas não ambos — este é o chamado *ou exclusivo*.[3]

Podemos também negar uma afirmativa. Se X é uma afirmativa verdadeira, então (*não* X) é falsa. Da mesma forma, se Y é uma afirmativa falsa, então (*não* Y) é verdadeira. Negar uma afirmativa pode ser útil pois para concluir que uma afirmativa Z é falsa, as vezes é mais fácil provar que (*não* Z) é verdadeira.

Seguramente, este papo poderia ir bem mais longe com a álgebra de Boole ou booleana (http://en.wikipedia.org/wiki/Boolean_algebra).

A.1.2. Implicações. Os passos de uma argumentação matemática são dados via implicações. Se de um fato conhecido, por exemplo uma afirmativa verdadeira X, eu possso concluir uma afirmativa verdadeira Y, então eu escrevo

$$X \implies Y, \tag{A.1.3}$$

e leio X *implica* Y, ou ainda *se* X *então* Y. Por exemplo

$$a > 0 \implies 2a > 0. \tag{A.1.4}$$

Abstraindo um pouco mais, note que (A.1.3) e (A.1.4) também são afirmativas. Outros exemplos de afirmativas:

$$0 = 0 \implies 0 = 0, \tag{A.1.5}$$

$$0 = 1 \implies 0 = 0, \tag{A.1.6}$$

$$0 = 1 \implies 0 = 1, \tag{A.1.7}$$

$$0 = 0 \implies 0 = 1. \tag{A.1.8}$$

As três primeiras afirmativas acima são verdadeiras. Somente a última é falsa. A primeira da lista é uma tautologia (redundância, do grego *tauto*, o mesmo), e é obviamente correta. Já a segunda é correta pois de hipóteses falsas pode-se concluir verdades (multiplique ambos os lados de (A.1.6) por zero). A terceira é verdade pois se a hipótese é verdadeira, a conclusão, sendo uma mera repetição da hipótese, também o é (este tipo de argumento é usado em demonstrações por contradição). Finalmente, (A.1.8) é falsa pois não se pode deduzir uma afirmativa verdadeira partindo-se de uma falsa.

A argumentação (e a demonstração) matemática baseia-se em supor que algumas hipóteses são verdadeiras e em concluir resultados através de implicações.

Note que a implicação não é "reversível", i.e., se $X \implies Y$, não podemos concluir que $Y \implies X$. Realmente, $x = -1 \implies x^2 = 1$, mas $x^2 = 1 \not\Longrightarrow x = -1$ (esta seta cortada é o símbolo de *não implica*), ou seja, não se pode concluir se $x = -1$ ou não a partir da hipótese $x^2 = 1$.

As vezes, tanto a implicação como seu reverso valem. Se por exemplo $X \implies Y$ e $Y \implies X$ escrevemos simplesmente $X \iff Y$, e lemos X *se e somente se* Y.

[3]Outro termo matemático que pode ter sentido diferente do uso diário é *em geral*. Na matemática, em geral quer dizer *sempre*, enquanto no dia-a-dia quer dizer "quase sempre"

A.1.3. Axiomas. E como começar a construção da matemática em si, i.e., quais são as hipóteses *básicas* que são necessariamente verdadeiras? Iso é importante pois, como vimos, partindo-se de hipóteses falsas pode-se chegar a conclusões falsas, *sem comprometer a lógica.* Aqui entram os *axiomas*, premissas verdadeiras consideradas "óbvias." É uma boa idéia que este conjunto de premissas seja o menor possível, i.e., um axioma do conjunto não pode ser demonstrada a partir dos outros.

A partir dos axiomas contrói-se via implicações toda uma matemática (mudando-se o conjunto de axiomas, muda-se a matemática).

Um exemplo de axioma vem a seguir.

Axioma A.1.1 (do conjunto vazio). Existe um conjunto que não contém nenhum elemento.

Suponha que se possa definir o que é uma pessoa careca, e considere o seguinte axioma.

Axioma A.1.2 (do fio extra). Um careca que ganhar um fio extra de cabelo continua careca.

Pode-se concluir então o seguinte resultado (tente demonstrá-lo).

Se o Axioma do fio extra vale, então todos os seres humanos são carecas.

O alerta que o resultado acima nos fornece é que devemos ter cuidado com os axiomas escolhidos. Resultados "patológicos" podem advir deles. E de fato, resultados "estranhos" permeiam a matemática...

A.1.4. Definições, lemas, teoremas. Uma das formas de se construir novos objetos matemáticos é através de *definições.* Por exemplo podemos definir o conjunto dos números naturais como $\mathbb{N} = \{1, 2, 3, \dots\}$[4]. Outro exemplo: seja

$$\begin{aligned} f : \mathbb{Z} &\to \mathbb{R} \\ x &\mapsto x^2. \end{aligned}$$

A expressão acima define uma função chamada "f" que associa a cada número inteiro o seu quadrado, levando-o nos reais.

E quanto a proposições dadas por lemas e teoremas[5]? Normalmente, lemas e teoremas são escritos à parte, sendo compostos por hipóteses, e conclusões explicitamente mencionadas.

Exemplos de lema e teorema vêm a seguir.

Lema A.1.3. Supondo que o Axioma do conjunto vazio vale, então existe somente um conjunto vazio.

[4]Alguns autores utilizam o símbolo := no lugar de = em definições. Esta é provavelmente uma boa idéia pouco utilizada, e eu não a seguirei.

[5]Uma dúvida comum: qual a diferença entre os três? Bom, normalmente *proposição* tem um caráter mais geral, sendo uma sentença lógica verdadeira (na matemática "usual"). Já um *lema* é proposição preliminar, que contribui na demonstração de um resultado principal, um *teorema.* Muitas vezes entretanto, o lema tem interese próprio. Em geral, o gosto e o estilo do autor determinam o que é proposição, lema ou teorema.

TEOREMA A.1.4 (de Fermat). [6] *Seja* $n \in \mathbb{N}$, *com* $n > 2$. *Então não existem inteiros positivos* x, y, z *tais que* $x^n + y^n = z^n$.

A hipótese do lema A.1.3 é o axioma do conjunto vazio (Axioma A.1.1), e a conclusão é de que só existe um conjunto vazio, isto é todos os conjuntos vazios são iguais. Este é um típico resultado de *unicidade*. Já no Teorema de Fermat A.1.4, impondo-se hipóteses sobre a potência n (ser inteiro e maior que dois), obtem-se um resultado de *não existência*.

Normalmente lemas e teoremas descrevem resultados de interesse e não triviais, i.e., as conclusões não se seguem trivialmente das hipóteses. Algumas vezes entretanto casos importantes particulares são facilmente obtidos de resultados mais gerais. Estes casos particulares são chamados de *corolários*. O Teorema de Fermat por exemplo é um corolário de um outro resultado mais poderoso (chamado Teorema da Modularidade). É claro que "trivialidade" não é um conceito rigoroso e é certamente relativa.

A.1.5. Prova ou demonstração. Uma *prova* ou *demonstração* são os passos lógicos para se concluir uma proposição. Algumas demonstrações são simples, outras nem tanto. Por exemplo, a demonstração por Andrew Wiles do Teorema de Fermat fechou com chave de ouro a matemática do século XX. A prova é uma intricada sequência de resultados publicada num artigo de 109 páginas na mais conceituada revista de matemática, os Anais de Matemática de Princeton [**31**].

Antes da demonstração de Wiles, o agora "Teorema de Fermat" era "somente" uma conjectura, um resultado que acredita-se verdadeiro mas que ninguém demonstrou. Uma ainda conjectura famosa é a de Goldbach, que afirma que *todo inteiro par maior que dois pode ser escrito como a soma de dois números primos*. Para números menores que 10^{18}, o resultado foi checado computacionalmente, mas o caso geral ainda não está provado.

A.2. Demonstração por indução e contradição

Primeiro revemos aqui, através de um exemplo, como é possível demonstrar alguns fatos usando argumentos indutivos.

Considere a afirmativa

$$\sum_{i=1}^{n} i = \frac{n}{2}(n+1) \tag{A.2.1}$$

para todo $n \in \mathbb{N}$.

Para demonstrar que (A.2.1) vale para todos os inteiros positivos, começamos observando que para $n = 1$, a afirmativa é obviamente verdadeira. Assuma então que (A.2.1) seja verdade para $n = N^*$, i.e,

$$\sum_{i=1}^{N^*} i = \frac{N^*}{2}(N^*+1). \tag{A.2.2}$$

[6] Enunciado de Fermat, na margem do livro *Arithmetica* de Diophantus: *Cubum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duos eiusdem nominis fas est dividere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.* (É impossível separar um cubo em dois cubos, ou a quarta potência em quartas potências, ou em geral qualquer potência em duas potências iguais. Eu descobri uma demonstração realmente maravilhosa disto, para a qual esta margem é por demais exígua para caber.)

Para $n = N^* + 1$ temos

$$\sum_{i=1}^{N^*+1} i = N^* + 1 + \sum_{i=1}^{N^*} i.$$

Usamos a hipótese indutiva (A.2.2) obtemos

$$\sum_{i=1}^{N^*+1} i = N^* + 1 + \frac{N^*}{2}(N^* + 1) = \frac{N^* + 1}{2}(N^* + 2),$$

e podemos concluir que (A.2.1) vale para $n = N^* + 1$, e portanto vale para todos os inteiros positivos.

Um dos passos fundamentais, e algumas vezes esquecido, da demonstração por indução é mostrar que o resultado vale para algum valor inicial (na demonstração acima, $n = 1$). De fato, sem isto, podemos erroneamente "provar" que

(A.2.3) $\qquad 2n$ é sempre ímpar para todo $n \in \mathbb{N}$,

com uma argumentação obviamente falsa. De fato supondo que $2N^*$ é ímpar, temos que $2(N^* + 1) = 2N^* + 2$ também é pois $2N^*$ é ímpar por hipótese, e somando 2 a um ímpar obtemos um ímpar. O problema desta demonstração é que não se mostrou (A.2.3) para nenhum número natural.

A demonstração por contradição segue os seguintes princípios lógicos: se queremos mostrar que uma afirmativa implica noutra, podemos simplesmente negar este fato e tentar chegar numa contradição. Considere a afirmativa

(A.2.4) $\qquad \emptyset \subseteq A \quad$ para qualquer conjunto A.

Talvez uma demonstração "direta" não seja tão fácil. Mas suponha que (A.2.4) seja falso. Então existe algum conjunto A tal que $\emptyset \nsubseteq A$. Portanto existe algum elemento no conjunto vazio que não está em A. Mas isto é um absurdo, pois o vazio não contém nenhum elemento. O que se vemos é que negar (A.2.4) (afirmar que (A.2.4) é falso) nos leva a concluir um absurdo, e portanto (A.2.4) só pode ser verdade.

A.3. Exercícios

EXERCÍCIO A.1. Mostre por indução que $n < 2^n$ para todo $n \in \mathbb{N}$.

EXERCÍCIO A.2. Prove que, para todo inteiro $n > 1$ tem-se que

$$1 + \sum_{i=2}^{n} \frac{1}{\sqrt{i}} = 1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{3}} + \cdots + \frac{1}{\sqrt{n}} > \sqrt{n}.$$

EXERCÍCIO A.3. Mostre por indução a desigualdade de Bernoulli: se $x > -1$, então $(1 + x)^n \geq 1 + nx$ para todo $n \in \mathbb{N}$.

EXERCÍCIO A.4. Mostre usando contradição que $\sqrt{2}$ não é racional.

EXERCÍCIO A.5. Mostre usando contradição que se $p_1, \dots, p_n$ são todos os números primos menores ou iguais a p_n, então $p_1 \times \cdots \times p_n + 1$ não é divisível por p_i para nenhum $i \in \{1, \dots, n\}$.

Exercício A.6. Mostre usando contradição que existem infinitos números primos.

Exercício A.7. Usando indução, mostre que existe $J \in \mathbb{N}$ tal que $j^2 - 10j > 0$ para todo inteiro $j > J$.

Exercício A.8. Seja $\lambda < 1$ e $n \in \mathbb{N}$. Mostre que

$$\sum_{i=n}^{k} \lambda^i = \lambda^n \frac{1 - \lambda^{k-n+1}}{1 - \lambda}$$

para todo inteiro $k \geq n$.

Referências Bibliográficas

[1] G. Ávila, *Cálculo,* Livros Técnicos e Científicos S.A., Rio de Janeiro, 1987. Vols. I, II e III.

[2] Robert G. Bartle, *The elements of real analysis*, 2nd ed., John Wiley & Sons, New York-London-Sydney, 1976. MR0393369 (52 #14179)

[3] Robert G. Bartle and Donald R. Sherbert, *Introduction to real analysis*, 2nd ed., John Wiley & Sons Inc., New York, 1992. MR1135107 (92i:26002)

[4] José Boldrini, Sueli Costa, Vera Ribeiro, and Henry Wetzler, *Álgebra Linear*, Editora Harper & Row di Brasil LTDA, São Paulo, 1978.

[5] William E. Boyce and Richard C. DiPrima, *Elementary differential equations and boundary value problems*, John Wiley & Sons, Inc., New York-London-Sydney, 1965. MR0179403

[6] A.C. Chiang, *Matemática para Economistas*, McGraw-Hill, São Paulo.

[7] Djairo G. de Figueiredo, *Funções Reais*, The PanAmerican Union, Washington D.C., 1970 (Portuguese).

[8] R. Fletcher, *Practical methods of optimization*, 2nd ed., A Wiley-Interscience Publication, John Wiley & Sons Ltd., Chichester, 1987. MR955799 (89j:65050)

[9] W.A. Granville, P.F. Smith, and W.R. Longley, *Elementos de cálculo diferencial e integral*, Editora Científica, Rio de Janeiro.

[10] H.L. Guidorizzi, *Um Curso de Cálculo*, Vol. Vols. 1 a 4, Forense-Universitária, Rio de Janeiro. 2ª ed.

[11] G. Hadley, *Álgebra Linear*, Forense-Universitária, Rio de Janeiro.

[12] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985. MR832183 (87e:15001)

[13] Tamara J. Lakins, *The tools of mathematical reasoning*, Pure and Applied Undergraduate Texts, vol. 26, American Mathematical Society, Providence, RI, 2016. MR3525355

[14] Elon Lages Lima, *Análise Real, Volume I*, Coleção Matemática Universitária, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 2009 (Portuguese).

[15] ———, *Álgebra Linear*, Coleção Matemática Universitária, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1996 (Portuguese).

[16] ———, *Curso de análise. Vol. 1*, Projeto Euclides [Euclid Project], vol. 1, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1976 (Portuguese). MR654861 (83h:26002a)

[17] ———, *Curso de análise. Vol. 2*, Projeto Euclides [Euclid Project], vol. 13, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1981 (Portuguese). MR654862 (83h:26002b)

[18] ———, *Espaços métricos*, Projeto Euclides [Euclid Project], vol. 4, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1977 (Portuguese). MR654506 (83d:54001)

[19] Paul R. Halmos, *Naive set theory*, Springer-Verlag, New York, 1974. Reprint of the 1960 edition; Undergraduate Texts in Mathematics. MR0453532 (56 #11794)

[20] S. Kesavan, *Nonlinear functional analysis*, Texts and Readings in Mathematics, vol. 28, Hindustan Book Agency, New Delhi, 2004.

[21] David G. Luenberger, *Introduction to linear and nonlinear programming*, Addison-Wesley, Reading, MA, 1973.Zbl 0297.90044

[22] W. Keith Nicholson, *Elementary Linear Algebra with applications*, 2nd ed., PWS-Kent Publishing Company, Boston, 1990.

[23] Walter Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1976. International Series in Pure and Applied Mathematics. MR0385023 (52 #5893)

[24] *Monkey saddle — Wikipedia, The Free Encyclopedia*, Wikipedia (2009).

[25] Carl Simon and L. Blume, *Mathematics for Economists*, Norton, New York, 1994.

[26] G.L. Reis and V.V. da Silva, *Geometria Analitica*, Livros Técnicos e Científicos Editora S.A., Rio de Janeiro, 1984 (Portuguese).
[27] I. M. Singer and J. A. Thorpe, *Lecture notes on elementary topology and geometry*, Springer-Verlag, New York, 1976. Reprint of the 1967 edition; Undergraduate Texts in Mathematics. MR0413152 (54 #1273)
[28] Terence Tao, *Analysis. I*, Texts and Readings in Mathematics, vol. 37, Hindustan Book Agency, New Delhi, 2006. MR2195040 (2006g:26002a)
[29] ______, *Analysis. II*, Texts and Readings in Mathematics, vol. 38, Hindustan Book Agency, New Delhi, 2006. MR2195041 (2006g:26002b)
[30] S. Viera, *Matemática Financeira*, Atlas, São Paulo.
[31] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), no. 3, 443–551, DOI 10.2307/2118559. MR1333035 (96d:11071)