

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

*UM ALGORITMO MATEMÁTICO PARA PROGRAMAÇÃO
VETORIAL*

FÁBIO JÚNIOR PIMENTEL DA SILVA

MANAUS - 2015

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

FÁBIO JÚNIOR PIMENTEL DA SILVA

UM ALGORITMO MATEMÁTICO PARA PROGRAMAÇÃO
VETORIAL

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

Orientador: Prof. Dr. Sandro Dimy Barbosa Bitar

MANAUS - 2015

S586a Silva, Fábio Junior Pimentel da
Um algoritmo matemático para programação vetorial/ Fábio Junior
Pimentel da Silva. - - 2015.
79 f.; ilust.; 21 cm x 29,7 cm

Dissertação (Mestrado em Matemática-Programa de Pós-
graduação). – Universidade Federal do Amazonas, 2015.

Orientador: Prof. Dr. Sandro Dimy Barbosa Bitar

1. Algoritmo. 2. Otimização vetorial 3. Método de descida 4. K-
crítico I. Título. II. Bitar, Sandro Dimy Barbosa

CDU 517.98 (043.3)

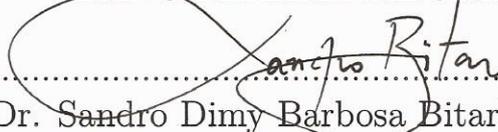
FÁBIO JÚNIOR PIMENTEL DA SILVA

UM ALGORITMO MATEMÁTICO PARA PROGRAMAÇÃO
VETORIAL

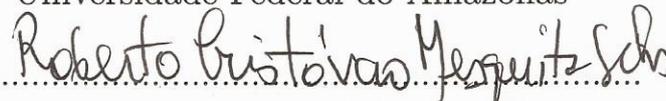
Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Matemática, área de concentração em Otimização.

Manaus, 27 de Abril de 2015.

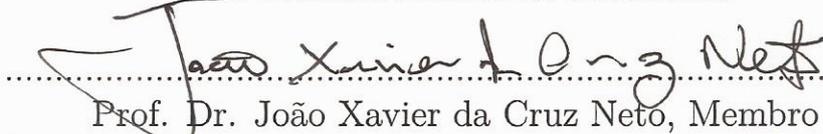
BANCA EXAMINADORA



.....
Prof. Dr. Sandro Dimy Barbosa Bitar, Presidente
Universidade Federal do Amazonas



.....
Prof. Dr. Roberto Cristóvão Mesquita Silva, Membro
Universidade Federal do Amazonas



.....
Prof. Dr. João Xavier da Cruz Neto, Membro
Universidade Federal do Piauí.

AGRADECIMENTOS

À Universidade Federal do Amazonas pela oportunidade de realizar este curso de mestrado.

Ao meu orientador professor Sandro Dimy Barbosa Bitar pela paciência, confiança e pelos conselhos que muito acrescentaram em minha vida profissional. Meu muito obrigado.

A minha verdadeira educadora que sempre acreditou e torceu por mim Ana Maria Eugênciã Pimentel da Silva (minha mãe).

Aos meus irmãos, Lenira, Hélio, Ray, Rosana, Jéssica, Tainá e em particular a Lenice Pimentel que me acolheu em sua humilde casa e vivenciou toda minha luta.

Ao meu sobrinho Pedro Luiz (Pedrinho) pela contribuição nas figuras.

A todos meus colegas de mestrado pela força e pela amizade durante esses dois anos de estudos.

Ao amigo, professor Mário Salvatierra Júnior pelo aval na implementação do AMKdG.

Ao meu grande amigo, professor Marcos Marreirro Salvatierra, pela ajuda na implementação do AMKdG e pelas sugestões que muito contribuíram para aprimoramento deste trabalho. A vitória é nossa mermão!

A todos meus amigos do ISB-Coari que torceram por mim.

RESUMO

UM ALGORITMO MATEMÁTICO PARA PROGRAMAÇÃO VETORIAL

Neste trabalho, apresenta-se um algoritmo que utiliza o método de descida para resolver um problema de otimização vetorial ou multiobjetivo irrestrito, onde as funções consideradas são continuamente diferenciáveis. Apresenta-se um estudo sobre os fundamentos teóricos, a saber: elementos da análise convexa, ordem parcial induzida por um cone K convexo, fechado, pontiagudo e com o interior não vazio bem como alguns fundamentos para programação multiobjetivo e vetorial, necessários para formulação do modelo matemático. Para o cálculo da direção de descida, utiliza-se uma função auxiliar fortemente convexa e, para o tamanho do passo, um procedimento tipo Armijo. Demonstra-se que todo ponto de acumulação da sequência gerada por esse algoritmo é K -crítico.

Palavras-chave Otimização vetorial, método de descida, cone, k -crítico.

ABSTRACT

ALGORITHM FOR MATHEMATICAL PROGRAMMING VECTOR

This paper presents an algorithm that uses the descent method to solve a vector optimization problem unconstrained multiobjective where the functions considered are continuously differentiable. It will also be a study on the theoretical foundations, namely: elements of convex analysis, induced partial order by a generic cone K , as well as multi-objective and vectorial programming fundamentals, required for formulation of the mathematical model. To calculate the direction of descent, an auxiliary function strongly convex and is used for the step size, the Armijo rule type. It is shown that the whole point of accumulation of the generated sequence the algorithm is K -critical for the vector.

Keywords Optimization vector, descent method, cone, k -critical.

Sumário

Introdução	3
1 Preliminares	5
1.1 Definições e resultados de otimização com único objetivo	5
1.2 Elementos de análise convexa	7
1.2.1 Funções convexas	10
1.2.2 Funções convexas não diferenciáveis	13
1.3 Fundamentos de programação multiobjetivo	18
1.4 Relação de ordem parcial	19
1.5 Fundamentos de programação vetorial	24
2 Um Algoritmo para Programação Vetorial	29
2.1 Algoritmo para o método K-direção de descida genérico (AMKdG) . .	29
2.1.1 Comentários sobre a convergência do algoritmo AMKdG	31
2.2 Cálculo da direção de descida	32
2.3 Método K-descida	36
2.4 Adaptações do Algoritmo(AMKdG)	41
2.5 Análise de Convergência:Caso Geral	45
3 Aplicação do Algoritmo (AMKdG) ao caso $K = \mathbb{R}_+^m$	49
3.1 Cálculo da direção de busca	51
3.1.1 Outra possibilidade para a direção de busca	55
3.2 Cálculo do Comprimento do Passo	58
3.3 O algoritmo	59
3.3.1 Comentários sobre o algoritmo	60
3.4 Convergência do Algoritmo	60
3.5 Implementação Computacional	63
Conclusão	67
3.6 Sugestões para trabalhos futuros	67
Referências Bibliográficas.	69

Lista de Figuras

1.1	Um exemplo de cone	8
1.2	Um exemplo de cone não convexo	8
1.3	Exemplo de cone normal de um conjunto convexo.	10
1.4	Relações \leq e $<$ no \mathbb{R}^n . Fonte <i>Sampaio</i> (2011).	21
1.5	Dominância de Pareto no espaço de objetivos .Fonte <i>Arroyo</i> (2002).	22
1.6	Exemplo de um cone polar positivo.	24
3.1	Comportamento do AMKdG quando $\beta = 0.9$	63
3.2	Comportamento do AMKdG quando $\beta = 0.6$	63
3.3	Comportamento do AMKdG quando $\beta = 0.5$	64
3.4	Conjunto das soluções eficientes de $f(x) = (x^2, (x - 1)^2)$	64
3.5	Aproximação da curva de Pareto de $f(x) = (x^2, (x - 1)^2)$ gerada pelo AMKdG	64
3.6	Conjunto das soluções eficientes de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2)$	65
3.7	Aproximação da curva de Pareto de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2)$ gerada pelo AMKdG	65
3.8	Conjunto das soluções eficientes de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2), (x - 1)^2 + (y - 3)^2)$	66
3.9	Aproximação da curva de Pareto de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2), (x - 1)^2 + (y - 3)^2)$ gerada pelo AMKdG	66

Introdução

Em otimização multiobjetivo ou vetorial, procuramos resolver o problema de minimização de várias funções objetivos, considerando que é muito raro a existência de uma solução que minimize todas as funções objetivos ao mesmo tempo, pois em geral esses objetivos são conflitantes. Dessa forma, resolver um problema com mais de um objetivo é buscar dentro de um conjunto de decisões admissíveis (variáveis de decisão), soluções que reúnam melhores características de um certo equilíbrio entre os objetivos, ou seja, temos que entender o conceito de solução eficiente (também conhecidas como não dominadas, ótimo Pareto) introduzido por Pareto (Oliveira, 2011).

Uma das características da otimização vetorial é a perda total da ordenação no espaço de objetivos, onde se encontra a fronteira de soluções ou fronteira de Pareto, que contém as soluções não dominadas ou ótimo Pareto. Em geral, não é possível comparar todas as soluções viáveis porque a comparação com base em uma função objetivo pode contradizer a comparação com base em outra função objetivo e isso a diferencia da otimização com um único objetivo (Cohon, 1978). Esta teoria se aplica em diversas áreas do conhecimento, tais como: economia, engenharia, design, estatísticas, ciências sociais, ciências naturais e várias outras. Devido a diversidade de aplicações e, por se tratar de um tema de extrema relevância, apresentaremos um método de máxima descida para otimização multiobjetivo.

Conceitos importantes como ponto K-crítico, Pareto-ótimo, são fundamentais para estabelecer condições necessárias de primeira ordem. Desta forma, quando garantirmos que o algoritmo que apresentamos gera uma sequência infinita sob a hipótese de existência de ponto de acumulação dizemos que é K-crítico. Portanto, faremos uma análise desse algoritmo, que foi inicialmente proposto em (Fliege e Svaiter, 2000) e generalizado por (Graña e Svaiter, 2005), fazendo, ao mesmo tempo, um estudo de caso juntamente com implementações computacionais.

Objetivos

Objetivo geral

Apresentar a partir de um método de descida com busca direcional linear, um algoritmo para otimização vetorial(multiobjetivo).

Objetivos específicos

Esta dissertação objetiva conceituar elementos da otimização multiobjetivo (vetorial) e da análise convexa, assim como ordem parcial induzida por um cone genérico,

fundamental para o modelo matemático investigado. Pretende-se investigar propriedades de convergência da sequência gerada pelo algoritmo estudado.

Além disso, aplicar o caso generalizado para otimização multiobjetivo onde o cone em estudo é o ortante não negativo.

Descrição da dissertação

Esta dissertação está organizada em três capítulos. O capítulo 1 trata de alguns elementos da análise convexa, fundamentos da otimização multiobjetivo (vetorial), bem como a formulação do problema multiobjetivo (vetorial), definimos o conceito de K-crítico e apresentamos um procedimento tipo Armijo (adaptado para otimização vetorial). No capítulo 2, apresenta-se o principal resultado deste trabalho, um método de descida para programação vetorial irrestrita, demonstrando, finalmente que todo ponto de acumulação da sequência gerada pelo algoritmo (AMKdG) é K-crítico. No capítulo 3, aplica-se o caso geral para otimização multiobjetivo e faz-se algumas implementações computacionais do Algoritmo (AMKdG). Por fim, algumas considerações relevantes ao método são feitas com perspectivas para trabalhos futuros.

Capítulo 1

Preliminares

A finalidade deste capítulo é facilitar a compreensão da leitura desta dissertação fornecendo as definições e propriedades que são necessárias para o desenvolvimento dos demais capítulos. Informações mais detalhadas podem ser encontradas nas referências bibliográficas.

1.1 Definições e resultados de otimização com único objetivo

Nesta seção estudaremos alguns resultados básicos sobre os problemas de otimização, que serão necessários ao longo desta dissertação. Serão apresentados os resultados de existência de soluções e as condições necessárias e suficientes de otimalidade para problemas irrestritos. Sejam $D \subset \mathbb{R}^n$ e $f : D \rightarrow \mathbb{R}$. Estamos interessados em encontrar um minimizador de f no conjunto D . Este problema será formulado como

$$\begin{cases} \min f(x) \\ \text{s.a } x \in D \end{cases} \quad (1.1)$$

Considerando $D = \mathbb{R}^n$, o problema (1.1) passa a ser chamado de *problema de minimização irrestrito* e escrevemos como

$$\begin{cases} \min f(x) \\ \text{s.a } x \in \mathbb{R}^n \end{cases} \quad (1.2)$$

Em (1.1) o conjunto D é o *conjunto viável* do problema, os pontos de D serão chamados de *viáveis*, e f é denominada *função objetivo*.

Definição 1. Dizemos que um ponto $\bar{x} \in D$ é

1. a) *minimizador global* de (1.2), se

$$f(\bar{x}) \leq f(x) \quad \forall x \in D \quad (1.3)$$

2. b) *minimizador local* de (1.2), se existir $\epsilon > 0$ tal que,

$$f(\bar{x}) \leq f(x) \quad \forall x \in B(\bar{x}, \epsilon) \quad (1.4)$$

Se para todo $x \neq \bar{x}$ a desigualdade (1.2) ou (1.3) é estrita, então \bar{x} será chamado de minimizador estrito global ou local, respectivamente.

Definição 2. Dizemos que $\bar{v} \in [-\infty, +\infty)$ definido por

$$\bar{v} = \inf_{x \in D} f(x) \quad (1.5)$$

é o valor ótimo do problema (1.1).

Uma função pode admitir vários minimizadores globais e locais, mas o valor ótimo do problema é sempre o mesmo.

A seguir apresentaremos condições que garantem a existência de solução global.

Teorema 1.1. (Teorema de Weierstrass) *Sejam $D \subset \mathbb{R}^n$ um conjunto compacto não-vazio e $f : D \rightarrow \mathbb{R}$ uma função contínua. Então o problema (1.2) tem solução global.*

Demonstração. Ver [22].

□

1.2 Elementos de análise convexa

Definição 3. Um conjunto $D \subset \mathbb{R}^n$ é chamado conjunto convexo, se dados quaisquer $x \in D$, $y \in D$ e $\alpha \in [0, 1]$, tem-se que $(1 - \alpha)x + \alpha y \in D$.

Exemplo 1. Seja $D = \{x \in \mathbb{R}^n \mid \langle a, x \rangle \leq c\}$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Afirmamos que D é um subconjunto convexo do \mathbb{R}^n .

Com efeito, dados $x \in D$, $y \in D$, temos $\langle a, x \rangle \leq c$ e $\langle a, y \rangle \leq c$. Para qualquer $\alpha \in [0, 1]$ vale $c = (1 - \alpha)c + \alpha c$, logo

$$\langle a, (1 - \alpha)x + \alpha y \rangle = (1 - \alpha)\langle a, x \rangle + \alpha\langle a, y \rangle \leq (1 - \alpha)c + \alpha c = c \quad (1.6)$$

Exemplo 2. Seja $D = \{(x, y) \in \mathbb{R}^2 \mid y \leq x^2\}$. Afirmamos que o conjunto $D \subset \mathbb{R}^2$ não é convexo. De fato,

basta tomarmos $x = (-1, 1)$, $y = (1, 1)$ e $\alpha = \frac{1}{2}$, pois $x \in D$, $y \in D$ mas $(1 - \frac{1}{2})x + \frac{1}{2}y = (0, 1)$ não pertence a D .

O conjunto vazio, o espaço \mathbb{R}^n , e um conjunto que contém um ponto só, são trivialmente convexos.

Definição 4. (Combinação Convexa) Sejam $x_1, x_2, \dots, x_n \in \mathbb{R}^n$ e $\alpha_i \in [0, 1]$, $i = 1, 2, \dots, p$. Chama-se combinação convexa de pontos $x^i \in \mathbb{R}^n$ com parâmetros $\alpha_i \in [0, 1]$, $i = 1, 2, \dots, p$, ao ponto

$$\sum_{i=1}^p \alpha_i x^i \text{ com } \sum_{i=1}^p \alpha_i = 1 \quad (1.7)$$

Definição 5. Um conjunto $K \subset \mathbb{R}^n$ é um **cone**, se para todo $d \in K$, $t \geq 0$, $td \in K$

Exemplo 3. O conjunto \mathbb{R}^n , e qualquer subespaço é cone.

Definição 6. Dizemos que um cone K é **pontiagudo** se $K \cap -K = \{0\}$

Definição 7. Seja $K \subset \mathbb{R}^n$ um cone. K é um **cone convexo**, se K é convexo.

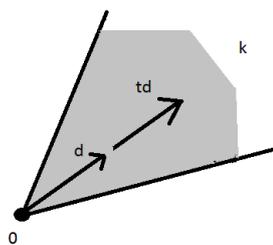


Figura 1.1: Um exemplo de cone

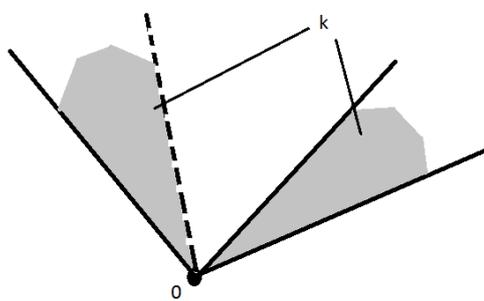


Figura 1.2: Um exemplo de cone não convexo

Exemplo 4. Em \mathbb{R}^n o ortante não-negativo

$$\mathbb{R}_+^n = \{x = (x_1, x_2, \dots, x_n) / x_i \geq 0, 1 \leq i \leq n\} \quad (1.8)$$

é exemplo de cone convexo, fechado e pontiagudo.

Proposição 1.2. *Seja $K \subset \mathbb{R}^n$ um cone. Então, K é convexo se, e somente se, $K = K + K$.*

Demonstração. Ver [15]. □

Definição 8. (Fecho convexo) *Seja $D \subset \mathbb{R}^n$ um conjunto qualquer. O fecho convexo de D , denotado $\text{conv}D$, é o menor conjunto convexo em \mathbb{R}^n que contém D (ou, equivalente, a interseção de todos os conjuntos convexos em \mathbb{R}^n que contém D).*

Exemplo 5. O Fecho convexo da esfera é a bola, isto é, para

$$D = \{x \in \mathbb{R}^n \mid \|x\| = c\}, \quad (1.9)$$

onde $c \in \mathbb{R}$, tem-se

$$\text{conv}D = \{x \in \mathbb{R}^n \mid \|x\| \leq c\}. \quad (1.10)$$

Definição 9. (Fecho cônico) Seja $D \subset \mathbb{R}^n$ um conjunto qualquer. O fecho cônico de D , denotado $\text{cone}D$, é o menor cone convexo em \mathbb{R}^n que contém D (ou, equivalente, a interseção de todos os cones convexos em \mathbb{R}^n que contém D).

Para um conjunto convexo, o fecho cônico é composto por todos os múltiplos não negativos de elementos do conjunto.

Proposição 1.3. (Fecho cônico de um conjunto convexo) Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Tem-se que

$$\text{cone}D = \{d \in \mathbb{R}^n \mid d = \alpha x, x \in D, \alpha \in \mathbb{R}_+\} \quad (1.11)$$

Definição 10. (Cone normal) Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in D$. O *cone normal* (cone das direções normais) no ponto \bar{x} em relação ao conjunto D é dado por

$$\mathcal{N}_{D(\bar{x})} = \{d \in \mathbb{R}^n \mid \langle d, x - \bar{x} \rangle \leq 0 \forall x \in D\}. \quad (1.12)$$

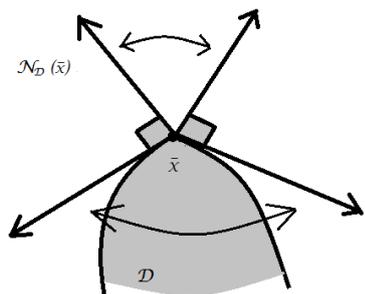


Figura 1.3: Exemplo de cone normal de um conjunto convexo.

1.2.1 Funções convexas

Definição 11. Seja $D \subset \mathbb{R}^n$ um conjunto convexo, diz-se que a função $f : D \rightarrow \mathbb{R}$ é convexa em D quando para quaisquer $x \in D$, $y \in D$ e $\alpha \in [0, 1]$, tem-se que

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y). \quad (1.13)$$

A função f diz-se estritamente convexa quando

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y) \quad \forall x \neq y \text{ e } \alpha \in (0, 1) \quad (1.14)$$

A função f diz-se fortemente convexa com módulo $\gamma > 0$, quando para quaisquer $x \in D$, $y \in D$ e $\alpha \in [0, 1]$, tem-se

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) - \gamma\alpha(1 - \alpha)\|x - y\|^2. \quad (1.15)$$

Exemplo 6. A função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f(x) = \|x\|^2$ é fortemente convexa com módulo $\gamma = 1$.

De fato, dados $x \in D$, $y \in D$ e $\alpha \in [0, 1]$

$$\begin{aligned}
 f(\alpha x + (1 - \alpha)y) &= \langle \alpha x + (1 - \alpha)y, \alpha x + (1 - \alpha)y \rangle \\
 &= \alpha^2 \|x\|^2 + 2\alpha(1 - \alpha)\langle x, y \rangle + (1 - \alpha)^2 \|y\|^2 \\
 &= \alpha^2 \|x\|^2 + \alpha(1 - \alpha)(\|x\|^2 + \|y\|^2 - \|x - y\|^2) + (1 - \alpha)^2 \|y\|^2 \\
 &= \alpha \|x\|^2 + (1 - \alpha)\|y\|^2 - \alpha(1 - \alpha)\|x - y\|^2 \\
 &= \alpha f(x) + (1 - \alpha)f(y) - \alpha(1 - \alpha)\|x - y\|^2
 \end{aligned}$$

Definição 12. (Epígrafo) Sejam $D \subset \mathbb{R}^n$ e $f : D \rightarrow \mathbb{R}$. O epígrafo de f é o conjunto

$$E_f = \{(x, c) \in D \times \mathbb{R} \mid f(x) \leq c\}. \quad (1.16)$$

Teorema 1.4. *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \rightarrow \mathbb{R}$ é convexa em D se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Demonstração. Suponhamos primeiro que E_f seja convexo. Sejam $x \in D$, $y \in D$ quaisquer. Obviamente, $(x, f(x)) \in E_f$ e $(y, f(y)) \in E_f$. Pela convexidade de E_f , para todo $\alpha \in [0, 1]$ temos que

$$\begin{aligned}
 (\alpha x + (1 - \alpha)y, \alpha f(x) + (1 - \alpha)f(y)) &= (\alpha x, \alpha f(x)) + ((1 - \alpha)y + (1 - \alpha)f(y)) \\
 &= \alpha(x, f(x)) + (1 - \alpha)(y, f(y)) \in E_f.
 \end{aligned}$$

Pela definição do epígrafo, segue que

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

logo, é convexa. Suponhamos agora que f seja convexa. Dados $(x, c_1) \in E_f$ e $(y, c_2) \in E_f$ temos que $f(x) \leq c_1$ e $f(y) \leq c_2$, para todo $\alpha \in [0, 1]$ vale

$$\alpha f(x) \leq \alpha c_1 \text{ e } (1 - \alpha)f(y) \leq (1 - \alpha)c_2 \quad (1.17)$$

e pela convexidade de f , tem-se

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &\leq \alpha f(x) + (1 - \alpha)f(y) \\ &\leq \alpha c_1 + (1 - \alpha)c_2 \end{aligned} \tag{1.18}$$

O que significa que

$$\alpha(x, c_1) + (1 - \alpha)(y, c_2) = (\alpha x + (1 - \alpha)y, \alpha c_1 + (1 - \alpha)c_2) \in E_f,$$

isto é, E_f é convexo.

□

Teorema 1.5. (Teorema de minimização convexa) *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \rightarrow \mathbb{R}$ uma função convexa em D .*

Então todo minimizador local do problema (1.1) é global. Além disso, o conjunto de minimizadores é convexo.

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

Demonstração. Seja $\bar{x} \in D$ um minimizador local da função convexa $f : D \rightarrow \mathbb{R}$. Se existisse $y \in D$ tal que $f(y) < f(\bar{x})$, então para todo $\alpha \in [0, 1]$ e pela convexidade de D teríamos, $(1 - \alpha)y + \alpha\bar{x} \in D$. Agora pela convexidade de f , para todo $\alpha \in (0, 1]$, teríamos, $f((1 - \alpha)y + \alpha\bar{x}) \leq (1 - \alpha)f(y) + \alpha f(\bar{x}) < (1 - \alpha)f(y) + \alpha f(y) = f(y)$. Pondo $x(\alpha) = (1 - \alpha)y + \alpha\bar{x}$ e assim tomando $\alpha > 0$ suficientemente pequeno, poderíamos garantir o ponto $x(\alpha)$ tão próximo de \bar{x} quando se queira, com $f(x(\alpha)) < f(\bar{x})$. Isto contradiz o fato de que \bar{x} é minimizador local do problema (1.1). Portanto qualquer solução local deve ser global.

Sejam $S \subset D$ o conjunto de minimizadores globais e $\bar{v} \in \mathbb{R}$ o valor ótimo do problema.

Para quaisquer $x \in S$, $\bar{x} \in S$ e $\alpha \in [0, 1]$, pela convexidade de f , obtemos

$$f(\alpha x + (1 - \alpha)\bar{x}) \leq \alpha f(x) + (1 - \alpha)f(\bar{x}) = \alpha\bar{v} + (1 - \alpha)\bar{v} = \bar{v} \Rightarrow f(\alpha x + (1 - \alpha)\bar{x}) = \bar{v}$$

e portanto, $\alpha x + (1 - \alpha)\bar{x} \in S$. Suponhamos agora que f seja estritamente convexa e que existam $x \in S$ e $\bar{x} \in S$, $x \neq \bar{x}$. Seja $\alpha \in (0, 1)$. Como x e \bar{x} são minimizadores

globais e $\alpha x + (1 - \alpha)\bar{x} \in D$, pela convexidade de D , segue que

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

Por outro lado, pela convexidade estrita,

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

O que resulta em contradição. Portanto, concluímos que o minimizador é único. \square

1.2.2 Funções convexas não diferenciáveis

A seguir estudaremos as propriedades de funções convexas não diferenciáveis. Esta classe de funções é muito importante pois de acordo com Solodov (2009) a perda da diferenciabilidade acontece em otimização com frequência. Mostraremos, no entanto, que uma função convexa possui derivadas direcionais e possui ainda uma derivada generalizada, que se chama subdiferencial.

Teorema 1.6. (*Derivada direcional de uma função convexa*) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então para todo $x \in \mathbb{R}^n$, f é diferenciável em cada direção $d \in \mathbb{R}^n$. Além disso,*

$$f(x + \alpha d) \geq f(x) + \alpha f'(x; d) \quad \forall \alpha \in \mathbb{R}_+.$$

Demonstração. Ver [22]. \square

Definição 13. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Dizemos que $y \in \mathbb{R}^n$ é um *subgradiente* de f no ponto $x \in \mathbb{R}^n$ se

$$f(z) \geq f(x) + \langle y, z - x \rangle \quad \forall z \in \mathbb{R}^n. \quad (1.19)$$

O conjunto de todos os subgradientes de f em x se chama o *subdiferencial* de f em x , o denotamos por $\partial f(x)$.

Caracterização do subdiferencial de uma função convexa

Seja $y \in \partial f(x)$ e pondo $z = x + \alpha d$, onde $d \in \mathbb{R}^n, \alpha > 0 \forall z \in \mathbb{R}^n$, tem-se

$$f(z) \geq f(x) + \langle y, z - x \rangle \Leftrightarrow f(x + \alpha d) \geq f(x) + \langle y, x + \alpha d - x \rangle \Leftrightarrow$$

$$f(x + \alpha d) \geq f(x) + \alpha \langle y, d \rangle \Leftrightarrow \frac{f(x + \alpha d) - f(x)}{\alpha} \geq \langle y, d \rangle$$

$$\lim_{\alpha \rightarrow 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha} \geq \lim_{\alpha \rightarrow 0^+} \langle y, d \rangle \quad (1.20)$$

Pelo teorema (1.5) o limite existe quando $\alpha \rightarrow 0^+$, logo $f'(x; d) \geq \langle y, d \rangle$. Portanto,

$$\partial f(x) = \{y \in \mathbb{R}^n \mid f'(x; d) \geq \langle y, d \rangle \forall d \in \mathbb{R}^n\}. \quad (1.21)$$

Exemplo 7. A função $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = |x|$ é convexa e não possui derivada no zero.

Note que o ponto $\bar{x} = 0$ é o único minimizador global irrestrito de f . Afirmamos que $0 \in \partial f(\bar{x}) \Leftrightarrow \bar{x} = 0$. Com efeito, pela caracterização do subdiferencial de f em \bar{x} , dado $y \in \partial f(\bar{x}) \Leftrightarrow f'(x; d) \geq \langle y, d \rangle \forall d \in \mathbb{R}$. Calculando a derivada direcional quando $d > 0$ e $d < 0$, respectivamente, obtemos

$$d = f'(0; d) \geq yd \text{ e } -d = f'(0; d) \leq yd \Rightarrow y \leq 1, y \geq -1 \Rightarrow \partial f(\bar{x}) = [-1, 1].$$

□

Proposição 1.7. *Uma função convexa $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é diferenciável no ponto $x \in \mathbb{R}^n$, se, e somente se, o conjunto $\partial f(x)$ contém um elemento só. Neste caso, $\partial f(x) = \{\nabla f(x)\}$*

Demonstração. Ver [22]. □

A seguir iremos mostrar alguns resultados do subdiferencial através de exemplificações, importantes para o desenvolvimento da teoria da otimização imprescindíveis para o entendimento dos resultados posteriores.

O subdiferencial de uma função convexa

Teorema 1.8. *(Condições de otimalidade para minimização de uma função convexa num conjunto compacto)* Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e $D \subset \mathbb{R}^n$ um conjunto convexo. Então $\bar{x} \in \mathbb{R}^n$ é um minimizador de f em D se, e somente se,

$$\exists y \in \partial f(\bar{x}) \text{ tal que } \langle y, x - \bar{x} \rangle \geq 0 \quad \forall x \in D, \quad (1.22)$$

ou equivalentemente,

$$0 \in \partial f(\bar{x}) + \mathcal{N}_{D(\bar{x})}. \quad (1.23)$$

Em particular, \bar{x} é um minimizador de f no \mathbb{R}^n se, e somente se, $0 \in \partial f(\bar{x})$.

Demonstração. Ver [22]. □

Proposição 1.9. *(O subdiferencial da soma de funções convexas)* Sejam $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, p$ funções convexas. Então

$$\partial\left(\sum_{i=1}^p f_i(x)\right) = \sum_{i=1}^p \partial f_i(x). \quad (1.24)$$

Demonstração. Ver [22]. □

Teorema 1.10. *(A derivada direcional e o subdiferencial do máximo de funções convexas)* Suponhamos que Z seja um conjunto compacto qualquer. Suponhamos que $\psi : \mathbb{R}^n \times Z \rightarrow \mathbb{R}$ seja uma função contínua tal que $\psi(\cdot, z) : \mathbb{R}^n \rightarrow \mathbb{R}$ seja convexa para todo $z \in Z$. Seja

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f(x) = \max_{z \in Z} \psi(x, z).$$

Então:

1. (a) A função f é convexa no \mathbb{R}^n , e para todo $x \in \mathbb{R}^n$, tem-se

$$f'(x; d) = \max_{\bar{z} \in Z(x)} \psi'(x, \bar{z}; d) \quad \forall d \in \mathbb{R}^n, \quad (1.25)$$

onde

$$Z(x) = \{\bar{z} \in Z \mid f(x) = \psi(x, \bar{z}) = \max_{z \in Z} \psi(x, z)\}. \quad (1.26)$$

Em particular, se $Z(x) = \{\bar{z}\}$ e $\psi(\cdot, \bar{z})$ é diferenciável no ponto $x \in \mathbb{R}^n$, então f é diferenciável em x e $\nabla f(x) = \psi'_x(x, \bar{z})$.

2. b) Se $\psi(\cdot, z)$ é diferenciável para todo $z \in Z$ e $\psi'_x(x, \cdot)$ é contínua em Z para todo $x \in \mathbb{R}^n$, então

$$\partial f(x) = \text{conv}\{\psi'_x(x, \bar{z}) \mid \bar{z} \in Z(x)\}, \quad x \in \mathbb{R}^n. \quad (1.27)$$

Demonstração. Ver [22]. □

Exemplo 8. Sejam $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ funções convexas diferenciáveis, $i = 1, \dots, p$ e

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f(x) = \max_{i=1, \dots, p} f_i(x).$$

onde $I(x) = \{i = 1, \dots, p \mid f(x) = f_i(x)\}$ é o conjunto de índices ativos em x . Pelo teorema acima, temos que

$$\partial f(x) = \text{conv}\{\nabla f_i(x), i \in I(x)\}, \quad x \in \mathbb{R}^n. \quad (1.28)$$

Exemplo 9. Resolver o problema de $\min f(x)$, $x \in \mathbb{R}^n$, onde

$$f(x) = \frac{1}{2} \|x\|^2 + |\langle a, x \rangle + 1| \quad (1.29)$$

e $a \in \mathbb{R}^n$ é um parâmetro.

Solução. Sejam $f_1(x) = \frac{1}{2}\|x\|^2$ e $f_2(x) = |\langle a, x \rangle + 1| = \max\{\langle a, x \rangle + 1, -(\langle a, x \rangle + 1)\}$. Temos que $f(x) = f_1(x) + f_2(x)$ é fortemente convexa, pois é a soma de uma função f_1 fortemente convexa com a função f_2 convexa. Logo, o problema irrestrito possui solução e é única para qualquer parâmetro $a \in \mathbb{R}^n$. Pelas proposições (1.7) e (1.9) temos que

$$\partial f(x) = \partial f_1(x) + \partial f_2(x)$$

Assim, $\partial f_1(x) = \nabla f_1(x) = x$, $\partial f_2(x) = \nabla f_2(x) = a$ para $\langle a, x \rangle > -1$ e $\partial f_2(x) = \nabla f_2(x) = -a$ para $\langle a, x \rangle < -1$

Agora pelo teorema (1.10)

$$\partial f_2(x) = \text{conv}\{-a; a\} \text{ desde que } \langle a, x \rangle = -1.$$

Pelo teorema (1.8) o zero pertence ao subdiferencial, segue que

$$0 \in \partial f(x) = x + \partial f_2(x)$$

Consideramos as três possibilidades na definição de $\partial f_2(x)$.

Para $\partial f_2(x) = a$, devemos satisfazer

$$0 = x + a \Rightarrow x = -a, \quad \langle a, x \rangle = \langle a, -a \rangle = -\|a\|^2 > -1 \Rightarrow \|a\|^2 < 1.$$

Assim, a solução do problema existe e é única para $\bar{x} = -a$ desde que $\|a\| < 1$.

Para $\partial f_2(x) = -a$, devemos satisfazer

$$0 = x - a \Rightarrow x = a, \quad \langle a, x \rangle = \langle a, a \rangle = \|a\|^2 < -1 \text{ absurdo.}$$

Para $\partial f_2(x) = \text{conv}\{-a; a\}$, devemos satisfazer

$$0 = x + (1-t)(-a) + ta = x + 2ta - a = x + (2t-1)a \Rightarrow 1-2t = -\frac{1}{\|a\|^2}, \text{ onde } t \in [0, 1]$$

$$\langle a, x \rangle = \langle a, (1-2t)a \rangle = (1-2t)\|a\|^2 = -1 \Rightarrow \|a\|^2 \geq 1, \forall t \in (0, 1].$$

Portanto, a solução do problema existe e é única para $\bar{x} = (1-2t)a = -\frac{a}{\|a\|^2}$ desde que $\|a\| \geq 1$.

1.3 Fundamentos de programação multiobjetivo

De acordo com Cohon(1978), a otimização multiobjetivo ou vetorial consiste em minimizar duas ou mais funções objetivos em geral conflitantes o que diferencia da otimização clássica, ou seja, enquanto o valor ótimo na clássica é único, não podemos dizer o mesmo para problemas multiobjetivos porque uma solução que minimiza uma das funções objetivos não será, em geral, qualquer minimização das demais funções objetivos em questão. O que é melhor em termos de um dos objetivos geralmente resulta na piora para outras. Dessa forma, pretendemos na realidade encontrar, dentro do conjunto de decisões admissíveis ou viáveis, um vetor de variáveis de decisão que otimize um vetor de funções cujas componentes representam as funções objetivos. Para entender esses conceitos, precisamos da definição do conceito de solução eficiente ou Pareto-ótimo introduzido por Vilfredo Pareto (1848 – 1923).

Formalmente, queremos resolver o seguinte problema

$$\begin{cases} \min F(x) = (F_1(x), \dots, F_m(x)) \\ \text{s.a } x \in D \end{cases} \quad (1.30)$$

$F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $D \subset \mathbb{R}^n$ é o conjunto viável ou espaço de decisões e $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ é a i -ésima função objetivo.

Resolver o problema (1.31), significa encontrar soluções com todos os valores dos objetivos que não podem ser melhorados simultaneamente. Neste contexto chamaremos $x \in \mathbb{R}^n$ de *vetor decisor*, $F(x)$ de *vetor objetivo* e $F(D)$ de *espaço de objetivos*.

1.4 Relação de ordem parcial

Um problema vetorial ou multiobjetivo difere do problema clássico, justamente porque o espaço vetorial \mathbb{R}^m não é ordenado, ou seja, dado dois vetores $x, y \in \mathbb{R}^m$, em geral são não comparáveis, no sentido de que não podemos dizer explicitamente se x é "menor do que" y ou y é "menor do que" x , como ocorre em \mathbb{R} e também porque na otimização clássica o valor ótimo é único. O mesmo não acontece na otimização vetorial. Assim, encontrar uma solução "eficiente" ou ótima para o problema vetorial ou multiobjetivo, tal qual na otimização clássica torna-se inviável. Dessa forma, conceitos como ordenação parcial induzida por um cone convexo fechado com interior não vazio e dominância (não-dominância, ou não (eficientes)), são essenciais para entendermos, por solução ótima, de um problema vetorial ou multiobjetivo.

Uma relação binária A em \mathbb{R}^n é um subconjunto do produto cartesiano $\mathbb{R}^n \times \mathbb{R}^n$. Assim, diz-se que um elemento $x \in \mathbb{R}^n$ está relacionado com $y \in \mathbb{R}^n$ por A se $(x, y) \in A$.

Definição 14. Seja A uma relação binária em \mathbb{R}^n . Dizemos que ela é;

1. reflexiva, se $(x, x) \in A$ para todo $x \in \mathbb{R}^n$;
2. anti-simétrica, se $(x, y) \in A$ e se $(y, x) \in A$ implica que $x = y$ para $x, y \in \mathbb{R}^n$;
3. transitiva, se $(x, y) \in A$ e se $(y, z) \in A$ implica que $(x, z) \in A$ para $x, y, z \in A$;
4. completo ou conectado, se $(x, y) \in A$ ou $(y, x) \in A$ para todo $x, y \in \mathbb{R}^n$ com $x \neq y$;
5. linear, se $(x_1, y_1), (x_2, y_2) \in A$ e $t \geq 0$ então (tx_1, ty_1) e $(x_1 + x_2, y_1 + y_2) \in A$.

Definição 15. (Conjunto parcialmente ordenado) Sejam $X \subset \mathbb{R}^n$. Dizemos que um conjunto X é parcialmente ordenado se para todo $x, y \in X$ valem as propriedades 1, 2 e 3 da definição (15).

Vale destacar que uma ordem parcial é uma relação binária que possui as propriedades (1); (2); (3). Se além disso, satisfaz a propriedade (5), dizemos que a ordem parcial é linear. Quando uma ordem parcial satisfaz a propriedade (4), ela é chamada de ordem total. Por exemplo, o conjunto dos números reais é totalmente ordenado.

É claro que o espaço vetorial \mathbb{R}^n não é totalmente ordenado, no entanto podemos estabelecer em \mathbb{R}^n várias relações de ordem parcial. A saber,

Definição 16. (Comparação entre vetores) Sejam $x, y \in \mathbb{R}^n$. As relações de " \leq ", " $<$ " e " $=$ ", são definidas da seguinte forma:

$$x \leq y \Leftrightarrow x_i \leq y_i, i = 1 \dots n; \quad (1.31)$$

$$x < y \Leftrightarrow x_i < y_i, i = 1 \dots n; \quad (1.32)$$

$$x = y \Leftrightarrow x_i = y_i, i = 1 \dots n; \quad (1.33)$$

as relações " \geq " e " $>$ " são definidas de forma análoga. Já a relação " \neq " é definida como:

$$x \neq y \Leftrightarrow \exists i \mid x_i \neq y_i. \quad (1.34)$$

Quando entre os vetores x e y não for possível relacioná-los através de " \leq " (ou seja, não for possível escrever $x \leq y$ ou $y \leq x$), será utilizada a notação $x \not\leq y$.

As Figuras 1(a) e 1(b) ilustram os casos (1.32) e (1.33), respectivamente.

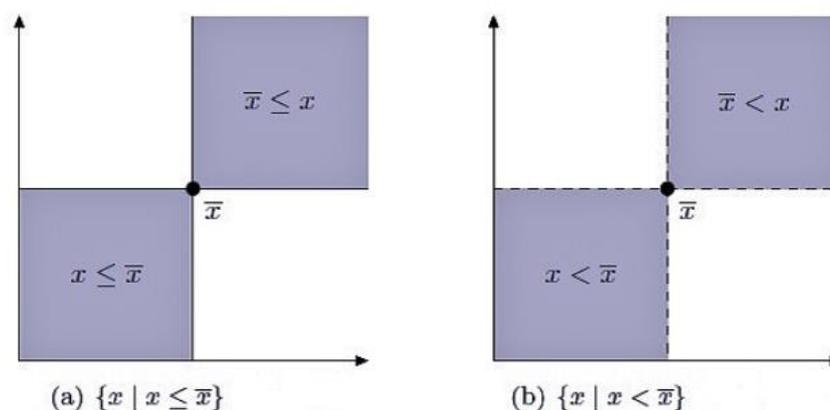


Figura 1.4: Relações \leq e $<$ no \mathbb{R}^n . Fonte *Sampaio* (2011).

Exemplo 10. [18] Considere os vetores $u, v, w \in X \subset \mathbb{R}^3$ tais que:

$$u = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad v = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}, \quad w = \begin{bmatrix} 7 \\ 4 \\ 8 \end{bmatrix}$$

Então verificam-se as seguintes relações: $u \leq v$, $u \leq w$, $v \not\leq w$ e $w \not\leq v$.

Definição 17. (Pareto Dominância no espaço de objetivos) Sejam $y_1, y_2 \in Y$, dois vetores no espaço de objetivos. Diz-se que y_1 domina y_2 se $y_1 \leq y_2$ e $y_1 \neq y_2$. Em outras palavras, $y_{1i} \leq y_{2i}$, $\forall i \in \{1, \dots, m\}$ e $y_{1i} < y_{2i}$ para pelo menos um $i \in \{1, \dots, m\}$. Caso aconteça que $y_1 \not\leq y_2$ e $y_2 \not\leq y_1$ (y_1 não domina y_2 e y_2 não domina y_1), então y_1, y_2 são chamados indiferentes.

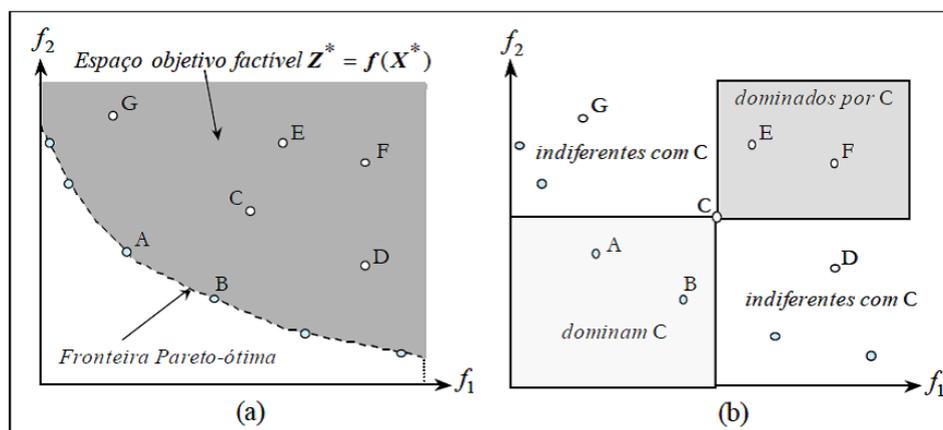


Figura 1.5: Dominância de Pareto no espaço de objetivos .Fonte Arroyo (2002).

Na Figura 1.5 (b), o ponto C domina os pontos pertencentes ao retângulo superior direito (subconjunto do espaço de objetivos). Os pontos pertencentes ao retângulo inferior esquerdo dominam o ponto C. Os pontos G, C e D são indiferentes.

Definição 18. (Dominância de Pareto no espaço de decisão) Sejam $x_1, x_2 \in X$. Diz-se que x_1 domina x_2 se $f(x_1)$ dominar $f(x_2)$, ou seja, $f(x_1) \leq f(x_2)$ e $f(x_1) \neq f(x_2)$. Caso ocorra $f(x_1) \not\leq f(x_2)$ e $f(x_2) \not\leq f(x_1)$, então x_1, x_2 são chamados indiferentes.

Definição 19. (Conjunto Pareto-ótimo) Um vetor de decisão $\bar{x} \in X$ é uma solução Pareto-ótima se não existe outro vetor de decisão $x \in X$ tal que $f(x) \leq f(\bar{x})$ e $f(x) \neq f(\bar{x})$, em outras palavras, se nenhum $x \in X$ dominar $\bar{x} \in X$, então \bar{x} é Pareto-ótimo. O conjunto de todas as soluções Pareto-ótimas é denominado Conjunto Pareto-ótimo. A imagem do conjunto Pareto-ótimo em $Y = f(X)$ é chamada fronteira Pareto-Ótima.

Definição 20. (Solução localmente Pareto-ótima) Um vetor de decisão $\bar{x} \in X$ é localmente Pareto-ótimo se existe $\delta > 0$ tal que \bar{x} é Pareto-ótimo em $X \cap B(\bar{x}, \delta)$. Ou seja, se \bar{x} não é dominado por nenhum ponto da bola $B(\bar{x}, \delta)$. Um vetor de objetivos $y^* \in Y$ é localmente Pareto-ótimo se seu vetor de decisão correspondente é localmente Pareto-ótimo.

Definição 21. (Programação Multiobjetivo Convexa) O problema de otimização multiobjetivo definido em (1.31) é convexo se todas as funções coordenadas da função

vetorial F e o conjunto $D \subset \mathbb{R}^n$ forem convexos.

Teorema 1.11. *Se o problema de otimização definido em ((1.31)) for convexo, então toda solução localmente Pareto-ótima é também globalmente Pareto-ótima.*

Demonstração. Seja $\bar{x} \in D$ localmente Pareto-ótimo, isto é, existem $\delta > 0$ e uma bola $B(\bar{x}; \delta)$ tal que não existe um $x \in D \cap B(\bar{x}; \delta)$ para os quais $f_j(x) \leq f_j(\bar{x})$ para todo $j = 1, 2, \dots, m$ e $f_i(x) < f_i(\bar{x})$ para algum i . Suponhamos que $\bar{x} \in D$ não seja Pareto-ótimo global, isto é, existe $\hat{x} \in D$ tal que

$$f_j(\hat{x}) \leq f_j(\bar{x}) \text{ para todo } j = 1, 2, \dots, m \text{ e } f_i(\hat{x}) < f_i(\bar{x}) \text{ para algum } i. \quad (1.35)$$

Defina $y(t) = (1-t)\hat{x} + t\bar{x}$, onde $t \in (0, 1)$ é tomado de modo que $y(t) \in B(\bar{x}; \delta)$. Pela convexidade de D tem-se $y(t) \in D$. Por outro lado, pela convexidade de cada f_j

$$f_j(y(t)) = f_j((1-t)\hat{x} + t\bar{x}) \leq (1-t)f_j(\hat{x}) + tf_j(\bar{x}) \quad (1.36)$$

Por (1.35) tem-se,

$$f_j(y(t)) \leq (1-t)f_j(\bar{x}) + tf_j(\bar{x}) \quad (1.37)$$

logo,

$$f_j(y(t)) \leq f_j(\bar{x}) \quad \forall j = 1, 2, \dots, m. \quad (1.38)$$

Assim, pela hipótese de \bar{x} ser localmente Pareto-ótimo e $y(t) \in B(\bar{x}; \delta)$, logo $f_j(y(t)) = f_j(\bar{x})$ para todo $j = 1, 2, \dots, m$, e por (1.38), obtemos $(1-t)f_j(\bar{x}) \leq (1-t)f_j(\hat{x})$ para todo $j = 1, 2, \dots, m$, sendo $(1-t) > 0$, logo $f_j(\bar{x}) \leq f_j(\hat{x}) \quad \forall j = 1, 2, \dots, m$, o que é uma contradição, pois supomos $f_i(\hat{x}) < f_i(\bar{x})$ para algum i . Portanto, $\bar{x} \in D$ é Pareto-ótimo global. \square

A seguir mostraremos exemplos de programação vetorial e as definições que as

compõem.

1.5 Fundamentos de programação vetorial

Definição 22. Seja $A \subseteq \mathbb{R}^m$. O cone polar positivo de A é definido por

$$A^* = \{w \in \mathbb{R}^m / \langle y, w \rangle \geq 0 \forall y \in A\} \quad (1.39)$$

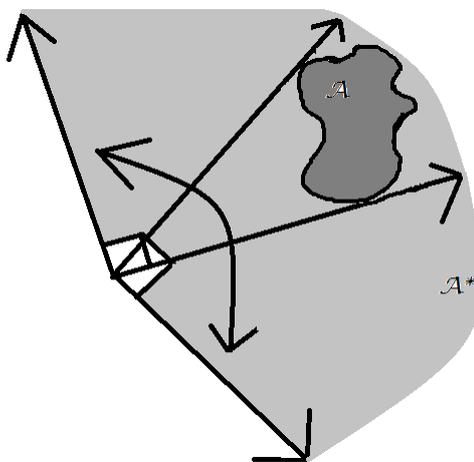


Figura 1.6: Exemplo de um cone polar positivo.

Afirmamos que A^* é fechado. Com efeito,

seja $y \in \overline{A^*}$, logo $y = \lim_k y^k$ onde $y^k \in A^* \forall k$. Assim, $\langle y^k, d \rangle \geq 0 \forall k$.

Passando ao limite quando $(k \rightarrow \infty)$, pela continuidade do produto interno obtemos que $\langle y, d \rangle \geq 0$. Portanto $y \in A^*$.

Exemplo 11. Seja $A = [0, +\infty)$. O cone polar positivo de A é o conjunto $A^* = [0, +\infty)$. De fato,

dado $z \in A^* \Rightarrow z \cdot d \geq 0 \forall d \in A$. Sendo $d \geq 0$, tomando $d = 1$, logo $z \cdot 1 = z \geq 0 \Rightarrow z \in A$.

Reciprocamente, dado $z \in A \Rightarrow z \geq 0 \Rightarrow d \cdot z \geq 0 \forall d \geq 0$. Portanto, $z \in A^*$.

Proposição 1.12. Seja $\emptyset \neq K \subset \mathbb{R}^m$ um cone convexo e fechado. Tem-se que

$$K = (K^*)^*. \quad (1.40)$$

Demonstração. Ver [15]. □

Pela proposição acima podemos reescrever o cone K da seguinte forma,

$$K = \{y \in \mathbb{R}^m / \langle y, w \rangle \geq 0 \forall w \in K^*\} \quad (1.41)$$

$$-K = \{y \in \mathbb{R}^m / \langle y, w \rangle \leq 0 \forall w \in K^*\} \quad (1.42)$$

e

$$\text{int}K = \{y \in \mathbb{R}^m / \langle y, w \rangle > 0 \forall w \in K^* \setminus \{0\}\} \quad (1.43)$$

$$-\text{int}K = \{y \in \mathbb{R}^m / \langle y, w \rangle < 0 \forall w \in K^* \setminus \{0\}\} \quad (1.44)$$

Caracterização do cone. Seja $C \subseteq \mathbb{R}^m$ um compacto tal que

1. $0 \notin C$
2. $\text{cone}(\text{conv}C) = K^*$

Como $\text{int}(K) \neq \emptyset$ e $C \subseteq K^* \setminus \{0\}$, segue que $0 \notin \text{conv}(C)$. Consequentemente

$$-K = \{u \in \mathbb{R}^m / \langle u, w \rangle \leq 0 \forall w \in C\}, \quad (1.45)$$

$$-\text{int}K = \{u \in \mathbb{R}^m / \langle u, w \rangle < 0 \forall w \in C\}. \quad (1.46)$$

Note que para otimização clássica, basta fazer $K = \mathbb{R}_+$ e $C = \{1\}$. Assim temos,

$$\mathbb{R}_+^* = \{w \in \mathbb{R}^m / y \cdot w \geq 0 \forall y \in \mathbb{R}_+\} = \mathbb{R}_+$$

$$0 \notin C = \{1\}$$

$$\text{conv}C = \text{conv}\{1\} = \{1\}$$

Por outro lado, $\text{cone}(\text{conv}C) = \text{cone}(\{1\}) = \mathbb{R}_+$, ou seja, $\text{int}K \neq \emptyset$, $C = \{1\} \subseteq K^* \setminus \{0\}$,

logo $0 \notin \text{conv}C$.

Além disso, por (1.46) e (1.47), obtemos

$$-K = \{u \in \mathbb{R} / u \cdot w \leq 0 \forall w \in C\} = -\mathbb{R}_+, \quad (1.47)$$

$$-\text{int}K = \{u \in \mathbb{R} / u \cdot w < 0 \forall w \in C\} = -\mathbb{R}_{++}. \quad (1.48)$$

No caso multiobjetivo, K , K^* são os ortantes positivos do \mathbb{R}^m e C uma base canônica do \mathbb{R}^m .

Para um cone K genérico(convexo, fechado, pontiagudo e com o interior não vazio), o conjunto

$$C = \{w \in K^* / \|w\| = 1\} \quad (1.49)$$

Cumpra as condições (1) e (2) da caracterização.

Definição 23. Dado $K \subset \mathbb{R}^m$, um cone não vazio, convexo, fechado e pontiagudo. A relação binária associada a K , denotada por \preceq_K , é definida como:

1. a) $x \preceq_K y \Leftrightarrow y - x \in K$

Neste caso, diz-se que x **domina** y

2. b) $x \prec_K y \Leftrightarrow y - x \in \text{int}(K)$

Neste caso, diz-se que x **domina estritamente** y .

Definição 24. Dizemos que um ponto $x^* \in \mathbb{R}^n$ é K -minimizador irrestrito(ou K -ótimo) se não existe um $y \in \mathbb{R}^n$ com $F(y) \preceq_K F(x^*)$ e $F(y) \neq F(x^*)$.

Exemplo 12. Sejam $K = \mathbb{R}_+^2$, $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $F(x, y) = (x^2 + 1, x^2 + y^2)$. Temos que $(0, 0)$ é um ponto K -ótimo de F , pois $x^2 + 1 < 1$ e $x^2 + y^2 < 0$ não possuem solução para nenhum par $(x, y) \in \mathbb{R}^2$. Assim, não existe um $(x, y) \in \mathbb{R}^2$ com $F((x, y)) \preceq_K F((0, 0))$ e $F((x, y)) \neq F((0, 0))$.

Definição 25. Dizemos que um ponto $x \in \mathbb{R}^n$ é K-crítico de F quando

$$\text{Im}JF(x) \cap -\text{int}(K) = \emptyset \quad (1.50)$$

Dizer que um ponto $x \in \mathbb{R}^n$ não é K-crítico significa que existe uma direção de descida $v \in \mathbb{R}^n$ satisfazendo $JF(x)v \in -\text{int}(K)$ isto é, $JF(x)v \prec_K 0$.

Definição 26. [10] Seja $x \in \mathbb{R}^n$ um ponto não K-crítico. Dizemos que $v \in \mathbb{R}^n$ é uma K-direção de descida de F em x quando

$$JF(x)v \in -\text{int}(K) \text{ ou seja, } JF(x)v \prec_K 0. \quad (1.51)$$

Definição 27. [10] Dizemos que um ponto $v \in \mathbb{R}^n$ é uma K-direção de descida em x , se satisfaz (1.52). Além disso, existe um $t \geq 0$ tal que $F(x + tv) \preceq_K F(x)$ para todo $t \in (0, \bar{t})$.

Definição 28. Seja v uma K-direção de descida em x dizemos que $t > 0$ satisfaz uma regra tipo Armijo para $\beta \in (0, 1)$ se $F(x + tv) \preceq_K F(x) + \beta t JF(x)v$.

Como no caso real, dada uma direção de descida, a regra tipo Armijo é satisfeita para alguns t_s .

Proposição 1.13. *Seja $\beta \in (0, 1)$. Se $JF(x)v \prec_K 0$, então existe um $\bar{t} > 0$ tal que,*

$$F(x + tv) \prec_K F(x) + \beta t JF(x)v \quad \forall t \in (0, \bar{t})$$

Demonstração. Sendo F diferenciável, temos que,

$$F(x + tv) = F(x) + tJF(x)v + tR(t) \quad (1.52)$$

onde $\lim_{t \rightarrow 0} R(t) = 0$. Assumimos que $JF(x)v \in -\text{int}(K)$. Sendo $\beta \in (0, 1)$ implica que $(1 - \beta)JF(x)v \in -\text{int}(K)$. De $\lim_{t \rightarrow 0} \frac{R(t)}{t} = 0$, existe $\bar{t} > 0$ tal que para todo $t \in (0, \bar{t})$ a

norma $\|R(t)\|$ é suficientemente pequena de modo que, $R(t) + (1 - \beta)JF(x)v \in -\text{int}(K)$ o que significa dizer que $tR(t) \prec_K -t(1 - \beta)JF(x)v$. Combinando esta K-desigualdade com (1.53) obtemos,

$$F(x + tv) = F(x) + tJF(x)v + tR(t) \quad (1.53)$$

$$\prec_K F(x) + tJF(x)v - t(1 - \beta)JF(x)v \quad (1.54)$$

$$\prec_K F(x) + tJF(x)v - tJF(x)v + t\beta JF(x)v \quad (1.55)$$

$$= F(x) + \beta tJF(x)v. \quad (1.56)$$

□

Capítulo 2

Um Algoritmo para Programação Vetorial

Neste capítulo consideramos uma função vetorial $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ continuamente diferenciável. Estamos interessados em resolver o problema de otimização vetorial irrestrito onde a ordem parcial é induzida por um cone K genérico (convexo, fechado, pontiagudo e com interior não vazio), ou seja, buscamos um minimizador irrestrito da função F no sentido de Pareto. Mais precisamente resolveremos o seguinte problema:

$$\begin{cases} \min_K F(x) = (F_1(x), \dots, F_m(x)) \\ \text{s.a } x \in \mathbb{R}^n \end{cases} \quad (2.1)$$

2.1 Algoritmo para o método K-direção de descida genérico (AMKdG)

1. Passo 1. Tomar $\beta \in (0, 1)$, $x^0 \in \mathbb{R}^n$. $k := 0$
2. Passo 2. Se x^k for K-crítico Pare! Caso contrário
3. Passo 3. Calcule v^k , uma K-direção de descida em x^k .
4. Passo 4. Calcule o comprimento do passo $t_k > 0$ tal que $F(x^k + t_k v^k) \preceq_K$

$$F(x^k) + \beta t_k JF(x^k)v^k$$

5. Passo 5. Fazer $x^{k+1} := x^k + t_k v^k, k := k + 1$ e volte passo 2.

Definição 29. Dizemos que uma seqüência $\{x^k\}_k \subset \mathbb{R}^n$ é K-decrescente com relação ao cone K quando, $x^k \prec_K x^l \quad \forall k, l$ e $k > l$.

Proposição 2.1. *Seja $\{x^k\}_k$ uma seqüência infinita gerada pelo algoritmo(AMKdG).*

Se \bar{x} é um ponto de acumulação de $\{x^k\}_k$, então

$F(\bar{x}) \prec_K F(x^k) \quad \forall k$ e $\lim_{x \rightarrow +\infty} F(x^k) = F(\bar{x})$. Em particular, F é constante no conjunto de pontos de acumulação de $\{x^k\}_k$.

Demonstração. Suponha que a seqüência $\{x^k\}_k$ gerada pelo algoritmo seja infinita. Portanto, para cada $k \in \mathbb{N}$, x^k não é K-crítico. Isto é, existe uma direção $v^k \in \mathbb{R}^n$ satisfazendo $JF(x^k)v^k \in -\text{int}(K)$, logo $\beta t_k JF(x^k)v^k \in -\text{int}(K) \quad \forall k \in \mathbb{N}$, o que equivale dizer que $JF(x^k)v^k \prec_K 0$ e pelo algoritmo no passo 4, temos,

$$F(x^k + t_k v^k) \preceq_K F(x^k) + \beta t_k JF(x^k)v^k \prec_K F(x^k).$$

Donde

$$F(x^{k+1}) \prec_K F(x^k).$$

Logo $\{F(x^k)\}_k$ é K-decrescente. Por hipótese \bar{x} é ponto de acumulação de $\{x^k\}_k$, isto é, existe uma subseqüência $\{x^{k_j}\}_j \subset \{x^k\}_k$ tal que $\lim_{j \rightarrow +\infty} x^{k_j} = \bar{x}$. Seja $\mathbb{N}_1 \subseteq \mathbb{N}$ um subconjunto infinito de índices. Assim para $k \in \mathbb{N} \exists k_j \in \mathbb{N}_1, k_j > k$ com $F(x^{k_j}) \preceq_K F(x^k)$ passando ao limite quando $(j \rightarrow \infty)$ e pela continuidade de F obtemos, $F(\bar{x}) \prec_K F(x^k) \quad \forall k \in \mathbb{N}$.

Seja \hat{x} outro ponto de acumulação de $\{x^k\}_k$. Assim, existe uma subseqüência $\{x^{k_p}\}_p \subset \{x^k\}_k$ tal que $\lim_{p \rightarrow +\infty} x^{k_p} = \hat{x}$. Seja $\mathbb{N}_2 \subset \mathbb{N}$, logo $F(\bar{x}) \preceq_K F(x^{k_p})$, pois $F(\bar{x}) \prec_K F(x^k) \quad \forall k \in \mathbb{N}$. Novamente, passando ao limite quando $(p \rightarrow \infty)$, pela continuidade de F obtemos, $F(\bar{x}) \preceq_K F(\hat{x}) \quad \forall k \in \mathbb{N}$, donde $F(\hat{x}) - F(\bar{x}) \in K$. Pelo mesmo argumento

anterior, mas começando por \hat{x} em lugar de \bar{x} , obtemos

$$F(\bar{x}) - F(\hat{x}) \in K$$

donde

$$-(F(\hat{x}) - F(\bar{x})) = F(\bar{x}) - F(\hat{x}) \in K \Rightarrow -(F(\hat{x}) - F(\bar{x})) \in K$$

Como $K \cap -K = \{0\}$, logo $F(\hat{x}) - F(\bar{x}) = 0 \Rightarrow F(\hat{x}) = F(\bar{x})$ e conseqüentemente

$$\lim_{k \rightarrow +\infty} F(x^k) = F(\bar{x}) \quad \square$$

2.1.1 Comentários sobre a convergência do algoritmo AMKdG

Em relação a escolha do tamanho dos passos, se forem tomados demasiadamente pequenos a seqüência $\{x^k\}_k$ gerada pelo algoritmo(AMKdG) pode não convergir para um ponto K-crítico. Assim, devemos caminhar sobre direções que promovam reduções da função, pois segundo Solodov2 (2009) a regra de Armijo, não deixa de ser uma busca linear, mas é mais econômica computacionalmente em relação a busca unidimensional.

A regra de Armijo consiste em computar um comprimento de passo que resulte em decréscimo suficiente da função objetivo, garantindo assim convergência. Dessa forma, afim de escolher um comprimento de passo adequado na iterada k , prescrevemos o procedimento *backtracking* usual que consiste em estimar o passo retornando a uma fração do passo anterior, se houver fracasso no passo vigente doo teste, ou seja, quando fracassa Armijo. Vejamos tal procedimento:

1. defina $t := 1$
2. Se $F(x^k + tv^k) \preceq_K F(x^k) + \beta t JF(x^k)v^k$. Então $t_k = t$ e (of backtracking)
Caso contrário,
3. Defina $t := \frac{t}{2}$, volte 2

Note que o procedimento (critério) *backtracking* acima é finito, no sentido que

termina em número finito de passos pela proposição (1.13)

Além disso,

$$t_k := \max\{2^{-j}/j \in \mathbb{N}, F(x^k + 2^{-j}v^k) \preceq_K F(x^k) + \beta 2^{-j} JF(x^k)v^k\} \quad (2.2)$$

Acabamos de ver como se dá a relação da escolha dos passos. Em seguida vamos apresentar, como se calcular a direção de descida através de uma função auxiliar fortemente convexa.

2.2 Cálculo da direção de descida

Algoritmos para otimização irrestrita se diferenciam pela maneira como se obtém a direção de descida e também pela maneira de como se realizam o tamanho do passo. Dessa forma, apresentaremos uma maneira de como calcular a direção de descida, ou seja, a escolha de v^k exigida no passo 2 do algoritmo (AMKdG).

Pelo que vimos na seção 1.5 um conjunto compacto $C \subseteq \mathbb{R}^m$ foi fundamental para caracterizar o cone polar positivo de K . Assim, para um cone genérico K (convexo, fechado, pontiagudo e com o interior não vazio) em um espaço m -dimensional definiremos o conjunto compacto C que também irá satisfazer as condições (1) e (2) da caracterização. Além disso, definiremos uma função auxiliar neste compacto fundamental para a escolha da direção K -descida.

Seja $C = \{w \in K^* / \|w\| = 1\}$.

O conjunto C cumpre as condições (1) e (2) da caracterização

Seja $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$, dada por

$$\varphi(y) := \sup_{w \in C} \langle y, w \rangle \quad (2.3)$$

Por (1.46), (1.47) e pela compacidade de C , temos

$$-K = \{y \in \mathbb{R}^m / \varphi(y) \leq 0\}, \quad (2.4)$$

$$-\text{int}K = \{y \in \mathbb{R}^m / \varphi(y) < 0\}. \quad (2.5)$$

A seguir o lema estabelece algumas propriedades fundamentais da função φ que será utilizada na seqüência.

Lema 2.2. 1. *Sejam $y, y' \in \mathbb{R}^m$, então*

$$\varphi(y + y') \leq \varphi(y) + \varphi(y')$$

e ainda

$$\varphi(y) - \varphi(y') \leq \varphi(y - y').$$

2. *Sejam $y, y' \in \mathbb{R}^m$, se $y \prec_K y'$ ($y \preceq_K y'$), então*

$$\varphi(y) < \varphi(y') \quad (\varphi(y) \leq \varphi(y')).$$

3. *A função $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}$ é Lipschitz contínua.*

Demonstração. 1. Pela definição da φ . Dados $y, y' \in \mathbb{R}^m$, temos

$$\begin{aligned} \varphi(y + y') &= \sup_{w \in C} \langle y + y', w \rangle \\ &= \sup_{w \in C} (\langle y, w \rangle + \sup_{w \in C} \langle y', w \rangle) \\ &\leq \sup_{w \in C} \langle y, w \rangle + \sup_{w \in C} \langle y', w \rangle \\ &= \varphi(y) + \varphi(y') \end{aligned}$$

Além disso, como $\varphi(\lambda y) = \lambda \varphi(y)$. Em particular, $\lambda = -1 \Rightarrow \varphi(-y) = -\varphi(y)$

logo,

$$\varphi(y + (-y')) \leq \varphi(y) + \varphi(-y')$$

$$\leq \varphi(y) - \varphi(y')$$

Por outro lado,

$$-\varphi(y + (-y')) \geq -(\varphi(y) - \varphi(y'))$$

$$\varphi(-y + y') \geq -\varphi(y) + \varphi(y')$$

$$\varphi(y' - y) \geq \varphi(y') - \varphi(y)$$

2. Sendo $y \prec_K y'$, temos que $y' - y \in \text{int}(K) \Rightarrow (y - y') \in -\text{int}(K)$ e por (i), temos

$$\varphi(y - y') < 0 \Rightarrow \varphi(y) - \varphi(y') < 0 \Rightarrow \varphi(y) < \varphi(y')$$

$$\text{Se } y \preceq_K y' \Leftrightarrow y - y' \in -K \Rightarrow \varphi(y - y') \leq 0 \Rightarrow \varphi(y) - \varphi(y') \leq 0 \Rightarrow \varphi(y) \leq \varphi(y')$$

3. Dados $y, y' \in \mathbb{R}^m$, por (i) tem-se

$$\varphi(y) - \varphi(y') \leq \varphi(y - y')$$

e

$$\varphi(y') - \varphi(y) \leq \varphi(y' - y)$$

Isto é, $\varphi(y - y')$, $\varphi(y' - y)$ são cotas superiores, logo

$$|\varphi(y) - \varphi(y')| \leq \sup\{\varphi(y - y'), \varphi(y' - y)\}$$

Suponha que o supremo seja atingido em $\varphi(y' - y)$, temos

$$|\varphi(y) - \varphi(y')| \leq \varphi(y' - y) = \sup_{w \in C} \langle y' - y, w \rangle$$

$$\leq \sup_{w \in C} |\langle y' - y, w \rangle| \leq \sup_{w \in C} \|y' - y\| \cdot \|w\| = \|y' - y\| \sup_{w \in C} \|w\|$$

Donde,

$$|\varphi(y) - \varphi(y')| \leq L \|y - y'\|$$

onde $L = \sup_{w \in C} \|w\|$ é a constante de Lipschitz.

□

Definamos agora a função $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$, dada por

$$f_x(v) := \varphi(JF(x)v) = \sup_{w \in C} \langle w, JF(x)v \rangle \quad (2.6)$$

Observação 2.2.1 Note que $v \in \mathbb{R}^n$ é uma K -direção de descida em $x \in \mathbb{R}^n$, se e somente se, $f_x(v) < 0$. De fato,

(\Rightarrow) Seja $v \in \mathbb{R}^n$ uma K -direção de descida em $x \in \mathbb{R}^n$, logo $JF(x)v \in -\text{int}K$ por (2.5) e (2.6), temos

$$f_x(v) := \varphi(JF(x)v) < 0 \Rightarrow f_x(v) < 0 \quad (2.7)$$

(\Leftarrow) Sendo $f_x(v) < 0 \Rightarrow \varphi(JF(x)v) < 0 \Rightarrow JF(x)v \in -\text{int}K \Rightarrow JF(x)v \prec_K 0$

Portanto, por (1.52) $v \in \mathbb{R}^n$ uma K -direção de descida em $x \in \mathbb{R}^n$. □

É claro que $x \in \mathbb{R}^n$ é K -crítico, se e somente se, $f_x(v) \geq 0 \forall v \in \mathbb{R}^n$.

Observação 2.2.2. A função definida em (2.6) é convexa. De fato,

dados $v, \bar{v} \in \mathbb{R}^n$ e $t \in [0, 1]$, tem-se que

$$\begin{aligned} f_x((1-t)v + t\bar{v}) &= \sup_{w \in C} \langle w, JF(x)((1-t)v + t\bar{v}) \rangle \\ &= \sup_{w \in C} (\langle w, JF(x)(1-t)v \rangle + \langle w, JF(x)t\bar{v} \rangle) \\ &\leq \sup_{w \in C} \langle w, JF(x)(1-t)v \rangle + \sup_{w \in C} \langle w, JF(x)t\bar{v} \rangle \\ &\leq (1-t) \sup_{w \in C} \langle w, JF(x)v \rangle + t \sup_{w \in C} \langle w, JF(x)\bar{v} \rangle \\ &= (1-t)f_x(v) + tf_x(\bar{v}). \end{aligned}$$

A seguir a definição de função homogênea é importante para continuidade do trabalho.

Definição 30. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Dizemos que f é **homogênea de grau n** quando

$$f(\lambda x) = \lambda^n f(x) \quad \forall \lambda \in \mathbb{R}.$$

Quando $\lambda \in \mathbb{R}_+$, dizemos que f é **homogênea positiva** .

Observação 2.2.3. A função definida em (2.6) é um exemplo de função homogênea de grau 1. A prova segue os passos da observação 2.2, com as devidas adaptações.

2.3 Método K-descida

Nesta seção, estenderemos a noção de direção de descida para o caso vetorial onde o cone K satisfaz as condições estabelecidas em (2.1) e apresentamos o método de K-descida (K-steepest descent method). Para isso, necessitamos das seguintes definições e resultados, a seguir.

Definição 31. Dado $x \in \mathbb{R}^n$, a K-direção de descida para F em x , denotada por v_x é a solução do problema

$$\begin{cases} \min f_x(v) + \frac{1}{2}\|v\|^2 \\ \text{s.a } v \in \mathbb{R}^n \end{cases} \quad (2.8)$$

e α_x o valor ótimo do problema.

Observação 2.3.1. Como $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$ dado por $f_x(v) = \sup_{w \in C} \langle w, JF(x)v \rangle$ é uma função real, convexa e fechada. Temos que α_x e v_x estão bem definidos. Além disso, como F é continuamente diferenciável e φ é Lipschitziana contínua, a aplicação $(x, v) \mapsto f_x(v)$ também é contínua.

Lema 2.3.

1. Se x é K-crítico então $v_x = 0$ e $\alpha_x = 0$
2. Se x não é K-crítico então $v_x \neq 0$ e $\alpha_x < 0$ $f_x(v_x) < -\frac{1}{2}\|v_x\| < 0$
e v_x é uma K-direção de descida.
3. As aplicações $x \mapsto v_x$, $x \mapsto \alpha_x$ são contínuas.

Demonstração. 1 Sabemos que x é K-crítico $\Leftrightarrow f_x(v) \geq 0 \quad \forall v \in \mathbb{R}^n$. Como $f_x(v) := \sup_{w \in C} \langle w, JF(x)v \rangle$. Tomamos $v = 0$ e obtemos $f_x(0) = 0$. Pela definição de v_x segue

que $f_x(v_x) + \frac{1}{2}\|v_x\|^2 \leq f_x(0) + \frac{1}{2}\|0\|^2 = 0 \Rightarrow f_x(v_x) + \frac{1}{2}\|v_x\|^2 \leq 0 \Rightarrow \frac{1}{2}\|v_x\|^2 \leq f_x(v_x) + \frac{1}{2}\|v_x\|^2 \leq 0 \Rightarrow \|v_x\|^2 = 0 \Leftrightarrow v_x = 0$ e como $\alpha_x = \min\{f_x(v_x) + \frac{1}{2}\|v_x\|^2\} \Rightarrow \alpha_x = 0$

□

Demonstração. 2 Se x não é K-crítico, logo existe uma direção $v \in \mathbb{R}^n$ tal que ($v \neq 0$) $f_x(v) < 0$. Tomando $\tilde{t} = -\frac{f_x(v)}{\|v\|^2}$, $\tilde{v} = \tilde{t}v$, obtemos,

$$\begin{aligned} f_x(\tilde{v}) + \frac{1}{2}\|\tilde{v}\|^2 &= f_x(\tilde{t}v) + \frac{1}{2}\|\tilde{t}v\|^2 = \tilde{t}f_x(v) + \frac{1}{2}\tilde{t}^2\|v\|^2 \\ &= -\frac{f_x(v)}{\|v\|^2}f_x(v) + \frac{1}{2}\left(-\frac{f_x(v)}{\|v\|^2}\right)^2\|v\|^2 = -\frac{[f_x(v)]^2}{\|v\|^2} + \frac{1}{2}\frac{[f_x(v)]^2}{\|v\|^2} \\ &= \left(-1 + \frac{1}{2}\right)\frac{[f_x(v)]^2}{\|v\|^2} = -\frac{1}{2}\frac{[f_x(v)]^2}{\|v\|^2} < 0 \Rightarrow \\ \alpha_x = \min\{f_x(v_x) + \frac{1}{2}\|v_x\|^2\} &\leq f_x(\tilde{v}) + \frac{1}{2}\|\tilde{v}\|^2 < 0 \Rightarrow \alpha_x < 0, v_x \neq 0 \text{ e} \\ f_x(v_x) &< -\frac{1}{2}\|v_x\|^2 < 0 \end{aligned}$$

□

Demonstração. 3 tome $x^0 \in \mathbb{R}^n$ e $\epsilon > 0$. Defina $S = \{v \in \mathbb{R}^n \mid \|v_{x^0} - v\| = \epsilon\}$, onde v_{x^0} é a solução do problema (2.8) com $x = x^0$. Pondo $g_x(v) = f_x(v) + \frac{1}{2}\|v\|^2$, dessa forma temos que g_x é fortemente convexa com módulo $\gamma = \frac{1}{2}$. Logo

$$\begin{aligned} g_{x^0}(v_{x^0} + t(v - v_{x^0})) &\leq g_{x^0}(v_{x^0}) + t(g_x(v) - g_x(v_{x^0})) - \frac{1}{2}t(1-t)\|v_{x^0} - v\|^2 \\ g_{x^0}(v_{x^0} + t(v - v_{x^0})) - g_{x^0}(v_{x^0}) &\leq t(g_x(v) - g_x(v_{x^0})) - \frac{1}{2}t(1-t)\|v_{x^0} - v\|^2 \end{aligned} \quad (2.9)$$

$$0 \leq \frac{g_{x^0}(v_{x^0} + t(v - v_{x^0})) - g_{x^0}(v_{x^0})}{t} \leq g_x(v) - g_x(v_{x^0}) - \frac{1}{2}(1-t)\|v_{x^0} - v\|^2 \quad (2.10)$$

(passando o limite quando $t \rightarrow 0^+$) obtemos,

$$0 \leq g_x(v) - g_x(v_{x^0}) - \frac{1}{2}\|v_{x^0} - v\|^2 \quad (2.11)$$

logo,

$$f_{x^0}(v_{x^0}) + \frac{1}{2}\|v_{x^0}\|^2 + \frac{1}{2}\epsilon^2 \leq f_{x^0}(v) + \frac{1}{2}\|v\|^2, \forall v \in S. \quad (2.12)$$

Como a aplicação $(x, v) \mapsto f_x(v)$ é contínua e S é compacto, logo

$$\exists \delta > 0 \text{ talque } \|x - x^0\| \leq \delta \Rightarrow f_x(v_{x^0}) + \frac{1}{2}\|v_{x^0}\|^2 < f_x(v) + \frac{1}{2}\|v\|^2 \quad \forall v \in S \quad (2.13)$$

Tomemos agora $x \in \mathbb{R}^n$, $\|x - x^0\| \leq \delta$. Da convexidade da função $v \mapsto f_x(v) + \frac{1}{2}\|v\|^2$

$$\begin{aligned} f_x(v_x) + \frac{1}{2}\|v_x\|^2 &\leq f_x(v) + \frac{1}{2}\|v\|^2 \quad \forall v \in \mathbb{R}^n. \\ f_x(v_x) + \frac{1}{2}\|v_x\|^2 &\leq f_x(v_{x^0}) + \frac{1}{2}\|v_{x^0}\|^2 \end{aligned} \quad (2.14)$$

Assim, pela desigualdade (2.14), concluímos que v_x , o minimizador de $f_x(\cdot) + \frac{1}{2}\|\cdot\|^2$ não pode pertencer a região $\|v - v_{x^0}\| \geq \epsilon$. Significa que temos que analisar dois casos, a saber

$$\|v_x - v_{x^0}\| \leq \epsilon \text{ ou } \|v_x - v_{x^0}\| \geq \epsilon. \text{ Se } \|v_x - v_{x^0}\| > \epsilon \quad (2.15)$$

Da desigualdade (2.12) fazendo $v = v_x$ obtemos,

$$f_x(v_{x^0}) + \frac{1}{2}\|v_{x^0}\|^2 \leq f_x(v_x) + \frac{1}{2}\|v_x\|^2 - \frac{1}{2}\|v_x - v_{x^0}\|^2 \quad (2.16)$$

Donde,

$$\frac{1}{2}\epsilon^2 < \frac{1}{2}\|v_x - v_{x^0}\|^2 \leq 0. \quad (2.17)$$

O que é uma contradição, logo

$$\|x - x^0\| \leq \delta \Rightarrow \|v_x - v_{x^0}\| < \epsilon. \quad (2.18)$$

Portanto, v_x é contínua e como consequência α_x também é contínua.

□

Uma possível escolha para v^k no AMKdG é v_{x^k} , ou seja, a K-direção de descida em x^k . Uma vez que o cálculo de v_x requer a solução de

$$\begin{cases} \min f_x(v) + \frac{1}{2}\|v\|^2 \\ \text{s.a } v \in \mathbb{R}^n \end{cases} \quad (2.19)$$

Seria interessante trabalhar com a solução aproximada deste problema.

Definição 32. Seja $\sigma \in [0, 1)$. Dizemos que v é uma K-direção de descida σ -aproximada em $x \in \mathbb{R}^n$ se $f_x(v) + \frac{1}{2}\|v\|^2 \leq (1 - \sigma)\alpha_x$ ou, equivalente,

$$f_x(v) + \frac{1}{2}\|v\|^2 - (f_x(v_x) + \frac{1}{2}\|v_x\|^2) \leq \sigma|\alpha_x|.$$

Observação 2.3.2. Observa-se que a K-direção de descida exata em x é sempre uma K-direção de descida σ -aproximada, porque assumimos $\sigma \in [0, 1)$. A K-direção de descida exata em x é a única K-direção de descida σ -aproximada. Basta tomar

$$\sigma = 0.$$

Lema 2.4. *Seja $\sigma \in [0, 1)$. Se v é uma K-direção de descida σ -aproximada em x , então*

$$\|v_x - v\|^2 \leq 2\sigma|\alpha_x|$$

.

Demonstração. Defina $g_x(v) = f_x(v) + \frac{1}{2}\|v\|^2$ e sendo g_x fortemente convexa com módulo $\gamma = \frac{1}{2}$. Assim,

dados $v_x, v \in \mathbb{R}^n$ e $t \in [0, 1)$ onde v_x é o minimizador de g_x , segue que

$$\begin{aligned} g_x(v_x + t(v - v_x)) &\leq g_x(v_x) + t(g_x(v) - g_x(v_x)) - \frac{1}{2}t(1-t)\|v_x - v\|^2 \\ 0 \leq g_x(v_x + t(v - v_x)) - g_x(v_x) &\leq t(g_x(v) - g_x(v_x)) - \frac{1}{2}t(1-t)\|v_x - v\|^2 \\ 0 \leq \frac{g_x(v_x + t(v - v_x)) - g_x(v_x)}{t} &\leq g_x(v) - g_x(v_x) - \frac{1}{2}(1-t)\|v_x - v\|^2 \end{aligned} \quad (2.20)$$

(passando o limite quando $t \rightarrow 0^+$) obtemos,

$$\begin{aligned}
 g_x(v_x) + \frac{1}{2}\|v_x - v\|^2 &\leq g_x(v) \\
 g_x(v_x) + \frac{1}{2}\|v_x - v\|^2 &\leq g_x(v) \\
 f_x(v_x) + \frac{1}{2}\|v_x\|^2 + \frac{1}{2}\|v_x - v\|^2 &\leq f_x(v) + \frac{1}{2}\|v\|^2
 \end{aligned}
 \tag{2.21}$$

Pela hipótese v é uma K -direção de descida σ -aproximada em x , logo

$$\frac{1}{2}\|v_x - v\|^2 \leq f_x(v) + \frac{1}{2}\|v\|^2 - (f_x(v_x) + \frac{1}{2}\|v_x\|^2) \leq \sigma|\alpha_x|
 \tag{2.22}$$

Portanto,

$$\|v_x - v\|^2 \leq 2\sigma|\alpha_x|
 \tag{2.23}$$

□

Observação 2.3.3. Note que $v = 0$ é uma K -direção de descida σ -aproximada em x , se e somente se, x é K -crítico. De fato,

(\Rightarrow) seja $v = 0$ é uma K -direção de descida σ -aproximada em x , logo

$$\|v_x - 0\|^2 \leq 2\sigma|\alpha_x| \quad \forall \sigma \in [0, 1)
 \tag{2.24}$$

Em particular, tomando $\sigma = 0$ obtemos,

$$\|v_x\|^2 \leq 2 \cdot 0 \cdot |\alpha_x| = 0 \Rightarrow \|v_x\|^2 = 0 \Rightarrow v_x = 0$$

Sendo que, $\alpha_x = f_x(v_x) + \frac{1}{2}\|v_x\|^2 = f_x(0) + \frac{1}{2}\|0\|^2 = 0 \Rightarrow \alpha_x = 0$

Logo, pelo lema(2.3) x é K -crítico.

(\Leftarrow) Seja $v \in \mathbb{R}^n$ uma K -direção de descida σ -aproximada em x . Pelo lema (2.4), temos

$$\|v_x - v\|^2 \leq 2\sigma|\alpha_x|$$

Por hipótese x é K -crítico, logo pelo lema (2.3), temos que $v_x = 0$ e $\alpha_x = 0$. Assim, tem-se

$$\|0 - v\|^2 \leq 2\sigma|0| = 0 \quad \forall \sigma \in [0, 1).$$

Concluimos que,

$$\|v\|^2 = 0 \Rightarrow v = 0$$

Portanto, $v = 0$ é uma K -direção de descida σ -aproximada em x .

Observação 2.3.4. Se x não é K -crítico, então pelo lema (2.3) temos que $v_x \neq 0$, $\alpha_x < 0$ e como v é uma K -direção de descida σ -aproximada em x , logo

$$f_x(v) < f_x(v) + \frac{1}{2}\|v\|^2 \leq (1 - \sigma)|\alpha_x| < 0 \quad (2.25)$$

O que resulta em $f_x(v) < 0$ e por (2.6) segue que v é uma K -direção de descida onde $v \neq 0$.

A seguir formalizaremos o método K -descida com a regra tipo K -Armijo, implementando o processo de retrocesso *backtracking*. Trata-se na verdade das adaptações do AMKdG.

2.4 Adaptações do Algoritmo(AMKdG)

1. Passo 1. Tomar $\beta \in (0, 1), \sigma \in [0, 1)$ $x^0 \in \mathbb{R}^n, k := 0$
2. Passo 2. Se x^k for K -crítico($f_{x^k}(v) \geq 0 \forall v \in \mathbb{R}^n$) Pare! Caso contrário

3. Passo 3 Calcule v^k , uma K -direção de descida σ -aproximada em x^k .
4. Passo 4. Calcule o comprimento do passo $t_k \in (0, 1]$ da seguinte maneira:

$$t_k := \max\{t = \frac{1}{2^j}/j \in \mathbb{N}, \varphi(F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j} JF(x^k)v^k) \leq 0\}$$

5. Passo 5. Fazer $x^{k+1} := x^k + t_k v^k$, $k := k + 1$ e volte passo 2.

Com relação ao algoritmo AMKdG adaptado, vemos que se x^k não for K -crítico, v^k obtida no passo 3 é uma K -direção de descida por (2.6) e o cálculo para o tamanho do passo t_k está bem definido, isto é, o conjunto dos t_k é não vazio. Além disso, cada t_k pode ser obtido por um procedimento de retrocesso, conhecido por *backtracking* tal como discutido anteriormente.

Concluimos nossas análises a respeito do AMKdG adaptado, notando que em virtude de (2.5) e pelo passo 4, obtemos

$$t_k := \max\{t = \frac{1}{2^j}/j \in \mathbb{N}, F(x^k + 2^{-j}v^k) \preceq_K F(x^k) + \beta 2^{-j} JF(x^k)v^k\} \quad (2.26)$$

Além disso, x^{k+1} satisfaz a desigualdade

$$F(x^{k+1}) \preceq_K F(x^k).$$

Assim, obtemos uma sequência $\{F(x^k)\}$ K -não crescente. Com efeito,

$$t_k := \max\{t = \frac{1}{2^j}/j \in \mathbb{N}, \varphi(F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j} JF(x^k)v^k) \leq 0\}$$

$$\varphi(F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j} JF(x^k)v^k) \leq 0$$

$$F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j} JF(x^k)v^k \in -K$$

$$-(F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j} JF(x^k)v^k) \in K$$

$$F(x^k) + \beta 2^{-j} JF(x^k)v^k - F(x^k + 2^{-j}v^k) \in K$$

$$F(x^k + 2^{-j}v^k) \preceq_K F(x^k) + \beta 2^{-j} JF(x^k)v^k$$

O próximo resultado é uma generalização da proposição (Armijo) que é uma simples consequência do fato de que F é continuamente diferenciável.

Proposição 2.5. *Sejam $\beta \in (0, 1)$, x e v tais que $JF(x)v \prec_K 0$. Então existem $\hat{t}, \delta, \delta' > 0$ tais que:*

1. v' é uma K -direção de descida em x' .
2. $F(x' + tv') \prec_K F(x') + \beta t JF(x')v'$, para todo $t \in (0, \hat{t})$, $x' \in B(x; \delta)$, $v' \in B(v; \delta')$.

Demonstração. Temos por hipótese que $JF(x)v \prec_K 0 \Leftrightarrow -JF(x)v \in \text{int}(K) \Leftrightarrow JF(x)v \in -\text{int}(K)$. Assim, $\exists \epsilon > 0$ tal que $JF(x)v + y \in -\text{int}(K) \forall y \in \mathbb{R}^m, \|y\| \leq \epsilon$.

Pela continuidade da jacobiana, existem $\delta, \delta' > 0$ tais que,

$\|x' - x\| \leq \delta_1, \|v' - v\| \leq \delta_2 \Rightarrow \|JF(x')v' - JF(x)v\| \leq \frac{\epsilon}{2} \Rightarrow JF(x')v' \in -\text{int}(K) \Rightarrow JF(x')v' \prec_K 0$. Logo, v' é uma K -direção de descida em x' e pela continuidade diferenciável de F , temos;

$$F(z + tu) = F(z) + tJF(z)u + tR(z, tu) \text{ com } \lim_{t \rightarrow 0} \|R(z, tu)\| = 0$$

uniformemente para z e u num compacto. Portanto, existe $\hat{t} > 0$ tal que, para $t \in (0, \hat{t})$, $\|x' - x\| \leq \delta_1, \|v' - v\| \leq \delta_2 \Rightarrow \|R(x', tv')\| \leq \frac{\epsilon}{2}(1 - \beta)$.

Agora, assumimos que $t \in (0, \hat{t}), \|x' - x\| \leq \delta_1, \|v' - v\| \leq \delta_2$, então,

$$\begin{aligned} F(x' + tv') &= F(x') + tJF(x')v' + tR(x', tv') \\ F(x' + tv') &= F(x') + tJF(x')v' + tR(x', tv') + t\beta JF(x')v' - t\beta JF(x')v' \\ &= F(x') + t\beta JF(x')v' + t(1 - \beta)JF(x')v' + tR(x', tv') \\ &= F(x') + t\beta JF(x')v' + t[(1 - \beta)JF(x')v' + R(x', tv')] \end{aligned}$$

Definindo

$$u := JF(x')v' + \frac{R(x', tv')}{(1 - \beta)}$$

temos que,

$$F(x' + tv') = F(x') + t\beta JF(x')v' + t[(1 - \beta)]u$$

Basta ver que $u \prec_K 0$. Com efeito,

pondo $u = JF(x)v + \tilde{y}$, onde,

$$\tilde{y} = JF(x')v' - JF(x)v + \frac{R(x', tv')}{(1 - \beta)}$$

$$\begin{aligned} \|\tilde{y}\| &= \|JF(x')v' - JF(x)v + \frac{R(x', tv')}{(1 - \beta)}\| \\ &\leq \|JF(x')v' - JF(x)v\| + \frac{1}{(1 - \beta)}\|R(x', tv')\| \\ &\leq \frac{\epsilon}{2} + \frac{1}{(1 - \beta)}\frac{\epsilon}{2}(1 - \beta) = \epsilon \Rightarrow \|\tilde{y}\| \leq \epsilon \end{aligned}$$

logo

$$u = JF(x)v + \tilde{y} \in -\text{int}(\mathbf{K}) \Rightarrow u \prec_K 0$$

Assim,

$$\begin{aligned} -t(1 - \beta)u &= F(x') - F(x') + t\beta JF(x')v' - t\beta JF(x')v' - t(1 - \beta)u \\ &= F(x') + t\beta JF(x')v' - (F(x') + t\beta JF(x')v' + t(1 - \beta)u) \in \text{int}(\mathbf{K}) \Rightarrow \\ &F(x') + t\beta JF(x')v' + t(1 - \beta)u \prec_K F(x') + t\beta JF(x')v'. \end{aligned}$$

Por outro lado, como

$$F(x' + tv') = F(x') + t\beta JF(x')v' + t(1 - \beta)u \prec_K F(x') + t\beta JF(x')v' \Rightarrow$$

$$F(x' + tv') \prec_K F(x') + t\beta JF(x')v'$$

2.5 Análise de Convergência:Caso Geral

A partir de agora, $\{x^k\}$, $\{v^k\}$ e $\{t_k\}$ são seqüências geradas pelo algoritmo (AMKdG) adaptado. Se o algoritmo termina depois de um número finito de iterações, ele termina em um ponto K-crítico. Nesta seção, supomos que a seqüência gerada pelo AMKdG adaptado seja infinita. Assim, em vista do lema (2.3), definição (32) e a proposição (2.5), para todo $k \in \mathbb{N}$, temos

1. $\alpha_{x^k} < 0$
2. $f_{x^k}(v^k) + \frac{1}{2}\|v^k\|^2 \leq (1 - \sigma)\alpha_{x^k} < 0$
3. $F(x^{k+1}) \preceq_K F(x^k) + \beta t_k JF(x^k)v^k \preceq_K F(x^k)$.

Em Particular a seqüência $\{F(x^k)\}$ é K-decrescente. Usando a K-desigualdade acima, lema (2.2) e a homogeneidade de φ , obtemos:

$$\begin{aligned}
 \varphi(F(x^{k+1})) &\preceq_K \varphi(F(x^k) + \beta t_k JF(x^k)v^k) \\
 &\preceq_K \varphi(F(x^k)) + \varphi(\beta t_k JF(x^k)v^k) \\
 &= \varphi(F(x^k)) + \beta t_k \varphi(JF(x^k)v^k) \\
 &\preceq_K \varphi(F(x^k)) + \beta t_k f_{x^k}(v^k) \\
 &\preceq_K \varphi(F(x^k)) + \beta t_k ((1 - \sigma)\alpha_{x^k} - \frac{1}{2}\|v^k\|^2). \quad (2.27)
 \end{aligned}$$

Como conseqüência deste resultado obtemos o seguinte lema.

Lema 2.6. *Se $\{F(x^k)\}$ é K-limitada inferiormente (isto é, se existe \bar{y} tal que $\bar{y} \preceq_K F(x^k)$ para todo $k \in \mathbb{N}$). Então,*

$$\sum_{k=0}^{\infty} t_k |\alpha_{x^k}| < \infty, \quad \sum_{k=0}^{\infty} t_k \|v^k\|^2 < \infty$$

Demonstração. Pela K-desigualdade (2.27), temos,

$$\varphi(F(x^{k+1})) \leq \varphi(F(x^k)) + \beta t_k [(1 - \sigma)\alpha_{x^k} - \frac{1}{2}\|v^k\|^2]$$

para $k = 0$,

$$\varphi(F(x^1)) \leq \varphi(F(x^0)) + \beta t_0 [(1 - \sigma)\alpha_{x^0} - \frac{1}{2}\|v^0\|^2]$$

para $k = 1$,

$$\varphi(F(x^2)) \leq \varphi(F(x^1)) + \beta t_1 [(1 - \sigma)\alpha_{x^1} - \frac{1}{2}\|v^1\|^2]$$

⋮

para $k = k + 1$, temos

$$\varphi(F(x^{k+1})) \leq \varphi(F(x^k)) + \beta t_k [(1 - \sigma)\alpha_{x^k} - \frac{1}{2}\|v^k\|^2]$$

Comparando as $K+1$ - desigualdades obtemos,

$$\begin{aligned} \varphi(F(x^{k+1})) &\leq \varphi(F(x^0)) + \beta t_0 [(1 - \sigma)\alpha_{x^0} - \frac{1}{2}\|v^0\|^2] + \beta t_1 [(1 - \sigma)\alpha_{x^1} - \frac{1}{2}\|v^1\|^2] + \cdots + \\ &\beta t_k [(1 - \sigma)\alpha_{x^k} - \frac{1}{2}\|v^k\|^2] = \varphi(F(x^0)) + \sum_{k=0}^n \beta t_k [(1 - \sigma)\alpha_k - \frac{1}{2}\|v^k\|^2] \end{aligned}$$

$$= \varphi(F(x^0)) - \sum_{k=0}^n \beta t_k ((1 - \sigma)(-\alpha_k) + \frac{1}{2}\|v^k\|^2) = \varphi(F(x^0)) - \sum_{k=0}^n \beta t_k ((1 - \sigma)|\alpha_k| + \frac{1}{2}\|v^k\|^2)$$

Como $\{F(x^k)\}$ é K -limitada inferiormente, logo existe \bar{y} tal que $\bar{y} \preceq_K F(x^k)$ para todo $k \in \mathbb{N} \Rightarrow \varphi(\bar{y}) \leq \varphi(F(x^k)) \Rightarrow \varphi(\bar{y}) \leq \varphi(F(x^0)) - \sum_{k=0}^n \beta t_k ((1 - \sigma)|\alpha_k| + \frac{1}{2}\|v^k\|^2)$

$$\leq \varphi(F(x^0)) - \sum_{k=0}^n \beta t_k (1 - \sigma)|\alpha_k| - \sum_{k=0}^n \beta t_k \frac{1}{2}\|v^k\|^2$$

$$\Rightarrow \varphi(\bar{y}) - \varphi(F(x^0)) \leq - \sum_{k=0}^n \beta t_k (1 - \sigma)|\alpha_k| - \sum_{k=0}^n \beta t_k \frac{1}{2}\|v^k\|^2$$

$$\Rightarrow \varphi(F(x^0)) - \varphi(\bar{y}) \geq \sum_{k=0}^n \beta t_k (1 - \sigma)|\alpha_k| + \sum_{k=0}^n \beta t_k \frac{1}{2}\|v^k\|^2 \text{ (Fazendo } n \rightarrow \infty \text{) obtemos,}$$

$$\varphi(F(x^0)) - \varphi(\bar{y}) \geq \sum_{k=0}^{\infty} \beta t_k (1 - \sigma)|\alpha_k| + \sum_{k=0}^{\infty} \beta t_k \frac{1}{2}\|v^k\|^2$$

O que prova o enunciado. \square

Agora estamos em condições de estudar as propriedades de convergência do Algoritmo (AMKdG).

Teorema 2.7. *Todo ponto de acumulação da seqüência $\{x^k\}$ é K -crítico.*

Demonstração. Seja \bar{x} um ponto de acumulação de $\{x^k\}_k$, isto é, existe uma subsequência $\{x^{k_j}\}_j \subset \{x^k\}_k$ tal que $\lim_{j \rightarrow +\infty} x^{k_j} = \bar{x}$. Pelo lema 2.3 item 3 segue que as seqüências $\{v_{x^{k_j}}\}$ e $\{\alpha_{x^{k_j}}\}$ são limitadas, pois $\lim_{j \rightarrow +\infty} v_{x^{k_j}} = v_{\bar{x}}$ e $\lim_{j \rightarrow +\infty} \alpha_{x^{k_j}} = \alpha_{\bar{x}}$, respectivamente.

Logo, $\|v_{x^{k_j}}\| \leq c_1$ e $|\alpha_{x^{k_j}}| \leq c_2$, onde $c_1, c_2 \in \mathbb{R}_+$. Pelo lema 2.5

$$\begin{aligned} \|v^{k_j}\| - \|v_{x^{k_j}}\| &\leq \|v^{k_j} - v_{x^{k_j}}\| \leq \sqrt{2\sigma|\alpha_{x^{k_j}}|} \\ \|v^{k_j}\| &\leq \sqrt{2\sigma|\alpha_{x^{k_j}}|} + \|v_{x^{k_j}}\| \leq \sqrt{2\sigma c_1} + c_2 \\ \|v^{k_j}\| &\leq c, \text{ onde } c = \sqrt{2\sigma c_1} + c_2. \end{aligned}$$

Logo $\{v^{k_j}\}$ é limitada. (passando a uma subseqüência se necessário). Assumimos,

$$\lim_{j \rightarrow +\infty} v^{k_j} = \bar{v}.$$

Como $f_{x^k}(v^k) + \frac{1}{2}\|v^k\|^2 \leq (1 - \sigma)\alpha_{x^k} \forall k \in \mathbb{N}$. Seja $\mathbb{N}_1 \subseteq \mathbb{N}$ um subconjunto infinito de índices. Assim para $k \in \mathbb{N} \exists k_j \in \mathbb{N}_1$ tal que para $k_j > k$ tem-se

$$f_{x^{k_j}}(v^{k_j}) + \frac{1}{2}\|v^{k_j}\|^2 \leq (1 - \sigma)\alpha_{x^{k_j}}.$$

Fazendo $j \rightarrow \infty$, temos

$$f_{\bar{x}}(\bar{v}) + \frac{1}{2}\|\bar{v}\|^2 \leq (1 - \sigma)\alpha_{\bar{x}}. \quad (2.28)$$

Pela proposição 2.1, $F(\bar{x}) \preceq_K F(x^k) \forall k \in \mathbb{N}$, ou seja, a seqüência $\{F(x^k)\}$ é K-limitada inferiormente, logo pelo lema 2.6

$$\sum_{k=0}^{\infty} t_k |\alpha_{x^k}| < \infty, \quad \sum_{k=0}^{\infty} t_k \|v^k\|^2 < \infty \Rightarrow \lim_{k \rightarrow +\infty} t_k |\alpha_{x^k}| = 0 \text{ e } \lim_{k \rightarrow +\infty} t_k \|v^k\| = 0$$

Passando a subseqüência, obtemos,

$$\lim_{j \rightarrow +\infty} t_{k_j} |\alpha_{x^{k_j}}| = 0 \quad (2.29)$$

e

$$\lim_{j \rightarrow +\infty} t_{k_j} \|v^{k_j}\| = 0 \quad (2.30)$$

Afirmamos que $\bar{v} = 0$. Com efeito, se fosse $\bar{v} \neq 0$ e como $\alpha_{\bar{x}} < 0$, usando (2.28)

$f_{\bar{x}}(\bar{v}) \leq f_{\bar{x}}(\bar{v}) + \frac{1}{2}\|\bar{v}\|^2 \leq (1-\sigma)\alpha_{\bar{x}} < 0 \Rightarrow f_{\bar{x}}(\bar{v}) < 0 \Rightarrow JF(\bar{x})\bar{v} \prec_K 0$ e pela proposição 2.5. Concluimos que existe $\hat{t} > 0$ tal que para j suficientemente grande (maior do que algum j_0).

$$F(x^{k_j} + tv^{k_j}) \preceq_K F(x^{k_j}) + \beta t JF(x^{k_j})v^{k_j} \quad \forall t \in (0, \hat{t}). \quad (2.31)$$

Agora vamos mostrar que para j maior que tal j_0 ,

$$2t_{k_j} \geq \min\{1, \hat{t}\}. \quad (2.32)$$

De fato, seja $j > j_0$. Se $t_{k_j} = 1$ é óbvio.

Se $t_{k_j} < 1$, então este tamanho de passo foi obtido por um procedimento *backtracking*, onde o tamanho de passo anterior possível $2t_{k_j}$, não pertence ao conjunto (2.26), pois t_{k_j} é o máximo dos comprimentos de passos que satisfaz a condição tipo Armijo. Por (2.31) concluímos que $2t_{k_j} \geq \hat{t}$.

Segue que da hipótese $\bar{v} \neq 0$ e por 2.30, tem-se $\lim_{j \rightarrow +\infty} t_{k_j} = 0$. Mas isso contradiz (2.32), logo $\bar{v} = 0$. Usando (2.28) segue que $(1-\sigma)\alpha_{\bar{x}} \geq 0 \Rightarrow \alpha_{\bar{x}} \geq 0$. Por outro lado, $\alpha_{x^k} < 0$, pois x^k não é K-crítico. Passando a subsequência (neste caso) $\alpha_{x^{k_j}} < 0 \Rightarrow \lim_{j \rightarrow +\infty} \alpha_{x^{k_j}} = \alpha_{\bar{x}} \leq 0$, logo $\alpha_{\bar{x}} = 0$. Portanto, \bar{x} é K-crítico.

□

Capítulo 3

Aplicação do Algoritmo (AMKdG) ao caso $K = \mathbb{R}_+^m$

Neste capítulo, aplicaremos o algoritmo (AMKdG), estudado no capítulo anterior, no caso quando o cone K considerado é o ortante não negativo do \mathbb{R}^m . Uma vez que fizemos esta consideração, a comparação entre seus vetores se reduz à definição (16). Assim, a solução para o problema definido em (1.31) é a de encontrar solução Pareto ótima, isto é, um ponto $z \in \mathbb{R}^n$ tal que não existe outro ponto $y \in \mathbb{R}^n$ com $F(y) \leq F(z)$ e $F(y) \neq F(z)$.

A seguir uma condição necessária para que um ponto $x \in \mathbb{R}^n$ seja Pareto ótimo de F .

Teorema 3.1. *Seja F diferenciável em $x \in \mathbb{R}^n$. Uma condição necessária para que x seja Pareto-ótimo é:*

$$\text{Im}(JF(x)) \cap (-\mathbb{R}_{++})^m = \emptyset \quad (3.1)$$

Demonstração. Suponha por absurdo que $\text{Im}(JF(x)) \cap (-\mathbb{R}_{++})^m \neq \emptyset$, isto é, existe uma direção de descida $v \in \mathbb{R}^n$ tal que $JF(x)v \in (-\mathbb{R}_{++})^m \Rightarrow \langle \nabla F_i(x), v \rangle < 0 \quad \forall i = 1, \dots, m$. Assim, v é uma direção de descida para cada F_i a partir de x , ou seja, existe $\bar{\lambda} > 0$ tal que $F_i(x + \lambda v) < F_i(x) \quad \forall i = 1, \dots, m, \quad \forall \lambda \in (0, \bar{\lambda}]$.

Logo,

$F(x + \lambda v) \leq_{\mathbb{R}_+^m} F(x)$ o que é uma contradição, pois supomos que x é pareto. \square

Os pontos que satisfazem a condição (3.1) acima são chamados de *Pareto crítico*.

Observação 3.1. Podemos observar que o teorema acima generaliza o conceito de ponto crítico para otimização mono-objetivo. Basta fazer $m = 1$ e $K = [0, +\infty)$. Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciável e seja $x \in \mathbb{R}^n$ um ponto Pareto-ótimo, tem-se $JF(x) = \nabla F(x)$ e como $\text{-int}K = -\mathbb{R}_{++}$, logo

$$\text{Im}(\nabla F(x)) \cap (-\mathbb{R}_{++}) = \emptyset \quad (3.2)$$

$\forall v \in \mathbb{R}^n, \nabla F(x)v \notin (-\mathbb{R}_{++}) \Rightarrow \langle \nabla F(x), v \rangle \geq 0$. Tomando $v = -\nabla F(x)$ temos que $\langle \nabla F(x), -\nabla F(x) \rangle \geq 0 \Rightarrow \|\nabla F(x)\|^2 \leq 0 \Rightarrow \|\nabla F(x)\|^2 = 0 \Rightarrow \nabla F(x) = 0$.

Observação 3.2. A condição (3.1) é necessária, mas não suficiente para que um ponto $x \in \mathbb{R}^n$ seja Pareto-ótimo de F .

Exemplo 13. Sejam $K = \mathbb{R}_+^2$, $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $F(x, y) = (x^2 - y^2, x^2 - y^3)$. Temos que $(0, 0)$ satisfaz (2.1), mas não é um ponto Pareto-ótimo de F . De fato, a jacobiana de F é

$$JF(x, y) = \begin{pmatrix} 2x & -2y \\ 2x & -3y^2 \end{pmatrix}$$

$$JF(0, 0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Logo, $JF(0, 0)v = 0 \forall v \in \mathbb{R}^2$.

Portanto,

$$\text{Im}JF(0, 0) \cap (-\mathbb{R}_{++}^2) = \emptyset$$

Porém, afirmamos que existe um par $(x, y) \in \mathbb{R}^2$ tal que $F(x, y) \leq_{\mathbb{R}_+^2} F(0, 0)$, ou seja, $F(0, 0) - F(x, y) \in \mathbb{R}_+^2$. Com efeito,

basta tomarmos qualquer solução das desigualdes $x^2 - y^2 \leq 0$ e $x^2 - y^3 \leq 0$. Neste caso $x = 1$ e $y = 2$. Portanto, existe um par $(1, 2) \in \mathbb{R}^2$ tal que $F(1, 2) \leq_{\mathbb{R}_+^2} F(0, 0)$.

3.1 Cálculo da direção de busca

Uma característica do método de descida em otimização é buscar uma direção $v \in \mathbb{R}^n$ e partindo de um ponto x^k ao longo desta direção obter outro ponto x^{k+1} que promova um descenso da função objetivo. Para isso, defina a seguinte função auxiliar: Sejam $x \in \mathbb{R}^n$ fixo (porém arbitrário) e $f_x : \mathbb{R}^n \rightarrow \mathbb{R}^m$ dada por

$$f_x(v) := \max_{i=1,2,\dots,m} \{\langle \nabla F_i(x), v \rangle\}. \quad (3.3)$$

Note que f_x é convexa e positivamente homogênea. De fato, a convexidade é pelo fato de f_x ser o máximo de funções lineares. A homogeneidade, segue de

$$f_x(v) := \max_{i=1,2,\dots,m} \{\langle \nabla F_i(x), v \rangle\} \Rightarrow f_x(\lambda v) := \max_{i=1,2,\dots,m} \{\langle \nabla F_i(x), \lambda v \rangle\}$$

$$f_x(\lambda v) := \max_{i=1,2,\dots,m} \{\lambda \langle \nabla F_i(x), v \rangle\} = \lambda \max_{i=1,2,\dots,m} \{\langle \nabla F_i(x), v \rangle\} \Rightarrow$$

$$f_x(\lambda v) := \lambda f_x(v) \quad (3.4)$$

Observação 3.2.1. Observamos que a função de definida em (3.3) é exatamente a mesma função definida em (2.6). Basta considerarmos $C = \{e_1, e_2, \dots, e_m\}$ uma base canônica do \mathbb{R}^m .

Uma estratégia para cálculo da direção de descida é considerar o seguinte problema de minimização irrestrito auxiliar:

$$\begin{cases} \min f_x(v) + \frac{1}{2}\|v\|^2 \\ \text{s.a } v \in \mathbb{R}^n \end{cases} \quad (3.5)$$

O problema auxiliar (3.5) possui solução(única), pois sua função objetivo é fortemente convexa. Dessa forma, chamaremos $v(x)$ a solução do problema (3.5), ou seja,

$$v(x) = \operatorname{argmin}\{f_x(v) + \frac{1}{2}\|v\|^2\} \quad (3.6)$$

e $\alpha(x)$ o valor ótimo do Problema (3.5), ou seja,

$$\alpha(x) = \min\{f_x(v) + \frac{1}{2}\|v\|^2\}. \quad (3.7)$$

Observe que fazendo $m = 1$ e $K = \mathbb{R}_+$ em (2.1) recuperamos o método de descida da otimização monobjetivo. Neste caso, a escolha natural é tomar a direção oposta do gradiente da função F , ou seja, $v^k = -\nabla F(x^k)$ (máxima descida) que passa a ser a solução do problema irrestrito:

$$\begin{cases} \min \langle v, \nabla F(x^k) \rangle + \frac{1}{2}\|v\|^2 \\ \text{s.a } v \in \mathbb{R}^n \end{cases} \quad (3.8)$$

Lema 3.2. *Sejam $v(x)$ e $\alpha(x)$ a solução e o valor ótimo do Problema (3.5) respectivamente, ou seja*

$$v(x) = \operatorname{argmin}\{f_x(v) + \frac{1}{2}\|v\|^2\}$$

$$\alpha(x) = \min\{f_x(v) + \frac{1}{2}\|v\|^2\}$$

1. *Se x é crítico Pareto, então $v(x) = 0 \in \mathbb{R}^n$ e $\alpha(x) = 0$.*
2. *Se x não é crítico Pareto, então $\alpha(x) < 0$ e*

$$f_x(v(x)) \leq -\frac{1}{2}\|v(x)\|^2 < 0$$

$$(JF(x)v(x))_i \leq f_x(v(x)) \quad i = 1, 2, \dots, n$$

3. *As funções $v(x)$ e $\alpha(x)$ são contínuas.*

Demonstração. 1. Suponha que $v(x) \neq 0$ e como $v(x)$ é solução de (3.5), logo $v(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq f_x(v) + \frac{1}{2}\|v\|^2 \forall v \in \mathbb{R}^n$. Em particular tomando $v = 0 \Rightarrow f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq f_x(0) + \frac{1}{2}\|0\|^2 = 0 \Rightarrow f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq 0 \Rightarrow f_x(v(x)) < 0$, pois $v(x) \neq 0$. Portanto, $JF(x)v(x) \in (-\mathbb{R}_{++})^m \Rightarrow JF(x)v(x) \in \text{Im}JF(x) \cap (-\mathbb{R}_{++})^m = \emptyset \Rightarrow x$ não é crítico Pareto. Agora se $\alpha(x) \neq 0$ e sendo $v(x)$ solução de (3.4). Segue que $\alpha(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq f_x(v) + \frac{1}{2}\|v\|^2 \forall v \in \mathbb{R}^n$. Tomando $v = 0 \Rightarrow \alpha(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq f_x(0) + \frac{1}{2}\|0\|^2 = 0 \Rightarrow \alpha(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq 0$, por hipótese $\alpha(x) \neq 0$, logo $f_x(v(x)) \leq f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 < 0 \Rightarrow f_x(v(x)) < 0$. Portanto x não é crítico Pareto.

2. Se x não é crítico Pareto, logo existe $0 \neq v \in \mathbb{R}^n$ tal que $JF(x)v \in (-\mathbb{R}_{++})^m \Rightarrow JF(x)(tv) \in (-\mathbb{R}_{++})^m \forall t > 0$. Por outro lado, $\alpha(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2 \leq f_x(tv) + \frac{1}{2}\|tv\|^2 \Rightarrow \alpha(x) \leq tf_x(v) + t^2\frac{1}{2}\|v\|^2 = t(f_x(v) + \frac{t}{2}\|v\|^2)$. Como $f_x(v) < 0$ e para $t > 0$ suficientemente pequeno, temos $f_x(v) + \frac{t}{2}\|v\|^2 < 0$, logo $\alpha(x) < 0$ onde $t \in (0, -\frac{2f_x(v)}{\|v\|^2})$ e conseqüentemente $f_x(v(x)) \leq -\frac{1}{2}\|v(x)\|^2$

3. Considere a função $g : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $g(v) = f_x(v) + \frac{1}{2}\|v\|^2$. Assim o problema de minimização irrestrito:

$$\begin{cases} \min g \\ \text{s.a } v \in \mathbb{R}^n \end{cases} \quad (3.9)$$

tem como solução $v(x)$, logo $0 \in \partial g(v(x))$. Contudo $\partial g(v(x)) = \partial f_x(v(x)) + \partial(\frac{1}{2}\|v(x)\|^2) \Rightarrow \partial g(v(x)) = \sum_{i=1}^m \lambda_i \nabla F_i(x) + v(x) \Rightarrow v(x) = -\sum_{i=1}^m \lambda_i \nabla F_i(x)$ o que mostra que v é contínua, pois F é continuamente diferenciável. Conseqüentemente a função $x \mapsto \alpha(x)$ é contínua, pois $\alpha(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2$.

□

Para efeito de implementação algorítmica é interessante trabalhar com soluções inexatas, ou seja, aproximadas ao invés de atacar o problema (3.5) de forma exata. A seguir damos alguns passos nesta direção.

Definição 33. Dizemos que $v \in \mathbb{R}^n$ é uma solução aproximada de (3.4) com tolerância

$\sigma \in (0, 1]$ se

$$f_x(v) + \frac{1}{2}\|v\|^2 \leq \sigma\alpha(x) \quad (3.10)$$

É claro que para $\sigma = 1$ somente a solução exata satisfaz a (3.10) uma vez que $\alpha(x)$ é o valor ótimo de (3.5).

Definição 34. Seja $A \in \mathbb{R}^{m \times n}$ e $x \in \mathbb{R}^n$. Definimos a norma $\|\cdot\|_{\infty,2}$ em $\mathbb{R}^{m \times n}$ por

$$\|A\|_{\infty,2} := \max_{x \neq 0} \frac{\|Ax\|_{\infty}}{\|x\|_2} \quad (3.11)$$

Lema 3.3. *Suponha que x não seja crítico Pareto e que v seja uma solução aproximada de (3.10) com tolerância $\sigma \in (0, 1]$. Então*

$$\|v\| \leq 2\|A\|_{\infty,2} \quad (3.12)$$

onde $A = JF(x)$.

Demonstração. Como x não é crítico Pareto pelo lema (3.2) $\alpha(x) < 0$ e como $\sigma \in (0, 1]$ $\Rightarrow \sigma\alpha(x) < 0$. Seja v uma solução aproximada de (3.10)

$$f_x(v) + \frac{1}{2}\|v\|^2 \leq \sigma\alpha(x) < 0 \Rightarrow \frac{1}{2}\|v\|^2 < -f_x(v) \leq |f_x(v)| \Rightarrow \|v\|^2 \leq 2|f_x(v)|$$

mas,

$$\begin{aligned} |f_x(v)| &= |\max(JF(x)v)_i| = \left| \max_{i=1,2,\dots,n} \left(\sum_{j=1}^m A_{ij}v_j \right) \right| \\ &\leq \max_{i=1,2,\dots,n} \left| \sum_{j=1}^m A_{ij}v_j \right| = \|JF(x)v\|_{\infty} \end{aligned} \quad (3.13)$$

$$\frac{1}{2}\|v\|^2 \leq \|JF(x)v\|_{\infty} \Rightarrow \|v\| \leq 2 \frac{\|JF(x)v\|_{\infty}}{\|v\|} \leq 2 \max_{v \neq 0} \frac{\|JF(x)v\|_{\infty}}{\|v\|} \quad (3.14)$$

Portanto,

$$\frac{1}{2}\|v\| \leq \|JF(x)\|_{\infty,2} = \|A\|_{\infty,2} \quad (3.15)$$

□

3.1.1 Outra possibilidade para a direção de busca

Uma outra possibilidade para a escolha da direção de descida em $x \in \mathbb{R}^n$ consiste em tomar $v \in \mathbb{R}^n$ como uma solução do seguinte problema restrito

$$\begin{cases} \min f_x(v) \\ \text{s.a } D = \{v \in \mathbb{R}^n / \|v\|_{\infty} \leq 1\} \end{cases} \quad (3.16)$$

Note ainda, que podemos reformular este problema como

$$\begin{cases} \min \beta \\ \text{s.a } \langle \nabla F_i(x), v \rangle \leq \beta, \quad i = 1, \dots, m \\ D = \{v \in \mathbb{R}^n / \|v\|_{\infty} \leq 1\} \end{cases} \quad (3.17)$$

Que é um problema linear.

Lema 3.4. *Sejam $V(x)$, $\beta(x)$ o conjunto solução e o valor ótimo do problema (3.16), respectivamente.*

1. *Se x é crítico Pareto, então $0 \in V(x)$ e $\beta(x) = 0$.*
2. *Se x não é crítico Pareto, então $\beta(x) < 0$ e para todo $v \in V(x)$ temos*

$$(JF(x)v)_i \leq f_x(v) < 0 \quad i = 1, 2, \dots, m$$

3. *A função $\beta(x)$ é contínua.*

4. Se $\{x^k\}$ converge para \bar{x} e $v^k \in V(x^k)$ converge para \bar{v} , então $v \in V(\bar{x})$.

Demonstração. 1. Como x é crítico Pareto, temos que $JF(x)v \geq 0 \forall v \in \mathbb{R}^n$. O que implica

$$f_x(v) = \max_{i=1,2,\dots,m} \{(JF(x)v)_i\} \geq 0$$

Segue da definição de f_x

$$f_x(0) = 0$$

donde, $f_x(0) = 0 \leq f_x(v) \forall v \in \mathbb{R}^n$, conseqüentemente $\beta(x) = f_x(0) = 0$. Portanto, $0 \in V(x)$ e $\beta(x) = 0$.

2. Se x não é crítico Pareto, logo existe uma direção $v \in \mathbb{R}^n$ tal que $JF(x)v \in (-\mathbb{R}_{++})^m \Rightarrow JF(x)(tv) \in (-\mathbb{R}_{++})^m \forall t > 0$. Por outro lado, $\beta(x)$ é o valor ótimo de (3.14). Logo

$$\beta(x) \leq f_x(v) \quad \forall v \in \mathbb{R}^n \text{ com } \|v\|_\infty \leq 1 \Rightarrow \beta(x) \leq f_x(tv) = tf_x(v) \quad \forall t \in (0, \frac{1}{\|v\|}]$$

$$\text{como } f_x(v) := \max\{\langle \nabla F_i(x), v \rangle | i = 1, \dots, m\} < 0 \Rightarrow \beta(x) < 0 \quad \forall t \in (0, \frac{1}{\|v\|}].$$

Conseqüentemente,

$$(JF(x)v)_i \leq f_x(v) < 0 \quad i = 1, 2, \dots, m.$$

3. A função $\beta : \mathbb{R}^n \mapsto \mathbb{R}$ é contínua. De fato,

sejam $x \in \mathbb{R}^n$ e $\{x^k\} \subset \mathbb{R}^n$, uma seqüência em \mathbb{R}^n tal que $\lim_{k \rightarrow +\infty} x^k = x$. Assim,

para cada $k \in \mathbb{N}$, $\beta(x^k)$ é o valor ótimo do problema (2.14), isto é, $\beta(x^k) =$

$$f_{x^k}(v) = \max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v \rangle \quad \text{para algum } v \in V(x^k).$$

Sabemos que o máximo é atingido para algum $i \in \{1, \dots, m\}$ fixo (porém arbitrá-

rio) e usando a desigualdade de Cauchy- Schwarz, tem-se

$$\begin{aligned}
|\beta(x^k) - \beta(x)| &= \left| \max_{i=1,2,\dots,m} \{\langle \nabla F_i(x^k), v \rangle\} - \max_{i=1,2,\dots,m} \{\langle \nabla F_i(x), v \rangle\} \right| \\
&= |\langle \nabla F_i(x^k), v \rangle - \langle \nabla F_i(x), v \rangle| \\
&= |\langle \nabla F_i(x^k) - \nabla F_i(x), v \rangle| \\
&\leq \|\nabla F_i(x^k) - \nabla F_i(x)\|_\infty \cdot \|v\|_\infty \\
&\leq \|\nabla F_i(x^k) - \nabla F_i(x)\|_\infty \quad \text{pois, } \|v\|_\infty \leq 1.
\end{aligned}$$

Agora, pela continuidade do ∇F_i ,

Dado $\epsilon > 0$, existe $\delta > 0$ tal que $\|x^k - x\| \leq \delta$ implica $\|\nabla F_i(x^k) - \nabla F_i(x)\|_\infty < \epsilon$

Portanto,

$$|\beta(x^k) - \beta(x)| < \epsilon \Leftrightarrow \lim_{k \rightarrow +\infty} \beta(x^k) = \beta(x).$$

4. Por hipótese para cada $k \in \mathbb{N}$, $v^k \in V(x^k)$, logo por definição de $V(x^k)$, temos

$$f_{x^k}(v^k) \leq f_{x^k}(v) \quad \forall v, \|v\|_\infty \leq 1$$

Assim, $f_{x^k}(v^k)$ é o valor ótimo de (3.14), logo

$$\beta(x^k) = f_{x^k}(v^k) = \max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v^k \rangle \leq \max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v \rangle$$

Além disso, por hipótese $\lim_{k \rightarrow +\infty} v^k = \bar{v}$, pela continuidade de β e $\nabla F_i \quad \forall i = 1, 2, \dots, m$, tem-se

$$\begin{aligned}
\lim_{k \rightarrow +\infty} \max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v^k \rangle &\leq \lim_{k \rightarrow +\infty} \max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v \rangle \\
\max_{i=1,2,\dots,m} \langle \nabla F_i(\bar{x}), \bar{v} \rangle &\leq \max_{i=1,2,\dots,m} \langle \nabla F_i(\bar{x}), v \rangle \\
f_{\bar{x}}(\bar{v}) &\leq f_{\bar{x}}(v) \Rightarrow \bar{v} \in V(\bar{x}).
\end{aligned}$$

□

3.2 Cálculo do Comprimento do Passo

Lema 3.5. *Se F é diferenciável e $JF(x)v < 0$, então existe algum $\epsilon > 0$ (depende de x, v e β) tal que $F(x + tv) < F(x) + \beta t JF(x)v$ para todo $t \in (0, \epsilon]$.*

Demonstração. Como F é diferenciável, temos que $F(x + h) = F(x) + JF(x)h + R(h)$ com $\lim_{h \rightarrow 0} \frac{R(h)}{\|h\|} = 0$ o que é equivalente a

$$F_i(x + h) = F_i(x) + \langle \nabla F_i(x), h \rangle + R_i(h) \quad (3.18)$$

com

$$\lim_{h \rightarrow 0} \frac{|R_i(h)|}{\|h\|} = 0, \quad i = 1, 2, \dots, m. \quad (3.19)$$

Defina

$$a = \max_{i=1,2,\dots,m} \{ \langle \nabla F_i(x), h \rangle \}$$

Note que

$$f_x(v) < 0 \quad e \quad v \neq 0.$$

Assim de $\lim_{h \rightarrow 0} \frac{|R_i(h)|}{\|h\|} = 0 \Leftrightarrow \forall \delta > 0, \exists \epsilon > 0$ tal que $0 < t \leq \epsilon \Rightarrow |R_i(tv)| \leq \delta \|tv\| = t\|v\|\delta$. Como

$$\beta \in (0, 1] \quad e \quad a < 0, \text{ pondo } \delta = \frac{(1-\beta) \cdot |a|}{\|v\|} > 0$$

$\forall i = 1, 2, \dots, m$. Logo, para $0 < t \leq \epsilon \Rightarrow |R_i(tv)| \leq \frac{(1-\beta) \cdot |a|}{\|v\|} \cdot t\|v\| = t(1-\beta)|a|$. Porém,

$$|a| = -a = -\max\{ \langle \nabla F_i(x), v \rangle \mid i = 1, 2, \dots, m \}.$$

$$|a| = -a = \min\{-\langle \nabla F_i(x), v \rangle \mid \forall i = 1, 2, \dots, m\}.$$

consequentemente, $\forall t \in (0, \epsilon]$, tem-se

$$|R_i(tv)| \leq t(1 - \beta) \min\{-\langle \nabla F_i(x), v \rangle \mid \forall i = 1, 2, \dots, m\} \quad (3.20)$$

$$|R_i(tv)| \leq t(1 - \beta)(-\langle \nabla F_i(x), v \rangle) = -t(1 - \beta)\langle \nabla F_i(x), v \rangle \quad (3.21)$$

Logo, por (3.18) e (3.19) temos,

$$\begin{aligned} F_i(x + tv) &= F_i(x) + \langle \nabla F_i(x), tv \rangle + R_i(tv) \\ &< F_i(x) + t\langle \nabla F_i(x), v \rangle - t(1 - \beta)\langle \nabla F_i(x), v \rangle \\ &= F_i(x) + t\langle \nabla F_i(x), v \rangle \quad \forall i = 1, 2, \dots, m. \end{aligned}$$

Portanto, $F(x + tv) < F(x) + t\beta JF(x)v$, $\forall t \in (0, \epsilon]$

□

3.3 O algoritmo

1. Passo 1. Escolha duas constantes $\beta \in (0, 1)$, $\sigma \in (0, 1]$ e $x^1 \in \mathbb{R}^n$, $k := 1$
2. Passo 2. Se x^k for crítico Pareto Pare! Caso contrário.
3. Passo 3 Calcule v^k , uma solução aproximada do problema (3.5) para $x = x^k$ com tolerância σ .
4. Passo 4. Calcule o comprimento do passo $t_k \in (0, 1]$ como o máximo

$$T_k := \{t = \frac{1}{2^j} / j \in \mathbb{N}, F(x^k + tv^k) \leq F(x^k) + \beta t JF(x^k)v^k\}$$
5. Passo 5. Fazer $x^{k+1} := x^k + t_k v^k$, $k := k + 1$ e volte passo 2.

3.3.1 Comentários sobre o algoritmo

Caso o passo 3 seja atingido na iteração k , então x^k não é crítico Pareto e portanto temos,

$$\max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v^k \rangle \leq \max_{i=1,2,\dots,m} \langle \nabla F_i(x^k), v^k \rangle + \frac{1}{2} \|v^k\|^2 < 0 \Rightarrow \langle \nabla F_i(x^k), v^k \rangle < 0 \quad \forall i = 1, 2, \dots, m$$

e pelo lema (3.5) vale a regra de armijo, logo $T_k \neq \emptyset$, ou seja, está bem definido.

3.4 Convergência do Algoritmo

Nesta seção faremos uma análise da convergência do AMKdG, ou seja, mostraremos que todo ponto de acumulação da sequência gerada pelo AMKdG é crítico Pareto. É importante observar que, se o Algoritmo pára após um número finito k de iterações, ele termina num ponto Pareto, pois $\alpha(x^k) = 0$ para algum $k \in \mathbb{N}$. Assim, vamos supor que a sequência gerada pelo Algoritmo 1 seja infinito, isto é, $\alpha(x^k) \neq 0 \quad \forall k \in \mathbb{N}$.

Teorema 3.6. *Todo o ponto de acumulação da sequência $\{x^k\}_{k \in \mathbb{N}}$ gerada pelo Algoritmo 1 é crítico Pareto. Se a função F tem conjunto de nível limitado, ou seja, $L = \{x \in \mathbb{R}^n \mid F(x) \leq F(x^1)\}$ é limitado, então a sequência $\{x^k\}_{k \in \mathbb{N}}$ é limitada e portanto, tem pelo menos um ponto de acumulação crítico Pareto.*

Demonstração. Seja y um ponto de acumulação da sequência $\{x^k\}_{k \in \mathbb{N}}$ e $\{x^{k_j}\}_{j \in \mathbb{N}}$ uma subsequência de $\{x^k\}_{k \in \mathbb{N}}$ com $\lim_{j \rightarrow +\infty} x^{k_j} = y$. Pelo Lema é suficiente mostrarmos que $\alpha(y) = 0$.

Como x^k não é crítico pelo passo 4, isto é,

$$F(x^k + t_k v^k) \leq F(x^k) + \beta t_k JF(x^k) v^k \quad \forall k \in \mathbb{N} \quad (3.22)$$

Consequentemente $\{F(x^k)\}_k$ é decrescente. Passando a subsequência, obtemos

$$F(x^{k_j+1}) \leq F(x^{k_j}) + \beta t_{k_j} JF(x^{k_j}) v^{k_j} \quad \forall j \in \mathbb{N}. \quad \text{Logo,} \quad (3.23)$$

$$0 < -\beta t_{k_j} JF(x^{k_j})v^{k_j} \leq F(x^{k_j}) - F(x^{k_j+1}). \quad (3.24)$$

Pela continuidade de F , $\lim_{j \rightarrow +\infty} F(x^{k_j}) = F(y) \Rightarrow \lim_{j \rightarrow +\infty} \|F(x^{k_j}) - F(y)\| = 0$,

sendo que $\|F(x^{k_j}) - F(x^{k_j+1})\| \leq \|F(x^{k_j}) - F(y)\| + \|F(x^{k_j+1}) - F(y)\|$

Passando o limite quando $(j \rightarrow \infty)$

logo, $\lim_{j \rightarrow +\infty} \|F(x^{k_j}) - F(x^{k_j+1})\| = 0$. Donde, por (3.24) concluímos

$$\lim_{j \rightarrow +\infty} t_{k_j} JF(x^{k_j})v^{k_j} = 0 \text{ onde } t_{k_j} \in (0, 1] \quad (3.25)$$

Agora, temos dois casos a analisar: a saber,

$$1. \lim_{j \rightarrow +\infty} \sup t_{k_j} > 0$$

Tomemos uma subsequência $\{t_{k_u}\}_u$ tal que $\lim_{u \rightarrow +\infty} t_{k_u} = \bar{t} > 0$ e por (3.25)

$$\lim_{u \rightarrow +\infty} JF(x^{k_u})v^{k_u} = 0.$$

Sendo, $f_{x^{k_j}}(v^{k_j}) = \max_{i=1,2,\dots,m} \langle \nabla F_i(x^{k_j}), v^{k_j} \rangle$ tem-se que,

$$\lim_{u \rightarrow +\infty} f_{x^{k_u}}(v^{k_u}) = \max_{i=1,2,\dots,m} \left(\lim_{u \rightarrow +\infty} \langle \nabla F_i(x^{k_u}), v^{k_u} \rangle \right) \Rightarrow \lim_{u \rightarrow +\infty} f_{x^{k_u}}(v^{k_u}) = 0$$

Como $f_{x^{k_u}}(v^{k_u}) \leq f_{x^{k_u}}(v^{k_u}) + \frac{1}{2}\|v^{k_u}\|^2 < 0$, mas $\alpha(x^{k_u}) = f_{x^{k_u}}(v^{k_u}) + \frac{1}{2}\|v^{k_u}\|^2 < 0$

donde concluímos que $\lim_{u \rightarrow +\infty} \alpha(x^{k_u}) = 0$ e pela continuidade de α $\lim_{u \rightarrow +\infty} \alpha(x^{k_u}) = \alpha(y) \Rightarrow \alpha(y) = 0 \quad \therefore y$ é crítico Pareto.

$$2. \lim_{j \rightarrow +\infty} \sup t_{k_j} = 0$$

Sabemos que pelo Lema (3.3) a sequência $\{v^{k_j}\}_j$ é limitada, pois

$$\|v^{k_j}\| \leq 2\|JF(x^{k_j})\|_{\infty,2}. \quad (3.26)$$

Assim, existe uma subsequência $\{v^{k_r}\}_r$ de $\{v^{k_j}\}_j$ tal que $\lim_{r \rightarrow +\infty} v_{k_r} = \bar{v}$ e $\lim_{r \rightarrow +\infty} x^{k_r} =$

y .

Note que para todo $r \in \mathbb{N}$, temos

$$\max_{i=1,2,\dots,m} \langle \nabla F_i(x^{k_r}), v^{k_r} \rangle \leq f_{x^{k_r}}(v^{k_r}) + \frac{1}{2}\|v^{k_r}\|^2 \leq \sigma \alpha(x^{k_r}) < 0 \quad (3.27)$$

Passando o limite quando ($r \rightarrow \infty$)

$$\max_{i=1,2,\dots,m} \left(\lim_{r \rightarrow +\infty} \langle \nabla F_i(x^{k_r}), v^{k_r} \rangle \right) \leq \sigma \left(\lim_{r \rightarrow +\infty} \alpha(x^{k_r}) \right) < \lim_{r \rightarrow +\infty} 0 \quad (3.28)$$

$$\max_{i=1,2,\dots,m} \langle \nabla F_i(y), \bar{v} \rangle \leq \sigma \alpha(y) \leq 0 \quad (3.29)$$

Seja $q \in \mathbb{N}$. Como $\lim_{j \rightarrow +\infty} \sup t_{k_j} = 0$, para $r \in \mathbb{N}$ suficientemente grande tem-se $t_{k_r} < \frac{1}{2^q}$. Assim, a condição de Armijo não é satisfeita para $t = \frac{1}{2^q}$, ou seja,

$$F(x^{k_r} + (\frac{1}{2^q})v^{k_r}) \not\leq F(x^{k_r}) + \beta(\frac{1}{2^q})JF(x^{k_r})v^{k_r} \quad (3.30)$$

logo, para algum $j \in \{1, 2, \dots, m\}$ temos,

$$F_j(x^{k_r} + (\frac{1}{2^q})v^{k_r}) \geq F_j(x^{k_r}) + \beta(\frac{1}{2^q})\langle \nabla F_j(x^{k_r}), v^{k_r} \rangle \quad (3.31)$$

Passando o limite quando ($r \rightarrow \infty$)

$$F_j(y + (\frac{1}{2^q})\bar{v}) \geq F_j(y) + \beta(\frac{1}{2^q})\langle \nabla F_j(y), \bar{v} \rangle \quad (3.32)$$

Pelo Lema (3.5) temos que,

$$\max_{i=1,2,\dots,m} \langle \nabla F_i(y), \bar{v} \rangle \geq 0 \quad (3.33)$$

Comparando (3.29) e (3.33)

$$0 \leq \max_{i=1,2,\dots,m} \langle \nabla F_i(y), \bar{v} \rangle \leq \frac{1}{\sigma} \max_{i=1,2,\dots,m} \langle \nabla F_i(y), \bar{v} \rangle \leq \alpha(y) < 0. \quad (3.34)$$

Donde concluímos que $\alpha(y) = 0$ e portanto, y é crítico Pareto.

Afirmamos que o conjunto $L = \{x \in \mathbb{R}^n \mid F(x) \leq F(x^1)\}$ é compacto. Com efeito, seja $y \in \bar{L}$, logo $y = \lim_{k \rightarrow +\infty} x^k$ onde $x^k \in L \forall k$. Assim, $F(x^k) \leq F(x^1) \forall k$.

Passando ao limite quando $(k \rightarrow \infty)$, pela continuidade de F obtemos que $F(y) \leq F(x^1)$. Logo $y \in \bar{L}$ sendo L limitado $\Rightarrow L$ é compacto. Por outro lado, $\{F(x^k)\}_k$ é decrescente $\Rightarrow \{x^k\}_k \subset L$ logo, $\{x^k\}_k$ é limitada e portanto possui pelo menos um ponto de acumulação.

□

3.5 Implementação Computacional

Nesta seção apresentaremos resultados de algumas simulações computacionais feitas em ambiente Matlab desenvolvidas com o intuito de mostrar a viabilidade do algoritmo (AMKdG). Em todas as simulações computacionais, o valor de tolerância utilizada no cálculo da direção de busca foi o usualmente adotado: $\sigma = 10^{-10}$.

As figuras 3.1, 3.2 e 3.3 mostram a convergência do AMKdG para o ponto crítico Pareto

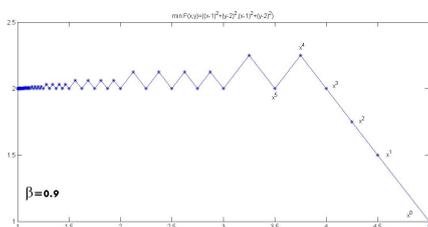


Figura 3.1: Comportamento do AMKdG quando $\beta = 0.9$

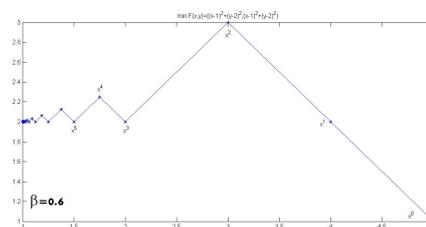


Figura 3.2: Comportamento do AMKdG quando $\beta = 0.6$

da função dada por $f(x, y) = ((x - 1)^2 + (y - 2)^2, (x - 1)^2 + (y - 2)^2)$. Observamos que quanto menor o valor de β , maior o tamanho do passo em cada iteração diminuindo, assim, o número de iterações necessárias para se chegar a uma boa aproximação do ponto crítico Pareto.

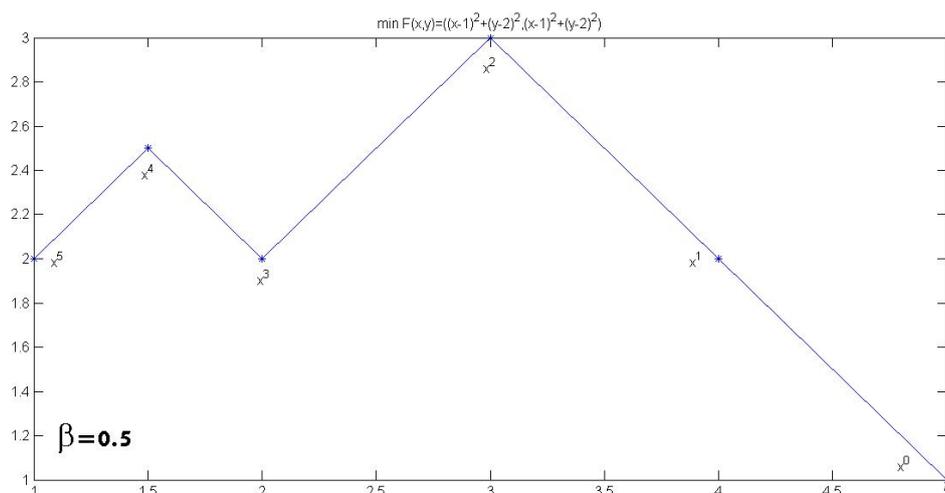


Figura 3.3: Comportamento do AMKdG quando $\beta = 0.5$

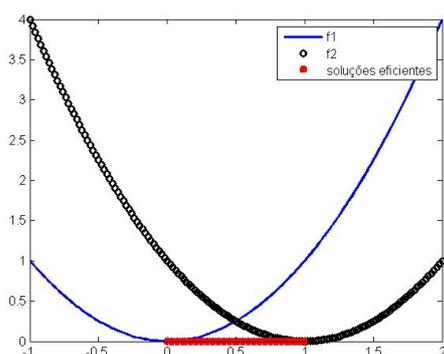


Figura 3.4: Conjunto das soluções eficientes de $f(x) = (x^2, (x-1)^2)$

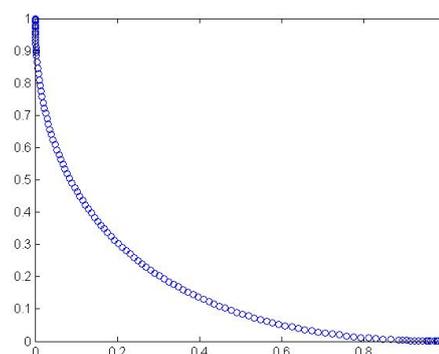


Figura 3.5: Aproximação da curva de Pareto de $f(x) = (x^2, (x-1)^2)$ gerada pelo AMKdG

Nas figuras 3.4 e 3.5 podemos ver o resultado do AMKdG na busca da curva de Pareto da função dada por $f(x) = (x^2, (x-1)^2)$. Observamos neste caso que o AMKdG produziu uma boa aproximação da curva de Pareto, como era de se esperar.

Aplicando o AMKdG na busca da curva de Pareto da função $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $f(x, y) = (x^2 + y^2, (x-1)^2 + (y-2)^2)$, cujos objetivos são conflitantes, o resultado também foi bastante satisfatório como mostram as figuras 3.6 e 3.7.

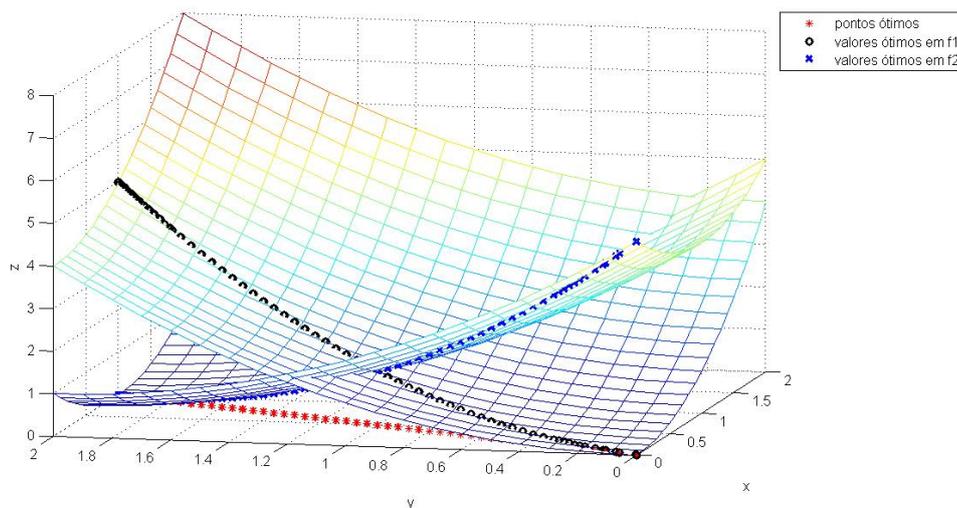


Figura 3.6: Conjunto das soluções eficientes de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2)$

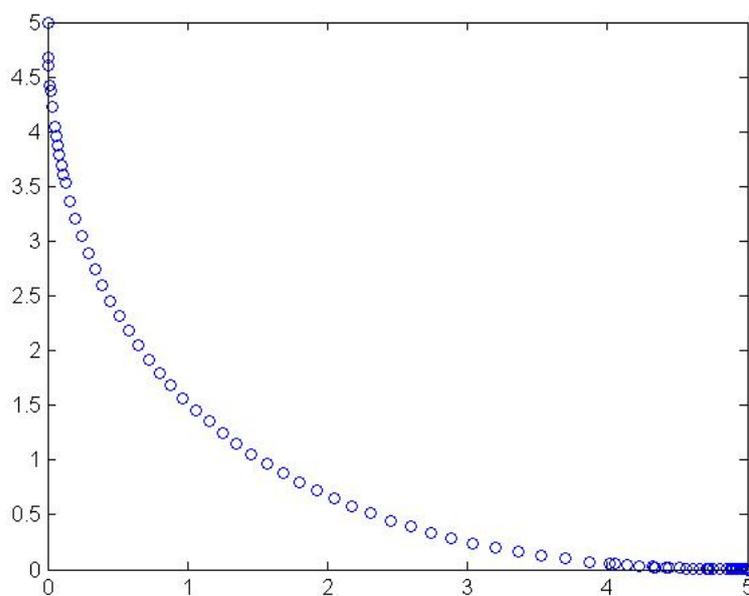


Figura 3.7: Aproximação da curva de Pareto de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2)$ gerada pelo AMKdG

Finalmente, aplicamos o AMKdG na busca das soluções eficientes e da curva de Pareto da função $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ dada por $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2, (x - 1)^2 + (y - 3)^2)$, atendendo nossas expectativas quanto aos resultados obtidos nas aproximações, como mostram as figuras 3.8 e 3.9.

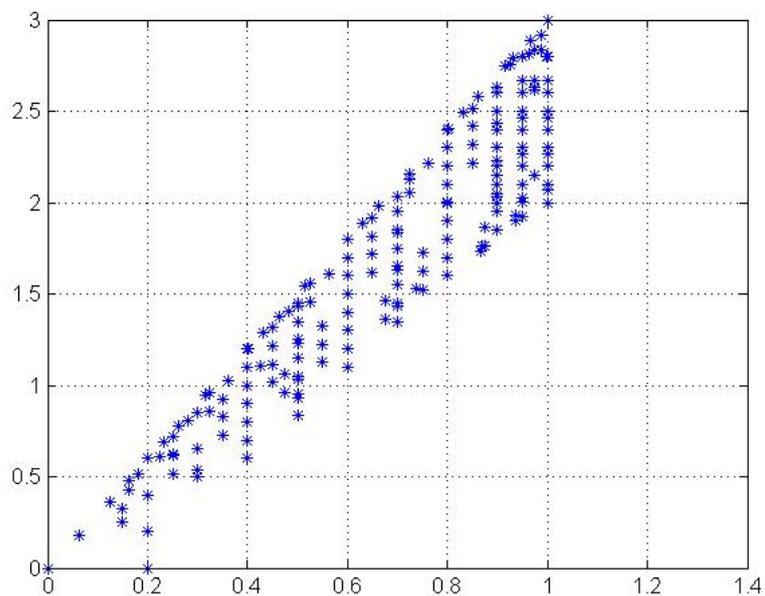


Figura 3.8: Conjunto das soluções eficientes de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2), (x - 1)^2 + (y - 3)^2)$

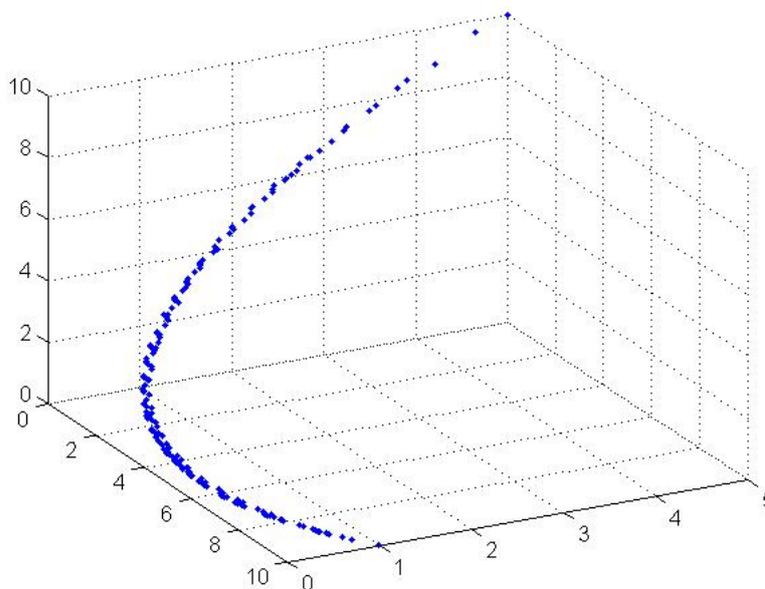


Figura 3.9: Aproximação da curva de Pareto de $f(x, y) = (x^2 + y^2, (x - 1)^2 + (y - 2)^2), (x - 1)^2 + (y - 3)^2)$ gerada pelo AMKdG

Conclusão

Neste trabalho, apresentou-se um método de descida com busca linear tipo Armijo para otimização vetorial irrestrito, onde a ordem parcial induzida por um cone K convexo, fechado, pontiagudo e com o interior não vazio foi fundamental para entendimento do conceito de dominância ou eficiência no sentido de Pareto. O resultado alcançado trata-se de uma extensão do método clássico (otimização de um único objetivo), o qual mostrou-se que sob certas condições necessárias de primeira ordem, K -otimalidade, o algoritmo (AMKdG) é globalmente convergente, ou seja, todo ponto de acumulação da sequência gerada pelo algoritmo é K -crítico.

Apresentou-se também como aplicação do caso vetorial um método de descida tipo Cauchy para otimização multiobjetivo. Neste caso, considerou-se a ordem parcial ortante não negativo do \mathbb{R}^m de forma a definir soluções do problema multiobjetivo, isto é, $K = \mathbb{R}_+^m$. Por fim, verificou-se que o algoritmo (AMKdG) gerou aproximações para as curvas de Pareto de forma bastante razoável cumprindo condições de tolerância predeterminadas. Uma outra vantagem que observamos do AMKdG é que, na sua implementação não houve a necessidade de se escalarizar as funções objetivos que demandam acréscimos de parâmetros desconhecidos a priori, deixando a escolha dos mesmos para o modelador ou tomador de decisão.

3.6 Sugestões para trabalhos futuros

Buscar mais robustez nas implementações computacionais do algoritmo estudado através de rotinas que observem as estruturas das matrizes e vetores envolvidos nos

cálculos abrangendo, assim, um maior número de casos.

Referências Bibliográficas

- [1] Arroyo, José Elias Cláudio. *Heurísticas e metaheurísticas para otimização combinatória multiobjetivo*. Tese de doutorado, Campinas: UNICAMP, 2002.
- [2] Burachik, R., Grana Drummond, L. M., Iusem, A. N., and Svaiter, B. F., *Full Convergence of the Steepest Descent Method with Inexact Line Searches*, *Optimization*, vol. 32, pp. 137-146, 1995.
- [3] Cohon, Jared L, *Multiobjective programming and planning*, Academic Press, New York San Francisco London, 1978.
- [4] COSTA, R. R. Algumas contribuições para a otimização multiobjetivo via teoria dos cones. Manaus, 2013. Dissertação de Mestrado. PPGM - Universidade Federal do Amazonas.
- [5] Ehrgott, M. *Multicriteria Optimization*, *Lecture Notes in Economics and Mathematical Systems*, vol. 491, Springer, Berlin, 2005.
- [6] Kiwiel, K. C., and Murty, K., *Convergence of the Steepest Descent Method for Minimizing Quasiconvex Functions*, *Journal of Optimization Theory and Applications*, vol. 89, pp. 221-226, 1996.
- [7] Lima, E. L., *Curso de Análise*, Instituto de Matemática Pura e Aplicada, vol. 1, terceira edição, Rio de Janeiro, 1995.
- [8] Fliege, J. and Svaiter, B. F. *Steepest descent methods for multicriteria optimization*. *Mathematical Methods of Operations Research*, 51(3):479-494, 2000.

- [9] Graña Drommond, L. M. and Svaiter, B. F., *A steepest descent method for vector optimization*, J. Comput. Appl. Math. 175(2005), no. 2, 395-414, 2005.
- [10] Luc, T.D., *Theory of Vector Optimization*, Lecture Notes in Economics and Mathematical Systems, vol. 319, Springer, Berlin, 1989.
- [11] Magron, V., Henrion, D., Lasserre, J. *Approximating Pareto Curves using Semidefinite Relaxations*. arXiv:1404.4772v2, 2014.
- [12] Mangasarian, O. L., *Nonlinear Programming*, Mc-Graw, New York, 1969; Reimpresso por SIAM, Filadélfia, Pennsylvania, 1994.
- [13] Miettinen, K. ; 1999. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers. Boston, Massachusetts.
- [14] Oliveira, Washington Alves de. *Condições de otimalidade em programação multi-objetivo fracional quadrático* . Tese de doutorado, Campinas: UNICAMP, 2011.
- [15] Rockafellar, R.T. *Convex Analysis*. New Jersey: Princeton University Press, 1970.
- [16] Takahashi, R. H. C., *Otimização Escalar e Vetorial*. Notas de Aula, Belo Horizonte, 2007. <http://www.mat.ufmg.br/taka/>.
- [17] Villacorta, K. D. V. *Método de Ponto Proximal para Otimização Vetorial e Método de Região de Confiança para Otimização multiobjetivo*. Tese de doutorado, Rio de Janeiro: COPPE/UFRJ, 2011.
- [18] Silva Neto, G.C., *Um Método para Solução de Problemas de Otimização Multi-objetivo em Ambiente Fuzzy*. Manaus, 2011. Dissertação de Mestrado. PPGM - Universidade Federal do Amazonas.
- [19] Quiroz, E. A. P., Quispe, E. M., and Oliveira, P. R., *Steepest Descent Method with a Generalized Armijo Search for Quasiconvex Functions on Riemannian Manifolds*, Journal of Mathematical Analysis and Applications, vol.341, pp. 467-477, Rio de Janeiro, 2007.

- [20] Sampaio, R.P., Teoria, métodos e aplicações de otimização multiobjetivo. São Paulo, 2011. Dissertação de Mestrado. IME - USP.
- [21] Solodov, M., e Izmailov, A., *Otimização : Métodos Computacionais*, Instituto de Matemática Pura e Aplicada , vol. 2, Rio de Janeiro, 2007.
- [22] Solodov, M., e Izmailov, A., *Otimização : Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*, Instituto de Matemática Pura e Aplicada , vol. 1, Rio de Janeiro, 2007.
- [23] Pareto, V., *Manual de Economia Política*. Tradução de João Guilherme Vargas Netto.[S.I]: Editora Nova Cultura, 1996.