



UM MÉTODO MULTIPLICADOR PROXIMAL PARA MINIMIZAÇÃO CONVEXA SEPARÁVEL

Orlando Sarmiento Chumbes

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Paulo Roberto Oliveira

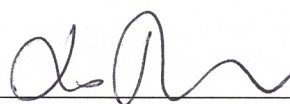
Rio de Janeiro
Janeiro de 2015

UM MÉTODO MULTIPLICADOR PROXIMAL PARA MINIMIZAÇÃO
CONVEXA SEPARÁVEL

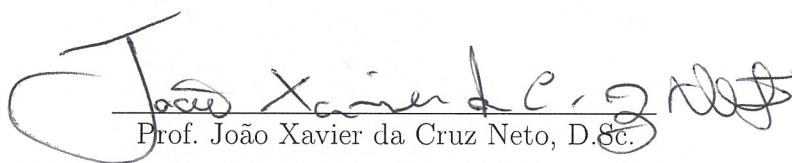
Orlando Sarmiento Chumbes

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE
SISTEMAS E COMPUTAÇÃO.

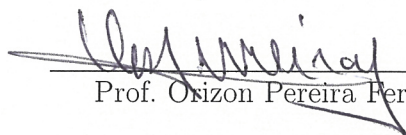
Examinada por:



Prof. Paulo Roberto Oliveira, D.Sc.



Prof. João Xavier da Cruz Neto, D.Sc.



Prof. Osizon Pereira Ferreira, D.Sc.



Prof. Ernesto Prado Lopes, Ph.D.

RIO DE JANEIRO, RJ – BRASIL
JANEIRO DE 2015

Chumbes, Orlando Sarmiento

Um método multiplicador proximal para minimização convexa separável/Orlando Sarmiento Chumbes. – Rio de Janeiro: UFRJ/COPPE, 2015.

XI, 70 p.: il.; 29,7cm.

Orientador: Paulo Roberto Oliveira

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2015.

Referências Bibliográficas: p. 68 – 70.

1. Método multiplicador proximal. 2. Problemas convexos separables. 3. Distâncias proximais. I. Oliveira, Paulo Roberto. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Dedico este trabalho à memória
de meus pais Alipio Sarmiento e
Marcosa Chumbes.*

*Aos meus irmãos César Alberto,
Marisol e José Luis, e à minha
filha Maria Isabel.*

Agradecimentos

Quero agradecer acima de tudo a Deus por estar conosco em cada momento de nossas vidas.

Agradeço à minha família, à minha mãe Marcosa e ao meu pai Alipio que fizeram tantos sacrifícios para que eu pudesse ter bom estudo e assim ter boas oportunidades como venho tendo. Se não fosse eles, hoje provavelmente não teria chegando onde cheguei.

Agradeço aos meus irmãos, César Alberto, José Luis e Marisol, pelo apoio direto ou indireto durante toda a minha vida acadêmica.

Ao Prof. Paulo Roberto Oliveira, pela orientação, apoio, amizade e sobre tudo pela confiança depositada.

Aos membros da banca examinadora Prof. João Xavier da Cruz Neto, Prof. Orizon Pereira Ferreira e Prof. Ernesto Prado Lopes, cujas observações, críticas e sugestões enriqueceram esta dissertação.

Ao Prof. Erik Alex Papa Quiroz que me orienta desde a graduação, onde tive grande motivação pelo tema de pesquisa que fui apresentado e estudo até hoje. E não só nisso mas também foi que abriu as portas para que eu começasse minha vida na pesquisa acadêmica, muito obrigado.

Agradeço à minha namorada Liliana, muito obrigado pelo apoio, carinho, companhia e consideração.

Agradeço aos meus colegas Francisco, Leonardo, Felipe, Pedro, Nancy, João e Regina, por todo esse período de estar ali todos os dias e partilhar momentos de estudos, histórias e tantos outros fatos.

Quero agradecer também à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo auxílio financeiro que possibilitou a realização deste trabalho, e aos professores do Programa de Engenharia de Sistemas e Computação da COPPE-UFRJ, em especial aos da linha de Otimização.

Considerando esta dissertação como resultado de uma caminhada que não começou na COPPE-UFRJ, agradecer pode não ser tarefa fácil, muito menos justa. Para não correr o risco da injustiça, agradeço a todos que de alguma forma passaram pela minha vida e contribuíram para a construção desta dissertação.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UM MÉTODO MULTIPLICADOR PROXIMAL PARA MINIMIZAÇÃO CONVEXA SEPARÁVEL

Orlando Sarmiento Chumbes

Janeiro/2015

Orientador: Paulo Roberto Oliveira

Programa: Engenharia de Sistemas e Computação

Atualmente, existem várias áreas que procuram resolver problemas nas Ciências e Engenharia, uma das quais é a Otimização. Esse trabalho considera os problemas de otimização convexa com uma estrutura separável, isto é, problemas de minimizar a soma de funções convexas sujeitas a restrições para cada variável independente. Para a solução desses problemas foram estudados vários métodos propostos na literatura, como os trabalhos de Chen e Teboulle (Método Multiplicador Preditor-Corretor Proximal), Kyono e Fukushima (Método Multiplicador Preditor-Corretor Proximal Não Linear) e Auslender e Teboulle (Método de Decomposição Proximal Entrópico).

Nesta dissertação propomos um método multiplicador proximal inexato usando distâncias proximais, este método unificou os métodos citados acima, e além disso, amplia as propriedades de convergência para o método usando a classe de distâncias φ -divergência.

Demonstramos, sob determinadas hipóteses, que as iterações geradas pelo método proposto, são bem definidas e toda sequência gerada por este método converge para uma solução ótima do problema com uma taxa de convergência linear. Além disso, mostramos alguns resultados computacionais comparando a funcionalidade de algumas distâncias proximais conhecidas na literatura.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A PROXIMAL MULTIPLIER METHOD FOR SEPARABLE CONVEX MINIMIZATION

Orlando Sarmiento Chumbes

January/2015

Advisor: Paulo Roberto Oliveira

Department: Systems Engineering and Computer Science

Actually, there are several areas that seek to solve problems in the Sciences and Engineering, one of which is Optimization. This work considers convex optimization problems with a separable structure, i.e., to minimize problems the sum of convex functions subject restrictions for each independent variable. To solve these problems proposed various methods have been studied in the literature, such as the works of Chen and Teboulle (Predictor-Corrector Proximal Multiplier Method), Kyono and Fukushima (Nonlinear Predictor-Corrector Proximal Multiplier Method) and Auslender and Teboulle (Entropic Proximal Decomposition Method).

This dissertation, we propose an inexact proximal multiplier method using proximal distances, this method unified the methods mentioned above, and furthermore, extends the convergence properties for the method using the class of φ -divergence distances.

We demonstrate, under appropriate assumptions, that the iterations generated by the method propose are well defined and every sequence generated by this method converges to an optimal solution of the problem with a linear convergence rate. Moreover, we show some computational results comparing the functionality of various proximal distances known in the literature.

Sumário

Lista de Figuras	ix
Lista de Símbolos	x
1 Introdução	1
2 Resultados Preliminares	5
2.1 Propriedades Topológicas e Elementos de Análise Convexa	6
2.2 Distâncias Proximais	12
3 Método Multiplicador Proximal	18
3.1 O Problema	18
3.2 Um Modelo de Rede de Telecomunicações	19
3.3 O Algoritmo (PMAPD)	24
3.4 Análise de Convergência do PMAPD	26
3.5 Convergência para o Caso Geral	34
4 Taxa de Convergência e Implementações do PMAPD	39
4.1 O Algoritmo (PMAPD) sob um Adequado Critério de Aproximação	40
4.2 Taxa de Convergência do PMAPD	42
4.3 Experimentos Numéricos com o Algoritmo (PMAPD)	55
5 Conclusões e Trabalhos Futuros	67
Referências Bibliográficas	68

Lista de Figuras

2.1	Conjunto convexo e não convexo.	7
2.2	Ilustração da definição de função convexa.	8
2.3	Os elementos y^1 e y^2 são subgradientes de f em x	9
2.4	Ilustração da definição de ϵ -subdiferencial de f em x	10
2.5	Construção de $D_h(x, y)$	14
3.1	Grafo direcionado de 4 nós e 5 arestas	20

Lista de Símbolos

$0^+(\mathbf{C})$	O cone de recessão do conjunto $\mathbf{C} \subseteq \mathbb{R}^n$, p. 11
$C^1(\Omega)$	O conjunto de funções continuamente diferenciáveis em Ω , p. 16
$C^2(\Omega)$	O conjunto $\{f : f' \in C^1(\Omega)\}$, p. 16
$L_f(\alpha)$	O conjunto de nível da função f , p. 8
$\bar{\mathbf{C}}$	O fecho do conjunto $\mathbf{C} \subseteq \mathbb{R}^n$, p. 6
$\delta_{\mathbf{C}}$	Função indicadora do conjunto $\mathbf{C} \subseteq \mathbb{R}^n$, p. 10
$\arg \min_{x \in \mathbb{R}^n} \{f(x)\}$	O conjunto dos minimizadores da função f , p. 29
$\liminf_{k \rightarrow +\infty} f(x^k)$	Limite inferior da sequência de números reais $\{f(x^k)\}$, p. 8
$\limsup_{k \rightarrow +\infty} f(x^k)$	Limite superior da sequência de números reais $\{f(x^k)\}$, p. 33
$\nabla f(x)$	O gradiente da função f no ponto x , p. 13
$\partial f(x)$	O subdiferencial da função convexa f no ponto x , p. 9
$\partial_{\epsilon} f(x)$	O ϵ -subdiferencial da função convexa f no ponto x , p. 10
\mathbb{R}	O conjunto dos números reais, p. 6
$\mathbb{R} \cup \{\pm\infty\}$	Significa $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$, p. 6
\mathbb{R}^n	O espaço euclidiano, p. 6
$\text{epi}(f)$	O epígrafo da função f , p. 8
$\text{ir}(\mathbf{C})$	O interior relativo do conjunto $\mathbf{C} \subseteq \mathbb{R}^n$, p. 7
$\{x^k\}$	Uma sequência de pontos no \mathbb{R}^n , p. 5
$\text{aff}(S)$	O envoltório afim de $S \subseteq \mathbb{R}^n$, p. 7

$dom(f)$	O domínio efetivo da função f , p. 8
f^0	Função de recessão da função f , p. 11
f^*	Função conjugada da função f , p. 11
$int(C)$	O interior do conjunto $C \subseteq \mathbb{R}^n$, p. 6

Capítulo 1

Introdução

Diversos problemas na programação matemática podem ser formulados pelo seguinte problema de otimização convexo separável:

$$(PC) \quad \min\{f(x) + g(z) : Ax + Bz = b, x \in \bar{\mathcal{C}}, z \in \bar{\mathcal{K}}\},$$

onde $\mathcal{C} \subset \mathbb{R}^n$ e $\mathcal{K} \subset \mathbb{R}^p$ são conjuntos convexos abertos não vazios, $\bar{\mathcal{C}}$ e $\bar{\mathcal{K}}$ denotam o fecho (na topologia euclidiana) de \mathcal{C} e \mathcal{K} respectivamente, $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ e $g : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ são funções próprias, semicontínuas inferiores, convexas e $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$, $b \in \mathbb{R}^m$.

O problema dual de Fenchel, veja [23], para o (PC) é definida como:

$$(D) \quad \max\{-(f + \delta_{\bar{\mathcal{C}}})^*(-A^T y) - (g + \delta_{\bar{\mathcal{K}}})^*(-B^T y) - \langle y, b \rangle : y \in \mathbb{R}^m\}$$

onde $\delta_{\mathcal{X}}$ denota a função indicadora, $(f + \delta_{\bar{\mathcal{C}}})^*$, $(g + \delta_{\bar{\mathcal{K}}})^*$ são as funções conjugadas de $f + \delta_{\bar{\mathcal{C}}}$ e $g + \delta_{\bar{\mathcal{K}}}$, respectivamente, e $\langle \cdot, \cdot \rangle$ denota o produto interno canônico.

Nas últimas décadas, há um grande interesse em estudar a estrutura separável do (PC), esse modelo foi encontrado em vários problemas de otimização, por exemplo, em Telecomunicações, veja Mahey et. al. [19], gerenciamento de Energia Elétrica, veja Lenoir [17], e em Ciência da Computação para resolver problemas de completção de uma matriz, veja o exemplo 2 de Goldfarb et. al. [13].

Na literatura existem diversos métodos para resolver o (PC). Os métodos clássicos estão baseados no algoritmo de ponto proximal (APP), como por exemplo, o método dos multiplicadores Bertsekas [6], que é obtido através da aplicação do APP para o problema dual (D), e o método dos multiplicadores proximal Rockafellar [22], que é obtido através da aplicação do APP para o problema primal-dual. No entanto, esses algoritmos não aproveitam ao máximo a estrutura separável do problema original (PC).

Vários métodos de decomposição que exploram a estrutura especial para o problema têm sido propostos. Alguns exemplos de tais métodos são: o método dos multiplicadores de direções alternadas [7, 11, 12], o método inverso parcial do Spingarn [10, 18, 26, 27], os métodos dos multiplicadores preditor-corretor proximais de Chen e Teboulle e suas extensões [3, 8, 16].

Para o caso particular quando $\mathbf{C} = \mathbb{R}^n$, $\mathbf{K} = \mathbb{R}^m$, $B = -I$ e $b = 0$ no (PC), a versão exata do Chen e Teboulle [8], chamado método Multiplicador Preditor-Corretor Proximal (MPCP), dado um ponto inicial (x^0, z^0, y^0) , gera uma sequência de pontos $\{(x^k, z^k, y^k)\}$ com o seguinte esquema iterativo:

$$p^{k+1} = y^k + \lambda_k(Ax^k - z^k) \quad (1.1)$$

$$x^{k+1} = \arg \min \{f(x) + \langle p^{k+1}, Ax \rangle + (1/2\lambda_k)\|x - x^k\|^2\} \quad (1.2)$$

$$z^{k+1} = \arg \min \{g(z) - \langle p^{k+1}, z \rangle + (1/2\lambda_k)\|z - z^k\|^2\} \quad (1.3)$$

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} - z^{k+1}) \quad (1.4)$$

onde λ_k é um parâmetro positivo. Assumindo que existe uma solução primal-dual ótima (x^*, z^*, y^*) , Chen e Teboulle [8] provaram que a sequência $\{(x^k, z^k, y^k)\}$ converge globalmente para (x^*, z^*, y^*) , isto é, converge para uma solução primal-dual respectivamente. Além disso, eles provaram uma taxa de convergência linear do método (MPCP).

Kyono e Fukushima [16], estudaram o mesmo problema, a versão exata do seu algoritmo, chamado Método Multiplicador Preditor-Corretor Proximal Não Linear (MMPCPN), substituindo (1.2) e (1.3) pelas seguintes iterações:

$$x^{k+1} = \arg \min \{f(x) + \langle p^{k+1}, Ax \rangle + (1/\lambda_k)D_h(x, x^k)\}$$

$$z^{k+1} = \arg \min \{g(z) - \langle p^{k+1}, z \rangle + (1/\lambda_k)D_h(z, z^k)\}$$

onde D_h é uma distância de Bregman. Uma vantagem deste método em relação ao algoritmo (MPCP) é que, por uma escolha apropriada da distância de Bregman, cada subproblema torna-se um problema sem restrições. Com base em pressupostos do mesmo tipo do que Chen e Teboulle, os autores provaram a convergência global das iterações a uma solução primal-dual respectivamente.

Independente do trabalho de Kyono e Fukushima [16], Auslender e Teboulle [3], estudaram problemas de desigualdade variacional e otimização convexa com uma estrutura separável. Para o caso de Otimização, introduziram o Método de Decomposição Proximal Entrópico (MDPE), onde eles consideravam o problema (PC) com $\bar{\mathbf{C}} = \mathbb{R}_+^n$ e $\bar{\mathbf{K}} = \mathbb{R}^m$. Os autores, motivados pelo trabalho anterior do

Auslender, Teboulle e Ben-Tiba [5], substituíram a iteração (1.2) por:

$$x^{k+1} = \arg \min \{f(x) + \langle p^{k+1}, Ax \rangle + (1/\lambda_k)d(x, x^k)\}$$

onde

$$d(u, v) = \begin{cases} \sum_{i=1}^n \frac{\sigma}{2}(u_i - v_i)^2 + \mu(v_i^2 \log(\frac{v_i}{u_i}) + u_i v_i - v_i^2) & \text{se } u \in \mathbb{R}_{++}^n \\ +\infty & \text{caso contrário.} \end{cases}$$

Sob o mesmo pressuposto acima, eles provaram a convergência global das iterações a uma solução primal-dual respectivamente.

Neste trabalho, considera-se o problema geral (PC) e propõe-se a seguinte iteração inexata: Dado $(x^k, z^k, y^k) \in \mathbb{C} \times \mathbb{K} \times \mathbb{R}^m$, calcular $p^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b)$ e encontrar $(x^{k+1}, v^{k+1}) \in \mathbb{C} \times \mathbb{R}^n$ e $(z^{k+1}, \xi^{k+1}) \in \mathbb{K} \times \mathbb{R}^p$ tais que

$$\begin{aligned} v^{k+1} &\in \partial_{a_k} f^k(x^{k+1}), & v^{k+1} + \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) &= 0, \\ \xi^{k+1} &\in \partial_{b_k} g^k(z^{k+1}), & \xi^{k+1} + \lambda_k^{-1} \nabla_1 d'(z^{k+1}, z^k) &= 0. \\ y^{k+1} &= y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b) \end{aligned}$$

onde as funções $f^k : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ e $g^k : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ são definidas por $f^k(x) = f(x) + \langle p^{k+1}, Ax \rangle$ e $g^k(z) = g(z) + \langle p^{k+1}, Bz \rangle$, respectivamente, com p^{k+1} dado acima, $\partial_{a_k} f^k$ e $\partial_{b_k} g^k$ são os a_k - e b_k - subdiferencial de Fenchel de f^k e g^k respectivamente, e d, d' são distâncias proximais generalizadas, veja Seções 2.2 e 3.3.

Este método é uma extensão dos métodos MPCP e MMPCPN e inclui a classe de distâncias φ -divergências regularizadas, veja Subseção 3.3 do [2], o que para o nosso conhecimento, ainda não foi estudado neste contexto. Observe também que o EPDM é um caso particular do nosso método, quando em nosso algoritmo consideramos iterações exatas e usamos a distância log-quadrática regularizada, veja Seção 2 do [3]. Assim, os principais resultados deste trabalho são: mostrar que a convergência global do método proposto ainda é válido quando usamos distâncias proximais regularizadas, mostrar que as sequências geradas pelo método converge com uma taxa linear e mostrar resultados computacionais comparando a funcionalidade das diversas distâncias proximais conhecidas na literatura.

Esta dissertação está organizada da seguinte forma: O Capítulo 2 está dividido em 2 seções, a seguir na Seção 2.1 apresentamos algumas propriedades topológicas, definições e resultados em Análise Convexa e na Seção 2.2 apresentamos a classe de distâncias proximais e distâncias proximais induzidas, esses conceitos serão fortemente utilizados no decorrer deste trabalho, com destaque para as Seções 3.4, 3.5 e 4.2. Nos Capítulos 3 e 4, apresentam-se os resultados mais importantes da

dissertação. O Capítulo 3 está dividido em 5 seções na Seção 3.1 é introduzido o problema (PC), que desejamos resolver, para logo propor na Seção 3.2 um exemplo de modelagem do tipo (PC). Na Seção 3.3 introduzimos o Algoritmo Multiplicador Proximal com Distâncias Proximais (por sua sigla em inglês, chamado PMAPD) e na Seção 3.4 é apresentado as hipóteses e a convergência global das iterações do algoritmo para a solução ótima do (PC). Na Seção 3.5, analisamos a convergência do algoritmo no caso geral, isto é, considerando as distâncias proximais induzidas no caso geral. No Capítulo 4 provamos nosso resultado da taxa de convergência do algoritmo proposto e mostramos alguns resultados computacionais comparando a funcionalidade de algumas distâncias proximais conhecidas na literatura. Finalmente, no Capítulo 5 concluímos esta dissertação e apresentamos desdobramentos futuros prováveis acerca deste trabalho.

Os Resultados desta dissertação foram divulgados em um artigo submetido à revista *Optimization* [25].

Capítulo 2

Resultados Preliminares

Com o intuito de facilitar a leitura e compreensão desta dissertação, neste capítulo fornecemos algumas definições e resultados de convexidade, assim como apresentamos a definição de distância proximal e distância proximal induzida, analisando suas propriedades e mostrando alguns exemplos. Estas duas noções são necessárias para o desenvolvimento dos demais capítulos. Em toda a dissertação $\mathbb{R}^n := \{x = (x_1, \dots, x_n) : x_i \in \mathbb{R}, \forall i = 1, \dots, n\}$ é o espaço euclidiano dotado do produto interno canônico $\langle \cdot, \cdot \rangle$ e a norma de x dada por $\|x\| := \langle x, x \rangle^{1/2}$. Além disso, $\{x^k\}$ denota uma sequência de pontos $x^k := (x_1^k, \dots, x_n^k)$ em \mathbb{R}^n e para qualquer matriz $M \in \mathbb{R}^{m \times n}$ definimos $\|M\| := \max_{\|x\| \leq 1} \|Mx\|$. Denotamos $\mathbb{R} \cup \{\pm\infty\}$ ao conjunto $\mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$. Para levar cálculos aritméticos que envolvem $+\infty$ e $-\infty$, adotamos as seguintes regras:

$$\alpha + \infty = \infty + \alpha = \infty \quad \text{para} \quad -\infty < \alpha \leq \infty,$$

$$\alpha - \infty = -\infty + \alpha = -\infty \quad \text{para} \quad -\infty \leq \alpha < \infty,$$

$$\alpha \cdot \infty = \infty \cdot \alpha = \infty, \quad \alpha(-\infty) = (-\infty)\alpha = -\infty \quad \text{para} \quad 0 < \alpha \leq \infty,$$

$$\alpha \cdot \infty = \infty \cdot \alpha = -\infty, \quad \alpha(-\infty) = (-\infty)\alpha = \infty \quad \text{para} \quad -\infty \leq \alpha < 0,$$

$$0 \cdot \infty = \infty \cdot 0 = 0 = 0(-\infty) = (-\infty)0, \quad -(-\infty) = \infty.$$

As combinações $\infty - \infty$ e $-\infty + \infty$ são indeterminadas e são evitadas.

De acordo com estas regras, as leis familiares da aritmética:

$$\alpha_1 + \alpha_2 = \alpha_2 + \alpha_1, \quad (\alpha_1 + \alpha_2) + \alpha_3 = \alpha_1 + (\alpha_2 + \alpha_3),$$

$$\alpha_1 \alpha_2 = \alpha_2 \alpha_1, \quad (\alpha_1 \alpha_2) \alpha_3 = \alpha_1 (\alpha_2 \alpha_3),$$

$$\alpha(\alpha_1 + \alpha_2) = \alpha \alpha_1 + \alpha \alpha_2,$$

são ainda válidas, desde que nenhuma das somas $\infty - \infty$ e $-\infty + \infty$ sejam adotadas.

2.1 Propriedades Topológicas e Elementos de Análise Convexa

Nesta seção lembramos algumas propriedades topológicas em \mathbb{R}^n , definições e resultados de Análise Convexa a fim de tornar essa dissertação autocontida. Mais detalhes podem ser encontrados em [23].

Definição 2.1 *Ao longo da seção as seguintes considerações serão muito úteis.*

1. A distância euclidiana entre dois pontos x e y em \mathbb{R}^n é definido por:

$$d(x, y) = \|x - y\| = \langle x - y, x - y \rangle^{1/2}.$$

2. Denotamos por B a bola unitária euclidiana em \mathbb{R}^n

$$B := \{x \in \mathbb{R}^n : \|x\| \leq 1\} = \{x \in \mathbb{R}^n : d(x, 0) \leq 1\}.$$

3. Para qualquer $s \in \mathbb{R}^n$, a bola com raio $\epsilon > 0$ e centro s é dado por

$$\{x : d(x, s) \leq \epsilon\} = \{s + y : \|y\| \leq \epsilon\} = s + \epsilon B.$$

4. Para qualquer conjunto \mathcal{C} em \mathbb{R}^n , o conjunto de pontos x de quem a distância de \mathcal{C} não excede ao ϵ é

$$\{x : \exists y \in \mathcal{C}, d(x, y) \leq \epsilon\} = \cup\{y + \epsilon B : y \in \mathcal{C}\} = \mathcal{C} + \epsilon B.$$

5. O fecho $\bar{\mathcal{C}}$ e o interior $\text{int}(\mathcal{C})$ do conjunto \mathcal{C} podem ser expressados pelas fórmulas:

$$\bar{\mathcal{C}} = \cap\{\mathcal{C} + \epsilon B : \epsilon > 0\}$$

$$\text{int}(\mathcal{C}) = \{x : \exists \epsilon > 0, x + \epsilon B \subset \mathcal{C}\}.$$

Definição 2.2 *Dizemos que um conjunto $M \subset \mathbb{R}^n$ é chamado conjunto afim se $(1 - \lambda)x + \lambda y \in M$ para cada $x, y \in M$ e $\lambda \in \mathbb{R}$.*

Exemplo 2.3 *Os subespaços de \mathbb{R}^n são conjuntos afins que contem a origem.*

De fato, cada subespaço contem a origem, além disso, são fechados sob adição e multiplicação por escalar, assim em particular são conjuntos afins.

Reciprocamente, supor M um conjunto afim contendo a origem. Para quaisquer $x, y \in M$ temos

$$\frac{1}{2}(x + y) = \frac{1}{2}x + (1 - \frac{1}{2})y \in M,$$

dai,

$$x + y = 2\left(\frac{1}{2}(x + y)\right) \in \mathbf{M}.$$

Assim, \mathbf{M} é fechado sob a adição e portanto é um subespaço.

Definição 2.4 *Seja $S \subset \mathbb{R}^n$. O envoltório (ou fecho) afim de S , denotado por $\text{aff}(S)$, é definido como*

$$\text{aff}(S) := \{\lambda_1 x_1 + \dots + \lambda_m x_m : x_1, \dots, x_m \in S, \lambda_1 + \dots + \lambda_m = 1\}$$

Definição 2.5 *Dizemos que um conjunto $C \subset \mathbb{R}^n$ é convexo se para cada $x, y \in C$ e $\lambda \in [0, 1]$ tem-se que $(1 - \lambda)x + \lambda y \in C$.*

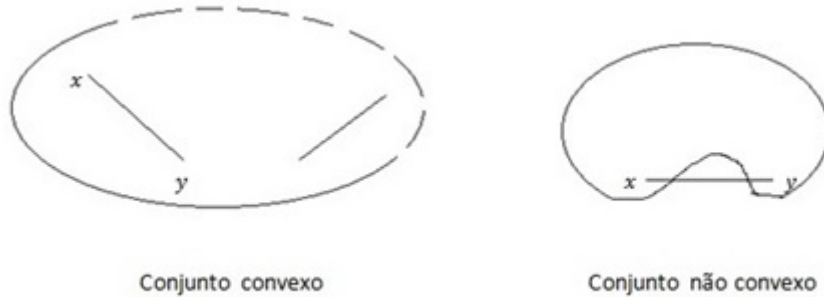


Figura 2.1: Conjunto convexo e não convexo.

Exemplo 2.6 *Todos os conjuntos afins, incluindo \emptyset e \mathbb{R}^n são conjuntos convexos.*

Definição 2.7 *Seja $C \subset \mathbb{R}^n$ um conjunto convexo. O interior relativo do conjunto C , denotado por $\text{ir}(C)$, é definido como*

$$\text{ir}(C) := \{x \in \text{aff}(C) : \exists \epsilon > 0, (x + \epsilon B) \cap \text{aff}(C) \subset C\}.$$

Observação 2.8 *Claramente, $\text{ir}(C) \subset C \subset \bar{C}$. A diferença $\bar{C} \setminus \text{ir}(C)$ é chamada a fronteira relativa de C . Naturalmente, C é chamado relativamente aberto se $\text{ir}(C) = C$. Notemos que, para qualquer conjunto convexo n -dimensional, $\text{aff}(C) = \mathbb{R}^n$, por definição, então $\text{ir}(C) = \text{int}(C)$.*

Definição 2.9 *Dada uma função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$.*

1. *O domínio de f é definido por $\text{dom } f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$;*
2. *f é uma função própria se:*

- a) $\text{dom } f \neq \emptyset$;
- b) $\forall x \in \text{dom } f : f(x) > -\infty$.
3. O epigrafo de f é definido por $\text{epi } f := \{(x, \beta) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \beta\}$.
4. O conjunto de nível de f é definido por $L_f(\alpha) := \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$.
5. f é uma função semicontínua inferior em \bar{x} se, para toda sequência $\{x^l\} \subset \mathbb{R}^n$ tal que $\lim_{l \rightarrow +\infty} x^l = \bar{x}$ tem-se que

$$f(\bar{x}) \leq \liminf_{l \rightarrow +\infty} f(x^l).$$

Dizemos que f é uma função semicontínua inferior se é semicontínua inferior em cada ponto de seu domínio.

Definição 2.10 Seja $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ uma função. Então f é convexa se, e somente se,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \quad \lambda \in (0, 1),$$

para cada $x, y \in \mathbb{R}^n$. A função f diz-se estritamente convexa quando a desigualdade acima é estrita para todos $x \neq y$ e $\lambda \in (0, 1)$.

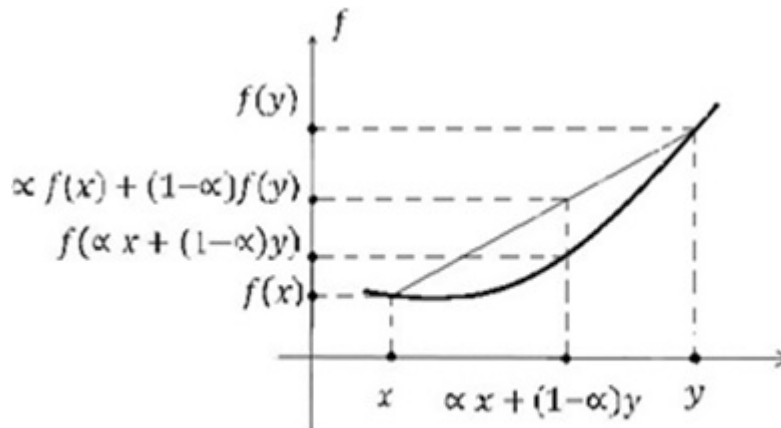


Figura 2.2: Ilustração da definição de função convexa.

Definição 2.11 Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função. Diz-se que f é uma função fortemente convexa com módulo $\beta > 0$ quando para quaisquer $x, y \in \mathbb{R}^n$ e para todo $\lambda \in (0, 1)$ temos que

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \frac{1}{2}\beta\lambda(1 - \lambda) \|x - y\|^2.$$

Definição 2.12 Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função convexa. Dizemos que $y \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in \mathbb{R}^n$ se

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

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

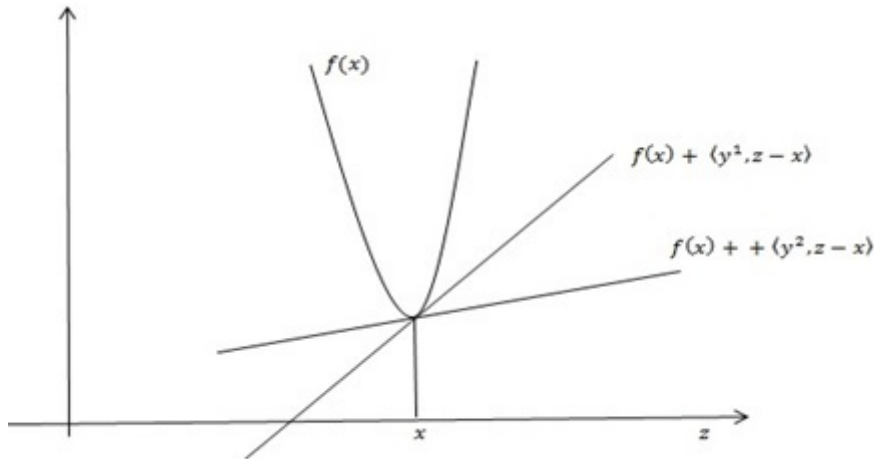


Figura 2.3: Os elementos y^1 e y^2 são subgradientes de f em x .

Proposição 2.13 ([24], Proposição 6) Seja $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ uma função própria. A função f é fortemente convexa com módulo β se, e somente se, para todo $x, y \in \mathbb{R}^n$ temos

$$f(y) \geq f(x) + \langle v, y - x \rangle + \frac{1}{2}\beta \|y - x\|^2, \text{ sempre que } v \in \partial f(x).$$

Definição 2.14 Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função convexa e $\epsilon \geq 0$. Dizemos que $y \in \mathbb{R}^n$ é um ϵ -subgradiente de f no ponto $x \in \mathbb{R}^n$ se

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

O conjunto de todos os ϵ -subgradientes de f em x , denotado por $\partial_\epsilon f(x)$, se chama o ϵ -subdiferencial de f em x .

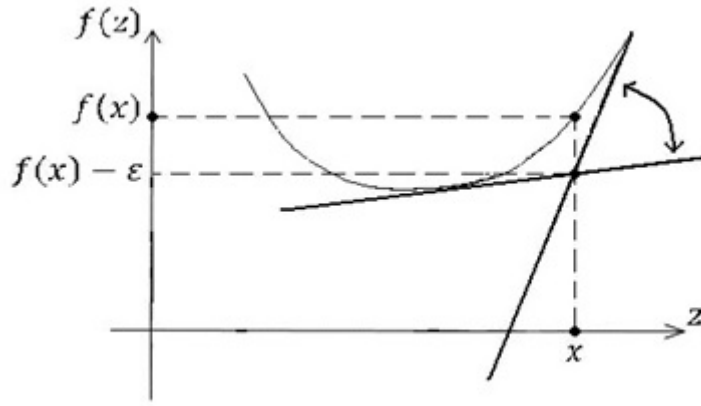


Figura 2.4: Ilustração da definição de ϵ -subdiferencial de f em x .

Observação 2.15 Usando a definição é fácil ver que

$$\partial_{\epsilon_1} f(x) \subset \partial_{\epsilon_2} f(x) \quad \forall x \in \mathbb{R}^n, \forall \epsilon_2 \geq \epsilon_1 \geq 0.$$

Em particular,

$$\partial f(x) = \partial_0 f(x) \subset \partial_{\epsilon} f(x) \quad \forall x \in \mathbb{R}^n, \forall \epsilon \geq 0.$$

Além disso,

$$0 \in \partial_{\epsilon} f(x) \Leftrightarrow f(x) \leq \inf_{z \in \mathbb{R}^n} f(z) + \epsilon,$$

o que representa uma condição de otimalidade aproximada.

Definição 2.16 Seja $X \subset \mathbb{R}^n$. A função indicadora $\delta_X : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ de X é definido por

$$\delta_X(u) = \begin{cases} 0, & \text{se } u \in X \\ +\infty, & \text{se } u \in \mathbb{R}^n \setminus X \end{cases}$$

Proposição 2.17 Seja $C \subset \mathbb{R}^n$ um conjunto convexo não vazio. Se $x_0 \in \text{int}(C)$, então $\partial \delta_C(x_0) = \{0\}$.

Demonstração. Sejam $x_0 \in \text{int}(C)$ e $x' \in \partial \delta_C(x_0)$, então

$$\delta_C(x) \geq \delta_C(x_0) + \langle x', x - x_0 \rangle, \quad \forall x \in \mathbb{R}^n$$

equivalentemente

$$\langle x', x - x_0 \rangle \leq 0, \quad \forall x \in C. \quad (2.1)$$

Como $x_0 \in \text{int}(\mathcal{C})$ e $x' \in \mathbb{R}^n$, podemos obter

$$x_0 + \epsilon x' \in \mathcal{C}, \text{ para algum } \epsilon > 0.$$

Assim, substituindo em 2.1, em particular para $x = x_0 + \epsilon x' \in \mathcal{C}$, obtemos

$$\langle x', x_0 + \epsilon x' - x_0 \rangle \leq 0, \text{ para algum } \epsilon > 0,$$

assim, $x' = 0$. É claro que, 0 satisfaz a desigualdade 2.1, isto é $0 \in \partial\delta_{\mathcal{C}}(x_0)$.

Portanto, $\partial\delta_{\mathcal{C}}(x_0) = \{0\}$. □

Definição 2.18 *Seja \mathcal{C} um conjunto convexo não vazio, o cone de recessão de \mathcal{C} é definido por*

$$0^+(\mathcal{C}) = \{z \in \mathbb{R}^n : w + tz \in \mathcal{C}, \forall w \in \mathcal{C}, t \geq 0\}.$$

Observação 2.19 *Se $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma função própria e convexa, então o epigrafo de f , é um conjunto convexo não vazio em \mathbb{R}^{n+1} e possui cone de recessão $0^+(\text{epi } f)$. Por definição,*

$$0^+(\text{epi } f) = \{(y, \nu) \in \mathbb{R}^n \times \mathbb{R} : (x, \mu) + \lambda(y, \nu) \in \text{epi } f, \text{ para todo } \lambda \geq 0\}$$

onde (x, μ) é um elemento arbitrário do $\text{epi } f$. Assim, isto pode se reescrever como

$$0^+(\text{epi } f) = \{(y, \nu) : \text{epi } f + \lambda(y, \nu) \subset \text{epi } f, \text{ para todo } \lambda \geq 0\}$$

e como $0^+(\text{epi } f)$ é um cone convexo, obtemos

$$0^+(\text{epi } f) = \{(y, \nu) : \text{epi } f + (y, \nu) \subset \text{epi } f\}.$$

Notemos que $0^+(\text{epi } f)$ é o epigrafo de alguma função, essa função será definida como segue.

Definição 2.20 *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria e convexa, sua correspondente função de recessão $f0^+ : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ é definida como*

$$\text{epi}(f0^+) = 0^+(\text{epi } f).$$

Da Observação 2.19, podemos obter o seguinte resultado.

Proposição 2.21 *([23], Teorema 8.5) Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função própria e convexa. Então a função de recessão $f0^+$ de f é uma função própria, convexa, positivamente homogênea e para cada vetor y , temos*

$$(f0^+)(y) = \sup\{f(x + y) - f(x) : x \in \text{dom } f\}.$$

Se f é semicontínua inferior, $f0^+$ é também semicontínua inferior, e para qualquer $x \in \text{dom}f$, $f0^+$ é dado pela fórmula

$$(f0^+)(y) = \sup_{\lambda > 0} \frac{f(x + \lambda y) - f(x)}{\lambda} = \lim_{\lambda \rightarrow +\infty} \frac{f(x + \lambda y) - f(x)}{\lambda}.$$

Proposição 2.22 ([23], Teorema 9.3) *Sejam f_1, \dots, f_m funções próprias e convexas em \mathbb{R}^n . Se cada f_i é semicontínua inferior e $f_1 + \dots + f_m$ não é identicamente $+\infty$, então $f_1 + \dots + f_m$ é uma função própria convexa, semicontínua inferior e*

$$(f_1 + \dots + f_m)0^+ = f_10^+ + \dots + f_m0^+.$$

Definição 2.23 *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função própria, a conjugada de f , denotada por f^* , é definida como*

$$f^*(x^*) = \sup_x [\langle x, x^* \rangle - f(x)].$$

Proposição 2.24 ([23], Corolário 13.3.4, (c)) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função própria, convexa e semicontínua inferior. Seja x^* um vetor fixo e seja $g(x) = f(x) - \langle x, x^* \rangle$. Então $x^* \in \text{int}(\text{dom}f^*)$ se, e somente se, $(g0^+)(y) > 0$ para cada $y \neq 0$.*

Proposição 2.25 ([23], Teorema 23.8) *Sejam f_1, \dots, f_m funções próprias e convexas em \mathbb{R}^n e seja $f = f_1 + \dots + f_m$. Se os conjuntos convexos $\text{ir}(\text{dom}f_i)$, $i = 1, 2, \dots, m$ têm um ponto em comum, então*

$$\partial f(x) = \partial f_1(x) + \dots + \partial f_m(x), \quad \forall x \in \mathbb{R}^n.$$

Proposição 2.26 ([23], Teorema 27.1, (d)) *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\pm\infty\}$ uma função própria, convexa e semicontínua inferior. O conjunto minimal de f é não vazio e limitado se, e somente se, $0 \in \text{int}(\text{dom}f^*)$.*

2.2 Distâncias Proximais

Nesta seção, apresentamos uma variante da definição de distância proximal e distância proximal induzida o qual foi introduzido por Auslender e Teboulle [2].

Definição 2.27 *Uma função $d : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ é chamada distância proximal em relação a um conjunto aberto convexo não vazio $\mathcal{C} \subset \mathbb{R}^n$ se para cada $y \in \mathcal{C}$ satisfaz as seguintes propriedades:*

- (i) $d(\cdot, y)$ é própria, semicontínua inferior, convexa e continuamente diferenciável em \mathcal{C} ;

- (ii) $\text{dom } d(\cdot, y) \subset \bar{\mathcal{C}}$ e $\text{dom } \partial_1 d(\cdot, y) = \mathcal{C}$, onde $\partial_1 d(\cdot, y)$ denota a função subdiferencial clássica de $d(\cdot, y)$ em relação à primeira variável;
- (iii) $d(\cdot, y)$ é coerciva em \mathbb{R}^n (isto é, $\lim_{\|u\| \rightarrow \infty} d(u, y) = +\infty$);
- (iv) $d(y, y) = 0$.

Denotamos por $\mathcal{D}(\mathcal{C})$ à família de funções que satisfazem esta definição.

Observação 2.28 A propriedade (i) é necessária para preservar a convexidade de $d(\cdot, y)$, a propriedade (ii) forçará a iteração do método para ficar em \mathcal{C} , e a propriedade (iii) será utilizado para garantir a existência das iterações proximais. Para cada $y \in \mathcal{C}$, $\nabla_1 d(\cdot, y)$ denota o gradiente da função $d(\cdot, y)$ em relação à primeira variável. Note que por definição $d(\cdot, \cdot) \geq 0$ e de (iv) o mínimo global de $d(\cdot, y)$ é obtido em y , isto mostra que $\nabla_1 d(y, y) = 0$.

Definição 2.29 Dado $d \in \mathcal{D}(\mathcal{C})$, uma função $H : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ é chamada distância proximal induzida de d , se existe $\gamma \in (0, 1]$ com H tomando valores finitos em $\mathcal{C} \times \mathcal{C}$ tal que para cada $a, b \in \mathcal{C}$, tem-se

$$(Ii) \quad H(a, a) = 0.$$

$$(Iii) \quad \langle c - b, \nabla_1 d(b, a) \rangle \leq H(c, a) - H(c, b) - \gamma H(b, a), \quad \forall c \in \mathcal{C}.$$

Escreveremos $(d, H) \in \mathcal{F}(\mathcal{C})$ a distância proximal e distância proximal induzida que satisfaz as premissas da Definição 2.29.

Denotamos também $(d, H) \in \mathcal{F}(\bar{\mathcal{C}})$ se existe H tal que:

$$(Iiii) \quad H \text{ toma valores finitos } \bar{\mathcal{C}} \times \mathcal{C} \text{ satisfazendo (Ii) e (Iii), para cada } c \in \bar{\mathcal{C}}.$$

$$(Iiv) \quad \text{Para cada } c \in \bar{\mathcal{C}}, H(c, \cdot) \text{ têm conjuntos de nível limitados em } \mathcal{C}.$$

Finalmente, escreveremos $(d, H) \in \mathcal{F}_+(\bar{\mathcal{C}})$ se

$$(Iv) \quad (d, H) \in \mathcal{F}(\bar{\mathcal{C}}).$$

$$(Ivi) \quad \forall y \in \bar{\mathcal{C}} \text{ e } \forall \{y^k\} \subset \mathcal{C} \text{ limitada com } \lim_{k \rightarrow +\infty} H(y, y^k) = 0, \text{ temos } \lim_{k \rightarrow +\infty} y^k = y.$$

$$(Ivii) \quad \forall y \in \bar{\mathcal{C}} \text{ e } \forall \{y^k\} \subset \mathcal{C} \text{ tal que } \lim_{k \rightarrow +\infty} y^k = y, \text{ obtemos } \lim_{k \rightarrow +\infty} H(y, y^k) = 0.$$

A seguinte condição adicional em H será útil para a convergência de nosso algoritmo.

Dado $(d, H) \in \mathcal{F}_+(\bar{\mathcal{C}})$, H satisfaz:

$$(Iviii) \quad \text{Para todo } c \in \bar{\mathcal{C}} \text{ e para todo } \{y^k\} \subset \mathcal{C} \text{ tal que } \lim_{k \rightarrow +\infty} y^k = y, \text{ obtemos}$$

$$\lim_{k \rightarrow +\infty} H(c, y^k) = H(c, y).$$

Exemplos

Alguns exemplos de distâncias proximais que satisfazem (Ii)-(Iviii) são os seguintes:

- (a) **Distâncias de Bregman.** Seja $S \subseteq \mathbb{R}^n$ subconjunto aberto e não vazio, e seja \bar{S} o fecho de S . Seja $h : \bar{S} \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, convexa e semicontínua inferior com $\text{dom } \nabla h = S$, estritamente convexa e contínua em $\text{dom } h$, e continuamente diferenciável em S .

Defina

$$H(x, y) := D_h(x, y) := \begin{cases} h(x) - [h(y) + \langle \nabla h(y), x - y \rangle], & \forall x \in \bar{S}, \forall y \in S \\ +\infty, & \text{caso contrário.} \end{cases}$$

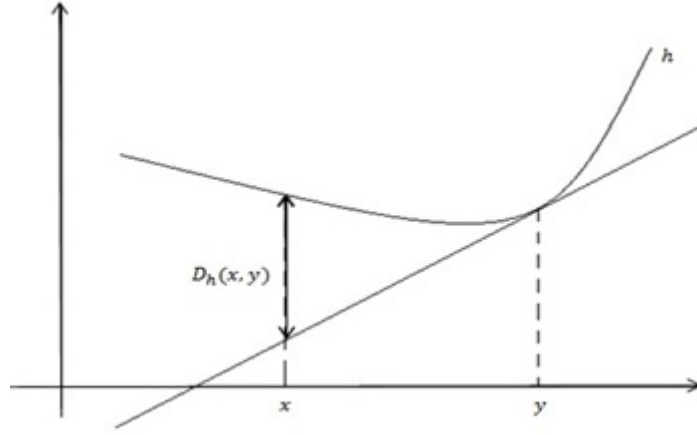


Figura 2.5: Construção de $D_h(x, y)$.

A função D_h satisfaz a identidade de três pontos (veja [9], Lema 3.1),

$$H(c, a) = H(c, b) + H(b, a) + \langle c - b, \nabla_1 H(b, a) \rangle \quad \forall a, b \in S, \forall c \in \text{dom } h.$$

A função h é chamada função de Bregman com zona S , se as seguintes condições são satisfeitas:

- (B₁) $\text{dom } h = \bar{S}$;
 (B₂) (i) $\forall x \in \bar{S}$, $D_h(x, \cdot)$ têm conjuntos de nível limitados em $\text{int}(\text{dom } h)$;

(ii) $\forall y \in S$, $D_h(\cdot, y)$ têm conjuntos de nível limitados;

- (B₃) $\forall y \in \text{dom } h$ e $\forall \{y^k\} \subset \text{int}(\text{dom } h)$ com $\lim_{k \rightarrow +\infty} y^k = y$, obtemos

$$\lim_{k \rightarrow +\infty} D_h(y, y^k) = 0;$$

- (B₄) Se $\{y^k\}$ é uma sequência limitada no $\text{int}(\text{dom } h)$ e $y \in \text{dom } h$ tal que

$$\lim_{k \rightarrow +\infty} D_h(y, y^k) = 0, \text{ então } \lim_{k \rightarrow +\infty} y^k = y.$$

Note que (B_4) é uma consequência direta das primeiras três propriedades, um fato provado por Kiwiel em ([14], Lema 2.16).

Lembre-se que a função de Bregman h com zona S é chamada essencialmente suave (veja [23]) se satisfaz a seguinte condição:

(B_5) Se $\{y^k\} \subset S$ é convergente a um ponto da fronteira de S e $u \in S$, então $\lim_{k \rightarrow +\infty} D_h(u, y^k) = +\infty$.

Da definição de função de Bregman, é claro que, para $\mathbf{C} = S$, definindo

$$d(x, y) := H(x, y) := D_h(x, y),$$

obtemos que $(d, H) \in \mathcal{F}_+(\bar{\mathbf{C}})$. Agora, verificaremos que, quando h é uma função de Bregman essencialmente suave, H satisfaz a condição (Iviii). De fato, seja $c \in \bar{\mathbf{C}}$ e uma sequência $\{y^k\} \subset \mathbf{C}$ tal que $\lim_{k \rightarrow +\infty} y^k = y$. Supondo que y é um ponto da fronteira de \mathbf{C} , pela condição (Ivii), obtemos

$$\lim_{k \rightarrow +\infty} D_h(y, y^k) = 0. \quad (2.2)$$

Por outro lado, definindo $u = y^k + (1 - \theta)(y - y^k)$ com $0 < \theta < 1$, claramente, $u \in \mathbf{C}$. Como ∇h é monótono,

$$\langle \nabla h(u), u - y^k \rangle \geq \langle \nabla h(y^k), u - y^k \rangle.$$

Daí, como $u - y^k = (1 - \theta)(y - y^k)$, obtemos

$$\langle \nabla h(u), y - y^k \rangle \geq \langle \nabla h(y^k), y - y^k \rangle.$$

Assim,

$$\begin{aligned} D_h(y, u) + D_h(u, y^k) &= h(y) - h(u) - \langle \nabla h(u), y - u \rangle \\ &\quad + h(u) - h(y^k) - \langle \nabla h(y^k), u - y^k \rangle \\ &= h(y) - h(y^k) - \theta \langle \nabla h(u), y - y^k \rangle - (1 - \theta) \langle \nabla h(y^k), y - y^k \rangle \\ &\leq h(y) - h(y^k) - \theta \langle \nabla h(y^k), y - y^k \rangle - (1 - \theta) \langle \nabla h(y^k), y - y^k \rangle \\ &= h(y) - h(y^k) - \langle \nabla h(y^k), y - y^k \rangle \\ &= D_h(y, y^k). \end{aligned}$$

Isto é,

$$D_h(y, u) + D_h(u, y^k) \leq D_h(y, y^k).$$

Tomando $k \rightarrow +\infty$, e usando o fato que h é essencialmente suave e igualdade (2.2), obtemos $+\infty \leq 0$, o que é uma contradição e portanto $y \in \mathbf{C}$.

Finalmente, pela continuidade de h e ∇h , a condição (Iviii) é satisfeita.

(b) **Distância proximal φ -divergência.** Seja $\varphi : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, convexa e semicontínua inferior tal que $\text{dom } \varphi \subset \mathbb{R}_+$ e $\text{dom } \partial\varphi = \mathbb{R}_{++}$. Supor adicionalmente que φ é C^2 , estritamente convexa e não negativa em \mathbb{R}_{++} com $\varphi(1) = \varphi'(1) = 0$. Denotemos por Φ à classe destas funções e por Φ_1 á subclasse de estas funções satisfazendo

$$\varphi''(1) \left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1) \log t \quad \forall t > 0.$$

A outra subclasse de Φ , de interesse, é denotado por Φ_2 , onde a desigualdade acima é substituída pela

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

Exemplos de funções em Φ_1, Φ_2 são (veja, por exemplo, [5, 29])

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

Correspondente à classe Φ_1 , definimos a distancia proximal φ -divergência por

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

Para qualquer $\varphi \in \Phi$, como o $\text{argmin}\{\varphi(t) : t \in \mathbb{R}\} = \{1\}$, φ é coerciva e segue-se $d_\varphi \in \mathcal{D}(\mathbf{C})$, com $\mathbf{C} = \mathbb{R}_{++}^n$.

Seja $\varphi \in \Phi_1$ e definimos a distância proximal φ -divergência regularizada (classe Φ_1) por

$$\bar{d}_\varphi(x, y) := \sum_{i=1}^n y_i \varphi \left(\frac{x_i}{y_i} \right) + \frac{\sigma}{2} \|x - y\|^2.$$

onde σ é um constante positiva. Note também que $\bar{d}_\varphi \in \mathcal{D}(\mathbf{C})$.

Tomando $\varphi(t) = t - \log t - 1$, obtemos

$$\bar{d}_\varphi(x, y) := \sum_{i=1}^n x_i - y_i - y_i \log \frac{x_i}{y_i} + \frac{\sigma}{2} \|x - y\|^2,$$

e tomando

$$\bar{H}(x, y) = \sum_{i=1}^n x_i \log \frac{x_i}{y_i} + y_i - x_i + \frac{\sigma}{2} \|x - y\|^2$$

obtemos que

$$\langle c - b, \nabla_1 \bar{d}_\varphi(b, a) \rangle \leq \bar{H}(c, a) - \bar{H}(c, b) - \gamma \bar{H}(b, a),$$

veja Auslender e Teboulle [4] (Subseção 2.3, Caso a2), e portanto $(\bar{d}_\varphi, \bar{H}) \in \mathcal{F}_+(\mathbb{R}_+^n)$. Notemos também que, definindo

$$h(x) := \sum_{j=1}^n x_j \log x_j + (\sigma/2) \|x\|^2$$

e adotando a convenção $0 \log 0 = 0$, obtemos $D_h(x, y) = \bar{H}$, isto é, \bar{H} satisfaz a condição (B_5) . Portanto, como h é uma função de Bregman essencialmente suave, obtemos que \bar{H} satisfaz a condição (Iviii).

(c) Distância proximal homogênea de segundo ordem. Seja $p \in \Phi_2$, $\varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2$ com $\nu \geq \mu p''(1) > 0$, e seja a distância proximal associada definida por

$$d_\varphi(x, y) = \sum_{j=1}^n y_j^2 \varphi\left(\frac{x_j}{y_j}\right).$$

É claro que, $d_\varphi \in \mathcal{D}(\mathbb{R}_{++}^n)$ e da desigualdade importante (veja fórmula (3.21) [5], Lema 3.4) para qualquer $p \in \Phi_2$ tem-se

$$\langle c - b, \nabla_1 d_\varphi(b, a) \rangle \leq \bar{\eta} (\|c - a\|^2 - \|c - b\|^2 - \gamma \|b - a\|^2) \quad \forall a, b \in \mathbb{R}_{++}^n, \quad \forall c \in \mathbb{R}_+^n$$

onde $\bar{\eta} = 2^{-1}(\nu + \mu p''(1))$. Portanto com

$$H(x, y) = \bar{\eta} \|x - y\|^2, \quad \gamma = \frac{\nu - \mu p''(1)}{\nu + \mu p''(1)}$$

segue que $(d_\varphi, H) \in \mathcal{F}_+(\mathbb{R}_+^n)$, da continuidade da $\|\cdot\|$, claramente também a condição (Iviii) é satisfeita.

O principal resultado do método será quando $(d, H) \in \mathcal{F}_+(\bar{\mathcal{C}})$ e a condição (Iviii) é satisfeita.

Capítulo 3

Método Multiplicador Proximal

Neste capítulo introduzimos o Algoritmo Multiplicador Proximal com Distâncias Proximais (PMAPD) (o nome PMAPD é por sua sigla em inglês: 'Proximal Multiplier Algorithm with Proximal Distances') e sob os pressupostos adequados, prova-se que as iterações dadas pelo algoritmo são bem definidas e estabelecemos sua convergência global para uma solução ótima do (PC).

3.1 O Problema

Nesta dissertação, estamos interessados em resolver o seguinte problema de otimização convexo separável:

$$(PC) \quad \min\{f(x) + g(z) : Ax + Bz = b, x \in \bar{\mathcal{C}}, z \in \bar{\mathcal{K}}\},$$

onde $\mathcal{C} \subset \mathbb{R}^n$ e $\mathcal{K} \subset \mathbb{R}^p$ são conjuntos convexos abertos não vazios, $\bar{\mathcal{C}}$ e $\bar{\mathcal{K}}$ denotam o fecho (na topologia euclidiana) de \mathcal{C} e \mathcal{K} respectivamente, $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ e $g : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ são funções próprias, semicontínuas inferiores, convexas e $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$, $b \in \mathbb{R}^m$. O problema dual de Fenchel, veja Rockafellar [23], para (PC) é definida como:

$$(D) \quad \max\{-(f + \delta_{\bar{\mathcal{C}}})^*(-A^T y) - (g + \delta_{\bar{\mathcal{K}}})^*(-B^T y) - \langle y, b \rangle : y \in \mathbb{R}^m\}$$

onde $\delta_{\mathcal{X}}$ denota a função indicadora, $(f + \delta_{\bar{\mathcal{C}}})^*$, $(g + \delta_{\bar{\mathcal{K}}})^*$ são as funções conjugadas de $f + \delta_{\bar{\mathcal{C}}}$ e $g + \delta_{\bar{\mathcal{K}}}$, respectivamente, e $\langle \cdot, \cdot \rangle$ denota o produto interno canônico.

O Lagrangiano para (PC) é definido por $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \rightarrow (-\infty, +\infty]$,

$$L(x, z, y) = (f + \delta_{\bar{\mathcal{C}}})(x) + (g + \delta_{\bar{\mathcal{K}}})(z) + \langle y, Ax + Bz - b \rangle.$$

onde y é o multiplicador de Lagrange associado com a restrição $Ax + Bz = b$.

Lembremos o seguinte importante resultado.

Seja $(x^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^p$ e $y^* \in \mathbb{R}^m$. Se (x^*, z^*) é uma solução ótima do (PC) e y^* um multiplicador de Lagrange ótimo, é necessário e suficiente que (x^*, z^*, y^*) seja um ponto de sela do Lagrangiano L do (PC) (veja [23], Teorema 28.3); isto é, para todo $x \in \text{dom } f \cap \bar{C}$, $z \in \text{dom } g \cap \bar{K}$ e $y \in \mathbb{R}^m$, temos

$$L(x^*, z^*, y) \leq L(x^*, z^*, y^*) \leq L(x, z, y^*). \quad (3.1)$$

3.2 Um Modelo de Rede de Telecomunicações

Nesta seção, para motivar o problema estudado, apresentamos um modelo de fluxo de vários bens e o expressamos como um problema do tipo (PC). Para os leitores interessados em uma revisão de modelos de fluxo de vários bens recomendo a leitura [1, 20].

Os problemas de fluxo de vários bens pertencem a uma classe de modelos de otimização em grafos e refere-se à distribuição de vários bens ou produtos em um comum (finito), a rede deve atender aos seguintes requisitos:

- a. Os bens devem ser indivisíveis, isto é, ao longo da rede não deve haver uma divisão da propriedade.
- b. Os produtos devem ser originários de nós-fonte específicos e devem chegar a nós-destino determinados.

O modelo matemático dos problemas de fluxo de tráfego de vários bens pode ser expressado da seguinte forma: Considere um grafo direcionado $G(V, E)$ onde V é o conjunto de nós e E é o conjunto de arestas, supor que o grafo tem m nós e n arestas com capacidades c_u em cada aresta u . Uma aresta de nó i ao nó j é denotada por (i, j) ou alternadamente por μ_l onde $l = 1, 2, \dots, n$. Por exemplo, a Figura 3.1 representa um grafo direcionado de 4 nós e 5 arestas, onde $V = \{1, 2, 3, 4\}$,

$$E = \{(1, 2), (2, 3), (3, 4), (4, 1), (4, 2)\} = \{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\},$$

e as capacidades em cada aresta são $c_{(1,2)} = c_{\mu_1} = 4$, $c_{(2,3)} = c_{\mu_2} = 3$, $c_{(3,4)} = c_{\mu_3} = 7$, $c_{(4,1)} = c_{\mu_4} = 1$, $c_{(4,2)} = c_{\mu_5} = 5$.

Para modelar o tráfego necessário, supor que K é o número de produtos que você deseja mover através do grafo e $k = 1, 2, \dots, K$. Para cada bem k você tem um nó de origem s_k e um nó destino t_k , também chamado um k -par (s_k, t_k) . Esses dados são dados pelo usuário. Por exemplo, considerar o grafo da Figura 3.1 que temos dois bens, ou seja, $K = 2$, para o bem $k = 1$ considerar como nó de origem ao nó 1 e o nó de destino o nó 3 e para o bem $k = 2$ considere que o nó de origem é 3 e o nó de destino é 2, respectivamente, neste caso, temos que $(s_1, t_1) = (1, 3)$ e $(s_2, t_2) = (3, 2)$.

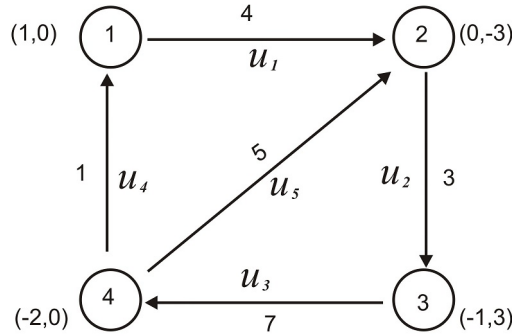


Figura 3.1: Grafo direcionado de 4 nós e 5 arestas

Para cada k , seja

$x_{ij}^k =$ o fluxo do bem k na aresta (i, j)

$d^k =$ o valor do fluxo a ser enviado s_k a t_k

$x_{ij} = \sum_{k=1}^K x_{ij}^k$ o fluxo total na aresta (i, j) .

Para cada $i \in N$, seja

$$A(i) = \{j \in V : (i, j) \in E\} \text{ e } B(i) = \{j \in V : (j, i) \in E\}.$$

Em nosso exemplo:

x_{12}^1 o fluxo do bem 1 na aresta $(1, 2)$

x_{23}^1 o fluxo do bem 1 na aresta $(2, 3)$

x_{34}^1 o fluxo do bem 1 na aresta $(3, 4)$

x_{41}^1 o fluxo do bem 1 na aresta $(4, 1)$

x_{42}^1 o fluxo do bem 1 na aresta (4, 2)

x_{12}^2 o fluxo do bem 2 na aresta (1, 2)

x_{23}^2 o fluxo do bem 2 na aresta (2, 3)

x_{34}^2 o fluxo do bem 2 na aresta (3, 4)

x_{41}^2 o fluxo do bem 2 na aresta (4, 1)

x_{42}^2 o fluxo do bem 2 na aresta (4, 2)

$d^1 = 1$ o valor do fluxo a ser enviado $s_1 = 1$ a $t_1 = 3$

$d^2 = 3$ o valor do fluxo a ser enviado $s_2 = 3$ a $t_2 = 2$.

$$x_{12} = x_{12}^1 + x_{12}^2$$

$$x_{23} = x_{23}^1 + x_{23}^2$$

$$x_{34} = x_{34}^1 + x_{34}^2$$

$$x_{41} = x_{41}^1 + x_{41}^2$$

$$x_{42} = x_{42}^1 + x_{42}^2$$

$$A(1) = \{2\}, A(2) = \{3\}, A(3) = \{4\}, A(4) = \{1, 2\}$$

$$B(1) = \{4\}, B(2) = \{1, 4\}, B(3) = \{2\}, B(4) = \{3\}$$

Note que no grafo as exigências de cada bem são representados pelo vetor (a, b) no lado do nó do grafo, neste caso a é a procura de um bem 1 e b é do bem 2, se a é positivo significa que este nó é um nó de origem e um valor negativo significa que é um nó de destino, de modo semelhante o valor de b representa a demanda do bem 2.

Um multi-fluxo é um conjunto de fluxos de arestas $\{x_{ij}^k : (i, j) \in E, k = 1, \dots, K\}$ satisfazendo:

$$\sum_{j \in A(i)} x_{ij}^k - \sum_{j \in B(i)} x_{ij}^k = \begin{cases} d^k, & \text{se } i = s^k \\ -d^k, & \text{se } i = t^k \\ 0, & \text{em outros casos} \end{cases}$$

$$x_{ij}^k \geq 0, \forall (i, j).$$

Escolhendo uma numeração de nós e arestas, podemos definir o vetor $x^k = (x_1^k, x_2^k, \dots, x_n^k)$ onde x_μ^k é o fluxo do bem k no aresta μ_l , $l = 1, \dots, n$. Com esta notação podemos definir a matriz de incidência nó-aresta a que será denotada por M .

No exemplo dado, temos

$$x^1 = (x_1^1, x_2^1, x_3^1, x_4^1, x_5^1) = (x_{\mu_1}^1, x_{\mu_2}^1, x_{\mu_3}^1, x_{\mu_4}^1, x_{\mu_5}^1)$$

$$x^2 = (x_1^2, x_2^2, x_3^2, x_4^2, x_5^2) = (x_{\mu_1}^2, x_{\mu_2}^2, x_{\mu_3}^2, x_{\mu_4}^2, x_{\mu_5}^2)$$

e a matriz de incidência nó-aresta é:

$$M = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

O problema geral de vários bens pode ser expresso como:

$$\begin{aligned} & \min f(x^1, \dots, x^K) \\ & \text{s.a :} \\ & \quad x^k \in \Omega_k, \text{ para } k = 1, 2, \dots, K \\ & \quad g(x^1, \dots, x^K) \leq 0 \end{aligned}$$

onde $f : \mathbb{R}^{Kn} \rightarrow \mathbb{R}$ é uma função (possivelmente não linear), $g : \mathbb{R}^{Kn} \rightarrow \mathbb{R}$ é uma função de restrições imposta nos fluxos e Ω_k é o conjunto de vetores de fluxo para a demanda k satisfazendo a conservação de fluxo:

$$\Omega_k = \{x^k \in \mathbb{R}^n : Mx^k = D^k, 0 \leq x^k\},$$

M é a matriz de incidência do grafo,

$$D^k = (0, \dots, d^k, \dots, -d^k, \dots, 0),$$

e d^k é a quantidade requerida de fluxo entre o k -ésimo par origem-destino.

Para o exemplo dado, o problema de fluxo de vários bens é expresso como:

$$\begin{aligned}
 & \min f(x^1, x^2) \\
 & \text{s.a :} \\
 & \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1^1 \\ x_2^1 \\ x_3^1 \\ x_4^1 \\ x_5^1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} \\
 & \begin{pmatrix} 1 & 0 & 0 & -1 & 0 \\ -1 & 1 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_4^2 \\ x_5^2 \end{pmatrix} = \begin{pmatrix} 0 \\ -3 \\ 3 \\ 0 \end{pmatrix} \\
 & g(x^1, x^2) \leq 0
 \end{aligned}$$

A maioria dos modelos não lineares de fluxo de vários bens são da forma

$$\min f(x) = \sum_{j=1}^n f_j(x_{0j}) + \sum_{k=1}^K \sum_{j=1}^n f_{kj}(x_j^k)$$

s.a :

$$\begin{aligned}
 x_{0j} &= \sum_{k=1}^K x_j^k, \quad \text{para } j = 1, 2, \dots, n \\
 Mx^k &= D^k, \quad \text{para } k = 1, 2, \dots, K \\
 0 &\leq x^k, \quad \text{para } k = 1, 2, \dots, K \\
 0 &\leq x_{0j} \leq c_j, \quad \text{para } j = 1, 2, \dots, n
 \end{aligned}$$

Entre os problemas de vários bens, em particular podemos encontrar os modelos relacionados a roteamento de dados em redes, onde o problema de roteamento de mensagens desempenha um papel significativo na otimização de redes. Este problema consiste em determinar o conjunto de caminhos em que os pacotes irão mover de tal maneira de otimizar algum custo que mede a qualidade global do serviço. Especialistas nesta área usam geralmente a função de atraso médio de tráfego de mensagens que foi introduzida por Kleinrock (1972), [15], definida por

$$f_j(x_{0j}) = \frac{x_{0j}}{c_j - x_{0j}}, \quad \text{e } f_{kj}(x_j^k) = 0$$

onde $c_j > x_{0j}$.

Assim, o problema de tráfego de envio de mensagens sob a utilização da função de Kleinrock é estabelecida por:

$$\begin{aligned} \min f(x) &= \sum_{j=1}^n \frac{x_{0j}}{c_j - x_{0j}} \\ \text{s.a :} \\ x_{0j} &= \sum_{k=1}^K x_j^k, \quad \text{para } j = 1, 2, \dots, n \\ Mx^k &= D^k, \quad \text{para } k = 1, 2, \dots, K \\ 0 &\leq x^k, \quad \text{para } k = 1, 2, \dots, K \\ 0 &\leq x_{0j} \leq c_j, \quad \text{para } j = 1, 2, \dots, n. \end{aligned}$$

O problema acima pode ser expresso como (PC), definindo $y_k = x^k$, $k = 1, \dots, K$, $z_j = x_{0j}$, para $j = 1, 2, \dots, n$ e

$$F(y, z) = \begin{cases} \sum_{j=1}^n \frac{z_j}{c_j - z_j} & \text{se } My_k = D^k, k = 1, 2, \dots, K, 0 \leq z_j < c_j, j = 1, 2, \dots, n \\ +\infty & \text{caso contrário.} \end{cases}$$

Assim, obtemos o problema

$$\min\{F(y, z) : Ay + Bz = b, y \in \bar{\mathcal{C}}, z \in \bar{\mathcal{K}}\},$$

onde $A = [I \ I \ I \ \dots \ I] \in \mathbb{R}^{n \times Kn}$, $B = -I$ com I matriz identidade $n \times n$, $b = 0 \in \mathbb{R}^n$, $\mathcal{C} := \mathbb{R}_{++}^{Kn}$ e $\mathcal{K} := \mathbb{R}_{++}^n$.

3.3 O Algoritmo (PMAPD)

No algoritmo proposto, usamos a classe de distâncias proximais $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$ $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$, satisfazendo a condição (Iviii) e dados $\mu > 0, \mu' > 0$ definem-se as seguintes funções :

$$d(x, y) := d_0(x, y) + (\mu/2) \|x - y\|^2, \quad (3.2)$$

$$H(x, y) := H_0(x, y) + (\mu/2) \|x - y\|^2, \quad (3.3)$$

$$d'(x, y) := d'_0(x, y) + (\mu'/2) \|x - y\|^2, \quad (3.4)$$

$$H'(x, y) := H'_0(x, y) + (\mu'/2) \|x - y\|^2. \quad (3.5)$$

É fácil verificar que $(d, H) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d', H') \in \mathcal{F}_+(\bar{\mathcal{K}})$ e ambos satisfazem a condição (Iviii).

O algoritmo, que será chamado Algoritmo Multiplicador Proximal com Distâncias Proximais (PMAPD) é como segue:

Algoritmo (PMAPD)

Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathbb{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathbb{K}})$ satisfazendo a condição (Iviii) e defina (d, H) , (d', H') dados por (3.2)-(3.3) e (3.4)-(3.5) respectivamente.

Sejam $\{a_k\}$, $\{b_k\}$ seqüências de escalares não negativos e $\{\lambda_k\}$ seqüência de escalares positivos.

Passo 0. Comece com um ponto de partida arbitrário $w^0 = (x^0, z^0, y^0) \in \mathbb{C} \times \mathbb{K} \times \mathbb{R}^m$. Seja $k = 0$.

Passo 1. Calcule

$$p^{k+1} = y^k + \lambda_k(Ax^k + Bz^k - b). \quad (3.6)$$

Passo 2. Encontre $(x^{k+1}, v^{k+1}) \in \mathbb{C} \times \mathbb{R}^n$ e $(z^{k+1}, \xi^{k+1}) \in \mathbb{K} \times \mathbb{R}^p$ resolvendo

$$v^{k+1} \in \partial_{a_k} f^k(x^{k+1}), \quad v^{k+1} + \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) = 0, \quad (3.7)$$

$$\xi^{k+1} \in \partial_{b_k} g^k(z^{k+1}), \quad \xi^{k+1} + \lambda_k^{-1} \nabla_1 d'(z^{k+1}, z^k) = 0. \quad (3.8)$$

onde as funções $f^k : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ e $g^k : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ são definidas por $f^k(x) = f(x) + \langle p^{k+1}, Ax \rangle$ e $g^k(z) = g(z) + \langle p^{k+1}, Bz \rangle$, respectivamente.

Passo 3. Calcule

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b). \quad (3.9)$$

Passo 4. Seja $w^{k+1} = (x^{k+1}, z^{k+1}, y^{k+1})$. Se $w^{k+1} = w^k$, parar; caso contrário fazer $k := k + 1$, logo ir ao Passo 1.

Observação 3.1 *Se o Passo 4 (critério de parada) é satisfeito, então o algoritmo (PMAPD) encontra um $(a_k + b_k)$ -solução do (PC). De fato, se $x^{k+1} = x^k$, $z^{k+1} = z^k$ e $y^{k+1} = y^k$, de (3.7) e (3.8), tem-se*

$$0 \in \partial_{a_k} f^k(x^k) \quad e \quad 0 \in \partial_{b_k} g^k(z^k) \quad (3.10)$$

e de (3.6) e (3.9), temos

$$Ax^k + Bz^k = b \quad e \quad p^{k+1} = y^k, \quad (3.11)$$

assim, de (3.10), obtemos

$$\begin{aligned} f^k(x) &\geq f^k(x^k) - a_k \quad \forall x \in \text{dom } f, \\ g^k(z) &\geq g^k(z^k) - b_k, \quad \forall z \in \text{dom } g, \end{aligned}$$

adicionando e reorganizando termos, usando (3.11), obtemos

$$f(x^k) + g(z^k) - (a_k + b_k) \leq f(x) + g(z) + \langle y^k, Ax + Bz - b \rangle,$$

então para todo $x \in \text{dom} f \cap \bar{\mathcal{C}}$ e $z \in \text{dom} g \cap \bar{K}$ tal que $Ax + Bz = b$, obtemos

$$f(x^k) + g(z^k) - (a_k + b_k) \leq f(x) + g(z).$$

Portanto,

$$f(x^k) + g(z^k) \leq \inf\{f(x) + g(z) : Ax + Bz = b, x \in \bar{\mathcal{C}}, z \in \bar{K}\} + (a_k + b_k),$$

assim, obtemos o resultado.

Observação 3.2 No Passo 4 do algoritmo, para realizar as implementações computacionais, pode considerar-se o critério de parada: $\|w^{k+1} - w^k\| \leq \text{tol}$, onde tol é um número positivo dado, muito pequeno.

Observação 3.3 Como estamos interessados na convergência assintótica das sequências geradas pelo algoritmo, então consideramos $w^{k+1} \neq w^k, \forall k$.

3.4 Análise de Convergência do PMAPD

No restante do capítulo, assumimos as seguintes hipóteses:

(\mathcal{H}_1) O problema (PC) tem uma solução ótima (x^*, z^*) e um correspondente multiplicador de Lagrange y^* .

(\mathcal{H}_2) Para cada $v \in \mathcal{C}$ e $v' \in \mathcal{K}$, existem $x \in \text{ir}(\text{dom} d(\cdot, v)) \cap \text{ir}(\text{dom} f)$ e $z \in \text{ir}(\text{dom} d'(\cdot, v')) \cap \text{ir}(\text{dom} g)$ tais que $Ax + Bz = b$.

(\mathcal{H}_3) As sequências $\{a_k\}$ e $\{b_k\}$ são não negativas e

$$\sum_{k=0}^{\infty} (a_k + b_k) < \infty.$$

(\mathcal{H}_4) Dados os parâmetros $\mu > 0$, $\mu' > 0$, definidos em (3.2) e (3.4) respectivamente, a sequência $\{\lambda_k\}$ satisfaz

$$\eta < \lambda_k < \bar{c} - \eta \tag{3.12}$$

onde $\eta \in (0, \bar{c}/2)$ com $\bar{c} := \min\{\frac{\sqrt{\gamma\mu}}{2\|A\|}, \frac{\sqrt{\gamma'\mu'}}{2\|B\|}\}$ e γ, γ' são constantes positivas relacionados com d e d' , respectivamente, na Definição 2.29, (Iii).

Observação 3.4 A hipótese (\mathcal{H}_1) implica que a função de Lagrange L possui um ponto de sela (x^*, z^*, y^*) . A hipótese (\mathcal{H}_2) garante a existência de um ótimo multiplicador dual de Lagrange y^* relacionado com a restrição $Ax + Bz = b$, e associado com a hipótese (\mathcal{H}_1) garantizam que as iterações dadas pelo PMAPD são bem definidas. As hipóteses acima não são muito fortes e estes foram utilizados em pesquisas anteriores, por exemplo veja [3, 8, 16]. A hipótese (\mathcal{H}_3) é uma condição natural quando trabalhamos com iterações inexatas. A hipótese (\mathcal{H}_4) será utilizada para assegurar a convergência do método.

O seguinte resultado mostra que o algoritmo (MPAPD) está bem definido.

Teorema 3.5 Seja $d_0 \in \mathcal{D}(\mathcal{C})$ e $d'_0 \in \mathcal{D}(K)$. Supor que as hipóteses (\mathcal{H}_1) e (\mathcal{H}_2) são satisfeitas, então para qualquer $(x^k, z^k, y^k) \in \mathcal{C} \times K \times \mathbb{R}^m$, $\lambda_k > 0$, existem únicos $(x^{k+1}, z^{k+1}) \in \mathcal{C} \times K$ satisfazendo (3.7) e (3.8).

Demonstração. Seja $F_k(x) = f(x) + \langle p^{k+1}, Ax \rangle + (1/\lambda_k)d(x, x^k)$, então da hipótese (\mathcal{H}_2) obtemos que $\text{dom } F_k = \text{dom } f \cap \text{dom } d(\cdot, x^k) \neq \emptyset$. Defina

$$S_k = \arg \min_{x \in \mathcal{C}} \{F_k(x)\} = \arg \min_{x \in \mathbb{R}^n} \{F_k(x) + \delta_{\mathcal{C}}(x)\},$$

onde $\delta_{\mathcal{C}}$ denota a função indicadora, provaremos que S_k é não vazio. De fato, pois $f(\cdot)$, $\langle A^T p^{k+1}, \cdot \rangle$ e $(1/\lambda_k)d(\cdot, x^k)$ são funções próprias, convexas e semicontínuas inferiores, então F_k é também própria, convexa e semicontínua inferior e das proposições 2.24 e 2.26, é suficiente mostrar que $((F_k + \delta_{\mathcal{C}})0^+)(y) > 0$, $\forall y \neq 0$.

Seja $y \neq 0$, então pela Proposição 2.22, obtemos

$$((F_k + \delta_{\mathcal{C}})0^+)(y) = (f0^+)(y) + (\langle A^T p^{k+1}, \cdot \rangle 0^+)(y) + (1/\lambda_k)(d(\cdot, x^k)0^+)(y) + (\delta_{\mathcal{C}}0^+)(y). \quad (3.13)$$

Mostraremos que

$$(f0^+)(y) > -\infty, \quad (\delta_{\mathcal{C}}0^+)(y) > -\infty, \quad (\langle A^T p^{k+1}, \cdot \rangle 0^+)(y) = \langle A^T p^{k+1}, y \rangle \quad \text{e}$$

$$(d(\cdot, x^k)0^+)(y) = +\infty.$$

De fato, como f e $\delta_{\mathcal{C}}$ são funções próprias e convexas, da Proposição 2.21, obtemos

$$(f0^+)(y) = \sup\{f(x+y) - f(x) : x \in \text{dom } f\}$$

$$(\delta_{\mathcal{C}}0^+)(y) = \sup\{\delta_{\mathcal{C}}(\bar{x}+y) - \delta_{\mathcal{C}}(\bar{x}) : \bar{x} \in \text{dom } \delta_{\mathcal{C}}\}.$$

Notemos que:

Se $x+y \in \text{dom } f$, então $(f0^+)(y) \geq f(x+y) - f(x) > -\infty$, (f é própria).

Se $x+y \notin \text{dom } f$, então $f(x+y) = +\infty$ e assim $(f0^+)(y) = +\infty > -\infty$.

Analogamente,

Se $\bar{x} + y \in \text{dom } \delta_{\bar{c}}$, então $(\delta_{\bar{c}}0^+)(y) = 0 > -\infty$, ($\delta_{\bar{c}}$ é própria).

Se $\bar{x} + y \notin \text{dom } \delta_{\bar{c}}$, então $(\delta_{\bar{c}}0^+)(y) = +\infty$ e assim $(\delta_{\bar{c}}0^+)(y) = +\infty > -\infty$.

Além disso, como $\langle A^T p^{k+1}, \cdot \rangle$ é própria, convexa e semicontínua inferior da Proposição 2.21, tem-se

$$\begin{aligned} (\langle A^T p^{k+1}, \cdot \rangle 0^+)(y) &= \lim_{\lambda \rightarrow +\infty} \frac{\langle A^T p^{k+1}, x + \lambda y \rangle - \langle A^T p^{k+1}, x \rangle}{\lambda} \\ &= \lim_{\lambda \rightarrow +\infty} \frac{\lambda \langle A^T p^{k+1}, y \rangle}{\lambda} = \langle A^T p^{k+1}, y \rangle. \end{aligned}$$

Assim,

$$(f0^+)(y) > -\infty, \quad (\delta_{\bar{c}}0^+)(y) > -\infty, \quad (\langle A^T p^{k+1}, \cdot \rangle 0^+)(y) = \langle A^T p^{k+1}, y \rangle.$$

agora provaremos que $(d(\cdot, x^k)0^+)(y) = +\infty$.

Da Proposição 2.21, para $\bar{x} \in \text{dom } d_0(\cdot, x^k)$, obtemos

$$(d(\cdot, x^k)0^+)(y) = \lim_{\lambda \rightarrow +\infty} \left(\frac{d_0(\bar{x} + \lambda y, x^k) - d_0(\bar{x}, x^k)}{\lambda} + \frac{\rho\lambda}{2} \|y\|^2 \right) - \mu \langle \bar{x} - x^k, y \rangle,$$

e como $d_0(\cdot, x^k)$ é uma função convexa, obtemos que $(d(\cdot, x^k)0^+)(y) = +\infty$ e portanto de (3.13)

$$((F_k + \delta_{\bar{c}})0^+)(y) = +\infty > 0, \quad \forall y \neq 0.$$

Assim, existe $\bar{x} := x^{k+1} \in \mathbb{R}^n$ tal que

$$0 \in \partial(f(\cdot) + \langle A^T p^{k+1}, \cdot \rangle + (1/\lambda_k)d(\cdot, x^k) + \delta_{\bar{c}}(\cdot))(x^{k+1}).$$

Também, a estrita convexidade de $F_k + \delta_{\bar{c}}$ (devido a que $d(\cdot, x^k)$ é estritamente convexa) garante a unicidade de x^{k+1} .

Agora, pela hipótese (\mathcal{H}_2) e da Definição de $d(\cdot, x^k)$, existe $x \in \text{ir}(\text{dom } f) \cap \text{ir}(\text{dom } d(\cdot, x^k)) \cap \mathbf{C}$, então da Proposição 2.25, obtemos que

$$0 \in \partial f(x^{k+1}) + A^T p^{k+1} + (1/\lambda_k)\partial_1 d(x^{k+1}, x^k) + \partial \delta_{\bar{c}}(x^{k+1}).$$

Por Definição 2.27, (ii) temos que $\text{dom } \partial_1 d(\cdot, x^k) = \mathbf{C}$, então $x^{k+1} \in \mathbf{C}$ e assim $\partial \delta_{\bar{c}}(x^{k+1}) = \{0\}$ (veja Proposição 2.17). Isto é, existe $v^{k+1} \in \partial f^k(x^{k+1})$ satisfazendo (3.7) com $a_k = 0$.

Como $\partial f^k(x^{k+1}) \subseteq \partial_\epsilon f^k(x^{k+1})$ para todo $\epsilon \geq 0$, então existem $x^{k+1} \in \mathbf{C}$ e $v^{k+1} \in \mathbb{R}^n$ satisfazendo (3.7).

Para a existência $z^{k+1} \in \mathbf{K}$ e $\xi^{k+1} \in \mathbb{R}^p$ satisfazendo (3.8) o análise é análogo.

Portanto, a demonstração está concluída. \square

O seguinte análise de convergência, segue um argumento semelhante ao análise de Kyono e Fukushima, [16].

Usaremos o seguinte resultado muito conhecido na literatura.

Lema 3.6 [8](, Lema 3.1) *Seja $F : \mathbb{R}^m \rightarrow (-\infty, +\infty]$ uma função própria, convexa e semicontínua inferior, $\tau > 0$ e defina:*

$$u^{k+1} = \arg \min_{u \in \mathbb{R}^m} \{F(u) + (1/(2\tau))\|u - u^k\|^2\}.$$

Então para qualquer $k \geq 0$,

$$2\tau[F(u^{k+1}) - F(u)] \leq \|u^k - u\|^2 - \|u^{k+1} - u\|^2 - \|u^{k+1} - u^k\|^2, \forall u \in \mathbb{R}^m.$$

Também usaremos a seguinte notação útil

$$\Delta_k(x, z) = H(x, x^k) - H(x, x^{k+1}) - \gamma H(x^{k+1}, x^k) + H'(z, z^k) - H'(z, z^{k+1}) - \gamma' H'(z^{k+1}, z^k) \quad (3.14)$$

O seguinte resultado dá algumas estimativas para as sequências geradas pelo algoritmo (PMAPD).

Lema 3.7 *Sejam $(d_0, H_0) \in \mathcal{F}(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{\mathcal{K}})$. Supor que as hipóteses (\mathcal{H}_1) e (\mathcal{H}_2) são satisfeitas. Então, para todo $x \in \text{dom} f \cap \bar{\mathcal{C}}$, $z \in \text{dom} g \cap \bar{\mathcal{K}}$, $y \in \mathbb{R}^m$ as seguintes desigualdades são satisfeitas:*

- (i) $\lambda_k \{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \Delta_k(x, z) + \lambda_k(a_k + b_k)$
- (ii) $2\lambda_k \{L(x^k, z^k, y) - L(x^k, z^k, p^{k+1})\} \leq \|y^k - y\|^2 - \|p^{k+1} - y\|^2 - \|p^{k+1} - y^k\|^2$
- (iii) $2\lambda_k \{L(x^{k+1}, z^{k+1}, y) - L(x^{k+1}, z^{k+1}, y^{k+1})\} \leq \|y^k - y\|^2 - \|y^{k+1} - y\|^2 - \|y^{k+1} - y^k\|^2.$

Demonstração. (i) Seja $x \in \text{dom} f \cap \bar{\mathcal{C}}$, $z \in \text{dom} g \cap \bar{\mathcal{K}}$, $y \in \mathbb{R}^m$ da definição do Lagrangiano L , é evidente que

$$L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1}) = (f^k(x^{k+1}) - f^k(x)) + (g^k(z^{k+1}) - g^k(z)). \quad (3.15)$$

Como $v^{k+1} \in \partial_{a_k} f^k(x^{k+1})$, então $f^k(x) \geq f^k(x^{k+1}) + \langle v^{k+1}, x - x^{k+1} \rangle - a_k$, assim, de (3.7) e da Definição 2.29, (Iii), obtemos

$$\begin{aligned} \lambda_k(f^k(x^{k+1}) - f^k(x)) &\leq \langle \nabla_1 d(x^{k+1}, x^k), x - x^{k+1} \rangle + \lambda_k a_k \\ &\leq H(x, x^k) - H(x, x^{k+1}) - \gamma H(x^{k+1}, x^k) + \lambda_k a_k. \end{aligned} \quad (3.16)$$

Analogamente, como $\xi^{k+1} \in \partial_{b_k} g^k(z^{k+1})$, obtemos $g^k(z) \geq g^k(z^{k+1}) + \langle \xi^{k+1}, z - z^{k+1} \rangle - b_k$, assim, de (3.8) e da Definição 2.29, (Iii), obtemos

$$\begin{aligned} \lambda_k(g^k(z^{k+1}) - g^k(z)) &\leq \langle \nabla_1 d'(z^{k+1}, z^k), z - z^{k+1} \rangle + \lambda_k b_k \\ &\leq H'(z, z^k) - H'(z, z^{k+1}) - \gamma' H'(z^{k+1}, z^k) + \lambda_k b_k, \end{aligned} \quad (3.17)$$

adicionando (3.16) e (3.17), de (3.15) e a notação em (3.14), obtemos

$$\lambda_k \{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \Delta_k(x, z) + \lambda_k(a_k + b_k).$$

Para provar (ii) e (iii) note que os passos 1 e 3 do (PMAPD) podem ser escritos de forma equivalente como:

$$\begin{aligned} p^{k+1} &= \arg \min_{y \in \mathbb{R}^m} \{-L(x^k, z^k, y) + (1/2\lambda_k)\|y - y^k\|^2\} \\ y^{k+1} &= \arg \min_{y \in \mathbb{R}^m} \{-L(x^{k+1}, z^{k+1}, y) + (1/2\lambda_k)\|y - y^k\|^2\}. \end{aligned}$$

Então, usando o Lema 3.6, dois vezes com $\tau = \lambda_k$, $F(y) = -L(x^k, z^k, y)$ e $F(y) = -L(x^{k+1}, z^{k+1}, y)$ respectivamente, o primeiro e segundo equações acima, respectivamente, deu as desigualdades desejadas (ii) e (iii). \square

Observação 3.8 *No resto da seção, denotamos $w^k = (x^k, z^k, y^k)$, $w^* = (x^*, z^*, y^*)$, $w = (x, z, y)$, $s = (l, q, r) \in \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ e defina a função $\hat{H} : X \times X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ por*

$$\hat{H}(w, s) = \hat{H}((x, z, y), (l, q, r)) := H(x, l) + H'(z, q) + (1/2)\|y - r\|^2, \quad (3.18)$$

onde $X = \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$, H e H' são definidas em (3.3) e (3.5) respectivamente.

Proposição 3.9 *Sejam $(d_0, H_0) \in \mathcal{F}(\bar{C})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{K})$. Supor que as hipóteses (\mathcal{H}_1) , (\mathcal{H}_2) e (\mathcal{H}_4) são satisfeitas. Seja $\{w^k\}$ e $\{p^k\}$ uma sequência gerada pelo (PMAPD); seja (x^*, z^*) uma solução ótima do (PC), e seja y^* seu correspondente multiplicador de Lagrange, então temos, para cada k ,*

$$\begin{aligned} \hat{H}(w^*, w^{k+1}) &\leq \hat{H}(w^*, w^k) - [\gamma H_0(x^{k+1}, x^k) + (1/2)(\gamma\mu - 4\lambda_k^2\|A\|^2)\|x^{k+1} - x^k\|^2] \\ &\quad - [\gamma' H'_0(z^{k+1}, z^k) + (1/2)(\gamma'\mu' - 4\lambda_k^2\|B\|^2)\|z^{k+1} - z^k\|^2] \\ &\quad - (1/2)(\|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2) + \lambda_k(a_k + b_k), \end{aligned} \quad (3.19)$$

onde $w^* = (x^*, z^*, y^*)$. Em particular, temos, para todo k ,

$$\hat{H}(w^*, w^{k+1}) \leq \hat{H}(w^*, w^k) + \lambda_k(a_k + b_k). \quad (3.20)$$

Demonstração. Usando $(x, z) = (x^*, z^*)$ no Lema 3.7, (i), obtemos

$$\begin{aligned} \lambda_k(L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})) &\leq H(x^*, x^k) - H(x^*, x^{k+1}) + H'(z^*, z^k) \\ &\quad - H'(z^*, z^{k+1}) - (\gamma H(x^{k+1}, x^k) + \\ &\quad + \gamma' H'(z^{k+1}, z^k)) + \lambda_k(a_k + b_k). \end{aligned} \quad (3.21)$$

Desde que (x^*, z^*, y^*) é um ponto de sela do Lagrangiano L , obtemos de (3.1)

$$\lambda_k(L(x^*, z^*, p^{k+1}) - L(x^{k+1}, z^{k+1}, y^*)) \leq 0. \quad (3.22)$$

adicionando (3.21) e (3.22) e reorganizando termos, obtemos

$$\begin{aligned} H(x^*, x^{k+1}) + H'(z^*, z^{k+1}) &\leq H(x^*, x^k) + H'(z^*, z^k) - [\gamma H(x^{k+1}, x^k) + \gamma' H'(z^{k+1}, z^k)] \\ &\quad - \lambda_k \langle p^{k+1} - y^*, Ax^{k+1} + Bz^{k+1} - b \rangle + \lambda_k(a_k + b_k). \end{aligned} \quad (3.23)$$

Agora, usando o Lema 3.7, (ii), em $y := y^{k+1}$ e $y := y^*$ para (iii), obtemos, respectivamente,

$$2\lambda_k(L(x^k, z^k, y^{k+1}) - L(x^k, z^k, p^{k+1})) \leq \|y^k - y^{k+1}\|^2 - \|p^{k+1} - y^{k+1}\|^2 - \|p^{k+1} - y^k\|^2 \quad (3.24)$$

$$2\lambda_k(L(x^{k+1}, z^{k+1}, y^*) - L(x^{k+1}, z^{k+1}, y^{k+1})) \leq \|y^k - y^*\|^2 - \|y^{k+1} - y^*\|^2 - \|y^{k+1} - y^k\|^2 \quad (3.25)$$

adicionando (3.24) e (3.25) e reorganizando termos, obtemos

$$\begin{aligned} (1/2) \|y^{k+1} - y^*\|^2 &\leq (1/2) \|y^k - y^*\|^2 - (1/2) [\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2] \\ &\quad - \lambda_k (\langle y^* - y^{k+1}, Ax^{k+1} + Bz^{k+1} - b \rangle \\ &\quad + \langle y^{k+1} - p^{k+1}, Ax^k + Bz^k - b \rangle). \end{aligned} \quad (3.26)$$

adicionando (3.23) e (3.26) e de (3.18), obtemos

$$\begin{aligned} \hat{H}(w^*, w^{k+1}) &\leq \hat{H}(w^*, w^k) - (\gamma H(x^{k+1}, x^k) + \gamma' H'(z^{k+1}, z^k)) \\ &\quad - (1/2) (\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) + \lambda_k(a_k + b_k) \\ &\quad + \lambda_k \langle y^{k+1} - p^{k+1}, A(x^{k+1} - x^k) + B(z^{k+1} - z^k) \rangle \\ &= \hat{H}(w^*, w^k) - (\gamma H(x^{k+1}, x^k) + \gamma' H'(z^{k+1}, z^k)) \\ &\quad - (1/2) (\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) + \\ &\quad + \lambda_k^2 \|A(x^{k+1} - x^k) + B(z^{k+1} - z^k)\|^2 + \lambda_k(a_k + b_k), \end{aligned} \quad (3.27)$$

onde a última igualdade segue de (3.6) e (3.9).

Usando, em (3.27), a desigualdade

$$\|A(x^{k+1} - x^k) + B(z^{k+1} - z^k)\|^2 \leq 2 \left(\|A\|^2 \|x^{k+1} - x^k\|^2 + \|B\|^2 \|z^{k+1} - z^k\|^2 \right),$$

e também,

$$\begin{aligned} H(x^{k+1}, x^k) &= H_0(x^{k+1}, x^k) + (\mu/2) \|x^{k+1} - x^k\|^2, \\ H'(z^{k+1}, z^k) &= H'_0(z^{k+1}, z^k) + (\mu'/2) \|z^{k+1} - z^k\|^2, \end{aligned}$$

obtemos

$$\begin{aligned}\hat{H}(w^*, w^{k+1}) &\leq \hat{H}(w^*, w^k) - [\gamma H_0(x^{k+1}, x^k) + (1/2)(\gamma\mu - 4\lambda_k^2 \|A\|^2)\|x^{k+1} - x^k\|^2] \\ &\quad - [\gamma' H'_0(z^{k+1}, z^k) + (1/2)(\gamma'\mu' - 4\lambda_k^2 \|B\|^2)\|z^{k+1} - z^k\|^2] \\ &\quad - (1/2)(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) + \lambda_k(a_k + b_k).\end{aligned}$$

A desigualdade (3.20) segue imediatamente de (3.19) e a Hipótese (\mathcal{H}_4) . \square

Lema 3.10 ([21], Lema 2) *Sejam $\{\alpha_k\}$ e $\{\beta_k\}$ sequências não negativas tais que $\alpha_{k+1} \leq \alpha_k + \beta_k$, para todo k e $\sum_{k=0}^{\infty} \beta_k < \infty$. Então $\{\alpha_k\}$ é convergente.*

Proposição 3.11 *Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ satisfazendo a condição (Iviii). Supor que as hipóteses (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3) e (\mathcal{H}_4) são satisfeitas. Seja $\{w^k\}$ uma sequência gerada pelo algoritmo (PMPD), então $\{w^k\}$ é limitada e cada ponto de acumulação de $\{w^k\}$ é um ponto de sela de L .*

Demonstração. Primeiro, note que a hipótese (\mathcal{H}_3) juntamente com a hipótese (\mathcal{H}_4) garantem que

$$\sum_{k=0}^{\infty} \lambda_k(a_k + b_k) < \infty. \quad (3.28)$$

Então, de (3.20), obtemos

$$w^k \in L_{\hat{H}}(w^*, \bar{\alpha}) := \{w : \hat{H}(w^*, w) \leq \bar{\alpha}\}, \text{ para todo } k,$$

onde $\bar{\alpha} = \hat{H}(w^*, w^0) + \sum_{k=0}^{\infty} \lambda_k(a_k + b_k)$.

Isto implica que $\{w^k\}$ é limitada graças à condição (Iiv) da Definição 2.29 para H e H' .

Agora, seja (x^*, z^*) uma solução ótima do (PC), seja y^* seu correspondente multiplicador de Lagrange, e seja $w^* = (x^*, z^*, y^*)$.

Então, de (3.20), (3.28) e do Lema 3.10, obtemos que $\{\hat{H}(w^*, w^k)\}$ é convergente, isto é, existe $\beta \geq 0$ tal que

$$\lim_{k \rightarrow +\infty} \hat{H}(w^*, w^k) = \beta, \quad (3.29)$$

tomando limite na desigualdade do Lema (3.19), da Hipótese (\mathcal{H}_3) e reordenando termos, obtemos

$$\begin{aligned}\limsup_{k \rightarrow +\infty} (\gamma H_0(x^{k+1}, x^k) + \frac{1}{2}(\gamma\mu - 4\lambda_k^2 \|A\|^2)\|x^{k+1} - x^k\|^2 + \frac{1}{2}(\gamma'\mu' - 4\lambda_k^2 \|B\|^2)\|z^{k+1} - z^k\|^2 \\ + \gamma' H'_0(z^{k+1}, z^k) + \frac{1}{2}(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2)) \leq 0.\end{aligned} \quad (3.30)$$

Segue-se então a partir de (3.30) e a Hipótese (\mathcal{H}_4) que

$$\begin{aligned} H_0(x^{k+1}, x^k) &\rightarrow 0, & \|x^{k+1} - x^k\| &\rightarrow 0 \\ H'_0(z^{k+1}, z^k) &\rightarrow 0, & \|z^{k+1} - z^k\| &\rightarrow 0 \\ \|p^{k+1} - y^{k+1}\| &\rightarrow 0, & \|p^{k+1} - y^k\| &\rightarrow 0. \end{aligned} \quad (3.31)$$

Agora, seja $\bar{w} = (\bar{x}, \bar{z}, \bar{y})$ um ponto de acumulação de $\{w^k\}$, e seja $\{w^k\}_{k \in \mathcal{K}}$ uma subsequência convergindo a \bar{w} . Usando o Lema 3.7, (i), para todo $x \in \text{dom} f \cap \bar{\mathcal{C}}$ e $z \in \text{dom} g \cap \bar{\mathcal{K}}$, obtemos

$$\lambda_k \{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \Delta_k(x, z) + \lambda_k(a_k + b_k). \quad (3.32)$$

Tomando o limite sobre as subsequências apropriadas de ambos lados de (3.32), usando (3.31) e das Hipóteses (\mathcal{H}_3) , (\mathcal{H}_4) , e a condição (Iviii) para H e H' , obtemos

$$L(\bar{x}, \bar{z}, \bar{y}) - L(x, z, \bar{y}) \leq 0, \quad (3.33)$$

desde que λ_k está limitada inferiormente por $\eta > 0$.

Similarmente, tomando limite, sobre as subsequências apropriadas, de ambos lados da desigualdade no Lema 3.7, (ii), obtemos que, para todo $y \in \mathbb{R}^m$,

$$L(\bar{x}, \bar{z}, y) - L(\bar{x}, \bar{z}, \bar{y}) \leq 0. \quad (3.34)$$

De (3.33) e (3.34), obtemos que, para todo $x \in \text{dom} f \cap \bar{\mathcal{C}}$, $z \in \text{dom} g \cap \bar{\mathcal{K}}$, $y \in \mathbb{R}^m$,

$$L(\bar{x}, \bar{z}, y) \leq L(\bar{x}, \bar{z}, \bar{y}) \leq L(x, z, \bar{y}).$$

Portanto, $\bar{w} = (\bar{x}, \bar{z}, \bar{y})$ é um ponto de sela do Lagrangiano L . □

Das proposições acima, obtemos o seguinte teorema de convergência para o PMAPD.

Teorema 3.12 *Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ satisfazendo a condição (Iviii). Supor que as Hipóteses (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3) , (\mathcal{H}_4) são satisfeitas e seja $\{(x^k, z^k, y^k)\}$ uma sequência gerada pelo algoritmo (PMAPD), então a sequência $\{(x^k, z^k, y^k)\}$ converge globalmente para (x^*, z^*, y^*) , com (x^*, z^*) ótimo para (PC) e y^* seu correspondente multiplicador de Lagrange.*

Demonstração. Da Proposição 3.11, $\{w^k\} = \{(x^k, z^k, y^k)\}$ é limitada e qualquer ponto de acumulação de $\bar{w} = (\bar{x}, \bar{z}, \bar{y})$ é um ponto de sela de L .

Seja $\{w^k\}_{k \in \mathcal{K}}$ uma subsequência convergindo para \bar{w} , então (\bar{x}, \bar{z}) é uma solução ótima para o (PC) e \bar{y} seu correspondente multiplicador de Lagrange (veja Rockafellar [23], Teorema 28.3). Então, é suficiente provar que $\{w^k\}$ converge para \bar{w} .

De fato, desde que

$$\lim_{k \rightarrow +\infty, k \in \mathcal{K}} x^k = \bar{x}, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} z^k = \bar{z}, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} y^k = \bar{y},$$

da Definição 2.29, (Ivii), obtemos que

$$\lim_{k \rightarrow +\infty, k \in \mathcal{K}} H(\bar{x}, x^k) = 0, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} H'(\bar{z}, z^k) = 0, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} (1/2) \|y^k - \bar{y}\|^2 = 0$$

tomando $w = \bar{w}$ e $s = w^k$ em (3.18), obtemos

$$\lim_{k \rightarrow +\infty, k \in \mathcal{K}} \hat{H}(\bar{w}, w^k) = 0, \quad (3.35)$$

Por outro lado, de (3.20) substituindo w^* por \bar{w} , obtemos

$$0 \leq \hat{H}(\bar{w}, w^{k+1}) \leq \hat{H}(\bar{w}, w^k) + \lambda_k(a_k + b_k), \quad \text{para todo } k,$$

da Hipótese (\mathcal{H}_3) e o Lema 3.10 obtemos que $\hat{H}(\bar{w}, w^k)$ converge e como existe uma subsequência convergindo para zero, (veja (3.35)), então toda a sequência converge, isto é,

$$\lim_{k \rightarrow +\infty} \hat{H}(\bar{w}, w^k) = 0.$$

Finalmente, de (3.18) e Definição 2.29, (Ivi), para H e H' , obtemos que

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

Isto é, obtemos o resultado desejado. \square

3.5 Convergência para o Caso Geral

Na seção anterior analisamos a convergência do algoritmo (PMAPD) quando $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$, em particular quando a distância proximal satisfaz a condição (Iii) da Definição 2.29, isto é,

$$\langle c - b, \nabla_1 d(b, a) \rangle \leq H(c, a) - H(c, b) - \gamma H(b, a), \quad \text{com } \gamma \in (0, 1],$$

$$\langle c - b, \nabla_1 d'(b, a) \rangle \leq H'(c, a) - H'(c, b) - \gamma' H'(b, a), \quad \text{com } \gamma' \in (0, 1].$$

As condições acima são satisfeitas pelas classes de distâncias de Bregman, distâncias homogêneas de segundo ordem e para algumas distâncias φ -divergências. Porém, é bem conhecido que todas as distâncias φ -divergências satisfazem a desigualdade de

acima quando $\gamma = 0$ and $\gamma' = 0$, isto é,

$$(Iii)' : \langle c - b, \nabla_1 d(b, a) \rangle \leq H(c, a) - H(c, b),$$

veja Teboulle [29], Lema 4.1, (ii).

Nesta seção analisamos a convergência do algoritmo (PMAPD) substituindo na Definição 2.29 a condição (Iii) por (Iii)'. Também substituímos a Hipótese (\mathcal{H}_4) pela seguinte:

$$(\mathcal{H}_4)' : \begin{aligned} & \kappa < \lambda_k < \bar{\lambda} \\ & \sum_{k=0}^{\infty} \|x^{k+1} - x^k\|^2 < +\infty \\ & \sum_{k=0}^{\infty} \|z^{k+1} - z^k\|^2 < +\infty \end{aligned}$$

onde κ e $\bar{\lambda}$ são constantes positivas.

Observação 3.13 *Podemos interpretar $(\mathcal{H}_4)'$ em termos computacionais, como uma certa medida, se as somas não explodem garantimos a convergência do método utilizando qualquer distância φ -divergência.*

Note que, sob as novas hipóteses, os resultados do Teorema 3.5, Lema 3.6, Lema 3.7 (com $\Delta_k(x, z) = H(x, x^k) - H(x, x^{k+1}) + H'(z, z^k) - H'(z, z^{k+1})$) continuam sendo válidas porém a Proposição 3.9 torna-se

Proposição 3.14 *Sejam $(d_0, H_0) \in \mathcal{F}(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{\mathcal{K}})$. Supor que as Hipóteses (\mathcal{H}_1) , (\mathcal{H}_2) são satisfeitas. Seja $\{w^k\}$ e $\{p^k\}$ sequências geradas pelo algoritmo (PMAPD); seja (x^*, z^*) uma solução ótima do (PC), e seja y^* seu correspondente multiplicador de Lagrange, então temos que, para cada k ,*

$$\begin{aligned} \hat{H}(w^*, w^{k+1}) & \leq \hat{H}(w^*, w^k) - (1/2)(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) + \\ & + 2\lambda_k^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) + \lambda_k(a_k + b_k), \end{aligned}$$

onde $w^* = (x^*, z^*, y^*)$. Em particular, temos, para todo k ,

$$\hat{H}(w^*, w^{k+1}) \leq \hat{H}(w^*, w^k) + 2\lambda_k^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) + \lambda_k(a_k + b_k). \quad (3.36)$$

Demonstração. Procedendo analogamente como a prova da Proposição 3.9, usando o Lema 3.7 com $\Delta_k(x, z) = H(x, x^k) - H(x, x^{k+1}) + H'(z, z^k) - H'(z, z^{k+1})$, obtemos o resultado. \square

Além disso, obtemos o seguinte resultado

Proposição 3.15 *Sejam $(d_0, H_0) \in \mathcal{F}(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{\mathcal{K}})$ (substituindo (Iii) por (Iii)') satisfazendo a condição (Iviii). Supor que as Hipóteses (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3)*

e $(\mathcal{H}_4)'$ são satisfeitas. Seja $\{w^k\}$ a sequência gerada pelo algoritmo (PMAPD), então $\{w^k\}$ é limitada e cada ponto de acumulação de $\{w^k\}$ é um ponto de sela do Lagrangiano L .

Demonstração. A prova é análoga à prova da Proposição 3.11. Primeiro, note que a Hipótese (\mathcal{H}_3) junto com a Hipótese $(\mathcal{H}_4)'$ implica que

$$\sum_{k=0}^{\infty} 2\lambda_k^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) + \lambda_k(a_k + b_k) < +\infty. \quad (3.37)$$

Assim, de (3.36), obtemos

$$w^k \in L_{\hat{H}}(w^*, \bar{\alpha}) := \{w : \hat{H}(w^*, w) \leq \bar{\alpha}\}, \text{ para todo } k,$$

onde

$$\bar{\alpha} = \hat{H}(w^*, w^0) + \sum_{k=0}^{+\infty} 2\lambda_k^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) + \lambda_k(a_k + b_k).$$

Isto implica que $\{w^k\}$ é limitada graças à condição (liv) da Definição 2.29 para H e H' .

Agora, seja (x^*, z^*) uma solução ótima do (PC), seja y^* seu correspondente multiplicador de Lagrange, e seja $w^* = (x^*, z^*, y^*)$,

assim, de (3.36), (3.37) e do Lema 3.10, obtemos que $\{\hat{H}(w^*, w^k)\}$ é convergente, isto é, existe $\beta \geq 0$ tal que

$$\lim_{k \rightarrow +\infty} \hat{H}(w^*, w^k) = \beta. \quad (3.38)$$

Tomando limsup na primeira desigualdade da Proposição 3.14, obtemos

$$\limsup_{k \rightarrow +\infty} \frac{1}{2}(\|p^{k+1} - y^{k+1}\|^2 + \|p^{k+1} - y^k\|^2) - 2\lambda_k^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) - \lambda_k(a_k + b_k) \leq 0, \quad (3.39)$$

dai, das Hipóteses (\mathcal{H}_3) e $(\mathcal{H}_4)'$, obtemos

$$\begin{aligned} \|x^{k+1} - x^k\| &\rightarrow 0, & \|z^{k+1} - z^k\| &\rightarrow 0 \\ \|p^{k+1} - y^{k+1}\| &\rightarrow 0, & \|p^{k+1} - y^k\| &\rightarrow 0. \end{aligned} \quad (3.40)$$

Por outro lado, seja $\bar{w} = (\bar{x}, \bar{z}, \bar{y})$ um ponto de acumulação de $\{w^k\}$, e seja $\{w^k\}_{k \in \mathcal{K}}$ uma subsequência convergindo para \bar{w} . Usando o Lema 3.7, (i), para todo $x \in \text{dom } f \cap \bar{\mathcal{C}}$ e $z \in \text{dom } g \cap \bar{\mathcal{K}}$, obtemos

$$\lambda_k\{L(x^{k+1}, z^{k+1}, p^{k+1}) - L(x, z, p^{k+1})\} \leq \Delta_k(x, z) + \lambda_k(a_k + b_k). \quad (3.41)$$

Tomando limite sobre as apropriadas subsequências em ambas desigualdades de (3.41), usando (3.40) e as Hipóteses (\mathcal{H}_3) , $(\mathcal{H}_4)'$, e condição (Iviii) para H e H' , obtemos

$$L(\bar{x}, \bar{z}, \bar{y}) - L(x, z, \bar{y}) \leq 0, \quad (3.42)$$

já que λ_k está limitada inferiormente por $\kappa > 0$.

Similarmente, tomando limite, sobre as subsequências apropriadas, em ambos lados da desigualdade do Lema 3.7, (ii), obtemos que, para todo $y \in \mathbb{R}^m$,

$$L(\bar{x}, \bar{z}, y) - L(\bar{x}, \bar{z}, \bar{y}) \leq 0. \quad (3.43)$$

De (3.42) e (3.43), obtemos que, para todo $x \in \text{dom } f \cap \bar{\mathcal{C}}$, $z \in \text{dom } g \cap \bar{\mathcal{K}}$, $y \in \mathbb{R}^m$,

$$L(\bar{x}, \bar{z}, y) \leq L(\bar{x}, \bar{z}, \bar{y}) \leq L(x, z, \bar{y}).$$

Portanto, \bar{w} é um ponto de sela de L . □

Finalmente, obtemos o seguinte principal resultado:

Teorema 3.16 *Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ (substituindo (Iii) por (Iii)') satisfazendo a condição (Iviii). Supor que as Hipóteses (\mathcal{H}_1) , (\mathcal{H}_2) , (\mathcal{H}_3) e $(\mathcal{H}_4)'$ são satisfeitas e seja $\{(x^k, z^k, y^k)\}$ uma sequência gerada pelo algoritmo (PMAPD), então a sequência $\{(x^k, z^k, y^k)\}$ converge globalmente a (x^*, z^*, y^*) , com (x^*, z^*) solução ótima para (PC) e y^* seu correspondente multiplicador de Lagrange.*

Demonstração. Procedendo analogamente à prova do Teorema 3.12, da Proposição 3.15, $\{w^k\} = \{(x^k, z^k, y^k)\}$ é limitada e qualquer ponto de acumulação de $\bar{w} = (\bar{x}, \bar{z}, \bar{y})$ é um ponto de sela de L .

Seja $\{w^k\}_{k \in \mathcal{K}}$ uma subsequência convergindo para \bar{w} , então (\bar{x}, \bar{z}) é uma solução ótima para o (PC) e \bar{y} seu correspondente multiplicador de Lagrange (veja Rockafellar [23], Teorema 28.3). Então, é suficiente provar que $\{w^k\}$ converge para \bar{w} .

De fato, desde que

$$\lim_{k \rightarrow +\infty, k \in \mathcal{K}} x^k = \bar{x}, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} z^k = \bar{z}, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} y^k = \bar{y},$$

da Definição 2.29, (Ivii), obtemos que

$$\lim_{k \rightarrow +\infty, k \in \mathcal{K}} H(\bar{x}, x^k) = 0, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} H'(\bar{z}, z^k) = 0, \quad \lim_{k \rightarrow +\infty, k \in \mathcal{K}} (1/2) \|y^k - \bar{y}\|^2 = 0$$

tomando $w = \bar{w}$ e $s = w^k$ em (3.18), obtemos

$$\lim_{k \rightarrow +\infty, k \in \mathcal{K}} \hat{H}(\bar{w}, w^k) = 0, \quad (3.44)$$

Por outro lado, de (3.36) substituindo w^* por \bar{w} , para todo k , obtemos

$$0 \leq \hat{H}(\bar{w}, w^{k+1}) \leq \hat{H}(\bar{w}, w^k) + 2\lambda_k^2(\|A\|^2\|x^{k+1} - x^k\|^2 + \|B\|^2\|z^{k+1} - z^k\|^2) + \lambda_k(a_k + b_k)$$

das Hipóteses (\mathcal{H}_3) , $(\mathcal{H}_4)'$ e o Lema 3.10 obtemos que $\hat{H}(\bar{w}, w^{k+1})$ converge e como existe uma subsequência convergindo para zero, (veja (3.44)), então toda a sequência converge, isto é,

$$\lim_{k \rightarrow +\infty} \hat{H}(\bar{w}, w^k) = 0.$$

Finalmente, de (3.18) e Definição 2.29, (Ivi), para H e H' , obtemos que

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

Isto é, obtemos o resultado desejado. □

Capítulo 4

Taxa de Convergência e Implementações do PMAPD

Neste capítulo estamos interessados em analisar a taxa de convergência do Método Multiplicador Proximal com Distâncias Proximais (PMAPD) e apresentar resultados computacionais comparando a funcionalidade das diversas distâncias proximais, o método (PMAPD) foi aplicado para resolver um programa convexo com estrutura separável da forma:

$$(PC) \quad \min\{f(x) + g(z) : Ax + Bz = b, x \in \bar{\mathbf{C}}, z \in \bar{\mathbf{K}}\},$$

onde $\mathbf{C} \subset \mathbb{R}^n$ e $\mathbf{K} \subset \mathbb{R}^p$ são conjuntos convexos abertos não vazios, $\bar{\mathbf{C}}$ e $\bar{\mathbf{K}}$ denotam o fecho (na topologia euclidiana) de \mathbf{C} e \mathbf{K} respectivamente, $f : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ e $g : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ são funções próprias, semicontínuas inferiores, convexas e $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times p}$, $b \in \mathbb{R}^m$.

Observação 4.1 *O Lagrangiano $L(x, z, y)$ para o (PC) é definido por $L(x, z, y) : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \rightarrow (-\infty, +\infty]$,*

$$L(x, z, y) = (f + \delta_{\bar{\mathbf{C}}})(x) + (g + \delta_{\bar{\mathbf{K}}})(z) + \langle y, Ax + Bz - b \rangle.$$

onde δ_x denota a função indicadora, $\langle \cdot, \cdot \rangle$ denota o produto interno e y é o multiplicador de Lagrange associado com a restrição $Ax + Bz = b$.

Observação 4.2 *O Lagrangiano $L(x, z, y)$ é uma função convexa-côncava e semicontínua inferior. Portanto o operador subdiferencial S definida sobre $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$ dado por*

$$S(x, z, y) = \partial_{x,z}L(x, z, y) \times \partial_y(-L(x, z, y)) \quad (4.1)$$

é monótono maximal, veja Rockafellar [22].

Lembremos que o par (x^, z^*) é solução ótima do (PC) e y^* é um ótimo multiplicador*

de Lagrange se e somente se

$$L(x^*, z^*, y) \leq L(x^*, z^*, y^*) \leq L(x, z, y^*), \quad \forall (x, z) \in \mathbb{R}^{n \times p}, \forall y \in \mathbb{R}^m.$$

Isto é se, e somente se, $0 \in S(x^*, z^*, y^*)$, (onde 0 denota o zero em $\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m$).

4.1 O Algoritmo (PMAPD) sob um Adequado Critério de Aproximação

Nesta seção lembraremos o Algoritmo Multiplicador Proximal com Distâncias Proximais (PMAPD) apresentado na Seção 3.3, onde se usa a classe de distâncias proximais $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$, satisfazendo a condição (Iviii) (veja Definição 2.29). Dados $\mu > 0, \mu' > 0$ definimos as seguintes funções :

$$d(x, y) := d_0(x, y) + (\mu/2) \|x - y\|^2, \quad (4.2)$$

$$H(x, y) := H_0(x, y) + (\mu/2) \|x - y\|^2, \quad (4.3)$$

$$d'(x, y) := d'_0(x, y) + (\mu'/2) \|x - y\|^2, \quad (4.4)$$

$$H'(x, y) := H'_0(x, y) + (\mu'/2) \|x - y\|^2. \quad (4.5)$$

É fácil verificar que $(d, H) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d', H') \in \mathcal{F}_+(\bar{\mathcal{K}})$ e ambos satisfazem a condição (Iviii) da Definição 2.29.

Algoritmo (PMAPD)

Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{\mathcal{K}})$ satisfazendo a condição (Iviii) e defina (d, H) , (d', H') dados por (3.2)-(3.3) e (3.4)-(3.5) respectivamente.

Sejam $\{a_k\}$, $\{b_k\}$ seqüências de escalares não negativos e $\{\lambda_k\}$ seqüência de escalares positivos.

Passo 0. Comece com um ponto de partida arbitrário $w^0 = (x^0, z^0, y^0) \in \mathbb{C} \times \mathbb{K} \times \mathbb{R}^m$. Seja $k = 0$.

Passo 1. Calcule

$$p^{k+1} = y^k + \lambda_k (Ax^k + Bz^k - b). \quad (4.6)$$

Passo 2. Encontre $(x^{k+1}, v^{k+1}) \in \mathbb{C} \times \mathbb{R}^n$ e $(z^{k+1}, \xi^{k+1}) \in \mathbb{K} \times \mathbb{R}^p$ resolvendo

$$v^{k+1} \in \partial_{a_k} f^k(x^{k+1}), \quad v^{k+1} + \lambda_k^{-1} \nabla_1 d(x^{k+1}, x^k) = 0, \quad (4.7)$$

$$\xi^{k+1} \in \partial_{b_k} g^k(z^{k+1}), \quad \xi^{k+1} + \lambda_k^{-1} \nabla_1 d'(z^{k+1}, z^k) = 0. \quad (4.8)$$

onde as funções $f^k : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ e $g^k : \mathbb{R}^p \rightarrow (-\infty, +\infty]$ são definidas por $f^k(x) = f(x) + \langle p^{k+1}, Ax \rangle$ e $g^k(z) = g(z) + \langle p^{k+1}, Bz \rangle$, respectivamente.

Passo 3. Calcule

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b). \quad (4.9)$$

Passo 4. Seja $w^{k+1} = (x^{k+1}, z^{k+1}, y^{k+1})$. Se $w^{k+1} = w^k$, parar; caso contrário fazer $k := k + 1$, logo ir ao Passo 1.

No Capítulo 3, se provaram os seguintes principais resultados:

Teorema 4.3 *Seja $d_0 \in \mathcal{D}(\mathcal{C})$ e $d'_0 \in \mathcal{D}(\bar{K})$. Supor que o (PC) tem uma solução ótima (x^*, z^*) e um correspondente multiplicador de Lagrange y^* e existem $x \in \text{ir}(\text{dom } d(\cdot, v)) \cap \text{ir}(\text{dom } f)$ e $z \in \text{ir}(\text{dom } d'(\cdot, v')) \cap \text{ir}(\text{dom } g)$ tais que $Ax + Bz = b$, então para qualquer $(x^k, z^k, y^k) \in \mathcal{C} \times K \times \mathbb{R}^m$, $\lambda_k > 0$, existem únicos $(x^{k+1}, z^{k+1}) \in \mathcal{C} \times K$ satisfazendo (4.7) e (4.8).*

Teorema 4.4 *Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{K})$ satisfazendo a condição (Iviii). Supor que as Hipóteses do Teorema 4.3 são satisfeitas e $\{a_k\}, \{b_k\}$ são sequências não negativas tais que $\sum_{k=0}^{\infty} (a_k + b_k) < \infty$. Seja $\{(x^k, z^k, y^k)\}$ uma sequência gerada pelo algoritmo (PMAPD). Se $\{\lambda_k\}$ satisfaz*

$$\eta < \lambda_k < \bar{c} - \eta$$

para algum $\eta \in (0, \bar{c}/2)$ com $\bar{c} := \min\{\frac{\sqrt{\gamma\mu}}{2\|A\|}, \frac{\sqrt{\gamma'\mu'}}{2\|B\|}\}$ onde γ, γ' são constantes positivas relacionados com d e d' , respectivamente, na Definição 2.29, (Iii) e μ, μ' são constantes positivas definidas em (3.2) e (3.4) respectivamente, então a sequência $\{(x^k, z^k, y^k)\}$ converge globalmente para (x^*, z^*, y^*) , com (x^*, z^*) ótimo para (PC) e y^* seu correspondente multiplicador de Lagrange.

Observação 4.5 *Neste Capítulo, a sequência $\{x^k, z^k, y^k\}$ é gerada pelo Algoritmo (PMAPD) sob o critério de aproximação*

$$\|x^{k+1} - \bar{x}^{k+1}\| \leq \eta_k \|x^{k+1} - x^k\|, \quad \|z^{k+1} - \bar{z}^{k+1}\| \leq \eta_k \|z^{k+1} - z^k\|,$$

onde $\bar{x}^{k+1}, \bar{z}^{k+1}$ denotam aos pontos obtidos do Passo 2 do Algoritmo (PMAPD), quando $a_k = b_k = 0 \forall k$, e $\eta_k \geq 0$ com $\sum_{k=0}^{+\infty} \eta_k < +\infty$.

Esta Observação será considerada na seguinte seção como uma das hipóteses para obter a taxa de convergência do PMAPD.

Observação 4.6 *Dados $\bar{x}^{k+1}, \bar{z}^{k+1}$ definidos na Observação 4.5. Definimos*

$$\bar{y}^{k+1} := y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b).$$

4.2 Taxa de Convergência do PMAPD

Nesta seção mostraremos que a sequência $\{(x^k, z^k, y^k)\}$ gerada pelo Algoritmo (PMAPD), satisfazendo o critério dado na Observação 4.5, converge a um ponto de sela do Lagrangiano $L(x, z, y)$ com uma taxa de convergência linear. Para provar este resultado, precisamos de algumas hipóteses adicionais ao problema dado, considerando-se a seguinte observação.

Observação 4.7 *Seja T um operador monótono maximal em \mathbb{R}^n . Segue-se de Rockafellar [24] que o operador T^{-1} é Lipschitz contínuo na origem com módulo $a \geq 0$, se existe uma única solução \bar{u} tal que $0 \in T(\bar{u})$ e para algum $\tau > 0$, temos $\|u - \bar{u}\| \leq a \|v\|$, sempre que $v \in T(u)$ e $\|u\| \leq \tau$.*

Hipóteses

Nesta seção assumimos as seguintes hipóteses:

(\mathcal{T}_1) S^{-1} é Lipschitz contínuo na origem com módulo $a \geq 0$.

(\mathcal{T}_2) A sequência $\{x^k, z^k, y^k\}$ é gerada pelo Algoritmo (PMAPD) sob o critério de aproximação

$$\|x^{k+1} - \bar{x}^{k+1}\| \leq \eta_k \|x^{k+1} - x^k\|, \quad \|z^{k+1} - \bar{z}^{k+1}\| \leq \eta_k \|z^{k+1} - z^k\| \quad (4.10)$$

onde $\bar{x}^{k+1}, \bar{z}^{k+1}$ denotam aos pontos obtidos do Passo 2 do Algoritmo (PMAPD), quando $a_k = b_k = 0 \forall k$, e $\eta_k \geq 0$ com $\sum_{k=0}^{+\infty} \eta_k < +\infty$.

As Hipóteses (\mathcal{T}_1) e (\mathcal{T}_2) foram sugeridas por Rockafellar (veja [22], pg. 100) para obter a taxa de convergência do método proximal dos multiplicadores, e aqui também serão usadas para obter a taxa de convergência para o Algoritmo (PMAPD).

Observação 4.8 *Também assumiremos hipóteses às distâncias $(d_0, H_0) \in \mathcal{F}_+(\bar{C})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{K})$. Ao longo da Seção consideremos as distâncias proximais d, d' definidas por*

$$\begin{aligned} d(x, y) &:= d_0(x, y) + (1/2)\|x - y\|^2, \\ d'(x, y) &:= d'_0(x, y) + (1/2)\|x - y\|^2. \end{aligned}$$

Isto é, da definição dada no Algoritmo (PMAPD), em particular, tomamos $\mu = \mu' = 1$.

(\mathcal{T}_3) Para (x^*, z^*) solução ótima do (PC), para algum $k_0 \in \mathbb{N}$, os pontos \bar{x}^{k+1} , \bar{z}^{k+1} satisfazem

$$d_0(x^*, x^k) \leq d_0(\bar{x}^{k+1}, x^k), \quad d'_0(z^*, z^k) \leq d'_0(\bar{z}^{k+1}, z^k), \quad \forall k \geq k_0.$$

(\mathcal{T}_4) As funções $\nabla_1 d_0(\cdot, u)$, $\nabla_1 d'_0(\cdot, v)$ são localmente Lipschitz com módulo α_1, α_2 em \mathbf{C} e \mathbf{K} respectivamente, isto é, para qualquer $x_0 \in \mathbf{C}$, $z_0 \in \mathbf{K}$ existem $\alpha_1 > 0, \alpha_2 > 0$ e $r_1 > 0, r_2 > 0$ tais que

$$\|\nabla_1 d_0(x, u) - \nabla_1 d_0(x', u)\| \leq \alpha_1 \|x - x'\|, \quad \forall x, x' \in B_{r_1}(x_0), \quad \forall u \in \bar{\mathbf{C}}$$

$$\|\nabla_1 d'_0(z, v) - \nabla_1 d'_0(z', v)\| \leq \alpha_2 \|z - z'\|, \quad \forall z, z' \in B'_{r_2}(z_0), \quad \forall v \in \bar{\mathbf{K}}$$

onde $B_{r_1}(x_0) := \{x \in \mathbf{C} : \|x - x_0\| < r_1\}$ e $B'_{r_2}(z_0) := \{z \in \mathbf{K} : \|z - z_0\| < r_2\}$.

Observação 4.9 *Das hipóteses podemos dar os seguintes comentários:*

1. Note que para o problema (PC) temos

$$S^{-1}(v_1, v_2, v_3) = \arg \min_{x, z} \max_y \{L(x, z, y) - \langle x, v_1 \rangle - \langle z, v_2 \rangle + \langle y, v_3 \rangle\}$$

e portanto a Hipótese (\mathcal{T}_1), considerando-se as Observações 4.2 e 4.7, pode ser interpretado em termos do problema dado como segue: existe um único ponto de sela w^* tal que para algum $\tau > 0$, temos $\|w - w^*\| \leq a \|v\|$, sempre que $\|v\| \leq \tau$ e $w = (x, z, y) \in S^{-1}(v_1, v_2, v_3)$.

2. Notemos que para simplificar a notação em nossa análise, usamos o mesmo η_k para o critério de aproximação em (4.10). Isto se deve porque se fossem sequências diferentes $\eta_k^i \geq 0$, $\sum_{k=0}^{+\infty} \eta_k^i < +\infty$, $i = 1, 2$, poderíamos definir $\eta_k = \max\{\eta_k^1, \eta_k^2\}$ em (4.10).

3. Note que o uso do mesmo k_0 na Hipótese (\mathcal{T}_3) se deve porque se fossem diferentes k_0^i , $i = 1, 2$, podemos definir $k_0 = \max\{k_0^1, k_0^2\}$.

4. Note que das distâncias proximais dadas na Seção 2.2 do Capítulo 2, a distância proximal φ -divergência e a distância proximal homogênea de segundo ordem, satisfazem a Hipótese (\mathcal{T}_4) usando seus Definições, isto é,

$$d_\varphi(x, y) := \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right) \text{ com } \varphi \in C^2(\mathbb{R}_{++}),$$

$$d_\varphi(x, y) := \sum_{i=1}^n y_i^2 \varphi\left(\frac{x_i}{y_i}\right) \text{ com } \varphi(t) = \mu p(t) + \frac{\nu}{2}(t-1)^2, \quad \nu \geq \mu p''(1) > 0, p \in C^2(\mathbb{R}_{++}),$$

respectivamente. Notamos que em ambos casos, $\nabla_1 d_\varphi \in C^1(\mathbb{R}_{++}^n)$.

Portanto, $\nabla_1 d_\varphi(\cdot, y)$ é localmente Lipschitz.

Além disso, adicionando uma condição às funções h Bregman, isto é, considerando funções h Bregman tais que $h \in C^2(\mathcal{C})$, e da definição de Distância de Bregman, $D_h(x, y) := h(x) - [h(y) + \langle \nabla h(y), x - y \rangle]$ temos que $\nabla_1 D_h(\cdot, y) = \nabla h(\cdot) - \nabla h(y)$ é continuamente diferenciável em \mathcal{C} e portanto, $\nabla_1 D_h(\cdot, y)$ é localmente Lipschitz.

O seguinte análise, da taxa de convergência do método (PMAPD), segue um argumento semelhante ao análise de Chen e Teboulle, [8].

Lema 4.10 *Seja $F : \mathbb{R}^n \rightarrow \mathbb{R} \cup (-\infty, +\infty]$ uma função própria, convexa e semi-contínua inferior e $d_0 \in \mathcal{D}(\mathcal{C})$, defina-se*

$$\bar{v}^{k+1} := \arg \min \{F(v) + (1/\lambda_k)d(v, v^k)\},$$

onde $d(x, y) := d_0(x, y) + (1/2)\|x - y\|^2$. Então para qualquer $k \geq 0$, tem-se que

$$2\lambda_k[F(\bar{v}^{k+1}) - F(v)] \leq \|v^k - v\|^2 - \|\bar{v}^{k+1} - v\|^2 - \|\bar{v}^{k+1} - v^k\|^2 - 2d_0(\bar{v}^{k+1}, v^k) + 2d_0(v, v^k).$$

Demonstração. Seja $\psi_k(v) := F(v) + (1/\lambda_k)d_0(v, v^k) + (1/2\lambda_k)\|v - v^k\|^2$. Por definição de \bar{v}^{k+1} , temos que $0 \in \partial\psi_k(\bar{v}^{k+1})$. Como $F(\cdot)$ e $d_0(\cdot, v^k)$ são funções convexas, então ψ_k é fortemente convexa com módulo $(1/\lambda_k)$, e portanto segue-se

$$2\lambda_k[\psi_k(v) - \psi_k(\bar{v}^{k+1})] \geq \|\bar{v}^{k+1} - v\|^2, \quad \forall v, \quad (4.11)$$

(veja Proposição 2.13), da definição de ψ_k , obtemos

$$2\lambda_k \left[F(\bar{v}^{k+1}) + \frac{1}{\lambda_k}d_0(\bar{v}^{k+1}, v^k) - F(v) - \frac{1}{\lambda_k}d_0(v, v^k) \right] \leq \|v^k - v\|^2 - \|\bar{v}^{k+1} - v\|^2 - \|\bar{v}^{k+1} - v^k\|^2.$$

Assim,

$$2\lambda_k[F(\bar{v}^{k+1}) - F(v)] \leq \|v^k - v\|^2 - \|\bar{v}^{k+1} - v\|^2 - \|\bar{v}^{k+1} - v^k\|^2 - 2d_0(\bar{v}^{k+1}, v^k) + 2d_0(v, v^k).$$

□

No seguinte resultado estabelecemos 2 estimativas fundamentais relacionando iterações exatas e inexatas de uma solução ótima.

Lema 4.11 *Sejam $(d_0, H_0) \in \mathcal{F}(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}(\bar{\mathcal{K}})$ e seja $\{(x^k, z^k, y^k)\}$ uma sequência gerada pelo PMAPD. Supor que o (PC) tem uma solução ótima (x^*, z^*) e um correspondente multiplicador de Lagrange y^* e existem $x \in \text{ir}(\text{dom } d(\cdot, v)) \cap \text{ir}(\text{dom } f)$ e $z \in \text{ir}(\text{dom } d'(\cdot, v')) \cap \text{ir}(\text{dom } g)$ tais que $Ax + Bz = b$. Então para qualquer $k \geq 0$*

$$\begin{aligned}
(i) \quad \|\bar{x}^{k+1} - x^*\|^2 + \|\bar{z}^{k+1} - z^*\|^2 &\leq \|x^k - x^*\|^2 + \|z^k - z^*\|^2 - \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2\} \\
&\quad - 2\lambda_k \langle p^{k+1} - y^*, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b \rangle \\
&\quad - 2[d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)] + 2[d_0(x^*, x^k) + d'_0(z^*, z^k)]; \\
(ii) \quad \|\bar{y}^{k+1} - y^*\|^2 &\leq \|y^k - y^*\|^2 - \{\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\
&\quad - 2\lambda_k \{\langle y^* - \bar{y}^{k+1}, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b \rangle + \langle \bar{y}^{k+1} - p^{k+1}, Ax^k + Bz^k - b \rangle\}.
\end{aligned}$$

Demonstração. (i) Do Passo 2 do Algoritmo (PMAPD) as sequências $\{\bar{x}^k\}$, $\{\bar{z}^k\}$ são obtidas quando $a_k = b_k = 0$, assim,

$$\bar{x}^{k+1} = \arg \min \{f^k(x) + \delta_{\bar{c}}(x) + (1/\lambda_k)d(x, x^k)\},$$

$$\bar{z}^{k+1} = \arg \min \{g^k(z) + \delta_{\bar{k}}(z) + (1/\lambda_k)d'(z, z^k)\}.$$

da Observação 4.8 e usando o Lema 4.10 dois vezes com $F(\cdot) := (f^k + \delta_{\bar{c}})(\cdot)$; $F(\cdot) := (g^k + \delta_{\bar{k}})(\cdot)$ nos pontos $x = x^*$, $z = z^*$, respectivamente, obtemos

$$2\lambda_k[f^k(\bar{x}^{k+1}) - f^k(x^*)] \leq \|x^k - x^*\|^2 - \|\bar{x}^{k+1} - x^*\|^2 - \|\bar{x}^{k+1} - x^k\|^2 - 2d_0(\bar{x}^{k+1}, x^k) + 2d_0(x^*, x^k),$$

$$2\lambda_k[g^k(\bar{z}^{k+1}) - g^k(z^*)] \leq \|z^k - z^*\|^2 - \|\bar{z}^{k+1} - z^*\|^2 - \|\bar{z}^{k+1} - z^k\|^2 - 2d'_0(\bar{z}^{k+1}, z^k) + 2d'_0(z^*, z^k),$$

adicionando ambas desigualdades e da definição do Lagrangiano L , obtemos

$$\begin{aligned}
2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, p^{k+1}) - L(x^*, z^*, p^{k+1})] &\leq \|x^k - x^*\|^2 + \|z^k - z^*\|^2 \\
&\quad - (\|\bar{x}^{k+1} - x^*\|^2 + \|\bar{z}^{k+1} - z^*\|^2) \\
&\quad - (\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2) \\
&\quad - 2[d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)] \\
&\quad + 2[d_0(x^*, x^k) + d'_0(z^*, z^k)].
\end{aligned} \tag{4.12}$$

Como (x^*, z^*, y^*) é um ponto de sela do Lagrangiano $L(x, z, y)$, obtemos

$$2\lambda_k[L(x^*, z^*, y^*) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, y^*)] \leq 0, \tag{4.13}$$

adicionando as desigualdades (4.12) com (4.13) e reordenando termos, obtemos

$$\begin{aligned}
\|\bar{x}^{k+1} - x^*\|^2 + \|\bar{z}^{k+1} - z^*\|^2 &\leq \|x^k - x^*\|^2 + \|z^k - z^*\|^2 - \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2\} \\
&\quad - 2\lambda_k \langle p^{k+1} - y^*, A\bar{x}^{k+1} + B\bar{z}^{k+1} - b \rangle \\
&\quad - 2[d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)] + 2[d_0(x^*, x^k) + d'_0(z^*, z^k)].
\end{aligned}$$

Por outro lado, notemos que o Passo 1 do Algoritmo (PMAPD) e a definição de

\bar{y}^{k+1} (veja Observação 4.6) podem se escrever equivalentemente como

$$p^{k+1} := \arg \min_{y \in \mathbb{R}^m} \{-L(x^k, z^k, y) + (1/2\lambda_k)\|y - y^k\|^2\},$$

$$\bar{y}^{k+1} := \arg \min_{y \in \mathbb{R}^m} \{-L(\bar{x}^{k+1}, \bar{z}^{k+1}, y) + (1/2\lambda_k)\|y - y^k\|^2\},$$

usando o Lema 3.6 dois vezes com $\tau = \lambda_k$, $F(y) = -L(x^k, z^k, y)$ e $F(y) = -L(\bar{x}^{k+1}, \bar{z}^{k+1}, y)$ respectivamente, obtemos

$$2\lambda_k[L(x^k, z^k, \bar{y}^{k+1}) - L(x^k, z^k, p^{k+1})] \leq \|y^k - \bar{y}^{k+1}\|^2 - \|p^{k+1} - \bar{y}^{k+1}\|^2 - \|p^{k+1} - y^k\|^2,$$

$$2\lambda_k[L(\bar{x}^{k+1}, \bar{z}^{k+1}, y^*) - L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})] \leq \|y^k - y^*\|^2 - \|\bar{y}^{k+1} - y^*\|^2 - \|\bar{y}^{k+1} - y^k\|^2.$$

Adicionando ambas desigualdades e reordenando termos, obtemos (ii). \square

Proposição 4.12 *Supor que as hipóteses do Lema 4.11 satisfazem-se e supor que $\{\lambda_k\}$ satisfaz $\eta < \lambda_k < \bar{c} - \eta$ (para todo k), para algum $\eta \in (0, \bar{c}/2)$ com $\bar{c} := \min\{\frac{\sqrt{\gamma}}{2\|A\|}, \frac{\sqrt{\gamma'}}{2\|B\|}\}$ onde γ, γ' são constantes positivas relacionados com d e d' , respectivamente, na Definição 2.29, (Iii). Sejam (x^*, z^*) uma solução ótima do (PC) e y^* seu correspondente multiplicador de Lagrange, então para qualquer $k \geq 0$*

$$\begin{aligned} \|\bar{w}^{k+1} - w^*\|^2 &\leq \|w^k - w^*\|^2 - D\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 \\ &\quad + \|p^{k+1} - y^k\|^2\} - 2\{d_0(\bar{x}^{k+1}, x^k) - d_0(x^*, x^k) + d'_0(\bar{z}^{k+1}, z^k) - d'_0(z^*, z^k)\} \end{aligned} \quad (4.14)$$

onde $D := \min\{1 - 4(\bar{c} - \eta)^2\|A\|^2, 1 - 4(\bar{c} - \eta)^2\|B\|^2\}$.

Demonstração. Denotemos $w = (x, z, y)$ com a norma associada $\|w\|^2 = \|x\|^2 + \|z\|^2 + \|y\|^2$. Adicionando as desigualdades (i)–(ii) derivados do Lema 4.11, obtemos

$$\begin{aligned} \|\bar{w}^{k+1} - w^*\|^2 &\leq \|w^k - w^*\|^2 - \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2\} - \{\|p^{k+1} - \bar{y}^{k+1}\|^2 \\ &\quad + \|p^{k+1} - y^k\|^2\} - 2\{d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)\} \\ &\quad + 2\{d_0(x^*, x^k) + d'_0(z^*, z^k)\} + \phi \end{aligned} \quad (4.15)$$

onde $\phi := 2\lambda_k \langle \bar{y}^{k+1} - p^{k+1}, A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k) \rangle$.

Como $\bar{y}^{k+1} := y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$ e $p^{k+1} := y^k + \lambda_k(Ax^k + Bz^k - b)$, e usando a desigualdade $(r + q)^2 \leq 2(r^2 + q^2)$, obtemos

$$\begin{aligned} \phi &:= 2\lambda_k \langle \lambda_k(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)), A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k) \rangle \\ &= 2\lambda_k^2 \|A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)\|^2 \\ &\leq 4\lambda_k^2 \|A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + 4\lambda_k^2 \|B\|^2 \|\bar{z}^{k+1} - z^k\|^2. \end{aligned} \quad (4.16)$$

De (4.15) e (4.16), obtemos

$$\begin{aligned}
\|\bar{w}^{k+1} - w^*\|^2 &\leq \|w^k - w^*\|^2 - (1 - 4\lambda_k^2\|A\|^2)\|\bar{x}^{k+1} - x^k\|^2 - (1 - 4\lambda_k^2\|B\|^2)\|\bar{z}^{k+1} - z^k\|^2 \\
&\quad - \{\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} - 2\{d_0(\bar{x}^{k+1}, x^k) + d'_0(\bar{z}^{k+1}, z^k)\} \\
&\quad + 2\{d_0(x^*, x^k) + d'_0(z^*, z^k)\}.
\end{aligned} \tag{4.17}$$

Note que da hipótese para λ_k , temos que $\lambda_k < \bar{c} - \eta$, com $\bar{c} := \min\{\frac{\sqrt{\gamma}}{2\|A\|}, \frac{\sqrt{\gamma'}}{2\|B\|}\}$, e junto à Definição 2.29, obtemos

$$\bar{c} - \eta < \bar{c} \leq \frac{\sqrt{\gamma}}{2\|A\|} \leq \frac{1}{2\|A\|},$$

$$\bar{c} - \eta < \bar{c} \leq \frac{\sqrt{\gamma'}}{2\|B\|} \leq \frac{1}{2\|B\|},$$

então

$$1 - 4(\bar{c} - \eta)^2\|A\|^2 > 0,$$

$$1 - 4(\bar{c} - \eta)^2\|B\|^2 > 0.$$

Além disso, como $\lambda_k < \bar{c} - \eta$ com $\eta \in (0, \bar{c}/2)$, temos que

$$1 - 4(\bar{c} - \eta)^2\|A\|^2 < 1 - 4\lambda_k^2\|A\|^2,$$

$$1 - 4(\bar{c} - \eta)^2\|B\|^2 < 1 - 4\lambda_k^2\|B\|^2,$$

assim,

$$0 < 1 - 4(\bar{c} - \eta)^2\|A\|^2 < 1 - 4\lambda_k^2\|A\|^2,$$

$$0 < 1 - 4(\bar{c} - \eta)^2\|B\|^2 < 1 - 4\lambda_k^2\|B\|^2.$$

Portanto, considerando $D := \min\{1 - 4(\bar{c} - \eta)^2\|A\|^2, 1 - 4(\bar{c} - \eta)^2\|B\|^2\} > 0$, na desigualdade (4.17), obtemos

$$\begin{aligned}
\|\bar{w}^{k+1} - w^*\|^2 &\leq \|w^k - w^*\|^2 - D\{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 \\
&\quad + \|p^{k+1} - y^k\|^2\} - 2\{d_0(\bar{x}^{k+1}, x^k) - d_0(x^*, x^k) + d'_0(\bar{z}^{k+1}, z^k) - d'_0(z^*, z^k)\}.
\end{aligned}$$

□

Lema 4.13 *Se cumpre-se a Hipótese (\mathcal{T}_2), então*

$$\|w^{k+1} - \bar{w}^{k+1}\| \leq \delta_k \|w^{k+1} - w^k\| \tag{4.18}$$

onde $\delta_k := \eta_k \max\{\sqrt{1 + 2\lambda_k^2\|A\|^2}, \sqrt{1 + 2\lambda_k^2\|B\|^2}\}$.

Demonstração. Temos

$$y^{k+1} = y^k + \lambda_k(Ax^{k+1} + Bz^{k+1} - b),$$

$$\bar{y}^{k+1} = y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b),$$

$$\|x^{k+1} - \bar{x}^{k+1}\| \leq \eta_k \|x^{k+1} - x^k\|,$$

$$\|z^{k+1} - \bar{z}^{k+1}\| \leq \eta_k \|z^{k+1} - z^k\|,$$

assim

$$\begin{aligned} \|y^{k+1} - \bar{y}^{k+1}\| &= \lambda_k \|A(x^{k+1} - \bar{x}^{k+1}) + B(z^{k+1} - \bar{z}^{k+1})\| \\ &\leq \eta_k \lambda_k (\|A\| \|x^{k+1} - x^k\| + \|B\| \|z^{k+1} - z^k\|) \end{aligned}$$

usando a desigualdade $(r + q)^2 \leq 2(r^2 + q^2)$, na desigualdade acima, obtemos

$$\|y^{k+1} - \bar{y}^{k+1}\|^2 \leq 2(\eta_k \lambda_k)^2 (\|A\|^2 \|x^{k+1} - x^k\|^2 + \|B\|^2 \|z^{k+1} - z^k\|^2).$$

Portanto, da Hipóteses (\mathcal{T}_2) , e da definição de δ_k dado no Lema, obtemos

$$\begin{aligned} \|w^{k+1} - \bar{w}^{k+1}\|^2 &= \|x^{k+1} - \bar{x}^{k+1}\|^2 + \|z^{k+1} - \bar{z}^{k+1}\|^2 + \|y^{k+1} - \bar{y}^{k+1}\|^2 \\ &\leq \eta_k^2 (\|x^{k+1} - x^k\| + \|z^{k+1} - z^k\|)^2 \\ &\quad + 2(\eta_k \lambda_k)^2 (\|A\|^2 \|x^{k+1} - x^k\|^2 + \|B\|^2 \|z^{k+1} - z^k\|^2) \\ &\leq \eta_k^2 \{ (1 + 2\lambda_k^2 \|A\|^2) \|x^{k+1} - x^k\|^2 + (1 + 2\lambda_k^2 \|B\|^2) \|z^{k+1} - z^k\|^2 \} \\ &\leq \delta_k^2 \|w^{k+1} - w^k\|^2. \end{aligned}$$

□

Agora podemos apresentar e provar nosso resultado da taxa de convergência.

Teorema 4.14 *Sejam $(d_0, H_0) \in \mathcal{F}_+(\bar{\mathcal{C}})$, $(d'_0, H'_0) \in \mathcal{F}_+(\bar{K})$ satisfazendo a condição (Iviii). Seja $\{x^k, z^k, y^k\}$ uma sequência limitada gerada pelo Algoritmo (PMAPD). Supor que as hipóteses do Lema 4.11 e (\mathcal{T}_1) - (\mathcal{T}_4) cumprem-se e λ_k satisfaz $\eta < \lambda_k < \bar{c} - \eta$ (para todo k), para algum $\eta \in (0, \bar{c}/2)$ com $\bar{c} := \min\{\frac{\sqrt{\gamma}}{2\|A\|}, \frac{\sqrt{\gamma'}}{2\|B\|}\}$ onde γ, γ' são constantes positivas relacionados com d e d' , respectivamente, na Definição 2.29, (Iii). Então $\{w^k\}$ converge linearmente para uma única solução ótima $w^* := (x^*, z^*, y^*)$; isto é, existe um inteiro \bar{k} tal que, para todo $k \geq \bar{k}$*

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|$$

onde $\theta_k := \frac{\delta_k}{1 - \delta_k} < \frac{\rho}{1 - \rho} < 1$ para $\rho \in (0, 1/2)$ ou $\theta_k \leq \frac{\sqrt{a^2 N + D} + a\sqrt{N}}{2\sqrt{a^2 N + D}} < 1$ com D definido em (4.14) e $N := \max\{4(\bar{c} - \eta)^2 (\|A^T A\|^2 + \|B^T A\|^2) + 2\alpha^2 \eta^{-2}, 4(\bar{c} - \eta)^2 (\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha')^2 \eta^{-2}\}$.

Demonstração. Sob nossas hipóteses, $\{w^k\}$ é limitada e considerando

$$a_k = \eta_k \|x^{k+1} - x^k\|, \quad b_k = \eta_k \|z^{k+1} - z^k\|$$

cumple-se $\sum_{k=1}^{+\infty} (a_k + b_k) < +\infty$, e portanto o Teorema 3.12 cumple-se, assim, $\{w^k\}$ converge a w^* . Agora estabeleceremos a taxa de convergência.

Do passo 2 do Algoritmo (PMAPD), quando $a_k = b_k = 0$,

$$-\lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k) \in \partial f^k(\bar{x}^{k+1}),$$

$$-\lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k) \in \partial g^k(\bar{z}^{k+1}),$$

logo

$$\begin{aligned} 0 &\in \partial f(\bar{x}^{k+1}) + A^T p^{k+1} + \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k) \\ &= \partial f(\bar{x}^{k+1}) + A^T \bar{y}^{k+1} - A^T (\bar{y}^{k+1} - p^{k+1}) + \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k) \end{aligned} \quad (4.19)$$

$$\begin{aligned} 0 &\in \partial g(\bar{z}^{k+1}) + B^T p^{k+1} + \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k) \\ &= \partial g(\bar{z}^{k+1}) + B^T \bar{y}^{k+1} - B^T (\bar{y}^{k+1} - p^{k+1}) + \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k) \end{aligned} \quad (4.20)$$

da Observação 4.6, temos $\bar{y}^{k+1} := y^k + \lambda_k (A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$, então

$$-\lambda_k^{-1} (\bar{y}^{k+1} - y^k) = b - A\bar{x}^{k+1} - B\bar{z}^{k+1}. \quad (4.21)$$

Além disso, como

$$\begin{aligned} \partial_x L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) &= \partial f(\bar{x}^{k+1}) + A^T \bar{y}^{k+1}, \\ \partial_z L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1}) &= \partial g(\bar{z}^{k+1}) + B^T \bar{y}^{k+1}, \\ \partial_y (-L(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})) &= b - A\bar{x}^{k+1} - B\bar{z}^{k+1}. \end{aligned}$$

Assim, de (4.19), (4.20), (4.21) e da definição de S , veja (4.1), obtemos,

$$(\pi_k, \sigma_k, \xi_k) \in S(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})$$

onde

$$\pi_k := A^T (\bar{y}^{k+1} - p^{k+1}) - \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k), \quad (4.22)$$

$$\sigma_k := B^T (\bar{y}^{k+1} - p^{k+1}) - \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k), \quad (4.23)$$

$$\xi_k := -\lambda_k^{-1} (\bar{y}^{k+1} - y^k). \quad (4.24)$$

Do Passo 1 do Algoritmo (PMAPD); $p^{k+1} := y^k + \lambda_k(Ax^k + Bz^k - b)$, e como $\bar{y}^{k+1} := y^k + \lambda_k(A\bar{x}^{k+1} + B\bar{z}^{k+1} - b)$ subtraindo, obtemos

$$\bar{y}^{k+1} - p^{k+1} := \lambda_k(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)). \quad (4.25)$$

Substituindo (4.25) em (4.22), (4.23), (4.24), obtemos

$$\pi_k := \lambda_k A^T(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)) - \lambda_k^{-1} \nabla_1 d(\bar{x}^{k+1}, x^k) \quad (4.26)$$

$$\sigma_k := \lambda_k B^T(A(\bar{x}^{k+1} - x^k) + B(\bar{z}^{k+1} - z^k)) - \lambda_k^{-1} \nabla_1 d'(\bar{z}^{k+1}, z^k) \quad (4.27)$$

$$\xi_k := -\lambda_k^{-1}(\bar{y}^{k+1} - y^k). \quad (4.28)$$

Por outro lado, seja (x^∞, z^∞) uma solução ótima do (PC) com y^∞ seu correspondente multiplicador de Lagrange tal que $w^k = (x^k, z^k, y^k)$ converge a $w^\infty = (x^\infty, z^\infty, y^\infty)$, temos

$$\|\bar{w}^{k+1} - w^\infty\| \leq \|\bar{w}^{k+1} - w^{k+1}\| + \|w^{k+1} - w^\infty\|,$$

como $\|w^{k+1} - w^\infty\| \rightarrow 0$ e do Lema 4.13, $\|\bar{w}^{k+1} - w^{k+1}\| \rightarrow 0$ quando $k \rightarrow +\infty$, então

$$\|\bar{w}^{k+1} - w^\infty\| \rightarrow 0, \quad (k \rightarrow +\infty)$$

usando este resultado, considerando a Hipótese (\mathcal{T}_3) e tomando limite em ambos lados de (4.14), obtemos

$$\begin{aligned} \|\bar{x}^{k+1} - x^k\| &\rightarrow 0, & \|\bar{z}^{k+1} - z^k\| &\rightarrow 0 \\ \|p^{k+1} - \bar{y}^{k+1}\| &\rightarrow 0, & \|p^{k+1} - y^k\| &\rightarrow 0. \end{aligned} \quad (4.29)$$

Notemos que, da eleição da distância proximal $d(x, y) := d_0(x, y) + (1/2)\|x - y\|^2$, obtemos

$$\begin{aligned} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\| &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) + (\bar{x}^{k+1} - x^k)\| \\ &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k) + (\bar{x}^{k+1} - x^k)\| \end{aligned}$$

e da Hipótese (\mathcal{T}_4) , $\nabla_1 d_0(\cdot, u)$ é localmente Lipschitz, e como $\|\bar{x}^{k+1} - x^\infty\| \rightarrow 0$, $\|x^k - x^\infty\| \rightarrow 0$ com $x^\infty \in \mathcal{C}$, então existem $\alpha_1 > 0$, $k'_1, k''_1 \in \mathbb{N}$ tais que

$$\|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k)\| \leq \alpha_1 \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k'''_1 := \max\{k'_1, k''_1\},$$

assim,

$$\begin{aligned}
\|\nabla_1 d(\bar{x}^{k+1}, x^k)\| &= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) + (\bar{x}^{k+1} - x^k)\| \\
&= \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k) + (\bar{x}^{k+1} - x^k)\| \\
&\leq \|\nabla_1 d_0(\bar{x}^{k+1}, x^k) - \nabla_1 d_0(x^k, x^k)\| + \|\bar{x}^{k+1} - x^k\| \\
&\leq \alpha_1 \|\bar{x}^{k+1} - x^k\| + \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k_1''' := \max\{k_1', k_1''\} \\
&= \alpha \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k_1''' := \max\{k_1', k_1''\}
\end{aligned}$$

onde $\alpha = \alpha_1 + 1$.

Analogamente, da Hipótese (\mathcal{T}_4) , e como $\|\bar{z}^{k+1} - z^\infty\| \rightarrow 0$, $\|z^k - z^\infty\| \rightarrow 0$ com $z^\infty \in \mathbf{K}$, temos que existem $\alpha_2 > 0$, $k_2', k_2'' \in \mathbb{N}$ tais que

$$\|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) - \nabla_1 d'_0(z^k, z^k)\| \leq \alpha_2 \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k_2', k_2''\},$$

e assim

$$\begin{aligned}
\|\nabla_1 d'(\bar{z}^{k+1}, z^k)\| &= \|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) + (\bar{z}^{k+1} - z^k)\| \\
&= \|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) - \nabla_1 d'_0(z^k, z^k) + (\bar{z}^{k+1} - z^k)\| \\
&\leq \|\nabla_1 d'_0(\bar{z}^{k+1}, z^k) - \nabla_1 d'_0(z^k, z^k)\| + \|\bar{z}^{k+1} - z^k\| \\
&\leq \alpha_1 \|\bar{z}^{k+1} - z^k\| + \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k_2', k_2''\} \\
&= \alpha \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k_2', k_2''\},
\end{aligned}$$

onde $\alpha' = \alpha_2 + 1$.

Portanto, obtemos

$$\begin{aligned}
\|\nabla_1 d(\bar{x}^{k+1}, x^k)\| &\leq \alpha \|\bar{x}^{k+1} - x^k\|, \quad \forall k \geq k_1''' := \max\{k_1', k_1''\} \\
\|\nabla_1 d'(\bar{z}^{k+1}, z^k)\| &\leq \alpha' \|\bar{z}^{k+1} - z^k\|, \quad \forall k \geq k_2''' := \max\{k_2', k_2''\},
\end{aligned} \tag{4.30}$$

onde α, α' são constantes positivas. Assim, usando (4.29), (4.30), em (4.26), (4.27) e (4.28) com $\eta < \lambda_k < \bar{c} - \eta$, obtemos

$$(\pi_k, \sigma_k, \xi_k) \rightarrow 0, \quad (k \rightarrow +\infty).$$

Isto é, existe \ddot{k} tal que $\|(\pi_k, \sigma_k, \xi_k)\| < \tau$ para todo $k \geq \ddot{k}$ e usando a Hipótese (\mathcal{T}_1) e os fatos que $0 \in S(x^*, z^*, y^*)$ e $(\pi_k, \sigma_k, \xi_k) \in S(\bar{x}^{k+1}, \bar{z}^{k+1}, \bar{y}^{k+1})$, com a eleição

$$w^* = (x^*, z^*, y^*) \quad v = (\pi_k, \sigma_k, \xi_k)$$

obtemos

$$\|\bar{w}^{k+1} - w^*\| \leq a \|(\pi_k, \sigma_k, \xi_k)\| \quad \forall k \geq \ddot{k}. \tag{4.31}$$

Notemos que, se $a = 0$ então $\bar{w}^{k+1} = w^* \forall k \geq \bar{k}$, e do Lema 4.13, obtemos

$$\begin{aligned} \|w^{k+1} - w^*\| &= \|w^{k+1} - \bar{w}^{k+1}\| \\ &\leq \delta_k \|w^{k+1} - w^k\| \\ &= \delta_k \|(w^{k+1} - w^*) + (w^* - w^k)\| \\ &\leq \delta_k \|w^{k+1} - w^*\| + \delta_k \|w^k - w^*\| \end{aligned}$$

e assim,

$$\|w^{k+1} - w^*\| \leq \theta_k \|w^k - w^*\|, \quad \text{com } \theta_k = \frac{\delta_k}{1 - \delta_k}.$$

Como $\delta_k \rightarrow 0$ (veja (4.18) e o fato que $\eta_k \rightarrow 0$), então existe $\bar{k} \in \mathbb{N}$ tal que

$$\theta_k = \frac{\delta_k}{1 - \delta_k} < \frac{\rho}{1 - \rho} < 1, \quad \forall k \geq \bar{k}$$

para $\rho \in (0, 1/2)$.

Agora, consideremos o caso quando $a > 0$.

Nós vamos estimar a parte direita da desigualdade (4.31). Usando a definição de (π_k, σ_k, ξ_k) e a desigualdade $(r + q)^2 \leq 2(r^2 + q^2)$, obtemos

$$\begin{aligned} \|\pi_k\|^2 &\leq 2\{[\lambda_k(\|A^T A\| \|\bar{x}^{k+1} - x^k\| + \|A^T B\| \|\bar{z}^{k+1} - z^k\|)]^2 + \lambda_k^{-2} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\|^2\} \\ &\leq 2\{2[\lambda_k^2 \|A^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \lambda_k^2 \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2] + \lambda_k^{-2} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\|^2\} \\ &= 4\lambda_k^2 [\|A^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2] + 2\lambda_k^{-2} \|\nabla_1 d(\bar{x}^{k+1}, x^k)\|^2 \\ \|\sigma_k\|^2 &\leq 4\lambda_k^2 [\|B^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \|B^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2] + 2\lambda_k^{-2} \|\nabla_1 d'(\bar{z}^{k+1}, z^k)\|^2 \\ \|\xi_k\|^2 &\leq \lambda_k^{-2} \|\bar{y}^{k+1} - y^k\|^2 = \lambda_k^{-2} \|(\bar{y}^{k+1} - p^{k+1}) + (p^{k+1} - y^k)\|^2 \\ &\leq 2\lambda_k^{-2} (\|\bar{y}^{k+1} - p^{k+1}\|^2 + \|p^{k+1} - y^k\|^2), \end{aligned}$$

assim, de (4.30), obtemos

$$\begin{aligned} \|\pi_k\|^2 &\leq 4\lambda_k^2 (\|A^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2) + 2\alpha^2 \lambda_k^{-2} \|\bar{x}^{k+1} - x^k\|^2 \quad \forall k \geq k_1''' \\ &= (4\lambda_k^2 \|A^T A\|^2 + 2\alpha^2 \lambda_k^{-2}) \|\bar{x}^{k+1} - x^k\|^2 + 4\lambda_k^2 \|A^T B\|^2 \|\bar{z}^{k+1} - z^k\|^2 \quad \forall k \geq k_1''' \\ \|\sigma_k\|^2 &\leq 4\lambda_k^2 \|B^T A\|^2 \|\bar{x}^{k+1} - x^k\|^2 + (2(\alpha')^2 \lambda_k^{-2} + 4\lambda_k^2 \|B^T B\|^2) \|\bar{z}^{k+1} - z^k\|^2 \quad \forall k \geq k_2''' \\ \|\xi_k\|^2 &\leq 2\lambda_k^{-2} (\|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2). \end{aligned}$$

Portanto, para todo $k \geq k_3''' := \max\{k_1''', k_2'''\}$, obtemos

$$\begin{aligned} \|(\pi_k, \sigma_k, \xi_k)\| &= \|\pi_k\|^2 + \|\sigma_k\|^2 + \|\xi_k\|^2 \\ &\leq D_k \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \end{aligned} \quad (4.32)$$

onde

$$D_k := \max\{4\lambda_k^2(\|A^T A\|^2 + \|B^T A\|^2) + 2\alpha^2\lambda_k^{-2}; 4\lambda_k^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha')^2\lambda_k^{-2}; 2\lambda_k^{-2}\},$$

como $\eta < \lambda_k < \bar{c} - \eta$, notemos que D_k é positivo e limitado.

De fato, da definição de D_k , $D_k \geq 2\lambda_k^{-2}$ assim, $D_k \geq M$ com $M := 2(\bar{c} - \eta)^{-2} > 0$. Por outro lado, como $\alpha > 1$, $\alpha' > 1$, então

$$D_k = \max\{4\lambda_k^2(\|A^T A\|^2 + \|B^T A\|^2) + 2\alpha^2\lambda_k^{-2}; 4\lambda_k^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha')^2\lambda_k^{-2}\},$$

porém, como $\eta < \lambda_k < \bar{c} - \eta$, então

$$\begin{aligned} 4\lambda_k^2(\|A^T A\|^2 + \|B^T A\|^2) + 2\alpha^2\lambda_k^{-2} &< 4(\bar{c} - \eta)^2(\|A^T A\|^2 + \|B^T A\|^2) + 2\alpha^2\eta^{-2} \\ 4\lambda_k^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha')^2\lambda_k^{-2} &< 4(\bar{c} - \eta)^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha')^2\eta^{-2} \end{aligned}$$

considerando

$$N := \max\{4(\bar{c} - \eta)^2(\|A^T A\|^2 + \|B^T A\|^2) + 2\alpha^2\eta^{-2}, 4(\bar{c} - \eta)^2(\|A^T B\|^2 + \|B^T B\|^2) + 2(\alpha')^2\eta^{-2}\},$$

obtemos que $D_k < N$, e portanto

$$0 < M \leq D_k < N. \quad (4.33)$$

Agora, de (4.31), (4.32) e (4.33), para todo $k \geq \hat{k} := \max\{\hat{k}, k_3'''\}$, obtemos

$$\begin{aligned} \|\bar{w}^{k+1} - w^*\|^2 &\leq a^2 D_k \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \\ &< a^2 N \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 + \|p^{k+1} - y^k\|^2\} \end{aligned} \quad (4.34)$$

além disso, de (4.14) e a Hipótese (\mathcal{T}_3) , para todo $k \geq k_0$, obtemos

$$\begin{aligned} \|\bar{w}^{k+1} - w^*\|^2 &\leq \|w^k - w^*\|^2 - D \{\|\bar{x}^{k+1} - x^k\|^2 + \|\bar{z}^{k+1} - z^k\|^2 + \|p^{k+1} - \bar{y}^{k+1}\|^2 \\ &\quad + \|p^{k+1} - y^k\|^2\}, \end{aligned} \quad (4.35)$$

assim, para $\check{k} := \max\{\hat{k}, k_0\}$, multiplicando por D e a^2N nas desigualdades (4.34) e (4.35), respectivamente e adicionando obtemos

$$a^2N\|\bar{w}^{k+1} - w^*\|^2 + D\|\bar{w}^{k+1} - w^*\|^2 \leq a^2N\|w^k - w^*\|^2 \quad \forall k \geq \check{k}.$$

Definindo $\nu := a\sqrt{N}/\sqrt{a^2N + D}$, a última desigualdade se reduce para

$$\|\bar{w}^{k+1} - w^*\| \leq \nu\|w^k - w^*\| \quad \forall k \geq \check{k}, \quad (4.36)$$

além disso, do Lema 4.13, tem-se que

$$\begin{aligned} \|w^{k+1} - \bar{w}^{k+1}\| &\leq \delta_k\|w^{k+1} - w^k\| = \delta_k\|(w^{k+1} - w^*) + (w^* - w^k)\| \\ &\leq \delta_k\|w^{k+1} - w^*\| + \delta_k\|w^k - w^*\|. \end{aligned} \quad (4.37)$$

Portanto, de (4.36) e (4.37), obtemos

$$\begin{aligned} \|w^{k+1} - w^*\| &= \|(w^{k+1} - \bar{w}^{k+1}) + (\bar{w}^{k+1} - w^*)\| \\ &\leq \|w^{k+1} - \bar{w}^{k+1}\| + \|\bar{w}^{k+1} - w^*\| \\ &\leq \delta_k\|w^{k+1} - w^*\| + \delta_k\|w^k - w^*\| + \nu\|w^k - w^*\|, \quad \forall k \geq \check{k}, \end{aligned}$$

então

$$\|w^{k+1} - w^*\| \leq \theta_k\|w^k - w^*\|, \quad \forall k \geq \check{k} \quad \text{com} \quad \theta_k = \frac{\nu + \delta_k}{1 - \delta_k}.$$

Como $\delta_k \rightarrow 0$ e de

$$\frac{\sqrt{a^2N + D} + a\sqrt{N}}{2\sqrt{a^2N + D}} > \frac{a\sqrt{N}}{\sqrt{a^2N + D}} = \nu,$$

para algum $\bar{k} \geq \check{k}$, obtemos

$$\|w^{k+1} - w^*\| \leq \theta_k\|w^k - w^*\|, \quad \forall k \geq \bar{k}$$

com

$$1 > \frac{\sqrt{a^2N + D} + a\sqrt{N}}{2\sqrt{a^2N + D}} \geq \theta_k.$$

□

4.3 Experimentos Numéricos com o Algoritmo (PMAPD)

Nesta seção apresentamos os resultados numéricos da implementação do algoritmo (PMAPD).

Para as implementações, usamos computador com as seguintes configurações: Intel Core i5 2.30 GHz, 3GB de RAM, o sistema operacional Windows 7 SP1 64 bits e o software MATLAB versão 7.10 (R2010a).

Para cada exemplo faremos as implementações com cada uma das distâncias proximais definidas na Seção 2.2 do Capítulo 2, as quais são: distância de Bregman, distância proximal φ -divergência e distância proximal homogênea de segundo ordem.

Observação 4.15 *Lembremos que se $a_k = b_k = 0$ para todo k , o Passo 2 do algoritmo (PMAPD) torna-se: encontrar (x^{k+1}, z^{k+1}) tais que*

$$x^{k+1} := \arg \min_{x \in \mathcal{C}} \{f(x) + \langle p^{k+1}, Ax \rangle + (1/\lambda_k)d(x, x^k)\}, \quad (4.38)$$

$$z^{k+1} := \arg \min_{z \in \bar{K}} \{g(z) + \langle p^{k+1}, Bz \rangle + (1/\lambda_k)d'(z, z^k)\}. \quad (4.39)$$

Nas tabelas seguintes denotamos por k à ordem de iteração do algoritmo, $N[x^k]$ ao número total de iterações em que se consegue resolver o subproblema 4.38 na k -ésima iteração, $N[z^k]$ ao número total de iterações em que se consegue resolver o subproblema 4.39 na k -ésima iteração, x^k e z^k são os pontos obtidos na iteração k , $f(x^k) + g(z^k)$ é o valor da função objetivo na iteração k e

$$|q^k - q^{k-1}|_{\bar{M}} = \max\{\|x^k - x^{k-1}\|, \|z^k - z^{k-1}\|, \|y^k - y^{k-1}\|\} \text{ para cada } k = 1, 2, \dots$$

Além disso, devemos ressaltar que em todas as implementações usamos a função *fmincon* (escolhendo o algoritmo do ponto interior) do software MATLAB para resolver os subproblemas.

Exemplo 4.16 *Consideremos o problema*

$$\left\{ \begin{array}{l} \min \quad (x_1 - 1)^2 + (x_2 - 1)^2 + (z_1 - 1)^2 + (z_2 - 1)^2 \\ \text{s.a} \\ \quad x_1 + 2x_2 + 2z_1 - z_2 = 4 \\ \quad -2x_1 + x_2 + z_1 + z_2 = 1 \\ \quad x_i \geq 0, \quad i = 1, 2. \\ \quad z_i \geq 0, \quad i = 1, 2. \end{array} \right.$$

É claro que a função $f + g : \mathbb{R}^4 \rightarrow \mathbb{R}$ definida por

$$f(x_1, x_2) + g(z_1, z_2) = (x_1 - 1)^2 + (x_2 - 1)^2 + (z_1 - 1)^2 + (z_2 - 1)^2,$$

é própria, contínua (assim é semicontínua inferior) e convexa. Além disso, note que o ponto $(x^*, z^*) = ((1, 1), (1, 1))$ é o ponto mínimo global da função $f + g$, $f(x^*, z^*) = 0$ e $y^* = (0, 0)$ é o ótimo multiplicador de Lagrange associado às restrições de igualdade.

O problema é da forma (PC), considerando:

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}, B = \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 4 \\ 1 \end{bmatrix}, \bar{C} = \mathbb{R}_+^n, \bar{K} = \mathbb{R}_+^n.$$

Com esses dados, $\|A\| = 2.2361$, $\|B\| = 2.3028$ e tomando $\mu = \mu' = 1$, $\gamma = \gamma' = 1$, obtemos $\bar{c} := \min\{\frac{\sqrt{\gamma\mu}}{2\|A\|}, \frac{\sqrt{\gamma'\mu'}}{2\|B\|}\} = 0.2171$, assim, devemos elegir $\eta \in (0, 0.1085)$ e $\lambda_k \in (\eta, \bar{c} - \eta)$.

Neste exemplo, tomaremos $\eta = 0.0543$, o erro $tol = 10^{-4}$, $a_k = 0$, $b_k = 0$, $\lambda_k = 0.125$ para todo k e o ponto inicial $(x^0, z^0, y^0) = ((1, 2), (3, 2), (1, 1))$.

(a) Escolhendo a distância de Bregman (Kullback-Leibler), definida por:

$$d_0(u, v) := \sum_{i=1}^n u_i \log\left(\frac{u_i}{v_i}\right) + v_i - u_i,$$

e usando o algoritmo (PMAPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	(x_1^k, x_2^k)	(z_1^k, z_2^k)	(y_1^k, y_2^k)	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	9	9	(1.07766, 1.53611)	(2.32791, 1.86739)	(1.36729, 1.32201)	0.68505	2.80916
2	8	7	(1.15916, 1.14951)	(1.77014, 1.75337)	(1.52293, 1.49135)	0.56931	1.20836
3	8	7	(1.23859, 0.85480)	(1.33815, 1.65208)	(1.51948, 1.53733)	0.44371	0.61756
4	9	8	(1.31067, 0.64616)	(1.02346, 1.56054)	(1.40565, 1.48843)	0.32773	0.53648
5	9	10	(1.37117, 0.50764)	(0.80677, 1.47744)	(1.22097, 1.36962)	0.23208	0.64547
6	9	8	(1.41710, 0.42185)	(0.66622, 1.40227)	(0.99484, 1.20164)	0.28169	0.78146
7	9	7	(1.44690, 0.37486)	(0.58284, 1.33478)	(0.74828, 1.00148)	0.31758	0.87662
8	9	8	(1.46026, 0.35745)	(0.54277, 1.27471)	(0.49653, 0.78328)	0.33315	0.90923
9	8	8	(1.45799, 0.36462)	(0.53712, 1.22171)	(0.25150, 0.55921)	0.33203	0.87687
10	9	8	(1.44174, 0.39429)	(0.56063, 1.17541)	(0.02352, 0.34007)	0.31623	0.78583
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
85	6	6	(0.99939, 0.99991)	(0.99995, 1.00041)	(-0.00023, -0.00063)	0.00025	0.00000
86	6	6	(0.99943, 0.99999)	(1.00002, 1.00037)	(-0.00034, -0.00044)	0.00022	0.00000
87	7	6	(0.99949, 1.00005)	(1.00009, 1.00032)	(-0.00041, -0.00025)	0.00020	0.00000
88	7	5	(0.99957, 1.00011)	(1.00013, 1.00026)	(-0.00044, -0.00008)	0.00017	0.00000
89	6	5	(0.99965, 1.00014)	(1.00017, 1.00020)	(-0.00043, 0.00007)	0.00015	0.00000
90	6	5	(0.99974, 1.00016)	(1.00018, 1.00014)	(-0.00040, 0.00020)	0.00013	0.00000
91	6	4	(0.99982, 1.00016)	(1.00018, 1.00009)	(-0.00034, 0.00029)	0.00011	0.00000
92	6	6	(0.99990, 1.00016)	(1.00017, 1.00004)	(-0.00028, 0.00037)	0.00010	0.00000

Na tabela anterior notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 92 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((1, 1), (1, 1))$, pois o critério de parada foi satisfeito em 0.00010.

(b) Escolhendo a distância proximal φ -divergência, definida por:

$$d_0(u, v) := \sum_{i=1}^n v_i \varphi\left(\frac{u_i}{v_i}\right),$$

com $\varphi(t) := t - \log t - 1$, e usando o algoritmo (PMAPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	(x_1^k, x_2^k)	(z_1^k, z_2^k)	(y_1^k, y_2^k)	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	10	9	(1.07896, 1.55436)	(2.34741, 1.86871)	(1.37672, 1.32657)	0.66567	2.88370
2	8	8	(1.16156, 1.18068)	(1.80330, 1.75623)	(1.54838, 1.50371)	0.55561	1.27593
3	8	8	(1.27593, 0.89171)	(1.37769, 1.65688)	(1.56384, 1.55905)	0.43705	0.64431
4	9	8	(1.31423, 0.68202)	(1.06233, 1.56757)	(1.46827, 1.51949)	0.32776	0.52587
5	9	9	(1.37483, 0.53774)	(0.83959, 1.48679)	(1.29860, 1.40879)	0.23694	0.61689
6	8	9	(1.42074, 0.44384)	(0.68979, 1.41381)	(1.08288, 1.24704)	0.26963	0.75381
7	8	8	(1.45052, 0.38807)	(0.59591, 1.34818)	(0.84166, 1.05093)	0.31087	0.86194
8	9	9	(1.46398, 0.36215)	(0.54534, 1.28951)	(0.59035, 0.83456)	0.33163	0.91266
9	6	9	(1.46200, 0.36151)	(0.52992, 1.23739)	(0.34128, 0.61017)	0.33524	0.89844
10	9	9	(1.44625, 0.38445)	(0.54491, 1.19142)	(0.10547, 0.38870)	0.32350	0.82178
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
85	6	6	(0.99940, 0.99990)	(0.99993, 1.00042)	(-0.00019, -0.00065)	0.00025	0.00000
86	7	6	(0.99943, 0.99998)	(1.00001, 1.00038)	(-0.00032, -0.00046)	0.00022	0.00000
87	7	6	(0.99949, 1.00004)	(1.00007, 1.00033)	(-0.00039, -0.00028)	0.00020	0.00000
88	6	6	(0.99956, 1.00010)	(1.00012, 1.00027)	(-0.00043, -0.00011)	0.00017	0.00000
89	7	6	(0.99964, 1.00013)	(1.00016, 1.00021)	(-0.00043, 0.00004)	0.00015	0.00000
90	6	6	(0.99973, 1.00015)	(1.00017, 1.00015)	(-0.00040, 0.00017)	0.00013	0.00000
91	6	4	(0.99981, 1.00016)	(1.00018, 1.00010)	(-0.00035, 0.00027)	0.00011	0.00000
92	6	4	(0.99989, 1.00016)	(1.00017, 1.00005)	(-0.00029, 0.00035)	0.00010	0.00000

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 92 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((1, 1), (1, 1))$, pois o critério de parada foi satisfeito em 0.00010.

(c) Escolhendo a distância proximal homogênea de segundo ordem, definida por:

$$d_0(u, v) := \sum_{i=1}^n v_i^2 \varphi\left(\frac{u_i}{v_i}\right),$$

com $\varphi(t) := \mu(-\log t + t - 1) + \frac{\nu}{2}(t - 1)^2$.

Para $\mu = \nu = 1$, usando o algoritmo (PMAPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	(x_1^k, x_2^k)	(z_1^k, z_2^k)	(y_1^k, y_2^k)	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	9	10	(1.05373, 1.75126)	(2.67557, 1.92870)	(1.49733, 1.40601)	0.64202	4.23727
2	8	8	(1.11324, 1.48502)	(2.33850, 1.86495)	(1.85925, 1.71376)	0.47507	2.78779
3	9	8	(1.17562, 1.21723)	(2.00409, 1.80668)	(2.08570, 1.92335)	0.33944	1.73696
4	8	9	(1.23823, 0.96155)	(1.68549, 1.75204)	(2.18323, 2.03868)	0.32325	1.09370
5	9	8	(1.29874, 0.72903)	(1.39332, 1.69957)	(2.16372, 2.06673)	0.29684	0.80676
6	9	9	(1.35515, 0.52794)	(1.13556, 1.64815)	(2.04297, 2.01690)	0.26284	0.78745
7	9	8	(1.40578, 0.36359)	(0.91755, 1.59707)	(1.83935, 1.90023)	0.23468	0.93296
8	10	10	(1.44925, 0.23836)	(0.74222, 1.54597)	(1.57240, 1.72874)	0.31729	1.14645
9	10	8	(1.48453, 0.15182)	(0.61048, 1.49483)	(1.26169, 1.51475)	0.37727	1.35076
10	12	9	(1.51088, 0.10154)	(0.52189, 1.44386)	(0.92592, 1.27044)	0.41524	1.49383
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
125	6	6	(1.00058, 0.99959)	(0.99962, 0.99980)	(0.00042, -0.00024)	0.00029	0.00000
126	5	6	(1.00049, 0.99961)	(0.99965, 0.99984)	(0.00032, -0.00048)	0.00026	0.00000
127	5	7	(1.00039, 0.99965)	(0.99968, 0.99989)	(0.00021, -0.00067)	0.00022	0.00000
128	5	7	(1.00029, 0.99970)	(0.99973, 0.99994)	(0.00012, -0.00082)	0.00018	0.00000
129	4	7	(1.00019, 0.99976)	(0.99979, 0.99998)	(0.00003, -0.00093)	0.00014	0.00000
130	7	6	(1.00010, 0.99982)	(0.99985, 1.00002)	(-0.00004, -0.00099)	0.00011	0.00000
131	7	7	(1.00001, 0.99989)	(0.99991, 1.00005)	(-0.00010, -0.00101)	0.00010	0.00000
132	8	6	(0.99994, 0.99995)	(0.99997, 1.00008)	(-0.00014, -0.00100)	0.00010	0.00000

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 132 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((1, 1), (1, 1))$, pois o critério de parada foi satisfeito em 0.00010.

Neste exemplo podemos ressaltar a seguinte diferença, usando as distâncias de Bregman e ϕ -divergência, com os mesmos dados iniciais, o programa levou 92 iterações para se aproximar à solução do problema enquanto a distância proximal homogênea de segundo ordem levou 132 iterações.

Exemplo 4.17 Consideremos o problema (apresentado por Stefanov, [28])

$$\left\{ \begin{array}{l} \min \quad F(u) = 4u_1^2 + 10u_2^2 + 4u_3^2 + 3u_4^2 + 7u_5^2 + 3u_6^2 + u_7^2 \\ \text{s.a} \\ \sum_{j=1}^7 u_j = 72 \\ 4 \leq u_1 \leq 7 \\ 4.5 \leq u_2 \leq 10 \\ 8 \leq u_3 \leq 13 \\ 5 \leq u_4 \leq 8 \\ 4 \leq u_5 \leq 7 \\ 30 \leq u_6 \leq 40 \\ 4 \leq u_7 \leq 7. \end{array} \right.$$

É claro que a função $F : \mathbb{R}^7 \rightarrow \mathbb{R}$ definida por

$$F(u) = 4u_1^2 + 10u_2^2 + 4u_3^2 + 3u_4^2 + 7u_5^2 + 3u_6^2 + u_7^2,$$

é própria, contínua (assim é semicontínua inferior) e convexa.

Denotando $x = (u_1, u_2, u_3, u_4)$, $z = (u_5, u_6, u_7)$, fazendo

$$f(x) := 4u_1^2 + 10u_2^2 + 4u_3^2 + 3u_4^2,$$

$$g(z) := 7u_5^2 + 3u_6^2 + u_7^2,$$

claramente $F(u) = f(x) + g(z)$. Assim, o problema é da forma (PC), considerando $A = [1, 1, 1, 1]$, $B = [1, 1, 1]$, $b = 72$,

$$\bar{C} = \{x \in \mathbb{R}^4 : 4 \leq u_1 \leq 7, 4.5 \leq u_2 \leq 10, 8 \leq u_3 \leq 13, 5 \leq u_4 \leq 8\}$$

$$\bar{K} = \{z \in \mathbb{R}^3 : 4 \leq u_5 \leq 7, 30 \leq u_6 \leq 40, 4 \leq u_7 \leq 7\}.$$

Usando a definição, é fácil provar que \bar{C} e \bar{K} são conjuntos convexos.

Podemos resolver o problema usando a função `fmincon` do software MATLAB, o qual dá: $(x^*, z^*) = ((7, 4.5, 9.8636, 8), (5.6363, 30, 7))$ e $F(x^*, z^*) = 3951.0454$.

Dos dados do problema, obtemos $\|A\| = 2$, $\|B\| = 1.7321$ e tomando $\mu = \mu' = 1$, $\gamma = \gamma' = 1$, obtemos $\bar{c} := \min\left\{\frac{\sqrt{\gamma\mu}}{2\|A\|}, \frac{\sqrt{\gamma'\mu'}}{2\|B\|}\right\} = 0.25$, assim, devemos elegir $\eta \in (0, 0.125)$ e $\lambda_k \in (\eta, \bar{c} - \eta)$.

Neste exemplo, tomaremos $\eta = 0.0625$, o erro $tol = 10^{-3}$, $a_k = 0$, $b_k = 0$, $\lambda_k = 0.125$ para todo k e o ponto inicial $(x^0, z^0, y^0) = ((3, 3, 2, 4), (3, 2, 5), 1)$.

(a) Escolhendo a distância de Bregman (Kullback-Leibler), definida por:

$$d_0(u, v) := \sum_{i=1}^n u_i \log\left(\frac{u_i}{v_i}\right) + v_i - u_i,$$

e usando o algoritmo (PMPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	$(x_1^k, x_2^k, x_3^k, x_4^k)$	(z_1^k, z_2^k, z_3^k)	y^k	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	8	9	(4, 4.5, 8, 5)	(4, 30, 4.59293)	-0.48838	28.02081	3430.59506
2	10	9	(4, 4.5, 8, 5)	(4, 30, 4.00001)	-2.05088	1.56250	3425.50028
3	10	11	(4, 4.5, 8, 5)	(4, 30, 4)	-3.61338	1.56250	3425.50002
4	10	10	(4, 4.5, 8, 5)	(4, 30, 4)	-5.17588	1.56250	3425.50002
5	11	10	(4, 4.5, 8, 5)	(4, 30, 4)	-6.73838	1.56250	3425.50002
6	10	10	(4, 4.5, 8, 5)	(4, 30, 4.02514)	-8.29774	1.55936	3425.70178
7	12	9	(4, 4.5, 8, 5)	(4, 30, 4.17633)	-9.83820	1.54046	3426.94176
8	12	10	(4, 4.5, 8, 5)	(4, 30, 4.43149)	-11.34676	1.50856	3429.13810
9	12	9	(4, 4.5, 8, 5)	(4, 30, 4.77157)	-12.81281	1.46605	3432.26791
10	12	9	(4, 4.5, 8, 5)	(4, 30, 5.17952)	-14.22787	1.41506	3436.32740
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
312	9	9	(7, 4.5, 9.85768, 8)	(5.63300, 30, 7)	-78.86272	0.00117	3950.30985
313	9	9	(7, 4.5, 9.85782, 8)	(5.63308, 30, 7)	-78.86386	0.00114	3950.32790
314	9	9	(7, 4.5, 9.85797, 8)	(5.63316, 30, 7)	-78.86497	0.00111	3950.34550
315	9	9	(7, 4.5, 9.85811, 8)	(5.63324, 30, 7)	-78.86605	0.00108	3950.36266
316	9	9	(7, 4.5, 9.85824, 8)	(5.63331, 30, 7)	-78.86710	0.00106	3950.37938
317	9	9	(7, 4.5, 9.85837, 8)	(5.63339, 30, 7)	-78.86813	0.00103	3950.39571
318	9	9	(7, 4.5, 9.85850, 8)	(5.63346, 30, 7)	-78.86914	0.00100	3950.41164
319	9	9	(7, 4.5, 9.85863, 8)	(5.63353, 30, 7)	-78.87012	0.00098	3950.42718

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 319 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((7, 4.5, 9.8636, 8), (5.6363, 30, 7))$, pois o critério de parada foi satisfeito em 0.00098.

(b) Escolhendo a distância proximal φ -divergência, definida por:

$$d_0(u, v) := \sum_{i=1}^n v_i \varphi\left(\frac{u_i}{v_i}\right),$$

com $\varphi(t) := t - \log t - 1$, e usando o algoritmo (PMAPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	$(x_1^k, x_2^k, x_3^k, x_4^k)$	(z_1^k, z_2^k, z_3^k)	y^k	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	9	9	(4, 4.5, 8, 5)	(4, 30, 4.59543)	-0.48807	28.02077	3430.61799
2	11	13	(4, 4.5, 8, 5)	(4, 30, 4.00061)	-2.05049	1.56242	3425.50490
3	11	9	(4, 4.5, 8, 5)	(4, 30, 4)	-3.61299	1.56250	3425.50003
4	10	10	(4, 4.5, 8, 5)	(4, 30, 4.00006)	-5.17549	1.56249	3425.50053
5	9	11	(4, 4.5, 8, 5)	(4, 30, 4.00006)	-6.73798	1.56249	3425.50053
6	11	12	(4, 4.5, 8, 5)	(4, 30, 4.02516)	-8.29733	1.55935	3425.70194
7	9	9	(4, 4.5, 8, 5)	(4, 30, 4.17676)	-9.83774	1.54040	3426.94537
8	9	9	(4, 4.5, 8, 5)	(4, 30, 4.43299)	-11.34611	1.50838	3429.15140
9	9	10	(4, 4.5, 8, 5)	(4, 30, 4.77460)	-12.81179	1.46567	3432.29684
10	10	10	(4, 4.5, 8, 5)	(4, 30, 5.18422)	-14.22626	1.41447	3436.37613
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
312	9	10	(7, 4.5, 9.85768, 8)	(5.63300, 30, 7)	-78.86271	0.00117	3950.30973
313	9	10	(7, 4.5, 9.85782, 8)	(5.63308, 30, 7)	-78.86385	0.00114	3950.32775
314	9	10	(7, 4.5, 9.85797, 8)	(5.63316, 30, 7)	-78.86496	0.00111	3950.34537
315	9	10	(7, 4.5, 9.85811, 8)	(5.63324, 30, 7)	-78.86604	0.00108	3950.36253
316	9	10	(7, 4.5, 9.85824, 8)	(5.63331, 30, 7)	-78.86710	0.00106	3950.37928
317	9	10	(7, 4.5, 9.85837, 8)	(5.63339, 30, 7)	-78.86813	0.00103	3950.39561
318	9	10	(7, 4.5, 9.85850, 8)	(5.63346, 30, 7)	-78.86913	0.00100	3950.41156
319	9	10	(7, 4.5, 9.85863, 8)	(5.63353, 30, 7)	-78.87011	0.00098	3950.42708

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 319 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((7, 4.5, 9.8636, 8), (5.6363, 30, 7))$, pois o critério de parada foi satisfeito em 0.00098.

(c) Escolhendo a distância proximal homogênea de segundo ordem, definida por:

$$d_0(u, v) := \sum_{i=1}^n v_i^2 \varphi\left(\frac{u_i}{v_i}\right),$$

com $\varphi(t) := \mu(-\log t + t - 1) + \frac{\nu}{2}(t - 1)^2$. Para $\mu = \nu = 1$, usando o algoritmo (PMPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	$(x_1^k, x_2^k, x_3^k, x_4^k)$	(z_1^k, z_2^k, z_3^k)	y^k	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	10	9	(4, 4.5, 8, 5)	(4, 30, 4.81939)	-0.46008	28.01843	3432.72653
2	12	12	(4, 4.5, 8, 5)	(4, 30, 4.52828)	-1.95654	1.49647	3430.00532
3	12	12	(4, 4.5, 8, 5)	(4, 30, 4.31597)	-3.47955	1.52300	3428.12761
4	13	13	(4, 4.5, 8, 5)	(4, 30, 4.17779)	-5.01982	1.54028	3426.95390
5	13	13	(4, 4.5, 8, 5)	(4, 30, 4.10908)	-6.56869	1.54886	3426.38458
6	13	13	(4, 4.5, 8, 5)	(4, 30, 4.10522)	-8.11803	1.54935	3426.35282
7	10	13	(4, 4.5, 8, 5)	(4, 30, 4.16149)	-9.66035	1.54231	3426.81800
8	13	13	(4, 4.5, 8, 5)	(4, 30, 4.27314)	-11.18871	1.52836	3427.75977
9	14	13	(4, 4.5, 8, 5)	(4, 30, 4.43538)	-12.69678	1.50808	3429.17263
10	13	13	(4, 4.5, 8, 5)	(4, 30, 4.64341)	-14.17886	1.48207	3431.06125
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
300	12	11	(7, 4.5, 9.85746, 8)	(5.63296, 30, 7)	-78.86357	0.00120	3950.29006
301	12	11	(7, 4.5, 9.85762, 8)	(5.63305, 30, 7)	-78.86474	0.00117	3950.30941
302	12	11	(7, 4.5, 9.85778, 8)	(5.63313, 30, 7)	-78.86587	0.00114	3950.32827
303	12	11	(7, 4.5, 9.85793, 8)	(5.63322, 30, 7)	-78.86698	0.00111	3950.34663
304	12	11	(7, 4.5, 9.85807, 8)	(5.63330, 30, 7)	-78.86806	0.00108	3950.36453
305	12	11	(7, 4.5, 9.85821, 8)	(5.63337, 30, 7)	-78.86911	0.00105	3950.38198
306	12	11	(7, 4.5, 9.85835, 8)	(5.63345, 30, 7)	-78.87014	0.00102	3950.39898
307	12	11	(7, 4.5, 9.85849, 8)	(5.63353, 30, 7)	-78.87113	0.00100	3950.41554

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 307 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((7, 4.5, 9.8636, 8), (5.6363, 30, 7))$, pois o critério de parada foi satisfeito em 0.001.

Neste exemplo podemos ressaltar a seguinte diferença, usando as distâncias de Bregman e ϕ -divergência, com os mesmos dados iniciais, o programa levou 319 iterações para se aproximar à solução do problema enquanto a distância proximal homogênea de segundo ordem levou 307 iterações.

Exemplo 4.18 *Consideremos o problema*

$$\left\{ \begin{array}{l} \min \quad |x_1 - 1| + |x_2 - 1| + |z_1 - 1| + |z_2 - 1| \\ \text{s.a} \\ \quad \quad \quad x_1 + 2x_2 + 2z_1 + z_2 = 6 \\ \quad \quad \quad 4x_1 + 3x_2 + 5z_1 = 12 \\ \quad \quad \quad 0.5 \leq x_1 \leq 2 \\ \quad \quad \quad 0.5 \leq x_2 \leq 2 \\ \quad \quad \quad z_1 \geq 0.5 \\ \quad \quad \quad z_2 \geq 0.5. \end{array} \right.$$

É claro que a função $f + g : \mathbb{R}^4 \rightarrow \mathbb{R}$ definida por

$$f(x_1, x_2) + g(z_1, z_2) = |x_1 - 1| + |x_2 - 1| + |z_1 - 1| + |z_2 - 1|,$$

é própria, contínua (assim é semicontínua inferior) e convexa.

Além disso, note que o ponto $(x^*, z^*) = ((1, 1), (1, 1))$ é o ponto mínimo global da função $f + g$ e $f(x^*, z^*) = 0$.

O problema é da forma (PC), considerando:

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}, B = \begin{bmatrix} 2 & 1 \\ 5 & 0 \end{bmatrix}, b = \begin{bmatrix} 6 \\ 12 \end{bmatrix},$$

$$\bar{C} = \{x \in \mathbb{R}^2 : 0.5 \leq x_1 \leq 2, 0.5 \leq x_2 \leq 2\}, \bar{K} = \{z \in \mathbb{R}^2 : z_1 \geq 0.5, z_2 \geq 0.5\}.$$

Com esses dados, $\|A\| = 5.3983$, $\|B\| = 5.3983$ e tomando $\mu = \mu' = 1$, $\gamma = \gamma' = 1$, obtemos $\bar{c} := \min\{\frac{\sqrt{\gamma\mu}}{2\|A\|}, \frac{\sqrt{\gamma'\mu'}}{2\|B\|}\} = 0.0926$, assim, devemos elegir $\eta \in (0, 0.0463)$ e $\lambda_k \in (\eta, \bar{c} - \eta)$.

Neste exemplo, tomaremos $\eta = 0.0232$, o erro $tol = 10^{-3}$, $a_k = 0$, $b_k = 0$, $\lambda_k = 0.0347$ para todo k e o ponto inicial $(x^0, z^0, y^0) = ((1, 2), (3, 2), (1, 1))$.

(a) Escolhendo a distância de Bregman (Kullback-Leibler), definida por:

$$d_0(u, v) := \sum_{i=1}^n u_i \log\left(\frac{u_i}{v_i}\right) + v_i - u_i,$$

e usando o algoritmo (PMPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	(x_1^k, x_2^k)	(z_1^k, z_2^k)	(y_1^k, y_2^k)	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	10	9	(0.89788, 1.82148)	(2.72384, 1.94834)	(1.20601, 1.37043)	0.42386	3.59579
2	8	9	(0.78118, 1.62296)	(2.41108, 1.89330)	(1.37058, 1.64973)	0.32418	3.14616
3	9	8	(0.66136, 1.41678)	(2.08193, 1.83604)	(1.49185, 1.83383)	0.33409	2.67339
4	11	9	(0.54801, 1.21429)	(1.75564, 1.77767)	(1.57046, 1.92450)	0.33148	2.19959
5	8	8	(0.50000, 1.02491)	(1.44889, 1.71920)	(1.61095, 1.93558)	0.31227	1.69300
6	9	8	(0.50000, 0.88694)	(1.17304, 1.66139)	(1.62071, 1.88443)	0.28184	1.44749
7	9	9	(0.50000, 0.76524)	(0.96831, 1.60478)	(1.60586, 1.78509)	0.21241	1.37124
8	9	9	(0.50000, 0.66079)	(0.79603, 1.54976)	(1.56989, 1.64499)	0.18085	1.59295
9	7	10	(0.50000, 0.57370)	(0.65626, 1.49668)	(1.51633, 1.47158)	0.18150	1.76673
10	10	12	(0.50000, 0.50309)	(0.54676, 1.44581)	(1.44851, 1.27181)	0.21096	1.89596
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
169	1	1	(1.01305, 1.00000)	(1.00000, 1.00000)	(0.61520, -0.24241)	0.00187	0.01305
170	1	1	(1.01305, 1.00000)	(1.00000, 1.00000)	(0.61566, -0.24060)	0.00187	0.01305
171	1	1	(1.01305, 1.00000)	(1.00000, 1.00000)	(0.61611, -0.23879)	0.00187	0.01305
172	1	1	(1.01305, 1.00000)	(1.00000, 1.00000)	(0.61656, -0.23698)	0.00187	0.01305
173	1	1	(1.01305, 1.00000)	(1.00000, 1.00000)	(0.61702, -0.23517)	0.00187	0.01305
174	1	1	(1.01305, 1.00000)	(1.00000, 1.00000)	(0.61747, -0.23336)	0.00187	0.01305
175	8	1	(1.00272, 1.00000)	(1.00000, 1.00000)	(0.61756, -0.23298)	0.01033	0.00272
176	1	1	(1.00272, 1.00000)	(1.00000, 1.00000)	(0.61766, -0.23260)	0.00039	0.00272

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 176 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((1, 1), (1, 1))$, pois o critério de parada foi satisfeito em 0.00039.

(b) Escolhendo a distância proximal φ -divergência, definida por:

$$d_0(u, v) := \sum_{i=1}^n v_i \varphi\left(\frac{u_i}{v_i}\right),$$

com $\varphi(t) := t - \log t - 1$, e usando o algoritmo (PMAPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	(x_1^k, x_2^k)	(z_1^k, z_2^k)	(y_1^k, y_2^k)	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	11	10	(0.90057, 1.82430)	(2.72728, 1.94857)	(1.20654, 1.37169)	0.42522	3.59957
2	9	9	(0.78764, 1.62955)	(2.41932, 1.89377)	(1.37238, 1.65400)	0.32742	3.15500
3	10	8	(0.67210, 1.42782)	(2.09611, 1.83674)	(1.49580, 1.84320)	0.32820	2.68857
4	11	9	(0.56278, 1.23000)	(1.77640, 1.77856)	(1.57749, 1.94116)	0.32497	2.22218
5	9	9	(0.50000, 1.04496)	(1.47616, 1.72021)	(1.62129, 1.95906)	0.30586	1.74133
6	7	8	(0.50000, 0.90850)	(1.20645, 1.66247)	(1.63491, 1.91595)	0.27582	1.46042
7	8	22	(0.50000, 0.78749)	(1.00000, 1.60774)	(1.62390, 1.82443)	0.21358	1.32025
8	9	12	(0.50000, 0.68295)	(0.82875, 1.55259)	(1.59183, 1.69231)	0.17991	1.54090
9	9	12	(0.50000, 0.59493)	(0.68839, 1.49931)	(1.54207, 1.52668)	0.17295	1.71598
10	8	12	(0.50000, 0.52266)	(0.57682, 1.44815)	(1.47778, 1.33416)	0.20297	1.84867
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
66	20	1	(1.00000, 0.89828)	(1.00000, 1.00000)	(0.60832, -0.37705)	0.01572	0.10172
67	15	1	(1.00000, 0.91359)	(1.00000, 1.00000)	(0.60233, -0.38605)	0.01531	0.08641
68	17	1	(1.00000, 0.93059)	(1.00000, 1.00000)	(0.59751, -0.39327)	0.01700	0.06941
69	9	3	(1.00016, 0.94775)	(1.00000, 1.00000)	(0.59389, -0.39869)	0.01716	0.05241
70	9	18	(1.00062, 0.96533)	(1.00069, 1.00000)	(0.59155, -0.40209)	0.01759	0.03597
71	12	1	(1.00118, 0.98319)	(1.00069, 1.00000)	(0.59048, -0.40356)	0.01787	0.01868
72	18	1	(1.00206, 1.00000)	(1.00069, 1.00000)	(0.59060, -0.40315)	0.01683	0.00275
73	6	1	(1.00243, 1.00000)	(1.00069, 1.00000)	(0.59073, -0.40270)	0.00048	0.00312

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 73 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((1, 1), (1, 1))$, pois o critério de parada foi satisfeito em 0.00048.

(c) Escolhendo a distância proximal homogênea de segundo ordem, definida por:

$$d_0(u, v) := \sum_{i=1}^n v_i^2 \varphi\left(\frac{u_i}{v_i}\right),$$

com $\varphi(t) := \mu(-\log t + t - 1) + \frac{\nu}{2}(t - 1)^2$.

Para $\mu = \nu = 1$, usando o algoritmo (PMAPD), obtemos os resultados apresentados na seguinte tabela.

k	$N[x^k]$	$N[z^k]$	(x_1^k, x_2^k)	(z_1^k, z_2^k)	(y_1^k, y_2^k)	$ q^k - q^{k-1} _{\bar{M}}$	$f(x^k) + g(z^k)$
1	14	13	(0.93172, 1.91072)	(2.87750, 1.97417)	(1.22494, 1.41107)	0.46859	3.83066
2	8	9	(0.84517, 1.80447)	(2.72968, 1.94597)	(1.42826, 1.77343)	0.41550	3.63495
3	9	9	(0.74366, 1.68392)	(2.56057, 1.91569)	(1.60691, 2.07980)	0.35466	3.41652
4	9	9	(0.63090, 1.55199)	(2.37455, 1.88365)	(1.75846, 2.32452)	0.28785	3.17930
5	17	9	(0.51090, 1.41173)	(2.17624, 1.85019)	(1.88120, 2.50357)	0.21708	2.92727
6	7	10	(0.50000, 1.26629)	(1.97034, 1.81567)	(1.97797, 2.63024)	0.20877	2.55230
7	7	10	(0.50000, 1.11765)	(1.75974, 1.78034)	(2.04859, 2.70491)	0.21355	2.15772
8	7	11	(0.50000, 0.98929)	(1.54805, 1.74450)	(2.09437, 2.72948)	0.21470	1.80326
9	7	9	(0.50000, 0.86204)	(1.33862, 1.70844)	(2.11553, 2.70447)	0.21251	1.68501
10	9	11	(0.50000, 0.73814)	(1.13485, 1.67241)	(2.11269, 2.63120)	0.20692	1.56913
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
111	14	1	(1.05743, 1.00000)	(1.00000, 1.00000)	(0.77571, -0.24287)	0.00933	0.05743
112	15	1	(1.04776, 1.00000)	(1.00000, 1.00000)	(0.77737, -0.23624)	0.00966	0.04776
113	11	1	(1.03788, 1.00000)	(1.00000, 1.00000)	(0.77869, -0.23098)	0.00988	0.03788
114	15	1	(1.02796, 1.00000)	(1.00000, 1.00000)	(0.77966, -0.22710)	0.00991	0.02796
115	25	1	(1.01794, 1.00000)	(1.00000, 1.00000)	(0.78028, -0.22461)	0.01002	0.01794
116	19	1	(1.00787, 1.00000)	(1.00000, 1.00000)	(0.78055, -0.22352)	0.01007	0.00787
117	21	1	(1.00000, 1.00000)	(1.00000, 1.00000)	(0.78055, -0.22352)	0.00787	0.00000
118	1	1	(1.00000, 1.00000)	(1.00000, 1.00000)	(0.78055, -0.22352)	0.00000	0.00000

Na tabela notamos o seguinte: Em cada iteração os subproblemas foram resolvidos em poucas iterações, isso podemos notar nas colunas $N[x^k]$ e $N[z^k]$, e notamos também que o algoritmo parou em 118 iterações com uma solução aproximada à solução ótima $(x^*, z^*) = ((1, 1), (1, 1))$, pois o critério de parada foi satisfeito em 0.00000.

Neste exemplo podemos ressaltar a seguinte diferença, usando a distância de Bregman o programa levou 176 iterações para se aproximar à solução do problema enquanto a distância ϕ -divergência levou 73 iterações e a distância proximal homogênea de segundo ordem levou 118 iterações.

Capítulo 5

Conclusões e Trabalhos Futuros

Nesta dissertação, propõe-se um método multiplicador proximal inexato (PMAPD), usando distâncias proximais regularizadas, para resolver problemas de minimização convexas com uma estrutura separável da forma:

$$(PC) \quad \min\{f(x) + g(z) : Ax + Bz = b, x \in \bar{C}, z \in \bar{K}\}.$$

O método proposto é uma extensão do método Multiplicador Preditor-Corretor Proximal (PCPM) e do Método Multiplicador Preditor-Corretor Proximal Não Linear (NPCPMM) dados por Chen-Teboulle e Kyono-Fukushima respectivamente, e inclui a classe das distâncias φ -divergência, que para nosso conhecimento não foram desenvolvidos ainda.

A extensão é baseada na distância proximal dada na Seção 2.2. Notamos também que o Método de Decomposição Proximal Entrópico (EPDM) é um caso particular do nosso método quando consideramos iterações exatas e usamos a distância log-quadrática regularizada.

Estabelecemos, sob as hipóteses (\mathcal{H}_1) - (\mathcal{H}_4) , que as iterações geradas pelo método são bem definidas e a sequência converge para uma solução ótima do problema.

Além disso, se adicionalmente as hipóteses (\mathcal{T}_1) - (\mathcal{T}_4) satisfazem-se, obtivemos que a taxa de convergência da sequência gerada pelo PMAPD é linear. Também mostramos alguns resultados computacionais comparando a funcionalidade de algumas distâncias proximais conhecidas na literatura.

Uma futura pesquisa pode incluir encontrar uma variante do método (PMAPD) e poder resolver problemas mais gerais, tais como, otimização quase-convexa, programação semidefinida e desigualdades variacionais.

Referências Bibliográficas

- [1] Assad, A. A., 1978, “Multicommodity network flows: a survey”, *Networks*, v. 8, n. 1, pp. 37–91.
- [2] Auslender, A., Teboulle, M., 2006, “Interior gradient and proximal methods for convex and conic optimization”, *SIAM Journal on Optimization*, v. 16, n. 3, pp. 697–725.
- [3] Auslender, A., Teboulle, M., 2001, “Entropic proximal decomposition methods for convex programs and variational inequalities”, *Mathematical programming*, v. 91, n. 1, pp. 33–47.
- [4] Auslender, A., Teboulle, M., 2005, “Interior projection-like methods for monotone variational inequalities”, *Mathematical programming*, v. 104, n. 1, pp. 39–68.
- [5] Auslender, A., Teboulle, M., Ben-Tiba, S., 1999, “Interior proximal and multiplier methods based on second order homogeneous kernels”, *Mathematics of Operations Research*, v. 24, n. 3, pp. 645–668.
- [6] Bertsekas, D. P., 1982, “Constrained optimization and Lagrange multiplier methods”, *Computer Science and Applied Mathematics, Boston: Academic Press, 1982*, v. 1.
- [7] Bertsekas, D. P., Tsitsiklis, J. N., 1989, *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc.
- [8] Chen, G., Teboulle, M., 1994, “A proximal-based decomposition method for convex minimization problems”, *Mathematical Programming*, v. 64, n. 1–3, pp. 81–101.
- [9] Chen, G., Teboulle, M., 1993, “Convergence analysis of a proximal-like minimization algorithm using Bregman functions”, *SIAM Journal on Optimization*, v. 3, n. 3, pp. 538–543.

- [10] Chifflet, J., Mahey, P., Reynier, V., 1994, “Proximal decomposition for multicommodity flow problems with convex costs”, *Telecommunication Systems*, v. 3, n. 1, pp. 1–10.
- [11] Gabay, D., 1983, “Chapter ix applications of the method of multipliers to variational inequalities”, *Studies in mathematics and its applications*, v. 15, pp. 299–331.
- [12] Gabay, D., Mercier, B., 1976, “A dual algorithm for the solution of nonlinear variational problems via finite element approximation”, *Computers & Mathematics with Applications*, v. 2, n. 1, pp. 17–40.
- [13] Goldfarb, D., Ma, S., Scheinberg, K., 2013, “Fast alternating linearization methods for minimizing the sum of two convex functions”, *Mathematical Programming*, v. 141, n. 1-2, pp. 349–382.
- [14] Kiwiel, K. C., 1997, “Proximal minimization methods with generalized Bregman functions”, *SIAM journal on control and optimization*, v. 35, n. 4, pp. 1142–1168.
- [15] Kleinrock, L., 2007, *Communication nets: Stochastic message flow and delay*. Courier Dover Publications.
- [16] Kyono, M., Fukushima, M., 2000, “Nonlinear proximal decomposition method for convex programming”, *Journal of optimization theory and applications*, v. 106, n. 2, pp. 357–372.
- [17] Lenoir, A., 2008, *Modèles et algorithmes pour la planification de production à moyen terme en environnement incertain*. Tese de Doutorado, Université Blaise Pascal-Clermont-Ferrand II.
- [18] Mahey, P., Oualibouch, S., Tao, P. D., 1995, “Proximal decomposition on the graph of a maximal monotone operator”, *SIAM Journal on Optimization*, v. 5, n. 2, pp. 454–466.
- [19] Mahey, P., Ouorou, A., LeBlanc, L., et al., 1998, “A new proximal decomposition algorithm for routing in telecommunication networks”, *Networks*, v. 31, n. 4, pp. 227–238.
- [20] Ouorou, A., Mahey, P., Vial, J.-P., 2000, “A survey of algorithms for convex multicommodity flow problems”, *Management Science*, v. 46, n. 1, pp. 126–147.
- [21] Polëïiak, B., 1987, *Introduction to optimization*. Optimization Software, Publications Division (New York).

- [22] Rockafellar, R. T., 1976, “Augmented Lagrangians and applications of the proximal point algorithm in convex programming”, *Mathematics of operations research*, v. 1, n. 2, pp. 97–116.
- [23] Rockafellar, R. T., 1997, *Convex analysis*, v. 28. Princeton university press.
- [24] Rockafellar, R. T., 1976, “Monotone operators and the proximal point algorithm”, *SIAM Journal on Control and Optimization*, v. 14, n. 5, pp. 877–898.
- [25] Sarmiento, C. O., Papa Quiroz, E. A., Oliveira, P. R., 2014, “A proximal multiplier method for separable convex minimization”. Submitted for publication in the journal *Optimization*, July.
- [26] Spingarn, J. E., 1983, “Partial inverse of a monotone operator”, *Applied mathematics and optimization*, v. 10, n. 1, pp. 247–265.
- [27] Spingarn, J. E., 1985, “Applications of the method of partial inverses to convex programming: decomposition”, *Mathematical Programming*, v. 32, n. 2, pp. 199–223.
- [28] Stefanov, S. M., 2007, “Minimization of a strictly convex separable function subject to a convex inequality constraint or linear equality constraints and bounds on the variables”, *Sci. Res*, v. 5, pp. 1–10.
- [29] Teboulle, M., 1997, “Convergence of proximal-like algorithms”, *SIAM Journal on Optimization*, v. 7, n. 4, pp. 1069–1083.