Universidade Federal de Goiás
Instituto de Matemática e Estatística

Kelvin Rodrigues Couto

Método do Gradiente para Funções Convexas Generalizadas

Goiânia
2009

Termo de Ciência e de Autorização para Disponibilizar as Teses e Dissertações Eletrônicas (TEDE) na Biblioteca Digital da UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás –UFG a disponibilizar gratuitamente através da Biblioteca Digital de Teses e Dissertações – BDTD/UFG, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: [x] Dissertação [ ] Tese

2. Identificação da Tese ou Dissertação

Autor(a):	Kelvin Rodrigues Couto				
CPF:		E-mail:	Kelvimrodrigues@yahoo.com		
Seu e-mail pode ser disponibilizado na página? [x]Sim [ ] Não					
Vínculo Empregatício do autor					
Agência de fomento:	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior	Sigla:	CAPES		
País:	Brasil	UF:	GO	CNPJ:	00889834/0001-08
Título:	Método do Gradiente para Funções Convexas Generalizadas				
Palavras-chave:	Método do gradiente, método do gradiente projetado, funções convexas generalizadas.				
Título em outra língua:	Gradiente method for generalized convex functions				
Palavras-chave em outra língua:	Gradient method, gradient projection method, generalized convex functions				
Área de concentração:	Matemática Aplicada				
Data defesa: (dd/mm/aa)	16/12/2009				
Programa de Pós-Graduação:	Mestrado				
Orientador(a):	Dr. Orizon Pereira Ferreira				
CPF:		E-mail:	orizon@mat.ufg.br		

3. Informações de acesso ao documento:

Liberação para disponibilização?[1] [x] total [ ] parcial

Em caso de disponibilização parcial, assinale as permissões:

[ ] Capítulos. Especifique: ____

[ ] Outras restrições: _____Gostaria que não fosse divulgado os anexos.

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC da tese ou dissertação.

O Sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses e ou dissertações, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.

____

Assinatura do(a) autor(a) Data: 23/12/2009

[1] Em caso de restrição, esta poderá ser mantida por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Todo resumo e metadados ficarão sempre disponibilizados.

KELVIN RODRIGUES COUTO

Método do Gradiente para Funções Convexas Generalizadas

Dissertação apresentada ao Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de concentração: Matemática Aplicada.

Orientador: Prof. Dr. Orizon Pereira Ferreira

Goiânia
2009

Dados Internacionais de Catalogação-na-Publicação (CIP)

(GPT/BC/UFG)

Couto, Kelvin Rodrigues..
C871m Método do gradiente para funções convexas generalizadas [manuscrito] / Kelvin Rodrigues Couto. – 2009.
75f.

Orientador: Prof. Dr. Orizon Pereira Ferreira.
Dissertação (Mestrado) – Universidade Federal de Goiás, Instituto de Matemática e Estatística, 2009.
Bibliografia.

1. Algoritmo - Gradiente, Método do. 2. Método do gradiente Projetado. 3. Funções convexas. 4. Otimização matemática I. Título.

CDU: 510.51

KELVIN RODRIGUES COUTO

Método do Gradiente para Funções Convexas Generalizadas

Dissertação defendida no Programa de Pós–Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Matemática, aprovada em 16 de dezembro de 2009, pela Banca Examinadora constituída pelos professores:

Prof. Dr. Orizon Pereira Ferreira
Instituto de Matemática e Estatística – UFG
Presidente da Banca

Prof. Dr. Paulo Roberto Oliveira
COPPE – UFRJ

Prof. Dr. José Yunier Bello Cruz
Instituto de Matemática e Estatística – UFG

Kelvin Rodrigues Couto

Graduou–se em Bacharel em Matemática pela UFG - Universidade Federal de Goiás. Durante a graduação foi monitor de Cálculo I, Cálculo II e Álgébra I, e foi bolsista do programa institucional de bolsas de iniciação científica/CNPq. Durante o Mestrado foi bolsista da CAPES. Atualmente é professor substituto no Instituto de Matemática e Estatística – UFG

Às pessoas que mais amo:
meus pais, minha irmã
e minha namorada.

Agradecimentos

Primeiramente a Deus e toda sua força divina, por me dar saúde, condições de estudar e por me cercar de tantas pessoas queridas que me auxiliaram muito. Dedico este trabalho a todos aqueles que me auxiliaram na minha jornada de estudos. Agradeço a minha família, minha mãe Lucilene, meu pai Elísio e minha irmã Katileen por sempre terem incentivado e apoiado meus estudos. Sou muito grato a minha namorada Natália por ter tido paciência e compreensão durante minhas jornadas de estudo.

Em especial ao professor Dr. Orizon Pereira Ferreira pela honra de ser seu orientando, pelo seu tempo dedicado a mim, sempre com muito profissionalismo, seriedade, paciência, generosidade, competência e dedicação em dividir seu conhecimento comigo.

Ao professor Glaydston de Carvalho Bento, que é uma pessoa maravilhosa e foi quem intermediou o meu primeiro contato com o professor Orizon.

Ao professor Dr. Geci José Pereira da Silva e à professora Dra. Rosely Maria Barbosa Goes, por proporcionarem o meu primeiro contato com a otimização durante a iniciação científica.

A todos que foram meus professores eu sou muito grato. Assim como sou grato a todos os funcionários do IME, que são fundamentais para o funcionamento da estrutura da graduação e da pós-graduação da Matemática na UFG.

A todos meus colegas de curso do mestrado e graduação, com os quais aprendi muito e fiz grandes amizades.

À (CAPES) pela bolsa de estudos, que me auxiliou muito no desenvolvimento do trabalho[1].

Em especial aos professores Dr. Paulo Roberto Oliveira, Dr. José Yunier Bello Cruz e Dr. Luis Román Lucambio Pérez, por me darem a honra de suas participações na banca, por todo tempo dedicado a leitura da dissertação e sugestões para sua melhoria.

[1] Este trabalho teve suporte financeiro da CAPES.

Resumo

Couto, Kelvin Rodrigues. **Método do Gradiente para Funções Convexas Generalizadas**. Goiânia, 2009. 74p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Neste trabalho trataremos da convergência do método do gradiente para minimizar funções continuamente diferenciáveis e convexas-generalizadas, isto é, pseudo-convexas ou quase-convexas. Veremos que sob certas condições o método do gradiente, assim como o método do gradiente projetado, gera uma sequência que converge para minimizador quando existe um e a função objetivo é pseudo-convexa. Quando a função objetivo é quase-convexa a sequência gerada converge para um ponto estacionário do problema quando existe um tal ponto.

Palavras–chave

Método do Gradiente, Método do Gradiente Projetado, Funções Convexas Generalizadas.

Abstract

Couto, Kelvin Rodrigues. **Gradient Method for Generalized Convex Functions**. Goiânia, 2009. 74p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

The Convergence theory of gradient method and gradient projection method, for minimization of continuously differentiable generalized convex functions, that is, pseudoconvex functions and quasiconvex functions is studied in this work. We shall see that under certain conditions the gradient method, as well as gradient projection method, generate a convergent sequence and the limit point is a minimizing, whenever the function has minimizing and is pseudoconvex functions. If the objective function is quasiconvex then the generated sequence converges to a stationary point whenever that point exists.

Keywords

Gradient Method, Gradient Projection Method, Generalized Convex Functions.

Sumário

CAPÍTULO 1

Introdução

Consideremos uma função $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável. Trataremos inicialmente, do seguinte problema irrestrito:

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{1-1}$$

Uma idéia natural para resolver esse problema, seria encontrar um $x^{k+1} \in \mathbb{R}^n$ a partir de $x^k \in \mathbb{R}^n$ de tal modo que:

$$f(x^{k+1}) < f(x^k), \qquad k = 0, 1, \ldots.$$

Uma estratégia natural para resolver (1-1) seria encontrar uma sequência $\{x^k\}$ de pontos em $\mathbb{R}^n$, de tal forma que a função objetivo f seja decrescente. Considerando-se que na direção $-\nabla f(x^k)$, a função f é decrescente a partir de x^k nessa direção, pelo menos para passos curtos $t_k > 0$, isto nos sugere a seguinte iteração:

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

É exatamente na iteração (1-2) que consiste o método do gradiente. Observemos que o método gera uma sequência de pontos $\{x^k\}$ em $\mathbb{R}^n$ de tal forma que $f(x^{k+1}) < f(x^k)$, ou seja, o método gera uma sequência de pontos $\{x^k\}$ em $\mathbb{R}^n$ de tal forma que ocorra um decrescimento da função objetivo, quando aplicada na sequência $\{x^k\}$.

O método do gradiente, também conhecido como método de Cauchy, é um dos métodos mais antigos e mais simples para minimizar uma função $f : \mathbb{R}^n \to \mathbb{R}$. Observemos que, até o momento, só sabemos que o método busca pontos, tais que ocorra um decrescimento da função objetivo, a princípio não temos nenhuma garantia que a sequência de pontos $\{x^k\}$ em $\mathbb{R}^n$ converge para a solução de (1-1), se a solução existir. É importante percebermos que para cada forma de calcular o comprimento do passo t_k, gera-se uma sequência $\{x^k\}$ diferente. Assim, a convergência do método está relacionada com a forma com que o comprimento do passo t_k é escolhido, além, é claro, com o tipo de função objetivo que estamos trabalhando.

Veremos no Capítulo 4 que para uma escolha de passo t_k que satisfaça a certas condições propostas em [7], garantiremos que se o problema (1-1) tiver solução então o método do gradiente convergirá para um ponto estacionário do problema em questão, no caso em que a função objetivo f é quase-convexa.

Para o caso em que nosso problema é restrito, ou seja, trataremos da minimização de uma função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável sobre um conjunto $D \in \mathbb{R}^n$, nosso problema pode ser proposto como:

$$\min f(x), \quad D \in \mathbb{R}^n. \tag{1-3}$$

Para os problemas restritos (1-3) estudaremos o método do gradiente projetado, que foi originalmete estabelecido por Goldstein [3] e por Levitin e Polyak [8]. O método consiste em definir uma sequência de pontos $\{x^k\}$ como:

$$x^{k+1} = x_k(t_k) := P(x^k - t_k \nabla f(x^k)), \tag{1-4}$$

onde $P(\cdot)$ denota a projeção sobre D, isto é, $P(x^k - t_k \nabla f(x^k))$ é o ponto pertencente a D que está mais próximo de $x^k - t_k \nabla f(x^k)$. Observemos que as iterações para o método do gradiente projetado são uma combinação natural do método do gradiente para problemas irrestritos, com a projeção das iterações obtidas sobre o conjunto viável $D \subset \mathbb{R}^n$ em questão que, no nosso caso, é um conjunto não-vazio, fechado e convexo.

Veremos, no Capítulo 5, que para escolha de passo t_k descrita em [12], a sequência $\{x^k\}$, gerada pelo método do gradiente projetado, convergirá para solução do problema (1-3), caso ela exista, f seja pseudo-convexa e os passos t_k sejam limitados superiormente. E no caso em que f for quase-convexa, veremos que a sequência $\{x^k\}$, gerada pelo método do gradiente projetado, convergirá para um ponto estacionário do problema (1-3), caso exista solução.

A respeito da organização deste trabalho, ele está organizado da seguinte forma:

Com o objetivo de tornar este trabalho auto-contido, no Capítulo 2 trataremos de conceitos básicos para a compreensão do trabalho, tais como: resultados de análise na Seção 2.1, resultados de otimização na Seção 2.2, análise convexa na Seção 2.3 e resultados sobre projeções na Seção 2.4. Os resultados de análise serão divididos em duas partes: na primeira trataremos de análise real e na segunda parte trataremos de análise em $\mathbb{R}^n$; os resultados de otimização são limitados a definições básicas como minimizadores e pontos estacionários; a Seção 2.3 é fundamental, pois definiremos funções convexas, quase-convexas, pseudo-convexas e resultados sobre essas classes. Toda essa discussão será amplamente utilizada nos Capítulos 4 e 5; a Seção 2.4 tratará de algumas das principais propriedades do operador projeção, e será exigida no tratamento do Capítulo 5.

O Capítulo 3 servirá de preparação para a leitura dos Capítulos 4 e 5, pois dará uma visão preliminar sobre o método do gradiente e sobre o método do gradiente projetado. No Capítulo 3 veremos que o método do gradiente é um tipo de método de descida, além disso, ficará claro que cada forma de calcular o comprimento do passo t_k, gera uma sequência $\{x^k\}$ diferente, pois discutiremos algumas buscas lineares, tanto para o caso de problemas irrestritos (tratados pelo método do gradiente), quanto para problemas restritos (tratados pelo método do gradiente projetado).

E como já foi dito anteriormente, estudaremos no Capítulo 4 algumas condições sob as quais a sequência $\{x^k\}$, gerada pelo método do gradiente, convergirá para um ponto estacionário do problema (1-1), quando a função objetivo f é quase-convexa. Além disso, veremos que o tipo de busca linear definida em [7] por Kiwiel e Murty, generaliza vários outros tipos de busca linear, como por exemplo: a busca linear definida pela regra de Armijo, as buscas linares definidas por Burachik *at al.* em [1] e a busca proximal definida no artigo [4] de Iusem e Svaiter.

No Capítulo 5, como já foi dito anteriormente, estudaremos algumas condições sob as quais a sequência $\{x^k\}$, gerada pelo método do gradiente projetado, convergirá para um ponto solução do problema (1-3), quando a função objetivo f é pseudo-convexa. Estudaremos também algumas condições sob as quais a sequência $\{x^k\}$, gerada pelo método do gradiente projetado, convergirá para um ponto estacionário do problema (1-3), caso a função objetivo f seja quase-convexa.

E por fim no Capítulo 6 faremos nossas considerações finais.

CAPÍTULO 2

Conceitos básicos

Neste capítulo faremos uma breve revisão sobre alguns conceitos, que serão exigidos para leitura deste trabalho, para isso dividimos este primeiro capítulo em quatro seções, nas quais trataremos, respectivamente, de conceitos de básicos de análise, otimização, análise convexa e projeções.

2.1 Resultados básicos de Análise

Com esta primeira seção queremos fazer um breve revisão sobre alguns conceitos de análise, para isso a dividimos em duas subseções. Na primeira subseção trataremos de conceitos relativos à análise real e na segunda abordaremos conceitos sobre análise no espaço euclidiano $\mathbb{R}^n$.

2.1.1 Resultados e conceitos de Análise real

Com o objetivo de tornar este trabalho auto-contido indicaremos, nesta subseção, alguns resultados básicos de análise real. Estes resultados são de caráter introdutório e podem ser encontrados em muitos livros de introdução à análise real, por exemplo em [9].

Considerando $X \subset \mathbb{R}$, dizemos que o número real a é um ponto de acumulação do conjunto X se para todo $\varepsilon > 0$, o intervalo $(a-\varepsilon, a+\varepsilon)$ contém algum ponto $x \in X$ diferente de a.

Seja $f : X \subset \mathbb{R} \to \mathbb{R}$ uma função real e a um ponto de acumulação de X, diremos que o número real L é o limite de $f(x)$ quando x tende para a para significar que: dado arbitrariamente $\varepsilon > 0$, sempre podemos encontrar $\delta > 0$ de tal modo que tenhamos $|f(x) - L| < \varepsilon$, sempre que $x \in X$ e $0 < |x-a| < \delta$. Adotaremos a notação:

$$\lim_{x \to a} f(x) = L$$

Estudando sobre limites de funções podemos verificar, dentre outros resultados importantes, o seguinte teorema.

Teorema 2.1 *Sejam a um ponto de acumulação de $X \subset \mathbb{R}$ e funções reais f, $g : X \to \mathbb{R}$. Se $\lim_{x\to a} f(x) = L$ e $\lim_{x\to a} g(x) = M$ com $L < M$, então existe $\delta > 0$ tal que:*

$$x \in X,\ 0 < |x-a| < \delta \Rightarrow f(x) < g(x).$$

Prova. Veja página 155 de [9]. □

Como consequência imediata do teorema anterior temos o seguinte corolário.

Corolário 2.2 (Permanência do sinal) *Sejam a um ponto de acumulação de $X \subset \mathbb{R}$ e $f : X \to \mathbb{R}$. Se $\lim_{x\to a} f(x) = L$ com $L < 0$, então existe $\delta > 0$ tal que:*

$$x \in X,\ 0 < |x-a| < \delta \Rightarrow f(x) < 0.$$

Prova. Basta fazer $g(x) = 0$ no último teorema, que a prova segue de forma direta. □

Uma sequência de números reais é formada pelo conjunto das imagens de uma função $x : \mathbb{N} \to \mathbb{R}$, definida no conjunto $\mathbb{N} = \{1,2,\ldots\}$ dos números naturais e tomando valores reais. O valor $x(k)$, para todo $k \in \mathbb{N}$, é chamado de k-ésimo termo da sequência e será representado por x_k. Quando nos referirmos à sequência x usaremos a notação $\{x_k\}$. A seguir definiremos um conjunto limitado.

Dizemos que uma sequência real $\{x_k\}$ é limitada quando o conjunto de seus termos é um conjunto limitado, isto é, existe $M \in \mathbb{R}$ tal que $|x_k| < M$ para todo $k \in \mathbb{N}$. E dizemos que uma sequência real é convergente ao ponto a quando para valores grandes de k, podemos garantir que os termos x_k tornam-se e se mantêm tão próximos de a quanto se deseje. Assim dizer que a sequência real $\{x_k\}$ converge ao número a significa que: para todo $\varepsilon > 0$ existe k_0 natural tal que $|x_k - a| < \varepsilon$, sempre que $k > n_0$. Em notação padrão temos

$$\lim_{k\to\infty} x_k = a.$$

Pela discussão acima pode ser feita de forma imediata a verificação dos seguintes teoremas.

Teorema 2.3 *O limite de uma sequência real $\{x_k\}$ convergente é único.*

Prova. Veja página 85 de [9]. □

Teorema 2.4 *Toda sequência real $\{x_k\}$ convergente é limitada.*

Prova. Veja página 86 de [9]. □

Uma subsequência da sequência real $\{x_k\}$ é a restrição da função x a um subconjunto infinito $\mathbb{N}' = \{k_1, k_2, \ldots\}$ de $\mathbb{N}$. Usaremos a notação x_{k_j} com $j = 1,2,\ldots$, para indicar a subsequência da sequência $\{x_k\}$. Dizemos que uma sequência $\{x_k\}$ é crescente quando $x_k < x_{k+1}$ para todo $k \in \mathbb{N}$, se pudermos afirmar que $x_k \leq x_{k+1}$ para todo $k \in \mathbb{N}$ então diremos que $\{x_k\}$ é uma sequência não-decrescente. De forma análoga definimos sequência decrescente e sequência não-crescente. Uma sequência que é crescente ou decrescente ou não-crescente ou não-decrescente, é dita uma sequência monótona. Com estas definições em mãos juntamente com as discussões anteriores, pode-se verificar os seguintes teoremas.

Teorema 2.5 *Se uma sequência de números reais $\{x_k\}$ converge para o número L, então toda subsequência de $\{x_k\}$ também converge para L.*

Prova. Veja página 85 de [9]. □

Teorema 2.6 *Uma sequência real, monótona e limitada é convergente.*

Prova. Veja página 86 de [9]. □

Teorema 2.7 *Se uma sequência monótona de números reais $\{x_k\}$ possui uma subsequência convergente, então $\{x_k\}$ é convergente.*

Prova. Veja página 86 de [9]. □

Para encerrar a discussão sobre sequências reais, enunciaremos um resultado bastante conhecido, e que será utilizado em demonstrações de fatos importantes em capítulos subsequentes.

Teorema 2.8 *Toda sequência limitada de números reais possui uma subsequência convergente.*

Prova. Veja página 96 de [9]. □

Antes de encerrar a subseção com o Teorema 2.10, precisaremos definir o conceito de função contínua.

Definição 2.9 *Sejam $f : X \subset \mathbb{R} \to \mathbb{R}$ uma função real e $a \in X$. Dizemos que f é contínua no ponto $a \in X$, se para todo $\varepsilon > 0$ dado arbitrariamente, existir $\delta > 0$ tal que:*

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

Para finalizar esta subseção, apresentaremos a seguir um resultado importante sobre Análise real. Este teorema garante que sob certas condições uma função contínua f admite valor máximo e valor mínimo.

Teorema 2.10 *Toda função contínua $f : [a,b] \to \mathbb{R}$ definida em um intervalo fechado é limitada e atinge seus extremos, isto é, existem $x_1, x_2 \in [a,b]$ tais que,*

$$f(x_1) \leq f(x) \leq f(x_2), \qquad \forall x \in [a,b].$$

Prova. Veja página 187 de [9]. □

2.1.2 Resultados e conceitos de Análise em $\mathbb{R}^n$

Nesta subseção abordaremos conceitos básicos de análise no espaço euclidiano $\mathbb{R}^n$ como produto interno e norma, o que nos levará a enunciar a Desigualdade de Cauchy-Schwarz. Além disso, trataremos de resultados clássicos sobre sequências, pois aparecerão com frequência nas demonstrações de teoremas de importância relevante nos últimos capítulos. E por fim discutiremos conceitos e resultados relacionados a funções $f : U \subset \mathbb{R}^n \to \mathbb{R}$, como derivadas direcionais, Teorema do Valor Médio e Diferenciabilidade de f.

O elemento inicial do estudo da topologia do espaço euclidiano $\mathbb{R}^n$ é a noção sobre produto interno, que definiremos a seguir.

Definição 2.11 *Um produto interno é uma função real simétrica, bilinear, positiva definida $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, ou seja, um produto interno é uma relação em que cada par de elementos de $x, y \in \mathbb{R}^n$ corresponde a um elemento real, indicado por $\langle x, y \rangle$, de tal forma que para quaisquer $x, y, z \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ tenhamos:*

[P1] $\langle x\,,\, y \rangle = \langle y\,,\, x \rangle$;

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

[P3] $\langle \alpha x\,,\, y \rangle = \alpha \langle x\,,\, y \rangle = \langle x\,,\, \alpha y \rangle$;

[P4] $x \neq 0 \Rightarrow \langle x\,,\, x \rangle > 0$.

A seguir definiremos uma norma no espaço euclidiano n-dimensional, que é o conceito necessário para definição de distância.

Definição 2.12 *Uma norma no espaço euclidiano n-dimensional $\mathbb{R}^n$ é definida como uma função real positiva $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ que para $x, y \in \mathbb{R}^n$ e $\alpha \in \mathbb{R}$ cumpre as seguintes condições:*

[N1] $\|x + y\| \leq \|x\| + \|y\|$;

[N2] $\|\alpha x\| = |\alpha| \|x\|$;

[N3] $x \neq 0 \Rightarrow \|x\| > 0$.

Enunciaremos, em seguida, o teorema que nos dará a expressão de uma desigualdade muito conhecida e importante, conhecida como Desigualdade de Cauchy-Schwarz.

Teorema 2.13 *Sejam os pontos $x, y \in \mathbb{R}^n$ quaisquer, então temos a desigualdade:*

$$|\langle x, y\rangle| \leq \|x\|\|y\|, \qquad \textit{para todo} \quad x, y \in \mathbb{R}^n.$$

E a igualdade em (2.13) ocorre se, e somente se, os vetores x, y são multiplos escalares, ou seja, se existe $\alpha \in \mathbb{R}$ tal que $x = \alpha y$.

Prova. Veja página 4 de [10]. □

Como foi dito anteriormente, usaremos o conceito de norma para definir distâncias entre pontos $x, y \in \mathbb{R}^n$. Definimos a distância de x a y por $d(x,y) = \|x-y\|$. Agora estamos preparados para uma definição de bolas em $\mathbb{R}^n$.

Definição 2.14 *Seja $a \in \mathbb{R}^n$ e seja $\delta > 0$ um número real. A bola aberta de centro a e raio δ, é o conjunto dos pontos $x \in \mathbb{R}^n$ cuja distância ao ponto a é menor do que δ, isto é,*

$$B(a, \delta) = \{x \in D;\ \|x-a\| < \delta\}.$$

Analogamente, $B[a, \delta] = \{x \in D;\ \|x-a\| \leq \delta\}$ é a bola fechada de centro a e raio δ.

Agora que sabemos identificar bola aberta, podemos definir um ponto interior a um conjunto $U \subset \mathbb{R}^n$ e, com isso, definir conjuntos abertos.

Definição 2.15 *Seja $U \subset \mathbb{R}^n$. Um ponto $x \in U$ chama-se ponto interior ao conjunto U, quando é centro de alguma bola aberta inteiramente contida em U. E dizemos que um conjunto $U \subset \mathbb{R}^n$ é um conjunto aberto, quando todos os pontos de U são pontos interiores.*

Com conhecimentos sobre distâncias e bolas podemos falar em sequências. Uma sequência $\{x^k\} \subset \mathbb{R}^n$ é formada pelo conjunto das imagens de uma função $x\colon \mathbb{N} \to \mathbb{R}^n$, que associa a cada número $k \in \mathbb{N}$ um vetor $x^k \in \mathbb{R}^n$, que denotaremos por $x^k = (x_{k_1}, x_{k_2}, \ldots, x_{k_n})$. Uma subsequência de $\{x^k\}$ é uma restrição da sequência a um subconjunto infinito $\mathbb{N}' = \{k_1 < k_2 < \ldots < k_i < \ldots\} \subset \mathbb{N}$. Dizemos que a sequência $\{x^k\}$ é limitada quando o conjunto dos seus termos é limitado em $\mathbb{R}^n$, ou seja, quando existe um número real $c > 0$ tal que $\|x_k\| \leq c$, para todo $k \in \mathbb{N}$.

Definimos $a \in \mathbb{R}^n$ como o limite da sequência de pontos $\{x^k\} \subset \mathbb{R}^n$, quando para todo $\varepsilon > 0$ existe $k_0 \in \mathbb{N}$, tal que, para $k > k_0$ temos $\|x^k - a\| < \varepsilon$, ou seja, para $k > k_0$ todos os elementos da sequência $\{x^k\}$ pertencem a bola $B(a, \varepsilon)$, para todo $\varepsilon > 0$.

A seguir enunciaremos um teorema que pode ser usado para generalização dos teoremas sobre sequências de números reais, que foram tratados anteriormente.

Teorema 2.16 *Uma sequência $\{x^k\} \subset \mathbb{R}^n$ converge para o ponto $a = (a_1, \ldots, a_n)$ se, e somente se, para cada $i = 1, 2, \ldots, n$, tem-se $\lim_{k\to\infty} x_i^k = a_i$, ou seja, cada coordenada de $\{x^k\}$ converge para a coordenada correspondente de a.*

Prova. Veja página 15 de [10]. □

Observação 2.17 *Utilizando o teorema anterior vemos que as conclusões dos Teoremas 2.3, 2.4 e 2.5 são verdadeiras para uma sequência $\{x^k\} \subset \mathbb{R}^n$. Para demonstração deste fato, veja página 14 de [10].*

Para encerrar a discussão sobre sequências $\{x^k\} \subset \mathbb{R}^n$, enunciaremos o resultado que generaliza o Teorema 2.8 da subseção anterior.

Teorema 2.18 *(Teorema de Bolzano Weierstrass) Toda sequência limitada em $\mathbb{R}^n$ possui uma subsequência convergente.*

Prova. Veja página 16 de [10]. □

Agora mudaremos o foco de nossa discussão e discorreremos sobre funções reais definidas em $\mathbb{R}^n$, isto é, trataremos de conceitos e resultados relacionados a funções $f : \mathbb{R}^n \to \mathbb{R}$. Primeiramente, definiremos uma função contínua em um ponto $a \in \mathbb{R}^n$.

Definição 2.19 *Uma função $f : X \to \mathbb{R}$, definida no conjunto $X \subset \mathbb{R}^n$, diz-se contínua no ponto $a \in X$ quando, para todo $\varepsilon > 0$ dado arbitrariamente, pode-se obter $\delta > 0$ de tal modo que $|f(x) - f(a)| < \varepsilon$, sempre que $x \in X$ e $\|x - a\| < \delta$. Diz-se que $f : X \to \mathbb{R}$ é uma função contínua, quando f é contínua em todos os pontos.*

Daremos agora a definição de derivada parcial para posteriormente formalizarmos o conceito de função diferenciável.

Definição 2.20 *Seja $f : U \to \mathbb{R}$ definida no aberto $U \subset \mathbb{R}^n$. Dado um ponto $a \in U$, a i-ésima derivada parcial de f no ponto a, com $1 \leq i \leq n$, é definida pelo limite*

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

Onde $e_i \in \mathbb{R}^n$ é um vetor de $\mathbb{R}^n$ composto pelo número 1 na i-ésima coordenada e por 0 nas demais coordenadas.

Definiremos a seguir o conceito de função diferenciável para funções de n variáveis, que constitui a extensão adequada para a definição de função derivável de uma só variável.

Definição 2.21 *A aplicação $f : U \to \mathbb{R}$, definida no aberto $U \subset \mathbb{R}^n$, diz-se diferenciável no ponto $x \in U$, quando existem as derivadas parciais em x, isto é, o vetor gradiente $\nabla f(x)$ em x, e além disso, para $x + v \in U$ tem-se*

$$f(x+v) - f(x) = \langle \nabla f(x), v \rangle + r(v), \qquad \text{onde} \quad \lim_{|v| \to 0} \frac{r(v)}{|v|} = 0.$$

Considerando $f : U \to \mathbb{R}$ definida no aberto $U \subset \mathbb{R}^n$, seja $a \in U$ e $v \in \mathbb{R}^n$. Definimos derivada direcional de f no ponto a, segundo o vetor v, como:

$$\frac{\partial f}{\partial v}(a) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t},$$

quando tal limite existe. Este conceito de derivada direcional é fundamental para o desenvolvimento do estudo sobre as funções $f : \mathbb{R}^n \to \mathbb{R}$. Um resultado de grande importância em Análise no espaço euclidiano $\mathbb{R}^n$ e que será exigido futuramente neste trabalho, é o Teorema do Valor Médio que enunciamos a seguir.

Teorema 2.22 *(Teorema do Valor Médio) Seja a função diferenciável $f : U \to \mathbb{R}$ definida no aberto $U \subset \mathbb{R}^n$. Suponhamos que o segmento de reta $[a, a+v]$ esteja contido em U. Então para $v \in \mathbb{R}^n$ existe $\theta \in (0,1)$ tal que,*

$$f(a+v) - f(a) = \langle \nabla f(a + \theta v), v \rangle.$$

Prova. Veja página 123 e 138 de [10]. □

No capítulo 4 será exigido o conceito sobre função Lipschitziana, que definiremos agora.

Definição 2.23 *Uma função $f : \mathbb{R}^n \to \mathbb{R}^n$ é dita Lipschitziana se existir uma constante L, chamada constante de Lipschitz, que satisfaça a seguinte condição*

$$\|f(x) - f(y)\| \leq L\|x - y\|, \qquad \forall x, y \in \mathbb{R}^n.$$

2.2 Resultados básicos de Otimização

Seremos breves nesta seção. Por se tratar de um assunto tão extenso, nos deteremos a conceitos e definições que aparecerão ao decorrer deste trabalho. Definiremos alguns conceitos introdutórios de otimização, como minimizadores locais e globais e pontos estacionários e alguns resultados básicos.

Consideremos um conjunto $\Omega \subset \mathbb{R}^n$ e uma função $f : \Omega \to \mathbb{R}$, gostaríamos de encontrar o minimizador de f em Ω. Este problema normalmente é exposto da seguinte maneira:

$$\min f(x), \qquad x \in \Omega. \tag{2-1}$$

O conjunto Ω é chamado conjunto viável do problema (2-1), e a função f é chamada função objetivo do problema (2-1).

Definição 2.24 *Seja a função $f : \Omega \to \mathbb{R}$ com D subconjunto de $\mathbb{R}^n$, dizemos que $\bar{x} \in \Omega$ é minimizador local do problema (2-1) se existir $\varepsilon > 0$, tal que,*

$$f(\bar{x}) \leq f(x), \qquad \forall\, x \in B(x, \varepsilon). \tag{2-2}$$

E dizemos que $\bar{x}$ é minimizador global do problema (2-1), se for possível verificar que,

$$f(\bar{x}) \leq f(x), \qquad \forall\, x \in \Omega. \tag{2-3}$$

Observação 2.25 *Se as desigualdades (2-2) e (2-3) forem estritas, então $\bar{x}$ será chamado, respectivamente, de minimizador estrito local e minimizador estrito global.*

Quando no problema (2-1) consideramos o conjuto Ω como sendo o próprio $\mathbb{R}^n$, então dizemos que temos um problema irrestrito e apresentamos este problema como:

$$\min f(x), \qquad x \in \mathbb{R}^n, \tag{2-4}$$

onde $f : \Omega \to \mathbb{R}$, com $\Omega \subset \mathbb{R}^n$ um conjunto aberto. Quando a função objetivo f é diferenciável podemos definir um ponto estacionário.

Definição 2.26 *Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é um ponto estacionário ou crítico para problema (2-4), quando $\nabla f(\bar{x}) = 0$.*

A definição de ponto estacionário é importante, pois estes pontos são os candidatos naturais a minimizadores do problema (2-4).

Teorema 2.27 *Seja $\Omega \subset \mathbb{R}^n$ um conjunto aberto e suponha que a função $f : D \to \mathbb{R}$ seja diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local do problema (2-4), então:*

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

Prova. Veja página 15 de [5]. □

Observação 2.28 *É importante notar que o resultado anterior também é válido quando tratamos de problemas restritos a um conjunto viável Ω, desde que, $\bar{x}$ esteja contido no interior do conjunto viável Ω.*

Quando no problema (2-1) não consideramos o conjuto Ω como sendo o próprio $\mathbb{R}^n$, então dizemos que temos um problema restrito e apresentamos este problema como em (2-1).

Consideraremos Ω um conjunto convexo e fechado, pois esta será a caracterização de Ω no capítulo 5.

Também definimos ponto estacionário para problemas restritos, mas agora a definição deve se adaptar a tais casos.

Definição 2.29 *Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é um ponto estacionário ou crítico para problema* (2-1)*, quando:*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0, \qquad \forall\, x \in \Omega.$$

O próximo teorema afirma que como no caso dos problemas irrestritos, um minimizador dos problemas restritos deve ser ponto estacionário do problema (2-1).

Teorema 2.30 *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in D$. Se $\bar{x}$ é minimizador local do problema* (2-1)*, então*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0, \qquad \forall\, x \in \Omega.$$

Prova. Veja página 66 de [5]. □

2.3 Resultados básicos de Análise Convexa

Nesta seção exibiremos resultados e definições relacionados à análise convexa. Para isso dividimos esta seção em duas subseções: na primeira trataremos de funções convexas e de conjuntos convexos, na segunda definiremos função pseudo-convexa e função quase-convexa, além de alguns resultados importantes.

2.3.1 Funções Convexas e Conjuntos Convexos

Começaremos esta subseção definindo um conjunto convexo em $\mathbb{R}^n$. Este conceito será exigido neste trabalho, quando estivermos trabalhando com problemas restritos de otimização.

Definição 2.31 *Um conjunto $D \subset \mathbb{R}^n$ é dito convexo, se para todo λ no intervalo $[0,1]$, tivermos:*

$$\lambda x + (1-\lambda) y \in D, \quad \forall\, x, y \in D.$$

Geometricamente, esta definição nos diz que o segmento de reta $[x,y]$ definido por,

$$[x,y] := \{\lambda x + (1-\lambda) y : 0 \leq \lambda \leq 1\},$$

está inteiramente contido em D, sempre que os pontos extremos x e y estão em D.

Exemplo 2.32 *O espaço euclidiano $\mathbb{R}^n$ e a bola aberta de centro $a \in \mathbb{R}^n$ e raio δ, denotada por $B(a, \delta)$, são conjuntos convexos.*

A seguir daremos a definição de função convexa, que é um conceito muito importante em otimização com o qual obtemos resultados fortes.

Definição 2.33 ***(Função Convexa)*** *Uma função real $f : D \to \mathbb{R}$ definida sobre um conjunto convexo $D \subset \mathbb{R}^n$ é dita convexa em D, se para todo $\lambda \in [0, 1]$ tivermos:*

$$(1-\lambda)f(\bar{x}) + \lambda f(x) \geq f\left((1-\lambda)\bar{x} + \lambda x\right), \quad \forall x, \bar{x} \in D.$$

A observação a seguir nos dará uma caracterização geometrica para a definição anterior.

Observação 2.34 *Segue da definição anterior que o gráfico de uma função convexa f, deve estar localizado abaixo do seguimento de reta que liga os pontos $(x, f(x))$ e $(y, f(y))$.*

A seguir definiremos epígrafo de uma função, e em seguida enunciaremos o Teorema 2.36, que nos dará outra caracterização para função convexa definida em um conjunto convexo.

Definição 2.35 *O epígrafo da função $f : D \subset \mathbb{R}^n \to \mathbb{R}$ é definido como o conjunto:*

$$E_f = \{(x, c) \in D \times \mathbb{R} : f(x) \leq c\}.$$

Teorema 2.36 *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \to \mathbb{R}$ é convexa em D se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Prova. Veja página 58 de [11]. □

Definiremos agora o conjunto de nível de uma função, em seguida enunciaremos um resultado, que nos dará uma condição necessária para que uma função real f, seja convexa no conjunto convexo $D \subset \mathbb{R}^n$.

Definição 2.37 *O conjunto de nível de uma função $f : D \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por*

$$L_{f,D}(c) := \{x \in D : f(x) \leq c\}.$$

Teorema 2.38 *Seja f uma função real definida no conjunto convexo $D \subset \mathbb{R}^n$. Uma condição necessária, mas não suficiente, para que f seja convexa em D, é que $L_{f,D}(c)$ seja convexo para todo número real c.*

Prova. Veja página 59 de [11]. □

A seguir daremos um exemplo de função que não é convexa, mas que $L_{f,D}(c)$ é convexo para todo número real c,

Exemplo 2.39 *A função real $f(x) = x^3$ é um exemplo de função em que $L_{f,D}(c)$ seja convexo para todo número real c, mas não é uma função convexa.*

Embora, como foi dito anteriormente, funções convexas sejam muito importantes, principalmente em otimização, não nos aprofundaremos neste assunto pois o objetivo deste trabalho é obter resultados a respeito de uma classe maior de funções: As funções convexas geralizadas, isto é, as funções quase-convexas e pseudo-convexas.

2.3.2 Funções Quase-convexas e Pseudo-convexas

Destinamos esta subseção a um estudo dos conceitos relacionados a funções ditas convexas generalizadas, que como já foi mencionado, são as funções quase-convexas e pseudo-convexas. Estes conceitos e resultados apresentados nesta subseção são fundamentais para o desenvolvimento posterior deste trabalho. Comecemos com a definição de função quase-convexa.

Definição 2.40 ***(Função Quase-convexa)*** *Uma função real f definida sobre um conjunto convexo $D \subset \mathbb{R}^n$ é dita quase-convexa em D, se para todo $\lambda \in [0,1]$ e $f(x) \leq f(\bar{x})$, tivermos:*

$$f[(1-\lambda)\bar{x} + \lambda x] \leq f(\bar{x}).$$

Observamos que dizer que uma função é quase-convexa significa, geometricamente, que para cada $x \in D$ tal que $f(x) \leq f(\bar{x})$, a função f não assume um valor maior que $f(\bar{x})$, sobre cada ponto da intersecção do segmento fechado $[\bar{x}, x]$.

Teorema 2.41 *Seja f uma função real definida no conjunto convexo $D \subset \mathbb{R}^n$. Uma condição necessária e suficiente para que f seja quase-convexa em D, é que $L_{f,D}(c)$ seja convexo para todo número real c.*

Prova. Veja página 133 de [11]. □

Observação 2.42 *Pelo Teorema 2.38 juntamente com o teorema 2.41, vemos que a classe de funções quase-convexas contém a classe de funções convexas.*

Para dar sentido à definição de função quase-convexa, isto é, para mostrar que a definição de função convexa não é equivalente à definição de função convexa, exibiremos um exemplo de uma função que é quase-convexa mas não é convexa.

Exemplo 2.43 *Pelo exemplo 2.39 já sabemos que $f(x) = x^3$ não é uma função convexa. Observando que $L_{f,D}(c)$ é convexo para todo número real c, pelo teorema 2.41, concluímos que f é uma função quase-convexa.*

Pela observação anterior percebemos que métodos de otimização desenvolvidos para trabalhar com funções quase-convexas, tratam de um número maior de problemas do que os métodos desenvolvidos para funções convexas. O resultado seguinte será importante em demostrações dos últimos capítulos.

Lema 2.44 *Sejam $\Omega \subset \mathbb{R}^n$ aberto, $x, y \in \Omega$ e $f : \Omega \to \mathbb{R}$ diferenciável e quase-convexa em x. Então:*

$$f(y) \leq f(x) \Rightarrow \langle \nabla f(x), y - x \rangle \leq 0.$$

Prova. Se $x = y$ a afirmação é direta. Consideremos então $x \neq y$. Como Ω é aberto, existe uma bola aberta $B(x, \delta)$ que está contida em Ω. Então para algum $\tilde{\mu}$ tal que $0 < \tilde{\mu} < 1$ e $\tilde{\mu} < \frac{\delta}{\|y-x\|}$, seja

$$\tilde{x} = x + \tilde{\mu}(y - x) = (1 - \tilde{\mu})x + \tilde{\mu}y. \tag{2-5}$$

Portanto $\tilde{x} \in B(x, \delta) \subset \Omega$. Considerando que $f(y) \leq f(x)$, temos pela quase-convexidade de f que $f(\tilde{x}) \leq f(x)$. E como $\tilde{x}, x \in B(x, \delta) \subset \Omega$, pela convexidade da bola $B(x, \delta)$ segue-se,

$$f[(1-\lambda)x + \lambda\tilde{x}] \leq f(x), \qquad \text{para} \quad 0 < \lambda \leq 1.$$

Pela diferenciabilidade da função f em $x \in \Omega$, obtemos a seguinte desigualdade,

$$\langle \lambda \nabla f(x), \tilde{x} - x \rangle + r(\lambda(\tilde{x} - x))\lambda \|\tilde{x} - x\| \leq 0,$$

onde $\lim_{\lambda \to 0} r(\lambda(\tilde{x} - x)) = 0$. Quando fazemos λ tender a zero na última desigualdade, chegamos a conclusão que $\langle \nabla f(x), \tilde{x} - x \rangle \leq 0$. E por fim substituindo a expressão de $\tilde{x}$, dada por (2-5) e usando $\tilde{\mu}$ temos o que queremos demonstrar. $\square$

Agora definiremos uma função pseudo-convexa. Este conceito será exigido para o estudo do capítulo 5, no qual trabalharemos com funções desta classe.

Definição 2.45 ***(Função Pseudo-Convexa)*** *Seja f uma função real definida sobre algum conjunto aberto do $\mathbb{R}^n$ contendo o conjunto D, f é dita pseudo-convexa em D se é diferenciável e para todo $x, \bar{x} \in D$ temos:*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \Rightarrow f(x) \geq f(\bar{x}).$$

Observação 2.46 *Observemos que se a função real $f : \mathbb{R} \to \mathbb{R}$ é pseudo-convexa então vemos que se $f'(\bar{x}) > 0$ então f é crescente para $x > \bar{x}$, se $f'(\bar{x}) < 0$ então f é decrescente para $x < \bar{x}$ e se $f'(\bar{x}) = 0$ então $f(\bar{x}) < f(x)$ para todo $x \in \mathbb{R}$.*

Usando somente a definição de função pseudo-convexa e a definição 2.24, podemos demonstrar o seguinte teorema.

Teorema 2.47 *Seja $f : D \to \mathbb{R}$ uma função pseudo-convexa em $\bar{x} \in D$, onde $D \subset \mathbb{R}^n$. Então:*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0, \quad \forall\, x \in D \;\Rightarrow\; f(\bar{x}) = \min_{x \in D} f(x).$$

Prova. Usando a definição de função pseudo-convexa temos $f(x) \geq f(\bar{x})$ para todo $x \in D$, o que conclui a demonstração. □

Corolário 2.48 *Seja $f : D \to \mathbb{R}$ uma função pseudo-convexa em $\bar{x} \in D$, onde $D \subset \mathbb{R}^n$. Então:*

$$\nabla f(\bar{x}) = 0 \Rightarrow f(\bar{x}) = \min_{x \in D} f(x).$$

Prova. Segue diretamente do Teorema 2.47. □

Teorema 2.49 *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo, e f uma função real definida em algum conjunto aberto que contém D. Se f é pseudo-convexa em D então f é quase-convexa em D.*

Prova. Veja página 143 de [11]. □

Observação 2.50 *A implicação inversa do teorema anterior não é válida, isto é, a classe de funções quase-convexas não está contida na classe de funções pseudo-convexas. Este fato, será discutido no próximo exemplo.*

Exemplo 2.51 *Considere a função real $f(x) = x^3$, pelo exemplo 2.43 sabemos que f é uma função quase-convexa. Como $x = 0$ é ponto estacionário de f, mas não é ponto de minimo, pelo Corolário 2.48 sabemos que f não pode ser pseudo-convexa.*

Até agora, já sabemos que a classe de funções pseudo-convexas está contida na classe de funções quase-convexas, mas não ao contrário. A seguir veremos que a classe de funções convexas, é um subconjunto do conjunto das funções pseudo-convexas.

Teorema 2.52 *Seja f uma função real definida em um conjunto aberto $D \subset \mathbb{R}^n$. Suponha que $\bar{x} \in D$ e que f seja diferenciável em $\bar{x}$. Se f for convexa em $\bar{x}$, então f é pseudo-convexa em $\bar{x}$. A afirmação inversa não é verdadeira.*

Prova. Veja página 144 de [11]. □

O próximo exemplo mostrará que a afirmação inversa do Teorema 2.52 não é verdadeira, isto é, que a classe de funções pseudo-convexas não está contida na classe de funções convexas.

Exemplo 2.53 *A função real $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = -e^{-x^2}$ é um exemplo de função pseudo-convexa, pois se $f'(\bar{x}) > 0$ então f é crescente para $x > \bar{x}$, se $f'(\bar{x}) < 0$ então f é decrescente para $x < \bar{x}$ e se $f'(\bar{x}) = 0$ então $f(\bar{x}) < f(x)$ para todo $x \in \mathbb{R}$, mas f não é convexa pois dados dois pontos $x, y \in \mathbb{R}$ não garantimos que o seguimento que os liga está abaixo do gráfico de f.*

Assim chegamos ao fim desta subseção, sabendo que a classe de funções quase-convexas amplia a classe de funções pseudo-convexas, que por sua vez, amplia a classe de funções convexas.

2.4 Projeções

Nesta seção trataremos sobre projeções ortogonais, de um ponto $x \in \mathbb{R}^n$ sobre um conjunto $D \subset \mathbb{R}^n$. Os resultados desta seção serão exigidos quando trabalharmos com problemas de otimização restritos no capítulo 5. Algumas demonstrações serão desenvolvidas e outras serão indicadas, por se tratarem de um assunto bastante conhecido na literatura.

Uma projeção (ortogonal) do ponto $x \in \mathbb{R}^n$ sobre um conjunto $D \subset \mathbb{R}^n$ é um ponto pertencente a D que está mais próximo de x, utilizando a distância medida pela norma euclidiana. Assim podemos dizer que a projeção de $x \in \mathbb{R}^n$ sobre D, é uma solução global do problema:

$$\min \|y - x\|, \quad y \in D.$$

Definição 2.54 *Definiremos a projeção de $x \in \mathbb{R}^n$ sobre D, do seguinte modo:*

$$P(x) := \arg\min\{\|y - x\| : y \in D\}. \tag{2-6}$$

Teorema 2.55 ***(Teorema da Projeção)***

(a) *Seja $D \subset \mathbb{R}^n$ um conjunto fechado e não-vazio. Então para todo ponto $x \in \mathbb{R}^n$, existe uma projeção de x sobre D.*

(b) *Se, além de ser fechado, o conjunto D for convexo, então para todo ponto $x \in \mathbb{R}^n$, existe uma única projeção de x sobre D.*

Prova. Veja página 10 e página 94 de [5] □

Consideraremos nos lemas subsequentes o conjunto $D \in \mathbb{R}^n$ como sendo não-vazio, fechado e convexo.

Lema 2.56 *Considerando P a projeção sobre D. Temos as seguintes propriedades:*

(a) *Se* $x \in D$*, então* $\langle P(y) - y\,, x - P(y) \rangle \geq 0$*, para todo* $y \in \mathbb{R}^n$*;*

(b) $\|P(y) - P(x)\|^2 \leq \langle P(y) - P(x)\,, y - x \rangle$*, para todo* $y, x \in \mathbb{R}^n$*;*

(c) $\langle P(y) - P(x)\,, y - x \rangle \geq 0$*, para* $x, y \in \mathbb{R}^n$*. Se* $P(y) \neq P(x)$*, então a desigualdade é estrita;*

(d) $\|P(y) - P(x)\| \leq \|y - x\|$*, para* $x\,, y \in \mathbb{R}^n$*, isto é, o operador projeção é não-expansivo.*

Prova. Consideremos $y(\alpha) = (1 - \alpha)P(y) + \alpha x$. Como D é um conjunto convexo e $P(y) \in D$, conclui-se que $y(\alpha) \in D$, para todo $x \in D$ e para $\alpha \in [0\,, 1]$. Pela definição de projeção (2-6), temos:

$$\|y - P(y)\| \leq \|P(y) - y(\alpha)\|, \qquad \alpha \in [0\,, 1].$$

Desenvolvendo a inequação anterior chegamos à seguinte inequação,

$$2\,\langle y - P(y)\,, x - P(y) \rangle - \alpha^2\,\|x - P(y)\|^2 \leq 0.$$

Agora, basta fazer α tender a zero pela direita e o item (a) está demonstrado. Para demontrar (b), sejam x, $y \in \mathbb{R}^n$, como $P(z) \in D$ temos pelo item (a) a desigualdade:

$$\langle P(y) - y\,, P(x) - P(y) \rangle \geq 0, \qquad \forall\, y \in \mathbb{R}^n.$$

Como $P(y) \in D$, obtemos de forma análoga à desigualdade anterior, a seguinte expressão:

$$\langle P(z) - x\,, P(y) - P(x) \rangle \geq 0, \qquad \forall\, x \in \mathbb{R}^n.$$

Fazendo uma manipulação conjunta das duas desigualdades anteriores, demonstra-se o resultado enunciado no item (b). O item (c) decorre diretamente do item (b). Para demonstração de (d) basta utilizar (a) duas vezes, a Desigualdade de Cauchy-Schwarz, e por manipulação algébrica simples conclui-se (d), para maiores detalhes veja página 98 de [5]. □

Lema 2.57 *Seja* P *a projeção sobre* D*. Considere* $x \in D$*, e defina:*

$$x(t) := P(x - t\,d), \qquad d \in \mathbb{R}^n.$$

Então as duas afirmações seguintes são verdadeiras:

(a) $\langle x(t) - x + td\,, y - x(t) \rangle \geq 0$*, para todo* $y \in D$ *e* $t > 0$*;*

(b) *Para todo* $t > 0$ *temos* $\langle d\,, x - x(t) \rangle \geq \|x(t) - x\|^2/t$*.*

Prova. Basta fazer uma aplicação direta do item (a) e do item (b) do Lema 2.56 para demonstrar, respectivamente, o item (a) e o item (b) deste teorema. □

Do último Lema concluímos os dois primeiros itens do próximo Corolário, que será usado muitas vezes durante o trabalho, principalmente no Capítulo 5.

Corolário 2.58 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável e P a projeção sobre D. Considere $x \in D$, e defina:*

$$x(t) := P(x - t\nabla f(x)).$$

Então as duas afirmações seguintes são verdadeiras:

(a) $\langle x(t) - x + t\nabla f(x), y - x(t)\rangle \geq 0$, *para todo $y \in D$ e $t > 0$;*

(b) *Para todo $t > 0$ temos $\langle \nabla f(x), x - x(t)\rangle \geq \|x(t) - x\|^2/t$.*

(c) *$P(x - t\nabla f(x)) = x$, para algum $t \in \mathbb{R}_+$ se, e somente se, $\langle \nabla f(x), y - x\rangle \geq 0$, para todo $y \in D$.*

Prova. Para os itens (a) e (b) basta fazer $d = \nabla f(x)$ no Lema anterior. Para demonstrar o item (c) consideremos primeiramente que $P(x - t\nabla f(x)) = x$, pelo item (a) do Lema 2.56 conclui-se diretamente que $\langle \nabla f(x), y - x\rangle \geq 0$, para todo $y \in D$. Para mostrar a implicação contrária, observemos que pelo item (b) do Lema 2.56 segue a desigualdade

$$\|x(t) - x\|^2 \leq -t\langle \nabla f(x), x(t) - x\rangle \leq 0$$

Assim $x(t) = x$, ou seja, $P(x - t\nabla f(x)) = x$. Portanto a demonstração está concluída. □

Lema 2.59 *Seja P a projeção sobre D. Dados $x, d \in \mathbb{R}^n$, a função ψ definida por*

$$\psi(\alpha) = \frac{\|P(x + \alpha d) - x\|}{\alpha}, \qquad \alpha > 0.$$

é monótona (não-crescente).

Prova. Sejam $\alpha > \beta > 0$ constantes reais. Se $\|P(x + \alpha d) - x\| = \|P(x + \beta d) - x\|$ o lema segue. Suponhamos então,

$$\|P(x + \alpha d) - x\| \neq \|P(x + \beta d) - x\|. \tag{2-7}$$

Através de um cálculo simples vemos que se $\langle v, u - v\rangle > 0$, então obtemos a desigualdade:

$$\frac{\|u\|}{\|v\|} \leq \frac{\langle u, u - v\rangle}{\langle v, u - v\rangle}. \tag{2-8}$$

Fazendo $u = P(x+\alpha d) - x$ e $v = P(x+\beta d) - x$, temos pelo Lema 2.56 (a) as seguintes desigualdades:

$$\langle u, u - v\rangle \leq \alpha \langle d, P(x+\alpha d) - P(x+\beta d)\rangle \tag{2-9}$$

$$\langle v, u - v\rangle \geq \beta \langle d, P(x+\alpha d) - P(x+\beta d)\rangle \tag{2-10}$$

Pelo Lema 2.56 (c) juntamente com (2-7) temos $\langle d, P(x+\alpha d) - P(x+\beta d)\rangle > 0$. Portanto, basta usar (2-8) juntamente com (2-9) e (2-10) para concluir o teorema. $\square$

CAPÍTULO 3

Método do gradiente com busca linear

Neste capítulo, faremos uma discussão bastante sucinta sobre problemas de otimização restritos e irrestritos. O objetivo é dar uma visão preliminar sobre o método do gradiente para problemas irrestritos, e sobre o método do gradiente projetado para problemas restritos. Assim este capítulo servirá de preparação para a leitura dos próximos capítulos. Primeiramente, discutiremos sobre métodos de descida em problemas irrestritos e buscas lineares para esses casos, pois como veremos o método do gradiente se enquadra nesta classe de métodos. Encerraremos tratando do método do gradiente projetado. Observamos que este capítulo foi baseado nas referências [5] e [6] de Izmailov e Solodov.

3.1 Métodos de descida em problemas irrestritos

Consideremos $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável. Trataremos do seguinte problema irrestrito,

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{3-1}$$

Uma idéia natural para resolver esse problema, seria encontrar um $x^{k+1} \in \mathbb{R}^n$ a partir de $x^k \in \mathbb{R}^n$ de tal modo que:

$$f(x^{k+1}) < f(x^k), \qquad \text{para} \quad k = 0, 1, \ldots.$$

Isto é, uma estratégia natural para resolver (3-1) seria encontrar uma sequência $\{x^k\}$ de pontos em $\mathbb{R}^n$, de tal forma que a função objetivo f seja decrescente. Considerando-se uma direção $d^k \in \mathbb{R}^n$, tal que f é decrescente (pelo menos para passos curtos $t_k > 0$) a partir de x^k nessa direção teremos

$$f(x^k + t_k d^k) < f(x^k).$$

Assim fazemos $x^{k+1} = x^k + t_k d^k$. Em seguida, repetimos o mesmo raciocínio para x^{k+1} e assim por diante. Esses tipos de métodos são chamados Métodos de descida, pois a cada iteração diminuímos o valor da função objetivo.

Observação 3.1 *Temos na verdade uma família de métodos para resolver o problema* (3-1)*, pois para cada forma de calcular a direção* d^k*, e para cada forma de calcular o tamanho do passo* t_k *temos um método específico.*

Definição 3.2 *Dizemos que* $d \in \mathbb{R}^n$ *é uma direção de descida da função* $f : \mathbb{R}^n \to \mathbb{R}$ *no ponto* $x \in \mathbb{R}^n$*, se existe* $\varepsilon > 0$ *tal que* $f(x+td) < f(x)$*, para todo* $t \in (0,\varepsilon]$*.*

Para facilitar a escrita, denotaremos neste capítulo, o conjunto de todas as direções de descida da função f no ponto x por $D_f(x)$.

Observação 3.3 *A definição* 3.2 *nos diz, em outras palavras, que se* $d \in D_f(x)$*, então temos algum decréscimo da função* f*, a partir do ponto* x*, para passos suficientemente curtos na direção* d*. Assim, é evidente que uma direção de descida de uma função* f *no ponto* x *pode não existir.*

No próximo resultado daremos uma caracterização para as direções de descida de uma função f no ponto x.

Lema 3.4 *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável no ponto* $x \in \mathbb{R}^n$*. As seguintes afirmações são verdadeiras:*

(a) Para toda direção de descida $d \in \mathbb{R}^n$ *da função* f *no ponto* x*, tem-se* $\langle \nabla f(x), d\rangle \leq 0$*.*

(b) Se $d \in \mathbb{R}^n$ *satisfaz* $\langle \nabla f(x), d\rangle < 0$*, então* d *é uma direção de descida da função* f *no ponto* x*, isto é,* $d \in D_f(x)$*.*

Prova. Seja $d \in \mathbb{R}^n$ direção de descida da função f no ponto x, ou seja, $d \in D_f(x)$. Então para todo $t > 0$ suficientemente pequeno e pela diferenciabilidade de f em x temos:

$$0 > f(x+td) - f(x) = t\langle \nabla f(x), d\rangle + r(v), \qquad onde \quad \lim_{|v|\to 0} \frac{r(v)}{|v|} = 0.$$

Fazendo $t \to 0+$, obtemos $\langle \nabla f(x), d\rangle \leq 0$ e assim o item (a) está demonstrado. A prova de (b) segue a mesma ideia do item anterior. □

Se a função objetivo f, no problema (3-1), for uma função diferenciável, então o esquema iterativo geral para os métodos de descida para problema irrestritos, pode ser enunciado formalmente pelo algoritmo a seguir.

Algorithm 1 (Algoritmo para métodos de descida para problemas irrestritos)
INICIALIZAÇÃO. *Tome* $x^0 \in \mathbb{R}^n$. *faça* $k = 0$.
CRITÉRIO DE PARADA. *Se* $D_f(x^k) = \emptyset$, *então pare. Caso contrário, execute o passo seguinte.*
PASSO ITERATIVO. *Escolha* $d^k \in D_f(x^k)$ *e tome* $t_k \in \mathbb{R}_+$ *de tal modo que,*

$$f(x^k + t_k d^k) < f(x^k), \tag{3-2}$$

e defina

$$x^{k+1} = x^k + t_k d^k.$$

Faça $k = k+1$ *e volte ao* CRITÉRIO DE PARADA.

Observemos pelo nosso algoritmo que se $D_f(x^k) = \emptyset$, então o algoritmo pára, o que faz sentido pois se $D_f(x^k) = \emptyset$, então o ponto x^k é um minimizador local para nosso problema (3-1), já que não existe direção de descida da função f a partir de x^k. Observemos também que a condição dada por (3-2), é satisfeita pela Definição 3.2.

O Algoritmo 1 nos dá uma caracterização geral para os métodos de descida, pois cada forma de escolher a direção $d^k \in D_f(x^k)$, e cada forma de calcular o comprimento do passo t_k nos dá um método de descida específico. O comprimento do passo t_k é calculado examinando o comportamento da função f, ao longo da semi-reta a partir de x^k na direção d^k. Por isso os procedimentos para o cálculo de t_k, recebem o nome de busca linear.

Apresentaremos em seguida algumas das principais regras de busca linear, supondo que para cada ponto x^k dado, uma direção $d^k \in D_f(x^k)$ já foi escolhida.

3.1.1 Buscas Lineares

Nesta subseção nos deteremos a comentar algumas regras de busca linear, sem nos aprofundarmos no assunto, pois nosso objetivo é chamar a atenção para o fato de que existem vários tipos de busca linear, e que cada uma dessas buscas nos dá um método de descida específico, como foi dito anteriormente. Suporemos, nesta subseção, que para cada ponto x^k dado, uma direção $d^k \in D_f(x^k)$ já foi escolhida.

Regra da minimização uni-dimensional

Esta regra consiste em minimizar a função objetivo diferenciável f sobre a semi-reta

$$0 < t \longmapsto x^k + t d^k. \tag{3-3}$$

Consideremos então a função $\phi_k : \mathbb{R}_+ \to \mathbb{R}$ definida por $\phi_k(t) = f(x^k + td^k)$, então o tamanho do passo t_k é escolhido como solução do problema:

$$\min \phi_k(t), \qquad t \in \mathbb{R}_+. \tag{3-4}$$

Observe que $d^k \in D_f(x^k)$ nos garante que as soluções do problema (3-4) estão contidas no interior do conjunto viável, ou seja, que $t_k > 0$. Assim pelo Teorema 2.27 temos:

$$0 = \phi'(t_k) = \left\langle \nabla f(x^{k+1}), d^k \right\rangle. \tag{3-5}$$

Portanto pela equação anterior (3-5), vemos que se $\nabla f(x^{k+1}) \neq 0$, então x^{k+1} será o ponto de interseção da semi-reta (3-3) com a curva de nível da função f que passa por x^{k+1}, em outras palavras, a direção da busca linear é ortogonal à direção do gradiente no iterando seguinte.

Observação 3.5 *Veja que se esta busca for utilizada para algum método de descida, então deve-se resolver em cada iteração do Algoritmo 3-2 um problema unidimensional, o que não torna esse processo interessante na prática.*

Na prática outras regras de busca linear são mais úteis e usadas, desenvolveremos algumas delas a seguir.

Regra de Armijo

Uma regra mais viável computacionalmente é a Regra de Armijo, que é uma regra bastante conhecida e consiste em garantir apenas um decréscimo suficiente da função f, a partir do ponto x^k na direção d^k. Assim não faremos como na regra unidimensional em que estávamos interessados no passo t_k que desse o valor mínimo da função f a partir do ponto x^k na direção d^k, mas sim em um comprimento de passo t_k que represente um certo decrescimento da função f, em relação ao valor $f(x^k)$ e que seja suficiente para garantir a convergência.

Supondo f diferenciável no ponto x^k, podemos definir formalmente a Regra de Armijo através do próximo algoritmo.

Algorithm 2 (Algoritmo para Regra de Armijo) *Fixe os parâmetros* $\hat{t} > 0$, α, $\theta \in (0,1)$.
INICIALIZAÇÃO. *Tome* $t := \hat{t}$.
CRITÉRIO DE PARADA. *Se*

$$f(x^k + td^k) \leq f(x^k) + \alpha t \left\langle \nabla f(x^k), d^k \right\rangle. \tag{3-6}$$

Então pare. Caso contrário, execute o passo seguinte.

PASSO ITERATIVO. *Tome $t = \theta t$ e Volte ao* CRITÉRIO DE PARADA.

Observação 3.6 *Examinando a Regra de Armijo percebemos que como $d^k \in D_f(x^k)$, então, pelo Lema 3.4, a desigualdade (3-6) significa que o decréscimo real que a função f deve sofrer será uma fração de $t\langle \nabla f(x^k)\,,\,d^k\rangle$, e essa fração é determinada pela constante $\alpha \in (0,1)$. Além disso, vemos que o comprimento de passo t_k, será o maior número entre todos os números da forma $t = \hat{t}\theta^i$, $i = 0,1,\ldots$, que satisfaz a desigualdade (3-6) dada no Algoritmo para Regra de Armijo. Assim, dados os parâmetros $\hat{t} > 0$, $\alpha, \theta \in (0,1)$, uma forma de reescrever a Regra de Armijo seria escolher $t_k = \hat{t}\theta^{i_k}$ onde,*

$$i_k := \min\left\{ i \in \mathbb{N} : \ f(x^k + \hat{t}\theta^i d^k) \leq f(x^k) + \alpha \hat{t}\theta^i \langle \nabla f(x^k)\,,\,d^k\rangle \right\}.$$

Neste caso, verifica-se que a sequência $\{t_k\}$ satisfaz as seguintes condições:

1. *$f(x^k + t_k d^k) \leq f(x^k) + \alpha t_k \langle \nabla f(x^k)\,,\,d^k\rangle$;*
2. *$t_k = \hat{t}$ ou $t_k < \hat{t}$ e $f(x^k + \hat{t}\theta^{i_k-1} d^k) > f(x^k) + \alpha \hat{t}\theta^{i_k-1} \langle \nabla f(x^k)\,,\,d^k\rangle$.*

Veremos a seguir que a Regra de Armijo está bem definida, quando a direção d^k satisfaz a condição suficiente de descida dada pelo Lema 3.4, isto é, quando temos:

$$\langle \nabla f(x^k)\,,\,d^k\rangle < 0. \tag{3-7}$$

Lema 3.7 *(**A Regra de Armijo está bem definida.**) Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $x^k \in \mathbb{R}^n$. Suponha que $d^k \in \mathbb{R}^n$ satisfaz (3-7). Então a desigualdade (3-6) é satisfeita para todo $t > 0$ suficientemente pequeno. Em particular, a Regra de Armijo está bem definida e termina com um $t_k > 0$.*

Prova. Como f é diferenciável e $\alpha \in (0,1)$ temos, utilizando a notação da Definição 2.21, para t suficientemente pequeno a desigualdade:

$$(1-\alpha)\langle \nabla f(x^k)\,,\,d^k\rangle + \frac{r(t)}{t} \leq \frac{1-\alpha}{2}\langle \nabla f(x^k)\,,\,d^k\rangle < 0. \tag{3-8}$$

Por outro lado, para t suficientemente pequeno obtemos:

$$f(x^k + td^k) - f(x^k) = \langle \nabla f(x^k)\,,\,t\,d^k\rangle + r(t). \tag{3-9}$$

E por manipulação algébrica simples em (3-9) temos:

$$f(x^k + td^k) - f(x^k) = t\alpha\langle \nabla f(x^k)\,,\,d^k\rangle + t\left((1-\alpha)\langle \nabla f(x^k)\,,\,d^k\rangle + \frac{r(t)}{t} \right).$$

Portanto, basta utilizar a desigualdade (3-8) para concluirmos que a desigualdade (3-6) é satisfeita para todo $t > 0$ suficientemente pequeno, isto é, existe $\varepsilon > 0$ tal que a desigualdade (3-6) é satisfeita para todo $t \in (0, \varepsilon]$. Logo, a Regra de Armijo está bem definida, pois os números $t = \hat{t}\theta^i$, gerados pela Regra de Armijo, pertencerão ao intervalo $(0, \varepsilon]$ em alguma iteração do Algoritmo 2. □

Assim pelo o que vimos no Lema anterior, supondo a condição (3-7), o Algoritmo 2 produz um valor $t_k > 0$ aceitável, após um número finito de iterações, ou seja, após um número finito de reduções do valor inicial $\hat{t}$.

Regra do comprimento de passo fixo

Esta é a regra mais simples e naturalmente é a menos eficiente. Ela consiste, como o próprio nome diz, em fixar um número $\bar{t} > 0$ que não dependa de k, para o comprimento t_k de todas iterações, ou seja, consiste em fazer $t_k = \bar{t}$, para $k = 0, 1, \ldots$. É claro que um $\bar{t}$ grande provavelmente não resultaria na convergência do método, por outro lado, a escolha de $\bar{t}$ pequeno resultaria em uma convergência lenta ou convergência antes de chegar a uma solução. Essa regra só é útil em casos que, por algum motivo, temos dificuldades para o cálculo da função f.

No Capítulo 4 veremos outros tipos de busca linear, que serão aplicadas a uma direção de descida específica, a direção do anti-gradiente $-\nabla f(x)$.

3.1.2 Método do gradiente

Pelo Lema 3.4 concluímos que $-\nabla f(x)$ é a direção de máxima descida. Os métodos de descida que se utilizam tal direção, são chamados de método do gradiente, ou método de máxima descida, e suas iterações são dadas por: Dado x^k, defina

$$x^{k+1} = x^k - t_k \nabla f(x^k), \qquad k = 0, 1, \ldots.$$

Em um certo sentido, o método do gradiente é o mais natural, visto que a direção do anti-gradiente $-\nabla f(x)$ é a direção de descida mais evidente, mas a escolha simplista da direção $d^k = -\nabla f(x^k)$ não torna o método eficiente na prática. Em geral, a convergência pode ser bastante lenta, principalmente se as curvas de nível da função objetivo f forem "alongadas", pois nestes casos $-\nabla f(x^k)$ é quase ortogonal à direção $d^k = x^k - x^*$, que leva a solução x^*. Assim a sequência, $\{x^k\}$ gerada pelo método do gradiente, se aproxima a solução seguindo uma trajetória de "zig-zag".

Observamos que o método do gradiente, que é um dos mais antigos e conhecidos métodos para minimizar funções de várias variáveis, é muito importante do ponto de vista teórico, pois serve de motivação para desenvolvimento e embasamento de métodos mais avançados que possibilitam uma convergência mais rápida. Além disso, a escolha da direção $d^k = -\nabla f(x^k)$ pode ser usada como um “último recurso”, quando num dado iterando x^k escolhas mais sofisticadas, por algum motivo, não funcionaram. Para encerrar esta subseção, iremos formalizar a discussão com o algoritmo teórico para o método do gradiente.

Algorithm 3 (Algoritmo teórico do gradiente)
INICIALIZAÇÃO. *Tome* $x^0 \in \mathbb{R}^n$. *Faça* $k = 0$.
CRITÉRIO DE PARADA. *Se* $\nabla f(x^k) = 0$, *então pare. Caso contrário, execute o passo seguinte.*
PASSO ITERATIVO. *Calcule o comprimento do passo* $t_k > 0$ *utilizando alguma busca linear e defina*

$$x^{k+1} := x^k - t_k \nabla f(x^k);$$

Faça $k = k+1$ *e volte ao* CRITÉRIO DE PARADA.

No Capítulo 4 iremos estudar a convergência do método do gradiente, sob condições especiais a serem definidas oportunamente.

3.2 Método do gradiente projetado

Consideremos $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável. Trataremos do seguinte problema restrito:

$$\min f(x), \quad x \in D, \tag{3-10}$$

onde D é um conjunto convexo, fechado e não-vazio. É bastante lógico que não poderíamos simplesmente aplicar o método do gradiente ao problema (3-10), pois o método poderia trabalhar com pontos não viáveis, isto é, pontos que não pertençam ao conjunto D. Uma idéia natural para resolver esse possível problema seria projetar, em cada iteração, o ponto x^k sobre o conjunto viável D; e esta é a filosofia do método do gradiente projetado. Assim as iterações, geradas pelo método do gradiente projetado, são dadas por: Dado x^k, defina

$$x^{k+1} = x_k(t_k) := P\left(x^k - t_k \nabla f(x^k)\right), \qquad k = 0, 1, \ldots.$$

onde $P(\cdot)$ denota a projeção sobre D, e os comprimentos do passo $t_k > 0$ são calculados utilizando extensões de técnicas de busca linear, como as discutidas na seção anterior. Formalmente podemos dar o método do gradiente projetado através do algoritmo teórico a seguir.

Algorithm 4 (Algoritmo teórico do gradiente projetado)

INICIALIZAÇÃO. *Tome* $x^0 \in D$*. Faça* $k = 0$*.*

CRITÉRIO DE PARADA. *Se* $x^{k+1} = x^k$*, então pare. Caso contrário, execute o passo seguinte.*

PASSO ITERATIVO. *Calcule o comprimento do passo* $t_k > 0$ *utilizando alguma busca linear e defina:*

$$x^{k+1} := P\left(x^k - t_k \nabla f(x^k)\right),$$

Faça $k = k+1$ *e volte ao* CRITÉRIO DE PARADA.

Observação 3.8 *O Algoritmo 4 deve parar em um ponto estacionário, pois pelo do Corolário 2.58 (c) temos que* $x^{k+1} = x^k$ *se, e somente se,* x^{k+1} *é ponto estacionário.*

3.2.1 Buscas Lineares

Nesta subseção iremos discorrer sobre a extensão da Regra do passo fixo, sobre a Regra da minimização uni-dimensional e sobre a extensão a regra de Armijo, para o caso dos problemas com restrição.

Regra do comprimento de passo fixo

Como já foi dito anteriormente, esta é a regra mais simples e naturalmente é a menos eficiente. Sua extensão para o caso de problemas restritos consiste em fixar um número $\bar{t} > 0$ que não dependa de k, para o comprimento t_k de todas iterações, ou seja, consiste em fazer $t_k = \bar{t}$, para $k = 0, 1, \ldots$.

Regra da minimização uni-dimensional no caso de problemas restritos

Esta regra consiste em minimizar a função objetivo diferenciável f sobre o arco de projeção

$$x_k(t) := P\left(x^k - t\nabla f(x^k)\right), \qquad t > 0.$$

Consideremos $\phi_k : \mathbb{R}_+ \to \mathbb{R}$ a função definida por $\phi_k(t) = f(x_k(t))$, então o tamanho do passo t_k é escolhido como solução do problema

$$\min \phi_k(t), \qquad t \in \mathbb{R}_+.$$

No entanto, assim como no caso irrestrito, as regras de minimização uni-dimensional não são muito interessantes do ponto de vista prático. Mais atrativas são as extensões da regra de Armijo, uma das quais citaremos neste instante.

Regra de Armijo para problemas restritos

Supondo f diferenciável no ponto x^k podemos definir a Regra de Armijo formalmente através do seguinte algoritmo.

Algorithm 5 (Algoritmo Regra de Armijo para problemas restritos) *Fixe os parâmetros $\hat{t} > 0$, α, $\theta \in (0,1)$.*
INICIALIZAÇÃO. *Tome $t := \hat{t}$.*
CRITÉRIO DE PARADA. *Se*

$$f(x_k(t)) \leq f(x^k) + \alpha \langle \nabla f(x^k), x_k(t) - x^k \rangle. \tag{3-11}$$

Então pare. Caso contrário, execute o passo seguinte.
PASSO ITERATIVO. *Tome $t = \theta t$ e Volte ao* CRITÉRIO DE PARADA.

Observação 3.9 *Um empecilho quando utilizamos a Regra de Armijo para problemas restritos no Algoritmo 4, é que pode ser exigido o cálculo da projeção em D várias vezes em cada iteração.*

Observação 3.10 *Pelo Corolário 2.58 item (b) sabemos que $\langle \nabla f(x^k), x_k(t) - x^k \rangle < 0$, desde que $\nabla f(x^k) \neq 0$. Assim a desigualdade (3-11) significa que na k-ésima iteração do Algoritmo 5, o decréscimo real que a função f deve sofrer será uma fração de $\langle \nabla f(x^k), x_k(t) - x^k \rangle < 0$, essa fração é determinada por $\alpha \in (0,1)$. Além disso, o comprimento de passo t_k, será o maior número entre todos os números da forma $t = \hat{t}\theta^i$, $i = 0, 1, \ldots$, que satisfaz a desigualdade (3-11) dada no Algoritmo 5. Assim, vemos que dados os parâmetros $\hat{t} > 0$, $\alpha, \theta \in (0,1)$, uma forma de reescrever a Regra de Armijo seria escolher $t_k = \hat{t}\theta^{i_k}$ onde,*

$$i_k := \max \left\{ i \in \mathbb{N} : \ f(x_k(\hat{t}\theta^i)) \leq f(x^k) + \alpha \langle \nabla f(x^k), x_k(\hat{t}\theta^i) - x^k \rangle \right\}.$$

Neste caso, verifica-se que a sequência $\{t_k\}$ satisfaz as seguintes condições:

1. $f(x_k(t_k)) \leq f(x^k) + \alpha \langle \nabla f(x^k), x_k(t_k) - x^k \rangle$;
2. $t_k = \hat{t}$ ou $t_k < \hat{t}$ e $f(x_k(\hat{t}\theta^{i_k - 1})) \leq f(x^k) + \alpha \langle \nabla f(x^k), x_k(\hat{t}\theta^{i_k - 1}) - x^k \rangle$.

Provaremos, a seguir, que a busca linear definida pela extensão da regra de Armijo está bem definida.

Lema 3.11 *(**A regra de Armijo para problemas restritos está bem definida.**) Seja uma função $f : \mathbb{R}^n \to \mathbb{R}$ diferenciável no ponto $x^k \in D$. A desigualdade (3-11) é satisfeita para $t > 0$ suficientemente pequeno. Em particular, a Regra de Armijo está bem definida e termina com um $t_k > 0$.*

Prova. Se x^k é ponto estacionário, então segue do item (c) do Corolário 2.58 que $x_k(t) = x^k$ e assim a desigualdade (3-11) é satisfeita para todo $t > 0$ suficientemente pequeno. Suponhamos então que x^k não seja ponto estacionário. Neste caso, usando novamente o item (c) do Corolário 2.58 temos $\|x_k(t) - x^k\|^2 \neq 0$, para todo $t > 0$. Combinando o Corolário 2.58 item (b) com o Lema 2.59 concluímos que

$$\left\langle \nabla f(x^k)\,, x_k(t) - x^k \right\rangle \leq -\|x_k(1) - x^k\| \cdot \|x_k(t) - x^k\|, \tag{3-12}$$

para $t \in (0\,,1]$. Por outro lado, dado $t > 0$, pelo Teorema do valor médio 2.22 existe ξ_t pertencente ao seguimento $[x^k\,, x_k(t)]$ tal que

$$f(x^k) - f(x_k(t)) = \left\langle \nabla f(\xi_t)\,, x^k - x_k(t) \right\rangle, \qquad t > 0.$$

Com uma manipulação algébrica simples na última equação,

$$f(x_k(t)) - f(x^k) = \alpha \left\langle \nabla f(x^k)\,, x_k(t) - x^k \right\rangle + (1-\alpha) \left\langle \nabla f(x^k)\,, x_k(t) - x^k \right\rangle + \left\langle \nabla f(\xi_t) - \nabla f(x^k)\,, x_k(t) - x^k \right\rangle.$$

Para facilitar a notação definiremos agora $b_k(t)$ como a expressão seguinte,

$$b_k(t) := -(1-\alpha)\|x_k(1) - x^k\| + \left\langle \nabla f(\xi_t) - \nabla f(x^k)\,, \frac{x_k(t) - x^k}{\|x_k(t) - x^k\|} \right\rangle.$$

Usando a desigualdade (3-12) juntamente com a igualdade anterior obtemos que

$$f(x_k(t)) - f(x^k) \leq \alpha \left\langle \nabla f(x^k)\,, x_k(t) - x^k \right\rangle + b_k(t)\|x_k(t) - x^k\|.$$

Como a função f é continuamente diferenciável e $\xi_t \in [x^k\,, x_k(t)]$ conclui-se o limite,

$$\lim_{t \to 0} \nabla f(\xi_t) = \nabla f(x^k),$$

Fazendo t tender a zero pela direita, na última desigualdade, temos o seguinte limite

$$\lim_{t \to 0} b_k(t) = \,-\,(1-\alpha)\|x_k(1) - x^k\| < 0.$$

Combinando as duas últimas desigualdades com o Teorema da Permanência do Sinal (Corolário 2.2), concluímos que existe $\varepsilon > 0$, tal que

$$f(x_k(t)) - f(x^k) \leq \alpha \left\langle \nabla f(x^k)\,, x_k(t) - x^k \right\rangle, \qquad t \in (0, \varepsilon].$$

Portanto a extensão da Regra de Armijo está bem definida. □

Assim pelo que vimos no Lema anterior, se f é uma função diferenciável no ponto $x^k \in \mathbb{R}^n$, então o Algoritmo 5 produz um valor $t_k > 0$ aceitável, após um número finito de iterações, ou seja, após um número finito de reduções do valor inicial $\hat{t}$.

CAPÍTULO 4

Método do gradiente para problemas irrestritos

Neste capítulo, estudaremos o método do gradiente para minimizar funções diferenciáveis definidas no espaço Euclideano n-dimensional. Mostraremos que o método gradiente converge sempre que o problema tem solução e a função objetivo for quase-convexa. Este capítulo foi fortemente baseado no artigo [7] devido a Kiwiel e Murty.

4.1 Método do gradiente para função quase-convexa

Nesta seção vamos assumir que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função continuamente diferenciável e quase-convexa. Consideremos o problema de otimização

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{4-1}$$

Seja S^* o conjunto solução do problema acima, o qual naturalmente pode ser vazio. O método do gradiente com busca unidimensional, para resolver o problema irrestrito acima, é definido formalmente da seguinte forma:

$$x^{k+1} = x^k - t_k \nabla f(x^k), \qquad k = 0, 1, \ldots. \tag{4-2}$$

onde, para $k = 0, 1, \ldots$, o passo t_k é escolhido após uma busca unidimensional na direção do vetor $-\nabla f(x^k)$. Para definir esta busca precisamos de uma função auxiliar $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ que satisfaça as seguintes hipóteses:

(H1) Existe $\alpha \in (0, 1)$ e $\tau_\alpha > 0$, tal que:

$$\phi(t) \leq \alpha t, \qquad \forall\, t \in (0, \tau_\alpha];$$

(H2) Existe $\beta > 0$ e $\tau_\beta > 0$, tal que:

$$\phi(t) \geq \beta t^2, \qquad \forall\, t \in (0, \tau_\beta].$$

Observação 4.1 *Tomando* $\beta = \alpha$ *e* $\tau_\beta = \tau_\alpha = 1$ *a função* $\phi(t) = \alpha t$ *satisfaz as hipóteses* **H1** *e* **H2**.

Usando uma função que satisfaz as condições **H1** e **H2**, escolha uma sequência $\{t_k\} \subset \mathbb{R}_+$ satisfazendo as seguintes condições:

(C1) $0 < t_k \leq \tau_\beta$ e

$$f(x^{k+1}) = f(x^k - t_k \nabla f(x^k)) \leq f(x^k) - \phi(t_k)\|\nabla f(x^k)\|^2, \qquad k = 0, 1, \ldots.$$

(C2) Existe $\gamma > 1$ e $\tau_\gamma > 0$ tal que, para $k = 0, 1, \ldots$

$$t_k \geq \tau_\gamma,$$

ou existe $\tilde{t}_k \in [t_k, \gamma t_k]$, tal que:

$$f(x^k - \tilde{t}_k \nabla f(x^k)) \geq f(x^k) - \phi(\tilde{t}_k)\|\nabla f(x^k)\|^2.$$

Observação 4.2 *Gostaríamos de observar que podemos ter* $\tau_\beta = +\infty$ *em* **H2** *e* **C1**.

Observação 4.3 *É sempre possível escolher uma sequência* $\{t_k\}$ *satisfazendo as condições* **C1** *e* **C2**. *De fato, tome* $0 < \hat{t} \leq \tau_\beta$ *e defina a sequência* $\{t_k\}$ *como:*

$$t_k := \max\left\{2^{-i}\hat{t} : i = 0, 1, \ldots,\ f(x^k - 2^{-i}\hat{t}\nabla f(x^k)) \leq f(x^k) - \phi(2^{-i}\hat{t})\|\nabla f(x^k)\|^2\right\}, \quad (4\text{-}3)$$

Para cada k fixo, usando a hipótese **H1** *concluímos o seguinte limite:*

$$\lim_{t \to 0}\left(\frac{f(x^k - t\nabla f(x^k)) - f(x_k)}{t} + \frac{\phi(t)}{t}\|f(x^k)\|^2\right) \leq -\|f(x^k)\|^2 + \alpha\|f(x^k)\|^2.$$

Como $\alpha \in (0, 1)$*, o limite acima é negativo. Assim, pelo Corolário 2.2, segue-se que existe* $\delta_k > 0$*, tal que:*

$$f(x^k - t\nabla f(x^k)) \leq f(x^k) - \phi(t)\|\nabla f(x^k)\|^2, \qquad t \in (0, \delta_k).$$

Usando a desigualdade anterior e o fato que $\lim_{i \to +\infty} 2^{-i}\hat{t} = 0$*, concluímos que existe* i_k*, tal que* $0 < t_k = 2^{-i_k}\hat{t} < \delta_k$*. Logo,* t_k *dado por (4-3) esta bem definido e satisfaz a condição* **C1***, isto é,*

$$f(x^{k+1}) \leq f(x^k) - \phi(t_k)\|\nabla f(x^k)\|^2, \qquad k = 0, 1, \ldots.$$

Agora, tomando $\gamma = 2$ *e* $\tau_\gamma = \hat{t}$ *temos, para* $k = 0, 1, \ldots,$

$$t_k = \hat{t},$$

ou $t_k < \hat{t}$ e, neste caso, tomando $\tilde{t}_k = 2t_k$ temos

$$f(x^k - \tilde{t}_k \nabla f(x^k)) > f(x^k) - \phi(\tilde{t}_k)\|\nabla f(x^k)\|^2,$$

e assim, a condição **C2** *também é satisfeita. Veja também Observação 3.6 e Lema 3.7.*

A Observações 4.1 e 4.3 mostram que o conjunto das sequências $\{t_k\} \subset \mathbb{R}_+$ satisfazendo as condições **C1** e **C2** é diferente do vazio. Assim, dado x^k, para cada escolha de $t_k \in \mathbb{R}_+$ satisfazendo as condições **C1** e **C2** a iteração x^{k+1} está bem definida. Então o algoritmo conceitual do gradiente é descrito formalmente da seguinte forma:

Algorithm 6 (Algoritmo conceitual do gradiente) *Tome as constantes $\gamma > 1$, $\tau_\gamma > 0$, $\alpha \in (0,1)$, $\tau_\alpha > 0$, $\beta > 0$ e $\tau_\beta > 0$ e uma função $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, tal que:*

$$\beta t^2 \leq \phi(t) \leq \alpha t, \qquad 0 < t \leq \min\{\tau_\alpha, \tau_\beta\},$$

INICIALIZAÇÃO. *Tome $x^0 \in \mathbb{R}^n$. Faça $k = 0$.*
CRITÉRIO DE PARADA. *Se $\nabla f(x^k) = 0$, então pare. Caso contrário, execute passo seguinte.*
PASSO ITERATIVO. *Tome $t_k \in (0, \tau_\beta]$ satisfazendo:*

1. $f(x^k - t_k \nabla f(x^k)) \leq f(x^k) - \phi(t_k)\|\nabla f(x^k)\|^2$,
2. $t_k \geq \tau_\gamma$ *ou* $t_k < \tau_\gamma$ *tal que existe* $\tilde{t}_k \in [t_k, \gamma t_k]$ *satisfazendo:*

$$f(x^k - \tilde{t}_k \nabla f(x^k)) \geq f(x^k) - \phi(\tilde{t}_k)\|\nabla f(x^k)\|^2,$$

e defina:

$$x^{k+1} := x^k - t_k \nabla f(x^k);$$

Faça $k = k+1$ e volte ao CRITÉRIO DE PARADA.

A Observação 4.3 mostra que o Algoritmo 6 esta bem definido. Com o objetivo de provar a convergência da sequência $\{x^k\}$, gerada gerada pelo Algoritmo 6, considere o seguinte conjunto:

$$\bar{S} := \{y \in \mathbb{R}^n : f(x^k) \geq f(y), k = 0, 1, \ldots\}.$$

Lema 4.4 *As seguintes afirmações valem:*

1) *Se $S^* \neq \emptyset$, isto é, o conjunto solução de (4-1) é diferente do conjunto vazio, então $\bar{S}$ é diferente do conjunto vazio e, neste caso, $S^* \subset \bar{S} \neq \emptyset$.*

2) *Se a sequência $\{x^k\}$ gerada pelo método do gradiente admite um ponto de acumulação, digamos $\bar{x}$, então $\bar{x} \in \bar{S} \neq \emptyset$.*

Prova. A definição de $\bar{S}$ implica imediatamente que este conjunto contém o conjunto solução. E assim, a primeira afirmação segue-se.

Para provar a segunda afirmação, suponhamos que $\bar{x}$ seja um ponto de acumulação de $\{x^k\}$. Se $\bar{x}$ é um ponto de acumulação de $\{x^k\}$, então existe uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ tal que:

$$\lim_{j\to\infty} x^{k_j} = \bar{x}.$$

Pela continuidade da f temos que $\lim_{j\to\infty} f(x^{k_j}) = f(\bar{x})$. A condição **C1** juntamente com a hipótese **H2** implicam que

$$f(x^{k+1}) \leq f(x^k) - \phi(t_k)\|\nabla f(x^k)\|^2 \leq f(x^k), \qquad k = 0, 1, \ldots.$$

A última desigualdade implica que a sequência $\{f(x^k)\}$ é monótona não crescente. Como $\{f(x^k)\}$ também possui uma subsequência $\{f(x^{k_j})\}$, que converge para $f(\bar{x})$, segue-se do Teorema 2.5 juntamente com o Teorema 2.7 que $\{f(x^k)\}$ também converge para $f(\bar{x})$. Como $\lim_{k\to\infty} f(x^k) = f(\bar{x})$, usando novamente a monotonicidade de $\{f(x^k)\}$ concluímos que:

$$f(\bar{x}) = \inf\{f(x^k), k = 1, 2, ...\},$$

ou equivalentemente, $f(x^k) \geq f(\bar{x})$, para $k = 0, 1, \ldots$, e isto conclui a prova do lema. □

Lema 4.5 *Se $f(x^k) \geq f(\hat{x})$ para algum $\hat{x}$ fixo e $k = 0, 1, \ldots$, então:*

$$\sum_{k=0}^{\infty} t_k^2 \|\nabla f(x^k)\|^2 \leq \frac{f(x^0) - f(\hat{x})}{\beta}.$$

Além disso, a sequência $\{x^k\}$ converge.

Prova. Combinando as hipóteses **H2** e **C1** concluímos que:

$$\beta t_k^2 \|\nabla f(x^k)\|^2 \leq \phi(t_k) \|\nabla f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}), \qquad k = 0, 1, \ldots.$$

A última desigualdade juntamente com a hipótese, $f(x^k) \geq f(\hat{x})$ para $k = 0, 1, \ldots$, implica que

$$\sum_{k=0}^{n} t_k^2 \|\nabla f(x^k)\|^2 \leq \frac{f(x^0) - f(x^{n+1})}{\beta} \leq \frac{f(x^0) - f(\hat{x})}{\beta},$$

o que implica a primeira parte do lema.

Agora vamos mostrar que a sequência $\{x^k\}$ converge. Observe que,

$$\|\hat{x}-x^{k+1}\|^2 = \|(\hat{x}-x^k)+(x^k-x^{k+1})\|^2, \qquad k=0,1,\ldots.$$

Assim, desenvolvendo o lado direito da igualdade anterior e usando (4-2) obtemos:

$$\|\hat{x}-x^{k+1}\|^2 = \|\hat{x}-x^k\|^2 + 2t_k\left\langle \hat{x}-x^k, \nabla f(x^k)\right\rangle + t_k^2\|\nabla f(x^k)\|^2, \qquad k=0,1,\ldots.$$

Novamente usando a hipótese $f(x^k)\geq f(\hat{x})$ para $k=0,1,\ldots$, juntamente com a quase-convexidade de f segue do Lema 2.44 que:

$$\left\langle \hat{x}-x^k, \nabla f(x^k)\right\rangle \leq 0, \qquad k=0,1,\ldots,$$

esta desigualdade combinada com a igualdade anterior implica imediatamente que:

$$\|\hat{x}-x^{k+1}\|^2 \leq \|\hat{x}-x^k\|^2 + t_k^2\|\nabla f(x^k)\|^2, \qquad k=0,1,\ldots.$$

Da desigualdade anterior e da primeira parte do lema concluímos a seguinte desigualdade:

$$\|\hat{x}-x^k\|^2 \leq \|\hat{x}-x^0\|^2 + \sum_{j=0}^{k-1} t_j^2\|\nabla f(x^j)\|^2 \leq \|\hat{x}-x^k\|^2 + \frac{f(x^0)-f(\hat{x})}{\beta}, \qquad k=1,2\ldots.$$

Assim com o resultado anterior implica que $\{x^k\}$ é limitada, e portanto possui ponto de acumulação, digamos $\bar{x}$. Portanto, utilizando o Lema 4.4 temos:

$$f(x^k)\geq f(\bar{x}), \qquad k=0,1,\ldots,$$

e assim $\bar{x}$ satisfaz a hipótese do lema. Logo, de modo análogo ao que fizemos acima, podemos mostrar também que

$$\|\bar{x}-x^{k+1}\|^2 \leq \|\bar{x}-x^k\|^2 + t_k^2\|\nabla f(x^k)\|^2, \qquad k=0,1,\ldots.$$

Para cada k fixado, obtemos da última desigualdade que

$$\|\bar{x}-x^\ell\|^2 \leq \|\bar{x}-x^k\|^2 + \sum_{j=k}^{\ell-1} t_j^2\|\nabla f(x^j)\|^2, \qquad \forall\, \ell > k.$$

Seja $\varepsilon > 0$. Como $\bar{x}$ é ponto de acumulação de $\{x^k\}$ e, pela primeira parte do lema $\sum_{j=k}^{\infty} t_j^2 \|\nabla f(x^j)\|^2$ é convergente, podemos tomar k suficientemente grande de modo que:

$$\|\bar{x} - x^k\| < \varepsilon/2, \qquad \sum_{j=k}^{\infty} t_j^2 \|\nabla f(x^j)\|^2 < \varepsilon/2.$$

Finalmente, combinado as três últimas desigualdades concluímos que $\{x^k\}$ é convergente. $\square$

Lema 4.6 *Se $\bar{x}$ é ponto de acumulação de $\{x^k\}$, então $\nabla f(\bar{x}) = 0$.*
Prova. Suponha por contradição que uma subsequência $\{x^{k_j}\}$ converge a $\bar{x}$, mas tenhamos $\nabla f(\bar{x}) \neq 0$. De **H2** *e* **C1** *obtemos:*

$$0 \leq \beta t_{k_j}^2 \|\nabla f(x^{k_j})\|^2 \leq f(x^{k_j}) - f(x^{k_j+1}).$$

Como, por continuidade, $\{\nabla f(x^{k_j})\}$ e $\{f(x^{k_j})\}$ convergem, respectivamente, a $\nabla f(\bar{x})$ e $f(\bar{x})$, utilizando a relação anterior vemos que $\{t_{k_j}\}$ converge a zero. Assim, existe j_0 tal que para $j \geq j_0$, temos $t_{k_j} < \tau_\gamma$ e assim segue de **C2** *que existe $\tilde{t}_{k_j} \in [t_{k_j}, \gamma t_{k_j}]$ tal que:*

$$f(x^{k_j} - \tilde{t}_{k_j} \nabla f(x^{k_j})) > f(x^{k_j}) - \phi(\tilde{t}_{k_j}) \|\nabla f(x^{k_j})\|^2, \qquad j = j_0, j_0+1, \ldots.$$

Utilizando **H1** *e a desigualdade anterior concluímos, após simples manipulações algébricas, que para $\alpha \in (0,1)$ temos:*

$$f(x^{k_j} - \tilde{t}_{k_j} \nabla f(x^{k_j})) - f(x^{k_j}) > -\alpha \tilde{t}_{k_j} \|\nabla f(x^{k_j})\|^2, \qquad j = j_0, j_0+1, \ldots.$$

Aplicando o teorema do Valor Médio 2.22 no lado esquerdo da desigualdade anterior:

$$-\left\langle \nabla f(x^{k_j} - \hat{t}_{k_j} \nabla f(x^{k_j})), \nabla f(x^{k_j}) \right\rangle > -\alpha \|\nabla f(x^{k_j})\|^2, \quad \hat{t}_{k_j} \in [0, \tilde{t}_{k_j}], \quad j = j_0, j_0+1, \ldots.$$

Tomando limite, com j tendendo a infinito, obtemos da última desigualdade e continuidade de ∇f que

$$-\|\nabla f(\bar{x})\|^2 \geq -\alpha \|\nabla f(\bar{x})\|^2.$$

Portanto, temos $\alpha \geq 1$, o que é uma contradição pois sabemos por **H1** *que $\alpha \in (0,1)$.* $\square$

O teorema a seguir trata da convergência global do Método do Gradiente para a classe de funções quase-convexas.

Teorema 4.7 *A sequência* $\{x^k\}$ *gerada pelo Algoritmo 6 converge para um ponto estacionário de* f*, ou o conjunto solução* S^* *de* (4-1) *é vazio,*

$$\lim_{k\to+\infty}\|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Suporemos por absurdo que o conjunto solução S^* de (4-1) é diferente do conjunto vazio ou $\lim_{k\to+\infty}\|x^k\| \neq +\infty$ ou $\lim_{k\to+\infty} f(x^k) \neq \inf f$.

Se $S^* \neq \emptyset$ então pelo Lema 4.4 item 1, temos que para algum $\hat{x}$ fixo

$$f(x^k) \geq f(\hat{x}), \qquad k = 0, 1, \ldots,$$

e usando o Lema 4.5 concluímos que $\{x^k\}$ converge para algum $\bar{x}$. Finalmente, pelo Lema 4.6, concluímos que $\nabla f(\bar{x}) = 0$. Assim a sequência $\{x^k\}$ converge para um ponto estacionário.

Se $\lim_{k\to+\infty}\|x^k\| \neq +\infty$, então $\{x^k\}$ possui uma subsequência limitada, e como consequência possui um ponto de acumulação, digamos $\bar{x}$. Portanto pelo pelo Lema 4.4 item 2, temos que para algum $\hat{x}$ fixo

$$f(x^k) \geq f(\hat{x}), \qquad k = 0, 1, \ldots,$$

e usando o Lema 4.5 concluímos que $\{x^k\}$ converge para algum $\bar{x}$. Novamente, pelo Lema 4.6 concluímos que $\nabla f(\bar{x}) = 0$. Assim a sequência $\{x^k\}$ converge para um ponto estacionário.

Se $\lim_{k\to+\infty} f(x^k) \neq \inf f$, então $\lim_{k\to\infty} f(x^k) > \inf f$. Logo, pela definição de ínfimo, existe $\hat{x}$ tal que

$$f(x^k) \geq f(\hat{x}), \qquad k = 0, 1, \ldots,$$

e usando o Lema 4.5 concluímos que $\{x^k\}$ converge para algum $\bar{x}$. Mais uma vez, usando o Lema 4.6, concluímos que $\nabla f(\bar{x}) = 0$. Assim a sequência $\{x^k\}$ converge para um ponto estacionário e o teorema esta provado. □

Os próximos corolários são aplicação direta do Teorema 4.7, observando-se que as classes de funções pseudo-convexas e de funções convexas estão contidas na classe de funções quase-convexas.

Corolário 4.8 *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função continuamente diferenciável e pseudo-convexa. A sequência* $\{x^k\}$*, gerada pelo Algoritmo 6, converge para uma solução de* (4-1)*, isto é, para um ponto de* S^* *ou* S^* *é vazio,*

$$\lim_{k\to+\infty}\|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Pelo Teorema 2.49 sabemos que uma função pseudo-convexa também é uma função quase-convexa. E pelo Corolário 2.48 sabemos que se f é pseudo-convexa, então os pontos estacionários são minimizadores. Assim basta aplicar o Teorema 4.7 para concluir a demonstração.

□

Corolário 4.9 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ é uma função continuamente diferenciável e convexa. A sequência $\{x^k\}$, gerada pelo Algoritmo 6, converge para uma solução de (4-1), isto é, para um ponto de S^* ou S^* é vazio,*

$$\lim_{k\to+\infty} \|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Basta utilizar o Teorema 2.47 juntamente com o Corolário anterior. □

Na próxima seção vamos apresentar várias maneiras de encontrar o passo t_k. Note que cada função $\phi : \mathbb{R}_+ \to \mathbb{R}_+$, e cada maneira de encontrar t_k dá origem a um novo método.

Para encerrar esta seção, daremos um exemplo que verificará que a condição de quase-convexidade da função objetivo f, é necessária para garantirmos o Teorema 4.7.

Exemplo 4.10 *Consideremos a função de duas variáveis $f(x,y) = e^x - y^2$. Mostraremos que esta função não é quase-convexa e que a conclusão do Teorema 4.7 não será verificada. Primeiramente observemos que f não é quase-convexa, pois para os pontos $\alpha = (a,b)$ e $\beta = (a,-b)$ temos $f(\alpha) \leq f(\beta)$, mas com cálculo simples vemos que:*

$$f[(1-\lambda)\beta + \lambda\alpha] > f(\beta) \qquad \textit{para} \quad \lambda \in [0,1).$$

Assim, de acordo com a definição 2.40, sabemos que f não é uma função quase-convexa. Verificaremos agora que a conclusão do Teorema 4.7 não será satisfeita, para isso observemos que a função f dada, não possui ponto estacionário, pois

$$\nabla f(x) = (e^x, -2y), \qquad \forall x,y \in \mathbb{R}^2.$$

Assim, se o teorema fosse válido para f teríamos conjunto solução S^* de (4-1) vazio e

$$\lim_{k\to+\infty} \|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Mas como veremos $\lim_{k\to+\infty} f(x^k) = 0$, enquanto $\inf f = -\infty$. De fato, aplicando o *algoritimo 6 geramos a sequência* $x^k = (x_1^k, 0)$*, com* $\lim_{k\to+\infty} x_1^k = -\infty$ *e*

$$\lim_{k\to+\infty} f(x_1^k, 0) = \lim_{k\to+\infty} e^{x_1^k} = 0 \Rightarrow \lim_{k\to+\infty} f(x^k) = 0.$$

Mas observe que $\inf f = -\infty$*. De fato, basta calcularmos* $f(0,k)$ *e fazermos* k *tender a infinito, para provar que* f *é ilimitada inferiormente. Portanto, fica assim verificada a necessidade da função objetivo* f *ser quase-convexa, para garantirmos as conclusões do Teorema 4.7.*

4.2 Casos especiais

Nesta seção vamos apresentar vários casos especiais do algoritmo apresentado na última seção. Vamos assumir que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função continuamente diferenciável. Consideremos o problema de otimização

$$\min f(x), \quad x \in \mathbb{R}^n. \tag{4-4}$$

Seja S^* o conjunto solução do problema (4-4), o qual naturalmente pode ser vazio. Do mesmo modo que na seção anterior o método do gradiente com busca unidimensional, para resolver o problema irrestrito acima, é definido formalmente da seguinte forma:

$$x^{k+1} = x^k - t_k \nabla f(x^k), \qquad k = 0, 1, \ldots.$$

onde, para $k = 0, 1, \ldots$, o passo t_k é escolhida após uma busca unidimensional na direção do vetor $-\nabla f(x^k)$. A seguir vamos apresentar várias maneiras de escolher o passo t_k satisfazendo as condições **C1** e **C2**, e assim obter vários algoritmos como caso particular do Algoritmo 6.

4.2.1 Caso 1: Método do gradiente com busca de Armijo

A seguir vamos apresentar o método do gradiente com busca de Armijo. Veremos que este método é um caso particular do Algoritmo 6.

Algorithm 7 (Algoritmo do gradiente com busca de Armijo) *Fixe parâmetros* $\hat{t} > 0$, $\alpha, \theta \in (0,1)$.

INICIALIZAÇÃO. *Tome* $x^0 \in \mathbb{R}^n$*. Faça* $k = 0$.

CRITÉRIO DE PARADA. *Se* $\nabla f(x^k) = 0$*, então pare. Caso contrário, execute o passo seguinte.*

PASSO ITERATIVO. *Tome* $t_k \in (0, \hat{t}]$ *satisfazendo:*

$$t_k := \max\left\{\hat{t}\theta^i : i = 0, 1, \ldots,\ f(x^k - \hat{t}\theta^i \nabla f(x^k)) \leq f(x^k) - \alpha \hat{t}\theta^i \|\nabla f(x^k)\|^2\right\},$$

e defina:

$$x^{k+1} := x^k - t_k \nabla f(x^k);$$

Faça $k = k+1$ *e volte ao* CRITÉRIO DE PARADA.

Para mostrar que o algoritmo acima é um caso particular do Algoritmo 6, escolha $\beta = \alpha$ e $\tau_\beta = \tau_\alpha = 1$ e a função $\phi(t) = \alpha t$. É imediato verificar que ϕ satisfaz as condições **H1** e **H2**.

Agora vamos mostrar que a sequência $\{t_k\}$, definida no Algoritmo 7, satisfaz as condições **C1** e **C2**. Primeiro note que para cada k fixo, usando a hipótese **H1** temos:

$$\lim_{t \to 0}\left(\frac{f(x^k - t\nabla f(x^k)) - f(x_k)}{t} + \alpha\|\nabla f(x^k)\|^2\right) \leq -\|\nabla f(x^k)\|^2 + \alpha\|\nabla f(x^k)\|^2.$$

Como $\alpha \in (0,1)$, o limite acima é negativo. Assim, pelo Teorema 2.1, segue-se que existe $\delta_k > 0$ tal que:

$$f(x^k - t\nabla f(x^k)) \leq f(x^k) - \alpha t\|\nabla f(x^k)\|^2, \qquad t \in (0, \delta_k).$$

Usando a desigualdade acima e o fato que $\lim_{i \to +\infty} \hat{t}\theta^i = 0$, concluímos que existe i_k tal que, $0 < t_k = \hat{t}\theta^{i_k}$ satisfaz a condição **C1**, isto é,

$$f(x^{k+1}) \leq f(x^k) - \alpha t_k \|\nabla f(x^k)\|^2, \qquad k = 0, 1, \ldots.$$

Agora, tomando $\gamma = \theta^{-1}$ e $\tau_\gamma = \hat{t}$ temos, para $k = 0, 1, \ldots,$

$$t_k = \hat{t},$$

ou $t_k < \hat{t}$ e, neste caso, tomando $\tilde{t}_k = \theta^{-1} t_k$ temos:

$$f(x^k - \tilde{t}_k \nabla f(x^k)) > f(x^k) - \alpha \tilde{t}_k \|\nabla f(x^k)\|^2,$$

e assim, a condição **C2** também é satisfeita.

A discussão acima mostra que Algoritmo 7 é um caso particular do Algoritmo 6. Portanto a sequência $\{x^k\}$, gerada pelo Algoritmo 7, está bem definida e temos os seguintes resultados de convergência.

Teorema 4.11 *A sequência $\{x^k\}$, gerada pelo método do gradiente com busca de Armijo, é convergente para um ponto estacionário, ou o conjunto solução S^* de (4-4) é vazio,*

$$\lim_{k\to+\infty} \|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Visto que o Algoritmo 7 é um caso particular do Algoritmo 6, o resultado segue do Teorema 4.7. $\square$

Corolário 4.12 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável e pseudo-convexa. A sequência $\{x^k\}$, gerada pelo Algoritmo 7, converge para uma solução de (4-1), isto é, para um ponto de S^* ou S^* é vazio,*

$$\lim_{k\to+\infty} \|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Visto que o Algoritmo 7 é um acaso particular do Algoritmo 6, o resultado segue-se do Corolário 4.8. $\square$

Corolário 4.13 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável e convexa. A sequência $\{x^k\}$, gerada pelo Algoritmo 7, converge para uma solução de (4-1), isto é, para um ponto de S^* ou S^* é vazio,*

$$\lim_{k\to+\infty} \|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Visto que o Algoritmo 7 é um acaso particular do Algoritmo 6, o resultado segue-se do Corolário 4.9. $\square$

4.2.2 Caso 2: Algoritmo A de Burachik *at al.* [1]

Nesta subseção discutiremos o Algoritmo A do artigo Burachik *at al.* [1]. Vamos mostrar que ele pode ser visto como um caso particular do Algoritmo 6.

Considere nesta subseção a função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ convexa e continuamente diferenciável, tal que ∇f seja uma função Lipschitziana com constante de Lipschitz L, isto é,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \forall x, y \in \mathbb{R}^n.$$

O Algoritmo A do artigo Burachik *at al.* [1] é definido da seguinte maneira:

Algorithm 8 (Algoritmo A [1]) *Tome* $\delta_1, \delta_2 > 0$, *tais que:*

$$\frac{L}{2}\delta_1 + \delta_2 < 1.$$

INICIALIZAÇÃO. *Tome* $x^0 \in \mathbb{R}^n$. *faça* $k = 0$.
CRITÉRIO DE PARADA. *Se* $\nabla f(x^k) = 0$, *então pare. Caso contrário, execute o passo seguinte.*
PASSO ITERATIVO. *Tome* $t_k \in \mathbb{R}_+$ *tal que,*

$$\delta_1 \leq t_k \leq \frac{2}{L}(1 - \delta_2), \tag{4-5}$$

e defina:

$$x^{k+1} := x^k - t_k \nabla f(x^k);$$

Faça $k = k+1$ *e volte ao* CRITÉRIO DE PARADA.

Note que o Algoritmo 8 está bem definido, isto é, todos os passos são factíveis. Para mostrar que o algoritmo anterior é um caso particular do Algoritmo 6, primeiramente precisaremos demonstrar o seguinte Lema.

Lema 4.14 *A sequência* $\{x^k\}$, *gerada pelo algoritmo 8, satisfaz:*

$$f(x^{k+1}) \leq f(x^k) - \frac{L\gamma_2}{2(1-\gamma_2)} t_k^2 \|\nabla f(x^k)\|^2, \qquad k = 0, 1, \ldots.$$

Prova. Utilizando-se do teorema fundamental do cálculo, a seguinte igualdade é verdadeira:

$$-t_k \int_0^1 \langle \nabla f(x^k - ut_k \nabla f(x^k)) - \nabla f(x^k), \nabla f(x^k)\rangle du = f(x^{k+1}) - f(x^k) + t_k \|f(x^k)\|^2.$$

Usando agora a desigualdade de Cauchy-Schwarz juntamente com o fato de que ∇f é uma função Lipschitziana, com constante de Lipschitz L, obtemos:

$$f(x^{k+1}) \leq f(x^k) - \left(1 - \frac{Lt_k}{2}\right) t_k \|\nabla f(x^k)\|^2.$$

Pela condição (4-5) temos, por simples manipulação algébrica, a desigualdade seguinte:

$$1 - \frac{Lt_k}{2} \geq \frac{L\gamma_2}{2(1-\gamma_2)} t_k.$$

Portanto, combinando as duas últimas desigualdades o resultado desejado segue. □
Agora com o Lema 4.14, somos capazes de mostrar que de fato o Algoritmo 8 é um caso particular do Algoritmo 6. Observando a desigualdade do Lema 4.14 vemos que basta

tomarmos:

$$\phi(t) = \beta t^2, \qquad onde \quad \beta = L\delta_2/2(1-\delta_2).$$

Podemos então escolher $\alpha \in (0,1)$ e tomar $\tau_\alpha = \alpha/\beta$ para satisfazer **H1** e **H2**. Pela condição (4-5) e pelo Lema 4.14 tomando

$$\tau_\beta = \frac{2}{L}(1-\delta_2), \qquad \tau_\gamma = \delta_1,$$

as condições **C1** e **C2** também são satisfeitas. Portanto, realmente o Algoritmo 8 é um caso particular do Algoritmo 6, e por esta razão concluímos os mesmos resultados do Teorema 4.11 e dos Corolários 4.12 e 4.13 da última subseção, para isto basta tomarmos a sequência $\{x^k\}$ gerada pelo Algoritmo 8.

4.2.3 Caso 3: Algoritimo B de Burachik *at al.* [1]

Nesta subseção discutiremos o Algoritmo B do artigo Burachik *at al.* [1]. Vamos mostrar que ele pode ser visto como um caso particular do Algoritmo 6. Para este caso precisaremos de uma função auxiliar $\psi : \mathbb{R}_+ \to \mathbb{R}_+$ convexa e continuamente diferenciável, que satisfaça as seguintes condições:

(B1) $\psi(0) = 0$ e $\psi'(0) < 1$;

(B2) $\lim\limits_{u \to 0+} \dfrac{\psi(u)}{u^2} > 0.$

Observação 4.15 *Durante toda esta subseção ψ será caracterizada como uma função que satisfaz as hipóteses discutidas acima.*

Observação 4.16 *Veja que* **B2** *implica que $\psi'(0) \geq 0$. Assim por* **B1** *temos:*

$$0 \leq \psi'(0) < 1.$$

E como ψ é uma função convexa, podemos concluir pela desigualdade anterior que ψ é uma função monótona não-decrescente.

Agora apresentamos o Algoritmo do gradiente com a busca linear dada pelo Algoritmo B de [1].

Algorithm 9 (Algoritmo do gradiente com busca linear do Algoritmo B de [1])
Tome as constantes reais positivas δ_1, δ_2 e $\psi : \mathbb{R}_+ \to \mathbb{R}$, uma função continuamente diferenciável e convexa tal que:

$$\psi(0) = 0, \quad \psi'(0) < 1, \qquad \lim_{u \to 0+} \frac{\psi(u)}{u^2} > 0.$$

INICIALIZAÇÃO. *Tome* $x^0 \in \mathbb{R}^n$. *faça* $k = 0$.
CRITÉRIO DE PARADA. *Se* $\nabla f(x^k) = 0$, *então pare. Caso contrário, execute o passo seguinte.*
PASSO ITERATIVO. *Calcule o passo* $t_k \in \mathbb{R}_+$ *da seguinte maneira:*

1. *Tome* $\delta_1 < t_0 < \delta_2$ *e faça* $j = 0$;
2. *Se* $f(x^k - t_j \nabla f(x^k)) \leq f(x^k) - \psi(t_j)\|\nabla f(x^k)\|^2$, *então faça* $t_k = t_j$. *Caso contrário, execute próximo passo 3.*
3. $t_{j+1} = t_j/2$ *e volte ao passo 2.*

Defina:

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

Faça $k = k+1$ *e volte ao* CRITÉRIO DE PARADA.

Mostraremos agora que o Algoritmo 9 é um caso particular do Algoritmo 6. Para isto, tomemos $\phi = \psi$. Por **B1** sabemos que existe $\varepsilon_1 > 0$ tal que para $\alpha \in (0,1)$ temos:

$$\frac{\psi(t)}{t} < \alpha < 1 \qquad t \in (0, \varepsilon_1].$$

Assim, basta tomar $\tau_\alpha = \varepsilon_1$ para satisfazermos a condição **H1**. Por **B2** sabemos que existe $\varepsilon_2 > 0$ tal que tal que para $\beta \in (0,1)$ temos:

$$\frac{\psi(t)}{t^2} > \beta > 0, \qquad t \in (0, \varepsilon_2].$$

Assim, basta tomar $\tau_\beta = \varepsilon_2$ para satisfazermos a condição **H2**. Para verificar as condições **C1** e **C2** basta tomar $\tau_\beta = \min\{\varepsilon_2, \delta_2\}$ e $\tau_\gamma = \delta_1$, veja Observação 4.3. Portanto, o Algoritmo 9 é um caso particular do Algoritmo 6, e as conclusões do Teorema 4.11 e dos Corolários 4.12 e 4.13 também são válidas, para a sequência $\{x^k\}$ gerada pelo Algoritmo 9.

Exemplo 4.17 *Um exemplo de uma função* ψ, *que satisfaça as condições dadas anteriormente, poderia ser* $\psi(u) = \alpha u$, *com* $\alpha \in (0,1)$.

4.2.4 Caso 4: Método do gradiente com busca proximal

Nesta seção relembremos, rapidamente, o método do gradiente com a regra de minimização-unidemensional, que consiste em tomar:

$$x^{k+1} = x^k + t_k \nabla f(x^k), \qquad k = 0, 1, \ldots,$$

onde o comprimento do passo t_k é dado como o argumento que minimiza f, sobre a semi-reta que parte do ponto x^k e tem direção $-\nabla f(x^k)$, ou seja,

$$t_k = \arg\min\left\{f(x^k - t\nabla f(x^k)) : t \geq 0\right\}.$$

O método do gradiente com busca proximal, que é definida no artigo [4] de Iusem e Svaiter, ao contrário do método do gradiente com regra de minimização-unidimensional, não consiste em tomar o argumento que minimiza f sobre a semi-reta que parte do ponto x^k e tem direção $-\nabla f(x^k)$, mas sim em tomar o argumento que minimiza a função regularizada

$$\mathbb{R}^n \ni x \longmapsto f(x) + \lambda_k \|x - x^k\|^2,$$

sobre esta mesma semi-reta. Assim, t_k é dado como segue,

$$t_k = \arg\min\left\{f(x^k - t\nabla f(x^k)) + t^2\lambda_k\|\nabla f(x^k)\|^2 : t \geq 0\right\}, \tag{4-6}$$

onde $\{\lambda_k\} \subset \mathbb{R}_+$ e satisfaz, para certas constantes $\hat{\lambda}, \tilde{\lambda} \in \mathbb{R}_+$, a condição seguinte:

$$\hat{\lambda} \leq \lambda_k \leq \tilde{\lambda},$$

e da mesma forma fazemos,

$$x^{k+1} = x^k - t_k\nabla f(x^k).$$

Portanto podemos definir formalmente o método do gradiente com busca proximal pelo seguinte algoritmo:

Algorithm 10 (Algoritmo do gradiente com busca proximal) *Tome as constantes reais positivas $\hat{\lambda}, \tilde{\lambda} \in \mathbb{R}_+$ e uma sequência $\{\lambda_k\}$ tal que:*

$$\hat{\lambda} \leq \lambda_k \leq \tilde{\lambda}, \qquad k = 0, 1, \ldots .$$

INICIALIZAÇÃO. *Tome $x^0 \in \mathbb{R}^n$. faça $k = 0$.*
CRITÉRIO DE PARADA. *Se $\nabla f(x^k) = 0$, então pare. Caso contrário, execute o passo seguinte.*
PASSO ITERATIVO.*Tome $t_k \in \mathbb{R}_+$ satisfazendo:*

$$t_k = \arg\min\left\{f(x^k - t\nabla f(x^k)) + t^2\lambda_k\|\nabla f(x^k)\|^2 : t \geq 0\right\},$$

e defina:

$$x^{k+1} := x^k - t_k\nabla f(x^k);$$

Faça $k = k + 1$ e volte ao CRITÉRIO DE PARADA.

Assumindo que nosso problema (4-4) tem solução denotaremos por f^* o valor mínimo de f. O Lema a seguir mostrará que a sequência $\{x^k\}$, gerada pelo Algoritmo 10, está bem definida, ou seja, o problema (4-6) tem solução.

Lema 4.18 *A busca proximal está bem definida, ou seja, o problema* (4-6) *tem solução.*

Prova. Para facilitar a discussão usaremos a seguinte notação,

$$\varphi_k(t) = f(x^k - t\nabla f(x^k)) + t^2\lambda_k\|\nabla f(x^k)\|^2, \qquad onde \quad \hat{\lambda} \leq \lambda_k \leq \tilde{\lambda}.$$

Se $\nabla f(x^k) = 0$, então $\varphi_k(t) = f(x^k)$ é constante, e temos $x^{k+1} = x^k$, para qualquer escolha de t_k. Observemos que temos a seguinte desigualdade:

$$\varphi_k(t) \geq f^* + t^2\hat{\lambda}\|\nabla f(x^k)\|^2.$$

E caso $\nabla f(x^k) \neq 0$, temos pela última desigualdade o limite que segue-se:

$$\lim_{t\to\infty} \varphi_k(t) = +\infty.$$

Logo, existe $\rho > 0$ tal que $\varphi_k(\rho) < \varphi_k(t)$, para todo $t > \rho$. Portanto a minimização em (4-6) se reduz ao seguinte problema:

$$t_k = \arg\min \left\{ f(x^k - t\nabla f(x^k)) + t^2\lambda\|\nabla f(x^k)\|^2 : t \in [0,\rho] \right\}. \tag{4-7}$$

Como φ_k é contínua então, pela Observação 2.17, o problema (4-7) tem solução. Assim a busca proximal está bem definida e o lema está demonstrado. $\square$

A seguir enunciaremos um lema que nos ajudará a mostrar que o Algoritmo 10 é um caso particular do Algoritmo 6.

Lema 4.19 *A sequência* $\{x^k\}$, *gerada pelo Algoritmo 10, satisfaz a seguinte propriedade:*

$$f(x^{k+1}) \leq f(x^k) - \hat{\lambda}t_k^2|\nabla f(x^k)\|^2, \qquad k = 0, 1, \ldots. \tag{4-8}$$

Prova. Seguindo a notação do último lema observe que $f(x^k) = \varphi_k(0)$. Portanto,

$$f(x^k) \geq f(x^k - t_k\nabla f(x^k)) + t_k^2\,\|\lambda_k\|\nabla f(x^k)\|^2.$$

Como $\hat{\lambda} \leq \lambda_k$ a desigualdade do lema segue-se. $\square$

Agora estamos preparados para mostrar que o Algoritmo 10 é um caso particular do Algoritmo 6. Assim, a desigualdade (4-8) nos sugere tomar $\phi(t) = \hat{\lambda}t^2$. Então podemos escolher algum $\alpha \in (0,1)$ e adotar $\tau_\alpha = \alpha\hat{\lambda}$ para satisfazermos **H1**. E para que **H2** seja satisfeita basta escolhermos $\beta = \hat{\lambda}$.

Segue do Lema 4.19 que a condição **C1** é sempre satisfeita e, neste caso, $\tau_\beta = +\infty$.

Observação 4.20 *Pelo artigo [7] devido a Kiwiel e Murty, a condição* **C2** *também deve ser satisfeita, pois segundo [7] o Algoritmo 10 é um caso particular do Algoritmo 6. Não conseguimos demonstrar que existem constantes adequadas para que a condição* **C2** *fosse satisfeita. Pensamos que a demonstração deste fato é técnica, pois a princípio a sequência de passos t_k pode convergir a zero ou até mesmo divergir.*

Assumamos que a condição **C2** é satisfeita. Pelo que foi discutido concluímos que o Algoritmo 10 é um caso particular do Algoritmo 6. Portanto as conclusões do Teorema 4.11 e dos Corolários 4.12 e 4.13 também são válidas para a sequência $\{x^k\}$ gerada pelo Algoritmo 10.

CAPÍTULO 5

Método do gradiente para problemas restritos

Neste capítulo estudaremos o método do gradiente projetado, para solução de problemas restritos a um conjunto viável $D \subset \mathbb{R}^n$ não-vazio, fechado e convexo, com função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável e convexa-generalizada, isto é, pseudo-convexa ou quase-convexa. Este capítulo foi baseado nos artigos [12] de Wang e Xiu, e no artigo [2] de Calamai e Moré.

5.1 Método do gradiente projetado para funções convexas-generalizadas

O Problema a ser estudado neste capítulo será:

$$\min f(x), \quad x \in D, \tag{5-1}$$

o qual consiste em minimizar a função objetivo f sobre o conjunto $D \subset \mathbb{R}^n$, onde D é um conjunto não-vazio, fechado e convexo e a função f é continuamente diferenciável em D. Assumiremos que f é uma função convexa-generalizada, isto é, pseudo-convexa ou quase-convexa. Um método bastante conhecido para procurar soluções do problema (5-1), é o método do gradiente projetado, que foi originalmete estabelecido por Goldstein [3] e por Levitin e Polyak [8], o método consiste em definir uma sequência de pontos $\{x^k\}$ como:

$$x^{k+1} = x_k(t_k) := P(x^k - t_k \nabla f(x^k)), \tag{5-2}$$

onde $P(\cdot)$ denota a projeção sobre D. Assim as iterações para o método do gradiente projetado são uma combinação natural do método do gradiente para problemas irrestritos, com a projeção das iterações obtidas sobre o conjunto viável D em questão. Este método possui algumas vantagens: em primeiro lugar o método é de fácil implementação (veja [12]), especialmente quando é fácil calcular a projeção sobre o conjunto D, e não exige grande capacidade de armazenamento; além disso, é possível acrescentar ou retirar restrições a cada iteração.

O tamanho do passo t_k será escolhido de tal forma que, para constantes reais $\mu_1, \mu_2 \in (0,1)$ e $\gamma_1, \gamma_2 \in \mathbb{R}_+$ tivermos:

$$f(x_k(t_k)) \leq f(x^k) + \mu_1 \langle \nabla f(x^k), x_k(t_k) - x^k \rangle, \tag{5-3}$$

e além disso,

$$t_k \geq \gamma_1 \qquad \text{ou} \qquad t_k \geq \gamma_2 \bar{t}_k, \tag{5-4}$$

onde $\bar{t}_k$ satisfaz

$$f(x_k(\bar{t}_k)) > f(x^k) + \mu_2 \langle \nabla f(x^k), x_k(\bar{t}_k) - x^k \rangle. \tag{5-5}$$

Pelo Corolário 2.58 item (b), a condição dada por (5-3) garante um decréscimo real que a função f deve sofrer. Esse decréscimo será uma fração de $\langle \nabla f(x^k), x_k(t_k) - x^k \rangle$, e essa fração será determinada por $\mu_1 \in (0,1)$. E a condição (5-4) garante que o comprimento do passo t_k não será tão pequeno.

Observação 5.1 *Segundo o artigo [12], de Wang e Xiu, a busca linear definida por* (5-3) - (5-5)*, satisfaz outras buscas lineares as buscas definidas em [3] e [8] e a extensão da Regra de Armijo, que foi dicutida na Seção 3.2.1 do Capítulo 3.*

A seguir veremos que a afirmação com respeito da extensão da Regra de Armijo, feita na observação anterior, realmente é verdadeira.

Observação 5.2 *É sempre possível escolher uma sequência $\{t_k\}$ satisfazendo as condições* (5-3) - (5-5)*. De fato, defina a sequência $\{t_k\}$ como $t_k = \hat{t}\theta^{i_k}$, onde*

$$i_k := \min \left\{ i \in \mathbb{N} : \; f(x_k(\hat{t}\theta^i)) \leq f(x^k) + \alpha \langle \nabla f(x^k), x_k(\hat{t}\theta^i) - x^k \rangle \right\}, \tag{5-6}$$

onde $\hat{t} > 0$, α, $\theta \in (0,1)$ são parâmetros fixados. Pelo Lema 3.11 sabemos que a busca dada por (5-6) *está bem definida. Se tomarmos $\mu_1 = \mu_2 = \alpha$ então a condição* (5-3) *será satisfeita para algum $t_k = \hat{t}\theta^{i_k}$ suficientemente pequeno, pois tomando $\gamma_1 = \hat{t}$, $\gamma_2 = \theta^2$ e $\bar{t} = \hat{t}\theta^{i_k - 1}$ as condições 5-4 e 5-5 são satisfeitas. Veja também Observação 3.10.*

O resultado principal a ser discutido sobre o método do gradiente projetado, aplicado ao nosso problema (5-1), é que podemos garantir resultados importantes sobre a convergência da sequência $\{x^k\}$, gerada por (5-2) - (5-5), sob a hipótese que a sequência de passos $\{t_k\}$ é limitada superiormente, isto é, existe $\gamma_3 > 0$ tal que:

$$t_k \leq \gamma_3, \qquad k = 0, 1, \ldots. \tag{5-7}$$

Observação 5.3 *Evidentemente existem muitas regras de busca linear que satisfazem a condição* (5-7)*, por exemplo a Regra de Armijo e a Regra definida por Goldstein em [3].*

A Observação 5.2 mostra que o conjunto das sequências $\{t_k\} \subset \mathbb{R}_+$ satisfazendo as condições (5-3) - (5-5) é diferente do vazio. Assim, dado x^k, para cada escolha de $t_k \in \mathbb{R}_+$ satisfazendo as condições (5-3) - (5-5) a iteração x^{k+1} está bem definida. Então o algoritmo conceitual do gradiente projetado, descrito pelas condições (5-3) - (5-5), pode ser enunciado formalmente da seguinte forma.

Algorithm 11 (Algoritmo do gradiente projetado) *Tome as constantes reais positivas $\mu_1, \mu_2 \in (0,1)$ e $\gamma_1, \gamma_2 \in \mathbb{R}_+$.*
INICIALIZAÇÃO. *Tome $x^0 \in D$. Faça $k = 0$.*
CRITÉRIO DE PARADA. *Se $x^{k+1} = x^k$, então pare. Caso contrário, execute o passo seguinte.*
PASSO ITERATIVO. *Tome $t_k \in \mathbb{R}_+$ satisfazendo:*

1. $f(x_k(t_k)) \leq f(x^k) + \mu_1 \langle \nabla f(x^k), x_k(t_k) - x^k \rangle$;
2. $t_k \geq \gamma_1$ *ou* $t_k \geq \gamma_2 \bar{t}_k$, *tal que* $\bar{t}_k$ *satisfaz*

$$f(x_k(\bar{t}_k)) > f(x^k) + \mu_2 \langle \nabla f(x^k), x_k(\bar{t}_k) - x^k \rangle.$$

Defina

$$x^{k+1} := P\left(x^k - t_k \nabla f(x^k)\right),$$

Faça $k = k+1$ e volte ao CRITÉRIO DE PARADA.

Observação 5.4 *O Algoritmo 4 deve parar em um ponto estacionário, pois pelo do Corolário 2.58 (c) temos que $x^{k+1} = x^k$ se, e somente se, x^{k+1} é ponto estacionário.*

Veremos adiante que assumindo (5-7) e trabalhando com uma função f, pseudo-convexa (respectivamente, quase-convexa) em D, então a sequência $\{x^k\}$, gerada pelo Algoritmo 11, será convergente a um minimizador (respectivamente, a um ponto estacionário) de f , ou $\|x^k\| \to \infty$, o conjunto solução de (5-1) é vazio e

$$\lim_{k \to \infty} f(x^k) = \inf\{f(x) : x \in D\}.$$

Assim, o que veremos neste capítulo será em certo sentido uma extensão do capítulo anterior, ao passo que trabalharemos com problemas restritos para funções quase-convexas. Além disso, trabalharemos com funções pseudo-convexas que, como foi discutido no capítulo 2, é uma classe de funções mais restrita que a classe de funções quase-convexas e por isso conseguiremos resultados mais fortes do que quando trabalhamos com funções quase-convexas.

Antes de exibir os principais resultados precisaremos estudar algumas considerações preliminares que serão exigidas posteriormente.

5.1.1 Resultados Preliminares

O primeiro teorema desta subseção foi baseado no artigo de Paul H. Calamai e Jorge J. Moré [2], e será de grande importância para a demonstração do Teorema 5.8, do Teorema 5.12 e do Teorema 5.16, que são os principais resultados deste capítulo.

Teorema 5.5 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 11. Se $\{x^{k_j}\}$ é alguma subsequência limitada de $\{x^k\}$, então*

$$\lim_{k_j\to\infty} \frac{\|x^{k_j+1}-x^{k_j}\|}{t_{k_j}} = 0,$$

Além disso, qualquer ponto de acumulação da sequência $\{x^k\}$ é um ponto estacionário do problema (5-1).

Prova. Suponha, por contradição, que exista uma subsequência infinita $\{x^{k_\ell}\}$ de $\{x^{k_j}\}$ de tal modo que

$$\frac{\|x^{k_\ell+1}-x^{k_\ell}\|}{t_{k_\ell}} \geq \varepsilon > 0, \qquad \ell = 0,1,\ldots. \tag{5-8}$$

Manipulando algebricamente (5-8), podemos concluir a seguinte desigualdade

$$\frac{\|x^{k_\ell+1}-x^{k_\ell}\|^2}{t_{k_\ell}} \geq \varepsilon \max\{\varepsilon t_{k_\ell}, \|x^{k_\ell+1}-x^{k_\ell}\|\}, \qquad \ell = 0,1,\ldots. \tag{5-9}$$

Pelo Corolário 2.58 (b) e por (5-3) conclui-se que $\{f(x^{k_\ell})\}$ é monótona não crescente e como por hipótese $\{x^{k_\ell}\}$ é limitada, então usando o Teorema 2.6 concluímos que $\{f(x^{k_\ell})\}$ é convergente. Por outro lado, o Corolário 2.58 (b) juntamente com as desigualdades (5-3) e (5-8) implicam que

$$0 \leq \frac{\|x^{k_\ell+1}-x^{k_\ell}\|^2}{t_{k_\ell}} \leq \langle \nabla f(x^{k_\ell}), x^{k_\ell}-x^{k_\ell+1}\rangle \leq \frac{f(x^{k_\ell})-f(x^{k_\ell+1})}{\mu_1}, \qquad \ell = 0,1,\ldots.$$

Como a sequência $\{f(x^{k_\ell})\}$ é convergente, as desigualdades anteriores implicam que

$$\lim_{k_\ell\to\infty} \frac{\|x^{k_\ell+1}-x^{k_\ell}\|^2}{t_{k_\ell}} = 0, \qquad \lim_{k_\ell\to\infty} \langle \nabla f(x^{k_\ell}), x^{k_\ell}-x^{k_\ell+1}\rangle = 0.$$

Combinando o primeiro limite acima e a desigualdade (5-9) concluímos que

$$\lim_{k_\ell\to\infty} t_{k_\ell} = 0, \qquad \lim_{k_\ell\to\infty} \|x^{k_\ell+1}-x^{k_\ell}\| = 0.$$

Assim, por (5-4) vemos que não existe $\gamma_1 > 0$ tal que $t_{k_\ell} \geq \gamma_1$, portanto existe $\gamma_2 > 0$ tal que $t_{k_\ell} \geq \gamma_2 \bar{t}_{k_\ell}$, onde $\bar{t}_{k_\ell}$ satisfaz (5-5). Seja

$$\bar{x}^{k_\ell+1} := x_{k_\ell}(\bar{t}_{k_\ell}), \qquad \beta_{k_\ell} := \min\left\{t_{k_\ell}, \bar{t}_{k_\ell}\right\}.$$

Pela definição acima e o Lema 2.59 temos a seguinte desigualdade

$$\frac{\|x_{k_\ell}(\beta_{k_\ell}) - x^{k_\ell}\|^2}{\beta_{k_\ell}} \geq \beta_{k_\ell} \frac{\|x^{k_\ell+1} - x^{k_\ell}\|}{t_{k_\ell}} \frac{\|\bar{x}^{k_\ell+1} - x^{k_\ell}\|}{\bar{t}_{k_\ell}}.$$

De (5-8) temos $\|x^{k_\ell+1} - x^{k_\ell}\| \geq \varepsilon t_{k_\ell}$. Como $t_{k_\ell} \geq \gamma_2 \bar{t}_{k_\ell}$ a última desigualdade implica que

$$\frac{\|x_{k_\ell}(\beta_{k_\ell}) - x^{k_\ell}\|^2}{\beta_{k_\ell}} \geq \varepsilon \min\{1, \gamma_2\} \|\bar{x}^{k_\ell+1} - x^{k_\ell}\|.$$

Como $\beta_{k_\ell} := \min\left\{t_{k_\ell}, \bar{t}_{k_\ell}\right\}$, a desigualdade anterior junto com o Corolário 2.58 (b) implica

$$\min\{\langle \nabla f(x^{k_\ell}), x^{k_\ell} - x^{k_\ell+1}\rangle, \langle \nabla f(x^{k_\ell}), x^{k_\ell} - \bar{x}^{k_\ell+1}\rangle\} \geq \varepsilon \min\{1, \gamma_2\} \|\bar{x}^{k_\ell+1} - x^{k_\ell}\|. \quad (5\text{-}10)$$

Como $\{\langle \nabla f(x^{k_\ell}), x^{k_\ell} - x^{k_\ell+1}\rangle\}$ converge para zero, então a última desigualdade fornece

$$\lim_{\ell\to\infty} \|\bar{x}^{k_\ell+1} - x^{k_\ell}\| = 0. \quad (5\text{-}11)$$

Para facilitar a notação definiremos agora $p_{k_\ell}(\alpha)$ como a expressão seguinte,

$$p_{k_\ell}(\alpha) := \frac{f(x^{k_\ell}) - f(x_{k_\ell}(\alpha))}{\langle \nabla f(x^{k_\ell}), x_{k_\ell} - x_{k_\ell}(\alpha)\rangle}.$$

Usando o Teorema do Valor Médio 2.22 e alguma manipulações algébricas concluímos que

$$p_{k_\ell}(\bar{t}_{k_\ell}) - 1 = \frac{\langle \nabla f(\xi^{k_\ell}), x^{k_\ell} - \bar{x}^{k_\ell+1}\rangle - \langle \nabla f(x^{k_\ell}), x^{k_\ell} - \bar{x}^{k_\ell+1}\rangle}{\langle \nabla f(x^{k_\ell}), x^{k_\ell} - \bar{x}^{k_\ell+1}\rangle}, \quad (5\text{-}12)$$

onde $\xi^{k_\ell} = x^{k_\ell} + \theta_k(x^{k_\ell+1} - x^{k_\ell}) \in D$ com $0 \leq \theta_{k_\ell} \leq 1$. Utilizando-se de (5-10) e de (5-12), conclui-se

$$|p_{k_\ell}(\bar{t}_{k_\ell}) - 1| \leq \frac{\|\nabla f(\xi^{k_\ell}) - \nabla f(x^{k_\ell})\| \, \|x^{k_\ell} - \bar{x}^{k_\ell+1}\|}{\varepsilon \min\{1, \gamma_2\} \|\bar{x}^{k_\ell+1} - x^{k_\ell}\|} = \frac{\|\nabla f(\xi^{k_\ell}) - \nabla f(x^{k_\ell})\|}{\varepsilon \min\{1, \gamma_2\}}.$$

Como $\xi^{k_\ell} = x^{k_\ell} + \theta_{k_\ell}(x^{k_\ell+1} - x^{k_\ell})$ com $0 \leq \theta_{k_\ell} \leq 1$, usando (5-11) temos

$$\lim_{\ell\to\infty} (\xi^{k_\ell} - x^{k_\ell}) = 0.$$

Portanto, fazendo ℓ tender a infinito na última desigualdade e usando a continuidade de ∇f e a igualdade anterior concluímos que $\lim_{\ell\to\infty}|p_{k_\ell}(\bar{t}_{k_\ell})-1|=0$. Assim, vemos que para ℓ suficientemente grande, temos $p_{k_\ell}(\bar{t}_{k_\ell})>\mu_2$. Mas isto é uma contradição, pois (5-5) garante que $p_k(\bar{t}_k)<\mu_2$ para $k=0,1,\ldots$ e assim o resultado do lema segue-se. □

O próximo lema é simples e não exigirá grande esforço, entretanto será útil em demonstrações de resultados posteriores.

Lema 5.6 *Seja a sequência $\{x^k\}$ gerada pelo Algoritmo 11. Supondo a condição (5-7), isto é, existe $\gamma_3>0$ tal que:*

$$t_k\le\gamma_3,\qquad k=0,1,\ldots.$$

Então para cada iteração do Algoritmo 11, teremos a desigualdade a seguir:

$$\|x^{k+1}-x\|^2\le\|x^k-x\|^2+\frac{2\gamma_3}{\mu_1}\{f(x^k)-f(x^{k+1})\}+2t_k\langle\nabla f(x^k),x-x^k\rangle,\quad x\in D.$$

Prova. Desenvolvendo $\|x^{k+1}-x\|^2$ obtemos a seguinte expressão,

$$\|x^{k+1}-x\|^2=\|x^k-x\|^2+2\langle x^{k+1}-x^k,x^k-x\rangle+\|x^{k+1}-x^k\|^2.$$

Por manipulação algébrica simples observamos que a seguinte desigualdade é verdadeira,

$$2\langle x^{k+1}-x^k,x^k-x\rangle+\|x^{k+1}-x^k\|^2\le 2\langle x^{k+1}-x^k,x^{k+1}-x\rangle.$$

Logo pelas duas últimas desigualgades podemos concluir a próxima desigualdade

$$\|x^{k+1}-x\|^2\le\|x^k-x\|^2+2\langle x^{k+1}-x^k,x^{k+1}-x\rangle.$$

Utilizando o item (a) do Corolário 2.58 e a desigualdade anterior obtemos,

$$\|x^{k+1}-x\|^2\le\|x^k-x\|^2+2t_k\langle\nabla f(x^k),x-x^{k+1}\rangle.$$

Por manipulações algébricas simples no lado direito da desigualdade anterior conclui-se

$$\|x^{k+1}-x\|^2\le\|x^k-x\|^2+2t_k\langle\nabla f(x^k),x^k-x^{k+1}\rangle+2t_k\langle\nabla f(x^k),x-x^k\rangle.$$

Assim combinando (5-3) e (5-7) com a última desigualdade concluímos a demonstração.
□

O Corolário a seguir é uma consequência direta do Lema 5.6, e aparecerá nas demonstrações dos próximos teoremas.

Corolário 5.7 *Considere a sequência $f_k(x)$ definida por*

$$f_k(x) = \|x^k - x\|^2 + \frac{2\gamma_3}{\mu_1} f(x^k).$$

(a) *Se $\langle \nabla f(x^k), x - x^k \rangle < 0$ para todo $k = 0, 1, \ldots$, então $f_k(x)$ é decrescente;*

(b) *Se $\langle \nabla f(x^k), x - x^k \rangle \leq 0$ para todo $k = 0, 1, \ldots$ então $f_k(x)$ é monótona não-crescente.*

Prova. É aplicação direta do Lema 5.6 juntamente com as condições dadas em cada item.
□

5.1.2 Resultados Principais

Os principais resultados sobre a convergência do Algoritmo 11 serão enunciados nesta subseção. Consideremos S^* o conjunto solução do problema (5-1), o qual naturalmente pode ser vazio. Primeiramente estudaremos a convergência do método quando a função objetivo f é pseudo-convexa.

Teorema 5.8 *Suponha f pseudo-convexa em D, e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 11. Supondo a condição (5-7), ou seja, existe $\gamma_3 > 0$ de tal modo que:*

$$t_k \leq \gamma_3, \qquad k = 0, 1 \ldots.$$

Então as seguintes afirmações são verdadeiras.

(a) *$S^* \neq \emptyset$, isto é, conjunto solução de (5-1) é diferente de vazio, se e somente se, a sequência $\{x^k\}$ admite pelo menos um ponto de acumulação. Neste caso, a sequência $\{x^k\}$ converge para solução de (5-1);*

(b) $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in D\}$.

Prova. Demonstraremos primeiramnete o item (a) do teorema. Suponhamos que x^* seja ponto de acumulação da sequência $\{x^k\}$. Pelo Teorema 5.5, x^* é ponto estacionário, ou seja, $\langle \nabla f(x^*), x - x^* \rangle \geq 0$ para todo $x \in D$. Como f é pseudoconvexa temos

$$f(x) \geq f(x^*), \quad x \in D.$$

Portanto x^* pertence ao conjunto solução de (5-1), isto é, $x^* \in S^*$. Logo $S^* \neq \emptyset$. Para demonstrar a implicação contrária, assumamos que o conjunto solução de (5-1) é diferente de vazio. Assim existe $\bar{x}$ pertencente ao conjunto solução de (5-1) tal que:

$$f(\bar{x}) < f(x^k), \qquad k = 0, 1 \ldots. \tag{5-13}$$

Pela pseudoconvexidade da função objetivo f concluímos diretamente o seguinte fato:

$$\langle \nabla f(x^k), \bar{x} - x^k \rangle < 0, \qquad k = 0, 1 \ldots.$$

Consideraremos a sequência $\{f_k(\bar{x})\}$ definida no Corolário 5.7 pela igualdade seguinte

$$f_k(\bar{x}) = \|x^k - \bar{x}\|^2 + m_1 f(x^k),$$

com $m_1 = 2\gamma_3/\mu_1$. Usando o Corolário 5.7 concluímos que $\{f_k(\bar{x})\}$ é uma sequência decrescente. Assim por (5-13) conjuntamente com o fato de $\{f_k(\bar{x})\}$ ser decrescente temos

$$m_1 f(\bar{x}) < \|x^k - \bar{x}\|^2 + m_1 f(\bar{x}) < \|x^0 - \bar{x}\| + m_1 f(x^0)$$

Portanto $\{x^k\}$ é limitada, assim $\{x^k\}$ admite pelo menos um ponto de acumulação x^*. Como foi demonstrado anteriormente x^* pertence ao conjunto solução de (5-1).

Para concluir o item (a) mostraremos agora, que se a sequência $\{x^k\}$, gerada pelo Algoritmo 11, admite um ponto de acumulação x^*, então a sequência $\{x^k\}$ converge para solução de (5-1). Consideremos assim:

$$\lim_{j \to \infty} x^{k_j} = x^*, \qquad j = 0, 1, \ldots.$$

Como $\{f(x^k)\}$ é não-crescente temos $\lim_{k \to \infty} f(x^k) = f(x^*)$. Logo concluímos que

$$f(x^*) < f(x^k), \qquad k = 0, 1, \ldots. \tag{5-14}$$

Novamente, pela pseudoconvexidade da função f segue-se a próxima desigualdade

$$\langle \nabla f(x^k), x^* - x^k \rangle < 0, \qquad k = 0, 1, \ldots.$$

Utilizando a última desigualdade juntamente com o Corolário 5.7 concluímos que $\{f_k(x^*)\}$ é uma sequência decrescente. De forma análoga ao que foi feito anteriormente, por (5-14) juntamente com o Corolário 5.7, conclui-se que $\{f_k(x^*)\}$ é uma sequência limitada. Consequentemente a sequência $\{f_k(x^*)\}$ é convergente. Portanto pelo Teorema 2.5 temos:

$$\lim_{k \to \infty} \{\|x^k - x^*\|^2 + m_1 f(x^k)\} = \lim_{j \to \infty} \{\|x^{k_j} - x^*\|^2 + m_1 f(x^{k_j})\} = m_1 f(x^*).$$

Logo da última equação, verificamos o seguinte limite para sequência $\{x^k\}$,

$$\lim_{k \to \infty} x^k = x^*, \qquad k = 0, 1, \ldots.$$

Portanto está demonstrado o item (a) do teorema. Demonstraremos agora o item (b), consideremos primeiramente $S^* \neq \emptyset$. Pelo item (a), a sequência $\{x^k\}$ converge para solução de (5-1) e portanto:

$$\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in D\}.$$

Neste caso basta analisar o caso em que $S^* = \emptyset$. Como foi demonstrado no item (a), $\{x^k\}$ não admite ponto de acumulação. Logo,

$$\lim_{k\to\infty} \|x^k\| = +\infty. \tag{5-15}$$

Suponha, por contradição, que $\lim_{k\to\infty} f(x^k) > \inf\{f(x) : x \in D\}$. Então existe $\bar{x} \in D$ de tal forma que:

$$f(\bar{x}) < f(x^k), \qquad k = 0, 1 \ldots.$$

Analogamente ao que foi feito no item (a), concluímos que $\{x^k\}$ é limitada. Portanto contraríamos (5-15). Portanto o item (b) fica demonstrado. $\square$

Corolário 5.9 *Suponha f convexa em D, e seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 11. Supondo a condição (5-7), ou seja, existe $\gamma_3 > 0$ de tal modo que:*

$$t_k \leq \gamma_3, \qquad k = 0, 1, \ldots.$$

Então as seguintes afirmações são verdadeiras.

(a) $S^* \neq \emptyset$*, isto é, conjunto solução de (5-1) é diferente de vazio, se e somente se, a sequência $\{x^k\}$ admite pelo menos um ponto de acumulação. E neste caso, a sequência $\{x^k\}$ converge para solução de (5-1);*

(b) $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in D\}$.

Prova. Basta observar que pelo Teorema 2.52 f é pseudo-convexa, portanto podemos aplicar o Teorema 5.8 à sequência $\{x^k\}$. $\square$

Observação 5.10 *A hipótese sobre f ser pseudo-convexa é necessária para garantir o item (b) do Teorema 5.8, como será mostrado no exemplo a seguir.*

Exemplo 5.11 *Consideremos novamente a função de duas variáveis $f(x,y) = e^x - y^2$. Veremos que f é um exemplo de função que não é pseudo-convexa, e que não satisfaz a afirmação (b) do Teorema 5.8 quando consideramos $D = \mathbb{R}^2$. De fato, utilizando a* *notação da Definição 2.45, façamos, por exemplo,* $\bar{x} = (2,-10)$ *e* $x = (0,20)$*, para verificar que* f *não é pseudo-convexa. No Exemplo 4.10 foi demonstrado que*

$$\lim_{k\to+\infty} f(x^k) = 0,$$

mas $\inf f = -\infty$*. Portanto a hipótese sobre* f *ser pseudo-convexa é necessária para garantir o item (b) do Teorema 5.8.*

O próximo teorema é extensão natural do Teorema 4.7, do capítulo anterior. Ele caracteriza a convergência de uma sequência $\{x^k\}$, gerada pelo Algoritmo 11, quando a função objetivo é quase-convexa.

Teorema 5.12 *Seja* f *uma função quase-convexa em* D*, e seja* $\{x^k\}$ *uma sequência gerada pelo Algoritmo 11. Supondo (5-7), ou seja, existe* $\gamma_3 > 0$ *de tal modo que,*

$$t_k \leq \gamma_3, \qquad k = 0,1,\ldots.$$

Então os seguintes itens serão verdadeiros.

(a) *Se existe um ponto de acumulação* x^* *na sequência* $\{x^k\}$*, então* x^* *é um ponto estacionário do problema (5-1), além disso* $\lim_{k\to\infty} x^k = x^*$*;*

(b) *Caso contrário, o conjunto solução* S^* *de (5-1) é vazio,*

$$\lim_{k\to+\infty} \|x^k\| = +\infty, \qquad \lim_{k\to+\infty} f(x^k) = \inf f.$$

Prova. Suponhamos que a sequência $\{x^k\}$, gerada pelo Algoritmo 11, admite ponto de acumulação x^*, ou seja,

$$\lim_{j\to\infty} x^{k_j} = x^*, \qquad j = 0,1,\ldots. \tag{5-16}$$

Como a sequência $\{f(x^k)\}$ é monótona não-crescente temos, por (5-16) e pela continuidade de f, que $\lim_{k\to\infty} f(x^k) = f(x^*)$. Assim,

$$f(x^*) \leq f(x^k), \qquad k = 0,1,\ldots. \tag{5-17}$$

Pelo Lema 2.44 juntamente com a desigualdade anterior (5-17), obtemos diretamente,

$$\langle f(x^k), x^* - x^k \rangle \leq 0. \tag{5-18}$$

Consideraremos a sequência $\{f_k(x^*)\}$, definida no Corolário 5.7 pela igualdade seguinte:

$$f_k(x^*) = \|x^k - x^*\|^2 + m_1 f(x^k),$$

com $m_1 = 2\gamma_3/\mu_1$. Utilizando (5-18) combinada ao Corolário 5.7 conclui-se que a sequência $\{f_k(x^*)\}$ é monotóna não-crescente. Logo pela desigualdae (5-17) temos,

$$m_1 f(x^*) \leq \|x^k - x^*\|^2 + m_1 f(x^k) \leq \|x^0 - x^*\| + m_1 f(x^0)$$

Pela desigualdade anterior conclui-se que $\{f_k(x^*)\}$ é limitada, portanto é convergente. Assim pelo Teorema 2.5, concluímos a seguinte expressão:

$$\lim_{k\to\infty} \{\|x^k - x^*\|^2 + m_1 f(x^k)\} = \lim_{j\to\infty} \{\|x^{k_j} - x^*\|^2 + m_1 f(x^{k_j})\} = m_1 f(x^*).$$

Portanto a sequência, gerada pelo Algoritmo 11, $\{x^k\}$ converge para x^*, ou seja,

$$\lim_{k\to\infty} x^k = x^*, \qquad k = 0, 1, \ldots.$$

Utilizando o Teorema 5.5, concluímos que x^* é ponto estacionário. Assim o item (a) está demonstrado.

Para demonstrar (b) suponhamos primeiramente que $\lim_{k\to+\infty} \|x^k\| \neq +\infty$. Assim $\{x^k\}$ é limitada. Portanto $\{x^k\}$ tem ponto de acumulação, consequentemente ocorre o item (a). Agora suponhamos que $\lim_{k\to\infty} f(x^k) > \inf\{f(x) : x \in D\}$, então existe $\bar{x} \in D$ tal que:

$$f(\bar{x}) < f(x^k), \qquad k = 0, 1, \ldots.$$

Analogamente a demonstração do item (a), podemos concluir que $\{x^k\}$ é limitada. Logo $\{x^k\}$ tem ponto de acumulação, assim ocorre o item (a). Por último suponha que $S^* \neq \emptyset$, isto é, o conjunto solução de (5-1) não é vazio, então existe x^* pertencente ao conjunto solução de (5-1). Daí:

$$f(x^*) \leq f(x^k), \qquad k = 0, 1, \ldots.$$

Analogamente a demonstração do item (a), concluímos que $\{x^k\}$ é limitada. Consequentemente ocorre o item (a). Portanto a demonstração do teorema está concluída. □

Observemos que o Teorema 5.12 não conseguiu afirmações tão fortes quanto as afirmações do Teorema 5.8, mas por outro lado, os resultados serão aplicáveis a uma classe maior de funções, já que trata de funções quase-convexas.

Observação 5.13 *Note que para $D = \mathbb{R}^2$, temos pelo exemplo 4.10, que a condição de quase-convexidade de f é necessária para garantirmos a conclusão do Teorema 5.12.*

A seguir enunciaremos dois lemas, que serão necessários para a demonstração do nosso último teorema.

Lema 5.14 *Para cada iteração do Algoritmo 11, temos a seguinte desigualdade:*

$$\frac{\langle \nabla f(x^k), x^k - x\rangle}{\|x^k - x\|} \leq \frac{\|x^k - x^{k-1}\|}{\alpha_{k-1}} + \|\nabla f(x^k) - \nabla f(x^{k-1})\|, \qquad \forall x \in D. \tag{5-19}$$

Prova. Pelo item (a) do Corolário 2.58 temos a seguinte desigualdade:

$$\langle \nabla f(x^{k-1}), x^k - x\rangle \leq \frac{\langle x^{k-1} - x^k, x^k - x\rangle}{\alpha_{k-1}}, \qquad x \in D.$$

Por simples manipulação algébrica, temos pela última desigualdade a próxima expressão:

$$\frac{\langle \nabla f(x^k), x^k - x\rangle}{\|x^k - x\|} \leq \frac{\langle x^{k-1} - x^k, x^k - x\rangle}{\alpha_{k-1}\|x^k - x\|} + \frac{\langle \nabla f(x^k) - \nabla f(x^{k-1}), x^k - x\rangle}{\|x^k - x\|}.$$

Utilizando-se da desigualdade de Cauchy-Schwarz, dada no Teorema 2.13, nas duas parcelas do lado direito da última inequação concluímos a desigualdade (5-19). Portanto a demonstração do lema está concluída. □

Lema 5.15 *Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo 11. Então a seguinte desigualdade é verdadeira:*

$$\left(\frac{\|x^{k+1} - x^k\|}{t_k}\right)^2 \leq \frac{\langle \nabla f(x^k), x^k - x^{k+1}\rangle}{t_k} \leq \mu^{-1}\frac{f(x^k) - f(x^{k+1})}{t_k}.$$

Prova. Pelo item (b) do Corolário 2.58 obtemos diretamente a primeira desigualdade. A segunda desigualdade é obtida de (5-3) por simples manipulação algébrica. □

O teorema subsequente encerrará este capítulo e mostrará que tanto a sequência $\{x^k\}$, gerada pelas iterações do Algoritmo 11, quanto a sequência $\{f(x^k)\}$ são sequências bem definidas.

Teorema 5.16 *Seja f uma função convexa generalizada em D, isto é, quase-convexa ou pseudoconvexa. Seja ainda $\{x^k\}$ uma sequência gerada pelo Algoritmo 11. Supondo a condição (5-7), ou seja, existe $\gamma_3 > 0$ tal que,*

$$t_k \leq \gamma_3, \qquad k = 0, 1, \ldots.$$

Se $\{x^k\}$ admitir ponto de acumulação, então a sequência $\{x^k\}$ é convergente e valem os seguintes limites equivalentes:

(a) $\displaystyle\lim_{k\to\infty} \frac{\|x^{k+1} - x^k\|}{t_k} = 0$;

(b) $\lim\limits_{k\to\infty} \dfrac{\langle \nabla f(x^k), x^k - x^{k+1}\rangle}{t_k} = 0$;

(c) $\lim\limits_{k\to\infty} \dfrac{f(x^k) - f(x^{k+1})}{t_k} = 0$;

(d) $\lim\limits_{k\to\infty} \dfrac{\langle \nabla f(x^k), x^k - x^{k+1}\rangle}{\|x^{k+1} - x^k\|} = 0.$

Prova. Pelos Teoremas 5.8 e 5.12 temos que $\{x^k\}$ converge para um ponto estacionário do problema (5-1). Assim $\{x^k\}$ é limitada; logo pelo Teorema 5.5 podemos concluir o item (a).

Para demonstrar o item (b) façamos $x = x^{k+1}$ no Lema 5.14 e dividamos a desigualdade (5-19) por t_k, assim juntamente com o item (b) do Corolário 2.58 temos:

$$0 \leq \frac{\langle \nabla f(x^k), x^k - x^{k+1}\rangle}{t_k\,\|x^k - x^{k+1}\|} \leq \frac{\|x^k - x^{k-1}\|}{t_k\,.\,\alpha_{k-1}} + \frac{\|\nabla f(x^k) - \nabla f(x^{k-1})\|}{t_k}. \tag{5-20}$$

Por manipulação algébrica simples em (5-20) obtemos a seguinte relação:

$$\begin{aligned} 0 &\leq \lim_{k\to\infty} \frac{\langle \nabla f(x^k), x^k - x^{k+1}\rangle}{t_k} \\ &\leq \lim_{k\to\infty} \left[\frac{\|x^k - x^{k-1}\|}{t_{k-1}} \frac{\|x^k - x^{k+1}\|}{t_k} + \|\nabla f(x^k) - \nabla f(x^{k-1})\| \frac{\|x^k - x^{k+1}\|}{t_k} \right]. \end{aligned}$$

Pela continuidade de $\nabla f(x)$ em D juntamente com o item (a), conclui-se que o último limite da desigualdade anterior converge para zero. Logo o item (b) está demonstrado.

Para demonstrar o item (c) observemos primeiramente que, por (5-7) juntamente com o item (a) obtemos,

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{5-21}$$

Por outro lado da condição (5-3) juntamente com o item (b) do Corolário 2.58 conclui-se

$$f(x^k) - f(x^{k+1}) \geq \mu_1 \langle \nabla f(x^k), x^k - x^{k+1} \geq \frac{\|x^k - x\|^2}{t_k} \geq 0. \tag{5-22}$$

Pelo Teorema do Valor Médio 2.22 existe $\xi^k = x^k + \theta_k(x^{k+1} - x^k)$, com $0 \leq \theta_k \leq 1$ tal que

$$f(x^k) - f(x^{k+1}) = \langle \nabla f(\xi^k), x^k - x^{k+1}\rangle. \tag{5-23}$$

Combinando as expressões dadas por (5-22) e (5-23) concluímos que,

$$0 \leq \frac{f(x^k) - f(x^{k+1})}{t_k} = \frac{\langle \nabla f(\psi^k), x^k - x^{k+1}\rangle}{t_k}, \tag{5-24}$$

onde $\psi^k = x^k + \theta_k(x^{k+1} - x^k) \in D$ com $0 \leq \theta_k \leq 1$. Por uma manipulação algébrica simples em (5-24) e a aplicação da desigualdade de Cauchy-Schwarz chegamos a seguinte relação,

$$\begin{aligned} 0 &\leq \lim_{k\to\infty} \frac{f(x^k) - f(x^{k+1})}{t_k} \\ &\leq \lim_{k\to\infty} \frac{\langle \nabla f(x^k)\,, x^k - x^{k+1}\rangle}{t_k} + \lim_{k\to\infty} \|\nabla f(\psi^k) - \nabla f(x^k)\| \frac{\|x^k - x^{k+1}\|}{t_k} \end{aligned}$$

Agora basta utilizarmos os itens (a) e (b) juntamente com (5-21) para concluírmos que o último limite convergirá para zero. Consequentemente o item (c) está demonstrado.

Por fim demonstraremos agora o item (d). Pelo item (b) do Corolário 2.58 juntamente com o Lema 5.14 conclui-se a desigualdade,

$$0 \leq \lim_{k\to\infty} \frac{\langle \nabla f(x^{k+1}), x^{k+1} - x^{k+2}\rangle}{\|x^{k+1} - x^{k+2}\|} \leq \lim_{k\to\infty} \frac{\|x^k - x^{k-1}\|}{t_k} + \lim_{k\to\infty} \|\nabla f(x^k) - \nabla f(x^{k+1})\|.$$

Assim pela continuidade de $\nabla f(x)$ e por (5-21), concluímos que o último limite converge para zero. Portanto fica demonstrado o item (d). A equivalência os limites é consequência do Lema 5.15. $\square$

CAPÍTULO 6

Considerações finais

Com este trabalho tentamos descrever bem a filosofia do método do gradiente. No Capítulo 3 tentamos deixar claro, que para problemas irrestritos o método do gradiente é um caso particular de método de descida, quando escolhemos a direção de descida $-\nabla f$. Além disso, no capítulo 3 evidenciamos que existem várias formas de calcular o comprimento de passo t_k. Assim percebe-se que a convergência do método do gradiente está intimamente relacionada com a busca linear escolhida, pois para cada busca linear o método do gradiente gera uma sequência $\{x^k\}$ distinta.

No Capítulo 4 estudamos o artigo [7] devido a Kiwiel e Murty. Trabalhamos com a busca linear definida pelas condições **C1** e **C2**, sob as quais o método do gradiente converge à solução de um problema de minimização irrestrito, caso exista solução e a função objetivo f seja continuamente diferenciável e quase-convexa. É importante notar que o Exemplo 4.10 mostra que a hipótese de quase-convexidade de f é necessária. Vimos na Seção 4.2 que a busca linear definida pelas condições **C1** e **C2**, generaliza outras buscas lineares como a Regra de Armijo e as buscas definidas nas referências [1] e [4]. Portanto pode-se extender os resultados de convergência do Teorema 4.7, às sequências do método do gradiente com buscas de Armijo e buscas definidas em [1] e [4].

Os resultados principais do Capítulo 5 foram basedos no artigo [12] de Wang e Xiu. O Teorema 5.5, que é de grande importância para conclusão dos demais resultados, foi extraido do artigo [2] de Calamai e Moré. Notemos que neste capítulo ampliamos nosso leque de problemas, pois tratamos de problemas restritos a um conjunto D convexo, fechado e não-vazio. Definimos uma busca linear pelas condições (5-3) - (5-5), e supomos que a sequência de passos $\{t_k\}$ fosse limitada superiormente. Assim concluímos resultados fortes, como os Teoremas 5.8 e 5.12, que são extensões do Teorema 4.7.

Um estudo interessante é analisar a condição **C2** para busca proximal. Também seria enriquecedor estudar as referências [3] e [8], pois segundo a Observação 4.20 a busca linear definida por (5-3) - (5-5), satisfaz as buscas definidas em [3] e [8], além de serem artigos clássicos por terem estabelecido o método do gradiente projetado.

Referências Bibliográficas

[1] Burachik, R., Graña Drummond, L.M., Iusem, A.N., and Svaiter, B.F.,*Full convergence of the steepest descent method with inexact line searches*, Optimization, **32**, p. 137-146 (1995)

[2] Calamai, P.H., Moré, J.J., *Projected Gradient Methods for Linearly Constrained Problems*, Math. Programming, **39**, p. 93-116 (1987).

[3] Goldstein, A.A., *Convex programming in Hilbert space*. Bull. Amer. Math. Soc., **70**, p. 709–710 (1964).

[4] Iusem, A.N., Svaiter, B.F. *A proximal regularization of the steepest descent method*, RAIRO Rech. Opér., **29**, no.2, p. 123-130 (1995).

[5] Izmailov,A., Solodov, M., *Otimização-Volume 1-Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade* , IMPA, Rio de Janeiro, 2005.

[6] Izmailov,A., Solodov, M., *Otimização-Volume 2-Métodos computacionais*, IMPA, Rio de Janeiro, 2007.

[7] Kiwiel, K.C., Murty, K., *Convergence of the Steepest Descent Method for Minimizing Quasiconvex Functions*, J. Optim. Theory Appl., **89**, no. 1, p. 221-226 (1996).

[8] Levitin, E.S., Polyak, B.T.,*Constrained minimization prloblems*, USSR. Comput. Math. Math. Phys., **6**, p. 1-50 (1966).

[9] Lima, E.L., *Curso de análise - Volume 1*. IMPA, 12ª edição, 2007.

[10] Lima, E.L., *Curso de análise - Volume 2*. IMPA, 6ª edição, 2006.

[11] Mangasarian, O.L., *Nonlinear Programming*, Classics In Applied Mathematics, SIAM, Philadelphia, 1994.

[12] Wang, C., Xiu, N., *Convergence of the Gradient Projection Method for Generalized Convex Minimization*, Comput. Optim. Appl., **16** no. 2, p. 111-120 (2000).