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

Um Método Interior Proximal Linearizado para Programação DC

Felipe de Sales Cavalcante

Teresina - 2019

Felipe de Sales Cavalcante

Dissertação de Mestrado:

Um Método Interior Proximal Linearizado 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 - 2019

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

Um Método Interior Proximal Linearizado para Programação DC

Felipe de Sales Cavalcante

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 20 de Agosto de 2019.

Banca Examinadora:

Prof. Dr. Jurandir de Oliveira Lopes - Orientador

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

Prof. Dr. Pedro Jorge Sousa dos Santos (UFDPar)

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

C377m Cavalcante, Felipe de Sales.
Um método interior proximal linearizado para programação DC / Felipe de Sales Cavalcante. – Teresina, 2019.
48 f.

Dissertação (Mestrado) – Universidade Federal do Piauí, Centro de Ciências da Natureza, Pós-Graduação em Matemática, 2019.
Orientador: Prof. Dr. Jurandir de Oliveira Lopes

1. Otimização. 2. Método do Ponto Proximal. 3. Programação DC. I. Título.

CDD 519.3

Bibliotecária : Caryne Maria da Silva Gomes / CRB3-1461

Dedico este trabalho a meus pais Francisca e Antônio José, a minha esposa Rosana, a minha filha Alicia Danielly, e familiares.

Agradecimentos

Agradeço primeiramente a DEUS pela vida, e por ter me dado força e coragem para persistir no meu objetivo, principalmente nos momentos mais difíceis.

A minha mãe que sempre me incentivou e me apoiou, ao longo dessa jornada, sempre acreditando que eu era capaz. Ao meu avó que sempre confiou em mim.

A minha esposa Rosana, que também sempre acreditou em mim e me apoiou durante esta jornada, tendo paciência. A minha filha Alicia Danielly (Dany) que sentiu muito a minha ausência.

A todos os meus amigos especialmente os alunos do mestrado, pelo companheirismo, amizade e solidariedade.

Aos professores do Departamento de Matemática da UFPI, especialmente aos professores: Jurandir, pela paciência, compreensão e incentivo, professor Newton por ter acreditado em mim, professores João Carlos e Pedro Jorge, pelas sugestões e correções.

Agradeço a CAPES, pelo apoio financeiro.

"*O SENHOR dos Exércitos jurou, dizendo: Como pensei, assim sucederá, e como determinei, assim se efetuará*".

Isaías 14;24.

Resumo

Apresentamos um Método Interior Proximal para resolver problemas não convexos, problemas de otimização onde a função objetivo é dada pela diferença de duas funções convexas (função DC). Para este fim, consideramos um método proximal linearizado com distância proximal como regularização. Análise de convergência de escolhas particulares de distância proximal, como distância proximal de segunda ordem e distância de Bregman, são consideradas.

Abstract

We present an interior proximal method for solving constrained nonconvex optimization problems where the objective function is given by the difference of two convex function (DC function). To this end, we consider a linearized proximal method with a proximal distance as regularization. Convergence analysis of particular choices of the proximal distance as second-order homogeneous proximal distances and Bregman distances are considered.

Sumário

Introdução

Nesta dissertação abordaremos o problema de encontrar o ponto crítico de uma função não convexa $f : \mathbb{R}^n \to \mathbb{R}$ restrita a um conjunto convexo, fechado e não vazio C. Consideremos o seguinte problema:

$$\min_{x \in C} f(x) \tag{1}$$

que é equivalente à

$$\min_{x \in \mathbb{R}^n} \left(f(x) + \delta_C(x) \right),$$

onde $\delta_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ é dada por

$$\delta_C(x) = \begin{cases} 0, & \text{se } x \in C \\ \infty, & \text{se } x \notin C, \end{cases}$$

é a função indicadora do conjunto C, a qual é própria, convexa e semicontínua inferior (ver Seção 1 do Capítulo 1 para maiores detalhes). Focaremos especialmente nas subclasses de funções localmente Lipschitz (não-convexas) $f : \mathbb{R}^n \to \mathbb{R}$ a qual pode ser escrita como diferença de duas funções convexas (Funções DC), isto é,

$$f(x) = g(x) - h(x)$$

com funções componentes $g, h : \mathbb{R}^n \to \mathbb{R}$ convexas e semicontínuas inferior. O problema (1) tem diversas aplicações em Ciências e Engenharias, além disso, generaliza o problema de minimização convexa (quando $h \equiv 0$). De modo geral o problema (1) não é fácil de ser resolvido, sendo assim nosso objetivo é encontrar $s \in \partial^\circ f(\overline{x})$ tal que

$$\langle s, y - \overline{x} \rangle \geqslant 0, \forall y \in C, \tag{2}$$

onde $\overline{x}$, é chamado de ponto crítico de f e $\partial^\circ f$ é denotado como subdiferencial de Clarke de f (ver seção 1 do Capítulo 1 para maiores detalhes). Vamos denotar por S o conjunto de todos os pontos críticos de f que nesta dissertação vamos considerar diferente de vazio.

Em 1986, o Algoritmo para funções DC foi introduzido por Pham Dinh e El Bernoussi [18]. E recentemente desenvolvido extensivamente para se tornar agora um campo de pesquisa frutífero em ambos pontos de vista teórico e de aplicação. Nosso foco principal é o Método do Ponto Proximal. Este método foi introduzido por Martinet [8] e Moreau [9] e popularizado por Rockafellar [13] no contexto de operadores monótonos. Vários autores desenvolveram o método de ponto proximal para resolver problema de programação DC, a saber, [17, 11, 15, 7]. Para resolver o problema (2) utilizaremos o Método Interior Proximal Linearizado desenvolvido por Cruz Neto et al.[7], e mostraremos a análise de convergência utilizando a distância proximal como regularização.

Na seção 1 apresentaremos algumas definições e propriedades básicas de funções convexas. Na seção 2 apresentaremos definições e propriedades sobre distância proximal e sua distância proximal induzida associada, bem como Distância de Bregman e Distância Proximal Homogênea de Segunda Ordem, exibiremos exemplos conhecidos de distância proximal. Na seção 3 apresentaremos o Método Interior Proximal (MIP) e faremos a análise de convergência para Distância Proximal Homogênea de Segunda Ordem e distância de Bregman.

Capítulo 1

Noções Preliminares

1.1 Fatos Básicos

Neste capítulo, apresentaremos noções básicas de análise convexa que serão abordadas ao longo dessa dissertação. Dado $C \subset \mathbb{R}^n$, denotaremos por $\mathtt{int}\ C$, $\mathtt{fr}\ C$ e $\overline{C}$ como o interior, a fronteira e o fecho de C, respectivamente. Consideremos $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n;\ x_j \geqslant 0,\ j \in \{1, \cdots, n\}\}$ e $\mathbb{R}^n_{++} = \mathtt{int}\ \mathbb{R}^n_+$.

Definição 1.1.1. *Uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *é dita própria se o domínio efetivo, denotado por* $\mathtt{dom}\ f := \{x \in \mathbb{R}^n\ :\ f(x) < \infty\}$ *é um conjunto não vazio.*

Definição 1.1.2. *Dizemos que a função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *é semicontínua inferiormente (sci) no ponto* $x \in \mathtt{dom}\ f$*, quando para qualquer sequência* $\{x^k\} \subset \mathbb{R}^n$ *tal que* $\lim_{k\to\infty} x^k = x$*, tem-se*

$$f(x) \leqslant \liminf_{k\to\infty} f(x^k).$$

Definição 1.1.3. *Dizemos que a função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *é semicontínua superiormente (scs) no ponto* $x \in \mathtt{dom}\ f$*, quando para qualquer sequência* $\{x^k\} \subset \mathbb{R}^n$ *tal que* $\lim_{k\to\infty} x^k = x$*, tem-se*

$$\limsup_{k\to\infty} f(x^k) \leqslant f(x).$$

Definição 1.1.4. *Dizemos que a função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *é coerciva no conjunto* C *se para toda sequência* $\{x^k\} \subset C$ *tal que* $\lim_{k\to\infty} \|x^k\| = \infty$ *ou* $\lim_{k\to\infty} x^k = \bar{x} \in \overline{C} \setminus C$ *implicar*

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

Teorema 1.1.1. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$*, própria, sci e coerciva em* $C \subset \operatorname{dom} f$*, com* $C \neq \emptyset$*, então* f *possui minimizador global em* C*.*

Demonstração. Seja $\overline{x} \in C$, então o conjunto de nível $L(f(\overline{x})) = \{x \in C \,:\, f(x) \leqslant f(\overline{x})\}$ é diferente de vazio. Além disso,

1) $L(f(\overline{x}))$ é limitado.

Com efeito, suponha por absurdo que $L(f(\overline{x}))$ não seja limitado, isto é, $\exists \{x^k\} \subset L(f(\overline{x}))$ tal que $\|x^k\| \to \infty$. Temos então,

$$f(x^k) \leqslant f(\overline{x}),$$

usando a coercividade de f,

$$\infty = \limsup_{k\to\infty} f(x^k) \leqslant f(\overline{x}).$$

Absurdo, logo $L(f(\overline{x}))$ é limitado.

2) $L(f(\overline{x}))$ é fechado.

De fato, suponha por absurdo que $\exists \{x^k\} \subset L(f(\overline{x}))$ tal que $x^k \to x^0 \in \overline{L(f(\overline{x}))} \setminus L(f(\overline{x}))$. Observe que,

$$L(f(\overline{x})) = C \cap \{x \in \mathbb{R}^n \,:\, f(x) \leqslant f(\overline{x})\},$$

o conjunto $\{x \in \mathbb{R}^n \,:\, f(x) \leqslant f(\overline{x})\}$ é fechado (pois f é sci). Então $x^k \to x^0 \in \overline{C} \setminus C$, pela coercividade de f em C, temos que

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

isto contradiz o fato de $\{x^k\} \subset L(f(\overline{x}))$, logo $L(f(\overline{x}))$ é fechado. Assim $L(f(\overline{x}))$ é compacto e não vazio. Por Weierstrass, existe um minimizador $\hat{x} \in L(f(\overline{x}))$. Seja $x \in C \setminus L(f(\overline{x}))$, logo

$$f(\hat{x}) \leqslant f(\overline{x}) < f(x),$$

isto mostra que $\hat{x}$ é minimizador global em C. □

Definição 1.1.5. *Seja* $C \subset \mathbb{R}^n$ *um conjunto convexo. Dizemos que* $f : C \to \mathbb{R} \cup \{\infty\}$ *é convexa em* C *se* $\forall x, y \in C$ *e* $\alpha \in [0, 1]$

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

A função f *diz-se estritamente convexa quando a desigualdade acima é estrita para todos* $x \neq y$ *e* $\alpha \in (0, 1)$*.*

Teorema 1.1.2. *Sejam* $C \subset \mathbb{R}^n$ *e* $f_i : C \to \mathbb{R} \cup \{\infty\}$, $i = 1, \ldots, p$ *funções convexas em* C*. Então para quaisquer* $\mu_i \in \mathbb{R}_+$*, a função* $f : C \to \mathbb{R} \cup \{\infty\}$ *dada por*

$$f(x) = \sum_{i=1}^{p} \mu_i f_i(x),$$

é convexa em C*.*

Demonstração. Para $x, y \in C$ e $\alpha \in [0,1]$ quaisquer, temos que

$$\begin{aligned} f(\alpha x + (1-\alpha)y) &= \sum_{i=1}^{p} \mu_i f_i(\alpha x + (1-\alpha)y) \\ &\leqslant \sum_{i=1}^{p} \mu_i \left(\alpha f_i(x) + (1-\alpha) f_i(y)\right) \\ &= \alpha \sum_{i=1}^{p} \mu_i f_i(x) + (1-\alpha) \sum_{i=1}^{p} \mu_i f_i(y) \\ &= \alpha f(x) + (1-\alpha) f(y). \end{aligned}$$

□

Teorema 1.1.3. *(Teorema de Minimização Convexa) Sejam* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função convexa num conjunto convexo, não vazio* $C \subset \mathrm{dom}(f)$*. Então:*

1. *Todo minimizador local do problema* $\min f(x)$ *sujeito a* $x \in C$ *é global;*

2. *O conjunto dos minimizadores é convexo;*

3. *Se* f *é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração. (1) Suponha por absurdo que exista um minimizador local $\overline{x}$ tal que não seja global, isto é,

$$\exists y \in C : \; f(y) < f(\overline{x}).$$

Considere $z(\alpha) = \alpha y + (1-\alpha)\overline{x} \in C, \; \forall \alpha \in [0,1]$. Por hipótese f é convexa em C, daí

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

Para $\alpha > 0$ suficientemente pequeno, temos que $z(\alpha)$ está próximo de $\bar{x}$, além disso temos que,

$$f(z(\alpha)) < f(\bar{x}).$$

Absurdo, pois isso contraria a hipótese de $\bar{x}$ ser minimizador local. Então todo minimizador local é global.

(2) Sejam $S \subset C$ o conjunto dos minimizadores de f em C e $v \in \mathbb{R}$ o valor ótimo do problema, ou seja, $f(x) = v, \ \forall x \in S$. Sejam $x, y \in S$, pela convexidade de f, para todo $\alpha \in [0, 1]$, temos que

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

Isto mostra que $(1-t)x + ty \in S, \ \forall \alpha \in [0, 1]$, ou seja, S é convexo.

(3) Suponha por absurdo que existam $x, \bar{x} \in S \subset C$ com $x \neq \bar{x}$. Por hipótese f é estritamente convexa, logo para $\alpha \in [0, 1]$,

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

Absurdo, pois contraria o fato de v ser o valor ótimo. Portanto o minimizador é único. $\square$

Definição 1.1.6. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função. Dizemos que* f *é localmente Lipschitz em* $x \in \mathbb{R}^n$ *se existem uma constante* $K_x > 0$ *e* $\varepsilon > 0$ *tais que*

$$|f(y) - f(z)| \leqslant K_x \|y - z\|, \ \forall y, z \in B(x, \varepsilon)$$

onde $B(x, \varepsilon) = \{w \in \mathbb{R}^n \ : \ \|w - x\| < \varepsilon\}$ *denota a bola aberta de centro em* x *e raio* $\varepsilon > 0$.

Definição 1.1.7. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função localmente Lipschitz em* $x \in \mathbb{R}^n$ *e* $v \in \mathbb{R}^n$. *A derivada direcional de* f *em* x *na direção* v *denotada por* $f^\circ(x, v)$ *(derivada direcional de Clarke) é definida por*

$$f^\circ(x, v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y + tv) - f(y)}{t},$$

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

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

A seguir, no item (a) da Proposição 1.1.1, mostraremos que o limite acima existe pelo fato de f *ser Lipschitz, e como consequência o subdiferencial está bem definido.*

Proposição 1.1.1. *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função localmente Lipschitz em* $x \in \mathbb{R}^n$*. Então:*

a) A função $v \to f^\circ(x,v)$ *é finita e subaditiva em uma bola.*

b) $f^\circ(x,-v) = -f^\circ(x,v), \ \forall v \in \mathbb{R}^n$;

c) $f^\circ(x,\lambda v) = \lambda f^\circ(x,v), \ \forall \lambda > 0, \ \forall v \in \mathbb{R}^n$;

d) $\partial^\circ(\beta \cdot f)(x) = \beta \cdot \partial^\circ f(x), \ \forall \beta \in \mathbb{R}$.

Demonstração. (a) Por hipótese f é localmente Lipschitz em $x \in \mathbb{R}^n$, entâo

$$|f(z) - f(y)| \leqslant K\|z - y\|, \ \forall z, y \in B(x,\varepsilon).$$

Defina $z = y + tv$ para t suficientemente próximo de 0. Seja y suficientemente próximo de x, temos que

$$|f(y + tv) - f(y)| \leqslant Kt\|v\|.$$

Assim,

$$\frac{f(y+tv) - f(y)}{t} \leqslant \frac{|f(y+tv) - f(y)|}{t} \leqslant K\|v\|,$$

isto é,

$$\limsup_{y\to x \ t\downarrow 0} \frac{f(y+tv) - f(y)}{t} \leqslant K\|v\|.$$

Vamos mostrar a subaditividade, por definição

$$\begin{aligned} f^\circ(x, v+w) &= \limsup_{y\to x \ t\downarrow 0} \frac{f(y+tv+tw) - f(y)}{t} \\ &\leqslant \limsup_{y\to x \ t\downarrow 0} \frac{f(y+tv+tw) - f(y+tw)}{t} + \limsup_{y\to x \ t\downarrow 0} \frac{f(y+tw) - f(y)}{t} \\ &= f^\circ(x,v) + f^\circ(x,w). \end{aligned}$$

(b) Por definição,

$$\begin{aligned} f^\circ(x,-v) &= \limsup_{y\to x \ t\downarrow 0} \frac{f(y-tv) - f(y)}{t} \\ &= \limsup_{u\to x \ t\downarrow 0} \frac{f(u) - f(u+tv)}{t} \\ &= -\limsup_{u\to x \ t\downarrow 0} \frac{f(u+tv) - f(u)}{t} \\ &= -f^\circ(x,v), \end{aligned}$$

onde na segunda igualdade fizemos $y - tv = u$.

(c) Por definição,

$$\begin{aligned} f^{\circ}(x, \lambda v) &= \limsup_{y \to x\ t \downarrow 0} \frac{f(y + t\lambda v) - f(y)}{t} \\ &= \lambda \cdot \limsup_{y \to x\ t \downarrow 0} \frac{f(y + t\lambda v) - f(y)}{t\lambda} \\ &= \lambda \cdot \limsup_{y \to x\ \mu \downarrow 0} \frac{f(y + \mu v) - f(y)}{\mu} \\ &= \lambda \cdot f^{\circ}(x, v), \end{aligned}$$

onde na terceira igualdade fizemos $t\lambda = \mu$.

(d) Para $\beta \geqslant 0$.

$$\begin{aligned} w \in \partial^{\circ}(\beta f)(x) &\Leftrightarrow \langle w, v \rangle \leqslant (\beta f)^{\circ}(x, v),\ \forall v \in \mathbb{R}^n \\ &\Leftrightarrow \langle w, v \rangle \leqslant f^{\circ}(x, \beta v) \\ &\Leftrightarrow \langle w, v \rangle \leqslant \beta \cdot f^{\circ}(x, v) \\ &\Leftrightarrow \left\langle \frac{w}{\beta}, v \right\rangle \leqslant f^{\circ}(x, v) \\ &\Leftrightarrow \frac{w}{\beta} \in \partial^{\circ} f(x) \\ &\Leftrightarrow w \in \beta \cdot \partial^{\circ} f(x). \end{aligned}$$

Para $\beta = -1$,

$$\begin{aligned} w \in \partial(-f)(x) &\Leftrightarrow \langle w, v \rangle \leqslant (-f)^{\circ}(x, v),\ \forall v \in \mathbb{R}^n \\ &\Leftrightarrow \langle w, v \rangle \leqslant f^{\circ}(x, -v) \\ &\Leftrightarrow \langle w, -u \rangle \leqslant f^{\circ}(x, u) \\ &\Leftrightarrow \langle -w, u \rangle \leqslant f^{\circ}(x, u) \\ &\Leftrightarrow -w \in \partial^{\circ} f(x) \\ &\Leftrightarrow w \in -\partial^{\circ} f(x), \end{aligned}$$

onde na terceira equivalência fizemos $u = -v$. □

Teorema 1.1.4. *Sejam $\varphi, \psi : \mathbb{R}^n \to \mathbb{R}$ funções tais que φ é localmente Lipschitz, ψ localmente Lipschitz e diferenciável. Então,*

$$\partial^{\circ}(\varphi - \psi) = \partial^{\circ}\varphi - \partial^{\circ}\psi = \partial^{\circ}\varphi - \{\nabla\psi\}.$$

Demonstração. Ver [5], página 39. □

Definição 1.1.8. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função convexa. A derivada direcional de* f *em* $x \in \operatorname{dom} f$ *na direção* v *denotada por* $f'(x, v)$ *(derivada direcional de Fenchel-Moreau) é definida por*

$$f'(x, v) = \lim_{t \to 0^+} \frac{f(x + tv) - f(x)}{t},$$

caso o limite acima exista (consideremos ∞ *como limite). Como na definição de derivada direcional de Clarke, o item (a) da Proposição 1.1.1 é satisfeito, pois o Subdiferencial de Clarke coincide com o Subdiferencial de Fenchel-Moreau. [ver Teorema 1.1.7].*

Definição 1.1.9. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função convexa, definimos o subdiferencial de Fenchel-Moreau de* f *em* x *(denotado por* $\partial f(x)$*) como*

$$\partial f(x) = \begin{cases} \{w \in \mathbb{R}^n \ : \ \langle w, z - x \rangle + f(x) \leqslant f(z), \ \forall z \in \mathbb{R}^n\}, & \textit{se } x \in \operatorname{dom} f \\ \emptyset, & \textit{se } x \notin \operatorname{dom} f, \end{cases}$$

onde o w é dito o subgradiente de f em x. Denotaremos o domínio do operador subdiferencial como,

$$\operatorname{dom} \partial f(x) = \{x \in \mathbb{R}^n ; \ \partial f(x) \neq \emptyset\}.$$

Teorema 1.1.5. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função própria e convexa. Se* $x \in \operatorname{dom} f$ *então* $\partial f(x) \neq \emptyset$*.*

Demonstração. Ver [12], página 217. □

Teorema 1.1.6. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função convexa e* $x \in \operatorname{dom} f$*. Então* w *é um subgradiente de* f *no ponto* x *se, e somente se,*

$$\langle w, v \rangle \leqslant f'(x, v), \ \forall v \in \mathbb{R}^n.$$

Demonstração. $(\Leftarrow)$ Por definição,

$$f'(x, v) = \lim_{t \to 0^+} \frac{f(x + tv) - f(x)}{t}.$$

como a função $t \mapsto \dfrac{f(x + tv) - f(x)}{t}$ é não decrescente, temos

$$\langle w, v \rangle \leqslant \frac{f(x + tv) - f(x)}{t},$$

Daí,

$$\langle w, tv\rangle \leqslant f(x+tv) - f(x).$$

Fazendo $z = x + tv$, temos que

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

$(\Rightarrow)$ Por hipótese $w \in \partial f(x)$, ou seja,

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

Defina $z = x + tv$,

$$\langle w, v\rangle \leqslant \frac{f(x+tv) - f(x)}{t}.$$

Fazendo $t \to 0^+$, obtemos $\langle w, v\rangle \leqslant f'(x, v), \ \forall v \in \mathbb{R}^n$. $\square$

O Teorema a seguir mostra que se f é convexa o subdiferencial de Clarke coincide com o subdiferencial clássico de Fenchel-Moreau.

Teorema 1.1.7. *Sejam* $C \subset \mathbb{R}^n$ *um conjunto convexo, não vazio e* $f : C \to \mathbb{R}$ *uma função localmente Lipschitz em* $x \in C$. *Se* f *é convexa em* C *então* $f^\circ(x, v) = f'(x, v)$ *para cada* $v \in \mathbb{R}^n$ *e* $\partial^\circ f(x) = \partial f(x)$.

Demonstração. Por hipótese, f é localmente Lipschitz em $x \in C$, ou seja, $\exists \delta > 0$ tal que,

$$|f(y) - f(x)| \leqslant K\|y - x\|, \ \forall y \in B\left(x, \frac{\delta\varepsilon}{2K}\right) \cap C,$$

com $\varepsilon > 0$ arbitrário e $K > 0$ constante de Lipschitz (dependendo de x). Note que, podemos escrever

$$f^\circ(x, v) = \lim_{\delta \downarrow 0} \sup_{||y-x||<\frac{\delta\varepsilon}{2K}} \sup_{0<t<\delta} \frac{f(y+tv) - f(y)}{t}.$$

Como f é convexa, temos que a função

$$t \mapsto \frac{f(y+tv) - f(y)}{y},$$

é não decrescente em $0 < t < \delta$. Consequentemente,

$$f^\circ(x, v) = \lim_{\delta \downarrow 0} \sup_{||y-x||\leqslant\frac{\delta\varepsilon}{2K}} \frac{f(y+\delta v) - f(y)}{\delta}.$$

Logo, para $0 < \delta$ suficientemente próximo de zero,

$$
\begin{aligned}
|f^{\circ}(x,v) - f'(x,v)| &= \left| \frac{f(y+\delta v) - f(y)}{\delta} - \frac{f(x+\delta v) - f(x)}{\delta} \right| \\
&= \left| \frac{f(y+\delta v) - f(x+\delta v)}{\delta} + \frac{f(x) - f(y)}{\delta} \right| \\
&\leqslant \left| \frac{f(y+\delta v) - f(x+\delta v)}{\delta} \right| + \left| \frac{f(x) - f(y)}{\delta} \right| \\
&\leqslant \frac{2K}{\delta} \cdot \|y - x\| \\
&< \frac{2K}{\delta} \cdot \frac{\delta\varepsilon}{2K} \\
&= \varepsilon.
\end{aligned}
$$

Finalmente,

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

□

Teorema 1.1.8. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *convexa. Então* $\overline{x}$ *é minimizador de* f *em* $\mathbb{R}^n$ *se, e somente se,* $0 \in \partial f(\overline{x})$.

Demonstração. Note que,

$$
\begin{aligned}
\overline{x} \text{ é } \mathrm{minimizador} &\Leftrightarrow f(\overline{x}) \leqslant f(x), \ \forall x \in \mathbb{R}^n \\
&\Leftrightarrow \langle 0, x - \overline{x}\rangle + f(\overline{x}) \leqslant f(x), \ \forall x \in \mathbb{R}^n \\
&\Leftrightarrow 0 \in \partial f(\overline{x}).
\end{aligned}
$$

□

Teorema 1.1.9. *Sejam* $\varphi, \psi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *funções convexas tais que* $\mathrm{int}(\mathrm{dom}\ \varphi) \cap \mathrm{int}(\mathrm{dom}\ \psi) \neq \emptyset$, *então*

$$\partial(\varphi + \psi) = \partial\varphi + \partial\psi.$$

Demonstração. Ver [19], página 79. □

Definição 1.1.10. *Dado um conjunto convexo, fechado e não vazio* $C \subset \mathbb{R}^n$, *o cone normal de* C *no ponto* $x \in \mathbb{R}^n$ *é dado por,*

$$
N_C(x) = \begin{cases} \{s \in \mathbb{R}^n \ : \ \langle s, y - x\rangle \leqslant 0, \ \forall y \in C\}, & \textit{se } x \in C \\ \emptyset, & \textit{se } x \notin C. \end{cases}
$$

Exemplo 1.1.1. *Seja* $C \subset \mathbb{R}^n$ *um conjunto convexo, fechado e não vazio. A função* $\delta_C : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *dada por:*

$$\delta_C(x) = \begin{cases} 0, & \text{se } x \in C \\ \infty, & \text{se } x \notin C, \end{cases}$$

chamada função indicadora de C. *Mostraremos que tal função é própria, convexa, sci e o subdiferencial coincide com o cone normal de* C.

1) *É própria, pois sendo* $C \neq \emptyset \Rightarrow \mathrm{dom}(\delta_C) \neq \emptyset$.

2) *É convexa, pois como* C *é convexo então* $\forall x, y \in C$ *e* $\alpha \in [0,1]$ *tem-se que*

$$\alpha x + (1-\alpha)y \in C,$$

assim,

$$\begin{aligned} 0 &= \delta_C(\alpha x + (1-\alpha)y) \\ &\leqslant \alpha \cdot 0 + (1-\alpha) \cdot 0 \\ &= \alpha \cdot \delta_C(x) + (1-\alpha)\delta_C(y). \end{aligned}$$

Ou seja, $\delta_C(\alpha x + (1-\alpha)y) \leqslant \alpha\delta_C(x) + (1-\alpha)\delta_C(y),\ \forall x, y \in C$ *e* $\alpha \in [0,1]$.

3) δ_C *é sci em* C.

De fato, seja $\{x^k\} \subset \mathbb{R}^n$ *tal que* $x^k \to x \in C$ *então,*

$$0 = \delta_C(x) \leqslant \liminf \delta_C(x^k),$$

onde esse limite inferior é 0 *ou* ∞.

4) $\partial\delta_C(\cdot) = N_C(\cdot)$.

Com efeito, seja $s \in \partial\delta_C(x)$, *por definição*

$$\langle s, y - x\rangle + \delta_C(x) \leqslant \delta_C(y)$$

$\mathrm{i})$ *Se* $x \notin C$, *temos dois casos para* y.

$\mathrm{a})$ *Se* $y \notin C$, *tem-se*

$$\infty \leqslant \infty.$$

$\mathrm{b})$ *Se* $y \in C$, *tem-se*

$$\infty \leqslant 0.$$

absurdo, logo $\partial\delta_C(x) = \emptyset$*, se* $x \notin C$.

$ii)$ *Se* $x \in C$*, temos dois casos para* y.

$a)$ *Se* $y \notin C$*, tem-se*

$$\langle s, y - x\rangle \leqslant \infty.$$

$b)$ *Se* $y \in C$*, tem-se*

$$\langle s, y - x\rangle \leqslant 0.$$

Logo, por (a) *e* (b) *temos que* $\langle s, y - x\rangle \leqslant 0$*. Portanto,*

$$\begin{aligned} \partial\delta_C(x) &= \begin{cases} \{s \in \mathbb{R}^n \ : \ \langle s, y - x\rangle \leqslant 0, \ \forall y \in C\}, & \textit{se } x \in C \\ \emptyset, & \textit{se } x \notin C \end{cases} \\ &= N_C(x). \end{aligned}$$

Teorema 1.1.10. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função convexa. Se* $\overline{x}$ *é um minimizador global em um conjunto convexo, fechado e não vazio* $C \subset \operatorname{dom} f$*, então*

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

Demonstração. Note que,

$$\min_{x \in C} f(x) \Leftrightarrow \min_{x \in \mathbb{R}^n} (f(x) + \delta_c(x))$$

Pelo Teorema 1.1.8 e 1.1.9, temos que $\overline{x}$ é solução do problema acima se, e somente se,

$$\begin{aligned} 0 \in \partial(f(\cdot) + \delta_c(\cdot))(\overline{x}) &= \partial f(\overline{x}) + \partial\delta_c(\overline{x}) \\ &= \partial f(\overline{x}) + N_c(\overline{x}). \end{aligned}$$

□

Teorema 1.1.11. *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função convexa, se* f *for diferenciável em* $x \in C \subset \operatorname{dom} f$*, então* $\partial f(x) = \{\nabla f(x)\}$.

Demonstração. Suponha f diferenciável em $x \in C$, logo temos que $f'(x, \cdot) = \langle \nabla f(x), \cdot\rangle$. Seja $w \in \partial f(x)$, pelo Teorema 1.1.6

$$\langle w, v\rangle \leqslant \langle \nabla f(x), v\rangle, \forall v \in \mathbb{R}^n.$$

Logo,

$$0 \leqslant \langle \nabla f(x) - w, v \rangle \ \forall v \in \mathbb{R}^n.$$

Escolha $v = -(\nabla f(x) - w)$, daí

$$0 \leqslant -\|\nabla f(x) - w\|^2.$$

Portanto, $\nabla f(x) - w = 0$, ou seja, $\nabla f(x) = w$. $\square$

Teorema 1.1.12. *Sejam $C \subset \mathbb{R}^n$ um conjunto não vazio e $f : C \to \mathbb{R}$ uma função localmente Lipschitz em $x \in C$. Então $f^\circ(x, v)$ é semicontínua superior.*

Demonstração. Sejam $\{x^k\} \subset C$ e $\{v^k\} \subset \mathbb{R}^n$ sequências com $\lim\limits_{k\to\infty} x^k = x$ e $\lim\limits_{k\to\infty} v^k = v$. Pela definição de limite superior, temos que $\forall k > 0$, existem x^k, y^k e $t_k > 0$, com $\{y^k\}$ suficientemente próximo de x tais que,

$$\begin{aligned} f^\circ(x^k, v^k) - \frac{1}{k} &< \frac{f(y^k + t_k v^k) - f(y^k)}{t_k} \\ &= \frac{f(y^k + t_k v^k) - f(y^k)}{t_k} - \frac{f(y^k + t_k v)}{t_k} + \frac{f(y^k + t_k v)}{t_k} \\ &= \frac{f(y^k + t_k v^k) - f(y^k + t_k v)}{t_k} + \frac{f(y^k + t_k v) - f(y^k)}{t_k}. \end{aligned}$$

Assim,

$$\begin{aligned} f^\circ(x^k, v^k) &\leqslant \frac{1}{k} + \left| \frac{f(y^k + t_k v^k) - f(y^k + t_k v)}{t_k} \right| + \frac{f(y^k + t_k v) - f(y^k)}{t_k} \\ &\leqslant \frac{1}{k} + \|v^k - v\| + \frac{f(y^k + t_k v) - f(y^k)}{t_k}. \end{aligned}$$

Passando o limite superior na desigualdade acima,

$$\limsup_{k\to\infty} f^\circ(x^k, v^k) \leqslant \limsup_{k\to\infty} \left(\frac{1}{k} + \|v^k - v\| + \frac{f(y^k + t_k v) - f(y^k)}{t_k} \right).$$

Portanto,

$$\limsup_{k\to\infty} f^\circ(x^k, v^k) \leqslant f^\circ(x, v).$$

$\square$

Teorema 1.1.13. *Sejam $C \subset \mathbb{R}^n$ um conjunto fechado, convexo e não vazio, $f : C \to \mathbb{R}$ uma função localmente Lipschitz em $x \in C$ e $D \subset C$ um conjunto limitado. Então o conjunto $\bigcup\limits_{x\in D} \partial^\circ f(x)$ é limitado. Além disso, se f é convexa em C, então $\bigcup\limits_{x\in D} \partial f(x)$ é limitado.*

Demonstração. Suponha por absurdo que $\bigcup_{x\in D} \partial^\circ f(x)$ não seja limitado, ou seja, existem $\{x^k\} \subset D$ e $\{w^k\} \subset \mathbb{R}^n$ tais que $w^k \in \partial^\circ f(x^k)$ e $\|w^k\| \to \infty$ quando $k \to \infty$. Defina,

$$v^k = \frac{w^k}{\|w^k\|}.$$

Logo a menos de subsequência temos que, $\lim_{k\to\infty} x^k = x$ e $\lim_{k\to\infty} v^k = v$. Então,

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

Tomando o limite superior na desigualdade acima,

$$\begin{aligned} \infty = \limsup_{k\to\infty} \|w^k\| &= \limsup_{k\to\infty} \langle w^k, v^k \rangle \\ &\leqslant \limsup_{k\to\infty} f^\circ(x^k, v^k) \\ &\leqslant f^\circ(x, v). \end{aligned}$$

Absurdo, pois $f^\circ(x, v)$ é finito. Portanto o conjunto $\bigcup_{x\in D} \partial^\circ f(x)$ é limitado. A última afirmação segue pelo Teorema 1.1.7. □

Teorema 1.1.14. *Sejam $C \subset \mathbb{R}^n$ um conjunto não vazio e $f : C \to \mathbb{R}$ uma função localmente Lipschitz em $x \in C$. Sejam $\{w^k\} \subset \mathbb{R}^n$, $\{x^k\} \subset C$ tais que $w^k \in \partial^\circ f(x^k)$, $\forall k > 0$ e $x^k \to \overline{x}$ quando $k \to \infty$. Então $\{w^k\}$ é limitada e todos os seus pontos de acumulação pertencem a $\partial^\circ f(\overline{x})$.*

Demonstração. Por definição,

$$w^k \in \partial^\circ f(x^k) \Leftrightarrow \langle w^k, v \rangle \leqslant f^\circ(x^k, v), \forall v \in \mathbb{R}^n.$$

Logo, w^k é limitado pois $\partial^\circ f(x^k)$ é, então $\lim_{j\to\infty} w^{k_j} = \overline{w}$. Aplicando o limite superior na desigualdade acima,

$$\begin{aligned} \langle \overline{w}, v \rangle &= \limsup_{j\to\infty} \langle w^{k_j}, v \rangle \\ &\leqslant \limsup_{j\to\infty} f^\circ(x^{k_j}, v) \\ &\leqslant f^\circ(\overline{x}, v). \end{aligned}$$

Isto mostra que $\overline{w} \in \partial^\circ f(\overline{x})$. □

Capítulo 2

Distância Proximal

2.1 Distância Proximal

Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função localmente Lipschitz, não necessariamente convexa, C um conjunto convexo, fechado e não vazio. Nem sempre o problema

$$\min_{x \in C} f(x)$$

está bem definido, no sentido de existir o minimizador e que tal minimizador pertença ao interior do conjunto C, necessitamos de uma função que regularize (isto é, que obrigue o problema a ter solução) e penalize tal problema (que obrigue a solução estar no interior do conjunto). Para isto, utilizaremos a função distância proximal, definida a seguir.

Definição 2.1.1. *Uma função* $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ *é chamada de distância proximal com respeito ao conjunto convexo, fechado e não vazio* $C \subset \mathbb{R}^n$ *se para cada* $y \in \operatorname{int} C$ *fixado, as seguintes propriedades são satisfeitas:*

(D.1) $d(\cdot, y)$ *é uma função própria, sci, convexa e de classe* C^1 *em* $\operatorname{int} C$*;*

(D.2) dom $d(\cdot, y) \subset C$ *e dom* $\nabla_1 d(\cdot, y) = \operatorname{int} C$*, onde* $\nabla_1 d(\cdot, y)$ *denota o gradiente da função* $d(\cdot, y)$ *com respeito a primeira variável;*

(D.3) $d(\cdot, y)$ *é nível limitada em* $\mathbb{R}^n$*, isto é,* $\lim_{||x|| \to \infty} d(x, y) = \infty$*;*

(D.4) $d(y, y) = 0$.

Observação 2.1.1. *Denotamos por* $\mathcal{D}(C)$ *a família de funções* d *que satisfazem a definição 2.1.1.*

1. *Por definição* $\mathrm{d}(\cdot,\cdot) \geqslant 0$*;*

2. *A propriedade (D.1) preserva a convexidade de* $\mathrm{d}(\cdot, y)$*;*

3. *A propriedade (D.2) força um processo interativo que usa a distância proximal para permanecer em* $\operatorname{int} C$ *e é equivalente à conhecida hipótese de coercividade na fronteira, ou seja,* $\forall x \in \operatorname{int} C$*, se* $\{y^k\} \subset \operatorname{int} C$ *e* $\lim_{k\to\infty} y^k = y \in \operatorname{fr} C$*, então*

$$\lim_{k\to\infty} \langle \nabla_1 \mathrm{d}(y^k, x), x - y^k \rangle = -\infty;$$

4. *A propriedade (D.3) é usada para garantir a existência de tal iteração para cada* $y \in \operatorname{int} C$*;*

5. *De (D.4), o mínimo global de* $\mathrm{d}(\cdot, y)$ *é atingido em* y *o que significa* $\nabla_1 \mathrm{d}(y, y) = 0$*.*

2.2 Distância Proximal Induzida

O próximo passo é associar cada função $\mathrm{d} \in \mathcal{D}(C)$ dada com uma distância proximal correspondente satisfazendo algumas propriedades desejadas.

Definição 2.2.1. *Dado* $\mathrm{d} \in \mathcal{D}(C)$*. Defina* $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+ \cup \{\infty\}$ *uma função tal que* $\operatorname{int} C \times \operatorname{int} C \subseteq \operatorname{dom} H$*. A aplicação* $H(\cdot,\cdot)$ *é chamada distância proximal induzida por* d *se as seguintes propriedades são satisfeitas:*

(H.1) Para cada $x \in \operatorname{int} C$*,* $H(x, \cdot)$ *é contínua em* $\operatorname{int} C$*;*

(H.2) $H(x, x) = 0, \ \forall x \in \operatorname{int} C$*;*

(H.3) Para todo $x \in C$ *e* $\alpha \in \mathbb{R}$ *o conjunto* $\{y \in \operatorname{int} C; \ H(x, y) \leqslant \alpha\}$ *é limitado;*

(H.4) Para cada $x, y \in \operatorname{int} C$ *e algum* $\gamma > 0$ *fixado tem-se que,*

$$\langle z - x, \nabla_1 \mathrm{d}(x, y) \rangle \leqslant H(z, y) - H(z, x) - \gamma H(x, y), \ \forall z \in C;$$

(H.5) Se $\{y^k\} \subset \operatorname{int} C$ *e* $\{y^k\} \to y \in C$*, então* $H(y, y^k) \to 0$*;*

(H.6) Seja $z \in C$ *e* $y \in \operatorname{int} C$ *e tome* $w = \alpha z + (1 - \alpha) y$*, com* $\alpha \in (0, 1)$*. Então,*

$$H(z, w) + H(w, y) \leqslant H(z, y);$$

(H.7) Se $\{x^k\}, \{y^k\} \subset \operatorname{int} C$ *são sequências tais que* $\{x^k\}$ *converge para* x *e* $\{y^k\}$ *converge para* y, *com* $x \neq y$, *então*

$$\liminf_{k\to\infty} H(x^k, y^k) > 0.$$

Observação 2.2.1. *Denotamos por* $\mathcal{F}(C)$ *o conjunto dos pares* (d, H) *da distância proximal e distância proximal induzida generalizada que satisfaz as condições da Definição 2.2.1.*

1. *A hipótese (H.1) é usada na definição de distância proximal induzida, por exemplo em [3] e a hipótese (H.2) é usada em [3, 1, 2].*

2. *As hipóteses (H.3) e (H.5) são hipóteses clássicas no contexto de distância de Bregman e (H.4) é usada por exemplo em [3], [2].*

3. *As condições (H.6) e (H.7) provaram satisfazer distâncias de Bregman em Lemas* 2.2 *e* 2.3 *de [14] e para homogênea de segunda ordem em Observações* 4.2 *e* 4.4 *de [3].*

2.3 Distância de Bregman

Seja $C \subset \mathbb{R}^n$ convexo, fechado e não vazio. Considere uma função $h : C \to \mathbb{R}$ convexa e diferenciável em $\operatorname{int} C$. Defina $D_h : C \times \operatorname{int} C \to \mathbb{R}$ dada por

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle,$$

induzida por h.

Definição 2.3.1. *Dizemos que* h *é uma função de Bregman e* D_h *a distância de Bregman induzida por* h *se as seguintes condições são satisfeitas:*

(D.B.1) h é de classe C^1 *em* $\operatorname{int} C$;

(D.B.2) h *é estritamente convexa e contínua em* C;

(D.B.3) $\forall \alpha > 0$ *os conjuntos de níveis*

$$\begin{aligned} \Gamma_1(y, \alpha) &:= \{x \in C \mid D_h(x, y) \leqslant \alpha\} \\ \Gamma_1(x, \alpha) &:= \{y \in \operatorname{int} C \mid D_h(x, y) \leqslant \alpha\}, \end{aligned}$$

são limitados $\forall y \in \operatorname{int} C$ *e* $\forall x \in C$, *respectivamente;*

(D.B.4) Se $\{y^k\} \subset \mathrm{int}\ C$ *tal que* $\lim_{k\to\infty} y^k = \bar{y}$ *então* $\lim_{k\to\infty} D_h(\bar{y}, y^k) = 0$*;*

(D.B.5) Se $\{z^k\} \subset C$ *e* $\{y^k\} \subset \mathrm{int}\ C$ *são sequências tais que* $\{z^k\}$ *é limitada,*

$$\lim_{k\to\infty} y^k = \bar{y} \text{ e } \lim_{k\to\infty} D_h(z^k, y^k) = 0.$$

Então, $\lim_{k\to\infty} z^k = \bar{y}$*;*

(D.B.6) (A Função de Bregman é Coerciva na Fronteira) Se $\{y^k\} \subset \mathrm{int}\ C$ *é tal que* $\lim_{k\to\infty} y^k = y \in \mathrm{fr}\ C$*, então*

$$\lim_{k\to\infty} \langle \nabla h(y^k), x - y^k \rangle = -\infty, \forall x \in \mathrm{int}\ C;$$

(D.B.7) Dizemos que C *é zona de* h *se para todo* $y \in \mathbb{R}^n$ *existe* $x \in \mathrm{int}\ C$ *tal que* $\nabla h(x) = y$.

Exemplo 2.3.1. *Sejam* $C \subset \mathbb{R}^n$ *um conjunto fechado, convexo, com* $\mathrm{int}\ C$ *não vazio. Considere* $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *uma função própria, estritamente convexa e sci com domínio* $\mathcal{D} = \mathrm{dom}\ \varphi$ *fechado e de classe* C^1 *no* $\mathrm{int}\ C$*. A distância de Bregman associada a* φ *é dada por,*

$$D_\varphi(x, y) = \begin{cases} \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x - y \rangle, & \forall y \in \mathrm{int}\ C, \\ \infty, & \text{caso contrário.} \end{cases}$$

As seguintes aplicações são exemplos de Distância de Bregman:

Ex.1: $\varphi(x) = \frac{1}{2}\|x\|^2$, *com* $C = \mathbb{R}^n$. *Então,* $\nabla\varphi(y) = y$, *daí,*

$$\begin{aligned} D_\varphi(x, y) &= \frac{1}{2}(\|x\|^2 - \|y\|^2) - \langle y, x - y \rangle \\ &= \frac{1}{2}\left(\langle x, x \rangle - \langle y, y \rangle - 2\langle y, x \rangle + 2\langle y, y \rangle\right) \\ &= \frac{1}{2}\left(\langle x, x \rangle - 2\langle x, y \rangle + \langle y, y \rangle\right) \\ &= \frac{1}{2}\|x - y\|^2. \end{aligned}$$

Ex.2: $\varphi(x) = -\sum_{i\in I(x)}^{n} \log x_i$, *com* $C = \mathbb{R}^n_+$. *Então,* $\nabla\varphi(y) = -\sum_{i\in I(x)} \frac{1}{y_i}$, *daí,*

$$\begin{aligned} D_\varphi(x, y) &= \sum_{i\in I(x)}^{n} -\log x_i + \log y_i - \left\langle -\frac{1}{y_i}, x_i - y_i \right\rangle \\ &= \sum_{i\in I(x)}^{n} \log\left(\frac{y_i}{x_i}\right) + \frac{x_1}{y_i} - 1. \end{aligned}$$

Onde $I(x) = \{i; x_i > 0\}$.

Ex.3: $\varphi(x) = \sum_{i=1}^{n} x_i \log x_i$, *com* $C = \mathbb{R}^n_+$. *Então,* $\nabla\varphi(x) = \sum_{i=1}^{n} \log x_i + 1$. *Daí,*

$$\begin{aligned} D_\varphi(x,y) &= \sum_{i=1}^{n} x_i \log x_i - y_i \log y_i - \langle \log y_i + 1, x_i - y_i \rangle \\ &= \sum_{i=1}^{n} x_i \log x_i - y_i \log y_i - x_i \log y_i + y_i \log y_i + y_i - x_i \\ &= \sum_{i=1}^{n} x_i \log\left(\frac{x_i}{y_i}\right) + y_i - x_i. \end{aligned}$$

Onde as duas últimas funções são conhecidas como entropia de Burg e entropia de Shannon respectivamente e a última distância induzida D_φ *é a conhecida distância de entropia relativa de Kullback-Leibler.*

Proposição 2.3.1. *Seja* φ *uma função de Bregman. Então:*

i) $D_\varphi(z,x) - D_\varphi(z,y) - D_\varphi(y,x) = \langle \nabla\varphi(x) - \nabla\varphi(y), y - z \rangle, \forall z \in C \text{ e } x, y \in \text{int } C;$

ii) $\nabla_1 D_\varphi(x,y) = \nabla\varphi(x) - \nabla\varphi(y), \ \forall x, y \in \text{int } C.$

Demonstração. (i) Pela definição de função de Bregman,

$$\begin{aligned} D_\varphi(z,x) - D_\varphi(z,y) - D_\varphi(y,x) &= -\langle \nabla\varphi(x), z - x \rangle + \langle \nabla\varphi(y), z - y \rangle + \langle \nabla\varphi(x), y - x \rangle \\ &= \langle \nabla\varphi(x), y - z \rangle + \langle \nabla\varphi(y), z - y \rangle \\ &= \langle \nabla\varphi(x) - \nabla\varphi(y), y - z \rangle. \end{aligned}$$

(ii) Pela definição de função de Bregman,

$$D_\varphi(x,y) = \varphi(x) - \varphi(y) - \langle \nabla\varphi(y), x - y \rangle.$$

Derivando a expressão acima em relação a primeira variável,

$$\nabla_1 D_\varphi(x,y) = \nabla\varphi(x) - \nabla\varphi(y).$$

□

Como em Burachik e Scheimberg [4], consideramos ao longo deste trabalho o seguinte conjunto de suposições sobre φ:

(B.1) Os conjuntos de níveis à direita de $D_\varphi(y,\cdot)$

$$S_{y,\alpha} := \{z \in \text{int } \mathcal{D} \mid D_\varphi(y,z) \leqslant \alpha\},$$

são limitados $\forall \alpha \geqslant 0$, e $\forall y \in \mathcal{D}$;

(B.2) Se $\{x^k\}$, $\{y^k\} \subset \operatorname{int} \mathcal{D}$ com,

$$\lim_{k\to\infty} x^k = x, \ \lim_{k\to\infty} y^k = x \ e \ \lim_{k\to\infty} D_\varphi(x^k, y^k) = 0.$$

Então, $\lim_{k\to\infty} D_\varphi(x, x^k) - D_\varphi(x, y^k) = 0$;

(B.3) Se $\{x^k\} \subset \mathcal{D}$ é limitada, $\{y^k\} \subset \operatorname{int} \mathcal{D}$ é tal que,

$$\lim_{k\to\infty} y^k = y \ e \ \lim_{k\to\infty} D_\varphi(x^k, y^k) = 0.$$

Então, $\lim_{k\to\infty} x^k = y$;

(B.4) Para cada $y \in C$, existe $x \in \operatorname{int} \mathcal{D}$ tal que $\nabla\varphi(x) = y$;

Seja $d = D_\varphi$ a distância de Bregman induzida por φ. É provado em [1] que a distância proximal induzida por d é $H_d = D_\varphi$. As hipóteses de φ implica que $H_d = D_\varphi$ verifica (H.1) e (H.2) bem como (H.4) por causa da a identidade dos três pontos (item (i) da Proposição 2.3.1). Como provado nos Lemas 2.2 e 2.3 de [14] cada distância de Bregman verifica (H.6) e (H.7). Portanto as condições (H.1),(H.4),(H.6) e (H.7) são satisfeitas automaticamente para o par $(d, H_d) = (D_\varphi, D_\varphi)$. Por outro lado as condições (H.3) e (H.5) são hipóteses padrão no contexto de distâncias de Bregman. A seguinte condição alternativa também será considera em nossa análise.

(B.5) Se $\{y^k\} \subset \operatorname{int} \mathcal{D}$ e $\lim_{k\to\infty} D_\varphi(y^k, y^{k+1}) = 0$, então

$$\liminf_{k\to\infty} [D_\varphi(x, y^{k+1}) - D_\varphi(x, y^k)] \geqslant 0, \ \forall x \in C.$$

As Distâncias de Bregman do Exemplo 2.3.1 são exemplos de funções que satisfazem a condição (B.5).

2.4 Distância Proximal Homogênea de Segunda Ordem

Seja $\varphi : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ uma função convexa, própria e sci tal que $\operatorname{dom} \varphi \subseteq [0, \infty)$. Assumimos que:

(i) φ é de classe C^2 em $\operatorname{int}(\operatorname{dom} \varphi) = (0, \infty)$;

(ii) φ é estritamente convexa em seu domínio;

(iii) $\lim_{t\to 0^+} \varphi^{'}(t) = -\infty$;

(iv) $\varphi(1) = \varphi^{'}(1) = 0$ e $\varphi^{''}(1) > 0$.

Nós denotamos por Φ a classe de funções satisfazendo as condições de (i) à (iv) acima. Dado $\varphi \in \Phi$, nós definimos as duas seguintes subclasses de Φ:

$$\Phi_1 := \left\{\varphi \in \Phi \ : \ \varphi^{''}(1)\left(1-\frac{1}{t}\right) \leqslant \varphi^{'}(t) \leqslant \varphi^{''}(1)\ln\, t,\ \forall t>0\right\}$$

e

$$\Phi_2 := \left\{\varphi \in \Phi \ : \ \varphi^{''}(1)\left(1-\frac{1}{t}\right) \leqslant \varphi^{'}(t) \leqslant \varphi^{''}(1)(t-1),\ \forall t>0\right\}.$$

Sendo que $\ln t \leqslant t-1$ para todo $t>0$ e $\varphi^{''}(1)>0$. Claramente $\Phi_1 \subseteq \Phi_2 \subseteq \Phi$. As seguintes funções

$$\begin{aligned} \varphi_1(t) &= t\ln t - t + 1,\ \mathrm{dom}\ \varphi_1 = [0,\infty) \\ \varphi_2(t) &= -\ln t + t - 1,\ \mathrm{dom}\varphi_2 = (0,\infty) \\ \varphi_3(t) &= (\sqrt{t}-1)^2,\ \mathrm{dom}\ \varphi_3 = [0,\infty), \end{aligned}$$

pertencem à Φ_1 e consequentemente pertencem à Φ_2. Nesse papel, focaremos na função distância homogênea de segunda ordem d_ϕ dada por:

$$d_\phi(x,y) = \begin{cases} \sum_{i=1}^n y_i^2\phi\left(\frac{x_i}{y_i}\right), & \forall x,y \in \mathbb{R}^n_{++}, \\ \infty, & \text{caso contrário}, \end{cases}$$

onde ϕ é definido como:

$$\phi(t) = \mu\varphi(t) + \frac{\nu}{2}(t-1)^2,$$

com $\varphi \in \Phi_2$ e parâmetros positivos $\nu,\ \mu$ satisfazendo a seguinte inequação:

$$\nu > \mu\varphi^{''}(1) > 0.$$

Lema 2.4.1. *Seja* $\varphi \in \Phi_2,\ \phi(t) = \mu\varphi(t) + \frac{\nu}{2}(t-1)^2$, *com* $\nu > \mu\varphi^{''}(1) > 0$. *Para cada* $x,y \in \mathbb{R}^n_{++}$ *e* $z \in \mathbb{R}^n_{+}$, *temos que*

$$\langle \nabla_1 d_\phi(x,y), x-z\rangle \geqslant \left(\frac{\nu+\varphi^{''}(1)\mu}{2}\right)(\|x-z\|^2 - \|y-z\|^2) + \left(\frac{\nu-\varphi^{''}(1)\mu}{2}\right)\|x-y\|^2. \quad (2.1)$$

Demonstração. Note que,

$$\phi^{'}(t) = \mu\varphi^{'}(t) + \nu(t-1),$$

e

$$\nabla_1 d_\phi(x,y) = \sum_{i=1}^{n} y_i \phi' \left(\frac{x_i}{y_i}\right).$$

Usando a definição de ϕ' temos que,

$$\begin{aligned}
\nabla_1 d_\phi(x,y) &= \sum_{i=1}^{n} y_i \left(\mu\varphi'\left(\frac{x_i}{y_i}\right) + \nu\left(\frac{x_i - y_i}{y_i}\right)\right) \\
&= \sum_{i=1}^{n} y_i \mu\varphi'\left(\frac{x_i}{y_i}\right) + \nu(x_i - y_i).
\end{aligned}$$

Portanto,

$$\nabla_1 d_\phi(x,y) = \sum_{i=1}^{n} y_i \mu\varphi'\left(\frac{x_i}{y_i}\right) + \nu(x_i - y_i). \tag{2.2}$$

Seja $\varphi \in \Phi_2$, então

$$\varphi''(1)\left(1 - \frac{1}{t}\right) \leqslant \varphi'(t) \leqslant \varphi''(1)(t-1), \ \forall t > 0.$$

Por um lado,

$$\varphi'\left(\frac{x_i}{y_i}\right) \leqslant \varphi''(1)\left(\frac{x_i - y_i}{y_i}\right).$$

Multiplicando a desigualdade acima por $z_i y_i \geqslant 0$, obtemos

$$z_i y_i \varphi'\left(\frac{x_i}{y_i}\right) \leqslant z_i \varphi''(1)(x_i - y_i). \tag{2.3}$$

Note que,

$$-\varphi''(1)(t-1) \leqslant -\varphi'(t) \leqslant -\varphi''(1)\left(1 - \frac{1}{t}\right), \ \forall t > 0.$$

Daí,

$$-\varphi'\left(\frac{x_i}{y_i}\right) \leqslant -\varphi''(1)\left(\frac{x_i - y_i}{x_i}\right).$$

Multiplicando a desigualdade acima por $x_i y_i \geqslant 0$, obtemos

$$-x_i y_i \varphi'\left(\frac{x_i}{y_i}\right) \leqslant -y_i \varphi''(1)(x_i - y_i). \tag{2.4}$$

Somando as desigualdades (2.3) e (2.4),

$$(z_i - x_i) y_i \varphi'\left(\frac{x_i}{y_i}\right) \leqslant (z_i - y_i)\varphi''(1)(x_i - y_i).$$

Somando de $i = 1, \cdots, n$, obtemos

$$\sum_{i=1}^{n} (z_i - x_i) y_i \varphi'\left(\frac{x_i}{y_i}\right) \leqslant \varphi''(1) \sum_{i=1}^{n} (z_i - y_i)(x_i - y_i).$$

Multiplicando a desigualdade acima por $\mu > 0$ e em seguida somando com $\sum_{i=1}^{n} \nu(x_i - y_i)(z_i - x_i) > 0$, obtemos

$$\sum_{i=1}^{n} (z_i - x_i)\left[y_i \mu \varphi'\left(\frac{x_i}{y_i}\right) + \nu(x_i - y_i)\right] \leqslant \mu\varphi''(1) \sum_{i=1}^{n} (z_i - y_i)(x_i - y_i) + \nu(x_i - y_i)(z_i - x_i).$$

Usando a equação (2.2),

$$\langle z - x, \nabla_1 d_\phi(x, \nu)\rangle \leqslant \mu\varphi''(1)\langle x - y, z - y\rangle + \nu\langle x - y, z - x\rangle. \tag{2.5}$$

Note que,

$$\langle x - y, z - y\rangle = \frac{1}{2}\left(||x - y||^2 + ||z - y||^2 - ||x - z||^2\right) \tag{2.6}$$

$$\langle x - y, z - x\rangle = -\frac{1}{2}\left(||y - x||^2 + ||z - x||^2 - ||y - z||^2\right). \tag{2.7}$$

Substituindo (2.6) e (2.7) em (2.5), obtemos

$$\begin{aligned}
\langle z - x, \nabla_1 d_\phi(x, \nu)\rangle &\leqslant \mu\varphi''(1)\langle x - y, z - y\rangle + \nu\langle x - y, z - x\rangle \\
&= \left(\frac{\mu\varphi''(1) + \nu}{2}\right)(||z - y||^2 - ||x - z||^2) + \left(\frac{\mu\varphi''(1) - \nu}{2}\right)||x - y||^2.
\end{aligned}$$

Multiplicando a desigualdade acima por (-1),

$$\langle \nabla_1 d_\phi(x, y), x - z\rangle \geqslant \left(\frac{\mu\varphi''(1) + \nu}{2}\right)(||x - z||^2 - ||y - z||^2) + \left(\frac{\nu - \mu\varphi''(1)}{2}\right)||x - y||^2.$$

$\square$

Assim, temos que

$$H_d(x, y) = \left(\frac{\mu\varphi''(1) + \nu}{2}\right)||x - y||^2,$$

verifica a condição (H.4). Todas as outra condições (H.1)-(H.3), (H.5)-(H.7) são trivialmente satisfeitas para essa H_d.

Capítulo 3

Método Interior Proximal (MIP)

Neste capítulo abordamos o problema de encontrar os pontos críticos de uma função $f : \mathbb{R}^n \to \mathbb{R}$ restrito a um conjunto convexo, fechado e $\operatorname{int} C \neq \emptyset$. Na introdução foi considerado o problema,

$$\min_{x \in C} f(x).$$

Concentraremos numa subclasse especial de funções localmente Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ que pode ser escrita como diferença de duas funções convexas (funções DC), ou seja, $f(x) = g(x) - h(x)$ com funções componentes $g, h : \mathbb{R}^n \to \mathbb{R}$ sendo semicontínuas inferior e convexas. Essa importante subclasse de funções localmente Lipschitz, desperta o interesse tanto da matemática pura como da matemática aplicada. Essa classe surge naturalmente como o menor subespaço vetorial contendo todas as funções contínuas e convexas definidas em um dado conjunto.

Hamdi [6] propôs um método proximal utilizando distância de Bregman para programação DC, no entanto, o autor não apresentou uma prova correta no resultado principal de convergência. A motivação principal desse trabalho é apresentar um contra exemplo para ilustrar esse fato, o qual será desenvolvido na seção 4 deste capítulo.

Ao longo dessa dissertação, assumimos que f é limitada inferiormente, h diferenciável e que a distância proximal d em relação à C satisfaz as hipóteses (A.1) à (A.4). Para encontrar os pontos críticos de f, realizamos o seguinte método iterativo:

3.1 Algoritmo do Método Interior Proximal - MIP

Definição 3.1.1. *(Algoritmo - Método Interior Proximal)*

Passo 1*: Dado um ponto inicial* $x^0 \in \mathtt{int}\ C$ *e uma sequência limitada de números positivos* $\{\lambda_k\}$ *tal que* $\liminf_{k\to\infty} \lambda_k > 0$.

Passo 2*: Computar*

$$x^{k+1} \in \arg\min_{x\in C} \left\{ g(x) - \langle \nabla h(x^k), x - x^k \rangle + \lambda_k d(x, x^k) \right\}. \tag{3.1}$$

Passo 3*: Se* $x^{k+1} = x^k$*, pare. Caso contrário, defina* $k = k + 1$ *e retorne para o Passo* 2.

Proposição 3.1.1. *A sequência gerada por MIP é bem definida e está contida em* $\mathtt{int}\ C$.

Demonstração. Dado $x^0 \in \mathtt{int}\ C$. Suponha que o algoritmo é válido para k, vamos mostrar que vale para $k + 1$. Defina,

$$f_k(x) = g(x) - \langle \nabla h(x^k), x - x^k \rangle + \lambda_k d(x, x^k). \tag{3.2}$$

Note que f_k é convexa e semicontínua inferior, pois g, $\langle \nabla h(x^k), \cdot - x^k \rangle$ e $d(\cdot, x^k)$ são funções convexas, semicontínuas inferior, além disso f_k é própria, pois

$$\mathrm{dom}\ f_k = \mathrm{dom}\ g \cap \mathrm{dom}\ (\langle \nabla h(x^k), \cdot - x^k \rangle) \cap\ \mathrm{dom}\ d(\cdot, x^k) = \mathrm{dom}\ d(\cdot, x^k) \subset C.$$

Por hipótese, h é diferenciável então $\forall x, x^k \in \mathbb{R}^n$,

$$h(x) \geqslant h(x^k) + \langle \nabla h(x^k), x - x^k \rangle.$$

Daí,

$$-\langle \nabla h(x^k), x - x^k \rangle \geqslant h(x^k) - h(x). \tag{3.3}$$

Portanto, usando (3.3) em (3.2), obtemos

$$\begin{aligned} f_k(x) &= g(x) - \langle \nabla h(x^k), x - x^k \rangle + \lambda_k d(x, x^k) \\ &\geqslant g(x) - h(x) + h(x^k) + \lambda_k d(x, x^k) \\ &= f(x) + h(x^k) + \lambda_k d(x, x^k) \\ &\geqslant M + h(x^k) + \lambda_k d(x, x^k), \end{aligned}$$

onde na última desigualdade usamos o fato de que f é limitada inferiormente. Assim,

$$f_k(x) \geqslant M + h(x^k) + \lambda_k d(x, x^k), \tag{3.4}$$

fazendo $||x|| \to \infty$ em (3.4),

$$\lim_{||x||\to\infty} f_k(x) \geqslant M + h(x^k) + \lambda_k \cdot \lim_{||x||\to\infty} d(x, x^k) = \infty,$$

onde usamos a coercividade de $d(\cdot, x^k)$. Assim temos que f_k é semicontínua inferiormente e coerciva, pelo Teorema 1.1.10, f_k possui um minimizador $\hat{x} \in C$, pois $\mathrm{dom}\ f_k = \mathrm{dom}\ d(\cdot, x^k) \subset C$. Note que, pelo Teorema 1.1.10, se $\hat{x}$ é minimizador de f_k então,

$$0 \in \partial f_k(\hat{x}) + \partial\delta_C(\hat{x}),$$

ou seja,

$$0 \in \partial\left(g(\cdot) - \langle \nabla h(x^k), \cdot - x^k\rangle + \lambda_k d(\cdot, x^k) + \delta_C(\cdot)\right)(\hat{x}).$$

Pelo Teorema 1.1.9,

$$0 \in \partial g(\hat{x}) - \nabla h(x^k) + \lambda_k \nabla_1 d(\hat{x}, x^k) + \partial\delta_C(\hat{x}).$$

Equivalentemente,

$$\nabla h(x^k) \in \partial\left(g(\cdot) + \lambda_k d(\cdot, x^k) + \delta_C(\cdot)\right)(\hat{x}).$$

Mostraremos a seguir que o ponto $\hat{x}$ pertence ao interior do conjunto C.

Afirmação: $\partial\left(g(\cdot) + \lambda_k d(\cdot, x^k) + \delta_C(\cdot)\right)(x) = \emptyset,\ \forall x \in \mathrm{fr}\ C$.

Suponha por absurdo que exista $\xi \in \partial\left(g(\cdot) + \lambda_k d(\cdot, x^k) + \delta_C(\cdot)\right)(x)$ para algum $x \in \mathrm{fr}\ C$. Seja $z \in \mathrm{int}\ C$, então pela convexidade de C

$$y^j = (1 - \varepsilon_j)x + \varepsilon_j z \in \mathrm{int}\ C.$$

Com $\lim\limits_{j\to\infty} \varepsilon_j = 0$, assim $\lim\limits_{j\to\infty} y^j = x \in \mathrm{fr}\ C$. Pela definição de subdiferencial, $\xi \in \partial\left(g(\cdot) + \lambda_k d(\cdot, x^k) + \delta_c(\cdot)\right)(x)$ se, e somente se,

$$\begin{aligned}
\varepsilon_j\langle \xi, z - x\rangle &= \langle \xi, y^j - x\rangle \\
&\leqslant g(y^j) - g(x) + \lambda_k\left[d(y^j, x^k) - d(x, x^k)\right] + \delta_C(y^j) - \delta_C(x) \\
&\leqslant (1 - \varepsilon_j)g(x) + \varepsilon_j g(z) - g(x) + \lambda_k\left\langle \nabla_1 d(y^j, x^k), y^j - x\right\rangle \\
&= \varepsilon_j[g(z) - g(x)] + \lambda_k \cdot \frac{\varepsilon_j}{1 - \varepsilon_j}\left\langle \nabla_1 d(y^j, x^k), z - y^j\right\rangle.
\end{aligned}$$

Onde na segunda desigualdade usamos a convexidade de g, o fato que $d(\cdot, x^k)$ é convexa, diferenciável e $\delta_C(y^j) - \delta_C(x) \leqslant 0$. Daí,

$$\frac{1 - \varepsilon_j}{\lambda_k}\left[g(x) - g(z) + \langle \xi, z - x\rangle\right] \leqslant \left\langle \nabla_1 d(y^j, x^k), z - y^j\right\rangle.$$

Usando o fato que $\lambda_k \leqslant \bar{\lambda}$ e passando o limite com $j \to \infty$,

$$\begin{aligned}\lim_{j\to\infty} \frac{1-\varepsilon_j}{\bar{\lambda}}\left[g(x) - g(z) + \langle \xi, z - x\rangle\right] &\leqslant \lim_{j\to\infty} \frac{1-\varepsilon_j}{\lambda_k}\left[g(x) - g(z) + \langle \xi, z - x\rangle\right] \\ &\leqslant \lim_{j\to\infty} \left\langle \nabla_1 d(y^j, x^k), z - y^j \right\rangle.\end{aligned}$$

Então, pelo fato da distância proximal ser coerciva na fronteira,

$$\frac{g(x) - g(z) + \langle \xi, z - x\rangle}{\bar{\lambda}} \leqslant -\infty.$$

Absurdo! Logo $\partial\left(g(\cdot) + \lambda_k d(\cdot, x^k) + \delta_C(\cdot)\right)(x) = \emptyset,\ \forall x \in \mathrm{fr}\ C$. Como

$$\hat{x} \in C = \mathrm{int}\ C \cup \mathrm{fr}\ C$$

e $\hat{x} \notin \mathrm{fr}\ C$ portanto, $\hat{x} \in \mathrm{int}\ C$ e a afirmação segue tomando $x^{k+1} = \hat{x}$.

$\square$

Observação 3.1.1. *Note que, pelo Teorema 1.1.10, temos que* $\left\{x^{k+1}\right\}$ *é minimizador de* f_k *em* C *se, e somente se,*

$$0 \in \partial(f_k(\cdot) + \delta_C(\cdot))(x^{k+1}),$$

ou seja,

$$0 \in \partial\left(g(\cdot) - \langle \nabla h(x^k), \cdot - x^k\rangle + \lambda_k d(\cdot, x^k) + \delta_C(\cdot)\right)(x^{k+1}).$$

Pelo Teorema 1.1.9,

$$0 \in \partial g(x^{k+1}) - \nabla h(x^k) + \lambda_k \nabla_1 d(x^{k+1}, x^k) + \partial\delta_C(x^{k+1}). \tag{3.5}$$

Como $x^{k+1} \in \mathrm{int}\ C$, *temos que* $\partial\delta_C(x^{k+1}) = N_C(x^{k+1}) = \{0\}$, *logo*

$$0 \in \partial g(x^{k+1}) - \nabla h(x^k) + \lambda_k \nabla_1 d(x^{k+1}, x^k).$$

Portanto,

$$\nabla h(x^k) - \lambda_k \nabla_1 d(x^{k+1}, x^k) \in \partial g(x^{k+1}),\ \forall k \geqslant 0. \tag{3.6}$$

Agora provaremos que MIP é um processo de descida.

Teorema 3.1.1. *A sequência* $\left\{x^k\right\}$ *gerada por MIP satisfaz:*

1) Ou o algoritmo para em um ponto crítico;

2) Ou f *decresce estritamente, isto é,* $f(x^{k+1}) < f(x^k), \forall k \in \mathbb{N}$.

Demonstração. (1) Se $x^{k+1} = x^k$ de (3.5), obtemos que

$$0 \in \partial g(x^k) - \nabla h(x^k) + \partial \delta_C(x^k).$$

Como $x^k \in \mathtt{int}\ C$ e sabemos que $\partial\delta_C(x^k) = N_C(x^k) = \{0\}$. Temos que,

$$0 \in \partial g(x^k) - \nabla h(x^k) = \partial^\circ f(x^k).$$

Temos que x^k é ponto crítico de f.

2) Suponha que $x^{k+1} \neq x^k$, usando (3.6), obtemos

$$g(x^{k+1}) + \langle \nabla h(x^k) - \lambda_k \nabla_1 d(x^{k+1}, x^k), x^k - x^{k+1} \rangle \leqslant g(x^k). \tag{3.7}$$

Por hipótese h é convexa e diferenciável, logo

$$h(x^k) + \langle \nabla h(x^k), x^{k+1} - x^k \rangle \leqslant h(x^{k+1}).$$

Daí,

$$-h(x^{k+1}) - \langle \nabla h(x^k), x^k - x^{k+1} \rangle \leqslant -h(x^k). \tag{3.8}$$

Somando (3.7) e (3.8), obtemos

$$f(x^{k+1}) - \lambda_k \langle \nabla_1 d(x^{k+1}, x^k), x^k - x^{k+1} \rangle \leqslant f(x^k). \tag{3.9}$$

Por (H4) obtemos,

$$\begin{aligned} \langle \nabla_1 d(x^{k+1}, x^k), x^k - x^{k+1} \rangle &\leqslant H(x^k, x^k) - H(x^k, x^{k+1}) \\ &= -H(x^k, x^{k+1}). \end{aligned}$$

Multiplicando a expressão acima por $-\lambda_k < 0$, obtemos

$$\lambda_k H(x^k, x^{k+1}) \leqslant -\lambda_k \langle \nabla_1 d(x^{k+1}, x^k), x^k - x^{k+1} \rangle. \tag{3.10}$$

Usando (3.10) em (3.9)

$$f(x^{k+1}) + \lambda_k H(x^k, x^{k+1}) \leqslant f(x^k). \tag{3.11}$$

Por (H.7) e pelo fato de $0 < \liminf_{k\to\infty} \lambda_k$, temos que $\lambda_k H(x^k, x^{k+1})$ é positivo, logo

$$f(x^{k+1}) < f(x^k).$$

□

Proposição 3.1.2. *Seja* $\{x^k\}$ *uma sequência gerada por MIP. Então,*

$$\lim_{k\to\infty} H(x^k, x^{k+1}) = 0.$$

Demonstração. Por (3.11), obtemos

$$\lambda_k H(x^k, x^{k+1}) \leqslant f(x^{k+1}) - f(x^k).$$

Fazendo $k \to \infty$ e usando (H.7) e o fato que $0 < \liminf_{k\to\infty} \lambda_k$, obtemos

$$0 \leqslant \lim_{k\to\infty} \lambda_k H(x^k, x^{k+1}) \leqslant \lim_{k\to\infty} [f(x^{k+1}) - f(x^k)].$$

Pelo item 2 do Teorema 3.1.1 (f é monótona) e pela hipótese de f ser limitada inferiormente, temos que a sequência $\{f(x^k)\}$ é convergente. Daí, $\lim_{k\to\infty} \lambda_k H(x^k, x^{k+1}) = 0$, usando o fato de que λ_k é limitada. Como $\liminf \lambda_k > 0$, segue que $\lim_{k\to\infty} H(x^k, x^{k+1}) = 0$. □

Lidando com métodos de descida para funções convexas ou quase-convexas, podemos esperar que o algoritmo atinja a convergência total (isto é, a convergência da totalidade da sequência). No entanto, se as funções sob consideração não são necessariamente convexas nem quase-convexas, o método não assegura a convergência total entretanto obtemos resultados de convergência parcial (os pontos de acumulação da sequência gerada pelo algoritmo são pontos críticos de f). Devido a condição de otimalidade (3.5) para obter resultados de convergência (mesmo parcial), é esperado que $\{x^k\}$ tenha um comportamento assintoticamente regular no sentido que $\|\nabla_1 d(x^{k+1}, x^k)\|$ se anule quando $k \to \infty$, veja por exemplo [16, 11] para o caso particular onde $d(x, y) = \|x - y\|^2$.

A seguir, fornecemos resultados de convergência para alguns exemplos específicos de distâncias proximais, a saber, para a escolhas de d como distâncias proximal homogênea de segunda ordem e distância de Bregman.

3.2 Método Proximal Homogêneo de Segunda Ordem

Aqui consideramos o algoritmo MIP com a distância proximal em (3.1) dada pela distância proximal homogênea de segunda ordem,

$$d_\phi(x, y) = \begin{cases} \sum_{i=1}^n y_i^2 \phi\left(\dfrac{x_i}{y_i}\right), & \forall x, y \in \mathbb{R}^n_{++}, \\ \infty, & \text{caso contrário}, \end{cases}$$

e distância proximal induzida dada por,

$$H_d(x,y) = \left(\frac{\mu\varphi''(1)+\nu}{2}\right)\|x-y\|^2, \tag{3.12}$$

e $C = \mathbb{R}^n_+$.

Teorema 3.2.1. *Assuma que a sequência* $\{x^k\}$ *gerada por MIP seja limitada. Então, cada ponto de acumulação de* $\{x^k\}$ *é um ponto crítico de* f.

Demonstração. Seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$, por definição existe uma subsequência $\{x^{k_j}\}$ tal que $\lim_{j\to\infty} x^{k_j} = \bar{x}$ e por (H7) juntamente com a Proposição 3.1.2, temos que $\lim_{j\to\infty} x^{k_j+1} = \bar{x}$. Por (3.1) e pela Observação (3.1.1), existe $\xi^{k_j+1} \in \partial g(x^{k_j+1})$ tal que,

$$\xi^{k_j+1} = \nabla h(x^{k_j}) - \lambda_{k_j}\nabla_1 d_\phi(x^{k_j+1}, x^{k_j}),$$

ou seja,

$$\nabla h(x^{k_j}) - \xi^{k_j+1} = \lambda_{k_j}\nabla_1 d_\phi(x^{k_j+1}, x^{k_j}). \tag{3.13}$$

Como $\xi^{k_j+1} \in \partial g(x^{k_j+1})$ e $\{x^{k_j+1}\}$ é limitado segue pelo Teorema 1.1.13 que $\{\xi^{k_j+1}\}$ é limitado. Portanto, temos que,

$$\lim_{j\to\infty} \xi^{k_j+1} = \bar{\xi} \text{ e } \lim_{j\to\infty} \nabla h(x^{kj}) = \nabla h(\bar{x}),$$

a menos de subsequência. Como ∂g é fechado segue pelo Teorema 1.1.14 que $\bar{\xi} \in \partial g(\bar{x})$. Por (H4), $\forall z \in \mathbb{R}^n_+$,

$$\langle \nabla_1 d_\phi(x^{k_j+1}, x^{k_j}), z - x^{k_j+1}\rangle \leqslant H_d(z, x^{k_j}) - H_d(z, x^{k_j+1}) - \gamma H_d(x^{k_j+1}, x^{k_j}). \tag{3.14}$$

Combinando a desigualdade (3.14) com (3.12),

$$\langle \nabla_1 d_\phi(x^{k_j+1}, x^{k_j}), z - x^{k_j+1}\rangle \;\leqslant\; \left(\frac{\mu\varphi''(1)+\nu}{2}\right)\left[\|z - x^{k_j}\| - \|z - x^{k_j+1}\| - \gamma\cdot\|x^{k_j+1} - x^{k_j}\|\right].$$

Sendo,

$$\gamma = \frac{\nu - \mu\varphi''(1)}{\mu\varphi''(1)+\nu} > 0.$$

Pelo Lema 2.4.1, obtemos

$$\begin{aligned}\langle \nabla_1 d_\phi(x^{k_j+1}, x^{k_j}), x^{k_j+1} - z\rangle \;&\geqslant\; \left(\frac{\nu + \varphi''(1)\mu}{2}\right)(\|z - x^{k_j+1}\|^2 - \|z - x^{k_j}\|^2)\\ &+\; \left(\frac{\nu - \varphi''(1)\mu}{2}\right)\|x^{k_j+1} - x^{k_j}\|^2.\end{aligned}$$

Multiplicando por λ_k e em seguida passando o limite inferior com $j \to \infty$ temos que

$$\liminf_{j\to\infty} \langle \lambda_{k_j} \nabla_1 d_\phi(x^{k_j+1}, x^{k_j}), x^{k_j+1} - z \rangle \geqslant 0, \ \forall z \in \mathbb{R}^n_+.$$

Por (3.13),

$$\liminf_{j\to\infty} \langle \nabla h(x^{k_j}) - \xi^{k_j+1}, x^{k_j+1} - z \rangle \geqslant 0, \ \forall z \in \mathbb{R}^n_+.$$

Daí,

$$\liminf_{j\to\infty} \langle \xi^{k_j+1} - \nabla h(x^{k_j}), z - x^{k_j+1} \rangle \geqslant 0, \ \forall z \in \mathbb{R}^n_+.$$

Portanto,

$$\langle \overline{\xi} - \nabla h(\overline{x}), z - \overline{x} \rangle \geqslant 0, \ \forall z \in \mathbb{R}^n_+.$$

Como $\overline{\xi} - \nabla h(\overline{x}) \in \partial^\circ f(\overline{x})$, temos que $\overline{x}$ é ponto crítico de f. □

3.3 Método Proximal Bregman

Agora considere o algoritmo MIP para escolha particular da distância proximal como a distância de Bregman dada por:

$$D_\varphi(x,y) = \begin{cases} \varphi(x) - \varphi(y) - \langle \nabla \varphi(y), x - y \rangle, & \forall y \in \text{int } C, \\ \infty, & \text{caso contrário.} \end{cases}$$

Teorema 3.3.1. *Assuma que a sequência $\{x^k\}$ gerada por MIP seja limitada. Então, cada ponto de acumulação de $\{x^k\}$ é um ponto crítico de f.*

Demonstração. Seja $\overline{x}$ um ponto de acumulação de $\{x^k\}$, por definição existe uma subsequência $\{x^{k_j}\}$ tal que $\lim\limits_{j\to\infty} x^{k_j} = \overline{x}$ e por (H7) juntamente com a Proposição 3.1.2, temos que $\lim\limits_{j\to\infty} x^{k_j+1} = \overline{x}$. Por (3.1) e pela Observação (3.1.1), existe $\xi^{k_j+1} \in \partial g(x^{k_j+1})$ tal que,

$$\nabla h(x^{k_j}) - \xi^{k_j+1} = \lambda_{k_j} \nabla_1 D_\varphi(x^{k_j+1}, x^{k_j}). \tag{3.15}$$

Como $\xi^{k_j+1} \in \partial g(x^{k_j+1})$ e $\{x^{k_j+1}\}$ é limitado segue pelo Teorema 1.1.13 que $\{\xi^{k_j+1}\}$ é limitado. Portanto, temos que,

$$\lim_{k\to\infty} \xi^{k_j+1} = \overline{\xi} \text{ e } \lim_{j\to\infty} \nabla h(x^{kj}) = \nabla h(\overline{x}),$$

a menos de subsequência. Como ∂g é fechado segue pelo Teorema 1.1.14 que $\overline{\xi} \in \partial g(\overline{x})$. Como a distância de Bregman induzida coincide com a própria distância de Bregman, temos pela Proposição 3.1.2,

$$\lim_{k\to\infty} D_\varphi(x^k, x^{k+1}) = 0,$$

em particular,

$$\lim_{j\to\infty} D_{\varphi}(x^{k_j}, x^{k_j+1}) = 0.$$

Além disso por (B5) temos que,

$$\liminf_{j\to\infty} D_{\varphi}(z, y^{k^j+1}) - D_{\varphi}(z, y^{k^j}) \geqslant 0, \forall z \in C.$$

Pela Proposição 2.3.1,

$$\begin{aligned}\left\langle \nabla_1 D_{\varphi}(x^{k_j+1}, x^{k_j}), x^{k_j} - z\right\rangle &= \left\langle \nabla\varphi(x^{k_j+1}) - \nabla\varphi(x^{k_j}), x^{k_j} - z\right\rangle \\ &= D_{\varphi}(z, x^{k_j+1}) - D_{\varphi}(z, x^{k_j}) - D_{\varphi}(x^{k_j}, x^{k_j+1}).\end{aligned}$$

Multiplicando por λ_k e em seguida passando o limite inferior com $j \to \infty$ temos que

$$\liminf_{j\to\infty} \langle \lambda_{k_j} \nabla_1 D_{\varphi}(x^{k_j+1}, x^{k_j}), x^{k_j} - z\rangle \geqslant 0, \ \forall z \in C.$$

Por (3.15),

$$\liminf_{j\to\infty} \langle \nabla h(x^{k_j}) - \xi^{k_j+1}, x^{k_j} - z\rangle \geqslant 0, \ \forall z \in C.$$

Daí,

$$\liminf_{j\to\infty} \langle \xi^{k_j+1} - \nabla h(x^{k_j}), z - x^{k_j}\rangle \geqslant 0, \ \forall z \in C.$$

Portanto,

$$\langle \overline{\xi} - \nabla h(\overline{x}), z - \overline{x}\rangle \geqslant 0, \ \forall z \in C.$$

Como $\overline{\xi} - \nabla h(\overline{x}) \in \partial^{\circ} f(\overline{x})$ segue que $\overline{x}$ é um ponto crítico de f. □

3.4 Exemplo

Como mencionado no início deste capítulo, Hamdi [6] propôs um Método Interior Proximal para diferença de funções convexas. Dado uma função DC, $f(x) = g(x) - h(x)$, o seguinte Método Interior Proximal Bregman:

$$x^{k+1} = (\nabla\varphi + \lambda\partial g)^{-1}(\nabla\varphi(x^k) + \lambda z_k), \tag{3.16}$$

onde $\nabla\varphi$ denota o gradiente de φ, ∂g denota o subdiferencial de g no sentido de análise convexa, $z^k \in \partial h(x^k)$ e $\lambda > 0$. Note que, aplicando $(\nabla\varphi + \lambda\partial g)$ na expressão (3.16), obtemos

$$(\nabla\varphi + \lambda\partial g)(x^{k+1}) = \nabla\varphi(x^k) + \lambda z^k.$$

Isto é,

$$\frac{\nabla\varphi(x^k) - \nabla\varphi(x^{k+1})}{\lambda} + z^k \in \partial g(x^{k+1}). \tag{3.17}$$

Assumindo que o conjunto S dos pontos críticos de f é não vazio e combinando (3.16) com a monotocidade de ∂g, temos que

$$0 \leqslant \left\langle \frac{\nabla\varphi(x^k) - \nabla\varphi(x^{k+1})}{\lambda} + z^k - \bar{z}, x^{k+1} - \bar{x} \right\rangle,$$

ou seja,

$$\frac{1}{\lambda}\langle \nabla\varphi(x^k) - \nabla\varphi(x^{k+1}), \bar{x} - x^{k+1}\rangle \leqslant \langle z^k - \bar{z}, x^{k+1} - \bar{x}\rangle. \tag{3.18}$$

Para todo $\bar{z} \in \partial h(\bar{x}) \cap \partial g(\bar{x})$, $\forall k \geqslant 0$.

Pergunta 1: A desigualdade (3.18) implica na seguinte desigualdade

$$\langle \nabla\varphi(x^k) - \nabla\varphi(x^{k+1}), \bar{x} - x^{k+1}\rangle \leqslant 0, \tag{3.19}$$

para algum $k_0 \in \mathbb{N}$ com $k > k_0$ e para qualquer ponto crítico $\bar{x}$ de f?

Através da desigualdade (3.19) o autor provou a convergência total da sequência para um ponto crítico de f. Sendo assim, forneceremos um contra exemplo para ilustrar que a resposta para a Pergunta 1 é negativa.

Exemplo 3.4.1. *Seja $f : \mathbb{R} \to \mathbb{R}$ uma função DC real dada por,*

$$f(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2.$$

Seja $D_\varphi(x, y) = \frac{1}{2}(x - y)^2$ a distância de Bregman associada à $\varphi(x) = \frac{1}{2}x^2$. Por simplicidade assumimos que $\lambda = 1$. Assim, o Método Interior Proximal Bregman (3.17) assume a seguinte forma,

$$x_k - x_{k+1} + x_k = x_{k+1}^3,$$

ou seja,

$$x_{k+1}^3 + x_{k+1} - 2x_k = 0. \tag{3.20}$$

É fácil provar que para cada $a \in (0, 1)$, a equação $x^3 + x - 2a = 0$ possui uma única raiz real o qual pertence à $(0, 1)$. Então, se (3.20) inicia no ponto $x_0 \in (0, 1)$, então $x_k \in (0, 1)$, $\forall k \in \mathbb{N}$. Provaremos por indução em k que a sequência $\{x_k\}$ gerada por (3.20) é crescente.

Suponha que $x_{k-1} \leqslant x_k$ *satisfaça,*

$$x_k^3 + x_k - 2x_{k-1} = 0, \tag{3.21}$$

escolha $x_{k+1} \in \mathbb{R}$ *tal que,*

$$x_{k+1}^3 + x_{k+1} - 2x_k = 0, \tag{3.22}$$

subtraindo (3.22) de (3.21), obtemos

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

Usando a hipótese, que $x_{k-1} \leqslant x_k$

$$x_{k+1}^3 - x_k^3 + x_{k+1} - x_k \geqslant 0,$$

então,

$$(x_{k+1} - x_k)(x_{k+1}^2 + x_{k+1}x_k + x_k^2 + 1) \geqslant 0.$$

Portanto, $0 \leqslant x_{k+1} - x_k$.

Agora note que, os pontos críticos de f são 0, -1 e 1. De (3.19), defina

$$\gamma_{k+1}(\overline{x}) := \frac{1}{\lambda}\langle \nabla\varphi(x_k) - \nabla\varphi(x_{k+1}), \overline{x} - x_{k+1}\rangle.$$

Aplicando em nosso exemplo com $\gamma = 1$,

$$\gamma_{k+1}(\overline{x}) = (x_{k+1} - x_k)(x_{k+1} - \overline{x}).$$

Portanto, $\gamma_{k+1}(0) = x_{k+1}(x_{k+1} - x_k)$ e se (3.20) inicia no ponto $x_0 \in (0,1)$, nós temos que $\gamma_k(0) > 0$, $\forall k \in \mathbb{N}$ com $\{x^k\}$ crescente e convergindo para 1 quando $k \to \infty$. Isto prova a resposta negativa para a Pergunta 1. O que podemos concluir como apresentado na seção 3.3 que sob a hipótese de limitação da sequência, os pontos de acumulação são pontos críticos da função f.

Capítulo 4

Considerações Finais

Nesta dissertação foi realizado o estudo de um Método Interior Proximal Linearizado para encontrar um ponto crítico de funções DC, $f(x) = g(x) - h(x)$ sob um conjunto fechado, convexo, não vazio C e com interior diferente de vazio. Sob as hipóteses de f ser localmente Lipschitz, semicontinua inferiormente, limitada inferiormente e g, h funções semicontinuas inferiormente, convexas com h diferenciável, mostramos que o método proposto gera um sequência a qual está bem definida, isto é, existe cada iterada e está contida no interior do conjunto C. Além disso, sob as hipóteses do conjunto dos pontos críticos de f em C ser não vazio e limitação da sequência, mostramos para o caso particular de escolhas de distância proximal, a saber: Distância Homogênea de Segunda Ordem e Distância de Bregman, que os pontos de acumulação da sequência gerada pelo método são pontos críticos de f em C. Por fim apresentamos um contra exemplo para a afirmação de Hamdi [6], o qual através dela ele provou que a sequência gerada pelo método converge para os pontos críticos de f em C.

Referências Bibliográficas

[1] Auslender A, Teboulle M. Interior gradient and proximal methods for convex and conic optimization. SIAM J Optim. 2006;16:697–725.

[2] Auslender A, Teboulle M. Interior projection-likemethods formonotone variational inequalities.Math Program.2005;104:39–68.

[3] Burachik R, Dutta J. Inexact Proximal point methods for variational inequality problems. SIAM J Optim. 2010;20(5):2653–2678.

[4] Burachik RS, Scheimberg S. A proximal point method for the variational inequality problem in Banach spaces. SIAM J Control Optim. 2000;39(5):1633–1649.

[5] Clarke FH. Optimization and nonsmooth analysis. Vol. 5, Classics in applied mathematics. Philadelphia: SIAM;1990.

[6] Hamdi A. Modified Bregman proximal scheme to minimize the difference of two convex functions. Appl Math e-notes. 2006;6:132–140.

[7] J. X. Cruz Neto, J. O. Lopes, P. S. M. Santos and J. C. O. Souza (2018): An interior proximal linearized method for DC programming based on Bregman distance or secondorder homogeneous kernels.

[8] Martinet B. Regularisation d'iné quations variationelles par approximations succesives. Rev Française d'Inform Recherche Oper. 1970;4:154–159.

[9] Moreau JJ. Proximité et dualité dans un espace Hilbertien. Bull Soc Math France. 1965;93:273–299.

[10] Moudafi, A., Maingé, P.-E.: On the convergence of an approximate proximalmethod for d.c. functions. J. Comput. Math. 24, 475–480 (2006)

[11] Moudafi A, Maingé P-E. On the convergence of an approximate proximal method for DC functions. J Comput Math. 2006;24:475–480.

[12] Rockafellar RT. Convex Analysis. New Jersey: Princeton University Press; 1972.

[13] Rockafellar RT. Monotone operators and the proximal point algorithm. SIAM J control optim. 1976;14:877–898.

[14] Solodov MV, Svaiter BF. An inexact hybrid generalized proximal point algorithm and some new results on the theory of bregman functions. Math Oper Res. 2000;25:214–230.

[15] Souza, J.C.O., Oliveira, P.R.: A proximal point algorithm for DC functions on Hadamard manifolds. J. Glob. Optim. (2015).

[16] Souza JCO, Oliveira PR, Soubeyran A. Global convergence of a proximal linearized algorithm for difference of convex functions. Optim Lett. 2016;10(7):1529–1539.

[17] Sun, W., Sampaio, R.J.B., Candido, M.A.B.: Proximal point algorithm for minimization of DC Functions. J. Comput. Math. 21, 451–462 (2003)

[18] Tao PD, Souad EB. Algorithms for solving a class of nonconvex optimization problems.Methods of subgradients. In: Hiriart-Urruty J-B, editor. FERMAT Days 85: mathematics for optimization. North-Holland mathematics studies;1986;129:249–271.

[19] Tiel, JV. Convex Analysis. De Bilt: John Wiley and Sons Ltd; 1984.