UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

LAYS GRAZIELLE CARDOSO SILVA DE JESUS

Método de Descida para Problemas de Otimização Multiobjetivo

Goiânia
2010

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR AS TESES E DISSERTAÇÕES ELETRÔNICAS (TEDE) NA BIBLIOTECA DIGITAL DA UFG

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

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

2. Identificação da Tese ou Dissertação

Autor (a):	Lays Grazielle Cardoso Silva de Jesus
E-mail:	laysgraziel-el-le@hotmail.com
Seu e-mail pode ser disponibilizado na página? [ ]Sim [x] Não	
Vínculo empregatício do autor	
Agência de fomento:	CNPQ — Sigla: CNPQ
País:	Brasil — UF:Go — CNPJ: CNPQ
Título:	Método de Descida para Problemas de Otimização Multiobjetivo
Palavras-chave:	Ponto pareto, método de descida.
Título em outra língua:	Descente Methods for Problem of Multiobjetivo Optimization.
Palavras-chave em outra língua:	Pareto points, steepest descent
Área de concentração:	Matemática Aplicada
Data defesa: (dd/mm/aaaa)	30/04/2010
Programa de Pós-Graduação:	Mestrado em Matemática
Orientador (a):	Prof. Dr. Luis Román Lucambio Perez.
E-mail:	lrlp@mat.ufg.br
Co-orientador (a):	
E-mail:	

3. Informações de acesso ao documento:

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

Em caso de disponibilização parcial, assinale as permissões:

[ ] Capítulos. Especifique: ______________________________

[ ] Outras restrições: ______________________________

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

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

Lays Grazielle C. S. de Jesus

Assinatura do (a) autor (a)

Data: 17 / 01 / 2011

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

LAYS GRAZIELLE CARDOSO SILVA DE JESUS

Método de Descida para Problemas de Otimização Multiobjetivo

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

Área de concentração: Matemática Aplicada.

Orientador: Prof. Dr. Luis Román Lucambio Péres

Goiânia
2010

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

J585m

Jesus, Lays Grazielle Cardoso Silva.
Método de Descida para problemas de otimização multiobjetivo. / Lays Grazielle Cardoso Silva de Jesus. - 2010.

Orientador: Prof. Dr. Luis Román Lucambio Perez.
Dissertação (Mestrado) – Universidade Federal de Goiás, Instituto de Matemática e Estatística, 2010.
Bibliografia.

1. Ponto Pareto. 2. Método de descida. 3. Otimização Multiobjetivo. I. Título.

CDU: 51-3

LAYS GRAZIELLE CARDOSO SILVA DE JESUS

Método de Descida para Problemas de Otimização Multiobjetivo

Dissertação defendida no Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal como requisito parcial para obtenção do título de Mestre em Matemática, aprovada no dia 30 de abril de 2010, pela Banca Examinadora constituída pelos professores:

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

Prof. Dr. Luis Mauricio Graña Drummond
Faculdade de Administração e Ciências Contábeis-UFRJ

Glaydston Carvalho Bento

Prof. Dr. Glaydston de Carvalho Bento
Instituto de Matemática e Estatística - UFG

Lays Grazielle Cardoso Silva de Jesus

Graduou–se em Licenciatura plena pela UFT - Universidade Federal do Tocantins. Durante sua graduação, foi monitora no departamento de Matemática da UFT. Durante o Mestrado, na UFG - Universidade de Goiás, foi bolsista da CNPQ .

Dedico esta dissertação a minha querida mãe, a Allan Ramos (meu namorado) e a meu sobrinho Guilherme que tanto amo.

Agradecimentos

A Deus, por tudo.

A pessoa a quem devo tudo o que sou, e que sempre esteve ao meu lado. Lêda Cardoso dos Santos (minha mãe).

Aos meus pais Leandro Epifânio, Édson Dias, e em especial a Mizahel Ramalho que tanto me ajudou nesta jornada.

Aos meus irmãos Livia, Lekyson, Andréia, Ivan e em especial a Letícia que sempre mim apoio durante meus estudos.

Aos meus padrinhos Emanoel e Terezinha.

Aos meus colegas pelo apoio e companherismo, e em especial a minha grande amiga Elaine que tanto me incentivou nesta jornada.

A Ouzina pelo apoio e pelos deliciosos almoços.

Ao professor Eudes Antônio que sempre acreditou em mim e incetivou-me a seguir em frete.

Ao meu namorado Allan Ramos pela atenção e o apoio incondicional.

A Gil de Souza, Alline e em especial a Maria Aparecida pelas orações.

Ao meu orientador Luis Román pela paciência e dedicação na realização deste trabalho.

Aos professores Luis Maurício e Claydston pela participação na banca examinadora, pela leitura criteriosa e pelas indicações de alguns deslizes.

A Cnpq, pelo apoio financeiro.

<Epígrafe é uma citação relacionada com o tópico do texto>

<Nome do autor da citação>,
<Título da referência à qual a citação pertence>.

Resumo

de Jesus Silva Cardoso, Lays Grazielle . **Método de Descida para Problemas de Otimização Multiobjetivo**. Goiânia, 2010. 62p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Neste trabalho, estudamos o método de descida para problemas de otimização multiobjetivo, para o qual introduzimos uma relação de ordem induzida por um cone fechado e convexo. Estudamos como calcular uma direção de descida e provamos que todo ponto de acumulação da sequência gerada pelo método de descida com busca de Armijo é fracamente eficiente.

Palavras–chave

Ponto pareto, método de descida

Abstract

de Jesus Silva Cardoso, Lays Grazielle . **Descente Methods for Problem of Multiobjetivo Optimization** . Goiânia, 2010. 62p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

In this work, we study the descent of methods for problem of optimization multiobjective which we introduce an order of relation induced by an closed convex cone. We study as it wiel calculate an descent of direction and we prove that every accumalation point of the sequence generated by the descent of methods with search of Armijo is weakly efficient.

Keywords

Pareto points, steepest descent.

Sumário

CAPÍTULO 1

Introdução

Dada $f : \mathbb{R}^n \to \mathbb{R}^m$ e considerando a ordem parcial induzida pelo orctante não negativo $\mathbb{R}^m_+$, estamos interessados no problema

$$(P) \left\{ \min_{\mathbb{R}^m_+} \quad f(x) \right. ,$$

com dois significados: (1) um ponto $\bar{x} \in \mathbb{R}^n$ é dito solução fracamente eficiente ou fracamente Pareto ótimo do problema acima se, e somente se, não existe $x \in \mathbb{R}^n$ tal que $f(x) < f(\bar{x})$ e (2) um ponto $\bar{x}$ é dito eficiente ou Pareto-ótimo de (P) se, não existe $x \in \mathbb{R}^n$ tal que $f(x) \leq f(\bar{x})$ e $f(x) \neq f(\bar{x})$; onde as desigualdades devem ser interpretadas componente a componente.

Este problema é chamado problema de otimização multiobjetivo ou otimização multicritéria. Usualmente nenhum ponto deverá ser minimizador de todas as funções simultaneamente, ou seja, não existirá um minimizador ideal, e portanto o conceito de otimalidade tem de ser substítuido pelo o conceito de eficiência ou otimalidade - Pareto.

O objetivo deste trabalho é apresentar um método de máxima descida com busca de Armijo proposto em [1] para resolver (P). Mostraremos que todo ponto de acumulação da sequência gerada por este método é ponto crítico da função objetivo.

No segundo capítulo, fazemos uma breve revisão de alguns conceitos básicos de análise e otimização convexa, com ênfase em alguns resultados que serão de grande importância no decorrer da exposição.

No capítulo 3, formulamos o problema multicritério, definimos otimalidade Pareto e discutimos uma condição necessária de otimalidade Pareto.

No capítulo 4, propomos um método de descida para minimizar uma função vetorial no $\mathbb{R}^n$ e, seguindo as idéias em [1], mostramos que todo ponto de acumulação da sequência gerada é Pareto crítico. Finalmente propomos um método para minimizar uma função vetorial restrita a um subconjunto do $\mathbb{R}^n$. Apresentamos o teorema de convergência e fazemos um breve comentário sobre o método para minimizar uma função vetorial restrita a um conjunto convexo fechado com a relação de ordem induzida por qualquer cone convexo fechado pontudo e com interior não vazio.

CAPÍTULO 2

Preliminares

2.1 Análise Convexa

2.1.1 Conjunto convexo

Nesta seção apresentaremos alguns resultados básicos de Análise Convexa que serão utilizados neste trabalho. Convexidade é uma noção muito importante na teoria de otimização. Com hipótese de convexidade, as condições necessárias de otimalidade passam a ser suficientes.

Definição 2.1 *Um subconjunto $C \subset \mathbb{R}^n$ é chamado convexo se para todo $x, y \in C$,*

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

Exemplo 1 *i O conjunto vazio e o espaço $\mathbb{R}^n$ são convexos.*

ii Toda bola $B = B[a;r] \subset \mathbb{R}^n$, fechada de centro a e raio $r > 0$ é convexa. De fato, sejam $x, y \in B$. Então $||x-a|| \leq r$ e $||y-a|| \leq r$. Para qualquer $\alpha \in [0,1]$, temos:

$$||(1-\alpha)x+\alpha y-a|| = ||(1-\alpha)(x-a)+\alpha(y-a)|| \leq (1-\alpha)||x-a||+\alpha||y-a|| \leq r.$$

Logo, $||(1-\alpha)x+\alpha y-a|| \leq r$ o que implica que $(1-\alpha)x+\alpha y \in B$.
No caso em que a bola é aberta a demostração é análoga.

iii Qualquer semi-espaço em $\mathbb{R}^n$ (isto é, um conjunto $D = \{x \in \mathbb{R}^n; \langle a\,,\,x\rangle \leq c\}$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$) é convexo. De fato, sejam $x, y \in D$, então $\langle a\,,\,x\rangle \leq c$ e $\langle a, y\rangle \leq c$. Assim para qualquer $\alpha \in [0,1]$, temos:

$$\begin{aligned}\langle a, \alpha x + (1-\alpha)y\rangle &= \langle a, \alpha x\rangle + \langle a, (1-\alpha)y\rangle = \alpha\,\langle a,x\rangle + (1-\alpha)\,\langle a,y\rangle \\ &\leq \alpha c + (1-\alpha)c = c.\end{aligned}$$

iv Qualquer hiperplano em $\mathbb{R}^n$ (isto é, um conjunto $\{x \in \mathbb{R}^n; \langle a\,,\,x\rangle = c\}$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$) é convexo.
A demonstração é análoga à do item (iii).

Proposição 2.2 *Sejam $D_i \subset \mathbb{R}^n$, $i \in I$, conjuntos convexos, onde I é um conjunto qualquer. Então a interseção $D = \bigcap_{i\in I} D_i$ também é um conjunto convexo.*

Prova. Sejam $x, y \in D$. Como $D = \bigcap_{i\in I} D_i$, temos que, $x, y \in D_i$, para todo $i \in I$. Como os conjuntos D_i, $i \in I$, são covexos, $\alpha x + (1-\alpha)y \in D_i$ para qualquer $\alpha \in [0,1]$ e todo $i \in I$. Logo, $\alpha x + (1-\alpha)y \in \bigcap_{i\in I} D_i$. Portanto $D = \bigcap_{i\in I} D_i$, é convexo. □

Definição 2.3 *Dados $x^i \in \mathbb{R}^n$, $\alpha_i \in [0,1]$, $i = 1,...,p$, tais que $\sum_{i=1}^{p} \alpha_i = 1$, o ponto $x = \sum_{i=1}^{p} \alpha_i x^i$ chama-se a combinição convexa dos pontos $x^i \in \mathbb{R}^n$ com parâmetros α_i, onde $i = 1,...,p$.*

Pela Definição 2.1, um conjunto convexo contém as combinações convexas de quaisquer dois pontos do conjunto. Mostremos agora que, se um conjunto é convexo, então ele contém todas as combinações convexas de qualquer número de pontos do conjunto.

Proposição 2.4 *Um conjunto $D \subset \mathbb{R}^n$ é convexo se, e somente se, ele contém todas as combinações convexas de seus pontos.*

Prova. Se D possui todas as combinações convexas de seus pontos, ele contém em particular as combinações convexas de quaisquer dois pontos (isto é, para p=2). Logo, D é convexo, pela Definição 2.1.

Suponhamos agora que D seja convexo. Dados $p \in \mathbb{N}$, $x^i \in D$ e $\alpha_i \in [0,1]$, $i = 1,...,p$ tais que $\sum_{i=1}^{p} \alpha_i = 1$, definamos $x = \sum_{i=1}^{p} \alpha_i x^i$. Mostremos, por indução em relação ao número de pontos, que $x \in D$.

Se $p = 1$, tem-se $\alpha_1 = 1$ e, portanto, $x = x^1 \in D$.

Tomemos $j \in \mathbb{N}$ e suponhamos que qualquer combinação convexa de quaisquer j pontos de D pertença a D, e consideremos o caso em que $p = j+1$ pontos em D.

Se $\alpha_{j+1} = 1$, então $\alpha_i = 0$ para todo $i = 1,..., j$. Neste caso, $x = x^{j+1} \in D$.

Seja $\alpha_{j+1} \in [0,1)$. Como $1-\alpha_{j+1} > 0$, podemos escrever

$$\begin{aligned} x &= \textstyle\sum_{i=1}^{j+1} \alpha_i x^i \\ &= (1-\alpha_{j+1}) \textstyle\sum_{i=1}^{j} \frac{\alpha_i}{1-\alpha_{j+1}} x^i + \alpha_{j+1} x^{j+1} \\ &= (1-\alpha_{j+1}) y + \alpha_{j+1} x^{j+1}, \end{aligned}$$

onde

$$y = \sum_{i=1}^{j} \beta_i x^i, \;\; \beta_i = \frac{\alpha_i}{1-\alpha_{j+1}} \geq 0, \;\; i = 1,..., j.$$

Como $1 = \sum_{i=1}^{j+1} \alpha_i = \sum_{i=1}^{j} \alpha_i + \alpha_{j+1}$. Temos que,

$$\sum_{i=1}^{j} \beta_i = \sum_{i=1}^{j} \frac{\alpha_i}{1-\alpha_{j+1}} = (1-\alpha_{j+1})^{-1} \sum_{i=1}^{j} \alpha_i = (1-\alpha_{j+1})^{-1}(1-\alpha_{j+1}) = 1.$$

Portanto y é uma combinação convexa de j pontos de D. Pela hipótese de indução, $y \in D$. Agora, de (2.3), temos que x é uma combinação convexa de $y \in D$ e $x^{j+1} \in D$. Como D é convexo, obtemos que $x \in D$, como queríamos ver. □

Definição 2.5 *Seja $D \in \mathbb{R}^n$ um conjunto qualquer. O fecho convexo de D, é o menor conjunto convexo no $\mathbb{R}^n$ que contém D (ou, equivalentemente, é a interseção de todos os conjuntos convexos do $\mathbb{R}^n$ que contém D).*

Proposição 2.6 *Seja $D \subset \mathbb{R}^n$ um conjunto qualquer. O fecho convexo de D é o conjunto de todas as combinações convexas de pontos de D.*

Prova. Definimos por $C \subset \mathbb{R}^n$, o conjunto de todas combinações convexas de pontos de D, isto é,

$$C = \left\{ x \in \mathbb{R}^n; x = \sum_{i=1}^{p} \alpha_i x^i,\ \sum_{i=1}^{p} \alpha_i = 1,\ \alpha_i \in \mathbb{R}_+,\ x^i \in D,\ i = 1, \ldots, p,\ p \in \mathbb{N} \right\},$$

onde $\mathbb{R}_+ = [0, +\infty)$. Devemos mostrar que $convD = C$, onde $convD$ denota o fecho convexo de D.

Da Proposição 2.4 segue que o fecho convexo é convexo, logo contém todas as combinações convexas dos seus pontos e, portanto, de D, pois $D \subset convD$. Concluímos: $C \subset convD$.

Observamos que, $D \subset C$, o que implica que $convD \subset convC$. Portanto, se C for convexo, pela Proposição 2.4, temos que $convC = C$, e nossa afirmação está provada. Sendo assim, basta mostrar que o conjunto C é convexo.

Sejam $z^1 \in C$ e $z^2 \in C$, isto é,

$$z^1 = \sum_{i=1}^{p_1} \alpha_i x^i,\ \alpha_i \in \mathbb{R}_+,\ \sum_{i=1}^{p_1} \alpha_i = 1, p_1 \in \mathbb{N},$$

$$z^2 = \sum_{i=1}^{p_2} \beta_i y^i,\ \beta_i \in \mathbb{R}_+,\ \sum_{i=1}^{p_2} \beta_i = 1, p_2 \in \mathbb{N}.$$

Para qualquer $\gamma \in [0, 1]$,

$$\gamma z^1 + (1-\gamma) z^2 = \sum_{i=1}^{p_1} \gamma \alpha_i x^i + \sum_{i=1}^{p_2} (1-\gamma) \beta_i y^i.$$

Notemos que $\gamma\alpha_i \geq 0,\ i = 1, \ldots, p_1$ e $(1-\gamma)\beta_i \geq 0,\ i = 1, \ldots, p_2$. E ainda,

$$\sum_{i=1}^{p_1} \gamma \alpha_i + \sum_{i=1}^{p_2} (1-\gamma) \beta_i = \gamma + (1-\gamma) = 1,$$

o que mostra que o ponto $\gamma z^1 + (1-\gamma)z^2$ é uma combinação convexa de pontos de D. Logo, $\gamma z^1 + (1-\gamma)z^2 \in C$. Portanto C é convexo. $\square$

Exemplo 2 *Consideremos a esfera $D = \{x \in \mathbb{R}^n; ||x|| = c\}$, onde $c > 0$. Tomemos $x, y \in D$ e $0 < \alpha < 1$. Então $||x|| = c$, $||y|| = c$ e $||\alpha x + (1-\alpha)y|| \leq \alpha||x|| + (1-\alpha)||y|| = c$. Logo, $\alpha x + (1-\alpha)y \in B(0;c)$, (ver exemplo(2.2), item(ii)). Portanto, o fecho convexo da esfera está contido na bola de raio c e centro na origem. Tomemos agora $x \in B(0;c)$, com $||x|| < c$. Escolha $y \in D$. A reta que passa por x e y intercepta D num outro ponto z e x pode ser escrito como combinação convexa de y e z. Logo $B(0;c) \subset convD$.*

Se $X \subset \mathbb{R}^n$ for fechado e convexo, dado $z \in \mathbb{R}^n$ podemos encontrar $y \in X$ tal que $\|z-y\| \leq \|z-x\|$ para todo $x \in X$, ou seja, o elemento de X mais próximo de z. Esse elemento y é dito projeção ortogonal de z sobre X. A existência e unicidade do elemento y é discutida no teorema seguinte.

Teorema 2.7 ***(Teorema da Projeção)*** *Dado o conjunto $X \subset \mathbb{R}^n$ fechado e convexo, a aplicação $P_X : \mathbb{R}^n \to X$ dada por $P_X(z) = \arg\min\{||z-x||; x \in X\}$, está bem definida, e vale que $y = P_X(z)$ se, e somente se, $y \in X$ e $\langle z-y, x-y\rangle \leq 0$ para todo $x \in X$.*

Prova. Tomemos $z \in \mathbb{R}^n$ e $\bar{x} \in X$. As soluções do problema $\min\{||z-x||; x \in X\}$ estão no conjunto $X \cap \{u \in \mathbb{R}^n; ||z-u|| \leq ||z-\bar{x}||\}$, que é um conjunto compacto. Pelo teorema de Weierstrass, ver [4], Teorema.20, Corolário.1, segue a existência das soluções.

Seja $y \in \arg\min\{||z-x||; x \in X\}$. Como $y \in X$ e X é convexo, $(1-\alpha)y + \alpha x = z(\alpha) \in X$, para todo $\alpha \in (0,1]$ e $x \in X$. Temos, então, que $\|z-y\| \leq \|z-z(\alpha)\|$ e portanto:

$$0 \geq \|z-y\|^2 - \|z-z(\alpha)\|^2 = 2\langle z-y, z(\alpha)-y\rangle - \|z(\alpha)-y\|^2$$

$$= 2\alpha\langle z-y, x-y\rangle - \alpha^2\|x-y\|^2$$

Dividindo os dois lados da desigualdade acima por $2\alpha > 0$ e passando o limite quando $\alpha \to 0^+$, obtemos:

$$0 \geq \langle z-y, x-y\rangle.$$

Suponhamos agora que $y \in X$ e $\langle z-y, x-y\rangle \leq 0$, para todo $x \in X$, ou seja,

$$0 \geq \langle z-y, x-y\rangle = \frac{1}{2}(\|z-y\|^2 + \|x-y\|^2 - \|z-x\|^2)$$

$$\geq \frac{1}{2}(\|z-y\|^2 - \|z-x\|^2)$$

Das desigualdades acima temos que $\|z-y\| \leq \|z-x\|$ para todo $x \in X$, isto é, y é uma projeção de z sobre X. Para mostrarmos que $P_X(z)$ é única, suponhamos que $\bar{y}$ seja uma outra projeção. Então:

$$\langle z-y, \bar{y}-y\rangle \leq 0 \text{ e } \langle z-\bar{y}, y-\bar{y}\rangle \leq 0$$

Somando, obtemos que:

$$0 \geq \langle z-y-(z-\bar{y}), \bar{y}-y\rangle = \|\bar{y}-y\|^2$$

Logo, $\bar{y}=y$, como queríamos. $\square$

Definição 2.8 *Dado $X \subset \mathbb{R}^m$, uma aplicação $f: X \to \mathbb{R}^n$ diz-se Lipschitziana ou Lipschitz - contínua quando existe $k > 0$ (constante de Lipschitz de f) tal que, para quaisquer $x, y \in X$, tem-se $||f(x)-f(y)|| \leq k||x-y||$. Dizemos que f é localmente Lipschitziana em a , se dado uma bola aberta $B(a;\varepsilon)$ existe $k > 0$ tal que, $||f(x)-f(y)|| \leq k||x-y||$, para todo $x, y \in B(a;\varepsilon)$.*

A seguir, mostraremos que o operador de projeção é não expansivo isto é, Lipschitz - contínuo com constante de Lipsthiz igual a 1.

Proposição 2.9 *Seja $D \subset \mathbb{R}^n$ um conjunto convexo e fechado. Então para $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^n$ quaisquer, tem-se*

$$||P_D(x)-P_D(y)|| \leq ||x-y||. \tag{2-1}$$

Prova. Como $P_D(x) \in D$ e $P_D(y) \in D$, segue do teorema anterior, para x e y respectivamente que,

$$\langle x-P_D(x), P_D(y)-P_D(x)\rangle \leq 0,$$

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

Somando as duas desigualdades, temos que

$$\begin{aligned} 0 &\geq \langle y-P_D(y)-x+P_D(x), P_D(x)-P_D(y)\rangle \\ &= ||P_D(x)-P_D(y)||^2 + \langle y-x, P_D(x)-P_D(y)\rangle\,, \end{aligned}$$

usando a desigualdade de Cauchy-Schwarz no termo $\langle y-x, P_D(x)-P_D(y)\rangle$, obtemos,

$$||P_D(x)-P_D(y)||||y-x|| \geq \langle x-y, P_D(x)-P_D(y)\rangle\,.$$

Agora de (2.1.1), $||P_D(x)-P_D(y)||||y-x|| \geq ||P_D(x)-P_D(y)||^2$.

Se $P_D(x) = P_D(y)$, (2-1) vale trivialmente. Caso contrário, obtemos (2-1) divindindo os dois lados da desigualdade acima por $||P_D(x)-P_D(y)|| > 0$. $\square$

Segue da proposição acima que o operador projeção é contínuo.

Teorema 2.10 ***(Teorema da Separação Estrita)***

Sejam $D_1 \subset \mathbb{R}^n$ e $D_2 \subset \mathbb{R}^n$ conjuntos convexos, fechados e não vazios. Suponhamos que um deles também seja limitado. Então $D_1 \cap D_2 = \emptyset$ se, e somente se, existem $a \in \mathbb{R}^n - 0$ e $c \in \mathbb{R}$ tais que

$$\left\langle a, x^1 \right\rangle < c < \left\langle a, x^2 \right\rangle \ \forall x^1 \in D_1, \ \forall x^2 \in D_2.$$

Prova. Suponhamos,

$$\left\langle a, x^1 \right\rangle < c < \left\langle a, x^2 \right\rangle \ \forall x^1 \in D_1, \ \forall x^2 \in D_2.$$

É imediato que $D_1 \cap D_2 = \emptyset$, pois caso contrário teríamos $\langle a, x \rangle < c < \langle a, x \rangle$, com $x \in D_1 \cap D_2$.

Suponhamos agora que D_2 seja limitado e $D_1 \cap D_2 = \emptyset$. Como o conjunto D_1 é convexo e fechado, o operador de projeção $P_{D_1} : \mathbb{R}^n \to D_1$ é bem definido (pelo Teorema da Projeção) e contínuo (veja Proposição 2.10). Logo, a função

$$\psi : \mathbb{R}^n \to \mathbb{R}_+, \ \psi(x) = dist(x, D_1) = ||x - P_{D_1}(x)||$$

é contínua. Portanto, como D_2 é compacto, o problema

$$\min \psi(x) \ s.a \ x \in D_2$$

tem solução global, que denotaremos por a^2. Denotamos $a^1 = P_{D_1}(a^2)$. Pelo Teorema da Projeção

$$\left\langle a^2 - a^1, x \right\rangle \leq \left\langle a^2 - a^1, a^1 \right\rangle = c_1 \ \ \forall x \in D_1. \tag{2-2}$$

Para todo $x \in D_2$, tem-se que

$$||x - a^1|| \geq dist(x, D_1) \geq \min_{u \in D_2} dist(u, D_1) = \psi(a^2).$$

Como $dist(a^2, D_1) = ||a^2 - P_{D_1}(a^2)|| = ||a^2 - a^1||$, concluímos da desigualdade acima que

$$||x - a^1|| \geq ||a^2 - a^1||, \ \forall x \in D_2,$$

o que significa que $a^2 = P_{D_2}(a^1)$. Usando novamente o Teorema da Projeção, obtemos que

$$\left\langle a^2 - a^1, x \right\rangle \geq \left\langle a^2 - a^1, a^2 \right\rangle = c_2 \ \forall x \in D_2. \tag{2-3}$$

Definimos $a = a^2 - a^1 \neq 0$, pois $a^1 \neq a^2$ já que $D_1 \cap D_2 \neq \emptyset$ e $c = \frac{(c_1+c_2)}{2}$. Observemos que $c_2 > c > c_1$, pois

$$c_2 - c_1 = ||a^1 - a^2||^2 > 0.$$

Portanto de (2-2) e (2-3), temos que

$$\langle a, x^1 \rangle \leq c_1 < c < c_2 \leq \langle a, x^2 \rangle, \ \forall x^1 \in D^1 \ e \ \forall x^2 \in D_2.$$

□

2.1.2 Funções convexas

Definição 2.11 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Dizemos que a função $f : C \to \mathbb{R}$ é convexa em C quando para quaisquer $x, y \in C$ e $\alpha \in [0,1]$, tem-se*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y). \tag{2-4}$$

Se a desigualdade na equação (2-4) é estrita para todo $x \neq y$ e todo $\alpha \in (0,1)$, dizemos que a função é estritamente convexa. A função $f : C \subset \mathbb{R}^n \to \mathbb{R}$, é dita fortemente convexa com módulo $\gamma > 0$, se para todo $x, y \in C$ e todo $\alpha \in [0,1]$ temos que

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y) - \frac{1}{2}\gamma\alpha(1-\alpha)||x-y||^2. \tag{2-5}$$

Notemos que uma função fortemente convexa é estritamente convexa, e uma função estritamente convexa é convexa.

Teorema 2.12 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$. A função f é convexa se, e somente se, para todo $x, y \in C$ e $\beta \geq 0$ tal que $y + \beta(y-x) \in C$, nós temos*

$$f(y + \beta(y-x)) \geq f(y) + \beta(f(y) - f(x)).$$

Prova. Suponhamos que f é convexa. Denotemos $\alpha = \frac{\beta}{1+\beta}$ e $u = y + \beta(y-x)$. Então

$$y = \frac{1}{1+\beta}(u + \beta x) = (1-\alpha)u + \alpha x.$$

Portanto

$$f(y) \leq (1-\alpha) f(u) + \alpha f(x) = \frac{1}{1+\beta} f(u) + \frac{\beta}{1+\beta} f(x).$$

Suponhamos agora que dados $x, y \in C$ vale

$$f(y + \beta(y-x)) \geq f(y) + \beta(f(y) - f(x)),$$

para todo $\beta \geq 0$ tal que $y+\beta(y-x) \in C$. Calculemos α tal que $\beta = \frac{1-\alpha}{\alpha}$. Então $\alpha \in (0,1)$. Denotamos $u = \alpha x + (1-\alpha)y$. Então $x = \frac{1}{\alpha}(u-(1-\alpha)y) = u+\beta(u-y)$. Logo,

$$f(x) \geq f(u) + \beta(f(u) - f(y)) = \frac{1}{\alpha}f(u) - \frac{1-\alpha}{\alpha}f(y).$$

$\square$

O seguinte lema mostra que as "seções unidimensionais" de uma função convexa são convexas, ou seja a restrição de uma função convexa a uma reta é uma função convexa. Mais adiante, no Teorema 2.23, mostraremos um resultado mais geral.

Lema 2.13 *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ convexa, $x, y \in \mathbb{R}^n$. Então $\psi(\alpha) = f(x+\alpha y)$ é uma função convexa.*

Prova. Para $\lambda \in [0,1]$ vale que

$$\begin{aligned} \psi(\lambda\alpha + (1-\lambda)\beta) &= f(x + (\lambda\alpha + (1-\lambda)\beta)y) \\ &= f(\lambda x + (1-\lambda)x + \lambda\alpha y + (1-\lambda)\beta y) \\ &= f(\lambda(x+\alpha y) + (1-\lambda)(x+\beta y)) \\ &\leq \lambda f(x+\alpha y) + (1-\lambda)f(x+\beta y)) \\ &= \lambda\psi(\alpha) + (1-\lambda)\psi(\beta). \end{aligned}$$

$\square$

Definição 2.14 *Seja $D \subset \mathbb{R}^n$ $f : D \to \mathbb{R}$. O epígrafo de f é o conjunto*

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

A seguir, apresentamos um resultado que relaciona convexidade de uma função com convexidade de seu epígrafo.

Teorema 2.15 *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \to \mathbb{R}$ é convexa em D se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Prova. Suponhamos que f é convexa. Sejam $(x,c_1),(y,c_2) \in E_f$. Assim, $f(x) \leq c_1$ e $f(y) \leq c_2$. Pela convexidade de f, para todo $\alpha \in [0,1]$ tem-se,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha c_1 + (1-\alpha)c_2,$$

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.

Sejam $x, y \in D$ quaisquer. Obviamente, $(x, f(x)) \in E_f$ e $(y, f(y)) \in E_f$. Sendo E_f convexo, temos que

$$(\alpha x + (1-\alpha)y, \alpha f(x) + (1-\alpha)f(y)) = \alpha(x, f(x)) + (1-\alpha)(y, f(y)) \in E_f.$$

Pela definição de epígrafo, isto equivale a dizer que

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

ou seja, que f é convexa. □

Proposição 2.16 ***(Convexidade da soma de funções convexas)***
Sejam $D \subset \mathbb{R}^n$ um conjunto convexo, e $f_i : D \to \mathbb{R}$, $i = 1, \ldots, m$, funções convexas em D. Então para quaisquer $\lambda_i \in \mathbb{R}_+$, $i = 1, \ldots, m$, a função $f(x) = \sum_{i=1}^m \lambda_i f_i(x)$ é convexa em D.

Prova. Como D é convexo, $\alpha x + (1-\alpha)y \in D$, para quaisquer $x, y \in D$ e $\alpha \in [0,1]$, assim

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \textstyle\sum_{i=1}^m \lambda_i f_i(\alpha x + (1-\alpha)y) \\ &\leq \textstyle\sum_{i=1}^m \lambda_i(\alpha f_i(x) + (1-\alpha)f_i(y)) \\ &= \alpha \textstyle\sum_{i=1}^m \lambda_i f_i(x) + (1-\alpha)\sum_{i=1}^m \lambda_i f_i(y) \\ &= \alpha f(x) + (1-\alpha)f(y), \end{aligned}$$

onde a desigualdade acima segue da convexidade de f_i e do fato de $\lambda_i \in \mathbb{R}_+$, $i = 1, \ldots, m$. □

Proposição 2.17 ***(Convexidade do supremo de funções convexas)***
Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f_i : D \to \mathbb{R}$ funções convexas em D, com $i \in I$, onde I é um conjunto de índices qualquer. Suponhamos que exista $\beta \in \mathbb{R}$ tal que $f_i(x) \leq \beta$ para todo $x \in D$ e $i \in I$. Então a função

$$f(x) = \sup_{i \in I} f_i(x)$$

é convexa em D.

Prova. Como existe $\beta \in \mathbb{R}$ tal que $f_i(x) \leq \beta$, $\forall x \in D$ e $i \in I$, vale que $domf = \{x \in D; f(x) < \infty\} = D$. Temos que,

$$\begin{aligned} E_f &= \{(x,c) \in D \times \mathbb{R}; f(x) \leq c\} \\ &= \{(x,c) \in D \times \mathbb{R}; f_i(x) \leq c \ \forall i \in I\} \\ &= \textstyle\bigcap_{i \in I} \{(x,c) \in D \times \mathbb{R}; f_i(x) \leq c\} \\ &= \textstyle\bigcap_{i \in I} E_{f_i}. \end{aligned}$$

Acabamos de mostrar que, o epígrafo do supremo é a interseção dos epígrafos das funções que definem o supremo. Como as funções f_i são convexas, segue do Teorema 2.15 que os conjuntos E_{f_i} são convexos. Pela Proposição 2.3 o conjunto E_f é convexo. Usando o Teorema 2.15, temos que f é convexa. $\square$

Proposição 2.18 *Sejam $g : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $\psi : \mathbb{R} \to \mathbb{R}$ uma função convexa não decrescente. Então a função*

$$f(x) = \psi(g(x))$$

é convexa.

Prova. Temos que

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \psi(g(\alpha x + (1-\alpha)y)) \\ &\le \psi(\alpha g(x) + (1-\alpha)g(y)) \\ &\le \alpha\psi(g(x)) + (1-\alpha)\psi(g(y)) \\ &= \alpha f(x) + (1-\alpha)f(y), \end{aligned}$$

onde a primeira desigualdade segue do fato que g é uma função convexa e que ψ é não decrescente, e a segunda decorre da convexidade de ψ. $\square$

Observação 2.19 *Se a hipótese da função ψ ser não-decrescente for retirada não podemos afirmar o mesmo.*

Exemplo 3 *Seja, $g, \psi : \mathbb{R} \to \mathbb{R}$, dada por $g(x) = x^2$, $\psi(x) = -x$. A função ψ é decrescente e $f(x) = \psi(g(x)) = -x^2$ não é convexa.*

O teorema a seguir nos mostra que, a composição de uma função convexa com um operador linear é uma função convexa.

Teorema 2.20 *Seja $g : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Consideremos o operador afim $A : \mathbb{R}^n \to \mathbb{R}^m$, definido por $A(x) = Ax + b$. Então a função $f(x) = g(A(x))$ é convexa.*

Prova. Seja $x_1, x_2 \in \mathbb{R}^n$ e $\alpha \in [0,1]$, temos que

$$\begin{aligned} f(\alpha x_1 + (1-\alpha)x_2) &= g(A(\alpha x_1 + (1-\alpha)x_2)) \\ &= g(\alpha A(x_1) + (1-\alpha)A(x_2)) \\ &\le \alpha g(A(x_1)) + (1-\alpha)g(A(x_2)) \\ &= \alpha f(x_1) + (1-\alpha)f(x_2) \end{aligned}$$

Portanto f é convexa. $\square$

Mostremos agora que os conjuntos de nível de uma função convexa são convexos.

Proposição 2.21 *(Convexidade de conjuntos de nível de funções convexas)*

Suponhamos que o conjunto $D \subset \mathbb{R}^n$ seja convexo e a função $f : D \to \mathbb{R}$ seja convexa. Então o conjunto de nível

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

é convexo para todo $c \in \mathbb{R}$.

Prova. Tomemos $c \in \mathbb{R}$ arbitrário. Se $L_{f,D}(c) = \emptyset$, a conclusão segue (o conjunto vazio é convexo trivialmente).

Sejam $x \in L_{f,D}(c)$ e $y \in L_{f,D}(c)$, isto é, $x \in D$, $f(x) \leq c$ e $y \in D$, $f(y) \leq c$. Pela convexidade de D, $\alpha x + (1-\alpha)y \in D$, com $\alpha \in [0,1]$. Como f é convexa em D por hipótese, temos que

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

o que mostra que $(\alpha x + (1-\alpha)y) \in L_{f,D}(c)$. $\square$

Observação 2.22 *A convexidade de todos os conjuntos de nível de uma função não é suficiente para dizer que esta função é convexa. Como exemplo, temos a função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^3$, todos os conjuntos de nível são convexos, porém ela não é convexa. A esse grupo de funções damos o nome de funções quase-convexa, as quais não serão mencionadas novamente neste trabalho, pois não faremos uso deste tipo de funções.*

Lema 2.23 *Seja f uma função convexa no $\mathbb{R}^n$. Para quaisquer $x_1,\dots,x_n \in \mathbb{R}^n$ e coeficientes $\alpha_1,\dots,\alpha_n$, tais que*

$$\sum_{i=1}^{n} \alpha_i = 1\,, \alpha_i \geq 0\,, i = 1,\dots n, \tag{2-6}$$

temos

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i).$$

Prova. Mostremos por indução sobre n. Se $n = 2$, por convexidade, temos

$$f(\alpha_1 x_1 + \alpha_2 x_2) \leq \alpha_1 f(x_1) + \alpha_2 f(x_2).$$

Assumimos que, para $n = k$ seja verdadeiro. Mostremos que para $n = k+1$ também é. Vale que

$$\sum_{i=1}^{k+1} \alpha_i x_i = \alpha_1 x_1 + (1-\alpha_1) \sum_{i=1}^{k} \beta_i x_{i+1}$$

onde $\beta_i = \frac{\alpha_{i+1}}{1-\alpha_1}$. Claramente,

$$\sum_{i=1}^{k} \beta_i = \frac{1}{1-\alpha_1} \sum_{i=2}^{k} \alpha_i = \frac{1}{1-\alpha_1}(1-\alpha_1) = 1.$$

$\beta_i \geq 0$ $, i = 1, .., n$. Como f é convexa, segue que

$$\begin{aligned} f\left(\sum_{i=1}^{k+1} \alpha_i x_i\right) &= f\left(\alpha_1 x_1 + (1-\alpha_1) \sum_{i=1}^{k} \beta_i x_{i+1}\right) \\ &\leq \alpha_1 f(x_1) + (1-\alpha_1) f\left(\sum_{i=1}^{k} \beta_i x_{i+1}\right) \\ &\leq \sum_{i=1}^{k+1} \alpha_i f(x_i). \end{aligned}$$

□

Corolário 2.24 *Seja f uma função convexa no $\mathbb{R}^n$ e x uma combinação convexa dos pontos $x^1, \ldots, x^n$. Então*

$$f(x) \leq max_{1 \leq i \leq n} f(x^i).$$

Prova. Por hipótese $x = \sum_{i=1}^{n} \alpha_i x^i$, com $\sum_{i=1}^{n} \alpha_i = 1,\ \alpha_i \geq 0$. Assim

$$f(x) = f\left(\sum_{i=1}^{n} \alpha_i x^i\right) \leq \sum_{i=1}^{n} \alpha_i f(x^i) \leq \max_{1 \leq i \leq n} f(x^i).$$

onde a primeira desigualdade segue do Lema 2.23. □

Observação 2.25 *O fato de $f(x) \leq max_{1 \leq i \leq n} f(x^i)$ não nos garante que a função f seja convexa. Em outras palavras, a recíproca do corolário anterior não é verdadeira.*

Exemplo 4 *Tome $x^i \in \mathbb{R}$, $i = 1, \ldots, n$, $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^3$, e $g : \mathbb{R} \to \mathbb{R}$ dada por $g(x) = 0$ se $x < 0$ e $g(x) = 1$ quando $x \geq 0$. É evidente que $f(x) \leq \max_{1 \leq i \leq n} f(x^i)$, $g(x) \leq \max_{1 \leq i \leq n} g(x^i)$ e que as funções f e g não são convexas.*

O resultado seguinte nos garante que, se uma função é convexa, então ela é Lipschitz contínua em qualquer ponto do interior do domínio da função.

Teorema 2.26 *Sejam $D \subset \mathbb{R}^n$ convexo, $f : D \to \mathbb{R}$ convexa e $x_0 \in intD$. Então f é localmente Lipschitz contínua em x_0.*

Prova. Seja $B_2(x_0;\varepsilon) \subseteq D$. Como f é convexa, f é localmente limitada (ver Lema 3.1.2 Cap.3 de [10]). Assim $\sup\{f(x); x \in B_2[x_0;\varepsilon]\} \leq M$, onde $B_2[x_0;\varepsilon]$ é o fecho de $B_2(x_0;\varepsilon)$. Consideremos $y \in B_2[x_0,\varepsilon]$, $y \neq x_0$. Denotemos, $\alpha = \frac{1}{\varepsilon}||y-x_0||$, $z = x_0 + \frac{1}{\alpha}(y-x_0)$. Notemos que $||z-x_0|| = \frac{1}{\alpha}||y-x_0|| = \varepsilon$. Portanto para $y = \alpha z + (1-\alpha)x_0$ e $\alpha \in [0,1]$ temos:

$$\begin{aligned} f(y) \leq \alpha f(z) + (1-\alpha)f(x_0) &\leq f(x_0) + \alpha(M - f(x_0)) \\ &= f(x_0) + \tfrac{M-f(x_0)}{\varepsilon}||y-x_0||. \end{aligned}$$

Denotemos $u = x_0 + \frac{1}{\alpha}(x_0-y)$. Então $||x_0-y|| = \varepsilon$ e $y = x_0 + \alpha(x_0-u)$. Assim, pelo teorema **??** temos

$$f(y) \geq f(x_0) + \alpha(f(x_0)-f(u)) \geq f(x_0) - \alpha(M-f(x_0)) = f(x_0) - \frac{M-f(x_0}{\varepsilon}||y-x_0||. \tag{2-7}$$

Denotemos $u = x_0 + \frac{1}{\alpha}(x_0-y)$. Então $||u-x_0|| = \varepsilon$ e $y = x_0 + \alpha(x_0-1)$. Portanto temos,

$$\begin{aligned} f(y) &\geq \alpha f(x) + \alpha(f(x_0)-f(u)) \geq f(x_0) - \alpha(M-f(x_0)) \\ &= f(x_0) - \tfrac{M-f(x_0)}{\varepsilon}||y-x_0||. \end{aligned}$$

Portanto, $|f(y)-f(x_0)| \leq \frac{M-f(x_0)}{\varepsilon}||y-x_0||$. □

Observação 2.27 *Acabamos de mostrar que toda função $f: D \to \mathbb{R}$, onde $D \subset \mathbb{R}^n$ com interior não vazio, convexa é contínua em todo ponto $x \in intD$, pois ela é localmente Lipschitz contínua em x.*

Exemplo 5 *Consideremos $f:[a,b] \to \mathbb{R}$, definida por $f(x) = 0$, se $a < x \leq b$ e $f(a) = 1$. f é convexa, mas f não é contínua em a, pois $\lim_{x\to a^+} f(x) = 0 \neq f(a)$.*

2.1.3 Funções convexas diferenciáveis

Seja $f: U \to \mathbb{R}$, onde U é um aberto de $\mathbb{R}^n$. Tomemos $a \in U$ e $v \in \mathbb{R}^n$, dizemos que f possui derivada no ponto a na direção do vetor v, se o limite abaixo existe,

$$\lim_{t\to 0} \frac{f(a+tv)-f(a)}{t}. \tag{2-8}$$

Caso exista o limite acima, o denotaremos por $f^{'}(a,v)$ e também o chamaremos de derivada direcional de f em a na direção de v.

Fazendo $v = e_i$ (o i-ésimo vetor da base canônica de $\mathbb{R}^n$) em (2-8), obtemos

$$\lim_{t\to 0} \frac{f(a+te_i)-f(a)}{t}.$$

Quando este limite existe, o chamamos da i-ésima derivada parcial de f no ponto a e denotamos por $\frac{\partial f}{\partial x_i}(a)$. Dizemos que a função $f : U \to \mathbb{R}$, definida no aberto $U \subset \mathbb{R}^n$, é diferenciável no ponto $a \in U$, quando existem as derivadas parciais $\frac{\partial f}{\partial x_1}(a), ..., \frac{\partial f}{\partial x_n}(a)$, e para todo $v = (v_1, ..., v_n)$ tal que $a + v \in U$, temos que

$$f(a+v) - f(a) = \sum_{i=1}^{n} \frac{\partial f(a)}{\partial x_i} v_i + r(v), \; onde \; \lim_{||v|| \to 0} \frac{r(v)}{||v||} = 0.$$

O vetor $f^{'}(a) = \left(\frac{\partial f}{\partial x_1}(a), ..., \frac{\partial f}{\partial x_n}(a) \right)$ é chamado de gradiente da função f no ponto a. Sendo f diferenciável no ponto $a \in U$, vale a fórmula de Taylor, ou seja,

$$f(y) = f(a) + f^{'}(a)(y-a) + r(y-a) \; com \; \lim_{y \to a} \frac{r(y-a)}{||y-a||} = 0,$$

(ver seção 8 de [4]). Logo, toda direção $d \in \mathbb{R}^n$ com

$$f^{'}(a)(y-a) = \left\langle f^{'}(a), d \right\rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) d_i < 0$$

é tal que, existe $\varepsilon > 0$ satisfazendo que $f(a + \delta d) \leq f(a)$ quando $\delta \in [0, \varepsilon)$. Estas direções são chamadas direções de descida.

Teorema 2.28 *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável em D. Então as seguintes propriedades são equivalentes:*

(1) A função f é convexa em D.

(2) Para todo $x, y \in D$,

$$f(y) \geq f(x) + \left\langle f^{'}(x), y - x \right\rangle.$$

(3) Para todo $x, y \in D$,

$$\left\langle f^{'}(y) - f^{'}(x), y - x \right\rangle \geq 0.$$

Prova. Mostremos que (1) é equivalente a (2).

Suponhamos que f é convexa. Assim, para $x, y \in D$ e $\alpha \in (0, 1]$ quaisquer, definindo $d = y - x$, temos que

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

donde $\alpha(f(y) - f(x)) \geq f(x + \alpha d) - f(x)$. Dividindo os dois lados da desigualdade acima por $\alpha > 0$, e passando ao limite quando $\alpha \to 0+$, obtemos

$$f(y) - f(x) \geq lim_{\alpha \to 0+} \frac{f(x + \alpha d) - f(x)}{\alpha} = \left\langle f^{'}(x), d \right\rangle = \left\langle f^{'}(x), y - x \right\rangle.$$

Portanto, para todo $x,y \in D$, $f(y) \geq f(x) + \left\langle f^{'}(x), y-x \right\rangle$ e vale (2).

Suponha agora que vale (2). Definimos novamente $d = y - x$. Usando (2), para os pontos x e $x+\alpha d$; y e $x+\alpha d$ respectivamente, obtemos que

$$f(x) \geq f(x+\alpha d) - \alpha \left\langle f^{'}(x+\alpha d), d \right\rangle,$$

$$f(y) \geq f(x+\alpha d) + (1-\alpha) \left\langle f^{'}(x+\alpha d), d \right\rangle.$$

Multiplicando a primeira desigualdade por $1-\alpha \geq 0$ e a segunda por $\alpha \geq 0$, e somando, obtemos

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

o que mostra que f é uma função convexa.

Com isto, podemos concluir que f é convexa se, e somente se, para todo $x,y \in D$, $f(y) \geq f(x) + \left\langle f^{'}(x), y-x \right\rangle$.

Mostremos agora que (2) é equivalente a (3)

Trocando os papéis de x e y no item (2), temos

$$f(x) \geq f(y) + \left\langle f^{'}(y), x-y \right\rangle.$$

Somando esta desigualdade com a de (2), obtemos que, para todo $x \in D$ e todo $y \in D$, $\left\langle f^{'}(y) - f^{'}(x), y-x \right\rangle \geq 0$, isto é vale (3).

Sejam $x,y \in D$. Pelo Teorema do Valor Médio, ver em [3], existe $\alpha \in (0,1)$ tal que

$$f(y) - f(x) = \left\langle f^{'}(x+\alpha(y-x)), y-x \right\rangle. \tag{2-9}$$

Usando (3) para os pontos $x+\alpha(y-x)$ e x, obtemos

$$\left\langle f^{'}(x+\alpha(y-x)), \alpha(y-x) \right\rangle - \left\langle f^{'}(x), \alpha(y-x) \right\rangle \geq 0,$$

o que implica,

$$\begin{aligned}
\left\langle f^{'}(x+\alpha(y-x)), y-x \right\rangle &\geq \alpha^{-1} \left\langle f^{'}(x), \alpha(y-x) \right\rangle \\
&= \left\langle f^{'}(x), y-x \right\rangle.
\end{aligned}$$

Combinando esta desigualdade com (2-9), obtemos

$$f(y) - f(x) \geq \left\langle f^{'}(x), y-x \right\rangle.$$

Como queríamos.

$\square$

2.1.4 Funções convexas não diferenciáveis. Subgradiente

Agora estudaremos as propriedades das funções convexas não diferenciáveis. Funções convexas sempre possuem derivadas direcionais, como mostraremos a seguir.

Teorema 2.29 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ função convexa. Então, f possui derivada direcional no ponto x em cada direção $v \in \mathbb{R}^n$, qualquer que seja $x \in \mathbb{R}^n$. Além disso,*

$$f(x+\alpha v) \geq f(x) + \alpha f^{'}(x,v) \;\; \forall \alpha \in \mathbb{R}_+.$$

Prova. Sejam $x \in \mathbb{R}^n$ e $v \in \mathbb{R}^n$ arbitrários. Definimos

$$\psi : \mathbb{R} \to \mathbb{R} \;\; , \psi(\alpha) = f(x+\alpha v).$$

Pela definição de derivada direcional, temos que mostrar a existência do limite

$$\lim_{\alpha \to 0+} \frac{f(x+\alpha v) - f(x)}{\alpha}. \tag{2-10}$$

Observemos que

$$\frac{f(x+\alpha v) - f(x)}{\alpha} = \frac{\psi(\alpha) - \psi(0)}{\alpha}.$$

Pelo Lema 2.26, temos que f é Lipschitz-contínua em torno de x, assim existem $L > 0$ e $\bar{\alpha} > 0$ tais que

$$\frac{||\psi(\alpha) - \psi(0)||}{\alpha} \leq L||v|| \;\; \forall \alpha \in [0, \bar{\alpha}]. \tag{2-11}$$

Seja $\beta \geq \alpha > 0$. Logo, existe um certo $t \in (0,1]$ tal que $\alpha = t\beta = (1-t)0 + t\beta$. Assim, pela convexidade de ψ (veja Lema 2.13), temos que

$$\begin{aligned}
\frac{\psi(\alpha)-\psi(0)}{\alpha} &\leq \frac{(1-t)\psi(0)+t\psi(\beta)-\psi(0)}{\alpha} \\
&= \frac{t(\psi(\beta)-\psi(0))}{t\beta} \\
&= \frac{\psi(\beta)-\psi(0)}{\beta}.
\end{aligned}$$

O que mostra que a função $\alpha \mapsto \alpha^{-1}(\psi(\alpha) - \psi(0))$ é monótona não decrescente para $\alpha > 0$ e, por (2-11), ela é limitada. Segue que o limite (2-10) existe (ver Cap.6, Teorema.12 de [3]). Assim,

$$\alpha^{-1}(\psi(\alpha) - \psi(0)) \geq f^{'}(x,v),$$

ou seja,

$$f(x+\alpha v) \geq f(x) + \alpha f^{'}(x,v).$$

$\square$

Proposição 2.30 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então, dadas quaisquer duas sequências $\{x^k\}$ e $\{v^k\}$ tais que $x^k \to x$, $v^k \to d$ $(k \to \infty)$, tem-se que*

$$\limsup_{k\to\infty} f^{'}(x^k, v^k) \leq f^{'}(x,d).$$

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

Prova. Pelo o que foi visto na demostração do Teorema 2.29, para todo $\varepsilon > 0$ existe $\bar{\alpha}$ tal que $0 < \alpha < \bar{\alpha}$ implica que

$$\frac{f(x+\alpha d) - f(x)}{\alpha} < f^{'}(x,d) + \varepsilon.$$

Pelo Teorema 2.26 f é contínua. Segue então que para todo k suficientemente grande vale que

$$\frac{f(x^k+\alpha v^k) - f(x^k)}{\alpha} < f^{'}(x,d) + \varepsilon.$$

Por outro lado, na demonstração do Teorema 2.29, vimos que

$$f^{'}(x^k, v^k) \leq \frac{f(x^k+\alpha v^k) - f(x^k)}{\alpha}, \ \forall k$$

Portanto,

$$f^{'}(x^k, v^k) < f^{'}(x,d) + \varepsilon,$$

donde concluimos que

$$\limsup_{k\to\infty} f^{'}(x^k, v^k) < f^{'}(x,d) + \varepsilon.$$

Como $\varepsilon > 0$ é arbritário, isto implica o resultado desejado.

Suponhamos agora que f é convexa e diferenciável no $\mathbb{R}^n$. Sejam $x^k \to x$ $(k \to \infty)$ e $d \in \mathbb{R}^n$ quaisquer. Pela proposição anterior, temos que

$$\begin{aligned} \limsup\nolimits_{k\to\infty} \left\langle f^{'}(x^k), d \right\rangle &= \limsup\nolimits_{k\to\infty} f^{'}(x^k, d) \\ &\leq f^{'}(x,d) = \left\langle f^{'}(x), d \right\rangle, \end{aligned}$$

$$
\begin{aligned}
-\liminf_{k\to\infty}\left\langle f'(x^k),d\right\rangle &= \limsup_{k\to\infty}\left\langle f'(x^k),-d\right\rangle \\
&= \limsup_{k\to\infty} f'(x^k,-d) \\
&\leq f'(x,-d) = -\left\langle f'(x),d\right\rangle .
\end{aligned}
$$

Portanto,

$$
\limsup_{k\to\infty}\left\langle f'(x^k,d\right\rangle \leq \left\langle f'(x),d\right\rangle \leq \liminf_{k\to\infty}\left\langle f'(x^k),d\right\rangle ,
$$

isto é,

$$
\liminf_{k\to\infty}\left\langle f'(x^k),d\right\rangle = \limsup_{k\to\infty}\left\langle f'(x^k),d\right\rangle .
$$

Logo, $\lim_{k\to\infty}\left\langle f'(x^k),d\right\rangle = \left\langle f'(x),d\right\rangle$, onde $x^k \to x$ $(k\to\infty)$, e $d\in\mathbb{R}^n$ são arbitrários. Isto significa que

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

isto é, f' é contínua. □

Definição 2.31 *Seja f uma função convexa no $\mathbb{R}^n$. Dizemos que o vetor w é um subgradiente de f no ponto $x\in\mathbb{R}^n$ se, para qualquer $y\in\mathbb{R}^n$, temos*

$$
f(y) \geq f(x) + \langle w, y-x\rangle . \tag{2-12}
$$

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

Exemplo 6 *Consideremos a função $f:\mathbb{R}\to\mathbb{R}$, com $f(x)=|x|$. Para todo $y\in\mathbb{R}$ e $w\in[-1,1]$, temos*

$$
f(y) \geq wy = f(0) + w(y-0).
$$

Logo, $\partial f(0)\supset[-1,1]$. Tome $|w|>1$. Vale que $|\frac{w}{2}|<w\frac{w}{2}$, ou seja, $f(\frac{w}{2})<f(0)+w(\frac{w}{2}-0)$. Logo $w\in\partial f(0)$. Concluimos então que $\partial f(0)=[-1,1]$.

Lema 2.32 *Se para todo $x\in\mathbb{R}^n$ o subdiferencial $\partial f(x)$ é não vazio, então f é uma função convexa.*

Prova. Seja $x,y\in domf$, $\alpha\in[0,1]$. Como o $domf$ é convexo podemos considerar $y_\alpha = x+\alpha(y-x)\in domf$. Tomemos $g\in\partial f(y_\alpha)$, então

$$
\begin{aligned}
f(y) &\geq f(y_\alpha) + \langle g, y-y_\alpha\rangle = f(y_\alpha) + (1-\alpha)\langle g, y-x\rangle , \\
f(x) &\geq f(y_\alpha) + \langle g, x-y_\alpha\rangle = f(y_\alpha) - \alpha\langle g, y-x\rangle .
\end{aligned}
$$

Multiplicando a primeira inequação por α e a segunda inequação por $(1-\alpha)$ respectivamente e somando, obtemos o resultado desejado. □

Teorema 2.33 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então para todo $x \in \mathbb{R}^n$, o conjunto $\partial f(x)$ é convexo, compacto e não vazio. Além disso, para todo $d \in \mathbb{R}^n$, tem-se*

$$f'(x,d) = \max_{y \in \partial f(x)} \langle y, d \rangle \tag{2-13}$$

Prova. Como mostraremos logo a seguir $\partial f(x) \neq \emptyset$, então podemos tomar $g \in \partial f(x)$ e $y = x + \alpha d$. Partindo da definição de subgradiente, divindindo ambos os lados da desigualdade por α, e aplicando o limite com $\alpha \to 0+$, obtemos que $f'(x,d) \geq \langle g, d \rangle$, para qualquer $d \in \mathbb{R}^n$ e $g \in \partial f(x)$. Assim,

$$\begin{aligned} \partial f(x) &= \{g \in \mathbb{R}^n; \langle g, d \rangle \leq f'(x,d) \ \forall d \in \mathbb{R}^n\} \\ &= \textstyle\bigcap_{d \in \mathbb{R}^n} \{g \in \mathbb{R}^n; \langle g, d \rangle \leq f'(x,d)\}. \end{aligned} \tag{2-14}$$

Como a interseção de semi-espaços fechados é convexo e fechado, obtemos que $\partial f(x)$ é convexo e fechado. Mostremos que $\partial f(x)$ também é limitado. Suponhamos que $\partial f(x)$ é ilimitado e tomemos $(g^k)_k \subset \partial f(x)$ tal que $||g^k|| \to \infty$ $(k \to \infty)$. Para todo k, seja $d^k = \frac{g^k}{||g^k||}$. Podemos admitir, sem perda de generalidade, que $d^k \to d$ $(k \to \infty)$. Por (2-14), obtemos que

$$\left\langle g^k, d^k \right\rangle \leq f'(x, d^k).$$

Como $||g^k|| = \left\langle g^k, d^k \right\rangle$ segue que,

$$\limsup_{k \to \infty} ||g^k|| \leq \limsup_{k \to \infty} f'(x, d^k) \leq f'(x,d) < +\infty,$$

onde a segunda desigualdade segue da Proposição 2.30 e a terceira do Teorema 2.34. Isto, contradiz o fato de $||g^k|| \to \infty$ $(k \to \infty)$. Logo $\partial f(x)$ é limitado e portanto compacto.

Sabemos da análise anterior que, $f'(x,d) \geq \langle g, d \rangle$ $\forall g \in \partial f(x)$. Como $\partial f(x)$ é compacto, temos que existe o máximo por Weierstrass, logo $f'(x,d) \geq \max_{g \in \partial f(x)} \langle g, d \rangle$. Suponhamos que (2-13) não ocorre, isto é,

$$f'(x,d) > \langle g, d \rangle \ \forall g \in \partial f(x). \tag{2-15}$$

Definimos

$$D_1 = \{(z,c) \in \mathbb{R}^n \times \mathbb{R}; c > f(z)\},$$

$$D_2 = \{(z,c) \in \mathbb{R}^n \times \mathbb{R}; c = f(x) + \alpha f'(x,d),\ z = x + \alpha d,\ \alpha \in \mathbb{R}_+\}.$$

Notemos que o conjunto D_2 é convexo e que o conjunto D_1 é o epígrafo de f sem a sua fronteira, portanto D_1 também é convexo pela convexidade de f, veja Proposição 2.15.

Se $D_1 \cap D_2 \neq \emptyset$, teríamos que,

$$f(x+\alpha d) < f(x) + \alpha f^{'}(x,d)$$

para algum $\alpha \in \mathbb{R}_+$, contradizendo o Teorema 2.29. Logo $D_1 \cap D_2 = \emptyset$.

Pelo Teorema da Separação (ver Teorema.3.2.6 Cap.3, de [5]), existe um certo $(u,\beta) \in (\mathbb{R}^n \times \mathbb{R}) - 0$ tal que

$$\langle u,z\rangle + \beta c \leq \langle u, x+\alpha d\rangle + \beta(f(x)+\alpha f^{'}(x,d)) \ \forall z \in \mathbb{R}^n, \ \forall \alpha \in \mathbb{R}_+, \ \forall c > f(z). \quad (2\text{-}16)$$

Se β fosse 0, teríamos

$$\langle u,z\rangle \leq \langle u, x+\alpha d\rangle \ \ \forall z \in \mathbb{R}^n,$$

o que só pode acontecer quando $u = 0$. Como $(u,\beta) \neq 0$, concluímos então que $\beta \neq 0$.

Suponhamos $\beta > 0$. Fazendo $z = x$ e $\alpha = 0$ em (2-16), temos que $\beta c \leq \beta f(x)$ para todo os c tais que $c > f(x)$, o que é uma contradição. Concluímos então que $\beta < 0$. Dividindo os dois lados da desigualdade (2-16) por β, obtemos

$$c + \left\langle \frac{u}{\beta}, z-x \right\rangle \geq f(x) + \alpha f^{'}(x,d) + \alpha \left\langle \frac{u}{\beta}, d \right\rangle \ \forall z \in \mathbb{R}^n, \ \forall \alpha \in \mathbb{R}_+, \ \forall c > f(z). \quad (2\text{-}17)$$

Tomando os limites $\alpha \to 0_+$ e $c \to f(z)_+$, temos que

$$f(z) \geq f(x) - \left\langle \frac{u}{\beta}, z-x \right\rangle \ \forall z \in \mathbb{R}^n,$$

isto é, $-\frac{u}{\beta} \in \partial f(x)$, pela definição de subgradiente. Em particular acabamos de mostrar que $\partial f(x) \neq \emptyset$.

Tomando ainda em (2-17) $\alpha = 1$ e $z = x$, e passando ao limite $c \to f(z)_+$, obtemos

$$0 \geq f^{'}(x,d) + \left\langle \frac{u}{\beta}, d \right\rangle.$$

Temos então que

$$-\left\langle \frac{u}{\beta}, d \right\rangle \geq f^{'}(x,d), \ \ -\frac{u}{\beta} \in \partial f(x),$$

o que contradiz (2-15). Portanto, (2-13) tem que ser verdadeira. □

Corolário 2.34 *Seja f uma função com domínio convexo. A função f é convexa se, e somenete se, $\partial f(x) \neq \emptyset$, $\forall x \in dom f$.*

Prova. No teorema anterior provamos que se f é convexa então $\partial f(x) \neq \emptyset$, $\forall x \in dom f$. Suponhamos agora que $\partial f(x) \neq \emptyset$, $\forall x \in dom f$. Mostremos que f é convexa. Tome

$x, y \in domf$, $\alpha \in (0,1)$ e $z = \alpha x + (1-\alpha)y$. Pela convexidade de $domf$ temos que $z \in domf$. Seja $w \in \partial f(z)$. Vale que $f(x) \geq f(z) + \langle w, x-z\rangle$ e $f(y) \geq f(z) + \langle w, y-z\rangle$. Multiplicando a primeira inequação por α, a segunda inequação por $1-\alpha$, e somando obtemos

$$\alpha f(x) + (1-\alpha)f(y) \geq f(z) + \langle w, \alpha x + (1-\alpha)y - z\rangle = f(z).$$

□

Teorema 2.35 *Temos que* $f(\bar{x}) = \min_{x\in\mathbb{R}^n} f(x)$ *se, e somente se,* $0 \in \partial f(\bar{x})$.

Prova. Suponhamos que $0 \in \partial f(\bar{x})$. Então $f(x) \geq f(\bar{x}) + \langle 0, x-\bar{x}\rangle = f(\bar{x})$, para todo $x \in \mathbb{R}^n$. Por outro lado se $f(x) \geq f(\bar{x})$, $\forall x \in \mathbb{R}^n$, segue imediatamente da Definição 2.31 que $0 \in \partial f(\bar{x})$. □

Proposição 2.36 *Uma função convexa* $f : \mathbb{R}^n \to \mathbb{R}$ *é diferenciável no ponto* $x \in \mathbb{R}^n$ *se, e somente se, o conjunto* $\partial f(x)$ *contém um só elemento. Neste caso,* $\partial f(x) = \{f'(x)\}$.

Prova. Pela diferenciabilidade de f no ponto x, temos que $\left\langle f'(x), d\right\rangle = f'(x,d)$, (ver Seção.5 Cap.3 de [4]). Considerando $w \in \partial f(x)$, segue do Teorema 2.37 que, $f'(x,d) \geq \langle w, d\rangle$, $\forall d \in \mathbb{R}^n$. Logo

$$\left\langle f'(x), d\right\rangle \geq \langle w, d\rangle.$$

Por outro lado se, $f'(x;d) \geq \langle w, d\rangle$, $\forall d \in \mathbb{R}^n$, temos que

$$f'(x,-d) \geq \langle w, -d\rangle \Rightarrow -\left\langle f'(x), d\right\rangle \geq -\langle w, d\rangle \Rightarrow \left\langle f'(x), d\right\rangle \leq \langle w, d\rangle.$$

Portanto $\left\langle f'(x), d\right\rangle = \langle w, d\rangle$ para todo $w \in \partial f(x)$. Finalmente considerando $d = e_k$, $k = 1, \ldots, n$, nós obtemos $w = f'(x)$.

Suponhamos agora que, $\partial f(x) = \{w\}$. Pelo Teorema 2.33, $\langle w, d\rangle = f'(x,d)$, $\forall d \in \mathbb{R}^n$. Escolhendo como d os elementos da base canônica do $\mathbb{R}^n$, vemos que $w_i = \frac{\partial f(x)}{\partial x_i}$, $i = 1, \ldots, n$. Temos que $f'(x,d)$ é uma função linear em d da forma $\left\langle f'(x), d\right\rangle$, o que por definição, implica a diferenciabilidade de f em x. □

Proposição 2.37 *Sejam* $f_i : \mathbb{R}^n \to \mathbb{R}$, *com* $i = 1, \ldots, p$ *funções convexas. Então*

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

Prova. É suficiente mostrarmos para $p = 2$. Seja $f : \mathbb{R}^n \to \mathbb{R},\ f(x) = f_1(x) + f_2(x)$.

Suponhamos que $g^i \in \partial f_i(x),\ i = 1,2$, temos que

$$f_i(y) \geq f_i(x) + \left\langle g^i, y - x \right\rangle,\ \forall y \in \mathbb{R}^n.$$

Somando as duas desigualdades, obtemos

$$f(y) \geq f(x) + \left\langle g^1 + g^2, y - x \right\rangle,$$

logo $g^1 + g^2 \in \partial f(x)$. Portanto $\partial f_1(x) + \partial f_2(x) \subset \partial f(x)$.

Suponhamos agora que $\partial f(x)$ não é subconjunto de $\partial f_1(x) + \partial f_2(x)$, isto é, existe $g \in \partial f(x)$ tal que $g \notin \partial f_1(x) + \partial f_2(x)$. Sabemos do Teorema 2.33 que os conjuntos $\partial f_1(x)$ e $\partial f_2(x)$ são convexos, compactos e não vazios. Portanto, $\partial f_1(x) + \partial f_2(x)$ é convexo, (ver, Proposição.3.2.3 Cap.3 de [5]). Como $g \notin \partial f_1(x) + \partial f_2(x)$, pelo Teorema da Separação Estrita 2.10, existem $a \in \mathbb{R}^n - \{0\}$ e $c \in \mathbb{R}$ tais que

$$\left\langle a, g^1 + g^2 \right\rangle < c < \langle a, g \rangle,\ \forall g^i \in \partial f_i(x),\ i = 1,2.$$

Portanto, usando o Teorema 2.33, obtemos que

$$\begin{aligned} \langle a, g \rangle &> \sup_{g^i \in \partial f_i(x), i=1,2} \left\langle g^1 + g^2, a \right\rangle \\ &= \sup_{g^1 \in \partial f_1(x)} \left\langle g^1, a \right\rangle + \sup_{g^2 \in \partial f_2(x)} \left\langle g^2, a \right\rangle \\ &= \max_{g^1 \in \partial f_1(x)} \left\langle g^1, a \right\rangle + \max_{g^2 \in \partial f_2(x)} \left\langle g^2, a \right\rangle \\ &= f_1^{'}(x,a) + f_2^{'}(x,a) \\ &= f^{'}(x,a). \end{aligned}$$

Mas, em vista do Teorema 2.39, esta desigualdade contradiz o fato de que $g \in \partial f(x)$. O que conclui nossa demonstração. $\square$

Seja $f : X \to Y$. Dizemos que f é fechada se para todo $V \subset X$ fechado temos que, $f(V) \subset Y$ é fechado.

Lema 2.38 *Se as funções f_i, com $i = 1, \ldots, p$, são convexas e fechadas, então a função $f(x) = \max_{1 \leq i \leq p} f_i(x)$ é convexa e fechada. Além disso para $x \in \int dom f = \cap_{i=1}^{p} \int dom f_i$, temos*

$$\partial f(x) = conv\{\partial f_i(x); i \in I(x)\},$$

onde $I(x) = \{i \in \{1, \ldots, p\}; f_i(x) = f(x)\}$.

Prova. A primeira afirmação segue do fato de que o máximo de funções convexas é convexa (veja Proposição 2.19). Consideremos $x \in \cap_{i=1}^{p} \int dom f_i$, mostremos que $\partial f(x) = conv\{\partial f_i(x); i \in I(x)\}$.

Sem perda de generalidade assumimos que $I(x) = \{1, ..., k\},\ k \leq p$. Então para $v \in \mathbb{R}^n$, temos

$$f'(x,v) = \max_{1 \leq i \leq k} f_i'(x,v) = \max_{1 \leq i \leq k} \max\{\langle g_i, v\rangle ; g_i \in \partial f_i(x)\},$$

onde a segunda igualdade segue do Teorema 2.33. Notemos que para cada conjunto de valores $a_1, ..., a_k$, temos

$$\max_{1 \leq i \leq k} a_i = \max\left\{\sum_{i=1}^{k} \lambda_i a_i;\ (\lambda_1, ..., \lambda_k \in \Delta_k\right\},$$

onde $\Delta_k = \{(\lambda_1, ..., \lambda_k); \lambda_i \geq 0; 1 \leq i \leq k, \sum_{i=1}^{k} \lambda_i = 1\}$. Portanto,

$$\begin{aligned} f'(x,v) &= \max_{\{\lambda_i\} \in \Delta_k} \left\{\sum_{i=1}^{k} \lambda_i \max\{\langle g_i, v\rangle ; g_i \in \partial f_i(x)\}\right\} \\ &= \max\left\{\left\langle \sum_{i=1}^{k} \lambda_i g_i, v\right\rangle ; g_i \in \partial f_i(x), \lambda_i \in \Delta_k\right\} \\ &= \max\left\{\langle g, v\rangle ; g = \sum_{i=1}^{k} \lambda_i g_i,\ g_i \in \partial f_i(x),\ \lambda_i \in \Delta_k\right\} \\ &= \max\{\langle g, v\rangle ; g \in conv\{\partial f_i(x); i \in I(x)\}\}. \end{aligned}$$

□

2.2 Condições de otimalidade

Seja $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ e consideremos o seguinte problema

$$(P)\begin{cases} \min & f(x) \\ s.a & x \in D. \end{cases} \tag{2-18}$$

Definição 2.39 *Dizemos que um ponto $a \in D$ é*

- *minimizador ou solução global do problema (2-18), se*

$$f(a) \leq f(x)\ \forall x \in D;$$

- *minimizador local do problema (2-18), se existe uma vizinhança U de a tal que*

$$f(a) \leq f(x)\ \forall x \in D \cap U.$$

Observação 2.40 *Se $\bar{v} \in [-\infty, +\infty]$ é tal que $\bar{v} = \inf_{x \in D} f(x)$, dizemos que $\bar{v}$ é valor ótimo do problema acima.*

A seguir, citaremos a condição que deve ser satisfeita quando um ponto $\bar{x} \in D$ dado é minimizador (local) do problema (2-18). Uma condição desse tipo recebe o nome de condição necessária de otimalidade.

Teorema 2.41 *Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$, é diferenciável no ponto $\bar{x} \in D \subset \mathbb{R}^n$. Se $\bar{x}$ é um minimizador local do problema (2-18), então*

$$\left\langle f'(\bar{x}), y - \bar{x} \right\rangle \geq 0, \ \forall y \in D \tag{2-19}$$

Prova. Seja $d = y - \bar{x} \in \mathbb{R}^n$ um ponto arbitrário, porém fixo. Pela definição de minimizador local, existe $\varepsilon > 0$ tal que

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

Pela diferenciabilidade de f em $\bar{x}$, temos que

$$f(\bar{x} + td) = f(\bar{x}) + t \left\langle f'(\bar{x}), d \right\rangle + o(t),$$

onde $\frac{o(t)}{t} \to 0$ quando $t \to 0$. Logo,

$$0 \leq t \left\langle f'(\bar{x}), d \right\rangle + o(t).$$

Dividindo por $t > 0$ e tomando o limite quando $t \to 0+$, obtemos

$$0 \leq \left\langle f'(\bar{x}), d \right\rangle.$$

□

Observação 2.42 *Se D é aberto e x é um minimizador local, então $f'(x) = 0$. Se $f'(x) = 0$, então x é dito estacionário (ou crítico) para o problema (2-18). Portanto, se f é diferenciável e D é aberto, as soluções locais do problema (2-18) devem ser pontos estacionários.*

A seguir, enunciaremos um resultado que nos garante que, quando a função é convexa, todo minimizador local é global.

Teorema 2.43 *(**Teorema de Minimazação Convexa**)*

Seja $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \to \mathbb{R}$ uma função convexa em D. Então todo minimizador local do problema (2-18) é global. Além disso o conjunto de minimizadores é convexo. Se f é estritamente convexa, então este minimizador é único.

Prova. Suponhamos que $x \in \mathbb{R}$ seja minimizador local e não global. Então existe $y \in D$ tal que $f(y) < f(x)$. Definimos $x(\alpha) = \alpha y + (1-\alpha)x$. Pela convexidade de D, $x(\alpha) \in D$ para todo $\alpha \in [0,1]$. Agora, pela convexidade de f, para todo $\alpha(0,1]$, tem -se

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

Tomando $\alpha > 0$ suficientemente pequeno, podemos garantir que o ponto $x(\alpha)$ é arbitrariamente próximo ao ponto x, e ainda tem-se que $f(x(\alpha)) < f(x)$ e $x(\alpha) \in D$. Isto contradiz o fato de que x é minimizador local de (2-18). Portanto, qualquer solução local deve ser global. As outras duas afirmações são imediatas. □

Observação 2.44 *Com a hipótese de convexidade da função f, a condição necessária citada no Teorema 2.41 passa a ser suficiente, isto é, um ponto $\bar{x} \in D$ é minimizador de uma função $f : C \to \mathbb{R}$, diferenciável e convexa em $C \supset D$ se, e somente se,*

$$\left\langle f'(x), x - \bar{x} \right\rangle \geq 0 \ \ \forall x \in D. \tag{2-20}$$

De fato, se $\bar{x}$ é minimizador de f em D, segue do teorema que $\left\langle f'(x), x - \bar{x} \right\rangle \geq 0 \ \forall x \in D$.

Suponhamos agora que vale (2-20). Como f é diferenciável em $\bar{x} \in D$ por hipótese, temos, pelo Teorema 2.28, que para qualquer $x \in D$,

$$f(x) \geq f(\bar{x}) + \left\langle f'(x), x - \bar{x} \right\rangle \geq f(\bar{x}),$$

isto é, $\bar{x}$ é minimizador global.

CAPÍTULO 3

Otimização Vetorial

Na otimização vetorial a função objetivo é do tipo $f : \mathbb{R}^n \to \mathbb{R}^m$. Diante da impossibilidade de definir uma relação de ordem total em $\mathbb{R}^m$, o primeiro passo é esclarecer o que entendemos por solução ótima de um problema de otimização vetorial.

3.1 Relação de ordem parcial

Iniciamos definindo cone.

Definição 3.1 *Dizemos que um subconjunto $K \subset \mathbb{R}^m$ é um cone quando $td \in K$ para cada $d \in K$ e para cada $t \in \mathbb{R}_+$. Quando o cone K não contém retas, dizemos que ele é pontudo. O cone dual de K é definido por*

$$K^* = \{y \in \mathbb{R}^m; \langle y, d\rangle \geq 0 \ \forall d \in K\}$$

Observação 3.2 *(a) Se K é um cone convexo e $x, y \in K$, então $x + y \in K$. Porque $x + y = 2(\frac{1}{2}x + \frac{1}{2}y)$*

(b) Se $K_1 \subset K_2$, então $K_2^* \subset K_1^*$. De fato, se $x \in K_2^*$, então $\langle x, d\rangle \geq 0, \ \forall d \in K_2$, logo $\langle x, d\rangle \geq 0, \ \forall d \in K_1$, pois $K_1 \subset K_2$, o que implica que $x \in K_1^*$. Portanto $K_2^* \subset K_1^*$.

(c) O conjunto K^* é sempre convexo e fechado. De fato, sejam $x, y \in K^*$, isto é, $\langle x, d\rangle \geq 0, \ \forall d \in K$, e $\langle y, d\rangle \geq 0, \ \forall d \in K$, e $\alpha \in [0, 1]$. Para qualquer $d \in K$, temos que

$$\langle \alpha x + (1-\alpha)y, d\rangle = \alpha \langle x, d\rangle + (1-\alpha)\langle y, d\rangle \geq 0,$$

isto é, $\alpha x + (1-\alpha)y \in K^*$. Portanto, K^* é convexo.
Seja $(y^k)_k \subset K^*$, com $\lim_{k\to\infty} y^k = y$. Fixando $d \in K$ arbitrário e, passando ao limite quando $k \to \infty$ na relação $\langle y^k, d\rangle \geq 0$, obtemos $\langle y, d\rangle \geq 0$. Portanto, como $d \in K$ era arbitrário, $y \in K^*$. O que mostra que K^* é fechado.

(d) Seja $\emptyset \neq K \subset \mathbb{R}^n$ um cone qualquer. Tem-se $(K^*)^* = clconvK$, o fecho da cápsula convexa de K. Em particular, se K é convexo e fechado, então $K = (K^*)^*$. De fato, seja $d \in K$, segue da definição de K^* que, $\langle y, d\rangle \geq 0$ para todo $y \in K^*$. Usando

novamente a definição de cone dual, isto significa que $d \in (K^*)^*$. *Temos então que*

$$K \subset (K^*)^*.$$

Como acabamos de mostrar, o cone dual é convexo e fechado, logo $convK \subset (K^*)^*$, *o que implica que,* $clconvK \subset cl(K^*)^* = (K^*)^*$, *ou seja,* $clconvK \subset (K^*)^*$.
Para verificar que $(K^*)^* \subset clconvK$, *provamos que se* $x \notin clconvK$, *então* $x \notin (K^*)^*$.
Assumimos que $x \notin clconvK$. *Como* $clconvK$ *é um conjunto convexo fechado e* $\{x\}$ *é convexo e compacto, segue do Teorema da Separação Estrita (Teorema 2.10) que existem* $a \in \mathbb{R}^n - \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

$$\langle a,x \rangle < c < \langle a,d \rangle \ \ \forall d \in clconvK. \tag{3-1}$$

Suponhamos que exista $y \in K$ *tal que* $\langle a,y \rangle < 0$. *Como,* $ty \in K \subset clconvK$ *para todo* $t \in \mathbb{R}_+$, *obtemos* $\langle a,ty \rangle = t\langle a,y \rangle < c$ *para* t *suficientemente grande, o que contradiz (3-1). Concluímos que* $\langle a,y \rangle \geq 0$ *para todo* $y \in K$, *isto é,* $a \in K^*$.
Tomando $d = 0 \in clconvK$ *em (3-1), obtemos que* $c < 0$. *Usando novamente (3-1), concluímos que* $\langle a,x \rangle < 0$, *onde* $a \in K^*$, *isto significa que* $x \notin (K^*)^*$. *Logo* $(K^*)^* \subset clconvK$. *Portanto* $(K^*)^* = clconvK$.

(e) Exemplo de cone pontudo é $\mathbb{R}^m_+ = \{(x_1,...,x_m) \in \mathbb{R}^m; x_i \geq 0, i = 1,..,m\}$. *É fácil verificar que* $(\mathbb{R}^m_+)^* = \mathbb{R}^m_+$.

Definição 3.3 *Dados* $K \subset \mathbb{R}^m$ *um cone convexo fechado pontudo, e* $x,y \in \mathbb{R}^m$, *dizemos que:*

(a) $x \leq_K y$ *se, e somente se,* $y - x \in K$;
(b) $x <_K y$ *se, e somente se,* $y - x \in \int K$

As relações $x \geq_K y$ *e* $x >_K y$ *se definem equivalentemente a* $y \leq_K x$ *e* $y <_K x$, *respectivamente.*

Observemos que valem:

(i) $x \leq_K y$ e $y \leq_K z$ implica $x \leq_K z$,
(ii) $x \leq_K y$ e $y \leq_K x$ implica que $x = y$, se K for pontudo,
(iii) $x \leq_K x$
(iv) A relação de ordem $\leq_K$, definida em (3.3), não é completa, ou seja, existem vetores não comparáveis.

Exemplo 7 *Tomemos* $K = \mathbb{R}^2_+$, $u = (2,2)$, $v = (1,1)$ *e* $w = (2,0)$, *vale que,* $u >_K v$, $u \geq_K w$, *e* v *e* w *não são comparáveis, ou seja, não vale* $v \leq_K w$ *nem* $w \leq_K v$.

Observemos que, $u \leq_{\mathbb{R}^m_+} v$ equivale a $u_i \leq v_i$ para $i = 1, \ldots, m$.

A partir de agora, $K \subset \mathbb{R}^m$ é um cone convexo fechado pontundo e com interior não vazio.

3.2 K-convexidade

Definição 3.4 *Seja $D \subset \mathbb{R}^n$ convexo. Dizemos que uma função $f : D \to \mathbb{R}^m$ é K-convexa se para todo $x, y \in D$ e $\alpha \in [0,1]$, vale*

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

Se para todo $x \neq y$ e $\alpha \in (0,1)$ a desigualdade acima for estrita dizemos que f é estritamente K-convexa.

Lema 3.5 *Dado $D \subset \mathbb{R}^n$ convexo e $f : D \to \mathbb{R}^m$. A função f é K-convexa se, e somente se, a função $\langle w, f \rangle : D \to \mathbb{R}$ é convexa para todo $w \in K^*$.*

Prova. Suponhamos que f é K-convexa. Dados $x, y \in D$, $\alpha \in [0,1]$, vale

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

Logo, $\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha)y) \in K$. Tomemos $w \in K^*$, então

$$\langle w, \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha)y) \rangle \geq 0,$$

ou seja, $\langle w, \alpha f(x) + (1-\alpha) f(y) \rangle \geq \langle w, f(\alpha x + (1-\alpha)y) \rangle$ e assim

$$\alpha \langle w, f(x) \rangle + (1-\alpha) \langle w, f(y) \rangle \geq \langle w, f(\alpha x + (1-\alpha)y) \rangle,$$

ou seja $\langle w, f \rangle$ é convexa.

Tomemos $w \in K^*$ e assumamos que a função $\langle w, f \rangle$ é convexa em D. Vale que

$$\langle w, f(\alpha x + (1-\alpha)y) \rangle \leq \langle w, \alpha f(x) + (1-\alpha) f(y) \rangle,$$

ou seja

$$\langle w, \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha)y) \rangle \geq 0, \ \forall w \in K^*.$$

Aplicando a definição do cone dual, temos que $\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha)y) \in (K^*)^*$. Segue da observação (3.2) item (d) que $\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha)y) \in K$. Portanto $f(\alpha x + (1-\alpha)y) \leq_K \alpha f(x) + (1-\alpha) f(y)$, ou seja f é K-convexa. $\square$

Claramente, vale um resultado análogo para K-convexidade estrita.

3.3 O problema de otimização vetorial

Definição 3.6 *Sejam $f:\mathbb{R}^n\to\mathbb{R}^m$, $K\subset\mathbb{R}^n$ e $D\subset\mathbb{R}^n$ dados. Dizemos que um ponto $z\in D$ é Pareto ótimo global de f, se não existe outro ponto $y\in D$ tal que $f(y)\leq_K f(z)$ e $f(y)\neq f(z)$. Um ponto $z\in\mathbb{R}^n$ é chamado Pareto ótimo local, se numa vizinhança $U\subset\mathbb{R}^n$ de z não existe nenhum ponto $y\in U\cap D$ tal que $f(y)\leq_K f(z)$ e $f(y)\neq f(z)$. Ainda dizemos que $z\in D$ é Pareto fraco global, se não existe outro ponto $y\in D$ tal que $f(y)<_K f(z)$, e um ponto $z\in\mathbb{R}^n$ é chamado Pareto fraco local, se numa vizinhança $U\subset\mathbb{R}^n$ de z não existe nenhum ponto $y\in U\cap D$ tal que $f(y)<_K f(z)$.*

Dados $D\subset\mathbb{R}^n$ e $f:\mathbb{R}^n\to\mathbb{R}$. Denotaremos o problema de achar ponto Pareto ótimo de f restrito a D por:

$$(POV)\begin{cases}\min_K & f(x)\\ s.a & x\in D\end{cases}$$

É claro, da definição, que se z é Pareto ótimo (global ou local), z é Pareto fraco (global ou local, respectivamente).

Exemplo 8 *Consideremos $K=\mathbb{R}^2_+$*

(1) Seja $f:\mathbb{R}^2\to\mathbb{R}^2$ definida por $f(x)=x$ e $D=\{(x_1,x_2);x_2\geq\frac{1}{x_1},x_1>0\}$.
Os pontos Pareto ótimos do problema

$$(P)\begin{cases}\min_{\mathbb{R}^2_+} & x\\ s.a & x\in D\end{cases}$$

são da forma $\left(x_1,\frac{1}{x_1}\right)$, $x_1>0$.

(2) Sejam $a,b\in\mathbb{R}^n$ e $f:\mathbb{R}^n\to\mathbb{R}^2$ dada por $f(x)=(||x-a||^2,||x-b||^2)$. O segmento de reta que une a com b é o conjunto de pontos Pareto ótimos.

(3) Sejam $a,b\in\mathbb{R}$ e $f:\mathbb{R}^2\to\mathbb{R}^2$ dada por $f(x)=(|x_1-a|,|x_1-b|)$, onde $x=(x_1,x_2)$. O conjunto Pareto ótimo de f é $S=\{(x_1,x_2);a\leq x_1\leq b\}$.

A partir de agora $K=\mathbb{R}^m_+$, e a função f é continuamente diferenciável, isto é, as funções f_i são continuamente diferenciáveis.

Teorema 3.7 *Seja f uma função continuamente diferenciável e z Pareto ótimo para f. Então*

$$Im(Jf(z))\cap(-\mathbb{R}^m_{++})=\emptyset \tag{3-2}$$

onde, $Jf(z)$ é a matriz Jacobiana de f no ponto z e $Im(Jf(z))=\{Jf(z)v:v\in\mathbb{R}^n\}$.

Prova. Suponhamos que $d \in Im(Jf(z)) \cap \left(-\mathbb{R}^m_{++}\right)$ então, existe $v \in \mathbb{R}^n$ tal que $d = \langle \nabla f_i(z)\,, v\rangle < 0$, $\imath = 1, \ldots, m$. Logo, d é direção de descida para todas as funções coordenada de f, e portanto existe $\lambda > 0$ tal que $f(z+\lambda d) <_{\mathbb{R}^m_+} f(z)$. □

A condição (3-2) é necessária, mas não é suficiente para otimalidade. Assim, um ponto $x \in \mathbb{R}^n$ satisfazendo (3-2) é chamado Pareto crítico.

Notemos que a condição (3-2) é uma generalização da definição de ponto crítico. De fato, seja x um ponto Pareto crítico de $f : \mathbb{R}^n \to \mathbb{R}^m$. Se $m = 1$, então $Jf(x) = \nabla f(x)$ e $K = [0, +\infty)$. Considerando x Pareto crítico com relação à ordem induzida por K temos, $Im(Jf(x)) \cap int(K) = \emptyset$. Logo, para todo $v \in \mathbb{R}^n$,

$$\langle \nabla f(x)\,, v\rangle \notin (-int(K)) = (-\infty, 0),$$

assim temos que $\langle \nabla f(x)\,, v\rangle \geq 0$, para todo $v \in \mathbb{R}^n$. Tomando $v = -\nabla f(x)$ obtemos $||\nabla f(x)||^2 \leq 0$, logo $\nabla f(x) = 0$.

3.4 As funções auxiliares

Seja dado um ponto $x \in \mathbb{R}^n$. Definimos:

$$\varphi(x, v) = \max\{(Jf(x)v)_i; i = 1, \ldots, m\} = \max\{\langle \nabla f_i(x), v\rangle\,; i = 1, \ldots, m\}.$$

Como $\varphi(x, v)$ é o máximo de funções lineares, ela é convexa (Proposição 2.17). Ela é positivamente homogênea, pois, $\varphi(x, \lambda v) = \lambda\varphi(x, v)$, $\forall \lambda \in \mathbb{R}_+$, e vale que $\varphi(x, 0) = 0$.

Consideremos agora o problema

$$\begin{cases} \min & \varphi(x, v) + \frac{1}{2}||v||^2 \\ s.a & v \in \mathbb{R}^n \end{cases} \tag{3-3}$$

Observemos que a função objetivo do problema (3-3) é fortemente convexa, pois é a soma de uma função convexa com uma função fortemente convexa, o que nós garante, pelo Teorema 2.43, que o problema (3-3), possui uma única solução. Definamos $v(x)$ a solução do problema (3-3), isto é

$$v(x) = \arg\min\left\{\varphi(x, v) + \frac{1}{2}||v||^2;\, v \in \mathbb{R}^n\right\}$$

e $\theta(x)$ o valor ótimo do problema (3-3), isto é

$$\theta(x) = \min\left\{\varphi(x, v) + \frac{1}{2}||v||^2;\, v \in \mathbb{R}^n\right\}.$$

Observação 3.8 *Para $m = 1$, $v(x)$ é a direção de Cauchy, isto é, $v(x) = -\nabla f(x)$. De fato, seja $f : \mathbb{R}^n \to \mathbb{R}^m$ uma função diferenciável e seja x um ponto K-crítico de f. Se $m = 1$, $Jf(x) = \nabla f(x)$. Como x é ponto K-crítico,*

$$Im(Jf(z)) \cap (-\mathbb{R}^m_{++}) = \emptyset$$

logo para todo $v \in \mathbb{R}^n$

$$\langle \nabla f(x), v \rangle \notin -\mathbb{R}^m_{++} = (-\infty, 0),$$

pois $K = \mathbb{R}^m_+$. Assim temos que

$$\langle \nabla f(x), v \rangle \geq 0,$$

para todo $v \in \mathbb{R}^n$. Tomando $v = -\nabla f(x)$ obtemos

$$||\nabla f(x)||^2 \leq 0,$$

logo $\nabla f(x) = 0$.

O resultado a seguir nos fornece uma condição necessária e suficiente, para que um vetor $v \in \mathbb{R}^n$ seja solução de (3-3).

Lema 3.9 *Um vetor $v \in \mathbb{R}^n$ é solução do problema em (3-3) se, e somente se, existe $\alpha_i \geq 0$, $i \in I(x,v)$, tal que*

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

onde $I(x,v) := \{i \in I; \langle \nabla f_i(x), v \rangle = \max_{i \in I} \langle \nabla f_i(x), v \rangle\}$, com $I = \{1, \ldots, m\}$.

Prova. Como a função objetivo em (3-3) é convexa, segue do Teorema 2.35, que v é solução do problema em (3-3) se, e somente se,

$$0 \in \partial \left(\varphi(x,.) + \frac{1}{2}||.||^2 \right)(v),$$

ou equivalentemente, pela Proposição 2.37,

$$-v \in \partial(\varphi(x,.))(v).$$

Portanto a afirmação segue do Lema 2.38. $\square$

Para $A \in \mathbb{R}^{m\times n}$ definimos:

$$||A||_{\infty,2} = \max_{x\neq 0} \frac{||Ax||_\infty}{||x||_2},$$

que é uma norma em $\mathbb{R}^{m\times n}$.

Lema 3.10 *Para as funções v e θ valem as seguintes afirmações:*

1. *Se x é Pareto crítico, então $v(x) = 0$ e $\theta(x) = 0$.*
2. *Se x não é Pareto crítico, então $\theta(x) < 0$, $\varphi(x, v(x)) \leq -\frac{||v(x)||^2}{2} < 0$.*
3. *$||v(x)||_2 \leq 2||Jf(x)||_{\infty,2}$, para qualquer $x \in \mathbb{R}^n$.*
4. *As funções $v(x)$ e $\theta(x)$ são contínuas.*

Prova.

1. Suponhamos por absurdo que $v(x) \neq 0$. Como $v(x)$ é solução de (3-3), temos

$$\varphi(x, v(x)) + \frac{1}{2}||v(x)||^2 \leq \varphi(x, v) + \frac{1}{2}||v||^2, \forall v \in \mathbb{R}^n. \tag{3-4}$$

Em particular para $v = 0$. Assim $\varphi(x, v(x)) + \frac{1}{2}||v(x)||^2 \leq \varphi(x, 0) = 0$. Logo, $\varphi(x, v(x)) < 0$, pois $v(x) \neq 0$, e portanto $Im(Jf(x)) \cap (-\mathbb{R}^m_{++}) \neq \emptyset$, isto é, x não é Pareto crítico.
Suponhamos agora que $\theta(x) \neq 0$, pelas definições de v e de θ temos:

$$\theta(x) = \varphi(x, v(x)) + \frac{1}{2}||v(x)||^2 \leq \varphi(x, v) + \frac{1}{2}||v||^2, \forall v \in \mathbb{R}^n,$$

em particular para $v = 0$. Logo, $\theta(x) \leq 0$, $\forall x \in \mathbb{R}^n$. Assim $\theta(x) \neq 0$, implica que $\theta(x) < 0$, ou seja, que $\varphi(x, v(x)) < 0$. Mas isto nos diz que $Im(Jf(x)) \cap (-\mathbb{R}^m_{++}) \neq \emptyset$. Logo, x não é Pareto crítico. Contradição, pois por hipótese x é Pareto crítico. O item i está provado.

2. Suponhamos agora que x não é Pareto crítico, então existe $w \in \mathbb{R}^n$ tal que

$$Jf(x)w \in Im(Jf(x)) \cap (-\mathbb{R}^m_{++}).$$

Observemos que $Jf(x)(tw) \in Im(Jf(x)) \cap (-\mathbb{R}^m_{++})$, $\forall t > 0$. Como $\theta(x) \leq \varphi(x, tw) + \frac{1}{2}||tw||^2$ e a função $\varphi(x, .)$ é positivamente homogênea de grau 1, temos que:

$$\theta(x) \leq \varphi(x, tw) + \frac{1}{2}||tw||^2 = t\varphi(x, w) + \frac{t^2}{2}||w||^2 = t\left(\varphi(x, w) + \frac{t}{2}||v||^2\right), \ t > 0.$$

Assim:

$$\theta(x) \leq t\left(\varphi(x, w) + \frac{t}{2}||w||^2\right) < 0, \ t \in \left(0, -2\frac{\varphi(x, w)}{||w||^2}\right).$$

Logo, $\theta(x) < 0$ e portanto $\varphi(x, v(x)) \leq -\frac{||v(x)||^2}{2} < 0$.

3. Se x é Pareto crítico não há o que fazer, pois $v(x) = 0$. Suponhamos que x não é Pareto crítico. Pelo item anterior, vale que $\theta(x) < 0$, e consequentemente

$$\varphi(x, v(x)) + \frac{1}{2}||v(x)||^2 = \theta(x) < 0.$$

Assim,

$$\begin{aligned} ||v(x)||_2^2 < 2(-\varphi(x, v(x))) &= 2(-\max\{(Jf(x)v(x))_i;\, i = 1, \ldots, m\}) \\ &= 2\min\{(-Jf(x)v(x))_i;\, i = 1, \ldots, m\} \\ &\leq 2\max\{|Jf(x)v(x)|_i;\, i = 1, \ldots, m\} \\ &= 2||Jf(x)v(x)||_\infty. \end{aligned}$$

Como $v(x) \neq 0$, pelo item anterior, obtemos

$$||v(x)||_2 < 2\frac{||Jf(x)v(x)||_\infty}{||v(x)||_2}.$$

Portanto

$$||v(x)||_2 < 2\max_{v \neq 0} \frac{||Jf(x)v||_\infty}{||v||_2} = 2||Jf(x)||_{\infty,2}.$$

4. Seja $x \in \mathbb{R}^n$ e $(x^k)_k \subset \mathbb{R}^n$ uma sequência tal que $\lim_{k\to\infty} x^k = x$. Pela definição de $\theta(x^k)$ e $v(x^k)$, temos

$$\theta(x^k) = \varphi(x^k, v(x^k)) + \frac{1}{2}||v(x^k)||^2 \leq \varphi(x^k, v(x)) + \frac{1}{2}||v(x)||^2.$$

Assim,

$$\limsup_{k\to\infty} \theta(x^k) \leq \limsup_{k\to\infty} \left[\varphi(x^k, v(x)) + \frac{1}{2}||v(x)||^2\right].$$

Como a função φ é contínua, temos que:

$$\limsup_{k\to\infty} \theta(x^k) \leq \varphi(x, v(x)) + \frac{1}{2}||v(x)||^2 = \theta(x).$$

Portanto:

$$\limsup_{k\to\infty} \theta(x^k) \leq \theta(x). \tag{3-5}$$

Por outro lado,

$$\theta(x) = \varphi(x, v(x)) + \frac{1}{2}||v(x)||^2 \leq \varphi(x, v(x^k)) + \frac{1}{2}||v(x^k)||^2, \tag{3-6}$$

pois $v(x)$ é solução de (3-3). Assim,

$$\begin{aligned}
\liminf_{k\to\infty} \varphi(x,v(x)) + \tfrac{1}{2}||v(x)||^2 &\leq \liminf_{k\to\infty} \varphi(x,v(x^k)) + \tfrac{1}{2}||v(x^k)||^2 \\
&= \liminf_{k\to\infty} (\theta(x^k) - \varphi(x^k,v(x^k)) + \varphi(x,v(x^k)) \\
&= \liminf_{k\to\infty} (\varphi(x,v(x^k)) - \varphi(x^k,v(x^k)) + \liminf_{k\to\infty} \theta(x^k) \\
&= \liminf_{k\to\infty} \max_i[(Jf(x) - Jf(x^k))v(x^k)]_i + \liminf_{k\to\infty} \theta(x^k) \\
&= \liminf_{k\to\infty} \theta(x^k).
\end{aligned}$$

onde a última igualdade segue do fato de que $\varphi(x,v(x^k)) - \varphi(x^k,v(x^k))$ é contínua, pois é a diferença de funções contínuas. Portanto, combinando a última desigualdade com (3-6), obtemos

$$\liminf_{k\to\infty} \theta(x^k) \geq \theta(x), \tag{3-7}$$

o que, junto com (3-5), nos dá

$$\limsup_{k\to\infty} \theta(x^k) \leq \theta(x) \leq \liminf_{k\to\infty} \theta(x^k).$$

Logo, $\lim_{k\to\infty} \theta(x^k) = \theta(x)$, isto é, a função θ é contínua.

Mostremos agora que a função v é contínua.

Seja $(x^k)_k \subset \mathbb{R}^n$ uma sequência que converge a $\bar{x}$. Definimos $v^k := v(x^k)$. Segue do item *iii*, que a sequência $(v^k)_k$ é limitada. Seja $\bar{v}$ um ponto de acumulação de $(v^k)_k$. Da definição da função v e do Lema 3.9, concluímos que existem $\alpha_i^k \geq 0$, $i \in I(x^k,v^k)$, tais que

$$v^k = -\sum_{i\in I(x^k,v^k)} \alpha_i^k \nabla f_i(x^k), \quad \sum_{i\in I(x^k,v^k)} \alpha_i^k = 1, \ \ k = 0,1,\ldots \tag{3-8}$$

onde $I(x^k,v^k) := \{i \in I; \langle \nabla f_i(x^k), v^k \rangle = \max_{i\in I} \langle \nabla f_i(x^k), v^k \rangle\}$, com $I = \{1,\ldots,m\}$. Usando as constantes acima e os índices associados, definimos a sequência $(\alpha^k)_k$ como

$$\alpha^k := (\alpha_1^k,\ldots,\alpha_m^k), \ \ \alpha_i^k = 0, \ \ i \in I - I(x^k,v^k), \ \ k = 0,1,\ldots. \tag{3-9}$$

Seja $||.||_1$ a norma da soma, isto é $||x||_1 = \sum_{i=1}^m |x_i|$. Visto que $\sum_{i\in I(x^k,v^k)} \alpha_i^k = 1$, temos $||\alpha^k||_1 = 1$ para todo k, o que implica que a sequência $(\alpha^k)_k$ é limitada. Seja $\bar{\alpha}$ um ponto de acumulação da sequência $(\alpha^k)_k$, e $(v^{k_s})_s$, $(\alpha^{k_s})_s$ subsequências de $(v^k)_k$ e $(\alpha^k)_k$, tais que

$$\lim_{s\to+\infty} v^{k_s} = \bar{v}, \quad \lim_{s\to+\infty} \alpha^{k_s} = \bar{\alpha}. \tag{3-10}$$

Como o conjunto de índices I é finito e $I(x^{k_s},v^{k_s}) \subset I$ para todo s, assumimos, sem perda de generalidade, que

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

Assim, concluímos de (3-8) e da última igualdade que

$$v^{k_s} = -\sum_{i\in\bar{I}} \alpha_i^{k_s} \nabla f_i(x^{k_s}), \;\; \sum_{i\in\bar{I}} \alpha_i^{k_s} = 1, \;\; s = 0,1,....$$

Fazendo s tender a $+\infty$ nas igualdades acima, obtemos

$$\bar{v} = -\sum_{i\in\bar{I}} \bar{\alpha}_i \nabla f_i(\bar{x}), \;\; \sum_{i\in\bar{I}} \bar{\alpha}_i = 1. \tag{3-11}$$

Por outro lado, $\bar{I} = I(x^{k_s},v^{k_s}) = \left\{ i\in I; \left\langle \nabla f_i(x^{k_s}), v^{k_s} \right\rangle = \max_{i\in I} \left\langle \nabla f_i(x^{k_s}), v^{k_s} \right\rangle \right\}$. Logo, por continuidade, $\bar{I} \subset I(\bar{x},\bar{v})$. Portanto, pelo Lema 3.9, (3-11) implica que $\bar{v}$ é minimizador de $\min \varphi(\bar{x},v) + \frac{1}{2}||v||^2$. A unicidade de solução do problema (3-3) implica que $\bar{v} = v(\bar{x})$, pela definição da função v. Logo, v é contínua.

□

Podemos reformular o problema (3-3) da seguinte forma:

$$\begin{cases} \min & \alpha + \frac{1}{2}||v||^2 \\ s.a & (Jf(x)v)_i \leq \alpha, i = 1,...,m \end{cases} \tag{3-12}$$

que é um problema quadrático, com função objetivo fortemente convexa. Logo, tem uma única solução que é relativamente fácil de calcular (com certeza muito mais fácil de calcular que a solução do problema (3.3)). As soluções aproximadas serão ainda mais fáceis de calcular.

Em vez da solução exata do problema (3-3), é interessante considerar soluções inexatas, isto é soluções aproximadas.

Definição 3.11 *Dizemos que $v \in \mathbb{R}^n$ é uma solução aproximada de (3-3) para x com tolerância $\sigma \in (0,1]$ se:*

$$\varphi(x,v) + \frac{1}{2}||v||^2 \leq \sigma\theta(x). \tag{3-13}$$

Notemos que para $\sigma = 1$, somente a solução exata satisfaz (3-13).

Lema 3.12 *Suponhamos que x não é Pareto crítico e que v é uma solução aproximada de (3-3) com tolerância $\sigma \in (0,1]$. Então $v \neq 0$ e*

$$||v||_2 \leq 2||Jf(x)||_{\infty,2}.$$

Prova. Pelo item (ii) do Lema 3.10, $\theta(x) < 0$ e como $\sigma \in (0,1]$, temos que $\sigma\theta(x) < 0$. Como v é uma solução aproximada de (3-3) com tolerância $\sigma \in (0,1]$, vale

$$\varphi(x,v) + \frac{1}{2}||v||_2^2 \leq \sigma\theta(x) < 0.$$

Assim,

$$\begin{aligned} ||v||_2^2 < 2(-\varphi(x,v)) \quad &= 2(-\max\{(Jf(x)v)_i : i = 1,\ldots,m\}) \\ &= 2\min\{(-Jf(x)v)_i : i = 1,\ldots,m\} \\ &\leq 2\max\{|Jf(x)v|_i : i = 1,\ldots,m\} \\ &= 2||Jf(x)v||_\infty. \end{aligned}$$

Observemos que $\varphi(x,0) + \frac{1}{2}||0||^2 = 0 > \sigma\theta(x)$, logo $v \neq 0$, portanto

$$||v||_2 < 2\frac{||Jf(x)v||_\infty}{||v||_2}.$$

Consequentemente vale que

$$||v||_2 < 2\max_{w\neq 0}\frac{||Jf(x)w||_\infty}{||w||_2} = 2||Jf(x)||_{\infty,2}.$$

$\square$

Destaquemos que se x não é Pareto ótimo e $v \in \mathbb{R}^n$ satisfaz $\varphi(x,v) + \frac{||v||^2}{2} < 0$, então existe um $\sigma \in (0,1]$ tal que v é solução aproximada de (3-3) com esta tolerância σ.

Observação 3.13 *Seja $\bar{x}$ não Pareto ótimo. Temos, por Lema* 3.12 *i, que* $\theta(\bar{x}) < 0$. *Tomemos* $\sigma \in (0,1]$ *e v uma solução aproximada de (3-3) para $\bar{x}$ com tolerância σ. Temos que* $\max\{\langle\nabla f_i(\bar{x}), v\rangle : i = 1,\ldots,m\} = \varphi(\bar{x},v) < -\frac{1}{2}||v||^2 < 0$. *Logo o vetor v é uma direção de descida em $\bar{x}$ para todas as funções $f_i(x)$.*

3.5 Computando o passo

Suponhamos agora que $v \in \mathbb{R}^n$ é tal que $Jf(x)v < 0$, isto é, v é uma direção de descida para f em x. Nosso objetivo é calcular $y = x + tv$, $t > 0$, tal que $f(y) <_{\mathbb{R}^m_+} f(x)$. Para computarmos o tamanho do passo $t > 0$, usaremos uma regra tipo Armijo. Seja $\beta \in (0,1]$, uma constante. A condição para adimitir t é:

$$f(x+tv) \leq_{\mathbb{R}^m_+} f(x) + \beta t Jf(x)v \tag{3-14}$$

Começamos com $t = 1$ e enquanto a condição (3-14) não é satisfeita, fazemos $t := \frac{t}{2}$. Esse procedimento é finito, porque pela continuidade da f, (3-14) vale estritamente para valores pequenos de t (num intervalo não degenerado do tipo $(0,\varepsilon]$), como será provado no seguinte lema.

Lema 3.14 *Se $Jf(x)v <_{\mathbb{R}^m_+} 0$ e $\beta \in (0,1)$, então existe $\varepsilon > 0$ pequeno (que depende continuamente de x, v e β) tal que:*

$$f(x+tv) <_{\mathbb{R}^m_+} f(x) + \beta t Jf(x)v$$

para algum $t \in (0,\varepsilon]$.

Prova. Asumindo que f é uma função diferenciável, temos que

$$f(x+h) = f(x) + Jf(x)h + R(h),$$

com $\lim_{h\to 0} \frac{|R_i(h)|}{||h||} = 0\,, i = 1,2,...,m$, onde $R(h) = (R_1(h),...,R_m(h)) \in \mathbb{R}^m$.

Observemos que $\varphi(x,v) < 0$ e $v \neq 0$. Desde que $\beta < 1$, existe um $\varepsilon > 0$ tal que

$$0 < t \leq \varepsilon \Rightarrow \frac{|R_i(tv)|}{||tv||} < \frac{(1-\beta)|\varphi(x,v)|}{||v||}\,, i = 1,2,...,m$$

Portanto, para $0 < t \leq \varepsilon$ obtemos:

$$|R_i(tv)| < t(1-\beta)|\varphi(x,v)|\,, i = 1,2,...,m.$$

Como $|\varphi(x,v)| = -\varphi(x,v) = \min_{i=1,...,m} -(Jf(x)v)_i$, vale, que

$$R(tv) <_{\mathbb{R}^m_+} -t(1-\beta)Jf(x)v.$$

Portanto,

$$f(x+tv) = f(x) + tJf(x)v + R(tv) <_{\mathbb{R}^m_+} f(x) + tJf(x)v - t(1-\beta)Jf(x)v = f(x) + t\beta Jf(x)v$$

para $0 < t \leq \varepsilon$. □

3.6 Como calcular uma direção de descida.

Nesta seção, discutiremos outras possibilidades para o cálculo de uma direção de descida em x.

Em vez de consideramos a solução do problema (3-3), podemos tomar $\tilde{v}(x)$ como a solução de:

$$\begin{cases} \min & \varphi(x,v) \\ s.a & \|v\|_\infty \leq 1 \end{cases} \tag{3-15}$$

O problema (3-15) pode ser reformulado como

$$\begin{cases} \min & \beta \\ s.a & (Jf(x)v)_i \leq \beta,\ i=1,\ldots,m \\ & \|v\|_\infty \leq 1, \end{cases} \tag{3-16}$$

que é um problema linear porque

$$\{v \in \mathbb{R}^n; \|v\|_\infty \leq 1\} = \{v \in \mathbb{R}^n; -1 \leq v_i \leq 1,\ i=1,\ldots,n\}.$$

Observação 3.15 1. *Sejam $\tilde{v}(x) \in \{v \in \mathbb{R}^n; \|v\|_\infty \leq 1\}$, $\varphi(x,\tilde{v}(x))$ o valor ótimo de (3-15) e $\beta(x)$ o valor ótimo de (3-16). Observe que $(Jf(x)\tilde{v}(x))_i \leq \beta(x)$, $i=1,\ldots,m$. Logo $\varphi(x,\tilde{v}(x)) \leq \beta(x)$. Suponha que $\beta(x) > \varphi(x,\tilde{v}(x))$. Existe então $\delta > 0$ tal que*

$$\langle \nabla f_i(x), \tilde{v}(x)\rangle \leq \beta(x) - \delta, \quad i=1,\ldots,m.$$

Contradição, pois $\beta(x)$ é valor ótimo de (3-16) por hipótese. Portanto $\beta(x) = \varphi(x,\tilde{v}(x))$. O que mostra que os problemas (3-15) e (3-16) são equivalentes.

2. *Como $\{v; \|v\|_\infty \leq 1\}$ é compacto, o problema (3-15) possui solução pois, $\varphi(x,v)$ é contínua em v.*
3. *Como 0 é elemento do conjunto $\{v; \|v\|_\infty \leq 1\}$, o valor ótimo do problema (3-15) é menor ou igual que zero $(\varphi(x,0)=0)$.*
4. *Se o valor ótimo de (3-15) é negativo, então qualquer vetor v aonde este valor seja alcançado, é direção de descida.*
5. *A escolha da restrição $\|v\|_\infty \leq 1$ pode ser substítuida por uma restrição mais geral: $v \in D$, com $D \subset \mathbb{R}^n$ um conjunto compacto fechado, com $0 \in \int(D)$.*
6. *Notemos que não precisamos especificar a escolha de $\frac{1}{2}\|.\|^2$ como a função que é adicionada à φ definida no problema 3-3, pois qualquer função estritamente convexa fechada e própria $g : \mathbb{R}^n \to \mathbb{R}$ com $g(0)=0$, $g(v)>0$ $(v \neq 0)$ e $g(v) = o(\|v\|)$ $(v \to 0)$ pode ser usada para que o problema de minimização $\min_{v\in\mathbb{R}^n} \varphi(x,v)+g(v)$, tenha solução única.*

Lema 3.16 *Sejam $V(x)$ e $\beta(x)$ o conjunto solução e o valor ótimo do problema 3-15, respectivamente.*

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

$$(JF(x)v)_i \leq \varphi(x,v) < 0\,, i = 1,\ldots,n$$

3) *Se* x^k *converge para* x, $v^k \in V(x^k)$ *e* v^k *converge para* v, *então* $v \in V(x)$.
4) *A aplicação* $x \mapsto \beta(x)$ *é contínua.*

Prova.

1) Suponhamos por absurdo que $0 \notin V(x)$. Seja $v_1 \in V(x)$, então $\varphi(x,v_1) \leq \varphi(x,v)\,, \forall v \in \mathbb{R}^n$, com $||v||_\infty \leq 1$. Em particular para $v = 0$. Assim $\varphi(x,v_1) \leq 0$. Como $0 \notin V(x)$, temos que $v_1 \neq 0$, logo $\varphi(x,v_1) < 0$ e portanto $v_1 \in (ImJf(x) \cap (-\mathbb{R}_+)^m)$, isto é, x não é pareto crítico. Contradição, logo $0 \in V(x)$.
Como acabamos de mostrar que $0 \in V(x)$, logo $\beta(x) = \varphi(x,0) = 0$.
2) Por hipótese $\beta(x)$ é valor ótimo do problema (3-15), logo $\beta(x) \leq \varphi(x,v)$ $\forall v$, com $||v||_\infty \leq 1$, em particular para $\beta(x) \leq \varphi(x,0) = 0$. Como x não é Pareto ótimo, então $ImJf(x) \cap (-\mathbb{R}^m_{++}) \neq \emptyset$. Seja $w \in ImJf(x) \cap (-\mathbb{R}^m_{++})$, temos que $tw \in ImJf(x) \cap (-\mathbb{R}^m_{++})$, $\forall t \geq 0$. Logo

$$\beta(x) \leq \varphi(x,tw), \;\; \forall t \in [0, \frac{1}{||w||}].$$

Como φ é positivamente homogênea, segue que,

$$\beta(x) \leq t\varphi(x,w) < 0, \;\; \forall t \in (0, \frac{1}{||w||}].$$

3) Por hipótese $v^{(k)} \in V(x^k)$, então, $\varphi(x^k,v^k) \leq \varphi(x^k,w)$, $\forall w \in \mathbb{R}^n$, com $||w||_\infty \leq 1$. Assim, aplicando o limite de ambos os lados da desigualdade e usando o fato que a função φ é contínua, temos que:

$$\varphi(\lim_{k\to\infty} x^k, \lim_{k\to\infty} v^k) \leq \varphi(lim_{k\to\infty} x^k, w).$$

Como x^k converge para x e v^k converge para $\bar{v}$, segue que,

$$\varphi(\lim_{k\to\infty} x^k, \lim_{k\to\infty} v^k) = \varphi(x,v) \leq \varphi(x,w) \; \forall w \in \mathbb{R}^n, com \; ||w||_\infty \leq 1.$$

O que implica que, $v \in V(x)$.
4) Seja $x \in \mathbb{R}^n$ e $\{x^k\} \in \mathbb{R}^n$, uma sequência em $\mathbb{R}^n$, tal que $\lim_{k\to\infty} x^k \to x$. Como $\beta(x^k)$ é valor ótimo do problema (2.12), $\beta(x^k) = \varphi(x^k, v(x^k)) \leq \varphi(x^k, v(x))$, assim aplicando o $\limsup$ de ambos os lados da desigualdade e usando o fato de que a

função φ é contínua fixando a segunda variável, temos que:

$$\limsup_{k\to\infty} \beta(x^k) \leq \limsup \varphi(x^k, v(x)) = \varphi(x, v(x)) = \beta(x).$$

Portanto

$$\limsup_{k\to\infty} \beta(x^k) \leq \beta(x). \tag{3-17}$$

Seja $v(x) \in V(x)$ então, $\varphi(x^k, v(x)) \leq \varphi(x^k, v(x^k)) = \beta(x^k) \Rightarrow \varphi(x^k, v(x)) \leq \beta(x^k)$. Aplicando $\liminf$ de ambos os lados da desigualdade e usando o fato que a função φ é contínua quando fixamos a segunda variável, temos que:

$$\beta(x) \leq \liminf \beta(x^k). \tag{3-18}$$

De (3-17) e (3-18) segue que:

$$\limsup \beta(x^k) \leq \beta(x) \leq \liminf \beta(x^k).$$

Portanto $\lim \beta(x^k) = \beta(x)$, ou seja β é uma função contínua.

$\square$

3.7 O caso geral

Este capítulo foi baseado nas idéias de FLIEGE e SVAITER desenvolvidas em [1], onde o problema é encontrar um ponto Pareto ótimo de uma função vetorial com a relação de ordem induzida pelo cone $\mathbb{R}^m_+$, e sem restrições. O chamamos problema de otimização multicritério ou multiobjetivo irrestrito. Podemos também considerar o problema de achar ótimos Pareto de uma função restrita a um conjunto convexo e fechado C, conhecido como problema multiobjetivo com restrições. Neste caso, as funções auxiliares, são definidas segundo:

$$\theta(x) = \min\{\varphi(x^k, v) + \frac{1}{2}||v||^2; v \in C - x^k\},$$

$$v(x^k) = \arg\min\{\varphi(x^k, v) + \frac{1}{2}||v||^2; v \in C - x^k\}.$$

Um problema ainda mais geral é dado quando a relação de ordem é definida por um cone convexo, fechado, pontudo e com interior não vazio K, que não o $\mathbb{R}^m_+$. O chamamos de problema de otimização vetorial, o qual é objeto de estudos em vários artigos como [6] e [11]. Para este problema é necessário definirmos a função φ, segundo:

$$\varphi(x,v) = \max\{y^t Jf(x)v; y \in G\},$$

onde G é um conjunto compacto de geradores normalizados do cone dual. No Lema 3.10, mostramos que as funções auxiliares v e θ são contínuas, e que, x é Pareto crítico se, e somente se, $\theta(x) = 0$. Estas afirmações continuam sendo verdadeiras neste caso.

Se verifica também que se $\varphi(x,v) < 0$, então v é uma direção de descida em x para f, ou seja, existe $\delta \in (0,1)$, tal que

$$x + \alpha v \in C \ e \ f(x+\alpha v) < f(x), \ \alpha \in (0,\delta].$$

CAPÍTULO 4

Algoritmo

Neste capítulo discutimos um método de descida para achar um ponto Pareto crítico de uma função vetorial em $\mathbb{R}^n$, proposto por FLIEGE e SVAITER em [1]. Mostraremos que todo ponto de acumulação da sequência produzida pelo algoritmo é ponto Pareto crítico.

4.1 Método de descida com busca Armijo

O algoritmo precisa de duas constantes, dadas: $\beta \in (0,1]$ e $\sigma \in (0,1]$:

Algoritmo 1
1. *Inicialização: Tome $x^0 \in \mathbb{R}^n$.*
2. *Passo iterativo: Para x^k:*
   (a) *Calculemos $\theta(x^k)$. Se $\theta(x^k) = 0$, então PARE.*
   (b) *Caso contrário, calculemos v^k, uma solução aproximada de (3-3) para $x = x^k$ com tolerância σ.*
   (c) *Calculemos o tamanho do passo $t_k \in (0,1]$ como o máximo de:*

$$T_k := \left\{ t = \frac{1}{2^j} : j \in \mathbb{N}, f(x^k + tv^k) \leq_{\mathbb{R}^m_+} f(x^k) + \beta t Jf(x^k)v^k \right\}.$$

   (d) *Façamos $x^{k+1} = x^k + t_k v^k$.*

Se x^k não é Pareto crítico, então $Jf(x^k)v^k <_{\mathbb{R}^m_+} 0$, pelo Lema 3.10. Portanto, pelo Lema 3.14, o conjunto T_k é não vazio, logo t_k está bem definido. Notemos que a sequência $(f(x^k))_k$ é $\mathbb{R}^m_+$ - não decrescente. Se o algoritmo (1) termina depois de um número finito de iterações, a sequência gerada termina em um ponto x^k, com $\theta(x^k) = 0$, ou seja, que é Pareto crítico, veja Lema (3.10).

4.2 Análise de convergência

De agora em diante, supomos que a sequência gerada pelo o algoritmo (1) é infinita.

Teorema 4.1 *Todos os pontos de acumulação da sequência $(x^k)_k$, gerada pelo algoritmo (1), são Pareto críticos. Se a função f tem conjunto de nível limitado, isto é, o conjunto $L_f(f(x^0)) = \left\{x \in \mathbb{R}^n; f(x) \leq_{\mathbb{R}^m_+} f(x^0)\right\}$ é limitado, então a sequência $(x^k)_k$ é limitada e, portanto tem pontos de acumulação Pareto críticos.*

Prova.

Seja y um ponto de acumulação da sequência $(x^k)_k$ e $(x^{k_\ell})_\ell$ uma subsequência de $(x^k)_k$ convergente a y. É suficiente mostrarmos que $\theta(y) = 0$ (veja o Lema 3.10).

Observemos que $f(x^k + t_k v^k) \leq_{\mathbb{R}^m_+} f(x^k) + \beta t_k Jf(x^k)v^k \, \forall k$. E portanto,

$$f(x^{k_{\ell+1}}) \leq_{\mathbb{R}^m_+} f(x^{k_\ell+1}) = f(x^{k_\ell} + t_{k_\ell} v^{k_\ell}) \leq_{\mathbb{R}^m_+} f(x^{k_\ell}) + \beta t_{k_\ell} Jf(x^{k_\ell})v^{k_\ell}, \forall \ell.$$

Logo, $f(x^{k_\ell}) - f(x^{k_{\ell+1}}) \geq_{\mathbb{R}^m_+} -\beta t_{k_\ell} Jf(x^{k_\ell})v^{k_\ell} \geq_{\mathbb{R}^m_+} 0$ e, consequentemente,

$$\lim_{\ell \to \infty} t_{k_\ell} Jf(x^{k_\ell})v^{k_\ell} = 0. \tag{4-1}$$

Temos dois casos a analisar.

- **Caso 1:** $\limsup_{\ell \to \infty} t_{k_\ell} > 0$.
  Tomemos $(t_{k_u})_u$ subsequência de $(t_{k_\ell})_\ell$ tal que $\lim_{u \to \infty} t_{k_u} = \limsup_{\ell \to \infty} t_{k_\ell} > 0$. Então, por (4-1), temos $\lim_{u \to \infty} Jf(x^{k_u})v^{k_u} = 0$. Logo

  $$\lim_{u \to \infty} \varphi(x^{k_u}, v^{k_u}) = \lim_{u \to \infty} \max_{i=1,\ldots,m} (Jf(x^{k_u} v^{k_u})_i = \max_{i=1,\ldots,m} (\lim_{u \to \infty} Jf(x^{k_u} v^{k_u})_i) = 0.$$

  Como
  $$\varphi(x^{k_u}, v^{k_u}) \leq \varphi(x^{k_u}, v^{k_u}) + \frac{||v^{k_u}||^2}{2} \leq \sigma\theta(x^{k_u}) < 0, \tag{4-2}$$
  obtemos $\theta(y) = 0$, porque θ é contínua como provado no item *iv* do Lema 3.10.
- **Caso 2:** $\limsup_{\ell \to \infty} t_{k_\ell} = 0$.
  Como $x^{k_\ell} \to y$ e f é continuamente diferenciável temos que $(Jf(x^{k_\ell}))_\ell$ é limitada. Logo, pelo Lema 3.12 vale que $(v^{k_\ell})_\ell$ é limitada. Seja $(v^{k_u})_u$ uma subsequência de $(v^{k_\ell})_\ell$ convergente a $\bar{v}$. Note que passando ao limite sobre a subsequência (k_l) em (4-2), obtemos
  $$\frac{1}{\sigma} \max_{i=1,\ldots,m} (Jf(y)\bar{v})_i \leq \theta(y) \leq 0. \tag{4-3}$$
  Seja $q \in \mathbb{N}$. Como $\limsup_{\ell \to \infty} t_{k_\ell} = 0$, para u suficientemente grande, vale que $t_{k_u} < \frac{1}{2^q}$. Pela definição de t_{k_u}, para u suficientemente grande tem

  $$f(x^{k_u} + \frac{1}{2^q} v^{k_u}) \not\leq_{\mathbb{R}^m_+} f(x^{k_u}) + \frac{\beta}{2^q} Jf(x^{k_u})v^{k_u}.$$

Logo para pelo menos um indíce $j \in 1,\ldots,m$ vale

$$f_j(y+\frac{1}{2^q}\bar{v}) \geq f_j(y)+\frac{\beta}{2^q}(Jf(y)\bar{v})_j. \tag{4-4}$$

Observemos que (4-4) vale para todo número natural maior que q. Então, o Lema 3.14 implica que $\max_i(Jf(y)\bar{v})_i \geq 0$ e, por (4-3), concluímos $\theta(y)=0$.

Visto que o conjunto de nível da função f associado a $f(x^0)$ é limitado, e que a função f é contínuamente diferenciável, temos que $L_f(f(x^0))$ é um conjunto compacto. Como $(f(x^k))_k$ é monótona decrescente, temos que $(x^k)_k \subset L_f(f(x^0))$ e, consequentemente, $(x^k)_k$ é limitada. □

4.3 Um método voltado para implementação

O Teorema 4.1 nos garante que, todo ponto de acumulação da sequência $(x^k)_k$ gerada pelo algoritmo (1) é Pareto crítico. Mas na prática o algoritmo deveria terminar em um número finito de passo. Assim, precisamos de outro critério de parada que não o que é usado teoricamente no passo iterativo do algoritmo (4.1). Um critério de parada aceitável e executavel é: Parar se $\theta(x^k) > -\tau$, onde $\tau > 0$ pequeno.

A direção de descida $v(x^k)$, é solução aproximada do problema de otimização auxiliar $3-3$ segundo a Definição 3.11. Notemos que essa Definição envolve o valor exato do minímo $\theta(x^k)$. Passando então, a uma implementação, se faz necessário substituír $\theta(x^k)$ por alguma cota superior aceitável.

Para isto usamos o problema dual. Como sabemos, o dual é dado pelo máximo do ínfimo do lagrangiano. Observamos que o lagrangiano de $(3-3)$ é dado por,

$$L((\alpha,v),\lambda) = f(x)+\sum_{i=1}^{m}\lambda_i[\alpha - Jf(x)v],$$

logo

$$\nabla L_{(\alpha,v)}((\alpha,v),\lambda)=0 \Leftrightarrow (1-\sum_{i=1}^{m}\lambda_i, v+Jf(x)^T\lambda)=0.$$

Portanto $\sum_{i=1}^{m}\lambda_i = 1$ e $Jf(x)^T\lambda = -v$. Assim temos que o dual de (3.3) é dado pelo problema:

$$\begin{cases} \max & -\frac{1}{2}||\sum_{i=1}^{m}\lambda_i\nabla f_i(x)||^2 \\ s.a & \\ & \sum_{i=1}^{m}\lambda_i = 1, \\ & \lambda \geq 0 \end{cases} \tag{4-5}$$

O problema (4-5) é um problema com função objetivo quadrática e com restrições lineares. Logo, pode-se obter uma boa aproximação do valor máximo sem muito esforço computacional e esta aproximação do valor máximo é uma cota superior de $\theta(x^k)$. Assim, podemos substituir os passos $ii(a)$ e $ii(b)$ do algoritmo por:

- Calculamos uma aproximação do valor ótimo de (4-5) em $x = x^k$, que chamamos de θ_k. Se $\theta_k > -\tau$, paramos.
- Calculamos v^k tal que $\varphi(x_k, v) + \frac{1}{2}||v||^2 \leq \sigma\theta_k$.

4.4 Caso restrito

Como observamos na Seção 2.2 do Capítulo 2, podemos considerar um problema com restrições, isto é, problemas do tipo:

$$(P)\begin{cases} \min_{\mathbb{R}^m_+} & f(x) \\ s.a & x \in G \end{cases}$$

onde $G \subset \mathbb{R}^n$ e $f : \mathbb{R}^n \to \mathbb{R}^m$. E minimizar esta função vetorial, consiste em encontar pontos Pareto ótimos, isto é, pontos $z \in G$, tal que não exista um outro ponto $y \in G$ com $f(y) \leq f(z)$ e $f(y) \neq f(z)$. Para isto, existem condições que nos garantem que um dado ponto z seja Pareto; são as chamadas condições de otimalidade. Para desenvolver estas condições, precisamos estudar não só o comportamento da função objetivo numa vizinhança de z, mas também a estrutura do conjunto viável nessa vizinhança. As noções de direção tangente e cone tangente são fundamentais neste contexto e importantes para um entendimento correto das condições de otimalidade.

Definição 4.2 *Dado $z \in G$, um vetor $v \in \mathbb{R}^n$ é dito direção tangente de G em z se existe uma sequência $(z^k)_k \subseteq G$ e um escalar $\lambda \in \mathbb{R}_+$ tal que:*

$$lim_{k\to\infty} z^k = z \ e \ lim_{k\to\infty} \lambda \frac{z^k - z}{||z^k - z||} = v.$$

O conjunto de todas as direções tangentes de G em z é chamado cone tangente (de Bouligard) G em z e o denotaremos por $T(G,z)$.

Dado $G = \{x \in \mathbb{R}^n; g_j(x) \leq 0, j = 1, ..., l\}$, onde as funções $g_j : \mathbb{R}^n \to \mathbb{R}$, $1 \leq j \leq l$ são funções continuamente diferenciáveis, e $I_0(z) = \{j; g_j(z) = 0\}$, o conjunto de restrições ativas no ponto viável z, definimos *cone linearizado* de G em z por,

$$C(z) = \{v \in \mathbb{R}^n; \nabla g_j(z)^T v \leq 0 \ \forall j \in I_0(z)\}.$$

Suponhamos que $C(z) = T(G,z)$. Sabemos, pelo Teorema 3.7, que, se z é Pareto crítico, então, para todo $v \in \mathbb{R}^n$, existe um $i = 1,\dots,m$ *tal que* $\langle \nabla f_i(z), v\rangle \geq 0$, em particular isso vale para $v \in T(G,z)$. Assim, a condição necessária para um ponto $z \in G$ ser Pareto crítico, $Jf(z)(G-z) \cap (-\mathbb{R}^m_{++}) = \emptyset$, pode ser escrita como

$$Jf(z)T(G,z) \cap (-\mathbb{R}^m_{++}) = \emptyset \tag{4-6}$$

que é equivalente a que o sistema

$$\begin{cases} (\nabla f_i(^x))^T v < 0 \ ,\forall i = 1,\dots,m \\ \\ (\nabla g_j(x))^T v \leq 0 \ ,\forall j \in I_0(z) \end{cases} \tag{4-7}$$

não admita solução $v \in T(G,z)$.

Consideremos agora o seguinte problema,

$$\begin{cases} \qquad \min \alpha \\ s.a. \\ \qquad (\nabla f_i(x))^T v \leq \alpha, \quad i = 1,\dots,m \\ \qquad (\nabla g_j(x))^T v \leq \alpha, \quad \forall j \in I_\varepsilon(x), \\ \qquad ||v||_\infty \leq 1 \end{cases} \tag{4-8}$$

onde, $I_\varepsilon(x) := \{j \in I; g_j(x) \geq -\varepsilon\}$ e $I := \{1,\dots,l\}$.

Denotamos por $\alpha(x,\varepsilon)$, o valor ótimo de (4-8). E definimos

$$(PA)\begin{cases} \qquad \min \alpha \\ s.a. \\ \qquad \langle \nabla f_i(x), v\rangle \leq \alpha, \quad \forall i = 1,\dots,m \\ \qquad \left\langle \nabla g_j(x), v\right\rangle \leq \alpha, \quad \forall j \in I_\varepsilon(x) \end{cases},$$

o problema auxiliar de (4-8). Observemos que o valor minímo de (PA) é não positivo, pois $(0,0)$ é sempre possível e que os problemas (4-8) e (PA) são equivalentes. De fato, suponhamos que $(\bar{v},\bar{\theta})$ é solução de (PA), com $||v||_\infty > 1$. Tomemos $\left(\frac{\bar{v}}{||v||_\infty}, \bar{\theta}\right)$. Notemos que este ponto é possível para (4-8), pois:

$$\left\langle \nabla fi(x), \frac{v}{||v||_\infty}\right\rangle = \frac{1}{||v||_\infty}\langle \nabla f_i(x), \bar{v}\rangle < \left\langle f_i'(x), \bar{v}\right\rangle \leq \bar{\theta} \ \forall i,$$

de forma análoga obtemos que $\left\langle \nabla g_j(x), \frac{\bar{v}}{||v||_\infty}\right\rangle \leq \bar{\theta}$. Suponhamos agora que (w,β) é possível para $(4-8)$, com $\beta < \bar{\theta}$, então (w,β) é possível para (PA) e como $\beta < \bar{\theta}$ teremos que $\bar{\theta}$ não é valor ótimo de (PA).

O método de máxima descida com regra de Armijo para resolver o problema de otimização restrito (4-8) é como segue.

Dadas as constantes $\beta \in (0,1)$, $\varepsilon_1 > 0$.

Algoritmo 2 *1 Inicialização: Tome* $x^1 \in G$.

2 Passo iterativo: Para $k = 1, \ldots$:

- *(a) Resolvemos* $(4-8)$, *com isso calculamos a direção* v^k *e o valor ótimo* $\theta(x^k, \varepsilon_k)$ *para o problema auxiliar.*
- *(b) Se* $\theta(x^k, \varepsilon_k) \geq -\varepsilon_k$, *fazemos* $\varepsilon_{k+1} := \frac{\varepsilon_k}{2}$, $x^{k+1} := x^k$, $k := k+1$, *e voltamos o passo a.*
- *(c) Caso contrário, calculamos o tamanho do passo* $t_k \in (0,1]$ *tal que:*

$$f(x^k + t_k v^k) \leq f(x^k) + \beta t_k J f(x^k) v^k$$

  e $x^k + t_k v^k \in G$.
- *(d) Fazemos* $x^{k+1} := x^k + t_k v^k$, $k := k+1$,*e voltamos ao passo a.*

O teorema de convergência é o seguinte:

Teorema 4.3 *Seja* Jf *e* ∇g_i *com* $i \in I$ *Lipshitz contínua e seja a sequência* $(x^k)_k$ *gerada pelo algoritmo (2) limitada. Então a sequência* $(x^k)_k$ *possui um ponto de acumulação* y *com* $\theta(y, 0) = 0$.

Veja a demonstração do Teorema 4.3 em [1]. Observemos que quando f possui conjunto de nível limitado, a sequência $(x^k)_k$ é necessariamente limitada. Como no caso irrestrito, não é necessário resolvermos $(4-8)$ exatamente. Em vez disso, é suficiente para paramos que o valor da função θ exista com $\theta < -\varepsilon_k$. Neste caso $\theta(x^k, \varepsilon_k) \geq -\varepsilon_k$ asegura que o valor ótimo da função $\theta(x^k, \varepsilon_k)$, pode ser usado para parar.

4.5 O caso geral

Como foi comentado na Seção 2.2 do Capítulo 2, podemos estudar o problema

$$\begin{cases} \min_K & f(v) \\ s.a & v \in C \end{cases} \tag{4-9}$$

onde f é uma função vetorial e C é um conjunto convexo fechado, e com a relação de ordem induzida por um cone qualquer convexo fechado pontudo e com interior não vazio. Neste caso, usamos o método do gradiente projetado para o caso vetorial proposto por GRAÑA DRUMMOND e IUSEM, em [6] para resolvermos o problema restrito.

O método consiste em tomarmos a direção de descida da forma

$$v^k = \arg\min_{v\in C_k}\left\{\frac{||v||^2}{2} + \beta_k \max_{y\in G}(y^t Jf(x^k)v\right\} \tag{4-10}$$

e fazermos $x^{k+1} = x^k + \alpha_k v^k$, onde a função φ é a mesma definida ao final do Capítulo 2, na Seção (2.7), $C_k = \{u - x^k; u \in C\}$, $\beta_k > 0$ e $\alpha_k \subset (0,1]$ é encontrado fazendo a busca Armijo.

Observemos que, se $(x^k)_k$ é uma sequência gerada pelo método acima, então $(x^k)_k \subset C$. De fato, o ponto inicial x^0, está em C pela a inicialização do método. Assumimos que $x^k \in C$, então pela equação (4-10), temos que $v^k \in C_k$, assim podemos expressar $v^k = x - x^k$ e consequentemente $x^{k+1} = x^k + \alpha_k(x - x^k)$, com $x \in C$. Notemos que $x^k + \alpha_k(x - x^k) = (1 - \alpha_k)x^k + \alpha_k x$. Como x^k *e* $x \in C$ e C é convexo, temos que $(1 - \alpha_k) + \alpha_k x \in C$, portanto $x^{k+1} = x^k + \alpha_k(x - x^k) \in C$.

Neste caso continua valendo um resultado semelhante ao Teorema 4.1, ou seja, todo ponto de acumulação da sequência gerada pelo método acima é ponto Pareto crítico. A demonstração é análoga a demonstração de Teorema 4.1 - veja .

Consideremos agora o método da projeção do gradiente sob a hipótese que a função f é K-convexa. Neste caso nós substituímos a equação (4-10) por

$$v^k = \arg\min_{v\in C_k}\left\{\frac{||v||^2}{2} + \frac{\beta_k}{\eta_k} \max_{y\in G}(y^t Jf(x^k)v\right\}, \tag{4-11}$$

com passos exógenos, $(\beta_k)_k \subset \mathbb{R}_{++}$, satisfazendo

$$\sum_{k=0}^{\infty} \beta_k = \infty, \quad \sum_{k=0}^{\infty} \beta_k^2 < \infty \tag{4-12}$$

$$\eta_k = \max_{y\in G} ||Jf(x^k)^t y||, \tag{4-13}$$

onde $C_k = \{u - x^k; u \in C\}$ e G é o conjunto de geradores normalizados do cone polar K^*.

Dada a constante $\sigma \in [0,1)$.

Algoritmo 3 *1 Inicialização: Tomemos* $x^0 \in C$.

2 Passo iterativo: Para $k = 0, 1, ...$*:*

(a) Se x^k *é K-crítico (ou* $\eta_k = 0$*) paramos.*

(b) Caso contrário, calculamos v^k *sobre* C *partindo de* x^k *e dado por (4-11), (4-12) e (4-13)*

(c) Façamos uma busca Armijo $j \in \mathbb{Z}_+$*, tal que*

$$l(k) = \min\{j \in \mathbb{Z}_+ : f(x^k + 2^{-j}v^k) \leq_k f(x^k) + \sigma 2^j Jf(x^k)v^k\}$$

e tomemos $\alpha_k = 2^{-l(k)}$

(d) Façamos $x^{k+1} = x^k + \alpha_k v^k$.

Observamos que a sequência $(x^k)_k$ produzida pelo algoritmo (3) é viável. De fato, a iteração inicial $x^0 \in C$ pela inicialização do algoritmo. Assumimos que $x^k \in C$ e como $\alpha_k \in (0,1]$ pelo passo (c), e $v^k \in C_k$ por (4-11). Logo pela convexidade de C concluímos que $x^{k+1} = x^k + \alpha_k v^k = \alpha_k(v^k + x^k) + (1-\alpha_k)x^k \in C$.

Se a função $f : \mathbb{R}^n \to \mathbb{R}^m$ é K-convexa e $T = \{x \in C; f(x) \leq f(x_k)\ \forall k\}$. é não vazio, então a sequência produzida pelo algoritmo (3) converge a uma solução fracamente eficiente do problema (4-9). A prova deste resultado pode ser encontrada no artigo de GRAÑA DRUMMOND e IUSEM, [6].

CAPÍTULO 5

Conclusão

Segundo a referência [8], problemas de otimização vetorial aparecem implícitos em vários processos de tomada de decisões, quando somos confrotados com fins conflitantes. Neste contexto, um primeiro desafio é definir qual seria a melhor decisão a tomar. Historicamente, a primeira definição satisfatória de uma solução ótima para (POV) foi formulada por um matemático e economista italiano chamado V. Pareto, em [9]. Nesta dissertação usamos dois conceitos aceitos atualmente: soluções Pareto e soluções fracamente Pareto.

As direções de descida para f em um ponto x, isto é, as direções em que todo passo suficientemente curto a partir de x nesta direção resulta em algum decréscimo da função f foram o objeto de estudo desta dissertação. Vale ressaltar que nos concentramos no caso mais simples: o caso sem restrições e com a ordem induzida por $\mathbb{R}^m_+$. Ao final do Capítulo 3 apresentamos brevemente a extensão para o caso geral. No Capítulo 4 apresentamos um método de descida. Todo ponto de acumulação da sequência gerada por este método é ponto Pareto crítico. Para isto a única hipótese colocada é que a função objetivo seja de classe C^1. No caso mais geral, se a função f do problema (4-9) é K-convexa e $T = \{x \in C; f(x) \leq f(x^k)\ \forall k\}$ é não vazio, a sequência produzida pelo algoritmo (3) converge a uma solução fracamente eficiente.

Referências Bibliográficas

[1] J. Fliege e B.F. Svaiter, “Steepest Descent Methods for Multicriteria Optimization", Mathematical Methods of Operations Research, vol.51, pp.479-494, 2000.

[2] E.L. Lima, “Curso de Análise", Volume 1, Rio de Janeiro, Impa 2002.

[3] E.L. Lima, “Curso de Análise", Volume 2, Rio de Janeiro, Impa 2005.

[4] A. Izmailov e S. Mikhail, “Otimização, volume 1, Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade", Rio de Janeiro, Impa 2005.

[5] B. Stephen e V. Lieven, “Convex Optimization", Cambridge University, 2004.

[6] L.M. Graña Drummond e A.N.Iusem, “A Projected Gradient Method for Vector Optimization Problems", Computational Optimization and Applications, 28, 5-29, 2004.

[7] G. C. Bento, “Métodos para Otimização em Variedades Riemannianas: Gradiente para Funções Vetoriais e Proximal Local para Funções Reais", Tese de Doutorado, PESC/COPPE-UFRJ, Rio de Janeiro 2009.

[8] G. Giorgi, A. Guerraggio e J. Thierfelder “ Mathematics of optimization: smooth and nonsmooth case". Elsevier Science B.V., Amsterdam, 2004.

[9] V. Pareto, “Manual of Political Economic". Tradução para o inglés de 1971 de “Manuale di Economia Politica", 1906.

[10] Yurii Nesterov, “Introductory Lectures on Convex Optimization. A basic course". Applied Optimization, 87. Kluwer Academic Publishers, Boston, MA, 2004.