

Universidade Estadual do Sudoeste da Bahia
Departamento de Ciências Exatas e Tecnológicas

Licenciatura em Matemática

Samara Viriato Vilar Dias

Uma contribuição para determinação de bacias
newtonianas e um estudo introdutório sobre
Otimização Contínua com foco na busca de Armijo

AD PLENAM VITAM

Vitória da Conquista
2024

Samara Viriato Vilar Dias

Uma contribuição para determinação de bacias
newtonianas e um estudo introdutório sobre
Otimização Contínua com foco na busca de Armijo

Monografia apresentada ao Departamento de Ciências Exatas e Tecnológicas da Universidade Estadual do Sudoeste da Bahia, Campus Vitória da Conquista - BA, como requisito parcial à obtenção do título de Licenciada em Matemática, sob orientação do Prof. Dr. Teles Araújo Fernandes.

Vitória da Conquista
2024

Samara Viriato Vilar Dias

Uma contribuição para determinação de bacias
newtonianas e um estudo introdutório sobre
Otimização Contínua com foco na busca de Armijo

Monografia apresentada ao Colegiado do Curso de Matemática como requisito parcial para aprovação na disciplina Seminário de Pesquisa II do Curso de Licenciatura em Matemática.

Trabalho aprovado em 11 de setembro de 2024.

BANCA EXAMINADORA

Prof. Dr. Teles Araújo Fernandes - UESB

Orientador

Prof. Dr. Orizon Pereira Ferreira - UFG

Convidado

Prof. Dr. André Nagamine - UESB

Convidado

Prof. Me. Lucas Venâncio da Silva Santos - UESB

Convidado

Vitória da Conquista

2024

“Pois dele, por ele e para ele são todas as coisas.”

Romanos 11:36a

AGRADECIMENTOS

A Deus, que, ainda na minha adolescência, respondeu às minhas orações, me guiou conforme Sua vontade e me conduziu a um lugar seguro de amor, paz e alegria. És digno, Senhor, de tudo o que tenho e do que sou.

Aos meus pais, Noel e Rosália, e à minha irmã, Láila, que são essenciais em minha vida. À minha avó, Ivete, por seu amor e cuidado, que refletem o amor de Deus por mim, e por tornar meu sonho parte do seu. Aos meus tios, Günter e Moema, por serem tão abençoadores e por todo o amor e suporte fornecidos a mim, e Heleno e Caukep, por me inspirarem em um caminho excelente a trilhar. Agradeço também aos demais familiares que, direta ou indiretamente, prestaram ajuda e apoio ao longo desta caminhada.

Ao professor orientador Teles Araújo Fernandes, por seu acompanhamento desde o primeiro semestre e por, além de ser um excelente profissional, ser um professor que acreditou nos meus estudos e me incentivou a avançar a cada semana. À banca examinadora, por aceitar participar deste momento e por contribuir com este trabalho.

Aos meus amigos Emanuel, Gabriel e Giselle, que tornaram meus dias dos últimos quatro anos mais felizes. Aos meus colegas de turma e a toda a equipe do PETIMAT, que tanto me ensinaram.

Ao Programa de Educação Tutorial Institucional (PETI/UESB), pela bolsa de estudos.

RESUMO

Este trabalho tem como objetivo investigar a resolução de dois problemas: a minimização de funções e a determinação de seus zeros, ambos com ampla aplicação em diversas áreas das ciências. Para a determinação dos zeros de funções, estudamos o Método de Newton e analisamos a convergência da sequência por ele gerada. Na literatura vigente, as bacias de atração do Método de Newton são determinadas apenas por estimativas numéricas. Contudo, neste trabalho, construímos um exemplo de um sistema não linear de equações, determinamos a sequência newtoniana, provamos a convergência dessa sequência para as soluções do sistema e estabelecemos, por meio de uma proposição, as bacias de atração do Método de Newton. Como complemento deste resultado, realizamos uma simulação numérica que evidencia o resultado teórico mencionado. Com relação à minimização de funções, estudamos a busca linear de Armijo, frequentemente utilizada na Otimização Contínua para aumentar a eficiência de métodos numéricos. Como segundo resultado deste trabalho, buscamos elucidar a aplicação da busca de Armijo na minimização irrestrita de uma função do tipo soma de quadrados.

Palavras-chave: Bacias de atração. Busca de Armijo. Método de Newton. Otimização Contínua. Zero de função.

ABSTRACT

This work aims to investigate the resolution of two problems: the minimization of functions and the determination of their zeros, both with widespread application in various areas of science. For determining the zeros of functions, we studied Newton's Method and analyzed the convergence of the sequence it generates. In the current literature, the basins of attraction of Newton's Method are determined only through numerical estimates. However, in this work, we constructed an example of a nonlinear system of equations, determined the Newtonian sequence, proved the convergence of this sequence to the system's solutions, and established, through a proposition, the basins of attraction of Newton's Method. As a complement to this result, we conducted a numerical simulation that highlights the aforementioned theoretical result. Regarding function minimization, we studied Armijo's line search, frequently used in Continuous Optimization to increase the efficiency of numerical methods. As the second result of this work, we seek to elucidate the application of Armijo's line search in the unconstrained minimization of a sum-of-squares function.

Keywords: Basins of attraction. Armijo's search. Newton's method. Continuous Optimization. Zero of function.

Lista de Figuras

2.1	Interpretação geométrica do conjunto de valores admissíveis para a busca de Armijo.	28
3.1	Interpretação geométrica do esquema iterativo (3.3).	31
3.2	Sequência gerada pelo Método de Newton para $x_0 = 11/10$	32
3.3	Sequência gerada pelo Método de Newton para $x_0 = 9/10$	33
4.1	Curvas de nível e representação gráfica da função f	39
4.2	Conjunto dos comprimentos de passo admissíveis.	40
4.3	Curvas de nível de f em $3/2$ e $691/1250$	40
4.4	Curvas de nível e representação gráfica da função f	41
4.5	Curvas que representam as equações do sistema (4.3).	42
4.6	Bacias de atração do Método de Newton para o problema (4.3).	45

Sumário

Introdução	7
1 Resultados preliminares	10
1.1 Noções de Álgebra Linear	10
1.2 Noções de Análise em Espaços Euclidianos	15
1.2.1 Continuidade e diferenciabilidade de funções	19
2 Aspectos da Otimização Contínua	23
2.1 Definições básicas e resultados auxiliares	23
2.2 A busca linear de Armijo	27
3 Método de Newton	29
3.1 O Método de Newton	29
3.2 Análise de convergência	33
4 Aplicações	38
4.1 Minimização irrestrita com a busca de Armijo	38
4.2 As bacias de atração do Método de Newton para um sistema não linear de equações	42
Considerações finais	46
A Implementações	50
A.1 Método de Newton	50

Introdução

Encontrar zeros de funções, isto é, determinar um elemento x no domínio de uma função F tal que $F(x) = 0$, é um dos problemas matemáticos mais antigos e que aparece em diversos contextos, tais como: encontrar raízes de polinômios, determinar as soluções de uma equação, estudar o comportamento de funções e modelar fenômenos da natureza. Resultados para a resolução desse problema são abordados já no Ensino Fundamental e foram propostos ao longo da história da Matemática.

Desde 1545, com o compilado de trabalhos relacionados às soluções de equações quadráticas feito por Girolamo Cardano, conforme documentado em [5], estudiosos têm se dedicado ao desenvolvimento de teorias para encontrar os zeros de funções quaisquer. Conforme Lima, [18], essas investigações apresentaram resultados expressivos ainda em 1545, quando matemáticos como S. Ferro, G. Cardano, N. Tartaglia e L. Ferrari desenvolveram métodos para determinar as raízes de polinômios de terceiro e quarto grau, utilizando a resolução por radicais. Em [9], Humes destaca que, no século XIX, E. Galois concluiu que o problema de encontrar a raiz de um polinômio de grau maior ou igual a cinco não pode ser resolvido por radicais. Além disso, Humes acrescenta que, nesses casos, são necessários outros métodos para determinar as raízes reais desses polinômios, caso elas existam. Neste contexto, surge o método de Newton, que é conhecido por ser uma ferramenta poderosa para encontrar zeros de funções.

O Método de Newton foi originalmente proposto por Isaac Newton no artigo intitulado *De analysi per aequationes numero terminorum infinitas*, [20]. Posteriormente, outra versão do Método de Newton foi apresentada em [21]. Nessas primeiras versões, Newton aplicou o método exclusivamente a polinômios, calculando explicitamente uma sequência de polinômios em vez de utilizar aproximações sucessivas, o que caracterizava o método como puramente algébrico. Somente em 1690, com as contribuições apresentadas por Joseph Raphson em [25] é que o método foi reformulado por ele em termos de aproximações sucessivas para encontrar as raízes de polinômios, tornando evidente sua relação com o Cálculo. Isto fez com que o Método de Newton também ficasse conhecido como Método de Newton-Raphson. Em 1740, Thomas Simpson desempenhou um papel crucial ao avançar ainda mais o método, considerando-o não apenas como uma ferramenta para encontrar raízes de polinômios, mas como uma técnica iterativa para resolver sistemas de equações não lineares, restrito ao caso de duas equações, conforme descrito em [27]. A partir disso,

o Método de Newton foi desenvolvido para encontrar zeros de funções diferenciáveis de várias variáveis, veja [23, Cap. 7, Sec. 7.1]. Para uma análise mais detalhada sobre a evolução do Método de Newton, veja [30]. É importante salientar que o Método de Newton já foi desenvolvido em contextos não lineares, por exemplo, em variedades Riemannianas, veja [6].

O Método de Newton é um método iterativo, o que significa que o seu princípio fundamental consiste em determinar, a partir de um ponto inicial, uma sequência de pontos que devem convergir para a solução do problema. Esse processo envolve a resolução de um sistema linear para determinar a direção de Newton, e a soma dessa direção com a iterada anterior resulta na nova iteração. O resultado principal desse método também nos diz que ele é um método local com taxa de convergência rápida, o que significa que a solução para a qual a sequência gerada por esse método converge depende da boa escolha do ponto inicial e, quando converge, possui uma taxa de convergência superlinear ou quadrática. As regiões nas quais o ponto inicial pode ser tomado, garantindo a convergência da sequência, são chamadas bacias de atração e têm sido determinadas apenas por estimativas numéricas, veja [7, 28].

Nesse sentido, a literatura vigente revela uma necessidade de investigação sobre as potencialidades do Método de Newton, especialmente no que tange à determinação das bacias de atração obtidas por meio da aplicação deste método. Neste trabalho, analisamos a convergência do Método de Newton para encontrar zeros de funções diferenciáveis e abordamos como contribuição a determinação das bacias de atração do Método de Newton para a resolução de um sistema não linear de equações, resultado que foi apresentado na 76^a Reunião Anual da Sociedade Brasileira para o Progresso da Ciência (SBPC), [3].

A segunda aplicação deste trabalho está no contexto da Otimização Contínua. Nesta área, busca-se determinar zeros de funções por meio da minimização ou maximização de funções objetivo. Assim, são desenvolvidas teorias para assegurar sequências que convergem para os extremos de funções. Em geral, para obter essas sequências é adotado o seguinte procedimento: toma-se um ponto no domínio da função, um vetor diretor a partir deste ponto, e a nova iterada da sequência é a soma do ponto com o vetor diretor. Em alguns casos, a determinação da nova iterada pode ser mais eficiente se for dada como a soma do ponto com uma contração do vetor diretor. Para obter essa contração, é necessário um comprimento de passo que ajuste o vetor diretor, e a busca de Armijo é uma ferramenta muito conhecida para essa estratégia.

A busca linear de Armijo foi introduzida por Louis Armijo em 1966 no artigo intitulado *Minimization of Functions Having Lipschitz-Continuous First Partial Derivatives*, [2]. Seu objetivo principal era determinar de forma eficiente um comprimento de passo adequado ao longo de uma direção de descida. Assim, a busca de Armijo assegura a existência de um comprimento de passo que forneça um decrescimento no valor da função objetivo. Esta abordagem é amplamente reconhecida na literatura e motivou o desenvol-

vimento de outras buscas lineares inexatas, como aquelas propostas por Wolfe, Goldstein e Grippo-Lampariello-Lucidi, veja [24]. Dessa forma, neste trabalho, buscamos elucidar a aplicação da busca de Armijo para a minimização irrestrita de uma função objetivo.

O presente trabalho está dividido em quatro capítulos. Com o objetivo de torná-lo mais completo, apresentamos no Capítulo 1 os principais resultados de Álgebra Linear e Análise em espaços euclidianos, que serão utilizados nos demais capítulos. Em seguida, o Capítulo 2 aborda aspectos relacionados à Otimização Contínua, com foco na minimização irrestrita de funções e na busca de Armijo. O Método de Newton para encontrar zeros de funções continuamente diferenciáveis e a demonstração de sua convergência local são discutidos no Capítulo 3. O Capítulo 4 mostra os resultados obtidos em duas aplicações: a busca de Armijo na minimização de uma função do tipo soma de quadrados e a determinação das bacias de atração do Método de Newton para um sistema não linear de equações. Por fim, apresentamos nossas considerações finais, perspectivas futuras em relação a este trabalho e, no Apêndice A, as implementações realizadas.

Capítulo 1

Resultados preliminares

O estudo dos teoremas clássicos da Álgebra Linear e da Análise em espaços euclidianos é fundamental para este trabalho, pois serve de base para todos os resultados apresentados nos demais capítulos. Portanto, apresentaremos neste capítulo um estudo introdutório, sendo importante que o leitor tenha familiaridade com conceitos de topologia do espaço \mathbb{R}^n , tais como completude da reta, cota superior/inferior, supremo e ínfimo, entre outros, os quais podem ser encontrados em [4, 16, 17].

1.1 Noções de Álgebra Linear

Entre os temas centrais da Álgebra Linear, destaca-se o estudo dos operadores lineares, que são funções que preservam as propriedades dos espaços vetoriais. Neste trabalho, estamos interessados em encontrar zeros de funções diferenciáveis definidas no espaço vetorial conhecido como espaço euclidiano n -dimensional, o que fundamenta a abordagem adotada nesta seção. Além disso, abordaremos as propriedades topológicas do espaço dos operadores lineares, incluindo a definição e a utilização da norma apropriada.

Inicialmente, denotaremos $\mathbb{R}^{m \times n}$ o conjunto das matrizes reais com m linhas e n colunas. Dada uma matriz $A \in \mathbb{R}^{m \times n}$, denotaremos a matriz transposta de A por A^\top . Sabemos que cada matriz pertencente a $\mathbb{R}^{m \times n}$ pode ser considerada como um vetor do espaço euclidiano \mathbb{R}^{mn} , basta escrevermos as colunas, uma após a outra, como um vetor linha. Assim, quando for conveniente, utilizaremos a matriz coluna $A \in \mathbb{R}^{m \times 1}$ para representar um vetor $x \in \mathbb{R}^m$. Nesse contexto, o transposto do vetor x , denotado por x^\top , é uma matriz linha.

O produto de $A \in \mathbb{R}^{m \times n}$ por $B \in \mathbb{R}^{n \times p}$ é uma matriz $C \in \mathbb{R}^{m \times p}$ onde cada elemento c_{ij} é definido pela soma dos produtos dos elementos correspondentes das linhas de A e das colunas de B . Formalmente, temos $C = A \cdot B$, em que $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$ para $i = 1, \dots, m$ e $j = 1, \dots, p$. Outra propriedade das matrizes é a inversibilidade, que definiremos a seguir.

Definição 1.1.1. Uma matriz $A \in \mathbb{R}^{n \times n}$ é dita invertível se existe uma matriz $B \in \mathbb{R}^{n \times n}$ tal que $A \cdot B = B \cdot A = I$, onde $I \in \mathbb{R}^{n \times n}$ é a matriz identidade. A matriz B é dita inversa de A e denotamos $B = A^{-1}$. Se A não tem inversa, dizemos que A é não invertível.

Além das propriedades do espaço de matrizes, a noção de produto interno e de norma de vetor é fundamental na Álgebra Linear, pois nos permite analisar ângulos entre vetores e definir a norma induzida pelo produto interno, que apresentaremos posteriormente. Assim, dado V um espaço vetorial, definimos o produto interno da seguinte forma:

Definição 1.1.2. Uma função $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ é dita produto interno sobre V se satisfaz as seguintes propriedades para quaisquer $x, y, z \in V$ e $\alpha \in \mathbb{R}$:

P1. $\langle x, x \rangle \geq 0$ e $\langle x, x \rangle = 0$ se, e somente se, $x = 0$;

P2. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$;

P3. $\langle \alpha x, y \rangle = \alpha \cdot \langle x, y \rangle$;

P4. $\langle x, y \rangle = \langle y, x \rangle$.

No espaço euclidiano n -dimensional, consideremos dois vetores $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$. Um exemplo de produto interno neste espaço é o produto interno usual, definido por $\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^\top y$.

Definição 1.1.3. Uma função $\| \cdot \| : V \rightarrow \mathbb{R}$ é dita norma em um espaço vetorial V se satisfaz as seguintes propriedades para quaisquer $x, y \in V$ e $\alpha \in \mathbb{R}$:

N1. $\|x\| \geq 0$ e $\|x\| = 0$ se, e somente se, $x = 0$;

N2. $\|\alpha x\| = |\alpha| \|x\|$, onde $|\cdot|$ denota o valor absoluto;

N3. $\|x + y\| \leq \|x\| + \|y\|$ (desigualdade triangular).

No espaço euclidiano n -dimensional, podemos encontrar diversas normas importantes. A seguir, apresentamos três exemplos:

Exemplo 1.1.4. Dado $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, a norma $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $\|x\| = \sqrt{\langle x, x \rangle}$ é denominada norma induzida pelo produto interno $\langle \cdot, \cdot \rangle$. Em particular, a norma induzida pelo produto interno usual é denominada norma euclidiana e denotaremos, por simplicidade, como $\| \cdot \|$.

Exemplo 1.1.5. Dado $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, a norma $\| \cdot \|_m : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $\|x\|_m = \max_{1 \leq i \leq n} |x_i|$ é denominada norma do máximo.

Exemplo 1.1.6. Dado $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, a norma $\| \cdot \|_s : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $\|x\|_s = \sum_{i=1}^n |x_i|$ é denominada norma da soma.

Duas normas quaisquer $\|\cdot\|_1$ e $\|\cdot\|_2$ do espaço vetorial V são ditas equivalentes quando existem constantes a e b positivas tais que, para todo $x \in V$, vale $a\|x\|_2 \leq \|x\|_1 \leq b\|x\|_2$. Considerando o \mathbb{R}^n , o resultado que pode ser encontrado em [17, Teo. 8] nos garante que duas normas quaisquer deste espaço são equivalentes. Assim, podemos agora definir operadores lineares e apresentar suas propriedades de interesse para este trabalho.

Definição 1.1.7. *Sejam V e W dois espaços vetoriais. Dizemos que $T : V \rightarrow W$ é um operador linear se satisfaz as seguintes propriedades para quaisquer $x, y \in V$ e $\alpha \in \mathbb{R}$:*

O1. $T(x + y) = T(x) + T(y)$;

O2. $T(\alpha x) = \alpha T(x)$.

Se considerarmos os espaços vetoriais \mathbb{R}^n e \mathbb{R}^m , o conjunto de todos os operadores lineares $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$, denotado por $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$, é um espaço vetorial. Este resultado segue das seguintes definições: se $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ é um operador linear, então $T(0) = 0$; dados $v_1, \dots, v_n \in \mathbb{R}^n$ e $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, vale $T(\alpha_1 v_1 + \dots + \alpha_n v_n) = \alpha_1 T(v_1) + \dots + \alpha_n T(v_n)$; a soma de dois operadores lineares $T, P : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e o produto de T por um escalar $\alpha \in \mathbb{R}$ são os operadores lineares $T + P : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $\alpha T : \mathbb{R}^n \rightarrow \mathbb{R}^m$, respectivamente. Eles são dados por $(T + P)(v) = T(v) + P(v)$ e $(\alpha T)(v) = \alpha T(v)$, para todo $v \in \mathbb{R}^n$.

Um resultado muito importante da Álgebra Linear nos diz que, fixando as bases canônicas de \mathbb{R}^n e \mathbb{R}^m , é possível encontrar um isomorfismo entre $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ e o conjunto $\mathbb{R}^{m \times n}$. Este resultado é provado em [14, Cap. 4, Teo. 3.3] e nos garante que podemos associar cada operador linear $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ a uma matriz de $\mathbb{R}^{m \times n}$ em relação às bases canônicas de \mathbb{R}^n e \mathbb{R}^m . A partir disso, as operações com operadores lineares se tornam operações matriciais. Outro conceito importante é o de autovalor e autovetor de um operador linear ou da matriz que o representa, que é definido a seguir.

Definição 1.1.8. *Dizemos que o número real λ é um autovalor da matriz $A \in \mathbb{R}^{n \times n}$ quando λ é um autovalor do operador $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ associado à matriz A em relação à base canônica de \mathbb{R}^n . Isto significa que existe um vetor $x \in \mathbb{R}^n$ não nulo tal que $T(x) = \lambda x$ ou, de forma equivalente, uma matriz $x \in \mathbb{R}^{n \times 1}$ não nula tal que $A \cdot x = \lambda x$. Dizemos também que x é o autovetor associado ao autovalor λ .*

Além disso, o polinômio característico de A é, por definição, $p_A(\lambda) = \det(A - \lambda I)$, onde I é a matriz identidade de mesma ordem que A . Segundo [14, Cap. 8, Teo. 2.2], para $\lambda \in \mathbb{R}$ ser um autovalor de A é necessário e suficiente que $p_A(\lambda) = 0$, de onde concluímos que A possui no máximo n autovalores. Como apresentado a seguir, as propriedades de autovalores se relacionam diretamente à análise de matrizes simétricas, isto é, matrizes do tipo $A \in \mathbb{R}^{n \times n}$ em que $A = A^\top$.

Lema 1.1.9. *Seja $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica com λ_1 e λ_n sendo o menor e o maior autovalor de A , respectivamente. Então $\lambda_1 \|x\|^2 \leq x^\top A x \leq \lambda_n \|x\|^2$, para todo $x \in \mathbb{R}^n$.*

Demonstração. Sabemos que a matriz A é simétrica, então é possível encontrar um conjunto ortonormal $\{x_1, x_2, \dots, x_n\}$ de autovetores de A . Qualquer vetor $x \in \mathbb{R}^n$ não nulo pode ser escrito como $x = \sum_{i=1}^n \alpha_i x_i$, onde $\alpha_i \in \mathbb{R}$ é um escalar apropriado. Usando a ortonormalidade dos autovetores, obtemos $\|x_i\| = 1$, para todo $i = 1, \dots, n$. Agora, usando a propriedade de que $Ax_i = \lambda_i x_i$, onde λ_i é o autovalor de A associado a x_i para todo $i = 1, \dots, n$, temos $x^\top Ax = \sum_{i=1}^n \lambda_i \alpha_i^2 \|x_i\|^2 = \sum_{i=1}^n \lambda_i \alpha_i^2$ e $\|x\|^2 = \sum_{i=1}^n \alpha_i^2 \|x_i\|^2 = \sum_{i=1}^n \alpha_i^2$. Observando que $\lambda_1 \leq \lambda_i \leq \lambda_n$ para todo $i = 1, \dots, n$, temos

$$\lambda_1 \|x\|^2 = \lambda_1 \sum_{i=1}^n \alpha_i^2 = \sum_{i=1}^n \lambda_1 \alpha_i^2 \leq \sum_{i=1}^n \lambda_i \alpha_i^2 \leq \sum_{i=1}^n \lambda_n \alpha_i^2 = \lambda_n \sum_{i=1}^n \alpha_i^2 = \lambda_n \|x\|^2.$$

Portanto, $\lambda_1 \|x\|^2 \leq x^\top Ax \leq \lambda_n \|x\|^2$. □

Definição 1.1.10. *Sejam $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica e $x \in \mathbb{R}^n$ não nulo. Dizemos que A é definida positiva se, para todo x , temos $x^\top Ax > 0$. Dizemos que A é semidefinida positiva se, para todo x , temos $x^\top Ax \geq 0$. Analogamente, A é definida negativa se, para todo x , temos $x^\top Ax < 0$, e A é semidefinida negativa se, para todo x , temos $x^\top Ax \leq 0$.*

O Lema 1.1.11 a seguir apresenta uma equivalência dessa classificação e, por meio dele, concluímos que para analisar se uma matriz é definida positiva, basta analisar o sinal dos seus autovalores.

Lema 1.1.11. *Uma matriz simétrica $A \in \mathbb{R}^{n \times n}$ é definida positiva se, e somente se, todos os seus autovalores são positivos.*

Demonstração. Suponha que A é definida positiva. Se λ é um autovalor de A com autovetor associado $x \neq 0$, então $0 < x^\top Ax = x^\top \lambda x = \lambda x^\top x = \lambda \|x\|^2$. Como $\|x\|^2 > 0$, temos $\lambda > 0$. Reciprocamente, suponha que todos os autovalores de A sejam positivos. Como vimos no Lema 1.1.9, qualquer vetor $x \in \mathbb{R}^n$ não nulo pode ser escrito como $x = \sum_{i=1}^n \alpha_i x_i$, onde $\alpha_i \in \mathbb{R}$ é um escalar apropriado. Usando a propriedade de que $Ax_i = \lambda_i x_i$, onde λ_i é o autovalor de A associado a x_i para todo $i = 1, \dots, n$, e a ortogonalidade dos autovetores, obtemos $x^\top Ax = \sum_{i=1}^n \alpha_i^2 \lambda_i \|x_i\|^2 > 0$. Portanto, A é definida positiva. □

No Exemplo 1.1.12, determinaremos os autovalores de um operador linear associado a uma matriz e a classificaremos de acordo com os resultados apresentados.

Exemplo 1.1.12. *Considere a função $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $T(x_1, x_2) = (x_1, 2x_2)$. Essa função é um operador linear, pois a função $T(x_1, x_2)$ pode ser descrita como a combinação linear dos vetores $(1, 0)$ e $(0, 2)$ com coeficientes x_1 e x_2 . Esse operador leva cada vetor $x \in \mathbb{R}^2$ no vetor $T(x) \in \mathbb{R}^2$. A matriz de T relativa à base canônica do \mathbb{R}^2 é dada por*

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

Para obter os autovalores de A , vamos encontrar as raízes do seu polinômio característico:

$$p_A(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 0 \\ 0 & 2 - \lambda \end{vmatrix} = (1 - \lambda)(2 - \lambda) = 0 \quad (1.1)$$

Resolvendo a equação (1.1), temos as soluções $\lambda = 1$ e $\lambda = 2$. Estes são os autovalores da matriz A . Note que A é uma matriz simétrica, então pelo Lema 1.1.11, é uma matriz definida positiva e, conseqüentemente, é semidefinida positiva.

Por fim, vamos definir a norma de um operador linear T como sendo a função $\|\cdot\| : \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \rightarrow \mathbb{R}$ dada por

$$\|T\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|T(x)\|}{\|x\|}.$$

A seguir, vamos verificar se esta é de fato uma norma. Inicialmente, temos que se $T \neq 0$ é um operador linear, existe $x \in \mathbb{R}^n \setminus \{0\}$, tal que $T(x) \neq 0$, então $\|T(x)\|/\|x\| > 0$. Usando a definição de supremo, podemos concluir que $\|T\| \geq 0$. Agora, se $\|T\| = 0$, temos $T(x) = 0$ para todo $x \in \mathbb{R}^n \setminus \{0\}$, que implica $T = 0$. Reciprocamente, se T é o operador nulo, então $\|0\| = 0$. Para a segunda propriedade, consideremos T um operador linear, $x \in \mathbb{R}^n \setminus \{0\}$ e $\alpha \in \mathbb{R}$, então

$$\|\alpha T\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|\alpha T(x)\|}{\|x\|} = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{|\alpha| \|T(x)\|}{\|x\|} = |\alpha| \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|T(x)\|}{\|x\|} = |\alpha| \|T\|.$$

Por fim, para a terceira propriedade da norma, se T e P são operadores lineares e $x \in \mathbb{R}^n \setminus \{0\}$, usando a desigualdade triangular da norma de vetores, temos

$$\|T + P\| = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|(T + P)(x)\|}{\|x\|} \leq \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|T(x)\|}{\|x\|} + \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|P(x)\|}{\|x\|} = \|T\| + \|P\|.$$

Além disso, o operador T é dito limitado se existir um número real c tal que $\|T(x)\| \leq c\|x\|$, para todo $x \in \mathbb{R}^n$ não nulo. Note que, se tomarmos $c = \|T\|$, obtemos

$$\|T(x)\| \leq \|T\| \|x\|. \quad (1.2)$$

Outras propriedades relacionadas à norma do operador podem ser encontradas em [13, Cap. 2]. Na próxima seção, discutiremos um exemplo importante de operador linear: a derivada de uma função. Em seguida, apresentaremos a matriz jacobiana que representa esse operador linear.

1.2 Noções de Análise em Espaços Euclidianos

Nesta seção, apresentaremos as definições e os resultados da Análise no espaço euclidiano n -dimensional que auxiliarão na demonstração de resultados posteriores deste trabalho. As definições e demonstrações mais básicas poderão ser encontradas em [4, 16, 17].

Inicialmente, sejam dados o ponto $a \in \mathbb{R}^n$ e o número real $\delta > 0$. O conjunto dos pontos $x \in \mathbb{R}^n$ cuja distância ao ponto a é menor do que δ é dito bola aberta, isto é, $B(a; \delta) = \{x \in \mathbb{R}^n : \|x - a\| < \delta\}$. Analogamente, temos o conjunto chamado de bola fechada, com centro a e raio δ , definido por $B[a; \delta] = \{x \in \mathbb{R}^n : \|x - a\| \leq \delta\}$. Um conjunto $U \subset \mathbb{R}^n$ é dito aberto quando para cada $a \in U$ existe $\delta > 0$ tal que $B(a; \delta) \subset U$, isto é, quando todos os seus pontos são interiores. De forma similar, um conjunto $U \subset \mathbb{R}^n$ é fechado quando contém todos os seus pontos de aderência, isto é, para todo ponto $a \in \mathbb{R}^n$, onde qualquer vizinhança de a contém algum elemento de U , então $a \in U$. Dizemos que um conjunto $U \subset \mathbb{R}^n$ é limitado quando existe um número real $c > 0$ tal que $|a| \leq c$ para todo $a \in U$. Além disso, dizemos que um conjunto $U \subset \mathbb{R}^n$ é compacto quando ele é limitado e fechado.

Esses conceitos são fundamentais para entender a topologia dos conjuntos no espaço euclidiano e são particularmente úteis ao lidar com sequências. A seguir, definimos formalmente o que é uma sequência e discutimos suas propriedades no contexto dos números reais, estendendo posteriormente essas definições para o espaço \mathbb{R}^n . Por conveniência, consideramos, neste estudo, que $0 \in \mathbb{N}$.

Definição 1.2.1. *Uma sequência de números reais é uma função $x : \mathbb{N} \rightarrow \mathbb{R}$ que associa a cada k natural um número real $x(k)$. Denotaremos o k -ésimo termo $x(k)$ por x_k e a sequência cujo k -ésimo termo é x_k por $\{x_k\}$ ou $\{x_0; x_1; \dots; x_k; \dots\}$.*

Definição 1.2.2. *Uma sequência $\{x_k\} \subset \mathbb{R}$ é dita limitada superiormente se existir $c \in \mathbb{R}$ tal que $x_k \leq c$ para todo $k \in \mathbb{N}$. Da mesma forma, a sequência $\{x_k\}$ é dita limitada inferiormente se existir $c \in \mathbb{R}$ tal que $x_k \geq c$ para todo $k \in \mathbb{N}$.*

Definição 1.2.3. *Uma sequência $\{x_k\} \subset \mathbb{R}$ é dita limitada se existir um número real $c > 0$ tal que $|x_k| \leq c$ para todo $k \in \mathbb{N}$.*

Exemplo 1.2.4. *Considere a sequência de números reais dada por $x_k = 1/k$ para todo $k \geq 1$. Temos que $\{x_k\}$ é limitada inferiormente por zero e superiormente por 1. Logo, considerando $c = 1$, sabemos que $|x_k| \leq 1$ para todo $k \geq 1$, isto é, $\{x_k\}$ é uma sequência limitada.*

Definição 1.2.5. *Uma sequência $\{x_k\} \subset \mathbb{R}$ é dita crescente quando, para todo $k \in \mathbb{N}$, temos $x_k < x_{k+1}$. No caso da desigualdade não ser estrita, isto é, quando para todo $k \in \mathbb{N}$ temos $x_k \leq x_{k+1}$, a sequência é denominada não-decrescente. Analogamente, se para todo $k \in \mathbb{N}$ temos $x_k > x_{k+1}$, a sequência é decrescente, e se $x_k \geq x_{k+1}$, é chamada não-crescente. Em todos estes casos, a sequência é dita monótona.*

Exemplo 1.2.6. Considere a sequência de números reais dada por $x_k = k$ para todo $k \geq 1$. Temos que $\{x_k\}$ é limitada inferiormente por 1, ilimitada superiormente e monótona crescente.

Definição 1.2.7. Diz-se que uma sequência $\{x_k\} \subset \mathbb{R}$ converge para $a \in \mathbb{R}$ quando, para todo $\varepsilon > 0$ dado, é possível obter $m \in \mathbb{N}$ tal que $k > m$ implica $|x_k - a| < \varepsilon$. Neste caso, denotaremos por $\lim_{k \rightarrow \infty} x_k = a$ ou dizemos que $x_k \rightarrow a$ quando $k \rightarrow \infty$.

As operações com limites de sequências podem ser consultadas em [16, Cap. 4, Sec. 3, Teo. 6]. Ademais, um resultado que relaciona as três últimas definições é o Teorema 1.2.9. Antes de demonstrar este teorema, precisamos provar o resultado a seguir.

Teorema 1.2.8. Se uma sequência $\{x_k\}$ de números reais é convergente, então é limitada.

Demonstração. Suponha que $\lim_{k \rightarrow \infty} x_k = a$ e seja $\varepsilon = 1$. Então, existe um número natural m tal que $|x_k - a| < 1$ para todo $k \geq m$. Aplicando a desigualdade triangular para $k \geq m$, temos $|x_k| = |x_k - a + a| \leq |x_k - a| + |a| < 1 + |a|$. Se tomarmos $M = \sup\{|x_0|, |x_1|, \dots, |x_{m-1}|, 1 + |a|\}$, temos $M \geq |x_k|$ para todo $k \in \mathbb{N}$. Logo, a sequência $\{x_k\}$ é limitada. \square

Teorema 1.2.9. Uma sequência $\{x_k\}$ de números reais monótona é convergente se, e somente se, for limitada. Além disso:

- (a) Se $\{x_k\}$ for uma sequência crescente limitada, então $\lim_{k \rightarrow \infty} x_k = \sup\{x_k : k \in \mathbb{N}\}$.
- (b) Se $\{x_k\}$ for uma sequência decrescente limitada, então $\lim_{k \rightarrow \infty} x_k = \inf\{x_k : k \in \mathbb{N}\}$.

Demonstração. Pelo Teorema 1.2.8, sabemos que toda sequência convergente é limitada. Assim, basta provarmos que uma sequência monótona e limitada é convergente. Consideremos os seguintes casos:

- (a) Seja $\{x_k\}$ uma sequência crescente limitada. Como $\{x_k\}$ é limitada, sabemos que $\{x_k : k \in \mathbb{N}\}$ é um conjunto não vazio e limitado superiormente. Portanto, existe $s = \sup\{x_k : k \in \mathbb{N}\} \in \mathbb{R}$. Queremos mostrar que $\lim_{k \rightarrow \infty} x_k = s$. Dado $\varepsilon > 0$, temos que $s - \varepsilon$ não é uma cota superior de $\{x_k : k \in \mathbb{N}\}$, isto é, existe $k' \in \mathbb{N}$ tal que $s - \varepsilon < x_{k'}$. Como $\{x_k\}$ é crescente, temos $x_{k'} \leq x_k$ para todo $k \geq k'$, então $s - \varepsilon < x_{k'} \leq x_k \leq s < s + \varepsilon$, para todo $k \geq k'$. Portanto, $|x_k - s| < \varepsilon$ para todo $k \geq k'$. Como $\varepsilon > 0$ é arbitrário, concluímos que $\lim_{k \rightarrow \infty} x_k = s$.
- (b) A demonstração é análoga ao caso anterior, considerando a sequência decrescente limitada e seu ínfimo. \square

Outro conceito fundamental no estudo de sequências é o de subsequência:

Definição 1.2.10. Uma subsequência da sequência $\{x_k\} \subset \mathbb{R}$ é a restrição desta sequência a um subconjunto infinito $\mathbb{N}' = \{k_0 < k_1 < \dots < k_n < \dots\} \subset \mathbb{N}$, e é indicada pela notação $\{x_{k_j}\}$, com $k_j \in \mathbb{N}'$, ou $\{x_{k_0}; x_{k_1}; \dots; x_{k_n}; \dots\}$.

Exemplo 1.2.11. Considere a sequência de números reais dada por $x_k = (-1)^k$. Note que essa sequência possui duas subsequências: tomando $\mathbb{N}_p = \{k_j \in \mathbb{N} : k_j = 2j, j \in \mathbb{N}\}$, temos a subsequência dos termos com índices pares $\{x_{k_j}\} = \{1; 1; \dots\}$, com $k_j \in \mathbb{N}_p$; tomando $\mathbb{N}_i = \{k_j \in \mathbb{N} : k_j = 2j + 1, j \in \mathbb{N}\}$, temos a subsequência dos termos com índices ímpares $\{x_{k_j}\} = \{-1; -1; \dots\}$, com $k_j \in \mathbb{N}_i$.

Agora, vamos estender a definição de sequências de números reais para sequências em \mathbb{R}^n da seguinte forma: uma sequência de pontos em \mathbb{R}^n é uma função $x: \mathbb{N} \rightarrow \mathbb{R}^n$ que associa a cada número natural k um ponto $x_k \in \mathbb{R}^n$. Uma sequência $\{x_k\} \subset \mathbb{R}^n$ representa n sequências $\{x_{(k)i}\}$, com $i = 1, \dots, n$, onde cada $\{x_{(k)i}\}$ é uma sequência de números reais. Para cada $i \in \{1, \dots, n\}$, indicamos por $x_{(k)i}$ a i -ésima coordenada do ponto $x_k \in \mathbb{R}^n$, isto é, $x_k = (x_{(k)1}, x_{(k)2}, \dots, x_{(k)n})$. As n sequências $\{x_{(k)i}\}$, com $i = 1, \dots, n$, são chamadas de sequências das coordenadas de $\{x_k\}$.

Exemplo 1.2.12. A sequência que associa a cada $k \in \mathbb{N}$ um ponto $x_k = (2^k, 2^{2k}, \dots, 2^{nk}) \in \mathbb{R}^n$, com $n \in \mathbb{N}$, é dada por $\{x_k\} = \{(2, 2^2, \dots, 2^n); \dots; (2^k, 2^{2k}, \dots, 2^{nk}); \dots\}$. Uma sequência das coordenadas de $\{x_k\}$ é $\{x_{(k)1}\}$, que é dada por $x_{(k)1} = 2^k$.

Podemos, então, estender também as noções de limitação e convergência de sequências para esse espaço. Uma sequência $\{x_k\} \subset \mathbb{R}^n$ é limitada se existir um número real $c > 0$ tal que $\|x_k\| \leq c$ para todo $k \in \mathbb{N}$. Ademais, a sequência $\{x_k\} \subset \mathbb{R}^n$ converge para um ponto $a \in \mathbb{R}^n$ se, para todo $\varepsilon > 0$, existe um número natural m tal que, para todo $k > m$, temos $\|x_k - a\| < \varepsilon$. Neste caso, dizemos que $\lim_{k \rightarrow \infty} \|x_k - a\| = 0$, o que implica que $\lim_{k \rightarrow \infty} x_k = a$. A relação entre a convergência de uma sequência em \mathbb{R}^n e a convergência de suas sequências das coordenadas é formalizada no resultado a seguir.

Teorema 1.2.13. Uma sequência $\{x_k\}$ em \mathbb{R}^n é convergente se, e somente se, as sequências $\{x_{(k)i}\}$ das coordenadas de $\{x_k\}$ convergem em \mathbb{R} .

Demonstração. Seja $\{x_k\} \subset \mathbb{R}^n$ convergente para $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. Vamos fixar um i , digamos i_0 , com $1 \leq i_0 \leq n$. Dado $\varepsilon > 0$ arbitrário, existe um número natural m tal que $\|x_k - a\| < \varepsilon$ para todo $k > m$. Assim, para todo $k > m$, temos $|x_{(k)i_0} - a_{i_0}| \leq \left(\sum_{i=1}^n |x_{(k)i} - a_i|^2\right)^{1/2} = \|x_k - a\| < \varepsilon$. Isto mostra que $\lim_{k \rightarrow \infty} x_{(k)i_0} = a_{i_0}$. Como i_0 é arbitrário, temos que $\lim_{k \rightarrow \infty} x_{(k)i} = a_i$ para todo $i = 1, \dots, n$. Reciprocamente, suponhamos que $\lim_{k \rightarrow \infty} x_{(k)i} = a_i$, para cada $i = 1, \dots, n$. Seja $\varepsilon > 0$ dado. Então, para cada $i = 1, \dots, n$ existe um número natural m_i tal que para todo $k > m_i$ tem-se $|x_{(k)i} - a_i| < \varepsilon/\sqrt{n} = \varepsilon_1$. Sejam $m = \max\{m_1, \dots, m_n\}$ e $a = (a_1, \dots, a_n) \in \mathbb{R}^n$. Também, para $k \geq m$, temos $\|x_k - a\| = \left(\sum_{i=1}^n |x_{(k)i} - a_i|^2\right)^{1/2} < \left(\sum_{i=1}^n \varepsilon_1^2\right)^{1/2} = \varepsilon$. Isto mostra que $\lim_{k \rightarrow \infty} x_k = a$. \square

Exemplo 1.2.14. Considere a sequência que associa a cada $k \in \mathbb{N}$ um ponto $x_k = (1/k, 2)$. Esta sequência converge para $(0, 2)$. Para provar isso pela definição, basta obter $m \in \mathbb{N}$ tal que, dado $\varepsilon > 0$ arbitrário, para todo $k > m$, a desigualdade $\|(1/k, 2) - (0, 2)\| < \varepsilon$ seja satisfeita. Assim, temos $\|(1/k, 2) - (0, 2)\| = \|(1/k, 0)\| = \sqrt{(1/k)^2 + 0^2} = 1/|k|$. Portanto, dado $\varepsilon > 0$, basta escolher $m > 1/\varepsilon$. Outra forma de provar isso é usar o Teorema 1.2.13 e os fatos que $\lim_{k \rightarrow \infty} x_{(k)1} = \lim_{k \rightarrow \infty} 1/k = 0$ e $\lim_{k \rightarrow \infty} x_{(k)2} = \lim_{k \rightarrow \infty} 2 = 2$.

Do Teorema 1.2.13, também segue que as operações envolvendo limites de sequências em \mathbb{R}^n podem ser reduzidas às operações com limites de sequências em \mathbb{R} . Ademais, a definição de subsequência segue da Definição 1.2.10 considerando $\{x_k\} \subset \mathbb{R}^n$.

Agora, considerando uma sequência $\{x_k\} \subset \mathbb{R}^n$, se existir um número real $c \in [0, 1)$ tal que $\|x_{k+1} - a\| \leq c\|x_k - a\|$ para todo $k \in \mathbb{N}$, então, aplicando isso recursivamente, temos $\|x_{k+1} - a\| \leq c\|x_k - a\| \leq \dots \leq c^{k+1}\|x_0 - a\|$. Segue disso que $0 < \|x_k - a\| \leq c^k\|x_0 - a\|$ para todo $k \in \mathbb{N}$. Como $c \in [0, 1)$, temos $\lim_{k \rightarrow \infty} \|x_k - a\| = 0$, o que significa que a sequência $\{x_k\}$ converge para a . A partir disso, temos o conceito de taxa de convergência.

Existem vários tipos de taxas com as quais uma sequência pode convergir. Apresentamos a seguir as taxas de convergência linear, superlinear e quadrática. As definições e exemplos apresentados a seguir são baseados em [22].

Definição 1.2.15. Considere a sequência $\{x_k\} \subset \mathbb{R}^n$. Dizemos que x_k converge linearmente para $a \in \mathbb{R}^n$ quando existe uma constante $c \in [0, 1)$ e um número natural m , tais que para todo $k \geq m$, tem-se

$$\frac{\|x_{k+1} - a\|}{\|x_k - a\|} \leq c.$$

Definição 1.2.16. Considere a sequência $\{x_k\} \subset \mathbb{R}^n$. Dizemos que x_k converge superlinearmente para $a \in \mathbb{R}^n$ quando

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - a\|}{\|x_k - a\|} = 0.$$

Exemplo 1.2.17. A sequência cujo k -ésimo termo é $x_k = 1/2^k$, com $k \in \mathbb{N}$, converge linearmente para $a = 0$, mas não superlinearmente. De fato, basta notar que

$$\frac{|x_{k+1} - a|}{|x_k - a|} = \frac{2^k}{2^{k+1}} = \frac{2^k}{2^k \cdot 2} = \frac{1}{2} \leq \frac{4}{5}.$$

Porém,

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - a|}{|x_k - a|} = \lim_{k \rightarrow \infty} \frac{1}{2} = \frac{1}{2} \neq 0.$$

Exemplo 1.2.18. A sequência cujo k -ésimo termo é $x_k = 1/2^{k^2}$, com $k \in \mathbb{N}$, converge superlinearmente para $a = 0$. De fato, basta notar que

$$\frac{|x_{k+1} - a|}{|x_k - a|} = \frac{2^{k^2}}{2^{(k+1)^2}} = \frac{1}{2^{2k} \cdot 2} = \frac{1}{2} \cdot \frac{1}{2^{2k}}$$

e sabemos que

$$\lim_{k \rightarrow \infty} \frac{1}{2} \cdot \frac{1}{2^{2k}} = \lim_{k \rightarrow \infty} \frac{1}{2} \cdot \lim_{k \rightarrow \infty} \frac{1}{2^{2k}} = \frac{1}{2} \cdot 0 = 0.$$

Definição 1.2.19. Considere a sequência $\{x_k\} \subset \mathbb{R}^n$. Dizemos que x_k converge quadraticamente para $a \in \mathbb{R}^n$ quando existe uma constante $c > 0$ e um número natural m , tais que para todo $k \geq m$, tem-se

$$\frac{\|x_{k+1} - a\|}{\|x_k - a\|^2} \leq c.$$

Exemplo 1.2.20. A sequência cujo k -ésimo termo é $x_k = \frac{1}{2^{2^k}}$, com $k \in \mathbb{N}$, converge quadraticamente para $a = 0$. De fato,

$$\frac{|x_{k+1} - a|}{|x_k - a|^2} = \frac{(2^{2^k})^2}{2^{2^{k+1}}} = \frac{2^{2^{k+1}}}{2^{2^{k+1}}} = 1.$$

Logo, basta tomar $c \geq 1$.

1.2.1 Continuidade e diferenciabilidade de funções

Nesta subseção, abordaremos os conceitos e alguns resultados clássicos de continuidade e diferenciabilidade de funções definidas em espaços euclidianos. Começaremos com a definição formal de continuidade de uma função em um ponto.

Definição 1.2.21. Seja $F : U \rightarrow \mathbb{R}^m$ uma função definida no conjunto $U \subset \mathbb{R}^n$. Dizemos que F é contínua no ponto $a \in U$ quando, para qualquer $\varepsilon > 0$ dado, se pode obter $\delta > 0$ tal que se $x \in U$ e $\|x - a\| < \delta$, então $\|F(x) - F(a)\| < \varepsilon$.

Se $F : U \rightarrow \mathbb{R}^m$ é contínua em todos os pontos do conjunto $U \subset \mathbb{R}^n$, diz-se que F é uma função contínua. Vale destacar que entre as hipóteses usuais para a análise da taxa de convergência do Método de Newton, que veremos no Capítulo 3, temos a de continuidade Lipschitz de uma função, que é uma forma de continuidade mais forte. Assim, apresentaremos sua definição a seguir.

Definição 1.2.22. Seja $F : U \rightarrow \mathbb{R}^m$ definida no conjunto $U \subset \mathbb{R}^n$. Dizemos que F é Lipschitz Contínua em U quando existe uma constante $L > 0$ (denominada constante de Lipschitz de F) tal que, para quaisquer $x, y \in U$, tem-se $\|F(x) - F(y)\| \leq L\|x - y\|$.

Além disso, relembremos ao leitor que nosso estudo tem como objetivo a determinação dos zeros de funções diferenciáveis, o que torna fundamental a compreensão do conceito de diferenciabilidade.

Definição 1.2.23. Sejam $f : U \rightarrow \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$ e um ponto $a \in U$. Definimos a i -ésima derivada parcial de f no ponto a (onde $1 \leq i \leq n$), como o limite

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + te_i) - f(a)}{t},$$

quando tal limite existe e e_i denota os vetores da base canônica do \mathbb{R}^n .

Em termos práticos, ao calcular a i -ésima derivada parcial de uma função, tratamos todas as variáveis como constantes, exceto a variável em relação à qual estamos diferenciando. Para essa variável, aplicamos as regras usuais de derivação, que podem ser consultadas em [17, Cap. 5]. Ademais, sabemos que as derivadas parciais sem hipóteses adicionais tornam-se casos particulares das derivadas direcionais, que definimos a seguir.

Definição 1.2.24. *Sejam $f : U \rightarrow \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$, um ponto $a \in U$ e uma direção $d \in \mathbb{R}^n$. Definimos a derivada direcional de f no ponto a , segundo a direção d , como o limite, quando tal limite existe,*

$$\frac{\partial f}{\partial d}(a) = \lim_{t \rightarrow 0} \frac{f(a + td) - f(a)}{t}.$$

Estenderemos agora essa definição para uma função $F : U \rightarrow \mathbb{R}^m$ definida no aberto $U \subset \mathbb{R}^n$. Definindo, para cada $j = 1, \dots, m$, a função $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $F(x) = (f_1(x), \dots, f_m(x))$ para todo $x \in \mathbb{R}^n$, a derivada direcional de F no ponto $a \in U$, segundo a direção $d \in \mathbb{R}^n$, é o vetor

$$\frac{\partial F}{\partial d}(a) = \left(\frac{\partial f_1}{\partial d}(a), \dots, \frac{\partial f_m}{\partial d}(a) \right) \in \mathbb{R}^m.$$

Considerando esses conceitos, podemos definir agora a diferenciabilidade de uma função, que implica na existência da derivada direcional da função para quaisquer direções.

Definição 1.2.25. *Seja $F : U \rightarrow \mathbb{R}^m$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Dizemos que F é diferenciável no ponto $a \in U$ quando existe um operador linear $F' : \mathbb{R}^n \rightarrow \mathbb{R}^m$ tal que, para todo $h \in \mathbb{R}^n$ com $a + h \in U$, tem-se*

$$F(a + h) = F(a) + F'(a)h + r(h), \text{ onde } \lim_{h \rightarrow 0} \frac{r(h)}{\|h\|} = 0.$$

Se $F : U \rightarrow \mathbb{R}^m$ é diferenciável em todos os pontos do conjunto aberto $U \subset \mathbb{R}^n$, diz-se simplesmente que F é uma função diferenciável. Assim, definimos a função derivada $F' : U \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ que associa a cada ponto $a \in U$ o operador linear $F'(a) : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Como mencionado na seção anterior, a partir do conceito de diferenciabilidade, obtemos um operador linear. A seguir, definimos a matriz que o representa.

Definição 1.2.26. *Seja $F : U \rightarrow \mathbb{R}^m$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Definindo, para cada $j = 1, \dots, m$, a função $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $F(x) = (f_1(x), \dots, f_m(x))$ para todo $x \in \mathbb{R}^n$, a representação matricial da derivada de F no ponto $a \in \mathbb{R}^n$ é deno-*

minada matriz jacobiana de F em a e é definida como

$$F'(a) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(a) & \cdots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \cdots & \frac{\partial f_m}{\partial x_n}(a) \end{bmatrix}.$$

No caso particular em que $m = 1$, temos $f : \mathbb{R}^n \rightarrow \mathbb{R}$. A matriz jacobiana de f no ponto $a \in U$, denotada por ∇f , é chamada de vetor gradiente de f em a e é dada por

$$f'(a) := \nabla f(a) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{bmatrix}.$$

Definição 1.2.27. *Seja $F : U \rightarrow \mathbb{R}^m$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Diz-se que F é continuamente diferenciável em U , ou que F é de classe \mathcal{C}^1 , e escreveremos $F \in \mathcal{C}^1$, quando F for diferenciável e, além disso, $F' : U \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ for contínua.*

Se F' é diferenciável em U , a sua derivada é chamada de segunda derivada de F e denotaremos por F'' . Nesse sentido, a k -ésima derivada de F é a derivada da derivada de ordem $k - 1$. De forma geral, se $F \in \mathcal{C}^k$, com $k \in \mathbb{N}$, dizemos que F é k vezes diferenciável e sua derivada F' é de classe \mathcal{C}^{k-1} . Além disso, as derivadas parciais de segunda ordem de uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável são denotadas por $\partial^2 f / \partial x_i \partial x_j$ para $i, j \in \{1, 2, \dots, n\}$.

Definição 1.2.28. *Dada $f : U \rightarrow \mathbb{R}$, definida no conjunto aberto $U \subseteq \mathbb{R}^n$, duas vezes diferenciável, a matriz jacobiana (derivada) do gradiente de f no ponto $a \in U$ é denominada matriz hessiana de f em a e é definida como*

$$\nabla^2 f(a) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(a) \end{bmatrix}.$$

Um resultado da Análise em \mathbb{R}^n , conhecido como Teorema de Schwarz, nos diz que se $f : U \rightarrow \mathbb{R}$, definida em $U \subset \mathbb{R}^n$, é duas vezes diferenciável no ponto $a \in U$, então as derivadas parciais mistas de segunda ordem são iguais. Assim, este resultado assegura que $\nabla^2 f(a)$ é uma matriz simétrica, desde que f seja duas vezes diferenciável em torno de $a \in U$. A prova desse resultado pode ser vista em [17, Cap. 3, Sec. 7].

Exemplo 1.2.29. *Consideremos a função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida por $f(x, y) = [(x-2)^2/2] + (y-1)^2$. Para determinar o gradiente da função f , calculamos as derivadas parciais de*

f em relação a x e y . Assim, o gradiente da função f é dado por

$$\nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x, y), \frac{\partial f}{\partial y}(x, y) \right) = (x - 2, 2(y - 1)).$$

Agora, para determinarmos a matriz hessiana de f , calculamos as segundas derivadas parciais de f . Daí, segue que

$$\nabla^2 f(x, y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial x \partial y}(x, y) \\ \frac{\partial^2 f}{\partial y \partial x}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}.$$

Note que, neste caso, as entradas da matriz hessiana de f não variam de acordo com ponto (x, y) e, como vimos no Exemplo 1.1.12, esta matriz é definida positiva.

Neste contexto, outro resultado importante, amplamente empregado na otimização e que será utilizado nos capítulos subsequentes, é a Fórmula de Taylor de Segunda Ordem. A demonstração deste resultado pode ser encontrada em [17, Cap. 3, Sec. 8].

Teorema 1.2.30 (Fórmula de Taylor de Segunda Ordem). *Seja a função $f : U \rightarrow \mathbb{R}$, definida no aberto $U \subset \mathbb{R}^n$, duas vezes diferenciável em $x \in U$ com $x + d \in U$. Então*

$$f(x + d) = f(x) + \nabla f(x)^\top d + \frac{1}{2} d^\top \nabla^2 f(x) d + r(d),$$

com $\lim_{\|d\| \rightarrow 0} r(d)/\|d\|^2 = 0$.

Por fim, apresentaremos a Desigualdade do Valor Médio, um resultado que será utilizado na prova da convergência quadrática da sequência gerada pelo Método de Newton. Sua demonstração pode ser encontrada em [17, Cap. 5, Sec. 5.5, p. 209].

Teorema 1.2.31 (Desigualdade do Valor Médio). *Dado $U \subset \mathbb{R}^n$ aberto, seja $F : U \rightarrow \mathbb{R}^m$ diferenciável em cada ponto do segmento de reta aberto (x, y) e tal que sua restrição ao segmento fechado $[x, y] \subset U$ seja contínua. Se $\|F'(z)\| \leq M$ para todo $z \in (x, y)$, então $\|F(y) - F(x)\| \leq M\|y - x\|$.*

Capítulo 2

Aspectos da Otimização Contínua

Na Otimização Contínua, pode-se determinar zeros de funções por meio dos processos de minimização ou maximização. Neste trabalho, estamos interessados na minimização irrestrita de funções diferenciáveis. Para resolver esse tipo de problema, uma classe importante de métodos numéricos utiliza buscas lineares, como os métodos do gradiente e de Newton, veja [11, 26]. As buscas lineares podem influenciar diretamente a eficiência desses métodos e incluem estratégias bem conhecidas, como as buscas da Seção Áurea, de Armijo, entre outras que podem ser consultadas em [24].

Neste capítulo, apresentamos as definições básicas e os resultados auxiliares da Otimização Contínua que serão utilizados neste trabalho, a busca de Armijo e sua interpretação geométrica.

2.1 Definições básicas e resultados auxiliares

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 $\Omega \subset \mathbb{R}^n$ não vazio e uma função $f : \Omega \rightarrow \mathbb{R}$. O problema de encontrar o minimizador x^* de f no conjunto Ω pode ser descrito como:

$$\text{minimizar } f(x) \text{ sujeito a } x \in \Omega, \quad (2.1)$$

em que f é chamada de função objetivo. Quando $\Omega = \mathbb{R}^n$, dizemos que o problema (2.1) é irrestrito. Já quando $\Omega \neq \mathbb{R}^n$, chamamos de problema restrito. Além disso, um ponto $x^* \in \Omega$ é um minimizador local de f se houver $\delta > 0$ tal que $f(x^*) \leq f(x)$ para todo $x \in \Omega \cap B(x^*, \delta)$. Enquanto que $x^* \in \Omega$ é chamado de minimizador global de f se $f(x^*) \leq f(x)$ para todo $x \in \Omega$. O Teorema 2.1.1 garante a existência de um minimizador global para uma classe de funções.

Teorema 2.1.1 (Teorema de Weierstrass). *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto compacto não vazio e $f : \Omega \rightarrow \mathbb{R}$ uma função contínua. Então existe minimizador global de f em Ω .*

Demonstração. Sabemos que o conjunto imagem de um conjunto compacto por uma função contínua é compacto, então o conjunto $f(\Omega)$ é compacto. Usando o fato que $f(\Omega)$ é limitado e a completude da reta, temos que existe o ínfimo $i = \inf f(\Omega)$, isto é, para todo $k \in \mathbb{N}$ existe $x_k \in \Omega$ tal que $i \leq f(x_k) \leq i + 1/k$. Fazendo $k \rightarrow \infty$, concluímos que $\lim_{k \rightarrow \infty} f(x_k) = i$. Como $\{x_k\} \subset \Omega$ e Ω é compacto, segue-se que $\{x_k\}$ é limitada. Então $\{x_k\}$ possui uma subsequência $\{x_{k_j}\}$ tal que $\lim_{j \rightarrow \infty} x_{k_j} = x^* \in \Omega$, já que Ω é fechado. Pela continuidade de f , temos $\lim_{j \rightarrow \infty} f(x_{k_j}) = f(x^*) = i$. Portanto, f assume o seu valor mínimo em Ω no ponto x^* , isto é, x^* é um minimizador global de f em Ω . \square

Abordaremos a seguir as condições de otimalidade de primeira e segunda ordem, que se baseiam nas derivadas da função objetivo para identificar os possíveis minimizadores do problema (2.1). O resultado apresentado a seguir é de grande importância para este trabalho, pois fundamenta a ideia de encontrar zeros de funções através do processo de minimização/maximização, em particular, a determinação do zero do gradiente da função.

Teorema 2.1.2 (Condição necessária de primeira ordem). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um minimizador local irrestrito de f , então*

$$\nabla f(x^*) = 0. \quad (2.2)$$

Demonstração. Seja $d \in \mathbb{R}^n$ uma direção não nula arbitrária, porém fixa. Pela definição de minimizador local, existe $\delta > 0$ tal que $f(x^*) \leq f(x^* + \alpha d)$, para todo $\alpha \in [0, \delta)$. Pela diferenciabilidade de f em x^* , temos $f(x^* + \alpha d) - f(x^*) = \alpha \nabla f(x^*)^\top d + r(\alpha)$, com $\lim_{\alpha \rightarrow 0} r(\alpha)/|\alpha| = 0$. Assim, temos $\alpha \nabla f(x^*)^\top d + r(\alpha) \geq 0$. Dividindo ambos os lados da desigualdade por $\alpha > 0$, temos $\nabla f(x^*)^\top d + r(\alpha)/\alpha \geq 0$ e passando o limite quando $\alpha \rightarrow 0$, obtemos $\nabla f(x^*)^\top d \geq 0$. Suponha, por absurdo, que $\nabla f(x^*)$ não é nulo. Como d é uma direção não nula arbitrária, podemos escolher $d = -\nabla f(x^*)$. Segue daí que $-\|\nabla f(x^*)\|^2 \geq 0$, o que é um absurdo. Logo, $\nabla f(x^*) = 0$. \square

A demonstração do resultado a seguir pode ser vista em [10, Cap. 1, Sec. 1.3, Teo. 1.3.1] ou [26, Cap. 2, Sec. 2, Teo. 2.12].

Teorema 2.1.3 (Condição necessária de segunda ordem). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* é um minimizador local irrestrito de f , então vale (2.2) e $\nabla^2 f(x^*)$ é semidefinida positiva.*

Teorema 2.1.4 (Condição suficiente de segunda ordem). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ duas vezes diferenciável no ponto $x^* \in \mathbb{R}^n$. Se x^* satisfaz (2.2) e a matriz hessiana de f é definida positiva em x^* , então x^* é minimizador local estrito de f .*

Demonstração. Seja λ o menor autovalor de $\nabla^2 f(x^*)$. Pelo Lema 1.1.11, temos $\lambda > 0$ e, pelo Lema 1.1.9, temos $d^\top \nabla^2 f(x^*) d \geq \lambda \|d\|^2$ para todo $d \in \mathbb{R}^n$ não nulo. Pela Fórmula

de Taylor de Segunda Ordem (veja o Teorema 1.2.30) e considerando a hipótese que $\nabla f(x^*) = 0$, obtemos

$$f(x^* + d) - f(x^*) = \frac{1}{2}d^T \nabla^2 f(x^*)d + r(d) \geq \frac{1}{2}\lambda \|d\|^2 + r(d) = \|d\|^2 \left(\frac{\lambda}{2} + \frac{r(d)}{\|d\|^2} \right),$$

onde $\lim_{d \rightarrow 0} r(d)/\|d\|^2 = 0$. Segue daí que

$$\lim_{d \rightarrow 0} \left(\frac{\lambda}{2} + \frac{r(d)}{\|d\|^2} \right) = \frac{\lambda}{2} > 0.$$

Assim, existe um $\delta > 0$ tal que para todo $d \in B(0, \delta) \setminus \{0\}$, vale

$$\frac{\lambda}{2} + \frac{r(d)}{\|d\|^2} > 0.$$

Isso implica que $f(x^* + d) > f(x^*)$ para todo $d \in B(0, \delta) \setminus \{0\}$, ou seja, $f(x) > f(x^*)$ para todo $x \in B(x^*, \delta) \setminus \{x^*\}$. Portanto, x^* é um minimizador local estrito de f . \square

Além dessas condições de otimalidade, na Otimização Contínua, a convexidade de uma função é um aspecto importante na análise de problemas de minimização. A convexidade garante, entre outras propriedades, que qualquer minimizador local de uma função convexa definida em um conjunto convexo é também um minimizador global. Para o leitor interessado em mais informações sobre análise convexa, recomendamos [10]. A seguir, definiremos conjuntos convexos e funções convexas e demonstraremos resultados que serão utilizados posteriormente.

Definição 2.1.5. O conjunto $\Omega \subset \mathbb{R}^n$ é dito convexo quando contém qualquer segmento de reta cujos extremos pertençam a Ω , isto é, se $x, y \in \Omega$, então $[x, y] = \{\alpha x + (1 - \alpha)y : \alpha \in [0, 1]\} \subset \Omega$.

Exemplo 2.1.6. Toda bola $B \subset \mathbb{R}^n$ é um conjunto convexo. De fato, se $B = B(a; \delta)$, isto é, B a bola aberta de centro $a \in \mathbb{R}^n$ e raio $\delta > 0$, então $\|x - a\| < \delta$ e $\|y - a\| < \delta$, para todo $x, y \in B$. Isso implica que para qualquer $\alpha \in [0, 1]$, temos $\|\alpha x + (1 - \alpha)y - a\| = \|\alpha(x - a) + (1 - \alpha)(y - a)\| \leq \alpha\|x - a\| + (1 - \alpha)\|y - a\| < \delta$. De forma análoga, temos que a bola fechada $B[a; \delta]$ também é convexa.

Definição 2.1.7. Seja $\Omega \subset \mathbb{R}^n$ convexo. A função $f : \Omega \rightarrow \mathbb{R}$ é dita convexa em Ω quando $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$, para todo $x, y \in \Omega$ e $\alpha \in [0, 1]$.

Uma relação importante que envolve as duas definições apresentadas anteriormente é o Teorema 2.1.8 e sua demonstração pode ser vista em [10, Teo. 3.4.30].

Teorema 2.1.8 (Caracterização de funções convexas diferenciáveis). *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \rightarrow \mathbb{R}$ uma função duas vezes diferenciável em Ω . A*

função f é convexa em Ω se, e somente se, a matriz hessiana de f é semidefinida positiva em todo ponto de Ω .

Teorema 2.1.9. *Sejam $\Omega \subset \mathbb{R}^n$ convexo e $f : \Omega \rightarrow \mathbb{R}$ uma função convexa. Se x^* é minimizador local de f , então x^* é minimizador global de f .*

Demonstração. Suponha, por absurdo, que $x^* \in \Omega$ é um minimizador local de f que não é minimizador global. Então, existe $x \in \Omega$ tal que $f(x) < f(x^*)$. Pela convexidade de Ω , temos $y = \alpha x + (1 - \alpha)x^* \in \Omega$ para todo $\alpha \in [0, 1]$. Agora, pela convexidade de f , para todo $\alpha \in (0, 1]$, tem-se $f(y) \leq \alpha f(x) + (1 - \alpha)f(x^*) = f(x^*) + \alpha(f(x) - f(x^*)) < f(x^*)$. Tomando $\alpha > 0$ suficientemente pequeno, podemos garantir que y é arbitrariamente próximo de x^* , e ainda tem-se que $f(y) < f(x^*)$ com $y \in \Omega$, o que é um absurdo. Portanto, qualquer minimizador local de f em Ω é também um minimizador global. \square

Com os resultados previamente estabelecidos, podemos agora explorar técnicas da Otimização Contínua para resolver o problema (2.1). Em geral, para obter sequências que convergem para o minimizador de uma função, utiliza-se o seguinte procedimento: toma-se um ponto no domínio da função, escolhe-se um vetor a partir desse ponto e obtém-se um novo ponto na sequência somando-se o ponto inicial com o vetor. Em alguns casos, é necessário ajustar o comprimento do vetor multiplicando-o por um número real, que é conhecido como comprimento de passo.

A escolha do vetor varia de acordo com o contexto do problema e o algoritmo utilizado. Uma das estratégias mais comuns para resolver o problema (2.1) é a seguinte: dada uma aproximação $x \in \Omega$ da solução do problema, com $\Omega \subset \mathbb{R}^n$, busca-se encontrar um ponto $y \in \Omega$ tal que $f(y) < f(x)$. Nesse sentido, em alguns algoritmos da Otimização Contínua, toma-se uma direção $d \in \mathbb{R}^n$ tal que f seja decrescente a partir do ponto x nessa direção. Uma definição que descreve uma direção para o decrescimento da função é a seguinte:

Definição 2.1.10. *Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida da função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $x \in \mathbb{R}^n$, se existe $\delta > 0$ tal que $f(x + \alpha d) < f(x)$ para todo $\alpha \in (0, \delta]$.*

Lema 2.1.11. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $x \in \mathbb{R}^n$. Então:*

- (a) *Para toda direção de descida $d \in \mathbb{R}^n$, tem-se que $\nabla f(x)^\top d \leq 0$.*
- (b) *Se $d \in \mathbb{R}^n$ satisfaz $\nabla f(x)^\top d < 0$, então d é uma direção de descida.*

Demonstração. Seja $d \in \mathbb{R}^n$ uma direção de descida. Para todo $\alpha > 0$ suficientemente pequeno, pela diferenciabilidade de f em x , temos $f(x + \alpha d) - f(x) = \alpha \nabla f(x)^\top d + r(\alpha)$, com $\lim_{\alpha \rightarrow 0} r(\alpha)/|\alpha| = 0$. Como $f(x + \alpha d) < f(x)$, temos $\alpha \nabla f(x)^\top d + r(\alpha) < 0$. Dividindo ambos os lados dessa desigualdade por $\alpha > 0$ e passando ao limite quando $\alpha \rightarrow 0$, obtemos $\nabla f(x)^\top d \leq 0$. Isto prova o item (a).

Suponhamos agora que $d \in \mathbb{R}^n$ satisfaz $\nabla f(x)^\top d < 0$. Temos que $f(x + \alpha d) - f(x) = \alpha(\nabla f(x)^\top d + r(\alpha)/\alpha)$. Em particular, para todo $\alpha > 0$ suficientemente pequeno, temos

$$\nabla f(x)^\top d + \frac{r(\alpha)}{\alpha} \leq \frac{1}{2} \nabla f(x)^\top d < 0,$$

o que implica que $f(x + \alpha d) - f(x) < 0$, isto é, d é direção de descida, o que prova (b). \square

Considerando a premissa de que a direção do gradiente representa a direção de maior crescimento de uma função e que, portanto, sua direção oposta resulta no maior decréscimo, essa direção é amplamente empregada em métodos numéricos para a minimização de funções. No exemplo a seguir, demonstraremos que, de fato, essa é uma direção de descida.

Exemplo 2.1.12. *Dada $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável, temos que $-\nabla f(x)$ é uma direção de descida de f em $x \in \mathbb{R}^n$. De fato, basta notar que $-\nabla f(x)^\top \nabla f(x) = -\|\nabla f(x)\|^2 < 0$.*

Exemplo 2.1.13. *Considere a função e o vetor gradiente do Exemplo 1.2.29. A partir do ponto $p = (1, 0)$, temos que $d = (3, 1)$ é uma direção de descida da função f . De fato, basta notar que $\nabla f(p)^\top d = (-1, -2)^\top (3, 1) = -5 < 0$.*

Como mencionado anteriormente, em alguns casos, a determinação do novo ponto da sequência pode ser mais eficiente se for dada como a soma do ponto com uma contração do vetor diretor. Na próxima seção, abordaremos uma técnica da Otimização Contínua para a obtenção dessa contração.

2.2 A busca linear de Armijo

Uma ferramenta eficaz que auxilia na determinação de comprimentos de passo é a chamada busca de Armijo. O resultado principal que caracteriza esta busca também é conhecido como condição de diminuição suficiente e assegura que, ao empregar a busca de Armijo a partir de um ponto x na direção de descida d , existe um comprimento de passo $\alpha > 0$ que proporciona um decréscimo suficiente da função f em seu novo ponto, dado por $x + \alpha d$. A busca de Armijo é formalmente apresentada pelo teorema a seguir.

Teorema 2.2.1. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável, $d \in \mathbb{R}^n$ uma direção de descida de f a partir de $x \in \mathbb{R}^n$ e $\eta \in (0, 1)$. Então existe $\delta > 0$ tal que para todo $\alpha \in [0, \delta)$, temos*

$$f(x + \alpha d) \leq f(x) + \eta \alpha \nabla f(x)^\top d. \quad (2.3)$$

Demonstração. Como d é direção de descida de f a partir de $x \in \mathbb{R}^n$, pelo Lema 2.1.11, temos $\nabla f(x)^\top d \leq 0$. Para $\alpha = 0$ ou quando $\nabla f(x)^\top d = 0$, é imediato que vale (2.3). Assim, queremos provar que se $\nabla f(x)^\top d < 0$, então existe $\delta > 0$ tal que (2.3) é válida para

todo $\alpha \in (0, \delta)$. Pela diferenciabilidade de f , temos que $f(x + \alpha d) - f(x) = \alpha \nabla f(x)^\top d + r(\alpha)$, onde $\lim_{\alpha \rightarrow 0} r(\alpha)/|\alpha| = 0$. Então, podemos escrever, para $\eta \in (0, 1)$,

$$f(x + \alpha d) - f(x) = \eta \alpha \nabla f(x)^\top d + \alpha \left((1 - \eta) \nabla f(x)^\top d + \frac{r(\alpha)}{\alpha} \right). \quad (2.4)$$

Além disso, temos que

$$\lim_{\alpha \rightarrow 0} \left((1 - \eta) \nabla f(x)^\top d + \frac{r(\alpha)}{\alpha} \right) = (1 - \eta) \nabla f(x)^\top d < 0.$$

Assim, existe $\delta > 0$ tal que, para todo $\alpha \in (0, \delta)$, a desigualdade

$$(1 - \eta) \nabla f(x)^\top d + \frac{r(\alpha)}{\alpha} < 0$$

é válida. De (2.4) e desta última desigualdade, concluímos que existe $\delta > 0$ tal que (2.3) é válida para todo $\alpha \in (0, \delta)$. \square

Uma interpretação geométrica do conjunto de comprimentos de passo admissíveis na busca de Armijo é que ele corresponde ao conjunto de todos os comprimentos de passo $\alpha > 0$ tais que $f(x + \alpha d) - f(x)$ esteja abaixo da reta $\eta \alpha \nabla f(x)^\top d$. Isso pode ser visto na Figura 2.1. Quando (2.3) não é satisfeita para algum comprimento de passo, ele pode ser rejeitado ou ajustado até que pertença ao intervalo de valores admissíveis.

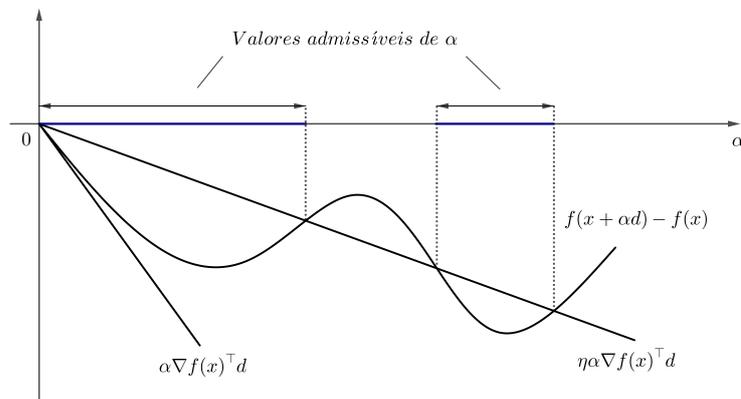


Figura 2.1: Interpretação geométrica do conjunto de valores admissíveis para a busca de Armijo.

Uma aplicação da busca de Armijo é apresentada no Capítulo 4, em que discutimos detalhadamente o processo de minimização irrestrita de uma função objetivo do tipo soma de quadrados.

Capítulo 3

Método de Newton

O Método de Newton é amplamente reconhecido como uma ferramenta poderosa para encontrar zeros de funções, desempenhando um papel fundamental na teoria matemática e em aplicações, como na resolução de problemas de programação linear, [19], e na sensibilidade da pressão exercida pela mão de um robô ao segurar um objeto, [8].

Neste capítulo, descrevemos o Método de Newton para encontrar zeros de funções continuamente diferenciáveis e analisamos a convergência local da sequência gerada por ele. Na primeira seção, abordamos o procedimento utilizado para obter a direção e a sequência newtoniana, bem como a interpretação geométrica desse método. Na última seção, demonstramos o teorema clássico de convergência local do Método de Newton.

3.1 O Método de Newton

Dado um conjunto aberto $\Omega \subseteq \mathbb{R}^n$ e uma função $F : \Omega \rightarrow \mathbb{R}^n$ continuamente diferenciável, vamos considerar o problema de encontrar um $x^* \in \Omega$ tal que

$$F(x^*) = 0. \quad (3.1)$$

A seguir, apresentaremos dois exemplos em que os zeros da função podem ser obtidos por meio de manipulações algébricas.

Exemplo 3.1.1. *Uma função afim é uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ definida por $f(x) = ax + b$, onde a e b são constantes, com $a \neq 0$. O zero da função afim é dado por $x = -b/a$.*

Exemplo 3.1.2. *Considere a função $g : \mathbb{R} \rightarrow \mathbb{R}$ definida por $g(x) = x^4 - 10x^2 - 24$. Queremos resolver a equação $x^4 - 10x^2 - 24 = 0$, isto é, determinar os zeros reais de g . Primeiro, substituímos a incógnita x^2 por y e obtemos $y^2 - 10y - 24 = 0$. Esta equação do segundo grau possui duas soluções: $y = 12$ e $y = -2$. Assim, vamos relacionar essas soluções com a substituição $x^2 = y$. Temos $x^2 = 12 \implies x = \pm 2\sqrt{3}$ e $x^2 = -2$ que não fornece uma solução real. Portanto, os zeros reais de g são $x_1 = 2\sqrt{3}$ e $x_2 = -2\sqrt{3}$.*

Nestes exemplos, a determinação dos zeros das funções pode ser feita de forma algébrica ou analítica com relativa facilidade. Entretanto, conforme destaca Humes, [9], para polinômios de grau maior ou igual a cinco, essa abordagem pode se tornar inviável. Além disso, para funções mais complexas ou sistemas de alta ordem, a solução algébrica/analítica pode não ser possível. Nesse contexto, não há teorias estabelecidas para encontrar os zeros de uma função diferenciável qualquer, tornando os métodos iterativos uma alternativa viável. Entre esses métodos, destaca-se o Método de Newton, cuja ideia para resolver o problema (3.1) é aproximar uma solução por meio de uma sequência de aproximações obtidas a partir da resolução de um sistema linear.

Para obtermos o sistema linear que determina a direção de Newton para a resolução do problema em questão, realizaremos o seguinte procedimento: dado um número natural k , seja $x_k \in \Omega$ tal que a derivada $F'(x_k)$ seja não nula. Pela diferenciabilidade de F , para todo $x \in \Omega$, temos que

$$F(x) = F(x_k) + F'(x_k)(x - x_k) + r(x - x_k),$$

onde $\lim_{x \rightarrow x_k} r(x - x_k)/\|x - x_k\| = 0$. A partir disto, vamos considerar a aproximação linear de F em x_k dada por

$$F(x) \approx F(x_k) + F'(x_k)(x - x_k).$$

Assim, para obtermos $F(x) = 0$, resolvemos a equação linear

$$0 = F(x_k) + F'(x_k)(x - x_k).$$

Esta última equação é chamada de equação de iteração do Método de Newton. Agora, fazendo a mudança de variável $d_k = x - x_k$, onde d_k será uma solução do sistema linear

$$F'(x_k)d_k = -F(x_k), \tag{3.2}$$

obtemos a direção de Newton a partir de x_k . Admitindo que exista $F'(x_k)^{-1}$, para todo $k \in \mathbb{N}$, o Método de Newton pode ser escrito na forma do esquema iterativo:

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k), \quad k = 0, 1, \dots \tag{3.3}$$

Para que este esquema possa ocorrer repetidas vezes, faz-se necessário que $x_k \in \Omega$ e $F'(x_k)^{-1}$ exista para todo $k \in \mathbb{N}$. A determinação de x_{k+1} em função de x_k pode ser justificada geometricamente para o seguinte caso: a inclinação da reta tangente ao gráfico de F no ponto $(x_k, F(x_k))$ é $F'(x_k)$. A única reta com a inclinação $F'(x_k)$ que passa por $(x_k, F(x_k))$ é $y = F(x_k) + F'(x_k)(x - x_k)$. O ponto x_{k+1} é definido como x que satisfaça $y = 0$, isto é, $0 = F(x_k) + F'(x_k)(x_{k+1} - x_k)$ ou (3.3). Isto implica que x_{k+1} é a interseção

da reta tangente com o eixo das abscissas. Essa interpretação geométrica pode ser vista na Figura 3.1.

Segundo Arenales e Darezzo, [1], também faz-se necessário que $F'(x_k)$ seja não nula quando $F(x_k) \neq 0$, pois se tivéssemos $F'(x_k) = 0$, a reta tangente à função no ponto x_k seria paralela ao eixo das abscissas e, portanto, x_{k+1} não estaria bem definido. Porém, se $F'(x^*) = 0$, em que $x^* \in \Omega$ é a solução do problema (3.1), e $F'(x_k) \neq 0$, podemos calcular x_{k+1} para todo $k \in \mathbb{N}$.

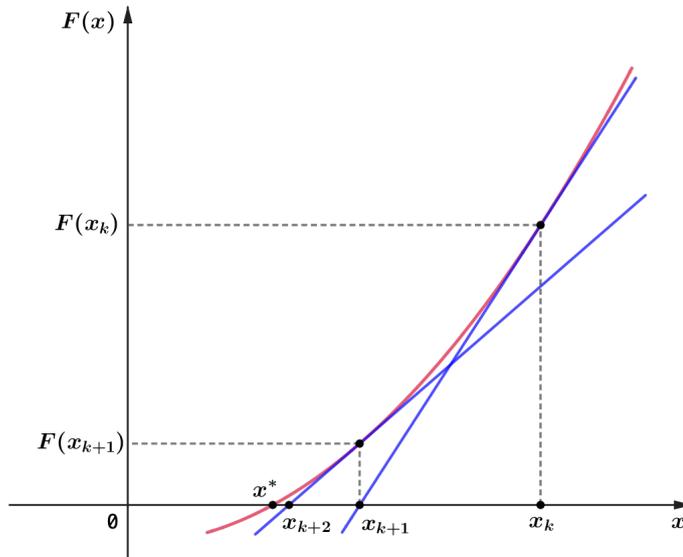


Figura 3.1: Interpretação geométrica do esquema iterativo (3.3).

Na Figura 3.1, a curva em vermelho é a representação gráfica de F , enquanto que as retas azuis são as tangentes ao gráfico de F nos pontos da forma $(x, F(x))$. Além disso, as interseções das retas tangentes ao gráfico de F com o eixo das abscissas se aproximam do zero da função. Vale ressaltar que, supondo apenas que $x_k \in \Omega$ e que exista $F'(x_k)^{-1}$, não há garantia que a sequência gerada pelo esquema iterativo (3.3) é convergente para a solução do problema. O exemplo a seguir ilustra este fato.

Exemplo 3.1.3. Seja $f : \mathbb{R} \rightarrow \mathbb{R}$, definida por

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

Essa função possui um único zero que é dado por $x^* = 0$. Assim, vamos analisar a convergência da sequência gerada pelo esquema iterativo (3.3) para x^* . Temos que f é continuamente diferenciável e sua derivada é dada por $f'(x) = 1/\sqrt{(1+x^2)^3}$, isto é, $f'(x) \neq 0$, para todo $x \in \mathbb{R}$. Por (3.3), temos

$$x_{k+1} = x_k - \left(\frac{1}{\sqrt{(1+x_k^2)^3}} \right)^{-1} \left(\frac{x_k}{\sqrt{1+x_k^2}} \right) = x_k - x_k(1+x_k^2) = -x_k^3. \quad (3.4)$$

Note que esta sequência é da forma $\{x_k\} = \{x_0; -x_0^3; x_0^9; -x_0^{27}; \dots\}$. A escolha do ponto inicial $x_0 \in \mathbb{R}$ pode ser dada a partir de três casos: $|x_0| = 1$ ou $|x_0| > 1$ ou $|x_0| < 1$. Analisaremos, a seguir, a convergência de $\{x_k\}$ em cada um dos casos citados.

(i) Se $|x_0| = 1$, então $\{x_k\}$ diverge;

De fato, tomando x_0 tal que $|x_0| = 1$, teremos uma sequência alternada entre os valores 1 e -1 que é uma sequência divergente, pois admite duas subsequências que convergem para limites distintos.

(ii) Se $|x_0| > 1$, então $\{x_k\}$ diverge;

Considerando a subsequência de índices pares, temos $\{x_0; x_0^9; \dots\}$, que é da forma $x_k = x_0^{3^k}$, para $k \geq 1$. Como $|x_0| > 1$, essa subsequência é crescente e não limitada superiormente, o que implica que $\{x_k\}$ não é limitada. Portanto, pela contrapositiva do Teorema 1.2.8, temos que $\{x_k\}$ é divergente.

Em particular, na Figura 3.2, representamos graficamente a sequência gerada pelo esquema iterativo (3.3) quando tomamos $x_0 = 11/10$. Observa-se que, à medida que k aumenta, o termo x_k torna-se mais distante de x^* .

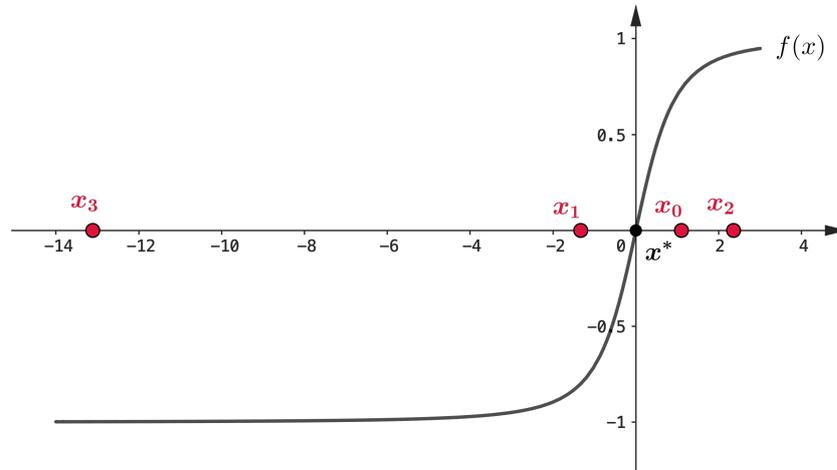


Figura 3.2: Sequência gerada pelo Método de Newton para $x_0 = 11/10$.

(iii) Se $|x_0| < 1$, então $\{x_k\}$ converge para $x^* = 0$.

Note que a sequência gerada por (3.4) tem termo geral $x_k = (-1)^k x_0^{3^k}$ para $k \in \mathbb{N}$. Se $0 < x_0 < 1$, então $|x_k| = |(-1)^k x_0^{3^k}| = x_0^{3^k}$, que é uma subsequência de $x_n = x_0^n$, com $n \in \mathbb{N}$, a qual converge para zero. Sabemos que $\lim |x_0^n| = \lim |x_0|^n = 0$. Logo, concluímos que x_k também converge para zero quando $-1 < x_0 < 1$.

Em particular, na Figura 3.3, representamos graficamente a sequência gerada pelo esquema iterativo (3.3) quando tomamos $x_0 = 9/10$.

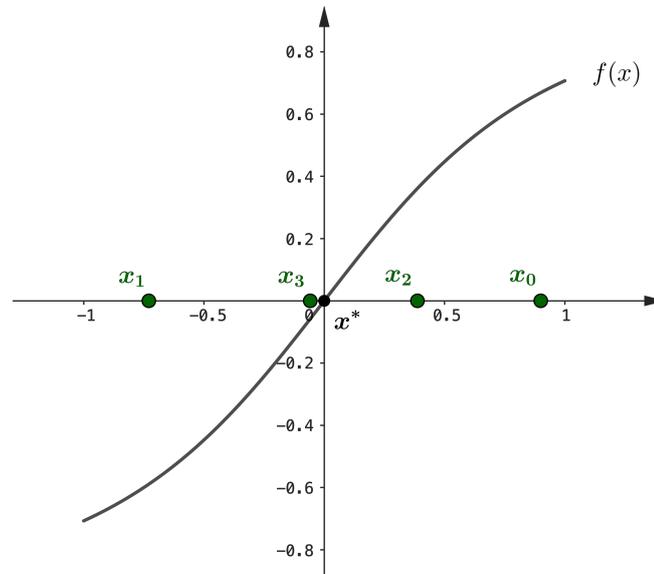


Figura 3.3: Sequência gerada pelo Método de Newton para $x_0 = 9/10$.

Neste exemplo, podemos concluir que a sequência newtoniana está bem definida para cada um dos casos. Porém, a convergência desta sequência para o zero da função só é dada quando o ponto inicial $x_0 \in \mathbb{R}$ é tal que $|x_0| < 1$. Assim, esse exemplo evidencia a convergência local do Método de Newton, mostrando que a convergência da sequência gerada por ele depende da boa escolha do ponto inicial.

Na próxima seção, abordaremos mais profundamente as hipóteses adequadas para a garantia da convergência local do Método de Newton. Para isso, apresentamos primeiramente o Algoritmo 1 do Método de Newton para resolver o problema (3.1).

Algoritmo 1: MÉTODO DE NEWTON

- 1 Considere um ponto inicial $x_0 \in \Omega$, com $\Omega \subset \mathbb{R}^n$, e $\varepsilon > 0$. Defina $k = 0$.
 - 2 Se $\|F(x_k)\| < \varepsilon$, então pare; caso contrário, continue.
 - 3 Resolva o sistema linear (3.2).
 - 4 Defina $x_{k+1} = x_k + d_k$ e $k = k + 1$. Retorne ao passo 2.
-

No passo 2, temos o critério de convergência: se $\|F(x_k)\| < \varepsilon$ para um $\varepsilon > 0$ pequeno, o algoritmo para, indicando que encontramos uma aproximação satisfatória para o zero da função F . Caso contrário, no passo 3, resolve-se o sistema linear que determina a direção de Newton que é usada no próximo passo para atualizar a iterada. No passo 4, atualizamos a iterada e retornamos ao passo 2 até que a convergência seja atingida.

3.2 Análise de convergência

A fim de realizarmos uma análise da convergência local do Método de Newton, apresentaremos inicialmente dois lemas que nos fornecerá uma vizinhança em que a sequência

gerada por este método estará bem definida. O Lema 3.2.1 é uma consequência do conhecido Teorema sobre perturbações pequenas de uma matriz invertível e sua demonstração pode ser encontrada em [23, Cap. 2, Sec. 2].

Lema 3.2.1. *Suponha que a função $A : \Omega \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ definida no aberto $\Omega \subset \mathbb{R}^n$ é contínua em um ponto $x^* \in \Omega$ tal que $A(x^*)$ é invertível. Então, existem $\delta > 0$ e $\gamma > 0$ tais que $A(x)$ é invertível e $\|A(x)^{-1}\| \leq \gamma$ para todo $x \in \Omega \cap B(x^*; \delta)$. Além disso, A^{-1} é contínua em x^* .*

Agora, para o Lema 3.2.2, vamos considerar a função de iteração do Método de Newton $N_F : \Omega \cap B(x^*; \delta) \rightarrow \mathbb{R}^n$ definida por $N_F(x) = x - F'(x)^{-1}F(x)$, onde $x^* \in \Omega$ é a solução do problema (3.1).

Lema 3.2.2. *Seja $F : \Omega \rightarrow \mathbb{R}^n$ uma função definida no aberto $\Omega \subset \mathbb{R}^n$ e diferenciável numa vizinhança do ponto $x^* \in \Omega$, com derivada $F' : \Omega \rightarrow \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ contínua neste ponto. Além disso, sejam x^* um zero de F , tal que $F'(x^*)$ seja invertível, e $\delta > 0$ dado pelo Lema 3.2.1. Então, a função N_F está bem definida e existe $\hat{\delta} > 0$, com $\delta \geq \hat{\delta}$, tal que $N_F(x) \in \Omega \cap B(x^*; \hat{\delta})$ para todo $x \in \Omega \cap B(x^*; \hat{\delta})$.*

Demonstração. Pelo Lema 3.2.1, temos que $F'(x)$ é invertível para todo $x \in \Omega \cap B(x^*; \delta)$. Assim, N_F está bem definida em $\Omega \cap B(x^*; \delta)$. Agora, vamos provar que existe $\hat{\delta} \leq \delta$ tal que $N_F(x) \in \Omega \cap B(x^*; \hat{\delta})$ para todo $x \in \Omega \cap B(x^*; \hat{\delta})$. Como $F(x^*) = 0$, temos

$$\begin{aligned} \|N_F(x) - x^*\| &= \|(x - F'(x)^{-1}F(x)) - x^*\| \\ &= \|x - x^* - F'(x)^{-1}F(x) + F'(x)^{-1}F(x^*)\| \\ &= \|x - x^* - F'(x)^{-1}(F(x) - F(x^*))\| \\ &= \|F'(x)^{-1}F'(x)(x - x^*) - F'(x)^{-1}(F(x) - F(x^*))\| \\ &= \|F'(x)^{-1}(F(x^*) - F(x) - F'(x)(x^* - x))\|, \end{aligned}$$

para todo $x \in \Omega \cap B(x^*; \delta)$. Usando a propriedade (1.2), temos

$$\|N_F(x) - x^*\| \leq \|F'(x)^{-1}\| \|F(x^*) - F(x) - F'(x)(x^* - x)\|. \quad (3.5)$$

Pela diferenciabilidade de F , podemos definir

$$R(x^* - x) = \frac{F(x^*) - F(x) - F'(x)(x^* - x)}{\|x^* - x\|},$$

onde $\lim_{x \rightarrow x^*} R(x^* - x) = 0$. Assim, $\|N_F(x) - x^*\| \leq \|F'(x)^{-1}\| \|R(x^* - x)\| \|x^* - x\|$. Pelo Lema 3.2.1, também temos que existe $\gamma > 0$ tal que $\|F'(x)^{-1}\| \leq \gamma$ para todo $x \in B(x^*; \delta)$.

Segue disto e da última desigualdade que

$$\begin{aligned} \|N_F(x) - x^*\| &\leq \gamma \|R(x^* - x)\| \|x - x^*\| \\ \frac{\|N_F(x) - x^*\|}{\|x - x^*\|} &\leq \gamma \|R(x^* - x)\|. \end{aligned} \quad (3.6)$$

Como $\|N_F(x) - x^*\|/\|x - x^*\| \geq 0$, passando o limite quando $x \rightarrow x^*$ em (3.6), temos

$$\lim_{x \rightarrow x^*} \frac{\|N_F(x) - x^*\|}{\|x - x^*\|} = 0. \quad (3.7)$$

Isso significa que existe $\hat{\delta} > 0$, com $\delta \geq \hat{\delta}$, tal que

$$\|N_F(x) - x^*\| \leq \|x - x^*\|$$

para todo $x \in \Omega \cap B(x^*; \hat{\delta})$. Portanto, $N_F(x) \in \Omega \cap B(x^*; \hat{\delta})$. \square

Além disso, os lemas a seguir nos auxiliarão na análise da taxa de convergência quadrática da sequência gerada pelo Método de Newton.

Lema 3.2.3. *Se $F : \Omega \rightarrow \mathbb{R}^n$ é uma função continuamente diferenciável definida no aberto e convexo $\Omega \subset \mathbb{R}^n$ e $M = \sup_{\alpha \in [0,1]} \|F'(\alpha x + (1 - \alpha)y) - F'(x)\|$ com $x, y \in \Omega$, então*

$$\|F(x) - F(y) - F'(x)(x - y)\| \leq M\|x - y\|.$$

Demonstração. Fixando $x \in \Omega$, considere $g : \Omega \rightarrow \mathbb{R}^n$ definida por $g(z) = F(z) - F'(x)z$. Sabemos que todo operador linear é diferenciável, então F' é diferenciável, o que implica que g é diferenciável em Ω . Como Ω é convexo, temos $[x, y] \subset \Omega$ para qualquer $y \in \Omega$, então g é diferenciável em $[x, y]$. Além disso, $\|g'(z)\| = \|F'(z) - F'(x)\| = \|F'(\alpha x + (1 - \alpha)y) - F'(x)\| \leq \sup_{\alpha \in [0,1]} \|F'(\alpha x + (1 - \alpha)y) - F'(x)\| = M$, para todo $z \in [x, y]$. Assim, segue da Desigualdade do Valor Médio (veja o Teorema 1.2.31) que $\|F(x) - F(y) - F'(x)(x - y)\| = \|(F(x) - F'(x)x) - (F(y) - F'(x)y)\| = \|g(x) - g(y)\| \leq M\|x - y\|$. \square

Lema 3.2.4. *Se $F : \Omega \rightarrow \mathbb{R}^n$ é uma função continuamente diferenciável definida no aberto e convexo $\Omega \subset \mathbb{R}^n$ com a derivada de F Lipschitz contínua com constante $L > 0$, então, para todo $x, y \in \Omega$,*

$$\|F(x) - F(y) - F'(x)(x - y)\| \leq L\|x - y\|^2.$$

Demonstração. Fixando $x, y \in \Omega$, considere $M = L\|x - y\|$ e $g : \Omega \rightarrow \mathbb{R}^n$ definida por $g(z) = F(z) - F'(x)z$. Como Ω é convexo, g é diferenciável em $[x, y] \subset \Omega$. Além disso, como F' é Lipschitz contínua com $L > 0$, temos $\|g'(z)\| = \|F'(z) - F'(x)\| \leq L\|z - x\| \leq L\|x - y\| = M$, para todo $z \in [x, y]$. Portanto, segue da Desigualdade do Valor Médio que

$$\|F(x) - F(y) - F'(x)(x - y)\| = \|(F(x) - F'(x)x) - (F(y) - F'(x)y)\| = \|g(x) - g(y)\| \leq M\|x - y\| = L\|x - y\|^2. \quad \square$$

Para usarmos estes lemas na demonstração do Teorema 3.2.5, relembramos ao leitor que toda bola aberta é um conjunto convexo, como vimos no Exemplo 2.1.6. A seguir, provaremos a convergência local do Método de Newton. Em outras palavras, demonstraremos que a sequência gerada pelo Algoritmo 1 está bem definida e converge com taxa superlinear ou quadrática para a solução do problema (3.1).

Teorema 3.2.5 (Convergência local do Método de Newton). *Seja $F : \Omega \rightarrow \mathbb{R}^n$ uma função continuamente diferenciável definida no aberto $\Omega \subset \mathbb{R}^n$. Se $x^* \in \Omega$ é uma solução da equação (3.1), tal que a matriz $F'(x^*)$ é invertível, então para qualquer ponto inicial $x_0 \in \Omega$ suficientemente próximo a x^* , o Algoritmo 1 gera uma sequência $\{x_k\}$ bem definida que converge superlinearmente para x^* . Além disso, se F' for Lipschitz contínua numa vizinhança de x^* , então a taxa de convergência é quadrática.*

Demonstração. Para provarmos que a sequência $\{x_k\}$ gerada pelo Algoritmo 1 está bem definida, basta tomarmos $x_0 \in B(x^*; \hat{\delta}) \subset \Omega$, com $\hat{\delta} > 0$ dado pelo Lema 3.2.2, e definirmos $x_{k+1} = N_F(x_k)$. Segue disso e do Lema 3.2.1, que $F'(x_k)$ é invertível e $x_{k+1} \in B(x^*; \hat{\delta})$, para todo $k \in \mathbb{N}$. Logo, $\{x_k\}$ é uma sequência bem definida. Segue diretamente da equação (3.7) que

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

Portanto, para x_0 suficientemente próximo de x^* , a sequência $\{x_k\}$ converge superlinearmente para x^* (veja a Definição 1.2.16). Reescrevendo a equação (3.5) em termos de x_k , temos que existe $\gamma > 0$ tal que $\|F'(x_k)^{-1}\| \leq \gamma$, então

$$\begin{aligned} \|x_{k+1} - x^*\| &\leq \|F'(x_k)^{-1}\| \|F(x_k) - F(x^*) - F'(x_k)(x_k - x^*)\| \\ &\leq \gamma \|F(x_k) - F(x^*) - F'(x_k)(x_k - x^*)\|. \end{aligned}$$

Se F' é Lipschitz contínua em $B(x^*; \hat{\delta})$ com constante $L > 0$, usando o Lema 3.2.4 e a última desigualdade, obtemos

$$\|x_{k+1} - x^*\| \leq \gamma L \|x_k - x^*\|^2.$$

Conforme a Definição 1.2.19, isso significa que a taxa de convergência da sequência $\{x_k\}$ para a solução x^* é quadrática. \square

Após analisar a convergência local do Método de Newton para uma solução x^* do problema (3.4), surgem as seguintes questões: qual conjunto de pontos iniciais x_0 garante que a sequência gerada convergirá para x^* ? Como determinar esse conjunto? Conforme discutido no Exemplo 3.1.3, é crucial que o ponto inicial esteja “suficientemente próximo”

da solução. Assim, o conjunto formado por todos os pontos iniciais x_0 a partir dos quais a sequência $\{x_k\}$ gerada pelo Método de Newton converge para x^* é conhecido como a bacia de atração do Método de Newton para x^* . A representação gráfica das bacias de atração permite visualizar como o Algoritmo 1 se comporta para encontrar os zeros da função, considerando vários pontos iniciais distintos.

No Capítulo 4, discutiremos sobre as questões levantadas anteriormente e abordaremos a aplicação do Método de Newton para encontrar zeros de uma função $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ continuamente diferenciável. Este problema é equivalente a encontrar as soluções de um sistema não linear de duas equações. Assim, apresentaremos a sequência newtoniana e, por meio de uma proposição, determinaremos as bacias de atração do Método de Newton para as soluções deste problema.

Capítulo 4

Aplicações

As técnicas de Otimização Contínua e o Método de Newton são ferramentas extremamente importantes na Matemática Aplicada, pois muitos problemas práticos, quando modelados matematicamente, podem ser resolvidos por meio da minimização de funções ou da determinação de seus zeros.

Neste capítulo, buscamos contribuir com as teorias apresentadas nos Capítulos 2 e 3 explorando duas aplicações em problemas. Na primeira seção, destacamos a aplicação da busca de Armijo na minimização de uma função do tipo soma de quadrados. Já na segunda seção, determinamos as bacias de atração do Método de Newton na resolução de um sistema não linear de equações. Esta última aplicação foi apresentada na 76^a Reunião Anual da Sociedade Brasileira para o Progresso da Ciência (SBPC), [3].

4.1 Minimização irrestrita com a busca de Armijo

Relembramos ao leitor que a minimização de uma função objetivo é o processo de encontrar pontos que determinem o menor valor desta função. Em alguns casos, faz-se necessário utilizar buscas lineares que determinem um comprimento de passo que forneça um decréscimo suficiente na função objetivo. Essas buscas lineares podem influenciar diretamente na eficiência de algoritmos de minimização. Nesse sentido, diversas buscas lineares foram propostas na literatura matemática. Alguns exemplos incluem a busca da Seção Áurea, a busca de Armijo, entre outras, veja [11, Cap. 3, Sec. 3.1], [24, Cap. 3] e [26, Cap. 4, Sec. 4.2].

Nesta seção, analisaremos iterativamente o comportamento de decréscimo de uma função objetivo usando o comprimento de passo definido pela busca de Armijo e destacaremos as decisões tomadas com base nos comprimentos de passo avaliados. Por fim, apresentaremos o resultado obtido, evidenciando a sequência de pontos gerada por este processo e como eles se aproximam do mínimo global da função.

Inicialmente, consideremos a função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ definida por

$$f(x, y) = \frac{1}{2}(x - 2)^2 + (y - 1)^2. \quad (4.1)$$

Os cálculos do vetor gradiente e da matriz hessiana de f podem ser vistos no Exemplo 1.2.29. Por este exemplo, também temos que o ponto $p^* = (2, 1)$ é um zero do gradiente de f e a matriz hessiana $\nabla^2 f$ é definida positiva para qualquer ponto de \mathbb{R}^2 . Então, pelo Teorema 2.1.4, temos que p^* é um minimizador local estrito de f . Além disso, conforme o Teorema 2.1.8, a função f é convexa, o que implica, pelo Teorema 2.1.9, que $p^* = (2, 1)$ é um ponto de mínimo global de f .

Agora, iremos aplicar a busca de Armijo para minimizar a função f ao longo de direções de descida. Dados $p_0 = (1, 0)$ e $d_0 = (3, 1)$, pelo Exemplo 2.1.13 sabemos que d_0 é direção de descida a partir de p_0 . Segue da desigualdade de Armijo (2.3) que

$$\begin{aligned} f \begin{bmatrix} 1 + 3\alpha_0 \\ \alpha_0 \end{bmatrix} &\leq f \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \eta\alpha_0 \begin{bmatrix} -1 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} \\ 11\alpha_0^2 &\leq 10\alpha_0 - 10\eta\alpha_0 \\ \alpha_0 &\leq \frac{10(1 - \eta)}{11}. \end{aligned} \quad (4.2)$$

Note que, se tomarmos $\eta = 1/4$, então qualquer $\alpha_0 \leq 15/22$ assegura a desigualdade (4.2). Assim, começando com $\alpha_0 = 1 > 15/22$, obteremos $p_0 + \alpha_0 d_0 = (4, 1) = p_1$ e, conseqüentemente, $f(p_1) = 2 > 3/2 = f(p_0)$. Observe na Figura 4.1(a) as curvas de nível de f nos níveis $3/2$ e 2 e, na Figura 4.1(b), as imagens da função f nos pontos p^* , p_0 e p_1 .

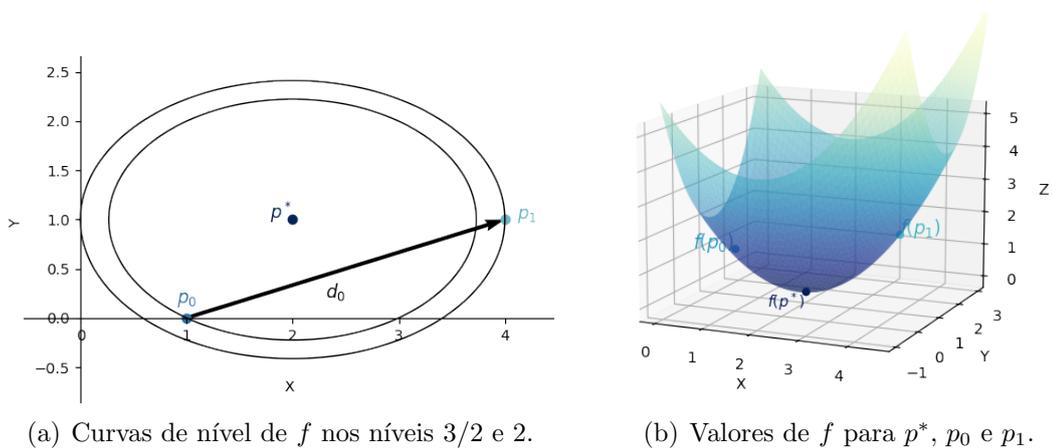


Figura 4.1: Curvas de nível e representação gráfica da função f .

Assim, o comprimento de passo $\alpha_0 = 1$ não fornece um decréscimo da função f em relação ao valor $f(p_0)$. Isso ocorre devido ao fato de que este comprimento de passo não pertence ao conjunto dos comprimentos de passo admissíveis, como pode ser observado na Figura 4.2. Rejeitando o comprimento de passo $\alpha_0 = 1$, calcularemos um novo p_1 .

Considerando $\alpha_0 = 16/25$, o passo será aceito, pois satisfaz (4.2) para $\eta = 1/4$; ou seja, esse passo pertence ao conjunto dos comprimentos de passo admissíveis.

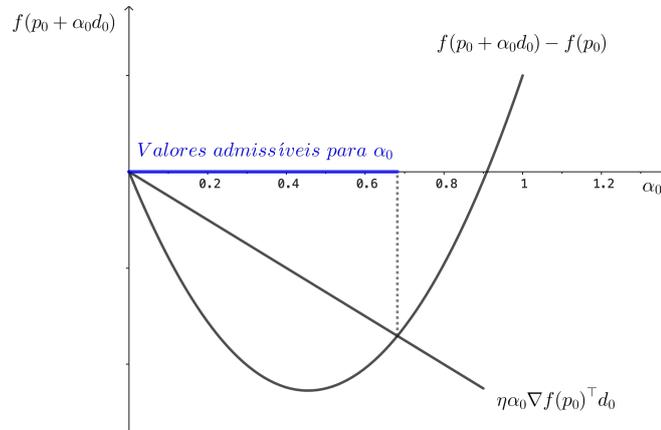


Figura 4.2: Conjunto dos comprimentos de passo admissíveis.

Considerando o novo comprimento de passo $\alpha_0 = 16/25$, temos $p_0 + \alpha_0 d_0 = (73/25, 16/25) = p_1$, de onde segue que $f(p_1) < f(p_0)$. Assim, o comprimento de passo $\alpha_0 = 16/25$ fornece um decréscimo suficiente da função f , a partir de p_0 na direção d_0 . Isso pode ser observado nas curvas de nível de f em $3/2$ e $691/1250$ apresentadas na Figura 4.3.

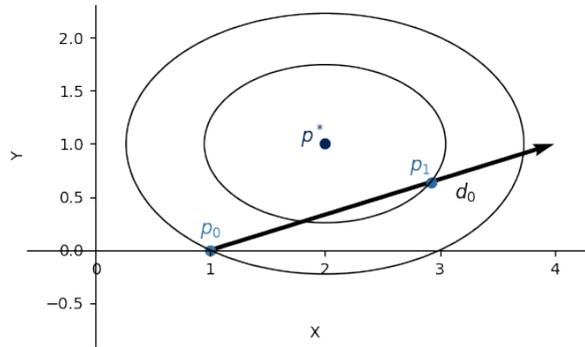


Figura 4.3: Curvas de nível de f em $3/2$ e $691/1250$.

Agora, vamos continuar decrescendo o valor da função f a partir da nova iterada p_1 na direção de descida $d_1 = (-1/2, 3/4)$. Note que d_1 é uma direção de descida de f a partir de p_1 , pois $\nabla f(p_1)^\top d_1 = (23/25, -18/25)^\top (-1/2, 3/4) = -1 < 0$. Usando a desigualdade de Armijo (2.3), temos

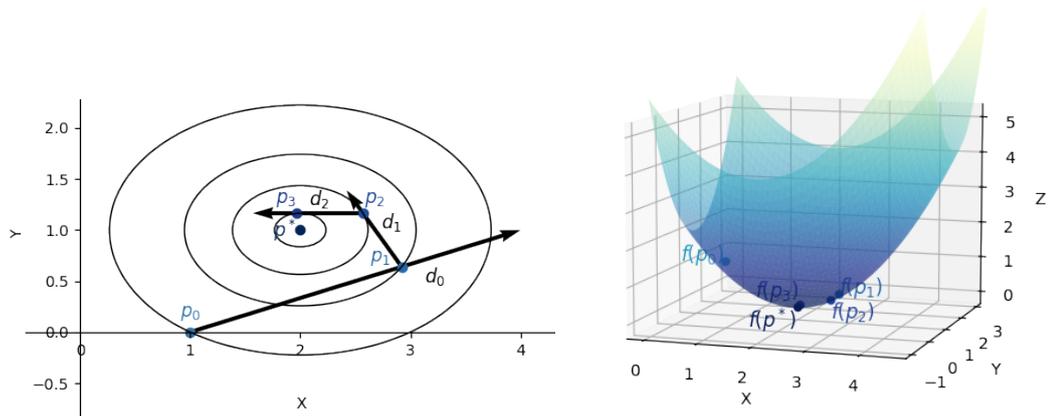
$$f \begin{bmatrix} (73/25) - (\alpha_1/2) \\ (16/25) + (3\alpha_1/4) \end{bmatrix} \leq f \begin{bmatrix} 73/25 \\ 16/25 \end{bmatrix} + \eta \alpha_1 \begin{bmatrix} 23/25 & -18/25 \end{bmatrix} \begin{bmatrix} -1/2 \\ 3/4 \end{bmatrix}$$

$$\alpha_1 \leq \frac{16(1 - \eta)}{11}.$$

Tomando $\eta = 1/2$, basta fazermos $\alpha_1 = 7/10$ e teremos $p_1 + \alpha_1 d_1 = (257/100, 233/200) = p_2$, de onde segue que $f(p_2) < f(p_1)$. Analogamente aos procedimentos anteriores, fazemos a última aplicação da busca de Armijo a partir de $p_2 = (257/100, 233/200)$, na

direção de descida $d_2 = (-1, 0)$. Tomando $\eta = 1/3$, basta fazermos $\alpha_2 = 6/10$ e teremos $p_2 + \alpha_2 d_2 = (197/100, 233/200) = p_3$, de onde segue que $f(p_3) < f(p_2)$. Portanto, temos $f(p_{k+1}) < f(p_k)$ para $k = 0, 1, 2, 3$.

Na Figura 4.4(a), podemos observar as curvas de nível de f nos níveis $3/2$, $691/1250$, $7587/40000$ e $1107/40000$. Além disso, na Figura 4.4(b) podemos observar as imagens da função f nos pontos p^* , p_0 , p_1 , p_2 e p_3 .



(a) Curvas de nível da função f para os pontos da sequência gerada pela busca de Armijo ao longo de direções de descida.

(b) Valores de f para os pontos da sequência gerada pela busca de Armijo ao longo de direções de descida.

Figura 4.4: Curvas de nível e representação gráfica da função f .

Ademais, a Tabela 4.1 apresenta a sequência dos pontos obtidos usando a busca de Armijo e as direções de descida utilizadas para a minimização da função definida em (4.1). Nesta tabela, k é o índice do termo da sequência, p_k é o termo da sequência, d_k é a direção de descida, α_k é o comprimento de passo e $f(p_k)$ é o valor da função.

k	p_k	d_k	α_k	$f(p_k)$
0	(1, 0)	(3, 1)	16/25	3/2
1	(73/25, 16/25)	(-1/2, 3/4)	7/10	691/1250
2	(257/100, 233/200)	(-1, 0)	6/10	7587/40000
3	(197/100, 233/200)	—	—	1107/40000
\vdots	\vdots	\vdots	\vdots	\vdots
?	(2, 1)	?	?	0

Tabela 4.1: Iteradas, direções de descida, comprimentos de passo e valores de f .

Na última coluna, $f(p_k)$ está diminuindo a cada iteração, indicando que $f(p_k)$ está cada vez mais próximo de 0, que corresponde ao valor de mínimo global da função f . Por fim, observe que preenchemos algumas células da última linha dessa tabela com o símbolo de interrogação. Isso decorre da abordagem adotada neste trabalho, onde buscamos elucidar a busca de Armijo de forma algébrica. Entretanto, ao empregar a busca de Armijo em métodos numéricos, é possível determinar o número de iterações para obter uma aproximação satisfatória da solução, com base em critérios de convergência da sequência.

4.2 As bacias de atração do Método de Newton para um sistema não linear de equações

O Método de Newton clássico assegura que, dado um ponto inicial “suficientemente próximo” da solução, uma sequência bem definida é gerada, convergindo para a solução do problema, veja o Teorema 3.2.5. A fim de determinar quão próximo da solução o ponto inicial deve estar, é estudada a convergência dessa sequência sob as condições do Teorema de Kantorovich ou condições relacionadas. Essa abordagem é conhecida como convergência semilocal, veja, por exemplo, [12, 22, 29]. Além disso, busca-se mapear as bacias de atração do Método de Newton por meio de simulações numéricas. Essas simulações possibilitam a visualização do comportamento das sequências geradas pelo Método de Newton a partir de um conjunto limitado de pontos iniciais, veja [7, 28].

Nesta seção, iremos determinar as bacias de atração do Método de Newton para um sistema não linear de equações. Para isso, vamos determinar a sequência newtoniana, provar a convergência dessa sequência para as soluções do sistema e, por fim, determinar as bacias de atração do problema.

Consideremos o sistema não linear de equações dado por

$$\begin{cases} x - y = -1 \\ -x^2 + y = 1 \end{cases}, \quad (4.3)$$

que admite duas soluções: $s_1 = (0, 1)$ e $s_2 = (1, 2)$. As curvas que representam as duas equações do sistema acima podem ser vistas na Figura 4.5.

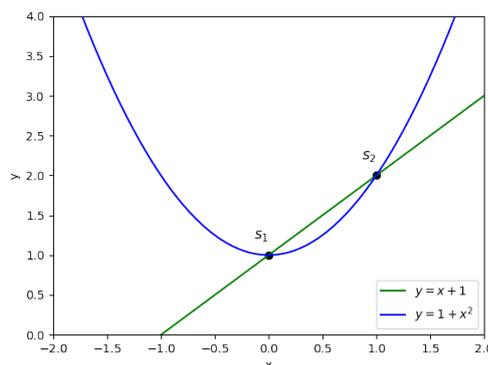


Figura 4.5: Curvas que representam as equações do sistema (4.3).

Encontrar as soluções do sistema (4.3) é equivalente à resolução do problema (3.1) para a função $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ definida por $F(x, y) = (x - y + 1, -x^2 + y - 1)$. Note que a função F é continuamente diferenciável, pois suas coordenadas são continuamente diferenciáveis. Assim, considerando $p_k = (x_k, y_k)$ e calculando $F'(p_k)^{-1}$, após algumas manipulações algébricas, chegamos ao resultado que devemos ter $x_k \neq 1/2$, pois F' não

é invertível para este caso e, portanto, a sequência newtoniana não seria bem definida. Pelo esquema iterativo (3.3), desde que $x_k \neq 1/2$ para todo $k \in \mathbb{N}$, obtemos

$$\begin{aligned}
p_{k+1} &= p_k - F'(p_k)^{-1}F(p_k) \\
&= \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} 1/(1-2x_k) & 1/(1-2x_k) \\ 2x_k/(1-2x_k) & 1/(1-2x_k) \end{bmatrix} \begin{bmatrix} x_k - y_k + 1 \\ -x_k^2 + y_k - 1 \end{bmatrix} \\
&= \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} (-x_k^2 + x_k)/(1-2x_k) \\ (x_k^2 - 2x_k y_k + 2x_k + y_k - 1)/(1-2x_k) \end{bmatrix} \\
&= \begin{bmatrix} x_k^2/(2x_k - 1) \\ [x_k^2/(2x_k - 1)] + 1 \end{bmatrix} \\
&:= \begin{bmatrix} p^{(k+1)1} \\ p^{(k+1)2} \end{bmatrix}.
\end{aligned} \tag{4.4}$$

Pelo Teorema 1.2.13, para provarmos que a sequência $\{p_k\}$ converge, temos que provar que as sequências das coordenadas $\{p^{(k)1}\}$ e $\{p^{(k)2}\}$ são convergentes. Assim, uma ideia fundamental para a prova de convergência dessa sequência é a da Observação 4.2.1.

Observação 4.2.1. *Pela sequência newtoniana definida em (4.4), temos*

$$p^{(k)1} - p^{(k+1)1} = x_k - \frac{x_k^2}{2x_k - 1} = \frac{2x_k^2 - x_k - x_k^2}{2x_k - 1} = \frac{x_k^2 - x_k}{2x_k - 1}, \tag{4.5}$$

desde que $x_k \neq 1/2$ para todo $k \in \mathbb{N}$. Além disso, temos que:

- (i) se $p_{(0)1} \in (-\infty, 0) \cup (0, 1/2)$, então $p^{(k)1} \in (-\infty, 0)$ para todo $k \geq 1$. Neste caso, analisando o sinal do quociente (4.5), temos que $\{p^{(k)1}\}$ será monótona crescente;
- (ii) se $p_{(0)1} \in (1/2, 1) \cup (1, +\infty)$, então $p^{(k)1} \in (1, \infty)$ para todo $k \geq 1$. Neste caso, analisando o sinal do quociente (4.5), temos que $\{p^{(k)1}\}$ será monótona decrescente;
- (iii) se $p_{(0)1} = 0$ ou $p_{(0)1} = 1$, então $\{p^{(k)1}\}$ será constante para todo $k \in \mathbb{N}$.

A partir destes resultados, a análise de convergência da sequência newtoniana e as bacias de atração do Método de Newton para as soluções do sistema são determinadas pela Proposição 4.2.2, que é provada diretamente pelos Lemas 4.2.3, 4.2.4 e 4.2.5.

Proposição 4.2.2. *Se o ponto inicial for tomado no semiplano $\{p = (x, y) \in \mathbb{R}^2 : x < 1/2\}$, então a sequência definida por (4.4) converge para $s_1 = (0, 1)$. Se o ponto inicial for tomado no semiplano $\{p = (x, y) \in \mathbb{R}^2 : x > 1/2\}$, então a sequência definida por (4.4) converge para $s_2 = (1, 2)$.*

Lema 4.2.3. *Se $p_{(0)1} < 1/2$, então $p^{(k)1} \rightarrow 0$ quando $k \rightarrow \infty$.*

Demonstração. Supondo que $p_{(0)1} < 0$, da definição de $p^{(k)1}$ para $k \geq 1$, temos que $\{p^{(k)1}\}$ é limitada superiormente por 0. Como $p_{(0)1} \in (-\infty, 0)$, de acordo com a Observação

4.2.1 (item (i)), temos que $\{p_{(k)1}\}$ é monótona crescente. Pelo Teorema 1.2.9 (a), essa sequência é convergente e converge para o seu supremo. Suponha, por absurdo, que zero não é o supremo de $\{p_{(k)1}\}$. Então existe $c > 0$ tal que $-c$ é uma cota superior, isto é, temos $p_{(k+1)1} = x_k^2/(2x_k - 1) < -c$, para todo $k \in \mathbb{N}$. Note que, se tomarmos $p_{(0)1} = x_0 \in (-c - \sqrt{c^2 + c}, -c)$, temos

$$p_{(1)1} + c = \frac{x_0^2}{2x_0 - 1} + c = \frac{x_0^2 + 2cx_0 - c}{2x_0 - 1} > 0. \quad (4.6)$$

De fato, analisando o numerador do quociente (4.6) quando $x_0 \in (-c - \sqrt{c^2 + c}, -c)$, temos $-\sqrt{c^2 + c} < x_0 + c < 0$ e, conseqüentemente,

$$\begin{aligned} |x_0 + c| < |-\sqrt{c^2 + c}| &\implies |x_0 + c|^2 < |-\sqrt{c^2 + c}|^2 \\ &\implies (x_0 + c)^2 < c^2 + c \\ &\implies (x_0^2 + 2cx_0 + c^2) - c^2 - c < 0 \\ &\implies x_0^2 + 2cx_0 - c < 0. \end{aligned}$$

Por outro lado, analisando o denominador de (4.6), como $x_0 < -c \implies 2x_0 < -2c \implies 2x_0 - 1 < -2c - 1 \leq 0$, pois $c > 0$. Logo, $p_{(1)1} > -c$, o que é um absurdo, pois estamos supondo $p_{(k+1)1} < -c$, para todo $k \in \mathbb{N}$. Portanto, zero é o supremo de $\{p_{(k)1}\}$ e se $p_{(0)1} < 0$, então $p_{(k)1} \rightarrow 0$ quando $k \rightarrow \infty$.

Agora, resta mostrar que, se $0 \leq p_{(0)1} < 1/2$, então $p_{(k)1} \rightarrow 0$ quando $k \rightarrow \infty$. No caso em que $p_{(0)1} = 0$, o resultado é imediato, uma vez que a sequência $\{p_{(k)1}\}$ será constante. No caso $0 < p_{(0)1} < 1/2$, pela Observação 4.2.1 (item (i)), temos que $p_{(k)1} \in (-\infty, 0)$ para todo $k \geq 1$. Portanto, $\{p_{(k)1}\}$ é limitada superiormente por 0 e é monótona crescente para todo $k \geq 1$. De forma análoga à demonstração do caso em que $p_{(0)1} < 0$, segue que $p_{(k)1} \rightarrow 0$ quando $k \rightarrow \infty$. \square

Lema 4.2.4. *Se $p_{(0)1} > 1/2$, então $p_{(k)1} \rightarrow 1$ quando $k \rightarrow \infty$.*

Demonstração. Supondo que $p_{(0)1} > 1$, da definição de $p_{(k)1}$ temos que $\{p_{(k)1}\}$ é limitada inferiormente por 1. Como $p_{(0)1} \in (1, +\infty)$, de acordo com a Observação 4.2.1 (item (ii)), temos que $\{p_{(k)1}\}$ é monótona decrescente. Pelo Teorema 1.2.9 (b), esta sequência é convergente e converge para o seu ínfimo. Suponha, por absurdo, que 1 não é o ínfimo de $\{p_{(k)1}\}$. Então existe uma cota inferior $c > 1$ tal que $p_{(k+1)1} = x_k^2/(2x_k - 1) > c$, para todo $k \in \mathbb{N}$. Porém, de forma análoga ao que fizemos na demonstração anterior, se tomarmos $p_{(0)1} = x_0 \in (c, c + \sqrt{c^2 - c})$, teremos

$$p_{(1)1} - c = \frac{x_0^2}{2x_0 - 1} - c = \frac{x_0^2 - 2cx_0 + c}{2x_0 - 1} < 0.$$

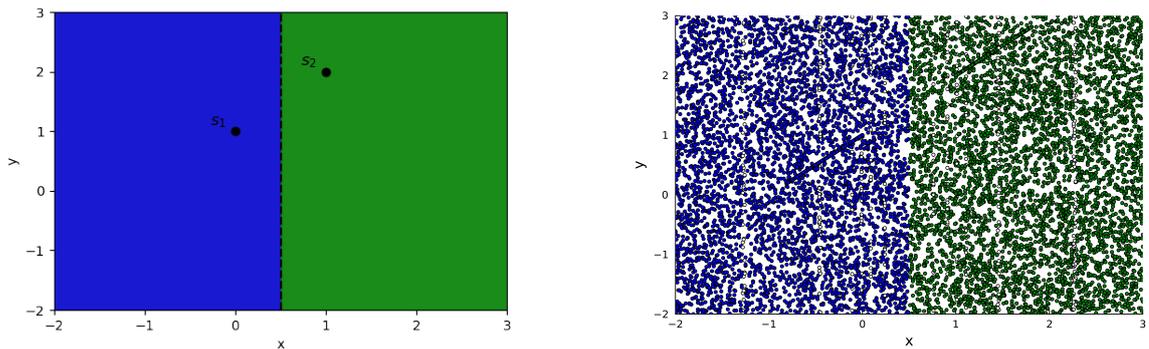
Isso implica $p_{(1)1} < c$, o que é um absurdo, pois estamos supondo $p_{(k+1)1} > c$, para todo

$k \in \mathbb{N}$. Portanto, 1 é o ínfimo de $\{p_{(k)1}\}$ e $p_{(k)1} \rightarrow 1$ quando $k \rightarrow \infty$. Agora, resta mostrar que, se $1/2 < p_{(0)1} \leq 1$, então $p_{(k)1} \rightarrow 1$ quando $k \rightarrow \infty$. No caso em que $p_{(0)1} = 1$, o resultado é imediato, uma vez que a sequência $\{p_{(k)1}\}$ será constante. No caso $1/2 < p_{(0)1} < 1$, pela Observação 4.2.1 (item (i)), temos que $p_{(k)1} \in (1, +\infty)$, com $k \geq 1$. Portanto, $\{p_{(k)1}\}$ é limitada inferiormente por 1 e é monótona decrescente para todo $k \geq 1$. De forma análoga à demonstração do caso em que $p_{(0)1} > 1$, segue que $p_{(k)1} \rightarrow 1$ quando $k \rightarrow \infty$. \square

Lema 4.2.5. *Se $p_{(0)1} < 1/2$, então $p_{(k)2} \rightarrow 1$ quando $k \rightarrow \infty$; e se $p_{(0)1} > 1/2$, então $p_{(k)2} \rightarrow 2$ quando $k \rightarrow \infty$.*

Demonstração. Note que, por (4.4), podemos escrever $p_{(k+1)2} = p_{(k+1)1} + 1$. Assim, temos $\lim_{k \rightarrow \infty} p_{(k+1)2} = \lim_{k \rightarrow \infty} (p_{(k+1)1} + 1) = (\lim_{k \rightarrow \infty} p_{(k+1)1}) + 1$, caso $\lim_{k \rightarrow \infty} p_{(k+1)1}$ exista. Portanto, pelos Lemas 4.2.3 e 4.2.4, temos o resultado desejado. \square

Abaixo, apresentamos duas figuras que representam graficamente os resultados obtidos nessa aplicação do Método de Newton. A Figura 4.6(a), apresenta as bacias de atração obtidas a partir da Proposição 4.2.2. Já a Figura 4.6(b), apresenta as bacias de atração do Método de Newton quando obtidas numericamente. Para isso, implementamos o Algoritmo 1 usando a linguagem de programação Julia em um processador Intel Core i5 Dual-Core de 1,4 GHz, 8 GB de RAM e sistema operacional macOS Big Sur. Foram gerados 10000 pontos iniciais por meio do comando *rand*. Consideramos convergência na iteração k quando $\|F(p_k)\| \leq 10^{-8}$ e critério de parada de máximo de 1000 iterações. As implementações realizadas estão livremente disponibilizadas nos Apêndices A.



(a) Os semiplanos em azul e verde, obtidos por meio da Proposição 4.2.2, representam as bacias de atração para $s_1 = (0, 1)$ e $s_2 = (1, 2)$.

(b) Os pontos em azul e verde, obtidos numericamente, representam as bacias de atração para $s_1 = (0, 1)$ e $s_2 = (1, 2)$, respectivamente.

Figura 4.6: Bacias de atração do Método de Newton para o problema (4.3).

Na Figura 4.1(b), é possível observar que, apesar de utilizarmos 10000 pontos iniciais, a região gerada não é suficiente para definir exatamente as bacias de atração, fornecendo apenas indícios das bacias obtidas por meio da Proposição 4.2.2. Contudo, obtivemos as bacias exatas, conforme apresentado na Figura 4.1(a).

Considerações finais

Neste trabalho, investigamos dois problemas: a minimização de funções e a determinação de seus zeros, ambos com ampla aplicação em diversas áreas das ciências. Analisamos a convergência da sequência gerada pelo Método de Newton para um zero de uma função diferenciável e apresentamos uma contribuição na determinação das bacias de atração deste método. Especificamente, construímos um exemplo de um sistema não linear de equações, determinamos a sequência newtoniana, provamos a convergência dessa sequência para as soluções do sistema e estabelecemos, por meio de uma proposição, as bacias de atração do Método de Newton, o que não é abordado na literatura atual. Para obtermos este resultado encontramos maior dificuldade na construção do sistema não linear de equações, uma vez que resolver este sistema deveria ser equivalente a assegurar a convergência da sequência newtoniana associada. Além disso, abordamos o processo de minimização irrestrita de uma função objetivo do tipo soma de quadrados. Para isso, buscamos elucidar algebricamente a técnica da busca de Armijo, que é uma abordagem amplamente utilizada em métodos numéricos da Otimização Contínua. Para obter esses resultados foi necessário o estudo de Álgebra Linear, Análise Matemática e Otimização Contínua. Para pesquisas futuras, seria interessante a generalização da determinação das bacias de atração do Método de Newton para sistemas de equações não lineares quaisquer.

Referências Bibliográficas

- [1] ARENALES, Selma; DAREZZO, Arthur. **Cálculo Numérico: aprendizagem com apoio de software**. São Paulo: Thonsom Learning, 2008.
- [2] ARMIJO, Larry. Minimization of functions having Lipschitz continuous first partial derivatives. **Pacific Journal of mathematics**, v. 16, n. 1, p. 1-3, 1966.
- [3] As bacias de atração do Método de Newton para um sistema de equações não lineares. [S.I.]: SBPCnet, 2024. 76ª RA - Sessão de Pôsteres. Disponível em: <https://www.youtube.com/watch?v=EMNzY04e2ng>. Acesso em: 30 jul. 2024.
- [4] BARTLE, Robert G.; SHERBERT, Donald R. **Introduction to real analysis**. 3 ed. New York: Wiley, 2000.
- [5] CARDANO, Girolamo; *et al.* **The rules of algebra: Ars Magna**. Courier Corporation, 2007.
- [6] FERNANDES, Teles A.; FERREIRA, Orizon P.; YUAN, Jinyun. On the superlinear convergence of Newton's method on Riemannian manifolds. **Journal of Optimization Theory and Applications**, v. 173, p. 828-843, 2017.
- [7] GUTIÉRREZ, José M. *et al.* Influence of the multiplicity of the roots on the basins of attraction of Newton's method. **Numerical Algorithms**, v. 66, p. 431-455, 2014.
- [8] HELMKE, Uwe; RICARDO, Sandra; YOSHIZAWA, Shintaro. **Newton's algorithm in Euclidean Jordan algebras, with applications to robotics**. Communications in Information and Systems, v. 2, n. 3, p. 283-298, 2002.
- [9] HUMES, Ana Flora P. de Castro; *et al.* **Noções de cálculo numérico**. São Paulo: McGraw-Hill do Brasil, 1984.
- [10] IZMAILOV, Alexey; SOLODOV, Mikhail. **Otimização, volume 1**. 4 ed. Rio de Janeiro: IMPA, 2020. 256 p. (Coleção Matemática Universitária).
- [11] IZMAILOV, Alexey; SOLODOV, Mikhail. **Otimização, volume 2: Métodos computacionais**. Rio de Janeiro: IMPA, 2018. 494 p. (Coleção Otimização).

- [12] KANTOROVICH, L. V. The majorant principle and Newton's method. In: **Dokl. Akad. Nauk SSSR**. 1951. p. 17-20. cni
- [13] KREYSZIG, Erwin. **Introductory functional analysis with applications**. John Wiley & Sons, 1978.
- [14] LANG, Serge. **Linear algebra**. New York: Springer Science & Business Media, 1987.
- [15] LAUWENS, Ben; DOWNEY, Allen B. **Think Julia: how to think like a computer scientist**. O'Reilly Media, 2019.
- [16] LIMA, Elon Lages. **Curso de análise, vol. 1**. 15. ed. Rio de Janeiro: Impa, 2022. 320 p. (Coleção Projeto Euclides).
- [17] LIMA, Elon Lages. **Curso de análise, vol. 2**. 12. ed. Rio de Janeiro: Impa, 2023. 452 p. (Coleção Projeto Euclides).
- [18] LIMA, Elon Lages; *et al.* A equação do Terceiro Grau. **Matemática Universitária**, Rio de Janeiro, n. 5, p. 9-23, jun. 1987.
- [19] MANGASARIAN, Olvi L. **A Newton method for linear programming**. Journal of Optimization Theory and Applications, v. 121, p. 1-18, 2004.
- [20] NEWTON, Isaac. **De analysi per aequationes numero terminorum infinitas**. 1711.
- [21] NEWTON, Isaac; COLSON, John. **The method of fluxions and infinite series: with its application to the geometry of curve-lines**. Nourse, 1736.
- [22] ORTEGA, James M. **The newton-kantorovich theorem**. The American Mathematical Monthly, v. 75, n. 6, p. 658-660, 1968.
- [23] ORTEGA, James M.; RHEINBOLDT, Werner C. **Iterative solution of nonlinear equations in several variables**. Society for Industrial and Applied Mathematics, 2000.
- [24] PANONCELI, Diego Manoel. **Um estudo de buscas unidirecionais aplicadas ao método BFGS**. 2015. 117 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Matemática, Universidade Federal do Paraná, Curitiba, 2015.
- [25] RAPHSOON, Joseph. **Analysis aequationum universalis seu ad aequationes algebraicas resolvendas methodus generalis, et expedita, ex nova inflnitarum serierum doctrina deducta ac demonstrata**. Londres, 1690, Microfilm cópia: University Microfilms, Ann Arbor, MI.

-
- [26] RIBERO, Ademir Alves; KARAS, Elizabeth Wegner. **Otimização Contínua: aspectos teóricos e computacionais**. São Paulo: Cengage Learning, 2013.
- [27] SIMPSON, Thomas. **Essays on Several Curious and Useful Subjects, in Speculative and Mix'd Mathematicks**. H. Woodfall, jun. 1740.
- [28] SUSANTO, Hadi; KARJANTO, Natanael. Newton's method's basins of attraction revisited. **Applied Mathematics and Computation**, v. 215, n. 3, p. 1084-1090, 2009.
- [29] YAMAMOTO, Tetsuro. A method for finding sharp error bounds for Newton's method under the Kantorovich assumptions. **Numerische Mathematik**, v. 49, p. 203-220, 1986.
- [30] YPMA, Tjalling J. Historical development of the Newton–Raphson method. **SIAM review**, v. 37, n. 4, p. 531-551, 1995.

Apêndice A

Implementações

Todas as implementações apresentadas a seguir foram realizadas na Linguagem de programação Julia, [15]. Os pacotes utilizados incluem LinearAlgebra para operações de Álgebra Linear, DataFrames para manipulação de dados tabulares, Random para geração de números aleatórios, Printf para formatação de saída e Plots para criação de gráficos.

A.1 Método de Newton

A seguir, apresentamos o código do Método de Newton para encontrar zeros de funções diferenciáveis de n variáveis.

```

1      # Newton's method to find zero of functions
2      #           of several variables
3
4      function newtonmethod(x,F,J_F,epsilon,maxiter)
5
6      Fnorm = Float64[]
7      info = DataFrame()
8      seqx = x
9      iter = 0
10     t0 = time()
11
12     while true
13         Fx = F(x)
14         normFx = norm(F(x))
15         push!(Fnorm, normFx)
16         invJ_Fx = inv(J_F(x))
17
18         if normFx < epsilon
19             info.norm = Fnorm
20             error=0

```

```

21         etime = time() - t0
22         println("iter = $iter normF = $(norm(F(x)))  F(x) = $(F
(x)) time = $etime")
23         println("Solutions has found!")
24         return(x,error,info,seqx,etime)
25     end
26
27     iter += 1
28
29     if iter > maxiter
30         info.norm = Fnorm
31         error=1
32         etime = time() - t0
33         println("Maximum of iterations was achieved!")
34         return(x,error,info,seqx,etime)
35     end
36
37     d = - invJ_Fx * Fx
38     x = x + d
39     seqx = [seqx x]
40     end
41 end
42

```

Por fim, apresentamos o código que gera as bacias de atração do Método de Newton na determinação dos zeros de uma função vetorial. Para a compilação deste código, faz-se necessário que o código anterior esteja salvo em um arquivo nomeado “newtonmethod.jl”, devido ao uso do comando *include*.

```

1     # Application of Newton's Method to determine basins of
2     # attraction of a system of nonlinear equations
3
4     using LinearAlgebra, DataFrames, Random, Printf, Plots
5
6     include("newtonmethod.jl")
7
8     eTIME = Float64[]
9     ITERS = Float64[]
10    Ierror = Float64[]
11    solinfo = DataFrame()
12
13    num_points = 10000
14    initial_points = [[rand([-2:0.01:0.49; 0.51:0.01:3]), rand

```

```
(-2:0.01:3)] for _ in 1:num_points]
15     nguess = length(initial_points)
16
17     function F(x)
18         f = [x[1] - x[2] + 1; -(x[1])^2 + x[2] - 1]
19         return f
20     end
21
22     function J_F(x)
23         j = [(1) (-1); (-2 * x[1]) (1)]
24         return j
25     end
26
27     println("-----")
28     println("Iter      CPU Time      Error Code      Last Point
29     Norm of F      Initial Point")
30     println("-----")
31
32     epsilon = 1.e-8
33     maxiter = 1000
34
35     for k in 1:nguess
36         local x0 = initial_points[k]
37         local x,error,info,seqx,etime = newtonmethod(x0,F,J_F,
38         epsilon,maxiter)
39
40         if error > 0
41             push!(eTIME, Inf)
42             push!(Iters, -1)
43             push!(Ierror, error)
44         else
45             push!(eTIME, etime)
46             push!(Iters, (size(info, 1) - 1))
47             push!(Ierror, error)
48         end
49
50         local last_point = seqx[:, end]
51         local F_norm = norm(F(last_point), 2)
52
53         @printf("%5d %12.8e %4d %s %12.8e %s\n", Iters[end],
54         eTIME[end], error, string(last_point), F_norm, string(x0))
```

```
52     end
53
54     solinfo.iters = ITERS
55     solinfo.time = eTIME
56     solinfo.error = Ierror
57     solinfo.initial_point = initial_points
58     println("-----")
59
60     bacia_x = Float64[]
61     bacia_y = Float64[]
62     bacia_color = Symbol[]
63
64     for k in 1:nguess
65         local x0 = initial_points[k]
66         local x,error,info,seqx,etime = newtonmethod(x0,F,J_F,
67             epsilon,maxiter)
68
69         if norm([0, 1] - x) < epsilon
70             color = :blue
71         elseif norm([1, 2] - x) < epsilon
72             color = :green
73         else
74             color = :white
75         end
76
77         for i in 1:length(seqx[1, :])
78             push!(bacia_x, seqx[1, i])
79             push!(bacia_y, seqx[2, i])
80             push!(bacia_color, color)
81         end
82     end
83
84     scatter(bacia_x, bacia_y, c=bacia_color, marker=:o,
85         markersize=2, xlabel="x", ylabel="y", legend=false, xlims=(-2,
86         3), ylims=(-2, 3), dpi=1000)
```