



NANCY BAYGORREA CUSIHUALLPA

**ALGORITMO DO PONTO PROXIMAL PARA
OPERADORES NÃO MONÓTONOS**

CAMPINAS

2013



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA
E COMPUTAÇÃO CIENTÍFICA

NANCY BAYGORREA CUSIHUALLPA

ALGORITMO DO PONTO PROXIMAL PARA OPERADORES NAO
MONOTONOS

Orientador: Prof. Dr. Roberto Andreani

Dissertação de mestrado apresentada ao Instituto de
Matemática, Estatística e Computação Científica da Unicamp para
obtenção do título de Mestra em matemática aplicada.

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO
DEFENDIDA PELA ALUNA NANCY BAYGORREA CUSIHUALLPA,
E ORIENTADA PELO PROF. DR. ROBERTO ANDREANI.

Assinatura do Orientador

A handwritten signature in black ink, appearing to be "R. Andreani", written over a horizontal line.

CAMPINAS
2013

FICHA CATALOGRÁFICA ELABORADA POR
MARIA FABIANA BEZERRA MULLER - CRB8/6162
BIBLIOTECA DO INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E
COMPUTAÇÃO CIENTÍFICA - UNICAMP

Baygorrea Cusihuallpa, Nancy, 1982-
B342a Algoritmo do ponto proximal para operadores não monótonos /
Nancy Baygorrea Cusihuallpa. – Campinas, SP : [s.n.], 2013.

Orientador: Roberto Andreani.
Dissertação (mestrado) – Universidade Estadual de Campinas,
Instituto de Matemática, Estatística e Computação Científica.

1. Operadores monótonos. 2. Dualidade (Matemática). 3.
Algoritmos. 4. Programação não-linear. 5. Análise variacional. I.
Andreani, Roberto, 1961-. II. Universidade Estadual de Campinas.
Instituto de Matemática, Estatística e Computação Científica. III.
Título.

Informações para Biblioteca Digital

Título em inglês: Proximal point algorithm for non-monotone operators

Palavras-chave em inglês:

Monotone operators

Duality theory (Mathematics)

Algorithms

Nonlinear programming

Variational analysis

Área de concentração: Matemática Aplicada

Titulação: Mestra em Matemática Aplicada

Banca examinadora:

Roberto Andreani [Orientador]

Paulo José da Silva e Silva

Roberto Cristóvão Mesquita Silva

Data de defesa: 20-02-2013

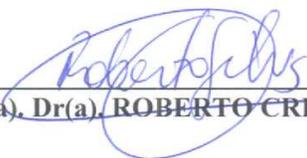
Programa de Pós-Graduação: Matemática Aplicada

Dissertação de Mestrado defendida em 20 de fevereiro de 2013 e aprovada

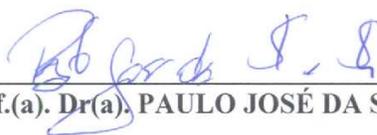
Pela Banca Examinadora composta pelos Profs. Drs.



Prof.(a). Dr(a). ROBERTO ANDREANI



Prof.(a). Dr(a). ROBERTO CRISTÓVÃO MESQUITA SILVA



Prof.(a). Dr(a). PAULO JOSÉ DA SILVA E SILVA

Ao meu filho Lucas Luciano,
a maior fonte de motivação na minha vida.

*“Mesmo as noites sem estrelas,
podem anunciar a aurora de uma grande realização ”*

Martin Luther King

Agradecimentos.

Para começar, dividir o tempo entre maternidade e o mestrado foi complicado demais, e muitos dos que me conheceram em Campinas sabem disso. Por isso, agradecer a todos aqueles que ajudaram a construir esta dissertação e colaboraram, quer em termos acadêmicos quer em termos pessoais, não é tarefa fácil. Em primeiro lugar, gostaria de agradecer ao meu orientador Roberto Andreani pela orientação e tempo, pelo ter acreditado em mim desde o começo pelos seus conselhos, por ter me paciência e me compreender na fase pela qual eu estava passando. Ao professor Paulo Jose da Silva e Silva, sem ele este trabalho não teria sido possível, pelo seu tempo, suas sugestões e observações feitas que contribuíram para a melhoria desta dissertação. Ao professor Roberto Cristovão, por ter participado da banca examinadora.

O meu maior agradecimento vai pra minha família, meus pais e minha irmã, Felix, Claudia e Irma, por terem sido o contínuo apoio em todos estes anos, sempre estiveram comigo, a pesar da distância. Também, a minha família campineira adotiva: os Souza, os levarei sempre em meu coração.

Aos meus amigos, os de lá e os de aqui, que não citarei os nomes porque é uma lista muito longa, mas que foram muito importante neste periodo da minha vida me dando a força necessária em todos os momentos difíceis e os momentos bons que compartilhamos, todos inesquecíveis. Aos meus colegas do mestrado, a todos os professores do instituto, aos funcionarios do Imecc-Unicamp, ao pessoal da biblioteca do Instituto, todos eles sempre disponíveis para ajudar em todo o que precisar. Por último deste grupo mas não menos importante, quero agradecer ao pessoal da SAE pela ajuda proporcionada e ao pessoal do DEdic-UNICAMP pela vaga concedida na creche pro meu bebezinho. Minha profunda gratidão a todos vocês!

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq, pela bolsa concedida durante os anos do curso.

Para finalizar, a todas as pessoas que direta ou indiretamente me ajudaram a chegar até aqui.

Resumo.

Esta dissertação desenvolve um estudo detalhado da convergência local do método de ponto proximal para resolver o problema de encontrar zeros de operadores maximais sem a condição de monotonicidade. Em particular, é estudado a convergência dos métodos de multiplicadores proximais para resolver problemas de otimização não linear sem a condição de convexidade.

Para obter os resultados desejados apresentaremos ferramentas de análise variacional para substituir a condição de monotonicidade maximal do operador como também, a teoria de dualidade generalizada para a aplicação do método de multiplicadores proximais.

Apresentamos também uma aplicação do algoritmo do ponto proximal aos métodos dos multiplicadores para uma classe de problemas gerais baseados num esquema de dualidade generalizada.

Palavras chaves: Operadores monótonos maximais, dualidade generalizada, algoritmo do ponto proximal, métodos multiplicadores, k -Localização Lipschitz, λ -regularização Yosida, λ -hipomonotonicidade.

Abstract.

In this dissertation we will develop a detailed study of local convergence of proximal point method for finding a root of a maximal operators without monotonicity. In particular, it is studied the convergence for proximal method of multipliers by solving nonlinear optimization problems without convexity conditions.

In order to obtain the desired results we will study some variational analysis tools to replace maximal monotonicity condition of operators as well as general duality theory which is treated to study an application to proximal method of multipliers.

Also, we show an application of the proximal point algorithm to the multipliers methods for a class of problems which is based in general duality scheme.

Key words: Maximal monotone operators, general duality, proximal point algorithm, multiplier methods, k -Lipschitz Localization, λ -Yosida regularization, λ -hypomonotone.

Sumário

Introdução	1
1 Noções Preliminares	5
1.1 Alguns resultados de análise convexa	5
1.2 Condições de otimalidade para o problemas (<i>PNL</i>)	12
1.3 Dualização geral.	18
2 Operadores monótonos maximais.	23
2.1 Operadores monótonos	24
2.2 Dualidade com monotonicidade maximal	44
3 O algoritmo do ponto proximal (APP)	47
3.1 O Algoritmo do Ponto Proximal Relaxado (APPR)	54
3.2 O APP com monotonicidade local.	64
3.3 O APP sem monotonicidade.	70
4 Aplicação do APP aos métodos de multiplicadores proximais	75
4.1 Métodos dos multiplicadores	76
4.2 Metodo multiplicadores proximais	83
4.3 Aplicação dos métodos multiplicadores proximais para problemas (PNL)	86
5 Conclusões	93
Referencias Bibliográficas	95

Introdução

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

$$\min_{x \in \mathbb{R}^n} f(x). \quad (0.0.1)$$

Segundo as condições de otimalidade e utilizando o subdiferencial da função f (no caso em que a função não for diferenciável), podemos escrever o problema (0.0.1) da seguinte forma:

$$\text{Encontrar } \bar{x} \in \mathbb{R}^n, \text{ tal que } 0 \in \partial f(\bar{x}). \quad (0.0.2)$$

Cabe notar que o subdiferencial de uma função própria, semi-contínua inferior e convexa é um operador monótono maximal, então ao substituímos o subdiferencial de f por qualquer outro operador monótono maximal T obtemos uma extensão natural do problema (0.0.2), isto é, dado um operador multivalorado monótono maximal $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, o problema é encontrar algum $z^* \in \mathbb{R}^n$ tal que

$$0 \in T(z^*). \quad (0.0.3)$$

O método mais conhecido para resolver esse tipo de problemas é o Algoritmo do Ponto Proximal (APP), o qual foi proposto por Martinet[23] em 1970 como um método de regularização no contexto de otimização convexa em espaços de Hilbert e desenvolvido por Rockafellar[36] para encontrar zeros de operadores monótonos maximais.

Segundo Rockafellar[36, pp. 878], o APP para resolver o problema de encontrar zeros de operadores multivalorados monótonos maximais em espaços de Hilbert é baseado no

fato de que para cada $z \in \mathcal{H}$ e $c > 0$, existe um único $u \in \mathcal{H}$ tal que $z - u \in cT(u)$, isto é,

$$z \in (I + cT)(u).$$

Assim, O APP gera, para qualquer ponto arbitrário $z^0 \in \mathcal{H}$, uma sequência $\{z^k\} \subset \mathcal{H}$ pela regra aproximativa:

$$z^{k+1} \approx J_{c_k}^T(z^k), \quad (0.0.4)$$

onde $J_{c_k}^T = (I + c_k T)^{-1}$ é chamado o resolvente do operador T e $\{c_k\}$ é uma sequência de números reais positivos.

Este algoritmo é interessante por muitas razões, dentre elas pelo papel que cumpre em determinados métodos computacionais baseados em dualidade tais como os métodos de multiplicadores e o método de multiplicadores proximais, os quais serão estudados nesta dissertação.

Além disso, considerando que o conjunto solução do problema (0.0.3), $Zer(T)$, seja não vazio e que as iterações geradas pelo algoritmo sejam limitadas tais que

$$\|z^{k+1} - J_{c_k}^T(z^k)\| \leq \epsilon_k, \quad \text{com} \quad \sum_{k=0}^{\infty} \epsilon_k < \infty$$

e $\{c_k\}$ uma sequência não decrescente limitada inferiormente, pode-se garantir a convergência fraca.

A convergência forte do algoritmo é garantida se $\{z^k\}$ é limitada, $\{c_k\}$ é uma sequência não decrescente tal que $c_k \uparrow \bar{c} \leq +\infty$ e T^{-1} é a -Lipschitz contínua (ver Rockafellar[36, Teorema 2]), ou se T é fortemente monótono com módulo $a > 0$ sem necessidade de $\{z^k\}$ ser limitada (ver Rockafellar[36, Proposição 5]). Mas a hipótese de T ser fortemente monótona é muito restritivo pois exclue aplicações importantes tais como os problemas típicos de otimização convexa.

Dentre das diversas variantes consideráveis que teve o APP clássico ao longo do tempo, existe um algoritmo resultante da combinação do APP clássico de Rockafellar com um resultado de Gol'shtein e Tret'yakov[14] que considera as iterações dadas por

$$z^{k+1} = (\rho_k J_c^T + (1 - \rho_k)I)(z^k),$$

onde em comparação com o algoritmo de Rockafellar[36], o método introduzido por Gol'stein e Tret'yakov[14] não permitiram o parâmetro c variar em cada passo k e não consideraram o caso no qual $Zer(T) = \emptyset$, mas sim permitiram que os resolventes do operador T foram achados aproximadamente. Por conseguinte, combinando esses dois algoritmos obtemos um algoritmo mais geral chamado o algoritmo do ponto proximal relaxado (APPR), o qual foi desenvolvido por Eckstein e Bertsekas[9], onde, a partir de um ponto arbitrário $z^0 \in \mathcal{H}$, é gerada uma sequência dada por

$$z^{k+1} = (\rho_k J_{c_k}^T + (1 - \rho_k)I)(z^k),$$

com $\rho_k \in (0, 2)$ como o parâmetro de relaxação.

Observe que se considerarmos $\rho_k = 1$ no APPR obtemos o APP clássico, ademais, quando $\rho_k \in (0, 1)$ e $\rho_k \in [1, 2)$ temos os métodos sub-relaxado e sobre-relaxado, respectivamente. Note que z^{k+1} é uma combinação linear do operador resolvente de T com o operador identidade, aplicados em z^k .

Por outro lado, Rockafellar[32] mostrou como os métodos dos multiplicadores, ou chamados também, método lagrangeano aumentado, para minimização convexa pode ser obtido como um caso particular do APP clássico. Além disso, é bem conhecido que o método multiplicador clássico para minimização muitas vezes tem convergência, ainda que a função não seja convexa. Mais ainda, Eckstein e Ferris[10] aplicaram com sucesso o seu método multiplicador a desigualdades variacionais não monótonas, embora a análise de convergência desse método fosse baseada em monotonicidade. As pesquisas anteriores sugerem que a monotonicidade não é necessária para garantir convergência local do APP, isso motiva a estudar condições sobre um operador multivalorado arbitrário T a fim de aplicar o APPR e garantir a convergência local. Vale a pena mencionar que Pennanen[29] obteve a convergência local do APP sem monotonicidade baseados em ferramentas de análise variacional tais como: a λ -regularização Yosida de T , definida por

$$T_\lambda = (\lambda I + T^{-1})^{-1},$$

no qual, a importância desse operador vem do fato que $Zer(T) = Zer(T_\lambda)$, onde $Zer(T_\lambda)$ denota o conjunto de soluções do problema (0.0.3) trocando T por T_λ . Além disso, T_λ

pode ser localmente monótona mesmo quando T não o for. Isso vai ocorrer quando T^{-1} for um operador localmente λ -hipomonótono, ou seja, quando para algum $\lambda > 0$ o operador T^{-1} acrescentado o operador λI for monótono. Neste caso, é possível aplicar o APPR ao novo operador T_λ .

Outro conceito muito importante para o desenvolvimento desta dissertação é a k -Localização Lipschitz de um operador. De fato, se tivermos que T^{-1} tem uma k -Localização Lipschitz num ponto, então podemos obter a monotonicidade maximal numa vizinhança desse ponto de T_λ , para algum $\lambda \geq k$. Então, com a hipótese de T^{-1} ser k -Localização Lipschitz num ponto, podemos aplicar o APP ao operador maximal T_λ , para algum $\lambda \geq k$.

Nesta dissertação, estudaremos os resultados apresentados por Pennanen[29] restritos ao espaço euclidiano \mathbb{R}^n detalhando os mesmos através de uma abordagem didática, com a apresentação de vários exemplos. Além disso, daremos ênfase ao estudo do APP sobre-relaxado pelo fato de que a taxa de convergência deste método tem melhores propriedades.

No Capítulo 1 desta dissertação daremos as noções preliminares de convexidade, uma breve teoria de dualização geral também é apresentada. No Capítulo 2, apresentaremos resultados da teoria de operadores monótono maximais com uma grande variedade de exemplos. No Capítulo 3, a partir do APPR de Eckstein e Bertsekas[9] para aplicações monótonas maximais, vamos mostrar que a monotonicidade global da aplicação pode ser substituída por uma monotonicidade local sem modificar o comportamento local do APPR. Depois, com condições mais fracas, daremos os resultados do APP sobre-relaxado para operadores multivalorados arbitrários. No Capítulo 4, mostramos uma aplicação do APP aos métodos dos multiplicadores baseado num esquema de dualidade generalizada, em particular, para os métodos de multiplicadores proximais para programação não linear sem assumir a condição de complementariedade estrita. Usando resultados do APP sem monotonicidade, vamos mostrar a convergência local linear dos métodos proximais dos multiplicadores, os quais têm melhores propriedades teóricas de convergência que os métodos dos multiplicadores clássicos.

Capítulo 1

Noções Preliminares

A finalidade deste capítulo é facilitar a leitura desta dissertação fornecendo alguns resultados que são necessários para o desenvolvimento dos demais capítulos. Apresentaremos algumas definições e resultados básicos de análise convexa, de análise variacional e dualidade os quais serão de grande importância para provar a convergência dos algoritmos.

Vamos denotar a bola aberta centrada em \bar{x} com raio δ como $B(\bar{x}, \delta)$, quando $\bar{x} = 0$ denotaremos simplesmente por $B(\delta)$. O interior, o fecho e a fronteira de um conjunto K denotaremos por $\text{int}(K)$, \bar{K} e $\text{Front}(K)$, respectivamente. Para mais detalhes dos resultados apresentados neste capítulo revisar referências tais como Izmailov[20], Bertsekas[6], Hiriart-Urruty[18, Vol. I-II], Rockafellar e Wets[38], Burachik[8], Auslender[4], entre outros.

1.1 Alguns resultados de análise convexa

Denotaremos por $\overline{\mathbb{R}} := [-\infty, +\infty]$ à reta real estendida completa.

As regras aritméticas são estendidas para incluir:

$$\begin{aligned} +\infty + \infty &= +\infty, & \alpha \cdot \pm\infty &= \pm\infty \quad \text{para} \quad \alpha > 0, & \inf \emptyset &= \infty, & \sup \emptyset &= -\infty, \\ -\infty - \infty &= -\infty, & 0 \cdot (\pm\infty) &= 0, & +\infty - (+\infty) &= \textit{indeterminado}. \end{aligned}$$

Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, o domínio efetivo de f é o conjunto dado como

$$\text{dom}(f) := \{x \in \mathbb{R}^n \mid f(x) < +\infty\}.$$

Uma função f é dita própria se $f(x) < +\infty$ para algum $x \in \text{dom}(f)$ e $f(x) > -\infty$ para todo $x \in \text{dom}(f)$, ou equivalentemente, se $\text{dom}(f)$ é um conjunto não vazio no qual f é uma função finita. Caso contrário, vamos chamar f de imprópria.

Dizemos que uma função $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ é semi-contínua inferior em $\bar{x} \in \mathbb{R}^n$ se para todo $\epsilon > 0$ existe $\delta > 0$ tal que para $x \in B(\bar{x}, \delta)$ temos $f(\bar{x}) \leq f(x) + \epsilon$, e é dita semi-contínua inferior em \mathbb{R}^n se é semi-contínua inferior em cada $x \in \mathbb{R}^n$. É simples ver que f é semi-contínua inferior em $\bar{x} \in \mathbb{R}^n$ se, e somente se, $\liminf_{x \rightarrow \bar{x}} f(x) \geq f(\bar{x})$.

Além disso, uma função própria f , é dita convexa quando para quaisquer $x, y \in \mathbb{R}^n$, e todo $\alpha \in (0, 1)$, temos

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

Seja $K \subset \mathbb{R}^n$, a função $\delta_K : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ definida por

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

é chamada função indicadora de K , o qual é semi-contínua inferior em \mathbb{R}^n se, e somente se, K é um conjunto fechado; é convexa se, e somente se, K é um conjunto convexo.

A regularização λ -Moreau-Yosida de uma função convexa é definida como

$$f_\lambda(x) = \inf_y \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\},$$

com $\lambda > 0$. Há abundante literatura para estudarmos propriedades da regularização Moreau-Yosida, podemos referenciar, por exemplo Fukushima[12] e Hiriart[18, Vol. II].

O ponto mínimo da função f_λ é denotado por

$$\text{prox}_{\lambda, f}(x) = \arg \min \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 : y \in \mathbb{R}^n \right\},$$

a aplicação $\text{prox}_{\lambda, f}$ é chamado o operador proximal e $\text{prox}_{\lambda, f}(x)$ é o ponto proximal de x com respeito à função f . Quando $\lambda = 1$, o ponto mínimo de f_λ será denotado por $\text{prox}_f(x)$, isto é,

$$\text{prox}_f(x) = \arg \min \left\{ f(y) + \frac{1}{2} \|y - x\|^2 : y \in \mathbb{R}^n \right\}.$$

Lema 1.1.1 [38, Teorema 2.26,b] *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, semi-continua e convexa e $\lambda > 0$. A função f_λ é finita em \mathbb{R}^n e diferenciável com gradiente dado como*

$$\nabla f_\lambda(x) = \frac{1}{2\lambda} (x - \text{prox}_{\lambda, f}(x)). \quad (1.1.1)$$

Consideremos K um conjunto convexo e fechado, a λ -regularização Moreau-Yosida da função indicadora de K e o ponto proximal de x com respecto dessa função são dados por

$$(\delta_K)_\lambda(x) = \frac{1}{2\lambda} d_K^2(x) \quad (1.1.2)$$

$$\text{prox}_{\lambda, \delta_K}(x) = P_K(x), \quad (1.1.3)$$

respectivamente, onde d_K é a função distância sobre o conjunto K definida por $d_K(x) = \inf_w f(w, x)$ tal que $f(w, x) = \|w - x\| + \delta_K(w)$ e P_K denota o operador projeção sobre K . Tal projeção desse ponto também pode ser denotado como

$$P_K(x) = \{z \in K \mid \|z - x\| \leq \|y - x\| \ \forall y \in K\}, \quad (1.1.4)$$

o qual é equivalente solucionar o seguinte problema:

$$(P1) \quad \min \{ \|x - y\| : y \in K \}.$$

Observe que a solução do problema (P1) também é solução do problema

$$(P2) \quad \min \left\{ \delta_K(y) + \frac{1}{2} \|x - y\|^2 \right\} = \min \left\{ \frac{1}{2} \|x - y\|^2 : y \in K \right\},$$

Observação 1.1.2 *A vantagem de trabalhar com o problema (P2) está na diferenciabilidade da função objetivo deste problema cujo gradiente é $x - y$, isto ajuda nas demonstrações da parte teórica da análise variacional. Daí, vamos considerar o problema (P2) quando trabalhar com projeções.*

Teorema 1.1.3 [20, Teorema 3.2.4] *Seja $K \subset \mathbb{R}^n$ um conjunto convexo e fechado. Então para todo $x \in \mathbb{R}^n$, a projeção de x sobre K , denotado por $P_K(x)$ existe e é única. Além disso, $\bar{x} = P_K(x)$ se, e somente se,*

$$\bar{x} \in K, \quad \langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \quad \forall y \in K$$

Podemos observar, dos resultados acima, que o fato de K ser fechado faz que exista a projeção de um ponto sobre o conjunto K . Além disso, a convexidade de K é a propriedade adicional que faz de P_K ser uma função ou operador simples valorado.

Seja f uma função própria e convexa. Dizemos que $s \in \mathbb{R}^n$ é um subgradiente de f no ponto $x \in \mathbb{R}^n$ se,

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

O conjunto de todos os subgradientes da função f em $x \in \mathbb{R}^n$ é chamado o subdiferencial de f em x , e é denotado como ∂f , isto é,

$$\partial f(x) = \{s \in \mathbb{R}^n \mid f(y) \geq f(x) + \langle s, y - x \rangle, \quad \forall y \in \mathbb{R}^n\}. \quad (1.1.5)$$

O conceito de subdiferenciabilidade foi introduzido nos anos 60 por Moreau[25], Rockafellar[34] entre outros.

Exemplo 1.1.4 Definimos a aplicação $h : \mathbb{R}^n \rightarrow [0, \infty]$ como

$$h(x) = \begin{cases} 0, & \|x\| \leq r, \\ +\infty, & \|x\| > r. \end{cases}$$

onde o subdiferencial da função h , ∂h , é dado por:

$$\partial h(x) = \begin{cases} 0, & \|x\| < r, \\ \alpha x, \alpha \geq 0 & \|x\| = r, \\ +\infty, & \|x\| > r. \end{cases}$$

Observe que esta função é própria e convexa pois é a função indicadora da bola $B(r)$.

Em seguida, vamos mostrar a relação existente entre o subdiferencial e minimizadores globais. A condição de otimalidade de Fermat afirma que para uma função própria e diferenciável $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, uma condição necessária para \bar{x} ser um mínimo local da função f é que a gradiente de f em \bar{x} seja igual a 0, isto é, $\nabla f(\bar{x}) = 0$. Esta condição vai ser suficiente se considerarmos f uma função convexa. Ademais, este resultado pode ser estendido para problemas mais gerais tais como para o caso não diferenciável onde a derivada será trocada pelo conceito de subdiferenciabilidade como daremos no seguinte teorema.

Teorema 1.1.5 [18, Teorema 2.2.1, Vol. I] *Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ uma função própria e convexa. O ponto $\bar{x} \in \mathbb{R}^n$ é mínimo da função f se, e somente se, $0 \in \partial f(\bar{x})$.*

Introduzimos o conceito de função conjugada, originalmente desenvolvido por Fenchel[11], muito útil no contexto de dualidade geral. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}^n \cup \{\infty\}$ uma função própria e convexa. A função $f^* : \mathbb{R}^n \rightarrow \mathbb{R}^n \cup \{\infty\}$ definida por

$$f^*(x) = \sup_w \{ \langle x, w \rangle - f(w) \},$$

é chamada a função conjugada da função f . Dizemos que a função $f^{**} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é a função conjugada de (f^*) .

Em seguida, alguns resultados que envolvem a conjugada de uma função o que vão ser necessários para nosso trabalho.

Proposição 1.1.6 [6, Proposição 7.1.1] *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, convexa e semicontínua inferior. Então, $f = f^{**}$.*

Teorema 1.1.7 [18, Teorema 1.4.1, Vol. II] *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, convexa e semi-contínua inferior. Então, $s \in \partial f(x)$ se, e somente se, $f^*(s) = \langle x, s \rangle - f(x)$.*

Proposição 1.1.8 [38, Proposição 11.3] *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função própria, convexa e semi-contínua inferior. Então, se tem que $(\partial f)^{-1} = \partial f^*$.*

Lema 1.1.9 *Seja f uma função própria, convexa e semi-contínua e $x, y, z \in \mathbb{R}^n$ temos que $z = x + y$ e $f(x) + f^*(y) = \langle x, y \rangle$ se, e somente se, $x = \text{prox}_f(z)$ e $y = \text{prox}_{f^*}(z)$.*

Prova. Consideremos $f(x) + f^*(y) = \langle x, y \rangle$. Da Proposição 1.1.6 e do Teorema 1.1.7 temos que $x \in \partial f^*(y)$. Do fato que $z = x + y$ obtemos que $0 \in \partial f^*(y) + (y - z)$, o qual implica que $y = \text{prox}_{f^*}(z)$. Analogamente obtemos que $x = \text{prox}_f(z)$.

Reciprocamente, seja $x = \text{prox}_f(z)$ então $0 \in \partial f(x) + (x - z)$, isto é, existe $y_0 \in \partial f(x)$ tal que

$$z = y_0 + x. \tag{1.1.6}$$

Além disso, da Proposição 1.1.6 e do Teorema 1.1.7, é válido que

$$f^*(y_0) = \langle x, y_0 \rangle - f(x) = \langle x, y_0 \rangle - (f^*)^*(x).$$

De novo, do Teorema 1.1.7 temos que $x \in \partial f^*(y_0)$ se, e somente se, $0 \in \partial f^*(y_0) + (y_0 - z)$, isto implica que $y_0 = \text{prox}_{f^*}(z)$. Daí, pela hipótese temos que $y = y_0$. Logo, de (1.1.6) temos que $z = y + x$ com

$$f^*(y) = \langle x, y \rangle - f(x). \quad \blacksquare$$

Por outro lado, um conjunto $K \subset \mathbb{R}^n$ é chamado de cone se $\alpha x \in K$ sempre que $x \in K$ e $\alpha > 0$.

Definição 1.1.10 *Seja $K \subset \mathbb{R}^n$ e $\bar{x} \in K$. Dizemos que $v \in \mathbb{R}^n$ é um vetor normal ao conjunto K em \bar{x} , no sentido regular, e denotado por $v \in \hat{N}_K(\bar{x})$, se*

$$\langle v, x - \bar{x} \rangle \leq o(\|x - \bar{x}\|), \quad \forall x \in K$$

onde $o(\|x - \bar{x}\|)$ satisfaz

$$\lim_{\substack{x \rightarrow \bar{x} \\ x \neq \bar{x}}} \frac{o(\|x - \bar{x}\|)}{\|x - \bar{x}\|} = 0.$$

Definição 1.1.11 *Seja $K \subset \mathbb{R}^n$ e $\bar{x} \in K$. Dizemos que $v \in \mathbb{R}^n$ é um vetor normal ao conjunto K , no sentido geral, e denotado por $v \in N_K(\bar{x})$, se existem seqüências $x^k \rightarrow \bar{x}$ e $v^k \rightarrow v$ tal que $v^k \in \hat{N}_K(x^k)$ com $\{x^k\} \subset K$.*

Definição 1.1.12 *Dizemos que um vetor $w \in \mathbb{R}^n$ é tangente ao conjunto K em $\bar{x} \in K$, denotado por $w \in T_K(\bar{x})$, se existem seqüências $\{x^k\} \subset K$ com $x^k \rightarrow \bar{x}$ e $\{\tau_k\}$ com $\tau_k \searrow 0$ tal que*

$$\frac{x^k - \bar{x}}{\tau_k} \rightarrow w.$$

Proposição 1.1.13 [38, Proposição 6.5] *Seja $K \subset \mathbb{R}^n$ não vazio e $\bar{x} \in K$, os conjuntos $\hat{N}_K(\bar{x})$ e $N_K(\bar{x})$ são fechados, convexos e caracterizados por*

$$v \in \hat{N}_K(\bar{x}) \iff \langle v, w \rangle \leq 0, \quad \forall w \in T_K(\bar{x}). \quad (1.1.7)$$

Além disso,

$$\hat{N}_K(\bar{x}) \subset \limsup_{x \rightarrow \bar{x}} \hat{N}_K(x) = N_K(\bar{x}). \quad (1.1.8)$$

Teorema 1.1.14 [38, Teorema 6.9] *Seja $K \subset \mathbb{R}^n$ um conjunto convexo. Então o cone normal a K é dado por*¹

$$N_K(x) = \begin{cases} \{s \in \mathbb{R}^n \mid \langle s, y - x \rangle \leq 0, \forall y \in K\}, & \text{se } x \in K, \\ \emptyset, & \text{se } x \notin K. \end{cases} \quad (1.1.9)$$

Definição 1.1.15 *O operador cone polar de $K \subset \mathbb{R}^n$ é definido por*

$$K^\otimes = \{s \in \mathbb{R}^n \mid \langle s, x \rangle \leq 0, \forall x \in K\}.$$

Lema 1.1.16 [38, Exemplo 11.4] *Seja $K \subset \mathbb{R}^n$ um cone convexo e fechado temos que $(\delta_K)^* = \delta_{K^\otimes}$. Além disso,*

$$N_K^{-1}(x) = N_{K^\otimes}(x). \quad (1.1.10)$$

Lema 1.1.17 *Para qualquer cone convexo, fechado e não vazio K temos*

$$\frac{1}{2}(d_K(u))^2 = \sup_{y \in K^\otimes} \left\{ \langle u, y \rangle - \frac{1}{2} \|y\|^2 \right\}.$$

Prova. Definimos

$$g(u) = \frac{1}{2}(d_K(u))^2,$$

onde d_K é a função distância sobre o conjunto K . Então, a conjugada de g é dado por

$$\begin{aligned} g^*(u) &= \sup_w \left\{ \langle w, u \rangle - \frac{1}{2} \inf_{v \in K} \|v - w\|^2 \right\} \\ &= \sup_w \left\{ \langle w, u \rangle - \frac{1}{2} \inf_v (\|v - w\|^2 + \delta_K(v)) \right\} \\ &= \sup_w \left\{ \langle w, u \rangle + \frac{1}{2} \sup_v \{-\|v - w\|^2 - \delta_K(v)\} \right\} \\ &= \sup_{w,v} \left\{ \langle w, u \rangle - \frac{1}{2} \|v - w\|^2 - \delta_K(v) \right\} \\ &= \sup_{w,v} \left\{ \left(\langle w - v, u \rangle - \frac{1}{2} \|u\|^2 - \frac{1}{2} \|w - v\|^2 \right) + \frac{1}{2} \|u\|^2 + \langle v, u \rangle - \delta_K(v) \right\} \end{aligned} \quad (1.1.11)$$

¹Em Burachik[8, Definição 3.6.1] foi definido o operador cone N_K , para K subconjunto arbitrário, como é dado em (1.1.9), o qual não sempre é válido, na verdade é apenas para conjuntos convexos. Por exemplo: sejam $S = \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\} \cup \{(0,0)\}$ e $\bar{x} = (0,0)$. Se considerarmos como foi definido em Burachik[8] temos que $N_S(\bar{x}) = 0$. Por outro lado, da Definição 1.1.12 temos que $T_S(\bar{x}) = 0$, e segundo (1.1.7) temos que $\hat{N}_S(\bar{x}) = \mathbb{R}^2$. Então, de (1.1.8), temos que $N_S(\bar{x}) = \mathbb{R}^2$.

Observe que é imediato verificar que

$$\langle w - v, u \rangle - \frac{1}{2} \|u\|^2 - \frac{1}{2} \|w - v\|^2 = -\frac{1}{2} \|w - v - u\|^2. \quad (1.1.12)$$

Substituindo (1.1.12) em (1.1.11), a conjugada de g fica como

$$\begin{aligned} g^*(u) &= \sup_{w,v} \left\{ -\frac{1}{2} \|w - (v + u)\|^2 + \langle v, u \rangle - \delta_K(v) \right\} + \frac{1}{2} \|u\|^2 \\ &= \sup_v \left\{ \sup_w \left\{ -\frac{1}{2} \|w - (v + u)\|^2 \right\} + \langle v, u \rangle - \delta_K(v) \right\} + \frac{1}{2} \|u\|^2 \\ &= \sup_v \{ \langle v, u \rangle - \delta_K(v) \} + \frac{1}{2} \|u\|^2 \\ &= \delta_K^*(u) + \frac{1}{2} \|u\|^2, \end{aligned}$$

onde o ótimo é atingido quando $w = v + u$. Além disso, como d_K é convexo pois K é convexo e fechado. Então, da Proposição 1.1.6 e do Lema 1.1.16 temos que

$$g(u) = \sup_y \{ \langle u, y \rangle - g^*(y) \} = \sup_y \{ \langle u, y \rangle - (\delta_{K^\otimes}(y) + \frac{1}{2} \|y\|^2) \} = \sup_{y \in K^\otimes} \left\{ \langle u, y \rangle - \frac{1}{2} \|y\|^2 \right\}. \quad \blacksquare$$

1.2 Condições de otimalidade para o problemas (PNL)

Seja $C \subseteq \mathbb{R}^n$ e uma função $f : C \rightarrow \overline{\mathbb{R}}$, o problema de encontrar um minimizador de f no conjunto C é expresso como

$$(P3) \quad \min \{ f(x) : x \in C \}.$$

Se $C = \mathbb{R}^n$, se diz que o problema de otimização é irrestrito.

Dizemos que $\bar{x} \in C$ é um minimizador global do problema (P3) se

$$f(\bar{x}) \leq f(x), \quad \forall x \in C. \quad (1.2.1)$$

Dizemos que $\bar{x} \in C$ é um minimizador local de (P3) se existe uma bola $B(\bar{x}, \epsilon)$, tal que

$$f(x) \geq f(\bar{x}), \quad \forall x \in B(\bar{x}, \epsilon). \quad (1.2.2)$$

Se para $x \neq \bar{x}$ a desigualdade (1.2.1) ou (1.2.2) é estrita, se diz que \bar{x} é um minimizador estrito global ou local, respectivamente.

Além disso, da análise convexa dizemos que um problema tem formulação convexa quando no problema o conjunto $C \subset \mathbb{R}^n$ é convexo e a função f é convexa. Nesse caso as condições necessárias de otimalidade são também condições suficientes de otimalidade.

Por outro lado, no caso de otimização com restrições de igualdade e desigualdade é necessário introduzir os denominados multiplicadores de Lagrange para obter uma condição análoga ao caso de minimizadores sem restrições.

Consideremos o problema de otimização não linear

$$(PNL) \left\{ \begin{array}{l} \min f_0(x), \\ s.a. \\ f_i(x) = 0, \quad i = 1, \dots, r \\ f_i(x) \leq 0, \quad i = r + 1, \dots, m \end{array} \right.$$

para funções reais $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ com $f_i \in \mathcal{C}^2$ para $i = 0, \dots, m$.

Dizemos que \bar{x} é um ponto estacionário (ou ponto *KKT*) do problema (PNL) se existem escalares $\bar{y}_i, i = 1, \dots, m$ com $\bar{y}_i \geq 0$ para $i = r + 1, \dots, m$ tais que

$$\begin{aligned} \nabla f_0(\bar{x}) + \sum_{i=1}^r \bar{y}_i \nabla f_i(\bar{x}) + \sum_{j=r+1}^m \bar{y}_j \nabla f_j(\bar{x}) &= 0, \\ f_i(\bar{x}) &= 0, \quad i = 1, \dots, r \\ f_i(\bar{x}) &\leq 0, \quad i = r + 1, \dots, m \\ \bar{y}_i f_i(\bar{x}) &= 0, \quad i = r + 1, \dots, m. \end{aligned}$$

Estas condições são conhecidas como condições Karush-Kuhn-Tucker (KKT) e as componentes de $\bar{y} = (\bar{y}_1, \dots, \bar{y}_m)$ são chamados os Multiplicadores de Lagrange e a última igualdade dada acima no sistema de equações é chamada “condição de complementariedade”.

A única dificuldade em relação às condições *KKT* é que nem todos os minimizadores locais as satisfazem. Para acontecer isso é preciso que o ponto cumpra alguma condição nas restrições. Tal condição é chamada condição de qualificação. Uma condição de qualificação é uma propriedade sobre os pontos do conjunto viável que, quando satisfeita por

um minimizador local do problema, implica que se satisfazem as condições KKT nesse ponto.

Existem várias condições de qualificação conhecidas na literatura: regularidade (o qual é a mais utilizada e também conhecido como independência linear dos gradientes das restrições de igualdade e das restrições ativas), Mangasarian-Fromovitz[22], Slater[39], dependência linear positiva constante[2], quase-normalidade[16], quase-regularidade conhecida também como a condição de Abadie[1] sendo esta uma das condições em PNL mais fracas.

Do fato que algumas condições de qualificações são caracterizadas por conjuntos linearmente independente positivos e linearmente dependente positivos é que há a necessidade de definir esses termos, os quais são similares ao conceito de conjunto linearmente independente e dependente de vetores.

Consideremos $A = \{a_1, \dots, a_r\}$ e $B = \{b_1, \dots, b_l\}$ dois conjuntos finitos em \mathbb{R}^n tal que $A \cap B \neq \emptyset$. Dizemos que o par (A, B) é linearmente dependente positivo se existem $\alpha \in \mathbb{R}^r$ e $\beta \in \mathbb{R}^l$ tais que $\beta \geq 0$, $(\alpha, \beta) \neq 0$ e

$$\sum_{i=1}^r \alpha_i a_i + \sum_{j=1}^l \beta_j b_j = 0.$$

Dizemos que o par (A, B) é linearmente independente positivo se o sistema nas variáveis $\alpha \in \mathbb{R}^r$ e $\beta \in \mathbb{R}^l$ com $\beta \geq 0$

$$\sum_{i=1}^r \alpha_i a_i + \sum_{j=1}^l \beta_j b_j = 0$$

tem unicamente a solução nula.

Para o problema (PNL) em questão, definimos $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e o conjunto $K \subset \mathbb{R}^m$ por

$$H(x) = (f_1(x), \dots, f_m(x)) \quad (1.2.3)$$

$$K = \{0\}^r \times \mathbb{R}_-^{m-r}, \quad (1.2.4)$$

onde o conjunto viável para esse problema é dado por $C = \{x \in \mathbb{R}^n \mid H(x) \in K\}$.

Denotemos a matriz Jacobiana de H no ponto x como $DH(x)$, onde em termos de

$H(x)$ para $x = (x_1, \dots, x_n)$, a jacobiana de H é a matriz

$$DH(x) = \left[\frac{\partial f_i}{\partial x_j}(x) \right]_{i,j=1}^{m,n} \in \mathbb{R}^{m \times n},$$

onde a matriz transposta $DH(x)^T$ tem como colunas às gradientes $\nabla f_i(x)$ para $i = 1, \dots, m$ tal que

$$DH(x)^T y = y_1 \nabla f_1(x) + \dots + y_m \nabla f_m(x)$$

para $y = (y_1, \dots, y_m)$.

Seja (\bar{x}, \bar{y}) um ponto satisfazendo a condição (KKT) para o problema (PNL). Definimos os conjuntos de índices

$$I_1(\bar{x}) = \{1, \dots, r\} \cup \{i \in [r+1, \dots, m] \mid f_i(\bar{x}) = 0, \bar{y}_i > 0\},$$

$$I_2(\bar{x}) = \{i \in [r+1, \dots, m] \mid f_i(\bar{x}) = 0, \bar{y}_i = 0\},$$

$$I_3(\bar{x}) = \{i \in [r+1, \dots, m] \mid f_i(\bar{x}) < 0, \bar{y}_i = 0\}.$$

Nesta dissertação vamos considerar a condição de Mangasarian-Fromovitz para o problema (PNL).

Definição 1.2.1 *Seja \bar{x} um ponto viável de (PNL). Dizemos que a condição de qualificação de Mangasarian-Fromovitz é satisfeita em $\bar{x} \in C$ quando $\{\nabla f_i(\bar{x}), i = 1, \dots, r\}$ é um conjunto linearmente independente e existe $d \in \mathbb{R}^n$ tal que*

$$\nabla f_i(\bar{x})^T d = 0, \quad i = 1, \dots, r$$

$$\nabla f_i(\bar{x})^T d < 0, \quad \forall i \in (I_1(\bar{x}) \cup I_2(\bar{x})) \setminus \{1, \dots, r\}.$$

A seguir apresentamos uma caracterização dual equivalente à condição de Mangasarian-Fromovitz.

Proposição 1.2.2 [20, Proposição 4.1.2] *Seja \bar{x} um ponto viável de (PNL). O ponto \bar{x} satisfaz a condição de Mangasarian-Fromovitz se, e somente se, o sistema*

$$\sum_{i=1}^r d_i \nabla f_i(\bar{x}) + \sum_{i \in (I_1(\bar{x}) \cup I_2(\bar{x})) \setminus \{1, \dots, r\}} d_i \nabla f_i(\bar{x}) = 0$$

com $d_i \geq 0$ para $i \in (I_1(\bar{x}) \cup I_2(\bar{x})) \setminus \{1, \dots, r\}$, tem apenas a solução nula.

Por outro lado, segundo o Teorema 1.1.14, o operador cone normal do conjunto $K = \{0\}^r \times \mathbb{R}_-^{m-r}$ para o problema (PNL) é dado por

$$\begin{aligned} N_K(H(\bar{x})) &= \{v \in \mathbb{R}^m \mid v_i \in \mathbb{R}, i = 1, \dots, r, \text{ e } v_j \geq 0, j = r+1, \dots, m\} \\ &= \mathbb{R}^r \times \mathbb{R}_+^{m-r} \end{aligned}$$

Logo, considerando (\bar{x}, \bar{y}) um ponto KKT e como as restrições inativas não influenciam uma vez que os seus multiplicadores $\lambda_i = 0$, para $i \in I_3(\bar{x})$ e do fato que $N_K(H(\bar{x})) = \mathbb{R}^r \times \mathbb{R}_+^{m-r}$ é que podemos caracterizar a condição de qualificação de Mangasarian-Fromovitz para o problema (PNL) em termos do cone normal do conjunto K , o qual vai ser utilizado ao longo desta dissertação, como é dado a continuação: Um ponto KKT (\bar{x}, \bar{y}) satisfaz a condição de qualificação de Mangasarian-Fromovitz se

$$\bar{y} \in N_K(H(\bar{x})), \quad DH(\bar{x})^T \bar{y} = 0 \implies \bar{y} = 0.$$

Como $\bar{y} \in N_K(H(\bar{x}))$, $H(\bar{x}) \in N_{K^\otimes}^{-1}(\bar{y})$ e do Lema 1.1.16 temos

$$H(\bar{x}) \in N_{K^\otimes}(\bar{y}), \tag{1.2.5}$$

com K^\otimes como o cone polar de $K = \{0\}^r \times \mathbb{R}_-^{m-r}$ dado por $K^\otimes = \mathbb{R}^r \times \mathbb{R}_+^{m-r}$.

De (1.2.5), o sistema KKT do problema (PNL) é dado como

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \nabla f_0(x) + DH(x)^T y \\ -H(x) + N_{K^\otimes}(y) \end{pmatrix}. \tag{1.2.6}$$

Além disso, este sistema KKT pode ser escrito, mais convenientemente, como uma equação generalizada dada por

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \nabla_x \mathcal{L}(x, y) \\ -H(x) \end{pmatrix} + \partial \delta_{\mathbb{R}^n \times K^\otimes} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{1.2.7}$$

onde $\delta_{\mathbb{R}^n \times K^\otimes}$ é a função indicadora sobre $\mathbb{R}^n \times K^\otimes$, \mathcal{L} é a função Lagrangiana do problema (PNL) dada por $\mathcal{L}(x, y) = f_0(x) + H(x)^T y$ e $y \in \mathbb{R}^r \times \mathbb{R}_-^{m-r}$ são os multiplicadores de Lagrange com respeito às funções f_i para $i = 1, \dots, m$. Observe que \mathcal{L} e H são diferenciáveis do fato que $f_i \in \mathcal{C}^2$ com $i = 0, 1, \dots, m$.

Por outro lado, precisamos algumas ferramentas para tentar procurar alguma condição que garanta o bom comportamento das soluções do problema não linear (1.2.7). Tal condição vai ser chamada de “regularidade forte” ou equivalentemente, a propriedade de “Localização Lipschitz”, propriedade muito importante a qual vai ser utilizado na última seção do Capítulo 4 nesta dissertação. Com este propósito, a seguir são enunciados conceitos e resultados necessários para o desenvolvimento desta dissertação, veja para mais detalhes Robinson[31].

Seja $K \subset \mathbb{R}^n$ um conjunto convexo e fechado e $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função diferenciável em \bar{x} . Consideremos que a equação generalizada

$$0 \in f(x) + \partial\delta_K(x). \quad (1.2.8)$$

Suponhamos que a equação generalizada dada em (1.2.8) tem uma solução em \bar{x} . Definimos a linearização do problema (1.2.8) dado por

$$T(x) = f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + \partial\delta_K(x). \quad x \in \mathbb{R}^n$$

Definição 1.2.3 Dizemos que (1.2.8) é fortemente regular em \bar{x} com constante k -Lipschitz associado, se existem vizinhanças $U \ni 0$ e $V \ni \bar{x}$ tal que para $u \in U$ a função

$$\tau : u \mapsto T^{-1}(u) \cap V \quad (1.2.9)$$

é simples valorada e k -Lipchitz contínua em U . Mais ainda, se existir funções como definidas em (1.2.9) satisfazendo as condições dadas para τ , dizemos que a inversa do operador $T : X \rightarrow Y$, T^{-1} , tem uma k -Localização Lipschitz em $(\bar{x}, 0)$.

Definição 1.2.4 Dizemos que o par KKT (\bar{x}, \bar{y}) satisfaz a condição suficientemente forte de segunda ordem para o problema (PNL) se

$$\langle w, \nabla_{xx}^2 \mathcal{L}(\bar{x}, \bar{y})w \rangle > 0$$

para todo $w \neq 0$ tal que $\langle \nabla f_i(\bar{x}), w \rangle = 0$, para todo $i \in I_1(\bar{x})$.

Proposição 1.2.5 [29] *A condição suficientemente forte de segunda ordem é válido se, e somente se, existe $c \in \mathbb{R}$ tal que*

$$\langle w, \nabla_{xx}^2 \mathcal{L}(\bar{x}, \bar{y})w \rangle + c \sum_{i \in I_1(\bar{x})} \langle \nabla f_i(\bar{x}), w_i \rangle^2 > 0, \quad \forall w \neq 0. \quad (1.2.10)$$

Teorema 1.2.6 [31, Teorema 4.1] *Consideremos o problema de otimização não linear (PNL). Suponha que (\bar{x}, \bar{y}) soluciona o sistema KKT de (PNL) dado por (1.2.7), que a condição suficientemente forte de segunda ordem é válido em (\bar{x}, \bar{y}) e que as gradientes das restrições ativas sejam linearmente independentes. Então, (1.2.7) é fortemente regular em (\bar{x}, \bar{y}) .*

1.3 Dualização geral.

Consideremos o problema de minimização

$$(\mathcal{P}) \quad \min\{f_0(x) : x \in \mathbb{R}^n\}$$

onde f_0 é uma função convexa real estendida.

Assumimos que uma parametrização para f_0 , $f(x, y)$, com u como uma perturbação, tenha sido especificada de alguma forma tal que

$$f_0(x) = f(x, 0),$$

para alguma função convexa real estendida f .

Considere (\mathcal{P}) o problema primal associado com f . Segundo a análise convexa, o problema primal associado com f é resolver o problema (\mathcal{P}) .

Além disso, o Lagrangeano associado com f é a função convexa-côncava, l , isto é, convexo com respeito à variável x e côncava com respeito à variável u e é definido por Rockafellar[33, página 18] por

$$l(x, z) := \inf_y \{f(x, y) + \langle y, z \rangle\},$$

onde o problema Lagrangeano é encontrar um ponto de sela (\bar{x}, \bar{z}) de l com respeito à minimização em x e maximização em z .

Por outro lado, a função g definida por

$$\begin{aligned} g(w, z) &= \inf_x \{l(x, z) - \langle x, w \rangle\} \\ &= \inf_{x,y} \{f(x, y) + \langle y, z \rangle - \langle x, w \rangle\} \\ &= -\sup_{x,y} \{\langle x, w \rangle + \langle y, -z \rangle - f(x, y)\} \\ &= -f^*(w, z), \end{aligned}$$

resulta uma família parametrizada de funções côncavas dadas por

$$g_w(z) := g(w, z),$$

onde o problema dual é encontrar um maximizador \bar{z} do dual objetivo definido por

$$g_0(z) = g(0, z).$$

Por conseguinte, os problemas primal, Lagrangeano e dual são respectivamente equivalentes às inclusões

$$0 \in \partial f_0(x), \tag{1.3.1}$$

$$(0, 0) \in \partial l(x, z), \tag{1.3.2}$$

$$0 \in \partial g_0(z). \tag{1.3.3}$$

Cabe lembrar que a teoria de dualidade em otimização convexa tem as suas raízes na transformada de Legendre-Fenchel, constituindo assim o contexto de dualidade conjugada de Rockafellar[33, 34].

Vale a pena mencionar também que uma grande quantidade de teoria e algoritmos para programação convexa são construídos sob a noção de dualidade. A teoria de dualidade que vamos apresentar nesta seção foi desenvolvida por Pennanen[27, 28], o qual está baseada no conceito de inversas parciais de aplicações sobre espaços produto e projeções.

Sejam X, Y, W e Z conjuntos arbitrários de diferentes dimensões no espaço euclidiano. Denotamos a aplicação projeção sobre X por $P_X : X \times Y \rightarrow X$. Consideremos as aplicações $P_{1,X,Y} : X \rightarrow X \times Y$ e $P_{2,X,Y} : X \rightarrow Y \times X$ tais que $P_{1,X,Y} = (x, 0)$ e $P_{2,X,Y} = (0, x)$ para todo $x \in X$. Também, consideremos uma aplicação $F : X \times Y \rightarrow W \times Z$, as inversas

parciais dessa aplicação, denotado como $F^{(-1,1)} : W \times Y \rightrightarrows X \times Z$ e $F^{(1,-1)} : X \times Z \rightrightarrows W \times Y$, são dadas por

$$\begin{aligned} F^{(-1,1)}(w, y) &= \{(x, z) \mid (w, z) \in F(x, y)\}, \\ F^{(1,-1)}(x, z) &= \{(w, y) \mid (w, z) \in F(x, y)\}. \end{aligned}$$

Isto significa que

$$(w, y) \in F^{(-1,1)}(x, z) \iff (w, z) \in F(x, y) \iff (x, z) \in F^{(1,-1)}(w, y). \quad (1.3.4)$$

Então, observamos que

$$\begin{aligned} (F^{(-1,1)})^{(-1,1)} &= (F^{(1,-1)})^{(1,-1)} = F \\ (F^{(-1,1)})^{(1,-1)} &= (F^{(1,-1)})^{(-1,1)} = F^{-1}. \end{aligned}$$

Mais geralmente, podemos observar que $(F^{(i,j)})^{(k,l)} = F^{(ik,jl)}$, para todo $i, j, k, l \in \{-1, 1\}$, onde $F^{(1,1)} = F$ e $F^{(-1,-1)} = F^{-1}$.

A idéia em generalizar um esquema de dualidade em otimização convexa é expressar o problema primal, lagrangeano e dual dados em (1.3.1), (1.3.2) e (1.3.3), respectivamente, por aplicações conjuntos valoradas mais gerais.

Definição 1.3.1 *Seja $F_0 : X \rightrightarrows W$ e suponhamos que uma parametrização de F_0 , $F : X \times Y \rightrightarrows W \times Z$, tenha sido especificada como*

$$F_0 = P_W \circ F \circ P_{1,X,Y}.$$

O lagrangeano correspondente a F é definido como $L = F^{(1,-1)}$.

Definimos também $G = F^{-1}$ e $G_0 = P_Y \circ G \circ P_{2,Z,W}$.

Os problemas

$$0 \in F_0(x), \quad (1.3.5)$$

$$(0, 0) \in L(x, z), \quad (1.3.6)$$

$$0 \in G_0(z), \quad (1.3.7)$$

são chamados o problema da inclusão primal, lagrangeano e dual, respectivamente.

A parametrização F pode ser vista como a descrição de uma família de aplicações $F_y : X \rightarrow W$ dada como

$$F_y(x) = P_W(F(x, y)),$$

onde cada família corresponde a um problema primal perturbado dado como

$$0 \in F_y(x).$$

Dualmente, a aplicação G dá lugar à família de aplicações: $G_w : Z \rightarrow Y$ definidas como

$$G_w(z) = P_Y(G(w, z)),$$

onde cada família corresponde a um problema dual perturbado

$$0 \in G_w(z).$$

Na verdade, para um operador F_0 dado, existem muitas formas de introduzir uma parametrização F , tal que $F_0 = P_W \circ F \circ P_{1,X,Y}$. Qualquer de tais parametrizações F define a mesma inclusão primal, mas as formas do lagrangeano e inclusões duais dependem crucialmente da escolha particular. Então, qualquer das aplicações F , L ou G são suficientes para definir os tres problemas unicamente desde que os outros dois possam ser obtidos por inversão parcial. De fato, de (1.3.4) e como $L = F^{(1,-1)}$ e $G = F^{-1}$ temos que

$$(w, z) \in F(x, y) \iff (w, y) \in L(x, z) \iff (x, y) \in G(w, z). \quad (1.3.8)$$

Daí, temos que

$$\begin{aligned} F_0(x) &= \{w \mid \exists z \in Z, (x^*, 0) \in L(x, z)\}, \\ G_0(z) &= \{y \mid \exists x \in X, (0, y) \in L(x, z)\}. \end{aligned} \quad (1.3.9)$$

O resultado seguinte corresponde ao Teorema Karush-Kun-Tucker (KKT) generalizado em programação convexa.

Teorema 1.3.2 [26, Teorema 2.7] *Os conjuntos soluções de (1.3.5) e (1.3.7) são as projeções do conjunto (1.3.6) sobre X e Z , respectivamente. Em particular, cada um dos conjuntos solução é não-vazio se, e somente se, os outros dois são não vazios.*

Capítulo 2

Operadores monótonos maximais.

O propósito deste Capítulo é dar uma introdução à teoria de operadores monótonos maximais em espaços euclidianos, os quais são úteis, entre outros, para analisar e resolver alguns métodos de otimização, em particular, o método do ponto proximal o qual será analisado nesta dissertação.

Desde que o subdiferencial de uma função, ∂f , associe para cada vetor $x \in \mathbb{R}^n$, não exatamente um vetor de \mathbb{R}^n , mas um subconjunto do \mathbb{R}^n , é que precisamos estender as noções de aplicações simples valoradas (aplicações ponto-a-ponto) à teoria de aplicações multivaloradas, o qual vai ser utilizada, se não dizer o contrário, em toda nesta dissertação

Se diz que T é uma aplicação multivalorada em \mathbb{R}^n se para todo $x \in \mathbb{R}^n$, $T(x) \subseteq \mathbb{R}^n$. Este tipo de aplicações é denotado como $T : \mathbb{R}^n \rightarrow 2^{\mathbb{R}^n}$, onde $2^{\mathbb{R}^n}$ denota a família de todos os subconjuntos não vazios de \mathbb{R}^n , ou também $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$.

O grafo do operador T é definido por $graf T := \{(x, y) \mid y \in T(x)\}$. Vale a pena mencionar que é natural identificar o operador T com o seu grafo, isto é, $graf T = T$.

Definimos o domínio, a imagem e o inverso de T como

$$\begin{aligned} dom(T) &= \{x \in \mathbb{R}^n \mid T(x) \neq \emptyset\}, \\ Im(T) &= \{y \in \mathbb{R}^n \mid \exists x \in \mathbb{R}^n, y \in T(x)\} = \bigcup \{T(x), x \in \mathbb{R}^n\}, \\ T^{-1}(y) &:= \{x \in \mathbb{R}^n \mid y \in T(x)\}, \end{aligned}$$

respectivamente. O operador inverso é algumas vezes chamado na literatura como a pseudo-inversa pois não tem a necessidade que o operador T corresponda a uma função

injetiva, ou equivalentemente, que T^{-1} seja simples valorada. Observe que, segundo a definição de inversa dada acima, $(x, y) \in T$ (ou equivalentemente $y \in T(x)$) se, e somente se, $(y, x) \in T^{-1}$. De fato, seja $(x, y) \in T$ se, e somente se, $y \in T(x)$, isto é, $x \in T^{-1}(y)$ se, e somente se, $(y, x) \in T$.

Além disso, a inclusão de dois operadores, denotado por $T_1 \subset T_2$, é tal que $T_1(x) \subset T_2(x)$ para todo $x \in \mathbb{R}^n$.

2.1 Operadores monótonos

Definição 2.1.1 *Se diz que o operador T é monótono se para todo $x, x' \in \mathbb{R}^n$,*

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

Definição 2.1.2 *Se diz que o operador T é monótono maximal se é monótono e dado qualquer operador monótono \bar{T} tal que $T \subset \bar{T}$, temos que $T = \bar{T}$.*

Observação 2.1.3 *Seja T monótono maximal. É válido os seguintes enunciados:*

- a. T^{-1} é monótono maximal.
- b. λT é monótono maximal para qualquer $\lambda > 0$.
- c. $T + d = \{(x, y + d) \mid (x, y) \in T\}$ é monótono maximal.

Vamos verificar as três condições necessárias dadas nessa observação. A verificação da condição suficiente é análoga.

Primeiro verificamos o item a. Sejam $(y, x), (y', x') \in T^{-1}$ então

$$\langle y - y', x - x' \rangle = \langle x - x', y - y' \rangle \geq 0,$$

onde a última desigualdade é obtida pela monotonicidade de T . Logo, T^{-1} é monótono. Logo, consideremos um operador monótono S tal que $T^{-1} \subseteq S$. É válido que se $T^{-1} \subseteq S$ então $T \subseteq S^{-1}$. De fato, seja $(x, y) \in T$ o qual é equivalente ter que $(y, x) \in T^{-1}$. Isso significa que $(y, x) \in S$ se, e somente se, $(x, y) \in S^{-1}$. Como S^{-1} é monótono e $T \subset S^{-1}$ e da maximalidade de T temos que $T = S^{-1}$, isto é, $T^{-1} = S$. Portanto, T^{-1} é um

operador monótono maximal.

Segundo, verificamos o item b. Sejam $(x, \frac{y}{\lambda}), (x', \frac{y'}{\lambda}) \in T$. Como $(x, \frac{y}{\lambda}) \in T$ temos que $(x, y) \in \lambda T$. Analogamente, temos que $(x', y') \in \lambda T$ e da monotonicidade de T obtemos

$$0 \leq \left\langle x - x', \frac{y}{\lambda} - \frac{y'}{\lambda} \right\rangle = \frac{1}{\lambda} \langle x - x', y - y' \rangle.$$

Como $\lambda > 0$, temos $0 \leq \langle x - x', y - y' \rangle$. Logo, λT é monótono. Seja S qualquer operador monótono tal que $\lambda T \subseteq S$, isto é, $T \subseteq \frac{1}{\lambda} S$. Como S é monótona, $\frac{1}{\lambda} S$ é também monótona. Logo, da maximalidade de T , temos que $T = \frac{1}{\lambda} S$, isto é, $\lambda T = S$. Logo, λT é maximal.

Terceiro, verificamos o item c., sejam $(x, y + d), (x', y' + d) \in T + d$ então

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

onde a desigualdade é obtida do fato que $(x, y), (x', y') \in T$ e T é monótono. Para verificar a maximalidade do operador $T + d$, vamos considerar um operador monótono S tal que $T + d \subseteq S$, isto é, $T \subseteq S - d$. Da maximalidade de T temos que $T = S - d$. Logo, $T + d = S$. Obtendo assim a maximalidade do operador $T + d$.

A seguir apresentamos alguns exemplos de operadores monótonos maximais.

Exemplo 2.1.4 Seja T um operador monótono maximal tal que $\text{dom}(T) = \{a\}$ para algum $a \in \mathbb{R}$. Daí, a única forma de definir o operador T está dado por

$$T(x) = \begin{cases} (-\infty, +\infty), & \text{se } x = a \\ \emptyset, & \text{se } x \neq a \end{cases}$$

Exemplo 2.1.5 Seja $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ própria, semi-contínua inferior e convexa. Temos que o operador subdiferencial de f , ∂f , é monótono maximal.

De fato, sejam $s \in \partial f(x)$ e $t \in \partial f(y)$ o que implica que

$$f(y) - f(x) \geq \langle s, y - x \rangle, \quad \text{para todo } y \in \mathbb{R}^n, \quad (2.1.1)$$

$$f(x) - f(y) \geq \langle t, x - y \rangle, \quad \text{para todo } x \in \mathbb{R}^n. \quad (2.1.2)$$

Somando (2.1.1) e (2.1.2),

$$\langle t, y - x \rangle \geq \langle s, y - x \rangle.$$

Então, $\langle t - s, y - x \rangle \geq 0$. Logo, como t e s foram escolhidos arbitrariamente, obtemos que ∂f é um operador monótono.

Logo, provaremos a maximalidade desse operador. Seja T monótono tal que $\partial f \subset T$. Vamos provar que $T \subset \partial f$. Seja $(x_0, y_0) \in T$ e consideremos x_1 e y_1 tais que

$$x_1 = \text{prox}_f(x_0 + y_0) \quad e \quad y_1 = \text{prox}_{f^*}(x_0 + y_0),$$

onde f^* é a função conjugada de f . Do Lema 1.1.9 temos que

$$x_0 + y_0 = x_1 + y_1, \tag{2.1.3}$$

e $f^*(y_1) = \langle x_1, y_1 \rangle - f(x_1)$. Do Teorema 1.1.7 obtemos que $y_1 \in \partial f(x_1)$, mas $\partial f \subset T$. Logo, $(x_1, y_1) \in T$. Daí, da monotonicidade de T , temos que $\langle x_0 - x_1, y_0 - y_1 \rangle \geq 0$. Logo, de (2.1.3) obtemos que

$$0 \leq \langle x_0 - x_1, x_1 - x_0 \rangle = -\|x_0 - x_1\|^2,$$

o que implica que $x_1 = x_0$. Então, de (2.1.3) temos que $y_0 = y_1$, mas $y_1 \in \partial f(x_1) = \partial f(x_0)$ pois $x_1 = x_0$. Logo, $(x_0, y_0) \in \partial f$ e do fato que (x_0, y_0) foram escolhidos arbitrariamente temos que $T \subset \partial f$. Portanto, ∂f é um operador monótono maximal.

Exemplo 2.1.6 Um caso particular do Exemplo 2.1.5 temos quando f é a função indicadora. Seja $K \subset \mathbb{R}^n$ um conjunto fechado e convexo o que implica que δ_K é uma função semi-contínua inferior e convexa. Por conseguinte, do Exemplo 2.1.5 temos que N_K é um operador monótono maximal.

A seguir, o seguinte resultado mostra que qualquer aplicação monótona tem uma extensão monótona maximal.

Lema 2.1.7 [38, Proposição 12.6] *Seja T monótona multivalorada, então existe uma aplicação \bar{T} monótono maximal (não necessariamente única) tal que $T \subset \bar{T}$, isto é, T possui uma extensão monótona maximal.*

Lema 2.1.8 *Seja T monótono. O operador T é maximal se, e somente se, para cada $(x, x') \in \mathbb{R}^n \times \mathbb{R}^n$,*

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

então, $(x, x') \in T$.

Prova. Seja T um operador monótono maximal, vamos provar que $(x, x') \in T$. Suponhamos que $x' \notin T(x)$. Definimos

$$\hat{T}(x) = \begin{cases} T(y), & \text{se } y \neq x \\ x', & \text{se } y = x \end{cases}$$

Logo, vamos verificar que o operador \hat{T} é monótono. De fato, precisamos verificar que para todo $(y, y'), (z, z') \in \hat{T}$ se cumpre

$$\langle y - z, y' - z' \rangle \geq 0. \quad (2.1.4)$$

Da definição de \hat{T} e sendo T monótono, basta verificarmos (2.1.4) para $z = x$. Logo, da hipótese temos que

$$\langle y - x, y' - x' \rangle \geq 0,$$

isto é, \hat{T} é um operador monótono. Além disso, da definição de \hat{T} podemos reparar que $T \subset \hat{T}$. Desde que T é maximal temos que $T = \hat{T}$ e como $x' \in \hat{T}$ obtemos que $x' \in T$, o qual é uma contradição. Logo, $(x, x') \in T$.

Reciprocamente, suponha que T não é maximal, isto é, existe \bar{T} monótono tal que $T \subsetneq \bar{T}$ com $T \neq \bar{T}$. Seja $(w, w') \in \bar{T}$, desde que T é monótono, temos que

$$\langle w - y, w' - y' \rangle \geq 0, \quad (y, y') \in \bar{T}.$$

Como $T \subsetneq \bar{T}$, tomamos em particular para todos os $(x, x') \in T$. Então,

$$\langle w - x, w' - x' \rangle \geq 0, \quad \forall (x, x') \in T$$

o que implica que $(w, w') \in T$. Logo, $\bar{T} \subseteq T$. Daí, $T = \bar{T}$, o que contradiz com a nossa suposição. Portanto, T é monótono maximal. ■

Exemplo 2.1.9 Toda aplicação monótona contínua $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ é monótona maximal.

De fato, seja $(x, u) \in \mathbb{R}^n \times \mathbb{R}^n$ tal que

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

Tomando $y = x - t(z - x)$ com $t > 0$ e $z \in \mathbb{R}^n$ arbitrário obtemos

$$\begin{aligned} 0 &\leq \langle u - T(x - t(z - x)), x - (x - t(z - x)) \rangle \\ &= t \langle u - T(x - t(z - x)), z - x \rangle. \end{aligned}$$

Como $t > 0$ podemos multiplicar por $1/t$. Tomando limite, quando $t \rightarrow 0$, e desde que T é um contínuo temos

$$0 \leq \langle u - T(x), z - x \rangle,$$

para todo $z \in \mathbb{R}^n$. Em particular para $z = x + T(x) - u$ temos

$$0 \leq \langle u - T(x), (x - T(x) + u) - x \rangle = -\|T(x) - u\|^2 \iff T(x) = u.$$

Logo, do Lema 2.1.8 obtemos que T é um operador maximal.

O exemplo acima não se cumpre se $\text{dom}(T) \subsetneq \mathbb{R}^n$. De fato, considere uma aplicação monótona $T : \mathbb{R} \rightarrow \mathbb{R}$ definida como

$$T(x) := \begin{cases} x, & \text{se } x < 0, \\ x, & \text{se } x > 0, \\ \emptyset, & \text{se } x = 0. \end{cases}$$

A monotonicidade vem do fato que para quaisquer $(x, x), (x', x') \in T$ temos

$$\langle x - x', x - x' \rangle = \|x - x'\|^2 \geq 0.$$

Observe que esta aplicação é contínua em cada ponto do seu domínio $\text{dom}(T) = \mathbb{R} \setminus \{0\}$, mas não é uma aplicação maximal pois sendo $\tilde{T}(x) = x$ monótona tal que $T \subset \tilde{T}$ temos que $T \neq \tilde{T}$.

A seguir, apresentamos uma propriedade topológica utilizado para provar alguns resultados do seguinte capítulo desta dissertação.

Lema 2.1.10 *Seja T monótono maximal. Para cada $x \in \mathbb{R}^n$, o conjunto $T(x)$ é fechado e convexo.*

Prova. Desde que T é monótono maximal e do Lema 2.1.8, para cada $(x, x^*) \in T$ temos

$$T(x) = \{x^* \in \mathbb{R}^n \mid \langle x^* - y^*, x - y \rangle \geq 0, \quad \forall (y, y^*) \in T\},$$

o qual é imediato provar que $T(x)$ é fechado e convexo. ■

Uma das caracterizações mais úteis e fundamentais das aplicações monótonas máximas é dado por o Teorema de Minty. Isto se reduz à questão de relacionar monotonicidade maximal de um operador monótono T com a sobrejetividade da correspondente aplicação perturbada $I + \lambda T$ com $\lambda > 0$ como é dado no seguinte resultado.

Proposição 2.1.11 [4, Teorema 6.2.2] *Seja T monótono e seja $\lambda > 0$. Então, T é monótono maximal se, e somente se, $I + \lambda T$ é sobrejetora, para cada $\lambda > 0$, isto é, $Im(I + \lambda T) = \mathbb{R}^n$.*

A seguir, uma consequência desta proposição é dado no seguinte lema, chamado como o Lema de Representação.

Lema 2.1.12 *Sejam $c > 0$ e T monótono. Então cada elemento $z \in \mathbb{R}^n$ pode ser escrito ao máximo da forma $x + cy$, onde $(x, y) \in T$. Se T é maximal, então cada elemento de $z \in \mathbb{R}^n$ pode ser escrito de forma única como $x + cy$, onde $(x, y) \in T$.*

Prova. Da Proposição 2.1.11, para cada $z \in \mathbb{R}^n$ existe $x \in \mathbb{R}^n$ tal que $(x, z) \in (I + cT)$, isto é, $z \in x + cy$ com $(x, y) \in T$ e $c > 0$. Logo, provaremos que a representação de qualquer elemento $z \in \mathbb{R}^n$ é única. Suponhamos que existem $x, x' \in \mathbb{R}^n$ com $x \neq x'$ tal que

$$(I + cT)(x) \cap (I + cT)(x') \neq \emptyset,$$

isto é, existe z tais que $(x, z) \in (I + cT)$ e $(x', z) \in (I + cT)$ com $c > 0$. Então,

$$\left(x, \frac{z-x}{c}\right) \in T \quad \text{e} \quad \left(x', \frac{z-x'}{c}\right) \in T.$$

Como T é um operador monótono, temos que

$$0 \leq \left\langle x - x', \frac{z-x}{c} - \frac{z-x'}{c} \right\rangle = \frac{1}{c} \langle x - x', x' - x \rangle.$$

Como $c > 0$, temos

$$0 \leq \langle x - x', x' - x \rangle = -\|x - x'\|^2.$$

Daí, $\|x - x'\| = 0$. Logo, $x = x'$. O qual contradiz com a nossa suposição. Então, qualquer $z \in \mathbb{R}^n$ pode ser apenas representado de uma única forma. ■

A seguir, é apresentado um teorema dado por Rockafellar[37]. Mas sua prova mais simples pode ser encontrada em Burachick[8] para espaços de Banach reflexivos, no qual o seguinte resultado, embora não aparece nesse livro, será útil para a demonstração dele.

Lema 2.1.13 *Seja $K \subset \mathbb{R}^n$ um conjunto fechado tal que para todo $x, y \in K$*

$$\frac{x + y}{2} \in K,$$

então K é convexo.

Prova. Dados $x, y \in K$ e $z = \alpha x + (1 - \alpha)y$ com $\alpha \in [0, 1]$. Vamos provar que $z \in K$.

Definimos os seguintes conjuntos

$$S_j = \left\{ \lambda x + (1 - \lambda)y \mid \lambda = \frac{i}{2^j}, \quad i = 0, 1, 2, \dots, 2^j \right\},$$

para todo $j \in \mathbb{N}$.

Observe que $S_j \subset K$ para cada $j \in \mathbb{N}$. Além disso, definimos a sequência

$$x_j = \arg \min_{w \in S_j} \|z - w\|,$$

o qual implica que $d(x_j, z) \leq \frac{1}{2^j}$.

Tomando limite, quando $j \rightarrow +\infty$, temos que $d(x_j, z) \rightarrow 0$, isto é, $x_j \rightarrow z$. Então, do fecho de K obtemos que $z \in K$. ■

Teorema 2.1.14 [8, Teorema 4.4.9] *Seja T monótona maximal. Então $\overline{Im(T)}$ é convexo.*

Proposição 2.1.15 *Seja T monótona maximal, então o conjunto $\overline{dom(T)}$ é convexo.*

Prova. Como T é monótono maximal então T^{-1} tem também essas propriedades. Aplicando o Teorema 2.1.14 ao operador T^{-1} temos que $\overline{Im(T^{-1})}$ é convexo, mas $Im(T^{-1}) = dom(T)$. Daí, $\overline{dom(T)}$ é convexo. ■

Definição 2.1.16 *Se diz que T é localmente limitado em $\bar{x} \in \mathbb{R}^n$ se existe $\delta > 0$ tal que $T(B(\bar{x}, \delta))$ é limitado.*

Observe que a definição dada acima que se T é localmente limitado em \bar{x} então o conjunto $T(\bar{x})$ é limitado.

Teorema 2.1.17 *Seja T monótono maximal. Se $x \in \overline{\text{dom}(T)}$ e T é localmente limitado em x , então $x \in \text{int}(\text{dom}(T))$.*

Proposição 2.1.18 [4, Proposição 6.6.2] *Seja T monótono maximal e seja $x \in \text{dom}(T)$. Então $T(x)$ é um conjunto compacto e, e somente se, $x \in \text{int}(\text{dom}(T))$.*

Proposição 2.1.19 *Seja T monótona maximal. Então T^{-1} é localmente limitado se, e somente se, T é sobrejetivo.*

Prova. Primeiro, vamos provar que T é sobrejetivo. Suponhamos que T^{-1} é localmente limitado, seja $x \in \overline{\text{dom}(T^{-1})}$ então, segundo o Teorema 2.1.17, temos que $x \in \text{int}(\text{dom}(T^{-1}))$, isto implica que $\overline{\text{dom}(T^{-1})} \subset \text{int}(\text{dom}(T^{-1}))$. Portanto, $\overline{\text{dom}(T^{-1})} = \text{int}(\text{dom}(T^{-1}))$, o que equivalente dizer que $\text{dom}(T^{-1})$ é um conjunto aberto e fechado à vez. Como $\text{dom}(T^{-1}) \neq \emptyset$, temos que $\text{Im}(T) = \mathbb{R}^n$.

Reciprocamente, consideremos T sobrejetivo e seja $x \in B(x, \epsilon) \subset \text{dom}(T^{-1})$, isto implica que $B(x, \epsilon) \in \text{int}(\text{dom}(T^{-1}))$. Da Proposição 2.1.18 temos que $T^{-1}(B(x, \epsilon))$ é limitado. Portanto, T^{-1} é localmente limitado. ■

Lema 2.1.20 *Seja T monótono maximal com domínio limitado. Então, T é sobrejetivo. Em particular, $0 \in T(z)$ para algum $z \in \mathbb{R}^n$.*

Prova. Vamos provar que T^{-1} é localmente limitado em $y \in \mathbb{R}^n$. Consideremos o conjunto

$$T^{-1}(B(y, \epsilon)) = \cup_{x \in B(y, \epsilon)} \{x \mid y \in T(x)\} \subseteq \text{dom}(T).$$

Pela hipótese, o $\text{dom}(T)$ é limitado o que implica que $T^{-1}(B(y, \epsilon))$ é limitado. Logo, da Proposição 2.1.19 temos que T é um operador sobrejetivo. ■

A seguir, são enunciados conceitos de operadores localmente monótono maximal.

Definição 2.1.21 *Se diz que o operador T é monótono em $U \subseteq \mathbb{R}^n \times \mathbb{R}^n$, ou localmente monótono, se existe um operador monótono \bar{T} tal que $\text{graf } T \cap U = \text{graf } \bar{T}$. Mais ainda, T é monótono maximal em $U \subseteq \mathbb{R}^n \times \mathbb{R}^n$, ou localmente monótono maximal, se existe \bar{T} monótono maximal tal que $\text{graf } \bar{T} \cap U = \text{graf } T \cap U$.*

Lema 2.1.22 *Seja T qualquer aplicação monótona em $U \subseteq \mathbb{R}^n \times \mathbb{R}^n$. Então o operador T possui, talvez não única, uma extensão monótona maximal em U*

Prova. Análogo ao feito no Lema 2.1.7. ■

No seguinte resultado, apresentamos uma versão local do critério de continuidade para aplicações multivaloradas monótonas maximais dado no Exemplo 2.1.9.

Proposição 2.1.23 *Seja T um operador tal que para alguns conjuntos abertos $X, Y \subset \mathbb{R}^n$, a aplicação $x \mapsto T(x) \cap Y$ é simples valorada e contínua sobre X . Se T é monótono sobre $X \times Y$ então, também é monótono maximal sobre $X \times Y$.*

Prova. Do Lema 2.1.22, T possui uma extensão monótona maximal em $X \times Y$. Consideremos \bar{T} tal extensão de $\text{graf } T \cap (X \times Y)$ e seja $(\tilde{x}, \tilde{y}) \in \text{graf } \bar{T} \cap (X \times Y)$.

Como X é aberto, para qualquer $z \in \mathbb{R}^n$, existe $\epsilon > 0$ para n suficientemente grande tal que $1/n < \epsilon$, com $\tilde{x} + (1/n)z \in X$.

Assim, pela monotonicidade de \bar{T} ,

$$0 \leq \langle T(\tilde{x} + (1/n)z) \cap Y - \tilde{y}, (\tilde{x} + (1/n)z) - \tilde{x} \rangle = \frac{1}{n} \left\langle T(\tilde{x} + \frac{1}{n}z) \cap Y - \tilde{y}, z \right\rangle.$$

Multiplicando por $n > 0$ e como T é contínuo,

$$0 \leq \left\langle T(\tilde{x} + \frac{1}{n}z) \cap Y - \tilde{y}, z \right\rangle \rightarrow \langle T(\tilde{x}) \cap Y - \tilde{y}, z \rangle \geq 0.$$

Então, para $z \in \mathbb{R}^n$ arbitrário, temos que $T(\tilde{x}) \cap Y = \tilde{y}$. Logo, $(\tilde{x}, \tilde{y}) \in \text{graf } T$. ■

Por outro lado, definimos a soma de duas aplicações multivaloradas T_1, T_2 por

$$T_1 + T_2 = \{(x, y + w) \mid (x, y) \in T_1, (x, w) \in T_2\},$$

onde $(T_1 + T_2)(x) = T_1(x) + T_2(x)$.

A seguir, damos o resultado que mostra que a família de operadores monótonos é fechado com respeito à adição.

Proposição 2.1.24 *Sejam T_1 e T_2 monótonos. Então, $T_1 + T_2$ é monótono.*

Prova. Sejam $(x, y), (w, z) \in (T_1 + T_2)$. Daí,

$$y \in (T_1 + T_2)(x) = T_1(x) + T_2(x),$$

isto é, existem $(x, y_1) \in T_1$ e $(x, y_2) \in T_2$ tal que $y = y_1 + y_2$. Analogamente, existem $(w, z_1) \in T_1$ e $(w, z_2) \in T_2$ tal que $z = z_1 + z_2$. Como T_1 e T_2 são operadores monótonos, temos que

$$\langle x - w, y_1 - z_1 \rangle \geq 0, \quad e \quad \langle x - w, y_2 - z_2 \rangle \geq 0.$$

Logo,

$$\langle x - w, y - z \rangle = \langle x - w, (y_1 + y_2) - (z_1 + z_2) \rangle = \langle x - w, y_1 - z_1 \rangle + \langle x - w, y_2 - z_2 \rangle \geq 0.$$

Portanto, $T_1 + T_2$ é um operador monótono. ■

Em geral, a soma de dois operadores maximais nem sempre é maximal como mostraremos no seguinte exemplo.

Exemplo 2.1.25 *Sejam os conjuntos*

$$A = \{(x, a) \in \mathbb{R} \times \mathbb{R} \mid x^2 \leq a\},$$

$$B = \mathbb{R} \times \{0\}.$$

Consideremos os operadores $T_1 = \partial\delta_A$, $T_2 = \partial\delta_B$ e $T_3 = \partial\delta_{\{(0,0)\}}$. Como A e B são convexos, do Exemplo 2.1.6 temos que T_1 , T_2 e T_3 são monótonos maximais.

Mais ainda,

$$\text{dom}(T_1 + T_2) = \{x \in \mathbb{R}^2 \mid (T_1 + T_2)(x) \neq \emptyset\} = \{(0, 0)\},$$

$$\text{dom}(T_3) = \{x \in \mathbb{R}^2 \mid \partial\delta_{\{(0,0)\}}(x) \neq \emptyset\} = \{(0, 0)\}.$$

Então, o subdiferencial da função indicadora para os conjuntos A , B e $\{(0, 0)\}$, são obtido como segue:

- i. $T_1(0, 0) = \{s \in \mathbb{R}^2 \mid \langle s, y - (0, 0) \rangle \leq 0, \forall y \in A\}$, isto é, $s_1 y_1 + s_2 y_2 \leq 0$ e $y_2 \geq 0$. Se $y_2 = 0$ temos que $s_1 y_1 \leq 0$. Se $y_1 > 0$, $s_1 \leq 0$. Se $y_1 < 0$, $s_1 \geq 0$. Logo, $s_1 = 0$. Como $y_2 \geq 0$ temos que $s_2 \geq 0$. Então $s = \{0\} \times \mathbb{R}_+$.

ii. $T_2(0,0) = \{s \in \mathbb{R}^2 \mid \langle s, y - (0,0) \rangle \leq 0, \forall y \in B\}$, isto é, $s_1y_1 + s_2y_2 \leq 0$ para todo $(y_1, y_2) \in \mathbb{R} \times \{0\}$. Logo, $s_1y_1 \leq 0$ e $s_2 \in \mathbb{R}$. Se $y_1 > 0$, $s_1 \leq 0$. Se $y_1 < 0$, $s_1 \geq 0$. Logo, $s_1 = 0$. Então $s = \{0\} \times \mathbb{R}$.

iii. $T_3(0,0) = \{s \in \mathbb{R}^2 \mid \langle s, y \rangle \leq 0, \forall y \in (0,0)\}$. Então, $s = \mathbb{R}^2$.

Além disso,

$$(T_1 + T_2)(0,0) = T_1(0,0) + T_2(0,0) = \{0\} \times \mathbb{R}_+ + \{0\} \times \mathbb{R} = \{0\} \times \mathbb{R}.$$

Então, $(T_1 + T_2)(0,0) = \{0\} \times \mathbb{R} \subsetneq \mathbb{R}^2 = T_3(0,0)$. Portanto, $T_1 + T_2$ não é maximal.

Na verdade, tem muitos critérios que garantem que a soma de operadores monótonos maximais é maximal. O nosso caso de interesse é dado no seguinte teorema, cuja prova foi dada por Rockafellar em espaços de Banach real.

Teorema 2.1.26 [35, Teorema 1,a] *Sejam T_1, T_2 operadores monótonos maximais. Se $\text{dom}(T_1) \cap \text{int}(\text{dom}(T_2)) \neq \emptyset$ então*

$$T_1 + T_2$$

é um operador monótono maximal.

Proposição 2.1.27 [38, Exemplo 12.48] *Seja $K \subset \mathbb{R}^n$ um conjunto convexo, fechado e não vazio e uma aplicação $F : K \rightarrow \mathbb{R}^n$. Se F é contínua e monótona sobre K , então*

$$T = F + N_K$$

é um operador monótono maximal.

Proposição 2.1.28 [4, Proposição 6.3.1] *Seja $K \subset \mathbb{R}^n$ um conjunto não vazio, convexo e fechado. Então para todo $\lambda > 0$ e para todo $x \in \mathbb{R}^n$ temos que*

$$(I + \lambda N_K)^{-1}(x) = P_K(x), \quad (2.1.5)$$

e portanto a projeção P_K sobre K é uma aplicação monótona maximal simples valorado.

Em seguida, apresentamos um resultado muito sobre o operador projetor, o qual vai ser utilizado no último capítulo desta dissertação.

Lema 2.1.29 *Consideremos $K \subset \mathbb{R}^n$ um conjunto não vazio e convexo. A projeção sobre o cone dual K^\otimes é dado como*

$$P_{K^\otimes}(x) = \frac{1}{2} \nabla d_K^2(x). \quad (2.1.6)$$

Prova. Primeiro, afirmamos que é válido a seguinte igualdade

$$I - (I + \lambda T)^{-1} = (I + \lambda T^{-1})^{-1}, \quad (2.1.7)$$

no qual a sua prova é imediata.

Logo, de (1.1.2), (1.1.3), (2.1.5), (2.1.7) e dos Lemas 1.1.1 e 1.1.16 temos que

$$\begin{aligned} P_{K^\otimes}(x) &= (I + \lambda N_{K^\otimes})^{-1}(x) \\ &= (I - (I + \lambda N_{K^\otimes}^{-1})^{-1})(x) \\ &= (I - (I + \lambda N_K)^{-1})(x) \\ &= (I - P_K)(x) \\ &= \lambda \partial \delta_\lambda(x) \\ &= \frac{1}{2} \nabla d_K^2(x) \quad \blacksquare \end{aligned} \quad (2.1.8)$$

A partir daqui, vamos definir operadores especiais, tais como os operadores fortemente monótonos, não expansivos, proximais, as λ -regularização Yosida, os λ -hipomonótonos os quais vão ser cruciais no desenvolvimento desta dissertação.

Definição 2.1.30 *Se diz que T é fortemente monótono, com módulo $\alpha > 0$, se para $(x, y), (x', y') \in T$, temos*

$$\langle x - x', y - y' \rangle \geq \alpha \|x - x'\|^2,$$

ou equivalentemente, se existe um $\alpha > 0$ tal que $T - \alpha I$ seja monótona.

Proposição 2.1.31 *Se a inversa do operador T é fortemente monótono, para algum $\alpha > 0$, então T é simples valorada.*

Prova. Seja $x \in \mathbb{R}^n$ tais que $(x, y), (x, y') \in T$, para algum $\alpha > 0$, temos que

$$0 = \langle x - x, y - y' \rangle \geq \alpha \|y' - y\|^2,$$

isto implica que $\|y' - y\| = 0$ se, e somente se, $y' = y$. \blacksquare

Definição 2.1.32 *Se diz que T é não-expansivo se,*

$$\|y - y'\| \leq \|x - x'\|, \quad \forall (x, y), (x', y') \in T.$$

Definição 2.1.33 *Se diz que T é firmemente não-expansivo se*

$$\|y' - y\|^2 \leq \|x' - x\|^2 - \|(x' - y') - (x - y)\|^2, \quad \forall (x, y), (x', y') \in T$$

onde com a norma usual, $\|x\| = \langle x, x \rangle^{1/2}$, pode ser redefinido como

$$\langle x' - x, y' - y \rangle \geq \|y' - y\|^2, \quad \forall (x, y), (x', y') \in T.$$

Observação 2.1.34 *Podemos observar que se T é um operador firmemente-expansivo é equivalente dizer que T^{-1} é fortemente monótono com módulo $\alpha = 1$.*

Na literatura, operadores firmemente não-expansivos são algumas vezes chamados como operadores pseudo-contrativos.

Observação 2.1.35 *A propriedade de não expansividade é preservada tanto para composições de aplicações quanto para soma de aplicações não expansivas sob certas condições. Sejam S_1, S_2 aplicações não expansivas, são válidas as seguintes propriedades:*

- a. $S_1 \circ S_2$ é não expansiva.
- b. $\lambda_1 S_1 + \lambda_2 S_2$ é não expansiva se $|\lambda_1| + |\lambda_2| \leq 1$.

De fato, para provarmos o item a. tomaremos $(x, y), (x', y') \in S_1 \circ S_2$, isto é, existem $a, a' \in \text{dom}(S_1)$ tais que $(a, y) \in S_1$ com $(x, a) \in S_2$ e $(a', y') \in S_1$ com $(x', a') \in S_2$.

Como S_1 e S_2 são aplicações não expansivas, temos

$$\|y' - y\| \leq \|a - a'\| \quad e, \quad \|a - a'\| \leq \|x - x'\|,$$

o que implica que $\|y' - y\| \leq \|x - x'\|$. Logo, $S_1 \circ S_2$ é uma aplicação não expansiva.

Agora vamos provar o item b. Sejam $(x, y), (x', y') \in \lambda_1 S_1 + \lambda_2 S_2$, então existem y_1, y_2 com $(x, y_1) \in S_1$ e $(x, y_2) \in S_2$ tais que

$$y = \lambda_1 y_1 + \lambda_2 y_2. \tag{2.1.9}$$

Analogamente, existem y'_1, y'_2 com $(x', y'_1) \in S_1$ e $(x', y'_2) \in S_2$ tais que

$$y' = \lambda_1 y'_1 + \lambda_2 y'_2. \quad (2.1.10)$$

De (2.1.9), (2.1.10) e como S_1 e S_2 são não-expansivos, temos

$$\begin{aligned} \|y' - y\| &= \|(\lambda_1 y'_1 + \lambda_2 y'_2) - (\lambda_1 y_1 + \lambda_2 y_2)\| \\ &\leq |\lambda_1| \|y'_1 - y_1\| + |\lambda_2| \|y'_2 - y_2\| \\ &\leq |\lambda_1| \|x' - x\| + |\lambda_2| \|x' - x\| \\ &= (|\lambda_1| + |\lambda_2|) \|x' - x\|. \end{aligned}$$

Logo, temos que $\lambda_1 S_1 + \lambda_2 S_2$ é uma aplicação não expansiva se $|\lambda_1| + |\lambda_2| \leq 1$.

Proposição 2.1.36 *Seja T uma aplicação não expansiva, então T é simples valorada e k -Lipschitz contínua com constante Lipschitz k .*

Prova. Seja $x \in \mathbb{R}^n$ tal que $(x, y), (x, y') \in T$. Como T é não-expansivo temos que

$$\|y' - y\| \leq \|x - x'\| = 0.$$

Logo, $\|y' - y\| = 0$ se, e somente se, $y = y'$. Portanto, T é simples valorada. Além disso, por definição, um operador não-expansivo é k -Lipschitz com constante $k = 1$. ■

Uma das vantagens das aplicações não expansivas é que estas podem ser utilizadas na construção de aplicações monótonas como veremos no seguinte resultado.

Proposição 2.1.37 *Seja uma aplicação não expansiva $T_1 : \mathbb{R}^n \rightarrow \mathbb{R}^n$, então $T_2 = I - T_1$ é monótono.*

Prova. Sejam $x, x' \in \mathbb{R}^n$. Como T_1 é não-expansivo, temos que

$$\begin{aligned} \langle T_2(x) - T_2(x'), x - x' \rangle &= \langle x - x', x - x' \rangle + \langle T_1(x') - T_1(x), x - x' \rangle \\ &\geq \|x - x'\|^2 - \|T_1(x') - T_1(x)\| \|x - x'\| \\ &= 0. \quad \blacksquare \end{aligned}$$

Proposição 2.1.38 *Um operador T é firmemente não-expansivo se, e somente se, $I - T$ é firmemente não-expansivo.*

Prova. Primeiro, vamos provar que $I - T$ é firmemente não-expansivo. Sejam $(x, y), (x', y') \in T$, é válido a seguinte desigualdade

$$\|(x' - y') - (x - y)\|^2 \leq \langle x' - x, (x' - y') - (x - y) \rangle$$

para todo $(x, x - y), (x', x' - y') \in I - T$. De fato, como T é firmemente não-expansivo,

$$\begin{aligned} \|(x' - y') - (x - y)\|^2 &= \langle x' - x, (x' - x) - (y' - y) \rangle - \langle y' - y, x' - x \rangle + \langle y' - y, y' - y \rangle \\ &\leq \langle x' - x, (x' - x) - (y' - y) \rangle - \|y' - y\|^2 + \|y' - y\|^2 \\ &= \langle x' - x, (x' - y') - (x - y) \rangle. \end{aligned}$$

Reciprocamente, sejam $(x, y), (x', y') \in (I - T)$. É válido que

$$\langle x' - x, (x - y) - (x' - y') \rangle \geq \|(x - y) - (x' - y')\|^2$$

para todo $(x, x - y), (x', x' - y') \in I - T$. De fato, como $I - T$ é firmemente não-expansivo,

$$\begin{aligned} -\langle x - x', y - y' \rangle &\geq \|y - y'\|^2 - 2\langle x - x', y - y' \rangle, \\ \|x - x'\|^2 - \langle x - x', y - y' \rangle &\geq \|y - y'\|^2 - 2\langle x - x', y - y' \rangle + \|x - x'\|^2 \\ \langle x - x', (x - x') - (y - y') \rangle &\geq \|(x - x') - (y - y')\|^2 \quad \blacksquare \end{aligned}$$

Definição 2.1.39 O conjunto zeros de T é definido como

$$\text{Zer}(T) := T^{-1}(0) = \{x \in \mathbb{R}^n : 0 \in T(x)\}.$$

Cabe notar que este conjunto é interessante em otimização e na teoria de pontos fixos do fato que se

- $T = \partial f$, onde f é uma função própria, convexa e semi-contínua inferior, então $\text{Zer}(T)$ é o conjunto de minimizadores de f .
- $T = I - \tilde{T}$ com \tilde{T} não-expansivo, então $\text{Zer}(T)$ é o conjunto de pontos fixos de T .

Definição 2.1.40 O resolvente do operador T , para $c > 0$, é definido por

$$J_c^T = (I + cT)^{-1}. \quad (2.1.11)$$

Se $c = 1$, o resolvente do operador T será denotado por J^T .

Proposição 2.1.41 [4, Proposição 6.2.1] *Seja T um operador monótono e seja $c > 0$. Então $J_c^T : \text{Im}(I + cT) \rightrightarrows \mathbb{R}^n$ é monótono, não-expansivo e simples valorado. Além disso, se T é maximal então J_c^T é também maximal.*

Proposição 2.1.42 *O operador T é monótono se, e somente se, J_c^T é um operador firmemente não-expansivo para cada $c > 0$.*

Prova. Sejam $(x, y), (x', y') \in T$ e seja $c > 0$.

$$\begin{aligned} T &\iff \langle x - x', cy - cy' \rangle \geq 0 \\ &\iff \langle x - x', cy - cy' + x - x' \rangle \geq \|x - x'\|^2 \\ &\iff \langle x - x', (x + cy) - (x' + cy') \rangle \geq \|x - x'\|^2. \end{aligned}$$

Então, $(I + cT)^{-1}$ é firmemente não-expansiva. ■

Lema 2.1.43 *Seja T monótona e J_c^T como definido em (2.1.11). O operador $(J_c^T)^{-1}$ é fortemente monótona com $\alpha = 1$.*

Prova. Como T é um operador monótono, segundo a Proposição 2.1.42, J_c^T é não-expansivo para cada $c > 0$. Logo, da Observação 2.1.34 temos que $(J_c^T)^{-1}$ é fortemente monótono com $\alpha = 1$. ■

Lema 2.1.44 *Seja T um operador monótono. Para qualquer $c > 0$ e $\rho \in (0, 2)$, a aplicação*

$$\rho J_c^T + (1 - \rho)I$$

é não expansiva, e os seus pontos fixos são os zeros de T .

Prova. Vamos provar que o operador $\rho J_c^T + (1 - \rho)I$ é não-expansivo. De fato,

$$\rho J_c^T + (1 - \rho)I = \frac{2 - \rho}{2}I + \frac{\rho}{2}N,$$

onde $N = 2J_c^T - I$.

Além disso, é imediato verificar que se J_c^T é não-expansivo então N é também não

-expansivo. Logo, desde que $\rho \in (0, 2)$, segundo a Observação 2.1.35, item *b*. temos que $\rho J_c^T + (1 - \rho)I$ é um operador não-expansivo.

Por outro lado, suponhamos que x é ponto fixo de $\rho J_c^T + (1 - \rho)I$. Então, $J_c^T(x) = x$. Daí, $x = (I + cT)(x)$ se, e somente se, $0 \in T(x)$. Portanto, $x \in \text{Zer}(T)$. ■

A seguir, definimos o operador

$$Q_c^T = I - J_c^T. \quad (2.1.12)$$

Observação 2.1.45 *Podemos observar que o operador Q_c^T é firmemente não-expansivo. De fato, da Proposição 2.1.42 temos que J_c^T é firmemente não-expansivo e, segundo a Proposição 2.1.38, temos que Q_c^T é também firmemente não-expansivo.*

Observe que os operadores J_c^T e Q_c^T estão íntimamente relacionados, pois um zero de T é um ponto fixo de J_c^T e este à vez é um zero de Q_c^T , isto é,

$$0 \in T(z) \iff J_c^T(z) = z \iff Q_c^T(z) = 0. \quad (2.1.13)$$

De fato,

$$0 \in T(z) \iff z \in (I + T)(z) \iff J_c^T(z) = z \iff (I - J_c^T)(z) = 0 \iff Q_c^T(z) = 0.$$

Proposição 2.1.46 *Para todo $z, w \in \mathbb{R}^n$ é válido as seguintes afirmações:*

a. $c^{-1}Q_c^T(z) \in T(J_c^T(z))$.

b. $\|z - w\|^2 \geq \|J_c^T(z) - J_c^T(w)\|^2 + \|Q_c^T(z) - Q_c^T(w)\|^2$

Prova. Primeiro, vamos provar o item *a.*. De (2.1.11) e (2.1.12), para todo $z \in \mathbb{R}^n$,

$$\begin{aligned} (I + cT)(J_c^T(z)) = z &\iff z \in J_c^T(z) + cT(J_c^T(z)) \\ &\iff c^{-1}Q_c^T(z) \in T(J_c^T(z)). \end{aligned}$$

Segundo, vamos provar o item *b.*,

$$\begin{aligned} \|z - w\|^2 &= \|(J_c^T + Q_c^T)(z) - (J_c^T + Q_c^T)(w)\|^2 \\ &\geq \|J_c^T(z) - J_c^T(w)\|^2 + \|Q_c^T(z) - Q_c^T(w)\|^2. \end{aligned}$$

A última desigualdade é obtida do fato que $\langle J_c^T(w) - J_c^T(z), Q_c^T(w) - Q_c^T(z) \rangle \geq 0$.

■

Definição 2.1.47 Para qualquer aplicação T e $\lambda > 0$, a correspondente λ -regularização Yosida de T definido por

$$T_\lambda = (\lambda I + T^{-1})^{-1} = (I + \lambda^{-1}T^{-1})^{-1} \circ \lambda^{-1}I.$$

É importante mencionar que o estudo dos operadores λ -regularizações Yosida são considerados muito vantajosos em otimização. De fato,

a. Para qualquer T e $\lambda \in \mathbb{R}$ temos que

$$T_\lambda^{-1}(0) = T^{-1}(0) + \lambda I(0) = T^{-1}(0), \quad (2.1.14)$$

isto é, $Zer(T_\lambda) = Zer(T)$.

b. O T_λ pode ser localmente monótona embora T não seja, isto facilita o nosso estudo porque a teoria dos operadores monótonos tem uma grande quantidade de propriedades.

Proposição 2.1.48 Cada aplicação T , para qualquer $\lambda > 0$, satisfaz

$$T_\lambda = \lambda^{-1} (I - J_\lambda^T) \quad (2.1.15)$$

Prova. Por um cálculo simples, temos que

$$\begin{aligned} (w, z) \in \lambda^{-1} (I - J_\lambda^T) &\iff (w, w - \lambda z) \in J_\lambda^T \\ &\iff (w - \lambda z, w) \in (I + \lambda T) \\ &\iff (z, w - \lambda z) \in T^{-1} \\ &\iff (z, w) \in (T^{-1} + \lambda I) \\ &\iff (w, z) \in T_\lambda. \quad \blacksquare \end{aligned}$$

Proposição 2.1.49 [4, Teorema 6.2.3, (b)] *Seja T monótono maximal. Então, T_λ , é um operador simples valorado, monótono maximal e λ^{-1} -Lipschitz contínua.*

Definição 2.1.50 Uma aplicação T é dita λ -hipomonótona em \bar{x} se existe uma vizinhança $X \ni \bar{x}$ e $\lambda > 0$ tal que $T + \lambda I$ é monótono em X , isto é,

$$\langle y' - y, x' - x \rangle \geq -\lambda \|x' - x\|^2, \quad (x, y), (x', y') \in T.$$

O operador T é um operador λ -hipomonótono em X se é λ -hipomonótono em cada $\bar{x} \in X$.

Exemplo 2.1.51 *Toda aplicação monótona é λ -hipomonótona. De fato, consideremos T uma aplicação monótona então $T + \lambda I$ é monótona. Daí, T é λ -hipomonótona.*

Corolário 2.1.52 *Seja $F : X \rightarrow \mathbb{R}^n$ uma aplicação simples valorada e estritamente contínua sobre $X \in \mathbb{R}^n$, então F é λ -hipomonótona sobre X .*

Prova. Seja k a constante Lipschitz de F sobre X .

Definimos

$$G = F + kI.$$

Então,

$$\begin{aligned} \langle G(x') - G(x), x' - x \rangle &= \langle F(x') - F(x), x' - x \rangle + k \langle x' - x, x' - x \rangle \\ &\geq - \|F(x') - F(x)\| \|x' - x\| + k \|x' - x\|^2. \\ &\geq 0, \end{aligned}$$

onde a última desigualdade é obtida do fato de F ter k -Lipschitz contínua. Logo, G é monótono sobre X , isto é, F é uma aplicação λ -hipomonótona com $\lambda = k$. ■

Exemplo 2.1.53 *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função própria e semi-contínua inferior. A aplicação subdiferencial, ∂f , é λ -hipomonótona em \bar{x} se, e somente se, existe $\lambda \in \mathbb{R}_+$ tal que $f + 1/2\lambda \|\cdot\|^2$ é convexa sobre uma vizinhança de $X \ni \bar{x}$.*

De fato, seja ∂f uma aplicação λ -hipomonótona em \bar{x} , então existe uma vizinhança $X \ni \bar{x}$ e $\lambda \in \mathbb{R}_+$ tal que $\partial f + \lambda I$ é monótona sobre X .

Logo, pelo Exemplo 2.1.5, a convexidade da função

$$f + \frac{\lambda}{2} \|\cdot\|^2$$

em X é condição necessária e suficiente para $\partial f + \lambda I$ ser monótona em X . O resultado seguinte mostra como obter os resolventes de $J_c^{T\lambda}$ mediante utilização dos resolventes de T .

Lema 2.1.54 *Para qualquer T e $c, \lambda \in \mathbb{R}$ tal que $c + \lambda \neq 0$ e $c \neq 0$, temos*

$$J_c^{T\lambda} = \frac{c}{c + \lambda} J_{(c+\lambda)}^T + \frac{\lambda}{c + \lambda} I.$$

Prova. Por um cálculo simples, temos

$$\begin{aligned}
\bar{x} \in J_{(c+\lambda)}^T(x) &\iff \bar{x} \in (I + (c + \lambda)T)^{-1}(x) \\
&\iff \bar{x} \in T^{-1}\left(\frac{x - \bar{x}}{c + \lambda}\right) \\
&\iff \frac{c\bar{x} + \lambda x}{c + \lambda} \in (\lambda I + T^{-1})\left(\frac{x - \bar{x}}{c + \lambda}\right) \\
&\iff \frac{x - \bar{x}}{c + \lambda} \in T\left(\frac{c\bar{x} + \lambda x}{c + \lambda}\right) \\
&\iff x \in (I + cT_\lambda)\left(\frac{c\bar{x} + \lambda x}{c + \lambda}\right) \\
&\iff \frac{c\bar{x} + \lambda x}{c + \lambda} \in J_c^{T_\lambda}(x) \\
&\iff \bar{x} \in \left(\frac{(c + \lambda)}{c} J_c^{T_\lambda} + \frac{\lambda}{c} I\right)(x). \quad \blacksquare
\end{aligned}$$

Definição 2.1.55 Se diz que T tem k -Localização Lipschitz em $(\bar{z}, v) \in \text{graf } T$ se existem vizinhanças $X \ni \bar{z}$ e $Y \ni v$ tal que a aplicação

$$x \mapsto T(x) \cap Y, \quad x \in X$$

é simples valorada e Lipschitz contínua sobre X .

Proposição 2.1.56 Se T^{-1} tem uma k -localização Lipschitz em $(0, \bar{z})$. Então, para $\lambda \geq k$, T_λ é monótono maximal numa vizinhança de $(\bar{z}, 0)$.

Prova. Como T^{-1} tem uma k -Localização Lipschitz em $(0, \bar{z})$, existem vizinhanças $Y \ni 0$ e $X \ni \bar{z}$ tal que a aplicação

$$y \mapsto T^{-1}(y) \cap X$$

é simples valorada e k -Lipschitz contínua em X .

Seja F uma aplicação multivalorada k -Lipschitz contínua sobre Y tal que

$$\text{graf } F = \text{graf } T^{-1} \cap (Y \times X).$$

É válido que $F + \lambda I$ é uma aplicação monótona para $\lambda \geq k$. De fato, sejam $y_1, y_2 \in Y$ e como F é uma aplicação k -Lipschitz contínua, temos

$$\begin{aligned}
\langle (F(y_1) + \lambda y_1) - (F(y_2) + \lambda y_2), y_1 - y_2 \rangle &\geq -\|F(y_1) - F(y_2)\| \|y_1 - y_2\| + \lambda \|y_1 - y_2\|^2 \\
&\geq -k \|y_1 - y_2\|^2 + \lambda \|y_1 - y_2\|^2 \\
&= (\lambda - k) \|y_1 - y_2\|^2.
\end{aligned}$$

Então, para $\lambda \geq k$, $F + \lambda I$ é uma aplicação monótona.

Por outro lado, para qualquer $y \in \text{dom}(F + \lambda I) = Y$, temos

$$(F + \lambda I)(y) = F(y) + \lambda y = T^{-1}(y) \cap X + \lambda y = (T^{-1}(y) + \lambda y) \cap (X + \lambda y).$$

Então,

$$\text{graf}(F + \lambda I) = \text{graf}(T^{-1} + \lambda I) \cap (Y \times (X + \lambda y)) = \text{graf}(T^{-1} + \lambda I) \cap U = \text{graf} T_\lambda^{-1} \cap U,$$

onde

$$U = \{(y, z) : y \in Y, z \in X + \lambda y\}.$$

Como F é k -Lipschitz contínua sobre Y , então $F + \lambda I$ também é k -Lipschitz contínua.

Em particular, $F + \lambda I$ é contínua sobre Y .

Além disso, da monotonicidade de $F + \lambda I$, temos que $T^{-1} + \lambda I$ é uma aplicação monótona sobre U . Logo, $T_\lambda = (T^{-1} + \lambda I)^{-1}$ é monótona sobre U . Portanto, pela Proposição 2.1.23, temos que T_λ é monótona maximal sobre U . ■

2.2 Dualidade com monotonicidade maximal

Nesta seção vamos considerar o modelo dual de composição em programação convexa estudado por Pennanen[27, Seção 5].

Sejam X e U dois espaços euclidianos, K um cone convexo em U , $f : X \rightarrow \mathbb{R}$ e $g : U \rightarrow \mathbb{R}$ convexas e $H : X \rightarrow U$ tal que H é K -convexa, isto é, para todo $x_1, x_2 \in X$ temos que

$$H(\alpha_1 x_1 + \alpha_2 x_2) - \alpha_1 H(x_1) - \alpha_2 H(x_2) \in K,$$

onde $\alpha_1, \alpha_2 \in [0, 1]$ e $\alpha_1 + \alpha_2 = 1$.

A composição de H e g é definida por

$$(g \circ H)(x) = g(H(x)),$$

para todo $x \in \text{dom}(g \circ H) = \{x \in \text{dom}(H) \mid H(x) \in \text{dom}(g)\}$.

O modelo de composição em programação convexa estudada por Pennanen[28, Seção 8] é o problema de minimização

$$\min \{f_0(x) = (f + g \circ H)(x), \quad x \in \text{dom}(f + g \circ H)\}.$$

Segundo Pennanen[28, Proposição 8.1], se existir $x \in ri(dom(f)) \cap ri(dom(H))$ tal que $H(x) \in ri(dom(g))$, então o subdiferencial de f_0 pode ser expresso como

$$\partial f_0(x) = \partial f(x) + DH(x)^T \partial g(H(x)),$$

onde $DH(x)^T$ é a transporta da matriz jacobiana de H .

Nestas condições, uma generalização deste modelo de otimização é obtida quando ∂f e ∂g são substituídos por aplicações monótonas maximais gerais, tendo assim o seguinte resultado de dualidade em programação de composição.

Proposição 2.2.1 [27, Proposição 5.1] *Sejam X, Y, W e Z dois espaços euclidianos e K um cone convexo de X . Sejam $S : X \rightrightarrows W$ e $T : Y \rightrightarrows Z$ operadores monótonos maximais e $H : X \rightarrow Y$ uma função K -convexa e K -fechado tal que $Im(T) \subset K^\otimes$.*

Dado um ponto $x \in X$, x resolve

$$(\mathcal{P}_{\text{comp}}) \quad 0 \in S(x) + DH(x)^T(T(H(x)))$$

se, e somente se, existe um ponto z tal que

$$(\mathcal{L}_{\text{comp}}) \quad (0, 0) \in S(x) \times T^{-1}(z) + \tilde{\partial}R(x, z),$$

onde R é a função sela definida como

$$R(x, z) = \begin{cases} H(x)^T z - \delta_K(z), & x \in dom(H) \\ +\infty, & \text{caso contrario.} \end{cases}$$

com $\tilde{\partial}R(x, z) = \partial_x R(x, z) \times \partial_z(-R)(x, z)$.

O modelo de dualidade de composição, dado na Proposição 2.2.1, corresponde ao contexto de dualidade geral da Definição 1.3.1 com

$$\begin{aligned} F_0(x) &= S(x) + DH(x)^T H(x), \\ F(x, z) &= \begin{pmatrix} S(x) \\ 0 \end{pmatrix} + \begin{pmatrix} DH(x)^T \\ I \end{pmatrix} (T(H(x) + z)), \\ L &= S \times T^{-1} + \tilde{\partial}R. \end{aligned} \tag{2.2.1}$$

Nestas considerações, designamos o problema primal, Lagrangeano e dual, no contexto de dualidade de composição por

$$0 \in F_0(x) = S(x) + DH(x)^T T(H(x)), \quad (2.2.2)$$

$$(0, 0) \in L(x, z) = \begin{pmatrix} S(x) \\ T^{-1}(z) \end{pmatrix} + \begin{pmatrix} DH(x)^T z \\ -H(x) \end{pmatrix}, \quad (2.2.3)$$

$$0 \in G_0(z), \quad (2.2.4)$$

onde $G_0(z)$ é definido em (1.3.9).

Cabe observar que nos casos gerais a aplicação G_0 não é uma expressão explícita em termos de S , T e H como as aplicações F_0 e L , mas ela pode ser calculada de acordo com a fórmula dada na equação (1.3.9), o qual vai depender da forma do lagrangeano.

Exemplo 2.2.2 *De novo, consideremos o problema (PNL) apresentado na Seção 1.2 do Capítulo 1. No contexto de programação de composição, com H e K como foram definidos em (1.2.3) e (1.2.4), o problema (PNL) pode ser reformulado como segue:*

$$\min \{f_0(x) + \delta_K(H(x)), \quad x \in \mathbb{R}^n\}.$$

Observe que uma condição de otimalidade desse problema é dado por

$$0 \in \nabla f_0(x) + DH(x)^T N_K(H(x)),$$

no qual corresponde ao problema primal no contexto de dualidade em programação composta se substituirmos $S(x) = \nabla f_0(x)$ e $T(x) = N_K(x)$ em (2.2.2).

Além disso, como $T(x) = N_K(x)$, segundo o Lema 1.1.16, temos que $T^{-1}(z) = N_{K^\otimes}(z)$. Então o problema lagrangeano no contexto de dualidade de composição, segundo (2.2.3), fica como

$$(0, 0) \in L(x, z) = \nabla f_0(x) \times N_{K^\otimes}(z) + (DH(x)^T z, -H(x)). \quad (2.2.5)$$

Observe que este problema coincide com o sistema KKT para o problema (PNL) dado em (1.2.6).

Capítulo 3

O algoritmo do ponto proximal (APP)

Neste Capítulo apresentaremos o algoritmo do ponto proximal clássico de Rockafellar[36] para logo estudar uma nova proposta de generalização deste algoritmo: o algoritmo do ponto proximal relaxado de Eckstein e Bertsekas[9], o qual é um algoritmo aplicado também para operadores monótonos maximais. Depois, analisaremos a versão local deste algoritmo, no qual, se tivermos problemas com operadores gerais então procuraremos ferramentas para ainda aplicar o algoritmo. Muitos dos resultados desse capítulo foram obtidos, principalmente, de Bertsekas[5], Eckstein e Bertsekas[9], Gol'shtein[13], Iusem[19], Pennanen[29], Rockafellar[32, 36, 37], entre outros.

Considere o problema de otimização sem restrições

$$\min\{f(z) : z \in \mathbb{R}^n\} \quad (3.0.1)$$

onde f é uma função convexa e limitada inferiormente.

Um dos métodos mais populares para resolver este problema convexo e não necessariamente diferenciável é o algoritmo do ponto proximal (APP). O APP clássico, para resolver o problema (3.0.1), gera uma sequência de pontos $\{z^k\} \subset \mathbb{R}^n$, a partir de $z^0 \in \mathbb{R}^n$, dado como

$$z^{k+1} \in \arg \min \left\{ f(z) + \frac{\lambda_k}{2} \|z - z^k\|^2, \quad z \in \mathbb{R}^n \right\}, \quad (3.0.2)$$

onde λ_k é um parâmetro positivo tal que $\lambda_k \in (0, \bar{\lambda}]$ para algum $\bar{\lambda} > 0$.

Cabe observar que

$$f_{\lambda_k}(z) = f(z) + \frac{\lambda_k}{2} \|z - z^k\|^2$$

é a regularização Moreau-Yosida de f , onde se espera que f_{λ_k} forneça algum tipo de aproximação do minimizador, caso existir, da função objetivo f .

Observe que, segundo (3.0.2), z^{k+1} são os pontos proximais de z^k com respeito à função f_{λ_k} , isto é, a iterada de ordem $k + 1$ do método de ponto proximal pode ser reescrita como segue

$$z^{k+1} \in \text{prox}_{\lambda_k, f}(z^k).$$

Mais ainda, se f é limitada inferiormente e semi-contínua inferior então a sequência $\{z^k\}$ gerada por (3.0.2), está bem definida. Além disso, é bem conhecido na literatura que sob algumas hipóteses razoáveis a sequência gerada por (3.0.2), converge para um minimizador de f . Uma abordagem para esta prova de convergência baseia-se na noção de Fejér convergente e é dado como segue.

Definição 3.0.3 *Seja $\{z^k\} \subset \mathbb{R}^n$ e um conjunto $U \subset \mathbb{R}^n$. Se diz que $\{z^k\}$ é Fejér convergente em U se para todo $z \in U$ temos*

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

para todo $k \in \mathbb{N}$.

Proposição 3.0.4 [19, Proposição 2.1] *Seja $U \subset \mathbb{R}^n$ um conjunto não vazio. Se $\{z^k\}$ é Fejér convergente em U então $\{z^k\}$ é limitada. Além disso, se existir um ponto de acumulação \bar{z} de $\{z^k\}$ tal que $\bar{z} \in U$, então toda sequência $\{z^k\}$ converge para o ponto limite \bar{z} .*

Nestas condições, se f for uma função convexa, limitada inferiormente, semi-contínua inferior e assumindo que o conjunto de minimizadores de f em \mathbb{R}^n , X^* , é não vazia então a prova de convergência de $\{z^k\}$, gerado por (3.0.2), para um ponto de X^* pode ser esquematizado conforme os seguintes passos:

- i. A boa definição do algoritmo, isto é, a existência das sequências $\{z^k\}$.

- ii. A sequência $\{z^k\}$ é Fejér convergente, o que implica que $\{z^k\}$ é limitada.
- iii. $\lim_{k \rightarrow \infty} (x^{k+1} - x^k) = 0$.
- iv. $\{z^k\}$ tem pontos de acumulação e todos eles pertencem a X^* .

Por outro lado, se f é convexa e o operador $T = \partial f$ temos que $x \in \text{Zer}(T)$ se, e somente se, x é minimizador de f . De fato, para $x \in \text{Zer}(T)$ temos que para todo $y \in \mathbb{R}^n$

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

o que implica que $f(x) \leq f(y)$ para todo $y \in \mathbb{R}^n$. Então o problema de encontrar zeros de operadores monótonos maximais generaliza o problema de minimização convexa.

Além disso, de acordo com o Teorema 1.1.5, o APP para solucionar o problema de minimização (3.0.1) pode ser caracterizado por

$$0 \in \partial f(z^{k+1}) + \lambda_k (z^{k+1} - z^k). \quad (3.0.3)$$

Segundo o Exemplo 2.1.5, ∂f é monótono maximal, o que conduz a uma extensão natural o qual é equivalente a

$$z^{k+1} \in (I + \frac{1}{\lambda_k} T)^{-1}(z^k) \quad (3.0.4)$$

para todo $k \in N$ com $T = \partial f$.

Mais ainda, da monotonicidade de T e da Proposição 2.1.41 temos que o operador $(I + \frac{1}{\lambda_k} T)^{-1}$ é simples valorado, isto é,

$$z^{k+1} = J_{c_k}^T(z^k),$$

onde $J_{c_k}^T$ é chamado de operador resolvente de T para uma sequência $\{c_k\}$ de números reais positivos tal que $c_k = 1/\lambda_k$. Vale a pena mencionar que Minty[24] provou que a sequência $\{z^k\}$ existe e é única devido ao fato da sobrejetividade e injetividade do operador $J_{c_k}^T$, respectivamente.

A seguir, vamos dar uma interpretação geométrica do comportamento do algoritmo para operadores monótonos maximais reais. Se $T : \mathbb{R} \rightarrow \mathbb{R}$ temos que

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

com $y^{k+1} = T(z^{k+1})$.

Como λ_k é um parâmetro positivo, o ângulo de inclinação da reta que passa por o ponto $(z^k, 0)$ é obtuso, isto significa que o par (z^{k+1}, y^{k+1}) , se aproxima, em cada iteração para o $(\bar{z}, 0)$ com $\bar{z} \in \text{Zer}(T)$.

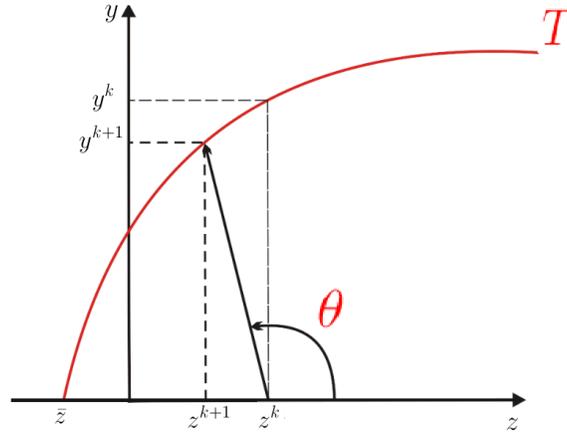


Figura 3.1: O ângulo de inclinação da reta que pasa por o ponto $(z^k, 0)$ é obtuso.

Por conseguinte, segundo o APP, o problema de encontrar zeros de operadores monótonos maximais

$$0 \in T(z) \quad (3.0.5)$$

é substituído por uma sequência de subproblemas dados por

$$0 \in T(z) + c_k^{-1}(z - z^k). \quad (3.0.6)$$

De fato, este algoritmo iterativo pode ser visto como uma regularização, possivelmente mal-comportado, do operador T como veremos no próximo exemplo: Seja $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que $T(z) = Az - b$ com $A \in \mathbb{R}^{n \times n}$ semidefinida positiva e $b \in \mathbb{R}^n$. Suponhamos que A for singular e $c_k^{-1} \rightarrow \infty$. Nestas condições, temos que o subproblema do APP dado por

$$c_k^{-1}z^k + b \in (A + c_k^{-1}I)(z)$$

é bem-comportado do fato que a condição de número da matriz $A + c_k^{-1}I$, $k(A + c_k^{-1}I)$, converge para 1, embora o problema original de solucionar o sistema linear $b \in Az$ não .

É importante mencionar que solucionar o problema (3.0.6), muitas vezes o algoritmo, no ponto de vista de esforço computacional, torna-se demorado e caro. Uma alternativa para esses obstáculos é considerar uma aproximação das soluções desse problema o qual é obtido por algum método iterativo eficiente. Daí, uma sequência de erro é relacionada com as soluções do problema (3.0.6). Obtendo assim uma versão inexata do APP. Vale a pena mencionar que trabalhos anteriores ao Rockafellar[36] basearam os estudos do APP na “forma exata” no qual as sequências de erro é considerado 0 para todo k .

Então, o Algoritmo de Ponto Proximal clássico (APP) foi proposto por Martinet[23] em 1970 como um método de regularização no contexto de otimização convexa em espaços de Hilbert e desenvolvido logo depois por Rockafellar[36] para solucionar o problema de otimização convexa (3.0.1) e para obter zeros de operadores monótonos maximais em espaços de Hilbert \mathcal{H} . A seguir apresentaremos um resultado de Rockafellar[36, Teorema 2], que foi dado em espaços de Hilbert \mathcal{H} no qual mostra o método APP para encontrar zeros de operadores monótonos maximais para resolver o problema (3.0.5), garantindo também a convergência linear desse algoritmo baseado na abordagem da k -Lipschitz contínua de T^{-1} em 0 ao invés da monotonicidade forte de T o qual está mais orientado para programação convexa.

Definição 3.0.5 *Dizemos que T^{-1} é k -Lipschitz contínua em $0 \in \mathbb{R}^n$ com módulo $k \geq 0$, se $Zer(T) = \{\bar{z}\}$ e existe $\tau > 0$ tal que se $z \in T^{-1}(w)$ com $\|w\| \leq \tau$ temos que*

$$\|z - \bar{z}\| \leq k \|w\|. \quad (3.0.7)$$

Teorema 3.0.6 [36, Teorema 2] *Seja $T : \mathcal{H} \rightrightarrows \mathcal{H}$ um operador monótono maximal. Considere que o APP gera uma sequência $\{z^k\}$, a partir de um ponto inicial arbitrário $z^0 \in \mathcal{H}$, dada como*

$$z^{k+1} \approx J_{c_k}^T(z^k) \quad (3.0.8)$$

tal que

$$\|z^{k+1} - J_{c_k}^T(z^k)\| \leq e_k,$$

onde a sequência escalar $\{e_k\}$ satisfaz que $\sum_{k=0}^{\infty} e_k < \infty$ e $\{c_k\}$ é uma sequência positiva não decrescente tal que $c_k \uparrow \bar{c} \geq +\infty$.

Suponha também que a sequência $\{z^k\}$ é gerada pelo APP tal que

$$\|z^{k+1} - J_{c_k}^T(z^k)\| \leq \delta_k \|z^{k+1} - z^k\|,$$

onde as sequências escalares $\{\delta_k\}$ e $\{c_k\}$ satisfazem $\sum_{k=0}^{\infty} \delta_k < \infty$ e $c_k \uparrow c \leq \infty$.

Além disso, assumamos que $\{z^k\}$ é limitada, o que ocorre sempre que $\text{Zer}(T) \neq \emptyset$, e que T^{-1} é a -Lipschitz contínua em 0 com $a > 0$.

Seja

$$u_k = \frac{a}{\sqrt{a^2 + c_k^2}} < 1.$$

Então, a sequência $\{z^k\}$ converge fortemente para a única solução \bar{z} para (3.0.5) com

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

para todo $k \geq k'$, onde

$$0 \leq \alpha_k = \frac{u_k + \delta_k}{1 - \delta_k} < 1.$$

Rockafellar[36] também mostrou que se e^k converge para 0 suficientemente rápido tal que

$$\sum_{k=1}^{+\infty} \|e^k\| < \infty,$$

então a sequência $\{z^k\}$ converge fracamente para um zero de T . Além disso, a convergência forte do algoritmo é garantido se $\{z^k\}$ é limitado, $\{c_k\}$ é uma sequência não decrescente tal que $c_k \uparrow \bar{c} \leq \infty$ e T^{-1} é a -Lipschitz contínua, ver Rockafellar[36, Teorema 2], ou se T é fortemente monótono com módulo $a > 0$ sem necessidade de $\{z^k\}$ ser limitada, ver Rockafellar[36, Proposição 5]. A importância de convergência forte é sublinhada por Güler[15], onde uma função convexa f é minimizada pelo APP: neste é mostrado que a taxa de convergência da sequência $\{f(z^k)\}$ é melhor quando $\{z^k\}$ converge fortemente do que quando convergir fracamente. Mas a hipótese que T é fortemente monótona excluem algumas aplicações mais importantes tais como os problemas típicos de programação convexa. Por outro lado, Güler[15] construiu um exemplo mostrando que o algoritmo dado por Rockafellar em (3.0.8) para espaços de Hilbert na versão exata não converge fortemente em geral. Nesta situação, como a convergência fraca não é suficiente para obter um algoritmo eficiente, robusto e com baixo tempo de execução e tendo em

conta que o APP não converge fortemente em geral é que muitos dos pesquisadores tem trabalhado para encontrar algoritmos que sempre convergem fortemente, ou pelo menos modificar o algoritmo de Rockafellar de tal forma que a convergência forte seja garantida. Uma de tais modificações foi obtida por Solodov e Svaiter[40].

Além disso, cabe destacar que a análise da taxa de convergência do APP foi melhorada por Luque[21] permitindo $T^{-1}(0)$ ser um conjunto como é dado no seguinte resultado.

Teorema 3.0.7 [21, Teorema 2.1] *Seja $\{c_k\}$ uma sequência não crescente de constantes positivas e $\{z^k\}$ as sequências geradas pelo APP. Suponhamos que temos uma vizinhança Y da origem e uma constante positiva $a \in \mathbb{R}$ tal que*

$$y \in Y, \quad y \in T(z) \rightarrow \text{dist}\{z, T^{-1}(0)\} \leq a\|y\|.$$

Então, $\text{dist}\{z^k, T^{-1}(0)\}$ converge linearmente para 0 com um taxa de convergência limitada inferiormente por

$$\frac{a}{\sqrt{a^2 + (\lim c_k)^2}}.$$

Vale a pena mencionar que Brézis e Lions[7] fizeram alguns refinamentos para a prova de convergência do APP.

Por outro lado, ao longo do tempo muitas variantes do APP foram consideradas, uma de tais considerações que se realizou ao APP foi a relaxação desse método, o qual é o resultado da combinação do APP clássico do Rockafellar com um algoritmo de Gol'shtein e Tret'yakov[14] que considera as iterações dadas por

$$z^{k+1} = (\rho_k J_c^T + (1 - \rho_k)I)(z^k),$$

onde em comparação com o algoritmo de Rockafellar[36], o parâmetro c não varia em cada passo k e não considera os casos nos quais $Zer(T) = \emptyset$, mas sim permite resolver aproximadamente. Assim, foi introduzido o APP relaxado para operadores monótonos maximais multivalorados, o qual, a partir de um ponto arbitrário z^0 , uma sequência dada por

$$z^{k+1} = (\rho_k J_{c_k}^T + (1 - \rho_k)I)(z^k),$$

onde $\rho_k \in (0, 2)$ é o parâmetro de relaxação

3.1 O Algoritmo do Ponto Proximal Relaxado (APPR)

Nesta seção apresentamos o APP relaxado, o qual foi desenvolvido por Eckstein e Bertsekas[9] em espaços de Hilbert real \mathcal{H} . Para os resultados e a análise destes nesta seção, será considerado o espaço euclidiano. Também, sob certas hipóteses mostraremos a convergência linear do APP relaxado. Além disso, ao longo desta seção vamos denotar os operadores $J_k^T = J_{c_k}^T$ e $Q_k^T = Q_{c_k}^T$ como foram definidos em (2.1.11) e (2.1.12), respectivamente.

Lema 3.1.1 *Seja $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ um operador monótono maximal e seja $\{z^k\}$ tal que*

$$z^{k+1} = (1 - \rho_k)z^k + \rho_k w^k$$

onde

$$\|w^k - J_k^T(z^k)\| \leq \epsilon_k, \quad \forall k \geq 0,$$

e $\{\epsilon_k\}, \{\rho_k\}, \{c_k\} \subset [0, +\infty)$ são sequências tais que

$$E_1 = \sum_{k=0}^{+\infty} \epsilon_k < \infty, \quad \Delta_1 = \inf_{k \geq 0} \rho_k > 0, \quad \Delta_2 = \sup_{k \geq 0} \rho_k < 2, \quad \bar{c} = \inf_{k \geq 0} c_k > 0.$$

Tal sequência $\{z^k\}$ gerada pelo algoritmo acima, é denominado ponto proximal relaxado. Se $Zer(T) \neq \emptyset$ então $\{z^k\}$ é uma sequência limitada. Mais ainda, para todo $k \in \mathbb{N}$ temos que $J_k^T(z^k)$ é limitado.

Prova. Primeiro vamos provar que $\{z^k\}$ é limitada sempre que $Zer(T) \neq \emptyset$. Para isso, definimos

$$\bar{z}^{k+1} = (1 - \rho_k)z^k + \rho_k J_k^T(z^k),$$

para todo k , isto é,

$$\bar{z}^{k+1} = (I - \rho_k Q_k^T)(z^k).$$

Seja $z^* \in Zer(T)$. De (2.1.13), temos que

$$\begin{aligned} \|\bar{z}^{k+1} - z^*\|^2 &= \|(I - \rho_k Q_k^T)(z^k) - z^*\|^2 \\ &= \|z^k - z^* - \rho_k Q_k^T(z^k)\|^2 \\ &= \|z^k - z^*\|^2 + \rho_k^2 \|Q_k^T(z^k)\|^2 - 2\rho_k \langle z^k - z^*, Q_k^T(z^k) - Q_k^T(z^*) \rangle. \end{aligned}$$

Como Q_k^T é um operador firmemente não-expansivo, segundo a Observação 2.1.45, obtemos

$$\begin{aligned} \|\bar{z}^{k+1} - z^*\|^2 &\leq \|z^k - z^*\|^2 + \rho_k^2 \|Q_k^T(z^k)\|^2 - 2\rho_k \|Q_k^T(z^k)\|^2 \\ &= \|z^k - z^*\|^2 + \rho_k(\rho_k - 2) \|Q_k^T(z^k)\|^2. \end{aligned} \quad (3.1.1)$$

Além disso, pela hipótese

$$0 < \Delta_1 \leq \rho_k, \quad (3.1.2)$$

$$2 > \Delta_2 \geq \rho_k \iff 2 - \rho_k \geq 2 - \Delta_2 > 0. \quad (3.1.3)$$

De (3.1.2) e (3.1.3),

$$\rho_k(\rho_k - 2) \leq \Delta_1(\Delta_2 - 2) < 0. \quad (3.1.4)$$

Então, de (3.1.1) e (3.1.4),

$$\|\bar{z}^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 + \Delta_1(\Delta_2 - 2) \|Q_k^T(z^k)\|^2 \quad (3.1.5)$$

$$\leq \|z^k - z^*\|^2. \quad (3.1.6)$$

Para todo $k \geq 0$, temos que

$$\begin{aligned} \|z^{k+1} - \bar{z}^{k+1}\| &= \|(1 - \rho_k)z^k + \rho_k w^k - ((1 - \rho_k)z^k + \rho_k J_k^T(z^k))\| \\ &= \rho_k \|w^k - J_k^T(z^k)\| \\ &\leq \rho_k \epsilon_k. \end{aligned} \quad (3.1.7)$$

De (3.1.6) e (3.1.7),

$$\|z^{k+1} - z^*\| \leq \|z^{k+1} - \bar{z}^{k+1}\| + \|\bar{z}^{k+1} - z^*\| \leq \rho_k \epsilon_k + \|z^k - z^*\|.$$

Calculando o somatório em ambos os lados da desigualdade, para todo k , temos

$$\begin{aligned} \|z^{k+1} - z^*\| &\leq \|z^0 - z^*\| + \sum_{i=0}^k \rho_i \epsilon_i \\ &\leq \|z^0 - z^*\| + \sum_{i=0}^k \sup\{\rho_i\} \epsilon_i \\ &\leq \|z^0 - z^*\| + 2 \sum_{i=0}^k \epsilon_i \\ &= M, \end{aligned} \quad (3.1.8)$$

onde $M = \|z^0 - z^*\| + 2E_1$. Logo, $\{z^k\}$ é uma sequência limitada.

Por outro lado, vamos provar que $J_k^T(z^k)$ é limitado. De fato, como $\{z^k\}$ é limitada, existe $\epsilon > 0$ tal que $\|z^k\| < s$ para todo $k \in \mathbb{N}$. Consideremos

$$\bar{e} = \sup_{k \geq 0} \{\epsilon_k\}, \quad e \quad r = \frac{2s}{\min\{1, \Delta_1\}} + \bar{e} + 1$$

então

$$\|z^k\| < s < \frac{2s}{\min\{1, \Delta_1\}} = r - 1 - \bar{e} \leq r - 1. \quad (3.1.9)$$

Como $w^k = \frac{1}{\rho_k} (z^{k+1} - (1 - \rho_k)z^k)$ temos que

$$\begin{aligned} \|w^k\| &= \frac{1}{\rho_k} \|z^{k+1} - (1 - \rho_k)z^k\| \\ &\leq \frac{1}{\rho_k} (\|z^{k+1}\| + |\rho_k - 1| \|z^k\|) \\ &< \frac{1}{\Delta_1} (s + s) = \frac{2s}{\Delta_1}. \end{aligned} \quad (3.1.10)$$

Desde que $\min\{1, \Delta_1\} \leq \Delta_1$ e de (3.1.9) temos que

$$\frac{2s}{\Delta_1} \leq \frac{2s}{\min\{1, \Delta_1\}} \leq r - 1. \quad (3.1.11)$$

Logo, de (3.1.10) e (3.1.11) obtemos que

$$\|w^k\| < \frac{2s}{\Delta_1} \leq r - 1.$$

Daí, como $\| \|w^k\| - \|J_k^T(z^k)\| \| \leq \|w^k - J_k^T(z^k)\| \leq \epsilon_k$, para todo k , temos que

$$\|J_k^T(z^k)\| \leq \|w^k\| + \epsilon_k < \frac{2s}{\Delta_1} + \bar{e} \leq r - 1. \quad \blacksquare$$

Teorema 3.1.2 *Nas condições do Lema 3.1.1, se T possui algum zero, $\{z^k\}$ converge para um zero de T . Se T não tem zeros, então $\{z^k\}$ será uma sequência ilimitada.*

Demonstração. Primeiro, vamos analisar o caso de T possuir algum zero. De (3.1.5), (3.1.6), (3.1.7) e (3.1.8) temos que

$$\begin{aligned} \|z^{k+1} - z^*\|^2 &= \|(\bar{z}^{k+1} - z^*) + (z^{k+1} - \bar{z}^{k+1})\|^2 \\ &\leq \|\bar{z}^{k+1} - z^*\|^2 + \|z^{k+1} - \bar{z}^{k+1}\|^2 + 2\|\bar{z}^{k+1} - z^*\| \|z^{k+1} - \bar{z}^{k+1}\| \\ &\leq \|z^k - z^*\|^2 + \Delta_1(\Delta_2 - 2) \|Q_k^T(z^k)\|^2 + (\rho_k \epsilon_k)^2 + 2\|z^k - z^*\| \rho_k \epsilon_k \\ &\leq \|z^k - z^*\|^2 + (\rho_k \epsilon_k)^2 + \Delta_1(\Delta_2 - 2) \|Q_k^T(z^k)\|^2 + 2\rho_k \epsilon_k M. \end{aligned} \quad (3.1.12)$$

Como $\sum_{k=0}^{+\infty} \epsilon_k$ é convergente então $\sum_{k=0}^{+\infty} \epsilon_k^2$ é convergente. Calculando o somatório em ambos os lados da desigualdade, para todo k , temos

$$\begin{aligned} \|z^{k+1} - z^*\|^2 &\leq \|z^0 - z^*\|^2 + 4 \sum_{i=0}^k \epsilon_i^2 + \Delta_1(\Delta_2 - 2) \sum_{i=0}^k \|Q_i^T(z^i)\|^2 + 4M \sum_{i=0}^k \epsilon_i \\ &\leq \|z^0 - z^*\|^2 + 4E_2 + \Delta_1(\Delta_2 - 2) \sum_{i=0}^k \|Q_i^T(z^i)\|^2 + 4ME_1, \end{aligned}$$

onde $E_2 = \sum_{i=0}^k \epsilon_i^2$.

Daí,

$$\Delta_1(2 - \Delta_2) \sum_{i=0}^k \|Q_i^T(z^i)\|^2 \leq \|z^0 - z^*\|^2 + 4(E_2 + ME_1) - \|z^{k+1} - z^*\|^2,$$

e segundo o Lema 3.1.1 temos que $\{z^k\}$ é limitado. Daí,

$$\Delta_1(2 - \Delta_2) \sum_{i=0}^k \|Q_i^T(z^i)\|^2$$

é limitado. Como $\sum_{i=0}^k \|Q_i^T(z^i)\|^2$ é monótono crescente e limitado existe o limite. Tomando limite, quando $k \rightarrow \infty$, temos

$$\lim_{k \rightarrow \infty} Q_k^T(z^k) = 0. \quad (3.1.13)$$

Como T é maximal, do Lema 2.1.12, z^k pode ser escrito de forma única como

$$z^k = x^k + c_k y^k, \quad (3.1.14)$$

onde $c_k > 0$ e $(x^k, y^k) \in T$. Então $z^k \in x^k + c_k T(x^k) = (I + c_k T)(x^k)$, isto é, $J_k^T(z^k) = x^k$, de onde

$$z^k - x^k = I(z^k) - J_k^T(z^k) = Q_k^T(z^k). \quad (3.1.15)$$

Logo, de (3.1.13) temos que

$$z^k - x^k \rightarrow 0.$$

Além disso, de (3.1.13), (3.1.14), (3.1.15) e como $\{c_k\}$ é uma sequência limitada inferiormente,

$$0 \leftarrow c_k^{-1} Q_k^T(z^k) = c_k^{-1} (z^k - x^k) = c_k^{-1} (x^k + c_k y^k - x^k) = y^k.$$

Como $\{z^k\}$ é uma sequência limitada, possui uma subsequência $\{z^{k_j}\} \subset \{z^k\}$ convergente, isto é, existe \bar{z}^∞ tal que $z^{k_j} \rightarrow \bar{z}^\infty$. Logo, $x^{k_j} \rightarrow \bar{z}^\infty$ do fato que $z^k - x^k \rightarrow 0$.

Seja (x, y) qualquer ponto em T . Pela monotonicidade do operador T temos que para todo k ,

$$\langle x - x^k, y - y^k \rangle \geq 0,$$

em particular, $\langle x - x^{k_j}, y - y^{k_j} \rangle \geq 0$. Tomando limite, quando $j \rightarrow \infty$, temos $\langle x - \bar{z}^\infty, y \rangle \geq 0$. Como $(x, y) \in T$ foi tomado arbitrariamente temos que $(z, 0)$ deve estar em T . Portanto, temos que $(\bar{z}^\infty, 0) \in T$, isto é, $\bar{z}^\infty \in \text{Zer}(T)$.

A seguir, vamos provar que existe um único ponto de acumulação de $\{z^k\}$ gerada pelo APP. De fato, consideremos qualquer zero z^* do operador T e de (3.1.12), para todo $k \in \mathbb{N}$, temos

$$\|z^{k+1} - z^*\|^2 \leq \|z^k - z^*\|^2 + r_k$$

onde $r_k = (\rho_k \epsilon_k)^2 + \Delta_1(\Delta_2 - 2) \|Q_k^T(z^k)\|^2 + 2\rho_k \epsilon_k M$.

Como $\sum_{i=0}^{\infty} r_k < \infty$, temos que $\|z^{k+1} - z^*\|$ converge para um ponto limite finito e não negativo. Em particular para $z^* = \bar{z}^\infty$ e do fato que existe $\{z^{k_j}\} \rightarrow \bar{z}^\infty$, temos que $\|z^{k_j} - \bar{z}^\infty\| \rightarrow 0$. Logo, $\|z^{k+1} - \bar{z}^\infty\|$ converge para zero, o que implica que $z^k \rightarrow \bar{z}^\infty$. Daí, o ponto de acumulação da sequência $\{z^k\}$ é única.

Por outro lado, vamos analisar o caso quando T não possui zero nenhum. Suponhamos que $\{z^k\}$ é uma sequência limitada e definimos a função $h : \mathbb{R}^n \rightarrow [0, +\infty]$ tal que

$$h(x) = \begin{cases} 0, & \|x\| \leq r, \\ +\infty, & \|x\| > r. \end{cases}$$

Do Exemplo 1.1.4, o subdiferencial da função h , ∂h , é dado como

$$\partial h(x) = \begin{cases} 0, & \|x\| < r, \\ ax, a \geq 0 & \|x\| = r, \\ \emptyset, & \|x\| > r. \end{cases}$$

Observe que

$$\text{int}(\text{dom}(\partial h)) = \{x \mid \|x\| < r\} = B_r(0) \quad (3.1.16)$$

Seja $T' = T + \partial h$ tal que

$$T'(x) = \begin{cases} T(x), & \|x\| < r, \\ \{y + ax : y \in T(x), a \geq 0\}, & \|x\| = r, \\ \emptyset, & \|x\| > r. \end{cases}$$

Tomemos $y^k = J_k^T(z^k)$ então $z^k = (J_k^T)^{-1}(y^k)$ se, e somente se,

$$z^k - y^k = c_k T(y^k),$$

isto é, $y^k \in \text{dom}(T)$ e do Lema 3.1.1, $J_k^T(z^k) \leq r - 1$, temos que

$$y^k \in B_r(0) \cap \text{dom}(T). \quad (3.1.17)$$

Logo, de (3.1.16) e (3.1.17),

$$\text{int}(\text{dom}(\partial h)) \cap \text{dom}(T) \neq \emptyset.$$

Ademais, segundo o Teorema 2.1.26, o operador $T' = T + \partial h$ é uma aplicação monótona maximal. Da maximalidade de T' e como $\text{dom}(T') = B(x, r)$ é limitado, então pelo Lema 2.1.20 temos que existe pelo menos um z tal que $0 \in T(z)$. Então, a sequência $\{z^k\}$ cumpre a iteração do ponto proximal para T' , isto significa que para todo $k \in N$ temos a iteração

$$z^{k+1} = (1 - \rho_k) z^k + \rho_k w^k,$$

onde

$$\|w^k - J_k^{T'}(z^k)\| \leq \epsilon_k$$

com $J_k^{T'} = (I + cT')^{-1}$.

Pela razão da primeira parte deste teorema, $\{z^k\}$ converge para algum zero $\bar{z}^\infty \in T'$. Além disso, como $\|z^k\| \leq r - 1$ para todo k , $\|\bar{z}^\infty\| < r$. Pela definição de T' , temos que $T'(\bar{z}^\infty) = T(\bar{z}^\infty)$ onde $\bar{z}^\infty \in \text{Zer}(T)$. O que é uma contradição do fato que foi suposto que $\text{Zer}(T) \neq \emptyset$. Portanto, a sequência $\{z^k\}$ não é limitada. ■

A seguir, vamos dar uma interpretação do papel que cumpre o parâmetro de relaxação no algoritmo $\rho_k \in (0, 2)$. Para isto, vamos considerar que os resolventes do operador T

quando aplicado a z^k podem ser achados em forma exata. Então, as iterações do APPR são da seguinte forma

$$z^{k+1}(\rho) = (1 - \rho)(z^k) + \rho(I + c_k T)^{-1}(z^k) = (1 - \rho)(z^k) + \rho w^k.$$

Observe que para $\rho = 1$ temos que $z^{k+1}(1) = w^k$ e para $\rho = 0$ temos que $z^{k+1}(0) = z^k$. Como $w^k = (I + c_k T)^{-1}(z^k)$ temos que

$$\frac{z^k - w^k}{c_k} \in T(w^k).$$

Pela monotonicidade de T , para $(w^k, \frac{z^k - w^k}{c_k})$, $(\bar{z}^\infty, 0) \in T$ com $\bar{z}^\infty \in \text{Zer}(T)$ e do fato que c_k é um parâmetro positivo temos que $\langle \bar{z}^\infty - w^k, z^k - w^k \rangle \leq 0$, o que implica que

$$\|z^{k+1}(1) - \bar{z}^\infty\| = \|w^k - \bar{z}^\infty\| \leq \|z^{k+1}(\rho) - \bar{z}^\infty\|$$

para todo $\rho < 1$. Conclui-se que sendo $\text{Zer}(T)$ o conjunto solução, se para todo $\bar{z}^\infty \in \text{Zer}(T)$ tal que

$$\langle \bar{z}^\infty - w^k, z^k - w^k \rangle < 0$$

então é válido o desenho ??, isto é, existe algum $\bar{\rho} \in (1, 2)$ tal que

$$\text{dist}(z^{k+1}(\bar{\rho}), \text{Zer}(T)) < \text{dist}(w^k, \text{Zer}(T)).$$

A eleição do parâmetro $\rho \in (1, 2)$ dá lugar aos métodos denominados de sobre-relaxação e serve para acelerar a convergência do método. Por este motivo vantajoso é de interesse estudar o método sobre-relaxado do APP.

Por outro lado, com as mesmas hipóteses utilizadas em Rockafellar[36]: a condição de Lipschitz contínua do operador T^{-1} numa vizinhança de zero e uma pequena tolerância ao erro, temos o seguinte resultado da taxa de convergência para o algoritmo sobre-relaxado.

Proposição 3.1.3 *Com as condições do Teorema 3.1.2, assumamos que $\rho_k \geq 1$, $c_k \nearrow \bar{c} \leq \infty$, e*

$$\epsilon_k \leq \delta_k \|w^k - z^k\|, \quad (3.1.18)$$

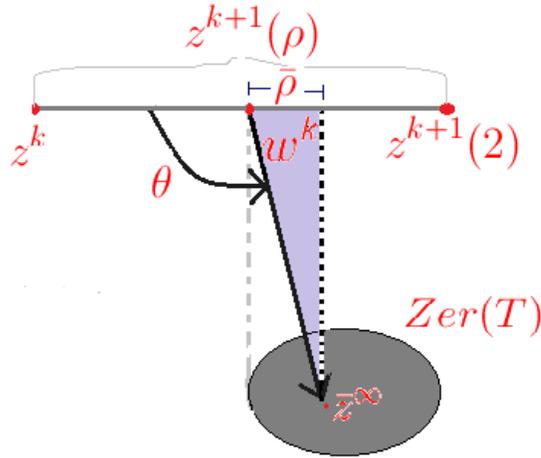


Figura 3.2: Como consequência da monotonicidade de T , o ângulo θ é obtuso.

onde $\delta_k \rightarrow 0$. Se \bar{z} é a única solução de $0 \in T(z)$ e que o operador T^{-1} é a -Lipschitz contínuo numa vizinhança do zero com $a \geq 0$, então a sequência $\{z^k\}$ converge linearmente a \bar{z} com uma taxa de

$$\sqrt{1 - \bar{\rho}(2 - \bar{\rho}) \frac{\bar{c}^2}{a^2 + \bar{c}^2}} < 1,$$

onde $\bar{\rho} = \limsup_{k \rightarrow \infty} \rho_k$.

Prova. Seja $\{z^k\}$ uma sequência gerada pelo APP. Da equação (3.1.13),

$$z^k - J_k^T(z^k) = Q_k^T(z^k) \rightarrow 0,$$

e do fato que $\bar{c} = \inf c_k > 0$ temos que

$$0 \leq \left\| \frac{z^k - J_k^T(z^k)}{c_k} \right\| \leq \frac{\|z^k - J_k^T(z^k)\|}{\bar{c}} \rightarrow 0,$$

isto é,

$$c_k^{-1} Q_k^T(z^k) = \frac{z^k - J_k^T(z^k)}{c_k} \rightarrow 0.$$

Logo, para k suficientemente grande,

$$c_k^{-1} \|Q_k^T(z^k)\| \leq \epsilon.$$

Da Proposição 2.1.46, item *a.*,

$$J_k^T(z^k) \in T^{-1}(c_k^{-1} Q_k^T(z^k)),$$

e pela condição de a -Lipschitz contínua em 0 do operador T^{-1} ,

$$\|J_k^T(z^k) - \bar{z}\|^2 \leq a^2 \|c_k^{-1} Q_k^T(z^k)\|^2. \quad (3.1.19)$$

Da Proposição 2.1.46, item b . e de (2.1.13)

$$\begin{aligned} \|z^k - \bar{z}\|^2 &\geq \|J_k^T(z^k) - J_k^T(\bar{z})\|^2 + \|Q_k^T(z^k) - Q_k^T(\bar{z})\|^2 \\ &\geq \|J_k^T(z^k) - \bar{z}\|^2 + \|Q_k^T(z^k)\|^2, \end{aligned}$$

então

$$\|Q_k^T(z^k)\|^2 \leq \|z^k - \bar{z}\|^2 - \|J_k^T(z^k) - \bar{z}\|^2. \quad (3.1.20)$$

De (3.1.19) e (3.1.20),

$$\|J_k^T(z^k) - \bar{z}\|^2 \leq \frac{a^2}{c_k^2} \left(\|z^k - \bar{z}\|^2 - \|J_k^T(z^k) - \bar{z}\|^2 \right),$$

rearranjando temos

$$\|J_k^T(z^k) - \bar{z}\|^2 \leq \frac{a^2}{c_k^2 + a^2} \|z^k - \bar{z}\|^2. \quad (3.1.21)$$

Mais uma vez, definimos $\bar{z}^{k+1} = (1 - \rho_k)z^k + \rho_k J_k^T(z^k)$, então

$$\begin{aligned} \|\bar{z}^{k+1} - \bar{z}\|^2 &= \|\rho_k J_k^T(z^k) + (1 - \rho_k)z^k - \bar{z}\|^2 \\ &= \rho_k^2 \|J_k^T(z^k) - \bar{z}\|^2 + 2\rho_k(1 - \rho_k) \langle J_k^T(z^k) - \bar{z}, z^k - \bar{z} \rangle \\ &\quad + (1 - \rho_k)^2 \|z^k - \bar{z}\|^2. \end{aligned} \quad (3.1.22)$$

Como $\rho_k \geq 1$, temos que $2\rho_k(1 - \rho_k) \leq 0$ e da Proposição 2.1.42, J_k^T é firmemente não-expansivo, então

$$\begin{aligned} 2\rho_k(1 - \rho_k) \langle J_k^T(z^k) - J_k^T(\bar{z}), z^k - \bar{z} \rangle &\leq 2\rho_k(1 - \rho_k) \|J_k^T(z^k) - J_k^T(\bar{z})\|^2 \\ &= 2\rho_k(1 - \rho_k) \|J_k^T(z^k) - \bar{z}\|^2. \end{aligned} \quad (3.1.23)$$

Substituindo (3.1.23) em (3.1.22),

$$\begin{aligned} \|\bar{z}^{k+1} - \bar{z}\|^2 &\leq \rho_k^2 \|J_k^T(z^k) - \bar{z}\|^2 + 2\rho_k(1 - \rho_k) \|J_k^T(z^k) - \bar{z}\|^2 + (1 - \rho_k)^2 \|z^k - \bar{z}\|^2 \\ &= \rho_k(2 - \rho_k) \|J_k^T(z^k) - \bar{z}\|^2 + (1 - \rho_k)^2 \|z^k - \bar{z}\|^2. \end{aligned} \quad (3.1.24)$$

Substituindo (3.1.21) em (3.1.24) temos

$$\begin{aligned} \|\bar{z}^{k+1} - \bar{z}\|^2 &\leq \rho_k(2 - \rho_k) \frac{a^2}{a^2 + c_k^2} \|z^k - \bar{z}\|^2 + (1 - \rho_k)^2 \|z^k - \bar{z}\|^2 \\ &= \left(\rho_k(2 - \rho_k) \frac{a^2}{a^2 + c_k^2} + (1 - \rho_k)^2 \right) \|z^k - \bar{z}\|^2. \end{aligned}$$

Seja

$$\alpha_k = \sqrt{\rho_k(2 - \rho_k) \frac{a^2}{a^2 + c_k^2} + (1 - \rho_k)^2} = \sqrt{1 - \frac{\rho_k(2 - \rho_k)c_k^2}{a^2 + c_k^2}},$$

então

$$\|\bar{z}^{k+1} - \bar{z}\| \leq \alpha_k \|z^k - \bar{z}\|, \quad (3.1.25)$$

onde

$$\begin{aligned} \|z^{k+1} - \bar{z}^{k+1}\| &= \|(\rho_k w^k + (1 - \rho_k)z^k) - (\rho_k J_k^T(z^k) + (1 - \rho_k)z^k)\| \\ &= \rho_k \|J_k^T(z^k) - w^k\| \\ &\leq \rho_k \epsilon_k. \end{aligned} \quad (3.1.26)$$

Além disso, de (3.1.18), (3.1.25) e (3.1.26),

$$\begin{aligned} \|z^{k+1} - \bar{z}\| &\leq \|\bar{z}^{k+1} - \bar{z}\| + \|z^{k+1} - \bar{z}^{k+1}\| \\ &= \|\bar{z}^{k+1} - \bar{z}\| + \rho_k \epsilon_k \\ &\leq \alpha_k \|z^k - \bar{z}\| + \rho_k \delta_k \|w^k - z^k\| \\ &= \alpha_k \|z^k - \bar{z}\| + \delta_k (\rho_k \|w^k - z^k\|) \\ &= \alpha_k \|z^k - \bar{z}\| + \delta_k \|z^{k+1} - z^k\| \\ &\leq \alpha_k \|z^k - \bar{z}\| + \delta_k \|z^{k+1} - \bar{z}\| + \delta_k \|\bar{z} - z^k\| \\ &= (\alpha_k + \delta_k) \|z^k - \bar{z}\| + \delta_k \|z^{k+1} - \bar{z}\|. \end{aligned}$$

Arranjando a desigualdade acima,

$$(1 - \delta_k) \|z^{k+1} - \bar{z}\| \leq (\alpha_k + \delta_k) \|z^k - \bar{z}\|,$$

então

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

onde

$$b_k = \frac{\alpha_k + \delta_k}{1 - \delta_k}.$$

Tomando limite, quando $k \rightarrow \infty$, temos que $\limsup b_k = \limsup \alpha_k$ pois, quando $k \rightarrow \infty$, $\delta_k \rightarrow 0$. Então

$$\limsup b_k = \sqrt{1 - \bar{\rho}(2 - \bar{\rho}) \frac{\bar{c}^2}{a^2 + \bar{c}^2}} < 1. \quad \blacksquare$$

Cabe observar que quando $\rho_k \in (1, 2)$ temos que $\bar{\rho} > 1$, então $\limsup_{k \rightarrow +\infty} b_k < 1$ mesmo que $\bar{c} = \infty$, o qual implica uma convergência linear para o algoritmo. Se $\rho_k = 1$ temos que $\bar{\rho} = 1$, o qual implica que $b_k \rightarrow 0$, obtendo assim uma convergência superlinear. Então, a melhor taxa de convergência para $\rho_k \in [1, 2)$ é obtida quando $\bar{\rho} = 1$.

3.2 O APP com monotonicidade local.

Muitas vezes exigir que o operador T do problema (3.0.5) seja monótono, com a finalidade de ainda aplicar o APP, é uma condição muito forte para muitas aplicações. Em geral, encontramos operadores que possuem monotonicidade pelo menos numa vizinhança. Esse fato sugere fazer uma análise de todos os resultados até agora obtidos no contexto local, mais uma vez, em espaços euclidianos. Para isso, precisamos do conceito de monotonicidade localmente maximal do operador multivalorado T sobre $X \times Y \subset \mathbb{R}^n \times \mathbb{R}^n$, onde X e Y são conjuntos abertos. Então, com a hipótese que o conjunto

$$T^{-1}(0) \cap X \neq \emptyset,$$

seja fechado e não vazio, e que o ponto inicial z^0 esteja suficientemente próximo do conjunto $T^{-1}(0) \cap X$, obtemos uma iteração que converge para um ponto que pertence ao conjunto $T^{-1}(0) \cap X$.

Com essas considerações, antes de apresentar o APP relaxado para operadores localmente monótonos maximais e a sua convergência local vamos apresentar dois resultados necessários para a prova deste resultado chave, para o qual será suficiente mostrar a convergência local do algoritmo sobre-relaxado, isto é, $\rho_k \in [1, 2)$.

Lema 3.2.1 *Seja $X, Y \subset \mathbb{R}^n$ subconjuntos abertos, T uma aplicação monótona maximal em $X \times Y$ e \bar{T} uma extensão monótona maximal de T tais que $0 \in Y$, $\bar{T}^{-1}(0) \subset X$, $\bar{T}^{-1}(0) = T^{-1}(0) \cap X$ é não vazio e fechado e seja $\{c_k\}$ tal que $\inf c_k > 0$, então existe $\epsilon > 0$ tal que para todo $z \in B(\bar{z}, \epsilon)$ temos que*

$$B(\bar{z}, \epsilon) \cap J_k^T(z) = J_k^{\bar{T}}(z),$$

onde $\bar{z} \in \bar{T}^{-1}(0)$.

Prova. De fato, como \bar{T} é uma extensão monótona maximal de T temos que

$$\text{graf}\bar{T} \cap (X \times Y) = \text{graf}T \cap (X \times Y).$$

Além disso, como X é aberto e $\bar{T}^{-1}(0) \subset X$, existe $\epsilon > 0$ tal que $\bar{T}^{-1}(0) + B(\epsilon) \subset X$, em particular, $X \supset \bar{z} + B(\epsilon) = B(\bar{z}, \epsilon)$. Como Y é aberto e $0 \in Y$, existe $\delta > 2\epsilon/\bar{c} > 0$ tal que $B(2\epsilon/\bar{c}) \subset Y$ onde $\bar{c} = \inf c_k$.

Por outro lado, Consideremos $z \in B(\bar{z}, \epsilon)$ arbitrário e do fato que \bar{T} é monótono maximal podemos definir

$$z' = J_k^{\bar{T}}(z). \tag{3.2.1}$$

Como $\bar{z} \in \bar{T}^{-1}(0) \subset X$, de (2.1.13), temos que $J_k^{\bar{T}}(\bar{z}) = \bar{z} \in X$ e da não expansividade de $J_k^{\bar{T}}$ obtemos

$$\|z' - \bar{z}\| = \left\| J_k^{\bar{T}}(z) - J_k^{\bar{T}}(\bar{z}) \right\| \leq \|z - \bar{z}\| < \epsilon,$$

o que implica que

$$z' \in B(\bar{z}, \epsilon) \subset X. \tag{3.2.2}$$

Então,

$$c_k^{-1}(z - z') \in c_k^{-1}[B(\bar{z}, \epsilon) - B(\bar{z}, \epsilon)] = B(2\epsilon/c_k) \subset B(2\epsilon/\bar{c}) \subset Y \tag{3.2.3}$$

Como $z = (J_k^{\bar{T}})^{-1}(z') = (I + c_k\bar{T})(z')$ então $\bar{T}(z') = c_k^{-1}(z - z')$, isto é

$$(z', c_k^{-1}(z - z')) \in \text{graf}\bar{T}. \tag{3.2.4}$$

De (3.2.2), (3.2.3), (3.2.4) e pela hipótese temos que

$$(z', c_k^{-1}(z - z')) \in \text{graf}\bar{T} \cap (X \times Y) = \text{graf}T \cap (X \times Y) \subset \text{graf}T,$$

de onde,

$$\begin{aligned}
c_k^{-1}(z - z') \in T(z') &\iff z - z' \in c_k T(z') \\
&\iff z \in (I + c_k T)(z') \\
&\iff z' \in J_k^T(z).
\end{aligned} \tag{3.2.5}$$

De (3.2.1), (3.2.2) e (3.2.5) temos que

$$J_k^{\bar{T}}(z) = z' \in J_k^T(z) \cap B(\bar{z}, \epsilon).$$

Reciprocamente, seja $z'' \in B(\bar{z}, \epsilon) \cap J_k^T(z)$, isto é,

$$z'' \in B(\bar{z}, \epsilon) \subset X \tag{3.2.6}$$

e

$$\begin{aligned}
z'' \in J_k^T(z) &\iff z \in (J_k^T)^{-1}(z'') \\
&\iff z \in z'' + c_k T(z'') \\
&\iff c_k^{-1}(z - z'') \in T(z'').
\end{aligned} \tag{3.2.7}$$

Mais ainda, como $z'' \in B(\bar{z}, \epsilon)$ e pelo mesmo raciocínio feito em (3.2.2) e (3.2.3), temos que

$$c_k^{-1}(z - z'') \subset Y. \tag{3.2.8}$$

De (3.2.6), (3.2.7) e (3.2.8) e pela hipótese,

$$\begin{aligned}
(z'', c_k^{-1}(z - z'')) &\in \text{graf } T \cap (X \times Y) \\
&= \text{graf } \bar{T} \cap (X \times Y) \\
&\subseteq \text{graf } \bar{T},
\end{aligned}$$

o qual implica que,

$$\begin{aligned}
c_k^{-1}(z - z'') \in \bar{T}(z'') &\iff z \in (I + c_k \bar{T})(z'') \\
&\iff z'' = J_k^{\bar{T}}(z).
\end{aligned}$$

Então, $B(\bar{z}, \epsilon) \cap J_k^T(z) \in J_k^{\bar{T}}(z)$. Portanto, $B(\bar{z}, \epsilon) \cap J_k^T(z) = J_k^{\bar{T}}(z)$. ■

Lema 3.2.2 *Nas condições do Lema 3.2.1, se $S_k = \rho_k J_k^T + (1 - \rho_k)I$ e $\bar{S}_k = \rho_k \bar{J}_k^T + (1 - \rho_k)I$ então existe $\epsilon > 0$ tal que para todo $z \in B(\bar{z}, \epsilon)$ temos que $\bar{S}_k(z) = B(\bar{z}, \epsilon) \cap S_k(z)$ para todo $k \in \mathbb{N}$, onde $\bar{z} \in \text{Zer}(T)$.*

Prova. Do Lema 2.1.44 o operador \bar{S}_k é não-expansivo, o qual necessariamente é simples valorado. Consideremos

$$x = \bar{S}_k(z). \quad (3.2.9)$$

Vamos provar que $x \in B(\bar{z}, \epsilon) \cap S_k(z)$. De fato, como $\bar{z} \in \bar{T}^{-1}(0)$ e de (2.1.13) temos que $\bar{z} = J_k^T(\bar{z})$. Logo, $\bar{z} = \bar{S}_k(\bar{z})$ e da não expansividade de \bar{S}_k temos que

$$d(x, \bar{z}) = d(\bar{S}_k(z), \bar{S}_k(\bar{z})) \leq d(z, \bar{z}) < \epsilon,$$

isto é,

$$x \in B(\bar{z}, \epsilon). \quad (3.2.10)$$

De (3.2.9) e do Lema 3.2.1,

$$\frac{x + (\rho_k - 1)z}{\rho_k} = J_k^T(z) = B(\bar{z}, \epsilon) \cap J_k^T(z) \subset J_k^T(z),$$

o que implica que

$$x \in \rho_k J_k^T(z) + (1 - \rho_k)z = S_k(z). \quad (3.2.11)$$

De (3.2.10) e (3.2.11) temos que

$$x \in S_k(z) \cap B(\bar{z}, \epsilon).$$

Reciprocamente, consideremos $w \in B(\bar{z}, \epsilon) \cap S_k(z)$ então isto é,

$$\frac{w + (\rho_k - 1)z}{\rho_k} \subset J_k^T(z). \quad (3.2.12)$$

Por outro lado, como $w \in B(\bar{z}, \epsilon)$, $z \in B(\bar{z}, \epsilon)$ e pela convexidade de $B(\bar{z}, \epsilon)$ para $\rho_k \in [1, 2]$ temos que

$$\frac{w + (\rho_k - 1)z}{\rho_k} \in B(\bar{z}, \epsilon). \quad (3.2.13)$$

De (3.2.12), (3.2.13) e do Lema 3.2.1 temos

$$\frac{w + (\rho_k - 1)z}{\rho_k} \subset J_k^T(z) \cap B(\bar{z}, \epsilon) = J_k^T(z),$$

e como $J_k^{\bar{T}}$ é não-expansiva, é simples valorado. Daí temos que

$$\frac{w + (\rho_k - 1)z}{\rho_k} = J_k^{\bar{T}}(z).$$

Então, temos que $w = \bar{S}_k(z)$. Portanto, $B(\bar{z}, \epsilon) \cap S_k(z) = \bar{S}_k(z)$. ■

Logo, com esses dois últimos resultados vamos mostrar, no seguinte resultado, que o APPR pode ser aplicado para operadores localmente monótono maximal.

Proposição 3.2.3 *Seja $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ uma aplicação monótona maximal em $X \times Y$, onde $X, Y \subset \mathbb{R}^n$ são conjuntos abertos tal que $0 \in Y$ e $T^{-1}(0) \cap X$ é não vazio e fechado com $T^{-1}(0) \cap X + B(\delta) \subset X$ para algum $\delta > 0$. Seja $\{c_k\}$ e $\{\rho_k\}$ tal que $\inf c_k > 0$, $\inf \rho_k \geq 1$ e $\sup \rho_k < 2$. Se $d_{T^{-1}(0) \cap X}(z^0)$ é suficientemente pequeno, então existe $\epsilon > 0$ e $\bar{z} \in T^{-1}(0) \cap X$ tal que $z^0 \in B(\bar{z}, \epsilon)$ e*

$$z^{k+1} = B(\bar{z}, \epsilon) \cap [\rho_k (I + c_k T)^{-1}(z^k) + (1 - \rho_k)z^k]$$

gera uma sequência única $\{z^k\}$ convergindo monotonicamente Fejér para um ponto $T^{-1}(0) \cap X \cap B(\bar{z}, \epsilon)$.

Se ademais, c_k, ρ_k e a aplicação $y \rightarrow T^{-1}(y) \cap X$ satisfazem as hipóteses da Proposição 3.1.3, então $\{z^k\}$ converge linearmente para $\bar{z} = T^{-1}(0) \cap X$ com taxa

$$\sqrt{1 - \bar{\rho}(2 - \bar{\rho}) \frac{\bar{c}^2}{a^2 + \bar{c}^2}} < 1. \quad (3.2.14)$$

Prova. Consideremos \bar{T} a extensão monótona maximal de T , isto é, \bar{T} é um operador monótono maximal tal que

$$\text{graf } \bar{T} \cap (X \times Y) = \text{graf } T \cap (X \times Y) \quad (3.2.15)$$

Pela hipótese, existe $x \in T^{-1}(0) \cap X$ e como $0 \in Y$ então temos que

$$(x, 0) \in \text{graf } T \cap (X \times Y). \quad (3.2.16)$$

De (3.2.15) e (3.2.16), é equivalente ter $(x, 0) \in \text{graf } \bar{T} \cap (X \times Y)$, isto é, $x \in \bar{T}^{-1}(0) \cap X$, o que implica que $T^{-1}(0) \cap X \subset \bar{T}^{-1}(0) \cap X$. Analogamente, usando (3.2.15) obtemos que $\bar{T}^{-1}(0) \cap X \subset T^{-1}(0) \cap X$. Logo,

$$T^{-1}(0) \cap X = \bar{T}^{-1}(0) \cap X, \quad (3.2.17)$$

o que implica que, segundo a hipótese, o conjunto $\bar{T}^{-1}(0) \cap X$ é fechado e não vazio.

Além disso, como \bar{T}^{-1} é monótono maximal e do Lema 2.1.10 temos que o conjunto $\bar{T}^{-1}(0)$ é convexo e fechado. Mais ainda, pode se dizer que o conjunto $\bar{T}^{-1}(0)$ é um conjunto não vazio e conexo. Lembre-se que a propriedade de convexidade é condição suficiente de conexidade.

A seguir, vamos mostrar que $\bar{T}^{-1}(0) = \bar{T}^{-1}(0) \cap X$. De fato, o conjunto $\bar{T}^{-1}(0)$ sempre pode se escrever nessa forma

$$\bar{T}^{-1}(0) = \left(\bar{T}^{-1}(0) \cap X \right) \cup \left(\bar{T}^{-1}(0) \cap X^c \right),$$

com X^c como o complemento do conjunto X . Observe que $\left(\bar{T}^{-1}(0) \cap X \right) \cap \left(\bar{T}^{-1}(0) \cap X^c \right) = \emptyset$. Isso significa que existe uma partição do conjunto $\bar{T}^{-1}(0)$ em duas partes fechadas, não vazias e disjuntas, o que equivale dizer que o conjunto $\bar{T}^{-1}(0)$ é desconexo, o qual é uma contradição. Então, como $\bar{T}^{-1}(0) = \bar{T}^{-1}(0) \cap X$ e de (3.2.17) podemos afirmar que

$$\bar{T}^{-1}(0) = T^{-1}(0) \cap X.$$

Por outro lado, para $z^0 \in X$ temos que

$$d_{T^{-1}(0) \cap X}(z^0) = d_{\bar{T}^{-1}(0)}(z^0),$$

e como o conjunto $T^{-1}(0)$ é convexo e fechado, segundo o Teorema 1.1.3, existe um único $\bar{z} \in \bar{T}^{-1}(0)$ o qual é a projeção de z^0 sobre o conjunto $\bar{T}^{-1}(0)$. Além disso, pela hipótese, $d_{\bar{T}^{-1}(0)}(z^0)$ é suficientemente pequeno, isto é, existe $\epsilon > 0$ suficientemente pequeno tal que $d_{\bar{T}^{-1}(0)}(z^0) < \epsilon$.

A partir de $z^0 \in B(\bar{z}, \epsilon)$, definimos para cada $k \in \mathbb{N}$,

$$z^{k+1} = \left(\rho_k J_k^{\bar{T}} + (1 - \rho_k)I \right) (z^k). \quad (3.2.18)$$

Observe que a unicidade da sequência $\{z^k\}$ é consequência do Lema 2.1.44.

Nestas condições, segundo o Lema 3.2.2 para cada $z^k \in B(\bar{z}, \epsilon)$, (3.2.18) pode se redefinir como

$$z^{k+1} = B(\bar{z}, \epsilon) \cap \left(\rho_k J_k^T + (1 - \rho_k)I \right) (z^k).$$

Além disso, de (3.2.18) temos que $\{z^k\}$ é gerado pelo APPR aplicado ao operador \bar{T} e do fato que $\bar{T}^{-1}(0)$ é não vazio, segundo o Teorema 3.1.2 temos que $\{z^k\}$ converge para um ponto de $Zer(\bar{T})$, isso é, $\{z^k\}$ converge para um ponto do conjunto $T^{-1}(0) \cap X \cap B(\bar{z}, \epsilon)$.

Mais ainda, $\{z^k\}$ é uma sequência Fejér-monótona com respecto ao conjunto $\bar{T}^{-1}(0) \cap B(\bar{z}, \epsilon)$. De fato, para $w \in \bar{T}^{-1}(0) \cap B(\bar{z}, \epsilon)$ e de (2.1.13) temos que $w = J_k^{\bar{T}}(w)$, o qual implica que $\bar{S}_k(w) = w$ e pela não expansividade de \bar{S}_k temos que

$$\|z^{k+1} - w\| = \|\bar{S}_k(z^k) - \bar{S}_k(w)\| \leq \|z^k - w\|,$$

provando assim que $\{z^k\}$ é Fejér-monótono em $\bar{T}^{-1}(0) \cap B(\bar{z}, \epsilon)$.

Além disso, de (3.2.15) temos que $\bar{T}^{-1}(y) \cap X = T^{-1}(y) \cap X$ para todo $y \in Y$. Do fato que $y \mapsto T^{-1}(y) \cap X$ é a -Lipschitz contínua e segundo a Proposição 3.1.3 temos que o algoritmo, quando T é localmente monótono maximal, possui uma taxa de convergência linear como é dado em (3.2.14). ■

Esta última proposição mostra que podemos aplicar o APPR, para $z^0 \in B(\bar{z}, \epsilon)$, cujas iterações estão dados por

$$z^{k+1} = (\rho_k J_k^T + (1 - \rho_k)I)(z^k),$$

os quais, para cada k , estão contidas em $B(\bar{z}, \epsilon)$.

3.3 O APP sem monotonicidade.

A existência de iterações para algoritmos proximais aplicados a operadores com certas características muitas vezes é sensível. No caso de operadores monótonos maximais, a principal ferramenta usada para estabelecer existência é o Teorema de Minty[24], o qual não trabalha sem monotonicidade. Então, para operadores mais gerais precisamos obter algumas características convenientes desse de tal forma que possamos utilizar ainda o APP descrito na seção anterior. Superar estes obstáculos requer algumas técnicas onde λ -hipomonotonicidade e k -localização Lipschitz se tornam cruciais.

Teorema 3.3.1 *Seja $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ tal que para algum $\lambda \geq 0$, T_λ é monótono maximal em $X \times Y$, onde $X, Y \subset \mathbb{R}^n$ são conjuntos abertos tais que $0 \in Y$, $T^{-1}(0) \cap X$ é não vazio*

e fechado com $T^{-1}(0) \cap X + B(\delta) \subset X$ para algum $\delta > 0$ e, $\inf c_k > 2\lambda$. Se $d_{T^{-1}(0) \cap X}(z^0)$ é suficientemente pequeno, então existe $\epsilon > 0$ e $\bar{z} \in T^{-1}(0) \cap X$ tal que $z^0 \in B(\bar{z}, \epsilon)$, e os iterados

$$z^{k+1} = B(\bar{z}, \epsilon) \cap J_k^T(z^k)$$

geram uma sequência única $\{z^k\}$ que converge monotonicamente Fejér para um ponto de $T^{-1}(0) \cap X$. Se ademais, c_k e a aplicação $y \rightarrow T^{-1}(y) \cap (X - \lambda Y)$ satisfazem as hipóteses da Proposição 3.1.3, então $\{z^k\}$ converge linearmente para $\{\bar{z}\} = T^{-1}(0) \cap X$ com taxa de convergência

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - \rho} \left(2 - \frac{\bar{c}}{\bar{c} - \rho}\right) \frac{\bar{c}^2}{(a + \rho)^2 + \bar{c}^2}} < 1.$$

Prova. Primeiro, vamos provar que para um z^0 suficientemente perto de $\bar{z} \in \text{Zer}(T)$, as iterações geradas pelo APP clássico estão contidas numa vizinhança de \bar{z} . De fato, como $\inf c_k > 2\lambda$, então para todo k temos que $c_k - \lambda > \lambda$. Denotemos $c'_k = c_k - \lambda$ e consideremos $\rho_k = \frac{c_k}{c_k - \lambda}$, então

$$\rho_k = 1 + \frac{\lambda}{c_k}.$$

Como $c'_k > \lambda$ temos que $1 > \lambda/c'_k$, então $\sup \rho_k < 2$.

Além disso,

$$\inf \rho_k = 1 + \inf \frac{\lambda}{c'_k} = 1 + \frac{\lambda}{\sup c'_k} \geq 1.$$

Por outro lado, como T_λ é um operador localmente monótono maximal em $X \times Y$, de (2.1.14), temos que $T_\lambda^{-1}(0) = T^{-1}(0)$ e do fato que as hipóteses nesse teorema são semelhantes à Proposição 3.2.3, é que podemos substituir o operador T por T_λ nessa proposição. Então, existe $\epsilon > 0$ e $\bar{z} \in T_\lambda^{-1}(0) \cap X = T^{-1}(0) \cap X$ tal que $z^0 \in B(\bar{z}, \epsilon)$ e os iterados dados por

$$z^{k+1} = B(\bar{z}, \epsilon) \cap \left[\rho_k \left(I + c'_k T_\lambda \right)^{-1} (z^k) + (1 - \rho_k) z^k \right], \quad (3.3.1)$$

geram uma sequência única $\{z^k\}$ que converge monotonicamente Fejér para um ponto de $T_\lambda^{-1}(0) \cap X = T^{-1}(0) \cap X$.

Mais ainda, do Lema 2.1.54, temos

$$\begin{aligned}
\rho_k \left(I + c'_k T_\lambda \right)^{-1} (z^k) + (1 - \rho_k) (z^k) &= \rho_k \left[\frac{c'_k}{c'_k + \lambda} \left(I + (c'_k + \lambda) T \right)^{-1} + \frac{\lambda}{c'_k + \lambda} I \right] (z^k) + (1 - \rho_k) z^k \\
&= \rho_k \left[\frac{c'_k}{c_k} \left(I + c_k T \right)^{-1} + \frac{\lambda}{c_k} I \right] (z^k) + (1 - \rho_k) z^k \\
&= \rho_k \left[\frac{1}{\lambda_k} \left(I + c_k T \right)^{-1} + \frac{\lambda}{c_k} I \right] (z^k) + (1 - \rho_k) z^k \\
&= \left(I + c_k T \right)^{-1} (z^k) + \left(\frac{\rho_k \lambda}{c_k} + 1 - \rho_k \right) z^k \\
&= \left(I + c_k T \right)^{-1} (z^k).
\end{aligned} \tag{3.3.2}$$

Então, de (3.3.1) e (3.3.2) temos que

$$z^{k+1} \in B(\bar{z}, \epsilon) \cap \left(I + c_k T \right)^{-1} (z^k).$$

Além disso, é simples verificar que é válido a seguinte igualdade

$$T_\lambda^{-1}(y) \cap X = T^{-1}(y) \cap (X - \lambda Y) + \lambda Y,$$

para $y \in Y$. E, segundo a hipótese temos que a aplicação $y \mapsto T^{-1}(y) \cap (X - \lambda Y)$ é a -Lipschitz contínua em 0. Daí

$$\begin{aligned}
x \in T_\lambda^{-1}(y) \cap X &\iff x = x' + \lambda y, \quad x' \in T^{-1}(y) \cap (X - \lambda Y) \\
&\iff \|x\| \leq \|x'\| + \lambda \|y\| \leq (a + \lambda) \|y\|,
\end{aligned}$$

o qual prova que a aplicação $y \mapsto T_\lambda^{-1}(y) \cap X$ é $(a + \lambda)$ -Lipschitz contínua. Então da Proposição 3.2.3 temos a taxa de convergência pedida. ■

Cabe observar que, sendo T um operador que prescinde da propriedade de monotonicidade, os iterados do Teorema 3.3.1 geram uma única sequência contida em $B(\bar{z}, \epsilon)$, o qual resolve a inclusão

$$0 \in T(z) + c_k^{-1}(z - z^k).$$

Uma situação particularmente conveniente ocorre quando, nos resultados, a condição de monotonicidade maximal de um operador é trocado por uma k -Localização Lipschitz como veremos no resultado seguinte.

Corolário 3.3.2 *Seja $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ tal que T^{-1} tem k -Localização Lipschitz em $(0, \bar{z})$, $\{c_k\}$ é uma sequência tal que $\inf c_k > 2k$ e $c_k \nearrow \bar{c} \geq \infty$. Se $\|z^0 - \bar{z}\|$ é suficientemente pequeno, então existe $\epsilon > 0$ tal que $z^0 \in B(\bar{z}, \epsilon)$ e os iterados dados como*

$$z^{k+1} = B(\bar{z}, \epsilon) \cap J_k^T(z^k)$$

gera uma única sequência $\{z^k\}$, convergindo linearmente e monotonicamente Fejér para \bar{z} com taxa de convergência

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - k} \left(2 - \frac{\bar{c}}{\bar{c} - k}\right) \frac{\bar{c}^2}{(2k)^2 + \bar{c}^2}} < 1. \quad (3.3.3)$$

Prova. Como T^{-1} tem uma k -Localização Lipschitz, segundo a Proposição 2.1.56, para algum $\lambda \geq k$ o operador T_λ é monótona maximal numa vizinhança $X \times Y \ni (\bar{z}, 0)$. Logo, do Teorema 3.3.1, existe $\epsilon > 0$ e $\bar{z} \in T_\lambda^{-1}(0) \cap X = T^{-1}(0) \cap X$ tal que para $z^0 \in B(\bar{z}, \epsilon)$ os iterados

$$z^{k+1} = B(\bar{z}, \epsilon) \cap J_k^T(z^k),$$

gera uma única sequência $\{z^k\}$ que converge monotonicamente Fejér para um ponto de $T^{-1}(0) \cap X$.

Além disso, como T^{-1} tem k -Localização Lipschitz em $(0, \bar{z})$, temos que

$$x \mapsto T^{-1}(0) \cap Y$$

é k -Lipschitz contínua, satisfazendo assim a hipótese da Proposição 3.1.3. Portanto, o algoritmo, quando aplicado a operadores T sem monotonicidade, possui uma taxa de convergência linear como é dado em (3.3.3). ■

Exemplo 3.3.3 *Consideremos a função $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ tal que*

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

Um problema clássico em otimização é procurar os pontos estacionários da função, o que equivale solucionar o problema dado por

$$(0, 0) \in \partial f(x, y). \quad (3.3.4)$$

Definimos T como o operador subdiferencial de f , isto é,

$$T(x, y) = \nabla f(x, y) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Então, uma alternativa para achar os zeros do subdiferencial dessa função é utilizar o APP o qual gera, a partir de um ponto $(x^k, y^k) = (\epsilon^k, \epsilon^k)$, uma sequência de iterações dados por

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = J_k^T \begin{pmatrix} \epsilon^k \\ \epsilon^k \end{pmatrix} = \begin{pmatrix} \frac{1}{1+c_k} \epsilon^k \\ \frac{1}{1-c_k} \epsilon^k \end{pmatrix},$$

o qual $\{y^{k+1}\}$ é uma sequência divergente se $c_k \in (0, 2)$. Portanto, a sequência do APP clássico diverge, o qual acontece porque o operador não é monótono.

Daí, uma proposta para regularizar esse operador é a λ -hipomonotonicidade de T^{-1} , isto é, que

$$(T^{-1} + \lambda I) = \begin{pmatrix} 1 + \lambda & 0 \\ 0 & -1 + \lambda \end{pmatrix}$$

seja monótona, o qual acontece quando $\lambda \geq 1$.

Com estas considerações, segundo a hipótese do Teorema 3.3.1 temos que $\inf c_k > 2\lambda$ o que implica que $c_k > 2$ para todo $k \in \mathbb{N}$. Observe que x^k , em cada iteração, cada vez mais vai se aproximando a zero enquanto y^k , em cada iteração, aproxima-se do zero oscilatoriamente. Portanto, para solucionar o problema (3.3.4) o APP, gera uma sequência (x^k, y^k) tal que $(x^k, y^k) \rightarrow (0, 0)$ mesmo que o operador T não seja monótono maximal.

Capítulo 4

Aplicação do APP aos métodos de multiplicadores proximais

Neste capítulo vamos fazer uma análise dos métodos multiplicadores proximais. Rockafellar[36], além de estudar o APP para o caso geral de encontrar zeros de operadores monótonos maximais, mostrou que o APP pode ser aplicado a problemas de otimização convexa, como pode se ver em Rockafellar[32], em três diferentes maneiras:

- a. O algoritmo para problemas de minimização.
- b. O método dual dos multiplicadores ou chamado método Lagrangeano Aumentado.
- c. O método de multiplicadores proximais, o qual é uma combinação dos métodos multiplicadores com uma função quadrática regularizante.

Cabe lembrar que uma das construções mais úteis em otimização é a teoria de dualidade, onde o problema de minimização é relacionado a dois problemas auxiliares: os problemas dual e lagrangeano. Além disso, como uma série de teorias e algoritmos para programação convexa é construída sob noções de dualidade, então é natural reformular uma teoria aplicável neste contexto para problemas de otimização mais gerais. Como a teoria de dualidade clássica falha em ausência de convexidade é que vamos apresentar, para problemas gerais, os métodos de multiplicadores no contexto de dualidade geral.

É importante observar que aplicar o APP ao problema primal, às vezes, não parece

muito útil, do fato que este método substitui uma inclusão simples por uma sequência de inclusões, o que pode ser tão complexo como solucionar a inclusão inicial. Isto motiva aplicar o APP ao problema dual ou lagrangeano e utilizar o contexto de dualidade para solucionar esses subproblemas.

Neste capítulo vamos utilizar todas as notações da teoria de dualidade da Seção 1.3 do Capítulo 1 e da Seção 2.2 do Capítulo 2.

4.1 Métodos dos multiplicadores

Rockafellar[32] mostrou como os métodos multiplicadores (ou chamado também método de lagrangeano aumentado) para minimização convexa pode ser derivado do APP. Daqui, usando o contexto de dualidade geral podemos usar a mesma idéia para derivar os métodos multiplicadores para inclusões monótonos mais gerais.

Lembremos que o problema da inclusão primal, lagrangeano e dual, dados em (1.3.5), (1.3.6) e (1.3.7), respectivamente, são dados por

$$\begin{aligned} 0 &\in F_0(x), \\ (0, 0) &\in L(x, z), \\ 0 &\in G_0(z). \end{aligned}$$

Assim, o método de multiplicadores é obtido quando o APP é aplicado ao problema dual $0 \in G_0(z)$, onde $G_0 = P_Y \circ G \circ P_Z^*$, gerando assim uma sequência $\{z^{k+1}\}$, a partir de $z^0 \in Z$, o qual resolve o problema dual modificado

$$0 \in G_{0,k}(z), \tag{4.1.1}$$

onde

$$G_{0,k}(z) = G_0(z) + c_k^{-1}(z - z^k),$$

cuja forma parametrizada é dado por

$$G_k(w, z) = G(w, z) + \{0\} \times c_k^{-1}(z - z^k)$$

Lembremos que é suficiente ter G_k para obter L_k e F_k . Então podemos definir o problema primal, lagrangeano e dual, correspondiente a F_k , dados por

$$0 \in (F_k)_0(x), \quad (4.1.2)$$

$$(0, 0) \in L_k(x, z), \quad (4.1.3)$$

$$0 \in (G_k)_0(z), \quad (4.1.4)$$

onde F_k e G_k são as parametrizações de $(F_k)_0$ e $(G_k)_0$, respectivamente, isto é, $(F_k)_0(x) = P_W(F_k(x, y))$ e $(G_k)_0(z) = P_Y(G(w, z))$.

Observe que (4.1.1) e (4.1.4) são problemas equivalentes. De fato,

$$(G_k)_0(z) = P_Y(G_k(0, z)) = P_Y(G(0, z) + \{0\} \times c_k^{-1}(z - z^k)) = G_0(z) + c_k^{-1}(z - z^k) = G_{0,k}(z).$$

Por outro lado, segundo o contexto de dualidade geral, o conjunto solução do problema dual (1.3.7) pode ser expresso como

$$\begin{aligned} G_0^{-1}(0) &= \{z \in Z \mid 0 \in G_0(z)\} \\ &= \{z \in Z \mid \exists x \in X \text{ tal que } (0, 0) \in L(x, z)\} \\ &= \{z \in Z \mid \exists x \in X \text{ tal que } (0, 0) \in F^{(1,-1)}(x, z)\} \\ &= \{z \in Z \mid \exists x \in X \text{ tal que } (0, z) \in F(x, 0)\} \\ &= \{z \in Z \mid \exists x \in X \text{ tal que } 0 \in F_0(x), z \in P_Z F(x, 0)\} \\ &= \cup_{x \in F_0^{-1}(0)} P_Z F(x, 0), \end{aligned}$$

no qual, a segunda igualdade é obtida de (1.3.9), a terceira igualdade da definição de L , a quarta pela definição do operador $F^{(1,-1)}$, a quinta pela projeção de F sobre W e Z e da definição de F_0 .

Se G_0 é monótono maximal, então $G_{0,k}$ também é monótono maximal e do fato da presença do termo de regularização $c_k^{-1}(z - z^k)$ temos que $G_{0,k}$ é também fortemente monótono maximal. Logo, a iteração gerada por (4.1.1) tem solução única.

Portanto, embora o problema (4.1.2) tenha múltiplas soluções, a aplicação $P_Z F_k(x^{k+1}, 0)$ é um conjunto unitário e é independente de x^{k+1} . Assim, se encontramos uma solução x^{k+1} ao problema (4.1.2), uma solução de (4.1.1) é $z^{k+1} = P_Z F_k(x^{k+1}, 0)$. Assim o algoritmo pode ser estabelecido da seguinte forma:

Algoritmo 4.1.1 .

1. Escolher $z^0 \in Z$ e estabelecer $k = 0$.

2. Resolver

$$0 \in (F_k)_0(x), \quad (4.1.5)$$

obtendo x^{k+1} , com $(F_k)_0(x) = P_W(F_k(x, y))$.

3. Estabelecer

$$z^{k+1} = P_Z F_k(x^{k+1}, 0),$$

$$k \leftarrow k + 1$$

com $F_k = G_k^{-1}$.

A seguir damos o resultado de convergência para o Algoritmo 4.1.1.

Teorema 4.1.2 *Seja F monótono maximal tal que $0 \in \text{Im}(F_0)$ e G_0 ser monótono maximal. Então, o Algoritmo 4.1.1 gera as sequências $\{z^k\}$ e $\{x^k\}$ tal que $\{z^k\}$ converge para uma solução \bar{z} do problema (1.3.7). Além disso, para cada ponto limite \bar{x} de $\{x^k\}$, (\bar{x}, \bar{z}) resolve o problema primal dual (1.3.6). Em particular, \bar{x} resolve (1.3.5).*

Demonstração. Como $0 \in \text{Im}(F_0)$, existe x tal que $0 \in F_0(x)$ e segundo o Teorema 1.3.2, existe z tal que $(0, 0) \in L(x, z)$, como também $0 \in G_0(z)$. Então, aplicando o APP ao operador monótono maximal G_0 geramos uma única sequência $\{z^k\}$, tal que soluciona o problema (4.1.1), o qual converge para uma solução do problema dual (1.3.7).

Aliás, o Algoritmo 4.1.1 gera uma sequência $\{x^k\}$ que resolve (4.1.5), então segundo o Teorema 1.3.2, z^k resolve o problema (4.1.4) tal que (x^k, z^k) resolve o problema (4.1.3), isto é, $(0, 0) \in L_k(x^k, z^k)^2$ onde

$$L_k(x, z) = G_k^{(-1,1)}(x, z) = L(x, z) + \{0\} \times c_k^{-1}(z - z^k)$$

Logo, $(x^{k+1}, z^{k+1}, 0, c_k^{-1}(z^k - z^{k+1})) \in L$.

Como F é monótona então $F^{-1} = G$ é monótona, o que implica que $G^{(-1,1)}$ é

²Em Pennanen[26], na quinta linha da demonstração do Teorema 6.5, pp. 55, por engano aparece L ao invés de L_k . Além disso, na seguinte linha aparece, mais uma vez por engano, o $G^{(-1,1)}_k$ aplicado ao (x^*, u^*) , o qual teria que ser em (x, u^*) .

monótona. Logo, L é monótona. Então para qualquer $(x, z, w, y) \in L$ temos

$$\begin{aligned} 0 &\leq \langle (x^{k+1}, z^{k+1}) - (x, z), (0, c_k^{-1}(z^k - z^{k+1})) - (w, y) \rangle \\ &= \langle x^{k+1} - x, 0 - w \rangle + \langle z^{k+1} - z, c_k^{-1}(z^{k+1} - z^k) - y \rangle. \end{aligned}$$

Tomando limite, quando $k \rightarrow \infty$, temos que $c_k^{-1}(z^k - z^{k+1})$ converge para 0.

Logo,

$$\begin{aligned} 0 &\leq \langle \bar{x} - x, 0 - w \rangle + \langle \bar{z} - z, 0 - y \rangle \\ &= \langle (\bar{x}, \bar{z}) - (x, z), (0, 0) - (w, y) \rangle, \end{aligned}$$

onde \bar{x} e \bar{z} são os pontos limites de $\{x^k\}$ e $\{z^k\}$, respectivamente.

Como $(x, z, w, y) \in L$ foi escolhido arbitrariamente, do Lema 2.1.8 temos que $(\bar{x}, \bar{z}, 0, 0) \in L$, isto é, $(0, 0) \in L(\bar{x}, \bar{z})$. Logo, (\bar{x}, \bar{z}) resolve o problema primal dual (1.3.6). Então, do Teorema 1.3.2, \bar{x} resolve (1.3.5). ■

Por outro lado, dado T um operador monótono maximal, definimos a aplicação

$$T_k^{-1}(u) = T^{-1}(u) + c_k^{-1}(u - z^k), \quad (4.1.6)$$

considerando $\psi(w) = w - c_k^{-1}z^k = u$ temos que $w = \psi^{-1}(u) = u + c_k^{-1}z^k$. Daí,

$$\begin{aligned} T_k(u) &= ((T^{-1} + c_k^{-1}I)^{-1} \circ \psi^{-1})(u) \\ &= (T^{-1} + c_k^{-1}I)^{-1}(u + c_k^{-1}z^k), \end{aligned} \quad (4.1.7)$$

onde $(T^{-1} + c_k^{-1}I)^{-1}$ é a c_k^{-1} -regularização Yosida de T , denotado por $T_{c_k^{-1}}$. Aliás, do fato que $T_k = T_{c_k^{-1}} \circ \psi^{-1}$ onde $\psi(z) = z - c_k^{-1}z^k$. Da Proposição 2.1.49 temos que $T_{c_k^{-1}}$ é simples valorada, e assim T_k é também simples valorada embora o operador T pode ser multivalorado.

No contexto da dualidade geral com monotonicidade segundo a Seção 2.2, se substituimos T^{-1} por T_k^{-1} na equação (2.2.3), o Lagrangeano correspondente a T_k pode ser

escrito como

$$\begin{aligned}
L_k(x, z) &= (DH(x)^T z, -H(x)) + S(x) \times T_k^{-1}(z) \\
&= (DH(x)^T z, -H(x)) + S(x) \times (T^{-1}(z) + c_k^{-1}(z - z^k)) \\
&= (DH(x)^T z, -H(x)) + S(x) \times T^{-1}(z) + 0 \times (c_k^{-1}(z - z^k)) \\
&= L(x, z) + (0, c_k^{-1}(z - z^k)).
\end{aligned} \tag{4.1.8}$$

Desde que $(G_k)_0 = G_{0,k}$ e segundo a equação (1.3.9) temos que

$$\begin{aligned}
G_{0,k}(z) &= \{y \in Y \mid \exists x \in X, (0, y) \in L_k(x, z)\} \\
&= \{y \in Y \mid \exists x \in X, (0, y) \in \{L(x, z) + (0, c_k^{-1}(z - z^k))\}\} \\
&= \{y \in Y \mid \exists x \in X, (0, y) \in L(x, z)\} + c_k^{-1}(z - z^k) \\
&= G_0(z) + c_k^{-1}(z - z^k).
\end{aligned}$$

Então o problema primal, lagrangeano e dual correspondente a F_k dados em (4.1.2), (4.1.3) e (4.1.4), respectivamente, vai ser dado na seguinte forma:

$$\begin{aligned}
(P_k) \quad & 0 \in (F_k)_0(x) = S(x) + DH(x)^T T_k(H(x)), \\
(PD_k) \quad & (0, 0) \in L_k(x, z) = L(x, z) + c_k^{-1}(z - z^k). \\
(D_k) \quad & 0 \in G_{0,k} = G_0(z) + c_k^{-1}(z - z^k),
\end{aligned}$$

o qual é chamado o problema primal, lagrangeano e dual associado ao operador T_k .

Lembre que, no contexto de dualidade geral da Seção 2.2, a aplicação F_k é definido como segue

$$F_k(x, z) = \begin{pmatrix} S(x) \\ 0 \end{pmatrix} + \begin{pmatrix} DH(x)^T \\ I \end{pmatrix} T_k(H(x) + z),$$

então segundo a equação (2.2.1) temos que

$$G_{0,k}(z) = P_Z F_k(x, 0) = T_k(H(x)).$$

e do fato que z^{k+1} é a única solução do problema (4.1.4), temos que

$$z^{k+1} = T_k(H(x^{k+1})).$$

A seguir apresentamos um resultado que relaciona as soluções do problema primal (P_k) , dual (D_k) e lagrangeano (PD_k) .

Lema 4.1.3 *As soluções do problema (D_k) são os pontos tais que $z^{k+1} = T_k(H(x^{k+1}))$, onde x^{k+1} é uma solução do problema (P_k) . Além disso, os pontos (x^k, z^k) são soluções do problema (PD_k) .*

Prova. Como (P_k) , (D_k) e (PD_k) são os problemas primal, dual e lagrangeano, respectivamente, no contexto de dualidade geral, o resultado segue do Teorema 1.3.2. ■

Então, o algoritmo 4.1.1 no contexto de dualidade geral com monotonicidade para os métodos multiplicadores e a sua convergência local é apresentada como segue.

Algoritmo 4.1.4.

1. Escolher $z^0 \in Z$ e estabelecer $k = 0$.

2. Resolver

$$0 \in S(x) + DH(x)^T T_k(H(x))$$

obtendo x^{k+1} .

3. Estabelecer

$$z^{k+1} = T_k(H(x^{k+1})),$$

$$k \leftarrow k + 1.$$

Teorema 4.1.5 *Assumimos que a aplicação L^{-1} tem k -Localização Lipschitz no ponto $(0, 0, \bar{x}, \bar{z})$. Seja $\{c_k\}$ uma sequência tal que $c_k > 2k$ e $c_k \nearrow \bar{c} \leq \infty$. Se $\|z^0 - \bar{z}\|$ é suficientemente pequeno, então existe uma sequência $\{(x^k, z^k)\}$ gerada pelo Algoritmo 4.1.4 e uma vizinhança aberta X_k tais que para cada k , x^{k+1} é a única solução em X_k para o problema primal (P_k) . Além disso,*

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - k} \left(2 - \frac{\bar{c}}{\bar{c} - k}\right) \frac{\bar{c}^2}{(2k)^2 + \bar{c}^2}} < 1. \quad (4.1.9)$$

Finalmente, a sequência $\{x^k\}$ converge para \bar{x} e existe uma constante $M > 0$ tal que

$$\|x^k - \bar{x}\| \leq M \|z^k - \bar{z}\|.$$

Demonstração. Como L^{-1} tem uma k -Localização Lipschitz no ponto $(0, 0, \bar{x}, \bar{z})$, existem vizinhanças $X \ni \bar{x}$, $Z \ni \bar{z}$, $W \ni 0$ e $Y \ni 0$ tais que a aplicação

$$(w, y) \mapsto L^{-1}(w, y) \cap (X \times Z), \quad (4.1.10)$$

é simples valorado e k -Lipschitz contínua sobre $W \times Y$, o que implica que a aplicação

$$B : y \mapsto P_V (L^{-1}(0, y) \cap (X \times Z)), \quad V \subseteq Z$$

é k -Lipschitz contínua sobre V e simples valorada.

Denotemos $A_0 = B^{-1}$. Então, A_0^{-1} tem k -Localização Lipschitz em $(0, \bar{z})$. Além disso, consideremos $\{c_k\}$ uma sequência tal que $\inf c_k > 2k$ com $c_k \nearrow \bar{c} \leq \infty$ e $\|z^0 - \bar{z}\|$ suficientemente pequeno. Então, aplicando o Corolário 3.3.2 ao operador A_0^{-1} temos que existe $\epsilon > 0$ tal que $z_0 \in B(\bar{z}, \epsilon)$ e a iteração dada por

$$z^{k+1} = B(\bar{z}, \epsilon) \cap (I + c_k A_0^{-1})^{-1}(z^k), \quad (4.1.11)$$

o qual gera uma sequência única $\{z^k\}$ convergindo linearmente e monotonicamente Fejér para \bar{z} com taxa de convergência dada por (3.3.3).

Observe que (4.1.11), para $z^{k+1} \in B(\bar{z}, \epsilon) \cap Z$, pode ser escrita como $c_k^{-1}(z^k - z^{k+1}) \in A_0(z^{k+1})$, ou equivalentemente,

$$z^{k+1} \in A_0^{-1}(c_k^{-1}(z^k - z^{k+1})).$$

Por outro lado, pela definição de A_0 , existe $x^k \in X$ tal que

$$\begin{aligned} (x^{k+1}, z^{k+1}) &\in A_0^{-1}(c_k^{-1}(z^k - z^{k+1})) \\ &= (X \times (B(\bar{z}, \epsilon) \cap Z)) \times P_V (L^{-1}(0, c_k^{-1}(z^k - z^{k+1})) \cap (X \times Z)) \\ &= L^{-1}(0, c_k^{-1}(z^k - z^{k+1})), \end{aligned}$$

para $x^{k+1} \in X$ e $z^{k+1} \in Z \cap B(\bar{z}, \epsilon)$. Então,

$$(0, c_k^{-1}(z^k - z^{k+1})) \in L(x^{k+1}, z^{k+1}), \quad (4.1.12)$$

ou equivalentemente

$$(0, 0) \in L_k(x^{k+1}, z^{k+1}),$$

o qual é o problema lagrangeano (PD_k) , no qual, segundo o Lema 4.1.3 temos que $z^{k+1} = T_k(H(x^{k+1}))$. Então,

$$x^{k+1} \in X_k = (T_k \circ H)^{-1}(Z \cap B(\bar{z}, \epsilon)).$$

Denotemos $X_{k+1} = X \cap (T_k \circ H)^{-1}(Z \cap B(\bar{z}, \epsilon))$. Como T_k é k -Lipschitz contínuo e $H \in C^2$ temos que $T_k \circ H$ é k -Lipschitz contínua, e como $Z \ni \bar{z}$ é uma vizinhança, temos que $(T_k \circ H)^{-1}(Z \cap B(\bar{z}, \epsilon))$ é aberto. Logo, X_k é um conjunto aberto. A unicidade de X_k é obtida do fato que a aplicação definida em (4.1.10) é simples valorado.

Além disso, como z^k converge linearmente para \bar{z} , existe $r_k \in (0, 1)$ tal que

$$\|z^{k+1} - \bar{z}\| \leq r_k \|z^k - \bar{z}\|. \quad (4.1.13)$$

Logo, como $x^{k+1} \in X$ e $z^{k+1} \in Z \cap B(\bar{z}, \epsilon)$, de (4.1.12) e da k -Lipschitz da aplicação definida em (4.1.10) temos que

$$\begin{aligned} \|(x^{k+1}, z^{k+1}) - (\bar{x}, \bar{z})\| &\leq k \|L^{-1}(0, c_k^{-1}(z^k - z^{k+1})) - L^{-1}(0, 0)\| \\ &\leq k \|(0, c_k^{-1}(y^k - y^{k+1}))\| \\ &\leq kc_k^{-1} \|y^k - y^{k+1}\|. \end{aligned} \quad (4.1.14)$$

Então, de (4.1.14) e (4.1.13),

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\leq \|(x^{k+1}, z^{k+1}) - (\bar{x}, \bar{z})\| \\ &\leq kc_k^{-1} \|z^k - z^{k+1}\| \\ &\leq kc_k^{-1} (\|z^k - \bar{z}\| + \|z^{k+1} - \bar{z}\|) \\ &\leq kc_k^{-1}(1 + r_k) \|z^k - \bar{z}\| \\ &\leq M \|z^k - \bar{z}\|, \end{aligned}$$

onde $M = \sup \left\{ \frac{k(1+r_k)}{c_k} \right\} < \infty$. ■

Outra possibilidade seria aplicar o APP à inclusão lagrangiana, isto leva a um algoritmo mais interessante e bem comportado que o método multiplicador apresentado acima e que Rockafellar[32] chamou o método de multiplicadores proximais, o qual difiere do método multiplicador usual pela presença de um termo quadrático regularizante no qual é acrescentado aos problemas primal e dual.

4.2 Metodo multiplicadores proximais

Muitas extensões e modificações dos métodos multiplicadores tem sido estudados desde que estes foram sugeridos por Hestenes[17] e Powell[30] em 1969, todos eles com o obje-

tivo de substituir o problema de otimização com restrições por subproblemas irrestritos que tenham para ser solucionados eficientemente por algoritmos poderosos e flexíveis, um desses algoritmos é o método dos multiplicadores proximais, o qual foi introduzido independentemente por Rockafellar[32] e Antipin[3]. Vale a pena mencionar que os métodos multiplicadores proximais tem muita semelhança com os tão conhecidos métodos de penalidade, mas geralmente são melhores comportados que esses métodos com uma boa taxa de convergência e estabilidade numérica, ver Bertsekas [5].

Mais uma vez, nesta dissertação vamos utilizar as ferramentas da teoria de dualidade geral com monotonicidade da Seção 2.2 para analisarmos os métodos dos multiplicadores proximais para problemas não convexos.

É importante mencionar que aplicar o método multiplicador proximal significa perturbar o problema primal e dual por um termo regularizante. Nesta situação, consideremos as aplicações S_k e T_k dados por

$$\begin{aligned} S_k(x) &= S(x) + c_k^{-1}(x - x^k), \\ T_k(u) &= (I + c_k T^{-1})^{-1}(z^k + c_k u). \end{aligned}$$

Assim, se substituirmos S e T por S_k e T_k no contexto da dualidade geral e segundo (4.1.6) temos que

$$\begin{aligned} L_k(x, z) &= (DH(x)^T z, -H(x)) + S_k(x) \times T_k^{-1}(z) \\ &= (DH(x)^T z, -H(x)) + S(x) \times T^{-1}(z) + c_k^{-1}(x - x^k) \times c_k^{-1}(z - z^k) \\ &= L(x, z) + c_k^{-1}(x - x^k, z - z^k). \end{aligned}$$

Assim, o método dos multiplicadores proximais é obtido quando o APP é aplicado ao problema lagrangeano $(0, 0) \in L(x, z)$, obtendo assim uma sequência $\{(x^{k+1}, z^{k+1})\}$ que soluciona o subproblema

$$(\overline{PD}_k) \quad (0, 0) \in L(x, z) + c_k^{-1}(x - x^k, z - z^k).$$

Além disso, o subproblema primal associado com a aplicação S_k e T_k , de acordo com a teoria da dualidade geral, é dado por

$$(\overline{P}_k) \quad 0 \in S_k(x) + DH(x)^T T_k(H(x)). \quad (4.2.1)$$

Lema 4.2.1 *As soluções do subproblema (\overline{PD}_k) são os pontos dados como $(x^{k+1}, T_k(H(x^{k+1})))$, onde x^{k+1} é uma solução do subproblema (\overline{P}_k) .*

Prova. O resultado se obtém quando aplicarmos o Teorema 1.3.2 aos problemas (\overline{P}_k) e (\overline{PD}_k) . ■

A seguir apresentamos o algoritmo dos métodos dos multiplicadores proximais e um resultado sobre a sua convergência local.

Algoritmo 4.2.2.

1. Escolher $(x^0, z^0) \in X \times Z$ arbitrário e estabelecer $k = 0$.

2. Resolver

$$0 \in S(x) + c_k^{-1}(x - x^k) + DH(x)^T T_k(H(x))$$

obtendo x^{k+1} .

3. Estabelecer

$$z^{k+1} = T_k(H(x^{k+1})), \quad (4.2.2)$$

$$k = k + 1.$$

Do Lema 4.2.1, as soluções (x^{k+1}, z^{k+1}) do Algoritmo 4.2.2, onde z^{k+1} é gerado segundo o passo 3, são soluções do problema (\overline{PD}_k) , no qual converge para uma solução do problema lagrangeano original $(0, 0) \in L(x, z)$.

Teorema 4.2.3 *Considere que a aplicação L^{-1} tem k -Localização Lipschitz no ponto $(0, 0, \bar{x}, \bar{z})$. Seja $\{c_k\}$ uma sequência tal que $\inf c_k > 2k$ e $c_k \nearrow \bar{c} \leq \infty$. Se $\|(x^0, z^0) - (\bar{x}, \bar{z})\|$ é suficientemente pequeno, então existe uma sequência $\{(x^k, z^k)\}$ gerado pelo Algoritmo 4.2.2 e uma vizinhança aberta X_k tal que para cada k , x^{k+1} é a única solução em X_k para o subproblema (\overline{P}_k) . Além disso, a sequência $\{(x^k, z^k)\}$ converge linearmente e monotonicamente Fejér para (\bar{x}, \bar{z}) com uma taxa de convergência de*

$$\sqrt{1 - \frac{\bar{c}}{\bar{c} - k} \left(2 - \frac{\bar{c}}{\bar{c} - k}\right) \frac{\bar{c}^2}{(2k)^2 + \bar{c}^2}} < 1. \quad (4.2.3)$$

Demonstração . Pela hipótese, L^{-1} tem k -Localização Lipschitz em $(0, 0, \bar{x}, \bar{z})$. Então, do Corolário 3.3.2 existe $\epsilon > 0$ tal que

$(x^0, z^0) \in B((\bar{x}, \bar{z}), \epsilon)$ e os iterados

$$(x^{k+1}, z^{k+1}) = B((\bar{x}, \bar{z}), \epsilon) \cap (I + c_k L)^{-1} (x^k, z^k) \quad (4.2.4)$$

geram uma única sequência $\{(x^k, z^k)\}$ convergindo linear e Fejér monótona para (\bar{x}, \bar{z}) com a taxa de convergência dada em (3.3.3).

Além disso, a equação (4.2.4) pode ser escrito como

$$(0, 0) \in L_k(x^{k+1}, z^{k+1}).$$

Observe que (x^k, z^k) satisfaz o problema lagrangeano (\overline{PD}_k) . Do Lema 4.2.1, temos que $(x^k, z^k) = (x^k, T_k(H(x^k)))$. Consideremos $X_k = H_k^{-1}(B((\bar{x}, \bar{z}), \epsilon))$ tal que

$$H_k(x) = (x, T_k(H(x))).$$

Afirmção. É válido que H_k é contínua.

De fato, é suficiente provar que T_k é k -Lipschitz contínua. De equação (4.1.7) podemos observar que T_k é a composição das funções $A : u \mapsto z^k + c_k u$ e os resolventes do operador T^{-1} , isto é,

$$T_k = J_k^{T^{-1}} \circ A.$$

Como T^{-1} é monótono e da Proposição 2.1.41, $J_k^{T^{-1}}$ é não-expansivo, isto implica, segundo a Proposição 2.1.36, que $J_k^{T^{-1}}$ é k -Lipschitz contínua. Então, do fato que A é uma função k -Lipschitz contínua (pois é linear) temos que T_k é k -Lipschitz contínua. Provando assim a afirmação.

Então, X_k é aberto tal que $x^{k+1} \in X_k$ soluciona (\overline{P}_k) . Agora, vamos provar a unicidade de x^{k+1} . Suponhamos que \tilde{x}^{k+1} é outra solução em X_k que soluciona (\overline{P}_k) , então do Lema 4.2.1, temos que $H_k(\tilde{x}^{k+1}) \in B((\bar{x}, \bar{z}), \epsilon)$ deve solucionar o problema (\overline{PD}_k) , o qual é uma contradição do fato que apenas (x^{k+1}, z^{k+1}) é a única solução desse problema. ■

4.3 Aplicação dos métodos multiplicadores proximais para problemas (PNL)

Nesta seção, vamos estudar os métodos multiplicadores proximais aplicados a problemas (PNL) no contexto de dualidade geral com monotonicidade, o qual foi dado na Seção

2.1. Consideremos $S(x) = \nabla f_0(x)$ e $T_k(x) = N_K(x)$ no Algoritmo 4.2.2 dos métodos multiplicadores proximais.

De (4.2.2), $z^{k+1} = T_k(H(x^{k+1}))$. Então, é importante mencionar que para calcular z^{k+1} no Algoritmo 4.2.2, é mais conveniente ter uma representação mais conhecida e amigável do T_k . Com este objetivo e considerando $T = N_K$, de (4.1.7), (1.1.10), (2.1.5) e (2.1.6) temos

$$\begin{aligned} T_k(u) &= (I + c_k T^{-1})^{-1}(z^k + c_k u) \\ &= (I + c_k N_K^{-1})^{-1}(z^k + c_k u) \\ &= (I + c_k N_{K^\otimes})^{-1}(z^k + c_k u) \\ &= P_{K^\otimes}(z^k + c_k u) \end{aligned} \tag{4.3.1}$$

$$= \frac{1}{2} \nabla d_K^2(z^k + c_k u). \tag{4.3.2}$$

Então, substituindo (4.3.2) em (4.2.1), o problema primal (\overline{P}_k) para o problema (PNL) pode ser escrito como

$$0 \in \nabla f_0(x) + c_k^{-1}(x - x^k) + DH(x)^T \frac{1}{2} \nabla d_K^2(z^k + c_k H(x)),$$

o qual pode ser generalizado no seguinte problema:

$$\min \{ \psi_k(x), \quad x \in X \},$$

onde

$$\psi_k(x) = f_0(x) + \frac{1}{2c_k} \|x - x^k\|^2 + \frac{1}{2c_k} d_K^2(y^k + c_k H(x)). \tag{4.3.3}$$

Mais ainda, do fato que $K^\otimes = \mathbb{R}^r \times \mathbb{R}_+^{m-r}$ temos que

$$\begin{aligned} y_i^{k+1} &= y_i^k + c_k f_i(x^{k+1}), & i = 1, \dots, r \\ y_i^{k+1} &= \max\{y_i^k + c_k f_i(x^{k+1}), 0\}, & i = r + 1, \dots, m \end{aligned}$$

Então, de acordo com o Algoritmo 4.2.2, segue o algoritmo dos métodos multiplicadores proximais para resolver o problema (PNL).

Algoritmo 4.3.1.

1. Escolher $(x^0, z^0) \in X \times Z$ arbitrário e estabelecer $k = 0$.

2. Resolver

$$\min \{ \psi_k(x), \quad x \in X \} \tag{4.3.4}$$

como foi definido em (4.3.3) para obter x^{k+1} .

3. Estabelecer

$$\begin{aligned} z_i^{k+1} &= z_i^k + c_k f_i(x^{k+1}), & i &= 1, \dots, r \\ z_i^{k+1} &= \max\{z_i^k + c_k f_i(x^{k+1}), 0\}, & i &= r+1, \dots, m \\ k &\leftarrow k+1 \end{aligned}$$

O próximo teorema garante a convergência do Algoritmo 4.3.1, antes de enunciarmos apresentaremos alguns resultados necessários para a prova deste teorema.

Definimos o conjunto.

$$\theta_{D,B}(u) = \sup_{d \in D} \{\langle d, u \rangle - \frac{1}{2} \langle d, Bd \rangle\}. \quad (4.3.5)$$

Lema 4.3.2 *Consideremos a função ψ_k como foi definida em (4.3.3), então uma função lagrangeana para uma parametrização dada de ψ_k é dada por*

$$l(x, z) = L(x, z) - \delta_{K^\otimes}(z) + \delta_X(x),$$

onde $L(x, z) = \tilde{f}_0(x) + \langle z, \tilde{H}(x) \rangle - \frac{1}{2} \langle z, Bz \rangle$ com $\tilde{f}_0(x) = f_0(x) + \frac{1}{2} \|x - x^k\|^2$.

Prova. De fato, do Lema 1.1.17 temos que

$$\begin{aligned} \frac{1}{2} d_K^2(z^k + c_k H(x)) &= \sup_{z \in K^\otimes} \{\langle z^k + c_k H(x), z \rangle - \frac{1}{2} \langle z, z \rangle\} \\ &= c_k \theta_{K^\otimes, B}(c_k^{-1} z^k + H(x)), \end{aligned}$$

onde $B = c_k^{-1} I$.

Por outro lado, o problema (4.3.4) pode ser formulado no contexto de programação composta como

$$\min \left(\tilde{f}_0 + (\theta_{K^\otimes, B} \circ \tilde{H}) \right) (x), \quad (4.3.6)$$

onde $\tilde{f}_0(x) = f_0(x) + \frac{1}{2c_k} \|x - x^k\|^2$ e $\tilde{H}(x) = c_k^{-1} z^k + H(x)$.

Consideremos uma parametrização de $\psi_k(x)$ dado por

$$\tilde{\psi}_k(x, u) = \tilde{f}_0(x) + \theta_{K^\otimes, B}(\tilde{H}(x) + u),$$

onde $\tilde{\psi}_k(x, 0) = \psi_k(x)$.

Logo, o lagrangeano associado com $\tilde{\psi}_k(x, z)$ é dado por:

$$\begin{aligned}
 l(x, z) &= \inf_u \{ \tilde{\psi}_k(x, u) - \langle z, u \rangle \} \\
 &= \tilde{f}_0(x) + \inf_u \{ \theta_{K^\otimes, B}(\tilde{H}(x) + u) - \langle z, u \rangle \} \\
 &= \tilde{f}_0(x) + \langle z, \tilde{H}(x) \rangle - \sup_u \{ \langle z, \tilde{H}(x) + u \rangle - \theta_{K^\otimes, B}(\tilde{H}(x) + u) \} \\
 &= \tilde{f}_0(x) + \langle z, \tilde{H}(x) \rangle - \theta_{K^\otimes, B}^*(z),
 \end{aligned} \tag{4.3.7}$$

onde $\theta_{K^\otimes, B}^*$ é a conjugada de $\theta_{K^\otimes, B}$.

Mais ainda, se considerarmos $\alpha(u) = \frac{1}{2} \langle u, Bu \rangle$ temos que

$$\theta_{K^\otimes, B}(u) = \sup_y \{ \langle y, u \rangle - (\alpha + \delta_{K^\otimes})(y) \} = (\alpha + \delta_{K^\otimes})^*(u).$$

Observe que $\alpha + \delta_{K^\otimes}$ é convexa. Então, da Proposição 1.1.6 temos que

$$\theta_{K^\otimes, B}^*(u) = \alpha(u) + \delta_{K^\otimes}(u) = \frac{1}{2} \langle u, Bu \rangle + \delta_{K^\otimes}(u) \tag{4.3.8}$$

Sustituindo (4.3.8) em (4.3.7) temos que

$$\begin{aligned}
 l(x, z) &= \tilde{f}_0(x) + \langle z, \tilde{H}(x) \rangle - \left(\frac{1}{2} \langle z, Bz \rangle + \delta_{K^\otimes}(z) \right) \\
 &= L(x, z) - \delta_{K^\otimes}(z),
 \end{aligned} \tag{4.3.9}$$

com $L(x, z) = \tilde{f}_0(x) + \langle z, \tilde{H}(x) \rangle - \frac{1}{2} \langle z, Bz \rangle$.

Proposição 4.3.3 [38, Exercício 13.26] *Considere o problema*

$$\min \{ f_0(x) + \theta_{D, B}(f_1(x), f_2(x), \dots, f_n(x)), \quad \forall x \in X \}$$

para funções $f_i \in \mathcal{C}^2$, onde $H(x) = (f_1(x), \dots, f_n(x))$.

Definimos

$$\begin{aligned}
 L(x, z) &= f_0(x) + \langle z, H(x) \rangle - \frac{1}{2} \langle z, Bz \rangle, \\
 Z(\bar{x}) &= \{ \bar{z} \mid -\nabla_x L(\bar{x}, \bar{z}) \in N_X(\bar{x}), \nabla_z L(\bar{x}, \bar{z}) \in N_Z(\bar{z}) \}, \\
 X'(\bar{x}) &= T_X(\bar{x}) \cap \nabla_x L(\bar{x}, \bar{z})^\perp, \quad Z'(\bar{x}) = T_Z(\bar{z}) \cap \nabla_z L(\bar{x}, \bar{z})^\perp.
 \end{aligned}$$

(a) Se \bar{x} é um ótimo local então $Z(\bar{x}) \neq \emptyset$ e

$$\max_{z \in Z(\bar{x})} \langle w, \nabla_{xx}^2 L(\bar{x}, \bar{z})w \rangle + 2\theta_{Z'(\bar{x}), B}(DH(x)^T w) \geq 0, \quad \forall w \in X'(\bar{x}).$$

(b) Reciprocamente, se essas condições tomam e a desigualdade do item (a) é estrita para todo $w \neq 0$ então \bar{x} é um ótimo local.

Teorema 4.3.4 *Seja (\bar{x}, \bar{z}) um ponto satisfazendo as condições KKT para o problema (PNL), satisfazendo também as condições suficientemente forte de segunda ordem e que $\{\nabla f_i(\bar{x})\}_{i \in I_1(\bar{x})}$ é linearmente independente. Se a sequência $\{c_k\}$ é suficientemente grande tal que $c_k \nearrow \bar{c} \leq \infty$ e se $\|(x^0, z^0) - (\bar{x}, \bar{z})\|$ é suficientemente pequeno, então existe uma sequência $\{(x^k, z^k)\}$ gerada pelo Algoritmo 4.3.1 numa vizinhança X_k tal que para cada k , x^{k+1} é a única solução em X_k ao problema primal (\bar{P}_k) . Também, a sequência $\{(x^k, z^k)\}$ converge linearmente e Fejér monótona para (\bar{x}, \bar{z}) com taxa de convergência $r(\bar{c}) < 1$.*

Demonstração. Vamos provar que para c_k suficientemente grande o único ponto KKT do problema (4.3.4) é um ótimo local da função ψ_k , o qual foi definido em (4.3.3).

Segundo o Lema 4.3.2, o lagrangeano da função ψ_k é dado por $l(x, z) = L(x, z) + \delta_X(x) - \delta_{K^\otimes}(z)$. Além disso, uma condição de otimalidade vem dado quando solucionamos os seguintes problemas:

$$0 \in \partial_x l(x, z), \quad 0 \in \partial_z(-l)(x, z), \quad (4.3.10)$$

isto é, $-\nabla_x L(x, y) = 0$ e $\nabla_y L(x, y) \in N_{K^\otimes}(\bar{y})$.

Definimos o conjunto

$$K^\otimes(x^{k+1}) = \{z \mid -\nabla_x L(x^{k+1}, z) = 0, \nabla_z L(x^{k+1}, z) \in N_{K^\otimes}(z)\}.$$

Então, $K^\otimes(x^{k+1})$ é exatamente o conjunto dos z satisfazendo o seguinte sistema

$$\begin{aligned} 0 &= \nabla f_0(x) + DH(x)^T z + c_k^{-1}(x - x^k), \\ 0 &\in N_{K^\otimes}(z) - H(x) + c_k^{-1}(z - z^k). \end{aligned}$$

Observe que o sistema dado acima coincide com o problema (\overline{PD}_k) para solucionar o problema (PNL). Segundo o Lema 4.2.1, as soluções deste sistema são os pontos

dados por $\{(x, T_k(H(x))) \mid \nabla \psi_k(x) = 0\}$. Mais ainda, de (4.3.1) temos que $T_k(u) = P_{K^\otimes}(z^k + c_k u)$ e como K^\otimes é convexo e fechado, do fato de K ser um cone, segundo o Teorema 1.1.3 temos que z^{k+1} é único, isto é,

$$K^\otimes(x^{k+1}) = \{z^{k+1}\}. \quad (4.3.11)$$

Seja

$$\begin{aligned} (K^\otimes)'(x^{k+1}) &= T_{K^\otimes}(z^{k+1}) \cap \nabla_y \tilde{L}(x^{k+1}, z^{k+1})^\perp \\ &= T_{K^\otimes}(z^{k+1}) \cap [H(x^{k+1}) + c_k^{-1}(z^{k+1} - z^k)]^\perp. \end{aligned} \quad (4.3.12)$$

Agora, precisamos estabelecer condições suficientemente para que x^{k+1} seja minimizador local do problema (4.3.6). Segundo Rockafellar[38, Exercício 13.26], uma condição de otimalidade de segunda ordem para funções com penalidade é dado como segue:

$$\max_{y \in K^\otimes(x^{k+1})} \langle w, \nabla_{xx}^2 \tilde{L}(x^{k+1}, y)w \rangle + 2\theta_{(K^\otimes)'(x^{k+1}), B}(D\tilde{H}(x^{k+1})^T w) > 0, \quad \forall w \neq 0,$$

ou equivalentemente,

$$\max_{y \in K^\otimes(x^{k+1})} \langle w, (\nabla_{xx}^2 l(x^{k+1}, y) + c_k^{-1}I)w \rangle + 2\theta_{(K^\otimes)'(x^{k+1}), B}(DH(x^{k+1})^T w) > 0, \quad \forall w \neq 0. \quad (4.3.13)$$

Definimos

$$Q = \{y \in \mathbb{R}^m \mid y_i = 0, \forall i \notin I_1\}.$$

Afirmação. É válido que $Q \subset (K^\otimes)'(x^{k+1})$.

De fato, primeiro vamos mostrar que $\overline{K^\otimes} \subset T_{K^\otimes}(y^{k+1})$. Seja $y^{k+1} \in Q$ o que implica que $y_i^{k+1} = 0$ para todo $i \notin I_1(\bar{x})$. Como $\{y^{k+1}\}$ é Fejér monótono temos que para todo i , $|y_i^{k+1} - \bar{y}_i| < \epsilon$. Daí, $\bar{y}_i \leq y_i^{k+1}$. Logo,

$$y_i^{k+1} > 0, \quad i \in I_1(\bar{x}) \setminus \{1, \dots, r\}. \quad (4.3.14)$$

então $Q \subset T_{K^\otimes}(y^{k+1})$. Ademais, como (x^k, y^k) resolve o problema primal-dual modificado para o problema (PNL) e de (4.3.14) temos que $f_i(x^{k+1}) - c_k^{-1}(y_i^{k+1} - y_i^k) = 0$. Daí, $Q = T_{K^\otimes}(y^{k+1})$. Portanto, $Q \subset (K^\otimes)'(x^{k+1})$. Obtendo assim a nossa afirmação.

Alem disso, do fato que $Q \subset (K^\otimes)'(x^{k+1})$, do Lema 1.1.17 e da definição de $\theta_{(K^\otimes)'(x^{k+1}),B}$ temos que

$$\theta_{(K^\otimes)'(x^{k+1}),B}(u) \geq \sup_{y \in Q} \left\{ \langle y, u \rangle - \frac{1}{2c_k} \|y\|^2 \right\} = \frac{c_k}{2} d_{Q^\otimes}^2(u) = \frac{c_k}{2} \sum_{i \in I_1} |u_i|^2.$$

Como $D\tilde{H}(x^{k+1}) = DH(x^{k+1})$, temos que

$$\theta_{(K^\otimes)'(x^{k+1}),B}(D\tilde{H}(x^{k+1})^T w) \geq \frac{c_k}{2} \sum_{i \in I_1(\bar{x})} |\langle \nabla f_i(x^{k+1}), w_i \rangle|^2 \geq 0. \quad (4.3.15)$$

Segundo a hipótese, o par (\bar{x}, \bar{z}) satisfaz a condição suficientemente forte de segunda ordem. Então, da Proposição 4.3.3, existe $c \in \mathbb{R}$ tal que

$$\langle w, \nabla_{xx}^2 l(\bar{x}, \bar{z}) w \rangle + c \sum_{i \in I_1(\bar{x})} \langle \nabla f_i(\bar{x}), w_i \rangle^2 > 0, \quad \forall w \neq 0.$$

Mais ainda, do fato que $\nabla_{xx}^2 l(\cdot, \cdot)$ e $\nabla f_i(\cdot)$ são funções contínuas e que a sequência (x^k, z^k) é Fejér monótona, então para (x^0, z^0) suficientemente perto de (\bar{x}, \bar{z}) temos que

$$\langle w, \nabla_{xx}^2 l(x^{k+1}, z^{k+1}) w \rangle + c_k \sum_{i \in I_1(\bar{x})} \langle \nabla f_i(x^{k+1}), w_i \rangle^2 > 0, \quad \forall w \neq 0. \quad (4.3.16)$$

Logo, de (4.3.15) e (4.3.16) temos que para todo $w \neq 0$,

$$\begin{aligned} 0 &< \langle w, \nabla_{xx}^2 l(x^{k+1}, z^{k+1}) w \rangle + 2\theta_{(K^\otimes)'(x^{k+1}),B}(D\tilde{H}(x^{k+1})^T w), \\ &= \langle w, \nabla_{xx}^2 L(x^{k+1}, z^{k+1}) w \rangle + 2\theta_{(K^\otimes)'(x^{k+1}),B}(D\tilde{H}(x^{k+1})^T w). \end{aligned}$$

A última desigualdade é obtida do fato que $l(x, z) = L(x, z) - \delta_{K^\otimes}(z)$. Então, segundo a Proposição 4.3.3, garante-se que x^{k+1} é minimizador local do problema (4.3.6).

Por outro lado, como o ponto $KKT(\bar{x}, \bar{z})$ satisfaz a condição suficientemente forte de segunda ordem e a independência linear dos gradientes $\nabla f_i(\bar{x})$ para $i \in I_1(\bar{x})$, segundo o Teorema 1.2.6, são condições suficientes para que o sistema KKT do problema (PNL) seja fortemente regular no ponto (\bar{x}, \bar{z}) .

Além disso, do Exemplo 2.2.2 vimos que o sistema KKT para el problema (PNL) coincide com o problema Lagrangeano $(0, 0) \in L(\bar{x}, \bar{z})$ com L definido em (2.2.5). Daí, L é um operador fortemente regular no ponto (\bar{x}, \bar{z}) , o que equivale dizer, segundo a Definição 1.2.3, que L^{-1} tem uma k -Localização Lipschitz. Portanto, do Teorema 4.2.3 obtemos a convêrgencia do algoritmo 4.3.1. ■

Capítulo 5

Conclusões

Neste trabalho apresentamos de forma didática a teoria de operadores monótonos máximos como também os resultados de convergência dos algoritmos:

- Algoritmo de ponto proximal relaxado (APPR) introduzido por Eckstein e Bertsekas, [9]
- Algoritmo de ponto proximal (APP) com monotonicidade local, estudado por Pennanen [29].
- (APP) sem monotonicidade, estudado também por Pennanen [29].

Alem disso, apresentamos as aplicações destes resultados aos métodos de multiplicadores e fazemos uma discussão da implementação dos algoritmos propostos.

Do ponto de vista teórico podemos concluir que estes algoritmos tem boas propriedades de convergência.

Bibliografia

- [1] ABADIE J.M.- *On the Kuhn-Tucker theorem*. Nonlinear Programming, J. Abadie ed., John Wiley, New York, p.21-36. 1967.
- [2] ANDREANI R., MARTINEZ J.M., SCHUVERDT M.L.- *The CPLD condition of Q_i and Wei implies the quasinormality constraint qualification*. Journal of Opt. theory and App., 125(2) p.473-485. 2005
- [3] ANTIPIN A.S.- *On a method for convex programs using a symetrical modification of the Lagrange function*. Ekonomika i Mat. Metody 12, p. 1164-1173. 1976.
- [4] AUSLENDER A., TEBOULLE M.- *Asymptotic cones and functions in optimization and variational inequalities*. Springer, New York, NY, USA. 2003
- [5] BERTSEKAS D.P.- *Multiplier methods : A survey*. Automatica, 12, p.133-145. 1976.
- [6] BERTSEKAS D.P., NEDIĆ A., OZDAGLAR E.- *Convex analysis and optimization* Athena scientific. 2003.
- [7] BRÉZIS H., LIONS P.L.- *Produits infinis de resolvantes*. Israel Journal of Mathematics 29(4), pp.329-345. 1978.
- [8] BURACHIK R.S., IUSEN A.N.- *Set-valued mappings and enlargements of monotone operators*. Springer-Verlag. NY. 2008.
- [9] ECKSTEIN J., BERTSEKAS D.- *On the Douglas-Ratford splitting method and the proximal point algorithm for masximal monotone operators*. Math. Programming 55, pp. 293-318. 1992.

- [10] ECKSTEIN J., FERRIS C.- *Smooth methods of multipliers for complementarity problems*. Math. Programming 86, pp.65-90. 1999.
- [11] FENCHEL W.,- *On conjugate convex functions*. Canadian Journal of Mathematics 1. pp.73-77. 1949.
- [12] FUKUSHIMA M., QI L.,- *A global and superlinear convergent algorithm for nonsmooth convex minimization*. SIAM J. Optim, 6 pp.1106-1120. 1996
- [13] GOL'SHTEIN E.G.- *Method for modification of monotone mappings*. Ekonomika i Matemacheskaya Metody 11(6): pp.1142-1159. 1975.
- [14] GOL'SHTEIN E. G., TRET'YAKOV N.V.- *Modified Lagrangians in convex programming and their generalizations* . Mathematical Prog. Study 10, p.86-97. 1979.
- [15] GÜLER O. - *On the convergence of the proximal point algorithm for convex minimization*. SIAM J. Control Optim.29, pp.403-419. 1991.
- [16] HESTENES, M.R. - *Optimization Theory, The finite dimensional case*. John Wiley, N.Y. 1975.
- [17] HESTENES, M.R. - *Multiplier and Gradient Methods*.J.O.T.A., v.4, n5,p.303-320.1969
- [18] HIRIART-URRUTY J.B., LEMARÉCHAL C.,- *Convex analysis and minimization algorithms. Vol. I, II* Springer Verlag, Berlin. 1993.
- [19] IUSEM, A.- *Métodos de ponto proximal em otimização* . 20º Colóquio Brasileiro de matemática. Rio de Janeiro. IMPA, 24-28 julho, 1995.
- [20] IZMAILOV A., SOLODOV E. - *Otimização . Volume 1. Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. Rio de Janeiro. IMPA, 2005.
- [21] LUQUE F.J. - *Asymptotic convergence analysis of the proximal point algorithm*. SIAM J. on Control and Optim. 22(2), pp. 277-293. 1984.

- [22] MANGASARIAN O.L., FROMOVITZ S.- *The Fritz-John necessary optimality conditions in presence of equality and inequality constraints* . Journal of Mathematical Analysis and Applications 17, p. 37-47. 1967.
- [23] MARTINET B.- *Regularisation d'inéquations variationnelles par approximations successives* . Rev. Francaise Inf. Rech. Oper., 4 p.154-159. 1970.
- [24] MINTY G.J. - *Monotone(nonlinear) operators in Hilbert space*. Duke Mathematics Journal 29, p.341-346. 1962.
- [25] MOREAU, J. - *Proximité et dualité dans un espace Hilbertien*. Bulletin de la Société Mathématique de France 93, pp.273-299. 1965.
- [26] PENNANEN T. - *Dualization of monotone generalized equations*. Thesis. University of Washington. 1999.
- [27] PENNANEN T. - *Dualization of generalized equations of maximal monotone type*. SIAM J. Optim. 10 (3), p.809-835. 2000.
- [28] PENNANEN T.- *Graph convex mappings and K-convex functions* J. Convex Anal.,6. 1999.
- [29] PENNANEN T. - *Local convergence of the proximal point algorithm and multiplier methods without monotonicity*. Math. oper. Res. 27, pp.170-191. 2002.
- [30] POWELL, M.J.D. - *A Method for Nonlinear Constraints in Minimization Problems*. In: R. Fletcher, Optimization, New York, Academic Press, p. 283-298. 1969.
- [31] ROBINSON S. M. - *Strongly regular generalized equations*. Math. Oper. Res. 5, pp. 43-62. 1980.
- [32] ROCKAFELLAR R.T. - *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*. Math. O.R. 1, p.97-116. 1976
- [33] ROCKAFELLAR R.T. - *Conjugate duality and optimization*. SIAM, Philadelphia, PA. 1974

- [34] ROCKAFELLAR R.T. - *Convex analysis*. Princeton Univ. Press, Princeton, 1970.
- [35] ROCKAFELLAR R.T. - *On the maximality of sums of nonlinear monotone operators*. Transactions of the American Math. Society. Vol. 149, p.75-89. 1970.
- [36] ROCKAFELLAR R.T. - *Monotone operators and the Proximal Point Algorithm*. SIAM J. Control Optim. 14, p.877-898. 1976.
- [37] ROCKAFELLAR R.T. - *The multiplier method of Hestenes and Powell applied to convex programming*. Journal of Opt. Theory and App. 12, 6 , p.555-562. 1973.
- [38] ROCKAFELLAR R.T., WETS J.B. - *Variational Analysis*. Springer. Berlin. 1998.
- [39] SLATER M.- *Lagrange multipliers revisited: a contribution to non-linear programming*. Cowles Comission Discussion paper 7. 1950.
- [40] SOLODOV M.V., SVAITER B.F.- *Forcing strong convergence of proximal point iterations in a Hilbert space*. Math. Program. Ser. A,14, p.189-202. 2000.