

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE MATEMÁTICA - DAMAT

MARIA ALICE DOS SANTOS

OTIMIZAÇÃO IRRESTRITA E ESTIMAÇÃO DE PARÂMETROS

TRABALHO DE CONCLUSÃO DE CURSO

CURITIBA

2017

MARIA ALICE DOS SANTOS

OTIMIZAÇÃO IRRESTRITA E ESTIMAÇÃO DE PARÂMETROS

Trabalho de Conclusão de Curso apresentada ao Departamento acadêmico de Matemática - DAMAT da Universidade Tecnológica Federal do Paraná como requisito parcial para obtenção do grau de “Licenciado em Matemática”.

Orientador: Profa. Dra. Diane Rizzotto Rossetto

CURITIBA

2017

TERMO DE APROVAÇÃO

“OTIMIZAÇÃO IRRESTRITA E ESTIMAÇÃO DE PARÂMETROS”

por

“MARIA ALICE DOS SANTOS”

Este Trabalho de Conclusão de Curso foi apresentado às 14 horas do dia 17 de outubro de 2017 na sala Q304 como requisito parcial à obtenção do grau de Licenciado em Matemática na Universidade Tecnológica Federal do Paraná - UTFPR - Câmpus Curitiba. A aluna foi arguida pela Banca de Avaliação de Defesa abaixo assinados. Após deliberação, de acordo com o parágrafo 1º do art. 37 do Regulamento Específico do Trabalho de Conclusão de Curso para o Curso de Licenciatura em Matemática da UTFPR do Câmpus Curitiba, a Banca de Avaliação considerou o trabalho APROVADO.

_____ Prof. Dra. Diane Rizzotto Rossetto (Presidente - UTFPR/Curitiba)	_____ Prof. Dr. Adriano Verdério (Avaliador 1 – UTFPR/Curitiba)
_____ Prof. Dr. Rodolfo Gotardi Begiato (Avaliador 2 – UTFPR/Curitiba)	_____ Prof. Ms. Violeta Maria Estephan (Professora Responsável pelo TCC – UTFPR/Curitiba)
_____ Prof. Dra. Neusa Nogas Tocha (Coordenadora do Curso de Licenciatura em Matemática – UTFPR/Curitiba)	

“A Folha de Aprovação assinada encontra-se na Coordenação do Curso.”

RESUMO

SANTOS, Maria Alice. OTIMIZAÇÃO IRRESTRITA E ESTIMAÇÃO DE PARÂMETROS. 65 f. Trabalho de Conclusão de Curso – Departamento acadêmico de Matemática - DAMAT, Universidade Tecnológica Federal do Paraná. Curitiba, 2017.

Neste trabalho apresentaremos um breve estudo a respeito do problema de otimização irrestrita, buscando resolvê-lo através de processos iterativos chamados métodos de direção de descida. Neste trabalho, são abordados dois métodos clássicos: Método Gradiente e Método de Newton. Após o estudo destes métodos, são feitas comparações entre ambos quando aplicados em diferentes funções. Apresentaremos também um breve estudo a respeito da estimação de parâmetros utilizando a técnica de mínimos quadrados e reduzindo o problema de estimação de parâmetros a um problema de otimização irrestrita. A solução deste novo problema é encontrada através de dois métodos de direção de descida: Método Gradiente e Método de Gauss-Newton.

Palavras-chave: otimização, direção de descida.

SUMÁRIO

1	INTRODUÇÃO	9
2	REVISÃO DE CONCEITOS	10
2.1	REVISÃO DE ÁLGEBRA LINEAR	10
2.2	REVISÃO DE ANÁLISE	11
3	O PROBLEMA DE OTIMIZAÇÃO IRRESTRITA	14
3.1	CONDIÇÕES DE OTIMALIDADE	15
4	MÉTODOS DE DIREÇÃO DE DESCIDA	18
4.1	MÉTODO DE BUSCA UNIDIRECIONAL COM A SEÇÃO ÁUREA	20
4.1.1	Algoritmo de direção de descida com Seção Áurea	25
4.2	MÉTODO DE BUSCA UNIDIRECIONAL USANDO A CONDIÇÃO DE ARMIJO	25
4.2.1	Algoritmo de direção de descida com a Condição de Armijo	27
4.3	CONVERGÊNCIA GLOBAL DOS MÉTODOS DE DIREÇÃO DE DESCIDA	27
5	MÉTODOS CLÁSSICOS DE DESCIDA	31
5.1	MÉTODO GRADIENTE	31
5.1.1	Algoritmo do Método Gradiente	32
5.1.2	Convergência Global do Método	32
5.2	MÉTODO DE NEWTON	33
5.2.1	Algoritmo do Método de Newton	34
5.2.2	Convergência Global do Método	35
5.3	EXEMPLOS PRÁTICOS	36
6	O PROBLEMA DE ESTIMAÇÃO DE PARÂMETROS	48
6.1	MÉTODO DOS MÍNIMOS QUADRADOS	49
6.2	MÉTODO GRADIENTE	50
6.3	MÉTODO DE GAUSS-NEWTON	51
6.3.1	Algoritmo do Método de Gauss-Newton	51
6.3.2	Convergência Global do Método	51
7	EXEMPLOS PRÁTICOS	53
8	CONCLUSÕES	64
	REFERÊNCIAS	65

1 INTRODUÇÃO

A Otimização está presente em diferentes áreas da Ciência, Engenharias e Administração. O objetivo principal dos estudos nesta área é resolver o seguinte problema:

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } x \in S \end{aligned} \tag{4}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função contínua e diferenciável de classe C^2 em S e $S \subset \mathbb{R}^n$ é o conjunto viável. Neste trabalho consideramos $S = \mathbb{R}^n$, que se trata de um problema de otimização irrestrita.

Há diferentes métodos para resolver o problema (4) através de processos iterativos. Neste trabalho abordaremos dois métodos clássicos iterativos: Método Gradiente e Método de Newton. Estes tipos de métodos constroem sequências da forma $x^{k+1} = x^k + \alpha_k d^k$ tal que $(f(x^k))$ é decrescente, onde d^k representa uma determinada direção definida, e α_k o tamanho do passo dado ao longo de d^k na respectiva iteração k , com $k \in \mathbb{N}$.

Abordaremos também um breve estudo a respeito da convergência dos métodos.

Neste trabalho trataremos também do problema de estimar parâmetros utilizando as técnicas de Mínimos Quadrados e Otimização Irrestrita para encontrar uma função que melhor se ajusta a um conjunto de dados. Resolvemos este problema a partir de processos iterativos utilizando o Método Gradiente e o Método de Gauss-Newton.

Por fim traremos comparações entre os métodos estudados aplicados em algumas funções escolhidas propositalmente com propriedades em particular. Através destas comparações, mostraremos a importância da fundamentação teórica de cada método.

2 REVISÃO DE CONCEITOS

Para o desenvolvimento deste trabalho, tanto na fundamentação teórica quanto nas demonstrações, precisamos rever algumas definições e teoremas de Álgebra Linear, Cálculo e Análise Matemática que me ajudaram a ter melhor compreensão dos tópicos estudados. Estas definições e teoremas estão compilados neste capítulo e foram retiradas de (GUIDORIZZI, 2011), (LIMA, 2016) e (et.al., 2006) e (RIBEIRO e KARAS, 2013).

2.1 REVISÃO DE ÁLGEBRA LINEAR

Definição 2.1. *Seja \mathbb{V} um espaço vetorial real. Um **produto interno** sobre \mathbb{V} é uma função que a cada par de vetores v_1 e v_2 , associa um número real, denotado $\langle v_1, v_2 \rangle$ satisfazendo as propriedades:*

$$i) \langle v, v \rangle \geq 0 \text{ para todo vetor } v \text{ e } \langle v, v \rangle = 0 \text{ se, e somente se } v = 0.$$

$$ii) \langle \alpha v_1, v_2 \rangle = \alpha \langle v_1, v_2 \rangle \text{ para todo real } \alpha.$$

$$iii) \langle v_1 + v_2, v_3 \rangle = \langle v_1, v_3 \rangle + \langle v_2, v_3 \rangle.$$

$$iv) \langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle.$$

• *Observação: Neste trabalho, consideramos $\langle v_1, v_2 \rangle = v_1^\top v_2$ como o produto interno usual.*

Definição 2.2. *Seja \mathbb{V} um espaço vetorial com produto interno $\langle \cdot, \cdot \rangle$. Dizemos que dois vetores v e w são **ortogonais** se $\langle v, w \rangle = 0$. Denotamos $v \perp w$ no caso em que esses dois vetores são ortogonais.*

Definição 2.3. *Seja \mathbb{V} um espaço com produto interno $\langle \cdot, \cdot \rangle$. Definimos a **norma** (ou comprimento) de um vetor por $\|v\| = \sqrt{\langle v, v \rangle}$. Se $\|v\| = 1$, dizemos que v é um **vetor unitário** ou que v está **normalizado**.*

Definição 2.4. *Seja $M \in \mathbb{R}^{n \times n}$ uma matriz. Dizemos que M é uma **matriz simétrica** quando $M = M^\top$.*

Definição 2.5. Seja $M \in \mathbb{R}^{n \times n}$ uma matriz. Dizemos que M é uma matriz **definida positiva** se para toda transformação linear M e para determinado produto interno $\langle v, Mv \rangle$, temos que $\langle v, Mv \rangle > 0$, exceto para $v = 0$.

Definição 2.6. Seja $M \in \mathbb{R}^{n \times n}$ uma matriz. Dizemos que M é uma matriz **semidefinida positiva** se para toda transformação linear M e para determinado produto interno $\langle v, Mv \rangle$, temos $\langle v, Mv \rangle \geq 0$.

Proposição 2.7. Uma matriz M é definida positiva se, e somente se, todos os seus autovalores forem positivos.

Proposição 2.8. Uma matriz M é semidefinida positiva se, e somente se, todos os seus autovalores forem não-negativos.

Definição 2.9. Dizemos que uma matriz $M \in \mathbb{R}^{n \times n}$ é **singular** se, e somente se, o determinante de M for nulo. Caso contrário, dizemos que a matriz M é **não-singular**.

Definição 2.10. Dizemos que a matriz $M \in \mathbb{R}^{n \times n}$ é **invertível** se existir uma matriz $B \in \mathbb{R}^{n \times n}$ tal que

$$MB = BM = I_n.$$

Se M for não-singular, então não admite inversa.

Proposição 2.11. Se M for uma matriz não-singular, então M admite inversa.

2.2 REVISÃO DE ANÁLISE

Os resultados a seguir são referentes a funções reais com uma variável real.

Teorema 2.12. (Teorema do Valor Médio.) Se f for uma função contínua em $[a, b]$ e derivável em $]a, b[$. Então existe pelo menos um c em $]a, b[$ tal que

$$\frac{f(b) - f(a)}{b - a} = f'(c).$$

Teorema 2.13. (Teorema de Bolzano.) Se f for uma função contínua no intervalo $[a, b]$ e se $f(a)$ e $f(b)$ tiverem sinais opostos, então existe pelo menos um c em $[a, b]$ tal que $f(c) = 0$.

Teorema 2.14. (Teorema do Valor Intermediário.) Se f for uma função contínua no intervalo $[a, b]$ e se γ for um real compreendido entre $f(a)$ e $f(b)$, então existe pelo menos um c em $[a, b]$ tal que $f(c) = \gamma$.

Teorema 2.15. (Regra da Cadeia.) Sejam $y = f(x)$ e $x = g(t)$ duas funções deriváveis, Im_g sendo o conjunto imagem de g e D_f o conjunto domínio de f . Se $Im_g \subset D_f$, então a função composta $h(t) = f(g(t))$ é derivável e vale a regra da cadeia:

$$h'(t) = f'(g(t))g'(t).$$

Definição 2.16. Seja a função f diferenciável em $x = x_0$. A equação da reta tangente ao ponto $(x_0, f(x_0))$ é dada por $f(x) = f(x_0) + f'(x_0)(x - x_0)$ e é dita **Aproximação linear** de f em x_0 .

Os resultados a seguir são referentes a funções reais de variáveis reais.

Definição 2.17. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Dizemos que f é uma **função de classe C^n** se todas as suas derivadas parciais, até a ordem n , forem contínuas.

Definição 2.18. O ponto \bar{x} pertencente ao domínio D_f de uma função f , onde a derivada primeira $f'(\bar{x}) = 0$ é chamado **ponto estacionário** de f . Em particular para funções $f : \mathbb{R}^n \rightarrow \mathbb{R}$, temos $\nabla f(x) = 0$.

Teorema 2.19. (Polinômio de Taylor.) Seja a função f de classe C^{n+1} no intervalo I e sejam $x, x_0 \in I$. Então existe pelo menos um \bar{x} no intervalo aberto de extremos x_0 e x tal que

$$f(\bar{x}) = P(x) + \frac{f^{n+1}(x_0)}{(n+1)!}(x - x_0)^{n+1},$$

onde

$$P(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{\nabla^2 f(x_0)}{2!}(x - x_0)^2 + \dots + \frac{\nabla^n f(x_0)}{n!}(x - x_0)^n.$$

Definição 2.20. Seja a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Considere $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ um vetor tal que $\|d\| = 1$. Dizemos que a **derivada direcional** de f em x na direção de d é dada por:

$$\frac{\partial f}{\partial d}(x, f(x)) = \lim_{h \rightarrow 0} \frac{f(x + hd) - f(x)}{h}.$$

Proposição 2.21. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Se f é diferenciável no ponto x então existem as derivadas direcionais de f em qualquer direção $d \neq 0$ e são dadas por

$$\frac{\partial f}{\partial d}(x) = \langle \nabla f(x), d \rangle.$$

Definição 2.22. Seja a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 . Definimos o **polinômio de Taylor** de ordem 2 centrado em $x = a$ como:

$$f(x) = f(a)(x - a) + \nabla f(a)(x - a) + \frac{1}{2!}\nabla^2 f(a)(x - a) + o(x),$$

onde $o(v)$ é o resíduo infinitesimal, e $\lim_{x \rightarrow a} \frac{o(v)}{(x-a)^2} = 0$.

Os resultados a seguir são referentes a funções vetoriais, ou seja, $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$

Teorema 2.23. (Regra da Cadeia para várias variáveis). *Sejam $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$ funções diferenciáveis nos pontos $a \in \mathbb{R}^m$, $b = f(a) \in \mathbb{R}^n$, com Im_f sendo o conjunto imagem de f e D_g o conjunto domínio de g . Se $Im_f \subset D_g$, então $g \circ f : \mathbb{R}^m \rightarrow \mathbb{R}^p$ é diferenciável em a e*

$$(g \circ f)'(a) = g'(b)f'(a),$$

sendo $g'(b)$ a matriz Jacobiana de $g(b)$.

Os resultados a seguir são referentes a Sequências.

Definição 2.24. *Dado $a \in \mathbb{X} \subset \mathbb{R}^n$, dizemos que a é um **ponto de acumulação** ou **valor de aderência** de uma sequência (x^k) de elementos de \mathbb{X} , dois a dois distintos, se, e somente se, $\lim x^{k_n} = a$, onde (x^{k_n}) é uma subsequência de (x^k) .*

Definição 2.25. *Dizemos que uma sequência é **limitada inferiormente** quando existe $a \in \mathbb{R}^n$ tal que $a \leq x^k$ para todo $k \in \mathbb{N}$.*

Definição 2.26. *Seja (x^k) uma sequência limitada inferiormente. Então $\lim inf x^k = a$ é o menor valor de aderência, ou ponto de acumulação de (x^k) .*

Definição 2.27. *Dizemos que uma sequência (x^k) é **monótona** quando (x^k) é crescente ou decrescente.*

Definição 2.28. *Toda sequência (x^k) tal que $\lim x^k$ existe, é chamada de **sequência convergente**.*

Teorema 2.29. *Toda sequência monótona limitada é convergente.*

Corolário 2.30. *Se uma sequência monótona (x^k) possui uma subsequência (x^{k_n}) , com $n \in \mathbb{N}$, convergente, então (x^k) é convergente.*

Teorema 2.31. Bolzano-Weierstrass. *Toda sequência (x^k) limitada em \mathbb{R}^n possui uma subsequência convergente.*

Corolário 2.32. *Uma sequência limitada (x^k) em \mathbb{R}^n é convergente se, e somente se, possui um único ponto de acumulação.*

3 O PROBLEMA DE OTIMIZAÇÃO IRRESTRITA

Um problema de otimização pode ser escrito como

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } x \in S \end{aligned} \tag{7}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função contínua e diferenciável em S e $S \subset \mathbb{R}^n$ é o conjunto viável. Neste trabalho consideramos $S = \mathbb{R}^n$, neste caso se trata de um problema de otimização irrestrita.

Para avançarmos no estudo deste tipo de problema aconselhamos rever algumas definições do Cálculo Diferencial, presentes no capítulo anterior. Além de revisar duas definições importantes para o desenvolvimento deste trabalho, retiradas de (FRIEDLANDER, 2009) e formalizadas a seguir.

Definição 3.1. Um ponto $x^* \in S \subset \mathbb{R}^n$ é um **minimizador local** de f em S se, e somente se, existe $\varepsilon > 0$ tal que $f(x) \geq f(x^*)$ para todo $x \in S$ tal que $\|x - x^*\| < \varepsilon$.

Se $f(x) > f(x^*)$ para todo $x \in S$ tal que $x \neq x^*$ e $\|x - x^*\| < \varepsilon$, diremos que se trata de um **minimizador local estrito** de f em S .

Definição 3.2. Um ponto $x^* \in S \subset \mathbb{R}^n$ é um **minimizador global** de f em S se, e somente se, $f(x) \geq f(x^*)$ para todo $x \in S$.

Se $f(x) > f(x^*)$ para todo $x \in S$, diremos que se trata de um **minimizador global estrito** de f em S .

Vejamos o exemplo a seguir:

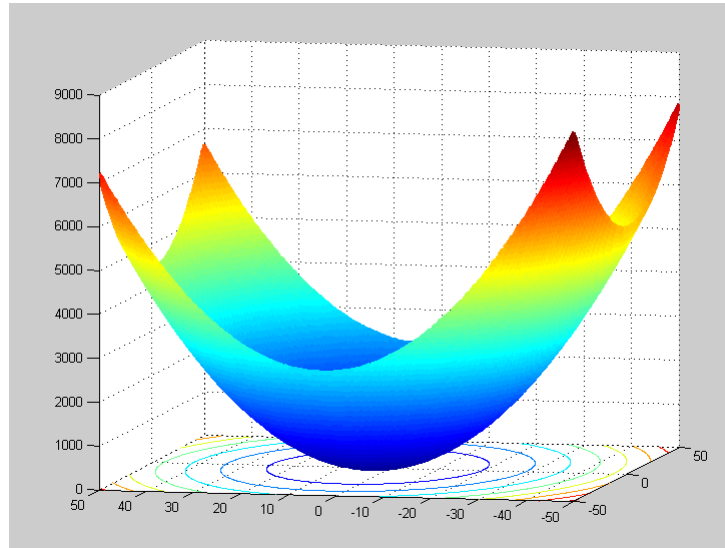


Figura 6: $f(x) = (x_1 - 3)^2 + 2(x_2 - 3)^2$

Note que a função apresenta seu valor mínimo em $x^* = (x_1, x_2) = (3, 3)$, pois para qualquer outro ponto (x_1, x_2) , a função f apresenta um valor maior que em $(x_1, x_2) = (3, 3)$. Note também que no ponto x^* temos $\nabla f(x^*) = 0$. Diante desta pequena análise intuitiva, vamos formalizar as condições que nos garantem se um ponto x é um minimizador de f . Seguindo o livro (FRIEDLANDER, 2009), as condições a seguir caracterizam uma solução do problema (7).

3.1 CONDIÇÕES DE OTIMALIDADE

As condições apresentadas aqui são referentes ao estudo do comportamento de $\nabla f(x)$ e também de $\nabla^2 f(x)$ para caracterizar soluções do problema (7).

Teorema 3.3. (Condição necessária de primeira ordem). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1$. Se x^* é um minimizador local de f em \mathbb{R} , então $\nabla f(x^*) = 0$.*

Demonstração. Fixamos $d \in \mathbb{R}^n$ arbitrário e consideramos a função $\phi : \mathbb{R} \rightarrow \mathbb{R}$ definida por

$$\phi(\alpha) = f(x^* + \alpha d).$$

Como x^* é um minimizador local de f , $\alpha = 0$ é um minimizador local de ϕ , logo $\phi'(0) = 0$. Pela regra da cadeia obtemos $\phi'(\alpha) = d^\top \nabla f(x + \alpha d)$. Substituindo em $\alpha = 0$ temos

$$0 = \phi'(0) = \nabla f(x)^{\top} d.$$

Como $d \in \mathbb{R}^n$ é arbitrário, esta igualdade significa que $\nabla f(x^*)$ é um vetor ortogonal a todos os vetores do espaço, portanto $\nabla f(x^*) = 0$. \square

Teorema 3.4. (Condição necessária de segunda ordem). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$. Se x^* é um minimizador local de f em \mathbb{R}^n , então*

1. $\nabla f(x^*) = 0$;
2. $\nabla^2 f(x^*)$ é semidefinida positiva.

Demonstração. A prova do item (1) segue da demonstração anterior. Para provar o item (2), considere $\phi(\alpha)$ como no teorema 3.3. Como $\alpha = 0$ é minimizador global e se trata de uma função $\phi(\alpha) : \mathbb{R} \rightarrow \mathbb{R}$, então $\phi''(0) \geq 0$. Pela regra da cadeia temos $\phi''(\alpha) = d^\top \nabla^2 f(x^* + \alpha d)d$. Logo,

$$\phi''(0) = d^\top \nabla^2 f(x^*)d \geq 0.$$

Portanto, $\nabla^2 f(x^*)$ é semidefinida positiva (definição 2.6). \square

Teorema 3.5. Condição suficiente de segunda ordem. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^2$, e $x \in \mathbb{R}^n$. Se $\nabla f(x^*) = 0$ e $\nabla^2 f(x^*) > 0$, então x^* é um minimizador local estrito de f em \mathbb{R}^n .*

Demonstração. Consideremos a função $\phi(\alpha) = f(x^* + \alpha d)$ como nos teoremas anteriores. A partir disso, seja $B = \{h \in \mathbb{R}^n / \|h\| = 1\}$. Consideremos a função $T : B \rightarrow \mathbb{R}$ dada por

$$T(h) = h^\top \nabla^2 f(x^*)h.$$

Como T é uma função contínua e B é um conjunto fechado e limitado, portanto T atinge um valor máximo e um valor mínimo em B . Chamemos a ao valor mínimo, então:

$$T(h) \geq a > 0, \forall h \in B.$$

Agora consideremos $d \in \mathbb{R}^n$ arbitrário não-nulo. Como $\frac{d}{\|d\|} \in B$, temos

$$d^\top \nabla^2 f(x^*)d \geq a\|d\|^2. \quad (8)$$

Desenvolvendo em série de Taylor a função $\phi(\alpha) = f(x^* + \alpha d)$ em torno de 0, temos

$$f(x^* + \alpha d) - f(x) = \nabla f(x^*)^\top d + \frac{1}{2}d^\top \nabla^2 f(x^*)d + o(\|d\|^2)$$

Pela hipótese, $\nabla f(x^*) = 0$, e pela equação (8), temos que

$$f(x^* + \alpha d) - f(x) \geq \frac{a}{2}\|d\|^2 + o(\|d\|^2).$$

Então, para todo d tal que $\|d\|$ é suficientemente pequeno, o primeiro termo do membro direito da desigualdade define o sinal deste lado. Mas,

$$\frac{a}{2}\|d\|^2 > 0,$$

pois $a > 0$. Assim, para $\|d\|$ suficientemente pequeno não-nulo (digamos $0 < \|d\| < \varepsilon$)

$$f(x^* + \alpha d) - f(x^*) > 0,$$

ou seja, $f(x^* + \alpha d) > f(x^*)$. Então para todo $x \in B(x^*, \varepsilon)$ (onde $B(x^*, \varepsilon)$ é uma bola aberta de raio ε e centro x^*), com $x \neq x^*$ temos que

$$f(x^* + \alpha d) > f(x^*).$$

Logo, x^* é um minimizador local estrito de f . □

Nos capítulos a seguir estudaremos métodos clássicos de minimização que procuram verificar as condições de otimalidade, conhecidos como métodos de direção de descida.

4 MÉTODOS DE DIREÇÃO DE DESCIDA

Antes de estudarmos os métodos propriamente ditos, vamos definir o que é uma direção de descida. Os resultados apresentados neste capítulo podem ser encontrados em (FRIEDLANDER, 2009) e (RIBEIRO e KARAS, 2013).

Os métodos clássicos de descida constroem sequências $x^{k+1} = x^k + \alpha_k d^k$ que permitem construir uma sequência $(f(x^k))$ decrescente, onde d^k é uma direção e α_k é o tamanho do passo dado a partir de x^k na direção d^k . Isso é possível se a direção d^k for de descida.

Vejam os a seguir a definição de uma direção de descida:

Definição 4.1. Dado $x \in \mathbb{R}^n$, $\alpha \in [0, 1]$ e uma direção $d \in \mathbb{R}^n / \{0\}$. Dizemos que d é uma *direção de descida* se $f(x + \alpha d) \leq f(x)$.

O próximo resultado fornece uma condição que garante que d seja uma direção de descida.

Proposição 4.2. Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $f \in C^1$, $x \in \mathbb{R}^n$ tal que $\nabla f(x) \neq 0$ e $d \in \mathbb{R}^n$ tal que $\nabla f(x)^\top d < 0$. Então existe $\delta > 0$ tal que $f(x + \alpha d) < f(x)$ para todo $\alpha \in (0, \delta)$.

Demonstração. Considere a função $\phi(\alpha) = f(x + \alpha d)$. Temos $\phi(0) = f(x)$, e pela regra da cadeia, $\phi'(0) = \nabla f(x)^\top d$. Sabemos por definição de derivada que

$$\phi'(0) = \lim_{\alpha \rightarrow 0} \frac{\phi(\alpha) - \phi(0)}{\alpha},$$

e em relação a derivada direcional,

$$\lim_{\alpha \rightarrow 0} \frac{\phi(\alpha) - \phi(0)}{\alpha} = \nabla f(x)^\top d.$$

Então, pela hipótese temos:

$$\nabla f(x)^\top d = \lim_{\alpha \rightarrow 0} \frac{\phi(\alpha) - \phi(0)}{\alpha} = \phi'(\alpha) < 0.$$

Assim, obtemos $\phi'(0) < 0$ e $\phi(\alpha) - \phi(0) < 0$, para $0 < \alpha < \delta$, e portanto,

$$f(x + \alpha d) < f(x).$$

□

Para chegar ao minimizador de f ao longo de uma direção de descida, devemos caminhar por esta direção. Entretanto, o tamanho do passo dado, ou seja, o quanto que devemos caminhar deve ser apenas o suficiente para chegar até o minimizador.

A figura a seguir nos auxilia a compreender melhor o significado do tamanho de passo a ser dado.

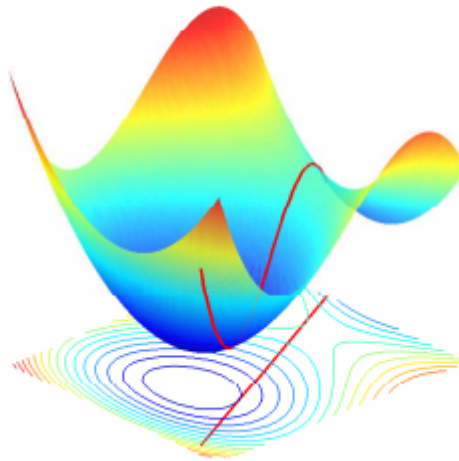


Figura 8: Direção de descida determinada

Queremos minimizar $f(x)$ ao longo da direção de descida determinada (em vermelho). Pelas curvas de nível da função, podemos ver que se andarmos muito ao longo desta direção, passamos do ponto minimizador. De maneira análoga, se andamos pouco não chegamos a este ponto.

Assim, precisamos escolher o tamanho do passo α , ou seja, o quanto andar ao longo da direção, de forma que encontremos o minimizador de $f(x)$ ao longo desta direção.

Como já vimos, direções de descida permitem construir sequências $(f(x^k))$ decrescentes, mas isto não é suficiente para garantir que chegaremos ao minimizador x^* de f . O exemplo seguinte ilustra este fato.

Exemplo 4.3. Considere a função unidimensional a seguir $f : \mathbb{R} \rightarrow \mathbb{R}$ definida por $f(x) = x^2$.

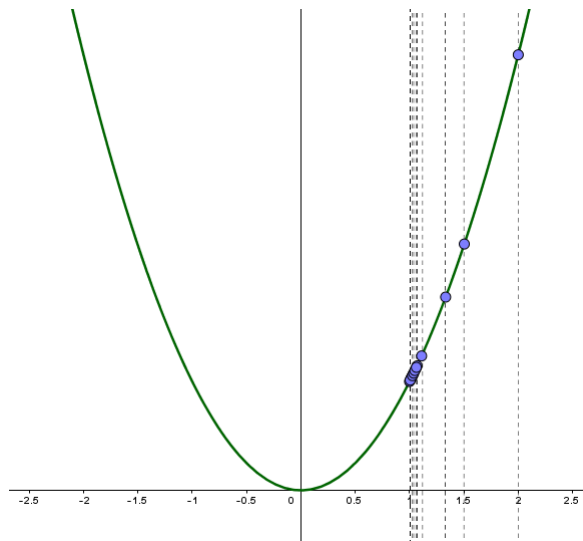


Figura 9: Redução insuficiente de $f(x) = x^2$

Considere a sequência (x^k) gerada pelos pontos $x^k = 1 + \frac{1}{k+1}$. A sequência $(f(x^k))$ gerada a partir de (x^k) é decrescente, mas (x^k) converge para $x = 1$ quando $k \rightarrow \infty$, e o minimizador desta função é $x^* = 0$. Assim, a redução de f realizada em cada iteração não foi suficiente para chegar ao minimizador x^* .

Como vimos no exemplo 4.3, apenas a direção de descida não é suficiente para encontrar o minimizador x^* , precisamos também ter o tamanho do passo α_k bem definido. Encontramos o tamanho deste passo através de processos iterativos, onde nos restringimos a estudar uma função $\phi(\alpha) : \mathbb{R} \rightarrow \mathbb{R}$ que estudaremos a seguir. A partir de agora buscamos minimizar a função ϕ a partir do ponto x na direção de d . Ou seja, precisamos resolver o problema

$$\begin{aligned} \text{minimizar } \phi(\alpha) &= f(x + \alpha d) \\ \text{sujeito a } \alpha &> 0 \end{aligned} \tag{10}$$

4.1 MÉTODO DE BUSCA UNIDIRECIONAL COM A SEÇÃO ÁUREA

Os resultados apresentados nesta seção se encontram em (RIBEIRO e KARAS, 2013) e (NOCEDAL, 2006).

O método de busca unidirecional com a Seção Áurea é eficaz para resolver o problema (10) somente se a função ϕ for unimodal, que será definida a seguir. Além disso, estudaremos

o método da Seção Áurea aplicando-o na função $\phi : [0, \infty] \rightarrow \mathbb{R}$ definida por

$$\phi(\alpha) = f(x^* + \alpha d)$$

Definição 4.4. Dizemos que a função $\phi : [0, \infty] \rightarrow \mathbb{R}$ é **unimodal** quando esta admite um conjunto de minimizadores $[t_1, t_2]$ e é estritamente crescente em $[t_2, \infty]$ e estritamente decrescente em $[0, t_1]$

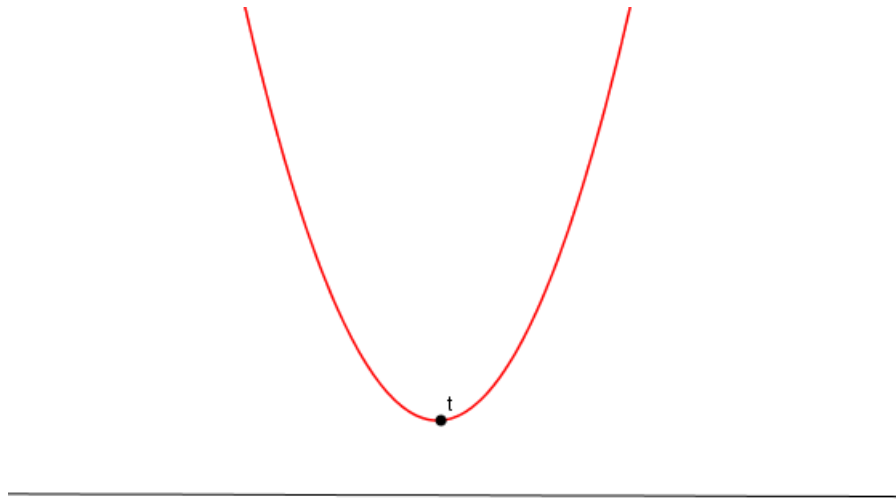


Figura 10: $f(x) = x^2 + 8$

Agora, para descrevermos o algoritmo da Seção Áurea, com o auxílio da figura a seguir listaremos os passos que devem ser seguidos respectivamente para encontrar o minimizador de ϕ e supor que este minimizador pertence a $[a, b]$.

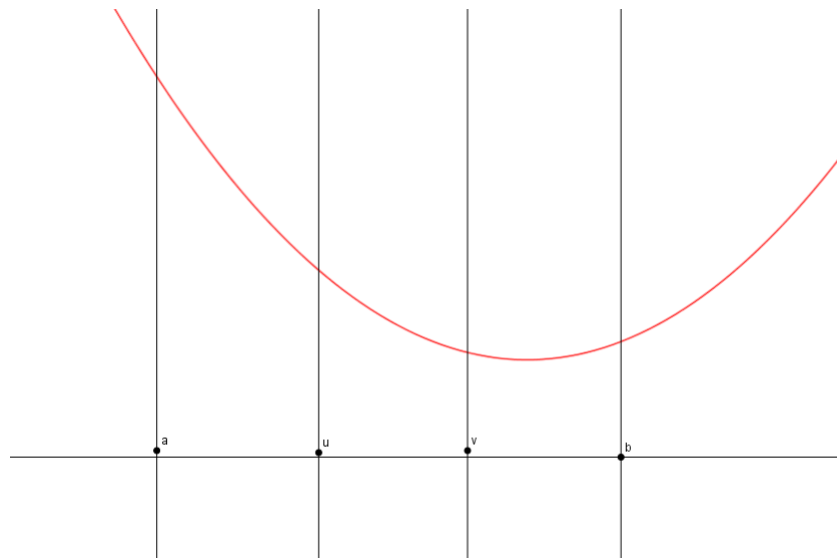


Figura 11: $f(x) = 3x^2 + 12$

Vamos seguir os seguintes passos:

- i) Considere $a < u < v < b$ em $[0, \infty]$.
- ii) Se $\phi(u) < \phi(v)$ então o intervalo $[v, b]$ deve ser descartado, pois não possui um minimizador.
- iii) Analogamente, se $\phi(u) \geq \phi(v)$ então $[a, u]$ deve ser descartado.
- iv) Particione novamente o intervalo que restou e repita o processo até encontrar um minimizador de ϕ .

Agora, estudaremos como particionar o intervalo $[a, b]$. Existem muitas formas de particioná-lo, porém algumas formas exigem maior trabalho computacional que pode ser reduzido se dividirmos o intervalo de forma estratégica. Uma das formas mais eficientes é dividir $[a, b]$ em relação à seção áurea, definida como:

Definição 4.5. Um ponto c divide o segmento $[a, b]$ na razão áurea quando a razão entre o maior segmento e o segmento todo é igual à razão entre o menor e o maior segmento. Esta razão é chamada **número de ouro** e vale $\frac{\sqrt{5}-1}{2} \approx 0,618$.

Assim, precisamos que u e v satisfaçam

$$\frac{b-u}{b-a} = \frac{u-a}{b-u} \quad e \quad \frac{v-a}{b-a} = \frac{b-v}{v-a}.$$

Agora, considerando θ_1 e θ_2 tais que

$$u = a + \theta_1(b - a) \quad e \quad v = a + \theta_2(b - a) \quad (11)$$

temos:

$$1 - \theta_1 = \frac{\theta_1}{1 - \theta_1} \quad e \quad \theta_2 = \frac{1 - \theta_2}{\theta_2} \quad (12)$$

e portanto,

$$\theta_1 = \frac{3 - \sqrt{5}}{2} = 0,382 \quad e \quad \theta_2 = \frac{\sqrt{5} - 1}{2} = 0,618,$$

ainda,

$$\theta_2 + \theta_1 = 1 \quad e \quad \theta_2^2 = \theta_1.$$

Observação 4.6. Denotaremos os novos pontos obtidos a cada iteração por r^* , onde r se refere ao ponto anterior à iteração.

A figura a seguir ilustra um processo de partição pela Seção Áurea, onde o intervalo $[v, b]$ é descartado. Note que após a primeira iteração, os novos pontos são $a = a^*$, $u = v^*$ e $v = b^*$.

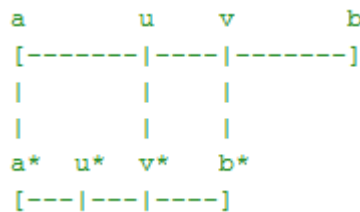


Figura 12: Partição pela Seção Áurea

Em (RIBEIRO e KARAS, 2013) é mostrado que uma das vantagens de particionarmos o intervalo desta forma é que a cada iteração, o intervalo descartado é maior do que se partionássemos, por exemplo, o intervalo em 3 partes iguais. Pois com o método da seção áurea, a cada iteração descartamos 38,2% do intervalo, ao invés de apenas 33,3% pela divisão em 3 partes iguais.

Outra vantagem é em relação a economia na avaliação da função. A cada iteração descartamos $[a, u]$ ou $[v, b]$ restando um dos segmentos que será particionado novamente. Este segmento será limitado por $[a^*, b^*]$ e particionado por u^* e v^* . Veremos no lema a seguir que a cada nova iteração o ponto u passa a ser v^* quando é descartado o intervalo $[v, b]$, dessa forma $\phi(u)$ é aproveitado.

Lema 4.7. *No método da Seção Áurea, se o segmento $[v, b]$ é descartado então $v^* = u$, e analogamente, se $[a, u]$ for descartado, então $u^* = v$.*

Demonstração. Como $[a, u]$ foi descartado $u = a^*$ e $b = b^*$, usando (11) temos que

$$v^* = a^* + \theta_2(b^* - a^*) = a + \theta_2(v - a),$$

usando agora (11) em relação a (12) temos:

$$v^* = a + \theta_2^2(b - a) = a + \theta_1(b - a) = u.$$

Portanto, $v^* = u$. A segunda afirmação é demonstrada de forma análoga. □

A seguir apresentamos um pseudocódigo do método da Seção Áurea.

4.1.1 ALGORITMO DE DIREÇÃO DE DESCIDA COM SEÇÃO ÁUREA

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

Fase 1: Obter o intervalo $[a, b]$

$a = 0$, $s = \rho$, e $b = 2\rho$

REPITA enquanto $\phi(b) < \phi(s)$

$a = s$, $s = b$ e $b = 2b$

Fase 2: Obter $\alpha \in [a, b]$

$u = a + \theta_1(b - a)$, $v = a + \theta_2(b - a)$

REPITA enquanto $(b - a) > \varepsilon$

SE $\phi(u) < \phi(v)$

$b = v$, $v = u$, $u = a + \theta_1(b - a)$

SENÃO

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

defina $\alpha = \frac{u+v}{2}$

Caso a função não seja unimodal, não podemos garantir a eficácia do método da Seção Áurea. Por isso, estudaremos a seguir um método eficaz para qualquer função.

4.2 MÉTODO DE BUSCA UNIDIRECIONAL USANDO A CONDIÇÃO DE ARMIJO

O método da Condição de Armijo encontra um passo α que garante um decréscimo suficiente da função, satisfazendo a seguinte desigualdade:

$$f(x + \alpha d) \leq f(x) + \eta \alpha \nabla f(x)^\top d \quad (13)$$

com $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ uma direção de descida e $\eta \in (0, 1)$, com $\alpha > 0$.

Vamos estudar este método mais minuciosamente, buscando uma interpretação geométrica. Primeiramente, considere a função $\phi : [0, \infty] \rightarrow \mathbb{R}$ definida como

$$\phi(\alpha) = f(x + \alpha d).$$

A aproximação linear de ϕ em torno de $\alpha = 0$ é dada por:

$$p(\alpha) = \phi(0) + \alpha \phi'(0).$$

Como $\phi(0) = f(x)$, pela regra da cadeia,

$$\phi'(\alpha) = d^\top \nabla f(x + \alpha d) \Rightarrow \phi'(0) = d^\top \nabla f(x).$$

Assim, a Condição de Armijo pode ser reescrita como:

$$\phi(\alpha) \leq q(\alpha),$$

onde temos:

$$q(\alpha) = \phi(0) + \eta \alpha \phi'(0).$$

Portanto, a Condição de Armijo é válida para os pontos α tais que o valor de $\phi(\alpha) = f(x + \alpha d)$ está abaixo de $q(\alpha) = \phi(0) + \eta \alpha \phi'(0)$. Veja a figura (13):

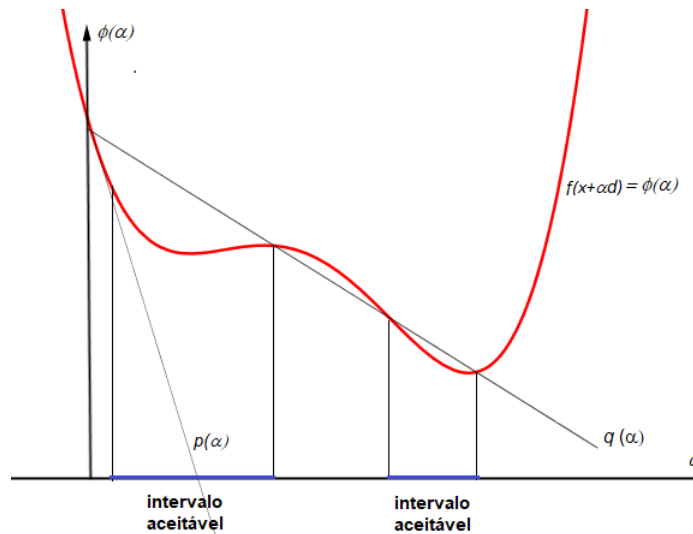


Figura 13: Interpretação gráfica do método

Por outro lado, temos que:

$$\phi(\alpha) - \phi(0) \leq \eta(p(\alpha) - p(0)),$$

implica que,

$$\phi(0) - \phi(\alpha) \leq \eta(p(0) - p(\alpha)).$$

Assim podemos concluir que o passo dado deve ser maior ou igual à fração η da redução obtida na aproximação linear.

O teorema a seguir garante que sempre encontraremos um α que satisfaça a Condição de Armijo. Para a prova deste resultado, foram necessários tópicos estudados em (GUIDO-RIZZI, 2011) e (RIBEIRO e KARAS, 2013).

Teorema 4.8. *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável, $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ uma direção de descida, com $\eta \in (0, 1)$. Então existe $\delta > 0$ tal que*

$$f(x + \alpha d) \leq f(x) + \eta \alpha \nabla f(x)^\top d \quad \forall \alpha \in [0, \delta].$$

Demonstração. Se $\nabla f(x)^\top d = 0$, o resultado segue da definição de direção de descida. Então suponhamos que $\nabla f(x)^\top d < 0$ e $\eta < 1$, temos:

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

Logo, existe $\delta > 0$ tal que

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

para todo $\alpha \in (0, \delta)$.

Portanto,

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

□

A seguir apresentamos um pseudocódigo do método de busca unidirecional usando a condição de Armijo, encontrado também em (RIBEIRO, 2013).

4.2.1 ALGORITMO DE DIREÇÃO DE DESCIDA COM A CONDIÇÃO DE ARMIJO

Dados: $x \in \mathbb{R}^n$, $d \in \mathbb{R}^n$ e $\gamma, \eta \in (0, 1)$

$\alpha = 1$

$k = 0$ REPITA enquanto $f(x + \alpha d) > f(x) + \eta \alpha \nabla f(x)^\top d$

$\alpha = \gamma \alpha$

FAÇA $k = k + 1$ Note que começamos com $\alpha = 1$, pois este é o maior tamanho de passo possível, podendo em alguns casos encontrar o minimizador já na primeira iteração, caso contrário, a cada iteração fazemos a sua redução até encontrar o α que satisfaça a desigualdade (13).

4.3 CONVERGÊNCIA GLOBAL DOS MÉTODOS DE DIREÇÃO DE DESCIDA

Para prosseguir no estudo da convergência global dos métodos, precisaremos de algumas definições e teoremas de Cálculo e Análise presentes no capítulo de revisão de conceitos. Neste trabalho, analisaremos apenas convergência global dos métodos.

Vejamos a definição a seguir, encontrada em (RIBEIRO, 2013), para podermos então estudar a convergência global dos métodos de descida.

Definição 4.9. Dizemos que um método é **Globalmente Convergente** quando para toda sequência (x^k) gerada e qualquer ponto de acumulação \bar{x} de (x^k) , temos que \bar{x} é um ponto estacionário.

Para analisar a convergência global do método de direção de descida vamos considerar uma direção de descida como $d = -H(x)\nabla f(x)$, onde $H : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ é uma função contínua que associa a cada $x \in \mathbb{R}^n$ uma matriz definida positiva $H \in \mathbb{R}^n$. Claramente,

$$\nabla f(x)^\top d = -\nabla f(x)^\top H(x)\nabla f(x) < 0.$$

Em seguida, vamos analisar o que ocorre com cada escolha para o tamanho do passo. O teorema a seguir, encontrado também em (RIBEIRO, 2013), garante a convergência global do método de direção de descida com Seção Áurea.

Teorema 4.10. *O método de direção de descida com o tamanho do passo calculado pelo método da Seção Áurea é globalmente convergente.*

Demonstração. Sejam (x^k) uma sequência gerada pelo algoritmo e \bar{x} um ponto de acumulação de (x^k) , além disso, $x^{k_n} \rightarrow \bar{x}$. Suponhamos por absurdo que \bar{x} não é um ponto estacionário, ou seja,

$$\nabla f(\bar{x}) \neq 0.$$

Por outro lado, $d = -H(\bar{x})\nabla f(\bar{x})$ é uma direção de descida, então existe $\bar{\alpha} > 0$ tal que

$$\beta = f(\bar{x}) - f(\bar{x} + \bar{\alpha}d) > 0.$$

Agora, consideremos $h(x) = f(x) - f(x - \bar{\alpha}H(x)\nabla f(x))$. Como h é contínua¹, temos que $h(x^k) \rightarrow h(\bar{x}) = \beta$, portanto:

$$f(x^k) - f(x^k + \bar{\alpha}d^k) = h(x^k) \geq \frac{\beta}{2}$$

para todo $k \in \mathbb{N}$ suficientemente grande. Como α_k foi obtido através de uma minimização local², temos que

$$f(x^{k+1}) = f(x^k + \alpha_k d^k) \leq f(x^k + \bar{\alpha}d^k) \leq f(x^k) - \frac{\beta}{2},$$

portanto,

$$f(x^k) - f(x^{k+1}) \geq \frac{\beta}{2} \tag{14}$$

para todo $k \in \mathbb{N}$ suficientemente grande. Em contrapartida, pela continuidade de f sabemos que $f(x^k) \rightarrow f(\bar{x})$ e por $(f(x^k))$ ser decrescente, $f(x^{k+1}) \rightarrow f(\bar{x})$. Assim,

$$f(\bar{x}) - f(\bar{x}) \geq \frac{\beta}{2}$$

¹ f é contínua, $H(x)$ é definida positiva e $H : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ é uma função contínua

² minimizar $\phi(\alpha) = f(x + \alpha d)$ s. a $\alpha > 0$

contradizendo $\beta > 0$. □

O próximo teorema garante a convergência global do método de direção de descida com a Condição de Armijo, também encontrado em (RIBEIRO, 2013).

Teorema 4.11. *O método de direção de descida que tem seu passo calculado pelo método da Condição de Armijo é globalmente convergente.*

Demonstração. Sejam (x^k) uma sequência gerada pelo método e \bar{x} um ponto de acumulação da sequência, além disso $x^{k_n} \rightarrow \bar{x}$. Suponhamos que \bar{x} não é ponto estacionário, ou seja, $\nabla f(\bar{x}) \neq 0$. Por outro lado, pela Condição de Armijo temos

$$f(x^{k+1}) = f(x^k + \alpha_k d^k) \leq f(x^k) + \eta \alpha_k \nabla f(x^k)^\top d^k.$$

Reescrevendo a partir da definição de d^k , obtemos

$$f(x^k) - f(x^{k+1}) \geq \eta \alpha_k \nabla f(x^k)^\top H(x^k) \nabla f(x^k) \geq 0.$$

Como f é contínua, sabemos que $f(x^k) \rightarrow f(\bar{x})$, assim como $f(x^{k+1}) \rightarrow f(\bar{x})$, portanto

$$\alpha_k \nabla f(x^k)^\top H(x^k) \nabla f(x^k) \rightarrow 0.$$

Mas $[(\nabla f(x^k)^\top H(x^k)) \rightarrow (\nabla f(\bar{x})^\top H(\bar{x})) \neq 0]^3$, resultando em $\alpha_k \rightarrow 0$. Logo, $\alpha_k < 1$ para todo $k \in \mathbb{N}$ suficientemente grande. Pelo método da Condição de Armijo, o passo $\frac{\alpha_k}{\gamma}$ existiu e foi recusado, assim

$$f(x^k + \alpha_k d^k) \leq f(x^k) + \eta \alpha_k \nabla f(x^k)^\top d^k \tag{15}$$

e

$$f\left(x^k + \frac{\alpha_k}{\gamma} d^k\right) \geq f(x^k) + \eta \alpha_k \nabla f(x^k)^\top d^k \tag{16}$$

sendo a inequação (15) em relação ao passo aceito e a inequação (16) em relação ao passo rejeitado.

Como a função $\xi(\alpha) = f(x^k + \alpha d^k) - f(x^k) - \eta \alpha \nabla f(x^k)^\top d^k$ é contínua, pelo teorema de Bolzano existe $s_k \in [\alpha_k, \frac{\alpha_k}{\gamma}]$ tal que $\xi(s_k) = 0$, ou seja,

$$f(x^k + \alpha d^k) - f(x^k) = \eta s_k \nabla f(x^k)^\top d^k.$$

Utilizando o teorema do valor médio em $\xi(\alpha)$,

$$\nabla f(x^k + \theta_k s_k d^k)^\top s_k d^k = f(x^k + s_k d^k) - f(x^k) = \eta s_k \nabla f(x^k)^\top d^k, \tag{17}$$

³O operador ∇f é uma função contínua e H é definida positiva. Supomos $\nabla f(\bar{x}) \geq 0$.

com $\theta \in (0, 1)$.

Como $s_k \rightarrow 0$, pois s_k é subsequência de α_k e como já vimos $\alpha_k \rightarrow 0$, podemos concluir de (17) que

$$\nabla f(\bar{x})^\top H(\bar{x}) \nabla f(\bar{x}) = \eta \nabla f(\bar{x})^\top H(\bar{x}) \nabla f(\bar{x})$$

o que é um absurdo. Portanto, \bar{x} é um ponto estacionário. \square

Após termos estudado o que é convergência global, vale ressaltarmos alguns fatos, listados a seguir.

Observação 4.12. *A escolha do ponto x inicial não interfere na convergência global de um método.*

Observação 4.13. *Um método ser globalmente convergente não garante que este método irá convergir para um minimizador global.*

Observação 4.14. *A convergência global depende da existência de pontos de acumulação.*

Observação 4.15. *Um ponto x^* ser estacionário não significa necessariamente que x^* é minimizador.*

Observação 4.16. *Um método ser globalmente convergente não implica que $x^k \rightarrow x^*$, onde x^* é minimizador.*

5 MÉTODOS CLÁSSICOS DE DESCIDA

Agora iremos estudar os métodos clássicos que possuem uma direção de descida bem definida, seguindo as referências de (NOCEDAL, 2006) e (RIBEIRO e KARAS, 2013). Não esquecendo que nosso objetivo é resolver o problema irrestrito

$$\begin{aligned} & \text{minimizar } f(x) \\ & \text{sujeito a } x \in \mathbb{R}^n \end{aligned} \tag{19}$$

5.1 MÉTODO GRADIENTE

No Método Gradiente definimos a direção de descida como $d = -\nabla f(x)$. Vamos analisar agora porque este método é eficiente.

Primeiramente, consideramos o vetor $v \in \mathbb{R}^n$ tal que $\|v\| = \|d\|$. Analisando a derivada direcional, temos

$$\frac{\partial f}{\partial d}(x) = \langle \nabla f(x), d \rangle = -\|\nabla f(x)\|^2,$$

agora em relação a v ,

$$-\|\nabla f(x)\|^2 = -\|\nabla f(x)\| \|d\| = -\|\nabla f(x)\| \|v\|.$$

Usando a desigualdade de Schwarz temos

$$-\|\nabla f(x)\| \|v\| \leq \langle \nabla f(x), v \rangle = \frac{\partial f}{\partial v}(x),$$

assim obtemos:

$$\frac{\partial f}{\partial d}(x) \leq \frac{\partial f}{\partial v}(x),$$

que nos garante que a direção oposta ao vetor gradiente é a que produz maior decrescimento na função f que todas as demais direções.

Uma das propriedades deste método é que duas iterações consecutivas são sempre ortogonais. Esta afirmação é formalizada no lema a seguir.

Proposição 5.1. *Seja o método de direção de descida com d^k escolhido de acordo com o método Gradiente. Se α_k é obtido por uma minimização local $f(x^k + \alpha_k d^k)$, então $d^{k+1 \top} d^k = 0$.*

Demonstração. Consideremos a função $\phi : \mathbb{R} \rightarrow \mathbb{R}$ como $\phi(\alpha) = f(x + \alpha d)$, temos:

$$\phi'(\alpha_k) = d^{k \top} \nabla f(x^k + \alpha_k d^k) = \nabla f(x^{k+1})^\top d^k.$$

Assim, como a busca é feita por uma minimização local, obtemos a seguinte igualdade:

$$d^{k+1 \top} d^k = -\nabla f(x^{k+1})^\top d^k = -\phi'(\alpha_k) = 0.$$

Portanto, as direções d^k e d^{k+1} são ortogonais entre si, para todo $k \in \mathbb{N}$. □

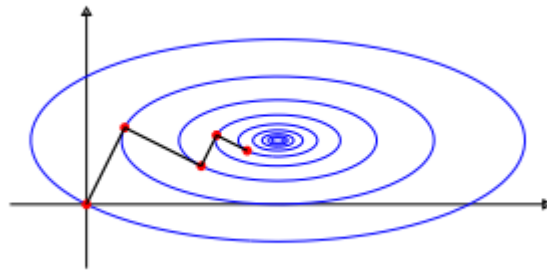


Figura 15: Ortogonalidade entre as iterações

Apresentamos a seguir um pseudocódigo do Método Gradiente.

5.1.1 ALGORITMO DO MÉTODO GRADIENTE

Dado: $x^k \in \mathbb{R}^n$

$k = 0$

REPITA enquanto $\nabla f(x^k) \neq 0$

Defina $d^k = -\nabla f(x^k)$

Obter $\alpha_k > 0$ usando Seção Áurea ou Armijo FAÇA $k = k + 1$

5.1.2 CONVERGÊNCIA GLOBAL DO MÉTODO

A convergência global deste método recorre diretamente dos teoremas 4.9 e 4.10 estudados no capítulo anterior e pode ser encontrada também em (RIBEIRO, 2013), veja:

Teorema 5.2. *O método de direção de descida com d estabelecido pelo Método Gradiente com o tamanho do passo α_k calculado pelo método da Seção Áurea é globalmente convergente. De forma análoga, o método de direção de descida com d estabelecido pelo Método Gradiente com o tamanho do passo α_k calculado pelo método da Condição de Armijo também é globalmente convergente.*

Demonstração. A demonstração segue diretamente dos teoremas 4.9 e 4.10 do capítulo anterior. Em particular, para este caso consideramos $H(x) = I_n(x)$. \square

A seguir estudaremos outro método clássico de direção de descida, muito utilizado em problemas de otimização.

5.2 MÉTODO DE NEWTON

Primeiramente, vamos estudar como funciona este método. Nosso objetivo é minimizar a função f dada por $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^2 , ou seja, encontrar $\nabla f(x) = 0$ para algum $x \in D_f$, sendo D_f o conjunto domínio da função f . Em outras palavras, precisamos resolver o sistema $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ de classe C^1 tal que

$$F(x) = 0.$$

Geralmente não conseguimos resolvê-lo diretamente, por isso o uso de processos de iteração pode ser a forma mais eficiente de resolver este problema.

Vamos aproximar a função F pelo polinômio de Taylor de primeira ordem. Considerando uma estimativa x_0 e a matriz Jacobiana J_F da função f , temos o sistema

$$F(x_0) + J_F(x_0)(x - x_0) = 0. \quad (20)$$

Note que se a matriz Jacobiana J_F for inversível, então o sistema (20) pode ser reescrito da seguinte forma:

$$\bar{x} = x_0 - (J_F(x_0))^{-1}F(x_0)$$

que é exatamente uma iteração do método Newton. Como nosso objetivo é minimizar f , reescrevemos o método acima para $F = \nabla f$ e $J_F = \nabla^2 f$, então, $\bar{x} = x_0 - (\nabla^2 f(x_0))^{-1}\nabla f(x_0)$, veja a figura:

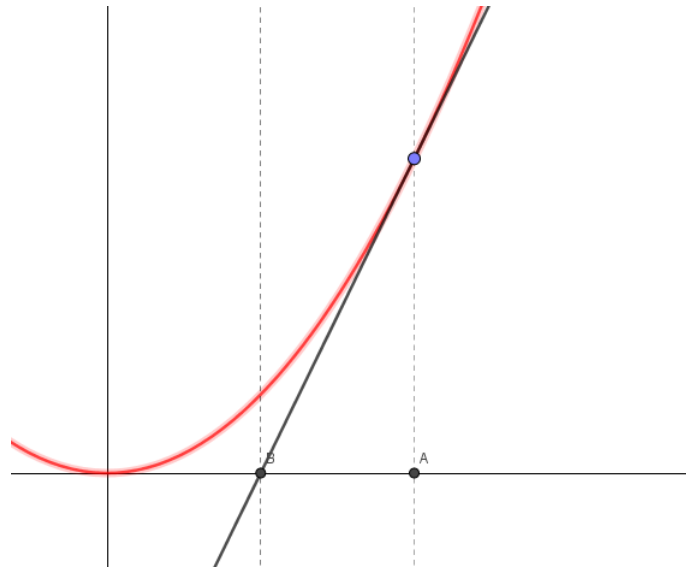


Figura 16: iteração do método no caso $f : \mathbb{R} \rightarrow \mathbb{R}$

Na figura 16 está representada uma iteração do Método de Newton para a função $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = x^2$. Note que quando aproximamos f pelo polinômio de Taylor de primeira ordem, obtemos uma redução no valor de f . Veja a seguir um pseudocódigo do Método de Newton.

5.2.1 ALGORITMO DO MÉTODO DE NEWTON

Dado: $x^k \in \mathbb{R}^n$

$k = 0$

REPITA enquanto $\nabla^2 f(x^k) \neq 0$

Defina $d^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k)$

Determinar o tamanho do passo $\alpha_k > 0$ usando Seção Áurea ou Condição de Armijo

FAÇA $k = k + 1$

Observe que neste método, a direção não está bem definida se a matriz $\nabla^2 f(x^k)$ for singular. Mas se $\nabla^2 f(x^k)$ for definida positiva, temos a garantia de que d^k está bem definida e é uma direção de descida.

Agora, vamos estudar uma forma diferente de obter a direção d^k , também originando o Método de Newton. Primeiramente, vamos considerar a aproximação de Taylor de segunda ordem de f :

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

Para minimizar p temos

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

o que nos fornece a direção d^k do algoritmo 5.2.1.

Notemos que o Método de Newton minimiza o modelo quadrático de f dado por

$$x^\top Ax + bx + c,$$

em torno de x^k usando apenas uma iteração. Vejamos o porque:

De fato, $\nabla f(x) = Ax + b$ e $\nabla^2 f(x) = A$. Reescrevendo a direção do Método de Newton:

$$d^0 = -[\nabla^2 f(x^0)]^{-1} \nabla f(x^0) = -A^{-1}(Ax^0 + b)$$

$$d^0 = -x^0 - A^{-1}b$$

$$x^1 - x^0 + x^0 = -A^{-1}b$$

$$x^1 = -A^{-1}b$$

Tal vantagem reduz consideravelmente o gasto computacional deste método quando se trata de funções quadráticas.

5.2.2 CONVERGÊNCIA GLOBAL DO MÉTODO

A convergência global do Método de Newton recorre diretamente dos teoremas 4.9 e 4.10 estudados no capítulo anterior e pode ser encontrada também em (RIBEIRO e KARAS, 2013), veja:

Teorema 5.3. *O algoritmo de descida com d estabelecido pelo Método de Newton com o tamanho do passo α calculado pelo método da Seção Áurea é globalmente convergente. De forma análoga, o algoritmo de descida com d estabelecida pelo Método de Newton com o tamanho do passo α calculado pelo Método da Condição de Armijo também é globalmente convergente.*

Demonstração. A demonstração segue diretamente dos teoremas 4.9 e 4.10 do capítulo anterior. Em particular, para este caso consideramos $H(x) = \nabla^2 f(x)$. □

5.3 EXEMPLOS PRÁTICOS

Nesta seção traremos alguns exemplos de funções que buscamos minimizar utilizando o Método Gradiente e o Método de Newton, utilizando a Condição de Armijo. Vamos discutir se é possível realizar as iterações e se a sequência gerada pelos métodos nos conduzem a minimizadores. Todos os resultados e gráficos foram obtidos a partir da execução dos métodos no software MatLab. Para estes exemplos, consideramos como critério de parada $\|\nabla f(x)\| < 10^{-6}$.

Exemplo 5.4. *Considere a função*

$$f(x_1, x_2) = 1 + (x_2)^2$$

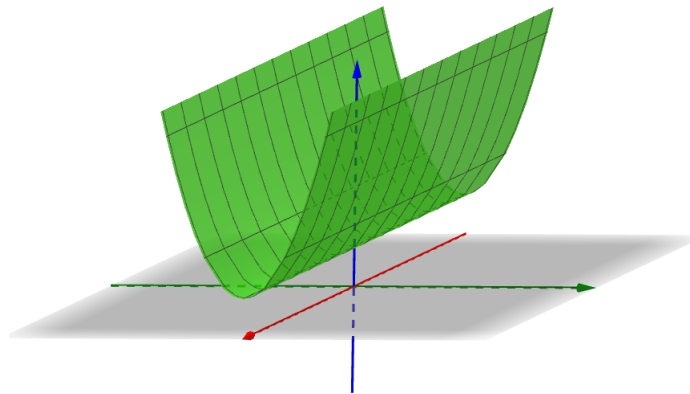


Figura 17: Imagem produzida pelo GeoGebra

- Notemos que $\nabla^2 f(x) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$ não é definida positiva, neste caso as iterações d^k para o Método de Newton não estão bem definidas.
- Os gráficos a seguir trazem a variação de $\|\nabla f(x)\|$ e a variação de $f(x)$ ao longo de cada iteração.
- Escolhemos como ponto inicial $x = (1, 1)$
- Para este exemplo, como o Método de Newton não está bem definido, os gráficos trazem somente a variação de $\|\nabla f(x)\|$ e de $f(x)$ para o método Gradiente.

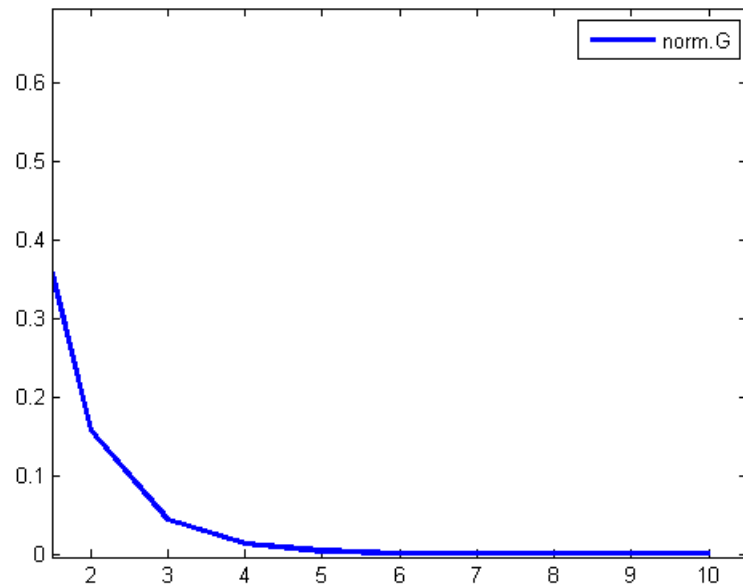


Figura 18: Variação da $\|\nabla f(x)\|$ para o Método Gradiente

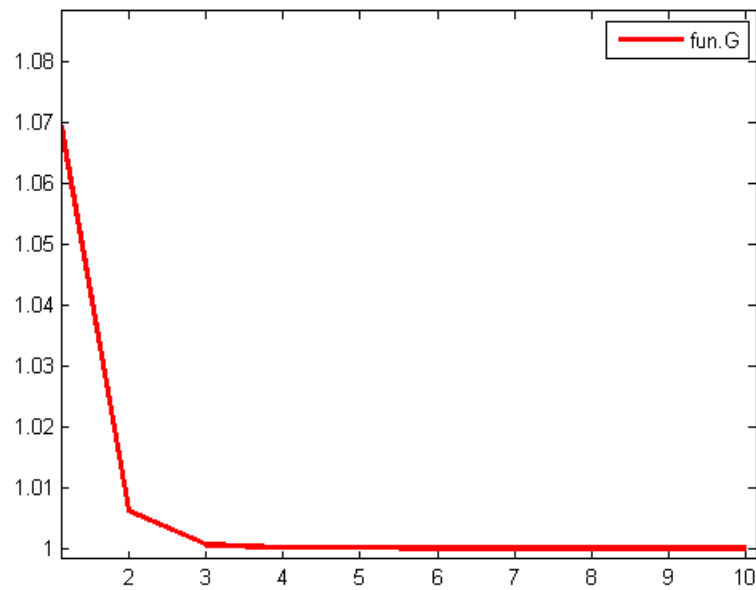


Figura 19: Variação da função para o Método Gradiente

- Podemos notar que utilizando o Método Gradiente, com apenas 10 iterações encontramos um minimizador de f , $x^* = (0,0)$.

Exemplo 5.5. Considere a função

$$f(x_1, x_2) = 2(x_1)^3 - 3(x_1)^2 - 6(x_1)(x_2)(x_1 - x_2 - 1)$$

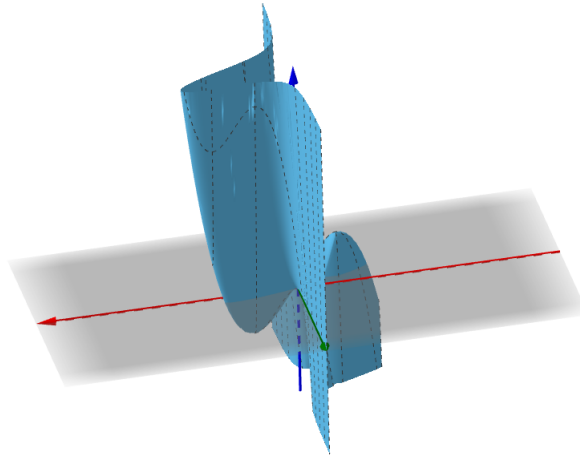


Figura 20: Imagem produzida pelo GeoGebra

- Pelo gráfico da função podemos notar que f não é limitada inferiormente. Embora a sequência $(f(x^k))$ seja decrescente, não é limitada, logo não há garantia da existência de ponto de acumulação.
- A matriz Hessiana desta função é da forma:

$$\nabla^2 f(x) = \begin{bmatrix} 12x_1 - 6 - 12x_2 & -12x_1 + 12x_2 + 6 \\ -12x_1 + 12x_2 + 6 & 12x_1 \end{bmatrix},$$

então pode haver pontos x^k tal que $\nabla^2 f(x^k)$ não seja definida positiva. Portanto, a direção de descida d^k para o Método de Newton não está bem definida para todos os pontos x^k

- Da mesma forma que o exemplo 5.7, as figuras (21) e (22) trazem os gráficos da variação de $\|\nabla f(x)\|$ e a variação de $f(x)$ ao longo de cada iteração para o método de Newton e Método Gradiente.
- Para esta função os dois métodos atingiram o número máximo de iterações $k = 1000$, não encontrando o minimizador global de f .
- Após a 6ª iteração, a sequência $(f(x^k))$ gerada pelo método Gradiente decresce rapidamente, pois o ponto x^k da próxima iteração se encontra na região onde a função é ilimitada inferiormente.

- Embora a $\|\nabla f(x)\|$, para o Método de Newton, pareça estar em zero, não está. Pois se estivesse, então o método teria encontrado o minimizador de f . Ocorre esta ilusão devido a grande escala do gráfico.

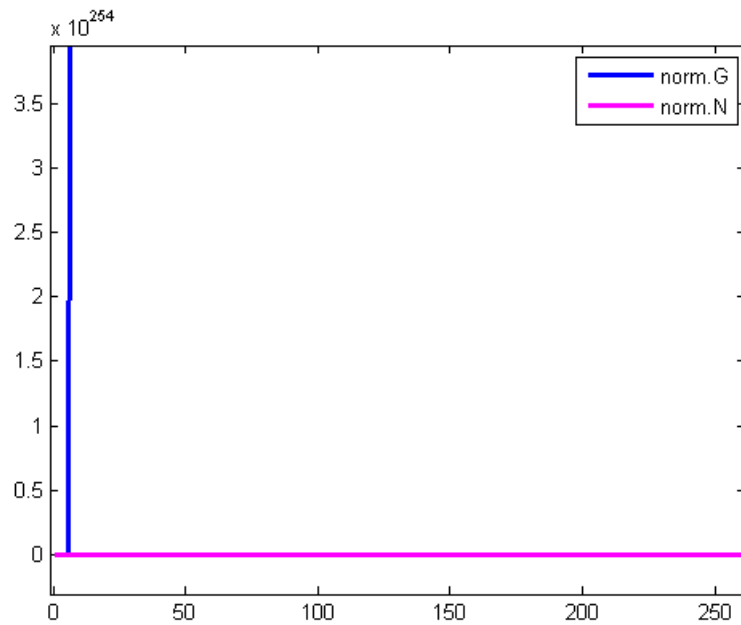


Figura 21: Variação da $\|\nabla f(x)\|$ para Gradiente (norm.G) e para Newton (norm.N)

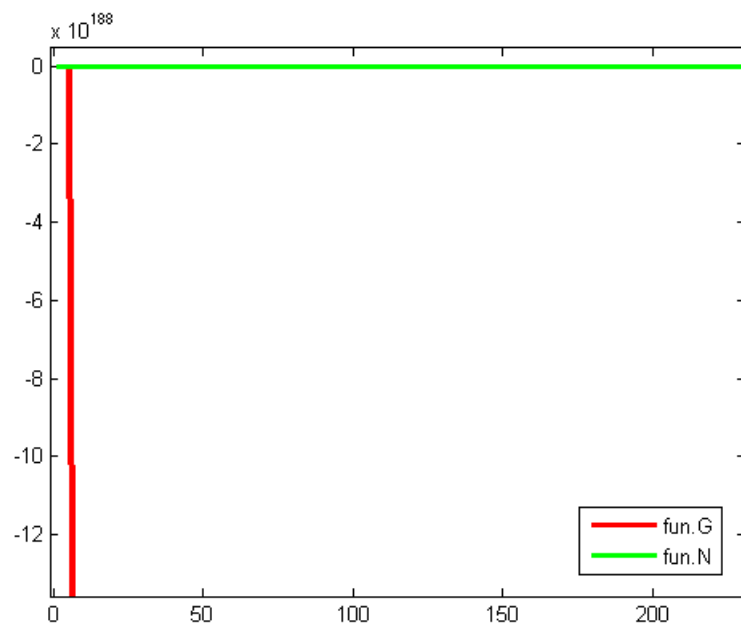


Figura 22: Variação da função para Gradiente (fun.G) e para Newton (fun.N)

Exemplo 5.6. Considere a função

$$f(x) = (x_1^2 - x_2^2)e^{[-(x_1^2 - x_2^2)/2]}$$

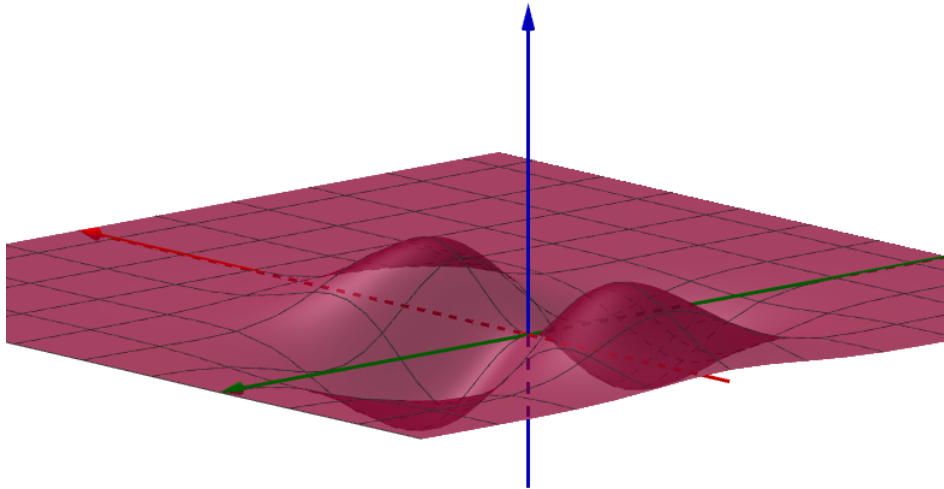


Figura 23: Imagem produzida pelo GeoGebra

- Esta função é limitada inferiormente, e possui pontos de mínimo, máximo e sela.
- Os gráficos a seguir trazem a variação da função f e a variação da $\|\nabla f(x)\|$ para o Método Gradiente e para o Método de Newton.
- Quando tomamos o ponto inicial¹ $x = (-2, 2)$, o Método Gradiente e o Método de Newton atingem o número máximo de iterações $k = 1000$.
- Embora no gráfico que traz a variação da norma do gradiente pareça $\|\nabla f(x)\|$ esteja em zero, não está. Pois se estivesse, então os métodos teriam encontrado um ponto estacionário de f .
- Nesta função, dependendo da escolha para o x inicial, os métodos podem encontrar um ponto de sela, de máximo ou mínimo. Pois a convergência global está relacionada apenas a pontos estacionários e não a pontos de mínimo.

¹O ponto $x = (-2, 2)$ está longe do ponto de mínimo de f .

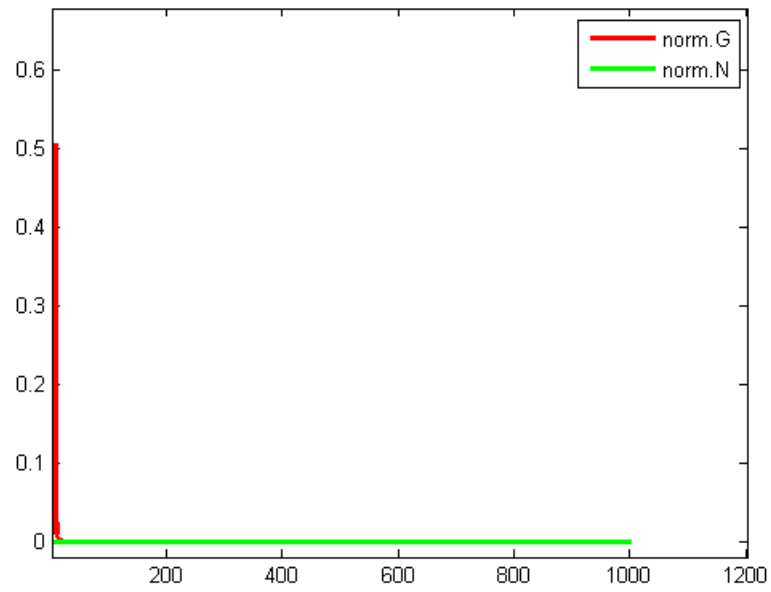


Figura 24: Variação da $\|\nabla f(x)\|$ para Gradiente (norm.G) e para Newton (norm.N)

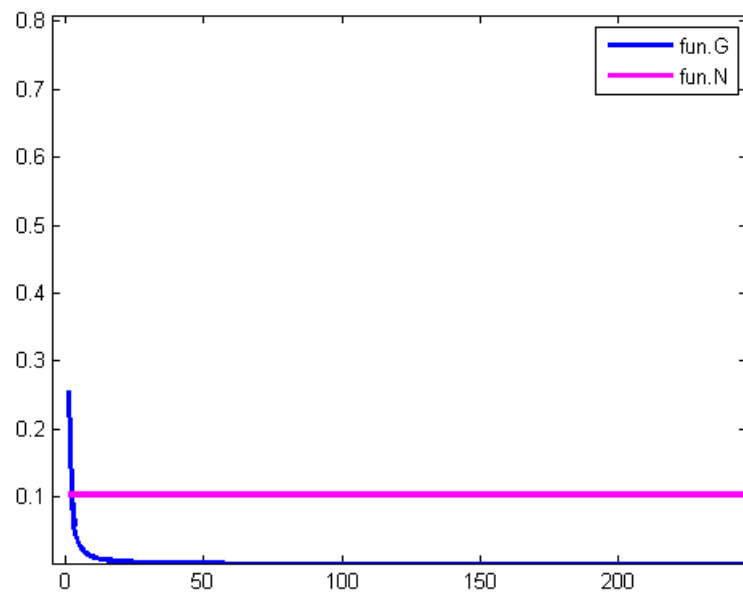


Figura 25: Variação da função para Gradiente (fun.G) e para Newton (fun.N)

- Se escolhermos o ponto inicial como $x = (-1, 1)$, o Método Gradiente encontra um minimizador em 14 iterações. Já o Método de Newton alcança o número máximo de iterações $k = 1000$.

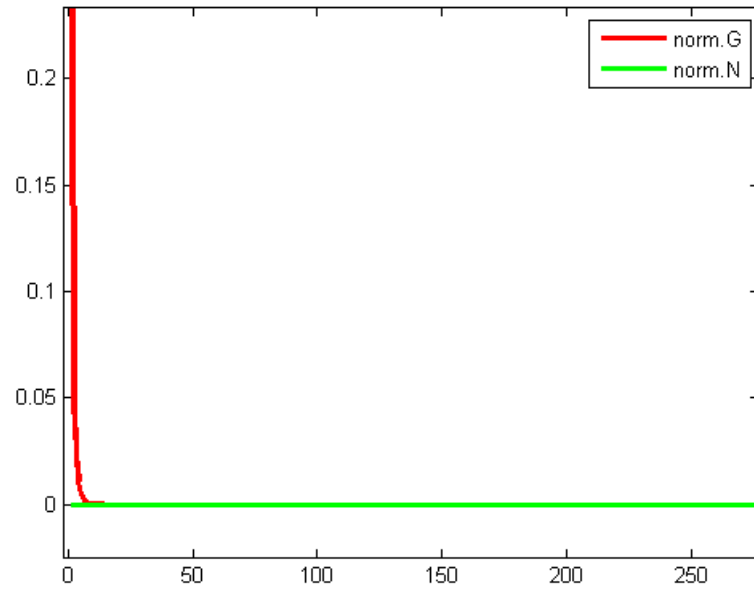


Figura 26: Variação da $\|\nabla f(x)\|$ para Gradiente (norm.G) e para Newton (norm.N)

Os gráficos a seguir trazem separadamente a variação do valor de f , para melhor visualização.

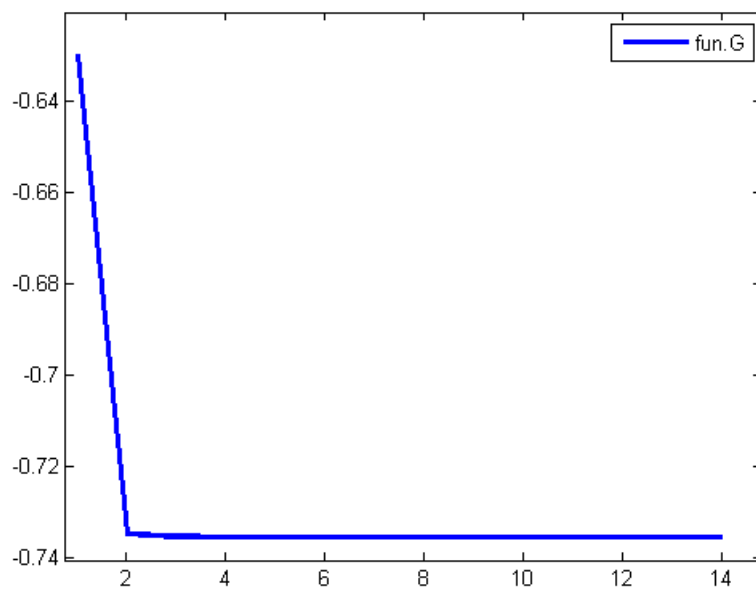


Figura 27: Variação da função para Gradiente (fun.G)

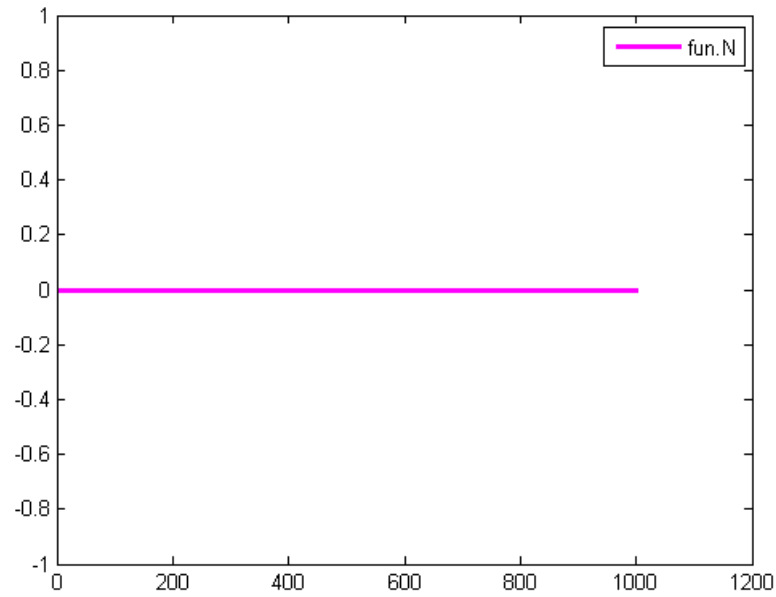


Figura 28: Variação da função para Newton (fun.N)

- *Embora pareça que o valor de f seja igual a zero quando aplicado o Método de Newton, isto não acontece. Se acontecesse isto, para esta função, então o Método de Newton teria encontrado o ponto de sela, que é um ponto estacionário. Esta falsa intuição ocorre devido a grande escala do gráfico.*

- Finalmente, se escolhermos o ponto inicial² como $x = (2,0)$, o Método de Newton encontra um ponto estacionário³ em 1 iteração, sendo este ponto $\bar{x} = (0,0)$, e temos $f(\bar{x}) = 0$. Já o Método Gradiente alcança o número máximo de iterações $k = 1000$.

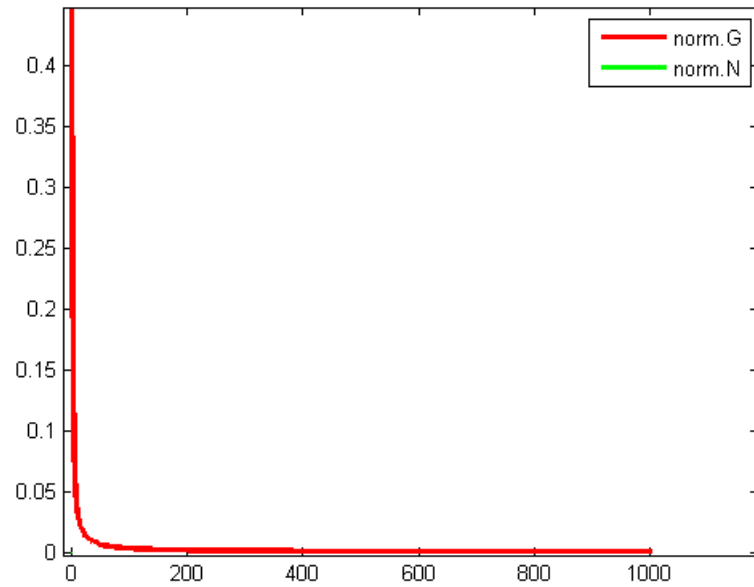


Figura 29: Variação da $\|\nabla f(x)\|$ para Gradiente (norm.G) e para Newton (norm.N)

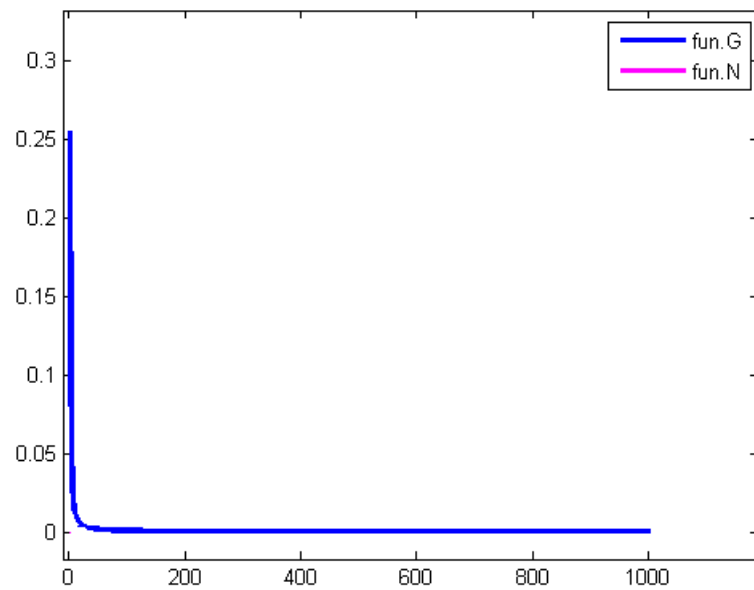


Figura 30: Variação da função para Gradiente (fun.G) e para Newton (fun.N)

²O ponto $x = (2,0)$ está próximo de um ponto de máximo.

³O ponto estacionário encontrado é o ponto de sela $x = (0,0)$

- Podemos notar que não está representado no gráfico a variação de f e a variação de $\|\nabla f(x)\|$ para o Método de Newton, pois o método resolveu o problema em apenas uma iteração. Assim, devido a grande escala do gráfico, não há a representação da variação para apenas uma iteração.
- Notemos que o minimizador $x^* = (0; 1,41)$ encontrado através do Método de Newton satisfaz as condições de otimalidade, pois é o mesmo ponto minimizador encontrado anteriormente, tomando o ponto inicial $x = (-1, 1)$.

Exemplo 5.7. Considere a função $f(x) = x_1^2 + 4x_1x_2 + 6x_2^2$

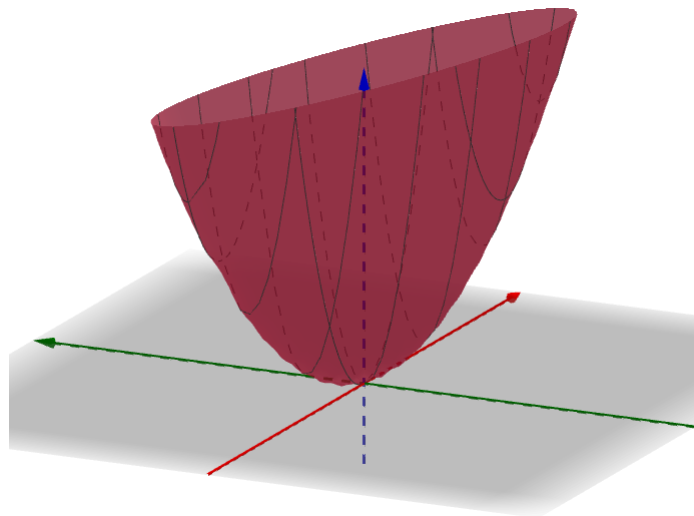


Figura 31: Imagem gerada pelo GeoGebra

- Esta função é quadrática, e além disso limitada inferiormente possuindo mínimo global.
- O Método Gradiente encontrou o minimizador x^* em 128 iterações.
- O Método de Newton encontrou o minimizador x^* em 1 iteração. Pois para funções quadráticas, o Método de Newton encontra x^* em uma iteração, como estudamos na seção 5.2.

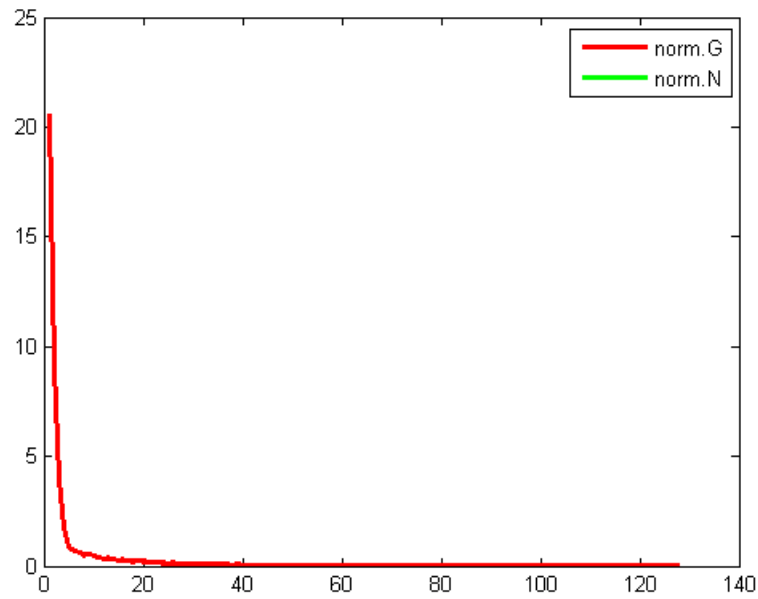


Figura 32: Variação da $\|\nabla f(x)\|$ para Gradiente (norm.G) e para Newton (norm.N)

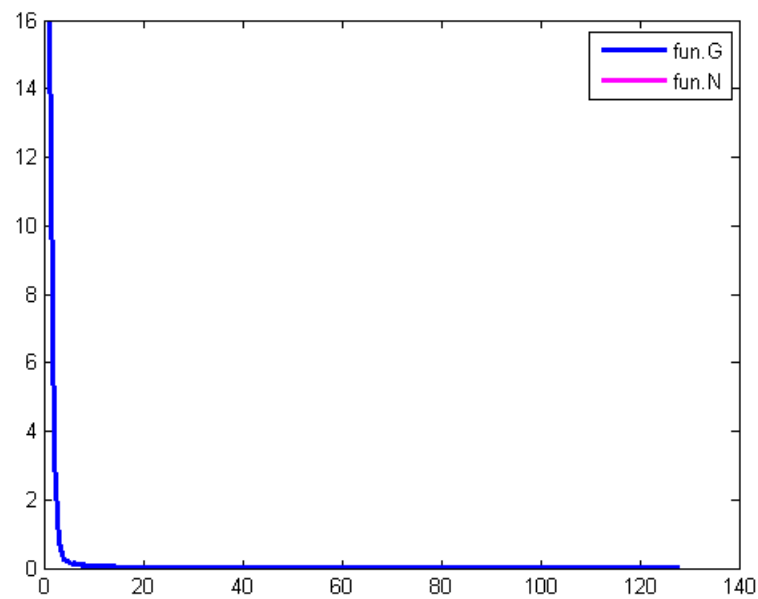


Figura 33: Variação da função para Gradiente (fun.G) e para Newton (fun.N)

- Novamente, no gráfico não está representada a variação para uma iteração. Desta forma, estão representadas somente as variações para o Método Gradiente.

COMPARAÇÕES

- O exemplo 5.8 enfatiza a principal condição para a convergência global dos Métodos de direção de descida. Só temos garantia que estes métodos de fato encontram o ponto \bar{x} estacionário, se a função objetivo f for limitada inferiormente. Em outras palavras, se a função não for limitada inferiormente, não há a garantia que a sequência $(f(x^k))$ possua ponto de acumulação. Assim, não sabemos nada a respeito da convergência global dos métodos aplicados a este problema.
- Nos interessamos em explorar neste trabalho exemplos problemáticos, mostrando que as propriedades dos métodos, como por exemplo, $\nabla^2 f(x)$ ser definida positiva para o Método de Newton, e a função objetivo ser limitada inferiormente para garantir a convergência global dos métodos, são fundamentais para que os métodos de fato encontrem um ponto estacionário \bar{x} da função f . Desta forma, precisamos ter garantia de que o problema a ser resolvido satisfaz a fundamentação teórica do método a ser aplicado, para então aplicar este método.
- O exemplo 5.9 mostra que a eficiência de um método e o número de iterações necessário depende do ponto inicial x escolhido. De fato, se escolhermos o ponto inicial $x = (2, 0)$, o Método de Newton encontra um ponto estacionário. Por outro lado, se escolhermos o ponto inicial $x = (-1, 1)$, o Método Gradiente encontra um minimizador x^* .

No próximo capítulo estudaremos uma classe de problemas que podem ser resolvidos usando técnicas de otimização irrestrita. Chamamos estes problemas de Estimção de Parâmetros.

6 O PROBLEMA DE ESTIMAÇÃO DE PARÂMETROS

Neste capítulo estudaremos como um problema de estimar parâmetros pode ser reduzido a um problema de otimização irrestrita. As principais referências utilizadas foram (NOCEDAL, 1999), (BENATTI, 2017) e (CROEZE, 2012).

Primeiramente, considere o seguinte conjunto de dados referentes a população dos Estados Unidos (em milhões) entre os anos de 1815 a 1885.

Tabela 9:

Tempo (t)	População (y)	Ano
1	8.3	1815
2	11.0	1825
3	14.7	1835
4	19.7	1845
5	26.7	1855
6	35.2	1865
7	44.4	1875
8	55.9	1885

Tabela 10: Retirado de (CROEZE, 2012).

Segundo o artigo (CROEZE, 2012), este conjunto de dados pode ser aproximado por uma função $\phi(t) = x_1 e^{x_2 t}$ onde t é uma variável independente e x são os parâmetros que queremos estimar. Convenientemente atribuímos 1815 a $t = 1$, 1825 a $t = 2$ e assim sucessivamente.

Na figura (35) representamos $\phi(t, y)$. O gráfico em preto representa $\phi(t, y) = 7e^{0,2t}$, e o gráfico em vermelho representa $\phi(t, y) = 6e^{0,3t}$. O eixo horizontal representa o valor de t e o eixo vertical representa o valor de y referentes aos dados reais.

e mudarmos o valor de x , a função que modela o conjunto de dados pode se aproximar ou se afastar dos valores reais do conjunto de dados em questão. Assim, formulamos um novo problema para aproximar a função $\phi(t_i)$ o máximo possível dos valores reais do conjunto de dados.

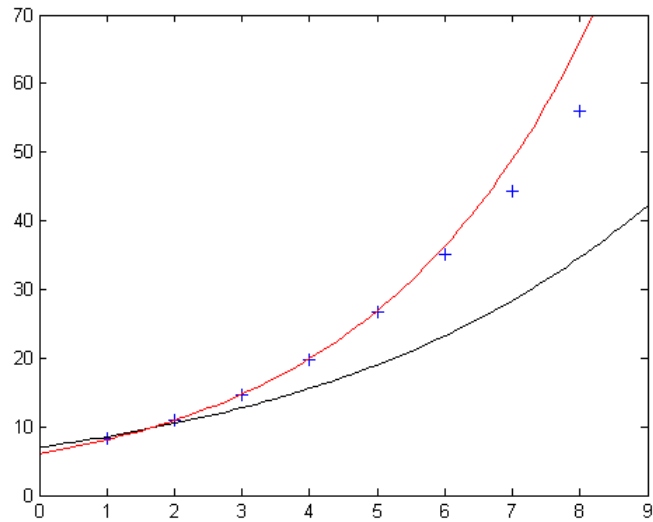


Figura 35: $\phi(t)$

Para formular este problema, consideramos primeiramente a função

$$F_i(x) = \phi(t_i) - y_i,$$

onde $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, em que cada ponto (t_i, y_i) representa o valor real de um ponto do conjunto de dados, com $i = 1, 2, \dots, m$.

Esta função $F_i(x)$ representa a distância de cada ponto (t_i, y_i) do conjunto de dados até a sua aproximação feita pela função $\phi(x, t_i)$. Assim, formulamos nosso problema como: minimizar soma de todas as distâncias de cada ponto (t_i, y_i) até a função modelo $\phi(x, t_i)$.

O problema de minimizar a soma das distâncias está descrito na seção a seguir, através da técnica dos Mínimos Quadrados.

6.1 MÉTODO DOS MÍNIMOS QUADRADOS

Como vimos anteriormente, nosso objetivo agora é minimizar as distâncias de cada ponto até o seu valor real. Assim, sem perda de generalidade, podemos minimizar o quadrado destas distâncias. Em outras palavras, podemos reescrever nosso problema como:

$$\begin{aligned} \text{minimizar } f(x) &= \frac{1}{2} \|F(x)\|^2 \\ \text{sujeito a } x &\in \mathbb{R}^n \end{aligned} \tag{22}$$

onde $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Como se trata de um problema de otimização, precisamos garantir que a função f seja contínua e diferenciável, por isso temos a norma euclidiana ao quadrado. Assim,

podemos reescrever o problema (22) como:

$$\text{minimizar } f(x) = \frac{1}{2} \left[\sum_{i=1}^m F_i(x)^2 \right]. \quad (23)$$

Sabemos que as condições de otimalidade envolvem o estudo do $\nabla f(x)$ e também da $\nabla^2 f(x)$. Assim, nos interessa analisar como será o $\nabla f(x)$ e a $\nabla^2 f(x)$ da função f formalizada no problema (22).

Primeiramente, para encontrarmos $\nabla f(x)$ devemos encontrar o vetor gradiente da função $F(x)$, que é dado pela Jacobiana:

$$J = \begin{bmatrix} \nabla F_1(x)^\top \\ \nabla F_2(x)^\top \\ \vdots \\ \nabla F_m(x)^\top \end{bmatrix}$$

Assim, pela regra da cadeia aplicada em $f(x)$, temos os seguintes resultados:

$$\nabla f(x) = \sum_{i=1}^m F_i(x) \nabla F_i(x) = J(x)^\top F(x) \quad (24)$$

$$\begin{aligned} \nabla^2 f(x) &= \sum_{i=1}^m \nabla F_i(x) \nabla F_i(x)^\top + \sum_{i=1}^m F_i(x) \nabla^2 F_i(x) \\ &= J(x)^\top J(x) + \sum_{i=1}^m F_i(x) \nabla^2 F_i(x) \end{aligned} \quad (25)$$

Estes resultados serão importantes para estudarmos nas próximas seções dois métodos iterativos para resolver o problema (23).

6.2 MÉTODO GRADIENTE

Para resolver um problema de Mínimos Quadrados podemos aplicar também o Método Gradiente já estudado no capítulo anterior. De maneira análoga, vamos tomar a direção de descida como $d = -\nabla f(x)$. Entretanto, para o problema de Mínimos Quadrados temos que

$$\nabla f(x) = \sum_{i=1}^m F_i(x) \nabla F_i(x) = J(x)^\top F(x)$$

Todos os resultados referentes a este método, estudados no capítulo 5, são válidos também para o caso onde:

$$f(x) = \frac{1}{2} \|F(x)\|^2.$$

6.3 MÉTODO DE GAUSS-NEWTON

O Método de Gauss-Newton é baseado no Método de Newton já estudado no capítulo 5, lembrando que a direção de descida obtida através deste método é dada por:

$$d^k = -(\nabla^2 f(x^k))^{-1} \nabla f(x^k).$$

Agora nosso objetivo é resolver o problema em particular dado por:

$$\text{minimizar } f(x) = \frac{1}{2} \|F(x)\|^2,$$

onde a Hessiana é da forma dada na equação (27). O Método de Gauss-Newton consiste em aproximar a matriz Hessiana por:

$$\nabla^2 f(x) \approx J(x)^\top J(x).$$

Desta forma, a direção de Gauss-Newton é dada por:

$$d^k = -(J(x^k)^\top J(x^k))^{-1} J(x^k)^\top F(x^k),$$

onde $J(x^k)^\top F(x^k) = \nabla f(x^k)$.

Para que o método esteja bem definido, $J(x^k)^\top J(x^k)$ deve ser definida positiva, mas nem sempre temos esta garantia.

Apresentamos a seguir o pseudocódigo do Método de Gauss-Newton.

6.3.1 ALGORITMO DO MÉTODO DE GAUSS-NEWTON

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

$k = 0$

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

defina $d^k = (-J(x^k)^\top J(x^k))^{-1} J(x^k)^\top F(x^k)$

obtenha $\alpha_k > 0$

faça $k = k + 1$

6.3.2 CONVERGÊNCIA GLOBAL DO MÉTODO

Não há garantia de convergência global para o Método de Gauss-Newton, de acordo com (AGUIAR, 2012). Sob certas condições é possível garantir a convergência local deste

método, mas este tópico não será estudado neste trabalho.

No artigo (CROEZE, 2012), os autores comentam que o Método de Gauss-Newton converge rapidamente quando o ponto inicial está próximo da solução.

Para o caso onde $J^T J$ não for definida positiva, temos outro método chamado Levenberg-Marquardt, que está relacionado com o Método da Região de Confiança, tópicos não abordados neste trabalho, que podem ser vistos em (NOCEDAL, 2006) e (RIBEIRO e KARAS, 2013).

Podemos encontrar os tópicos citados aqui mas não abordados, nas seguintes referências: (NOCEDAL, 2006), (CROEZE, 2012) e (BENATTI, 2017).

7 EXEMPLOS PRÁTICOS

Os dois exemplos a seguir foram inspirados nos modelos trazidos no artigo de (CROEZE, 2012). Foi realizada a implementação dos métodos utilizando o software MatLab, e reproduzido todos os resultados trazidos nos modelos de (CROEZE, 2012). De fato, os resultados obtidos conferem com os trazidos no artigo. Aqui novamente utilizamos como critério de parada $\|\nabla f(x)\| < 10^{-6}$ e $k = 500$.

Exemplo 7.1. *Consideremos o conjunto de dados exposto no início desta seção, que representa a população dos Estados Unidos (em milhões), entre os anos de 1815 a 1885:*

Tabela 12:

Tempo (t)	População	Ano
1	8.3	1815
2	11.0	1825
3	14.7	1835
4	19.7	1845
5	26.7	1855
6	35.2	1865
7	44.4	1875
8	55.9	1885

Cada ponto é da forma (t_i, y_i) onde atribuímos o ano 1815 a $t = 1$ e assim sucessivamente para todos os pontos do conjunto de dados, e y_i representa a população (em milhões) para cada ano t_i .

- **Estimativa 1**

Como já vimos, estes dados podem ser modelados pela função $\phi(x, t) = x_1 e^{x_2 t}$. Tomaremos como estimativa inicial $(x) = (6; 0, 3)$. Assim, temos $\phi(t, y) = 6e^{0,3t}$ como primeira estimativa.

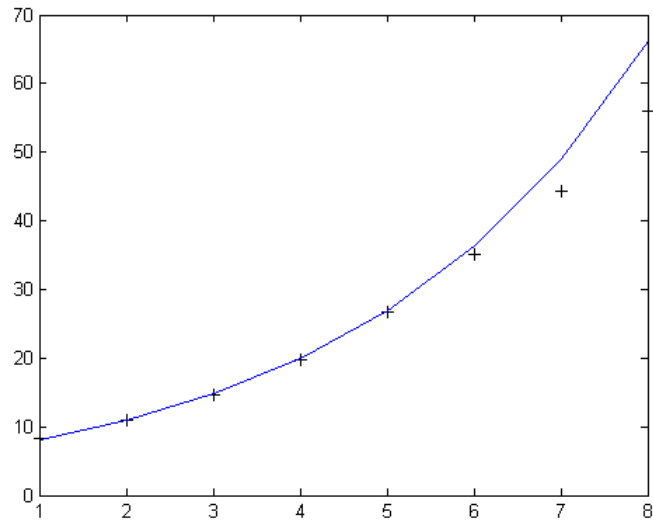


Figura 37: $\phi(t) = 6e^{0,3t}$

• **Estimativa 1 com a Condição de Armijo**

O gráfico a seguir traz a nova aproximação $\phi(t, y)$ a partir do parâmetro ótimo x determinado pelo Método Gradiente e pelo Método de Gauss-Newton simultaneamente, tomando como x inicial $x = (6; 0, 3)$ e tomando α pela Condição de Armijo.

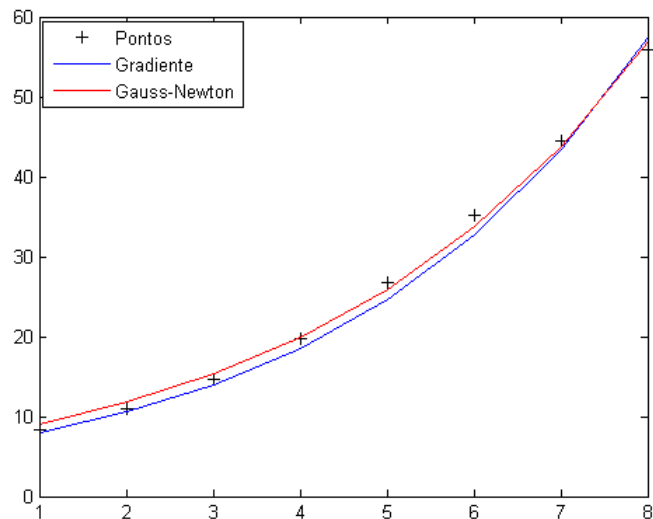


Figura 38: Aproximação a partir do x ótimo

Podemos notar que, para este problema, o Método de Gauss-Newton encontrou o parâmetro x que melhor aproxima a função $\phi(t, y)$ do conjunto de dados inicial. Embora o Método Gradiente também tenha aproximado, o resíduo $F_i(x) = \phi(x, y_i) - y_i$ para

$i = 1, \dots, 8$ resultante após aplicar este método, ainda é maior que o resíduo resultante do Método de Gauss-Newton.

- **Estimativa 1 com passo constante $\alpha = 1$**

Agora, aplicaremos o Método Gradiente e o Método de Gauss-Newton simultaneamente, tomando o tamanho do passo $\alpha = 1$ para a estimativa inicial $x = (6; 0, 3)$.

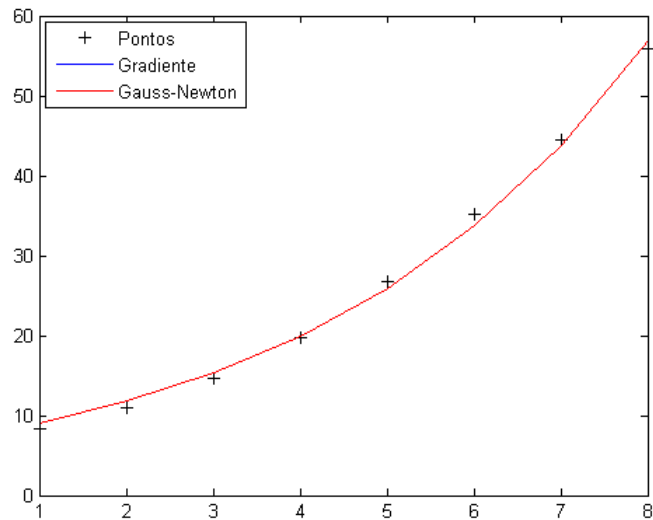


Figura 39: Aproximação a partir do x ótimo

Podemos notar que o Método Gradiente aplicado com o tamanho do passo $\alpha = 1$ não é eficiente para este problema, pois não resultou em uma aproximação. Entretanto o Método de Gauss-Newton foi eficiente, resultando em uma boa aproximação do modelo $\phi(t, y)$ em relação ao conjunto de dados.

- **Estimativa 2**

Tomaremos também $(x) = (7; 0, 2)$ como estimativa inicial, e analisaremos o que acontece quando aplicamos o Método Gradiente e o Método de Gauss-Newton para este problema.

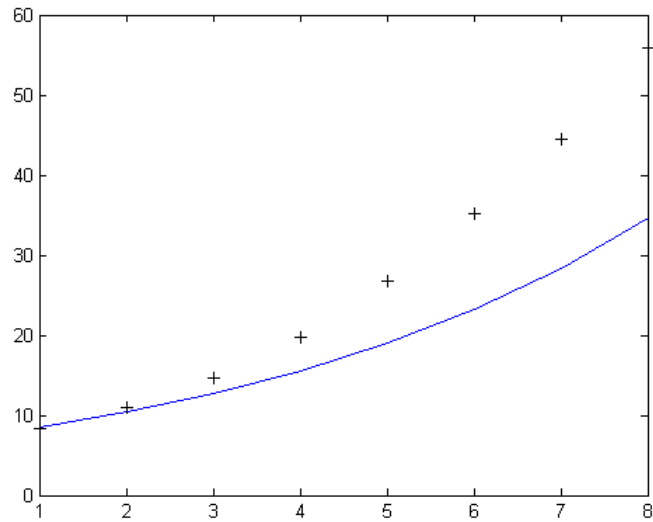


Figura 40: $\phi(t,y) = 7e^{0,2t}$

- **Estimativa 2 com a Condição de Armijo**

O gráfico a seguir traz a nova aproximação $\phi(t,y)$ a partir do parâmetro ótimo x determinado pelo Método Gradiente e pelo Método de Gauss-Newton simultaneamente, tomando como x inicial $x = (7;0,2)$ e tomando α pela Condição de Armijo, com $\eta = 0.7$.

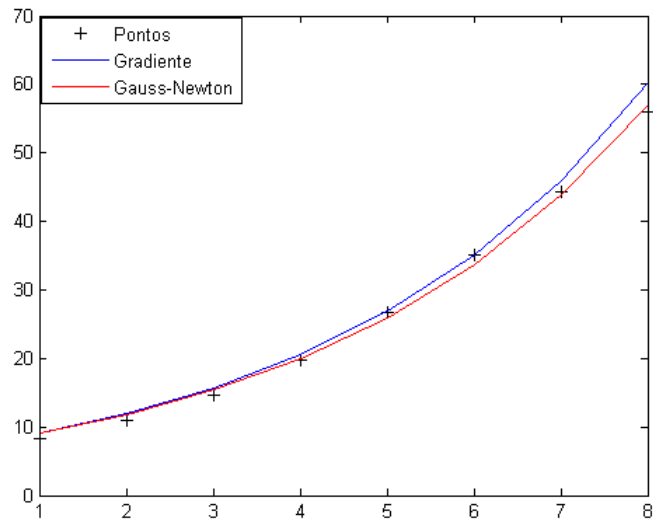


Figura 41: Aproximação a partir do x ótimo

- **Estimativa 2 com passo constante $\alpha = 1$**

De maneira análoga, aplicaremos o Método Gradiente e o Método de Gauss-Newton, tomando $\alpha = 1$ para a estimativa inicial $x = (7;0,2)$. Da mesma forma, o

Método Gradiente não foi eficiente.

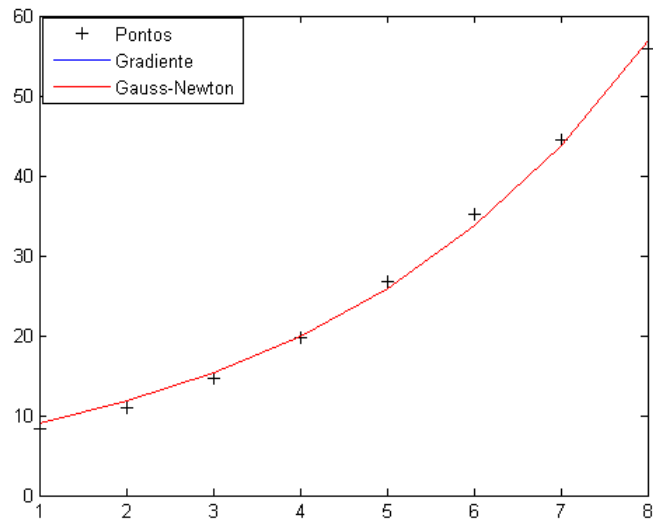


Figura 42: Aproximação a partir do x ótimo

As tabelas a seguir trazem os resultados obtidos, tomando um parâmetro x inicial e obtendo o x ótimo, $f(x)$ ótimo e número de iterações, quando aplicado o Método Gradiente e quando aplicado o Método de Gauss-Newton.

Tabela 13: Método Gradiente com α determinado por Armijo

x inicial	k	x^*	$f(x^*)$
(6;0,3)	59	(5,99;0,28)	8,140
(7;0,2)	55	(7;0,27)	12,483

Tabela 14: Método Gauss-Newton com α determinado por Armijo

x inicial	k	x^*	$f(x^*)$
(6;0,3)	3	(6,99;0,26)	3,009
(7;0,2)	19	(7;0,26)	3,006

Tabela 15: Método Gauss-Newton com $\alpha = 1$

x inicial	k	x^*	$f(x^*)$
(6;0,3)	8	(7;0,26)	3,006
(6;0,3)	10	(7;0,26)	3,006

Note que o Método de Gauss-Newton é mais eficiente quando tomamos o tamanho do passo determinado pela Condição de Armijo. Tal resultado também está em (CROEZE, 2012). Note que não há a tabela com os resultados obtidos quando aplicamos o Método Gradiente com $\alpha = 1$, pois o mesmo atingiu o número máximo de iterações ($k = 500$) não encontrando o x ótimo do problema. Note que neste exemplo não poderíamos utilizar o método da Seção Áurea para determinar α , pois a função $\phi(\alpha) = f(x + \alpha d)$ não é unimodal.

Exemplo 7.2. Considere a seguir um conjunto de dados a respeito das maiores temperaturas (em F°) registradas a cada mês, na cidade de Baton Rouge, em Los Angeles:

Tabela 16:

Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
61	65	72	78	85	90	92	92	88	81	72	63

Da mesma forma que o exemplo (6.1), este conjunto de dados pode ser modelado pela função $\phi(x,t) = x_1 \text{sen}(x_2 t + x_3) + x_4$ onde janeiro corresponde a $t = 1$, fevereiro corresponde a $t = 2$, e assim sucessivamente.

- **Estimativa 1**

Como estimativa inicial, vamos considerar $(x_1, x_2, x_3, x_4) = (17; 0,5; 10,5; 77)$. Assim,

$$\phi(t,y) = 17\text{sen}(0,5t + 10,5) + 77$$

é a estimativa inicial para o modelo.

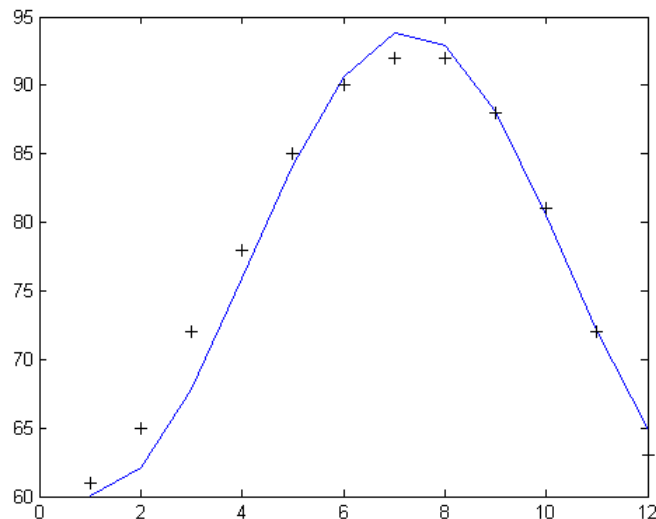


Figura 43: $\phi(t,y) = 17\text{sen}(0,5t + 10,5) + 77$

Vamos aplicar o Método Gradiente e o Método de Gauss-Newton para encontrar a melhor aproximação $\phi(t,y)$ em relação ao nosso conjunto de dados.

- **Estimativa 1 com passo determinado pela condição de Armijo**

O gráfico a seguir traz a aproximação encontrada aplicando o Método Gradiente e o Método de Gauss-Newton simultaneamente para x inicial $(x_1, x_2, x_3, x_4) = (17; 0,5; 10,5; 77)$.

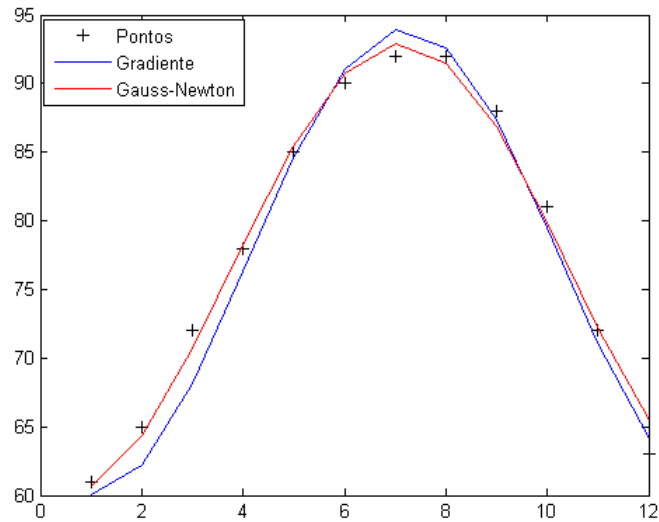


Figura 44: Estimativa a partir do x ótimo

Podemos notar que o Método Gradiente resultou em uma aproximação não muito diferente da aproximação inicial. Por outro lado, o Método de Gauss-Newton obteve uma aproximação melhor em relação à estimativa inicial.

- **Estimativa 1 com passo constante $\alpha = 1$**

O gráfico a seguir traz a aproximação encontrada pelos Métodos Gradiente e Gauss-Newton, mas agora com o tamanho do passo $\alpha = 1$.

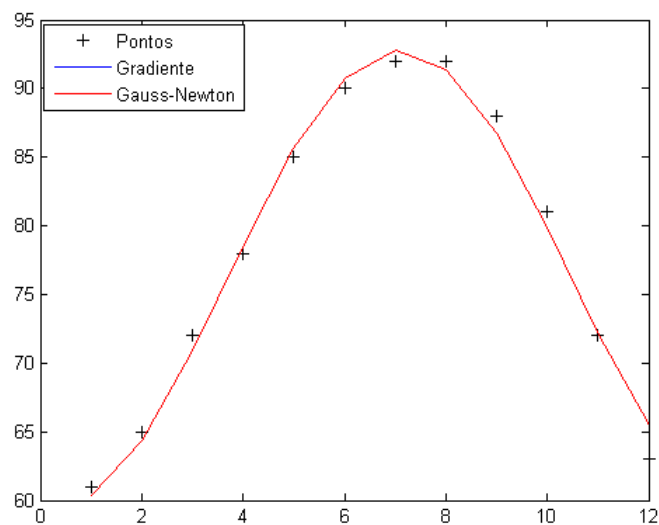


Figura 45: Aproximação a partir do x ótimo

Percebemos que o Método Gradiente não é eficaz quando tomamos $\alpha = 1$, pois não temos

garantia da convergência global do Método Gradiente quando tomamos o tamanho do passo como constante $\alpha = 1$.

- **Estimativa 2**

Assim como no exemplo anterior, analisaremos o que acontece quando tomamos uma estimativa inicial diferente. Tomemos x inicial como $(x_1, x_2, x_3, x_4) = (15; 0, 6; 10, 5; 70)$.

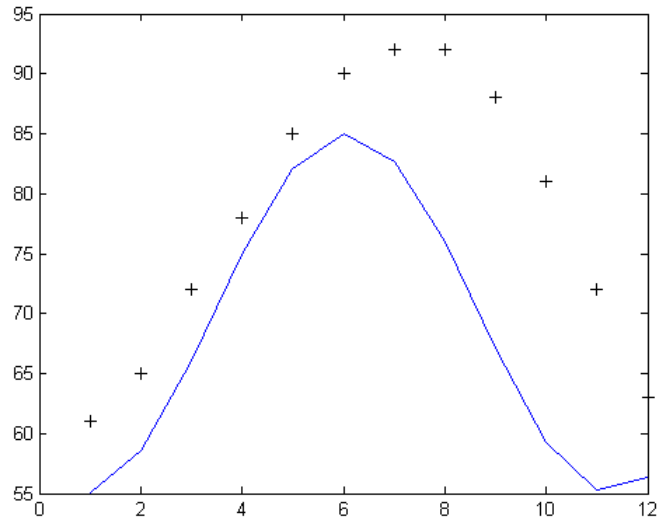


Figura 46: $\phi(t, y) = 15\text{sen}(0,6t + 10,5) + 70$

- **Estimativa 2 com a Condição de Armijo**

O gráfico a seguir traz a aproximação obtida quando aplicamos o Método Gradiente e o Método de Gauss-Newton simultaneamente para este problema, a partir de $(x_1, x_2, x_3, x_4) = (15; 0, 6; 10, 5; 70)$.

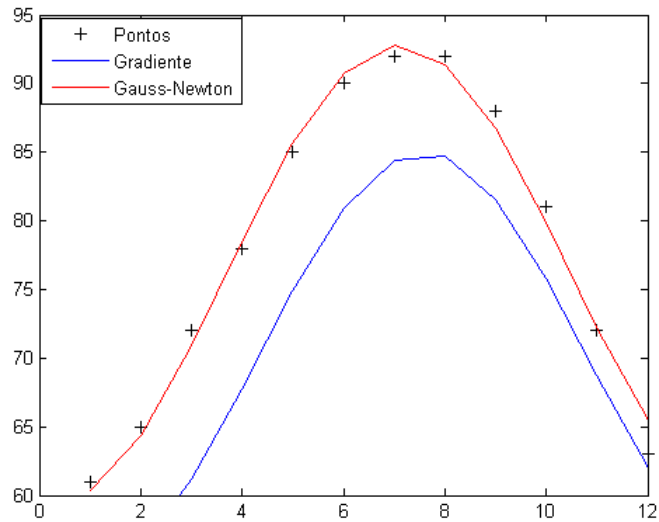


Figura 47: Aproximação a partir de x ótimo

Podemos ver que tomando a estimativa inicial a partir de

$$(x_1, x_2, x_3, x_4) = (15; 0, 6; 10, 5; 70)$$

o Método Gradiente não resulta em uma boa aproximação. Isto nos mostra que para o problema de Mínimos Quadrados o Método Gradiente é eficiente apenas localmente.

- **Estimativa 2 com tamanho de passo $\alpha = 1$**

Temos agora a aproximação obtida com os Métodos Gradiente e Gauss-Newton com o tamanho do passo $\alpha = 1$, tomando $(x_1, x_2, x_3, x_4) = (15; 0, 6; 10, 5; 70)$ como estimativa inicial.

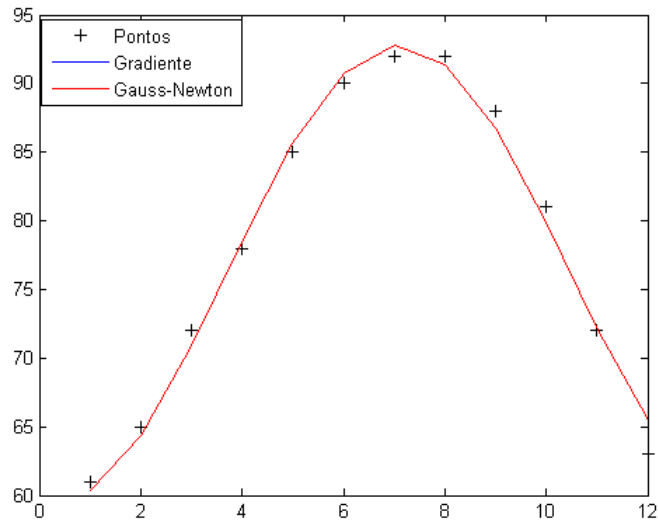


Figura 48: Aproximação a partir de x ótimo

Novamente, para $\alpha = 1$ o Método Gradiente não encontra uma aproximação $\phi(x, t)$ para o conjunto de dados em questão.

As tabelas a seguir, assim como no exemplo anterior, trazem os resultados obtidos, tomando um parâmetro x inicial e obtendo o x ótimo, $f(x)$ ótimo e número de iterações, quando aplicado o Método Gradiente e quando aplicado o Método de Gauss-Newton.

Tabela 17: Método Gradiente com α determinado por Armijo

x inicial	k	x^*	$f(x^*)$
(15; 0, 6; 10, 5; 70)	máx=500	(14, 99; 0, 48; 10, 48; 70)	352, 020
(17; 0, 5; 10, 5; 77)	53	(16, 99; 0, 505; 10, 501; 77)	18, 191

Tabela 18: Método Gauss-Newton com α determinado por Armijo

x inicial	k	x^*	$f(x^*)$
(15; 0, 6; 10, 5; 70)	19	(16, 63; 0, 46; 10, 85; 76, 19)	6, 511
(17; 0, 5; 10, 5; 77)	14	(16, 63; 0, 46; 10, 85; 76, 19)	6, 511

Tabela 19: Método Gauss-Newton com $\alpha = 1$

x inicial	k	x^*	$f(x^*)$
(15; 0, 6; 10, 5; 70)	15	(16, 63; 0, 46; 10, 85; 76, 19)	6, 511
(17; 0, 5; 10, 5; 77)	14	(16, 63; 0, 46; 10, 85; 76, 19)	6, 511

Note que não há a tabela com os resultados obtidos quando aplicamos o Método Gradiente com $\alpha = 1$, pois o mesmo atingiu o número máximo de iterações ($k = 500$) não encontrando o x ótimo do problema.

De fato, quando tomamos $\alpha = 1$ para o Método Gradiente não temos garantia de convergência global deste método. Como estudamos anteriormente, o Método Gradiente possui convergência global quando tomamos α determinado a partir da Condição de Armijo ou a partir da Seção Áurea. Para tamanhos de passo α tomados de outra forma, não sabemos nada a respeito da convergência global.

8 CONCLUSÕES

Neste trabalho estudamos inicialmente a forma de um problema de otimização irrestrita-problema (7), e as condições que caracterizam a solução deste problema. Em seguida estudamos como determinar o tamanho do passo α de maneira que a função f possua um decréscimo suficiente até que encontremos seu minimizador x^* .

Posteriormente, estudamos dois métodos clássicos de descida para resolver o problema (7): Método Gradiente e o Método de Newton. Abordamos as principais características destes métodos, e aplicando em problemas práticos, analisamos as principais diferenças entre eles.

Na segunda parte do trabalho, estudamos como estimar parâmetros utilizando a técnica de Mínimos Quadrados, reduzindo o problema inicial a um problema de otimização irrestrita. Para estimar estes parâmetros, estudamos dois métodos iterativos: Método Gradiente e Método de Gauss-Newton, para encontrar o parâmetro x^* ótimo. Em seguida, os aplicamos em alguns exemplos práticos e comparamos os resultados obtidos.

A partir do estudo e das comparações realizadas, percebemos a importância de ter toda a bagagem teórica bem clara para então compreender o que acontece com cada método, quando aplicado em problemas diferentes. E diante disso, notamos também que não há um método melhor que outro, mas de acordo com a forma do problema a ser resolvido, um método pode ser mais eficiente que outro sob certas condições.

Embora tenhamos abordado somente os dois métodos clássicos de otimização irrestrita (Método Gradiente e Método de Newton), existem diferentes métodos que resolvem o problema (7). Entre os principais, temos: Método de Direções Conjugadas, Métodos Quase-Newton e também o Método da Região de Confiança

Os métodos apenas citados aqui, o estudo a respeito da velocidade de convergência dos métodos que não foi estudada neste trabalho, além da otimização restrita, ficam como inspiração para trabalhos futuros.

REFERÊNCIAS

BOLDRINI, L. et. al. **Álgebra Linear**. 2 ed. São Paulo: Editora Harbra, 1980.

GUIDORIZZI, H. **Um curso de cálculo vol.2**. 5 ed. São Paulo: LTC, 2011.

LIMA, E. **Curso de Análise vol.1**. 14 ed. Rio de Janeiro: Editora IMPA, 2016.

FRIEDLANDER, A. **Elementos de Programação não-linear**. São Paulo: Editora Unicamp, 2009.

RIBEIRO, A.; KARAS, E. **Otimização Contínua - Aspectos teóricos e computacionais**. São Paulo: Cengage Learning, 2013.

BENATTI, K. *O método de Levenberg-Marquardt para o Problema de Quadrados Mínimos não-Linear*. 2017. 101f. Dissertação de Mestrado- Universidade Federal do Paraná, Curitiba, 2017.

NOCEDAL, J. **Numerical Optimization**. New York: Springer-Verlag, 2006.

MARTINEZ, J. **Métodos Computacionais de Otimização**. São Paulo: Editora Unicamp, 1998.

CROEZE, A. *Solving nonlinear Least-Squares problems with the Gauss-Newton and Levenberg-Marquardt Methods*. 2012. 18f. Artigo Acadêmico- Duke University, Durham, 2012.

AGUIAR, A. *Análise de Convergência Local do Método de Gauss-Newton sob condição Lipschitz*. 2012. 48f. Monografia de especiação- Universidade Federal de Goiás, Goiânia, 2012.