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

Algoritmo do Ponto proximal Inercial para Programação DC

Sillas Augusto Ferreira Carvalho Frazão

Teresina - 2022

Sillas Augusto Ferreira Carvalho Frazão

Dissertação de Mestrado:

Algoritmo do Ponto Proximal Inercial para Programação DC

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 - 2022

FICHA CATALOGRÁFICA
Universidade Federal do Piauí
Biblioteca Comunitária Jornalista Carlos Castello Branco
Divisão de Representação da Informação

F848a Frazão, Sillas Augusto Ferreira Carvalho.
Algoritmo do Ponto Proximal Inercial para Programação DC / Sillas Augusto Ferreira Carvalho Frazão. -- 2022.
53 f.

Dissertação (Mestrado) – Universidade Federal do Piauí, Programa de Pós-Graduação em Matemática, Teresina, 2022.
"Orientador: Prof. Dr. Jurandir de Oliveira Lopes".

1. Programação DC. 2. Algoritmo de Ponto Proximal. 3. Algoritmo Inercial. I. Lopes, Jurandir de Oliveira. II. Título.

CDD 511.8

Bibliotecária: Francisca das Chagas Dias Leite - CRB3/1004

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

Algoritmo do Ponto Proximal Inercial para Programação DC

Sillas Augusto Ferreira Carvalho Frazão

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 24 de agosto de 2022.

Banca Examinadora:

Jurandir de Oliveira Lopes

Prof. Dr. Jurandir de Oliveira Lopes - Orientador

Pedro Antônio Soares Júnior

Prof. Dr. Pedro Antônio Soares Júnior - UESPI

joão Xavier da Cruz Neto

Prof. Dr. João Xavier da Cruz Neto - UFPI

José Gomes Carvalho. (In memoriam).

Agradecimentos

Agradeço aos meus familiares, em especial a Ana cláudia (mãe), Selles Gustavo (irmão), Maria de Lourdes (avó), Maria de Jesus (tia), Kelle Cristina (tia), Raimundo Batista (tio), Anderson Reis e Leandro Reis (primos), por todo o apoio prestado durante minha graduação e mestrado, sem esse apoio talvez não seria possível concluir os cursos.

Agradeço aos meus amigos que fizeram parte da minha turma de mestrado pelos vários momentos de estudo e apoio, em especial a Raquel Lemos e ao Paulo Sergio.

Não poderia deixar de agradecer ao apoio dos que se fizeram mais presentes durante todo esse tempo de curso: Natália Tiffany, Suerlan Oliveira, Gabriel Barbosa e Jefferson Galvão. Obrigado por todo carinho, todas as brincadeiras e todas as piadas ruins contadas durante esse tempo de universidade. Ao lado de vocês tudo ficou mais leve e divertido.

Agradeço ao Evandro Junior e, novamente, ao Jefferson Galvão por ter me cedido e me acolhido em suas casas nos momentos que mais precisei.

Agradeço aos meus amigos de graduação que sempre me animou para continuar na caminhada, em especial a Iara Lima, Danilo Silva, Vitor Gabriel (Ousado Resolve), Débora Rios, Daniel Sertanejo, Edvan Cardoso e Brenner Matos.

Agradeço a Jurandir de Oliveira Lopes pela orientação, por compartilhar seus conhecimentos comigo, pelo carinho e, principalmente, pela confiança que depositou em mim durante todo esse tempo.

Agradeço aos meus professores de graduação por ter me dado o primeiros contato com a matemática, em especial a Kelton Bezerra, Marcos Vinicius, Barnabé, Roger Peres de Moura, Paulo Alexandre e Mário Gomes.

Agradeço aos meus professores de mestrado, em especial a Gleison Santos, João Carlos, Newton Luis, Jefferson Leite, Antonio Wilson e Ítalo Dowell, por ter se esforçado tanto em compartilhar seus conhecimentos nesse momento pandêmico tão dificil que atrvessamos.

Agradeço também ao professor Leandro Pessoa pela valiosa ajuda na preparação para seleção de mestrado.

Agradeço aos professores Pedro Soares e João Xavier por aceitarem a partiicipar da banca.

Agradeço ao CNPq pelo apoio financeiro.

"A gente não precisa ser bom em tudo nem muito menos o tempo todo".

Desconhecido

Resumo

Neste trabalho, apresentamos um algoritmo de ponto proximal inercial que tem como objetivo resolver o problema de minimização DC (diferenças de funções convexas). Mostramos a boa definição da sequência gerada pelo algoritmo, que cada iterada dessa sequência soluciona um subproblema proposto e que todo ponto de acumulação de $\{x^k\}$ é um ponto crítico do problema de minimização DC.

Abstract

In this work, we present an inertial proximal point algorithm that aims to solve the DC minimization problem (differences of convex functions). We show the good definition of the sequence generated by the algorithm, that each iteration of this sequence solves a proposed subproblem and that every accumulation point of $\{x^k\}$ is a critical point of the DC minimization problem.

Sumário

Introdução

As funções DC, ou seja, funções que podem ser representadas como a diferença de duas funções convexas, vem sendo estudada a bastante tempo e atualmente tem tido grande notoriedade na área da Otimização (ver Andrade et al. [4], Oliveira e Tcheou [9], Lima [15] e, Tao e An [20]). Esse interesse tem sido pelo fato dos problemas de otimização DC englobar os problemas de otimização convexa, pela grande aplicabilidade de tais funções em diferentes contextos e também pela estabilidade de tais funções em relação as operações usuais da otimização (ver Fernandes [10], Hartman [11] e Tuy [21]).

O Algoritmo do Ponto Proximal para resolver problemas de minimização foi introduzido por Martinet [16], popularizado por Rockafellar [17] e consiste em, a partir de um ponto inicial dado, gerar uma sequência em que cada iterada é solução do problema de minimizar a função objetivo acrescida da função distância ao quadrado . Já o Algoritmo do Ponto Proximal Inercial consiste em, a partir de um ponto inicial dado, gerar uma sequência de pontos que soluciona um subproblema que possui uma força inercial, ou seja, além desse subproblema lidar com a iterada atual, ele também utiliza iteradas anteriores. Algoritmos inerciais tem sido bastante estudados atualmente, principalmente por sua aplicabilidade. Justificativas para o uso de Algoritmos Inerciais e aplicações podem ser encontrados em Alvarez e Attouch [3], Andrade et al [4], Oliveira e Tcheou [9].

2003, Sun et al. [19] apresentaram uma versão do Algoritmo do Ponto Proximal para resolver problemas de otimização DC. Como pode ser verificado, uma função DC admite infinitas decomposições e estas podem interferir na qualidade do algoritmo. Encontrar uma decomposição "ótima" para se obter melhor qualidade do algoritmo ainda é um problema em aberto, veja Oliveira Souza [8].

Nesta dissertação, estudamos um Algoritmo do Ponto Proximal Inercial, no espaço Euclideano, que foi apresentado por Andrade et al. em [4] no contexto de Variedade de Hadamard. Desta forma, seja $f : \mathbb{R}^n \to \mathbb{R}$, tal que $f = g-h$, onde g e h são funções convexas. O problema de minimização DC consiste no seguinte:

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

Assim, apresentamos um algoritmo que gera uma sequência em que cada iterada soluciona um subproblema de minização e que cada ponto de acumulação será um ponto crítico (ver Definição 13) de f.

A dissertação está organizada do seguinte modo: no capítulo 1, apresentamos resultados de análise no $\mathbb{R}^n$, análise convexa diferenciável e não diferenciável, e funções DC. No capítulo 2, apresentamos o algoritmo e analisamos a convergência da sequência gerada pelo algoritmo. No capítulo 3, apresentamos as considerações finais.

Capítulo 1

Noções Preliminares

O texto apresentado nesta seção não pretende ser um curso dos temas em questão, traz apenas os resultados necessários para a compreensão dos assuntos posteriores. Os resultados e notações podem ser encontrados em [1, 2, 5, 7, 13, 14]

1.1 Análise no $\mathbb{R}^n$

Seja $f : \Omega \to \mathbb{R}$, com $\Omega \subset \mathbb{R}^n$. Considere o seguinte problema

$$\underset{x \in D}{min}\ f(x) \quad \text{onde} \quad D \subset \Omega. \tag{1.1}$$

Neste caso, D é denominado conjunto viável do problema, os pontos de D são ditos pontos viáveis e f é a função objetivo.

Definição 1 *Sejam $f : \Omega \to \mathbb{R}$, $D \subset \Omega$ e $x^* \in D$.*

- *x^* é mínimo local (resp. local estrito) quando existe uma vizinhança U de x^* tal que*

$$f(x^*) \leq f(x), \quad \forall x \in D \cap U \quad (resp. <).$$

- *x^* é mínimo global (resp. global estrito) quando*

$$f(x^*) \leq f(x), \quad x \in D \quad (resp. <).$$

- *O número $\overline{v} \in [-\infty, +\infty)$ dado por*

$$\overline{v} = \inf_{x \in D} f(x)$$

*é o **valor ótimo** do problema* (1.1).

Nosso objetivo é encontrar pontos críticos (ver Definição 13) da função f, em particular, as soluções do problema (1.1) são pontos críticos.

Exemplo 1 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ definida por $f(x) = \sum_{i=1}^n e^{x_i}$. Assim, o problema de minimizar f em $\mathbb{R}^n$ não possui solução, entretanto e seu valor ótimo é 0.*

Definição 2 *Dizemos que uma função $f : D \subset \mathbb{R}^n \to \mathbb{R}$ é semicontínua inferiormente no ponto $x \in D$, quando para qualquer sequência $\{x^k\} \subset D$ tal que $x^k \to x$ tem-se que*

$$\lim_{k \to \infty} \inf f(x^k) \geq f(x).$$

Note que toda função contínua é semicontínua inferiormente. O próximo teorema garante a existência mínimos.

Teorema 1 *(Weierstrass) Seja $f : \mathbb{R}^n \to \mathbb{R}$ semicontínua inferiormente e $D \subset \mathbb{R}^n$ compacto. Então o problema de minimizar f em D possui solução global.*

Demonstração: Veja que $\{f(D)\}$ é limitado inferiormente. De fato, se supormos o contrário, existiria um sequência $\{x^k\} \subset D$ de modo que $f(x^k) \to -\infty$. Pela compacidade de D, existe $x^* \in D$ de modo que $x^{k_j} \to x^*$ para alguma subsequência $\{x^{k_j}\}$. Daí, pela semiconitnuidade inferior temos que

$$-\infty = \lim_{j \to +\infty} \inf f(x^{k_j}) \geq f(x^*).$$

Absurdo. Portanto $\{f(D)\}$ é limitado inferiormente. Assim, existe $f^* = \inf_{x \in D} f(x)$, que implica em existir também uma sequência $\{y^k\} \subset D$ de modo que

$$f(y^k) \to f^*.$$

Novamente pela compacidade de D, existe $y^* \in D$ satisfazendo $f(y^{k_j}) \to f^*$ para alguma subsequência $\{y^{k_j}\}$, com $y^{k_j} \to y^*$. Assim, y^* é um solução global.

$\blacksquare$

Exemplo 2 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = \langle Ax, x \rangle$, onde $A \in \mathbb{R}^{n \times n}$ é simétrica. Mostre que f tem um minimizador global em $S = \{x \in \mathbb{R}^n; \| x \| = 1\}$. Mostre também que existe $\lambda \in \mathbb{R}$ de modo que $\langle Ax, x \rangle \geq \lambda \| x \|^2, \quad \forall x \in \mathbb{R}^n$.*

Solução: Note que S é compacto e f é contínua, portanto pelo Teorema de Weierstrass existe x^* minimizador global de f em S. Agora dado $x \in \mathbb{R}^n \backslash \{0\}$, temos que $\dfrac{x}{\| x \|} \in S$ e daí

$$f(\frac{x}{\| x \|}) \geq f(x^*)$$

$$\Rightarrow \langle A\frac{x}{\| x \|}, \frac{x}{\| x \|} \rangle \geq \langle Ax^*, x^* \rangle$$

$$\Rightarrow \langle Ax,\, x \rangle \geq \langle Ax^*, x^* \rangle \| x \|^2 .$$

Fazendo $\lambda = \langle Ax^*, x^* \rangle$, temos o resultado.

Corolário 1.1 *Seja $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ uma função semicontínua em D. Suponhamos que existe $c \in \mathbb{R}$ de modo que $L_{f,D}(c) := \{x \in D |\, f(x) \leq c\}$ seja não vazio e compacto. Então o problema de minimizar f em D tem solução global.*

Demonstração: Pelo Teorema de Weierstrass (Teorema 1), o problema de minimizar f em $L_{f,D}(c)$ tem solução global, digamos x^*, ou seja, $f(x^*) \leq f(x)$ para todo $x \in L_{f,D}(c)$. Por outro lado, se $y \in D \backslash L_{f,D}(c)$, então

$$f(x^*) \leq c < f(y).$$

Portanto, todos os pontos de D que não estão no conjunto de nível $L_{f,D}(c)$ possuem valores maiores que $f(x^*)$. Assim x^* é um ponto de mínimo global.

■

A noção de coercividade será crucial para a demonstração da boa definição do algoritmo, pois é com ela que iremos garantir a existência de soluções dos subproblemas.

Definição 3 *Dizemos que uma função $f : D \to \mathbb{R}$ é* ***coerciva no conjunto D****, quando para cada sequência $\{x^k\} \subset D$ tal que ou $\| x^k \| \to +\infty$ ou $x^k \to x \in \overline{D} \backslash D$, tem-se que $\lim_{k\to\infty} \sup f(x^k) = +\infty$.*

Corolário 1.2 *Sejam $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ uma função semicontíuna inferiormente e coerciva em $D \neq \emptyset$. Então o problema de minimizar f em D possui solução global.*

Demonstração: Como $D \neq \emptyset$, podemos encontrar $c \in \mathbb{R}$ de modo que $L_{f,D}(c)$ seja não vazio, basta fazer $c = f(x)$, para algum $x \in D$. Note que tal conjunto é limitado, pois caso contrário existiria uma sequência $\{x^k\} \in L_{f,D}(c)$ tal que $\| x^k \| \to +\infty$ e assim, pela coercividade, teríamos $\limsup_{k\to\infty} f(x^k) \to +\infty$, que é um absurdo, já que $f(x^k) \leq c, \quad \forall k \in \mathbb{N}$. Note também que $L_{f,D}(c)$ é fechado, pois caso contrário existiria uma sequência $\{x^k\} \subset L_{f,D}(c)$ convergindo para um ponto x que não pertence a $L_{f,D}(c)$. Assim temos duas opções: x pertence a D ou x não pertence a D. Supondo $x \in D$ e usando que f é semicontínua inferiormente, teríamos que $\liminf_{k\to\infty} f(x^k) \geq f(x) > c$, absurdo, pois $f(x^k) \leq c, \quad \forall k \in \mathbb{N}$. Supondo agora que x não pertence a D, teríamos então que $x \in \overline{D}\backslash D$ e, pela coercividade, teríamos $\limsup_{k\to+\infty} f(x^k) = +\infty$, absurdo, já que $f(x^k) \leq c, \quad \forall k \in \mathbb{N}$.

■

Agora veremos as condições de otimalidade, necessárias e suficientes, que caracterizam um minimizador de uma função f. A partir de agora denotaremos o gradiente e a Hessiana da função f, respectivamente, por f' e f''.

Teorema 2 *(Condição necessária de 1^a ordem) Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto x^*. Se x^* é um minimizador local de f, então*

$$f'(x^*) = 0.$$

Demonstração: Suponha que x^* seja um minimizador local de f. Por definição, existe uma vizinhança U de x^* tal que $f(x^*) \leq f(x), \quad \forall x \in U$. Fixado $d \in \mathbb{R}^n$, note que existirá $\epsilon > 0$ tal que $x^* + td \in U, \quad \forall t \in (0, \epsilon]$. Assim, usando a diferenciabilidade de f em x^*, temos

$$0 \leq f(x^* + td) - f(x^*) = tf'(x^*)d + r(t), \quad \forall t \in (0, \epsilon],$$

onde $\lim_{t\to 0} \frac{r(t)}{t} = 0$. Dividindo tudo pot $t > 0$ temos que

$$0 \;\; \leq \;\; f'(x^*)d + \frac{r(t)}{t}.$$

Agora fazendo $t \to 0^+$ temos que $0 \leq f'(x^*)d$. Como d é arbitrário, escolha $d = -f'(x^*)$ e teremos $0 \leq - \| f'(x^*) \|^2$, implicando que $f'(x^*) = 0$.

■

Teorema 3 *(Condição necessária de 2^a ordem) Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função duas vezes diferenciável no ponto x^*. Se x^* é um minimizador local de f, então a matriz Hessiana de f no ponto x^* é semidefinida positiva, ou seja,*

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

Demonstração: Suponha que x^* seja um minimizador local de f. Fixe $d \in \mathbb{R}^n$. Assim existirá $\epsilon > 0$ de tal modo que

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

$\forall t \in (0, \epsilon]$ com $\lim_{t\to 0} \dfrac{r(t)}{t^2} = 0$. Dividindo tudo por $t^2 > 0$ temos que

$$0 \leq \frac{1}{2}\langle f''(x^*)d, d\rangle + \frac{r(t)}{t^2}.$$

Fazendo $t \to 0^+$, temos que $\langle f''(x^*)d, d\rangle \geq 0$.

■

Definição 4 *Um ponto x^* é dito **estacionário** ou **crítico** de uma funcão diferenciável $f : \mathbb{R}^n \to \mathbb{R}$ quando $f'(x^*) = 0$.*

Teorema 4 *(Condição suficiente de 2^a ordem) Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função duas vezes diferenciável no ponto x^*. Se x^* é um ponto estacionário e $f''(x^*)$ é definida positiva, então x^* é um minimizador local estrito de f.*

Demonstração: Seja $\lambda > 0$ o menor autovalor de $f''(x^*)$, pois tal matriz é simétrica e positiva definida. Assim, $\langle f''(x^*)d,\, d\rangle \geq \lambda \parallel d \parallel^2, \forall d \in \mathbb{R}^n$. Por Taylor,

$$f(x^* + d) = f(x^*) + \frac{t^2}{2}\langle f''(x^*)d,\, d\rangle + r(d) \geq f(x^*) + \frac{t^2}{2}\lambda \parallel d \parallel^2 + r(d),$$

onde $\lim_{d\to 0} \dfrac{r(d)}{\parallel d \parallel^2} = 0$. Dividindo tudo por $\parallel d \parallel^2 > 0$ temos que

$$\frac{f(x^* + d) - f(x^*)}{\| d \|^2} \geq \frac{\lambda}{2} + \frac{r(d)}{\| d \|^2}.$$

Portanto existe $\epsilon > 0$ de modo que $\dfrac{\lambda}{2} + \dfrac{r(d)}{\| d \|^2} > 0, \quad \forall d \in B(0, \epsilon) \setminus \{0\}$, implicando que $f(x^* + d) - f(x^*) > 0, \quad \forall d \in B(0, \epsilon) \setminus \{0\}$. Como $d \in \mathbb{R}^n$ é arbitrário, temos o resultado.

■

1.2 Análise Convexa

A noção de função convexa é essencial para o decorrer deste trabalho. Com ela alguns teoremas que eram apenas necessários tornam-se também suficientes. Os resultados e notação podem ser escontrados em [1].

Definição 5 *Um conjunto $C \subset \mathbb{R}^n$ é dito convexo se $[x, y] = \{(1-t)x + ty;\ t \in [0,1]\} \subset C$, para todo $x, y \in C$. Neste caso, $[x, y]$ é dito segmento de extremidades x e y.*

Exemplo 3 *Considere $C = B(0; r) = \{x \in \mathbb{R}^n;\ \| x \| < r\}$ com $r > 0$. Então, C é convexo. De fato, dados $x,\ y \in C$ e $t \in [0, 1]$*

$$\begin{aligned}
\| tx + (1-t)y \| &\leq \| tx \| + \| (1-t)y \| \\
&= t \| x \| + (1-t) \| y \| \\
&< tr + (1-t)y \\
&= r.
\end{aligned}$$

Exemplo 4 *O conjunto $S = \{x \in \mathbb{R}^n;\ \| x \| = r\}$, para $r > 0$, não é convexo. De fato, sabemos que $S \neq \emptyset$, basta considerar um vetor com uma entrada r e as demais iguais a 0. Daí, dado $x \in S$, temos que $-x \in S$. Portanto, se S fosse convexo teríamos que $\frac{1}{2}x + \frac{1}{2}(-x) \in S$ que implicaria que $0 \in S$, absurdo. Logo, S não é convexo.*

Teorema 5 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto x^*. Se x^* é um minimizador local de f no conjunto C, então*

$$\langle f'(x^*),\ x - x^* \rangle \geq 0,\ \forall x \in C.$$

Demonstração: Sendo x^* um minimizador local, existe $\delta > 0$ de modo que $f(y) \geq f(x^*)$ para todo $y \in B(x;\delta)$. Agora fixe $x \in C$. Daí, pela convexidade de C, $(1-t)x^* + tx \in C$ para todo $t \in (0,1)$. Note que para $t \in (0,1)$ com $t < \dfrac{\delta}{\|x^* - x\|}$, temos que

$$\begin{aligned} \| \, x^* - [(1-t)x^* + tx] \, \| &= \| \, tx^* - tx \, \| \\ &= t\|x^* - x\| \\ &< \delta. \end{aligned}$$

Logo, para $t \in (0,1)$ com $t < \dfrac{\delta}{\|x^* - x\|}$ temos que $0 \leq f((1-t)x^* + tx) - f(x^*)$. Portanto, usando a diferenciabilidade em x^*, com $t \in (0,1)$ e $t < \dfrac{\delta}{\|x^* - x\|}$,

$$\begin{aligned} 0 &\leq f((1-t)x^* + tx) - f(x^*) \\ &= f(x^* + t(x - x^*) - f(x^*) \\ &= t\langle f'(x^*), \, x - x^* \rangle + r(t), \end{aligned}$$

onde $\lim_{t\to 0} \dfrac{r(t)}{t} = 0$. Assim, dividindo a expressão toda por $t > 0$ e passando o limite $t \to 0^+$, temos $\langle f'(x^*), \, x - x^* \rangle \geq 0, \; \forall x \in C$. Como $x \in C$ é arbitrário, segue o resultado.

$\blacksquare$

Definição 6 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \to \mathbb{R}$ é dita convexa se para quaisquer $x, y \in C$ e $t \in [0,1]$ tem-se que*

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

Além disso, f será estritamente convexa quando a desigualdade acima for estrita para todo $x \neq y$ e $t \in (0, 1)$.

Exemplo 5 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = \| \, x \, \|^2$. Então é convexa. Com efeito, sejam $x, \, y \in \mathbb{R}^n$ e $t \in [0, \, 1]$. Assim, mostraremos que*

$$f(tx + (1-t)y) - [tf(x) + (1-t)f(y)] \leq 0.$$

De fato,

$$
\begin{aligned}
f(tx+(1-t)y)-[tf(x)+(1-t)f(y)] &= \|tx+(1-t)y\|^2 - t\|x\|^2 - (1-t)\|y\|^2 \\
&= t^2\|x\|^2 + (1-t)^2\|y\|^2 + 2t(1-t)\langle x,\, y\rangle \\
&- t\|x\|^2 - (1-t)\|y\|^2 \\
&= t(t-1)\|x\|^2 - t(1-t)\|y\|^2 + 2t(1-t)\langle x,\, y\rangle \\
&= t(1-t)[2\langle x,\, y\rangle - (\|x\|^2 + \|y\|^2)] \\
&= -t(1-t)\|x-y\|^2 \\
&\leq 0.
\end{aligned}
$$

Logo, f é convexa.

Definição 7 *A função $f : C \subset \mathbb{R}^n \to \mathbb{R}$, onde C é convexo, é dita fortemente convexa com módulo (ou parâmetro) $\rho > 0$ quando para quaisquer $x, y \in C$ e $t \in [0,1]$ tem-se*

$$f(tx+(1-t)y) \leq tf(x) + (1-t)f(y) - \rho t(1-t) \parallel x-y \parallel^2 .$$

Exemplo 6 *A função $f : \mathbb{R} \to \mathbb{R}$ dada por $f(x) = x^2$ é um exemplo de função fortemente convexa com $\rho = 1$ (segue do Exemplo 5).*

Teorema 6 *Sejam $D \subset \mathbb{R}^n$ convexo, $g : D \to \mathbb{R}$ uma função convexa, $f : g(D) \to \mathbb{R}$ uma função convexa não decrescente. Então a função $h : \mathbb{R}^n \to \mathbb{R}$ dada por $h(x) = f(g(x))$ é convexa.*

Demonstração: De fato, como g é convexa, temos que

$$g(tx+(1-t)y) \leq tg(x) + (1-t)g(y).$$

Daí,

$$f(g(tx+(1-t)y)) \leq f(tg(x)+(1-t)g(y)) \leq tf(g(x)) + (1-t)f(g(y)),$$

onde na primeira desigualdade foi usado o fato de f ser não decrescente e na segunda desigualdade foi usada a convexidade.

■

Agora daremos uma primeira caracterização para uma função ser convexa.

Teorema 7 *Sejam $C \subset \mathbb{R}^n$ convexo, $\lambda > 0$ e $f, g : C \subset \mathbb{R}^n \to \mathbb{R}$. Então, $f + \lambda g$ é convexa. Além disso, se f ou g são estritamente convexa, então $f + \lambda g$ também será.*

Demonstração: Dados $x, y \in C$ e $t \in [0, 1]$, temos que

$$\begin{aligned}
(f + \lambda g)(tx + (1-t)y) &= f(tx + (1-t)y) + \lambda g(tx + (1-t)y) \\
&\leq tf(x) + (1-t)f(y) + \lambda t g(x) + \lambda(1-t)g(y) \\
&= t(f(x) + \lambda g(x)) + (1-t)(f(y) + \lambda g(y)) \\
&= t(f + \lambda g)(x) + (1-t)(f + \lambda g)(y).
\end{aligned}$$

Logo, $f + \lambda g$ é convexa. Se f ou g é estritamente convexa, a prova é análoga.

■

Definição 8 *O Epígrafo de uma função $f : C \to \mathbb{R}$ é o conjunto*

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

Teorema 8 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : C \to \mathbb{R}$ é convexa em C se, e somente se, o seu epígrafo é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.*

Demonstração: De fato. Suponha que f seja convexa e considere (x, c), $(y, d) \in E_f$. Seja, $A = (1-t)(x, c) + t(y, d) = ((1-t)x + ty, (1-t)c + td)$. Agora note que

$$\begin{aligned}
f((1-t)x + ty) &\leq (1-t)f(x) + tf(y) \\
&\leq (1-t)c + td,
\end{aligned}$$

que implica que $A \in E_f$. Logo o epígrafo é convexo. Reciprocamente, suponha que E_f seja convexo. Note que $(x, f(x)), (y, f(y)) \in E_f$. Pela convexidade, $((1-t)x+ty, (1-t)f(x)+tf(y)$ pertence ao epígrafo para $t \in [0, 1]$. Daí,

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

que implica que f é convexa.

■

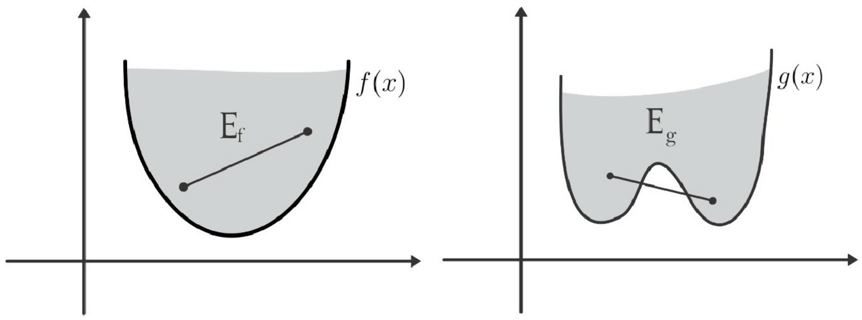

Figura 1.1: Ilustrando o Teorema 8. Na imagem f é convexa e g não é convexa.

Teorema 9 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : C \to \mathbb{R}$ uma função convexa em C. Então f é localmente Lipschitz-contínua em C. Em particular, f é contínua em C.*

Demonstração: Ver [1], página 147.

$\blacksquare$

O teorema a seguir mostra a importância e quão boa é a classe das funções convexas

Teorema 10 *(Teorema de minimização convexa) Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função convexa. Se $x^* \in C$ é um minimizador local de f, então x^* é minimizador global. Além disso, o conjunto dos minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração: Considere que $x^* \in C$ seja um minimizador local mas não global, ou seja, existem $r > 0$ e $y \in C$ tais que $f(x^*) \leq f(x)$ para todo $x \in B(x^*, r)$ e $f(y) < f(x^*)$. Pela convexidade de C, $(1-t)x^* + ty$ pertence a C sempre que $t \in [0, 1]$. Agora note que para t suficientemente pequeno temos que $(1-t)x^* + ty$ pertence a $B(x^*, r)$. De fato,

$$\begin{aligned} \| x^* - [(1-t)x^* + ty] \| &= \| t(x^* - y) \| \\ &= t \| x^* - y \|, \end{aligned}$$

basta então fazer $t < \dfrac{r}{\| x^* - y \|}$ que teremos $\| x^* - [(1-t)x^* + ty] \| = t \| x^* - y \| < r$. Agora usando a convexidade de f, segue que

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

Absurdo. Portanto, x^* é mínimo global.

Considere agora que $S \subset C$ seja o conjunto dos minimizadores de f. Fixe $x, y \in S$ e faça $\overline{x} = (1-t)x + ty$, com $t \in [0,1]$. Pela convexidade de C, temos que $\overline{x} \in C$ e, além disso, pelo fato provado acima, $f(x) = f(y)$. Daí, pela convexidade de f,

$$f(\overline{x}) \leq (1-t)f(x) + tf(y) = (1-t)f(x) + tf(x) = f(x)$$

Portanto $\overline{x} \in S$, o que mostra que S é convexo.

Suponha agora que f é estritamente convexa. Daí, sejam $x, y \in S$, $x \neq y$. Como S é convexo, $(1-t)x + ty = \overline{x} \in S$, com $t \in [0,1]$. Assim,

$$f(\overline{x}) < (1-t)f(x) + tf(y) = (1-t)f(x) + tf(x) = f(x).$$

Absurdo. Logo, neste caso, f só pode admitir um minimizador.

■

1.2.1 Funções Convexas Diferenciáveis

Agora iremos caracterizar as funções convexas diferenciáveis e seus pontos de mínimo.

Teorema 11 *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \to \mathbb{R}$ uma função diferenciável em Ω. Então são equivalentes:*

(a) A função f é convexa em Ω.

(b) Para todo x, $y \in \Omega$,

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

(c) Para todo x, $y \in \Omega$,

$$\langle f'(y) - f'(x), \, y - x \rangle \geq 0.$$

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

(d) A matriz Hessiana de f é semi-definida positiva em todo ponto de Ω:

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

Demonstração:
(a) $\Rightarrow$ **(b)**: Sejam x, $y \in \Omega$. Para $t \in (0,1)$ temos que $ty + (1-t)x = x + t(y-x) \in \Omega$, pela convexidade. Usando a diferenciabilidade, temos que

$$\begin{array}{rcl} tf(y) + (1-t)f(x) & \geq & f(ty + (1-t)x) \\ \Rightarrow t(f(y) - f(x)) + f(x) & \geq & f(x + t(y-x)) \\ & = & f(x) + t\langle f'(x), y - x\rangle + r(t) \end{array}$$

assim,

$$t(f(y) - f(x)) \geq t\langle f'(x), y - x\rangle + r(t).$$

Dividindo tudo por $t > 0$ e fazendo tal tender a zero pela direita, temos que

$$f(y) - f(x) \geq \langle f'(x), y - x\rangle$$

$$\Rightarrow f(y) \geq f(x) + \langle f'(x), y - x\rangle$$

(b) $\Rightarrow$ **(c)**: Temos que $f(y) \geq f(x) + \langle f'(x), y - x\rangle$ e também que $f(x) \geq f(y) + \langle f'(y), x - y\rangle$. Substituindo uma desigualdade noutra

$$f(x) \quad \geq \quad f(y) + \langle f'(y), x - y\rangle \quad \geq \quad f(x) + \langle f'(x), y - x\rangle + \langle f'(y), x - y\rangle$$

Daí,

$$0 \geq \langle f'(x), y - x\rangle + \langle f'(y), x - y\rangle$$

implicando em

$$0 \geq \langle f'(x) - f'(y), y - x\rangle,$$

ou seja,

$$\langle f'(y) - f'(x), y - x\rangle \geq 0.$$

(c) $\Rightarrow$ **(b)**: Pelo Teorema do Valor Médio, existe $t \in (0,1)$ tal que

$$f(y) - f(x) = \langle f'(x + t(y - x)), y - x \rangle.$$

Para os pontos $(x + t(y - x))$ e x, temos

$$\begin{array}{rcl} f(y) - f(x) = \langle f'(x + t(y - x)), y - x \rangle & = & \frac{1}{t}\langle f'(x + t(y - x)), t(y - x) \rangle \\ & \geq & \frac{1}{t}\langle f'(x), t(y - x) \rangle \\ & = & \langle f'(x), y - x \rangle. \end{array}$$

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

(b) $\Rightarrow$ **(a)**: Defina $d = y - x$ e aplique em (b) para os pontos x e $x + td$; y e $x + td$. Assim,

$$f(x) \geq f(x + td) - t\langle f'(x + td),\, d \rangle,$$

$$f(y) \geq f(x + td) + (1 - t)t\langle f'(x + \lambda d),\, d \rangle,$$

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

$$\begin{array}{rcl} (1 - t)f(x) + tf(y) & \geq & (1 - t)(f(x + td) - t\langle f'(x + td,\, d \rangle) \\ & + & t(f(x + td) + (1 - t)\langle f'(x + td,\, d \rangle) \\ & = & f(x + td) \\ & = & f((1 - t)x + ty). \end{array}$$

Logo f é convexa. Suponhamos agora que f seja duas vezes diferenciável em Ω. É suficiente mostrar que $(b) \Leftrightarrow (d)$. Fixemos $x \in \Omega$ e $d \in \mathbb{R}^n$ quaisquer. Como Ω é aberto, $x + td \in \Omega$ para todo t suficientemente pequeno. Por (b),

$$f(x + td) - f(x) \geq t\langle f'(x),\, d \rangle.$$

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

$$\begin{aligned} 0 &\leq f(x+td) - f(x) - t\langle f'(x),\, d\rangle \\ &= \frac{t^2}{2}\langle f''(x)d,\, d\rangle + r(t^2). \end{aligned}$$

Dividindo por $t^2 > 0$ e fazendo $t \to 0^+$, obtemos (d).

Sejam $x,\, y \in \Omega$ quaisquer. Pelo Teorema do Valor Médio, existe $t \in (0,1)$ tal que

$$\begin{aligned} & f(y) - f(x) - \langle f'(x), y - x\rangle \\ = \; & \frac{1}{2}\langle f''(x + t(y-x))(y-x),\, y - x\rangle \geq 0, \end{aligned}$$

onde a desiguadade segue de (d). Portanto, $(d) \Leftrightarrow (b)$.

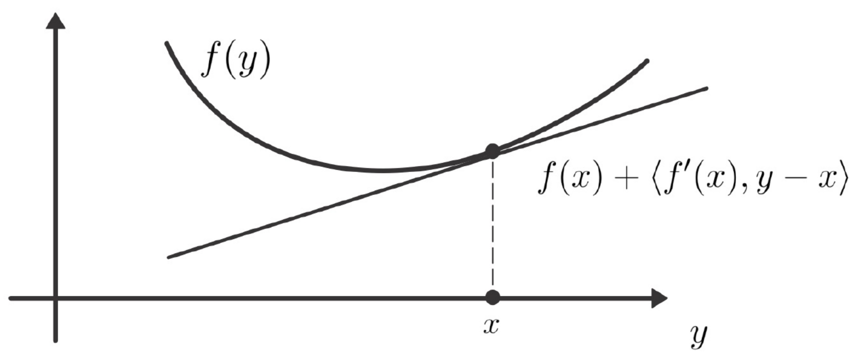

Figura 1.2: Ilustrando o Teorema 11.

Exemplo 7 *A função $f : \mathbb{R}^n \to \mathbb{R}$ definida por $f(x) = \sum_{i=1}^n e^{x_i}$ é convexa. De fato, como*

$$f''(x) = \begin{pmatrix} e^{x_1} & 0 & \cdots & 0 \\ 0 & e^{x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{x_n} \end{pmatrix},$$

temos então que $\langle f''(x)d,\ d\rangle = \sum_{i=1}^n e^{x_i} d_i^2 \geq 0 \quad \forall x, d \in \mathbb{R}^n$.

Teorema 12 *(Caracterizações de Funções Fortementes Convexas Diferenciáveis) Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \to \mathbb{R}$ uma função diferenciável em Ω, com derivada contínua. Então as propriedades seguintes são equivalentes:*

(a) A função f é fortemtente convexa em Ω com módulo $\rho > 0$;

(b) Para todo $x,\ y \in \Omega$, $f(y) \geq f(x) + \langle f'(x),\ y - x\rangle + \rho \parallel y - x \parallel^2$;

(c) Para todo $x,\ y \in \Omega$, $\langle f'(y) - f'(x), y - x\rangle \geq 2\rho \parallel y - x \parallel^2$;

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

(d) A matriz Hessiana de f *é definida positiva uniformemente em todo ponto de* Ω, *isto é,*

$$\langle f''(x)d,\ d\rangle \geq 2\rho \parallel d \parallel^2 \qquad \forall x \in \Omega, \quad \forall d \in \mathbb{R}^n.$$

Demonstração: Análogo ao teorema anterior.

■

Exemplo 8 *A função* $f : \mathbb{R}^n \to \mathbb{R}$ *definida por* $f(x) = \cosh x := \sum_{i=1}^{n} \cosh x_i$ *é fortemente convexa. De fato, como*

$$f''(x) = \begin{pmatrix} \cosh x_1 & 0 & \cdots & 0 \\ 0 & \cosh x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cosh x_n \end{pmatrix},$$

temos então que $\langle f''(x)d,\ d\rangle = \sum_{i=1}^{n} \cosh x_i d_i^2 \geq \parallel d \parallel^2 = 2 \cdot \dfrac{1}{2} \parallel d \parallel^2 \quad \forall x, d \in \mathbb{R}^n$, *já que* $\cosh x_i = \dfrac{e^{x_i} + e^{-x_i}}{2} \geq 1, \quad \forall x_i \in \mathbb{R}$.

Exemplo 9 *A função* $f : \mathbb{R}^n \to \mathbb{R}$ *definida por* $f(x) = \parallel x \parallel^2$ *é fortemente convexa módulo* $\rho = 1$.

Segue do item (d) do Teorema 12, onde, pela Hessiana, temos que $\langle f''(x)d,\ d\rangle = 2\|d\|^2$.

Exemplo 10 *A função* $f : \mathbb{R}^n \to \mathbb{R}$ *definida por* $f(x) = \parallel x \parallel^4$ *é convexa.*

De fato, a função $g : \mathbb{R}^n \to \mathbb{R}$ dada por $g(x) = \parallel x \parallel^2$ é convexa. Além disso, a função $h : [0, \infty) \to \mathbb{R}$ dada por $h(x) = x^2$ é convexa e crescente. Agora note que $f(x) = h(g(x))$ e, pelo Teorema 6, f é convexa.

Definição 9 *Uma projeção de* $x \in \mathbb{R}^n$ *sobre* $\Omega \subset \mathbb{R}^n$ *é uma solução (global) do problema* $\min\limits_{y \in \Omega} \|y - x\|$.

Teorema 13 *(Teorema da Projeção) Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo, fechado e não-vazio. Então para todo $x \in \mathbb{R}^n$, a projeção de x sobre Ω existe e é única. Além disso, $\overline{x}$ é a projeção se, e somente se, $\overline{x} \in \Omega$ e $\langle x - \overline{x},\, y - \overline{x}\rangle \leq 0, \quad \forall y \in \Omega$.*

Demonstração: Ver [1].

■

Lema 1 *(Lema de Minkowski) Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo não-vazio. Se $x \notin \overline{\Omega}$, então existem $a \in \mathbb{R}^n - \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a,\, x\rangle = c, \quad \langle a,\, y\rangle > c \quad \forall y \in \Omega.$$

Demonstração: Como Ω é convexo, temos que $\overline{\Omega}$ é convexo e fechado. Daí, seja $\overline{x}$ a projeção de x sobre $\overline{\Omega}$, cuja existência e unicidade é assegurada pelo Teorema da Projeção (Teorema 13). Definindo $a = \overline{x} - x($ $a \neq 0$ pois $x \notin \overline{\Omega})$ e $c = \langle \overline{x} - x, x\rangle$. Pela definição, $\langle a,\, x\rangle = c$. Seja $y \in \Omega$. Temos que

$$\langle a,\, y\rangle = \langle \overline{x} - x, y\rangle \geq \langle \overline{x} - x,\, \overline{x}\rangle, \tag{1.2}$$

onde a última desigualdade vale para todo $y \in \overline{\Omega}$ e, portanto, para todo $y \in \Omega$, pelo Teorema da Projeção (Teorema 13). Obtemos que

$$\begin{aligned} \langle \overline{x} - x,\, \overline{x}\rangle &= \|x - \overline{x}\|^2 + \langle \overline{x} - x,\, x\rangle \\ &= \|x - \overline{x}\|^2 + c > c. \end{aligned}$$

Combinando a última relação com (1.2) obtemos o resultado.

■

Lema 2 *Seja $\Omega \subset \mathbb{R}^n$ um conjunto convexo e não-vazio. Se $x \in fr\Omega := \overline{\Omega} \cap \overline{\mathbb{R}^n \setminus \Omega}$, então existem $a \in \mathbb{R}^n \setminus \{0\}$ e $c \in \mathbb{R}$ tais que*

$$\langle a,\; x\rangle = c, \quad \langle a,\; y\rangle \geq c \quad \forall y \in \Omega.$$

Demonstração: Como $x \in fr\Omega = fr\overline{\Omega}$, podemos escolher uma sequência $\{x^k\} \to x$ tal que $x^k \notin \overline{\Omega}, \quad \forall k$.. Pelo Lema 1, todo ponto x^k pode ser separado de Ω estritamente, isto é, para todo k existem $a^k \in \mathbb{R}^n \setminus \{0\}$ e $c_k \in \mathbb{R}$ tais que

$$\langle a^k, x^k\rangle = c_k, \quad \langle a^k,\, y\rangle > c_k \quad \forall y \in \Omega. \tag{1.3}$$

Podemos admitir (tomando uma subsequência se necessário) que $\left\{ \dfrac{a^k}{\|a^k\|} \right\} \to a \neq 0$. Obtemos que

$$\lim_{k\to\infty} \frac{c_k}{\|a^k\|} = \lim_{k\to\infty} \frac{\langle a^k, x^k\rangle}{\|a^k\|} = \langle a\, x\rangle.$$

Definindo $c = \langle a, x\rangle$, dividindo a segunda relação em (1.3) por $\|a^k\| > 0$, e passando ao limite quando $k \to \infty$, obtemos que para todo $y \in \Omega$ tem-se

$$\langle a,\, y\rangle \geq c.$$

■

Teorema 14 *(Teorema de Separação) Sejam $D_1, D_2 \subset \mathbb{R}^n$ 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*

$$\langle a, x^1\rangle \leq c \leq \langle a, x^2\rangle \quad \forall x^i \in D_i.$$

Demonstração: O conjunto $D = D_2 \setminus D_1$ é convexo e $0 \notin D$, pois $D_1 \cap D_2 = \emptyset$. Nesta situação há duas possibilidades: ou $0 \notin \overline{D}$, ou $0 \in frD$. Para separar 0 do conjunto D, no primeiro caso usamos o Lema 1 e no segundo caso usamos o Lema 2. Concluímos que existe $a \in \mathbb{R}^n \setminus \{0\}$ tal que

$$\langle a,\, x\rangle \geq 0 \quad \forall x \in D,$$

isto é, $\langle a, x^2 - x^1\rangle \geq 0$ para todos $x^2 \in D_2$, $x^1 \in D_1$. Ou ainda,

$$\langle a, x^2\rangle \geq \langle a, x^1\rangle \quad \forall x^i \in D_i.$$

Em particular, a função $\langle a, \cdot\rangle$ é limitada inferiormente em D_2 e limitada superiormente em D_1. Ainda da relação acima obtemos que

$$\gamma_2 = \inf_{x^2\in D_2} \langle a, x^2\rangle \geq \sup_{x^1\in D_1} \langle a, x^1\rangle = \gamma_1.$$

Definindo $c = \dfrac{\gamma_1 + \gamma_2}{2}$, temos que $\gamma_2 \geq c \geq \gamma_1$ e

$$\forall x^1 \in D_1, \quad \langle a, x^1 \rangle \le \sup_{x^1 \in D_1} \langle a, x^1 \rangle = \gamma_1 \le c,$$

$$\forall x^2 \in D_2, \quad \langle a,\, x^2 \rangle \ge \inf_{x^2 \in D_2} \langle a,\, x^2 \rangle = \gamma_2 \ge c.$$

■

Teorema 15 *(Teorema da Separação Estrita) Sejam $D_1, D_2 \subset \mathbb{R}^n$ 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*

$$\langle a, x^1 \rangle < c < \langle a, x^2 \rangle \qquad \forall x^i \in D_i.$$

Demonstração: Se temos separação estrita mas existe algum $x \in D_1 \cap D_2$, logo chegamos à contradição: $\langle a, x \rangle < c < \langle a, x \rangle$.

Suponhamos agora que D_2 seja limitado e $D_1 \cap D_2 = \emptyset$. Como o conjunto D_1 é convexo e fechado, o operador projeção $P_{D_1} : \mathbb{R}^n \to D_1$ é bem definido pelo Teorema da Projeção (Teorema 13) e contínuo. Logo a função

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

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

$$\min_{x \in D_2} \phi(x)$$

tem uma solução global, que denotamos por a^2. Denotamos $a^1 = P_{D_1}(a^2)$. Pelo Teorema da Projeção (Teorema 13),

$$\langle a^2 - a^1, x \rangle \le \langle a^2 - a^1, a^1 \rangle = c_1, \qquad \forall x \in D_1. \tag{1.4}$$

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

$$\begin{aligned} \|x - a^1\| &\ge dist(x, D_1) \ge \min_{x \in D_2}\{dist(x, D_1)\} = \phi(a^2) \\ &= dist(a^2, D_1) = \|a^2 - P_{D_1}(a^2)\| = \|a^2 - a^1\|, \end{aligned}$$

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

$$\langle a^2 - a^1, x\rangle \geq \langle a^2 - a^1, a^2\rangle = c_2 \quad \forall x \in D_2. \tag{1.5}$$

Definindo $a = a^2 - a^1 \neq 0$ e $c = \dfrac{c_1 + c_2}{2}$ observamos que $c_2 > c > c_1$ pois

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

Agora, o resultado segue de (1.4) e (1.5).

■

1.2.2 Subdiferencial

Constantemente lidamos com funções convexas com as quais não podemos contar com a propriedade de diferenciabilidade. Nestes casos, procuramos estudar o conjunto dos vetores que são "parecidos", como definiremos mais adiante, com o vetor gradiente e este é denominado subdiferencial.

Para começar temos o seguinte

Teorema 16 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então para todo $x \in \mathbb{R}^n$, f é diferenciável em cada direção $d \in \mathbb{R}^n$. Além disso,*

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

Demonstração: Ver [1], página 171.

■

Além disso, temos ainda

Proposição 1 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Então para quaisquer duas sequências $\{x^k\} \to x$ e $\{d^k\} \to d$, tem-se que*

$$\limsup f'(x^k; d^k) \leq f'(x; d).$$

Demonstração: Ver [1].

■

A próxima definição norteia este trabalho.

Definição 10 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa. Dizemos que $y \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in \mathbb{R}^n$ se*

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

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

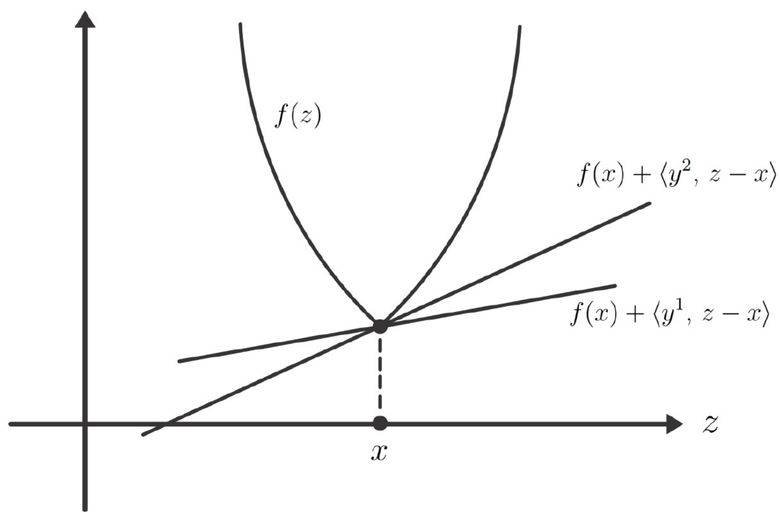

Figura 1.3: Subdiferencial

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

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

Demonstração: Primeiro note que

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

Assim, fazendo $z = x + \alpha d$, onde $d \in \mathbb{R}^n$ está fixado e $\alpha > 0$, tem-se que

$$\begin{aligned} y \in \partial f(x) \quad &\Leftrightarrow \quad f(x + \alpha d) - f(x) \geq \langle y, \alpha d \rangle \\ &\Leftrightarrow \quad \frac{f(x + \alpha d) - f(x)}{\alpha} \geq \langle y, d \rangle. \end{aligned}$$

Usando teorema anterior e fazendo $\alpha \to 0^+$ obtém-se

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

Logo,

$$\begin{aligned} \partial f(x) &= \{y \in \mathbb{R}^n;\ f'(x;d) \geq \langle y,\, d\rangle, \quad \forall d \in \mathbb{R}^n\} \\ &= \cap_{d\in\mathbb{R}^n}\{y \in \mathbb{R}^n;\ \langle y,\, d\rangle \leq f'(x;d)\}. \end{aligned} \tag{1.6}$$

Deste modo, como a interseção de semi-espaços fechados é convexo e fechado, obtém-se que $\partial f(x)$ é convexo e fechado.

Veja agora que $\partial f(x)$ é limitado. De fato, se $\partial f(x) = \emptyset$, não há o que fazer. Suponha agora que $\partial f(x)$ seja ilimitado. Neste caso existirá uma sequência $\{y^k\} \subset \partial f(x)$ de modo que $\| y^k \| \to \infty$. Considere agora a sequência auxiliar $d^k = \dfrac{y^k}{\| y^k \|}$ que, sem perda de generalidade, podemos considerar $d^k \to d$. Por (1.6),

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

Portanto,

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

o que contradiz o fato de $\| y^k \| \to \infty$. Foi usado a proposição anterior na segunda desigualdade. Deste modo, fica provado que $\partial f(x)$ é limitado. Para ver que $\partial f(x) \neq \emptyset$, fixe $d \in \mathbb{R}^n$. Pela compacidade do subdiferencial, existe $\max\limits_{y\in\partial f(x)} \langle y,\, d\rangle$ e por (1.6), tem-se que

$$f'(x;d) \geq \max_{y\in\partial f(x)} \langle y,\, d\rangle. \tag{1.7}$$

Defina os seguintes conjuntos:

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

$$D_2 = \left\{ (z,\, c) \in \mathbb{R}^n \times \mathbb{R} \;\middle|\; \begin{array}{l} c = f(x) + \alpha f'(x;\, d), \\ \\ z = x + \alpha d, \quad \alpha \in \mathbb{R}_+ \end{array} \right\},$$

Veja que o conjunto D_1 é o epígrafo de f sem sua fronteira e, portanto, é convexo. Além disso, o conjunto D_2 também é convexo.

Se $D_1 \cap D_2 \neq \emptyset$, então

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

para algum $\alpha > 0$, o que contradiz o Teorema 16. Logo, $D_1 \cap D_2 = \emptyset$. Pelo Teorema da Separação (Teorema 14), existe $(u, \beta) \in (\mathbb{R}^n \times \mathbb{R}) \setminus \{(0, 0)\}$ tal que

$$\begin{gathered} \langle u,\, z\rangle + \beta c \leq \langle u,\, x + \alpha d\rangle + \beta(f(x) + \alpha f'(x; d)) \\ \forall z \in \mathbb{R}^n, \quad \forall \alpha \in \mathbb{R}_+, \quad \forall c > f(z). \end{gathered} \tag{1.8}$$

Se fosse $\beta = 0$, obteria-se que

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

que implicaria em $u = 0$. Como $(u,\, \beta) \neq 0$, tem-se que $u \neq 0$.

Suponha que $\beta > 0$. Escolha $z = x$ e $\alpha = 0$ em (1.4), obtendo que $\beta c \leq \beta f(x)$ para todos os c tais que $c > f(x)$, o que é uma contradição. Conclui-se que $\beta < 0$. Dividindo ambos os lados de (1.8) por β, obtém-se

$$\begin{gathered} c + \langle \frac{u}{\beta}, z - x\rangle \geq f(x) + \alpha f'(x;\, d) + \alpha\langle \frac{u}{\beta},\, d\rangle \\ \forall z \in \mathbb{R}^n, \quad \forall \alpha \in \mathbb{R}_+, \quad \forall c > f(z). \end{gathered} \tag{1.9}$$

Fazendo $\alpha \to 0^+$ e $c \to f(z)^+$, obtém-se

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

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

Tomando ainda em (1.9) $\alpha = 1$ e $z = x$. e passando ao limite $c \to f(z)^+$, obtém-se que

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

Assim,

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

Por (1.7),

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

■

Exemplo 11 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ definida por $f(x) = \| x \|$. Sabemos que f não é diferenciável em 0 e, além disso, pelo Teorema 17, $\partial f(0)$ é diferente de vazio. Assim, dado $w \in \partial f(0)$,*

$$\begin{aligned} f(y) \geq f(0) + \langle w,\ y - 0\rangle \ &\Rightarrow\ \| y \| \geq \| 0 \| + \langle w,\ y\rangle \\ &\Rightarrow\ \| y \| \geq \langle w,\ y\rangle, \quad \forall y \in \mathbb{R}^n. \end{aligned}$$

Portanto, $\partial f(0) = \{w \in \mathbb{R}^n |\ \| y \| \geq \langle w,\ y\rangle, \ \ \forall y \in \mathbb{R}^n\}$, onde se pode concluir que $\partial f(0) = B[0; 1] = \{x \in \mathbb{R}^n |\ \| x \| \leq 1\}$.

Proposição 2 *Uma função convexa é diferenciável no ponto $x \in \mathbb{R}^n$ se, e somente se, o $\partial f(x) = \{f'(x)\}$.*

Demonstração: Ver [1], página 179.

■

O próximo teorema dará uma caracterização para um minimizador de uma função num determinado conjunto.

Teorema 18 *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e $D \subset \mathbb{R}^n$ um conjunto convexo. Então $\overline{x}$ é um minimizador de f em D se, e somente se,*

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

Em particular, $\overline{x}$ é um minimizador de f no $\mathbb{R}^n$ se, e somente se,

$$0 \in \partial f(\overline{x}).$$

Demonstração: Ver [1], página 180.

■

Proposição 3 ***(Continuidade do Subdiferencial)*** *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa, $\{x^k\} \to x$ e $y^k \in \partial f(x^k)$ para todo $k \in \mathbb{N}$. Então a sequência $\{y^k\}$ é limitada e todos os seus pontos de acumulação pertencem a $\partial f(x)$.*

Demonstração: Suponha que $\{y^k\}$ seja ilimitada. Daí, sem perda de generalidade, podemos considerar que $\| y^k \| \to \infty$. Definindo $d^k = \dfrac{y^k}{\| y^k \|}$ podemos considerar, a menos de uma subsequência, que $d^k \to d$. Pelo Teorema 17, temos

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

Por outro lado, pela Proposição 1,

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

que é uma contradição. Logo, $\{y^k\}$ é limitada.

Considere agora $y \in \mathbb{R}^n$ um ponto de acumulação de $\{y^k\}$, digamos $\{y^{k_j}\} \to y$. Tomando limsup quando $j \to \infty$, temos

$$\langle y,\, d \rangle \leq \limsup_{k\to\infty} f'(x^k; d) \leq f'(x; d) \quad \forall d \in \mathbb{R}^n,$$

onde foi usado novamente a Proposição 1. Usando (1.6), concluimos que $y \in \partial f(x)$.

■

Teorema 19 ***(O Subdiferencial da soma de funções convexas)*** *Sejam $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., p$, 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 para $p = 2$. Seja $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = f_1(x) + f_2(x)$. Supondo que $y^i \in \partial f_i(x)$, $i = 1, 2$, temos que

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

Somando as duas desigualdades, obtemos

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

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

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

Suponha agora que exista $y \in \partial f(x)$, de modo que $y \notin \partial f_1(x) + \partial f_2(x)$. Como $\partial f_i(x)$ é convexo, compacto e não vazio, temos que $\partial f_1(x)+\partial f_2(x)$ também é convexo compacto e não vazio. Como $y \notin \partial f_1(x) + \partial f_2(x)$, pelo Teorema da Separação Estrita, existem $a \in \mathbb{R}^n - \{0\}$ e $c \in \mathbb{R}$ tais que

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

Portanto, usando o Teorema 17, 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 isso contradiz o fato de que $y \in \partial f(x)$.

■

Teorema 20 *Seja $h : \mathbb{R}^n \to \mathbb{R}$ uma função fortemente convexa com parâmetro $\rho > 0$. Então para todo $x, y \in \mathbb{R}^n$ e $w \in \partial h(x)$ tem-se que*

$$h(y) \geq h(x) + \langle w,\, y - x\rangle + \frac{\rho}{2} \parallel x - y \parallel^2 .$$

Demonstração: De fato, por definição de subdiferencial,

$$h(\frac{x+y}{2}) \geq h(x) + \langle w,\, \frac{x+y}{2} - x\rangle,$$

ou seja,

$$h(\frac{x+y}{2}) \geq h(x) + \frac{1}{2}\langle w,\, y - x\rangle, \quad \forall y \in \mathbb{R}^n. \tag{1.10}$$

Por outro lado, por definição de fortemente convexa,

$$\frac{h(x)+h(y)}{2} - \frac{\rho}{4} \parallel x-y \parallel^2 \geq h(\frac{x+y}{2}) \quad \forall x,y \in \mathbb{R}^n \tag{1.11}$$

Agora combinando (1.10) e (1.11), temos

$$\frac{h(x)+h(y)}{2} - \frac{\rho}{4} \parallel x-y \parallel^2 \geq h(x) + \frac{1}{2}\langle w,\, y-x\rangle$$

$$\Rightarrow h(x)+h(y) - \frac{\rho}{2} \parallel x-y \parallel^2 \geq 2h(x) + \langle w,\, y-x\rangle$$

$$\Rightarrow h(y) \geq h(x) + \langle w,\, y-x\rangle + \frac{\rho}{2} \parallel x-y \parallel^2 .$$

Isso para todo $x, y \in \mathbb{R}^n$ e $w \in \partial h(x)$.

■

1.3 Funções DC

Agora estudaremos o espaço das funções que podem ser escritas como diferença de duas funções convexas e veremos o quão é importante tal espaço.

Definição 11 *Seja $f : C \to \mathbb{R}$, onde $C \subset \mathbb{R}^n$ é um conjunto convexo. Diremos que f é DC se existirem duas funções $g, h : C \to \mathbb{R}$ convexas de modo que $f = g - h$.*

Exemplo 12 *Sejam $g, h : \mathbb{R}^n \to \mathbb{R}$ dadas por $g(x) = \sum_{i=1}^n e^{x_i}$ e $h(x) = \parallel x \parallel$. Então a função $f : \mathbb{R}^n \to \mathbb{R}$ dada por $f(x) = g(x) - h(x)$ é DC.*

Exemplo 13 *A função $f(x) = \frac{1}{4} \parallel x \parallel^4 - \frac{1}{2} \parallel x \parallel^2$ é DC. Além disso, f é coerciva. De fato, as funções $g(x) = \frac{1}{4} \parallel x \parallel^4$ e $h(x) = \frac{1}{2} \parallel x \parallel^2$ são convexas. Portanto, f é DC. Além disso,*

$$\begin{aligned} f(x) &= \frac{1}{4} \parallel x \parallel^4 - \frac{1}{2} \parallel x \parallel^2 \\ &= \frac{1}{4}\left(\parallel x \parallel^2 - 1\right)^2 - \frac{1}{4}. \end{aligned}$$

Logo, $f(x) \to +\infty$ quando $\parallel x \parallel \to +\infty$. Portanto, f é coerciva. Note que f não é convexa, pois caso fosse considerando os pontos $x = (1, 0, ..., 0)$ e $y = (-1, 0, ..., 0)$ teríamos que

$$f(\frac{1}{2}x + \frac{1}{2}y) = f(0) = 0$$

e

$$\begin{aligned}
\frac{1}{2}f(x) + \frac{1}{2}f(y) &= \frac{1}{2}[f(x) + f(y)] \\
&= \frac{1}{2}[\frac{1}{4}\|x\| - \frac{1}{2}\|x\| + \frac{1}{4}\|y\| - \frac{1}{2}\|y\|] \\
&= \frac{1}{2}[\frac{1}{4} - \frac{1}{2} + \frac{1}{4} - \frac{1}{2}] \\
&= \frac{1}{2}[\frac{1}{2} - 1] \\
&= -\frac{1}{4},
\end{aligned}$$

logo,

$$f(\frac{1}{2}x + \frac{1}{2}y) > \frac{1}{2}f(x) + \frac{1}{2}f(y).$$

Teorema 21 *Seja f, f_i $(i = 1, 2, ..., m)$ funções DC. Então as seguintes funções também são DC:*

i) $\sum_{i=1}^{m} \alpha_i f_i$, *para todo* α_i *real;*

ii) $g(x) = \max\{f_1(x), ..., f_m(x)\}$;

iii) $h(x) = \min\{f_1(x), ..., f_m(x)\}$.

Demonstração:

i) É imediato.

ii) Seja $f_i = g_i + h_i \quad (i = 1, ..., m)$ uma decomposição DC de f_i. Então

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

Daí,

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

que ainda é uma função DC, desde que o máximo, assim como a soma de um número finito de funções convexas sejam funções convexas.

iii) Análogo ao item anterior.

■

Definição 12 *Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é localmente DC em $a \in \mathbb{R}^n$ se existe uma vizinhança U de a tal que f é DC em U.*

Teorema 22 *Toda função $f : \mathbb{R}^n \to \mathbb{R}$ localmente DC é DC.*

Demonstração: Ver Hartman [11].

■

Teorema 23 *Toda função $f : \mathbb{R}^n \to \mathbb{R}$ de classe C^2 é DC.*

Demonstração: Os elementos da Hessiana $\nabla^2 f$ de f são limitados em toda vizinhança fechada $B[a;\epsilon]$. Logo, para $\mu > 0$ suficientemente grande, a função $f(x) + \mu \parallel x \parallel^2$ é convexa em $B[a;\epsilon]$, desde que a Hessiana $\nabla^2 f + 2\mu I$ é semidefinida positiva em $B(a;\epsilon)$ para μ suficientemente grande. Então $(f(x) + \mu \parallel x \parallel^2) - \mu \parallel x \parallel^2$ é DC em $B[a;\epsilon]$. Daí, pelo Teorema 22, f é DC.

■

Corolário 23.1 *Toda função f de classe C^2 é DC em qualquer conjunto convexo e compacto.*

Corolário 23.2 *Para toda função contínua f num conjunto Ω, compacto e convexo, e todo $\epsilon > 0$ existe uma função DC g tal que*

$$\max_{x \in \Omega} |f(x) - g(x)| < \epsilon$$

Demonstração: Pelo Teorema da Aproximação de Weierstrass (ver [5]) existe um polinômio g satifazendo a condição exigida.

■

A seguir, apresentamos as condições necessárias para a otimização local na programação DC (Ver Teroema 2.1 de [12]).

Teorema 24 *Sejam $g, h : \mathbb{R}^n \to \mathbb{R}$ funções convexas . Se $x^* \in \mathbb{R}^n$ minimizador local de $f = g - h$, então*

$$\partial h(x^*) \subset \partial g(x^*). \tag{1.12}$$

A condição (1.12) não é fácil de verificar, pois na prática não é fácil calcular todos os subdiferenciais das funções g e h. Portanto, em algoritmos numéricos, uma forma relaxada da condição (1.12) frequentemente usada é uma condição exigindo que em $x^* \in \mathbb{R}^n$ seja um ponto crítico de f, que é definido do seguinte modo:

Definição 13 *Sejam $g, h : \mathbb{R}^n \to \mathbb{R}$ funções convexas . Dizemos que $x^* \in \mathbb{R}^n$ é um ponto crítico de $f = g - h$, se*

$$\partial h(x^*) \cap \partial g(x^*) \neq \emptyset. \tag{1.13}$$

Exemplo 14 *Sejam as função $g, h : \mathbb{R}^n \to \mathbb{R}$ definidas por:*

$$g(x) = \cosh x := \sum_{i=1}^{n} \cosh x_i \text{ e } h(x) = \|x\|^2 = \sum_{i=1}^{n} {x_i}^2.$$

Como g e h são convexas, então $f = g - h$ é uma função DC. Além disso, f é coerciva. De fato,

$$e^{x_i} = \sum_{m=0}^{\infty} \frac{x^m}{m!} \quad e \quad e^{-x_i} = \sum_{m=0}^{\infty} \frac{(-x_i)^m}{m!} \quad \Rightarrow \quad e^{x_i} + e^{-x_i} = \sum_{m=0}^{\infty} \frac{x^m}{m!} + \sum_{m=0}^{\infty} \frac{(-x_i)^m}{m!}$$
$$\Rightarrow \quad e^{x_i} + e^{x_i} = 2 \cdot \sum_{m=0}^{\infty} \frac{(x_i)^{2m}}{(2m)!}.$$

Daí,

$$\begin{aligned} \cosh x_i - x_i^2 &= \sum_{m=0}^{\infty} \frac{(x_i)^{2m}}{(2m)!} - x_i^2 \\ &\geq \frac{x_i^4}{4!} - x_i^2 \\ &= \left(\frac{x_i^2}{\sqrt{4!}} - \frac{\sqrt{4!}}{2} \right)^2 - 12. \end{aligned}$$

Portanto,

$$\begin{aligned} \cosh x - \parallel x \parallel^2 &= \sum_{n=1}^{n} \left(\cosh x_i - x_i^2 \right) \\ &\geq \sum_{i=1}^{n} \left(\left(\frac{x_i^2}{\sqrt{4!}} - \frac{\sqrt{4!}}{2} \right)^2 - 12 \right). \end{aligned}$$

Como $\| x \|^2 \to \infty$, *então* $x_i^2 \to \infty$, *para pelo menos algum* i. *Logo,* $f(x) \to +\infty$ *e* f *é coerciva. Assim, existe* x^* *mínimo de* f *e, como* g *e* h *são diferenciáveis, temos ainda que* $\sinh x_i^* = 2x_i^*$. *Neste caso, concluimos que* $x^* = 0$ *é um ponto crítico de* f. *Note que* f *não é convexa. De fato,*

$$f''(x) = \begin{pmatrix} \cosh x_1 - 2 & 0 & \cdots & 0 \\ 0 & \cosh x_2 - 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \cosh x_n - 2 \end{pmatrix},$$

logo

$$f''(0) = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix},$$

ou seja, a Hessiana de f *no ponto 0 é negativa definida.*

Observação 1: Dada uma função DC $f = g - h$ podemos considerar, sem perda de generalidade, que as funções componentes g e h sejam fortemente convexas. De fato, se para tal decomposição DC as funções componentes não forem fortemente convexas, podemos somar e subtrair uma terceira função fortemente convexa, digamos ψ, e teremos o resultado, pois f teria a seguinte decomposição $f = [g + \psi] - [h + \psi]$.

Capítulo 2

Algoritmo do Ponto Proximal Inercial para Programação DC

Nesta seção apresentamos o algoritmo para resolver o problema de minimização DC descrito como:

$$(P) \qquad \min_{x \in \mathbb{R}^n} f(x),$$

onde $f = g - h$, com g e h convexas. Denotamos por S o conjunto dos pontos críticos de f e consideremos $S \neq \emptyset$. Pela **Observação 1** iremos considerar que a função h é fortemennte convexa com parâmetro $\rho > 0$.

2.1 Algoritmo

PASSO 1: Seja $x^0 \in \mathbb{R}^n$ o ponto inicial, $\gamma_k \in [0, \frac{\rho}{2})$ e uma sequência limitada de números positivos $\{\mu_k\}$ tal que $0 < a \leq \mu_k \leq b$. Defina $x^{-1} = x^0$,

PASSO 2: Para $k = 0, 1, 2, ...$, defina $d^k = \gamma_k(x^k - x^{k-1})$ e calcule

$$w^k \in \partial h(x^k) \quad e \quad y^k = x^k + \mu_k(w^k + d^k). \tag{2.1}$$

PASSO 3: Encontre $x^{k+1} \in \mathbb{R}^n$ de modo que

$$x^{k+1} = argmin\left\{ g_k(x) := g(x) + \frac{1}{2\mu_k} \parallel x - y^k \parallel^2 \right\}. \tag{2.2}$$

CRITÉRIO DE PARADA: Se $x^k = x^{k+1}$ e $d^k = 0$, pare. Caso contrário, faça $k := k + 1$ e retorne ao *Passo 2*.

2.2 Boa Definição

Analisaremos agora a boa definição do algoritmo.

Teorema 25 *A sequência $\{x^k\}$ gerada pelo Algoritmo está bem definida.*

Demonstração: A demonstração será feita por indução. Como $x^0 \in \mathbb{R}^n$ é um ponto dado, suponha que já seja encontrados os pontos $x^0, x^1, ..., x^k$. Pelo Teorema 17 temos que $\partial h(x^k) \neq \emptyset$ e, portanto, existe $w^k \in \partial h(x^k)$, consequentemente existe $y^k = x^k + \mu_k(w^k + d^k)$. Agora, seja $v^k \in \partial g(y^k)$, assim

$$\begin{aligned} g(x) &\geq g(y^k) + \langle v^k,\ x - y^k \rangle \\ &\geq g(y^k) - \parallel v^k \parallel \cdot \parallel x - y^k \parallel . \end{aligned}$$

Daí,

$$\begin{aligned} g_k(x) &= g(x) + \frac{1}{2\mu_k} \parallel x - y^k \parallel^2 \\ &\geq g(y^k) - \parallel v^k \parallel \cdot \parallel x - y^k \parallel + \frac{1}{2\mu_k} \parallel x - y^k \parallel^2 \\ &= g(y^k) + \parallel x - y^k \parallel \left[\frac{1}{2\mu_k} \parallel x - y^k \parallel - \parallel v^k \parallel \right]. \end{aligned}$$

Portanto, $g_k(x)$ é continua e coerciva. Logo, pelo Corolário 1.2, existe $x^* \in \mathbb{R}^n$ mínimo de g_k. Sendo g_k estritamente contexa (Teorema 7), temos que x^k é único (Teorema 10). Faça $x^{k+1} = x^*$.

■

2.3 Análise da Convergência

Vamos analisar primeiramente o critério de parada.

Proposição 4 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo. Suponha que o Algoritmo termine na iterada k, então $\{x^k\}$ é um ponto crítico de f.*

Demonstração: Como $d^k = 0$, temos que $w^k = \dfrac{1}{\mu_k}(y^k - x^k) \in \partial h(x^k)$. Além disso,

$$\begin{aligned}\partial g_k(x) &= \partial g(x) + \partial(\frac{1}{2\mu_k} \parallel x - y^k \parallel^2) \\ &= \partial g(x) + \{\frac{1}{\mu_k}(x - y^k)\}.\end{aligned}$$

Observando que $x^k = x^{k+1}$ satifaz a condição (2.2) e, portanto, $0 \in \partial g_k(x^k)$ temos que existe $\overline{w} \in \partial g(x^k)$ de modo que

$$0 = \overline{w} + \frac{1}{\mu_k}(x^k - y^k) = \overline{w} - w^k,$$

ou seja, $w^k \in \partial g(x^k)$ implicando $\partial g(x^k) \cap \partial h(x^k) \neq \emptyset$. Portanto, x^k é um ponto crítico.

■

A partir de agora assumiremos que $x^k \neq x^{k+1}$ para todo $k \in \mathbb{N}$ e faremos a análise de convergência.

Proposição 5 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo. Então vale a seguinte desigualdade*

$$f(x^{k+1}) + \left(\frac{1}{b} + \frac{\rho}{4}\right) \parallel x^{k+1} - x^k \parallel^2 \leq f(x^k) + \frac{\rho}{4} \parallel x^k - x^{k-1} \parallel^2 .$$

Em particular, a sequência $\{f(x^k) + \frac{\rho}{4} \parallel x^k - x^{k-1} \parallel^2\}$ é decrescente.

Demonstração: Por (2.1) temos que

$$w^k = \frac{1}{\mu_k}(y^k - x^k) - d^k \in \partial h(x^k).$$

Pelo Teorema 20,

$$h(x) \geq h(x^k) + \langle \frac{1}{\mu_k}(y^k - x^k) - d^k, x - x^k \rangle + \frac{\rho}{2} \parallel x^k - x \parallel^2$$

Fazendo $x = x^{k+1}$,

$$h(x^{k+1}) \geq h(x^k) + \langle \frac{1}{\mu_k}(y^k - x^k) - d^k, x^{k+1} - x^k \rangle + \frac{\rho}{2} \parallel x^k - x^{k+1} \parallel^2 \tag{2.3}$$

Como x^{k+1} minimiza $g_k(x)$ temos que existe $\xi^k \in \partial g(x^{k+1})$ tal que

$$\xi^k = \frac{1}{\mu_k}(y^k - x^{k+1}).$$

Pela convexidade de g,

$$g(x) \geq g(x^{k+1}) + \langle \frac{1}{\mu_k}(y^k - x^k), x - x^{k+1} \rangle.$$

Fazendo $x = x^k$,

$$g(x^k) \geq g(x^{k+1}) + \langle \frac{1}{\mu_k}(y^k - x^k), x^k - x^{k+1} \rangle. \tag{2.4}$$

Agora, somando (2.3) e (2.4),

$$\begin{aligned} h(x^{k+1}) + g(x^k) \quad &\geq \quad h(x^k) + \langle \frac{1}{\mu_k}(y^k - x^k) - d^k,\, x^{k+1} - x^k \rangle + \frac{\rho}{2} \parallel x^{k+1} - x^k \parallel^2 \\ &+ \quad g(x^{k+1}) + \langle \frac{1}{\mu_k}(y^k - x^{k+1},\, x^k - x^{k+1} \rangle, \end{aligned}$$

implicando em

$$\begin{aligned} f(x^k) - f(x^{k+1}) + \langle d^k,\, x^{k+1} - x^k \rangle \quad &\geq \quad \frac{1}{\mu_k}\left[\langle y^k - x^k,,\; x^{k+1} - x^k \rangle + \langle y^k - x^{k+1},\, x^k - x^{k+1} \rangle\right] \\ &+ \quad \frac{\rho}{2} \parallel x^{k+1} - x^k \parallel^2 \\ &= \quad \frac{1}{\mu_k}\left[\langle x^k - y^k,\, x^k - x^{k+1} \rangle + \langle y^k - x^{k+1},\, x^k - x^{k+1} \rangle\right] \\ &+ \quad \frac{\rho}{2} \parallel x^{k+1} - x^k \parallel^2 \\ &= \quad \frac{1}{\mu_k}\langle x^k - x^{k+1},\, x^k - x^{k+1} \rangle + \frac{\rho}{2} \parallel x^{k+1} - x^k \parallel^2 \\ &= \quad \left(\frac{1}{\mu_k} + \frac{\rho}{2}\right) \parallel x^{k+1} - x^k \parallel^2, \end{aligned}$$

ou seja,

$$f(x^k) - f(x^{k+1}) + \langle d^k,\, x^{k+1} - x^k \rangle \quad \geq \quad \left(\frac{1}{\mu_k} + \frac{\rho}{2}\right) \parallel x^{k+1} - x^k \parallel^2 .$$

Como $0 < a < \mu_k \leq b$, segue que

$$f(x^k) - f(x^{k+1}) + \langle d^k,\, x^{k+1} - x^k \rangle \geq \left(\frac{1}{b} + \frac{\rho}{2} \right) \| \, x^{k+1} - x^k \, \|^2 .$$

Agora, note que

$$\begin{aligned} \langle d^k,\, x^{k+1} - x^k \rangle &= \langle \gamma_k (x^{k-1} - x^k),\, x^{k+1} - x^k \rangle \\ &\leq \frac{\gamma_k}{2} \| \, x^{k-1} - x^k \, \|^2 + \frac{\gamma_k}{2} \| \, x^{k+1} - x^k \, \|^2 . \end{aligned}$$

Como $\gamma_k \in [0, \frac{\rho}{2})$, temos que

$$\begin{aligned} \left(\frac{1}{b} + \frac{\rho}{2} \right) \| \, x^{k+1} - x^k \, \|^2 &\leq f(x^k) - f(x^{k+1}) + \langle d^k,\, x^{k+1} - x^k \rangle \\ &\leq f(x^k) - f(x^{k+1}) + \frac{\rho}{4} \| \, x^{k-1} - x^k \, \|^2 + \frac{\rho}{4} \| \, x^{k+1} - x^k \, \|^2 . \end{aligned}$$

Implicando em,

$$\left(\frac{1}{b} + \frac{\rho}{4} \right) \| \, x^{k+1} - x^k \, \|^2 \leq f(x^k) - f(x^{k+1}) + \frac{\rho}{4} \| \, x^{k-1} - x^k \, \|^2 .$$

Logo,

$$f(x^{k+1}) + \frac{\rho}{4} \| \, x^{k+1} - x^k \, \|^2 \leq f(x^k) + \frac{\rho}{4} \| \, x^{k-1} - x^k \, \|^2,$$

Portanto, $\{f(x^k) + \frac{\rho}{4} \| \, x^k - x^{k-1} \, \|^2\}$ é decrescente.

■

Teorema 26 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo. Se $\inf_{x \in \mathbb{R}^n} f(x) = f^* > -\infty$, então $\lim_{k \to \infty} \| \, x^{k+1} - x^k \, \| = 0$.*

Demonstração:

Pela Proposição 5 a sequência $\{f(x^k) + \frac{\rho}{4} \| \, x^k - x^{k-1} \, \|^2\}$ é decrescente. Além disso,

$$f(x^k) \leq f(x^k) + \frac{\rho}{4} \| \, x^k - x^{k-1} \, \|^2 \leq f(x^0) + \frac{\rho}{4} \| \, x^{-1} - x^0 \, \|^2 = f(x^0) \tag{2.5}$$

Daí, pela desigualdade da Proposição 5, temos

$$
\begin{aligned}
\sum_{k=0}^{n} \frac{1}{b} \parallel x^{k+1} - x^k \parallel^2 \;&\leq\; \sum_{k=0}^{n} \Big(f(x^k) + \frac{\rho}{4} \parallel x^k - x^{k-1} \parallel^2 - f(x^{k+1}) + \frac{\rho}{4} \parallel x^{k+1} - x^k \parallel^2 \Big) \\
&= \; f(x^0) - \Big(f(x^{n+1}) + \frac{\rho}{4} \parallel x^{n+1} - x^n \parallel^2 \Big).
\end{aligned}
$$

De (2.5) e que $\inf_{x \in \mathbb{R}^n} f(x) = f^* > -\infty$,

$$
\begin{aligned}
\sum_{k=1}^{\infty} \frac{1}{b} \parallel x^{k+1} - x^k \parallel^2 \;&\leq\; f(x^0) - \lim_{n \to \infty} \Big(f(x^{n+1}) + \frac{\rho}{4} \parallel x^{n+1} - x^n \parallel^2 \Big) \\
&\leq\; f(x^0) - \inf_{x \in \mathbb{R}^n} f(x),
\end{aligned}
$$

onde a última desigualdade vem da equação (2.5). Portanto, $\lim_{k \to \infty} \parallel x^{k+1} - x^k \parallel = 0$.

■

Teorema 27 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo. Suponha válida as hipóteses do Teorema 26. Se a sequência gerada pelo Algoritmo é limitada, então todo ponto de acumulação de $\{x^k\}$ é um ponto crítico de f.*

Demonstração: Sendo $\{x^k\}$ limitada, existe uma $x^{k_j} \to x$. a Proposição 3 garante que $\{w^{k_j}\}$ e consequentemente $\{y^{k_j}\}$ são limitadas. Como também μ_k é limitada, podemos supor, a menos de uma subsequência, que $w^{k_j} \to w$, $y^{k_j} \to y$ e $\mu_{k_j} \to \mu$.
Do Teorema 26 segue que $x^{k_j+1} \to x$ e $x^{k_j-1} \to x$. Por (2.1),

$$
w^{k_j} = \frac{1}{\mu_{k_j}}(y^{k_j} - x^{k_j}) - d^{k_j}.
$$

Fazendo $j \to \infty$, obtemos

$$
w = \frac{1}{\mu}(y - x),
$$

Pois $d^{k_j} \to 0$. Por (2.2) temos que

$$
\frac{1}{\mu_{k_j}}(y^{k_j} - x^{k_j+1}) \in \partial g(x^{k_j+1})
$$

Fazendo $j \to \infty$, obtemos

$$
\frac{1}{\mu}(y - x) = w \in \partial g(x).
$$

Portanto, x é um ponto crítico de f.

$\blacksquare$

Corolário 27.1 *Seja $\{x^k\}$ a sequência gerada pelo Algoritmo. Suponha válido as hipóteses do Teorema 26. Se $L_f(f(x_0))$ for limitado, então todo ponto de acumulação de $\{x^k\}$ é ponto crítico de f.*

Demonstração: Se $L_f(f(x^0))$ for limitado, então, pela equação (2.5), $\{x^k\}$ também é limitada. Portanto, pelo Teorema 27, todo ponto de acumulação da sequência é ponto crítico.

$\blacksquare$

Exemplo 15 *As funções DC, $f_1(x) = \dfrac{1}{4} \| x \|^4 - \dfrac{1}{2} \| x \|^2$ e $f_2(x) = \cosh x - \| x \|^2$ satisfazem todas as hipóteses exigidas na análise de convergência do Algoritmo. Da coercividade (ver exemplos e 13 e 14) temos que $S \neq \emptyset$, $\inf\limits_{x \in \mathbb{R}^n} f_i(x) > -\infty$, $L_{f_i}(x^0)$ é limitado. Portanto, o Algoritmo do Ponto Proximal Inercial gera uma sequência $\{x^k\}$ em que cada ponto de acumulação x^* é um ponto crítico dessas funções.*

Capítulo 3

Considerações Finais

Neste trabalho estudamos o Algoritmo do Ponto Proximal Inercial para programação DC que visa encontrar pontos críticos de funções $f : \mathbb{R}^n \to \mathbb{R}$ que são escritas como diferença de funções convexas, ou seja, $f = g - h$, onde as funções componentes g e h são convexas. Mostramos que a sequência gerada pelo Algoritmo está bem definida; o critério de parada está bem definido; sob hipótese do Teorema 20, obtemos que $\{f(x^k) + \dfrac{1}{2\mu_k} \parallel x^k - x^{k-1} \parallel^2\}$ é decrescente; sob as hipóteses do Teorema 20, $f^* \in \mathbb{R}$, $S \neq \emptyset$ e $\{x^k\}$ limitada, obtemos que os pontos de acumulação de $\{x^k\}$ são pontos críticos de f. Além disso, sob todas as hipóteses anteriores e que $L_f(f(x^0))$ limitado, garantimos que os pontos de acumulação de $\{x^k\}$ são críticos. Neste sentido, uma melhoria dos resultados seria mostrar que a sequência $\{x^k\}$ convergiria para um ponto crítico de f.

Referências Bibliográficas

[1] A. F. Izmailov e M. V. Solodov, *Otimização volume 1. Condições de otimalidade, elementos de análise convexa e de dualidade.*. 3 ed. Rio de Janeiro: IMPA, 2014.

[2] A. F. Izmailov e M. V. Solodov, *Otimização volume 2. Métodos computacionais,*..IMPA, Rio de Janeiro, 2007 (segunda edição 2012).

[3] Alvarez, Felipe and Attouch, Hedy - An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping - Set-Valued Analysis - pages 3-11 - Sprigner, 2011.

[4] Andrade, João; Lopes, Jurandir; Sousa, João Carlos - *Um algoritmo inercial para funções DC em variedades de Hadamard.* XL CNMAC 2021, 2021, Cuiabá. São Carlos/SP: Sociedade de Matemática Aplicada e Computacional, 2021. v. 8.

[5] Cipolatti, Rolci - *Cálculo Avançado 2ª Edição.* Textos Universitários - SBM, 2020.

[6] da Silva Souza, Sissy and Oliveira, Paulo Roberto and da Cruz Neto, João Xavier e Soubeyran, Antonie - Um Algoritmo Proximal com Distância de Bregman para Otimização Quuase Convexa - XXXIX SBPO, 2007.

[7] de Lima, Ronaldo Freire - *Topologia e Análise no espaço Rn.*Textos Universitário - SBM, 2015.

[8] de Oliveira Souza, Joao Carlos - Convergência de Métodos de Descida para Funções não Convexas com Aplicações a Teoria de Comportamento - Rio de Janeiro: UFRJ/COPPE, 2016.

[9] de Oliveira, W., Tcheou, M.P. - *An Inertial Algorithm for DC Programming.* Set-Valued Var. Anal 27, 895–919 (2019).

[10] Fernandes, M. Pontes - *Otimização por funções representavéis com a diferença entre funções convexas com aplicação em um problema de arranjo Físico.* Dissertação de Mestrado, UFPB. 2012.

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

[12] Joki, Kaisa and Bagirov, A.M. and Karmitsa, Napsu and Mäkelä, Marko and Taheri, Sona - *Double Bundle Method for finding Clarke Stationary Points in Nonsmooth DC Programming* SIAM Journal on Optimization, vol 28. 2018.

[13] Lima, Elon Lages - *Curso de análise vol.2*/Elon Lages Lima. 11.ed. Rio de Janeiro: IMPA, 2015. 1997.

[14] Lima, Elon Lages - *Álgebra Linear 8ª Edição.*Coleção Matemática Universitátia - SBM, 2018.

[15] Lima, Fernando Santana - *Um Método Ponto Proximal para Minimização da Diferença de Funções Convexas.* Projeto Euclides - Discertação de Mestrado - UPFI, 2017.

[16] Martinet, B., "Regularisation d'in´equations variationelles par approximations succesives", Rev. Fran¸caise d'Inform. Recherche Oper., v. 4, pp. 154– 159, 1970.

[17] Rockafellar, R. T., "Monotone operators and the proximal point algorithm", SIAM J. Control. Optim., v. 14, pp. 877–898, 1976

[18] Souza, Lara Thaise Bezerra Lima - Método do Ponto Proximal com Distância de Bregman para Problemas de Minimização Quase Convexos - UFRJ, 2017.

[19] Sun, W., Sampaio, R. J. B., Candido, M. A. B., "A proximal point algorithm for minimization of DC function", J. of Comput. Math., v. 21, pp. 451–462, 2003.

[20] Tao, P.D., An, L.T.H. - *Convex analysis approach to DC programming: theory, algorithms and applications..* Acta Math.Vietnamica 22(1), 289–355 (1997).

[21] Tuy, Hoang - *Convex Analysis and Global Optimization.* Springer Optimization and Its Applications 110, 2010.