



UNIVERSIDADE ESTADUAL PAULISTA

A large, abstract geometric design in shades of blue and white. It features a central circle with several overlapping triangles and lines, creating a complex, star-like pattern. The design is set against a white background with blue corner accents.

**PROGRAMA DE
PÓS-GRADUAÇÃO EM
MATEMÁTICA**

Métodos Computacionais de Otimização

Bruna Alves Ferraz

INSTITUTO DE GEOCIÊNCIAS E CIÊNCIAS EXATAS

RIO CLARO/SP

2017



UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
Instituto de Geociências e Ciências Exatas
Campus de Rio Claro

Métodos Computacionais de Otimização

Bruna Alves Ferraz

Dissertação apresentada ao Programa de Pós-Graduação – Mestrado Profissional em Matemática como requisito parcial para a obtenção do grau de Mestre

Orientadora
Profa. Dra. Suzete Maria Silva Afonso

2017

510.285 Ferraz, Bruna Alves
F381m Métodos computacionais de otimização / Bruna Alves
Ferraz. - Rio Claro, 2018
73 f. : il., figs., gráfs., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista,
Instituto de Geociências e Ciências Exatas
Orientadora: Suzete Maria Silva Afonso

1. Matemática - Programas de computador. 2. Otimização
irrestrita. 3. Métodos. 4. Computacionais. I. Título.

TERMO DE APROVAÇÃO

Bruna Alves Ferraz

MÉTODOS COMPUTACIONAIS DE OTIMIZAÇÃO

Dissertação APROVADA como requisito parcial para a obtenção do grau de Mestre no Curso de Pós-Graduação Mestrado Profissional em Matemática Universitária do Instituto de Geociências e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, pela seguinte banca examinadora:

Profa. Dra. Suzete Maria Silva Afonso
Orientadora

Profa. Dra. Renata Zotin Gomes de Oliveira
Departamento de Matemática - Unesp

Profa. Dra. Ligia Laís Fêmina
Faculdade de Matemática / UFU - Uberlândia/MG

Rio Claro, 19 de dezembro de 2017

Às Leis do Universo, as quais me trouxeram até aqui.

Agradecimentos

Agradeço aos meus familiares pela compreensão e apoio durante a minha transição profissional. Agradeço aos professores do Departamento de Matemática da Unesp Rio Claro, pela didática, pelo incentivo e cobrança; em especial à minha orientadora, Suzete Maria Silva Afonso, pela paciência e disposição em esclarecer minhas infinitas dúvidas.

Resumo

Neste trabalho discutiremos alguns métodos clássicos para otimização irrestrita, a saber o Método de Cauchy e o Método de Newton, e analisaremos a convergência desses métodos. Veremos que o Método de Cauchy, que faz a cada iteração uma busca unidirecional na direção de máxima descida, ou seja, na direção oposta ao gradiente, tem convergência linear. O método de Newton, por outro lado, minimiza, em cada iteração, a aproximação quadrática da função objetivo. Nos métodos de busca unidirecional é preciso minimizar uma função a partir de um certo ponto, segundo uma direção dada, que é a direção de busca. Por essa razão, estudaremos o Método da Seção Áurea, que fornece uma minimização exata de uma função real de uma variável real.

Palavras-chave: Otimização Irrestrita, Métodos de Otimização, Convergência.

Abstract

In this work we will discuss some classic methods for unrestricted optimization, namely the Cauchy Method and Newton's Method, and we will analyze the convergence of those methods. We will see that the Cauchy Method, that realizes on each iteration a unidirectional search in the direction of maximum descent, that is, in the direction opposite to the gradient, has linear convergence. The Newton Method, on the other hand, minimizes, in each iteration, the quadratic approximation of the objective function. In unidirectional search methods, one must minimize a function from a certain point in a given direction, which is the search direction. For that reason, we will study the Golden Section Method, which provides the exact minimization of a real function of a real variable.

Keywords: Unrestricted Optimization, Optimization Methods, Convergence.

Lista de Figuras

1.1	Termos da sequência (x_k) na reta	20
1.2	Termos da sequência (x_k) na reta	21
1.3	Valores de $r_k \times a_k$	24
1.4	Funções convexa e não convexa	30
3.1	Passos do Algoritmo de Cauchy	37
3.2	Passos do Algoritmo de Cauchy	42
4.1	Uma iteração do Método de Newton para equações em $\mathbb{R}$	44
4.2	Uma iteração do Método de Newton	47
5.1	Busca unidimensional exata	53
5.2	Funções unimodais	55
5.3	Seção Áurea	56
5.4	Intervalo dividido em três partes iguais	56
5.5	Partição do intervalo - Lema 5.4	59
5.6	Partição do intervalo - Lema 5.5	59
5.7	Análise da primeira etapa do Algoritmo da Seção Áurea	62
5.8	Análise do item i) do Teorema 5.7	63
5.9	Intervalos que contêm o minimizador de uma função quadrática	65

Sumário

Introdução	17
1 Preliminares	19
1.1 Convergência de sequências	19
1.2 Ordem de convergência	24
1.3 Revisão de conceitos e resultados de Álgebra Linear	27
1.4 Mais alguns conceitos e resultados importantes de Análise no $\mathbb{R}^n$. .	28
2 Introdução à Otimização	31
3 Método de Cauchy	35
3.1 Algoritmo	35
3.2 Convergência	37
3.3 Velocidade de convergência	38
4 Método de Newton	43
4.1 Método de Newton para resolução de equações	43
4.2 Método de Newton para otimização irrestrita	44
4.3 Algoritmo	45
4.4 Convergência	47
5 Método da Seção Áurea	53
5.1 Algoritmo	60
5.2 Convergência	63
5.3 Velocidade de convergência	66
Conclusão	69

Introdução

Problemas reais em diversas áreas podem ser formulados matematicamente, dentre os quais: corte de barras e chapas, diagnóstico de doenças, minimização de custo e maximização de lucro. Esses dois últimos dizem respeito à otimização, ou seja, encontrar pontos de mínimo e máximo de uma função real. Neste trabalho abordaremos alguns métodos clássicos de otimização irrestrita, tais como: Cauchy, Newton e Seção Áurea. Em otimização, a solução de um problema de minimizar uma função é obtida por meio de um processo iterativo. Consideramos um ponto inicial x_0 , obtemos um ponto melhor x_1 , isto é, que diminui o valor da função objetivo e repetimos o processo gerando uma sequência na qual a função objetivo decresce.

Pode-se dizer que existem três aspectos atinentes aos métodos de otimização. O primeiro consiste na criação do algoritmo propriamente dito, o qual deve levar em conta a estrutura do problema e as propriedades satisfeitas pelas soluções, entre outras coisas. O segundo aspecto se refere às sequências geradas pelo algoritmo, onde a principal questão é saber se tais sequências realmente convergem para uma solução do problema. O terceiro ponto a ser considerado é a velocidade com que a sequência converge para uma solução; não basta que essa sequência seja convergente, é necessário que uma aproximação do limite possa ser obtida em um tempo razoável. Dessa forma, bons algoritmos são os que geram sequências que convergem rapidamente para uma solução.

Uma forma geral de construir um algoritmo consiste em escolher, a partir de cada ponto obtido, uma direção para dar o próximo passo. A direção escolhida no Algoritmo de Cauchy é a oposta ao gradiente, pois essa é a de maior decrescimento da função. Já o Algoritmo de Newton, minimiza, em cada iteração, a aproximação

quadrática da função objetivo. Quando a direção de busca já é dada e precisamos minimizar uma função a partir de um certo ponto, segundo essa direção, recaímos em um problema de minimizar uma função real de apenas uma variável. Um dos métodos que podem ser usados para resolver esse problema é o Método da Seção Áurea, que faz a minimização exata dessa função. Os três métodos mencionados serão objeto de estudo deste trabalho.

O texto está organizado da seguinte forma: inicialmente, revisitaremos resultados e conceitos importantes para o desenvolvimento deste trabalho. Então, antes de entrarmos no assunto "otimização", teremos um capítulo com os seguintes apontamentos: convergência de sequências, ordem de convergência, conceitos e resultados de Álgebra Linear e Análise no $\mathbb{R}^n$. Em seguida, um problema de otimização irrestrita é formalizado, destacando a condição necessária de 1ª ordem, na qual um ponto pode ser candidato a minimizador local. Esse capítulo destaca a necessidade de se aplicar um método iterativo para escolher a direção de descida da função, assim como, o tamanho do passo. Os demais capítulos apresentam os Métodos de Cauchy, de Newton e da Seção Áurea, seus algoritmos, exemplos e suas convergências.

No final do trabalho, concluímos as respostas que cada método forneceu, seus comportamentos em relação às velocidades de convergência, comparando o Algoritmo de Newton com o de Cauchy e destacando a particularidade do valor de convergência do método de busca exata, o da Seção Áurea.

1 Preliminares

Neste capítulo apresentaremos uma breve revisão sobre conceitos e resultados básicos de Análise no $\mathbb{R}^n$ e Álgebra Linear. Exporemos resultados, muitos sem demonstração, que serão relevantes ao longo do trabalho. As principais referências utilizadas para a confecção deste capítulo foram [3], [4] e [7].

Consideraremos o espaço euclidiano $\mathbb{R}^n$ munido de uma norma $\|\cdot\|$.

1.1 Convergência de sequências

Definição 1.1. Uma sequência em $\mathbb{R}^n$ é uma aplicação $x : \mathbb{N} \rightarrow \mathbb{R}^n$ que associa a cada $k \in \mathbb{N}$ um vetor $x_k \in \mathbb{R}^n$. Denotaremos uma sequência por $(x_k)_{k \in \mathbb{N}}$, ou simplesmente por (x_k) . Por conveniência, consideraremos $\mathbb{N} = \{0, 1, 2, 3, \dots\}$.

Definição 1.2. Um número real “ a ” é dito limite de uma sequência (x_k) , quando, para todo $\varepsilon > 0$ dado, é possível obter $\bar{k} \in \mathbb{N}$ tal que

$$k \geq \bar{k} \Rightarrow \|x_k - a\| < \varepsilon.$$

Neste caso, dizemos que (x_k) é convergente e converge para “ a ”, e escrevemos

$$\lim_{k \in \mathbb{N}} x_k = a \text{ ou } \lim_{k \rightarrow \infty} x_k = a \text{ ou } x_k \rightarrow a.$$

Pela Definição 1.2, o ponto $a \in \mathbb{R}^n$ é o limite da sequência (x_k) se, para cada $\varepsilon > 0$, o conjunto $N_1 = \{k \in \mathbb{N} : \|x_k - a\| \geq \varepsilon\}$ é finito, ou seja, fora da bola $B(a, \varepsilon) = \{x \in \mathbb{R}^n : \|x - a\| < \varepsilon\}$ só poderão estar, no máximo, os termos $x_0, x_1, \dots, x_{\bar{k}-1}$.

Definição 1.3. Dada uma sequência (x_k) , chamaremos de subsequência de (x_k) a toda restrição da função que define (x_k) a um conjunto infinito $\mathbb{N}' = \{k_0, k_1, \dots, k_i, \dots\} \subset \mathbb{N}$. Escreveremos $(x_k)_{k \in \mathbb{N}'}$ ou $(x_{k_i})_{i \in \mathbb{N}}$ ou simplesmente (x_{k_i}) , em que $(k_i)_{i \in \mathbb{N}}$ é uma sequência crescente de números inteiros positivos, para denotar uma subsequência de (x_k) . Note que $k_i \geq i$ para todo $i \in \mathbb{N}$.

Teorema 1.4. Se uma sequência (x_k) converge para um limite a , então toda subsequência (x_{k_i}) também converge para a .

Demonstração. Dado $\varepsilon > 0$ existe $\bar{k}$ tal que, para todo $k \geq \bar{k}$, tem-se $\|x_k - a\| < \varepsilon$. Sendo assim, dado $i \geq \bar{k}$, temos $k_i \geq \bar{k} \geq \bar{k}$. Por conseguinte, $\|x_{k_i} - a\| < \varepsilon$. $\square$

O limite de uma subsequência (x_{k_i}) é chamado *valor de aderência* ou *ponto de acumulação* da sequência (x_k) .

Pelo Teorema 1.4, se uma sequência possui dois pontos de acumulação, então ela não é convergente. Observe o exemplo a seguir.

Exemplo 1.5. Considere a sequência $x_k = (-1)^k + \frac{1}{k+1}$.

Para $k = 2i$, com $i \in \mathbb{N}$, temos $x_k \rightarrow 1$ e, para $k = 2i + 1$, com $i \in \mathbb{N}$, temos $x_k \rightarrow -1$, conforme se pode observar na Figura 1.1.

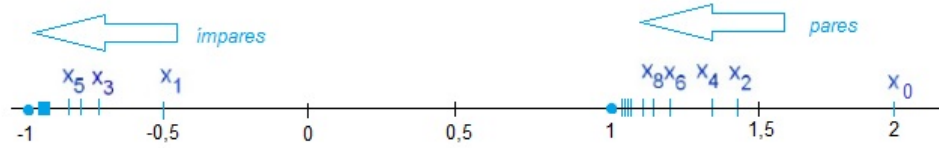


Figura 1.1: Termos da sequência (x_k) na reta.

Temos, pois, que essa sequência tem dois pontos de acumulação e, portanto, não é convergente.

Exemplo 1.6. Seja (x_k) uma sequência em $\mathbb{R}$. Se $x_k \rightarrow a > 0$ então existe $\bar{k} \in \mathbb{N}$ tal que, para $k \geq \bar{k}$, tem-se $x_k \geq \frac{a}{2}$.

Com efeito, para $\varepsilon = \frac{a}{2}$, existe $\bar{k}$ tal que, para $k \geq \bar{k}$, tem-se $|x_k - a| < \frac{a}{2}$. Então, para $k \geq \bar{k}$, tem-se

$$-\frac{a}{2} < x_k - a < \frac{a}{2},$$

ou seja,

$$\frac{a}{2} < x_k < \frac{3a}{2},$$

de onde segue que $x_k > \frac{a}{2}$, conforme a Figura 1.2.

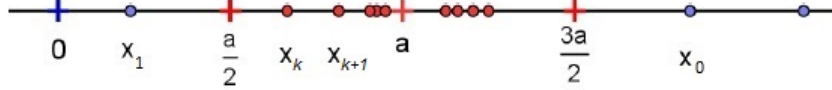


Figura 1.2: Termos da seqüência (x_k) na reta.

Definição 1.7. Uma seqüência $(x_k) \subset \mathbb{R}^n$ é limitada quando o conjunto formado pelos seus elementos é limitado, ou seja, quando existe um número real $M > 0$ tal que $\|x_k\| \leq M$ para todo $k \in \mathbb{N}$.

Definição 1.8. Seja $(x_k) \subset \mathbb{R}^n$ uma seqüência limitada. Definimos o limite inferior da seqüência (x_k) como seu menor ponto de acumulação e o denotamos por $\liminf(x_k)$. Analogamente, definimos o limite superior da seqüência como seu maior ponto de acumulação e o denotamos por $\limsup(x_k)$.

A seguir enunciaremos alguns resultados importantes da análise de convergência de seqüências. Para verificar as provas desses resultados, sugerimos ao leitor que consulte a referência [4].

Teorema 1.9. Toda seqüência $(x_k) \subset \mathbb{R}^n$ convergente é limitada.

Demonstração. Seja (x_k) uma seqüência convergente. Digamos que $\lim_{k \in \mathbb{N}} x_k = a$. Então, dado $\varepsilon = 1$, existe $k_0 \in \mathbb{N}$ tal que $\|x_k - a\| < 1$ para todo $k > k_0$.

Se $r = \max\{1, \|x_1 - a\|, \|x_2 - a\|, \dots, \|x_{k_0} - a\|\}$, então $\|x_k - a\| \leq r$ para todo $k \in \mathbb{N}$, ou seja, $\|x_k\| \leq r + \|a\|$ para todo $k \in \mathbb{N}$, de onde segue que (x_k) é limitada. $\square$

Teorema 1.10. Toda seqüência $(x_k) \subset \mathbb{R}$ monótona e limitada é convergente.

Demonstração. Seja (x_k) uma seqüência monótona e limitada. Suponhamos, sem perda de generalidade, que (x_k) seja não decrescente. Então, $x_1 \leq x_2 \leq x_3 \leq \dots \leq$

$x_k \leq \dots$. Como (x_k) é limitada, o conjunto $X = \{x_1, x_2, x_3, \dots, x_k, \dots\}$ possui supremo. Digamos que $\sup X = a$. Dado qualquer $\varepsilon > 0$, temos $a - \varepsilon < a$. Pela definição de supremo, existe algum $k_0 \in \mathbb{N}$ tal que $a - \varepsilon < x_{k_0}$. Como $x_k \leq a$ para todo $k \in \mathbb{N}$, vemos que

$$k \geq k_0 \implies a - \varepsilon < x_{k_0} \leq x_k \leq a < a + \varepsilon,$$

uma vez que (x_k) é não decrescente. Daí, concluímos que $x_k \rightarrow a$. $\square$

As demonstrações dos próximos resultados, Teorema 1.11 e Teorema 1.12, podem ser encontradas na página 17 da referência [4].

Teorema 1.11. *Uma sequência limitada em $\mathbb{R}^n$ é convergente se, e somente se, possui um único ponto de acumulação.*

Teorema 1.12 (Bolzano-Weierstrass). *Toda sequência limitada em $\mathbb{R}^n$ possui uma subsequência convergente.*

Teorema 1.13. *Uma sequência monótona $(x_k) \subset \mathbb{R}$ é limitada se, e somente se, possui uma subsequência limitada.*

Demonstração. É claro que se (x_k) é limitada, então qualquer subsequência de (x_k) também é limitada. Reciprocamente, suponhamos que (x_k) seja uma sequência monótona não decrescente e $(x_k)_{k \in \mathbb{N}'}$ seja uma subsequência limitada de (x_k) , ou seja, existe uma constante $M > 0$ tal que $x'_k \leq M$ para todo $k' \in \mathbb{N}'$. Como $\mathbb{N}'$ é ilimitado, dado $k \in \mathbb{N}$, existe $k' \in \mathbb{N}'$ tal que $k' > k$. Como (x_k) é não decrescente, obtemos

$$x_1 \leq x_k \leq x_{k'} \leq M.$$

Dessa forma, podemos concluir que $x_1 \leq x_k \leq M$ para todo $k \in \mathbb{N}$, ou seja, (x_k) é limitada. $\square$

Teorema 1.14. *Se $(x_k) \subset \mathbb{R}$ é uma sequência monótona que possui uma subsequência $(x_k)_{k \in \mathbb{N}'}$ convergente, digamos $x_k \xrightarrow{\mathbb{N}'} a$, então $x_k \rightarrow a$.*

Demonstração. Seja (x_k) uma sequência monótona que possui uma subsequência $(x_k)_{k \in \mathbb{N}'}$ convergente. Pelo Teorema 1.13, (x_k) é limitada porque possui uma subsequência convergente e, portanto, limitada (veja o Teorema 1.9). E, pelo Teorema 1.10, segue que (x_k) é convergente. $\square$

Teorema 1.15 (Teorema do Confronto). *Sejam (x_k) , (y_k) e (z_k) seqüências em $\mathbb{R}$. Se $x_k \leq y_k \leq z_k$ para todo $k \in \mathbb{N}$, $x_k \rightarrow a$ e $z_k \rightarrow a$, então $y_k \rightarrow a$.*

Demonstração. Seja $\varepsilon > 0$ arbitrário. Como $x_k \rightarrow a$, existe $k_1 = k_1(\varepsilon) \in \mathbb{N}$ tal que

$$k \geq k_1 \implies a - \varepsilon < x_k < a + \varepsilon. \quad (1.1)$$

De forma análoga, como $z_k \rightarrow a$, existe $k_2 = k_2(\varepsilon) \in \mathbb{N}$ tal que

$$k \geq k_2 \implies a - \varepsilon < z_k < a + \varepsilon. \quad (1.2)$$

Seja $k_0 = \max\{k_1, k_2\}$. Pela hipótese, por (1.1) e (1.2), obtemos

$$k \geq k_0 \implies a - \varepsilon < x_k \leq y_k \leq z_k < a + \varepsilon \implies a - \varepsilon < y_k < a + \varepsilon.$$

Logo, $y_k \rightarrow a$. □

O próximo exemplo pode ser encontrado, com mais detalhes, na referência [6].

Exemplo 1.16. *Considere a seqüência $r_k = \frac{a_{k+1}}{a_k}, k \geq 0$, em que os a'_k s são os termos da seqüência $\{1, 1, 2, 3, 5, 8, 13, 21, \dots\}$. A seqüência (r_k) é dada por:*

$$\frac{1}{1}, \frac{2}{1}, \frac{3}{2}, \frac{5}{3}, \frac{8}{5}, \dots, \text{ ou seja, } 1, 2, 1.5, 1.666\dots, 1.6, \dots$$

Por meio do gráfico abaixo (Figura 1.3) em que o eixo horizontal indica os termos a_k e o eixo vertical indica os respectivos valores de r_k , pode-se perceber que a seqüência tende a um valor entre 1.5 e 2.

A seqüência $(r_k), k \geq 0$, é dada recursivamente por

$$r_0 = 1 \text{ e } r_k = \frac{1}{r_{k-1}} + 1, k \geq 1.$$

Consideremos, agora, as subsequências (r_{2k}) e (r_{2k-1}) de (r_k) , ou seja, as subsequências de índices pares e ímpares, respectivamente. É possível provar por indução que (r_{2k}) é crescente e (r_{2k-1}) é decrescente. É importante observar que (r_{2k}) é limitada superiormente por 2 e (r_{2k-1}) é limitada inferiormente por 1. Ambas as seqüências são convergentes. Além disso, como satisfazem a mesma relação de

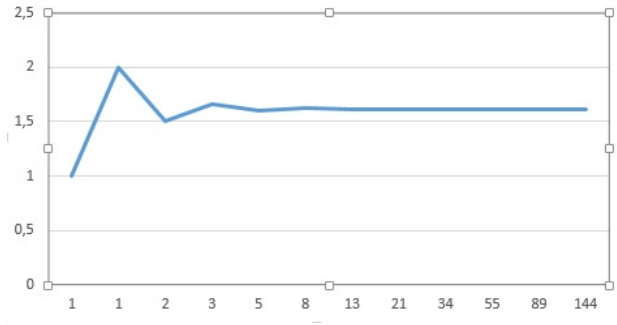


Figura 1.3: Valores de $r_k \times a_k$.

recorrência

$$r_k = 2 - \frac{1}{r_{k-2} + 1}, \quad k \geq 2,$$

pode-se concluir que seus limites são iguais e, conseqüentemente, a sequência (r_k) converge para esse mesmo limite, que é

$$r = \frac{1 + \sqrt{5}}{2}.$$

O número r é o conhecido como “Número de Ouro”.

O Número de Ouro (também conhecido como *Razão Áurea*) é representado pela letra grega phi (ϕ) em homenagem à Fídias, famoso escultor grego que utilizou a proporção de ouro na construção do Partenón, em Atenas. Esse número irracional ($\phi = 1,618\dots$) é considerado por muitos o símbolo da harmonia. Ele pode ser identificado em muitas obras de arte, construções, objetos e na natureza.

1.2 Ordem de convergência

No contexto de otimização, existe outro aspecto importante a ser analisado em sequência: a velocidade de convergência. Considere, por exemplo, as sequências a seguir:

$$x_k = \frac{1}{k+5}, \quad y_k = \frac{1}{3^k}, \quad w_k = \frac{1}{2^{k^2}} \quad \text{e} \quad z_k = \frac{1}{2^{2^k}}.$$

Todas elas convergem para 0, mas não da mesma maneira, conforme se pode verificar na tabela abaixo.

k	x_k	y_k	w_k	z_k
0	0,2000	1,0000	1,0000	0,5000
1	0,1667	0,3333	0,5000	0,2500
2	0,1429	0,1111	0,0625	0,0625
3	0,1250	0,0370	$1,95 \cdot 10^{-3}$	$3,91 \cdot 10^{-3}$
4	0,1111	0,0123	$1,53 \cdot 10^{-5}$	$1,53 \cdot 10^{-5}$
5	0,1000	0,0041	$2,98 \cdot 10^{-8}$	$2,33 \cdot 10^{-10}$
6	0,0909	0,0014	$1,46 \cdot 10^{-11}$	$5,42 \cdot 10^{-20}$

Diante disso, é conveniente estabelecer uma maneira de medir a velocidade de sequências convergentes. Considere, então, uma sequência $(x_k) \subset \mathbb{R}^n$ que converge para $\bar{x} \in \mathbb{R}^n$. Assim, $e_k = \|x_k - \bar{x}\| \rightarrow 0$. O que faremos é avaliar como o erro (e_k) tende à 0.

Definição 1.17. Dizemos que a sequência $(x_k) \subset \mathbb{R}^n$ converge linearmente para $\bar{x} \in \mathbb{R}^n$ quando existe uma constante $r \in [0, 1)$ e um número $k_0 \in \mathbb{N}$ tais que

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \leq r \quad (1.3)$$

sempre que $k \geq k_0$.

Vale observar que a condição (1.3) implica que $x_k \rightarrow \bar{x}$, pois

$$\|x_{k_0+p} - \bar{x}\| \leq r^p \|x_{k_0} - \bar{x}\|,$$

para quaisquer $p \in \mathbb{N}$ e $r \in [0, 1)$.

Exemplo 1.18. A sequência $x_k = \frac{1}{k+5}$ converge para 0, mas não linearmente. De fato:

$$\frac{\|x_{k+1}\|}{\|x_k\|} = \frac{k+5}{k+6} \rightarrow 1.$$

Exemplo 1.19. A sequência $y_k = \frac{1}{3^k}$ converge linearmente para 0, pois

$$\frac{\|y_{k+1}\|}{\|y_k\|} = \frac{1}{3}.$$

Exemplo 1.20. As sequências $w_k = \frac{1}{2^{k^2}}$ e $z_k = \frac{1}{2^{2^k}}$ também convergem linearmente para 0.

Vejamos agora uma forma mais rápida de convergência.

Definição 1.21. A sequência $(x_k) \subset \mathbb{R}^n$ converge superlinearmente para $\bar{x} \in \mathbb{R}^n$ quando

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} \rightarrow 0. \quad (1.4)$$

Veja que a condição (1.4) também implica que $x_k \rightarrow \bar{x}$.

Note que $y_k = \frac{1}{3^k}$ não converge superlinearmente. Porém, a sequência $w_k = \frac{1}{2^{k^2}}$ converge superlinearmente para 0. Com efeito,

$$\frac{\|w_{k+1}\|}{\|w_k\|} = \frac{2^{k^2}}{2^{(k+1)^2}} = \frac{1}{2^{2k+1}} \rightarrow 0.$$

A sequência $z_k = \frac{1}{2^{2^k}}$ também converge superlinearmente para 0.

Outra forma de convergência, ainda mais rápida, será exposta abaixo.

Definição 1.22. Suponha que $x_k \rightarrow \bar{x}$. A convergência é dita quadrática quando existe uma constante $M > 0$, tal que

$$\frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|^2} \leq M. \quad (1.5)$$

É importante observar que apenas a condição (1.5) não implica que $x_k \rightarrow \bar{x}$.

A sequência $z_k = \frac{1}{2^{2^k}}$ converge quadraticamente para 0, pois,

$$\frac{\|z_{k+1}\|}{\|z_k\|^2} = \frac{\frac{1}{2^{2^{k+1}}}}{\left(\frac{1}{2^{2^k}}\right)^2} = \frac{2^{2^k \cdot 2^1}}{2^{2 \cdot 2^k}} = 2^0 = 1.$$

As demais sequências já exibidas não convergem quadraticamente, por exemplo, $w_k = \frac{1}{2^{k^2}}$:

$$\frac{\|w_{k+1}\|}{\|w_k\|^2} = \frac{(2^{k^2})^2}{2^{(k+1)^2}} = \frac{2^{2k^2}}{2^{k^2+2k+1}} = \frac{2^{k^2}}{2^{2k+1}} \rightarrow \infty.$$

Portanto, não existe $M > 0$ tal que $\frac{\|w_{k+1}\|}{\|w_k\|^2} \leq M$.

1.3 Revisão de conceitos e resultados de Álgebra Linear

Começaremos esta revisão com a Desigualdade de Cauchy-Schwarz, cuja demonstração pode ser encontrada em [3], página 169.

Teorema 1.23 (Desigualdade de Cauchy-Schwarz). *Se E é um espaço vetorial com produto interno $\langle \cdot, \cdot \rangle$, então*

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|, \quad \text{para quaisquer } x, y \in E,$$

em que $\|x\| = \sqrt{\langle x, x \rangle}$ e $\|y\| = \sqrt{\langle y, y \rangle}$. A igualdade é válida se, e somente se, x e y são linearmente dependentes.

A próxima definição será importante para os próximos capítulos.

Definição 1.24. *Seja $A \in \mathbb{R}^{n \times n}$ uma matriz simétrica. Dizemos que A é positiva definida (respectivamente, negativa definida) quando $x^T A x > 0$ (respectivamente, $x^T A x < 0$), para todo $x \in \mathbb{R}^n \setminus \{0\}$, em que x^T é o vetor transposto de x . Tal propriedade é denotada por $A > 0$ (respectivamente, por $A < 0$). Se $x^T A x \geq 0$ (respectivamente, $x^T A x \leq 0$), para todo $x \in \mathbb{R}^n$, A será dita positiva semidefinida (respectivamente, negativa semidefinida), fato este denotado por $A \geq 0$ (respectivamente, por $A \leq 0$).*

Cabe salientar que a definição geral de positividade de uma matriz não exige que ela seja simétrica. No entanto, no contexto deste trabalho vamos supor a simetria quando considerarmos matrizes positivas.

Para verificar as demonstrações dos dois resultados seguintes, sugerimos ao leitor que consulte a referência [3].

Lema 1.25 (Decomposição de Cholesky). *Se $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica positiva definida, então A pode ser fatorada em um produto GG^T , em que $G \in \mathbb{R}^{n \times n}$ é uma matriz triangular inferior com elementos diagonais positivos.*

Lema 1.26. *Se $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica com λ_1 e λ_n sendo seu menor e seu maior autovalor, respectivamente, então*

$$\lambda_1 \|x\|^2 \leq x^T A x \leq \lambda_n \|x\|^2,$$

para todo $x \in \mathbb{R}^n$.

1.4 Mais alguns conceitos e resultados importantes de Análise no $\mathbb{R}^n$

Definição 1.27. *Dada uma função vetorial de várias variáveis reais diferenciável $F : \mathbb{R}^m \rightarrow \mathbb{R}^n$, com $F(X) = (f_1(X), \dots, f_n(X))$ para $X \in \mathbb{R}^m$, a representação matricial de sua derivada é denominada matriz Jacobiana de F e é definida como sendo:*

$$J_F = F' = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_m} \end{bmatrix}.$$

Definição 1.28. *Dada uma função real de várias variáveis, $f : \mathbb{R}^n \rightarrow \mathbb{R}$, duas vezes diferenciável, a matriz Jacobiana (derivada) do gradiente de f (que é uma função vetorial) é denominada matriz Hessiana de f . Assim,*

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

∇f , denominado vetor gradiente de f , é o vetor cujas componentes são as derivadas parciais de f , ou seja, $\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right)$. Em notação matricial,

temos

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}.$$

Definição 1.29. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função real de n variáveis. Considere $a \in \mathbb{R}^n$ e $u \in \mathbb{R}^n$ um vetor com $\|u\| = 1$. A derivada direcional de f no ponto a , na direção u , é*

$$D_u f(a) = \lim_{h \rightarrow 0} \frac{f(a + hu) - f(a)}{h},$$

se esse limite existir.

A Desigualdade do Valor Médio, que será bastante útil em alguns resultados do Capítulo 4, pode ser encontrada na referência [4], página 263.

Teorema 1.30 (Desigualdade do Valor Médio). *Seja $U \subset \mathbb{R}^n$ um conjunto aberto e convexo. Se $f : U \rightarrow \mathbb{R}^n$ é diferenciável e $\|J_f(x)\| \leq \beta$ para todo $x \in U$, então f é lipschitziana, com $\|f(x) - f(y)\| \leq \beta \cdot \|x - y\|$ para quaisquer $x, y \in U$.*

Definição 1.31. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável. Dizemos que $a \in \mathbb{R}^n$ é um ponto crítico ou estacionário de f quando $\nabla f(a) = 0$.*

Definição 1.32. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável. Dizemos que f tem um máximo (respectivamente, um mínimo) local em $a \in \mathbb{R}^n$ quando existe $\delta > 0$ tal que*

$$\|v\| < \delta \implies f(a + v) \leq f(a) \text{ (respectivamente, } f(a) \leq f(a + v)).$$

A demonstração do próximo resultado pode ser verificada na referência [4], página 158.

Teorema 1.33. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função duas vezes diferenciável no ponto crítico $a \in \mathbb{R}^n$ e $\nabla^2 f(a)$ a forma quadrática Hessiana de f no ponto a .*

i) *Se $\nabla^2 f(a)$ é positiva, então a é um ponto de mínimo local;*

- ii) Se $\nabla^2 f(a)$ é negativa, então a é um ponto de máximo local;
- iii) Se $\nabla^2 f(a)$ não é positiva, nem negativa, então a não é ponto de mínimo local, nem ponto de máximo local.

Definição 1.34. Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa em C quando

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y),$$

para todos $x, y \in C$ e $t \in [0, 1]$.

Geometricamente, podemos dizer que qualquer arco no gráfico de uma função convexa está sempre abaixo do segmento que liga as extremidades. Veja na Figura 1.4 uma função convexa e outra não convexa.

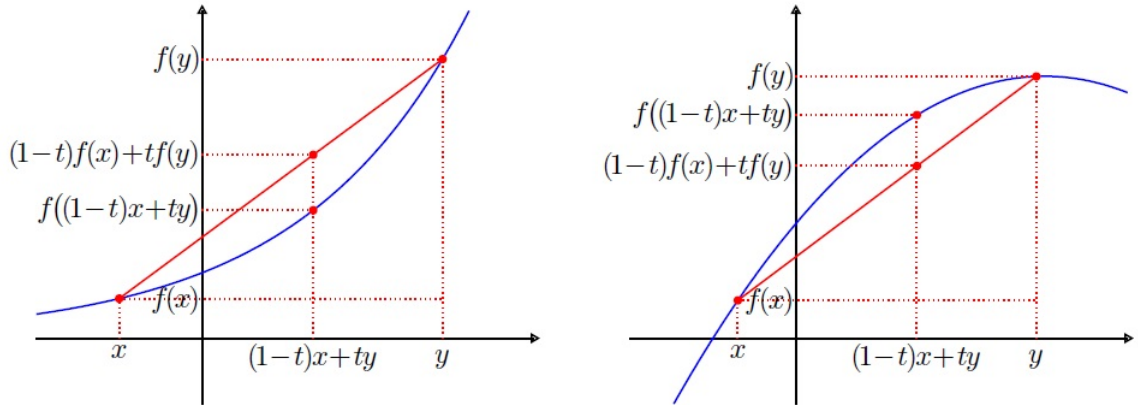


Figura 1.4: Funções convexa e não convexa.

No próximo capítulo, formalizaremos um problema de otimização irrestrita, tendo como base as referências [2] e [7].

2 Introdução à Otimização

Formalmente, um problema de otimização irrestrita pode ser escrito como

$$\text{minimizar } f(x) \tag{2.1}$$

$$\text{sujeito à } x \in \mathbb{R}^n$$

em que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$, denominada *função objetivo*, é de classe C^1 . Ao considerar tal problema, o objetivo é determinar minimizadores locais de f .

Definição 2.1. Dizemos que $\bar{x} \in \mathbb{R}^n$ é um *minimizador local* de f se $f(\bar{x}) \leq f(x)$ para todo x pertencente à uma vizinhança de $\bar{x}$, ou seja, se existe $\delta > 0$ tal que $f(\bar{x}) \leq f(x)$ para todo $x \in B(\bar{x}, \delta) = \{y \in \mathbb{R}^n; \|y - \bar{x}\| < \delta\}$.

Definição 2.2. Dizemos que $\bar{x} \in \mathbb{R}^n$ é um *minimizador global* de f se $f(\bar{x}) \leq f(x)$ para todo $x \in \mathbb{R}^n$.

No caso em que a função objetivo está definida sobre um subconjunto Ω de $\mathbb{R}^n$, determinado por restrições de igualdade e/ou desigualdade, temos então um caso de problema de otimização restrita. Esse tipo de problema não será abordado neste trabalho, entretanto, para estudá-lo, o leitor pode consultar, por exemplo, a referência [7], que aborda o problema geral de Programação Não Linear (PNL).

Em muitas aplicações, o interesse é obter o valor de máximo de uma função, como por exemplo, no caso em que se considera uma função que representa o lucro de uma empresa. No entanto, do ponto de vista matemático, não existe diferença relevante entre minimização e maximização de uma função f , uma vez que maximizar f é equivalente a minimizar $-f$. Dessa forma, a teoria desenvolvida para

problemas de otimização leva em consideração problemas da forma apresentada em (2.1). No entanto, a fim de completar o presente texto, apresentaremos também as definições de maximizador local e global de f .

Definição 2.3. Dizemos que $\bar{x} \in \mathbb{R}^n$ é um maximizador local de f se $f(\bar{x}) \geq f(x)$ para todo x pertencente à uma vizinhança de $\bar{x}$, ou seja, se existe $\delta > 0$ tal que $f(\bar{x}) \geq f(x)$ para todo $x \in B(\bar{x}, \delta)$.

Definição 2.4. Dizemos que $\bar{x} \in \mathbb{R}^n$ é um maximizador global de f se $f(\bar{x}) \geq f(x)$ para todo $x \in \mathbb{R}^n$.

Para que um ponto $\bar{x} \in \mathbb{R}^n$ seja um minimizador local é necessário que ele satisfaça algumas condições, as quais são denominamos condições de otimalidade. Vejamos então a condição necessária de otimalidade de 1ª ordem, estabelecida no teorema a seguir.

Teorema 2.5. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local de f , então $\nabla f(\bar{x}) = 0$.

Demonstração. Considere $d \in \mathbb{R}^n \setminus \{0\}$ arbitrário. Como $\bar{x}$ é um minimizador local, existe $\delta > 0$ tal que

$$f(\bar{x}) \leq f(\bar{x} + td), \quad (2.2)$$

para todo $t \in (0, \delta)$. Pela expansão de Taylor, temos

$$f(\bar{x} + td) = f(\bar{x}) + t\nabla f(\bar{x})^T \cdot d + r(t), \quad (2.3)$$

com $\lim_{t \rightarrow 0} \frac{r(t)}{t} = 0$. Usando (2.2) e (2.3), vemos que

$$0 \leq t\nabla f(\bar{x})^T \cdot d + r(t). \quad (2.4)$$

Dividindo a desigualdade (2.4) por t ($t > 0$), obtemos $0 \leq \nabla f(\bar{x})^T \cdot d + \frac{r(t)}{t}$. Passando o limite quando $t \rightarrow 0$, concluimos que

$$\nabla f(\bar{x})^T \cdot d \geq 0. \quad (2.5)$$

Se $\nabla f(\bar{x})$ não fosse nulo, poderíamos escolher $d = -\nabla f(\bar{x})$, resultando em

$$\|\nabla f(\bar{x})\|^2 = -\nabla f(\bar{x})^T \cdot d \leq 0,$$

o que está em contradição com (2.5). Logo, $\nabla f(\bar{x}) = 0$. $\square$

Assim, vemos que a condição estabelecida no Teorema 2.5 garante que se um ponto $\bar{x}$ for o minimizador local de f , então $\bar{x}$ será solução do sistema homogêneo, na maioria das vezes não-linear, $\nabla f(\bar{x}) = 0$. Os pontos que satisfazem tal condição, como já vimos no Capítulo 1, são denominados pontos críticos ou estacionários da função f . É importante observar que esta é apenas uma condição necessária, mas não suficiente. Portanto, se $\nabla f(\bar{x}) = 0$, então $\bar{x}$ pode ser maximizador, minimizador, ou nenhum dos dois. Os pontos que satisfazem essa condição, mas não são maximizadores nem minimizadores de f , são denominados *pontos de sela*.

Embora a condição $\nabla f(x) = 0$ seja necessária, mas não suficiente, a resolução desse sistema faz parte de muitos métodos de otimização. Porém, dificilmente é possível resolvê-lo por métodos diretos, sendo necessário o emprego de métodos iterativos que consistem em, dado um ponto inicial x_0 , gerar uma sequência de iterandos $(x_k) \subset \mathbb{R}^n$ na qual a função objetivo decresce.

Um algoritmo básico de otimização irrestrita consiste em, a partir de cada ponto obtido, determinar uma direção para dar o próximo passo. Como o objetivo é minimizar f , é razoável que a função decresça na direção escolhida. Uma direção que apresenta essa propriedade é denominada *direção de descida*.

Definição 2.6. Considere uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$, um ponto $\bar{x} \in \mathbb{R}^n$ e uma direção $d \in \mathbb{R}^n \setminus \{0\}$. Dizemos que d é uma direção de descida para f , a partir de $\bar{x}$, quando existe $\delta > 0$ tal que $f(\bar{x} + td) < f(\bar{x})$, para todo $t \in (0, \delta)$.

Apresentamos abaixo uma condição suficiente para uma direção ser de descida.

Teorema 2.7. Se $\nabla f(\bar{x})^T \cdot d < 0$, então d é uma direção de descida para f , a partir de $\bar{x}$.

Demonstração. Sabemos que

$$\nabla f(\bar{x})^T d = \frac{\partial f}{\partial d}(\bar{x}) = \lim_{t \rightarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t}.$$

Pela hipótese e pela preservação do sinal (f é continuamente diferenciável), existe $\delta > 0$ tal que

$$\frac{f(\bar{x} + td) - f(\bar{x})}{t} < 0,$$

para todo $t \in (-\delta, \delta)$, $t \neq 0$. Portanto, $f(\bar{x} + td) < f(\bar{x})$, para todo $t \in (0, \delta)$, e a prova está completa. $\square$

Exemplo 2.8. *Sejam $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $f(x) = \frac{1}{2}(x_1^2 - x_2^2)$ e $\bar{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Afirmamos que $d = -\nabla f(\bar{x})$ é uma direção de descida para f . Com efeito, temos que*

$$\nabla f(\bar{x}) = \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad e \quad d = -\nabla f(\bar{x}) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}.$$

Portanto,

$$\nabla f(\bar{x})^T d = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \end{pmatrix} = -1 < 0$$

e, pelo Teorema 2.7, temos a veracidade da afirmação.

As direções que formam um ângulo maior do que 90 com o gradiente de f avaliado em $\bar{x}$, são direções de descida para f a partir desse ponto.

Após escolher a direção de descida, é necessário definir o quanto caminhar nessa direção. Pode-se dar um passo completo, ou seja, considerar o tamanho $t = 1$, no entanto, essa escolha pode não ser a melhor. Dessa forma, existem métodos denominados métodos de busca que podem ser empregados para esse fim.

O que diferencia os métodos de otimização é a escolha da direção de descida d e a forma com que é calculado o tamanho do passo a ser dado. Nos próximos capítulos veremos o método de busca unidirecional, do tipo exata, que é o Método da Seção Áurea, e os métodos de otimização irrestrita: de Cauchy (do Gradiente) e de Newton. Já na sequência estudaremos o Método de Cauchy.

3 Método de Cauchy

Abordaremos agora um dos métodos para resolver o problema de minimizar uma função em $\mathbb{R}^n$.

O Método de Cauchy, também conhecido como o Método do Gradiente, é um processo iterativo que a cada etapa faz uma busca na direção oposta ao vetor gradiente da função objetivo no ponto corrente. A justificativa dessa escolha se baseia no fato de que o gradiente da função objetivo avaliado em um ponto x aponta para a direção de maior crescimento de f a partir desse ponto. Sendo assim, a direção oposta ao vetor gradiente é a que fornece um maior decréscimo na função objetivo. De fato, se $d = -\nabla f(x)$ e $v \in \mathbb{R}^n$ é tal que $\|v\| = \|d\|$, então

$$\frac{\partial f}{\partial d}(x) = \nabla f(x)^T d = -\|\nabla f(x)\|^2 = -\|\nabla f(x)\|\|v\| \leq \nabla f(x)^T v = \frac{\partial f}{\partial v}(x).$$

As principais referências para este capítulo são [1], [5] e [7].

3.1 Algoritmo

Apresentamos a seguir o algoritmo associado ao Método de Cauchy.

Algoritmo

Dado: $x_0 \in \mathbb{R}^n$

$k = 0$

REPITA enquanto $\nabla f(x_k) \neq 0$

Defina $d_k = -\nabla f(x_k)$

Obtenha $t_k > 0$ tal que $f(x_k + t_k d_k) < f(x_k)$

Faça $x_{k+1} = x_k + t_k d_k$

$k = k + 1$

O Algoritmo acima deixa em aberto a determinação do comprimento do passo t , e nessa etapa pode ser usado qualquer algoritmo de busca exata, como o Método da Seção Áurea (a ser visto no Capítulo 5), ou de busca inexata, como a busca de Armijo, que não será abordada neste trabalho. Para maiores detalhes sobre métodos baseados em busca inexata, o leitor pode consultar, por exemplo, a referência [5].

Exemplo 3.1. Considere $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, definida por $f(x_1, x_2) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2$ para $(x_1, x_2) \in \mathbb{R}^2$.

Temos $\nabla f(x) = \begin{pmatrix} x_1 \\ x_2^3 - x_2 \end{pmatrix}$ e $\nabla^2 f(x) = \begin{pmatrix} 1 & 0 \\ 0 & 3x_2^2 - 1 \end{pmatrix}$.

Os pontos estacionários de f são $\bar{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ e $x' = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$.

Através de cálculos simples, pode-se constatar que $\nabla^2 f$ não é positiva definida, nem negativa definida em $\bar{x}$, o que significa que $\bar{x}$ é um ponto de sela. Além disso, $\nabla^2 f$ é positiva definida em x^* e x' , de onde segue que esses dois são minimizadores locais de f (veja o Teorema 1.33).

Vamos fazer uma iteração do Método do Gradiente a partir de $x^0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

Pois bem, no ponto x^0 , a direção de Cauchy é $d^0 = -\nabla f(x^0) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$. Dessa

forma, o novo ponto é $x^1 = x^0 + t_0 d^0 = \begin{pmatrix} 1 - t_0 \\ 0 \end{pmatrix}$.

Note que se $x = \begin{pmatrix} a \\ 0 \end{pmatrix}$, então $d = -\nabla f(x) = \begin{pmatrix} -a \\ 0 \end{pmatrix}$ e $f(x + td) = \frac{1}{2}(1 - t)^2 a^2$.

Portanto, a busca exata fornece $t = 1$ e $x^+ = x + d = \bar{x}$. Ou seja, uma iteração

do método de Cauchy encontra o ponto estacionário $\bar{x}$.

A Figura 3.1 mostra 4 iterações do algoritmo com busca exata aplicado para minimizar uma função quadrática, onde as curvas de nível desta função são elipses.

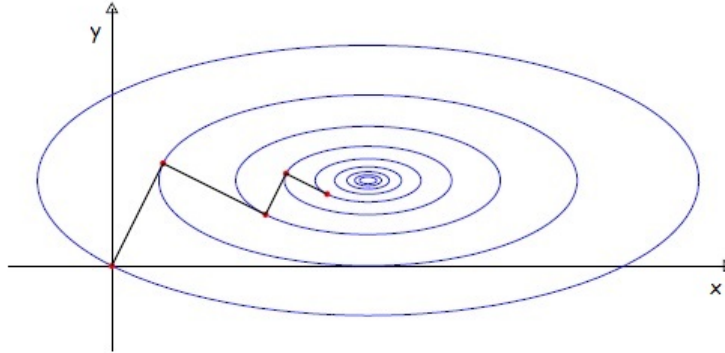


Figura 3.1: Passos do Algoritmo de Cauchy.

Ela sugere duas propriedades do algoritmo. Uma delas, é o fato de duas direções consecutivas serem ortogonais.

De fato, definindo $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ por $\varphi(t) = f(x_k + td_k)$, temos

$$\varphi'(t_k) = \nabla f(x_k + t_k d_k)^T d_k = \nabla f(x_{k+1})^T d_k.$$

Portanto, como a busca é feita por uma minimização local, concluímos que

$$(d_{k+1})^T d_k = -\nabla f(x_{k+1})^T d_k = -\varphi'(t_k) = 0.$$

A outra propriedade refere-se à convergência, que será discutida na próxima seção.

3.2 Convergência

Teorema 3.2. *O Algoritmo de Cauchy, com o tamanho do passo t_k calculado pela busca exata, é globalmente convergente, isto é, para qualquer sequência (x_k) gerada pelo Algoritmo de Cauchy e qualquer ponto de acumulação $\bar{x}$ de (x_k) , tem-se que $\bar{x}$ é estacionário.*

Demonstração. Sejam (x_k) uma sequência gerada pelo Algoritmo de Cauchy e $\bar{x}$ um ponto de acumulação de (x_k) , digamos que $x_k \xrightarrow{\mathbb{N}'} \bar{x}$.

Suponha que $\bar{x}$ não seja estacionário, isto é, que $\nabla f(\bar{x}) \neq 0$. Assim, $\bar{d} = -\nabla f(\bar{x})$ é uma direção de descida, o que garante a existência de $\bar{t} > 0$ tal que $f(\bar{x} + \bar{t}\bar{d}) < f(\bar{x})$. Considere $h : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $h(x) = f(x) - f(x - \bar{t}\nabla f(x))$. Como h é contínua, pois f é diferenciável, temos que $h(x_k) \xrightarrow{\mathbb{N}'} h(\bar{x})$. Chamando $h(\bar{x}) = \delta > 0$, temos que $h(x_k) \xrightarrow{\mathbb{N}'} \delta$. Assim, para todo $k \in \mathbb{N}'$ suficientemente grande, temos que

$$h(x_k) \geq \frac{\delta}{2}, \quad (3.1)$$

como vimos no Exemplo (1.6). Então, por (3.1) e pela definição de h , vemos que

$$f(x_k) - f(x - t_k \nabla f(x_k)) \geq \frac{\delta}{2}. \quad (3.2)$$

Mas, como t_k foi obtido pela busca exata, a relação (3.2) equivale a

$$f(x_k) - f(x + t_k d_k) \geq \frac{\delta}{2}.$$

Por conseguinte,

$$f(x_k) - f(x_{k+1}) \geq \frac{\delta}{2}, \quad (3.3)$$

para todo $k \in \mathbb{N}'$ suficientemente grande. Por outro lado, pela continuidade da f , temos $f(x_k) \xrightarrow{\mathbb{N}'} f(\bar{x})$. Como a sequência $(f(x_k))_{n \in \mathbb{N}}$ é monótona decrescente, pois $f(x_{k+1}) < f(x_k)$, o Teorema 1.14 garante que $f(x_k) \xrightarrow{\mathbb{N}} f(\bar{x})$, contradizendo (3.3). $\square$

3.3 Velocidade de convergência

Os resultados mais relevantes sobre a velocidade de convergência do algoritmo de Cauchy são obtidos quando a função objetivo é quadrática. Vamos, portanto, considerar

$$f(x) = \frac{1}{2}x^T A x + b^T x + c, \quad x \in \mathbb{R}^n,$$

em que $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

É possível provar que f é convexa e tem um único minimizador $\bar{x}$, que é global

e satisfaz

$$\nabla f(\bar{x}) = A\bar{x} + b.$$

Para verificar uma prova dessa última afirmação, veja [7], página 38.

Provaremos agora que, usando a norma euclidiana, a sequência gerada pelo método de Cauchy com busca exata converge linearmente para $\bar{x}$, com taxa de convergência $\sqrt{1 - \frac{\lambda_1}{\lambda_n}}$, em que λ_1 e λ_n são o menor e o maior autovalor de A , respectivamente.

Inicialmente, note que o passo ótimo é dado por $t_k = \frac{(d_k)^T d_k}{(d_k)^T A d_k}$. Com efeito, como

$$\begin{aligned} \frac{d}{dt} f(x_k + t d_k) &= [\nabla f(x_k + t d_k)] d_k \\ &= [A(x_k + t d_k) + b]^T d_k \\ &= [(A x_k + b) + t A d_k]^T d_k \\ &= [(\nabla f(x_k)) + t A d_k]^T d_k \\ &= \nabla f(x_k)^T d_k + t (d_k)^T A d_k \end{aligned}$$

e t_k é o passo ótimo, devemos ter $\nabla f(x_k)^T d_k + t_k (d_k)^T A d_k = 0$. Portanto,

$$t_k = \frac{-\nabla f(x_k)^T d_k}{(d_k)^T A d_k} = \frac{(d_k)^T d_k}{(d_k)^T A d_k}.$$

No que segue, para simplificar a notação, vamos admitir que $\bar{x} = 0$ e $f(\bar{x}) = 0$, isto é,

$$f(x) = \frac{1}{2} x^T A x.$$

Lema 3.3. *Dado $x \in \mathbb{R}^n$, $x \neq 0$, considere $d = -Ax$. Então,*

$$\frac{d^T d}{d^T A d} \leq \frac{x^T A x}{x^T A^2 x}.$$

Demonstração. Afirmamos que $x^T A x = d^T A^{-1} d$ e $x^T A^2 x = d^T d$. De fato,

$$d^T A^{-1} d = (-Ax)^T A^{-1} (-Ax) = -(x^T) A^T A^{-1} (-Ax) = x^T A x$$

e

$$d^T d = (-Ax)^T (-Ax) = -(x^T) A^T (-Ax) = x^T A^2 x.$$

Portanto,

$$\frac{d^T dx^T A^2 x}{d^T A dx^T A x} = \frac{(d^T d)^2}{(d^T A d)(d^T A^{-1} d)} \quad (3.4)$$

Como $A > 0$, pela decomposição de Cholesky (Lema 1.25), existe $G \in \mathbb{R}^{n \times n}$ tal que $A = GG^T$, em que G é uma matriz triangular inferior com elementos diagonais positivos. Fazendo $u = G^T d$ e $v = G^{-1} d$, temos que

$$\begin{aligned} \star \quad u^T v &= (G^T d)^T (G^{-1} d) = d^T G G^{-1} d = d^T d; \\ \star \quad u^T u &= (G^T d)^T (G^T d) = d^T G G^T d = d^T A d; \\ \star \quad v^T v &= (G^{-1} d)^T (G^{-1} d) = d^T (G^{-1})^T G^{-1} d = d^T A^{-1} d. \end{aligned}$$

Pela Desigualdade de Cauchy- Schwarz, obtemos

$$\frac{(d^T d)^2}{(d^T A d)(d^T A^{-1} d)} = \frac{(u^T v)^2}{(u^T u)(v^T v)} = \frac{|\langle u, v \rangle|^2}{\langle u, u \rangle \langle v, v \rangle} = \frac{|\langle u, v \rangle|^2}{\|u\|^2 \|v\|^2} \leq 1. \quad (3.5)$$

De (3.4) e (3.5) segue que

$$\frac{d^T dx^T A^2 x}{d^T A dx^T A x} \leq 1,$$

o que completa a prova. □

Teorema 3.4. *Considere a função quadrática*

$$f(x) = \frac{1}{2} x^T A x$$

e a sequência (x_k) gerada pelo Algoritmo de Cauchy, com busca exata. Se $\gamma = \sqrt{1 - \frac{\lambda_1}{\lambda_n}}$, em que λ_1 e λ_n são o menor e o maior autovalor de A , respectivamente, então $\|x_{k+1}\| \leq \gamma \|x_k\|$, para todo $k \in \mathbb{N}$.

Demonstração. Como $d_k = -\nabla f(x_k) = -A x_k$, temos

$$\begin{aligned} \|x_{k+1}\|^2 &= (x_k + t_k d_k)^T (x_k + t_k d_k) \\ &= (x_k)^T x_k + (x_k)^T t_k d_k + (d_k)^T t_k x_k + (d_k)^T t_k^2 d_k \\ &= \|x_k\|^2 + 2t_k (x_k)^T d_k + (t_k)^2 (d_k)^T d_k \end{aligned}$$

$$\begin{aligned}
&= \|x_k\|^2 + 2t_k(x_k)^T(-Ax_k) + (t_k)^2(-Ax_k)^T(-Ax_k) \\
&= \|x_k\|^2 - 2t_k(x_k)^T Ax_k + (t_k)^2(x_k)^T A^2 x_k.
\end{aligned}$$

Pelo Lema 3.3, temos que

$$\frac{(d_k)^T d_k}{(d_k)^T A d_k} (x_k)^T A^2 x_k \leq (x_k)^T A x_k.$$

Como $t_k = \frac{(d_k)^T d_k}{(d_k)^T A d_k} > 0$, temos

$$(t_k)^2 (x_k)^T A^2 x_k = t_k \frac{(d_k)^T d_k}{(d_k)^T A d_k} (x_k)^T A^2 x_k \leq t_k (x_k)^T A x_k.$$

Assim,

$$\|x_{k+1}\|^2 = \|x_k\|^2 - 2t_k(x_k)^T A x_k + (t_k)^2 (x_k)^T A^2 x_k \leq \|x_k\|^2 - t_k(x_k)^T A x_k. \quad (3.6)$$

Se $x^k = 0$, então não há o que fazer. Suponhamos, então, $x^k \neq 0$. Da relação (3.6), obtemos

$$\frac{\|x_{k+1}\|^2}{\|x_k\|^2} \leq \frac{\|x_k\|^2 - t_k(x_k)^T A x_k}{\|x_k\|^2} = 1 - \frac{t_k(x_k)^T A x_k}{\|x_k\|^2} = 1 - \frac{(d_k)^T d_k}{(d_k)^T A d_k} \cdot \frac{(x_k)^T A x_k}{(x_k)^T x_k}. \quad (3.7)$$

Pelo Lema 1.26, sabemos que

$$\frac{(d_k)^T d_k}{(d_k)^T A d_k} \geq \frac{1}{\lambda_n} \text{ e } \frac{(x_k)^T A x_k}{(x_k)^T x_k} \geq \lambda_1.$$

Usando isso em (3.7), temos

$$\left(\frac{\|x_{k+1}\|}{\|x_k\|} \right)^2 \leq 1 - \frac{\lambda_1}{\lambda_n}.$$

De acordo com a Definição 1.17, concluímos que a velocidade de convergência da sequência gerada pelo Algoritmo de Cauchy é linear, com constante $\sqrt{1 - \frac{\lambda_1}{\lambda_n}}$. $\square$

O teorema precedente tem uma interpretação geométrica interessante. As curvas

de nível de f são elipsoides cuja excentricidade depende da diferença entre o maior e o menor autovalor de A . Se $\lambda_1 \approx \lambda_n$, então as curvas de nível são esferas e a convergência ocorre em um único passo. Entretanto, se $\lambda_1 \ll \lambda_n$ os elipsoides ficam muito excêntricos e a convergência se dá de forma lenta. Veja a Figura 3.2.

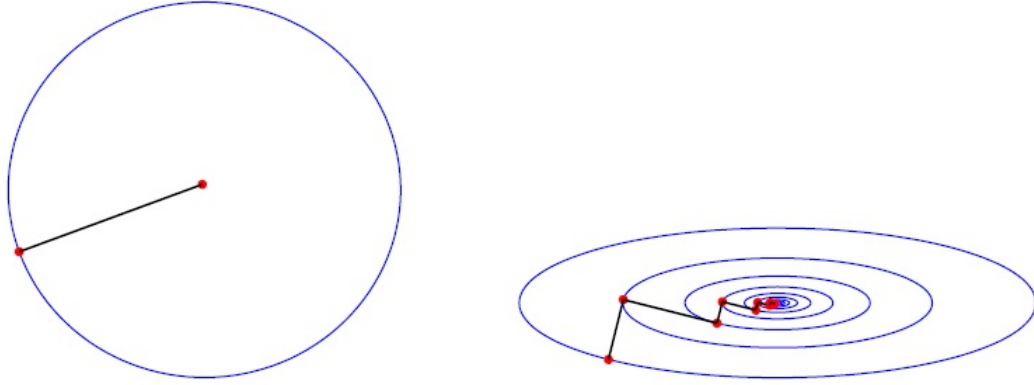


Figura 3.2: Passos do Algoritmo de Cauchy

No capítulo seguinte abordaremos o Método de Newton Puro.

4 Método de Newton

O Método de Newton é uma das ferramentas mais importantes em otimização. Tanto o algoritmo básico, chamado de Newton Puro, quanto suas variantes, que incorporam busca linear, são muito utilizados para resolver sistemas não-lineares e também para minimização de funções. Neste capítulo estudaremos o Método de Newton “Puro”.

As principais referências para este capítulo são [2], [5] e [7].

4.1 Método de Newton para resolução de equações

Considere $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ de classe C^1 e o problema de resolver o sistema (normalmente não linear)

$$F(x) = 0.$$

Como na maioria das vezes não conseguimos resolvê-lo de forma direta, os processos iterativos constituem a forma mais eficiente de lidar com tais situações. A ideia é aproximar F por seu polinômio de Taylor de primeira ordem. Dada uma estimativa $\bar{x}$, considere o sistema linear

$$F(\bar{x}) + J_F(\bar{x})(x - \bar{x}) = 0, \tag{4.1}$$

em que J_F representa a matriz jacobiana de F . Caso $J_F(\bar{x})$ seja inversível, o sistema (4.1) pode ser resolvido e fornece

$$x = \bar{x} - (J_F(\bar{x}))^{-1}F(\bar{x}).$$

Isso corresponde a uma iteração do Método de Newton para resolução de equações; veja a Figura 4.1.

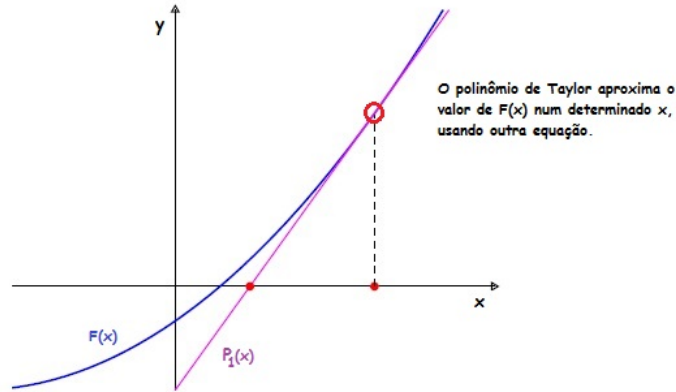


Figura 4.1: Uma iteração do Método de Newton para equações em $\mathbb{R}$.

4.2 Método de Newton para otimização irrestrita

Agora, vamos considerar o problema de minimização irrestrita

$$\min f(x) \text{ e } x \in \mathbb{R}^n \quad (4.2)$$

em que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função de classe C^2 . Os pontos estacionários desse problema são caracterizados pela equação $\nabla f(x) = 0$. Vamos aplicar a relação (4.1) para $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ dada por

$$F(x) = \nabla f(x).$$

Seja $x_k \in \mathbb{R}^n$ uma aproximação de um ponto estacionário $\bar{x}$ do problema (4.2). A aproximação seguinte x_{k+1} é computada como solução do sistema de equações lineares

$$\nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0 \quad (4.3)$$

em relação a $x \in \mathbb{R}^n$. Supondo que $\nabla^2 f(x_k)$ seja não-singular para todo $k \in \mathbb{N}$,

obtemos o seguinte esquema iterativo:

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k), \quad k = 0, 1, \dots \quad (4.4)$$

4.3 Algoritmo

Com base na relação (4.4), podemos formalizar o Método de Newton para minimizar a função f . Basicamente, temos três variantes do algoritmo. Uma delas é o Método “Puro”, onde não fazemos busca unidirecional e aceitamos o passo completo ($t_k = 1$ para todo $k \in \mathbb{N}$). As outras duas (Exata ou Armijo) podem ser encontradas em [7].

Algoritmo

Dado: $x_0 \in \mathbb{R}^n$

$k = 0$

REPITA enquanto $\nabla f(x_k) \neq 0$

Defina $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

Determine o tamanho do passo $t_k > 0$

Faça $x_{k+1} = x_k + t_k d_k$

$k = k + 1$

Exemplo 4.1. *Vamos determinar, usando o método de Newton, a menor raiz positiva da equação $4 \cos(x) - e^x = 0$, com erro relativo $\left(= \frac{|x_{k+1} - x_k|}{|x_{k+1}|} \right)$ inferior a 10^{-2} , na vizinhança do ponto $x_0 = 1$. Temos*

$$f(x) = 4 \cos(x) - e^x \quad e \quad f'(x) = -4 \operatorname{sen}(x) - e^x.$$

Portanto,

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{4 \cos(x_k) - e^{x_k}}{-4 \operatorname{sen}(x_k) - e^{x_k}},$$

de onde segue que

$$x_1 = 1 - \frac{4 \cos(1) - e^1}{-4 \sin(1) - e^1} = 0,908 \quad e \quad \text{erro relativo} = \frac{|x_1 - x_0|}{|x_1|} = 0,101.$$

Como devemos ter o erro relativo inferior a 10^{-2} , busquemos por x_2 . Com efeito,

$$x_2 = 0,908 - \frac{4 \cos(0,908) - e^{0,908}}{-4 \sin(0,908) - e^{0,908}} = 0,905 \quad e \quad \text{erro relativo} = \frac{|x_2 - x_1|}{|x_2|} = 0,0033.$$

O Algoritmo de Newton pode não estar bem definido, caso a matriz Hessiana $\nabla^2 f(x_k)$ seja singular. Além disso, mesmo que o passo d_k seja calculado, essa direção pode não ser de descida. Entretanto, se $\nabla^2 f(x_k)$ é definida positiva, então o passo d_k está bem definido e é uma direção de descida.

O passo do Método de Newton também pode ser obtido por uma abordagem diferente da que foi exposta anteriormente. Para isso, considera-se a aproximação de Taylor de segunda ordem de f , dada por

$$p(x) = f(x_k) + \nabla f(x_k)^T(x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).$$

Com o objetivo de minimizar p , fazemos

$$\nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = \nabla p(x) = 0,$$

obtendo exatamente o passo d_k do Algoritmo de Newton (veja a Figura 4.2).

Essa última abordagem sugere que se o Método de Newton for aplicado em uma função quadrática, então basta uma iteração para resolver o problema. Com efeito, considere a função quadrática $f(x) = \frac{1}{2}x^T A x + b^T x + c$. Dado $x_0 \in \mathbb{R}^n$, obtemos:

$$d_0 = -(\nabla^2 f(x_0))^{-1} \nabla f(x_0) = -A^{-1}(Ax_0 + b) = -x_0 - A^{-1}b.$$

Portanto, o minimizador $\bar{x}$ é obtido em um só passo, pois

$$x_1 = x_0 + d_0 = -A^{-1}b = \bar{x}.$$

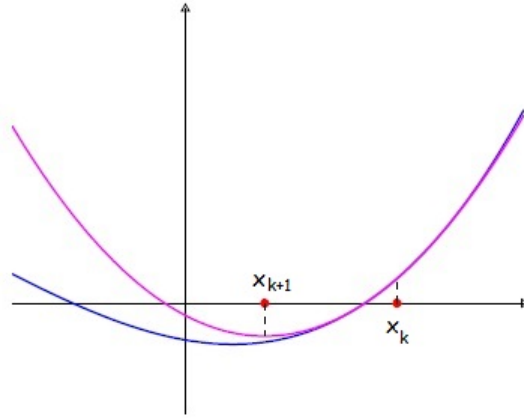


Figura 4.2: Uma iteração do Método de Newton.

4.4 Convergência

A direção de Newton pode não ser de descida. Se assim for, não garantimos convergência global quando o problema a ser resolvido envolve uma função arbitrária.

Lema 4.2. *Suponha que $\nabla^2 f(\bar{x}) > 0$. Então, existem constantes $\delta_1 > 0$ e $M > 0$ tais que*

$$\nabla^2 f(x) > 0 \text{ e } \|(\nabla^2 f(x))^{-1}\| \leq M,$$

para todo $x \in B(\bar{x}, \delta_1)$.

Demonstração. Seja $\lambda > 0$ o menor autovalor de $\nabla^2 f(\bar{x})$. Pela continuidade de $\nabla^2 f$, dado $\varepsilon = \frac{\lambda}{2}$, existe $\delta_1 > 0$ tal que

$$\|\nabla^2 f(x) - \nabla^2 f(\bar{x})\| \leq \frac{\lambda}{2}, \quad (4.5)$$

para todo $x \in B(\bar{x}, \delta_1)$. Assim, dado $d \in \mathbb{R}^n$, com $\|d\| = 1$, temos que

$$d^T \nabla^2 f(x) d = d^T \nabla^2 f(\bar{x}) d + d^T [\nabla^2 f(x) - \nabla^2 f(\bar{x})] d. \quad (4.6)$$

Podemos afirmar através do Lema 1.26 que $d^T \nabla^2 f(x) d \geq \lambda$.

Agora, através da Desigualdade de Cauchy-Schwarz (Teorema 1.23), podemos garantir que:

$$d^T[\nabla^2 f(x) - \nabla^2 f(\bar{x})]d \leq \|d\| \|\nabla^2 f(x) - \nabla^2 f(\bar{x})\| \|d\| \leq \frac{\lambda}{2}.$$

Portanto, pela relação (4.6), concluímos que

$$d^T \nabla^2 f(x) d \geq \lambda - \frac{\lambda}{2} = \frac{\lambda}{2},$$

provando que $\nabla^2 f(x)$ é definida positiva para todo $x \in B(\bar{x}, \delta_1)$. Para provarmos a desigualdade $\|(\nabla^2 f(x))^{-1}\| \leq M$, tomemos $x \in B(\bar{x}, \delta_1)$. Chamemos $A = \nabla^2 f(\bar{x})$ e $B = \nabla^2 f(x)$. Usando novamente o Lema 1.26, agora aplicado em A^2 , obtemos:

$$\|Ad\|^2 = \langle Ad, Ad \rangle = (Ad)^T Ad = d^T A^t Ad = d^T A^2 d \geq \lambda^2 \|d\|^2 = \lambda^2 d^T d$$

para todo $d \in \mathbb{R}^n$. Através da relação (4.5), concluímos que

$$\|Bd\| = \|Ad + (B - A)d\| \geq \|Ad\| - \|(B - A)d\| \geq \lambda \|d\| - \frac{\lambda}{2} \|d\| = \frac{\lambda}{2} \|d\|.$$

Considere agora $y \in \mathbb{R}^n$, com $\|y\| = 1$. Aplicando a relação acima para $d = B^{-1}y$, obtemos

$$1 = \|y\| = \|BB^{-1}y\| \geq \frac{\lambda}{2} \|B^{-1}y\|.$$

Mas,

$$\|B^{-1}\| = \sup_{x \neq 0} \frac{\|B^{-1}x\|}{\|x\|} = \sup_{x \neq 0} \left\| B^{-1} \frac{x}{\|x\|} \right\| \leq \frac{2}{\lambda}.$$

Então, pela desigualdade acima, chamando $M = \frac{2}{\lambda}$, segue que

$$\|(\nabla^2 f(x))^{-1}\| = \|B^{-1}\| \leq M,$$

como queríamos demonstrar. □

Lema 4.3. *Sejam $U \subset \mathbb{R}^n$ aberto convexo e $\beta = \sup_{x,y \in U} \|\nabla^2 f(x) - \nabla^2 f(y)\|$. Então,*

$$\|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| \leq \beta \|x - y\|,$$

para quaisquer $x, y \in U$.

Demonstração. Fixemos $y \in U$ e definamos $h(x) = \nabla f(x) - \nabla^2 f(y)x$. Temos:

$$\|J_h(x)\| = \|\nabla^2 f(x) - \nabla^2 f(y)\| \leq \beta,$$

para todo $x \in U$. Usando a Desigualdade do Valor Médio (Teorema 1.30), obtemos

$$\|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| = \|h(x) - h(y)\| \leq \beta\|x - y\|,$$

de onde segue o resultado. $\square$

Lema 4.4. *Seja $U \subset \mathbb{R}^n$ aberto e convexo. Se $\nabla^2 f$ é lipschitziana com constante de Lipschitz L , então*

$$\|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| \leq L\|x - y\|^2,$$

para quaisquer $x, y \in U$.

Demonstração. Fixados $x, y \in U$, definamos $\beta = L\|x - y\|$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^n$ por $h(z) = \nabla f(z) - \nabla^2 f(y)z$. Como $\nabla^2 f$ é lipschitziana com constante L , temos:

$$\|J_h(z)\| = \|\nabla^2 f(z) - \nabla^2 f(y)\| \leq L\|z - y\| \leq L\|x - y\| = \beta,$$

para todo $z \in [x, y]$. Usando a Desigualdade do Valor Médio (Teorema 1.30), obtemos

$$\|\nabla f(x) - \nabla f(y) - \nabla^2 f(y)(x - y)\| = \|h(x) - h(y)\| \leq \beta\|x - y\| = L\|x - y\|^2,$$

o que completa a demonstração. $\square$

Teorema 4.5. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 . Suponha que $\bar{x} \in \mathbb{R}^n$ seja um minimizador local de f , com $\nabla^2 f(\bar{x})$ definida positiva. Então, existe $\delta > 0$ tal que, se $x_0 \in B(\bar{x}, \delta)$, o Algoritmo de Newton, aplicado com $t_k = 1$ para todo $k \in \mathbb{N}$, gera uma sequência (x_k) tal que:*

i) $\nabla^2 f(x_k)$ é definida positiva, para todo $k \in \mathbb{N}$;

ii) (x_k) converge superlinearmente para $\bar{x}$;

iii) Se $\nabla^2 f$ é lipschitziana, então a convergência é quadrática.

Demonstração. Sejam δ_1 e M as constantes positivas fornecidas pelo Lema 4.2 e $U_1 = B(\bar{x}, \delta_1)$. Assim, se $x_k \in U_1$, o passo de Newton está bem definido e, como $\nabla f(\bar{x}) = 0$, vale

$$x_{k+1} - \bar{x} = (\nabla^2 f(x_k))^{-1}(\nabla f(\bar{x}) - \nabla f(x_k) - \nabla^2 f(x_k)(\bar{x} - x_k)). \quad (4.7)$$

Pela continuidade de $\nabla^2 f$, para $\varepsilon = \frac{1}{4M}$, existe $\delta_2 > 0$ tal que

$$\|\nabla^2 f(x) - \nabla^2 f(\bar{x})\| \leq \frac{1}{4M},$$

para todo $x \in B(\bar{x}, \delta_2)$. Se $y \in B(\bar{x}, \delta_2)$, vale

$$\|\nabla^2 f(y) - \nabla^2 f(\bar{x})\| \leq \frac{1}{4M}.$$

Então,

$$\begin{aligned} \|\nabla^2 f(x) - \nabla^2 f(y)\| &= \|\nabla^2 f(x) - \nabla^2 f(\bar{x}) + \nabla^2 f(\bar{x}) - \nabla^2 f(y)\| \\ &\leq \|\nabla^2 f(x) - \nabla^2 f(\bar{x})\| + \|\nabla^2 f(y) - \nabla^2 f(\bar{x})\| \\ &\leq \frac{1}{4M} + \frac{1}{4M} \\ &= \frac{1}{2M}. \end{aligned}$$

Portanto, $\sup_{x,y \in U} \|\nabla^2 f(x) - \nabla^2 f(y)\| \leq \frac{1}{2M}$, em que $U = B(\bar{x}, \delta)$ e $\delta = \min\{\delta_1, \delta_2\}$. Pelos Lemas 4.2 e 4.3, concluímos que

$$\|x_{k+1} - \bar{x}\| \leq \|(\nabla^2 f(x_k))^{-1}\| \|\nabla f(\bar{x}) - \nabla f(x_k) - \nabla^2 f(x_k)(\bar{x} - x_k)\| \leq M\beta \|x_k - \bar{x}\|$$

e, consequentemente,

$$\|x_{k+1} - \bar{x}\| \leq \frac{1}{2} \|x_k - \bar{x}\|.$$

Isso prova que a sequência (x_k) está bem definida, que $x_k \in U$, para todo $k \in \mathbb{N}$ e que $x_k \rightarrow \bar{x}$. Pelo Lema 4.2, vemos que $\nabla^2 f(x_k) > 0$ e, por conseguinte, comprovamos a validade de *i*).

Vamos, agora, provar o item *ii*), ou seja, que a convergência é superlinear. Dado $\varepsilon > 0$, considere $\delta_0 < \delta$ tal que $\sup_{x,y \in U_0} \|\nabla^2 f(x) - \nabla^2 f(y)\| < \frac{\varepsilon}{M}$, em que $U_0 = B(\bar{x}, \delta_0)$. Tome $k_0 \in \mathbb{N}$ tal que $x_k \in U_0$, para todo $k \geq k_0$. Aplicando novamente os Lemas 4.2 e 4.3 na relação (4.7), obtemos

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

concluindo o desejado.

Finalmente, se $\nabla^2 f$ é Lipschitziana, podemos usar os Lemas 4.2 e 4.4 em (4.7) para verificar que

$$\|x_{k+1} - \bar{x}\| \leq ML \|x_k - \bar{x}\|^2,$$

obtendo *iii*). □

Na sequência, estudaremos o Método da Seção Áurea, tão comentado nos últimos capítulos.

5 Método da Seção Áurea

O Método da Seção Áurea será apresentado neste capítulo. Para estudar esse método, utilizamos as referências [1] e [7].

Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$, um ponto $\bar{x} \in \mathbb{R}^n$ e uma direção de descida $d \in \mathbb{R}^n$, queremos encontrar $\bar{t} > 0$ tal que

$$f(\bar{x} + \bar{t}d) < f(\bar{x}).$$

Como deixamos em aberto o tamanho do passo t , precisamos balanceá-lo com o decréscimo promovido em f . Veremos, nas próximas linhas, a abordagem de uma busca exata a partir do ponto $\bar{x}$ segundo a direção d .

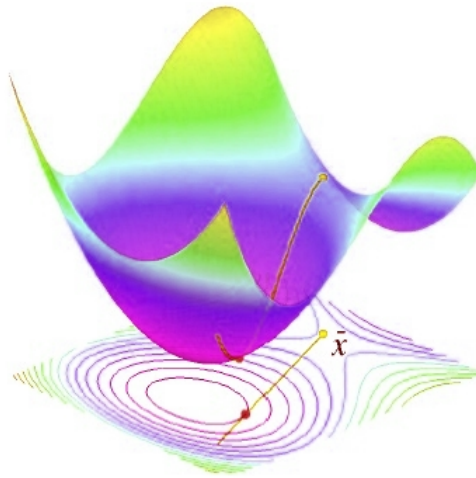


Figura 5.1: Busca unidimensional exata.

Nosso objetivo nesse caso consiste, então, em minimizar f a partir do ponto $\bar{x}$ na direção d . Mais precisamente, temos que resolver o problema

$$\text{minimizar } f(\bar{x} + td) \quad (5.1)$$

sujeito a $t > 0$.

Esse problema é, em geral, difícil de ser resolvido de forma exata. Entretanto, para certas funções especiais, existem algoritmos para resolvê-lo. Antes porém, vejamos o exemplo a seguir, resolvido de forma direta.

Exemplo 5.1. Considere $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, definida por $f(x_1, x_2) = \frac{1}{2}(x_1 - 2)^2 + (x_2 - 1)^2$, para $(x_1, x_2) \in \mathbb{R}^2$, $\bar{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ e $d = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$. Façamos a busca exata de um minimizador de f a partir de $\bar{x}$, na direção d .

- A direção d é de descida, pois $\nabla f(\bar{x})^T d = (-1 - 2) \begin{pmatrix} 3 \\ 1 \end{pmatrix} = -5 < 0$.
- Definamos $\varphi(t) = f(\bar{x} + td) = f(1 + 3t, t) = \frac{11t^2}{2} - 5t + \frac{3}{2}$, cujo minimizador satisfaz $\varphi'(t) = 11t - 5 = 0$.
- Assim, $\bar{t} = \frac{5}{11}$ e $\bar{x} + \bar{t}d = \frac{1}{11} \begin{pmatrix} 26 \\ 5 \end{pmatrix} \approx \begin{pmatrix} 2,36 \\ 0,45 \end{pmatrix}$.

Na prática é claro que os problemas são bem mais complexos que o exemplo anterior e só podem ser resolvidos por meio de algoritmos. Ao aplicarmos o Método da Seção Áurea em funções unimodais, obtemos resultados satisfatórios, por isso definiremos a seguir esse tipo de função. Na sequência, veremos o Algoritmo da Seção Áurea, que encontra um ponto próximo de um minimizador com a precisão que se queira (conforme ilustrado na Figura 5.1).

Uma outra abordagem, que não será tratada neste trabalho, é a busca inexata, através de uma redução suficiente de f que seja de certo modo proporcional ao tamanho do passo, como por exemplo, o método de Armijo. Essa opção é bastante

útil em situações onde não convém aplicar a busca exata, ou porque φ não é unimodal, ou pelo alto custo computacional.

Definição 5.2. *Uma função contínua $\varphi : [0, \infty) \rightarrow \mathbb{R}$ é dita unimodal quando admite um intervalo de minimizadores $[t_1, t_2]$, é estritamente decrescente em $[0, t_1]$ e estritamente crescente em $[t_2, \infty)$.*

Veja, na Figura 5.2, exemplos de funções unimodais a seguir e note que o intervalo de minimizadores $[t_1, t_2]$ pode ser degenerado, como ilustrado no segundo gráfico.

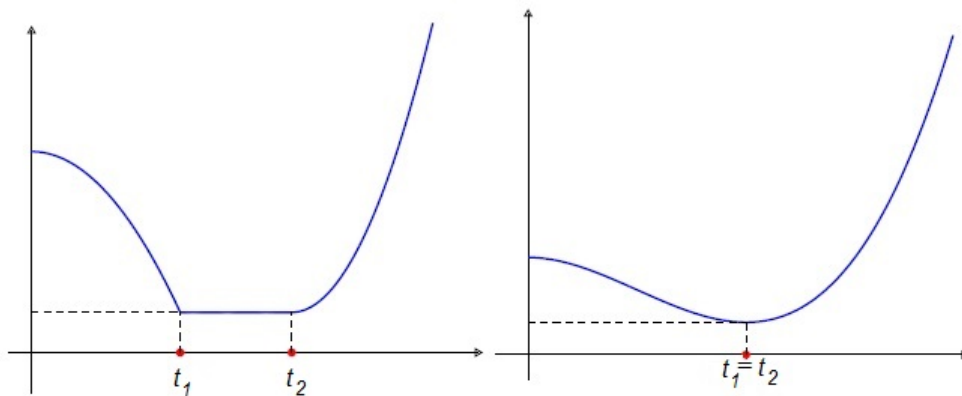


Figura 5.2: Funções unimodais.

Para facilitar a descrição do algoritmo, vamos considerar a Figura 5.3.

Suponha que o minimizador de φ pertença ao intervalo $[a, b]$.

- i) Considere $a < u < v < b$ em $[0, \infty)$;
- ii) Se $\varphi(u) < \varphi(v)$ então o trecho $(v, b]$ não pode conter o minimizador e pode ser descartado;
- iii) Se $\varphi(u) \geq \varphi(v)$ então o trecho $[a, u)$ pode ser descartado;
- iv) Particione o intervalo que restou e repita o processo.

Agora, vamos analisar como o intervalo $[a, b]$ deve ser particionado. A obtenção desse intervalo, que deve conter um minimizador de φ , será tratada adiante. A

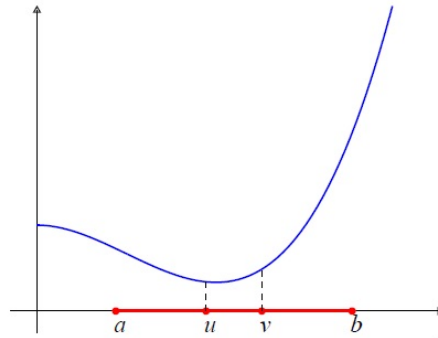


Figura 5.3: Seção Áurea.

estratégia mais natural é dividir o intervalo em três partes iguais, ou seja, definir

$$u = a + \frac{1}{3}(b - a) \quad \text{e} \quad v = a + \frac{2}{3}(b - a).$$

Dessa forma, descartamos $\frac{1}{3}$ do intervalo a cada iteração, conforme ilustrado na Figura 5.4. Além disso, se o intervalo descartado for o $(v, b]$, temos como novo intervalo $[a^+, b^+]$, onde $a^+ = a$ e $b^+ = v$, mas não podemos utilizar o antigo ponto u , calculado na iteração anterior, o que é uma desvantagem.

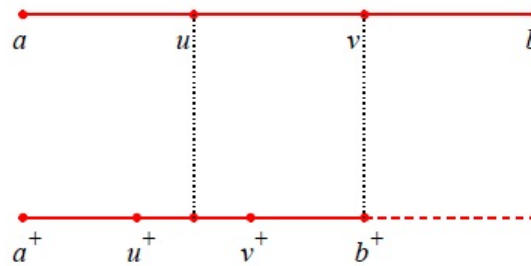


Figura 5.4: Intervalo dividido em três partes iguais.

Uma outra estratégia é escolher u e v que dividem o segmento $[a, b]$ na razão áurea, de acordo com a definição dada abaixo.

Definição 5.3. Um ponto c divide o segmento $[a, b]$ na razão áurea quando a razão entre o segmento todo e o maior segmento é igual a razão entre o maior e o menor segmento. Tal razão é conhecida como o número de ouro e vale $\frac{\sqrt{5}+1}{2} \approx 1,618$.

Portanto, a maior medida equivale a 1,618 da menor, ou a menor equivale a 0,618 (inverso do número de ouro) da maior.

Dessa forma, temos que u e v devem satisfazer

$$\frac{b-u}{b-a} = \frac{u-a}{b-u} \quad \text{e} \quad \frac{v-a}{b-a} = \frac{b-v}{v-a}. \quad (5.2)$$

Considerando θ_1 e θ_2 tais que

$$u = a + \theta_1(b-a) \quad \text{e} \quad v = a + \theta_2(b-a) \quad (5.3)$$

e substituindo u em (5.2), temos:

$$\begin{aligned} \frac{b-(a+\theta_1(b-a))}{b-a} &= \frac{a+\theta_1(b-a)-a}{b-(a+\theta_1(b-a))} \\ \frac{(b-a)(1-\theta_1)}{b-a} &= \frac{\theta_1(b-a)}{(1-\theta_1)(b-a)}, \end{aligned}$$

de onde se obtém

$$1-\theta_1 = \frac{\theta_1}{1-\theta_1}. \quad (5.4)$$

De forma análoga, se substituirmos v em (5.2), obtemos:

$$\theta_2 = \frac{1-\theta_2}{\theta_2}. \quad (5.5)$$

Como $u, v \in [a, b]$, encontramos $\theta_1 = \frac{3-\sqrt{5}}{2} \approx 0,382$ e $\theta_2 = \frac{\sqrt{5}-1}{2} \approx 0,618$.

Note que

$$\theta_1 + \theta_2 = 1 \quad \text{e} \quad (\theta_2)^2 = \theta_1. \quad (5.6)$$

Uma das vantagens da divisão na razão áurea em relação à divisão em três partes iguais é que descartamos mais de 38% do intervalo ao invés de 33,33%.

Outra vantagem é que podemos aproveitar o ponto u ou o ponto v após termos descartado o intervalo $[v, b]$ ou $[a, u]$ na iteração anterior. Indicamos por $[a^+, b^+]$ o novo intervalo que será particionado pelos pontos u^+ e v^+ . Conforme veremos no próximo resultado, o ponto v é aproveitado na próxima etapa e passa a ser u^+ quando descartamos $[a, u)$. Assim, o valor da função $\varphi(v)$ é aproveitado para a próxima etapa.

Lema 5.4. *Na seção áurea, se $[a, u)$ é descartado, então $u^+ = v$.*

Demonstração. Como $[a, u)$ foi descartado, então $a^+ = u$ e $b^+ = b$. Portanto, temos que:

$$\begin{aligned} u^+ &= a^+ + \theta_1(b^+ - a^+) \\ &= u + \theta_1(b - u) \\ &= a + \theta_1(b - a) + \theta_1(b - (a + \theta_1(b - a))) \\ &= a + (2\theta_1 - (\theta_1)^2)(b - a). \end{aligned}$$

Da relação (5.4) vem que $(\theta_1)^2 = 3\theta_1 - 1$. Por conseguinte,

$$\begin{aligned} u^+ &= a + (2\theta_1 - 3\theta_1 + 1)(b - a) \\ &= a + (1 - \theta_1)(b - a), \end{aligned}$$

e, por (5.6), concluímos que

$$u^+ = a + \theta_2(b - a) = v.$$

□

A Figura 5.5 ilustra o que foi abordado no Lema 5.4.

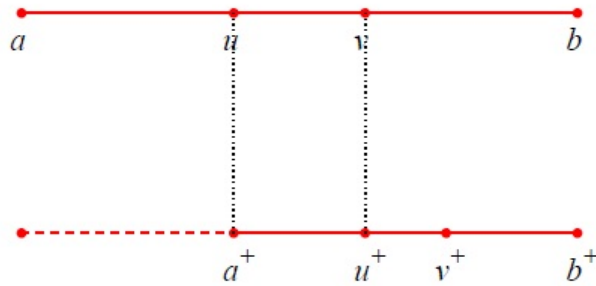


Figura 5.5: Partição do intervalo $[a, b]$ - Lema 5.4.

Lema 5.5. Na seção áurea, se $(v, b]$ é descartado, então $v^+ = u$.

Demonstração. Como $(v, b]$ é descartado, então $a^+ = a$ e $b^+ = v$. Das relações (5.3) e (5.4), vem que:

$$\begin{aligned}
 v^+ &= a^+ + \theta_2(b^+ - a^+) \\
 &= a + \theta_2(v - a) \\
 &= a + \theta_2(a + \theta_2(b - a) - a) \\
 &= a + \theta_2^2(b - a) \\
 &= a + \theta_1(b - a) \\
 &= u,
 \end{aligned}$$

conforme queríamos demonstrar. □

A Figura 5.6 ilustra o que foi estabelecido no Lema 5.5.

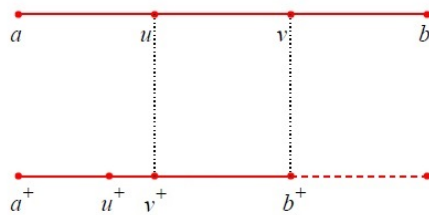


Figura 5.6: Partição do intervalo $[a, b]$ - Lema 5.5.

Apresentamos agora o Algoritmo da Seção Áurea, o qual tem duas fases. Na primeira, obtemos um intervalo $[a, b]$ que contém um minimizador de φ . A idéia desta etapa é considerar um intervalo inicial $[0, 2\rho]$, com $\rho > 0$, e ampliá-lo, deslocando para a direita, até que um crescimento de φ seja detectado. Na segunda fase, o intervalo $[a, b]$ é reduzido, por meio do descarte de subintervalos, até que reste um intervalo de tamanho suficiente para que uma precisão ε seja alcançada.

5.1 Algoritmo

Dados: $\varepsilon > 0$, $\rho > 0$, $\theta_1 = \frac{3 - \sqrt{5}}{2}$, $\theta_2 = 1 - \theta_1$

Fase 1: Obtenção do intervalo $[a, b]$

$a_0 = 0$, $s_0 = \rho$ e $b_0 = 2\rho$

$k = 0$

REPITA enquanto $\varphi(b_k) < \varphi(s_k)$

$a_{k+1} = s_k$, $s_{k+1} = b_k$ e $b_{k+1} = 2b_k$

$k = k + 1$

$a = a_k$ e $b = b_k$

Fase 2: Obtenção de $\bar{t} \in [a, b]$

$a_0 = a$, $b_0 = b$

$u_0 = a_0 + \theta_1(b_0 - a_0)$, $v_0 = a_0 + \theta_2(b_0 - a_0)$

$k = 0$

REPITA enquanto $(b_k - a_k) > \varepsilon$

SE $\varphi(u_k) < \varphi(v_k)$

$a_{k+1} = a_k$, $b_{k+1} = v_k$, $u_{k+1} = a_{k+1} + \theta_1(b_{k+1} - a_{k+1})$

SENÃO

$a_{k+1} = u_k$, $b_{k+1} = b_k$, $u_{k+1} = a_{k+1} + \theta_2(b_{k+1} - a_{k+1})$

$k = k + 1$

Defina $\bar{t} = \frac{u_k + v_k}{2}$

Na primeira fase, após um número finito de etapas, é possível encontrar um intervalo $[a, b]$ que contém pelo menos um minimizador de φ . Isto é o que provaremos

na sequência.

Teorema 5.6. *O Algoritmo da Seção Áurea, na primeira fase, encontra um intervalo $[a, b]$ que contém pelo menos um minimizador da função unimodal $\varphi : [0, +\infty) \rightarrow \mathbb{R}$ em um número finito de iterações.*

Demonstração. Seja $[t_1, t_2]$ o conjunto de minimizadores de φ .

Inicialmente, afirmamos que o loop da primeira fase do Algoritmo da Seção Áurea é finito. Com efeito, suponha que $\varphi(b_k) < \varphi(s_k)$ para todo $k \in \mathbb{N}$. Então, $s_k \rightarrow \infty$ quando $k \rightarrow \infty$, pois $s_{k+1} = b_k = 2b_{k-1} = 2s_k$. Assim, existe $k_0 \in \mathbb{N}$ tal que $s_{k_0} \geq t_1$. Como φ é unimodal, φ é não-decrescente em $[t_1, \infty)$, de onde segue que $\varphi(b_{k_0}) \geq \varphi(s_{k_0})$, uma contradição. Portanto, a Fase 1 do algoritmo termina em um certo $k \in \mathbb{N}$. Basta-nos verificar que o intervalo obtido contém, de fato, um minimizador de φ . Temos dois casos a considerar:

i) Caso 1: $s_k < t_1$.

Neste caso, temos que $b_k > t_2$, pois do contrário teríamos $\varphi(b_k) < \varphi(s_k)$. Assim,

$$[t_1, t_2] \subset [s_k, b_k] \subset [a_k, b_k].$$

ii) Caso 2: $s_k \geq t_1$.

Neste caso, temos que $a_k < t_1$. De fato, note que

$$\begin{cases} a_k = 0, & \text{se } k = 0 \\ a_k = s_{k-1}, & \text{caso contrário.} \end{cases} \quad (5.7)$$

Se $a_k = s_{k-1} \geq t_1$, então $\varphi(b_{k-1}) \geq \varphi(s_{k-1})$ e a Fase 1 terminaria na iteração $k-1$ ao invés da iteração k . Portanto, temos $a_k < t_1 \leq s_k < b_k$, o que implica que $t_1 \in [a_k, b_k]$.

O primeiro gráfico da Figura 5.7 ilustra o primeiro caso da demonstração e o segundo gráfico ilustra o segundo caso.

□

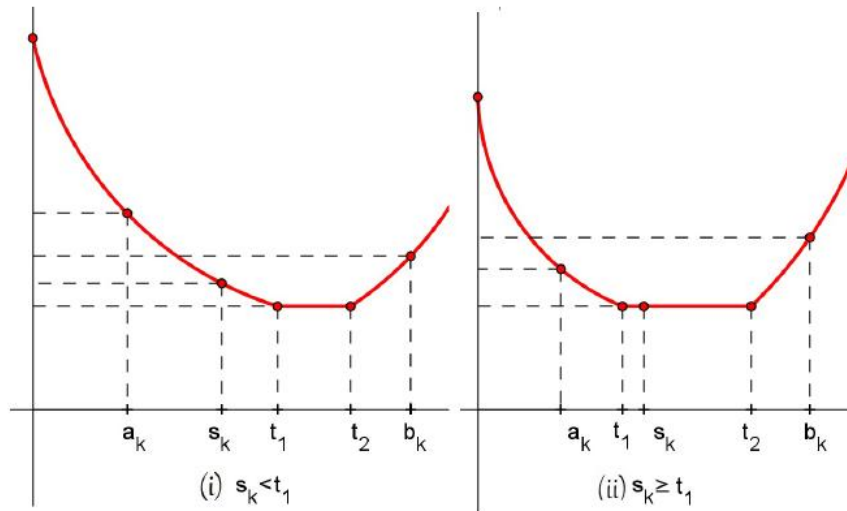


Figura 5.7: Análise da primeira etapa do algoritmo da Seção Áurea.

A seguir, provaremos um resultado que analisa se na segunda fase do algoritmo, ao descartar um dos intervalos, aquele que sobrou contém um minimizador.

Teorema 5.7. *Seja $\varphi : [0, +\infty) \rightarrow \mathbb{R}$ uma função unimodal.*

- i) Se $\varphi(v) \leq \varphi(u)$ e o intervalo $[a, u)$ é descartado, então o intervalo que sobrou $[u, b]$ contém pelo menos um minimizador.*
- ii) Se $\varphi(v) > \varphi(u)$ e o intervalo $(v, b]$ é descartado, então o intervalo que sobrou $[a, v]$ contém pelo menos um minimizador.*

Demonstração. Considere $[t_1, t_2]$ o intervalo de minimizadores da função φ .

- i)* Suponha que não exista um minimizador de φ em $[u, b]$. Portanto, existe um mínimo t^* no intervalo $[a, u)$. Note que $t_2 < u$ pois do contrário teríamos $t_2 > b$ e assim $[u, b] \subset [t^*, t_2] \subset [t_1, t_2]$, o que é uma contradição, visto que estamos supondo que não existe minimizador em $(u, b]$. Para entender melhor os argumentos utilizados, veja a Figura 5.8. Como φ é unimodal, ou seja, é estritamente crescente em $[t_2, \infty)$, temos que $t_2 < u < v$ implica $\varphi(u) < \varphi(v)$, o que contradiz a hipótese.

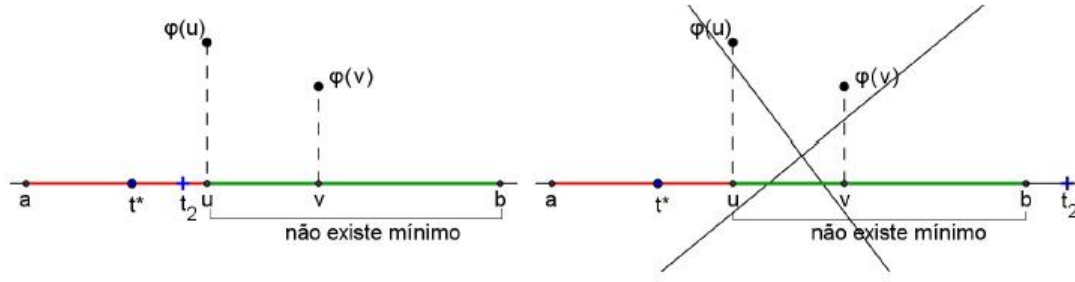


Figura 5.8: Análise do item *i*) do Teorema 5.7.

- ii*) Suponha que não exista um minimizador de φ no intervalo $[a, v]$. Então, existe um mínimo t^* no intervalo $(v, b]$. Note que $t_1 > v$, pois do contrário teríamos $t_1 < a$ e assim $[a, v] \subset [t_1, t^*] \subset [t_1, t_2]$, o que é uma contradição, uma vez que estamos supondo que não existe minimizador de φ em $[a, v]$. Como φ é unimodal, ou seja, é estritamente decrescente em $[0, t_1]$, temos que $u < v < t_1$ implica $\varphi(u) > \varphi(v)$, o que contradiz a hipótese.

□

5.2 Convergência

Antes de analisarmos a convergência do método, provaremos um resultado auxiliar que assegura que o tamanho do intervalo $[a_k, b_k]$, obtido na primeira etapa do algoritmo que contém o minimizador da função, tende a zero quando k tende a ∞ .

Teorema 5.8. *Se $[a_k, b_k]$ é o intervalo obtido pelo Algoritmo da Seção Áurea, então $b_k - a_k \rightarrow 0$, quando $k \rightarrow \infty$.*

Demonstração. Seja r_k o tamanho do intervalo $[a_k, b_k]$, ou seja, $r_k = b_k - a_k$. Como o Método da Seção Áurea descarta mais de 38% do intervalo $[a_0, b_0]$, ou melhor, descarta $\theta_1 = \frac{3 - \sqrt{5}}{2} \approx 0.382$ do intervalo $[a_0, b_0]$, temos que:

$$b_1 - a_1 = r_1 = r_0 - \theta_1 r_0 = r_0 \theta_2.$$

Repetindo o processo com o intervalo de cada iteração, obtemos:

$$b_2 - a_2 = r_2 = r_0\theta_2 - \theta_1(r_0\theta_2) = r_0\theta_2(1 - \theta_1) = r_0(\theta_2)^2$$

...

$$b_k - a_k = r_0(\theta_2)^k.$$

Como r_0 é uma constante positiva e $\theta_2 \in (0, 1)$, temos que

$$\lim_{k \rightarrow \infty} r_0(\theta_2)^k = r_0 \lim_{k \rightarrow \infty} (\theta_2)^k = 0.$$

Consequentemente,

$$\lim_{k \rightarrow \infty} b_k - a_k = 0$$

e a prova está completa. $\square$

Podemos afirmar que as sequências (a_k) , (u_k) , (v_k) e (b_k) convergem? Pois bem, como as sequências (a_k) e (b_k) são monótonas, limitadas inferiormente por a_0 e superiormente por b_0 , de acordo com o Teorema 1.10, podemos inferir que essas sequências convergem. O próximo teorema estabelece a convergência dessas sequências para um minimizador de φ .

Teorema 5.9. *Se $\varphi : [0, \infty) \rightarrow \infty$ é uma função unimodal, então as sequências (a_k) , (u_k) , (v_k) e (b_k) convergem para um minimizador de φ em $[t_1, t_2]$.*

Demonstração. Como (a_k) é não-decrescente e limitada, então (a_k) é uma sequência convergente, digamos que $a_k \rightarrow \bar{a}$. Da mesma forma, como (b_k) é não-crescente e limitada, temos que (b_k) é convergente. Digamos que $b_k \rightarrow \bar{b}$. Pelo Teorema 5.8, sabemos que $a_k - b_k \rightarrow 0$. Mas, $a_k - b_k \rightarrow \bar{a} - \bar{b}$. Logo, $\bar{a} = \bar{b}$. Chamemos $\bar{t} = \bar{a} = \bar{b}$.

Como $a_k \leq u_k \leq v_k \leq b_k$, o Teorema do Confronto (Teorema 1.15) garante que

$$u_k \rightarrow \bar{t} \quad e \quad v_k \rightarrow \bar{t}.$$

Afirmamos que $\bar{t}$ é um minimizador de φ . Com efeito, seja r_k um minimizador de φ em $[a_k, b_k]$. Pelo Teorema do Confronto, temos que $r_k \rightarrow \bar{t}$. Como (r_k) é uma sequência de minimizadores da função φ que pertencem ao conjunto $[t_1, t_2]$ e ainda,

$r_k \rightarrow \bar{t}$, temos que $\bar{t}$ é ponto de acumulação de $[t_1, t_2]$. Como $[t_1, t_2]$ é fechado, concluímos que $\bar{t} \in [t_1, t_2]$. $\square$

Agora, vejamos um caso particular de função unimodal. Considere que a função φ seja quadrática e, portanto, tenha apenas um minimizador x^* . Nesse caso, $t_1 = t_2 = x^*$. Conforme sugere a Figura 5.9, provaremos que o minimizador da função φ pertence a todo intervalo $[a_k, b_k]$.

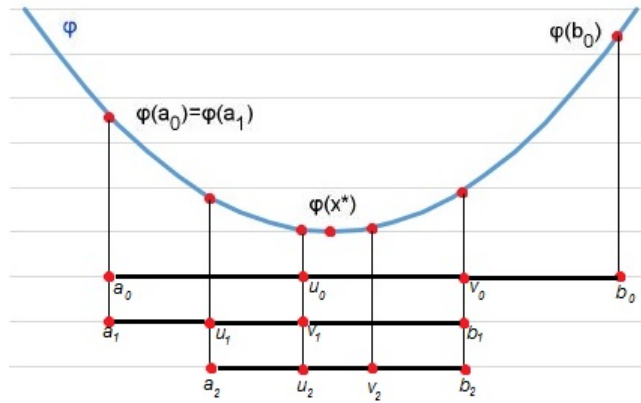


Figura 5.9: Intervalos que contêm o minimizador de uma função quadrática.

Teorema 5.10. Se x^* é o minimizador da função quadrática φ , então $x^* \in [a_k, b_k]$, para todo $k \in \mathbb{N}$.

Demonstração. Suponha que, na primeira iteração do Método da Seção Áurea, o intervalo $[a_0, u_0)$ tenha sido descartado - o outro caso é análogo. Pelo Lema 5.4, temos que $a_1 = u_0$ e $b_1 = b_0$. Assim, $x^* \in [a_1, b_1] \subset [a_0, b_0]$. Aplicando o algoritmo sucessivamente, concluímos que

$$x^* \in [a_k, b_k] \subset [a_{k-1}, b_{k-1}] \subset \cdots \subset [a_0, b_0],$$

para todo $k \in \mathbb{N}$, de onde segue o resultado. $\square$

Teorema 5.11. Se (t_k) é a sequência definida por $t_k = \frac{u_k + v_k}{2} = \frac{a_k + b_k}{2}$ e x^* o minimizador da função quadrática φ , então $t_k \rightarrow x^*$.

Demonstração. Do Teorema 5.8 segue que $b_k - a_k \rightarrow 0$. E, pelo Teorema 5.10, sabemos que $x^* \in [a_k, b_k]$, para todo $k \in \mathbb{N}$. Como $(t_k) \subset [a_k, b_k]$, concluimos que (t_k) converge para o minimizador x^* . $\square$

5.3 Velocidade de convergência

Provaremos que a velocidade de convergência da sequência (b_k) é linear com taxa de convergência igual ao inverso do número de ouro. Suponha que o intervalo $(v_k, b_k]$ tenha sido descartado. Dessa forma, temos que $a_{k+1} = a_k$, $v_{k+1} = u_k$ e $b_{k+1} = v_k$.

Teorema 5.12. *As sequências (a_k) e (b_k) geradas pelo Algoritmo da Seção Áurea têm convergência linear e a taxa de convergência é θ_2 , ou seja, $\|b_{k+1} - x^*\| \leq \theta_2 \|b_k - x^*\|$ para todo $k \in \mathbb{N}$.*

Demonstração. Considere a função $g : [a_k, v_k] \rightarrow \mathbb{R}$, definida por $g(t) = \frac{v_k - t}{b_k - t}$. Temos que

$$\begin{aligned} g'(t) &= \frac{(-1)(b_k - t) - (a_k + \theta_2(b_k - a_k) - t)(-1)}{(b_k - t)^2} \\ &= \frac{-b_k + t + a_k + \theta_2(b_k - a_k) - t}{(b_k - t)^2} \\ &= \frac{(b_k - a_k)(-1 + \theta_2)}{(b_k - t)^2} \\ &= \frac{-\theta_1(b_k - a_k)}{(b_k - t)^2} < 0, \end{aligned}$$

para todo $t \in [a_k, v_k]$. Como g' é negativa, segue que g é decrescente. Consequentemente,

$$g(t) \leq g(a_k) = \frac{v_k - a_k}{b_k - a_k} = \frac{a_k + \theta_2(b_k - a_k) - a_k}{b_k - a_k} = \theta_2,$$

para todo $t \in [a_k, v_k]$.

Em particular, para $x^* \in [a_k, v_k]$, temos que

$$g(x^*) = \frac{v_k - x^*}{b_k - x^*} = \frac{\|b_{k+1} - x^*\|}{\|b_k - x^*\|} \leq \theta_2.$$

Por conseguinte, a sequência (b_k) gerada pelo Algoritmo da Seção Áurea tem convergência linear e a taxa de convergência é o valor $\theta_2 = 0,618$ (inverso do número de ouro). Para provar que a sequência (a_k) gerada pelo Algoritmo da Seção Áurea tem convergência linear e a taxa de convergência é também o valor $\theta_2 = 0,618$, basta seguir um procedimento análogo. $\square$

Conclusão

Nesse trabalho conceitos e resultados de Análise no $\mathbb{R}^n$ e Álgebra Linear foram necessários para introduzir o estudo de convergência das sequências geradas por alguns algoritmos clássicos de otimização irrestrita que foram aqui abordados.

O método de Cauchy faz a cada iteração uma busca unidirecional na direção de maior decrescimento da função, ou seja, na direção oposta ao gradiente. A sequência gerada por esse algoritmo tem convergência global e a velocidade de convergência é linear. Se a função objetivo for de classe C^2 e o ponto inicial estiver próximo de um minimizador, o método de Newton gera uma sequência que converge superlinearmente. Caso a Hessiana da função a ser minimizada seja lipschitziana, então a convergência do método de Newton é quadrática. Portanto, o Algoritmo de Newton encontra o minimizador mais rapidamente que o Algoritmo de Cauchy.

Nos métodos de busca unidimensional, precisamos minimizar uma função a partir de um certo ponto, segundo uma direção dada, que é a direção de busca. Esse problema é equivalente a minimizar uma função real de uma variável. Um dos métodos que pode ser usado para resolver esse problema é o da Seção Áurea, que faz a minimização exata desta função. Analisamos as etapas desse algoritmo; vimos que na primeira fase o algoritmo encontra o intervalo com pelo menos um minimizador. Além disso, ao descartar intervalos em cada iteração do algoritmo, o intervalo que sobrou contém pelo menos um minimizador. Mostramos também que o algoritmo realmente converge para um minimizador. Finalmente, demonstramos que a sequência (a_k) ou (b_k) gerada pelo Algoritmo da Seção Áurea converge linearmente, com taxa de convergência de 0,618 (inverso do número de ouro).

Referências

- [1] FERNANDES, F. M. *Velocidade de Convergência de Métodos de Otimização Irrestrita*. Trabalho de conclusão de curso, Universidade Federal do Paraná, 2010.
- [2] IZMAILOV, A.; SOLODOV, M. *Otimização: Métodos Computacionais*. Volume 2. Rio de Janeiro: IMPA, 2012.
- [3] LEON, S. J. *Álgebra Linear com Aplicações*. Rio de Janeiro: LTC, 1999.
- [4] LIMA, E. L. *Curso de Análise*. Volume 2. Rio de Janeiro: IMPA, 1981.
- [5] MARTÍNEZ, J. M.; SANTOS, S. A. *Métodos Computacionais de Otimização*. Campinas: Departamento de Matemática Aplicada, IMECC - UNICAMP, 1995.
- [6] PEREIRA, L. C.; FERREIRA, M. V. Sequência de Fibonacci: História, propriedades e relações com a razão áurea. *Disc. Scientia. Série: Ciências Naturais e Tecnológicas*, v. 9, n. 1, p. 67-81, 2008.
- [7] RIBEIRO, A. A.; KARAS, E. W. *Otimização Contínua: Aspectos Teóricos e Computacionais*. São Paulo: Cengage Learning, 2013.

Índice Remissivo

Algoritmo

- da Seção Áurea, 60
- do Método de Cauchy, 35
- do Método de Newton, 45

convergência

- do Algoritmo da Seção Áurea , 63
- do Algoritmo de Cauchy, 37
- do Algoritmo de Newton, 47
- linear, 25
- quadrática, 26
- superlinear, 26

Decomposição de Cholesky, 27

derivada

- direcional, 29

Desigualdade

- de Cauchy-Schwarz, 27
- do Valor Médio, 29

direção de descida, 33

função

- convexa, 30
- unimodal, 55

limite de sequência, 19

mínimo local, 29

máximo local, 29

Método

da Seção Áurea , 53

de Cauchy, 35

de Newton, 43

do Gradiente, 35

matriz

- Hessiana, 28
- Jacobiana, 28
- negativa definida, 27
- negativa semidefinida, 27
- positiva definida, 27
- positiva semidefinida, 27

maximizador

- global, 32
- local, 32

minimizador

- global, 31
- local, 31

Número de ouro, 24

ponto

- crítico, 29
- de sela, 33
- estacionário, 29

ponto de acumulação, 20

razão áurea, 56

sequência, 19

limitada, 21
subsequência, 19

Teorema
 Bolzano-Weierstrass, 22
 do Confronto, 23

valor de aderência, 20
vetor gradiente, 28