Universidade Federal do Piauí

Centro de Ciências da Natureza

Pós-Graduação em Matemática

Mestrado em Matemática

Um Método Ponto Proximal para Minimização da Diferença de Funções Convexas

Fernando Santana Lima

Teresina - 2017

Fernando Santana Lima

Dissertação de Mestrado:

Um Método Ponto Proximal para Minimização da Diferença de Funções Convexas

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.

Orientador:
Prof. Dr. Jurandir de Oliveira Lopes

Teresina - 2017

FICHA CATALOGRÁFICA
Serviço de Processamento Técnico da Universidade Federal do Piauí
Biblioteca Setorial do CCN

L733m Lima, Fernando Santana.
Um método ponto proximal para minimização da diferença de funções convexas / Fernando Santana Lima. – Teresina, 2017.
42f. il.

Dissertação (Mestrado) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Pós-Graduação em Matemática, 2016.
Orientador: Prof. Dr. Jurandir de Oliveira Lopes.

1. Otimização Matemática. 2.Metódo do Ponto Proximal. 3. Funções DC. I. Título

CDD 519.3

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO PIAUÍ
CENTRO DE CIÊNCIAS DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Um Método Ponto Proximal para Minimização da Diferença de Funções Convexas

FERNANDO SANTANA LIMA

Esta Dissertação foi submetida como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**, outorgado pela Universidade Federal do Piauí.

A citação de qualquer trecho deste trabalho é permitida, desde que seja feita em conformidade com as normas da ética científica.

Dissertação aprovada em 23 de fevereiro de 2017.

Banca Examinadora:

Prof. Dr. Jurandir de Oliveira Lopes - Orientador

Prof. Dr. João Carlos de Oliveira Souza (UFPI)

Prof. Dr. Arnaldo Silva Brito - Membro externo

Aos meus pais Maria Odete e Adão, à minha esposa Jóskarlleny(Joice) e aos meus irmãos (Nairo,Thiago e Sérgio).

Agradecimentos

Agradeço primeiramente a Deus, Criador da vida e Soberano sobre todo o universo, que sempre esteve ao meu lado me dando força e coragem para seguir em frente a cada momento, produzindo em nós uma força interior para dar mais um passo e acreditar que é possível.

Aos meus pais, Maria Odete Santana Lima e Adão Francisco Vieira Lima, pelo apoio e incentivo a buscar sempre uma vida melhor atravez dos estudos buscando andar sempre em integridade de vida com dignidade e honra, que é o meio mais simples de dignificar o homem, que nunca deixaram de acreditar em mim e sempre trazendo uma palavra de ânimo. Aos meus irmãos Nairo Alex, Thiago Santana e Sérgio Henrique que presenciaram minha jornada com alegria e orgulho.

À minha querida e amada esposa Jóskarlleny(Joice), que desde de o início desta carreira esteve ao meu lado me dando apoio e incentivo, mesmo quando estava desanimado sempre esteve me encorajando a continuar. Por seu companherismo e em especial sua paciência pelas noite em que ela despertava com meu sonambolismo falando dos mais diversos tipos de funções. Pelo companherismo e amizade que tem proporcionado momentos de alegria e felicidade. Enfim, só tenho a agradecer esta pessoa maravilhosa que tenho em minha vida.

Agradeço a todos os meus amigos que direta ou indiretamente fizeram parte desta jornanda. Aos companheiros de curso pelo apoio e horas de estudo e troca de conhecimento que tivemos juntos. Aos meus amigos de graduação Abimael Augusto e Romario Leite que estiveram sempre por perto.

Aos meus professores que me ajudaram sejam com transmissão do seus conhecimento como também com palavras de incentivo. Em especial aos meus professores do curso de graduação Arnaldo Brito, que sempre estave pronto a ajudar a buscar cada vez mais um pouco de conhecimento e esclarencendo todas as eventuais dúvidas, também os professores Andreson Fabian e Gildo Jesus.

Ao professor Jurandir pelos seus trabalhos de orientação de mestrado.

Agradeço à CAPES pelo apoio financeiro.

"O Senhor teu Deus é misericordioso e não te desamparará".

Deuteronônimo 4: 31.

"Sonhe e ouse sonhar. Você nunca irá além de seus sonhos".

Marco Feliciano.

Resumo

Nesta dissertação é estudado um método de ponto proximal para minimização da diferença de duas funções convexas (Funções DC). O algoritmo apresentado baseia-se no mesmo processo do algoritmo do ponto proximal clássico, onde a partir de um ponto inicial dado, gera-se uma sequência de pontos a qual seus pontos de acumulação são pontos críticos da função objetivo, sob hipótese de limitação da mesma. Usando o método do ponto proximal para funções DC apresentaremos uma demonstração da convergência linear do método do ponto proximal clássico para funções fortemente convexas.

Palavras-chave: Problema de Minimização, Método do Ponto Proximal, Funções DC.

Abstract

In this work is studied a proximal point method for minimizing the difference of two convex functions (DC functions). The present algorithm is based a the same process as the classical proximal point method, where from a given starting point it generates a sequence whichets cluster points are critical point of the objective function. By using a proximal point method for DC functions it is showr a proof of linear convergence of the classical proximal point method.

Keywords: Minimization Problem , Proximal point method, DC Functions.

Conteúdo

Introdução

Neste trabalho consideramos o problema de otimização não necessariamente convexa, denotado por:

$$\min_{x \in \mathbb{R}^n} f(x).$$

Se a função $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ é dada pela diferença de duas funções convexas, $f(x) = g(x) - h(x)$, neste caso denominamos este problema de Otimização DC, com $g, h : \mathbb{R}^n \longrightarrow \mathbb{R}$ funções convexas. Quando h é identicamente nula, retornamos ao caso clássico de minimização convexa. Problemas de otimização DC tem um grande número de aplicações, como nas áreas: economia, engenharia, programações geométricas e outros campos de aplicações que podem ser vistos em[16, 17, 3].

O método clássico do algoritmo de ponto proximal (PPA) é uma ótima ferramenta que auxilia na resolução de problemas em otimização convexa irrestrita denotado por:

$$\min_{x \in \mathbb{R}^n} \varphi(x),$$

com $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}$, onde o algoritmo gera uma sequência $\{x^n\}$ dada por

$$x^{k+1} = argmin_{x \in \mathbb{R}^n} \left\{ \varphi(x) + \gamma_k \|x - x^k\|^2 \right\},$$

sendo $\{\gamma_k\}$ uma sequência de números reais positivos. Este método foi introduzido inicialmente por Martine [9] em 1970 e desenvolvido posteriormente por Rockafeller [11] em 1976 para encontrar zero de um operador monótono máximal T com o seguinte algoritmo,

$$0 \in c_k T(x^{k+1}) + x^{k+1} - x^k. \tag{1}$$

Em particular se $T(\cdot) = \partial\varphi(\cdot)$, onde $\partial\varphi$ denota o subdiferencial de uma função convexa, então (1) é equivalente à

$$x^{k+1} = arg \min_{x \in \mathbb{R}^n} \left\{ \varphi(x) + \frac{1}{2c_k} \|x - x^k\|^2 \right\}.$$

Mais tarde o método foi desenvolvido em vários contextos como podemos ver em [5, 6].

Alguns métodos para resover problemas de otimização DC são variações do PPA como podemos ver em: Sun et al.[15], Moudafi e Maingé[10], e recentemente Souza et al.[12]. Nos estudos feitos por Souza foi verificado o PPA para funções DC em variedades de Hadamart como podemos ver em [12, 14].

Iremos trabalhar baseado nos estudos feitos por Souza et al.[13], para o caso em que $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Assim o algoritmo será definido da seguinte forma:

$$\begin{gathered} \text{dado } x^0 \in \mathbb{R}^n, \text{ calcular} \\ w^k \in \partial h\left((x^k\right) \text{ e} \\ x^{k+1} = \arg\min\left\{ g\left(x\right) - \langle w^k, x - x^k \rangle + \frac{1}{2\lambda_k} \| \ x - x^k \ \|^2 \right\}, \end{gathered}$$

onde $\{\lambda_k\}$ é uma sequência de números reais positivos tal que $\liminf \lambda_k > 0$, onde mostraremos que a sequência quando limitada seus pontos de acumulação são pontos críticos da função objetivo.

No capítulo 1 apresentaremos alguns resultados sobre convexidade, otimização e funções DC. No capítulo 2 apresentaremos o algoritmo, do qual mostraremos sua boa definição e análise de convergência, bem como uma aplicação sobre a convergência linear do método para o caso fortemente convexo. E por último apresentaremos as considerações finais. .

Capítulo 1

Noções Preliminares

Apresentaremos agora as definições e resultados básicos que serão utilizados no decorrer deste trabalho, essenciais para o desenvolvimento do mesmo.

1.1 Noções de Otimização

Dado uma função $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, o problema de minimizaçõa da função f irrestrito será denotado da seguinte maneira:

$$\min_{x \in \mathbb{R}^n} f(x). \tag{1.1}$$

Definição 1.1. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *:*

(a) *Um ponto* $\bar{x} \in \mathbb{R}^n$ *é minimizador global de* f*,se*

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

(b) *Um ponto* $\bar{x} \in \mathbb{R}^n$ *é minimizador local de* f*,se existe uma vizinhaça* U *de* $\bar{x}$ *tal que*

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

(c) *Seja* $\bar{z} \in [-\infty, +\infty)$ *definido por,*

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

é chamado de valor ótimo do problema (1.1).

(d) *O conjunto de nível da função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *associado a* $c \in \mathbb{R}$*, é o conjunto dado por*

$$L_f(c) = \{x \in \mathbb{R}^n \mid f(x) \leqslant c\}.$$

(e) *A função* f *é contínua no ponto* $\bar{x} \in \mathbb{R}^n$, *quando para toda sequência* $\{x^k\} \subset \mathbb{R}^n$ *tal que* $x^k \longrightarrow \bar{x}$, *tem-se*

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

A função f *é contínua quando ela é contínua em todos os pontos do* $\mathbb{R}^n$.

Observação: *A soma de funções contínuas é contínua.*

Teorema 1.1. *(Teorema de Weierstrass) Sejam* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função contínua e* C *um conjunto compacto não vazio. Então o problema de minimizar* f *em* C *tem solução global.*

Corolário 1.1. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *contínua. Se* $\exists\ c \in \mathbb{R}$ *tal que o conjunto de nível* $L_f(c)$ *seja não vazio e compacto. Então o problema de minimizar* f *tem uma solução global.*

Demonstração. De fato, do Teorema (1.1) existe $\bar{x} \in L_f(c)$ tal que $f(\bar{x}) \leqslant f(x), \quad \forall x \in L_f(c)$. Agora se $x \notin L_f(c)$ então $f(x) \geqslant c \geqslant f(\bar{x})$. Portanto existe um mínimo. □

Definiremos agora o conceito de função coerciva que será útil para mostrar a boa definição do algoritmo.

Definição 1.2. *A função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é coerciva se* $\lim\limits_{\|x\|\to\infty} f(x) = +\infty$.

Exemplo 1.1. *Dado* $x^0 \in \mathbb{R}^n$, *se* f *é limitada inferiormente. Então a função* $F : \mathbb{R}^n \longrightarrow \mathbb{R}$ *dada por*

$$F(x) := f(x) + \frac{1}{2}\| x - x_0 \|^2,$$

é coerciva .

De fato: como f *é limitada inferiormente existe um* $M \in \mathbb{R}$ *tal que* $M \leqslant f(x), \forall x \in \mathbb{R}^n$. *Assim*

$$\lim_{\|x\|\to+\infty} F(x) = \lim_{\|x\|\to+\infty} (f(x) + \frac{1}{2}\| x - x_0 \|^2) \geqslant \lim_{\|x\|\to+\infty} (M + \frac{1}{2}\| x - x_0 \|^2) = +\infty.$$

Portanto, F *é coerciva.*

O resultado que apresentaremos a seguir irá nos auxiliar a provar a boa definição do algoritmo.

Teorema 1.2. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função contínua e coerciva. Então, o problema (1.1) possui uma solução.*

Demonstração. Seja $\bar{x} \in \mathbb{R}^n$ e definamos o seguinte conjunto de nível

$$L_f f(\bar{x}) = \{x \in \mathbb{R}^n : f(x) \leqslant f(\bar{x})\}\ .$$

Note que $L_f f(\bar{x}) \neq \emptyset$, pois $\bar{x} \in L_f f(\bar{x})$. Se mostrarmos que $L_f f(\bar{x})$ é compacto, como f é contínua, o resultado seguirá do Corolário (1.1).

Suponha, por absurdo, que $L_f f(\bar{x})$ seja ilimitado. Desta forma existirá uma sequência $\{x^k\} \subset L_f f(\bar{x})$ tal que

$$\lim_{k\to\infty} \|x\| \ = +\infty.$$

Como $\{x^k\} \subset L_f f(\bar{x})$, então $f(x^k) \leqslant f(\bar{x})$. Da coercividade de f e de $\|x^k\| \to +\infty$ temos

$$f(\bar{x}) \geqslant f(x^k) \to \ +\infty,$$

o que é absurdo. Logo $L_f f(\bar{x})$ é limitado.

Agora suponha por absurdo que $L_f f(\bar{x})$ seja não fechado. Então vai existir uma seqência $\{x^k\} \subset L_f f(\bar{x})$ tal que $x^k \to \hat{x}$, mas $\hat{x} \notin L_f f(\bar{x})$, assim temos que $f(\hat{x}) > f(\bar{x})$. Por outro lado, dacontinuidade de f temos

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

daí $f(\hat{x}) \leqslant f(\bar{x})$, implicando que $\hat{x} \in L_f f(\bar{x})$. O que é absurdo. Portanto $L_f f(\bar{x})$ é fechado.

Desta forma concluímos que $L_f f(\bar{x})$ é compacto. Logo segue-se o resultado. □

Teorema 1.3. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função diferenciável no ponto* $\bar{x} \in \mathbb{R}^n$*. Se* $\bar{x}$ *é um minimizador local(global) de* f*, então,* $\nabla f(\bar{x}) = 0$.

Demonstração. Seja $z \in \mathbb{R}^n$ qualquer. Como $\bar{x}$ é um mínimo local, $\exists \varepsilon > 0$ tal que

$$f(\bar{x}) \leqslant f(\bar{x} + tz) \qquad \forall t \in [0, \varepsilon].$$

Da diferenciabilidade de f em $\bar{x}$,

$$f(\bar{x} + tz) = f(\bar{x}) + t\langle \nabla f(\bar{x}), z\rangle + o(t)$$

onde $\displaystyle\lim_{t\to 0^+} \frac{o(t)}{t} = 0$.

Dividindo por $t > 0$, temos que

$$0 \leqslant \langle \nabla f(\bar{x}), z\rangle + \frac{o(t)}{t},$$

e tomando o limite quando $t \to 0^+$, obtemos

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

Como $z \in \mathbb{R}^n$ é qualquer, podemos tomar $z = -\nabla f(\bar{x})$. Daí segue-se que

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

logo $\nabla f(\bar{x}) = 0$. □

Iremos agora apresentar o conceito de sequência Féjer convergente que irá nos ajudar mostrar que os pontos de acumulação da sequência gerada pelo algoritmo são pontos críticos da função f para o caso fortemente convexo.

Definição 1.3. *Uma sequência* $\left\{x^k\right\} \subset \mathbb{R}^n$ *é dida Féjer convergente ao conjunto não vazio* $U \subset \mathbb{R}^n$ *quando*

$$\parallel u - x^{k+1} \parallel \leqslant \parallel u - x^k \parallel \ , \quad \forall u \in U \text{ e } \forall \ k \in \mathbb{N}.$$

Exemplo 1.2. *Seja* $\{x^k\} = -\frac{1}{n}$, $\forall \ n \in \mathbb{N}$ *e* $U = [1, M]$ *com* $M > 1$. *A sequência* $\{x^k\}$ *é Féjer convergente ao conjunto* U.
De fato, dado $u \in U$ *temos* $\|u - x^{k+1}\| = \|u + \frac{1}{k+1}\|$, *sendo* $k + 1 > k \Rightarrow \dfrac{1}{k} > \dfrac{1}{k+1}$, *daí*

$$\|u + \frac{1}{k+1}\| < \|u + \frac{1}{k}\| \Rightarrow \|u - x^{k+1}\| < \|u - x^k\|.$$

Mostrando assim o resultado.

Proposição 1.1. *Se* $\left\{x^k\right\} \subset \mathbb{R}^n$ *é Féjer convergente ao conjunto não vazio* $U \subset \mathbb{R}^n$, *então* $\left\{x^k\right\}$ *é limitada. Além disso, se existe uma subsequência* $\left\{x^{k_j}\right\}$ *de* $\left\{x^k\right\}$ *tal que* $x^{k_j} \to x^* \in U$, *então* $x^k \to x^*$.

Demonstração. Da definição de sequência Féjer convergente, segue-se que

$$\parallel u - x^k \parallel \leqslant \ \|u - x^{k-1}\| \ \leqslant \ \ ... \ \ \leqslant \parallel u - x^0 \parallel \ , \quad \forall u \in U.$$

Logo $\left\{x^k\right\}$ está contida na bola de centro u e raio $\parallel u - x^0 \parallel$, assim a sequência é limitada. Agora seja $\left\{x^{k_j}\right\}$ uma subsequência de $\left\{x^k\right\}$ tal que $x^{k_j} \to x^* \in U$. Como $x^* \in U$ temos que a sequência $\parallel x^* - x^k \parallel$ é decrescnte e não negativa. Além disso a subsequência $\parallel x^* - x^{k_j} \parallel \to 0$. Daí $\parallel x^* - x^k \parallel \to 0$, portanto $x^k \to x^*$. □

1.2 Conjuntos Convexos e Funções Convexas

Apresentaremos agora os conceitos e resultados básicos de conjuntos convexos e funções convexas.

Definição 1.4. *Um conjunto* $C \subset \mathbb{R}^n$ *é dito convexo se para todo* $x, y \in C$ *e* $\alpha \in [0,1]$, *tem-se* $\alpha x + (1-\alpha)\, y \in C$.

Exemplo 1.3. *Os conjuntos vazio, unitário e o* $\mathbb{R}^n$ *são conjuntos trivialmente convexos.*

Exemplo 1.4. *O hiperplano* $H = \{x \in \mathbb{R}^n : \langle a, x\rangle \leqslant c\}$, *onde* $a \in \mathbb{R}^n - \{0\}, c \in \mathbb{R}$ *é convexo.*
De fato, seja $x, y \in R^n$ e $\alpha \in (0,1)$, *então*

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

portanto H *é convexo.*

Proposição 1.2. *Sejam* $D_i \subset \mathbb{R}^n,\ i \in I$, *conjuntos convexos, onde* I *é um conjunto de índices qualquer. Então a interseção* $D = \cap_{i\in I} D_i$ *também é um conjunto convexo.*

Demonstração. Sejam $x, y \in D$. Pela definição de interseção, $x, y \in D_i$. Como os conjuntos D_i, com $i \in I$, são convexos, segue-se que $\alpha x + (1-\alpha)y \in D_i$ para qualquer $\alpha \in [0,1]$ e todo $i \in I$. Lodo $\alpha x + (1-\alpha)y \in D$, portanto D é convexo. □

Definição 1.5. *Uma função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *diz-se convexa quando para todo* $x, y \in C$ *e* $\alpha \in [0,1]$, *tem-se*

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

A função f diz-se *estritamente convexa* se a desigualdade acima é estrita para todo $x \neq y$ e $\alpha \in (0,1)$. A função f diz-se *fortemente convexa com módulo* $\gamma > 0$, quando $x, y \in C$ e $\alpha \in [0,1]$, temos

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

Observação: Se uma função f é fortemente convexa implica que ela é estritamente convexa, e assim convexa. A recíproca nem sempre é verdadeira.

Exemplo 1.5. *A função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *dada por* $f(x) = \langle x, \bar{x}\rangle$*, com* $\bar{x} \in \mathbb{R}^n$ *fixado é convexa.*

De fato, seja $x, y \in \mathbb{R}^n$ e $\alpha \in (0,1)$*, então*

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

portanto f *é convexa.*

Teorema 1.4. *A função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é convexa se, somente se, seu epígrafo dado por*

$$\mathrm{Epi}(f) = \{(x, \alpha) \in \mathbb{R}^{n+1} \mid f(x) \leqslant \alpha\}$$

é convexo.

Demonstração. Ver [8]. □

O resultado a seguir embora simples de se vereficar será muito útil para mostrar a boa definição do algoritmo.

Proposição 1.3. *Sejam* $g, h : \mathbb{R}^n \longrightarrow \mathbb{R}$ *tais que* g *é convexa e* h *estritamente convexa, então a função* $g + h$ *é estritamente convexa.*

Demonstração. De fato, seja $f(x) = g(x)+h(x)$, com g convexa e h estritamente convexa. Assim dados $x, y \in \mathbb{R}^n$ e $\alpha \in [0,1]$. temos

$$\begin{aligned} g(\alpha x + (1-\alpha y)) &\leqslant \alpha g(x) + (1-\alpha) g(y) \\ h(\alpha x + (1-\alpha y)) &< \alpha h(x) + (1-\alpha) h(y), \end{aligned}$$

daí somando $g + h$ podemos concluir que $f(\alpha x + (1-\alpha y)) < \alpha f(x) + (1-\alpha) f(y)$.

□

Teorema 1.5. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então todo minimizador local é glogal. Além disso, o conjunto dos minimizadores é convexo. E se* f *é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração. Suponha $\bar{x} \in \mathbb{R}^n$ um minimizador local mas não global. Então existe um $y \in \mathbb{R}^n$ tal que $f(y) < f(\bar{x})$. Definamos $x(\alpha) = \alpha y + (1-\alpha)\bar{x}$, agora pela convexidade de f, para todo $\alpha \in (0,1]$, tem-se

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

Tomando $\alpha > 0$ suficientemente pequeno, nos garante que $x(\alpha)$ é arbitrariamente próximo de $\bar{x}$, e que $f(x(\alpha)) < f(\bar{x})$ e $x(\alpha) \in \mathbb{R}^n$. Contrariando o fato de $\bar{x}$ ser mínimo local.

Agora tomando $S \subset \mathbb{R}^n$ o conjunto dos minimizadores globais e $\bar{v} \in \mathbb{R}$ o valor ótimo de f. Assim $f(x) = \bar{v}, \quad \forall x \in S$. Tomando $x, \bar{x} \in S$ e $\alpha \in [0, 1]$ temos que

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

assim $f(\alpha x + (1-\alpha)\bar{x}) = \bar{v}$, logo $(\alpha x + (1-\alpha)\bar{x}) \in S$, consequentemente S é convexo. Supondo agora f estritamente convexa e que $x, \bar{x} \in S$, $x \neq \bar{x}$ minimizadores. Seja $\alpha \in (0, 1)$. Assim segue-se $f(\alpha x + (1-\alpha)\bar{x}) \geqslant f(x) = f(\bar{x}) = \bar{v}$. Mas pela convexidade estrita temos

$$\begin{aligned} f(\alpha x + (1-\alpha)\bar{x}) &< \alpha f(x) + (1-\alpha) f(\bar{x}) \\ &= \alpha\bar{v} + (1-\alpha)\bar{v} = \bar{v}, \end{aligned}$$

o que gera uma contradição. Portanto o minimizador é único. □

Teorema 1.6. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então* f *é contínua.*

Demonstração. Ver [7] pág.136. □

1.3 Funções Convexas Diferenciáveis

A seguir veremos alguns resultados que nos será útil na determinação de funções convexas diferenciáveis.

Teorema 1.7. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *convexa e diferenciável. Então as propriedades abaixo são equivalentes:*

(a) *A função* f *é convexa.*

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

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

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

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

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

(d) *A matriz Hessiana de f é semidefinida positiva, isto é*

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

Demonstração. Primeiramente veremos que (a)⇒(b)⇒(c).

Suponha f é convexa, assim $\forall x, y \in \mathbb{R}^n$ e $\alpha \in (0,1]$, definindo $d = y - x$, temos

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

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

Dividindo os dois lados da por $\alpha > 0$, e fazendo $\alpha \to 0^+$, obtemos que

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

Assim, obtemos a primeira parte. Agora substituindo x por y no item (b), temos que

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

somando a desigualdade com o ítem (b), concluimos a veracidade de (c).

Vamos mostrar agora que (c)⇒(b)⇒(a).

Sejam $x, y \in \mathbb{R}^n$. Pelo Teorema do Valor Médio, existe$\alpha \in (0,1)$ tal que

$$f(y) - f(x) = \langle \nabla f(x + \alpha(y-x)), y - x\rangle \tag{1.2}$$

Usando (c) para os pontos $(x + \alpha(y-x))$ e x, temos

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

Combinando esta desigualdade com (1.2), obtemos (b).

Definindo $d = y - x$, temos

$$f(x) \geqslant f(x+\alpha d) - \alpha\langle \nabla f(x+\alpha d), d\rangle, \tag{1.3}$$

$$f(y) \geqslant f(x+\alpha d) + (1-\alpha)\langle \nabla f(x+\alpha d), d\rangle, \tag{1.4}$$

onde usando (b) para os pontos x e $x+\alpha d$; y e $x+\alpha d$, respectivamente. Multiplicando (1.3) por $1-\alpha \geqslant 0$ e (1.4) por $\alpha \geqslant 0$, e somando, obtemos

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

logo f é convexa.

Suponhamos agora que f seja duas vezes diferenciável. Mostraremos que $(b) \Longleftrightarrow (d)$.

Seja $x, d \in \mathbb{R}^n$ com x fixo. Tomomeos $x + \alpha d$ para todo $\alpha > 0$ suficientemente pequeno. Por (b),

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

Da diferenciabilidade de f,

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

Dividindo por α^2 e tomando o limite quando $\alpha \to 0+$, obtemos (d).

Seja $x, y \in \mathbb{R}^n$ quaisquer. Pelo Teorema do Valor Médio, existe $\alpha \in (0, 1)$ tal que

$$\begin{aligned} & f(x) - f(y) - \langle f'(x), y - x \rangle \\ = \; & \frac{1}{2} \langle f''(x + \alpha(y - x))(y - x), y - x \rangle \geqslant 0, \end{aligned}$$

onde a desigualdade segue de (d). Portanto, $(d) \Longleftrightarrow (b)$. $\square$

Para funções diferenciáveis fortemente convexas, temos critérios que são análogos do Teorema 1.7 para funções convexas.

Teorema 1.8. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *convexa e diferenciável, com derivada contínua. Então as propriedades abaixo são equivalentes:*

(a) *A função* f *é fortemente convexa com módulo* $\gamma > 0$.

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

$$f(y) \geqslant f(x) + \langle \nabla f(x), y - x \rangle + \gamma \parallel y - x \parallel^2 .$$

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

$$\langle \nabla f(y) - \nabla f(x), y - x \rangle \geqslant 2\gamma \parallel y - x \parallel^2 .$$

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

(d) *A matriz Hessiana de f é definida positiva uniformemente, isto é,*

$$\langle f''(x) d, d \rangle \geqslant 2\gamma \|d\|^2 \qquad \forall x, d \in \mathbb{R}^n$$

Demonstração. Segue-se de forma similar da demonstração do Teorema 1.7 $\square$

Exemplo 1.6. *A função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *dado por* $f(x) = \| x - x_0 \|^2, x_0 \in \mathbb{R}^n$ *é fortemente convexa, consequentemente estritamente convexa.*

De fato: basta usarmos o ítem (d) do Teorema 1.8.

Exemplo 1.7. *A função* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *dada por,* $f(x) = |x| = \max\{x, -x\}$ *é convexa mas não diferenciável em* $x = 0$*. Entretando, a mesma possui propriedades muito similares a funções convexas diferenciáveis.*

1.4 Funções Convexas não Diferenciáveis

Veremos um pouco de funções convexas não diferenciáveis, haja vista que a diferenciabilidade nem sempre é preservada na convexidade, como no caso do supremo de funções convexas.

Os dois resultados apresentados a seguir serão muito útil para verificar algumas propriedades essênciais do subdiferencial de uma função.

Teorema 1.9. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então para todo* $x \in \mathbb{R}^n$*,* f *possui derivada direcional em cada direção* $d \in \mathbb{R}^n$*. Além disso,*

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

onde $f'(x; d)$ *representa a derivada direcional de* f*.*

Demonstração. Em [7] pág. 159. □

Proposição 1.4. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então para quaisquer duas sequências* $x^k \longrightarrow x, d^k \longrightarrow d$*, tem-se que* $\lim \sup_{k \to \infty} f'\left(x^k, d^k\right) \leqslant f'(x, d)$.

Demonstração. Em [7] pág. 161. □

Definição 1.6. *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*

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

O conjunto de todos os subgradientes de f em x é chamado de *subdiferencial* de f em x, e denotaremos por

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

Observação: No caso em que tomarmos $z \in \mathbb{R}^n$ como $z = x + \alpha d$, onde $d \in \mathbb{R}^n, \alpha > 0$, podemos definir o subgradiente como sendo

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

daí usando o Teorema 1.9 e fazendo $\alpha \to 0^+$ temos que

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

Assim as seguintes definições equivalentes:

$$\begin{aligned} \partial f(x) &= \{y \in \mathbb{R}^n \mid f(z) \geqslant f(x) + \langle y, z - x \rangle \qquad \forall z \in \mathbb{R}^n\} \\ &= \{y \in \mathbb{R}^n \mid f'(x; d) \geqslant \langle y, d \rangle \qquad \forall d \in \mathbb{R}^n\}, \end{aligned} \tag{1.5}$$

Um fato muito importante sobre o subdiferencial na teoria de otimização é o que apresentaremos agora.

Proposição 1.5. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então* $\bar{x}$ *é um minimizador de* f *se, somente se,* $0 \in \partial f(\bar{x})$.

Demonstração. De fato, seja $\bar{x}$ um minimizador de f e $z \in \mathbb{R}^n$ qualquer, então

$$f(\bar{x}) \leqslant f(z) \quad \Leftrightarrow \langle 0, z - \bar{x} \rangle + f(\bar{x}) \leqslant f(z),$$

garantindo assim o resultado desejado. □

Mostraremos que para uma função convexa definida em todo o espaço, seu subdiferencial é compacto, convexo e não vazio. O fato de ser não vázio irá ajudar na demonstração do algoritmo.

Teorema 1.10. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa. Então* $\forall x \in \mathbb{R}^n$*, o conjunto* $\partial f(x)$ *é convexo, compacto e não vazio.*

Demonstração. Como

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

é uma interseção de semiespaços fechados e convexos, portanto é fechado e convexo. Vamos mostrar agora a limitação, caso $\partial f(x) = \emptyset$, então ele é limidado trivialmente. Suponha que $\partial f(x)$ não seja limitado, logo existirá $\{y^k\} \subset \partial f(x)$ tal que $\| y^k \| \longrightarrow \infty$. Para todo k tomemos $d^k = \dfrac{y^k}{\| y^k \|}$. Sem perda de generalidade, podemos admitir que $d^k \longrightarrow d$. Assim obtemos de (1.5) que

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

portanto

$$\limsup_{k\to\infty} \|y^k\| \leqslant \limsup_{k\to\infty} f'\left(x; d^k\right) \leqslant f'(x; d) < +\infty,$$

onde a segunda desigualdade segue da Proposição 1.4, contradizendo o fato de que $\|y^k\| \longrightarrow \infty$, portanto $\partial f(x)$ é limitado. Finalmente mostraremos que $\partial f(x) \neq \emptyset$. De fato, de (1.5) podemos concluir que

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

Suponhamos que

$$f'(x; d) > \langle y, d \rangle \qquad \forall y \in \partial f(x).$$

Definamos os seguintes conjuntos:

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

$$D_2 := \left\{ (z, c) \in \mathbb{R} \times \mathbb{R} \mid \begin{array}{c} c = f(x) + \alpha f'(x; d), \\ z = x + \alpha d, \quad \alpha \in \mathbb{R}_+ \end{array} \right\}.$$

É claro que D_2 é convexo. O conjunto D_1 é o epigrafo de f sem sua fronteira, logo também convexo.

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

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

para algum $\alpha \in \mathbb{R}_+$, em contradição com o Teorema 1.9. Portanto $D_1 \cap D_2 = \emptyset$.

Pelo Teorema da Separação (Ver [7] pág. 104), existe um certo $(u, \beta) \in \mathbb{R}^n \times \mathbb{R} \setminus \{0\}$ tal que

$$\begin{aligned} \langle u, z \rangle + \beta c \quad &\leqslant \quad \langle u, x + \alpha d \rangle + \beta \left(f(x) + \alpha f'(x; d)\right) \\ &\forall\ z \in \mathbb{R}^n, \forall\ \alpha \in \mathbb{R}_+, \forall\ c > f(z). \end{aligned} \tag{1.6}$$

Se $\beta = 0$, teríamos que

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

o que só ocorre se $u = 0$. Como $\langle u, \beta \rangle \neq 0$, conclui-se que $\beta \neq 0$.
Supondo $\beta > 0$. Então escolhendo $z = x$ e $\alpha = 0$ em (1.6), temos que $\beta c \leqslant \beta f(x)$ para todos os $c > f(x)$, o que é uma contradição.
Assim $\beta < 0$. Dividindo os dois lados em (1.6) por $\beta < 0$, obtemos

$$\begin{array}{c} c + \langle u/\beta, z - x \rangle \geqslant f(x) + \alpha f'(x; d) + \alpha \langle u/\beta, d \rangle \\ \forall z \in \mathbb{R}^n \text{ e } \forall \alpha \in \mathbb{R}_+, \forall c > f(z). \end{array}.$$

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

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

ou seja, $-u/\beta \in \partial f(x)$, pela definição de subgradiente. Em particular, $\partial f(x) \neq \emptyset$. □

Para o caso da função ser diferenciável, o seu subdiferencial se reduz a um único ponto. E usaremos este fato para mostrar a convergência linear do método para o caso fortemente convexo.

Proposição 1.6. *Uma função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é diferenciável no ponto* $x \in \mathbb{R}^n$ *se, somente se, o conjunto* $\partial f(x)$ *contém um só elemento. Neste caso,* $\partial f(x) = \{\nabla f(x)\}$.

Demonstração. Seja f diferenciável no ponto $x \in \mathbb{R}^n$. Pelo Teorema 1.7 temos:

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

ou seja, $\nabla f(x) \in \partial f(x)$
Agora tomemos $y \in \partial f(x)$. Para todo $d \in \mathbb{R}^n$,

$$\begin{aligned} f(z) + \langle y, d \rangle &\leqslant f(x + d) \\ &= f(x) + \langle \nabla f(x), d \rangle + o(\| d \|) \end{aligned},$$

logo

$$\langle y - \nabla f(x), d \rangle \leqslant o(\| d \|),$$

tomando

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

temos que $d^k \longrightarrow 0 (k \to \infty)$ e

$$\| y - \nabla f(x) \| / k \leqslant o(\| 1/k \|),$$

o que só ocorre quando $y - \nabla f(x) = 0$. Consequentemente $\partial f(x) = \{\nabla f(x)\}$.

Seja $\partial f(x) = \{y\}$. Pelo Teorema 1.10

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

Tomando d os elementos da base canônica do $\mathbb{R}^n$, tem-se que $y_i = \partial f(x) / \partial x_i, i = 1, ...n$. Temos que $f'(x, d)$ é uma função linear em d de forma $\langle \nabla f(x), d \rangle$, o que implica a diferenciabilidade de f em x. □

Definição 1.7. *Um operador ponto-conjunto* $T : \mathbb{R}^n \longrightarrow P(\mathbb{R}^n)$ *é dito monótono (fortemente monótomo módulo* $\rho > 0$ *) quando para quaisquer* $x, y \in \mathbb{R}^n, u \in T(x)$ e $v \in T(y)$ *tem-se que*

$$\langle u - v, x - y \rangle \geqslant 0 \quad \left(\langle u - v, x - y \rangle \geqslant \rho \| x - y \|^2\right).$$

Teorema 1.11. *Seja* $\partial f : \mathbb{R}^n \longrightarrow P(\mathbb{R}^n)$*, onde* f *é uma função convexa. Então* ∂f *é um operador ponto-conjunto monótono.*

Demonstração. De fato, seja $x, y \in \mathbb{R}^n$ e $\xi \in \partial f(x)$ e $\eta \in \partial f(y)$. Da definição de subdiferencial temos

$$\langle \xi, y - x \rangle \leqslant f(y) - f(x) \text{ e } - \langle \eta, y - x \rangle = \langle \eta, x - y \rangle \leqslant f(x) - f(y),$$

somando as desigualdades temos

$$\langle \xi, y - x \rangle - \langle \eta, y - x \rangle \leqslant 0,$$

logo

$$\langle \xi - \eta, y - x \rangle \leqslant 0 \Rightarrow \langle \xi - \eta, x - y \rangle \geqslant 0.$$

Portanto, ∂f é um operador monótono. □

De forma semelhante verifica-se que se f é fortemente convexa, então ∂f é fortemente monótono.

A seguir mostraremos que o subdiferencial de uma função convexa é limitado em cunjuntos limitados e que tem certas propriedades de continuidade no seguinte sentido.

Proposição 1.7. (Continuidade do Subdiferencial) *Sejam* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função convexa e* $\{x^k\}$ *limitada com* $y^k \in \partial f(x^k)$ *para todo* k*, então a sequência* $\{y^k\}$ *é limitada. Além disso, se* $x^k \to x$*, então todos os seus pontos de acumulação de* $\{y^k\}$ *pertencem a* $\partial f(x)$.

Demonstração. Suponha que $\{x^k\}$ seja limitada e $\{y^k\} \subset \mathbb{R}^n$ tais que $y^k \in \partial f\left(x^k\right)$ para todo k e $\|y^k\| \to \infty (k \to \infty)$. Definindo $d^k = \dfrac{y^k}{\|y^k\|}$ e tomando subsequências se necessário, podemos admitir que $x^k \to x$ e $d^k \to d(k \to \infty)$, sendo $\|d\| \to 1$. Do Teorema (1.10),

$$f'(x^k, d^k) \geqslant \langle y^k, d^k \rangle = \|y^k\| \to +\infty.$$

Por outro lado, pela Proposição (1.4),

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

O que gera uma contradição, portanto $\partial f(x^k)$ é limitado e consequentemente $\{y^k\}$.

Como $y^k \in \partial f\left(x^k\right)$, pela relação (1.5) temos que

$$f'\left(x^k; d\right) \geqslant \langle y^k, d \rangle \qquad \forall d \in \mathbb{R}^n.$$

Seja $y \in \mathbb{R}^n$ qualquer ponto de acumulação de $\left\{y^k\right\}$, isto é, existe $y^{k_j} \longrightarrow y$. Tomando o limite superior com $j \longrightarrow \infty$ na relação acima, temos que

$$\langle y, d \rangle \leqslant \limsup_{k \longrightarrow \infty} f'\left(x^k, d\right) \leqslant f'(x, d) \qquad \forall d \in \mathbb{R}^n,$$

daí novamente da relação (1.5) conclui-se que $y \in \partial f(x)$. □

Proposição 1.8. *Sejam* $f_i : \mathbb{R}^n \longrightarrow \mathbb{R}, i = 1, 2, ..., p$ *funções convexas. Então*

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

Demonstração. É suficiente provar o resultado para $p = 2$. Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, onde $f(x) = f_1(x) + f_2(x)$.

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

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

Somando as duas desigualdades, obtemos

$$\begin{aligned} f(z) &= f_1(z) + f_2(z) \\ &\geqslant f_1(x) + f_2(x) + \langle y^1 + y^2, z - x \rangle \\ &= f(x) + \langle y^1 + y^2, z - x \rangle, \end{aligned}$$

o que verifica que $y^1 + y^2 \in \partial f(x)$. Portanto

$$\partial f_1(x) + \partial f_2(x) \subset \partial f(x).$$

Suponhamos agora que $\partial f(x) \not\subset \partial f_1(x) + \partial f_2(x)$, isto é, que exita um $y \in \partial f(x)$ tal que $y \notin \partial f_1(x) + \partial f_2(x)$. Do Teorema 1.10 temos que os conjuntos $\partial f_1(x)$ e $\partial f_2(x)$ são convexos, compactos e não vazios. Portanto $\partial f_1(x) + \partial f_2(x)$ é convexo, compacto e não vazio. Como $y \notin \partial f_1(x) + \partial f_2(x)$, pelo Teorema da Separação Estrita (ver [7] pág. 105), existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tal que

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

Portanto, usando o Teorema 1.10, obtemos que

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

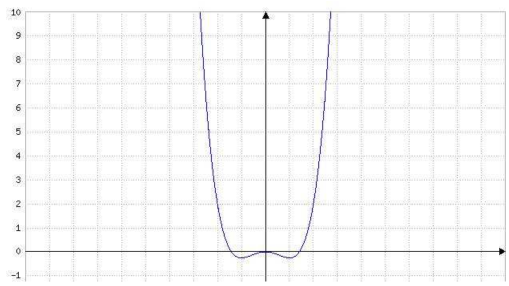
Mas esta desigualdade contradiz o fato de $y \in \partial f(x)$. □

1.5 Otimização DC

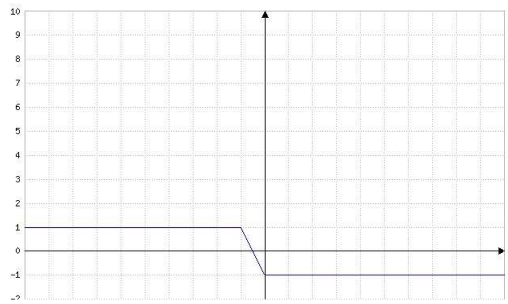
A Otimização DC é um campo interessante e importante em otimização não convexa. Isto porque vários problemas de otimização podem ser reescritos como problemas de otimização DC. Além disso, um grande número de funções podem ser representadas como a diferença de funções convexas pois um resultado muito importante é o fato de que toda função de classe C^2 tem uma representação DC, embora não haja uma representação geral desta forma, bastando exibir uma representação DC. Estas funções são bastantes interessante e importante da otimização não convexa, pelos seus aspectos teóricos e também pelo grande número de aplicações.

A seguir mostraremos alguns exemplos de funções DC.

Exemplo 1.8. *Considere a função* $f : \mathbb{R} \longrightarrow \mathbb{R}$ *dada por,* $f(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2$. *Note que* f *é uma função DC, onde* $g(x) = \frac{1}{4}x^4$ *e* $h(x) = \frac{1}{2}x^2$.
E também $f'(x) = x^3 - x = 0 \Leftrightarrow x = 0$ ou $x = \pm 1$, *logo estes serão seus pontos críticos. Além disso* f *não é convexa, pois nem sempre* $f''(x) = 3x^2 - 1 \geqslant 0$. *Como podemos verificar este fato observando o seu gráfico (epígrafo).*



Exemplo 1.9. *Seja* $f_1, f_2 : \mathbb{R} \longrightarrow \mathbb{R}$, *com* $f_1(x) = \mid x \mid$ *e* $f_2(x) = \mid x + 1 \mid$. *Temos que* $f_1 - f_2$ *também não é convexa, embora suas componentes sejam, como vemos em seu gráfico (epígrafo) logo abaixo*



Definição 1.8. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *uma função DC. Um ponto* $x^* \in \mathbb{R}^n$ *é dito crítico de* f *se* $\partial h(x^*) \cap \partial g(x^*) \neq \emptyset$.

No caso de termos $\partial f(x^*) \subset \partial g(x^*) - \partial h(x^*)$, temos que x^* é um mínimo global.

Nas secções seguintes serão apresentadas algumas propriedades das funções DC.

1.5.1 O Espaço das funções DC

Veremos algumas propriedades básicas de funções DC e mostrar que elas são fechadas para algumas operações usuais em otimização. É fácil ver que o espaço das funções convexas está contido no espaço das funções DC.

Teorema 1.12. *Seja* f, f_i $(i = 1, 2, ..., m)$ *funções DC. Então as seguintes funções também são DC:*

(a) $\sum_{i=1}^{m} \lambda_i f_i (x)$ *para algum número real* λ_i;

(b) $max_{i=1,...,m} f_i (x)$ *e* $min_{i=1,...,m} f_i (x)$

(c) $\mid f (x) \mid, f^+ := max \{0, f_i (x)\}, f^- := min \{0, f_i (x)\}$;

(d) *o produto* $\prod_{i=1}^{m} f_i (x)$.

Demonstração. (a). É imediato.

(b). Seja $f_i = g_i - h_i, (i = 1, ..., m)$ uma decomposição de funções DC de f_i. Então:

$$f_i = g_i + \sum_{j=1, j\neq i}^{m} h_j - \sum_{j=1}^{m} h_j \ ,$$

daí,

$$max\ f_i = max\left\{g_i + \sum_{j=1, j\neq i}^{m} h_j\right\} - \sum_{j=1}^{m} h_j$$

que ainda é uma função DC, desde que o máximo, assim como a soma de um número finito de funções convexas sejam funções convexas. De forma análoga se mostra que o $min\ f_i$ também o é.

(c).Como $f = g - h$ com g, h convexas. Tomemos $\mid f \mid = 2max\{g, h\} - (g + h)$ que é uma decomposição DC de $\mid f \mid$. Assim os casos de $f^+ e\ f^-$ segue-se diretamente do ítem (b).

Para mostrar (d) vejamos as funções como sendo $f_i = f^+ + f^-$, do ponto de vista do ítem (c) pode-se reduzir a encontrar uma decomposição DC para um produto de funções DC não negativa, $f_i = g_i - h_i$ com $g_i e\ h_i$ não negativas e convexas, para i =(1, 2). Então,

$$f_1.f_2 = (g_1 - h_1) . (g_2 - h_2) =$$
$$= \tfrac{1}{2}\left[(h_1 + h_2)^2 + (g_1 + g_2)^2\right] - \tfrac{1}{2}\left[(g_1 + h_2)^2 + (g_2 + h_2)^2\right]$$

é uma decomposição DC de $f_1.f_2$, desde que o quadrado de uma função nao negativa convexas seja convexa. O rersultado de $\prod_{i=1}^{m} f_i\,(x)$ segue-se por indução. ☐

Definição 1.9. *Uma função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *é localmente DC em* x_0 *se existe uma vizinhaça* V_{x_0} *onde* f *é DC em* V_{x_0}*.*

Proposição 1.9. *Toda função localmente DC é DC.*

Demonstração. Ver Harmart [2]. ☐

Outra classe interessante de funções DC é a classe das funções C^2.

Proposição 1.10. *Toda função* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *de classe* C^2 *é DC.*

Demonstração. Os elementos da Hessiana $\nabla^2 f$ de f são limitados em toda vizinhança fechada de $B\,[x_0, \varepsilon]$. Logo, para $\mu > 0$ suficientment grande, a função $f\,(x) + \mu\| \ x \ \|^2$ é convexa em $B\,[x_0, \varepsilon]$, desde que a Hessiana $\nabla^2 f + 2\mu I$ é semidefinida positiva em $B\,(x_0, \varepsilon)$ para μ suficientemente grande. Então $\left(f\,(x) + \mu\| \ x \ \|^2\right) - \mu\| \ x \ \|^2$ é DC em $B\,[x_0, \varepsilon]$. Daí pelo Proposição 1.9 f é DC. ☐

Outros exemplos de funções DC são as funções do tipo $C^{1,1}$(funções diferenciáveis cujo o gradiente é Lipschitz) e as lunções lower-C^2. E também temos que o espaço das funções DC é denso no espaço das funções contínuas em compacto, (ver [12]).

Proposição 1.11. *Seja* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *DC e* $g : \mathbb{R} \longrightarrow \mathbb{R}$ *convexa. Então a composta* $g \circ f$ *é DC.*

Demonstração. Ver [1] □

Teorema 1.13. *O quadrado da função distância de um ponto* $x \in \mathbb{R}^n$ *à um subconjunto fechado não vazio* $S \subset \mathbb{R}^n$*, denotado por* $d_S^2(x)$ *é DC.*

Demonstração. Temos que

$$\begin{aligned} d_S^2(x) &= \inf\left\{\| x - y \|^2 : y \in S\right\} \\ &= \| x \|^2 + \inf\left\{- \| x \|^2 + \| x - y \|^2 : y \in S\right\} \\ &= \| x \|^2 - \sup\left\{\| x \|^2 - \| x - y \|^2 : y \in S\right\} \\ &= \| x \|^2 - \sup\left\{2\langle x, y\rangle - \| y \|^2 : y \in S\right\}. \end{aligned}$$

A norma de $g(x) = \| x \|^2$ é convexa e a função $h(x) := \sup\left\{2\langle x, y\rangle - \| y \|^2 : y \in S\right\}$ é o supremo local de uma família de funções afins, logo é convexa. Portanto, $d_S^2(x)$ é DC. □

Capítulo 2

Algoritmo de Ponto Proximal

Neste capítulo apresentaremos um algoritimo proximal para minimizar a classe de funções DC, onde dado um ponto $x_0 \in \mathbb{R}^n$ encontra-se uma sequência $\{x^n\} \subset \mathbb{R}^n$ a qual se ela for limitada os pontos de acumulação são os crítico da função objetivo. Os resultados apresentados neste capítulo foram demonstrados em [12, 13] para o caso das funções DC $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$.

2.1 O Algoritmo

Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma função DC limitada inferiormente, onde $f(x) = g(x) - h(x)$.

Algoritmo 1. *Passo 1: Dado um ponto* $x^0 \in \mathbb{R}^n$ *e uma sequência de números positivos limitada* λ_k *tal que* $\liminf \lambda_k > 0$. *Para* $k = 1, 2, ...,$

Passo 2: Calcular

$$w^k \in \partial h\left(x^k\right) \tag{2.1}$$

Passo 3: Calcular

$$x^{k+1} = \arg\min_{x \in \mathbb{R}^n} \left\{ g\left(x\right) - \langle w^k, x - x^k \rangle + \frac{1}{2\lambda_k} \| \, x - x^k \, \|^2 \right\} \tag{2.2}$$

Passo 4: Se $x^k = x^{k+1}$, *pare. Caso contrário, defina* $k := k + 1$ *e retorne ao passo 2.*

Observação*: Quando* $h\left(x\right) \equiv 0$, *teremos exatamente o algoritmo ponto proximal clássico para funções convexas.*

2.2 Boa Definição do Algoritmo

Mostraremos que o algoritmo proposto está bem definido.

Teorema 2.1. *O Algoritmo 1 está bem definido.*

Demonstração. Provaremos por indução sobre k. Para $k = 0$ é verdadeiro. Suponha válido para k, ou seja, existe x^k, temos também que existe

$$w^k \in \partial h\left(x^k\right),$$

pois o subdiferencial de uma função convexa é não vazio. Seja

$$f_k\left(x\right) = g\left(x\right) - \langle w^k, x - x^k\rangle + \frac{1}{2\lambda_k}\| x - x^k \|^2,$$

como $f\left(x\right)$ é limitada inferiormente existe $M \in \mathbb{R}$ tal que $M \leqslant f\left(x\right)$. Da convexidade de h, temos

$$\begin{aligned} h(x) &\geqslant h(x^k) + \langle w^k, x - x^k\rangle \\ -h(x) &\leqslant -h(x^k) - \langle w^k, x - x^k\rangle, \end{aligned}$$

assim $M \leqslant f(x) = g(x) - h(x) \leqslant g(x) - h(x^k) - \langle w^k, x - x^k\rangle$. Somando $\dfrac{1}{2\lambda_k}\| x - x^k \|^2$ em ambos os lados da desigualdade teremos,

$$f_k\left(x\right) \geqslant M + h(x^k) + \frac{1}{2\lambda_k}\| x - x^k \|^2,$$

logo garantimos que

$$\lim_{\|x\|\to\infty} f_k\left(x\right) = \infty.$$

Portanto, sendo $f_k\left(x\right)$ contínua e coerciva, então pelo Teorema 1.2 garantimos a existência de um mínimo. Sendo $\langle w^k, \cdot - x^k\rangle$, $g(\cdot)$ convexas e $\| \cdot - x^k \|^2$ estritamente convexa temos que $f_k\left(x\right)$ é estritamente convexa, logo pelo Teorema 1.5 o minímo é único, ou seja, existe um único x^{k+1}. Assim, $\{x^k\}$ está bem definido. □

2.3 Análise de Convergência

Apresentaremos agora os resultados de convergência do algoritmo.

Proposição 2.1. (Criterio de Parada) *Se $x^k = x^{k+1}$, então x^{k+1} é um ponto crítico de f.*

Demonstração. De (2.2) temos que

$$0 \in \partial[g(x^{k+1}) - w^k + \frac{1}{\lambda_k}(x^{k+1} - x^k)],$$

logo

$$w^k \in \partial[g(x^{k+1}) + \frac{1}{\lambda_k}(x^{k+1} - x^k)], \tag{2.3}$$

assim se $x^k = x^{k+1}$, segue-se de (2.1) e (2.3) que $w^k \in \partial h\left(x^k\right) \cap \partial g\left(x^k\right)$, daí x^{k+1} é um ponto crítico de f. $\square$

Apartir de agora, vamos considerar que $x^k \neq x^{k+1}, \quad \forall\, k \in \mathbb{N}$.

Teorema 2.2. *A sequência* $\left\{f(x^k)\right\}$ *é estritamente decrescente.*

Demonstração. De (2.1) e da convexidade h temos que

$$h\left(x^{k+1}\right) \geqslant h\left(x^k\right) + \langle w^k, x^{k+1} - x^k\rangle, \tag{2.4}$$

por outro lado de (2.2) temos

$$g\left(x^k\right) \geqslant g\left(x^{k+1}\right) - \langle w^k, x^{k+1} - x^k\rangle + \frac{1}{2\lambda_k}\| \, x^{k+1} - x^k \, \|^2 \ . \tag{2.5}$$

Somando as desigualdades (2.4) e (2.5), obtemos

$$f\left(x^k\right) \geqslant f\left(x^{k+1}\right) + \frac{1}{2\lambda_k}\| \, x^{k+1} - x^k \, \|^2 \ , \tag{2.6}$$

como λ_k é uma sequência de termos positivos e $x^k \neq x^{k+1}$, segue-se que $f\left(x^{k+1}\right) < f\left(x^k\right)$. $\square$

Corolário 2.1. *Seja* $\left\{x^k\right\}$ *gerada pelo Algoritmo . Então a sequência* $\left\{f\left(x^k\right)\right\}$ *é convergente. Além disso, se* $\left\{x^k\right\}$ *é limitada, então* $f\left(x^k\right) \to f\left(\bar{x}\right)$*, onde* $\bar{x}$ *é algum ponto de acumulação de* $\left\{x^k\right\}$*.*

Demonstração. Do Teorema 2.2 temos que $\left\{f\left(x^k\right)\right\}$ é monótona decrescente, e por f ser limitada inferiormente logo convergirá. Sendo $\left\{x^k\right\}$ limitada ela será convergente a menos de uma subsequência, e como f é contínua segue-se que $f\left(x^k\right) \to f\left(\bar{x}\right)$. $\square$

Proposição 2.2. *Seja* $\left\{x^k\right\}$ *gerada pelo Algoritmo . Então* $\lim_{k\to\infty} \| \, x^{k+1} - x^k \, \| = 0$.

Demonstração. De (2.6), somando de $k = 0, 1, \cdots, n-1$ temos que:

$$\sum_{k=0}^{n-1} \frac{1}{2\lambda_k} \| x^{k+1} - x^k \|^2 \leqslant f\left(x^0\right) - f\left(x^n\right). \tag{2.7}$$

Como f é limitada inferiormente e aplicando o limite com $n \to \infty$ em (2.7) obtemos que

$$\sum_{k=0}^{n-1} \frac{1}{2\lambda_k} \| x^{k+1} - x^k \|^2 < \infty,$$

sendo $\liminf \lambda_k > 0$ temos que

$$\sum_{k=0}^{n-1} \| x^{k+1} - x^k \|^2 < \infty,$$

desta forma $\lim\limits_{k\to\infty} \| x^{k+1} - x^k \| = 0$. □

Teorema 2.3. *Se* $\left\{x^k\right\}$ *é limitada. Então, cada ponto de acumulação de* $\left\{x^k\right\}$ *é um ponto crítico da função* f.

Demonstração. Sejam x^* e w^* pontos de acumulação de $\left\{x^k\right\}$ e $\left\{w^k\right\}$ respectivamente, onde w^* existirá pois sendo $\left\{x^k\right\}$ limitada temos que $\left\{w^k\right\}$ também será, pois h é convexa. Assim, existem subsequências $\left\{x^{k_j}\right\}$ e $\left\{w^{k_j}\right\}$ que convergen para os pontos de acumulação dados. Sendo h convexa, segue-se de (2.1) e da continuidade do subdiferencial (Proposição 1.7) que $w^* \in \partial h\left(x^*\right)$. Resta mostrar que $w^* \in \partial g\left(x^*\right)$. De (2.3), existe $z^{k_j+1} \in \partial g\left(x^{k_j+1}\right)$ tal que

$$\| w^{k_j} - z^{k_j+1} \| = \frac{1}{2\lambda_{k_j}} \| x^{k_j+1} - x^{k_j} \|, \tag{2.8}$$

sendo λ_k limitada e de (2.8) e usando a Proposição 2.2, resulta que

$$\lim_{k\to\infty} z^{k_j+1} = \lim_{k\to\infty} w^{k_j} = w^*.$$

Como g é convexa, e da continuidade do subdiferencial (Proposição 1.7) segue-se que $w^* \in \partial g\left(x^*\right)$. Consequentemente, $w^* \in \partial g\left(x^*\right) \cap \partial h\left(x^*\right)$. Portanto, da Definição (1.8) x^* é ponto crítico de f. □

Observação: Se $\left\{x^k\right\}$ é limitada, do Corolário 2.1 temos que $f\left(x^k\right) \to f\left(x^*\right)$, onde x^* é algum ponto de acumulação da sequência.

Ilustração do Método: Seja a função $f : \mathbb{R} \longrightarrow \mathbb{R}$ dada por, $f(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2$ como no Exemplo 1.8. Já sabemos que a função possui como pontos críticos $x = 0$ e $x = \pm 1$. Aplicaremos o Algoritmo para verificar o fato. Note que $\partial g(x) = \{x^3\}$ e $\partial h(x) = \{x\}$, logo $\partial g(x^{k+1}) = (x^{k+1})^3$ e $\partial h(x^k) = x^k$, por h ser diferenciável temos que no algoritmo que $w^k = x^k$. Aplicando o Algoritimo e fazendo $\lambda_k = 1$ segue-se que

$$0 = \partial g(x^{k+1}) - w^k + \frac{1}{\lambda_k}(x^{k+1} - x^k),$$

ou ainda,

$$0 = (x^{k+1})^3 + x^{k+1} - 2x^k.$$

Se tomarmos um ponto $p \in (0, 1)$ temos que a equação $x^3 + x - 2p = 0$ terá solução $x^* \in (0, 1)$ pelo Teorema do Valor Intermediário e única pelo Teorema de Rolle. Logo se $x^0 \in (0, 1)$ então $x^k \in (0, 1)$ para todo $k \in \mathbb{N}$, ou seja, $\{x^k\}$ é limitada. Mostraremos agora por indução sobre k que $\{x^k\}$ é crescente. Tomemos $x^{k-1} \leqslant x^k$ com $(x^k)^3 + x^k - 2x^{k-1} = 0$ e x^{k+1} tal que $(x^{k+1})^3 + x^{k+1} - 2x^k = 0$. Fazendo a diferença entre as duas igualdades temos

$$(x^{k+1})^3 - (x^k)^3 + x^{k+1} - x^k + 2(x^{k-1} - x^k) = 0,$$

como $x^{k-1} - x^k \leqslant 0$ então $(x^{k+1})^3 - (x^k)^3 + x^{k+1} - x^k \geqslant 0$, logo segue-se que

$$\begin{aligned}
0 &\leqslant (x^{k+1})^3 - (x^k)^3 + x^{k+1} - x^k \\
&= (x^{k+1})^3 - (x^{k+1})^2.x^k + (x^{k+1})^2.x^k - (x^{k+1}).(x^k)^2 + (x^{k+1}).(x^k)^2 - (x^k)^3 + x^{k+1} - x^k \\
&= x^{k+1}.\left[(x^{k+1})^2 + x^{k+1}.x^k + (x^k)^2 + 1\right] - x^k.\left[(x^{k+1})^2 + x^{k+1}.x^k + (x^k)^2 + 1\right] \\
&= \left[(x^{k+1})^2 + x^{k+1}.x^k + (x^k)^2 + 1\right].(x^{k+1} - x^k).
\end{aligned}$$

Logo $\left[(x^{k+1})^2 + x^{k+1}.x^k + (x^k)^2 + 1\right].(x^{k+1} - x^k) \geqslant 0$, como $x^k \in (0, 1)$ teremos que $(x^{k+1})^2 + x^{k+1}.x^k + (x^k)^2 + 1 \geqslant 0$ consequentemente $(x^{k+1} - x^k) \geqslant 0$, desta forma $(x^{k+1} \geqslant x^k)$ o que prova a indução.

Aplicando assim o algoritmo temos que $\{x^k\}$ é limitada em $(0, 1)$ e crescente, portanto a menos de uma subsequência irá convergir para 1 que é um ponto crítico da função em questão. Além disso se tomarmos $x^0 \in (-1, 0)$ mostra-se também de modo analogo que a menos de subsequência $\{x^k\}$ irá convergir à -1.

2.4 Aplicação: Convergência Linear

Anteriormente foi mostrado que o Algoritmo é uma boa ferramenta para encontrar os pontos críticos de uma funções DC. Consideraremos agora o mesmo problema com algumas condições especiais sobre as funções componentes. Tomaremos a função h diferenciável e ∇h uma função Lipschitz contínua com módulo L, e a g fortemente convexa com constante ρ. Assim (2.1) e (2.2) será reduzido a:

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \left\{ g(x) - \langle \nabla h\left(x^k\right), x - x^k\rangle + \frac{1}{2\lambda_k} \| x - x^k \|^2 \right\} . \tag{2.9}$$

É importante lembrar se a sequência $\left\{x^k\right\}$ é limitada então os pontos de acumulação são pontos críticos de f. Chamando de S o conjunto dos pontos críticos de f, com $S \neq \emptyset$, então os pontos de acumulação x^* de $\{x^k\}$ pertencerão S, neste caso serão um ponto crítico de f se $\nabla h(x^*) \in \partial g(x^*)$. Não teremos nenhum problema para este caso, pois sendo g fortemente convexa podemos escolher uma função φ fortemente convexa, onde a funçã admite f uma decomposição em componentes fortemente convexas da seguinte forma: $f(x) = (g(x) + \varphi(x)) - (h(x) + \varphi(x))$, uma vez que a soma de funções fortemente convexa é fortemente convexa.

Teorema 2.4. *Seja $\left\{x^k\right\}$ gerada pelo Algoritmo com (2.1) e (2.2) substituído por (2.9). Assuma que g é uma função fortemente convexa com constante $\rho > 0$ e $\nabla h(x)$ é uma função Lipschitz contínua com constante $L > 0$. Se $\rho > 2L$, então existe uma constante $0 < r < 1$ tal que*

$$\| x^{k+1} - x^* \| \leqslant r \| x^k - x^* \| , \quad \forall x^* \in S . \tag{2.10}$$

Demonstração. De (2.9), existe um $z^{k+1} \in \partial g\left(x^{k+1}\right)$ tal que

$$\nabla h\left(x^k\right) = z^{k+1} + \frac{1}{\lambda_k}\left(x^{k+1} - x^k\right) \Longrightarrow z^{k+1} = \nabla h\left(x^k\right) + \frac{\left(x^k - x^{k+1}\right)}{\lambda_k} . \tag{2.11}$$

Tomemos $x^* \in S$, de acordo com a Definição (1.8) temos que $\partial h(x^*) \cap \partial g(x^*) \neq \emptyset$, sendo h diferenciável obtemos que $\partial h(x^*) = \nabla h(x^*)$, daí $\nabla h(x^*) \in \partial g(x^*)$. Como g é fortemente convexa *módulo* ρ segue-se que ∂g é fortemente monótono (Teorema 1.11), combinando (2.11) com o fato de $\nabla h(x^*) \in \partial g(x^*)$ temos

$$\begin{aligned} 0 &\leqslant \langle x^{k+1} - x^*, z^{k+1} - \nabla h(x^*)\rangle - \rho \| x^{k+1} - x^* \|^2 \\ &= \langle x^{k+1} - x^*, \frac{\left(x^k - x^{k+1}\right)}{\lambda_k} + \nabla h\left(x^k\right) - \nabla h(x^*)\rangle - \rho \| x^{k+1} - x^* \|^2, \end{aligned}$$

como $\lambda_k > 0$ temos que

$$0 \leqslant 2\lambda_k \langle x^{k+1} - x^*, \nabla h\left(x^k\right) - \nabla h(x^*)\rangle - 2\langle x^{k+1} - x^*, x^{k+1} - x^k\rangle - 2\lambda_k \rho \| x^{k+1} - x^* \|^2,$$

usando a desigualdade de Cauchy-Schwarz

$$\begin{aligned} 0 &\leqslant 2\lambda_k \parallel x^{k+1} - x^* \parallel . \parallel \nabla h\left(x^k\right) - \nabla h\left(x^*\right) \parallel -2\langle x^{k+1} - x^*, x^{k+1} - x^k\rangle \\ &\quad -2\lambda_k \rho \parallel x^{k+1} - x^* \parallel^2, \end{aligned}$$

sendo ∇h Lipschitz temos

$$\begin{aligned} 0 &\leqslant 2\lambda_k L \parallel x^{k+1} - x^* \parallel . \parallel x^k - x^* \parallel -2\langle x^{k+1} - x^*, x^{k+1} - x^k\rangle \\ &\quad -2\lambda_k \rho \parallel x^{k+1} - x^* \parallel^2. \end{aligned} \tag{2.12}$$

Note que

$$\begin{aligned} 0 &\leqslant \|x^k - x^*\|^2 = \|x^k - x^{k+1} + x^{k+1} - x^*\|^2 \\ &= |x^{k+1} - x^*\|^2 - 2\langle x^{k+1} - x^*, x^{k+1} + x^k\rangle + \|x^k - x^*\|^2 \ , \end{aligned}$$

assim

$$-2\langle x^{k+1} - x^*, x^{k+1} + x^k\rangle = \|x^k - x^*\|^2 - |x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2. \tag{2.13}$$

E dado $a, b \in \mathbb{R}$, com $a, b \geqslant 0$ temos que $0 \leqslant (a-b)^2 = a^2 - 2ab + b^2$, consequentemente $a^2 + b^2 \geqslant 2ab \geqslant ab$, deste fato temos

$$(\|x^{k+1} - x^*\|^2 + \|x^k - x^*\|^2) \geqslant \|x^{k+1} - x^*\|^2 . \|x^k - x^*\|^2 \tag{2.14}$$

Combinado (2.12) com (2.13) e (2.14)

$$\begin{aligned} 0 &\leqslant 2\lambda_k L \left(\parallel x^{k+1} - x^* \parallel^2 + \parallel x^k - x^* \parallel^2\right) - \parallel x^{k+1} - x^* \parallel^2 - \parallel x^k - x^{k+1} \parallel^2 \\ &\quad + \parallel x^k - x^* \parallel^2 - 2\lambda_k \rho \parallel x^{k+1} - x^* \parallel^2 \\ &= (1 + 2\lambda_k L) \parallel x^k - x^* \parallel^2 - [1 + 2\lambda_k (\rho - L)] . \parallel x^{k+1} - x^* \parallel^2 - \parallel x^k - x^{k+1} \parallel^2, \end{aligned}$$

daí,

$$[1 + 2\lambda_k (\rho - L)] . \parallel x^{k+1} - x^* \parallel^2 \leqslant (1 + 2\lambda_k L) \parallel x^k - x^* \parallel^2,$$

desta forma temos que

$$\parallel x^{k+1} - x^* \parallel^2 \leqslant \sqrt{\frac{(1 + 2\lambda_k L)}{1 + 2\lambda_k (\rho - L)}} . \parallel x^k - x^* \parallel^2,$$

como $\rho > 2L$ podemos definir

$$0 < r := \sqrt{\frac{1 + 2\lambda_k L}{1 + 2\lambda_k (\rho - L)}} < 1.$$

Com isso garantimos a convergência linear para o caso em que f é uma função fortemente convexa. $\square$

Observação: Note que a hipótese de g ser fortemente convexa é razoavel pois dada uma função fortemente convexa $\varphi(x)$, e $f(x) = g(x) - h(x)$, temos que f sempre admite uma decomposição fortemente convexa, isto é, $f(x) = (g(x) + \varphi(x)) - (h(x) + \varphi(x))$, sabendo que a soma de uma função convexa com uma fortemente convexa é uma função fortemente convexa. A referência citada acima esta feito em [12] pág. 63.

A condição $\rho > 2L$ implica que f é fortemente convexa pois,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle = \langle \nabla g(x) - \nabla g(y), x - y \rangle - \langle \nabla h(x) - \nabla h(y), x - y \rangle, \quad (2.15)$$

como g é fortemente convexa módulo ρ temos

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geqslant \rho \|x - y\|^2. \quad (2.16)$$

Usando a desigualdade de Schuwrs e o fato de ∇h ser Lipschtz temos que,

$$\langle \nabla h(x) - \nabla h(y), x - y \rangle \leqslant \rho \|\nabla h(x) - \nabla h(y)\|.\|x - y\| \leqslant L.\|x - y\|, \quad (2.17)$$

daí,

$$-\langle \nabla h(x) - \nabla h(y), x - y \rangle \geqslant L.\|x - y\|^2. \quad (2.18)$$

Substituindo 2.16 e 2.18 em 2.15 temos

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geqslant \rho \|x - y\|^2 - L\|x - y\|^2,$$

sendo $\rho > 2L \Rightarrow -L > -\dfrac{\rho}{2}$, assim

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle > \frac{\rho}{2} \|x - y\|^2.$$

Logo f é fortemente convexa. Assim atravéz da convergência linear do método proximal para funções DC no caso fortemente convexo podemos dar uma nova demonstração da convergência linear do APP clássico. Para isto basta tomar a função convexa original do método clássico, digamos f e tomar uma função fortemente convexa h nas condições do teorema acima. Assim podemos escrever $f(x) = (f(x) + h(x)) - h(x)$ e aplicar o método. Podemos verificar esta referência em [12] pág. 65.

Observação: Note que por $0 < r < 1$ temos que $\| x^{k+1} - x^* \| \leqslant \| x^k - x^* \|$,ou seja, $\{x^k\}$ é Fejer Convergente para o conjunto S. Então usando a Proposição 1.1 e o Teorema 2.3 obtemos resultado análogo ao anterior.

Capítulo 3

Considerações Finais

Neste trabalho foi realisado o estudo de um método proximal para minimizar funções DC, $f(x) = g(x) - h(x)$, as quais contém a classe de funções convexas.

Mostramos que o Algoritmo proposto gera uma sequência a qual está bem definida e sob a hipótese de ser limitada, garantimos que seus pontos de acumulação são pontos críticos de f. Também fizemos uma demonstração da convergência linear do método quando as componentes g for fortemente convexa e a h for diferenciável de tal modo que a f seja fortemente convexa.

É fato que, quando a segunda componente de f for identicamente nula, $h(x) \equiv 0$, retornamos ao método do PPA clássico para funções convexas, como consequência obtemos a limitação da sequência. Neste sentido, como trabalhos futuros seria obter a limitação da sequência no caso em que a segunda componente de f não seja identicamente nula.

Bibliografia

[1] Fernandes, M.O.;*Otimização Por Funções Representáveis Como a Diferença Entre Funções Convexas Com Aplicação em um Problema de Arranjo Físico*, 2012.

[2] Hartman, P.; *On Functions Reresentable as Difference of Convex Functions*, Pacific Journal of Mathematics, Vol. 9, pp.707-713, 1959.

[3] Hoai An, L.T.; TAO, P.D.; *A continuos approaach for globally solving linearly constained quadratic zero-one programation problems, Optimization.* 50, 93120, 2001.

[4] Hönig, C.S.; *Aplicações da Topologia à Análise.* Projeto Euclides - IMPA, 1976.

[5] Iusem, A. N.;*Augmented Lagrangian Methods and Proximal Point Methods for Convex Optimization.* Investigación Operativa, v. 8, p. 11-49, 1999.

[6] Iusem, A. N.; *Métodos de ponto proximal em otimização.* IMPA, Rio de Janeiro, 1995.

[7] Izmailov, Alexey.; *Otimização volume 1. Condições de otimalidade, elementos de análise convexa e dualidade* / Alexey Izmailove, Mikhail Solodov. Rio de Janeiro: IMPA, 2005.

[8] Mangasarian, O.L.; *Nonlinear Programming.* McGraw-Hill, New York, 1994.

[9] Martine, B.;*Regularisation, d'inéquations variationelles par approximations successives.* (French) Rev. Française Informat. Recherche Opérationnelle 4, Ser. R-3, 154-158, 1970.

[10] Moudafi, A., and Maingé, P-E.;*On the convergence of an approximate proximal method for d.c. functions.* Journal of Computational Mathematics, 24, 475-480, 2006.

[11] Rockafellar, R. T.;*Monotone Operators and the Proximal Point Algorithm.* SIAM Journal on Control and Optimization, 14, 877-898, 1976.

[12] Souza,J.C.O.; *Convergência de Método de Descida Para Funções não Convexas com Aplicações à Teoria de Comportamento*, Tese, PESC-COPPG, UFRJ, 2016.

[13] Souza,J.C.O. , Oliveira,P.R. , A. Soubeyran ; *Global convergence of a proximal linearized algorithm for difference of convex functions*, Optimization Letters, 10(7), 1529-1539, 2016.

[14] Souza,J.C.O. , Oliveira,P.R.; *A proximal point method for DC functions on Hadamard manifolds*, Jornal of Global Optimization, 63(4), 797-810, 2015.

[15] Sun, W., Sampaio, R.J.B., and Candido, M.A.B.; *Proximal point algorithm for minimization of DC Functions*. Journal of Computational Mathematics, 21, 451-462, 2003.

[16] Tuy, H., Al-Khayyal, F., ZHOU, F.; *A DC Optimization Method for Single Facility Location Problems J. Global Optim*. 7: 209-227, 1995.

[17] Tuy, H.; *Convex Analysis and Global Optimization*, Honai, Vietnam: Kluwer Academic Publisher, 1998.