

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

*Método do Ponto Proximal Inexato e uma Técnica de Busca
Linear Não Monótona para Otimização Irrestrita*

Suellen Paulino Lima

Manaus

2015

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

Suellen Paulino Lima

Método do Ponto Proximal Inexato e uma Técnica de Busca
Linear Não Monótona para Otimização Irrestrita

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da
Universidade Federal do Amazonas, como requisito parcial para obtenção
do título de Mestre em Matemática, área de concentração em Otimização.

Orientador: Prof. Dr. Roberto Cristóvão Mesquita Silva

Manaus, 15 de dezembro de 2015.

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

Suellen Paulino Lima

Método do Ponto Proximal Inexato e uma Técnica de Busca
Linear Não Monótona para Otimização Irrestrita

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

BANCA EXAMINADORA

Prof. Dr. Roberto Cristóvão Mesquita Silva
Presidente/Orientador- UFAM

Prof. Dr. Sandro Dimy Barbosa Bitar
Membro - UFAM - ICE

Prof. Dr. Mário Salvatierra Júnior
Membro - UFAM - ICOMP

Manaus, 15 de dezembro de 2015.

Agradecimentos

No decorrer do período de Pós graduação, inúmeras pessoas passaram por minha vida, deixando marcas e lições para toda ela, proporcionando-me alegrias, conhecimento e crescimento pessoal. Neste momento gostaria de agradecer-las, pois, de alguma forma, contribuíram para a conclusão desta etapa.

Agradeço a Deus, pela força espiritual para realização desse trabalho. Aos meus pais e a toda minha grande família, pelo amor e o apoio de sempre. Ao meu amado esposo Hydelo Wagner, pelo carinho, dedicação, paciência e incentivo.

Agradeço ao professor Roberto Cristóvão, que além de colega e excelente professor, me encorajou nesta dissertação, sendo orientador, e acima de tudo por ter contribuído para minha formação pessoal, e também ao professor Sandro Bitar, por ter me mostrado as formas mais belas da matemática.

A todos os professores da Licenciatura em matemática do Instituto Federal de Educação Ciência e Tecnologia do Ceará que me estimularam a buscar este mestrado, bem como a todos os professores do Departamento de Matemática da Universidade Federal do Amazonas, e todos os amigos que adquiri nessa caminhada.

Finalmente, agradeço a todos aqueles que de alguma forma estiveram e estão próximos de mim, fazendo esta vida valer cada vez mais a pena.

RESUMO

Método do Ponto Proximal Inexato e uma Técnica de Busca Linear Não Monótona para Otimização Irrestrita

Apresentaremos neste trabalho algoritmos para resolução de problemas irrestritos. Inicialmente será abordado o Algoritmo do Ponto Proximal Inexato com a utilização de algoritmos clássicos para resolução do problema de regularização da função convexa, continuamente diferenciável e com determinante da hessiana próximo de zero. Em seguida, o Algoritmo de Busca Linear não monótona que tem o objetivo de melhorar a probabilidade de encontrar um ótimo global, utilizando métodos de descida tradicionais para obter o tamanho do passo, além disso, eles podem melhorar a velocidade de convergência em casos específicos do esquema monótono. Ao final faremos a implementação de funções quadráticas e a análise dos resultados obtidos.

Palavras-chave: Ponto proximal. Busca linear. Minimizador

ABSTRACT

Inexact Proximal Point Method and A Nonmonotone Line Search Technique for unconstrained optimization.

This paper shows algorithms for solving unconstrained problems. Firstly it will approach the Inexact Proximal Point Algorithm with the use of classical algorithms for solving the convex function regularization problem, continuously differentiable with the Hessian determinant near zero. In sequence, the non-monotone Linear Search Algorithm that aims to improve the probability of finding a global optimum, using traditional descent methods to obtain the step size, in addition, they can improve the speed of convergence in specific cases of monotonous scheme. At the end we will make the implementation of quadratic functions and analysis of the results.

Keywords: Proximal point. Linear search. Minimizer.

Sumário

Introdução	5
1 Resultados Preliminares	9
1.1 Análise	9
1.2 Fundamentos da Otimização	10
1.3 Elementos de Análise Convexa	12
1.3.1 Funções Convexas Diferenciáveis	12
1.4 Busca Linear	15
1.4.1 A Regra de Armijo	16
1.4.2 A Regra de Wolfe	17
2 Método do Ponto Proximal Inexato	20
2.1 Métodos do Ponto Proximal Clássico	20
2.1.1 Algoritmo do Ponto Proximal em \mathbb{R}^n	20
2.1.2 Análise de Convergência do Algoritmo	21
2.2 Método de Descida	23
2.2.1 Esquema geral dos métodos de descida.	24
2.2.2 Propriedade de Descida	26
2.2.3 Algoritmo do Ponto Proximal Inexato para Funções Diferenciáveis	29
3 Uma Técnica de Busca Linear não monótona	31
3.1 Introdução	31
3.1.1 Algoritmo de Busca Linear Não Monótona	32
3.2 Convergência Global	34
4 Resultados Numéricos	40
4.1 Algoritmos	40
4.1.1 Algoritmo do Ponto Proximal Inexato para Funções Diferenciáveis	40

4.1.2	Teste do Algoritmo do Método do Ponto Proximal In-	
	xato	41
4.1.3	Algoritmo de Busca Linear Não Monótona	48
4.1.4	Teste do Algoritmo Busca Linear não Monótona	48
4.1.5	Função de Rosenbrock	51
4.1.6	Análise dos Resultados	52
5	Considerações Finais	55
5.1	Conclusão	55
5.2	Sugestões para trabalhos futuros	56
	Referências Bibliográficas	57

Lista de Notações

$\{x^k\} \rightarrow x$ ($k \rightarrow \infty$) - a sequência $\{x^k\}$ converge ao ponto x (também escrevemos $\lim_{k \rightarrow \infty} x^k = x$ e para sequências de números reais $\{t_k\}$, ainda

$\{t_k\} \rightarrow t$ ($k \rightarrow \infty$));

$t_k \rightarrow 0$ - a sequência de números reais $\{t_k\}$, todos positivos, converge ao zero;

$x^k = o(t)$ - a sequência $\{x^k\}$ possui a propriedade de que $\{\frac{x^k}{t_k}\} \rightarrow 0$, onde

$t_k \rightarrow 0+$; $f(t) = o(t)$ - a função $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ possui a propriedade de que $\frac{f(t)}{t} \rightarrow 0$ quando $t \rightarrow 0+$;

$f'(x)$ - o gradiente da função f no ponto x ;

$f'(x, d)$ - a derivada direcional da função f no ponto x na direção d .

$N_X(\bar{x})$ - o cone de direções normais no ponto $\bar{x} \in X$ em relação ao conjunto X ;

$T_X(\bar{x})$ - o cone de direções tangentes no ponto $\bar{x} \in X$ em relação ao conjunto X ;

clX - o fecho do conjunto x ;

$D_f(x)$ - o conjunto de direções de descida da função f a partir do ponto x ;

$\nabla f(x)^T$ - Gradente transposto de f (vetor linha), $\nabla f(x)^T = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right)$

Lista de Tabelas

4.1 - Tabela de Iterações do APPIGC - 1º função.....	51
4.2 - Tabela de Iterações do APPIMN - 1º função.....	54
4.3 - Tabela de Iterações do APPIGC - 2º função.....	55
4.4 - Tabela de Iterações do APPIMN - 2º função.....	56
4.5 - Tabela de Iterações do ABLNM - 1º função.....	58
4.6 - Tabela de Iterações do ABLNM - 2º função.....	59
4.7 - Tabela de Iterações do APPIGC - Função de Rosenbrock	60
4.8 - Tabela de Análise dos Resultados - 1º função	62
4.8 - Tabela de Análise dos Resultados - 2º função	63

Lista de Figuras

2.1 - Representação geométrica da propriedade de descida	35
4.1 - Gráfico do APPIGC - 1º função	53
4.2 - Gráfico do APPIMN - 1º função	54
4.3 - Gráfico do APPIGC - 2º função	56
4.4 - Gráfico do APPIMN - 2º função	57
4.5 - Gráfico do ABLNM - 1º função	58
4.6 - Gráfico do ABLNM - 2º função	59
4.7 - Gráfico Função de Rosenbrock	60
4.8 - Gráfico do Confronto entre os Métodos - 1º função	62
4.9 - Gráfico do Confronto entre os Métodos - 2º função	63

INTRODUÇÃO

Inúmeras questões na programação matemática podem ser formulados pelo seguinte problema de minimização irrestrita:

$$\min_{s.a. x \in \mathbb{R}^n} f(x)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa e continuamente diferenciável.

Para resolvermos este problema existe um algoritmo que é uma abordagem padrão de regularização na otimização, denominado Algoritmo do Ponto proximal e que foi desenvolvido por Rockfellar [17] e aplicado pela primeira vez em otimização convexa por Martinet [14], na década de 70.

Nesta dissertação estudamos o Algoritmo do Ponto Proximal Inexato, formulado no artigo publicado por Carlos Humes JR e Paulo J. S. Silva em [7]. Esse algoritmo consiste em minimizar uma função convexa, continuamente diferenciável e com o determinante da hessiana próximo de zero. O processo gera uma sequência de pontos $\{x_k\} \subset \mathbb{R}^n$ a partir de um ponto inicial $x_0 \in \mathbb{R}^n$, segundo uma iteração da regularização do problema:

$$\min_{s.a. x \in \mathbb{R}^n} f(x) + \frac{\alpha_k}{2} \|x - x_k\|^2 \quad (1)$$

Métodos do ponto proximal seguem essa idéia de regularização ao extremo. Em vez de tomar um único passo de Newton, o algoritmo do ponto proximal clássico pede uma solução completa de (1). Provaremos a convergência do algoritmo a partir de um método de descida proposto pelos autores.

Outro método utilizado para resolver problemas irrestritos de otimização, muito utilizado durante as duas grandes guerras, porém, a maioria de suas publicações conhecidas surgiram a partir da década de 1970, é o de Busca Linear que consiste em encontrar uma direção d_k proposta por:

$$x_{k+1} = x_k + \alpha_k d_k$$

e o passo é estipulado pela busca linear.

Nesta dissertação estudaremos um método de Busca Linear Não Monótona fundamentado no artigo proposto por ZHANG, H. HAGER, W. em [19]. Diferentemente dos métodos de busca linear monótona, nesse método o tamanho do passo é escolhido de tal forma que se permita algum crescimento no valor da função. Dessa forma, muitos pesquisadores apontam esse esquema não monótono como uma melhoria na probabilidade de encontrar um ótimo global. Nesse contexto, a prova de convergência global será apresentada para funções não convexas e suaves.

Organizamos este trabalho da seguinte forma:

No capítulo 1 apresentamos alguns conceitos de grande importância para fundamentação teórica desta dissertação.

No capítulo 2 consideramos os seguintes métodos de otimização: Método do Ponto Proximal Clássico, Método de Descida e por último o Método do Ponto Proximal Inexato.

No capítulo 3 apresentamos as características e a convergência do Algoritmo de Busca Linear não Monótona.

No Capítulo 4 apresentaremos a implementação dos métodos citados nesta dissertação. Fazendo o confronto dos resultados obtidos pelos referidos algoritmos minimizando uma função com características específicas destacando as peculiaridades dos métodos bem como as semelhanças entre eles.

No Capítulo 5, finalmente são realizadas as considerações finais e são apresentadas algumas perspectivas de pesquisas futuras.

Objetivos

Objetivo geral

Apresentar e confrontar métodos irrestritos de otimização com especificações parcialmente diferentes para a função.

Objetivo específico

O principal objetivo deste trabalho é destacar dois métodos de minimização irrestrita, o Método do Ponto Proximal Inexato e o de Busca Linear não monótona, mostrando a fundamentação teórica de ambos. Outro objetivo é a implementação dos métodos e a análise sobre os resultados obtidos.

Capítulo 1

Resultados Preliminares

Para garantirmos a boa compreensão dos assuntos apresentados nesta dissertação, vamos durante este capítulo rever alguns conceitos de Análise, Análise convexa e Otimização.

1.1 Análise

Nesta seção veremos alguns resultados provenientes da Análise encontradas nos livros do Elon Lages [11], [12] e [10].

Teorema 1.1.1. Teorema do Valor Médio

Dada $f : U \rightarrow \mathbb{R}$ diferenciável no aberto $U \subset \mathbb{R}^n$, se o segmento de reta $[a, a + v]$ estiver contido em U então existe $\theta \in (0, 1)$ tal que

$$f(a + v) - f(a) = \nabla f(a + \theta v)^T v$$

onde $v = (\alpha_1, \alpha_2, \dots, \alpha_n)$.

Ver demonstração em [10].

Definição 1.1.2. Quando a função $f : I \subset \mathbb{R} \rightarrow \mathbb{R}$ possui derivada em todos os pontos do intervalo I , considera-se a função derivada $f'(x) : I \rightarrow \mathbb{R}$, que associa a cada $x \in I$ a derivada $f'(x)$. Quando f' é contínua diz-se que f é uma função continuamente diferenciável no intervalo I .

Teorema 1.1.3. Teorema Fundamental do Cálculo

Se uma função integrável $f : [a, b] \rightarrow \mathbb{R}$ possui uma primitiva $F : [a, b] \rightarrow \mathbb{R}$, então

$$\int_a^b f(x) dx = F(b) - F(a)$$

Em outros termos, se uma função $F : [a, b] \rightarrow \mathbb{R}$ possui derivada integrável, então

$$F(b) - F(a) = \int_a^b F'(t) dt$$

Ver demonstração em [11].

1.2 Fundamentos da Otimização

Para direcionar melhor nosso estudo em Otimização, nesta seção utilizamos as referências Solodov [8], Izmailov [9] e Ademir Ribeiro [1]

Teorema 1.2.1. Teorema de Weierstrass

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ contínua e $D \subset \mathbb{R}^n$ compacto não vazio. Então existe minimizador global de f em D .

Demonstração. Vejamos primeiro que o conjunto $f(D) = \{f(x) | x \in D\}$ é limitado inferiormente. Suponha por absurdo que não. Então, para todo $k \in \mathbb{N}$, existe $x_k \in D$ tal que $f(x_k) \leq -k$. Como a sequência $\{x_k\}$ está no compacto D , ela possui uma subsequência convergente para um ponto de D , digamos $\{x_k\} \rightarrow x^* \in D$. Pela continuidade de f , temos $f(x_k) \rightarrow f(x^*) \in D$, uma contradição. Portanto, $f(D) = \{f(x) | x \in D\}$ é limitado inferiormente. Considere $f^* = \inf\{f(x) | x \in D\}$. Então, para todo $k \in \mathbb{N}$, existe $\{x_k\} \in D$ tal que

$$f^* \leq f(x_k) \leq f^* + \frac{1}{k},$$

o que implica $f(x_k) \rightarrow f^*$. Repetindo o argumento acima, obtemos $f(x_k) \rightarrow f(x^*)$, com $x^* \in D$. Pela unicidade do limite, temos $f(x^*) = f^* \leq f(x)$, para todo $x \in D$ o que completa a demonstração. \square

Definição 1.2.2. O conjunto de nível da função $f : D \rightarrow \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por

$$\mathcal{L}_{f,D}(c) = \{x \in D | f(x) \leq c\}.$$

Definição 1.2.3. Dizemos que uma sequência $\{x_k\} \subset \mathbb{R}^n$ é crítica em relação ao conjunto D , se $\{x_k\} \subset D$ e ou $\|x_k\| \rightarrow \infty$ ou $\{x_k\} \rightarrow x \in clD \setminus D (k \rightarrow \infty)$. Dizemos que a função $f : D \rightarrow \mathbb{R}$ é coerciva no conjunto D , quando para toda sequência $\{x_k\}$ crítica em relação a D , tem-se $\limsup_{k \rightarrow \infty} f(x_k) = +\infty$.

Corolário 1.2.4. Sejam $D \subset \mathbb{R}^n$, não vazio e $f : D \rightarrow \mathbb{R}$ uma função contínua coerciva. Então, o problema de minimizar f em D possui uma solução global.

Definição 1.2.5. Dizemos que a função $f : D \rightarrow \mathbb{R}$ é semi-contínua inferiormente no ponto $x \in D \subset \mathbb{R}^n$, quando para qualquer sequência $\{x_k\} \subset D$ tal que $\{x_k\} \rightarrow x$ ($k \rightarrow \infty$) tem-se:

$$\liminf_{k \rightarrow \infty} f(x_k) \geq f(x).$$

Dizemos que f é semi-contínua superiormente quando, nas mesmas condições,

$$\limsup_{k \rightarrow \infty} f(x_k) \leq f(x)$$

A função f é semi-contínua inferiormente (superiormente) no conjunto D , quando ela é semi-contínua inferiormente (superiormente) em todos os pontos de D .

Teorema 1.2.6. (Condição necessária de primeira ordem) Suponhamos que a função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja diferenciável, no ponto $\bar{x} \in \mathbb{R}^n$. Se \bar{x} é um minimizador local irrestrito de f , então

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

Demonstração. Seja $d \in \mathbb{R}^n$ arbitrário, porém fixo. Pela definição de minimizador local, existe $\epsilon > 0$ tal que

$$f(\bar{x}) \leq f(\bar{x} + td) \quad \forall t \in [0, \epsilon].$$

Pela diferenciabilidade de f em \bar{x} ,

$$f(\bar{x} + td) = f(\bar{x}) + t\nabla f(\bar{x})^T d + o(t).$$

Logo,

$$0 \leq t\nabla f(\bar{x})^T d + o(t).$$

Dividindo por $t > 0$, temos que

$$0 \leq \nabla f(\bar{x})^T d + \frac{o(t)}{t},$$

e tomando o limite quando $t \rightarrow 0+$, obtemos

$$0 \leq \nabla f(\bar{x})^T d.$$

Como $d \in \mathbb{R}^n$ é arbitrário, podemos escolher $d = -\nabla f(\bar{x})$, o que resulta na condição $0 \leq \nabla f(\bar{x})^T d = -\|\nabla f(\bar{x})\|^2$. Donde segue que $\nabla f(\bar{x}) = 0$. \square

Definição 1.2.7. Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $\bar{x} \in \mathbb{R}^n$, se existe $\epsilon > 0$ tal que

$$f(\bar{x} + td) < f(\bar{x}) \quad \forall t \in (0, \epsilon].$$

Denotamos por $D_f(\bar{x})$ o conjunto de direções de descida da função f no ponto \bar{x} .

1.3 Elementos de Análise Convexa

Nesta seção veremos algumas ferramentas da Análise Convexa que serão necessárias para o desenvolvimento deste trabalho. Poderemos observar a importância das projeções na obtenção de minimizadores e da convexidade no estudo da otimização para funções diferenciáveis e não diferenciáveis. Conheceremos também com mais profundidade as propriedades dos subdiferenciais. Nesta seção utilizamos as referências Solodov [8] e Lemaréchal [2]

1.3.1 Funções Convexas Diferenciáveis

Nesta seção trataremos da caracterização de funções convexas diferenciáveis.

Teorema 1.3.1. Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \rightarrow \mathbb{R}$ uma função diferenciável em D . Então, f é uma função convexa em D se, e somente se:

$$\forall x, y \in D, \quad f(y) \geq f(x) + \nabla f(x)^T(y - x)$$

.

Demonstração. Seja f convexa. Para $x \in D, y \in D$ e $\alpha \in (0, 1]$ quaisquer, definindo $d = y - x$, pela convexidade de f temos que:

$$\begin{aligned} f(x + \alpha d) = f(x + \alpha(y - x)) &= f(\alpha y + (1 - \alpha)x) \\ &\leq \alpha f(y) + (1 - \alpha)f(x), \end{aligned}$$

donde

$$\alpha(f(y) - f(x)) \geq f(x + \alpha d) - f(x).$$

Dividindo os dois lados desigualdade acima por $\alpha > 0$, e passando ao limite quando $\alpha \rightarrow 0+$, obtemos

$$\begin{aligned} f(y) - f(x) &\geq \lim_{\alpha \rightarrow 0+} \frac{f(x + \alpha d) - f(x)}{\alpha} \\ &= \nabla f(x)^T d = \nabla f(x)^T(y - x). \end{aligned}$$

Assim temos:

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) \quad (1.1)$$

Reciprocamente

Definindo novamente $d = y - x$, de (1.1), obtemos:

$$f(x) \geq f(x + \alpha d) - \alpha(f(y) - f(x)). \quad (1.2)$$

Sejam $x, y \in D$. Pelo Teorema do Valor Médio, existe $\alpha \in (0, 1)$, tal que:

$$f(y) - f(x) = \nabla f(x + \alpha(y - x))^T(y - x)$$

Voltando para (1.2) e substituindo $y - x$ por d , temos:

$$f(x) \geq f(x + \alpha d) - \alpha \nabla f(x + \alpha d)^T d$$

Agora, somando $-f(y)$ em ambos os lados da desigualdade

$$f(x) - f(y) \geq f(x + \alpha d) - f(y) - \alpha \nabla f(x + \alpha d)^T d \quad (1.3)$$

Utilizando novamente o Teorema do Valor Médio, temos

$$f(x) - f(y) = -\nabla f(x + \alpha d)^T d$$

Substituindo em (1.3)

$$f(y) \geq f(x + \alpha d) + (1 - \alpha) \nabla f(x + \alpha d)^T d$$

Onde usamos $f(y) \geq f(x) + \nabla f(x)^T(y - x)$ para os pontos x e $x + \alpha d$; y e $x + \alpha d$, respectivamente:

$$f(x) \geq f(x + \alpha d) - \alpha \nabla f(x + \alpha d)^T d \quad (1.4)$$

e

$$f(y) \geq f(x + \alpha d) + (1 - \alpha) \nabla f(x + \alpha d)^T d \quad (1.5)$$

Multiplicando (1.4) por $1 - \alpha \geq 0$ e (1.5) por $\alpha \geq 0$, obtemos:

$$(1 - \alpha)f(x) \geq (1 - \alpha)(f(x + \alpha d) - \alpha \nabla f(x + \alpha d)^T d)$$

e

$$\alpha f(y) \geq \alpha(f(x + \alpha d) + (1 - \alpha) \nabla f(x + \alpha d)^T d)$$

Somando os resultados, teremos:

$$(1 - \alpha)f(x) + \alpha f(y) \geq f(x + \alpha d)$$

Sendo $d = y - x$

$$(1 - \alpha)f(x) + \alpha f(y) \geq f((1 - \alpha)x + \alpha y)$$

o que mostra que f é convexa. □

Observação: Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e f uma função convexa, uma condição necessária para \bar{x} ser minimizador é $\nabla f(\bar{x}) = 0$. Vejamos:

$$f(y) \geq f(\bar{x}) + 0(y - \bar{x}) \Rightarrow f(y) \geq f(\bar{x})$$

Logo \bar{x} é minimizador global.

Proposição 1.3.2. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então para quaisquer duas sequências $\{x_k\} \rightarrow x$, $\{d_k\} \rightarrow d$ ($k \rightarrow \infty$), tem-se que

$$\limsup_{k \rightarrow \infty} f'(x_k; d_k) \leq f'(x; d).$$

Mais ainda, se f é diferenciável no \mathbb{R}^n , então a derivada de f é contínua no \mathbb{R}^n .

Ver demonstração em [8]

Teorema 1.3.3. Teorema de minimização convexa

Sejam $D \subseteq \mathbb{R}^n$ um conjunto convexo e $f : D \rightarrow \mathbb{R}$ uma função convexa em D . Então todo minimizador local de:

$$\min_{s.a. x \in D \subset \mathbb{R}^n} f(x)$$

é global. Além disso, o conjunto de minimizadores é convexo.

Se f é estritamente convexa, não pode haver mais de um minimizador.

Demonstração. Vamos iniciar provando que todo minimizador local de f em D é global. Suponhamos que $\bar{x} \in D$ seja um minimizador local que não é global. Então existe um $y \in D$, tal que $f(y) < f(\bar{x})$.

Definindo $x(\alpha) = \alpha y + (1 - \alpha)\bar{x}$, pela convexidade de D , temos:

$x(\alpha) \in D \quad \forall \alpha \in [0, 1]$. Pela convexidade de f , teremos, $\forall \alpha \in (0, 1]$

$$\begin{aligned} f(x(\alpha)) &= f(\alpha y + (1 - \alpha)\bar{x}) \leq \alpha f(y) + (1 - \alpha)f(\bar{x}) \\ &\leq f(\bar{x}) + \alpha(f(y) - f(\bar{x})) \end{aligned}$$

mas, $f(y) < f(\bar{x})$, então:

$$f(x(\alpha)) \leq f(\bar{x}) + \alpha(f(y) - f(\bar{x})) < f(\bar{x})$$

Para $\alpha > 0$ suficientemente pequeno, $x(\alpha)$ pertence a vizinhança de \bar{x} . Como $f(x(\alpha)) < f(\bar{x})$, contradição, pois \bar{x} não seria minimizador local. Então, todo minimizador local é global.

Agora vamos mostrar que o conjunto de minimizadores é convexo. Chamaremos de $S \subset D$ o conjunto de minimizadores da função e seja $\bar{v} \in \mathbb{R}$ o valor ótimo do problema (seja $f(x) = \bar{v}$ para qualquer $x \in S$). Para quaisquer $\bar{x} \in S$, $\bar{y} \in S$, e $\alpha \in [0, 1)$, pela convexidade de f , temos:

$$\begin{aligned} f(\alpha\bar{x} + (1 - \alpha)\bar{y}) &\leq \alpha f(\bar{x}) + (1 - \alpha)f(\bar{y}) \\ &\leq \alpha\bar{v} + (1 - \alpha)\bar{v} \\ &\leq \bar{v} \end{aligned}$$

Como todos os minimizadores são globais e \bar{v} é valor ótimo então, $\bar{v} \leq f(\alpha\bar{x} + (1 - \alpha)\bar{y}) \leq \bar{v}$. Assim, $\bar{v} = f(\alpha\bar{x} + (1 - \alpha)\bar{y})$, logo $\alpha\bar{x} + (1 - \alpha)\bar{y} \in S$ e S é convexo.

Suponhamos agora que a função é estritamente convexa e que seja $\bar{x} \in S$, $\bar{y} \in S$, tal que $\bar{x} \neq \bar{y}$. Seja $\bar{v} \in \mathbb{R}$ o valor ótimo tal que para qualquer $\bar{x} \in S$, $f(\bar{x}) = \bar{v}$. Então, pela convexidade de D , temos:

$$f(\alpha\bar{y} + (1 - \alpha)\bar{x}) \geq f(\bar{x}) = \bar{v}.$$

No entanto, pela convexidade estrita, temos:

$$f(\alpha\bar{y} + (1 - \alpha)\bar{x}) < \alpha f(\bar{y}) + (1 - \alpha)f(\bar{x}) = \alpha\bar{v} + (1 - \alpha)\bar{v} = \bar{v}$$

O que resulta em contradição. Concluímos que neste caso o minimizador é único.

□

1.4 Busca Linear

Nesta seção apresentaremos algumas das principais regras de busca linear que determinam o comprimento do passo, sendo que o cálculo desse passo dá-se observando o comportamento da função f ao longo da semi-reta a partir de x_k na direção d_k , ou ao longo de um intervalo limitado na mesma direção. Mais informações podem ser obtidas em [9].

1.4.1 A Regra de Armijo

A regra de Armijo tem como estratégia computar um comprimento de passo que resulta em decréscimo suficiente da função f em relação ao valor $f(x_k)$. A idéia é tentar contornar algumas patologias que ocorrem devido ao processo.

Algoritmo para Regra de Armijo

1. Fixe os parâmetros $\hat{\alpha} > 0$, $\sigma, \theta \in (0, 1)$.
2. Inicialização: Tome $\alpha := \hat{\alpha}$
3. Critério de Parada: Se α satisfaz:

$$f(x_k + \alpha d_k) \leq f(x_k) + \sigma \alpha \nabla f(x_k)^T d_k. \quad (1.6)$$

Então pare. Caso contrário, execute o passo seguinte.

4. Passo Iterativo: Tome $\alpha := \theta \alpha$ e volte ao critério de parada.

Caso contrário, aceitamos $\alpha_k = \alpha$ como o valor do comprimento do passo. Em outras palavras, α_k é o maior número entre todos os números da forma $\alpha := \hat{\alpha} \theta^i$, $i = 0, 1, 2, \dots$, que satisfaz a desigualdade (1.6). O resultado seguinte mostra que quando d_k satisfaz a condição

$$\nabla f(x_k)^T d_k < 0, \quad (1.7)$$

então a regra de Armijo está bem definida (ela produz um valor $\alpha_k > 0$ aceitável, após um número finito de reduções do valor inicial $\hat{\alpha}$).

Lema 1.4.1. (A regra de Armijo está bem definida)

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $x_k \in \mathbb{R}^n$. Suponhamos que $d_k \in \mathbb{R}^n$ satisfaça (1.7).

Então a desigualdade (1.6) é satisfeita para todo $\alpha > 0$ suficientemente pequeno. Em particular, a regra de Armijo está bem definida e termina com um $\alpha_k > 0$.

Demonstração. Para todo α suficientemente pequeno,

$$\begin{aligned} f(x_k + \alpha d_k) - f(x_k) &= \nabla f(x_k)^T \alpha d_k + o(\alpha) \\ &= \sigma \alpha \nabla f(x_k)^T d_k + (1 - \sigma) \alpha \nabla f(x_k)^T d_k + o(\alpha) \\ &= \sigma \alpha \nabla f(x_k)^T d_k + \alpha ((1 - \sigma) \nabla f(x_k)^T d_k + \frac{o(\alpha)}{\alpha}) \\ &\leq \sigma \alpha \nabla f(x_k)^T d_k, \end{aligned}$$

onde a desigualdade segue de

$$(1 - \sigma)\nabla f(x_k)^T d_k + \frac{o(\alpha)}{\alpha} \leq \frac{1 - \sigma}{2}\nabla f(x_k)^T d_k < 0,$$

que vale para $\alpha > 0$ suficientemente pequeno, pela condição (1.7). \square

Portanto, sob a condição (1.7), o cálculo de α_k pela regra de Armijo garante a propriedade $f(x_{k+1}) < f(x_k)$ de descida para os valores da função objetivo. Mais ainda, a desigualdade (1.6), com α_k , fornece uma estimativa de quanto o valor $f(x_{k+1})$ é menor do que $f(x_k)$. Esta estimativa é importante para a prova de convergência do método.

1.4.2 A Regra de Wolfe

O comprimento do passo $\alpha_k > 0$ é calculado para satisfazer as seguintes desigualdades (em relação a α):

$$f(x_k + \alpha d_k) \leq f(x_k) + \sigma_1 \alpha \nabla f(x_k)^T d_k, \quad (1.8)$$

$$\nabla f^T(x_k + \alpha d_k) d_k \geq \sigma_2 \nabla f^T(x_k) d_k, \quad (1.9)$$

onde $0 < \sigma_1 < \sigma_2 < 1$.

A primeira condição acima é, de novo, aquela da regra de Armijo. A segunda condição controla, de certo modo, a precisão da minimização de f na semi-reta $x_k + \alpha d_k$, $\alpha \geq 0$. Como já vimos acima, se $\alpha > 0$ minimiza esta função exatamente tem-se que $\nabla f^T(x_k + \alpha d_k) d_k = 0$. Portanto, a desigualdade (1.9) é um teste de qualidade de um $\alpha > 0$ dado como aproximação de minimizadores de f na semi-reta ao longo da direção d_k . Quando o cálculo de derivada de f não é muito caro, a regra de Wolfe é considerada a mais eficiente regra de busca linear. A eficiência da regra é garantida pelo controle do passo α . A regra de Wolfe pode ser implementada seguindo os procedimentos abaixo: Fixamos os parâmetros $0 < \sigma_1 < \sigma_2 < 1$. Tomar $\hat{\alpha} := 0$ e $\bar{\alpha} := 0$. Tomar um valor inicial $\alpha > 0$.

1. Se (1.8) e (1.9) se satisfazem, passar ao Item 6.
2. Se (1.8) é violada, tomar $\hat{\alpha} := \alpha$ e passar ao Item 4.
3. Se (1.9) é violada, tomar $\bar{\alpha} := \alpha$.

4. Se $\hat{\alpha} = 0$, escolher um novo valor de $\alpha > \bar{\alpha}$ (extrapolação) e passar ao Item 1.
5. Escolher um novo valor $\alpha \in (\bar{\alpha}, \hat{\alpha})$ (interpolação) e passar ao Item 1.
6. Tomar $\alpha_k := \alpha$.

A violação da condição (1.8) significa que o valor de α é grande demais, enquanto a violação da condição (1.9) significa que este valor é pequeno demais. O procedimento acima funciona da seguinte maneira. Primeiro, se fazem passos de extrapolação até que $\hat{\alpha}$ se torne positivo. Após isso, se fazem passos de interpolação. Enquanto isso, $\hat{\alpha}$ pode só diminuir, mais ficando positivo, e $\bar{\alpha}$ pode só aumentar, mais sempre ficando menor que $\hat{\alpha}$.

Lema 1.4.2. A regra de Wolfe está bem definida

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função limitada inferiormente no \mathbb{R}^n , diferenciável com derivada contínua. Suponhamos que $x_k \in \mathbb{R}^n$ e $d_k \in \mathbb{R}^n$ satisfaçam $\nabla f(x_k)^T d_k < 0$.

Então qualquer implementação da regra de Wolfe, para a qual tem-se que $\bar{\alpha} \rightarrow +\infty$ no caso de número infinito de extrapolações e $(\hat{\alpha} - \bar{\alpha}) \rightarrow 0$ no caso de número infinito de interpolações, está bem definida, no sentido de que ela termina com um valor de comprimento de passo $\alpha_k > 0$.

Demonstração. Suponhamos primeiro que o número de extrapolações seja infinito. Neste caso, temos uma sequência de valores $\bar{\alpha}_k \rightarrow +\infty$, para os quais

$$f(x_k + \bar{\alpha}d_k) \leq f(x_k) + \sigma_1 \bar{\alpha} \nabla f(x_k)^T d_k. \quad (1.10)$$

Levando em conta $\nabla f(x_k)^T d_k < 0$, isto contradiz o fato de que f é limitada inferiormente. Portanto, o número de extrapolações é finito.

Suponhamos agora que o número de interpolações seja infinito. Neste caso, as sequências monótonas de valores de $\bar{\alpha}$ e de $\hat{\alpha}$ convergem a limite comum α^* . Além disso, os elementos da primeira sequência satisfazem (1.10) e

$$\nabla f(x_k + \bar{\alpha}d_k)^T d_k < \sigma_2 \nabla f(x_k)^T d_k, \quad (1.11)$$

e os elementos da segunda sequência satisfazem

$$f(x_k + \tilde{\alpha}d_k) > f(x_k) + \sigma_1 \tilde{\alpha} \nabla f(x_k)^T d_k. \quad (1.12)$$

Passando ao limite em (1.10) e (1.12), obtemos a igualdade

$$f(x_k + \tilde{\alpha}d_k) = f(x_k) + \sigma_1 \tilde{\alpha} \nabla f(x_k)^T d_k \quad (1.13)$$

Levando em conta (1.12) e o fato de que os valores de $\hat{\alpha}$ são decrescentes, concluimos que estes valores ficam estritamente maiores de α^* . Usando (1.13), escrevemos (1.12) como

$$\begin{aligned} f(x_k + \hat{\alpha}d_k) &> f(x_k) + \sigma_1(\tilde{\alpha} + \hat{\alpha} - \tilde{\alpha})\nabla f(x_k)^T d_k \\ &= f(x_k + \tilde{\alpha}d_k) + \sigma_1(\hat{\alpha} - \tilde{\alpha})\nabla f(x_k)^T d_k. \end{aligned}$$

E como $\hat{\alpha} - \tilde{\alpha} > 0$, temos que

$$\frac{f(x_k + \hat{\alpha}d_k) - f(x_k + \tilde{\alpha}d_k)}{\hat{\alpha} - \tilde{\alpha}} > \sigma_1 \nabla f(x_k)^T d_k.$$

Passando ao limite (quando $(\hat{\alpha} - \tilde{\alpha}) \rightarrow 0$) e usando que $\sigma_1 < \sigma_2$ e $\nabla f(x_k)^T d_k < 0$, obtemos

$$\nabla f(x_k + \tilde{\alpha}d_k)^T d_k \geq \sigma_1 \nabla f(x_k)^T d_k > \sigma_2 \nabla f(x_k)^T d_k \quad (1.14)$$

Por outro lado, passando ao limite em (1.11), obtemos a desigualdade

$$\nabla f(x_k + \tilde{\alpha}d_k)^T d_k \leq \sigma_2 \nabla f(x_k)^T d_k,$$

□

em contradição com (1.14). Analisando a desigualdade acima observamos que as hipóteses feitas sobre o comportamento da função objetivo f no \mathbb{R}^n , também podem ser extendidas de forma análoga sobre a função $f(x_k + \alpha d_k)$ na semi reta $\mathbb{R}_+ \setminus \{0\}$.

Capítulo 2

Método do Ponto Proximal Inexato

Neste capítulo apresentaremos o algoritmo do Ponto Proximal Clássico que irá servir como base para exposição do algoritmo do Ponto Proximal Inexato para funções diferenciáveis com motivação geométrica para justificar a propriedade de descida, que foi proposto por Humes e Silva [7].

2.1 Métodos do Ponto Proximal Clássico

Exibiremos o Algoritmo do Ponto Proximal Clássico que foi proposto inicialmente por Martinet [14] e desenvolvido por Rockafellar [17] para problemas de minimização convexa no \mathbb{R}^n .

2.1.1 Algoritmo do Ponto Proximal em \mathbb{R}^n .

Seja o problema

$$\min_{s.a. x \in \mathbb{R}^n} f(x) \quad (2.1)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa, diferenciável e limitada inferiormente.

O Algoritmo do Ponto Proximal para o problema (2.1) em \mathbb{R}^n gera uma sequência $\{x_k\} \subset \mathbb{R}^n$, da seguinte maneira :

Inicialização: Escolha um $x_0 \in \mathbb{R}^n$ e $\lambda_0 \in (0, \bar{\lambda})$.

Iteração: Para $k = 1, 2, 3, \dots$

1. Escolha o parâmetro de regularização $\lambda_k \in (0, \bar{\lambda})$, e encontre x_{k+1} :

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x_k\|^2\} \quad (2.2)$$

2. Faça $k := k + 1$ e volte para 1.

Observamos que a sequência $\{x_k\}$ gerada pelo Algoritmo 2.1.1 está bem definida. Pois, definindo $f_k(x) := f(x) + \lambda_k \|x - x_k\|^2$ temos que

$$\lim_{\|x\| \rightarrow \infty} f_k(x) = +\infty,$$

ou seja, f_k é coerciva, pois f é limitada inferiormente. Assim, a minimização do problema se reduz a um conjunto compacto, e como f_k é contínua, pelo Teorema de Weierstrass ela atinge seu mínimo. Temos também por hipótese que f é convexa e como $\lambda_k \|x - x_k\|^2$ é estritamente convexa segue que f_k é estritamente convexa, e assim tem um único minimizador, x_{k+1} .

2.1.2 Análise de Convergência do Algoritmo

Nesta seção provaremos o Teorema de Convergência do Algoritmo do Ponto Proximal. Para provarmos esse resultado, necessitamos da definição da sequência Fejér convergente.

Definição 2.1.1. Uma sequência $\{y_k\}$ em \mathbb{R}^n é dita ser Fejér convergente a um conjunto $U \subset \mathbb{R}^n$, com respeito à norma Euclideana, se

$$\|y_{k+1} - u\| \leq \|y_k - u\|, \forall k \geq 0, \forall u \in U.$$

Os resultados apresentados a seguir serão de grande importância para provarmos a convergência do Algoritmo do Ponto Proximal.

Proposição 2.1.2. Se $\{y_k\}$ é uma sequência Fejér convergente a um conjunto $U \subset \mathbb{R}^n$, então $\{y_k\}$ é limitada. Além disso, se um ponto de acumulação y de $\{y_k\}$ pertence a U , então:

$$\lim_{k \rightarrow +\infty} y_k = y.$$

Demonstração. Fixando um ponto $u \in U$ e usando a hipótese que a sequência é Fejér convergente, temos que:

$$\|y_{k+1} - u\| \leq \|y_k - u\| \leq \dots \leq \|y_0 - u\|.$$

Portanto, a sequência $\{y_k\}$ está contida em uma bola de centro u e raio $\|y_0 - u\|$, logo $\{y_k\}$ é limitada. Para a segunda parte considere uma subsequência $\{y_{k_j}\}$ de $\{y_k\}$, tal que $\lim_{j \rightarrow \infty} y_{k_j} = \bar{y}$. Como $\bar{y} \in U$ e $\{y_k\}$ é Fejér convergente, pela Definição 2.1.1 a sequência $\{\|y_k - \bar{y}\|\}$ é não-crescente e não-negativa, portanto convergente.

Como, a subsequência $\{\|y_{k_j} - \bar{y}\|\}$ converge a zero, segue que $\{\|y_k - \bar{y}\|\}$ converge a zero. Isto é, $\lim_{k \rightarrow \infty} \|y_k - \bar{y}\| = 0$. Logo, $\lim_{k \rightarrow \infty} y_k = \bar{y}$. \square

Proposição 2.1.3. Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa e continuamente diferenciável, $U \subset \mathbb{R}^n$, o conjunto de minimizadores de f , é não vazio. Então:

(i) A sequência $\{x_k\}$ gerada pelo algoritmo 2.1.1 satisfaz a seguinte desigualdade:

$$\|x_{k+1} - \bar{x}\|^2 \leq \|x_k - \bar{x}\|^2 - \|x_{k+1} - x_k\|^2, \forall k \geq 0 \text{ e } \forall \bar{x} \in U;$$

(ii) A sequência $\{x_{k+1} - x_k\}$ converge para zero;

(iii) A sequência $\{x_k\}$ gerada pelo Algoritmo 2.1.1 tem pontos de acumulação e todos pertencem a U .

Demonstração. Temos que:

$$\begin{aligned} \|x_k - \bar{x}\|^2 &= \|x_k - x_{k+1} + x_{k+1} - \bar{x}\|^2 \\ &= \|x_k - x_{k+1}\|^2 + \|x_{k+1} - \bar{x}\|^2 + 2\langle x_k - x_{k+1}, x_{k+1} - \bar{x} \rangle. \end{aligned}$$

Como x_{k+1} satisfaz (2.2), temos:

$$0 = \nabla f_k(x_{k+1}) = \nabla f(x_{k+1}) + 2\lambda_k(x_{k+1} - x_k).$$

Assim,

$$\begin{aligned} \|x_k - \bar{x}\|^2 &- \|x_{k+1} - x_k\|^2 - \|x_{k+1} - \bar{x}\|^2 = 2\langle x_k - x_{k+1}, x_{k+1} - \bar{x} \rangle \\ &= \frac{1}{\lambda_k} \nabla f(x_{k+1})^T (x_{k+1} - \bar{x}) \\ &\geq \frac{1}{\lambda_k} [f(x_{k+1}) - f(\bar{x})] \\ &\geq 0 \end{aligned}$$

onde a primeira desigualdade segue pelo fato de f ser convexa e a segunda pelo fato de \bar{x} ser um minimizador de (2.1). Desta forma, a afirmação (i) é estabelecida.

Agora, por (i) temos:

$$0 \leq \|x_{k+1} - x_k\|^2 \leq \|x_k - \bar{x}\|^2 - \|x_{k+1} - \bar{x}\|^2$$

Como $\{\|x_{k+1} - \bar{x}\|\}$ é não crescente e não negativa então é convergente. Assim, o lado direito da desigualdade converge para 0, logo a afirmação (ii)

é obtida.

Agora note que por (i), $\{x_k\}$ é uma sequência Fejér convergente em U , assim pela Proposição 2.1.2, ela é limitada. Desta forma, seja \bar{x} um ponto de acumulação de $\{x_k\}$ e $\{x_{j_k}\}$ uma subsequência de $\{x_k\}$, tal que

$$\lim_{k \rightarrow +\infty} (x_{j_k}) = \bar{x}.$$

Por (2.2), temos que

$$0 = \nabla f_k(x_{k+1}) = \nabla f(x_{k+1}) + 2\lambda_k(x_{k+1} - x_k),$$

assim

$$0 = \nabla f(x_{j_k+1}) = 2\lambda_{j_k}(x_{j_k} - x_{j_k+1}). \quad (2.3)$$

Mas, por (ii), segue que:

$$\lim_{k \rightarrow +\infty} (x_{j_k+1}) = \lim_{k \rightarrow +\infty} (x_{j_k}) = \bar{x}.$$

Usando que $\lambda_k \leq \bar{\lambda}$, a diferenciabilidade contínua de f e aplicando o limite em (2.3), obtemos que $\nabla f(\bar{x}) = 0$. Portanto, pela convexidade de f , $\bar{x} \in U$ e a prova da afirmação (iii) está concluída. □

Com auxílio dos resultados anteriores, provaremos a seguir o teorema da convergência do algoritmo do ponto proximal.

Teorema 2.1.4. Teorema da Convergência do Algoritmo do Ponto Proximal

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e continuamente diferenciável e $U \subset \mathbb{R}^n$, o conjunto de minimizadores de f , não vazio. Então a sequência $\{x_k\}$, obtida pelo Algoritmo 2.1.1, converge para um ponto $x^* \in U$.

Demonstração. Constatamos que o item (i) e (iii) da Proposição 2.1.3 asseguram que a sequência $\{x_k\}$ é Fejér convergente e que todos os seus pontos de acumulação pertencem a U . Assim, pela Proposição 2.1.2, a sequência $\{x_k\}$ converge a $x^* \in U$. □

2.2 Método de Descida

Com o objetivo minimizar uma função f , é coerente que a mesma decresça na direção escolhida. Uma direção que apresenta essa característica é denominada direção de descida.

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função dada. Uma das estratégias mais naturais para resolver o problema irrestrito

$$\min_{s.a. x \in \mathbb{R}^n} f(x) \quad (2.4)$$

é a seguinte. Dada uma aproximação $x_k \in \mathbb{R}^n$ da solução do problema, encontramos um ponto $x_{k+1} \in \mathbb{R}^n$ tal que

$$f(x_{k+1}) < f(x_k).$$

Obviamente, isto pode ser feito de várias maneiras. Uma realização desta estratégia básica é tomar uma direção $d_k \in \mathbb{R}^n$ tal que f é decrescente, pelo menos para passos curtos, a partir do ponto x_k nessa direção, e computar um comprimento do passo $\alpha_k > 0$ que fornece um valor de f menor do que no ponto x_k ,

$$f(x_k + \alpha_k d_k) < f(x_k).$$

Assim obtemos o iterando seguinte $x_{k+1} = x_k + \alpha_k d_k$. Repetimos o processo para o novo ponto a partir de x_{k+1} . Métodos deste tipo se chamam métodos de descida, descreveremos alguns resultados já que os métodos são diversos.

2.2.1 Esquema geral dos métodos de descida.

Nesta subseção veremos o Esquema geral dos métodos de descida descrito em [9].

Seja $x \in \mathbb{R}^n$ um ponto qualquer, tal que $\nabla f(x) \neq 0$. Consideremos os pontos de forma $x + td$, onde $d = -\nabla f(x)$. Pela fórmula de Taylor de primeira ordem, para todo $t \geq 0$ temos que

$$\begin{aligned} f(x + td) &= f(x) - t\|\nabla f(x)\|^2 + o(t\|\nabla f(x)\|) \\ &= f(x) - t(\|\nabla f(x)\|^2 + \frac{o(t)}{t}). \end{aligned}$$

Para valores de $t > 0$ suficientemente pequenos, tem-se que $\|\nabla f(x)\|^2 > \frac{o(t)}{t}$, considerando esses valores vale que $f(x + td) < f(x)$. De fato, o mesmo raciocínio pode ser usado para qualquer direção $d \in \mathbb{R}^n$ satisfazendo

$$\nabla f(x)^T d < 0$$

A seguir, formalizaremos estas observações.

Definição 2.2.1. Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida da função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $x \in \mathbb{R}^n$, se existe $\epsilon > 0$ tal que

$$f(x + td) < f(x) \quad \forall t \in (0, \epsilon].$$

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

Em outras palavras, $d \in D_f(x)$ se, e somente se, todo passo suficientemente curto a partir do ponto x na direção d resulta em algum decréscimo da função f .

Lema 2.2.2. Direções de descida

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

- (a) Para todo $d \in D_f(x)$, tem-se $\nabla f(x)^T d \leq 0$.
- (b) Se $d \in \mathbb{R}^n$ satisfaz $\nabla f(x)^T d < 0$, tem-se que $d \in D_f(x)$.

Demonstração. Seja $d \in D_f(x)$. Para todo $t > 0$ suficientemente pequeno, pela diferenciabilidade de f em \bar{x} ,

$$0 > f(x + td) - f(x) = t(\nabla f(x)^T d + \frac{o(t)}{t}).$$

Dividindo os dois lados da desigualdade acima por $t > 0$ e passando ao limite quando $t \rightarrow 0+$, obtemos $0 \geq \nabla f(x)^T d$, o que prova o item (a). Suponhamos agora que $\nabla f(x)^T d < 0$. Temos que

$$f(x + td) - f(x) = t(\nabla f(x)^T d + \frac{o(t)}{t}).$$

Em particular, para todo $t > 0$ suficientemente pequeno, temos

$$\nabla f(x)^T d + \frac{o(t)}{t} \leq \frac{1}{2} \nabla f(x)^T d < 0,$$

o que implica que $f(x + td) - f(x) < 0$, i.e., $d \in D_f(x)$. □

2.2.2 Propriedade de Descida

Agora trataremos da propriedade de descida, apresentada por Humes e Silva [7], importante para provarmos a convergência do Método do Ponto Proximal Inexato para funções diferenciáveis. Vamos considerar o problema de minimização irrestrito.

$$\min_{s.a. x \in \mathbb{R}^n} f(x) \quad (2.5)$$

onde f é uma função convexa e continuamente diferenciável.

Vários algoritmos usados para resolver (2.5), geram uma sequência minimizante, isto é uma sequência $\{x_k\}$, tal que $\{f(x_k)\}$ é decrescente.

Suponha que queiramos relacionar o decrescimento de f com uma determinada sequência $\{x_k\}$ e a norma do gradiente de f nesta mesma sequência. Isto deve ser feito usando a seguinte desigualdade fundamental:

$$f(x_k) \geq f(x_{k+1}) + g_{k+1}^T(x_k - x_{k+1}) \quad (2.6)$$

onde g_{k+1} denota $\nabla f(x_{k+1})$. Note que (2.6) é equivalente a:

$$f(x_k) \geq f(x_{k+1}) + \|g_{k+1}\| \|x_k - x_{k+1}\| \cos \theta \quad (2.7)$$

onde θ é o ângulo entre g_{k+1} e $x_k - x_{k+1}$. Veja a figura [2.1].

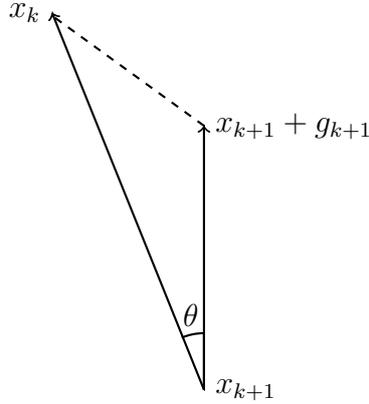


Figura 2.1: Representação geométrica

Então, x_{k+1} é melhor do que o ponto anterior, sempre que θ é agudo. Se considerarmos o triângulo $[x_k, x_{k+1}, x_{k+1} + g_{k+1}]$, mostrado na figura 2.1, θ será agudo se o lado oposto a ele é o menor lado.

Uma forma de garantir essa propriedade é concentrar-se em sequências que verifiquem o seguinte critério de aceitação:

$$\|g_{k+1} + (x_{k+1} - x_k)\| \leq \sigma \max\{\|g_{k+1}\|, \|x_{k+1} - x_k\|\}, \quad (\text{CA1})$$

onde σ é um número positivo menor que 1. Adiante vamos mostrar que o método do ponto proximal gera uma sequência que satisfaz (CA1). Vamos começar a nossa análise estudando algumas propriedades simples do triângulo $[x_k, x_{k+1}, x_{k+1} + g_{k+1}]$.

Lema 2.2.3. Se (CA1) vale, então: $0 \leq \text{sen } \theta \leq \sigma$ e conseqüentemente $\cos \theta \geq \sqrt{1 - \sigma^2}$.

Demonstração. Denotamos por M , o comprimento do maior dos lados do triângulo, ($M := \max\{\|g_{k+1}\|, \|x_{k+1} - x_k\|\}$) e seja α seu ângulo oposto. Segue pela lei dos senos que:

$$\frac{M}{\text{sen } \alpha} = \frac{\|g_{k+1} + x_{k+1} - x_k\|}{\text{sen } \theta}.$$

Utilizando (CA1), obtemos:

$$\frac{\text{sen } \theta}{\text{sen } \alpha} = \frac{\|g_{k+1} + x_{k+1} - x_k\|}{M} \leq \sigma$$

o que é equivalente a:

$$\text{sen } \theta \leq \sigma \text{sen } \alpha,$$

pois, $\alpha < \pi$, então:

$$0 \leq \text{sen } \theta \leq \sigma.$$

Usando a equação fundamental $\text{sen}^2 \theta + \text{cos}^2 \theta = 1$ e a última expressão obtida, concluímos que :

$$\text{cos } \theta \geq \sqrt{1 - \sigma^2}.$$

□

Observação 2.2.4. Note que (CA1) não implica somente que θ é agudo; na verdade afirma que θ está longe de $\frac{\pi}{2}$.

O critério de aceitação também implica que a relação entre os lados $\|g_{k+1}\|$ e $\|x_{k+1} - x_k\|$ é bem comportada.

Lema 2.2.5. Seja $M = \max\{\|g_{k+1}\|, \|x_{k+1} - x_k\|\}$ e m o mínimo destes dois números. Se (CA1) é satisfeita, obtemos;

$$(1 - \sigma)M \leq m \leq M.$$

Demonstração. De (CA1) , temos:

$$M - m \leq \|x_k - x_{k+1} - g_{k+1}\| \leq \sigma M$$

Então;

$$(1 - \sigma)M \leq m \leq M.$$

□

Temos agora todos os elementos para obter um decréscimo de f com qualquer sequência satisfazendo (CA1). De (2.7), temos:

$$\begin{aligned} f(x_k) &\geq f(x_{k+1}) + \|g_{k+1}\| \|x_k - x_{k+1}\| \cos \theta \\ &\geq f(x_{k+1}) + mM \cos \theta \end{aligned}$$

Do Lema 2.2.3 temos:

$$f(x_k) \geq f(x_{k+1}) + mM\sqrt{1 - \sigma^2}$$

Utilizando o Lema 2.2.5, obtemos,

$$f(x_k) \geq f(x_{k+1}) + (1 - \sigma)\sqrt{1 - \sigma^2}M^2$$

Usando o fato que $M \geq \|g_{k+1}\|$; Concluímos que:

$$f(x_k) \geq f(x_{k+1}) + (1 - \sigma)\sqrt{1 - \sigma^2}\|g_{k+1}\|^2 \quad (2.8)$$

Essa desigualdade será de grande importância para provarmos a convergência do método do ponto proximal.

2.2.3 Algoritmo do Ponto Proximal Inexato para Funções Diferenciáveis

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função continuamente diferenciável e convexa. Seja σ um número em $[0, 1)$.

1- Inicialização: Seja x_1 qualquer ponto em \mathbb{R}^n .

2- Iteração: Calcular x_{k+1} tal que, para:

$$g_{k+1} = \nabla f(x_{k+1}) \quad e \quad e_k = g_{k+1} + x_{k+1} - x_k, \\ \|e_k\| \leq \sigma \max\{\|g_{k+1}\|, \|x_{k+1} - x_k\|\}$$

É importante observar que a condição de aceitação apresentada na etapa iterativa é exatamente (CA1). Além disso, se minimizarmos aproximadamente a função regularizada

$$f(x) + \frac{1}{2}\|x - x_k\|^2$$

com precisão suficiente, podemos encontrar um ponto que satisfaz essa condição. Note que, aplicando (2.8) sucessivamente uma sequência calculada pelo Algoritmo do Ponto Proximal Inexato, obtemos:

$$f(x_1) \geq f(x_k) + (1 - \sigma)\sqrt{1 - \sigma^2} \sum_{j=2}^k \|g_j\|^2 \quad (2.9)$$

Teorema 2.2.6. Suponhamos que f é convexa, continuamente diferenciável e limitada inferiormente. Seja $\{x_k\}$ qualquer sequência gerada pelo algoritmo do ponto proximal inexato. Então, todos os seus pontos de acumulação são soluções para o problema (2.4).

Demonstração. Note que, por (2.8) a sequência $\{f(x_k)\}$ é decrescente, sendo f limitada inferiormente esta sequência converge para algum valor real. Sendo assim por [2.9], temos:

$$\nabla f(x_k) \rightarrow 0.$$

Seja \bar{x} um ponto de acumulação da sequência $\{x_k\}$, então existe uma subsequência $\{x_{k_j}\} \subset \{x_k\}$, tal que

$$\lim_{j \rightarrow \infty} x_{k_j} = \bar{x}.$$

A continuidade do ∇f , implica que:

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

Como f é convexa então todas as condições necessárias de otimalidade são suficientes. Então \bar{x} é solução do problema.

□

Capítulo 3

Uma Técnica de Busca Linear não monótona

3.1 Introdução

Neste capítulo faremos um estudo sobre a convergência do algoritmo de busca linear não monótona proposto por Zhang e Hager [19]. Consideremos o problema de minimização irrestrita:

$$\min_{s.a. x \in \mathbb{R}^n} f(x) \quad (3.1)$$

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

Muitos métodos iterativos para (3.1) produzem uma sequência x_0, x_1, x_2, \dots , onde x_{k+1} é gerado a partir de x_k , a direção atual d_k , e o comprimento do passo $\alpha_k > 0$ pela regra:

$$x_{k+1} = x_k + \alpha_k d_k$$

Nos métodos de busca linear monótona, o comprimento do passo α_k é escolhido de modo que $f(x_{k+1}) < f(x_k)$. No método de busca linear não monótona algum crescimento no valor da função é permitido. Como apontado por muitos pesquisadores [3,18], esquemas não monótonos podem melhorar a probabilidade de encontrar um ótimo global. A estrutura mais antiga deste método foi desenvolvida por Grippo, Lampariello e Lucidi em [6] para o Método de Newton. Sua abordagem era aproximadamente a seguinte: introduzir os seguintes parâmetros $\lambda_1, \lambda_2, \sigma$ e δ , onde $0 < \lambda_1 < \lambda_2$ e $\sigma, \delta \in (0, 1)$, M inteiro não negativo, e incluíram $\alpha_k = \overline{\alpha}_k \sigma^{h_k}$ onde $\overline{\alpha}_k \in (\lambda_1, \lambda_2)$ é o julgamento do passo, e h_k é o menor número negativo tal que:

$$f(x_k + \alpha_k d_k) \leq \max_{0 \leq j \leq m_k} [f(x_{k-j})] + \delta \alpha_k \nabla f(x_k)^T d_k, \quad (3.2)$$

onde $m_0 = 0$ e $0 \leq m_k \leq \min[m_{k-1} + 1, M]$, $k \geq 1$.
O algoritmo de busca linear não monótona é o seguinte:

3.1.1 Algoritmo de Busca Linear Não Monótona

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função continuamente diferenciável.

Inicialização: Escolha um ponto inicial x_0 , e os parâmetros :
 $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $0 < \delta < \sigma < 1 < \rho$ e $\mu > 0$, usaremos
 $C_0 = f(x_0)$, $Q_0 = 1$, e $k = 0$.

Teste de Convergência: Se $\|\nabla f(x_k)\|$ é suficientemente pequeno,
então pare.

Atualização da Busca Linear: Faça $x_{k+1} = x_k + \alpha_k d_k$, onde α_k
satisfaz tanto as condições (não-monótonas) de Wolfe:

$$f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k \nabla f(x_k)^T d_k, \quad (3.3)$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f(x_k)^T d_k \quad (3.4)$$

ou as condições de Armijo (não monótonas) : $\alpha_k = \bar{\alpha}_k \rho^{h_k}$, onde $\bar{\alpha}_k > 0$,
é o comprimento do passo e h_k é o maior inteiro, tal que (3.3) e acontece
 $\alpha_k \leq \mu$.

Atualização de Custo: Escolha $\eta_k \in [\eta_{\min}, \eta_{\max}]$, e faça:

$$Q_{k+1} = \eta_k Q_k + 1, \quad C_{k+1} = \frac{(\eta_k Q_k C_k + f(x_{k+1}))}{Q_{k+1}} \quad (3.5)$$

Substituir k por $k + 1$ e retorna para o teste de convergência.
Observe que C_{k+1} é uma combinação convexa de C_k e $f(x_{k+1})$. Desde que,
 $C_0 = f(x_0)$, segue-se que C_k é uma combinação convexa dos valores das
funções $f(x_0), f(x_1), f(x_2), \dots, f(x_k)$. A escolha de η_k controla o grau de
não monotonicidade. Se $\eta_k = 0$ para cada k , então a busca linear se torna a
busca linear de Wolfe ou de Armijo monótona usual. Se $\eta_k = 1$ para cada k ,
então $C_k = A_k$, onde:

$$A_k = \frac{1}{k+1} \sum_{i=0}^k f_i, \quad f_i = f(x_i),$$

É a média dos valores da função. Assim, temos:

$$\begin{aligned}
A_0 &= f(x_0) \\
A_1 &= \frac{1}{2} \sum_{i=0}^1 f_i = \frac{f(x_0) + f(x_1)}{2} \Rightarrow 2A_1 = f(x_0) + f(x_1) \\
A_2 &= \frac{1}{3} \sum_{i=0}^2 f_i = \frac{f(x_0) + f(x_1) + f(x_2)}{3} \Rightarrow A_2 = \frac{2A_1 + f(x_2)}{3} \\
&\vdots \\
A_k &= \frac{kA_{k-1} + f(x_k)}{k+1}
\end{aligned}$$

O esquema com $C_k = A_k$, foi sugerido por Yu-Hong Dai. Em [5], a possibilidade de comparar o valor da função atual com uma média de M valores da função anterior também foi analisada. No entanto uma vez que M é fixo, nem todos os valores anteriores são uma média de funções em conjunto como em (3.5). Como mostraremos a seguir no Lema (3.1.1), para qualquer escolha de $\eta_k \in [0, 1]$, C_k fica entre f_k e A_k , o que implica que a atualização da busca linear está bem definida. Quando η_k se aproxima de 0 a busca linear se aproxima da busca linear monótona usual, e quando η_k tende para 1, o esquema torna-se mais não monótono, tratando de todos os valores anteriores da função com peso igual quando calculamos o custo C_k .

Lema 3.1.1. Se $\nabla f(x_k)^T d_k \leq 0$ para cada k , então para os iterados gerados pelo algoritmo de busca linear não monótona, temos $f_k \leq C_k \leq A_k$ para cada k . Além disso, se $\nabla f(x_k)^T d_k < 0$ e $f(x)$ é limitada inferiormente, então existe α_k que satisfaz as condições de Wolfe ou Armijo na atualização da busca linear.

Demonstração. Definindo $D_k : \mathbb{R} \rightarrow \mathbb{R}$ por:

$$D_k(t) = \frac{tC_{k-1} + f_k}{t+1}$$

Derivando $D_k(t)$, temos;

$$D'_k(t) = \frac{C_{k-1} - f_k}{(t+1)^2}$$

Desde que $\nabla f(x_k)^T d_k \leq 0$, segue de (3.3) que $f(x_{k+1}) \leq C_k$, o que implica que $D'_k(t) \geq 0$, para todo $t \geq 0$. Assim, D_k é não decrescente, e $f_k = D_k(0) \leq D_k(t)$, para todo $t \geq 0$. Em particular, tomando $t = \eta_{k-1}Q_{k-1}$, e por (3.5), obtemos:

$$D_k(\eta_{k-1}Q_{k-1}) = C_k$$

Logo:

$$f_k = D_k(0) \leq D_k(\eta_{k-1}Q_{k-1}) = C_k \quad (3.6)$$

Isso estabelece o limite inferior para C_k .

O limite superior $C_k \leq A_k$ será provado por indução. Para $k = 0$, isto acontece pela inicialização $C_0 = f(x_0)$. Suponhamos que $C_j \leq A_j$, para todo $0 \leq j < k$. Por (3.5), a inicialização $Q_0 = 1$, e o fato de que $\eta_k \in [0, 1]$, temos:

$$Q_1 = \eta_0 Q_0 + 1 = \eta_0 + 1.$$

$$Q_2 = \eta_1 Q_1 + 1 = \eta_1(\eta_0 + 1) + 1 = \eta_1 \eta_0 + \eta_1 + 1 = 1 + \sum_{i=0}^1 \prod_{m=0}^i \eta_{1-m}.$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$Q_{j+1} = 1 + \sum_{i=0}^j \prod_{m=0}^i \eta_{j-m} \leq j + 2 \quad (3.7)$$

Desde que D_k seja monótona não decrescente, (3.7), implica que:

$$C_k = D_k(\eta_{k-1}Q_{k-1}) = D_k(Q_k - 1) \leq D_k(k). \quad (3.8)$$

Usando a hipótese de indução temos, $C_{k-1} \leq A_{k-1}$, logo:

$$D_k(k) = \frac{kC_{k-1} + f_k}{k+1} \leq \frac{kA_{k-1} + f_k}{k+1} = A_k \quad (3.9)$$

Relacionando (3.8) e (3.9) $C_k \leq A_k$. Uma vez que ambas as condições padrão de Wolfe (1.4.2) e Armijo (1.4.1) podem ser satisfeitas quando $\nabla f(x_k)^T d_k < 0$ e $f(x)$ é limitada inferiormente, e uma vez que $f_k \leq C_k$, α_k pode ser escolhido para satisfazer quer a de Wolfe ou a busca linear de Armijo, condições do Algoritmo de busca linear não monótona. \square

3.2 Convergência Global

Para começar, vamos dar um limite inferior para o passo gerado pelo algoritmo de busca linear não monótona. Aqui e em outros lugares, $\|\cdot\|$ indica o Norma euclidiana, e $g_k = \nabla f(x_k)$.

Lema 3.2.1. Suponhamos que o algoritmo de busca linear não monótonos é utilizado no caso em que $g_k^T d_k \leq 0$ e ∇f satisfaz as seguintes condições de Lipschitz com a constante de Lipschitz L :

1. $\|\nabla f(x_{k+1}) - \nabla f(x_k)\| \leq L\|x_{k+1} - x_k\|$ se forem utilizadas as condições de Wolfe, ou
2. $\|\nabla f(x) - \nabla f(x_k)\| \leq L\|x - x_k\|$, para todo x no segmento linear que liga x_k e $x_k + \alpha_k \rho d_k$ se a condição de Armijo é usada e $\rho \alpha_k \leq \mu$.

Se as condições de Wolfe são satisfeitas, então

$$\alpha_k \geq \left(\frac{1 - \sigma}{L} \right) \frac{|g_k^T d_k|}{\|d_k\|^2}. \quad (3.10)$$

Se as condições de Armijo são satisfeitas, então

$$\alpha_k \geq \min \left\{ \frac{\mu}{\rho}, \left(\frac{2(1 - \delta)}{L\rho} \right) \frac{|g_k^T d_k|}{\|d_k\|^2} \right\}. \quad (3.11)$$

Demonstração. Consideramos os limites inferiores (3.10) e (3.11) nos dois casos seguintes.

Caso 1. Suponha que α_k satisfaz as condições Wolfe. Por (3.4), e subtraindo $\nabla f(x_k)^T d_k$ em ambos os lados, temos:

$$(g_{k+1}^T - g_k^T) d_k \geq (\sigma - 1) g_k^T d_k. \quad (3.12)$$

Visto que $g_k^T d_k \leq 0$ e $\sigma < 1$, então $(\sigma - 1) g_k^T d_k \geq 0$, e pela continuidade de Lipschitz de f , temos:

$$\begin{aligned} (\sigma - 1) \nabla g_k^T d_k &\leq (g_{k+1}^T - g_k^T) d_k \\ &\leq \|g_{k+1}^T - g_k^T\| \|d_k\| \\ &\leq L \alpha_k \|d_k\|^2 \end{aligned}$$

assim, obtemos:

$$\alpha_k \geq \frac{(1 - \sigma) |g_k^T d_k|}{L \|d_k\|^2}$$

o que implica em (3.10).

Caso 2. Suponha que α_k satisfaz as condições de Armijo. Se $\rho \alpha_k \geq \mu$, então obtemos $\alpha_k \geq \frac{\mu}{\rho}$ satisfazendo (3.11). De modo contrário, se $\rho \alpha_k < \mu$, então desde que h_k seja o maior inteiro tal que $\alpha_k = \bar{\alpha}_k \rho^{h_k}$ satisfaz (3.3).

Agora, note que $\rho\alpha_k = \bar{\alpha}_k\rho^{h_k+1}$ não satisfaz a condição de Armijo e desde que $f_k \leq C_k$, temos:

$$f(x_k + \rho\alpha_k d_k) > C_k + \delta\rho\alpha_k g_k^T d_k \geq f(x_k) + \delta\rho\alpha_k g_k^T d_k. \quad (3.13)$$

Pelo Teorema Fundamental do Cálculo (1.1.3), somando e subtraindo $\alpha g_k^T d_k$, temos:

$$g_{k+1} - g_k = \int_0^\alpha g_{k+1}^T d_k dt + \alpha g_k^T d_k - \alpha g_k^T d_k$$

Sendo $\alpha g_k^T d_k = \int_0^\alpha g_k^T d_k dt$, chegamos a:

$$g_{k+1} - g_k = \alpha g_k^T d_k + \int_0^\alpha [g_{k+1} - g_k]^T d_k dt$$

Pela continuidade de Lipschitz de ∇f ;

$$\begin{aligned} g_{k+1} - g_k &\leq \alpha g_k^T d_k + \int_0^\alpha Lt \|d_k\|^2 dt \\ &= \alpha g_k^T d_k + \frac{1}{2} L\alpha^2 \|d_k\|^2. \end{aligned}$$

Tomando $\alpha = \rho\alpha_k$ e combinando com (3.13), obtemos (3.11).

$$g_{k+1} - g_k \leq \rho\alpha_k g_k^T d_k + \frac{1}{2} L\rho^2 \alpha_k^2 \|d_k\|^2 \quad (3.14)$$

Combinando (3.13) com (3.14), chegamos a:

$$\begin{aligned} \delta\rho\alpha_k g_k^T d_k &\leq \rho\alpha_k g_k^T d_k + \frac{1}{2} L\rho^2 \alpha_k^2 \|d_k\|^2 \\ \frac{1}{2} L\rho^2 \alpha_k^2 \|d_k\|^2 &\geq \delta\rho\alpha_k g_k^T d_k - \rho\alpha_k g_k^T d_k \\ \frac{1}{2} L\rho^2 \alpha_k^2 \|d_k\|^2 &\geq (\delta - 1)\rho\alpha_k g_k^T d_k \end{aligned}$$

Como $g_k d_k \leq 0$, concluímos que:

$$\alpha_k \geq \frac{2(1 - \delta) |g_k^T d_k|}{L\rho \|d_k\|^2}$$

portanto reunindo os casos 1 e 2, obtemos (3.11). \square

O resultado de convergência global utiliza as seguintes hipóteses (veja por exemplo [6, 3]) com respeito a busca de direções.

Hipótese sobre a direção: Existem constantes positivas c_1 e c_2 , tal que:

$$g_k^T d_k \leq -c_1 \|g_k\|^2, \quad (3.15)$$

e

$$\|d_k\| \leq c_2 \|g_k\| \quad (3.16)$$

Para todo k suficientemente grande.

Teorema 3.2.2. Suponha que $f(x)$ é limitada inferiormente e que (3.15) e (3.16) são satisfeitas. Além disso, se forem utilizadas as condições de Wolfe suponhamos que ∇f é Lipschitz contínuo, com a constante de Lipschitz L , no conjunto de nível

$$\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}.$$

Seja $\bar{\mathcal{L}}$ o símbolo que denota o conjunto de $x \in \mathbb{R}^n$, cuja distância para \mathcal{L} é no máximo μd_{\max} onde $d_{\max} = \sup_k \|d_k\|$. Se forem utilizadas as condições de Armijo, suponhamos que $\nabla f(x)$ é Lipschitz contínuo, com a constante de Lipschitz L em $\bar{\mathcal{L}}$. Então os iterados x_k gerados pelo algoritmo de busca linear não monótona, satisfazem

$$\liminf_{k \rightarrow \infty} \|f(x_k)\| = 0. \quad (3.17)$$

Além disso, se $\eta_{\max} < 1$, então

$$\lim_{k \rightarrow \infty} \nabla f(x_k) = 0. \quad (3.18)$$

Assim, cada subsequência convergente de iterados se aproxima de um ponto x^* , onde $\nabla f(x^*) = 0$.

Demonstração. Mostraremos primeiro que:

$$f_{k+1} \leq C_k - \beta \|g_k\|^2 \quad (3.19)$$

onde,

$$\beta = \min \left\{ \frac{\delta \mu c_1}{\rho}, \frac{2\delta(1-\delta)c_1^2}{L\rho c_2^2}, \frac{\delta(1-\sigma)c_1^2}{Lc_2^2} \right\} \quad (3.20)$$

Caso 1: Se usarmos a condição de Armijo e $\rho\alpha_k \geq \mu$, então $\alpha_k \geq \frac{\mu}{\rho}$. Por (3.3) e (3.15), temos que:

$$f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k \nabla f(x_k)^T d_k \leq C_k - \delta \alpha_k c_1 \|g_k\|^2.$$

então:

$$f(x_{k+1}) \leq C_k - \delta \frac{\mu}{\rho} c_1 \|g_k\|^2.$$

Tomando $\beta < \delta \frac{\mu}{\rho} c_1$ chegamos em (3.19).

Caso 2: Se forem utilizadas as condições de Armijo e $\rho \alpha_k \leq \mu$, segue de (3.11) que,

$$\alpha_k \geq \left(\frac{2(1-\delta)}{L\rho} \right) \frac{g_k^T d_k}{\|d_k\|^2} \quad (3.21)$$

Por (3.3), e utilizando (3.20) temos:

$$f_{k+1} \leq C_k - \delta \left(\frac{2(1-\delta)}{L\rho} \right) \left(\frac{g_k^T d_k}{\|d_k\|} \right)^2 \quad (3.22)$$

Finalmente, por (3.15) e (3.16), temos:

$$f_{k+1} \leq C_k - \left(\frac{2\delta(1-\delta)c_1^2}{L\rho c_2^2} \right) \|g_k\|^2, \quad (3.23)$$

Tomando, $\beta < \left(\frac{2\delta(1-\delta)c_1^2}{L\rho c_2^2} \right)$, chegamos em (3.19).

Caso 3: Se forem utilizadas as condições de Wolfe, então a análise é semelhante ao caso 2 exceto que o limite inferior (3.21) passa a ter a expressão dada como em (3.10), daí segue que $-\alpha_k \leq -\left(\frac{1-\sigma}{L} \right) \frac{|g_k^T d_k|}{\|d_k\|^2}$. Segue de (3.3) e (3.15) que:

$$f_{k+1} \leq C_k - \delta \left(\frac{1-\sigma}{L} \right) \frac{c_1^2}{\|d_k\|^2} \|g_k\|^4$$

De (3.16), temos $\|d_k\|^2 \leq c_2^2 \|g_k\|^2$:

$$f(x_{k+1}) \leq C_k - \delta \left(\frac{(1-\sigma)c_1^2}{Lc_2^2} \right) \|g_k\|^2,$$

Tomando $\beta < \delta \left(\frac{(1-\sigma)c_1^2}{Lc_2^2} \right)$, chegamos em (3.19). Juntando os casos 1,2 e 3, obtemos (3.19) e (3.20).

Combinando a relação de atualização de custo (3.5) e o limite superior (3.19), temos:

$$\begin{aligned} C_{k+1} &= \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}} \\ &\leq \frac{C_k(\eta_k Q_k + 1) - \beta \|g_k\|^2}{Q_{k+1}} = C_k - \frac{\beta \|g_k\|^2}{Q_{k+1}} \end{aligned} \quad (3.24)$$

Como f é limitada inferiormente e $f_k \leq C_k$, para todo k , concluímos que C_k é limitada inferiormente. Segue-se a partir da equação (3.24) que:

$$C_{k+1} \leq C_k - \beta \frac{\|g_k\|^2}{Q_{k+1}} \leq C_{k-1} - \beta \frac{\|g_{k-1}\|^2}{Q_k} - \beta \frac{\|g_k\|^2}{Q_{k+1}} \leq \dots \leq C_0 - \beta \sum_{j=0}^k \frac{\|g_j\|^2}{Q_{j+1}}$$

Aplicando o limite quando $k \rightarrow \infty$, na expressão anterior, obtemos:

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^2}{Q_{k+1}} < \infty. \quad (3.25)$$

Se $\|g_k\|$ for limitada longe de 0, (3.25) seria violada visto que $Q_{k+1} \leq k+2$ por (3.7). Sendo, $\prod_{i=0}^j \eta_{k-i} = \eta_k \cdot \eta_{k-1} \cdot \dots \cdot \eta_{k-j}$, e $0 \leq \eta_{\min} \leq \eta_{\max} < 1$, então:

$$\prod_{i=0}^j \eta_{k-i} = \eta_k \cdot \eta_{k-1} \cdot \dots \cdot \eta_{k-j} \leq \eta_{\max} \cdot \eta_{\max} \cdot \dots \cdot \eta_{\max} = \eta_{\max}^{j+1}$$

Então por (3.7), e pelas propriedades de séries convergentes, temos:

$$Q_{k+1} = 1 + \sum_{j=0}^k \prod_{i=0}^j \eta_{k-i} \leq 1 + \sum_{j=0}^k \eta_{\max}^{j+1} \leq \sum_{j=0}^{\infty} \eta_{\max}^j = \frac{1}{1 - \eta_{\max}}. \quad (3.26)$$

Consequentemente, (3.25) implica em (3.18). □

Capítulo 4

Resultados Numéricos

Neste capítulo analisaremos os testes numéricos obtidos a partir dos algoritmos anteriormente apresentados, serão exemplos elementares com o propósito de melhorar a compreensão dos leitores. O primeiro algoritmo mostrado por Humes e Silva [7] e o segundo proposto por Zhang e Hager [19]. Ambos abordam problemas de minimização irrestrita com funções continuamente diferenciáveis. Utilizaremos para a implementação dos algoritmos o software Matlab, os testes foram realizados usando o compilador (64-bits), versão 7.7.0.471, em um Lenovo, Intel Core I3, 1.50GHz, memória RAM de 4Gb.

4.1 Algoritmos

4.1.1 Algoritmo do Ponto Proximal Inexato para Funções Diferenciáveis

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função continuamente diferenciável e convexa. Seja σ um número em $[0, 1]$.

1- Início: Seja x_0 algum ponto de \mathbb{R}^n .

1.1- Encontrar

$$x_{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ f(x) + \frac{1}{2} \|x - x_k\|^2 \right\} \quad (4.1)$$

Utilizando algoritmos clássicos para resolução do problema regularizado (4.1).

2- Calcular:

$$g_{k+1} = \nabla f(x_{k+1}) \quad e \quad e_k = g_{k+1} + x_{k+1} - x_k,$$

3- Verificar o critério de aceitação:

$$\|e_k\| \leq \sigma \max\{\|g_{k+1}\|, \|x_{k+1} - x_k\|\}$$

- 3.1- Caso seja aceito, fazer $k = k + 1$ e volta para (4.1).
 3.2- Caso não seja satisfeito . PARE.

4.1.2 Teste do Algoritmo do Método do Ponto Proximal Inexato

Análise I

A primeira função $f : \mathbb{R}^2 \Rightarrow \mathbb{R}$, que analisaremos será:

$$f(x) = 0,3x_1^2 + x_2^2 + 0,916x_1x_2 - 0,05x_1 + 0,01x_2$$

convexa e continuamente diferenciável com solução $\bar{x} = (0,32429, -0,143513)$ e determinante da matriz hessiana igual a 0,361, quase singular. Partindo do ponto inicial $x_1 = (100, 100)$ e $\sigma = 0.1$, e adotando $\nabla f(x) < 10^{-5}$. Realizaremos testes para o Método do Ponto Primal Inexato, aplicando o método do Gradiente Conjugado e o Método de Newton para resolução do problema interno.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100,100)	328,639	22156,0	141,309
17	(4,7869, -2,3605)	0,7361	1,8330	5,0025
33	(0,8010, -0,3900)	0,0818	0,0145	0,5562
50	(0,3508, -0,1674)	0,0079	-0,0081	0,0539
67	(0,3071, -0,1458)	0,0007	-0,0083	0,0052
83	(0,3029, -0,1438)	$8,5980 \cdot 10^{-5}$	-0,0083	$5,8207 \cdot 10^{-4}$
97	(0,3027, -0,1436)	$4,3502 \cdot 10^{-5}$	-0,0083	$2,9449 \cdot 10^{-4}$

Tabela 4.1: Iterações do Método do Ponto Proximal Inexato (MPPI) - Gradiente Conjugado

Abaixo veremos a interpretação geométrica da tabela apresentada, que utiliza o Método do Gradiente Conjugado para solucionar o problema interior do MPPI.

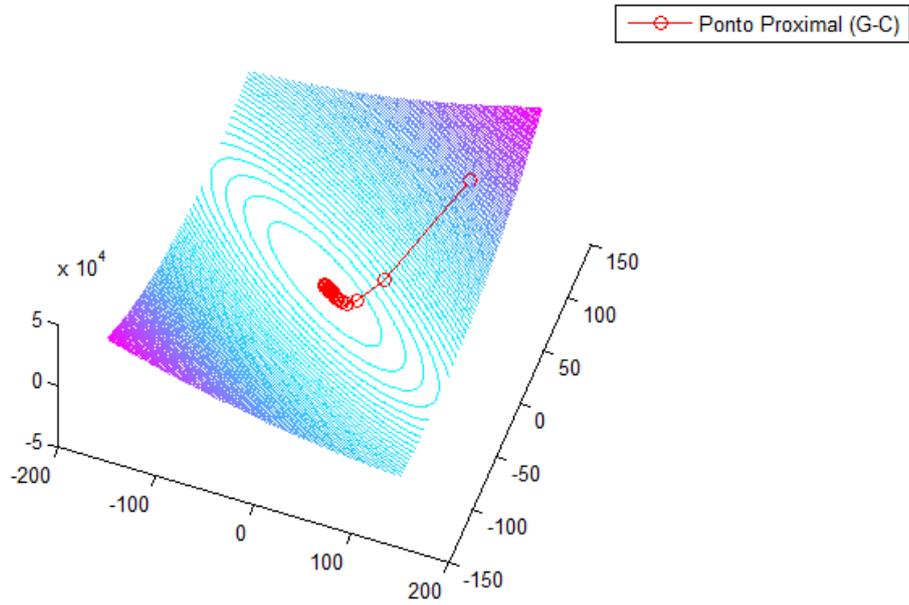


Figura 4.1: Gráfico do Algoritmo do Ponto Proximal Inexato- GC

Na tabela seguinte traçamos a solução do MPPI usando o Método de Newton para resolução do problema interno, com o mesmo ponto inicial $x_1 = (100, 100)$ e $\sigma = 0.1$.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100, 100)	328,6393	22156,0	141,30934
45	(0,3984,-0,191)	0,0158	-0,0074343	0,1071
90	(0,303,-0,144)	$3,2704 \cdot 10^{-5}$	-0,00828	0,00022
135	(0,30243,-0,1435)	$6,7863 \cdot 10^{-8}$	-0,00828	$4,6117 \cdot 10^{-7}$
180	(0,3024,-0,1435)	$1,4082 \cdot 10^{-10}$	-0,00828	$9,5694 \cdot 10^{-10}$
225	(0,3024, -0,1435)	$2,9224 \cdot 10^{-13}$	-0,008278	$1,9511 \cdot 10^{-12}$
269	(0,3024, -0,1435)	$5,8787 \cdot 10^{-16}$	-0,008278	$4,3298 \cdot 10^{-14}$

Tabela 4.2: Iterações do Método do Ponto proximal Inexato - Método de Newton

Posteriormente teremos a visão geométrica das informações apresentadas na tabela (4.2)

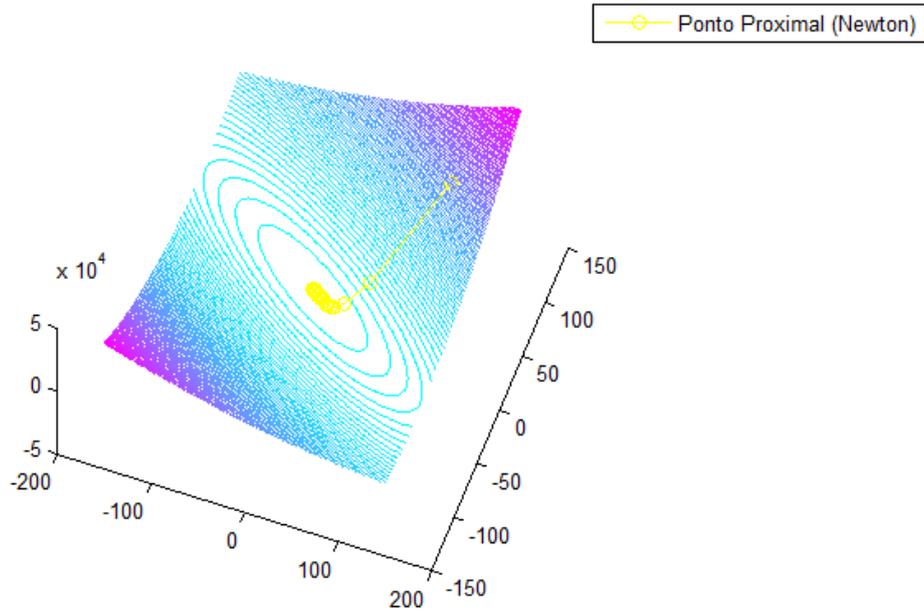


Figura 4.2: Gráfico do Algoritmo do Ponto Proximal Inexato- MN.

A seguir teremos o Método do Ponto Proximal Inexato com Método de Newton, utilizando como critério de parada $\nabla f(\bar{x}) < 10^{-5}$, obtemos:

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100, 100)	328,6393	22156,0	141,30934
17	(4,7868,-2,3605)	0,7361	1,8330	5,0025
34	(0,7371,-0,3584)	0,0713	0,0090	0,4849
50	(0,3508,-0,1674)	0,0079	-0,0081	0,0539
67	(0,3071,-0,1458)	$7,6893 \cdot 10^{-4}$	-0,0083	0,0052
83	(0,3030, -0,1438)	$8,5496 \cdot 10^{-5}$	-0,0083	$5,8100 \cdot 10^{-4}$
99	(0,3025, -0,1435)	$9,5061 \cdot 10^{-6}$	-0,0083	$6,4600 \cdot 10^{-5}$

Tabela 4.3: Iterações do Método do Ponto proximal Inexato - Método de Newton- critério de parada: tamanho do gradiente

Análise II

A segunda função que verificaremos será:

$$f(x) = 1,3x_1^2 + 0,4x_2^2 + 0,01x_1 - 0,3x_2 + 1,4x_1x_2$$

com solução $\bar{x} = (-3,5667, 6,6167)$ e o valor do determinante da hessiana igual a 0,12, considerado quase singular, a função é convexa e continuamente diferenciável como exigido no Método do Ponto Proximal Inexato (MPPI). Começando do ponto inicial $x_1 = (100, 100)$, $\sigma = 0,1$ e utilizando o Método do gradiente conjugado para resolução do problema interno.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100, 100)	456,3727	30971	44,0554
48	(-6,5508, 12,0826)	0,2221	-0,3187	6,0130
97	(-4,1025, 7,5981)	0,0399	-0,9880	1,0796
145	(-3,6663, 6,7992)	0,0074	-1,0096	0,2008
194	(-3,5846, 6,6494)	0,0013	-1,0103	0,0360
242	(-3,5700, 6,6228)	$2,4772 \cdot 10^{-4}$	-1,0103	0,0067
291	(-3,5675, 6,6182)	$6,1443 \cdot 10^{-5}$	-1,0103	0,0017

Tabela 4.4: Iterações do Método do Ponto proximal Inexato - Segunda função - Gradiente Conjugado

Adiante exibiremos o gráfico para melhor compreensão da tabela apresentada anteriormente.

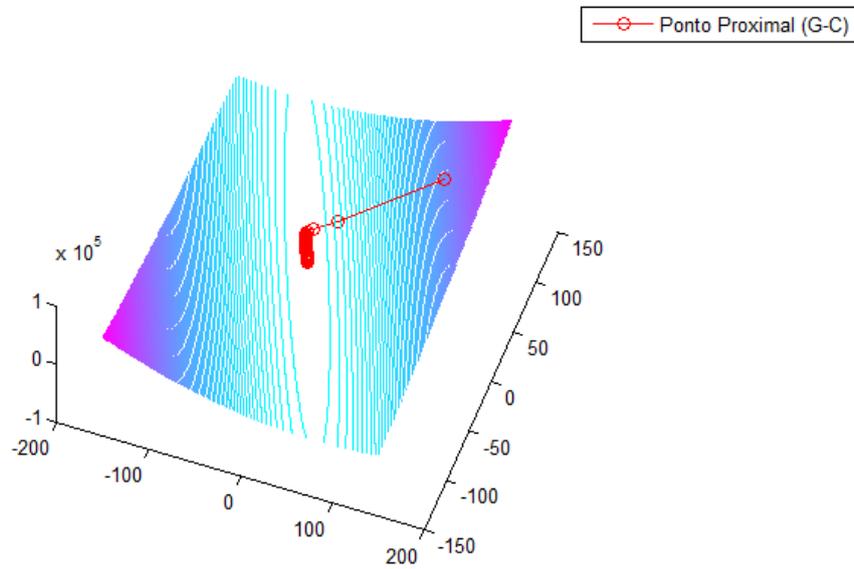


Figura 4.3: Gráfico do Algoritmo do Ponto Proximal Inexato- GC - 2º função.

De modo análogo ao da primeira função, veremos a análise do caso em que o Algoritmo do Ponto Proximal Inexato utiliza o Método de Newton para resolver o problema interno. Adotando o ponto inicial $x_1 = (100, 100)$ e $\sigma = 0, 1$.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100, 100)	456,3727535	30971	44,0554
113	(-3,8725, 7,1768)	0,0228	-1,0031	0,06162
226	(-3,5725, 6,6273)	$4,3382 \cdot 10^{-4}$	-1,0103	0,0117
339	(-3,5668, 6,6169)	$8,2675 \cdot 10^{-6}$	-1,0103	$2,2292 \cdot 10^{-4}$
452	(-3,5667, 6,6167)	$1,5756 \cdot 10^{-7}$	-1,0103	$3,3841 \cdot 10^{-6}$
565	(-3,5667, 6,6167)	$3,0026 \cdot 10^{-9}$	-1,0103	$7,9977 \cdot 10^{-7}$
678	(-3,5667, 6,6167)	$5,7222 \cdot 10^{-11}$	-1,0103	$7,9976 \cdot 10^{-7}$
791	(-3,5667, 6,6167)	$1,8440 \cdot 10^{-12}$	-1,0103	$7,9971 \cdot 10^{-7}$
904	(-3,5667, 6,6167)	$2,0701 \cdot 10^{-14}$	-1,0103	$7,9965 \cdot 10^{-7}$

Tabela 4.5: Iterações do Método do Ponto proximal Inexato - Método de Newton - 2º função

A seguir veremos a interpretação gráfica da tabela.

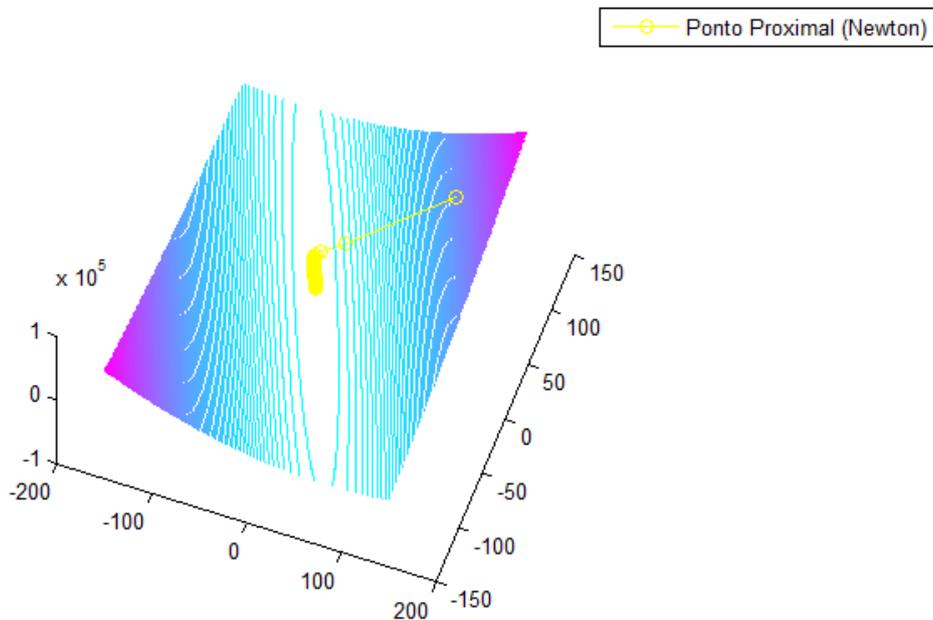


Figura 4.4: Gráfico do Algoritmo do Ponto Proximal Inexato- MN - 2º função.

Verificaremos posteriormente o Algoritmo do Ponto Proximal Inexato com Método de Newton para resolução do problema interno e com critério de parada $\nabla f(\bar{x}) < 10^{-5}$.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100, 100)	456,3727535	30971	44,0554
56	(-5,8212,10,7462)	0,1678	-0,6156	4,5428
110	(-3,9064,7,2389)	0,0253	-1,0014	0,6846
167	(-3,6128,6,7011)	0,0034	-1,0102	0,0929
223	(-3,5731,6,6285)	$4,8192 \cdot 10^{-4}$	-1,0103	0,0130
280	(-3,5675,6,6183)	$6,5372 \cdot 10^{-5}$	-1,0103	0,0018
334	(-3,5668,6,6169)	$9,8509 \cdot 10^{-6}$	-1,0103	$2,6579 \cdot 10^{-4}$

Tabela 4.6: Iterações do Método do Ponto proximal Inexato - Método de Newton- critério de parada: tamanho do gradiente- 2ª função

Nessas implementações do Algoritmo do Ponto Proximal Inexato, com o Método do Gradiente Conjugado e o Método de Newton para resolver o problema interno do algoritmo, utilizamos funções convexas e continuamente diferenciável. Obtivemos uma diferença considerável entre o número de iterações, e uma precisão satisfatória com relação a aproximação do ponto mínimo. Posteriormente veremos uma análise mais detalhada sobre os casos apresentados.

4.1.3 Algoritmo de Busca Linear Não Monótona

Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função continuamente diferenciável.

Inicialização: Escolha um ponto inicial x_0 , e os parâmetros $0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$, $0 < \delta < \sigma < 1 < \rho$ e $\mu > 0$. Usaremos $C_0 = f(x_0)$, $Q_0 = 1$, e $k = 0$.

Teste de Convergência: Se $\|\nabla f(x_k)\|$ suficientemente pequeno ($\|\nabla f(x_k)\| < 10^{-5}$), então PARE.

Atualização da Busca Linear: Definir $x_{k+1} = x_k + \alpha_k d_k$, onde α_k satisfaz tanto as condições (não-monótonas) de Wolfe:

$$f(x_k + \alpha_k d_k) \leq C_k + \delta \alpha_k \nabla f(x_k) d_k, \quad (4.2)$$

$$\nabla f(x_k + \alpha_k d_k) d_k \geq \sigma \nabla f(x_k) d_k \quad (4.3)$$

ou as condições (não monótonas) de Armijo: $\alpha_k = \overline{\alpha}_k \rho^{h_k}$, onde $\overline{\alpha}_k > 0$, é o passo da verificação e h_k é o maior inteiro, tal que (4.2) e possui $\alpha_k \leq \mu$.

Atualização de Custo: Escolha $\eta_k \in [\eta_{\min}, \eta_{\max}]$, e defina:

$$Q_{k+1} = \eta_k Q_k + 1, \quad C_{k+1} = \frac{(\eta_k Q_k C_k + f(x_{k+1}))}{Q_{k+1}} \quad (4.4)$$

Substituir k por $k + 1$ e retorna para o teste de convergência.

4.1.4 Teste do Algoritmo Busca Linear não Monótona

Utilizaremos inicialmente a função

$$f(x) = 0,3x_1^2 + x_2^2 + 0,916x_1x_2 - 0,05x_1 + 0,01x_2$$

continuamente diferenciável. Com o ponto inicial $x_1 = (100, 100)$, $\eta_{\min} = 0,2$, $\eta_{\max} = 0,8$, $\delta = 0,7$, $\sigma = 0,8$, $\rho = 3$ e $\mu = 2$.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100,100)	328,6393	22156,0	141,3093
21	(21,4678, 33,2491)	96,4339	1896,8	39,5353
41	(0,9579, 0,2533)	1,5860	0,5162	0,7661
61	(0,3431, -0,1570)	0,0157	-0,0081	0,0428
81	(0,30635, -0,1454)	0,0007	-0,0083	0,0044
101	(0,3028, -0,1438)	$6,0527 \cdot 10^{-5}$	-0,0083	0,0004
121	(0,3025, -0,1435)	$6,1800 \cdot 10^{-6}$	-0,0083	$3,317 \cdot 10^{-5}$
139	(0,3024, -0,1435)	$9,1881 \cdot 10^{-7}$	-0,0083	$3,7038 \cdot 10^{-6}$

Tabela 4.7: Iterações do Algoritmo Busca Linear não Monótona

Exibiremos a seguir o gráfico para explanarmos as informações adquiridas anteriormente.

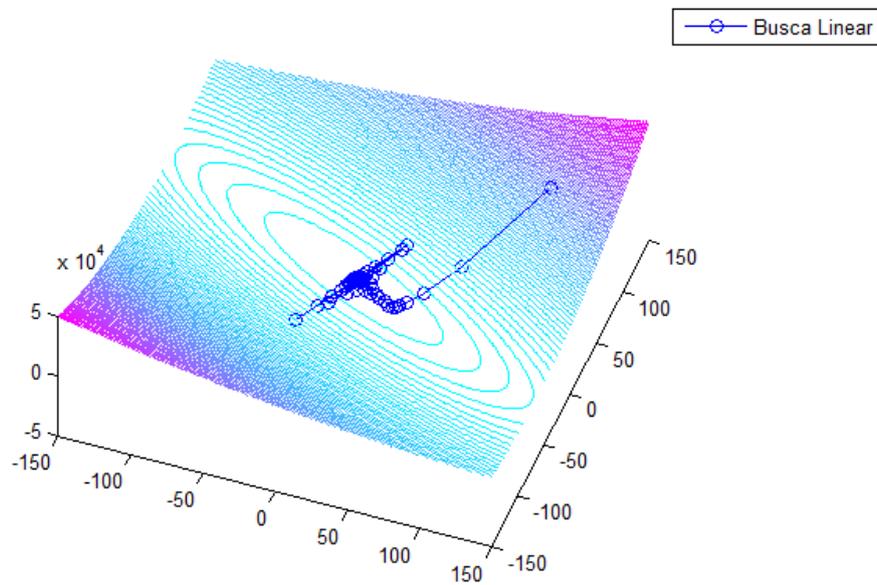


Figura 4.5: Gráfico do Algoritmo do Método de Busca Linear Não Monótona.

Agora iremos verificar o Método de busca linear não monótona na seguinte função:

$$f(x) = 1,3x_1^2 + 0,4x_2^2 + 0,01x_1 - 0,3x_2 + 1,4x_1x_2$$

que é continuamente diferenciável. Partiremos do ponto inicial $x_1 = (100, 100)$, $\eta_{\min} = 0,2$, $\eta_{\max} = 0,8$, $\delta = 0,7$, $\sigma = 0,8$, $\rho = 3$ e $\mu = 2$.

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(100, 100)	456,3728	30971	141,309
125	(3,1403, 12,0556)	28,5734	120,3702	8,6351
250	(-4,7398, 6,1007)	4,2958	1,7325	1,2815
380	(-3,7262, 6,5374)	0,5987	-0,9571	0,1781
500	(-3,5923, 6,6033)	0,0971	-1,0089	0,0289
633	(-3,5633, 6,6185)	0,0129	-1,0103	0,0038
758	(-3,5672, 6,6164)	0,0019	-1,0103	$5,7787 \cdot 10^{-4}$
879	(-3,5666, 6,6167)	$3,1056 \cdot 10^{-4}$	-1,0103	$9,2314 \cdot 10^{-5}$
999	(-3,5667, 6,6167)	$5,0368 \cdot 10^{-5}$	-1,0103	$1,4996 \cdot 10^{-5}$

Tabela 4.8: Iterações de Busca Linear não Monótona - 2º função

Posteriormente, veremos as informações obtidas expostas em gráfico.

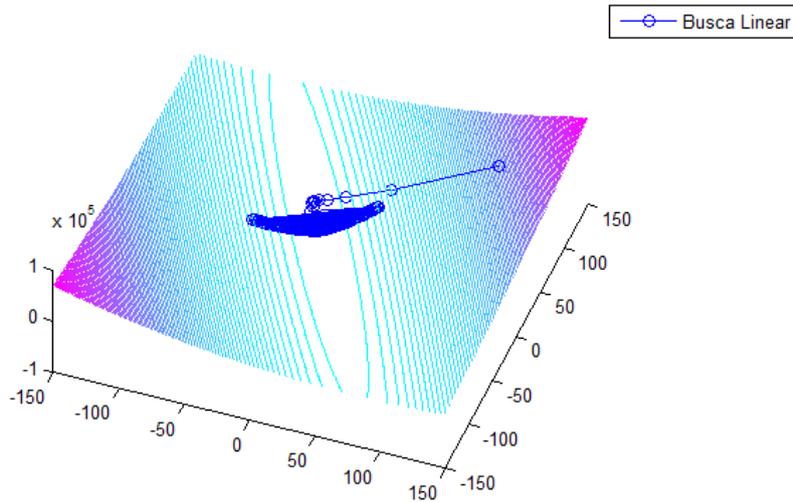


Figura 4.6: Gráfico do Algoritmo do Método de Busca Linear Não Monótona- 2º função.

4.1.5 Função de Rosenbrock

Para ilustrar melhor os exemplos apresentados utilizaremos o Algoritmo de Busca Linear não Monótona para encontrarmos os minimizadores da função de Rosenbrock, também conhecida como função banana, onde a solução é $\bar{x} = (1, 1)$:

$$f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

com ponto inicial $x_1 = (-2, -1)$, $\eta_{\min} = 0,1$, $\eta_{\max} = 0,5$, $\delta = 0,9$, $\sigma = 0,9$, $\rho = 3$ e $\mu = 2$

Número da Iteração k	x_k	$\ g_{k+1}\ $	$f(x_{k+1})$	$\ x_k - \bar{x}\ $
1	(-2, -1)	$4,1289 \cdot 10^{-3}$	2509	3,6056
97	(0,4217, 0,1728)	1,0621	0,3370	1,0093
194	(0,6457, 0,4157)	0,4604	0,1257	0,6834
290	(0,7257, 0,5258)	0,3572	0,0753	0,5478
387	(0,8021, 0,6418)	0,3430	0,0394	0,4092
484	(0,8461, 0,7152)	0,1577	0,0237	0,3237
581	(0,8720, 0,7589)	0,4091	0,0166	0,2294
677	(0,8932, 0,7969)	0,2018	0,0115	0,2294
714	(0,8989, 0,8075)	0,0995	0,0102	0,2174

Tabela 4.9: Função de Rosenbrock

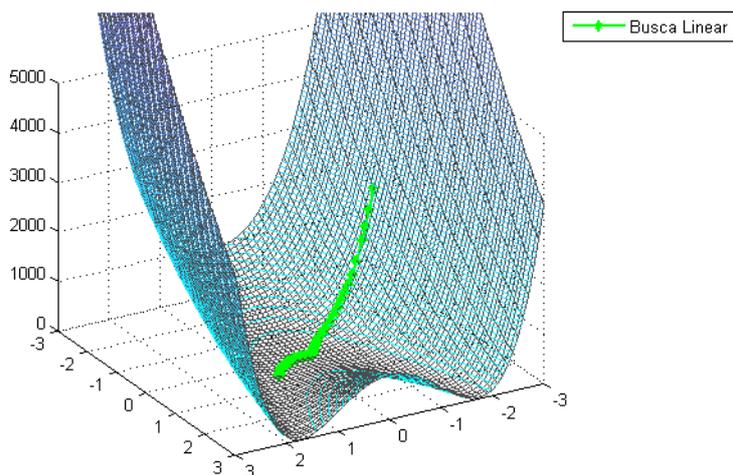


Figura 4.7: Gráfico Função de Rosenbrock.

4.1.6 Análise dos Resultados

Considerando as tabelas e gráficos apresentados anteriormente, observamos algumas características que serão descritas posteriormente, verificando as peculiaridades das funções objetivo, os valores ótimos e as velocidades de convergência das funções.

1. Na função $f(x) = 0,3x_1^2 + x_2^2 + 0,916x_1x_2 - 0,01x_1 + 0,01x_2$, não obtemos o mesmo ponto nos três casos estudados. Na implementação do Algoritmo do Ponto Proximal Inexato com o método do Gradiente Conjugado (APPIGC) para resolução do problema interno, foram 97 iterações em 0,0474 segundos para obter o ponto mais próximo do minimizador da função chegando a uma precisão de $2,9449 \cdot 10^{-4}$. Neste caso quando tentamos modificar o critério de parada do algoritmo, passando a utilizar o critério de parada clássico que provém do tamanho do gradiente (adotamos $\nabla f(\bar{x}) < 10^{-5}$), não obtemos sucesso pois, o algoritmo parou antes que o gradiente alcançasse o resultado. Ainda no Algoritmo do Ponto Proximal Inexato (APPIMN), agora com Método de Newton conseguimos chegar no ponto $x_k = (0,3024, -0,1435)$, próximo ao minimizador, em 0,0783 segundos e 269 iterações porém, neste caso quando utilizamos o gradiente como critério de parada alcançamos em 99 iterações o ponto $x_k = (0,3025, -0,1435)$. Na implementação do Algoritmo de Busca Linear não Monótona, em 139 iterações e 0,1066 segundos chegamos a um ponto próximo ao minimizador da função. Conseguimos verificar que para esta função o Método do Ponto Proximal Inexato (MPPIGC) é "melhor" em termos computacionais pois, chegou próximo ao ponto de mínimo em menor tempo e número de iterações, no segundo caso (APPIMN) o número de iterações foi superior aos demais e ainda assim o tempo foi menor que o do Algoritmo de Busca Linear Não monótona (ABLNM). O (APPIMN) foi o que teve maior precisão ($4,3298 \cdot 10^{-14}$) na distância entre os pontos encontrados em cada iteração e \bar{x} .
2. Com relação a segunda função, que possui o determinante da Hessiana mais próximo de zero (quase singular) que a mostrada anteriormente. Assim, $f(x) = 1,3x_1^2 + 0,4x_2^2 + 0,01x_1 - 0,3x_2 + 1,4x_1x_2$, não obtemos o mesmo ponto nas três simulações realizadas. Foram 291 iterações em 0,1452 segundos pelo Algoritmo do Ponto Proximal Inexato com Gradiente Conjugado, neste caso quando utilizamos o gradiente como critério de parada, não conseguimos êxito pois, o algoritmo para antes de obtermos o gradiente adotado. Com o Algoritmo do Ponto Proximal Inexato através do Método de Newton em 904 iterações e 0,3812 segundos che-

gamos ao ponto $x_k = (-3, 5667, 6, 6167)$, quando usamos o gradiente como critério de parada atingimos o ponto $x_k = (-3, 5668, 6, 6169)$ em 334 iterações. Já o Algoritmo de Busca Linear não Monótona fez em 0,6516 segundos 999 iterações. Novamente o (APPIGC) foi o "melhor" porém, percebemos que quando inserimos um outro critério de parada podemos ter uma diferença considerável no número de iterações. Para esta função a diferença no número de iterações entre os métodos é ampla e verificando as tabelas acima observamos que o APPIMN chega ao ponto em iterações anteriores mas o critério de parada só é satisfeito na última iteração.

3. Sabendo da importância da resolução de problemas clássicos de otimização, conseguimos também a implementação da Função de Rosembrock pelo Método de Busca Linear não Monótona, já que se trata de uma função continuamente diferenciável. O fato desta função não ser convexa não nos permite implementá-la no APPI. Não conseguimos chegar em um ponto próximo ao minimizador, das diversas implementações encontradas na internet sabemos que o ponto mínimo desse problema é $\bar{x} = (1, 1)$. Chegamos a esse ponto $x_k = (0, 8989, 0, 8075)$ em 0,2174 segundos e 714 iterações e com critério de parada $\nabla f(\bar{x}) < 10^{-1}$. Quando minimizamos o valor do gradiente no critério de parada conseguimos nos aproximar muito do ponto de mínimo porém, o número de iterações aumenta bastante.

A seguir, veremos a análise descrita acima em formas gráficas.

Algoritmos	Número de Iterações	Tempo de Convergência
Ponto Proximal Inexato- GC	97	0,0474
Ponto Proximal Inexato- MN	269	0,0783
Busca Linear não monótona	139	0,1066

Tabela 4.10: Comparação entre os métodos na função $f(x) = 0,3x_1^2 + x_2^2 + 0,916x_1x_2 - 0.05x_1 + 0.01x_2$.

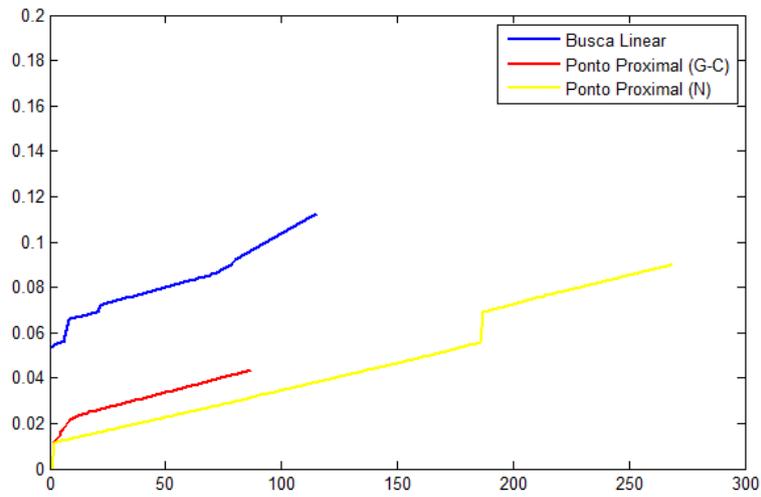


Figura 4.8: Gráfico do Confronto entre os Métodos.

Algoritmos	Número de Iterações	Tempo de Convergência
Ponto Proximal Inexato- GC	291	0,1452
Ponto Proximal Inexato- MN	904	0,3812
Busca Linear não monótona	999	0,6516

Tabela 4.11: Comparação entre os métodos na função $f(x) = 1, 3x_1^2 + 0, 4x_2^2 + 0, 01x_1 - 0, 3x_2 + 1, 4x_1x_2$.

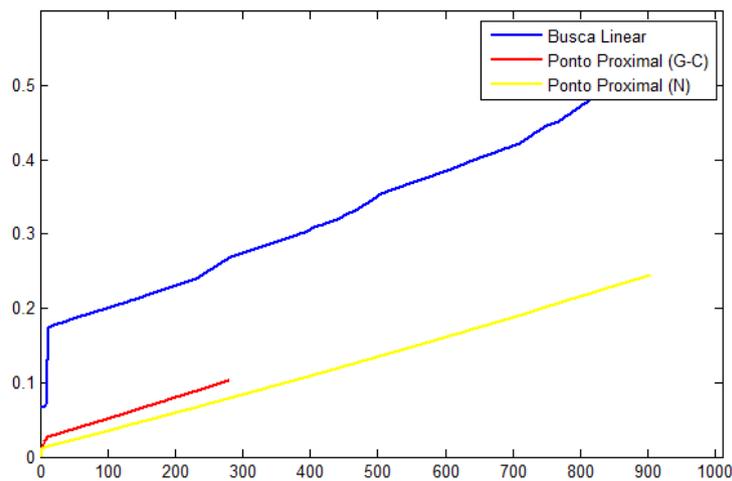


Figura 4.9: Gráfico do Confronto entre os Métodos - 2º função.

Capítulo 5

Considerações Finais

5.1 Conclusão

Os problemas de minimização irrestrita, tem o objetivo de minimizar uma função de variáveis reais a um valor real sem que haja restrições aos valores das variáveis.

Neste trabalho abordamos dois métodos irrestritos, onde o primeiro obtém a direção de descida a partir de propriedades do triângulo, e minimiza funções continuamente diferenciáveis e convexas. E o segundo trata da busca linear onde podem acontecer crescimento do valor da função objetivo antes de obtermos o valor ótimo. Em ambos algoritmos estudados obtivemos resultados relevantes. O Algoritmo do Ponto Proximal Inexato é desenvolvido a partir de hipóteses bem específicas o que torna mais difícil a obtenção de funções para implementação nesse método, assim constatamos sua importância para determinados tipos de funções. Já o Algoritmo de Busca Linear não monótona pode ser utilizado para funções mais simples. Por esse motivo conseguimos a implementação de um problema de otimização clássico utilizando o ABLNM.

Realizamos os testes utilizando o software Matlab, e assim conseguimos observar o desempenho dos dois algoritmos apresentados, suas características individuais e suas semelhanças. Desta forma sabemos que os métodos apresentados ainda tem muito a ser estudado.

5.2 Sugestões para trabalhos futuros

Com o objetivo prosseguir com o estudo dos referidos métodos, aprimorando os problemas apresentados e explorando casos não analisados nesta pesquisa. Buscando as melhores formulações para as questões abordadas.

- Buscar problemas clássicos com funções que se ajustam as hipóteses do Algoritmo do Ponto Proximal Inexato.
- Implementar o Algoritmo do Ponto Proximal Inexato para funções não diferenciáveis.
- Implementar o Algoritmo do Ponto Proximal Inexato utilizando outros algoritmos para resolução dos problemas internos.
- Implementar o Algoritmo de Busca Linear não monótona utilizando outras funções.

Referências Bibliográficas

- [1] ALVES, A.R.; KARAS, E.W. **Um Curso de Otimização**, Curitiba, 2011.
- [2] BAPTISTE, J; URRUTY, H; LEMARÉCHAL, C. **Fundamental of Convex Analysis**, Springer- Verlag Berlin Heidelberg GmbH, New York, pp: 201-205, 2001.
- [3] DAI, Y.H. **On the nonmonotone line search**, J. Optim. Theory Appl., 112, pp. 315–330 (2002).
- [4] GODOI, G. H. **Proximal Inexato Tipo Descida para Otimização Suave**, 2013, 59 P, Dissertação (Mestrado em Matemática), Otimização, Instituto de Matemática e Estatística, Universidade Federal de Goiás.
- [5] GRIPPO, L.; LAMPARIELLO, F.; LUCIDI. S. **A class of nonmonotone stabilization methods in unconstrained optimization**, Numer. Math., 59 (1991), pp. 779–805.
- [6] GRIPPO, L.; LAMPARIELLO, F.; LUCIDI. S. **A nonmonotone line search technique for Newton’s method**, SIAM J. Numer. Anal., 23 (1986), pp. 707–716.
- [7] HUMES, C.; SILVA, P.J.S. **Inexact proximal point algorithms and descent methods in optimization**. Optim. Eng., pp:257-271, 2005.
- [8] IZMAILOV, A.; SOLODOV, M. **Otimização - volume 1. Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade** 2.ed. Rio de Janeiro: IMPA, 2009.
- [9] IZMAILOV, A.; SOLODOV, M. **Otimização - volume 2. Métodos Computacionais** 2.ed. Rio de Janeiro: IMPA, 2009.
- [10] LIMA, E.L. **Análise Real - volume 2**. 5. ed.- Rio de Janeiro, IMPA, 2010, (Coleção matemática universitária).

- [11] LIMA, E.L. **Curso de Análise - volume 1**. 12. ed.- Rio de Janeiro, IMPA, 2008, (Projeto Euclides).
- [12] LIMA, E.L. **Curso de Análise - volume 2**. 11. ed.- Rio de Janeiro, IMPA, 2012, (Projeto Euclides).
- [13] LOPES, J.O. **Métodos de Ponto Proximal em Otimização** . Departamento de Matemática, Universidade Federal do Piauí, Abril-2006.
- [14] MARTINET, B. **Régularisation d'inéquations variationnelles par approximation sucessives** . Rev. Fr. Inform. Rech. Oper.,R3:154-158, 1970.
- [15] **MATLAB**. The Language of Technical Computing. 2015. <http://www.mathworks.com/products/matlab/>. Acessado em maio de 2015.
- [16] NETO, G.C. **Método para Solução de Problemas de Otimização Multiobjetivo em Ambientes Fuzzy**,83p , Dissertação (Mestrado em Matemática) , Otimização - Instituto de Ciências Exatas, Universidade Federal do Amazonas
- [17] ROCKFERLLAR, R.T, **Augmented lagrangians and applications of the proximal point algorithm in convex programming**. Math. Oper. Res.,1:97-116,1976 .
- [18] TOINT, P. L. **An assessment of non-monotone line search techniques for unconstrained optimization**, SIAM J. Sci. Comput., 17 (1996), pp. 725–739.
- [19] ZHANG, H. HAGER, W. **A Nonmonotone Line Search Technique and its Application to Unconstrained Optimization**, SIAM J. 2004, pp. 1043–1056.