Universidade Federal do Piauí
Centro de Ciências da Natureza
Pós-Graduação em Matemática
Mestrado em Matemática

Método do ponto proximal para um problema de otimização multiobjetivo não-convexo

Ray Victor Guimarães Serra

Teresina - 2016

Ray Victor Guimarães Serra

Dissertação de Mestrado:

Método do ponto proximal para um problema de otimização multiobjetivo não-convexo

Dissertação submetida à Coordenação do Programa 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 - 2016

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Serviço de Processamento Técnico

S487m Serra, Ray Victor Guimarães.
Método do Ponto Proximal para um problema de otimização multiobjetivo não-convexo / Ray Victor Guimarães Serra. -- 2016.
80 f. : il.

Dissertação (Mestrado) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Pós-Graduação em Matemática, Teresina, 2016.
"Orientação: Profª. Drª. Sissy da Silva Souza."

1. Método do Ponto Proximal. 2. Otimização multiobjetivo. 3. Otimização não-convexa. I. Título.

CDD 512.5

Dedico este trabalho aos meus pais e irmã;
às minhas avós, Antônia Serra (In memoriam) e Raimunda Paixão; à toda minha família.

Agradecimentos

Chegando ao fim de mais uma etapa, agradeço primeiramente a Deus por me dar força de vontade o suficiente para chegar até aqui. São tantas as pessoas que quero e devo agradecer pela importância que tiveram para conseguir mais esse objetivo. Agradeço primeiramente meus pais e irmã, por serem a minha base e pelo apoio incondicional em todos os momentos, não existem palavras que descrevam meu sentimento e agradecimento por vocês, muito obrigado!

Agradeço a toda minha família a qual compartilho essa vitória, pelo apoio e confiança sem igual, em especial às minhas avós Antônia Serra (In memoriam) e Raimunda Paixão, às minhas tias (Cristina, Edmea, Iranilda, Josefa, Maria José), tios (Artur, Iranildo, Jesus, Marco Jorge) e primos.

Aos meus professores/mestres que tive até chegar aqui, entre eles o Professor Gilson Santana pelo grande incentivo e apoio desde o ensino médio até o presente momento. A todos os Professores do campus de Parnaíba, em especial Cleyton Cunha, Pedro Jorge, Clóris Violeta, Paulo Sérgio e Sissy Souza pelos ensinamentos, apoio e principalmente amizade ao longo desses anos. Obrigado!

A todos os professores do mestrado, em especial os professores Jurandir, José Francisco, Marcondes e Barnabé. Aqui incluo novamente meu agradecimento para Sissy Souza e Paulo Sérgio pela excelente orientação, ensinamentos repassados e amizade desde a graduação.

Aos meus amigos de Parnaíba, em especial Herbet Micael, Amanda Thaína, Victória Régia (In memoriam) pela amizade desde sempre. Aos amigos de graduação, com ênfase aos parceiros Maciel Santos e Julliette Araújo pela enorme amizade. Por fim, agradeço algumas pessoas que tive o privilégio de ter em minha vida e que serei eternamente grato por isso, são elas Renata Brito, George Santos, Fernando Santos e em especial Carla Brito a quem agradeço por toda a confiança depositada em mim, meu muito obrigado pela amizade e companheirismo incondicional, apesar da distância.

Aos amigos do mestrado Rafael (Coxinha), Veludex (Antônio Aguiar), Raul, Atécio (Rico), Hércules (Semi-Deus), Tiago (Balada), Jeferson (Poker), Kelvin (Cabelim), Juliana, Quaresma, Lívio, Bruno, Fernando Lima, Fernando Gomes e à todos os outros com ênfase aos amigos Pádua (Pádua) e Lucas (Foto) companheiros de graduação.

Agradeço a CAPES pelo apoio financeiro.

"A alegria está na luta, na tentativa, no sofrimento envolvido e não na vitória propriamente dita.".

Mahatma Gandhi.

Resumo

Neste trabalho, apresentamos uma generalização do Método do Ponto Proximal para minimizar uma função vetorial, não necessariamente convexa, no contexto de espaços euclidianos. No problema estudado, as coordenadas da função vetorial são definidas como o máximo de funções continuamente diferenciáveis. Iniciamos estudando o Método do Ponto Proximal clássico para o Problema de Otimização Multiobjetivo e ferramentas de derivação generalizada no sentido de Clarke. Como resultados, temos que qualquer ponto de acumulação da sequência gerada pelo método é um ponto crítico Pareto-Clarke. Além disso apresentamos a convergência do Método do Ponto Proximal para um ponto crítico Pareto fraco.

Abstract

In this work, we present a generalized proximal point method for minimizing a vectorial function, not necessarily convex, in the setting of Euclidean Spaces. In the problem studied, the coordinates functions of the vectorial function are defined by the maximum of continuously differentiable functions. For this, we started studying of the Classical Proximal Point Method for the Multiobjective Optimization Problem and tools of Clarke's generalized derivate. As results, we have that any cluster point of the sequence generated by the method is a Pareto-Clarke stationary point. Moreover, we present a convergence of the Proximal Point Method for a weak Pareto stationary point.

Sumário

Introdução

Em nosso cotidiano temos várias tomadas de decisões, onde é natural querer que as escolhas sejam as melhores possíveis, em outras palavras, desejamos sempre tomar a decisão ideal para cada problema. A dificuldade na maioria dos casos está no conflito existente entre nossos objetivos e metas. Por isso nas situações do cotidiano acabamos escolhendo nossas decisões baseadas na intuição, sem ter uma precisão de que tal escolha é a melhor possível. Porém em alguns casos a intuição não é o bastante para garantir um resultado satisfatório. O presente trabalho busca fazer um breve estudo desses problemas com múltiplos objetivos. Daí, fala-se de Otimização Multiobjetivo, onde buscamos otimizar simultaneamente duas ou mais funções, as quais chamamos de funções objetivos. Em geral, não há um único ponto que irá minimizar ou maximizar todas as funções objetivos ao mesmo tempo. Nesses casos estamos interessados em buscar um ponto que chamamos de ótimo Pareto, ou ainda como iremos ver, ponto crítico Pareto-Clarke, ver [6]. Os problemas de Otimização Multiobjetivo têm inúmeras aplicações, entre elas na biologia, estatística e engenharia como pode ser visto em [15], [2] e [9] respectivamente.

Nosso objetivo aqui será o de apresentar o que é um problema de Otimização Multiobjetivo e além disso abordar um método para resolver tal problema. A solução em questão será uma generalização do chamado Método do Ponto Proximal clássico, onde trabalhamos com uma função objetivo escalar e convexa. O Método do Ponto Proximal para problemas de Otimização Multiobjetivo é introduzido por Bonnel em [5], onde trabalha-se com uma certa classe de funções não-convexas. Esta dissertação tem como inspiração os trabalhos feitos por Bonnel, Iusem e Svaiter em [5], G. Bento, O. Ferreira e V. L. Sousa Junior em [4], G. Bento, J. X. Cruz Neto e P. R. Oliveira em [12]. É importante ressaltar que outros métodos, associados à otimização escalar, já foram estendidos a Otimização Multiobjetivo, entre eles [10], [11], [13] e [14].

Para entendermos tal generalização do Método do Ponto Proximal Clássico o presente trabalho apresenta no Capítulo 1 alguns resultados preliminares vistos em [17] a cerca de Análise convexa, em especial a noção de subdiferencial de uma função convexa, entre outros resultados que serão utilizados no decorrer deste trabalho.

No Capítulo 2, baseado em [16] recordamos o Método do Ponto Proximal clássico, bem como a noção de Féjer convergência que será também de grande importância para o resultado principal deste trabalho.

No Capítulo 3 abordaremos alguns tópicos de Otimização Multiobjetivo. Entre eles iremos definir um problema de Otimização Multiobjetivo e caracterizar a solução desse problema. Além disso, com o objetivo de facilitar a demonstrção do resultado principal deste trabalho iremos introduzir também a noção de Escalarização. Tal técnica tem como objetivo substituir um problema de Otimização Multiobjetivo, onde não temos uma relação de ordem total, por um problema de Otimização Escalar ou uma série de problemas escalares, onde podemos aplicar alguns dos resultados clássicos da literatura abordados no Capítulo 1.

No Capítulo 4 baseado em [6] buscamos generalizar alguns dos resultados vistos no Capítulo 1, em especial o conceito de subdiferencial de uma função convexa. Generalizamos a derivada usual do $\mathbb{R}^n$, abordando a derivada generalizada no sentindo de Clarke, buscando fazer um comparativo de ambas.

No Capítulo 5 apresentamos a generalização do Método do Ponto Proximal clássico para um problema de Otimização Multiobjetivo não-convexo abordado em [4]. Inicialmente mostramos que qualquer ponto de acumulação da sequência gerada pelo método é um ponto crítico Pareto-Clarke. Por fim, adaptando alguns resultados de [5] fazemos a análise da convergência do método, provando que tal método converge para uma solução fraca, isto é, um ponto crítico Pareto fraco.

Capítulo 1

Preliminares

Neste capítulo estudaremos alguns fatos básicos que serão necessários ao longo deste trabalho. Iremos abordar resultados que garantem a existência de soluções do seguinte problema:

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

onde $f : \mathbb{R}^n \to \mathbb{R}$. Veremos sob quais hipóteses nosso problema tem solução. Além disso, estudaremos as condições necessárias e suficientes de otimalidade para o Problema (1.1). Por fim, abordaremos alguns tópicos de Análise convexa, procurando esclarecer a importância da convexidade na otimização. Maiores detalhes dos resultados abordados neste Capítulo poderão ser encontrados em [17], [21], [22] e [4].

1.1 Definições e alguns fatos básicos

Definição 1.1.1. *([17]) Dizemos que um ponto* $\bar{x} \in \mathbb{R}^n$ *é*

1. ***Minimizador global*** *de* (1.1) *se*

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

2. ***Minimizador local*** *de* (1.1) *se existe uma vizinhança* U *de* $\bar{x}$ *tal que*

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

Se $\bar{x}$ é minimizador global então $f(\bar{x})$ é chamado **valor ótimo global**.

Definição 1.1.2. *([17]) Dizemos que* $\bar{v} \in [-\infty, +\infty)$ *definido por*

$$\bar{v} = \inf_{x \in \mathbb{R}^n} f(x),$$

é o valor ótimo do Problema (1.1).

Uma função pode admitir vários minimizadores globais, mas o valor ótimo (global) do problema é único. Apresentamos a seguir uma condição necessária de primeira ordem para caracterizar um minimizador para o Problema (1.1).

Teorema 1.1.1. *([17])* ***(Condição necessária de primeira ordem)*** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *diferenciável no ponto* $\bar{x} \in \mathbb{R}^n$*. Se* $\bar{x}$ *é um minimizador local de* f*, então*

$$f'(\bar{x}) = 0. \tag{1.2}$$

Demonstração. Seja $d \in \mathbb{R}^n$ arbitrário, pela definição de minimizador local, temos que existe $\epsilon > 0$ tal que

$$f(\bar{x}) \leqslant f(\bar{x} + td), \ \forall t \in (0, \epsilon].$$

Daí, pela diferenciabilidade de f em $\bar{x}$,

$$f(\bar{x} + td) = f(\bar{x}) + t\langle f^{'}(\bar{x}), d\rangle + \theta(t),$$

onde $\theta(t)$ segue da definição de função diferenciável e é tal que $\lim_{t \to 0^+} \dfrac{\theta(t)}{t} = 0$.

$$0 \leqslant t\langle f^{'}(\bar{x}), d\rangle + \theta(t).$$

Dividindo por $t > 0$, obtemos

$$0 \leqslant \langle f^{'}(\bar{x}), d\rangle + \frac{\theta(t)}{t},$$

passando o limite quando $t \longrightarrow 0^+$, temos que

$$0 \leqslant \langle f^{'}(\bar{x}), d\rangle.$$

Como $d \in \mathbb{R}^n$ é fixo, porém arbitrário, podemos escolher $d = -f^{'}(\bar{x})$, o que resulta na condição $0 \leqslant \langle f^{'}(\bar{x}), d\rangle = -\|f^{'}(\bar{x})\|^2$. Com isso, segue que $f^{'}(\bar{x}) = 0$. □

Definição 1.1.3. *Um ponto* $\bar{x} \in \mathbb{R}^n$ *que satisfaz a condição* (1.2) *é chamado de ponto crítico ou estacionário da função* f.

A partir do Teorema 1.1.1. e Definição 1.1.3., temos que se f é diferenciável então as soluções locais do problema de minimização (1.1) são pontos estacionários. O mesmo valendo para problemas de maximização.

Abordaremos a seguir um resultado que é conhecido na literatura como Fórmula de Taylor, onde a prova poderá ser encontrada em ([21]).

Teorema 1.1.2. *Seja* $f : U \longrightarrow \mathbb{R}^n$ *de classe* C^2 *no aberto* $U \subset \mathbb{R}^n$. *Fixado* $a \in U$, *para todo* $v = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$ *tal que* $a + v \in U$, *escrevamos*

$$f(a+v) - f(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\alpha_i + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\alpha_i\alpha_j + \theta(v),$$

as derivadas sendo calculadas no ponto a. *Então* $\lim_{v \to 0} \frac{\theta(v)}{|v|^2} = 0$.

Teorema 1.1.3. *([17]) (**Condição necessária de segunda ordem**) Suponhamos que* $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciável em* $\bar{x} \in \mathbb{R}^n$. *Se* $\bar{x}$ *é um minimizador local irrestrito de* f, *então vale* (1.2) *e a matriz hessiana de* f *em* $\bar{x}$ *é semidefinida positiva, ou seja,*

$$\langle f''(\bar{x})d, d\rangle \geqslant 0, \quad \forall d \in \mathbb{R}^n. \tag{1.3}$$

Demonstração. A condição (1.2) é satisfeita pelo Teorema 1.1.1.. Agora, seja $d \in \mathbb{R}^n$ qualquer, porém fixo. Se $\bar{x}$ é minimizador local do Problema (1.1), então para todo t estritamente positivo e suficientemente pequeno temos que

$$\begin{aligned} 0 &\leqslant f(\bar{x}+td) - f(\bar{x}) \\ &= \langle f'(\bar{x}), td\rangle + \frac{1}{2}\langle f''(\bar{x})td, td\rangle + \theta(t^2) \\ &= \frac{t^2}{2}\langle f''(\bar{x})d, d\rangle + \theta(t^2), \end{aligned}$$

onde acima usamos a expansão da função em série de Taylor (Ver [17]) e novamente o Teorema 1.1.1. Daí, dividindo ambos os lados por $t^2 > 0$, temos que

$$0 \leqslant \frac{1}{2}\langle f''(\bar{x})d, d\rangle + \frac{\theta(t^2)}{t^2}.$$

Por fim, passando o limite quando $t \to 0^+$ e usando o Teorema 1.1.2. citado acima temos que $\langle f''(\bar{x})d, d)\rangle \geqslant 0, \forall d \in \mathbb{R}^n$. $\square$

A recíproca do teorema anterior é verdadeira sob uma condição mais forte sobre a matriz hessiana de f, como enunciaremos abaixo.

Teorema 1.1.4. *([17]) (**Condição suficiente de segunda ordem**)*
Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$ *seja duas vezes diferenciáveis em* $\bar{x} \in \mathbb{R}^n$. *Se* $\bar{x}$ *é um ponto estacionário e se a matriz hessiana de* f *em* $\bar{x}$ *é definida positiva então existem uma vizinhança* U *de* $\bar{x}$ *e* $\beta > 0$ *tais que*

$$f(x) - f(\bar{x}) \geqslant \beta\|x - \bar{x}\|^2, \ \forall x \in U. \tag{1.4}$$

Em particular, $\bar{x}$ *é um minimizador local estrito do Problema* (1.1)*. Reciprocamente, se vale* (1.4)*, então* $\bar{x}$ *é um ponto estacionário do Problema* (1.1) *e tem matriz hessiana definida positiva.*

Demonstração. Seja $\bar{x}$ ponto estacionário do Problema (1.1) e que a hessiana seja definida positiva, i.e.,

$$\langle f''(x)d, d\rangle > 0, \ \forall d \in \mathbb{R}^n \setminus \{0\}.$$

Suponha por absurdo que não ocorra (1.4). Então, existe $\{x^k\} \subset \mathbb{R}^n \setminus \{x^k\}$ tal que $x^k \to \bar{x}$ e

$$f(x^k) - f(\bar{x}) < \frac{1}{k}\|x^k - \bar{x}\|^2, \quad \forall k.$$

Daí, como $\left\{\dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|}\right\}$ é limitada, ela possui pontos de acumulação. Portanto, a menos de subsequência, podemos supor sem perda de generalidade que $\left\{\dfrac{x^k - \bar{x}}{\|x^k - \bar{x}\|}\right\} \to d \in \mathbb{R}^n \setminus \{0\}$. Logo, temos que

$$\begin{aligned} \frac{1}{k}\|x^k - \bar{x}\|^2 &> f(x^k) - f(\bar{x}) \\ &= \langle f'(\bar{x}), x^k - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + \theta(\|x^k - \bar{x}\|^2) \\ &= \frac{1}{2}\langle f''(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + \theta(\|x^k - \bar{x}\|^2), \end{aligned}$$

onde acima usamos que $f'(\bar{x}) = 0$. Dividindo ambos os membros desta desigualdade por $\|x^k - \bar{x}\|^2$ e tomando o limite quando $k \to +\infty$, obtemos

$$0 \geqslant \langle f''(\bar{x})d, d\rangle,$$

o que gera uma contradição, pois temos por hipótese que a hessiana é definida positiva. Portanto, temos que vale (1.4). Em particular,

$$f(x) \geqslant f(\bar{x}), \ \forall x \in U,$$

ou seja, $\bar{x}$ é um minimizador local.

Reciprocamente, suponha que valha (1.4), então como tal condição implica que $\bar{x}$ é minimizador local de f então obtemos que tal ponto é estacionário (vide Teorema 1.1.1.). Seja $d \in \mathbb{R}^n$ arbitrário porém fixo, então para todo $t \in \mathbb{R}$ suficientemente pequeno $\bar{x}+td \in U$. Logo, usando (1.4) e o fato de $\bar{x}$ ser ponto estacionário obtemos que

$$\begin{array}{rcl} \beta t^2\|d\|^2 & = & \beta\|(\bar{x}+td)-\bar{x}\|^2 \\ & \leqslant & f(\bar{x}+td)-f(\bar{x}) \\ & = & t\langle f'(\bar{x}), d\rangle + \dfrac{t^2}{2}\langle f''(\bar{x})d, d\rangle + \theta(t^2) \\ & = & \dfrac{t^2}{2}\langle f''(\bar{x})d, d\rangle + \theta(t^2). \end{array}.$$

Dividindo ambos os lados por t^2 e fazendo $t \to 0$, temos que

$$\beta\|d\|^2 \leqslant \frac{1}{2}\langle f''(\bar{x})d, d\rangle.$$

Portanto, como d é arbitrário temos que a hessiana é definida positiva. □

Note que as condições (1.2) e (1.3) não são suficientes para garantir que um ponto $\bar{x}$ é minimizador. Por exemplo, a função $f : \mathbb{R} \longrightarrow \mathbb{R}$ dada por $f(x) = x^3$, satisfaz tais condições em $\bar{x} = 0$, mas tal ponto não é minimizador local do Problema (1.1). Por outro lado, (1.2) e o fato da hessiana ser definida positiva não são condições necessárias para que um ponto $\bar{x}$ seja minimizador. Como exemplo, podemos tomar $f : \mathbb{R} \longrightarrow \mathbb{R}$ dada por $f(x) = x^6$. Note que $\bar{x} = 0$ é ponto minimizador de (1.1). Mas essa função não possui hessiana definida positiva nesse ponto.

1.2 Existência de soluções

No caso em que f é ilimitada inferiormente na Definição 1.1.2, temos que $\bar{v} = -\infty$, e o problema de minimização não possui solução global. Porém, mesmo quando $\bar{v}$ é finito o minimizador pode não existir, como é o caso de $f(x) = e^x$ em $\mathbb{R}$. Nesta seção teremos por objetivo estudar alguns resultados básicos, porém indispensáveis para garantir a existência de soluções para nosso Problema (1.1). Por exemplo, a continuidade da função estudada e a compacidade do conjunto em questão garantem a existência de um minimizador em tal conjunto, como veremos a seguir. Alguns desses resultados são bem conhecidos na literatura e por isso algumas demonstrações serão omitidas.

O primeiro resultado desta seção traz a importância da compacidade do domínio e continuidade da função estudada para garantirmos a existência de solução do problema de maximização e minimização. Tal resultado é bem conhecido na literatura, em especial na Análise Matemática, como Teorema de Weierstrass, onde sua demonstração pode ser encontrada em ([17]).

Teorema 1.2.1. *([17]) Sejam* $D \subset \mathbb{R}^n$ *um conjunto compacto não-vazio e* $f : D \to \mathbb{R}^n$ *uma função contínua. Então, os problemas de minimização e maximização de* f *em D têm soluções globais.*

Como explica o autor em [17] o exemplo da função $f(x) = e^x$ mencionada acima reforça a importância do conjunto(viável) em questão ser compacto. Daí, tal hipótese só poderá ser retirada e ainda assim garantirmos a existência de soluções se reforçamos as hipóteses com respeito a função objetivo em questão. Daí, surge a importância da seguinte definição:

Definição 1.2.1. *([17]) O conjunto de nível da função* $f : D \to \mathbb{R}$ *associado a* $c \in \mathbb{R}$*, é o conjunto dado por*

$$L_{f,D}(c) = \{x \in D \mid f(x) \leqslant c\}.$$

Corolário 1.2.1. *([17]) Sejam* $D \subset \mathbb{R}^n$ *e* $f : D \to \mathbb{R}$ *contínua no conjunto D. Suponhamos que existe* $c \in \mathbb{R}$ *tal que o conjunto de nível* $L_{f,D}(c)$ *seja não-vazio e compacto. Então o problema de minimização* f *em D possui uma solução global.*

Demonstração. Pelo Teorema 1.2.1 o Problema (1.1) tem solução global em $L_{f,D}(c)$, digamos $\bar{x}$, isto é, $f(\bar{x}) \leqslant f(x), \forall x \in L_{f,D}(c)$. Agora seja $x \in D \setminus L_{f,D}(c)$ qualquer, temos que $f(x) > c \geqslant f(\bar{x})$, em outras palavras $\bar{x}$ é um minimizador global de f em todo conjunto D.

□

Definição 1.2.2. *Dizemos que* $f : D \to \mathbb{R}$ *é coerciva no conjunto D, quando para cada sequência* $\{x^k\} \subset D$ *tal que ou* $\| x^k \| \to \infty$ *ou* $\{x^k\} \to x \in \bar{D} \setminus D (k \to \infty)$*,tem-se* $\lim_{k\to\infty} \sup\, f(x^k) = +\infty$.

Corolário 1.2.2. *([17]) Sejam* $D \subset \mathbb{R}^n$ *e* $f : D \to \mathbb{R}$ *uma função contínua coerciva em* $D \neq \emptyset$*. Então, o problema de minimizar* f *em D posui uma solução global.*

Demonstração. Com $x \in D$ arbitrário, definimos $c = f(x)$. Daí, obtemos que $L_{f,D}(c) \neq \emptyset$. Iremos mostrar que tal conjunto é compacto. Suponhamos inicialmente que exista $\{x^k\} \subset L_{f,D}(c)$ tal que $\{x^k\} \longrightarrow x$, onde $x \notin L_{f,D}(c)$. Usando que

$$L_{f,D}(c) = D \cap \{x \in \mathbb{R}^n \mid f(x) \leqslant c\},$$

onde o segundo conjunto é fechado pelo fato de f ser contínua por hipótese. Isto nos diz que $\{x^k\} \longrightarrow x \in \bar{D} \setminus D$. Com isso, devido a coercividade de f em D temos que $\lim_{k\to\infty} \sup f(x^k) = +\infty$. Mas isso contradiz o fato de $\{x^k\} \subset L_{f,D}(c)$. Portanto, obtemos que $L_{f,D}(c)$ é fechado.

Agora suponha $L_{f,D}(c)$ ilimitado, isto é, podemos tomar uma sequência $\{x^k\} \subset L_{f,D}(c)$ tal que $\{x^k\} \longrightarrow +\infty$. Então pela fato de f ser coerciva em D e $\{x^k\} \subset D$, $\lim_{k\to\infty} \sup f(x^k) = +\infty$. Mas $f(x^k) \leqslant c, \ \forall\ k \in \mathbb{N}$. Portanto, $L_{f,D}(c)$ é limitado. □

1.3 Elementos de Análise convexa

Nesta seção apresentamos tópicos de Análise Convexa, abordando alguns fatos básicos sobre conjuntos convexos e funções convexas que serão importantes para o desenvolvimento deste trabalho. É grande a importância da convexidade na otimização, pois com tal hipótese, as condições necessárias de otimalidade se tornam suficientes . Isto é, todo ponto estacionário é solução do problema de minimização. Além disso, o estudo de alguns resultados de análise convexa são de grande valia para termos um maior entendimento do teorema principal deste trabalho.

Definição 1.3.1. *([17]) Um conjunto $C \subset \mathbb{R}^n$ é convexo se, para quaisquer elementos $x, \ y \in C$, temos*

$$tx + (1-t)y \in C \text{ para todo } 0 \leqslant t \leqslant 1.$$

Em outras palavras, se x e y são pontos quaisquer do conjunto C então o segmento de reta com extremos x e y pertence à C. É chamado de combinação convexa de x e y o ponto $tx + (1-t)y$. Exemplos triviais de conjunto convexo é o conjunto vazio, o espaço n-dimensional $\mathbb{R}^n$, um conjunto que contém só um ponto.

Exemplo 1.3.1. *Todo semi-espaço em $\mathbb{R}^n$, isto é,*

$$C = \{x \in \mathbb{R}^n : \langle a, x \rangle \leqslant c\},$$

onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$ é convexo.

Demonstração. De fato, sejam $x, y \in C$, logo $\langle a, x\rangle \leqslant c$ e $\langle a, y\rangle \leqslant c$. Então, para $tx + (1-t)y$, com $t \in [0, 1]$, temos:

$$\langle a, tx + (1-t)y\rangle \leqslant tc + (1-t)c = c.$$

Portanto, $tx + (1-t)y \in C$ e, assim, C é um conjunto convexo. □

Definição 1.3.2. *([17]) Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita **convexa** quando para todos $x, y \in \mathbb{R}^n$ tivermos*

$$f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y), \tag{1.5}$$

para todo $t \in [0, 1]$. A função f diz-se estritamente convexa quando a desigualdade acima é estrita para qualquer $x \neq y$ e $t \in (0, 1)$.

Observação 1.3.1. *Note que a condição (1.5) é equivalente a $f(x + t(y - x)) \leqslant f(x) + t\,(f(y) - f(x))$.*

Definição 1.3.3. *([17]) Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita **fortemente convexa** com módulo $L > 0$ quando para todos $x, y \in \mathbb{R}^n$ tivermos*

$$f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y) - Lt(1-t)\|\, x - y \,\|^2, \tag{1.6}$$

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

Observação 1.3.2. *Note que se f é fortemente convexa então é estritamente convexa, em particular também é convexa.*

Exemplo 1.3.2. *A função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$ é convexa.*

Demonstração. De fato,

$$\begin{aligned} f(x + t(y-x)) &= x^2 + 2tx(y-x) + t^2(y-x)^2 \\ &\leqslant x^2 + 2tx(y-x) + t(y-x)^2 \\ &= x^2 + t(y-x)^2 \ \forall\ t \in [0, 1]. \end{aligned}$$

□

Exemplo 1.3.3. *([17]) Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função convexa e sejam $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ quaisquer. Então a função $\psi : \mathbb{R} \longrightarrow \mathbb{R}$, $\psi(\alpha) = f(x + \alpha d)$, é convexa.*

Demonstração. De fato, sejam $\alpha_1,\ \alpha_2 \in \mathbb{R}$ e $\lambda \in [0,1]$, temos que

$$\begin{aligned}
\psi(\lambda\alpha_1 + (1-\lambda)\alpha_2) &= f(x + (\lambda\alpha_1 + (1-\lambda)\alpha_2)d) \\
&= f(x + \lambda\alpha_1 d + (1-\lambda)\alpha_2 d) \\
&= f(\lambda(x + \alpha_1 d) + (1-\lambda)(x + \alpha_2 d)) \\
&\leqslant \lambda f(x + \alpha_1 d) + (1-\lambda) f(x + \alpha_2 d) \\
&= \lambda\psi(\alpha_1) + (1-\lambda)\psi(\alpha_2).
\end{aligned}$$

Portanto, ψ é convexa. □

Exemplo 1.3.4. *A função* $f : \mathbb{R}^n \to \mathbb{R}$ *dada por* $f(x) = \| x \|^2$ *é fortemente convexa com módulo* $L = 1$.

Demonstração. Dados $x, y \in \mathbb{R}^n$ e $t \in [0,1]$ temos que,

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

□

Proposição 1.3.1. *([4]) Seja* $h_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função fortemente convexa com constante* L_i *para* $i \in I = \{1, ..., p\}$. *Então* $h : \mathbb{R}^n \longrightarrow \mathbb{R}$ *definida por* $h(x) = \max_{i\in I} h(x)$ *é fortemente convexa com constante* $L := \min_{i \in L_i}$. *Em particular, se* h_i *é convexa para cada* $i \in I$, h *é convexa em* Ω.

Demonstração. Temos por hipótese que

$$h_i(\alpha x + (1-\alpha)y) \leqslant \alpha h_i(x) + (1-\alpha)h_i(y) - \frac{L_i}{2}\alpha(1-\alpha)\|x - y\|^2.$$

Daí, como

$$h(\alpha x + (1-\alpha)y) = \max_{i\in I} h_i(\alpha x + (1-\alpha)y), \tag{1.7}$$

suponha que o máximo seja atingido em algum $j \in I$. Por (1.5),

$$\begin{aligned} h((\alpha x + (1-\alpha)y) &= h_j((\alpha x + (1-\alpha)y) \\ &\leqslant \alpha h_j(x) + (1-\alpha)h_j(y) - \frac{L_j}{2}\alpha(1-\alpha)\|x-y\|^2 \\ &\leqslant \alpha h(x) + (1-\alpha)h(y) - \frac{L}{2}\alpha(1-\alpha)\|x-y\|^2 \end{aligned}$$

onde $L := \min_{i\in I} L_i$, i.e., $L \leqslant L_j$, $\forall j \in I$, então $-L_j \leqslant -L$. Concluimos então que h é fortemente convexa.

$\square$

Definição 1.3.4. *([17]) O epígrafo da função $f : D \to \mathbb{R}$ é o conjunto*

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

Teorema 1.3.1. *([17]) Seja $D \in \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \to \mathbb{R}$ é convexa em D se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Demonstração. Seja f uma função convexa e (x, c_1) e (y, c_2) pontos de E_f. Como pela definição 1.3.4., $f(x) \leqslant c_1$ e $f(y) \leqslant c_2$ temos pela convexidade de f que para todo $\alpha \in [0, 1]$ ocorre o seguinte:

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

logo,

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

Em outras palavras, temos que E_f é convexo.

Reciprocamente, suponha E_f convexo. Daí, dados $x, y \in D$, como $(x, f(x)), (y, f(y)) \in E_f$ temos que para todo $\alpha \in [0, 1]$

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

Logo, pela definição anterior temos que

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

ou seja, f é convexa. $\square$

Levando em conta que estamos interessados em resolver o Problema (1.1), a seguinte definição que envolve o comportamento da função objetivo se torna essencial.

Definição 1.3.5. *Dizemos que* $d \in \mathbb{R}^n$ *é uma direção de descida de* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *no ponto* $\bar{x} \in \mathbb{R}^n$*, se existir* $\epsilon > 0$ *tal que*

$$f(\bar{x} + td) < f(\bar{x}) \ \forall t \in (0, \epsilon].$$

Ou ainda,

$$\langle \text{grad } f(\bar{x}), d \rangle < 0.$$

Denotamos por $D_f(\bar{x})$ *o conjunto de todas as direções de descidas da função* f *no ponto* $\bar{x}$.

Em geral, $D_f(\bar{x})$ pode ser vazio. Por exemplo, quando $\bar{x}$ é um minimizador de f.

Definição 1.3.6. *([17]) Um conjunto* $K \subset \mathbb{R}^n$ *chama-se cone quando ele contem todos os múltiplos não-negativos de seus elementos, i.e.,*

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

Note que se K é um cone não-vazio, podemos concluir que $0 \in K$ e que de certa maneira, podemos interpretar um cone como um conjunto de direções. Daí, temos que quando $D_f(\bar{x})$ é não-vazio, ele não é um cone, pois $0 \notin D_f(\bar{x})$. No entanto, $D_f(\bar{x}) \cup \{0\}$ é um cone não-vazio, no caso em que $D_f(\bar{x}) \neq \emptyset$. Vejamos agora um exemplo de cone que será utilizado no decorrer deste trabalho.

Exemplo 1.3.5. *O ortante não-negativo* $\mathbb{R}^n_+$ *é um cone.*

Demonstração. De fato, tome $y = (y_1, ..., y_n) \in \mathbb{R}^n_+$, logo temos que $y_i \in \mathbb{R}_+$ para todo $i = 1, ..., n$. Daí, dado $t \in \mathbb{R}_+$ então

$$ty_i \in \mathbb{R}_+,$$

para todo $i = 1, ..., n$. Portanto, $ty \in \mathbb{R}^n_+$. □

Definição 1.3.7. *([17]) Sejam* $D \subset \mathbb{R}^n$ *um conjunto convexo e* $\bar{x} \in D$*. O cone normal no ponto* $\bar{x}$ *em relação ao conjunto D é dado por:*

$$\aleph_D(\bar{x}) = \{d \in \mathbb{R}^n \mid \langle d, x - \bar{x} \rangle \leqslant 0, \forall x \in D\}.$$

Outros dois tipos específicos de cone bastante usados na literatura, os quais iremos nos utilizar ao decorrer deste capítulo são os chamados cone tangente e cone de Bouligand, abordados abaixo.

Definição 1.3.8. *([17]) Dado* $\mathsf{D} \subset \mathbb{R}^n$ *e ponto* $\bar{x} \in \mathsf{D}$*, chama-se de cone tangente no ponto* $\bar{x}$ *o conjunto de todas as direções tangentes ao conjunto* D *no ponto* $\bar{x}$ *definido por*

$$\mathcal{T}_D(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{c} \forall \{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0^+, \\ \exists \{d^k\} \subset \mathbb{R}^n, \{d^k\} \to d, \ \textit{tal que} \\ \bar{x} + t_k d^k \in D \ \textit{para todo}\ k \in \mathbb{N} \end{array} \right\}.$$

Outra noção útil é a de cone (tangente) de Bouligand ou contigent cone, dado por:

$$\mathcal{B}_D(\bar{x}) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{c} \exists \{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0^+, \\ \exists \{d^k\} \subset \mathbb{R}^n, \{d^k\} \to d, \ \textit{tal que} \\ \bar{x} + t_k d^k \in D \ \textit{para todo}\ k \in \mathbb{N} \end{array} \right\}.$$

Note que pela Definição 1.3.8. temos que $\tau_D(\bar{x}) \subset B_D(\bar{x})$. Ou seja, a noção de cone (tangente) de Bouligand é mais geral que a de cone tangente.

Teorema 1.3.2. *([17]) Se* $f : \mathbb{R}^n \to \mathbb{R}$ *é convexa então todo minimizador local é global. Além disso, o conjunto dos minimizadores é convexo.*

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

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

Tomando t suficientemente pequeno temos que $(1-t)x^* + ty$ é arbitrariamente próximo de x^* com $f((1-t)x^*+ty) < f(x^*)$. Isso contradiz o fato de x^* ser mínimo local. Portanto, se f é convexa todo mínimo local é global.

Agora, seja $S \subset \mathbb{R}^n$ o conjunto dos minimizadores e $\bar{v} \in \mathbb{R}$ o valor ótimo do nosso Problema (1.1), isto é, $f(x) = \bar{v}, \forall x \in S$. Tomando $x, y \in S$ e $\alpha \in [0,1]$, pela convexidade de f obtemos

$$\begin{aligned} f(\alpha x + (1-t)y) &\leqslant \alpha f(x) + (1-\alpha)f(y) \\ &= \alpha\bar{v} + (1-\alpha)\bar{v} \\ &= \bar{v}. \end{aligned}$$

Como $\bar{\nu}$ é valor ótimo temos que

$$f(\alpha x + (1-t)y) = \bar{\nu}.$$

Em outras palavras $\alpha x + (1-t)y \in S$. □

Teorema 1.3.3. *([17]) Se* $f : \mathbb{R}^n \to \mathbb{R}$ *é estritamente convexa. Então não pode haver mais de um minimizador.*

Demonstração. Seja $S \subset \mathbb{R}^n$ o conjunto dos minimizadores de f. Suponha por contradição que exista $x \in S$ e $\bar{x} \in S$, com $x \neq \bar{x}$. Daí, seja $\alpha \in (0,1)$ pela fato de f ser estritamente convexa temos que

$$\begin{aligned} f(\alpha x + (1-t)\bar{x}) &< \alpha f(x) + (1-\alpha)f(\bar{x}) \\ &= \alpha\bar{\nu} + (1-\alpha)\bar{\nu} \\ &= \bar{\nu}. \end{aligned}$$

Como S é convexo (pelo teorema anterior) temos que $\alpha x + (1 - \alpha\bar{x})$ é uma solução para nosso problema com valor menor do que $\bar{\nu}$. Mas isso gera uma contradição. Portanto, o minimizador deve ser único. □

O próximo resultado nos diz que toda função convexa é contínua em qualquer subconjunto aberto do seu domínio. Além disso, ainda podemos garantir que tal função será também localmente Lipschitz-contínua no interior do seu domínio como pode ser visto em [17].

Teorema 1.3.4. *(**Continuidade de funções convexas**) Sejam* $\Omega \subset \mathbb{R}^n$ *um conjunto convexo e aberto e* $f : \Omega \longrightarrow \mathbb{R}$ *função convexa em* Ω*. Então* f *é localmente Lipschitz-contínua em* Ω*. Em particular, segue que* f *é contínua em* Ω*.*

No caso em que a função estudada é diferenciável, a convexidade admite algumas caracterizações que são utéis para se saber se uma função é convexa ou não. Tais caracterizações também serão essenciais no prosseguimento deste trabalho.

Teorema 1.3.5. *([17]) Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função diferenciável. Então as seguintes propriedades são equivalentes:*

(i) A função f *é convexa;*

(ii) Para todo $x, y \in \mathbb{R}^n$,

$$f(y) \geqslant f(x) + \langle f'(x),\ y - x\rangle.$$

Quando f *é duas vezes diferenciável, as propriedades acima também são equivalentes a*

(iii) A matriz hessiana de f *é semi-definida positiva em todo* $x \in \mathbb{R}^n$*,*

$$\langle H(x)d, d\rangle \geqslant 0, \quad \forall\ d \in \mathbb{R}^n.$$

Demonstração. Suponha f convexa, então para $x, y \in \mathbb{R}^n$, $\alpha \in (0,1]$ quaisquer, definindo $d = y - x$, temos que

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

donde

$$\alpha(f(y) - f(x)) \geqslant f(x+\alpha d) - f(x).$$

Dividindo ambos os membros por $\alpha > 0$,e passando o limite com $\alpha \to 0^+$, obtemos

$$\begin{aligned} f(y) - f(x) &\geqslant \lim_{\alpha\to 0^+} \frac{f(x+\alpha d) - f(x)}{\alpha} \\ &= \langle f'(x), d\rangle = \langle f'(x), y - x\rangle. \end{aligned}$$

Logo,

$$f(y) \geqslant f(x) + \langle f'(x), y - x\rangle.$$

Reciprocamente, fazendo novamente $d = y - x$ e usando $b)$ para os pontos x e $x + \alpha d$; y e $x + \alpha d$ temos que

$$f(x) \geqslant f(x+\alpha d) - \alpha\langle f'(x+\alpha d), d\rangle$$

e

$$f(y) \geqslant f(x+\alpha d) - (1-\alpha)\langle f'(x+\alpha d), d\rangle.$$

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

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

ou seja, f é convexa.

No caso em que f é duas vezes diferenciável é suficiente mostrar que $(ii) \Leftrightarrow (iii)$. De fato, fixemos $x, d \in \mathbb{R}^n$ quaisquer, daí por (ii),

$$f(x+\alpha d) - f(x) \geqslant \alpha\langle f'(x), d\rangle.$$

Usando ainda o fato de que f é duas vezes diferenciável,

$$\begin{aligned} 0 &\geqslant f(x+\alpha d) - f(x) - \alpha\langle f'(x), d\rangle \\ &= \frac{\alpha^2}{2}\langle f''(x)d, d\rangle + \theta(\alpha^2). \end{aligned}$$

Dividindo por $\alpha^2 > 0$ e fazendo $\alpha \to 0^+$, obtemos (iii). Supondo agora que ocorre (iii), sejam $x, y \in \mathbb{R}^n$, pelo Teorema do Valor Médio existe $\alpha \in (0,1)$ tal que

$$f(y) - f(x) - \langle f'(x), y-x\rangle = \frac{1}{2}\langle f''(x+\alpha(y-x))(y-x), y-x\rangle \geqslant 0.$$

Logo, segue (ii). $\square$

Exemplo 1.3.6. *Considere a seguinte função,* $f(x_1, x_2) = 2x_1^2 + 3x_1x_2 + 5x_2^2 - 2x_1 + 6x_2$. *Note que*

$$y^T \bigtriangledown^2 f(x)y = (y_1 \;\; y_2)\begin{pmatrix} 4 & 3 \\ 3 & 10 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 4y_1^2 + 6y_1y_2 + 10y_2^2 \geqslant 0.$$

Então, pelo resultado anterior, a função f *é convexa.*

Teorema 1.3.6. *([17])* ***(Derivada direcional de uma função convexa)*** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *convexa. Então para todo* $x \in \mathbb{R}^n$, f *é diferenciável em cada direção* $d \in \mathbb{R}^n$. *Além disso,*

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

Demonstração. Fixamos $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ arbitrários. Defina,

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

Daí, precisamos mostrar a existência do seguinte limite

$$\begin{aligned} f(x; d) &= \lim_{\alpha \to 0_+} \frac{f(x+\alpha d) - f(x)}{\alpha} \\ &= \lim_{\alpha \to 0_+} \frac{\psi(\alpha) - \psi(0)}{\alpha}. \end{aligned} \tag{1.8}$$

Como f é localmente lipschitz, pelo Teorema 1.3.4., existem $L > 0$ e $\bar{\alpha} > 0$ tais que

$$|f(x+\alpha d) - f(x)| \leqslant L\alpha||d|| \;\; \forall \alpha \in [0, \bar{\alpha}].$$

Portanto,

$$\frac{|\psi(\alpha) - \psi(0)|}{\alpha} \leqslant L||d||,$$

ou seja, a função $\alpha^{-1}(\psi(\alpha)-\psi(0))$ é limitada no conjunto $[0,\bar{\alpha}]$. Iremos mostrar agora que essa mesma função também é não-decrescente em $[0,\bar{\alpha}]$. De fato, sejam $\beta \geqslant \alpha > 0$, então existe $t \in (0,1]$ tal que $\alpha = t\beta = (1-t)0 + t\beta$. Daí, devido a convexidade de ψ, Exemplo 1.3.3., temos que

$$\psi(\alpha) \leqslant (1-t)\psi(0) + t\psi(\beta).$$

Portanto,

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

Isso nos diz que a função $\alpha^{-1}(\psi(\alpha)-\psi(0))$ é não-decrescente quando $\alpha \to 0_+$. Como já vimos que ela é limitada, em particular inferiormente, segue que o limite (1.7) existe. Além disso,

$$\alpha^{-1}(\psi(\alpha)-\psi(0)) \geqslant f^{'}(x;d),$$

ou seja,

$$f(x+\alpha d) \geqslant f(x) + \alpha f^{'}(x;d).$$

□

Definição 1.3.9. *([17]) Seja* $f:\mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Dizemos que* $y \in \mathbb{R}^n$ *é um subgradiente de* f *no ponto* $x \in \mathbb{R}^n$ *se*

$$f(z) \geqslant f(x) + \langle y, z-x\rangle \quad \forall z \in \mathbb{R}^n. \tag{1.9}$$

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

Considere agora $z = x + \alpha d$, onde $d \in \mathbb{R}^n$, $\alpha > 0$, a definição acima de subgradiente pode ser escrita da seguinte forma,

$$y \in \partial f(x) \Leftrightarrow \frac{f(x+\alpha d)-f(x)}{\alpha} \geqslant \langle y, d\rangle,\ \forall d \in \mathbb{R}^n, \forall \alpha > 0. \tag{1.10}$$

Daí, pelo que foi visto na demonstração do Teorema 1.3.6, $\alpha^{-1}(f(x+\alpha d)-f(x))$ é uma função não-decrescente; além disso, ela possui um limite quando fazemos $\alpha \to 0_+$. Logo, (1.9) é equivalente à:

$$f^{'}(x;d) = \lim_{\alpha\to 0_+}\frac{f(x+\alpha d)-f(x)}{\alpha} \geqslant \langle y, d\rangle, \forall d \in \mathbb{R}^n.$$

Em resumo temos que a definição de subgradiente pode também ser vista de uma forma mais prática, que é a seguinte:

$$\partial f(x) = \{y \in \mathbb{R}^n \mid f^{'}(x; d) \geqslant \langle y, d\rangle, \ \forall d \in \mathbb{R}^n\}. \tag{1.11}$$

Proposição 1.3.2. *([4]) Seja* $\Omega \subset \mathbb{R}^n$ *não-vazio e convexo. Dada* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$*, temos que* f *é fortemente convexa com constante* $L > 0$ *se, e somente se,*

$$\langle u - v, x - y\rangle \geqslant L||x - y||^2, \quad u \in \partial f(x), \quad v \in \partial f(y) \tag{1.12}$$

para quaisquer $x, y \in \Omega$*.*

Demonstração. Sabendo que f é fortemente convexa, i.e.,

$$f(\alpha(x - y) + y) - f(y) \leqslant \alpha f(x) - \alpha f(y) - L\alpha(1 - \alpha)||x - y||^2,$$

dividindo ambos os lados por $\alpha(1 - \alpha)$ obtemos,

$$\frac{f(\alpha(x - y) + y) - f(y)}{\alpha(1 - \alpha)} \leqslant \frac{f(x)}{(1 - \alpha)} - \frac{f(y)}{(1 - \alpha)} - L||x - y||^2.$$

Agora, fazendo $\alpha \to 0$:

$$f^{'}(y, d) \leqslant f(x) - f(y) - L||x - y||^2; \quad d = x - y.$$

Portanto,

$$L||x - y||^2 \leqslant -f^{'}(y, d) + f(x) - f(y).$$

Trocando x por y e vice-versa, conseguimos

$$L||x - y||^2 \leqslant -f^{'}(x, d^{'}) - f(x) + f(y); \quad d^{'} = y - x.$$

Somando as duas desigualdades acima temos que

$$L||x - y||^2 \leqslant -f^{'}(x.d) - f^{'}(y, d). \tag{1.13}$$

Por outro lado, como

$$\begin{cases} f(z) & \geqslant & f(x) + \langle u, z - x\rangle \\ f(z) & \geqslant & f(y) + \langle v, z - y\rangle \end{cases} \tag{1.14}$$

$\forall z \in \mathbb{R}^n$, logo tomando $z = \alpha y + (1 - \alpha)x = x + \alpha(y - x)$ na primeira desigualdade de (1.14), obtemos

$$f(x + \alpha d^{'}) - f(x) \geqslant \langle u, x + \alpha d^{'} - x\rangle.$$

Dividindo por α e fazendo $\alpha \to 0$ obtemos que

$$f'(x, d') \geqslant \langle u, y - x \rangle. \tag{1.15}$$

De forma análoga fazendo $z = \alpha x + (1-\alpha)y = y + \alpha(x - y)$

$$f'(y, d) \geqslant \langle v, x - y \rangle. \tag{1.16}$$

Logo, multiplicando (1.15) e (1.16) por (-1) e somando

$$-f'(x, d') - f'(y, d) \leqslant \langle u - v, x - y \rangle. \tag{1.17}$$

Então combinando (1.17) com (1.13) segue que

$$L\|x - y\|^2 \leqslant \langle u - v, x - y \rangle.$$

Reciprocamente, como $u \in \partial f(x),\ v \in \partial f(y)$ então

$$\begin{cases} f(z) & \geqslant & f(x) + \langle u, z - x \rangle \\ f(z) & \geqslant & f(y) + \langle v, z - y \rangle \end{cases} \tag{1.18}$$

Fazendo $z = x = \alpha x + (1-\alpha)y$ na primeira desigualdade de (1.18) e $z = y = \alpha x + (1-\alpha)y$ na segunda desigualdade,

$$\begin{cases} f(x) & \geqslant & f(\alpha x + (1-\alpha)y) + \langle u, x - \alpha x - (1-\alpha)y \rangle \\ f(y) & \geqslant & f(\alpha x + (1-\alpha)y) + \langle -v, -y + \alpha x + (1-\alpha)y \rangle \end{cases} \tag{1.19}$$

Daí, multiplicando a primeira desigualdade de (1.19) por α e a segunda por $(1-\alpha)$ e somando obtemos que

$$\alpha f(x) + (1-\alpha)f(y) \geqslant f(\alpha x + (1-\alpha)y) + \alpha(1-\alpha)\langle u - v, x - y \rangle,$$

logo,

$$f(\alpha x + (1-\alpha)y) \leqslant \alpha f(x) + (1-\alpha)f(y) - L\alpha(1-\alpha)\|x - y\|^2.$$

Ou seja, valendo (1.12) implica que f é fortemente convexa. □

Os próximos dois resultados nos dizem que dois conjuntos convexos sem pontos em comum são separáveis. Tais resultados serão utilizados em algumas demonstrações no decorrer deste capítulo, mas omitiremos suas demonstrações, que poderão ser encontrada em [17].

Teorema 1.3.7. ***(Separação)*** *Sejam* $D_1 \subset \mathbb{R}^n$ *e* $D_2 \subset \mathbb{R}^n$ *conjuntos convexos não-vazios tais que* $D_1 \cap D_2 = \emptyset$. *Então existem* $a \in \mathbb{R}^n \setminus \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

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

Teorema 1.3.8. ***(Separação Estrita)*** *Sejam* $D_1 \subset \mathbb{R}^n$ *e* $D_2 \subset \mathbb{R}^n$ *conjuntos convexos, fechados e não-vazios. Suponhamos que um deles também seja limitado (logo, compacto). Então* $D_1 \cap D_2 = \emptyset$ *se, e somente se,existem* $a \in \mathbb{R}^n \setminus \{0\}$ *e* $c \in \mathbb{R}$ *tais que*

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

Teorema 1.3.9. *([17]) Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então para todo* $x \in \mathbb{R}^n$, *o conjunto* $\partial f(x)$ *é convexo, compacto e não-vazio. Além disso, para todo* $d \in \mathbb{R}^n$, *tem-se*

$$f'(x; d) = max_{y \in \partial f(x)} \langle y, d \rangle. \tag{1.20}$$

Demonstração. Por (1.11), temos que

$$\begin{aligned} \partial f(x) &= \{y \in \mathbb{R}^n \mid \langle y, d \rangle \leqslant f'(x, d),\ \forall d \in \mathbb{R}^n\} \\ &= \cap_{d \in \mathbb{R}^n} \{y \in \mathbb{R}^n \mid \langle y, d \rangle \leqslant f'(x, d)\}. \end{aligned}$$

Como a intersecção de semi-espaços fechados é convexa e fechada, obtemos que $\partial f(x)$ é convexo e fechado. Resta mostrarmos que $\partial f(x)$ é limitado. Até então, ainda não sabemos que $\partial f(x) \neq \emptyset$, daí caso $\partial f(x) = \emptyset$, o subdiferencial é limitado. Suponhamos que exista uma sequência $\{y^k\} \subset \partial f(x)$ tal que $||y^k|| \to \infty$. Daí, considere para todo k, $d^k = \dfrac{y^k}{||y^k||}$, onde podemos supor, a menos de subsequência que $\{d^k\} \to d$, pois d^k é limitada. Então por (1.11),

$$||y^k|| = \langle y^k, d^k \rangle \leqslant f'(x; d^k).$$

Portanto,

$$\lim_{k \to \infty} \sup ||y^k|| \leqslant \lim_{k \to \infty} \sup\ f'(x; d^k) \leqslant f'(x; d) < \infty,$$

onde a segunda desigualdade segue de [17, pag 173]. Isto contradiz o fato de $||y^k|| \to \infty$. Portanto, segue que $\partial f(x)$ é limitado.

Por fim provemos que $\partial f(x) \neq \emptyset$. Para isso, mostremos (1.20) e ao mesmo tempo verifiquemos que $\partial f(x) \neq \emptyset$. De fato, fixemos $d \in \mathbb{R}^n$ arbitrário, então pelos cálculos acima

$$f'(x; d) \geqslant \max_{y \in \partial f(x)} \langle y, d \rangle,$$

onde a existência do máximo acima segue pelo fato de $\partial f(x)$ ser compacto, junto com o Teorema 1.2.1.. Defina os seguintes dois conjuntos:

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

e

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

Note que ambos os conjuntos são convexos, onde D_1 é o epígrafo de f sem a sua fronteira, logo é convexo, pois f é convexa.

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

$$f(x + \alpha d) < f(x) + \alpha f'(x; d),$$

para algum $\alpha \in \mathbb{R}_+$, mas isso gera uma contradição com o Teorema 1.3.6. Portanto, temos que $D_1 \cap D_2 = \emptyset$. Então, pelo Teorema 1.3.7., existe $(u, \beta) \in \mathbb{R}^n \times \mathbb{R} \setminus \{0\}$ tal que

$$\langle u, z\rangle + \beta c \leqslant \langle u, x + \alpha d\rangle + \beta(f(x) + \alpha f'(x; d)), \tag{1.21}$$

$\forall z \in \mathbb{R}^n,\ \forall \alpha \in \mathbb{R}_+,\ \forall c > f(x)$. Se $\beta = 0$, então

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

mas isso só pode acontecer quando $u = 0$. Como $(u, \beta) \neq 0$, logo $\beta \neq 0$.

Suponha $\beta > 0$, então escolhendo $z = x$ e $\alpha = 0$ em (1.21), temos que $\beta c \leqslant \beta f(x)$ para todos os c tais que $c > f(x)$, o que gera uma contradição. Portanto, obtemos que $\beta < 0$. Dividindo ambos os lados em (1.21) por $\beta < 0$, temos que

$$c + \langle \frac{u}{\beta}, z - x\rangle \geqslant f(x) + \alpha f'(x; d) + \alpha\langle \frac{u}{\beta}, d\rangle, \tag{1.22}$$

$\forall z \in \mathbb{R}^n,\ \forall \alpha \in \mathbb{R}_+,\ \forall c > f(z)$. Tomando os limites quando $\alpha \to 0^+$ e $c \to f(z)^+$, obtemos que

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

ou seja, $\dfrac{-u}{\beta} \in \partial f(x)$. Em outras palavras, segue que $\partial f(x) \neq \emptyset$.

Tomando ainda $\alpha = 1$ e $z = x$ em (1.22) e passando o limite com $c \to f(z)^+$, obtemos

$$0 \geqslant f'(x; d) + \langle \frac{u}{\beta}, d\rangle.$$

Logo,

$$-\langle\frac{u}{\beta},d\rangle \geqslant f'(x;d),\ \frac{-u}{\beta}\in\partial f(x).$$

Daí, como já sabemos que $f'(x;d)\geqslant \max_{y\in\partial f(x)}\langle y,d\rangle$, então

$$f'(x;d)=\max_{y\in\partial f(x)}\langle y,d\rangle.$$

Segue o resultado. □

Proposição 1.3.3. *([17]) Uma função convexa* $f:\mathbb{R}^n\longrightarrow\mathbb{R}$ *é diferenciável no ponto* $x\in\mathbb{R}^n$ *se, e somente se, o conjunto* $\partial f(x)$ *contém um elemento só. Neste caso,* $\partial f(x)=\{f^{'}(x)\}$ *(onde* $f^{'}(x)$ *denota o gradiente de* f *em* x*).*

Demonstração. Considere f diferenciável no ponto $x\in\mathbb{R}^n$. Então, pelo Teorema 1.3.5.

$$f(z)\geqslant f(x)+\langle f'(z),z-x\rangle,\forall z\in\mathbb{R}^n,$$

ou seja, $f'(x)\in\partial f(x)$. Seja agora $y\in\partial f(x)$, então para todo $d\in\mathbb{R}^n$, fazendo $z=x+d$ na Definição 1.3.8. temos

$$\begin{aligned} f(x)+\langle y,d\rangle &\leqslant f(x+d)\\ &= f(x)+\langle f'(x),d\rangle+\theta(\|d\|). \end{aligned}$$

Portanto,

$$\langle y-f'(x),d\rangle\leqslant\theta(\|d\|).$$

Tomando,

$$d^k=\frac{y-f'(x)}{k\|y-f'(x)\|},$$

temos que $d^k\to 0$ e

$$\frac{\|y-f'(x)\|}{k}\leqslant\theta(\frac{1}{k}),$$

mas isso implica que $y-f'(x)=0$. Concluímos então que $\partial f(x)=\{f'(x)\}$.

Seja $\partial f(x)=\{y\}$, pelo Teorema 1.3.7.,

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

Escolhendo como d os elementos da base canônica do $\mathbb{R}^n$, temos que $y_i=\dfrac{\partial f(x)}{\partial x_i}$, $i=1,...,n$. Temos que $f'(x;d)$ é uma função linear em d de forma $\langle f'(x),d\rangle$, o que implica a diferenciabilidade de f em x.

□

A importância do estudo do subdiferencial e suas propriedades na teoria de otimização pode ser vista no seguinte resultado, onde iremos nos utilizar da Definição 1.3.7.. Como foi visto anteriormente temos que sempre vale a inclusão $\tau_D(\bar{x}) \subset B_D(\bar{x})$. Mas ainda, se o conjunto D em questão for convexo temos por [17] que vale a igualdade $\tau_D(\bar{x}) = B_D(\bar{x})$. Usando este fato, vejamos o próximo resultado.

Teorema 1.3.10. *([17]) (**Condição de otimalidade para minimização de uma função convexa num conjunto convexo**) Sejam $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função convexa e $D \subset \mathbb{R}^n$ um conjunto convexo. Então $\bar{x} \in \mathbb{R}^n$ é um minimizador de f em D se, e somente se,*

$$\exists y \in \partial f(\bar{x}) \text{ tal que } \langle y, x - \bar{x} \rangle \geqslant 0 \quad \forall x \in D, \tag{1.23}$$

ou equivalentemente,

$$0 \in \partial f(\bar{x}) + \aleph_D(\bar{x}). \tag{1.24}$$

Em particular, $\bar{x}$ é minimizador de f em $\mathbb{R}^n$ se, e somente se, $0 \in \partial f(\bar{x})$.

Demonstração. Suponha inicialmente que $\langle y, x - \bar{x} \rangle \geqslant 0$ para todo $x \in D$, então temos que $-y \in \aleph_D(\bar{x})$. Pela definição de subgradiente, segue que

$$f(x) \geqslant f(\bar{x}) + \langle y, x - \bar{x} \rangle \geqslant f(\bar{x}), \ \forall x \in D,$$

isto é, $\bar{x}$ é minimizador de f em D.

Reciprocamente, suponhamos que $\bar{x}$ é um minimizador de f em D. Escolhemos qualquer $d \in \tau_D(\bar{x}) = B_D(\bar{x}), \ d \neq 0$. Logo, existem sequências $\{t_k\} \subset \mathbb{R}_+ \setminus \{0\}$ e $\{d^k\} \subset \mathbb{R}^n$ tais que $\{d^k\} \to d, \ \{t_k\} \to 0^+, \ x + t_k d^k \in D$ para todo k. Daí, para todo k temos que

$$0 \leqslant f(\bar{x} + t_k d^k) - f(\bar{x}) = t_k f'(\bar{x}, d^k) + \theta(t_k).$$

Dividindo ambos os lados por $t_k > 0$ e passando o limite com $k \to +\infty$ concluímos que

$$f'(\bar{x}, d) \geqslant 0, \ \forall d \in \tau_D(\bar{x}). \tag{1.25}$$

Suponhamos que não ocorra (1.22), logo não existe $y \in \mathbb{R}^n$ tal que $-y \in \partial f(\bar{x})$ e $y \in \aleph_D(\bar{x})$. Logo,

$$(-\partial f(\bar{x})) \cap \aleph_D(\bar{x}) = \emptyset.$$

Como $\partial f(\bar{x})$ é convexo, compacto e não-vazio e $\aleph_d(\bar{x})$ é convexo, fechado e não vazio pelo Teorema 1.3.8., existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que

$$-\langle a, y \rangle > c > \langle a, d \rangle, \ \forall y \in \partial f(\bar{x}), \ \forall d \in \aleph_D(\bar{x}). \tag{1.26}$$

Como $0 \in \aleph_D(\bar{x})$, temos que $c > 0$. Portanto, $\langle a, y \rangle < c < 0$ para todo $y \in \partial f(\bar{x})$. Daí, pelo Teorema 1.3.7., obtemos que

$$f'(\bar{x}, a) = \max_{y \in \partial f(\bar{x})} \langle a, y \rangle < 0. \tag{1.27}$$

Suponhamos que $\langle a, d \rangle > 0$ para algum $d \in \aleph_D(\bar{x})$. Tomando $td \in \aleph_D(\bar{x})$, $t > 0$ e fazendo $t \to +\infty$, temos uma contradição em (1.24), pois $c \in \mathbb{R}$ é fixo. Então, $\langle a, d \rangle \leqslant 0$ para todo $d \in \aleph_D(\bar{x})$. Portanto,

$$a \in (\aleph_D(\bar{x}))^* = ((\tau_D(\bar{x}))^*)^* = \tau_D(\bar{x}), \tag{1.28}$$

onde as desigualdades seguem de ([17], pag 118).
Daí, temos que (1.27) e (1.28) contradizem (1.25). Por fim, se $\bar{x}$ é um minimizador de f em $\mathbb{R}^n$ então

$$f(\bar{x}) \leqslant f(x), \ \forall x \in \mathbb{R}^n.$$

Logo,

$$f(x) \geqslant f(\bar{x}) + \langle 0, x - \bar{x} \rangle,$$

isto é, $0 \in \partial f(\bar{x})$. Reciprocamente, se $0 \in \partial f(\bar{x})$ segue que $\bar{x}$ é minimizador de f em $\mathbb{R}^n$. □

Ainda no que diz respeito ao subdiferencial de uma função convexa, e usando os resultados abordados acima, nosso objetivo agora será provar o seguinte resultado:

Proposição 1.3.4. *([25]) Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função convexa. Então vale que*

$$\partial(\lambda f)(x) = \lambda \partial f(x), \quad \forall x, \ \forall \lambda \geqslant 0. \tag{1.29}$$

Demonstração. É imediato o caso $\lambda = 0$. Agora, seja $\lambda > 0$ e $\alpha \in \partial(\lambda f)(x)$, i.e.,

$$\lambda f(z) \geqslant \lambda f(x) + \langle \alpha, z - x \rangle, \ \forall z \in \mathbb{R}^n. \tag{1.30}$$

Nosso objetivo inicialmente será escrever $\alpha = \lambda y$; $y \in \partial f(x)$. Daí, como estamos supondo $\lambda > 0$ por (1.19) temos que

$$f(z) \geqslant f(x) + \langle \frac{\alpha}{\lambda}, z - x \rangle, \ \forall z \in \mathbb{R}^n,$$

isto implica que $\frac{\alpha}{\lambda} \in \partial f(x)$. Por outro lado, veja que $\alpha = \lambda \frac{\alpha}{\lambda}$, logo $\alpha \in \lambda \partial f(x)$.
Reciprocamente, dado $\beta \in \lambda \partial f(x)$ então $\beta = \lambda y$ com $y \in \partial f(x)$, i.e.,

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

Multiplicando por $\lambda > 0$ obtemos que,

$$\lambda f(z) \geqslant \lambda f(x) + \langle \beta, z - x \rangle,$$

para todo $z \in \mathbb{R}^n$. Portanto, $\beta \in \partial(\lambda f)(x)$.

$\square$

Com o objetivo de facilitar algumas contas feitas neste trabalho mais adiante, veremos a seguinte propriedade em relação ao subdiferencial de funções convexas não-diferenciáveis.

$$\partial(\lambda_1 f_1 + \lambda_2 f_2)(x) \subset \lambda_1 \partial f_1(x) + \lambda_2 \partial f_2(x), \ \forall x \in \Omega, \tag{1.31}$$

onde $\lambda \geqslant 0$ e $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $(i = 1, 2)$ é uma função convexa em $\Omega \subset \mathbb{R}^n$ com Ω não-vazio e convexo. Note que devido a Proposição 1.3.4. basta mostrarmos que $\partial(f_1 + f_2)(x) \subset \partial f_1(x) + \partial f_2(x)$, $\forall x \in \Omega$. De fato, suponha por contradição que não vale a inclusão dada, i.e., existe $y \in \partial f(x)$, com $f(x) = f_1(x) + f_2(x)$ tal que $y \notin \partial f_1(x) + \partial f_2(x)$. Os conjuntos $\partial f_1(x)$ e $\partial f_2(x)$ são convexos, compactos e não-vazios (Teorema 1.3.6.). Portanto, $\partial f_1(x) + \partial f_2(x)$ é convexo, compacto e não-vazio. Como $y \notin \partial f_1(x) + \partial f_2(x)$, pelo Teorema 1.3.8., existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que

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

Daí,

$$\begin{aligned}
\langle y, a \rangle &< \sup_{y^i \in \partial f_i(x), i=1,2} \langle y^1 + y^2, a \rangle \\
&= \sup_{y^1 \in \partial f_1(x)} \langle y^1, a \rangle + \sup_{y^2 \in \partial f_2(x)} \langle y^2, a \rangle \\
&= max_{y^1 \in \partial f_1(x)} \langle y^1, a \rangle + smax_{y^2 \in \partial f_2(x)} \langle y^2, a \rangle \\
&= f_1^{'}(x; a) + f_2^{'}(x; a) \\
&= f^{'}(x; a).
\end{aligned}$$

Porém, esta desigualdade gera uma contradição pelo fato de $y \in \partial f(x)$ (vide (1.8)). Vejamos agora um resultado de uma função convexa não-diferenciável, que é de grande importância para o desenvolvimento deste trabalho. Tal resultado é fornecido pelo máximo de funções convexas (mesmo no caso em que elas sejam diferenciáveis).

Teorema 1.3.11. *([17]) **(Derivada direcional e o subdiferencial do máximo de funções convexas)***

Considere Z *um conjunto compacto qualquer. Suponhamos que* $\psi : \mathbb{R}^n \times Z \longrightarrow \mathbb{R}$ *seja uma função contínua tal que* $\psi(., z) : \mathbb{R}^n \longrightarrow \mathbb{R}$ *seja convexa para todo* $z \in Z$*. Seja*

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

Então:

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

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

onde

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

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

$$grad \ f(x) = grad \ \psi_x(x, \bar{z}).$$

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

$$\partial f(x) = conv \ \{grad \ \psi_x(x, \bar{z}) \mid \bar{z} \in Z(x)\}, \ x \in \mathbb{R}^n.$$

Demonstração. Por Weierstrass, temos que $Z(x) \neq \emptyset$ e $f(x)$ é finito para todo $x \in \mathbb{R}^n$. Daí, como Z é compacto temos que f é convexa, (ver [17],pag 142). Pela definição de f, $\forall \bar{z} \in Z(x)$

$$\begin{cases} f(x) & = & \psi(x, \bar{z}) \\ f(x + \alpha d) & \geqslant & \psi(x + \alpha d, \bar{z}) \end{cases}$$

Note que a segunda equação acima seria uma igualdade se $\bar{z} \in Z(x + \alpha d)$, mas $\bar{z} \in Z(x)$. Então, temos que

$$\frac{f(x + \alpha d) - f(x)}{\alpha} \geqslant \frac{\psi(x + \alpha d, \bar{z}) - \psi(x, \bar{z})}{\alpha}, \ \forall d \in \mathbb{R}^n, \ \forall \alpha > 0.$$

Fazendo $\alpha \to 0^+$, temos que

$$f'(x, d) \geqslant \psi'(x, \bar{z}, d), \ \forall \bar{z} \in Z(x), \ \forall d \in \mathbb{R}^n.$$

Portanto,

$$f'(x,d) \geqslant \sup_{\bar{z}\in Z(x)} \ \psi'(x,\bar{z},d), \ \forall d \in \mathbb{R}^n. \tag{1.32}$$

Agora, seja $d \in \mathbb{R}^n$ qualquer e $\{\alpha_k\} \to 0^+$. Para todo $k \in \mathbb{N}$ definimos $x^k = x + \alpha_k d$ e $\bar{z}^k \in Z(x^k)$. Como $\{\bar{z}^k\} \subset Z(x^k)$ e Z é compacto, podemos supor que $\{\bar{z}^k\} \to z \in Z$. Daí,

$$\psi(x^k,\bar{z}^k) = \max_{z\in Z} \psi(x^k, z) \geqslant \psi(x^k, z), \ \forall z \in Z.$$

Portanto, pelo fato de ψ ser continua, $\psi(x,\bar{z}) \geqslant \psi(x,z), \ \forall z \in Z$. Logo,

$$\bar{z} \in Z(x). \tag{1.33}$$

Pelo Teorema 1.3.6. e usando a definição de f e (1.27) obtemos que

$$\begin{aligned} f'(x,d) &\leqslant \frac{f(x+\alpha_k d) - f(x)}{\alpha_k} \\ &= \frac{\psi(x+\alpha_k d, \bar{z}^k) - \psi(x,\bar{z})}{\alpha_k} \\ &= \frac{\psi(x+\alpha_k, \bar{z}^k) - \psi(x,\bar{z}^k)}{\alpha_k}. \end{aligned}$$

Fazendo $k \to +\infty$,

$$f'(x,d) \leqslant \psi'(x,\bar{z},d),$$

ou seja,

$$f'(x,d) \leqslant \sup_{\bar{z}\in Z(x)} \psi'(x,\bar{z},d). \tag{1.34}$$

Portanto de (1.26) e (1.28), segue o item a).

Agora o item b). Para todo $u \in \mathbb{R}^n$ e $\bar{z} \in Z(x)$ temos

$$\begin{aligned} f(u) &= \max_{z\in Z} \psi(u,z) \geqslant \psi(u,\bar{z}) \\ &\geqslant \psi(x,\bar{z}) + \langle \psi'_x(x,\bar{z}), u - x\rangle \\ &= f(x) + \langle \psi'_x(x,\bar{z}), u - x\rangle. \end{aligned}$$

Logo, $\psi'_x(x,\bar{z}) \in \partial f(x), \ \forall \bar{z} \in Z(x)$. Como $\partial f(x)$ é convexo temos que

$$conv\{\psi'_x(x,\bar{z}) \ / \ \bar{z} \in Z(x)\} \subset conv\partial f(x) = \partial f(x),$$

isto é, $conv\{\psi'_x(x,\bar{z}) \ / \ \bar{z} \in Z(x)\} \subset \partial f(x)$. Note agora, que $Z(x) = Z \cap \{z \ / \ \psi(x,z) = f(x)\}$. Como Z é compacto e o segundo é fechado, pela continuidade de ψ, temos que $Z(x)$

é compacto. Agora, pela continuidade de $\psi'_z(.,\bar{z})$ temos que $\{\psi'_x(x,\bar{z})\ /\ \bar{z} \in Z(x)\}$ é compacto, então por ([17], pag 92), obtemos que $\mathbf{conv}\ \{\psi'_x(x,\bar{z})\ /\ \bar{z} \in Z(x)\}$ é compacto. Suponha então que exista $y \in \partial f(x) \setminus \mathbf{conv}\ \{\psi'_x(x,\bar{z})\ /\ \bar{z} \in Z(x)\}$. Então pelo Teorema 1.3.8., existem $\alpha \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tal que

$$\langle a, y\rangle > c > \langle a, \psi'_x(x,\bar{z})\rangle,\ \forall \bar{z} \in Z(x).$$

Logo,

$$\begin{aligned}\langle y, a\rangle &> \max_{z\in Z(x)}\langle \psi'_x(x,\bar{z}), a\rangle\\ &= \max_{\bar{z}\in Z(x)} \psi'(x,\bar{z}), a)\\ &= f'(x,a),\end{aligned}$$

onde acima usamos o item a) e o fato de $\psi(.,\bar{z})$ ser diferenciável. Mas isso gera uma contradição pelo fato de $y \in \partial f(x)$ e pelo Teorema 1.3.7.. Portanto,

$$\partial f(x) \subset \mathbf{conv}\{\psi'_x(x,\bar{z})\ /\ \bar{z} \in Z(x)\}.$$

Logo, segue o resultado. □

Como aplicação imediata do teorema anterior apresentamos o seguinte exemplo:

Exemplo 1.3.7. *([17]) Sejam $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ funções convexas diferenciáveis, $i = 1, ..., p$ e*

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

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

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

Pelo Teorema 1.3.11, temos que

$$\partial f(x) = \left\{ y \in \mathbb{R}^n \ \middle|\ y = \sum_{i\in I(x)} \alpha_i \ \mathrm{grad}\ f_i(x)\ :\ \sum_{i\in I(x)} \alpha_i = 1, \alpha_i \geqslant 0,\ i \in I(x) \right\}.$$

Em particular, pelo Teorema 1.3.10, $\bar{x}$ é um minimizador de f no $\mathbb{R}^n$ se, e somente se, existem $\bar{\alpha}_i \in I(\bar{x})$, tais que

$$0 = \sum_{i\in I(\bar{x})} \bar{\alpha}_i \mathrm{grad}\ f_i(\bar{x}),\ \sum_{i\in I(\bar{x})} \bar{\alpha}_i = 1.$$

Capítulo 2

Método do ponto proximal para otimização escalar

Neste capítulo iremos abordar o chamado Método do Ponto Proximal clássico. Tal método busca resolver o problema de minimização de uma função $f : \mathbb{R}^n \to \mathbb{R}$ convexa. Na literatura existem diversas variações do método do ponto proximal, de acordo com a função distância adotada. Por exemplo, em [16] além do caso clássico, que iremos abordar, onde adota-se a distância euclidiana, é possível encontrar quando adota-se a função de Bregman, como é trabalhado [7].

Nosso objetivo aqui será além de definir o método clássico, mostrar que o mesmo está bem definido e analisar a sua convergência. Em outras palavras, estamos interessados em saber sob quais hipóteses o método do ponto proximal clássico converge para uma solução do Problema (1.1). Para maiores informações veja [16].

2.1 Método do ponto proximal clássico

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Considere a sequência $\{x^k\} \subset \mathbb{R}^n$ dada por:

$$\begin{aligned} & x^0 \in \mathbb{R}^n \\ & x^{k+1} := \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \parallel x - x^k \parallel^2\} \end{aligned} \tag{2.1}$$

onde $\lambda_k \in \mathbb{R}$; $0 < \lambda_k \leqslant \tilde{\lambda}$, para algum $\tilde{\lambda} > 0$. Acima $\parallel . \parallel$ denota a norma euclidiana.

Proposição 2.1.1. *O método do ponto proximal definido em* (2.1) *está bem definido.*

Demonstração. Por indução, considere $f_k(x) = f(x) + \lambda_k \parallel x - x^k \parallel^2$. Como estamos supondo que f atinge o mínimo, temos que f é limitada inferiormente, logo $\lim\limits_{j\to\infty} f_k(y^j) = +\infty$, onde $\{y^j\} \subset \mathbb{R}^n$ com $\parallel y^j \parallel \longrightarrow +\infty$. De fato, como

$$\parallel y^j \parallel \leqslant \parallel y^j - x^k \parallel + \parallel x^k \parallel .$$

Então, $\parallel y^j - x^k \parallel \longrightarrow +\infty$, quando $j \longrightarrow +\infty$. Como sabemos que $\{f(y^j)\}$ é limitada inferiormente, usando a definição de f_k temos que $\lim\limits_{j\to\infty} f_k(y^j) = +\infty$. Em outras palavras, f_k é coerciva. Portanto, como $f_k(x)$ é claramente contínua atinge o mínimo e é único, pois pelo Teorema 1.3.3 , f_k é estritamente convexa, logo x^{k+1} é único.

□

2.2 Análise de convergência

A seguir mostraremos que sob algumas hipóteses a sequência gerada por (2.1) converge para o minimizador de f. Para isso iremos precisar da seguinte definição e proposição.

Definição 2.2.1. *([16]) Uma sequência $\{y^k\} \subset \mathbb{R}^n$ é dita Fejér convergente para o conjunto $U \subset \mathbb{R}^n$, com respeito a norma euclidiana se,*

$$\parallel y^{k+1} - u \parallel \leqslant \parallel y^k - u \parallel \quad \forall k \geqslant 0, \forall u \in U. \tag{2.2}$$

Proposição 2.2.1. *([16]) Se $\{y^k\}$ é Fejér convergente para $U \neq \emptyset$ então $\{y^k\}$ é limitada. Se um ponto de acumulação y de $\{y^k\}$,com $y \in U$ então $\lim\limits_{k\to\infty} y^k = y$.*

Demonstração. De (2.2) temos que,

$$\parallel y^{k+1} - u \parallel \leqslant \parallel y^0 - u \parallel,$$

$\forall u \in U$, isto é, $\{y^k\} \in B[u, r]$, para cada $k \in \mathbb{N}$, com $r = \parallel y^0 - u \parallel$. Onde $B[u, r]$ denota a bola fechada de centro em u e raio r, isto é, $\{y^k\}$ é limitada.

Agora seja $y \in U$ um ponto de acumulação de $\{y^k\}$, logo $\exists \{y^{k_j}\}$ tal que $y^{k_j} \longrightarrow y$. Por (2.2) temos que a sequência $\{\parallel y^k - y \parallel\}$ é decrescente e não-negativa (aqui fazemos $u = y$ em (2.2), já que por hipótese $y \in U$), logo temos que tal sequência converge com

$$\lim_{k\to\infty} \parallel y^k - y \parallel = 0.$$

Portanto, $\lim_{k\to\infty} y^k = y$.

□

Teorema 2.2.1. *([16]) Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *convexa e continuamente diferenciável. Suponha que o conjunto* U *dos minimizadores de* f *em* $\mathbb{R}^n$ *é não-vazio. Então a sequência* $\{x^k\}$ *gerada por* (2.1) *converge para algum ponto* $x^* \in U$.

Demonstração. Sendo $\{x^k\}$ a sequência gerada em (2.1), dividiremos a demonstração em três etapas. Primeiramente iremos mostrar que a sequência é Fejér convergente para U, na segunda etapa mostraremos que $\lim_{k\to\infty} \| x^{k+1} - x^k \| = 0$ e na terceira etapa iremos mostrar que qualquer ponto de acumulação da sequência em (2.1) é solução do nosso problema e a sequência converge para tal ponto.

$1°\mathtt{Etapa}$:

Afirmação: $\| x^{k+1} - \bar{x} \|^2 \leqslant \| x^k - \bar{x} \|^2 - \| x^{k+1} - x^k \|, \forall k \geqslant 0, \forall \bar{x} \in U$.

Note que,

$$\| x^k - \bar{x} \|^2 \leqslant \| x^k - x^{k+1} \|^2 + \| x^{k+1} - \bar{x} \|^2 + 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x} \rangle. \tag{2.3}$$

Como já sabemos, x^{k+1} é solução de (2.1) logo,

$$0 = \mathrm{grad} f_k(x^{k+1}) = \mathrm{grad} f(x^{k+1}) + 2\lambda_k (x^{k+1} - x^k) \tag{2.4}$$

De (2.3), (2.4) e pela convexidade de f

$$\begin{aligned} \| x^k - \bar{x} \|^2 - \| x^{k+1} - x^k \|^2 - \| x^{k+1} - \bar{x} \|^2 &= 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x} \rangle \\ &= \tfrac{1}{\lambda_k} \langle \mathrm{grad}\ f(x^{k+1}), x^{k+1} - \bar{x} \rangle \\ &\geqslant 0, \end{aligned}$$

onde a última desigualdade é devido ao Teorema 1.3.4 e pelo fato de $\bar{x}$ ser minimizador. Com isso obtemos que,

$$\begin{aligned} \| x^{k+1} - \bar{x} \|^2 &\leqslant \| x^k - \bar{x} \|^2 - \| x^{k+1} - x^k \|^2 \\ &\leqslant \| x^k - \bar{x} \|^2 . \end{aligned}$$

Então, $\| x^{k+1} - \bar{x} \| \leqslant \| x^k - \bar{x} \|, \forall \bar{x} \in U$. Portanto, $\{x^k\}$ é Fejér convergente para U.

$2°\mathtt{Etapa}$: Da etapa 1 temos que

$$0 \leqslant \| x^{k+1} - x^k \|^2 \leqslant \| x^k - \bar{x} \|^2 - \| x^{k+1} - \bar{x} \|^2 . \tag{2.5}$$

Pelo fato de $\{x^k\}$ ser Fejér convergente temos que $\{\| x^k - \bar{x} \|\}$ é não-crescente e não-negativa, logo convergente, isto é,

$$\| x^k - \bar{x} \| \longrightarrow 0. \tag{2.6}$$

Daí, por (2.5)

$$\lim_{k \to \infty} (x^{k+1} - x^k) = 0. \tag{2.7}$$

$3^{\circ}\mathtt{Etapa}$: A existência de pontos de acumulação de $\{x^k\}$ segue da etapa 1, onde mostramos que $\{x^k\}$ é Fejer convergente e da Proposição 2.1.1. Agora, seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$, então $x^{k_j} \longrightarrow \bar{x}$ para alguma subsequência x^{k_j}. Daí, por (2.4)

$$\mathrm{grad}\ f(x^{k_j+1}) = 2\lambda_{k_j}(x^{k_j} - x^{k_j+1}).$$

Fazendo $j \longrightarrow +\infty$ e usando (2.6) e (2.7) temos que

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

Como $f \in C^1$ e é convexa, pelo Teorema 1.3.5 temos que $\bar{x} \in U$. Portanto, pela Proposição 2.1.1

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

$\square$

Capítulo 3

Otimização multiobjetivo

Neste capítulo abordaremos algumas informações que serão de grande importância na sequência desse trabalho. Apresentaremos conceitos relacionados à Otimização Multiobjetivo, iniciando com definições sobre ordenação de conjuntos.

3.1 Ordem parcial

Nesta seção apresentaremos uma noção de ordem parcial no espaço $\mathbb{R}^n$ que será útil quando definirmos o problema de Otimização Multiobjetivo. Tal noção de ordem também se faz necessário quando falamos de convexidade de uma função $F : \mathbb{R}^m \longrightarrow \mathbb{R}^n$. Os resultados e definições aqui abordados serão de grande valia quando formos estudar o resultado principal de [4], abordado mais adiante no Capítulo 5.

Definição 3.1.1. *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:*

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

Segue abaixo uma definição importante quando estudamos otimização multiobjetivo.

Definição 3.1.2. *Dizemos que* 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* $y \preceq x$.

As operações de comparação entre vetores pertencentes ao $\mathbb{R}^n$ serão definidas da seguinte forma:

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

O operador $\neq$ é dado por:

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

isto é,

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

Note então que o espaço $\mathbb{R}^n$ é parcialmente ordenado com a ordem definida acima.

Exemplo 3.1.1. *Considere os vetores* $x, y, z \in \mathbb{R}^3$*. Temos:*

$$x = (4, 6, 8), y = (3, 5, 7), z = (4, 5, 6)$$

$$\begin{aligned} y \prec x \Rightarrow y \preceq x \\ z \prec x \Rightarrow z \preceq x \end{aligned}$$

Porém, os vetores z *e* y *não podem ser comparados.*

Seguindo esse raciocínio, neste trabalho iremos adotar os cones:

$$\mathbb{R}^m_+ := \{x \in \mathbb{R}^m | x_i \geqslant 0, \ j \in I\},$$

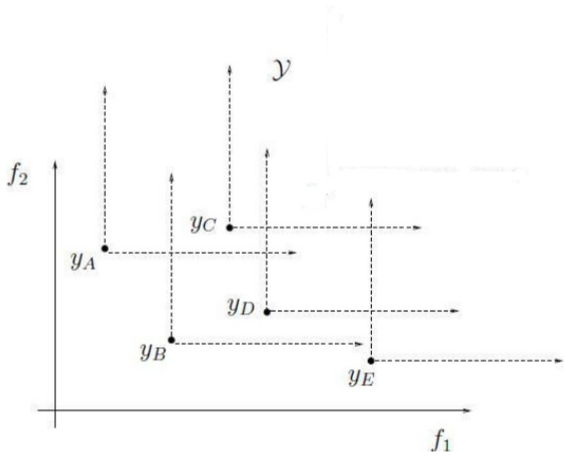
e

$$\mathbb{R}^m_{++} := \{x \in \mathbb{R}^m | x_i > 0, \ j \in I\}, \ I := \{1, ..., m\}.$$

Daí, para $y, z \in \mathbb{R}^m$, escrevemos $z \succeq y$ (ou $y \preceq z$) para significar $z - y \in \mathbb{R}^m_+$ e $z \succ y$ (ou $y \prec z$) para $z - y \in \mathbb{R}^m_{++}$.

Definição 3.1.3. ***(Dominância)*** *Diz-se que o ponto* $x_1 \in \mathbb{R}^n$ *domina outro ponto* $x_2 \in \mathbb{R}^n$ *se* $F(x_1) \preceq F(x_2)$ *e* $F(x_1) \neq F(x_2)$.

A figura abaixo ilustra o conceito de dominância.



Sejam x_A, x_B, x_C, x_D e x_E, pontos de modo que $F(x_A) = y_A, F(x_B) = y_B, F(x_C) = y_C, F(x_D) = y_D$ e $F(x_E) = y_E$. Note que x_C é dominado por x_A, enquanto x_B, x_A e x_E não são dominados por nenhuma outra solução.

3.2 Problema de Otimização Multiobjetivo

Nessa seção apresentaremos algumas propriedades e notações em otimização multiobjetivo. Iniciamos com o problema de Otimização Multiobjetivo, objeto de estudo deste trabalho, definido por:

$$\begin{array}{ll} \min & F(x) \\ \text{s.a} & x \in \mathbb{R}^n, \end{array} \tag{3.1}$$

onde $F : \mathbb{R}^n \to \mathbb{R}^m$.

Podemos dizer que a otimização multiobjetivo é um processo que busca otimizar simultaneamente duas ou mais funções-objetivo de várias variáveis em valores reais. Como é mencionado em [1], no que diz respeito as soluções de um problema de otimização multiobjetivo, há dois possíveis tipos:

(i) Soluções que, quando comparadas às demais apresentam pior desempenho sob todos os objetivos simultaneamente considerados.

(ii) Soluções que, quando comparadas às demais, são melhores em um ou mais objetivos.

Obviamente, as soluções do primeiro grupo não nos interessa, enquanto as soluções do segundo grupo são as que possivelmente irão otimizar as funções objetivos estudadas.

As primeiras definições estudadas no ambiente de Otimização Multiobjetivo são dadas abaixo.

Definição 3.2.1. *Um ponto* $\bar{x} \in \mathbb{R}^n$ *é dito crítico Pareto quando*

$$\mathrm{Im}(JF(\bar{x})) \cap (-\mathbb{R}_{++})^m = \emptyset, \tag{3.2}$$

onde $\mathrm{Im}(JF(\bar{x}))$ *é a imagem da transformação linear* $JF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$. *Ou seja, sendo* $\bar{x} \in \mathbb{R}^n$ *um ponto crítico Pareto e* $F = (F_1, ..., F_m)$ *então para todo* $v \in \mathbb{R}^n$, $\exists\, i \in \{1, ..., m\}$ *tal que*

$$\langle \mathrm{grad}\ F_i(\bar{x}), v\rangle \geqslant 0.$$

Uma importante observação que segue da definição anterior é a seguinte: se um ponto $x \in \mathbb{R}^n$ não é crítico Pareto, então existe uma direção $v \in \mathbb{R}^n$ tal que

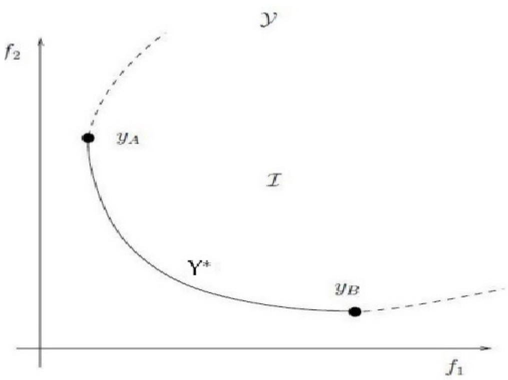
$$JF(x)v \in (-\mathbb{R}_{++})^m,$$

isto é, v é uma direção de descida para a função objetivo F. Em outras palavras temos que

$$\langle \mathrm{grad}\ F_i(x), v\rangle < 0,\ \forall i \in \{1, ..., m\}.$$

Definição 3.2.2. *Um ponto* $x \in \mathbb{R}^n$ *é dito ótimo Pareto se não existe um ponto* $y \in \mathbb{R}^n$ *com* $F(y) \preceq F(x)$ *e* $F(y) \neq F(x)$.
De forma análoga, $x \in \mathbb{R}^n$ *chama-se ótimo Pareto fraco ou Pareto fraco se não existe um ponto* $y \in \mathbb{R}^n$ *com* $F(y) \prec F(x)$.

Denotaremos por $\mathrm{argmin}\{F(x)\ :\ x \in \mathbb{R}^n\}$ o conjunto dos pontos ótimo Pareto e $\mathrm{argmin}_w\{F(x)\ :\ x \in \mathbb{R}^n\}$ o conjunto dos pontos ótimo Pareto fracos para o Problema (3.1).
Note que a Definição 3.3.2 nos diz que se um ponto $x \in \mathbb{R}^n$ é ótimo Pareto significa dizer que x não é dominado por nenhum outro ponto.
A próxima figura ilustra o conceito de ótimo Pareto,



onde Y^* é o conjunto Pareto-ótimo representado em linha contínua.

Exemplo 3.2.1. *Seja* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *dada por*

$$F(x, y) = (x^2 + 2, y^2 + 2),$$

então o ponto $(0, 0)$ *é ótimo Pareto.*

Demonstração. De fato, caso contrário existiria $(x, y) \in \mathbb{R}^2$;

$$F(x, y) \preceq F(0, 0) \ ; \quad F(x, y) \neq (2, 2)$$

Logo, $x^2 + 2 < 2$ e $y^2 + 2 < 2$. Isto implicaria em

$$x^2 + 2 < 2 \Rightarrow x^2 < 0$$

e

$$y^2 + 2 < 2 \Rightarrow y^2 < 0,$$

ou seja, um absurdo. $\square$

Observe que vale a seguinte inclusão:

$$\mathrm{argmin}\{F(x) \ : \ x \in \mathbb{R}^n\} \subset \mathrm{argmin}_w\{F(x) \ : \ x \in \mathbb{R}^n\}.$$

De fato, dado $x \in \mathrm{argmin}\{F(x) \ : \ x \in \mathbb{R}^n\}$ então não existe $y \in \mathbb{R}^n$ tal que $F(y) \preceq F(x)$, com $F(x) \neq F(y)$. Em particular, não existe $y \in \mathbb{R}^n$ tal que $F(y) \prec F(x)$. Ou seja, x é um ponto Pareto fraco.

Definição 3.2.3. *Seja* $C \subset \mathbb{R}^n$ *um conjunto não-vazio e convexo.*
$i)$ F *é dita convexa em C (ou C-convexa) se, e somente se, para todo* $x, y \in C$ *ocorrer o seguinte:*

$$F((1-t)x + ty) \preceq (1-t)F(x) + tF(y), \ \ t \in [0, 1].$$

$ii)$ F *é chamada fortemente convexa com constante* $\nu \in \mathbb{R}^m_{++}$ *em* C *se, e somente se, para cada* $x, y \in C$*, ocorrer o seguinte:*

$$F((1-t)x + ty) \preceq (1-t)F(x) + tF(y) - t(1-t) \parallel x - y \parallel^2 \nu, \ \ t \in [0, 1].$$

Observação 3.2.1. *Note que* F *é convexa (resp. fortemente convexa) se, e somente se, cada função coordenada de* F *é convexa (resp. fortemente convexa). Isso decorre da definição de ordem parcial* $\preceq$ *introduzida anteriormente, e pela Definição* 3.2.3..

Daí, lembrando que se $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ é fortemente convexa em particular é também convexa, logo temos que se $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ é fortemente convexa então também é convexa, pois cada função coordenada o será. Mas a recíproca é falsa como foi visto no Capítulo 1. O próximo resultado nos dá uma condição necessária de otimalidade de primeira ordem para o Problema (3.1) de um ponto $x^* \in \mathbb{R}^n$.

Lema 3.2.1. *Seja* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$*, uma aplicação continuamente diferenciável e* x^* *um ponto ótimo Pareto para* (3.1)*, então*

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

Demonstração. Temos por hipótese que x^* é um ponto ótimo Pareto para o problema (3.1). Daí, temos que não existe $\bar{x} \in \mathbb{R}^n$ tal que $F_i(\bar{x}) \leqslant F_i(x^*)$, para todo $i \in \{1, 2, ..., m\}$ e $F_j(\bar{x}) \leqslant F_j(x^*)$ para algum índice $j \in \{1, 2, ..., m\}$. Agora, suponha por contradição que $d \in \mathrm{Im}(JF(x^*)) \cap (-\mathbb{R}_{++})^m$, então existe $v \in \mathbb{R}^n$ tal que $\langle \mathrm{grad}\ F_i(x^*), v\rangle < 0$, para todo $i \in \{1, 2, ..., m\}$. Como G é continuamente diferenciável temos que,

$$\lim_{t\to 0} \frac{F_i(x^* + tv) - F_i(x^*)}{t} = \langle \mathrm{grad}\ F_i(x^*), v\rangle < 0,\ \forall i \in \{1, 2, ..., m\}.$$

Mas assim, v é uma direção de descida, para cada função componente F_i, isto é, existe $t_0 > 0$ tal que

$$F_i(x^* + tv) < F_i(x^*),\ \forall t \in (0, t_0],\ \forall i \in \{1, 2, ..., m\}.$$

Em outras palavras,

$$F(x^* + tv) \prec F(x^*),\ \forall t \in (0, t_0].$$

Mas isso contradiz o fato de x^* ser um ponto ótimo Pareto de (3.1). □

O Lema 3.3.1. nos diz que todo ponto ótimo Pareto é um ponto crítico Pareto.

Observação 3.2.2. *Note que o lema anterior continua sendo válido para o caso em que* x^* *é um ponto Pareto fraco do Problema* (3.1)*. Em geral, a expressão* (3.2) *é uma condição necessária, mas não suficiente para otimalidade. Por exemplo:*

Exemplo 3.2.2. *Considere* $F : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, *tal que* $F(x_1, x_2) = (x_1^3, x_2^3)$. *Daí,*

$$JF(x_1, x_2) = \begin{pmatrix} 3x_1^2 & 0 \\ 0 & 3x_2^2 \end{pmatrix}$$

e

$$Im(JF(x_1, x_2)) = \{(\langle(3x_1^2, 0), \nu\rangle, \langle(0, 3x_2^2), \nu\rangle),\ \forall \nu \in \mathbb{R}^n\}.$$

Agora, note que para $x^* = (0, 0)$ *temos que* $Im(JF(x^*) \cap (-\mathbb{R}^m_{++})) = \emptyset,\ \forall \nu \in \mathbb{R}^n$. *Porém, se considerarmos* $\bar{x} = (-1, -1)$ *obtemos que* $F(\bar{x}) \prec F(x^*)$. *Portanto,* x^* *não é um ponto Pareto fraco.*

3.2.1 Escalarização

Anteriormente definimos uma relação de ordem parcial denotada por $\preceq$. Pelo fato de tal ordem não ser total, encontramos algumas dificuldades e para tentar contornar tais dificuldades apresentamos agora um raciocínio bastante útil quando se está estudando Otimização Multiobjetivo, a chamada Escalarização. Tal técnica tem como objetivo substituir o problema de Otimização Multiobjetivo, onde não temos uma relação de ordem total, por um problema de otimização escalar, ou em uma série de problemas escalares. Com essa técnica nós podemos usar não só a relação de ordem total que temos quando trabalhamos com uma função escalar, mas também nos permite usar os resultados existentes na literatura para otimização escalar. Para esta subseção procuramos adaptar para o $\mathbb{R}^n$, foco principal deste trabalho, os resultados obtidos em [23] onde adota-se um espaço vetorial qualquer.

Definição 3.2.4. *Uma função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é dita uma representação escalar estrita de uma função vetorial* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *quando dados* $x, \bar{x} \in \mathbb{R}^n$:

$$F(x) \preceq F(\bar{x}) \Longrightarrow f(x) \leqslant f(\bar{x}) \quad \text{e} \quad F(x) \prec F(\bar{x}) \Longrightarrow f(x) < f(\bar{x}).$$

Além disso, dizemos que f *é uma representação escalar fraca de* F *se*

$$F(x) \prec F(\bar{x}) \Longrightarrow f(x) < f(\bar{x}).$$

Note que toda representação escalar estrita é também uma representação escalar fraca. O próximo resultado é de grande importância no prosseguimento deste trabalho, pois ele propõe uma maneira de obter uma representação escalar estrita para funções $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

Proposição 3.2.1. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. *Temos que* f *é uma representação escalar estrita de* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *se, e somente se, existe uma função estritamente crescente* $g : F(\mathbb{R}^n) \longrightarrow \mathbb{R}$ *tal que* $f = g \circ F$.

Demonstração. Seja $g : F(\mathbb{R}^n) \longrightarrow \mathbb{R}$ uma função crescente. Então, para todo $x, y \in \mathbb{R}^n$ temos que,

$$F(x) \succeq F(y) \Longrightarrow (g \circ F)(x) \geqslant (g \circ F)(y)$$

e

$$F(x) \succ F(y) \Longrightarrow (g \circ F)(x) > (g \circ F)(y).$$

Em outras palavras, $f = g \circ F$ é representação escalar estrita de F.
Reciprocamente, seja f uma representação escalar estrita de F. Defina a seguinte função crescente g em $F(\mathbb{R}^n)$ dada por

$$g(a) = f(x),$$

onde $a \in F(\mathbb{R}^n)$, i.e., $a = F(x)$, para algum $x \in \mathbb{R}^n$. Observe que escrevendo dessa maneira temos $f = g \circ F$. Além disso, tal função está bem definida. De fato, seja $y \in \mathbb{R}^n$ outro ponto tal que $F(y) = a$, então $F(x) = F(y) = a$. Daí, pela definição 3.2.4, temos que $f(x) \leqslant f(y)$ e $f(y) \leqslant f(x)$, i.e., $f(x) = f(y)$. Concluimos então que $g(a)$ não depende do ponto x tal que $F(x) = a$. Provemos agora que g é uma função crescente. De fato, pois se $a, b \in F(\mathbb{R}^n)$, i.e., $a = F(x),\ b = F(y)$ para algum $x, y \in \mathbb{R}^n$ e $a \succeq b$ então,

$$g(a) = f(x) \geqslant f(y) = g(b).$$

□

Exemplo 3.2.3. *Seja* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, *então* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *definida por* $f(x) = \langle F(x), z \rangle$, *com* $z \in \mathbb{R}^m \setminus \{0\}$ *fixo é uma representação escalar estrita de* F.

Demonstração. De fato, sejam $x, \bar{x} \in \mathbb{R}^n$. Se $F(x) \preceq F(\bar{x})$, então $F(\bar{x}) - F(x) \in \mathbb{R}^m_+$, ou seja, $F_i(\bar{x}) - F_i(x) \geqslant 0$ para todo $i \in \{1, 2, ..., m\}$. Daí, para todo $z \in \mathbb{R}^m_+ \setminus \{0\}$,

$$|F_i(\bar{x}) - F_i(x)| z_i \geqslant 0,\ \forall i \in \{1, 2, ..., m\}.$$

Logo, $\langle F(\bar{x}) - F(x), z \rangle \geqslant 0,\ \forall i \in \{1, 2, ..., m\}$, ou de forma equivalente,

$$\langle F(x), z \rangle \leqslant \langle F(\bar{x}), z \rangle.$$

De maneira análoga, se $F(x) \prec F(\bar{x})$, então $F(\bar{x}) - F(x) \in \mathbb{R}^m_{++}$. Portanto, $F_i(\bar{x}) - F_i(x) > 0$ para todo $i \in \{1, 2, ..., m\}$. Isto é, para todo $z \in \mathbb{R}^m_+ \setminus \{0\}$,

$$\langle F(\bar{x}) - F(x), z\rangle > 0 \Longrightarrow \langle F(x), z\rangle < \langle F(\bar{x}), z\rangle.$$

Ou seja, $f(x) = \langle F(x), z\rangle$, com $z \in \mathbb{R}^m \setminus \{0\}$ fixo é uma representação escalar estrita de F. □

Proposição 3.2.2. *Seja* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *uma função convexa, então*

$$\bigcup_{z \in \mathbb{R}^m_+ \setminus \{0\}} \mathrm{argmin}\{\langle F(x), z\rangle \mid x \in \mathbb{R}^n\} = \mathrm{argmin}_w\{F(x) \mid x \in \mathbb{R}^n\}.$$

Demonstração. Seja $\bar{x} \in \bigcup_{z \in \mathbb{R}^m_+ \setminus \{0\}} \mathrm{argmin}\{\langle F(x), z\rangle \mid x \in \mathbb{R}^n\}$, logo existe $z \in \mathbb{R}^m_+ \setminus \{0\}$ tal que,

$$\langle F(\bar{x}), z\rangle \leqslant \langle F(x), z\rangle, \ \forall x \in \mathbb{R}^n. \tag{3.3}$$

Daí, suponha por absurdo que $\bar{x} \notin \mathrm{argmin}_w\{F(x) \mid x \in \mathbb{R}^n\}$, então existe $x^* \in \mathbb{R}^n$ tal que $F(x^*) \prec F(\bar{x})$. Usando o fato de que $\langle F(.), z\rangle$ é uma representação escalar estrita de F (vide Exemplo 3.2.3.) temos que,

$$\langle F(x^*), z\rangle < \langle F(\bar{x}), z\rangle.$$

Mas isso é uma contradição com (3.3). Portanto, $\bar{x} \in \mathrm{argmin}_w\{F(x) \mid x \in \mathbb{R}^n\}$.
Reciprocamente, seja $\bar{x} \in \mathrm{argmin}_w\{F(x) \mid x \in \mathbb{R}^n\}$ temos que não existe $y \in \mathbb{R}^n$ tal que $F(y) \prec F(\bar{x})$. Agora, suponha por absurdo que $\bar{x} \notin \bigcup_{z \in \mathbb{R}^m_+ \setminus \{0\}} \mathrm{argmin}\{\langle F(x), z\rangle \mid x \in \mathbb{R}^n\}$. Então para todo $z \in \mathbb{R}^m_+ \setminus \{0\}$, existe $x_z \in \mathbb{R}^m$ tal que

$$\langle F(x_z), z\rangle < \langle F(\bar{x}), z\rangle.$$

Em particular, se tormarmos os vetores da base canônica do $\mathbb{R}^m$ temos que,

$$F_i(x_z) < F_i(\bar{x}), \ \forall i \in \{1, 2, ..., m\}.$$

Mas isso gera uma contradição, pois assim teríamos que $F(x_z) \prec F(\bar{x})$. □

Capítulo 4

Derivada Generalizada

Neste capítulo iremos generalizar alguns conceitos vistos no Capítulo 1, considerando funções localmente Lipschitz. Mais precisamente neste capítulo iremos generalizar o conceito usual de derivada direcional no espaço $\mathbb{R}^n$, bem como o de subdiferencial de uma função vista no Capítulo 1. Os resultados aqui apresentados são baseados em [6], [20] e [24].

4.1 Derivada direcional generalizada

Definição 4.1.1. *([6]) Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função localmente lipschitz,* $x \in \mathbb{R}^n$ *e* $d \in \mathbb{R}^n$. *A derivada direcional de Clarke em* x *na direção* d*, denotada por* $f^\circ(x; d)$*, é definida por*

$$f^\circ(x; d) := \lim_{t \downarrow 0} \sup_{y \to x} \frac{f(y + td) - f(y)}{t}. \tag{4.1}$$

O subdiferencial de Clarke de f *em* x*, denotado por* $\partial^\circ f(x)$*, é definido por*

$$\partial^\circ f(x) := \{w \in \mathbb{R}^n \ : \ \langle w, d\rangle \leqslant f^\circ(x, d), \ \forall d \in \mathbb{R}^n\}.$$

Proposição 4.1.1. *([6]) Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *localmente lipschitz,* $x \in \mathbb{R}^n$ *e* K_x *a constante de lipschitz para uma vizinhança* V *de* x. *Então,*

$$|f^\circ(x, d)| \leqslant K_x \|d\|, \ \forall d \in \mathbb{R}^n.$$

Demonstração. Considere $d \in \mathbb{R}^n$ e $t > 0$ tal que $y + td, y \in V$. Daí, temos

$$\begin{aligned} |f^\circ(x, d)| &= \left| \limsup_{t\downarrow 0} \sup_{y\to x} \frac{f(y+td) - f(y)}{t} \right| \\ &\leqslant \lim_{t\downarrow 0} \sup_{y\to x} \frac{|f(y+td) - f(y)|}{t} \\ &\leqslant \lim_{t\downarrow 0} \sup_{y\to x} \frac{K_x \|td\|}{t} \\ &\leqslant K_x \|d\|. \end{aligned}$$

□

Exemplo 4.1.1. *([6]) Considere* $f : \mathbb{R} \longrightarrow \mathbb{R}$*, dada por* $f(x) = |x|$*. Então,*
$\mathrm{i)Para}\ x > 0:$

$$f^\circ(x; d) := \lim_{t\downarrow 0} \sup_{y\to x} \frac{y + td - y}{t} = d.$$

Daí, $\partial^\circ f(x) := \{w \in \mathbb{R}^n \ : \ d \geqslant wd, \ \forall d \in \mathbb{R}\}$*. Logo, se* $d \geqslant dw$ *então* $w \leqslant 1$*. Agora fazendo para* $d = -1$*, tem-se que* $w \geqslant 1$*. Portanto,* $\partial^\circ f(x) := \{1\}$*.*
$\mathrm{ii)Para}\ x < 0:$

$$f^\circ(x; d) := \lim_{t\downarrow 0} \sup_{y\to x} \frac{-(y + td) - (-y)}{t} = -d.$$

De forma análoga chegamos que $\partial^\circ f(x) := \{-1\}$*.*
$\mathrm{iii)Para}\ x = 0:$

$$f^\circ(x; d) := \begin{cases} d, & d \geqslant 0 \\ -d, & d < 0 \end{cases}$$

i.e., $f^\circ(x; d) := |d|$*. Agora note que* $\partial^0 f(0) = [-1, 1]$*. De fato, dado* $w \in \partial^0 f(0)$ *então* $|d| \geqslant wd, \ \forall d \in \mathbb{R}$*. Suponha que* $|w| > 1$*, i.e.,* $w > 1$ *ou* $w < -1$

•$w > 1$: *Então, para* $d > 0$*, temos que* $dw > d = |d| \Longrightarrow dw > |d|$*.*

•$w < -1$: *Então, para* $d < 0$*, temos que* $dw > -d = |d| \Longrightarrow dw > |d|$*.*

Mas em ambos os casos temos uma contradição. Agora, dado $w \in [-1, 1]$ *temos que*

$$wd \leqslant |wd| = |w||d| \leqslant |d|, \ \mathrm{logo}\ |d| \geqslant wd.$$

Portanto, $w \in \partial^0 f(0)$*.*

Proposição 4.1.2. *([6]) Se* f *é uma função convexa, então* $f^\circ(x, d) = f'(x, d)$*, onde* $f'(x, d)$ *é a derivada direcional usual de* f *em* x *na direção* d*. Portanto,* $\partial^\circ f(x) =$

$\partial f(x)$, $\forall x \in \mathbb{R}^n$, *i.e., o subdiferencial de Clarke generaliza o subdiferencial usual para funções convexas.*

Demonstração. Inicialmente temos pelo Teorema 1.3.4. que f é localmente lipschitz, em particular contínua. Por outro lado, de (4.1) obtemos

$$f^{\circ}(x;d) := \lim_{t\downarrow 0} \sup_{y\to x} \frac{f(y+td)-f(y)}{t} \geqslant \lim_{t\downarrow 0} \frac{f(y+td)-f(y)}{t} = f^{'}(x,d),$$

Agora, dado $\delta > 0$ como $\phi(t) = t^{-1}(f(y+td)-f(y))$ é não-decrescente então $\phi(t) = \sup\ \phi(s)_{0<s<t}$. Então,

$$\begin{aligned} f^{\circ}(x,d) = \lim_{t\to 0}\left[\sup_{\|y-x\|<t\delta} \frac{f(y+td)-f(y)}{t}\right] &= \lim_{t\to 0}\left[\sup_{\|y-x\|<t\delta}\ \sup_{0<s<t} \frac{f(y+sd)-f(y)}{s}\right] \\ &= \lim_{\varepsilon\to 0}\left[\sup_{\|y-x\|<\varepsilon\delta}\ \sup_{0<s<\varepsilon} \frac{f(y+sd)-f(y)}{s}\right] \\ &= \lim_{\varepsilon\downarrow 0}\ \sup_{\|y-x\|<\varepsilon\delta} \frac{f(y+\varepsilon d)-f(y)}{\varepsilon}. \end{aligned}$$

Como f é localmente lipschitz em x, então para ε suficientemente pequeno,
$|f(y+\varepsilon d)-f(y)-f(x+\varepsilon d)+f(x)| \leqslant |f(y+\varepsilon d)-f(x+\varepsilon d)|+|f(y)-f(x)| \leqslant 2K_x|y-x|$.
Então,

$$\left|\frac{f(y+\varepsilon d)-f(y)}{\varepsilon} - \frac{f(x+\varepsilon d)-f(x)}{\varepsilon}\right| \leqslant \frac{2K_x}{\varepsilon}|y-x| \leqslant 2K_x\delta.$$

Daí,

$$f^{\circ}(x,d) = \lim_{\varepsilon\downarrow 0}\left[\sup_{\|y-x\|<\varepsilon\delta}\left[\frac{f(y+\varepsilon d)-f(y)}{\varepsilon} + \frac{f(x+\varepsilon d)-f(x)}{\varepsilon} - \frac{f(x+\varepsilon d)-f(x)}{\varepsilon}\right]\right]$$

$$\leqslant \lim_{\varepsilon\downarrow 0}\ \sup_{\|y-x\|<\varepsilon\delta}\left[\frac{f(y+\varepsilon d)-f(y)}{\varepsilon} - \frac{f(x+\varepsilon d)-f(x)}{\varepsilon}\right] + \lim_{\varepsilon\downarrow 0} \frac{f(x+\varepsilon d)-f(x)}{\varepsilon}.$$

Portanto, $f^{\circ}(x,d) \leqslant \lim_{\varepsilon\downarrow 0}\ \sup_{\|y-x\|<\varepsilon\delta}(2K_x\delta) + f^{'}(x,d)$, i.e.,

$$f^{\circ}(x,d) \leqslant 2K_x\delta + f^{'}(x,d),\ \forall\delta > 0.$$

Logo, $f^{\circ}(x,d) \leqslant f^{'}(x,d)$. Provando que $f^{\circ}(x,d) = f^{'}(x,d)$.

$\square$

Lema 4.1.1. *([6]) Seja* $\Omega \subset \mathbb{R}^n$ *aberto e convexo. Se* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é localmente lipschitz em* Ω *e* $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é convexa em* Ω*, então* $(f+g)^{\circ}(x, d) = f^{\circ}(x, d) + g^{'}(x, d)$*, para cada* $x \in \Omega$ *e* $d \in \mathbb{R}^n$*. Como consequência, se* $g : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é continuamente diferenciável em* Ω*,* $\partial^{\circ}(f + g)(x) = \partial^{\circ}f(x) + \mathtt{grad}\ g(x)$*, para cada* $x \in \Omega$*.*

Demonstração. Pela Definição 4.1.1. temos que

$$\begin{aligned}
(f+g)^{\circ}(x, d) &= \lim_{t\downarrow 0} \sup_{y\to x} \frac{f(y+td) + g(y+td) - f(y) - g(y)}{t} \\
&\leqslant \lim_{t\downarrow 0} \sup_{y\to x} \frac{f(y+td) - f(y)}{t} + \lim_{t\downarrow 0} \sup_{y\to x} \frac{g(y+td) - g(y)}{t} \\
&= f^{\circ}(x, d) + g^{'}(x, d).
\end{aligned}$$

onde acima usamos a proposição anterior e uma propriedade de supremo. Por outro lado, note que $f^{\circ}(x, d) = (f + g + (-g))^{\circ}(x, d)$, logo usando novamente a proposição anterior e a desigualdade acima temos que

$$f^{\circ}(x, d) \leqslant (f + g)^{\circ} + (-g)^{'}(x, d),$$

isto é, $(f + g)^{\circ}(x, d) = f^{\circ}(x, d) + g^{'}(x, d)$. □

Lema 4.1.2. *([3])Seja* $\Omega \subset \mathbb{R}^n$ *um conjunto aberto e convexo e* $I = \{1, ..., m\}$*. Seja* $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função continuamente diferenciável em* Ω *para todo* $i \in I$*. Definindo* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *por* $f(x) := \mathtt{max}_{i\in I} f_i(x)$*, e* $I(x) = \{i \in I \ : \ f_i(x) = f(x)\}$*. Então,* f *é localmente lipschitz em* Ω *e* $\mathtt{conv}\{\mathtt{grad}\ f_i(x) \ : \ i \in I(x)\} \subset \partial^{\circ}f(x)$*, para cada* $x \in \Omega$*.*

Demonstração. Como $f_i \in C^1$ por hipótese, então f_i é localmente lipschitz em $\Omega, \forall i \in I$, [21]. Portanto, dado $\bar{p} \in \Omega$ e $i \in I$, $\exists \delta_i, L_i > 0$;

$$|f_i(p) - f_i(q)| \leqslant L_i|p - q|, \forall p, q \in B(\delta_i, \bar{p}). \tag{4.2}$$

Como vale que,

$$|\max_{i\in I} f_i(p) - \max_{i\in I} f_i(q)| \leqslant \max_{i\in I} |f_i(p) - f_i(q)|. \tag{4.3}$$

Por (4.2) e (4.3) temos que

$$|f(p) - f(q)| \leqslant L|p - q|, \forall p, q \in B(\delta, \bar{p}),$$

onde $\delta = \mathtt{min}\delta_i$ e $L = \mathtt{max}L_i$. Logo, f é localmente lipschitz. Agora, dado $x \in \Omega$ tome $u \in \mathtt{conv}\{\mathtt{grad}\ f_i(x),\ i \in I(x)\}$ e $v \in \mathbb{R}^n$. Logo, existem $\alpha_i \geqslant 0$; $\sum_{i=1}^{m} \alpha_i = 1$ e $i \in I(x)$;

$$u = \sum_{i\in I(x)} \alpha_i \mathtt{grad}\ f_i(x).$$

Então,

$$\langle u, v\rangle = \sum_{i\in I(x)} \alpha_i \langle \mathrm{grad}\ f_i(x), v\rangle = f_i'(x, v) \Longrightarrow \langle u, v\rangle = f_i'(x, v), \tag{4.4}$$

onde acima usamos o fato de f ser C^1, em particular diferenciável, logo usando que $f_i'(x, v) \leqslant f^\circ(x, v)$, temos por (4.4) que

$$\langle u, v\rangle \leqslant f^\circ(x, v).$$

Logo, $u \in \partial^\circ f(x)$.

□

No capítulo anterior, introduzimos a idéia de ponto crítico Pareto no caso de aplicações continuamente diferenciáveis. A partir de agora, com a teoria abordada neste capítulo sobre o subdiferencial de Clarke, iremos abordar a definição de crítico Pareto-Clarke para aplicações localmente Lipschitz em $\mathbb{R}^n$, que será de grande importância neste trabalho.

Definição 4.1.2. *Seja* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *localmente lipschitz em* $\mathbb{R}^n$. *Dizemos que* $x^* \in \mathbb{R}^n$ *é ponto crítico Pareto-Clarke de* F *se, para todas as direções* $d \in \mathbb{R}^n$, *existe* $i_0 = i_0(d) \in \{1, ..., m\}$ *tal que* $f^\circ_{i_0}(x^*, d) \geqslant 0$.

Observação 4.1.1. *Note que a Definição* 4.1.2 *generaliza a Definição* 3.2.1 *vista no capítulo anterior. Pela definição de ponto crítico Pareto-Clarke temos que não existe* $d \in \mathbb{R}^n$ *que seja direção de descida para todas as funções objetivo. Observe também que se um ponto é Pareto ótimo, então é necessariamente um ponto crítico Pareto-Clarke.*

Segue da definição acima que se x não é um ponto crítico Pareto-Clarke, então existe uma direção $d \in \mathbb{R}^n$ tal que

$$f^\circ_i(x; d) < 0,\ \forall i \in \{1, 2, ..., m\}..$$

Daí, como visto na demonstração da Proposição 4.1.2., $f'(x; d) \leqslant f^\circ(x; d)$, logo d é uma direção de descida para cada função coordenada de F, isto é, existe $\epsilon > 0$ tal que Daí, como visto na demonstração da Proposição 4.1.2., $f'(x; d) \leqslant f^\circ(x; d)$, logo d é uma direção de descida para cada função coordenada de F, isto é, existe $\epsilon > 0$ tal que

$$F(x + td) < F(x),\ \forall t \in (0, \epsilon].$$

Veremos agora, um resultado apresentado em [24] sobre o subdiferencial de Clarke de uma função definida pela diferença entre funções convexas.

Teorema 4.1.1. *Seja* $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, *com* $i = 1, 2$ *funções convexas e diferenciáveis. Considere* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *dada por* $f = f_1 - f_2$. *Então,* $\partial^\circ f(x) = \partial f_1(x) - \partial f_2(x)$.

Exemplo 4.1.2. *Considere* $f, g : \mathbb{R}_{++} \longrightarrow \mathbb{R}$ *dadas por* $f(x) = \ln\,(x) + \frac{1}{x}$, $g(x) = 2\sqrt{x} + \frac{1}{x}$. *Daí, temos que* $\partial^\circ f(x)$ *e* $\partial^\circ g(x)$ *se reduzem as derivadas de* f *e* g *respectivamente. Note que* $f(x) = \frac{1}{x} - (-\ln\,(x))$ *e* $g(x) = \frac{1}{x} - (-2\sqrt{x})$, *ou seja,* f *e* g *são diferenças entre funções convexas e diferenciáveis. Daí, pelo Teorema* 4.1.1. *temos que* $\partial^\circ f(x) = \left\{ \frac{1}{x} - \frac{1}{x^2} \right\}$ *e* $\partial^\circ g(x) = \left\{ \frac{1}{\sqrt{x}} - \frac{1}{x^2} \right\}$.

O exemplo anterior é de grande relevância para ilustrar que o conjunto das funções que satisfazem as hipóteses do teorema principal deste trabalho é não-vazio.

Capítulo 5

Método do ponto proximal para otimização multiobjetivo não-convexo

Neste capítulo abordaremos o resultado principal deste trabalho, motivado por [4]. Apresentamos uma generalização do método do ponto proximal para minimizar uma função $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, não necessariamente convexa, onde as funções coordenadas de F são obtidas pelo máximo de uma certa classe de funções continuamente diferenciáveis. Inicialmente iremos abordar o método do ponto proximal apresentado em [4], em seguida mostraremos que qualquer ponto de acumulação da sequência gerada pelo método é um ponto crítico Pareto-Clarke. Por fim, adaptando alguns resultados obtidos em [5] para o espaço n-dimensional $\mathbb{R}^n$ iremos mostrar que a sequência gerada pelo método converge para um ponto crítico Pareto fraco.

5.1 Método do ponto proximal para otimização multiobjetivo não-convexo

Nesta seção apresentamos o método do ponto proximal para otimização multiobjetivo no caso não-convexo, onde o objetivo principal será provar o seguinte teorema:

Teorema 5.1.1. *([4]) Seja* $\Omega \subset \mathbb{R}^n$ *um conjunto aberto e convexo,* $y \in \mathbb{R}^n$, $\hat{I} :=$ $\{1, 2, ..., m\}$ *e* $I_j := \{1, 2, ..., l_j\}$, *com* $l_j \in \mathbb{Z}_+$. *Seja* $F : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ *definida por* $F(x) :=$

$(f_1(x), ..., f_m(x))$ *onde*

$$f_j(x) := max_{i \in I_j} f_{ij}(x), \ j \in \hat{I},$$

e $f_{ij} : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é uma função continuamente diferenciável em* Ω *e contínua em* $\bar{\Omega}$*, para todo* $i \in I_j$*. Assumimos ainda que para todo* $j \in \hat{I}$*,* $-\infty < inf_{x \in \mathbb{R}^n} f_j(x), gradf_{ij}$ *é lipschitz em* Ω *com constante* L_{ij} *para cada* $i \in I_j$ *e*

$$S_F(F(\bar{y})) := \{x \in \mathbb{R}^n : F(x) \preceq F(\bar{y})\} \subset \Omega.$$

Seja $\bar{\lambda} > 0$ *e* $\bar{\mu} > 0$ *tal que* $\bar{\mu} < 1$*. Tome as sequências* $\{e^k\} \subset \mathbb{R}^m_{++}$ *e* $\{\lambda_k\} \subset \mathbb{R}_{++}$ *satisfazendo*

$$\|e^k\| = 1, \quad \bar{\mu} < e^k_j, \quad \frac{1}{\bar{\mu}} max_{i \in I_j} L_{ij} < \lambda_k \leqslant \bar{\lambda}, \quad j \in \hat{I}, \quad k = 0, 1, 2, ... \tag{5.1}$$

Seja $\hat{x} \in S_F F(\bar{y})$*. Se* $\Omega_k := \{x \in \mathbb{R}^n \ : \ F(x) \preceq F(x^k)\}$*, então o método do ponto proximal*

$$x^{k+1} \in argmin_w \ \{F(x) + \frac{\lambda_k}{2}\|x - x^k\|^2 e^k \ : \ x \in \Omega_k\}, \quad k = 0, 1, 2, ... \tag{5.2}$$

com ponto inicial $x^0 = \hat{x}$ *está bem definido, a sequêcia gerada* $\{x^k\}$ *está contida em* $S_F(F(\bar{y}))$ *e qualquer ponto de acumulação de* $\{x^k\}$ *é um ponto crítico Pareto-Clarke de* F*, contanto que* Ω_k *seja convexo, para cada* k*.*

Denotaremos o método definido acima de Método do Ponto Proximal para Otimização multiobjetivo (MPPOM). Inicialmente, nosso objetivo será investigar a boa definição e se tal método é realmente útil para resolver (POM) que satisfaça as hipóteses do Teorema 5.1.1.. Antes disso, vejamos o seguinte lema.

Lema 5.1.1. *([4]) Para todo* $\tilde{x} \in \mathbb{R}^n, \ \nu := (\nu_1, ...\nu_m) \in \mathbb{R}^m_{++}, \ j \in \hat{I}$ *e* λ *satisfazendo* $sup_{i \in I_j} L_{ij} < \lambda \nu_j$*, a função* $f_{ij} + \frac{\lambda}{2}\nu_j\|. - \tilde{x}\|^2$ *é fortemente convexa em* Ω *com constante* $\lambda\nu_j - sup_{i \in I_j} L_{ij}$*. Consequentemente, a função* $f_j + \frac{\lambda}{2}\nu_j\|. - \tilde{x}\|^2$ *é fortemente convexa em* Ω *com constante* $\lambda\nu_j - sup_{i \in I_j} L_{ij}$ *e* $F + \frac{\lambda}{2}\|. - \tilde{x}\|^2 \nu$ *é fortemente convexa em* Ω*. Portanto,* $\langle F(.), z\rangle + \frac{\lambda}{2}\langle \nu, z\rangle\|. - \tilde{x}\|^2$ *é fortemente convexa em* Ω *para cada* $z \in \mathbb{R}^m_+ \setminus \{0\}$*.*

Demonstração. Dados $j \in \hat{I}, \ i \in I_j, \ \tilde{x} \in \mathbb{R}^n, \ \nu_j \in \mathbb{R}_{++}$, defina

$$h_{ij} = f_{ij} + \frac{\lambda}{2}\nu_j\|. - \tilde{x}\|^2.$$

Logo,

$$\mathrm{grad}\ h_{ij}(x) = \mathrm{grad}\ f_{ij}(x) + \lambda\ \nu_j\ (x - \tilde{x}),$$

temos então que,

$$\langle \mathrm{grad}\ h_{ij}\ (x) - \mathrm{grad}\ h_{ij}(y), x - y\rangle = \langle \mathrm{grad}\ f_{ij}(x) - \mathrm{grad}\ f_{ij}(y), x - y\rangle + \lambda\ \nu_j\ \|x - y\|^2.$$

Daí, por Cauchy-Schwarz

$$\begin{aligned}\langle \mathrm{grad} h_{ij}(x) - \mathrm{grad} h_{ij}(y), x - y\rangle &\geqslant -\|\mathrm{grad} f_{ij}(x) - \mathrm{grad} f_{ij}(y)\|\ \|x - y\| + \lambda\ \nu_j\ \|x - y\|^2\\ &\geqslant (\lambda\ \nu_j - L_{ij})\|x - y\|^2\\ &\geqslant (\lambda\ \nu_j - \sup_{i\in I_j} L_{ij})\|x - y\|^2\end{aligned}$$

onde na última desigualdade acima se deve ao fato de $\mathrm{grad}\ f_{ij}$ ser Lipschitz em Ω e a hipótese que $\lambda\ \nu_j > \mathrm{sup}_{i\in I_j} L_{ij}$. Portanto, concluimos que h_{ij} é fortemente convexa (vide Proposição 1.3.2). A prova restante resulta da Proposição 1.3.1. e Observação 3.2.1.. □

Observe que as hipóteses do Teorema (5.1.1.), implicam na condição $\mathrm{sup}_{i\in I_j} L_{ij} < \lambda\nu_j$, com $\lambda_k = \lambda$ e $e_j^k = \nu_j$. Mostraremos agora a boa definição do (MPPOM):

Proposição 5.1.1. *([4]) O método do ponto proximal definido em* (5.2) *aplicado a função* F *com ponto inicial* $x^0 = \hat{x}$ *está bem definido.*

Demonstração. A prova será feita por indução em k. O objetivo aqui será mostrar a existência de cada iteração e que $\{x^k\} \subset S_F(F(\bar{y}))$. Daí, seja $\{x^k\}$ a sequência definida em (5.2), por hipótese temos que $\hat{x} := x^0 \in S_F(F(\bar{y}))$. Agora, suponha que $x^k \in S_F(F(\bar{y}))$ para algum k. Tomando $z \in \mathbb{R}^m_+ \setminus \{0\}$, defina $\phi_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ dada por

$$\phi_k(x) = \langle F(x), z\rangle + \frac{\lambda_k}{2}\langle e^k, z\rangle\|x - x^k\|^2.$$

Daí, como temos que $-\infty < \inf_{x\in\mathbb{R}^n} f_j(x),\ \forall j \in \hat{I}$, a função $\langle F(.), z\rangle$ é limitada inferiormente; além disso $\langle e^k, z\rangle > 0$ (pois $\{e^k\} \subset \mathbb{R}^m_{++}$ e $z \in \mathbb{R}^m_+ \setminus \{0\}$). Segue de forma análoga ao caso escalar que ϕ_k é coerciva e que existe $\tilde{x} \in \mathbb{R}^n$ tal que

$$\tilde{x} = \mathrm{argmin}\ \phi_k(x) \tag{5.3}$$

Pois, ϕ_k é fortemente convexa, pelo Lema (5.1.1.), em paricular, estritamente convexa. Mas, como Ω_k é fechado temos que $\tilde{x} \in \Omega_k$, logo podemos reescrever (5.3) como

$$\tilde{x} = \mathrm{argmin}_{x\in\Omega_k}\ \phi_k(x) \tag{5.4}$$

Portanto, usando a Proposição 3.3.2. podemos tomar $x^{k+1} := \tilde{x}$. Note que a hipótese de indução foi usada para garantir que $x^{k+1} \in S_F(F(\hat{y}))$. □

Note que, seja $\{x^k\}$ a sequência gerada em (5.2), então pela Proposição 3.3.2. existe uma sequência $\{z^k\} \subset \mathbb{R}^m_+ \setminus \{0\}$ tal que

$$x^{k+1} = argmin_{x\in\Omega_k}\psi_k(x), \tag{5.5}$$

onde $\psi_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ é dada por

$$\psi_k(x) := \langle F(x), z^k\rangle + \frac{\lambda_k}{2}\langle e^k, z^k\rangle \|x - x^k\|^2. \tag{5.6}$$

Como a solução do problema (5.5) não se altera com a multiplicação de z^k escalares positivos, então podemos assumir que $\|z^k\| = 1$, (multiplicando por $\frac{z^k}{\|z^k\|}$) para $k = 0, 1, ...$

Agora, com base nos resultados vistos anteriormente, em especial a Proposição 5.1.1. que mostra a boa definição do método, estamos em condições de provar o Teorema 5.1.1.

Demonstração do Teorema 5.1.1. : Seja $\{x^k\}$ a sequência gerada pelo método. Como $x^0 = \hat{x} \in S_F(F(\hat{y})) \subset \Omega$, então por (5.2) temos que $F(x^{k+1}) \preceq F(x^k)$, logo $\{x^k\} \subset S_F(F(\hat{y}))$. Agora, seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$, assumimos que Ω_k é um conjunto convexo e que $\bar{x}$ não seja um ponto crítico Pareto-Clarke em $\mathbb{R}^n$. Então, existe $d \in \mathbb{R}^n$ tal que

$$f_i^\circ(\bar{x}, d) < 0, \quad i \in \{1, ..., m\} \tag{5.7}$$

Logo, d é uma direção de descida para função multiobjetivo F em $\bar{x}$, i.e., existe $\delta > 0$ tal que $F(\bar{x} + td) \prec F(\bar{x})$ para todo $t \in (0, \delta]$. Daí, $\bar{x} + td \in \Omega_k$ para $k = 0, 1, ...$

Seja $\{z^k\}$ uma sequência satisfazendo (5.5), então usando o Lema 5.1.1. e Teorema 1.3.10 obtemos

$$0 \in \partial\left(\langle F(.), z^k\rangle + \frac{\lambda_k}{2}\langle e^k, z^k\rangle \|. - x^k\|^2\right)(x^{k+1}) + N_{\Omega_k}(x^{k+1}),\ k = 0, 1, ...$$

Escrevendo $z^k = (z^k_1, ..., z^k_m)$, $e^k = (e^k_1, ..., e^k_m)$, por (1.31) obtemos que,

$$0 \in \sum_{j=1}^m z^k_j \partial\left(f_j + \frac{\lambda_k}{2} e^k_j \|. - x^k\|^2\right)(x^{k+1}) + N_{\Omega_k}(x^{k+1}),\ k = 0, 1, ...$$

Então, existe $v^{k+1} \in N_{\Omega_k}(x^{k+1})$ tal que

$$0 = \sum_{j=1}^m z^k_j \partial\left(f_j + \frac{\lambda_k}{2} e^k_j \|. - x^k\|^2\right)(x^{k+1}) + v^{k+1},\ k = 0, 1, ...$$

Lembrando que por hipótese temos que $\bar{\mu} < e_j^k$ e $\frac{1}{\bar{\mu}} max_{i\in I_j} L_{ij} < \lambda_k$, então $max_{i\in I_j} L_{ij} < \lambda_k e_j^k$. Daí, pelo Lema 5.1.1 temos que $f_{ij} + \frac{\lambda_k}{2} e_j^k \|. - x^k\|^2$ e $f_j + \frac{\lambda_k}{2} e_j^k \|. - x^k\|^2$ são funções fortemente convexas, para todo $j \in \hat{I}$ e $k = 0, 1, ...$ Então para cada $j \in \hat{I}$, usando o Exemplo 1.3.7 existem constantes $\alpha_{ij}^{k+1} \geqslant 0$ com $i \in I_j(x^{k+1})$ tal que

$$0 = \sum_{j=1}^{m} z_j^k \left(\sum_{i \in I_j(x^{k+1})} \alpha_{ij}^{k+1} \, grad \, \left(f_{ij} + \frac{\lambda_k e_j^k}{2} \|. - x^k\|^2 \right) (x^{k+1}) \right) + \nu^{k+1}, \quad \sum_{i \in I_j(x^{k+1})} \alpha_{ij}^{k+1} = 1,$$

para todo $k = 0, 1, ...$ Isso nos diz que

$$0 = \sum_{j=1}^{m} z_j^k \left(\sum_{i \in I_j(x^{k+1})} \alpha_{ij}^{k+1} \, \left(grad \, f_{ij}(x^{k+1}) + \lambda_k e_j^k (x^{k+1} - x^k) \right) \right) + \nu^{k+1} \tag{5.8}$$

com $\sum_{i \in I_j(x^{k+1})} \alpha_{ij}^{k+1} = 1$, para todo $k = 0, 1, ...$ Agora, para cada $j \in \hat{I}$, seja $\{\alpha_{ij}^{k+1}\} \subset \mathbb{R}$ a sequência definida por

$$\alpha_j^{k+1} = (\alpha_{1j}^{k+1}, \alpha_{2j}^{k+1}, ..., \alpha_{mj}^{k+1}), \ \alpha_{ij}^{k+1} = 0, \ i \in I_j \setminus I_j(x^{k+1}), \ k = 0, 1, ...$$

Como $\sum_{i \in I_j(x^{k+1})} \alpha_{ij}^{k+1} = 1$ e $\alpha_{ij}^{k+1} \geqslant 0$, temos que $\|\alpha_{ij}^{k+1}\|_1 = 1$ para todo k, onde $\|.\|_1$ denota a norma da soma em $\mathbb{R}^n$. Logo, $\{\alpha_j^{k+1}\}$ é limitada. Daí, por hipótese e por estarmos supondo que $\bar{x}$ é ponto de acumulação e $\|z^k\| = 1$, existem subsequências, $\{z^{k_s+1}\}$, $\{x^{k_s+1}\}$, $\{e_j^{k_s+1}\}$, $\{\lambda_{k_s+1}\}$ e $\{\alpha_j^{k_s+1}\}$ de $\{z^{k+1}\}$, $\{x^{k+1}\}$, $\{e_j^{k+1}\}$, $\{\lambda_{k+1}\}$ e $\{\alpha_j^{k+1}\}$, respectivamente, tais que

$$\begin{aligned} &\lim_{s\to\infty} z^{k_s+1} = \bar{z}, \ \lim_{s\to\infty} x^{k_s+1} = \bar{x}, \ \lim_{s\to\infty} e_j^{k_s+1} = \bar{e}_j, \\ &\lim_{s\to\infty} \lambda_{k_s+1} = \bar{\lambda}, \quad \lim_{s\to\infty} \alpha_j^{k_s+1} = \bar{\alpha}_j \end{aligned} \tag{5.9}$$

Como ja foi visto $\{x^k\} \subset S_F(F(\bar{y}))$, então pelo fato de F ser contínua em Ω, concluimos que $\bar{x} \in S_F(F(\bar{y}))$. Agora, desde que I_j seja finito, podemos assumir sem perda de generalidade que

$$I_j(x^{k_1+1}) = I_j(x^{k_2+1}) = ... =: \bar{I}_j, \tag{5.10}$$

e (5.8) se reescreve como

$$0=\sum_{j=1}^{m} z_j^{k_s}\left(\sum_{i\in \bar{I}_j(x^{k+1})} \alpha_{ij}^{k_s+1}\ \left(\mathrm{grad}\ f_{ij}(x^{k_s+1})+\lambda_{k_s}e_j^{k_s}(x^{k_s+1}-x^{k_s})\right)\right)+\nu^{k_s+1} \tag{5.11}$$

com $\sum_{i\in\bar{I}_j}\alpha_{ij}^{k_s+1} = 1$, $s = 0, 1, \ldots$. Como F é contínua, segue que Ω_k é fechado. Daí, levando em conta que $x^{k_s}\in\Omega_{k_s}, \Omega_{k_s}$ é um conjunto convexo e que $\Omega_{k_s+1}\subset\Omega_{k_s}$, para $s = 0, 1, \ldots$, obtemos que

$$\tilde{\Omega} := \bigcap_{s=0}^{+\infty}\Omega_{k_s}, \tag{5.12}$$

é um conjunto não-vazio fechado e convexo. Como $\nu^{k_s+1}\in N_{\Omega_{k_s}}(x^{k_s+1})$ e $\tilde{\Omega}\subset\Omega_{k_s}$, então pela Definição 1.3.7

$$\langle \nu^{k_s+1}, x-x^{k_s+1}\rangle \leqslant 0, \quad \forall x\in\tilde{\Omega}, \quad s=0,1,\ldots. \tag{5.13}$$

Agora, a partir de (5.9) e (5.11), podemos supor que $\lim_{s\to+\infty}\nu^{k_s+1} = \bar{\nu}$. Daí, fazendo $s\to+\infty$ de (5.13) concluimos que $\bar{\nu}\in N_{\tilde{\Omega}}(\bar{x})$. Finalmente, por (5.11),

$$0=\sum_{j=1}^{m}\bar{z}_j\sum_{i\in\bar{I}_j}\bar{\alpha}_{ij}\ \mathrm{grad}\ f_{ij}(\bar{x})+\bar{\nu},\quad \sum_{i\in\bar{I}_j}\bar{\alpha}_{ij}=1.$$

Daí, $x\in\tilde{\Omega}$ e tomando $u_j=\sum_{i\in\bar{I}_j}\bar{\alpha}_{ij}\ \mathrm{grad}\ f_{ij}(\bar{x})$ a igualdade acima se torna

$$0=\sum_{j=1}^{m}\bar{z}_j\langle u_j, x-\bar{x}\rangle+\langle\bar{\nu}, x-\bar{x}\rangle. \tag{5.14}$$

Agora, como $\bar{x}+td\in\Omega_k$, para todo $k=0,1,\ldots$, (5.12) implica que $\bar{x}+td\in\tilde{\Omega}$, $t\in(0,\delta]$. Desde que $u_j=\sum_{i\in\bar{I}_j}\bar{\alpha}_{ij}\ \mathrm{grad}\ f_{ij}(\bar{x})$ e $\sum_{i\in\bar{I}_j}\bar{\alpha}_{ij}=1$, temos pelo Exemplo 1.3.7 e Lema 4.1.2, $u_j\in\partial^\circ f_j(\bar{x})$. Portanto, usando que $\bar{\nu}\in N_{\tilde{\Omega}}(\bar{x})$, a Definição 4.1.1. temos por (5.14) com $x=\bar{x}+td$ que

$$\begin{aligned}0\leqslant\sum_{j=1}^{m}\bar{z}_j\langle u_j, td\rangle &\leqslant \sum_{j=1}^{m}\bar{z}_j\langle u_j, d\rangle\\ &\leqslant \sum_{j=1}^{m}\bar{z}_j f_j^\circ(\bar{x}, d)\end{aligned}.$$

onde acima usamos que $t\in(0,\delta]$. Então, existe $j\in\{1,\ldots,m\}$ tal que $f_j^\circ(\bar{x},d)\geqslant 0$, que é uma contradição por (5.7). Concluimos então que $\bar{x}$ é um ponto crítico Pareto-Clarke e o teorema está provado.

5.2 Análise de convergência

Iremos agora para mais uma importante etapa neste trabalho, que será a análise da convergência do método definido em (5.2). Como foi dito anteriormente os resultados aqui apresentados foram obtidos através da adaptação para o $\mathbb{R}^n$ do trabalho feito em [5], onde a prova será baseada em quatro afirmações que veremos adiante. Para isso vamos introduzir algumas hipóteses no Teorema 5.1.1., a fim de garantir a convergência da sequência gerada em (5.2) para um ponto Pareto fraco. Suponha que

(H1) $U = \{y \in \mathbb{R}^n \mid F(y) \preceq F(x^k),\ k = 0, 1, ...\} \neq \emptyset$

(H2) Existe $c \in \mathbb{R}$ satisfazendo as seguintes condições:

(a) $S_F(c) := \{x \in \mathbb{R}^n \mid F(x) \preceq ce\} \neq \emptyset$ e $S_F(c) \subsetneq S_F(F(\bar{y}))$;
(b) $S_F(c)$ é convexo e F é convexa em $S_F(c)$, onde $e := (1, ..., 1) \in \mathbb{R}^m$;

(H3) Existe $\delta > 0$ tal que, para todo $z \in \mathbb{R}^m_+ \setminus \{0\}$, $x \in S_F(F(\bar{y})) \setminus S_F(c)$ e $w_z(x) \in \partial^\circ\big(\langle F(.), z\rangle\big)(x) + N_{\Omega_k}(x)$, ocorre $\|w_z(x)\| > \delta > 0$.

Note que em geral o conjunto U definido acima pode ser vazio. De fato, podemos ilustrar tal afirmação com o seguinte exemplo.

Exemplo 5.2.1. *Considere* $f(x) = x$, $\lambda_k \equiv 1$ *e o método do ponto proximal clássico para o problema de minimizar* f *em* $\mathbb{R}$. *A sequência gerada satisfaz* $x^{k+1} = x^k - 1$ *e* $\lim_{k\to\infty} f(x^k) = -\infty$. *Mas isso resulta em* $U = \emptyset$.

No caso em que U é não-vazio, provaremos agora que (H1) implica em dizer que $\{x^k\}$ possui ponto de acumulação. Temos então a seguinte afirmação:

Afirmação 1: A sequência $\{x^k\}$ definida em (5.2) é Fejér convergente em relação ao conjunto U.

De fato, note que por (5.2) temos que $F(x^{k+1}) \preceq F(x^k)$, $\forall k \in \mathbb{N}$. Agora, como

$$x^{k+1} \in argmin_w\{F(x) + \frac{\lambda_k}{2}\|x - x^k\|^2 e^k \ : \ x \in \Omega_k\},$$

então pela Proposição 3.2.2., existe $z^k \in \mathbb{R}^n_+ \setminus \{0\}$ tal que x^{k+1} é solução do problema

$$\begin{aligned} &\min \eta_k(x) \\ &x \text{ s.a. } \Omega_k \end{aligned} \tag{5.15}$$

onde $\eta_k : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{\infty\}$ é dada por

$$\eta_k(x) = \langle F(x), z^k \rangle + \frac{\lambda_k}{2}\langle e^k, z^k \rangle \|x - x^k\|^2.$$

Desde que a solução de (5.15) não seja alterada pela multiplicação de z^k por escalares positivos, nós assumimos que $\|z^k\| = 1, \forall n \in \mathbb{N}$ (multiplicando por $\frac{z^k}{\|z^k\|}$). Note que por definição

$$\Omega_k \subset \mathrm{dom}(\eta_k) = \mathrm{dom}(F).$$

Segue que x^{k+1} satisfaz a condição de otimalidade de 1° ordem, i.e., $\exists u^k \in \mathbb{R}^n$ tal que

$$u^k \in \partial \eta^k(x^{k+1}) \tag{5.16}$$

$$0 \leqslant \langle u^k, x - x^{k+1} \rangle, \ \forall x \in \Omega_k \tag{5.17}$$

De fato, temos que $0 \in \partial\eta_k(x^{k+1}) + N_{\Omega_k}(x^{k+1})$, i.e., $\exists u^k \in \partial\eta_k(x^{k+1})$ e $w^k \in N_{\Omega_k}(x^{k+1})$ tal que,

$$0 = u^k + w^k \tag{5.18}$$

Como $w^k \in N_{\Omega_k}(x^{k+1})$ então temos que

$$\langle w^k, x - x^{k+1} \rangle \leqslant 0, \ \forall x \in \Omega_k,$$

logo, $\langle -w^k, x - x^{k+1} \rangle \geqslant 0$. Portanto, por (5.18)

$$\langle u^k, x - x^{k+1} \rangle \geqslant 0, \ \forall x \in \Omega_k.$$

Agora, definimos $\psi_k : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ como

$$\psi_k(x) = \langle F(x), z^k \rangle. \tag{5.19}$$

Daí, por definição de η_k e ψ_k temos que

$$u^k = v^k + \lambda_k \langle e^k, z^k \rangle (x^{k+1} - x^k), \tag{5.20}$$

para algum $\nu^k \in \partial\psi_k(x^{k+1})$. Agora, como $U \neq \emptyset$ por hipótese, tomando $x^* \in U$, temos que $x^* \in \Omega_k$, $\forall k \in \mathbb{N}$. Daí, fazendo $x = x^*$ em (5.17) e fazendo interno com $x^* - x^{k+1}$ em (5.20), obtemos

$$\begin{aligned} 0 &\leqslant \langle u^k, x^* - x^{k+1}\rangle = \langle \nu^k, x^* - x^{k+1}\rangle + \lambda_k\langle e^k, z^k\rangle\langle x^{k+1} - x^k, x^* - x^{k+1}\rangle \\ &\leqslant \langle F(x^*) - F(x^{k+1}), z^k\rangle + \lambda_k\langle e^k, z^k\rangle\langle x^{k+1} - x^k, x^* - x^{k+1}\rangle \\ &\leqslant \lambda_k\langle e^k, z^k\rangle\langle x^{k+1} - x^k, x^* - x^{k+1}\rangle \end{aligned}$$

Onde acima usamos que, $\nu^k \in \partial\psi_k(x^{k+1})$. Além disso, usamos que $x^* \in U, z^k \in \mathbb{R}^n_+ \setminus \{0\}$, pois $F(x^*) - F(x^{k+1}) \preceq 0$ (já que $x^* \in U$), logo $\langle F(x^*) - F(x^{k+1}), z^k\rangle \leqslant 0$. Fazendo $\delta_k = \lambda_k\langle e^k, z^k\rangle$, pelo que já vimos temos que $\delta_k > 0$, logo

$$0 \leqslant \delta_n\langle x^{k+1} - x^k, x^* - x^{k+1}\rangle = \delta_k(\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 - \|x^k - x^{k+1}\|^2).$$

Daí,

$$\|x^{k+1} - x^*\|^2 \leqslant \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2. \tag{5.21}$$

Portanto, $\|x^{k+1} - x^*\| \leqslant \|x^k - x^*\|$, $\forall x^* \in U$. Provando assim, a Fejér convergência em relação ao conjunto U.

Em particular, em virtude da Proposição 2.1.1., obtemos que $\{x^k\}$ é limitada, isto é, possui ponto de acumulação. Daí, segue a seguinte observação:

Observação 5.2.1. *Se $\bar{x}$ é ponto de acumulação da sequência $\{x^k\}$ definida em (5.2), então $\bar{x} \in U$. De fato, como $\bar{x}$ é ponto de acumulação de $\{x^k\}$, existe $\{x^{k_j}\}$, tal que $\{x^{k_j}\} \longrightarrow \bar{x}$. Como F é contínua, temos que $F(\bar{x})$ é um ponto de acumulação da sequência $\{F(x^k)\}$. Daí, relembrando que $\{F(x^k)\}$ é uma sequência decrescente concluimos que $\{F(x^k)\} \longrightarrow F(\bar{x})$. Portanto, $\bar{x} \in U$.*

De forma análoga ao caso escalar visto no Capítulo 2, é possível provar que duas iterações consecutivas de (5.2) são tão próximas quanto se queira. Daí, segue a segunda afirmação:

Afirmação 2 **:** $\lim_{k\to\infty}\|x^{k+1} - x^k\| = 0$.
Segue de (5.21) que para todo $x^* \in U$, a sequência $\left\{\|x^k - x^*\|^2\right\}$ é não-crescente e limitada, logo convergente. Agora, reescrevendo (5.21) segue que

$$\|x^k - x^{k+1}\|^2 \leqslant \|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2$$

e observando que o lado direito da desigualdade anterior converge para zero (vide Proposição 2.1.1.). Portanto,

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

Provaremos agora que qualquer ponto de acumulação de $\{x^k\}$ é um ponto Pareto fraco. Para isso, segue a terceira afirmação:

Afirmação 3 **:** Optimalidade fraca dos pontos de acumulação de $\{x^k\}$.

De fato, como já foi visto temos que $\{x^k\}$ é limitada logo possui pelo menos um ponto de acumulação. Daí, seja $\bar{x}$ tal que existe subsequência $\{x^{k_j}\} \to \bar{x}$. Considere, $\psi_z : \mathbb{R}^n \longrightarrow \mathbb{R}$ dada por $\psi_z(x) = \langle F(x), z\rangle$. Segue que,

$$\psi_z(\bar{x}) \leqslant \psi_z(x^k), \tag{5.22}$$

para todo $z \in \mathbb{R}^m_+$ e para todo $k \in \mathbb{N}$. Pois,

$$\psi_z(\bar{x}) = \lim_{j\to+\infty} \psi_z(x^{k_j}) = \inf\, \{\psi_z(x^k)\} \leqslant \psi_z(x^k),$$

logo $\psi_z(\bar{x}) \leqslant \psi_z(x^k)$. Daí, tomando z como sendo a base canônica do $\mathbb{R}^m$, segue de (5.22) que

$$F(\bar{x}) \preceq F(x^k), \tag{5.23}$$

para todo $k \in \mathbb{N}$. Agora, suponha que $\bar{x}$ não seja ponto Pareto fraco, isto é, existe $\hat{x} \in \mathbb{R}^n$ tal que $F(\hat{x}) \prec F(\bar{x})$. Tomando $\{z^k\}$ satisfazendo (5.15) com $\|z^k\| = 1$, $\forall k \in \mathbb{N}$, segue que z^k é limitada. Portanto, existe $z^{k_j} \to \bar{z}$, para algum $\bar{z} \in \mathbb{R}^m$. Daí, temos que

$$\begin{aligned}
\langle F(\hat{x}) - F(\bar{x}), z^{k_j}\rangle &\geqslant \langle F(\hat{x}) - F(x^{k_{j+1}}), z^{k_j}\rangle = \psi_{k_j}(\hat{x}) - \psi_{k_j}(x^{k_{j+1}}) \\
&\geqslant \psi_{k_j}(x^{k_{j+1}}) + \langle v_{k_j}, \hat{x} - x^{k_{j+1}}\rangle - \psi_{k_j}(x^{k_{j+1}}) \\
&= \langle v_{k_j}, \hat{x} - x^{k_{j+1}}\rangle = \langle u_{k_j}, \hat{x} - x^{k_{j+1}}\rangle - \delta_k\langle x^{k_{j+1}} - x^{k_j}, \hat{x} - x^{k_{j+1}}\rangle\ . \\
&\geqslant -\delta_k\langle x^{k_{j+1}} - x^{k_j}, \hat{x} - x^{k_{j+1}}\rangle \\
&\geqslant -\delta_k\|x^{k_{j+1}} - x^{k_j}\|\|\hat{x} - x^{k_{j+1}}\|,
\end{aligned}$$

usando (5.22) na primeira desigualdade, (5.20) na segunda, (5.17) na terceira. Além disso, usamos que

$$F(\hat{x}) \prec F(\bar{x}) \preceq F(x^k),\ \forall k \in \mathbb{N}.$$

Portanto,

$$\langle F(\hat{x}) - F(\bar{x}), z^{k_j} \rangle \geqslant -\delta_k \|x^{k_{j+1}} - x^{k_j}\| \|\hat{x} - x^{k_{j+1}}\|. \tag{5.24}$$

Mas como $\|z^k\| = \|e^k\| = 1$, temos que $\{\delta_k\}$ é limitada. Além disso, como $\{x^k\}$ é limitada então $\{\|\hat{x} - x^{k_{j+1}}\|\}$ é limitada. Então, fazendo $j \to +\infty$ em (5.24) obtemos que

$$\langle F(\hat{x}) - F(\bar{x}), \bar{z} \rangle \geqslant 0. \tag{5.25}$$

Mas isso gera uma contradição com o fato de $F(\hat{x}) \prec F(\bar{x})$ e $\bar{z} \in \mathbb{R}^m_+$. Portanto, $\bar{x}$ é ponto Pareto fraco.

Afirmação 4: Unicidade do ponto de acumulação de $\{x^k\}$.

Suponha que $\bar{x}$ e $\hat{x}$ sejam pontos de acumulação de $\{x^k\}$. Como $F(\bar{x}) \preceq F(x^k)$, $\forall k \in \mathbb{N}$, segue de maneira análoga que $F(\hat{x}) \preceq F(x^k)$, $\forall k \in \mathbb{N}$. Ou seja, $\bar{x}$ e $\hat{x}$ pertencem a U. Além disso, pelo comentário feito no início da Afirmação 2, segue que $\{\|\bar{x} - x^k\|\}$ e $\{\|\hat{x} - x^k\|\}$ convergem. Logo, existem $\bar{\beta}, \hat{\beta} \in \mathbb{R}$ tais que

$$\lim_{k\to+\infty} \|\bar{x} - x^k\| = \bar{\beta} \qquad \lim_{k\to+\infty} \|\hat{x} - x^k\| = \hat{\beta}. \tag{5.26}$$

Como,

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

Concluimos de (5.26) que

$$\lim_{k\to+\infty} \langle x^k - \hat{x}, \hat{x} - \bar{x} \rangle = \frac{1}{2}\left(\bar{\beta}^2 - \hat{\beta}^2 - \|\hat{x} - \bar{x}\|^2\right). \tag{5.27}$$

Como $\hat{x}$ é ponto de acumulação de $\{x^k\}$ temos de (5.27) que

$$\bar{\beta}^2 - \hat{\beta}^2 = \|\hat{x} - \bar{x}\|^2. \tag{5.28}$$

De forma análoga, trocando $\bar{x}$ por $\hat{x}$ obtemos que $\hat{\beta}^2 - \bar{\beta}^2 = \|\hat{x} - \bar{x}\|^2$. Então, $\|\bar{x} - \hat{x}\| = 0$, ou seja, $\bar{x} = \hat{x}$. Provando assim a unicidade do ponto de acumulação de $\{x^k\}$.

Segue das afirmações acima que assumindo as hipóteses $(H1), (H2)$ e $(H3)$, a sequência gerada em (5.2) converge para um ponto Pareto fraco.

Lema 5.2.1. *([4])Assumindo que as hipóteses* (H1), (a) *em* (H2) *e* (H3) *acontecem e* λ_k *satisfaz* (5.1). *Então, depois de um número finito de passos as iterações do método do ponto proximal para otimização multiobjetivo* (5.2), *pertence ao conjunto* $S_F(c)$. *Em outras palavras, existe* k_0 *tal que* $\{x^k\} \subset S_F(c)$, *para todo* $k \geqslant k_0$.

Demonstração. Tendo por hipótese que $S_F(c) \neq \emptyset$, suponha por contradição que $x^k \in S_F(F(\bar{y})) \setminus S_F(c)$ para todo k. Seja $\{z^k\}$ a sequência definida em (5.15), combinando os Lemas 5.1.1., 4.1.1. e o Teorema 1.3.7 obtemos

$$0 \in \partial^\circ\big(\langle F(.), z^k\rangle\big)(x^{k+1}) + \frac{\lambda_k}{2}\langle e^k, z^k\rangle(x^{k+1} - x^k) + N_{\Omega_k}(x^{k+1}),\ k \geqslant 0.$$

Então,

$$-\frac{\lambda_k}{2}\langle e^k, z^k\rangle(x^{k+1} - x^k) \in \partial^\circ\big(\langle F(.), z^k\rangle\big)(x^{k+1}) + N_{\Omega_k}(x^{k+1}),\ k \geqslant 0.$$

Agora, como estamos supondo que $x^k \in S_F(F(\bar{y})) \setminus S_F(c)$, então pela (H3)

$$\frac{\lambda_k}{2}\langle e^k, z^k\rangle\|x^{k+1} - x^k\| > \delta, \quad k \geqslant 0. \tag{5.29}$$

Por outro lado, como $x^{k+1} = \mathrm{argmin}_{x \in \mathbb{R}^n}\psi_k(x)$ (vide (5.15)) e $\|z^k\| = 1$, então $\psi_k(x^{k+1}) \leqslant \psi_k(x),\ \forall x \in \mathbb{R}^n$. Em particular para $x = x^k$, logo

$$\begin{array}{rcl} \frac{\lambda_k}{2}\langle e^k, z^k\rangle\|x^{k+1} - x^k\| & \leqslant & \langle F(x^k), z^k\rangle - \langle F(x^{k+1}), z^k\rangle \\ & = & \langle F(x^k) - F(x^{k+1}), z^k\rangle \\ & = & \|\langle F(x^k) - F(x^{k+1}), z^k\rangle\| \\ & \leqslant & \|F(x^k) - F(x^{k+1})\| \end{array}.$$

onde acima usamos que $F(x^{k+1}) \preceq F(x^k),\ z \in \mathbb{R}^m_+\setminus\{0\}$. Daí, por (H1) temos que $\{\|F(x^k) - F(x^{k+1})\|\}$ converge para zero, logo pela desigualdade acima teriamos que $\|x^{k+1} - x^k\|$ também convegiria para zero, entrando em contradição com (5.22). Portanto, o lema está provado. □

5.2.1 Exemplo

Apresentaremos agora uma função que satisfaz as hipóteses do Teorema 5.1.1., além de (H1), (H2) e (H3). O objetivo será mostrar que o conjunto das funções que satisfazem tais hipóteses é não-vazio.

Exemplo 5.2.2. *Tome* $0 < \epsilon < 0.4$, $\Omega = (\epsilon, +\infty)$, $\hat{I} := \{1, 2\}$ *e* $\bar{y} = 2.718...$ *com* $\ln \bar{y} = 1$. *Seja* $F : \mathbb{R} \longrightarrow \mathbb{R}^2$ *dada por*

$$F(x) = \begin{cases} (0,0) & se \quad x \in \mathbb{R} \setminus \mathbb{R}_{++} \\ (f_1(x), f_2(x)) & se \quad x \in \mathbb{R}_{++} \end{cases}$$

onde $f_j(x) := \max_{i\in\hat{I}} f_{ij}(x)$ *para* $j \in \hat{I}$ *e*

$$f_{11}(x) = \ln x + \frac{1}{x},\quad f_{21}(x) = \ln x - \frac{1}{x},\quad f_{12}(x) = 2\sqrt{x} + \frac{1}{x}.\quad f_{22}(x) = 2\sqrt{x} - \frac{1}{x},\quad x \in \mathbb{R}_{++}.$$

Observe que f_{1j}, f_{2j} *são continuamente diferenciáveis em* Ω *e contínuas em* $\bar{\Omega}$*, para qualquer* $j \in \hat{I}$*. Daí, pelo fato de* $f''_{1j},\ f''_{2j}$ *serem limitadas em* Ω*, por [22] temos que* $f'_{1j},\ f'_{2j}$ *são Lipschitz em* Ω*, para qualquer* $j \in \hat{I}$*. Desde que* $\max\{a, b\} = \dfrac{a+b}{2} + \dfrac{|a-b|}{2}$*, para quaisquer* $a, b \in \mathbb{R}$ *concluimos que*

$$F(x) = \left(\ln x + \frac{1}{x}, 2\sqrt{x} + \frac{1}{x}\right),\ x \in \mathbb{R}_{++}.$$

Note agora que para $x = 5, y = 3$ *e* $\alpha = \dfrac{1}{2}$ *concluimos que* f_2 *não é convexa, em particular* F *nao é convexa.*

Consideremos agora o seguinte problema de otimização multiobjetivo

$$\min_w \{F(x)\ :\ x \in \Omega\}. \tag{5.30}$$

Note que esse problema tem $x^* = 1$ *como solução única. De fato, temos que*

$$f'_1(x) = \frac{x-1}{x^2} \qquad (d_1)$$

$$f'_2(x) = \frac{x^2 - \sqrt{x}}{x^2\sqrt{x}} \qquad (d_2)$$

Daí, observe que $f'_j(x) < 0$ *para* $x \in (\epsilon, 1)$*,* $f'_j(x) > 0$ *para* $x \in (1, +\infty)$ *e* $f'_j(1) = 0$ *para* $i = 1, 2$*. Portanto, segue que* $x^* = 1$ *é solução de* (5.30)*. Em outras palavras,* $F(1) \prec F(x)$*, para qualquer* $x \in \mathbb{R}_{++}$*. Assim,* $-\infty < \inf_{x\in\mathbb{R}} f_j(x)$*, para todo* $j \in \hat{I}$*. Desde que* $0 < \epsilon < \bar{y}$ *temos que* $S_F(F(\bar{y})) \subset \Omega$ *e* $S_F(F(\bar{y})) \neq \emptyset$*, pela análise feita acima e pelo fato de* $F(1) = (1, 3),\ F(\bar{y}) \cong (1.36, 3.66)$*.*

Além disso, temos que $S = \{x \in \mathbb{R}\ :\ F(x) \preceq \zeta\} = (-\infty, 0]$*, em particular* S *é convexo e não-vazio, para qualquer* $\zeta \geqslant 1$*. De fato, se* $x \in (-\infty, 0]$ *então por definição de* F*, temos que* $F(x) = (0, 0) \prec \zeta,\ \forall \zeta \geqslant 1$*. Mostraremos agora que se* $x \notin (-\infty, 0]$ *então* $x \notin S$*. Para isso, tome* $\zeta = 1$ *e consideremos os casos:*

- $0 < x < 1,$

$$f_1(x) > f_1(1) = 1 = \zeta,$$

$\bullet x = 1$,

$$f_2(1) = 3 > 1 = \zeta,$$

$\bullet x > 1$,

$$f_1(x) > f_1(1) = 1 = \zeta,$$

logo em todos os casos $x \notin S$. *Portanto,* $S = \{x \in \mathbb{R} \, : \, F(x) \preceq \zeta\} = (-\infty, 0]$.
Assim, levando em conta que $x^* = 1$ *é solução única de* (5.30) *temos que* Ω_k *é convexo, pois para cada valor de* $k, F(x^k) \succeq 1$. *Portanto,* F *satisfaz as hipóteses do Teorema* 5.1.1..
Verificaremos agora que F *satisfaz as hipóteses* $(H1), (H2)$ *e* $(H3)$. *Inicialmente, como já foi visto* $F(1) \prec F(x)$ *para qualquer* $x \in \mathbb{R}_{++}$, *logo* F *satisfaz* $(H1)$, *isto é,* $U \neq \emptyset$. *Tome agora* $c = f_2(2) \cong 3.32$. *Daí, temos considere a seguinte afirmação.*

Afirmação: $S_F(c) \subseteq [0.5, 2.7] \subset S_F(F(\bar{y}))$.
Para mostrarmos isso, vamos verificar inicialmente que se $x \notin [0.5, 2.7]$ *então* $x \notin S_F(c)$. *De fato, se* $0 < x < 0.5$ *então por* $(d_2), f_2(x) > f_2(0.5) \cong 3.41$. *Logo* $f_2(x) > c$, *isto é,* $x \notin S_F(c)$. *Analogamente, se* $x > 2.7$ *então novamente por* (d_2), $f_2(x) > f_2(2.7)$. *Ou seja,* $f_2(x) > 3.65 > c$. *Daí, como* $S_F(c) \subseteq S_{f_2}(c) \subseteq [0.5, 2.7]$, *temos que* $S_F(c) \subseteq [0.5, 2.7]$. *Observe também que* $0.5 \notin S_F(c)$. *Agora, note que*

$$f_1([0.5, 2.7]) = f_1([0.5, 1]) \cup f_1([1, 2.7]).$$

Além disso, por (d_1) *temos que*

$$\sup \; f_1([0.5, 1]) = f_1(0.5) < f_1(\bar{y}), \quad \sup \; f_1([1, 2.7]) = f_1(2.7) < f_1(\bar{y}).$$

Logo, $\sup(f_1([0.5, 2.7])) < f_1(\bar{y})$. *Analogamente, por* (d_2), $\sup \; f_2([0.5, 2.7]) < f_2(\bar{y})$. *Concluimos então que* $[0.5, 2.7] \subset S_F(F(\bar{y}))$.
Falta verificarmos agora que $S_F(c)$ *é um conjunto convexo e que* F *é convexa em* $S_F(c)$. *Para isso, note que* $f_2(x) = c$ *só admite como soluções* $\alpha_1 \cong 0.53$ *e* $\alpha_2 = 2$. *Vamos mostrar agora que* $S_F(c) = [\alpha_1, \alpha_2]$. *De fato, note que* $S_F(c) \subset [\alpha_1, \alpha_2]$, *pois se* $x < \alpha_1$ *por* (d_2) *temos que* $f_2(x) > f_2(\alpha_1) = c$, *isto é,* $x \notin S_F(c)$. *Analogamente, se* $x > \alpha_2$ *então* $f_2(x) > f_2(\alpha_2) = c$, *isto é,* $x \notin S_F(c)$. *Logo,* $S_F(c) \subset [\alpha_1, \alpha_2]$. *Por outro lado, note que* $[\alpha_1, 1] \subset S_F(c)$, *já que* f_j *é decrescente nesse intervalo e* $[1, \alpha_2] \subset S_F(c)$, *pois* f_j *é crescente nesse intervalo. Conluimos então que* $S_F(c) = [\alpha_1, \alpha_2]$, *logo convexo. Por fim, analisando o sinal da segunda derivada com* $x \in S_F(c)$ *temos*

$$f_1''(x) = \frac{-x + 2}{x^3} \geqslant 0, \quad f_2''(x) = \frac{-x^3 + 4x^{\frac{3}{2}}}{2x^{\frac{9}{2}}} \geqslant 0$$

Portanto, F *é convexa em* $S_F(c)$*. Aqui, podemos justificar a escolha feita anteriormente para mostrar que* F *não é convexa. Pois* $x = 5$ *e* $y = 3$ *não estão contidos em* Ω*, mas fora de* $S_F(c)$ *onde a função é convexa.*

Agora, para mostrarmos que (H3) *também é satisfeita note inicialmente que se* $x \in S_F(F(\bar{y}))$ *então* $F(x) \preceq F(\bar{y}) \cong (1.36, 3.66)$*, isto é,*

$$\begin{cases} f_1(x) & \leqslant & 1.36 \\ f_2(x) & \leqslant & 3.66 \end{cases}$$

Mas pelo estudo da derivada feito acima e usando o fato de que $F(1) = (1, 3)$ *temos que* $S_F(F(\bar{y})) \subset [\beta_1, \beta_2]$*, onde*

$$\begin{cases} \beta_1 & = & \max\{f_1^{-1}(f_1(\bar{y})), f_2^{-1}(f_2(\bar{y}))\} \cong 0.47 \\ \beta_2 & = & \min\{f_1^{-1}(f_1(\bar{y})), f_2^{-1}(f_2(\bar{y}))\} \cong 2.72 \end{cases}$$

Pois para $x < 1$*,* $f_j(x)$ *é decrescente e* $x > 1$*,* $f_j(x)$ *é crescente, como foi visto anteriormente. Em particular,* $S_F(F(\bar{y})) \setminus S_F(c) \subset [0.47, \alpha_1) \cup (\alpha_2, 2.72]$*. Daí, para cada* $z = (z_1, z_2) \in \mathbb{R}^2_+ \setminus \{0\}$ *com* $\|z\|_1 = z_1 + z_2 = 1$*, tome* $x \in S_F(F(\bar{y})) \setminus S_F(c)$ *e* $w_z(x) \in \partial^\circ(\langle F(.), z\rangle)(x) + N_{\Omega_k}(x)$*. Então, pelo Exemplo* 4.1.2 *existe* $\nu \in N_{\Omega_k}(x)$ *tal que*

$$w_z(x) = z_1\Big(\frac{1}{x} - \frac{1}{x^2}\Big) + z_2\Big(\frac{1}{\sqrt{x}} - \frac{1}{x^2}\Big) + \nu. \tag{5.31}$$

Analisando quando $x \in [0.47, \alpha_1)$ *temos pela Definição* 1.3.6. *que* $N_{\Omega_k}(x) \subset \mathbb{R}_-$*. Daí, por* (5.24) *obtemos*

$$\begin{aligned} w_z(x) & \leqslant & z_1\Big(\frac{1}{x} - \frac{1}{x^2}\Big) + z_2\Big(\frac{1}{\sqrt{x}} - \frac{1}{x^2}\Big) \\ & < & -\frac{0.4}{(0.47)^2}z_1 - \frac{0.2}{(0.47)^{\frac{3}{2}}}z_2, \end{aligned}$$

visto que $(x-1)/x^2 < -0.4/(0.47)^2$ *e* $(x^2 - \sqrt{x})/x^{\frac{3}{2}} < -0.2/(0.47)^{\frac{3}{2}}$*. Então, concluimos da desigualdade acima que*

$$|w_z(x)| > \frac{0.4}{(0.47)^2}z_1 + \frac{0.2}{(0.47)^{\frac{3}{2}}}z_2 > \|z\|_1\frac{0.2}{(0.47)^{\frac{3}{2}}} = \frac{0.2}{(0.47)^{\frac{3}{2}}} > \frac{1}{(2.72)^2}. \tag{5.32}$$

Analogamente, se $x \in (\alpha_2, 2.72]$ *temos que* $N_{\Omega_k}(x) \subset \mathbb{R}_+$*. Portanto, de* (5.24) *ocorre*

$$w_z(x) \geqslant z_1\Big(\frac{1}{x} - \frac{1}{x^2}\Big) + z_2\Big(\frac{1}{\sqrt{x}} - \frac{1}{x^2}\Big) > \frac{1}{(2.72)^2}z_1 + \frac{2.3}{(2.72)^{\frac{3}{2}}}z_2.$$

Logo, para qualquer $w_z(x) \in \partial^\circ(\langle F(.), z\rangle)(x) + N_{\Omega_k}(x)$ *temos que*

$$|w_z(x)| > \frac{1}{(2.72)^2}z_1 + \frac{2.3}{(2.72)^{\frac{3}{2}}}z_2 > \|z\|_1\frac{1}{(2.72)^2} = \frac{1}{(2.72)^2}, \ x \in (2, 2.72].$$

Concluimos então que F *satisfaz* $(\mathsf{H3})$ *com* $\delta = \dfrac{1}{(2.72)^2}$.

Capítulo 6

Considerações Finais

Neste trabalho abordamos o Método do Ponto Proximal para Otimização Mutiobjetivo não-convexo, onde as entradas da função vetorial estudada são dadas pelo máximo de uma certa classe de funções continuamente diferenciáveis. A grande vantagem deste método se encontra no fato de não termos necessariamente a hipótese de convexidade no problema estudado. Como já foi mencionado, tal hipótese facilita de diversas maneiras a obtenção de uma solução para nosso problema e a retirada de tal condição torna o Método mais forte.

Uma perspectiva para trabalhos futuros é a de trabalhar com o Método do Ponto Proximal para Otimização Multiobjetivo em espaços de dimensão infinita, motivados por [5]. Além disso, uma outra possibilidade é a de estudar a generalização de outros métodos para resolver um problema de Otimização Multiobjetivo.

Referências Bibliográficas

[1] Ávila, M. A., Pfeilsticker, P. C. Notas de aula: Problemas de Otimização Multiobjetivo. Disponível em :¡ http://www.cpdee.ufmg.br/ jramirez/disciplinas/otimizacao/t5.pdf ¿. Acesso em: 26 de juhlo (2013).

Marcelo de Azevedo Ávila, Pedro Carvalho Pfeilsticker

[2] Arora, J. S. and Marler, R. T. Survey of multi-objective optimization methods for engineering. Struct Multidisc Optim. 26, 369-395. (2004).

[3] Bento, G. C., Ferreira, O. P., Oliveira, P.,R. Proximal point method for a special class of nonconvex functions on Hadamard manifolds. (2012)

[4] Bento, G. C., Ferreira, O. P., Sousa Junior, V. L. Proximal point method for a special class of nonconvex multiobjective optimization problem. December 8, arXiv: 1504.00424v4, (2015).

[5] Bonnel, H., Iusem, A. N., and Svaiter, B. F. Proximal methods in vector optimization. SIAM J. Optim., 15(4):953-970 (eletronic), (2005).

[6] Clarke, F., H. Optimization and nonsmooth analysis, volume 5 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, (1990).

[7] Cruz Neto, J. X., Oliveira, P.R., Souberyan, A., Souza, S. S. A proximal method with separable Bregman distances for quasiconvex minimization over the nonnegative orthant. European Journal of Operational. 201, 365-376, (2010).

[8] Custodio, A.L., Madeira, J.F.A., Vaz, A.I.F e Vicente, L.N.: Direct Multisearch for multiobjective optimization. SIAM J. Optim., 21, 1109-1140, (2011).

[9] E. Carrizosa, J.B.G. Frenk, Dominating sets for convex functions with some applications. J. Optim. Theory Appl. 97(2), 281-295, (1998).

[10] E.H.Fukuda and L. M. Graña Drummond. Inexact projected gradient method for vector optimization. Comput. Optim. Appl., $54(3) : 473 - 493, (2013)$.

[11] E. H. Fukuda and L. M. Graña Drummond. On the convergence of the projected gradient method for vector optimization. Optimization, $60(8-9) : 1009-1021, (2011)$.

[12] G. C. Bento, O. P. Ferreira, P. R. Oliveira. Proximal point method for a special class of nonconvex functions on Hadamard manifolds. Optimization, 64(2):289-319, (2015).

[13] G. C. Bento and J. X. Cruz Neto. A subgradient method for multiobjective optimization on Riemannian manifolds. J. Optim, Theory Appl., 159(1):125-137,(2013).

[14] G. C. Bento and J. X. Cruz Neto, P. R. Oliveira, and A. Souberyan. The self regulation problem as an inexact steepest descent method for multicriteria optimization. European J. Oper. Res., 235(3):494-502, (2014).

[15] Handl, J., Kell, B. D., and Knowles, J. Multiobjective Optimization in Bioinformatics and Computational Biology. IEEE/ACM Transactions on computational biology and bioinformatics, Vol. 4, no. 2, April-June, (2007)

[16] Iusem, A, N,. Métodos de Ponto Proximal em Otimização, 20º Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, (1995).

[17] Izmailov, A., Solodov, M. Otimização, Volume 1, Rio de Janeiro: IMPA, (2009).

[18] Izmailov, A., Solodov, M. Otimização, volume 2, Rio de Janeiro, IMPA, (2007).

[19] Jahn, J., Vector Optimization: Theory, Applications and Extensions. Second Edition. Springer. (2011).

[20] J. B. Hiriart-Urruty: Generalized differentiability, duality and optimization for problems dealing with difference of convex functions. Convexity and duality in opti-

mization (Groningen, 1984), 37-70, Lecture Notes in Econom. and Math. Systems, Springer, Berlin, (1985).

[21] Lima, Elon L.: Análise Real, Volume 2, Funções de n várias variáveis. 11ªed. Rio de Janeiro: IMPA, (2012).

[22] Lima, E. L., Curso de Análise, volume 2. Projeto Euclides. IMPA, Rio de Janeiro, Brasil, (2009).

[23] Luc, T. D. Theory of Vector Optimization, volume 319 of Lect. Notes in Econ. and Math. Systems. Springer-Verlag, (1989).

[24] M. Bacák and M. B. Jonathan. On difference convexity of locally lipschitz functions (Prepared for Optimization in Honour of Alfredo Iusen's Sixtieth birthday).

[25] Rockafellar, R. T.: Convex Analysis. Princeton Univ. Press (1970) .

[26] Rockafellar, R, T. Monotone Operator and the Proximal Point Algorithm. SIAM J. Optim., 14, 877-898, (1976)