

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

UM ALGORITMO DE BUSCA LINEAR PARA OTIMIZAÇÃO IRRESTRITA

Manaus-AM

2016

UNIVERSIDADE FEDERAL DO AMAZONAS - UFAM
INSTITUTO DE CIÊNCIAS EXATAS - ICE
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

DANIELE ALENCAR FABRÍCIO DA SILVA

UM ALGORITMO DE BUSCA LINEAR PARA OTIMIZAÇÃO IRRESTRITA

Dissertação submetida ao corpo docente da coordenação do Programa de Pós Graduação em Matemática da Universidade Federal do Amazonas como parte dos requisitos necessários para obtenção do grau de Mestre em Matemática, área de concentração em Otimização.

Orientador: Prof. Dr. Roberto Cristóvão Mesquita Silva

Manaus-AM

2016

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).

S586u Silva, Daniele Alencar Fabrício da
Um algoritmo de busca linear para otimização irrestrita / Daniele Alencar Fabrício da Silva. 2016
50 f.: il.; 31 cm.

Orientador: Dr. Roberto Cristóvão Mesquita Silva
Dissertação (Matemática - Licenciatura em Matemática) -
Universidade Federal do Amazonas.

1. Otimização Irrestrita. 2. Busca Linear. 3. Convergência Global.
4. Taxa de convergência linear. I. Silva, Dr. Roberto Cristóvão Mesquita II. Universidade Federal do Amazonas III. Título

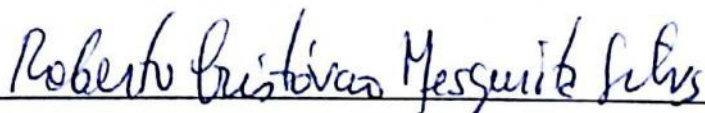
DANIELE ALENCAR FABRÍCIO DA SILVA

UM ALGORITMO DE BUSCA LINEAR PARA OTIMIZAÇÃO IRRESTRITA

Dissertação apresentada ao Curso de Mestrado em Matemática da Universidade Federal do Amazonas, como requisito parcial para obtenção do Grau de Mestre. Área de Concentração: Matemática Aplicada.

Aprovada em Novembro de 2016.

BANCA EXAMINADORA



Prof. Dr. ROBERTO CRISTÓVÃO MESQUITA SILVA

Presidente /Orientador - UFAM - ICE



Prof. Dr. SANDRO DIMY BARBOSA BITAR

Membro - UFAM - ICE



Prof. Dr. PAULO ROBERTO OLIVEIRA

Membro - UFRJ - COPPE

Manaus-Am

2016

*Dedico este trabalho a minha mãe,
Mary Carvalho, por todo o incentivo
e ajuda para que isso fosse possível.*

"A meta é só um começo de uma nova
corrida."

(autor desconhecido)

Agradecimentos

Ao concluir este trabalho, agradeço:

Primeiramente a Deus, pela vida e por ter me dado forças para vencer as barreiras.

A meus pais Mary e Frank e meus irmãos Leandro e Guilherme pelo esforço que fizeram para que eu pudesse chegar até aqui, obrigada pelo apoio e incentivo.

Ao meu esposo Sérgio pelo apoio, carinho, amor, paciência e incentivo.

Ao meu orientador, Roberto Cristóvão Mesquita Silva pela orientação, tolerância e atenção. Agradeço todos os conselhos e a ajuda na preparação deste trabalho. Por ter contribuído substancialmente na minha formação, sempre terá meu respeito e gratidão.

Aos meus amigos do mestrado que, em especial, a Márcia, Osenildo e Ezequiel, pela amizade e pelos inúmeros momentos de alegria. Estejam certos de que este sonho realizado constitui a presença marcante de cada um.

Aos professores do Departamento de Matemática, pelo bom relacionamento e por me tornarem capaz de realizar este trabalho. Em particular, aos professores Sandro Bitar e Flávia Morgana, que foram essenciais em minha formação. E ao professor Anselmo, por ser o responsável por me fazer escolher a matemática há dez anos.

Aos funcionários da secretaria, ao Ari e ao Euclimar. À Capes pelo suporte financeiro.

Resumo

Neste trabalho apresentamos um algoritmo de busca linear para problemas de otimização irrestrita proposto por Gonglin Yuan, Sha Lu e Zengxi Wei [1], denominado aqui por Algoritmo GSZ. Este algoritmo é concebido sob a perspectiva de herdar a simplicidade e o baixo custo computacional do método do gradiente conjugado. Neste contexto, uma prova detalhada da análise de convergência global para funções não necessariamente convexas é apresentada. Ressaltamos ainda a obtenção da taxa de convergência linear para o caso em que a função é fortemente convexa.

Palavras-chave: Otimização Irrestrita, Busca Linear, Convergência Global, Taxa de convergência linear.

Abstract

This work presents a linear search algorithm for unconstrained optimization problems proposed by Gonglin Yuan, Sha Lu Wei and Zengxi [1], called here by Algorithm GSZ. This algorithm is designed from the perspective of inheriting the simplicity and low computational cost of the conjugate gradient method. In this context, a detailed proof of the global convergence analysis for functions not necessarily convex is presented. We also emphasize the achievement of the linear convergence rate for the case where the function is strongly Convex.

Keywords: Unconstrained Optimization, Linear Search, Global Convergence, R-linear Convergence.

Lista de Figuras

2.1	Conjunto convexo.	11
2.2	Conjunto não convexo.	11
2.3	Função convexa.	12
2.4	Função não convexa.	12
2.5	O ponto $\bar{x} = 0$ não é minimizador local de f	20
2.6	O ponto $\bar{x} = 0$ é minimizador local do problema (1.1)	20
2.7	Ilustração do Teorema 2.14.	21
2.8	Interpretação da condição de Armijo	22
3.1	Busca linear exata.	32
3.2	Conjunto admissível pelas condições WWP	39
3.3	Conjunto admissível pelas condições de SWP	40

Sumário

Agradecimentos	viii
Resumo	ix
Abstract	x
Lista de Tabelas	1
Introdução	4
1 Introdução	4
2 Conceitos básicos	7
2.1 Resultados de Análise no \mathbb{R}^n	7
2.2 Conceitos e resultados de Análise Convexa	10
2.2.1 Conjuntos e Funções Convexas	11
2.3 Conceitos e resultados de Otimização	17

	3
2.3.1 O problema de otimização	18
2.4 Métodos de descida	20
2.4.1 Busca inexata - condição de Armijo	21
2.4.2 Algoritmo dos gradientes conjugados (GC)	22
3 Algoritmo de Yuan, Lu e Wei	26
3.1 Algoritmo GSZ	26
3.2 Convergência global do Método GSZ	31
3.2.1 Algoritmo GSZ com Busca Linear Exata	31
3.2.2 Algoritmo GSZ com o critério de Goldstein	35
3.2.3 Algoritmo GSZ com o critério de Wolf-Powell	38
3.3 Taxa de convergência	41
4 Considerações Finais	48
Referências Bibliográficas	49

Capítulo 1

Introdução

Neste trabalho, abordamos um algoritmo de busca linear para otimização irrestrita que foi proposto por Gonglin Yuan, Sha Lu e Zengxi Wei, em [15]; o qual denominamos de Algoritmo GSZ, conforme as iniciais dos nomes dos autores. Este algoritmo está voltada para os problemas de otimização irrestrito, que é denotado da seguinte maneira:

$$\begin{aligned} \min f(x) \\ \text{s.a. } x \in \mathbb{R}^n, \end{aligned} \tag{1.1}$$

onde a função objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável. Procuramos, portanto, se existir, um ponto \bar{x} tal que $f(\bar{x}) \leq f(x), \forall x \in \mathbb{R}^n$.

A maior parte dos algoritmos de busca linear exigem que d seja uma direção de descida. Conforme apresentado em [10], existe uma condição suficiente para que uma direção seja de descida, a saber:

$$\nabla f(\bar{x})^T d < 0. \tag{1.2}$$

Existem, na literatura, muitos métodos [8,12] que resolvem (1.1) cumprindo (1.2). Estes métodos, normalmente, são iterativos: a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, geram recursivamente uma sequência de pontos $\{x_k\} \in \mathbb{R}^n$ pela iteração:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \dots,$$

onde x_k é o ponto de iteração corrente, d_k é uma direção de busca e $\alpha_k > 0$ é o comprimento do passo.

O método do Gradiente puro (ou Cauchy), de Newton, os quase-Newton (por exemplo: BFGS, DFP) e dos Gradientes Conjugados (GC), são exemplos clássicos de métodos que utilizam buscas lineares. A essência de cada método está no cálculo de sua direção de descida, o que implica diretamente na eficiência e convergência dos mesmos. A direção escolhida no método do Gradiente é a oposta ao gradiente da função a cada iteração, ou seja, $d_k = -\nabla f(x_k)$. Este método é de primeira ordem e globalmente convergente, porém, possui convergência bastante lenta. Por outro lado, o método de Newton tem boa convergência, mas uma das desvantagens é uso de derivadas de segunda ordem, o que o torna caro. Os métodos dos GC e quase-Newton, por sua vez, são mais velozes que o gradiente puro e mais baratos que o Newton, tornando-se uma boa alternativa para os problemas com velocidade de convergência e complexidade de cálculos intermediários. De acordo com Izmailov e Solodov [5], os métodos GC e quase-Newton, são vistos como os mais adequados para minimização irrestrita de funções diferenciáveis. No entanto, quando se trata em resolver os problemas de grandes porte, o método GC tornou-se uma ferramenta poderosa, devido a sua simplicidade e sua exigência de baixa memória [8,9]. Sob este aspecto, o método GC com algumas condições de busca linear [3], tem sido amplamente estudado ao longo dos últimos anos, como exemplo veja [13,16,17].

Nesse contexto, faremos um estudo teórico do Algoritmo GSZ, que foi inspirado em gradientes conjugados como citado em [15]. Abordamos de forma detalhada as-

pectos de convergência global, seguindo a ideia dos autores [15]. O estudo da taxa convergência foi realizado para o caso no qual a função é fortemente convexa, cuja demonstração foi baseado em [11]. Vale ressaltar, que as direções geradas pelo algoritmo GSZ, além de cumprirem a condição (1.2), também têm a seguinte propriedade, que é fundamental na análise de convergência do Algoritmo GSZ:

$$\|d_k\| \leq c_2 \|\nabla f_k\|, \forall k \geq 0, c_2 \in (0, 1). \quad (1.3)$$

A fim de obter a conclusão apresentada acima, destinamos o Capítulo 2, às definições e resultados teóricos básicos de análise no \mathbb{R}^n , de análise convexa e de otimização. Além disso, apresentamos o algoritmo do gradiente conjugado.

Por fim, no capítulo 3, serão expostos os principais resultados deste trabalho: supondo a existência de solução para o problema (1.1) apresentaremos o Algoritmo GSZ, descrevendo os aspectos de convergência global, bem como a taxa de convergência linear obtida para o caso em que a função objetivo é fortemente convexa.

Capítulo 2

Conceitos básicos

Neste capítulo, veremos conceitos relacionados à topologia do espaço euclidiano, à análise convexa e à otimização irrestrita, no intuito de esclarecer e fixar a notação. Ressaltamos que as demonstrações dos resultados básicos serão omitidas, entretanto podem ser encontradas em [10], [12], [7], [4], [5].

2.1 Resultados de Análise no \mathbb{R}^n

Definição 1. *O produto interno de x por y é uma aplicação $\langle \cdot, \cdot \rangle : \mathbb{R}^n \rightarrow \mathbb{R}$ que cumpre as condições abaixo, para $x, y, z \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ arbitrários:*

PI.1 $\langle x, y \rangle = \langle y, x \rangle;$

PI.2 $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle;$

PI.3 $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle;$

PI.4 $x \neq 0 \Rightarrow \langle x, x \rangle > 0.$

Exemplo 2.1. O produto interno canônico do espaço euclidiano \mathbb{R}^n , o qual é dado por

$$\langle x, y \rangle = x_1y_1 + \cdots + x_ny_n,$$

onde $x = (x_1, \dots, x_n)$ e $y = (y_1, \dots, y_n)$.

Definição 2. A função real $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ é chamada norma em um espaço vetorial \mathbb{R}^n se, e somente se, satisfaz as seguintes propriedades para quaisquer $x, y \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$

N1. Se $x \neq 0 \Rightarrow \|x\| \neq 0$;

N2. $\|\alpha x\| = |\alpha| \|x\|$;

N3. $\|x + y\| \leq \|x\| + \|y\|$.

A partir do produto interno, definimos a norma euclidiana de um vetor $x \in \mathbb{R}^n$ por $\|x\|^2 = \langle x, x \rangle$.

Proposição 2.1. (Cap. 1, p. 5, [7]) **[Desigualdade de Cauchy-Schwarz]** Seja $\langle \cdot, \cdot \rangle$ um produto interno e $\|\cdot\|$ a norma por ele induzida. Então, para quaisquer $x, y \in \mathbb{R}^n$, temos

$$\|\langle x, y \rangle\| \leq \|x\| \cdot \|y\|.$$

Além disso, vale a igualdade se, e somente se, um dos vetores x, y é múltiplo escalar do outro.

Uma sequência em \mathbb{R}^n é uma aplicação $x : \mathbb{N} \rightarrow \mathbb{R}^n$, definida no conjunto $\mathbb{N} = \{0, 1, 2, \dots, n, \dots\}$. O valor que essa função assume no número k é indicado com x_k e chama-se o k -ésimo termo da sequência. No decorrer deste trabalho, adotamos a notação $\{x_k\}$ para indicar a sequência cujo k -ésimo termo é $x_k \in \mathbb{R}^n$. Uma subsequência de $\{x_k\}$ é a restrição da sequência a um subconjunto infinito $N' = \{k_1 < k_2 < \dots < k_i < \dots\} \subset N$. Indicaremos uma subsequência pela notação $\{x_{k_j}\}$.

Definição 3. Dizemos que o ponto $\bar{x} \in \mathbb{R}^n$ é o limite da sequência $\{x_k\}$ quando, para todo $\varepsilon > 0$ dado, é possível obter $\bar{k} \in \mathbb{N}$ tal que

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

Neste caso, a sequência $\{x_k\}$ converge para \bar{x} e denotamos por $x_k \rightarrow \bar{x}$ ou $\lim_{k \rightarrow \infty} x_k = \bar{x}$.

Definição 4. Seja $f : \Omega \rightarrow \mathbb{R}$ uma função definida no conjunto $\Omega \subset \mathbb{R}^n$. Dizemos que f é contínua no ponto $a \in \Omega$ quando, para todo $\varepsilon > 0$ dado arbitrariamente, pode-se obter $\delta > 0$ tal que

$$x \in \Omega, \|x - a\| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon.$$

Dizemos que f é contínua em Ω quando é contínua em todos os pontos de Ω .

Definição 5. Seja $f : \Omega \rightarrow \mathbb{R}$ uma função definida no conjunto $\Omega \subset \mathbb{R}^n$. Dizemos que f é Lipschitz contínua em Ω quando existe $L > 0$ (denominada constante Lipschitz de f) tal que, para quaisquer $x, y \in \Omega$ tem-se

$$|f(x) - f(y)| \leq L\|x - y\|.$$

Corolário 2.1. (Cap.5, p140, [6]) Seja $f : \Omega \rightarrow \mathbb{R}$ uma função definida no subconjunto $\Omega \subset \mathbb{R}^n$. Dizemos que f é contínua no ponto $a \in \Omega$ se, e somente se, para toda sequência de pontos $\{x_k\} \in \mathbb{R}^n$, com $\lim_{k \rightarrow \infty} x_k = a$ tem-se $\lim_{k \rightarrow \infty} f(x_k) = f(a)$.

Teorema 2.1. (Cap.2, p.27, [10])(**Weierstrass**) Sejam $f : \Omega \rightarrow \mathbb{R}$ contínua e $\Omega \subset \mathbb{R}^n$ compacto não vazio. Então existe minimizador global de f em Ω .

As aproximações de Taylor para uma função constituem uma das mais importantes ferramentas em otimização. A seguir apresentamos alguns teoremas que serão importantes no desenvolvimento deste trabalho.

Teorema 2.2. (Cap. 3, p. 123, [7])(**Teorema do Valor Médio**) Considere $f :$

$\mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável no segmento $(x, x+d)$. Então, dados $x, y \in \mathbb{R}^n$ existe um ponto z no segmento ligando x a y tal que

$$f(y) - f(x) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(z)(y-x)_i$$

onde $(y-x)_i$ é a i -ésima componente do vetor $(y-x) \in \mathbb{R}^n$.

Observação 2.1. Podemos escrever a equação anterior na forma

$$f(y) - f(x) = (y-x)^T \nabla f(z).$$

Teorema 2.3. (Cap. 1, p.22, [10])[Taylor com resto de Lagrange] Considere $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 e $x, d \in \mathbb{R}^n$. Se f é duas vezes diferenciável no segmento $(x, x+d)$, então existe $t \in (0, 1)$ tal que

$$f(x+d) = f(x) + \nabla f(x)^T d + \frac{1}{2} d^T \nabla^2 f(x+td)^T d.$$

Teorema 2.4. (Cap. 1, p.26, [5])(Fórmula de Newton-Leibnitz) Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}^l$ uma função continuamente diferenciável. Então para todo $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$ tem-se que

$$F(x+y) = F(x) + \int_0^1 F'(x+ty) y dt.$$

2.2 Conceitos e resultados de Análise Convexa

O conceito de convexidade é fundamental na otimização, pois quando os problemas possuem esta propriedade são mais fáceis de resolver. Nesta seção, apresentamos os conceitos fundamentais e resultados de conjuntos convexos e funções convexas.

2.2.1 Conjuntos e Funções Convexas

Os conjuntos convexos constituem o domínio natural para as funções convexas, conforme veremos agora.

Definição 6. *Seja o conjunto $\Omega \subset \mathbb{R}^n$. Se, para todo $x, y \in \Omega$, tivermos*

$$\alpha x + (1 - \alpha)y \in \Omega, \quad \forall \alpha \in [0, 1]$$

então Ω é dito ser um conjunto convexo.

Geometricamente, esta definição nos diz que o segmento de reta $[x, y] := \{\alpha x + (1 - \alpha)y : 0 \leq \alpha \leq 1\}$ está inteiramente contido em Ω , sempre que os pontos extremos x e y estão em Ω . Na figura ilustramos, um conjunto convexo e outro não.

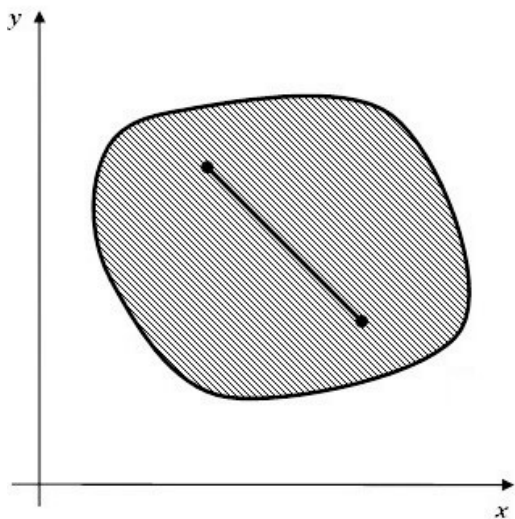


Figura 2.1: Conjunto convexo.

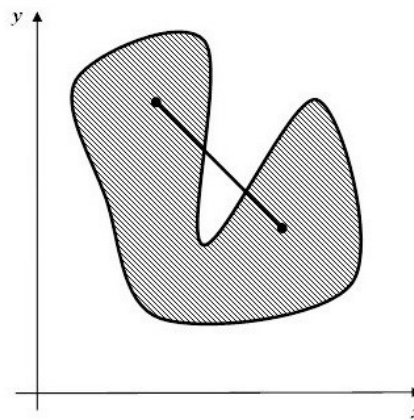


Figura 2.2: Conjunto não convexo.

Definição 7. (Função Convexa) *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo não vazio. Dizemos que a função $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa em Ω quando, para quaisquer $x, y \in \Omega$ e todo $\alpha \in [0, 1]$, tem-se*

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

Diz-se que f é *estritamente convexa* quando a desigualdade acima é estrita para todo $x \neq y$ e $\alpha \in (0, 1)$.

Diz-se que f é **fortemente convexa** em Ω com *módulo* c quando existe uma constante $c > 0$ onde para quaisquer $x, y \in \Omega$ e todo $\alpha \in [0, 1]$, tem-se

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \frac{1}{2}c\alpha(1 - \alpha)\|x - y\|^2 \quad (2.1)$$

Observação 2.2.

- (a) Uma função fortemente convexa é estritamente convexa, e uma função estritamente convexa é convexa.
- (b) Uma vez que \mathbb{R}^n é convexo, dizemos que o problema (1.1) é um problema de minimização convexo quando f é uma função convexa em \mathbb{R}^n .

A seguir veremos uma ilustração de uma função convexa e outra não-convexa.

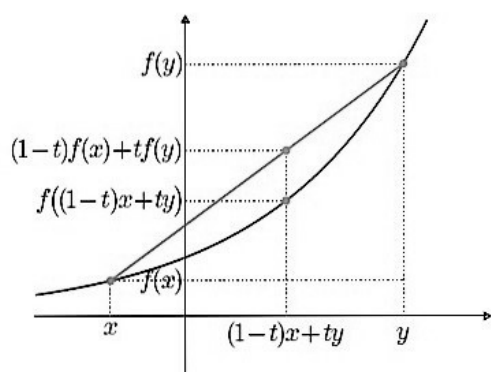


Figura 2.3: Função convexa.

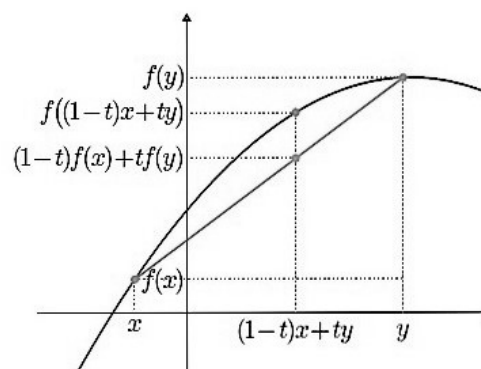


Figura 2.4: Função não convexa.

Exemplo 2.2. Seja $D \subset \mathbb{R}^n$ convexo. A função $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ é fortemente convexa com *módulo* γ se, e somente se, $h : D \subset \mathbb{R}^n$ definida por $h(x) = f(x) - \gamma\|x\|^2$ é convexa.

Primeiramente, note que por hipótese a função f é fortemente convexa e $\|x\|^2$ é fortemente convexa com módulo um. Sendo assim, a função h é uma soma de funções convexas e, conseqüentemente, uma função convexa. Reciprocamente, para todo $x, y \in D$ e $\gamma \in [0, 1]$ temos,

$$\begin{aligned}
f(\alpha x + (1 - \alpha)y) &= h(\alpha x + (1 - \alpha)y) + \gamma \|\alpha x + (1 - \alpha)y\|^2 \\
&\leq \alpha h(x) + (1 - \alpha)h(y) + \gamma \|\alpha x + (1 - \alpha)y\|^2 \\
&\Downarrow \\
&\leq \alpha (f(x) - \gamma \|x\|^2) + (1 - \alpha) (f(y) - \gamma \|y\|^2) + \gamma \|\alpha x + (1 - \alpha)y\|^2 \\
&\Downarrow \\
&= \alpha f(x) + (1 - \alpha)f(y) + \gamma [\|\alpha x + (1 - \alpha)y\|^2 - \alpha \|x\|^2 - (1 - \alpha)\|y\|^2] \\
&\Downarrow \\
&= \alpha f(x) + (1 - \alpha)f(y) + \gamma [\alpha^2 \|x - y\|^2 + 2\alpha \langle x, y \rangle - 2\alpha \|y\|^2 \\
&\quad + \|y\|^2 - \alpha \|x\|^2 - (1 - \alpha)\|y\|^2] \\
&\Downarrow \\
&= \alpha f(x) + (1 - \alpha)f(y) + \gamma [\alpha^2 \|x - y\|^2 + 2\alpha \langle x, y \rangle - \alpha \|x\|^2 - \alpha \|y\|^2] \\
&\Downarrow \\
&= \alpha f(x) + (1 - \alpha)f(y) + \alpha \gamma [\alpha \|x - y\|^2 - (\|x\|^2 - 2\alpha \langle x, y \rangle - \|y\|^2)] \\
&\Downarrow \\
&= \alpha f(x) + (1 - \alpha)f(y) + \alpha \gamma [\alpha \|x - y\|^2 - \|x - y\|^2] \\
&\Downarrow \\
&= \alpha f(x) + (1 - \alpha)f(y) - \alpha \gamma (1 - \alpha) \|x - y\|^2 \\
&\Downarrow \\
f(\alpha x + (1 - \alpha)y) &\leq \alpha f(x) + (1 - \alpha)f(y) - \alpha \gamma (1 - \alpha) \|x - y\|^2.
\end{aligned}$$

Logo, a função f é fortemente convexa.

Teorema 2.5. (Cap. 1, p. 47, [12]) *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e não vazio, f uma função convexa definida em Ω , $\alpha \in \mathbb{R}$. Então o conjunto de nível $L_0 = \{x \mid \in \Omega, f(x) \leq f(x_0)\}$ é um conjunto convexo.*

Dizemos que (1.1) é um problema de minimização convexa, quando $\Omega \subset \mathbb{R}^n$ é um conjunto convexo e $f : \Omega \rightarrow \mathbb{R}$ é uma função convexa em Ω . A importância da convexidade já pode ser vista no seguinte resultado.

Teorema 2.6. (Cap. 3, p. 77, [4]) **(Teorema de Minimização Convexa)** *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e $f : \Omega \rightarrow \mathbb{R}$ uma função convexa. Então todo minimizador local do problema (1.1) é global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Teorema 2.7. (Cap. 3, p 144, [4]) *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto. Seja $f : \Omega \rightarrow \mathbb{R}$ convexa em Ω . Então f é localmente Lipschitz-contínua em Ω . Em particular, f é contínua em Ω .*

Teorema 2.8. (Cap. 3, p 148, [4]) *Suponhamos que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja fortemente convexa no \mathbb{R}^n . Então o conjunto de nível*

$$L_0 = \{x \mid x \in \mathbb{R}^n, f(x) \leq f(x_0)\}$$

é compacto para todo $x_0 \in \mathbb{R}$.

Corolário 2.2. (Cap. 3, p 149, [4]) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função fortemente convexa e $\Omega \subset \mathbb{R}^n$ não vazio, um conjunto fechado qualquer. Então f tem um minimizador em Ω . Se Ω é convexo, o minimizador é único.*

Teorema 2.9. (Cap. 1, p. 47, [12]) **(Compacidade do conjunto de nível de uma função fortemente convexa)** *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função duas vezes continuamente diferenciável em $\Omega \in \mathbb{R}^n$, onde Ω é um conjunto convexo não vazio. Suponha-se que existe $\gamma > 0$ tal que*

$$2\gamma\|d\|^2 \leq d^T \nabla^2 f(x)d, \quad \forall x \in L_0, d \in \mathbb{R}^n$$

Então o conjunto de nível $L_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ é um conjunto convexo compacto.

Quando uma função é diferenciável, a convexidade admite caracterizações que são muito úteis para determinar se uma função é convexa ou não.

Teorema 2.10. (Cap. 3, p. 155, [4])(**Caracterização de Funções Convexas Diferenciáveis**) *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto não vazio, convexo e aberto e $f : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável em Ω . Então as propriedades seguintes são equivalentes:*

(a) *A função f é convexa em Ω .*

(b) *Para todo $x \in \Omega$ e todo $y \in \Omega$,*

$$f(y) \geq f(x) + (y - x)^T \nabla f(x).$$

(c) *Para todo $x \in \Omega$ e todo $y \in \Omega$,*

$$(y - x)^T \nabla f(x) \geq 0$$

Quando f é duas vezes diferenciável em Ω , as propriedades acima também são equivalentes a

(d) *A matriz Hessiana de f é semidefinida positiva em todo ponto de Ω :*

$$d^T \nabla^2 f(x) \geq 0 \quad \forall d \in \mathbb{R}^n$$

Para as funções diferenciáveis fortemente convexas, temos critérios que são análogos àqueles do Teorema 2.10 para as funções convexas.

Teorema 2.11. (Cap. 3, p. 164, [4])(**Caracterização de Funções Fortemente Convexas**) *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto não vazio, convexo e aberto e $f : \Omega \rightarrow \mathbb{R}$ uma função diferenciável em Ω , com derivada contínua em Ω . Então as propriedades seguintes são equivalentes:*

- (a) A função f é fortemente convexa em Ω com módulo $\gamma > 0$.
 (b) Para todo $x \in \Omega$ e todo $y \in \Omega$,

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + \gamma\|y - x\|^2$$

- (c) Para todo $x \in \Omega$ e todo $y \in \Omega$,

$$(\nabla f(y) - \nabla f(x))^T(y - x) \geq 2\gamma\|y - x\|^2$$

Quando f é duas vezes diferenciável em Ω , as propriedades acima também são equivalentes a

- (d) A matriz Hessiana de f é definida positiva uniformemente em todo ponto de Ω , i.e.,

$$\forall x \in \Omega, d^T \nabla^2 f(x) d \geq 2\gamma\|d\|^2 \quad \forall d \in \mathbb{R}^n$$

Demonstração. Primeiramente, mostraremos que (a) \Leftrightarrow (b). Seja $h : \Omega \rightarrow \mathbb{R}$, definida por

$$h(x) = f(x) - \gamma\|x\|^2 \text{ onde } \gamma > 0.$$

Pelo exemplo (2.2), vimos que $h(x)$ é convexa $\Leftrightarrow f(x)$ é fortemente convexa em Ω com módulo $\gamma > 0$. Agora, note que

$$\begin{aligned} h(x) = f(x) - \gamma\|x\|^2 &\text{ é convexa} \\ \Leftrightarrow h(y) &\geq h(x) + \nabla h(x)^T(y - x), \quad \forall x, y \in \Omega \\ \Leftrightarrow f(y) - \gamma\|y\|^2 &\geq f(x) - \gamma\|x\|^2 + (\nabla f(x) - 2\gamma x)^T(y - x) \\ \Leftrightarrow f(y) &\geq f(x) + \nabla f(x)^T(y - x) + (\gamma\|y\|^2 - \gamma\|x\|^2 - 2\gamma x)^T(y - x) \\ \Leftrightarrow f(y) &\geq f(x) + \nabla f(x)^T(y - x) + \gamma\|x - y\|^2 \end{aligned}$$

Concluimos assim que $a \Leftrightarrow b$. A seguir mostraremos que $a \Leftrightarrow c$. Agora, note que f é

fortemente convexa em Ω com módulo $\gamma > 0$, se e somente se,

$$\begin{aligned}
& h(x) = f(x) - \gamma\|x\|^2 \text{ é convexa em } \Omega \\
& \Leftrightarrow (\nabla h(y) - \nabla h(x))^T (y - x) \geq 0, \quad \forall x, y \in \Omega \\
& \Leftrightarrow ((\nabla f(y) - 2\gamma y) - (\nabla f(x) - 2\gamma x))^T (y - x) \geq 0, \quad \forall x, y \in \Omega \\
& \Leftrightarrow (\nabla f(y) - \nabla f(x))^T (y - x) \geq 2\gamma\|y - x\|^2 \quad \forall x, y \in \Omega.
\end{aligned}$$

Mostraremos agora que (a) \Leftrightarrow (d). Suponhamos agora que f seja duas vezes diferenciável em Ω com módulo $\gamma > 0$, se e somente se,

$$\begin{aligned}
& h(x) = f(x) - \gamma\|x\|^2 \text{ é convexa em } \Omega \\
& \Leftrightarrow d\nabla^2 h(x)^T(d) \geq 0, \quad \forall x, y \in \Omega \\
& \Leftrightarrow d(\nabla^2 f(x) - 2\gamma I)^T d \geq 0 \\
& \Leftrightarrow d\nabla^2 f(x)^T(d) \geq 2\gamma\|d\|^2.
\end{aligned}$$

□

2.3 Conceitos e resultados de Otimização

Nesta seção nossa atenção está voltada para alguns fatos básicos de otimização, que serão necessários ao longo deste trabalho. Algumas referências para este capítulo são [12], [10], [4] e [8], onde é possível encontrar as demonstrações omitidas, como também, as demais definições aqui apresentadas.

2.3.1 O problema de otimização

Vamos considerar aqui o problema

$$\begin{aligned} \min f(x) \\ \text{s.a. } x \in \mathbb{R}^n, \end{aligned}$$

onde a função objetivo $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é continuamente diferenciável.

Definição 8. *Seja a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Dizemos que \bar{x} é um minimizador local do problema (1.1), se existe $\delta > 0$, tal que*

$$f(\bar{x}) \leq f(x), \quad \forall x \in B(\bar{x}, \delta). \quad (2.2)$$

E dizemos que \bar{x} é minimizador global do problema (1.1) se

$$f(\bar{x}) \leq f(x), \quad \forall x \in \mathbb{R}^n. \quad (2.3)$$

Se as desigualdades (2.2) e (2.3) forem estritas, então será chamado, respectivamente, de minimizador estrito local e minimizador estrito global.

Definição 9. *Dizemos que $\bar{v} \in [-\infty, +\infty)$ definido por*

$$\bar{v} = \inf_{x \in \mathbb{R}^n} f(x) \quad (2.4)$$

é o valor ótimo do problema (1.1).

Uma função pode admitir vários minimizadores globais, mas o valor ótimo do problema é sempre o mesmo. Em seguida veremos condições que garantem a existência de solução global para o problema (1.1).

Teorema 2.12. *(Cap.2, p. 27, [10])(Weierstrass) Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ contínua e $\Omega \in \mathbb{R}^n$, onde Ω compacto não vazio. Então existe minimizador global de f em Ω .*

Corolário 2.3. (Cap.2, p. 28, [10]) *Sejam $\Omega \subset \mathbb{R}^n$ e $f : \Omega \rightarrow \mathbb{R}$ contínua no conjunto Ω . Suponhamos que existe $\alpha \in \mathbb{R}$ tal que o conjunto de nível $\mathcal{L}(\alpha)$ seja não vazio e compacto. Então o problema de minimizar f em Ω possui uma solução global.*

Veremos a seguir as condições que devem ser satisfeitas quando um $\bar{x} \in \mathbb{R}^n$ dado é minimizador (local) do problema (1.1). E as condições que garantem que um ponto é minimizador local do problema.

Teorema 2.13. (Cap.2, p. 29, [10]) **(Condição de otimalidade no caso irrestrito)**

(a) *Suponhamos que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Suponhamos também que \bar{x} é um minimizador local do problema (1.1). Então*

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

(b) *Se f é duas vezes diferenciável no ponto $\bar{x} \in \mathbb{R}^n$, então além de (Teorema 2.11 (a)) tem-se que a matriz Hessiana de f no ponto \bar{x} é semidefinida positiva, isto é,*

$$d^T \nabla^2 f(\bar{x}) d \geq 0 \quad \forall d \in \mathbb{R}^n.$$

Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja duas vezes diferenciável no ponto $\bar{x} \in \mathbb{R}^n$.

(c) *Se $\bar{x} \in \mathbb{R}^n$ satisfaz o item (a) e se a matriz Hessiana de f em $\bar{x} \in \mathbb{R}^n$ é definida positiva, isto é, se*

$$d^T \nabla^2 f(\bar{x}) d > 0 \quad \forall d \in \mathbb{R}^n.$$

Observação 2.3. *Um ponto $x \in \mathbb{R}^n$ que cumpre a condição (Teorema 2.11 (a)) é dito ponto crítico ou estacionário do problema (1.1).*

Observação 2.4. *Para um problema de programação convexo irrestrito diferenciável (Ω é um conjunto convexo e f uma função convexa e diferenciável), a condição (Teorema 2.11 (a)) é necessária e suficiente para otimalidade do problema.*

Exemplo 2.3. Seja $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^3$. Essa função satisfaz a desigualdade (Teorema 2.11 (b)) no ponto $\bar{x} = 0$, mas este ponto não é minimizador local do problema 1.1.

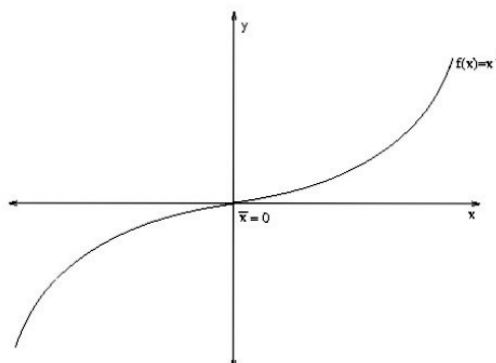


Figura 2.5: O ponto $\bar{x} = 0$ não é minimizador local de f .

Exemplo 2.4. Seja $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^4$. Essa função satisfaz a desigualdade (Teorema 2.11 (c)) no ponto $\bar{x} = 0$, mas este ponto é minimizador local do problema (1.1).

Figura 2.6: O ponto $\bar{x} = 0$ é minimizador local do problema (1.1)

2.4 Métodos de descida

Definição 10. 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 $f(\bar{x} + \alpha d) < f(\bar{x})$, para todo $\alpha \in (0, \delta)$

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

Teorema 2.14. (Cap.2, p. 47, [10]) Se $\nabla f(x_k)^T d < 0$, então d é uma direção de descida para f , a partir de \bar{x} .

Quando $n = 2$ ou $n = 3$, geometricamente estamos dizendo que as direções formam um ângulo obtuso com $\nabla f(x_k)$ são de descida. Veja a figura

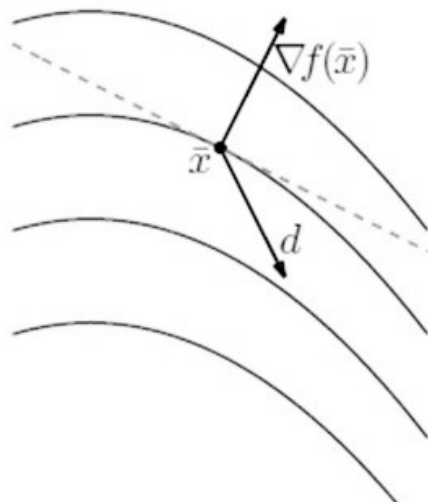


Figura 2.7: Ilustração do Teorema 2.14.

2.4.1 Busca inexata - condição de Armijo

A condição de Armijo é definida pela seguinte desigualdade:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k \quad (2.5)$$

onde $\sigma_1 \in (0, 1)$. Em outras palavras, a condição estabelece que redução na função objetivo deve ser proporcional ao comprimento do passo α_k e à derivada direcional $\nabla f(x_k)^T d_k$.

Geometricamente, o conjunto admissível para a condição de Armijo consiste de todos os comprimentos de passos α tais que o valor da imagem do ponto $x_k + d_k$ esteja abaixo da reta $l(\alpha) = f(x_k) + \alpha \sigma \nabla f(x_k)^T d_k$, conforme mostrado na Figura 2.2.

Na figura 2.2 consideramos a função $\varphi : [0, \infty) \rightarrow \mathbb{R}$ dada por $\varphi(\alpha) = f(x_k + d_k)$ e p como sendo o polinômio de Taylor de primeira ordem em torno do ponto $\alpha = 0$, também chamada de modelo linear que é dada por

$$p(\alpha) = \varphi(0) + \varphi'(0) = f(x_k) + \alpha \nabla f(x_k)^T d$$

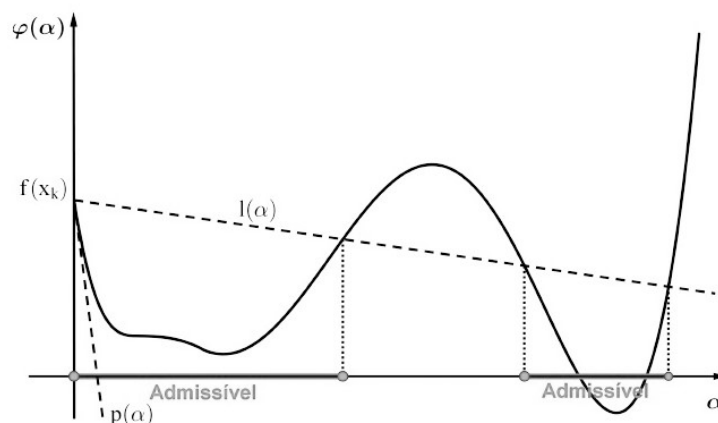


Figura 2.8: Interpretação da condição de Armijo

. Assim, podemos escrever a relação 2.5 como

$$\varphi(0) + \varphi(\alpha) = f(x_k) + f(x_k + \alpha d) \geq \sigma (p(0) - p(\alpha))$$

. Note também que na Figura 2.8 que a reta $l(\alpha)$ é dada por

$$l(\alpha) = f(x_k) + \sigma \alpha \nabla f(x_k)^T d_k$$

.

2.4.2 Algoritmo dos gradientes conjugados (GC)

O método geral do gradiente conjugado é projetado para resolver o problema (1.1). De acordo com Nocedal [8], sua primeira vertente foi proposta, pra problemas lineares, por Hestenes e Stiefel em 1950 como um método iterativo para resolver sistemas lineares com o coeficiente de definida positiva matrizes, que é bem adequado para a resolução de grande problemas. Já o primeiro método do gradiente conjugado não-linear foi introduzida por Fletcher e Reeves na década de 1960. Ao longo dos anos, muitas variantes deste regime inicial ter sido proposto, e apresentamos uma delas neste

trabalho, no próximo capítulo.

As principais características deste método é que são mais rápidos do que o método do Gradiente e por não necessitarem de armazenamento de matriz, o custo computacional torna-se menor do que os métodos de Newton. Outra característica deste algoritmo é que minimizam uma função quadrática definida em \mathbb{R}^n usando no máximo n passos. Porém, umas das propriedades notáveis deste método é a sua capacidade de gerar, de uma forma muito econômica, um conjunto de vetores com uma propriedade conhecida como direções conjugadas ou A-ortogonais. Diz-se que os vetores de um conjunto $\{d_1, d_2, \dots, d_n\}$ não nulos do \mathbb{R}^n com $m \leq n$ são A-ortogonais ou conjugados com respeito a matriz simétrica definida positiva A se

$$d_i^T A d_j = 0, \quad \forall i \neq j$$

Note que, no caso particular onde A é a matriz identidade, vetores A-conjugados são ortogonais no sentido usual. No caso geral, pode-se provar a independência linear de vetores A-conjugados, por exemplo veja [10].

- **O Algoritmo**

Algoritmo GC (Algoritmo do Gradiente Conjugado)

Inicialização. Dado $x_0 \in \mathbb{R}^n$, faça $d_0 = -\nabla f(x_0)$,

$k := 0$

Enquanto $\|g_k\| \neq 0$ fazer

1 $t_k = -\frac{\nabla f(x_k)^T d_k}{(d_k)^T A d_k}$

2 $x_{k+1} = x_k + t_k d_k$

3 $\beta_k = \frac{(d_k)^T A (\nabla f(x_{k+1}))}{(d_k)^T A d_k}$

4 $d_{k+1} = -\nabla f(x_{k+1}) + \beta_k (d_k)$. Faça $k = k + 1$.

- **Obtendo β_k**

Considere o seguinte problema

$$\min q(x) = \frac{1}{2}x^T Ax + b^T x + c, \quad x \in \mathbb{R}^n \quad (2.6)$$

onde $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica definida positiva, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Assim, dado $x^0 \in \mathbb{R}^n$, defina $d^0 = -\nabla f(x^0)$ e, para $k = 0, 1, \dots, n-2$,

$$d^{k+1} = -\nabla f(x^{k+1}) + \beta_k d^k,$$

onde $x^{k+1} = x^k + t_k d^k$, $t_k = -\frac{\nabla f(x^k)^T d^k}{(d^k)^T A d^k}$ e β_k calculado de modo que d^k e d^{k+1} sejam A-conjugadas:

$$(d^k)^T A(-\nabla f(x^{k+1}) + \beta_k d^k) = (d^k)^T A d^{k+1} = 0,$$

o que fornece

$$\beta_k = -\frac{(d^k)^T A(-\nabla f(x^{k+1}))}{(d^k)^T A d^k}$$

Outras formas de calcular este coeficiente foram posposto por Daniel, e Polak e Ribière e Fletcher e Reeves, respectivamente, é dada por

$$\beta_k = \frac{(d_k)^T A(\nabla f(x_{k+1}))}{(d_k)^T A d_k}$$

$$\beta_k^{PRP} = \frac{\nabla f(x_{k+1})^T (\nabla f(x_{k+1}) - \nabla f(x_k))}{\|\nabla f(x_k)\|^2}$$

$$\beta_k^{FR} = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}.$$

Tais expressões têm a vantagem computacional de utilizar apenas produto de vetores e coincidem no caso quadrático. Atualmente, a fórmula considerada mais eficiente a partir do ponto de vista da computação é o β_k^{PRP} [8,9].

Assim, a direção de busca do método do gradiente que será usada no Algoritmo GSZ terá a seguinte forma,

$$d_{k+1} = \begin{cases} -\nabla f(x_{k+1}) + \beta_k^{PRP} d_k, & \text{se } k \geq 1 \\ -\nabla f(x_k), & \text{se } k = 0 \end{cases} \quad (2.7)$$

Capítulo 3

Algoritmo de Yuan, Lu e Wei

Neste capítulo apresentamos o resultado principal deste trabalho, onde iniciamos descrevendo o esquema geral do Algoritmo GSZ abordado em [15]. Este algoritmo é proposto para resolver problemas do tipo (1.1), sob a perspectiva de herdar a simplicidade e o baixo custo computacional do método do gradiente conjugado. Em seguida, apresentamos na Seção 3.2 um estudo detalhado da análise de convergência do algoritmo. Além disso, veremos na Seção 3.3, que se a f for fortemente convexa, então teremos taxa de convergência linear.

3.1 Algoritmo GSZ

Nesta seção, por simplicidade, usaremos a notação ∇f_k para denotar o gradiente de f em x_k , isto é, $\nabla f_k = \nabla f(x_k)$.

Algoritmo GSZ

1: Escolha um ponto inicial $x_0 \in \mathbb{R}^n$ e constantes

$$0 < \varepsilon < 1, \quad 0 < \sigma_1 < \frac{1}{2}, \quad \sigma_1 < \sigma_2 < 1,$$

Tome $d_0 = -\nabla f(x_0)$, $k := 0$.

2: Se $\|\nabla f_k\| \leq \varepsilon$, PARE; caso contrário vá para o passo 3.

3: Calcule o comprimento de passo α_k através de busca linear. Faça $x_{k+1} = x_k + \alpha_k d_k$.

4: Se $\|\nabla f_{k+1}\| \leq \varepsilon$ PARE; caso contrário vá para o passo 5.

5: Calcule d_{k+1} ,

$$d_{k+1} = \begin{cases} -\nabla f_k + \beta_k^{PRP} d_k, & \text{se } k \geq 1 \\ -\nabla f_k, & \text{se } k = 0 \end{cases}$$

onde $\beta_k^{PRP} = \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\|\nabla f_k\|^2}$. Vá para 6.

6: Calcule a nova direção,

$$d_{k+1}^N = d'_{k+1} + \min \left\{ 0, \frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2} \right\} \nabla f_{k+1} - \nabla f_{k+1},$$

onde $d'_{k+1} = \frac{\|y_k^*\| \|\nabla f_{k+1}\|}{\|s_k\| \|d_{k+1}\|} d_{k+1}$, $s_k = x_{k+1} - x_k$, $y_k = \nabla f_{k+1} - \nabla f_k$ e $\|y_k^*\| = \max\{\|s_k\|, \|y_k\|\}$. Faça $k := k + 1$, e vá para o passo 3.

A fim de assegurar a convergência do Algoritmo GSZ a uma solução do problema (1.1), é necessário escolher não apenas passos apropriados, mas também direções adequadas. O Lema, a seguir, garante que as direções d_k obtidas pelo Algoritmo GSZ são direções de descida.

Para obtermos os resultados desejados, atribuímos duas condições necessárias à

função f :

(A 1) A função objetivo f é de classe C^1 e limitada inferiormente no conjunto de nível

$$L_0 = \{x \in \mathbb{R}^n \mid f(x) \leq f(x^0)\}$$

(A 2) O gradiente ∇f é Lipschitz contínuo em um conjunto aberto e convexo Ω contendo L_0 , isto é, existe $L > 0$ tal que

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega$$

Observação 3.1. As hipóteses (A1) e (A2) são padrões em análise de convergência.

Lema 3.1. Considere a sequência $\{x_k\}$ gerada pelo Algoritmo GSZ, independente de qualquer critério de busca linear. Suponha que a hipótese (A 2) seja satisfeita. Então existem números positivos c_1 e c_2 tais que

$$(i) \nabla f_k^T d_k \leq -c_1 \|\nabla f_k\|^2, \quad \forall k \geq 0, \quad c_1 \in (0, 1);$$

$$(ii) \|d_k\| \leq c_2 \|\nabla f_k\|, \quad \forall k \geq 0, \quad c_2 = 2\max\{1, L\} + 1, \quad \text{onde } L > 0 \text{ é a constante de Lipschitz.}$$

Demonstração. Para $k = 0$ são válidos (i) e (ii), trivialmente. Agora, suponhamos $k \geq 1$. Pelo passo 6 do Algoritmo GSZ, decorre que

$$d_{k+1}^N = d'_{k+1} + \min \left(0, -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}^T\|} \right) \nabla f_{k+1}^T - \nabla f_{k+1}^T. \quad (3.1)$$

Para provar (i), o produto $\nabla f_{k+1}^T d'_{k+1}$ deve ser considerado nos dois casos seguintes:

caso 1: Se $\nabla f_{k+1}^T d'_{k+1} \leq 0$ então $-\nabla f_{k+1}^T d'_{k+1} > 0$, daí (3.1) é reescrita como segue,

$$d_{k+1}^N = d'_{k+1} - \nabla f_{k+1}.$$

Tomando o produto interno com ∇f_{k+1} pela esquerda na igualdade acima temos,

$$\nabla f_{k+1}^T d_{k+1}^N = \nabla f_{k+1}^T d'_{k+1} - \|\nabla f_{k+1}\|^2 \leq -\|\nabla f_{k+1}\|^2. \quad (3.2)$$

caso 2: Se $\nabla f_{k+1}^T d'_{k+1} > 0$, então $\min\left(0, -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|}\right) = -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|}$. Neste caso, em (3.1), verifica-se que

$$d_{k+1}^N = d'_{k+1} - \frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2} \nabla f_{k+1} - \nabla f_{k+1} \quad (3.3)$$

De modo análogo, tomando o produto interno com ∇f_{k+1} pela esquerda em ambos os lados da igualdade (5.3) tem-se,

$$\nabla f_{k+1}^T d_{k+1}^N = \nabla f_{k+1}^T d'_{k+1} - \frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2} \|\nabla f_{k+1}\|^2 - \|\nabla f_{k+1}\|^2 = -\|\nabla f_{k+1}\|^2. \quad (3.4)$$

Em ambos os casos, observamos que $\nabla f_{k+1}^T d_{k+1}^N \leq -\|\nabla f_{k+1}\|^2$. Assim, tomando $c_1 \in (0, 1)$ chegamos ao resultado desejado, o *item (i)*. Agora, será feita a verificação do *item (ii)*, para tal, considere o passo 6 do Algoritmo GSZ, isto é,

$$d_{k+1}^N = d'_{k+1} + \min\left\{0, \frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2}\right\} \nabla f_{k+1} - \nabla f_{k+1}.$$

Tomando a norma e utilizando a Desigualdade Triangular na equação anterior obtemos,

$$\begin{aligned} \|d_{k+1}^N\| &= \left\| d'_{k+1} + \min\left\{0, -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2}\right\} \nabla f_{k+1} - \nabla f_{k+1} \right\| \\ &\leq \|d'_{k+1}\| + \left| \min\left\{0, -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2}\right\} \right| \|\nabla f_{k+1}\| + \|\nabla f_{k+1}\|. \end{aligned} \quad (3.5)$$

Decorre da definição de d'_{k+1} que

$$\begin{aligned} \|d'_{k+1}\| &= \frac{\|y_k^*\| \|\nabla f_{k+1}\|}{\|s_k\| \|d_{k+1}\|} \|d_{k+1}\| = \frac{\|y_k^*\| \|\nabla f_{k+1}\|}{\|s_k\|} \\ &= \frac{\max\{\|s_k\|, \|\nabla f_{k+1} - \nabla f_k\|\}}{\|s_k\|} \|\nabla f_{k+1}\|. \end{aligned} \quad (3.6)$$

Note que em (3.6), se $\max\{\|s_k\|, \|\nabla f_{k+1} - \nabla f_k\|\} = \|s_k\|$ então $\|d'_{k+1}\| = \|\nabla f_{k+1}\|$.

Agora, caso o máximo seja $\|\nabla f_{k+1} - \nabla f_k\|$, usando **(A 2)** segue que

$$\|d'_{k+1}\| \leq \frac{L\|s_k\|}{\|s_k\|} \|\nabla f_{k+1}\| = L\|\nabla f_{k+1}\|$$

Assim,

$$\|d'_{k+1}\| \leq \max\{1, L\} \|\nabla f_{k+1}\| \quad (3.7)$$

Então em (3.5), se tivermos $\min\left\{0, -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2}\right\} = 0$, ao combinarmos com (3.7) obteremos

$$\|d_{k+1}^N\| \leq \|d'_{k+1}\| + \|\nabla f_{k+1}\| \leq \max\{1, L\} \|\nabla f_{k+1}\| + \|\nabla f_{k+1}\| = (\max\{1, L\} + 1) \|\nabla f_{k+1}\|.$$

Ainda em (3.5), se o $\min = -\frac{\nabla f_{k+1}^T d'_{k+1}}{\|\nabla f_{k+1}\|^2}$, combinado com (3.7) tem-se que,

$$\begin{aligned} \|d_{k+1}^N\| &\leq \|d'_{k+1}\| + \frac{\|\nabla f_{k+1}\| \|d'_{k+1}\|}{\|\nabla f_{k+1}\|^2} \|\nabla f_{k+1}\| + \|\nabla f_{k+1}\| \\ &= 2\|d'_{k+1}\| + \|\nabla f_{k+1}\| = (2\max\{1, L\} + 1) \|\nabla f_{k+1}\| \end{aligned}$$

Portanto, basta tomarmos $c_2 = (2\max\{1, L\} + 1)$ e o passo 7 do Algoritmo GSZ para obtermos o resultado desejado.

□

A observação, a seguir, é uma consequência imediata do Lema 3.1. Tal propriedade terá um papel importante em nossa análise de convergência.

Observação 3.2. *Se existem constantes positivas c_1 e c_2 sob as mesmas condições do Lema 3.1, então existe $\tau > 0$ satisfazendo a hipótese direcional. Vejamos,*

Do item (i), segue que $-\nabla f_k^T d_K \geq c_1 \|\nabla f_k\|^2$ donde

$$-\frac{\nabla f_k^T d_k}{\|\nabla f_k\| \|d_k\|} \|d_k\| \geq c_1 \|\nabla f_k\|.$$

Por (ii) temos que $\frac{\|\nabla f_k\|}{\|d_k\|} \geq \frac{1}{c_2}$. Então utilizando este resultado na equação acima concluímos que

$$-\frac{\nabla f_k^T d_k}{\|\nabla f_k\| \|d_k\|} \geq c_1 \frac{\|\nabla f_k\|}{\|d_k\|} = \frac{c_1}{c_2}, \quad (3.8)$$

Então, para todo k , fazendo $\tau = \frac{c_1}{c_2} > 0$, chegamos ao nosso resultado,

$$-\frac{\nabla f_k^T d_k}{\|\nabla f_k\| \|d_k\|} \geq \tau > 0, \quad \forall k \in \mathbb{N}. \quad (3.9)$$

3.2 Convergência global do Método GSZ

Nesta seção faremos um estudo detalhado do teorema de convergência global do Algoritmo GSZ, seguindo a proposta dos autores, os chineses Goglin Yuan, Sha Lu e Zengxin Wei em [15]. Para efeito didático discutimos suas variações segundo os critérios de Busca Linear Exata, Goldstein e Wolfe.

3.2.1 Algoritmo GSZ com Busca Linear Exata

Nesta subseção apresentamos o Algoritmo GSZ satisfazendo a condição de busca linear exata, ou seja, buscamos um comprimento do passo α_k , o qual sempre existe, satisfazendo a seguinte condição

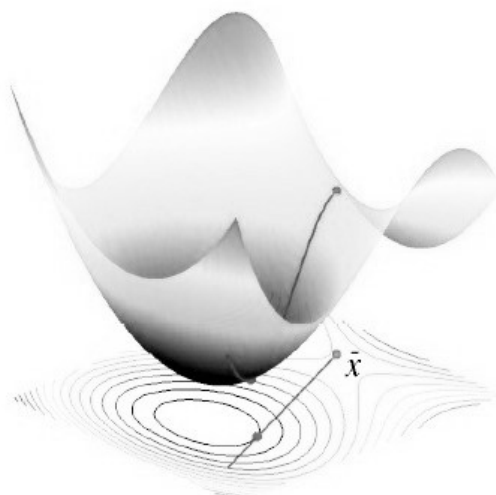


Figura 3.1: Busca linear exata.

$$f(x_k + \alpha_k d_k) = \min_{\alpha \geq 0} f(x_k + \alpha d_k) \quad (3.10)$$

isto é, α_k é uma solução do problema

$$\varphi(\alpha_k) = \min_{\alpha > 0} \varphi(\alpha)$$

onde $\varphi(\alpha_k) : \mathbb{R}_+ \rightarrow \mathbb{R}$, $\varphi(\alpha_k) = f(x_k + \alpha_k d_k)$.

Em seguida, mostraremos que a sequência gerada pelo algoritmo Algoritmo GSZ com busca Linear Exata, converge para o minimizador da função objetivo, caso exista algum.

Teorema 3.1. (Teorema de convergência global) *Suponhamos que f satisfaça as hipótese (A 1) e (A 2). Além disso, suponhamos também que $x_k \in \mathbb{R}^n$ e $d_k \in \mathbb{R}^n$ estão sob as condições do Lema 3.1. Assumimos que α_k é definido pelo critério de busca linear exata (3.10). Então as sequências $\{\alpha_k\}$, $\{d_k\}$, $\{x_{k+1}\}$, $\{\nabla f_{k+1}\}$ geradas pelo Algoritmo GSZ satisfazem*

$$\lim_{k \rightarrow \infty} \|\nabla f_k\| = 0. \quad (3.11)$$

Demonstração. Consideremos uma iteração de índice k e que α_k é a solução de (3.10). Tomemos ainda, um α_k^* qualquer

$$\frac{1}{5L} \frac{|\nabla f_k^T d_k|}{\|d_k\|^2} \leq \alpha_k^* \leq \frac{2}{5L} \frac{|\nabla f_k^T d_k|}{\|d_k\|^2} \quad (3.12)$$

Pela definição de minimizador tem-se que $f(x_k + \alpha_k d_k) \leq f(x_k + \alpha_k^* d_k)$, pelo Teorema do Valor Médio com resto integral [x], existe um $t \in (0, 1)$ de tal modo que

$$f(x_k + \alpha_k d_k) - f(x_k) \leq f(x_k + \alpha_k^* d_k) - f(x_k) = \int_0^1 \nabla f(x_k + t\alpha_k^* d_k)^T (\alpha_k^* d_k) dt$$

Ao subtrairmos $\alpha_k^* \nabla f(x_k)^T d_k$ em ambos os lados da equação acima, usando a desigualdade de Cauchy-Schwartz, o fato de que o gradiente é Lipschitz contínuo com constante L , temos

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) - \alpha_k^* \nabla f(x_k)^T d_k &\leq \int_0^1 [\nabla f(x_k + t\alpha_k^* d_k) - \nabla f(x_k)^T d]^T (\alpha_k^* d_k) dt \\ &\leq \int_0^1 \|[\nabla f(x_k + t\alpha_k^* d_k) - \nabla f(x_k)]\| \|(\alpha_k^* d_k)\| dt \\ &\leq |\alpha_k^*|^2 \|d_k\|^2 L \int_0^1 t dt \\ &= \frac{1}{2} L (\alpha_k^*)^2 \|d_k\|^2. \end{aligned} \quad (3.13)$$

Podemos escrever esta relação como

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \alpha_k^* \nabla f_k^T d_k + \frac{1}{2} L (\alpha_k^*)^2 \|d_k\|^2 \quad (3.14)$$

Notamos que, se d_k é uma direção tal que $\nabla f(x_k)^T d_k < 0$, pela esquerda de (3.12) usando propriedade de módulo e algumas manipulações algébricas obtemos

$$\alpha_k^* \nabla f(x_k)^T d_k \leq \frac{1}{5L} \frac{|\nabla f_k^T d_k|}{\|d_k\|^2} \nabla f(x_k)^T d_k = -\frac{1}{5L} \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \quad (3.15)$$

Por outro lado, em (3.12), ao elevarmos (α_k^*) ao quadrado e multiplicarmos toda a

inequação pelo termo $(L/2\|d_k\|^2)$ obtemos

$$\frac{L}{2}(\alpha_k^*)^2\|d_k\|^2 \leq \frac{2}{25L} \frac{|\nabla f_k^T d_k|^2}{\|d_k\|^2} = \frac{2}{25L} \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \quad (3.16)$$

pela direita, onde consideramos a propriedade de módulo e o fato que $(\nabla f_k^T d_k)^2 = (-\nabla f_k^T d_k)^2$ para obtermos a igualdade. Combinando (5.14), (5.15) e (5.16) temos

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq -\frac{1}{5L} \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 + \frac{2}{25L} \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \\ &= -\frac{3}{25L} \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \end{aligned}$$

Em particular, isto mostra que $f(x_k + \alpha_k d_k) \leq f(x_k)$ e como k foi arbitrário, concluímos que a sequência $\{f(x_k)\}$ é monótona decrescente para todo $k \in \mathbb{N}$. Podemos escrever a relação obtida acima do seguinte modo

$$f(x_k) - f(x_{k+1}) \geq \delta_0 \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \quad \forall k. \quad (3.17)$$

onde $\delta_0 = \frac{3}{25L}$ é uma constante positiva que é independente de k .

Em seguida, consideramos a soma dos k primeiros termos desta sequência, obtemos

$$f(x_0) - f(x_k) = \sum_{j=0}^{k-1} [f(x_j) - f(x_{j+1})] \geq \delta_0 \sum_{j=0}^{k-1} \left(\frac{-\nabla f_j^T d_j}{\|d_j\|} \right)^2 \quad (3.18)$$

Sabemos que f_k é monótona, além disso, é limitada inferiormente, então $\{f(x_k)\}$ converge para o ínfimo. Logo, usando o critério de comparação de série, ao fazermos $k \rightarrow \infty$ segue que

$$\delta_0 \sum_{j=0}^{k-1} \left(\frac{-\nabla f_j^T d_j}{\|d_j\|} \right)^2 \rightarrow 0,$$

ou seja, temos que

$$\sum_{k=0}^{\infty} \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 < +\infty. \quad (3.19)$$

Observamos que a condição (3.18) implica, pela condição necessária de convergência de séries que $\lim_{k \rightarrow \infty} \left(-\frac{\nabla f_k^T d_k}{\|d_k\|} \right)^2 = 0$. Logo,

$$0 = \lim_{k \rightarrow \infty} \frac{(-\nabla f_k^T d_k)^2}{\|d_k\|^2} = \lim_{k \rightarrow \infty} \left(-\frac{\nabla f_k^T d_k}{\|d_k\|} \right)^2 \geq \lim_{k \rightarrow \infty} (\tau \|\nabla f_k\|)^2 \geq \left(\lim_{k \rightarrow \infty} \tau \|\nabla f_k\| \right)^2 \quad (3.20)$$

onde a primeira desigualdade vem mediante uso do Lema 3.1. Ora, por (3.9) temos que $\tau > 0$ e portanto, usando também a propriedade de norma, segue que

$$\lim_{k \rightarrow \infty} \|\nabla f_k\| = 0.$$

Assim, se \bar{x} é um ponto de acumulação de $\{x_k\}$, a continuidade de ∇f implica que $\nabla f(\bar{x}) = 0$. Resulta que \bar{x} é minimizador de f . Isto é, todos seus pontos de acumulação são soluções de (1.1). \square

3.2.2 Algoritmo GSZ com o critério de Goldstein

Nesta subsecção apresentamos o caso em que o Algoritmo GSZ utiliza o comprimento de passo α_k calculado pelo critério de Goldstein, que consiste em encontrar um comprimento do passo α_k que satisfaça às seguintes desigualdades,

$$\sigma_1 \leq \frac{f(x_k + \alpha_k d_k) - f(x_k)}{\alpha \nabla f(x_k)^T d_k} \leq \sigma_2, \quad (3.21)$$

onde $x_k \in \mathbb{R}^n$, $d_k \in \mathbb{R}^n$ é uma direção de descida, e $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$ são os parâmetros dados. Equivalentemente, os critérios de Goldstein podem ser expressas como:

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k \\ f(x_k + \alpha_k d_k) &\geq f(x_k) + \sigma_2 \alpha_k \nabla f(x_k)^T d_k \end{aligned} \quad (3.22)$$

A primeira desigualdade é a de Armijo, i.e., (2.5), ela garante o decréscimo suficiente da função objetivo. A desigualdade (3.22), conhecida como condição de curvatura, é introduzida por este critério para evitar passos curtos demais.

Em seguida, mostraremos que a sequência gerada pelo algoritmo Algoritmo GSZ com critério de Goldstein, converge para o minimizador da função objetivo.

Teorema 3.2. (Teorema de convergência global)

Suponhamos que f satisfaz as hipóteses (A 1) e (A 2). Consideramos uma direção de busca $d_k \in \mathbb{R}^n$ que satisfaçam o Lema 3.1 e o α_k obtido pelo critério de Goldstein, 3.21. Então as sequências $\{\alpha_k\}$, $\{d_k\}$, $\{x_{k+1}\}$, $\{\nabla f_{k+1}\}$ geradas pelo Algoritmo GSZ satisfazem

$$\lim_{k \rightarrow \infty} \|\nabla f_k\| = 0. \quad (3.23)$$

Demonstração. Para todo $k \in \mathbb{N}$ temos, de (3.22), que

$$\alpha_k \sigma_2 \nabla f(x_k)^T d_k \leq f(x_k + \alpha_k d_k) - f(x_k). \quad (3.24)$$

Agora, pela diferenciabilidade de f e usando o Teorema do Valor Médio com resto integral (ref) obtemos

$$f(x_k + \alpha_k d_k) - f(x_k) - \alpha_k \nabla f(x_k)^T d_k = \int_0^1 [\nabla f(x_k + \theta_k \alpha_k d_k) - \nabla f(x_k)]^T (\alpha_k d_k) d\theta_k$$

onde $\theta_k \in (0, 1)$. Agora, os argumentos são semelhantes aos usados na prova do Teorema 4.1 na (p.33), pela desigualdade de Cauchy-Schwartz e por (A 2), verificamos que

$$\begin{aligned} f(x_k + \alpha_k d_k) - f(x_k) &\leq \alpha_k \nabla f(x_k)^T d_k + \int_0^1 [\nabla f(x_k + \theta_k \alpha_k d_k) - \nabla f(x_k)]^T (\alpha_k d_k) d\theta_k \\ &\leq \alpha_k \nabla f(x_k)^T d_k + \frac{L}{2} \alpha_k^2 \|d_k\|^2 \end{aligned} \quad (3.25)$$

Combinando (3.24) e (3.25), obtemos

$$\sigma_2 \alpha_k \nabla f(x_k)^T d_k \leq \alpha_k \nabla f(x_k)^T d_k + \frac{L}{2} L \alpha_k^2 \|d_k\|^2,$$

ou equivalentemente,

$$(\sigma_2 - 1) \alpha_k \nabla f(x_k)^T d_k \leq \frac{L}{2} L \alpha_k^2 \|d_k\|^2.$$

Dividindo ambos os lados da inequação acima por $\alpha_k > 0$, usando o fato que d_k é uma direção de descida e $-(1 - \sigma_2) = (\sigma_2 - 1)$, obtemos

$$\alpha_k \geq \frac{2(1 - \sigma_2) (-\nabla f(x_k)^T d_k)}{L \|d_k\|^2} > 0,$$

com $\sigma_1 < \sigma_2 < 1$. Daí, ao multiplicarmos ambos os lados da desigualdade acima por $\sigma_1 (-\nabla f_k^T d_k) > 0$ segue que

$$-\sigma_1 \alpha_k \nabla f_k^T d_k \geq \frac{2\sigma_1(1 - \sigma_2) (-\nabla f(x_k)^T d_k)}{L \|d_k\|^2} (-\nabla f(x_k)^T d_k).$$

E pela hipótese (5.21), temos que

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \frac{2\sigma_1(1 - \sigma_2)}{L} \left(-\frac{\nabla f(x_k)^T d_k}{\|d_k\|} \right)^2.$$

Fazendo $\delta_1 = \frac{2\sigma_1(1 - \sigma_2)}{L} > 0$, podemos reescrever esta inequação da seguinte forma,

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \delta_1 \left(\frac{\nabla f(x_k)^T d_k}{\|d_k\|^2} \right)^2,$$

ou seja, a sequência $\{f(x_k)\}$ é monótona decrescente. Além disso, é limitada inferiormente, portanto convergente. Como já foi visto na Seção 3.1, ao somarmos a expressão

acima para todos os índices menores ou iguais a k , obteremos que

$$f(x_0) - f(x_k) \geq \delta_1 \sum_{j=0}^{k-1} \left(\frac{-\nabla f(x_j)^T d_j}{\|d_j\|} \right)^2$$

Assim, o termo geral desta série será convergente devido o lado esquerdo da inequação acima ser convergente, onde isso implica que

$$\lim_{k \rightarrow \infty} \left(\frac{\nabla f(x_k)^T d_k}{\|d_k\|} \right)^2 = 0.$$

Consequentemente teremos (3.23), ao usarmos o corolário Lema 3.1.

□

3.2.3 Algoritmo GSZ com o critério de Wolf-Powell

Nesta subseção, apresentamos o Algoritmo GSZ baseado nos critérios fraco e forte de Wolf-Powell, nos referimos a este critério como (WWP) e (SWP), respectivamente. Dados um ponto $x_k \in \mathbb{R}^n$, uma direção de descida $d_k \in \mathbb{R}^n$, e parâmetros σ_1 e σ_2 tais que $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$ queremos encontrar um comprimento do passo α_k que satisfaça, simultaneamente, às condições

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k \\ \nabla f(x_k + \alpha_k d_k)^T d_k &\geq \sigma_2 \nabla f(x_k)^T d_k \end{aligned} \quad (3.26)$$

As condições de decréscimo suficiente (Armijo, 2.5) para a função objetivo e de curvatura (3.26) são conhecidas como critério fraco de Wolfe-Powell (**WWP**).

O critério de Wolfe-Powell, denominadas fortes (**SWP**), requerem que α_k satis-

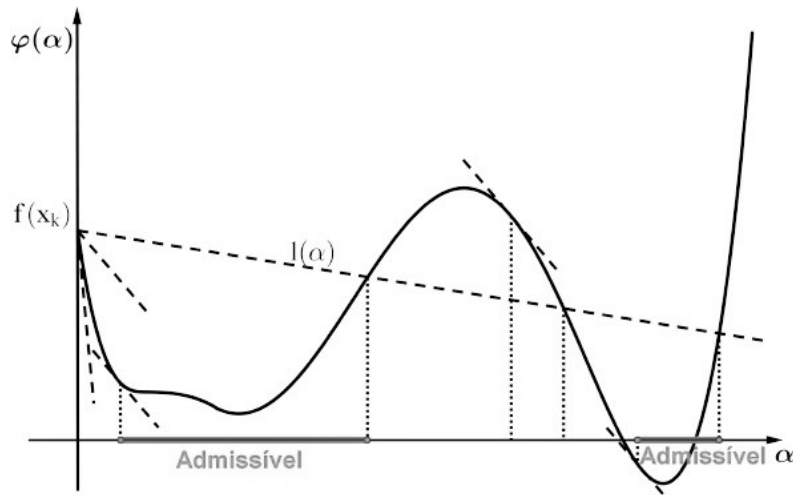


Figura 3.2: Conjunto admissível pelas condições WWP

faça simultaneamente às desigualdades

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + \sigma_1 \alpha_k \nabla f(x_k)^T d_k \\ |\nabla f(x_k + \alpha_k d_k)^T d_k| &\leq \sigma_2 |\nabla f(x_k)^T d_k| \end{aligned} \quad (3.27)$$

onde $0 < \sigma_1 < \frac{1}{2}$, $\sigma_1 < \sigma_2 < 1$. Com a condição de Armijo dada na primeira desigualdade, o critério tende a eliminar os comprimentos de passos que estão próximos aos máximos locais, embora isto nem sempre aconteça. A representação das condições fortes de Wolfe é dada na Figura (3.3). A prova de que existem tamanho de passos que satisfaçam as condições de Wolfe pode ser encontrada em Nocedal [8].

Em seguida, mostraremos que a sequência gerada pelo algoritmo Algoritmo GSZ com critério de Wolfe-Powell, converge para o minimizador da função objetivo. Resaltamos que o enfoque da demonstração do Teorema a seguir será para o caso em que o comprimento de passo α_k é obtido através do critério (SWP), uma vez que o caso (WWP) demonstra-se de forma análoga.

Teorema 3.3. (Teorema de convergência global)

Suponhamos que f satisfaça as hipóteses (A 1) e (A 2) e que d_k satisfaça as as

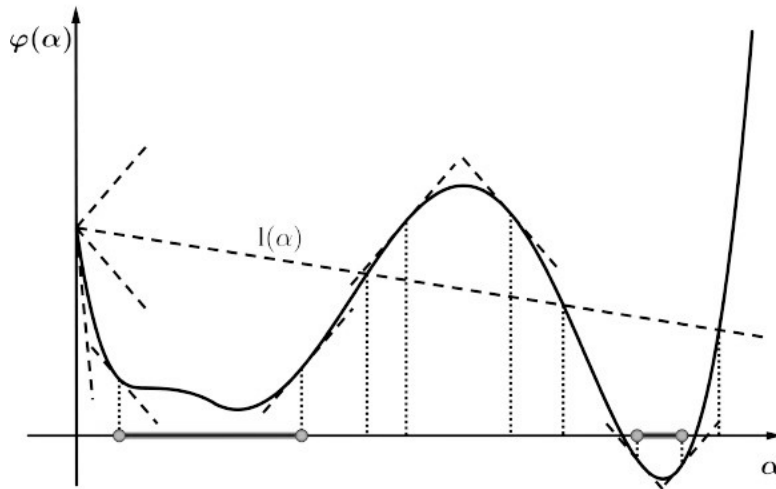


Figura 3.3: Conjunto admissível pelas condições de SWP

hipóteses direcionais do Lema 3.1. Assumimos que α_k é definido por um dos critérios de Wolfe-Powell, (SWP) ou (WWP). Então as sequências $\{\alpha_k\}$, $\{d_k\}$, $\{x_{k+1}\}$, $\{\nabla f_{k+1}\}$ geradas pelo pelo Algoritmo GSZ satisfazem

$$\lim_{k \rightarrow \infty} \|\nabla f_k\| = 0. \quad (3.28)$$

Demonstração. Consideremos que o comprimento do passo α_k está sendo calculado pelo critério (SWP). Da segunda condição de (SWP), (3.27) e usando o fato de d_k ser uma direção de descida, temos que

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq -|\nabla f(x_k + \alpha_k d_k)^T d_k| \geq \sigma_2 \nabla f(x_k)^T d_k. \quad (3.29)$$

Adicionando o termo $-\nabla f(x_k)^T d_k$ em ambos os membros da desigualdade acima obtemos

$$(\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k))^T d_k \geq (\sigma_2 - 1) (\nabla f(x_k)^T d_k) = (1 - \sigma_2)(-\nabla f(x_k)^T d_k).$$

Enquanto que a condição Lipschitz (A 2) e a desigualdade de Schwartz implicam, para

todo $k \in \mathbb{N}$, que

$$\begin{aligned} (\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k))^T d_k &\leq |(\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k))^T d_k| \\ &\leq \|[\nabla f(x_k + \alpha_k d_k) - \nabla f(x_k)]\| \|d_k\| \\ &\leq L\alpha_k \|d_k\|^2 \end{aligned}$$

Combinando as duas relações acima, obtemos:

$$\alpha_k \geq \frac{(1 - \sigma_2)}{L} \left(-\frac{\nabla f(x_k)^T d_k}{\|d_k\|^2} \right) > 0.$$

Substituindo esta desigualdade no primeiro critério de Goldstein, (2.5), verificamos

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \sigma_1 \alpha_k (-\nabla f(x_k)^T d_k) \geq \delta_2 \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2,$$

onde $\delta_2 = \frac{\sigma_1(1 - \sigma_2)}{L} > 0$. E a partir daqui, prosseguimos de maneira similar à prova do Teorema 4.1. E assim obtemos o resultado desejado, isto é, (3.28).

□

Em suma, temos que a sequência gerada pelo Algoritmo GSZ independentemente do critério de busca linear (Exata, Goldstein ou Wolfe-Powell) convergirá a uma solução do problema (1.1). À vista disso, apresentamos a taxa de convergência do algoritmo na próxima seção.

3.3 Taxa de convergência

Nesta seção apresentamos a taxa de convergência do Algoritmo GSZ, baseado em [11], reescrito com detalhes adicionais. Salientamos que o enfoque das demonstrações nesta subseção serão as funções fortemente convexas. Para isto, a função f deve

satisfazer, além de **(A 1)** e **(A 2)**, a seguinte condição:

(A 3) f é uma função de classe \mathcal{C}^2 e fortemente convexa em $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto.

Apresentamos agora, um resultado auxiliar que será essencial na prova do resultado principal, Teorema 3.4.

Lema 3.2. *Suponha que f satisfaz **(A 1)**, **(A 2)** e **(A 3)** e existem $0 < m < M$ tal que*

$$m\|d\|^2 \leq d^T \nabla^2 f(x) d \leq M\|d\|^2, \quad \forall x, d \in \Omega^n. \quad (3.30)$$

Então f tem um único minimizador x^* , e as seguintes afirmações são verdadeiras:

$$(i) \quad \frac{1}{2}m\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2}M\|x - x^*\|^2, \quad \forall x, y \in \Omega$$

$$(ii) \quad M\|x - y\|^2 \geq (\nabla f(x) - \nabla f(y))^T(x - y) \geq m\|x - y\|^2, \quad \forall x, y \in \Omega$$

$$(iii) \quad M\|x - x^*\|^2 \geq \nabla f(x)^T(x - x^*) \geq m\|x - x^*\|^2, \quad \forall x \in \Omega$$

$$(iv) \quad M\|x - x^*\| \geq \|\nabla f(x)\| \geq m\|x - x^*\| \quad \forall x \in \Omega$$

Demonstração. Pelo Teorema 2.11 (b) temos

$$\frac{1}{2}m\|x - x^*\|^2 \leq m\|x - x^*\|^2 + \langle f'(x^*), x - x^* \rangle \leq f(x) - f(x^*)$$

Por outro lado, fazendo $y = x^* + d$ no Teorema 2.3 segue que

$$f(y) - f(x^*) = \nabla f(x^*)^T d + \frac{1}{2}(y - x^*)^T \nabla^2 f(x^* + ty - x^*)^T (y - x^*)$$

Como $y - x^* \in \Omega$ e $x^* + t(y - x^*) \in \Omega$ usando a desigualdade do lado direito de (3.30)

$$m\|y - x^*\|^2 \leq 2m\|x - y\|^2 \leq (\nabla f(x) - \nabla f(y))^T(x - y) \leq \|\nabla f(x) - \nabla f(y)\|\|x - y\| \leq M\|x - y\|^2$$

E assim segue a afirmação (i). Agora, verificaremos o item (ii). Pelo Teorema 2.11 (c), temos que ∇f é uma função uniformemente monótona, logo,

$$\begin{aligned} m\|x - y\|^2 &\leq (\nabla f(x) - \nabla f(y))^T(x - y) \leq |(\nabla f(x) - \nabla f(y))^T(x - y)| \\ &\leq \|\nabla f(x) - \nabla f(y)\| \|x - y\| \\ &\leq M\|x - y\| \|x - y\| = M\|x - y\|^2. \end{aligned} \quad (3.31)$$

onde usamos a desigualdade de Cauchy-Schwartz, a hipótese **(A 2)** e o item (3.30) para obtermos o lado direito da desigualdade acima. Ao considerarmos na equação (3.31), $y = x^*$, no qual x^* é minimizador local de f , teremos de imediato o item (iii),

$$M\|x - x^*\|^2 \geq \nabla f(x)^T(x - x^*) \geq m\|x - x^*\|^2, \quad \forall x \in \Omega \quad (3.32)$$

Além disso, se aplicarmos a desigualdade de Cauchy-Schwartz na equação acima, obtemos

$$\begin{aligned} m\|x - x^*\|^2 &\leq \nabla f(x)^T(x - x^*) \leq |\nabla f(x)^T(x - x^*)| \\ &\leq \|\nabla f(x)\| \|x - x^*\| \leq M\|x - x^*\|^2, \quad \forall x \in \Omega \end{aligned}$$

e dividindo por $\|x - x^*\|^2 > 0$, temos que

$$M\|x - x^*\| \geq \|\nabla f(x)\| \geq m\|x - x^*\|, \quad \forall x \in \Omega$$

e assim chegamos à quinta relação, i.e, ao item (iv).

□

Demonstraremos a seguir o principal resultado desta seção, a convergência linear do Algoritmo GSZ.

Teorema 3.4. *Suponha que valem (A 3), (A 2) e (A 1). Sejam $\{\alpha_k\}$, $\{d_k\}$, $\{x_{k+1}\}$ e $\{\nabla f_k\}$ sequências geradas pelo Algoritmo GSZ sob qualquer busca linear (exata, Goldstein, SWP ou WWP). Então a sequência infinita $\{x_k\}$ converge linearmente para x^* , onde x^* único ponto estacionário de f .*

Demonstração. A partir da iteração $x_{k+1} = x_k + \alpha_k d_k$ e sob qualquer busca linear, vimos na Seção 3.2 que existe um número positivo δ tal que

$$f_k - f_{k+1} \geq \delta \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \quad \forall k. \quad (3.33)$$

Manipulando algebricamente o lado esquerdo da desigualdade acima, obtemos

$$f_k - f_{k+1} \geq \delta \left(\frac{-\nabla f_k^T d_k}{\|d_k\|} \right)^2 \frac{\|\nabla f_k\|^2}{\|\nabla f_k\|^2} = \delta \left(\frac{-\nabla f_k^T d_k}{\|\nabla f_k\| \|d_k\|} \right)^2 \|\nabla f_k\|^2 .$$

Assim, temos que

$$f_k - f_{k+1} \geq \delta \tau^2 \|\nabla f_k\|^2 \geq \delta \tau^2 m^2 \|x_k - x^*\|^2, \quad \forall k, \quad (3.34)$$

onde usamos o Lema 3.1 para obtermos a primeira desigualdade e o item (iv) do Lema 3.2 para obtermos a segunda desigualdade.

Por (A 3), existe um único minimizador da função f , digamos x^* . Assim, por definição $f(x^*) < f(x_k)$, $\forall x_k \in \Omega$. Logo, pelo lado direito do Lema 3.2 item (i) temos

$$f_k - f(x^*) \leq \frac{M}{2} \|x_k - x^*\|^2 \quad \forall x_k \in \Omega.$$

Sendo $M > 0$, podemos reescrever a desigualdade acima como segue,

$$\frac{2}{M} (f_k - f(x^*)) \leq \|x_k - x^*\|^2.$$

Agora usando o fato que $0 < m < M$, obtemos

$$\frac{2m^2}{M}(f_k - f(x^*)) \leq m^2 \|x_k - x^*\|^2. \quad (3.35)$$

Assim, combinando (3.34)-(3.35), segue que

$$f_k - f_{k+1} \geq \delta\tau^2 m^2 \|x_k - x^*\|^2 \geq \delta\tau^2 \frac{2m^2}{M}(f_k - f(x^*)). \quad (3.36)$$

Por (3.33) e **(A 1)** temos que $\{f(x_k)\}$ é monótona e limitada inferiormente. Tomando o limite da expressão acima, concluímos que

$$f_k - f_{k+1} \longrightarrow 0 \quad (k \longrightarrow \infty)$$

Portanto, pela continuidade da f , $\{x_k\}$ converge para x^* .

Agora, verificaremos que a taxa de convergência da sequência $\{x_k\}$ é linear. Para isto, definamos

$$\rho = \sqrt{2 \frac{\delta\tau^2 m^2}{M}} \quad (3.37)$$

$$\theta = \sqrt{1 - \rho^2} \quad (3.38)$$

Reescrevendo (3.37), obtemos

$$f_k - f_{k+1} \geq \rho^2 (f_k - f(x^*)). \quad (3.39)$$

Então usando (3.39) e o Lema 3.2 provaremos que $\rho \leq 1$. De fato, segue a partir do item (ii) do Lema 3.2 que

$$(\nabla f(x) - \nabla f(y))^T (x - y) \geq m \|x - y\|^2 \quad (3.40)$$

Para todo $x \in \mathbb{R}^n$, vale

$$\begin{aligned} (\nabla f(x) - \nabla f(y))^T(x - y) &\leq \|(\nabla f(x) - \nabla f(y))^T(x - y)\| \\ &\leq \|\nabla f(x) - \nabla f(y)\| \|x - y\| \leq L\|x - y\|^2, \end{aligned} \quad (3.41)$$

onde usamos Cauchy-Schwarz e a hipóteses **(A 2)** e **(A 1)** obtemos as duas últimas desigualdades. Assim, por (3.41)-(3.42), obtemos

$$L\|x - y\|^2 \geq m\|x - y\|^2,$$

onde L é a constante de Lipchitz. Isto implica que $\frac{1}{m} \geq \frac{1}{L}$. Vamos considerar agora, $\delta = \delta_0 = \frac{3}{25L}$ da equação 3.17. Então, por (3.37) temos,

$$\rho^2 = \frac{6}{25L} \frac{\tau^2 m^2}{M} \leq \frac{6}{25} \frac{\tau^2 m}{M} \leq 1,$$

Para os demais casos temos

$$\delta = \frac{\sigma_1(1 - \sigma_2)}{L},$$

onde $0 < \sigma_1 < \frac{1}{2}$ e $\sigma_1 < \sigma_2 < 1$. Então, por (3.37) obtemos,

$$\rho^2 = 2\sigma_1(1 - \sigma_2) \frac{\tau^2 m^2}{LM} \leq 2\sigma_1(1 - \sigma_2) \frac{\tau^2 m}{M} \leq 2\sigma_1(1 - \sigma_2) < 1.$$

Por (3.40), temos para todo k

$$\begin{aligned} f_{k+1} &\leq f_k - \rho^2(f(x) - f(x^*)) \\ f_{k+1} - f(x^*) &\leq f_k - f(x^*) - \rho^2 f(x) + \rho^2 f(x^*) \\ &= f_k(1 - \rho^2) - f(x^*)(1 - \rho^2) \\ &\leq (1 - \rho^2)(f_k - f(x^*)). \end{aligned}$$

Fazendo $k \geq 1$, segue que

$$f_k - f(x^*) \leq (1 - \rho^2)(f_{k-1} - f(x^*)) = \theta^2(f_{k-1} - f(x^*)).$$

Note que,

$$\begin{aligned} f_0 - f(x^*) &= \theta^2 \\ f_1 - f(x^*) &= \theta^2(f_0 - f(x^*)) = \theta^2\theta^2 = \theta^4 \\ f_2 - f(x^*) &= \theta^2(f_1 - f(x^*)) = \theta^2\theta^4 = \theta^6 \\ &\vdots \\ &= \theta^{2k}(f_0 - f(x^*)). \end{aligned}$$

Logo,

$$f_k - f(x^*) \leq \theta^{2k}(f_0 - f(x^*)).$$

Assim pelo Item (i) do Lema 3.2 e fazendo $\omega = \sqrt{\frac{2(f_0 - f(x^*))}{m}}$, segue-se que

$$\|x_k - x^*\| \leq \frac{2}{m}(f_k - f(x^*)) \leq \frac{2(f_0 - f(x^*))}{m}\theta^{2k} = \omega^2\theta^{2k}.$$

O que implica que $\|x_k - x^*\| \leq \omega^2\theta^{2k}$. Portanto, $\{x_k\}$ converge para x^* linearmente. □

Capítulo 4

Considerações Finais

Neste trabalho concentramos nossos estudos no Algoritmo GSZ, um algoritmo de primeira ordem (usam apenas informações da função e do gradiente). O embasamento teórico para entender este algoritmo, assim como seus respectivos resultados de análise de convergência e taxa de convergência, foi obtido através de um estudo categórico do método dos gradientes conjugados (GC). Assim, verificamos que o Algoritmo GSZ, de fato, não desperdiça informações, pois os cálculos necessários para computar os sucessivos iterandos são simples e sem requerimentos de muita memória, tornando-o computacionalmente "barato". Embora tenha sido inspirado no método (GC), seu grande diferencial reside em duas coisas: no controle da direção d_k garantida pelo Lema 3.1, e na convergência global que é dada independentemente da busca linear (Exata, Goldstein, Wolf-Powell fraca e forte) utilizada.

Terminamos este trabalho ressaltando que o algoritmo aqui apresentado tem boas propriedades de convergência sob hipóteses razoáveis. Como intenções futuras, pretendemos realizar experimentos computacionais para comparar a sua eficiência (velocidade, precisão de resultados) em relação a outros métodos.

Referências Bibliográficas

- [1] Bertsekas, D.P. , Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York, 1982.
- [2] Bertsekas, D.P. , Nonlinear Programming, Athena Belmont, 1999 (segunda edição)
- [3] Gilbert, J.C., Nocedal,J., Numerical optimization (Global Convergence Properties of Conjugate Gradient Method for Optimization, SIAM Journal on Optimization, Vol. 2, No 1, 1992, pp. 21-42
- [4] Izmailov, A.,Solodov, M.,Otimização volume 1. Condições de Otimalidade, elementos de análise convexa e de dualidade. 2ed.Rio de Janeiro:IMPA, 2009.
- [5] Izmailov, A.,Solodov, M., Otimização volume 2. Métodos computacionais. Rio de Janeiro: IMPA, 2007.
- [6] Lima, Elon Lages, Espaços métricos, 5. ed. Rio de Janeiro: IMPA,2013.
- [7] Lima, Elon Lages, Curso de análise vol.2, 11. ed. Rio de Janeiro: IMPA, 2009.
- [8] Nocedal,J.,Wright, Stephen J., Numerical optimization (Springer Series in Operations Research and Financial Engineering) 2 ed., Springer.
- [9] Borges, Paulo R.T., Aspectos sobre a aplicação de Métodos de Gradientes Conjugados em Otimização Irrestrita, Rio de Janeiro, 1985.
- [10] Ribeiro, A.R. e Karas,E.W. , Otimização Contínua: Aspectos Teóricos e Computacionais,Ed. Cengage Learning Curitiba, 2012.

- [11] Shi, Zhen-Jun, "Convergence of line search methods for unconstrained optimization", Applied Mathematics and Computation, n. 157, pp. 393-405, 2004.
- [12] Sun, W., Yuan. YX., Optimization Theory and Methods: Nonlinear Programming. Springer Optimization and Its Applications, Vol I, 2006.
- [13] Y. Dai and Y. Yun., "A Nonlinear Conjugate Gradient with a Strong Convergence Properties", Journal of Software Engineering and Applications, 2010, 3: 503- 509 Scientific Research.
- [14] Y. Dai and Y. Yun., Nonlinear Conjugate Gradient Methods, Shanghai Science and Technology Press, 2000.
- [15] Yuan. G., Lu, S., Wei, Z., "A Line Search Algorithm for Unconstrained Optimization", Journal of Software Engineering and Applications, 2010, 3: 503- 509 Scientific Research.
- [16] Yuan. G., Lu, X., "A Modified PPR Conjugate Gradient Methods", Annals Research, Vol 166, No. 1, 2009, pp. 73 - 90.
- [17] Wei, Z. Li, G., Qi, L., "New Nonlinear Conjugate Gradient formulas Large-Scale Unconstrained Optimization Problems", Applied Mathematics and Computation, Vol. 179, No. 2, 2006, pp. 407 - 430.