

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CURSO DE PÓS-GRADUAÇÃO EM MATEMÁTICA E
COMPUTAÇÃO CIENTÍFICA

UM TEOREMA DE EQUIVALÊNCIA ENTRE MÉTODOS
LAGRANGEANO AUMENTADO E ALGORITMOS DE PONTOS
PROXIMAIS

Dissertação apresentada ao curso de Pós-graduação em Matemática e Computação Científica, do Centro de Ciências Exatas da Universidade Federal de Santa Catarina, para a obtenção do grau de Mestre em Matemática. Área de Concentração : Otimização.

A todos os professores que tive a sorte de encontrar no caminho da minha formação. Em especial a Renee Espinoza e Guillermo Rezza, meus pais, Esperanza, Victor Echevarria, Rosa Hidalgo, Armando Venero, Pedro Canales, Felix Escalante e Veronique Colin.

Agradecimentos

Aos meus irmãos, meus amigos e a meus amigos, meus irmãos por escolha. Agradecimentos especiais para Mica, Magen, Celia, Ceci, William, Marlene, Darcia, Ana Claudia, Laura e Michelli.

Agradecimentos especiais também para meu orientador o professor Clovis Gonzaga e para o professor Eduardo Alarcon.

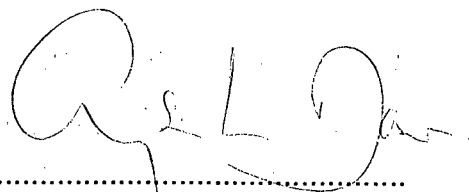
A Elizabeth, pelos desenhos.

Ao CNPQ pelo apoio recebido.

UM TEOREMA DE EQUIVALÊNCIA ENTRE MÉTODOS
LAGRANGEANO AUMENTADO E ALGORITMOS DE
PONTOS PROXIMAIS

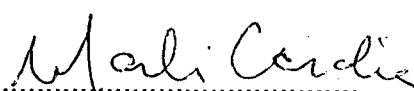
YNA CONSUELO REZZA ESPINOZA

Esta Dissertação foi julgada adequada para a obtenção do Título de
Mestre, Área de concentração em Otimização e aprovada em sua
forma final pelo Curso de Pós-Graduação em Matemática e
Computação Científica

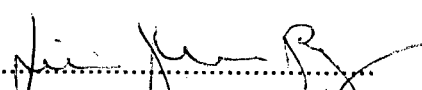


Coordenador

Comissão Examinadora

Prof. Dr. .....

(Membro-UFPR)

Prof. Dr. .....

(Membro-UFSC)

Prof. Dr. .....

(Membro-UFSC)

Florianópolis, 25 de Setembro de 1998

Resumo

Este trabalho apresenta uma prova geral do conhecido Teorema de Equivalência que relaciona o Método de Ponto Proximal e o Método Lagrangeano Aumentado, a qual inclui quase todos os casos existentes na literatura.

Seguindo um caminho lógico, fazemos primeiramente um estudo detalhado do Método de Ponto Proximal e sua evolução desde sua introdução em 1970 até a atualidade, aplicado ao problema específico de minimizar uma função convexa própria fechada. Dentro deste tema que se encontra bastante disperso, damos ordem e simplicidade ao dar condições suficientes para a boa definição do algoritmo de Ponto Proximal.

Em seguida, descrevemos o Método Lagrangeano Aumentado incluindo penalidades não estritamente convexas e provamos a boa definição do algoritmo correspondente. Sem precisar fazer uso da relação existente entre os Métodos Lagrangeano Aumentado e os de Ponto Proximal, obtemos novos resultados de unicidade a respeito da seqüência de multiplicadores (seqüência dual) gerada pelo algoritmo Lagrangeano Aumentado.

Para provar o Teorema de Equivalência, definimos o núcleo a ser utilizado no Método de Ponto Proximal como o somatório das conjugadas das penalidades usadas no Método Lagrangeano Aumentado e provamos a boa definição do algoritmo de Ponto Proximal respectivo, aplicado ao problema dual. Este núcleo pode não ser coercivo na fronteira. Finalmente, provamos que as seqüências $\{\mu^k\}$ geradas por cada um dos métodos são realmente as mesmas.

Abstract

We present here a general proof of the well known Equivalence Theorem which relates the Proximal Point and the Augmented Lagrangian methods. This proof includes almost all the cases existent in the literature.

Following a logical path we first do a detailed study of the Proximal Point method and its evolution since its introduction in 1970 until now when applied to the specific problem of minimizing a closed proper convex function. This subject is very dispersed and so we have attempted to put order and simplicity in it by giving sufficient conditions for the Proximal Point algorithm to be well defined.

Next we have described the Generalized Augmented Lagrangian method and we have proved the well-definiteness of the corresponding algorithm. Without making use of the relationship between the Proximal Point and the Augmented Lagrangian methods we have obtained new uniqueness results on the sequence of multipliers (dual sequence) generated by the Augmented Lagrangian algorithm.

In order to prove the Theorem of Equivalence we defined the kernel to be used in the Proximal Point method as the summation of the conjugates of the penalties of the Augmented Lagrangian method and then we proved the well-definiteness of the associated Proximal Point algorithm when applied to the dual problem. Finally we proved that the sequences $\{\mu^k\}$ generated by each of the methods are indeed the same.

Índice

1	Introdução	1
2	Estudo do Método de Ponto Proximal para o problema irrestrito	3
2.1	Descrição do Método do Ponto Proximal	4
2.2	Condições Suficientes para a boa definição do Algoritmo de P.P. com quase-distâncias como núcleo	6
2.3	Núcleo quadrático	7
2.3.1	Noções relativas à convergência	8
2.4	Distâncias de Bregman	9
2.4.1	Definição. A boa definição do Algoritmo de Ponto Proximal com distâncias de Bregman	10
2.4.2	Resultados preliminares	12
2.4.3	Convergência	14
2.5	Núcleos estritamente convexos ϕ	17
2.5.1	A boa definição do algoritmo	17
3	Estudo do Método de Ponto Proximal para o problema restrito	20
3.1	Distâncias de Bregman induzidas por funções de Bregman coercivas na fronteira	21
3.1.1	A boa definição	22
3.1.2	Convergência	24
3.2	φ -divergências	25
3.2.1	A boa definição	27
3.2.2	Convergência	29

4	O Método Aumentado	30
4.1	As funções Lagrangeano e Lagrangeano Aumentado	31
4.2	Resultados preliminares à descrição do método	32
4.3	Descrição do Método Lagrangeano Aumentado	36
4.4	Definições e Resultados Preliminares	38
4.5	A boa definição do Algoritmo Lagrangeano Aumentado com penali- dades $P_i \in \mathcal{P}$	41
4.6	Tipos de penalidades usadas pelo Método Lagrangeano Aumentado .	43
4.6.1	Penalidade Tipo I ($b = +\infty$)	44
4.6.2	Penalidade Tipo II ($b > 0$ finito)	48
4.7	Uma análise da seqüência $\{\mu^k\}$	50
4.8	A relação entre o Método L.A. com Penalidades $P_i \in \mathcal{P}$ e o Método de Ponto Proximal	52
4.9	A boa definição do Algoritmo de Ponto Proximal com núcleo N . . .	55
4.10	A equivalência entre o Método Lagrangeano Aumentado e o Método de Ponto Proximal	57
	Bibliografia	60

Capítulo 1

Introdução

Problemas não lineares de otimização surgem espontaneamente em muitos campos de aplicação. O primeiro destes problemas que estudaremos é o seguinte:

$$\begin{array}{ll} \text{minimizar } f(x) & (I) \\ \text{s.a. } x \in \bar{S} \end{array}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa própria fechada, S é aberto e convexo e \bar{S} é sua aderência.

Um método, muito elegante do ponto de vista teórico, para resolver este problema é o chamado Método de Ponto Proximal introduzido por Martinet [17] lá pelos anos 70, o qual consiste em substituir o problema inicial (I) por uma seqüência de problemas regularizados (ou seja, com solução única). É deste método que nos ocupamos nos Capítulos 2 e 3 apresentando os diferentes núcleos existentes na literatura tais como o quadrático, as distâncias de Bregman [2], as φ -divergências [23] e os núcleos estritamente convexos [11], assim como seus respectivos resultados de convergência, analisando, por separado, o caso irrestrito (onde $S = \mathbb{R}^n$) e depois o problema restrito.

Já para o problema seguinte:

$$\begin{aligned} &\text{minimizar } f(x) && (II) \\ &\text{s.a } g_i(x) \leq 0 \quad , \quad i = 1, \dots, m \\ & \quad \quad \quad x \in \mathbb{R}^n \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ para $i = 1, \dots, m$ são funções convexas, existe um método de otimização muito utilizado devido à sua facilidade de implementação, estamos-nos referindo ao Método Lagrangeano Aumentado (ou também chamado de multiplicadores), introduzido por Hestenes [10] e Powell [19] em 1969. No Capítulo 4 nos ocupamos deste método. As penalidades que definimos incluem as usadas neste método por diversos pesquisadores tais como Rockafellar [22], Bertsekas [1], Polyak-Teboulle [18] e Gonzaga-Castillo [7]. Analisamos a seqüência de multiplicadores e mostramos um interessante resultado de unicidade a seu respeito.

Com relação a estes métodos, Polyak-Teboulle [18] mostraram um caso no qual o M.L.A. (aplicado ao problema II) e o de P.P. (aplicado ao dual de II) podem ser vistos como equivalentes. Outras provas desta equivalência, para o caso do Lagrangeano Aumentado Clássico e distâncias de Bregman separáveis encontram-se em [12]. A prova do Teorema de equivalência que apresentamos inclui todos estes casos. Nós definimos o núcleo do método de P.P. como o somatório das conjugadas das penalidades usadas no M.L.A. A maior parte do desenvolvimento teórico que realizamos baseia-se nos textos de Lemaréchal-Hiriart-Urruty [8] e [9].

Capítulo 2

Estudo do Método de Ponto Proximal para o problema irrestrito

Definição 2.1 Diz-se que a função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa se para todos $x, x' \in \mathbb{R}^n$ e todo $\alpha \in (0, 1)$ temos: $f(\alpha x + (1 - \alpha)x') \leq \alpha f(x) + (1 - \alpha)f(x')$, considerando esta desigualdade em $\mathbb{R} \cup \{+\infty\}$.

Se além disso, f não é identicamente $+\infty$, diz-se que f é convexa própria.

Esta classe de funções será denotada por $\text{Conv}\mathbb{R}^n$.

Definição 2.2 O domínio (ou domínio efetivo) de $f \in \text{Conv}\mathbb{R}^n$ é o conjunto não vazio seguinte: $\text{dom} f = \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

Definição 2.3 A função $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ se diz fechada, se para cada $x \in \mathbb{R}^n$ verifica-se $\liminf_{y \rightarrow x} f(y) \geq f(x)$.

Se $f \in \text{Conv}\mathbb{R}^n$ e além disso f é fechada, denotaremos $f \in \overline{\text{Conv}}\mathbb{R}^n$.

O problema que estamos interessados em resolver neste capítulo é o seguinte:

$$\begin{aligned} &\text{minimizar } f(x) && (P) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma função convexa própria e fechada.

Hipótese 2.4 *Suponhamos que o problema (P) tem solução, ou seja, o mínimo de f existe; isto é, $\exists x^* \in \mathbb{R}^n$ tal que $f(x^*) \leq f(x) \forall x \in \mathbb{R}^n$.*

2.1 Descrição do Método do Ponto Proximal

Antes definimos algumas ferramentas que usaremos para descrevê-lo.

Definição 2.5 *Seja A um subconjunto de \mathbb{R}^n . Uma função $D : A \times A \rightarrow \mathbb{R}$ é uma quase-distância em A se:*

- i) $D(x, y) \geq 0 \forall x, y \in A$*
- ii) $D(x, y) = 0$ se e somente se $x = y$*

Seja $\{\lambda_k\}$ uma seqüência limitada de números reais positivos e $N : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ uma quase-distância em \mathbb{R}^n , conhecida como **núcleo**.

O método aproxima uma solução x^* do problema de otimização convexa (P) mediante a seqüência $\{x^k\} \subset \mathbb{R}^n$ gerada ao resolver iterativamente partindo de algum $x^0 \in \mathbb{R}^n$ dado, problemas do tipo:

$$\begin{aligned} &\text{minimizar } \{f(x) + \lambda_k N(x, x^k)\} && (P_k) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

onde, para cada $k \in \mathbb{N}$, o problema (P_k) deverá ter solução única, a qual denotaremos x^{k+1} .

Para conseguir este objetivo e também para obter os resultados de convergência que completariam o método (convergência da seqüência $\{x^k\}$ ou pelo menos de uma subseqüência dela a uma solução do problema (P)), a quase-distância N deverá adquirir certas características que darão origem a núcleos particulares os quais serão estudados posteriormente (nas seções (2.3), (2.4) e (2.5)).

Antes, ocupar-nos-emos de conseguir o primeiro objetivo, para isso, lembramos resultados que serão de utilidade, colocamos uma hipótese sobre o núcleo N e provamos a afirmação que aparece a seguir.

Teorema 2.6 *Sejam (M, d) um espaço métrico, A um subconjunto de M e $f : A \rightarrow \mathbb{R}$ uma função contínua. Se $K \subset A$ é compacto, então f é limitada em K , isto é : $B = \{f(x) : x \in K\} \subset \mathbb{R}$ é limitado. Portanto, existem $x^0, x^1 \in K$ tais que $f(x^0) = \inf(B)$ e $f(x^1) = \sup(B)$. Observamos que $\sup(B)$ é o máximo de f em K e $\inf(B)$ é o mínimo de f em K .*

Prova:

Ver Teorema 4.4.1, pag. 189 em [16].

Para $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, denotemos por $S_r(f)$, para algum $r \in \mathbb{R}$ aos conjuntos de nível de f dado por: $S_r(f) = \{x \in \mathbb{R}^n : f(x) \leq r\}$.

Seja $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria fechada, lembremos sua função de recessão $h'_\infty(d)$:

$$d \in \mathbb{R}^n \mapsto h'_\infty(d) = \lim_{\lambda \rightarrow \infty} \frac{h(x + \lambda d) - h(x)}{\lambda}, \quad \text{para } x \in \text{dom} h.$$

Proposição 2.7 *Para $f \in \overline{\text{Conv}} \mathbb{R}^n$, os seguintes itens são equivalentes:*

- i) *existe $r \in \mathbb{R}$ para o qual $S_r(f)$ é não vazio e compacto,*
- ii) *todos os conjuntos de nível de f são compactos,*
- iii) *$f'_\infty(d) > 0$ para todo $d \neq 0$ em \mathbb{R}^n .*

Prova:

Ver Proposição 3.2.5, pag. 180 em [8].

2.2 Condições Suficientes para a boa definição do Algoritmo de P.P. com quase-distâncias como núcleo

Hipótese 2.8 *Seja $D : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ uma quase-distância. Suponhamos que $D(\cdot, y)$ é estritamente convexa $\forall y \in \mathbb{R}^n$.*

Afirmção 2.9 *Seja $D : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ uma quase-distância que verifica a Hipótese 2.8. Então $\lim_{\|x\| \rightarrow \infty} D(x, y) = +\infty$ para cada $y \in \mathbb{R}^n$. Ou seja, $D(\cdot, y)$ é 0-coerciva.*

Prova:

Por contradição. Suponhamos que para $y \in \mathbb{R}^n$ fixo, $D(\cdot, y)$ não é 0-coerciva, então existe uma seqüência $\{x^k\} \subset \mathbb{R}^n$ tal que $\|x^k\| \rightarrow \infty$ e $D(x^k, y) \leq M$ para algum $M \in \mathbb{R}$. Isto quer dizer que o conjunto de nível

$$\Gamma(y, M) = \{x \in \mathbb{R}^n : D(x, y) \leq M\} \text{ é ilimitado.}$$

Pela Proposição 2.7, $\Gamma(y, 0) = \{x \in \mathbb{R}^n : D(x, y) \leq 0\}$ também o será, o que é uma contradição, pois $\Gamma(y, 0) = \{y\}$, já que D é uma quase-distância.

Portanto, $\lim_{\|x\| \rightarrow \infty} D(x, y) = +\infty$ para cada $y \in \mathbb{R}^n$.

□

Consideremos o problema (P) sob a Hipótese 2.4. Seja $\{x^k\} \subset \mathbb{R}^n$ uma seqüência gerada por:

$$\begin{aligned} x^0 &\in \mathbb{R}^n \\ x^{k+1} &= \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k N(x, x^k)\}, \quad \forall k \in \mathbb{N} \end{aligned}$$

onde o núcleo N verifica a Hipótese 2.8 e $\{\lambda_k\} \subset \mathbb{R}$ é tal que $\forall k \in \mathbb{N}$, $0 < \lambda_k \leq \tilde{\lambda}$ para algum real fixo $\tilde{\lambda} > 0$.

Afirmção 2.10 $\{x^k\} \subset \mathbb{R}^n$ está bem definida.

Prova

Vejam que para cada $k \in \mathbb{N}$ o problema (P_k) tem solução única.

Para $\bar{x} \in \mathbb{R}^n$ arbitrário. Da Hipótese 2.8 e a Afirmção 2.9 obtemos que $N(\cdot, \bar{x})$ é 0-coerciva. Usando além disto a Hipótese 2.4 temos que para qualquer $\lambda > 0$, $f(\cdot) + \lambda N(\cdot, \bar{x})$ também será 0-coerciva o que reduz o problema de minimizar a função contínua $\tilde{f}_{\bar{x}}(\cdot) = f(\cdot) + \lambda N(\cdot, \bar{x})$, a um conjunto compacto da forma $\Omega = \{x \in \text{dom}f : (f + \lambda N_{\bar{x}})(x) \leq r\}$ onde r é algum real tal que $r \geq f(\bar{x})$. Assim, pelo Teorema 2.6, temos garantida a existência de $\tilde{x} \in \text{dom}f$, solução do problema:

$$\begin{aligned} &\text{minimizar } \{f(x) + \lambda N(x, \bar{x})\} && (P_{\bar{x}}) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

Finalmente, da Hipótese 2.8 e do fato de f ser convexa obtemos a convexidade estrita de $\tilde{f}_{\bar{x}}$ a qual, acabamos de ver, tem mínimo. Portanto, o minimizador de $\tilde{f}_{\bar{x}}$ é único. Ou seja, a solução \tilde{x} do problema $(P_{\bar{x}})$ é **única**.

Fazendo $x^k = \bar{x}$ e $\lambda^k = \lambda$, para cada $k \in \mathbb{N}$, teremos $x^{k+1} = \tilde{x}$. Logo a afirmação estará provada.

□

A seguir ocupar-nos-emos dos principais núcleos usados no método de P.P., de seus algoritmos e dos respectivos resultados de convergência .

2.3 Núcleo quadrático

O algoritmo de ponto proximal com núcleo quadrático

$$x, y \in \mathbb{R}^n \rightarrow N(x, y) = \frac{1}{2} \|x - y\|^2$$

gera a partir de $x^0 \in \mathbb{R}^n$ dado, a seqüência $\{x^k\}$ mediante a regra:

$$\begin{aligned} x^{k+1} &= \operatorname{argmin}\{f(x) + \frac{\lambda_k}{2} \|x - x^k\|^2\}, \quad \forall k \in \mathbb{N} \\ &x \in \mathbb{R}^n \end{aligned}$$

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

Observemos que $x, y \in \mathbb{R}^n \rightarrow \frac{1}{2}\|x - y\|^2$ é uma quase-distância que verifica a Hipótese 2.8, portanto, considerando o problema (P) sob a Hipótese 2.4, pela Afirmção 2.10, a seqüência $\{x^k\}$ gerada pelo algoritmo de ponto proximal com núcleo quadrático está bem definida.

2.3.1 Noções relativas à convergência

Definição 2.11 *Seja H um espaço de Hilbert. $V : H \rightarrow H$ é um operador firmemente não expansivo se*

$$\|V(x) - V(y)\|^2 \leq \|x - y\|^2 - \|(I - V)(x) - (I - V)(y)\|^2, \quad \forall x, y \in H$$

onde $I : H \rightarrow H$ é o operador identidade.

Definição 2.12 *Seja $\{y^k\}$ uma seqüência em \mathbb{R}^n e $U \subset \mathbb{R}^n$ um conjunto. Diz-se que $\{y^k\}$ é Fejér convergente a U com respeito à distância Euclídeana se:*

$$\|y^{k+1} - u\| \leq \|y^k - u\|, \quad \forall k \in \mathbb{N}, \forall u \in U.$$

Definição 2.13 *Seja $T : \mathbb{R}^n \rightarrow \mathcal{P}(\mathbb{R}^n)$ um operador. T é monótono se:*

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

T é monótono maximal se além disso para todo operador monótono T' tal que $\forall x \in \mathbb{R}^n, T(x) \subset T'(x)$, tem-se $T(x) = T'(x)$.

Dado que o núcleo quadrático representa um caso particular das distâncias de Bregman (o núcleo que veremos na seguinte seção), a prova da convergência da seqüência $\{x^k\}$ a uma solução do problema (P) será omitida nesta. Fazendo um pouco de história, podemos dizer que os primeiros resultados de convergência do algoritmo de P.P. com núcleo quadrático foram obtidos por Martinet [17] no contexto dos operadores firmemente não expansivos em espaços de Hilbert e para o caso particular $\lambda_k = \frac{1}{c}$, $\forall k \in \mathbb{N}$, onde c é algum real positivo fixo.

Outra abordagem para provar a convergência deste algoritmo é mediante a noção de convergência Fejér, mais fraca do que a firme não expansividade. Esta foi realizada por Iusem [12] no âmbito dos operadores monótonos maximais.

2.4 Distâncias de Bregman

Foi em 1992 que Censor e Zenios [2] propuseram substituir o núcleo quadrático por uma quase-distância, conhecida como **distância de Bregman**.

Para $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, uma função convexa própria, dado $x \in \text{dom} f$, definamos o conjunto $\partial f(x) = \{s \in \mathbb{R}^n : f(y) \geq f(x) + \langle s, y - x \rangle \forall y \in \mathbb{R}^n\}$ chamado o subdiferencial de f em x .

Proposição 2.14 *Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Se f é diferenciável em x , então seu único subgradiente em x é seu gradiente $\nabla f(x)$.*

Prova

Ver Corolário 2.1.4, pag. 251 em [8].

Teorema 2.15 *Seja f uma função diferenciável num aberto $\Omega \subset \mathbb{R}^n$ e seja $C \subset \Omega$. Logo, f é estritamente convexa em C se e só se $f(x) > f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle, \forall (x, x^0) \in C \times C$ tal que $x \neq x^0$.*

Prova:

Ver Teorema 4.1.1, pag. 183 em [8]

Apesar de estar tratando o problema irrestrito, definiremos a seguir a distância de Bregman sobre o conjunto $\bar{S} \times S$, onde $S \subset \mathbb{R}^n$ é um conjunto aberto e convexo. Faremos isto simplesmente para evitar repeti-la no próximo capítulo.

Seja $S \subset \mathbb{R}^n$ aberto e convexo, (\bar{S} a aderência de S). Para $\varphi : \bar{S} \rightarrow \mathbb{R}$ uma função estritamente convexa em \bar{S} e diferenciável em S , usando a definição de subdiferencial, a Proposição 2.14 e o fato de φ ser diferenciável em S , temos:

$$\varphi(x) \geq \varphi(y) + \langle \nabla \varphi(y), x - y \rangle \quad \forall x \in \bar{S}, \forall y \in S.$$

Ou seja, φ induz uma quase-distância $D_\varphi : \bar{S} \times S \rightarrow \mathbb{R}$ dada pela regra:

$$D_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle x - y, \nabla\varphi(y) \rangle \quad (2.1)$$

Quando φ adquire certas características adicionais (e torna-se uma função de Bregman ψ) induzirá uma distância de Bregman.

Vejamos que na relação (2.1) temos realmente uma quase-distância.

Com efeito:

De antes: $D_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle x - y, \nabla\varphi(y) \rangle \geq 0, \quad \forall x \in \bar{S}, \forall y \in S.$

Para $x = y$, $D_\varphi(x, y) = \varphi(x) - \varphi(y) - \langle x - y, \nabla\varphi(y) \rangle = 0.$

Sejam $x, y \in S$. Se $D_\varphi(x, y) = 0$ então $x = y$, pois se $x \neq y$, como φ é estritamente convexa em S , usando o Teorema 2.15 teríamos $D_\varphi(x, y) > 0$ o que é uma contradição.

Assim D_φ é mesmo uma quase-distância em $\bar{S} \times S$.

2.4.1 Definição. A boa definição do Algoritmo de Ponto Proximal com distâncias de Bregman

Seja $S \subset \mathbb{R}^n$ um conjunto aberto e convexo. A função $\psi : \bar{S} \rightarrow \mathbb{R}$ é chamada uma função de Bregman se verifica:

i) ψ é continuamente diferenciável em S ,

ii) ψ é estritamente convexa e contínua em \bar{S} ,

iii) para todo $\alpha \in \mathbb{R}$: $L_1(y, \alpha) = \{x \in \bar{S} : D_\psi(x, y) \leq \alpha\}$,

$L_2(x, \alpha) = \{y \in S : D_\psi(x, y) \leq \alpha\}$ são limitados $\forall y \in S$ e $\forall x \in \bar{S}$, respectivamente;

iv) se $\{y^k\} \subset S$ converge a y^* então $D(y^*, y^k) \rightarrow 0$

v) se $\{x^k\} \subset \bar{S}$ e $\{y^k\} \subset S$ são seqüências tais que $\{x^k\}$ é limitada,

$\lim_{k \rightarrow \infty} y^k = y^*$ e $\lim_{k \rightarrow \infty} D_\psi(x^k, y^k) = 0$, então $\lim_{k \rightarrow \infty} x^k = y^*$.

Assim, $D_\psi : \bar{S} \times S \rightarrow \mathbb{R}$ definida como em (2.1):

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

onde ψ é uma função de Bregman, é chamada distância de Bregman induzida por ψ .

Como mencionamos antes para a função quadrática $\psi(x) = \|x\|^2$ (que é uma função de Bregman) temos $D_\psi(x, y) = \|x - y\|^2$ sua distância de Bregman correspondente.

Com efeito:

$$D_\psi(x, y) = \|x\|^2 - \|y\|^2 - \langle x - y, 2y \rangle = \|x\|^2 + \|y\|^2 - 2y^t x = \|x - y\|^2.$$

□

Para o caso que estamos estudando, no qual $S = \mathbb{R}^n$, basta assumir (i)-(iii) já que (iv) e (v) se verifica trivialmente. Então, o algoritmo de Ponto Proximal com distâncias de Bregman $D_\psi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, gerará a seqüência $\{x^k\}$ da seguinte maneira:

$$\begin{aligned} x^0 &\in \mathbb{R}^n \\ x^{k+1} &= \operatorname{argmin}\{f(x) + \lambda_k D_\psi(x, x^k)\} \quad \forall k \in \mathbb{N} \\ &x \in \mathbb{R}^n \end{aligned}$$

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

Afirmção 2.16 *A distância de Bregman D_ψ verifica a Hipótese 2.8. Logo, considerando o problema (P) sob a Hipótese 2.4 e pela Afirmção 2.10, a definição da seqüência $\{x^k\}$ gerada pelo algoritmo de P.P. com distâncias de Bregman (P.P.D.B.) está garantida.*

Prova:

Por (ii) da definição de função de Bregman, ψ é estritamente convexa em $S = \mathbb{R}^n$, então para $x, y \in \mathbb{R}^n$ e $\lambda \in (0, 1)$ temos para todo $z \in \mathbb{R}^n$:

$$\begin{aligned}
D_\psi(\alpha x + (1 - \alpha)y, z) &= \psi(\alpha x + (1 - \alpha)y) - \psi(z) \\
&\quad - \langle \alpha x + (1 - \alpha)y - z, \nabla\psi(z) \rangle \quad \text{para todo } z \in \mathbb{R}^n \\
D_\psi(\alpha x + (1 - \alpha)y, z) &< \alpha\psi(x) + (1 - \alpha)\psi(y) - \alpha\psi(z) - \alpha\langle x - z, \nabla\psi(z) \rangle \\
&\quad - (1 - \alpha)\psi(z) - (1 - \alpha)\langle y - z, \nabla\psi(z) \rangle \\
D_\psi(\alpha x + (1 - \alpha)y, z) &< \alpha D_\psi(x, y) + (1 - \alpha)D_\psi(y, z)
\end{aligned}$$

A última passagem é obtida reagrupando os termos da direita. Logo $D_\psi(\cdot, z)$ é estritamente convexa, $\forall z \in \mathbb{R}^n$.

□

2.4.2 Resultados preliminares

Lema 2.17 *Seja $\psi : \bar{S} \rightarrow \mathbb{R}$ uma função de Bregman. Para quaisquer três pontos $a, b \in S$ e $c \in \bar{S}$ a seguinte identidade se verifica:*

$$D_\psi(c, a) + D_\psi(a, b) - D_\psi(c, b) = \langle \nabla\psi(b) - \nabla\psi(a), c - a \rangle.$$

Prova

Da definição de D_ψ , temos:

$$\begin{aligned}
D_\psi(c, a) &= \psi(c) - \psi(a) - \langle c - a, \nabla\psi(a) \rangle \\
D_\psi(a, b) &= \psi(a) - \psi(b) - \langle a - b, \nabla\psi(b) \rangle \\
D_\psi(c, b) &= \psi(c) - \psi(b) - \langle c - b, \nabla\psi(b) \rangle.
\end{aligned}$$

Logo

$$\begin{aligned}
D_\psi(c, a) + D_\psi(a, b) - D_\psi(c, b) &= -\langle c - a, \nabla\psi(a) \rangle - \langle a - b, \nabla\psi(b) \rangle \\
&\quad + \langle c - b, \nabla\psi(b) \rangle \\
&= -\langle c - a, \nabla\psi(a) \rangle - \langle c - a, \nabla\psi(b) \rangle \\
&= \langle \nabla\psi(b) - \nabla\psi(a), c - a \rangle
\end{aligned}$$

□

Proposição 2.18 *Seja $\psi : \bar{S} \rightarrow \mathbb{R}$ uma função de Bregman. Então:*

$$\nabla_x D_\psi(x, y) = \nabla\psi(x) - \nabla\psi(y), \quad \forall x, y \in S$$

Prova:

Direta, da definição de D_ψ .

□

Proposição 2.19 *Para $f \in \text{Conv}\mathbb{R}^n$, tem-se:*

$$x \in \mathbb{R}^n \text{ minimiza } f \text{ se e só se } 0 \in \partial f(x).$$

Prova:

Consequência do Corolário 1.4.4, pag. 48 dado em [9].

Teorema 2.20 *Seja $f \in \text{Conv}\mathbb{R}^n$. Logo*

$$\partial f(x) \neq \emptyset \text{ para todo } x \in \text{ri} \text{dom} f$$

onde $\text{ri} \text{dom} f$ denota o interior relativo do domínio de f .

Prova:

Ver Teorema 1.4.2, pag. 47 em [9].

Corolário 2.21 *Sejam f_1, f_2 , duas funções em $\text{Conv}\mathbb{R}^n$. Assumamos que $\text{ri} \text{dom} f_1 \cap \text{ri} \text{dom} f_2 \neq \emptyset$. Logo*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \text{ para todo } x \in \text{dom} f_1 \cap \text{dom} f_2.$$

Prova:

Ver Corolário 3.1.2, pag. 114 em [9].

2.4.3 Convergência

Teorema 2.22 *Consideremos o problema (P) sob a Hipótese 2.4, então a seqüência $\{x^k\}$ gerada pelo algoritmo P.P.D.B. converge a uma solução do problema (P).*

Prova:

- a) $\{x^k\}$ está bem definida (ver Afirmação 2.16).
- b) $\{x^k\}$ é Fejér convergente ao conjunto de soluções do problema (P) com respeito à distância de Bregman D_ψ .

Seja \bar{x} uma solução do problema (P), arbitrária. Usando o Lema 2.17, com $S = \mathbb{R}^n$, para $a = x^{k+1}$, $b = x^k$, $c = \bar{x}$, pontos em \mathbb{R}^n com $k \in \mathbb{N}$ também arbitrário, temos:

$$D_\psi(\bar{x}, x^{k+1}) + D_\psi(x^{k+1}, x^k) - D_\psi(\bar{x}, x^k) = \langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), \bar{x} - x^{k+1} \rangle \quad (2.2)$$

Da Proposição 2.19: $0 \in \partial[f + \lambda_k D_\psi(\cdot, x^k)](x^{k+1})$, pelo Corolário 2.21, o Teorema 2.20 e a Proposição 2.18:

$$0 \in \partial f(x^{k+1}) + \underbrace{\lambda_k \nabla D_\psi(x^{k+1}, x^k)}_{\lambda_k (\nabla\psi(x^{k+1}) - \nabla\psi(x^k))}$$

ou seja:

$$y^k \in \partial f(x^{k+1}) \quad \text{onde} \quad y^k = \lambda_k (\nabla\psi(x^k) - \nabla\psi(x^{k+1})) \quad (2.3)$$

Da definição de subdiferencial, temos:

$$\langle y^k, \bar{x} - x^{k+1} \rangle \leq f(\bar{x}) - f(x^{k+1}) \quad (2.4)$$

Usando (2.3) e (2.4) em (2.2):

$$\begin{aligned} D_\psi(\bar{x}, x^{k+1}) + D_\psi(x^{k+1}, x^k) - D_\psi(\bar{x}, x^k) &= \langle \nabla\psi(x^k) - \nabla\psi(x^{k+1}), x^{k+1} - \bar{x} \rangle \\ &= \langle \lambda_k^{-1} y^k, x^{k+1} - \bar{x} \rangle \\ &= \frac{1}{\lambda_k} \langle y^k, x^{k+1} - \bar{x} \rangle \\ &\geq \frac{1}{\lambda_k} (f(x^{k+1}) - f(\bar{x})) \end{aligned} \quad (2.5)$$

Além disso, $f(\bar{x}) \leq f(x^{k+1})$ pois \bar{x} é solução do problema (P), logo:

$$D_\psi(\bar{x}, x^k) - D_\psi(x^{k+1}, x^k) - D_\psi(\bar{x}, x^{k+1}) \geq 0 \quad (2.6)$$

e finalmente,

$$D_\psi(\bar{x}, x^{k+1}) \leq D_\psi(\bar{x}, x^k), \quad \forall k \in \mathbb{N}.$$

c) $\{x^k\}$ é limitada. Além disso, se $\lim_{k \rightarrow \infty} x^{j_k} = \bar{x}$ para $\{x^{j_k}\}$ uma subseqüência de $\{x^k\}$ então $\lim_{k \rightarrow \infty} x^{j_k+1} = \bar{x}$.

Para \bar{x} uma solução arbitrária do problema (P) e $k \in \mathbb{N}$ também arbitrário, temos por b que $\{D_\psi(\bar{x}, x^k)\}$ é decrescente e não negativa (pois D_ψ é uma quase-distância). Logo

$$\exists A \in \mathbb{R} \text{ tal que } D_\psi(\bar{x}, x^k) \rightarrow A \text{ quando } k \rightarrow \infty \quad (2.7)$$

Tomando limite em (2.6), quando $k \rightarrow \infty$ obtemos:

$$A - \lim_{k \rightarrow \infty} D_\psi(x^{k+1}, x^k) - A \geq 0, \quad \text{onde } A = \lim_{k \rightarrow \infty} D_\psi(\bar{x}, x^k)$$

então

$$\lim_{k \rightarrow \infty} D_\psi(x^{k+1}, x^k) = 0 \quad (2.8)$$

Como $\{D_\psi(\bar{x}, x^k)\}$ é decrescente, de b:

$$D_\psi(\bar{x}, x^k) \leq D_\psi(\bar{x}, x^0) = \delta \quad \forall k \in \mathbb{N} \quad (2.9)$$

onde x^0 é o ponto inicial dado.

De (2.9) temos que $\{x^k\} \subset L_1(\bar{x}, \delta) = \{y \in \mathbb{R}^n : D_\psi(\bar{x}, y) \leq \delta\}$ como $L_1(\bar{x}, \delta)$ é limitado. Logo $\{x^k\}$ também o será.

Dado que toda seqüência limitada possui uma subseqüência convergente (tem pontos limites), existirá $\{x^{j_k}\} \subset \{x^k\}$ tal que $x^{j_k} \rightarrow \hat{x}$, para algum $\hat{x} \in \mathbb{R}$.

Usando (2.8) em (v) da definição de ψ , com a subseqüência $\{x^{j_k+1}\}$ como a **seqüência limitada** e $\{x^{j_k}\}$ como a **seqüência convergente**, concluímos que $\lim_{k \rightarrow \infty} x^{j_k+1} = \hat{x}$.

d) *Os pontos limites da seqüência $\{x^k\}$ são soluções do problema (P).*

Seja \hat{x} um ponto limite de $\{x^k\}$, ou seja, $\lim_{k \rightarrow \infty} x^{j_k} = \hat{x}$. Para \bar{x} solução do problema (P), de (2.5) para $k = j_k$ temos:

$$0 \leq \frac{1}{\lambda_{j_k}}(f(x^{j_k+1}) - f(\bar{x})) \leq D_\psi(\bar{x}, x^{j_k}) - D_\psi(x^{j_k+1}, x^{j_k}) - D_\psi(\bar{x}, x^{j_k+1}).$$

Usando o limitante superior de λ_k ($\tilde{\lambda}$) e tomando limite quando $k \rightarrow \infty$:

$$0 \leq \frac{1}{\tilde{\lambda}}(f(\hat{x}) - f(\bar{x})) \leq D_\psi(\bar{x}, \hat{x}) - 0 - D_\psi(\bar{x}, \hat{x}) = 0.$$

Logo $f(\hat{x}) = f(\bar{x})$ onde \bar{x} é a solução do problema (P) e também $f(\hat{x}) \leq f(x)$, $\forall x \in \mathbb{R}^n$.

e) *$\{x^k\}$ tem um só ponto limite.*

Seja \hat{x} o ponto limite da seqüência, $\{x^k\}$, então $\lim_{k \rightarrow \infty} x^{j_k} = \hat{x}$, por (iv) da definição de ψ , temos:

$$D_\psi(\hat{x}, x^{j_k}) \rightarrow 0 \tag{2.10}$$

De (2.7) e (2.10) temos que

$$D_\psi(\hat{x}, x^k) \rightarrow 0 \text{ quando } k \rightarrow \infty. \tag{2.11}$$

Suponhamos que existe outro ponto limite de $\{x^k\}$ além de \hat{x} e denotemos este de z , então existirá $\{x^{l_k}\} \subset \{x^k\}$ tal que $\lim_{k \rightarrow \infty} x^{l_k} = z$.

Usando (v) da definição de ψ com $\{y^k = z\}$ como a seqüência limitada, $\{x^{j_k}\}$ como a seqüência convergente e tendo em consideração (2.11) ($D_\psi(z, x^{j_k}) \rightarrow 0$) temos $\lim_{k \rightarrow \infty} z = \hat{x}$, ou seja, $z = \hat{x}$ o que é uma contradição.

Finalmente, como $\{x^k\}$ é limitada (foi provado no item c) e tem um único ponto limite, ela será convergente e convergirá (por d) a uma solução do problema (P).

□

2.5 Núcleos estritamente convexos $\phi(\cdot)$

Hipótese 2.23 *Seja $\phi : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função estritamente convexa que satisfaz o seguinte:*

- a) *tem o vetor zero como único subgradiente na origem e*
- b) $\phi(0) = 0$.

O algoritmo de Ponto Proximal com funções estritamente convexas $\phi(\cdot)$ como núcleo gera a seqüência $\{x^k\} \subset \mathbb{R}^n$ da seguinte maneira:

$$\begin{aligned} x^0 &\in \mathbb{R}^n \\ x^{k+1} &= \operatorname{argmin}\{f(x) + \lambda_k \phi(x - x^k)\} \\ &x \in \mathbb{R}^n \end{aligned}$$

onde f é a função objetivo do problema (P) e $\{\lambda_k\}$ é uma seqüência limitada de números reais positivos.

2.5.1 A boa definição do algoritmo

Proposição 2.24 *Uma condição necessária e suficiente para que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ convexa, seja estritamente convexa num convexo $C \subset \mathbb{R}^n$, é que para todos $x, x' \in C$ com $x \neq x'$, $f(x') > f(x) + \langle s, x' - x \rangle$ para todo $s \in \partial f(x)$*

Prova:

Ver Proposição 6.1.3, pag 281 em [8].

Afirmção 2.25 *Dado $x, y \in \mathbb{R}^n$, a função $D(x, y) = \phi(x - y)$ é uma quase-distância.*

Prova:

- i) Como o vetor zero é o único subgradiente da função convexa ϕ em $x = 0$, então da definição de subgradiente

$$0(z - 0) + \phi(0) \leq \phi(z) \quad \forall z \in \mathbb{R}^n$$

Da Hipótese 2.23 b e fazendo $z = x - y$, para $x, y \in \mathbb{R}^n$ temos:

$$0 \leq \phi(x - y) = D(x - y) \quad \text{para } x, y \in \mathbb{R}^n.$$

- ii) Se $x = y$ então $D(x, y) = D(y, y) = \phi(0)$ e novamente pela Hipótese 2.23 b temos $D(x, y) = 0$.

Se $D(x, y) = 0$ (também $\phi(x - y) = 0$) logo $x - y = 0$, pois se $x - y \neq 0$, da Proposição 2.24 obtemos: $\phi(x - y) > \phi(0) + \langle 0, x - y - 0 \rangle = 0$, o que é uma contradição. Assim $x = y$.

□

Afirmção 2.26 Para $z \in \mathbb{R}^n$, $D(\cdot, z) = \phi(\cdot, z)$ é estritamente convexa. Portanto, considerando o problema (P) sob a Hipótese 2.4, da Afirmção 2.10 a seqüência $\{x^k\}$ gerada pelo algoritmo de P.P. com funções estritamente convexas $\phi(\cdot)$ está bem definida.

Prova:

Para quaisquer $x, y \in \mathbb{R}^n$ e $\alpha \in \langle 0, 1 \rangle$

$$D(\alpha x + (1 - \alpha)y, z) = \phi(\alpha z + (1 - \alpha)y - z) \quad \text{para } z \in \mathbb{R}^n \text{ fixo.}$$

Fazendo $x = a + z$ e $y = b + z$, para $a, b \in \mathbb{R}^n$ quaisquer temos:

$$\begin{aligned} D(\alpha x + (1 - \alpha)y, z) &= \phi(\alpha(a + z) + (1 - \alpha)(b + z) - z) \\ &= \phi(\alpha a + (1 - \alpha)b) < \alpha\phi(a) + (1 - \alpha)\phi(b) \end{aligned}$$

dado que ϕ é estritamente convexa. Voltando aos vetores iniciais:

$$D(\alpha x + (1 - \alpha)y, z) < \alpha\phi(x - z) + (1 - \alpha)\phi(y - z).$$

Finalmente,

$$D(\alpha x + (1 - \alpha)y, z) < \alpha D(x, z) + (1 - \alpha)D(y, z)$$

para todo $x, y \in \mathbb{R}^n$, $\forall \alpha \in (0, 1)$, $z \in \mathbb{R}^n$.

□

Recentemente Humes-Silva [11] apresentaram uma prova de convergência a respeito usando a noção de separadores. Diversamente dos casos anteriores, não se consegue provar que a seqüência gerada pelo algoritmo de P. P. correspondente seja limitada.

Teorema 2.27 *Consideremos o problema (P) sob a Hipótese 2.4. Logo, todo ponto de acumulação da seqüência gerada pelo algoritmo de P.P. com núcleo estritamente convexo será solução do problema (P).*

Prova:

Ver Teorema 2 em [11].

□

Capítulo 3

Estudo do Método de Ponto Proximal para o problema restrito

Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria fechada e $S \subset \mathbb{R}^n$ um conjunto não vazio, aberto e convexo.

Hipótese 3.1 *Assumamos que $\text{ri} \text{dom} f \cap S \neq \emptyset$.*

O problema que estamos interessados em resolver agora é o seguinte:

$$\begin{aligned} & \text{minimizar } f(x) && (P1) \\ & \text{s.a. } x \in \bar{S} \end{aligned}$$

onde \bar{S} é a aderência de S , a qual é convexa e fechada.

Hipótese 3.2 *Suponhamos que o problema (P1) tem solução. Ou seja, existe $x^* \in \bar{S}$ tal que $f(x^*) \leq f(x)$, para todo $x \in \bar{S}$.*

Para resolver este problema, mais geral do que o problema (P), também podemos usar o método de Ponto Proximal. Os núcleos que usaremos serão umas distâncias de Bregman particulares e as chamadas φ -divergências.

3.1 Distâncias de Bregman induzidas por funções de Bregman coercivas na fronteira

Resultados prévios:

Definição 3.3 Uma função de Bregman $\psi : \bar{S} \rightarrow \mathbb{R}$ se diz coerciva na fronteira se: Para $\{y^k\} \subset S$ tal que $y^k \rightarrow y \in \partial S$ temos que para todo $x \in S$,

$$\lim_{k \rightarrow \infty} \nabla \psi(y^k)^t(x - y^k) = -\infty \quad (3.1)$$

onde ∂S é a fronteira de S .

Observação 3.4 Seja $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria fechada e $\psi : \bar{S} \rightarrow \mathbb{R}$, uma função de Bregman coerciva na fronteira. Então para todo $x \in \text{dom} f \cap \partial S$:

$$\partial(f + \lambda\psi)(x) = \emptyset. \quad (3.2)$$

Prova:

Por contradição. Suponhamos que existe $\gamma \in \mathbb{R}^n$ tal que

$$\gamma \in \partial(f + \lambda\psi)(x), \quad \text{para } x \in \text{dom} f \cap \partial S \quad (3.3)$$

Tomando $z \in S$ e $\{x^j\} \subset S$ tal que $x^j \rightarrow x$, definimos

$$y^j = (1 - \varepsilon_j)x^j + \varepsilon_j z, \quad \text{onde } 0 \leq \varepsilon_j \leq 1 \quad \forall j \in \mathbb{N} \text{ e } \lim_{j \rightarrow \infty} \varepsilon_j = 0. \quad (3.4)$$

Logo, pela convexidade de S temos que $y^j \in S$ e $\lim_{j \rightarrow \infty} y^j = x$. Assim, usando (3.3) e (3.4):

$$\begin{aligned} \varepsilon_j \gamma^t(z - x^j) &= \gamma^t(\varepsilon_j z - \varepsilon_j x^j) = \gamma^t(y^j - x^j) \\ &\leq f(y^j) - f(x^j) + \lambda(\psi(y^j) - \psi(x^j)) \end{aligned}$$

Da convexidade de ψ obtemos

$$\varepsilon_j \gamma^t(z - x^j) \leq f(y^j) - f(x^j) + \lambda \nabla \psi(y^j)^t(y^j - x^j)$$

Usando (3.4) e a convexidade de f :

$$\begin{aligned} \varepsilon_j \gamma^t(z - x^j) &\leq (1 - \varepsilon_j)f(x^j) + \varepsilon_j f(z) - f(x^j) + \lambda \nabla \psi(y^j)^t (y^j - x^j) \\ &\leq \varepsilon_j (f(z) - f(x^j)) + \lambda \frac{\varepsilon_j}{1 - \varepsilon_j} \nabla \psi(y^j)^t (z - y^j) \end{aligned}$$

Finalmente, essa desigualdade poderá-se exprimir

$$\left(\frac{1 - \varepsilon_j}{\lambda} \right) [f(x^j) - f(z) + \gamma^t(z - x^j)] \leq \nabla \psi(y^j)^t (z - y^j)$$

Fazendo tender $j \rightarrow \infty$, como ψ é coerciva na fronteira, temos $\frac{1}{\lambda}[f(x) - f(z) + \gamma^t(z - x)] = -\infty$ o que não é uma verdade, pois este é um valor real. Portanto, (3.2) verifica-se.

□

Retomando a definição de distância de Bregman dada na subseção (2.4.1) o algoritmo de P.P. com distâncias de Bregman $D_\psi : \bar{S} \times S \rightarrow \mathbb{R}$ induzidas por funções de Bregman ψ -coercivas na fronteira, gerará a seqüência $\{x^k\} \subset \mathbb{R}^n$ da seguinte forma:

$$x^0 \in S \cap \text{dom} f \tag{3.5}$$

$$x^{k+1} = \operatorname{argmin}\{f(x) + \lambda_k D_\psi(x, x^k)\} \tag{3.6}$$

$$x \in \bar{S}$$

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

3.1.1 A boa definição

Como se pode observar em (3.6) agora devemos garantir um detalhe adicional. Devemos provar que $x^{k+1} \in S, \forall k \in \mathbb{N}$ (já que se $x^{k+1} \in \bar{S}$, não poderíamos avaliar no passo seguinte D_ψ naquele ponto, pois ela está definida sobre $\bar{S} \times S$). A coercividade na fronteira nos permitirá garantir isto.

Consequentemente na minimização da relação (3.6) bem se pode escrever S em lugar \bar{S} .

Consideremos como de costume o problema (P1) sob as hipóteses 3.1 e 3.2.

Afirmção 3.5 A seqüência $\{x^k\} \subset \mathbb{R}^n$ gerada por (3.5) e (3.6) está bem definida, ou seja, para cada $k \in \mathbb{N}$ o problema de minimizar $\{f(x) + \lambda_k D_\psi(x, x^k)\}$ tem solução única e $x^{k+1} \in S \quad \forall k \in \mathbb{N}$.

Prova:

Para $y \in S \cap \text{dom}f$ arbitrário e $\lambda > 0$, escolhamos $M \in \mathbb{R}$ tal que $M \geq f(y)$ e definamos $\Omega = \{x \in \bar{S} \cap \overline{\text{dom}f} : f(x) + \lambda D_\psi(x, y) \leq M\}$

O conjunto Ω é não vazio já que $y \in \Omega$. (D_ψ é uma quase-distância em $\bar{S} \times S$).

Ω é limitado. Seja $x \in \Omega$ arbitrário, então $f_y(x) = f(x) + \lambda D_\psi(x, y) \leq M$, logo

$$D_\psi(x, y) \leq (M - f(x))/\lambda. \quad (3.7)$$

Como se considera a Hipótese 3.2, f será limitada inferiormente em \bar{S} , logo em $\bar{S} \cap \overline{\text{dom}f}$. Seja $\beta \in \mathbb{R}$ aquele limitante e usando-o em (3.7) teremos:

$$D_\psi(x, y) \leq (M - \beta)/\lambda.$$

Ou seja, $x \in L_1(y, (M - \beta)/\lambda)$. Como aquel conjunto de nível é limitado (ver a definição na subseção 2.4.1), conclui-se que Ω também o será.

Ω é fechado. Pois f é contínua em \bar{S} e D_ψ também é contínua com respeito à primeira variável dado que ψ é contínua em \bar{S} (ver subseção. 2.4.1(ii)).

Portanto, o problema de minimizar $\{f(x) + \lambda D_\psi(x, x^k)\}$ pode se reduzir a um compacto, logo existirá $\bar{x} \in \bar{S}$ tal que $f_y(\bar{x}) \leq f_y(x) \quad \forall x \in \bar{S}$. Além disso, $f_y(\cdot)$ é estritamente convexa em \bar{S} e para todo $y \in S$ então $\bar{x} \in \bar{S}$, o mínimo de $f_y(\cdot)$, é único.

Provemos que $\bar{x} \in S$:

O problema dado em (3.6) é equivalente ao problema irrestrito

$$\begin{aligned} &\text{minimizar } f_y(x) + I_{\bar{S}}(x), \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

onde $I_{\bar{S}}$ denota a função indicador de f sobre \bar{S} .

Usando agora a Proposição 2.19, o Corolário 2.21 e a Hipótese 3.1 temos que: $0 \in \partial(f + \lambda D_{\psi,y})(\bar{x}) + I_{\bar{S}}(\bar{x})$, ou seja,

$$\lambda \nabla \psi(y) - v \in \partial(f + \lambda \psi)(\bar{x}) \quad (3.8)$$

onde $v \in \partial I_{\bar{S}}(\bar{x})$.

Suponhamos que $\bar{x} \in \partial S$, então de (3.8) e da Observação 3.4 temos que $\lambda \nabla \psi(y) - v$ não existirá, o que é uma contradição pois ψ é diferenciável em S e $\partial I_{\bar{S}}(x) \neq \emptyset$ para $x \in \bar{S}$. Logo $\bar{x} \in S$.

Fazendo $x^k = y$ e $\lambda_k = \lambda$, para cada $k \in \mathbb{N}$, teremos $x^{k+1} = \bar{x}$. Logo a afirmação estará provada.

□

3.1.2 Convergência

Teorema 3.6 *Consideremos o problema (P1) sob as hipóteses (3.1) e (3.2), logo a seqüência $\{x^k\} \subset S$ gerada pelo algoritmo de P.P com distâncias de Bregman induzidas por funções de Bregman coercivas na fronteira converge a uma solução do problema (P1).*

Prova:

A boa definição da seqüência $\{x^k\}$ está garantida pela Afirmação 3.5.

As provas dos itens *b, c, d, e* são feitas de maneira similar as realizadas no Teorema 2.22, usando o Lema 2.17 e a Proposição 2.18, além dos itens *iv* e *v* da definição de função de Bregman ψ . Portanto, a convergência está garantida.

□

Um caso relevante de S é: $S = \mathbb{R}_{++}^n = \{x \in \mathbb{R}^n : x_i > 0, \forall i = 1, \dots, n\}$, o octante positivo de \mathbb{R}^n , no qual trabalharemos o resto do capítulo.

Assim

$$\bar{S} = \mathbb{R}_+^n = \{x \in \mathbb{R}^n : x_i \geq 0, \forall i = 1, \dots, n\}.$$

O problema correspondente a resolver será :

$$\begin{aligned} & \text{minimizar } f(x) && (P2) \\ & \text{s.a. } x \in \mathbb{R}_+^n \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é uma função convexa própria e fechada.

Hipótese 3.7 *Assumamos que $\text{ri dom } f \cap \mathbb{R}_{++}^n \neq \emptyset$.*

Este problema pode ser resolvido, como acabamos de ver, mediante a aplicação do Método de P.P. com distâncias de Bregman induzidas por funções de Bregman coercivas na fronteira, mas também pode ser abordado substituindo no método aquela distância de Bregman por uma certa φ -divergência.

3.2 φ -divergências

Introduzidas conjuntamente com as distâncias de Bregman em Teboulle [23], seus resultados de convergência foram obtidos por Iusem, Svaiter e Teboulle [13].

Enfatizamos que nesta seção $S = \mathbb{R}_{++}^n$, logo $\bar{S} = \mathbb{R}_+^n$.

Definição 3.8 *Seja $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ (ou $\varphi : \mathbb{R}_{++} \rightarrow \mathbb{R}$), uma função estritamente convexa e de classe C^2 que satisfaz:*

- i) $\varphi(1) = 0$
- ii) $\varphi'(1) = 0$
- iii) $\varphi''(1) > 0$.

A função $d\varphi : \mathbb{R}_+^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ (ou $d\varphi : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}$) definida por:

$$d\varphi(x, y) = \sum_{i=1}^n y_i \varphi(x_i/y_i)$$

e induzida por φ , é chamada φ -divergência.

Exemplo. Para $\varphi(t) = t \log t - t + 1$, uma função estritamente convexa e de classe C^2 em \mathbb{R}_+ , que satisfaz (i)-(iii) obtemos:

$$d_\varphi(x, y) = \sum_{j=1}^n x_j \log(x_j/y_j) + y_j - x_j$$

a φ -divergência de Kullback-Leibler.

Resultados Preliminares

Proposição 3.9 *Seja $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ uma função que satisfaz as condições acima, logo:*

- i) d_φ é uma quase-distância em $\mathbb{R}_+^n \times \mathbb{R}_{++}^n$
- ii) $L_1(y, \alpha) = \{x \in \mathbb{R}_+^n : d_\varphi(x, y) \leq \alpha\}$ são limitados $\forall y \in \mathbb{R}_{++}^n$ e todo $\alpha > 0$

Prova:

Similar à prova da Proposição 2.1 dada em [13], considerando $x \in \mathbb{R}_+^n$ em lugar de $x \in \mathbb{R}_{++}^n$.

□

Hipótese 3.10 *Seja $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}$ uma função dada pela Definição 3.8. Suponhamos que $\lim_{t \rightarrow 0} \varphi'(t) = -\infty$.*

Se esta hipótese verifica-se, para $\{x^k\} \subset \mathbb{R}_{++}^n$ tal que $x^k \rightarrow x \in \partial(\mathbb{R}_{++}^n)$ e $y \in \mathbb{R}_{++}^n$ temos

$$\lim_{k \rightarrow \infty} \nabla d_{\varphi, y}(x^k)^t (x - x^k) = -\infty \text{ para todo } x \in \mathbb{R}_{++}^n.$$

Observação 3.11 *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria fechada e d_φ uma φ -divergência induzida por φ que verifica a Hipótese 3.10, então*

$$\partial(f + d_{\varphi,y})(x) \text{ para todo } x \in \text{dom} f \cap \partial(\mathbb{R}_{++}^n).$$

Prova:

Similar à feita na Observação 3.4, usando na prova $d_{\varphi,y}$ em lugar da função de Bregman ψ .

□

Seja d_φ , uma φ -divergência induzida por φ dada pela Definição 3.8. Suponhamos que φ verifica a Hipótese 3.10. O algoritmo de ponto proximal com φ -divergências para o problema (P2) gerará a seqüência $\{x^k\}$ da seguinte maneira:

$$x^0 \in \mathbb{R}_{++}^n \cap \text{dom} f \tag{3.9}$$

$$x^{k+1} = \operatorname{argmin}\{f(x) + \lambda_k d_\varphi(x, x^k)\} \tag{3.10}$$

$$x \in \mathbb{R}_+^n$$

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

3.2.1 A boa definição

Consideremos o problema (P2) sob as hipóteses (3.2), com $S = \mathbb{R}_{++}^n$ e (3.7).

Afirmção 3.12 $\{x^k\} \subset \mathbb{R}^n$ está bem definida e para todo $k \in \mathbb{N}$, $x^k \in S = \mathbb{R}_{++}^n$.

Prova:

Para $y \in \mathbb{R}_{++}^n \cap \text{dom} f$ e $\lambda > 0$, escolhamos $M > 0$ tal que $M \geq f(y)$ e definamos

$$\Omega = \{x \in \mathbb{R}_+^n \cap \overline{\text{dom} f} : f(x) + \lambda d_\varphi(x, y) \leq M\}$$

O conjunto Ω é não vazio já que $y \in \Omega$. (d_φ , é uma quase-distância em $\mathbb{R}_+^n \times \mathbb{R}_{++}^n$).

Ω é limitado. Seja $x \in \Omega$ arbitrário, então $f_y(x) = f(x) + \lambda d\varphi(x, y) \leq M$, logo

$$d\varphi(x, y) \leq (M - f(x))/\lambda. \quad (3.11)$$

Como se considera a Hipótese 3.2, f será limitada inferiormente em \mathbb{R}_+^n , logo em $\mathbb{R}_+^n \cap \overline{\text{dom}f}$. Seja $\beta \in \mathbb{R}$ aquele limitante e usando-o em (3.11) teremos:

$$d\varphi(x, y) \leq (M - \beta)/\lambda.$$

Ou seja, $x \in L_1(y, (M - \beta)/\lambda)$, logo pela Proposição 3.9 (ii), Ω é limitado.

Ω é fechado. Pois f_y é contínua em $\mathbb{R}_+^n \cap \overline{\text{dom}f}$.

Portanto, o problema de minimizar $x \in \mathbb{R}_+^n \{f(x) + \lambda d\varphi(x, y)\}$ pode se reduzir a um compacto, logo existirá $\bar{x} \in \mathbb{R}_+^n$ tal que $f_y(\bar{x}) \leq f_y(x) \quad \forall x \in \mathbb{R}_+^n$. Além disso, $f_y(\cdot)$ é estritamente convexa em \mathbb{R}_+^n , então $\bar{x} \in \mathbb{R}_+^n$, é o único minimizador de $f_y(\cdot)$.

Provemos que $\bar{x} \in \mathbb{R}_{++}^n$

O problema dado em (3.10) é equivalente ao problema irrestrito

$$\begin{aligned} &\text{minimizar } f_y(x) + I_{\mathbb{R}_+^n}(x) \\ &\text{s.a. } x \in \mathbb{R}^n \end{aligned}$$

usando agora a Proposição 2.19, o Corolário 2.21 e a Hipótese 3.7 temos que: $0 \in \partial(f + \lambda d_{\varphi, y})(\bar{x}) + \partial I_{\mathbb{R}_+^n}(\bar{x})$, ou seja,

$$-v \in \partial(f + \lambda d_{\varphi, y})(\bar{x}) \quad (3.12)$$

onde $v \in \partial I_{\mathbb{R}_+^n}(\bar{x})$.

Suponhamos que $\bar{x} \in \partial \mathbb{R}_{++}^n$, logo pela Hipótese 3.10 temos $\partial(f + \lambda d_{\varphi, y})(\bar{x}) = \emptyset$, o que contradiz a relação (3.12) pois $\partial I_{\mathbb{R}_+^n}(x) \neq \emptyset$ para todo $x \in \mathbb{R}_+^n$. Assim $\bar{x} \in \mathbb{R}_{++}^n$. Fazendo $x^k = y$ e $\lambda_k = \lambda$, para cada $k \in \mathbb{N}$, teremos $x^{k+1} = \bar{x}$. Logo a afirmação estará provada.

□

3.2.2 Convergência

Para conseguir provar a convergência, consideram-se funções φ ainda mais particulares. Elas devem verificar a seguinte condição:

$$i) \quad \varphi'(t)/\varphi''(1) \leq \log(t) \text{ para todo } t > 0$$

Teorema 3.13 *Consideremos o problema (P2) sob as hipóteses (3.7), com $S = \mathbb{R}_{++}^n$ e (3.2) e a seqüência $\{x^k\}$ gerada pelo algoritmo de P.P. com φ -divergências, onde φ dada pela Definição 3.8 verifica a Hipótese 3.10. Se a condição (i) é satisfeita, então dita seqüência convergirá a uma solução do problema (P2).*

Prova:

Ver Corolário 4.4 em [13].

Capítulo 4

O Método Lagrangeano Aumentado

Consideremos o seguinte problema de otimização convexa:

$$\begin{aligned} & \text{minimizar } f(x) && (P3) \\ & \text{s.a. } g_i(x) \leq 0, \quad i = 1, \dots, m \\ & x \in \mathbb{R}^n \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g_i : \mathbb{R}^n \rightarrow \mathbb{R}$ para $i = 1, \dots, m$ são funções convexas.

Hipótese 4.1: *Suponhamos que o conjunto de soluções ótimas do problema (P3) é não vazio e limitado.*

Hipótese 4.2 (Condição de Qualificação de Slater). *Suponhamos que existe $\bar{x} \in \mathbb{R}^n$ tal que $g_i(\bar{x}) < 0 \quad \forall i = 1, \dots, m$.*

Definição 4.3 *Dado $b > 0$ (possivelmente $+\infty$), denotemos por \mathcal{P} a família de funções $P : \mathbb{R} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ ou $P : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ que satisfazem o seguinte:*

Se $b = +\infty$ (respectivamente $0 < b < +\infty$)

1.- $P(\cdot, u)$ *é continuamente diferenciável em \mathbb{R} , estritamente convexa em $[-b, +\infty)$ e constante em $(-\infty, -b]$*

2.- $P(0, u) = 0$, $\frac{\partial P}{\partial t}(0, u) = u$

$$3.- \lim_{t \rightarrow -\infty} \frac{\partial P}{\partial t}(t, u) = 0 \text{ e}$$

$$4.- \lim_{t \rightarrow \infty} \frac{\partial P}{\partial t}(t, u) = \infty$$

para qualquer $u > 0$ e $\forall t \in \mathbb{R}$, (respectivamente para qualquer $u \geq 0$ e $\forall t \in \mathbb{R}$).

Por comodidade daqui em diante escreveremos $P'(\cdot, u)$ em lugar de $\frac{\partial P}{\partial t}(\cdot, u)$, para qualquer $u \geq 0$.

4.1 As funções Lagrangeano e Lagrangeano Aumentado

A função Lagrangeano associada ao problema (P3) é definida por

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \rightarrow l(x, \mu) = f(x) + \sum_{i=1}^m \mu_i g_i(x),$$

aparece na teoria da otimização ao tratar das condições necessárias e suficientes que uma solução do problema (P3) verifica. Como veremos mais adiante, estas condições (e a função lagrangeano clássico) serão parte fundamental na construção do método Lagrangeano Aumentado em estudo. De maneira similar definimos, associada ao problema (P3), L , a função Lagrangeano aumentado (com penalidades $P_i \in \mathcal{P}$) dada por:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m, \lambda > 0 \rightarrow L(x, \mu, \lambda) = f(x) + \lambda \sum_{i=1}^m P_i \left(\frac{g_i(x)}{\lambda}, \mu_i \right) \quad (4.1)$$

onde as m funções $P_i : \mathbb{R} \times \mathbb{R}_{++} \rightarrow \mathbb{R}$ (ou as m funções $P_i : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$) chamadas **penalidades** pertencerão à família \mathcal{P} definida antes.

A primeira função Lagrangeano aumentado introduzida tinha aplicação a problemas com restrições de igualdade (ver [10] e [19]). Rockafellar [22] generalizou a metodologia para o caso onde as restrições são de desigualdade. Esta primeira função L.A. é dada por:

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m, \lambda > 0 \rightarrow L(x, \mu, \lambda) = f(x) + \lambda \sum_{i=1}^m \left[\left(\max\left\{0, \mu_i + \frac{g_i(x)}{\lambda}\right\} \right)^2 - \mu_i^2 \right].$$

4.2 Resultados preliminares à descrição do método

Teorema 4.4 *Seja f uma função diferenciável num aberto $\Omega \subset \mathbb{R}^n$ e seja $C \subset \Omega$ convexo. Logo f é convexa (respectivamente estritamente convexa) em C se e só se ∇f é monótono (respectivamente estritamente monótono) em C .*

Prova:

Ver Teorema 4.1.4, pag 185 em [8].

Observação 4.5 *Seja $f: \mathbb{R} \rightarrow \mathbb{R}$ convexa (resp. estritamente convexa) e diferenciável. Para $\mu \geq 0$ (resp. $\mu > 0$), se $f'(0) = \mu$ e $\lim_{t \rightarrow -\infty} f'(t) = 0$, então $f'(t) \geq 0$ (resp. $f'(t) > 0$) para todo $t \in \mathbb{R}$.*

Prova:

a) Para f convexa. Dado que $f'(t) = \mu \geq 0$ para $t = 0$, analisaremos dois casos.

i) Se $t \geq 0$ então pelo Teorema 4.4 $f'(t) \geq f'(0)$. Assim, $f'(t) \geq \mu \geq 0$.

ii) Caso $t < 0$. Suponhamos que existe $\hat{t} < 0$ tal que $f'(\hat{t}) < 0$, então, para todo $t < \hat{t}$ temos $f'(t) \leq f'(\hat{t}) < 0$ pelo mesmo teorema. Fazendo tender $t \rightarrow -\infty$ obtemos

$$\lim_{t \rightarrow -\infty} f'(t) \leq f'(\hat{t}) < 0$$

o que contradiz a hipótese. Logo, para todo $t < 0$, $f'(t) \geq 0$. Portanto,

$$f'(t) \geq 0 \quad \forall t \in \mathbb{R}. \quad (4.2)$$

b) Para f estritamente convexa. Suponhamos que existe $\hat{t} \in \mathbb{R}$ tal que $f'(\hat{t}) = 0$. Então para todo $t < \hat{t}$, novamente pelo Teorema 4.4, temos $f'(t) < f'(\hat{t})$. Assim, $f'(t) < f'(\hat{t}) = 0$.

Pelo resultado anterior (caso f convexa) sabemos que $f'(t) \geq 0 \quad \forall t \in \mathbb{R}$. Logo, $0 \leq f'(t) < 0$, o que é uma contradição. Portanto, não existe $t \in \mathbb{R}$ tal que $f'(t) = 0$, ou seja, tiramos a igualdade em (4.2).

□

Proposição 4.6 *Seja $f \in \text{Conv}\mathbb{R}^n$ (resp. $\overline{\text{Conv}}\mathbb{R}^n$) e seja $g \in \text{Conv}\mathbb{R}$ (resp. $\overline{\text{Conv}}\mathbb{R}$) crescente. Suponhamos que existe $x^o \in \mathbb{R}^n$ tal que $f(x^o) \in \text{dom } g$ e que $g(+\infty) = +\infty$. Logo, a função composta $x \mapsto g(f(x))$ está em $\text{Conv}\mathbb{R}^n$ (resp. $\overline{\text{Conv}}\mathbb{R}^n$).*

Prova:

Ver Proposição 2.1.8, pag 160, em [8].

Quando f e g são convexas e tomam valores finitos em \mathbb{R}^n , a Proposição 4.6 se segue sem precisar assumir que $g(+\infty) = +\infty$.

Observação 4.7 *Para cada $u \geq 0$ e $\lambda > 0$ fixos. Considere $P \in \mathcal{P}$ e $g: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Então:*

$$i) \frac{\partial P}{\partial t}(t, u) = P'_u(t) \geq 0 \quad \forall t \in \mathbb{R}. \text{ Ou seja, } P_u \text{ é crescente.}$$

$$ii) P(g(\cdot)/\lambda, u) = P_u \circ g/\lambda \text{ é convexa.}$$

Além disso, se $P \in \mathcal{P}$ é tal que para qualquer $u > 0$, $P(\cdot, u)$ é estritamente convexa (o caso em que $b = +\infty$), então $P'_u(t) > 0, \forall t \in \mathbb{R}$. Ou seja, P_u será estritamente crescente.

Prova:

i) Como $P \in \mathcal{P}$, para $u \geq 0$ fixo $P(\cdot, u) = P_u(\cdot)$ é convexa e diferenciável em \mathbb{R} . Além disso P_u verifica as condições da Observação 4.5 então $P'_u(t) \geq 0 \forall t \in \mathbb{R}$. Logo, P_u será crescente.

ii) Dado que g é convexa em \mathbb{R}^n , para $\lambda > 0$ fixo, g/λ também o será. Como, para $u \geq 0$ fixo. P_u é convexa e crescente, temos pela Proposição 4.6 que a composta $P_u \circ g/\lambda$ também será convexa em \mathbb{R}^n .

A última afirmação obtém-se do caso alternativo na Observação 4.5.

□

Observação 4.8 Para $\mu \geq 0$ em \mathbb{R}^m e $\lambda > 0$ fixos, consideremos a função Lagrangeano aumentado com penalidades $P_i \in \mathcal{P}$ definida em (4.1). Então,

$$x \in \mathbb{R}^n \mapsto L(x, \mu, \lambda) = L_{\mu, \lambda}(x) = f(x) + \lambda \sum_{i=1}^m P_i\left(\frac{g_i(x)}{\lambda}, \mu_i\right)$$

é convexa.

Prova:

Para $\mu \geq 0$ em \mathbb{R}^m fixo, temos $\mu_i \geq 0 \forall i = 1, \dots, m$. Então como cada $P_i \in \mathcal{P}$ e g_i são convexas, pela Observação 4.7 (i), $P_{i\mu_i} \circ g_i/\lambda$ será convexa em \mathbb{R}^n , para $\lambda > 0$ fixo.

Finalmente, como $L_{\mu, \lambda}$ é uma combinação positiva de funções convexas, $L_{\mu, \lambda}$ também será convexa em \mathbb{R}^n . □

Teorema 4.9 Sejam f_1, \dots, f_m funções convexas de \mathbb{R}^n em \mathbb{R} . Definamos F por:

$$x \in \mathbb{R}^n \mapsto F(x) = (f_1(x), \dots, f_m(x)) \in \mathbb{R}^m.$$

Seja $g: \mathbb{R}^m \rightarrow \mathbb{R}$ convexa e crescente componente a componente, isto é,

$$y_i \geq z_i \text{ para } i = 1, \dots, m \Rightarrow g(y) \geq g(z).$$

Logo, para todo $x \in \mathbb{R}^n$,

$$\partial(g \circ F)(x) = \left\{ \sum_{i=1}^m \rho_i s^i : (\rho_1, \dots, \rho_m) \in \partial g(F(x)), s^i \in \partial f_i(x) \text{ para } i = 1, \dots, m \right\}.$$

Se além disso g é diferenciável em $F(x)$ a relação anterior pode-se exprimir por:

$$\partial(g \circ F)(x) = \sum_{i=1}^m \frac{\partial g}{\partial y_i}(F(x)) \partial f_i(x) \text{ onde } y_i = f_i(x)$$

Prova:

Ver Teorema 4.3.1, pag 264 e 265, em [8].

Observação 4.10 Consideremos as m penalidades $P_i \in \mathcal{P}$ da relação (4.1) e as m restrições g_i do problema (P_2) . Dados $\bar{x} \in \mathbb{R}^n$, e $\mu \geq 0$ então

$$P'_1(g_1(\bar{x})/\lambda, \mu_1), \dots, P'_m(g_m(\bar{x})/\lambda, \mu_m)$$

são não negativos para qualquer $\lambda > 0$ fixo.

Além disso, se as m penalidades $P_i \in \mathcal{P}$ são tais que, para qualquer $\mu > 0$, $P_i(\cdot, \mu)$ é estritamente convexa, então $P'_i(g_i(\bar{x})/\lambda, \mu_i) > 0 \quad \forall i = 1, \dots, m$ e para quaisquer $\bar{x} \in \mathbb{R}^n$ e $\lambda > 0$ fixos.

Prova:

Se $\mu \geq 0$ em \mathbb{R}^m temos que $\mu_i \geq 0 \quad \forall i = 1, \dots, m$. Consideremos as penalidades $P_i \in \mathcal{P}$ e associemos para cada $i = 1, \dots, m$ fixo, P_i com μ_i . Logo, pela Observação 4.7 (i), $P'_i(t, \mu_i) \geq 0$ para todo $t \in \mathbb{R}$.

Finalmente, considerando as m restrições g_i do problema (P_2) e avaliando cada $P'_i(\cdot, \mu_i)$ em $t_i = g_i(\bar{x})/\lambda$ (para $\bar{x} \in \mathbb{R}^n$ dado e $\lambda > 0$ fixo), obtemos:

$$P'_1(g_1(\bar{x})/\lambda, \mu_1) \geq 0, \dots, P'_m(g_m(\bar{x})/\lambda, \mu_m) \geq 0$$

A última afirmação obtém-se do caso alternativo na Observação 4.7.

□

Definição 4.11 Um ponto $\bar{x} \in \mathbb{R}^n$ é chamado um ponto de Karush-Kuhn-Tucker (K.K.T) se existe algum ponto $\bar{u} \in \mathbb{R}^m$ tal que se verificam

a) $0 \in \partial f(\bar{x}) + \sum_{i=1}^m \bar{u}_i \partial g_i(\bar{x}) = \partial l(\bar{x}, \bar{u})$

b) $\bar{u} \geq 0$

c) $g_i(\bar{x}) \leq 0 \quad \forall i = 1, \dots, m$

d) $\bar{u}_i g_i(\bar{x}) = 0 \quad \forall i = 1, \dots, m.$

Se se verifica apenas a) diremos que (\bar{x}, \bar{u}) cumprem a condição de Lagrange.

Definição 4.12 Uma função definida da seguinte maneira: $x \in \mathbb{R}^n \mapsto \langle s, x \rangle + k$, para algum $s \in \mathbb{R}^n$ e algum $k \in \mathbb{R}$, é chamada função afim.

Seja $X = \{x \in \mathbb{R}^n : g_i(x) \leq 0\}$ o conjunto de pontos viáveis para o problema (P3). Denotemos por J_a o conjunto de índices correspondentes às restrições g_i afins:

$$J_a = \{i = 1, \dots, m : g_i \text{ é uma função afim}\}$$

Definição 4.13 Dizemos que X satisfaz a hipótese fraca de Slater se existe um ponto $x^o \in \mathbb{R}^n$ no qual acontece:

i) $g_i(x^o) \leq 0$ para $i \in J_a$

ii) $g_i(x^o) < 0$ para $i \notin J_a$

Ao assumir a Hipótese 4.2, estar-se-á satisfazendo a hipótese fraca de Slater.

Teorema 4.14 Suponhamos que se verifica a hipótese fraca de Slater. Então, para qualquer $\bar{x} \in \mathbb{R}$, o fato de \bar{x} ser um ponto de KKT é suficiente e necessário para que \bar{x} seja uma solução do problema (P3).

Prova:

Ver Teorema 2.2.5, pag 310, em [8]

4.3 Descrição do Método Lagrangeano Aumentado

Consideremos a função Lagrangeano Aumentado com penalidades $P_i \in \mathcal{P}$ associada ao problema (P3) e $\{\lambda_k\} \subset \mathbb{R}$ uma seqüência limitada de números reais positivos.

O método Lagrangeano aumentado para a otimização convexa tenta aproximar um ponto de KKT (que sob a hipótese (4.1), (4.2) e pelo Teorema 4.14, será também uma solução do problema (P3)) mediante a seqüência $\{x^k\} \subset \mathbb{R}^n$ gerada ao resolver iterativamente, partindo de algum ponto $\mu^0 \in \mathbb{R}_{++}^m$ dado, problemas do tipo:

$$\begin{aligned} &\text{minimizar} && L(x, \mu_k, \lambda_k) = f(x) + \lambda_k \sum_{i=1}^m P_i\left(\frac{g_i(x)}{\lambda_k}, \mu_i^k\right) && (\hat{P}_k) \\ &&& x \in \mathbb{R}^n \end{aligned}$$

onde a atualização de μ^k se realiza pela fórmula:

$$\mu_i^{k+1} = \frac{\partial P_i}{\partial y_i}(y_i^{k+1}, \mu_i^k), \quad y_i^{k+1} = \frac{g_i(x^{k+1})}{\lambda_k}, \quad \forall i = 1, \dots, m \quad (4.3)$$

(x^{k+1} denota uma solução do problema (\hat{P}_k))

A idéia é procurar verificar as condições dadas pela Definição 4.11.

Primeiro impomos a verificação da condição de Lagrange para todos os iterados. Expliquemos isto.

Para x^{k+1} , uma solução do problema (\hat{P}_k) (que por enquanto suporemos que existe), temos pela Observação 4.8 e a Proposição 2.19

$$0 \in \partial L_{\mu^k, \lambda_k}(x^{k+1}) \quad (4.4)$$

Avaliando a relação (4.1) em (μ_k, λ_k) obtemos:

$$L_{\mu^k, \lambda_k}(x) = f(x) + \lambda_k \sum_{i=1}^m P_{i, \mu_i^k}\left(\frac{g_i(x)}{\lambda_k}\right),$$

usando agora o Corolário 2.21:

$$\partial L_{\mu^k, \lambda_k}(x) = \partial f(x) + \lambda_k \sum_{i=1}^m \partial \left[P_{i, \mu_i^k}\left(\frac{g_i(x)}{\lambda_k}\right) \right]. \quad (4.5)$$

Como P_{i, μ_i^k} é crescente (Observação 4.7) e diferenciável em \mathbb{R} , pelo Teorema 4.9,

$$\partial \left(P_{i, \mu_i^k} \circ \frac{g_i(x)}{\lambda_k} \right) (x) = \frac{\partial}{\partial y_i} P_{i, \mu_i^k}\left(\frac{g_i(x)}{\lambda_k}\right) \frac{\partial g_i(x)}{\lambda_k}, \quad y_i = \frac{g_i(x)}{\lambda_k}$$

Substituindo isto em (4.5), avaliando em x^{k+1} e retomando (4.4)

$$0 \in \partial L_{\mu^k, \lambda_k}(x^{k+1}) = \partial f(x^{k+1}) + \sum_{i=1}^m \underbrace{\frac{\partial}{\partial y_i} P_{i, \mu_i^k}\left(\frac{g_i(x^{k+1})}{\lambda_k}\right)}_{\mu_i^{k+1}} \partial g_i(x^{k+1}) \quad (4.6)$$

Assim, (x^{k+1}, μ^{k+1}) verificam a condição de Lagrange, o que garante que um ponto limite da sequência $\{x^k, \mu^k\}$ (se existir) também satisfará dita condição.

Por outro lado, usando a Observação 4.10 (com $\mu = \mu^k \geq 0$, $\lambda = \lambda_k > 0$ e $\bar{x} = x^{k+1}$, para um dado $k \in \mathbb{N}$) e a fórmula (4.3) obtemos $\mu^{k+1} \geq 0$ (resp. $\mu^{k+1} > 0$ para todo $k \in \mathbb{N}$, para penalidades $P_i(\cdot, \mu_i^k)$ estritamente convexas).

Logo, a viabilidade dual se verifica para um ponto limite \bar{u} da sequência $\{\mu^k\}$.

Viabilidade primal e complementaridade (itens *c* e *d* da Definição 4.11 respectivamente), não se satisfazem ao longo dos iterados. No entanto deverão verificar-se num ponto limite (\bar{x}, \bar{u}) (isto acontece por exemplo no caso das penalidades de Polyak-Teboulle [18]).

Para conseguir esses objetivos, assim como para garantir a existência de soluções do problema (\hat{P}_k) e a convergência das sequências $\{x^k\}$ e $\{\mu^k\}$, deveremos assumir algumas hipóteses, mas isto não será tratado neste trabalho. Nós estamos interessados especificamente em provar a relação entre o método L.A. (aplicado ao problema (P3)) e o método do Ponto Proximal (aplicado ao problema dual). O teorema de equivalência que apresentaremos inclui quase todos os resultados de equivalência análogos, existentes na literatura, como os que se encontram em [12] e [18].

Começaremos analisando a boa definição do algoritmo L.A. com penalidades $P_i \in \mathcal{P}$. Para isso primeiramente apresentamos algumas definições e resultados que serão de utilidade.

4.4 Definições e Resultados Preliminares

Definição 4.15 *Uma direção $d \neq 0$ em \mathbb{R}^n é uma direção de recessão de h se $h'_\infty(d) \leq 0$, onde $h'_\infty(\cdot)$, denota a função de recessão de h definida na seção 2.1.*

Observação 4.16 *Se o conjunto de soluções ótimas do problema (P3) é não vazio e limitado (Hipótese 4.1), então as funções f, g_1, \dots, g_m não têm nenhuma direção de recessão comum.*

Prova:

Por contradição. Suponhamos que h é direção de recessão comum. Então, para x^* , uma solução ótima, temos que $\forall \lambda \geq 0$, $x^* + \lambda h$ será viável e $f(x^* + \lambda h) \leq f(x^*)$, o que é uma contradição, pois o conjunto ótimo é limitado. Logo, não existe direção de recessão comum.

□

Proposição 4.17 *Sejam f_1, \dots, f_m m funções em $\text{Conv} \mathbb{R}^n$ e t_1, \dots, t_m m números positivos. Assumamos que existe $x^0 \in \mathbb{R}^n$ no qual f_j é finito $\forall j = 1, \dots, m$. Logo, para $f = \sum_{j=1}^m t_j f_j$ temos $f'_\infty = \sum_{j=1}^m t_j f'_{j\infty}$.*

Prova:

Ver Proposição 3.2.9, pag 182 em [8].

Observação 4.18 *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa. Logo para $\bar{x} \in \mathbb{R}^n$ e $\forall h \in \mathbb{R}^n$ tal que $h \neq 0$,*

$$f'_\infty(h) = \lim_{\lambda \rightarrow \infty} f'(\bar{x} + \lambda h, h)$$

onde $f'(x, h)$ denota a derivada direcional de f em x , na direção h .

Prova:

Para todo $x \in \mathbb{R}^n$ e para qualquer $h \in \mathbb{R}^n$ tal que $h \neq 0$, pela definição de $f'(x, h)$ temos $f(x + \lambda h) \geq f(x) + \lambda f'(x, h)$ para todo $\lambda > 0$. Então

$$\lim_{\lambda \rightarrow \infty} \frac{f(x + \lambda h) - f(x)}{\lambda} \geq f'(x, h)$$

Ou seja, $f'_\infty(h) \geq f'(x, h)$. Fazendo $x = \bar{x} + \lambda h$ para $\bar{x} \in \mathbb{R}^n$ fixo e para todo $\lambda > 0$ temos:

$$f'_\infty(h) \geq f'(\bar{x} + \lambda h, h) \quad \forall \lambda > 0.$$

Tomando limite quando $\lambda \rightarrow \infty$

$$f'_\infty(h) \geq \lim_{\lambda \rightarrow \infty} f'(\bar{x} + \lambda h, h) \quad \forall \lambda > 0$$

Para $y = \bar{x} + \lambda h \in \mathbb{R}^n \quad \forall \lambda > 0$ então $f(\bar{x}) \geq f(y) + \lambda f'(y, -h)$. Substituindo y , $f(\bar{x}) \geq f(\bar{x} + \lambda h) + \lambda f'(y, -h)$. Mas por ser f convexa verifica-se também $f'(y, h) \geq -f'(y, -h)$.

Substituindo isso na relação anterior: $f(\bar{x} + \lambda h) - f(\bar{x}) \leq \lambda f'(\bar{x} + \lambda h, h) \quad \forall \lambda > 0$.

Tomando limite quando $\lambda \rightarrow \infty$:

$$\lim_{\lambda \rightarrow \infty} \frac{f(\bar{x} + \lambda h) - f(\bar{x})}{\lambda} \leq \lim_{\lambda \rightarrow \infty} f'(\bar{x} + \lambda h, h)$$

ou seja,

$$f'_\infty(h) \leq \lim_{\lambda \rightarrow \infty} f'(\bar{x} + \lambda h, h)$$

Logo temos provada a igualdade.

□

Observação 4.19 *Sejam $f: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa, $h \in \mathbb{R}^n$ e $x \in \mathbb{R}^n$.*

Então

$$f'(x, h) = \max_{s \in \partial f(x)} \{ \langle s, h \rangle \}$$

onde $\partial f(x)$ é não vazio compacto e convexo.

Prova:

Ver resultado (0.1), pag 236 em [8].

Observação 4.20 *Para $P: \mathbb{R} \rightarrow \mathbb{R}$ uma função convexa e diferenciável e $g: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa, definamos a composta $x \in \mathbb{R}^n \mapsto f(x) = P(g(x))$.*

Se $P'(t) \geq 0 \quad \forall t \in \mathbb{R}$ então $f'(x, h) = P'(g(x))g'(x, h) \quad \forall h \in \mathbb{R}^n, x \in \mathbb{R}^n$.

Prova:

Para $x \in \mathbb{R}^n$, arbitrário, $h \in \mathbb{R}^n$ fixo, da Observação 4.19 temos:

$$\begin{aligned} f'(x, h) &= \max\{ \langle s, h \rangle, s \in \partial f(x) \} \\ &= \max\{ s^t h : s = P'(g(x))\gamma, \gamma \in \partial g(x) \} \\ &= \max\{ P'(g(x))\gamma^T h : \gamma \in \partial g(x) \} \\ &= P'(g(x)) \max\{ \gamma^T h : \gamma \in \partial g(x) \} \end{aligned}$$

usando novamente a Observação 4.19 agora com g : $f'(x, h) = P'(g(x))g'(x, h)$, que é o que queríamos provar.

□

4.5 A boa definição do Algoritmo Lagrangeano Aumentado com penalidades $P_i \in \mathcal{P}$

Consideremos o problema (P3) sob a Hipótese 4.1. Seja $\{x^k\} \subset \mathbb{R}^n$ uma seqüência gerada pelo algoritmo Lagrangeano Aumentado com penalidades $P_i \in \mathcal{P}$ seguinte:

$$\mu^0 \in \mathbb{R}_{++}^m \quad (4.7)$$

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{L(x, \mu^k, \lambda_k)\} \quad (4.8)$$

$$\mu_i^{k+1} = \frac{\partial P_i}{\partial y_i}(y_i^{k+1}, \mu_i^k) \quad \forall i = 1, \dots, m \quad \forall k \in \mathbb{N} \quad (4.9)$$

onde $\{\lambda_k\}$ é uma seqüência limitada de números reais positivos, L é a função Lagrangeano Aumentado com penalidades $P_i \in \mathcal{P}$ e $y_i^{k+1} = g_i(x^{k+1})/\lambda_k$ para $i = 1, \dots, m$.

Afirmção 4.21 *A seqüência $\{x^k\}$ está bem definida.*

Prova:

Dados $\lambda_k > 0$ e $\mu^k \geq 0$ ($k \in \mathbb{N}$ arbitrário), provaremos que L_{μ^k, λ_k} não tem direção de recessão (o que segundo a Proposição 2.7 será equivalente a afirmar que os subconjuntos de nível de L_{μ^k, λ_k} são compactos. Logo L_{μ^k, λ_k} alcança seu mínimo em \mathbb{R}^n).

Da definição de L_{μ^k, λ_k} , como f e $P_i \mu_i^k \circ g_i/\lambda$ estão em $C\overline{\operatorname{conv}} \mathbb{R}^n$, pela Proposição 4.17 temos:

$$L'_{\infty}(z, \mu^k, \lambda_k) = f'_{\infty}(z) + \lambda_k \sum_{i=1}^m h'_{i\infty}(z) \quad \forall z \in \mathbb{R}^n$$

onde $h_i(x) = P_i(g_i(x)/\lambda_k, \mu_i^k) \quad \forall i = 1, \dots, m$ ou também usando a Observação 4.18:

$$\lim_{\lambda \rightarrow \infty} L'_{\mu^k, \lambda_k}(x + \lambda z, z) = \lim_{\lambda \rightarrow \infty} f'(x + \lambda z, z) + \lambda_k \sum_{i=1}^m \lim_{\lambda \rightarrow \infty} h_i(x + \lambda z, z) \quad (4.10)$$

Por outro lado, para $z \neq 0$ em \mathbb{R}^n , temos duas alternativas para uma restrição g_i arbitrária: z é direção de recessão de g_i ou não é.

i) Suponhamos que $g'_{i\infty}(z) \leq 0$. Então para \bar{x} um ponto viável do problema (P3) e $\lambda > 0$ temos, pela Observação 4.20,

$$h'_i(\bar{x} + \lambda z, z) = P'_i(g_i(\bar{x} + \lambda z)/\lambda_k, \mu_i^k)[g'_i(\bar{x} + \lambda z, z)]/\lambda_k.$$

Quando $\lambda \rightarrow \infty$ temos duas alternativas:

- i) $g_i(\bar{x} + \lambda z) \rightarrow -\infty$. Então $P'_i(g_i(\bar{x} + \lambda z)/\lambda_k, \mu_i^k) \rightarrow 0$ e $g'_i(\bar{x} + \lambda z, z)$ é limitada, ou
- ii) $g_i(\bar{x} + \lambda z)$ é limitada. Logo $P'_i(g_i(\bar{x} + \lambda z)/\lambda_k, \mu_i^k)$ é limitada e como g_i é convexa $g'_i(\bar{x} + \lambda z) \rightarrow 0$.

Em ambos os casos,

$$h'_i(\bar{x} + \lambda z, z) \rightarrow 0 \text{ quando } \lambda \rightarrow \infty. \quad (4.11)$$

ii) Se $g'_{i\infty}(z) > 0$, $g_i(\bar{x} + \lambda z) \rightarrow \infty$ e $P'_i(g_i(\bar{x} + \lambda z)/\lambda_k, \mu_i^k) \rightarrow \infty$ quando $\lambda \rightarrow \infty$.

Ou seja,

$$h'_i(\bar{x} + \lambda z, z) \rightarrow \infty. \quad (4.12)$$

Substituindo estes resultados em (4.10) e usando a Observação 4.18, temos:

$$L'_\infty(z, \mu^k, \lambda_k) = \begin{cases} f'_\infty(z) & \text{se } z \text{ é direção de recessão das funções } g_i, \\ \infty & \text{caso contrário.} \end{cases} \quad (4.13)$$

Finalmente, suponhamos que existe $z^0 \neq 0$ direção de recessão de L_{μ^k, λ_k} , logo, de (4.13), $f'_\infty(z^0) \leq 0$ (ou seja, z^0 é direção de recessão das g_i e da f), o que contradiz a Observação 4.16.

Portanto, $L'_\infty(z, \mu^k, \lambda_k) > 0 \quad \forall z \neq 0$ em \mathbb{R}^n .

□

4.6 Tipos de penalidades usadas pelo Método Lagrangeano Aumentado

Para funções $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$, convexas próprias define-se a função conjugada, denotada f^* e expressa por:

$$s \in \mathbb{R}^n \rightarrow f^*(s) = \sup_{x \in \text{dom} f} \{\langle s, x \rangle - f(x)\}$$

Dado que, para $P_i \in \mathcal{P}$ e $u \geq 0$ fixo, $P_i(\cdot, u)$ é convexa, então podemos definir $P_i^*(\cdot, u)$, a função conjugada de $P_i(\cdot, u)$, como

$$s \in \mathbb{R} \mapsto P_i^*(s, u) = \sup_{t \in \mathbb{R}} \{st - P_i(t, u)\}.$$

Proposição 4.22 *Para $f \in \overline{\text{Conv}}\mathbb{R}^n$ verifica-se o seguinte:*

$$f(x) + f^*(s) - \langle s, x \rangle = 0 \text{ (ou } \leq 0) \iff s \in \partial f(x) \iff x \in \partial f^*(s)$$

Prova:

Ver Corolário 1.4.4 pag. 48 em [9].

Teorema 4.23 *Seja $f \in \overline{\text{Conv}}\mathbb{R}^n$ uma função estritamente convexa. Logo*

i) $\text{intdom } f^* \neq \emptyset$

ii) f^* é continuamente diferenciável no $\text{intdom } f^*$

Prova:

Teorema 4.1.1, pag.79 em [9].

Para uma melhor compreensão das seções subsequentes, detalharemos os tipos de penalidades usadas pelo Método Lagrangeano Aumentado com penalidades $P_i \in \mathcal{P}$. As duas possibilidades com respeito a $b > 0$ (ver Definição (4.3)) darão lugar a duas famílias de penalidades, a primeira das quais pode por sua vez ser subdividida em dois casos como veremos a seguir.

4.6.1 Penalidade Tipo I ($b = +\infty$)

P_1 é uma penalidade Tipo I, se $\forall u > 0$ se satisfaz:

- 1.- $P_1(., u)$ é continuamente diferenciável e estritamente convexa em \mathbb{R} ,
- 2.- a) $P_1(., u)$ é ilimitada ou
b) $P_1(., u)$ é limitada inferiormente
- 3.- $P_1(0, u) = 0, P_1'(0, u) = u$
- 4.- $\lim_{t \rightarrow -\infty} P_1'(t, u) = 0$
- 5.- $\lim_{t \rightarrow +\infty} P_1'(t, u) = +\infty$.

Exemplos deste tipo de Penalidade são $P(t, u) = u\psi(t)$, para $u > 0$, usadas por Polyak-Teboulle [18] onde para ilustrar podemos escolher:

$$\psi_1(t) = \exp(t) - 1$$

para o caso limitado e

$$\psi_2(t) = \begin{cases} \exp(t) - 1, & \text{para } t > 0, \\ -\log(1 - t) & \text{para } t \leq 0 \end{cases}$$

para o caso ilimitado.

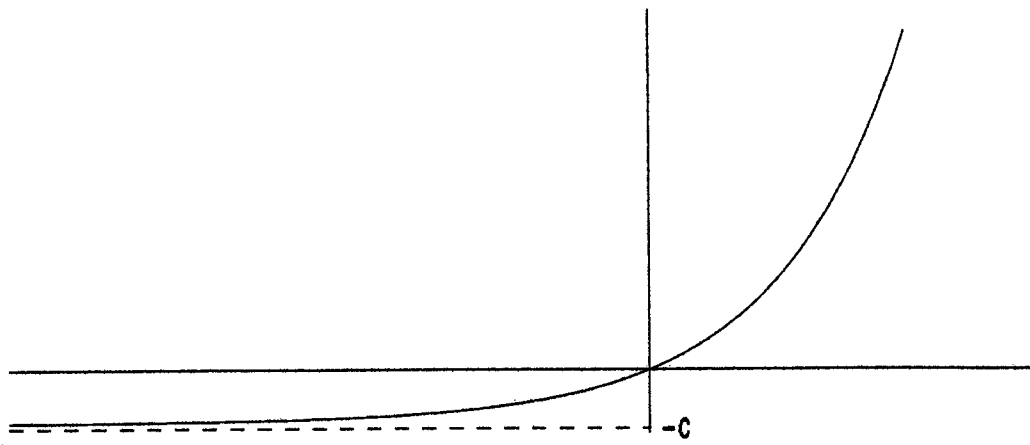
Nas Figuras 1 e 2 mostramos as funções $P_i(t, 1) = \psi_i(t)$, $i = 1, 2$ (com $u = 1 > 0$) e suas correspondentes funções conjugadas,

$$P_1^*(s, 1) = \begin{cases} s \log(s) - s + 1, & \text{para } s \geq 0. \end{cases}$$

e

$$P_2^*(s, 1) = \begin{cases} s - 1 + \log(1/s), & \text{para } 0 < s < 1, \\ s \log(s) - s + 1, & \text{para } s \geq 1 \end{cases}$$

Funcao



Conjugada

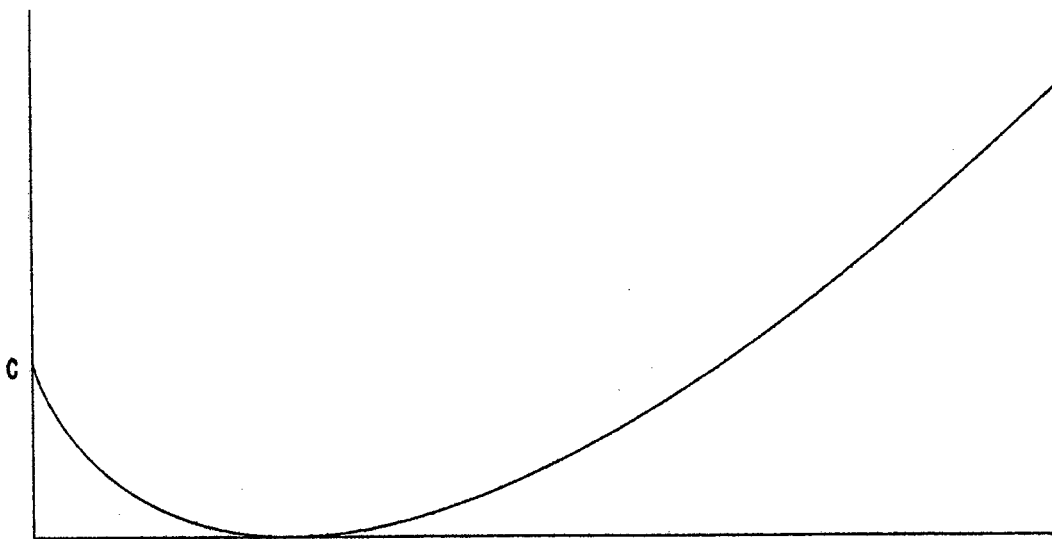
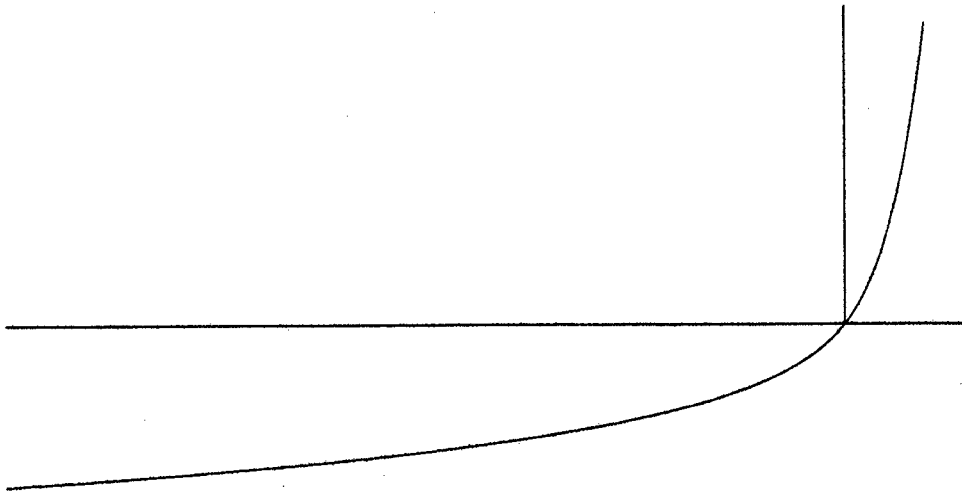


Figura 1 - Função conjugada definida e coerciva na fronteira

Funcao



Conjugada

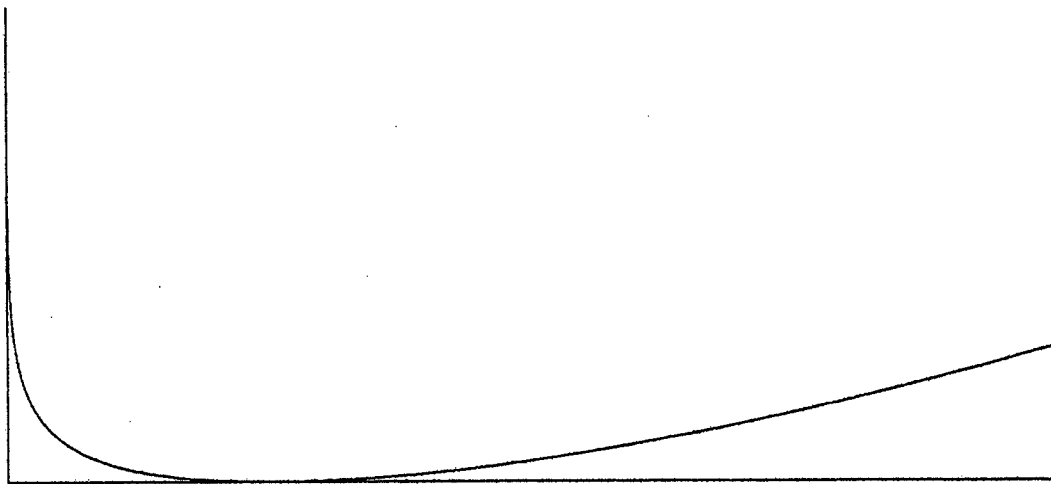


Figura 2 - Função conjugada não definida, mas coerciva na fronteira

Propriedades da conjugada $P_1^*(., u)$, com $P_1 \in \mathcal{P}$, de Tipo I.

Caso (a) ($P_1(., u)$ ilimitada)

1.- $\text{dom } P_1^*(., u) = \langle 0, +\infty \rangle$

Da Observação 4.7 sabemos que $P_1'(t, u) > 0$, para todo $t \in \mathbb{R}$. Logo, usando a Proposição 4.22, a imagem de $P_1'(., u) : P_1'(\mathbb{R}, u) \subset \text{dom } P_1^*(., u)$. Dado que $P_1'(., u)$ é contínua, $P_1'(\mathbb{R}, u) = \langle 0, +\infty \rangle \subset \text{dom } P_1^*(., u)$. Isto corresponde a $s = P_1'(t, u) > 0$.

Analisemos para $s = 0$. $P_1^*(0, u) = \sup_{t \in \mathbb{R}} \{0 - P_1(t, u)\} = -\inf_{t \in \mathbb{R}} \{P_1(t, u)\}$, dado que $P_1(., u)$ é ilimitada $P_1^*(0, u) = -(-\infty) = +\infty$.

Para $s < 0$. $P_1^*(s, u) = +\infty$ pois $P_1(., u)$ é crescente.

Assim, $\text{dom } P_1^*(., u) = \langle 0, +\infty \rangle$.

2.- $P_1^*(., u)$ é coerciva na fronteira (de seu domínio), ou seja, $\lim_{s \rightarrow 0} P_1^*(s, u) = -\infty$.

Denotemos $s = P_1'(t, u) > 0$ para $t \in \mathbb{R}$ e $u > 0$ fixos. Tomando limite quando $t \rightarrow -\infty$ temos $\lim_{t \rightarrow -\infty} s = \lim_{t \rightarrow -\infty} P_1'(t, u) = 0$. Ou seja,

$$s \rightarrow 0 \text{ quando } t \rightarrow -\infty \quad (4.14)$$

Pela Proposição 4.22, o Teorema 4.23 e a Proposição 2.14 $P_1^{*'}(s, u) = t$, pois $P_1^*(., u)$ é diferenciável em $\langle 0, +\infty \rangle$, então tomando limite quando $t \rightarrow -\infty$, obtemos que $\lim_{t \rightarrow -\infty} P_1^{*'}(s, u) = -\infty$. Finalmente, usando (4.14) para mudar de variável obtemos $\lim_{s \rightarrow 0} P_1^{*'}(s, u) = -\infty$.

Caso (b) ($P_1(., u)$ limitada inferiormente por uma constante $c < 0$)

1.- $\text{dom } P_1^*(., u) = [0, +\infty \rangle$

Similar ao caso anterior, com a diferença de que, para $s = 0$, $P_1^*(0, u) = \sup_{t \in \mathbb{R}} \{0 - P_1(t, u)\} = -\inf_{t \in \mathbb{R}} \{P_1(t, u)\} = -c$.

Assim, $\text{dom } P_1^*(., u) = [0, +\infty \rangle$.

2.- $P_1^*(., u)$ é coerciva na fronteira (de seu domínio), ou seja,

$$\lim_{s \rightarrow 0} P_1^*(s, u) = -\infty.$$

A prova é a mesma, pois $\text{intdom } P_1^*(., u) = \langle 0, +\infty \rangle$ também.

4.6.2 Penalidade Tipo II ($b > 0$ finito)

P_2 é uma penalidade Tipo II, se $\forall u \geq 0$ se satisfaz:

- 1.- $P_2(., u)$ é continuamente diferenciável em \mathbb{R} , estritamente convexa em $[-b, +\infty)$ e constante em $\langle -\infty, -b]$
- 2.- $P_2(0, u) = 0$, $P_2'(0, u) = u$
- 3.- $\lim_{t \rightarrow -\infty} P_2'(t, u) = 0$
- 4.- $\lim_{t \rightarrow +\infty} P_2'(t, u) = +\infty$

Exemplo deste tipo de penalidade são $P(t, u) = \theta(t + t_u) - \theta(t_u)$, para $u \geq 0$ introduzidas em Gonzaga-Castillo [7], onde t_u é tal que $\theta'(t_u) = u$, com $t_u = -2$ para $u = 0$.

Para ilustrar, escolhemos:

$$\theta(t) = \begin{cases} -2/3, & \text{para } t < -2, \\ t^3/12 + t^2 + t, & \text{para } -2 \leq t \leq 0, \\ \exp(t) - 1, & \text{para } t > 0. \end{cases}$$

$P_3(t, 1) = \theta(t + t_1) - \theta(t_1)$ para $t_1 = 0$ e sua conjugada são mostradas na Figura 3.

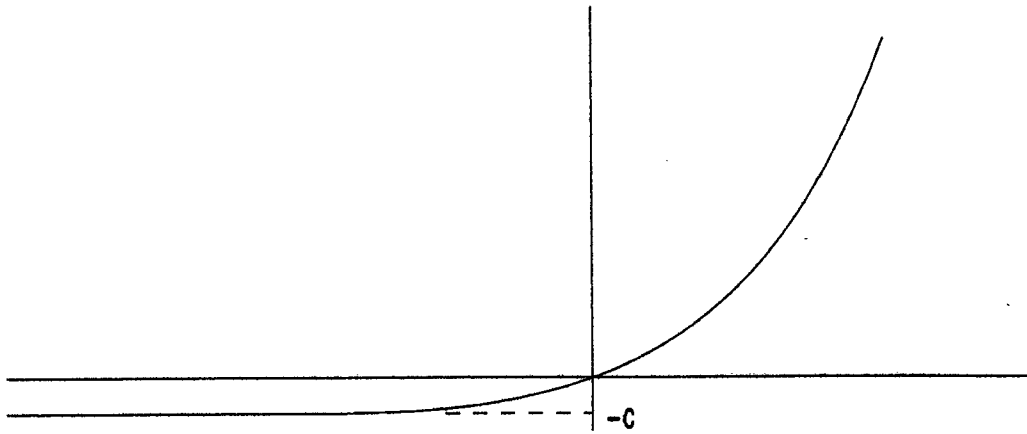
Propriedades da conjugada $P_2^*(., u)$, com $P_2 \in \mathcal{P}$, de Tipo II.

- 1.- $\text{dom } P_2^*(., u) = [0, +\infty)$

Similar aos casos anteriores, já que para $s < 0$ $P_2^*(s, u) = +\infty$ pois $P_2(., u)$ também é crescente.

Assim, $\text{dom } P_2^*(., u) = [0, +\infty)$.

Funcao



Conjugada

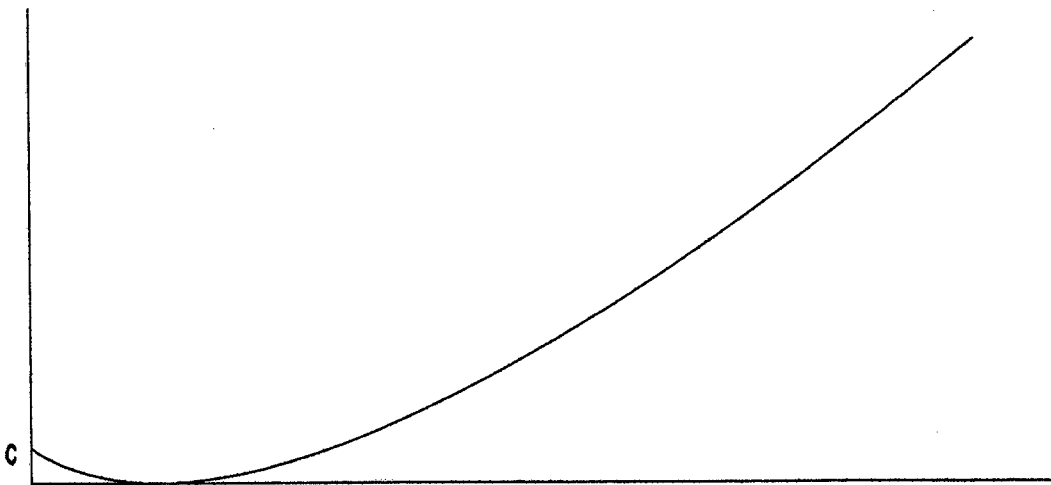


Figura 3 - Função conjugada definida e não coerciva na fronteira

2.- Se denotamos $A = (-\infty, -b]$, então $A = \partial P_2^*(0, u)$ para qualquer $u \geq 0$.

Por definição $P_2(t, u)$ é constante $\forall t \leq -b$ logo $P_2'(t, u) = 0 \forall t \leq -b$. Usando a Proposição 4.22, temos que todo elemento de $A = (-\infty, -b]$ pertencerá ao $\partial P_2^*(0, u)$. Suponhamos agora que existe $t \in \partial P_2^*(0, u)$ tal que $t > -b$, logo $0 \in \partial P_2(t, u)$ ou seja, $P_2'(t, u) = 0$ para $t > -b$, o que contradiz a definição de $P_2 \in \mathcal{P}$. Portanto, $(-\infty, -b] = \partial P_2^*(0, u)$.

4.7 Uma análise da seqüência $\{\mu^k\}$

A seguir apresentamos um resultado novo sobre a seqüência de multiplicadores $\{\mu^k\}$ gerada pelo Método L.A. com Penalidades $P_i \in \mathcal{P}$.

Lema 4.24 *Consideremos o algoritmo L.A. com penalidades $P_i \in \mathcal{P}$, dado pelas equações (4.7), (4.8) e (4.9). Em cada iteração k e para cada $i = 1, \dots, m$,*

μ_i^k é único e y_i^k é único, no caso das penalidades P_i de Tipo I,

μ_i^k é único e y_i^k é único ou $\mu_i^k = 0$ e $y_i^k \leq -b$, no caso das P_i de Tipo II.

Prova:

Para as penalidades de Tipo II:

- i) Se o problema dado em (4.8) tem mínimo único. Obviamente y^{k+1} e μ^{k+1} serão únicos.
- ii) Caso $L_{\mu^k, \lambda_k}(\cdot)$ não tenha minimizador único.

Sejam x^1, x^2 duas soluções diferentes do problema dado em (4.8). Denotemos por $y_i^1 = g_i(x^1)/\lambda_k$, $y_i^2 = g_i(x^2)/\lambda_k$ para cada $i = 1, \dots, m$ e comparemos y_i^1 com y_i^2 .

Para $y_i^1 < y_i^2$, com i arbitrário fixo. Suponhamos que $y_i^2 > -b$. Logo, para $\alpha \in (0, 1)$ temos

$$P_i(\alpha y_i^1 + (1 - \alpha)y_i^2, \mu_i^k) < \alpha P_i(y_i^1, \mu_i^k) + (1 - \alpha)P_i(y_i^2, \mu_i^k)$$

para aquele i , pois nesse intervalo $P_i(\cdot, \mu_i^k)$ é estritamente convexa. Dado que $P_i(\cdot, \mu_i^k)$ é convexa para todo $i \in \{1, \dots, m\}$, fazendo o somatório para obter L :

$$\begin{aligned} L(\bar{x}, \mu^k, \lambda_k) &< \alpha(f(x^1) + \sum_{i=1}^m P_i(y_i^1, \mu_i^k)) + (1 - \alpha)(f(x^2) + \sum_{i=1}^m P_i(y_i^2, \mu_i^k)) \\ &< \alpha L(x^1, \mu^k, \lambda_k) + (1 - \alpha)L(x^2, \mu^k, \lambda_k) \\ &< L(x^1, \mu^k, \lambda_k) \end{aligned}$$

onde $\bar{x} = \alpha x^1 + (1 - \alpha)x^2$, o que é uma contradição dado que $L(x^1, \mu^k, \lambda_k)$ é valor ótimo. Portanto, $y_i^2 \leq -b$, do que se deduz também que $\mu_i^{k+1} = 0$ pois $P_i'(t, u) = 0$ para $t \leq -b$ e $u \geq 0$ fixo. Além disso, como já tínhamos $y_i^1 < y_i^2$, podemos afirmar que $y_i^{k+1} \leq -b$, do que se deduz também que $\mu_i^{k+1} = 0$.

Se $y_i^1 = y_i^2$, y_i^{k+1} é único. Logo também μ_i^{k+1} será único.

Para as penalidades de Tipo I:

- i) Se o problema dado em (4.8) tem mínimo único. Obviamente y^{k+1} e μ^{k+1} serão únicos.
- ii) Caso $L_{\mu^k, \lambda_k}(\cdot)$ não tenha minimizador único.

Sejam x^1, x^2 duas soluções diferentes do problema dado em (4.8). Suponhamos que para i arbitrário, $y_i^1 \neq y_i^2$. Logo, para $\alpha \in (0, 1)$ temos

$$P_i(\alpha y_i^1 + (1 - \alpha)y_i^2, \mu_i^k, \lambda_k) < \alpha P_i(y_i^1, \mu_i^k) + (1 - \alpha)P_i(y_i^2, \mu_i^k)$$

pois $P_i(\cdot, \mu_i^k)$ é estritamente convexa para todo i . Fazendo o somatório para obter L , temos:

$$\begin{aligned} L(\bar{x}, \mu^k, \lambda_k) &< \alpha(f(x^1) + \sum_{i=1}^m P_i(y_i^1, \mu_i^k)) + (1 - \alpha)(f(x^2) + \sum_{i=1}^m P_i(y_i^2, \mu_i^k)) \\ &< \alpha L(x^1, \mu^k, \lambda_k) + (1 - \alpha)L(x^2, \mu^k, \lambda_k) = L(x^1, \mu^k, \lambda_k) \end{aligned}$$

onde $\bar{x} = \alpha x^1 + (1 - \alpha)x^2$, o que é uma contradição dado que $L(x^1, \mu^k, \lambda_k)$ é valor ótimo. Portanto, $y_i^2 = y_i^1$ para todo i , do que se deduz que $\mu_i^1 = \mu_i^2$, ou seja, y_i^{k+1} é único, logo μ_i^{k+1} também é único.

□

Referências que apresentam resultados de convergência para o método Lagrangeano Aumentado para algum tipo particular de penalidades são : Rockafellar [22], Polyak-Teboulle [18], Iusem [14] e Gonzaga-Castillo [7].

4.8 A relação entre o Método L.A. com Penalidades $P_i \in \mathcal{P}$ e o Método de Ponto Proximal

Associado ao problema inicial (P3) definamos seu problema dual como:

$$\begin{aligned} & \text{maximizar } \theta(\mu) & (D) \\ & \text{s.a. } \mu \geq 0 \end{aligned}$$

onde $\theta(\mu) = \inf_{x \in \mathbb{R}^n} l(x, \mu)$ e $l(x, \mu) = f(x) + \sum_{i=1}^m \mu_i g_i(x)$, para $x \in \mathbb{R}^n$, $\mu \in \mathbb{R}^m$ é a função Lagrangeano definida na seção (4.1).

Denotemos por $S = \{\mu \in \mathbb{R}^m : \mu_i > 0, \forall i = 1, \dots, m\}$ ao octante positivo de \mathbb{R}^m e por \bar{S} seu fecho; também denotemos por $P_i^*(\cdot, u)$ a conjugada de $P_i(\cdot, u)$, para $P_i \in \mathcal{P}$ e $u \geq 0$ fixo. Assim,

$$s \in \mathbb{R} \mapsto P_i^*(s, u) = \sup_{t \in \mathbb{R}} \{st - P_i(t, u)\}.$$

Lembrando o estudado nos capítulos anteriores, consideremos o método de Ponto Proximal aplicado neste caso ao problema (D) (que se pode escrever também como $\text{minimizar}_{\mu \in \bar{S}} -\theta(\mu)$) esta vez com núcleo N expresso por:

$$N(\mu, \hat{\mu}) = \sum_{i=1}^m P_i^*(\mu_i, \hat{\mu}_i)$$

onde $\mu \in S$, $\hat{\mu} \in S$ se as penalidades forem do Tipo I (a), $\mu \in \bar{S}$, $\hat{\mu} \in S$ se forem do Tipo I (b) e $\mu \in \bar{S}$, $\hat{\mu} \in \bar{S}$ no caso das de Tipo II.

Provaremos a seguir que o algoritmo de Ponto Proximal com núcleo N respectivo está bem definido.

Antes, alguns resultados a serem usados:

Observação 4.25 *Suponhamos que $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ é convexa própria. Então sua conjugada*

$$s \in \mathbb{R}^n \mapsto f^*(s) = \sup_{x \in \text{dom} f} \{ \langle s, x \rangle - f(x) \} \text{ satisfaz } f^* \in \text{Conv } \mathbb{R}.$$

Prova:

Ver Teorema 1.1.2, pag.38 em [9].

Definição 4.26 *Uma função $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ convexa própria é 0-coerciva (respectivamente 1-coerciva) quando*

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty \quad (\text{resp.} \quad \lim_{\|x\| \rightarrow +\infty} \frac{f(x)}{\|x\|} = +\infty).$$

Proposição 4.27 *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria. Logo,*

- i) *Se $x_0 \in \text{intdom} f$ então $f^* - \langle x_0, \cdot \rangle$ é 0-coerciva.*
- ii) *Em particular, se f é finita sobre \mathbb{R}^n então f^* é 1-coerciva.*

Prova:

Ver Proposição 1.39, pag.46 em [9].

Observação 4.28 *Seja $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ uma função convexa própria. Se f é 1-coerciva então é 0-coerciva.*

Prova:

Direta, da definição de limite. □

Teorema 4.29 *Seja $f \in \text{Conv } \mathbb{R}^n$. Se f é diferenciável em $\Omega = \text{intdom} f$ então f^* é estritamente convexa em cada subconjunto $C \subset \nabla f(\Omega)$.*

Prova:

Ver Teorema 4.1.3, pag. 81 em [9].

Observação 4.30 *O núcleo N expresso por:*

$$N(\mu, \hat{\mu}) = \sum_{i=1}^m P_i^*(\mu_i, \hat{\mu}_i)$$

é uma quase-distância sobre $S \times S$, se as penalidades forem do Tipo I (a) ou sobre $\bar{S} \times S$ se forem do Tipo I (b) ou em $\bar{S} \times \bar{S}$ para as penalidades de Tipo II.

Prova:

Para as penalidades Tipo II (resp. Tipo I (a))

i) $N(\mu, \mu) = 0$ para todo $\mu \in \bar{S}$ (resp. para todo $\mu \in S$).

Dado $\mu \in \bar{S}$ (resp. para todo $\mu \in S$). Da definição de P_i , temos que cada $\hat{\mu}_i \in \partial P_i(0, \mu_i)$ logo pela Proposição 4.22 (primeira equivalencia),

$$P_i^*(\mu_i, \mu_i) = \langle \mu_i, 0 \rangle - P_i(0, \mu_i) = 0$$

para cada $i = 1, \dots, m$.

Fazendo o somatório obtemos $N(\mu, \mu) = 0$

ii) Para $\mu, \hat{\mu} \in \bar{S}$ (resp. $\mu, \hat{\mu} \in S$). Se $\mu \neq \hat{\mu}$ então $N(\mu, \hat{\mu}) > 0$

Sejam $\mu, \hat{\mu} \in \bar{S}$ (resp. $\mu, \hat{\mu} \in S$) tais que $\hat{\mu}_i \neq \mu_i$, para algum $i \in \{1, \dots, m\}$.

Logo, como $0 \in \partial P_i^*(\hat{\mu}_i, \hat{\mu}_i)$ pela definição de subdiferencial temos:

$$\langle 0, \mu_i - \hat{\mu}_i \rangle \leq P_i^*(\mu_i, \hat{\mu}_i) - P_i^*(\hat{\mu}_i, \hat{\mu}_i), \quad \forall i = 1, \dots, m \quad (4.15)$$

Por outro lado, como as $P_i^*(\cdot, \hat{\mu}_i)$ são estritamente convexas em $[0, +\infty)$ (Teorema 4.29), (resp. estritamente convexas em $\langle 0, +\infty \rangle$), $\hat{\mu}_i \neq \mu_i$ e $0 \in \partial P_i^*(\hat{\mu}_i, \hat{\mu}_i)$ então

$$P_i^*(\mu_i, \hat{\mu}_i) > P_i^*(\hat{\mu}_i, \hat{\mu}_i) + \langle 0, \mu_i - \hat{\mu}_i \rangle \quad (4.16)$$

para algum $i \in \{1, \dots, m\}$.

Fazendo o somatório considerando (4.15), (4.16) e o item (i): $N(\mu, \hat{\mu}) > 0$.

Para as penalidades Tipo II (b)

i) $N(\mu, \mu) = 0$ para todo $\mu \in S$ já foi provado.

ii) $\mu \in \bar{S}, \hat{\mu} \in S$. Se $\mu \neq \hat{\mu}$ então $N(\mu, \hat{\mu}) > 0$

O caso $\mu \in S$ e $\hat{\mu} \in S$, já foi provado.

Caso $\mu \in \partial S$ e $\hat{\mu} \in S$, $\hat{\mu}_i \neq \mu_i$ para todo i , logo (4.16) vale para todo i .

Assim, somando obtemos também $N(\mu, \hat{\mu}) > 0$.

□

4.9 A boa definição do Algoritmo de Ponto Proximal com núcleo N

Consideremos o problema (D) sob as Hipóteses (4.1) e (4.2). Se $\{\hat{\mu}^k\}$ é uma seqüência gerada por:

$$\hat{\mu}^0 \in S \quad (4.17)$$

$$\hat{\mu}^{k+1} = \operatorname{argmin}_{\mu \in \bar{S}} \left\{ -\theta(\mu) + \lambda_k \sum_{i=1}^m P_i^*(\mu_i, \hat{\mu}_i^k) \right\} \quad \forall k \in \mathbb{N} \quad (4.18)$$

onde $\{\lambda_k\}$ é uma seqüência limitada de números reais positivos e as $P_i \in \mathcal{P}$ são todas de um mesmo tipo.

Afirmção 4.31 *A seqüência $\{\hat{\mu}^k\}$ está bem definida.*

Prova:

Para as penalidades de Tipo II:

Para $\hat{\mu} \in \bar{S}$ arbitrário e $\lambda > 0$, escolhamos $M \in \mathbb{R}$ tal que $M \geq -\theta(\hat{\mu})$ e definamos

$$\Omega = \{ \mu \in \bar{S} : -\theta(\mu) + \lambda N(\mu, \hat{\mu}) \leq M \}.$$

i) Ω é não vazio já que $\hat{\mu} \in \Omega$ (pois N é uma quase-distância).

ii) Ω é limitado. Para $\hat{\mu}_i \geq 0$, $P_i(\cdot, \hat{\mu}_i)$ é convexa. Como, além disto, $P_i(\cdot, \hat{\mu}_i)$ é finita, então, pela Proposição 4.27, $P_i^*(\cdot, \hat{\mu}_i)$ será 1-coerciva (logo pela Observação 4.25 e a Proposição 4.28, 0-coerciva).

Agora: $P_i^*(\mu_i, \hat{\mu}_i) \rightarrow +\infty$ quando $|\mu_i| \rightarrow +\infty$ implica $N(\mu, \hat{\mu}) \rightarrow +\infty$ quando $\|\mu\| \rightarrow +\infty$. (Ou seja, $N(\cdot, \hat{\mu})$ é 0-coerciva).

Por outro lado, usando as hipóteses (4.1) e (4.2) temos que o problema (D) tem solução, o que nos permite concluir que $-\theta(\cdot) + \lambda N(\cdot, \hat{\mu})$ também será 0-coerciva, o qual a sua vez garante que o conjunto de nível não vazio:

$$\hat{\Omega} = \{\mu \in \mathbb{R}^m : -\theta(\mu) + \lambda N(\mu, \hat{\mu}) \leq M\}$$

é compacto (logo limitado).

Por último, como $\Omega \subset \hat{\Omega}$ então Ω também será limitado.

iii) Dado que $-\theta(\cdot) + \lambda N(\cdot, \hat{\mu})$ é contínua em \bar{S} , Ω é fechado.

Assim, pelo Teorema 2.6, $\bar{\theta}_{\hat{\mu}}(\cdot) = -\theta(\cdot) + \lambda N(\cdot, \hat{\mu})$ alcança seu mínimo em \bar{S} . Além disso, $N(\cdot, \hat{\mu})$ é estritamente convexa em \bar{S} pois as $P_i^*(\cdot, \hat{\mu}_i)$ são estritamente convexas em $[0, +\infty)$ (Ver Teorema 4.29). Logo $\bar{\theta}_{\hat{\mu}}(\cdot)$ será estritamente convexa em \bar{S} já que $-\theta(\cdot)$ é convexa. Portanto, $\bar{\mu} \in \bar{S}$ o minimizador de $\bar{\theta}_{\hat{\mu}}(\cdot)$ é único.

Fazendo $\hat{\mu}^k = \hat{\mu}$ e $\lambda_k = \lambda$ para cada $k \in \mathbb{N}$, teremos que $\hat{\mu}^{k+1} = \bar{\mu}$. Logo a afirmação está provada.

Para as penalidades de Tipo I:

Tomando $\hat{\mu} \in S$ arbitrário, por um desenvolvimento análogo ao anterior, obtemos que $\bar{\mu} \in \bar{S}$ o minimizador de $\bar{\theta}_{\hat{\mu}}(\cdot)$ é único.

Vejamos que neste caso $\bar{\mu} \in S$.

Da Propriedade 2 das penalidades de Tipo I temos que $\lim_{s \rightarrow 0} P_i^*(s, \hat{u}) = -\infty$, para $\hat{u} > 0$ fixo, o que significa que $\partial N(\mu, \hat{\mu}) = \emptyset$ para $\mu \in \partial S$, logo

$$\partial \theta_{\hat{\mu}}(\mu) = \emptyset \text{ para todo } \mu \in \partial S. \quad (4.19)$$

Como o problema dado em (4.18), equivale a: minimizar $\mu \in \mathbb{R}^m \theta_{\hat{\mu}}(\mu) + I_{\bar{S}}(\mu)$, temos

$$0 \in \partial \theta_{\hat{\mu}}(\bar{\mu}) + \partial I_{\bar{S}}(\bar{\mu}), \text{ onde } \partial I_{\bar{S}}(\bar{\mu}) \neq \emptyset \text{ pois } \bar{\mu} \in \bar{S} \quad (4.20)$$

Se $\bar{\mu} \in \partial S$, por (4.20) $-b \in \partial\theta_{\bar{\mu}}(\bar{\mu})$ (com $b \in \partial I_{\bar{S}}(\bar{\mu})$) o que contradiz (4.19). Logo $\bar{\mu} \in S$.

Fazendo $\hat{\mu}^k = \hat{\mu}$ e $\lambda_k = \lambda$ para cada $k \in \mathbb{N}$, teremos que $\hat{\mu}^{k+1} = \bar{\mu}$. Logo $\hat{\mu}^k \in S$ para todo $k \in \mathbb{N}$.

□

4.10 A equivalência entre o Método Lagrangeano Aumentado e o Método de Ponto Proximal

Observação 4.32 *Seja $Y \subset \mathbb{R}^m$, definamos a função*

$$x \in \mathbb{R}^n, y \in Y \mapsto f(x, y) \in \mathbb{R}, \text{ que verifica:}$$

i) $f(x, \cdot)$ é diferenciável e convexa $\forall x \in \mathbb{R}^n$

ii) $f(\cdot, y)$ é contínua $\forall y \in Y$.

Para $y \in Y$, definamos $\varphi(y) = \sup_{x \in \mathbb{R}^n} f(x, y)$ e $I(y) = \{x \in \mathbb{R}^n, f(x, y) = \varphi(y)\}$.

Logo,

$$\nabla_y f(x, y) \in \partial\varphi(y) \text{ para todo } x \in I(y).$$

Prova:

φ também é convexa.

Dado que $f(x, \cdot)$ é diferenciável e convexa $\forall x \in \mathbb{R}^n$, pela definição de subgradiente num ponto dado $\bar{y} \in Y$ temos: $f(x, y) \geq f(x, \bar{y}) + \nabla_y f(x, \bar{y})^t(y - \bar{y})$ para todo $y \in Y$ e para qualquer $x \in \mathbb{R}^n$. Particularizando para $\bar{x} \in I(\bar{y})$, temos

$$f(\bar{x}, y) \geq f(\bar{x}, \bar{y}) + \nabla_y f(\bar{x}, \bar{y})^t(y - \bar{y}) \text{ para todo } y \in Y$$

Como $f(\bar{x}, \bar{y}) = \varphi(\bar{y})$ e além disto, da definição de φ temos que $\varphi(y) \geq f(\bar{x}, y)$, então

$$\varphi(y) \geq \varphi(\bar{y}) + \nabla_y f(\bar{x}, \bar{y})^t(y - \bar{y}) \text{ para todo } y \in Y$$

Portanto, $\nabla_y f(x, y) \in \partial\varphi(y)$ para todo $x \in I(y)$.

□

Teorema 4.33 (*I, de Equivalência*)

Consideremos o problema (P3) sob as hipóteses (4.1) e (4.2). Seja $\{\mu^k\} \subset \mathbb{R}^m$ a seqüência gerada pelo algoritmo L.A. com penalidades $P_i \in \mathcal{P}$ de Tipo I aplicado a (P3) e $\{\hat{\mu}^k\} \subset \mathbb{R}^m$ a seqüência gerada pelo método de P.P. com núcleo N aplicado esta vez ao problema dual (D).

Se $\mu^0 = \hat{\mu}^0 > 0$, então $\mu^k = \hat{\mu}^k > 0 \quad \forall k \in \mathbb{N}$. Ou seja, as seqüências são as mesmas.

Prova:

Por indução. Dados $\mu^k > 0$ em \mathbb{R}^m e $\lambda_k > 0$ ($k \in \mathbb{N}$ arbitrário), consideremos x^{k+1} uma solução do problema (\hat{P}_k). Então $0 \in \partial L(x^{k+1}, \mu^k) = \partial l(x^{k+1}, \mu^{k+1})$ onde μ^{k+1} é dada pela fórmula (4.3). Ou seja,

$$\theta(\mu^{k+1}) = l(x^{k+1}, \mu^{k+1}) \quad (4.21)$$

Da definição de θ podemos ter:

$$-\theta(\mu) = \sup_{x \in \mathbb{R}^n} -l(x, \mu)$$

a qual é uma função convexa.

Por outro lado $-l(\cdot, \mu)$ é contínua para todo $\mu \in \mathbb{R}^m$, além disso, $-l(x, \cdot)$ é convexa e diferenciável, para cada $x \in \mathbb{R}^n$ logo pelo Observação 4.32:

$$\nabla_{\mu}(-l(x, \mu)) \in \partial(-\theta(\mu))$$

para $x = x^{k+1} \in I(\mu) = I(\mu^{k+1})$.

Assim, usando (4.21):

$$(-g_1(x^{k+1}), \dots, -g_m(x^{k+1}))^t \in \partial(-\theta(\mu^{k+1})). \quad (4.22)$$

Sabemos também que $\mu_i^{k+1} = P'_i(g_i(x^{k+1})/\lambda_k, \mu_i^k) > 0$, então pela Proposição 4.22, $g_i(x^{k+1})/\lambda_k \in \partial P_i^*(\mu_i^{k+1}, \mu_i^k)$. Como $P_i^*(\cdot, \mu_i^k)$ é diferenciável em $\langle 0, +\infty \rangle$ teremos: $(g_i(x^{k+1})/\lambda_k) = P_i^{*'}(\mu_i^{k+1}, \mu_i^k)$ para cada $i = 1, \dots, m$.

Substituindo isto em (4.22) obtemos :

$$-\lambda_k (P_1^{*'}(\mu_1^{k+1}, \mu_1^k), \dots, P_m^{*'}(\mu_m^{k+1}, \mu_m^k))^t \in \partial(-\theta(\mu^{k+1}))$$

ou equivalentemente:

$$0 \in \partial(-\theta(\mu^{k+1}) + \lambda_k \partial N(\mu^{k+1}, \mu^k)). \quad (4.23)$$

Denotemos por (\bar{P}_k) ao problema dado em (4.18). Seja $\hat{\mu}^k$ uma solução do problema (\bar{P}_{k-1}) e suponhamos que $\mu^k = \hat{\mu}^k$. Logo substituindo em (4.23):

$$\gamma = 0 \in \partial(-\theta(\mu^{k+1})) + \lambda_k \partial N(\mu^{k+1}, \hat{\mu}^k).$$

Como também verifica-se $\gamma = 0$, $-\mu_i^{k+1} < 0$ (propriedade da seqüência de multiplicadores do método L.A. com penalidades de Tipo I), e $\gamma_i(-\mu_i^{k+1}) = 0$, então μ^{k+1} será um ponto de K.K.T. para o problema dado em (4.18). Como todas as restrições deste problema são afins, μ^{k+1} será também uma solução de dito problema. Finalmente, pela Afirmacão 4.31, $\mu^{k+1} = \hat{\mu}^{k+1}$. Portanto, as seqüências são equivalentes.

□

Teorema 4.34 (II, de Equivalência)

Consideremos o problema (P3) sob as hipóteses (4.1) e (4.2). Seja $\{\mu^k\} \subset \mathbb{R}^m$ a seqüência gerada pelo algoritmo L.A. com penalidades $P_i \in \mathcal{P}$ de Tipo II aplicado a (P3) e $\{\hat{\mu}^k\} \subset \mathbb{R}^m$ a seqüência gerada pelo método de P.P. com núcleo N aplicado esta vez ao problema dual (D).

Se $\mu^0 = \hat{\mu}^0 \geq 0$, então $\mu^k = \hat{\mu}^k \geq 0 \quad \forall k \in \mathbb{N}$. Ou seja, as seqüências são as mesmas.

Prova:

Por indução. Dados $\mu^k \geq 0$ em \mathbb{R}^m e $\lambda_k > 0$ ($k \in \mathbb{N}$ arbitrário), consideremos x^{k+1} uma solução do problema (\hat{P}_k) . Então, $0 \in \partial L(x^{k+1}, \mu^k) = \partial l(x^{k+1}, \mu^{k+1})$ onde μ^{k+1} é dada pela fórmula (4.3). Ou seja,

$$\theta(\mu^{k+1}) = l(x^{k+1}, \mu^{k+1}). \quad (4.24)$$

Da definição de θ podemos ter:

$$-\theta(\mu) = \sup_{x \in \mathbb{R}^n} -l(x, \mu),$$

a qual é uma função convexa.

Por outro lado, $-l(\cdot, \mu)$ é contínua para todo $\mu \in \mathbb{R}^m$, além disso, $-l(x, \cdot)$ é convexa e diferenciável para cada $x \in \mathbb{R}^n$, logo, pela Observação 4.32,

$$\nabla_{\mu}(-l(x, \mu)) \in \partial(-\theta(\mu))$$

para $x = x^{k+1} \in I(\mu) = I(\mu^{k+1})$.

Assim, usando (4.24):

$$(-g_1(x^{k+1}), \dots, -g_m(x^{k+1}))^t \in \partial(-\theta(\mu^{k+1})). \quad (4.25)$$

Sabemos também que $\mu_i^{k+1} = P'_i(g_i(x^{k+1})/\lambda_k, \mu_i^k) \geq 0$ para todo $i = 1, \dots, m$.

Assim a respeito de μ_i^{k+1} , temos duas possibilidades:

- i) Se $\mu_i^{k+1} > 0$ então $P'_i(\cdot, \mu_i^k)$ é diferenciável em μ_i^{k+1} e $P''_i(\mu_i^{k+1}, \mu_i^k) = g_i(x^{k+1})/\lambda_k > -b$
- ii) Se $\mu_i^{k+1} = 0$ então $P'_i(g_i(x^{k+1})/\lambda_k, \mu_i^k) = 0$. Logo $g_i(x^{k+1})/\lambda_k \in \partial P'_i(0, \mu_i^k)$ e pela Propriedade 2 das penalidades de tipo II: $g_i(x^{k+1})/\lambda_k \leq -b$.

Tomando isto em consideração, existirá $\gamma \in \mathbb{R}^m$ tal que

- i) $\gamma_i = -g_i(x^{k+1}) + \lambda_k(g_i(x^{k+1})/\lambda_k) = 0$ se $\mu_i^{k+1} > 0$
- ii) $\gamma_i = -g_i(x^{k+1}) + \lambda_k(-b) \geq 0$ se $\mu_i^{k+1} = 0$.

Como $-b \in \partial P'_i(0, \mu_i^k)$ e $P''_i(\mu_i^{k+1}, \mu_i^k) = g_i(x^{k+1})/\lambda_k$ para $\mu_i^{k+1} > 0$, então

$$\gamma \in (-g_1(x^{k+1}), \dots, -g_m(x^{k+1}))^t + \lambda_k \partial N(\mu^{k+1}, \mu^k). \quad (4.26)$$

Suponhamos que $\mu^k = \hat{\mu}^k$, onde $\hat{\mu}^k$ é uma solução do problema (\bar{P}_{k-1}) e substituindo em (4.26) teremos:

$$\gamma \in \partial(-\theta(\mu^{k+1})) + \lambda_k \partial N(\mu^{k+1}, \hat{\mu}^k).$$

Como também verifica-se $\gamma \geq 0$, $-\mu_i^{k+1} \leq 0$ e $\gamma_i(-\mu_i^{k+1}) = 0$ e como antes, as restrições do problema (\bar{P}_k) são afins, μ^{k+1} será também uma solução de (\bar{P}_k) . Finalmente, pela Afirmacão 4.31, $\mu^{k+1} = \hat{\mu}^{k+1}$.

□

Bibliografia

- [1] Bertsekas, D. P., *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [2] Censor, Y. e Zenios, S., "The proximal minimization algorithm with D-functions", *JOTA*, vol 73, pp.451–464, 1992.
- [3] Chen, G. e Teboulle, M., "Convergence analysis of a proximal-like minimization algorithm using Bregman function", *SIAM J. Optimization*, 3, pp. 538–543, 1993.
- [4] Csiszar, I., "Information-type measures of difference of probability distributions and indirect observations", *Studia Sci. Math. Hungar*, 2, pp. 299–318, 1967.
- [5] Eggermont, P., "Multiplicative Iterative Algorithms for convex programming", *Linear Algebra Appl.*, pp. 25–42, 1990.
- [6] Eckstein, J., "Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming", *Mathematics of Operations Research*, 18, pp. 202–226, 1993.
- [7] Gonzaga, C. e Castillo, R., "Métodos de Lagrangeano Aumentado usando Penalidades Generalizadas para Programação não linear " Tese, COPPE, UFRJ 1998.
- [8] Hiriart-Urruty, J.-B. e Lemarechal, C., *Convex Analysis and Minimization Algorithms I*, 1 ed. New York, Springer-Verlag, 1993.
- [9] Hiriart-Urruty, J.-B. e Lemarechal, C., *Convex Analysis and Minimization Algorithms II*, 1 ed. New York, Springer-Verlag, 1993.

- [10] Hestenes, M., "Multiplier and Gradient Methods", JOTA, vol. 4, pp.303–320, 1.969.
- [11] Humes, C. e da Silva, P., *Proximal point, separators and augmented lagrangians*, Work paper, U.S.P. São Paulo, Brasil, 1997.
- [12] Iusem, A., *Métodos de ponto proximal em Otimização*, 20mo Colóquio Brasileiro de matemática, IMPA, R.J. Brasil, 1995.
- [13] Iusem, A., Teboulle, M. e Svaiter, B., "Entropy-like proximal methods in convex programming", Mathematics of Operations Research, 19(4), pp. 790–814, 1994.
- [14] Iusem, A., *Augmented Lagrangian methods and proximal point methods for convex optimization*, Minicurso I workshop em otimização, Florianópolis, Brasil, DEzembro 1.997.
- [15] Iusem, A. e Teboulle, M., "On the convergence rate of entropic proximal optimization methods", Computational and Applied Mathematics, 12, pp. 153–168, 1.993.
- [16] Marsden, J., Hoffman, M., *Elementary Classical Analysis*, 2 ed. W. H.Freeman & Company , New York, 1993.
- [17] Martinet, B., "Regularisation D'inequations variationnelles par approximations successives", Revue Française de Informatiqué et Recherche Opérationelle 2, pp. 154–159, 1970.
- [18] Polyak, R. e Teboulle, M., "Nonlinear rescaling and proximal-like methods in convex programming", 76, pp. 697–739, 1.995.
- [19] Powell, M., "A method for nonlinear constraints in minimizations problems", in Optimization, R.Fletcher, Ed.,Academic Press, N.Y, pp.283–298, 1.969.
- [20] Rockafellar, R. T., *Convex Analysis*, 1 ed. Princeton University Press, Princeton, N.J.1970.

- [21] Rockafellar, R. T., "Monotone operators and the proximal point algorithm in convex programming", *SIAM J. on Control and Optimization* 14, pp. 877–898, 1976.
- [22] Rockafellar R. T., "Augmented Lagrangians and applications of the proximal point algorithm in convex programming", *Mathematics of Operations Research* 1, pp. 97–116, 1976.
- [23] Teboulle, M., "Entropic proximal mappings with applications to nonlinear programming", *Mathematics of Operations Research* 17, pp. 97–116, 1992.