Allan Amorim Oliveira, Emanuel Mendes Queiroz, Giselle Lopes da Cruz, João Matheus Lira Luiz, Marcio Antônio de Andrande Bortoloti, Maria Clara Brito dos Reis, Samara Viriato Vilar Dias, Wéllington Moutinho Dias

Minicurso PETIMAT

Introdução aos Métodos Computacionais em Otimização Contínua

18 de novembro de 2024

Agradecimentos

Os autores agradecem ao Programa de Educação Tutorial Institucional da Univesidade Estadual do Sudoeste da Bahia (PETI/UESB) pelas bolsas de estudo.

Apresentação

O Programa de Educação Tutorial Institucional (PETI/UESB) tem como objetivo aprimorar os cursos regulares de graduação na Universidade Estadual do Sudoeste da Bahia (UESB), com o objetivo de proporcionar uma formação abrangente e de alta qualidade para os alunos envolvidos, mantendo a integração entre ensino, pesquisa e extensão como um princípio fundamental do ambiente universitário.

O Programa de Educação Tutorial Institucional de Matemática (PETIMAT) se dedica a elevar o padrão da formação dos estudantes de graduação e promover o sucesso acadêmico. Para alcançar esses objetivos, o grupo oferece, a cada semestre, minicursos voltados principalmente para os alunos matriculados no curso de Licenciatura em Matemática da UESB. Esses minicursos visam estimular a capacitação de futuros profissionais e docentes, proporcionando-lhes uma qualificação acadêmica, científica, técnica e tecnológica.

Dentro desse contexto, o PETIMAT desenvolveu o minicurso ”Introdução aos Métodos Computacionais em Otimização Contínua”. O objetivo do minicurso é promover a compreensão de conceitos do Cálculo e de Métodos Computacionais vinculados à Otimização Contínua. Pretende-se implementar algoritmos na Linguagem de programação Julia para resolução de problemas de minimização/maximização de funções.

Este material corresponde às notas de aula e serve como um recurso de consulta para os participantes do minicurso. Nesta apostila, estão disponíveis os conteúdos matemáticos e os exercícios que serão abordados nas aulas. Para ter acesso aos códigos utilizados em cada aula e aos códigos dos métodos computacionais já implementados nas pesquisas produzidas pelo PETIMAT, basta consultar o nosso repositório no GitHub:

`https://github.com/petimatematica/minicurso_intro_otimizacao`

Para conhecer mais sobre o trabalho e as publicações do PETIMAT, acesse o nosso site:

`http://www2.uesb.br/programa/petimatematica/`

e nossa conta oficial no Instagram:

`https://www.instagram.com/petimatuesb/`

onde estão disponíveis informações detalhadas sobre nossas atividades, projetos, eventos e conteúdos relacionados ao programa.

Conteúdo

Introdução

Este material foi elaborado pelo grupo do Programa de Educação Tutorial Institucional de Matemática (PETIMAT) para o minicurso ”Introdução aos Métodos Computacionais em Otimização Contínua”, oferecido durante o período letivo de 2024.1. A finalidade desta apostila é fornecer aos participantes do minicurso uma base teórica sobre os conceitos fundamentais da Otimização Contínua.

Para abordar os conceitos de Otimização Contínua, é necessário discutir sobre conceitos da Análise, que oferece ferramentas essenciais para analisar o comportamento das funções e descrever suas variações. Através do estudo de limites e derivadas, podemos identificar pontos críticos e compreender o comportamento de funções reais. Também são explorados os conceitos de continuidade e convergência de sequências por exemplo, fornecendo uma compreensão mais profunda das propriedades das funções e suas relações. Assim, os conceitos e técnicas da Análise formam uma base teórica essencial para abordar os problemas da Otimização Contínua.

A Otimização Contínua desenvolve teorias que visam encontrar soluções ótimas para problemas nos quais se busca maximizar ou minimizar uma função. Tais problemas estão presentes em diversos campos, como a Engenharia e Economia, e a habilidade de resolvê-los de forma eficiente é crucial para o progresso dessas áreas.

Como parte integrante do minicurso, recomendamos a utilização da linguagem de programação Julia para a implementação dos algoritmos discutidos e para a visualização gráfica dos conceitos abordados.

Capítulo 1
Revisão sobre funções reais

Neste capítulo, abordaremos os fundamentos das funções reais, desde sua definição até a compreensão dos conjuntos de domínio, contradomínio e imagem. Exploraremos os diferentes tipos de funções, como as constantes e afins, incluindo análises de seus gráficos, zeros da função e estudo de sinais. Ademais, serão apresentados os conceitos de sobrejetividade, injetividade, bijetividade, função inversa e composição de funções.

1.1 Funções reais

Definição 1.1.1. *Uma função é uma relação da forma*

$$\begin{aligned} f : A &\to B \\ a &\mapsto b \end{aligned}$$

em que A e B são dois conjuntos não vazios e $a \mapsto b$ é uma regra que nos permite associar a cada elemento $a \in A$ um único elemento $b \in B$. O conjunto A é o domínio e B é o contradomínio. O único $b \in B$ associado ao elemento $a \in A$ é indicado por $f(a)$.

Em outras palavras, uma função é uma relação definida por um conjunto de partida (domínio) e um conjunto de chegada (contradomínio).

Definição 1.1.2. *Chamamos de imagem da função f o conjunto $Im(f)$ dos elementos $b \in B$ para os quais existem $a \in A$ tal que $f(a) = b$. Temos que $Im(f) \subseteq B$.*

Definição 1.1.3. *Dizemos que uma função é real se ela é do tipo $f : A \to \mathbb{R}$, onde $A \subset \mathbb{R}$.*

Definição 1.1.4. *Dada uma função $f : A \to B$ o gráfico de f é o conjunto $\{(x, f(x)) : x \in A\}$.*

1.2 Função constante

Definição 1.2.1. *Uma função real $f : A \subset \mathbb{R} \to \mathbb{R}$ recebe o nome de função constante quando qualquer elemento $x \in A$ está associado sempre o mesmo elemento $c \in \mathbb{R}$.*

Em outras palavras, uma função constante associa todos os elementos do domínio a um mesmo número no contradomínio, ou seja, $f(x) = c, \forall x \in A$. O gráfico de uma função constante é uma reta paralela ao eixo das abscissas e que passa pelo ponto $(0, c)$. Abaixo, apresentamos um exemplo de função constante.

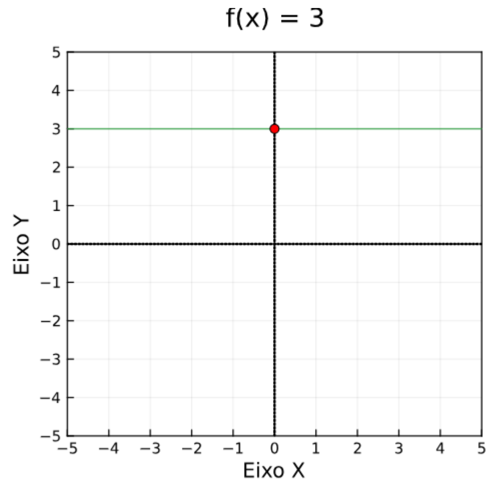


Figura 1.1: Função constante para $c = 3$.

1.3 Função afim

Definição 1.3.1. *Uma função $f : X \subset \mathbb{R} \to \mathbb{R}$ recebe o nome de função afim quando cada elemento $x \in X$ está associado ao elemento $(ax + b) \in \mathbb{R}$, em que $a \neq 0$, ou seja, $f(x) = ax + b$.*

O gráfico de uma função afim é uma reta que possui inclinação a em relação ao eixo das abscissas e que passa pelos pontos $(0, b)$ e $(-b/a, 0)$. O coeficiente a da função f é denominado coeficiente angular, enquanto b é denominado coeficiente linear. Abaixo, apresentamos um exemplo de função afim.

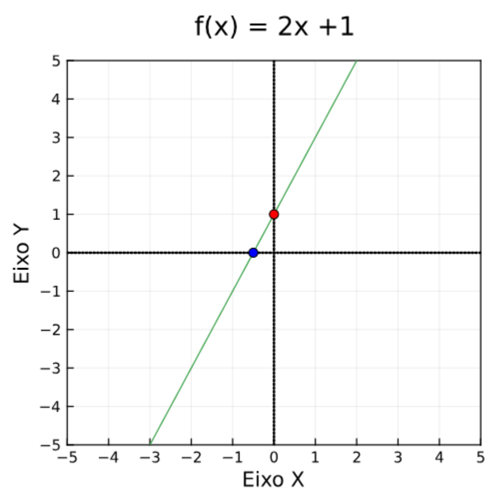


Figura 1.2: Função afim em que $a = 2$ e $b = 1$.

Exemplo 1.3.1. *Na função*

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R} \\ x &\mapsto 2x + 1 \end{aligned}$$

o coeficiente angular é igual a 2 *e o coeficiente linear é igual a* 1*. Note que, quando* $x = 0$*, temos* $f(0) = 2 \cdot 0 + 1 = 1$*. Portanto,* 1 *é a ordenada do ponto de interseção entre a reta e o eixo* y*.*

Definição 1.3.2. *Chamamos de zero da função* $f : X \subset \mathbb{R} \to \mathbb{R}$ *o número* $x \in X$ *tal que* $f(x) = 0$*. O conjunto dos zeros da função* f *é definido por* $Z_f = \{x \in X; f(x) = 0\}$*.*

Observação 1.3.3. *Para determinar os zeros de uma função afim* $f : \mathbb{R} \to \mathbb{R}, x \mapsto ax + b$*, basta resolver a equação* $ax + b = 0$*.*

Exemplo 1.3.2. *O zero da função afim* $f : \mathbb{R} \to \mathbb{R}, x \mapsto 2x - 1$ *é dado por* $x = 1/2$*, pois* $f(1/2) = 2 \cdot (1/2) - 1 = 1 - 1 = 0$*.*

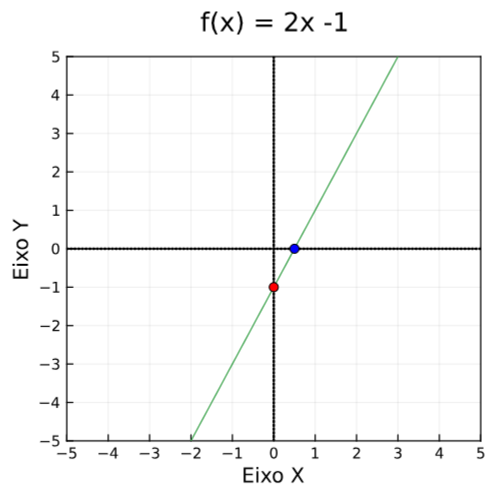


Figura 1.3: Função afim em que $a = 2$ e $b = -1$.

Definição 1.3.4 (Crescimento e descrescimento local). *Uma função* $f : X \subset \mathbb{R} \to \mathbb{R}$ *é dita crescente no conjunto* $X_1 \subset X$ *quando, dados quaisquer* $x_1, x_2 \in X_1$*, com* $x_1 < x_2$*, tivermos* $f(x_1) < f(x_2)$*. Por outro lado,* f *é dita decrescente no conjunto* $X_1 \subset X$ *quando, dados* $x_1, x_2 \in X_1$*, com* $x_1 < x_2$*, tivermos* $f(x_1) > f(x_2)$*.*

Isso significa que a função f é crescente no conjunto X_1 se, ao aumentarmos o valor de x, o valor de $f(x)$ também aumenta. Da mesma forma, a função é decrescente no conjunto X_1 se, ao aumentarmos o valor de x, o valor de $f(x)$ diminui.

Teorema 1.3.1. *Seja* $f : X \subset \mathbb{R} \to \mathbb{R}$ *a função afim definida por* $f(x) = ax + b$*, com* $x \in X$*. Valem as seguintes afirmações:*

(i) É crescente se, e somente se, o coeficiente angular for positivo.
(ii) É decrescente se, e somente se, o coeficiente angular for negativo.

Demonstração: Tomemos $x_1, x_2 \in X$, com $x_2 > x_1$. Então:

(i) Se $x_2 > x_1$ então $x_1 - x_2 < 0$, por f ser crescente $f(x_2) > f(x_1)$, que implica $f(x_1) - f(x_2) < 0$, logo $\frac{f(x_1)-f(x_2)}{x_1-x_2} > 0$. Portanto, se f é crescente, então $\frac{f(x_1)-f(x_2)}{x_1-x_2} > 0 \iff \frac{(ax_1+b)-(ax_2+b)}{x_1-x_2} > 0 \iff \frac{a(x_1-x_2)}{x_1-x_2} > 0 \iff a > 0$.

(ii) Se $x_2 > x_1$ então $x_1 - x_2 < 0$, por f ser decrescente $f(x_2) < f(x_1)$, que implica $f(x_2) - f(x_1) < 0$, logo $\frac{f(x_1)-f(x_2)}{x_1-x_2} < 0$. Portanto, se f é decrescente, então $\frac{f(x_1)-f(x_2)}{x_1-x_2} < 0 \iff \frac{(ax_1+b)-(ax_2+b)}{x_1-x_2} < 0 \iff \frac{a(x_1-x_2)}{x_1-x_2} < 0 \iff a < 0$.

□

Considerando que $x = -b/a$ é o valor de x tal que $f(x) = 0$, examinaremos os valores tais que $f(x) > 0$ e $f(x) < 0$.

(i) Para $x > -b/a$ temos que, se $a > 0$ então $f(x) = ax + b > 0$. Se $a < 0$, então $f(x) = ax + b < 0$.
(ii) Para $x < -b/a$ temos que, se $a > 0$ então $f(x) = ax + b < 0$. Se $a < 0$, então $f(x) = ax + b > 0$.

Exemplo 1.3.3. *A função $f : \mathbb{R} \to \mathbb{R}$; $x \mapsto 2x + 1$ é crescente, pois seu coeficiente angular $a = 2 > 0$*

Exemplo 1.3.4. *A função $f : \mathbb{R} \to \mathbb{R}$; $x \mapsto -2x + 1$ é decrescente em $\mathbb{R}$, pois seu coeficiente angular $a = -2 < 0$*

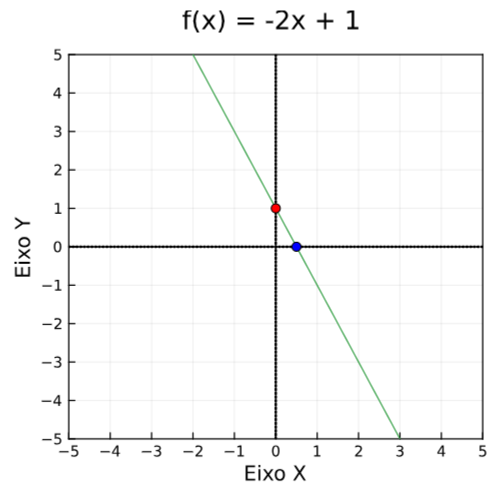


Figura 1.4: Função afim em que $a = -2$ e $b = 1$.

1.4 Composição de funções

Definição 1.4.1. *Sejam $f : A \to B$, $g : B \to C$ funções tais que o contradomínio de f é igual ao domínio de g. Então a composição de função de f com g é a função $g \circ f : A \to C$, definida por $(g \circ f)(x) = g(f(x))$*

Para calcular $(g \circ f)(x)$ por exemplo, procedemos da seguinte forma:
1°) Aplica-se a função f a x, obtendo-se $f(x)$.
2°) Aplica-se a função g a $f(x)$, obtendo-se $g(f(x))$.

Exemplo 1.4.1. *Sejam $A = \{-1, 0, 1, 2\}$, $B = \{0, 1, 2, 3, 4\}$ e $C = \{1, 3, 5, 7, 9\}$ e sejam as funções $f : A \to B$ e $g : B \to C$ definidas por $f(x) = x^2$ e $g(x) = 2x + 1$.*
Observe que $f(2) = 4$, $g(4) = 9$ e $h(2) = 9$, isto é, $h(2) = (g \circ f)(2) = g(f(2)) = g(4) = 9$. A lei de formação de $(g \circ f)(x)$ é dada por:

$$(g \circ f) = g(f(x)) = 2f(x) + 1 = 2x^2 + 1.$$

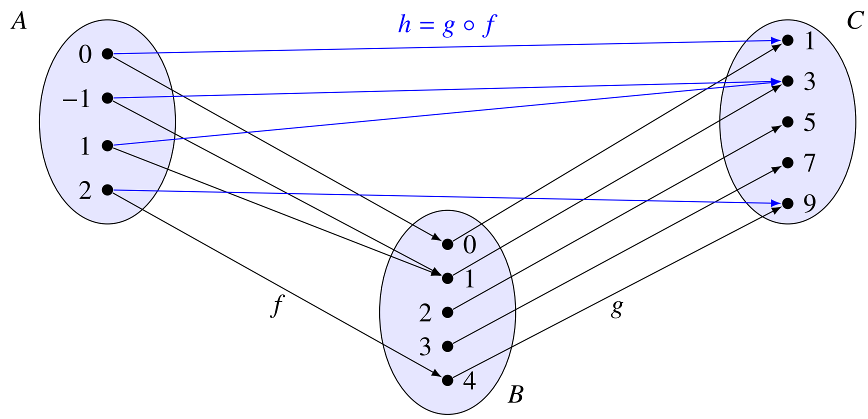


Figura 1.5: Diagrama ilustrativo da função composta $g \circ f : A \to C$.

1.5 Funções inversas

Definição 1.5.1. *Uma função $f : A \to B$ se diz injetora quando para quaisquer $x_1, x_2 \in A$ com $x_1 \neq x_2$, tem-se $f(x_1) \neq f(x_2)$*

Exemplo 1.5.1. *A função f de $A = \{0, 1, 2, 3\}$ em $B = \{1, 3, 5, 7, 9\}$ definida pela lei $f(x) = 2x + 1$ é injetora.*

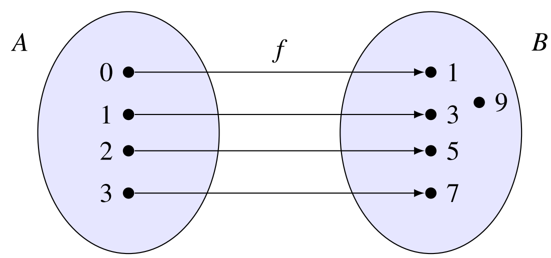


Figura 1.6: Diagrama ilustrativo de uma função $f : A \to B$ injetora.

Exemplo 1.5.2. *A função $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x$ é injetora.*

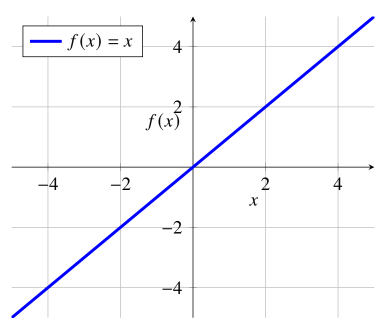


Figura 1.7: Exemplo de função real injetora.

Definição 1.5.2. *Uma função $f : A \to B$ é sobrejetora se, para para qualquer $b \in B$, existe pelo um elemento $a \in A$ tal que $f(a) = b$. Como consequência temos que $B = Im(f)$*

Exemplo 1.5.3. *A função f de $A = \{-1, 0, 1, 2\}$ em $B = \{0, 1, 4\}$ definida pela lei $f(x) = x^2$ é sobrejetora.*

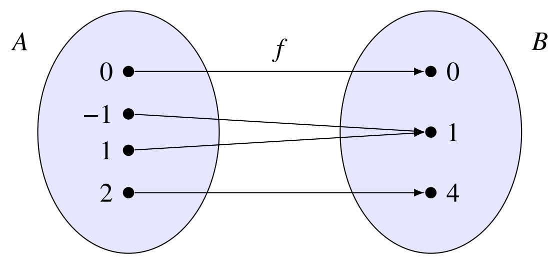


Figura 1.8: Diagrama ilustrativo de uma função $f : A \to B$ sobrejetora.

Exemplo 1.5.4. *A função $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x - 1$ é sobrejetora.*

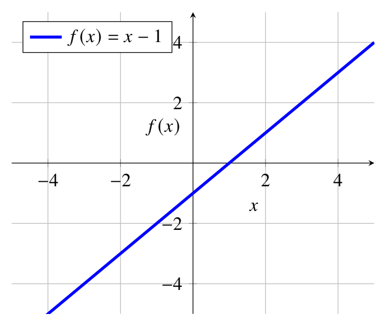


Figura 1.9: Exemplo de função real sobrejetora.

Definição 1.5.3. *Uma função $f : A \to B$ é bijetora se, e somente se, é injetora e sobrejetora.*

Exemplo 1.5.5. *A função f de $A = \{0, 1, 2, 3\}$ em $B = \{1, 2, 3, 4\}$ definida por $f(x) = x + 1$ é bijetora.*

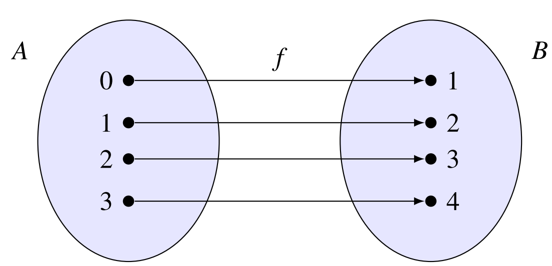


Figura 1.10: Diagrama ilustrativo de uma função $f : A \to B$ bijetora.

Exemplo 1.5.6. *A função $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = 2x$ é bijetora.*

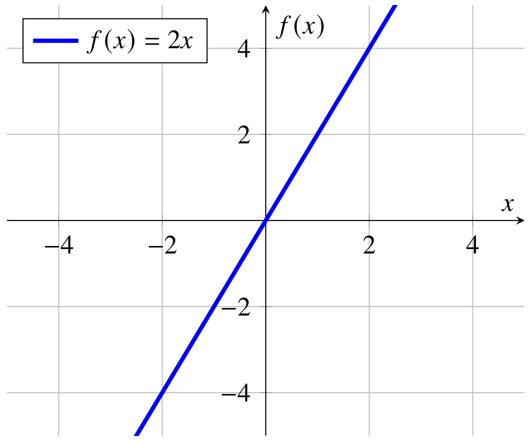


Figura 1.11: Exemplo de função real bijetora.

Exemplo 1.5.7. ***(Exemplo preliminar de função inversa)*** *Dados* $A = \{1, 2, 3, 4\}$, $B = \{1, 3, 5, 7\}$ *e* $f : A \to B$ *definida por* $f(x) = 2x - 1$, *note que essa é uma função bijetora.*

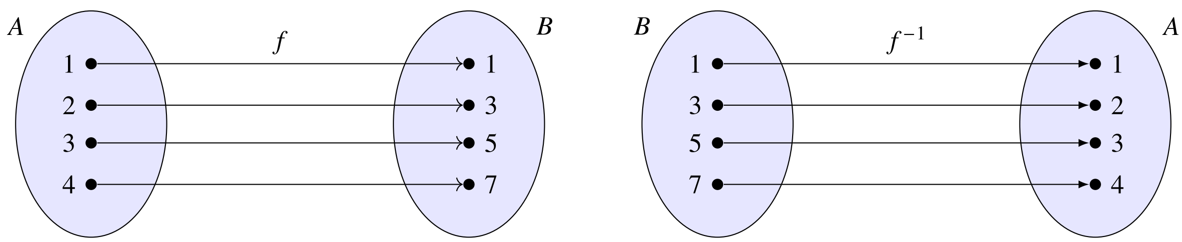


Figura 1.12: Diagramas ilustrativos de uma função $f : A \to B$ e sua inversa $f^{-1} : B \to A$.

Definição 1.5.4. *Dado um conjunto* X, *chamamos de função identidade em* X *a função* $Id_X : X \to X$ *tal que* $x \mapsto x$.

Definição 1.5.5. *Dada a bijeção* $f : A \to B$, *a função* $g : B \to A$ *é a inversa de* f *se, e somente se,* $(f \circ g)(x) = Id_B(x)$ *e* $(g \circ f)(x) = Id_A(x)$.

Exemplo 1.5.8. *Dada a função bijetora* $f : \mathbb{R} \to \mathbb{R}$ *definida por* $f(x) = 2x - 1$, *sua inversa* $f^{-1} : \mathbb{R} \to \mathbb{R}$ *é dada por* $f^{-1}(x) = \frac{x-1}{2}$, *pois* $(f \circ g)(x) = x$ *e* $(g \circ f)(x) = x$.

1.6 Exercícios

Exercício 1.6.1. *Esboce o gráfico das funções* $f : \mathbb{R} \to \mathbb{R}$ *dadas por:*
a) $f(x) = 2x + 1$.
b) $f(x) = \frac{2x-3}{2}$.
c) $f(x) = -3x - 4$.

Exercício 1.6.2. *Obtenha as funções do tipo afim que passam pelos pontos a seguir. Determine os zeros dessas funções e analise os intervalos em que a função é positiva, negativa ou nula. Por fim, esboce o gráfico.*
a) $(2, 3)$ *e* $(3, 5)$.
b) $(1, -1)$ *e* $(-1, 2)$.
c) $(3, -2)$ *e* $(2, -3)$.

Exercício 1.6.3. *Obtenha as funções do tipo afim tais que:*
a) Passa pelo ponto $(-2, 4)$ *e tem coeficiente angular igual a* -3.
b) Tenha o coeficiente linear igual a -3 *e passa pelo ponto* $(-3, -2)$.
c) Passa pelo ponto $(1, 3)$ *e tem coeficiente angular igual a* 2.
Agora, ache os zeros dessas funções e faça o estudo de sinal delas. Por fim, faça a visualização gráfica dessas funções.

Exercício 1.6.4. *Analise cada função abaixo e veja se é injetora, sobrejetora ou bijetora. Se for bijetora, determine sua inversa.*

(a) $f : \mathbb{R} \to \mathbb{R}$ *tal que* $f(x) = 2x + 1$;
(b) $g : \mathbb{R} \to \mathbb{R}$ *tal que* $g(x) = x^2$;
(c) $f : \mathbb{R} \to \mathbb{R}$ *tal que* $f(x) = 2x + 7$;
(d) $g : \mathbb{R} \to \mathbb{R}$ *tal que* $g(x) = \dfrac{1}{x^2 + 1}$;
(e) $n : \mathbb{Z} \to \mathbb{N}$ *tal que* $n(x) = x^2 - 2x + 1$;
(f) $p : \mathbb{N} \to \mathbb{N}$ *tal que* $p(x) = |x|$;
(g) $q : (-\infty, -10) \to (-\infty, 0)$ *tal que* $q(x) = -|x + 4|$;
(h) $r : (-\infty, 0) \to (-\infty, 0)$ *tal que* $r(x) = -|x + 4|$.

Capítulo 2
Problemas da Otimização que envolvem funções quadráticas

Neste capítulo, abordaremos conceitos elementares de otimização através das funções quadráticas. Para isso, veremos que tais funções, que serão definidas a seguir, possuem um mínimo ou um máximo global. Além disso, este ponto de máximo ou mínimo é único. Por fim, veremos a resolução de alguns problemas que podem ser modelados por uma função quadrática. Na parte de implementação em Julia, serão apresentados exemplos que utilizam o pacote *Plots*, o qual é capaz de esboçar gráficos de funções. A última seção destina-se a exercícios de fixação.

2.1 Função Quadrática

Definição 2.1.1. *Uma função $f : X \subset \mathbb{R} \to \mathbb{R}$ recebe o nome de **função quadrática** quando associa a cada $x \in \mathbb{R}$ o elemento $f(x) \in \mathbb{R}$, tal que $f(x) = ax^2 + bx + c$ em que a, b e c são números reais fixados e $a \neq 0$.*

Observação 2.1.2. *Se fosse $a = 0$ teríamos a função afim $f(x) = bx + c$ com $b, c \in \mathbb{R}$. Como tal função já foi estudada, consideraremos a restrição $a \neq 0$, com o intuito de obter novos resultados.*

Observe que a, b, c podem assumir qualquer valor real, excetuando-se a restrição no coeficiente a mencionada acima. Logo, existe uma infinidade de funções quadráticas. Apresentaremos abaixo alguns exemplos:

Exemplos 2.1.3. *Sejam $f, g, h, i : X \subset \mathbb{R} \to \mathbb{R}$*

1. $f(x) = x^2 - 3x + 2$ *em que* $a = 1$, $b = -3$ *e* $c = 2$
2. $g(x) = x^2 - 4$ *em que* $a = 1$, $b = 0$ *e* $c = -4$
3. $h(x) = -2x^2 + 5x$ *em que* $a = -2$, $b = 5$ *e* $c = 0$
4. $i(x) = -3x^2$ *em que* $a = -3$, $b = 0$ *e* $c = 0$

Observação 2.1.4. *Note que enfatizamos quais valores os coeficientes a, b e c assumiram em cada regra $x \mapsto ax^2+bx+c$. A motivação para tal é que as fórmulas que deduziremos a seguir serão escritas em termos dos coeficientes. Logo, tal habilidade é desejável uma vez que essas fórmulas são relativamente simples de serem memorizadas e nos ajudarão a resolver rapidamente os problemas propostos.*

2.2 Gráfico de uma Função Quadrática

Vimos na Definição 1.1.4 que o gráfico de uma função $f : A \to B$ é o conjunto $\{(x, f(x)) : x \in A\}$. Em particular, considerando a Definição 2.1.1, o gráfico de uma função quadrática é o conjunto $\{(x, f(x)) : x \in X\}$, como $f(x) = ax^2 + bx + c$ segue que $\{(x, ax^2 + bx + c) : x \in X\}$. Utilizando o plano cartesiano, podemos representar graficamente este conjunto através do esboço de uma curva que passa pelas coordenadas da forma $(x, ax^2 + bx + c)$. Vejamos o esboço do gráfico das funções f, g, h e i apresentadas no exemplo anterior.

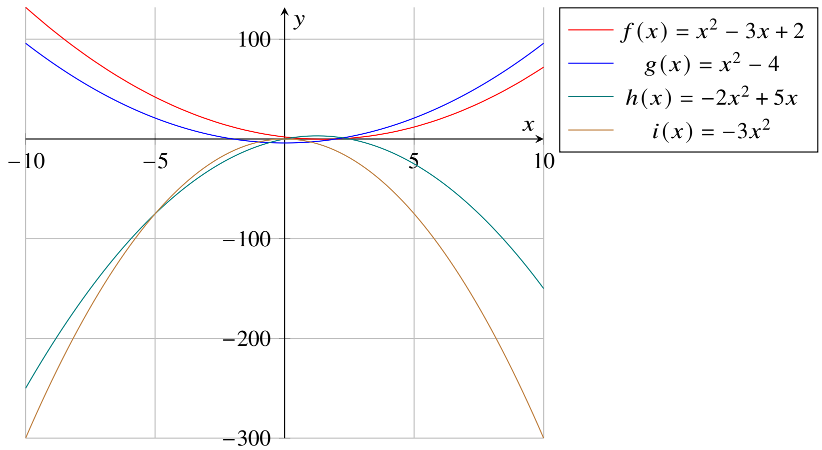


Figura 2.1: Esboço dos gráficos das funções quadráticas definidas no Exemplo 2.1.3.

A Figura 2.1 pode nos sugerir similaridades nos esboços dos gráficos das funções h e i. Além disso, a mesma similalidade é notada no esboço dos gráficos das funções f e g. Isso não é por acaso, visto que é sabido da Geometria Analítica que a curva que representa o gráfico da função quadrática é uma parábola, para mais detalhes, veja [2]. Apesar de ser extremamente interessante, não justificaremos essa afirmação pois, é possível otimizar uma função quadrática algebricamente. Entretanto, essa informação pode nos fornecer interpretações geométricas que podem ser úteis a depender do problema. Por isso, daremos a definição de uma parábola.

Definição 2.2.1. *Dado um ponto F (foco) e uma reta d (diretriz), pertencentes a um plano α, com $F \notin d$, seja p a distância entre F e d. Uma* ***parábola*** *é um conjunto de pontos de α que são equidistantes de F e de d simultaneamente.*

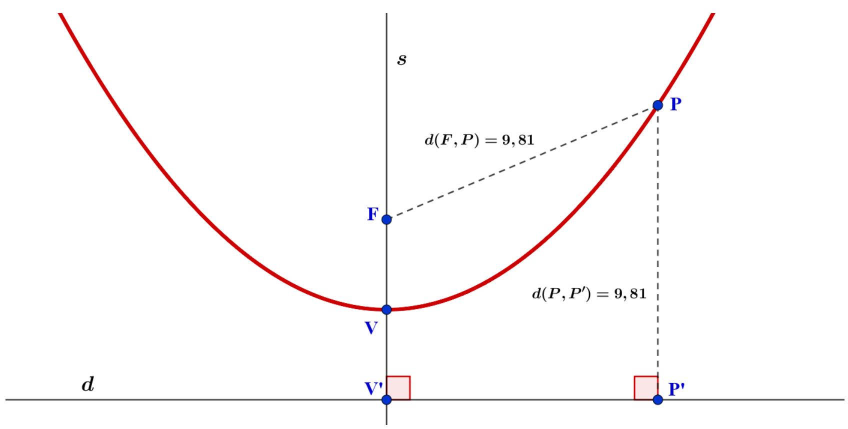


Figura 2.2: Parábola definida pela reta diretriz d e foco F.

2.3 Propriedades da Função Quadrática

Começaremos esta seção reescrevendo a equação $y = ax^2 + bx + c$ com $a \neq 0$ de forma conveniente, com o intuito de deduzir algumas propriedades da função quadrática. Para isso, note que:

$$
\begin{aligned}
y &= ax^2 + bx + c && (2.1)\\
&= a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right)\\
&= a\left(x^2 + \frac{b}{a}x + 0 + \frac{c}{a}\right)\\
&= a\left(x^2 + \frac{b}{a}x + \left(\frac{b^2}{4a^2} - \frac{b^2}{4a^2}\right) + \frac{c}{a}\right)\\
&= a\left(\left(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2}\right) - \left(\frac{b^2}{4a^2} - \frac{c}{a}\right)\right)\\
&= a\left(\left(x + \frac{b}{2a}\right)^2 - \left(\frac{b^2 - 4ac}{4a^2}\right)\right) && (2.2)
\end{aligned}
$$

Tomando, $\Delta = b^2 - 4ac$ obtemos a igualdade:

$$y = a\left(\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\Delta}{4a^2}\right)\right). \tag{2.3}$$

A seguir, abordaremos as propriedades relacionadas às raízes da função quadrática. Assim como a função afim, a função quadrática possui uma fórmula fechada para encontrar suas raízes. Fazendo $f(x) = 0$ temos:

$$ax^2 + bx + c = 0,$$

usando (2.3) obteremos:

$$a\left(\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\Delta}{4a^2}\right)\right) = 0, \tag{2.4}$$

porque $a \neq 0$. Multiplicando ambos os membros da igualdade acima por a^{-1} e somando os por $\frac{\Delta}{4a^2}$ sucede em:

$$\left(x + \frac{b}{2a}\right)^2 = \frac{\Delta}{4a^2}.$$

Considerando que $\Delta \geq 0$, podemos afirmar a existência de um número real que seja a raíz quadrada de $\frac{\Delta}{4a^2}$, daí segue que:

$$x + \frac{b}{2a} = \pm\frac{\sqrt{\Delta}}{2a}. \tag{2.5}$$

Por fim, ao somar $\frac{-b}{2a}$ em ambos os membros da igualdade, obtemos uma das fórmulas mais difundidas e lembradas do ensino básico:

$$x = \frac{-b \pm \sqrt{\Delta}}{2a}. \tag{2.6}$$

Caso $\Delta < 0$, a equação (2.6) não possui soluções reais pois, não existe um número real que seja raiz quadrada de um número negativo. Por isso, podemos explicitar as soluções complexas da seguinte maneira, seja $\Phi = -\Delta$. Note que $\Delta = -\Phi$ e $\Phi > 0$. Usando o fato de que $\sqrt{-1} = i$ verificamos a seguinte igualdade,

$$x = \frac{-b \pm \sqrt{\Delta}}{2a} = \frac{-b \pm \sqrt{-\Phi}}{2a} = \frac{-b \pm \sqrt{\Phi}i}{2a}.$$

Com isso, podemos concluir que uma função quadrática possui no máximo, duas raízes reais distintas. Dado o gráfico de uma função, podemos identificar facilmente as raízes raízes reais verificando os elementos do domínio em que a sua representação gráfica intersecta o eixo das abscissas ou a existência de raízes complexas, caso não haja interseção.

De (2.6), podemos escrever as raízes da função quadrática:

$$x_1 = \frac{-b-\sqrt{\Delta}}{2a} \text{ e } x_2 = \frac{-b+\sqrt{\Delta}}{2a}.$$

Logo, é possível deduzir uma forma explícita para a soma das raízes:

$$x_1 + x_2 = \frac{-b-\sqrt{\Delta}}{2a} + \frac{-b+\sqrt{\Delta}}{2a} = -\frac{b}{a}.$$

E consequentemente do ponto médio das ráizes:

$$\frac{x_1+x_2}{2} = -\frac{b}{2a}. \tag{2.7}$$

Observe que em (2.6) temos

$$x = \frac{-b \pm \sqrt{\Delta}}{2a} = -\frac{b}{2a} \pm \frac{\sqrt{\Delta}}{2a} = \frac{x_1+x_2}{2} \pm \frac{\sqrt{\Delta}}{2a}$$

isto é, caso a função quadrática possua raízes reais, podemos interpretar (2.6) como a soma de duas parcelas, em que uma delas corresponde ao ponto médio das raízes. Dito isso, podemos nos perguntar se a parcela $\pm\frac{\sqrt{\Delta}}{2a}$ possui algum significado geométrico. A resposta para essa questão é afirmativa e deixaremos para o leitor a tarefa de encontrá-lo no terceiro exercício proposto para este capítulo. Independentemente se as raízes são reais ou complexas, o valor $-\frac{b}{2a}$ possui um outro significado nas funções quadráticas. Abordaremos esse tema na seção a seguir.

2.4 Máximo e mínimo

Nesta seção verificaremos que uma função quadrática

$$\begin{aligned} f :&\mathbb{R} \to \mathbb{R} \\ &x \mapsto ax^2 + bx + c \end{aligned}$$

possui um ponto de máximo ou de mínimo a depender do sinal de a. Para isso, consideraremos definições provisórias do que futuramente serão chamados de **máximo global** e **mínimo global**.

Definição 2.4.1. *Dado a função $f : A \to B$. Dizemos que o número $M \in Im(f)$ é o **valor máximo** de f se, e somente se, $M \geq f(x)$ para todo $f(x) \in Im(f)$. Damos o nome **ponto de máximo** para o $x_M \in A$ tal que $f(x_M) = M$.*

Definição 2.4.2. *Dado a função $f : A \to B$. Dizemos que o número $m \in Im(f)$ é o **valor mínimo** de f se, e somente se, $m \leq f(x)$ para todo $f(x) \in Im(f)$. Damos o nome **ponto de mínimo** para o $x_m \in A$ tal que $f(x_m) = m$.*

Encontraremos a seguir, o ponto de máximo ou de mínimo da função quadrática $f(x) = ax^2 + bx + c$.

Proposição 2.4.1. *Se $a < 0$, a função quadrática*

$$\begin{aligned} f :&\mathbb{R} \to \mathbb{R} \\ &x \mapsto ax^2 + bx + c \end{aligned}$$

admite o valor máximo $M = -\frac{\Delta}{4a}$ para $x_M = -\frac{b}{2a}$.

Demonstração: Suponha que exista um $x \in \mathbb{R}$ tal que $f(x) > -\dfrac{\Delta}{4a}$. Usando a igualdade (2.3) temos que:

$$f(x) = a\left(\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\Delta}{4a^2}\right)\right) > -\frac{\Delta}{4a}.$$

Como $a < 0$ segue que $a^{-1} < 0$. Ao multiplicarmos ambos os membros da desigualdade acima por a^{-1}, sucede-se:

$$\left(x + \frac{b}{2a}\right)^2 - \left(\frac{\Delta}{4a^2}\right) < -\left(\frac{\Delta}{4a^2}\right). \tag{2.8}$$

Acarretando em:

$$\left(x + \frac{b}{2a}\right)^2 < 0. \tag{2.9}$$

O que não pode acontecer, visto que $\left(x + \dfrac{b}{2a}\right)^2 \geq 0$ para todo $x \in \mathbb{R}$. Portanto, $f(x) \leq -\dfrac{\Delta}{4a}$ para qualquer x real. Além disso, repare que

$$f\left(-\frac{b}{2a}\right) = a\left(\left(\left(-\frac{b}{2a}\right) + \frac{b}{2a}\right)^2 - \left(\frac{\Delta}{4a^2}\right)\right) = -\frac{\Delta}{4a}. \tag{2.10}$$

Ou seja, $x_M = \left(-\dfrac{b}{2a}\right)$ e $M = -\dfrac{\Delta}{4a}$.

A proposição a seguir demonstra-se de maneira análoga.

Proposição 2.4.2. *Se $a > 0$, a função quadrática*

$$\begin{aligned} f :&\mathbb{R} \to \mathbb{R} \\ &x \mapsto ax^2 + bx + c \end{aligned}$$

admite o valor mínimo $M = -\dfrac{\Delta}{4a}$ para $x_M = -\dfrac{b}{2a}$.

As duas últimas proposições mostram a existência de pelo menos um ponto de máximo ou de mínimo para funções quadráticas. Para trabalharmos com problemas de otimização com funções quadráticas precisamos demonstrar a unicidade desses pontos.

Proposição 2.4.3. *A função quadrática*

$$\begin{aligned} f :&\mathbb{R} \to \mathbb{R} \\ &x \mapsto ax^2 + bx + c \end{aligned}$$

admite um único valor mínimo se $a < 0$ ou um único valor máximo se $a > 0$.

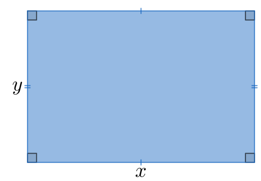
A demonstração da proposição acima será deixada como exercício para o leitor. Devido a unicidade, o ponto (M, x_M) é denominado de **vértice** da função quadrática.

2.5 Problema de otimização envolvendo uma função quadrática

Utilizando as ferramentas matemáticas desenvolvidas nas seções anteriores, podemos rapidamente resolver problemas de otimização modelados por funções quadráticas. A questão principal é identificar em quais problemas tal abordagem é viável. Como por exemplo:

Questão 2.5.1. *Dentre todos os retângulos de perímetro 20 cm, determine o de área máxima.*

Modelagem: Sabemos que o retângulo é um quadrilátero plano convexo que possui os quatro ângulos congruentes e lados opostos paralelos. Assim, para a fixação de ideias vamos desenhar um retângulo arbitrário de comprimento y e largura x:



Usando a figura acima, podemos afirmar que o perímetro (representado pela letra P) será dado (em centímetros) por $P = 2(x + y)$ e a área (em centímetros quadrados) por $A = xy$. Além disso, como o perímetro vale 20 cm, podemos escrever $P = 20$ cm. Consequentemente, por transitividade $2(x + y) = 20$, logo $x + y = 10$.

Observe que vamos determinar a área máxima mediante a restrição $x + y = 10$. Uma estratégia utilizada neste caso é reescrever A em termos de uma única variável, isto é, se notarmos que $y = 10 - x$ temos que $A = xy = x(10 - x) = -x^2 + 10x$, ou seja, a área pode ser descrita como uma função quadrática, repare ainda que a função $-x^2 + 10x$ possui um máximo, pois $a = -1$. Como $x_M = -\frac{b}{2a}$ segue que $x_M = -\frac{10}{2(-1)} = 5$. Sucede-se então que $y = 5$ porque $x + y = 10$. Portanto, o retângulo de perímetro 20 cm de área máxima é o quadrado cujo lado mede 5 cm.

2.6 Exercícios

Exercício 2.6.1. *Explique o motivo do uso de* $\pm$ *em* (2.5).

Exercício 2.6.2. *Mostre que o produto das raízes da uma função* $f : \mathbb{R} \to \mathbb{R}$ *tal que* $f(x) = ax^2 + bx + c$ *é dado pela fórmula* $\frac{c}{a}$.

Exercício 2.6.3. *Qual é o significado geométrico da parcela* $\pm\frac{\sqrt{\Delta}}{2a}$ *em* (2.6) *quando a função quadrática possui raízes reais?*

Exercício 2.6.4. *Explique porque as definições* (2.4.1) *e* (2.4.2) *são provisórias.*

Exercício 2.6.5. *Determine o retângulo de área máxima localizado no primeiro quadrante, com dois lados nos eixos cartesianos e um vértice na reta* $y = -4x + 5$.

Exercício 2.6.6. *Determine o retângulo de maior área contido num triângulo equilátero de lado* 4 *cm, estando a base do retângulo num lado do triângulo.*

Exercício 2.6.7. *Num triângulo isósceles de base* 6 *cm e altura* 4 *cm está inscrito um retângulo. Determine o retângulo de área máxima, sabendo que a base do retângulo está sobre a base do triângulo.*

Exercício 2.6.8. *Uma parede de tijolos será usada como um dos lados de um curral retangular. Para os outros lados iremos usar* 400 *metros de tela de arame, de modo a produzir área máxima. Qual é o quociente de um lado pelo outro?*

Exercício 2.6.9. *Determine o retângulo com área de* 1000 m^2 *cujo perímetro seja o menor possível.*

Exercício 2.6.10. *Considere o seguinte problema: um fazendeiro com* 300 *m de cerca quer cercar uma área retangular e então dividi-la em quatro partes com cercas paralelas a um lado do retângulo. Qual é a maior área possível das quatro partes.*

Capítulo 3
Derivadas e Otimização

Neste capítulo, abordaremos alguns conceitos fundamentais do Cálculo Diferencial tendo em vista uma de suas aplicações: encontrar valores máximos e mínimos de uma função real. Veremos que a derivada, principal ideia do Cálculo Diferencial, é um limite e por isso, discutiremos brevemente este conceito. Por fim, exibiremos a definição de continuidade, condição indispensável para otimizar funções com os recursos da Otimização Contínua

3.1 Retas secantes e tangentes

Na geometria plana, sabemos que a reta é um conceito primitivo, isto é, o adotamos sem definição. Em contrapartida, podemos definir circunferência.

Definição 3.1.1. ***Circunferência*** *é o conjunto dos pontos de um plano equidistantes a um ponto dado. O ponto dado é o* ***centro****, e a distância dada (não nula) é o* ***raio*** *da circunferência.*

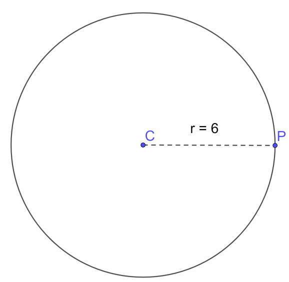


Figura 3.1: Circunferência de centro C e raio $r = 6$ unidades de medida.

Fixado um plano podemos classificar as posições relativas entre uma reta e uma circunferência.

Definição 3.1.2. *Uma reta* ***secante*** *a uma circunferência é uma reta que intercepta a circunferência em dois pontos distintos.*

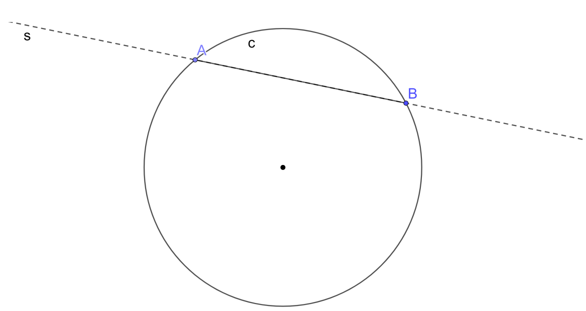


Figura 3.2: Reta s secante a circunferência c.

Definição 3.1.3. *Uma reta* ***tangente*** *a uma circunferência é uma reta que intercepta a circunferência num único ponto.*

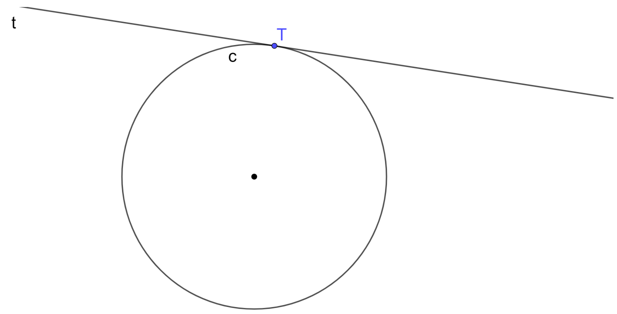


Figura 3.3: Reta t secante a circunferência c.

Note que a reta tangente a uma circunferência tem apenas um ponto de interseção com ela, sendo os demais pontos da reta externos à circunferência. O ponto comum é o ponto de tangência. Além disso, dizemos que a reta e a circunferência são tangentes.

Podemos estender os conceitos de reta secante e reta tangente a uma circunferência para funções. Vamos definir retas secantes e tangentes ao gráfico de uma função com o auxílio das definições anteriores.

Definição 3.1.4. *Seja $f : A \subset \mathbb{R} \to \mathbb{R}$. Uma reta t é tangente ao gráfico de f quando, ao fixarmos um ponto $P = (x, f(x))$ para algum $x \in A$, existe uma circunferência tangente a t com ponto de tangência P de modo que a interterseção entre a circunferência e o gráfico de f seja o ponto P, isto é, o gráfico de f e a circunferência possuem apenas um único ponto em comum.*

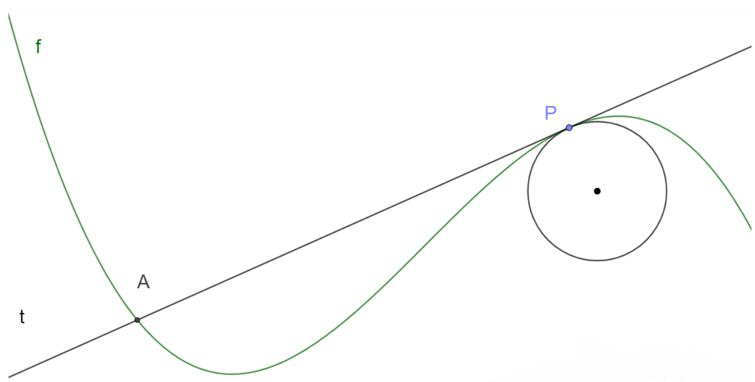


Figura 3.4: Reta t tangente ao gráfico de f.

Definição 3.1.5. *Seja $f : A \subset \mathbb{R} \to \mathbb{R}$. Uma reta s é secante ao gráfico de f quando, ao fixarmos um ponto $P = (x, f(x))$ para algum $x \in A$, não existe uma circunferência tangente a t com ponto de tangência P de modo que a interseção entre a circunferência e o gráfico de f seja o ponto P.*

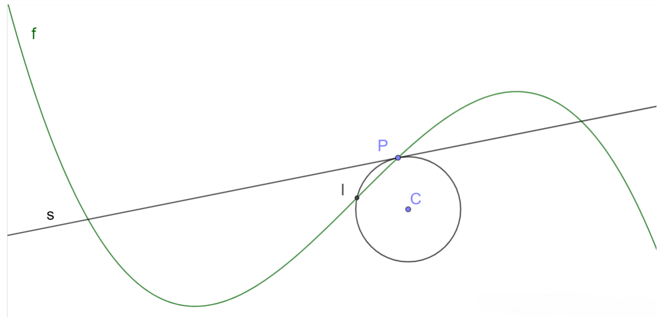


Figura 3.5: Reta s secante ao gráfico de f.

Observe que com a definição de reta tangente a uma circunferência, poderíamos ser tentados a definir uma reta tangente ao gráfico de uma função como a reta que intercepta o gráfico de f num único ponto. Mas note que com essa definição perderíamos generalidade em um certo grau, pois, na Figura 3.4 a reta t não seria tangente ao gráfico de f porque t passa por A e P, ambos pertencem ao gráfico de f. Entretanto, com a Definição 3.1.4, t é uma reta tangente ao gráfico de f que passa por P. Assim, a tangência entre uma reta e uma função é considerada localmente e isso é altamente desejado, já que a maioria das funções se comportam semelhante a retas em um intervalo suficientemente pequeno e dentre todas as retas é justamente a reta tangente que mais se aproxima desse comportamento, no sentido de que o gráfico da função e a reta tornam-se indistinguíveis.

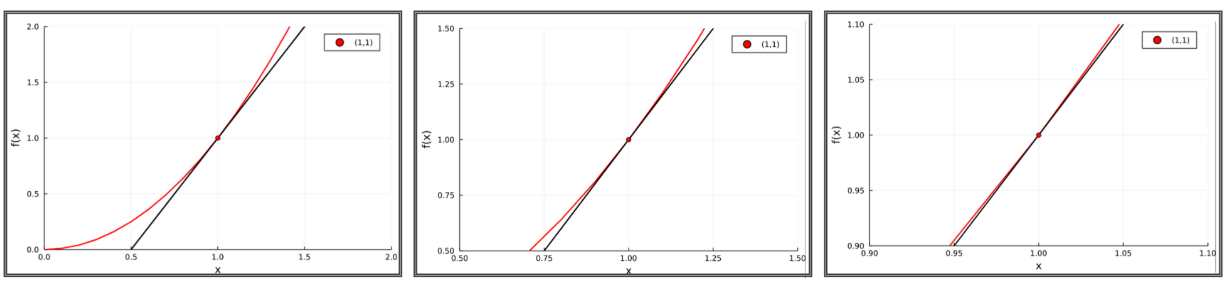


Figura 3.6: Comportamento localmente linear de uma função.

Antes de mostrarmos o motivo pelo qual determinar retas tangentes ao gráfico de funções é tão importante, mostraremos como tal procedimento é realizado.

3.2 Retas tangentes a gráficos de funções reais

Considere o problema de tentar determinar a reta tangente t ao gráfico de uma função f, em um dado ponto P. Uma vez que sabemos ser P um ponto sobre a reta tangente, podemos encontrar a equação de t se conhecermos o seu coeficiente angular, ou seja, a sua inclinação m. O problema está no fato de que, para calcular a inclinação, é necessário conhecer dois pontos e, sobre t, temos somente o ponto P. Para contornar esse problema, determinamos primeiro uma aproximação para m, tomando sobre o gráfico um ponto próximo Q e calculando a inclinação m_{PQ} da reta secante.

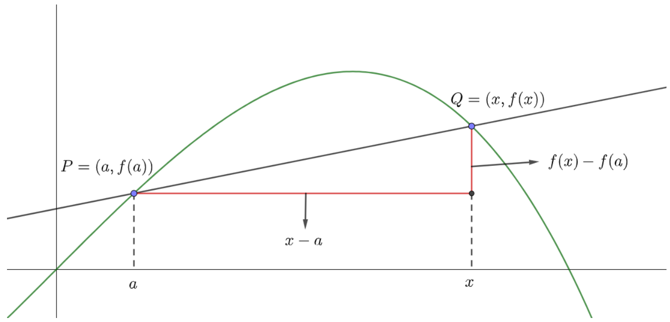


Figura 3.7: Aproximação da reta tangente.

Pela Figura 3.7 temos que

$$m_{PQ} = \frac{\Delta y}{\Delta x} = \frac{f(x) - f(a)}{x - a}. \tag{3.1}$$

Imagine agora o ponto Q movendo-se ao longo do gráfico em direção a P, conforme a Figura 3.8.

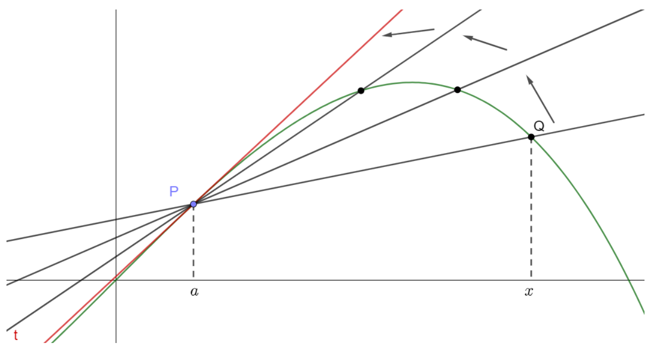


Figura 3.8: Aproximação dos pontos P e Q.

Podemos visualizar pela figura acima que a reta secante gira e aproxima-se da reta tangente como sua posição limite. Isso significa que a inclinação m_{PQ} da reta secante fica cada vez mais próxima da inclinação m da reta tangente. Isso é expresso por

$$m = \lim_{Q \to P} m_{PQ}$$

e dizemos que m é o limite de m_{PQ} quando Q tende a P ao longo do gráfico. Pela equação (3.1) podemos reescrever a equação acima como

$$m = \lim_{Q \to P} m_{PQ} = \lim_{Q \to P} \frac{f(x) - f(a)}{x - a}.$$

Lembre-se de que o ponto $P = (a, f(a))$ está fixado, logo os números reais a e $f(a)$ são fixos. Por outro lado, o ponto $Q = (x, f(x))$ está variando quando se aproxima de P, assim os números reais x e $f(x)$ estão variando. Como o número $f(x)$ depende de x, podemos reescrever $Q \to P$ em termos de x somente. Observe que se Q aproxima de P ao longo do gráfico, o valor da primeira coordenada de P se aproxima do valor da primeira coordenada de Q portanto, x se aproxima de a e consequentemente podemos escrever $x \to a$. Dessa forma,

$$m = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}. \tag{3.2}$$

O problema da reta tangente deu origem ao ramo do cálculo chamado cálculo diferencial, que possui inúmeras aplicações, dentre elas a determinação da velocidade instantânea de um corpo, mesmo quando a aceleração não é constante, o esboço de curvas definidas por funções e a resolução de problemas de otimização. Podemos dizer que muitos problemas práticos e teóricos que apresentam taxas de variação em relação ao tempo ou alguma outra variável foram resolvidos através do Cálculo Diferencial.

Vamos agora entender com mais detalhes o conceito matemático de limite, do qual utilizamos para o cálculo da inclinação m da reta tangente. Com isso, poderemos definir o que é a derivada de uma função e a função derivada, que associa a inclinação da reta tangente a cada ponto de uma determinada função previamente fixada. Dessa forma, resolveremos o mesmo problema porém de maneira algébrica. Graças a isso, será possível encontrar a derivada de funções sofisticadas atráves das regras de derivação.

3.3 Limite de uma função

Seja $f : A \subset \mathbb{R} \to \mathbb{R}$. Fixado um ponto $a \in A$, nosso objeivo é estudar o comportamento da função para valores próximos de a, mas não iguais a a.

Exemplos 3.3.1. *Vamos analisar o comportamento da função $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^2 - x + 2$. A Tabela 3.1 fornece os valores de $f(x)$ para valores de x próximos de 2, mas não iguais a 2.*

x	$f(x)$	x	$f(x)$
1	2,000000	3	8,000000
1,5	2,750000	2,5	5,750000
1,8	3,440000	2,2	4,640000
1,9	3,710000	2,1	4,310000
1,95	3,852500	2,05	4,152500
1,99	3,970100	2,01	4,030100
1,995	3,985025	2,005	4,015025
1,999	3,997001	2,001	4,003001

Tabela 3.1: Valores de $f(x)$ para valores de x próximos de 2.

Apresentamos abaixo a representação gráfica da função f.

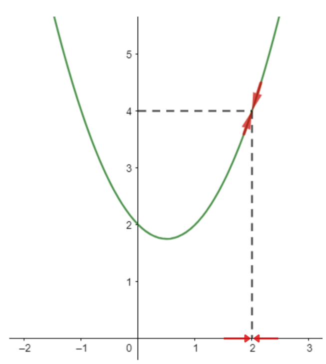


Figura 3.9: Comportamento da função f para valores próximos a 2.

Da tabela e do gráfico de f, vemos que quanto mais próximo x estiver de 2 (seja por valores menores ou maiores do que 2), mais próximo de $f(x)$ estará de 4. De fato, parece que podemos tornar os valores $f(x)$ tão próximos de 4 quanto quisermos, ao tornar x suficientemente próximos de 2. Expressamos isso dizendo que "o limite da função $f(x) = x^2 - x + 2$ quando x tende a 2 é igual a 4". A notação para isso é

$$\lim_{x\to 2}(x^2 - x + 2) = 4.$$

Exemplos 3.3.2. *Vamos calcular o limite das funções $g : \mathbb{R} \to \mathbb{R}$ e $h : \mathbb{R} \setminus \{2\} \to \mathbb{R}$ dadas pelas regras*

$$g(x) = \begin{cases} x^2 - x + 2 & \text{se } x \neq 2 \\ 1,5 & \text{se } x = 2 \end{cases} \quad e \quad h(x) = x^2 - x + 2$$

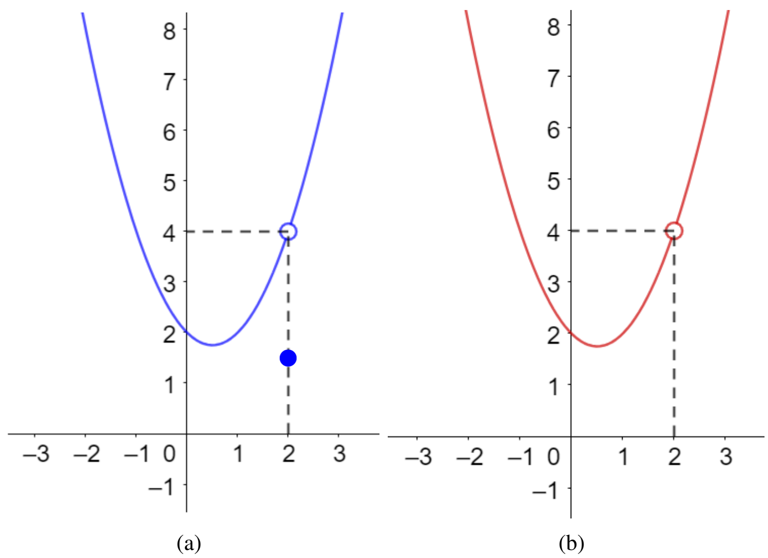


Figura 3.10: Gráfico da função g (a) e gráfico da função h (b).

Apesar de $g(2) = 1.5$ e h não estar definida em 2, observa-se que podemos tornar os valores $g(x)$ e $h(x)$ tão próximos de 4 quanto quisermos ao tornar x suficientemente próximo de 2. Portanto,

$$\lim_{x\to 2} g(x) = \lim_{x\to 2} h(x) = 4.$$

O exemplo acima nos mostra uma característica fundamental do limite de uma função, ao procurar o limite de $f(x)$ quando x tende a a, nunca consideramos $x = a$. Na verdade, $f(x)$ não precisa sequer estar definida quando $x = a$. A única coisa que importa é como f está definida próximo de a.

Com o que vimos até então, poderíamos tentar definir ingenuamente o limite de uma função da seguinte forma: suponha que f esteja definida em algum intervalo aberto que contenha a, exceto possivelmente no próprio a. Então escrevemos

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

e dizemos "o limite de $f(x)$, quando x tende a a, é igual a L" se pudermos tornar os valores de $f(x)$ arbitrariamente próximos de L ao tomar x suficientementemente próximo de a, mas não iguais a a. Entretanto, daremos um exemplo de uma função em que a definição dada acima falha.

Exemplos 3.3.3. *Analise o limite* $\lim_{x\to 0} \sin\frac{\pi}{x}$.
Solução: *A função* $f(x) = \sin(\pi/x)$ *não está definida em* 0*. Calculando a função para valores pequenos de x, temos*

x	1	1/2	1/3	1/4	1/10	1/100	1/1000	1/10000
$f(x)$	0	0	0	0	0	0	0	0

Tabela 3.2: Valores de $f(x)$ para valores de x próximos de 0.

Com base nessa informação, "ficaríamos tentados" a conjecturar que

$$\lim_{x\to 0} \sin\frac{\pi}{x} = 0.$$

Dessa vez, no entanto, nossa conjectura ***está errada!*** *Observe que, embora* $f(1/n) = \sin n\pi = 0$ *para todo número inteiro n, é também verdadeiro que* $f(x) = 1$ *para infinitos valores de x (tais como* 2/5 *ou* 2/10*) que tendem a* 0*. Podemos ver isso a partir do gráfico de f.*

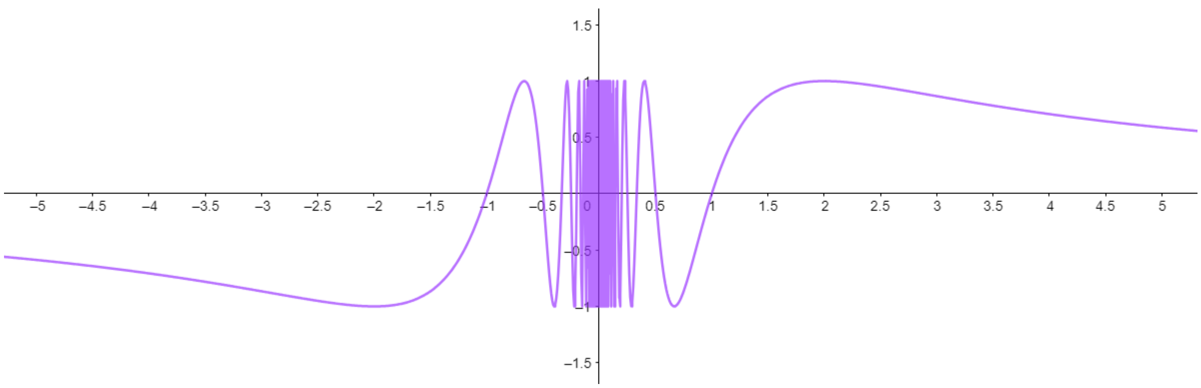


Figura 3.11: Gráfico da função $f(x) = \sin\frac{\pi}{x}$.

O comportamento do gráfico da função próximo ao eixo das ordenadas indica que os valores de $\sin(\pi/x)$ *oscilam entre* 1 *e* -1 *infinitas vezes quando x tende a* 0*. Uma vez que os valores* $f(x)$ *não tendem a um número fixo quando x tende a* 0 *portanto,*

$$\lim_{x\to 0} \sin\frac{\pi}{x} \textbf{ não existe}.$$

O exemplo acima ilustra algumas da armadilhas na conjectura sobre o valor de um limite. É fácil conjecturar um valor falso se usarmos valores não apropriados de x e, além disso, é difícil saber quando parar de calcular valores. Por isso, uma definição rigorosa é necessaria para obtermos um método eficaz no cálculo de limites.

O principal motivo para que a definição de limite de uma função dada seja falha está no fato de que frases como "x está próximo de 2" e "$f(x)$ aproxima-se cada vez mais de L" são vagas. Para se chegar a definição precisa de limite, consideremos a função

$$f(x) = \begin{cases} 2x-1 & \text{se } x \neq 3 \\ 6 & \text{se } x = 3 \end{cases}.$$

É intuitivamente claro que quando x está próximo de 3, mas $x \neq 3$, então $f(x)$ está próximo de 5 e, sendo assim, $\lim_{x\to 3} f(x) = 5$.

Para obter informações mais detalhadas sobre como $f(x)$ varia quando x está próximo de 3, faremos a seguinte pergunta: o quão próximo de 3 deverá estar x para que a diferença entre $f(x)$ e 5 seja menor que 0,1? A distância de x a 3 é $|x-3|$ e a distância de $f(x)$ a 5 é $|f(x)-5|$, logo nosso problema é achar um número δ tal que

$$|f(x)-5| < 0,1 \text{ se } |x-3| < \delta \text{ mas } x \neq 3.$$

Se $|x-3| > 0$, então $x \neq 3$; portanto uma formulação equivalente de nosso problema é achar um número δ tal que

$$|f(x)-5| < 0,1 \text{ se } 0 < |x-3| < \delta.$$

Observe que, se $0 < |x-3| < (0,1)/2 = 0,05$, então

$$|f(x)-5| = |(2x-1)-5| = |2x-6| = 2|x-3| < 2(0,05) = 0,1,$$

isto é,

$$|f(x)-5| < 0,1 \text{ se } 0 < |x-3| < 0,05.$$

Assim, uma resposta para o problema é dada por $\delta = 0,05$. Isso significa que se x estiver a uma distância de no máximo 0,05 de 3, então $f(x)$ estará a uma distância de no máximo 0,1 de 5. Se mudarmos o número 0,1 em nosso problema para o número menor 0,01, então, usando o mesmo método, achamos que $f(x)$ diferirá de 5 por menos que 0,01, desde que x difira de 3 por menos que $(0,01)/2 = 0,005$. De forma análoga,

$$|f(x)-5| < 0,001 \text{ se } 0 < |x-3| < 0,0005.$$

Os números 0,1, 0,01 e 0,001, anteriormente considerados, podem ser interpretadas como tolerâncias de erro que podemos admitir. Devemos agora ser capazes de resolver esse problema para qualquer tolerância de erro estritamente positiva! Se chamamos de ε (lê-se: épsilon) a um número positivo arbitrário, então, encontramos

$$|f(x)-5| < \varepsilon \text{ se } 0 < |x-3| < \delta = \frac{\varepsilon}{2}.$$

Se a afirmação acima vale para qualquer $\varepsilon > 0$, temos que $f(x)$ está próximo de 5 quando x está próximo de 3, pois podemos fazer os valores de $f(x)$ ficarem dentro de uma distância arbitrária ε de 5 restringindo os valores de x a uma distância $\varepsilon/2$ de 3 (sendo $x \neq 3$).

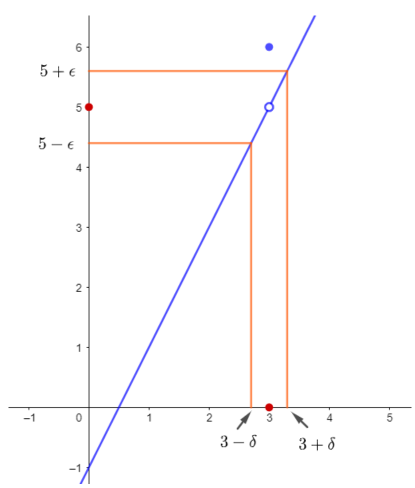


Figura 3.12: Valores de $f(x)$ possíveis para $0 < |x-3| < \delta$.

Com isso, podemos definir precisamente o limite de uma função.

Definição 3.3.4. *Seja f uma função definida em algum intervalo aberto que contenha o número a, exceto possivelmente a. Então dizemos que o **limite de** $f(x)$ **quando** x **tende a** a **é** L, e escrevemos*

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

se para todo número $\varepsilon > 0$ houver um número $\delta > 0$ tal que

$$0 < |x - a| < \delta \text{ implique } |f(x) - L| < \varepsilon.$$

3.4 Derivadas

A equação (3.2) ocorre sempre que calculamos uma taxa de variação. Uma vez que esse tipo de limite ocorre amplamente, ele recebe nome e notação especiais.

Definição 3.4.1. *A **derivada de uma função f em um número** a, denotada por $f'(a)$, é definida por*

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a},$$

se o limite existir.

Se considerarmos $h = x - a$ podemos reescrever $f'(a)$ como um limite em h da seguinte forma. Se $h = x - a$, então $x = a + h$ além disso, quanto $x \longrightarrow a$ temos que $h \longrightarrow 0$. Daí, segue que

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

se o limite existir. Uma das vantagens dessa mudança de varíavel no limite acima é que poderemos definir a função derivada, como veremos a seguir. Antes, apresentaremos alguns exemplos.

Exemplo 3.4.1. *Encontre a derivada da função $f(x) = x^2 - 8x + 9$ em um número a.*
Solução:

$$\begin{aligned} f'(a) &= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \\ &= \lim_{x \to a} \frac{(x^2 - 8x + 9) - (a^2 - 8a + 9)}{x - a} \\ &= \lim_{x \to a} \frac{(x^2 - a^2) - 8(x - a)}{x - a} \\ &= \lim_{x \to a} \frac{(x + a)(x - a) - 8(x - a)}{x - a} \\ &= \lim_{x \to a} (x + a - 8) \\ &= 2a - 8. \end{aligned}$$

Definimos a reta tangente ao gráfico de uma função f no ponto $P = (a, f(a))$ como a reta que passa em P e tem inclinação m dada pela Equação (3.2). Uma vez que, pela Definição 3.4.1, isso é o mesmo que a derivada $f'(a)$, podemos agora dizer o seguinte: A reta tangente ao gráfico de uma função f em $(a, f(a))$ é a reta que passa em $(a, f(a))$, cuja inclinação é igual a $f'(a)$, a derivada de f em a.

Se usarmos a forma ponto-inclinação da equação de uma reta, poderemos escrever uma equação da reta tangente ao gráfico de f no ponto $(a, f(a))$,

$$y - f(a) = f'(a)(x - a).$$

Exemplo 3.4.2. *Encontre uma equação da reta tangente à parábola* $y = x^2 - 8x + 9$ *no ponto* $(3, -6)$.

Solução: *Do exemplo anterior, sabemos que a derivada de* $f(x) = x^2 - 8x + 9$ *no número a é* $f'(a) = 2a - 8$. *Portanto, a inclinação da reta tangente em* $(3, -6)$ *é* $f'(3) = 2(3) - 8 = -2$. *Dessa forma, uma equação da reta tangente, é*

$$y = -(-6) = (-2)(x - 3) \text{ ou } y = -2x.$$

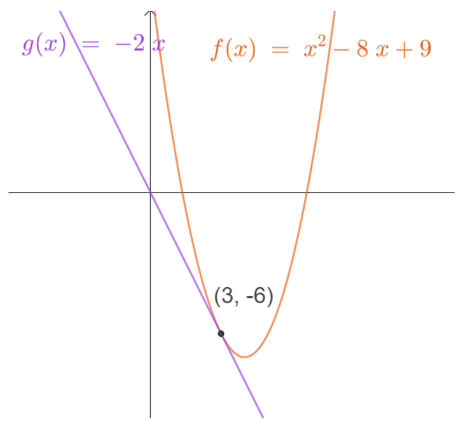


Figura 3.13: Reta tangente à parábola no ponto $(3, -6)$.

Até então, consideramos a derivada de uma função $f : A \subset \mathbb{R} \to \mathbb{R}$ em um número fixo a:

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

Se substituirmos a por uma variável x, podemos definir uma função $f' : B \subset A \to \mathbb{R}$ de modo que

$$f'(x) = \lim_{h \to 0} \frac{f(x + h) - f(x)}{h}, \tag{3.3}$$

onde $B = \{x : f'(x) \text{ existe}\}$.

Exemplo 3.4.3. *Se* $f(x) = x^3 - x$, *encontre uma fórmula para* $f'(x)$. *Em seguida, esboçe os gráficos de* f *e* f'.

Solução:

$$\begin{aligned}
f'(x) &= \lim_{x \to 0} \frac{f(x + h) - f(x)}{h} \\
&= \lim_{x \to 0} \frac{((x + h)^3 - (x + h)) - (x^3 - x)}{h} \\
&= \lim_{x \to 0} \frac{(x^3 + 3x^2h + 3xh^2 + h^3) - (x + h) - (x^3 - x)}{h} \\
&= \lim_{x \to 0} \frac{3x^2h + 3xh^2 + h^3 - h}{h} \\
&= \lim_{x \to 0} (3x^2 + 3xh + h^2 - 1) \\
&= 3x^2 - 1.
\end{aligned}$$

Abaixo mostramos o esboço dos gráficos:

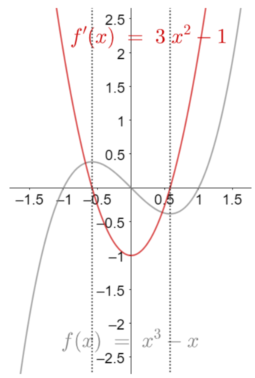


Figura 3.14: Gráfico das funções f e f'.

Observe que f decresce sempre que o gráfico de f' está abaixo do eixo das abscissas além disso, f cresce quando o gráfico de f' está acima do eixo das abscissas, ou seja, se $f'(x) < 0$ em um intervalo, então f é decrescente nele, caso $f'(x) > 0$ temos que f é crecente no intervalo.

Com o exemplo acima, podemos observar que dada uma função real, um problema natural seria determinar qual é a sua função derivada. Usamos (3.3) para calcular as derivadas de funções definidas por fórmulas. Mas seria muito trabalhoso se sempre usássemos a definição. Por isso, mostraremos as regras de derivação, que nos permitirão calcular com mais praticidade as derivadas de funções polinomiais e de outros tipos.

3.5 Regras de derivação

Inicialmente, vamos considerar a função constante $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = c$.

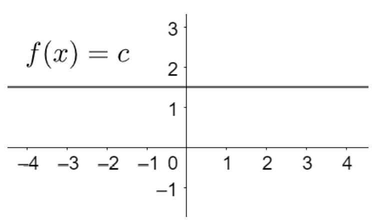


Figura 3.15: Função constante.

Observe que o gráfico de uma função constante tem inclinação igual a 0, pois os valores das ordenadas não variam. Logo, é plausível supor que $f'(x) = 0$. De fato,

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

Isso prova o seguinte resultado:

Proposição 3.5.1. *Seja $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = c$, para algum $c \in \mathbb{R}$. A função derivada $f' : \mathbb{R} \to \mathbb{R}$ será dada por $f'(x) = 0$.*

A seguir, apresentaremos outras regras de derivação igualmente úteis, das quais faremos uso sem a necessidade de prova, uma vez que estas podem ser consultadas em [6].

Proposição 3.5.2 (Regra da Potência)**.** *Seja $f : A \subset \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^n$, para algum $n \in \mathbb{R} \setminus \{0\}$. A função derivada $f' : B \subset A \to \mathbb{R}$ será dada por $f'(x) = nx^{n-1}$ se $n > 0$. Caso $n < 0$, o domínio da f' será $B \setminus \{0\}$.*

Exemplos 3.5.1. *Derive as funções $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ e $g : \mathbb{R} \to \mathbb{R}$, onde $f(x) = \frac{1}{x^2}$ e $g(x) = \sqrt[3]{x^2}$.*
Solução: *Em cada caso, reescreveremos a função como uma potência de x. Uma vez que $f(x) = x^{-2}$, usamos a Regra da Potência com $n = -2$. Logo,*

$$f'(x) = (-2)x^{(-2)-1} = -2x^{-3} = -\frac{2}{x^3},$$

Analogamente, como $g(x) = x^{\frac{2}{3}}$, podemos derivar g usando a regra da potência para $n = \frac{2}{3}$. Dessa forma,

$$g'(x) = \left(\frac{2}{3}\right)x^{\frac{2}{3}-1} = \frac{2}{3}x^{-\frac{1}{3}} = \frac{2}{3x^{\frac{1}{3}}} = \frac{2}{3\sqrt[3]{x}}.$$

Proposição 3.5.3 (Regra da Multiplicação por Constante)**.** *Seja $f : A \subset \mathbb{R} \to \mathbb{R}$ uma função derivável, isto é, a derivada de f existe, e seja $g : A \to \mathbb{R}$ tal que $g(x) = cf(x)$ para algum $c \in \mathbb{R}$. A função derivada $g' : B \subset A \to \mathbb{R}$ será dada pela regra $g'(x) = cf'(x)$.*

Exemplo 3.5.1. *Derive a função $g : \mathbb{R} \to \mathbb{R}$ definida por $g(x) = 3x^4$.*
Solução: *Note que se definirmos $f(x) = x^4$ para todo x pertencente ao domínio de g, podemos escrever $g(x) = 3f(x)$. Assim, pela regra da multiplicação por constante, $g'(x) = 3f'(x)$. Como $f'(x) = 4x^3$, pela regra da potência, concluímos que*

$$g'(x) = 3f'(x) = 3(4x^3) = 12x^3.$$

Proposição 3.5.4 (Regra da Soma e da Diferença)**.** *Sejam $f, g : A \subset \mathbb{R} \to \mathbb{R}$ funções deriváveis e $F : A \to \mathbb{R}$ tal que $F(x) = f(x) \pm g(x)$. A função derivada $F' : B \subset A \to \mathbb{R}$ será dada pela regra $F'(x) = f'(x) \pm g'(x)$.*

Combinando todas as regras de derivação dadas acima, podemos derivar qualquer polinômio.

Exemplo 3.5.2. *Derive a função $f : \mathbb{R} \to \mathbb{R}$ definida por $f(x) = x^8 + 12x^5 - 4x^4 + 10x^3 - 6x + 5$.*
Solução:

$$\begin{aligned}
f'(x) &= (x^8 + 12x^5 - 4x^4 + 10x^3 - 6x + 5)' \\
&= (x^8)' + (12x^5)' + (4x^4)' + (10x^3)' - (6x)' + (5)' && \textit{(Regra da Soma e da Diferença)} \\
&= (x^8)' + (12x^5)' + (4x^4)' + (10x^3)' - (6x)' + 0 && \textit{(Regra da Função constante)} \\
&= (x^8)' + 12(x^5)' + 4(x^4)' + 10(x^3)' - 6(x^1)' && \textit{(Regra da Multiplicação por Constante)} \\
&= 8x^{8-1} + 12(5x^{5-1}) + 4(4x^{4-1}) + 10(3x^{3-1}) - 6(1x^{1-1}) && \textit{(Regra da Potência)} \\
&= 8x^7 + 60x^4 + 16x^3 + 30x^2 - 6.
\end{aligned}$$

Proposição 3.5.5 (Regra do Produto)**.** *Sejam $f, g : A \subset \mathbb{R} \to \mathbb{R}$ funções deriváveis e $F : A \to \mathbb{R}$ tal que $F(x) = f(x)g(x)$. A função derivada $F' : B \subset A \to \mathbb{R}$ será dada pela regra $F'(x) = f(x)g'(x) + f'(x)g(x)$.*

Exemplo 3.5.3. *Derive a função $F(x) = (3x^2 + 5x)(x^2 + 2x + 1)$.*
Solução: *Considerando $f(x) = 3x^2 + 5$ e $g(x) = x^2 + 2x + 1$, obtemos que $f'(x) = 6x$ e $g'(x) = 2x + 2$. Assim, pela regra do produto,*

$$\begin{aligned}
F'(x) &= f(x)g'(x) + f'(x)g(x) \\
&= (3x^2 + 5)(2x + 2) + 6x(x^2 + 2x + 1) \\
&= (6x^3 + 6x^2 + 10x + 10) + (6x^3 + 12x^2 + 6x) \\
&= 12x^3 + 18x^2 + 16x + 10.
\end{aligned}$$

Proposição 3.5.6 (Regra do Quociente)**.** *Sejam $f, g : A \subset \mathbb{R} \to \mathbb{R}$ funções deriváveis e $F : A \to \mathbb{R}$ tal que $F(x) = f(x)/g(x)$ com $g(x) \neq 0$. A função derivada $F' : B \subset A \to \mathbb{R}$ será dada pela regra*

$$F'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2}.$$

Exemplos 3.5.2. *Derive a função $F(x) = (x^2 + x - 2)/(x^3 + 6)$.*
Solução: *Considerando $f(x) = x^2 + x - 2$ e $g(x) = x^3 + 6$, obtemos que $f'(x) = 2x + 1$ e $g'(x) = 3x^2$. Assim, pela regra do quociente,*

$$\begin{aligned} F'(x) &= \frac{g(x)f'(x) - f(x)g'(x)}{(g(x))^2} \\ &= \frac{(x^3 + 6)(2x + 1) - (x^2 + 2 - 2)(3x^2)}{(x^3 + 6)^2} \\ &= \frac{(2x^4 + x^3 + 12x + 6) - (3x^4 + 3x^3 - 6x^2)}{(x^3 + 6)^2} \\ &= \frac{-x^4 - 2x^3 + 6x^2 + 12x + 6}{(x^3 + 6)^2}. \end{aligned}$$

Proposição 3.5.7 (Regra da Cadeia)**.** *Se f for derivável em x e f for derivável em $g(x)$, então a função composta $F = f \circ g$ definida por $F(x) = f(g(x))$ é derivável em x, e F' é dada pelo produto*

$$F'(x) = f'(g(x)) \cdot g'(x).$$

Exemplos 3.5.3. *Encontre $F'(x)$ com $F(x) = \sqrt{x^2 + 1}$.*
Solução: *Primeiro, vamos encontrar duas funções f e g tais que $F = f \circ g$. Para isso, note que se $f(x) = \sqrt{x}$ e $g(x) = x^2 + 1$, temos $(f \circ g)(x) = f(g(x)) = \sqrt{g(x)} = \sqrt{x^2 + 1} = F(x)$. Como $f'(x) = \frac{1}{2\sqrt{x}}$ (pela regra da potência) e $g'(x) = 2x$, pela regra da cadeia, temos*

$$F'(x) = f'(g(x))g'(x) = \frac{1}{2\sqrt{g(x)}}(2x) = \frac{x}{\sqrt{x^2 + 1}}.$$

3.6 Valores de máximo e de mínimo de funções reais

Uma das aplicações mais importantes do Cálculo está na resolução de problemas da Otimização, nos quais buscamos encontrar a maneira mais eficiente (ou uma solução ótima) de realizar uma determinada tarefa. Esses problemas são frequentemente reduzidos ao problema de encontrar os valores máximos ou mínimos de uma função. Antes de prosseguirmos, é fundamental esclarecer o que significam os termos “valor máximo” e “valor mínimo” locais. Para isso, vamos considerar o gráfico de uma função dado pela Figura 3.16.

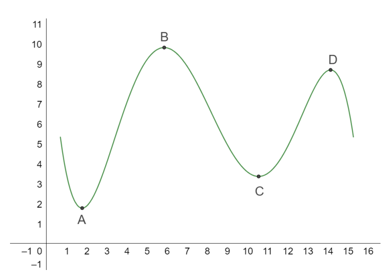


Figura 3.16: Gráfico de uma função.

Observe que o valor mais alto da função mostrado na figura 3.16 é o ponto B. Da mesma forma, o menor valor é a ordenada do ponto A. Em geral, usamos a seguinte definição que é análoga a Definição 2.4.1.

Definição 3.6.1. *Seja c um elemento do domínio D de uma função f. Então $f(c)$ é o valor* ***máximo absoluto*** *de D em f se $f(c) \geq f(x)$ para todo x em D.*

Analogamente, podemos definir o que é o mínimo global de uma função.

Definição 3.6.2. *Seja c um elemento do domínio D de uma função f. Então $f(c)$ é o valor* ***mínimo absoluto*** *de D em f se $f(c) \leq f(x)$ para todo x em D.*

Exemplo 3.6.1. *Analisando a Figura 3.16, observamos que a ordenada do ponto A corresponde ao mínimo absoluto, enquanto a ordenada do ponto B representa o máximo absoluto.*

Um máximo ou mínimo absoluto também é chamado de máximo ou mínimo **global**. Os valores máximos e mínimos de f são chamados de **valores extremos** de f.

Observe que os pontos C e D da Figura 3.16 não são necessariamente os valores extremos da função. Entretanto, note que no intervalo $(9, 13)$, C é o ponto mais abaixo do gráfico, enquanto no intervalo $(13, 15)$, D é o ponto mais acima do gráfico. Para destarcarmos esta observação definiremos o seguinte:

Definição 3.6.3. *Seja c um elemento do domínio D de uma função f. Então $f(c)$ é o valor* ***mínimo local*** *de D em f se existe um $C \subsetneq D$ tal que $f(c) \leq f(x)$ para todo x em C.*

Definição 3.6.4. *Seja c um elemento do domínio D de uma função f. Então $f(c)$ é o valor* ***máximo local*** *de D em f se existe um $C \subsetneq D$ tal que $f(c) \geq f(x)$ para todo x em C.*

Note que, pelas definições acima, todo mínimo global também é um mínimo local, pois tal valor é o menor valor assumido pela função em quaisquer subconjuntos de seu domínio. O mesmo argumento pode ser usado para mostrar que todo máximo global também é um máximo local.

Vale ressaltar que os pontos A, B, C e D presentes na Figura 3.16 possuem algo em comum. Para isso, vamos traçar retas tangentes à função em pontos próximos de A e D.

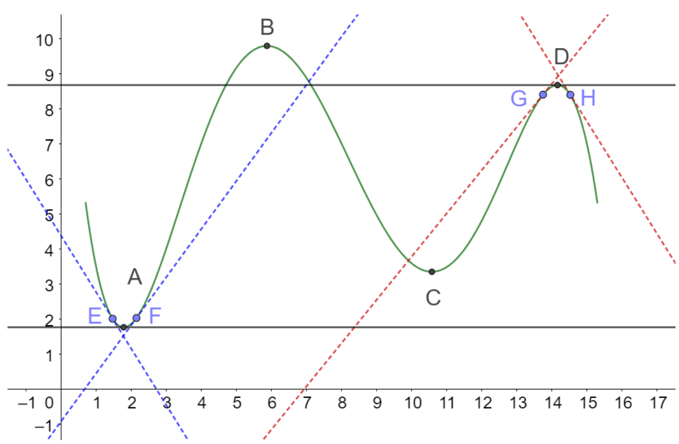


Figura 3.17: Retas tangentes ao gráfico da função.

Perceba que na Figura 3.17, as retas tangentes (retas tracejadas) à função que passam pelos pontos E e F possuem inclinações positiva e negativa, respectivamente. À medida que esses pontos se aproximam de A, a inclinação das retas tangentes desses pontos se aproxima da inclinação da reta tangente à função que passa por A, a qual é nula. O mesmo ocorre em relação aos pontos D, G e H, com a inclinação da reta tangente à função que passa pelo ponto D sendo nula. Por fim, o mesmo ocorre com os pontos B e C. Portanto, a inclinação da reta tangente nula é o fator comum aos pontos de máximo ou mínimo. Isso é garantido pelo seguinte teorema, do qual admitiremos sua veracidade.

Teorema 3.6.1 (Teorema de Fermat)**.** *Se uma função f tiver um máximo ou mínimo local em c e se $f'(c)$ existir, então $f'(c) = 0$.*

Se existir um c tal que $f'(c)$ não exista, isso não implica necessariamente que $f(c)$ não seja um valor extremo. Por exemplo, na função $f : \mathbb{R} \to \mathbb{R}$ definida como $f(x) = |x|$, $f'(0)$ não existe; entretanto, $f(0) = 0$ é mínimo global.

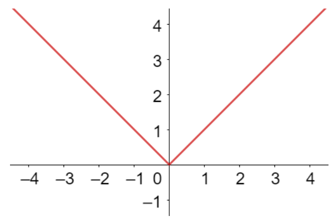


Figura 3.18: Representação gráfica da função modular $f(x) = |x|$.

Assim, dada uma função $f : A \subset \mathbb{R} \to \mathbb{R}$, os elementos de A que são candidatos a valores extremos são aqueles que possuem derivadas nulas ou os pontos nos quais as funções derivadas não existem. Dessa forma, o Teorema de Fermat sugere que devemos, pelo menos, começar procurando por valores extremos de f nos números c onde $f'(c) = 0$ ou onde $f'(c)$ não existe. Esses números recebem um nome especial, como veremos a seguir.

Definição 3.6.5. *Um **número crítico** de uma função f é um número c no domínio de f tal que ou $f'(c) = 0$ ou $f'(c)$ não existe.*

Com isso, podemos encontrar os valores de máximo e de mínimo absolutos de qualquer função f contínua (veja a Seção 3.8) em um intervalo fechado através do seguinte procedimento:

1. Encontre os valores de f nos números críticos de f em (a, b);
2. Encontre os valores de f nas extremidades do intervalo;
3. O maior valor entre as etapas 1 e 2 é o valor máximo absoluto. Por outro lado, o menor desses valores é o valor mínimo absoluto.

Exemplo 3.6.2. *Encontre os valores máximo e mínimo absolutos da função*

$$f(x) = x^3 - 3x^2 + 1, \quad -\frac{1}{2} \leq x \leq 4.$$

Solução: *Vamos encontrar os pontos críticos de f. Note que*

$$f'(x) = (x^3 - 3x^2 + 1)' = 3x^2 - 6x = 3x(x - 2).$$

Fazendo $f'(x) = 0$, temos a equação $3x(x-2) = 0$, que possui solução $x = 0$ ou $x = 2$. Estes são os números críticos de f. Os valores de f nestes números críticos são $f(0) = 1$ e $f(2) = -3$. Os valores de f nas extremidades do intervalo são $f\left(-\frac{1}{2}\right) = \frac{1}{8}$ e $f(4) = 17$. Comparando esses quatro números, vemos que o valor máximo absoluto é $f(4) = 17$ e o valor mínimo absoluto é $f(2) = -3$.

Observe que neste exemplo, o máximo absoluto ocorre em uma extremidade, enquanto o mínimo absoluto acontece em um número crítico.

3.7 Problemas de Otimização

Problemas que envolvem encontrar pontos extremos de funções têm muitas aplicações práticas. Por exemplo, um empresário pode desejar minimizar os custos e maximizar os lucros, enquanto um viajante pode buscar minimizar o tempo de transporte, entre outros problemas.

Os problemas de otimização apresentados no Capítulo 2, nos quais procuramos determinar extremos de funções, foram resolvidos facilmente utilizando as coordenadas do vértice da parábola. Entretanto, esse método se restringe a funções quadráticas e nem sempre será simples, dependendo da função que se pretende minimizar. Os conceitos apresentados neste capítulo podem ser utilizados para encontrar valores extremos de funções de qualquer tipo. Nesta seção, mostraremos um problema de minimização e uma proposta de resolução que envolve o uso de derivadas.

Exemplo 3.7.1. *Uma lata cilíndrica é feita para receber um litro de óleo. Encontre as dimensões que minimizarão o custo do metal para produzir a lata.*

Resolução: A figura a seguir ilustra a lata cilíndrica do problema, onde r é o raio da base e h é a altura da lata.

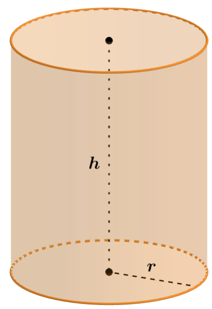


Figura 3.19: Cilindro de raio r e altura h.

A fim de minimizar o custo do metal, minimizamos a área da superfície total do cilindro (tampa, base e lado). A superfície lateral é uma folha retangular com dimensões $2\pi r$ e h e área $2\pi r h$ e cada uma das bases possui raio r e área πr^2. Logo, a área da superfície é

$$A = 2\pi r^2 + 2\pi r h.$$

Gostaríamos de expressar A em termos da variável r apenas. Para eliminarmos h, usamos o fato de que o volume é de $1L = 1000cm^3$. Logo

$$\pi r^2 h = 1000 \Longrightarrow h = \frac{1000}{\pi r^2}.$$

Substituindo na expressão da área A da superfície da lata, temos

$$A = 2\pi r^2 + 2\pi r\left(\frac{1000}{\pi r^2}\right) = 2\pi r^2 + \frac{2000}{r}.$$

A única restrição é que $r > 0$ (pois as soluções com $r \leq 0$ não possuem significado geométrico na aplicação). Portanto, a função que gostaríamos de minimizar é

$$A(r) = 2\pi r^2 + \frac{2000}{r}, \quad r > 0.$$

Para encontrarmos os números críticos, calculamos a derivada da função A :

$$A'(r) = 4\pi r - \frac{2000}{r^2} \Longrightarrow A'(r) = \frac{4(\pi r^3 - 500)}{r^2}.$$

Fazendo $A'(r) = 0$, obtemos

$$A'(r) = 0 \Longrightarrow \frac{4(\pi r^3 - 500)}{r^2} = 0 \Longrightarrow \pi r^3 - 500 = 0 \Longrightarrow r^3 = \frac{500}{\pi} \Longrightarrow r = \sqrt[3]{\frac{500}{\pi}}.$$

Podemos observar que $A'(r) < 0$ para $r < \sqrt[3]{500/\pi}$ e $A'(r) > 0$ para $r > \sqrt[3]{500/\pi}$. Portanto, A está decrescendo para todo r à esquerda do número crítico e crescendo para todo r à direita. Assim, $r = \sqrt[3]{500/\pi}$ deve originar um mínimo absoluto. O valor de h correspondente a $r = \sqrt[3]{500/\pi}$ é

$$h = \frac{1000}{\pi r^2} = \frac{1000}{\pi(500/\pi)^{2/3}} = 2\sqrt[3]{\frac{500}{\pi}} = 2r.$$

Dessa forma, para minimizar o custo da lata, o raio deve ser $\sqrt[3]{500/\pi}\ cm$ e a altura igual a duas vezes o raio, ou seja, $h = 2\sqrt[3]{500/\pi}\ cm$.

3.8 Continuidade de funções reais

Nesta seção, apresentaremos uma propriedade crucial para o campo da Otimização Contínua: a continuidade de uma função. Considere a função $f : \mathbb{R} \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} x+1 & \text{se } x \neq 2 \\ 1 & \text{se } x = 2 \end{cases}.$$

Considere a sua representação gráfica dada na Figura 3.20

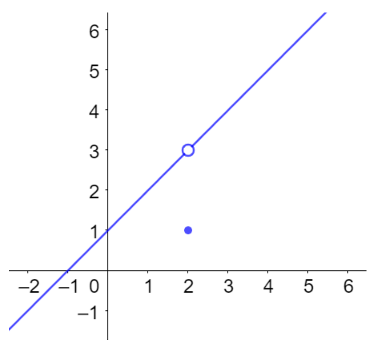


Figura 3.20: Gráfico da f.

Vamos analisar o comportamento da distância entre valores que a função assume, isto é, $|f(x) - f(y)|$ para $x \neq y$. Fixando $x = 2$ e tomando $y = 1$, temos $|f(x) - f(y)| = |f(2) - f(1)| = |1 - 2| = 1$, ou seja, uma mudança em uma unidade em x produziu uma alteração de uma unidade em $f(x)$. Analogamente, fazendo $y = 1,9$ observamos que $|f(x) - f(y)| = 1,9$ além disso, considerando $y = 1,99$ segue que $|f(x) - f(y)| = 1,99$. Com essas informações, podemos criar a seguinte tabela:

y	$\|x-y\|$	$\|f(x)-f(y)\|$	$\frac{\|f(x)-f(y)\|}{\|x-y\|}$
1	1	1	1
1,9	0,1	1,9	19
1,99	0,01	1,99	199

Tabela 3.3: Comparação das distâncias considerando $x = 2$.

Usando a Tabela 3.3, podemos verificar que pequenas variações em x produzem grandes variações em $f(x)$. Essa variação de $f(x)$ pode ser mantida tão grande quanto desejarmos, mantendo-se a variação em x suficientemente pequena.

Agora, iremos reanalisar o problema. Vamos considerar $x = 1$ e, assim, obter as seguintes informações:

y	$\lvert x-y\rvert$	$\lvert f(x)-f(y)\rvert$	$\frac{\lvert f(x)-f(y)\rvert}{\lvert x-y\rvert}$
0	1	1	1
0,9	0,1	0,1	1
0,99	0,01	0,01	1

Tabela 3.4: Comparação das distâncias considerando $x = 1$.

Usando a Tabela 3.4, podemos verificar que pequenas variações em x produzem pequenas variações em $f(x)$. Essa variação de $f(x)$ pode ser mantida tão pequena quanto desejarmos, mantendo-se a variação em x suficientemente pequena.

Observe que, ao analisarmos a variação de $f(x)$ em função da variação de x em dois pontos, conseguimos verificar dois comportamentos distintos. Em particular, quando uma função se comporta como no segundo caso em um certo ponto, digamos a, dizemos que a função f é contínua em a. Em termos de limites, podemos definir a continuidade de uma função em um ponto da seguinte forma:

Definição 3.8.1. *Uma função $f : A \subset \mathbb{R} \to \mathbb{R}$ é **contínua** em $a \in A$ se*

$$\lim_{x \to a} f(x) = f(a).$$

Quando uma função se comporta como no primeiro caso em um determinado ponto, dizemos que a função é **descontínua** neste ponto. Há uma grande variedade de funções que são contínuas em todos os pontos de seu domínio, dentre elas, as funções polinomiais, racionais, trigonométricas e exponenciais.

Muitos resultados importantes do Cálculo levam em consideração a continuidade da função como hipótese, como por exemplo o Teorema do Valor Intermediário, o Teorema do Valor Médio e o próprio Teorema Fundamental do Cálculo. Dentre todos esses resultados, segue o enunciado de um resultado extremamente importante na Otimização Contínua:

Teorema 3.8.1 (Teorema de Weierstrass)**.** *Se f for contínua em um intervalo fechado $[a, b]$, então existem $c, d \in [a, b]$ tais que $f(c) \leq f(x)$ e $f(d) \geq f(x)$ para todo $x \in [a, b]$.*

Observemos que o teorema acima garante a existência de valores extremos em funções contínuas definidas em intervalos fechados. Por fim, o teorema acima é uma versão particular da que é usada em Otimização Contínua do qual considera funções $f : A \subset \mathbb{R}^n \to \mathbb{R}$ com $n \in \mathbb{N}$, tais funções serão abordadas no próximo capítulo.

3.9 Exercícios

Exercício 3.9.1. *Encontre a derivada da função* $f(x) = x^2 - 8x + 9$ *em um número a usando*

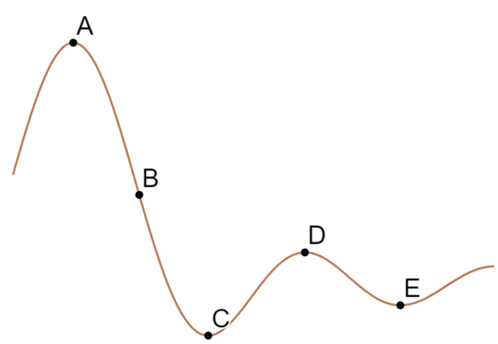
$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Exercício 3.9.2. *Encontre uma equação da reta tangente à parábola* $y = 4x - 3x^2$ *no ponto* $(2, -4)$*. Em seguida, esboçe a reta e o gráfico no mesmo plano cartesiano.*

Exercício 3.9.3. *Seja* $f : \mathbb{R} \to \mathbb{R}$ *tal que* $f(x) = 5x^2 + 4x$*, encontre* $f'(x)$*.*

Exercício 3.9.4. *Seja* $f : \mathbb{R} \to \mathbb{R}$ *tal que* $f(x) = \sqrt{x^2 + 4}$*, encontre* $f'(x)$*.*

Exercício 3.9.5. *Para cada um dos pontos indicados A, B, C, D e E diga se este ponto é de máximo ou mínimo absoluto, máximo ou mínimo local, ou nem máximo nem mínimo de acordo com a figura abaixo.*



Exercício 3.9.6. *Encontre os números críticos da função* $f : \mathbb{R} \to \mathbb{R}$ *dada por* $f(x) = 2x^3 - 3x^2 - 36x$*.*

Exercício 3.9.7. *Encontre os valores máximo e mínimo absolutos da função* $f : \mathbb{R} \to \mathbb{R}$ *cuja regra é* $f(x) = 2x^3 - 3x^2 - 12x + 1$ *no intervalo* $[-2, 3]$*.*

Exercício 3.9.8. *Encontre os valores máximo e mínimo absolutos da função* $f : \mathbb{R} \to \mathbb{R}$ *cuja regra é* $f(x) = 3x^4 - 4x^3 - 12x^2 + 1$ *no intervalo* $[-2, 3]$*.*

Exercício 3.9.9. *A taxa (em* mg *de carbono/*m^3*/h) na qual a fotossíntese ocorre para uma espécie de fitoplâncton é modelada pela função*

$$P = \frac{100I}{I^2 + I + 4}$$

em que I é a intensidade da luz (medida em milhares de candelas). Para qual intensidade de luz P é máximo.

Exercício 3.9.10. *Encontre o ponto sobre a reta* $y = 2x + 3$ *que está mais próximo da origem.*

Capítulo 4
Cálculo Diferencial com Duas Variáveis reais

Neste capítulo, abordaremos tópicos relacionados a funções de várias variáveis, expandindo os conceitos e resultados do Cálculo Diferencial estudados no capítulo anterior.

4.1 Funções de duas variáveis reais

Funções matemáticas são utilizadas para modelar diversos fenômenos do cotidiano. Estes problemas podem ser complexos e necessitarem de modelagem com funções que dependem de mais de uma variável. Um exemplo disso é a produção de uma loja de roupas (P), que depende tanto da quantidade de trabalhadores x quanto da quantidade de máquinas y que a empresa dispõe. Assim, a produção pode ser representada pela função $P = f(x, y) = 3x + 4y$, de modo que para avaliar a produção dessa loja, é necessário considerar ambas as variáveis. Outro exemplo é o cálculo da área de um retângulo de largura l e comprimento c. Podemos representar sua área pela função $A(l, c) = l \cdot c$. Para $l = 2$ e $c = 3$ teremos $A(2, 3) = 2 \cdot 3 = 6$ u.a..

Formalmente, uma função de duas variáveis reais pode ser definida como segue.

Definição 4.1.1. *Uma função de duas variáveis $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ é uma regra que associa a cada par ordenado de números reais $(x, y) \in D$ um único valor real $f(x, y) \in \mathbb{R}$.*

Escrevemos $z = f(x, y)$, em que x e y são as váriaveis independentes e z a variável dependente. A seguir, veremos como uma função de várias variáveis pode ser representada pelo diagrama de flechas.

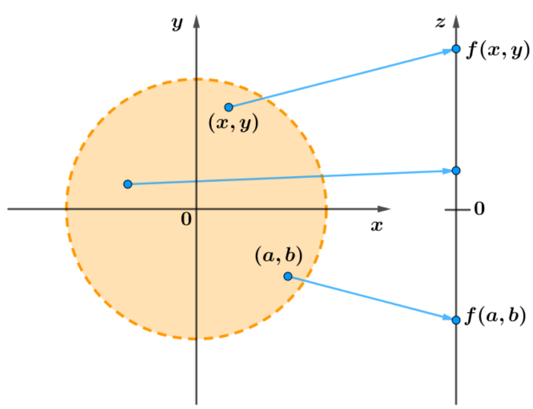


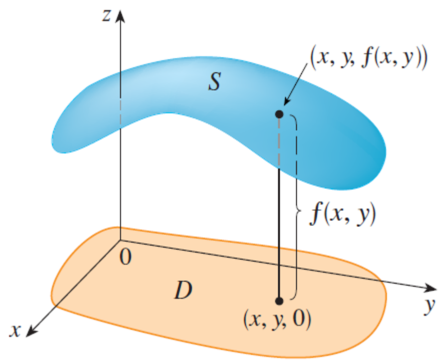
Figura 4.1: Representação de uma função de duas variáveis pelo diagrama de flechas.

4.2 Gráfico de funções

O gráfico é uma forma de representarmos a imagem de funções. É definido como segue.

Definição 4.2.1. *O gráfico da função $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ é o conjunto G de todos os pontos $(x, y, z) \in \mathbb{R}^3$ tal que $z = f(x, y)$ com $(x, y) \in D$, isto é, $G = \{(x, y, z) \in \mathbb{R} : z = f(x, y)\}$.*

A representação gráfica das funções de duas variáveis reais é o esboço de uma superfície, que nos permite entender como a função se comporta conforme variamos os valores das variáveis independentes. A Figura 4.2 apresenta um exemplo.



s

Figura 4.2: Representação gráfica da superfície S.(Fonte:STEWART, James. Cálculo volume 2. 2013, 2013.)

Exemplo 4.2.1. *Seja $f : \mathbb{R}^2 \to \mathbb{R}$, definida como $f(x, y) = x^2 + y^2 + 3$. A representação gráfica dessa função é a superfície apresentada na Figura 4.3.*

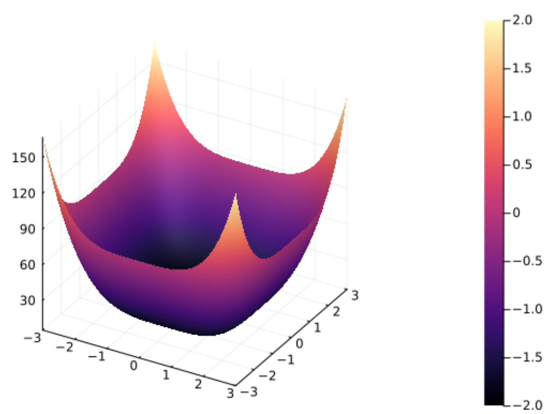


Figura 4.3: Superfície de f.

Exemplo 4.2.2. *Seja $g : \mathbb{R}^2 \to \mathbb{R}$, definida como $g(x, y) = sin(x) + cos(x)$.A representação gráfica dessa função é a superfície apresentada na Figura 4.4.*

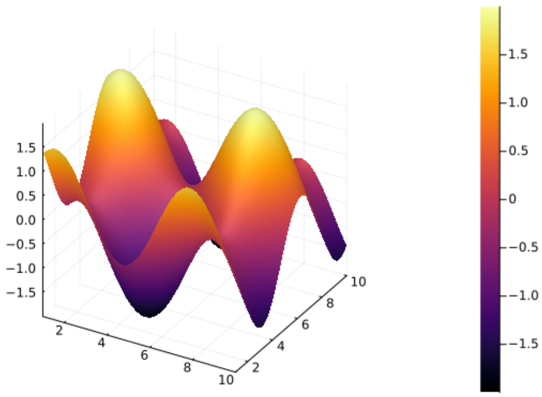
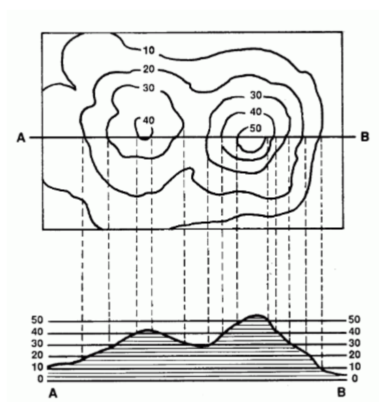


Figura 4.4: Superfície de g.

4.3 Curvas de nível

O gráfico de uma função de duas variáveis reais é representado por uma superfície contida no $\mathbb{R}^3$. Se quisermos representar o comportamento da função em duas dimensões, podemos traçar um conjunto de curvas chamado de curvas de nível. Uma área onde as curvas de nível são muito utilizadas é na Cartografia, como por exemplo no mapeamento do relevo de uma área, veja a Figura 4.5. A definição é dada como segue.



s

Figura 4.5: Representação das curvas de nível na Cartografia. (Fonte:https://eandreoli.blogspot.com/2011/10/curva-de-nivel.html.)

As curvas de nível de uma função $f(x, y)$ de duas variáveis reais são conjuntos de pontos no plano xy que satisfazem a equação $f(x, y) = k$, onde k é uma constante pertencente ao domínio de f.

Exemplo 4.3.1. *Seja* $f : \mathbb{R}^2 \to \mathbb{R}$ *definida por* $f(x, y) = x^2 + y^2$*, a curva de nível para* $k = 4$ *pode ser representada da seguinte maneira*

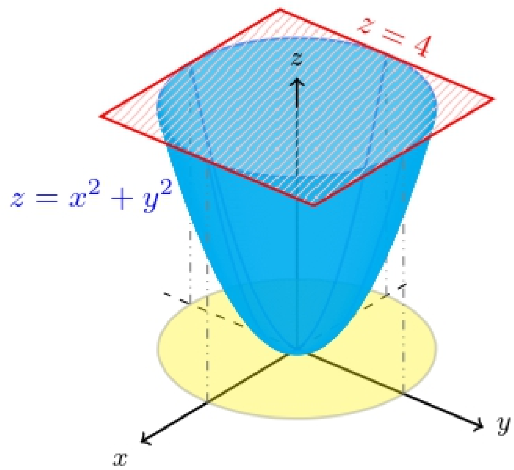


Figura 4.6: Curva de nível de f.

Exemplo 4.3.2. *Retomando o Exemplo 4.2.1, as curvas de nível da função f são dadas por:*

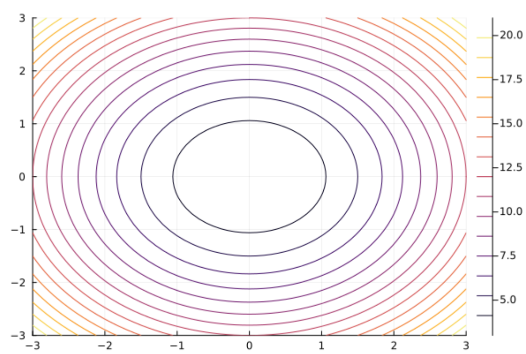


Figura 4.7: Curvas de nível de f.

Exemplo 4.3.3. *Retomando o Exemplo 4.2.2, as curvas de nível da função g são dadas por:*

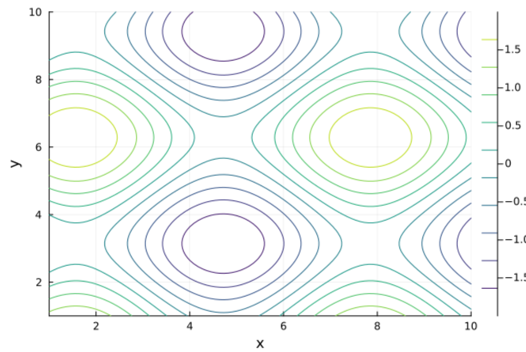


Figura 4.8: Curvas de nível de g.

4.4 Derivadas Parciais

Nesta seção, iremos estender os conceitos de derivadas abordados no Capítulo 3 para o caso de funções de duas váriaveis reais. Veremos que, neste caso, basta derivar cada uma das variáveis, considerando a outra variável como constante. A seguir apresentamos a definição.

Definição 4.4.1. *Se $f : \mathbb{R}^2 \to \mathbb{R}$ é uma função de duas variáveis, suas derivadas parcias em (a, b) são as funções $\frac{\partial f}{\partial x}$ e $\frac{\partial f}{\partial y}$ definidas por*

$$\frac{\partial f}{\partial x} = \lim_{h\to 0} \frac{f(a+h, b) - f(a, b)}{h} \quad e \quad \frac{\partial f}{\partial y} = \lim_{h\to 0} \frac{f(a, b+h) - f(a, b)}{h}.$$

As derivadas parciais das funções elementares são calculadas usando as regras de derivação do cálculo de uma variável.

No cálculo da derivada f_x, olhamos y temporariamente como constante e derivamos a função f como se ela dependesse apenas da variável x.

Exemplo 4.4.1. *Vamos calcular as derivadas parciais da função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = 3x^2 + 5xy - 4y^2$ no ponto $P(1, 3)$. Para isso, calculemos f_x:*

$$f_x = 6x + 5y \Rightarrow f_x(1, 3) = 6 \cdot 1 + 5 \cdot 3 = 6 + 15 = 21.$$

Agora, iremos calcular f_y:

$$f_y = 5x - 8y \Rightarrow f_y(1, 3) = 5 \cdot 1 - 8 \cdot 3 = 5 - 24 = -19.$$

As derivadas parciais $f_x(a, b)$ e $f_y(a, b)$ representam as inclinações das retas tangentes à superfície S em $P(a, b, c)$, com $c = f(a, b)$, com intersecção dos planos $y = b$ e $x = a$, respectivamente.

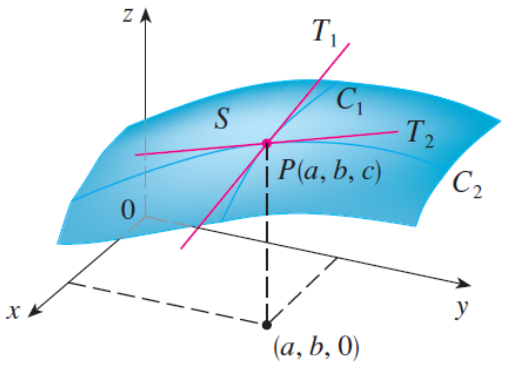


Figura 4.9: Interpretação geométrica das derivadas parciais.(Fonte: STEWART, James. Cálculo, Volume 2. 2013.)

4.5 Derivadas Direcionais e Vetor gradiente

As derivadas direcionais e o vetor gradiente são conceitos fundamentais na análise de funções com várias variáveis. Enquanto as derivadas parciais nos dão informações sobre a taxa de variação em relação a cada variável, as derivadas direcionais expandem essa análise para qualquer direção específica. O vetor gradiente é intensivamente utilizado na área da Otimização, pois pode ser empregado para identificar os pontos críticos de uma função e determinar as direções de maior crescimento e decrescimento. Nesta seção, vamos explorar esses conceitos e, em capítulos posteriores, discutiremos como aplicá-los em problemas da Otimização Contínua.

4.5.1 Derivada Direcional

A derivada direcional nos permite encontrar a taxa de variação de uma função de duas ou mais variáveis reais em qualquer direção. Formalmente, ela é definida como segue.

Definição 4.5.1. *A derivada direcional de f em (x_0, y_0) na direção do vetor $u = (a, b)$ é*

$$D_u f(x_0, y_0) = \lim_{h \to 0} \frac{f(x_0 + ha, y_0 + bh) - f(x_0, y_0)}{h}$$

se esse limite existir.

Para calcularmos as derivadas direcionais utilizamos o seguinte teorema:

Teorema 4.5.1. *Se $f : \mathbb{R}^2 \to \mathbb{R}$ é uma função diferenciável, então f tem derivada direcional na direção de qualquer vetor $u = (a, b)$ e*

$$D_u f(x, y) = f_x(x, y)a + f_y(x, y)b.$$

Exemplo 4.5.1. *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^3 y^2$. Iremos calcular a derivada direcional de f no ponto $P(-1, 2)$ na direção do vetor $u = (4, -3)$.*
Queremos determinar $D_u f(-1, 2)$ na direção de a. Como $f_x(x, y) = 3x^2 y^2$ e $f_y(x, y) = 2x^3 y$. Logo, em $P(-1, 2)$, temos

$$D_u f(-1, 2) = 3 \cdot (-1)^2 \cdot 2^2 \cdot 4 + 2 \cdot (-1)^3 \cdot 2 \cdot (-3) = -36.$$

4.5.2 Vetor Gradiente

O vetor gradiente indica a direção de maior crescimento da função e na otimização essa informação é importantíssima, visto que para minimização a direção de maior decréscimento será o $-\nabla f$.

Definição 4.5.2. *Se $f : \mathbb{R}^2 \to \mathbb{R}$ é uma função de duas variáveis reais, então o gradiente de f é a função vetorial, denotada por ∇f, definida por*

$$\nabla f(x, y) = (f_x(x, y), f_y(x, y)).$$

Exemplo 4.5.2. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ definida por $f(x, y) = x^2 + xy + y^2$ e tomamos o vetor gradiente $\Delta f(5, 2) = (12, 9)$.*

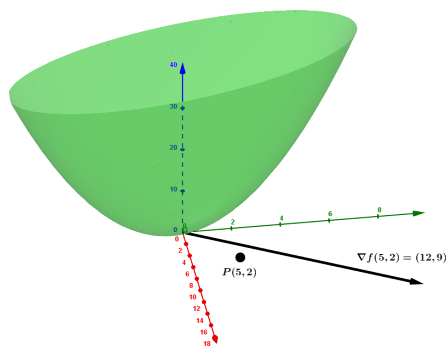


Figura 4.10: Vetor Gradiente da função f no ponto $P(5, 2)$.

4.6 Máximos e mínimos de funções de duas variáveis

Na Otimização Contínua, o foco dos estudos é encontrar máximos ou mínimos de funções. Para funções de mais de uma variável, faz-se necessário o uso de estrátegias do Cálculo Diferencial para encontrar estes valores.

Definição 4.6.1. *Uma função de duas variáveis reais tem o ponto de máximo local em (a, b) se $f(x, y) \leq f(a, b)$ quando (x, y) está próximo de (a, b). O valor $f(a, b)$ é chamado de valor de máximo local. Se $f(x, y) \geq f(a, b)$ quando (x, y) está próximo de (a, b), então f tem um mínimo local em (a, b) e $f(a, b)$ é um valor de mínimo local.*

Exemplo 4.6.1. *Note que a função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = sin(x)sin(y)$ possui pontos de máximo e mínimo.*

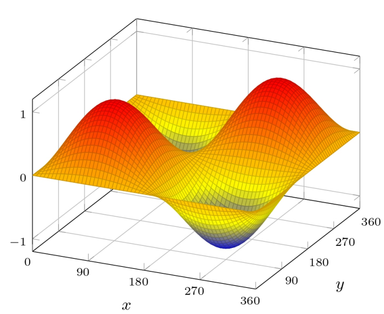


Figura 4.11: Esboço de f.

As derivadas parciais e os pontos de mínimo ou máximo estão estritamente ligados, é o que podemos observar no teorema abaixo.

Teorema 4.6.1. *Se $f : \mathbb{R}^2 \to \mathbb{R}$ tem um máximo ou mínimo local em (a, b) e as derivadas parciais de primeira ordem de f existem nesses pontos, então $f_x(a, b) = 0$ e $f_y(a, b) = 0$.*

Além disso, (a, b) é chamado ponto crítico ou ponto estacionário de f se $f_x(a, b) = 0$ e $f_y(a, b) = 0$, ou quando uma das derivadas parciais não existir. Para que um ponto (a, b) seja de máximo ou de mínimo de f, ele deve ser um ponto crítico, mas a recíproca não é verdadeira.

Exemplo 4.6.2. *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^2 + y^2 - 2x - 6y + 14$. O ponto crítico de f é dado a partir do cálculo das derivadas parciais:*

$$f_x(x, y) = 2x - 2 \text{ e } f_y(x, y) = 2y - 6. \tag{4.1}$$

Logo, $(1, 3)$ é o único ponto crítico.

4.7 Exercícios

Exercício 4.7.1. *Seja $F : D \subset \mathbb{R}^2 \to \mathbb{R}$ definida por $F(x, y) = 1 + \sqrt{4 - y^2}$.*

(a) Calcule $F(3, 1)$.
(b) Determine e esboce o domínio D de F.
(c) Determine a imagem de F.

Exercício 4.7.2. *Seja $g : D \subset \mathbb{R}^2 \to \mathbb{R}$ definida por $g(x, y) = cos(x + 2y)$.*

(a) Calcule g(2,-1).
(b) Determine e esboce o domínio D de g.
(c) Determine a imagem de g.

Exercício 4.7.3. *Determine as derivadas parciais de primeira ordem das seguintes funções $f : D \subset \mathbb{R}^2 \to \mathbb{R}$.*

(a) $f(x, y) = y^5 - 3xy$
(b) $f(x, t) = \sqrt{x} \ln t$
(c) $f(x, y) = \frac{x}{y}$
(d) $f(u, v) = (u^2 v - v^3)^5$

Exercício 4.7.4. *Determine a derivada direcional das funções $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ no ponto dado na direção do vetor $\boldsymbol{v}$.*

(a) $f(x, y) = e^x sin(y)$, $(0, \pi/3)$, $\boldsymbol{v}=(-6, 8)$
(b) $f(x, y) = \frac{x}{x^2+y^2}$, $(1, 2)$, $\boldsymbol{v}=(3, 5)$

Exercício 4.7.5. *Determine a derivada direcional de $f : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = \sqrt{xy}$ em $P(2, 8)$ na direção de $Q(5, 4)$.*

Exercício 4.7.6. *Determine os valores máximos e mínimos locais das seguintes funções $f : \mathbb{R}^2 \to \mathbb{R}$, definidas por:*

(a) $f(x, y) = 9 - 2x + 4y - x^2 - 4y^2$
(b) $f(x, y) = x^3 y + 12x^2 - 8y$
(c) $f(x, y) = (x^2 + y^2)e^{y^2 - x^2}$

Capítulo 5
Sequências Numéricas

Na abordagem dos problemas de Otimização Contínua, é frequente a necessidade de encontrar máximos ou mínimos de funções em um conjunto. Para solucionar esses problemas, recorremos a métodos iterativos que geram sequências numéricas. Por exemplo, para determinar o valor mínimo de uma função em um intervalo dado, escolhemos um ponto inicial nesse intervalo e, a partir dele, geramos novos pontos que avançam em direção ao mínimo da função. Esse processo iterativo resulta em uma sequência. Assim, é evidente que as sequências desempenham um papel fundamental na resolução de problemas de Otimização Contínua, fornecendo uma estrutura para analisar e compreender o processo de convergência em direção à solução ótima.

Neste capítulo, iremos explorar os conceitos fundamentais das sequências, incluindo subsequências, monotonicidade e limitação. Apresentaremos também a noção intuitiva de limite de uma sequência, sua definição formal e exemplos de sequências convergentes e divergentes.

5.1 Sequências Numéricas

Sequências numéricas são utilizadas na modelagem e resolução de problemas de diversas áreas. Considere, por exemplo, uma partícula em movimento em que registramos sua posição em intervalos regulares de tempo. Essas posições, ordenadas conforme o tempo avança, formam uma sequência de números que descreve o movimento da partícula ao longo do tempo. Na Figura abaixo, podemos perceber uma sequência ordenada de diferentes alturas para a bola ao longo de intervalos de tempo.

Figura 5.1: Sequência de fotografia em ação que registra o movimento da bola em intervalos.
(Fonte: Mathigon.)

A seguir, apresentamos formalmente a definição de sequências numéricas.

Definição 5.1.1. *Uma sequência de números reais é uma função $x : \mathbb{N} \to \mathbb{R}$, isto é, uma função que associa a cada número natural n um número real x_n que é chamado de n-ésimo termo da sequência. Podemos denotar uma sequência das seguintes maneiras:*

$$(x_1, ..., x_n, \dots) \text{ ou } (x_n)_{n\in\mathbb{N}} \text{ ou } (x_n).$$

As sequências são frequentemente definidas fornecendo-se uma fórmula para o n-ésimo termo x_n. Também pode-se listar os termos de uma sequência em ordem, parando quando a regra de formação parecer evidente. Outra maneira de definir uma sequência é especificar o valor de x_1 e criar uma fórmula para obter x_{n+1} $(n \geq 1)$ em termos de x_1. As sequências definidas desta maneira são chamadas de recursivas.

Exemplo 5.1.1. *Seja $b \in \mathbb{R}$, a sequência $(x_n) = (b, b, b, \ldots)$, cujos termos são iguais a b, é chamada de sequência constante. Na Figura 5.4, podemos visualizar graficamente uma sequência constante em que $x_n = 3$, para todo $n \in \mathbb{N}$.*

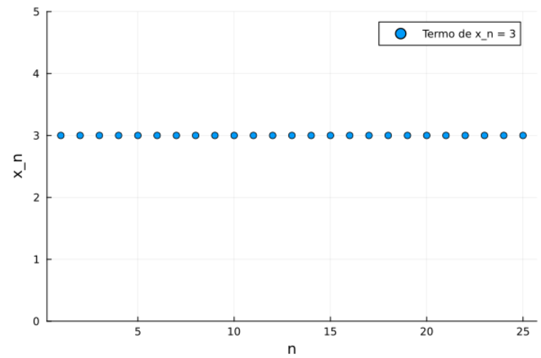


Figura 5.2: Sequência constante.

Exemplo 5.1.2. *A sequência de Fibonacci $(f_n) = (1, 1, 2, 3, 5, 8, 13, \ldots)$ é dada pela seguinte definição recursiva:*

$$f_1 = 1, \; f_2 = 1, \; f_{n+1} = f_{n-1} + f_n \text{ para } n \geq 2.$$

Assim, após o segundo termo, cada termo é a soma de seus dois antecessores.

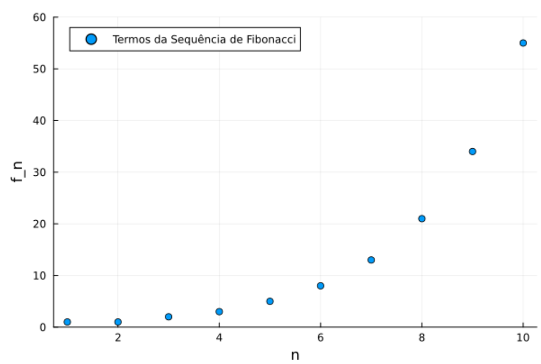


Figura 5.3: Sequência de Fibonacci.

Definição 5.1.2. *Uma sequência de números reais (x_n) diz-se crescente quando $x_1 < x_2 < x_3 < \ldots$, ou seja, quando, para todo $n \in \mathbb{N}$, temos $x_n < x_{n+1}$. No caso da desigualdade não ser estrita, isto é, quando para todo $n \in \mathbb{N}$ temos $x_n \leq x_{n+1}$, a sequência diz-se não-descescente. Analogalmente, se para todo $n \in \mathbb{N}$ temos $x_n > x_{n+1}$ a sequência é descrescente, e no caso a desigualdade não ser estrita, ou seja, $x_n \geq x_{n+1}$, diz-se não-crescente. Em todos os casos, dizemos que a sequência é monótona.*

Exemplo 5.1.3. *A sequência $x_n = n$, para todo $n \in \mathbb{N}$, é uma sequência estritamente crescente.*

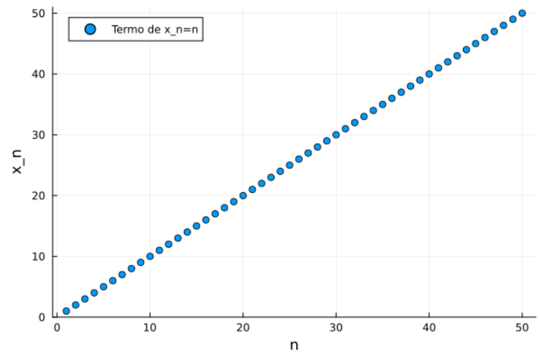


Figura 5.4: Sequência crescente.

Exemplo 5.1.4. *A sequência de Fibonacci que vimos no Exemplo 5.3 é uma sequência não-decrescente.*

Definição 5.1.3. *Uma subsequência de* $(x_n)_{n\in\mathbb{N}}$ *é a restrição desta sequência a um subconjunto infinito* $\mathbb{N}' = \{n_1 < n_2 < \ldots < n_m < \ldots\} \subset \mathbb{N}$*, e é indicada pela notação* $(x_n)_{n\in\mathbb{N}'}$.

Exemplo 5.1.5. *Considere a sequência dada por* $x_n = (-1)^n$*. Note que essa sequência possui duas subsequências:*

- *Tomando* $\mathbb{N}_p = \{n \in \mathbb{N} : n = 2k$ *para algum* $k \in \mathbb{N}\}$*, temos a subsequência dos termos com índices pares* $(x_n)_{n\in\mathbb{N}_p} = (1, 1, 1, \ldots)$.
- *Tomando* $\mathbb{N}_i = \{n \in \mathbb{N} : n = 2k + 1$ *para algum* $k \in \mathbb{N}\}$*, temos a subsequência dos termos com índices ímpares* $(x_n)_{n\in\mathbb{N}_i} = (-1, -1, -1, \ldots)$.

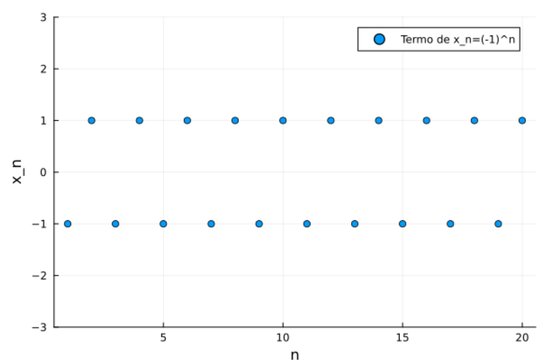


Figura 5.5: Exemplos de subsequências.

Definição 5.1.4. *Uma sequência* (x_n) *de números reais é limitada superiormente se houver um número* $M \in \mathbb{R}$ *tal que* $x_n \leq M$ *para todo* $n \in \mathbb{N}$*. Da mesma forma, a sequência é dita limitada inferiormente se existir um número* $P \in \mathbb{R}$ *tal que* $x_n \geq P$ *para todo* $n \in \mathbb{N}$.

Definição 5.1.5. *Uma sequência* (x_n) *de números reais é dita limitada se existe um número real* $M > 0$ *tal que* $|x_n| \leq M$ *para todo* $n \in \mathbb{N}$.

Exemplo 5.1.6. *Considere a sequência dada por* $x_n = 2^n$*. Temos que essa sequência é limitada inferiormente por* 2 *e não é limitada superiormente.*

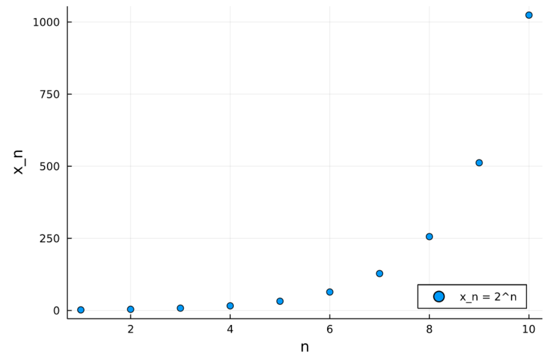


Figura 5.6: Sequência limitada apenas inferiormente.

Exemplo 5.1.7. *A sequência dada por* $x_n = 2 + (1/n)$ *é limitada inferiormente por* 2 *e superiormente por* 3*.*

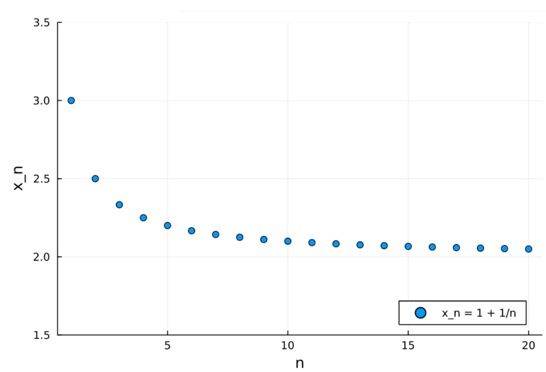


Figura 5.7: Sequência limitada.

Definição 5.1.6. *Seja* $X \subset \mathbb{R}$*. Dizemos que* X *é limitado superiormente quando existe algum* $s \in \mathbb{R}$ *tal que, para cada* $x \in X$ *ocorre* $x \leq s$*. Neste caso, diz-se que* s *é uma cota superior de* X*. Da mesma forma, diz-se que o conjunto* X *é limitado inferiormente quando existe algum* $p \in \mathbb{R}$ *tal que, para cada* $x \in X$ *ocorre* $x \geq p$*. O número* p *chama-se então uma cota inferior de* X*. Se* X *é limitado superior e inferiormente, dizemos que* X *é um conjunto limitado.*

Exemplo 5.1.8. *Seja* $X = \{1, 1/2, 1/3, 1/4, \ldots\}$*. Note que este conjunto é limitado. Temos, por exemplo, que* 1 *é uma cota superior e* 0 *é uma cota inferior de* X*.*

Definição 5.1.7. *Seja* $X \subset \mathbb{R}$ *limitado superiormente e não-vazio. Um número* $b \in \mathbb{R}$ *chama-se supremo do conjunto* X *quando* b *é a menor das cotas superiores de* X*. Escreveremos* $\sup X$ *para indicá-lo. As condições que caracterizam o supremo podem, portanto, ser escritas assim:*

$S1.$ $x \in X \implies x \leq \sup X$*;*
$S2.$ $b \geq x$ *para todo* $x \in X \implies b \geq \sup X$*;*
$S2'.$*Se* $b < \sup X$ *então existe* $x \in X$ *tal que* $b < x$*.*

Portanto, o supremo de um conjunto, quando existe, é único.

Definição 5.1.8. *Seja* $X \subset \mathbb{R}$ *limitado inferiormente e não-vazio. Um número* $a \in \mathbb{R}$ *chama-se ínfimo do conjunto* X*, quando* a *é a maior das cotas inferiores de* X*. Para que* a *seja o ínfimo de* X*, é necessário e suficiente que as condições abaixo sejam satisfeitas:*

*I*1. *Para todo* $x \in X$ *tem-se* $a \leq x$;
*I*2. *Se* $c \in \mathbb{R}$ *é tal que* $c \leq x$ *para todo* $x \in X$, *então* $c \leq a$.

O ínfimo de X, *quando existe, é único e escreve-se* $a = \inf X$.

Exemplo 5.1.9. *Vamos determinar o supremo e o ínfimo do conjunto* $A = \{1/n : n \in \mathbb{N}\}$.

- *Primeiro, vamos determinar o ínfimo de* A. *Note que todos os elementos de* A *são positivos e, portanto,* 0 *é uma cota inferior. Por definição, sabemos que o ínfimo é a maior das cotas inferiores do conjunto. Assim, temos* $\inf A \geq 0$. *Suponha, por absurdo, que* $\inf A > 0$. *Então, existiria* $c > 0$ *tal que* $\inf A > c > 0$. *Porém, pela propriedade arquimediana, existe* $n_0 > 0$ *tal que* $n_0 c > 1$ *que implica* $c > 1/n_0 \in A$. *Assim, teríamos um elemento de* A *que é menor que o ínfimo de* A, *o que é um absurdo. Como esse aburso surgiu quando supomos que* $\inf A > 0$, *segue-se que devemos ter* $\inf A = 0$.

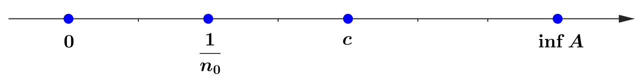


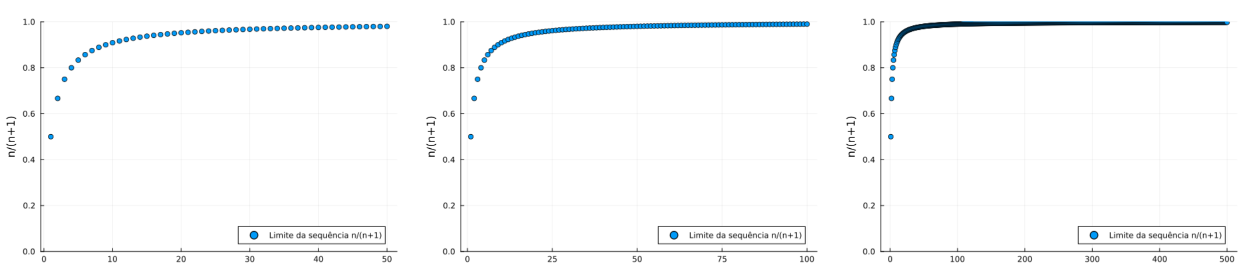
Figura 5.8: Construção geométrica do que ocorre se tivermos $\inf A > 0$.

- *Agora, vamos determinar o supremo de* A. *Note que* $1/n \leq 1$, *para todo* $n \in \mathbb{N}$, *o que significa que* 1 *é uma cota superior para o conjunto* A. *Em particular, devemos ter* $\sup A \leq 1$, *pois por definição o supremo é a menor das cotas superiores. Como* $1 \in A$, *temos que* $\sup A \geq 1$. *Logo,* $\sup A = 1$.

5.2 Noção intuitiva do limite de sequências

Intuitivamente, uma sequência $(x_n)_{n \in \mathbb{N}}$ possui limite L se, à medida que o índice n cresce, o elemento x_n vai se tornando arbitrariamente próximo de L.

Exemplo 5.2.1. *Seja* $(x_n)_{n \in \mathbb{N}} = \left(\frac{n}{n+1}\right) = \left(\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots, \frac{n}{n+1}, \ldots\right)$. *Vamos analisar a representação gráfica da sequência com a variação de* n:



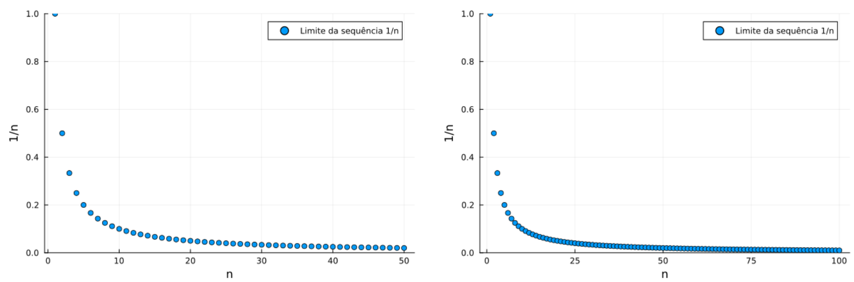
(a) n variando de 0 a 50. (b) n variando de 0 a 100. (c) n variando de 0 a 500.

Figura 5.9: Gráfico da sequência $(x_n)_{n \in \mathbb{N}} = (\frac{n}{n+1})$ com a variação de n.

Observações 5.2.1. *A sequência* $(x_n)_{n \in \mathbb{N}} = \left(\frac{n}{n+1}\right)$ *é crescente e limitada (limitada inferiormente por* $1/2$ *e superiormente por* 1*); Dizer que* 1 *é o limite de tal sequñcia, significa que para valores muito grandes de* n, *os termos de* x_n

ficam tão próximos de 1 *quanto quisermos; Nenhum termo da sequência será igual a* 1*, pois não existe* $n \in \mathbb{N}$ *tal que* $\frac{n}{n+1} = 1$.

Exemplo 5.2.2. *Seja* $(x_n)_{n\in\mathbb{N}} = \left(\frac{1}{n}\right) = \left(1, \frac{1}{2}, \frac{1}{3}, ..., \frac{1}{n}, ...\right)$. *Vamos analisar o gráfico da sequência com a variação de n:*



(a) n variando de 0 a 50. (b) n variando de 0 a 100.

Figura 5.10: Representação gráfica da sequência $(x_n)_{n\in\mathbb{N}} = (\frac{1}{n})$ com a variação de n.

Observações 5.2.2. *A sequência* $(x_n)_{n\in\mathbb{N}} = \left(\frac{1}{n}\right)$ *é decrescente e limitada (limitada inferiormente por* 0 *e superiormente por* 1*); O limite dessa sequência é* 0*, mas não existe* $n \in \mathbb{N}$ *tal que* $\frac{1}{n} = 0$*. Logo,* 0 *não é um termo de* $(x_n)_{n\in\mathbb{N}}$.

Exemplo 5.2.3. *Seja a sequência constante* $(x_n)_{n\in\mathbb{N}} = (1, 1, 1, ...)$. *Vamos analisar o gráfico da sequência com a variação de n:*

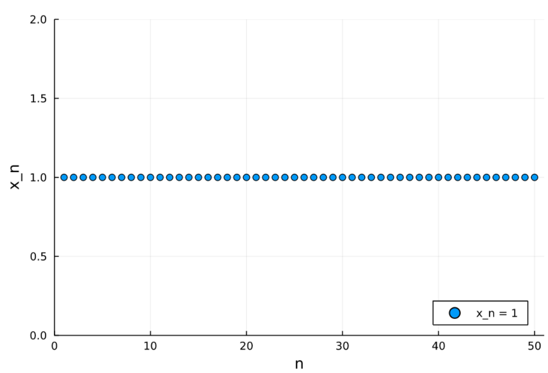


Figura 5.11: Limite de sequência constante

Observação 5.2.3. *Neste caso, o limite é um termo da sequência e é igual a* 1.

Portanto, dizer que um número é o limite de uma sequência nos leva a pensar que quanto maior a posição do termo, mais ele se aproxima desse número, arbitrariamente.

5.3 Limite de uma Sequência

Trazendo mais precisão à noção de limite apresentada na seção anterior, se for estipulado um “erro” por meio de um número real $\varepsilon > 0$, então existe um índice n_0 tal que todos os termos x_n da sequência que têm índice n maior do que

n_0 são valores aproximados de a com erro inferior a ε. O índice n_0 depende de ε, sendo de se esperar que, para valores cada vez menores de ε, necessite-se tomar n_0 cada vez maior. Isto nos leva à seguinte definição.

Definição 5.3.1. *Diz-se que o número real a é limite da sequência (x_n) de números reais, e escreve-se $a = \lim x_n$, quando para cada número real $\varepsilon > 0$, dado arbitrariamente, for possível obter um $n_0 \in \mathbb{N}$ tal que $|x_n - a| < \varepsilon$, sempre que $n > n_0$.*

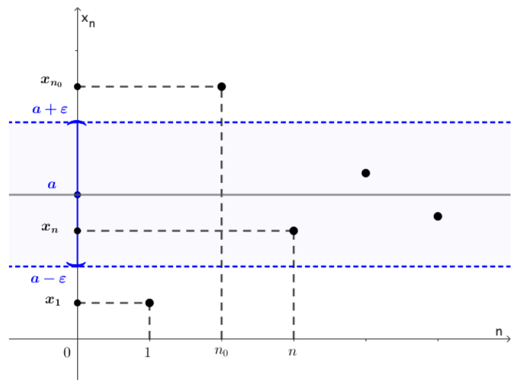


Figura 5.12: Ilustração da definição de limite de uma sequência.

Observe a relação com a definição de limites de funções apresentada no Capítulo 3.

Observação 5.3.2. *Se $\lim x_n = a$ então qualquer intervalo $(a - \varepsilon, a + \varepsilon)$, de centro a e raio $\varepsilon > 0$, contém todos os termos x_n da sequência, com exceção no máximo de um número finito de índices n, que são justamente os termos $x_1, x_2, \ldots, x_{n_0}$. Reciprocamente, se qualquer intervalo de centro a contém todos os x_n, salvo talvez para um número finito de índices n, então $\lim x_n = a$.*

Quando $\lim x_n = a$, dizemos que a sequência (x_n) *converge* para a. Uma sequência que possui limite chama-se *convergente*. Do contrário, ela chama-se *divergente*.

Ao analisarmos o limite de uma sequência, usualmente emprega-se desigualdades que são de grande valia para limitar superiormente o termo geral de uma sequência por ε, por exemplo. Como vimos anteriormente, cada termo da sequência $(x_n)_{n\in\mathbb{N}}$ em nosso contexto é um número real. Dentre as propriedades referentes aos números reais, vale destacar que $\mathbb{R}$ é um corpo arquimediano, isto é, as seguintes condições equivalentes são satisfeitas:

(i) $\mathbb{N} \subset \mathbb{R}$ é ilimitado superiormente (isto é, para todo número real x, existe um número natural n tal que $x < n$);
(ii) Dados $a, b \in \mathbb{R}$, com $a > 0$, existe $n \in \mathbb{N}$ tal que $n \cdot a > b$;
(iii) Dado qualquer $a > 0$ em $\mathbb{R}$, existe $n \in \mathbb{N}$ tal que $0 < \dfrac{1}{n} < a$.

No exemplo a seguir, já começaremos a utilizar a propriedade arquimediana em $\mathbb{R}$.

Exemplo 5.3.1. *Retomando o Exemplo 5.2.2, em que $(x_n)_{n\in\mathbb{N}} = \dfrac{1}{n}$, podemos mostrar que $\lim x_n = 0$. De fato, dado $\varepsilon > 0$ arbitrário, pela propriedade arquimediana existe $n_0 \in \mathbb{N}$ tal que $n_0 > \dfrac{1}{\varepsilon}$. Então $n > n_0 \Longrightarrow \dfrac{1}{n} < \dfrac{1}{n_0} < \varepsilon$, ou seja, $n > n_0 \Longrightarrow \left|\dfrac{1}{n} - 0\right| < \varepsilon$.*

Exemplo 5.3.2. *Seja a sequência $(x_n)_{n\in\mathbb{N}} = n^3$. Na figura abaixo, temos a representação gráfica dessa sequência:*

Observação 5.3.3. *Neste caso, a sequência é divergente, ou seja, dado qualquer número a real, existe um $\varepsilon > 0$ tal que, seja qual for $n_0 \in \mathbb{N}$, teremos $|x_n - a| \geq \varepsilon$ sempre que $n > n_0$.*

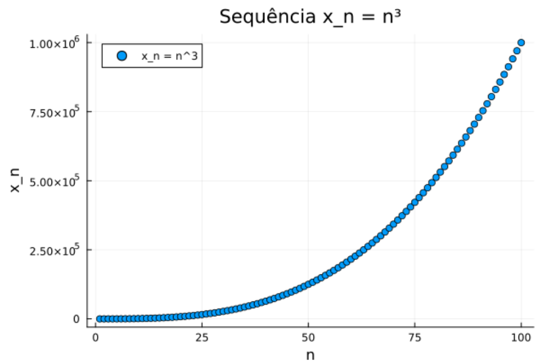


Figura 5.13: Representação gráfica da sequência $x_n = n^3$.

Definição 5.3.4. *Um corpo ordenado K chama-se completo quando todo subconjunto não-vazio, limitado superiormente, $X \subset K$, possui supremo em K.*

Resulta da definição que, num corpo ordenado completo, todo conjunto não-vazio, limitado inferiormente, $Y \subset K$, possui um ínfimo. Com efeito, dado Y, seja $X = -Y$, isto é, $X = \{y; y \in Y\}$. Então X é não-vazio e limitado superiormente; logo existe $a = supX$. Como se vê facilmente, tem-se $-a = infY$.

Axioma da Completude. *Todo conjunto não vazio e limitado superiormente de números reais possui um supremo que também é um número real.*

Demonstraremos, agora, alguns resultados simples sobre limites, a fim de podermos analisar de modo inteligente os exemplos futuros.

Teorema 5.3.1. ***(Unicidade do limite)*** *Se $\lim x_n = a$ e $\lim x_n = b$ então $a = b$.*

Demonstração: Seja $\lim x_n = a$. Dado qualquer número real $b \neq a$, mostraremos que não pode ocorrer $\lim x_n = b$. Para isso, tomemos $\varepsilon = \dfrac{|b-a|}{2}$. Vemos que $\varepsilon > 0$ e notamos ainda que os intervalos $(a-\varepsilon, a+\varepsilon)$ e $(b-\varepsilon, b+\varepsilon)$ são disjuntos. (Se existisse $x \in (a-\varepsilon, a+\varepsilon) \cap (b-\varepsilon, b+\varepsilon)$ teríamos $|a-x| < \varepsilon$ e $|x-b| < \varepsilon$, donde $|a-b| \leq |a-x|+|x-b| < 2\varepsilon = |a-b|$, um absurdo.) Como, $\lim x_n = a$, existe $n_0 \in \mathbb{N}$ tal que $n > n_0 \Longrightarrow x_n \in (a-\varepsilon, a+\varepsilon)$ e, portanto, $x_n \notin (b-\varepsilon, b+\varepsilon)$ para todo $n > n_0$. Logo, não ocorre $\lim x_n = b$. $\square$

Teorema 5.3.2. *Se $\lim x_n = a$ então toda subsequência de (x_n) converge para o limite a.*

Demonstração: Seja $(x_{n_1}, x_{n_2}, \ldots, x_{n_i}, \ldots)$ uma subsequência de (x_n). Dado $\varepsilon > 0$, existe $n_0 \in \mathbb{N}$ tal que $n > n_0 \Longrightarrow |x_n - a| < \varepsilon$. Como os índices da subsequência formam um conjunto infinito, existe entre eles um $n_{i_0} > n_0$. Então $n_i > n_{i_0} \Longrightarrow n_i > n_0 \Longrightarrow |x_{n_i} - a| < \varepsilon$. Logo, $\lim x_{n_i} = a$. $\square$

Teorema 5.3.3. *Toda sequência convergente é limitada.*

Demonstração: Seja $\lim x_n = a$. Então, tomando $\varepsilon = 1$, vemos que existe $n_0 \in \mathbb{N}$ tal que $n > n_0 \Longrightarrow x_n \in (a-1, a+1)$. Consideremos o conjunto finito definido por $F = \{x_1, x_2, \ldots, x_{n_0}, a-1, a+1\}$. Sejam c o menor e d o maior elemento de F. Então todos os termos x_n da sequência estão contidos no intervalo $[c, d]$; logo a sequência é limitada. $\square$

Observação 5.3.5. *A proposição que acabamos de demonstrar nos diz que basta verificar que uma sequência não é limitada para concluir que ela não converge. A proposição a seguir nos fornece um "critério de convergência", ou seja, nos permite concluir que uma sequência (x_n) converge, mesmo sem conhecermos, a princípio, seu limite.*

Teorema 5.3.4. *Toda sequência monótona limitada é convergente.*

Demonstração: Sem perda de generalidade, seja $(x_1 \leq x_2 \leq \ldots \leq x_n \leq \ldots)$ uma sequência não decrescente limitada. Pelo Axioma da Completude, podemos tomar $a = \sup\{x_n; n = 1, 2, \ldots\}$. Afirmamos que $\lim x_n = a$. Com efeito, dado qualquer $\varepsilon > 0$, como $a - \varepsilon < a$, o número $a - \varepsilon$ não é cota superior do conjunto dos x_n. Logo, existe um $n_0 \in \mathbb{N}$ tal que $a - \varepsilon < x_n$ para $n > n_0$. Como a sequência é monótona, $n > n_0 \Longrightarrow x_{n_0} \leq x_n$ e, portanto, $a - \varepsilon < x_n$. Como $x_n \leq a$ para todo n, vemos que $n > n_0 \Longrightarrow a - \varepsilon < x_n < a + \varepsilon$. Assim, temos de fato $\lim x_n = a$, como queríamos demonstrar. $\square$

Exemplo 5.3.3. *A sequência* $(1, 0, 2, 0, 3, 0, \ldots)$ *não é convergente porque é ilimitada superiormente. Nota-se que ela possui uma subsequência convergente, que é a subsequência constante dos índices pares onde os termos são todos iguais a* 0.

Exemplo 5.3.4. *Seja* $0 < a < 1$. *Então a sequência* $x_n = a^n$ *é monótona decrescente limitada; logo, converge. O gráfico a seguir mostra os* 50 *primeiros termos dessa sequência quando* $a = \dfrac{1}{2}$.

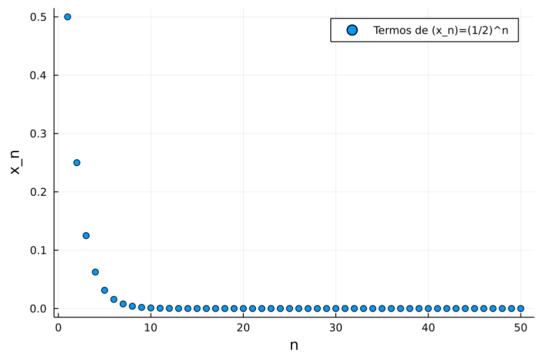


Figura 5.14: Representação gráfica da sequência $x_n = \left(\dfrac{1}{2}\right)^n$.

Exemplo 5.3.5. *Considere a sequência* $x_n = \dfrac{(-1)^{n+1}}{n}$. *Todos os elementos desta sequência são diferentes de zero, sendo positivos os termos de ordem ímpar e negativos os termos de ordem par. Intuitivamente, percebe-se que os termos da sequência se aproximam cada vez mais de zero, positiva e negativamente, o que é reforçado ao vermos o gráfico com os* 200 *primeiros termos da sequência:*

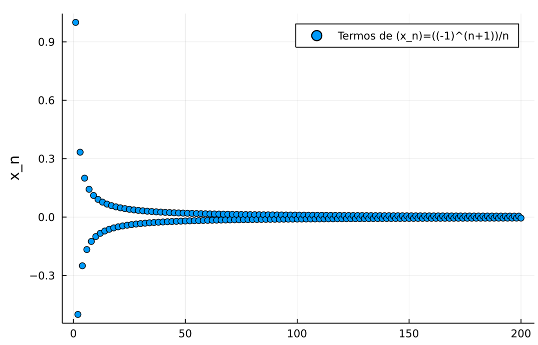


Figura 5.15: Representação gráfica da sequência $(x_n)_{n\in\mathbb{N}} = \dfrac{(-1)^{n+1}}{n}$.

Vamos mostrar que, de fato, $\lim x_n = 0$. *Com efeito, seja* ε *um número real positivo qualquer, pela propriedade arquimediana existe um natural* n_0 *tal que* $\dfrac{1}{n_0} < \varepsilon$. *Então* $\dfrac{(-1)^{n_0+1}}{n_0} \in (-\varepsilon, \varepsilon)$, *pois* $\left|\dfrac{(-1)^{n_0+1}}{n_0}\right| = \dfrac{1}{n_0}$. *Além disso, se*

$n > n_0$, então $\left|\dfrac{(-1)^{n+1}}{n}\right| = \dfrac{1}{n} < \dfrac{1}{n_0} < \varepsilon$*. Acabamos de verificar que* $\left|\dfrac{(-1)^{n+1}}{n}\right| < \varepsilon$ *para todo $n > n_0$. Isso implica que* $\lim x_n = 0$.

Observamos que as propriedades aritméticas dos limites de funções, apresentadas no Capítulo 4 deste material, são válidas também para limites de sequências de números reais, o que nos permite calcular o limite da sequência do exemplo a seguir.

Exemplo 5.3.6. *Considere a sequência $x_n = \sqrt[n]{n}$ que é decrescente a partir do seu terceiro termo. Segue-se que existe* $\lim \sqrt[n]{n} = a$*. O gráfico abaixo representa os* 300 *primeiros termos desta sequência.*

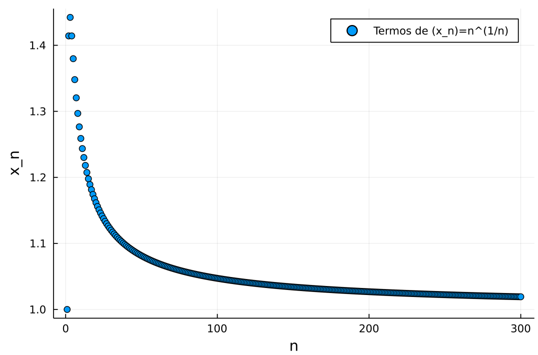


Figura 5.16: Representação gráfica da sequência $(x_n)_{n\in\mathbb{N}} = x_n = \sqrt[n]{n}$.

Mostremos agora que $\lim \sqrt[n]{n} = 1$. Escrevendo $l = \lim n^{1/n}$, vemos que $l = \inf\{n^{1/n}; n \in \mathbb{N}\}$. Como $n^{1/n} \geq 1$ para todo $n \in \mathbb{N}$, temos que $l \geq 1$. Em particular, $l > 0$. Considerando a subsequência $(2n)^{1/2n}$, temos

$$l^2 = \lim[(2n)^{1/2n}]^2 = \lim[(2n)^{1/n}] = \lim[2^{1/n} \cdot n^{1/n}] = lim 2^{1/n} \cdot \lim n^{1/n} = l. \tag{5.1}$$

Como $l \neq 0$, de $l^2 = l$ concluímos que $l = 1$.

5.4 Exercícios

Exercício 5.4.1. *Encontre a fórmula fechada para cada uma das sequências a seguir. Assuma que o primeiro termo dado seja a_1.*

(a) $(2, 5, 10, 17, 26, \ldots)$
(b) $(0, 2, 5, 9, 14, 20, \ldots)$
(c) $(1, 5, 23, 119, 719, \ldots)$

Exercício 5.4.2. *Dê um exemplo de:*

(a) Uma sequência estritamente crescente.
(b) Uma sequência não-decrescente.
(c) Uma sequência estritamente decrescente.
(d) Uma sequência não-crescente.

Exercício 5.4.3. *Prove que $(x_n)_{n\in\mathbb{N}} = \left(\frac{1}{n}\right)$ é uma sequência estritamente decrescente.*

Exercício 5.4.4. *Determine o supremo e o ínfimo do conjunto $B = \{\frac{n}{n+1} : n \in \mathbb{N}\}$.*

Exercício 5.4.5. *Dê um exemplo de:*

(a) Uma sequência limitada superiormente que não seja limitada.
(b) Uma sequência limitada inferiormente que não seja limitada.
(c) Uma sequência limitada.
(d) Uma sequência ilimitada.

Exercício 5.4.6. *Considerando os exemplos usados para responder os exercícios 5.4.2 e 5.4.5;*

(a) Encontre um intervalo $X \subset \mathbb{R}$ *tal que* $x_n \in X, \forall n \in \mathbb{N}$ *para cada exemplo;*
(b) Se possível, dê tres cotas superiores e três cotas inferiores distintas para cada intervalo obtido no item a);
(c) Se possível, encontre $\inf X$ *e* $\sup X$ *para cada intervalo obtido no item a).*

Exercício 5.4.7. *Prove que* $\lim x_n = 0$ *no Exemplo 5.3.4.*

Exercício 5.4.8. *Se* $\lim x_n = a$*, prove que* $\lim |x_n| = |a|$.

Exercício 5.4.9. *Considere a sequência* $x_n = \dfrac{n^2 - 3n}{2n+1}$. *Determine* $\lim x_n$.

Exercício 5.4.10. *Considere a sequência* $x_n = \dfrac{3n^2 + 2}{2n^2 + 5}$. *Determine* $\lim x_n$.

Exercício 5.4.11. *Represente graficamente as sequências dos dois exercícios anteriores utilizando a linguagem de programação Julia com* 10*,* 100 *e* 1000 *termos.*

Capítulo 6
Introdução à Otimização

Otimizar é um dos conceitos matemáticos que poderíamos perguntar a qualquer pessoa e ela, com uma alta probabilidade, nos daria uma boa noção do que se trata. Isso porque, segundo o Dicionário Aurélio "otimizar é um verbo que significa criar condições mais favoráveis para; tirar o melhor partido possível de algo". Ao passo que, matematicamente, poderíamos dizer que otimizar é um processo que permite encontrar a melhor maneira de realizar algo, desde que esse 'melhor' tenha uma definição matemática bem determinada.

Neste capítulo, serão abordadas noções fundamentais da teoria de Otimização Contínua, incluindo o problema de otimização, condições de otimalidade, introdução aos métodos e elementos essenciais para algoritmos de otimização. O objetivo é formalizar e aprofundar os conhecimentos previamente adquiridos neste material, visando a compreensão dos conceitos apresentados até o momento.

6.1 Noções Introdutórias

Considerando que os próximos conceitos estão ligadas ao estudo de algumas noções iniciais de conjuntos e vetores, abordaremos de forma sucinta alguns fundamentos dessa teoria.

6.1.1 Conjunto Aberto

Definição 6.1.1. *Diz-se que o ponto a é **interior** ao conjunto $X \subset \mathbb{R}$ quando existe um número $\epsilon > 0$ tal que o intervalo aberto $(a + \epsilon, a - \epsilon)$ está contido em X. Denotamos o conjunto de pontos interioes ao conjunto X como int X. Neste caso, $a \in int\ X$.*

Definição 6.1.2. *O conjunto $X \subset \mathbb{R}$ chama-se **aberto** quando todos os pontos de X são interiores a X, isto é, $X = int\ X$.*

Exemplo 6.1.1. *Todo intervalo aberto é um conjunto aberto. Os pontos a e b, extremos do intervalo fechado $[a, b]$ não são interiores a $[a, b]$. O int $[a, b] = (a, b)$. Todo ponto c do intervalo aberto (a, b) é um ponto interior a (a, b).*

Definição 6.1.3. *Diz-se que um ponto a é **aderente** ao conjunto $X \subset \mathbb{R}$ quando a é limite de alguma sequência de pontos $x_n \in X$.*

Definição 6.1.4. *Chama-se **fecho** de um conjunto X ao conjunto $\bar{X}$ formado por todos os pontos aderentes a X.*

Definição 6.1.5. *Um conjunto X diz-se **fechado** quando $X = \bar{X}$, isto é, quando todo ponto aderente a X pertence a X.*

Exemplo 6.1.2. *O fecho dos intervalos (a, b), $[a, b)$ e $(a, b]$ é o intervalo $[a, b]$.*

6.1.2 Revisão sobre vetores

Neste material, vamos considerar vetores do $\mathbb{R}^n$. Isto é, uma lista ordenada de n elementos aqui representado por letras minúsculas. O vetor $v \in \mathbb{R}^n$ é dado por $v = (v_1, v_2, ..., v_n)$. A fim de evitar ambiguidades ao longo deste capítulo, estabelecemos as seguintes convenções: letras minúsculas serão utilizadas para denotar vetores, enquanto letras com índices indicarão as coordenadas correspondentes deste vetor. Além disso, letras gregas, como α e β, serão reservadas para representar números reais, que também chamamos de escalares.

Inicialmente, abordaremos a adição de vetores. Seja $u = (u_1, u_2, ..., u_n)$, $v = (v_1, v_2, ..., v_n) \in \mathbb{R}^n$. Definimos o vetor $u + v$ da seguinte forma:

$$u + v = (u_1 + v_1, u_2 + v_2, ..., u_n + v_n).$$

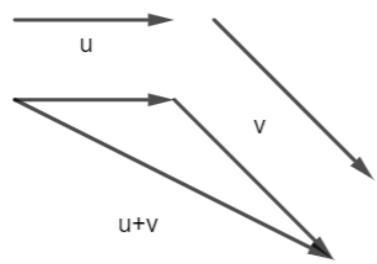


Figura 6.1: Representação geométrica da adição entre os vetores u e v.

Exemplo 6.1.3. *Seja $u, v \in \mathbb{R}^2$ em que $u = (1, 4)$ e $v = (1, 1)$. O vetor $u + v$ é dado por:*

$$u + v = (1, 4) + (1, 1) = (1 + 1, 4 + 1) = (2, 5).$$

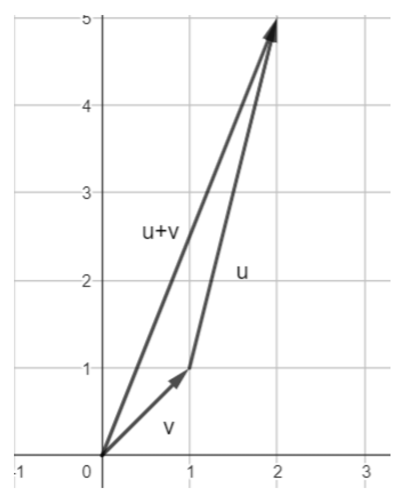


Figura 6.2: Representação geométrica do vetor $u + v$.

Propriedades 6.1.1. *A soma de vetores satisfaz as seguintes propriedades:*

(a) *Associatividade: $(u + v) + w = u + (v + w)$;*
(b) *Comutatividade: $u + v = v + u$;*
(c) *Elemento neutro: $u + 0 = u$.*

(d) *Inverso aditivo: Para cada vetor u, existe o vetor $-u \in \mathbb{R}^n$, chamado de* **vetor oposto** *do vetor u, tal que $u+(-u) = 0$. Com isso, $u - v$ é, por definição, a soma do vetor u com o vetor oposto do vetor v, isto é, $u - v = u + (-v)$.*

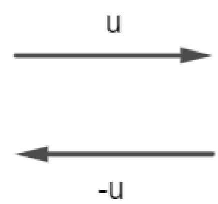


Figura 6.3: Representação geométrica do vetor oposto ao vetor u.

6.1.2.1 Norma

A norma de um vetor $v = (v_1, v_2, ..., v_n)$, denotada por $||v||$, é definida como a medida do vetor v e é calculada pela raiz quadrada da soma dos quadrados das suas coordenadas. Isto é,

$$||v|| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}.$$

Exemplo 6.1.4. *Considere o vetor $v = (3, -4)$ no plano cartesiano. Para calcular $||v||$, temos*

$$||v|| = \sqrt{3^2 + (-4)^2} = \sqrt{9 + 16} = \sqrt{25} = 5$$

Portanto, a norma do vetor v é $||v|| = 5$. Isso significa que o comprimento do vetor v é de 5 unidades.

6.1.2.2 Multiplicação de vetor por um número real (escalar)

Seja $v = (v_1, v_2, ..., v_n) \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$. Definimos o vetor αv da seguinte forma:

$$\alpha v = \alpha(v_1, v_2, ..., v_n) = (\alpha v_1, \alpha v_2, ..., \alpha v_n).$$

Geometricamente, temos que

- O vetor αv é paralelo a v;
- O vetor αv e v tem mesmo sentido se $\alpha > 0$, e sentido contrátrio se $\alpha < 0$;
- O comprimento de αv é $|\alpha|$ vezes o comprimento de v, isto é, $\|\alpha v\| = |\alpha|\|v\|$.

Propriedades 6.1.2. *A multiplicação de um vetor por um escalar satisfaz as seguintes propriedades:*

(a) $\alpha(u + v) = \alpha u + \alpha v$;
(b) $(\alpha + \beta)v = \alpha v + \beta v$;
(c) $1 \cdot v = v$;
(d) $\alpha(\beta v) = (\alpha\beta)v$.

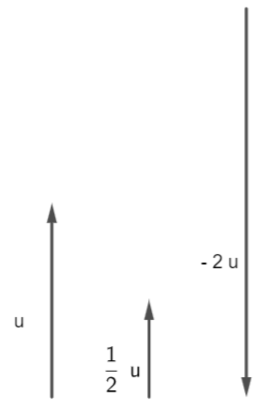


Figura 6.4: Representação geométrica da multiplicação do vetor v por escalares.

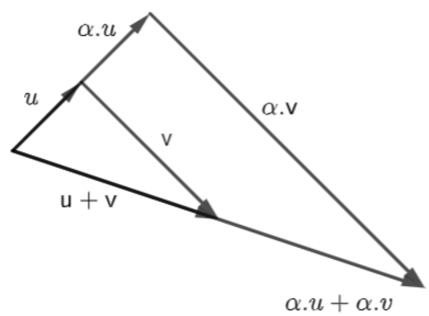


Figura 6.5: Representação geométrica da propriedade (a).

6.1.2.3 Produto interno

Definição 6.1.6. *Um produto interno é uma função que associa cada par de vetores* $u = (u_1, u_2, ..., u_n)$, $v = (v_1, v_2, ..., v_n) \in \mathbb{R}^n$ *um número real, denotado* $\langle u, v\rangle$*, que satisfaz as seguintes condições:*

(i) $\langle v, v\rangle \geq 0$, $\forall v \in \mathbb{R}^n$ *e* $\langle v, v\rangle = 0$ *se, e somente se,* $v = 0$;
(ii) $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$, $\alpha \in \mathbb{R}$;
(iii) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$, *em que* $w \in \mathbb{R}^n$;
(iv) $\langle u, v\rangle = \langle v, u\rangle$.

Ao longo deste material, consideraremos o produto interno usual, definido da seguinte forma:

$$\langle u, v\rangle = \langle (u_1, u_2, ..., u_n), (v_1, v_2, ..., v_n)\rangle = u_1 \cdot v_1 + u_2 \cdot v_2 + \ldots + u_n \cdot v_n.$$

Exemplo 6.1.5. *Considere a função* $f : \mathbb{R}^3 \to \mathbb{R}$*, em que* $f(x, y, z) = x^2 + 2y + 3z$*. O gradiente de* f *no ponto* $(2, 3, 8)$ *é o vetor* $(4, 2, 3)$*. Ao calcular o produto interno* $\langle (4, 2, 3), (1, 1, 0)\rangle$*, obtemos*

$$\langle (4, 2, 3), (1, 1, 0)\rangle = 4 \cdot 1 + 2 \cdot 1 + 3 \cdot 0 = 6.$$

Observação: Seja $v = (v_1, v_2, ..., v_n) \in \mathbb{R}^n$, temos que

$$\begin{aligned}||v||^2 = (\sqrt{v_1^2 + v_2^2 + \ldots + v_n^2})^2 &= v_1^2 + v_2^2 + \ldots + v_n^2 \\ &= v_1 v_1 + v_2 v_2 + \ldots + v_n v_n \\ &= \langle v, v \rangle.\end{aligned}$$

6.2 O Problema de Otimização

O conceito de Otimização refere-se ao processo de encontrar os pontos mínimos e/ou máximos de funções. Em termos formais, consideremos um conjunto aberto $\Omega \subset \mathbb{R}^n$ e uma função $f : \Omega \to \mathbb{R}$. O problema de encontrar x tal que f assuma seu menor valor nesse ponto pode ser descrito como:

$$\text{minimizar } f(x) \text{ sujeito a } x \in \Omega,$$

onde f é denominada **função objetivo**.

Em particular, trabalharemos ao longo dos próximos capítulos com a **Otimização Irrestrita**, isto é, $\Omega = \mathbb{R}^n$. Neste caso, podemos escrever

$$\text{minimizar } f(x) \text{ sujeito a } x \in \mathbb{R}^n. \tag{6.1}$$

Isto é, queremos encontrar o minimizador global de f corresponde ao ponto em que a função assume seu menor valor em todo domínio. Formalmente, ele é apresentado como segue.

Definição 6.2.1. *Dizemos que o ponto $\bar{x} \in \Omega$ é **minimizador global** de f, se*

$$f(\bar{x}) \leq f(x), \forall\, x \in \Omega.$$

O Teorema a seguir fornece condições sob as quais podemos afirmar a existência de um minimizador global.

Teorema 6.2.1 (Weierstrass). *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ contínua e $\Omega \subsetneq \mathbb{R}^n$ um conjunto não vazio, limitado e fechado. Então o problema de minimizar f tem solução global em Ω.*

Confira a demonstração do teorema acima em [5].

A implicação fundamental do Teorema de Weierstrass é que, sob as condições estabelecidas, sempre podemos encontrar uma solução global para problemas de minimização.

6.3 Condições de Otimalidade

Os princípios fundamentais que norteiam a busca por soluções ótimas em problemas de otimização envolvem a compreensão das condições que os pontos críticos de uma função devem satisfazer. Essas condições são importantes ao examinar se um determinado ponto é um minimizador da função, isto é, a solução ótima do problema.

Teorema 6.3.1. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ diferenciável no ponto $\bar{x}$. Se $\bar{x}$ for um minimizador local de f, então*

$$\nabla f(\bar{x}) = 0. \tag{6.2}$$

A demonstração do teorema acima está fora do escopo deste material e pode ser vista em [5]. Em essência, o teorema em questão garante que, em torno de $\bar{x}$, a função f não varia significativamente na direção do vetor gradiente, indicando que $\bar{x}$ é um ponto onde f atinge um mínimo local. Em termos práticos, essa interpretação sugere que, ao buscar minimizadores de uma função, podemos procurar pontos onde o gradiente se anula.

Cabe ressaltar que esta condição refere-se ao caso irrestrito e é denominada *condição necessária de primeira ordem* para o problema (6.1). Os pontos que satisfazem essa condição são chamados de **pontos estacionários** do problema.

6.4 Métodos

Resolver problemas de otimização nem sempre é uma tarefa simples devido à complexidade das funções envolvidas e das restrições impostas ao domínio da função. Por esse motivo, recorremos a diferentes métodos com o objetivo de gerar uma sequência que convirja para a solução do problema. Esses métodos começam com um ponto inicial x_0, frequentemente chamado de "**chute**" e, a partir de diferentes técnicas, obtemos um ponto melhor x_1. As informações derivadas dos pontos anteriores definem uma sequência que gradualmente se aproxima da solução ótima do problema. Durante este processo, é gerada uma sequência (x_k) de pontos do $\mathbb{R}^n$, onde k representa a iteração. Neste caso, para um k suficientemente grande, x_k deve ser uma boa aproximação da solução $\bar{x}$ do problema de minimizar f. Isto é,

$$(x_k) \to \bar{x} \ (k \to \infty).$$

Uma estratégia muito usual nos métodos de otimização é escolher, a partir de cada ponto obtido, uma direção $d_k \in \mathbb{R}^n$ e calcular um comprimento de passo $\alpha_k > 0$, que resulta em um valor menor de f em x_{k+1} do que no ponto x_k, ou seja,

$$f(x_k + \alpha_k d_k) < f(x_k). \tag{6.3}$$

Dessa forma, obtemos o próximo iterando $x_{k+1} = x_k + \alpha_k d_k$, repetindo o processo para o novo ponto x_{k+1}. Essa estratégia de atualização dos pontos permite que nos movamos em direção às regiões onde a função objetivo é minimizada, utilizando uma combinação apropriada de direção e tamanho de passo para guiar o processo de iterativo em direção à solução desejada.

Definição 6.4.1. *Dizemos que $d \in \mathbb{R}^n$ é uma* ***direção de descida*** *da função $f : \mathbb{R}^n \to \mathbb{R}$ no ponto $x \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que*

$$f(x + \alpha d) < f(x), \ \forall \alpha \in (0, \varepsilon].$$

Denotamos por $D_f(x)$ o conjunto de todas as direções de descida da função f no ponto x.

Teorema 6.4.1 (Direções de descida). *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $x \in \mathbb{R}^n$. Então:*
(i) Para todo $d \in D_f(x)$, tem-se $\langle \nabla f(x), d \rangle \leq 0$.
(ii) Se $d \in \mathbb{R}^n$ satisfaz $\langle \nabla f(x), d \rangle < 0$, tem-se que $d \in D_f(x)$.

Confira a demonstração em [3].

6.5 Condições de parada

Ao implementar métodos computacionais de Otimização, utilizamos condições de parada que garantam que o algoritmo alcance uma boa aproximação da solução do problema. Na prática computacional, são empregadas regras de parada baseadas em informações obtidas no ponto x_k. Em casos irrestritos com a função objetivo f diferenciável, uma condição de parada frequentemente utilizada é dada por:

$$||\nabla f(x_{k+1})|| < \varepsilon,$$

onde ε é um número pequeno. Essa condição de parada reflete a ideia de que, em torno de um ponto de mínimo local, o gradiente da função se aproxima de zero. Portanto, se a norma do gradiente se torna menor do que ε, isso indica que estamos em uma região próxima do minimizador local. Em geral, quanto menor o valor de ε, maior será a precisão da solução obtida, mas isso também implica em um custo computacional mais elevado.

6.6 Algoritmos

Já sabemos que ao implementar um método em Otimização, o processo começa com a escolha de um ponto inicial $x_0 \in \mathbb{R}^n$. Este ponto pode ser determinado aleatoriamente, com base em conhecimento prévio do problema. Após isso, o algoritmo passa a gerar uma sequência de pontos (x_k). Durante cada iteração, o algoritmo avalia a função objetivo f no ponto atual x_k para determinar o valor da função nesse ponto.

Além disso, em cada iteração, o algoritmo procura uma direção que ajuste o ponto atual x_k com o objetivo de reduzir o valor da função objetivo. Essa direção é importante para guiar o algoritmo na direção da solução do problema e é escolhida com base no método selecionado.

Posteriormente, estudaremos o Método do Gradiente, que utiliza o anti-gradiente da função $-\nabla f(x_k)$ como direção de descida. O objetivo é encontrar uma direção na qual a função decresça, permitindo que o algoritmo se aproxime do mínimo desejado. Após identificar essa direção, o algoritmo determina o comprimento do passo α_k a ser dado nessa direção. Esse comprimento de passo é escolhido de forma a garantir que o próximo ponto, x_{k+1}, esteja mais próximo do mínimo da função.

O processo de atualização dos pontos é repetido até que uma condição de parada seja alcançada. Neste material, essa condição será baseada no comportamento da norma do gradiente da função. Quando essa condição é satisfeita, o algoritmo termina e retorna o último ponto x_k como uma aproximação da solução do problema de minimizar f.

6.7 Exercícios

Exemplo 6.7.1. *Seja $u, v, w \in \mathbb{R}^n$. Prove que $u + v = u + w \Rightarrow v = w$.*

Exemplo 6.7.2. *Seja $v \in \mathbb{R}^n$ e α um número real. Prove as seguintes regras de sinais: a) $(-\alpha) \cdot v = -(\alpha \cdot v)$ b) $\alpha \cdot (-v) = -(\alpha \cdot v)$ c) $-\alpha - v = \alpha \cdot v$*

Exemplo 6.7.3. *Considerando o $\mathbb{R}^3$, com o produto interno usual, calcule $\langle u, v\rangle$ nos seguintes casos: a) $u = (\frac{1}{2}, 2, 1)$ e $v = (4, 1, -3)$. b) $u = (2, 1, 0)$ e $v = (4, 0, 2)$. c) $u = (1, 0, 1)$ e $v = (1,$ d)*

Exercício 6.7.1. *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x, y) = sen\, x\; sen\, y + e^{x^2+y^2}$. Mostre que $\bar{x} = (0, 0)$ é ponto estacionário de f.*

Exercício 6.7.2. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x, y) = x^2 - xy + 2y^2 - 2x + \frac{2}{3}y + e^x + y$. Mostre que $\bar{x} = \frac{1}{3}\begin{pmatrix}1\\-1\end{pmatrix}$ é um ponto estacionário de f;*

Exercício 6.7.3. *Considere o problema irrestrito*

$$\text{minimizar } f(x, y) = x^2 - xy + 2y^2 - 32x + e^x + y, \text{ sujeito a } x \in \mathbb{R}^2.$$

(a) Verifique que o ponto $\bar{x} = (0, 0)$ não é ótimo;
(b)Minimize a função a partir de $\bar{x}$ na direção $d = -\nabla f(\bar{x}))$.

Exercício 6.7.4. *Considere $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x, y) = \frac{1}{2}(x - 2)^2 + (y - 1)^2$ e $\bar{x} = \begin{pmatrix}1\\0\end{pmatrix}$. Mostre que $d = \begin{pmatrix}0\\1\end{pmatrix}$ é uma direção de descida para f.*

Exercício 6.7.5. *Sejam $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x, y) = \frac{1}{2}(x^2 + y^2)$, $\bar{x} = \begin{pmatrix}1\\0\end{pmatrix}$ e $d = \begin{pmatrix}d_1\\d_2\end{pmatrix}$. Mostre que se $d_1 < 0$, então d é uma direção de descida para f, a partir de $\bar{x}$. Estude o caso $d_1 = 0$.*

Capítulo 7
Aplicações e Algoritmos

Neste capítulo, apresentamos os princípios básicos que envolvem a estruturação de um algoritmo. Além disso, apresentaremos um método utilizado para encontrar o mínimo d uma função. Este método é chamado Método da Bisseção.

7.1 Estrutura de um algoritmo

Um algoritmo pode ser definido como uma sequência fnita de instruções para resolver determinado problema. Para que um algoritmo seja bem formulado é preciso:

- Definir ações simples e sem ambiguidade.
- Organizar as ações de forma ordenada.
- Estabelecer as ações dentro de uma sequência finita de passos.

Além do mais, os algoritmos são capazes de:

- Avaliar expressões algébricas, relacionais e lógicas.
- Tomar decisões com base nos resultados das expressões avaliadas.
- Repetir um conjunto de ações de acordo com uma condição.

Veja no Algoritmo 1 um exemplo prático de como funciona um algoritmo.

Algoritmo 1: TROCA DE PNEU DO CARRO

```
1  Puxe o freio de mão
2  Desligue o carro
3  Ligue o pisca alerta
4  Pegue as ferramentas
5  Pegue o estepe
6  Suspenda o carro com o macaco
7  Desenrosque os parafusos do pneu
8  Coloque o estepe
9  Enrosque os parafusos
10 Abaixe o carro com o macaco
11 Guarde as ferramentas
```

No exemplo acima, podemos perceber que o algoritmo apresentado tem como função instruir a forma de trocar um pneu em um número finito de passos. Na linha 1, por exemplo é do o primeiro comando para puxar o freio de mão e em seguida as demais informações são dadas de forma ordenada visando chegar ao fim que é a troca do pneu.

A estrutura dos algoritmos é feita de forma organizada logicamente. A estrutura básica de um algoritmo inclui várias partes essenciais que ajudam a garantir que ele funcione de forma eficiente e correta.

7.1.0.1 Entrada

Na entrada serão fornecidos dados ao algoritmo para que ele possa realizar suas operações, esses dados, neste caso, são o ponto de partida para q o processo seja iniciado. Exemplo desses dados são : lista de números, um conjunto de coordenadas, valores de tolerância, entre outros.

7.1.0.2 Saída

Serão os resultados que são fornecidos após o processamento das entradas. Assim, é preciso que na saída esteja a resolução do problema ou a realização do que foi pedido. Exemplos de sádas pode ser: soma de uma lista de números, ordenação de um conjunto de valores, o mínimo de uma função, entre outros.

7.1.0.3 Laço

Um laço é uma estrutura de controle que faz com que um bloco do código seja executado repetidamente um bloco de código enquanto uma condição específica for verdadeira. Exemplo de laço é o *while*.

O *while* é uma estrutura de controle de repetição que executa um bloco de código enquanto uma condição específica for verdadeira.

```
while (condição)
    // código a ser executado
end
```

7.1.0.4 For

O *for* é uma estrutura de controle de repetição que executa um bloco de código um número específico de vezes.

```
for (inicialização; condição; incremento/decremento) faça
    // código a ser executado
end
```

7.1.0.5 If

O *if* é uma estrutura de controle condicional que executa um bloco de código de uma condição específica for verdadeira. Pode ser seguido, se for o caso de um *else* para executar um bloco de código se a condição for falsa.

```
if (condição) then
    // código a ser executado se a condição for verdadeira
else
    // código a ser executado se a condição for falsa
end
```

Essas estrututuras são de fundamental importância para a construção e implementação dos códigos.

Com o intuito de nos aprofundarmos nos métodos da Otimização Contínua, apresentaremos o Método da Bisseção na próxima seção.

7.2 Método da Bisseção

Nesta seção, apresentaremos o Método da Bisseção, que é uma técnica utilizada para encontrar o ponto onde uma função contínua atinge seu mínimo em um intervalo fechado. As funções que serão utilizadas possuem um único mínimo local no intervalo dado e devem ser contínuas.

Intuitivamente, imagine duas crianças brincando de encontrar números inteiros entre 1 e 100. A brincadeira consiste em o menino pensar em um determinado número e a menina precisa descobrir. De início a menina chuta o número 50 e o menino responde que o número é maior. A menina chuta agora o número 75, e o menino responde que o número é menor do que 75. Agora a menina chuta o número 67, mas o menino responde que o número é menor do que 67. Podemos observar que a medida que a menina chuta ela se aproxima do ponto médio dos intervalos possíveis. E com esse procedimento, a menina em algum momento irá achar o número pensado pelo menino. O Método da Bisseção consiste nessa ideia para encontrar mínino de funções de apenas uma variável.

O método fará os valores de uma função $f : [a, b] \to \mathbb{R}$ em dois pontos de $[a, b]$ se aproximarem do minimizador global em intervalos de comprimentos cada vez menores. Repetindo esse procedimento, podemos diminuir cada vez mais o intervalo que contém a solução.

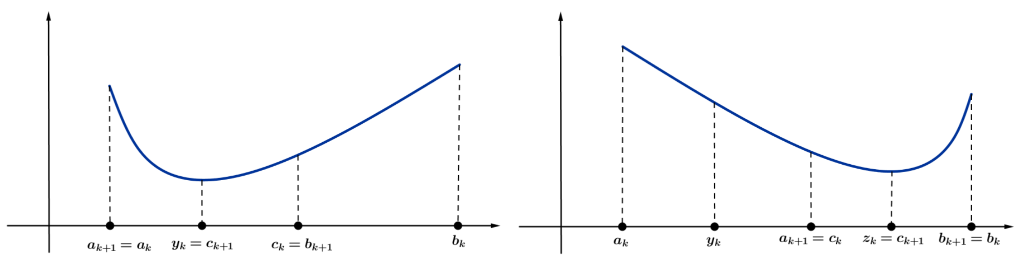
Tomamos um intervalo $[a_1, b_1]$, seja $c_1 = (a_1 + b_1)/2$, e calcula-se $f(c_1)$. Definimos $y_k = (a_k + c_k)/2$ e calcula-se $f(y_k)$. Assim teremos:

- Se $f(y_k) \leq f(c_k)$, definir $a_{k+1} = a_k, b_{k+1} = c_k, c_{k+1} = y_k$.
- Se $f(y_k) > f(c_k)$, definir $z_k = (c_k + b_k)/2$ e calcular $f(z_k)$.

Nota-se que descartamos o intervalo $[c_k, b_k]$. Vamos considerar o caso em que descartamos o intervalo $[a_k, c_k]$. Temos:

- Se $f(c_k) \leq f(z_k)$, definir $a_{k+1} = y_k, b_{k+1} = z_k, c_{k+1} = c_k$.
- Se $f(c_k) > f(z_k)$, definir $a_{k+1} = c_k, b_{k+1} = b_k, c_{k+1} = z_k$.

Assim, repetimos o processo iterativo até que seja possível encontrar a aproximação desejada. A seguir, apresentamos dois exemplos geométricos de iterações do Método da Bisseção.



(a) (b)

Figura 7.1: Uma iteração do Método da Bisseção

O descarte de um dos intervalos é pelo fato de que o intervalo descartado os valores são maiores, logo não irá convergir para o mínimo.

Como todo método numérico, o Método da Bisseção possui pontos positivos e negativos. Uma vantagem que podemos citar é que esse método possui um algoritmo simples e direto, de fácil compreensão e não exige a diferenciabilidade da função. Além disso, possui convergência definida, garantindo que uma solução será encontrada se ela estiver dentro de um determinado intervalo. Entretanto, esse método pode ser relativamente lento, especialmente quando comparado com outros métodos numéricos e também requer que a função seja contínua no intervalo de interesse, e isso pode não ocorrer. Por fim, enfatizamos que o Método da Bisseção fornece uma solução aproximada e múltiplas iterações são necessárias para uma maior precisão.

7.2.1 Algoritmo

Agora, iremos apresentar o algoritmo do Método da Bisseção.

Algoritmo 2: Método da Bisseção para Encontrar o Mínimo

Defina a função f
Defina os extremos a e b do intervalo inicial
Defina a tolerância $\epsilon > 0$
while $(b - a)/2 > \epsilon$ **do**
    Calcule os pontos: $c = \frac{a+b}{2}$, $d = c - \epsilon$, $e = c + \epsilon$
    Avalie a função nos pontos: $f(c)$, $f(d)$, $f(e)$
    if $f(d) < f(c)$ **then**
        Atualize $b = c$
    else if $f(e) < f(c)$ **then**
        Atualize $a = c$
    else
        Atualize $a = d$ e $b = e$
    end if
end while=0

7.2.2 Aplicações

Exemplo 7.2.1. *Uma empresa está estudando o comportamento de suas despesas em função do custo de seu produto. A despesa D (em milhares de dólares) em função do custo x (em reais) é dada pela função:*

$$D(x) = x^4 + 4x^3 - 6x^2 + 4x + 14$$

A empresa deseja encontrar o custo c que minimiza suas despesas. Vamos utilizar o método da bisseção para encontrar este custo com uma precisão de $\epsilon = 1 \times 10^{-6}$.

Passos do Método da Bisseção

1. ***Definir a Função e o Intervalo Inicial:***
   - *Função de despesas:* $D(x) = x^4 + 4x^3 - 6x^2 + 4x + 14$
   - *Intervalo inicial:* $[-5, 5]$
2. ***Calcular o Ponto Médio Inicial:***
   - *Ponto médio:* $c_1 = \frac{-5+5}{2} = 0$
   - *Avaliar a função:* $D(c_1) = D(0) = 14$
3. ***Primeira Divisão:***
   - *Novo ponto médio:* $y_1 = \frac{-5+0}{2} = -2.5$
   - *Avaliar a função:* $D(y_1) = D(-2.5) = -56.9375$
4. ***Comparação e Redução do Intervalo:***
   - *Como* $D(y_1) \leq D(c_1)$*:*
   - *Novo intervalo:* $[a_2, b_2] = [-5, 0]$
   - *Atualizar o ponto médio:* $c_2 = y_1$

Repetindo esse processo no intervalo $[-5, 0]$

1. ***Calcular o Ponto Médio Inicial:***
   - *Ponto médio:* $c_2 = -2.5$
   - *Avaliar a função:* $D(c_2) = D(-2.5) = -56.9375$
2. ***Segunda Divisão:***

- *Novo ponto médio:* $y_2 = \frac{-5+(-2.5)}{2} = -3.75$
- *Avaliar a função:* $D(y_2) = D(-3.75) = -98.55859375$

3. ***Comparação e Redução do Intervalo:***

- *Como* $D(y_2) \leq D(c_2)$*:*
- *Novo intervalo:* $[a_3, b_3] = [-5, -2.5]$
- *Atualizar o ponto médio:* $c_3 = y_2$

Continuamos esse processo até que a largura do intervalo seja menor que a precisão desejada. Fazendo isso temos que $x \approx -3.8473221054300666$*.*

Exercício 7.2.1. *Utilize o método da bisseção para minimizar as seguintes funções:*

(a) $f_1(x) = e^{-x} + cos(x)$ *no intervalo* $[2, 5]$
(b) $f_2(x) = 10xe^{-x^2} - 1$ *no intervalo* $[-1, 1]$
(c) $f_3(x) = x^3 - 2x + 2$ *no intervalo* $[-3, 3]$
(d) $f_4(x) = x^2 - e^x - 3x + 2$ *no intervalo* $[-2, 2]$
(e) $f_5(x) = sin(x) - \frac{x}{2}$ *no intervalo* $[-3, 5]$
(f) $f_6(x) = sin(x) - x^5 + x^3 - 1$ *no intervalo* $[-5, 2]$

Capítulo 8
Buscas lineares

As buscas lineares desempenham o papel de capacitar os algoritmos a discernir o comprimento de passo mais apropriado a cada nova iteração, em direção ao minimizador da função objetivo. Neste capítulo, discutiremos como o comprimento de passo influencia na escolha dos pontos da sequência gerada, abordando dois tipos de buscas comuns em Otimização Contínua: comprimento de passo fixo e Armijo.

8.1 Interpretação geométrica das direções de descida

Nas seções anteriores, discutimos sobre direções de descida e sua importância na determinação do caminho em busca do minimizador de f a partir de um determinado ponto. Agora, vamos explorar a interpretação geométrica dessas direções.

Quando definimos uma direção de descida, estamos essencialmente escolhendo um vetor que nos leva em direção ao ponto onde a função objetivo f decresce. Matematicamente, isso é expresso pelo Teorema 6.4.1. Considere a função $f : \mathbb{R} \to \mathbb{R}$, definida por $f(x) = (x-3)^2$, cujo esboço está representado na figura a seguir.

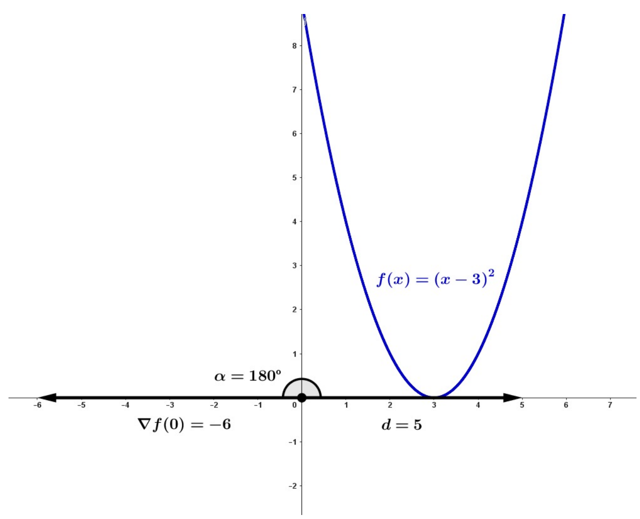


Figura 8.1: Representação geométrica da função $f(x) = (x-3)^2$

Para esta função, o vetor gradiente é calculado como segue:

$$\nabla f(x) = (x-3)^2 = 2(x-3).$$

Em $x = 0$, temos:

$$\nabla f(0) = 2(0-3) = -6.$$

O vetor gradiente ∇f calculado no ponto $x = 0$ aponta na direção de maior crescimento da função f em $x = 0$. Na imagem, observamos o vetor **d** representando uma possível direção a partir do ponto $x = 0$.

Na imagem, o ângulo α entre o vetor gradiente $\nabla f(0)$ e a direção **d** é 180°. Como $\cos(180°) = -1$, temos:

$$\langle \nabla f(0), \mathbf{d} \rangle = |\nabla f(0)| \cdot |d| \cdot \cos(\alpha) = 6 \cdot 5 \cdot (-1) = -30 < 0.$$

O resultado negativo nos diz que **d** é uma direção de descida.

Exemplo 8.1.1. *Vamos mostrar que $d_k = (-2x_k + 6)$ é direção de descida da função $f(x) = (x-3)^2$.*

Primeiro, vamos calcular o gradiente da função

$$f(x) : \nabla f(x) = \frac{\partial f}{\partial x} = 2(x-3).$$

Agora, verifiquemos se o produto escalar entre o gradiente e a direção d_k é negativo:

$$\nabla f(x) \cdot d_k = 2(x-3) \cdot (-2x_k + 6).$$

Expandindo o produto escalar: $\nabla f(x) \cdot d_k = -4x(x_k - 3) + 12(x-3)$. Simplificando:

$$\nabla f(x) \cdot d_k = -4xx_k + 12x + 12 - 36\nabla f(x) \cdot d_k = -4xx_k + 12x - 24.$$

Para que d_k seja uma direção de descida, precisamos que $\nabla f(x) \cdot d_k < 0 : -4xx_k + 12x - 24 < 0$.

Dividindo todos os termos por 4:

$$-xx_k + 3x - 6 < 0.$$

Reorganizando:

$$x_k > 3 - \frac{3x}{x}.$$

Portanto, a direção $d_k = -2x_k + 6$ é uma direção de descida para a função $f(x) = (x-3)^2$. □

Na Equação (6.3), faz-se necessário não apenas o cálculo da direção de descida d, mas também de um comprimento de passo α. O comprimento de passo determina o quanto iremos avançar ao longo dessa direção a partir de um determinado ponto. Quando este comprimento de passo é apropriado, ele pode auxiliar no desempenho do algoritmo. Veremos nas próximas seções que a escolha de um comprimento de passo muito longo pode fazer com que o algoritmo ultrapasse a solução do problema. Por outro lado, um comprimento de passo muito curto pode levar a um processo de convergência extremamente lento, aumentando o tempo computacional necessário para alcançar uma aproximação satisfatória da solução. Neste contexto, existem alternativas para ajustar o comprimento do passo de maneira mais eficaz.

8.2 Busca do comprimento de passo fixo

Ao escolher a busca do comprimento de passo fixo, um número real $\alpha > 0$ pré-determinado, que não depende de k, é mantido constante ao longo das iterações. A atualização dos pontos da sequência (x_k) ao longo da direção de descida d_k é dada por:

$$x_{k+1} = x_k + \alpha d_k,$$

O algoritmo da busca do comprimento de passo fixo é fácil de implementar, pois não requer cálculos adicionais ou complexos para determinar o comprimento do passo em cada iteração. Além disso, como o comprimento do passo é constante, o tempo computacional por iteração é reduzido, uma vez que não há necessidade de avaliação adicional da função objetivo para ajustar o passo. No entanto, existem algumas limitações ao se utilizar essa busca. A eficácia do passo fixo depende fortemente da escolha do valor de α. Se α for muito grande, o algoritmo pode ultrapassar o minimizador da função objetivo. Caso, α seja muito pequeno, a convergência se torna extremamente lenta, exigindo muitas iterações para alcançar uma solução satisfatória, o que aumenta o custo computacional total.

Exemplo 8.2.1. *Suponha que queremos minimizar a função quadrática $f : \mathbb{R} \to \mathbb{R}$, definida por $f(x) = (x-3)^2$. Ao escolhemos um ponto inicial $x_0 = 0$, definimos $\alpha = 0.1$ e considerarmos o vetor $d_k = (-2x_k + 6)$ como direção de descida para cada iteração, obtemos a seguinte atualização de pontos da sequência:*

$$x_{k+1} = x_k + 0.1 \cdot (-2x_k + 6)$$

Para a primeira iteração, tome $x_0 = 0$ e $d_0 = (-2 \cdot 0 + 6) = 6$. Pelo esquema iterativo:

$$x_1 = 0 + 0.1 \cdot 6 = 0 + 0.6 \Longrightarrow x_1 = 0.6.$$

Na segunda iteração, $d_1 = (-2 \cdot 0.6 + 6) = 4.8$. Assim:

$$x_2 = 0.6 + 0.1 \cdot (-4.8) = 0.6 + 0.48 \Longrightarrow x_2 = 1.08.$$

Na terceira iteração, temos $d_2 = (-2 \cdot 1.08 + 6) = 3.84$:

$$x_3 = 1.08 + 0.1 \cdot (3.84) = 1.08 + 0.384 \Longrightarrow x_3 = 1.464.$$

Veja que $f(0) = 9 > f(0.6) = 5.76 > f(1.08) = 3.6864$.

Repetindo este processo, podemos ver que o valor de x vai se aproximando lentamente de 3, o minimizador da função f. Neste exemplo, escolhemos $\alpha = 0.1$. Se tivéssemos escolhido um valor maior, como $\alpha = 1.5$, o algoritmo poderia ultrapassar o ponto ótimo, resultando em oscilações. Se tivéssemos escolhido um valor menor, como $\alpha = 0.01$, a convergência seria ainda mais lenta. Embora a simplicidade do passo fixo seja uma vantagem, a falta de adaptabilidade pode ser significativa em problemas mais complexos. Deixaremos como exercício para o leitor realizar algumas iterações, seguindo o exemplo anterior, com os valores de α sugeridos, a fim de observar o comportamento da sequência gerada.

8.3 Busca de Armijo

A busca de Armijo, em contraste com a busca do comprimento de passo fixo, onde o comprimento do passo permanece constante em todas as iterações, ajusta o comprimento do passo a cada nova iteração. Em essência, essa busca consiste em computar um comprimento de passo que resulta em um decréscimo suficiente da função f em relação ao valor $f(x_k)$, isto é, determinar α tal que

$$f(x_k + \alpha d_k) \leq f(x_k) + \sigma\alpha\langle \nabla f(x_k), d_k\rangle, \tag{8.1}$$

em que f é diferenciável no ponto x_k, $\alpha > 0$ e $\sigma \in (0, 1)$.

Se a busca de Armijo não for atendida para um determinado comprimento de passo α, então ele é multiplicado por um parâmetro $\theta \in (0, 1)$, até que a desigualdade (8.1) seja satisfeita. Formalmente, essa busca é apresentada pelo resultado a seguir.

Teorema 8.3.1. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $x_k \in \mathbb{R}^n$. Suponhamos que $d_k \in \mathbb{R}^n$ seja direção de descida. Então, a desigualdade*

$$f(x_k + \alpha d_k) \leq f(x_k) + \sigma\alpha\langle \nabla f(x_k), d_k\rangle$$

é satisfeita para todo α suficientemente pequeno. Em particular, a busca de Armijo está bem definida e termina com um $\alpha_k > 0$.

Confira a demonstração em [3, Capítulo 3, Lema 3.1.4]. O Teorema acima garante, se a direção d_k for de descida em relação à função objetivo f no ponto x_k, então a busca de Armijo é bem definida, pois garante a existência de um comprimento de passo adequado $\alpha_k > 0$ que satisfaz a desigualdade (8.2).

8.3.1 Interpretação geométrica

A Imagem abaixo representa uma interpretação geométrica da busca de Armijo.

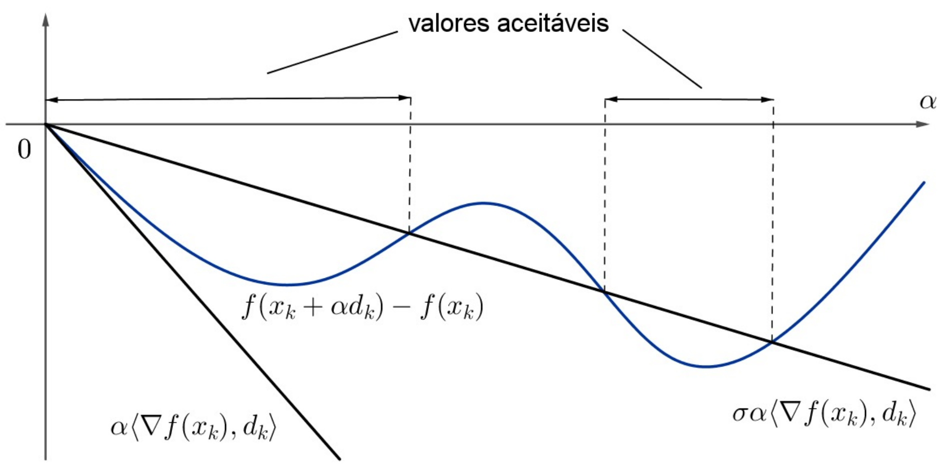


Figura 8.2: Valores de α que satisfazem a desigualdade de Armijo.

Note que a desigualdade (8.1) é satisfeita quando o gráfico de $f(x_k+\alpha d_k)-f(x_k)$ está abaixo da reta $\sigma\alpha\langle\nabla f(x_k), d_k\rangle$, isto é,

$$f(x_k + \alpha d_k) - f(x_k) \leq \sigma\alpha\langle\nabla f(x_k), d_k\rangle$$
$$f(x_k + \alpha d_k) \leq f(x_k) + \sigma\alpha\langle\nabla f(x_k), d_k\rangle.$$

Quando a desigualdade não é satisfeita, o valor do comprimento de passo α é ajustado até que α caia dentro do intervalo de valores de que satisfazem a desigualdade de Armijo.

Exemplo 8.3.1. *Considere $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x_1, x_2) = \frac{1}{2}(x_1 - 2)^2 + (x_2 - 1)^2$, $x_0 = (1, 0)$, $d = (3, 1)$ e $\sigma = 0,8$. Faça uma busca de Armijo a partir de x_0, na direção d. Temos que d é uma direção de descida, pois*

$$\langle\nabla f(x_0), d\rangle = \langle(-1, -2), (3, 1)\rangle = -5 < 0.$$

Começando com $\alpha = 1$, teremos o passo recusado, pois

$$f(x_0 + \alpha d) > f(x_0) + \sigma\alpha\langle\nabla f(x_0), d\rangle.$$

Então fazemos $\alpha = 0,8 \cdot 1$, que também é recusado. Enfim, fazendo $\alpha = 0,8 \cdot 0,8 = 0,64$, teremos o passo aceito. Assim,

$$\alpha = 0,64 \quad e \quad x_0 + \alpha d = (2,92\ 0,64).$$

Capítulo 9
Método do Gradiente equipado com Busca Linear

9.1 Introdução

Como anunciado na segunda seção do Capítulo 6, trabalharemos com um método de **Otimização Irrestrita**, ou seja, dada uma função $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, Ω será o maior conjunto do qual o valor $f(x)$, $x \in \Omega$ faça sentido.

Exemplo 9.1.1. *Seja $f : \Omega \subset \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$. Como todo número real pode ser elevado ao quadrado, concluímos que $\Omega = \mathbb{R}$ e, assim, o problema irrestrito pode ser enunciado da seguinte forma:*

$$\textit{minimizar } f(x) \textit{ sujeito a } x \in \Omega = \mathbb{R}.$$

Exemplo 9.1.2. *Seja $g : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ tal que $g(x, y) = x^2\sqrt{y-1}$. Observe que $y - 1 \geq 0$, isto é, $y \geq 1$ para que $x^2\sqrt{y-1}$ faça sentido. Portanto, o problema de otimização irrestrita de g é formalmente descrito como:*

$$\textit{minimizar } g(x) \textit{ sujeito a } x \in \Omega = \{(x, y) \in \mathbb{R}^2 : y \geq 1\}.$$

Observe que Ω é um semiplano, como vemos na Figura 9.1.

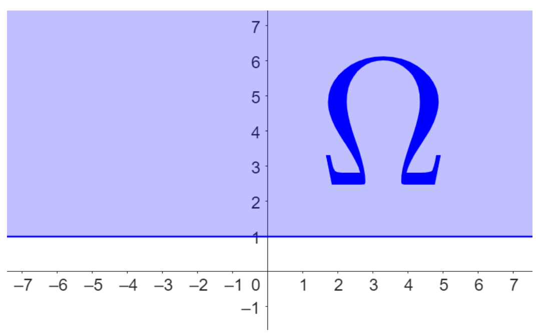


Figura 9.1: Semiplano Ω.

Para minimizarmos uma função f, é necessário encontrarmos os seus pontos críticos, por isso utilizaremos um método iterativo que irá gerar uma sequência que converge para um ponto crítico. Com esse intuito, a partir do chute $x_0 \in \Omega$ queremos um ponto $x_1 \in \Omega$ de tal forma que $f(x_0) > f(x_1)$. Como podemos obtê-lo? Pela soma do ponto com um vetor temos que x_1 é da forma $x_1 = x_0 + \alpha d_0$ com $\alpha > 0$, sendo o parâmetro de comprimento de passo e d_0 uma direção de descida. Para ilustrar a importância de uma boa escolha de α e d_0, considere a função $f : \mathbb{R}^2 \to \mathbb{R}$ definida por $f(x, y) = x^2 + y^2$. Evidentemente, o min $f(x)$, com $x \in \mathbb{R}^2$, é $(0, 0)$, já que $x^2 + y^2 \geq 0$ para todo $x, y \in \mathbb{R}$ e $f(0, 0) = 0$. As curvas de nível de f são apresentadas na Figura 9.2.

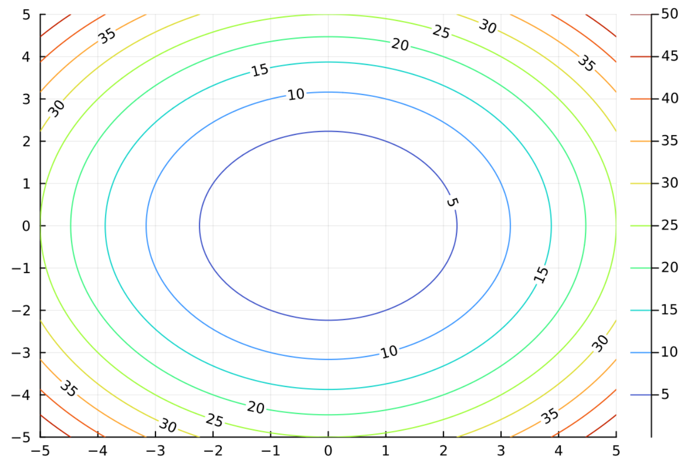


Figura 9.2: Curvas de nível da função f.

Considerando $x_0 = (3, 2)$, observe que não podemos tomar $d_0 = (0, 1)$ e $\alpha = 2$, pois teríamos $x_1 = x_0 + \alpha d_0 = (3, 2) + 2 \cdot (0, 1) = (3, 4)$, $10 < f(3, 2) < 15$ e $f(3, 4) = 25$, isto é, $f(x_0) < f(x_1)$. Assim, obteríamos um resultado contrário ao que gostaríamos que acontecesse. Isso pode ser visto na Figura 9.3

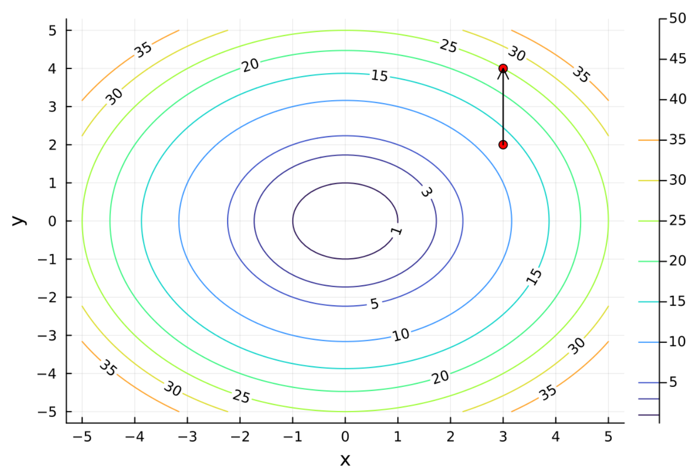


Figura 9.3: Exemplo de direção que não fornece decrescimento para o valor da função.

Por outro lado, sempre podemos utilizar a direção do anti-gradiente como uma direção de descida (veja a Definição em 6.4.1). Isto segue do fato que se $d = -\nabla f(x)$ com $x \in \mathbb{R}^2$ sendo não estacionário, isto é, $\|\nabla f(x)\| \neq 0$, então

$$\langle \nabla f(x), d \rangle = \langle \nabla f(x), -\nabla f(x) \rangle = -\langle \nabla f(x), \nabla f(x) \rangle = -\|\nabla f(x)\|^2 < 0.$$

Portanto, pelo Teorema 6.4.1, concluímos que a direção do anti-gradiente é uma direção de descida em um ponto não estacionário. Formalmente, o que acabamos de mostrar é enunciado pelo seguinte resultado:

Proposição 9.1.1. *Seja $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ diferenciável em $x \in \Omega$. Se x não for estacionário, então $d = -\nabla f(x)$ é uma direção de descida no ponto x.*

Observação 9.1.1. *Se x for ponto estacionário, então x é um candidato a minimizador de f. O que justifica a afirmação de que "sempre podemos utilizar a direção do anti-gradiente como direção de descida (desde que f seja diferenciável)".*

Em relação ao comprimento do passo, como já obtivemos uma direção de descida, podemos agora encontrar um α que garanta a condição de decrescimento $f(x_0) > f(x_1)$ usando a Busca de Armijo.

Assim, usando a direção do anti-gradiente e calculando α pela Busca de Armijo com $\eta = 0,7$, obtemos o seguinte resultado: $10 < f(x_0) < 15$ e $3 < f(x_1) < 5$, ou seja, $f(x_0) > f(x_1)$ como gostaríamos. A Figura 9.4 ilustra este resultado.

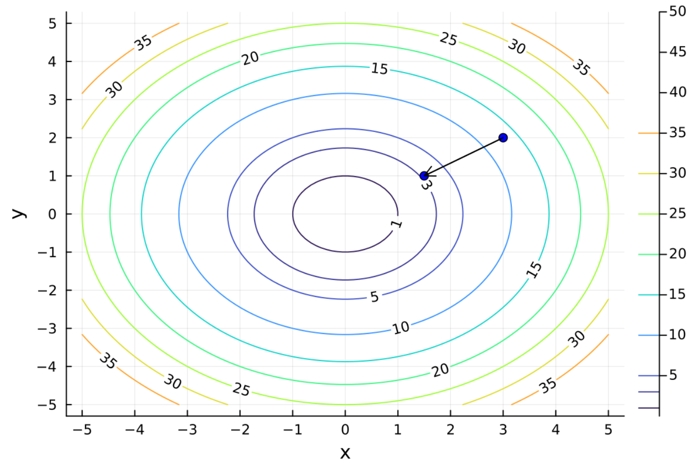


Figura 9.4: Direção de descida com comprimento de passo aceitável.

Por fim, cabe salientar que o valor de α adequado é muito importante, pois se α fosse o quíntuplo do calculado pela Busca de Armijo obteríamos a seguinte situação: $10 < f(x_0) < 15$ e $25 < f(x_1) < 30$, ou seja, $f(x_0) < f(x_1)$ obtendo uma situação indesejada novamente. Observe na Figura 9.5 que o novo ponto estaria mais distante da solução do problema que o ponto inicial.

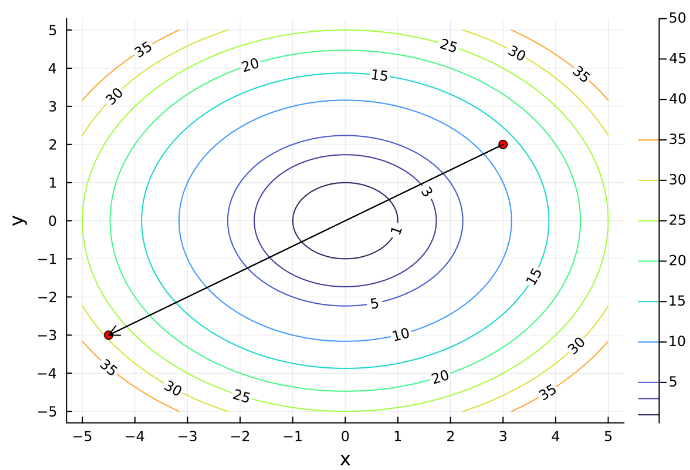


Figura 9.5: Direção de descida correta com comprimento de passo não aceitável.

De forma geral, dado um chute $x_0 \in \Omega$ podemos definir o restante dos termos da sequência de pontos $\{x_k\}_{k\in\mathbb{N}} \in \Omega$ recursivamente por

$$x_{k+1} = x_k + \alpha_k d_k, \text{ com } k \geq 1. \tag{9.1}$$

Se tomarmos $d_k = -\nabla f(x_k)$ para todo $k \geq 1$, então dizemos que a sequência $\{x_k\}_{k\in\mathbb{N}}$ gerada foi obtida pelo **Método do Gradiente**. Caso α_k seja calculado pela Busca de Armijo para todo $k \in \mathbb{N}$, trata-se do **Método do Gradiente equipado com a Busca de Armijo**. Por fim, se $\alpha_k = \bar{\alpha} > 0$ para todo $k \in \mathbb{N}$ temos o **Método do Gradiente equipado com a busca de comprimento de passo fixo**.

Nosso objetivo é que a sequência $\{x_k\}_{k\in\mathbb{N}}$ gerada pelo método convirja para um ponto estacionário de f, digamos $\bar{x}$, satisfazendo algum critério de parada. Infelizmente, a função objetivo não pode ser arbitrária, ou seja, precisamos que tal função satisfaça algumas condições para que a convergência seja garantida. Além disso, a escolha do chute inicial pode manter ou revogar essa garantia. Veremos o desdobramento dessas questões na próxima seção. Prosseguindo, nas próximas seções apresentaremos e discutiremos os algoritmos em pseudolinguagem do Método do Gradiente equipado com a Busca de Armijo e com passo fixo. Por fim, estudaremos o comportamento da sequência gerado pelo Método do Gradiente.

9.2 Função objetivo

Para que o Método do Gradiente com passo fixo seja eficiente e garanta convergência da sequência gerada, a função objetivo $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ deve satisfazer certas condições. Funções interessantes para esse método geralmente possuem as características enunciadas no Teorema 9.2.1 de convergência. Mas antes, vejamos as definições:

Definição 9.2.1. *Uma função $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ é limitada inferiormente se existir um $m \in \mathbb{R}$ tal que $f(x) \geq m$ para todo $x \in \Omega$*

Definição 9.2.2. *Uma função $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ possui derivada Lipschitz-contínua em Ω com módulo $L > 0$ se*

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \text{ para todo } x, y \in \Omega.$$

Teorema 9.2.1. *Seja $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ limitada inferiormente e derivada Lipschitz-contínua em Ω com módulo $L > 0$. Se a sequência $\{x_k\}_{k\in\mathbb{N}}$ for gerada pelo Método do Gradiente com passo fixo $\alpha_k = \bar{\alpha} > 0$, para todo $k \in \mathbb{N}$, de modo que $\bar{\alpha} < 2/L$, então*

$$\lim_{k\to\infty} \nabla f(x_k) = 0.$$

Isto é, a sequência $\{x_k\}_{k\in\mathbb{N}}$ converge para um ponto estacionário.

Observação 9.2.3. *A prova do teorema acima encontra-se em [3].*

Exemplo 9.2.1. *A função $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x, y) = x^2 + y^2$ é limitada inferiormente, pois como já vimos $f(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}^2$.*

Exemplo 9.2.2. *A função f dada no Exemplo 9.2.1 possui derivada Lipschitz-contínua em $\mathbb{R}^2$ com módulo $L = 2$, pois*

$$\frac{\partial f}{\partial x}(x, y) = 2x \text{ e } \frac{\partial f}{\partial y}(x, y) = 2y.$$

Assim,

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y)\right) = (2x, 2y) = 2(x, y).$$

Portanto,

$$\begin{aligned}
\|\nabla f(x_1, y_1) - \nabla f(x_2, y_2)\| &= \|2(x_1, y_1) - 2(x_2, y_2)\| \\
&= \|2\left((x_1, y_1) - (x_2, y_2)\right)\| \\
&= |2|\|(x_1, y_1) - (x_2, y_2)\| \\
&= 2\|(x_1, y_1) - (x_2, y_2)\| \\
&\leq 2\|(x_1, y_1) - (x_2, y_2)\|.
\end{aligned}$$

Dessa forma, pelo Teorema 9.2.1, segue que a sequência $\{x_k\}_{k\in\mathbb{N}}$ gerada pelo Método do Gradiente com passo fixo, onde $0 < \bar{\alpha} < 2/L = 2/2 = 1$, converge para um ponto estacionário do problema. Basta tomar, nesse caso, $0 < \bar{\alpha} < 1$. Observe ainda que tomar o maior valor possível de $\bar{\alpha}$ não garantirá uma convergência mais rápida (no sentido que veremos na próxima seção) em relação aos valores menores de $\bar{\alpha}$.

Outro detalhe importante é que para a maioria das funções não é uma tarefa fácil encontrar o valor de L. Por isso uma estratégia que pode ser adotada é tomar um $\bar{\alpha}$ suficientemente pequeno. Entretanto, isso pode impactar negativamente a velocidade de convergência da sequência $\{x_k\}_{k\in\mathbb{N}}$ ao ponto estacionário.

Por fim, para que o Método do Gradiente equipado com a Busca de Armijo seja eficiente e garanta convergência da sequência gerada, a função objetivo $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ deve satisfazer certas condições. Funções interessantes para esse método geralmente possuem as seguintes características enunciadas no Teorema 9.2.2 de convergência.

Teorema 9.2.2. *Seja $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável em Ω com derivada contínua. Se a sequência $\{x_k\}_{k\in\mathbb{N}}$ gerada pelo Método do Gradiente equipado com a Busca de Armijo for limitada, então*

$$\lim_{k\to\infty} \nabla f(x_k) = 0.$$

Isto é, a sequência $\{x_k\}_{k\in\mathbb{N}}$ converge para um ponto estacionário.

Observação 9.2.4. *A prova do teorema acima pode ser consultada em [3]*

Para uma função $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ seja diferenciável em Ω basta que, para todo $(x_1, x_2, \ldots, x_n) \in \Omega$, as suas derivadas parciais

$$\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}$$

existam e sejam contínuas no ponto. Ou seja, para todo $a \in \Omega$ temos que

$$\lim_{x\to a} \frac{\partial f}{\partial x_i}(x) = \frac{\partial f}{\partial x_i}(a)$$

para cada $i = 1, 2, \ldots, n$. A definição de sequência limitada encontra-se no Teorema 5.1.5.

Exemplo 9.2.3. *A função $g : \mathbb{R}^2 \to \mathbb{R}$ tal que $g(x,y) = 100(y - x^2)^2 + (x-1)^2$ é diferenciável em $\mathbb{R}^2$, pois*

$$\frac{\partial g}{\partial x}(x,y) = -400x(y - x^2) + 2(x-1) \;\; e \;\; \frac{\partial g}{\partial y}(x,y) = 200(y - x^2)$$

existem para quaisquer $(x,y) \in \mathbb{R}^2$. Além disso, as derivadas parciais da função g são funções polinomiais, logo contínuas. Portanto, se a sequência $(x_k)_{k\in\mathbb{N}} \subset Im(g)$ gerada pelo Método do Gradiente equipado com Busca de Armijo for limitada, então ela converge para um ponto estacionário da função g.

Observação 9.2.5. *A função g apresentada no exemplo anterior trata-se de uma versão particular da função Rosenbrock. Tal função é usada com interesses acadêmicos para testar algoritmos da Otimização Contínua.*

Considerando a função g do exemplo anterior, note que a sequência $\{x_k\}_{k\in\mathbb{N}}$ gerada pelo Método do Gradiente equipado com a Busca de Armijo é tal que $g(x_0) > g(x_1) > \cdots > g(x_k) > \ldots$, isto é, a sequência $\{g(x_k)\}_{k\in\mathbb{N}}$ é decrescente. Ademais, de acordo com o Exercício 9.5.1 a função g é limitada inferiormente. Portanto, pelo Teorema 9.2.2, temos que $\{x_k\}_{k\in\mathbb{N}}$ converge para um ponto estacionário de g.

Tendo em vista os exemplos e as observações já feitas, é teoricamente garantido encontrar os pontos estacionários das funções f e g pelo Método do Gradiente com passo fixo ou com a Busca de Armijo, respectivamente. Na próxima seção apresentaremos em pseudolinguagem o algoritmo que implementa o Método do Gradiente com busca linear.

9.3 Algoritmo do Método do Gradiente

Nesta seção, apresentaremos o algoritmo do Método do Gradiente equipado com uma busca linear. Este algoritmo é baseado na ideia de usar a direção oposta ao gradiente da função objetivo no ponto para encontrar um mínimo local ou global. A busca linear é incorporada para determinar o comprimento do passo em cada iteração, garantindo que o método decresça o valor da função em direção ao mínimo.

A seguir, detalhamos o algoritmo do Método do Gradiente equipado com uma busca linear, descrevendo cada passo e discutindo suas propriedades.

Algoritmo 3: MÉTODO DO GRADIENTE EQUIPADO COM BUSCA LINEAR

1 Tome um ponto inicial $x_0 \in \mathbb{R}^n$ e $\varepsilon > 0$.
2 Faça $k = 0$.
3 Defina $d_k = -\nabla f(x_k)$.
4 Se $\|d_k\| > \varepsilon$, obtenha $\alpha_k > 0$ tal que $f(x_k + \alpha_k d_k) < f(x_k)$.
5 Defina $x_{k+1} = x_k + \alpha_k d_k$.
6 Faça $k = k + 1$ e retorne para o passo 3.

O passo 4 nos mostra que o critério de parada utilizado no algoritmo é baseado na norma do gradiente da função f no ponto x_k. Como vimos no Capítulo 6, esse critério é comum em métodos de otimização, pois quando essa norma é pequena, isso sugere que estamos próximos de um ponto de mínimo (local ou global) da função, já que o gradiente é zero em um ponto crítico.

Além disso, a obtenção de α_k garante que o método decresce o valor da função em cada iteração. Existem várias técnicas para a busca linear, como as apresentadas no capítulo anterior, que podem ser implementadas para encontrar um α_k adequado.

O Método do Gradiente possui uma taxa de convergência linear. Essa taxa ocorre quando a distância entre as aproximações sucessivas diminui linearmente a cada iteração. Isto é,

$$\|x_{k+1} - \bar{x}\| \leq c\|x_k - \bar{x}\|,$$

onde $0 < c < 1$ e $\bar{x}$ é a solução do problema. Em outras palavras, a cada iteração, a aproximação melhora em uma proporção constante em relação à sua distância até a solução.

Outra propriedade importante é que o Método do Gradiente possui convergência global. Isso significa que, começando o processo iterativo com qualquer ponto no domínio da função (desde que a função satisfaça as hipóteses do Teorema 9.2.1 ou 9.2.2), o Algoritmo 3 encontrará uma aproximação do ponto de mínimo da função.

O Método do Gradiente pode ser aplicado a uma vasta gama de problemas de otimização irrestrita, especialmente quando o gradiente da função é de fácil cálculo. Entretanto, sua eficiência pode ser limitada por escolha inadequada de α_k ou pela complexidade da função objetivo.

9.4 Visualização gráfica do Método do Gradiente

A visualização gráfica pode auxiliar na análise e avaliação de algoritmos de otimização. Nesta seção, exploraremos algumas ferramentas de visualização gráfica e discutiremos como elas podem ajudar a compreender o comportamento dos algoritmos em diferentes cenários. Vamos explorar as representações gráficas das funções objetivo apresentadas na Seção 9.2 e a construção da sequência gerada pelo Método do Gradiente.

As representações gráficas das funções objetivo são importantes para compreendermos o comportamento das superfícies a serem otimizadas. Elas nos permitem analisar pontos críticos, como mínimos locais e globais, e entender as características da função em diferentes regiões do domínio. Além disso, as curvas de nível são utilizadas para visualizar o comportamento da função objetivo no espaço bidimensional. Vale ressaltar que essa segunda análise só é possível para funções com várias variáveis. Elas nos permitem identificar padrões, como vales e outros.

Exemplo 9.4.1. *Retomando o Exemplo 9.2.1 podemos analisar a representação gráfica e observar o ponto de mínimo dessa função.*

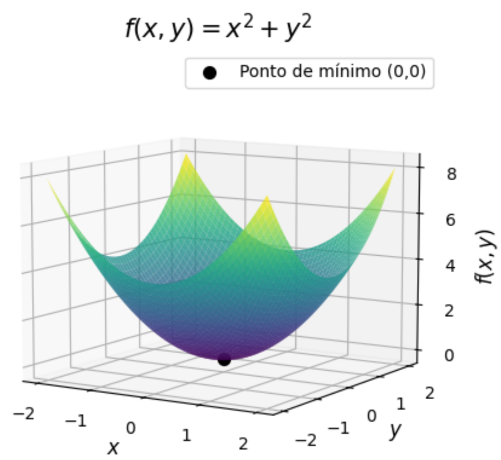


Figura 9.6: Representação gráfica da função do Exemplo 9.2.1.

Exemplo 9.4.2. *Retomando a função de Rosenbrock do Exemplo 9.2.3, temos que essa função tem um comportamento mais complexo de ser analisado que a função anterior.*

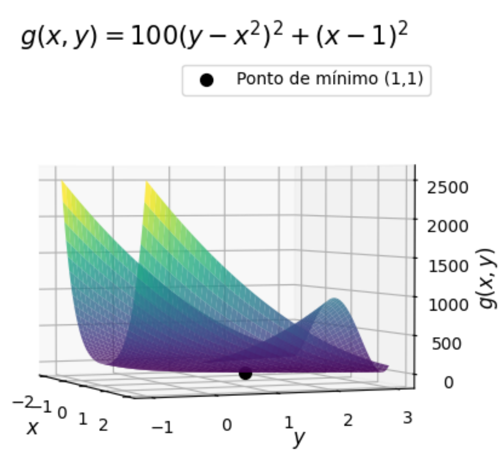


Figura 9.7: Representação gráfica da função de Rosenbrock.

Próximo ao mínimo global, as curvas de nível desta função formam uma estrutura de vale estreito e curvo semelhante a uma parábola, veja na Figura 9.8. Esta característica pode dificultar a convergência das sequências geradas pelos algoritmos de otimização para o ponto de mínimo. Se uma sequência se afastar do mínimo e o algoritmo não possuir garantia de decrescimento, o valor da função pode aumentar gradualmente, dificultando a determinação da aproximação do mínimo. Por outro lado, devido à natureza estreita e curva do vale, os algoritmos de descida do gradiente podem oscilar ou progredir lentamente quando a sequência está próxima do mínimo, uma vez que a norma do gradiente se tornará pequena. Ademais, eles podem exigir comprimentos de passo muito pequenos para permanecerem estáveis, o que pode ocasionar em muitas iterações para a determinação de boas aproximações.

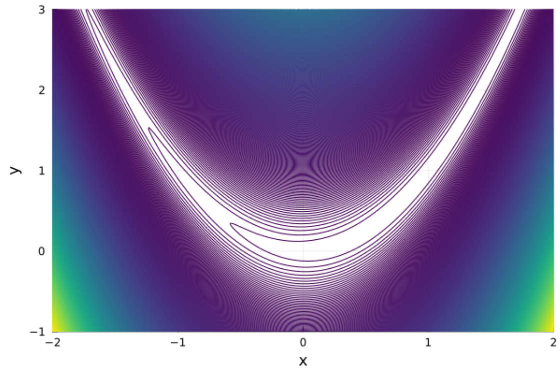


Figura 9.8: Curvas de nível da função de Rosenbrock.

A sequência gerada pelo algoritmo de otimização ao longo de suas iterações nos permite analisar a trajetória em direção ao mínimo e identificar padrões de convergência ou divergência.

Exemplo 9.4.3. *Aplicando o Método do Gradiente equipado com a Busca de Armijo na função de Rosenbrock (Exemplo 9.2.1), teremos que considerar alguns parâmentros. Vamos tomar o comprimento de passo inicial da Busca de Armijo como* $\alpha_0 = 1.0$*, o fator de redução do comprimento de passo como* $\theta = 0.5$*,* $\sigma = 0.0001$*, número máximo de iterações igual a* 5000 *e tolerância para a norma do gradiente como* $\varepsilon = 0.01$*. Podemos analisar o comportamento da sequência gerada pelo método variando esse parâmetros ou variando o ponto inicial. Como vimos anteriormente, o Método do Gradiente possui convergência global, isso significa que podemos tomar* $x_0 \in \mathbb{R}^2$ *qualquer.*

Por exemplo, se tomarmos $x_0 = (3/2, -1/2)$*, o comportamento da sequência será o mostrado na Figura 9.9. Neste caso, foram necessárias* 2838 *iteradas.*

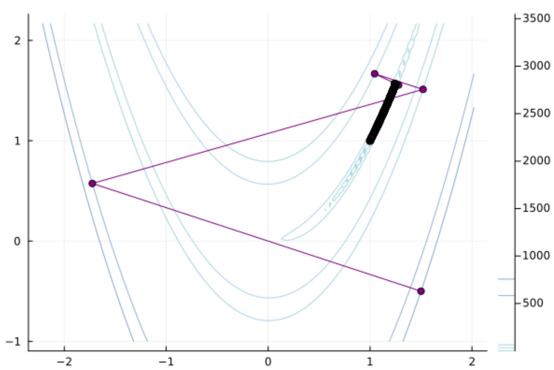


Figura 9.9: Sequência gerada pelo Algoritmo 3 quando $x_0 = (3/2, -1/2)$.

Agora, vamos alterar o ponto inicial tomando um ponto mais distante da solução $x^* = (1, 1)$*, como* $x_{0_2} = (0, -1/2)$*. Para verificar que este ponto inicial é mais distante que o primeiro ponto* $x_{0_1} = (3/2, -1/2)$ *apresentado anteriormente, basta fazermos*

$$\|x_{0_1} - x^*\| = \sqrt{\left(\frac{3}{2} - 1\right)^2 + \left(-\frac{1}{2} - 1\right)^2} = \sqrt{\frac{5}{2}} \approx 1.58 \text{ e } \|x_{0_2} - x^*\| = \sqrt{(0 - 1)^2 + \left(-\frac{1}{2} - 1\right)^2} = \sqrt{\frac{13}{4}} \approx 1.83.$$

Portanto, para o segundo ponto inicial a norma é maior, indicando que ele está mais distante do ponto de mínimo.

O comportamento da sequência e das curvas de nível podem ser vistas na Figura 9.10. Neste caso, foram necessárias 3514 *iteradas.*

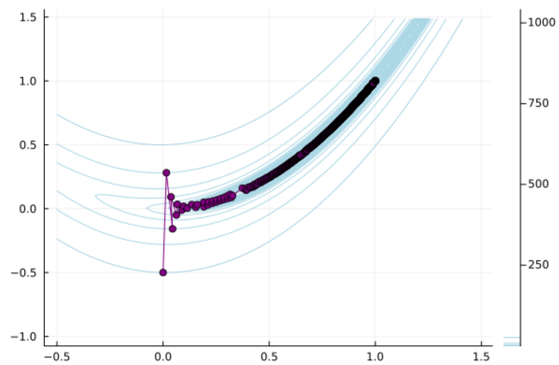


Figura 9.10: Sequência gerada pelo Algoritmo 3 quando $x_0 = (0, -1/2)$.

Note que o fato do ponto inicial ser mais distante do ponto de mínimo implicou que o número de iteradas foi maior, o que discutiremos mais intensivamente na próxima seção.

Da mesma forma que alteramos o ponto inicial, podemos variar os outros parâmetros. Por exemplo, podemos tornar a tolerância da norma do gradiente mais estrita. Para o segundo caso que fizemos anteriormente, se tomarmos $\varepsilon = 0.00000001 = 1 \times 10^{-8}$ *teríamos que alterar o número máximo de iteradas para* 20515 *para que o algoritmo atingisse convergência. Assim, nos estudos computacionais da Otimização Contínua, a variação destes parâmetros constituem uma importante análise numérica uma vez que a escolha dos parâmetros apropriados pode diminuir o número de iteradas, tempo de CPU gasto para resolver o problema e outros fatores.*

9.4.1 Comportamento gerado pelo Método do Gradiente

Para compreendermos o comportamento da sequência gerada pelo Método do Gradiente, é necessário relembrar alguns conceitos do Capítulo 6 e entender os conceitos apresentados a seguir.

Definição 9.4.1. *Dois vetores* $u, v \in \mathbb{R}^n$ *são ortogonais se* $\langle u, v\rangle = 0$. *Denotamos essa relação como* $u \perp v$.

Geometricamente, isso significa que o ângulo entre u e v é de $90°$, ou seja, eles formam um ângulo reto.

Exemplo 9.4.4. *Considere os vetores do* $\mathbb{R}^2$ *dados por* $u = (1, 2)$ *e* $v = (2, -1)$. *Note que eles são ortogonais, pois* $\langle u, v\rangle = 1 \cdot 2 + 2 \cdot (-1) = 2 - 2 = 0$.

Propriedades 9.4.1. *Dados* $u, v, w \in \mathbb{R}^n$*, temos as seguintes propriedades:*

(i) $0 \perp u$*, para todo* $u \in \mathbb{R}^n$*;*
(ii) $u \perp v$ *implica que* $v \perp u$*;*
(iii) *Se* $u \perp v$ *para todo* $v \in \mathbb{R}^n$, *então* $u = 0$*;*
(iv) *Se* $u \perp v$ *e* $w \perp v$, *então* $(u + w) \perp v$*;*
(v) *Se* $u \perp v$ *e* $\lambda \in \mathbb{R}$, *então* $\lambda u \perp v$.

A direção do gradiente no sentido oposto pode oscilar rapidamente ou ziguezaguear durante o processo de descrescimento do valor da função. Este comportamento de zig-zag pode levar um tempo considerável para atingir um ponto próximo ao mínimo. Além disso, quando a sequência gerada pelo método se aproxima de pontos estacionários, o vetor de direção de descida pode ter medida pequena, fazendo com que o método rasteje lentamente para pontos de mínimo. Estes dois problemas - embora não estejam presentes ao minimizar todas as funções - apresentam-se no

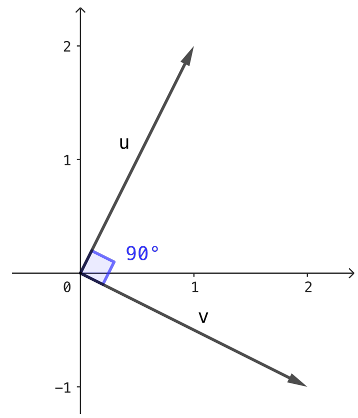


Figura 9.11: Ilustração do Exemplo 9.4.4.

processo iterativo, pois algumas funções que pretendemos minimizar podem ter vales longos e estreitos, e as curvas de nível podem se tornar cada vez mais paralelas.

O que justifica a trajetória de zig-zag do Método do Gradiente para a busca linear exata (que explicaremos a seguir) é a seguinte proposição:

Proposição 9.4.1. *Seja (x_k) uma sequência gerada pelo Algoritmo Método do Gradiente equipado com a busca exata. Então, quaisquer duas direções de descida consecutivas são ortogonais.*

Demonstração: Para a sequência (x_k) gerada pelo Método do Gradiente equipado com a busca exata, a direção de descida no passo i é dada por $d_i = -\nabla f(x_i)$. A atualização da iterada é dada por $x_{i+1} = x_i - \alpha_i \nabla f(x_i)$, onde o comprimento do passo α_i é escolhido de modo a minimizar a função f ao longo da direção d_i. Ou seja, α_i é a solução do problema unidimensional:

$$\alpha_i = \arg\min_{\alpha \geq 0} f(x_i + \alpha d_i) = \arg\min_{\alpha \geq 0} f(x_i - \alpha \nabla f(x_i)).$$

Queremos mostrar que $\nabla f(x_i)^\top \nabla f(x_{i+1}) = 0$. Para isso, definimos a função $q(\alpha) = f(x_i - \alpha \nabla f(x_i))$. O ponto com $\alpha = 0$ é onde começamos, e $\alpha = \alpha_i$ é o passo que escolhemos. A derivada de $q(\alpha)$ em relação a α é dada por:

$$q'(\alpha) = \nabla f(x_i - \alpha \nabla f(x_i))^\top (-\nabla f(x_i)).$$

No ponto $\alpha = \alpha_i$, que minimiza $q(\alpha)$, a derivada $q'(\alpha)$ deve ser zero:

$$q'(\alpha_i) = \nabla f(x_i - \alpha_i \nabla f(x_i))^\top (-\nabla f(x_i)) = \nabla f(x_{i+1})^\top (-\nabla f(x_i)) = 0.$$

Assim, o gradiente no ponto x_{i+1} é ortogonal ao gradiente no sentido oposto no ponto x_i. Portanto, quaisquer duas direções de descida consecutivas no Método do Gradiente com busca exata são ortogonais.

A Figura 9.12 mostra 3 iterações do algoritmo com busca exata aplicado para minimizar uma função quadrática, onde as curvas de nível desta função são elipses.

Note que provamos a Proposição 9.4.1 para a busca exata. Entretanto, essa propriedade vale para qualquer busca linear que o Método do Gradiente esteja equipado, uma vez que a direção se mantém a mesma e variamos apenas a forma de obtenção do comprimento de passo, que influencia apenas na medida do vetor gradiente. Assim, vale a mesma propriedade para a busca de passo fixo e para a Busca de Armijo.

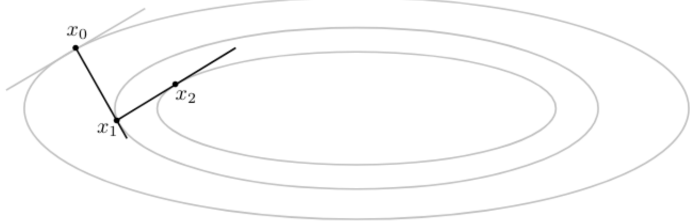


Figura 9.12: Ilustração da Proposição 9.4.1.

9.5 Exercícios

Exercício 9.5.1. *Mostre que a função $g : \mathbb{R}^2 \to \mathbb{R}$ tal que $g(x, y) = 100(y - x^2)^2 + (x - 1)^2$ é limitada inferiormente.*

Exercício 9.5.2. *Mostre que a função $h : \mathbb{R}^2 \to \mathbb{R}$ dada por $h(x, y) = 3x^2 + 3y^2$ satisfaz as hipóteses do Teorema 9.2.1.*

Exercício 9.5.3. *Mostre que a função Dixon-Price dada por $d : \mathbb{R}^2 \to \mathbb{R}$ dada por $d(x, y) = (x - 1)^2 + 2(2y^2 - x)^2$ satisfaz as hipóteses do Teorema 9.2.2.*

Exercício 9.5.4. *Considere a função X, que tem o gradiente Y. Implemente está função e encontre seu mínimo global utilizando o Método do Gradiente.*

Referências

1. Iezzi, Gelson; Murakami, C. Fundamentos da Matemática Elementar, volume 1. São Paulo: Atual, 1999.
2. Iezzi, Gelson; Dolce, Osvaldo; MURAKAMI, CARLOS. Fundamentos de Matemática Elementar, volume 7. São Paulo: Atual, 1997.
3. Izmailov, Alexey; Solodov, Mikhail. Otimização: métodos computacionais, volume 2. Rio de Janeiro: IMPA, 2018.
4. Lima, E. L. Curso de Análise, volume 1, 12º edição. Rio de Janeiro: Projeto Euclides, 2010.
5. Ribeiro, Ademir Alves; Karas, Elizabeth Wegner. Otimização Contínua: aspectos teóricos e computacionais. São Paulo: Cengage Learning, 2013.
6. Stewart, James. Cálculo, volume 1. 7ª ed. Boston: Cengage Learning, 2008. ISBN 978-0495392766.
7. Surjanovic, S. Bingham, D. (2013). Virtual Library of Simulation Experiments: Test Functions and Datasets. Retrieved May 27, 2024, from http://www.sfu.ca/ ssurjano.
8. Oliveira, Ivan de Camargo; Boulos, Paulo. Geometria analítica: um tratamento vetorial, 2º edição. São Paulo: Pearson, 2003.