



UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PÓS-GRADUAÇÃO EM MATEMÁTICA
MESTRADO EM MATEMÁTICA

**Um Método de Descida em Otimização
Multiobjetivo**

Felipe Marreiros Mesquita

Teresina - 2013

Felipe Marreiros Mesquita

Dissertação de Mestrado:

Um Método de Descida em Otimização Multiobjetivo

Dissertação submetida à Coordenação do Curso de Pós-Graduação em Matemática, da Universidade Federal do Piauí, como requisito parcial para obtenção do grau de Mestre em Matemática.

Orientadora:
Prof^a. Dr^a. Sissy da Silva Souza

Teresina - 2013

xxxx Mesquita ,F M.
Um Método de Descida em Otimização Multiobjetivo

Felipe Marreiros Mesquita – Teresina: 2013.

Orientadora: Prof^a. Dr^a.Sissy da Silva Souza.

1. Otimização

CDD 516.36

Agradecimentos

Inicialmente agradeço a Deus por ter me ajudado em mais esta conquista, agradeço por ter tido saúde, apoio e muita determinação para completar mais esta etapa de estudos. Agradeço aos meus pais Dorutéa e Jacob, aos meus irmãos Thyago e Wlysses, à minha esposa Rosa Neide que me apoiou bastante e me deu muita força e ao meu amigo de promessa Amando. Agradeço também aos meus professores de graduação Afonso Norberto, Alessandro Wilk e Halyson Baltazar, aos professores de mestrado Paulo Sérgio, Paulo Alexandre, Jurandir Lopes, Barnabé Pessoa Lima, Carlos Humberto e Marcos Vinícius. Agradeço especialmente à professora Sissy da Silva Souza por ter aceitado ser minha orientadora, por ter tido paciência e ter me ajudado nestes dois anos de mestrado, ela foi junto com os professores Paulo Sérgio e Jurandir boas amizades que fiz ao longo destes dois anos neste departamento, tenho muito a agradecer à eles. Agradeço ao meus colegas de graduação, em especial ao meu amigo Valdinês pelas inúmeras ajudas. Agradeço à todos meus amigos de mestrado, Irrael, Mikael, Bernardo, Diego, Edvalter, Valdir, Jefferson, Renata, Ailton, Samara, Leonardo, Vitaliano e Baiano por terem dividido o sofrimento e a dureza destes dois anos. Agradeço aos professores da Banca examinadora, Roberto Cristovão, Afonso Norberto, Jurandir Lopes e novamente à minha orientadora Sissy Souza. Agradeço à CAPES, ao CNPq e à FAPEPI pelo apoio financeiro.

*Aos meus pais, à minha esposa, aos meus irmãos e
todos meus amigos.*

*“Somos quem somos, pelas amizades que
construímos”.*

Resumo

No cotidiano, diversas situações surgem de maneira simultânea e é natural querermos as escolhas melhores possíveis. Tais cenários são comuns em várias áreas de conhecimentos como economia e engenharias e, são modelados como problemas de funções vetoriais de várias variáveis. Nesta dissertação estudaremos métodos de descida para resolver problemas de otimização multiobjetivo irrestrito e, também, problemas de otimização vetorial. Em ambos os casos, as funções objetivo consideradas são continuamente diferenciáveis e convexas; além disso, a busca linear considerada é do tipo Armijo para calcular o tamanho do passo de decrescimento. Mostraremos que o método converge para um ponto satisfazendo certas condições necessárias de primeira ordem para otimalidade Pareto no caso multiobjetivo e, no caso vetorial obtemos que a sequência gerada pelo método converge para um ponto K-crítico.

Abstract

In everyday life, many situations arise simultaneously and it is natural to want the best possible choices. Such scenarios are common in various areas of knowledge such as economics and engineering, and problems are modeled as vector functions of several variables. In this dissertation study descent methods for solving unconstrained multi-objective optimization problems and also problems of vector optimization. In both cases, the objective functions are considered continuously differentiable convex and, in addition, the linear search type Armijo is considered to calculate the step size decrease. Show that the method converges to a point satisfying certain conditions of Pareto optimality first order multiobjective the case, and if we obtain the vector sequence generated by the method converges to a point K-critical.

Sumário

Resumo	iii
Abstract	iv
Notações	1
Introdução	2
1 Preliminares	4
1.1 Definições e Resultados Básicos para Minimização em \mathbb{R}^n	4
1.1.1 Fatos Básicos	4
1.1.2 Convexidade de Funções	6
1.1.3 Subdiferencial de Funções Convexas	9
1.1.4 Espaços de Matrizes ($\mathbb{R}^{m \times n}$)	11
1.2 Definições e Resultados Básicos para Otimização Multiobjetivo	15
1.3 Definições e Resultados Básicos Para Otimização Vetorial	19
1.3.1 Definições Básicas	19
1.3.2 Introdução a Otimização Vetorial	21
2 Método de Descida em Otimização Multiobjetivo	23
2.1 Algoritmo Geral	23
2.2 Direção de Busca	23
2.3 Tamanho do Passo	29
2.4 Definição e Análise de Convergência do Algoritmo	30
2.4.1 Algoritmo Completo	30
2.4.2 Análise de Convergência do Algoritmo	31
3 Método de Descida em Otimização Vetorial	35
3.1 Introdução	35

3.2	Tamanho do Passo e Direção de Busca	37
3.3	Análise de Convergência do Algoritmo	51
	Referências Bibliográficas	57

Notações

\mathbb{R}^n	Espaço Euclidiano n-dimensional.
\mathbb{R}_+^n	Ortante não negativo de \mathbb{R}^n .
\mathbb{R}_{++}^n	Interior de \mathbb{R}_+^n .
$\text{conv}(\mathbf{U})$	Fecho convexo do conjunto \mathbf{U} .
$\text{cone}(\mathbf{U})$	Cone gerado por \mathbf{U} .
\mathbf{K}	Cone.
\mathbf{K}^*	Cone polar positivo do cone \mathbf{K} .
$\text{int}\mathbf{K}$	Interior do cone \mathbf{K} .
$\text{int}\mathbf{K}^*$	Interior do cone polar positivo \mathbf{K}^* .
$x \preceq_{\mathbf{K}} y$	$y - x \in \mathbf{K}$.
$x \prec_{\mathbf{K}} y$	$y - x \in \text{int}\mathbf{K}$.
$\text{Im}(F)$	Conjunto imagem de F .
$JF(x)$	Jacobiano da função F no ponto x .
$\nabla F_i(x)$	Gradiente da função F_i em x .
$\partial f(x)$	Subdiferencial de f em x .
argmin	Conjunto dos pontos minimizadores.
\min	Mínimo.
\max	Máximo.
AKdG	Algoritmo de K-descida Genérico.
AKd	Algoritmo de K-descida.

Introdução

Nesta dissertação estudaremos a minimização de funções $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ convexas e continuamente diferenciáveis sob duas abordagens. Trataremos inicialmente o problema de encontrar um ponto crítico Pareto para F , isto é, um ponto $\bar{x} \in \mathbb{R}^n$ tal que

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

onde $\text{Im}JF(\bar{x})$ é a imagem da aplicação jacobiano de F em \bar{x} . Para isso utilizaremos um algoritmo de descida cuja definição e análise de convergência apresentaremos no Capítulo 2. Esse problema é conhecido na literatura como problema de otimização multiobjetivo irrestrito.

Diversos problemas da vida real são formulados como problemas de otimização multiobjetivo, como podem ser vistos na teoria econômica e engenharias. Daí a importância de estudar estes problemas e a existência de diversos métodos para resolvê-los, dentre eles: método de Newton como o estudado por Fliege, Grana Drummond e Svaiter em [5], método de Região de Confiança como o estudado por Villacorta e Oliveira em [20], método do Subgradiente como o trabalho de Da Cruz Neto et al em [2] e método de Descida trabalhado por exemplo por Fliege e Svaiter em [4] no qual nossa abordagem de otimização multiobjetivo está baseada.

Nossa outra abordagem, também muito importante, é a da otimização vetorial. Neste caso, consideraremos o problema de encontrar um ponto K-crítico, isto é, um ponto $\bar{x} \in \mathbb{R}^n$ tal que

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

Para solucionar os problemas de otimização vetorial podemos utilizar vários métodos de resolução, tais como: método de ponto proximal como estudado por Villacorta e Oliveira em [19], método do gradiente projetado como o estudado por Grana Drummond e Iusem em [7] e também método de descida trabalhado por exemplo por Grana Drummond e Svaiter em [6] no qual está baseado nossa abordagem de minimização vetorial. Sendo assim, uma definição importante é a de

K -direção de descida, isto é, $v \in \mathbb{R}^n$ é uma K -direção de descida em $x \in \mathbb{R}^n$ quando $JF(x)v \preceq_K 0$, onde \preceq_K representa a ordenação parcial induzida por um cone $K \subset \mathbb{R}^m$ convexo, fechado e pontiagudo com interior diferente do vazio.

Esta dissertação está dividida como segue. No primeiro capítulo, apresentaremos as definições e resultados básicos preliminares para o desenvolvimento deste trabalho. No segundo capítulo, estudaremos um método de descida para a resolução de um problema de otimização multiobjetivo irrestrito. No terceiro capítulo, abordaremos um método de descida para otimização vetorial que generaliza o método visto no Capítulo 2, para um cone $K \subset \mathbb{R}^n$ convexo, fechado e pontiagudo com interior diferente do vazio.

Capítulo 1

Preliminares

Neste capítulo apresentaremos definições e resultados que serão essenciais para compreendermos o estudo de problemas de otimização tanto para problemas de minimização de funções escalares de várias variáveis como também problemas de otimização multiobjetivo e otimização vetorial. Para melhor compreensão de algumas definições e resultados, apresentaremos alguns exemplos e gráficos. Informações mais detalhadas poderão ser encontradas em [8], [14], [17] e [19].

1.1 Definições e Resultados Básicos para Minimização em \mathbb{R}^n

Nesta seção consideraremos o problema irrestrito

$$\begin{array}{ll} \min & f(x) \\ \text{s.a} & x \in \mathbb{R}^n, \end{array} \quad (1.1)$$

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

1.1.1 Fatos Básicos

Definição 1.1.1 ([8]). *Dizemos que um ponto $x \in \mathbb{R}^n$ é*

1. *Minimizador global de (1.1) se*

$$f(x) \leq f(y) \quad \forall x \in \mathbb{R}^n;$$

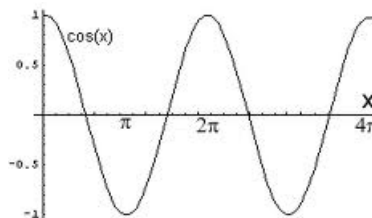
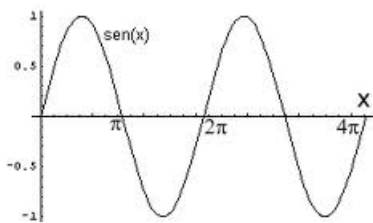
2. **Minimizador local** de (1.1) se existe uma vizinhança U de x tal que

$$f(x) \leq f(y) \quad \forall y \in U \cap \mathbb{R}^n.$$

Se x é minimizador global então $f(x)$ é chamado **valor ótimo global**.

Observação 1.1.1 ([8]). *Uma função pode admitir vários minimizadores globais, mas o valor ótimo global do problema sempre é o mesmo.*

Exemplo 1.1.1. *As funções $\sin(x)$ e $\cos(x)$, $x \in \mathbb{R}$.*



Definição 1.1.2. *Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é dita **homogênea de grau n** quando*

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

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

Proposição 1.1.1. *Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é o máximo entre funções lineares então f é homogênea de grau 1.*

Demonstração. Seja $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função linear e seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$f(x) = \max_{i=1,2,\dots,m} \{g_i(x) \mid x \in \mathbb{R}^n\}.$$

Pela linearidade de g_i

$$\begin{aligned} f(\lambda x) &= \max_{i=1,2,\dots,m} \{g_i(\lambda x) \mid x \in \mathbb{R}^n\} \\ &= \max_{i=1,2,\dots,m} \{\lambda g_i(x) \mid x \in \mathbb{R}^n\} \\ &= \lambda \max_{i=1,2,\dots,m} \{g_i(x) \mid x \in \mathbb{R}^n\} \\ &= \lambda f(x), \end{aligned}$$

portanto f é homogênea de grau 1. □

1.1.2 Convexidade de Funções

Definição 1.1.3 ([8]). *Uma função $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é dita*

1. **convexa** quando para todos $x, y \in \mathbb{R}^n$ tivermos

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

para todo $t \in [0, 1]$.

2. **estritamente convexa** quando para todos $x, y \in \mathbb{R}^n$ com $x \neq y$,

$$f((1-t)x + ty) < (1-t)f(x) + tf(y)$$

para todo $t \in (0, 1)$.

3. **fortemente convexa** com módulo $\gamma > 0$ quando para todos $x, y \in \mathbb{R}^n$ tivermos

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) - \gamma t(1-t)\|x - y\|^2$$

para todo $t \in [0, 1]$.

Naturalmente, toda função fortemente convexa é estritamente convexa e consequentemente convexa.

Exemplo 1.1.2 ([8]). *A função $f: \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = x^2$ é uma função fortemente convexa com módulo $\gamma = 1$.*

Exemplo 1.1.3 ([16]). *A função $f: \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = e^x$ é estritamente convexa (mas não fortemente) convexa.*

Para mostrar isso, notemos primeiramente que $e^x \geq 1 + x \forall x \in \mathbb{R}$. Com efeito, definimos a função auxiliar $f: \mathbb{R} \rightarrow \mathbb{R}$ dada por $f(x) = e^x - x$. Temos

$$x \geq 0 \Rightarrow f'(x) = e^x - 1 \geq 0.$$

Logo f é crescente para $x \geq 0$, assim

$$x \geq 0 \Rightarrow f(x) \geq f(0) \Rightarrow e^x - x \geq 1 \Rightarrow e^x \geq 1 + x.$$

Analogamente,

$$x \leq 0 \Rightarrow f'(x) = e^x - 1 \leq 0.$$

Logo f é decrescente para $x \leq 0$, assim

$$x \leq 0 \Rightarrow f(0) \leq f(x) \Rightarrow 1 \leq e^x - x \Rightarrow 1 + x \leq e^x.$$

Agora, para $z = (1 - t)x + ty$ temos

$$e^{x-z} \geq 1 + x - z \Rightarrow e^x \geq e^z + e^z(x - z), \quad (1.2)$$

e de maneira análoga

$$e^y \geq e^z + e^z(y - z). \quad (1.3)$$

Multiplicando (1.2) por $(1 - t)$ e (1.3) por t e somando ambas temos que

$$e^z \leq (1 - t)e^x + te^y \quad \forall t \in [0, 1].$$

Logo

$$e^z < (1 - t)e^x + te^y \quad \forall t \in (0, 1).$$

Portanto a função $f(x) = e^x$ é estritamente convexa mas, não fortemente convexa.

Exemplo 1.1.4. A função $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ dada por $f(x) = 2x_1 + x_2 - x_3$ é convexa (mas não estritamente convexa).

De fato, dados $x = (x_1, x_2, x_3)$ e $y = (y_1, y_2, y_3) \in \mathbb{R}^n$ temos

$$\begin{aligned} f((1 - t)x + ty) &= 2((1 - t)x_1 + ty_1) + (1 - t)x_2 + ty_2 - (1 - t)x_3 - ty_3 \\ &= (1 - t)2x_1 + t2y_1 + (1 - t)x_2 + ty_2 - (1 - t)x_3 - ty_3 \\ &= (1 - t)(2x_1 + x_2 - x_3) + t(2y_1 + y_2 - y_3) \\ &= (1 - t)f(x) + tf(y). \end{aligned}$$

Logo f é convexa mas, não estritamente convexa.

Exemplo 1.1.5. A função $f: \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f(x) = \|x\|^2$ é fortemente convexa. De fato,

$$\begin{aligned}
 f((1-t)x + ty) &= \|(1-t)x + ty\|^2 \\
 &\leq (1-t)^2\|x\|^2 + 2t(1-t)\|x\|\|y\| + t^2\|y\|^2 \\
 &= (1-t)\|x\|^2 + (t-1)\|x\|^2 + \|x\|^2 - 2t\|x\|^2 \\
 &\quad + t^2\|x\|^2 + 2t(1-t)\|x\|\|y\| + t\|y\|^2 - t\|y\|^2 + t^2\|y\|^2 \\
 &= (1-t)f(x) + tf(y) + t\|x\|^2(t-1) + t\|y\|^2(t-1) \\
 &\quad + 2t(1-t)\|x\|\|y\| \\
 &= (1-t)f(x) + tf(y) - t(1-t)(\|x\|^2 - 2\|x\|\|y\| + \|y\|^2) \\
 &= (1-t)f(x) + tf(y) - t(1-t)(\|x\| - \|y\|)^2 \\
 &\leq (1-t)f(x) + tf(y) - t(1-t)\|x - y\|^2.
 \end{aligned}$$

Proposição 1.1.2. Se $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa então todo minimizador local é global.

Demonstração. Suponhamos que x^* seja mínimo local porém não global, então existe $y \in \mathbb{R}^n$ tal que $f(y) < f(x^*)$. Como f é convexa temos que

$$\begin{aligned}
 f((1-t)y + tx^*) &\leq (1-t)f(y) + tf(x^*) \\
 &< (1-t)f(x^*) + tf(x^*) \\
 &= f(x^*), \quad \forall t \in [0, 1].
 \end{aligned}$$

Logo é uma contradição pois x^* é mínimo local. Portanto se f é convexa todo mínimo local é global. □

Proposição 1.1.3. Se $f: \mathbb{R}^n \rightarrow \mathbb{R}$ é o máximo entre funções lineares então f é convexa.

Demonstração. Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$f(x) = \max_{i=1,2,\dots,m} \{g_i(x) \mid x \in \mathbb{R}^n\}$$

onde cada $g_i: \mathbb{R}^n \rightarrow \mathbb{R}$ são funções lineares. Então, para todos $x, y \in \mathbb{R}^n$ temos

$$\begin{aligned}
 f((1-t)x + ty) &= \max_{i=1,2,\dots,m} \{g_i((1-t)x + ty)\} \\
 &= \max_{i=1,2,\dots,m} \{(1-t)g_i(x) + tg_i(y)\} \\
 &\leq \max_{i=1,2,\dots,m} (1-t)g_i(x) + \max_{i=1,2,\dots,m} tg_i(y) \\
 &= (1-t) \max_{i=1,2,\dots,m} g_i(x) + t \max_{i=1,2,\dots,m} g_i(y) \\
 &= (1-t)f(x) + tf(y).
 \end{aligned}$$

Portanto f é convexa. □

Definição 1.1.4. *Uma função $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, onde $F = (F_1, F_2, \dots, F_m)$, é dita convexa quando cada $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ for convexa.*

Exemplo 1.1.6. *Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ convexa e continuamente diferenciável. A função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $f_x(\mathbf{v}) = \max\{(\mathbf{J}F(\mathbf{x})\mathbf{v})_i \mid i = 1, 2, \dots, m\}$, onde $\mathbf{x} \in \mathbb{R}^n$ é fixo porém arbitrário, é convexa.*

De fato, isto segue imediatamente da Proposição 1.1.3 pois f_x é o máximo de uma função linear.

Proposição 1.1.4. *Sejam $f, g : \mathbb{R}^n \rightarrow \mathbb{R}$ tais que f é convexa e g é fortemente convexa com módulo $\gamma > 0$, então a função $h : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que $h = f + g$ é fortemente convexa com módulo $\gamma > 0$.*

Demonstração. Com efeito, sendo f convexa e g fortemente convexa temos para todos $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ que

$$\begin{aligned} h((1-t)\mathbf{x} + t\mathbf{y}) &= f((1-t)\mathbf{x} + t\mathbf{y}) + g((1-t)\mathbf{x} + t\mathbf{y}) \\ &\leq (1-t)f(\mathbf{x}) + tf(\mathbf{y}) + (1-t)g(\mathbf{x}) + tg(\mathbf{y}) - \gamma t(1-t)\|\mathbf{x} - \mathbf{y}\|^2 \\ &= (1-t)(f(\mathbf{x}) + g(\mathbf{x})) + t(f(\mathbf{y}) + g(\mathbf{y})) - \gamma t(1-t)\|\mathbf{x} - \mathbf{y}\|^2 \\ &= (1-t)h(\mathbf{x}) + th(\mathbf{y}) - \gamma t(1-t)\|\mathbf{x} - \mathbf{y}\|^2. \end{aligned}$$

Portanto, h é fortemente convexa com módulo $\gamma > 0$. □

1.1.3 Subdiferencial de Funções Convexas

Definição 1.1.5 ([8]). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Dizemos que $\mathbf{y} \in \mathbb{R}^n$ é um **subgradiente** de f no ponto $\mathbf{x} \in \mathbb{R}^n$ se*

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

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

Exemplo 1.1.7. *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por, $f(x) = |x|$. O subdiferencial de f no ponto $x = 0$ é dado por*

$$\begin{aligned} \partial f(0) &= \{\mathbf{y} \in \mathbb{R} \mid f(\mathbf{z}) - f(0) \geq \mathbf{y}(\mathbf{z} - 0) \quad \forall \mathbf{z} \in \mathbb{R}\} \\ &= \{\mathbf{y} \in \mathbb{R} \mid f(\mathbf{z}) \geq \mathbf{y}\mathbf{z}\} \\ &= \{\mathbf{y} \in \mathbb{R} \mid |\mathbf{z}| \geq \mathbf{y}\mathbf{z}\}. \end{aligned}$$

Se $z \geq 0$, $|z| = z$ e daí

$$\partial f(0) = \{y \in \mathbb{R} \mid z \geq yz\} = \{y \in \mathbb{R} \mid y \leq 1\}.$$

Se $z < 0$, $|z| = -z$ assim

$$\partial f(0) = \{y \in \mathbb{R} \mid -z \geq yz\} = \{y \in \mathbb{R} \mid y \geq -1\}.$$

Logo $\partial f(0) = [-1, 1]$.

Lema 1.1.1 ([8]). *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. f é diferenciável num ponto $x \in \mathbb{R}^n$ se, e somente se, $\partial f(x) = \{\nabla f(x)\}$.*

Exemplo 1.1.8 ([8]). *Seja $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por, $f(x) = |x|$. Vimos anteriormente que subdiferencial de f no ponto $x = 0$ é o intervalo $[-1, 1]$. Como f é diferenciável em todo $x \neq 0$ temos pelo Lema 1.1.1 anterior que o subdiferencial de f nesses pontos é dado por*

$$\partial f(x) = f'(x) = 1$$

ou ainda

$$\partial f(x) = f'(x) = -1$$

respectivamente quando $x > 0$ ou $x < 0$.

Veremos agora um resultado sobre o subdiferencial da soma de funções convexas cuja demonstração será omitida mas pode ser encontrada em [8], tal resultado será utilizado posteriormente no capítulo de otimização multiobjetivo.

Proposição 1.1.5. *Sejam $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, 2, \dots, p$, funções convexas. Então*

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

Demonstração. Veja [8]. □

Exemplo 1.1.9 ([8]). *Sejam $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ funções convexas diferenciáveis, $i = 1, 2, \dots, p$, e*

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad f(x) = \max f_i(x), \quad i = 1, 2, \dots, p$$

*Para $x \in \mathbb{R}^n$, definimos o conjunto de **índices ativos** em x por*

$$I(x) = \{i = 1, 2, \dots, p \mid f(x) = f_i(x)\}$$

(ver [8]). Temos que

$$\begin{aligned} \partial f(x) &= \text{conv} \left\{ f'_i(x), i \in I(x) \right\} \\ &= \left\{ y \in \mathbb{R}^n \mid y = \sum_{i \in I(x)} \alpha_i f'_i(x), \sum_{i \in I(x)} \alpha_i = 1, \alpha_i \in \mathbb{R}_+, i \in I(x) \right\}. \end{aligned}$$

Em particular, x é um minimizador de f no \mathbb{R}^n se, e somente se, existem $\alpha_i \in \mathbb{R}_+, i \in I(x)$, tais que

$$0 = \sum_{i \in I(x)} \alpha_i f'_i(x), \sum_{i \in I(x)} \alpha_i = 1.$$

1.1.4 Espaços de Matrizes ($\mathbb{R}^{m \times n}$)

O espaço $\mathbb{R}^{m \times n}$ é conhecido como o espaço das matrizes com m -linhas e n -colunas. Apresentaremos aqui algumas definições e resultados básicos para este tipo de espaço.

Definição 1.1.6 ([13]). *Sejam $x \in \mathbb{R}^n$ e $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função convexa e continuamente diferenciável. Definimos o jacobiano de F em x por $JF(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m$, uma matriz $m \times n$ cujas entradas são dadas por $(JF(x))_{i,j} = \frac{\partial F_i}{\partial x_j}(x)$.*

Exemplo 1.1.10. *Seja $F : \mathbb{R}^3 \rightarrow \mathbb{R}^4$ dada por $F(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3, x_1^2, x_1^3 + x_2^2 + x_3^2)$. O jacobiano de F no ponto $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ é dado por*

$$JF(x)_{4 \times 3} = \frac{\partial F_i}{\partial x_j}(x) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \frac{\partial F_1}{\partial x_2} & \frac{\partial F_1}{\partial x_3} \\ \frac{\partial F_2}{\partial x_1} & \frac{\partial F_2}{\partial x_2} & \frac{\partial F_2}{\partial x_3} \\ \frac{\partial F_3}{\partial x_1} & \frac{\partial F_3}{\partial x_2} & \frac{\partial F_3}{\partial x_3} \\ \frac{\partial F_4}{\partial x_1} & \frac{\partial F_4}{\partial x_2} & \frac{\partial F_4}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 2x_1 & 0 & 0 \\ 3x_1^2 & 2x_2 & 2x_3 \end{bmatrix}$$

Definição 1.1.7 ([11]). *Sejam $A \in \mathbb{R}^{m \times n}$ e $x \in \mathbb{R}^n$. Definiremos a norma $\|Ax\|_\infty$ como*

$$\|Ax\|_\infty = \max_{i=1,2,\dots,m} \left| \sum_{j=1}^n A_{i,j} x_j \right|.$$

Exemplo 1.1.11. *Sejam $A \in \mathbb{R}^{2 \times 2}$ e $x \in \mathbb{R}^2$ dados por*

$$A = \begin{bmatrix} 2 & 1 \\ 2 & 2 \end{bmatrix}, \quad x = (1, 2).$$

A norma $\|Ax\|_\infty$ é dada por

$$\begin{aligned} \|Ax\|_\infty &= \max_{i=1,2} \left| \sum_{j=1}^2 A_{i,j}x_j \right| \\ &= \max_{i=1,2} |A_{i,1}x_1 + A_{i,2}x_2| \\ &= \max_{i=1,2} |A_{i,1}1 + A_{i,2}2| \\ &= \max \{ |(A_{1,1}1 + A_{1,2}2)|; |(A_{2,1}1 + A_{2,2}2)| \} \\ &= \max \{ |4|; |6| \} = 6. \end{aligned}$$

Definição 1.1.8. Para $A \in \mathbb{R}^{m \times n}$, o seguinte número $\|A\|_{\infty,2}$ é definido como

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

Lema 1.1.2. Seja $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ então,

$$\|A\|_{\infty,2} = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_\infty$$

Demonstração. De fato, suponhamos que

$$M = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} \Rightarrow M \geq \frac{\|Ax\|_\infty}{\|x\|_2} \quad \forall x \neq 0.$$

Em particular

$$M \geq \|Ax\|_\infty \quad \forall x \in \mathbb{R}^n, \|x\|_2 = 1,$$

logo

$$M \geq \max_{\|x\|_2=1} \|Ax\|_\infty.$$

Reciprocamente, se

$$M = \max_{\|x\|_2=1} \|Ax\|_\infty \Rightarrow M \geq \|Ax\|_\infty, \|x\|_2 = 1.$$

Seja $x \in \mathbb{R}^n$ não nulo e arbitrário, então $\frac{x}{\|x\|_2}$ é unitário, de modo que

$$M \geq \left\| A \frac{x}{\|x\|_2} \right\|_\infty = \frac{1}{\|x\|_2} \|Ax\|_\infty,$$

logo

$$M \geq \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2}.$$

Portanto

$$\|A\|_{\infty,2} = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_\infty.$$

□

Proposição 1.1.6. *Seja $A \in \mathbb{R}^{m \times n}$, $\|A\|_{\infty,2}$ é uma norma em $\mathbb{R}^{m \times n}$, e ainda,*

$$\begin{aligned} \|A\|_{\infty,2} &= \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} \\ &= \max_{i=1,2,\dots,m} \left(\sum_{j=1}^n A_{i,j}^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Demonstração. Primeiramente verificaremos as três propriedades da definição de norma.

1. Se $A \neq 0$ existe $x \neq 0 \in \mathbb{R}^n$ tal que $Ax \neq 0$, logo $\|Ax\|_\infty > 0 \Rightarrow \|A\|_{\infty,2} = \max_{\|x\|_2=1} \|Ax\|_\infty > 0$.

2. Para todo $\lambda \in \mathbb{R}$,

$$\|\lambda A\|_{\infty,2} = \max_{\|x\|_2=1} \|\lambda Ax\|_\infty = \max_{\|x\|_2=1} |\lambda| \|Ax\|_\infty = |\lambda| \max_{\|x\|_2=1} \|Ax\|_\infty = |\lambda| \|A\|_{\infty,2}.$$

3. Agora, se $A, B \in \mathbb{R}^{m \times n}$ então

$$\begin{aligned} \|A + B\|_{\infty,2} &= \max_{\|x\|_2=1} \|(A + B)x\|_\infty = \max_{\|x\|_2=1} \|Ax + Bx\|_\infty \\ &\leq \max_{\|x\|_2=1} (\|Ax\|_\infty + \|Bx\|_\infty) \leq \max_{\|x\|_2=1} \|Ax\|_\infty + \max_{\|x\|_2=1} \|Bx\|_\infty \\ &= \|A\|_{\infty,2} + \|B\|_{\infty,2}. \end{aligned}$$

Sendo

$$\|A\|_{\infty,2} = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_\infty$$

temos que

$$\|Ax\|_\infty = \max \left| \sum_{j=1}^n A_{i,j} x_j \right| \leq \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}} \cdot \left(\sum_{j=1}^n x_j^2 \right)^{\frac{1}{2}},$$

pela desigualdade de Holder, assim

$$\|Ax\|_\infty \leq \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}} \cdot \|x\|_2 \Rightarrow \frac{\|Ax\|_\infty}{\|x\|_2} \leq \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}}$$

$\forall x \in \mathbb{R}^n; x \neq 0$ logo,

$$\|A\|_{\infty,2} = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} \leq \max_{i=1,2,\dots,m} \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}}.$$

Reciprocamente, tomemos $x_j = \frac{A_{i,j}}{(\sum_{k=1}^n (A_{i,k}^2))^{\frac{1}{2}}}$ notemos que

$$\begin{aligned} \|x\|_2 &= \left(\sum_{j=1}^n (x_j)^2 \right)^{\frac{1}{2}} = \left[\sum_{j=1}^n \left(\frac{(A_{i,j}^2)}{(\sum_{k=1}^n (A_{i,k}^2)^2)} \right) \right]^{\frac{1}{2}} \\ &= \left(\sum_{k=1}^n (A_{i,k}^2) \right)^{-\frac{1}{2}} \cdot \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}} \\ &= 1. \end{aligned}$$

Assim,

$$\begin{aligned} \left| \sum_{j=1}^n A_{i,j} x_j \right| &= \left| \sum_{j=1}^n \left[\frac{A_{i,j} \cdot A_{i,j}}{(\sum_{k=1}^n (A_{i,k}^2)^{\frac{1}{2}})} \right] \right| \\ &= \left| \left(\sum_{k=1}^n (A_{i,k}^2) \right)^{-\frac{1}{2}} \right| \cdot \left| \left(\sum_{j=1}^n (A_{i,j}^2) \right) \right| \\ &= \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}} \cdot \|x\|_2 \end{aligned}$$

Logo,

$$\begin{aligned} \|Ax\|_\infty &= \max_{i=1,2,\dots,m} \left| \sum_{j=1}^n A_{i,j} x_j \right| = \max_{i=1,2,\dots,m} \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}} \cdot \|x\|_2 \Rightarrow \\ \Rightarrow \|A\|_{\infty,2} &= \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_\infty \geq \max_{i=1,2,\dots,m} \left(\sum_{j=1}^n (A_{i,j}^2) \right)^{\frac{1}{2}}. \end{aligned}$$

Portanto

$$\|A\|_{\infty,2} = \max_{x \neq 0} \frac{\|Ax\|_{\infty}}{\|x\|_2} = \max_{i=1,2,\dots,m} \left(\sum_{j=1}^n A_{i,j}^2 \right)^{\frac{1}{2}}.$$

□

1.2 Definições e Resultados Básicos para Otimização Multiobjetivo

Nesta seção, consideraremos funções $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ convexas e continuamente diferenciáveis. Apresentaremos definições de ordenação de conjuntos e também definições e resultados básicos relacionados a problemas de otimização multiobjetivo.

Definição 1.2.1 ([18]). *Um conjunto C é dito ordenado com relação de ordem \preceq se dados quaisquer elementos x, y e $z \in C$ sempre vale $x \preceq y$ ou $y \preceq x$ e as seguintes propriedades são satisfeitas:*

$$\begin{aligned} x \preceq x & \quad (\text{reflexividade}) \\ x \preceq y, y \preceq z \Rightarrow x \preceq z & \quad (\text{transitividade}) \\ x \preceq y, y \preceq x \Rightarrow x = y & \quad (\text{antisimetria}). \end{aligned}$$

Apresentaremos agora uma definição muito importante quando tratamos de otimização multiobjetivo.

Definição 1.2.2 ([18]). *Dizemos que um conjunto C é parcialmente ordenado quando valem as propriedades acima porém, nem sempre dois elementos $x, y \in C$ são comparáveis, isto é, $x \preceq y$ ou $y \preceq x$.*

Definição 1.2.3 ([18]). *As operações de comparação entre vetores pertencentes ao \mathbb{R}^n serão definidas da seguinte forma:*

$$\begin{aligned} x \preceq y & \Rightarrow \{x_i \leq y_i, i = 1, 2, \dots, n\} \\ x \prec y & \Rightarrow \{x_i < y_i, i = 1, 2, \dots, n\} \\ x = y & \Rightarrow \{x_i = y_i, i = 1, 2, \dots, n\}. \end{aligned}$$

Observemos que o operador \neq é definido de outra maneira:

$$x \neq y \Rightarrow \{\exists i \mid x_i \neq y_i\},$$

isto é,

$$\mathbf{x} \neq \mathbf{y} \not\Rightarrow \{x_i \neq y_i, i = 1, 2, \dots, n\}.$$

Exemplo 1.2.1 ([18]). *Consideremos os vetores $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^3$ tais que:*

$$\mathbf{x} = (3, 5, 7) \quad , \quad \mathbf{y} = (2, 4, 6) \quad , \quad \mathbf{z} = (3, 4, 5).$$

Notemos que

$$\mathbf{y} \prec \mathbf{x} \Rightarrow \mathbf{y} \preceq \mathbf{x}$$

$$\mathbf{z} \prec \mathbf{x} \Rightarrow \mathbf{z} \preceq \mathbf{x}.$$

Contudo, podemos ver que os vetores \mathbf{z} e \mathbf{y} não podem ser comparados.

Definição 1.2.4. *Um ponto $\mathbf{x} \in \mathbb{R}^n$ é dito **crítico Pareto** quando*

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

onde $\text{Im}(\text{JF}(\mathbf{x}))$ é a imagem da transformação linear $\text{JF}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$.

Para o estudo desta dissertação onde abordamos direção de descida observamos que da Definição 1.2.4, se um ponto $\mathbf{x} \in \mathbb{R}^n$ não é crítico Pareto então existe uma direção $\mathbf{v} \in \mathbb{R}^n$ satisfazendo

$$\text{JF}(\mathbf{x})\mathbf{v} \in (-\mathbb{R}_{++})^m,$$

isto é, \mathbf{v} é uma direção de descida para a função objetivo F .

Definição 1.2.5. *Um ponto $\mathbf{x} \in \mathbb{R}^n$ é dito **ótimo Pareto** se não existe um ponto $\mathbf{y} \in \mathbb{R}^n$ com $F(\mathbf{y}) \preceq F(\mathbf{x})$ e $F(\mathbf{y}) \neq F(\mathbf{x})$.*

Exemplo 1.2.2. *Consideremos a função $F : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por*

$$F(\mathbf{x}, \mathbf{y}) = (x^2, y^2).$$

Notemos que $(0, 0)$ é ótimo Pareto para F uma vez que não existe $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^2$ tal que

$$F(\mathbf{x}, \mathbf{y}) \preceq F(0, 0) \quad \text{e} \quad F(\mathbf{x}, \mathbf{y}) \neq F(0, 0),$$

isto é,

$$(x^2, y^2) \preceq (0, 0) \quad \text{e} \quad (x^2, y^2) \neq (0, 0),$$

pois do contrário teríamos

$$x^2 < 0 \quad \text{ou}$$

$$y^2 < 0,$$

o que não é verdade.

Definição 1.2.6. Um ponto $x \in \mathbb{R}^n$ é chamado **localmente ótimo Pareto**, se existe uma vizinhança $U \subseteq \mathbb{R}^n$ de x tal que nessa vizinhança x é ótimo Pareto, isto é, não existe um ponto $y \in U$ com $F(y) \preceq F(x)$ e $F(y) \neq F(x)$.

Quando $m = 1$, isto é, se considerarmos uma função $F : \mathbb{R}^n \rightarrow \mathbb{R}$, dizer que $x \in \mathbb{R}^n$ é ótimo Pareto para F é equivalente a não existir $y \in \mathbb{R}^n$ tal que

$$F(y) < F(x) \Rightarrow F(x) \leq F(y) \quad \forall y \in \mathbb{R}^n$$

e dizer que $x \in \mathbb{R}^n$ é localmente ótimo Pareto é dizer que existe uma vizinhança $U \subseteq \mathbb{R}^n$ de x tal que

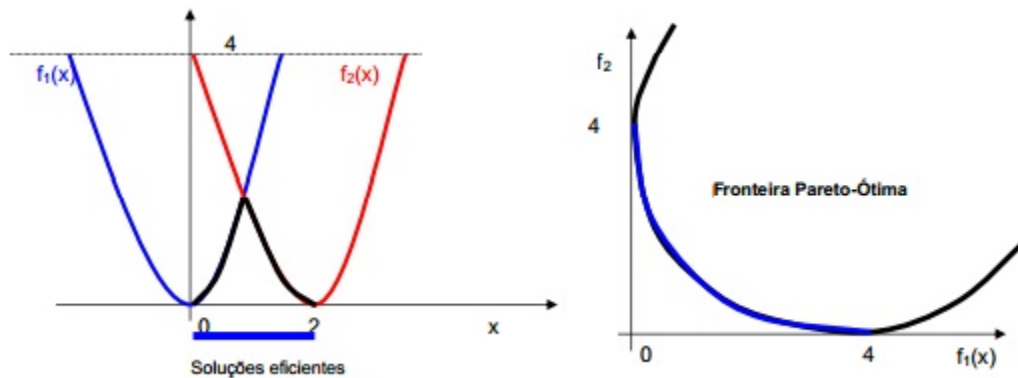
$$F(x) \leq F(y) \quad \forall y \in U.$$

Assim, quando $m = 1$ as definições de ótimo Pareto e localmente ótimo Pareto coincidem com as de mínimo global e local respectivamente.

Exemplo 1.2.3. Se considerarmos a função $F : \mathbb{R} \rightarrow \mathbb{R}^2$ dada por

$$F(x) = (f_1(x), f_2(x)) = (x^2, (x - 2)^2),$$

podemos visualizar o conjunto de pontos críticos Pareto e a fronteira ótimo Pareto.



Lema 1.2.1. Seja $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ convexa, então cada ponto ótimo Pareto local é ótimo Pareto.

Demonstração. Com efeito, seja \bar{x} um ponto ótimo Pareto local de $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, então existe $U \subset \mathbb{R}^n$ vizinhança de \bar{x} tal que não existe $y \in U$ satisfazendo

$$F(y) \preceq F(\bar{x}) \text{ e } F(y) \neq F(\bar{x}),$$

isto é, não existe $\mathbf{y} \in \mathbf{U}$ com

$$F_i(\mathbf{y}) \leq F_i(\bar{\mathbf{x}}) \text{ e } F_i(\mathbf{y}) \neq F_i(\bar{\mathbf{x}}). \quad (1.4)$$

Observemos que para cada $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$, (1.4) pode ser escrita como

$$F_i(\bar{\mathbf{x}}) \leq F_i(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbf{U},$$

ou seja, $\bar{\mathbf{x}}$ é mínimo local das F_i e como cada $F_i : \mathbb{R}^n \rightarrow \mathbb{R}$ é convexa temos

$$F_i(\bar{\mathbf{x}}) \leq F_i(\mathbf{y}) \quad \forall \mathbf{y} \in \mathbb{R}^n.$$

Logo voltando a (1.4) temos que não existe $\mathbf{y} \in \mathbb{R}^n$ tal que

$$F_i(\mathbf{y}) < F_i(\bar{\mathbf{x}}),$$

isto é, não existe $\mathbf{y} \in \mathbb{R}^n$ satisfazendo

$$F_i(\mathbf{y}) \leq F_i(\bar{\mathbf{x}}) \text{ e } F_i(\mathbf{y}) \neq F_i(\bar{\mathbf{x}}).$$

Portanto $\bar{\mathbf{x}}$ é ótimo Pareto. □

Uma outra demonstração para este resultado pode ser encontrada em [15].

Lema 1.2.2. *Se um ponto $\bar{\mathbf{x}} \in \mathbb{R}^n$ é ótimo Pareto então*

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

Demonstração. Suponhamos que

$$\text{Im}(\text{JF}(\bar{\mathbf{x}})\mathbf{v}) \cap (-\mathbb{R}_{++})^m \neq \emptyset,$$

isto é, existe $\mathbf{v} \in \mathbb{R}^n$ tal que

$$\text{JF}(\bar{\mathbf{x}})\mathbf{v} \in (-\mathbb{R}_{++})^m \text{ ou } \text{JF}(\bar{\mathbf{x}})\mathbf{v} \prec 0,$$

isto significa, que existe $\mathbf{v} \in \mathbb{R}^n$ tal que para qualquer $i = \{1, 2, \dots, m\}$

$$\langle \nabla F_i(\bar{\mathbf{x}}), \mathbf{v} \rangle < 0.$$

Como F é continuamente diferenciável temos

$$\begin{aligned} F_i(\bar{\mathbf{x}} + t\mathbf{v}) - F_i(\bar{\mathbf{x}}) &= t \langle \nabla F_i(\bar{\mathbf{x}}), \mathbf{v} \rangle + o(t) \\ &= t \left(\langle \nabla F_i(\bar{\mathbf{x}}), \mathbf{v} \rangle + \frac{o(t)}{t} \right). \end{aligned}$$

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

$$\langle \nabla F_i(\bar{x}), v \rangle + \frac{o(t)}{t} \leq \frac{\langle \nabla F_i(\bar{x}), v \rangle}{2} < 0.$$

Daí,

$$F_i(\bar{x} + tv) - F_i(\bar{x}) \leq t \left(\frac{\langle \nabla F_i(\bar{x}), v \rangle}{2} \right) < 0,$$

logo

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

Portanto \bar{x} não é ótimo Pareto. □

1.3 Definições e Resultados Básicos Para Otimização Vetorial

Abordaremos aqui definições e resultados que serão úteis nos próximos capítulos. Apresentaremos definições de cones, bem como a definição de ordem parcial induzida por um cone. Além disso veremos definições e um resultado relacionados a otimização vetorial.

1.3.1 Definições Básicas

Definição 1.3.1 ([8], [18]). 1. Um conjunto $K \subset \mathbb{R}^n$ é dito **cone** quando

$$\forall d \in K \Rightarrow td \in K \quad \forall t \in \mathbb{R}_+.$$

2. Seja $K \subset \mathbb{R}^n$ um cone, dizemos que K é **convexo** quando é um conjunto convexo, isto é, quando para quaisquer $x, y \in K$ e $t \in [0, 1]$ tivermos

$$(1 - t)x + ty \in K.$$

3. Seja $K \subset \mathbb{R}^n$ um cone, dizemos que K é **pontiagudo** quando

$$K \cap (-K) = \{0\}.$$

Exemplo 1.3.1. O \mathbb{R}_+^m é um cone convexo, fechado e pontiagudo.

De fato,

$$tx \in \mathbb{R}_+^m \quad \forall x \in \mathbb{R}_+^m \quad e \quad \forall t \in \mathbb{R}_+.$$

Para todos $x, y \in \mathbb{R}_+^m$ temos

$$(1 - t)x + ty \in \mathbb{R}_+^m \quad \forall t \in [0, 1]$$

e ainda,

$$\mathbb{R}_+^m \cap (-\mathbb{R}_+)^m = \{0\}.$$

Definição 1.3.2 ([19]). *Seja $A \subseteq \mathbb{R}^n$. Chamamos de **cone polar positivo** de A ao conjunto*

$$A^* = \{w \in \mathbb{R}^n \mid \langle y, w \rangle \geq 0 \quad \forall y \in A\}$$

Exemplo 1.3.2. *O cone polar positivo do \mathbb{R}_+^n é o próprio \mathbb{R}_+^n .*

De fato,

$$(\mathbb{R}_+^n)^* = \{w \in \mathbb{R}_+^n \mid \langle y, w \rangle \geq 0 \quad \forall y \in \mathbb{R}_+^n\}.$$

Seja $w = (w_1, w_2, \dots, w_n)$ e $y = (y_1, y_2, \dots, y_n)$ temos

$$(\mathbb{R}_+^n)^* = \{w \in \mathbb{R}_+^n \mid w_1 y_1 + w_2 y_2 + \dots + w_n y_n \geq 0 \quad \forall y_i \geq 0, i = 1, 2, \dots, n\}.$$

Note que para $w = (w_1, w_2, \dots, w_n)$ com pelo menos um $w_i < 0$ podemos escolher um y_i suficientemente grande tal que

$$w_1 y_1 + w_2 y_2 + \dots + w_n y_n < 0.$$

Logo $(\mathbb{R}_+^n)^ = \mathbb{R}_+^n$.*

Propriedade 1.3.1 ([17], [19]). *Seja K um cone convexo, fechado e pontiagudo com $\text{int}K \neq \emptyset$ as seguintes afirmações são válidas:*

1. K^* é convexo, fechado e pontiagudo.
2. $K^{**} = (K^*)^* = K$
3. $\text{int}K^* \neq \emptyset$
4. $K^* = \mathbb{R}_+^m$, quando $K = \mathbb{R}_+^m$

Definição 1.3.3. *Seja K um cone convexo, fechado e pontiagudo com interior não vazio. Como visto na Propriedade 1.3.1 item 2, temos que*

$$K = \{y \in \mathbb{R}^n \mid \langle y, w \rangle \geq 0 \quad \forall w \in K^*\} \Rightarrow -K = \{y \in \mathbb{R}^n \mid \langle y, w \rangle \leq 0 \quad \forall w \in K^*\} \quad \text{e}$$

$$\text{int}K = \{y \in \mathbb{R}^n \mid \langle y, w \rangle > 0 \quad \forall w \in K^*\} \Rightarrow -\text{int}K = \{y \in \mathbb{R}^n \mid \langle y, w \rangle < 0 \quad \forall w \in K^*\}.$$

1.3.2 Introdução a Otimização Vetorial

Nesta parte abordaremos as definições de um ponto K -minimizador irrestrito (ou K -ótimo), ponto K -crítico e a definição de direção de descida com a ordem induzida por um cone genérico (K -direção de descida). Estudaremos o seguinte problema

$$\min_{\mathbf{K}} F(\mathbf{x})$$

Definição 1.3.4. *Seja $K \subset \mathbb{R}^m$ um cone convexo, fechado e pontiagudo com interior não-vazio. A ordem parcial em \mathbb{R}^m induzida por K , \preceq_K , é definida por*

$$\mathbf{u} \preceq_K \mathbf{v}, \quad \text{se } \mathbf{v} - \mathbf{u} \in K,$$

$$\mathbf{u} \prec_K \mathbf{v}, \quad \text{se } \mathbf{v} - \mathbf{u} \in \text{int}K.$$

Quando $K = \mathbb{R}_+^m$, a definição resume-se a

$$\mathbf{u} \leq \mathbf{v}, \quad \text{se } \mathbf{v} - \mathbf{u} \in \mathbb{R}_+^m,$$

$$\mathbf{u} < \mathbf{v}, \quad \text{se } \mathbf{v} - \mathbf{u} \in \mathbb{R}_{++}^m.$$

Definição 1.3.5. *Um ponto $\mathbf{x} \in \mathbb{R}^n$ é dito **K -crítico** de uma função $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ quando*

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

Observe da definição acima que se \mathbf{x} é não K -crítico então existe $\mathbf{v} \in \mathbb{R}^n$ satisfazendo

$$JF(\mathbf{x})\mathbf{v} \in -\text{int}(K),$$

isto é, $JF(\mathbf{x})\mathbf{v} \prec_K 0$. Neste fato, tal \mathbf{v} é conhecido como uma K -direção de descida para a função objetivo F em \mathbf{x} . Outro fato que caracteriza \mathbf{v} como uma K -direção de descida é o fato de existir $\bar{t} > 0$ tal que

$$F(\mathbf{x} + t\mathbf{v}) \prec_K F(\mathbf{x}) \quad \forall t \in (0, \bar{t}).$$

Definição 1.3.6. *Um ponto $\mathbf{x} \in \mathbb{R}^n$ é dito **K -minimizador irrestrito (ou K -ótimo)** se não existe um ponto $\mathbf{y} \in \mathbb{R}^n$ com $F(\mathbf{y}) \preceq_K F(\mathbf{x})$ e $F(\mathbf{y}) \neq F(\mathbf{x})$.*

Definição 1.3.7. *Seja \mathbf{v} uma K -direção de descida em \mathbf{x} . Para $t > 0$ e $\beta \in (0, 1)$, uma regra tipo Armijo é dada por*

$$F(\mathbf{x} + t\mathbf{v}) \preceq_K F(\mathbf{x}) + \beta t JF(\mathbf{x})\mathbf{v}.$$

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

Proposição 1.3.1. *Seja $\beta \in (0, 1)$. Se $JF(x)v \prec_K 0$, então existe $\bar{t} > 0$ tal que*

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

Demonstração. Sendo f diferenciável temos

$$F(x + tv) = F(x) + tJF(x)v + R(t)$$

com $\lim_{t \rightarrow 0} R(t) = 0$. Assumimos que $JF(x)v \in -\text{int}K$. Sendo $\beta \in (0, 1)$, $t(1 - \beta)JF(x)v \in -\text{int}K$. Como $\lim_{t \rightarrow 0} R(t) = 0$ existe $\bar{t} > 0$ tal que para todo $t \in (0, \bar{t}]$, $\|R(t)\|$ é suficientemente pequena, de modo que

$$R(t) + t(1 - \beta)JF(x)v \in -\text{int}K,$$

logo

$$R(t) \prec_K -t(1 - \beta)JF(x)v \quad \forall t \in (0, \bar{t}].$$

Assim

$$\begin{aligned} F(x + tv) &= F(x) + tJF(x)v + R(t) \\ &\prec_K F(x) + tJF(x)v - t(1 - \beta)JF(x)v \\ &= F(x) + tJF(x)v - tJF(x)v + \beta tJF(x)v \\ &= F(x) + \beta tJF(x)v. \end{aligned}$$

□

Para otimização multiobjetivo, isto é, quando $K = \mathbb{R}_+^m$, a hipótese de $JF(x)v \prec_K 0$ é equivalente a $JF(x)v \leq 0$ portanto, podemos obter um resultado semelhante.

Capítulo 2

Método de Descida em Otimização Multiobjetivo

Neste capítulo apresentaremos o problema de minimização multiobjetivo irrestrito de uma função $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ convexa e continuamente diferenciável. Veremos alguns resultados relacionados a direção de descida e a busca linear. Além disso, analisaremos a convergência do método.

2.1 Algoritmo Geral

A idéia geral do algoritmo é escolher um ponto $x \in \mathbb{R}^n$ e verificar se é crítico Pareto, se não, calcularemos uma direção v e daremos um passo com comprimento de passo convenientemente escolhido a partir de x ao longo de v . Isto resulta num novo ponto, e o processo será repetido.

2.2 Direção de Busca

Com o objetivo de encontrar condições para definir uma direção de busca satisfatória, consideraremos um ponto $x \in \mathbb{R}^n$ e uma função $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

$$f_x(v) = \max\{(JF(x)v)_i \mid i = 1, 2, \dots, m\}.$$

Observamos que f_x é convexa pois é o máximo de uma função linear e homogênea positiva, como visto na Proposição 1.1.3.

Seja o problema de minimização irrestrita

$$\begin{aligned} \min \quad & f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \\ \text{s.a} \quad & \mathbf{x} \in \mathbb{R}^n. \end{aligned} \tag{2.1}$$

Lema 2.2.1. *Sejam $\mathbf{v}(\mathbf{x})$ e $\alpha(\mathbf{x})$ a solução e o valor ótimo do Problema (2.1) respectivamente.*

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

2. *Se \mathbf{x} não é crítico Pareto então $\alpha(\mathbf{x}) < 0$,*

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

e ainda

$$(\mathbf{JF}(\mathbf{x})\mathbf{v}(\mathbf{x}))_i \leq f_x(\mathbf{v}(\mathbf{x})), i = 1, 2, \dots, m.$$

3. *As aplicações $\mathbf{x} \rightarrow \mathbf{v}(\mathbf{x})$ e $\mathbf{x} \rightarrow \alpha(\mathbf{x})$ são contínuas.*

Demonstração. 1. Se \mathbf{x} é crítico Pareto ,

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

isto é, para todo $\mathbf{v} \in \mathbb{R}^n$ existe $i = 1, 2, \dots, m$ tal que $(\mathbf{JF}(\mathbf{x})\mathbf{v})_i \geq 0$. Assim

$$f_x(\mathbf{v}) = \max\{(\mathbf{JF}(\mathbf{x})\mathbf{v})_i \mid i = 1, 2, \dots, m\} \geq 0 \quad \forall \mathbf{v} \in \mathbb{R}^n,$$

daí

$$f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \geq 0$$

e, no caso que $\mathbf{v} = \mathbf{0}$ temos $f_x(\mathbf{0}) + \frac{1}{2}\|\mathbf{0}\|^2 = 0$. Suponhamos que $\mathbf{v}(\mathbf{x}) \neq \mathbf{0}$, logo

$$\alpha(\mathbf{x}) = f_x(\mathbf{v}(\mathbf{x})) + \frac{1}{2}\|\mathbf{v}(\mathbf{x})\|^2 > 0 = f_x(\mathbf{0}) + \frac{1}{2}\|\mathbf{0}\|^2,$$

o que é uma contradição pois $\mathbf{v}(\mathbf{x})$ é uma solução do problema. Portanto $\mathbf{v}(\mathbf{x}) = \mathbf{0}$ e, conseqüentemente, $\alpha(\mathbf{x}) = 0$.

2. Com efeito, como $\mathbf{v}(\mathbf{x})$ é solução, segue que

$$\alpha(\mathbf{x}) = f_x(\mathbf{v}(\mathbf{x})) + \frac{1}{2}\|\mathbf{v}(\mathbf{x})\|^2 \leq f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \quad \forall \mathbf{v} \in \mathbb{R}^n.$$

Por outro lado, como x não é crítico Pareto, existe $\hat{v} \in \mathbb{R}^n$ tal que

$$f_x(\hat{v}) = \max\{(JF(x)\hat{v})_i \mid i = 1, 2, \dots, m\} < 0.$$

Consideremos $\tilde{v} = \frac{-f_x(\hat{v})}{\|\hat{v}\|^2}\hat{v}$, então

$$\begin{aligned} \alpha(x) &\leq f_x(\tilde{v}) + \frac{1}{2}\|\tilde{v}\|^2 \\ &= f_x\left(\frac{-f_x(\hat{v})}{\|\hat{v}\|^2}\hat{v}\right) + \frac{1}{2}\left\|\frac{-f_x(\hat{v})}{\|\hat{v}\|^2}\hat{v}\right\|^2. \end{aligned}$$

Agora, como f_x é homogênea positiva, a última desigualdade pode ser escrita como

$$\begin{aligned} \alpha(x) &\leq \frac{-f_x(\hat{v})}{\|\hat{v}\|^2}f_x(\hat{v}) + \frac{1}{2}\frac{|-f_x(\hat{v})|^2\|\hat{v}\|^2}{\|\hat{v}\|^4} \\ &= \frac{-f_x(\hat{v})^2}{\|\hat{v}\|^2} + \frac{1}{2}\left(\frac{-f_x(\hat{v})^2}{\|\hat{v}\|^2}\right) \\ &= \frac{f_x(\hat{v})^2}{\|\hat{v}\|^2}\left(-1 + \frac{1}{2}\right) \\ &= \frac{f_x(\hat{v})^2}{\|\hat{v}\|^2}\left(-\frac{1}{2}\right) < 0. \end{aligned}$$

Portanto $\alpha(x) < 0$.

3. Sendo $v : \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma função que associa a cada ponto $x \in \mathbb{R}^n$ uma direção $v(x) \in \mathbb{R}^n$, consideremos uma função g definida por

$$g(v) = f_x(v) + \frac{1}{2}\|v\|^2.$$

Como $v(x)$ é a solução do problema de minimizar $g(v)$ com $v \in \mathbb{R}^n$, temos $0 \in \partial g(v(x))$. Logo,

$$\sum_{i=1}^m \lambda_i \nabla F_i(x) + v(x) = 0 \Rightarrow v(x) = -\sum_{i=1}^m \lambda_i \nabla F_i(x)$$

e, como F é continuamente diferenciável obtemos a continuidade de v . A função $\alpha : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por

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

também é contínua uma vez que f_x é contínua, a norma é contínua e $v(x)$ é contínua.

□

Definição 2.2.1. *Seja $x \in \mathbb{R}^n$ um ponto não crítico Pareto, dizemos que $v \in \mathbb{R}^n$ é uma solução aproximada do Problema (2.1) com tolerância $\sigma \in (0, 1]$ quando,*

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

Observação 2.2.1. *Observe que, sendo $\alpha(x)$ o valor ótimo do Problema (2.1), para $\sigma = 1$ somente a solução exata satisfaz a inequação acima.*

Lema 2.2.2. *Seja $A \in \mathbb{R}^{m \times n}$. Suponhamos que x não seja crítico Pareto e que $v \in \mathbb{R}^n$ seja uma solução aproximada do Problema (2.1) com tolerância $\sigma \in (0, 1]$ então,*

$$\|v\| \leq 2\|A\|_{\infty,2}$$

Demonstração. Seja $\|v\| = \|v\|_2$. Se x não é crítico Pareto, então pelo Lema 2.2.1, o valor ótimo $\alpha(x)$ do Problema (2.1) satisfaz $\alpha(x) < 0$, onde $\alpha(x) = f_x(v(x)) + \frac{1}{2}\|v(x)\|^2$. Agora, como v é solução aproximada, temos que

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

Daí,

$$f_x(v) + \frac{1}{2}\|v\|^2 < 0,$$

pois $\sigma \in (0, 1]$. Logo,

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

consequentemente

$$\|v\|^2 \leq 2|f_x(v)|. \tag{2.2}$$

Por outro lado, temos

$$\begin{aligned} |f_x(v)| &= \left| \max_{i=1,2,\dots,m} \{[Av]_i\} \right| = \left| \max_{i=1,2,\dots,m} \left(\sum_{j=1}^n A_{i,j}v_j \right) \right| \\ &\leq \max_{i=1,2,\dots,m} \left| \sum_{j=1}^n A_{i,j}v_j \right| \\ &= \|Av\|_{\infty}. \end{aligned} \tag{2.3}$$

Portanto, de (2.2) e (2.3)

$$\|\mathbf{v}\|^2 \leq 2\|\mathbf{A}\mathbf{v}\|_\infty.$$

Logo

$$\|\mathbf{v}\| \leq 2 \frac{\|\mathbf{A}\mathbf{v}\|_\infty}{\|\mathbf{v}\|} \leq 2 \max_{\mathbf{v} \neq 0} \frac{\|\mathbf{A}\mathbf{v}\|_\infty}{\|\mathbf{v}\|} = 2\|\mathbf{A}\mathbf{v}\|_{\infty,2}.$$

□

Discutiremos agora outra possibilidade para direção de busca de descida em \mathbf{x} . Em vez de resolver o Problema (2.1) poderíamos tomar \mathbf{v} como sendo solução de

$$\begin{aligned} \min \quad & f_{\mathbf{x}}(\mathbf{v}) \\ \text{s.a.} \quad & \|\mathbf{v}\|_\infty \leq 1 \end{aligned} \tag{2.4}$$

Lema 2.2.3. *Sejam $V(\mathbf{x})$ e $\beta(\mathbf{x})$ o conjunto solução e o valor ótimo do Problema (2.4) respectivamente.*

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

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

3. A aplicação $\mathbf{x} \rightarrow \beta(\mathbf{x})$ é contínua.
4. Se $\{\mathbf{x}^k\}$ converge para $\bar{\mathbf{x}}$ e $\mathbf{v}^k \in V(\mathbf{x}^k)$ converge para $\bar{\mathbf{v}}$, então $\bar{\mathbf{v}} \in V(\bar{\mathbf{x}})$.

Demonstração. 1. Se \mathbf{x} é crítico Pareto, temos

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

isto é, para todo $\mathbf{v} \in \mathbb{R}^n$ existe $i = 1, 2, \dots, m$ tal que $(\mathbf{JF}(\mathbf{x})\mathbf{v})_i \geq 0$. Logo,

$$f_{\mathbf{x}}(\mathbf{v}) = \max_{i=1,2,\dots,m} \{(\mathbf{JF}(\mathbf{x})\mathbf{v})_i\} \geq 0 \quad \forall \mathbf{v} \in \mathbb{R}^n.$$

Por outro lado,

$$f_{\mathbf{x}}(0) = \max_{i=1,2,\dots,m} \{(\mathbf{JF}(\mathbf{x})0)_i\} = 0 \leq f_{\mathbf{x}}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbb{R}^n.$$

Assim $0 \in V(\mathbf{x})$, conseqüentemente $f_{\mathbf{x}}(0) = 0$ é valor ótimo e, portanto, $\beta(\mathbf{x}) = 0$.

2. Se x não é crítico Pareto então existe $\bar{v} \in \mathbb{R}^n$ tal que

$$JF(x)\bar{v} \in (-\mathbb{R}_{++})^m,$$

isto é, $(JF(x)\bar{v})_i < 0 \forall i = 1, 2, \dots, m$. Logo,

$$f_x(\bar{v}) = \max\{(JF(x)\bar{v})_i\} < 0. \quad (2.5)$$

Notemos agora que, como $\beta(x)$ é valor ótimo do problema, temos

$$\beta(x) \leq f_x(v) \quad \forall v \in \mathbb{R}^n.$$

Assim, tomando $\tilde{v} = \frac{\bar{v}}{\|\bar{v}\|_\infty}$, temos $\|\tilde{v}\|_\infty = 1$ e

$$\begin{aligned} \beta(x) &\leq f_x(\tilde{v}) = f_x\left(\frac{\bar{v}}{\|\bar{v}\|_\infty}\right) \\ &= \frac{1}{\|\bar{v}\|_\infty} \cdot (f_x(\bar{v})) \\ &< 0, \end{aligned}$$

onde a última desigualdade segue de (2.5). Conseqüentemente, para todo $v \in V(x)$ temos $f_x(v) = \beta(x) < 0$ e, portanto,

$$(JF(x)v)_i \leq f_x(v) < 0 \quad \forall v \in V(x).$$

3. Sendo $\beta : \mathbb{R}^n \rightarrow \mathbb{R}$ dada por $\beta(x) = f_x(v)$ para cada $v \in V(x)$, temos

$$\beta(x) = \max_{i=1,2,\dots,m} \{(JF(x)v)_i\} = \max_{i=1,2,\dots,m} \langle \nabla F_i, v \rangle.$$

Como F é continuamente diferenciável, $\nabla F_i(x)$ é contínuo e, da continuidade do produto interno temos que β é contínua.

4. Com efeito, se $v^k \in V(x^k)$ então,

$$f_{x^k}(v^k) \leq f_{x^k}(v) \quad \forall v \in \mathbb{R}^n \text{ com } \|v\|_\infty \leq 1 \quad (2.6)$$

e $f_{x^k}(v^k)$ é valor ótimo, isto é, $f_{x^k}(v^k) = \beta(x^k)$. Assim, temos

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

onde a desigualdade segue de (2.6). Usando agora que β é contínua, F é continuamente diferenciável e fazendo $k \rightarrow \infty$, temos

$$\beta(\bar{x}) \leq \max_{i=1,2,\dots,m} \{\langle \nabla F_i(\bar{x}), v \rangle\} = f_{\bar{x}}(v) \quad \forall v \in \mathbb{R}^n \text{ com } \|v\|_\infty \leq 1.$$

Portanto $\bar{v} \in V(\bar{x})$.

□

2.3 Tamanho do Passo

Suponhamos que temos uma direção $\mathbf{v} \in \mathbb{R}^n$ com $\mathbf{JF}(\mathbf{x})\mathbf{v} \prec 0$. Para calcular o comprimento de passo t usamos uma regra tipo Armijo.

Seja $\beta \in (0, 1)$ uma constante. Buscamos $t > 0$ tal que

$$F(\mathbf{x} + t\mathbf{v}) \preceq F(\mathbf{x}) + \beta t \mathbf{JF}(\mathbf{x})\mathbf{v}. \quad (2.7)$$

1. Inicializamos com $t = 1$ e verificamos se a desigualdade (2.7) é satisfeita.
2. Se (2.7) não é satisfeita, fazemos $t = \frac{t}{2}$ e retornamos ao passo 1. Caso contrário, tomamos $t_k = t$ como o comprimento do passo.

O fim do procedimento decorre do fato de (2.7) ser satisfeita para $t > 0$ suficientemente pequeno, como mostraremos a seguir.

Lema 2.3.1. *Se F é diferenciável e $\mathbf{JF}(\mathbf{x})\mathbf{v} \prec 0$, então existe $\epsilon > 0$ (que pode depender de \mathbf{x}, \mathbf{v} e β) tal que,*

$$F(\mathbf{x} + t\mathbf{v}) \prec F(\mathbf{x}) + \beta t \mathbf{JF}(\mathbf{x})\mathbf{v}$$

para todo $t \in (0, \epsilon]$.

Demonstração. De fato, como F é diferenciável, temos

$$F(\mathbf{x} + \mathbf{h}) = F(\mathbf{x}) + \mathbf{JF}(\mathbf{x})\mathbf{h} + \mathbf{R}(\mathbf{h}), \quad (2.8)$$

com $\lim_{\mathbf{h} \rightarrow 0} \frac{|\mathbf{R}_i(\mathbf{h})|}{\|\mathbf{h}\|} = 0, i = 1, 2, \dots, m$. Definamos

$$\mathbf{a} := \max_{i=1,2,\dots,m} (\mathbf{JF}(\mathbf{x})\mathbf{v})_i.$$

Observemos que $\mathbf{a} < 0$ uma vez que $\mathbf{JF}(\mathbf{x})\mathbf{v} \prec 0$ é expandido num sentido de coordenada a coordenada e ainda $\mathbf{v} \neq 0$. Fazendo $\mathbf{h} = t\mathbf{v}$, temos $\lim_{t \rightarrow 0} \frac{|\mathbf{R}_i(t\mathbf{v})|}{\|t\mathbf{v}\|} = 0$, isto é,

$$\forall \delta > 0 \quad \exists \epsilon > 0 ; 0 < |t| \leq \epsilon \Rightarrow \frac{|\mathbf{R}_i(t\mathbf{v})|}{\|t\mathbf{v}\|} < \delta.$$

Como $\beta < 1$ podemos tomar

$$\delta = \frac{(1 - \beta) \cdot |\mathbf{a}|}{\|\mathbf{v}\|}, i = 1, 2, \dots, m.$$

Logo para $0 < t \leq \epsilon$ temos

$$\frac{|R_i(tv)|}{\|tv\|} < \frac{(1 - \beta) \cdot |\alpha|}{\|v\|}, i = 1, 2, \dots, m.$$

Assim, para $0 < t \leq \epsilon$

$$|R_i(tv)| < t(1 - \beta)|\alpha|, i = 1, 2, \dots, m.$$

Como $\alpha < 0$, $|\alpha| = -\alpha = \min_i (-JF(x)v)_i$, a expressão acima pode ser escrita como

$$|R_i(tv)| < t(1 - \beta) \min_{i=1,2,\dots,m} [-(JF(x)v)_i].$$

Portanto,

$$R_i(tv) < |R_i(tv)| < -t(1 - \beta)(JF(x)v)_i, \forall i = 1, 2, \dots, m,$$

e, conseqüentemente,

$$R(tv) \prec -t(1 - \beta)JF(x)v. \tag{2.9}$$

Logo, de (2.8) e (2.9)

$$\begin{aligned} F(x + tv) &= F(x) + tJF(x)v + R(tv) \\ &\prec F(x) + tJF(x)v - t(1 - \beta)JF(x)v \\ &= F(x) + t\beta JF(x)v \end{aligned}$$

para todo $0 < t \leq \epsilon$. □

2.4 Definição e Análise de Convergência do Algoritmo

Iniciaremos esta seção apresentando a definição do algoritmo e, em seguida, estudaremos a convergência do método.

2.4.1 Algoritmo Completo

Consideremos as constantes $\beta \in (0, 1)$ e $\sigma \in (0, 1]$.

1. Inicializamos com $x^0 \in \mathbb{R}^n$.

2. Se \mathbf{x}^k é crítico Pareto pare. Caso contrário,
3. Calcule \mathbf{v}^k , uma solução aproximada do Problema (2.1) com tolerância σ .
4. Calcule o tamanho de passo $t_k \in (0, 1]$ como sendo o máximo de T_k onde,

$$T_k = \left\{ t = \frac{1}{2^j} \mid j \in \mathbb{N}, F(\mathbf{x}^k + t_k \mathbf{v}^k) \leq F(\mathbf{x}^k) + \beta t JF(\mathbf{x}^k) \mathbf{v}^k \right\}.$$

5. Encontre

$$\mathbf{x}^{k+1} = \mathbf{x}^k + t_k \mathbf{v}^k$$

e volte ao passo 2.

Observação 2.4.1. *Se o passo 3 é atingido na iteração k , \mathbf{x}^k não é crítico Pareto, pois do contrário pararíamos no passo 2 e, do passo 3*

$$\max_{i=1,2,\dots,m} (JF(\mathbf{x}^k) \mathbf{v}^k)_i \leq \max_{i=1,2,\dots,m} (JF(\mathbf{x}^k) \mathbf{v}^k)_i + \frac{1}{2} \|\mathbf{v}^k\|^2 \leq \sigma \alpha(\mathbf{x}^k) < 0.$$

Assim, $JF(\mathbf{x}^k) \mathbf{v}^k \prec 0$. Portanto, nesta iteração, do Lema 2.3.1 T_k é não vazio, logo o passo 4 está bem definido. Além disso, esse passo é feito usando uma regra tipo Armijo como descrito acima.

2.4.2 Análise de Convergência do Algoritmo

Analisaremos agora a convergência do algoritmo. Observemos que se o algoritmo para após um número finito k de iterações então, do passo 2, o ponto \mathbf{x}^k é crítico Pareto. A partir de agora supomos que uma sequência infinita é gerada de modo que $\alpha(\mathbf{x}^k) \neq 0$ para todo $k \in \mathbb{N}$.

Teorema 2.4.1. *Todo ponto de acumulação da sequência $\{\mathbf{x}^k\}$, gerada pelo algoritmo é crítico Pareto. Se a função F tem conjunto de nível limitado, no sentido de que o conjunto*

$$A = \{\mathbf{x} \in \mathbb{R}^n \mid F(\mathbf{x}) \leq F(\mathbf{x}^0)\}$$

é limitado, então a sequência $\{\mathbf{x}^k\}$ também é limitada e tem pelo menos um ponto de acumulação.

Demonstração. Seja \mathbf{y} um ponto de acumulação da sequência $\{\mathbf{x}^k\}$, isto é,

$$\mathbf{y} = \lim_{j \rightarrow \infty} \mathbf{x}^{k_j},$$

onde $\{x^{k_j}\} \subset \{x^k\}$. Sejam $v(y)$ e $\alpha(y)$ a solução e o valor ótimo do Problema (2.1) em y , respectivamente.

$$v(y) = \operatorname{argmin} \left\{ f_y(v) + \frac{1}{2} \|v\|^2 \right\}, \quad \alpha(y) = \min_{v \in \mathbb{R}^n} \left\{ f_y(v) + \frac{1}{2} \|v\|^2 \right\},$$

onde

$$f_y(v) = \max_{i=1,2,\dots,m} (JF(y)v)_i.$$

Do Lema 2.2.1 é suficiente provar que $\alpha(y) = 0$. Notemos que a sequência $\{F(x^k)\}$ é, componente a componente, estritamente decrescente. Como F é contínua,

$$\lim_{k \rightarrow \infty} F(x^k) = F(y),$$

logo

$$\lim_{k \rightarrow \infty} \|F(x^k) - F(x^{k+1})\| = 0.$$

Porém,

$$F(x^k) - F(x^{k+1}) \succeq -t_k \beta JF(x^k)v^k \succeq 0,$$

e conseqüentemente

$$\lim_{k \rightarrow \infty} t_k JF(x^k)v^k = 0, \tag{2.10}$$

onde $t_k \in (0, 1] \quad \forall k$. Agora tomemos uma subsequência x^{k_j} convergindo para y . Consideraremos duas possibilidades

$$\limsup_{j \rightarrow \infty} t_{k_j} > 0 \quad \text{ou}$$

$$\limsup_{j \rightarrow \infty} t_{k_j} = 0.$$

1º caso : Neste caso existe uma subsequência $\{x^{k_l}\} \rightarrow y$ satisfazendo

$$\lim_{l \rightarrow \infty} t_{k_l} = \bar{t} > 0.$$

Usando (2.10) concluímos que

$$\lim_{l \rightarrow \infty} (JF(x^{k_l})v^{k_l}) = 0$$

e

$$\lim_{l \rightarrow \infty} v^{k_l} = \lim_{l \rightarrow \infty} \frac{x^{k_{l+1}} - x^{k_l}}{t_{k_l}} = \frac{y - y}{\bar{t}} = 0.$$

Assim,

$$\alpha(\mathbf{y}) = \alpha\left(\lim_{l \rightarrow \infty} \mathbf{x}^{kl}\right) = \lim_{l \rightarrow \infty} \alpha(\mathbf{x}^{kl}) = \lim_{l \rightarrow \infty} \left(\max_i (\mathbf{JF}(\mathbf{x}^{kl})\mathbf{v}^{kl})_i\right) + \frac{1}{2} \lim_{l \rightarrow \infty} \|\mathbf{v}^{kl}\|^2 = 0.$$

Portanto, \mathbf{y} é crítico Pareto.

2º caso: Neste caso, pelo Lema 2.2.2, temos que $\{\mathbf{v}^{kj}\}$ é limitada pois

$$\|\mathbf{v}^{kj}\| \leq 2\|\mathbf{JF}(\mathbf{x}^{kj})\|_{\infty,2}$$

e $\{\mathbf{x}^{kj}\}$ é uma subsequência convergente. Portanto podemos tomar uma subsequência $\{\mathbf{x}^{kr}\}$ de $\{\mathbf{x}^{kj}\}$ tal que a sequência $\{\mathbf{v}^{kr}\}$ também converge para algum $\bar{\mathbf{v}}$. Note que para todo $r \in \mathbb{N}$ temos

$$\max_{i=1,2,\dots,m} (\mathbf{F}(\mathbf{x}^{kr})\mathbf{v}^{kr})_i = f_{\mathbf{x}^{kr}}(\mathbf{v}^{kr}) \leq f_{\mathbf{x}^{kr}}(\mathbf{v}^{kr}) + \frac{1}{2}\|\mathbf{v}^{kr}\|^2 \leq \sigma\alpha(\mathbf{x}^{kr}) < 0.$$

Logo, fazendo $r \rightarrow \infty$, devido a continuidade das funções envolvidas, obtemos

$$\frac{1}{\sigma} \max_i (\mathbf{F}(\mathbf{y})\bar{\mathbf{v}})_i \leq \alpha(\mathbf{y}) \leq 0. \quad (2.11)$$

Tomando algum $q \in \mathbb{N}$ e considerando r suficientemente grande, segue que

$$t_{kr} < \frac{1}{2q},$$

o qual nos diz que a condição de Armijo não é satisfeita para $t = \frac{1}{2q}$. Assim, para $t = \frac{1}{2q}$ temos

$$\mathbf{F}\left(\mathbf{x}^{kr} + \left(\frac{1}{2q}\right)\mathbf{v}^{kr}\right) \not\leq \mathbf{F}(\mathbf{x}^{kr}) + \beta\left(\frac{1}{2q}\right)\mathbf{JF}(\mathbf{x}^{kr})\mathbf{v}^{kr},$$

isto é, existe j tal que

$$F_j\left(\mathbf{x}^{kr} + \left(\frac{1}{2q}\right)\mathbf{v}^{kr}\right) \geq F_j(\mathbf{x}^{kr}) + \beta\left(\frac{1}{2q}\right)(\mathbf{JF}(\mathbf{x}^{kr})\mathbf{v}^{kr})_j.$$

Portanto, fazendo $r \rightarrow \infty$ obtemos

$$F_j\left(\mathbf{y} + \left(\frac{1}{2q}\right)\bar{\mathbf{v}}\right) \geq F_j(\mathbf{y}) + \beta\left(\frac{1}{2q}\right)(\mathbf{JF}(\mathbf{y})\bar{\mathbf{v}})_j$$

para pelo menos um $j \in \{1, 2, \dots, m\}$. Note que esta desigualdade é válida para todo $q \in \mathbb{N}$. Daí, o Lema 2.3.1 implica que

$$JF(\mathbf{y})\bar{\mathbf{v}} \succeq 0,$$

isto é,

$$\max_{i=1,2,\dots,m} (JF(\mathbf{y})\bar{\mathbf{v}})_i \geq 0. \quad (2.12)$$

Logo, de (2.11) e (2.12) obtemos

$$0 \leq \max_{i=1,2,\dots,m} (JF(\mathbf{y})\bar{\mathbf{v}})_i \leq \frac{1}{\sigma} \max_i (F(\mathbf{y})\bar{\mathbf{v}})_i \leq \alpha(\mathbf{y}) \leq 0.$$

Portanto $\alpha(\mathbf{y}) = 0$ e \mathbf{y} é crítico Pareto.

Como a sequência $\{F(\mathbf{x}^k)\}$ é estritamente decrescente, componente a componente, concluímos que

$$F(\mathbf{x}^k) \prec F(\mathbf{x}^{k-1}) \prec F(\mathbf{x}^{k-2}) \prec \dots \prec F(\mathbf{x}^0),$$

$\text{logo}\{\mathbf{x}^k\} \in \mathbf{A}$. Assim, como \mathbf{A} é limitado (pois F é nível limitada) então $\{\mathbf{x}^k\}$ é limitada. Portanto, pelo Teorema de Bolzano-Weierstrass, $\{\mathbf{x}^k\}$ admite pelo menos um ponto de acumulação. \square

Capítulo 3

Método de Descida em Otimização Vetorial

Neste capítulo abordaremos um problema de otimização vetorial. Consideraremos aqui, uma ordenação parcial induzida por um cone $K \subset \mathbb{R}^m$ convexo, fechado e pontiagudo com interior diferente do vazio. O algoritmo utilizado é um algoritmo de descida. Além disso, apresentaremos definições e resultados importantes para a garantia de convergência do método. Mais detalhes podem ser encontrados em [6], [10], [14] e [18].

3.1 Introdução

Nesta seção estudaremos o problema

$$\min_K F(x) \tag{3.1}$$

A função $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ estudada, é convexa e continuamente diferenciável. O problema descrito é de minimizar a função F onde a ordenação é induzida por um cone K convexo, fechado e pontiagudo com interior diferente do vazio. Um método genérico de K -descida (AKdG) pode ser escrito da seguinte maneira.

Consideremos a constante $\beta \in (0, 1)$.

1. Inicializamos com $x^0 \in \mathbb{R}^n$.
2. Se x^k é K -crítico pare. Caso contrário,
3. Calcule v^k , uma K -direção de descida em x^k .

4. Calcule o tamanho de passo $t_k > 0$ tal que,

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

5. Encontre

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

e volte ao passo 2.

Definição 3.1.1 ([19]). Seja I um conjunto de índices e $\{x^k\}_{k \in I}$ uma sequência em \mathbb{R}^n . Diz-se que $\{x^k\}_{k \in I}$ é K -decrecente com relação a K quando

$$x^k \prec_K x^l \quad \forall k, l \in I; k > l.$$

Proposição 3.1.1. Seja $\{x^k\}$ uma sequência gerada pelo Algoritmo (AKdG). Se \bar{x} é um ponto de acumulação de $\{x^k\}$ então

$$F(\bar{x}) \preceq_K F(x^k)$$

Demonstração. Suponhamos $\{x^k\}$ uma sequência gerada pelo Algoritmo (AKdG), assim todos os x^k são não K -críticos. Logo existe v^k tal que

$$JF(x^k) v^k \in -\text{int}K.$$

Sendo assim,

$$\beta t_k JF(x^k) v^k \in -\text{int}K$$

e, conseqüentemente,

$$-\beta t_k JF(x^k) v^k \in \text{int}K. \tag{3.2}$$

Logo, pela definição do algoritmo temos que

$$F(x^{k+1}) \preceq_K F(x^k) + \beta t_k JF(x^k) v^k \prec_K F(x^k),$$

onde a última sentença segue de (3.2). Portanto, $\{F(x^k)\}$ é K -decrecente. Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $x^{k_j} \rightarrow \bar{x}$. Tomemos $k \in \mathbb{N}$ qualquer, para j suficientemente grande $k_j > k$ e

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

Assim, fazendo $j \rightarrow \infty$ obtemos

$$F(\bar{x}) \preceq_K F(x^k) \quad \forall k \in \mathbb{N}.$$

Tomemos agora \tilde{x} um outro ponto de acumulação da sequência $\{x^k\}$ e seja $\{x^{k_p}\}$ uma subsequência de $\{x^k\}$ tal que $x^{k_p} \rightarrow \tilde{x}$. Como

$$F(\bar{x}) \preceq_K F(x^k) \quad \forall k \in \mathbb{N},$$

temos que

$$F(\bar{x}) \preceq_K F(x^{k_p}).$$

Fazendo $p \rightarrow \infty$ temos

$$F(\bar{x}) \preceq_K F(\tilde{x}). \tag{3.3}$$

Invertendo os papéis de \bar{x} e \tilde{x} e procedendo analogamente, obteremos

$$F(\tilde{x}) \preceq_K F(\bar{x}). \tag{3.4}$$

Notemos que de (3.3)

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

e ainda de (3.4)

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

logo

$$-(F(\bar{x}) - F(\tilde{x})) = F(\tilde{x}) - F(\bar{x}) \in -K.$$

Como K é pontiagudo,

$$F(\tilde{x}) - F(\bar{x}) = 0$$

e, conseqüentemente,

$$F(\tilde{x}) = F(\bar{x}).$$

Assim F é constante no conjunto dos pontos de acumulação de $\{x^k\}$ e ainda

$$\lim_{k \rightarrow \infty} F(x^k) = F(\bar{x}).$$

□

3.2 Tamanho do Passo e Direção de Busca

Nesta seção discutiremos a escolha do tamanho de passo. Usaremos o procedimento *backtracking* usual para garantir que a escolha do comprimento de passo seja adequada em cada iteração k .

Consideraremos conhecida uma K -direção de descida v^k . Tome um $\beta \in (0, 1)$. Utilizaremos o seguinte procedimento para encontrar o tamanho de passo t :

1. Inicializamos com $t_0 = 1$ e verificamos se a seguinte condição é satisfeita:

$$F(x^k + tv^k) \preceq_K F(x^k) + \beta t JF(x^k)v^k. \quad (3.5)$$

2. Se (3.5) não é satisfeita, fazemos $t = \frac{t}{2}$ e retornamos ao passo 1. Caso contrário, tomamos $t_k = t$ como comprimento do passo.

Da Proposição 1.3.1 obtemos que o procedimento tem sempre uma terminação finita. Além disso,

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

Nosso objetivo agora é determinar uma direção de descida v^k . A seguir citaremos dois casos clássicos de determinação de direção de descida em otimização.

1. Na otimização escalar, isto é, quando $m = 1$ e $K = \mathbb{R}_+$ temos o conhecido método do Gradiente, $v^k = -\nabla F(x^k)$, que é solução do problema

$$\min \langle v, \nabla F(x^k) \rangle + \frac{1}{2} \|v\|^2, \quad v \in \mathbb{R}^n. \quad (3.6)$$

2. Para otimização multiobjetivo, isto é, quando $m > 1$ e $K = \mathbb{R}_+^m$, como foi visto no Capítulo anterior, v^k é solução do problema

$$\min f_x(v) + \frac{1}{2} \|v\|^2, \quad v \in \mathbb{R}^n \quad (3.7)$$

que pela definição de f_x é equivalente ao problema

$$\min \max_{i=1,2,\dots,m} \langle v, \nabla F_i(x^k) \rangle + \frac{1}{2} \|v\|^2, \quad v \in \mathbb{R}^n$$

onde $F(x) = (F_1(x), \dots, F_m(x))$.

Observe que (3.6) é um caso particular de (3.7) quando $m = 1$.

A partir de agora assumiremos que temos é um conjunto compacto $C \subseteq \mathbb{R}^n$ tal que

$$0 \notin C, \quad (3.8)$$

$$C \subset \text{Cone}(\text{Conv}C) = K^*. \quad (3.9)$$

Como $\text{int}K \neq \emptyset$ e $C \subseteq K^* \setminus \{0\}$, segue que $0 \notin \text{Conv}C$ e obtemos uma nova caracterização para os conjuntos $-K$ e $-\text{int}K$, sendo

$$-K = \{u \in \mathbb{R}^m \mid \langle u, w \rangle \leq 0 \quad \forall w \in C\} \quad (3.10)$$

e

$$-\text{int}K = \{u \in \mathbb{R}^m \mid \langle u, w \rangle < 0 \quad \forall w \in C\}. \quad (3.11)$$

Exemplo 3.2.1. Sendo $K = \mathbb{R}_+$ e $C = \{1\}$ retomamos a otimização escalar neste caso, temos

$$\begin{aligned} K^* &= \{w \in \mathbb{R} \mid \langle y, w \rangle \geq 0 \quad \forall y \in K\} = \mathbb{R}_+ \\ 0 &\notin C \\ \text{Conv}C &= \left\{ x \in \mathbb{R} \mid x = \sum_{i=1}^p \alpha_i x^i, \sum_{i=1}^p \alpha_i = 1, x^i \in C, \alpha_i \in \mathbb{R}_+ \right\} = \mathbb{R}_{++}. \end{aligned}$$

Além disso,

$$\begin{aligned} C &\subset \text{Cone}(\text{Conv}C) = \text{Cone}(\mathbb{R}_{++}) = \mathbb{R}_{++} \\ \text{int}K &\neq \emptyset, C \subset K^* \setminus \{0\} \Rightarrow 0 \notin \text{Conv}C. \end{aligned}$$

Pela definição, temos ainda:

$$\begin{aligned} -K &= \{u \in \mathbb{R} \mid \langle u, w \rangle \leq 0 \quad \forall w \in C\} = -\mathbb{R}_+ \\ -\text{int}K &= \{u \in \mathbb{R} \mid \langle u, w \rangle < 0 \quad \forall w \in C\} = -\mathbb{R}_{++}. \end{aligned}$$

Observação 3.2.1. Em otimização multiobjetivo, K e K^* são o \mathbb{R}_+^m e C é a base canônica do \mathbb{R}^m . Para um K genérico (convexo, fechado e pontiagudo com $\text{int}K \neq \emptyset$) temos

$$\begin{aligned} C &= \{w \in K^* \mid \|w\|_1 = 1\} \\ &= \{w \in K^* \mid |w|_1 + |w|_2 + \dots + |w|_m = 1\} \end{aligned}$$

Satisfaz as condições (3.8) e (3.9).

Definamos agora uma função $\phi : \mathbb{R}^m \rightarrow \mathbb{R}$ dada por

$$\phi(y) = \sup_{w \in C} \langle y, w \rangle.$$

Observação 3.2.2. *Seja C um conjunto compacto. A função $\phi : \mathbb{R}^m \rightarrow \mathbb{R}$ dada por*

$$\phi(\mathbf{y}) = \sup_{\mathbf{w} \in C} \langle \mathbf{y}, \mathbf{w} \rangle$$

é homogênea positiva.

Como C é compacto temos por (3.10) e (3.11) a seguinte caracterização para os conjuntos $-K$ e $-\text{int}K$:

$$-K = \{\mathbf{y} \in \mathbb{R}^m \mid \phi(\mathbf{y}) \leq 0\}, \quad (3.12)$$

$$-\text{int}K = \{\mathbf{y} \in \mathbb{R}^m \mid \phi(\mathbf{y}) < 0\}. \quad (3.13)$$

Lema 3.2.1. 1. *Sejam $\mathbf{y}, \bar{\mathbf{y}} \in \mathbb{R}^m$, então*

$$\phi(\mathbf{y} + \bar{\mathbf{y}}) \leq \phi(\mathbf{y}) + \phi(\bar{\mathbf{y}})$$

e ainda

$$\phi(\mathbf{y}) - \phi(\bar{\mathbf{y}}) \leq \phi(\mathbf{y} - \bar{\mathbf{y}}).$$

2. *Se $\mathbf{y}, \bar{\mathbf{y}} \in \mathbb{R}^m$, são tais que $\mathbf{y} \prec_K \bar{\mathbf{y}}$ ($\mathbf{y} \preceq_K \bar{\mathbf{y}}$), então $\phi(\mathbf{y}) < \phi(\bar{\mathbf{y}})$ ($\phi(\mathbf{y}) \leq \phi(\bar{\mathbf{y}})$).*

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

Demonstração. 1. De fato,

$$\begin{aligned} \phi(\mathbf{y} + \bar{\mathbf{y}}) &= \sup_{\mathbf{w} \in C} \langle \mathbf{y} + \bar{\mathbf{y}}, \mathbf{w} \rangle \\ &= \sup_{\mathbf{w} \in C} (\langle \mathbf{y}, \mathbf{w} \rangle + \langle \bar{\mathbf{y}}, \mathbf{w} \rangle) \\ &\leq \sup_{\mathbf{w} \in C} \langle \mathbf{y}, \mathbf{w} \rangle + \sup_{\mathbf{w} \in C} \langle \bar{\mathbf{y}}, \mathbf{w} \rangle \\ &= \phi(\mathbf{y}) + \phi(\bar{\mathbf{y}}). \end{aligned}$$

Por outro lado,

$$\begin{aligned} \phi(-\bar{\mathbf{y}} + \mathbf{y}) &= \sup_{\mathbf{w} \in C} \langle -\bar{\mathbf{y}} + \mathbf{y}, \mathbf{w} \rangle \\ &= \sup_{\mathbf{w} \in C} (\langle -\bar{\mathbf{y}}, \mathbf{w} \rangle + \langle \mathbf{y}, \mathbf{w} \rangle) \\ &\leq \sup_{\mathbf{w} \in C} \langle -\bar{\mathbf{y}}, \mathbf{w} \rangle + \sup_{\mathbf{w} \in C} \langle \mathbf{y}, \mathbf{w} \rangle. \end{aligned}$$

Assim,

$$\begin{aligned}\phi(\bar{\mathbf{y}} - \mathbf{y}) &= \phi[-(-\bar{\mathbf{y}} + \mathbf{y})] = -\phi(-\bar{\mathbf{y}} + \mathbf{y}) \\ &\geq \sup_{\mathbf{w} \in C} \langle \bar{\mathbf{y}}, \mathbf{w} \rangle - \sup_{\mathbf{w} \in C} \langle \mathbf{y}, \mathbf{w} \rangle \\ &= \phi(\bar{\mathbf{y}}) - \phi(\mathbf{y}).\end{aligned}$$

2. Se $\mathbf{y} \preceq_K \bar{\mathbf{y}}$ temos que $\mathbf{y} - \bar{\mathbf{y}} \in -K$ e, por (3.12)

$$\phi(\mathbf{y} - \bar{\mathbf{y}}) \leq 0.$$

Analogamente, se $\mathbf{y} \prec_K \bar{\mathbf{y}}$ temos que $\mathbf{y} - \bar{\mathbf{y}} \in -\text{int}K$ e, por (3.13)

$$\phi(\mathbf{y} - \bar{\mathbf{y}}) < 0.$$

3. Pelo item 1, temos que

$$\phi(\mathbf{y}) - \phi(\bar{\mathbf{y}}) \leq \phi(\mathbf{y} - \bar{\mathbf{y}}) \text{ e}$$

$$\phi(\bar{\mathbf{y}}) - \phi(\mathbf{y}) \leq \phi(\bar{\mathbf{y}} - \mathbf{y}).$$

Logo

$$|\phi(\mathbf{y}) - \phi(\bar{\mathbf{y}})| \leq \sup\{\phi(\mathbf{y} - \bar{\mathbf{y}}), \phi(\bar{\mathbf{y}} - \mathbf{y})\}.$$

Suponhamos sem perda de generalidade que

$$\sup\{\phi(\mathbf{y} - \bar{\mathbf{y}}), \phi(\bar{\mathbf{y}} - \mathbf{y})\} = \phi(\mathbf{y} - \bar{\mathbf{y}}).$$

Assim

$$\begin{aligned}|\phi(\mathbf{y}) - \phi(\bar{\mathbf{y}})| &\leq \phi(\mathbf{y} - \bar{\mathbf{y}}) \\ &= \sup_{\mathbf{w} \in C} \langle \mathbf{y} - \bar{\mathbf{y}}, \mathbf{w} \rangle \\ &\leq \sup_{\mathbf{w} \in C} (\|\mathbf{y} - \bar{\mathbf{y}}\| \cdot \|\mathbf{w}\|),\end{aligned}$$

pela desigualdade de Cauchy-Schwartz. Daí

$$|\phi(\mathbf{y}) - \phi(\bar{\mathbf{y}})| \leq \sup_{\mathbf{w} \in C} \|\mathbf{w}\| \cdot \|\mathbf{y} - \bar{\mathbf{y}}\| = L \|\mathbf{y} - \bar{\mathbf{y}}\|,$$

onde $L = \sup_{\mathbf{w} \in C} \|\mathbf{w}\|$ é a constante de Lipschitz. Portanto $\phi : \mathbb{R}^m \rightarrow \mathbb{R}$ é Lipschitz contínua.

□

Definamos agora para $x \in \mathbb{R}^n$ a função $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$, como sendo

$$\begin{aligned} f_x(v) &= \phi(JF(x)v) \\ &= \sup_{w \in C} \langle w, JF(x)v \rangle. \end{aligned} \quad (3.14)$$

Lema 3.2.2. *A função $f_x : \mathbb{R}^n \rightarrow \mathbb{R}$ dada na Definição 3.14 é convexa.*

Demonstração. De fato, se tomarmos o segmento $(1-t)z + ty$ com $t \in [0, 1]$ temos

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

Logo f_x é convexa. □

Lema 3.2.3. *Uma direção $v \in \mathbb{R}^n$ é uma K-direção de descida em $x \in \mathbb{R}^n$ se, e somente se, $f_x(v) < 0$.*

Demonstração. Se $v \in \mathbb{R}^n$ é uma K-direção de descida então

$$JF(x)v \in -\text{int}K.$$

logo por (3.13) e(3.14)

$$\phi(JF(x)v) = f_x(v) < 0.$$

Reciprocamente, se

$$f_x(v) = \phi(JF(x)v) < 0,$$

temos por (3.13)

$$JF(x)v \in -\text{int}K.$$

Logo $v \in \mathbb{R}^n$ é uma K-direção de descida em x . □

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

$$\begin{aligned} \min \quad & f_x(v) + \frac{1}{2}\|v\|^2 \\ \text{s.a} \quad & v \in \mathbb{R}^n. \end{aligned} \quad (3.15)$$

Denotamos o valor ótimo do problema por α_x .

Observação 3.2.3. 1. No caso escalar, onde $F : \mathbb{R}^n \rightarrow \mathbb{R}$ e $K = \mathbb{R}_+$ com $C = \{1\}$, a K -direção de descida é exatamente a clássica direção de descida $\mathbf{v}_x = -\nabla F(x)$.

2. Como $\mathbf{v} \rightarrow f_x(\mathbf{v})$ é uma função real, convexa e fechada, \mathbf{v}_x e α_x estão bem definidos. Além disso, como F é continuamente diferenciável e ϕ é Lipschitz contínua, a aplicação $(x, \mathbf{v}) \rightarrow f_x(\mathbf{v})$ é contínua.

Lema 3.2.4. 1. Se x é K -crítico, então $\mathbf{v}_x = 0$ e $\alpha_x = 0$.

2. Se x não é K -crítico, então $\mathbf{v}_x \neq 0$ e $\alpha_x < 0$ e

$$f_x(\mathbf{v}) < -\frac{1}{2}\|\mathbf{v}_x\|^2 < 0,$$

isto é, \mathbf{v}_x é uma K -direção de descida.

3. As aplicações $x \rightarrow \mathbf{v}_x$ e $x \rightarrow \alpha_x$ são contínuas.

Demonstração. 1. Se x é K -crítico temos

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

isto é, para todo $\mathbf{v} \in \mathbb{R}^n$,

$$JF(x)\mathbf{v} \notin -\text{int}K.$$

Assim,

$$f_x(\mathbf{v}) = \phi(JF(x)\mathbf{v}) \geq 0.$$

Como $f_x(0) = 0$ temos

$$0 = f_x(0) + \frac{1}{2}\|0\|^2 \leq f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \quad \forall \mathbf{v} \in \mathbb{R}^n.$$

Logo $\mathbf{v}_x = 0$ e $\alpha_x = 0$.

2. Se x não é K -crítico então existe $\bar{\mathbf{v}} \in \mathbb{R}^n$ tal que

$$JF(x)\bar{\mathbf{v}} \in -\text{int}K$$

e, conseqüentemente,

$$f_x(\bar{\mathbf{v}}) < 0.$$

Logo, fazendo $\lambda = \frac{-f_x(\mathbf{v})}{\|\mathbf{v}\|^2}$ e $\bar{\mathbf{v}} = \lambda\mathbf{v}$, obtemos

$$\begin{aligned} \alpha_x &\leq f_x(\bar{\mathbf{v}}) + \frac{1}{2}\|\bar{\mathbf{v}}\|^2 \\ &= f_x(\lambda\mathbf{v}) + \frac{1}{2}\|\lambda\mathbf{v}\|^2 \\ &= \lambda f_x(\mathbf{v}) + \frac{\lambda^2}{2}\|\mathbf{v}\|^2 \\ &= \frac{-f_x(\mathbf{v})}{\|\mathbf{v}\|^2} f_x(\mathbf{v}) + \frac{\left(\frac{-f_x(\mathbf{v})}{\|\mathbf{v}\|^2}\right)^2}{2}\|\mathbf{v}\|^2 \\ &= \frac{-f_x(\mathbf{v})^2}{\|\mathbf{v}\|^2} + \frac{f_x(\mathbf{v})^2}{2\|\mathbf{v}\|^4}\|\mathbf{v}\|^2 \\ &= \frac{-1}{2} \frac{f_x(\mathbf{v})^2}{\|\mathbf{v}\|^2} < 0. \end{aligned}$$

Assim,

$$\alpha_x = f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 < 0$$

e, conseqüentemente,

$$f_x(\mathbf{v}_x) < -\frac{1}{2}\|\mathbf{v}_x\|^2 < 0.$$

Portanto, do Lema 3.2.3, \mathbf{v}_x é uma K-direção de descida.

3. Tomemos $\mathbf{x}^0 \in \mathbb{R}^n$ e $\epsilon > 0$. Definamos

$$S := \{\mathbf{v} \in \mathbb{R}^n \mid \|\mathbf{v}_{x^0} - \mathbf{v}\| = \epsilon\},$$

onde \mathbf{v}_{x^0} é solução do problema (3.15) com $\mathbf{x} = \mathbf{x}^0$. Como f_x é convexa, a função objetivo do problema (3.15) é fortemente convexa com módulo $\frac{1}{2}$. Daí,

$$\begin{aligned} f_{x^0}((1-t)\mathbf{v} + t\mathbf{v}_{x^0}) &+ \frac{1}{2}\|(1-t)\mathbf{v} + t\mathbf{v}_{x^0}\|^2 \\ &\leq (1-t) \left(f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \right) + t \left(f_x(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 \right) \\ &\quad - \frac{1}{2}t\|\mathbf{v} - \mathbf{v}_{x^0}\|^2, \quad \forall t \in [0, 1]. \end{aligned}$$

Fazendo agora $t = 1$, obtemos

$$\begin{aligned} f_{x^0}(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 &\leq f_{x^0}(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 - \frac{1}{2}\|\mathbf{v} - \mathbf{v}_{x^0}\|^2 \\ &\leq f_{x^0}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 - \frac{1}{2}\|\mathbf{v} - \mathbf{v}_{x^0}\|^2 \quad \forall \mathbf{v} \in \mathbb{R}^n \\ &= f_{x^0}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 - \frac{1}{2}\epsilon^2 \quad \forall \mathbf{v} \in S. \end{aligned}$$

Portanto

$$f_{x^0}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \geq f_{x^0}(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 + \frac{1}{2}\epsilon^2 \quad \forall \mathbf{v} \in S.$$

Como a aplicação $(\mathbf{x}, \mathbf{v}) \rightarrow f_{\mathbf{x}}(\mathbf{v})$ é contínua e S é compacto concluímos que existe $\delta > 0$ tal que $\|\mathbf{x} - \mathbf{x}^0\| \leq \delta$ implica que

$$f_{\mathbf{x}}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 > f_{\mathbf{x}}(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 \quad \forall \mathbf{v} \in S. \quad (3.16)$$

Tomemos agora $\mathbf{x} \in \mathbb{R}^n$ tal que $\|\mathbf{x} - \mathbf{x}^0\| \leq \delta$. Da convexidade de $\mathbf{v} \rightarrow f_{\mathbf{x}}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2$, e da desigualdade (3.16) concluímos que \mathbf{v}_x , o minimizador de $f_{\mathbf{x}}(\cdot) + \frac{1}{2}\|\cdot\|^2$, não está na região $\|\mathbf{v} - \mathbf{v}_{x^0}\| \geq \epsilon$. De fato, seja $\mathbf{x} \in \mathbb{R}^n$ tal que $\|\mathbf{x} - \mathbf{x}^0\| \leq \delta$. Temos para todo $\mathbf{v} \in \mathbb{R}^n$ que

$$f_{\mathbf{x}}(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 \leq f_{\mathbf{x}}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2$$

e, conseqüentemente,

$$f_{\mathbf{x}}(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 \leq f_{\mathbf{x}}(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2. \quad (3.17)$$

Assim $\mathbf{v}_x \notin S$ pois, do contrário, (3.17) contradiria (3.16). Desse modo, ou $\|\mathbf{v}_x - \mathbf{v}_{x^0}\| > \epsilon$ ou $\|\mathbf{v}_x - \mathbf{v}_{x^0}\| < \epsilon$. Suponhamos que $\|\mathbf{v}_x - \mathbf{v}_{x^0}\| > \epsilon$. Como a função $\mathbf{v} \rightarrow f_{\mathbf{x}}(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2$ é fortemente convexa com módulo $\frac{1}{2}$ temos

$$\begin{aligned} f_{\mathbf{x}}((1-t)\mathbf{v}_x + t\mathbf{v}_{x^0}) + \frac{1}{2}\|(1-t)\mathbf{v}_x + t\mathbf{v}_{x^0}\|^2 \\ \leq (1-t) \left(f_{\mathbf{x}}(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 \right) + t \left(f_{\mathbf{x}}(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 \right) \\ - \frac{1}{2}t\|\mathbf{v}_x - \mathbf{v}_{x^0}\|^2, \quad \forall t \in [0, 1]. \end{aligned}$$

Fazendo agora $t = 1$ temos

$$f_x(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 \leq f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 - \frac{1}{2}\|\mathbf{v}_x - \mathbf{v}_{x^0}\|^2,$$

ou ainda,

$$f_x(\mathbf{v}_{x^0}) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 \geq f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_{x^0}\|^2 + \frac{1}{2}\|\mathbf{v}_x - \mathbf{v}_{x^0}\|^2.$$

Assim,

$$0 \geq \frac{1}{2}\|\mathbf{v}_x - \mathbf{v}_{x^0}\|^2 > \frac{1}{2}\epsilon^2,$$

o que é uma contradição. Logo,

$$\|\mathbf{x} - \mathbf{x}^0\| \leq \delta \Rightarrow \|\mathbf{v}_x - \mathbf{v}_{x^0}\| < \epsilon.$$

Portanto, \mathbf{v}_x é contínua e, conseqüentemente, α_x é contínua. □

Uma possível escolha para \mathbf{v}^k no Algoritmo (AKdG) é \mathbf{v}_{x^k} , isto é, a K -direção de descida em \mathbf{x}^k . No entanto como o cálculo de \mathbf{v}_x requer a solução do problema (3.15), é importante trabalharmos com a solução aproximada deste problema.

Definição 3.2.1. *Seja $\sigma \in [0, 1)$. Dizemos que \mathbf{v} é uma K -direção de descida σ -aproximada em $\mathbf{x} \in \mathbb{R}^n$ quando*

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

ou, equivalentemente,

$$f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 - \left(f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 \right) \leq \sigma|\alpha_x|.$$

Observação 3.2.4. *A K -direção de descida exata em \mathbf{x} é sempre uma K -direção de descida σ -aproximada, pois assumimos $\sigma \in [0, 1)$. A K -direção de descida exata é a única K -direção de descida σ -aproximada para $\sigma = 0$ pois, α_x é o valor ótimo.*

Lema 3.2.5. *Seja $\sigma \in [0, 1)$. Se \mathbf{v} é uma K -direção de descida σ -aproximada em \mathbf{x} então*

$$\|\mathbf{v}_x - \mathbf{v}\|^2 \leq 2\sigma|\alpha_x|.$$

Demonstração. De fato, a convexidade forte da função $\mathbf{v} \rightarrow f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2$ nos garante para $t = 1$ que

$$\begin{aligned} f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 &\leq f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 \\ &\leq f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 - \frac{1}{2}\|\mathbf{v}_x - \mathbf{v}\|^2, \end{aligned}$$

logo obtemos

$$\frac{1}{2}\|\mathbf{v}_x - \mathbf{v}\|^2 \leq f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 - \left(f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 \right).$$

Daí, como \mathbf{v} é uma K-direção de descida σ -aproximada temos

$$\|\mathbf{v}_x - \mathbf{v}\|^2 \leq 2 \left[f_x(\mathbf{v}) + \frac{1}{2}\|\mathbf{v}\|^2 - \left(f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 \right) \right] \leq 2\sigma|\alpha_x|.$$

□

Proposição 3.2.1. 1. A direção $\mathbf{v} = 0$ é uma K-direção de descida σ -aproximada em $\mathbf{x} \in \mathbb{R}^n$ se, e somente se, \mathbf{x} é K-crítico.

2. Se \mathbf{x} não é K-crítico e \mathbf{v} é uma K-direção de descida σ -aproximada em \mathbf{x} então \mathbf{v} é uma K-direção de descida. Em particular $\mathbf{v} \neq 0$.

Demonstração. 1. Se $\mathbf{v} = 0$ é uma K-direção de descida σ -aproximada em \mathbf{x} então pelo Lema 3.2.5

$$\|\mathbf{v}_x - 0\|^2 \leq 2\sigma|\alpha_x| \quad \forall \sigma \in [0, 1).$$

Logo, para $\sigma = 0$,

$$\|\mathbf{v}_x\|^2 \leq 0 \Rightarrow \mathbf{v}_x = 0$$

e, conseqüentemente,

$$\alpha_x = f_x(\mathbf{v}_x) + \frac{1}{2}\|\mathbf{v}_x\|^2 = 0.$$

Portanto, pelo Lema 3.2.4, \mathbf{x} é K-crítico. Reciprocamente, suponhamos que \mathbf{x} seja K-crítico. Então, pelo Lema 3.2.4, $\mathbf{v}_x = 0$ e $\alpha_x = 0$. Assim

$$f_x(0) + \frac{1}{2}\|0\|^2 = 0 \leq 0 = (1 - \sigma)\alpha_x \quad \forall \sigma \in [0, 1).$$

Logo, $\mathbf{v} = 0$ é uma K-direção de descida σ -aproximada em \mathbf{x} .

2. Sendo x não K -crítico então $v_x \neq 0$ e $\alpha_x < 0$ e sendo v uma K -direção de descida σ -aproximada em x temos

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

Portanto, pelo Lema 3.2.3, v é uma K -direção de descida e em particular $v \neq 0$. \square

Formalizaremos agora o método de K -descida (como método de K -Armijo) o qual denotaremos de Algoritmo (AKd). Este algoritmo é um caso particular do Algoritmo (AKdG) visto anteriormente. Consideraremos o seguinte:

Seja $\beta \in (0, 1)$ e $\sigma \in [0, 1)$.

1. Inicializamos com $x^0 \in \mathbb{R}^n$.
2. Se x^k é K -crítico, isto é, $f_{x^k} \geq 0$ para todo $v \in \mathbb{R}^n$ pare. Caso contrário,
3. Calcule v^k , uma um K -direção de descida σ -aproximada em x^k .
4. Calcule o tamanho de passo $t_k \in (0, 1]$ como sendo

$$t_k := \max \{2^{-j}, j \in \mathbb{N} \mid \Phi(F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j}JF(x^k)v^k) \leq 0\}.$$

5. Encontre

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

e retorne ao passo 2.

Observe que se x^k não é K -crítico então v^k obtida no passo 3 é uma K -direção de descida e assim t_k no passo 4 está bem definido. Além disso, t_k pode ser obtido por um procedimento *backtracking* usual como visto anteriormente.

Observação 3.2.5. Note que por (3.12), no passo 4, temos

$$t_k := \max \{2^{-j}, j \in \mathbb{N} \mid F(x^k + 2^{-j}v^k) \preceq_K F(x^k) + \beta 2^{-j}JF(x^k)v^k\}. \quad (3.18)$$

Além disso, x^{k+1} satisfaz

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

De fato, como

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

temos por (3.12) que

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

Logo,

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

e, conseqüentemente,

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

Assim, como visto na Proposição 3.1.1 temos

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

Portanto, a seqüência $\{F(x^k)\}$ é K -não-crescente.

Finalizaremos esta seção com uma generalização da Proposição 1.3.1, como conseqüência do fato de F ser continuamente diferenciável.

Proposição 3.2.2. *Sejam $\beta \in (0, 1)$, x e v tais que $JF(x)v \prec_K 0$. Então existem \bar{t}, δ e δ' constantes positivas tais que:*

1. v' é uma K -direção de descida em x' .

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

isto para todo $t \in (0, \bar{t})$, $x' \in B(x; \delta)$ e $v' \in B(v; \delta')$.

Demonstração. 1. Por hipótese, $JF(x)v \prec_K 0$, ou seja, $JF(x)v \in -\text{int}K$. Assim existe $\epsilon > 0$ tal que

$$JF(x)v + y \in -\text{int}K \quad \forall y \in \mathbb{R}^n, \|y\| \leq \epsilon. \quad (3.19)$$

Como JF é contínuo, existem δ_1 e δ_2 positivos tais que se

$$\|x - x'\| \leq \delta_1, \|v - v'\| \leq \delta_2 \Rightarrow \|JF(x')v' - JF(x)v\| \leq \frac{\epsilon}{2}. \quad (3.20)$$

Assim por (3.19),

$$JF(x')v' \in -\text{int}K \Rightarrow JF(x')v' \prec_K 0.$$

Logo v' é uma K -direção de descida em x' .

2. Como F é continuamente diferenciável, temos

$$F(z + tu) = F(z) + tJF(z)u + tR(z, tu),$$

com $\lim_{t \rightarrow 0} \|R(z, tu)\| = 0$ uniformemente para z e u num conjunto compacto.

Logo existe $\bar{t} > 0$ tal que

$$t \in (0, \bar{t}) \Rightarrow \|R(z, tu)\| \leq \bar{\epsilon},$$

em particular, para $t \in (0, \bar{t})$,

$$\|x - x'\| \leq \delta_1, \|v - v'\| \leq \delta_2 \Rightarrow \|R(x', tv')\| \leq \frac{\epsilon}{2}(1 - \beta). \quad (3.21)$$

Agora assumindo que $t \in (0, \bar{t})$, $\|x - x'\| \leq \delta_1$, $\|v - v'\| \leq \delta_2$, temos

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

Definamos agora

$$u := JF(x')v' + (1 - \beta)^{-1}R(x', tv'), \quad (3.22)$$

daí obtemos

$$F(x' + tv') = F(x') + t\beta JF(x')v' + t(1 - \beta)u. \quad (3.23)$$

Notemos agora que $u \prec_K 0$. De fato,

$$u = JF(x)v + \tilde{y}, \quad (3.24)$$

onde

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

Usando (3.20) e (3.21) obtemos

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

Logo por (3.19) e (3.24) concluímos que

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

Agora observemos que

$$\mathbf{F}(\mathbf{x}') + t\beta\mathbf{JF}(\mathbf{x}')\mathbf{v}' - \left(\mathbf{F}(\mathbf{x}') + t\beta\mathbf{JF}(\mathbf{x}')\mathbf{v}' + t(1-\beta)\mathbf{u} \right) = -t(1-\beta)\mathbf{u} \in \text{int}K,$$

isto é,

$$\mathbf{F}(\mathbf{x}') + t\beta\mathbf{JF}(\mathbf{x}')\mathbf{v}' + t(1-\beta)\mathbf{u} \prec_K \mathbf{F}(\mathbf{x}') + t\beta\mathbf{JF}(\mathbf{x}')\mathbf{v}', \quad (3.25)$$

para todo $t \in (0, \bar{t})$, $\|\mathbf{x} - \mathbf{x}'\| \leq \delta_1$, $\|\mathbf{v} - \mathbf{v}'\| \leq \delta_2$. Logo, como

$$\mathbf{F}(\mathbf{x}' + t\mathbf{v}') = \mathbf{F}(\mathbf{x}') + t\beta\mathbf{JF}(\mathbf{x}')\mathbf{v}' + t(1-\beta)\mathbf{u},$$

obtemos por (3.25):

$$\mathbf{F}(\mathbf{x}' + t\mathbf{v}') \prec_K \mathbf{F}(\mathbf{x}') + t\beta\mathbf{JF}(\mathbf{x}')\mathbf{v}'.$$

□

3.3 Análise de Convergência do Algoritmo

Nesta seção apresentaremos resultados auxiliares necessários para garantir a convergência do método (AKd). Consideraremos aqui as sequências $\{\mathbf{x}^k\}$, $\{\mathbf{v}^k\}$ e $\{t_k\}$ geradas pelo algoritmo. Note que se o algoritmo para após um número finito de iterações, este para em um ponto K-crítico.

Como visto no Lema 3.2.4, Definição 3.2.1 e Proposição 3.2.2,

$$\left\{ \begin{array}{l} \bullet \alpha_{x^k} < 0, \\ \bullet f_{x^k}(\mathbf{v}^k) + \frac{1}{2}\|\mathbf{v}^k\|^2 \leq (1-\sigma)\alpha_{x^k} < 0, \\ \bullet \mathbf{F}(\mathbf{x}^{k+1}) \preceq_K \mathbf{F}(\mathbf{x}^k) + \beta t_k \mathbf{JF}(\mathbf{x}^k)\mathbf{v}^k \preceq_K \mathbf{F}(\mathbf{x}^k). \end{array} \right.$$

Em particular, $\{\mathbf{F}(\mathbf{x}^k)\}$ é K-decrescente.

Lema 3.3.1. *Sejam $\{\mathbf{x}^k\}$, $\{\mathbf{v}^k\}$ e $\{t_k\}$ sequências geradas pelo Algoritmo (AKd). Então*

$$\phi(\mathbf{F}(\mathbf{x}^{k+1})) \leq \phi(\mathbf{F}(\mathbf{x}^k)) + \beta t_k \left((1-\sigma)\alpha_{x^k} - \frac{1}{2}\|\mathbf{v}^k\|^2 \right) \quad (3.26)$$

Demonstração. Com efeito, como

$$F(x^{k+1}) \preceq_K F(x^k) + \beta t_k JF(x^k)v^k,$$

temos pelo Lema 3.2.1

$$\begin{aligned} \phi(F(x^{k+1})) &\leq \phi(F(x^k) + \beta t_k JF(x^k)v^k) \\ &= \sup_{w \in C} \langle F(x^k) + \beta t_k JF(x^k)v^k, w \rangle \\ &= \sup_{w \in C} (\langle F(x^k), w \rangle + \beta t_k \langle JF(x^k)v^k, w \rangle) \\ &\leq \sup_{w \in C} \langle F(x^k), w \rangle + \beta t_k \sup_{w \in C} \langle JF(x^k)v^k, w \rangle \\ &= \phi(F(x^k)) + \beta t_k \phi(JF(x^k)v^k) \\ &= \phi(F(x^k)) + \beta t_k f_{x^k}(v^k) \\ &\leq \phi(F(x^k)) + \beta t_k \left((1 - \sigma)\alpha_{x^k} - \frac{1}{2}\|v^k\|^2 \right). \end{aligned}$$

□

Definição 3.3.1. Dizemos que uma sequência $\{x^k\}$ é *K-limitada* quando existe $\bar{y} \in \mathbb{R}^n$ tal que $\bar{y} \preceq_K x^k$ para todo k .

Exemplo 3.3.1. Toda sequência $\{x^k\}$ contida no cone K é *K-limitada*. De fato, se $\{x^k\} \subset K$, sempre existe um $\bar{y} \in \mathbb{R}^n$ tal que

$$x^k - \bar{y} \in K \quad \forall k.$$

Lema 3.3.2. Se $\{F(x^k)\}$ é *K-limitada* então

$$\sum t_k |\alpha_{x^k}| < \infty, \quad \sum t_k \|v^k\|^2 < \infty.$$

Demonstração. Com efeito, do Lema 3.3.1, obtemos

$$\begin{aligned}
 \phi(F(x^{n+1})) &\leq \phi(F(x^n)) + \beta t_n \left((1 - \sigma)\alpha_{x^n} - \frac{1}{2}\|v^n\|^2 \right) \\
 &\leq \phi(F(x^{n-1})) + \beta t_{n-1} \left((1 - \sigma)\alpha_{x^{n-1}} - \frac{1}{2}\|v^{n-1}\|^2 \right) \\
 &\quad + \beta t_n \left((1 - \sigma)\alpha_{x^n} - \frac{1}{2}\|v^n\|^2 \right) \leq \dots \\
 &\leq \phi(F(x^0)) + \sum_{k=0}^n \beta t_k \left((1 - \sigma)\alpha_{x^k} - \frac{1}{2}\|v^k\|^2 \right) \\
 &= \phi(F(x^0)) - \sum_{k=0}^n \beta t_k \left((1 - \sigma)|\alpha_{x^k}| - \frac{1}{2}\|v^k\|^2 \right).
 \end{aligned}$$

Como $\{F(x^k)\}$ é K -limitada, existe \bar{y} tal que

$$\bar{y} \preceq_K F(x^k) \quad \forall k.$$

Daí, pelo Lema 3.2.1, temos

$$\phi(\bar{y}) \leq \phi(F(x^k)).$$

Assim,

$$\phi(\bar{y}) \leq \phi(F(x^{n+1})) \leq \phi(F(x^0)) - \beta(1 - \sigma) \sum_{k=0}^n t_k |\alpha_{x^k}| - \frac{\beta}{2} \sum_{k=0}^n t_k \|v^k\|^2.$$

Logo,

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

e, equivalentemente,

$$\phi(F(x^0)) - \phi(\bar{y}) \geq \beta(1 - \sigma) \sum_{k=0}^n t_k |\alpha_{x^k}| + \frac{\beta}{2} \sum_{k=0}^n t_k \|v^k\|^2.$$

Fazendo $n \rightarrow \infty$ obtemos

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

e, conseqüentemente,

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

□

Agora estamos com as condições necessárias para estudar a convergência do Algoritmo (AKd).

Teorema 3.3.1. *Todo ponto de acumulação da seqüência $\{x^k\}$ gerada pelo Algoritmo (AKd) é K-crítico.*

Demonstração. Seja \bar{x} um ponto de acumulação de $\{x^k\}$. Então existe uma subsequência $\{x^{k_j}\}$ convergindo para \bar{x} , isto é,

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

Notemos que como $\{v_{x^{k_j}}\}$ e $\{\alpha_{x^{k_j}}\}$ são contínuas e $\{x^{k_j}\}$ converge para \bar{x} então $\{v_{x^{k_j}}\}$ e $\{\alpha_{x^{k_j}}\}$ convergem respectivamente para $v_{\bar{x}}$ e $\alpha_{\bar{x}}$ e, conseqüentemente, são limitadas. Logo, existem $a, b \in \mathbb{R}$ tais que

$$\|v_{x^{k_j}}\| \leq a \text{ e } |\alpha_{x^{k_j}}| \leq b.$$

Assim, pelo Lema 3.2.5, obtemos que

$$\|v^{k_j}\| - \|v_{x^{k_j}}\| \leq \|v^{k_j} - v_{x^{k_j}}\| \leq 2\sigma |\alpha_{x^{k_j}}|$$

e, portanto,

$$\|v^{k_j}\| \leq 2\sigma |\alpha_{x^{k_j}}| + \|v_{x^{k_j}}\| \leq 2\sigma b + a.$$

Refinando a subsequência original se necessário, assumimos que $v^{k_j} \rightarrow \bar{v}$. Como v^k é uma K-direção de descida σ -aproximada temos,

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

Logo, passando para a subsequência e fazendo $j \rightarrow \infty$ concluímos que

$$f_{\bar{x}}(\bar{v}) + \frac{1}{2} \|\bar{v}\|^2 \leq (1 - \sigma) \alpha_{\bar{x}}. \quad (3.27)$$

Como o Algoritmo (AKd) é um caso particular do Algoritmo (AKdG), como visto anteriormente, aplicando a Proposição 3.1.1, obtemos que $F(\bar{x})$ é K-limite inferior para $\{F(x^k)\}$. Agora, pelo Lema 3.3.2, segue que

$$\lim_{j \rightarrow \infty} t_{k_j} \alpha_{x^{k_j}} = 0, \quad (3.28)$$

$$\lim_{j \rightarrow \infty} t_{k_j} \|v^{k_j}\| = 0. \quad (3.29)$$

Afirmamos que

$$\bar{v} = 0. \quad (3.30)$$

Suponhamos por contradição que $\bar{v} \neq 0$. Como $\alpha_{\bar{x}} \leq 0$, usando (3.27) temos

$$f_{\bar{x}}(\bar{v}) \leq f_{\bar{x}}(\bar{v}) + \frac{1}{2} \|\bar{v}\|^2 \leq (1 - \sigma) \alpha_{\bar{x}} \leq 0$$

e, portanto, $f_{\bar{x}} \bar{v} < 0$. Logo, pelo Lema 3.2.3,

$$JF(\bar{x})\bar{v} \prec_K 0.$$

Pela Proposição 3.2.2 concluímos que existe $\bar{t} > 0$ tal que para todo j suficientemente grande (maior do que algum j_0)

$$F(x^{k_j} + tv^{k_j}) \preceq_K F(x^{k_j}) + \beta t JF(x^{k_j})v^{k_j} \quad \forall t \in [0, \bar{t}]. \quad (3.31)$$

Agora mostraremos que para j maior que tal j_0 ,

$$2t_{k_j} \geq \min\{1, \bar{t}\}. \quad (3.32)$$

De fato, seja $j > j_0$. Se $t_{k_j} = 1$, a desigualdade (3.32) é válida. Suponha $t_{k_j} < 1$, como este tamanho de passo é obtido por um procedimento *backtracking*, o tamanho de passo anterior possível $2t_{k_j}$ não satisfaz a condição de descida, uma vez que t_{k_j} é o máximo que satisfaz a condição. Assim, por (3.31)

$$2t_{k_j} \geq \min\{1, \bar{t}\}.$$

Daí, como $\bar{v} \neq 0$ (por hipótese), segue de (3.29) que

$$\lim_{j \rightarrow \infty} t_{k_j} = 0,$$

o que contradiz (3.32). Logo $\bar{v} = 0$. Portanto, de (3.27) e (3.30),

$$0 = f_{\bar{x}}(\bar{v}) + \frac{1}{2}\|\bar{v}\|^2 \leq (1 - \sigma)\alpha_{\bar{x}}$$

e, conseqüentemente,

$$\alpha_{\bar{x}} \geq 0. \tag{3.33}$$

Como x_{k_j} não é K-crítico, temos

$$\alpha_{x_{k_j}} < 0$$

logo,

$$\lim_{j \rightarrow \infty} \alpha_{x_{k_j}} = \alpha_{\bar{x}} \leq 0. \tag{3.34}$$

Portanto, de (3.33) e (3.34) obtemos $\alpha_{\bar{x}} = 0$ e, conseqüentemente, \bar{x} é K-crítico. \square

Mostramos portanto, que o algoritmo de descida assim como foi definido, gera uma seqüência $\{x^k\}$, onde qualquer ponto de acumulação desta seqüência é um ponto K-crítico de $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$, convexa e continuamente diferenciável, o qual era o objetivo deste capítulo.

Referências Bibliográficas

- [1] Burachick, R., Graña Drummond, L. M., Iusem, A. N. and Svaiter, B. F. *Full convergence of the steepest descent method with inexact line searches*. Optimization, 32:137-146, 1995.
- [2] Da Cruz Neto, J. X., Da Silva G. J. P., Ferreira, O. P. and Lopes, J. O. *A sub-gradient method for multiobjective optimization*. Computational Optimization and Applications, vol. 54, issue 3, pp 461-472, 2013.
- [3] De Weck, Olivier. L. *Multiobjective Optimization: History and Promise*. The Third China-Japan-Korea Joint Symposium on Optimization of Structural and Mechanical Systems. Kanazawa. CJK-OSM3, 2004.
- [4] Fliege, J. and Svaiter, B. F. *Steepest descent methods for multicriteria optimization*. Mathematical Methods of Operations Research, 51(3):479-494, 2000.
- [5] Fliege, J., Grana Drummond, L. M. and Svaiter, B. F., *Newton's method for multiobjective optimization*, SIAM J. Optim. 20, no. 2, 602-626. MR 2515788 (2010b:90117), 2009.
- [6] Graña Drommond, L. M. and Svaiter, B. F., *A steepest descent method for vector optimization*, J. Comput. Appl. Math. 175(2005), no. 2, 395-414, 2005.
- [7] Graña Drommond, L. M. and Iusem, A. N., *A projected gradient method for vector optimization problems*, Comput. Optim. Appl. 28, no. 1, 5-29. MR 2049673 (2005a:90139), 2004.
- [8] Izmailov, A. and Solodov, M. *Otimização - Volume 1: Condições de Otimalidade, Elementos de Análise Convexa e Dualidade*. IMPA, Rio de Janeiro, 2005.
- [9] Izmailov, A. and Solodov, M. *Otimização - Volume 2: Métodos Computacionais*. IMPA, Rio de Janeiro, 2007.

-
- [10] Jahn, J., *Vector Optimization: Theory, Applications and Extensions*. Second Edition. Springer. 2011.
- [11] Kreyszig, E., *Introductory Functional Analysis with Applications*, John Wiley, Sons, 1978.
- [12] Lima, E. L., *Curso de Análise, volume 1*. IMPA, Rio de Janeiro, Brasil, 2009.
- [13] Lima, E. L., *Curso de Análise, volume 2*. Projeto Euclides. IMPA, Rio de Janeiro, Brasil, 2009.
- [14] Luc, T. D. *Theory of Vector Optimization*, volume 319 of Lect. Notes in Econ. and Math. Systems. Springer-Verlag, 1989.
- [15] Pardalos, P. M. and Chinchuluun, A., *A survey of recent developments in multiobjective optimization*. Ann Oper Res, 154: 29-50, 2007.
- [16] Ribeiro, A. A. and Karas, E. W., *Um Curso de Otimização*. Cengage Learning Editora, A ser lançado em 2013.
- [17] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, New Jersey, 1970.
- [18] Takahashi, R. H. C., *Otimização Escalar e Vetorial*. Notas de Aula, Belo Horizonte, 2007. <http://www.mat.ufmg.br/taka/>.
- [19] Villacorta, K. D. V. *Método de Ponto Proximal para Otimização Vetorial e Método de Região de Confiança para Otimização multiobjetivo*. Tese de doutorado, Rio de Janeiro: COPPE/UFRJ, 2011.
- [20] Villacorta, K. D. V and Oliveira, P. R., *An interior proximal method in vector optimization*. European Journal of Operational Research, v. 214, p. 485-492, 2011.