UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

MILTON GABRIEL GARCIA DOS SANTOS

Problemas de Otimização Quase Convexos: Método do Gradiente para Funções Escalares e Vetoriais

Goiânia
2011

sistema de bibliotecas ufg

UFG

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, por meio 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):	Milton Gabriel Garcia dos Santos
E-mail:	matematica_gabriel@yahoo.com.br
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:	Problemas de Otimização Quase Convexos: Método do Gradiente para Funções Escalares e Vetoriais
Palavras-chave:	Método do Gradiente, Método do Gradiente Projetado, Funções Convexas Generalizadas
Título em outra língua:	Optimization Problems Quasi-Convex: Gradient Method for Vector and Scalar Functions
Palavras-chave em outra língua:	Gradient Method, Projected Gradient Method, Generalized convex functions.
Área de concentração:	Otimização
Data defesa:	(27/10/2011)
Programa de Pós-Graduação:	Mestrado em Matemática
Orientador (a):	Prof. Dr. Orizon Pereira Ferreira
E-mail:	orizon@mat.ufg.br
Co-orientador (a):*	
E-mail:	

*Necessita do CPF quando não constar no SisPG

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: ____________________

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.

Milton Gabriel Garcia dos Santos
Assinatura do (a) autor (a)

Data: 07 / 12 / 2011

[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.

MILTON GABRIEL GARCIA DOS SANTOS

Problemas de Otimização Quase Convexos: Método do Gradiente para Funções Escalares e Vetoriais

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: Otimização.

Orientador: Prof. Dr. Orizon Pereira Ferreira

Goiânia
2011

Dados Internacionais de Catalogação na Publicação na (CIP)
GPT/BC/UFG

S237p
Santos, Milton Gabriel Garcia dos.
Problemas de Otimização Quase Convexos: Método do Gradiente para Funções Escalares e Vetoriais [manuscrito] / Milton Gabriel Garcia dos Santos. -2011.
53 f.

Orientador: Prof. Dr. Orizon Pereira Ferreira.
Dissertação (Mestrado) – Universidade Federal de Goiás, Instituto de Matemática e Estatística, 2011.
Bibliografia.
Inclui lista de figuras, abreviaturas, siglas e tabelas.
Apêndices.

1. Método do Gradiente. 2. Método do Gradiente Projetado. 3. Funções Convexas Generalizadas. I. Título.

CDU:519.853

MILTON GABRIEL GARCIA DOS SANTOS

PROBLEMAS DE OTIMIZAÇÃO QUASE CONVEXOS: MÉTODO DO GRADIENTE PARA FUNÇÕES ESCALARES E VETORIAIS

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 no dia 27 de outubro de 2011, pela Banca Examinadora constituída pelos professores:

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

Prof. Dr. Roberto Andreani
Instituto de Matemática Estatística e Ciência da Computação-UNICAMP

Prof. Dr. Luis Román Lucambio Perez
Instituto de Matemática e Estatística-UFG

Milton Gabriel Garcia dos Santos

Graduou–se em Matemática pela UFMS - Universidade Federal do Mato Grosso do Sul. Durante a graduação foi monitor voluntário de ensino de graduação das disciplinas de Fundamentos de Algebra I e Fundamentos de Geometria e foi voluntário do programa institucional de voluntariado em iniciação científica/CNPq e bolsista do Programa Institucional de Bolsa de Iniciação à Docência - PIBID. Durante o Mestrado foi bolsista da CAPES.

Às pessoas que mais amo:
minha família, amigos
e minha namorada.

Agradecimentos

Primeiramente a Deus, por mais esta conquista e por toda sua força nos momentos difíceis, vindas não somente pelas minhas orações mas também através das pessoas que colocastes em meu caminho.

Sou grato a todos os meus familiares, por terem me apoiado desde quando "me entendo por gente"e por terem acreditado mesmo quando eu já não acreditava. À minha Mãe Maria Lucia Garcia sempre guerreira, que me ensinou que conquistas muitas vezes sãos resultados de lutas e sacrifícios. E a um dos maiores presentes que a vida me deu, à minha irmã Mariana Garcia dos Santos.

À meu Pai Milton Matias dos Santos que com sua simplicidade e poucas palavras me dava muita força e ânimo para seguir lutando. E por isso e por muitas outras histórias deixa enorme e eterna saudade. É como diz a Canção "Só enquanto eu respirar vou me lembrar de você(O Teatro Mágico)".

Agradeço a meus professores da graduação por tudo que me ensinaram. Em especial quero agradecer a Silvia Regina Vieira da Silva, Andreia Cristina Ribeiro e Márcio Ricardo Alves Gouveia pelo apoio, incentivo durante a graduação e também durante o Mestrado.

Agradeço à Kamila Andrade, Brunna e Ubirajara Castro pelo convite e incentivo a cursar o Mestrado aqui em Goiânia na UFG. À Gabriella Barros, Diogo Gonçalves, Lidiane Mayumi e Emerson pelo apoio em um dos momentos mais difíceis que vivi aqui em Goiãnia e por terem me mostrando que não estaria sozinho. Ao meu grande amigo peruano José Carlos Rubianes Silva, que muitas vezes me empurrava e me animava a seguir estudando, momentos estes em que o desanimo e a saudade insistiam em ficar.

À minha amiga Silvana Rodrigues que me alegrava com seu jeito descontraído e extrovertido de ser. À Sergio Reis e Lucimeire Carvalho, Bruno Freitas, Flavio P. Vieira, Fernando Soares, Marcio Lima, Mayer Solorzano, Victor Hugo, Alex Neri, Edwin Lopez, Tharsis Souza e seu amigo Joe, Rosane Gomes, Juliana Canella, Benedito, Agenor, Alfredo.

De maneira especial quero agradecer de coração a todos os amigos que, de alguma forma me ajudaram nessa caminhada e agradeço não somente a ajuda mas o simples e gratificante fato de estarem presentes em meu caminho. Pessoas estas que me ensinaram muitas coisas, conhecimentos estes que vão muito além de Matemática. Pelos momentos alegres e pelos momentos de discussões teóricas, acrescentando juntos nossos conhecimentos.

Agradeço a minha namorada Taiza Moura Silva que sempre com seu carinho, paciência e atenção me ajudou nessa caminhada. Grato pela sua compreensão nos momentos de correria. Serei sempre grato a Deus pelo seu caminho ter um dia encontrado o meu.

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. Em especial quero agradecer a Célia Guimarães pelos serviços prestados, pela competência e por todos os serviços prestados durante o curso.

Em especial ao Professor Dr. Orizon Pereira Ferreira pela honra de ser seu orientando, por tudo que apreendi durante as orientações e pela paciência nas correções na fase da escrita deste trabalho.

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

Em especial aos professores Dr. Roberto Andreani(IMECC/UNICAMP-titular), Dr. Luis Román Lucambio Pérez(MAT/UFG-titular), por me darem a honra de suas participações na banca, por todo tempo dedicado a leitura da dissertação e sugestões para sua melhoria. Agradeço também ao Professor José Yunier Bello Cruz(MAT/UFG-titular) pelas sugestões e correções nas apresentações dos seminários.

[1] Este trabalho teve suporte financeiro da CAPES.

"Nossa atitude em relação ao que a vida nos traz é o que nos difere como pessoas. Frente ao inevitável e de tudo o que é imutável, é sempre desafiador reencontrar o equilíbrio para prosseguir".

Fábio de Melo,
Extraído do Livro: Quando o sofrimento bater à sua porta..

Resumo

Santos, Milton Gabriel Garcia dos. **Problemas de Otimização Quase Convexos: Método do Gradiente para Funções Escalares e Vetoriais**. Goiânia, 2011. 53p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Neste trabalho faremos um estudo das propriedades de convergência do Método do Gradiente Projetado e do Método de Descida para otimização Multi-objetivo. No primeiro momento, o nosso problema de otimização será o de minimizar uma função real de n-variáveis, continuamente diferenciável e restrita a um conjunto de estrutura simples e acrescentaremos sobre a função objetivo a hipótese de quase-convexidade ou pseudo-convexidade. Em seguida iremos considerar o problema de otimização Multi-Objetivo irrestrito e adicionar algumas hipóteses sobre a função vetorial, como a convexidade ou quase-convexidade, além de ser continuamente diferenciável. É importante salientar que em ambos os problemas será utilizado a busca inexata de armijo ao longo de direções viáveis.

Palavras–chave

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

Abstract

Santos, Milton Gabriel Garcia dos. **I**. Goiânia, 2011. 53p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

n this work we study the convergence properties of the Gradient Method Designed and Descent Method for Multi-objective optimization. At first, our optimization problem is to minimize a real function of n-variables, continuously differentiable and restricted to a set of simple structure and add on the objective function of the hypothesis of pseudo-convexity or quasi-convexity. Then we consider the problem of unconstrained multi-objective optimization and add some hypotheses about the function vector, such as convexity or quasi-convexity, and is continuously differentiable. It is noteworthy that in both problems will be used to search for inexact Armijo over viable directions.

Keywords

Gradient Method, Gradient Projection Method, Generalized Convex Functions.

Sumário

Introdução

No primeiro momento iremos considerar o seguinte problema de otimização restrito:

$$\min f(x), \qquad \text{sujeito a} \quad x \quad \text{em} \quad C,$$

onde $f:\mathbb{R}^n \to \mathbb{R}$ é continuamente diferenciável, $C \subset \mathbb{R}^n$ é não vazio, fechado e convexo.

O método mais simples e intuitivo para resolver o problema acima é o Método do Gradiente Projetado. Formalmente este método, a partir de um ponto inicial $x^0 \in C$, gera uma sequência $\{x^k\} \subset C$ de modo que a função objetivo nesses pontos, isto é, a sequência dos valores funcionais $\{f(x^k)\}$ decresce. Estudamos este método não pelo seu valor prático. No entanto, é aquele que vem primeiro à mente, quando se inicia o estudo de métodos de otimização, na tentativa de lidar com métodos mais eficientes. Por outro lado, sabemos que é o Método do Gradiente é básico, isto é, ele é o ponto de partida para a construção de muitos outros algoritmos mais sofisticados e eficientes, por exemplo, ele é parcialmente utilizado em algumas modificações "globalmente convergente"do método de Newton para otimização irrestrita. A simples idéia de diminuir o valor da função objetivo também é usado em muitas outras modificações do método de Newton (ver [**?**] para uma exposição classica sobre o tema) .

Cada iteração do método do gradiente projetado consiste basicamente em duas etapas. A partir de um ponto $x^k \in C$ e número real positivo β_k o próximo ponto x^{k+1} é gerado do seguinte modo:

$$z^k = P_C(x^k - \beta_k \nabla f(x^k)), \qquad\qquad x^{k+1} = x^k + \gamma_k(z^k - x^k),$$

onde γ_k é escolhido segundo a regra de Armijo e P_C é a projeção sobre o conjunto viável. Está presente em diversas literaturas e artigos como em [2, **?**, 4, 6, 7], muitos resultados relevantes a cerca da sequência $\{x^k\}$ gerada pelo método do gradiente projetado e seu comportamento. Por exemplo, prova-se que todo ponto de acumulação da sequência, quando existem, são estacionários, considerando apenas a hipótese de existência de solução. Um outro resultado mais importante, é que adicionando a hipótese de convexidade na função objetivo podemos provar que a sequência gerada pelo método do gradiente projetado converge a uma solução $\bar{x}$ do problema. A pergunta que surge é que se podemos conseguir resultados relevantes enfraquecendo as hipóteses sobre função objetivo. A primeira parte da dissertação é feita com base no Artigo [7], onde foi provado que, sob a hipóteses de quase convexidade, a sequência $\{x^k\}$ converge a um ponto estacionário ou o conjunto solução do problema é vazio.

Na segunda parte da dissertação estudaremos um método de descida para resolver o seguinte problema:

$$\min F(x), \qquad \text{sujeito a } \; x \in \mathbb{R}^n, \tag{0-1}$$

onde a função $F : \mathbb{R}^n \to \mathbb{R}^m$ é continuamente diferenciável. Resolver o problema acima significa encontrar um *Pareto Ótimo* ou, respectivamente, um *Pareto Fraco*, isto é, um ponto $\bar{x} \in \mathbb{R}^n$ com a seguinte propriedade, não existe $y \in \mathbb{R}^n$ tal que $F(y) \preceq F(\bar{x})$ e $F(\bar{x}) \neq F(y)$, respectivamente, $F(y) \prec F(\bar{x})$. A ordem $\preceq$ definida em $\mathbb{R}^m$ é coordenada à coordenada.

A iteração do método do gradiente para otimização multi objetivo consiste basicamente em: A partir de um ponto $x^k \in \mathbb{R}^n$ o próximo ponto x^{k+1} é gerado do seguinte modo:

$$x^{k+1} = x^k + t_k x^{k+1} = x^k + t_k v^k,$$

onde v^k é uma direção de descida e t_k é escolhido segundo a regra de Armijo. até aqui As iterações para o método de descida multi-objetivo consiste em tomar uma direção de descida $v^k \in \mathbb{R}^n$ e calcular $x^k \in \mathbb{R}^n$ nessa direção com comprimento de passo t_k e este é calculado utilizando a busca inexata de Armijo gerando assim a sequência $\{x^k\}$ que será por nós analisada e principalmente quando a função objetivo esta aplicada nos pontos da sequência, gerando a sequência $\{F(x^k)\}$ dos valores vetoriais. No estudo dessa classe de funções, quando necessário analisaremos o seu comportamento quando nesta adicionarmos a hipótese de quase convexidade.

Provaremos que a sequência dos valores vetoriais $\{F(x^k)\}$ é decrescente, além disso todo ponto de acumulação da sequencia $\{x^k\}$ gerada pelo método acima será um ponto Pareto Crítico. Ainda sob hipótese de quase convexidade e sendo $U = \{x \in \mathbb{R}^n | F(x) \preceq F(x^k)\}$ não vazio, a sequência $\{x^k\}$ gerada por esse mesmo método converge a um ponto Pareto Crítico. E por últiimo apresentaremos as conclusões e possíveis perguntas para um estudo futuro: Como se comportaria o Método Gradiente Projetado para resolver um problema de otimização Multi Objetivo?

CAPÍTULO 1

Conceitos Básicos

1.1 Resultados de Análise no $\mathbb{R}^n$

Nesta seção 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 importantes nos próximos 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 de produto interno que definiremos a seguir.

Definição 1.1 *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 $x, y \in \mathbb{R}^n$ 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, ou seja, um conceito necessário para definição de distância.

Definição 1.2 *Uma norma no espaço euclidiano n-dimensional $\mathbb{R}^n$ é definida como uma função real $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ 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] $x \neq 0 \Rightarrow \|x\| > 0$.

A seguir enunciaremos um teorema que nos dará a expressão para uma desigualdade muito e importante para análise em $\mathbb{R}^n$, que é conhecida como Desigualdade de Cauchy-Schwarz.

Teorema 1.3 *(Desigualdade de Cauchy-Schwarz) Sejam os pontos $x, y \in \mathbb{R}^n$ quaisquer, então para todo $x, y \in \mathbb{R}^n$ temos a seguinte desigualdade*

$$|\langle xy \rangle| \leq \|x\|\|y\|. \tag{1-1}$$

E a igualdade em (1-1) ocorre se, e somente se, os vetores x e y são múltiplos escalares, ou seja, se existe $\alpha \in \mathbb{R}$ tal que $x = \alpha y$.

Prova. Veja página 5 de [9]. □

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

Definição 1.4 *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 a distância ao ponto a é menor do que δ, isto é,*

$$B(a,\delta) = \{x \in \mathbb{R}^n;\ \|x - a\| < \delta\}.$$

Analogamente a bola fechada de centro a e raio δ é o conjunto

$$B[a,\delta] = \{x \in \mathbb{R}^n;\ |x - a| \leq \delta\}.$$

Agora que sabemos identificar bolas abertas em $\mathbb{R}^n$, podemos definir um ponto interior a um conjunto $U \subset \mathbb{R}^n$ e definir também conjunto aberto.

Definição 1.5 *Seja $U \subset \mathbb{R}^n$. Dizemos que $x \in U$ é um ponto interior a U quando é centro de alguma bola aberta inteiramente contida em U. E dizemos que um conjunto $U \subset \mathbb{R}^n$ é aberto quando todos os pontos de U são pontos interiores a U.*

Partindo dos conhecimentos sobre distâncias e bolas abertas e fechadas podemos falar em sequências. Uma *sequência* $\{x^k\} \subset \mathbb{R}^n$ é uma função $x\colon \mathbb{N} \to \mathbb{R}^n$, que associa a cada número $k \in \mathbb{N}$ um vetor $x^k \in \mathbb{R}^n$. 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 $\|x^k - a\| < \varepsilon$. Isto significa que, para qualquer $\varepsilon > 0$ dado, todos os elementos da sequência $\{x^k\}$ com exceção de k_0 elementos pertencem a bola $B(a, \varepsilon)$.

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

Teorema 1.6 *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 $\lim_{k_i \to \infty} x_i^k = a_i$, ou seja, cada coordenada de $\{x^k\}$ converge para a coordenada correspondente de a.*

Prova. Veja página 16 de [9]. □

Teorema 1.7 *O limite de uma sequência $\{x^k\} \subset \mathbb{R}^n$ convergente é único.*

Prova. Veja página 15 de [9]. □

Teorema 1.8 *Toda sequência $\{x^k\} \subset \mathbb{R}^n$ convergente é limitada.*

Prova. Veja página 15 de [9]. □

Teorema 1.9 *Se uma sequência $\{x^k\} \subset \mathbb{R}^n$ converge para $L \in \mathbb{R}^n$, então toda subsequência de $\{x^k\}$ também converge para L.*

Prova. Veja página 15 de [9]. □

Teorema 1.10 *(Teorema de Bolzano Weierstrass) Toda sequência limitada em $\mathbb{R}^n$ possui uma subsequência convergente.*

Prova. Veja página 17 de [9]. □

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

Teorema 1.12 *(Permanência do Sinal) Seja $\bar{a}$ um ponto de acumulação de $X \subset \mathbb{R}^n$ e $f : X \to \mathbb{R}$ uma função real. Se $b = \lim_{x\to\bar{a}} f(x)$ um número positivo (resp. negativo). Então existe $\delta > 0$ tal que $x \in X$ e $0 < |x-\bar{a}| < \delta$ implicam $f(x) > 0 (resp. f(x) < 0)$*

Prova. Ver página 34 de [9] □

Definição 1.13 *A função $f : U \to \mathbb{R}$, 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*

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

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 $\mathbb{R}^n$ é o Teorema do Valor Médio que enunciamos a seguir.

Teorema 1.14 *(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 124 e 139 de [9]. □

Definição 1.15 *Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita Lipschitziana se existir uma constante L, chamada constante de Lipschitz, que satisfaça para todo $x, y \in \mathbb{R}^n$ a seguinte condição*

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

1.2 Resultados de Otimização

Seremos breves nesta seção por se tratar de um assunto muito extenso, nos deteremos a conceitos e definições que aparecerão no decorrer deste trabalho. Definiremos alguns conceitos introdutórios de otimização, como minimizadores locais e globais e pontos estacionários e alguns resultados básicos.

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

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

O conjunto C é chamado conjunto viável do Problema (1-2) e a função f é chamada função objetivo do Problema (1-2).

Definição 1.16 *Seja a função $f : C \to \mathbb{R}$ com C subconjunto de $\mathbb{R}^n$, dizemos que $\bar{x} \in C$ é minimizador local do Problema* (1-2) *se existir $\varepsilon > 0$, tal que para todo $x \in B(x, \varepsilon) \subset C$ verificamos*

$$f(\bar{x}) \leq f(x). \tag{1-3}$$

E dizemos que $\bar{x}$ é minimizador global do Problema (1-2)*, quando a condição acima é satisfeita para todo $x \in C$.*

Observações 1.1 *Se a desigualdade* (1-3) *for estrita, então $\bar{x}$ será chamado, respectivamente, de minimizador estrito local e minimizador estrito global.*

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

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

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

Definição 1.17 *Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é um ponto estacionário ou crítico para Problema* (1-4)*, quando*

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

A definição de ponto estacionário é importante pois estes pontos são os candidatos naturais a minimizadores do Problema (1-4).

Teorema 1.18 *Seja $C \subset \mathbb{R}^n$ um conjunto aberto e suponha que a função $f : C \to \mathbb{R}$ seja diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local do Problema* (1-4)*, então*

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

Prova. Veja página 15 de [5] $\square$

Observações 1.2 *É importante ressaltar que o resultado anterior também é válido quando tratamos de problemas restritos a um conjunto viável C, desde que $\bar{x}$ esteja contido no interior do conjunto viável C.*

Quando no Problema (1-2) não consideramos o conjunto C como sendo o próprio $\mathbb{R}^n$ dizemos que temos um problema restrito a C e apresentamos este problema como em (1-2). A partir de agora o conjunto C será não vazio, convexo e fechado.

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

Definição 1.19 *Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é um ponto estacionário ou crítico para Problema (1-2) quando, para todo $x \in C$ obtemos*

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

Teorema 1.20 *(Condição Necessária de Primeira Ordem) Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in C$. Se $\bar{x}$ é um minimizador local de f no conjunto C, então para todo $x \in C$ obtemos que*

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0. \tag{1-5}$$

Prova. Vide página 66 em [5] □

Como no caso dos problemas irrestritos, um minimizador dos problemas restritos deve ser ponto estacionário do Problema (1-2).

Definição 1.21 *Dizemos que uma sequência $\{x^k\} \subset \mathbb{R}^n$ é dita Féjer convergente a um conjunto não vazio $U \neq \emptyset$ se, para todo $z \in U$ e todo $k > 0$ obtemos a seguinte desigualdade*

$$\|x^{k+1} - z\|^2 \leq \|x^k - z\|^2.$$

Definição 1.22 *Um sequência $\{x^k\} \subset \mathbb{R}^n$ é dita quase-Féjer convergente a um conjunto não vazio U se, para todo $z \in U$ e para todo k, existe uma sequência $\{\varepsilon_k\} \subset \mathbb{R}_+$ somável, tal que*

$$\|x^{k+1} - z\|^2 \leq \|x^k - z\|^2 + \varepsilon_k.$$

Teorema 1.23 *Sejam $T \subset \mathbb{R}^n$ um conjunto não vazio e $\{x^k\} \subset \mathbb{R}^n$ uma sequência satisfazendo para todo $x \in T$ e todo k a seguinte desigualdade*

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \varepsilon_k, \tag{1-6}$$

onde $\{\varepsilon_k\} \subset \mathbb{R}_+$ é uma sequência somável. Então

i) *A sequência $\{x^k\}$ é limitada;*

ii) *Se $\bar{x}$ é um ponto de acumulação de $\{x^k\}$ e $\bar{x} \in T$, então $\{x^k\}$ converge a $\bar{x}$.*

Prova. Iremos provar então o primeiro item do teorema. Tomemos inicialmente um ponto fixo $x \in T$. Utilizando a Desigualdade (1-6) repetidas vezes obtemos

$$\|x^k - x\|^2 \leq \|x^0 - x\|^2 + \sum_{j=0}^{\infty} \varepsilon_j.$$

Como por hipótese, a série $\sum_{j=0}^{\infty} \varepsilon_j$ é convergente, sendo a^0 e x fixos podemos concluir que a sequência $\{x^k\}$ é limitada e

$$\|x^k - x\| \leq R,$$

onde tomamos $R = \left(\|x^0 - x\|^2 + \sum_{j=0}^{\infty} \varepsilon^j\right)^{1/2}$.

Provaremos agora o segundo item do teorema. Pelo item anterior temos que a sequência admite pontos de acumulação. Seja então $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$, isto implica que existe uma subsequência $\{x^{k_s}\}$ convergindo a $\bar{x}$.

Como $\{\varepsilon_j\}$ é uma sequência somável então existe s_0 tal que

$$\sum_{j=s_0}^{\infty} \varepsilon_{k_j} < \delta/2.$$

E existe um s_1 de modo que $k_{s_1} > s_0$ e então para todo $s > s_1$ vale a seguinte desigualdade

$$\|x^{k_s} - \bar{x}\|^2 \leq \delta/2.$$

Portanto, utilizando as duas desigualdades acima e a desigualdade (1-6) concluímos que para qualquer $s > k_{s_1}$ obtemos

$$\|x^k - \bar{x}\| < \delta,$$

e isto implica que a sequência $\{x^k\}$ converge a $\bar{x}$. Concluindo assim a demonstração do teorema. □

1.3 Resultados de Análise Convexa

Nesta seção exibiremos resultados e definições relacionados à análise Convexa. Para isso dividiremos esta seção em duas subseções: na primeira trataremos de funções convexas e de conjuntos convexos e na segunda definiremos função pseudo convexa e quase convexa, além de alguns resultados importantes. Começaremos definindo um conjunto convexo em $\mathbb{R}^n$. Este conceito será exigido no próximo capítulo deste trabalho, quando estivermos trabalhando com problemas de otimização restritos.

Definição 1.24 *Um conjunto $C \subset \mathbb{R}^n$ é dito convexo, se para todo λ no intervalo $[0,1]$ e para todo $x, y \in C$ tivermos*

$$\lambda x + (1-\lambda)y \in C.$$

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 C, sempre que os pontos extremos x e y estão em C.

Exemplo 1.3 *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.*

Definição 1.25 *Um conjunto $K \subset \mathbb{R}^n$ chama-se cone quando para todo $t \in \mathbb{R}_+$ obtemos*

$$d \in K \qquad \Rightarrow \qquad td \in K.$$

Definição 1.26 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in C$. O cone normal(cone de direções normais) no ponto $\bar{x}$ em relação ao conjunto C é dado por*

$$\aleph_C(\bar{x}) = \{d \in \mathbb{R}^n : \langle d, x - \bar{x}\rangle \leq 0 \quad x \in C\}.$$

Outro aspecto relevante a ser destacado é a definição de função convexa, que é um conceito bastante importante em otimização com o qual obtemos resultados também muito importantes.

Definição 1.27 ***(Funções Convexas)*** *Seja C um conjunto convexo. Uma função f é dita convexa em se para todo $x, y \in C$ e $\lambda \in [0,1]$ obtemos*

$$f((1-\lambda)y + \lambda x) \leq (1-\lambda)f(y) + \lambda f(x). \tag{1-7}$$

Definição 1.28 *O epígrafo da função $f : C \subset \mathbb{R}^n \to \mathbb{R}$ é definido como o conjunto*

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

Teorema 1.29 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \to \mathbb{R}$ é convexa em C se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Prova. Veja página 58 de [10]. $\square$

Definição 1.30 *O conjunto de nível de uma função $f : C \to \mathbb{R}$ associado ao número real $\bar{a} \in \mathbb{R}$, é o conjunto dado por*

$$L_{f,C}(\bar{a}) := \{x \in C : f(x) \leq \bar{a}\}.$$

Teorema 1.31 *Seja f uma função real definida no conjunto convexo $C \subset \mathbb{R}^n$. Uma condição necessaria mas não suficiente para que f seja convexa em C, é que o conjunto de nível $L_{f,C}(\bar{a})$ seja convexo para todo número real $\bar{a}$.*

Prova. Veja página 59 de [10]. $\square$

Exemplo 1.4 *A função real $f(x) = x^3$ é um exemplo de função em que $L_{f,C}(a)$ seja convexo para todo número real a, mas não é uma função convexa. Nos mostrando assim que condição não é suficiente para que f seja convexa.*

Embora, considerando o fato de que as 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 generalizadas, isto é, as funções quase convexas ou pseudo convexas.

Definição 1.32 *Uma função f diz-se estritamente convexa quando a desigualdade em (1-7) é estrita para todos $x \neq y$ e $\alpha \in (0,1)$.*

Observações 1.5 *Note que se uma função f é estritamente convexa, então ela será convexa. Porém a recíproca não é verdadeira.*

Exemplo 1.6 *Seja $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x$. Podemos observar que f é convexa, mas não é estritamente convexa.*

Definição 1.33 *Uma função f diz-se fortemente convexa em C com módulo $\gamma > 0$ quando para quaisquer $x, y \in C$ e $\alpha \in [0,1]$, obtemos que*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) - \gamma\alpha(1-\alpha)\|x-y\|^2.$$

Exemplo 1.7 *A função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$ é fortemente convexa.*

Observações 1.8 *Podemos observar que se uma dada função f é fortemente convexa implica que ela é estritamente convexa e assim convexa. Agora tomando f como sendo $f(x) = x$ e utilizando o exemplo anterior obtemos que ela é convexa, mas não é fortemente convexa.*

Proposição 1.34 *Uma função f é fortemente convexa em C com módulo $\gamma > 0$ se, e somente se a função $f - \gamma\|.\|^2/2$ é convexa em C.*

Prova. Ver página 73, Proposição 1.1.2 de [**?**]. $\square$

Teorema 1.35 *(Teorema de Minimização Convexa) Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função convexa. Então todo minimizador local do problema é global e se f é estritamente convexa a solução é única.*

Prova. Ver página. 77 de [5]. $\square$

Definição 1.36 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Dizemos que $y \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in \mathbb{R}^n$ quando para todo $z \in \mathbb{R}^n$ obtemos*

$$f(z) \geq f(x) + \langle y, z - x \rangle.$$

O conjunto de todos os subgradientes de f em x denomina-se subdiferencial de f em x e denotado por $\partial f(x)$.

Teorema 1.37 *(Condições de Otimalidade) Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $C \subset \mathbb{R}^n$. Então $v \in \mathbb{R}^n$ é minimizaodr de f em C se, e somente se, existe $y \in \partial f(x)$ tal que para todo $x \in C$ obtemos*

$$\langle y, x - v \rangle \geq 0,$$

ou, equivalentemente,

$$0 \in \partial f(v) + \aleph_C(v).$$

Em particular, concluímos que o ponto v será um minimizador de f em $\mathbb{R}^n$ se, e somente se, ocorrer a seguinte condição

$$0 \in \partial f(v).$$

Prova. Ver página 177 em [5]. $\square$

Proposição 1.38 *Sejam $f_1, ..., f_m$ funções convexas de $\mathbb{R}^n$ em $\mathbb{R}$ e definimos a função f como*

$$f := \max\{f_1, ..., f_m\},$$

onde $I(x) = \{i \in I | f_i(x) = f(x)\}$ e $I = \{1, 2, ...m\}$ são conjuntos de índices. Então obtemos que

$$\partial f(x) = co\left\{\bigcup \partial f_i(x) : i \in I(x)\right\}.$$

Prova. Ver página 266 de [**?**] $\square$

1.3.1 Funções Convexas Generalizadas

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

Definição 1.39 ***(Funções Quase-Convexas)*** *Seja C um conjunto não vazio, fechado e convexo. Uma função real f definida sobre um conjunto $C \subset \mathbb{R}^n$ é dita quase convexa em C se para todo $x, y \in C$ obtemos*

$$f((1-\lambda)\bar{x}+\lambda x) \leq \max\{f(x), f(y)\}.$$

Exemplo 1.9 *Pelo Exemplo 1.4 já sabemos que $f(x) = x^3$ não é uma função convexa. Mas podemos notar que ela satisfaz a definição de quase convexidade.*

A seguir enunciaremos um teorema que nos dará uma forma equivalente de identificar uma função quase-convexa.

Teorema 1.40 *Seja f uma função real definida no conjunto convexo $C \subset \mathbb{R}^n$. Uma condição necessária e suficiente para que f seja quase convexa em C, é que o conjunto $L_{f,C}(a)$ seja convexo para todo número real a.*

Prova. Veja página 133 de [10]. □

Observações 1.10 *Pelo Teorema 1.31 juntamente com o Teorema 1.40 vemos que a classe de funções quase convexas contém a classe de funções convexas.*

Exemplo 1.11 *Ainda analisando a função $f(x) = x^3$ observamos que $L_{f,C}(a)$ é convexo para todo número real a, então pelo Teorema 1.40 concluímos que f é uma função quase convexa.*

Diante da 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 demonstrações dos últimos capítulos.

Lema 1.41 *Sejam $C \subset \mathbb{R}^n$ aberto e $x, y \in C$ e $f : C \to \mathbb{R}$ diferenciável e quase convexa em x. Se $f(y) \leq f(x)$ então obtemos*

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

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

$$\tilde{x} = x + \tilde{\mu}(y-x) = (1-\tilde{\mu})x + \tilde{\mu}y. \tag{1-8}$$

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

$$f((1-\lambda)x + \lambda\tilde{x}) \leq f(x), \qquad 0 < \lambda \leq 1.$$

Pela diferenciabilidade da função f em $x \in C$ obtemos a seguinte desigaldade,

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

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 (1-8) temos o que queremos demonstrar. □

Definição 1.42 ***(Funções Estritamente Quase-Convexas)*** *Seja C um conjunto não vazio, fechado e convexo. Uma função real f definida sobre um conjunto $C \subset \mathbb{R}^n$ é dita estritamente quase-convexa em C se para todo $x, y \in C$ e $f(x) \neq f(y)$ obtemos*

$$f((1-\lambda)\bar{x} + \lambda x) < \max\{f(x), f(y)\}.$$

Segue abaixo um exemplo de uma função que é estritamente quase-convexa e no entanto não é quase-convexa

Exemplo 1.12 *Seja f uma função real definida da seguinte maneira*

$$f(x) = \begin{cases} 1 & se\ x = 0 \\ 0 & se\ x \neq 0 \end{cases}$$

Observações 1.13 *Podemos notar que esta função é estritamente quase-convexa. No entanto, ela não é quase-convexa, e para verificar isso basta tomar $x_1 = -1$, $x_2 = 1$ e $\lambda = 1/2$. Assim para que uma função estritamente quase-convexa seja quase-convexa é necessários alguma hipótese sobre a função f como, por exemplo, a continuidade.*

Agora definiremos uma função pseudo-convexa. Este conceito será exigido para o estudo do próximo capítulo, no qual trabalharemos com funções com essa característica.

Definição 1.43 ***(Função Pseudo-Convexa)*** *Seja f uma função real definida sobre algum conjunto aberto do $\mathbb{R}^n$ contendo o conjunto C. A função f é dita pseudo-convexa sobre $\bar{x} \in C$ se é diferenciável em $\bar{x}$ e para algum $x \in C$ satisfazendo*

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

implica que

$$f(x) \geq f(\bar{x}).$$

Dizemos que f é pseudo-convexa sobre C, se é pseudo-convexa em cada ponto x de C.

Exemplo 1.14 *Seja $f : \mathbb{R}^2 \to \mathbb{R}$ dada por $f(x,y) = x^2 + y^2$. Tomando $\bar{x} = (0,0)$ e $x = (1,1)$ obtemos*

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

e implica que

$$f(x) \geq f(\bar{x}).$$

Satisfazendo assim a nossa definição, ou seja, a função f é pseudo convexa.

Teorema 1.44 *Sejam $f : C \to \mathbb{R}$ uma função pseudo-convexa em $\bar{x} \in C$ e $C \subset \mathbb{R}^n$ tal que para todo $x \in C$ satisfaça*

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

Então obtemos que

$$f(\bar{x}) = \min_{x \in C} f(x).$$

Prova. Pela definição de função pseudo-convexa para algum $x \in C$ satisfazendo

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

implica que $f(x) \geq f(\bar{x})$. Mas como a desigualdade acontece para todo $x \in C$ obtemos o resultado desejado. □

Corolário 1.45 *Seja $f : C \to \mathbb{R}$ uma função pseudo-convexa em $\bar{x} \in C$, onde $C \subset \mathbb{R}^n$. Se o gradiente da função f em $\bar{x}$ se anula então*

$$f(\bar{x}) = \min_{x \in C} f(x).$$

Prova. Segue diretamente do teorema anterior. □

Teorema 1.46 *(Karamardian 67) Seja f uma função real semi continua superiormente(inferiormente) definida em um convexo $C \subset \mathbb{R}^n$. Se f é estritamente quase-convexa em C, então f é quase-convexa em C, mas não vale a recíproca.*
Prova. Veja página 139 de [10]. $\square$

Teorema 1.47 *Sejam f uma função definida em um conjunto convexo $C \subset \mathbb{R}^n$ e $\bar{x} \in C$ um mínimo local da função. Se f é estritamente quase-convexa, então $\bar{x}$ é um mínimo global de f em C.*

Prova. Veja página 139 de [10]. $\square$

Teorema 1.48 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo, e f uma função real definida em algum conjunto aberto que contém C. Se f é pseudo-convexa em C então f é estritamente quase-convexa em C e assim será quase-convexa em C.*

Prova. Veja página 143 de [10]. $\square$

Observações 1.15 *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.*

No próximo exemplo, exibiremos uma função que ilustrará o que foi exposto na observação anterior.

Exemplo 1.16 *Considere a função real $f(x) = x^3$. Já sabemos que f é uma função quase-convexa. Como $x = 0$ é ponto estacionário de f e não é ponto de mínimo, pelo Corolário 1.45 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 o contrário. A seguir veremos que a classe de funções convexas é um subconjunto do conjunto das funções pseudo-convexas.

Teorema 1.49 *Seja f uma função real definida em um conjunto aberto $C \subset \mathbb{R}^n$. 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 [10]. $\square$

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.

1.4 Projeções

Nesta seção trataremos sobre os operadores de projeção ortogonal de um ponto $x \in \mathbb{R}^n$ sobre um conjunto $C \subset \mathbb{R}^n$. Os resultados desta seção serão exigidos quando estivermos trabalhando com problemas de otimização restritos no próximo camítulo.

Uma projeção (ortogonal) do ponto $x \in \mathbb{R}^n$ sobre um conjunto $C \subset \mathbb{R}^n$ é um ponto pertencente a C que está mais próximo de x. Utilizamos aqui a distância como sendo a métrica induzida pela norma euclidiana. Assim podemos dizer que a projeção de $x \in \mathbb{R}^n$ sobre C é uma solução global do seguinte problema

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

Definição 1.50 *Definiremos a projeção de $x \in \mathbb{R}^n$ sobre C por*

$$P_C(x) := argmin\{\|y - x\| : y \in C\}. \tag{1-9}$$

Teorema 1.51 ***(Teorema da Projeção)*** *Seja $C \subset \mathbb{R}^n$ um conjunto fechado e não-vazio. Então para todo ponto $x \in \mathbb{R}^n$, existe uma projeção de x sobre C. Além disso, se C for convexo então para todo ponto $x \in \mathbb{R}^n$, existe uma única projeção de x sobre C.*

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

Consideraremos nos lemas subsequentes que o conjunto $C \in \mathbb{R}^n$ seja não-vazio, fechado e convexo

Lema 1.52 *Considerando P a projeção sobre C. Temos as seguintes propriedades*

(a) *Se $x \in C$, 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) *O operador projeção é não expansivo, isto é:*

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

Prova. **(a)** Consideremos $y(\alpha) = (1-\alpha)P(y) + \alpha x$. Como C é um conjunto convexo e $P(y) \in C$, conclui-se que $y(\alpha) \in C$ para todo $x \in C$ e para $\alpha \in [0,1]$. E pela definição de projeção (1-9), temos

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

Desenvolvendo a expressão anterior obtemos a seguinte desigualdade

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

Agora, basta fazer α tender a zero e nosso primeiro resultado está demonstrado.

Provaremos agora o item (b) do lema. **(b)** Sejam $x,\,y\in\mathbb{R}^n$, como $P(z)\in C$, 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 $P(y)\in C$, obtemos de forma análoga a desigualdade anterior a seguinte expressão

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

Agora, combinando as desigualdades anteriores, obtemos o resultado desejado neste item.. O item (c) decorre segue diretamente o item anterior. (d) Veja página 98 de [5]. □

Lema 1.53 *Seja P a projeção sobre C. Considere $x\in C$, e defina*

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

Então

(a) *$\bar{x}$ é ponto estacionário se, e somente se, $\bar{x}=P_C(\bar{x}-\alpha\nabla f(x))$;*

(b) $\langle x(\alpha)-x+\alpha\nabla f(x)\,,\,y-x(\alpha)\rangle\geq 0$ *para todo $y\in C$ e $\alpha>0$;*

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

*Prova.***(a)** Suponha inicialmente que valha $\bar{x}=P_C(\bar{x}-\alpha\nabla f(x))$, mas isso acontece se, e somente se, $\bar{x}$ é uma solução do problema

$$\min\psi(x)\quad\text{sujeito a } x\in C, \tag{1-10}$$

onde a função $\psi:\mathbb{R}^n\to\mathbb{R}$ é definida da seguinte maneira

$$\psi(x)=\frac{1}{2}\|x-(\bar{x}-\alpha\nabla f(\bar{x})\|^2,\qquad \alpha>0.$$

Pelo Teorema (1.20), para todo $x\in C$ obtemos a seguinte desigualdade

$$0\leq\langle\nabla\psi(\bar{x}),x-\bar{x}\rangle.$$

Calculando o gradiente da função ψ em $\bar{x}$ e usando o fato de α ser maior que zero obtemos a desigualdade

$$0 \leq \langle \nabla f(x), x - \bar{x} \rangle,$$

concluindo que $\bar{x}$ é ponto estacionário. □

Prova. **(b)** e **(c)** Basta fazer uma aplicação direta do item (a) e do item (b) do Lema 1.52 para demonstrar, respectivamente, o item (a) e o item (b) deste teorema. □

CAPÍTULO 2

Método do Gradiente Projetado para Problemas de Otimização Escalares

Neste Capítulo estudaremos o Método do Gradiente Projetado para buscar soluções em problemas com restrição a um conjunto viável $C \subset \mathbb{R}^n$ não vazio, fechado e convexo e com a função objetivo $f : \mathbb{R}^n \to \mathbb{R}$ continuamente diferenciável e convexa no sentido generalizado, isto é, quase-convexa ou pseudo-convexa. Analisaremos as propriedades de convergência do método com busca linear de Armijo. Considere o seguinte problema de otimização

$$\min f(x) \quad \text{sujeito a} \quad x \in C. \tag{2-1}$$

O Método do Gradiente Projetado com regra de Armijo para resolver o problema (2-1) é descrito formalmente da seguinte forma:

Algoritmo 1 (Método do Gradiente Projetado com regra de Armijo)

INICIALIZAÇÃO. *Tome* $0 < \widehat{\beta} < \overline{\beta}$ *e* $x^k \in C$. *Faça* $k = 0$.
CRITÉRIO DE PARADA. *Se* x^k *é estacionário então PARE. Caso contrário.*
PASSO ITERATIVO. *Calcule*

$$z^k = P_C(x^k - \beta_k \nabla f(x^k)), \qquad x^{k+1} = x^k + \gamma_k(z^k - x^k),$$

onde $\{\beta_k\} \subset [\widehat{\beta}, \tilde{\beta}]$ *tal que* $\gamma_k = 2^{-j(k)}$ *sendo que*

$$j(k) = \min\left\{ j \in \mathbb{N} : f(x^{k+1}) \leq f(x^k) + \sigma 2^{-j} \langle \nabla f(x^k), z^k - x^k \rangle \right\}, \qquad \sigma \in (0,1).$$

e VOLTE para CRITÉRIO DE PARADA.

Nosso objetivo de agora em diante é estudar a boa definição da sequência $\{x^k\}$ gerada pelo método e suas propriedades de convergência. A sequência $\{x^k\}$ é gerada basicamente da seguinte forma: Dado $x^k \in C$, damos uma passo β_k na direção oposta ao gradiente, isto é, $-\nabla f(x^k)$ o vetor resultante

$$x^k - \beta_k \nabla f(x^k),$$

é projetado sobre o conjunto viável C. Se a projeção for o próprio ponto x^k, isto é,

$$P_C(x^k - \beta_k \nabla f(x^k)) = x^k,$$

segue do Lema 1.53 que x^k é ponto estacionário para o problema e assim o método para. Caso contrário, defina

$$z^k = P_C(x^k - \beta_k \nabla f(x^k)),$$

e faça uma iteração a partir de x^k na direção do vetor $z^k - x^k$ com comprimento de passo γ_k, obtendo o próximo ponto

$$x^{k+1} = x^k + \gamma_k (z^k - x^k).$$

Note que para mostrar a boa definição da sequência $\{x^k\}$ devemos mostar que número inteiro positivo $j(k)$ pode ser calculado a cada passo, isto será mostrado na Proposição 2.2 abaixo.

2.1 Resultados Preliminares

Nesta seção iremos estudar alguns resultados necessários a análise de convergência do Algoritmos 1. Começamos com o seguinte resultado sobre projeções em conjuntos convexo.

Proposição 2.1 *Tomemos* $\sigma \in (0,1)$*, um ponto* $x \in C$ *arbitrário e* $v \in \mathbb{R}^n$ *tal que* $\langle \nabla f(x), v \rangle < 0$*. Então, existe* $\bar{\gamma} < 1$ *tal que para todo* $\gamma \in (0, \bar{\gamma}]$ *obtemos*

$$f(x + \gamma v) < f(x) + \sigma \gamma \langle \nabla f(x), v \rangle.$$

Prova. Como a função objetivo f é diferenciável, temos imediatamente pela definição que

$$f(x + \gamma v) - f(x) = \langle \nabla f(x), \gamma v \rangle + o(\gamma), \qquad \textit{onde} \qquad \lim_{\gamma \to 0} \frac{o(\gamma)}{\gamma} = 0$$

A partir da igualdade acima e com simples manipulações algébricas e em seguida fazendo γ ir a zero temos, utilizando o Teorema 1.12 que para todo $\gamma \in (0, \bar{\gamma}]$ a seguinte desigualdade

$$f(x + \gamma v) < f(x) + \sigma \gamma \langle \nabla f(x), v \rangle.$$

Justamente o que queríamos demonstrar. $\square$

2.2 Análise de Convergência

Iremos dividir a análise da convergência do Algoritmo 1 em duas partes, a convergência parcial e a convergência total da sequência gerada pelo algoritmo.

2.2.1 Convergência parcial

Nesta seção vamos estudar propriedades de convergência da sequência $\{x^k\}$ gerada pelo Algoritmo 1, apenas com a já mencionada diferenciabilidade, assumindo assim somente existência de solução.

Proposição 2.2 *Sejam $\{x^k\}$ e $\{z^k\}$ as sequências geradas pelo Algoritmo 1. Então:*

i. *A iteração $x^k \in C$ para todo k;*

ii. *Se x^k não é ponto estacionário, então $\langle \nabla f(x^k), z^k - x^k \rangle < 0$;*

iii. *Para cada k, $j(k)$ está bem definido.*

Prova. Vamos provar o item (i). Usaremos para essa prova indução no índice k das sequências $\{x^k\}$ e $\{z^k\}$. É imediato que $x^o \in C$, logo $z^o \in C$. Suponha então que a afirmação seja válida para k, ou seja, $x^k \in C$, e daí segue imediatamente que $z^k \in C$, nos restando provar que $x^{k+1} \in C$. Pelo algoritmo 1 temos que:

$$x^{k+1} = x^k + \gamma_k (z^k - x^k).$$

Reescrevendo de maneira conveniente a igualdade acima obtemos a seguinte igualdade

$$x^{k+1} = (1 - \gamma_k) x^k + \gamma_k z^k.$$

Como $x^k, z^k \in C$, utilizando a convexidade do conjunto C podemos concluir que $x^{k+1} \in C$. Desta forma, acabamos de mostrar que $x^k \in C$ para todo k.

Provaremos agora o item (ii). Como $x^k \in C$ para todo k e $\beta_k > 0$, aplicando o item (c) do Lema 1.53 obtemos a seguinte desigualdade

$$\langle \nabla f(x^k), x^k - z^k \rangle \geq \frac{\|z^k - x^k\|^2}{\beta_k}, \qquad k = 0, 1, \ldots.$$

Como x^k por hipótese não é ponto estacionário para o problema, utilizando novamente o Lema 1.53 e a desigualdade acima concluímos facilmente a demonstração.

Provaremos enfim o último item da proposição. Pelo item anterior, para todos os pontos x^k e $z^k \in C$ com x^k não sendo estacionário para o problema temos que

$$\langle \nabla f(x^k), z^k - x^k \rangle \; < \; 0.$$

Usando a última desigualdade e tomando $x = x^k \in C$ e $v = z^k - x^k \in \mathbb{R}^n$ na Proposição 2.1 concluímos que existe $j(k)$ tal que

$$f(x^k + \gamma_j(z^k - x^k)) \leq f(x^k) + \sigma 2^{-j(k)} \langle \nabla f(x^k), z^k - x^k \rangle, \qquad \sigma \in (0,1).$$

Logo $j(k)$ está bem definido. □

Proposição 2.3 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 1. Se o problema inicial (2-1) possui solução, então todo ponto de acumulação da sequencia $\{x^k\}$ é um ponto estacionário para o problema.*

Prova. Sejam $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$ e $\{x^{k_s}\}$ uma subsequência de $\{x^k\}$ convergindo para $\bar{x}$. Como C é fechado e convexo, pela Proposição 2.2 item (i) temos $\bar{x} \in C$. Além disso, Como f é contínua, segue imediatamente que

$$\lim_{k_s \to \infty} f(x^{k_s}) = f(\bar{x}). \tag{2-2}$$

Sabemos também que $\gamma_k \in (0,1)$ e $\beta_k \in [\widetilde{\beta}, \bar{\beta}]$, assim sem perda de generalidade podemos considerar

$$\lim_{k_s \to \infty} \gamma_{k_s} = \bar{\gamma} \in [0,1] \quad e \quad \lim_{k_s \to \infty} \beta_{k_s} = \bar{\beta} \geq \widetilde{\beta} > 0.$$

Note que sendo f continuamente diferenciável, o operador projeção contínuo, que $\{x^{k_s}\}$ converge para $\bar{x}$ e que $\{\beta_{k_s}\}$ converge para $\bar{\beta}$ obtemos que

$$\lim_{k_s \to \infty} z^{k_s} = P_C(\bar{x} - \bar{\beta} \nabla f(\bar{x})) =: \bar{z}.$$

Por outro lado, usando a definição de γ_k no Algoritmo 1 obtemos

$$0 < -\sigma \gamma_k \langle \nabla f(x^k), z^k - x^k \rangle \leq f(x^k) - f(x^{k+1}). \tag{2-3}$$

Da desigualdade acima concluímos facilmente que a sequência $\{f(x^k)\}$ é monótona decrescente, e como por hipótese f é contínua e o problema possui ao menos uma solução segue-se que $\{f(x^k)\}$ converge. Portanto, usando (2-2) concluímos que

$$\lim_{k \to \infty} f(x^k) = f(\bar{x}).$$

Tomando limite quando k_s tende ao infinito na desigualdade (2-3) e usando a igualdade anterior e que f é continuamente diferenciável temos

$$\sigma \bar{\gamma} \langle \nabla f(\bar{x}), \bar{z} - \bar{x} \rangle = 0. \tag{2-4}$$

Existem dois casos a serem considerados: quando $\bar{\gamma} > 0$ e quando $\bar{\gamma} = 0$.
1º Caso: Iremos considerar primeiro o caso em que $\bar{\gamma} > 0$. Neste caso, como $\sigma > 0$ segue imediatamente de (2-4) que

$$\langle \nabla f(\bar{x}), \bar{z} - \bar{x} \rangle = 0. \tag{2-5}$$

Fazendo $\bar{u} = \bar{x} - \bar{\beta}\nabla f(\bar{x})$ e usando o item (a) do Lema (1.52) concluímos facilmente que

$$0 \leq \langle \bar{z} - \bar{u}, \bar{x} - \bar{z} \rangle.$$

Combinando a desigualdade anterior e a igualdade (2-5) com simples manipulação chegamos a

$$\bar{z} = \bar{x},$$

e utilizando o item (a) do Lema 1.53 concluímos assim que $\bar{x}$ é um ponto estacionário para o problema.
2º caso: Iremos considerar agora o caso em que $\bar{\gamma} = 0$. Neste caso, como $\bar{\gamma} = 0$ a definição de γ_{k_s} no Algoritmo (1) implica que para todo $q \in \mathbb{N}$ fixo e como $\gamma_{k_s} = 2^{-j(k_s)}$, existe um k_o tal que para todo $k_s > k_o$ temos $j(k_s) > q$ e vale

$$f(x^{k_s} + 2^{-q}(z^{k_s} - x^{k_s})) > f(x^{k_s}) + \sigma 2^{-q}\langle \nabla f(x^{k_s}), z^{k_s} - x^{k_s} \rangle. \tag{2-6}$$

Agora tomando limite quando k_s tende ao infinito na expressão (2-6), usando que $\{x^{k_s}\}$ converge para $\bar{x}$ e a hipótese que f é continuamente diferenciável temos

$$f(\bar{x} + 2^{-q}(\bar{z} - \bar{x})) \geq f(\bar{x}) + \sigma 2^{-q}\langle \nabla f(\bar{x}), \bar{z} - \bar{x} \rangle.$$

Desde que a última desigualdade vale para todo q, usando a Proposição (2.1) e o fato de $\sigma > 0$ obtemos que

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

Como $\{x^{k_s}\}$ converge para $\bar{x}$ e $\{z^{k_s}\}$ converge para $\bar{z}$ o item (ii) da Proposição (2.2) implica

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

Finalmente, combinado as duas últimas desigualdade obtemos a igualdade

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

Portanto, de modo análogo ao caso anterior concluímos que $\bar{x}$ é um ponto estacionário para o problema e assim a prova esta concluida. □

2.2.2 Convergência Total

Já sabemos que se o problema inicial (2-1) possui solução, então todo ponto de acumulação da sequência gerada pelo Algoritmo 1 é um ponto estacionário para o problema. Agora iremos mostrar que a sequência gerada pelo Algoritmo 1 converge, isto sob a hipótese de quase-conexidade ou pseudo-conexidade da função objetivo. Começaremos apresentando a seguinte desigualdade fundamental para nossa análise.

Lema 2.4 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 1. Para todo $x \in C$ e todo k temos*

$$\|x^{k+1}-x\|^2 \leq \|x^k-x\|^2+\eta[f(x^k)-f(x^{k+1})]+2\gamma_k\beta_k\langle\nabla f(x^k),x-x^k\rangle, \tag{2-7}$$

onde $\eta=(2\tilde{\beta})/\sigma$.

Prova. Usando a definição de x^{k+1} no Algoritmo 1, após algumas manipulações algébricas obtemos a seguinte igualdade

$$\|x^{k+1}-x\|^2=\|x^{k+1}-x^k\|^2+\|x^k-x\|^2-2\gamma_k\langle z^k-x^k,x-x^k\rangle. \tag{2-8}$$

Segue do Algoritmo 1 que z^k é a projeção de $x^k-\beta_k\nabla f(x^k)$ sobre C. Então fazendo $\alpha=\beta_k$, $y=x$ e $x=x^k$ no Lema 1.53 item (ii) e usando a definição de z^k no Algoritmo 1 obtemos

$$0\leq\langle z^k-x^k+\beta_k\nabla f(x^k),x-z^k\rangle.$$

Após alguns cálculos simples, obtemos que a desigualdade acima é equivalente a

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

assim, passados os dois últimos termos para o lado esquerdo da desigualdade acima temos

$$\langle z^k-x^k,x-x^k\rangle\geq\beta_k\langle\nabla f(x^k),z^k-x\rangle+\|z^k-x^k\|^2.$$

Usando a definição de x^{k+1} no Algoritmo 1, a desigualdade acima é equivalente a

$$2\gamma_k\langle z^k-x^k,x-x^k\rangle \geq\ 2\gamma_k\beta_k\langle\nabla f(x^k),z^k-x\rangle+2\frac{\|x^{k+1}\ -\ x^k\|^2}{\gamma_k}.$$

Combinando a última desigualdade com a igualdade em (2-8) obtemos

$$\|x^{k+1}-x\|^2\leq\ \|x^k-x\|^2+(1-\frac{2}{\gamma_k})\|x^{k+1}-x^k\|^2+2\gamma_k\beta_k\langle\nabla f(x^k),x-z^k\rangle.$$

Pela definição de γ_k no Algoritmo 1 temos que $1 - 2/\gamma_k < 0$. Então da desigualdade anterior obtemos

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + 2\gamma_k\beta_k\langle \nabla f(x^k), x - z^k\rangle. \tag{2-9}$$

Para finalizar a prova, vamos estimar o último termo da desigualdade anterior. Primeiro note que

$$2\gamma_k\beta_k\langle \nabla f(x^k), x - z^k\rangle = 2\gamma_k\beta_k\langle \nabla f(x^k), x - x^k\rangle + 2\gamma_k\beta_k\langle \nabla f(x^k), x^k - z^k\rangle.$$

Por outro lado, a definição de γ_k no Algoritmo 1 implica que

$$2\gamma_k\beta_k\langle \nabla f(x^k), x^k - z^k\rangle \leq \frac{2\beta_k}{\sigma}[f(x^k) - f(x^{k+1})].$$

Combinando a última igualdade com a desigualdade anterior e levando em consideração que $\{f(x^k)\}$ é monótona decrescente e $\beta_k \leq \tilde{\beta}$ obtemos que

$$2\gamma_k\beta_k\langle \nabla f(x^k), x - z^k\rangle \leq \frac{2\tilde{\beta}}{\sigma}[f(x^k) - f(x^{k+1})] + 2\gamma_k\beta_k\langle \nabla f(x^k), x - x^k\rangle.$$

Portanto, a desigualdade desejada seque de (2-9) e da desigualdade anterior. □

Definiremos agora o conjunto $T \subset \mathbb{R}^n$ da seguinte forma

$$T = \{x \in C \ : \ f(x) \leq f(x^k), \quad k = 0, 1, 2, ...\}.$$

O conjunto definido acima será de fundamental importância para as demonstrações dos próximos resultados.

Observações 2.1 *Podemos notar que se o problema possuir ao menos uma solução ou a sequência $\{x^k\}$ gerada pelo Algoritmo 1 possui pontos de acumulação então $T \neq \emptyset$.*

Corolário 2.5 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 1. Se f é quase-convexa e $T \neq \emptyset$ então a sequência $\{x^k\}$ é convergente.*

Prova. Seja $x \in T$. A definição de T resulta que $f(x) \leq f(x^k)$ para todo k. Como f é quase-convexa implica pelo Lema 1.41 que

$$\langle \nabla f(x^k), x - x^k\rangle \leq 0.$$

Como por hipótese $x \in C$, combinando o Lema 2.4 com a desigualdade anterior obtemos

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \eta[f(x^k) - f(x^{k+1})], \qquad k = 0, 1, 2 \ldots. \tag{2-10}$$

onde $\eta = (2\tilde{\beta})/\sigma$. Agora definimos abaixo a sequência $\{\varepsilon_k\}$ como sendo

$$\varepsilon_k = \eta(f(x^k) - f(x^{k+1})).$$

Como a sequência $\{f(x^k)\}$ é monótona decrescente segue-se que $\varepsilon_k > 0$ para todo k. Agora note que

$$\sum_{k=0}^{s} \varepsilon_k = \eta(f(x^0) - f(x^{s+1})) \leq \eta f(x^0) - f(\bar{x}),$$

o que implica que a sequência $\{\varepsilon_k\}$ é somável. Portanto, concluímos de (2-10) que a sequência $\{x^k\}$ é quase-Féjer convergente a T. Como $T \neq \emptyset$ segue-se que $\{x^k\}$ é limitada.

Seja $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$ e $\{x^{k_s}\}$ uma subsequência de $\{x^k\}$ convergindo a $\bar{x}$. Note que sendo f contínua a sequência $\{f(x^{k_s})\}$ converge para $f(\bar{x})$. Como a sequência $\{f(x^k)\}$ é monótona decrescente e possui uma subsequência $\{f(x^{k_s})\}$ convergindo para $f(\bar{x})$ concluímos que

$$\inf\{f(x^k) \ : \ k = 0, 1, 2, \ldots\} = \lim_{k\to\infty} f(x^k) = f(\bar{x}),$$

e isto implica que $\bar{x} \in T$. Portando, pelo Teorema 1.23 concluímos que toda a sequência $\{x^k\}$ converge a $\bar{x}$. O que finaliza a demonstração do corolário. □

Teorema 2.6 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 1. Se a função f é quase convexa então uma, e somente uma, das alternativas abaixo ocorre:*

i) *A sequência $\{x^k\}$ converge para um ponto estacionário do problema* (2-1)*;*

ii) *O problema* (2-1) *não tem solução,* $\lim_{k\to\infty} \|x^k\| = \infty$*, e além disso*

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

Prova. Para a demonstração do teorema iremos analisar dois casos. O primeiro quando a sequência $\{x^k\}$ possui pontos de acumulação e o segundo quando não possui pontos de acumulação.

1^o **Caso:** Seja $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$ e $\{x^{k_s}\}$ uma subsequência de $\{x^k\}$ convergindo a $\bar{x}$. Note que sendo f contínua a sequência $\{f(x^{k_s})\}$ converge para $f(\bar{x})$. Como a sequência $\{f(x^k)\}$ é monótona decrescente e possui uma subsequência $\{f(x^{k_s})\}$ convergindo para $f(\bar{x})$ concluímos que

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

e isto implica que $\bar{x} \in T$. Portando, pela Teorema 1.23 concluímos que toda a sequência $\{x^k\}$ converge a $\bar{x}$, e pela Proposição 2.3 temos que $\bar{x}$ é um ponto estacionário para o problema.

2^o **Caso:** Agora estudaremos o caso em que a sequência $\{x^k\}$ não possua pontos de acumulação, ou seja,

$$\lim_{k\to\infty} \|x^k\| = \infty. \tag{2-11}$$

Suponhamos que o problema inicial possui um minimizador, isto é, existe um $\widehat{x} \in C$ tal que para todo k

$$f(\widehat{x}) \leq f(x^k),$$

o que implica que $T \neq \emptyset$. Utilizando o Corolário 2.4 obtemos que a sequência $\{x^k\}$ é convergente, o que contradiz a hipótese (2-11), logo o conjunto solução é vazio. Basta provar agora a última igualdade do teorema. Já sabemos que a sequência $\{f(x^k)\}$ é monótona decrescente. Então, suponhamos por absurdo que

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

Como $\{f(x^k)\}$ é monótona decrescente , a definição de ínfimo e desigualdade acima implicam que existe um $\bar{x} \in C$ tal que para todo k obtemos

$$f(\bar{x}) \leq f(x^k), \qquad \forall k.$$

Isto garante que $T \neq \emptyset$, e assim pelo Corolário 2.5 obtemos que $\{x^k\}$ é convergente, contradizendo novamente a hipótese (2-11), ou seja, a igualdade no teorema é válida. $\square$

Corolário 2.7 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 1. Se f é quase-convexa e $T \neq \emptyset$ então a sequência $\{x^k\}$ converge a um ponto estacionário.*

Prova. Se $T \neq \emptyset$, então podemos tomar $\bar{x} \in T$ implicando que para todo k temos

$$f(\bar{x}) \leq f(x^k).$$

Pelo Corolário 2.5 a sequência $\{x^k\}$ gerada pelo Algoritmo 1 converge a um ponto $\tilde{x} \in C$ e pelo Teorema 2.6 esse ponto é estacionário para o problema. $\square$

Teorema 2.2 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 1. Se f é pseudo-convexa então as seguintes alternativas são equivalentes:*

i) O conjunto solução do problema (2-1) *é não vazio;*

ii) A sequência $\{x^k\}$ possui um ponto de acumulação.

Além disso, qualquer uma dessas alternativas implica que sequência $\{x^k\}$ converge a uma solução do problema (2-1).

Prova. Pelo Teorema 1.48 f ser pseudo-convexa implica que f é quase-convexa.

Suponhamos que o conjunto solução do problema (2-1) seja não vazio. Seja $\bar{x} \in C$ uma solução do problema (2-1), em particular para todo k obtemos

$$f(\bar{x}) \leq f(x^k),$$

o que implica que $T \neq \emptyset$. Como f é quase-convexa e $T \neq \emptyset$, pelo Corolário 2.7 a sequência $\{x^k\}$ é convergente e pelo teorema anterior o ponto limite da sequência é estacionário, digamos, $\bar{x}$ e claramente esse ponto é de acumulação. Portanto, a condição (i) implica (ii). Além disso, sendo $\bar{x}$ ponto estacionário do problema (2-1), temos

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

Mas, como f é continuamente diferenciável e pseudo-convexa em C pelo Teorema 1.49 obtemos que

$$f(\bar{x}) = \min\{f(x) \ : \ x \in C\}.$$

Portanto, a sequência $\{x^k\}$ converge a uma solução do problema (2-1).

Agora, vamos mostrar que a condição (ii) implica (i). Sejam $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$ e $\{x^{k_s}\}$ uma subsequência de $\{x^k\}$ convergindo a $\bar{x}$. Note que, sendo f contínua a sequência $\{f(x^{k_s})\}$ converge para $f(\bar{x})$. Como a sequência $\{f(x^k)\}$ é monótona decrescente e possui uma subsequência $\{f(x^{k_s})\}$ convergindo para $f(\bar{x})$, concluímos que

$$\inf\{f(x^k) \ : \ k = 0, 1, 2, \ldots\} = \lim_{k \to \infty} f(x^k) = f(\bar{x}).$$

e isto implica que $\bar{x} \in T$. Pelo Corolário 2.5 e o Teorema 2.6 a sequência $\{x^k\}$ converge a um ponto estacionário do problema (2-1). Como $\bar{x}$ é ponto de acumulação de $\{x^k\}$ e $\{x^k\}$ converge, segue-se que $\{x^k\}$ converge a $\bar{x}$ e, sendo $\bar{x}$ ponto estacionário do problema (2-1) implica que para todo $x \in C$ obtemos

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

Utilizando novamente o Teorema 1.49 obtemos então que

$$f(\bar{x}) = \min\{f(x) \ : \ x \in C\}.$$

Portanto, o conjunto solução é não vazio e a sequência $\{x^k\}$ converge a uma solução do problema (2-1). $\square$

CAPÍTULO 3

Método do Gradiente para Problemas de Otimização Multi-Objetivo

Nesta segunda parte da dissertação iremos mudar nosso objeto de estudo. No capítulo anterior estudamos o Método do Gradiente Projetado, onde o problema era o de minimizar uma função real de n variáveis continuamente diferenciável, provando assim diversas propriedades de convergência da sequência gerada pelo Algoritmo 1.

O objetivo agora é estudar o Método de Descida para problemas de minimização multi-objetivo. O problema agora é o de minimizar uma aplicação de n variáveis que toma valores em $\mathbb{R}^m$. Analisaremos nesse contexto as propriedades de convergência para este caso.

3.1 O Problema Multi-Objetivo

Iremos agora estudar o seguinte problema de Otimização Irrestrito

$$\min F(x), \qquad \text{sujeito a } \ x \in \mathbb{R}^n. \tag{3-1}$$

Assumimos a partir de agora que a aplicação $F : \mathbb{R}^n \to \mathbb{R}^m$ é continuamente diferenciável.

3.2 Conceitos Preliminares

Nesta seção apresentaremos alguns conceitos básicos e definições muito importantes que irão nos auxiliar no estudo e resolução do nosso problema no decorrer deste capítulo.

Definiremos agora os subconjuntos $\mathbb{R}^m_+$ e $\mathbb{R}^m_{++}$ do espaço euclidiano $\mathbb{R}^m$ da seguinte forma

$$\mathbb{R}^m_+ := \{x \in \mathbb{R}^m | x_i \geq 0, i \in I\}, \qquad \mathbb{R}^m_{++} := \{x \in \mathbb{R}^m | x_i > 0, i \in I\},$$

onde $I = \{1, 2, ..., m\}$ é um conjunto de índices.

Agora tomando $x, y \in \mathbb{R}^m_+$, dizemos que $x \preceq y$ se, e somente se $y - x \in \mathbb{R}^m_+$. De modo semelhante dizemos que $x \prec y$ se, e somente se $y - x \in \mathbb{R}^m_{++}$.

Definição 3.1 *Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é Pareto Ótimo para o Problema 3-1 quando não existe outro $x \in \mathbb{R}^n$ satisfazendo*

$$F(x) \preceq F(\bar{x}), \qquad\qquad F(x) \neq F(\bar{x}).$$

Nosso estudo se baseará em encontrar os pontos Pareto Ótimo para o problema multi-objetivo 3-1 especificado acima. Como a aplicação F toma valores em $\mathbb{R}^m$ ela pode ser escrita como $F(x) = (f_1(x), ..., f_m(x))$, e denotamos então o Jacobiano da F como

$$J_F(x)^t = (\nabla f_1(x), ..., \nabla f_m(x)), \quad x \in \mathbb{R}^n.$$

Definimos também o conjunto imagem do Jacobiano da aplicação F em x dado pelo conjunto

$$Im(J_F(x)) = \left\{ J_F(x)v^t = (\langle \nabla f_1(x), v \rangle, ..., \langle \nabla f_m(x), v \rangle) \ : v \in \mathbb{R}^n \right\}.$$

Analisando a igualdade acima, percebemos que uma das condições para que um ponto $\bar{x} \in \mathbb{R}^n$ seja Pareto Ótimo é que ele satisfaça para $x \in \mathbb{R}^n$ a seguinte condição

$$Im\,(J_F(x)) \cap \left(-\mathbb{R}^m_{++}\right) = \emptyset.$$

Essa condição garante que não existirá nenhuma direção de descida para todas as funções que compõem a função vetorial, ou seja, eu nunca consigo decrescer simultaneamente todas as funções coordenadas por uma mesma direção v.

Observações 3.1 *A condição acima em geral não é suficiente para que um ponto seja Pareto Ótimo, essa condição é somente necessária para que o seja. Um ponto $\bar{x}$ que satisfaça a condição acima é chamado Pareto Crítico.*

Por outro lado, diremos que um ponto $\bar{x}$ não é Pareto Crítico quanto existe uma direção $v \in \mathbb{R}^n$ satisfazendo

$$J_F(\bar{x})v^t \in \left(-\mathbb{R}^m_{++}\right).$$

Pela condição acima combinada com a definição de $(-\mathbb{R}^m_{++})$ temos imediatamente a seguinte desigualdade

$$J_F(\bar{x})v^t \prec 0.$$

Dizemos que o vetor $v \in \mathbb{R}^n$ que satisfaz a condição anterior é uma direção de descida para a aplicação F em $\bar{x}$.

3.3 Método de Descida para Problemas Multi-Objetivo

Nesta seção apresentaremos alguns resultados relevantes que irão nos auxiliar na prova do teorema principal e o Método de Descida Multi Objetivo que será utilizado para resolver o Problema 3-1.

Tomando $\bar{x} \in \mathbb{R}^n$ iremos considerar agora o seguinte problema de otimização irrestrito

$$\min_{v \in \mathbb{R}^n} \left\{ \max_{i \in I} \langle \nabla f_i(\bar{x}), v \rangle + \frac{\|v\|^2}{2} \right\}. \tag{3-2}$$

Lema 3.2 *O Problema de otimização 3-2 tem somente uma solução. Além disso, o vetor v é a solução do problema 3-2 se, e somente se, existem $\alpha_i \geq 0, i \in I(\bar{x}, v)$, tal que*

$$v = -\sum_{i \in I(\bar{x},v)} \alpha_i \nabla f_i(\bar{x}), \quad e \quad \sum_{i \in I(\bar{x},v)} \alpha_i = 1,$$

onde definimos o conjunto $I(\bar{x}, v)$ da seguinte maneira

$$I(\bar{x}, v) = \{i \ \in I | \langle \nabla f_i(\bar{x}), v \rangle = \max_{i \in I} \langle \nabla f_i(\bar{x}), v \rangle \}.$$

Prova. Iremos estudar o problema de minimizar a função $\psi : \mathbb{R}^n \to \mathbb{R}$ onde ela é dada da seguinte forma

$$\psi(v) = \max_{i \in I} \langle \nabla f_i(\bar{x}), v \rangle + \frac{\|v\|^2}{2}.$$

Podemos notar que a função ψ é a soma de uma função convexa com uma função fortemente convexa, o que implica que ψ é fortemente convexa e daí pelo Teorema 1.35 o problema 3-2 possui uma única solução.

Agora como ψ é convexa pelo Teorema 1.37 obtemos que v é solução do problema 3-2 se, e somente se

$$0 \in \partial \left(\max_{i \in I} \langle \nabla f_i(\bar{x}), . \rangle + \frac{\|.\|^2}{2} \right)(v),$$

ou equivalentemente,

$$-v \in \partial (\max_{i \in I} \langle \nabla f_i(\bar{x}), . \rangle)(v).$$

Portanto a segunda condição segue da fórmula para o subdiferencial do máximo de funções lineares dada pela Proposição 1.38, concluindo assim nossa demonstração. $\square$

Lema 3.3 *Tomemos $x \in \mathbb{R}^n$. Se x não é Pareto Ótimo e v é uma solução do Problema 3-2, então*

$$\max_{i \in I} \langle \nabla f_i(x), v \rangle + \frac{\|v\|^2}{2} < 0,$$

em particular v é uma direção de descida para F em x.

Prova. Seja $x \in \mathbb{R}^n$. Sabemos por hipótese que x não é Pareto-Ótimo, então existe $\widehat{v} \in \mathbb{R}^n$ tal que

$$J_F(x)\widehat{v} \prec 0.$$

A desigualdade acima implica em particular que

$$\beta = \max_{i \in I} \langle \nabla f_i(x), \widehat{v} \rangle < 0.$$

Mas como por hipótese v é uma solução para o Problema 3-2 podemos tomar $\bar{v}$ como sendo

$$\bar{v} = \frac{-\beta}{\|\widehat{v}\|^2} \widehat{v},$$

e obtemos dessa forma a seguinte desigualdade

$$\max_{i \in I} \langle \nabla f_i(x), v \rangle + \frac{\|v\|^2}{2} \leq max_{i \in I} \langle \nabla f_i(x), \bar{v} \rangle + \frac{\|\bar{v}\|^2}{2}. \tag{3-3}$$

Combinando o segundo membro da desigualdade acima com a definição de β obtemos a desigualdade desejada do lema. □

Definição 3.4 *A função direção de descida para F é definida como*

$$v(\bar{x}) = \arg\min_{v \in \mathbb{R}^n} \left\{ \max_{i \in I} \langle \nabla f_i(\bar{x}), v \rangle + \frac{\|v\|^2}{2} \right\}.$$

Algoritmo 2 (Método de Descida para problemas Multi-Objetivo)

INICIALIZAÇÃO. *Tome $\beta \in (0,1)$ e $x^k \in \mathbb{R}^n$. E faça $k = 0$.*
CRITÉRIO DE PARADA. *Se x^k é um Pareto Crítico, então Pare. Caso contrario.*
PASSO ITERATIVO. *Computar a direção de descida para F em x^k, isto é, analisar a viabilidade de resolver um segundo problema que é o de encontrar*

$$v^k := v(x^k), \tag{3-4}$$

e em seguida o comprimento de passo $t_k = 2^{-j(k)} \in (0,1]$ onde $j(k)$ é definido da seguinte forma

$$j(k) = \min \left\{ j \in \mathbb{N} \,:\, F(x^k + 2^{-j} v^k) \preceq F(x^k) + \beta 2^{-j} J_F(x^k) v^k \right\}, \tag{3-5}$$

logo o próximo passo x^{k+1} é dado por

$$x^{k+1} = x^k + t_k v^k, \tag{3-6}$$

e VOLTE para CRITÉRIO DE PARADA.

Proposição 3.5 *A sequência gerada pelo Algoritmo 2 com a busca de Armijo está bem definida.*

Prova. Suponha que x^k não seja um ponto Pareto Crítico. Pela Definição 3.4 e pelo Lema 3.3 existe $v^k \in \mathbb{R}^n$ tal que

$$J_F\left(x^k\right)v^k \prec 0.$$

Como F é por hipótese continuamente diferenciável e $\beta \in (0,1)$, obtemos que

$$\lim_{t\to 0^+} \frac{F\left(x^k + tv^k\right) - F\left(x^k\right)}{t} \prec \beta J_F\left(x^k\right)v^k,$$

e pelo Teorema 1.12 existe uma vizinhança $(0,\delta)$ satisfazendo a seguinte desigualdade

$$F\left(x^k + tv^k\right) \prec F\left(x^k\right) + \beta t J_F\left(x^k\right)v^k, \quad t \in (0,\delta).$$

Como $\lim_{j\to\infty} 2^{-j} = 0$, basta tomar t dessa forma para concluir que a sequência gerada pelo Algoritmo 2 com busca de Armijo esta bem definido. □

3.4 Análise Parcial de Convergência

Provaremos nesta seção resultados importantes que dizem respeito a propriedades da função direção de descida para F e além disso mostraremos alguns resultados parciais de convergência da sequência gerada pelo Algoritmo 2.

Lema 3.6 *A função direção de descida para F, dada na Definição 3.4 é continua.*

Observações 3.2 *Este resultado é de fundamental importância, por que ele garante que podemos trabalhar com uma sequencia de funções de direções de descida sem que haja problema, justamente pela continuidade desta função.*

Prova. Seja a sequência $\{x^k\} \subset \mathbb{R}^n$ gerada pelo Algoritmo 2 convergindo a $\bar{x}$. Definindo $v^k := v(x^k)$ e combinando a definição 3.4 e o Lema 3.3 obtemos a seguinte desigualdade

$$\frac{\|v^k\|^2}{2} < -max_{i\in I}\langle \nabla f_i(x^k), v^k\rangle. \tag{3-7}$$

Desenvolvendo o lado direito da desigualdade acima e utilizando a Desigualdade de Cauchy-Schwarz obtemos

$$\|v^k\| < 2max_{i\in I}\|\nabla f_i(x^k)\|.$$

Como F é uma aplicação continuamente diferenciável e $\{x^k\}$ é convergente, então a desigualdade acima implica que a sequência $\{v^k\}$ é limitada.

Seja $\bar{v}$ um ponto de acumulação da sequência $\{v^k\}$. Pela Definição 3.4 e do Lema 3.2 temos que existem $\alpha_i{}^k \geq 0$ e $i \in I(x^k, v^k)$ tal que

$$v^k = -\sum_{i\in I(x^k,v^k)} \alpha_i{}^k \nabla f_i(x^k), \quad \text{e} \quad \sum_{i\in I(x^k,v^k)} \alpha_i{}^k = 1, \quad k = 1,2,..., \tag{3-8}$$

onde $I\left(x^k, v^k\right) = \left\{i \in I | max_{i\in I}\langle\nabla f_i(x^k), v^k\rangle = \langle\nabla f_i(x^k), v^k\rangle\right\}$. Usando as constantes acima e índices associados, defina a sequência α^k como sendo

$$\alpha^k := \left(\alpha_1{}^k, ..., \alpha_m{}^k\right), \quad \text{onde} \quad \alpha_i{}^k = 0 \quad se \quad i \in I \backslash I\left(x^k, v^k\right).$$

Seja $\|.\|_1$ a norma da soma em $\mathbb{R}^n$. Visto que $\sum_{i\in I(x^k,v^k)} \alpha_i{}^k = 1$ para todo k o que implica que a sequência $\left\{\alpha^k\right\}$ é limitada. Seja $\bar{\alpha}$ um ponto de acumulação da sequência $\left\{\alpha^k\right\}$ e $\left\{v^{k_s}\right\}$, $\left\{x^{k_s}\right\}$ subsequências de $\left\{v^k\right\}$ e $\left\{x^k\right\}$ respectivamente, tal que

$$\lim_{s\to\infty} v^{k_s} = \bar{v}, \qquad \lim_{s\to\infty} \alpha^{k_s} = \bar{\alpha}.$$

Como o conjunto de índices I é finito, obtemos que $I(x^{k_s}, v^{k_s}) \subset I$ para todo s. E podemos assumir sem perda de generalidade que

$$I(x^{k_1}, v^{k_1}) = I(x^{k_2}, v^{k_2}) = ... = \bar{I}. \tag{3-9}$$

Assim pela igualdade anterior e do Lema 3.2 podemos concluir que

$$v^{k_s} = -\sum_{i\in I(x^{k_s},v^{k_s})} \alpha_i{}^{k_s} \nabla f_i(x^{k_s}) \quad \text{com} \quad \sum_{i\in I(x^{k_s},v^{k_s})} \alpha_i{}^{k_s} = 1.$$

Agora fazendo s tender ao infinito na igualdade anterior obtemos que

$$\bar{v} = -\sum_{i\in\bar{I}} \bar{\alpha}_i \nabla f_i(\bar{x}), \quad \text{e} \quad \sum_{i\in\bar{I}} \bar{\alpha}_i = 1. \tag{3-10}$$

Por outro lado, combinando a definição de $I(x^{k_s}, v^{k_s})$ e a continuidade de ∇F obtemos que

$$\max_{i\in I}\langle\nabla f_i(\bar{x}), \bar{v}\rangle = \langle\nabla f_i(\bar{x}), \bar{v}\rangle, \quad i \in \bar{I}.$$

Portanto, pela definição de $I(\bar{x},\bar{v})$ temos que $\bar{I} \subset I(\bar{x},\bar{v})$. Além disso, combinando novamente a Definição 3.4, o Lema 3.2 e 3-10 concluímos que $\bar{v} = v(\bar{x})$, e o resultado desejado esta provado. □

Teorema 3.7 *As seguintes afirmações são verdadeiras*

i. *A sequência $\{F(x^k)\}$ é estritamente decrescente, ou seja, sempre podemos encontrar uma direção de descida que decresça simultaneamente todas as funções coordenadas da função vetorial F;*

ii. *Cada ponto de acumulação da sequência é um ponto Pareto Crítico.*

Prova. Provaremos o item i do teorema.

Seja $\{x^k\}$ a sequência gerada pelo Algoritmo 2. Suponha que x^k não seja Pareto Crítico, e daí obtemos pelo passo iterativo x^{k+1} no método a seguinte desigualdade

$$F\left(x^{k+1}\right) \preceq F\left(x^k\right) + \beta 2^{-j} J_F\left(x^k\right) v^k, \qquad k = 0,1,2,\ldots$$

Portanto o item (i) segue da utilização da Definição 3.4, do Lema 3.3 na desigualdade anterior. Demonstraremos agora o item (ii) do teorema.

Seja $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$ e $\{x^{k_s}\}$ uma subsequência convergindo a $\bar{x}$ e como F é contínua obtemos que

$$\lim_{s\to\infty} F(x^{k_s}) = F(\bar{x}).$$

Como pelo item (i) a sequência $\{F(x^k)\}$ é decrescente e observando que $F(\bar{x})$ é um ponto de acumulação da sequência $\{F(x^k)\}$ podemos concluir que a sequência toda $\{F(x^k)\}$ converge a $F(\bar{x})$.

Além disso, podemos notar pela definição de t_k no Algoritmo 2 que $\{t_k\} \subset (0,1]$, e isto implica que a sequência $\{t_k\}$ é limitada. Seja então $\bar{t} \in (0,1]$ um ponto de acumulação da sequência $\{t_k\}$ e uma subsequência $\{t_{k_s}\}$ convergindo a $\bar{t}$.

Por (i) e pela definição de t_k no Algoritmo 2 obtemos a seguinte desigualdade

$$F(x^{k+1}) - F(x^k) \preceq \beta t_k J_F(x^k) v^k.$$

Como F é continuamente diferenciável, $\beta \in (0,1)$, $\{x^{k_s}\}$ converge a $\bar{x}$ e $\{t_{k_s}\}$ converge a $\bar{t}$ podemos tomar limite sob a subsequência na desigualdade anterior obtendo que

$$\bar{t} J_F(\bar{x}) \bar{v} = 0.$$

Podemos notar que existem dois casos para serem analisados, quando $\bar{t} > 0$ e $\bar{t} = 0$.

(Primeiro Caso) Estudaremos primeiro o caso quando $\bar{t} > 0$, e isso implica que

$$J_F(\bar{x})\bar{v} = 0,$$

e utilizando a igualdade anterior podemos concluir em particular que

$$\max_{i\in I}\langle \nabla f_i(\bar{x}), \bar{v}\rangle = 0.$$

Por outro lado, agora utilizando o Lema 3.3 e a Definição 3.4 obtemos que

$$max_{i\in I}\langle \nabla f_i(x^{k_s}), v^{k_s}\rangle + \frac{\|v^{k_s}\|^2}{2} < 0.$$

Tendendo s ao infinito na desigualdade anterior e combinando novamente o Lema 3.3, a Definição 3.4, a continuidade de F e ∇f_i obtemos finalmente que

$$\bar{v} = v(\bar{x}) = 0,$$

e concluímos portanto que $\bar{x}$ é um ponto Pareto Crítico.

(Segundo Caso) Agora finalmente estudaremos o caso quando $\bar{t} = 0$.

Primeiramente como x^{k_s} não é Pareto Crítico utilizando a Definição 3.4, o Lema 3.3 e fazendo s tender ao infinito nós obtemos a seguinte desigualdade

$$\max_{i\in I}\langle \nabla f_i(\bar{x}), \bar{v}\rangle \leq 0.$$

Como $\lim_{s\to\infty} t_{k_s} = 0$ temos que para qualquer $q \in \mathbb{N}$ fixo irá existir s_0 tal que $j(k_s) > q$ para todo $k_s > k_{s_0}$. Assim pela definição de t_{k_s} no Algoritmo 2 existe algum $i_0 \in I$ obtemos que

$$f_{i_0}(x^{k_s} + 2^{-q}v^{k_s}) > f_{j_0}(x^{k_s}) + \beta 2^{-q}\langle \nabla f_i(x^{k_s}, v^{k_s})\rangle.$$

Utilizando agora a continuidade da função F e de ∇f_i e fazendo s tender ao infinito na desigualdade acima obtemos

$$\frac{f_{i_0}(\bar{x} + 2^{-q}v(\bar{x})) - f_{i_0}(\bar{x})}{2^{-q}} \geq \beta\langle \nabla f_{i_0}(\bar{x}), v(\bar{x})\rangle.$$

Utilizando o fato de F ser continuamente diferenciável, e que $\beta > 0$, fazendo q ir ao infinito na desigualdade anterior obtemos que

$$\langle \nabla f_{i_0}(\bar{x}), v(\bar{x})\rangle \geq 0,$$

e combinando a desigualdade anterior com a primeira desigualdade chegamos a

$$\max_{i \in I} \langle \nabla f_i(\bar{x}), \bar{v} \rangle = 0.$$

Concluímos assim de modo análogo ao caso anterior que $\bar{x}$ é um ponto Pareto Crítico. □

3.5 Análise Total da Convergência

Nesta seção faremos o estudo das propriedades de convergência da sequência gerada pelo Algoritmo 2 acrescentando a hipótese de quase-convexidade sobre função vetorial F.

Definição 3.8 *Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ dada por $F(x) = (f_1(x), ..., f_m(x))$. Temos que F é dita convexa em $\mathbb{R}^n$, se cada função coordenada f_i é convexa em $\mathbb{R}^n$ para todo $i \in I$.*

Definição 3.9 *Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ dada por $F(x) = (f_1(x), ..., f_m(x))$. Dizemos que F é quase-convexa em $\mathbb{R}^n$ se cada função coordenada f_i é quase-convexa em $\mathbb{R}^n$ para todo $i \in I$.*

Observação 3.3 *No caso em que F é convexa diferenciável implica que para todo $x, y \in \mathbb{R}^n$ obtemos que*

$$J_F(x)(y - x) \preceq F(y) - F(y).$$

Proposição 3.4 *Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função vetorial quase convexa diferenciável. Então para todo $x, y \in \mathbb{R}^n$ tal que $F(x) \preceq F(y)$ obtemos que*

$$J_F(y)(x - y) \preceq 0.$$

Prova. Combinando o Lema 1.41 juntamente com a definição de quase convexidade dada anteriormente obtemos o resultado desejado. □

Definição 3.10 *Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é um ponto Pareto Ótimo fraco se não existe $x \in \mathbb{R}^n$ satisfazendo*

$$F(x) \prec F(\bar{x}).$$

Isto significa que não existirá $x \in \mathbb{R}^n$ tal que $f_i(x) < f_i(\bar{x})$ para todo $i \in I$.

Proposição 3.11 *Seja $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função convexa continuamente diferenciável. Então, $x \in \mathbb{R}^n$ é um Pareto Crítico de F, isto é,*

$$Im(J_F(x)) \bigcap (-\mathbb{R}^m_+ +) = \emptyset,$$

se, e somente se é um ponto Pareto Ótimo Fraco para F.

Prova. ($\Rightarrow$) Seja $\bar{x}$ um ponto Pareto Crítico para o problema 3-1. Suponha por absurdo que $\bar{x}$ não seja Pareto Fraco para a função F. Como $\bar{x}$ não é Pareto Fraco então existe $x \in \mathbb{R}^n$ tal que

$$F(x) \preceq F(\bar{x}).$$

Como F é continuamente diferenciável e convexa, utilizando o fato de $\bar{x}$ não ser Pareto Fraco concluímos que

$$J_F(x)(\bar{x} - x) \preceq 0,$$

o que contradiz o fato de $\bar{x}$ ser Pareto Crítico para F, e a primeira parte esta concluída.

($\Leftarrow$) Seja $\bar{x}$ um ponto Pareto Ótimo Fraco para a função F. Suponha por absurdo que $\bar{x}$ não seja Pareto Crítico para F. Então obtemos que

$$Im(J_F(x)) \bigcap (-\mathbb{R}^m_{++}) \neq \emptyset,$$

ou seja, existe uma direção $v \in \mathbb{R}^n$ de descida para F em $\bar{x}$. Pela hipótese de diferenciabilidade de F em $\bar{x}$ temos que

$$\lim_{t \to 0^+} \frac{F(\bar{x} + tv) - F(\bar{x})}{t} = J_F(\bar{x})v \prec 0,$$

o que implica pelo Teorema 1.12 que existe uma vizinhança $(0, \delta)$ e $\delta > 0$ tal que

$$F(\bar{x} + tv) \prec F(\bar{x}) + tJ_F(\bar{x})v, \quad t \in (0, \delta).$$

Como v é uma direção de descida para F em $\bar{x}$ e $t \in (0, \delta)$, contradizendo o fato de $\bar{x}$ ser um ponto Pareto Ótimo Fraco para F, o que conclui a demonstração. $\square$

Defina agora o conjunto U da seguinte forma

$$U = \{x \in \mathbb{R}^n | F(x) \preceq F(x^k), \quad \forall k\}. \tag{3-11}$$

Observação 3.5 *O conjunto U nos auxiliará muito em nossa análise e é importante ressaltar que ele pode ser vazio. Por isso é necessário acrescentar alguma hipótese sobre a sequência $\{x^k\}$ para que ele seja não vazio. Para garantir que U seja não vazio basta supor que a sequência $\{x^k\}$ possui ao menos um ponto de acumulação $\bar{x}$ ou supor que o problema 3-1 possua ao menos uma solução.*

Lema 3.12 *Sejam F continuamente diferenciável e quase convexa e a sequencia $\{x^k\} \subset \mathbb{R}^n$ gerada pelo Algoritmo 2. Suponha que $U \neq \emptyset$. Então para todo $\bar{x} \in U$ obtemos que*

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 + t_k^2\|v^k\|.$$

Prova. Utilizando a Lei dos cossenos e os Passos Iterativos definidos no Algoritmo 2 obtemos a seguinte desigualdade

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

Como F é diferenciável e quase convexa, combinando o Lema 3.2 e a Definição 3.4 obtemos a seguinte desigualdade

$$\langle \nabla f_i(x^k), \bar{x} - x^k\rangle \leq 0.$$

Utilizando agora a desigualdade anterior na primeira igualdade chegamos facilmente na igualdade desejada do Lema.

□

Proposição 3.13 *Sejam F quase-convexa e U definido em 3-11 não vazio. Então a sequência $\{x^k\}$ gerada pelo Algoritmo 2 é quase Féjer convergente a U.*

Prova. Definiremos primeiramente a função auxiliar $\varphi : \mathbb{R}^m \to \mathbb{R}$ dada por

$$\varphi(y) = \max_{i \in I}\langle y, e_i\rangle \quad \text{onde} \quad I = \{1, \ldots m\},$$

e cada e_i pertence à base canônica do espaço euclidiano $\mathbb{R}^m$.

Podemos observar que a definição da φ implica que

i. $\varphi(x+y) \leq \varphi(x) + \varphi(y)$.

ii. $\varphi(tx) = t\varphi(x)$ onde $x, y \in \mathbb{R}^m$ e $t \in \mathbb{R}$

Pela definição de t_k e x^k no Algoritmo 2 obtemos que

$$F(x^{k+1}) \preceq F(x^k) + \beta t_k J_F(x^k)v^k, \quad k = 1, 2, \ldots$$

Utilizando as propriedades da função φ obtemos que

$$\varphi(F(x^{k+1})) \leq \varphi(F(x^k)) + \beta t_k \varphi(J_F(x^k)v^k).$$

Agora combinando a Definição 3.4 e o Lema 3.3 com a desigualdade anterior obtemos facilmente que

$$t_k{}^2\|v^k\|^2 < 2.\frac{[\varphi(F(x^k)) - \varphi(F(x^{k+1}))]}{\beta},$$

pois pelo Algoritmo 2 $t_k \in (0,1)$. Assim aplicando a soma telescópica na desigualdade anterior e utilizando novamente a definição de φ obtemos que

$$\sum_{k=0}^{n} t_k{}^2\|v^k\|^2 < \frac{2}{\beta}[\varphi(F(x^0)) - \varphi(F(\bar{x})],$$

pois $\bar{x} \in U$. Portanto, concluímos que quando n tende ao infinito a série anterior converge, provando assim que a sequência $\{x^k\}$ é quase Féjer convergente. $\square$

Teorema 3.14 *Sejam F quase-convexa e U não vazio. Então a sequência $\{x^k\}$ converge a um ponto Pareto Crítico para o problema 3-1.*

Prova. Pela Proposição 3.13 a sequência $\{x^k\}$ gerada pelo Algoritmo 2 é quase-Féjet convergente a U. E pelo item (i) do Teorema 1.23 sabemos que a sequência é limitada, o que nos garante a existência de pelo menos um ponto de acumulação.

Sejam $\bar{x}$ um ponto de acumulação da sequência $\{x^k\}$ e $\{x^{k_s}\}$ uma subsequência convergindo a $\bar{x}$. Como a função F é continua obtemos que

$$\lim_{s\to\infty} F(x^{k_s}) = F(\bar{x}).$$

Como pelo item i a sequência é decrescente e observando que $F(\bar{x})$ é um ponto de acumulação da sequência $\{F(x^k)\}$ podemos concluir que a sequência toda $\{F(x^k)\}$ converge a $F(\bar{x})$. Isto implica que para todo k tem-se

$$F(\bar{x}) \preceq F(x^k),$$

ou seja, $\bar{x} \in U$ e então pelo item (ii) do Teorema 1.23 toda sequência converge a $\bar{x}$ e este é Pareto Crítico para F. $\square$

Corolário 3.15 *Se F é convexa e U definido como em 3-11, então a sequência $\{x^k\}$ converge para um ponto Pareto Ótimo Fraco.*

Prova. Como por hipótese F é convexa temos que ela é quase-convexa. Portanto pelo Teorema 3.14 a sequência $\{x^k\}$ converge a um ponto crítico e utilizando a Proposição 3.13 concluímos que a sequência $\{x^k\}$ converge a um ponto Pareto Ótimo Fraco. $\square$

CAPÍTULO 4

Considerações Finais

Neste trabalho mostramos de maneira detalhada o estudo de dois problemas de otimização quase-convexa, onde em ambos utilizamos a Busca inexata de Armijo. Utilizamos dois métodos para resolver os problemas: o Método do Gradiente Projetado e o Método de Descida para Problemas Multi-Objetivo.

No Capítulo 3, estudamos o primeiro, isto é, como resolver o problema de otimização quase convexa cuja função objetivo f é real de n variáveis e usamos o Método do Gradiente Projetado para resolver o problema.

Era sabido por [4] que no caso em que f é convexa obtemos que a sequência converge a alguma solução do problema. Provamos no Capitulo 3 que assumindo a quase-convexidade da função f ou a sequência converge a um ponto estacionário ou diverge e, além disso se assumirmos que o conjunto T seja não vazio, então a sequência gerada pelo Método irá convergir a um ponto estacionário e se adicionarmos a hipótese de pseudo-convexidade sobre f a sequência converge para uma solução do nosso problema.

No Capitulo anterior estudamos um problema de otimização multi-objetivo também quase-convexo porém a função objetivo agora é vetorial e o Método utilizado foi o de Descida Multi-objetivo. O estudo se baseia na procura dos pontos Pareto Ótimo para o problema multi-objetivo.

Demonstramos nesta parte que se existe um ponto de acumulação da sequência gerada pelo Método, este é um ponto Pareto Crítico. Além disso, sob hipóteses de quase-convexidade sobre a função F e sobre U definido no capítulo a sequência converge a um ponto Pareto Crítico que necessariamente é também ponto Pareto Fraco.

Referências Bibliográficas

[1] BENTO, G. C; FERREIRA, O. P; OLIVEIRA, P. R. **Unconstrained steepest descent method for multicriteria optimization on Riemmanian manifolds**.

[2] BERTSEKAS, D. P. **Nonlinear Programming**. Athena Scientific, Belmont, 2ª edition, 2003.

[3] DENNIS, J. E. JR., S. R. B. **Numerical Methods for Unconstrained Optimization and Nonlinear Equations**. SIAM Classics in Applied Mathematics, 1996.

[4] IUSEM, A. N. **On the convergence properties of the projected gradient method for convex optimization**. Computational and Applied Mathematics, no. 1, 37–52, 2003.

[5] IZMAILOV, A; SOLODOV, M. **Otimização volume 1. Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade**. IMPA, Rio de Janeiro, 2ª edition, 2005.

[6] IZMAILOV, A; SOLODOV, M. **Otimização volume 2. Métodos Computacionais**. IMPA, Rio de Janeiro, 2ª edition, 2007.

[7] J. Y. BELLO CRUZ, L. R. L. P. **Convergence of a projected gradient method variant for quasiconvex objectives**. Nonlinear Analysis, no. 9, 2917–2922, 2010.

[8] LIMA, E. L. **curso de análise Vol. 1**. IMPA, Rio de Janeiro, 12ª edition, 2007.

[9] LIMA, E. L. **Curso de Análise Vol. 2**. IMPA, Rio de Janeiro, 2009.

[10] MANGASARIAN, O. L. **Nonlinear Programming**. McGraw-Hill, New York, 1994.

[11] URRUTY H.; JEAN, B; LEMARÉCHAL, C. **Convex Analysis and Minimization Algorithms I**. Springer, Berlin, 1993.

[12] URRUTY H.; JEAN, B; LEMARÉCHAL, C. **Convex Analysis and Minimization Algorithms I**. Springer, Berlin, 1993.