

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

Método do Ponto Proximal Usando Distâncias Generalizadas
Separáveis - Reescala e Seleção do Comprimento do Passo

por

Alvaro Moreira Neto

Orientador: Dr. Geci José Pereira da Silva

Dissertação de Mestrado em Matemática
Goiânia - Goiás
2008



Termo de Ciência e de Autorização para Disponibilizar as Teses e Dissertações Eletrônicas (TEDE) na Biblioteca Digital da UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás-UFG a disponibilizar gratuitamente através da Biblioteca Digital de Teses e Dissertações - BDTD/UFG, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

1 1. Identificação do material bibliográfico: **Dissertação** **Tese**

1 2. Identificação da Tese ou Dissertação

Autor(a):	Alvaro Moreira Neto		
CPF:		E-mail:	alvaromneto3@hotmail.com
Seu e-mail pode ser disponibilizado na página? <input checked="" type="checkbox"/> Sim <input type="checkbox"/> Não			
Vínculo Empregatício do autor			
Agência de fomento:	Coord. de Aperf. de Pessoal de Nível Superior	Sigla:	CAPES
País:	Brasil	UF:	GO
CNPJ:	00889834/0001-08		
Título:	Método do Ponto Proximal Usando Distâncias Generalizadas Separáveis - Reescala e Seleção do Comprimento do Passo.		
Palavras-chave:	Método do Ponto Proximal com Reescala.		
Título em outra língua:	Rescaling and Stepsize Selection in Proximal Methods using Separable Generalized Distances.		
Palavras-chave em outra língua:	Proximal Methods with Rescaling.		

Área de concentração:	Matemática Aplicada		
Data defesa:	(29/05/2008)		
Programa de Pós-Graduação:	Mestrado em Matemática.		
Orientador(a):	Geci José Pereira da Silva.		
CPF:		E-mail:	geci@mat.ufg.br
Co-orientador(a):	-----		
CPF:	-----	E-mail:	-----

3. Informações de acesso ao documento:

Liberação para disponibilização?¹ **total** **parcial**

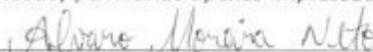
Em caso de disponibilização parcial, assinale as permissões:

Capítulos. Especifique: _____

Outras restrições: _____

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC da tese ou dissertação.

O Sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses e ou dissertações, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.



Assinatura do (a) autor (a).

Data: 11 / 06 / 2008.

¹ Em caso de restrição, esta poderá ser mantida por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Todo resumo e metadados ficarão sempre disponibilizados.

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
COORDENAÇÃO DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Método do Ponto Proximal Usando Distâncias Generalizadas
Separáveis - Reescala e Seleção do Comprimento do Passo

por

Alvaro Moreira Neto

Área de concentração: Matemática Aplicada
Orientador: Prof. Dr. Geci José Pereira da Silva

Dissertação submetida à Banca Examinadora designada pelo Conselho Diretor do Instituto de Matemática e Estatística, como parte dos requisitos necessários à obtenção do grau de **Mestre em Matemática**.

Goiânia, Goiás
2008

Dados Internacionais de Catalogação-na-Publicação (CIP)
(GPT/BC/UFG)

Moreira Neto, Alvaro.

M838m Método do ponto proximal usando distâncias generalizadas separáveis – reescala e seleção do comprimento do passo [manuscrito] / Álvaro Moreira Neto. – 2008.
47 f.

Orientador: Dr. Geci José Pereira da Silva.

Dissertação (Mestrado) – Universidade Federal de Goiás, Instituto de Matemática e Estatística, 2008.

1. Algoritmos [Ponto proximal] 2. MMPR (Método de minimização proximal com reescala) 3. Bregman, Método de I. Silva, Geci José Pereira I. Universidade Federal de Goiás. **Instituto de Matemática e Estatística.** III. Título.

CDU: 510.5

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
COORDENAÇÃO DO PROGRAMA DE PÓS-GRADUAÇÃO EM
MATEMÁTICA-MESTRADO

**"MÉTODO DO PONTO PROXIMAL USANDO DISTÂNCIAS
GENERALIZADAS SEPARÁVEIS-REESCALA E SELEÇÃO DO
COMPRIMENTO DO PASSO"**

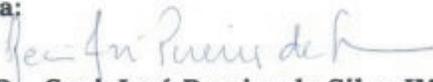
por

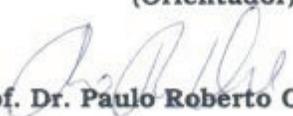
Alvaro Moreira Neto

Dissertação submetida à Banca Examinadora designada pelo Conselho Diretor do Instituto de Matemática e Estatística, como parte dos requisitos necessários à obtenção do grau de Mestre em Matemática.

Goiânia, 29 de maio de 2008.

Banca Examinadora:


Prof. Dr. Geci José Pereira da Silva-IME/UFG
(Orientador)


Prof. Dr. Paulo Roberto Oliveira-UFRJ


Prof. Dr. Orizon Pereira Ferreira-IME/UFG

Agradecimento

Primeiramente a Deus pela força e sabedoria, ao meu orientador Dr. Geci José Pereira da Silva pela paciência e escolha do tema, aos meus pais que me ajudaram nos momentos de dúvidas e à minha esposa Lívia pelo amor e compreensão.

Sumário

Introdução	5
1 Preliminares	8
2 Método do Algoritmo do Ponto Proximal	13
2.1 Algoritmo Clássico do Ponto Proximal	13
2.1.1 Análise da Convergência	14
2.2 Algoritmo do Ponto Proximal para Operadores Monótonos Maximais-(APPOMM)	16
2.2.1 Análise da Convergência	17
2.3 Algoritmo do Ponto Proximal com Distâncias de Bregman-(APPDB)	20
2.3.1 Função e Distância de Bregman	20
2.4 Descrição do Método-APPDB	25
2.4.1 Análise da Convergência	25
2.5 Distância φ -divergente	28
2.5.1 Método Ponto Proximal com φ -divergente-(MPPD)	29
3 Método de Minimização Proximal com Reescala-MMPR	32
3.1 Método Proximal com Reescala para Desigualdade Variacional- (MPRDV)	32
3.2 Análise Fundamental	33

3.2.1	Análise da Convergência	35
3.3	Método de Minimização Proximal com Reescala-MMPR	39
3.4	Descrição do Método-MMPR	39
3.5	Análise da Convergência	39
4	Método Ponto Proximal Interior de Bregman para Desigualdade Variacional	44
4.1	Análise da Convergência	45
4.2	Conclusão	48
	Referências Bibliográficas	49

Resumo

Nessa dissertação, inicialmente, apresentamos o Algoritmo do Ponto Proximal Clássico para resolver o problema de programação não-linear e para obter zeros de operadores monótonos maximais. Em seguida, substituímos a norma Euclideana pela Distância de Bregman e pela φ -divergência para apresentar o Algoritmo do Ponto Proximal Generalizado. Finalmente, estudamos o Método de Minimização Proximal com Reescala para resolver o problema de otimização com restrições e o Método Proximal com Reescala para resolver o problema da Desigualdade Variacional, proposto por Silva e Silva et al [21], onde permitimos diferentes comprimento do passo, os quais são atualizados em cada iteração e considerando a curvatura da função distância. Além disso, apresentamos, também, o Método do Ponto Proximal Interior de Bregman para resolver o problema da Desigualdade Variacional, e observamos que temos hipóteses menos restritivas sobre o operador do que as hipóteses anteriormente utilizadas na literatura.

Abstrat

In this work, initially, we will present the Classic Proximal Point Method to solve the problem of nonlinear programming and to find zeros of maximal monotone operators. Soon afterwards, we replaces the Euclidean norm for the distance Bregman and for the φ -divergent for to bring the Generalized Proximal Point Method. Finally, we study the Proximal Minimization Methods with Rescaling for to solve the optimization problem with stringent and the Proximal Methods with Rescaling for to solve the Inequality Variational Problem, proposed for Silva e Silva et al [21], wherein we afford different stepsizes, those are up-to-date in which iteration and considering the curvature of function distance. Moreover, we will present, also, the Bregman Interior Point Proximal Method for to solve the Inequality Variational Problem, and we observe that we have less stringent assumption about the operator than assumption before utilized in the literature.

Introdução

Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa e limitada inferiormente, o problema de otimização convexo irrestrito é definido como

$$\min_{x \in \mathbb{R}^n} f(x). \quad (1)$$

Agora, para um conjunto fechado e convexo $D \subset \mathbb{R}^n$, o problema de otimização com restrições é dado por

$$\min_{x \in D} f(x). \quad (2)$$

Utilizando o subdiferencial da função f , podemos escrever os problemas (1) e (2), respectivamente, da seguinte forma

$$\text{Encontre } x^* \in \mathbb{R}^n, \text{ tal que } 0 \in \partial f(x^*) \quad (3)$$

e

$$\text{Encontre } x^* \in D, \text{ tal que existe um } w \in \partial f(x^*) \text{ que satisfaz} \quad (4)$$

$$\langle w, y - x^* \rangle \geq 0, \quad \forall y \in D.$$

Com estas considerações, obtemos extensões naturais dos problemas acima substituindo o operador ∂f por qualquer outro operador monótono maximal T . Tem-se então os seguintes problemas

$$\text{Encontre } x^* \in \mathbb{R}^n \text{ tal que}$$

$$0 \in T(x^*) \quad (5)$$

e Encontre $x \in D$ tal que existe um $w \in T(x^*)$ que satisfaz

$$\langle w, y - x^* \rangle \geq 0, \quad \forall y \in D, \quad (6)$$

onde T é um operador monótono maximal e D é um subconjunto convexo, fechado e não-vazio do \mathbb{R}^n .

Desta forma, o problema (5) associado a um operador monótono maximal T , é o problema de encontrar zeros de T . E o problema (6) é chamado problema de desigualdade variacional para T e D .

Nessa dissertação, inicialmente no capítulo 2, apresentaremos o Algoritmo Clássico do Ponto Proximal para resolver os problemas (1) e (5). Tal Algoritmo foi proposto, inicialmente, por Martinet e desenvolvido por Rockafellar [27], o qual gera uma seqüência de pontos $\{x^k\}$ da seguinte maneira

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\}. \quad (7)$$

e

$$0 \in T(x^{k+1}) + \lambda_k(x^{k+1} - x^k), \quad (8)$$

onde $\{\lambda_k\}$ é uma seqüência de números positivos.

Mostraremos que a seqüência gerada por (7) e (8) converge a uma solução dos problemas (1) e (5), respectivamente, ou seja, a um zero de T , sempre que os problemas tiverem soluções e a seqüência de parâmetros $\{\lambda_k\}$ for limitada superiormente. Vimos que para o problema de otimização convexo irrestrito e para o problema de obter zeros de operadores monótonos maximais, o Algoritmo Clássico do Ponto Proximal pode ser, respectivamente, visto como tomar um $x^{k+1} \in \mathbb{R}^n$, tal que

$$0 \in \nabla f(x^{k+1}) + \lambda_k \nabla_x (\|x - x^k\|^2)(x^{k+1}).$$

e obter $x^{k+1} \in \mathbb{R}^n$, tal que

$$0 \in T(x^{k+1}) + \lambda_k \nabla_x (\|x - x^k\|^2)(x^{k+1}).$$

Portanto, consideramos, na seção 2.3, uma generalização do Algoritmo Clássico do Ponto Proximal onde a distância Euclidiana é substituída por uma distância mais geral, chamada distância de Bregman.

No capítulo 3 apresentaremos dois Métodos, o primeiro é o de Minimização Proximal para Desigualdade Variacional–**MPRDV** o qual se propõe a resolver o problema

$$0 \in T(x) + N_B(x) \quad (9)$$

da desigualdade variacional, onde T é o operador monótono maximal, e $N_B(x)$ denota o cone de vetores normal a B em x ; e o de Minimização Proximal com Reescala–**MMPR** que é utilizado para resolver o problema

$$\min_{x \in B} f(x), \quad (10)$$

onde $B \subset \mathbb{R}^n$ é uma caixa n-dimensional fechada $([a_1, b_1] \times \dots \times [a_n, b_n]) \cap \mathbb{R}^n$ tal que $-\infty \leq a_i \leq b_i \leq +\infty$ para todo $i = 1, 2, \dots, n$, e $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ é uma função convexa própria fechada. Cada método utilizará uma função de regularização para obter um melhor comportamento dos problemas (9) e (10) e um vetor $\alpha_k \in \mathbb{R}_{++}^n$, cujos elementos são chamados de comprimento do passo. Permitiremos diferentes comprimento do passo, os quais serão atualizados em cada iteração observando a curvatura da função d_i em cada ponto x_i^k corrente da iterada.

No capítulo 4, finalmente, apresentaremos o Método Ponto Interior de Bregman para Desigualdade Variacional, onde o vetor α^k será tomado observando a curvatura da função h_i da função distância de Bregman. É importante ressaltar que a análise da convergência destes três últimos métodos se difere, dos estudos clássicos, por não ser necessário utilizar hipóteses adicionais sobre T , como paramonotonicidade, pseudo-monotonicidade, semi-continuidade superior e hemi-continuidade [29].

Capítulo 1

Preliminares

Neste capítulo apresentaremos resultados e conceitos básicos que serão utilizados na dissertação.

Definição 1.1. *Seja $D \subset \mathbb{R}^n$ um conjunto qualquer. Diremos que D é convexo se $\alpha y + (1 - \alpha)x \in D$ para todo $x, y \in D$ e $\alpha \in [0, 1]$.*

Esta definição mostra que todo segmento de reta com ponto inicial x e ponto final y está contido em D . Um exemplo de conjunto convexo em \mathbb{R}^n é por exemplo o ortante positivo, isto é, $D = \mathbb{R}_+^n = \{x \in \mathbb{R}^n; x_i \geq 0, i = 1, \dots, n\}$.

Definição 1.2. *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \rightarrow \mathbb{R}$ é dita convexa se,*

$$f(\alpha y + (1 - \alpha)x) \leq \alpha f(y) + (1 - \alpha)f(x)$$

para todo $x, y \in D$ e $\alpha \in [0, 1]$. Diremos que f é estritamente convexa se a desigualdade acima for estrita para todos $x \neq y \in D$ e $\alpha \in (0, 1)$.

Teorema 1.3. *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : D \rightarrow \mathbb{R}$ uma função diferenciável em D . Então as seguintes propriedades são equivalentes:*

- (a) *A função f é convexa em D .*
- (b) *Para todo $x \in D$ e todo $y \in D$,*

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

(c) Para todo $x \in D$ e todo $y \in D$,

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

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

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

$$\langle \nabla^2 f(x)d, d \rangle \geq 0, \forall x \in D, \forall d \in \mathbb{R}^n.$$

Demonstração. Veja [1], Teorema 3.4.7. □

Definição 1.4. Dizemos que a seqüência $\{x^k\} \subset \mathbb{R}^n$ é crítica em relação ao conjunto D se $\{x^k\} \subset D$ e $\lim_{k \rightarrow \infty} \|x^k\| = +\infty$ ou $\lim_{k \rightarrow \infty} x^k = \bar{x}$ onde \bar{x} pertence à fronteira de D .

Definição 1.5. Seja $D \subset \mathbb{R}^n$ e $f : D \rightarrow \mathbb{R}$ uma função contínua. Dizemos que f é coerciva se toda seqüência $\{x^k\}$ crítica com relação a D tem-se que $\lim_{k \rightarrow \infty} \sup f(x^k) = +\infty$.

Definição 1.6. O conjunto $C \subset \mathbb{R}^n$ é dito afim se

$$(1 - \lambda)x + \lambda y \in C, \forall x, y \in C \text{ e } \lambda \in \mathbb{R}.$$

Obviamente, a intersecção de uma coleção de conjuntos afins é ainda um conjunto afim. Portanto, dado algum $C \subset \mathbb{R}^n$ existe um único menor conjunto afim que contém C , o qual é uma intersecção de uma coleção de conjuntos afins M tal que $M \supset C$. Este conjunto é chamado de envoltória afim de C e será denotado por $aff C$, e pode ser provado que $aff C$ é formado por todos os vetores da forma $\theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k$, tais que $x_i \in C$, e $\sum_{i=1}^k \theta_i = 1$.

Definição 1.7. Chamamos de interior relativo de C ao conjunto dado por

$$ri C := \{x \in C : B(x, \varepsilon) \cap aff C \subseteq C, \text{ para algum } \varepsilon > 0\}.$$

Um operador monótono maximal é uma generalização de uma transformação linear semi-definida positiva para o caso não-linear, o que inclui o gradiente de uma função convexa diferenciável. Seja $A \in \mathbb{R}^{n \times n}$, então A é semi-definida positiva se, e somente se, $0 \leq x^T A x = \langle x, A x \rangle$ para todo $x \in \mathbb{R}^n$. Seja agora $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ não necessariamente linear. Como, nem sempre ocorre que, $0 \leq \langle x, T(x) \rangle$, mas como $0 \leq \langle x, A x \rangle$ para todo x , se, e somente se, $0 \leq \langle (x - y), A(x - y) \rangle = \langle (x - y), Ax - Ay \rangle$ para todo x, y temos

Definição 1.8. $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ é monótono se, e somente se,

$$0 \leq \langle x - y, T(x) - T(y) \rangle \text{ para todo } x, y \in \mathbb{R}^n.$$

Um exemplo de operador monótono é quando tomamos $T = \nabla f$ com f convexa e diferenciável, veja Teorema (1.3)(c).

Definição 1.9. ξ é um subgradiente de f em x se, e somente se,

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

Desta maneira o subdiferencial de f em x será o conjunto

$$\partial f(x) = \{\xi : \xi \text{ é um subgradiente de } f \text{ em } x\}.$$

Desta forma as seguintes propriedades são válidas:

1. Se f diferenciável e convexa então $\partial f(x) = \{\nabla f(x)\}$.

2. Se f é convexa então $\partial f(x) \neq \emptyset$ para todo x no interior relativo do domínio efetivo de f . Percebamos que ∂f associa para cada x não apenas um vetor, mas um subconjunto de \mathbb{R}^n , então precisamos estender a noção de operador monótono para operadores ponto-conjunto $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$.

Definição 1.10. T é monótono se, e somente se, $0 \leq \langle x - y, u - v \rangle$ para todos $x, y \in \mathbb{R}^n, u \in T(x), v \in T(y)$.

Por exemplo se $T = \partial f$ com f convexa. Tome $\xi \in T(x), \eta \in T(y)$. Pela definição (1.9) $\langle \xi, y - x \rangle \leq f(y) - f(x)$ e $\langle -\eta, y - x \rangle \leq f(x) - f(y)$, implicando $0 \leq \langle \xi - \eta, x - y \rangle$.

Definição 1.11. T é monótono maximal se, e somente se,

(i) T é monótono;

(ii) Para todo T' monótono tal que $T(x) \subset T'(x)$ para todo x , segue que $T = T'$.

Pode ser verificado que ∂f é monótono maximal para f convexa.

Um vetor x^* é dito normal a um conjunto convexo $D \subset \mathbb{R}^n$ num ponto $a \in D$, se x^* não faz um ângulo agudo com qualquer segmento de reta em D com a como ponto extremo, isto é, se

$\langle x - a, x^* \rangle \leq 0$ para todo $x \in D$. Por exemplo se D é um semi-espço do tipo $D = \{x : \langle x, b \rangle \leq \beta\}$ e a satisfaz $\langle a, b \rangle = \beta$, então b é normal a D em a . Em geral o conjunto de todos os vetores x^* normais a D em a é chamado cone normal a D em a , ou seja,

$$N_D(a) := \{x^* \mid \langle x - a, x^* \rangle \leq 0, \forall x \in D\}.$$

Definição 1.12. *Seja $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$, X um espaço topológico. Dizemos que a função f é semi-contínua inferior em $a \in X$ se para cada $\lambda \in \mathbb{R}$, $\lambda < f(a)$, existe uma vizinhança U de a , tal que $f(U) > \lambda$. f é semi-contínua inferior se ela o for em cada ponto de X .*

Uma função convexa f é dita ser própria se

$$f(x) < +\infty, \text{ pelo menos para um } x;$$

$$f(x) > -\infty, \text{ para todo } x.$$

Observemos que o problema (2) é equivalente ao problema

$$\min_{x \in D} \{f(x) + \delta_D(x)\} \tag{1.1}$$

onde D é um conjunto convexo e δ_D é a função indicador de um conjunto convexo D , ou seja:

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

Teorema 1.13. *Sejam f_1, \dots, f_m funções convexas próprias em \mathbb{R}^n , e seja $f = f_1 + \dots + f_m$. Então*

$$\partial f(x) \supset \partial f_1(x) + \dots + \partial f_m(x), \forall x.$$

Se os conjuntos convexos $ri(\text{dom } f_i)$, $i = 1, \dots, m$, possuem algum ponto em comum, então vale

$$\partial f(x) = \partial f_1(x) + \dots + \partial f_m(x), \forall x.$$

Demonstração. Ver [23], Teorema 23.8. □

Se $ri(\text{dom } f) \cap ri(\text{dom } D) \neq \emptyset$, então o problema (1.1) é equivalente ao problema de encontrar um \bar{x} tal que $0 \in \partial f(\bar{x}) + N_D(\bar{x})$ de acordo com o Teorema (1.13).

Definição 1.14. O conjunto de nível da função $f : D \rightarrow \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por

$$L_{f,D}(c) = \{x \in D : f(x) \leq c\}.$$

Teorema 1.15. Seja $f : D \rightarrow \mathbb{R}$ uma função própria e convexa. Se o conjunto de nível $L_{f,D}(c) = \{x \in D : f(x) \leq c\}$ é não-vazio e limitado para algum c , então o mesmo será limitado para todo $t \in \mathbb{R}$.

Demonstração. Ver [23], Corolário 8.7.1. □

Definição 1.16. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dizemos que f é essencialmente suave em $C = \text{int}(\text{dom } f)$ se:

- (a) C é não vazio;
- (b) f é diferenciável em C ;
- (c)

$$\lim_{x_i \rightarrow \infty} |\nabla f(x_i)| = +\infty$$

para toda seqüência $\{x_1, x_2, \dots\}$ convergindo a um \bar{x} na fronteira de C .

Enunciaremos, agora, um Lema técnico, porém muito importante, que será utilizado posteriormente nos nossos resultados.

Lema 1.17. Suponha que $\{a_k\}, \{\gamma_k\} \subset \mathbb{R}$ sejam seqüências tais que $\{a_k\}$ é limitada inferiormente, $\sum_{k=1}^{\infty} \gamma_k$ existe e é finito e que ocorre $a_{k+1} \leq a_k + \gamma_k$, para todo k . Então, $\{a_k\}$ é convergente.

Demonstração. Ver [22], seção 2.2 □

Capítulo 2

Método do Algoritmo do Ponto Proximal

2.1 Algoritmo Clássico do Ponto Proximal

Apresentaremos nesta seção o método do ponto proximal clássico que foi proposto por Martinet (1970) e desenvolvido por Rockafellar, para resolver o problema de otimização convexa em \mathbb{R}^n como definido em (1) e para obter zeros de Operadores Monótonos Maximais num espaço de Hilbert.

Algoritmo Clássico do Ponto Proximal-ACPP

Este algoritmo gera uma seqüência $\{x^k\} \subset \mathbb{R}^n$, da seguinte forma:

1.**Inicialização.** Escolha um $x^0 \in \mathbb{R}^n$ e $\lambda_0 \in (0, \bar{\lambda})$.

2.**Iteração.** Para $k = 1, 2, 3, \dots$

(a) Escolha o parâmetro de regularização $\lambda_k \in (0, \bar{\lambda})$, e encontre x^{k+1} tal que

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\}. \quad (2.1)$$

3.Faça $k := k + 1$ e volte para 1.

Observamos que a seqüência $\{x^k\}$ gerada pelo algoritmo **ACPP** está bem definida. Pois, definindo $f_k(x) := f(x) + \lambda_k \|x - x^k\|^2$ temos que $\lim_{\|x\| \rightarrow \infty} f_k(x) = \infty$, ou seja, f_k é coerciva

(definição (1.5)), pois f é limitada inferiormente. Assim, a minimização do problema (1) se reduz a um conjunto compacto, e como f_k é contínua, pelo Teorema de Weirstrass [1], Teorema 1.2.1, ela atinge seu mínimo. Temos também por hipótese que f é convexa e como $\lambda_k \|x - x^k\|^2$ é estritamente convexa segue que f_k é estritamente convexa, e assim possui um único minimizador, x^{k+1} .

2.1.1 Análise da Convergência

Agora mostraremos que a seqüência obtida pelo algoritmo **ACPP** converge a uma solução do problema (1). Primeiramente, necessitaremos da seguinte definição.

Definição 2.1. *Uma seqüência $\{y^k\} \subset \mathbb{R}^n$ é dita Fejér convergente num conjunto $U \subset \mathbb{R}^n$, com respeito à norma Euclideana, se*

$$\|y^{k+1} - u\| \leq \|y^k - u\|, \forall k \geq 0, \forall u \in U.$$

Proposição 2.2. *Se $\{y^k\}$ é uma seqüência Fejér convergente, então $\{y^k\}$ é limitada. Além disso, se um ponto de acumulação \bar{y} de $\{y^k\}$ pertence a U , então $\lim_{k \rightarrow \infty} y^k = \bar{y}$.*

Demonstração. Seja $u \in U$, um ponto fixo. Como, por hipótese, a seqüência $\{y^k\}$ é Fejér convergente, temos que

$$\|y^{k+1} - u\| \leq \|y^k - u\| \leq \dots \leq \|y^0 - u\|.$$

Portanto a seqüência está contida numa bola de centro u e raio $\|y^0 - u\|$, logo é limitada.

Resta provarmos a segunda parte. Consideremos uma subsequência $\{y^{j_k}\}$ de $\{y^k\}$, tal que $\lim_{k \rightarrow \infty} y^{j_k} = \bar{y}$. Como $\bar{y} \in U$ e $\{y^k\}$ é Fejér convergente temos pela definição (2.1) que a seqüência $\{\|y^k - \bar{y}\|\}$ é não-negativa e não-crescente, portanto convergente. Como a subsequência $\{\|y^{j_k} - \bar{y}\|\}$ converge a zero, segue que $\{y^k\}$ também converge a zero, ou seja, $\lim_{k \rightarrow \infty} \|y^k - \bar{y}\| = 0$, assim $\lim_{k \rightarrow \infty} y^k = \bar{y}$. \square

Lema 2.3. *Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa e continuamente diferenciável e o conjunto dos minimizadores de f , $U \subset \mathbb{R}^n$, é não-vazio. Então a seqüência $\{x^k\}$, gerada por*

ACPP satisfaz a seguinte desigualdade:

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2, \forall k \geq 0 e \forall \bar{x} \in U.$$

Demonstração. Temos que

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

Como x^{k+1} satisfaz (2.1), temos

$$0 = \nabla f_k(x^{k+1}) = \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k).$$

Assim,

$$\begin{aligned} \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2 - \|x^{k+1} - \bar{x}\|^2 &= 2\langle x^k - x^{k+1}, x^{k+1} - \bar{x} \rangle \\ &= \frac{1}{\lambda_k} \langle \nabla f(x^{k+1}), x^{k+1} - \bar{x} \rangle \\ &\geq \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})) \geq 0, \end{aligned}$$

onde a primeira desigualdade segue do fato de f ser convexa, Teorema (1.3)(b), e a segunda, pelo fato de \bar{x} ser um minimizador do problema (1). \square

Lema 2.4. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e continuamente diferenciável, o conjunto dos minimizadores de f , $U \subset \mathbb{R}^n$, não-vazio e a seqüência $\{x^{k+1}\}$, gerada por **ACPP**. Então a seqüência $\{x^{k+1} - x^k\}$ converge para zero.*

Demonstração. Pelo Lema (2.3), temos

$$0 \leq \|x^{k+1} - x^k\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2.$$

Como $\{\|x^{k+1} - \bar{x}\|\}$ é não-crescente e não-negativa, então é convergente. Logo o lado direito da desigualdade converge para zero, e segue o resultado. \square

Lema 2.5. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e continuamente diferenciável e o conjunto dos minimizadores de f , $U \subset \mathbb{R}^n$, não-vazio. Então a seqüência $\{x^{k+1}\}$, gerada por **ACPP**, possui pontos de acumulação e todos pertencem a U .*

Demonstração. Pelo Lema (2.3), temos que $\{x^k\}$ é uma seqüência Fejér convergente em U , e assim pela Proposição (2.2), ela é limitada.

Seja \bar{x} um ponto de acumulação de $\{x^k\}$ e $\{x^{j_k}\}$ uma subsequência de $\{x^k\}$, tal que

$$\lim_{k \rightarrow \infty} (x^{j_k}) = \bar{x}.$$

Por (2.1),temos que

$$0 = \nabla f_k(x^{k+1}) = \nabla f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k),$$

assim

$$\nabla f(x^{j_k+1}) = 2\lambda_{j_k}(x^{j_k} - x^{j_k+1}). \quad (2.2)$$

Mas pelo Lema (2.4), segue que

$$\lim_{k \rightarrow \infty} (x^{j_k+1}) = \lim_{k \rightarrow \infty} (x^{j_k}) = \bar{x}.$$

Usando que $\lambda_k \leq \bar{\lambda}$, o fato de f ser continuamente diferenciável e aplicando o limite em (2.2), obtemos que $\nabla f(\bar{x}) = 0$. Portanto, pela convexidade de f , $\bar{x} \in U$. \square

Observamos que o Lema (2.3) garante que a seqüência $\{x^k\}$ gerada por **ACPP** é Fejér convergente e o Lema (2.4) que todos os seus pontos de acumulação pertencem a U . Portanto, pela Proposição (2.2), a seqüência $\{x^k\}$ converge a $x^* \in U$. Com todas estas considerações podemos exibir o Teorema da Convergência, onde sua demonstração segue dos Lemas anteriores.

Teorema 2.6. *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e continuamente diferenciável e o conjunto dos minimizadores de f , $U \subset \mathbb{R}^n$, não-vazio. Então a seqüência $\{x^k\}$, gerada por **ACPP**, converge para um ponto $x^* \in U$.*

2.2 Algoritmo do Ponto Proximal para Operadores Monótonos Maximais-(APPOMM)

Notemos que o Algoritmo Clássico do Ponto Proximal-**ACPP** gera para o problema (1) uma seqüência $\{x^k\} \subset \mathbb{R}^n$, pela iteração:

$$x^0 \in \mathbb{R}^n,$$

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x^k - x\|^2\},$$

onde $\lambda_k > 0$.

Assim, considerando a função f convexa, temos que

$$0 \in \partial f(x^{k+1}) + 2\lambda_k(x^{k+1} - x^k)$$

ou seja,

$$2\lambda_k(x^k - x^{k+1}) \in \partial f(x^{k+1})$$

assim,

$$x^k \in \left(I + \frac{1}{2\lambda_k} \partial f \right) (x^{k+1}). \quad (2.3)$$

Agora, motivados pela observação (2.3) e lembrando que o subdiferencial de uma função convexa, ∂f , é um operador monótono maximal, podemos estender o Algoritmo Clássico do Ponto Proximal para encontrar zeros de Operadores Monótonos Maximais $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$.

Algoritmo do Ponto Proximal para Operadores Monótonos Maximais-(APPOMM)

Este algoritmo gera uma seqüência $\{x^k\} \subset \mathbb{R}^n$, da seguinte forma:

1. **Iniciação:** Escolha um vetor $x^0 \in \mathbb{R}^n$ e $\lambda_0 \in (0, \bar{\lambda})$.
2. **Iteração:** Para $k = 0, 1, 2, \dots$ escolha o parâmetro $\lambda_k \in (0, \bar{\lambda})$, e encontre x^{k+1} tal que

$$x^{k+1} \in \left(I + \frac{1}{\lambda_k} T \right)^{-1} (x^k).$$

3. Faça $k = k + 1$ e volte a 1.

2.2.1 Análise da Convergência

Primeiramente, considerando T um operador ponto-conjunto, dizemos que T é sobrejetivo se, para todo $y \in \mathbb{R}^n$, existe $x \in \mathbb{R}^n$, tal que $y \in T(x)$. E o operador T será injetivo se para $x \neq y$, tivermos que $T(x) \cap T(y) = \emptyset$.

Teorema 2.7. (Teorema de Minty) *Se $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ é monótono maximal e $\lambda > 0$, então o operador $I + \lambda T$ é injetor e sobrejetor.*

Demonstração. Ver [6]. □

O teorema (2.7) garante que o algoritmo **APPOMM** está bem definido, pois sabemos que um operador ponto-conjunto T é sobrejetivo se existe $x \in \mathbb{R}^n$, tal que $y \in T(x)$, $\forall y \in \mathbb{R}^n$. Além disso, o Algoritmo **APPOMM** gera uma seqüência $\{x^k\}$ pela iteração

$$x^{k+1} \in \left(I + \frac{1}{\lambda_k} T \right)^{-1} (x^k).$$

Finalmente, o Teorema (2.7) garante que $I + \lambda T$ é sobrejetivo, garantindo a existência de $\{x^{k+1}\}$ tal que

$$x^k \in \left(I + \frac{1}{\lambda_k} T \right) (x^{k+1}).$$

Portanto,

$$x^{k+1} \in \left(I + \frac{1}{\lambda_k} T \right)^{-1} (x^k).$$

e a injetividade de $\left(I + \frac{1}{\lambda_k} T \right)$ garante a unicidade de x^{k+1} .

O teorema a seguir garante que os Operadores Monótonos Maximais são fechados.

Lema 2.8. *Se $\lim_{k \rightarrow \infty} y^k = \bar{y}$, $\lim_{k \rightarrow \infty} z^k = \bar{z}$, T é um operador monótono maximal e $y^k \in T(z^k)$, então*

$$\bar{y} \in T(\bar{z}).$$

Demonstração. Defina

$$T'(z) := \begin{cases} T(z), & \text{se } z \neq \bar{z} \\ T(\bar{z}) \cup \{\bar{y}\}, & \text{se } z = \bar{z}. \end{cases} \quad (2.4)$$

Então, precisamos verificar que:

$$\langle y - y', z - z' \rangle \geq 0, \forall z, z', \forall y \in T'(z), \forall y' \in T'(z'). \quad (2.5)$$

Pela definição de T' e como T é um operador monótono, basta verificarmos (2.5) para $y' = \bar{y}$, $z' = \bar{z}$. Por hipótese temos que, $y^k \in T(z^k)$, logo

$$\langle y - y^k, z - z^k \rangle \geq 0, \forall z, \forall y \in T(z). \quad (2.6)$$

Tomando o limite com $k \rightarrow \infty$ em (2.6), obtemos que

$$\langle y - \bar{y}, z - \bar{z} \rangle \geq 0, \forall z, \forall y \in T(z).$$

Assim T' é um operador monótono. Como, $T(x) \subset T'(x)$, para todo $x \in \mathbb{R}^n$ e por hipótese, T é maximal, concluímos que $T = T'$. Em particular, $T(\bar{z}) = T'(\bar{z}) = T(\bar{z}) \cup \{\bar{y}\}$, isto é, $\bar{y} \in T(\bar{z})$. \square

Apresentaremos, a seguir, o Teorema da Convergência do Algoritmo do Ponto Proximal para Operadores Monótonos Maximais. Observamos que de forma análoga ao Lema (2.3), apenas substituindo na demonstração o Teorema (1.3)(b) pela monotonicidade de T , segue que

A1–

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - x^k\|^2, \forall k \geq 0 \text{ e } \forall \bar{x} \text{ tal que } 0 \in T(\bar{x}).$$

Portanto, de forma análoga ao Lema (2.4), temos que

A2–A seqüência $\{x^{k+1} - x^k\}$ converge para zero.

Agora, usaremos o Lema (2.8) para mostrar o análogo ao Lema (2.5)

Lema 2.9. *Sejam $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ monótono maximal e o conjunto $U = \{x \in \mathbb{R}^n \mid 0 \in T(x)\}$, não-vazio. Então, a seqüência $\{x^k\}$, gerada por **APPOMM**, tem pontos de acumulação e todos pertencem a U .*

Demonstração. Temos, pelo item A1, que a seqüência $\{\|x^{k+1} - \bar{x}\|\}$ é monótona, decrescente e limitada inferiormente, então é convergente. Sejam \bar{x} um ponto de acumulação da seqüência $\{x^k\}$ e $\{x^{k_j}\} \rightarrow \bar{x}$. Por hipótese, temos que

$$x^{k+1} \in \left(I + \frac{1}{\lambda_k} T \right)^{-1} (x^k).$$

então

$$\lambda_{k_j} (x^{k_j} - x^{k_j+1}) \in T(x^{k_j+1}).$$

Como $\lambda_{k_j} < \bar{\lambda}$, ou seja, λ_{k_j} é limitado e $\{x^{k_j} - x^{k_j+1}\}$ converge para zero, pelo item A2, temos que

$$\lambda_{k_j} (x^{k_j} - x^{k_j+1})$$

também converge para zero. Assim podemos aplicar o Lema (2.8) e concluir que $0 \in T(x^*)$. Portanto, pelos itens A1 e A2 observamos que a seqüência $\{x^k\}$ é Fejér convergente, então, pela Proposição (2.2), esta converge para \bar{x} . \square

Teorema 2.10. *Sejam $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ monótono maximal e o conjunto $U = \{x \in \mathbb{R}^n \mid 0 \in T(x)\}$, não-vazio. Então, a seqüência $\{x^k\}$, gerada por **APPOMM**, converge a um elemento de U .*

Demonstração. Por A1, segue que $\{x^k\}$ é Fejér convergente em U e, portanto, pelo Lema (2.9) juntamente com a Proposição (2.2), esta converge a um elemento de U . \square

2.3 Algoritmo do Ponto Proximal com Distâncias de Bregman- (APPDB)

Nesta seção, apresentaremos um Algoritmo do Ponto Proximal Generalizado, isto é, substituindo a norma Euclidiana pela distância de Bregman, para resolver o problema de otimização convexo irrestrito. Primeiramente apresentaremos a função e distância de Bregman.

2.3.1 Função e Distância de Bregman

A noção da função distância de Bregman foi introduzida em [9] e tem sido estudada no contexto do método proximal em trabalhos subseqüentes.

Definição 2.11. *São dados S um subconjunto aberto e convexo em \mathbb{R}^n e $h : \bar{S} \rightarrow \mathbb{R}$, onde \bar{S} é o fecho de S , uma função convexa e diferenciável em S . Definimos a distância de Bregman, $D_h : \bar{S} \times S \rightarrow \mathbb{R}$, que é induzida por h , por*

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle. \quad (2.7)$$

Definição 2.12. *Dado S um subconjunto aberto e convexo em \mathbb{R}^n , dizemos que $h : \bar{S} \rightarrow \mathbb{R}$ é uma função de Bregman com zona S , se satisfaz as seguintes condições*

- (B1) h é estritamente convexa e contínua em \bar{S} ;
- (B2) h é continuamente diferenciável em S ;
- (B3) Para quaisquer $x \in \bar{S}$ e $\delta \in \mathbb{R}$, o conjunto de nível parcial à direita,

$$\Gamma_1(x, \delta) = \{y \in S; D_h(x, y) \leq \delta\},$$

é limitado;

(B4) Se a seqüência $\{y^k\} \subset S$, converge para y^* , então

$$D_h(y^*, y^k) \text{ converge para zero};$$

(B5) Sejam as seqüências $\{x^k\}$ e $\{y^k\}$, contidas em \bar{S} e S , respectivamente. Se $\{x^k\}$ é limitada, $\lim_{k \rightarrow \infty} y^k = y^*$ e $\lim_{k \rightarrow \infty} D_h(x^k, y^k) = 0$, então $\lim_{k \rightarrow \infty} x^k = y^*$.

Esta última condição é chamada consistência de convergência e pode ser provada usando as condições anteriores.

Lema 2.13. (A Desigualdade Triangular Restrita) *Seja h uma função que satisfaz as condições (B1) e (B2) da Definição (2.12). Se $x \in \bar{S}$, $y \in S$ e w é uma combinação própria convexa de x e y , isto é,*

$$w = (1 - \alpha)x + \alpha y \tag{2.8}$$

com $\alpha \in (0, 1)$, então

$$D_h(x, w) + D_h(w, y) \leq D_h(x, y).$$

Demonstração. Como, por hipótese, $w = (1 - \alpha)x + \alpha y$, com $\alpha \in (0, 1)$, então $w \in S$, pois S é convexo. Sabemos que ∇h é monótono, assim

$$\langle \nabla h(w), w - y \rangle \geq \langle \nabla h(y), w - y \rangle. \tag{2.9}$$

Substituindo, $w - y = (1 - \alpha)(x - y)$ na desigualdade (2.9), obtemos

$$\langle \nabla h(w), (1 - \alpha)(x - y) \rangle \geq \langle \nabla h(y), (1 - \alpha)(x - y) \rangle \langle \nabla h(w), x - y \rangle \geq \langle \nabla h(y), x - y \rangle. \tag{2.10}$$

Portanto, de (2.8) e (2.10), segue

$$\begin{aligned} D_h(x, w) + D_h(w, y) &= [h(x) - h(w) - \langle \nabla h(w), x - w \rangle] + [h(w) - h(y) - \langle \nabla h(y), w - y \rangle] \\ &= h(x) - h(y) - \alpha \langle \nabla h(w), x - y \rangle - (1 - \alpha) \langle \nabla h(y), x - y \rangle \\ &\leq h(x) - h(y) - \langle \nabla h(y), x - y \rangle = D_h(x, y). \end{aligned}$$

□

Lema 2.14. *Seja h uma função convexa que satisfaz os itens (B1) e (B2) da Definição (2.12). Sejam $\{x^k\}$ e $\{y^k\}$ seqüências em S , convergindo para \bar{x} e \bar{y} , respectivamente, e $\bar{x} \neq \bar{y}$, então*

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

Demonstração. Defina

$$z^k := \frac{1}{2}(x^k + y^k).$$

Claramente $\{z^k\}$ é uma seqüência em S convergindo a $\bar{z} = (1/2)(\bar{x} + \bar{y}) \in \bar{S}$. Pela convexidade de h e pelo Teorema (1.3)(b), segue que para todo k

$$h(z^k) \geq h(y^k) + \langle \nabla h(y^k), z^k - y^k \rangle = h(y^k) + \frac{1}{2} \langle \nabla h(y^k), x^k - y^k \rangle.$$

Portanto

$$\frac{h(x^k) + h(y^k)}{2} - h(z^k) \leq \frac{h(x^k) + h(y^k)}{2} - h(y^k) - \frac{1}{2} \langle \nabla h(y^k), x^k - y^k \rangle = \frac{D_h(x^k, y^k)}{2} \quad (2.11)$$

Fazendo $k \rightarrow \infty$, na desigualdade (2.11), segue

$$\frac{h(\bar{x}) + h(\bar{y})}{2} - h\left(\frac{\bar{x} + \bar{y}}{2}\right) \leq \frac{1}{2} \liminf_{k \rightarrow \infty} D_h(x^k, y^k).$$

Utilizando a convexidade estrita de h e a hipótese que $\bar{x} \neq \bar{y}$, segue o resultado desejado \square

O teorema, a seguir, substitui a propriedade (B5) da definição (2.12). Podemos observar, entretanto, que a referida propriedade é uma conseqüência imediata do Teorema (2.15)

Teorema 2.15. *Seja h uma função convexa que satisfaz os itens (B1) e (B2) da Definição (2.12). Sejam $\{x^k\}$ e $\{y^k\}$ seqüências em S e \bar{S} , respectivamente, e $\lim_{k \rightarrow \infty} D_h(x^k, y^k) = 0$. Se uma das seqüências $\{x^k\}$ ou $\{y^k\}$ convergir, a outra convergirá, também, para o mesmo limite.*

Demonstração. Suponhamos, para efeito de contradição, que uma das seqüências converge e que a outra não, ou não converge para o mesmo limite. Então existe algum $\epsilon > 0$ e uma subseqüência de índices $\{k_j\}$ que satisfaz

$$\|x^{k_j} - y^{k_j}\| > \epsilon.$$

Suponhamos, primeiramente, que $\lim_{k \rightarrow \infty} y^k = \bar{y}$. Defina

$$\tilde{x}_j := y^{k_j} + \frac{\epsilon}{\|x^{k_j} - y^{k_j}\|} (x^{k_j} - y^{k_j}).$$

Assim, \tilde{x}_j é uma combinação própria convexa de x^{k_j} e y^{k_j} . Utilizando o Lema (2.13), concluímos que

$$D_h(\tilde{x}_j, y^{k_j}) \leq D_h(x^{k_j}, y^{k_j}).$$

Então

$$\lim_{j \rightarrow \infty} D_h(\tilde{x}_j, y^{k_j}) = 0.$$

Como $\|\tilde{x}_j - y^{k_j}\| = \epsilon$ e y^{k_j} converge, segue que $\{\tilde{x}^{j_i}\}$ é limitada e existe uma subsequência $\{\tilde{x}^{j_i}\}$ convergindo para \tilde{x} . Portanto, temos as seguintes relações

$$\begin{aligned} \lim_{i \rightarrow \infty} D_h(\tilde{x}^{j_i}, y^{k_{j_i}}) &= 0 \\ \lim_{i \rightarrow \infty} y^{k_{j_i}} &= \bar{y} \\ \lim_{i \rightarrow \infty} \tilde{x}^{j_i} &= \tilde{x} \\ \|\tilde{x} - \bar{y}\| &= \epsilon > 0. \end{aligned}$$

que estão em contradição com o Lema (2.14). Admitindo que a seqüência $\{x^k\}$ converge, obtemos, analogamente, uma contradição ao Lema (2.14). \square

Definição 2.16. *Seja $h : \bar{S} \rightarrow \mathbb{R}$ uma função de Bregman.*

(i) *Dizemos que h é coerciva na fronteira se satisfaz a seguinte condição*

(B6) *Se a seqüência $\{y^k\} \subset S$ é tal que*

$$\lim_{k \rightarrow \infty} y^k = y \in \partial S,$$

então

$$\lim_{k \rightarrow \infty} \langle \nabla h(y^k), x - y^k \rangle = -\infty, \quad \forall x \in S.$$

(ii) *Dizemos que h é zona coerciva se satisfaz a seguinte condição*

(B7) *Para todo $y \in \mathbb{R}^n$ existe um x pertencente a S , tal que*

$$\nabla h(x) = y.$$

Apresentaremos agora dois Lemas úteis que serão usados na demonstração da convergência do algoritmo **APPDB** a ser definido.

Lema 2.17. *Seja h uma função de Bregman. Se h satisfaz (B6), então $D(\partial h) = S$.*

Demonstração. Ver em [2]. □

Lema 2.18. (Lema dos Três-Pontos) *Seja f uma função de Bregman com zona S , que satisfaz as condições (B1) e (B2) da definição (2.12). Para quaisquer $x, z \in S$ e $y \in \bar{S}$, garante-se que*

$$D_f(x, y) = D_f(z, x) + D_f(y, z) + \langle \nabla f(x) - \nabla f(z), z - y \rangle.$$

Demonstração. Segue da definição de distância de Bregman. □

Exibiremos, agora, alguns exemplos de funções de Bregman.

Exemplo 2.19. *Sejam $S = \mathbb{R}^n$ e*

$$h(x) = \frac{1}{2} \langle x, Mx \rangle,$$

onde $M \in \mathbb{R}^{n \times n}$ é simétrica e positiva definida. Neste caso,

$$D_h(x, y) = \frac{1}{2} \langle x - y, M(x - y) \rangle := \frac{1}{2} \|x - y\|_{Max}^2.$$

Exemplo 2.20. *Sejam $S = \mathbb{R}_{++}^n = \{x \in \mathbb{R}^n : x_i > 0, \forall i = 1, \dots, n\}$ e*

$$h(x) = \sum_{i=1}^n x_i \log x_i,$$

estendida com continuidade a $\partial \mathbb{R}_{++}^n$, com a condição que $0 \log 0 = 0$. Neste caso,

$$D_h(x, y) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{y_i} + y_i - x_i \right).$$

Exemplo 2.21. *Sejam $S = \mathbb{R}_{++}^n$ e*

$$h(x) = \sum_{i=1}^n (x_i^\alpha - x_i^\beta),$$

com $\alpha \geq 1$, $\beta \in (0, 1)$. Para o caso particular, no qual $\alpha = 1$, $\beta = 1/2$, temos que

$$D_h(x, y) = \sum_{i=1}^n \frac{1}{2\sqrt{y_i}} \left(\sqrt{x_i - \sqrt{y_i}} \right)^2.$$

2.4 Descrição do Método-APPDB

O algoritmo que apresentaremos agora destina-se à resolução do problema de otimização convexo restrito,

$$\min_{x \in \bar{S}} f(x) \quad (2.12)$$

onde $S \subset \mathbb{R}^n$ é aberto e convexo e f é uma função convexa, limitada inferiormente e contínua em \bar{S} .

1. **Inicialização:** Escolha $x_0 \in S$ e $\lambda_0 \in (0, \tilde{\lambda}]$.

2. **Iteração:** Para $k = 1, 2, 3, \dots$, escolha o parâmetro de regularização $\lambda_k \in (0, \tilde{\lambda}]$ e encontre x^{k+1} , tal que

$$x^{k+1} := \arg \min_{x \in \bar{S}} \{f(x) + \lambda_k D_h(x, x^k)\}, \quad (2.13)$$

onde h é uma função de Bregman com zona S .

3. Faça $k = k + 1$ e volte para 1.

2.4.1 Análise da Convergência

A demonstração da convergência será feita através dos quatro Lemas que se seguirão.

Lema 2.22. *A seqüência $\{x^k\}$ gerada por APPDB, está bem definida e contida em S .*

Demonstração. Seja β o limite inferior para f em \bar{S} e

$$f_k(x) := f(x) + \lambda_k D_h(x, x^k).$$

Assim

$$f_k(x) \geq \beta + \lambda_k D_h(x, x^k).$$

e segue de (B3) que o conjunto de nível é limitado, então a minimização em (2.13) se reduz a um conjunto compacto e o mínimo é atingido. Como f_k é estritamente convexa, então o mínimo é único e x^{k+1} é unicamente determinado.

Resta provarmos que $x^{k+1} \in S$. De (2.13), observamos que x^{k+1} é o único $x \in \bar{S}$, tal que

$$\lambda_k \nabla h(x^k) \in \partial(f + \lambda_k h)(x^{k+1}). \quad (2.14)$$

Observando que

$$x^{k+1} \in D(\partial f) \cap D(\partial h) = D(\partial f) \cap S$$

e portanto, pelo Lema (2.17), $x^{k+1} \in S$. \square

Lema 2.23. *Se \bar{x} é uma solução do problema (2.12), então*

$$D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k), \forall k.$$

Demonstração. Utilizando o Lema (2.18) com $x = \bar{x}$, $y = x^k$ e $z = x^{k+1}$, obtemos que

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \langle \nabla h(x^k) - \nabla h(x^{k+1}), x^{k+1} - \bar{x} \rangle. \quad (2.15)$$

Por (2.13), temos que

$$0 \in \partial(f + \lambda_k D_h(\cdot, x^k))(x^{k+1}). \quad (2.16)$$

Temos, também, por (2.16), que

$$\lambda_k(\nabla h(x^k) - \nabla h(x^{k+1})) \in \partial f(x^{k+1}). \quad (2.17)$$

Tome $y^k = \lambda_k(\nabla h(x^k) - \nabla h(x^{k+1}))$. Pela definição do subgradiente e por (2.15), temos que

$$D_h(\bar{x}, x^k) - D_h(\bar{x}, x^{k+1}) - D_h(x^{k+1}, x^k) = \frac{1}{\lambda_k} \langle y^k, x^{k+1} - \bar{x} \rangle \geq \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})). \quad (2.18)$$

Como $f(\bar{x}) \leq f(x^{k+1})$, segue a desigualdade desejada. \square

Lema 2.24. *Se a seqüência $\{x^k\}$, gerada por **APPDB** é limitada e possui uma subseqüência $\{x^{k_j}\}$ que converge para \hat{x} , então $\{x^{k_j+1}\}$ também converge para \hat{x} .*

Demonstração. Como a seqüência $\{D_h(\bar{x}, x^k)\}$ é monótona e

$$D_h(\bar{x}, x^{k+1}) \leq D_h(\bar{x}, x^k) - D_h(x^{k+1}, x^k), \forall k,$$

então

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

Como $D_h(\bar{x}, x^k)$ é não-crescente, temos que $D_h(\bar{x}, x^k) \leq D_h(\bar{x}, x^0)$ e portanto, $\{x^k\}$ é limitada, por (B3). Logo, se $\lim_{k \rightarrow \infty} x^{k_j} = \hat{x}$, então

$$\lim_{k \rightarrow \infty} x^{k_j+1} = \hat{x}.$$

□

Lema 2.25. *Todos os pontos de acumulação da seqüência $\{x^k\}$, gerada por **APPDB**, são soluções do problema (2.12).*

Demonstração. Sejam \bar{x} uma solução do problema (2.12) e \hat{x} um ponto de acumulação da seqüência $\{x^k\}$, tal que

$$\lim_{k \rightarrow \infty} x^{k_j} = \hat{x}.$$

Por (2.18), temos que

$$0 \leq \frac{1}{\lambda} \left(f(x^{k_j+1}) - f(\bar{x}) \right) \leq \frac{1}{\lambda_k} \left(f(x^{k_j+1}) - f(\bar{x}) \right) \leq D_h(\bar{x}, x^{k_j}) - D_h(\bar{x}, x^{k_j+1}) - D_h(x^{k_j+1}, x^{k_j}).$$

Como a seqüência $D_h(\bar{x}, x^k)$ é convergente e por (B5), temos que a última desigualdade converge para zero e, então $f(\hat{x}) = f(\bar{x})$. Como \bar{S} é fechado e $\{x^k\} \subset \bar{S}$, temos que $\hat{x} \in \bar{S}$, logo \hat{x} é solução de (2.12). □

Apresentaremos, a seguir, o teorema para a análise da convergência do Algoritmo **APPDB**, observando que sua demonstração é obtida dos resultados dos quatro Lemas anteriores.

Teorema 2.26. *Se o problema (2.12) tem solução e h é coerciva na fronteira, com respeito a S , então a seqüência $\{x^k\}$, gerada por **APPDB**, converge para uma solução x^* do problema (2.12).*

Demonstração. Como a seqüência $\{D_h(\bar{x}, x^{k_j})\}$ converge para zero, então $\{D_h(\bar{x}, x^k)\}$ também converge para zero. Assim, pela condição (B5), obtemos que x^k converge para uma solução, x^* , do problema (2.12). □

2.5 Distância φ -divergente

Apresentaremos nesta seção outra classe importante de "distâncias", a φ -divergente a qual representaremos por $d_\varphi(\cdot, \cdot)$, definida no \mathbb{R}_{++}^n .

Definição 2.27. Considere $\varphi : \mathbb{R}_{++} \rightarrow \mathbb{R}$, convexa e duas vezes continuamente diferenciável satisfazendo

$$\varphi(1) = \varphi'(1) = 0, \quad \varphi''(1) > 0, \quad \lim_{t \rightarrow 0} \varphi'(t) = -\infty.$$

Então a distância φ -divergente $d_\varphi : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ induzida por φ é definida por

$$d_\varphi(x, y) = \sum_{j=1}^n y_j \varphi\left(\frac{x_j}{y_j}\right).$$

As propriedades que se seguem são um resultado da Definição (2.27)

Proposição 2.28. Seja d_φ como na definição (2.27), então são válidas as propriedades:

(a) $d_\varphi(x, y) \geq 0$ para todo $x, y \in \mathbb{R}_{++}^n$, e $d_\varphi(x, y) = 0$, se, e só se, $x = y$.

(b) Os conjuntos de níveis de $d_\varphi(\cdot, y)$ e $d_\varphi(x, \cdot)$ são limitados para todo $y \in \mathbb{R}_{++}^n$ e todo $x \in \mathbb{R}_{++}^n$, respectivamente.

(c) $d_\varphi(x, y)$ é convexa em x, y , e estritamente convexa em x .

(d) $\lim_{k \rightarrow \infty} d_\varphi(\bar{y}, y^k) = 0$, se $\lim_{k \rightarrow \infty} y^k = \bar{y}$.

Seguem, agora, alguns exemplos de funções φ -divergente.

Exemplo 2.29. $\varphi_1(t) = t \log t - t + 1$. Então

$$d_{\varphi_1}(t) = \sum_{j=1}^n \left(x_j \log \frac{x_j}{y_j} + y_j - x_j \right) \quad (2.19)$$

d_{φ_1} é chamada de divergente de Kullback–Leiber. Note que podemos estendê-la para $\mathbb{R}_+^n \times \mathbb{R}_{++}^n$. Observando os termos lineares aditivos em h e as constantes multiplicativas em φ, h , o par (φ_1, h_1) com $h_1(x) = \sum_{j=1}^n x_j \log x_j$ é o único par (φ, h) tal que $d_\varphi = D_h$.

Exemplo 2.30. $\varphi_2(t) = t - \log t - 1$. Então

$$d_{\varphi_2}(x, y) = d_{\varphi_1}(y, x). \quad (2.20)$$

Exemplo 2.31. $\varphi_3(t) = (\sqrt{t} - 1)^2$. Então

$$d_{\varphi_3}(x, y) = \sum_{j=1}^n (\sqrt{x_j} - \sqrt{y_j})^2 \quad (2.21)$$

2.5.1 Método Ponto Proximal com φ -divergente-(MPPD)

O problema, agora, de nosso interesse é

$$\min_{x \in \mathbb{R}_+^n} f(x) \quad (2.22)$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa.

O método **MPPD** para resolver o problema (2.22) gera uma seqüência $\{x^k\} \subset \mathbb{R}^n$ do seguinte modo:

1.Inicialização: Escolha um vetor $x^0 \in \mathbb{R}_{++}^n$ e um parâmetro de regularização $\tilde{\lambda} > 0$.

2.Iteração: Encontre x^{k+1} de maneira que

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \{f(x) + \lambda_k d_{\varphi}(x, x^k)\} \quad (2.23)$$

onde $\lambda_k \in (0, \tilde{\lambda}]$.

Das condições de otimalidade para (2.23) temos que $u_k \in \partial f(x^{k+1})$ com

$$u_j^k = -\lambda_k \varphi' \left(\frac{x_j^{k+1}}{x_j^k} \right).$$

Se f é diferenciável, então $x^{k+1} \rightarrow \bar{x}$ onde \bar{x} é a solução do sistema

$$\nabla f(x)_j + \lambda_k \varphi' \left(\frac{x_j}{x_j^k} \right) = 0. \quad (2.24)$$

Note que (2.24) é não-linear possuindo n equações e n incógnitas x_1, \dots, x_n .

Percebamos que quando o problema (2.22) possui solução, a iteração (2.23) reduz-se a um subconjunto compacto, pela Definição (2.27), garantindo a existência de x^{k+1} , a unicidade de x^{k+1} segue da convexidade de f e da Proposição (2.28)(c). O problema é que a seqüência $\{x^{k+1}\}$ gerada por (2.23) não é Fejér-convergente para o conjunto solução do problema (2.22) com respeito à d_{φ} , isto é, talvez tenhamos um \bar{x} que resolva (2.22) tal que $d_{\varphi}(\bar{x}, x^k) > d_{\varphi}(\bar{x}, x^{k+1})$. Por este motivo, para a análise da convergência do **MPPD**, necessitaremos, inicialmente, da definição

Definição 2.32. Uma seqüência $\{y^k\} \subset \mathbb{R}_{++}^n$ é dita *quasi-Fejér convergente* ao conjunto $U \subset \mathbb{R}_{++}^n$ com respeito à distância φ -divergente d_φ se para cada $u \in U$ existe uma seqüência $\{\varepsilon_k\} \subset \mathbb{R}_{++}^n$ tal que

$$\sum_{k=0}^{\infty} \varepsilon_k < \infty,$$

e para todo $k \geq 0$ temos, também, que

$$d_\varphi(u, y^{k+1}) \leq d_\varphi(u, y^k) + \varepsilon_k.$$

Lema 2.33. Se $\{y^k\} \subset \mathbb{R}_{++}^n$ é *quasi-Fejér convergente* a $U \subset \mathbb{R}_{++}^n$ com respeito a uma distância φ -divergente d_φ então $\{y^k\}$ é limitada. Se um ponto de acumulação \bar{y} de $\{y^k\}$ pertence a U então

$$\lim_{k \rightarrow \infty} y^k = \bar{y}.$$

Demonstração. Ver em [15]. □

Voltemos nossa atenção à seqüência $\{x^k\}$ gerada por (2.23). Ela não é *quasi-Fejér convergente* ao conjunto solução do problema (2.22) com respeito a d_φ de fato, se uma solução \bar{x} de (2.23) pertence a $\partial\mathbb{R}_+^n$ então $d_\varphi(\bar{x}, x^k)$ não está bem definida. Portanto, sob algumas condições em φ , $\{x^k\}$ é *quasi-Fejér convergente* ao conjunto solução de (2.22) com respeito ao divergente de Kullback–Leiber, isto é, com respeito a d_ψ onde $\psi(t) = \varphi_1(t) = t \log t - t + 1$, que admite pontos em $\partial\mathbb{R}_+^n$ como um primeiro argumento e para o qual o Lema (2.33) está assegurado com $U \subset \mathbb{R}_+^n$, $\bar{y} \in \mathbb{R}_+^n$. Mais precisamente, considere a seqüência $\{x^k\}$ gerada por (2.23) com algum d_φ , seja U o conjunto solução do problema (2.22), tome $\bar{x} \in U$ e defina δ_k como

$$\delta_k := d_\varphi(\bar{x}, x^{k+1}) - d_\varphi(\bar{x}, x^k). \quad (2.25)$$

Lema 2.34. (i) Se $\varphi'(t) \leq \varphi''(1) \log t$ e existe um $\hat{\lambda}$ tal que $0 < \hat{\lambda} \leq \lambda_k$ para todo k então

$$\sum_{k=0}^{\infty} \delta_k < \infty.$$

Assim $\{x^k\}$ é *quasi-Fejér convergente* em U .

(ii) Se

$$\frac{\varphi''(1)}{t} \leq \varphi'(t) \leq \varphi''(1) \log t$$

então $\delta_k \leq 0$. Assim $\{x^k\}$ é *Fejér convergente* em U .

Demonstração. Ver [3]. □

Observamos que $\varphi_1, \varphi_2, \varphi_3$ dos exemplos (2.29), (2.30) e (2.31) satisfazem as desigualdades da hipótese do Lema (2.34). Em vista da primeira afirmação do Lema (2.33) e do Lema (2.34), a seqüência $\{x^k\}$ como em (2.23) é limitada quando $\varphi'(t) \leq \varphi''(1) \log t$ e $0 < \hat{\lambda} \leq \lambda_k$ ou quando $\frac{\varphi''(1)}{t} \leq \varphi'(t) \leq \varphi''(1) \log t$ e pelo Lema (2.33), é suficiente para provar que os pontos de acumulação da seqüência $\{x^k\}$ pertencem a U para garantir a convergência de $\{x^k\}$ a uma solução do problema (2.22). Vale mencionar, também, que se a seqüência $\{x^k\}$ converge ao ponto $\bar{x} > 0$ e f é diferenciável então é imediato ver que \bar{x} resolve o problema (2.22), porque de (2.24)

$$\nabla f(x^{k+1}) = -\lambda_k \varphi' \left(\frac{x_j^{k+1}}{x_j^k} \right)$$

e $\lim_{k \rightarrow \infty} x_j^{k+1} = \lim_{k \rightarrow \infty} x_j^k = \bar{x}_j > 0$ então

$$\lim_{k \rightarrow \infty} \varphi' \left(\frac{x_j^{k+1}}{x_j^k} \right) = \varphi'(1) = 0$$

e portanto $\nabla f(\bar{x})_j = 0$ usando $\lambda_k \leq \tilde{\lambda}$ para todo j , isto é, \bar{x} é o minimizador irrestrito de f . Em geral \bar{x} é apenas um ponto de acumulação o qual pode ter algum componente zero e f não é diferenciável.

Segue agora o principal Teorema da convergência do **MPPD**

Teorema 2.35. *Se um dos dois casos*

$$(i) \varphi'(t) \leq \varphi'(1) \log t \text{ e } \hat{\lambda} \leq \lambda_k \text{ para algum } \hat{\lambda} > 0,$$

ou

$$(ii) \varphi'(1) < \varphi'(t) \leq \varphi''(1) \log t,$$

ocorre, então a seqüência $\{x^k\}$ gerada por (2.23) converge a uma solução do problema (2.22).

Demonstração. Ver [15]. □

Capítulo 3

Método de Minimização Proximal com Reescala-MMPR

3.1 Método Proximal com Reescala para Desigualdade Variacional- (MPRDV)

Seja $B \subset \mathbb{R}^n$ uma caixa n-dimensional fechada $([a_1, b_1] \times \dots \times [a_n, b_n]) \cap \mathbb{R}^n$ onde $-\infty \leq a_i \leq b_i \leq +\infty$ para todo $i = 1, 2, \dots, n$. Consideraremos dois problemas, primeiramente

$$\min_{x \in B} f(x), \quad (3.1)$$

onde $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ é uma função convexa própria fechada, e a desigualdade variacional

$$0 \in T(x) + N_B(x) \quad (3.2)$$

onde T é o operador monótono maximal, e $N_B(x)$ denota o cone de vetores normal a B em x . É do nosso conhecimento que, segundo o Teorema (1.13), se $ri(dom f) \cap ri(B) \neq \emptyset$, temos que (3.1) é o caso especial de (3.2) para $T = \partial f$, uma vez que $N_B(x) = \partial \delta_B(x)$. Na literatura existem vários trabalhos sobre o método do ponto proximal baseado em distâncias generalizadas, como por exemplo em [11, 12, 14, 7, 16]. Cada método utiliza uma função de regularização para obter um melhor comportamento dos problemas (3.1) e (3.2). Nesta dissertação vamos considerar

termos de regularização separáveis da seguinte forma

$$D(x, y) = \sum_{i=1}^n d_i(x_i, y_i)$$

onde d_1, \dots, d_n são funções escalares satisfazendo certas condições. Também com as funções $d_i, i = 1, \dots, n$, teremos que para $x \in B$ próximo à sua fronteira, $\|\nabla_1 D(x, y)\| \rightarrow \infty$, onde ∇_1 denota o gradiente com respeito ao primeiro vetor argumento, ou seja $D(x, y)$ é coerciva na fronteira, isto faz com que a regularização não só aja como um termo estabilizador proximal mas também como um tipo de função barreira mantendo as iterações dentro de B .

Usando estes termos de regularização, o método proximal para (3.1) possui a forma

$$x^{k+1} := \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \right\}, \quad (3.3)$$

onde α^k é um vetor n -dimensional estritamente positivo, ou seja, pertencente ao \mathbb{R}_{++}^n , no qual seus elementos são chamados de comprimento do passo. Perceba que no capítulo 3 todos os termos de regularização eram números reais pertencentes a um intervalo em \mathbb{R} . No caso da desigualdade variacional (3.2), (3.3) generaliza buscando x^{k+1} satisfazendo

$$0 \in T(x^{k+1}) + \text{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k). \quad (3.4)$$

3.2 Análise Fundamental

Esta seção elabora a análise fundamental necessária para nossos resultados. Concentraremos nossa atenção no problema variacional (3.2). Para simplificar a notação, denotaremos para cada $i = 1, \dots, n$:

$$d'_i(x_i, y_i) := \frac{\partial d_i}{\partial x_i}(x_i, y_i),$$

$$d''_i(x_i, y_i) := \frac{\partial^2 d_i}{\partial x_i^2}(x_i, y_i).$$

Consideremos as seguintes hipóteses sobre as funções d_i :

Hipótese 3.1. *Para cada $i = 1, \dots, n$, a função $d_i : \mathbb{R} \times (a_i, b_i) \rightarrow (-\infty, \infty]$ possui as seguintes propriedades:*

- (a) Para todo $y_i \in (a_i, b_i)$, $d_i(\cdot, y_i)$ é fechado e estritamente convexa, com mínimo ocorrendo em y_i . Além disto, $\text{int dom } d_i(\cdot, y_i) = (a_i, b_i)$.
- (b) d_i é continuamente diferenciável sobre $(a_i, b_i) \times (a_i, b_i)$, e para todo $y_i \in (a_i, b_i)$, $d_i''(y_i, y_i)$ existe e é estritamente positivo.
- (c) Para todo $y_i \in (a_i, b_i)$, $d_i(\cdot, y_i)$ é essencialmente suave.
- (d) Existe $\rho, \epsilon > 0$ tal que se qualquer $-\infty < a_i < y_i \leq x_i < a_i + \epsilon$ ou $b_i - \epsilon < x_i \leq y_i < b_i < +\infty$ então $\rho |d_i'(x_i, y_i)| \leq d_i''(y_i, y_i) |x_i - y_i|$.

Exemplo 3.2. Seja $B = \mathbb{R}_+^n$ e

$$d_i(x_i, y_i) = y_i \varphi\left(\frac{x_i}{y_i}\right)$$

onde $\varphi : \mathbb{R} \rightarrow (-\infty, +\infty]$, tal que $\varphi(1) = \varphi'(1) = 0$. Onde podemos ter

- $\varphi(t) = -\log t + t - 1$, com $a_i = 0$ e $b_i = +\infty$.
- $\varphi(t) = 2(\sqrt{t} - 1)^2$, com $a_i = 0$ e $b_i = +\infty$.

A hipótese de convexidade estrita de d_i é padrão na generalização do método proximal. A hipótese de d_i ser duas vezes diferenciável é também bastante comum, no entanto muitos resultados exigem apenas que d_i seja diferenciável. A parte (c) da hipótese (3.1) faz com que a distância D aja como uma função barreira, forçando as iterações definida por (3.4), e conseqüentemente a versão (3.5) adiante, a permanecer no interior de B . Finalmente a quarta parte da hipótese (3.1) é nova para teoria do método proximal usando distâncias generalizadas. Para apresentarmos o método necessitaremos da

Hipótese 3.3. $\text{dom } T \cap \text{int } B \neq \emptyset$.

Apresentaremos, agora, o algoritmo:

Método Proximal com Reescala para Desigualdade Variacional- (MPRDV)

1. **Inicialização:** Faça $k = 0$. Escolha um escalar $c > 0$, e uma iteração inicial $x^0 \in \text{int } B$.

2. **Iteração:**

- (a) Escolha $\alpha^k \in \mathbb{R}_{++}^n$ tal que $\alpha_i^k \geq c \max\{1, d_i''(x_i^k, x_i^k)\}$ para todo $i = 1, \dots, n$.
- (b) Encontre x^{k+1} e e^{k+1} tal que

$$e^{k+1} \in T(x^{k+1}) + \text{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k). \quad (3.5)$$

(c) Faça $k = k + 1$, e repita a iteração.

A fim de obter a convergência do **MPRDV**, precisamos da seguinte hipótese no comprimento do passo $\{\alpha_i^k\}$ e na seqüência erro $\{e^k\}$; veja a hipótese (3.4) abaixo.

Definimos

$$\gamma^k := e^k - \text{diag}(\alpha^{k-1})^{-1} \nabla_1 D(x^k, x^{k-1}) \quad (3.6)$$

onde fica claro de (3.5) que $\gamma^k \in T(x^k)$ para todo $k \geq 1$.

Hipótese 3.4. *Considere $\{\beta_k\}$ uma seqüência de números reais convergindo a zero. A seqüência erro, $\{e^k\}$, as funções regularização d_1, \dots, d_n , e o comprimento do passo $\{\alpha_i^k\}$, $i = 1, \dots, n$, devem ser escolhidos na ordem para garantir que:*

$$(a) |e_i^k| \leq \frac{1}{\alpha_i^{k-1}} |d_i'(x_i^k, x_i^{k-1})| + \beta_k.$$

(b) *Se \bar{x} é um ponto de acumulação da seqüência $\{x^k\}$, isto é, existe um conjunto infinito $\mathcal{K} \subseteq \mathbb{N}$ tal que $x^k \rightarrow_{\mathcal{K}} \bar{x}$, então para cada $i = 1, \dots, n$ um dos dois casos ocorre $\gamma_i^k \rightarrow_{\mathcal{K}} 0$, ou existe um conjunto infinito $\mathcal{K}' \subseteq \mathcal{K}$ tal que $x_i^{k-1} \rightarrow_{\mathcal{K}'} \bar{x}_i$.*

3.2.1 Análise da Convergência

Assumiremos através desta seção que as Hipóteses (3.1) e (3.3) sejam válidas, e que as seqüências $\{\alpha_k\}$, $\{x^k\}$, e $\{e^k\}$ geradas pelo algoritmo **MPRDV** e satisfazendo a Hipótese (3.4) existem.

Lema 3.5. *Seja $\bar{x} \in \mathbb{R}^n$ tal que \bar{x} seja um ponto de acumulação da seqüência $\{x^k\}$, isto é $x^k \rightarrow_{\mathcal{K}} \bar{x}$, onde \mathcal{K} é algum subconjunto infinito dos naturais.*

Então para $i = 1, \dots, n$,

$$\begin{aligned} \lim_{k \rightarrow_{\mathcal{K}} \infty} \gamma_i^k &= 0, \text{ se } \bar{x}_i \in (a_i, b_i) \\ \lim_{k \rightarrow_{\mathcal{K}} \infty} \inf \gamma_i^k &\geq 0, \text{ se } \bar{x}_i = a_i \\ \lim_{k \rightarrow_{\mathcal{K}} \infty} \sup \gamma_i^k &\leq 0, \text{ se } \bar{x}_i = b_i \end{aligned} \quad (3.7)$$

Demonstração. Para cada i , consideremos os 3 possíveis casos: Primeiro, suponha que i é tal que $\bar{x}_i \in (a_i, b_i)$. Para efeito de contradição assumamos que $\gamma_i^k \not\rightarrow 0$, com $k \in \mathcal{K}$. Então usando a

hipótese (3.4)(b), existe um conjunto infinito $\mathcal{K}' \subseteq \mathcal{K}$ e um $\zeta > 0$ tal que para todo $k \in \mathcal{K}'$, $|\gamma_i^k| \geq \zeta$ e $x_i^{k-1} \rightarrow \bar{x}_i$. Portanto

$$|\gamma_i^k| = \left| e_i^k - \frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^{k-1}) \right| \leq \frac{1}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| + |e_i^k|$$

E pela hipótese (3.4)(a) podemos escrever esta última desigualdade do seguinte modo,

$$|\gamma_i^k| \leq \frac{2}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| + \beta_k$$

Escolhendo $\alpha_i^k = c$ e fazendo $k \rightarrow_{\mathcal{K}'} \infty$, teremos:

$$|\gamma_i^k| \leq \left(\frac{2}{c} \right) |d'_i(\bar{x}_i, \bar{x}_i)| + 0.$$

Agora como o mínimo de $d_i(\cdot, \bar{x}_i)$ é \bar{x}_i , então $|\gamma_i^k| = 0$, o que contraria o fato de $|\gamma_i^k| \geq \zeta$.

No próximo caso considere $\bar{x}_i = a_i$, e suponha que $\lim_{k \rightarrow_{\mathcal{K}'} \infty} \inf \gamma_i^k < 0$. Então usando a hipótese (3.4)(b), existirá $\zeta > 0$ e um conjunto infinito $\mathcal{K}' \subseteq \mathcal{K}$ tal que para todo $k \in \mathcal{K}'$, $\gamma_i^k \leq -\zeta$ e $x_i^{k-1} \rightarrow_{\mathcal{K}'} \bar{x}_i$, então

$$\zeta \leq |\gamma_i^k| = \left| e_i^k - \frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^{k-1}) \right|$$

Com as mesmas suposições finais do primeiro caso, temos

$$\zeta \leq \frac{2}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| + \beta_k \leq \frac{2}{cd_i''(x_i^{k-1}, x_i^{k-1})} |d'_i(x_i^k, x_i^{k-1})| + \beta_k$$

Seja ϵ como na hipótese (3.1)(d). Se existe um conjunto infinito $\mathcal{K}'' \subseteq \mathcal{K}'$ tal que $x_i^{k-1} \leq x_i^k \leq a_i + \epsilon$ para todo $k \in \mathcal{K}''$, podemos concluir daí que:

$$\begin{aligned} \zeta &\leq \frac{2}{cd_i''(x_i^{k-1}, x_i^{k-1})} |d'_i(x_i^k, x_i^{k-1})| + \beta_k \leq \frac{2d_i''(x_i^{k-1}, x_i^{k-1})}{\rho cd_i''(x_i^{k-1}, x_i^{k-1})} |(x_i^k - x_i^{k-1})| + \beta_k = \\ &= \frac{2}{\rho c} |(x_i^k - x_i^{k-1})| + \beta_k \rightarrow 0 \end{aligned}$$

desde que $x_i^{k-1} \rightarrow_{\mathcal{K}'} \bar{x}_i$, e $\beta_k \rightarrow 0$; mas esta conclusão contradiz $\zeta > 0$. Portanto $x_i^k \leq x_i^{k-1}$ para $k \in \mathcal{K}'$ suficientemente grande.

Como $d_i(\cdot, x_i^k)$ atinge o mínimo em x_i^{k-1} , tendo $x_i^k \leq x_i^{k-1}$ implica que $d'_i(x_i^k, x_i^k) \leq 0$. Deste modo

$$\gamma_i^k = e_i^k - \frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^{k-1}) \geq \frac{1}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| - |e_i^k| \geq -\beta_k > -\zeta$$

para $k \in \mathcal{K}'$ suficientemente grande, o que é uma contradição com o fato de $\gamma_i^k \leq -\zeta < 0, k \in \mathcal{K}'$.

Finalmente, o caso de $\bar{x}_i = b_i$ é análogo ao caso $\bar{x}_i = a_i$. \square

Lema 3.6. *Seja \bar{x} um ponto de acumulação da seqüência $\{x^k\}$, isto é $x^k \rightarrow \bar{x}$ para algum conjunto infinito $\mathcal{K} \subseteq \mathbb{N}$. Então, $\{\gamma^k\}_{\mathcal{K}}$ é limitada.*

Demonstração. Pela hipótese (3.3), existe algum $\tilde{x} \in \text{dom } T \cap \text{int } B$. Seja $\tilde{\gamma} \in T(\tilde{x})$. A monotonicidade de T implica que, para todo $k \geq 0$,

$$0 \leq \langle x^k - \tilde{x}, \gamma^k - \tilde{\gamma} \rangle = \sum_{i=1}^n (x_i^k - \tilde{x}_i)(\gamma_i^k - \tilde{\gamma}_i) \quad (3.8)$$

Mostraremos que se fosse a seqüência $\{\gamma^k\}_{\mathcal{K}}$ ilimitada, então a desigualdade (3.8) não seria verdadeira para um k suficientemente grande.

Se $\{\gamma^k\}_{\mathcal{K}}$ é ilimitada, então existirá algum conjunto $\mathcal{K}' \subseteq \mathcal{K}$ tal que $\{\gamma^k\}_{\mathcal{K}'}$ será divergente, ou seja, pelo menos um $\{\gamma_i^k\}_{\mathcal{K}'}$ será ilimitado. O Lema (3.5) implica que para cada coordenada ilimitada i , um dos dois casos ocorrerá:

$$\{\gamma_i^k\} \rightarrow_{\mathcal{K}'} +\infty \text{ e } \bar{x}_i = a_i$$

ou

$$\{\gamma_i^k\} \rightarrow_{\mathcal{K}'} -\infty \text{ e } \bar{x}_i = b_i$$

Portanto, para cada coordenada ilimitada de $\{\gamma_i^k\}$, temos:

$$(x_i^k - \tilde{x}_i)(\gamma_i^k - \tilde{\gamma}_i) \rightarrow_{\mathcal{K}'} (a_i - \tilde{x}_i)(+\infty) = -\infty$$

ou

$$(x_i^k - \tilde{x}_i)(\gamma_i^k - \tilde{\gamma}_i) \rightarrow_{\mathcal{K}'} (b_i - \tilde{x}_i)(-\infty) = -\infty$$

Por outro lado, para coordenadas tal que $\{\gamma_i^k\}_{\mathcal{K}'}$ é limitada, $(x_i^k - \tilde{x}_i)(\gamma_i^k - \tilde{\gamma}_i)$ é também limitada. Desta forma, para k suficientemente grande $k \in \mathcal{K}' \subseteq \mathcal{K}$, $\langle x^k - \tilde{x}, \gamma^k - \tilde{\gamma} \rangle$ deve ser negativo, o que contraria (3.8). \square

Apresentaremos, agora, o principal teorema da convergência do **MPRDV**.

Teorema 3.7. *Se $\{x^k\}$, é uma seqüência gerada pelo algoritmo **MPRDV** com as hipóteses (3.1), (3.3) e (3.4) válidas, então todo ponto de acumulação de $\{x^k\}$ é solução para o problema (3.2).*

Demonstração. Seja \bar{x} um ponto de acumulação da seqüência $\{x^k\}$, isto é $x^k \rightarrow_{\mathcal{K}} \bar{x}$, para algum conjunto infinito $\mathcal{K} \subseteq \mathbb{N}$. Pelo Lema (3.6), sabemos que seqüência correspondente $\{\gamma^k\} \in T(x^k)$ é limitada. Então deverá existir algum conjunto $\mathcal{K}' \subseteq \mathcal{K}$ com $\gamma^k \rightarrow_{\mathcal{K}'} \bar{\gamma} \in \mathbb{R}^n$. Pelo Lema (2.8), segue-se que $\bar{\gamma} \in T(\bar{x})$. Das conseqüências do Lema (3.5) temos que

$$\bar{\gamma} = 0 \text{ se } \bar{x}_i \in (a_i, b_i),$$

$$\bar{\gamma} \geq 0 \text{ se } \bar{x}_i = a_i,$$

$$\bar{\gamma} \leq 0 \text{ se } \bar{x}_i = b_i,$$

e estas condições são equivalentes a $0 \in T(\bar{x}) + N_B(\bar{x})$. \square

Vejamos, agora, que a condição de $d_i(\cdot, y_i)$, ser duas vezes diferenciável, pode ser desconsiderada pelo custo de uma complexidade adicional na descrição do método. Especificamente, considere a substituição da hipótese (3.1)(d) com a condição de que existe $\delta, \epsilon > 0$ e funções $L_i : (a_i, b_i) \rightarrow (\delta, +\infty)$ tal que, se um dos dois fatos ocorre $-\infty < a_i < y_i \leq x_i < a_i + \epsilon$ ou $b_i - \epsilon < x_i \leq y_i < b_i < +\infty$, então

$$|d'_i(x_i, y_i)| \leq L_i(y_i)|x_i - y_i|$$

Se o comprimento do passo é agora selecionado, então para algum escalar $c > 0$, temos para todo $i = 1, \dots, n$ e $k \leq 0$ que $\alpha_i^k \geq cL_i(x_i^k)$, então as conclusões do Teorema (3.7) continuam asseguradas. A presente aproximação é equivalente a tomar $L_i(y_i) = (1/\rho)d''(y_i, y_i)$, uma natural escolha visto que $d''(y_i, y_i)$ mede a taxa de mudança de $d'(\cdot, y_i)$ em torno de y_i .

3.3 Método de Minimização Proximal com Reescala-MMPR

Esta seção aplica a análise do método **MPRDV** para a minimização do problema (3.1). Deixaremos a hipótese (3.1) como uma hipótese de apoio, faremos também a hipótese de regularização padrão, em vista das propriedades da função barreira D , que é requerido para alguma aplicação de (3.3).

Hipótese 3.8. $\text{dom } f \cap \text{int } B \neq \emptyset$.

3.4 Descrição do Método-MMPR

1. **Inicialização:** Escolha $c > 0$ e $\sigma \in [0, 1]$. Escolha seqüências escalares não negativas s_k e z_k com $\sum_{k=1}^{\infty} s_k < \infty$ e $z_k \rightarrow 0$. Seja $k = 0$ e $x^0 \in \text{int } B$.

2. **Iteração:**

(a) Escolha $\alpha^k \in \mathbb{R}_{++}^n$ tal que $\alpha_i^k \geq c \max\{1, d_i''(x_i^k, x_i^k)\}$ para $i = 1, 2, \dots, n$.

(b) Procure $x^{k+1}, e^{k+1} \in \mathbb{R}^n$ tal que

$$e^{k+1} \in \partial f(x^{k+1}) + \text{diag}(\alpha^k)^{-1} \nabla_1 D(x^{k+1}, x^k) \quad (3.9)$$

$$|e_i^{k+1}| \leq \frac{\sigma}{\alpha_i^k} |d_i'(x_i^{k+1}, x_i^k)| + \min \left\{ \frac{s_{k+1}}{\|x^{k+1} - x^k\|}, z_{k+1} \right\}, \quad i = 1, 2, \dots, n, \quad (3.10)$$

Com a convenção de que $\min\{s_{k+1}/\|x^{k+1} - x^k\|, z_{k+1}\}$ é z_{k+1} sempre que $x^{k+1} = x^k$.

(c) Seja $k = k + 1$, e repita a iteração.

Note que, se uma das escolhas para $s_k = 0$, ou $z_k = 0$ para todo k , então (3.10) reduz-se ao critério construtivo

$$|e_i^{k+1}| \leq \frac{\sigma}{\alpha_i^k} |d_i'(x_i^{k+1}, x_i^k)|,$$

uma vez que $\min\{s_{k+1}/\|x^{k+1} - x^k\|, z_{k+1}\} = 0$.

3.5 Análise da Convergência

Iniciaremos mostrando que a iteração passo está bem definida se f é limitada inferiormente em B .

Lema 3.9. *Se f é limitada inferiormente em B , então o algoritmo **MMPR** está bem definido com $e^{k+1} = 0$. Desta maneira, a solução para (3.9) e (3.10) existe se f é limitada inferiormente em B .*

Demonstração. Considere l o limite inferior de f em B . Dado $\zeta \in \mathbb{R}$, considere o conjunto de nível

$$\left\{ x \in B : f(x) + \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \leq \zeta \right\} \subseteq \left\{ x \in B : \sum_{i=1}^n \frac{1}{\alpha_i^k} d_i(x_i, x_i^k) \leq \zeta - l \right\}.$$

Este último conjunto é um conjunto de nível de $\sum_{i=1}^n (1/\alpha_i^k) d_i(\cdot, x_i^k)$ em B , o qual será limitado, visto que pela hipótese (3.1)(a) esta função atinge seu mínimo no único ponto x^k , logo o primeiro conjunto de nível será limitado para todo $\zeta \in \mathbb{R}$, pelo Teorema (1.15). Portanto, $f(\cdot) + \sum_{i=1}^n (1/\alpha_i^k) d_i(\cdot, x_i^k)$ atinge um mínimo em B . A unicidade do mínimo segue da convexidade estrita de $D(\cdot, x^k)$. \square

Para aplicar a análise da convergência da seqüência $\{x^k\}$ do **MMPR**, é suficiente mostrar que a hipótese (3.4) está assegurada, ou seja, que existe a seqüência $\{\beta_k\}$ e que a seqüência erro satisfaça (3.10), para isto faremos:

Lema 3.10. *Definindo*

$$\beta_k := \min \left\{ \frac{s_k}{\|x^k - x^{k-1}\|}, z_k \right\}$$

para todo $k \geq 1$, a hipótese (3.4)(a) está assegurada para o **MMPR**.

Demonstração. Como s_k e z_k são não negativas segue-se que β_k é também não negativa. Visto que $z_k \rightarrow 0$, temos também que $\beta_k \rightarrow 0$. Além disso, como $\sigma \in [0, 1]$,

$$|e_i^k| \leq \frac{\sigma}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| + \beta_k,$$

para todo k , então a hipótese (3.4)(a) é válida. \square

Como em (3.6), definimos para todo $k \geq 0$ e $i = 1, \dots, n$,

$$\gamma_i^k = e_i^k - \frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^{k-1}),$$

e seja $\gamma^k \in \mathbb{R}^n$ um vetor com coordenadas γ_i^k .

Lema 3.11. $\gamma^k \in \partial f(x^k)$ e $\gamma_i^k(x_i^{k-1} - x_i^k) \geq -s_k$ para todo $k \geq 0$ e $i = 1, \dots, n$.

Demonstração. A afirmação que $\gamma^k \in \partial f(x^k)$ segue da definição de γ^k . Para a segunda afirmação, temos, usando a convexidade de $d_i(\cdot, x_i^{k-1})$,

$$\gamma_i^k(x_i^{k-1} - x_i^k) = \left(e_i^k - \frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^k) \right) (x_i^{k-1} - x_i^k)$$

Podemos escrever

$$\left(e_i^k - \frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^k) \right) (x_i^{k-1} - x_i^k) \geq -\frac{1}{\alpha_i^{k-1}} d'_i(x_i^k, x_i^{k-1})(x_i^{k-1} - x_i^k) - |e_i^k| |x_i^{k-1} - x_i^k|$$

Como $d'_i(x_i^k, x_i^{k-1})(x_i^{k-1} - x_i^k) \leq 0$, obtemos:

$$\gamma_i^k(x_i^{k-1} - x_i^k) = \left(\frac{1}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| - |e_i^k| \right) |x_i^{k-1} - x_i^k|.$$

Usando (3.10), segue-se então que:

$$\gamma_i^k(x_i^{k-1} - x_i^k) \geq \left(\frac{1-\sigma}{\alpha_i^k} |d'_i(x_i^k, x_i^{k-1})| - \min \left\{ \frac{s_k}{\|x^k - x^{k-1}\|}, z_k \right\} \right) |x_i^{k-1} - x_i^k|$$

Desprezando a primeira parcela dentro do parênteses, temos:

$$\gamma_i^k(x_i^{k-1} - x_i^k) \geq -\min \left\{ \frac{s_k}{\|x^k - x^{k-1}\|}, z_k \right\} |x_i^{k-1} - x_i^k|$$

Agora considerando a norma, obtemos, finalmente:

$$\gamma_i^k(x_i^{k-1} - x_i^k) \geq -\min \left\{ \frac{s_k}{\|x^k - x^{k-1}\|}, z_k \right\} \|x_i^{k-1} - x_i^k\| \geq -s_k.$$

□

Agora é possível determinar que a hipótese (3.4)(b) também é válida para o **MMPR**:

Lema 3.12. *Se f é limitada inferiormente em B , então $\{f(x^k)\}$ é convergente e*

$$|\gamma_i^k| |x_i^{k-1} - x_i^k| \rightarrow 0, \quad \forall i = 1, \dots, n.$$

*Deste modo a hipótese (3.4)(b) é válida para o **MMPR**.*

Demonstração. Usando o Lema (3.11),

$$f(x^{k-1}) \geq f(x^k) + \langle \gamma^k, x^{k-1} - x^k \rangle \geq f(x^k) - ns_k.$$

Agora, lembrando que $\{s_k\}$ é somável, do Lema (1.17) temos que $\{f(x^k)\}$ é uma seqüência convergente. Para todo $i = 1, \dots, n$, temos também que

$$f(x^{k-1}) \geq f(x^k) + \langle \gamma^k, x^{k-1} - x^k \rangle \geq f(x^k) - (n-1)s_k + \gamma_i^k(x_i^{k-1} - x_i^k).$$

E usando o Lema (3.11) novamente, segue-se que

$$f(x^{k-1}) - f(x^k) + (n-1)s_k \geq \gamma_i^k(x_i^{k-1} - x_i^k) \geq -s_k.$$

Passando o limite com $k \rightarrow \infty$, concluímos que $\gamma_i^k(x_i^{k-1} - x_i^k) \rightarrow 0$. \square

Desta maneira, o teorema (3.7) implica na otimalidade de todos os pontos de acumulação da seqüência $\{x^k\}$. Reforçamos esta idéia no seguinte

Teorema 3.13. *Suponha que as hipóteses (3.1) e (3.8) sejam válidas e que f seja limitada inferiormente em B . Se $\{x^k\}$ possui um ponto de acumulação, então $\{f(x^k)\}$ converge ao ínfimo de f em B e todo ponto de acumulação de $\{x^k\}$ será um mínimo de f em B . Uma condição que garante a existência do ponto de acumulação de $\{x^k\}$ é o fato do conjunto solução ser limitado, ou qualquer outro conjunto de nível de f .*

Demonstração. Como acabamos de ver, Lema (3.12) implica que a hipótese (3.4)(b) é válida, e então a hipótese (3.4) também é válida em sua totalidade. A condição (3.1) assegurada por hipótese, e, tomando $T = \partial f$, a hipótese (3.8) implica a hipótese (3.3). Desta forma, as conclusões do Teorema (3.7) se aplicam. Seja \bar{x} um ponto de acumulação da seqüência $\{x^k\}$, isto é, $x^k \rightarrow_{\mathcal{K}} \bar{x}$, para algum conjunto infinito $\mathcal{K} \subset \mathbb{N}$. O Teorema (3.7) afirma que $0 \in \partial f(\bar{x}) + N_B(\bar{x})$; pela hipótese (3.8), \bar{x} é um minimizador de f em B . No entanto, visto que o Lema (3.6) estabelece que $\{\gamma^k\}_{\mathcal{K}}$ é limitada, e visto que $\{f(x^k)\}$ é convergente pelo Lema (3.12),

$$\min_{x \in B} f(x) = f(\bar{x}) \geq f(x^k) + \sum_{i=1}^n \gamma_i^k(\bar{x} - x_i^k) \xrightarrow{\mathcal{K}} \lim_{k \rightarrow \infty} f(x^k) \geq f(\bar{x}).$$

Portanto, $\lim_{k \rightarrow \infty} f(x^k) = f(\bar{x})$.

Finalmente, do fato de qualquer conjunto de nível de uma função convexa fechada própria ser limitado implica na limitação de todos os conjuntos de níveis, Teorema (1.15) e o Lema (3.12) estabelece que $\{f(x^k)\}$ é convergente, conseqüentemente limitada. Então, $\{x^k\}$ é também limitada e portanto possui ponto de acumulação. \square

Capítulo 4

Método Ponto Proximal Interior de Bregman para Desigualdade Variacional

Voltemos nossa atenção para a caixa do problema (3.2) da desigualdade variacional, onde $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ é um operador monótono maximal. Neste capítulo usaremos a distância de Bregman, como definido em (2.11). Estenderemos a hipótese (3.3) como segue

Hipótese 4.1. *T é monótono maximal, o conjunto solução do problema (3.2) é não vazia, e existe algum $\tilde{x} \in \text{dom} T \cap \text{int} B$.*

Nossa meta é mostrar a convergência de uma versão aproximada da iterada (3.4); sem hipóteses adicionais sobre T .

Hipótese 4.2. *Para $i = 1, 2, 3, \dots, n$, as funções $h_i : \mathbb{R} \rightarrow (-\infty, \infty]$ possui as seguintes propriedades:*

(a) *h_i é fechada, $\text{int dom } h = (a_i, b_i)$, e h_i é duas vezes continuamente diferenciável, com segunda derivada estritamente positiva em (a_i, b_i) .*

(b) *h_i é essencialmente suave.*

(c) *Existe um $\rho > 0$ e $\epsilon > 0$ tal que se ocorre $-\infty < a_i < y_i \leq x_i < a_i + \epsilon$ ou $b_i - \epsilon < x_i \leq y_i <$*

$b_i < +\infty$, então $\rho|h'_i(x_i) - h'_i(y_i)| \leq h''_i(y_i)|x_i - y_i|$. Além disso h_i é contínua em $[a_i, b_i] \cap \mathbb{R}$. E mais, definimos $h(x) = \sum_{i=1}^n h_i(x_i)$ e $D_h(x, y) = \sum_{i=1}^n h_i(x_i) - h_i(y_i) + h'_i(y_i)(x_i - y_i)$:

(d) Para todo $x \in B$ e $\alpha \in \mathbb{R}$, o conjunto de nível $\{y \in \text{int } B \mid D_h(x, y) \leq \alpha\}$ é limitado.

(e) Se $\{x^k\} \subset B$ converge para $x \in \mathbb{R}^n$, então $\lim_{k \rightarrow \infty} D_h(x, x^k) = 0$.

(f) $\text{Im}(h') = \mathbb{R}$.

Exemplo 4.3. Considere $d_i(x_i, y_i) = h_i(x_i) - h_i(y_i) - h'_i(y_i)(x_i - y_i)$ onde h_i pode ser:

- $h_i(x) = \frac{1}{2}x^2$, com $a_i = -\infty$, e $b_i = +\infty$;
- $h_i(x) = x \log x$, com $a_i = 0$, e $b_i = +\infty$;
- $h_i(x) = x \log(e^x - 1)$, com $a_i = 0$, e $b_i = +\infty$;

Note que para os a_i 's e b_i 's finitos, a correspondente h_i é agora requerida a tomar um valor finito. Podemos agora estabelecer o algoritmo:

Algoritmo Ponto Proximal Interior da Caixa– (APPIC)

1. **Inicialização:** Seja $k = 0$, e fixe algum escalar $c > 0$. Tome $x^0 \in \text{int } B$.

2. **Iteração:** Escolha $\alpha_i^k \geq c \max\{1, h''_1(x_1^k), \dots, h''_n(x_n^k)\}$. Encontre vetores $x^{k+1}, e^{k+1} \in \mathbb{R}^n$ tal que

$$e^{k+1} \in T(x^{k+1}) + \frac{1}{\alpha_k} \nabla_1 D_h(x^{k+1}, x^k) = T(x^{k+1}) + \frac{1}{\alpha_k} \left(\nabla h(x^{k+1}) - \nabla h(x^k) \right) \quad (4.1)$$

3. Seja $k = k + 1$ e repita a iteração.

4.1 Análise da Convergência

Primeiro citaremos um resultado mostrando que a iteração do passo **APPIC** é bem definida:

Lema 4.4. Sob a hipótese (4.2), há um único ponto x^{k+1} que resolve a iteração passo (4.1) do **APPIC** com $e^{k+1} = 0$.

Demonstração. Ver [12], Teorema 4(i). □

Para garantir a convergência do **APPIC**, assumiremos alguns comportamentos para $\{e^k\}$. Embora não como um critério geral usado no **MMPR**, estas condições são melhores situadas para nossas análises, visto que elas permitirão usar propriedades associadas com monotonicidade Fejér.

Hipótese 4.5. *A seqüência erro $\{e^k\}$ satisfaz as seguintes propriedades:*

$$\sum_{k=0}^{\infty} \alpha_k \|e^{k+1}\| < +\infty;$$

e

$$\sum_{k=0}^{\infty} \alpha_k \langle e^{k+1}, x^{k+1} \rangle$$

existe e é finito.

Perceba que a Hipótese (4.5) implica que $\|e^k\| \rightarrow 0$, com $k \rightarrow \infty$, e desta forma a hipótese (3.4)(a) está assegurada, pois tomando $\beta_k = \{\|e^1\|, \|e^2\|, \|e^3\|, \dots, \|e^k\|, \dots\}$, teremos que $|e_i^k| \leq \frac{1}{\alpha_i^{k-1}} |d'_i(x_i^k, x_i^{k-1})| + \beta_k$. Apresentaremos agora dois lemas necessários:

Lema 4.6. *Seja $z \in (T + N_B)^{-1}(0)$. Então, para todo $k \geq 0$,*

$$D_h(z, x^{k+1}) \leq D_h(z, x^k) - D_h(x^{k+1}, x^k) + \alpha_k \langle e^{k+1}, x^{k+1} - z \rangle. \quad (4.2)$$

Demonstração. Análogo ao Lema (2.23). □

Lema 4.7. *Se a hipótese (4.5) está assegurada, então a seqüência $\{x^k\}$ é limitada e $D_h(x^{k+1}, x^k) \rightarrow 0$.*

Demonstração. Como $\sum_{k=0}^{\infty} \alpha_k \|e^{k+1}\|$ e $\sum_{k=0}^{\infty} \alpha_k \langle e^{k+1}, x^{k+1} \rangle$ convergem, então $\langle e^{k+1}, x^{k+1} \rangle \rightarrow 0$ e $\|e^{k+1}\| \rightarrow 0$, logo $\{x^k\}$ é limitada. Seja, agora, $z \in (T + N_B)^{-1}(0)$, então

$$E(z) := \sum_{k=0}^{\infty} \alpha_k \langle e^{k+1}, x^{k+1} - z \rangle = \sum_{k=0}^{\infty} \left(\alpha_k \langle e^{k+1}, x^{k+1} \rangle - \alpha_k \langle e^{k+1}, z \rangle \right)$$

existe e é finito. Mas,

$$\sum_{k=0}^{\infty} |\alpha_k \langle e^{k+1}, z \rangle| \leq \sum_{k=0}^{\infty} \alpha_k \|e^{k+1}\| \|z\|.$$

e a hipótese (4.5) implicam que o lado direito desta inequação é finito. Deste modo, $\sum_{k=0}^{\infty} \alpha_k \langle e^{k+1}, z \rangle$ existe e é finito. Usando a hipótese (4.5) mais uma vez, concluímos que $E(z)$ existe e é finito. Desta forma fazendo $k \rightarrow \infty$ em (4.2), temos que, como $\alpha_k \langle e^{k+1}, x^{k+1} - z \rangle \rightarrow 0$, $D_h(x^{k+1}, x^k) \rightarrow 0$. \square

Teorema 4.8. *Seja h_i satisfazendo a hipótese (4.2). Dada duas seqüências $\{x^k\} \subset B$ e $\{y^k\} \subset \text{int } B$, se uma das seqüências é convergente, com $\lim_{k \rightarrow \infty} D_h(x^k, y^k) = 0$, então a outra seqüência também converge ao mesmo limite.*

Demonstração. Veja o Teorema (2.15). \square

Este teorema implica que $h(x) = \sum_{i=1}^n h_i(x_i)$ é uma função de Bregman no sentido clássico. Usando o Teorema (4.8) e o Lema (4.7), obtemos:

Corolário 4.9. *Sob as hipóteses (4.1), (4.2) e (4.5), a seqüência $\{x^k\}$ possui pelo menos um ponto de acumulação. No entanto, se para algum conjunto infinito $\mathcal{K} \subseteq \mathbb{N}$, tivermos $x^k \rightarrow_k \bar{x}$, então $x^{k-1} \rightarrow_k \bar{x}$. Portanto a hipótese (3.4)(b) é válida.*

Antes de apresentar o principal teorema da convergência do **APPIC**, apresentaremos um lema técnico final que nos ajudará a provar a unicidade do ponto de acumulação de $\{x^k\}$.

Lema 4.10. *Sob a hipótese (4.5), para todo $z \in (T + N_B)^{-1}(0)$, $D_h(z, x^k)$ converge a um valor em $[0, +\infty)$ o qual denotaremos por $d(z)$.*

Demonstração. Considere algum $z \in (T + N_B)^{-1}(0)$. Então o Lema (4.6) implica que a desigualdade (4.2) está assegurada. Usando a hipótese (4.5) e $D_h(x^{k+1}, x^k) \geq 0$, para todo k , então as hipóteses do Lema (1.17) são satisfeitas com $a_k = D_h(z, x^k)$ e $\gamma_k = \alpha_k \langle e^{k+1}, x^{k+1} - z \rangle$. Portanto, $\{D_h(z, x^k)\}$ converge, necessariamente a um valor não negativo. \square

Teorema 4.11. *Sob as hipóteses (4.1), (4.2) e (4.5), a seqüência $\{x^k\}$ converge a uma solução de $0 \in T(x) + N_B(x)$.*

Demonstração. Deixe \bar{x} ser um ponto de acumulação de $\{x^k\}$, isto é, $x^k \rightarrow_{\mathcal{K}} \bar{x}$, para algum conjunto infinito $\mathcal{K} \subseteq \mathbb{N}$. Tal ponto existe pelo Lema (4.7). Do Teorema (3.7), $0 \in T(\bar{x}) + N_B(\bar{x})$.

Provaremos agora a unicidade do ponto de acumulação: da hipótese (4.2)(b), sabemos que $D_h(\bar{x}, x^k) \rightarrow_{\mathcal{K}} 0$. Então, $d(\bar{x})$, como definido no Lema (4.10), é zero. Suponha que $\{x^k\}$ possui outro ponto de acumulação x' , ou seja, $x^k \rightarrow_{\mathcal{K}'} x'$ para algum conjunto infinito $\mathcal{K}' \subseteq \mathbb{N}$. Temos então que $D_h(\bar{x}, x^k) \rightarrow_{\mathcal{K}'} d(\bar{x}) = 0$, e segue-se do Teorema (4.8) que $x' = \bar{x}$. \square

4.2 Conclusão

Os métodos **MPRDV**, **MMPR** e **APPIC** utilizam um vetor $\alpha_i^k \in \mathbb{R}_{++}^n$ onde cada coordenada é atualizada em cada iteração, observando a curvatura da função d_i ou da função h_i , (derivada segunda). Deste modo, nas provas de convergência, temos menos hipóteses sobre o operador T , como, por exemplo, Paramonotonicidade e Pseudo-monotnicidade [26].

Outra possível aplicação de nossa análise fundamental é tentar generalizar a idéia de adicionar a norma Euclideana à uma distância generalizada para obter monotonicidade Fejér para solucionar o problema (3.2), como proposto em [4, 5] para um caso especial de φ -divergente.

Referências Bibliográficas

- [1] Izmailov, Alexey, Solodov, Mikhail, Otimização volume 1. Condições de Otimidade, Elementos de Análise Convexa e de Dualidade, Instituto de Matemática Pura e Aplicada, Rio de Janeiro (2005).
- [2] Iusem, Alfredo N. On some properties of generalized proximal point methods for quadratic and linear programming. *Journal of Optimization Theory and Applications* 85, 593–612 (1995)
- [3] Iusem, Alfredo. Steepest descent methods with generalized distances for constrained optimization.
- [4] A. Auslender, M. Teboulle, and S. Ben-Tiba. Interior proximal and multiplier methods based on second order homogeneous kernels. *Mathematics of Operations Research*, 24:645–668, (1999).
- [5] A. Auslender, M. Teboulle, and S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications* 12:31-40, 1999.
- [6] Minty, B. Monotone nonlinear operators in Hilbert space. *Duke Mathematical Journal* 29, 341–346 (1978).
- [7] A. Ben-Tal and M. Zibulevsky. Penalty/ barrier multiplier methods for convex programming problems. *SIAM Journal of Optimization*, 7:347–366, (1997).
- [8] D. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Athena Scientific, (1996).

-
- [9] L.M.Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217 (1967).
- [10] G.Chen and M.Tebboulle. A convergence analysis of proximal-like minimization algorithms using Bregman functions.*Mathematics of Operations Research*,18:202–226,(1993).
- [11] J.Eckstein, Approximate iterations in Bregman-functions-based proximal algorithms,*Mathematical Programming*, 83:113–124, (1998).
- [12] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, 18:202–226, (1993).
- [13] J.B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms I and II*, Springer-Verlag (1993).
- [14] C. Humes e P.Silva. Strict convex regularizations, proximal point and augmented Lagrangians. *RAIRO Operations Research*, 34:283-303, (2000).
- [15] A. N. Iusem, Proximal point methods in optimization, 20⁰ Colóquio Brasileiro de Matemática, IMPA, (1995).
- [16] A.N. Iusem, B.F. Svaiter and M. Teboulle. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research*, 19:790–814, (1994).
- [17] E. L. Lima, *Curso de análise, Volume 2*, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1981.
- [18] G. J. Minty, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.*, **29** (1962), 341–346.
- [19] J. J. Moreau, Proximité et dualité dans un espace Hilbertien, *Bull. Soc. Math., France*, **93** (1965), 273–299.
- [20] J. M. Ortega, W.C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, Academic Press, Inc, New York (1970).

-
- [21] José, Paulo da Silva e Silva, Eckstein, Jonathan, Humes, Carlos Jr, Rescaling and Step-size Selection in Proximal Methods using Separable Generalized Distances, RUTCOR Research Report (1999), 35–99.
- [22] B. Polyak. Introduction to Optimization. Optimization Software Inc., New York, (1987).
- [23] R. T. Rockafellar, Convex analysis, Princeton University Press, Princeton, 1970.
- [24] R. T. Rockafellar. Monotone operators and the proximal point algorithm. SIAM J. cont. optim. **14** (1976), 877–898.
- [25] R. T. Rockafellar. Extension of Fenchel’s duality theorem for convex functions. Duke Math. J., 33:81-89, (1996).
- [26] R. T. Rockafellar and R. J. B. Wets. Variational Analysis Springer-Verlag, Berlin, (1998).
- [27] Rockafellar, R. Tyrrell. Monotone Operators and the Proximal Point Algorithm. SIAM J. Control and Optimization 14, 877-898 (1976).
- [28] M. V. Solodv and B. F. Svaiter. An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. Mathematics of Operations Research, 25:214–230, (2000).
- [29] Fonseca, Wilton Rodrigues. Uma abordagem sobre o Algoritmo do Ponto Proximal. Dissertação- Universidade Federal de Goiás, IME–2003.