Universidade Federal de Campina Grande
Centro de Ciências e Tecnologia
Programa de Pós-Graduação em Matemática
Curso de Mestrado em Matemática

Trajetória Central, Métodos de Ponto Proximal Generalizado e Trajetória de Cauchy em Variedades Riemannianas

por

Marco Antonio Lázaro Velásquez †

sob orientação do

Prof. Dr. João Xavier da Cruz Neto

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

†Este trabalho contou com apoio financeiro da CAPES.

Trajetória Central, Métodos de Ponto Proximal Generalizado e Trajetória de Cauchy em Variedades Riemannianas

por

Marco Antonio Lázaro Velásquez

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Matemática

Aprovada por:

Prof. Dr. Paulo Roberto Oliveira

Prof. Dr. José de Arimatéia Fernandes

Prof. Dr. João Xavier da Cruz Neto

Orientador

Universidade Federal de Campina Grande

Centro de Ciências e Tecnologia

Programa de Pós-Graduação em Matemática

Curso de Mestrado em Matemática

Março/2007

Agradecimentos

- A minha mãe Victorina que do céu guia e abençoa meus passos.
- A meu pai Alvaro, por seus conselhos e esforço.
- A minha tia Susana, por sua dedicação. Ela se comportou como uma mãe.
- A meus irmãos Julio, David e José, pelo apoio contínuo. Eles são exemplos de dedicação e persistência.
- A menha esposa Yvonne, pelo carinho, incentivo e compreensão durante a realização do curso de Mestrado.
- A meu amigo Antonio Luiz Soares Santos e sua família: Tania, Allyson, Ramon e Vitor, pelo apoio constante desde o primeiro dia que cheguei ao Brasil. Eu aprendi muito com cada um deles.
- Ao professor João Xavier que me orientou no desenvolvimento deste trabalho. Sempre paciente e compreensivo.
- Aos professores Paulo Roberto Oliveira e José de Arimatéia Fernandes pela disponibilidade nesta tarefa de me avaliar, fazendo parte da banca examinadora.
- Aos colegas de curso, pela amizade e ajuda mútua.
- A todos que fazem parte do Departamento de Matemática e Estatística da Universidade Federal de Campina Grande.
- A todos que fazem parte do Departamento de Matemática da Universidade Federal de Piauí.
- Finalmente, a todos que direta ou indiretamente contribuíram para a conclusão deste trabalho.

Dedicatória

À memória de minha mãe Victorina.

À minha esposa Yvonne.

Resumo

Em problemas de otimização convexa e, de maneira geral, em problemas de inequações variacionais aparecem os conceitos de: trajetória central (definida por uma função barreira), algoritmo de ponto proximal generalizado (com distâncias de Bregman) e trajetória de Cauchy em variedades de Riemannianas.

Nesta disertação são estudados os três conceitos e suas possíveis relações. Estas relações são dadas principalmente para programação linear.

Primeiro é mostrado, com hipóteses adequadas, que a trajetória central está bem definida, é limitada, contínua, possui pontos de acumulação e converge para o centro analítico do conjunto de soluções.

Depois, também com hipóteses adequadas, é provado que a seqüência gerada pelo algoritmo de ponto proximal generalizado converge para uma solução do problema de inequações varacionais.

Um fato importante é quando a trajetória central é definida pela distância de Bregman como função barreira. Nestas considerações, é mostrado que a trajetória central e a seqüência gerada pelo algoritmo de ponto proximal generalizado convergem para o mesmo ponto.

Além disso, para programação linear é mostrado que a seqüência gerada pelo algoritmo de ponto proximal generalizado está contida na trajetória central.

Finalmente, é mostrado para programação linear que a trajetória central também coincide com a trajetória de Cauchy em variedades Riemannianas definidas em subconjuntos abertos de $\mathbb{R}^n$ com métrica dada pelo hessiano da função barreira.

Abstract

In convex otimization problems and, more generally, in variational inequality problems appears concepts of: central paths defined by a barrier function, generalized proximal point algorithm with Bregman's distances and Cauchy trajectory in Riemannian manifolds.

In this work are studed these three concepts and its possible relationships. These relationships are showed principally to linear programming.

First is showed, with adequate hypotheses, that a central path is well defined, is bounded, is continuos, have cluster points, these cluster points are solutions of variational inequality problems and converge to the analytic center of the solution set.

Next, with adequate hypotheses too, is showed that a sequence generated by the generalized proximal point algorithm converge to someone solution of variational inequality problem.

An important fact is when a central path is defined by the Bregman's distance as a barrier function. In these cases, is showed that a central path and the sequence generated by the generalized proximal point algorithm converges to the same point.

Furthermore, to linear programming is showed that the sequence generated by the generalized proximal point algorithm is contained in the central path.

Finally, is showed to linear programming that a central path also coincides with a Cauchy trajectory in the Riemannian manifold defined on the open subsets of $\mathbb{R}^n$ with metric given by the hessian of the barrier function.

Conteúdo

Notações

$B(x,\varepsilon)$	:	bola aberta em $I\!R^n$ com centro em x e raio ε.
$B[x,\varepsilon]$	:	bola fechada em $I\!R^n$ com centro em x e raio ε.
$I\!R^n_+$	:	conjunto de vetores de $I\!R^n$ cujas coordenadas são não negativas.
$I\!R^n_{++}$	:	conjunto de vetores de $I\!R^n$ cujas coordenadas são positivas.
$I\!R^{m\times n}$	:	conjunto de matrizes reais de ordem $m \times n$.
$\mathcal{P}(I\!R^n)$	:	conjunto das partes de $I\!R^n$.
$dom(T)$	:	domínio do operador T.
$R(T)$	:	imagem do operador T.
$VIP(T,C)$	:	problema de inequações variacionais para T no conjunto C.
$S(T,C)$	:	conjunto de soluções de $VIP(T,C)$.
$Gap_{\,T,C}$	:	função gap de $VIP(T,C)$.
$rk(A)$	:	posto da matriz A.
A^t	:	trasposta da matriz A.
A^s	:	parte simetrica da matriz A.
$Ker(A)$	:	núcleo da matriz A.
$Im(A)$	:	imagem da matriz A.
$J_T(x)$	:	matriz Jacobiana de T em x.
$C^{\,0}$	:	interior do conjunto C.
$\mathrm{cl}(C)$	:	fecho do conjunto C.
∂C	:	fronteira do conjunto C.
∇f	:	gradiente de f.
∂f	:	subdiferencial de f.

δ_V	:	função indicadora do conjunto V.
N_V	:	operador normalizado do conjunto V.
$D_g(\cdot,\cdot)$	:	distância de Bregman induzida por g.
$ED(f)$	:	domínio efetivo de h.
$GPPA$	:	algoritmo ponto proximal generalizado.
$L_{f,D}(c)$	:	conjunto de nivel de f associado a $c \in I\!R$ no conjunto $D \subset I\!R^n$.
$\mathrm{conv}(D)$	:	fecho convexo do conjunto $D \subset I\!R^n$.
$aff(S)$	:	fecho afim do conjunto S.
$ri(C)$	:	interior relativo do conjunto C.
$P_C(x)$	:	projeção de $x \in I\!R^n$ ao conjunto C.
$H_{s,r}$	:	hiperplano suporte ao conjunto C em $x \in \partial C$.
E_f	:	epígrafo de f.
$\dfrac{\partial f}{\partial d}(x)$	:	derivada parcial de f em x na direcção d.
f^*	:	função conjugada da função f.
f^{**}	:	função conjugada da função conjugada de f.
KKT	:	Karush-Kuhn-Tucker.
PF	:	factibilidade primal.
DF	:	factibilidade dual.
CS	:	condição complementar de folga.
(U_α, x_α)	:	sistema de coordenadas de uma variedade diferenciável no ponto p.
T_pM	:	conjunto dos vetores tangentes da variedade M em p.
$C^\infty(M)$	:	conjunto das funções $f : M \to I\!R$ diferenciáveis em p.
$d\varphi_p$	:	diferencial de φ em p.
TM	:	fibrado tangente da variedade M.
$\mathfrak{X}(M)$	:	conjunto dos campos de vetores diferenciáveis em M.
$\langle\cdot,\cdot\rangle$	:	métrica Riemanniana.
g_{ij}	:	expresão da métrica Riemanniana num sistema de coordenadas.
∇	:	conexão afim em uma variedade diferenciável.
$\mathrm{grad}\, f$	:	grafiente de $f \in C^\infty(M)$.
H_p^f	:	hessiano de $f \in C^\infty(M)$ em p.

Introdução

Començamos com uma breve exposição daqueles tópicos que serão essenciais para uma boa compreenssão dos problemas e algoritmos que serão tratados neste trabalho.

Dada uma função $f : I\!\!R^n \to I\!\!R$ convexa e limitada inferiormente, o problema de otimização convexa é definido como

$$(P_0) \qquad \left\{ \min_{x \in I\!\!R^n} f(x). \right.$$

Quando x está restrito a um conjunto convexo e fechado $C \subset I\!\!R^n$, obtem-se o problema de otimização convexo com restrições

$$(P_1) \qquad \left\{ \min_{x \in C} f(x). \right.$$

Traduzindo esses problemas em termos de ∂f, isto é, do subdiferencial de f, eles podem ser escritos da seguinte forma

$$(P_2) \qquad \left\{ \begin{array}{rl} Encontre & x^* \in I\!\!R^n \\ tal\ \ que & 0 \in \partial f(x^*), \end{array} \right.$$

e

$$(P_3) \qquad \left\{ \begin{array}{cccc} Encontre & x^* \in C & tal\ que\ para\ algum & \\ w \in \partial f(x^*) & tem - se & \langle w, y - x^* \rangle \geq 0, & para\ todo\ y \in C; \end{array} \right.$$

respectivamente.

Para uma função convexa f é um fato conhecido (veja Teorema (1.1)) que o operador subdiferencial $\partial f : I\!\!R^n \to \mathcal{P}(I\!\!R^n)$ é monótono maximal. Obtemos assim, extensões naturais do problema sem restrições (P_2), e do problema com restrições (P_3), substituindo o operador ∂f por qualquer operador monótono maximal $T : I\!R^n \to \mathcal{P}(I\!R^n)$. Logo, são definidos os seguintes problemas

$$(P_4) \qquad \begin{cases} Encontre \quad x^* \in I\!R^n \\ \quad tal \; que \quad 0 \in T(x^*), \end{cases}$$

e

$$(P_5) \qquad \begin{cases} Encontre \quad x^* \in C \quad tal \; que \; para \; algum \\ w \in T(x^*) \quad tem-se \quad \langle w, y - x^* \rangle \geq 0, \quad para \; todo \; y \in C; \end{cases}$$

respectivamente.

O problema (P_4) é um problema canônico associado a um operador monótono maximal T, de forma específica: encontrar um zero de T.

Nesta disertação, a nossa atenção estará centrada no problema (P_5); isto é, o chamado *"Problema de Inequações Variacionais"* para o operador T no conjunto C, que de agora em diante será denotado por $VIP(T, C)$.

Para definir *"Trajetória Central"* para $VIP(T, C)$ é preciso considerar uma função barreira h para o conjunto C; isto é, uma função convexa e diferenciável no interior de C e tal que a norma de seu gradiente ∇h tende para $+\infty$ na fronteira de C. A trajetória central, denotada por $\{x(\mu)/\mu > 0\}$, para $VIP(T, C)$ com barreira h é definida pela relação

$$-\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu)),$$

para cada $\mu > 0$.

Esta definição é uma generalização das trajetórias centrais que aparecem em métodos de pontos interiores para programação linear, onde as funções barreiras são da forma logarítmica $h(x) = -\sum_{j=1}^{n} w_j log x_j$, $w_j > 0$ (Adler e Monteiro [1]). Segue da definição de trajetória central a seguinte pergunta:

(Q1) *Sob que condições a trajetória central existe, possui ponto de acumulação e este ponto é solução do $VIP(T, C)$?*

Quando $VIP(T, C)$ possui mais de uma solução, é preciso conhecer o comportamento de $x(\mu)$ quando μ tende para $+\infty$. Assim, é apresentada a seguinte pergunta:

(Q2) *Qual é o comportamento da trajetória central com relação ao conjunto de soluções de $VIP(T,C)$?*

Por outro lado, um algoritmo importante para encontrar zeros de operadores monótonos maximais é o *"Algoritmo de Ponto Proximal"*, cujas propiedades fundamentais foram provadas por Rockafellar [17]. Este algoritmo gera, para qualquer ponto inicial x^0, uma seqüência $\{x^k\}_{k\in I\!N}$ da seguinte forma

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

onde $\{\lambda_k\}_{k\in I\!N}$ é uma seqüência de números positivos. Prova-se em Rockafellar [17] que a seqüência gerada por (1) converge a uma solução de $(P4)$, isto é, a um zero de T, sempre que o conjunto desses zeros seja não vazio e os parâmetros λ_k sejam limitados superiormente.

Se $T = \partial f$, para alguma função convexa f, então (1) pode ser escrito da forma

$$x^{k+1} = \underset{x\in I\!R^n}{arg\ min}\left\{ f(x) + \frac{1}{2}\lambda_k\|x - x^k\|^2 \right\}, \tag{2}$$

que é equivalente a

$$0 \in \partial\left\{ f(\,\cdot\,) + \frac{1}{2}\lambda_k\|\cdot - x^k\|^2 \right\}(x^{k+1}). \tag{3}$$

O caráter irrestrito do algoritmo dado por (2) está refletido na distância euclidiana ([9], [11]).

Para o caso restrito, isto é, para problemas de inequações varacionais, é preciso substituir a distância euclidiana por outra cujo comportamento no conjunto de restrições C seja análogo àquele da norma em $I\!R^n$. Esquematicamente, o *"Algoritmo de Ponto Proximal Generalizado"* $(GPPA)$ para $VIP(T,C)$ pode ser escrito como:

$$0 \in T(x^{k+1}) + \lambda_k \nabla_x D_g\left(x^{k+1}, x^k\right),$$

onde o operador subdiferencial da expressão (3) foi substituido por um operador monótono maximal T arbitrario, e a distância euclidiana por uma distância generalizada $D_g(\cdot,\cdot)$ chamada distância de Bregman induzida por g. Este algoritmo foi estudado por Regina Burachik e Alfredo Iusem ([4], [5], [12]).

Assim, são apresentadas as seguintes questões:

(Q3) *Sob que condições a seqüência gerada por $GPPA$ existe e converge para uma solução de $VIP(T,C)$?*

(Q4) *Qual é o comportamento da seqüência gerada por $GPPA$ com relação ao conjunto de soluções de $VIP(T,C)$?*

As possíveis respostas para as perguntas (Q1) e (Q3) estariam resolvendo o problema de inequações variacionais (P_5).

Suponhamos, para alguns casos, que os limites $\lim_{\mu\to+\infty} x(\mu)$ e $\lim_{k\to+\infty} x^k$ existem e são soluções de $VIP(T,C)$ onde $x(\mu)$ são os pontos da trajetória central e x^k os elementos da seqüência gerada por $GPPA$. Logo, de forma natural, aparece a seguinte questão:

(Q5) *Dado um problema $VIP(T,C)$, em que casos existe alguma relação entre trajetória central e $GPPA$?*

Outro conceito importante em otimização é *"Trajetória de Cauchy"*. Dada uma função f definida numa variedade Riemanniana M, é possível definir seu gradiente, denotado por grad f, com respeito à métrica de M. A trajetória de Cauchy é a curva $x : [0,\beta] \to M$ tal que

$$(P_6) \qquad \begin{cases} x(0) &= \quad x^0 \\ \dot{x}(t) &= -\operatorname{grad} f(x(t))\,, \end{cases}$$

onde $\beta > 0$. É bem conhecido que para qualquer x^0, existe $\beta > 0$ tal que (P_6) possui uma única solução; isto é, a trajetória de Cauchy existe e é única dado um ponto inicial. Por último, é feita a seguinte pergunta:

(Q6) *Em que casos existe alguma relação entre trajetória central para problemas de inequações variacionais e trajetória de Cauchy em variedades Riemannianas?*

A seguir é dada uma descrição do conteúdo dos capítulos que são direcionados para dar respostas a todas as questões feitas anteriormente.

O primeiro capítulo contém definições e alguns resultados da teoria de operadores monótonos. É mostrado que o subdiferencial de uma função convexa é um operador monótono maximal e paramonótono. No final deste capítulo são mostradas propiedades de operadores ponto a ponto monótonos continuamente diferenciáveis que serão utilizados nos capítulos seguintes.

No segundo capítulo são analisadas as hipóteses que têm que satisfazer T, C, e h para que a trajetória central exista. Com estas hipóteses é mostrado que a trajetória central está bem definida, é limitada, está contida no interior de C e possui pontos de acumulação que são soluções de $VIP(T, C)$.

No início do capítulo 3 são definidas as funções e distâncias generalizadas de Bregman e são mostradas suas principais propriedades. O algoritmo de ponto proximal generalizado é analisado quando C é um conjunto convexo fechado qualquer e quando C é um conjunto poliedral. Neste último, o operador é considerado ponto a ponto monótono e continuamente diferenciável. No final desta seção são analisadas as possíveis relações entre trajetória central e $GPPA$ para problemas de programação linear.

No início do quarto capítulo é definida a trajetória de Cauchy em variedades Riemannianas e depois, são estabelecidas possíveis relações entre trajetória de Cauchy e trajetória central para problemas de programação linear.

No apêndice A estão as definições e resultados de análise convexa que são utilizadas em toda a disertação.

No apêndice B encontra-se uma ferramenta muito poderosa em otimização, as condições de otimalidade de Karush-Kuhn-Tucker, que são freqüentemente utilizadas nesta disertação.

O apêndice C contém definições e resultados de geometria Riemanniana que são necessárias para o capítulo 4.

Por último, as conclusões desta disertação são dadas depois do quarto capítulo.

Capítulo 1

Operadores Monótonos

Neste capitulo são apresentadas às definições e principais resultados de operadores monótonos. A maioria dos fatos foram obtidos dos trabalhos de Iusem [12], [13].

1.1 Operadores Monótonos Maximais

Seja $T : \mathbb{R}^n \to \mathbb{R}^n$ uma transformação linear semidefinida positiva, isto é, $\langle x, T(x) \rangle \geq 0$, para todo $x \in \mathbb{R}^n$. Logo, para qualquer x e y em $\mathbb{R}^n$, temos

$$0 \leq \langle x - y, T(x - y) \rangle = \langle x - y, T(x) - T(y) \rangle.$$

Um operador monótono é uma generalização de uma transformação linear semidefinida positiva ao caso não linear.

Definição 1.1.1 *Seja $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador (não necessariamente linear).*

i) *T é chamado monótono quando*

$$\langle T(x) - T(y), x - y \rangle \geq 0\,, \ \ \forall x, y \in \mathbb{R}^n. \tag{1.1}$$

ii) *T é chamado estritamente monótono quando é monótono e, em adição*

$$\langle T(x) - T(y), x - y \rangle = 0 \quad \Longrightarrow \quad x = y. \tag{1.2}$$

Exemplo 1.1.1 *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função convexa e diferenciável. Então $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ é um operador monótono (segue do Teorema A.9).*

Seja $f : I\!R^n \to I\!R$ uma função convexa, não necessariamente diferenciável. Pela Proposição A.6.1, para cada $x \in I\!R^n$, temos que o conjunto $\partial f(x)$ é não vazio, convexo e compacto. Seja $\mathcal{P}(I\!R^n)$ o conjunto das partes de $I\!R^n$. Como $\partial f : I\!R^n \to \mathcal{P}(I\!R^n)$ asocia para cada $x \in I\!R^n$ não exatamente um vetor, mas um subconjunto de $I\!R^n$, então é preciso estender a noção de operador monótono ponto a ponto para operadores ponto - conjunto.

Definição 1.1.2 *Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador ponto - conjunto.*

i) *T é chamado monótono quando para todo $x \in I\!R^n$, $y \in I\!R^n$, $u \in T(x)$ e $v \in T(y)$, é válido:*

$$\langle u - v, x - y \rangle \geq 0. \tag{1.3}$$

ii) *T é chamado estritamente monótono quando é monótono e, em adição, para qualquer $u \in T(x)$ e $v \in T(y)$:*

$$\langle u - v, x - y \rangle = 0 \qquad \Longrightarrow \qquad x = y. \tag{1.4}$$

Exemplo 1.1.2 *$\partial f : I\!R^n \to \mathcal{P}(I\!R^n)$ é um operador monótono, onde $f : I\!R^n \to I\!R$ é uma função convexa.*

De fato. Sejam $x \in I\!R^n$, $y \in I\!R^n$, $u \in \partial f(x)$ e $v \in \partial f(y)$. Da Definição A.6.1, $\langle u, y - x \rangle \leq f(y) - f(x)$ e $\langle v, x - y \rangle \leq f(x) - f(y)$. Logo, $\langle u, y - x \rangle + \langle v, x - y \rangle \leq 0$. Isto é, $\langle u - v, y - x \rangle \leq 0$. Assim, $\langle u - v, x - y \rangle \geq 0$. ■

Definição 1.1.3 *Um operador $T : I\!R^n \to \mathcal{P}(I\!R^n)$ é chamado monotóno maximal quando*

i) *T é monótono.*

ii) *Se existe operador monótono $T' : I\!R^n \to \mathcal{P}(I\!R^n)$ tal que $T(x) \subset T'(x)$, para todo $x \in I\!R^n$, então $T(x) = T'(x)$, para todo $x \in I\!R^n$.*

O objetivo principal desta seção é mostrar que o subdiferencial de uma função convexa é um operador monótono maximal.

Seja $f : I\!R^n \to I\!R$ uma função convexa e suponha que o problema

$$\left\{ \min_{x \in I\!R^n} f(x) \right. \tag{1.5}$$

possui solução. Logo, f é limitada inferiormente. Seja $z \in I\!R^n$ fixo e defina

$$\begin{array}{rcccl} F & : & I\!R^n & \longrightarrow & I\!R \\ & & x & \longmapsto & F(x) \ := \ f(x) + \frac{1}{2} \parallel x - z \parallel^2 . \end{array} \tag{1.6}$$

Se $h(u) = \frac{1}{2} \parallel u - z \parallel^2$, $\forall u \in I\!R^n$, então $F(x) = f(x) + h(x)$, $\forall x \in I\!R^n$.

Temos que h é estritamente convexa. De fato. Sejam $u_1 \in I\!R^n$ e $u_2 \in I\!R^n$, $u_1 \neq u_2$. Então $\nabla h(u_1) = u_1 - z$ e $\nabla h(u_2) = u_2 - z$. Logo,

$$\langle \nabla h(u_1) - \nabla h(u_2), u_1 - u_2 \rangle = \langle u_1 - u_2, u_1 - u_2 \rangle = \parallel u_1 - u_2 \parallel^2 > 0.$$

Pelo Corolario A.10, h é estritamente convexa.

Como f é convexa e h é estritamente convexa, pela Proposição A.5.1, $F = f + h$ é estritamente convexa. Logo, pelo Teorema A.14, F é contínua.

Além disso, F é coerciva em $I\!R^n$ (veja Definição A.2.2). De fato. Como f é limitada inferiormente, existe $\beta \in I\!R$ tal que $f(x) \geq \beta$, $\forall x \in I\!R^n$. Seja $\{x^k\}_{k \in I\!N}$ uma seqüência em $I\!R^n$ tal que $\lim\limits_{k \to +\infty} \parallel x^k \parallel = +\infty$. Segue que

$$\lim_{k \to +\infty} h(x^k) = \lim_{k \to +\infty} \frac{1}{2} \parallel x^k - z \parallel^2 = +\infty.$$

Daí,

$$+\infty = \lim_{k \to +\infty} \beta + h(x^k) \leq \lim_{k \to +\infty} f(x^k) + h(x^k) = \lim_{k \to +\infty} F(x^k).$$

Logo, $\limsup\limits_{k \to +\infty} F(x^k) = +\infty$ e portanto F é coerciva em $I\!R^n$.

Segue do Corolário A.3 que a função F possui um minimizador global. Como F é estritamente convexa, o minimizador global é único (veja Teorema A.12). Logo, podemos definir

$$\begin{array}{rcccl} prox_f & : & I\!R^n & \longrightarrow & I\!R^n \\ & & z & \longmapsto & prox_f(z) \ := \ \underset{x \in I\!R^n}{arg\,min} \left\{ f(x) + \frac{1}{2} \parallel x - z \parallel^2 \right\}. \end{array} \tag{1.7}$$

De forma análoga, defina

$$\begin{array}{rcccl} G & : & I\!R^n & \longrightarrow & I\!R \\ & & y & \longmapsto & G(y) \ := \ f^*(y) + \frac{1}{2} \parallel y - z \parallel^2 . \end{array} \tag{1.8}$$

onde f^* é a função conjugada de f (veja Definição A.7.5).

Como f é convexa, então f^* é convexa (veja Teorema A.17). Assim, $G = f^* + h$ é estritamente convexa (veja Proposição A.5.1). Logo, pelo Teorema A.14, G é função contínua.

Com argumentos similares feitos para F, é possível mostrar que G também é coerciva em $I\!R^n$.

Segue do Corolario A.3 que a função G possui um minimizador global. Como G é estritamente convexa, o minimizador global é único (veja Teorema A.12). Logo, podemos definir

$$\begin{array}{rcccl} prox_{f^*} & : & I\!R^n & \longrightarrow & I\!R^n \\ & & z & \longmapsto & prox_{f^*}(z) \;:=\; \underset{y\in I\!R^n}{arg\,min} \left\{ f^*(y) + \frac{1}{2} \parallel y - z \parallel^2 \right\}. \end{array} \tag{1.9}$$

Lema 1.1 *Seja $f : I\!R^n \to I\!R$ uma função convexa e considere x, y e z em $I\!R^n$. São equivalentes:*

i) $z = x + y$ *e* $f(x) + f^*(y) = \langle x, y\rangle$.

ii) $x = prox_f(z)$ *e* $y = prox_{f^*}(z)$.

Demonstração. Suponha que (i) é válido. Como $f(x) + f^*(y) = \langle x, y\rangle$ então, pelo Teorema A.16, $y \in \partial f(x)$. Como $0 = y + (x - z)$ e $y \in \partial f(x)$, então

$$0 \in \partial f(x) + \{x - z\} = \partial f(x) + \partial h(x) = \partial F(x).$$

Pelo Teorema A.15, x é minimizador de F. Sendo F estritamente convexa então $x = prox_f(z)$.

Por outro lado, pela Proposição A.7.2, $f = f^{**}$. Logo, $f^*(y) + f^{**}(x) = \langle y, x\rangle$ implica que $x \in \partial f^*(y)$ (veja Teorema A.16). Como $0 = x + (y - z)$ e $x \in \partial f^*(y)$, então

$$0 \in \partial f^*(y) + \{y - z\} = \partial f^*(y) + \partial h(y) = \partial G(y).$$

Pelo Teorema A.15, y é minimizador de G. Sendo G estritamente convexa então $x = prox_{f^*}(z)$.

Reciprocamente, suponha que (ii) é válido. Como $x = prox_f(z)$ então $0 \in$

$\partial F(x) = \partial f(x) + \{x - z\}$ (veja Teorema A.15). Logo, existe $y \in \partial f(x)$ tal que $0 = y + x - z$. Daí, $z = y + x$. Além disso, pelo Teorema A.16, como $y \in \partial f(x)$ então $f^*(y) = \langle x, y\rangle - f(x)$. Logo,

$$f(x) + f^*(y) = \langle x, y\rangle \tag{1.10}$$

De forma análoga, como $y = prox_{f^*}(z)$, então $0 \in \partial G(y) = \partial f^*(y) + \{y - z\}$ (veja Teorema A.15). Logo, existe $\xi \in \partial f^*$ tal que $0 = \xi + (y - z)$. Daí, $z = \xi + y$. Além disso, pelo Teorema A.16, como $\xi \in \partial f^*$, então $f^{**}(\xi) = \langle y, \xi\rangle - f(y)$. Pela Proposição A.7.2, $f^{**} = f$. Assim,

$$f(\xi) + f^*(y) = \langle y, \xi\rangle \tag{1.11}$$

Como $z = x + y$ e $z = \xi + y$ então $x = \xi$ e assim as equações (1.10) e (1.11) coincidem. Portanto $z = x + y$ e $f(x) + f^*(y) = \langle x, y\rangle$. ■

Teorema 1.1 *Se $f : I\!R^n \to I\!R$ é uma função convexa então $\partial f : I\!R^n \to \mathcal{P}(I\!R^n)$ é um operador monótono maximal.*

Demonstração. Pelo Exemplo 1.1.2, ∂f é operador monótono. Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono tal que $\partial f(x) \subset T(x)$, para todo $x \in I\!R^n$. Mostremos que $T(x) \subset \partial f(x)$, para todo $x \in I\!R^n$.

De fato. Seja $y_0 \in T(x_0)$, com $x_0 \in I\!R^n$. Considere $x_1 = prox_f(x_0 + y_0)$ e $y_2 = prox_{f^*}(x_0 + y_0)$. Pelo Lema 1.1, $x_0 + y_0 = y_1 + x_1$ e $f(x_1) + f^*(y_1) = \langle x_1, y_1\rangle$. Logo, $x_0 - x_1 = y_1 - y_0$, $\;y_1 \in \partial f(x_1)$, $\;x_1 \in \partial f^*(y_1)$ (veja Teorema A.16).

Sendo T monótono,

$$0 \leq \langle x_1 - x_0, y_1 - y_0\rangle = -\langle x_0 - x_1, y_1 - y_0\rangle = -\langle y_1 - y_0, y_1 - y_0\rangle = -\parallel y_1 - y_0 \parallel^2 \leq 0.$$

Segue que $y_1 = y_0$ e $x_1 = x_0$. Logo, $y_0 = y_1 \in \partial f(x_1) = \partial f(x_0)$. Como y_0 e x_0 são arbitrários, $T(x) \subset \partial f(x)$, $\forall x \in I\!R^n$. Assim, $T(x) = \partial f(x)$, para todo $x \in I\!R^n$.

Portanto $\partial f : I\!R^n \to \mathcal{P}(I\!R^n)$ é um operador monótono maximal. ■

Definição 1.1.4 *Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono. O domínio de T é o conjunto*

$$dom(T) = \{x \in I\!R^n \;/\; T(x) \neq \phi\}.$$

Definição 1.1.5 *Seja $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ um operador monótono. A imagem de T é o conjunto*

$$R(T) = \bigcup_{x \in \, dom(T)} T(x).$$

Definição 1.1.6 *Dado $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$, o operador $T^{-1} : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é definido pela seguinte relação:* $\quad y \in T^{-1}(x) \Leftrightarrow x \in T(y)$.

Definição 1.1.7 *Um ponto $\bar{x} \in \mathbb{R}^n$ é chamado zero de $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ quando $0 \in T(\bar{x})$.*

Observação 1.1.1 *Se $T = \partial f$, para alguma função convexa $f : \mathbb{R}^n \to \mathbb{R}$, temos que $0 \in \partial f(\bar{x})$ se, e somente se, $\bar{x}$ é solução do problema (1.5) (veja Teorema A.15). Assim, o problema de procurar zeros de um operador monótono maximal generaliza o problema de minimizar funções convexas em $\mathbb{R}^n$.*

1.2 Operadores Paramonótonos

Definição 1.2.1 *Um operador $T : \mathbb{R}^n \to \mathbb{R}^n$ é paramonótono quando é monótono e, em adição,*

$$\langle T(x) - T(y), x - y \rangle = 0 \qquad \Longrightarrow \qquad T(x) = T(y). \tag{1.12}$$

A extensão desta definição para operadores ponto - conjunto é

Definição 1.2.2 *Um operador $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ é paramonótono quando é monótono e, em adição, para $u \in T(x)$ e $v \in T(y)$*

$$\langle u - v, x - y \rangle = 0 \qquad \Longrightarrow \qquad u \in T(y)\,,\ v \in T(x). \tag{1.13}$$

O seguinte resultado mostra que o subdiferencial de uma função convexa satisfaz a relação (1.13).

Teorema 1.2 *Se $T = \partial f$, para alguma função convexa $f : \mathbb{R}^n \to \mathbb{R}$, então T é paramonótono.*

Demonstração. Pelo Exemplo 1.1.2, ∂f é um operador monótono. Sejam $x, y \in \mathbb{R}^n$ e considere $\langle u - v, x - y \rangle = 0$, com $u \in \partial f(x)$, $v \in \partial f(y)$. Defina

$$\begin{array}{rcccl} \widehat{f} & : & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & & z & \longmapsto & \widehat{f}(z) \ := \ f(z) + \langle u, x - z \rangle. \end{array} \tag{1.14}$$

Temos que $\widehat{f}$ é convexa pois se $z \in I\!\!R^n$, $w \in I\!\!R^n$ e $\alpha \in [0,1]$ então

$$\begin{aligned}
\widehat{f}(\alpha z + (1-\alpha)w) &= f(\alpha z + (1-\alpha)w) + \langle u, x - \alpha z - (1-\alpha)w\rangle \\
&\leq \alpha f(z) + (1-\alpha)f(w) + \langle u, x + \alpha x - \alpha x - \alpha z - (1-\alpha)w\rangle \\
&= \alpha f(z) + (1-\alpha)f(w) + \alpha\langle u, x - z\rangle + (1-\alpha)\langle u, x - w\rangle \\
&= \alpha\left(f(z) + \langle u, x - z\rangle\right) + (1-\alpha)\left(f(w) - \langle u, x - w\rangle\right) \\
&= \alpha\widehat{f}(z) + (1-\alpha)\widehat{f}(w).
\end{aligned}$$

Além disso, $\partial\widehat{f}(z) = \partial f(z) - \{u\} = \{w - u / w \in \partial f(z)\}$. Quando $z = x$ tem-se que $\widehat{f}(x) = f(x)$. Quando $w = u \in \partial f(x)$ então $0 \in \partial\widehat{f}(x)$. Assim, x é um minimizador de $\widehat{f}$ em $I\!\!R^n$ (veja Teorema A.15).

Como $\langle u - v, x - y\rangle = 0$ então $\langle u, x - y\rangle = \langle v, x - y\rangle$. Como $u \in \partial f(x)$ então $\langle u, y - x\rangle \leq f(y) - f(x)$. Daí, $f(x) - f(y) \leq \langle u, x - y\rangle$. Analogamente, como $v \in \partial f(y)$ então $\langle v, x - y\rangle \leq f(x) - f(y)$. Logo,

$$f(x) - f(y) \leq \langle u, x - y\rangle = \langle v, x - y\rangle \leq f(x) - f(y). \tag{1.15}$$

De (1.15) segue que $f(x) = f(y) + \langle u, x - y\rangle = \widehat{f}(y)$. Mas $\widehat{f}(x) = f(x)$. Logo, $\widehat{f}(x) = \widehat{f}(y)$. Assim, y é tambem um minimizador de $\widehat{f}$ em $I\!\!R^n$. Pelo Teorema A.15, $0 \in \partial\widehat{f}(y)$.

Além disso, $0 = w - u$, para algum $w \in \partial f(y)$. Logo, $u \in \partial f(y)$, pois $w = u$ para algum $w \in \partial f(y)$.

Defina

$$\begin{array}{rcccl}
\widehat{g} : & I\!\!R^n & \longrightarrow & I\!\!R & \\
& z & \longmapsto & \widehat{g}(z) & := f(z) + \langle v, y - z\rangle.
\end{array} \tag{1.16}$$

Temos que $\widehat{g}$ é convexa e $\partial\widehat{g}(z) = \partial f(z) - \{v\} = \{w - v / w \in \partial f(z)\}$. Quando $z = y$ tem-se que $\widehat{g}(y) = f(y)$. Quando $w = v \in \partial f(y)$ então $0 \in \partial\widehat{g}(y)$. Pelo Teorema A.15, y é um minimizador de $\widehat{g}$ em $I\!\!R^n$.

De (1.15) segue que $f(y) = f(x) + \langle v, y - x\rangle = \widehat{g}(x)$. Mas $\widehat{g}(y) = f(y)$. Logo, $\widehat{g}(x) = \widehat{g}(y)$. Assim, x é também um minimizador de $\widehat{g}$ em $I\!\!R^n$. Pelo Teorema A.15, $0 \in \partial\widehat{g}(x)$.

Além disso, $0 = w - v$, para algum $w \in \partial f(x)$. Logo, $v \in \partial f(x)$, pois $w = v$ para algum $w \in \partial f(x)$.

Assim, se $\langle u - v, x - y \rangle = 0$, com $u \in \partial f(x)$ e $v \in \partial f(y)$, então $u \in \partial f(y)$ e $v \in \partial f(x)$. Portanto, $T = \partial f$ é operador paramonótono. ■

Proposição 1.2.1 *Se $T_1 : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ e $T_2 : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ são operadores paramonótonos, então $T_1 + T_2$ é paramonótono.*

Demonstração. Sejam $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, $u \in (T_1 + T_2)(x)$ e $v \in (T_1 + T_2)(y)$. Como $u \in T_1(x) + T_2(x)$, existem $u_1 \in T_1(x)$ e $u_2 \in T_2(x)$ tais que $u = u_1 + u_2$. Da mesma forma, existem $v_1 \in T_1(y)$ e $v_2 \in T_2(y)$ tais que $v = v_1 + v_2$. Sendo T_1 e T_2 operadores monótonos, $\langle u_1 - v_1, x - y \rangle \geq 0$ e $\langle u_2 - v_2, x - y \rangle \geq 0$. Logo,

$$\begin{aligned} \langle u - v, x - y \rangle &= \langle u_1 + u_2 - v_1 - v_2, x - y \rangle \\ &= \langle u_1 - v_1, x - y \rangle + \langle u_2 - v_2, x - y \rangle \geq 0. \end{aligned}$$

Assim, $T_1 + T_2$ é monótono.

Além disso, se $\langle u - v, x - y \rangle = 0$ então $\langle u_1 + u_2 - v_1 - v_2, x - y \rangle = 0$. Daí, $\langle u_1 - v_1, x - y \rangle + \langle u_2 - v_2, x - y \rangle = 0$. Sendo T_1 e T_2 monótonos, $\langle u_1 - v_1, x - y \rangle = 0$ e $\langle u_2 - v_2, x - y \rangle = 0$. Como T_1 e T_2 são paramonótonos, $u_1 \in T_1(y)$, $v_1 \in T_1(x)$, $u_2 \in T_2(y)$ e $v_2 \in T_2(x)$. Logo, $u = u_1 + u_2 \in T_1(y) + T_2(y) = (T_1 + T_2)(y)$ e $v = v_1 + v_2 \in T_1(x) + T_2(x) = (T_1 + T_2)(x)$.

Portanto, $T_1 + T_2$ é paramonótono. ■

Proposição 1.2.2 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e fechado com interior não vazio e $T : \mathbb{R}^n \to \mathbb{R}^n$ um operador monótono continuamente diferenciável em C. Seja $J_T(x)$ a matriz Jacobiana de T em x. Então:*

i) *$J_T(x)$ é semidefinida positiva.*

ii) *$Ker\,(J_T(x)) = Ker\,(J_T(x)^t)$, $Im\,(J_T(x)) = Im\,(J_T(x)^t)$, para todo $x \in C$.*

iii) *Se $Ker\,(J_T(x)) = Ker\,(J_T(x)^s)$, $\forall x \in C$, então T é paramonótono em C.*

Demonstração.

i) Seja $z \in \mathbb{R}^n$, fixe $x \in C^0$ e considere $y = x - \lambda z$, com $\lambda \in (0, 1)$, tal que $y \in C$. Seja $w_\alpha = x + \alpha(y - x)$, com $\alpha \in (0, \lambda) \subset (0, 1)$. Pela convexidade de C, $w_\alpha \in C$. Como T é monótono,

$$0 \leq \langle T(w_\alpha) - T(x), w_\alpha - x \rangle = \alpha \langle T(w_\alpha) - T(x), y - x \rangle.$$

Então $0 \leq \langle \frac{1}{\alpha} \, [\, T(w_\alpha) - T(x)\,]\, , y - x\, \rangle$. Daí, $0 \leq \langle \lim_{\alpha \to 0} \dfrac{1}{\alpha} [\, T(w_\alpha) - T(x)\,]\, , y - x \rangle$. Assim, $0 \leq (x-y)^t J_T(x)(x-y)$. Logo, $J_T(x)$ é semidefinida positiva.

ii) Seja M uma matriz semidefinida positiva. Afirmamos que $Ker(M) = Ker(M^t)$.

De fato. Observe que

$$
\begin{aligned}
x^t M x &= \frac{1}{2} x^t M x + \frac{1}{2} x^t M x = \frac{1}{2} x^t M x + \frac{1}{2} (Mx)^t x \\
&= \frac{1}{2} x^t M x + \frac{1}{2} x^t M^t x = x^t \left[\frac{1}{2} \left(M + M^t \right) \right] x = x^t M^s x,
\end{aligned}
$$

onde $M^s = \frac{1}{2}\left(M + M^t\right)$. Tem-se que M^s é simétrica e semidefinida positiva.

Seja $x \in Ker(M)$; isto é $Mx = 0$. Então, $0 = x^t M x = x^t M^s x$. Sendo M^s simétrica e semidefinida positiva, $M^s x = 0$. Isto é, $\frac{1}{2}\left(Mx + M^t x\right)$. Como $Mx = 0$ então $M^t x = 0$. Assim, $x \in Ker(M^t)$.

Seja $x \in Ker(M^t)$; isto é $M^t x = 0$. Então $x^t M^t x = 0$. Daí, $x^t (x^t M^t)^t = 0$. Segue que $x^t M x = 0$. Sendo M^s simétrica e semidefinida positiva, $M^s x = 0$. Isto é, $\frac{1}{2}\left(Mx + M^t x\right)$. Como $M^t x = 0$ então $Mx = 0$. Assim, $x \in Ker(M)$.

Portanto, $Ker(M) = Ker(M^t)$, para toda matriz M semidefinida positiva.

Observe que $Im(M^t) = Ker(M)^\perp = Ker(M^t)^\perp = Im((M^t)^t) = Im(M)$, para toda matriz M semidefinida positiva. Pelo item (i), $J_T(x)$ é semidefinida positiva.

Portanto, $Ker\left(J_T(x)\right) = Ker\left(J_T(x)^t\right)$ e $Im\left(J_T(x)\right) = Im\left(J_T(x)^t\right)$, $\forall x \in C$.

iii) Sejam $x \in C$, $y \in C$ e considere $0 = \langle T(x) - T(y), x - y \rangle$. Para $\alpha \in (0,1)$ seja $w_\alpha = y + \alpha(x-y) = \alpha x + (1-\alpha) y$. Pela convexidade de C, $w_\alpha \in C$.

Como $w_\alpha = y + \alpha(x-y)$ então $x - y = \frac{1}{\alpha}(w_\alpha - y)$. Como $w_\alpha = \alpha x + (1-\alpha)y$ então $x - y = \frac{1}{1-\alpha}(x - w_\alpha)$. Logo,

$$
\begin{aligned}
0 &= \langle T(x) - T(y), x - y \rangle \\
&= \langle T(x) - T(w_\alpha), x - y \rangle + \langle T(w_\alpha) - T(y), x - y \rangle \\
&= \frac{1}{1-\alpha} \langle T(x) - T(w_\alpha), x - w_\alpha \rangle + \frac{1}{\alpha} \langle T(w_\alpha) - T(y), w_\alpha - y \rangle.
\end{aligned} \tag{1.17}
$$

Sendo T monótono, $\langle T(x) - T(w_\alpha), x - w_\alpha \rangle \geq 0$ e $\langle T(w_\alpha) - T(y), w_\alpha - y \rangle \geq 0$. Logo, de (1.17) segue

$$
0 = \frac{1}{1-\alpha} \langle T(x) - T(w_\alpha), x - w_\alpha \rangle = \langle T(x) - T(w_\alpha), x - y \rangle,
$$

$$0 = \frac{1}{\alpha}\langle T(w_\alpha) - T(y), w_\alpha - y\rangle = \langle T(w_\alpha) - T(y), w_\alpha - y\rangle.$$

Daí,

$$0 = \langle T(x) - T(w_\alpha), x - y\rangle = \langle T(w_\alpha) - T(y), w_\alpha - y\rangle. \tag{1.18}$$

Considere $\beta \in (0, 1)$ e $\gamma \in (0, 1 - \beta)$. Assim, $\beta + \gamma < 1$. De (1.18), $0 = \langle T(w_\beta) - T(y), x - y\rangle = \langle T(w_{\beta+\gamma}) - T(y), x - y\rangle$. Então

$$0 = \langle T(w_\beta) - T(y) - T(w_{\beta+\gamma}) + T(y), x - y\rangle.$$

Daí, $0 = \langle \lim_{\gamma\to 0} \frac{1}{\gamma} [\, T(w_\beta) - T(w_{\beta+\gamma})\,] , x - y\rangle$. Assim, $0 = (x - y)J_T(w_\beta)(x - y)$, para todo $\beta \in (0, 1)$. Logo, $0 = (x - y)J_T(w_\beta)^s(x - y)$, para todo $\beta \in (0, 1)$. Como $J_T(w_\beta)^s$ é semidefinida positiva então $J_T(w_\beta)^s(x-y) = 0$. Logo, $x-y \in Ker(J_T(w_\beta)^s)$. Por hipótese, $Ker(J_T(w_\beta)^s) = Ker(J_T(w_\beta))$. Assim, $x - y \in Ker(J_T(w_\beta))$, isto é, $J_T(w_\beta)(x - y) = 0$, para todo $\beta \in (0, 1)$. Sendo T continuamente diferenciável, $J_T(w_\beta)(x - y) = 0$, para todo $\beta \in [0, 1]$.

Sejam agora T_j , $1 \leq j \leq n$, as componentes de T. Então $T_j(y) = T_j(x) + \nabla T_j(w_{\beta_j})^t(x - y)$, para algum $\beta_j \in [0, 1]$. Note que $\nabla T_j(w_{\beta_j})$ é a j-ésima linha de $J_T(w_\beta)$. Logo, $\nabla T_j(w_{\beta_j})^t(x - y) = 0$. Assim, $T_j(y) = T_(x)$. Segue que $T(x) = T(y)$. Portanto, T é paramonótono. ■

1.3 Problemas de Inequações Variacionais

A extensão natural do problema

$$\begin{cases} \min\limits_{x\in C} f(x), \end{cases} \tag{1.19}$$

onde f é uma função convexa e $C \subset I\!R^n$ é um conjunto convexo, para operadores monótonos é chamada problema de inequações variacionais.

Definição 1.3.1 *Sejam $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono maximal e $C \subset I\!R^n$ um conjunto convexo e fechado. O problema de inequações variacionais, denotado por $VIP(T, C)$, consiste em procurar $z \in C$ tal que para algum $u \in T(z)$, tem-se*

$$\langle u, x - z\rangle \geq 0, \tag{1.20}$$

para todo $x \in C$. Denota-se ao conjunto de soluções do problema $VIP(T, C)$ por $S(T, C)$.

Observação 1.3.1 *Se* $T = \partial f$, *para alguma função convexa* $f : C \subset \mathbb{R}^n \to \mathbb{R}$, *temos que* $u \in \partial f(z)$ *se, e só se* $0 \leq \langle u, x - z \rangle \leq f(x) - f(z), \forall x \in C$ *(veja Definição A.6.1). Isto é,* $z \in C$ *é solução do problema (1.19). Assim, o problema* $VIP(T, C)$ *generaliza o problema de minimizar funções convexas em conjuntos convexos.*

Proposição 1.3.1 *Sejam* $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ *um operador paramonótono e* $C \subset \mathbb{R}^n$ *um conjunto convexo e fechado. Suponha que* $z \in S(T, C)$. *Então* $x \in C$ *é solução do problema* $VIP(T, C)$ *se, e somente se, existe* $u \in T(x)$ *tal que* $\langle u, z - x \rangle \geq 0$.

Demonstração. Suponha que $x \in S(T, C)$. Então existe $u \in T(x)$ tal que $\langle u, y - x \rangle \geq 0$, para todo $y \in C$. Como $z \in C$ então $\langle u, z - x \rangle \geq 0$.

Reciprocamente, suponha que $\langle u, z - x \rangle \geq 0$, para algum $u \in T(x)$. Como $z \in S(T, C)$, existe $v \in T(z)$ tal que $\langle v, y - z \rangle \geq 0$, para todo $y \in C$. Em particular, $\langle v, x - z \rangle \geq 0$, pois $x \in C$.

Como T é monótono, $\langle u - v, x - z \rangle \geq 0$. Daí, $0 \leq \langle v, x - z \rangle \leq \langle u, x - z \rangle$. Assim, $\langle v, x - z \rangle = \langle u, x - z \rangle = 0$. Logo, $\langle u - v, x - z \rangle = 0$. Sendo T paramonótono, $u \in T(z)$ e $v \in T(x)$. Então, para todo $y \in C$, temos

$$\langle v, y - x \rangle = \langle v, y - z \rangle + \langle v, z - x \rangle = \langle v, y - z \rangle + 0 \geq 0,$$

com $v \in T(x)$. Logo, $x \in S(T, C)$. ■

Proposição 1.3.2 *Sejam* $C \subset \mathbb{R}^n$ *um conjunto convexo e fechado,* $T : \mathbb{R}^n \to \mathbb{R}^n$ *um operador paramonótono em* C, $\tilde{x} \in S(T, C)$ *e*

$$U = \left\{ x \in C \,/\, T(x) = T(\tilde{x}), T(\tilde{x})^t x = T(\tilde{x})^t \tilde{x} \right\}.$$

Então $S(T, C) = U$.

Demonstração. Seja $x \in U$. Então $T(\tilde{x})^t x = T(\tilde{x})^t \tilde{x}$. Daí, $0 = \langle T(\tilde{x}), \tilde{x} - x \rangle$. Assim, $0 = \langle T(x), \tilde{x} - x \rangle$. Como $\tilde{x} \in S(T, C)$ e $0 = \langle T(x), \tilde{x} - x \rangle$, pela Proposição 1.3.1, $x \in S(T, C)$. Logo, $U \subset S(T, C)$.

Seja $x \in S(T, C)$. Então $x \in C$. Sendo T monótono, $0 \leq \langle T(\tilde{x}) - T(x), \tilde{x} - x \rangle$. Daí, $\langle T(x), \tilde{x} - x \rangle \leq \langle T(\tilde{x}), \tilde{x} - x \rangle$. Como $x \in S(T, C)$ então $\langle T(x), \tilde{x} - x \rangle \geq 0$. Como $\tilde{x} \in S(T, C)$ então $\langle T(\tilde{x}), x - \tilde{x} \rangle \geq 0$.

Logo, $0 \leq \langle T(x), \tilde{x} - x \rangle \leq \langle T(\tilde{x}), \tilde{x} - x \rangle \leq 0$. Assim, $\langle T(\tilde{x}), \tilde{x} - x \rangle = 0$ e segue que $\langle T(\tilde{x}), \tilde{x} \rangle = \langle T(\tilde{x}), x \rangle$. Logo,

$$T(\tilde{x})^t \tilde{x} = T(\tilde{x})^t x. \tag{1.21}$$

Além disso, como T é paramonótono,

$$\langle T(x) - T(\tilde{x}), x - \tilde{x}\rangle = 0 \quad\Longrightarrow\quad T(x) = T(\tilde{x}). \tag{1.22}$$

De (1.21) e (1.22) segue que $x \in U$. Logo, $S(T, C) \subset U$.

Portanto, $S(T, C) = U$. ■

Proposição 1.3.3 *Sejam $C \subset I\!R^n$ um conjunto convexo e fechado com interior não vazio, $T : I\!R^n \to I\!R^n$ um operador monótono continuamente diferenciável tal que para algum subespaço V temos*

$$Ker(J_T(x)) = Ker(J_T(x))^s = V\ ,\ \ \forall x \in C\,;$$

$\tilde{x} \in S(T, C)$, $\widehat{x} \in C$ arbitrário, $B = J_T(\widehat{x})$ e

$$\overline{U} = \left\{\, x \in C \,/\, Bx = B\tilde{x}\ ,\ T(\widehat{x})^t x = T(\widehat{x})^t \tilde{x} \,\right\}.$$

Então

i) *$T(x') - T(x) \in Ker(B)^{\perp}$, para todo x e x' em C.*

ii) *$\overline{U} = S(T, C)$.*

Demonstração. Pelos items (iii) e (ii) da Proposição 1.2.2, temos que T é paramonótono em C e $Ker\,(J_T(x)) = Ker\,(J_T(x)^t)$, $Im\,(J_T(x)) = Im\,(J_T(x)^t)$, $\forall x \in C$.

i) Sejam $x \in C$, $x' \in C$ e considere $\tau \in [0, 1]$. Observe que $T(x') - T(x)$ pode ser escrito da forma

$$T(x') - T(x) = \int_0^1 J_T(x + \tau(x' - x))(x' - x)d\tau. \tag{1.23}$$

Seja $v(\tau) = J_T(x + \tau(x' - x))(x' - x)$. Veja que $x + \tau(x' - x) \in C$, $\forall \tau \in [0, 1]$, pois C é convexo. Então

$$\begin{aligned} v(\tau) \in Im(J_T(x + \tau(x' - x))) &= Im(J_T(x + \tau(x' - x))^t) \\ &= Ker(J_T(x + \tau(x' - x)))^{\perp} = Ker(B)^{\perp}, \end{aligned}$$

$\forall \tau \in [0, 1]$. Segue que $\int_0^1 v(\tau)d\tau \in Ker(B)^{\perp}$. Logo, de (1.23), $T(x') - T(x) \in Ker(B)^{\perp}$, para todo x e x' em C.

ii) Como T é paramonótono, pela Proposição 1.3.2, é suficiente mostrar que $\overline{U} = U$. De fato.

Seja $x \in \overline{U}$. Então $Bx = B\tilde{x}$ e $T(\widehat{x})^t x = T(\widehat{x})^t \tilde{x}$. Segue que $0 = B(\tilde{x} - x)$ e assim, $\tilde{x} - x \in Ker(B)$. Pelo item (i), $T(\tilde{x}) - T(\widehat{x}) \in Ker(B)^{\perp}$ e $T(\tilde{x}) - T(x) \in Ker(B)^{\perp}$. Então, $\langle T(\tilde{x}) - T(\widehat{x}), \tilde{x} - x\rangle = \langle T(\tilde{x}) - T(x), \tilde{x} - x\rangle = 0$.

Como T é paramonótono, $T(x) = T(\tilde{x})$. Além disso,

$$\begin{aligned} 0 = [T(\tilde{x}) - T(\widehat{x})]^t (\tilde{x} - x) &= T(\tilde{x})^t(\tilde{x} - x) - T(\widehat{x})^t(\tilde{x} - x) \\ &= T(\tilde{x})^t(\tilde{x} - x) - T(\widehat{x})^t\tilde{x} + T(\widehat{x})^t x = T(\tilde{x})^t(\tilde{x} - x). \end{aligned}$$

Assim, $T(x) = T(\tilde{x})$ e $T(\tilde{x})^t x = T(\tilde{x})^t \tilde{x}$. Daí, $x \in U$. Logo, $\overline{U} \subset U$.

Seja $x \in U$. Pela Proposição 1.3.2, $U = S(T, C)$. Então $x \in S(T, C)$. Isto é, $\langle T(x), \tilde{x} - x\rangle \geq 0$. Além disso, como $\tilde{x} \in S(T, C)$ então $\langle T(\tilde{x}), x - \tilde{x}\rangle \geq 0$. Logo, $\langle T(x), \tilde{x} - x\rangle + \langle T(\tilde{x}), x - \tilde{x}\rangle \geq 0$. Daí, $\langle T(x) - T(\tilde{x}), \tilde{x} - x\rangle \geq 0$.

Assim, $\langle T(\tilde{x}) - T(x), \tilde{x} - x\rangle \leq 0$. Como T é monótono, $0 \leq \langle T(\tilde{x}) - T(x), \tilde{x} - x\rangle$. Logo,

$$\langle T(\tilde{x}) - T(x), \tilde{x} - x\rangle = 0. \tag{1.24}$$

Seja $x(\tau) = \tau x + (1 - \tau)\tilde{x} = \tilde{x} + \tau(x - \tilde{x})$, com $\tau \in (0, 1)$. Sendo

$$x(\tau) = \tilde{x} + \tau(x - \tilde{x})$$

então $x - \tilde{x} = \dfrac{1}{\tau}(x(\tau) - \tilde{x})$. Também, como

$$x - x(\tau) = x - \tau x - (1 - \tau)\tilde{x}$$

segue que $x - \tilde{x} = \dfrac{1}{1 - \tau}(x - x(\tau))$. De (1.24), temos

$$\begin{aligned} 0 &= \langle T(x) - T(x(\tau)), x - \tilde{x}\rangle + \langle T(x(\tau)) - T(\tilde{x}), x - \tilde{x}\rangle \\ &= \frac{1}{1 - \tau} \langle T(x) - T(x(\tau)), x - x(\tau)\rangle + \frac{1}{\tau} \langle T(x(\tau)) - T(\tilde{x}), x(\tau) - \tilde{x}\rangle. \end{aligned}$$

Sendo T monótono,

$$\langle T(x) - T(x(\tau)), x - x(\tau)\rangle \geq 0\ ,\ \ \langle T(x(\tau)) - T(\tilde{x}), x(\tau) - \tilde{x}\rangle \geq 0.$$

Logo,

$$\begin{aligned} 0 &= \frac{1}{\tau} \langle T(x(\tau)) - T(\tilde{x}), x(\tau) - \tilde{x}\rangle \\ &= \frac{1}{1 - \tau} \langle T(x) - T(x(\tau)), x - x(\tau)\rangle \\ &= \frac{1}{1 - \tau} \langle T(x) - T(x(\tau)), (1 - \tau)(x - \tilde{x})\rangle \\ &= \langle T(x) - T(x(\tau)), x - \tilde{x}\rangle. \end{aligned}$$

Assim, $0 = \langle T(x) - T(x(\tau)), x - \tilde{x} \rangle$. Diferenciando com relação a τ segue que $0 = (x-\tilde{x})^t J_T(x(\tau))(x-\tilde{x}) = (x-\tilde{x})^t J_T(x(\tau))^s (x-\tilde{x})$.

Como $J_T(x(\tau))^s$ é simétrica e semidefinida positiva, $J_T(x(\tau))^s(x-\tilde{x}) = 0$. Isto é, $x - \tilde{x} \in Ker(J_T(x(\tau))^s) = V = Ker(B)$. Daí, $B(x-\tilde{x}) = 0$. Logo, $Bx = B\tilde{x}$.

Pelo item (i), $T(\tilde{x}) - T(\widehat{x}) \in Ker(B)^{\perp}$. Logo, como $x - \tilde{x} \in Ker(B)$ então

$$\begin{aligned} 0 &= \langle T(\tilde{x}) - T(\widehat{x}), x - \tilde{x} \rangle = \langle T(\tilde{x}), x - \tilde{x} \rangle + \langle T(\widehat{x}), \tilde{x} - x \rangle \\ &= \langle T(\tilde{x}), x \rangle - \langle T(\tilde{x}), \tilde{x} \rangle + \langle T(\widehat{x}), \tilde{x} - x \rangle \\ &= T(\tilde{x})^t x - T(\tilde{x})^t \tilde{x} + \langle T(\widehat{x}), \tilde{x} - x \rangle = \langle T(\widehat{x}), \tilde{x} - x \rangle, \end{aligned}$$

pois $x \in U$. Assim, $0 = T(\widehat{x})^t(\tilde{x} - x)$ e daí $T(\widehat{x})^t x = T(\widehat{x})^t \tilde{x}$.

Como $Bx = B\tilde{x}$ e $T(\widehat{x})^t x = T(\widehat{x})^t \tilde{x}$, então $x \in \overline{U}$. Logo, $U \subset \overline{U}$.

Portanto $U = \overline{U}$. ■

Capítulo 2

Trajetória Central para Problemas de Inequações Variacionais

Neste capitulo são estudadas as condições que tem que satisfazer T e C para que a trajetória central do problema $VIP(T, C)$ exista. Também é estudado o comportamento da trajetoria central no conjunto de soluções de $VIP(T, C)$. A maioria dos fatos foram obtidos de [6].

2.1 Trajetória Central e Centro Analítico

Considere o problema $VIP(T, C)$, para algum operador monótono maximal $T : I\!R^n \to \mathcal{P}(I\!R^n)$ e algum conjunto convexo fechado $C \subset I\!R^n$, com $C^0 \neq \phi$.

Considere uma função barreira h para C. A função $h : C^0 \to I\!R$ é considerada estritamente convexa e diferenciável tal que a norma de seu gradiente ∇h tende para $+\infty$ em pontos da fronteira de C.

Definição 2.1.1 *Para cada $\mu \in I\!R_{++}$, seja $x(\mu) \in C$ tal que*

$$-\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu)). \tag{2.1}$$

O conjunto $\{x(\mu)/\mu > 0\}$ é chamado trajetória central para o problema $VIP(T, C)$ com barreira h.

Na seção 2.2 são analizadas as condições sobre T, C e h que são necessárias para garantir a existência e unicidade da trajetória central.

Observação 2.1.1 *Se $T = \partial f$, para alguma função convexa $f : C \subset I\!R^n \rightarrow I\!R$, então* (2.1) *equivale a*

$$x(\mu) = argmin_{\,y\in\, C}\left\{\mu f(y) + h(y)\right\}. \tag{2.2}$$

De fato. De (2.1), $-\frac{1}{\mu}\nabla h(x(\mu)) \in \partial f(x(\mu))$. Logo, para todo $y \in C$, temos

$$-\frac{1}{\mu}\langle\nabla h(x(\mu)), y - x(\mu)\rangle \leq f(y) - f(x(\mu)).$$

Como h é estritamente convexa, $[h(y) - h(x(\mu))] > \langle\nabla h(x(\mu)), y - x(\mu)\rangle$. Logo,

$$-\frac{1}{\mu}\left[h(y) - h(x(\mu))\right] < -\frac{1}{\mu}\langle\nabla h(x(\mu)), y - x(\mu)\rangle.$$

Segue que,

$$-\frac{1}{\mu}h(x(\mu)) - \frac{1}{\mu}h(y) < f(y) - f(x(\mu)), \forall y \in C.$$

Assim,

$$\mu f(x(\mu)) + h(x(\mu)) < \mu f(y) + h(y), \forall y \in C.$$

Portanto, (2.1) é equivalente a (2.2). ■

O problema de interesse é o comportamento de $x(\mu)$ quando μ tende para $+\infty$. Isto é, analizar se existe o limite de $x(\mu)$ quando μ tende para $+\infty$; e no caso afirmativo, analizar se $\lim\limits_{\mu\rightarrow+\infty} x(\mu)$ é solução de $VIP(T, C)$.

Definição 2.1.2 *Seja h uma função barreira para $C \subset I\!R^n$. A solução do problema*

$$\left\{\ \min_{x\in S(T,C)} h(x)\ ,\right. \tag{2.3}$$

única se existe pois h é estritamente convexa, é chamada centro analítico de $S(T, C)$ com relação a barreira h.

Na seção 2.3 é mostrado que se a interseção do domínio efetivo de h com $S(T, C)$ é não vazia, então $\lim\limits_{\mu\rightarrow+\infty} x(\mu)$ é solução do problema (2.3).

Para programação linear, a maioria das trajetórias centrais são da forma barreira logarítmica (Adler e Monteiro [1]), isto é, $h(x) = -\sum_{j=1}^{n} log(x_j), x = (x_1, ..., x_n) \in I\!R^n_{++}$, que em geral diverge nas soluções ótimas pois elas encontram-se em $\partial I\!R^n_{++}$. Para casos como este, na seção 2.4 é mostrado que $\lim\limits_{\mu\rightarrow+\infty} x(\mu)$ é solução do problema

$$\left\{\ \min_{x\in S(T,C)} \tilde{h}(x)\ ,\right.$$

onde $\tilde{h} : \partial I\!R^n_{++} \to I\!R$, é definida por $\tilde{h}(x) = \sum_{j \in J} h_j(x_j)$, com

$$J = \{j \in \{1, ..., n\} \, / z_j > 0 \, para \, algum \; z \in S(T, C)\} \, .$$

2.2 Hipóteses Adotadas

O objetivo principal deste capítulo é mostrar, com hipóteses adequadas, que a trajetória central $\{x(\mu)/\mu > 0\}$ converge, quando μ tende para $+\infty$, para uma solução do problema $VIP(T, C)$ e que $\lim\limits_{\mu \to +\infty} x(\mu)$ é o centro analítico do conjunto de soluções $S(T, C)$.

Nesta seção são enumeradas as várias hipóteses que devem satisfazer T, C e h para obter os resultados descritos anteriormente.

O conjunto $C \subset I\!R^n$ é considerado convexo e fechado, com $C^0 \neq \phi$.

Sobre a função barreira $h : I\!R^n \to I\!R \cup \{+\infty\}$ são feitos o primeiro grupo das hipóteses.

H1. h é estritamente convexa, fechada e contínua em seu domínio efetivo. Além disso, h é continuamente diferenciável em C^0.

H2. Se $\{x^k\}_{k \in I\!N}$ é uma seqüência contida em C^0 que converge para um ponto $\bar{x} \in \partial C$ e y é um ponto qualquer em C^0, então $\lim\limits_{k \to +\infty} \langle \nabla h(x^k), y - x^k \rangle = +\infty$.

Estas duas hipóteses são necessárias em quase todos os resultados obtidos. Uma função satisfazendo H2 é chamada *"coerciva na fronteira"*.

Alguns resultados obtidos precisam das seguintes hipóteses adicionais.

H3. h é finita, contínua e estritamente convexa em C.

Uma função satisfazendo H3 é chamada *"finita na fronteira"*.

H4. Para cada $y \in I\!R^n$, existe $x \in C^0$ tal que $\nabla h(x) = y$.

Uma função satisfazendo H4 é chamada *"zona coerciva"*.

A hipótese H3 junto com a finitude de h no conjunto de soluções de $VIP(T, C)$, simplificam a demonstração da convergência da trajetória central para o centro analítico.

Com a hipótese H4 e a monotonicidade de T é possível mostrar a existência da trajetória central.

Para o caso em que C é uma caixa generalizada; isto é, $C = [\alpha_1, \beta_1] \times, ..., \times [\alpha_n, \beta_n]$, com $\alpha_j \in [-\infty, +\infty)$, $\beta_j \in (-\infty, +\infty]$, $\alpha_j < \beta_j$, $j \in \{1, ..., n\}$, é considerada a seguinte hipótese.

H5. $h(x) = \sum_{j=1}^{n} h_j(x_j), \forall x \in I\!R^n$, com $h_j : (\alpha_j, \beta_j) \to I\!R$.

Uma função satisfazendo H5 é chamada *"separável"*. Esta hipótese é usada para mostrar a convergência da trajetória central para o centro analítico, em ausência de H3.

Para o caso $C = I\!R^n_+$, dois exemplos de funções barreiras são $h(x) = -\sum_{j=1}^{n} log(x_j)$, satisfazendo H1, H2 e H5, e $h(x) = \sum_{j=1}^{n} x_j log(x_j)$, satisfazendo H1, H2, H3, H4 e H5, com a convenção que $0log0 = 0$

O segundo grupo das hipóteses é feito sobre o operador $T : I\!R^n \to \mathcal{P}(I\!R^n)$. As primeiras duas delas são necessárias quase sempre.

H6. T é monótono maximal (veja Definição 1.1.2).

H7. T é paramonótono (veja Definição 1.2.2).

A hipótese seguinte é necessária para mostrar a existência da trajetória central em ausência de H4.

São necessárias algumas definições e observações preliminares.

Definição 2.2.1 *Seja $V \subset I\!R^n$ um conjunto não vazio, convexo e fechado. A função $\delta_V : I\!R^n \to I\!R \cup \{+\infty\}$, definida por*

$$\delta_V(x) = \begin{cases} 0, & se \quad x \in V \\ +\infty, & em \;\; outro \;\; caso \end{cases}, \tag{2.4}$$

é chamada a função indicadora do conjunto V.

Observação 2.2.1 *Claramente, δ_V é uma função convexa e fechada.*

Observação 2.2.2 *Sejam $f : I\!R^n \to I\!R \cup \{+\infty\}$ uma função convexa e $C \subset I\!R^n$ um conjunto não vazio e convexo. Se $dom(f) \cap C \neq \phi$ então $\varphi : I\!R^n \to I\!R \cup \{+\infty\}$, definida por:*

$$\varphi(x) = \begin{cases} f(x), & se \quad x \in C \\ +\infty, & em \;\; outro \;\; caso \end{cases},$$

é também uma função convexa. Neste caso, $\varphi = f + \delta_C$.

Definição 2.2.2 *O operador normalizado do conjunto $V \subset I\!R^n$, denotado por N_V, é definido por $N_V(x) = \partial\delta_V(x)$, para todo $x \in I\!R^n$.*

Observação 2.2.3 *Tem-se que,* $N_V(x) = \phi$ *se* $x \notin V$*, e* $N_V(x) = \{0\}$ *se* $x \in V^0$*.*

Observação 2.2.4 *Claramente,* $N_V(x)$ *é um cone conexo, para todo* x*.*

Observação 2.2.5 *Dado* $x \in V$*,* $v \in N_V(x)$ *se, e somente se,* $\langle v, x' - x\rangle \leq 0$*, para todo* $x' \in V$*.*

De fato. Se $x \in V$ e $v \in N_V(x)$ então $\langle v, x' - x\rangle \leq \delta_V(x') - \delta_V(x) = 0$, para todo $x' \in V$. Reciprocamente, se $\langle v, x' - x\rangle \leq 0$, para todo $x' \in V$, então $\langle v, x' - x\rangle \leq \delta_V(x') - \delta_V(x)$. Logo, $v \in N_V(x)$. ■

A hipótese é a seguinte.

H8. $T = \widehat{T} + N_V$, para algum operador contínuo e paramonótono $\widehat{T} : I\!R^n \to I\!R^n$ e algum conjunto $V \subset I\!R^n$ não vazio, convexo e fechado.

A importância de H8 é que o conjunto de soluções do problema $VIP(T, C)$ é o conjunto de zeros do operador $T + N_C$. De forma mais específica

Proposição 2.2.1 *Sejam* $T : I\!R^n \to \mathcal{P}(I\!R^n)$ *operador monótono maximal,* C *e* V *conjuntos convexos e fechados em* $I\!R^n$*. Então* $S(T, V \cap C) = S(T + N_V, C)$*.*

Demonstração. Suponha que $x \in S(T, V \cap C)$. Então existe $u \in T(x)$ tal que $\langle u, x' - x\rangle \geq 0$, para todo $x' \in V \cap C$. Daí, $\langle v, x' - x\rangle \leq 0$, para todo $x' \in V \cap C$, onde $v = -u$. Pela Observação 2.2.5, $v \in N_V(x)$. Assim, $0 = u + v$, com $u \in T(x)$ e $v \in N_V(x)$. Logo, $0 \in T(x) + N_V(x) = (T + N_V)(x)$.

Reciprocamente, se $0 \in (T + N_V)(x) = T(x) + N_V(x)$ então existem $u \in T(x)$ e $v \in N_V(x)$ tais que $0 = u + v$. Como $v \in N_V(x)$ então $\langle v, x' - x\rangle \leq 0$, para todo $x' \in V \cap C$ (veja Observação 2.2.5). Segue que, $\langle -u, x' - x\rangle \leq 0$, para todo $x' \in V \cap C$. Assim, $\langle u, x' - x\rangle \geq 0$, para todo $x' \in V \cap C$. Logo, $x \in S(T, V \cap C)$. ■

Um fato importante para programação linear é dado pela seguinte proposição.

Proposição 2.2.2 *Sejam* $A \in I\!R^{m \times n}$*,* $b \in I\!R^m$*. Se* $C = I\!R^n_+$ *e* $V = \{x \in I\!R^n / Ax = b\}$*, então*

$$N_V(x) = \begin{cases} Im(A^t), & se \quad x \in V \\ \phi, & em \;\; outro \;\; caso. \end{cases}$$

Demonstração. Da Observação 2.2.3, se $x \notin V$ então $N_V(x) = \phi$. Resta mostrar que $N_V(x) = Im(A^t)$ se $x \in V$.

De fato, seja $w \in Im(A^t)$. Então existe $v \in I\!R^m$ tal que $w = A^t v$. Considere $x \in V$ e $x' \in V$, isto é, $Ax = b$ e $Ax' = b$. Logo,

$$\langle w, x' - x \rangle = \langle A^t v, x' - x \rangle = v^t A x' - v^t A x = v^t b - v^t b = 0.$$

Pela Observação 2.2.5, $w \in N_V(x)$. Logo, $Im(A^t) \subset N_V(x)$, para todo $x \in V$.

Sejam agora, $x \in V$ e $w \in N_V(x)$. Pela Observação 2.2.5, $\langle w, x' - x \rangle \leq 0$, para todo $x' \in V$. Observe que $A(x - x') = Ax - Ax' = b - b = 0$. Assim, $x - x' \in Ker(A)$.

Se $\langle w, x' - x \rangle = 0$ então $w \in Ker(A)^{\perp} = Im(A^t)$. Se $\langle w, x' - x \rangle < 0$ para algum $x' \in V$, então $\langle w, x' - x \rangle + \langle w, x - x' \rangle = 0$. Logo, $0 = \langle w, x' - x + x - x' \rangle = \langle w, 0 \rangle$. Daí, $w \in Ker(A)^{\perp} = Im(A^t)$. Logo, $N_V(x) \subset Im(A^t)$, para todo $x \in V$.

Portanto $N_V(x) = Im(A^t)$, para todo $x \in V$. ■

A seguinte hipótese é necessária para mostrar a convergência da trajetória central para o centro analítico quando H3 não é satisfeita.

H9. Sejam $\widehat{T}$ e V como na hipótese H7. V é uma variedade afim, $\widehat{T}$ é continuamente diferenciável e existe um subespaço W tal que $Ker\left(J_{\widehat{T}}(x)\right) = W$, para todo $x \in C \cap V$.

Finalmente, o terceiro grupo de condições envolvem T e C (isto é, $VIP(T, C)$) e em alguns casos também h.

Primeiro consideramos duas hipóteses elementares sobre $VIP(T, C)$.

H10. $dom(T) \cap C^0 \neq \phi$.

O problema $VIP(T, C)$ satisfazendo H10 é chamado *"regular"*.

H11. $S(T, C) \neq \phi$.

Observação 2.2.6 *Suponha que* H8 *é satisteita. Então* $dom(T) = dom(\widehat{T}) \cap dom(N_V) = I\!R^n \cap V = V$. *Assim, neste caso* H10 *é equivalente a* $V \cap C^0 \neq \phi$.

Para garantir a existência da trajetória central, em ausência de H4, é preciso que sejam satisfeitas as seguintes duas hipóteses.

H12. h atinge seu mínimo em $dom(T) \cap C^0$ em algum $\tilde{x}$.

A hipótese seguinte requer da função gap.

Definição 2.2.3 *A função gap para o problema $VIP(T, C)$ é definida por*

$$\begin{array}{rcl} Gap_{\,T,C} : dom(T) \cap C & \to & I\!R \cup \{+\infty\} \\ x & \mapsto & Gap_{\,T,C}(x) := sup\,\{\langle v, x - y\rangle / (v, y) \in G_C T\}\,, \end{array} \tag{2.5}$$

onde $G_C T = \{(v, y)/v \in T(y), y \in C\}$.

Proposição 2.2.3 *Se $T : I\!R^n \to \mathcal{P}(I\!R^n)$ é monótono maximal e $C \subset I\!R^n$ é não vazio, convexo e fechado, então*

i) *$Gap_{\,T,C}$ é convexa em $dom(T) \cap C$.*

ii) *$Gap_{\,T,C}(x) \geq 0, \forall x \in dom(T) \cap C$.*

Demonstração.

i) Para $x \in dom(T) \cap C$ tem-se que $Gap_{\,T,C}(x)$ é o supremo de uma familia de transformações afins. Como as transformações afins são funções convexas e o supremo de funções convexas é uma função convexa (veja Proposição A.5.5), então $Gap_{\,T,C}$ é convexa em $dom(T) \cap C$.

ii) Basta tomar $y = x$ em (2.5). ■

A hipótese é a seguinte.

H13. $Gap_{\,T,C}(x) < +\infty, \forall x \in dom(T) \cap C$.

Uma relação importante entre a função gap com o conjunto de soluções $S(T, C)$ é dada pela seguinte proposição.

Proposição 2.2.4 *Suponha que* H10 *e* H11 *são satisfeitas. Então, $Gap_{\,T,C}(x) = 0$ se, e somente se, $x \in S(T, C)$*

Demonstração. Se $Gap_{\,T,C}(x) = 0$ então $\langle v, x - y\rangle \leq 0$, para todo $y \in C$ e todo $v \in T(y)$. Assim, $\langle v, y - x\rangle \geq 0$, para todo $y \in C$ e todo $v \in T(y)$. Seja $u \in T(x)$. Como T é monótono, $\langle u - v, x - y\rangle \geq 0$. Daí, $\langle u, x - y\rangle \geq \langle v, x - y\rangle$. Assim, $0 \leq \langle v, y - x\rangle \leq \langle u, y - x\rangle, \forall y \in C$. Logo, pela Proposição 1.3.1, $x \in S(T, C)$.

Reciprocamente, se $x \in S(T, C)$, então existe $u \in T(x)$ tal que $\langle u, y - x\rangle \geq 0$, para todo $y \in C$. Seja $v \in T(y)$. Como T é monótono, $\langle v - u, y - x\rangle \geq 0$. Daí, $\langle v, y - x\rangle \geq \langle u, y - x\rangle \geq 0$, para todo $y \in C$ e todo $v \in T(y)$. Logo, $\langle v, x - y\rangle \leq 0$, para todo $y \in C$ e todo $v \in T(y)$. Portanto, $Gap_{\,T,C}(x) = 0$. ■

A última hipótese é necessária para mostrar que a trajetória central é limitada quando H4 não é satisfeita.

H14. Uma das seguintes condições é satisfeita:

i) $S(T, C)$ é limitada e $T = \partial f$, para alguma função convexa $f : I\!R^n \to I\!R$.

ii) Existe $y \in C^0 \cap dom(T)$ tal que $\lim\limits_{k\to+\infty} \langle u^k, x^k - y\rangle = +\infty$, para toda seqüência $\left(x^k\right)_{k\in I\!N}$ satisfazendo $\lim\limits_{k\to+\infty} \|x^k\| = +\infty$, e toda seqüência $\left(u^k\right)_{k\in I\!N}$ com $u^k \in T(x^k)$, para todo $k \in I\!N$.

iii) $dom(T) \cap C$ é limitado.

Esta seção finaliza com alguns resultados sobre operadores monotónos que serão utilizados para mostrar a existência e unicidade da trajetória central.

Proposição 2.2.5 *Se h satisfaz* H1 *e, ou zona coerciva ou coerciva na fronteira (veja* H4 *e* H2*, respectivamente), então $dom(\partial h) = C^0$.*

Demonstração. Como h é estritamente convexa, $\partial h(x) \neq \phi$, para todo $x \in C^0$. Além disso, $\partial h(x) = \phi$, se $x \notin C$. Resta analizar em pontos da fronteira de C. Seja $x \in \partial C$ e suponha que existe $\xi \in \partial h(x)$.

Se h é zona coerciva, existe $z \in C^0$ tal que $\nabla h(z) = \xi$. Logo,

$$\langle \partial h(x) - \nabla h(z), x - z\rangle = \langle \xi - \xi, x - z\rangle = 0.$$

Como h é estritamente convexa então $x = z$, o que é um absurdo pois $z \in C^0$ e $x \in \partial C$. Logo, $\partial h(x) = \phi$, se $x \in \partial C$.

Se h é coerciva na fronteira, considere $\epsilon_k \in (0, 1)$ tal que $\lim\limits_{k\to+\infty} \epsilon_k = 0$. Seja $x^k = (1 - \epsilon_k)x + \epsilon_k y$, $k \in I\!N$, com $x \in \partial C$ e $z \in C^0$. Observe que $\lim\limits_{k\to+\infty} x^k = x \in \partial C$. Por convexidade de C, $x^k \in C$, $\forall k \in I\!N$. Logo,

$$\epsilon_k \langle \xi, y - x\rangle = \langle \xi, x^k - x\rangle \leq h(x^k) - h(x) < \langle \nabla h(x^k), x^k - x\rangle, \tag{2.6}$$

onde foi usada a convexidade estrita de h. Observe que,

$$
\begin{aligned}
x^k - x = x - \epsilon_k x + \epsilon_k y - x = \epsilon_k (y - x) &= \frac{\epsilon_k}{1-\epsilon_k}(1-\epsilon_k)(y-x) \\
&= \frac{\epsilon_k}{1-\epsilon_k}(1-\epsilon_k)(y - x - \epsilon_k y + \epsilon_k x) \\
&= \frac{\epsilon_k}{1-\epsilon_k}\left[y - ((1-\epsilon_k)x + \epsilon_k y)\right] = \frac{\epsilon_k}{1-\epsilon_k}(y - x^k).
\end{aligned}
$$

Logo, em (2.6), temos

$$\epsilon_k \langle \xi, y - x\rangle < \frac{\epsilon_k}{1-\epsilon_k}\langle \nabla h(x^k), y - x^k\rangle. \tag{2.7}$$

Cancelando ϵ_k e passando ao limite em (2.7), quando k tende para $+\infty$, temos

$$\langle \xi, y - x\rangle < \lim_{k\to+\infty} \frac{1}{1-\epsilon_k}\langle \nabla h(x^k), y - x^k\rangle = -\infty,$$

o que é um absurdo. Logo, $\partial h(x) = \phi$, se $x \in \partial C$.

Portanto, $dom(\partial h) = C^0$. ■

A demostração da seguinte proposição é muito elaborada e foge dos objetivos deste trabalho. Para uma prova do item (i), veja Teorema 1 em Rockafellar [16]; e para o item (ii), veja Lema 5 em Burachik [4].

Proposição 2.2.6 i) *Se T_1 e T_2 são operadores monótonos maximais e $dom(T_1) \cap dom(T_2)^0 \neq \phi$, então $T_1 + T_2$ é operador monótono maximal.*
ii) *Em adição, se T_2 é o subdiferencial de uma função convexa propria fechada e T_2 é sobrejetivo, então $T_1 + T_2$ é sobrejetivo.*

O seguinte resultado pode ser encontrado no Corolario 2.2 em Brézis [3].

Proposição 2.2.7 *Se T é operador monótono maximal e $dom(T)$ é limitado, então o operador T é sobrejetivo.*

2.3 Existência, Unicidade e Comportamento da Trajetória Central

Sejam $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono e $h : I\!R^n \to I\!R \cup \{+\infty\}$ uma função convexa. Para cada $\mu > 0$, considere $T_\mu = \mu T + \partial h$.

Lema 2.1 *Suponha que h satisfaz* H1*, T é monótono maximal (veja* H6*), $VIP(T,C)$ é regular (veja* H10*), e uma das seguintes condições é satisfeita:*
i) *h é zona coerciva (veja* H4*), ou*
ii) *h é coerciva na fronteira (veja* H2*) e atinge seu mínimo em $dom(T) \cap C^0$ (veja* H12*), a função gap é finita (veja* H13*), e T satisfaz* H8*.*
Então, o operador T_μ possui um único zero $x(\mu)$ que pertence a $dom(T) \cap C^0$.

Demonstração. Suponha que o caso (i) é válido. Sejam $T_1 = \mu T$ e $T_2 = \partial h$. Pela Proposição 2.2.5, $dom(T_2) = C^0$. Pela regularidade de $VIP(T,C)$, $dom(T_1) \cap dom(T_2)^0 \neq \phi$. Pelo item (i) da Proposição 2.2.6, $T_1 + T_2$ é monótono maximal. Como h é zona coerciva, T_2 é sobrejetivo. Pelo item (ii) da Proposição 2.2.6, T_μ é sobrejetivo. Logo, para cada $\mu > 0$, existe $x(\mu)$ tal que $0 \in T_\mu(x(\mu))$. Como h é estritamente convexo então T_2 é estritamente monótono. Logo, T_μ é estritamente monótono. Assim, $x(\mu)$ é o único zero de T_μ, para cada $\mu > 0$. Além disso, $dom(T_\mu) = dom(T_1) \cap dom(T_2) = dom(T) \cap C^0$.

Suponha que o caso (ii) é válido. Seja $\tilde{x}$ o mínimo de h em $dom(T) \cap C^0$. Como a função gap é finita em $dom(T) \cap C$, temos dois casos para analizar.

a) Quando $Gap_{\,T,C}(\tilde{x}) = 0$. Neste caso, pela Proposição 2.2.4, $\tilde{x} \in S(T,C)$. Então, pela Proposição 2.2.1, $\tilde{x}$ é zero de $T + N_C$; isto é, $0 \in T(\tilde{x}) + N_C(\tilde{x})$. Como $\tilde{x} \in C^0$ então $N_C(\tilde{x}) = \{0\}$. Logo, para $u \in T(\tilde{x})$, $0 = u + 0 = u$. Assim, $0 \in T(\tilde{x})$.

Pela hipótese H8, para algum $v \in N_V(\tilde{x})$, temos

$$0 = \widehat{T}(\tilde{x}) + v. \tag{2.8}$$

Por outro lado, como $\tilde{x}$ minimiza h em $dom(T)$ e $\tilde{x} \in C^0$, então $\tilde{x}$ é solução de $VIP(\partial h, dom(T))$. Logo, $0 \in (\partial h + N_{dom(T)})(\tilde{x})$. Observe que $\partial h(\tilde{x}) = \{\nabla h(\tilde{x})\}$ e $dom(T) = dom(\widehat{T}) \cap dom(N_V) = I\!R^n \cap V = V$. Assim, para algum $v' \in N_V(\tilde{x})$,

$$0 = \nabla h(\tilde{x}) + v'. \tag{2.9}$$

Seja $\bar{v} = \mu v + v'$. Como $N_V(\tilde{x})$ é um cone convexo, $\bar{v} \in N_V(\tilde{x})$. De (2.8), tem-se:

$$0 = \mu(\widehat{T}(\tilde{x}) + v). \tag{2.10}$$

Adicionando (2.9) e (2.10), temos $0 = \mu\widehat{T}(\tilde{x}) + \mu v + \nabla h(\tilde{x}) + v'$. Daí, $0 = \mu\widehat{T}(\tilde{x}) + \bar{v} + \nabla h(\tilde{x})$. Isto é, $0 \in (\mu T + \partial h)(\tilde{x}) = T_\mu(\tilde{x})$. Observe que $\tilde{x} \in dom(T) \cap C^0$.

Portanto, T_μ possui um único zero que pertence a $dom(T) \cap C^0$.

b) Quando $Gap_{T,C}(\tilde{x}) > 0$. Seja $W = \{ x \in C \ / \ h(x) \leq \mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x}) \}$.

Afirmamos que W é convexo.

De fato, sejam $x_1 \in W$, $x_2 \in W$ e $\alpha \in [0,1]$. Pela convexidade de h, temos

$$\begin{aligned} h(\alpha x_1 + (1-\alpha)x_2) &\leq \alpha h(x_1) + (1-\alpha)h(x_2) \\ &\leq \alpha\,[\mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x})] + (1-\alpha)\,[\mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x})] \\ &= \mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x}). \end{aligned}$$

Assim, $\alpha x_1 + (1-\alpha)x_2 \in W$. Logo, W é convexo.

Também, W é fechado.

De fato, seja $x \in \mathrm{cl}(W)$. Então, existe $\{x_n\}_{n\in I\!N} \subset W$ tal que $\lim\limits_{n\to+\infty} x_n = x$. Como $x_n \in W$, $\forall n \in I\!N$, então $h(x_n) \leq \mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x})$, $\forall n \in I\!N$. Como h é contínua, $h(x) = h(\lim\limits_{n\to+\infty} x_n) = \lim\limits_{n\to+\infty} h(x_n) \leq \mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x})$. Assim, $x \in W$. Logo, W é fechado.

Observe que $W^0 = \{x \in C / \ h(x) < \mu\,(Gap_{T,C}(\tilde{x})) + h(\tilde{x})\}$.

Afirmamos que $V \cap W$ é limitado.

Como $T = \widehat{T} + N_V$, então $dom(T) = I\!R^n \cap V = V$. Logo, $V \cap W$ é a interseção de um conjunto de nível de h com $dom(T) \cap C$, isto é, um conjunto de nível de $\widehat{h}$, definido por

$$\widehat{h}(x) = \begin{cases} h(x), & \text{se} \quad x \in dom(T) \cap C \\ +\infty, & \text{em outro caso} \end{cases} .$$

Pela hipótese H12, $\widehat{h}$ atinge seu mínimo. Seja $z \in dom(T) \cap C$ o mínimo de $\widehat{h}$ então $h(z) = \widehat{h}(z) \leq \widehat{h}(x) = h(x)$, $\forall x \in dom(T) \cap C$. Como $\tilde{x}$ é mínimo de h em $dom(T) \cap C$ e h é estritamente convexa, então $z = \tilde{x}$. Assim, $\tilde{x} \in dom(T) \cap C$ é único. Logo o conjunto de nível, $\left\{x \in C \ / \ \widehat{h}(x) \leq \widehat{h}(\tilde{x})\right\} = \{\tilde{x}\}$ é limitado. Como $\widehat{h}$ possui um conjunto de nível limitado, então todos seus conjuntos de níveis são limitados (veja Proposição A.5.4). Em particular $V \cap W$ é limitado.

Seja $U = T_\mu + N_W$. Como foi mostrado no item (i), T_μ é monótono maximal. Observe que $\tilde{x} \in dom(T_\mu) = dom(T_1) \cap dom(T_2) = dom(T) \cap dom(\partial h) = dom(T) \cap C^0$. Como $Gap_{T,C}(\tilde{x}) > 0$ então $\tilde{x} \in W^0 = dom(N_W)^0$. Assim, $dom(T_\mu) \cap dom(N_W)^0 \neq \phi$. Pelo item (i) da Proposição 2.2.6, U é monótono maximal.

Observe que

$$\begin{aligned} dom(U) &= dom(T_\mu) \cap dom(N_W) = dom(T_1) \cap dom(T_2) \cap dom(N_W) \\ &= dom(T) \cap dom(\partial h) \cap dom(N_W) = V \cap C^0 \cap W \subset V \cap W. \end{aligned}$$

Como $V \cap W$ é limitado então $dom(U)$ é limitado. Da Proposição 2.2.7, segue que o operador U é sobrejetivo.

Assim, existe $x \in dom(U)$ tal que

$$\begin{aligned} 0 \in U(x) &= T_\mu(x) + N_W(x) = \mu T(x) + \partial h(x) + N_W(x) \\ &= \mu \widehat{T}(x) + \mu N_V(x) + \partial h(x) + N_W(x). \end{aligned}$$

Logo, existem $v \in N_V(x)$ e $w \in N_W(x)$ tais que

$$0 = \mu \widehat{T}(x) + \mu v + w + \nabla h(x). \tag{2.11}$$

Resta mostrar que $x \in W^0$.

De fato, se $x = \tilde{x}$, claramente $x \in W^0$, pois $Gap_{\,T,C}(\tilde{x}) > 0$. Caso contrário, multiplicando a equação (2.11) por $x - \tilde{x}$, temos

$$0 = \mu\langle \widehat{T}(x), x - \tilde{x}\rangle + \mu\langle v, x - \tilde{x}\rangle + \langle w, x - \tilde{x}\rangle + \langle \nabla h(x), x - \tilde{x}\rangle.$$

Então,

$$\langle \nabla h(x), x - \tilde{x}\rangle = \mu\langle \widehat{T}(x), \tilde{x} - x\rangle + \mu\langle v, \tilde{x} - x\rangle + \langle w, \tilde{x} - x\rangle. \tag{2.12}$$

Como h é estritamente convexa, $\langle \nabla h(x), \tilde{x} - x\rangle < h(\tilde{x}) - h(x)$. Daí, $h(x) - h(\tilde{x}) < \langle \nabla h(x), x - \tilde{x}\rangle$. Logo, em (2.12),

$$h(x) - h(\tilde{x}) < \mu\langle \widehat{T}(x), \tilde{x} - x\rangle + \mu\langle v, \tilde{x} - x\rangle + \langle w, \tilde{x} - x\rangle. \tag{2.13}$$

Observe que $x \in V \cap W$, $\tilde{x} \in V \cap W$,$v \in N_V(x)$ e $w \in N_W(x)$. Então, pela Observação 2.2.5, $\langle v, \tilde{x} - x\rangle \leq 0$ e $\langle w, \tilde{x} - x\rangle \leq 0$. Assim, da equação (2.13),

$$h(x) - h(\tilde{x}) < \mu\langle \widehat{T}(x), \tilde{x} - x\rangle \leq 0 < \mu\left(Gap_{\,T,C}(\tilde{x})\right).$$

Segue que $h(x) < h(\tilde{x}) + \mu\left(Gap_{\,T,C}(\tilde{x})\right)$. Logo, $x \in W^0$.

Como $x \in W^0$, então $N_W(x) = \{0\}$. Daí, $w = 0$. Assim, $0 = \mu\widehat{T}(x) + \mu v + \nabla h(x)$, com $v \in N_V(x)$. Logo, $0 \in \mu T(x) + \partial h(x) = T_\mu(x)$. A unicidade de x sigue do item (i).

Portanto, T_μ possui um único zero que pertence a $dom(T) \cap C^0$. ■

Observação 2.3.1 *Para cada $\mu > 0$, o único zero de $T_\mu = \mu T + \partial h$ é um ponto da trajetória central $\{x(\mu)/\mu > 0\}$.*

De fato. Pelo Lema 2.1, $x(\mu)$ é o único zero de $T_\mu = \mu T + \partial h$, para cada $\mu > 0$. Entao, $0 \in T_\mu(x(\mu)) = \mu T(x(\mu)) + \partial h(x(\mu))$. Observe que $\partial h(x(\mu)) = \{\nabla h(x(\mu))\}$, pois h é diferenciável. Logo, existe $v \in (T(x(\mu)))$ tal que $0 = \mu v + \nabla h(x(\mu))$. Assim, $-\frac{1}{\mu}\nabla h(x(\mu)) = v \in T(x(\mu))$, como em (2.1). ∎

Observação 2.3.2 *O Lema 2.1 estabelece que a trajetória central $\{x(\mu)/\,\mu > 0)\}$ está bem definida e contida em C^0.*

Definição 2.3.1 *Um ponto de acumulação de $\{x(\mu)/\,\mu > 0)\}$ é um ponto $\bar{x}$ tal que $\bar{x} = \lim_{k\to+\infty} x(\mu_k)$, para alguma seqüência $\{\mu_k\}_{k\in\mathbb{N}}$ tal que $\lim_{k\to+\infty} \mu_k = +\infty$.*

Para mostrar que a trajetória central $\{x(\mu)/\,\mu > 0)\}$ possui pontos de acumulação é preciso o seguinte resultado.

Lema 2.2 *Com as hipóteses do Lema 2.1, tem-se:*
i) *$h(x(\mu))$ é não decrescente em μ.*
ii) *$x(\mu)$ é contínua em qualquer $\mu > 0$.*

Demonstração. Sejam $\mu_1 > 0$ e $\mu_2 > 0$. Defina $x^1 = x(\mu_1)$ e $x^2 = x(\mu_2)$. Por definição de $x(\mu)$, $0 \in T_\mu(x(\mu))$; equivalentemente $-\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu))$. Sejam $u^1 = -\frac{1}{\mu_1}\nabla h(x^1)$ e $u^2 = -\frac{1}{\mu_2}\nabla h(x^2)$. Tem-se que $u^1 \in T(x^1)$ e $u^2 \in T(x^2)$.

Pela convexidade de h,

$$\langle \nabla h(x^1), x^2 - x^1\rangle \le h(x^2) - h(x^1)\,,\ \ \langle \nabla h(x^2), x^1 - x^2\rangle \le h(x^1) - h(x^2).$$

Daí,

$$\frac{1}{\mu_1}\big(h(x^1) - h(x^2)\big) \le \frac{1}{\mu_1}\langle \nabla h(x^1), x^1 - x^2\rangle = \langle u^1, x^2 - x^1\rangle, \tag{2.14}$$

$$\frac{1}{\mu_2}\big(h(x^2) - h(x^1)\big) \le \frac{1}{\mu_2}\langle \nabla h(x^2), x^2 - x^1\rangle = \langle u^2, x^1 - x^2\rangle. \tag{2.15}$$

Adicionando (2.14) e (2.15):

$$\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)h(x^1) - \left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)h(x^2) \le \langle u^1, x^2 - x^1\rangle + \langle u^2, x^1 - x^2\rangle.$$

Como T é monótono,

$$\left(\frac{1}{\mu_1} - \frac{1}{\mu_2}\right)(h(x^1) - h(x^2)) \le \langle u^1 - u^2, x^2 - x^1\rangle \le 0. \tag{2.16}$$

i) Suponha que $\mu_1 > \mu_2$. Então, $\left(\dfrac{1}{\mu_1} - \dfrac{1}{\mu_2}\right) < 0$. Logo, de (2.16), temos $h(x^1) - h(x^2) \geq 0$. Asism,

$$h(x(\mu_1)) = h(x^1) \geq h(x^2) = h(x(\mu_2)). \tag{2.17}$$

Portanto, $h(x(\mu))$ é não decrescente em μ.

ii) De (2.16) segue que $\langle u^1 - u^2, x^1 - x^2\rangle \geq 0$. Daí, $\langle u^1, x^1 - x^2\rangle \geq \langle u^2, x^1 - x^2\rangle$. Assim, $-\dfrac{1}{\mu_1}\langle \nabla h(x^1), x^1 - x^2\rangle \geq -\dfrac{1}{\mu_2}\langle \nabla h(x^2), x^1 - x^2\rangle$. Logo,

$$\frac{1}{\mu_1}\langle \nabla h(x^1), x^1 - x^2\rangle \leq \frac{1}{\mu_2}\langle \nabla h(x^2), x^1 - x^2\rangle.$$

De (2.14) e (2.17),

$$0 \leq \frac{1}{\mu_1}\big[h(x^1) - h(x^2)\big] \leq \frac{1}{\mu_1}\langle \nabla h(x^1), x^1 - x^2\rangle \leq \frac{1}{\mu_2}\langle \nabla h(x^2), x^1 - x^2\rangle. \tag{2.18}$$

Seja $\bar{\mu} > 0$ fixo e considere $\hat{\mu}$ e $\tilde{\mu}$ tais que $\tilde{\mu} > \bar{\mu} > \hat{\mu}$. Como $h(x(\mu))$ é não decrescente, para $\mu \in (\hat{\mu}, \tilde{\mu})$, tem-se:

$$h(x(\mu)) \leq h(\hat{\mu}). \tag{2.19}$$

Usando (2.18) com $\mu_1 = \mu$ e $\mu_2 = \hat{\mu}$, tem-se:

$$0 \leq \langle \nabla h(x(\hat{\mu})), x(\mu) - \hat{\mu}\rangle. \tag{2.20}$$

Sejam $L_1 = \{x \in C / \, h(x) \leq h(x(\tilde{\mu}))\}$, $L_2 = \{x \in C / \, \langle \nabla h(x(\hat{\mu})), x - x(\hat{\mu})\rangle \geq 0\}$ e $L = L_1 \cap L_2 = \{x \in C / \, h(x) \leq h(x(\tilde{\mu})) + \langle \nabla h(x(\hat{\mu})), x - x(\hat{\mu})\rangle\}$. De (2.19) e (2.20), sigue que $\{x(\mu) / \, \hat{\mu} \leq \mu \leq \tilde{\mu}\} \subset L$.

Tem-se que L é limitado. De fato. Defina $\bar{h}$ como:

$$\bar{h}(x) = \begin{cases} h(x), & \text{se} \quad x \in L_2 \\ +\infty, & \text{em outro caso} \end{cases}.$$

Para $x \in L_2$, como h é estritamente convexa, tem-se

$$0 \leq \langle \nabla h(x(\hat{\mu})), x - x(\hat{\mu})\rangle < h(x) - h(x(\hat{\mu})).$$

Daí, $h(x(\hat{\mu})) < h(x)$, para todo $x \in L_2$. Logo, $x(\hat{\mu})$ é o único minimizador de $\bar{h}$. Assim, o conjunto de nível $\big\{x/\bar{h}(x) \leq \bar{h}(\hat{x})\big\} = \{x(\hat{x})\}$ de $\bar{h}$, é limitado. Isto implica que todos os conjuntos de nível de $\bar{h}$ são limitados (veja Proposição A.5.4). Logo, L pode ser escrito como $L = \big\{x \in I\!R^n / \bar{h}(x) \leq \bar{h}(\hat{x})\big\} = \{x(\hat{x})\}$. Assim, L é limitado.

Como $\{x(\mu)/ \, \hat{\mu} \leq \mu \leq \tilde{\mu}\} \subset L$, então $\{x(\mu)/ \, \hat{\mu} \leq \mu \leq \tilde{\mu}\}$ é limitado.

Para mostrar que $x(\mu)$ é contínua em $\bar{\mu}$ é suficiente mostrar que $\lim\limits_{k\to+\infty} x(\mu_k) = x(\bar{\mu})$, para alguma seqüência $\{\mu_k\}_{k\in I\!N}$ tal que $\lim\limits_{k\to+\infty} \mu_k = \bar{\mu}$.

Seja $\{\mu_k\}_{k\in I\!N}$ uma seqüência tal que $\lim\limits_{k\to+\infty} \mu_k = \bar{\mu}$, com $\mu_k \in (\hat{\mu}, \tilde{\mu})$ para todo k suficientemente grande. Assim, $\{x(\mu_k)\}_{k\in I\!N}$ é limitada, para todo k suficientemente grande.

Seja y um ponto de acumulação de $\{x(\mu_k)\}_{k\in I\!N}$. Logo, existe uma subseqüência $\{\mu_k\}_{k\in A}$, com A infinito, $A \subset I\!N$, tal que $\lim\limits_{k\in A} x(\mu_k) = y$. Denotando $\bar{x} = x(\bar{\mu})$ e $x^k = x(\mu_k)$ temos que $y = \lim\limits_{k\in A} x^k$.

Seja $K_1 = \{ \, k \in A \,/ \, \mu_k \leq \bar{\mu} \, \}$. Usando (2.18) com $\mu_1 = \bar{\mu}$ e $\mu_2 = \mu_k$, temos

$$\frac{1}{\bar{\mu}}\big[h(\bar{x})-h(x^k)\big] \leq \frac{1}{\bar{\mu}}\langle \nabla h(\bar{x}), \bar{x}-x^k\rangle \leq \frac{1}{\mu_k}\langle \nabla h(x^k), \bar{x}-x^k\rangle \leq \frac{1}{\mu_k}\big[h(\bar{x})-h(x^k)\big], \quad (2.21)$$

onde a convexidade de h foi utilizada na última desigualdade.

Observe que $(\mu_k)_{k\in A} \subset L_1 \subset ED(h)$. Pela hipótese H1, h é contínua em $y = \lim\limits_{k\in A} x^k$. Logo, se K_1 é infinito, tomando limite em (2.21) quando k tende para $+\infty$, com $k \in K_1$, tem-se:

$$\frac{1}{\bar{\mu}}\big[h(\bar{x}) - h(y)\big] \leq \frac{1}{\bar{\mu}}\langle \nabla h(y), \bar{x} - y\rangle \leq \frac{1}{\bar{\mu}}\big[h(\bar{x}) - h(y)\big]. \quad (2.22)$$

Segue que $h(\bar{x}) - h(y) = \langle \nabla h(y), \bar{x} - y\rangle$. Como h é estritamente convexa, então $\bar{x} = y$.

Se K_1 é finito ou vazio, considere $K_2 = A - K_1$. Seja $k \in K_2$. Usando (2.18) com $\mu_1 = \mu_k$ e $\mu_2 = \bar{\mu}$, tem-se:

$$\frac{1}{\mu_k}\big[h(x^k)-h(\bar{x})\big] \leq \frac{1}{\mu_k}\langle \nabla h(x^k), x^k-\bar{x}\rangle \leq \frac{1}{\bar{\mu}}\langle \nabla h(\bar{x}), x^k-\bar{x}\rangle \leq \frac{1}{\bar{\mu}}\big[h(x^k)-h(\bar{x})\big]. \quad (2.23)$$

Tem-se que K_2 é infinito, pois K_1 é finito. Logo, tomando limite em (2.23) quando k tende para $+\infty$, com $k \in K_2$, tem-se:

$$\frac{1}{\bar{\mu}}\big[h(y) - h(\bar{x})\big] \leq \frac{1}{\bar{\mu}}\langle \nabla h(\bar{x}), y - \bar{x}\rangle \leq \frac{1}{\bar{\mu}}\big[h(y) - h(\bar{x})\big]. \quad (2.24)$$

Segue que $h(y) - h(\bar{x}) = \langle \nabla h(\bar{x}), y - \bar{x}\rangle$. Como h é estritamente convexa, então $y = \bar{x}$.

Logo, $\lim_{k\in A} x(\mu_k) = \bar{x} = x(\bar{\mu})$, onde $(\mu_k)_{k\in A}$ é seqüência tal que $\lim_{k\in A} \mu_k = \bar{\mu}$. Redefinindo a seqüência como sendo a mesma subseqüência, tem-se que $\lim_{k\to\infty} x(\mu_k) = x(\bar{\mu})$, onde $\{\mu_k\}_{k\in I\!N}$ é seqüência tal que $\lim_{k\to\infty} \mu_k = \bar{\mu}$.

Portanto, $x(\mu)$ é contínua em qualquer $\mu > 0$. ∎

Lema 2.3 *Suponha que T satisfaz* H8, *que $VIP(T,C)$ possui soluções (veja* H11*), que h atinge seu mínimo em $dom(T)\cap C^0$ (veja* H12*) e que todas as hipóteses do Lema 2.1 são válidas. Se*
i) *h é finita na fronteira, ou*
ii) H14 *é válido,*
então, $\{x(\mu)/\mu > \bar{\mu}\}$ possui pontos de acumulação, para todo $\bar{\mu} > 0$.

Demonstração. Pela definição de $x(\mu)$ e pela hipótese H8,

$$-\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu)) = \widehat{T}(x(\mu)) + N_V(x(\mu)).$$

Pela Observação 2.2.6, $dom(T) = V$. Logo, existe $v \in N_V(x(\mu))$ satisfazendo $-\frac{1}{\mu}\nabla h(x(\mu)) = \widehat{T}(x(\mu)) + v$. Então, para $y \in V$, tem-se:

$$\langle -\frac{1}{\mu}\nabla h(x(\mu)), y - x(\mu)\rangle = \langle \widehat{T}(x(\mu)), y - x(\mu)\rangle + \langle v, y - x(\mu)\rangle.$$

Como $v \in N_V(x(\mu))$, $\langle v, y - x(\mu)\rangle \leq 0$. Logo, como T é monótono, segue que

$$\frac{1}{\mu}\langle \nabla h(x(\mu)), x(\mu) - y\rangle \leq \langle \widehat{T}(x(\mu)), y - x(\mu)\rangle \leq \langle \widehat{T}(y), y - x(\mu)\rangle. \qquad (2.25)$$

Suponha que o caso (i) é válido. Pela hipótese H11, $S(T,C) \neq \phi$. Seja $z \in S(T,C)$. Daí, $z \in C$. Como h é finita na fronteira (veja H3), então $z \in ED(h)$. Tomando $y = z$ em (2.25), segue que $\frac{1}{\mu}\langle \nabla h(x(\mu)), x(\mu) - z\rangle \leq \langle \widehat{T}(z), z - x(\mu)\rangle$. Pela hipótese H1, $\langle \nabla h(x(\mu)), z - x(\mu)\rangle \leq h(z) - h(x(\mu))$. Logo,

$$\frac{1}{\mu}\left[h(x(\mu) - h(z)\right] \leq \frac{1}{\mu}\langle \nabla h(x(\mu)), x(\mu) - z\rangle \leq \langle \widehat{T}(z), z - x(\mu)\rangle. \qquad (2.26)$$

Como $z \in S(T,C)$, existe $u = \widehat{T}(z) + v$ tal que $\langle u, x(\mu) - z\rangle \geq 0$, onde $v \in N_V(z)$. Logo, $\langle \widehat{T}(z), x(\mu) - z\rangle + \langle v, x(\mu) - z\rangle \geq 0$. Assim, $\langle \widehat{T}(z), z - x(\mu)\rangle \leq \langle v, x(\mu) - z\rangle$. Como $z \in V$ e $v \in N_V(z)$, então $\langle v, x(\mu) - z\rangle \leq 0$. Segue que $\langle \widehat{T}(z), z - x(\mu)\rangle \leq 0$. Assim, de (2.26), $h(x(\mu)) \leq h(z)$, $\forall \mu > 0$. Logo, $\{x(\mu)/\mu > 0\}$ está contida na interseção de V com um conjunto de nível de h. Como na demonstração do Lema 2.1, H12 implica que a interseção de V com os conjuntos de nível de h são limitados. Assim, $\{x(\mu)/\mu > 0\}$ é limitado. Portanto, $\{x(\mu)/\mu > 0\}$ possui pontos de acumulação.

Suponha que o caso (ii) é válido. Se h não é finita na fronteira pode acontecer que $h(z) = +\infty$, para algum $z \in S(T, C)$, e o argumento anterior não é válido.

Suponha que o item (i) da hipótese H14 é válido. Isto é, $S(T, C)$ é limitada e $T = \partial f$, para alguma função convexa $f : I\!R^n \to I\!R$. Logo, $-\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu)) = \partial f(x(\mu))$. Fixe $\bar{\mu} > 0$ e considere $\mu \geq \bar{\mu}$. Pela convexidade de h,

$$\langle \nabla h(x(\mu)), x(\bar{\mu}) - x(\mu)\rangle \leq h(x(\bar{\mu})) - h(x(\mu)).$$

Então, $\dfrac{1}{\mu}\left[h(x(\mu)) - h(x(\bar{\mu}))\right] \leq \langle -\frac{1}{\mu}\nabla h(x(\mu)), x(\bar{\mu}) - x(\mu)\rangle$. Como $-\frac{1}{\mu}\nabla h(x(\mu)) \in \partial f(x)$ então $\langle -\frac{1}{\mu}\nabla h(x(\mu)), x(\bar{\mu}) - x(\mu)\rangle \leq f(x(\bar{\mu})) - f(x(\mu))$. Logo,

$$\frac{1}{\mu}\left[h(x(\mu)) - h(x(\bar{\mu}))\right] \leq \frac{1}{\mu}\langle \nabla h(x(\mu)), x(\mu) - x(\bar{\mu})\rangle \leq f(x(\bar{\mu})) - f(x(\mu)). \qquad (2.27)$$

Pelo item (i) do Lema 2.2, $h(x(\mu)) \geq h(x(\bar{\mu}))$. Daí, $\frac{1}{\mu}\left[h(x(\mu)) - h(x(\bar{\mu}))\right] \geq 0$. Assim, de (2.27) segue que $f(x(\mu)) \leq f(x(\bar{\mu}))$. Logo, $\{x(\mu)/\mu > \bar{\mu}\}$ está contida num conjunto de nível de f. Observe que o problema $VIP(T, C)$, com $T = \partial f$, é equivalente ao problema

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

e o conjunto de soluções $S(T, C)$ deste problema, que são conjuntos de nível da restrição de f a C, é limitado.. Segue que todos os conjuntos de nível de f são limitados (veja Proposição A.5.4). Assim, $\{x(\mu)/\mu > \bar{\mu}\}$ é limitado. Logo, $\{x(\mu)/\mu > \bar{\mu}\}$ possui pontos de acumulação, para todo $\bar{\mu} > 0$.

Suponha que o item (ii) da hipótese H14 é válido. Fixe $\bar{\mu}$. Seja $u = -\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu))$. Considere $\mu \geq \bar{\mu}$. Pela hipótese, existe $y \in C^0 \cap dom(T)$ tal que

$$\lim_{k \to +\infty} \langle u^k, x^k - y\rangle = +\infty,$$

para toda seqüência $\left\{x^k\right\}_{k \in I\!N}$ satisfazendo $\lim_{k \to +\infty} \parallel x^k \parallel = +\infty$, e toda seqüência $\left\{u^k\right\}_{k \in I\!N} \subset T(x^k)$. Pela convexidade de h, $\langle \nabla h(x(\mu)), y - x(\mu)\rangle \leq h(y) - h(x(\mu))$. Então,

$$-\frac{1}{\mu}\left[h(x(\mu)) - h(y)\right] \leq \langle -\frac{1}{\mu}\nabla h(x(\mu)), y - x(\mu)\rangle = \langle u, y - x(\mu)\rangle.$$

Daí, $\langle u, y - x(\mu)\rangle \leq \frac{1}{\mu}[h(y) - h(x(\mu))] \leq \frac{1}{\mu}[h(y) - h(x(\bar{\mu}))]$, pois pelo item (i) do Lema 2.2, $h(x(\mu)) \geq h(x(\bar{\mu}))$. Logo,

$$\begin{aligned} \langle u, y - x(\mu)\rangle &\leq \frac{1}{\mu}[h(y) - h(x(\bar{\mu}))] \\ &\leq \frac{1}{\mu} \mid h(y) - h(x(\bar{\mu})) \mid \leq \frac{1}{\bar{\mu}} \mid h(y) - h(x(\bar{\mu})) \mid . \end{aligned} \tag{2.28}$$

Seja $\theta = \frac{1}{\bar{\mu}} \mid h(y) - h(x(\bar{\mu})) \mid$. Daí, $\langle u, y - x(\mu)\rangle \leq \theta$. Se $\{x(\mu)/\mu > \bar{\mu}\}$ não é limitada, existe uma seqüência $\{x^k\}_{k\in I\!N}$ tal que $\lim\limits_{k\to+\infty} \| x^k \| = +\infty$ e $\langle u^k, x(\mu_k) - y\rangle \leq \theta$, $\forall k \in I\!N$, com $u^k = -\frac{1}{\mu_k}\nabla h(x(\mu_k)) \in T(x(\mu_k))$, em contradição com a hipótese.

Logo, $\{x(\mu)/\mu > \bar{\mu}\}$ é limitado. Portanto, $\{x(\mu)/\mu > \bar{\mu}\}$ possui pontos de acumulação, para todo $\bar{\mu} > 0$.

Suponha que o item (iii) da hipótese H14 é válido. Isto é, $dom(T) \cap C$ é limitado. Pelo Lema 2.1, $\{x(\mu)/\mu > 0\} \subset dom(T) \cap C^0 \subset dom(T) \cap C$. Logo, $\{x(\mu)/\mu > 0\}$ é limitado. Portanto, $\{x(\mu)/\mu > 0\}$ possui pontos de acumulação. ■

Observação 2.3.3 *Seja $T = \widehat{T} + N_V$, como na hipótese* H8*, Como $\widehat{T}$ é paramonótono e $N_V = \partial\delta_V$ é paramonótono então T é paramonótono.*

Agora é mostrado que os pontos de acumulação de $\{x(\mu)/\mu > 0\}$ são soluções do problema $VIP(T, C)$.

Lema 2.4 *Com as hipóteses do Lema 2.3, todo ponto de acumulação de $\{x(\mu)/\mu > \bar{\mu}\}$ é solução do problema $VIP(T, C)$.*

Demonstração. Pelas hipóteses H10 e H11 segue que $S(T, C) \neq \phi$ e $dom(T) \cap C^0 = V \cap C^0 \neq \phi$. Sejam $z \in S(T, C)$ e $y \in V \cap C^0$. Considere um ponto de acumulação $\bar{x}$ de $\{x(\mu)/\mu > \bar{\mu}\}$ e uma seqüência $\{\mu_k\}_{k\in I\!N}$ tal que $\lim\limits_{k\to+\infty} \mu_k = +\infty$ e $\lim\limits_{k\to+\infty} x(\mu_k) = \bar{x}$. Seja $y(\epsilon) = (1-\epsilon)z + \epsilon y$, com $\epsilon \in (0, 1)$. Então $y(\epsilon) \in V \cap C^0$. De (2.25), tem-se

$$\frac{1}{\mu_k}\langle \nabla h(x(\mu_k)), x(\mu_k) - y(\epsilon)\rangle \leq \langle \widehat{T}(x(\mu_k)), y(\epsilon) - x(\mu_k)\rangle. \tag{2.29}$$

Como $\lim\limits_{k\to+\infty} \langle \nabla h(x(\mu_k)), x(\mu_k) - y(\epsilon)\rangle = \langle \nabla h(\bar{x}), \bar{x} - y(\epsilon)\rangle < +\infty$ e $\lim\limits_{k\to+\infty} \mu_k = +\infty$, então $\lim\limits_{k\to+\infty} \frac{1}{\mu_k}\langle \nabla h(x(\mu_k)), x(\mu_k) - y(\epsilon)\rangle = 0$. Logo, tomando limite em (2.29), quando k tende para $+\infty$, tem-se:

$$0 \leq \langle \widehat{T}(\bar{x}), y(\epsilon) - \bar{x}\rangle. \tag{2.30}$$

Daí, $0 \leq \lim_{\epsilon \to 0} \langle \widehat{T}(\bar{x}), y(\epsilon) - \bar{x} \rangle = \langle \widehat{T}(\bar{x}), z - \bar{x} \rangle$. Como V é fechado, $\bar{x} \in V$. Como $N_V(\bar{x})$ é um cone, $0 \in N_V(\bar{x})$. Assim, $\widehat{T}(\bar{x}) \in T(\bar{x})$.

Tem-se que $T = \widehat{T} + N_V$ é paramonótono, $z \in S(T, C)$ e $\langle \widehat{T}(\bar{x}), z - \bar{x} \rangle \geq 0$, com $\widehat{T}(\bar{x}) \in T(\bar{x})$. Pela Proposição 1.3.1, $\bar{x} \in S(T, C)$. Portanto, todo ponto de acumulação de $\{x(\mu)/\mu > \bar{\mu}\}$ é solução do problema $VIP(T, C)$. ■

No próximo resultado, é estabelecida a convergência de $\{x(\mu)/\mu > \bar{\mu}\}$ para o centro analítico do conjunto $S(T, C)$, quando h é finita na fronteira. Neste caso, H9 não é necessário.

Lema 2.5 *Com as hipóteses do Lema 2.4, se h é finita na fronteira (veja* H3*), então $\lim_{\mu \to +\infty} x(\mu)$ existe e é solução do problema:*

$$\left\{ \min_{x \in S(T,C)} h(x). \right. \tag{2.31}$$

Demonstração. Pelo Lema 2.3, $\{x(\mu)/\mu > \bar{\mu}\}$ possui pontos de acumulação. Seja $\bar{x}$ um ponto de acumulação de $\{x(\mu)/\mu > \bar{\mu}\}$. Pelo Lema 2.4, $\bar{x} \in S(T, C)$. Seja $\{\mu_k\}_{k \in I\!N}$ uma seqüência tal que $\lim_{k \to +\infty} \mu_k = +\infty$ e $\lim_{k \to +\infty} x(\mu_k) = \bar{x}$. Considere $z \in S(T, C)$. Então, usando (2.25), temos

$$\frac{1}{\mu_k} \langle \nabla h(x(\mu_k)), x(\mu_k) - z \rangle \leq \langle \widehat{T}(z), z - x(\mu_k) \rangle. \tag{2.32}$$

Observe que $x(\mu_k) \in V$. Como $z \in S(T, C)$, $\exists\, u = \widehat{T}(z) + v$ tal que $\langle u, x(\mu_k) - z \rangle \geq 0$, onde $v \in N_V(z)$. Logo, $\langle \widehat{T}(z), x(\mu_k) - z \rangle + \langle v, x(\mu_k) - z \rangle \geq 0$. Daí,

$$\langle \widehat{T}(z), z - x(\mu_k) \rangle \leq \langle v, x(\mu_k) - z \rangle.$$

Tem-se que $v \in N_V(z)$ se, e somente se $\langle v, x(\mu_k) - z \rangle \leq 0$. Assim, $\langle \widehat{T}(z), z - x(\mu_k) \rangle \leq 0$. Logo, em (2.32) temos que $\langle \nabla h(x(\mu_k)), x(\mu_k) - z \rangle \leq 0$. Pela convexidade de h,

$$h(x(\mu_k)) - h(z) \leq \langle \nabla h(x(\mu_k)), x(\mu_k) - z \rangle \leq 0. \tag{2.33}$$

Pela hipótese, h é finita na fronteira. Logo, tomando limite em (2.33), quando k tende para $+\infty$, temos que $h(\bar{x}) \leq h(z)$. Como z é um elemento arbitrário de $S(T, C)$, $\bar{x}$ é um minimizador de h em $S(T, C)$. Como $S(T, C)$ é convexo e h é estritamente convexa em C (veja H3), então $\bar{x}$ é único. Assim, todos os pontos de acumulação coincidem e $\{x(\mu)/\mu > \bar{\mu}\}$ converge, quando μ tende para $+\infty$, à solução de (2.31). ■

Finalmente, a existência, unicidade e comportamento da trajetória central para o problema $VIP(T, C)$ são resumidos no seguinte resultado.

Teorema 2.1 *Suponha que h satisfaz* H1*, é coerciva na fronteira (veja* H2*), e atinge seu mínimo em $dom(T) \cap C^0$ (veja* H12*); que T satisfaz* H8*, que $VIP(T, C)$ é regular e possui soluções (veja* H10 *e* H11*, respectivamente). Em adição, ou*
i) *h é zona coerciva e finita na fronteira (veja* H4 *e* H3*, respectivamente), ou*
ii) *a função gap é finita (veja* H13*) e que alguma das alternativas em* H14 *é válida. Então, para cada $\bar{\mu} > 0$, a trajetória central $\{x(\mu)/\mu > \bar{\mu}\}$ está bem definida, é contínua, é limitada e está contida em C^0, e todos seu pontos de acumulação são soluções do problema $VIP(T, C)$. No caso em que h seja finita na fronteira (veja* H3*), a trajetória central $\{x(\mu)/\mu > \bar{\mu}\}$ converge para o centro analítico do conjunto de soluções $S(T, C)$.*

Demonstração. Pelo Lema 2.1, $\{x(\mu)/\mu > \bar{\mu}\}$ está bem definida e contida em C^0. Observe que as hipóteses do Lema 2.1 são válidas em todos os casos neste teorema, pois H8 implica que T é monótono maximal. Pelo item (ii) do Lema 2.2, $\{x(\mu)/\mu > \bar{\mu}\}$ é contínua e limitada. Logo, $\{x(\mu)/\mu > \bar{\mu}\}$ possui pontos de acumulação. Pelo Lema 2.3, $\{x(\mu)/\mu > \bar{\mu}\}$ possui pontos de acumulação. Pelo Lema 2.4, todos os pontos de acumulação de $\{x(\mu)/\mu > \bar{\mu}\}$ são soluções do problema $VIP(T, C)$. Pelo Lema 2.5, $\{x(\mu)/\mu > \bar{\mu}\}$ converge, quando μ tende para $+\infty$, para o centro analítico do conjunto de soluções $S(T, C)$, sempre que h seja finita na fronteira. ■

Observação 2.3.4 *Para problemas de otimização convexa da forma*

$$\begin{cases} \min\limits_{x \in C} f(x), \end{cases}$$

onde f é a restrição de uma função convexa e diferenciável $\widehat{f} : I\!R^n \to I\!R$ a um conjunto convexo fechado V. Neste caso, no item (iv) tem-se $\widehat{T} = \nabla \widehat{f}$ e o Teorema 2.1 é válido. Quando $\widehat{f}$ não é diferenciável tem-se que $\widehat{T} = \partial \widehat{f}$ não é necessariamente um operador ponto a ponto e o Teorema 2.1 não pode ser aplicado.

2.4 Comportamento da Trajetória Central com Funções Barreiras Separáveis

Nesta seção, é analizado o comportamento de $\{x(\mu)/\mu > \bar{\mu}\}$ quando h não é finita na fronteira. Para isto precisamos de algumas considerações. Primeiro suponha que $C = I\!R^n_+$ e que h é separável (veja H5). Além disso, T é considerado satisfazendo H8 e H9.

Os seguintes dois resultados são necessários para mostrar um resultado similar ao Lema 2.5 sem que h seja finita na fronteira.

Lema 2.6 *Suponha que T satisfaz H8 e H9, e que $VIP(T, C)$ é regular e possui soluções (veja H10 e H11, respectivamente). Fixe qualquer $x \in S(T, C)$ e qualquer $\tilde{x} \in C \cap V$. Então*

$$S(T, C) = \left\{ x \in V \cap C \ / \ \widehat{T}(\tilde{x})^t x = \widehat{T}(\tilde{x})^t z, \ J_{\widehat{T}}(\tilde{x})x = J_{\widehat{T}}(\tilde{x})z \right\}. \tag{2.34}$$

Demonstração. Segue da Proposição 1.3.3, página 12. ■

Lema 2.7 *Seja $C = I\!R^n_+$. Com as hipóteses do Lema 2.1 e do Lema 2.6, o ponto $x(\mu)$ pertencente à trajetória central é a solução do problema:*

$$\left\{ \min_{x \in S(\mu)} h(x), \right. \tag{2.35}$$

onde

$$S(\mu) = \left\{ x \in I\!R^n \ / \widehat{T}(\tilde{x})^t x = \widehat{T}(\tilde{x})^t x(\mu), J_{\widehat{T}}(\tilde{x})x = J_{\widehat{T}}(\tilde{x})x(\mu), Ax = b, x \geq 0 \right\}, \tag{2.36}$$

$\tilde{x}$ é qualquer ponto em $C \cap V$, $V = \{x \in I\!R^n / Ax = b\}$, $A \in I\!R^{m \times n}$ e $b \in I\!R^m$.

Demonstração. Pela convexidade de h, basta verificar que $x(\mu)$ satisfaz as condições de otimalidade de Karush Kuhn Tucker (veja seção B.2), que são:

$$x(\mu) \in S(\mu), \tag{2.37}$$

$$\nabla h(x(\mu)) + A^t u + J_{\widehat{T}}(\tilde{x})^t w + \xi \widehat{T}(\tilde{x}) \geq 0, \tag{2.38}$$

$$x(\mu)^t \left[\nabla h(x(\mu)) + A^t u + J_{\widehat{T}}(\tilde{x})^t w + \xi \widehat{T}(\tilde{x}) \right] = 0, \tag{2.39}$$

para algum $u \in I\!R^m$, algum $w \in I\!R^n$ e algum $\xi \in I\!R$. De fato.

Pelo Lema 1.4, $x(\mu) \in dom(T) \cap C^0 = V \cap C^0$. Logo, (2.37) é válido. Para mostrar que (2.38) e (2.39) são válidos, basta mostrar que existem u, w e ξ satisfazendo (2.38) como igualdade e, conseqüêntemente (2.39) é válido.

Por definição de trajetória central, $-\frac{1}{\mu}\nabla h(x(\mu)) \in T(x(\mu))$. Isto é, existe $v \in N_V(x)$ tal que

$$-\frac{1}{\mu}\nabla h(x(\mu)) = \widehat{T}(x(\mu)) + v. \tag{2.40}$$

Pela Proposição 2.2.2, $N_V(x) = Im(A^t)$, $\forall x \in V$. Pela Observação 2.2.4, $N_V(x)$ é um cone. Então, $\mu v \in Im(A^t)$. Daí, existe $u \in I\!R^m$ tal que $\mu v = A^t u$. Assim, em (2.40), temos

$$\nabla h(x(\mu)) + \mu\widehat{T}(x(\mu)) + A^t u = 0. \tag{2.41}$$

Seja $\tau \in [0,1]$. Observe que $\widehat{T}(x(\mu))$ pode ser escrito como

$$\widehat{T}(x(\mu)) = \widehat{T}(\tilde{x}) + \int_0^1 J_{\widehat{T}}(y(\tau))(x(\mu) - \tilde{x})d\tau, \tag{2.42}$$

onde $y(\tau) = \tilde{x} + \tau(x(\mu) - \tilde{x})$. Como $\widehat{T}$ é um operador ponto a ponto monótono e diferenciável, $Ker\left(J_{\widehat{T}}(x)\right) = Ker\left(J_{\widehat{T}}(x)^t\right)$ (veja item (ii) na Proposição 1.2.2).

Usando H9, temos

$$J_{\widehat{T}}(y(\tau))(x(\mu) - \tilde{x}) \in Im\left[J_{\widehat{T}}(y(\tau))\right] = Ker\left[J_{\widehat{T}}(y(\tau))^t\right]^{\perp} = Ker\left[J_{\widehat{T}}(y(\tau))\right]^{\perp} = W^{\perp}.$$

Segue que, $\int_0^1 J_{\widehat{T}}(y(\tau))(x(\mu) - \tilde{x})d\tau \in W^{\perp}$. De (1.45), $\widehat{T}(x(\mu)) = \widehat{T}(\tilde{x}) + p$, com $p \in W^{\perp}$. Pela hipótese H9, $p \in W^{\perp} = Ker\left[J_{\widehat{T}}(\tilde{x})\right]^{\perp} = Im\left[J_{\widehat{T}}(\tilde{x})^t\right]$. Logo, existe $w \in I\!R^n$ tal que $p = J_{\widehat{T}}(\tilde{x})^t w$. Assim,

$$\mu\widehat{T}(x(\mu)) = \mu\widehat{T}(\tilde{x}) + J_{\widehat{T}}(\tilde{x})^t w. \tag{2.43}$$

De (2.41) e (2.43), segue que:

$$\nabla h(x(\mu)) + \mu\widehat{T}(\tilde{x})) + J_{\widehat{T}}(\tilde{x})^t w + A^t u = 0, \tag{2.44}$$

que é a igualdade (2.38), com $\xi = \mu$. Assim, a factibilidade dual e a condição complementar são satisfeitas.

Portanto, $x(\mu)$ é solução do problema (2.35). ■

A idea é aproximar, quando μ tende para $+\infty$, a otimalidade de $x(\mu)$ dada no Lema 2.7, de forma que $S(\mu)$ possa chegar a ser $S(T, C)$, como no Lema 2.6.

A dificuldade é que $\lim_{\mu\to+\infty} x(\mu)$ pode pertencer a ∂C onde h é possivelmente $+\infty$.

Agora devemos trabalhar com a face ótima de $I\!R^n_+$ e com a separabilidade de h (veja H5). Seja $J = \{j \in \{1,...,n\}\,/z_j > 0, \text{ para algum } z \in S(T,C)\}$. Pela convexidade de $S(T,C)$, o conjunto $Z = \{z \in S(T,C)/z_j > 0, \forall j \in J\}$ é não vazio. Observe que Z é o interior relativo de $S(T,C)$. Defina $\tilde{h} : \partial I\!R^n_{++} \to I\!R$, por

$$\tilde{h}(x) = \sum_{j\in J} h_j(x_j). \tag{2.45}$$

A seguir é mostrada a otimalidade de $\lim_{\mu\to+\infty} x(\mu)$ quando h não é finita na fronteira (veja H3).

Teorema 2.2 *Seja $C = I\!R^n_+$. Suponha que h satisfaz* H1*, é coerciva na fronteira (veja* H2*), é separável (veja* H5*) e atinge seu mínimo em dom$(T) \cap C^0$ (veja* H12*); que T satisfaz* H8 *e* H9*; que VIP(T,C) é regular (veja* H10*) e possui soluções (veja* H11*); que a função gap é finita (veja* H13*); e que alguma das alternativas de* H14 *é válida. Então a trajetória central $\{x(\mu)/\mu > 0\}$ converge, quando μ tende para $+\infty$, para a solução x^* do problema*

$$\left\{ \min_{x\in S(T,C)} \tilde{h}(x) \right. , \tag{2.46}$$

com $\tilde{h}$ como em (2.45)*.*

Demonstração. Pelo Teorema 2.1, $\{x(\mu)/\mu > 0\}$ possui pontos de acumulação. Sejam $\bar{x}$ um ponto de acumulação de $\{x(\mu)/\mu > 0\}$ e uma seqüência $\{\mu_k\}_{k\in I\!N}$ tal que $\lim_{k\to+\infty} \mu_k = +\infty$, $\lim_{k\to+\infty} x(\mu_k) = \bar{x}$. Denote-se $x^k = x(\mu_k)$, $\forall k \in I\!N$. Fixe algum $\tilde{x} \in V \cap C$ como nos Lemas 2.6 e 2.7. Considere qualquer $z \in S(T,C)$.

Afirmamos que

$$\tilde{h}(\bar{x}) \le \tilde{h}(z). \tag{2.47}$$

Primeiro é considerado quando z pertence ao interior relativo de $S(T,C)$, isto é, $z_j > 0$, $\forall j \in J$. Defina $y^k = z - \bar{x} + x^k$. Temos que $y^k \in S(\mu_k)$, para k suficientemente grande, com $S(\mu_k)$ como em (2.36). De fato. Seja $I = \{1,...,n\} \setminus J$. Da definição de J segue que $x_j = 0$ para todo $j \in I$ e todo $x \in S(T,C)$. Como $\bar{x}$ é ponto de acumulação de $\{x(\mu)/\mu > 0\}$ então, pelo Teorema 2.1, $\bar{x} \in S(T,C)$. Logo, como $z \in S(T,C)$, temos

$$y^k_j = x^k_j, \;\; (j \in I\,). \tag{2.48}$$

Pelo Lema 2.1, $x^k \in C^0$. Segue que $y_j^k > 0$ para $j \in I$. Para $j \in J$ temos que $y_j^k = z_j - \bar{x}_j + x_j^k$. Como $\lim\limits_{k \to +\infty} x_j^k = \bar{x}_j$ e $z_j > 0$ para $j \in J$ então $y_j^k > 0$ para $j \in J$ e k suficientemente grande. Conclua-se que $y_j^k > 0$ para todo k suficientemente grande.

Seja V o conjunto dado pelas hipóteses H8 e H9, igual a $\{x \in I\!R^n / Ax = b\}$. Como $\bar{x}, z \in S(T, C) \subset dom(T) = V$, então $Az = A\bar{x} = b$ e assim $Ay^k = Ax^k = b$, pois $x(\mu) \in dom(T) \subset V$. Logo, é suficiente verificar $\widehat{T}(\tilde{x})^t(x^k - y^y) = 0$ e $J_{\widehat{T}}(\tilde{x})(x^k - y^y) = 0$. Como $x^k - y^k = \bar{x} - z$, estes fatos seguem do Lema 2.6, pois pelo Teorema 2.1 segue que $\bar{x} \in S(T, C)$. Assim, $y^k \in S(\mu_k)$.

Pelo Lema 2.7, $h(x^k) \leq h(y^k)$. Logo,

$$\tilde{h}(x^k) + \sum_{j \in I} h_j(x_j^k) \leq \tilde{h}(y^k) + \sum_{j \in I} h_j(y_j^k). \tag{2.49}$$

De (2.48) e (2.49),

$$\tilde{h}(x^k) \leq \tilde{h}(y^k). \tag{2.50}$$

Agora, como h não é finita na fronteira tem-se que ter muito cuidado com o comportamento de $\tilde{h}$ na fronteira de $I\!R_+^n$.

Segue das hipóteses H1 e H5 que $h_j : I\!R_{++} \to I\!R$ é contínua e fechada. Assim, $\lim\limits_{t \to 0} h_j(t)$ está bem definido (possivelmente infinito) e pode-se tomar limite em (2.50) quando k tende para $+\infty$. Como $\lim\limits_{k \to +\infty} y^k = z$, $\lim\limits_{k \to +\infty} x^k = \bar{x}$ e $z_j > 0$ para todo $j \in J$, então

$$\tilde{h}(\bar{x}) \leq \tilde{h}(z) < +\infty, \tag{2.51}$$

para qualquer z no interior relativo de $S(T, C)$. Pelo mesmo argumento, (2.51) é válido para todo $z \in S(T, C)$ (note que $\tilde{h}$ pode ser infinito para algum x na fronteira relativa de $S(T, C)$, isto é, $z_j = 0$ para algum $j \in J$). Assim, $\bar{x}$ é solução do problema (2.46). Observe que $\tilde{h}$ não é estritamente convexa em $I\!R_+^n$, mas é estritamente convexa em $S(T, C)$ pois h_j é estritamente convexa por H1 e $x_j = 0$ para todo $x \in S(T, C)$, por definição de J. Assim, o problema (2.46) possui uma única solução e como todos os pontos de acumulação de $\{x(\mu)/\mu > 0\}$ são iguais a esta solução, conclui-se que $\lim\limits_{\mu \to +\infty} x(\mu)$ existe e converge para a solução x^* do problema (2.46). ■

Capítulo 3

Métodos de Ponto Proximal Generalizado para Problemas de Inequações Variacionais

Neste Capitulo apresentaremos resultados de existência e convergência do Algoritmo de Ponto Proximal Genaralizado para problemas $VIP(T, C)$. A maioria dos fatos foram obtidos de [4], [5] e [12].

3.1 Funções e Distâncias Generalizadas de Bregman

Seja $C \subset I\!R^n$ um conjunto fechado e convexo, com interior não vazio.

Definição 3.1.1 *Dizemos que $g : C \to I\!R$ é função de Bregman e $D_g : C \times C^0 \to I\!R$, definida por*

$$D_g(x, y) = g(x) - g(y) - \langle \nabla g(y), x - y \rangle, \tag{3.1}$$

é distância generalizada de Bregman induzida por g se as hipóteses H1 *e* H3 *são satisfeitas e, em adição, são válidas*

B1. *Para todo $\delta \in I\!R$, os conjuntos de nível $\Gamma_1(y, \delta) = \{x \in C / D_g(x, y) \leq \delta\}$ e $\Gamma_2(x, \delta) = \{y \in C^0 / D_g(x, y) \leq \delta\}$, são limitados, para todo $y \in C^0$ e todo $x \in C$, respectivamente.*

B2. *Se $\{y^k\}_{k \in I\!N}$ é seqüência em C^0 tal que $\lim\limits_{k \to +\infty} y^k = y^*$, então $\lim\limits_{k \to +\infty} D_g(y^*, y^k) = 0$.*

B3. *Se* $\{x^k\}_{k\in I\!N} \subset C$ *e* $\{y^k\}_{k\in I\!N} \subset C^0$ *são seqüências tais que* $\{x^k\}_{k\in I\!N}$ *é limitada,* $\lim_{k\to+\infty} y^k = y^*$ *e* $\lim_{k\to+\infty} D_g(x^k, y^k) = 0$, *então* $\lim_{k\to+\infty} x^k = y^*$.

Nestas considerações, C^0 *é chamada zona de* g.

Observação 3.1.1 *Pela convexidade estrita de* g, *para* $x \in C$ *e* $y \in C^0$, *temos que* $D_g(x,y) \geq 0$. *Além disso,* $D_g(x,y) = 0$ *se, e somente se,* $x = y$.

Observação 3.1.2 B2 *e* B3 *são válidos automaticamente quando* $\{x^k\}_{k\in I\!N} \subset C^0$ *e* $y^* \in C^0$. *Assim,* B2 *e* B3 *necessitam ser verificados somente em* ∂C.

Observação 3.1.3 *Se* g *é função de Bregman com zona* C^0 *e coerciva na fronteira (veja* H2*) então, para cada* $y \in \partial C$ *segue que* $D_g(x,y) = +\infty$, *para todo* $x \in C^0$.

De fato. Como g é coerciva na fronteira então $\lim_{k\to+\infty} \langle \nabla g(y^k), x - y^k \rangle = -\infty$, para todo $x \in C^0$ e alguma seqüência $\{y^k\}_{k\in I\!N} \subset C^0$ tal que $\lim_{k\to+\infty} y^k = y \in \partial C$. Logo,

$$\begin{aligned} D_g(x,y) &= g(x) - g(y) - \langle \nabla g(y), x - y \rangle \\ &= g(x) - g(\lim_{k\to+\infty} y^k) - \langle \nabla g(\lim_{k\to+\infty} y^k), x - \lim_{k\to+\infty} y^k \rangle \\ &= \lim_{k\to+\infty} g(x) - g(y^k) - \langle \nabla g(y^k), x - y^k \rangle = +\infty, \end{aligned}$$

para todo $x \in C^0$. ■

Exemplo 3.1.1 *Seja* $C = I\!R^n_+$ *e considere* $g(x) = \sum_{j=1}^n x_j \, log(x_j)$, *estendido para* $\partial I\!R^n_+$ *com a convenção* $0 \, log \, 0 = 0$. *Neste caso,*

$$D_g(x,y) = \sum_{j=1}^n x_j \, log\left(\frac{x_j}{y_i}\right) + y_j - x_j$$

(chamada divergência de Kullback-Leibler). Esta função de Bregman satisfaz H1, H2, H3, H4, H5, B1, B2 *e* B3.

Exemplo 3.1.2 *Seja* $C = I\!R^n_+$ *e considere* $g(x) = \sum_{j=1}^n \left(x_j^\alpha - x_j^\beta\right)$, *com* $\alpha \geq 1$, $0 < \beta < 1$. *Para* $\alpha = 2$ *e* $\beta = \frac{1}{2}$, *temos*

$$D_g(x,y) = \| x - y \|^2 + \frac{1}{2} \sum_{j=1}^n \frac{1}{\sqrt{y_j}} \left(\sqrt{x_i} - \sqrt{y_j}\,\right)^2,$$

e para $\alpha = 1$ *e* $\beta = \frac{1}{2}$, *temos*

$$D_g(x,y) = \frac{1}{2}\sum_{j=1}^{n} \frac{1}{\sqrt{y_j}} \left(\sqrt{x_i} - \sqrt{y_j}\,\right)^2.$$

Esta função de Bregman satisfaz H1, H2, H3, H4, H5, B1, B2 *e* B3, *exceto quando* $\alpha = 1$, *que não satisfaz* H4.

Exemplo 3.1.3 *Seja* $C = I\!R^n_+$. *Se* H3 *fosse descartada, é possível considerar a função* $g(x) = -\sum_{j=1}^{n} log(x_j)$ *que satisfaz* H1, H2, H5, B1, B2 *e* B3. *Neste caso,*

$$\begin{array}{rcl} D_g \;:\; C^0 \times C^0 & \longrightarrow & I\!R \\ x \qquad\quad & \longmapsto & D_g(x,y) \;:=\; \displaystyle\sum_{j=1}^{n} \left\{ \frac{x_j}{y_j} - log\left(\frac{x_j}{y_i}\right) - 1 \right\}. \end{array}$$

Esta distância é chamada de Itakura - Saitu.

Proposição 3.1.1 *Se* g *é uma função de Bregman com zona* C^0 *então*

i) $D_g(x,y) - D_g(x,z) - D_g(z,y) = \langle \nabla g(y) - \nabla g(z), z - x\rangle$, $\forall x \in C$, $\forall y, z \in C^0$.

ii) $\nabla_x D_g(x,y) = \nabla g(x) - \nabla g(y)$, $\forall x, z \in C^0$.

iii) $D_g(\cdot, y)$ *é estritamente convexa para todo* $y \in C^0$.

Demonstração.

i) De (2.1), para x em C, y e z em C^0, temos

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

ii) Sejam x e z em C^0. Como $D_g(x,y) = g(x) - g(y) - \langle \nabla g(y), x - y\rangle$ então

$$\begin{aligned} \nabla_x D_g(x,y) &= \nabla g(x) - 0 - \langle \frac{\partial}{\partial x}\nabla g(y), x - y\rangle - \langle \nabla g(y), \frac{\partial}{\partial x}(x - y)\rangle \\ &= \nabla g(x) - \langle 0, x - y\rangle - \langle \nabla g(y), 1\rangle \\ &= \nabla g(x) - \nabla g(y). \end{aligned}$$

iii) Para cada $y \in C^0$, defina

$$\begin{array}{rcl} G : C & \longrightarrow & I\!R \\ x & \longmapsto & G(x) := D_g(x,y). \end{array}$$

Sejam x_1 e x_2 em C, com $x_1 \neq x_2$. Pelo item (ii), $\nabla G(x_1) = \nabla h(x_1) - \nabla h(y)$ e $\nabla G(x_2) = \nabla h(x_2) - \nabla h(y)$. Logo,

$$\begin{aligned} \langle \nabla G(x_1) - \nabla G(x_2), x_1 - x_2 \rangle &= \langle \nabla h(x_1) - \nabla h(y) - \nabla h(x_2) + \nabla h(y), x_1 - x_2 \rangle \\ &= \langle \nabla h(x_1) - \nabla h(x_2), x_1 - x_2 \rangle > 0, \end{aligned}$$

pois h é estritamente convexa em C. Portanto, pelo Corolario A.10, $G = D_g(\cdot, y)$ é estritamente convexa em C, para cada $y \in C^0$. ■

3.2 $GPPA$ para $VIP(T,C)$

O algoritmo ponto proximal generalizado ($GPPA$) para o problema $VIP(T,C)$ é definido da seguinte forma.

Considere uma função de Bregman g com zona C^0 e uma seqüência de números positivos $\{\lambda_k\}$ limitada por algum $\tilde{\lambda} > 0$. Seja $\{x^k\}$ a seqüência definida por:

Inicialização.

$$x^0 \in C^0. \tag{3.2}$$

Iteração. Dado $x^k \in I\!R^n$, defina $T_k : I\!R^n \to \mathcal{P}(I\!R^n)$ por $T_k(\cdot) = T(\cdot) + \lambda_k \partial_x D_g(\cdot, x^k)$. Então considere $x^{k+1} \in I\!R^n$ tal que

$$0 \in T_k(x^{k+1}). \tag{3.3}$$

Lema 3.1 *Seja* $\{x^k\}$ *a seqüência gerada por* (3.2) *e* (3.3). *Se* $VIP(T,C)$ *é regular (veja* H10*),* T *é monótono maximal e a função de Bregman* g *é zona coerciva (veja* H4*), então* $\{x^k\}$ *está bem definida e contida em* C^0.

Demonstração. A prova é feita por indução em k.

De fato. De (3.2), $x^0 \in C^0$. Suponha que $x^k \in C^0$. Seja $B := \lambda_k \partial_x D_g(\cdot, x^k)$. Assim, $T_k = T + B_k$. Pela Proposição 2.2.5, $dom(B_k) = C^0$. Pela regularidade de $VIP(T,C)$, $dom(T) \cap dom(B_k) \neq \phi$. Pelo item (i) da Proposição 2.2.6, T_k é monótono maximal. Como g é zona coerciva, B_k é sobrejetivo. Pelo item (i) da Proposição 2.2.6, T_k é sobrejetivo. Logo, T_k possui um zero em $dom(T_k)$. Temos que B_k é estritamente monótono, pois g é estritamente convexa. Daí, T_k é estritamente monótono. Logo, o zero de T_k em $dom(T_k)$ é único. Denote este único zero de T_k em $dom(T_k)$ por x^{k+1}. Observe que $dom(T_k) = dom(T) \cap dom(B_k) = dom(T) \cap C^0$. Daí, $x^{k+1} \in dom(T) \cap C^0$, $\forall k$. Isto implica que $x^{k+1} \in C^0$, $\forall k$.

Portanto, $\{x^k\}$ está bem definida e contida em C^0. ■

Lema 3.2 *Seja* $\{x^k\}$ *a seqüência gerada por* (3.2) *e* (3.3). *Suponha que* $VIP(T,C)$ *seja regular e possua soluções (veja* H10 *e* H11, *respectivamente),* $Gap_{\,T,C}(x) < +\infty$, $\forall x \in dom(T) \cap C$ *(veja* H13*), g é coerciva na fronteira e T satisfaz* H6. *Então* $\{x^k\}$ *está bem definida e contida em* C^0.

Demonstração. A prova é feita por indução em k.

De fato. De (3.2), $x^0 \in C^0$. Por hipótese de indução, suponha que existe $x^k \in dom(T) \cap C^0$ tal que $0 \in T_{k-1}(x^k)$. Em particular, $x^k \in dom(T) \cap C$. Como a função gap é finita em $dom(T) \cap C$, $0 \leq Gap_{\,T,C}(x^k) < +\infty$. Assim, temos dois casos para analizar.

i) Se $Gap_{\,T,C}(x^k) = 0$ então, pela Proposição 2.2.4, $x^k \in S(T,C)$. Afirmamos que, se tomamos $x^{k+1} = x^k$ então $0 \in T_k(x^{k+1})$.

De fato, pelo item (ii) da Proposição 3.1.1 segue

$$\begin{aligned} T_k(x^{k+1}) &= T(x^{k+1}) + \lambda_{k+1}\partial_{x^{k+1}} D_g(x^{k+1}, x^k) \\ &= T(x^k) + \lambda_{k+1}[\nabla g(x^{k+1}) - \nabla g(x^k)] = T(x^k) \\ &= T(x^k) + \lambda_{k+1}[\nabla g(x^k) - \nabla g(x^k)] = T(x^k). \end{aligned} \tag{3.4}$$

Daí, $T_k(x^{k+1}) = T(x^k)$, com $x^k \in dom(T) \cap C^0 \cap S(T,C)$. Pela Proposição 2.2.1, x^k é zero de $T + N_V(x^k)$ em C. Isto é, $0 \in (T + N_V)(x^k) = T(x^k) + N_V(x^k)$. Para $u \in T(x^k)$ temos que $0 = u + 0 = u$. Segue que $0 \in T(x^k)$. Logo, de (3.4), $0 \in T_k(x^{k+1})$.

Assim, neste caso, $x^{k+1} \in C^0$, para todo k.

ii) Considere o caso $Gap_{\,T,C}(x^k) > 0$. Defina o conjunto

$$S_k := \left\{ x \in C \;/\; D_g(x, x^k) \leq \frac{Gap_{\,T,C}(x^k)}{\lambda_k} \right\}.$$

Afirmamos que S_k é convexo. De fato. Sejam $x_1, x_2 \in S_k$. Então

$$D_g(x_1, x^k) \leq \frac{Gap_{\ T,C}(x^k)}{\lambda_k},\ D_g(x_2, x^k) \leq \frac{Gap_{\ T,C}(x^k)}{\lambda_k}.$$

Temos que $D_g(\cdot, x^k)$ é convexa (veja Proposição 3.1.1). Logo, para $\alpha \in [0,1]$, temos

$$\begin{aligned} D_g(\alpha x_1 + (1-\alpha)x_2, x^k) &\leq \alpha\, D_g(x_1, x^k) + (1-\alpha)\, D_g(x_2, x^k) \\ &\leq\ \alpha \frac{Gap_{\ T,C}(x^k)}{\lambda_k} + (1-\alpha)\, \frac{Gap_{\ T,C}(x^k)}{\lambda_k} \\ &= \frac{Gap_{\ T,C}(x^k)}{\lambda_k}. \end{aligned}$$

Assim, $\alpha x_1 + (1-\alpha)x_2 \in S_k$. Logo, S_k é convexo.

Afirmamos que S_k é fechado. De fato. Seja $y \in cl(S_k)$. Então, existe $\{y^n\}_{n\in I\!N}$ em S_k tal que $\lim_{n\to+\infty} y^n = y$. Como $y^n \in S_k$, $\forall n \in I\!N$, então

$$D_g(y_n, x^k) \leq \frac{Gap_{\ T,C}(x^k)}{\lambda_k},\ \forall n \in I\!N.$$

Logo,

$$D_g(y, x^k) = \lim_{n\to+\infty} D_g(y^n, X^k) \leq \frac{Gap_{\ T,C}(x^k)}{\lambda_k}.$$

Assim, $y \in S_k$. Portanto, S_k é fechado.

Afirmamos que S_k é limitado, pois S_k é um conjunto de nível da distância de Bregman e, pela hipótese B2, todos os conjuntos de nível são limitados.

Além disso $x \in {S_k}^0$ se, e somente se,

$$x \in C^0 \ \text{ e } \ D_g(x, x^k) \leq \frac{Gap_{\ T,C}(x^k)}{\lambda_k}\,. \tag{3.5}$$

Observe que $x^k \in {S_k}^0$, pois $D_g(x^k, x^k) = 0 \leq \dfrac{Gap_{\ T,C}(x^k)}{\lambda_k}$ e $x^k \in C^0$, pela hipótese de indução.

Seja $N_k := N_{S_k}$ o operador normalizado de S_k. Pela Observação 2.2.3, $dom(N_k) = S_k$. Logo, $dom(N_k)$ é limitado. Agora defina $B_k(\cdot) := N_k(\cdot) + \lambda_k \partial_x D_g(\cdot, x^k)$. Para mostrar B_k é monótono maximal, pelo item (i) da Proposição 2.2.6, basta mostrar que $dom(N_k) \cap dom(\nabla_x D_g(\cdot, x^k))^0 \neq \phi$. De fato. Da hipótese de indução, $x^0 \in C^0$. Além disso, $x^k \in {S_k}^0 \subset S_k = dom(N_k)$. Logo, $dom(N_k) \cap C^0 \neq \phi$. Pela Proposição 2.2.5, $C^0 = dom(\partial_x D_g(\cdot, x^k)) = dom(\nabla_x D_g(\cdot, x^k))^0$. Assim, $dom(B_k) \cap dom(\nabla_x D_g(\cdot, x^k))^0 \neq \phi$. Logo, B_k é monótono maximal. Observe que, $dom(B_k)$ é limitado, pois $dom(B_k)$

é um subconjunto de S_k. Logo, pelo item (ii) da Proposição 2.2.6, B_k é operador sobrejetivo.

Considere agora o operador $A_k := T + B_k$. Para mostrar que A_k é monótono maximal, pelo item (i) da Proposição 2.2.6, basta mostrar que $dom(T) \cap dom(B_k)^0 \neq \phi$. De fato, no momento

$$x^k \in dom(T) \cap C^0 \cap S_k^0 = dom(T) \cap dom(B_k)^0.$$

Assim, $dom(T) \cap dom(B_k)^0 \neq \phi$. Logo, A_k é monótono maximal. Observe que $dom(A_k)$ é um subconjunto de $dom(B_k)$. Como $dom(B_k)$ é limitado então $dom(A_k)$ é limitado. Segue, do item (ii) da Proposição 2.2.6, que A_k é operador sobrejetivo.

Logo, existe $y \in dom(A_k) = dom(T) \cap dom(B_k) \subset dom(T) \cap C^0$, tal que

$$0 \in T(y) + N_k(y) + \lambda_k \partial_x D_g(y, x^k). \tag{3.6}$$

Afirmamos que $y \in S_k^0$. De fato. Se $y = x^k$, é imediato que $y \in S_k^0$. Caso contrário sejam u^k, w^k e v^k os elementos em $I\!R^n$ tais que $u^k \in T(y)$, $w^k \in N_k(y)$, $v^k \in \lambda_k \partial_x D_g(y, x^k)$ e

$$0 = u^k + w^k + v^k \tag{3.7}$$

Como $y \in dom(A_k) \subset dom(B_k) \subset dom(\partial_x D_g(\cdot, x^k)) = C^0$, basta mostrar, devido a (3.5), que

$$D_g(y, x^k) \leq \frac{Gap_{\,T,C}(x^k)}{\lambda_k}. \tag{3.8}$$

Como $D(\cdot, x^k)$ é estritamente convexa, $0 = D_g(x^k, x^k) > D_g(y, x^k) + \langle v, x^k - y\rangle$, onde $v \in \partial_x D_g(y, x^k)$ tal que $v^k = \lambda_k v$. Logo,

$$0 > D_g(y, x^k) + \frac{1}{\lambda_k}\langle v^k, x^k - y\rangle. \tag{3.9}$$

De (3.7) e (3.9),

$$\frac{1}{\lambda_k}\left[\langle u^k, x^k - y\rangle + \langle w^k, x^k - y\rangle\right] > D_g(y, x^k). \tag{3.10}$$

Como $w^k \in N_k(y)$ e $x^k \in S_k$ então, da Observação 2.2.5, segue

$$\langle w^k, x^k - y\rangle \leq 0. \tag{3.11}$$

De (3.11) e (3.10),

$$0 \leq \lambda_k D_g(y, x^k) < \langle u^k, x^k - y\rangle. \tag{3.12}$$

Como $\langle u^k, x^k - y\rangle \le sup\{\langle v, x^k - y\rangle \,/\, z \in C \cap dom(T)\,,\, v \in T(z)\} = Gap_{\,T,C}(x^k)$, então de (3.12) segue

$$D_g(y, x^k) \le \frac{Gap_{\,T,C}(x^k)}{\lambda_k}\,.$$

Logo, $y \in S_k^{\,0}$.

Pela Observação 2.2.3, $N_k(y) = \{0\}$. Assim, $w^k = 0$. De (3.7), $0 = u^k + v^k$ onde $u^k \in T(y)$ e $v^k \in \lambda_k \partial_x D_g(y, x^k)$. Logo, $0 \in T_k(y)$. Como T_k é estritamente monótono então y é o único zero de T_k. De (3.3), $y = x^{k+1}$. Como $y \in C^{\,0}$ então $x^{k+1} \in C^{\,0}$. A indução está completa.

Portanto, $\{x^k\}$ está bem definida e contida em $C^{\,0}$. ∎

Para estabelecer a convergência de $GPPA$, dada pelas relações (3.2) e (3.3), é preciso introduzir a seguinte definição.

Definição 3.2.1 *Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador tal que $dom(T)$ seja convexo e fechado. Dizemos que T é pseudomonótono se, e somente se, satisfaz a seguinte condição:*

Considere uma seqüência $\{x^k\}_{k\in I\!N} \subset dom(T)$ tal que $\lim\limits_{k\to+\infty} x^k = x^0 \in dom(T)$ e para qualquer seqüência $\{w^k\}_{k\in I\!N}$, com $w^k \in T(x^k)$ para todo $k \in I\!N$, tem-se

$$\limsup_k \langle w^k, x^k - x^0\rangle \le 0.$$

Então para todo $y \in dom(T)$ existe um elemento $w^0 \in T(x^0)$ tal que

$$\langle w^0, x^0 - y\rangle \le \liminf_k \langle w^k, x^k - y\rangle.$$

Proposição 3.2.1 *Seja $T : I\!R^n \to I\!R^n$ um operador monótono contínuo. Então T é pseudomonótono.*

Demonstração. Seja $\{x^k\}_{k\in I\!N}$ uma seqüência em $dom(T)$ tal que $\lim\limits_{k\to+\infty} x^k = x^0 \in dom(T)$. Como T é contínuo, $\lim\limits_{k\to+\infty}\langle T(x^k), x^k - x^0\rangle = \langle T(x^0), x^0 - x^0\rangle = 0$. Assim, $\limsup\limits_k \langle T(x^k), x^k - x^0\rangle = 0$. Seja $y \in dom(T)$. Então

$$\lim_{k\to+\infty}\langle T(x^k), x^k - y\rangle = \langle T(x^0), x^0 - y\rangle.$$

Logo, $\langle T(x^0), x^0 - y\rangle = \liminf\limits_k \langle T(x^k), x^k - y\rangle$. Portanto, T é pseudomonótono. ∎

Proposição 3.2.2 *Seja $T = \partial f : I\!\!R^n \to \mathcal{P}(I\!\!R^n)$, para alguma função convexa f. Então T é pseudomonótono.*

Demonstração. Seja $\{x^k\}_{k \in I\!\!N}$ uma seqüência em $dom(\partial f)$ tal que $\lim_{k \to +\infty} x^k = x^0 \in dom(\partial f)$ e $\limsup_k \langle w^k, x^k - x^0 \rangle \leq 0$, onde $w^k \in \partial f(x^k)$.

Seja $y \in dom(\partial f)$. Pela Definição A.6.1, $\langle w^k, y - x^k \rangle \leq f(y) - f(x^k)$, para todo $k \in I\!\!N$. Como f é convexa então f é contínua (veja Teorema A.14). Logo,

$$\liminf_k \langle T(x^k), x^k - y \rangle \geq \liminf_k \left(f(x^k) - f(y) \right) = f(x^0) - f(y)$$

Como f é convexa, existe $w^0 \in \partial f(x^0)$ tal que $f(y) \geq f(x^0) + \langle w^0, y - x^0 \rangle$ (veja Teorema A.11). Assim,

$$\langle w^0, x^0 - y \rangle \leq f(x^0) - f(y).$$

Logo, existe $w^0 \in \partial f(x^0)$ tal que

$$\langle w^0, x^0 - y \rangle \leq \liminf_k \langle w^k, x^k - y \rangle.$$

Portanto, ∂f é pseudomonótono. ■

Proposição 3.2.3 *Sejam T, $S : I\!\!R^n \to \mathcal{P}(I\!\!R^n)$ operadores pseudomonótonos, com $dom(T) \cap dom(S) \neq \phi$. Então $T + S$ é pseudomonótono.*

Demonstração. Seja $\{x^k\}_{k \in I\!\!N}$ uma seqüência em $dom(T+S)$ tal que $\lim_{k \to +\infty} x^k = x^0 \in dom(T+S)$ e

$$\limsup_k \langle w^k, x^k - x^0 \rangle \leq 0, \tag{3.13}$$

onde $w^k \in \partial f(x^k)$.

Seja $y \in dom(T+S)(x^k) = T(x^k) + S(x^k)$. Logo, existem $w_T^k \in T(x^k)$ e $w_S^k \in S(x^k)$ tais que $w^k = w_T^k + w_S^k$. De (3.13), $\limsup_k \langle w_T^k + w_S^k, x^k - x^0 \rangle \leq 0$. Então

$$\limsup_k \langle w_T^k, x^k - x^0 \rangle \leq \limsup_k \langle w_T^k + w_S^k, x^k - x^0 \rangle \leq 0,$$

$$\limsup_k \langle w_S^k, x^k - x^0 \rangle \leq \limsup_k \langle w_T^k + w_S^k, x^k - x^0 \rangle \leq 0.$$

Como T e S são pseudomonótonos, existem $w_T^0 \in T(x^0)$ e $w_S^0 \in S(x^0)$ tais que

$$\langle w_T^0, x^0 - y \rangle \leq \liminf_k \langle w_T^k, x^k - x^0 \rangle, \quad \langle w_S^0, x^0 - y \rangle \leq \liminf_k \langle w_S^k, x^k - x^0 \rangle,$$

para todo $y \in dom(T+S)$. Logo, $\langle w_T^0 + w_S^0, x^0 - y\rangle \leq \liminf_k \langle w_T^k + w_S^k, x^k - x^0\rangle$.

Seja $w^0 = w_T^0 + w_S^0 \in T(x^0) + S(x^0) = (T+S)(x^0)$. Então

$$\langle w^0, x^0 - y\rangle \leq \liminf_k \langle w^k, x^k - x^0\rangle,$$

para todo $y \in dom(T+S)$. Portanto $T+S$ é pseudomonótono. ■

Teorema 3.1 *Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono maximal com domínio fechado. Considere o problema $VIP(T,C)$, onde $C \subset I\!R^n$ é convexo e fechado. Seja g uma função de Bregman com zona C^0, isto é, g satisfaz* H1*,* H3*,* B1*,* B2 *e* B3*. Suponha que as seguintes condições são válidas:*

i) *$dom(T) \cap C^{,0} \neq \phi$.*

ii) *$VIP(T,C)$ possui soluções.*

iii) *T é pseudomonótono.*

iv) *$\lambda_k \in (0, \tilde{\lambda}]$, para algum $\tilde{\lambda} > 0$.*

v) *Ou,*

v1) *g é zona coerciva, ou*

v2) *g é coerciva na fronteira e $Gap_{T,C}(x) < +\infty$, $\forall x \in C \cap dom(T)$.*

Então a seqüência $\{x^k\}$, gerada por (3.2) e (3.3), satisfaz:

a) *A seqüência $\{D_g(z, x^k)\}$ é não crescente, $\forall z \in S(T,C)$.*

b) *$\{x^k\}$ é limitada e possui pontos de acumulação.*

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

d) *Se $\bar{x}$ é um ponto de acumulação de $\{x^k\}$ então $\exists\ \bar{u} \in T(\bar{x})$ tal que $\langle \bar{u}, x^* - \bar{x}\rangle = 0$, $\forall x^* \in S(T,C)$.*

Demonstração.

a) Afirma-se que para todo $z \in S(T,C)$,

$$D_g(z, x^{k+1}) \leq D_g(z, x^k) - D_g(x^{k+1}, x^k). \tag{3.14}$$

De fato. Observe que a existência de z está garantida pela condição (ii). De (3.3) e pela Proposição 2.2.5,

$$0 \in T_k(x^{k+1}) = T(x^{k+1}) + \lambda_k \partial_x D_g(x^{k+1}, x^k) = T(x^{k+1}) + \lambda_k \left\{\nabla g(x^{k+1}) - \nabla g(x^k)\right\}.$$

Logo, existe $u^k \in T(x^{k+1})$ tal que

$$u^k = \lambda_k(\nabla g(x^k) - \nabla g(x^{k+1})). \tag{3.15}$$

Pelo item (i) da Proposição 3.1.1, $D_g(x,y) - D_g(x,z) - D_g(z,y) = \langle \nabla g(y) - \nabla g(z), z - x \rangle$, para todo $x \in C$ e todo $y, z \in C^0$. Tomando $y = x^k$, $z = x^{k+1}$ e $x = y$, tem-se

$$\langle \nabla g(x^k) - \nabla g(x^{k+1}), x^{k+1} - y \rangle = D_g(y, x^k) - D_g(y, x^{k+1}) - D_g(x^{k+1}, x^k). \tag{3.16}$$

De (3.15) e (3.16),

$$\langle u^k, x^{k+1} - y \rangle = \lambda_k \left[D_g(y, x^k) - D_g(y, x^{k+1}) - D_g(x^{k+1}, x^k) \right]. \tag{3.17}$$

Seja $z \in S(T, C)$ e tome $v^* \in T(z)$ tal que

$$\langle v^*, x - z \rangle \geq 0, \forall x \in C. \tag{3.18}$$

Como T é monótono, $\langle u^k - v^*, x^{k+1} - z \rangle \geq 0$. Logo, de (3.18),

$$\langle u^k, x^{k+1} - z \rangle \geq \langle v^*, x^{k+1} - z \rangle \geq 0. \tag{3.19}$$

Tomando $y = z$ em (3.17) e usando (3.19),

$$D_g(z, x^k) - D_g(z, x^{k+1}) - D_g(x^{k+1}, x^k) \geq 0.$$

Logo, a desigualdade (3.14) é válida.

De (3.14), para todo $z \in S(T, C)$, tem-se que

$$D_g(z, x^{k+1}) \leq D_g(z, x^k). \tag{3.20}$$

Portanto, a seqüência $\left\{ D_g(z, x^k) \right\}$ é não crescente, $\forall z \in S(T, C)$.

b) De (3.20), para $z \in S(T, C)$ fixo, segue que

$$\cdots \leq D_g(z, x^k) \leq D_g(z, x^{k-1}) \leq \cdots \leq D_g(z, x^1) \leq D_g(z, x^0). \tag{3.21}$$

Logo, $\left\{ x^k \right\}$ está contida no conjunto de nível $\{x \in C \;/\; D_g(z, x) \leq D_g(z, x^0)\}$. Pela hipótese B1, este conjunto de nível é limitado. Portanto, $\left\{ x^k \right\}$ é limitada e possui pontos de acumulação.

c) Pela desigualdade (3.14), tem-se que

$$0 \leq D_g(x^{k+1}, x^k) \leq D_g(z, x^k) - D_g(z, x^{k+1}), \tag{3.22}$$

para todo $z \in S(T, C)$.

Pelo item (a), $\{D_g(z, x^k)\}$ é não crescente. De (3.21), $\{D_g(z, x^k)\}$ é limitada. Logo, $\{D_g(z, x^k)\}$ é convergente.

Seja $\alpha = \lim_{k \to +\infty} D_g(z, x^k)$, onde $\alpha > 0$. Tomando limite em (3.22), quando k tende para $+\infty$, segue que

$$0 \leq \lim_{k \to +\infty} D_g(x^{k+1}, x^k) \leq \lim_{k \to +\infty} D_g(z, x^k) - \lim_{k \to +\infty} D_g(z, x^{k+1}) = \alpha - \alpha = 0.$$

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

d) Suponha agora que $\{x^k\}$ possui pontos de acumulação. Observe que todo ponto de acumulação de $\{x^k\}$ está em C, pois C é fechado. Seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$ e $\{x^{k_j}\}_{j \in I\!N}$ uma subseqüência de $\{x^k\}$ que converge para $\bar{x}$.

Em (3.17), tomando $y = \bar{x}$ e $k = k_j$, tem-se

$$\langle u^{k_j}, x^{k_j+1} - \bar{x} \rangle = \lambda_{k_j} \left[D_g(\bar{x}, x^{k_j}) - D_g(\bar{x}, x^{k_j+1}) - D_g(x^{k_j+1}, x^k_j) \right]. \tag{3.23}$$

Observe que

d1) $\{x^{k_j+1}\}$ é limitada, pelo item (b).

d2) $\lim_{j \to +\infty} x^{k_j} = \bar{x}$, pela definição de $(x^{k_j})_{j \in I\!N}$.

d3) $\lim_{k \to +\infty} D_g(x^{k_j+1}, x^{k_j}) = 0$, pelo item (c).

Então, aplicando a hipótese B3 para as seqüências $\{x^{k_j+1}\}$ e $\{x^{k_j}\}$, temos

$$\lim_{j \to +\infty} x^{k_j+1} = \bar{x}. \tag{3.24}$$

A vista de (3.24), (d2) e (d3), é possível aplicar a hipótese B2 para as seqüências $\{x^{k_j+1}\}$ e $\{x^{k_j}\}$, isto é,

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

Assim,

$$\lim_{k \to +\infty} \left(D_g(\bar{x}, x^{k_j+1}) - D_g(\bar{x}, x^{k_j}) \right) = 0. \tag{3.25}$$

Segue do item (iv) que $\{\lambda_k\}_{k\in I\!N}$ é limitado. Logo, usando (d3) e (3.25) em (3.23), temos que

$$\lim_{j\to+\infty}\langle u^{k_j}, x^{k_j+1}-\bar{x}\rangle = 0. \tag{3.26}$$

Agora é possível usar a pseudomonotonicidade do operador T. De (3.26),

$$\limsup_j \langle u^{k_j}, x^{k_j+1}-\bar{x}\rangle = 0,$$

onde $\bar{x} = \lim\limits_{j\to+\infty} x^{k_j+1}$ e $u^{k_j} \in T(x^{k_j+1})$, para todo $j \in I\!N$. Pela Definição 3.2.1, para todo $z \in S(T,C)$ existe $\bar{u} \in T(\bar{x})$ tal que

$$\langle \bar{u}, \bar{x}-z\rangle \leq \liminf_j \langle u^{k_j}, x^{k+1}-z\rangle. \tag{3.27}$$

Tomando $y = z$ em (3.17),

$$\langle u^k, x^{k+1}-z\rangle = \lambda_k\left[D_g(z,x^k) - D_g(z,x^{k+1}) - D_g(x^{k+1},x^k)\right]. \tag{3.28}$$

Pelo item (a), $\{D_g(z,x^k)\}$ é decrescente. Logo, como $\{\lambda_k\}$ é limitado então, em (3.28), temos

$$\lim_{k\to+\infty}\langle u^k, x^{k+1}-z\rangle = \lim_{k\to+\infty}\lambda_k\left[D_g(z,x^k) - D_g(z,x^{k+1}) - D_g(x^{k+1},x^k)\right] = 0.$$

Assim, $\liminf\limits_j \langle u^{k_j}, x^{k+1}-z\rangle = 0$. De (3.27),

$$\langle \bar{u}, \bar{x}-z\rangle \leq 0, \tag{3.29}$$

com $\bar{u} \in T(\bar{x})$. Como $z \in S(T,C)$, existe $v^* \in T(z)$ tal que $\langle v^*, y-z\rangle \geq 0$ para todo $y \in C$. Além disso, pela monotonicidade de T, $\langle \bar{u}-v^*, \bar{x}-z\rangle \geq 0$. Logo,

$$\langle \bar{u}, \bar{x}-z\rangle \geq \langle v^*, \bar{x}-z\rangle \geq 0. \tag{3.30}$$

De (3.29) e (3.30) segue que $\langle \bar{u}, \bar{x}-z\rangle = 0$ e (d) é estabelecido. ■

Teorema 3.2 *Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono maximal . Considere o problema $VIP(T,C)$, onde $C \subset I\!R^n$ é convexo e fechado. Seja g uma função de Bregman com zona C^0, isto é, g satisfaz* H1, H3, B1, B2 *e* B3. *Suponha que as seguintes condições são válidas:*

i) $dom(T) \cap C^{,0} \neq \phi$.

ii) $S(T,C) \neq \phi$.

iii) *T é pseudomonótono em $dom(T)$.*

iv) *$\lambda_k \in (0, \tilde{\lambda}]$, para algum $\tilde{\lambda} > 0$.*

v) *Ou,*

v1) *g é zona coerciva, ou*

v2) *g é coerciva na fronteira e $Gap_{T,C}(x) < +\infty$, $\forall x \in C \cap dom(T)$.*

vi) *T é paramonótono em C.*

Então a seqüência $\{x^k\}$, gerada por (3.2) e (3.3), converge para uma solução $\bar{x}$ de $VIP(T, C)$.

Demonstração. Pelo item (b) do Teorema 3.1, $\{x^k\}$ possui pontos de acumulação. Seja $\bar{x}$ um ponto de acumulação de $\{x^k\}$. Pelo item (d), do Teorema 3.1, existe $\bar{u} \in T(\bar{x})$ tal que $\langle \bar{u}, x^* - \bar{x} \rangle = 0$, $\forall x^* \in S(T, C)$. Como T é paramonótono então, pela Proposição 1.3.1, $\bar{x} \in S(T, C)$.

Portanto, $\{x^k\}$ gerada por (3.2) e (3.3)converge para uma solução $\bar{x}$ de $VIP(T, C)$.

■

3.3 $GPPA$ para $VIP(T, C)$ num Conjunto Poliedral

Considere o problema $VIP(T, C)$ para o caso

$$C = \{x \in \mathbb{R}^n \;/\; Ax = b,\; x \geq 0\}, \tag{3.31}$$

onde $A \in \mathbb{R}^{m \times n}$ e $b \in \mathbb{R}^n$. Neste caso, não é possível aplicar os resultados obtidos na seção 3.2 pois $C^0 = \phi$.

A idéia, para dar solução a este problema, é desenvolver uma aproximação e transferir as restrições lineares para o operador.

A aproximação consiste em substituir o problema $VIP(T, C)$ pelo problema $VIP(\widehat{T}, \mathbb{R}^n_+)$, onde $\widehat{T} = T + N_E$, para algum operador monótono contínuo T, e

$$E = \{x \in \mathbb{R}^n \;/\; Ax = b\}. \tag{3.32}$$

Da Proposição 2.2.1, temos

$$S(\widehat{T}, \mathbb{R}^n_+) = S(T, E \cap \mathbb{R}^n_+) = S(T, C). \tag{3.33}$$

Por outro lado, da Proposição 2.2.2, segue

$$N_E(x) = \begin{cases} Im(A^t) & , \quad \text{se } x \in E, \\ \phi & , \quad \text{em outro caso.} \end{cases} \tag{3.34}$$

Considere uma função de Bregman g com zona $\mathbb{R}^n_+$ e coerciva na zona (veja H4). Seja $\{x^k\}$ a seqüência gerada por (3.2) e (3.3). Então x^{k+1} satisfaz

$$
\begin{aligned}
0 \in T_k(x^{k+1}) &= (T+N_E)(x^{k+1}) + \lambda_k \partial_x D_g(x^{k+1}, x^k) \\
&= T(x^{k+1}) + N_E(x^{k+1}) + \lambda_k \left\{\nabla g(x^{k+1}) - \nabla g(x^k)\right\} \\
&= T(x^{k+1}) + Im(A^t) + \lambda_k \left\{\nabla g(x^{k+1}) - \nabla g(x^k)\right\}.
\end{aligned}
$$

Logo, existe $y^{k+1} \in Im(A^t)$ tal que

$$
T(x^{k+1}) - A^t(y^{k+1}) = \lambda_k \left\{\nabla g(x^{k+1}) - \nabla g(x^k)\right\}. \tag{3.35}
$$

Além disso, x^{k+1} satisfaz

$$
A(x^{k+1}) - b = 0. \tag{3.36}
$$

Resumindo, o problema é procurar x^{k+1} para dar solução ao seguinte sistema de $n+m$ equações e $n+m$ variáveis (as variáveis são x_j e y_i com $1 \leq j \leq n$, $1 \leq i \leq m$):

$$
\left\{
\begin{array}{rcl}
T(x) + \lambda_k \nabla g(x) - A^t(y) &=& \lambda_k \nabla g(x^k), \\
Ax &=& b.
\end{array}
\right. \tag{3.37}
$$

Observe que x^{k+1} é unicamente determinado pelo sistema (3.37). Segue do Lema 3.1, que isto não acontece para y^{k+1}. Mas se A possui posto máximo, então y^{k+1} é unicamente determinado, pois neste caso (3.35) pode ser escrito como

$$
T(x^{k+1}) + \lambda_k \left\{\nabla g(x^{k+1}) - \nabla g(x^k)\right\} \in Im(A^t). \tag{3.38}
$$

As relações (3.35) e (3.36) definem o chamado *"Algoritmo Primal - Factível"*. O seguinte corolário apresenta o resultado do Teorema (3.2) para este caso específico.

Corolário 3.3 *Seja $\widehat{T} = T + N_E$, onde $T : I\!R^n \to I\!R^n$ é um operador monótono contínuo e $E = \{x \in I\!R^n \ / \ Ax = b\}$, $A \in R^{m \times n}$, $b \in I\!R^m$. Considere $VIP(T, C)$ com $C = \{x \in I\!R^n \ / \ Ax = b, \ x > 0\}$. Suponha que:*

i) $S(T, C) \neq \phi$.

ii) *Existe $x > 0$ tal que $Ax = b$.*

iii) $\lambda_k \in (0, \tilde{\lambda}]$, *para algum $\tilde{\lambda} > 0$.*

iv) *g é função de Bregman com zona $I\!R^n_{++}$ e zona coerciva.*

v) *T é paramonótono em $I\!R^n_+$.*

Então a seqüência $\{x^k\}$, gerada por $GPPA$, converge para uma solução $\bar{x}$ de $VIP(T, C)$.

Demonstração. Considere o algoritmo dado por (3.2) e (3.3) para o problema $VIP(\widehat{T}, I\!R^n_+)$. Observe que $dom(\widehat{T}) = dom(T+N_E) = dom(T) \cap dom(N_E) = I\!R^n \cap E = E$. Logo, do item (ii), segue $dom(\widehat{T}) \cap I\!R^n_+ \neq \phi$.

Pela Proposição 3.2.1, T é pseudomonótono. Também, N_E é pseudomonótono (segue da Proposição 3.2.2). Logo, $\widehat{T} = T + N_E$ é pseudomonótono (segue da Proposição 3.2.3).

Assim, as hipóteses do Teorema 3.2 são satisfeitas. Logo, a seqüência $(x^k)_{k \in I\!N \cup \{0\}}$, gerada por (3.35) e (3.36), converge para $\bar{x} \in S(\widehat{T}, I\!R^n_+)$.

De (3.33), a seqüência $\{x^k\}$ gerada por (3.35) e (3.36) converge para $\bar{x} \in S(T, C)$.

■

3.4 Comportamento do limite da seqüência $GPPA$ num Conjunto Poliedral

O objetivo desta seção, com as considerações feitas na seção 3.3 e com hipóteses adequadas, é mostrar que o limite da seqüência $\{x^k\}$, gerada pelo algoritmo primal - factível, isto é, a seqüência $GPPA$ num conjunto poliedral gerada por (3.35) e (3.36), é o centro analítico do conjunto de soluções do problema de inequações variacionais com relação a uma certa função barreira.

Sejam C e E os conjuntos definidos em (3.31) e (3.32).

Lema 3.3 *Seja $\widehat{T} = T + N_E$, onde $T : I\!R^n \to I\!R$ é um operador monótono continuamente diferenciável. Além disso, suponha que existe um subespaço W tal que $Ker(J_{\widehat{T}}(x)) = Ker(J_{\widehat{T}}(x)^s) = W$, para todo $x \in C \cap E$. Seja $\widehat{x} \in C$ fixo e denote $B = J_T(\widehat{T})$. Considere $\{x^k\}$ como a seqüência gerada por (3.35) e (3.36). Então, para cada k, x^k é solução do problema:*

$$
\begin{aligned}
& min \quad D_g(x, x^0) \\
s.a. \quad & Bx = Bx^k\,, && (3.39) \\
& T(\widehat{x})^t x = T(\widehat{x})^t x^k\,, && (3.40) \\
& Ax = Ax^k\,, && (3.41) \\
& x \geq 0\,. && (3.42)
\end{aligned}
$$

Demonstração. As condições de otimalidade de KKT (suficientes, pois $D_g(\cdot, x^0)$ é convexa) são dadas pelas relações (3.39), (3.40), (3.41), (3.42), e em adição, existência de $u^k \in I\!R^n$, $w^k \in I\!R^m$ e $\eta_k \in I\!R$ tais que

$$\nabla g(x) - \nabla g(x^0) + B^t u^k + A^t w^k + \eta_k T(\widehat{x}) \geq 0, \tag{3.43}$$

$$x^t \left[\nabla g(x) - \nabla g(x^0) + B^t u^k + A^t w^k + \eta_k T(\widehat{x}) \right] = 0. \tag{3.44}$$

Observe que x^k satisfaz trivialmente (3.39), (3.40) e (3.41), para todo k. Pelo Lema 3.1, x^k também satisfaz (3.42), para todo k.

Para mostrar que (3.43) e (3.44) são válidas para x^k, procuremos u^k, w^k e η_k tais que (3.43) é válido como igualdade e de forma imediata (3.44) também é válido.

De (3.35), para todo $\ell \in I\!N \cup \{0\}$, existe $y^{\ell+1} \in I\!R^m$ tal que

$$\nabla g(x^{\ell+1}) - \nabla g(x^\ell) + \lambda_\ell^{-1} [\, T(x^{\ell+1}) - A^t(y^{\ell+1}) \,] = 0. \tag{3.45}$$

Pela hipótese, existe um subespaço W tal que $Ker(B) = Ker(B^s) = W$. Logo, pelo item (i) da Proposição 1.3.3, segue

$$T(x^{\ell+1}) - T(x^\ell) \in Ker(B)^\perp = Im(B^t). \tag{3.46}$$

Assim, existe $v^\ell \in I\!R^n$ tal que

$$T(x^{\ell+1}) = T(x^\ell) + B^t v^\ell. \tag{3.47}$$

Substituindo (3.47) em (3.45), e somando de $\ell = 0$ até $\ell = k-1$, temos

$$\begin{aligned} \sum_{\ell=0}^{k-1} \left\{ \nabla g(x^{\ell+1}) - \nabla g(x^\ell) \right\} + \left\{ \sum_{\ell=0}^{k-1} \lambda_\ell^{-1} \right\} T(\widehat{x}) + B^t \left\{ \sum_{\ell=0}^{k-1} \lambda_\ell^{-1} v^\ell \right\} & \\ + A^t \left\{ \sum_{\ell=0}^{k-1} -\lambda_\ell^{-1} y^{\ell+1} \right\} = 0 & \end{aligned} \tag{3.48}$$

Defina

$$u^k = \sum_{\ell=0}^{k-1} \lambda_\ell^{-1} v^\ell \in I\!R^n, \quad w^k = \sum_{\ell=0}^{k-1} -\lambda_\ell^{-1} y^{\ell+1} \in I\!R^m, \quad \eta_n = \sum_{\ell=0}^{k-1} \lambda_\ell^{-1} \in I\!R.$$

Então, em (3.48), temos

$$\nabla g(x^k) - \nabla g(x^0) + \eta_k T(\widehat{x}) + B^t u^k + A^t w^k = 0, \tag{3.49}$$

que é a relação (3.43) para x^k como igualdade. Logo, x^k também satisfaz (3.44). Assim, as condições de KKT são satisfeitas para x^k. O resultado está estabelecido. ■

O seguinte fato pode ser encontrado em Hoffman [10].

Proposição 3.4.1 *Sejam $H \in I\!R^{s\times n}$, $p^k \in I\!R^s$ e $p \in I\!R^s$. Suponha que $p = \lim_{k\to+\infty} p^k$. Considere $L_k = \left\{x \in I\!R^r \ / \ Hx \le p^k\right\}$ e $L = \{x \in I\!R^r \ / \ Hx \le p\}$. Se $L \ne \phi$ e $L_k \ne \phi$, para todo $k \in I\!N$, então para todo $x \in L$ existe $\widetilde{x}^k \in L_k$ tal que $x = \lim_{k\to+\infty} \widetilde{x}^k$.*

Seja $\bar{x}$ o limite da seqüência $\left\{x^k\right\}$ gerada por (3.35) e (3.36). Observe que a existência de $\bar{x}$ é garantida pelo Corolario 3.3.

Finalmente, segue o resultado mais importante desta seção.

Teorema 3.4 *Sejam $A \in R^{m\times n}$, $b \in I\!R^m$. Considere $E = \{x \in I\!R^n \ / \ Ax = b\}$, e $C = \{x \in I\!R^n \ / \ Ax = b,\ x > 0\}$. Suponha que*

i) *$\widehat{T} = T + N_E$, onde $T : I\!R^n \to I\!R^n$ é um operador monótono continuamente diferenciável.*

ii) *Existe um subespaço W tal que $Ker(J_T(x)) = Ker(J_T(x)^s) = W$, para todo $x \in C \cap V$.*

iii) *$\lambda_k \in (0, \tilde{\lambda}]$, para algum $\tilde{\lambda} > 0$.*

iv) *g é função de Bregman com zona $I\!R^n_{++}$ e coerciva na zona.*

v) *$S(T, C) \ne \phi$.*

Então $\bar{x} = \lim_{k\to+\infty} x^k$ é solução do problema

$$\left\{ \min_{x\in S(T,C)} D_g(x, x^0), \right.$$

onde $\left\{x^k\right\}$ é a seqüência gerada por (3.35) e (3.36).

Demonstração. Considere qualquer $\widehat{x} \in C$ e seja $B = J_T(\widehat{x})$. Pelo Corolário 3.3, segue que $\bar{x} \in S(T, C)$. Pela hipótese (ii) e pelo item (iii) da Proposição 1.2.2, T é paramonótono em $I\!R^n_{++}$.

Além disso, pela hipótese (ii), é possível aplicar a Proposição 1.3.3 com $\tilde{x} = \bar{x}$. Assim, o problema

$$\left\{ \min_{x\in S(T,C)} D_g(x, x^0), \right.$$

equivale ao problema

$$\begin{aligned} & min \quad D_g(x, x^0) \\ s.a. \qquad & Bx = B\bar{x}\,, && (3.50) \\ & T(\widehat{x})^t x = T(\widehat{x})^t \bar{x}\,, && (3.51) \\ & Ax = A\bar{x}\,, && (3.52) \\ & x \geq 0\,. && (3.53) \end{aligned}$$

Sejam L e L_k os conjuntos de vetores satisfazendo (3.50)-(3.53) e (3.39)-(3.42), respectivamente. Isto é,

$$L = \left\{x \in I\!R^n \;/\; Bx = B\bar{x}\,, T(\widehat{x})^t x = T(\widehat{x})^t \bar{x}\,, Ax = A\bar{x}\,, x \geq 0\right\},$$

$$L_k = \left\{x \in I\!R^n \;/\; Bx = Bx^k\,, T(\widehat{x})^t x = T(\widehat{x})^t x^k\,, Ax = Ax^k\,, x \geq 0\right\}.$$

Seja $\left\{x^{k_j}\right\}_{j \in I\!N}$ uma subseqüência de $\left\{x^k\right\}$ tal que $\bar{x} = \lim\limits_{j \to +\infty} x^{k_j}$. É claro que L e L_k podem ser escritos como na Proposição 3.4.1, para algum H, p^k e p adequados, e tal que $\lim\limits_{k \to +\infty} p^{k_j} = p$. Além disso, L e L_k são não vazios pois $\bar{x} \in L$ e $x^k \in L_k$.

Seja agora $x \in S(T, C)$; isto é, $x \in L$. Logo, pela Proposição 3.4.1, existe $\widetilde{x}^{k_j} \in L_{k_j}$ tal que $x = \lim\limits_{j \to +\infty} \widetilde{x}^{k_j}$.

Pelo Lema 3.3,

$$D_g(x^{k_j}, x^0) \leq D_g(\widetilde{x}^{k_j}, x^0). \tag{3.54}$$

Logo, pela continuidade de $D_g(\cdot, x^0)$, segue

$$D_g(\bar{x}, x^0) = \lim_{j \to +\infty} D_g(x^{k_j}, x^0) \leq \lim_{j \to +\infty} D_g(\widetilde{x}^{k_j}, x^0) = D_g(x, x^0).$$

Portanto, como $x \in S(T, C)$ é arbitrário, segue que $\bar{x}$ é solução do problema $min\,\{D_g(x, x^0) \;/\; x \in S(T, C)\}$. ■

3.5 Relações entre Trajetória Central e $GPPA$ em Programação Linear

Nesta seção é feita uma ligação entre algoritmo de ponto proximal generalizado e trajetória central para problemas de programação linear, quando a função barreira é dada pela distância de Bregman.

O primeiro resultado mostra que a seqüência $GPPA$, com distância de Bregman como função barreira, e a trajetória central convergem para o mesmo ponto em problemas de programação linear.

Corolário 3.5 *Suponha que* T *satisfaz* H8 *e* H9*,* g *satisfaz* H1, B1, B2, B3*, é zona coerciva (veja* H4*) e finita na fronteira (veja* H3*); e* $VIP(T, C)$ *é regular e possui soluções (veja* H10 *e* H11*, respectivamente). Sejam* $\{x^k\}$ *a seqüência gerada por* $GPPA$ *com função de Bregman* g *começando em* $x^0 \in dom(T) \cap C$ *e* $\{x(\mu)/\mu > 0\}$ *a trajetória central com barreira* $h(x) = D_g(x, x^0)$*, para todo* $x \in C$*. Então* $\lim_{k \to +\infty} x^k$ *e* $\lim_{\mu \to +\infty} x(\mu)$ *coincidem e é a única solução do problema*

$$\left\{ \min_{x \in S(T,C)} D_g(x, x^0), \right.$$

isto é, eles são o centro analítico de $S(T, C)$ *com respeito à barreira* h*.*

Demonstração. O resultado para $\{x^k\}$ segue do Teorema 3.4, e o resultado para $\{x(\mu)/\mu > 0\}$ segue do item (i) do Teorema 2.1, pois $g(\cdot)$ e $D_g(\cdot, x^0)$ diferem pelo termo afim $g(x^0) + \langle \nabla g(x^0), x - x^0 \rangle$, e H1, H3, H4 são invariantes por adições de funções afins; deste modo que h também satisfaz H1, H3 e H4.

Além disso, neste caso H12 também é válido pois a função $h(x) = D_g(x, x^0)$, $x \in C$, atinge seu mínimo em $x^0 \in dom(T) \cap C^0$.

Portanto, $\lim_{k \to +\infty} x^k = \lim_{\mu \to +\infty} x(\mu)$ é o centro analítico de $S(T, C)$ com respeito à barreira h. ■

O Corolário 3.5 origina a seguinte pergunta:

A seqüência $\{x^k\}$*, gerada por* $GPPA$*, está contida na trajetória central* $\{x(\mu)/\mu > 0\}$ *com barreira* $h(x) = D_g(x, x^0)$*?*

A resposta é afirmativa quando o operador $\widehat{T}$ é constante e V é uma variedade afim (por exemplo, para problemas de programação linear), e negativo para outro caso.

O resultado da seguinte proposição é importante para mostrar as afirmações feitas anteriormente.

Proposição 3.5.1 *Suponha que* H11 *é válido, que existem* $\alpha > 0$ *e* $r \in \mathbb{R}$ *tais que para todo* $y \in dom(T) \cap C$*, com* $\| y \| > \alpha$*, é válido*

$$\langle v, y - x \rangle \geq -r, \tag{3.55}$$

para todo $v \in T(y)$*. Então* $Gap_{\,T,C}(x) < +\infty$*, para todo* $x \in dom(T) \cap C$*, isto é,* H13 *é válido.*

Demonstração. Como H11 é válido, para todo y em $S(T, C)$ existe $v \in T(y)$ tal que $\langle v, x - y \rangle \geq 0$, para todo $x \in dom(T) \cap C$. De (3.55), $0 \leq \langle v, x - y \rangle \leq r$, para todo $x \in dom(T) \cap C$, para todo $y \in dom(T) \cap C$ com $\| y \| > \alpha$ e todo $v \in T(y)$.

Seja $u \in T(x)$. Como T é monótono, $0 \leq \langle v, x - y \rangle \leq \langle u, x - y \rangle$. Logo, para $x \in dom(T) \cap C$, temos

$$\begin{aligned} 0 \leq Gap_{\,T,C}(x) &= \sup_{y \in\, dom(T) \cap C,\, v \in T(y)} \langle v, x - y \rangle \\ &= max \left\{ r \, , \sup_{y \in\, C,\, \|y\| \leq \alpha,\, v \in T(y)} \langle v, x - y \rangle \right\} \\ &\leq max \left\{ r \, , \sup_{y \in\, C,\, \|y\| \leq \alpha} \langle u, x - y \rangle \right\}, \end{aligned}$$

onde a segunda equação é uma conseqüência direta de (3.55) e a última desigualdade é válida por monotonicidade de T. Mas

$$max \left\{ r \, , \sup_{y \in C,\, \|y\| \leq \alpha} \langle u, x - y \rangle \right\} < +\infty,$$

pois o supremo é tomado num conjunto limitado. Portanto $Gap_{\,T,C}(x) < +\infty$, para todo $x \in dom(T) \cap C$. ■

O seguinte teorema é uma resposta a nossa pergunta feita anteriormente.

Teorema 3.6 *Considere* $VIP(T, C)$ *com* $T = c + N_V$*,* $V = \{x \in \mathbb{R}^n \;/\; Ax = b\}$*, e* C *un conjunto convexo e fechado em* $\mathbb{R}^n$ *. Suponha que:*

i) $dom(T) \cap C^0 \neq \phi$

ii) $S(T, C) \neq \phi$.

iii) *g é função de Bregman com zona* $\mathbb{R}^n_{++}$ *e coerciva na fronteira.*

iv) *Existe* $\alpha > 0$ *e* $r \in \mathbb{R}$ *tais que para todo* $y \in dom(T) \cap C$*, com* $\|y\| > \alpha$*, é válido:* $\langle v, y - x \rangle \geq r$ *para todo* $v \in T(y)$*.*

Sejam $\left\{x^k\right\}$ *a seqüência gerada por* $GPPA$ *com função de Bregman* g *começando em* $x^0 \in dom(T) \cap C$ *e* $\{x(\mu)/\mu > 0\}$ *a trajetória central com barreira* $h(x) = D_g(x, x^0)$, *com* $x \in C$. *Então*

$$\left\{x^k\right\} \subset \{x(\mu)/\mu > 0\},$$

e para qualquer seqüência crescente $\left\{\mu^k\right\} \subset I\!R_{++}$ *existe uma seqüência* $\{\lambda_k\}$ *em* $I\!R_{++}$ *tal que* $x(\mu_k) = x^k$, *para todo* k, *onde* $\left\{x^k\right\}$ *é a seqüência gerada por* $GPPA$ *com função de Bregman* g *e parâmetros de regularização* λ_k.

Demonstração. Primeiro é preciso verificar se $\left\{x^k\right\}$ e $\{x(\mu)/\mu > 0\}$ estão bem definidos.

Para $\{x(\mu)/\mu > 0\}$, observe que as hipóteses do Lema 2.1(ii) são satisfeitas, pois H13 segue da Proposição 3.5.1. Logo, a trajetória central está bem definida.

Para $\left\{x^k\right\}$, observe que as hipóteses do Lema 3.2 são satisfeitas. Logo, a seqüência $GPPA$ está bem definida.

Pela Proposição 2.2.2, $N_V(x) = Im(A^t)$, $\forall x \in V$. Logo, de (3.3),

$$\begin{aligned} 0 \in T_\ell(x^{\ell+1}) &= T(x^{\ell+1}) + \lambda_\ell \left[\nabla g(x^{\ell+1}) - \nabla g(x^\ell)\right] \\ &= c + N_V(x^{\ell+1}) + \lambda_\ell \left[\nabla g(x^{\ell+1}) - \nabla g(x^\ell)\right] \\ &= c + Im(A^t) + \lambda_\ell \left[\nabla g(x^{\ell+1}) - \nabla g(x^\ell)\right]. \end{aligned}$$

Assim, esiste $w^\ell \in Im(A^t)$ tal que

$$\begin{aligned} 0 &= c + A^t w^\ell + \lambda_\ell \left[\nabla g(x^{\ell+1}) - \nabla g(x^\ell)\right] \\ &= \frac{1}{\lambda_\ell}\left[c + A^t w^\ell\right] + \nabla g(x^{\ell+1}) - \nabla g(x^\ell). \end{aligned} \tag{3.56}$$

Somando em (3.56) de $\ell = 0$ até $\ell = k-1$,

$$\begin{aligned} 0 &= \sum_{\ell=0}^{k-1} \frac{1}{\lambda_\ell}\left[c + A^t w^\ell\right] + \sum_{\ell=0}^{k-1}\left[\nabla g(x^{\ell+1}) - \nabla g(x^\ell)\right] \\ &= c\sum_{\ell=0}^{k-1}\frac{1}{\lambda_\ell} + A^t\left(\sum_{\ell=0}^{k-1}\frac{1}{\lambda_\ell}w^\ell\right) + \left[\nabla g(x^1) - \nabla g(x^0)\right] + \cdots + \left[\nabla g(x^k) - \nabla g(x^{k-1})\right] \\ &= \mu_k c + \mu_k\, A^t\left((\mu_k)^{-1}\sum_{\ell=0}^{k-1}\frac{1}{\lambda_\ell}w^\ell\right) + \nabla g(x^k) - \nabla g(x^0) \\ &= \mu_k\left[c + A^t \bar{w}^\ell\right] + \nabla h(x^k), \end{aligned} \tag{3.57}$$

onde $\quad \mu_k = \sum_{\ell=0}^{k-1}(\lambda_\ell)^{-1} \quad$ e $\quad \bar{w}^k = (\mu_k)^{-1}\sum_{\ell=0}^{k-1} w^\ell.$

Por outro lado, como $x(\mu)$ é zero de $T_\mu = \mu T + \partial h$ em $dom(T) \cap C^0$, então

$$\begin{aligned} 0 \in T_\mu(x(\mu)) &= \mu T(x(\mu)) + \partial h(x(\mu)) \\ &= \mu \left[c + N_V(x(\mu))\right] + \nabla g(x(\mu)) - \nabla g(x^0) \\ &= \mu \left[c + Im(A^t))\right] + \nabla h(x(\mu)). \end{aligned} \tag{3.58}$$

De (3.57) e (3.58), segue que $x^k = x(\mu_k)$, para todo k.

Se uma seqüência $\{\mu_k\}_{k \in I\!N} \subset R_{++}$ dada é crescente então, pelo mesmo argumento, $x^k = x(\mu_k)$, onde $\{x^k\}$ é gerada com parâmetros $\lambda_k = (\mu_k - \mu_{k-1})^{-1} > 0$. Isto completa a prova. ■

O Teorema 3.6 diz que para o caso de programação linear, isto é quando $C = I\!R^n_{++}$, as noções de trajetória central e seqüência ponto proximal generalizado coincidem. O resultado depende de um modo essencial do fato que $\widehat{T}$ é constante.

Em geral, não é válido que a trajetória central e seqüência ponto proximal generalizada coincidam, nem mesmo para programação quadrática, como mostra o seguinte resultado.

Exemplo 3.5.1 *Seja $n = 2$. Considere o problema*

$$\begin{cases} min & \frac{1}{2} \parallel x \parallel^2 \\ s.a & x \geq 0 \end{cases}$$

Considere a distância de Itakura-Saitu (veja Exemplo 3.1.3), $x^0 = (1/8, 1/2)$, $\lambda_0 = 1/48$ e $\lambda_0 = 1/9$. Então a seqüência GPPA não está contida na trajetória central.

De fato. Neste caso, a função barreira é dada por

$$h(x) = D_g(x, x^0) = 8x_1 + 2x_2 - log(8x_1) - log(2x_2) - 2,$$

com $x_1 > 0$ e $x_2 > 0$. Logo, para cada $\mu > 0$, a trajetória central é dada por

$$x_\mu = arg\,min \left\{ \frac{\mu}{2} \parallel x \parallel^2 + h(x) \right\}$$
$$x_\mu = arg\,min \left\{ \frac{\mu}{2} x_1^2 + \frac{\mu}{2} x_2^2 + 8x_1 + 2x_2 - log(8x_1) - log(2x_2) - 2 \right\}.$$

Igualando as derivadas parciais a zero tem-se o sistema

$$\begin{cases} \mu x_1 + 8 - \frac{1}{x_1} &= 0, \\ \mu x_2 + 2 - \frac{1}{x_2} &= 0, \end{cases} \iff \begin{cases} x_1^2 + \frac{8}{\mu}x_1 - \frac{1}{\mu} &= 0, \\ x_2^2 + \frac{2}{\mu}x_2 - \frac{1}{\mu} &= 0. \end{cases}$$

Equivalentemente,

$$\begin{cases} \left[x_1 + \frac{4}{\mu}\right]^2 - \frac{16}{\mu^2} - \frac{\mu}{\mu^2} &= 0, \\ \left[x_2 + \frac{1}{\mu}\right]^2 - \frac{1}{\mu^2} - \frac{\mu}{\mu^2} &= 0, \end{cases} \iff \begin{cases} x_1 = -\frac{4}{\mu} + \frac{\sqrt{16+\mu}}{\mu}, \\ x_2 = -\frac{1}{\mu} + \frac{\sqrt{1+\mu}}{\mu}. \end{cases}$$

Assim, a trajetória central é

$$x(\mu) = (x_1(\mu), x_2(\mu)) = \mu^{-1}(\sqrt{16+\mu} - 4, \sqrt{1+\mu} - 1), \;\; \mu > 0.$$

Por outro lado, a seqüência $GPPA$ é dada por

$$\begin{cases} x^0 &= (\,\frac{1}{8}, \frac{1}{2}\,), \\ x^{k+1} &= arg\,min\left\{\,\frac{1}{2}\,\|\,x\,\|^2 + \lambda_k D_g(x, x^k)\,\right\}, \end{cases}$$

$$\Updownarrow$$

$$\begin{cases} x^0 &= (\,\frac{1}{8}, \frac{1}{2}\,), \\ x^{k+1} &= arg\,min\left\{ \frac{1}{2}x_1^2 + \frac{1}{2}x_2^2 + \lambda_k \left[\frac{x_1}{x_1^k} + \frac{x_2}{x_2^k} - log(\frac{x_1}{x_1^k}) - log(\frac{x_2}{x_2^k}) - 2\right]\right\}. \end{cases}$$

Igualando as derivadas parciais a zero tem-se o sistema

$$\begin{cases} x_1 + \frac{\lambda_k}{x_1^k} - \frac{\lambda_k}{x_1} &= 0, \\ x_2 + \frac{\lambda_k}{x_2^k} - \frac{\lambda_k}{x_2} &= 0, \end{cases} \iff \begin{cases} x_1^2 + \frac{\lambda_k}{x_1^k}x_1 - \lambda_k &= 0, \\ x_2^2 + \frac{\lambda_k}{x_2^k}x_2 - \lambda_k &= 0. \end{cases}$$

Equivalentemente,

$$\begin{cases} \left[x_1 + \frac{\lambda_k}{2x_1^k}\right]^2 - \frac{\lambda_k^2}{(2x_1^k)^2} - \lambda_k &= 0, \\ \left[x_2 + \frac{\lambda_k}{2x_2^k}\right]^2 - \frac{\lambda_k^2}{(2x_2^k)^2} - \lambda_k &= 0, \end{cases} \iff \begin{cases} x_1 = \frac{1}{2x_1^k}\left[-\lambda_k + \sqrt{\lambda_k^2 + \lambda_k(2x_1^k)^2}\;\right], \\ x_2 = \frac{1}{2x_2^k}\left[-\lambda_k + \sqrt{\lambda_k^2 + \lambda_k(2x_2^k)^2}\;\right]. \end{cases}$$

Assim, a seqüência $GPPA$ é dada por

$$\begin{cases} x^0 &= (\,\frac{1}{8}, \frac{1}{2}\,), \\ x^{k+1} &= \left(\frac{1}{2x_1^k}\left[-\lambda_k + \sqrt{\lambda_k^2 + \lambda_k(2x_1^k)^2}\;\right], \frac{1}{2x_2^k}\left[-\lambda_k + \sqrt{\lambda_k^2 + \lambda_k(2x_2^k)^2}\;\right]\right). \end{cases}$$

Para $\lambda_0 = \frac{1}{48}$ tem-se $x^1 = (x_1^1, x_2^1) = (\frac{1}{12}, \frac{1}{8})$.

A pergunta é: *Existe* $\mu > 0$ *tal que* $x(\mu) = x^1$*?* Resolvendo o sistema

$$\begin{cases} \mu^{-1}\sqrt{16+\mu} - 4 & = & \frac{1}{12}, \\ \mu^{-1}\sqrt{1+\mu} - 1 & = & \frac{1}{8}, \end{cases}$$

tem-se que existe $\mu = 48 = \lambda_0^{-1} > 0$ tal que $x(\mu) = x^1$.

Para $\lambda_1 = \frac{1}{48}$ tem-se $x^2 = (x_1^2, x_2^2) = (\frac{\sqrt{5}-2}{3}, \frac{1}{9})$.

A pergunta é: Existe $\mu > 0$ tal que $x(\mu) = x^2$? Resolvendo o sistema

$$\begin{cases} \mu^{-1}\sqrt{16+\mu} - 4 & = & \frac{\sqrt{5}-2}{3}, \\ \mu^{-1}\sqrt{1+\mu} - 1 & = & \frac{1}{9}, \end{cases}$$

tem-se que $x(\mu) \neq x^2$ para todo $\mu > 0$.

Portanto, a seqüência $GPPA$ não está contida na trajetória central. $\blacksquare$

Capítulo 4

Relações entre Trajetória Central e Trajetória de Cauchy em Programação Linear

Neste capítulo é feita uma ligação entre trajetória central e trajetória de Cauchy em uma variedade Riemanniana. Na seção 4.1 é definida trajetória de Cauchy em uma variedade Riemanniana fazendo as considerações necessárias que serão utilizadas na seção 4.2, onde é mostrado que a trajetória central e trajetória de Cauchy coincidem para problemas de programação linear. Estes fatos foram obtidos de [6] e [7].

4.1 Trajetórias de Cauchy em Variedades Riemannianas

Seja M uma variedade Riemanniana de dimensão s com métrica $\langle\cdot,\cdot\rangle$. Dado um sistema de coordenadas e uma vizinhança U do ponto $p \in M$, a métrica em M é dada pela matriz definida positiva e simétrica $H(q)$, com $H(q)_{ij} = \langle \frac{\partial}{\partial x_i}|_q , \frac{\partial}{\partial x_j}|_q \rangle$ para todo $q \in U$ (veja Definição C.4.1), onde $\{\frac{\partial}{\partial x_i}\}$ é a base do espaço tangente de M.

Se $f : M \to I\!R$ é diferenciável, o gradiente de f (veja Definição C.6.1) é o campo vetorial $\operatorname{grad} f$ definido por

$$\langle \operatorname{grad} f, X\rangle = df(X) = \frac{\partial f}{\partial X},$$

onde X é qualquer campo vetorial e $\frac{\partial f}{\partial X}$ é a derivada de f na direção X.

Da Proposição C.6.2, temos que

$$\operatorname{grad} f(q) = H(q)^{-1}\nabla f(q), \tag{4.1}$$

onde $\nabla f(q) = \left(\frac{\partial f}{\partial x_1}(q), \cdots, \frac{\partial f}{\partial x_s}(q)\right)$.

Seja $N \subset M$ uma subvariedade de M (veja Definição C.2.4). Para cada $x \in N$, sejam T_xM e T_xN os espaços tangentes de M e N em x, respectivamente. Seja $\Pi_x : T_xM \to T_xN$ a projeção ortogonal sobre T_xN. Se $f|_N$ é a restrição de f a N, então o gradiente de $f|_N$ é dado por

$$\operatorname{grad} f|_N(x) = \Pi_x(\operatorname{grad} f(x)). \tag{4.2}$$

Definição 4.1.1 *A trajetória de Cauchy para f en N é a curva parametrizada $x : [0, \beta] \to N$ dada por:*

$$x(0) = x^0\ , \tag{4.3}$$

$$\dot{x}(t) = -\operatorname{grad} f|_N(x(t))\ , \tag{4.4}$$

para algum $x^0 \in N$ dado e algum $\beta > 0$.

Dada uma função $f \in C^1(M)$, é bem conhecido que para $x^0 \in N$, existe $\beta > 0$ tal que o problema (4.3)-(4.4) possui uma única solução.

Observação 4.1.1 *Quando M é um subconjunto de $I\!R^s$, então a representação da métrica, dada por $H(x)$, $\forall x \in M$, é global. Assim, $H : M \to I\!R^{s\times s}$ é diferenciável. Além disso,* grad f *coincide com o vetor gradiente de funções reais de várias variáveis (veja Exemplo C.6.1).*

4.2 Trajetória de Cauchy e Trajetória Central em Programação Linear

O seguinte teorema estabelece, com as hipóteses do Teorema 3.6, que a trajetória de Cauchy em subvariedades euclidianas com métrica induzida e a trajetória central para uma certa barreira associada, coincidem.

Teorema 4.1 *Sejam M um subconjunto aberto de $I\!R^n$, $N = M \cap \{x \in I\!R^n / Ax = b\}$ e $f(x) = c^t x$, $\forall x \in M$. Seja H a representação da métrica em M e suponha que existe $h : M \to I\!R$ tal que $\nabla^2 h(x) = H(x)$, $\forall x \in M$ e $\nabla h(x^0) = 0$. Se a trajetória de Cauchy $\{x(t)/0 \leq t \leq \beta\}$, com $\beta > 0$, e a trajetória central $\{x(\mu)/\mu > 0\}$ existem, então elas coincidem.*

Demonstração. A trajetória central, neste caso, como em (3.58), satisfaz

$$0 = \mu \left[c + A^t w(\mu)\right] + \nabla h(x(\mu)), \tag{4.5}$$

para algum $w(\mu) \in Im(A^t)$.

De (4.5), $\nabla h(x(0)) = 0$. Como $\nabla h(x^0) = 0$ então, pela convexidade estrita de h, $x(0) = x^0$. Assim (4.3) é estabelecido. Resta mostrar que (4.4) é válido.

Seja P_A a projeção ortogonal sobre $Ker(A)$. Por (4.5), $\mu c + \nabla h(x(\mu)) \in Im(A^t)$. Assim, para todo $\mu > 0$,

$$0 = P_A \left[\mu c + \nabla h(x(\mu))\right]. \tag{4.6}$$

Observe que P_A é linear. Logo, diferenciando com relação a μ em (4.6), tem-se

$$0 = P_A \left[c + \nabla^2 h(x(\mu))\dot{x}(\mu)\right] = P_A \left[c + H(x(\mu))\dot{x}(\mu)\right]. \tag{4.7}$$

ou equivalente,

$$H(x(\mu))\dot{x}(\mu) - c \in Im(A^t). \tag{4.8}$$

Por outro lado,

$$\begin{aligned} \operatorname{grad} f|_N(x(\mu)) &= \Pi_{x(\mu)}\left(\operatorname{grad} f(x(\mu))\right) \\ &= \Pi_{x(\mu)}\left(H(x(\mu))^{-1}\nabla f(x(\mu))\right) = \Pi_{x(\mu)}\left(H(x(\mu))^{-1}c\right), \end{aligned} \tag{4.9}$$

onde $\Pi_{x(\mu)}$ é a projeção ortogonal de $T_{x(\mu)}M$ sobre $T_{x(\mu)}N$ com relação ao produto interno induzido pela métrica; isto é, $\langle u, v\rangle = u^t H(x(\mu))v$.

Como M é um aberto em $I\!R^n$ e N é uma variedade afim, tem-se que $T_{x(\mu)}M = I\!R^n$ e $T_{x(\mu)}N = Ker(A)$. Assim, para todo $y \in I\!R^n$, $\Pi_{x(\mu)}(y)$ é a única solução z do problema

$$min \;\; (z - y)^t H(x(\mu))(z - y) \tag{4.10}$$

$$s.a. \quad Az = 0, \tag{4.11}$$

cujas condições de otimalidade de Karush-Kuhn-Tucker (veja seção B.1) são a equação (4.11) e em adição

$$H(x(\mu))(z - y) = A^t w, \tag{4.12}$$

para algum $w \in I\!R^n$. Observe que $H(x(\mu))$ é o hessiano de h em $x(\mu)$ (veja seção C.7), $H(x(\mu))$ é simétrica e definida positiva. Para $y = H(x(\mu))^{-1}c$, a equação (4.12) é reduzida

$$-H(x(\mu))z + c \in Im(A^t). \tag{4.13}$$

De (4.8), $z = -\dot{x}(\mu)$ satisfaz (4.13). Como $x(\mu) \in dom(T) \subset V = N$, tem-se que $Ax(\mu) = b$, para todo $\mu > 0$. Isto implica, $0 = A\dot{x}(\mu) = A(-\dot{x}(\mu))$. Assim, $z = -\dot{x}(\mu)$ satisfaz (4.11) e (4.13) e é então igual a $\Pi_{x(\mu)}y = \Pi_{x(\mu)}H(x(\mu))^{-1}c$.

De (4.9) segue que $\dot{x}(\mu) = -\mathrm{grad}\, f|_N(x(\mu))$; isto é, (4.4) é válido.

Portanto, $\{x(\mu)/\mu > 0\}$ coincide com a trajetória de Cauchy. ■

Como foi mencionado previamente, o resultado do Teorema 4.1, cobre o caso de programação linear; isto é $M = I\!R^n_{++}$.

Observação 4.2.1 *Vale a pena observar que somente uma das duas condições impostas sobre h nas hipóteses do Teorema 4.1 é realmente importante, que $\nabla^2 h(x) = H(x)$.*

De fato. Se esta relação é satisfeita para alguma função $\bar{h}$, então é possível definir $h(x) = \bar{h}(x) - \nabla\bar{h}(x^0)^t x$; e assim $\nabla^2 h(x) = H(x)$ e $\nabla h(x^0) = 0$. ■

Observação 4.2.2 *Se $A \in I\!R^{m\times n}$ tem posto máximo, então* (4.4) *pode ser escrito da forma*

$$\dot{x}(t) = -H(x(t))^{-1}\left[I - A^t\left(AH(x(t))^{-1}A^t\right)^{-1}AH(x(t))^{-1}\right]\nabla f(x(t)). \tag{4.14}$$

De fato. Da equação (4.12) temos

$$z - y = H(x(\mu))^{-1}A^t w, \tag{4.15}$$

para algum $w \in I\!R^m$. Logo, $Az - Ay = AH(x(\mu))^{-1}A^t w$. Como z satisfaz (4.11), então $-Ay = AH(x(\mu))^{-1}A^t w$. Assim, $w = -\left(AH(x(\mu))^{-1}A^t\right)^{-1}Ay$.

Sustituindo w em (4.15) temos

$$z = \left[I - H(x(\mu))^{-1}A^t\left(AH(x(\mu))^{-1}A^t\right)^{-1}A\right]y.$$

Daí,

$$z = H(x(\mu))^{-1}\left[H(x(\mu)) - A^t\left(AH(x(\mu))^{-1}A^t\right)^{-1}A\right]y.$$

Tomando $y = H(x(\mu))^{-1}c = H(x(\mu))^{-1}\nabla f(x(\mu))$ temos

$$z = H(x(\mu))^{-1}\left[I - A^t\left(AH(x(\mu))^{-1}A^t\right)^{-1}AH(x(\mu))^{-1}\right]\nabla f(x(\mu)). \tag{4.16}$$

Como a trajetória central e a trajetória de Cauchy coincidem então (4.14) segue de (4.16), (4.9) e (4.4). ■

Observação 4.2.3 *O Teorema 4.1 diz que, quando* $H(x) = \nabla^2 h(x)$ *e* $\nabla h(x^0) = 0$, *a curva dada por* (4.3) *e* (4.14) *coincide com a curva*

$$\begin{cases} x(\mu) & = & arg\,min\,\{\,\mu c + h(x)\,\} \\ & s.a. & Ax = b\,. \end{cases}$$

Observação 4.2.4 *Se h é a função do Exemplo (3.1.1), isto é* $h(x) = \sum_{j=1}^{n} x_j\, log(x_j)$ *tem-se*

$$H(x)^{-1} = diag\,(x_1, \cdots, x_n)\,. \tag{4.17}$$

Neste caso, a curva dada por (4.3) *e* (4.14) *contem a seqüência gerada por* $GPPA$ *para problemas de programação linear com divergência de Kullback-Leibler.*

Observação 4.2.5 *Se h é a função do Exemplo (3.1.2), com* $\alpha = 1$ *e* $\beta = \frac{1}{2}$, *tem-se*

$$H(x)^{-1} = 4diag\left(x_1^{3/2}, \cdots, x_n^{3/2}\right). \tag{4.18}$$

Este caso ainda não foi estudado particularmente, nem sob a perspectiva de métodos de pontos interiores para programação linear nem sob a visão de método ponto proximal generalizado.

Observação 4.2.6 *Se h é a função do Exemplo (3.1.3), isto é* $h(x) = -\sum_{j=1}^{n} log(x_j)$ *tem-se*

$$H(x)^{-1} = diag\left(x_1^2, \cdots, x_n^2\right). \tag{4.19}$$

Neste caso, a curva dada por (4.3) *e* (4.14) *é a trajetória afim escala, amplamente estudado do ponto de vista dos métodos de pontos interiores para programação linear (Adler-Monteiro [1]).*

Conclusões

Feita a discussão do tema, apresentamos as seguintes conclções:

Para a pergunta (Q1): uma resposta é dada pelo

Teorema 2.1. *Suponha que:*

i) *h é estritamente convexa e contínua em $ED(h)$, e diferenciável em C^0.*

ii) *h é coerciva na fronteira.*

iii) *h atinge seu mínimo em $dom(T) \cap C^0$.*

iv) *$T = \widehat{T} + N_V$, onde $\widehat{T} : I\!\!R^n \to I\!\!R^n$ é operador contínuo e paramonótono, e V é um conjunto convexo e fechado em $I\!\!R^n$.*

v) *$dom(T) \cap C^0 \neq \phi$.*

vi) *$S(T, C) \neq \phi$.*

vii) *Ou,* **1)** *h é coerciva na zona e finita na fronteira, ou*
2) *A função gap é finita e alguma das alternativas em H14 é válida.*

Então, para cada $\bar{\mu} > 0$, a trajetória central $\{x(\mu)/\mu > \bar{\mu}\}$ está bem definida, é contínua, é limitada, está contida em C^0 e possui pontos de acumulação que são soluções de $VIP(T, C)$. No caso que h seja finita na fronteira, a trajetória central converge para o centro analítico de $S(T, C)$.

Para a pergunta (Q2): uma resposta é dada pelo mesmo Teorema 2.1 quando a função barreira é finita na fronteira. Outra resposta, quando a função barreira é separável, é dada pelo

Teorema 2.2. *Seja $C = I\!R^n_{++}$. Suponha que:*

i) *h é estritamente convexa e contínua em $ED(h)$, e diferenciável em C^0.*

ii) *h é coerciva na fronteira.*

iii) *h atinge seu mínimo em $dom(T) \cap C^0$.*

iv) *h é separável.*

v) *$T = \widehat{T} + N_V$, onde $\widehat{T} : I\!R^n \to I\!R^n$ é operador contínuo e paramonótono, e V é um conjunto convexo e fechado em $I\!R^n$.*

vi) *$\widehat{T}$ é continuamente diferenciável e existe subespaço W tal que $Ker(J_{\widehat{T}}(x)) = W$, para todo $x \in C \cap V$.*

vii) *$dom(T) \cap C^0 \neq \phi$.*

viii) *$S(T, C) \neq \phi$.*

ix) *A função gap é finita e alguma das alternativas em H14 é válida.*

Então a trajetória central $\{x(\mu)/\mu > 0\}$ converge, quando μ tende para $+\infty$, a uma solução x^* do problema:

$$\left\{ \min_{x \in S(T,C)} \tilde{h}(x), \right.$$

onde $\tilde{h}(x) = \sum_{j \in J} h_j(x_j)$ *e* $J = \{j \in \{1, ..., n\} / z_j > 0, \text{ para algum } z \in S(T, C)\}$.

Para a pergunta (Q3): uma resposta é dada pelo

Teorema 3.2. *Seja $T : I\!R^n \to \mathcal{P}(I\!R^n)$ um operador monótono maximal . Considere $VIP(T, C)$, onde C é convexo e fechado em $I\!R^n$. Seja g uma função de Bregman com zona C^0. Suponha que:*

i) *$dom(T) \cap C^{,0} \neq \phi$.*

ii) *$S(T, C) \neq \phi$.*

iii) *T é pseudomonótono em $dom(T)$.*

iv) *$\lambda_k \in (0, \tilde{\lambda}]$, para algum $\tilde{\lambda} > 0$.*

v) *Ou,* **v1)** *g é coerciva na zona, ou*

v2) *g é coerciva na fronteira e $Gap_{T,C}(x) < +\infty$, $\forall x \in C \cap dom(T)$.*

vi) *T é paramonótono em C.*

Então a seqüência $\{x^k\}$, gerada por $GPPA$, converge para uma solução $\bar{x}$ de $VIP(T, C)$.

Outra resposta, quando C é um conjunto poliedral, é dado pelo

Corolario 3.3. *Seja $\widehat{T} = T + N_E$, onde $T : I\!R^n \to I\!R^n$ é um operador monótono contínuo e $E = \{x \in I\!R^n \ / \ Ax = b\}$, $A \in R^{m\times n}$, $b \in I\!R^m$. Considere $VIP(T, C)$ com $C = \{x \in I\!R^n \ / \ Ax = b, \ x > 0\}$. Suponha que:*

i) *$S(T, C) \neq \phi$.*

ii) *Existe $x > 0$ tal que $Ax = b$.*

iii) *$\lambda_k \in (0, \tilde{\lambda}]$, para algum $\tilde{\lambda} > 0$.*

iv) *g é função de Bregman com zona $I\!R^n_{++}$ e coerciva na zona.*

v) *T é paramonótono em $I\!R^n_{+}$.*

Então a seqüência $\{x^k\}$, gerada por $GPPA$, converge para uma solução $\bar{x}$ de $VIP(T, C)$.

Para a pergunta (Q4): uma resposta, quando C é um conjunto poliedral e T é operador ponto a ponto monótono e continuamente diferenciável, é dado pelo

Teorema 3.4. *Sejam $A \in R^{m\times n}$, $b \in I\!R^m$. Considere $E = \{x \in I\!R^n \ / \ Ax = b\}$, e $C = \{x \in I\!R^n \ / \ Ax = b, \ x > 0\}$. Suponha que:*

i) *$\widehat{T} = T + N_E$, onde $T : I\!R^n \to I\!R^n$ é um operador monótono continuamente diferenciável.*

ii) *Existe subespaço W tal que $Ker(J_T(x)) = Ker(J_T(x)^s) = W$, para todo $x \in C \cap V$.*

iii) *$\lambda_k \in (0, \tilde{\lambda}]$, para algum $\tilde{\lambda} > 0$.*

iv) *g é função de Bregman com zona $I\!R^n_{++}$ e coerciva na zona.*

v) *$S(T, C) \neq \phi$.*

Então $\bar{x} = lim_{k\to+\infty}x^k$ *é solução do problema*

$$\left\{ \min_{x\in S(T,C)} D_g(x, x^0), \right.$$

onde $\{x^k\}$ *é a seqüência gerada por* $GPPA$.

Para a pergunta (Q5): Em primeiro lugar, o Corolario 3.5 estabelece que $\lim_{k\to+\infty} x^k$ e $lim_{\mu\to+\infty}x(\mu)$ coincidem com o centro analítico de $S(T, C)$, onde $\{x^k\}$ é a seqüência gerada por $GPPA$ e $x(\mu)$ sào os pontos da trajetória central que tem como função barreira a distância de Bregman. Especificamente:

Corolario 3.5. *Suponha que:*

i) $T = \widehat{T} + N_V$, *onde* $\widehat{T} : I\!R^n \to I\!R^n$ *é um operador contínuo e* V *um conjunto convexo e fechado de* $I\!R^n$.

ii) $\widehat{T}$ *é continuamente diferenciável e existe um subespaço* W *tal que* $Ker(J_T(x)) = W$, *para todo* $x \in C \cap V$.

iii) g *é função de Bregman com zona* $I\!R^n_{++}$, *coerciva na zona e finita na fronteira.*

iv) $dom(T) \cap C \neq \phi$.

v) $S(T, C) \neq \phi$.

Sejam $\{x^k\}$ *a seqüência gerada por* $GPPA$ *com função de Bregman* g *começando em* $x^0 \in dom(T) \cap C$ *e* $\{x(\mu)/\mu > 0\}$ *a trajetória central com barreira* $h(x) = D_g(x, x^0)$, *para todo* $x \in C$. *Então* $\lim_{k\to+\infty} x^k$ *e* $lim_{\mu\to+\infty}x(\mu)$ *coincidem e é a única solução do problema*

$$\left\{ \min_{x\in S(T,C)} D_g(x, x^0). \right.$$

Agora, quando o operador $\widehat{T} : I\!R^n \to I\!R^n$ é constante, $V = \{x \in I\!R^n \ / \ Ax = b\}$ e a função barreira é dada pela distância de Bregman, então a seqüência gerada por $GPPA$ está contida na trajetória central. Estes resultados podem ser encontrados especificamente no

Teorema 3.6. *Considere* $VIP(T, C)$ *com* $T = c + N_V$, $V = \{x \in I\!R^n \ / \ Ax = b\}$, *e* C *un conjunto convexo e fechado em* $I\!R^n$. *Suponha que:*

i) *$dom(T) \cap C^0 \neq \phi$*

ii) *$S(T, C) \neq \phi$.*

iii) *g é função de Bregman com zona $\mathbb{R}^n_{++}$ e coerciva na fronteira.*

iv) *Existe $\alpha > 0$ e $r \in \mathbb{R}$ tais que para todo $y \in dom(T) \cap C$, com $\|y\| > \alpha$, é valido: $\langle v, y - x\rangle \geq r$ para todo $v \in T(y)$.*

Sejam $\{x^k\}$ a seqüência gerada por $GPPA$ com função de Bregman g começando em $x^0 \in dom(T) \cap C$ e $\{x(\mu)/\mu > 0\}$ a trajetória central com barreira $h(x) = D_g(x, x^0)$, com $x \in C$. Então $\{x^k\} \subset \{x(\mu)/\mu > 0\}$, e para qualquer seqüência crescente $(\mu^k)_{k\in \mathbb{N}} \subset \mathbb{R}_{++}$ existe uma seqüência $(\lambda_k)_{k\in \mathbb{N}} \subset \mathbb{R}_{++}$ tal que $x(\mu_k) = x^k$, $\forall k \in \mathbb{N}$, onde $\{x^k\}$ é a seqüência gerada por $GPPA$ com função de Bregman g e parâmetros de regularização λ_k.

O Teorema 3.6 cobre o caso de programação linear, isto é, quando $C = \mathbb{R}^n_{++}$.

Para a pergunta (Q6): uma resposta é dada quando trabalhamos em programaçào linear. Esta conclução é encontrada especificamente no

Teorema 4.1. *Seja M um subconjunto aberto de $\mathbb{R}^n$, $N = M \cap \{x \in \mathbb{R}^n/Ax = b\}$ e $f(x) = c^t x$, $\forall x \in M$. Seja H a representação da métrica em M. Suponha que:*

i) *Existe $h : M \to \mathbb{R}$ tal que $\nabla^2 h(x) = H(x)$, $\forall x \in M$ e $\nabla h(x^0) = 0$, onde $x^0 \in M$.*

ii) *Existe a trajetória de Cauchy $\{x(t)/0 \leq t \leq \beta\}$, com $\beta > 0$.*

iii) *Existe a trajetória central $\{x(\mu)/\mu > 0\}$ com respeito a h.*

Entào a trajetória de Cauchy e a trajetória central coincidem.

Apêndice A

Elementos de Análise Convexa

Neste apêndice, apresentaremos alguns definições e resultados de análise convexa, os quais são relevantes nesta disertação. Os livros de Bazaraa [2], Solodov [14], Lemaréchal [18] são referências para este apêndice.

A.1 Minimização de Funções

Definição A.1.1 *Seja $D \subset I\!R^n$ um conjunto e $f : D \to I\!R$ uma função. Um problema de otimização é um problema da forma*

$$\min_{x \in D} f(x). \tag{A.1}$$

O conjunto D é chamado conjunto factível do problema (A.1)*, os pontos de D são chamados pontos factíveis e f é chamada função objetivo. Quando $D = I\!R^n$, dizemos que o problema de otimização é irrestrito, e quando $D \neq I\!R^n$ dizemos que o problema de otimização é com restrições.*

Definição A.1.2 *Dizemos que $\bar{x} \in D$ é*

i) *um minimizador global do problema* (A.1) *se*

$$f(\bar{x}) \leq f(x), \forall x \in D, \tag{A.2}$$

ii) *um minimizador local do problema* (A.1) *se existe $\varepsilon > 0$ tal que*

$$f(\bar{x}) \leq f(x), \forall x \in D \cap B[\bar{x}, \varepsilon]. \tag{A.3}$$

Se para todo $x \neq \bar{x}$ a desigualdade (A.2) *ou* (A.3) *é estrita, $\bar{x}$ é chamado minimizador estrito (global ou local, respectivamente).*

Observação A.1.1 *Pela Definição A.1.2, é claro que todo minimizador global é também um minimizador local, mas não reciprocamente.*

Definição A.1.3 *Dizemos que* $\bar{v} \in [-\infty, +\infty)$, *definido por* $\bar{v} = \inf_{x \in D} f(x)$ *é o valor ótimo do problema* (A.1)

Observação A.1.2 *Todo problema da forma*

$$\max_{x \in D} f(x) \tag{A.4}$$

é equivalente ao problema

$$\min_{x \in D} - f(x).$$

Em particular, as soluções locais e globais de ambos os problemas são as mesmas, com sinais opostos para os valores ótimos.

A.2 Existência de Soluções

Teorema A.1 (Teorema de Weierstrass) *Sejam* $D \subset I\!R^n$ *um conjunto compacto não vazio e* $f : D \to I\!R$ *uma função contínua. Então os problemas* (A.1) *e* (A.4) *possuem soluções globais.*

Demonstração. Veja Corolário 1, da seção 12, do Capítulo 1, de [15]. ■

Observação A.2.1 *A hipótese de que o conjunto* D *seja compacto só pode ser eliminada ao custo de fortalecimento das hipóteses sobre a função* f.

Definição A.2.1 *O conjunto de nível da função* $f : D \subset I\!R^n \to I\!R$ *associado a* $c \in I\!R$, *é o conjunto dado por*

$$L_{f,D}(c) = \{x \in D / f(x) \leq c\}.$$

Corolário A.2 *Sejam* $D \subset I\!R^n$ *um conjunto e* $f : D \to I\!R$ *uma função contínua em* D. *Suponha que existe* $c \in I\!R$ *tal que o conjunto de nível* $L_{f,D}(c)$ *seja não vazio e compacto. Então o problema* (A.1) *possui uma solução global.*

Demonstração. Como $L_{f,D}(c)$ é compacto e f é contínua então, pelo Teorema A.1, o problema

$$\min \{f(x) / x \in L_{f,D}(c)\}$$

possui uma solução global. Seja $\bar{x}$ a solução global deste problema. Então, $f(\bar{x}) \leq f(x)$, $\forall x \in L_{f,D}(c)$. Considere agora $x \in D \setminus L_{f,D}(c)$. Então, $f(x) > c \geq f(\bar{x})$. Logo, $f(\bar{x}) \leq f(x)$, $\forall x \in D$. Portanto, o problema (A.1) possui uma solução global. ■

Definição A.2.2 *Dizemos que uma seqüência* $(x^k)_{k\in I\!N} \subset I\!R^n$ *é crítica em relação ao conjunto* D*, se* $(x^k)_{k\in I\!N} \subset D$ *e* $\lim_{k\to+\infty} \| x^k \| = +\infty$ *ou* $\lim_{k\to+\infty} x^k = x \in \mathrm{cl}(D)\backslash D$*. Dizemos que a função* $f : D \to I\!R$ *é coerciva no conjunto* D*, quando para toda seqüência crítica* $(x^k)_{k\in I\!N}$ *em relação ao conjunto* D*, tem-se que* $\limsup_{k\to+\infty} f(x^k) = +\infty$*.*

Observação A.2.2 i) *Quando* D *é fechado, a Definição* (A.2.2) *pode ser simplificada afirmando que* $x^k \in D$*,* $\forall k \in I\!N$ *e* $\lim_{k\to+\infty} \| x^k \| = +\infty$ *implicam* $\limsup_{k\to+\infty} f(x^k) = +\infty$*.*
ii) *Quando* D *é limitado, a Definição* (A.2.2) *pode ser simplificada afirmando que* $x^k \in D$*,* $\forall k \in I\!N$ *e* $\lim_{k\to+\infty} x^k = x \in \mathrm{cl}(D) \setminus D$ *implicam* $\limsup_{k\to+\infty} f(x^k) = +\infty$*.*
iii) *Quando* D *é compacto, não existem seqüências críticas e, portanto, qualquer função* $f : D \to I\!R$ *é coerciva em* D *trivialmente.*

Corolário A.3 *Sejam* $D \subset I\!R^n$ *e* $f : D \to I\!R$ *uma função contínua e coerciva en* D*. Então o problema* (A.1) *possui uma solução global.*

Demonstração. Seja $\hat{x} \in D$ arbitrário e considere $c = f(\hat{x})$ fixo. Assim, conjunto de nível $L_{f,D}(c)$ é não vazio.

Afirmamos que $L_{f,D}(c)$ é limitado. De fato. Suponha que $L_{f,D}(c)$ seja não limitado. Entao existe seqüência $(x^k)_{k\in I\!N}$ em $L_{f,D}(c)$ tal que $\lim_{k\to+\infty} \| x^k \| = +\infty$. Como $x^k \in D$, $\forall k \in I\!N$, então $f(x^k) \leq c$, $\forall k \in I\!N$. Daí, $\sup_{k\geq m} f(x^k) \leq c$, $m \in I\!N$. Logo, $\limsup_{k\to+\infty} f(x^k) = \inf_m \sup_{k\geq m} f(x^k) \leq c$, o que é um absurdo, pois f é coerciva. Portanto, $L_{f,D}(c)$ é limitado.

Afirmamos que $L_{f,D}(c)$ é fechado. De fato. Seja $(x^k)_{k\in I\!N}$ em $L_{f,D}(c)$ tal que $\lim_{k\to+\infty} x^k = x$. Como f é contínua, $f(x) = \lim_{k\to+\infty} f(x^k) \leq c$. . Logo, $x \in L_{f,D}(c)$, conclui-se que $L_{f,D}(c)$ é fechado.

Portanto, $L_{f,D}(c)$ é compacto. Pelo Corolario A.2, o problema (A.1) possui uma solução global. ■

A.3 Conjuntos Convexos

Um conjunto convexo se caracteriza por conter todos os segmentos cujos extremos pertencem ao conjunto. Mas precisamente:

Definição A.3.1 *Um conjunto* $C \subset I\!R^n$ *é chamado convexo se para qualquer* $x \in C$*,* $y \in C$ *e* $\alpha \in [0,1]$ *tem-se:*

$$\alpha x + (1-\alpha)y \in C. \tag{A.5}$$

O ponto $\alpha x + (1-\alpha)y$ é chamado combinação convexa de x e y.

Exemplo A.3.1 *Qualquer subespaço de $I\!R^n$ é um conjunto convexo.*

Exemplo A.3.2 *Um semi-espaço em $I\!R^n$ é um conjunto da forma $\{x \in I\!R^n / \langle a, x\rangle \leq c\}$, onde $a \in I\!R^n$ e $c \in I\!R$. Seque da Definição A.3.1 que qualquer semi-espaço em $I\!R^n$ é um conjunto convexo.*

Exemplo A.3.3 *Um hiperplano em $I\!R^n$ é um conjunto da forma $\{x \in I\!R^n / \langle a, x\rangle = c\}$, onde $a \in I\!R^n$ e $c \in I\!R$. Seque da Definição A.3.1 que qualquer hiperplano em $I\!R^n$ é um conjunto convexo.*

Proposição A.3.1 *Sejam $C_j \subset I\!R^n$, $j \in I$, conjuntos convexos, onde I é um conjunto de índices. Então $C = \bigcap_{j \in I} C_j$ é um conjunto convexo.*

Demonstração. Sejam x e y em C. Então $x \in C_j$ e $y \in C_j$, $\forall j \in I$. Como C_j é convexo $\forall j \in I$, então para $\alpha \in [0,1]$, tem-se que $\alpha x + (1-\alpha)y \in C_j$, $\forall j \in I$. Logo, $\alpha x + (1-\alpha)y \in C$. Portanto, C é convexo. $\blacksquare$

Definição A.3.2 *Um conjunto $C \subset I\!R^n$ é chamado poliedral quando pode ser representado como o conjunto de soluções de um sistema finito de equações e inequações lineares.*

Proposição A.3.2 *Todo conjunto poliedral em $I\!R^n$ é convexo.*

Demonstração. Segue da Definição A.3.2 que todo conjunto poliedral é a interseção de hiperplanos e semi-espaços em $I\!R^n$. Pelos Exemplos A.3.2 e A.3.3 tem-se que os hiperplanos e semi-espaços em $I\!R^n$ são conjuntos convexos. Pela Proposição A.3.1 segue que todo conjunto poliedral em $I\!R^n$ é convexo.

Proposição A.3.3 *Seja $C \subset I\!R^n$ um conjunto convexo. Então $\mathrm{cl}(C)$ e C^0 são conjuntos convexos.*

Demonstração. Sejam $x \in \mathrm{cl}(C)$, $y \in \mathrm{cl}(C)$ e $\alpha \in [0,1]$. Então existem seqüências $(x_k)_{k \in I\!N}$ e $(y_k)_{k \in I\!N}$ em C tais que $\lim\limits_{k \to +\infty} x_k = x$ e $\lim\limits_{k \to +\infty} y_k = y$. Pela convexidade de C, $\alpha x_k + (1-\alpha)y_k \in C$, $\forall k \in I\!R$. Logo, $\alpha x + (1-\alpha)y = \lim\limits_{k \to +\infty} [\alpha x_k + (1-\alpha)y_k]$. Assim, $\alpha x + (1-\alpha)y \in \mathrm{cl}(D)$. Portanto, $\mathrm{cl}(C)$ é convexo.

Sejam agora $x \in C^0$ e $y \in C^0$. Então existem $\varepsilon_1 > 0$ e $\varepsilon_2 > 0$ tais que $B(x, \varepsilon_1) \subset C$ e $B(y, \varepsilon_2) \subset C$. Considerando $\varepsilon = min\,\{\varepsilon_1, \varepsilon_2\}$ tem-se que $B(x, \varepsilon) \subset C$ e $B(y, \varepsilon) \subset$

C. Para mostrar que $\alpha x + (1-\alpha)y \in C^0$ basta verificar que $B(\alpha x + (1-\alpha)y, \varepsilon) \subset C$. De fato.

Seja $z \in B(\alpha x + (1-\alpha), \varepsilon)$. Então $\| z - \alpha x - (1-\alpha)y \| < \varepsilon$. Se $q = z - \alpha x - (1-\alpha)y$, tem-se que $z = \alpha x + (1-\alpha)y + q$ e $\| q \| < \varepsilon$. Logo, $z = \alpha x + \alpha q + (1-\alpha)y + q - \alpha q = \alpha(x-q) + (1-\alpha)(y-q)$. Como $\| x + q - x \| = \| q \| < \varepsilon$, então $x + q \in B(x, \varepsilon) \subset C$. Da misma forma, $\| y + q - y \| = \| q \| < \varepsilon$ implica $y + q \in B(y, \varepsilon) \subset C$. Pela convexidade de C, $\alpha q = \alpha(x-q) + (1-\alpha)(y-q) \in C$. Assim, $z \in C$ e tem-se $B(\alpha x + (1-\alpha)y, \varepsilon) \subset C$. Logo, $\alpha x + (1-\alpha)y \in C^0$. Portanto, C^0 é convexo. ■

Definição A.3.3 *Sejam $x^i \in I\!\!R^n$ e $\alpha_i \in [0,1]$, $i = 1, ...p$, tais que $\sum_{i=1}^p \alpha_i = 1$. O ponto $\sum_{i=1}^p \alpha_i x^i$ é chamado a combinação convexa dos pontos $x^i \in I\!\!R^n$ com parámetros α_i.*

Pela Definição A.3.1, um conjunto convexo contém as combinações convexas de dois pontos arbitrários do conjunto. O seguinte resultado mostra que um conjunto convexo contém todas as combinações convexas de qualquer número de pontos do conjunto.

Teorema A.4 *Um conjunto $C \subset I\!\!R^n$ é convexo se, e somente se, para qualquer $p \in I\!\!N$, $x^i \subset C$ e $\alpha_i \in [0,1]$, $i = 1, ...p$, tais que $\sum_{i=1}^p \alpha_i = 1$, a combinação convexa $\sum_{i=1}^p \alpha_i x^i$ pertence a C.*

Demonstração. Suponha que C é convexo. A prova é feita por indução em $p \in I\!\!N$. Dados $p \in I\!\!N$, $x^i \in C$ e $\alpha_i \in [0,1]$ tais que $\sum_{i=1}^p \alpha_i = 1$. Defina $x = \sum_{i=1}^p \alpha_i x^i$.

Se $p = 1$ tem-se que $\alpha_1 = 1$ e $x = 1.x^1 \in C$.

Suponha que o argumento vale para $p = n$. Mostremos que o argumento vale para $p = n+1$. De fato.

Temos $\sum_{i=1}^{n+1} \alpha_i = 1$. Então $1 - \sum_{i=1}^p \alpha_i = \alpha_{n+1}$. Se $\alpha_{n+1} = 1$ então $\alpha_i = 0$, $\forall i = 1, ..., n$. Logo, $x = \sum_{i=1}^n 0.x^i + 1.x^{n+1} = x^{n+1} \in C$.

Considere $\alpha_{n+1} \in [0,1)$. Então $1 - \alpha_{n+1} > 0$. Logo,

$$\begin{aligned} x = \sum_{i=1}^{n+1} \alpha_i x^i &= \sum_{i=1}^{n} \alpha_i x^i + \alpha_{n+1} x^{n+1} \\ &= (1-\alpha_{n+1}) \sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} x^i + \alpha_{n+1} x^{n+1} \\ &= (1-\alpha_{n+1}) y + \alpha_{n+1} x^{n+1}, \end{aligned}$$

onde $y = \sum_{i=1}^{n} \beta_i x^i$, com $\beta_i = \dfrac{\alpha_i}{1-\alpha_{n+1}} \geq 0$, $\forall i = 1, ..., n$.

Tem-se que $\sum_{i=1}^{n} \beta_i = (1-\alpha_{n+1})^{-1} \sum_{i=1}^{n} \alpha_i = (1-\alpha_{n+1})^{-1}(1-\alpha_{n+1}) = 1$. Pela hipótese de indução, $y \in C$. Como $y \in C$ e $x^{n+1} \in C$ então, pela convexidade de C, $x \in C$. Portanto, $\sum_{i=1}^{p} \alpha_i x^i \in C$, $\forall p \in I\!N$, com $x^i \subset C$ e $\alpha_i \in [0,1]$, tais que $\sum_{i=1}^{p} \alpha_i = 1$.

Reciprocamente, suponha que $\sum_{i=1}^{p} \alpha_i x^i \in C$, $\forall p \in I\!N$, com $x^i \subset C$ e $\alpha_i \in [0,1]$, tais que $\sum_{i=1}^{p} \alpha_i = 1$. Sejam $x \in C$, $y \in C$ e $\alpha \in [0,1]$. Então $\alpha x + (1-\alpha)y = \alpha x + \beta y$, com $\beta = 1 - \alpha$. Pela hipótese, $\alpha x + \beta y \in C$. Portanto, C é convexo. ■

Definição A.3.4 *Seja $C \subset I\!R^n$ um conjunto qualquer. O fecho convexo de C, denotado por* conv(C)*, é o menor conjunto convexo em $I\!R^n$ que contém C (ou equivalentemente, a interseção de todos os conjuntos convexos em $I\!R^n$ que contem C).*

Observação A.3.1 *Como* conv(C) *é a interseção de conjuntos convexos em $I\!R^n$ que contém C então, pela Proposição A.3.1,* conv(C) *é um conjunto convexo. Além disso, se C é convexo então* conv$(C) = C$.

Proposição A.3.4 *Seja $C \subset I\!R^n$ um conjunto qualquer. O fecho convexo de C é o conjunto de todas as combinações convexas de pontos de C.*

Demonstração. Seja $D \subset I\!R^n$ o conjunto de todas as combinações convexas de pontos de C; isto é

$$D = \left\{ x \in I\!R^n \;\middle|\; x = \sum_{i=1}^{p} \alpha_i x^i,\ \sum_{i=1}^{p} \alpha_i = 1,\ x^i \in C,\ \alpha_i \geq 0,\ p \in I\!R \right\}. \tag{A.6}$$

Mostraremos que $D = \text{conv}(C)$. De fato. Como conv(C) é convexo então, pelo Teorema A.4, conv(C) contém todas as combinações convexas de pontos de conv(C). Como conv$(C) \supset C$ segue que conv(C) contém todas as combinações convexas de C. Logo $D \subset \text{conv}(C)$.

Por outro lado, é óbvio que $C \subset D$. Portanto, se D for convexo, necessariamente conv$(C) \subset$ conv$(D) = D$. Resta mostrar que D é convexo.

Sejam $z_1 \in D$ e $z_2 \in D$. Então, existem $p \in I\!N$ e $q \in I\!N$ tais que $z_1 = \sum_{i=1}^{p} \alpha_i x^i$ e $z_2 = \sum_{i=1}^{q} \beta_i y^i$, com $x^i \in C$, $y^i \in C$, $\alpha_i \geq 0$, $\beta_i \geq 0$ tais que $\sum_{i=1}^{p} \alpha_i = 1 = \sum_{i=1}^{q} \beta_i$. Para qualquer $\lambda \in [0,1]$ tem-se $\lambda z_1 + (1-\lambda) z_2 = \sum_{i=1}^{p} \lambda \alpha_i x^i + \sum_{i=1}^{q})(1-\lambda)\beta_i y^i$. É claro que $\lambda \alpha_i \geq 0$, $\forall i = 1, ..., p$ e $(1-\lambda)\beta_i \geq 0$, $\forall i = 1, ..., q$. Além disso,

$$\sum_{i=1}^{p} \lambda \alpha_i + \sum_{i=1}^{q} (1-\lambda)\beta_i = \lambda + (1-\lambda) = 1.$$

Isto mostra que o ponto $\lambda z_1+(1-\lambda)z_2$ é uma combinação convexa de pontos de D; em particular dos pontos $\{x^i/i=1,...,p\}\cup\{y^i/i=1,...,q\}$. Logo, $\lambda z_1+(1-\lambda)z_2\in D$.

Assim, D é convexo. Portanto, $D=\text{conv}(C)$. ■

Definição A.3.5 **i)** *Uma combinação afim de elementos* $x_1,...,x_k$ *de* $I\!R^n$ *é um elemento da forma* $\sum_{i=1}^k \alpha_i x_i$, *onde os coeficientes* α_i *satisfazem* $\sum_{i=1}^k \alpha_i = 1$.
ii) *Uma variedade afim em* $I\!R^n$ *é um conjunto contendo todas suas combinações afins. É facil verificar que a interseção de variedades afins é ainda uma variedade afim.*
iii) *Seja* $S\subset I\!R^n$ *um conjunto não vazio. O fecho afim de* S, *denotado por* $aff(S)$, *é a variedade afim generada por* S, *isto é, a interseção de todas as variedades afins que contém* S.

Observação A.3.2 *O fecho afim de* S *é a menor variedade afim que contém* S *e pode ser construida diretamente de* S *colecionando todas as combinações afins de elementos de* S.

Definição A.3.6 *Seja* $C\subset I\!R^n$ *um conjunto convexo. O interior relativo de* C, *denotado por* $ri(C)$ *é o interior de* C *com relação a topologia relativa do fecho afim de S; isto é,* $x\in ri(C)$ *se, e somente se,* $x\in aff(C)$ *e existe* $\delta>0$ *tal que* $aff(C)\cap B(x,\delta)\subset C$.

Exemplo A.3.4 *Se* $C=\{x\}$, *com* $x\in I\!R^n$, *então* $aff(C)=\{x\}$ *e* $ri(C)=\{x\}$. *Se* $C=\{\alpha x+(1-\alpha)y\ /\ 0\leq\alpha\leq 1\}$, *com* $x\in I\!R^n$ *e* $y\in I\!R^n$ *tais que* $x\neq y$, *então* $aff(C)=\{\alpha x+(1-\alpha)y\ /\ \alpha\in I\!R^n\}$ *e* $ri(C)=\{\alpha x+(1-\alpha)y\ /\ 0<\alpha<1\}$. *Se* $C=B(x_0,\delta)$, *com* $x_0\in I\!R^n$ *e* $\delta>0$, *então* $aff(C)=I\!R^n$ *e* $ri(C)=B(x_0,\delta)$.

Definição A.3.7 *Um conjunto* $K\subset I\!R^n$ *chama-se um cone quando*

$$d\in K \quad\Rightarrow\quad td\in K,\quad \forall\, t\in I\!R_+.$$

Em adição, se K *é convexo então* K *é chamado cone convexo.*

Observação A.3.3 *Pela Definição* A.3.7, *se* K *é um cone não vazio então necessariamente* $0\in K$. *Intuitivamente,* K *é um conjunto de direções.*

Exemplo A.3.5 *Alguns exemplos de cone são: o espaço* $I\!R^n$, *qualquer subespaço de* $I\!R^n$, *o ortante não negativo* $I\!R^n_+$.

A.4 Projeção sobre Conjuntos Convexos Fechados

Seja $C\subset I\!R^n$ um conjunto convexo não vazio e fechado. Para $x\in I\!R^n$ fixado, considere o seguinte problema:

$$\inf_{y\in C}\ \|\, y-x\,\|\,. \tag{A.7}$$

Dado $c \in C$, considere o conjunto $S = \{x \in I\!\!R^n \; / \; \| y - x \| \leq \| c - x \|\}$. Então (A.7) é equivalente ao problema:

$$\inf_{y \in C \cap S} \| y - x \|, \tag{A.8}$$

que tem uma solução pois a aplicação $y \longmapsto \| y - x \|$ é contínua e $C \cap S$ é um conjunto compacto. Portanto deduzimos a existência de um ponto em C que minimiza a distância ao ponto x.

Os dois teoremas seguintes mostram as afirmações feitas.

Teorema A.5 *Um ponto* $y_x \in C$ *é solução do problema* (A.7) *se, e somente se,*

$$\langle x - y_x, y - y_x \rangle \leq 0,$$

para todo $y \in C$.

Demonstração. Suponha que $y_x \in C$ é solução do problema (A.7). Sejam $y \in C$ e $\alpha \in (0,1)$. Pela convexidade de C, $\alpha y + (1-\alpha) y_x = y_x + \alpha(y - y_x) \in C$. Logo,

$$\begin{aligned} 0 \leq \| y_x - x \|^2 &\leq \| y_x + \alpha(y - y_x) - x \|^2 \\ &= \| (y_x - x) + \alpha(y - y_x) \|^2 \\ &= \| y_x - x \|^2 + 2\alpha \langle y_x - x, y - y_x \rangle + \alpha^2 \| y - y_x \|^2 . \end{aligned}$$

Assim, $0 \leq \alpha \langle y_x - x, y - y_x \rangle + \frac{1}{2}\alpha^2 \| y - y_x \|^2$. Como $\alpha > 0$,

$$0 \leq \langle y_x - x, y - y_x \rangle + \frac{1}{2}\alpha \| y - y_x \|^2 .$$

Logo, $0 \leq \langle y_x - x, y - y_x \rangle + \frac{1}{2} \lim_{\alpha \to 0} \alpha \| y - y_x \|^2$. Portanto, $\langle x - y_x, y - y_x \rangle \leq 0$, para todo $y \in C$.

Reciprocamente, suponha que $\langle x - y_x, y - y_x \rangle \leq 0$, $\forall y \in C$. Se $y_x = x$ então y_x é solução do problema (A.7). Se $y_x \neq x$ então para $y \in C$

$$\begin{aligned} 0 \geq \langle x - y_x, y - y_x \rangle &= \langle x - y_x, y - x + x - y_x \rangle \\ &= \| x - y_x \|^2 + \langle x - y_x, y - x \rangle \\ &\geq \| x - y_x \|^2 - \| x - y_x \| \| y - x \|, \end{aligned}$$

onde a Desigualdade de Cauchy Schawarz foi utilizada. Dividindo por $\| y_x - x \| > 0$, tem-se que $0 \geq \| x - y_x \| - \| y - x \|$, para todo $y \in C$. Logo, $\| x - y_x \| \leq \| x - y \|$, para todo $y \in C$. Portanto, $y_x \in C$ é solução do problema (A.7). $\blacksquare$

Teorema A.6 *Com as mesmas hipóteses do Teorema* A.5, y_x *é único.*

Demonstração. Suponha que existem $y_x \in C$ e $y'_x \in C$ tais que $\| y_x - x \| \leq \| y - x \|$ e $\| y'_x - x \| \leq \| y - x \|$. Pelo Teorema A.5, valem a desigualdades

$$\langle x - y_x, y - y_x \rangle \leq 0, \quad \langle x - y'_x, y - y'_x \rangle \leq 0,$$

para todo $y \in C$. Em particular, $\langle x - y_x, y'_x - y_x \rangle \leq 0$ e $\langle x - y'_x, y_x - y'_x \rangle \leq 0$. Somando, $\langle x - y_x, y'_x - y_x \rangle + \langle x - y'_x, y_x - y'_x \rangle \leq 0$. Então, $\langle x - y_x, y'_x - y_x \rangle - \langle x - y'_x, y'_x - y_x \rangle \leq 0$. Daí, $\langle x - y_x - x + y'_x, y'_x - y_x \rangle \leq 0$. Assim, $0 \leq \| y'_x - y_x \|^2 = \langle y'_x - y_x, y'_x - y_x \rangle \leq 0$. Segue que, $0 = \| y'_x - y_x \|^2$. Isto acontece só quando, $y'_x = y_x$. ■

Definição A.4.1 *O único ponto y_x em C que minimiza a distância de x ao conjunto C é chamado a projeção de $x \in I\!R^n$ ao conjunto C, e é denotado por $P_C(x)$.*

Teorema A.7 *Seja $C \subset I\!R^n$ um conjunto não vazio convexo e fechado, e seja $x \neq C$. Então existe $s \in I\!R^n$ tal que*

$$\langle s, x \rangle \geq sup \left\{ \langle s, y \rangle / y \in C \right\}. \tag{A.9}$$

Demonstração. Pelos Teoremas A.5 e A.6, existe um único ponto $P_C(x)$ que é solução do problema A.7. Defina $s := x - P_C(x) \neq 0$. Pelo Teorema A.5,

$$\langle x - P_C(x), y - P_C(x) \rangle \leq 0.$$

Daí, $0 \geq \langle s, y - x + s \rangle = \langle s, y \rangle - \langle s, x \rangle + \| s \|^2$. Logo, $\langle s, x \rangle \geq \langle s, y \rangle + \| s \|^2 \geq \langle s, y \rangle$, para todo $y \in C$. Portanto, existe $s = x - P_C(x)$ satisfazendo (A.9). ■

Definição A.4.2 *Dado um conjunto $C \subset I\!R^n$, com fronteira não vazia, e um ponto $x \in \partial C$, considere o hiperplano afim $H_{s,r} = \{ y \in I\!R^n / \langle s, y \rangle = r \}$, onde $s \in I\!R^n$ e $r \in I\!R$. O hiperplano $H_{s,r}$ é chamado hiperplano suporte ao conjunto C em x quando $\langle s, x \rangle \leq r$, $\forall x \in C$ e, além disso, $x \in H_{s,r}$.*

Teorema A.8 *Seja $x \in \partial C$, onde C é um conjunto não vazio e convexo em $I\!R^n$ (naturalmente $C \neq I\!R^n$). Então existe um hiperplano suporte ao conjunto C em x.*

Demonstração. Como C, o fecho de C e seus complementos possuem a mesma fronteira, uma seqüência $(x_k)_{k \in I\!N}$ pode ser encontrada tal que x_k não pertença ao fecho de C, $\forall k \in I\!N$ e $\lim_{k \to +\infty} x_k = x \in \partial C$. Para cada $k \in I\!N$ tem-se, pelo Teorema A.7, que existem $s_k \in I\!R^n$ tal que $\langle s_k, x_k \rangle > \langle s_k, y \rangle$, $\forall y \in C$. Logo, $\langle s_k, x_k - y \rangle > 0$, $\forall y \in C$. Segue que $s_k \neq 0$, $\forall k \in I\!N$. Seja $z_k = \dfrac{s_k}{\|s_k\|}$, $\forall k \in I\!N$. Então $\| z_k \| = 1$, $\forall k \in I\!N$. Assim, $(z_k)_{k \in I\!N}$ possui uma subseqüência $(z_k)_{k \in A}$ convergente, $A \subset I\!N$. Seja $s = \lim_{k \in A} z_k$. Logo, $0 < \lim_{k \in A} \langle z_k, x_k - y \rangle = \langle s, x - y \rangle$, $\forall y \in C$. Tomando $r = \langle s, x \rangle$ tem-se que $H_{s,r} = \{ y \in I\!R^n / \langle s, y \rangle = r \}$ é um hiperplano suporte ao conjunto C em x. ■

A.5 Funções Convexas

Definição A.5.1 *Seja $C \subset I\!R^n$ um conjunto convexo. Uma função $f : C \longrightarrow I\!R$ é chamada convexa em C quando para qualquer $x \in C$, $x \in C$ e $\alpha \in [0, 1]$, tem-se:*

$$f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y). \tag{A.10}$$

A função f é chamada estritamente convexa quando a desigualdade (A.10) *é estrita, para todo $x \neq y$ e $\alpha \in (0, 1)$.*

Proposição A.5.1 *Se $f : I\!R^n \longrightarrow I\!R$ é uma função convexa e $h : I\!R^n \longrightarrow I\!R$ é uma função estritamente convexa, então $f + g$ é uma função estritamente convexa.*

Demonstração. Segue da Definição A.5.1. ■

Definição A.5.2 *O epígrafo de uma função $f : C \subset I\!R^n \longrightarrow I\!R$ é o conjunto*

$$E_f = \{ (x, c) \in C \times I\!R / f(x) \leq c \} .$$

Proposição A.5.2 *Seja $C \subset I\!R^n$ um conjunto convexo. Uma função $f : C \longrightarrow I\!R$ é convexa em C se, e somente se, o epígrafo de f é um conjunto convexo em $I\!R^n \times I\!R$.*

Demonstração. Suponha que $f : C \longrightarrow I\!R$ é convexa em C. Sejam $(x, c_1) \in E_f$, $(x, c_1) \in E_f$. Isto é $f(x) \leq c_1$ e $f(y) \leq c_2$. Pela convexidade de f, para $\alpha \in [0, 1]$, segue que $f(\alpha x + (1 - \alpha) y) \leq \alpha f(x) + (1 - \alpha) f(y) \leq \alpha c_1 + (1 - \alpha) c_2$. Pela Definição A.5.2, $(\alpha x + (1 - \alpha) y, \alpha c_1 + (1 - \alpha) c_2) \in E_f$. Logo, $\alpha (x, c_1) + (1 - \alpha)(y, c_2) \in E_f$. Portanto, E_f é um conjunto convexo em $I\!R^n \times I\!R$.

Reciprocamente, suponha que E_f é um conjunto convexo em $I\!R^n \times I\!R$. Sejam $x \in C$ e $y \in C$. Tem-se que $(x, f(x)) \in E_f$, pois $f(x) \leq f(x)$. De forma análoga, $(y, f(y)) \in E_f$. Pela convexidade de E_f em $I\!R^n \times I\!R$, $\alpha(x, f(x)) + (1 - \alpha)(y, f(y)) \in E_f$,

$\alpha \in [0,1]$. Daí, $(\alpha x + (1-\alpha)y, \alpha f(x) + (1-\alpha)f(y)) \in E_f$. Pela Definição A.5.2, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)$. Portanto, f é convexa em C. ■

Proposição A.5.3 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo e $f : C \to \mathbb{R}$ uma função convexa em C. Então o conjunto de nível $L_{f,C}(t)$ é convexo, para todo $t \in \mathbb{R}$.*

Demonstração. Seja $t \in \mathbb{R}$. Se $L_{f,C}(t) = \phi$ então a conclusão é óbvia. Suponha que $L_{f,C}(t) \neq \phi$. Sejam $x \in L_{f,C}(t)$ e $y \in L_{f,C}(t)$. Segue, da Definição A.2.1, que $x \in C$, $f(x) \leq t$, $y \in C$ e $f(y) \leq t$. Pela convexidade de C, para todo $\alpha \in [0,1]$, tem-se que $\alpha x + (1-\alpha)y \in C$. Pela convexidade de f em C, $f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) \leq \alpha t + (1-\alpha)t = t$. Logo, $\alpha x + (1-\alpha)y \in L_{f,C}(t)$. Portanto, $L_{f,C}(t)$ é convexo, para todo $t \in \mathbb{R}$. ■

Proposição A.5.4 *Seja $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ uma funçao convexa. Suponha que exista $c \in \mathbb{R}$ tal que o conjunto de nível $L_{f,\mathbb{R}^n}(c)$ é não vazio e limitado. Então $L_{f,\mathbb{R}^n}(t)$ é limitado para todo $t \in \mathbb{R}^n$.*

Demonstração. Veja Teorema 3.4.4 em [14]. ■

Proposição A.5.5 *Sejam $C \subset \mathbb{R}^n$ um conjunto convexo e $f_i : C \longrightarrow \mathbb{R}$, $i \in I$, funções convexas em C, onde I é um conjunto qualquer de índices. Suponha que $\beta \in \mathbb{R}$ tal que $f_i(x) \leq \beta$ para todo $x \in C$ e $i \in I$. Então a função*

$$\begin{array}{rcccl} f & : & C & \longrightarrow & \mathbb{R} \\ & & x & \longmapsto & f(x) \;:=\; \sup\limits_{i \in I} f_i(x) \end{array}$$

é convexa em C.

Demonstração. Seja $c \in \mathbb{R}$ arbitrário. Pela Definição A.5.2 tem-se que

$$\begin{aligned} E_f &= \{\, (x,c) \in C \times \mathbb{R} \;/\; f(x) \leq c \,\} = \{\, (x,c) \in C \times \mathbb{R} \;/\; f_i(x) \leq c, \forall i \in I \,\} \\ &= \bigcap_{i \in I} \{\, (x,c) \in C \times \mathbb{R} \;/\; f_i(x) \leq c \,\} = \bigcap_{i \in I} E_{f_i}\,. \end{aligned}$$

Pela convexidade de f_i em C, segue da Proposição A.5.2, que os epígrafos E_{f_i} são convexos, para cada $i \in I$. Logo, pela Proposição A.3.1, a interseção destos epígrafos é um conjunto convexo. Novamente, pela Proposição A.5.2, a convexidade de E_f implica a convexidade de f. ■

Observação A.5.1 *Na Proposição* A.5.5*, a condição de que as funções que definem o supremo sejam limitadas superiormente é necessária somente para garantir que f tenha valores finitos no conjunto C. Em particular, esta hipótese não é necessária quando I é um conjunto infinito.*

Teorema A.9 (Caracterizações de Funções Convexas Diferenciáveis) *Sejam $C \subset I\!R^n$ um conjunto convexo aberto e $f : C \to I\!R$ uma função diferenciável em C. Então as seguintes proposições são equivalentes.*

i) *A função f é convexa em C.*

ii) $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$*, para todo $x, y \in C$.*

iii) $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$*, para todo $x, y \in C$.*

Demonstração. **i) $\Rightarrow$ ii)** Pela hipótese, para $x \in C$, $y \in C$ e $\alpha \in (0, 1]$, tem-se que $f(\alpha y + (1-\alpha)x) \leq \alpha f(y) + (1-\alpha) f(x)$. Considere $d = y - x$. Então $\alpha y + (1-\alpha)x = x + \alpha(y - x) = x + \alpha d$. Assim, $f(x + \alpha d) \leq \alpha f(y) + (1-\alpha) f(x)$. Então $f(x+\alpha d) - f(x) \leq \alpha(f(y) - f(x))$. Isto implica que $f(y) - f(x) \geq \dfrac{1}{\alpha}(f(x+\alpha d) - f(x))$. Logo, $f(y) - f(x) \geq \lim\limits_{\alpha \to 0} \dfrac{1}{\alpha}(f(x + \alpha d) - f(x)) = \dfrac{\partial f}{\partial d}(x) = \langle \nabla f(x), d \rangle$. Portanto, para todo $x, y \in C$, tem-se:

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle \tag{A.11}$$

ii) $\Rightarrow$ iii) Trocando x por y no item (ii), tem-se:

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \tag{A.12}$$

Somando (A.11) e (A.12), tem-se que $0 \geq \langle \nabla f(x), y - x \rangle + \langle \nabla f(y), x - y \rangle$. Portanto, $\langle \nabla f(y) - \nabla f(x), y - x \rangle \geq 0$, para todo $x, y \in C$.

iii) $\Rightarrow$ i) Pelo Teorema do Valor Médio [15], existe $\alpha \in (0, 1)$ tal que

$$f(y) - f(x) = \frac{\partial f}{\partial d}(x) = \langle \nabla f(x + \alpha d), d \rangle, \tag{A.13}$$

onde $d = y - x$. Aplicando a hipótese para $x + \alpha d$ e x, tem-se que $\langle \nabla f(x + \alpha d) - \nabla f(x), x + \alpha d - x \rangle = \langle \nabla f(x + \alpha d) - \nabla f(x), \alpha d \rangle \geq 0$. Então, $\langle \nabla f(x + \alpha d), \alpha d \rangle - \langle \nabla f(x), \alpha d \rangle \geq 0$. Daí, $\langle \nabla f(x + \alpha d), d \rangle \geq \langle \nabla f(x), d \rangle$. Assim, em A.13, tem-se que $f(y) - f(x) = \langle \nabla f(x+\alpha d), d \rangle \geq \langle \nabla f(x), y - x \rangle$. Portanto, $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$, para todo $x, y \in C$.

ii) $\Rightarrow$ *i)* Aplicando a hipótese para x e $x + \alpha d$, com $d = y - x$, segue que

$f(x) \geq f(x+\alpha d)+\langle \nabla f(x+\alpha d), x-x-\alpha d\rangle$. Então, $f(x) \geq f(x+\alpha d)-\alpha\langle \nabla f(x+\alpha d), d\rangle$. Daí,

$$(1-\alpha)f(x) \geq (1-\alpha)f(x+\alpha d) - (1-\alpha)\alpha\langle \nabla f(x+\alpha d), d\rangle. \tag{A.14}$$

De forma análoga, aplicando a hipótese para y e $x+\alpha d$, segue que

$$\alpha f(y) \geq \alpha f(x+\alpha d) + \alpha(1-\alpha)\langle \nabla f(x+\alpha d), d\rangle. \tag{A.15}$$

Somando (A.14) e (A.15), tem-se que $(1-\alpha)f(x)+\alpha f(y) \geq f(x+\alpha d)$. Logo, $f(\alpha y + (1-\alpha)x) = f(x+\alpha(y-x)) = f(x+\alpha d) \leq \alpha f(y)+(1-\alpha)f(x)$. Portanto, f é convexa em C. ■

Corolário A.10 *Sejam $C \subset I\!R^n$ um conjunto convexo aberto e $f : C \to I\!R$ uma função diferenciável em C. Então as seguintes proposições são equivalentes.*

i) *A função f é estritamente convexa em C.*

ii) *Para todo $x \in C$ e todo $y \in C$ tais que $x \neq y$, $f(y) > f(x) + \langle \nabla f(x), y-x\rangle$.*

iii) *Para todo $x \in C$ e todo $y \in C$ tais que $x \neq y$, $\langle \nabla f(y) - \nabla f(x), y-x\rangle > 0$.*

Demonstração. Segue da Definição A.5.1 e do Teorema A.9. ■

Teorema A.11 *Seja $f : I\!R^n \longrightarrow I\!R$ uma função convexa. Para todo $x \in I\!R^n$, existe $s = s(x) \in I\!R^n$ tal que $f(y) \geq f(x) + \langle s, y-x\rangle$, $\forall y \in I\!R^n$.*

Demonstração. Pela Proposição A.5.2, E_f é um conjunto convexo em $I\!R^n \times I\!R$. dados $x \in I\!R^n$ e $(x, f(x)) \in \partial E_f$, podemos tomar, pelo Teorema A.8, um hiperplano suporte ao conjunto E_f em $(x, f(x))$. Pelo Teorema A.7, existem $s = s(x) \in I\!R^n$ e $\alpha \in I\!R$, tais que

$$\langle s, y\rangle + \alpha r \leq \langle s, x\rangle + \alpha f(x),\ \forall (y, r) \in R^n \times I\!R. \tag{A.16}$$

Então $\langle s, y-x\rangle + \alpha(r - f(x)) \leq 0$, $\forall (y,r) \in R^n \times I\!R$. Escolha $r \in I\!R$ e $\delta > 0$ tais que $\delta s = y - x$ e $f(y) = f(x+\delta s) = r$. Logo, $\langle s, \delta s\rangle + \alpha(f(x+\delta s) - f(x)) \leq 0$. Assim, $0 \leq \delta \|s\|^2 \leq \alpha(f(x) - f(x+\delta s))$. Isto mostra que $\alpha \neq 0$, caso contrário $s = 0$. Sem perda de generalidade, considere $\alpha = -1$. Logo, em A.16, tem-se que $\langle s, y\rangle - f(y) = \langle s, r\rangle - r \leq \langle s, x\rangle - f(x)$, $\forall y \in I\!R^n$. Daí, $f(y) - \langle s, y\rangle \geq f(x) - \langle s, x\rangle$, $\forall y \in I\!R^n$. Portanto, Para todo $x \in I\!R^n$, existe $s = s(x) \in I\!R^n$ tal que $f(y) \geq f(x) + \langle s, y-x\rangle$, $\forall y \in I\!R^n$. ■

Observação A.5.2 *Segue, do Teorema* A.11*, que toda função convexa é minorizada por uma função afim.*

Teorema A.12 (Teorema de Minimização Convexa) *Sejam $C \subset I\!R^n$ um conjunto convexo e $f : C \longrightarrow I\!R$ uma função convexa em C. Então todo minimizador local do problema* (A.1) *é minimizador global. Além disso, o conjunto dos minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração. Seja $\bar{x}$ um minimizador local do problema (A.1) e suponha que $\bar{x}$ não é minimizador global. Sendo $\bar{x}$ minimizador local, existe $\varepsilon > 0$ tal que $f(\bar{x}) \leq f(x), \forall x \in D \cap B(\bar{x}, \varepsilon)$. Como $\bar{x}$ não é minimizador global, existe $y \in C$ tal que $f(y) < f(\bar{x})$. Sendo C convexo, $\lambda\bar{x} + (1-\lambda)z \in C$, para todo $\lambda \in (0,1)$. Escolha λ suficientemente próximo da unidade tal que $\lambda\bar{x} + (1-\lambda)y \in B(\bar{x}, \varepsilon)$. Pela convexidade de f, segue que $f(\lambda\bar{x} + (1-\lambda)y) \leq \lambda f(\bar{x}) + (1-\lambda)f(y) < \lambda f(\bar{x}) + (1-\lambda)f(\bar{x}) = f(\bar{x})$. Assim, $f(\lambda\bar{x} + (1-\lambda)y) < f(\bar{x})$. Mas como $\lambda\bar{x} + (1-\lambda)y \in B(\bar{x}, \varepsilon)$, então $f(\bar{x}) \leq f(\lambda\bar{x} + (1-\lambda)y)$; o que é um absurdo. Portanto, qualquer minimizador local do problema (A.1) é minimizador global.

Sejam $S \subset C$ o conjunto dos minimizadores (globais) e $\bar{v} \in I\!R$ o valor ótimo do problema (A.1); isto é $f(x) = \bar{v}$, $\forall x \in S$. Pela convexidade de f, para $x_1 \in S$, $x_2 \in S$ e $\alpha \in [0,1]$, tem-se que $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) = \alpha\bar{v} + (1-\alpha)\bar{v} = \bar{v}$. Como $\bar{v}$ é o valor ótimo do problema (A.1), $f(\alpha x_1 + (1-\alpha)x_2) = \bar{v}$. Assim, $\alpha x_1 + (1-\alpha)x_2 \in S$. Portanto, S é conevxo.

Suponha agora que f seja estritamente convexa e que existem $\bar{x}_1 \in C$ e $\bar{x}_2 \in C$, com $\bar{x}_1 \neq \bar{x}_2$. Seja $\alpha \in (0,1)$. Como C é convexo, $\alpha\bar{x}_1 + (1-\alpha)\bar{x}_2 \in C$. Como $\bar{x}_1$ e $\bar{x}_2$ são minimizadores, $f(\alpha\bar{x}_1 + (1-\alpha)\bar{x}_2) \geq f(\bar{x}_1) = f(\bar{x}_2) = \bar{v}$. Por outro lado, como f é estritamente convexa, $f(\alpha\bar{x}_1 + (1-\alpha)\bar{x}_2) < \alpha f(\bar{x}_1) + (1-\alpha)f(\bar{x}_2) = \alpha\bar{v} + (1-\alpha)\bar{v} = \bar{v}$. Assim, $f(\alpha\bar{x}_1 + (1-\alpha)\bar{x}_2) \geq \bar{v}$ e $f(\alpha\bar{x}_1 + (1-\alpha)\bar{x}_2) < \bar{v}$, o que é um absurdo. Portanto, Se f é estritamente convexa, não pode haver mais de um minimizador. ■

Teorema A.13 (Desigualdade de Jensen) *Sejam $C \subset I\!R^n$ um conjunto convexo e $f : C \longrightarrow I\!R$ uma função convexa em C. Então para quaisquer $p \in I\!N$, $x^i \in D$ e $\alpha_i \geq 0$, $i \in \{1, ..., p\}$, tais que $\sum_{i=1}^{p} \alpha_i = 1$, tem-se:*

$$f\left(\sum_{i=1}^{p} \alpha_i x^i\right) \leq \sum_{i=1}^{p} \alpha_i f(x^i)$$

Demonstração. Pela Definição A.5.2, $(x^i, f(x^i)) \in E_f$, para todo $i \in \{1, ..., p\}$. Pela Proposição A.5.2, E_f é convexo em $I\!R^n \times I\!R$. Pelo Teorema A.4, $\sum_{i=1}^p \alpha_i(x^i, f(x^i)) \in E_f$. Novamente, pela Definição A.5.2, $f\left(\sum_{i=1}^p \alpha_i x^i\right) \leq \sum_{i=1}^p \alpha_i f(x^i)$. ■

Teorema A.14 (Continuidade de Funções Convexas) *Sejam $C \subset I\!R^n$ um conjunto convexo aberto e $f : C \longrightarrow I\!R$ uma função convexa em C. Então f é localmente Lipschitz-contínua em D. Em particular, f é contínua em D.*

Demonstração. Seja $\bar{x} \in C$. Como C é aberto, existe $\delta > 0$ (que depende de $\bar{x}$) tal que $U \subset C$, onde $U = \{x \in I\!R^n / -\delta \leq x_i - \bar{x}_i \leq \delta, i = 1, ..., n\}$. Seja V o conjunto de vértices de U; V esta composto por 2^n pontos e $U = conv(V)$.

Seja $\beta = \underset{x \in V}{max}\, f(x)$, existe pois V tem um número finito de pontos. Pela convexidade de f, segue da Proposição A.5.3, que o conjunto de nível $L_{f,C}(\beta)$ é convexo. Como $V \subset L_{f,C}(\beta)$, segue-se que

$$U = conv(V) \subset conv(L_{f,C}(\beta)) = L_{f,C}(\beta) \tag{A.17}$$

Seja $x \in U$ tal que $0 < \|x - \bar{x}\| < \delta$. Defina

$$d := \frac{\delta(x - \bar{x})}{\|x - \bar{x}\|} = \frac{x - \bar{x}}{\alpha}, \text{ com } \alpha = \frac{\|x - \bar{x}\|}{\delta} \in (0, 1).$$

Dai, $|d_i| \leq \delta$, $\forall i = 1, ..., n$, e $-\delta \leq \bar{x}_i + d_i - \bar{x}_i \leq \delta$, $\forall i = 1, ..., n$. Assim, $\bar{x} + d \in U$ e $\bar{x} - d \in U$. Além disso,

$$x = \bar{x} + \frac{\alpha(x - \bar{x})}{\alpha} = \bar{x} + \alpha d = \alpha\bar{x} + \alpha d + \bar{x} - \alpha\bar{x} = \alpha(\bar{x} + d) + (1 - \alpha)\bar{x},$$

e pela convexidade de f,

$$f(x) \leq \alpha f(\bar{x} + d) + (1 - \alpha) f(\bar{x}) \leq \alpha\beta + (1 - \alpha) f(\bar{x}), \tag{A.18}$$

pois $\bar{x} + d \in U \subset L_{f,C}(\beta)$. De modo similar,

$$\bar{x} = x + \frac{\alpha(x - \bar{x})}{\alpha} = x + \alpha d = x + \alpha\bar{x} - \alpha d - \alpha\bar{x} = x + \alpha(\bar{x} - d) - \alpha\bar{x}.$$

Então $\alpha\bar{x} + \bar{x} = x + \alpha(\bar{x} - d)$. Assim $\bar{x} = \dfrac{1}{1+\alpha} x + \dfrac{\alpha}{1+\alpha}(\bar{x} - d)$.

Pela convexidade de f,

$$f(\bar{x}) \leq \frac{1}{1+\alpha} f(x) + \frac{\alpha}{1+\alpha} f(\bar{x} - d) \leq \frac{f(x) + \alpha\beta}{1+\alpha}, \tag{A.19}$$

pois $\bar{x} - d \in U \subset L_{f,C}(\beta)$. De (A.18), segue que $f(x) \leq \alpha\beta + f(\bar{x}) - \alpha f(\bar{x})$. Daí,

$$f(x) - f(\bar{x}) \leq \alpha(\beta - \alpha f(\bar{x})). \tag{A.20}$$

De (A.19), segue que $f(\bar{x}) + \alpha f(\bar{x}) \leq f(x) + \alpha\beta$. Daí,

$$-\alpha\beta + \alpha f(\bar{x}) \leq f(x) - f(\bar{x}). \tag{A.21}$$

De (A.21) e (A.20), $-\alpha(\beta - f(\bar{x})) \leq f(x) - f(\bar{x}) \leq \alpha(\beta - \alpha f(\bar{x}))$. Então

$$| f(x) - f(\bar{x})| \leq \alpha(\beta - f(\bar{x})) = M \|x - \bar{x}\|,$$

onde $M = \delta^{-1}(\beta - f(\bar{x}))$ depende de $\bar{x}$ e da escolha de δ, mas não depende de $x \in \{x \in I\!R^n / 0 < \| x - \bar{x} \| < \delta\}$. Portanto, f é localmente Lipschitz-contínua em D. ■

A.6 Subgradiente de uma Função

Definição A.6.1 *Seja $f : I\!R^n \longrightarrow I\!R$ uma função. O vetor $s \in I\!R^n$ é um subgradiente de f no ponto $x \in I\!R^n$ se*

$$f(y) \geq f(x) + \langle s, y - x \rangle, \tag{A.22}$$

para todo $y \in I\!R^n$. O conjunto de todos os subgradientes de f em x, denotado por $\partial f(x)$, é chamado o subdiferencial de f em x.

Observação A.6.1 *Seja $s \in \partial f(x)$. Considere $z \in I\!R^n$ como $z = x + \alpha d$, e $\alpha > 0$. Pela Definição A.6.1, tem-se que*

$$\begin{aligned}
s \in \partial f(x) \quad &\Leftrightarrow \quad \langle s, z - x \rangle \leq f(z) - f(x), \ \forall z \in I\!R^n \\
&\Leftrightarrow \quad f(x + \alpha d) - f(x) \geq \langle s, \alpha d \rangle, \ \forall d \in I\!R^n \ e \ \forall \alpha > 0 \\
&\Leftrightarrow \quad \frac{f(x + \alpha d) - f(x)}{\alpha} \geq \langle s, d \rangle, \ \forall d \in I\!R^n \ e \ \forall \alpha > 0 \\
&\Leftrightarrow \quad \lim_{\alpha \to 0^+} \frac{f(x + \alpha d) - f(x)}{\alpha} \geq \langle s, d \rangle, \ \forall d \in I\!R^n \\
&\Leftrightarrow \quad \frac{\partial f}{\partial d}(x) \geq \langle s, d \rangle, \ \forall d \in I\!R^n.
\end{aligned}$$

Assim, são equivalentes as seguintes definições do subdiferencial de f em x

$$\begin{aligned}
\partial f(x) &:= \{ s \in I\!R^n \ / \ f(z) \geq f(x) + \langle s, z - x \rangle, \ \forall z \in I\!R^n \} \\
&:= \left\{ s \in I\!R^n \ / \ \frac{\partial f}{\partial d}(x) \geq \langle s, d \rangle, \ \forall d \in I\!R^n \right\}.
\end{aligned}$$

Proposição A.6.1 *Seja $f : I\!R^n \rightarrow I\!R$ uma função convexa. Então o conjunto $\partial f(x)$ é não vazio, convexo e compacto, para todo $x \in I\!R^n$.*

Demonstração. Pelo Teorema A.11, para todo $x \in I\!R^n$, existe $s = s(x) \in I\!R^n$ tal que a desigualdade (A.22) é válida. Logo, $\partial f(x) \neq \phi$, $x \in I\!R^n$.

Sejam $s \in \partial f(x)$, $w \in \partial f(x)$ e $t \in [0,1]$. Então, $f(x) + \langle s, y-x\rangle \leq f(y)$ e $f(x) + \langle w, y-x\rangle \leq f(y)$, $y \in I\!R^n$. Logo,

$$\begin{aligned}
f(x) + \langle (1-t)s + tw, y-x\rangle &= f(x) + (1-t)\langle s, y-x\rangle + t\langle w, y-x\rangle \\
&= f(x) - tf(x) + tf(x) + (1-t)\langle s, y-x\rangle + t\langle w, y-x\rangle \\
&= (1-t)\left(f(x) + \langle s, y-x\rangle\right) + t\left(f(x) + \langle s, y-x\rangle\right) \\
&\leq (1-t)f(y) + tf(y) = f(y), \quad \forall y \in I\!R^n.
\end{aligned}$$

Assim, $(1-t)s + tw \in \partial f(x)$, $\forall t \in [0,1]$. Portanto, $\partial f(x)$ é convexo para todo $x \in I\!R^n$.

Resta mostrar que para cada $x \in I\!R^n$, $\partial f(x)$ é fechado e limitado. De fato.

Seja $s \in cl(\partial f(x))$. Então existe seqüência $(s_k)_{k\in I\!N}$ em $\partial f(x)$ tal que $\lim\limits_{k\to+\infty} s_k = s$. Como $s_k \in \partial f(x)$, $\forall k \in I\!N$, então $f(y) \geq f(x) + \langle s_k, y-x\rangle$, $\forall k \in I\!N$ e $\forall y \in I\!R^n$. Logo, $f(y) \geq f(x) + \lim\limits_{k\to+\infty}\langle s_k, y-x\rangle = f(x) + \langle s, y-x\rangle$, $\forall y \in I\!R^n$. Assim, $y \in \partial f(x)$. Portanto, $\partial f(x)$ é fechado para todo $x \in I\!R^n$.

Seja agora $s \in \partial f(x)$, $s \neq 0$, e considere $\delta > 0$ tal que $y = x + \delta\frac{s}{\|s\|}$ pertença ao compacto $B[x,\delta]$. Pelo Teorema A.14, f é localmente Lipschitz-contínua. Logo, para $y \in B[x,\delta]$, existe uma constante de Lipschitz $M > 0$ tal que $f(y) - f(x) \leq M\ \|y - x\|$. Daí,

$$f(y) - f(x) \leq M\|x + \delta\frac{s}{\|s\|} - x\| = M\delta. \tag{A.23}$$

Por outro lado, como $s \in \partial f(x)$ então $f(y) \geq f(x) + \langle s, y-x\rangle$. Daí,

$$f(y) \geq f(x) + \langle s, \delta\frac{s}{\|s\|}\rangle = f(x) + \delta\|s\|. \tag{A.24}$$

De (A.24) e (A.23), $\delta\|s\| \leq f(y) - f(x) \leq M\delta$. Isto implica que $\|s\| \leq M$. Portanto, $\partial f(x)$ é limitado para todo $x \in I\!R^n$. ■

Proposição A.6.2 *Seja $f : I\!R^n \rightarrow I\!R$ uma função convexa. Então, para todo $d \in I\!R^n$, tem-se*

$$\frac{\partial f}{\partial d}(x) = \max_{s\in\partial f(x)} \langle s, d\rangle$$

Demonstração. Por exemplo, veja Teorema 3.4.12 em [14]. ∎

Proposição A.6.3 *Uma função convexa $f : I\!R^n \to I\!R$ é diferenciável no ponto $x \in I\!R^n$ se, e somente se, o conjunto $\partial f(x)$ contém um único elemento. Neste caso, $\partial f(x) = \{\nabla f(x)\}$.*

Demonstração. Suponha que f é diferenciável em $x \in I\!R^n$. Então pelo Teorema A.9, $f(z) \geq f(x) + \langle \nabla f(x), z - x \rangle$, $z \in I\!R^n$. Assim, $\nabla f(x) \in \partial f(x)$. Seja $s \in \partial f(x)$. Logo, $f(z) \geq f(x) + \langle s, z - x \rangle$, $z \in I\!R^n$. Considere $\lambda d = z - x$, com $\lambda > 0$. Então $f(z) \geq f(x) + \langle s, \lambda d \rangle$. Como f é diferenciável em x,

$$f(x + \lambda d) \geq f(x) + \langle \nabla f(x), \lambda d \rangle + \|\lambda d\| r(\lambda d), \quad \lim_{\lambda \to 0} r(\lambda d) = 0.$$

Logo, $f(x) + \langle s, \lambda d \rangle \leq f(x) + \langle \nabla f(x), \lambda d \rangle + \|\lambda d\| r(\lambda d)$. Então, $\langle s, d \rangle \leq \langle \nabla f(x), d \rangle + \|d\| \lim_{\lambda \to 0} r(\lambda d)$. Assim, $\langle s - \nabla f(x), d \rangle \leq 0$, para todo $d \in I\!R^n$.

Em particular, quando $d = s - \nabla f(x)$, tem-se que $0 \leq \|s - \nabla f(x)\|^2 \leq 0$. Logo, $s = \nabla f(x)$. Portanto, $\partial f(x) = \{s\}$.

Reciprocamente, Suponha que $\partial f(x) = \{s\}$. Pela Proposição A.6.2, $\dfrac{\partial f}{\partial d}(x) = \langle s, d \rangle$, $\forall d \in I\!R^n$. Escolhendo d como os elementos da base canônica de $I\!R^n$, tem-se que $s_i = \partial f(x)/\partial x_i$, $i = 1, ..., n$. Assim $\dfrac{\partial f}{\partial d}(x) = \langle \nabla f(x), d \rangle$, $\forall d \in I\!R^n$, o que implica a diferenciabilidade de f em x. ∎

Teorema A.15 *Seja $f : I\!R^n \to I\!R$ uma função convexa. O ponto $\bar{x} \in I\!R^n$ é minimizador de f se, e somente se, $0 \in \partial f(\bar{x})$.*

Demonstração. Suponha que $\bar{x} \in I\!R^n$ é minimizador de f. Então, $f(\bar{x}) \leq f(x)$, $\forall x \in I\!R^n$. Logo, $f(x) \geq f(\bar{x}) + \langle 0, x - \bar{x} \rangle$, $\forall x \in I\!R^n$. Pela Definição A.6.1, $0 \in \partial f(\bar{x})$.

Reciprocamente, suponha que $0 \in \partial f(\bar{x})$. Então pela Definição A.6.1, $f(x) \geq f(\bar{x}) + \langle 0, x - \bar{x} \rangle = f(\bar{x})$, $\forall x \in I\!R^n$. Logo, $\bar{x} \in I\!R^n$ é minimizador de f. ∎

A.7 Função Conjugada de uma Função Convexa

Uma estenção da Definição A.5.1 é dada pela seguinte definição.

Definição A.7.1 *Uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *não identicamente* $+\infty$, *é chamada convexa quando, para todo* $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ *e todo* $\alpha \in (0, 1)$, *é valido*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \tag{A.25}$$

como uma desigualdade em $\mathbb{R} \cup \{+\infty\}$.

Observação A.7.1 *Seja* $C \subset \mathbb{R}^n$ *um conjunto convexo. Uma função convexa* $f : C \to \mathbb{R}$ *como da Definição A.5.1 pode ser estendida a uma função convexa como na definição A.7.1 da forma seguinte*

$$\begin{array}{rcl} \widehat{f} : \mathbb{R}^n & \to & \mathbb{R} \cup \{+\infty\} \\ x & \mapsto & \widehat{f}(x) \quad := \left\{ \begin{array}{ll} f(x), & se \quad x \in C \\ +\infty, & se \quad x \neq C \end{array} \right. . \end{array}$$

Definição A.7.2 *Uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é chamada convexa propria se* $f(x) < +\infty$ *para pelo menos un ponto* $x \in \mathbb{R}^n$, *e* $f(x) > -\infty$ *para todo* $x \in \mathbb{R}^n$.

Definição A.7.3 *O domínio efetivo de uma função convexa* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *denotado por* $ED(f)$, *é o conjunto* $\{x \in \mathbb{R}^n / f(x) < +\infty\}$.

Definição A.7.4 *Uma função* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *é chamada fechada se seu epígrafo é fechado em* $\mathbb{R}^n \times \mathbb{R}$; *ou equivalentemente, se seus conjuntos de nível são fechados.*

Definição A.7.5 *Seja* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa, não identicamente* $+\infty$. *A função* $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, *definida por*

$$f^*(s) = \sup_{x \in \mathbb{R}^n} \{\langle x, s\rangle - f(x)\}, \tag{A.26}$$

para todo $s \in \mathbb{R}^n$, *é chamada a conjugada da função* f.

Observação A.7.2 *Sejam* $x \in \mathbb{R}^n$ *e* $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ *uma função convexa. Tem-se que* $\partial f(x) = \{s \in \mathbb{R}^n / \langle s, z - x\rangle \leq f(z) - f(x)\}, \forall z \in \mathbb{R}^n$. *Daí,* $\langle s, z\rangle - f(z) \leq \langle s, x\rangle - f(x) \leq f^*(s)$. *Logo,* $\forall s \in \mathbb{R}^n$ *tem-se que* $f^*(s) = \sup_{x \in \mathbb{R}^n} \{\langle x, s\rangle - f(x)\}$.

Exemplo A.7.1 *Seja f(x)=* $|x|$, $\forall x \in \mathbb{R}$. *Tem-se que*

$$\partial f(x) = \{s \in \mathbb{R} / s(z - x) \leq |z| - |x|, \forall z \in \mathbb{R}\}.$$

Se $x > 0$, *então* $sz - sx \leq |z| - x$; *logo* $s = 1$, *pois* $1.z - 1.x \leq |z| - x$ *implica* $z \leq |z|$. *Se* $x < 0$, *então* $sz - sx \leq |z| + x$; *logo* $s = -1$, *pois* $-1.z + 1.x \leq |z| + x$ *implica* $-z \leq |z|$. *Se* $x = 0$, *então* $sz \leq |z|$; *logo* $s \in [0, 1]$. *Assim, o subdiferencial é dado por*

$$\partial f(x) = \left\{ \begin{array}{rl} \{1\}, & se \quad x > 0 \\ [-1, 1], & se \quad x = 0 \\ \{-1\}, & se \quad x < 0 \end{array} \right.$$

Portanto, a função conjugada é dada por

$$f^*(s) = \sup_{x\in \mathbb{R}^n} \{\ xs - |x|\ \} = \begin{cases} +\infty, & se \quad |x| > 1 \\ 0, & se \quad |x| \leq 1 \end{cases}$$

Observação A.7.3 *Segue, da Definição* A.7.5*, que para cada $s \in \mathbb{R}^n$, tem-se que $f^*(s) + f(x) \geq \langle s, x\rangle$, $\forall x \in \mathbb{R}^n$. Esta relação é chamada a "desigualdade de Fenchel".*

Teorema A.16 *Sejam $x \in \mathbb{R}^n$ e $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa, não identicamente $+\infty$. Então $s \in \partial f(x)$ se, e somente se, $f^*(s) = \langle s, x\rangle - f(x)$.*

Demonstração. Suponha que $s \in \partial f(x)$. Então $\langle s, z\rangle - f(z) \leq \langle s, x\rangle - f(x)$, $\forall z \in \mathbb{R}^n$. Logo, $f^*(s) = \sup_{z\in \mathbb{R}^n} \{\langle z, s\rangle - f(z)\} \leq \langle s, x\rangle - f(x)$. Portanto, $f^*(s) = \langle s, x\rangle - f(x)$.

Reciprocamente, suponha que $f^*(s) = \langle s, x\rangle - f(x)$. Então $f^*(s) < +\infty$. Logo, $\langle s, z\rangle - f(z) \leq \langle s, x\rangle - f(x)$, $\forall z \in \mathbb{R}^n$. Assim, $\langle s, z - x\rangle \leq f(z) - f(x)$, $\forall z \in \mathbb{R}^n$. Portanto, $s \in \partial f(x)$. ■

Observação A.7.4 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa, não identicamente $+\infty$. Segue, do Teorema A.17, que se $s \in \partial f(x)$, para algum $x \in \mathbb{R}^n$, então $f^* < +\infty$.*

Teorema A.17 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função convexa, não identicamente $+\infty$. Então a conjugada f^* é uma função convexa.*

Demonstração. Sejam s_1 e s_2 em $\mathbb{R}^n$. Se $f^*(s_1) < +\infty$ e $f^*(s_2) < +\infty$, então para $t \in [0, 1]$ tem-se

$$\begin{aligned} f^*((1-t)s_1 + ts_2) &= \sup_{x\in \mathbb{R}^n} \{\langle x, (1-t)s_1 + ts_2\rangle - f(x)\} \\ &= \sup_{x\in \mathbb{R}^n} \{(1-t)\langle x, s_1\rangle + t\langle x, s_2\rangle - f(x) + tf(x) - tf(x)\} \\ &= \sup_{x\in \mathbb{R}^n} \{(1-t)\,[\langle x, s_1\rangle - f(x)] + t\,[\langle x, s_2\rangle - f(x)]\} \\ &\leq (1-t) \sup_{x\in \mathbb{R}^n} \{\langle x, s_1\rangle - f(x)\} + t \sup_{x\in \mathbb{R}^n} \{\langle x, s_2\rangle - f(x)\} \\ &= (1-t)f^*(s_1) + tf^*(s_2). \end{aligned}$$

Se $f^*(s_1) = +\infty$ ou $f^*(s_2) = +\infty$ então, de forma trivial, $f^*((1-t)s_1 + ts_2) \leq (1-t)f^*(s_1) + tf^*(s_2)$. Portanto, f^* é uma função convexa. ■

Definição A.7.6 *Seja* $f : I\!R^n \to I\!R \cup \{+\infty\}$ *uma função convexa, não identicamente* $+\infty$. *A função* $f^{**} : I\!R^n \to I\!R \cup \{+\infty\}$ *é chamada a conjugada da conjugada de* f.

Proposição A.7.1 *Seja* $f : I\!R^n \to I\!R \cup \{+\infty\}$ *uma função convexa, não identicamente* $+\infty$. *Então* f^{**} *é o supremo do conjunto de todas as funções afins que minoran* f.

Demonstração. Seja A o conjunto de todas as funções afins que minoran f. Considere $\Gamma = sup\,\{g/g \in A\}$. Mostraremos que $f^{**} = \Gamma$. De fato.

Para cada $s \in I\!R^n$, a função $g(x) = \langle x, s\rangle - f^*(s)$, $\forall x \in I\!R^n$, é uma função afim. Pela desigualdade de Fenchel, $f(x) + f^*(s) \geq \langle x, s\rangle$, $x \in I\!R^n$. Logo, $f(x) \geq \langle x, s\rangle - f^*(s) = g(x)$, $x \in I\!R^n$. Daí, g é uma função afim que minora f. Assim, $g \in A$. Logo, para cada $x \in I\!R^n$, tem-se

$$f^{**}(x) = \underset{s\in I\!R^n}{sup}\ \{\langle x, s\rangle - f^*(s)\} \leq \Gamma(x). \tag{A.27}$$

Seja $h \in A$. Então para algum $s \in I\!R^n$ e algum $\alpha \in I\!R$, $h(x) = \langle x, s\rangle - \alpha$, $\forall x \in I\!R^n$. Como h minora f, $\langle x, s\rangle - \alpha \leq f(x)$, $\forall x \in I\!R^n$. Logo, $\alpha \geq \langle x, s\rangle - f(x)$ $\forall x \in I\!R^n$. Daí, $\alpha \geq \underset{sx\in I\!R^n}{sup}\ \{\langle x, s\rangle - f(x)\} = f^*(s)$. Assim, $h(x) = \langle x, s\rangle - \alpha \leq \langle x, s\rangle - f^*(s)$, $x \in I\!R^n$. Logo, para cada $x \in I\!R^n$, tem-se

$$\Gamma(x) = \underset{x\in I\!R^n}{sup}\ \{\langle x, s\rangle - \alpha\} \leq \underset{x\in I\!R^n}{sup}\ \{\langle x, s\rangle - f^*(s)\} = f^{**}(x) \tag{A.28}$$

De (A.27) e (A.28) segue que $\Gamma(x) = f^{**}(x)$, $\forall x \in I\!R^n$. Portanto, $f^{**} = \Gamma$. ■

Proposição A.7.2 *Seja* $f : I\!R^n \to I\!R \cup \{+\infty\}$ *uma função convexa, não identicamente* $+\infty$. *Então* $f = f^{**}$.

Demonstração. Mostraremos primeiro que $f^{**} \leq f$. De fato. Pela Definição A.7.6, $f^{**}(w) = \underset{y\in I\!R^n}{sup}\ \{\langle w, y\rangle - f^*(y)\}$, $\forall w \in I\!R^n$. Pela Definição A.7.5, para cada $y \in I\!R^n$, tem-se que $f^*(y) = \underset{x\in I\!R^n}{sup}\ \{\langle y, x\rangle - f(x)\} \geq \langle y, w\rangle - f(w)$, $\forall w \in I\!R^n$. Logo, $\langle w, y\rangle - f^*(y) \leq f(w)$, $\forall w \in I\!R^n$. Assim, $f^{**}(w) = \underset{y\in I\!R^n}{sup}\ \{\langle w, y\rangle - f^*(y)\} \leq f(w)$, $\forall w \in I\!R^n$. Logo, $f^{**} \leq f$.

Mostraremos agora que $f \leq f^{**}$. Suponha, pelo absurdo, que existe $x_0 \in I\!R^n$ tal que $f(x_0) > f^{**}(x_0)$. Logo, existe $\alpha \in I\!R$ tal que $f^{**}(x_0) < \alpha < f(x_0)$. Como f é convexa então, pelo Teorema A.11, existe uma função afim h minorando f tal que $h(x_0) = \alpha$. Defina $h(x) = \alpha + \langle s, x_0 - x\rangle$, $\forall x \in I\!R^n$. Pela Proposição A.7.1, $f^{**} \geq h$. Daí $f^(**)(x_0) \geq h(x_0) = \alpha$, o que é um absurdo. Logo, $f \leq f^{**}$.

Portanto, $f = f^{**}$.

Apêndice B

Condições de Otimalidade de Karush-Kuhn-Tucker

Neste apêndice, é apresentado uma ferramenta poderosa para dar solução a problemas de otimização, as chamadas Condições Necessárias de Otimalidade de Karush-Kuhn-Tucker. Além disso, é analizado a forma destas condições para problemas de programação linear, as quais são relevantes na seções 2.4 e 4.2. Um estudo mais profundo deste apêndice pode ser encontrado em Bazaraa [2].

B.1 Condições Necessárias de Otimalidade de Karush-Kuhn-Tucker

Sejam $X \subset I\!R^n$ um conjunto não vazio aberto e $f : I\!R^n \to I\!R$ e $g_i : I\!R^n \to I\!R$, $i = 1, ..., m$, funções. Considere o problema

$$(P) \left\{ \begin{array}{lll} min & f(x) & \\ s.a. & g_i(x) \leq 0 \,, & i = 1, ..., m \\ & x \in X \,. & \end{array} \right.$$

Nesta seção são presentadas as condições que devem ser satisfeitas quando um $\bar{x} \in I\!R^n$ dado é minimizador local do problema (P). Condições deste tipo se chaman condições necessárias de otimalidade.

Teorema B.1 *Seja $\bar{x}$ um ponto factível do problema (P) e denote $I = \{\ i \ / \ g_i(\bar{x}) = 0\}$. Suponha que f e g_i são diferenciáveis em $\bar{x}$ para todo $i \in I$, e que g_i é contínua em*

$\bar{x}$ para todo $i \notin I$. Além disso, suponha que $\nabla g_i(\bar{x})$, com $i \in I$, sejam linearmente independientes. Se $\bar{x}$ é uma solução local do problema (P), então existem escalares u_i, com $i \in I$, tais que

$$\begin{cases} \nabla f(\bar{x}) \;+\; \displaystyle\sum_{i \in I} u_i \nabla g_i(\bar{x}) \;=\; 0\,, \\ \qquad\qquad\qquad\qquad u_i \;\geq\; 0\,, \quad i \in I\,. \end{cases}$$

Em adição, se g_i, com $i \notin I$, são também diferenciáveis, então existem escalares u_i, com $i = 1, ..., m$, tais que

$$\begin{cases} \nabla f(\bar{x}) \;+\; \displaystyle\sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) \;=\; 0\,, \\ \qquad\qquad\qquad u_i g_i(\bar{x}) \;=\; 0\,, \quad i = 1, ..., m\,. \\ \qquad\qquad\qquad\qquad u_i \;\geq\; 0\,, \quad i = 1, ..., m\,. \end{cases}$$

Demonstração. Veja Teorema 4.2.13 em Bazaraa-Sherali-Shetty [2]. ∎

Os escalares u_i são chamados *"Multiplicadores de Lagrange"*.

O requerimento que $\bar{x}$ seja factível ao problema (P) é chamado *"Factibilidade Primal"* (PF).

O requerimento $\nabla f(\bar{x}) + \displaystyle\sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) = 0$, $u_i \geq 0$, para $i = 1, ..., m$, é chamado *"Factibilidade Dual"* (DF).

O requerimento $u_i g_i(\bar{x}) = 0$, para todo $i = 1, ..., m$, é chamado *"Condição Complementar de Folga"* (CS).

As condições PF, DF e CS são chamadas *"Condições Necessárias de Otimalidade de Karush - Kuhn - Tucker"* (KKT).

Observação B.1.1 *As condições necessárias de otimalidade KKT podem ser escritas em forma vectorial como*

$$\begin{cases} PF: & g(\bar{x}) \;\leq\; 0\,, \quad \bar{x} \;\geq 0\,. \\ DF: & \nabla f(\bar{x}) + \nabla g(\bar{x})^t u \;=\; 0\,, \quad u \;\geq 0\,. \\ CS: & u^t g(\bar{x}) \;=\; 0\,, \end{cases}$$

onde $g = (g_1, ..., g_m)$, $u = (u_1, ..., u_m)$ e $\nabla g(\bar{x})^t$ é uma matriz $n \times m$ cuja i-ésima coluna é $\nabla g_i(\bar{x})$ (isto é, a trasposta do Jacobiano de g em $\bar{x}$).

B.2 Condições Necessárias de Otimalidade de KKT em Programação Linear

Considere o problema de programção linear

$$(P)\left\{\begin{array}{lrcl} min & c^t x & & \\ s.a. & Ax & = & b \\ & x & \geq & 0 \end{array}\right.$$

onde $c \in I\!R^n$, $x \in I\!R^n$, $A = (a_{ij})_{m\times n}$ e $b \in I\!R^m$.

Observe que $Ax = b \Leftrightarrow -Ax \leq -b$ e $Ax \leq b$. Logo, o problema (P) é equivalente ao problema:

$$(P')\left\{\begin{array}{lrcl} min & c^t x & & \\ s.a. & -Ax & \leq & -b \\ & Ax & \leq & b \\ & -x & \leq & 0 \end{array}\right.$$

Defina

$$\begin{array}{rcccl} f(x) & := & c^t x & , & x \in I\!R^n. \\ g_1(x) & := & -Ax + b & , & x \in I\!R^n. \\ g_2(x) & := & Ax - b & , & x \in I\!R^n. \\ g_3(x) & := & -x & , & x \in I\!R^n. \end{array}$$

Denotemos por y^+, y^- e v como os multiplicadores de Lagrange para g_1, g_2 e g_3, respectivamente. Logo, as condições de otimalidade KKT para o problema (P') são

$$\left\{\begin{array}{ll} PF: & Ax = b \ , \ x \geq 0. \\ DF: & \nabla f(x) + y^+ \nabla g_1(x) + y^- \nabla g_2(x) + v \nabla g_3(x) = 0 \ ; \ y^+ \geq 0, y^- \geq 0, v \geq 0. \\ CS: & (b - Ax)^t y^+ = 0 \ , \ (Ax - b)^t y^- = 0 \ , \ -x^t v = 0. \end{array}\right.$$

Equivalentemente

$$\left\{\begin{array}{ll} PF: & Ax = b \ , \ x \geq 0. \\ DF: & c - A^t y^+ + A^t y^- + (-1)v = 0 \ ; \ y^+ \geq 0, \ y^- \geq 0, \ v \geq 0. \\ CS: & (b - b)^t y^+ = 0 \ , \ (b - b)^t y^- = 0 \ , \ x^t v = 0. \end{array}\right.$$

Da misma forma

$$\begin{cases} PF: & Ax = b \ , \ x \geq 0. \\ DF: & c - A^t(y^+ - y^-) - v = 0 \ ; \ y^+ \geq 0, \ y^- \geq 0, \ v \geq 0. \\ CS: & x^t v = 0. \end{cases}$$

Ou ainda

$$\begin{cases} PF: & Ax = b \ , \ x \geq 0. \\ DF: & c - A^t y - v = 0 \ ; \ y = (y^+ - y^-) \in I\!R^m, \ v \geq 0. \\ CS: & x^t v = 0. \end{cases}$$

Resumindo, as condições de otimalidade KKT para o problema de programação linear (P) são dadas por

$$\begin{cases} PF: & Ax = b \ , \ x \geq 0. \\ DF: & c = A^t y + v \ ; \ y \in I\!R^m, \ v \geq 0. \\ CS: & x_j v_j = 0 \ , \ j = 1, ..., n. \end{cases}$$

Apêndice C

Elementos de Geometria Riemanniana

Neste apêndice, apresentaremos alguns resultados e definições relativos às variedades diferenciáveis, os quais são relevantes no quarto capítulo. Mais detalhes em Manfredo [8] e em [7].

C.1 Variedades Diferenciáveis

Definição C.1.1 *Uma variedade diferenciável de dimensão n é um conjunto M e uma família de aplicações biunívocas $x_\alpha : U_\alpha \subset \mathbb{R}^n \to M$ de abertos U_α de $\mathbb{R}^n$ em M tais que:*

1) $\bigcup_\alpha x_\alpha(U_\alpha) = M$;

2) *Para todo par (α, β) com $x_\alpha(U_\alpha) \bigcap x_\beta(U_\beta) = W \neq \phi$, os conjuntos $x_\alpha^{-1}(W)$ e $x_\beta^{-1}(W)$ são abertos em $\mathbb{R}^n$ e as aplicações $x_\beta^{-1} \circ x_\alpha : x_\alpha^{-1}(W) \to x_\beta^{-1}(W)$ são diferenciáveis;*

3) *A família $F = \{(U_\alpha, x_\alpha)\}$ é máxima relativamente às condições (1) e (2).*

O par (U_α, x_α), ou a aplicação x_α, é chamado uma parametrização, um sistema de coordenadas, ou uma carta local, de M em torno do ponto $p \in x_\alpha(U_\alpha)$. Uma família $\{(U_\alpha, x_\alpha)\}$ satisfazendo as condições **(1)** *e* **(2)** *é chamada atlas, ou estrutura diferenciável de M.*

Observação C.1.1 *O atlas da variedade diferenciável M faz dela um espaço topológico, onde $A \subset M$ é aberto se $x^{-1}(A \cap x(U))$ é um aberto de $\mathbb{R}^n$, qualquer que seja a parametrização $x : U \subset \mathbb{R}^n \to M$.*

Definição C.1.2 *Dizemos que uma família* $\{f_\alpha\}$ *de funções diferenciáveis* $f_\alpha : M \to \mathbb{R}$ *é uma partição diferenciável da unidade se:*

(i) Para todo α, $f_\alpha \geq 0$ *e o suporte de* f_α *está contido em uma vizinhança coordenada* $V_\alpha = x_\alpha(U_\alpha)$ *de uma estrutura diferenciável* $\{(U_\alpha, x_\alpha)\}$ *de* M*;*

(ii) A família $\{V_\alpha\}$ *é localmente finita;*

(iii) $\sum_\alpha f_\alpha(p) = 1$, *para todo* $p \in M$.

Dizemos que a partição $\{f_\alpha\}$ *da unidade está subordinada à cobertura* $\{V_\alpha\}$.

Restrições quanto à topologia de M:

Axioma de Hausdorff. Dados dois pontos distintos de M, existem vizinhanças destes dois pontos as quais não se intersectam.

Axioma de base enumerável. A variedade M pode ser coberta por uma quantidade enumerável de vizinhanças coordenadas.

A variedade M é conexa.

Estas considerações são importantes devido ao resultado abaixo.

Teorema C.1 *Uma variedade diferenciável M possui uma partição diferenciável da unidade se, e somente se, toda componente conexa de M é de Hausdorff e possui base enumerável.*

C.2 Espaço Tangente

Definição C.2.1 *Sejam M_1 e M_2 variedades diferenciáveis de dimensões n e m, respectivamente. Uma aplicação $\varphi : M_1 \to M_2$ é dita ser diferenciável em $p \in M_1$ se, dada uma parametrização $y : V \subset \mathbb{R}^m \to M_2$ em $\varphi(p)$, existe uma parametrização $x : U \subset \mathbb{R}^n \to M_1$ em p tal que $\varphi(x(U)) \subset y(V)$ e a aplicação*

$$y^{-1} \circ \varphi \circ x : U \subset \mathbb{R}^n \to \mathbb{R}^m$$

é diferenciável em $x^{-1}(p)$. Mais ainda, diz-se que φ é diferenciável em um aberto de M_1 se é diferenciável em todos os pontos desse aberto.

Definição C.2.2 *Seja M uma variedade diferenciável. Uma aplicação diferenciável $\alpha : (-\varepsilon, \varepsilon) \to M$ é chamada uma curva (diferenciável) em M. Suponha que $\alpha(0) =$*

$p \in M$, e seja $C^\infty(M)$ o conjunto das funções $f : M \to \mathbb{R}$ diferenciáveis em p. O vetor tangente à curva α em $t = 0$ é a função $\alpha'(0) : C^\infty(M) \to \mathbb{R}$ dada por

$$\alpha'(0)f = \left.\frac{d(f \circ \alpha)}{dt}\right|_{t=0}, \quad f \in C^\infty(M). \tag{C.1}$$

Um vetor tangente em $p \in M$ é um vetor tangente, em $t = 0$, a alguma curva $\alpha : (-\varepsilon, \varepsilon) \to M$ com $\alpha(0) = p$. O conjunto dos vetores tangentes a M em p será denotado por T_pM.

De acordo com as notações na definição C.2.2, escolha uma parametrização $x : U \to M$ em $p = x(q)$ e escreva

$$\begin{aligned} & (x^{-1} \circ \alpha)(t) = (x_1(t), \cdots, x_n(t)) \\ \text{e} \quad & (f \circ x)(q) = \tilde{f}(x_1, \cdots, x_n), \; q = (x_1, \cdots, x_n) \in U. \end{aligned}$$

Assim,

$$\alpha'(0)f = \left.\frac{d}{dt}(f \circ \alpha)\right|_{t=0} = \left.\frac{d}{dt}\tilde{f}(x_1(t), \cdots, x_n(t))\right|_{t=0} = \sum_i x_i'(0)\frac{\partial \tilde{f}}{\partial x_i}(q).$$

Portanto, o vetor $\alpha'(0)$ pode ser expresso, através da parametrização x, por

$$\alpha'(0) = \sum_i x_i'(0)\frac{\partial}{\partial x_i}, \tag{C.2}$$

onde $\dfrac{\partial}{\partial x_i}$ denota o vetor pertencente ao conjunto T_pM tal que

$$\frac{\partial}{\partial x_i}(f) = \frac{\partial (f \circ x)}{\partial x_i}(x^{-1}(p)), \quad \forall f \in C^\infty(M). \tag{C.3}$$

Observação C.2.1 *Com as operações usuais de funções, nota-se que T_pM é espaço vetorial. Além disso, escolhida uma parametrização $x : U \to M$ em $p = x(q)$, pode-se ver que o conjunto $\{\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n}\}$ é uma base de T_pM, a qual chamamos de base associada à parametrização (U, x).*

Definição C.2.3 *O espaço vetorial T_pM é chamado espaço tangente de M em p.*

Definição C.2.4 *Sejam M^m e N^n variedades diferenciáveis. Uma aplicação $\varphi : M \to N$ é uma imersão se $d_{\varphi_p} : T_pM \to T_{\varphi(p)}N$ é injetiva para todo $p \in M$. Se, além disso, φ é um homeorfismo sobre $\varphi(M) \subset N$, onde $\varphi(M)$ tem a topologia induzida por N, diz-se que φ é um mergulho. Se $M \subset N$ e a inclusão $i : M \to N$ é um mergulho, diz-que M é uma subvariedade de N.*

Proposição C.2.1 *Sejam M_1 e M_2 variedades de dimensões n e m, respectivamente. Seja $\varphi : M_1 \to M_2$ uma aplicação diferenciável. Para cada $p \in M$ e cada $v \in T_p(M_1)$, escolha uma curva diferenciável $\alpha : (-\varepsilon, \varepsilon) \to M_1$ com $\alpha(0) = p$ e $\alpha'(0) = v$. Faça $\beta = \varphi \circ \alpha$. A aplicação $d\varphi_p : T_p(M_1) \to T_{\varphi(p)}(M_2)$ dada por $d\varphi_p(v) = \beta'(0)$ é uma aplicação linear que não depende da escolha de α.*

Definição C.2.5 *A aplicação linear $d\varphi_p$ dada pela Proposição C.2.1 é chamada diferencial de φ em p.*

Definição C.2.6 *Sejam M_1 e M_2 variedades diferenciáveis. Uma aplicação φ de M_1 em M_2 é um difeomorfismo se ela é diferenciável, biunívoca e sua inversa φ^{-1} é diferenciável.*

C.3 Campo de Vetores

Definição C.3.1 *Seja M uma variedade diferenciável. O conjunto $TM = \{(p, v); p \in M,\ v \in T_pM\}$ é chamado de fibrado tangente da variedade M.*

Proposição C.3.1 *O fibrado tangente TM é uma variedade diferenciável de dimensão $2n$, onde M tem dimensão n.*

Definição C.3.2 *Um campo de vetores X em uma variedade diferenciável M é uma correspondência que a cada ponto $p \in M$ associa um vetor $X(p) \in T_pM$. O campo é dito diferenciável se a aplicação $X : M \to TM$ é diferenciável.*

Notação. Denotaremos por $\mathfrak{X}(M)$ o conjunto dos campos de vetores diferenciáveis em M. Também é conveniente pensar em um campo de vetores como uma aplicação $X : C^\infty(M) \to C^\infty(M)$ definida por

$$(Xf)(p) = \sum_i a_i(p) \frac{\partial f}{\partial x_i}(p), \quad f \in C^\infty(M),$$

onde f indica, por um abuso de notação, a expressão de f na parametrização (U, x).

Lema C.1 *Sejam X e Y campos diferenciáveis de vetores em uma variedade diferenciável M. Então, existe um único campo vetorial Z tal que, para todo $f \in C^\infty(M)$, $Zf = (XY - YX)f$.*

Definição C.3.3 *O campo vetorial diferenciável Z dado no lema C.1 é chamado o colchete(de Lie) dos campos X e Y, e é denotado por $[X, Y]$; ou seja, escrevemos*

$$[X, Y] = XY - YX.$$

Definição C.3.4 *Um campo vetorial ao longo de uma curva* $c : I \to M$ *é uma aplicação que a cada* $t \in I$ *associa um vetor tangente* $V(t) \in T_pM$, *a qual é diferenciável no seguinte sentido: se* f *é uma função diferenciável em* M, *então a função* $t \mapsto V(t)f$ *é diferenciável em* I. *O campo vetorial* $dc\left(\frac{d}{dt}\right)$, *indicado por* $\frac{dc}{dt}$ *é chamado campo velocidade (ou tangente) de* c.

Observação C.3.1 *Sempre suporemos um atlas* $\{(U_\alpha, x_\alpha)\}$ *para* M *de modo que os campos* $X_i^\alpha = \frac{\partial}{\partial x_i^\alpha} \in \mathfrak{X}(x_\alpha(U_\alpha))$ *sejam passíveis de extensão em* M.

C.4 Métricas Riemannianas

Definição C.4.1 *Uma métrica Riemanniana ou estrutura Riemanniana em uma variedade diferenciável* M *é uma correspondência* g *que associa a cada ponto* $p \in M$ *um produto interno no espaço tangente* T_pM, *que varia diferenciavelmente no seguinte sentido: se* $x : U \subset \mathbb{R}^n \to M$ *é um sistema de coordenadas em torno de* p, *então, para cada* (i, j), *a função* $g_{ij} : U \to \mathbb{R}$ *definida por*

$$g_{ij}(x_1, \ldots, x_n) = \left\langle \frac{\partial}{\partial x_i}(q), \frac{\partial}{\partial x_j}(q) \right\rangle_q, \tag{C.4}$$

onde $q = x(x_1, \cdots, x_n)$, *é diferenciável.*

As funções g_{ij} *são chamadas expressões da métrica Riemanniana no sistema de coordenadas* (U, x). *Uma variedade diferenciável com uma dada métrica Riemanniana chama-se uma variedade Riemanniana.*

Proposição C.4.1 *Considere* $(g^{ij})_{n\times n}$ *a matriz inversa da matriz* $(g_{ij})_{n\times n}$. *Então,*

$$\frac{\partial g^{il}}{\partial x_j} = -\sum g^{ik} g^{ml} \frac{\partial g_{km}}{\partial x_j}. \tag{C.5}$$

Demonstração. Observe que

$$\sum_k g_{ik} g^{kl} = \delta_{il}.$$

Logo,

$$\frac{\partial}{\partial x_j}\left(\sum_k g_{ik} g^{kl}\right) = 0,$$

o que implica

$$\sum_k \frac{\partial g_{ik}}{\partial x_j} g^{kl} = -\sum_k g_{ik} \frac{\partial g^{kl}}{\partial x_j}.$$

Denotando-se $D = \left(\dfrac{\partial g_{ik}}{\partial x_j}\right)_{n\times n}$, $B = (g_{ik})_{n\times n}$ e $C = \left(\dfrac{\partial g^{ik}}{\partial x_j}\right)_{n\times n}$, temos, a partir da última iguadade,

$$DB^{-1} = -BC.$$

Assim,

$$C = -B^{-1}DB^{-1}.$$

Daí, efetuando-se o produto no lado direito da igualdade, pode-se concluir (C.5). ∎

Definição C.4.2 *O comprimento ou norma de um vetor tangente $u \in T_pM$ é definido por*

$$\|u\| = \|u\|_p = \sqrt{\langle u, u\rangle_p}.$$

Observação C.4.1 *Seja $f : M \to N$ uma imersão, isto é, f é diferenciável e $df_p : T_pM \to T_{f(p)}N$ é injetiva para todo $p \in M$. Se N tem uma estrutura Riemanniana, podemos munir M com uma estrutura Riemanniana definindo*

$$\langle u, v\rangle_p = \langle df_p(u), df_p(v)\rangle_{f(p)}, \quad u, v \in T_pM.$$

A métrica de M obtida dessa maneira é dita induzida por f.

Definição C.4.3 *Sejam M e N variedades Riemannianas. Um difeomorfismo $f : M \to N$ é chamado isometria se*

$$\langle u, v\rangle_p = \langle df_p(u), df_p(v)\rangle_{f(p)}, \tag{C.6}$$

quaisquer que sejam $p \in M$ e $u, v \in T_pM$.

Teorema C.2 *Uma variedade diferenciável M (de Hausdorff e com base enumerável) possui uma métrica Riemanniana.*

C.5 Conexões Riemannianas

Definição C.5.1 *Uma conexão afim ∇ em uma variedade diferenciável M é uma aplicação $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$, indicada por $(X, Y) \mapsto \nabla_X Y$, a qual satisfaz às seguintes propriedades:*

i) $\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z$

ii) $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$

iii) $\nabla_X(fY) = f\nabla_X Y + X(f)Y$,

onde $X, Y, Z \in \mathfrak{X}(M)$ *e* $f, g \in C^\infty(M)$.

Observação C.5.1 *A partir de (iii), pode-se mostrar que a conexão afim é uma noção local, isto é, se os campos* $X, Y \in \mathfrak{X}(M)$ *coincidem, em algum aberto* $A \subset M$, *com campos* $\overline{X}, \overline{Y} \in \mathfrak{X}(M)$, *respectivamente, então* $\nabla_X Y$ *e* $\nabla_{\overline{X}}\overline{Y}$ *coincidem em* A.

Observação C.5.2 *Pode-se mostrar também que* $\nabla_X Y(p)$ *depende apenas do valor de* $X(p)$ *e do valor de* Y *ao longo de uma curva tangente a* X.

Observação C.5.3 *Considerando-se os campos* $X_i = \dfrac{\partial}{\partial x_i} \in \mathfrak{X}(M)$, *podemos escrever* $\nabla_{X_i} X_j$ *em* $x(U)$ *como*

$$\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k.$$

Como $\nabla_{X_i} X_j$ *é um campo diferenciável, temos que as funções* Γ_{ij}^k *são diferenciáveis. Tais funções são chamadas símbolos de Christoffel associados à parametrização* (U, x).

Proposição C.5.1 *Seja* M *uma variedade diferenciável com uma conexão afim* ∇. *Então, existe uma única correspondência que associa a um campo vetorial* V *ao longo de uma curva diferenciável* $c : I \to M$ *um outro campo vetorial ao longo de* c, *denotado por* $\dfrac{DV}{dt}$, *tal que:*

a) $\dfrac{D}{dt}(V + W) = \dfrac{DV}{dt} + \dfrac{DW}{dt}$;

b) $\dfrac{D}{dt}(fV) = \dfrac{df}{dt}V + f\dfrac{Dv}{dt}$;

Onde W *é um campo de vetores ao longo de* c, *e* f *é uma função diferenciável em* I.

c) *Se* V *é induzido por um campo de vetores* Y, *isto é,* $V(t) = Y(c(t))$, *então* $\dfrac{DV}{dt} = \nabla_{\frac{dc}{dt}} Y$.

Definição C.5.2 *Um campo vetorial* V *ao longo de uma curva* $c : I \to M$ *é chamado paralelo quando* $\dfrac{Dv}{dt} = 0$ *para todo* $t \in I$.

Definição C.5.3 *Seja* M *uma variedade diferenciável com uma conexão afim* ∇ *e uma métrica Riemanniana* $\langle , \rangle$. *A conexão é dita compatível com a métrica* $\langle , \rangle$ *quando, para toda curva diferenciável* c *e quaisquer pares de campos de vetores paralelos* P, P' *ao longo de* c, *tivermos* $\langle P, P' \rangle = constante$.

Proposição C.5.2 *Suponha que uma variedade Riemanniana M tem uma conexão ∇ compatível com a métrica. Sejam V e W campos de vetores ao longo da curva diferenciável $c : I \longrightarrow M$. Então,*

$$\frac{d}{dt}\langle V, W\rangle = \left\langle \frac{dV}{dt}, W \right\rangle + \left\langle V, \frac{dW}{dt} \right\rangle, \quad t \in I. \tag{C.7}$$

Corolário C.3 *Uma conexão afim ∇ numa variedade Riemanniana M é compatível com a métrica se, e somente se,*

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \tag{C.8}$$

onde $X, Y, Z \in \mathfrak{X}(M)$.

Definição C.5.4 *Uma conexão afim ∇ em uma variedade diferenciável M é dita simétrica quando, quaisquer que sejam $X, Y \in \mathfrak{X}(M)$,temos*

$$\nabla_X Y - \nabla_Y X = [X, Y]. \tag{C.9}$$

Teorema C.4 (Levi-Civita) *Dada uma variedade Riemanniana M, existe uma única conexão afim ∇ em M que é simétrica e compatível com a métrica. Tal conexão é chamada conexão Riemanniana.*

Observação C.5.4 *Dada uma parametrização (U, x) de M, os símbolos de Christoffel da conexão Riemanniana em termos de componentes da métrica são dados por*

$$\Gamma_{ij}^m = \frac{1}{2}\sum_k \left(\frac{\partial g_{jk}}{\partial x_i} + \frac{\partial g_{ik}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_k} \right) g^{km}. \tag{C.10}$$

Observação C.5.5 *Usando-se (C.10) e (C.5), podemos concluir que*

$$\frac{\partial g^{il}}{\partial x_j} = -\sum_k g^{ik}\Gamma_{jk}^l - \sum_m g^{ml}\Gamma_{jm}^i. \tag{C.11}$$

C.6 Gradiente

Definição C.6.1 *Seja $f \in C^\infty(M)$, onde M é uma variedade Riemanniana. O gradiente de f é o campo de vetores em M, denotado por* $\operatorname{grad} f$, *definido pela seguinte condição:*

$$\langle \operatorname{grad} f(p), v\rangle = df_p(v), \quad p \in M, \quad v \in T_pM. \tag{C.12}$$

Proposição C.6.1 *Se $f, g \in C^\infty(M)$, então:*

(i) $\operatorname{grad} f + g = \operatorname{grad} f + \operatorname{grad} g$.

(ii) $\operatorname{grad} f \cdot g = f \cdot \operatorname{grad} g + g \cdot \operatorname{grad} f$.

Demonstração. Omitiremos esta demonstração em virtude de se tratar de um simples uso das propriedades da diferencial de uma aplicação. ■

Proposição C.6.2 *Sejam $x : U \to M$ uma parametrização de M, e $f \in C^\infty(M)$. Então, na vizinhança $x(U)$, temos*

$$\operatorname{grad} f = \sum_{ij} g^{ij} \frac{\partial f}{\partial x_j} \frac{\partial}{\partial x_i} \cdot \tag{C.13}$$

Consequentemente, $\operatorname{grad} f \in \mathfrak{X}(M)$.

Demonstração. Nesta demonstração, iremos usar o fato de que

$$df_p(v) = v(f), \quad p \in M, \quad v \in T_pM.$$

Suponha que, nesta parametrização, tenhamos

$$\operatorname{grad} f = \sum_i a_i \frac{\partial}{\partial x_i} \cdot$$

Portanto, em $x(U)$,

$$\left\langle \operatorname{grad} f, \frac{\partial}{\partial x_j} \right\rangle = \sum_i a_i \left\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right\rangle = \sum_i a_i g_{ij}. \tag{C.14}$$

Assim, denotando $F = \left(\left\langle \operatorname{grad} f, \frac{\partial}{\partial x_j} \right\rangle\right)_{n\times 1}$, $A = (a_i)_{n\times 1}$ e $B = (g_{ij})_{n\times n}$, decorre de (C.14) que $F = BA$. Sendo B invertível, temos $A = B^{-1}F$. Desse modo,

$$a_i = \sum_j g^{ij} \left\langle \nabla f, \frac{\partial}{\partial x_j} \right\rangle = \sum_j g^{ij} df \left(\frac{\partial}{\partial x_j} \right) = \sum_j g^{ij} \frac{\partial f}{\partial x_j},$$

de onde segue o resultado afirmado. ■

Observação C.6.1 *Utilizando a expressão (C.13) do gradiente, obtemos*

$$\langle \operatorname{grad} u, \operatorname{grad} v \rangle = \sum_{i,j=1}^{n} g^{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} \cdot \tag{C.15}$$

Com efeito, temos

$$\operatorname{grad} u = \sum_{i,j=1}^{n} g^{ij} \frac{\partial u}{\partial x_i} \frac{\partial}{\partial x_j} \quad \text{e} \quad \operatorname{grad} v = \sum_{k,\ell=1}^{n} g^{k\ell} \frac{\partial v}{\partial x_k} \frac{\partial}{\partial x_\ell}.$$

Logo,

$$\begin{aligned}
\langle \operatorname{grad} u, \operatorname{grad} v \rangle &= \sum_{i,j,k,\ell=1}^{n} g^{ij} \frac{\partial u}{\partial x_i} g_{j\ell} g^{k\ell} \frac{\partial v}{\partial x_k} \\
&= \sum_{i,j,k=1}^{n} g^{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_k} \left(\sum_{\ell=1}^{n} g_{j\ell} g^{\ell k} \right) \\
&= \sum_{i,j,k=1}^{n} \delta_{jk} g^{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_k} \\
&= \sum_{i,j=1}^{n} g^{ij} \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j}.
\end{aligned}$$

Exemplo C.6.1 *Se $M = I\!R^n$ com a métrica euclidiana, então*

$$g^{ij} = \begin{cases} 1 & , \quad se \quad i = j \\ 0 & , \quad se \quad i \neq j \end{cases}$$

e a base do espaço tangente coincide com a base canônica de $I\!R^n$. Assim, de (C.13), tem-se que

$$\operatorname{grad} f = \nabla f.$$

Isto é, o gradiente de funções em variedades de Riemann é uma generalização do vetor gradiente de uma função real de várias variáveis.

Exemplo C.6.2 *Seja $M = I\!R^n_{++} = \{x \in I\!R^n \ / \ x_i > 0, \ 1 \leq i \leq n\}$. Denota-se $G(x) = (g_{i,j}(x))_{n \times n}$. Neste caso, a métrica é dada pela métrica afim escala*

$$G(x) = D^{-2}(x),$$

onde $D(x) = diag(x_1, ..., x_n)$. Assim, de (C.13), tem-se que

$$\operatorname{grad} f(x) = D^2 \nabla f(x).$$

C.7 Hessiano

Definição C.7.1 *Seja M uma variedad Riemanniana com conexão ∇ dada pelo Teorema de Levi-Civita. Considere $f : M \to I\!R$ de classe C^r, $r \geq 2$. O Hessiano de f em $p \in M$, denotado por H_p^f, é definido como a derivada do campo gradiente; isto é:*

$$\begin{aligned}
H_p^f \ : \ T_pM &\longrightarrow T_pM \\
v &\longmapsto H_p^f \cdot v \ := \ \nabla_v \operatorname{grad} f.
\end{aligned}$$

Proposição C.7.1 *Para cada $p \in M$, operador H_p^f é auto-adjunto; isto é:*

$$\langle H_p^f \cdot v, w\rangle = \langle v, H_p^f \cdot w\rangle,$$

para todo $u \in T_pM$ e todo $v \in T_pM$.

Demonstração. Como ∇ é compatível com a métrica,

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle. \tag{C.16}$$

Como ∇ é simétrica,

$$\nabla_X Y - \nabla Y X = [X, Y]. \tag{C.17}$$

Tomando $X = v$, $Y = w$ e $Z = \operatorname{grad} f$ em (C.16), tem-se:

$$v\langle w, \operatorname{grad} f\rangle = \langle \nabla_v w, \operatorname{grad} f\rangle + \langle w, H_p^f \cdot v\rangle.$$

Tomando $X = w$, $Y = v$ e $Z = \operatorname{grad} f$ em (C.16), tem-se:

$$w\langle v, \operatorname{grad} f\rangle = \langle \nabla_w v, \operatorname{grad} f\rangle + \langle v, H_p^f \cdot w\rangle.$$

Logo,

$$\begin{aligned}
\langle H_p^f \cdot v, w\rangle - \langle v, H_p^f \cdot w\rangle &= (v\langle w, \operatorname{grad} f\rangle - \langle \nabla_v w, \operatorname{grad} f\rangle) \\
&\quad - (w\langle v, \operatorname{grad} f\rangle - \langle \nabla_w v, \operatorname{grad} f\rangle) \\
&= [v(wf) - w(vf)] - [\langle \nabla_v w - \nabla_w v, \operatorname{grad} f\rangle] \\
&= (vw - wv)f - (\nabla_v w - \nabla_w v,)f \\
&= [v, w]f - [v, w]f = 0,
\end{aligned}$$

onde (C.17) foi utilizada . Portanto, H_p^f é auto-adjunto. ■

Observação C.7.1 *Pela Propisição C.7.1, é possível definir a forma quadrática simétrica:*

$$\begin{array}{rcccl}
H_p^f & : & T_pM \times T_pM & \longrightarrow & I\!R \\
& & (u, v) & \longmapsto & H_p^f(u, v) \;:=\; \langle H_p^f \cdot u, v\rangle = \langle \nabla_u \operatorname{grad} f, v\rangle,
\end{array}$$

ou mais generalmente:

$$\begin{array}{rcccl}
H_p^f & : & \mathfrak{X}(M) \times \mathfrak{X}(M) & \longrightarrow & I\!R \\
& & (X, Y) & \longmapsto & H_p^f(X, Y) \;:=\; \langle \nabla_X \operatorname{grad} f, Y\rangle.
\end{array}$$

Exemplo C.7.1 *Seja $M = I\!R^n$ con a métrica usual. Então os coeficientes de Christoffel são $\Gamma_{i,j}^m = 0$, $\forall i, j, m \in \{1, ..., n\}$. Assim, $\nabla_{\partial/\partial x_i} \partial/\partial x_j = 0$, $\forall i, j \in \{1, ..., n\}$. Logo,*

$$H_p^f\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j}\right) = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Isto é, o operador da Definição C.7.1 coincide com o Hessiano usual.

Bibliografia

[1] ADLER I., MONTEIRO R.D.C., *Limiting Behavior of the Affine Scaling Contiuous Trajectories for Programming Problems*, Mathematical Programming 50 p.29-51, 1991.

[2] BAZARAA, M. S., SHERALI, H. D., SHETTY, C. M., *Nonlinear Programming: Theory and Algorithms*, Second Edition, John Wiley & Sons, Inc. 1993.

[3] BRÉZIS H., *Opérateurs Monotones Maximaux et Semigrups de Contraction dans les Espaces de Hilbert*, Mathematics Studies 5, North-Holland, New York, 1973.

[4] BURACHIK, R. S., IUSEM, A. N., *A Generalized Proximal Point Algorithm for the Variational Inequality Problem in a Hilbert Space*, SIAM J. Optim., 8, p.197-216, 1998.

[5] BURACHIK, R. S., *Generalized Proximal Point Methods for the Variational Inequality Problem*, Tese do Grau de Doutor em Matemática, IMPA, Rio de Janeiro, 1995.

[6] CRUZ NETO, J. X., IUSEM, A. N., SVAITER, B. F., *Central Paths, Generalized Proximal Point Methods And Cauchy Trajectories In Riemannian Manifolds*, SIAM Journal on Control and Optimization. Estados Unidos: , vol.37, n.2, p.566-588, 1999.

[7] CRUZ NETO, J. X., LIMA, L. L., OLIVEIRA, P. R., *Geodesic Methods in Riemannian Manifolds*, Balkan Journal in Geometry and Application, Romênia, v. 3, n. 2, p. 89-100, 1998.

[8] CARMO, M. P., *Geometria Riemanniana*, Terceira Edição, IMPA, Rio de Janeiro, 2005.

[9] DA SILVA, S. R. P., *Algoritmo de Ponto Proximal para Otimização em $I\!R^n$*, Disertação de Mestrado em Matemática, UFG, Goiania, 1999.

[10] HOFFMAN, A. J., *On Aproximate Solutions of Systems of nLinear Inequalities*, Jurnal of Research of the National Bureau of Standards, vol. 49, pp. 263-265, 1952.

[11] IUSEM, A. N., *Métodos de Ponto Proximal em Otimização*, 20° Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro, 1995.

[12] IUSEM, A. N., *On some Properties of Generalized Proximal Point Methods for Variational Inequalities*, Jurnal of Optimization Theory and Applications: v.96, n.2, p.337-362, 1998.

[13] IUSEM, A. N., *On some Properties of Paramonotone Operators*, Jurnal of Convex Analysis, 1998.

[14] IZMAILOV, A., SOLODOV, M., *Otimização – volume 1*, IMPA, Rio de Janeiro, 2005.

[15] LIMA, E. L., *Curso de Análise - volume 2*, Sexta Edição, IMPA, Rio de Janeiro, 2000.

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

[17] ROCKAFELLAR, R. T., *On the Maximality of Sums of Nonlinear monotone Operators*, Transactions of the American Mathematical Society 149, p.75-88, 1970.

[18] URRUTY J. B. H., LEMARÉCHAL C., *Fundamentals of Convex Analysis*, Springer, Berlin, 2001.