

PENALIDADES GENERALIZADAS E MÉTODOS DE LAGRANGEANO
AUMENTADO PARA PROGRAMAÇÃO NÃO LINEAR

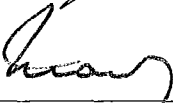
Rómulo Alberto Castillo Cárdenas

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PRO-
GRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NE-
CESSÁRIOS PARA OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS
EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

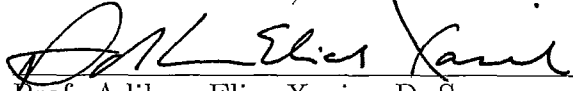
Aprovada por:



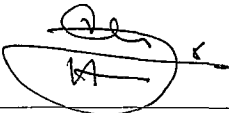
Prof. Clóvis Caesar Gonzaga, D.Sc.



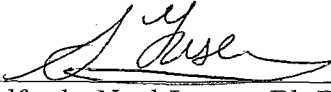
Prof. Nelson Maculan Filho, D.Habil.



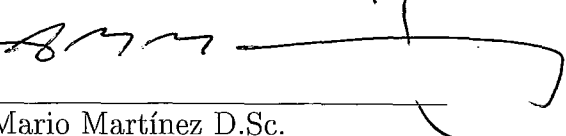
Prof. Adilson Elias Xavier, D. Sc.



Prof. Carlos Humes Júnior Ph.D



Prof. Alfredo Noel Iusem, Ph.D



Prof. José Mario Martínez D.Sc.

RIO DE JANEIRO, RJ – BRASIL
FEVEREIRO DE 1998

CASTILLO CÁRDENAS, RÓMULO ALBERTO

Penalidades generalizadas e métodos de
lagrangeano aumentado para programação
não-linear [Rio de Janeiro] 1998

IX, 120 p., 29.7 cm, (COPPE/UFRJ, D. Sc.,
Engenharia de Sistemas e Computação

TESE – Universidade Federal do Rio de
Janeiro, COPPE

1. Método de lagrangeano aumentado
 2. Programação Não Linear 3. Quase-distâncias
- I. COPPE/UFRJ II. Título(Série)

Pouco poderia ofertar a quem com infinita sabedoria mora no meu coração, mas na minha ignorância, por não compreender Sua Majestade, dedico o fruto da Sua Vontade através de meus atos.

À Bhagavan Sri Sathya Sai Baba.

Agradecimentos

Ao Prof. Clóvis Gonzaga, como um gesto de respeito e admiração por sua grande contribuição na minha formação acadêmica e nesta tese, por suas qualidades humanas e trabalho abnegado.

A Malón por sua compreensão e apoio durante minha ausência. A minha esposa Mercedes, meus filhos Javier, Alejandra e Prema pelo carinho e dificuldades afrontadas, a minha mãe e irmãos por sua solidariedade e a meu falecido pai quem sempre preocupou-se com minha educação.

A meu colega Hugo Lara com quem compartilho desde o início desta jornada e a Marco Antonio com quem compartilhei grande parte de minha estadia . A ambos agradeço sua sinceridade e amizade.

Às professoras Susana e Celina pelos conhecimentos transmitidos durante minha passagem pelo programa de pós-graduação na COPPE e em especial ao professor Adilson Xavier quem me motivou e introduziu na temática estudada nesta tese.

A Capes/Brasil, Conicit/Venezuela e UCLA/Venezuela pelo apoio recebido.

Aos professores participantes da prestigiosa banca.

Resumo da Tese apresentada à COPPE como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D. Sc.)

PENALIDADES GENERALIZADAS E MÉTODOS DE LAGRANGEANO AUMENTADO PARA PROGRAMAÇÃO NÃO-LINEAR

Rómulo Alberto Castillo Cárdenas

Fevereiro / 1998

Orientador: Clóvis Caesar Gonzaga

Programa: Engenharia de Sistemas e Computação

Apresentamos um algoritmo para resolver problemas de programação não linear não convexos e diferenciáveis usando uma família de funções de penalidade que dependem de dois parâmetros β e r . O algoritmo gera uma seqüência cujos pontos limites são soluções do problema considerado. A cada itera o parâmetro r diminui enquanto β aumenta somente se o iterado é inviável. Mostramos que se a condição de qualificação de Mangasarian-Fromovitz é satisfeita, os iterados permanecem viáveis para β suficientemente grande.

Consideramos também duas famílias de funções de penalidade a ser usadas em métodos de lagrangeano aumentado, uma delas com a novidade que a derivada no infinito na direção $+1$ é finita. Introduzimos uma relevante e conveniente mudança de variáveis em ambas as classes de penalidades tais que a derivada na origem da penalidade é igual a $\mu > 0$. Um fato interessante a destacar, é que estas mudanças nas penalidades conduzem a distâncias de Bregman via dualidade de Fenchel. Apresentamos um algoritmo de lagrangeano aumentado para cada uma das penalidades, cuja convergência está garantida pela teoria existente para distâncias de Bregman. Uma prova de convergência ergódica primal é mostrada e, sob condições apropriadas, a seqüência primal é limitada e todos os seus pontos limites são soluções ótimas do problema considerado.

Abstract of Thesis presented to COPPE as partial fulfillment of the requirements for degree of Doctor of Science (D. Sc.)

GENERALIZED PENALTY FUNCTIONS AND AUGMENTED
LAGRANGIAN METHODS FOR NON LINEAR PROGRAMMING

Rómulo Alberto Castillo Cárdenas

Fevereiro / 1998

Thesis Supervisor: Clóvis Caesar Gonzaga

Program: Systems Engineering and Computer Science

We present an algorithm for solving nonlinear and nonconvex differentiable programming problems using a family of penalty functions depending on the parameters β and r . The algorithm generates a sequence of points whose limit points are optimal solutions for the considered problem. At each iteration, the parameter r is decreased and the parameter β is augmented only if the iterate is infeasible. We show that if the Mangasarian-Fromovitz constraint qualification holds, then the iterates remain feasible for β sufficiently large.

We consider two families of penalty functions used in the augmented Lagrangian approach. One of the families has the property that the derivative at infinite in the direction $+1$ is finite. We introduce a convenient and relevant shift in both classes of penalty functions such that the derivative at the origin is equal to $\mu > 0$. It is interesting to emphasize that these shifts in the penalty functions lead to Bregman distances using the Fenchel duality theory. We present an augmented Lagrangian algorithm for each penalty. We show an ergodic convergence result for the primal sequence, furthermore, under suitable conditions this sequence is bounded and all of its limit points are optimal solutions of the considered problem.

Índice

1	Introdução	1
2	Métodos de Penalização	6
2.1	Métodos de Penalidade	8
2.1.1	Método de Barreira	8
2.1.2	Método de penalização externa	9
2.2	Penalidades generalizadas	12
2.2.1	Penalidade Tipo I	12
2.2.2	Penalidade Tipo II	13
2.3	Suavização usando penalidades generalizadas	13
2.3.1	Suavização da penalidade exata	13
2.3.2	Suavização da função indicadora do ortante negativo	14
2.4	Esquema geral de minimização	15
2.5	Técnicas de Suavização	16
2.5.1	Primeira Técnica	17
2.5.2	Segunda Técnica	20
2.5.3	Terceira Técnica	21
2.5.4	Quarta Técnica	23
2.6	Construção de penalidades tipo I	24
2.7	Método de penalidades generalizadas para programação não-linear	27
2.7.1	A função penalizada	29
2.7.2	Algoritmo.	30
2.7.3	Teoremas	31

3	Métodos de ponto proximal e minimização sobre R_+^n.	35
3.1	Método de Ponto Proximal Clássico.	35
3.2	Quase-distâncias	40
3.2.1	Distâncias de Bregman	40
3.2.2	φ -divergências	44
3.2.3	Outra quase-distância	47
3.2.4	Método de ponto proximal com regularização estritamente convexa	49
3.3	Minimização sobre R_+^n	49
3.3.1	Penalidade barreira tipo III	50
3.3.2	Penalidade barreira tipo IV	50
4	Convexidade e métodos de lagrangeano aumentado	52
4.1	Convexidade e Dualidade	52
4.2	Métodos de lagrangeano aumentado	57
4.2.1	Famílias de penalidades para o lagrangeano aumentado	58
4.3	Relação entre o método lagrangeano aumentado e o método de ponto proximal	67
5	Método de multiplicadores gerado por mudanças nas penalidades	73
5.1	Famílias de penalidades AL1 e AL2	74
5.1.1	Família de penalidades nãocoercivas	74
5.1.2	Família de penalidades coercivas pela direita	75
5.2	Método de multiplicadores baseados em shifts	80
5.2.1	Método de multiplicadores baseados em shifts para a família AL1	80
5.2.2	Método de multiplicadores baseados em shifts para a família AL2	82
5.3	Dualidade	88
5.3.1	Quase-distâncias geradas por $P_{\beta,r,\mu}^*$ e $P_{\tau,\mu}^*$	90
5.3.2	Métodos de ponto proximal usando distâncias de Bregman	96

5.3.3	Convergência	101
5.4	Relaxação da família de penalidades AL2	108
6	Conclusões	111
	Bibliografia	114

Capítulo 1

Introdução

Inumeráveis problemas surgem naturalmente nas áreas de engenharia e economia onde se precisa maximizar ganhos ou minimizar custos sob restrições que variam com cada problema dependendo da sua natureza. Estes problemas podem-se formular ou modelar em termos de funções geralmente não-lineares. Surgem assim os problemas de programação não-linear. Nesta tese estamos interessados em resolver inicialmente o problema de programação não linear não-convexa diferenciável com restrições de desigualdade dado por

$$\begin{aligned} & \text{minimizar} && f_0(x) \\ (P) \quad & \text{sujeito a} && f_i(x) \leq 0 \\ & && x \in R^n \end{aligned}$$

onde $f_i : R^n \rightarrow R$ para $i = 0, 1, \dots, m$ são funções continuamente diferenciáveis.

Posteriormente abordaremos o caso em que as funções envolvidas sejam convexas e não-diferenciáveis.

O método de lagrangeano aumentado foi introduzido no final dos anos 60 por Hestenes [30] e Powell [52] e nos últimos anos tem sido objeto de estudo por diversos pesquisadores que tem contribuído com grandes avanços teóricos e computacionais. Inicialmente foi introduzido dentro de um enfoque de penalidade para problemas com igualdades com a novidade de nãoexigir ao parâmetro de penalização crescer indefinidamente. O método foi amplamente conhecido e generalizado para o caso com desigualdades por Rockafellar [56], mostrando uma

interessante relação de dualidade com o método de ponto proximal. A penalidade usada por Rockafellar tem a desvantagem de não ser duas vezes diferenciável. Surgem então diversos enfoques que tentam melhorar as características das funções de penalidade e por outra parte, também o método de ponto proximal inicialmente com núcleo quadrático é generalizado com o uso de distâncias de Bregman [10] e φ -divergências [59]. Diversos artigos tem estudado a relação entre os métodos de lagrangeano aumentado e os métodos de ponto proximal abordando outros tipos de problemas como o de desigualdade variacional e problemas com restrições lineares, ver por exemplo, [8], [33], [6], [11], [4], [21], [50] e referências ali citadas. Uma das dificuldades encontradas nos métodos de lagrangeano aumentado está na prova de convergência da seqüência primal onde é introduzida a convergência num sentido ergódico, ver [61], [50], [53], [36]. Recentes trabalhos que abordam a ligação e situação atual de ambas metodologias com resultados importantes sobre a convergência primal são [35] e [37].

No capítulo 2 introduzimos duas famílias de penalidades (penalidade tipo I e penalidade tipo II) que por suas características motivaram seu uso nos métodos de lagrangeano aumentado estudados no capítulo 5, a principal diferença entre elas está no seu comportamento no infinito pela direita. Uma delas pode ser usada para aproximar a função de penalidade exata e a outra para aproximar a função indicadora do segundo quadrante. Assim, neste capítulo usamos a família de penalidade tipo I para aproximar a função de penalidade exata através de um parâmetro r . Fazemos um resumo das principais técnicas de suavização mais importantes que aproximam também a penalidade exata e mostramos que as mesmas seguem o formato *padrão* intruduzido. Em [14] e [15] mostra-se que a classe de penalidades que consideramos pode ser construída por dupla integração de funções de densidade de probabilidade e é aplicada a problemas de complementaridade linear, não-linear e mista. Métodos de penalidade baseados nesta família foram considerados por Auslender, Cominetti e Haddou [3] para o caso de programação convexa. Apresentamos um algoritmo para resolver o problema (P) nãoconvexo e diferenciável usando uma família de funções de penalidade general-

izadas que aproximam a função de penalidade exata envolvendo os parâmetros β e r . O algoritmo é simples de programar e gera uma seqüência de minimizadores da função critério cujos pontos limites são soluções ótimas para o problema (P). O parâmetro $\beta \geq 1$ é responsável pelo aumento da penalidade e é incrementado somente em iterados inviáveis. O parâmetro $r \in (0, 1]$ diminui em cada iteração controlando a precisão da aproximação. Um fato a resaltar é que em nossa abordagem, a partir de um valor de β suficientemente grande, os iterados permanecem viáveis e assim β não precisa crescer indefinidamente. Isto supera uma dificuldade mostrada em [3] onde se precisa que o parâmetro encarregado de aumentar a penalidade aumente indefinidamente. A metodologia de penalidade apresentada generaliza o método de penalização hiperbólica desenvolvido por Xavier [62] com diferentes hipóteses.

O objetivo principal dos capítulos 3 e 4 é apresentar as ferramentas básicas que facilitam a compreensão da temática introduzida no capítulo 5. Assim, no capítulo 3 fazemos um resumo do desenvolvimento feito durante os últimos anos no estudo de métodos de ponto proximal que usam núcleo quadrático, distâncias de Bregman e φ -divergências. Uma documentação com respeito a esta metodologia pode ser feita em [46], [55], [10], [12], [21], [32], [33], [34] e nas referências bibliográficas ali citadas. Destacamos diferentes tipos de provas que neste contexto podem ser abordadas para minimizar uma função convexa no ortante positivo. Um enfoque diferente para as provas de convergência do método usando separadores é feita em [31], neste caso a convergência é provada através de subsequências. Definimos as penalidades do tipo III e do tipo IV que serão relacionadas com as penalidades tipo I e tipo II.

No capítulo 4 estudamos o problema (P) sob hipótese de convexidade. Apresentamos os conceitos e teoremas principais sobre convexidade e dualidade que são usados em nossa abordagem. A teoria pode ser aprofundada em [54], [28] e [29]. Mostramos uma relação de dualidade entre a função de penalidade exata e a função Indicadora do intervalo $[0, \beta]$ onde β é determinado pela inclinação da penalidade exata o que motiva a relação de dualidade entre suas respectivas

aproximações. É considerado o método de lagrangeano aumentado para problemas com igualdades e para problemas com desigualdades assim como generalizações do mesmo. Um excelente texto nesta abordagem é apresentado por Bertsekas [8]. Fazemos um seguimento, de um ponto de vista primal, das distintas famílias de penalidades envolvidas em diferentes enfoques, que motiva a abordagem no capítulo 5.

O capítulo 5 introduz duas famílias de penalidades. Usando *shifts* ou mudanças de variáveis nas penalidades tipo I construímos uma nova família de penalidades até agora nãoconsiderada em métodos deste tipo e caracterizada por ter a derivada no infinito finita e ser duas vezes continuamente diferenciável. Além disso mantém na origem a derivada igual a μ , o multiplicador de lagrange, que constitui uma característica fundamental nesta metodologia. Esta mudança também é feita para penalidades do tipo II. Em ambos casos pedimos que suas correspondentes funções conjugadas estejam definidas na origem e damos exemplos. Um interessante fato é que ambas conduzem, via dualidade de Fenchel, a distâncias de Bregman no dual. Com isto se garante a convergência da seqüência de multiplicadores a uma solução ótima do problema dual de (P) . Uma prova de convergência ergódica para a seqüência primal que envolve os *shifts* considerados é mostrada.

Notação

Usaremos a seguinte notação nos próximos capítulos.

$\text{int } S$	Interior do conjunto S .
$\text{ri int } S$	Interior relativo do conjunto S .
\bar{S}	Fecho do conjunto S .
∂S	Fronteira do conjunto S .
$\text{dom } f$	Dominio da função f .
D_h	Distância de Bregman induzida por h .
d_φ	φ -divergência induzida por φ .
$y^+(\cdot)$	Função de penalidade exata $y \in R \mapsto y^+(y) = \max\{0, y\}$.
f'_∞	Função de recessão na direção d .
f^*	Conjugada da função f .
$I_S(\cdot)$	Função indicadora do conjunto S .
e	Vetor com todas as componentes iguais a 1.
$\ \cdot\ $	Norma Euclidiana de R^n .
R_+^n	Denota ao conjunto $\{x \in R^n \mid x \geq 0\}$.
R_{++}^n	Denota ao conjunto $\{x \in R^n \mid x > 0\}$.
$y \downarrow 0$	Significa que $y \rightarrow 0$ em R_{++}^n .

Capítulo 2

Métodos de Penalização

O objetivo principal deste capítulo é apresentar um método para resolver problemas de programação não-linear nãoconvexa diferenciável com restrições de desigualdade usando funções de penalidade generalizadas parametrizadas. O problema de interesse ao longo da presente tese é

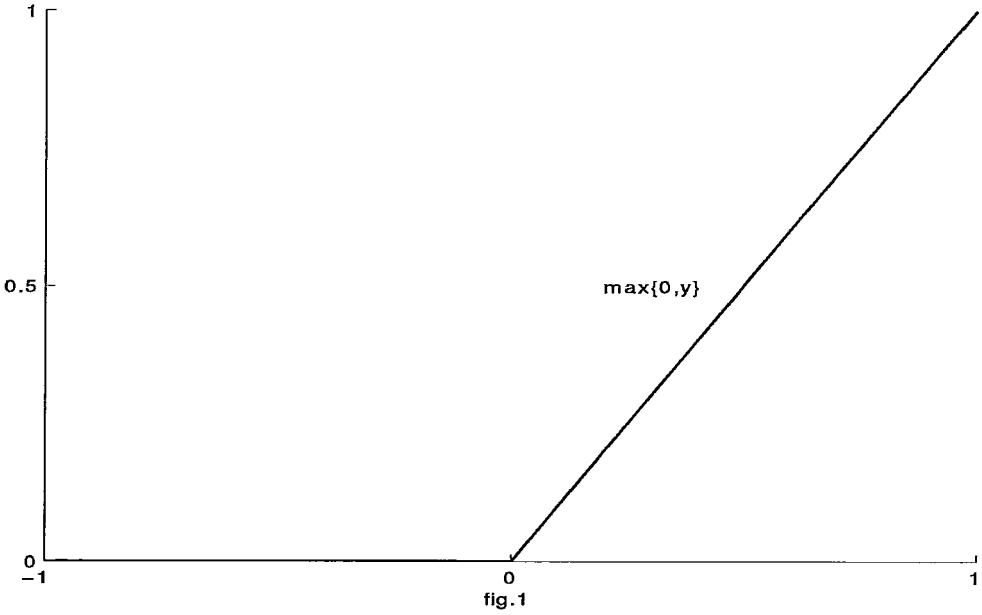
$$\begin{aligned} & \text{minimizar } f_0(x) \\ (P) \quad & \text{sujeito a } f_i(x) \leq 0 \\ & x \in R^n \end{aligned}$$

onde $f_i : R^n \rightarrow R$ para $i = 0, 1, \dots, m$ são funções continuamente diferenciáveis. Introduzimos brevemente os métodos de penalidade clássicos e duas importantes famílias de penalidades que usaremos no transcurso da presente tese.

Uma função de penalidade bem conhecida e com destaque é a função de penalidade exata ou função de penalidade l_1 , mostrada na figura fig.1 e definida como:

$$y \in R \mapsto y^+(y) = \max\{0, y\}. \quad (2.1)$$

Onde nãohouver confusão possível faz-se o abuso de notação $y^+(y) = y^+$.



Apresentamos um esquema para aproximar esta função de penalidade exata construindo funções do tipo

$$y \in R \mapsto \theta_r(y) = r\theta\left(\frac{y}{r}\right) \quad (2.2)$$

onde $r \in (0, 1]$, $\theta : R \rightarrow R$ é convexa, estritamente crescente, de classe C^2 e pertence a uma das famílias de penalidades introduzidas.

Mostramos diferentes técnicas de suavização conhecidas que aproximam a função de penalidade exata e que são usadas para resolver uma variedade de problemas em otimização. ressaltamos que estas técnicas seguem o esquema (2.2) e algumas das penalidades usadas apresentam a desvantagem de não ser duas vezes diferenciável.

Finalmente é apresentado um novo algoritmo que sob hipóteses razoáveis resolve o problema (P) através de minimizações sobre R^n da função critério

$$f_0(x) + \beta r \sum_{i=1}^m \theta\left(\frac{f_i(x)}{r}\right)$$

onde $\beta \geq 1$ e $r \in (0, 1]$ são parâmetros e θ é uma função de penalidade apresentada na seção (2.2.1). Observe a presença da aproximação (2.2) da penalidade exata na função critério. Uma análise de convergência do algoritmo proposto é mostrada.

2.1 Métodos de Penalidade

Consideremos sob hipóteses gerais o problema de programação não-linear:

$$\begin{aligned} & \text{minimizar } f_0(x) \\ & \text{sujeito a } x \in S \end{aligned} \tag{2.3}$$

onde $f : R^n \rightarrow R$ é continuamente diferenciável e $S \subset R^n$ é o conjunto de pontos viáveis com $\text{int}S \neq \emptyset$.

Existem dois bem conhecidos grupos de métodos clássicos a saber: métodos de penalização externa, os quais impõem um *peso* ou penalidade por pontos que não sejam viáveis, e os métodos de penalização interna ou de barreira em que se penaliza para manter-se longe da fronteira de S mas sempre no seu interior até atingir a otimalidade.

Ambos tipos de métodos resolvem o problema (2.3) através de uma seqüência de problemas irrestritos nos quais se adiciona à função objetivo uma função auxiliar ou de penalização gerando uma seqüência de pontos dos quais se espera que convirjam a uma solução ótima do problema original.

2.1.1 Método de Barreira

Considere o problema

$$\begin{aligned} & \text{minimizar } f_0(x) \\ (P) \quad & \text{sujeito a } f_i(x) \leq 0 \\ & x \in R^n \end{aligned}$$

onde $f_i : R^n \rightarrow R$ para $i = 0, 1, \dots, m$ são funções continuamente diferenciáveis. O conjunto viável é $S = \{x \in R^n / f_i(x) \leq 0 \text{ para } i = 1, \dots, m\}$. Supõe-se que o interior de S , $\text{int}S = \{x \in S / f_i(x) < 0 \text{ para } i = 1, \dots, m\} \neq \emptyset$

A estratégia a seguir é resolver iterativamente uma seqüência de problemas irrestritos da forma :

$$\text{minimizar}_{x \in \text{int}S} F_{r_k}(x)$$

com

$$F_{r_k}(x) = f_0(x) + r_k P(x)$$

onde usualmente definimos

$$x \in \text{int } S \mapsto P(x) = \sum_{i=1}^m -\log(-f_i(x))$$

ou

$$x \in \text{int } S \mapsto P(x) = \sum_{i=1}^m \frac{1}{f_i(x)}$$

e $\{r_k\}$ define uma seqüência de parâmetros positivos que decrescem monotonicamente a zero.

Sob condições bastante razoáveis, este método gera uma seqüência de pontos estritamente viáveis que convergem a uma solução ótima do problema através do interior (é por esta razão que são chamados de métodos de pontos interiores) criando uma barreira que mantém os iterados afastados da região inviável. Observe que quando $f(x)$ se aproxima da fronteira de S , $P(x) \rightarrow \infty$. O efeito da barreira é diminuído quando diminuimos o parâmetro r_k e assim os iterados podem gradualmente aproximar-se da fronteira da região viável. Um problema presente nesta metodologia é que na medida que r_k decresce, as minimizações tornam-se mais difíceis de resolver, criando problemas de mal condicionamento. Geralmente precisa-se de uma rotina inicial que forneça um ponto inicial viável. Métodos de barreira não serão mais discutidos neste capítulo.

2.1.2 Método de penalização externa

Métodos de penalização externa resolvem uma seqüência de problemas irrestritos cujas soluções usualmente são inviáveis ao problema original restrito, iterativamente a penalidade é incrementada por violação de cada restrição forçando aos iterados aproximar-se da região viável.

Considere o problema

$$(P1) \quad \begin{array}{ll} \text{minimizar} & f_0(x) \\ \text{sujeito a} & f_i(x) = 0 \\ & x \in R^n \end{array}$$

onde $f_i : R^n \rightarrow R$ para $i = 0, 1, \dots, m$ são funções continuamente diferenciáveis.

A função de penalidade clássica para este problema é a penalidade quadrática dada por

$$x \in R^n \mapsto P(x) = \frac{1}{2} \sum_{i=1}^m (f_i(x))^2. \quad (2.4)$$

A idéia inicial de penalizar com esta função tem sido generalizada a uma ampla classe de funções penalizadoras, ver [8].

Resolvemos então uma seqüência de problemas irrestritos da forma

$$\underset{x \in R^n}{\text{minimizar}} F_{\beta_k}(x)$$

com

$$F_{\beta_k}(x) = f(x) + \beta_k P(x)$$

onde $P(\cdot)$ é a função penalizadora (2.4) e $\{\beta_k\}$ define uma seqüência de parâmetros que crescem indefinidamente.

A função penalizadora depende da forma das restrições do problema . Para o nosso problema de interesse (P), a penalidade quadrática não é útil pois penaliza tanto pontos viáveis como inviáveis. Assim em geral, os problemas com desigualdades precisam de um tipo de função penalizadora diferente dos problemas com igualdades, isto é, precisamos de uma função que penalize somente pontos inviáveis. Neste sentido, a função de penalidade mais conhecida é a função de penalidade exata dada por

$$x \in R^n \mapsto P(x) = \sum_{i=1}^m y^+(f_i(x)) = \sum_{i=1}^m \max \{0, f_i(x)\}. \quad (2.5)$$

A função penalizada ou função criterio para os subproblemas é construída como

$$x \in R^n \mapsto f_{\beta_k}(x) = f_o(x) + \beta_k \sum_{i=1}^n (f_i(x))^+,$$

onde $\{\beta_k\}$ é uma seqüência crescente e positiva de parâmetros de penalização.

Esta função é bem conhecida, ver por exemplo Zangwill [64]. Se a mesma tem um minimizador para algum $\beta_0 > 0$, então também será verdade para todo $\beta_k > \beta_0$. Sob hipóteses razoáveis , é conhecido que se β_k é suficientemente grande, então os minimizadores de $f_{\beta_k}(\cdot)$ serão viáveis e conseqüentemente ótimos, isto indica que a seqüência de parâmetros $\{\beta_k\}$ não precisa crescer infinitamente, o

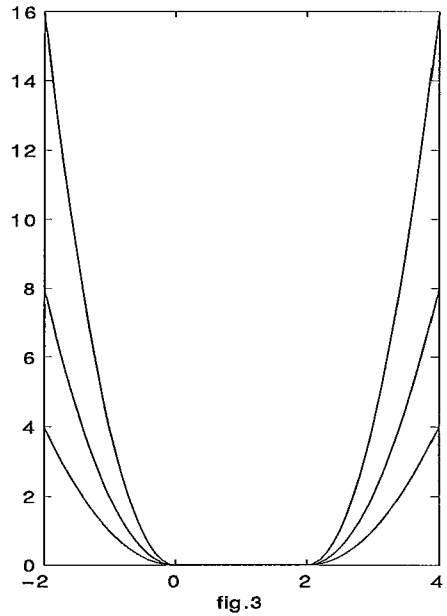
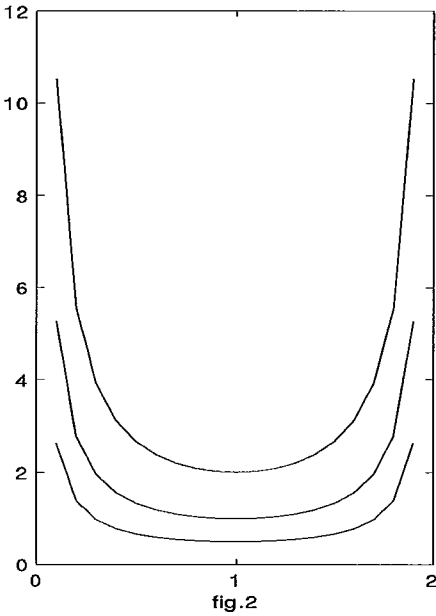
que evita, ao menos teoricamente, o problema de mal condicionamento. Assim podemos encontrar uma solução do problema original com um valor finito do parâmetro. No obstante, a principal desvantagem da função de penalidade exata consiste em que não é diferenciável dificultando assim o uso de eficientes métodos computacionais para resolver os subproblemas.

Para superar esta dificuldade tem-se proposto como penalidade a função

$$x \in R^n \mapsto P(x) = \frac{1}{2} \sum_{i=1}^m (\max \{0, f_i(x)\})^2.$$

Esta função é diferenciável e penaliza somente pontos inviáveis mas persiste o problema de mal condicionamento pois é necessário que $\beta_k \rightarrow \infty$.

A seguir mostra-se o comportamento geométrico das penalidades interna e externa aplicadas às restrições dadas por $f_1(x) = -x$ e $f_2(x) = x - 2$. Na fig.2 ilustra-se a função barreira $P(x) = \sum_{i=1}^2 \frac{1}{f_i(x)}$ e na fig.3 a penalidade quadrática $P(x) = \frac{1}{2} \sum_{i=1}^2 (\max \{0, f_i(x)\})^2$.



Na seção (2.7) apresentamos uma nova abordagem que usa aproximações suaves da função de penalidade exata que sob hipóteses razoáveis resolve o problema (P) sem precisar que $\beta_k \rightarrow \infty$. Este esquema gera tanto pontos viáveis como inviáveis e a partir de um valor suficientemente grande de β_k os iterados permanecem viáveis.

2.2 Penalidades generalizadas

Descrevemos dois tipos de funções de penalidade que mantém uma interessante relação de dualidade com outros tipos de penalidades que apresentamos no capítulo 3. Usaremos alguns conceitos básicos da análise convexa que podem ser encontrados nas referências [28], [29] e [54].

Para $f : R^n \rightarrow R \cup \{+\infty\}$ convexa própria fechada definimos sua função de recessão (ver pag. 179 em [28]) como

$$d \in R^n \mapsto f'_\infty(d) = \lim_{\lambda \rightarrow +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda} \quad (2.6)$$

onde x é arbitrário em $\text{dom} f$.

Consideremos $\theta : R \rightarrow R \cup \{+\infty\}$ convexa, própria, fechada estritamente crescente e de classe C^2 no int $\text{dom} \theta$. Classificaremos as penalidades em dois tipos:

2.2.1 Penalidade Tipo I

Este tipo de funções satisfazem

- a) $\text{dom} \theta = R$.
- b) $\lim_{y \rightarrow -\infty} \theta(y) = 0$.
- c) $\theta'_\infty(1) = 1$.

Alguns exemplos de funções deste tipo são:

$$\begin{aligned} \theta_1(y) &= \frac{1}{2}(y + \sqrt{y^2 + 4}) \\ \theta_2(y) &= \ln(1 + e^y) \\ \theta_3(y) &= \begin{cases} \frac{1}{2}e^y & \text{se } y \leq 0 \\ y + \frac{1}{2}e^{-y} & \text{se } y \geq 0 \end{cases} \end{aligned}$$

Resaltamos que a família de penalidades tipo I pode ser construída mediante dupla integração de funções de densidade de probabilidade, ver [14] e [15], fato que será abordado na seção 2.6. Auslender [3] usou funções semelhantes às do tipo I dentro de um esquema de penalidade.

2.2.2 Penalidade Tipo II

Esta família satisfaz

$$\text{a) } \text{dom } \theta = (-\infty, b), \quad 0 < b \leq +\infty.$$

$$\text{b) } \theta'_{\infty}(-1) = 0.$$

$$\text{b) } \theta'_{\infty}(1) = +\infty.$$

Exemplos:

$$\theta_4(y) = e^y - 1.$$

$$\theta_5(y) = -\log(1 - y) \text{ com } -\infty < y < 1.$$

Estes tipos de funções aparecem ao longo da tese e serão usados no capítulo 5 para definir dois tipos de métodos de Lagrangeano aumentado. A idéia fundamental está no fato de que elas podem ser usadas para aproximar a função de penalidade exata definida em (2.5) no caso da penalidade tipo I e no caso da penalidade tipo II a função indicadora de R_- , isto é, à função

$$y \in \mathbb{R} \mapsto I_{(R_-)}(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ +\infty & \text{se } y > 0. \end{cases} \quad (2.7)$$

Aproximações a função indicadora de R_- foram abordadas paralelamente por Auslender em [2].

2.3 Suavização usando penalidades generalizadas

A seguir apresentamos um esquema para aproximar a função de penalidade exata y^+ dada em (2.1) e a função indicadora do ortante negativo definida em (2.7) mediante a parametrização de funções de penalidades do tipo I e do tipo II respectivamente.

2.3.1 Suavização da penalidade exata

Seja θ uma função de penalidade do tipo I com $0 \in \text{dom } \theta$ e considere o parâmetro $r \in (0, 1]$. Definimos a função

$$y \in \mathbb{R} \mapsto \theta_r(y) = r\theta(y/r). \quad (2.8)$$

Note que para cada $r \in (0, 1]$, $\theta_r(\cdot)$ satisfaz as propriedades a),b) e c) para as penalidades tipo I.

A aproximação está determinada pela proposição a seguir.

Proposição 2.1 $\theta_r(\cdot)$ converge pontualmente a $y^+(\cdot)$ quando $r \rightarrow 0^+$.

Prova:

Para $y \in \mathbb{R}$ fixo, temos usando (2.8) e observando que $0 \in \text{dom } \theta$,

$$\lim_{r \rightarrow 0^+} \theta_r(y) = \lim_{r \rightarrow 0^+} r\theta(y/r) = \theta'_\infty(y).$$

Se $y < 0$, então claramente fazendo $u = \frac{y}{r}$ temos

$$\lim_{r \rightarrow 0^+} \theta\left(\frac{y}{r}\right) = \lim_{u \rightarrow -\infty} \theta(u) = 0 = y^+.$$

Se $y \geq 0$, então

$$\lim_{r \rightarrow 0^+} \theta_r(y) = y\theta'_\infty(1) = y = y^+.$$

Isto completa a prova.

O interesse deste capítulo é usar esta aproximação como uma penalidade num processo iterativo para resolver o problema (P) através da função critério

$$f_0(x) + \beta r \sum_{i=1}^m \theta\left(\frac{f_i(x)}{r}\right)$$

onde $\beta \geq 1$ é um parâmetro responsável pelo aumento da penalidade. Isto é feito na seção (2.7).

2.3.2 Suavização da função indicadora do ortante negativo

Considerando θ como uma função de penalidade do tipo II com $0 \in \text{dom } \theta$, procedemos em forma semelhante ao caso anterior. Considere o parâmetro $r \in (0, 1]$ e definamos a função

$$y \in \mathbb{R} \mapsto \theta_r(y) = r\theta(y/r). \quad (2.9)$$

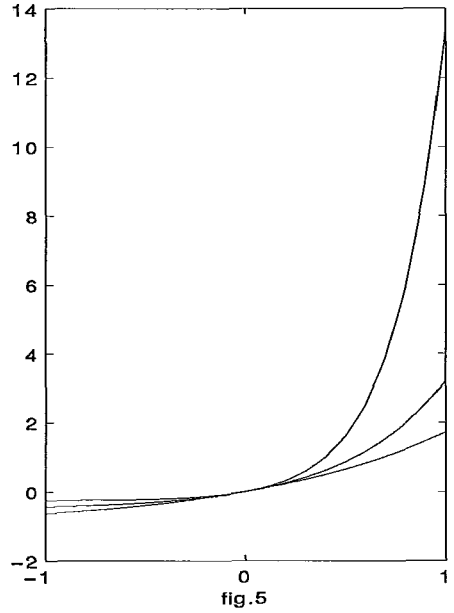
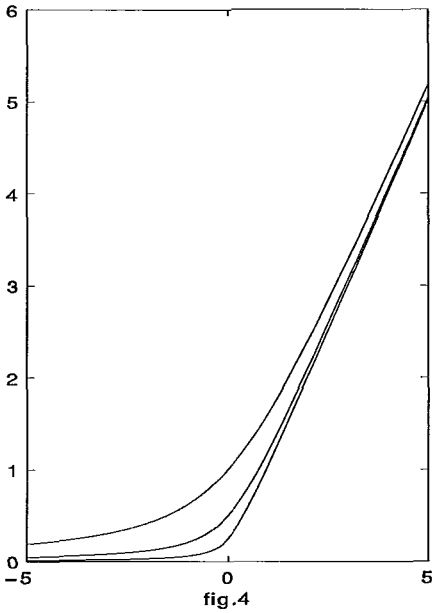
Notemos de novo que para cada $r \in (0, 1]$, $\theta_r(\cdot)$ satisfaz as propriedades a),b) e c) estabelecidas para as penalidades tipo II.

Enunciamos sem demonstração a proposição a seguir por ser semelhante à proposição 2.1.

Proposição 2.2 $\theta_r(\cdot)$ definida em (2.9) converge pontualmente à função indicadora definida em (2.7) quando $r \rightarrow 0^+$.

Como observação destacamos que a convergência formulada é uniforme no caso em que θ seja limitada por baixo.

Na fig.4 ilustramos o comportamento geométrico do esquema de suavização para aproximar a penalidade exata na medida que r decresce. Na fig.5 se aproxima a função indicadora do ortante positivo.



2.4 Esquema geral de minimização

Apresentamos um esquema iterativo geral de minimização que será seguido pelos métodos de penalidade considerados para resolver principalmente o problema com desigualdades (P). Incluem-se neste enfoque os métodos de Lagrangeano aumentado, que podem ser considerados como um caso especial de penalidade com atualização dos multiplicadores e serão discutidos no capítulo 4. A seguir apresentamos uma descrição do esquema iterativo que será usado na tese. Aqui as funções de penalidade são escolhidas a cada iteração e podem ser distintas para diferentes restrições.

Esquema iterativo.

Entrada:

Um ponto inicial x^0 e parâmetros iniciais.

Inicialização:

Faça $k = 0$.

Iteração k:

Escolha a penalidade θ_i^k para $i = 1, \dots, m$, $k = 0, 1, 2, \dots$ e resolva:

$$\text{minimizar}_x \{f_0(x) + \sum_{i=1}^m \theta_i^k(f_i(x))\}$$

para obter x^{k+1} .

Atualização:

Atualização de parâmetros usados, $k = k + 1$.

Comentário:

Os parâmetros iniciais e as atualizações podem variar dependendo da metodologia usada, por exemplo, em métodos de penalidade poderíamos trabalhar para todo i com a mesma penalidade da forma $\theta_i^k(y) = \beta^k \theta(y)$ ou $\theta_i^k(y) = \beta^k r^k \theta(\frac{y}{r^k})$ onde para todo k , β^k e r^k são reais positivos e θ é uma conveniente função de penalidade. De forma similar, para métodos de Lagrangeano aumentado poderíamos usar para todo i , por exemplo, $\theta_i^k(y) = r^k \mu_i^k \theta(\frac{y}{r^k})$ onde μ é o vetor de multiplicadores de Lagrange. Este esquema geral de minimização pode ser usado para formulações primais e para suas respectivas formulações duais.

2.5 Técnicas de Suavização

Um dos problemas surgidos com a função de penalidade exata é sua não diferenciabilidade, o que impossibilita a aplicação de eficientes métodos computacionais. Uma das idéias usadas para enfrentar esta dificuldade, consiste em substituir a função de penalidade exata (ou problema considerado) por uma aproximação suave (ou problema perturbado) controlada por parâmetros que determinam a precisão da aproximação. Esta aproximação suave permite o uso de eficientes algoritmos em otimização podendo ser aplicada em diversos tipos de problemas entre

os que mencionamos o problema min-max, problemas que incluem aproximações da norma l_1 , problemas de estruturas em redes, etc.

A seguir, introduzimos as principais e recentes técnicas de suavização conhecidas para problemas de otimização diferenciável e não diferenciável baseadas em aproximações da penalidade exata.

2.5.1 Primeira Técnica

Esta foi apresentada por Zang [63] para resolver basicamente o problema min-max definido em (2.10). Será evidente o uso de uma aproximação da função de penalidade exata.

O problema min-max discreto-contínuo é definido como:

$$\text{minimizar}_{x \in R^n} F(x) \tag{2.10}$$

onde

$$x \in R^n \mapsto F(x) = \max \{f_i(x)\}, \quad i = 1, \dots, m \tag{2.11}$$

com $f_i : R^n \rightarrow R$ $i = 1, \dots, m$ funções pelo menos suaves.

Bertsekas [8] escreveu a expressão (2.11) como uma composição da função barreira exata $y^+(y) = \max \{0, y\}$ e as funções $f_i(x)$, $i = 1, \dots, m$.

Assim escrevemos

$$F(x) = f_1(x) + y^+[f_2(x) - f_1(x) + y^+[f_3(x) - f_2(x) + \dots + y^+[f_m(x) - f_{m-1}(x)] \dots]]. \tag{2.12}$$

A diferença de Bertsekas, Zang usou uma suavização local da função $y^+(\cdot)$ introduzida na função F e controlada por um parâmetro r . Como se mostra na figura fig.5, y^+ é substituída na vizinhança $-r \leq y \leq r$ por uma função $b_r(y)$. Fora deste intervalo F não muda. Na medida que r é reduzido a zero, também diminui a região onde y^+ é aproximada e assim temos uma boa aproximação de F . A função que suaviza $y^+(\cdot)$ é dada por

$$y_r^+(y) = \begin{cases} 0 & \text{se } y \leq -r \\ b_r(y) & \text{se } -r \leq y \leq r \\ y & \text{se } y \geq r \end{cases}$$

onde $b_r(y)$ é uma função monótona crescente. A aproximação da F é dada por

$$\tilde{F}_r(x) = f_1(x) + y_r^+[f_2(x) - f_1(x) + y_r^+[f_3(x) - f_2(x) + \dots + y_r^+[f_m(x) - f_{m-1}(x)] \dots]].$$

Algumas expressões usadas para b são:

a) $b_r(y) = \frac{1}{4r}y^2 + \frac{1}{2}y + \frac{1}{4}r.$

b) $b_r(y) = -\frac{1}{16r^3}y^4 + \frac{3}{8r}y^2 + \frac{1}{2}y + \frac{3}{16}r.$

Uma observação importante é que estas aproximações podem-se expressar na forma *padrão* (2.8) correspondente às penalidades do tipo I, assim

Para a penalidade em a) temos

$$\begin{aligned} b_r(y) &= r \left(\frac{1}{4} \left(\frac{y}{r} \right)^2 + \frac{1}{2} \left(\frac{y}{r} \right) + \frac{1}{4} \right) \\ &= r \hat{b} \left(\frac{y}{r} \right) \end{aligned}$$

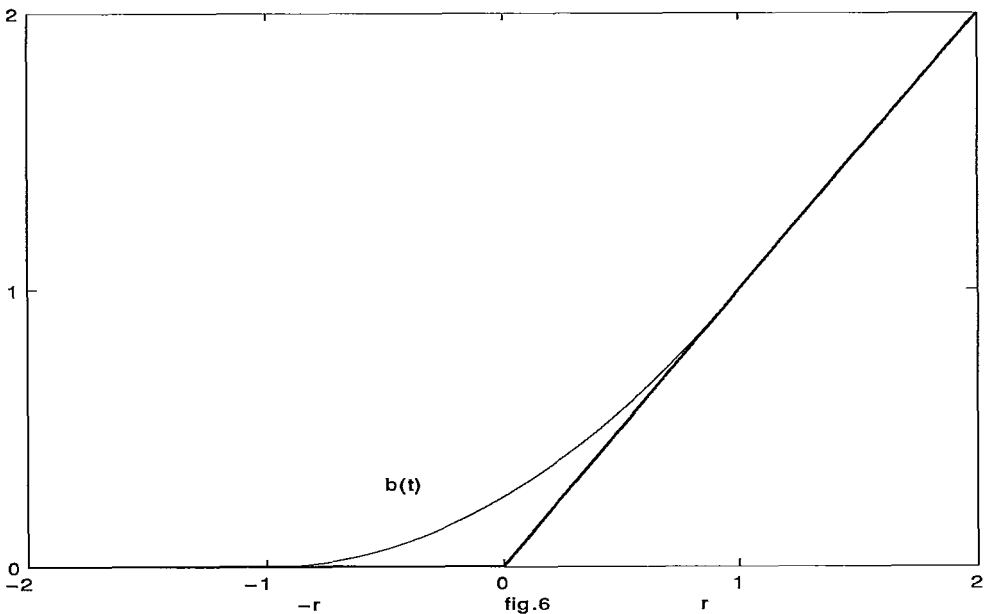
onde $\hat{b}(y) = \frac{1}{4}y^2 + \frac{1}{2}y + \frac{1}{4}.$

Em forma análoga, para a penalidade em b) temos

$$\begin{aligned} b_r(y) &= -r \left(\frac{1}{16} \left(\frac{y}{r} \right)^4 + \frac{3}{8} \left(\frac{y}{r} \right)^2 + \frac{1}{2} \frac{y}{r} + \frac{3}{16} \right) \\ &= r \hat{b} \left(\frac{y}{r} \right) \end{aligned}$$

onde $\hat{b}(y) = -\frac{1}{16}y^4 + \frac{3}{8}y^2 + \frac{1}{2}y + \frac{3}{16}.$

Podemos visualizar a função y^+ e a sua aproximação $b_r(y)$ para $r = 1$ na figura a seguir.



Supondo que $y_r^+(\cdot)$ é convexa e pelo menos de classe C^1 com respeito a y , pode-se provar que:

$$\lim_{r \rightarrow 0} y_r^+(y) = y^+$$

o que constitui uma propriedade desejável. Assim, controlando o parâmetro r fazemos $\tilde{F}_r(x)$ muito próxima a $F(x)$.

Por exemplo, para $m=2$ e usando (2.12) temos

$$F(x) = f_1(x) + y^+[f_2(x) - f_1(x)]$$

e assim

$$F_r(x) = f_1(x) + y_r^+[f_2(x) - f_1(x)]$$

$$= \begin{cases} f_1(x) & \text{se } f_2(x) - f_1(x) \leq -r \\ \frac{1}{4} \frac{[f_2(x) - f_1(x)]}{r} + \frac{1}{2} [f_2(x) + f_1(x)] + \frac{1}{4} r & \text{se } -r \leq f_2(x) - f_1(x) \leq r \\ f_2(x) & \text{se } r \leq f_2(x) - f_1(x) \end{cases}$$

Note que $F(x)$ é substituída por uma aproximação somente na região que satisfaz $-r \leq f_2(x) - f_1(x) \leq r$.

Esta aproximação foi usada por Tishler e Zang [60] para resolver alguns problemas de estimação por regressão onde aparecem operadores do tipo *max*.

Dentro de um esquema algorítmico, a idéia é começar com um ponto inicial $x_0 \in R^n$ e diminuir iterativamente r_k a zero para $k = 1, 2, \dots$ minimizando sobre R^n a função $\tilde{F}_{r_k}(x)$ obtendo a seqüência $\{x_k\}$ de minimizadores.

Detalhes sobre este enfoque encontram-se em [63], mas podemos mencionar os principais resultados:

Teorema 2.3 (ver Zang [63]) *Seja $\{r_k\} \rightarrow 0$ uma seqüência de números positivos e assumamos que x_k é uma solução de minimizar $\tilde{F}_{r_k}(x)$. Seja $\bar{x} \in R^n$ um ponto limite da seqüência $\{x_k\}$, então $F(\bar{x}) = F(x^*)$ com x^* ponto de mínimo da F .*

Teorema 2.4 (ver Zang [63]) *Seja x^* um minimizador local de $F(x)$, com $F(x^*) = f_l(x^*)$ para um único valor de $1 \leq l \leq m$. Então existe um número real positivo \bar{r} tal que x^* é um mínimo local de $\tilde{F}_r(x)$ para r satisfazendo $0 < r < \bar{r}$.*

2.5.2 Segunda Técnica

A. Ben-Tal e M. Teboulle [7] introduziram uma técnica de suavização para problemas de otimização não diferenciáveis fazendo ênfase no problema de minimizar a norma l_1 e no problema min-max discreto-contínuo definido em (2.10). A idéia consiste em substituir o problema original por aproximações suaves que controladas por um parâmetro determinam a precisão da aproximação.

Considere o problema

$$\inf_{x \in \mathbb{R}^n} F'_\infty(f_1(x), f_2(x), \dots, f_m(x)) \quad (2.13)$$

onde F'_∞ é a função de recessão de $\theta : \mathbb{R}^m \rightarrow \mathbb{R}$ convexa própria fechada com $0 \in \text{dom } \theta$ e $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$, para $i = 1, \dots, m$.

Assim, definimos

$$y \in \mathbb{R}^m \mapsto F_r(y) = r \theta\left(\frac{y}{r}\right).$$

Observe que

$$\lim_{r \rightarrow 0} F_r(y) = \lim_{r \rightarrow 0} r \theta\left(\frac{y}{r}\right) = F'_\infty(y) \quad \forall y \in \mathbb{R}^m$$

(ver [54] Corolário 8.5.2)

logo podemos aproximar o problema (2.13) através de F_r usando o problema perturbado

$$\inf_{x \in \mathbb{R}^n} \left\{ r \theta\left(\frac{f_1(x)}{r}, \dots, \frac{f_m(x)}{r}\right) \right\}$$

com a vantagem que freqüentemente θ é uma função suave e assim são suas aproximações.

Exemplos de aplicação desta abordagem são:

a) Problema de minimização da norma l_1 .

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m |f_i(x)|.$$

Aqui $F'_\infty(y) = \sum_{i=1}^m |y_i|$ é a função de recessão de $\theta(y) = \sum_{i=1}^m \sqrt{1 + y_i^2}$ que é suave.

Assim resolvemos o problema suavizado

$$\min_{x \in R^n} \sum_{i=1}^m \sqrt{(f_i(x))^2 + r^2}.$$

Esta última aproximação foi sugerida por El-Attar et al. (ver referência em [7]).

b) Problema min-max discreto-contínuo.

Considere novamente o problema

$$\text{minimizar}_{x \in R^n} F(x)$$

onde

$$x \in R^n \mapsto F(x) = \max \{f_i(x)\}, \quad i = 1, \dots, m$$

com $f_i : R^n \rightarrow R$ $i = 1, \dots, m$ funções pelo menos suaves.

Temos que $F'_\infty(y) = \max_{1 \leq i \leq m} y_i$, é a função de recessão de

$$y \in R^m \mapsto \theta(y) = \log \sum_{i=1}^m e^{(y_i)}.$$

O problema suavizado resultante é

$$\min_{x \in R^n} r \log \sum_{i=1}^m e^{\left(\frac{f_i(x)}{r}\right)}.$$

Aproximações semelhantes para esta abordagem foram consideradas em [9] e [51] no contexto de métodos de multiplicadores.

Finalmente mostra-se que se x^* e x_r^* são respectivamente soluções ótimas do problema original e do problema perturbado, e se

$$\forall r > 0, \forall z, r \theta\left(\frac{z}{r}\right) \geq \theta_\infty(z),$$

então se satisfaz para $r > 0$

$$0 \leq F_r(x_r^*) - F'_\infty(x^*) \leq r \theta(0).$$

Isto constitui um limitante para o erro na solução.

2.5.3 Terceira Técnica

Esta consiste basicamente em aproximar a função de penalidade l_1 mediante uma função linear-quadrática suave controlada por um parâmetro. Esta aproximação

foi apresentada por M. Pinar e S. Zenios [49] e distintamente da usada por Zang, não é simétrica ao redor da quina.

A r -suavização é dada por

$$y_r^+(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ \frac{y^2}{2r} & \text{se } 0 \leq y \leq r \\ y - \frac{r}{2} & \text{se } y \geq r. \end{cases} \quad (2.14)$$

Esta aproximação é conhecida em estatística computacional como estimador de Huber (ver referência em [49]). Madsen K. e Nielsen B. [44] usaram uma versão simétrica desta função linear-quadrática para desenvolver um algoritmo finito aplicado a problemas de regressão linear que envolvem minimizações da norma l_1 . Funções lineares-quadráticas também são usadas em estatística para desenvolver procedimentos de estimação robustos.

Observe além disso que esta penalidade mantém o formato padrão dado por 2.8 pois:

$$y_r^+(y) = r\hat{y}\left(\frac{y}{r}\right)$$

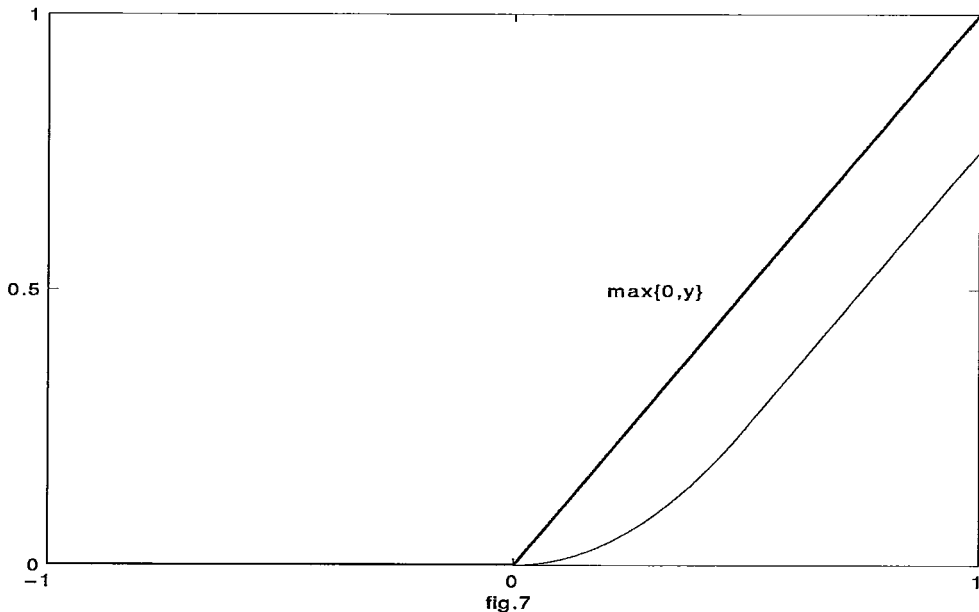
onde

$$\hat{y}(y) = \begin{cases} 0 & \text{se } y \leq 0 \\ \frac{y^2}{2} & \text{se } 0 \leq y \leq r \\ y - \frac{1}{2} & \text{se } y \geq r. \end{cases} \quad (2.15)$$

Prova-se que :

$$\lim_{r \rightarrow 0} y_r^+(y) = y^+$$

A figura fig.7 mostra esta função de aproximação para $r = \frac{1}{2}$.



Prova-se também que existe uma limitante superior para o erro de aproximação em função dos parâmetros de penalização. Desenvolve-se um algoritmo simples com resultados computacionais satisfatórios aplicados a problemas de programação não linear convexos de classe C^1 . Também é aplicado ao problema min-max discreto-contínuo previamente transformado em um problema não linear convexo. Reportam-se experiências computacionais interessantes na solução de problemas de fluxos em redes de grande porte.

2.5.4 Quarta Técnica

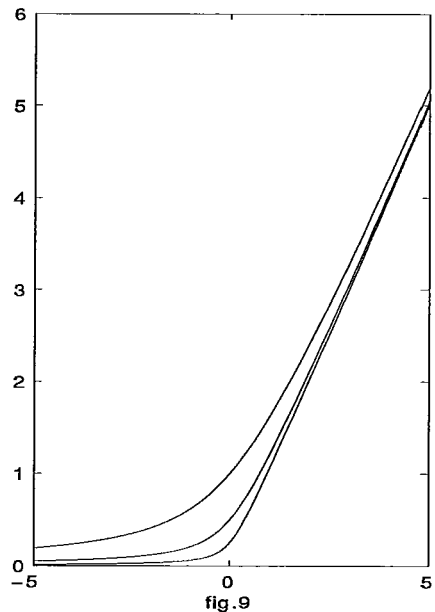
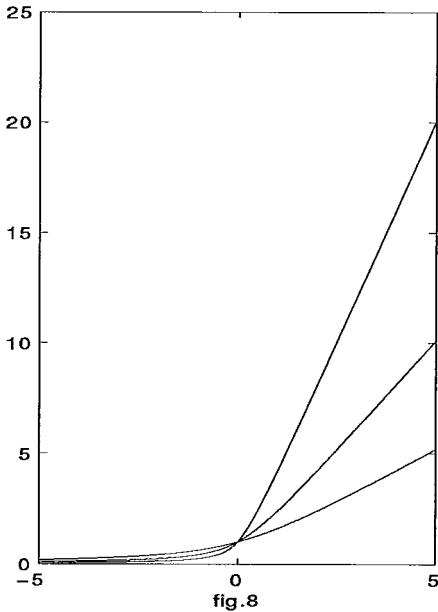
O método de penalização hiperbólica foi desenvolvido por Xavier [62] a partir da idéia de construir uma penalização intermediária entre a penalização exterior e a penalização interior para assim ser aplicado ao problema com desigualdades (P) via um enfoque de penalização. Usou a função de penalidade:

$$P(t, \alpha, r) = -\left(\frac{1}{2}\tan(\alpha)\right)t + \sqrt{\left(\frac{1}{2}\tan(\alpha)y\right)^2 + r^2}$$

onde $\alpha \in [0, \pi/2)$, $r \geq 0$.

Esta função corresponde à hipérbole $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ rodada ao redor da origem e está parametrizada por α e r . O parâmetro α é responsável pelo aumento da inclinação da penalidade para pontos inviáveis enquanto que r controla a precisão

da aproximação. Na fig.8 mostra-se o comportamento quando aumentamos o parâmetro α e na fig.9 quando α é fixo e r diminui. O algoritmo formulado consiste em, a partir de valores iniciais de α e r , resolver o problema irrestrito correspondente, diminuir r em cada iteração e aumentar o parâmetro α somente em iterados inviáveis.



Xavier [62] mostra teoremas referentes à existência de soluções viáveis e à convergência do algoritmo. Posteriormente faz uma ligação entre esta penalidade em particular com o método de lagrangeano aumentado.

Na próxima seção generalizamos estas idéias de suavização para resolver problemas de programação não-linear nãoconvexa diferenciável usando penalidades do tipo I. No capítulo 5 usamos estas idéias usando as penalidades tipo I e tipo II dentro de uma metodologia de lagrangeano aumentado para problemas de programação convexa usando mudanças de variáveis.

2.6 Construção de penalidades tipo I

Chen M. e Mangasarian apresentam em [14] e [15] uma interessante abordagem para construir aproximações suaves da função de penalidade exata mediante dupla integração de funções de densidade de probabilidade que são usadas para resolver problemas de complementaridade linear, não-linear e mista.

Consideremos a função de penalidade exata

$$y \in R \mapsto y^+(y) = \max\{0, y\}.$$

Note que y^+ pode-se expressar como

$$y^+(y) = \int_{-\infty}^y \sigma(x) dx$$

onde $\sigma(\cdot)$ é a função degrau definida por

$$y \in R \mapsto \sigma(y) = \begin{cases} 1 & \text{se } y > 0 \\ 0 & \text{se } y \leq 0. \end{cases}$$

Observe além disso que a função degrau pode-se expressar como

$$\sigma(y) = \int_{-\infty}^y \delta(x) dx$$

onde $\delta(\cdot)$ é a função delta de Dirac que satisfaz

a) $\delta(y) \geq 0$, para todo $y \in R$,

b) $\int_{-\infty}^{+\infty} \delta(x) dx = 1$.

Assim, a função de penalidade exata pode ser obtida por dupla integração da função delta de Dirac. Surge então a idéia de propor funções de densidade de probabilidade que aproximem a função delta de Dirac para assim obter aproximações suaves de suas integrais e conseqüentemente da função de penalidade exata. De esta forma, considere funções de densidade $d(\cdot)$ contínuas por partes tais que satisfaçam as condições a) e b) satisfeitas pela função $\delta(\cdot)$. Definimos a função de densidade parametrizada por r , como

$$d_r(y) = \frac{1}{r} d\left(\frac{y}{r}\right)$$

onde r é positivo.

Verifica-se que $d_r(y) \rightarrow \delta(y)$ quando $r \rightarrow 0$.

Isto dá origem à consideração das seguintes aproximações

$$\sigma_r(y) = \int_{-\infty}^y d_r(x) dx \approx \sigma(y)$$

e

$$y_r^+(y) = \int_{-\infty}^y \sigma_r(x) dx \approx y^+(y).$$

Desta forma podemos obter aproximações da função $y^+(\cdot)$ mediante dupla integração de funções de densidade. Em [15] mostra-se a consistência destas relações e sua aplicação a problemas de complementaridade linear, não-linear e mista. Métodos de penalização baseados em funções desta classe foram estudados em Auslender, Cominetti e Haddou [3], que descrevem algoritmos para o caso convexo.

Alguns exemplos desta abordagem são:

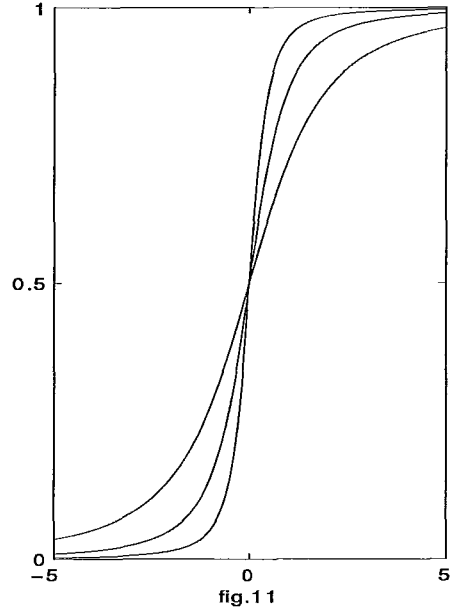
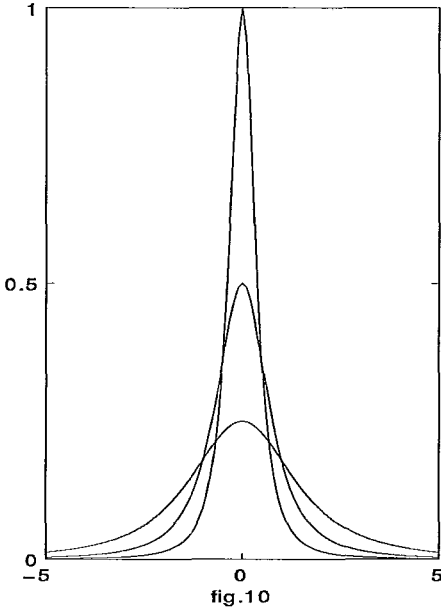
a) Para $d(y) = \frac{2}{(x^2+4)^{3/2}}$, temos

$$y_r^+(y) = \frac{1}{2}(y + \sqrt{y^2 + 4r^2}).$$

b) Para $d(y) = \frac{e^{-y}}{(1+e^{-y})^2}$, temos

$$y_r^+(y) = r \log(1 + e^{\frac{y}{r}}).$$

Note que estas aproximações correspondem às penalidades do tipo I. Na figura fig.10 mostra-se o comportamento geométrico da função de densidade $d_r(\cdot)$ para o exemplo em a). Na figura fig.11 mostra-se $\sigma_r(\cdot)$. Ambas para alguns valores do parâmetro r . O gráfico da correspondente aproximação $y_r(\cdot)$ é similar ao apresentado na fig. 9.



2.7 Método de penalidades generalizadas para programação não-linear

Esta seção constitui a parte principal deste capítulo e introduz um algoritmo para resolver problemas de programação não-linear nãoconvexa diferenciável dentro de um enfoque de penalidade usando a família de funções do tipo I que aproximam a função de penalidade exata apresentada na seção (2.2.1). Mediante a introdução dos parâmetros positivos β_k e r_k e com hipóteses usuais, o algoritmo gera uma seqüência de pontos que minimizam os subproblemas irrestritos considerados e cujos pontos limites são soluções ótimas para o problema original. O parâmetro $\beta_k \geq 1$ é responsável pelo aumento da penalidade e é incrementado somente em iterados inviáveis. O parâmetro $r_k \in (0, 1]$ diminui em cada iteração controlando a precisão da aproximação.

Auslender, Cominetti e Haddou [3] descrevem um método de penalização para problemas de programação convexa baseado em funções que aproximam a penalidade exata onde o parâmetro encarregado de aumentar a penalidade precisa crescer até infinito. Um fato importante a resaltar é que em nossa abordagem, a partir de um valor do parâmetro β_k suficientemente grande, os iterados

permanecem viáveis e assim β_k não precisa crescer indefinidamente. A metodologia apresentada generaliza o método de penalização hiperbólica desenvolvido por Xavier [62].

Considere o problema de programação não linear nãoconvexo

$$(P) \quad \begin{aligned} & \text{minimizar } f_0(x) \\ & \text{sujeito a } f_i(x) \leq 0 \\ & x \in R^n \end{aligned}$$

onde $f_i : R^n \rightarrow R$ para $i = 0, 1, \dots, m$ são funções continuamente diferenciáveis.

Para $\epsilon \geq 0$ defina o conjunto viável relaxado

$$S_\epsilon = \{x \in R^n \mid f_i(x) \leq \epsilon, i = 1, \dots, m\}.$$

Então S_0 é o conjunto viável para (P). Usaremos as seguintes hipóteses:

(H1) Existe $\epsilon > 0$ tal que S_ϵ é limitado.

(H2) Todas as soluções ótimas de (P) satisfazem a condição de qualificação de Cottle (ou Mangasarian-Fromovitz), digamos; para toda solução ótima x , existe $h \in R^n$ tal que

$$\nabla f_i(x)^T h < 0 \text{ para todo } i \in I(x) \equiv \{i = 1, \dots, m \mid f_i(x) = 0\} \text{ (ver [45]).}$$

(H3) Para todo $r \in (0, 1]$, $\beta \geq \beta^0 > 1$, suponhamos que existe *argmin* da função penalizada $f_{\beta,r}(\cdot)$ definida em (2.16).

Um ponto $x \in R^n$ será chamado *interior* se $f_i(x) < 0$ para $i = 1, \dots, m$. Observe que a condição de qualificação implica a existência de pontos interiores.

Consideremos novamente as penalidades $\theta(\cdot)$ do tipo I mostradas na seção (2.2.1) tal que

$\theta : R \rightarrow R$ é convexa, própria, fechada estritamente crescente, de classe C^2 e satisfaz

a) $\lim_{y \rightarrow -\infty} \theta(y) = 0.$

b) $\theta'_\infty(1) = 1.$

Seja $\theta(\cdot)$ a função $\theta(\cdot)$ definida anteriormente com $\theta(0)$ finito e $r \in (0, 1]$. Considere a função definida em (2.8)

$$y \in \mathbb{R} \mapsto \theta_r(y) = r\theta(y/r).$$

Para todo $r \in (0, 1]$, $\theta_r(\cdot)$ é uma penalidade do tipo I. Além disso, temos:

i) Para $y \geq 0$, $r \in (0, 1]$, $\theta'_r(y) = \theta'(y/r) \geq \theta'(y)$.

ii) Para $r \in (0, 1]$, $\theta_r(y) \geq \theta(y) - \theta(0)$.

iii) $\theta_r(\cdot)$ converge pontualmente a t^+ quando $r \rightarrow 0^+$ (ver proposição 2.1).

A propriedade (i) é imediata, e nos leva a (ii).

Exemplos específicos destas funções θ (ver [3]) são:

$$\theta_1(y) = 1/2 \left(y + \sqrt{y^2 + 4} \right).$$

$$\theta_2(y) = \ln(1 + e^y).$$

$$\theta_3(y) = \begin{cases} 1/2e^y & \text{se } y \leq 0 \\ y + 1/2e^{-y} & \text{se } y \geq 0. \end{cases}$$

2.7.1 A função penalizada

A função penalizada é construída usando $\theta_r(\cdot)$ para obter

$$f_{\beta,r}(x) = f_0(x) + \beta \sum_{i=1}^m \theta_r(f_i(x)) \quad (2.16)$$

onde $\beta \geq 1$ é o peso da penalização.

Denotemos

$$P_r(x) = \sum_{i=1}^m \theta_r(f_i(x)) = r \sum_{i=1}^m \theta \left(\frac{f_i(x)}{r} \right).$$

Note que na função penalizada estão presentes os parâmetros β e r . Quando aumentamos β , muda a inclinação da penalização $\beta \theta_r(\cdot)$. O parâmetro r controla a precisão da suavização.

O algoritmo trabalhará com β e r da seguinte forma: r decresce em cada iteração (digamos, $r := r/2$). Em pontos inviáveis β cresce (digamos, $\beta := 2\beta$); em pontos viáveis β permanece constante. O produto βr nunca cresce, o que garante a seguinte propriedade:

iv) Se βr é limitado, temos para todo ponto interior x ,

$$\lim_{r \rightarrow 0^+} \beta P_r(x) = 0.$$

Prova:

$$\lim_{r \rightarrow 0^+} \beta P_r(x) = \lim_{r \rightarrow 0^+} \beta r \sum_{i=1}^m \theta \left(\frac{f_i(x)}{r} \right),$$

para um ponto interior x , $f_i(x) < 0$ para $i = 1, \dots, m$. Assim temos

$$\lim_{r \rightarrow 0^+} \theta \left(\frac{f_i(x)}{r} \right) = 0.$$

O resultado segue-se pois βr é limitado. ■

Apresentamos um algoritmo para resolver o problema (P) baseado em funções de penalidade suaves do tipo I. Em cada iteração resolvemos um problema irrestrito obtido ao adicionar à função critério uma função de penalidade parametrizada que aproxima a penalidade exata.

2.7.2 Algoritmo.

Suponha que para $\beta \geq 1$, $r \in (0, 1]$, $f_{\beta, r}(\cdot)$ tem minimizadores. Esta suposição será discutida adiante.

Algoritmo 2.1

Dados: $\beta_0 = 1$, $r_0 = 1$. Para $k = 0, 1, \dots$

Encontre $x^k \in \operatorname{argmin}_{x \in R^n} \{f_0(x) + \beta_k \sum_{i=1}^m \theta_{r_k}(f_i(x))\}$.

$r^{k+1} := r^k/2$.

Se x^k é viável então $\beta_{k+1} := \beta_k$

senão $\beta_{k+1} := 2\beta_k$.

Este algoritmo mantém o formato do esquema geral de minimização considerado em (2.4) anteriormente com $\theta_i^k(\cdot) = \beta^k \theta_{r^k}(\cdot)$ para $i = 1, \dots, m$ e para todo $k \in N$. Neste caso a atualização dos parâmetros é dada para β e r .

As seguintes afirmações são trivialmente certas para seqüências geradas por este algoritmo:

- $r_k \rightarrow 0$, segue-se por (iii) que $\theta_{r_k} \rightarrow y^+$.
- $\beta_k r_k \leq 1$ e assim por (iv), $\beta_k P_{r_k}(\tilde{x}) \rightarrow 0$ para todo ponto interior \tilde{x} .
- Se (x^k) tem uma subsequência inviável infinita, então $\beta_k \rightarrow +\infty$.

Outro destaque é que o algoritmo não pertence à classe estudada por Auslender, Cominetti e Haddou [3]. Eles usam a mesma função penalizada, mas requerem que $\beta_k r_k \rightarrow 0$ e $\beta_k \rightarrow +\infty$. Em nosso enfoque, não precisamos da validade de este último limite, de fato, mostramos que sob a condição de qualificação de Mangasarian-Fromovitz $\{\beta_k\}$ está sempre limitado, isto é, β_k é igual a uma constante $\bar{\beta}$ a partir de certo valor de k , significando que $\beta_k P_{r_k}(\cdot)$ convergirá pontualmente a $\bar{\beta} \sum_{i=1}^m (f_i(\cdot))^+$, a penalidade exata com um valor fixo para $\bar{\beta}$.

Podemos afirmar, a partir do fato de que neste algoritmo se aproxima a função de penalidade exata usando as funções diferenciáveis θ , que estamos frente a uma forte ferramenta para tentar resolver uma considerável quantidade de problemas de interesse em otimização. Chen-Mangasarian [14, 15] construíram aproximações da função de penalidade exata para resolver problemas de complementaridade linear, não-linear e mista numa abordagem completamente diferente ao aqui apresentado.

De agora em diante assumiremos que $\{x^k\}, \{r_k\}, \{\beta_k\}$ são seqüências geradas pelo algoritmo. A geometria da função θ é semelhante ao específico caso da penalidade hiperbólica considerada por Xavier [62] parametrizada por os parâmetros α e r . De esta forma, quando β_k aumenta e r_k diminui estamos penalizando pontos inviáveis (fig.8). Na fig.9 β_k permanece fixo enquanto r_k varia (iterados viáveis) aumentando a precisão da aproximação.

2.7.3 Teoremas

Lema 2.5 *A seqüência (x^k) é limitada, e todos seus pontos de acumulação são soluções viáveis.*

Prova: Em x^0 temos por definição, para $k \in \mathbb{N}$

$$f_0(x^k) + p_{r_k}(x^k) \geq f_0(x^0) + p_{r_k}(x^0). \quad (2.17)$$

Usando (vi), $\theta(y) \leq \theta_{r_k}(y) + \theta(0)$, para todo $\theta \in \mathbb{R}$. Daqui, para $k \in \mathbb{N}$,

$$p_{r_0}(x^k) \leq p_{r_k}(x^k) + m\theta(0).$$

Usando (2.17),

$$f_0(x^k) + p_{r_k}(x^k) \geq f_0(x^0) + p_{r_0}(x^0) - m\theta(0). \quad (2.18)$$

Agora consideremos um ponto interior \tilde{x} . Por definição de x^k para $k \in \mathbb{N}$,

$$f_0(x^k) + \beta_k p_{r_k}(x^k) \leq f_0(\tilde{x}) + \beta_k p_{r_k}(\tilde{x}). \quad (2.19)$$

De (2.18) e (2.19) temos,

$$(\beta_k - 1)p_{r_k}(x^k) \leq \beta_k p_{r_k}(\tilde{x}) + K, \quad (2.20)$$

onde K é uma constante.

Se (x^k) é viável para valores grandes de k , então não temos nada que provar. Suponha que isto não é verdade. Segue-se que $\beta_k \rightarrow +\infty$. Por (iv), $\beta_k p_{r_k}(\tilde{x}) \rightarrow 0$. Daqui, $p_{r_k}(x^k) \rightarrow 0$ e conseqüentemente $\theta_{r_k}(f_i(x^k)) \rightarrow 0$ para $i = 1, \dots, m$. Isto implica que $(f_i(x^k))^+ \rightarrow 0$ porque $\theta_{r_k}(\cdot) \rightarrow t^+$ pontualmente.

— Por (H1), vemos que (x^k) é limitada, caso contrário para alguma subsequência infinita e algum $i = 1, \dots, m$, $f_i(x^k) > \epsilon$.

— Se \bar{x} é um ponto de acumulação de (x^k) , então pela continuidade das funções e do fato que $(f_i(x^k))^+ \rightarrow 0$, concluímos que $(f_i(\bar{x}))^+ = 0$, completando a prova. ■

Lema 2.6 : *Todo ponto limite de (x^k) é uma solução ótima de (P) .*

Prova: Seja \bar{x} um ponto limite de (x^k) , e seja $\mathcal{K} \subset \mathbb{N}$ tal que $x^k \xrightarrow{\mathcal{K}} \bar{x}$. Pelo lemma 2.5, \bar{x} é viável.

Consideremos agora um ponto interior arbitrário \tilde{x} . Por definição de x^k , para $k \in \mathcal{K}$,

$$f_0(x^k) \leq f_0(\tilde{x}) + \beta_k p_{r_k}(\tilde{x}) - \beta_k p_{r_k}(x^k).$$

Tomando limites para $k \in \mathcal{K}$, temos: $\beta_k p_{r_k}(\tilde{x}) \xrightarrow{\mathcal{K}} 0$, por causa de (vii), e sempre $\beta_k p_{r_k}(x^k) \geq 0$. Segue-se que

$$f_0(\bar{x}) \leq f_0(\tilde{x})$$

para todo ponto interior \tilde{x} . Pela restrição de qualificação, todas as soluções ótimas estão no fecho do conjunto de pontos interiores. Isto implica que $f_0(\bar{x}) \leq f_0(x^*)$ para toda solução ótima x^* , completando a prova. ■

Dado um ponto $x \in R^n$, definamos os conjuntos

$$I(x) = \{i = 1, \dots, m \mid f_i(x) = 0\}, I^+(x) = \{i = 1, \dots, m \mid f_i(x) \geq 0\}.$$

O seguinte lema é conhecido e apresentado em [5] sem prova.

Lema 2.7 *Se $x^k \rightarrow \bar{x}$, então para k suficientemente grande, $I^+(x^k) \subset I(\bar{x})$.*

Prova: Assuma que $x^k \rightarrow \bar{x}$, e seja $i = 1, \dots, m$ arbitrário. Devemos provar que se $i \notin I(\bar{x})$, então para k grande, $i \notin I^+(x^k)$.

Se $i \notin I(\bar{x})$, então $f_i(\bar{x}) < 0$. Então, para um k grande, e pela continuidade de $f_i(\cdot)$, $f_i(x^k) < 0$, ou equivalentemente, $i \notin I^+(x^k)$, completando a prova. ■

Teorema 2.8 *Seja (x^k) uma seqüência gerada pelo algoritmo. Então existe $\bar{k} \in \mathbb{N}$ tal que para $k \geq \bar{k}$, x^k é viável.*

Prova: Suponha por contradição que existe um conjunto infinito $\mathcal{K}_1 \subset \mathbb{N}$ tal que para $k \in \mathcal{K}_1$, x^k é inviável. Pelo lema 2.5, (x^k) é limitada, segue-se $(x^k)_{k \in \mathcal{K}_1}$ tem um ponto de acumulação x^* , o qual é ótimo pelo lema 2.6.

Seja $\mathcal{K}_2 \subset \mathcal{K}_1$ tal que $x^k \xrightarrow{\mathcal{K}_2} x^*$.

Associe com cada $k \in \mathbb{N}$ o conjunto $I^+(x^k)$ definido acima, as seguintes três afirmações são certas para $k \in \mathcal{K}_2$ grande:

- a) $I^+(x^k) \subset I(x^*)$, pelo lema 2.7.
- b) Para $i \in I(x^*)$, $\nabla f_i(x^k)^T h \leq -\rho/2$, onde $h \in R^n$ é dada pela restrição de qualificação. Isto segue da continuidade dos gradientes e do fato que $x^k \xrightarrow{\mathcal{K}_2} x^*$.
- c) Para $i \notin I(x^*)$, $f_i(x^k) \leq -\mu < 0$, onde $\mu > 0$ é tal que

$$-\mu > \max\{f_i(x^*) \mid i \notin I(x^*)\}.$$

Defina o conjunto $\mathcal{K} \subset \mathcal{K}_2$ como o conjunto de índices dos iterados onde estes três fatos são satisfeitos.

Consideremos a derivada direcional na direção h . Por definição de x^k , $\nabla f_{\beta_k, r_k}(x^k) = 0$, e daqui, usando o fato que $\theta'_r(y) = \theta'(y/r)$,

$$\nabla f_0(x^k)^T h + \beta_k \sum_{i=1}^n \theta' \left(\frac{f_i(x^k)}{r_k} \right) \nabla f_i(x^k)^T h = 0.$$

Dividindo por β_k e separando convenientemente os índices do somatório,

$$\frac{\nabla f_0(x^k)^T h}{\beta_k} + \sum_{i \notin I(x^*)} \theta' \left(\frac{f_i(x^k)}{r_k} \right) \nabla f_i(x^k)^T h = - \sum_{i \in I(x^*)} \theta' \left(\frac{f_i(x^k)}{r_k} \right) \nabla f_i(x^k)^T h \quad (2.21)$$

Tomando limites para $k \xrightarrow{\mathcal{K}} +\infty$, temos:

Lado esquerdo:

$$\frac{\nabla f_0(x^k)^T h}{\beta_k} \xrightarrow{\mathcal{K}} 0, \quad \text{porque } \beta_k \xrightarrow{\mathcal{K}} +\infty.$$

$$\frac{f_i(x^k)}{r_k} \leq -\frac{\mu}{r_k} \quad \text{para } i \notin I(\bar{x}), \quad \text{devido a (c). Daqui, por a)}$$

$$\frac{f_i(x^k)}{r_k} \xrightarrow{\mathcal{K}} -\infty, \quad \text{e assim } \theta' \left(\frac{f_i(x^k)}{r_k} \right) \xrightarrow{\mathcal{K}} 0.$$

concluimos que o lado esquerdo converge a 0 em \mathcal{K} .

Lado direito:

Para $i \in I(\bar{x})$, $\nabla f_i(x^k)^T h \leq -\rho/2$, por (b). Assim:

$$\begin{aligned} - \sum_{i \in I(x^*)} \theta' \left(\frac{f_i(x^k)}{r_k} \right) \nabla f_i(x^k)^T h &\leq -\frac{\rho}{2} \sum_{i \in I(x^*)} \theta' \left(\frac{f_i(x^k)}{r_k} \right) \\ &\leq -\frac{\rho}{2} \sum_{i \in I^+(x^k)} \theta' \left(\frac{f_i(x^k)}{r_k} \right) \\ &\leq -\frac{\rho}{2} \theta'(0) |I^+(x^k)|, \end{aligned}$$

porque $I^+(x^k) \subset I(\bar{x})$ por (a), e para $i \in I^+(x^k)$, $f_i(x^k) \geq 0$ e $\theta(\cdot)$ é crescente. Como x^k é inviável para $k \in \mathcal{K}$, $|I^+(x^k)| \geq 1$, o lado direito tem uma limitante superior negativa, contradizendo (2.21) e completando a prova. \blacksquare

Capítulo 3

Métodos de ponto proximal e minimização sobre R_+^n .

Neste capítulo apresentamos um resumo sobre as principais abordagens feitas sobre o método de ponto proximal. A intenção final é relacioná-las com o problema de minimização de uma função convexa definida em R_+^n . Finalmente apresentamos algumas funções de penalidade que usaremos em uma abordagem posterior.

3.1 Método de Ponto Proximal Clássico.

Seja $f : R^n \rightarrow R \cup \{+\infty\}$ uma função convexa própria fechada. Consideremos a função

$$z \in R^n \mapsto F(z) = \inf \left\{ f(x) + \frac{1}{2} \|z - x\|^2 : x \in R^n \right\}.$$

Esta função foi introduzida e estudada por Moreau [48]. O único ponto onde o ínfimo é atingido foi denotado por $prox_f(z)$. Se f é a função indicadora de um conjunto convexo e fechado $C \subset R^n$, então $prox_f(z)$ coincide com a projeção de z sob o conjunto C . Posteriormente Yosida fez a primeira modificação introduzindo um parâmetro positivo λ para obter a conhecida regularização Moreau-Yosida:

$$z \in R^n \mapsto F_\lambda(z) = \inf_x \left\{ f(x) + \frac{1}{2\lambda} \|z - x\|^2 \right\}.$$

Cabe destacar que esta regularização constitui uma convolução ínfima, no sentido de análise convexa (ver [54]), entre a função f e o termo ou núcleo quadrático, preservando o conjunto de minimizadores e tendo um efeito de suavização entre outras propriedades. Um estudo genérico de regularizações para problemas de minimização convexa encontra-se em Sosa [58].

Consideremos o problema

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } x \in R^n \end{aligned} \tag{3.1}$$

com $f : R^n \rightarrow R$ convexa.

O método de ponto proximal para este problema gera uma seqüência $\{x^k\}$ em R^n tal que

$$x_0 \in R^n \quad \text{e} \\ x^{k+1} = \operatorname{argmin}_{x \in R^n} \left\{ f(x) + \frac{\lambda_k}{2} \|x - x^k\|^2 \right\}, \tag{3.2}$$

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

A idéia geométrica do método é mostrada na fig.12 através de um simples exemplo, trata-se de minimizar $f(x) = x^2$ (linha grossa). A partir de $x^0 = -1.5$ e $\lambda_k = 1$ fixo, os minimizadores de $f(x) + \frac{\lambda_k}{2} \|x - x^k\|^2$ para duas iterações do método aproximam-se do ótimo $x^* = 0$.

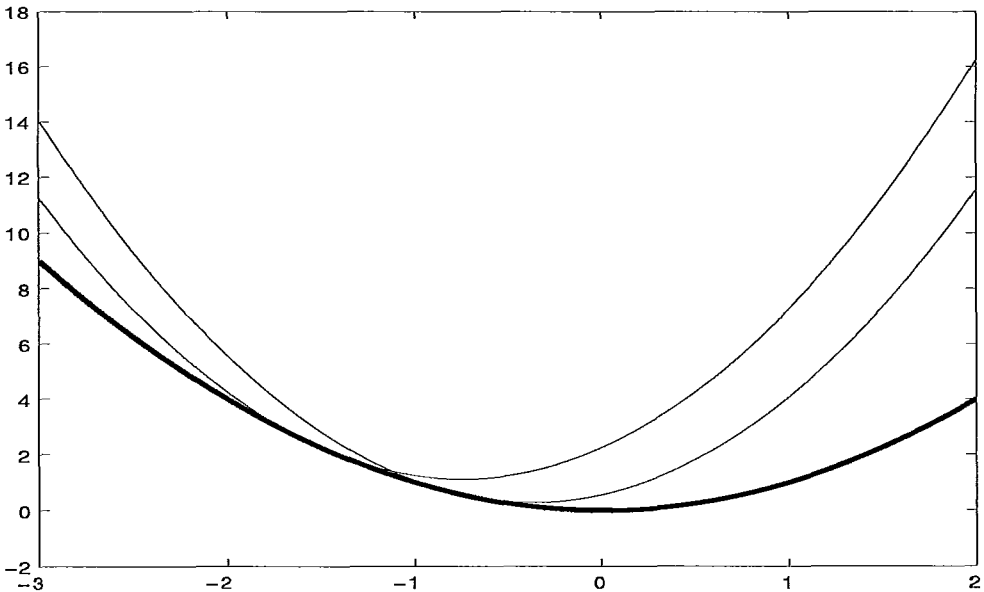


fig.12

Este método foi introduzido em otimização convexa por Martinet [46], sendo aprofundado por Rockafellar [55] no contexto de operadores monótonos maximais. Lemaire [42] apresenta um *survey* com as principais propriedades e aplicações do método. A seguir apresentamos alguns conceitos e resultados interessantes para uma melhor compreensão da temática subsequente.

Definição 3.1 Um operador $T : R^n \rightarrow \mathcal{P}(R^n)$ é monótono se e somente se

$$(x - y, u - v) \geq 0 \text{ para todo } x, y \in R^n, u \in T(x), v \in T(y).$$

Definição 3.2 Um operador T é monótono maximal se seu gráfico não está incluído propriamente no gráfico de outro operador monótono distinto, isto é, se T é monótono e todo operador monótono $T' : R^n \rightarrow \mathcal{P}(R^n)$ satisfaz: para todo $x \in R^n$, se $T(x) \subset T'(x)$ então $T(x) = T'(x)$.

Observação: O conceito de operador monótono maximal constitui uma generalização do conceito de transformação linear semidefinida positiva para o caso não linear.

Definição 3.3 Um operador $T : R^n \rightarrow R^n$ é não expansivo se e somente se $\forall x, y \in R^n, \|T(x) - T(y)\| \leq \|x - y\|$.

Definição 3.4 Um operador $T : R^n \rightarrow R^n$ é firmemente não expansivo se e somente se $\forall x, y \in R^n, \|T(x) - T(y)\|^2 \leq \|x - y\|^2 - \|(x - y) - (T(x) - T(y))\|^2$.

Definição 3.5 Seja $T : R^n \rightarrow \mathcal{P}(R^n)$ monótono maximal. T^{-1} é Lipschitz contínuo em 0 com módulo $\sigma > 0$ se existe β , tal que $\|x - \bar{x}\| \leq \sigma \|y\|$ sempre que $0 \in T(\bar{x})$, $y \in T(x)$ e $\|y\| \leq \beta$.

Rockafellar interessou-se inicialmente em achar zeros de um operador monótono maximal T , assim dado $T : R^n \rightarrow \mathcal{P}(R^n)$, x é um zero de T se e somente se $0 \in T(x)$. O método de ponto proximal para achar zeros de um operador monótono maximal gera uma seqüência definida por:

Dado $x^0 \in R^n$, encontre

$$x^{k+1} \in (I + \frac{1}{\lambda_k} T)^{-1}(x^k), \quad (3.3)$$

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

O problema de encontrar zeros de um operador monótono maximal generaliza o problema de minimizar uma função convexa pois se fazemos $T = \partial f$, $0 \in T(x)$ é equivalente a $0 \in \partial f(x)$ implicando

$$\forall y \in R^n, f(y) - f(x) \geq \langle 0, y - x \rangle \geq 0$$

e assim

$$\forall y \in R^n, f(x) \leq f(y).$$

Observe além disso que aplicando condições de otimalidade na expressão (3.2) com x^{k+1} como minimizador e usando um operador monótono maximal T em lugar de ∂f temos:

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

ou equivalentemente

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

obtendo assim a iteração principal do método de ponto proximal para encontrar zeros de um operador monótono maximal.

Podemos então considerar uma versão aproximada do método através da expressão

$$x^{k+1} \approx P_k(x^k) \tag{3.4}$$

com

$$P_k = \left(I + \frac{1}{\lambda_k} T \right)^{-1},$$

assim consideramos

$$\|x^{k+1} - P_k(x^k)\| \leq \epsilon_k$$

ou

$$\|x^{k+1} - P_k(x^k)\| \leq \epsilon_k \|x^{k+1} - x^k\| \tag{3.5}$$

onde $\{\epsilon_k\}$ satisfaz $\sum_{k=0}^{\infty} \epsilon_k < \infty$.

As provas de convergência usando esta versão aproximada (ver [55]) estão relacionadas com as idéias de não expansividade. Minty [47] mostrou que se $\lambda > 0$ e T é monótono maximal então $I + \lambda T$ é uma bijeção e $(I + \lambda T)^{-1}$ é firmemente não expansivo, tudo isto assegura que x^{k+1} em (3.3) é unicamente determinado.

O teorema 2 em Rockafellar [55] estabelece que se $\{x^k\}$ satisfazendo (3.5) é limitada e T^{-1} é Lipschitz contínuo em 0, com uma adequada escolha do módulo σ definido em (3.5), então o método converge com uma razão de convergência linear para λ_k suficientemente grande e superlinear se $\lambda_k \rightarrow +\infty$. No teorema 3 da mesma referência estabelecem-se condições mediante as quais o método em sua forma exata converge em um número finito de passos. Outros estudos sobre a taxa de convergência do método em sua forma exata ou inexacta foram dados em Luque [43] e Lemaire [42].

Nos últimos anos algumas sugestões tem surgido com o fim de estender a idéia original do método de ponto proximal substituindo o núcleo quadrático por outros não quadráticos. Attouch and Wets [1], no contexto de regularização consideraram núcleos da forma $k(z - x)$ onde $k : \mathbb{R}^n \rightarrow \mathbb{R}_+$ é uma função estritamente convexa e coerciva tal que $k(0) = 0$ recuperando assim a aproximação de Moreau-Yosida. Humes-Da Silva [31] consideram o método de ponto proximal com regularizações da forma $\theta(x - x^k)$ com $\theta : \mathbb{R}^n \rightarrow \mathbb{R}_+$ estritamente convexa e diferenciável na origem. Usando a idéia de separadores [20], mostram que todo ponto de acumulação da seqüência gerada pelo correspondente método de ponto proximal é solução ótima do problema (3.1).

Iusem [32] considera os conceitos de convergência no sentido de Fejér com as quais constrói provas mais simples para a convergência do método de ponto proximal com a vantagem de poder estender a idéia para o caso de núcleos não quadráticos, especificamente o caso das quase-distâncias que definiremos posteriormente.

Definição 3.6 *A seqüência $\{y^k\} \subset \mathbb{R}^n$ é Fejér convergente a um conjunto $U \subset \mathbb{R}^n$ com respeito à norma Euclidiana se*

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

Mostra-se que se uma seqüência $\{x^k\}$ é Fejér convergente a um conjunto $U \neq \emptyset$, então $\{x^k\}$ é limitada. Se um ponto limite \bar{x} desta seqüência pertence a U , então $\lim_{k \rightarrow +\infty} x^k = \bar{x}$. No teorema a seguir usa-se este fato com respeito à seqüência gerada por (3.2).

Teorema 3.7 (ver teorema 2.1 em Iusem [32]) *Seja $f : R^n \rightarrow R$ convexa e continuamente diferenciável. Suponha que o conjunto de minimizadores de f sobre R^n é não vazio. Então a seqüência gerada por (3.2) converge a um ponto $x^* \in U$.*

Podemos enfatizar que o método de ponto proximal não representa um esquema computacional implementável com eficiência, mas sim uma forte ferramenta conceitual para o análise da convergência de algoritmos em otimização.

3.2 Quase-distâncias

Definição 3.8 *Seja $X \subset R^n$ um conjunto não vazio. Uma função $D : X \times X \rightarrow R$ é uma quase-distância em X se e somente se verificam-se as condições:*

- a) $\forall x, y \in X, D(x, y) \geq 0$.
- b) $D(x, y) = 0$ se e somente se $x = y$.

O método de ponto proximal tem sido generalizado basicamente substituindo o núcleo quadrático por uma quase-distância. Entre elas abordaremos as conhecidas distâncias de Bregman e as φ -divergências. Mostra-se que as restrições do problema original podem ser incorporadas na função critério usando a propriedade de coercividade na fronteira que definiremos adiante. Na atualidade o estudo deste tipo de métodos está ligado aos métodos de lagrangeano aumentado, os quais possuem uma interessante relação de dualidade que será analisada no capítulo 4.

3.2.1 Distâncias de Bregman

Sejam $S \subset R^n$ um conjunto aberto com fecho \bar{S} e $h : \bar{S} \rightarrow R$ uma função convexa e diferenciável em S . Considere a função $D_h : \bar{S} \times S \rightarrow S$ definida por

$$D_h(x, y) = h(x) - h(y) - \nabla h(y)^t(x - y). \quad (3.6)$$

Dizemos que h é uma função de Bregman com zona S se satisfaz as seguintes condições:

- B1) h é continuamente diferenciável em S .

B2) h é estritamente convexa e contínua em \bar{S} .

B3) Para todo $\delta \in \mathbb{R}$, os conjuntos de nível

$$L_1(y, \delta) = \{x \in \bar{S} : D_h(x, y) \leq \delta\} \text{ e}$$

$$L_2(x, \delta) = \{y \in S : D_h(x, y) \leq \delta\}$$

são limitados $\forall y \in S, \forall x \in \bar{S}$ respectivamente.

B4) Se $\{y^k\} \subset S$ é uma seqüência que converge a y^* então $D_h(y^*, y^k)$ converge a 0.

B5) Se $\{x^k\} \subset \bar{S}$, $\{y^k\} \subset S$ são seqüências tais que $\{x^k\}$ é limitada, $\lim_{k \rightarrow \infty} y^k = y^*$ e $\lim_{k \rightarrow \infty} D_h(x^k, y^k) = 0$ então $\lim_{k \rightarrow \infty} x^k = y^*$.

Observações:

a) D_h é chamada distância de Bregman induzida por h , também é conhecida como D-função.

b) As condições iv) e v) se satisfazem automaticamente em S como consequência de i), ii) e iii), assim elas precisam ser verificadas somente na fronteira de S .

c) D_h satisfaz as condições a) e b) da definição (3.8) mas não é uma distância (não é simétrica e não satisfaz a desigualdade triangular).

d) Dada uma função de Bregman e um conjunto convexo fechado $C \subset \mathbb{R}^n$, podemos definir um operador projeção $P_{h,C}$ substituindo a norma usual euclidiana por $D_h(x, y)$ com solução única pela convexidade estrita de h .

Seja $S \subset \mathbb{R}^n$ um conjunto aberto e convexo e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e contínua em \bar{S} . Considere o problema

$$\begin{aligned} &\text{minimizar } f(x) \\ &\text{sujeito a } x \in \bar{S} \end{aligned} \tag{3.7}$$

O método de ponto proximal usando distâncias de Bregman gera uma seqüência $\{x^k\}$ contida em S tal que

$$x^0 \in S \quad \text{e}$$

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k D_h(x, x^k)\}, \tag{3.8}$$

com λ_k satisfazendo $0 < \lambda_k \leq \bar{\lambda}$ para algum $\bar{\lambda}$ e $k = 0, 1, \dots$

Definição 3.9 Uma função de Bregman h é coerciva na fronteira se verifica: Se $\{y^k\} \subset S$ é uma seqüência tal que $\lim_{k \rightarrow \infty} y^k = y \in \partial S$, então

$$\lim_{k \rightarrow \infty} \nabla h(y^k)^t(x - y^k) = -\infty \text{ para todo } x \in S.$$

Definição 3.10 Uma função h é coerciva na zona S se para cada $y \in R^n$, existe $x \in S$ tal que $\nabla h(x) = y$.

A coercividade na fronteira foi introduzida por Iusem [34] e constitui uma propriedade importante pois garante que os iterados de (3.8) permanecem no interior de S . O conceito de coercividade na zona de h é usado para a análise de convergência do método quando é aplicado ao problema de desigualdade variacional.

Teorema 3.11 (ver lema 3.1 em Iusem [34]) Suponha que $S = (a_1, b_1) \times \dots \times (a_n, b_n)$, com $a_j \in R \cup \{-\infty\}$, $b_j \in R \cup \{\infty\}$, $a_j < b_j$ para todo j e h é separável. Então h é coerciva na zona se e somente se é coerciva na fronteira e os conjuntos de níveis de $\|\nabla h(\cdot)\|$ são limitados.

As funções de Bregman foram introduzidas por L.Bregman [10] no contexto de encontrar pontos comuns entre conjuntos convexos em espaços de Hilbert e sua aplicação em otimização. Censor e Zenios [12] propuseram o método com D-funções para o caso de dimensão finita. Posteriores estudos do método podem ser encontrados nas referências [19],[13], [21], [59]. Burachik e Iusem [11] apresentam o método aplicado a problemas de desigualdade variacional em espaços de Hilbert. Em [40] relacionam-se os conceitos de trajetória central derivadas de funções barreira, de trajetórias de Cauchy em variedades de Riemann e de métodos de ponto proximal com distâncias de Bregman.

Exemplo 1:

Sejam $S = R_{++}^n = \{x \in R^n : x_i > 0, i = 1, \dots, n\}$ e a função

$$x \in \bar{S} \mapsto h(x) = \sum_{i=1}^n x_i \log x_i$$

estendida com continuidade na origem definindo $0 \log 0 = 0$. Assim obtemos a conhecida distância entrópica ou divergência de Kullback-Leibler:

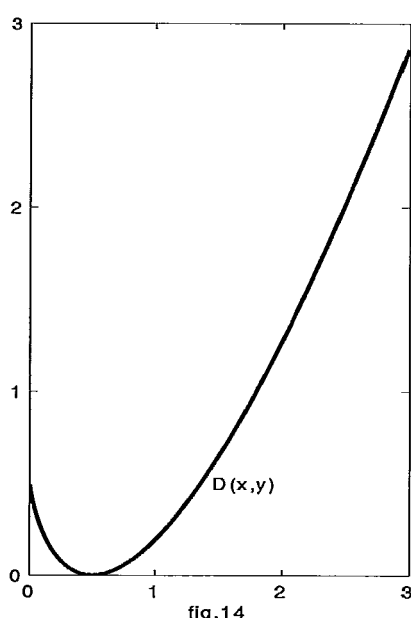
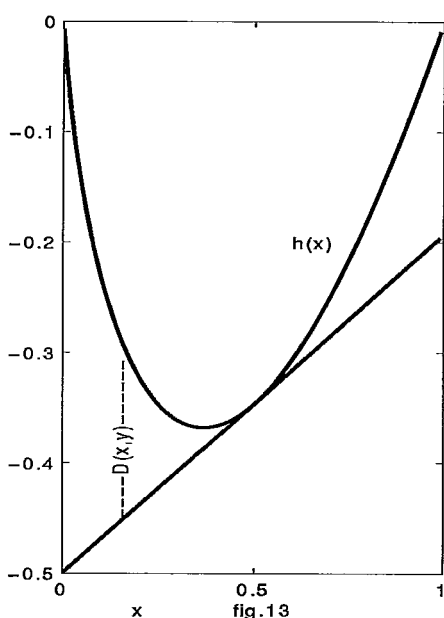
$$D_h(x, y) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{y_i} + y_i - x_i \right). \quad (3.9)$$

Exemplo 2:

Se $S = \mathbb{R}^n$ e $h(x) = \frac{1}{2} \|x\|^2$, obtemos $D_h(x, y) = \frac{1}{2} \|x - y\|^2$. Recupera-se o núcleo quadrático usado no método de ponto proximal clássico.

Outros exemplos podem-se encontrar em [21] e [32]. No capítulo 5 mostramos novos exemplos de funções de Bregman.

A seguir ilustramos o comportamento geométrico das funções D_h . Na fig.10 apresentamos uma interpretação geométrica para D_h como a diferença entre h e sua aproximação linear no ponto y . Na fig.11 mostramos o gráfico de D_h dada por (3.9).



Finalizamos apresentando três propriedades importantes das distâncias de Bregman e um resultado de convergência do método.

Proposição 3.12 (ver proposição 9.1 em Iusem [32]) *Seja h uma função de Bregman com zona S , então:*

- i) $D_h(x, y) - D_h(x, z) - D_h(z, y) = (\nabla h(y) - \nabla h(z), z - x)$ para todo $x \in \bar{S}$, $y, z \in S$,
- ii) $\nabla_x D_h(x, y) = \nabla h(x) - \nabla h(y)$ para todo $x, y \in S$,
- iii) $D_h(\cdot, y)$ é estritamente convexa para todo $y \in S$.

A propriedade i), é uma variação da conhecida identidade de três-pontos apresentada em Chen-Teboulle [13], a qual facilita as provas de convergência mostradas

em [59] e em [32] no sentido de Fejér convergência.

Teorema 3.13 (ver teorema 4.1 em Iusem [34]) *Sejam $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa com domínio efetivo $\text{dom}(f)$ fechado tal que $\text{dom}(f) \cap S \neq \emptyset$ e $\{\lambda_k\}$ uma seqüência em \mathbb{R}_{++}^n limitada superiormente. Considere $h : \bar{S} \rightarrow \mathbb{R}$ coerciva na fronteira. Então a seqüência gerada por (3.8) está contida em S e converge a uma solução do problema (3.7).*

Outras provas com diferentes hipóteses podem ser encontradas em [21] e [35]. Recentemente, Iusem e Monteiro [37], consideraram o problema de minimizar uma função convexa diferenciável restrita ao conjunto $C = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ com $A \in \mathbb{R}^{m \times n}$ e b um m -vector real, e seu correspondente problema dual. O algoritmo estudado gera uma seqüência $\{x^k\}$ através de

$$x^{k+1} = \operatorname{argmin}\{f(x) + \lambda_k D_h(x, x^k) : Ax = b\},$$

com $x^0 > 0 \in \mathbb{R}^n$, $\{\lambda_k\} \subset \mathbb{R}_{++}^n$ satisfazendo $\sum_{k=0}^{\infty} \lambda_k^{-1} = \infty$ e h uma família de funções de Bregman mais geral. Mostra-se principalmente que a seqüência de médias ponderadas da seqüência dual converge ao centroide do conjunto ótimo dual S^* . Distâncias de Bregman para o problema de desigualdade variacional são analisadas em [11] e [35].

3.2.2 φ -divergências

Consideremos $\varphi : (0, +\infty) \mapsto [0, +\infty)$ tal que:

- i) φ é estritamente convexa e três vezes continuamente diferenciável.
- ii) $\varphi(1) = 0$.
- iii) $\varphi'(1) = 0$.
- iv) $\varphi''(1) > 0$.
- v) $\lim_{t \rightarrow 0} \varphi'(t) = -\infty$.

Definamos $d_\varphi : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}_+$ como:

$$d_\varphi(x, y) = \sum_{i=1}^n y_i \varphi\left(\frac{x_i}{y_i}\right), \quad (3.10)$$

d_φ com φ satisfazendo i)-v) é chamada φ -divergência.

Proposição 3.14 (ver proposição 21.1 em Iusem [32])

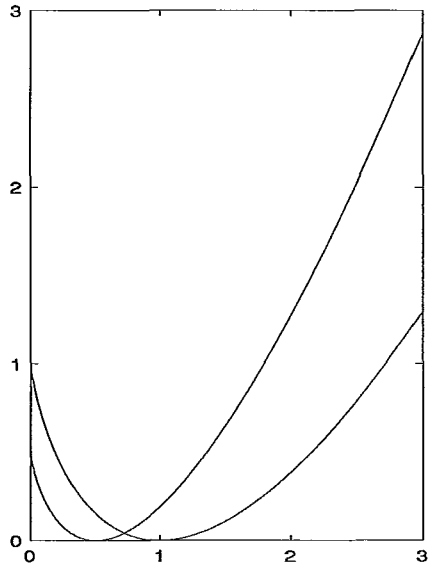
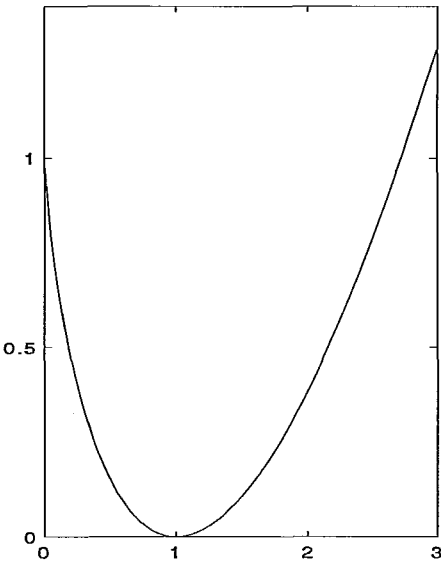
- i) $d_\varphi(x, y) \geq 0$ para todos $x, y \in \mathbb{R}_{++}^n$.
- ii) $d_\varphi(x, y) = 0$ se e somente se $x = y$.
- iii) Os conjuntos de nível de $d_\varphi(\cdot, y)$ são limitados para todo $y \in \mathbb{R}_{++}^n$.
- iv) Os conjuntos de nível de $d_\varphi(x, \cdot)$ são limitados para todo $x \in \mathbb{R}_{++}^n$.
- v) $d_\varphi(x, y)$ é juntamente convexa em x, y e estritamente convexa em x .
- vi) $\lim_{k \rightarrow \infty} d_\varphi(y, y^k) = 0$ se e somente se $\lim_{k \rightarrow \infty} y^k = y$.

As φ -divergências, assim como as distâncias de Bregman são quase-distâncias. Não satisfazem a simetria nem a desigualdade triangular. Csiszar [16] introduziu as φ -divergências como medidas generalizadas de informação sobre o conjunto de distribuições de probabilidade $D_n = \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i \geq 0\}$ para medir a proximidade entre dois elementos do conjunto D_n .

Observe que para $\varphi(t) = t \log t - t + 1$, a φ -divergência correspondente é

$$d_\varphi(x, y) = \sum_{i=1}^n \left(x_i \log \frac{x_i}{y_i} + y_i - x_i \right) \tag{3.11}$$

conhecida como a distância de entropia relativa de Kullback-Leibler e coincide com (3.9) como distância de Bregman. φ é mostrada na fig.8 e sua correspondente d_φ na fig.9 para dois valores distintos de y .



Consideremos novamente o problema (3.7) com $\bar{S} = [0, +\infty)$:

$$\begin{aligned} & \text{minimizar } f(x) \\ (P2) \quad & \text{sujeito a } x \geq 0 \end{aligned}$$

com $f : R^n \rightarrow R$ convexa.

O método de ponto proximal entrópico com φ -divergência para este problema e dado por:

Dado $x^0 \in R^n$,

$$x^{k+1} = \operatorname{argmin}_{x \geq 0} \{f(x) + \lambda_k d_\varphi(x, x^k)\} \quad (3.12)$$

com $\{\lambda_k\} \subset R$ satisfazendo para todo $k \in N$, $0 < \lambda_k \leq \bar{\lambda}$ para algum $\bar{\lambda}$ positivo.

Um estudo completo da convergência do método se encontra em Iusem, Svaiter e Teboulle [33] baseada na convergência quase-Fejér introduzida por Ermol'ev [26] no contexto de variáveis aleatórias. Eggermont [22] mostra um esquema iterativo baseado nas condições de otimalidade para o problema (P2) para f continuamente diferenciável sobre R^n e prova a convergência da seqüência gerada por (3.12) para d_φ como em (3.11).

Apresentamos alguns dos resultados principais sobre a seqüência gerada por (3.12) e que podem ser encontrados em [32] e [33].

Definição 3.15 *A seqüência $\{y^k\} \subset \mathbb{R}_{++}^n$ é quase-Fejér convergente ao conjunto $U \subset R_+^n$ com respeito à φ -divergência d_φ se para cada $u \in U$ existe uma seqüência $\{\epsilon_k\} \subset R_{++}$ tal que $\sum_{k=0}^{\infty} \epsilon_k < \infty$ e $\forall k \geq 0$*

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

A necessidade desta nova definição deve-se a que a seqüência gerada por (3.12) não é sempre Fejér convergente ao conjunto de soluções U com respeito à d_φ para φ arbitrária.

Teorema 3.16 *:(ver proposição 13.1 em Iusem [32]) Se $\{y^k\} \in \mathbb{R}_{++}^n$ é quase-Fejér convergente a $U \subset R_+^n$ com respeito à φ -divergência d_φ então $\{y^k\}$ é limitada. Se um ponto de acumulação \bar{y} pertence a U então $\lim_{k \rightarrow \infty} y^k = \bar{y}$.*

Uma condição suficiente para garantir a convergência da seqüência gerada por (3.12) a uma solução ótima de (P2) é que

$$\varphi'(t) \leq \varphi''(1) \log t \text{ para todo } t > 0. \quad (3.13)$$

Algumas funções que satisfazem esta condição podem-se ver em [33]. Particularmente a função $\varphi(t) = t \log t - t + 1$ satisfaz esta condição. Temos assim o seguinte resultado.

Teorema 3.17 *(ver corolário 4.4 em Iusem-Svaiter-Teboulle [33])*

Se $\varphi'(t) \leq \varphi''(1) \log t$ para todo $t > 0$ então a seqüência gerada por (3.12) converge a uma solução ótima de (P2).

Resultados referentes à taxa de convergência do método podem ser encontrados em [38] e em [39] para o problema de programação linear com restrições de igualdade. Auslender e Haddou [4] estendem o método com objetivo f convexa sobre R^n e restrita ao poliedro $C = \{x \in R^n / Ax \leq b\}$ com A de ordem $m \times n$ e $b \in R^m$. Também considera o problema de desigualdade variacional mostrando convergência.

Fora do enfoque de ponto proximal, as funções de entropia foram usadas inicialmente para resolver problemas de programação linear. Eriksson [24] estudou o problema de programação linear com restrições de igualdade adicionando à função objetivo ϵ vezes uma função entrópica que coincide com a φ -divergência (3.11). Outros usos de funções entrópicas para programação linear podem ser consultados em [23], [25] e referências ali citadas.

3.2.3 Outra quase-distância

Ben-Tal e Zibulevsky [6] consideram uma família de funções convexas

$\varphi : R^n \rightarrow R \cup \{+\infty\}$ com $\text{dom} \varphi = (0, +\infty)$ tal que:

- a) φ é estritamente convexa e diferenciável em $(0, +\infty)$,
- b) $\varphi(1) = 0$,
- c) $\varphi'(1) = 0$,

d) $\lim_{y \rightarrow 0^+} \varphi'(y) = -\infty$,

e) $\lim_{y \rightarrow \infty} \varphi(y) = \infty$,

f) $\varphi''(y) \leq M$ para $y \geq 1$.

É definida uma quase-distância como

$$d_\varphi(x, y) = \sum_{i=1}^m \rho(x_i, y_i), \quad x \geq 0, y > 0 \quad (3.14)$$

com

$$\rho(x, y) = y \pi(y) \varphi\left(\frac{x}{y}\right), \quad (3.15)$$

onde $\pi(y) > 0$ é crescente em R_+ , positiva e sublinear, isto é,

$$\forall y > 0: \pi(y) \leq c y \quad \text{para algum } c > 0.$$

A seguinte proposição caracteriza as principais propriedades satisfeitas por d_φ .

Proposição 3.18 (ver proposição 3 em [6]) *Seja $f : R_+^n \rightarrow R$ uma função côncava limitada superiormente e φ satisfazendo condições a)-f). Considere a quase-distância $d_\varphi : R_+^n \times R_+^n$ definida por (3.14) e (3.15). Então d_φ satisfaz cada uma das seguintes afirmações:*

1) $\forall y > 0$, $d_\varphi(\cdot, y)$ é estritamente convexa sobre R_+^n respeito do primeiro argumento.

2) Os conjuntos de níveis de $d_\varphi(\cdot, y)$ são limitados para todo $y > 0$.

3) $d_\varphi(x, y) \geq 0$, $d_\varphi(x, x) = 0$.

4) $\forall y > 0$, $d_\varphi(x, y) = \infty$ se $x < 0$, e $\lim_{\|x\| \rightarrow \infty} d_\varphi(x, y) = \infty$.

Além disso, a seqüência $\{x^k\}$ gerada por

$$x^{k+1} = \operatorname{argmax}_x \{f(x) - d_\varphi(x, x^k)\}, \quad u^0 > 0, \quad (3.16)$$

é bem definida, positiva e a seqüência de valores $f(x^k)$ é não decrescente.

O processo iterativo definido por (3.16) pode ser usado para resolver o problema (P2). Adiamos sua discussão para o próximo capítulo.

3.2.4 Método de ponto proximal com regularização estritamente convexa

Humes e Silva [31] apresentam um método de ponto proximal baseado em uma função $\varphi : R^n \rightarrow R \cup \{+\infty\}$ estritamente convexa, fechada e diferenciável na origem, cujo domínio efetivo contém uma bola fechada centrada na origem e $\varphi(0) = 0$.

Seja $f_0 : R^n \rightarrow R \cup \{+\infty\}$ convexa, própria, fechada e limitada por baixo. O método de ponto proximal com regularização estritamente convexa para minimizar f_0 sobre R^n gera uma seqüência $\{x^k\}$ tal que a partir de $x^0 \in \text{dom } f$ obtém-se

$$x^{k+1} \in \operatorname{argmin}_{x \in R^n} \{f_0(x) + r^k \varphi(x - x^k)\} \quad (3.17)$$

e r^k é uma seqüência de reais positivos limitada superiormente.

Uma novidade a destacar nesta abordagem é o uso de separadores (ver [63]) nas provas de convergência. Uma relação entre este método e o método de Lagrangeano aumentado será abordada no próximo capítulo.

3.3 Minimização sobre R_+^n

Considere o problema

$$(P2) \quad \begin{array}{ll} \text{minimizar} & d(s) \\ \text{sujeito a} & s \geq 0 \end{array}$$

com $d : R^n \rightarrow R$ convexa.

Para resolver este problema usando distâncias de Bregman tem-se usado a propriedade de coercividade na fronteira, que faz com que os iterados permaneçam no interior do ortante positivo. A função de Kullback-Leibler (3.9) e outras mostradas em [33] satisfazem esta condição.

Para as φ -divergências, a coercividade está incluída na definição da φ . Neste caso basicamente encontramos na literatura três tipos de provas para a convergência do método de ponto proximal associado. A primeira é encontrada em

Eggermont [22] que também usa a φ -divergência de Kullback-Leibler (3.9), desta vez num esquema iterativo obtido a partir das condições de otimalidade do problema (P2) penalizado com d_φ . As hipóteses usadas foram que a função objetivo seja de classe C^1 com conjuntos de nível limitados e gradiente localmente lipschitziano. A segunda abordagem é discutida por vários autores e é amplamente estudada em Iusem, Svaiter e Teboulle [33] e foi discutida nos teoremas 3.16 e 3.17 da seção anterior. No próximo capítulo consideramos uma hipótese formulada por Polyak e Teboulle [50] que conduz à condição (3.13), a qual é indispensável para as provas de convergência usadas em [33]. Finalmente Ben-Tal e Zibulesky [6] usaram a quase-distância definida em (3.16) obtendo resultados semelhantes de convergência com hipóteses diferentes, isto será retomado no capítulo 4.

As seguintes famílias de penalidades serão usadas no próximo capítulo onde se estabelecerá uma relação de dualidade com as penalidades tipo I e tipo II definidas no capítulo 2.

3.3.1 Penalidade barreira tipo III

São funções $\theta : R \rightarrow R \cup \{+\infty\}$ convexas, próprias e fechadas tais que $\text{ri int } \theta = (0, b)$ com $b > 0$ que satisfazem:

- i) θ é diferenciável em $(0, b)$.
- ii) $\theta'(1) = 0$,
- iii) $\theta(1) = 0$
- iv) $\lim_{t \rightarrow 0^+} \theta'(t) = -\infty$,
- v) $\lim_{t \rightarrow b^-} \theta'(t) = +\infty$.

Os extremos do intervalo $[0, b]$ podem ou não pertencer a $\text{dom } \theta$.

Exemplos

$$\theta(y) = 2(1 - \sqrt{y(2-y)}), \text{ dom}(\theta) = [0, 2].$$

3.3.2 Penalidade barreira tipo IV

São as funções $\theta : R \rightarrow R \cup \{+\infty\}$ convexas diferenciáveis tais que:

- i) $\text{ri int } \theta = (0, +\infty)$.

ii) $\theta'(1) = 0$,

iii) $\theta(1) = 0$,

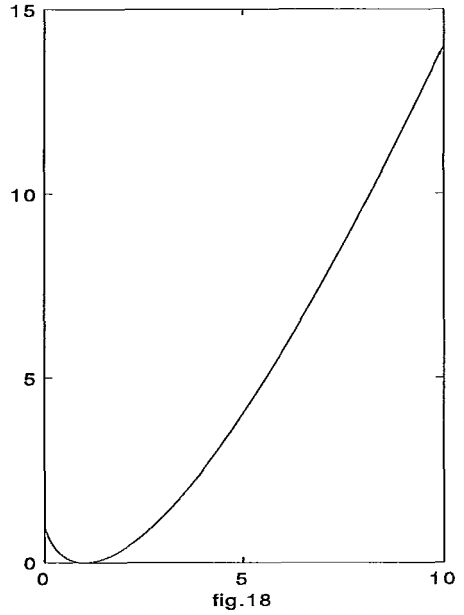
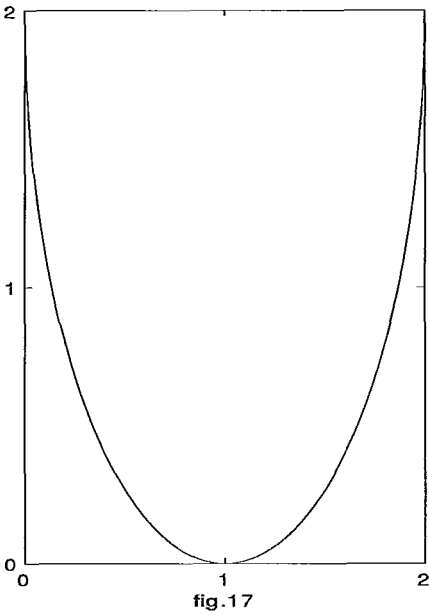
iv) $\lim_{t \rightarrow 0} \theta'(t) = -\infty$,

v) $\lim_{t \rightarrow +\infty} \theta(t) = +\infty$.

O ponto $y = 0$ pode ou não pertencer a $\text{dom } \theta$.

Exemplos Qualquer φ -divergência

Na fig.14 mostra-se uma penalidade do tipo III e na fig.15 uma do tipo IV. No próximo capítulo se estabelecerá uma relação de dualidade entre estas penalidades e as penalidades tipo I e tipo II.



Nos próximos capítulos usaremos estas famílias de penalidades III e IV. Será clara uma relação de dualidade com as famílias de penalidades I e II respectivamente.

Capítulo 4

Convexidade e métodos de lagrangeano aumentado

4.1 Convexidade e Dualidade

Neste capítulo retomamos o problema de programação não-linear (P) introduzido no capítulo 2, agora com hipótese de convexidade. Definimos o problema dual (D) associado a (P) e formulamos os conceitos básicos de análise convexa para assim introduzir o teorema de dualidade de Fenchel que usaremos como enlace entre os métodos de lagrangeano aumentado e os métodos de ponto proximal. Estes resultados podem ser encontrados em [54], [28] ou [29]. Fazemos um resumo das diversas famílias de penalidades e resultados principais considerados nos métodos de lagrangeano aumentado.

Considere o problema:

$$\begin{aligned} & \text{minimizar } f_0(x) \\ (P) \quad & \text{sujeito a } f_i(x) \leq 0 \quad i = 1, \dots, m \\ & x \in R^n \end{aligned}$$

com $f_i : R^n \rightarrow R \cup \{+\infty\}$ funções convexas próprias e fechadas para $i = 0, 1, \dots, m$. Seja S^* o conjunto de soluções ótimas de (P) e suponha que:

(A1) S^* é não vazio e compacto.

(A2) Existe $\bar{x} \in \text{dom} f$ tal que $f_i(\bar{x}) < 0$, $i = 1, \dots, m$ (Condição de qualificação de Slater).

Associado ao problema (P) temos o lagrangeano clássico dado por:

$$x \in R^n, y \in R^m \mapsto l(x, \mu) = f_0(x) + \sum_{i=1}^m \mu_i f_i(x).$$

Definimos a função dual lagrangeana como

$$\mu \in R^m \mapsto d(\mu) = \inf\{l(x, \mu) : x \in R^n\} \quad (4.1)$$

e o problema dual lagrangeano associado a P como:

$$\begin{aligned} & \text{maximizar } d(\mu) \\ (D) \quad & \text{sujeito a } \mu \geq 0. \end{aligned}$$

Observe que $d(\cdot)$ é côncava e assim $-d(\cdot)$ é convexa.

Denotemos por \hat{f} e \hat{d} os valores ótimos dos problemas (P) e (D) respectivamente e (D^*) o conjunto de soluções ótimas do problema (D).

É conhecido que sob a condição de qualificação de Slater, as condições de otimalidade (ou condições de Karush-Kuhn-Tucker) para o problema (P) são satisfeitas, isto é; para todo par $(x^*, \mu^*) \in S^* \times D^*$ cumpre-se:

$$\begin{aligned} 0 & \in \partial f_0(x^*) + \sum_{i=1}^m \mu_i^* \partial f_i(x^*), \\ \mu_i^* & \geq 0, \\ \mu_i^* f_i(x^*) & = 0 \quad (1 \leq i \leq m), \\ f_i(x^*) & \leq 0 \quad (1 \leq i \leq m). \end{aligned}$$

Além disso, (D^*) é não-vazio e compacto (ver ([28] VII-teorema 2.3.2), $\hat{f} = \hat{d}$ e para todo $\bar{d} < \hat{d}$, o conjunto de nível

$$\{\mu \in R^m \mid \mu \geq 0 \text{ e } d(\mu) \geq \bar{d}\}$$

é compacto.

Definição 4.1 Um vetor μ^* é um multiplicador de Lagrange para o problema (P) se $\mu^* \geq 0$ e

$$\inf_{x \in R^n} l(x, \mu^*) = \hat{f}.$$

Definição 4.2 *Seja $f : R^n \rightarrow R \cup \{+\infty\}$ uma função convexa. A convexa conjugada de f é a função f^* dada por:*

$$s \in R^n \mapsto f^*(s) = \sup_x \{x^t s - f(x) : x \in \text{dom} f\}.$$

Prova-se que f^* é fechada e convexa, própria se e somente se f é própria. A conjugada g^* de uma função côncava fechada $g : R^n \rightarrow R \cup \{-\infty\}$ é definida como

$$g^*(s) = \inf_{x \in R^n} \{x^t s - g(x)\}.$$

A função g^* é côncava e fechada. Observe que g^* não é definida como $-f^*$ para $f = -g$, de fato, definindo $f = -g$, obtém-se

$$\begin{aligned} f^*(s) &= \sup_{x \in R^n} \{x^t s - f(x)\} \\ &= \sup_{x \in R^n} \{-(-x^t s - g(x))\} \\ &= -\inf_{x \in R^n} \{-x^t s - g(x)\} \\ &= -g^*(-s). \end{aligned}$$

Proposição 4.3 *(ver proposição X.1.3.1 em [29])*

Sejam $f : R \rightarrow R \cup \{+\infty\}$ e $f_j : R \rightarrow R \cup \{+\infty\}$ para $j = 1, \dots, m$ funções convexas próprias fechadas. Tem-se:

- i) Se $h(x) = f(x) + \alpha$ então $h^*(s) = f^*(s) - \alpha$ para $\alpha \in R$.*
- ii) Se $h(x) = \alpha f(x)$, $\alpha > 0$, então $h^*(s) = \alpha f^*(\frac{s}{\alpha})$.*
- iii) Se $h(x) = f(\alpha x)$, $\alpha \neq 0$ então $h^*(s) = f^*(\frac{s}{\alpha})$.*
- iv) Se $h(x) = f(x - z)$ então $h^*(s) = f^*(s) + s^t z$, para $z \in R^n$.*
- v) Se $f(x) = \sum_{j=1}^m f_j(x_j)$ então*
 $f^*(s_1, \dots, s_m) = \sum_{j=1}^m f_j^*(s_j)$, para $f_j^* : R \rightarrow R \cup \{+\infty\}$, $j = 1, \dots, m$.

Definição 4.4 *Seja $f : R^n \rightarrow R \cup \{+\infty\}$ convexa. O vetor $s \in R^n$ é um subgradiente de f no ponto $x \in R^n$ se satisfaz:*

$$\forall y \in R^n, \quad f(y) \geq f(x) + s^t(y - x).$$

O conjunto de todos os subgradientes de f no ponto x , denotado por $\partial f(x)$, é chamado o subdiferencial de f em x .

Para $f : R^n \rightarrow R \cup \{-\infty\}$ côncava temos uma definição semelhante invertendo a desigualdade.

Proposição 4.5 (ver corolário X.1.4.4 em [29]) *Seja $f : R^n \rightarrow R \cup \{+\infty\}$ convexa própria fechada, as seguintes implicações são satisfeitas para $x, s \in R^n$:*

$$x \in \partial f^*(s) \Leftrightarrow s \in \partial f(x) \Leftrightarrow f(x) + f^*(s) - s^t x = 0 \quad (4.2)$$

Consideremos a função marginal ou função perturbação $v : R^m \rightarrow R$ definida por

$$u \in R^m \mapsto v(u) = \inf_{x \in R^n} \{f_0(x) \mid f(x) \leq u\}. \quad (4.3)$$

Sabe-se que esta função e a função dual d definida em (4.1) satisfazem

$$\forall \mu \geq 0, \quad d(\mu) = -v^*(-\mu), \quad (4.4)$$

de fato: (ver Bertsekas [8])

$$\begin{aligned} d(\mu) &= \inf_{x \in R^n} \{f_0(x) + \mu^t f(x)\} \\ &= \inf_{u \in R^m} \inf_{x \in R^n} \{f_0(x) + \mu^t f(x), f(x) \leq u\} \\ &= \inf_{u \in R^m} \inf_{x \in R^n} \{f_0(x) + \mu^t u, f(x) \leq u\} \\ &= \inf_{u \in R^m} \{\inf_{x \in R^n} \{f_0(x) \mid f(x) \leq u\} + \mu^t u\} \\ &= \inf_{u \in R^m} \{v(u) + \mu^t u\} \\ &= -\sup_{u \in R^m} \{(-\mu)^t u - v(u)\} \\ &= -v^*(-\mu) \end{aligned}$$

Teorema 4.6 (Dualidade de Fenchel) (ver teorema 31.1 em [54])

Seja f uma função convexa própria sobre R^n , e g uma função côncava própria sobre R^n , f^ a conjugada convexa de f e g^* a conjugada côncava de g . Verifica-se:*

$$\inf_{x \in R^n} \{f(x) - g(x)\} = \sup_{s \in R^n} \{g^*(s) - f^*(s)\} \quad (4.5)$$

se alguma das seguintes condições é satisfeita:

- a) $ri(\text{dom } f) \cap ri(\text{dom } g) \neq \emptyset$,
- b) f e g são fechadas, e $ri(\text{dom } g^*) \cap ri(\text{dom } f^*) \neq \emptyset$.

Se a) é satisfeita o supremo é atingido em algum s , se b) é satisfeita o ínfimo é atingido em algum x ; se a) e b) são satisfeitas simultaneamente, o ínfimo e o supremo são necessariamente finitos.

Estudaremos uma interessante relação de dualidade entre os métodos de penalidade que usam aproximações de penalidades tipo I com aqueles que aproximam penalidades do tipo III o que pode ser inferido a partir do exemplo a seguir:

Exemplo: Sejam as funções $f_i : R^n \rightarrow R \cup \{+\infty\}$ convexas próprias e fechadas para $i = 0, \dots, m$ e consideremos a função de penalidade barreira exata

$$\theta(y) = \sum_{i=1}^m \beta_i y_i^+$$

com

$$y_i^+ = \max\{0, y_i\} \text{ e } \beta_i > 0 \text{ para } i = 1, \dots, m.$$

Observe que se $\forall x \in R, f(x) = \max\{0, x\}$ então para todo

$$s \in R, f^*(s) = I_{[0,1]}(s),$$

onde I é a função indicadora do intervalo $[0, 1]$ dada por

$$I_{[0,1]}(y) = \begin{cases} 0 & \text{se } y \in [0, 1] \\ +\infty & \text{caso contrário.} \end{cases}$$

Usando as propriedades ii) e v) de funções conjugadas obtemos

$$\begin{aligned} \theta^*(w) &= \sum_{i=1}^m \beta_i I_{[0,1]} \left(\frac{w_i}{\beta_i} \right) \\ &= \sum_{i=1}^m I_{[0,\beta_i]}(w_i) \\ &= \begin{cases} 0 & \text{se } w_i \in [0, \beta_i], \quad i = 1, \dots, m. \\ +\infty & \text{caso contrário.} \end{cases} \end{aligned}$$

Consideremos o problema (P) e o correspondente problema penalizado

$$\text{minimizar}_{x \in R^n} \left\{ f_0(x) + \sum_{i=1}^m \beta_i (f_i(x))^+ \right\}, \quad (4.6)$$

sabe-se que sob a condição de Slater o problema (P) tem solução, e se $\hat{\mu}$ é um multiplicador ótimo então para $\beta_i > \hat{\mu}_i, i = 1, \dots, m$. as soluções de (4.6) coincidem com o conjunto de soluções de (P).

Finalmente, podemos associar a (4.6) um problema dual dado por

$$\begin{aligned} &\text{maximizar } d(w) \\ (D) \quad &\text{sujeito a } w_i \in [0, \beta_i] \quad i = 1, \dots, m. \end{aligned}$$

que é equivalente a considerar o problema

$$\text{maximizar}_{w \in R^n} \left\{ d(w) + \sum_{i=1}^m I_{[0, \beta_i]}(w_i) \right\}. \quad (4.7)$$

4.2 Métodos de lagrangeano aumentado

Na atualidade, existe uma forte tendência a estudar os métodos de lagrangeano aumentado motivada por seu bom desempenho em diversos problemas práticos e sobretudo pela rica e elegante ligação teórica com os métodos de ponto proximal via dualidade de Fenchel. Um excelente *survey* neste sentido pode se encontrar em Iusem [36] e referências ali citadas, principalmente associado a problemas com restrições de desigualdade. No entanto, podemos dizer que o interesse no estudo de lagrangeanos aumentados foi iniciado por Rockafellar [56] estendendo o método de Hestenes-Powell [57] formulado originalmente para o caso com igualdades.

Consideremos o problema com restrições de igualdade

$$\begin{aligned} & \text{minimizar} && f_0(x) \\ (P1) \quad & \text{sujeito a} && f_i(x) = 0 \quad i = 1, \dots, m \\ & && x \in R^n \end{aligned}$$

onde as funções $f_i : R^n \rightarrow R$ para $i = 1, \dots, m$ são continuamente diferenciáveis.

A idéia original de Hestenes-Powell foi usar a penalidade quadrática para o lagrangeano deste problema atualizando os multiplicadores em cada iteração. Assim consideraram o lagrangeano aumentado como a função

$$\begin{aligned} x \in R^n, \mu \in R^m \mapsto L_c(x, \mu) &= L(x, \mu) + \sum_{i=1}^m c f_i(x)^2 \\ &= f_0(x) + \sum_{i=1}^m \mu_i f_i(x) + c f_i(x)^2 \end{aligned}$$

onde $c > 0$ é o parâmetro de penalização e l o lagrangeano de (P1). Resolvemos o problema (P1) através de problemas irrestritos que a partir de μ^0 geram seqüências $\{x^k\}$, $\{\mu^k\}$ e $\{c^k\}$ dadas por

$$x^{k+1} \in \text{argmin} \{ L_{c^k}(x, \mu^k), \quad x \in R^n \} \quad \text{e}$$

$$\mu_i^{k+1} = \mu_i^k + 2c^k f_i(x^k) \quad \text{para } i = 1, \dots, m$$

A regra de atualização dos multiplicadores obedece a

$$\nabla L_{c^k}(x^{k+1}, \mu^k) = \nabla l(x^{k+1}, \mu^{k+1}) = 0 \quad (4.8)$$

O algoritmo tem uma regra de incremento para o parâmetro c^k . Todavía, a principal característica deste método é que satisfeitas condições de segunda ordem, temos convergência linear de $\{x^k\}$ a uma solução local de (P1) e de $\{\mu_k\}$ aos multiplicadores ótimos numa vizinhança destes sem precisar crescer ilimitadamente o parâmetro c . Desta forma melhora-se a estabilidade numérica dos métodos de penalidade.

4.2.1 Famílias de penalidades para o lagrangeano aumentado

A seguir mostramos uma família de funções de penalidade que são usadas pelo método de lagrangeano aumentado para o problema com igualdades (P1).

Família de penalidades (P_E) descrita por Bertsekas

Seja $\theta : \mathbb{R} \rightarrow \mathbb{R}$ tal que:

- i) θ é continuamente diferenciável e estritamente convexa sobre \mathbb{R} .
- ii) $\theta(0) = 0$.
- iii) $\theta'(0) = 0$.
- iv) $\lim_{y \rightarrow -\infty} \theta'(y) = -\infty$.
- v) $\lim_{y \rightarrow +\infty} \theta'(y) = \infty$.

Exemplos:

- a) $\theta(y) = \frac{1}{2}y^2$.
- b) $\theta(y) = \cosh(y) - 1$.
- c) $\theta(y) = \frac{1}{\rho}|y|^\rho, \quad \rho > 1$.

O lagrangeano aumentado usando esta família de funções é dado por

$$x \in \mathbb{R}^n, \mu \in \mathbb{R}^m \mapsto L_r(x, \mu) = f_0(x) + \sum_{i=1}^m \mu_i f_i(x) + r \sum_{i=1}^m \theta\left(\frac{f_i(x)}{r}\right).$$

onde $r \in (0, 1]$. O método de multiplicadores correspondente ou método de lagrangeano aumentado para resolver o problema (P1) gera seqüências $\{x^k\}$, $\{\mu^k\}$ e $\{r^k\}$ a partir de μ^0 tal que

$$x^{k+1} \in \operatorname{argmin}\{L_{r^k}(x, \mu^k), x \in R^n\} \quad \text{e}$$

$$\mu_i^{k+1} = \mu_i^k + \nabla\theta\left(\frac{f_i(x^k)}{r^k}\right), \quad i = 1, \dots, m.$$

Novamente a regra de atualização dos multiplicadores satisfaz

$$\nabla L_{r^k}(x^{k+1}, \mu^k) = \nabla l(x^{k+1}, \mu^{k+1}) = 0.$$

Para o caso com desigualdades, o primeiro enfoque foi dado por Rockafellar [56] e [57]. Considerando o problema (P), o clássico lagrangeano aumentado surge naturalmente ao considerar folgas nas restrições. Mostra-se, ver por exemplo [8], que a minimização com respeito a estas folgas pode ser escrita em termos da variável x obtendo uma expressão para L_c em termos de x, μ e c dada por

$$x \in R^n, \mu \in R^m \mapsto L_c(x, \mu) = f_o(x) + \frac{1}{2c} \sum_{i=1}^m \{\max\{0, \mu_i + cf_i(x)\}^2 - \mu_i^2\} \quad (4.9)$$

e novamente a regra de atualização dos multiplicadores num processo iterativo pode-se obter para (4.9) satisfazendo (4.8) onde a seqüência de parâmetros positivos $\{c_k\}$ é limitada.

Na figura seguinte mostra-se o gráfico de $\frac{1}{2c}(\max\{0, \mu + cy\}^2 - \mu^2)$, que passa pelo ponto $(-\frac{\mu}{c}, -\frac{\mu^2}{2c})$ e tem na origem a inclinação μ .

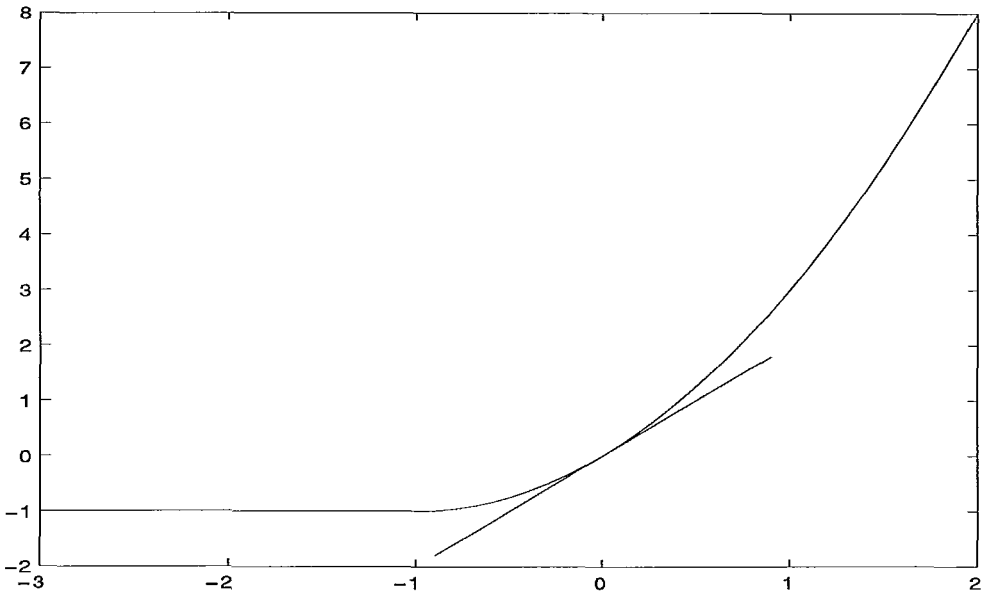


fig. 19

A principal desvantagem da função de penalidade introduzida por Rockafellar está em não ter segundas derivadas contínuas ainda que as restrições sejam de classe C^2 .

Família de penalidades (P_I) descrita por Bertsekas

Para o problema com desigualdades (P), Bertsekas [8] considera uma família de penalidades

$$y \in R, \mu \in R_+ \mapsto p(y, \mu) \in R \quad \text{tal que:}$$

i) p é contínua em $R \times R_+$, continuamente diferenciável em $R \times R_{++}$ e existe $\lim_{\mu \rightarrow 0^+} \frac{p(y, \mu) - p(y, 0)}{\mu} \quad \forall y \in R$ e $p(\cdot, 0)$ é de classe C^1 com respeito a $y \in R$.

ii) $p(y, \cdot)$ é côncava em $[0, +\infty)$ para todo y fixo.

iii) Para todo $\mu > 0$, $p(\cdot, \mu)$ é convexa em R e satisfaz a condição: (convexidade estrita)

Se $y_0 > 0$ ou $\nabla_y p(y_0, \mu) > 0$ então $p(y, \mu) - p(y_0, \mu) > (y - y_0) \nabla_y p(y_0, \mu)$ para todo $y \neq y_0$.

iv) $p(0, \mu) = 0, \quad \forall \mu > 0$.

v) $\nabla_y p(0, \mu) = \mu, \quad \forall \mu > 0$.

vi) $\lim_{\mu \rightarrow -\infty} \nabla_y p(y, \mu) = 0, \quad \forall \mu \geq 0$.

vii) $\lim_{\mu \rightarrow +\infty} \nabla_y p(y, \mu) = +\infty, \quad \forall \mu \geq 0$.

viii) $\inf_{y \in R} p(y, \mu) > -\infty, \quad \forall \mu \geq 0$.

Dentro desta família de penalidades P_I está a subfamília (P_E^+), caracterizada por funções da forma

$$y \in R, \mu \in R \mapsto p(y, \mu) = \begin{cases} \mu y + \theta(y) & \text{se } \mu + \theta'(y) \geq 0 \\ \min_{\tau \in R} \{\mu \tau + \theta(y)\} & \text{caso contrário} \end{cases}$$

onde θ pertence à família (P_E) considerada para o caso de igualdades. Para $\theta(y) = \frac{1}{2}y^2$ recupera-se a função clássica de Rockafellar com sua correspondente função lagrangeano aumentado mostrada em (4.9), pode-se escrever para todo i como

$$p(y_i, \mu_i) = \begin{cases} \mu_i y_i + \frac{1}{2} y_i^2 & \text{se } \mu_i + y_i \geq 0 \\ -\frac{1}{2} \mu_i^2 & \text{se } \mu_i + y_i < 0 \end{cases}$$

Esta subfamília será relacionada com outra na seção (4.2.1).

Também considera a família de penalidades (\hat{P}_I) que conduz a lagrangeanos aumentados duas vezes continuamente diferenciáveis que descrevemos a seguir.

Família de penalidades (\hat{P}_I) descrita por Bertsekas

Seja $p : R^2 \rightarrow R$ da forma $\mu\theta(y)$, onde $\theta : R \rightarrow R$ é de classe C^2 , $\nabla\theta(y) > 0$ para todo $y \in R$ e:

- i) $\theta(0) = 0$.
- ii) $\theta'(0) = 1$.
- iii) $\lim_{y \rightarrow -\infty} \theta(y) > -\infty$.
- iv) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$ e
- v) $\lim_{y \rightarrow \infty} \theta'(y) = \infty$.

Considerando o problema (P), o lagrangeano aumentado correspondente é dado por

$$\begin{aligned} x \in R^n, \mu \in R^m \mapsto L_r(x, \mu) &= f_0(x) + r \sum_{i=1}^m p\left(\frac{f_i(x)}{r}, \mu_i\right) \\ &= f_0(x) + r \sum_{i=1}^m \mu_i \theta\left(\frac{f_i(x)}{r}\right). \end{aligned}$$

onde $p(y, \mu) = \mu\theta(y)$ e $r \in [0, 1)$.

O algoritmo consiste em, a partir de um vetor de multiplicadores inicial $\mu^0 > 0$, gerarmos seqüências $\{x^k\}$, $\{\mu^k\}$ e $\{r^k\}$ tais que

$$x^{k+1} \in \operatorname{argmin}_x L_{r^k}(\cdot, \mu^k) \quad \text{e} \quad (4.10)$$

$$\mu_i^{k+1} = \mu_i^k \theta' \left(\frac{f_i(x^{k+1})}{r^k} \right) \quad \text{para } i = 1, \dots, m \quad (4.11)$$

A seqüência $\{r^k\}$ é decrescente e positiva podendo ser constante a partir de um valor apropriado. Observe que $\mu^k > 0$ para todo k .

Pode-se observar que a regra de atualização dos multiplicadores satisfaz

$$\nabla L_r(x^{k+1}, \mu^k) = \nabla l(x^{k+1}, \mu^{k+1}) = 0 \quad (4.12)$$

Como exemplo de uma penalidade na família (\hat{P}_I) temos o método de multiplicador exponencial determinado por

$$p(y, \mu) = \mu(\exp(y) - 1).$$

Tseng-Bertsekas [61] mostram a convergência da seqüência $\{\mu^k\}$ a uma solução ótima do problema dual (D) e para a seqüência primal $\{x^k\}$ consegue-se provar convergência ao conjunto viável num sentido ergódico, isto é, a convergência da seqüência média dada por $y^k = \frac{c^k x^k + \dots + c^0 x^0}{c^k + \dots + c^0}$ e $c^k = \frac{1}{r^k}$ para todo k . O dual da função p é a quase-distância de Kullback-Leibler (3.9), assim se abrem as portas para posteriores estudos relacionando os métodos de lagrangeano aumentado com as quase-distâncias.

Posteriormente, Polyak-Teboulle [50] consideram famílias de penalidades com características similares à família (\hat{P}_r) conseguindo incluir outras penalidades conhecidas.

Família de penalidades descritas por Polyak-Teboulle:

Considera-se o problema (P) com desigualdades do tipo \geq e a família de funções θ tais que:

$\theta : R \rightarrow R \cup \{-\infty\}$ estritamente crescente, de classe C^2 com $dom \theta = (a, +\infty)$ e $a \in [-\infty, 0)$, $\theta(a) = -\infty$, $\theta'(a) = +\infty$ satisfazendo:

- i) $\theta(0) = 0$,
- ii) $\theta'(0) = 1$,
- iii) $\theta''(y) < 0$,
- iv) $\lim_{y \rightarrow +\infty} \theta'(y) = 0$,
- v) $\lim_{r \rightarrow 0^+} r\theta\left(\frac{y}{r}\right) = 0$. para todo $y > 0$.

Entre as penalidades incluídas nesta família encontram-se:

- a) $\theta(y) = 1 - e^{-y}$, $y \in R$ (método exponencial).
- b) $\theta(y) = \log(1 + y)$, $y > -1$ (barreira modificada logarítmica).
- c) $\theta(y) = \frac{y}{1+y}$, $y > -1$ (barreira modificada hiperbólica).
- d) $\theta(y) = \begin{cases} \log(1 + y) & \text{se } y \geq -\frac{1}{2} \\ -e^{-2y-1} + 1 - \log(2) & \text{se } y \leq -\frac{1}{2}. \end{cases}$
(barreira modificada com penalidade exponencial).

Dada θ satisfazendo as propriedades i)-v) e $r \in (0, 1]$ observe que

$$y \geq 0 \text{ se e somente se } r\theta\left(\frac{y}{r}\right) \geq 0,$$

logo as restrições do problema (P) considerado, podem ser expressadas equivalentemente como

$$r\theta\left(\frac{f_i(x)}{r}\right) \geq 0 \text{ para } i = 1, \dots, m. \quad (4.13)$$

Considerando o lagrangeano clássico correspondente a minimizar f_0 sujeito a (4.13) obtemos o lagrangeano aumentado dado por

$$x \in R^n, y \in R^m \mapsto L_r(x, \mu) = f_0(x) - r \sum_{i=1}^m \mu_i \theta\left(\frac{f_i(x)}{r}\right). \quad (4.14)$$

O algoritmo usado segue o mesmo formato de (4.10) e (4.11). Esta abordagem com respeito ao lagrangeano aumentado foi chamado princípio de reescalamento não-linear e conduz naturalmente a um problema dual que envolve φ -divergências (ver seção (3.2.2)). Dada θ , define-se a função φ como

$$y \in R_{++} \mapsto \varphi(y) = -\theta^*(y).$$

um fato importante é que φ satisfaz as condições necessárias que definem uma φ -divergência (ver proposição 3.1 em [50]).

Uma hipótese feita sobre as funções θ consideradas é que θ' seja logaritmicamente convexa, ou equivalentemente que θ' satisfaça:

$$\log \theta'(t) \geq \log \theta'(t_0) + \frac{\theta''(t_0)}{\theta'(t_0)} (t - t_0) \quad \forall t > -1, t_0 \neq -1. \quad (4.15)$$

A condição anterior é usada para deduzir a condição

$$\varphi'(t) \leq \varphi''(1) \log(t), \quad \forall t > 0,$$

usada em [33] nas provas de convergência.

O principal resultado refere-se à convergência num sentido ergódico da seqüência gerada pelo lagrangeano aumentado. Isto é que todo ponto limite da seqüência de médias $\bar{x}^s = \sum_{k=1}^s \frac{x^k}{s}$ geradas pelo lagrangeano aumentado converge a uma solução ótima de (P) .

Família de penalidades descritas por Ben-Tal e Zibulevsky

Considere o problema (P) e as funções dadas por $\theta : (-\infty, b) \rightarrow R$ estritamente convexas, estritamente crescentes e de classe C^2 com $0 < b \leq \infty$ tais que:

i) $\theta(0) = 0$.

ii) $\theta'(0) = 1$.

iii) $\lim_{y \rightarrow b} \theta'(y) = \infty$.

iv) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$.

v) $\theta''(y) \geq \frac{1}{M}$ para todo $y \in [0, b]$, para algum $M > 0$.

Em [6] considera-se o lagrangeano aumentado da forma

$$x \in R^n, y \in R^m \mapsto L_r(x, \mu) = f_0(x) + \sum_{i=1}^m \mu_i r_i \theta\left(\frac{f_i(x)}{r_i}\right), \quad (4.16)$$

onde $r_i = \pi(\mu_i)$ para $i = 1, \dots, m$ e $\pi : R_{++} \rightarrow R_{++}$ é uma função não decrescente e sublinear, isto é,

$$\forall \mu > 0 : \quad \pi(\mu) \leq c\mu \text{ para algum } c > 0.$$

O algoritmo gera seqüências como em (4.10) e (4.11) adicionando a atualização

$$r^{k+1} = \pi^k(\mu^{k+1}). \quad (4.17)$$

Observe que neste caso r está em função dos multiplicadores, o que foi sugerido em [61] sem análise de convergência. No dual aparece uma quase-distância mostrada em (3.14) que é uma variação das θ -divergências consideradas em [33] (ver proposição 3.18 no capítulo 3). Com respeito à seqüência primal, a mesma tem pontos limites que pertencem ao conjunto de soluções ótimas de (P) .

Entre as penalidades que satisfazem este enfoque temos:

a) $\theta(y) = e^y - 1$. (método exponencial)

b) $\theta(y) = -\log(1 - y)$, $-\infty < y < 1$. (barreira modificada)

c) $\theta(y) = \begin{cases} y + \frac{1}{2}y^2 & \text{se } y \geq -\frac{1}{2} \\ -\frac{1}{4}\frac{1}{1-y} - \frac{7}{6} & \text{se } y \leq -\frac{1}{3} \end{cases}$ (quadrática logarítmica)

d) $\theta(y) = \begin{cases} y + \frac{1}{2}y^2 & \text{se } y \geq -\frac{1}{3} \\ \frac{32}{27}\log(-2y) - \frac{3}{8} & \text{se } y \leq -\frac{1}{3} \end{cases}$ (recíproca quadrática)

Considere novamente os problema primal (P) e dual (D) com as hipóteses:

(A1) O conjunto de soluções S^* de (P) é não vazio e compacto.

(A2) Existe $\bar{x} \in \text{dom}f$ tal que $f_i(\bar{x}) < 0$, $i = 1, \dots, m$ (Condição de qualificação de Slater).

Seja

$$\hat{f} = \inf\{f_0(x) \mid f_i(x) \leq 0, i = 1, \dots, m\}$$

e considere a família de funções φ introduzidas na seção (3.2.3) que determinam a quase-distância definida em (3.15):

$$d_\varphi(s, \mu) = \sum_{i=1}^m \mu_i \pi(\mu_i) \varphi\left(\frac{s_i}{\mu_i}\right), \quad x \geq 0, y > 0 \quad (4.18)$$

onde a função

$$\mu \in R_+ \mapsto \pi(\mu) \in R_+$$

é crescente e sublinear, isto é,

$$\forall \mu > 0 : \pi(\mu) \leq c \mu \quad \text{para algum } c > 0. \quad (4.19)$$

Considere além a seqüência $\{\mu^k\}$ gerada no seguinte processo

$$\mu^{k+1} = \operatorname{argmax}_x \{d(\mu) - d_\varphi(\mu, \mu^k)\}, \quad u^0 > 0, \quad (4.20)$$

onde d é a função dual e d_φ definida em (4.18).

Mostra-se que esta seqüência é decrescente. Uma condição adicional pedida é :

(A3) Seja d um vetor no conjunto de subgradientes de d_φ com respeito ao primeiro argumento, isto é, $d \in \partial_s d_\varphi(s, \mu)$. Para todo $\epsilon > 0$, $\exists \delta > 0$ tal que se para algum $i \in \{1, \dots, m\}$, $d_i > \epsilon$ então $d_\varphi(s, \mu) > \delta$.

A condição de sublinearidade de π e o lema a seguir são importantes para que (4.18) satisfaça a propriedade (A3).

Lema 4.7 (ver lema 2 em [6]) Se (i) $0 < y < x$ e $0 < \frac{\pi(y)}{y} < c$ para algum $c > 0$.

Então

$$\rho(x, y) \geq \frac{1}{2} \frac{[\rho'_1(x, y)]^2}{cM}, \quad \text{onde } 0 < M < \infty.$$

O teorma principal é o seguinte:

Proposição 4.8 (ver teorema 1 em [6]) Suponha que são satisfeitas (A1) e (A2) e θ satisfaz as condições i)-v). Para cada k , considere π^k verificando a condição

(4.19). Então as seqüências $\{x^k\}$ e $\{\mu^k\}$ geradas pelo método lagrangeano aumentado satisfazem:

$$i) \max\{0, f_i(x^k)\} \rightarrow 0 \quad \text{para } i = 1, \dots, m,$$

$$ii) f(x^k) \rightarrow \hat{f},$$

$$iii) \mu_i^k f_i(x^k) \rightarrow 0 \quad \text{para } i = 1, \dots, m,$$

iv) As seqüências $\{x^k\}$ e $\{\mu^k\}$ são limitadas e todos seus pontos limite são soluções dos problems (P) e (D) respectivamente.

A condição (A3) é importante para a convergência do método. A hipótese (A2) garante que a seqüência de valores duais implícita em (4.20) é não decrescente, isto junto a (A3) tornam a válida a afirmação i) do teorema anterior, que junto ao fato de que a função π é sublinear conduzem a iii) e ii).

Família de penalidades descrita por Humes-Da Silva

Humes e Silva [31]) apresentam um método de lagrangeano aumentado envolvendo penalidades com características semelhantes às penalidades da família (P_E^+) . A principal diferença é que nestas novas penalidades a derivada pela direita não precisa ser infinita. Além disso, as penalidades podem não ser separáveis. Uma desvantagem desta família de penalidades é que não são duas vezes diferenciáveis.

Consideremos novamente as funções φ introduzidas na seção (3.2.3) onde $\varphi : R^n \rightarrow R \cup \{+\infty\}$ é estritamente convexa, fechada, diferenciável na origem, cujo domínio efetivo contém uma bola fechada centrada na origem tal que $\varphi(0) = 0$.

As funções de penalidade tem a forma

$$y \in R^m \mapsto P_r(y, \mu) = \left(r\varphi(\cdot - \mu^k) + I_{R_+^m}(\cdot) \right)^* (y) \quad (4.21)$$

onde I é a função indicadora.

Exemplo:

$$a) P(y, \mu) = \frac{4}{3}y^4/3 + \mu y \text{ obtida a partir de } \varphi(y) = \frac{1}{4}y^4.$$

Dado o problema (P), o método lagrangeano aumentado proposto gera se-

seqüências $\{x^k\}$, $\{\mu^k\}$ e $\{r^k\}$ tais que

$$x^{k+1} \in \operatorname{argmin}\{f_0(x) + P_{r^k}(f(x), \mu^k) | x \in R^n\},$$

$$\mu^{k+1} \in \operatorname{argmax}\{(\mu, f_i(x^k)) - \varphi(\mu - \mu^k) | \mu \geq 0\}$$

e $\{r^k\}$ é uma seqüência de reais positivos.

Mostramos o teorema principal com respeito à convergência da seqüência $\{x^k\}$.

Teorema 4.9 *(ver teorema 6 em Humes-Da Silva [31]) Se a condição de qualificação de Slater é satisfeita e se f_0 e f_i são contínuas para $i = 1, \dots, m$, então todo ponto limite da seqüência gerada pelo lagrangeano aumentado converge a uma solução ótima do problema (P).*

Finalmente, observe que as famílias de penalidades (P_r) , as apresentadas por Polyak e Teboulle e as consideradas por Ben-Tal e Zibulevsky possuem em comum a condição de que a função de penalidade deve crescer mais rápido que uma aproximação linear. Poderíamos chamar esta propriedade coercividade lateral. No próximo capítulo relaxaremos esta propriedade usando funções de penalidade que aproximam a penalidade exata $\beta \max\{0, y\}$ com $\beta > 0$ mostrando uma interessante relação dual com os métodos que usam distâncias de Bregman. Por outra parte, consideramos uma família de funções com coercividade lateral que através de uma conveniente mudança de variáveis ou shifts nos conduzem a métodos de lagrangeano aumentado com distâncias de Bregman no dual.

4.3 Relação entre o método lagrangeano aumentado e o método de ponto proximal

No capítulo 3 introduzimos o método de ponto proximal clássico baseado na distância euclidiana sendo após generalizado com o uso de quase-distâncias, em forma semelhante fizemos um resumo das diferentes famílias de penalidades usadas nos métodos de lagrangeano aumentado e dos principais teoremas de convergência

que respaldam ambos métodos. Por uma parte, os métodos de lagrangeano aumentado podem ser implementados de forma eficiente e aplicados assim a diferentes tipos de problemas práticos, alguns pacotes computacionais para problemas de grande porte são o LANCELOT ([31]) e o BOX-QUACAN em suas diferentes versões desenvolvido na Universidade de Campinas por M.Martínez e colaboradores. Os métodos de ponto proximal não possuem estas características computacionais mas disfrutam de uma rica e elegante teoria que sustenta suas provas de convergência e as de outros métodos como é o caso do lagrangeano aumentado.

Dado o problema (P) , Rockafellar ([56]) introduz o lagrangeano aumentado clássico

$$x \in R^n, \mu \in R^m \mapsto L_c(x, \mu) = f_o(x) + \frac{1}{2c} \sum_{i=1}^m \{ \max\{0, \mu_i + cf_i(x)\}^2 - \mu_i^2 \}$$

onde $c > 0$. Consideremos a função de penalidade

$$y \in R^m, \mu \in R^m \mapsto P_c(y, \mu) = \frac{1}{2c} \sum_{i=1}^m p(cy_i, \mu_i),$$

onde

$$p(y_i, \mu_i) = \begin{cases} \mu_i y_i + \frac{1}{2} y_i^2 & \text{se } \mu_i + y_i \geq 0 \\ -\frac{1}{2} \mu_i^2 & \text{se } \mu_i + y_i < 0. \end{cases}$$

Observe que

$$P_c^*(s, \mu) = \begin{cases} \frac{1}{2c} \sum_{i=1}^m |s_i - \mu_i|^2 = \frac{1}{2c} \|s - \mu\|^2 & \text{se } s \geq 0 \\ +\infty & \text{caso contrário,} \end{cases}$$

e

$$L_c(x, \mu) = f_o(x) + P_c(f_i(x), \mu).$$

Usando o teorema de dualidade de Fenchel (4.5) e (4.4) obtemos

$$\begin{aligned} \inf_{x \in R^n} \{ f_o(x) + P_c(f_i(x), \mu) \} &= \inf_{u \in R^m} \{ v(u) + P_c(u, \mu) \} \\ &= \sup_{s \geq 0} \{ d(s) - \frac{1}{2c} \|s - \mu\|^2 \}, \end{aligned}$$

onde v é a função perturbação definida em (4.3) e d é a função dual. Assim, se sob hipóteses razoáveis, associamos o lado direito da expressão anterior com a

k -iteração do método de ponto proximal dada equivalentemente por

$$\mu^{k+1} = \operatorname{argmax}_{\mu \geq 0} \left\{ d(\mu) - \frac{1}{2c^k} \|\mu - \mu^k\|^2 \right\},$$

temos que se x^{k+1} minimiza o lagrangeano aumentado, então μ^{k+1} atinge o máximo no lado direito. Observe que esta última igualdade é equivalente a

$$\mu_i^{k+1} = \max\{0, \mu_i^k + c^k f_i(x^k)\} \quad \text{para todo } i = 1, \dots, m$$

que corresponde a atualização do multiplicador no método lagrangeano aumentado clássico. Assim se estabelece uma estreita ligação entre o método de ponto proximal clássico e o método de lagrangeano aumentado clássico para problemas com restrições de desigualdade. A convergência da seqüência gerada pelo método de ponto proximal a uma solução ótima do problema dual conduz à convergência da seqüência de multiplicadores gerados pelo método de ponto proximal clássico. Isto foi observado pela primeira vez por Rockafellar [56].

Bertsekas na seção 5.5 em [8], mostra uma relação similar para as famílias de penalidades (P_I) e (\hat{P}_I) consideradas nas seções (4.2.1) e (4.2.1) dada por

$$\begin{aligned} \inf\{L_r(x, \mu) \mid x \in R^n\} &= \inf_{x \in R^n} \{f_0(x) + r \sum_{i=1}^m p[\frac{f_i(x)}{r}, \mu_i]\} \\ &= \inf_{u \in R^m} \{v(u) + P_r(u, \mu)\} \\ &= \max_{s \in R^m} \{-v^*(-s) - P_r^*(s, \mu)\} \quad (\text{usando (4.5)}) \\ &= \max_{s \in R^m} \{d(s) - P_r^*(s, \mu)\} \quad (\text{por (4.4)}) \end{aligned}$$

onde $P_r(y, \mu) = r \sum_{i=1}^m p[\frac{y_i}{r}, \mu_i]$. A segunda igualdade pode-se obter em forma similar ao desenvolvimento feito para obter (4.4). Note que a penalidade exponencial

$$y \in R, \mu \in (0, +\infty) \mapsto p(y, \mu) = \mu(e^y - 1) \text{ está incluída na família } (\hat{P}_I).$$

Teboulle ([59]) considerou lagrangeanos aumentados construídos a partir de quase-distâncias e estabeleceu uma direta ligação entre ambas metodologias.

Consideremos novamente o nosso problema de interesse (P) junto com o seu respectivo problema dual (D) e a função lagrangeano definida como

$$x \in R^n, \mu \in R^m \mapsto L(x, \mu) = \begin{cases} f_0(x) + \sum_{i=1}^m \mu_i f_i(x) & \text{se } \mu \geq 0 \\ -\infty & \text{caso contrário.} \end{cases}$$

Teboulle define a função lagrangeano aumentado generalizado como:

$$x \in R^n, \mu \in R^m \mapsto L_r(x, s) = \sup_{s \in R_+^m} \{l(x, s) - rD(s, \mu)\}$$

onde $r > 0$ e D é uma φ -divergência ou uma distância de Bregman. Por exemplo, para o caso em que D é uma φ -divergência como definida em (3.10), define

$$\begin{aligned} L_r(x, \mu) &= \sup_{s \in R_+^m} \left\{ l(x, s) - r \sum_{i=1}^m \mu_i \varphi \left(\frac{s_i}{\mu_i} \right) \right\} \\ &= f_0(x) + \sum_{i=1}^m \sup_{s_i \in R_+} \left\{ s_i f_i(x) - r \mu_i \varphi \left(\frac{s_i}{\mu_i} \right) \right\} \\ &= f_0(x) + r \sum_{i=1}^m \mu_i \sup_{s_i \in R_+} \left\{ \frac{f_i(x)}{r} \frac{s_i}{\mu_i} - \varphi \left(\frac{s_i}{\mu_i} \right) \right\} \\ &= f_0(x) + r \sum_{i=1}^m \mu_i \varphi^* \left(\frac{f_i(x)}{r} \right). \end{aligned}$$

e sob hipóteses razoáveis, o método de multiplicadores é definido como:

Dada uma seqüência $\{r^k\}$ de números positivos, gere seqüências $\{x^k\} \subset R^n$ e $\{\mu^k\} \subset R^m$ tais que

$$x^{k+1} = \operatorname{argmin} \{L_{r^k}(x, \mu^k) : x \in R^n\},$$

$$\mu^{k+1} = \operatorname{argmax} \{(\mu, f_i(x^{k+1})) - r^k d_\varphi(\mu, \mu^k) : \mu > 0\}. \quad (4.22)$$

Observe que se μ^{k+1} satisfaz (4.22), então

$$f_i(x^{k+1}) - r^k \varphi' \left(\frac{\mu_i^{k+1}}{\mu_i^k} \right) = 0 \quad \text{para } i = 1, \dots, m$$

para assim obter

$$\begin{aligned} \mu_i^{k+1} &= \mu_i^k (\varphi')^{-1} \left(\frac{f_i(x^{k+1})}{r^k} \right) \\ &= \mu_i^k (\varphi^*)' \left(\frac{f_i(x^{k+1})}{r^k} \right) \quad \text{para } i = 1, \dots, m. \end{aligned} \quad (4.23)$$

Esta última igualdade corresponde à atualização dos multiplicadores no método de lagrangeano aumentado usado por Bertsekas [8] para a penalidade exponencial fazendo $\varphi(y) = y \log y - y + 1$.

Posteriormente, Polyak e Teboulle [50] mostram que a conjugada de φ corresponde à família de penalidades para o método de lagrangeano aumentado considerado na seção (4.2.1).

Note que a atualização (4.22) é equivalente a

$$\mu^{k+1} = \operatorname{argmax}_{s \in R_+^m} \{d(\mu) - r^k d_\varphi(\mu, \mu^k)\} \quad (4.24)$$

onde d é a função dual.

Desta forma, o iterado μ^{k+1} em (4.22) correspondente à atualização dos multiplicadores no método de lagrangeano aumentado considerado também é solução do lado direito de (4.24) que corresponde à iteração k do método de ponto proximal usando φ -divergências para minimizar a função $-d(\cdot)$ no ortante positivo.

Podemos explicitar ainda mais a relação anterior observando que, para o caso convexo e com hipóteses razoáveis, temos usando (4.23) que

Se x^{k+1} minimiza a $L_{r^k}(x, \mu^k)$ então x^{k+1} minimiza $l(x, \mu^{k+1})$, assim $d(\mu^{k+1}) = L(x^{k+1}, \mu^{k+1}) = f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} f_i(x^{k+1})$ mas

$$\begin{aligned} d(\mu) &= \min_x \{f_0(x) + \sum_{i=1}^m \mu_i f_i(x)\} \\ &\leq f_0(x^{k+1}) + \sum_{i=1}^m \mu_i f_i(x^{k+1}) \\ &= f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} f_i(x^{k+1}) + \sum_{i=1}^m (\mu_i - \mu_i^{k+1}) f_i(x^{k+1}) \\ &= d(\mu^{k+1}) + \sum_{i=1}^m (\mu_i - \mu_i^{k+1}) f_i(x^{k+1}) \end{aligned}$$

Assim, por ser d côncava temos

$$f(x^{k+1}) \in \partial d(\mu^{k+1}) \quad (4.25)$$

onde $f(\cdot) = (f_1(\cdot), \dots, f_m(\cdot))$ e $\partial d(\mu)$ é o conjunto de subgradientes em u da função côncava d (ver 4.4).

A atualização dos multiplicadores dada por (4.22) é

$$\mu_i^{k+1} = \mu_i^k (\varphi^*)' \left(\frac{f_i(x^{k+1})}{r^k} \right) \quad \text{para } i = 1, \dots, m$$

e assim usando a relação $(\varphi')^{-1} = (\varphi^*)'$ temos para $i = 1, \dots, m$

$$f_i(x^{k+1}) = r_k \varphi' \left(\frac{\mu_i^{k+1}}{\mu_i^k} \right) \quad (4.26)$$

De (4.25) e (4.26) observamos que

$$r_k \begin{pmatrix} \varphi' \left(\frac{\mu_1^{k+1}}{\mu_1^k} \right) \\ \vdots \\ \varphi' \left(\frac{\mu_m^{k+1}}{\mu_m^k} \right) \end{pmatrix} \in \partial d(\mu^{k+1})$$

ou equivalentemente

$$0 \in \partial d(\mu^{k+1}) - r_k \begin{pmatrix} \varphi' \left(\frac{\mu_i^{k+1}}{\mu_i^k} \right) \\ \vdots \\ \varphi' \left(\frac{\mu_m^{k+1}}{\mu_m^k} \right) \end{pmatrix} \quad (4.27)$$

que é equivalente a (4.24)

Iusem, Svaiter e Teboulle [33] mostram um enfoque similar com exemplos de funções φ e φ^* que satisfazem as relações anteriores. Para o caso em que D é uma distância de Bregman, além do trabalho de Teboulle, Eckstein [21] considera problemas com restrições de igualdades e problemas com restrições de desigualdade para definir métodos de lagrangeano aumentados em forma similar a Teboulle. Um estudo completo sobre a interrelação entre estas duas metodologias pode-se encontrar em Iusem [36]. Humes e Da Silva [31] relacionam o método de lagrangeano aumentado generalizado que usam as funções de penalidades mostradas em (4.21) e o método de ponto proximal definido em (3.17) via dualidade de Fenchel.

Capítulo 5

Método de multiplicadores gerado por mudanças nas penalidades

Introduzimos duas novas famílias de penalidades separáveis que através de apropriadas mudanças ou shifts conduzem a métodos de multiplicadores os quais por sua vez geram distâncias de Bregman no dual. Isto representa um novo enfoque no contexto deste tipo de métodos. Em primeira instância consideramos uma família de penalidades que aproximam a função de penalidade exata e posteriormente incorporamos uma família de penalidade coercivas pela direita. Estas famílias estão em correspondência respectivamente com as penalidades tipo I e tipo II introduzidas no capítulo 2.

Consideremos novamente o problema

$$\begin{aligned} & \text{minimizar } f_o(x) \\ (P) \quad & \text{sujeito a: } f_i(x) \leq 0, \quad i = 1, \dots, m. \\ & x \in R^n \end{aligned}$$

onde $f_i : R^n \rightarrow R \cup \{+\infty\}$ para $i = 0, 1, \dots, m$ são funções convexas próprias fechadas e S^* é o conjunto de soluções ótimas de (P) .

Suponhamos válidas as hipóteses (A1) e (A2) introduzidas no capítulo 4:

(A1) S^* é não vazio e compacto.

(A2) Existe $\bar{x} \in \text{dom} f$ tal que $f_i(\bar{x}) < 0$, $i = 1, \dots, m$ (Condição de Slater).

Para $f : R^n \rightarrow R \cup \{+\infty\}$ convexa própria fechada consideramos de novo sua função de recessão como:

$$d \in R^n \mapsto f'_\infty(d) = \lim_{\lambda \rightarrow +\infty} \frac{f(x + \lambda d) - f(x)}{\lambda}, \quad (5.1)$$

onde x é arbitrário em $\text{dom} f$.

Definição 5.1 *Uma direção $d \neq 0$ é uma direção de recessão de f se $f'_\infty(d) \leq 0$.*

É conhecido que (A1) é equivalente a:

$$(f_i)'_\infty(d) \leq 0, \quad \forall i = 0, 1, \dots, m. \Rightarrow d = 0, \quad (5.2)$$

isto é, que não existem direções de recessão não nulas.

5.1 Famílias de penalidades AL1 e AL2

5.1.1 Família de penalidades não coercivas

Consideremos inicialmente uma função $\theta : R \rightarrow R$ convexa e crescente tal que:

- i) θ é estritamente convexa em R .
- ii) θ é diferenciável em R .
- iii) $\theta'_\infty(1) = 1$. (0-coerciva pela direita na classificação dada em [28])
- iv) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$.
- v) $\theta(\cdot)$ limitada por baixo.
- vi) $\sup_{y > 0} \{y - \theta(y)\}$ é finito.

Comentários:

Observe que θ é estritamente crescente, própria e fechada.

Pelas condições i), ii) e iv) deduzimos que $\theta_\infty(-1) = 0$. (ver lema 1 em ([6]).

A condição v) é equivalente a $\theta^*(0)$ ser finito. (ver teorema 27.1.a em ([54]).

De igual forma vi) é dada para garantir a existência de $\theta^*(1)$. Estas duas condições garantem que $\theta(\cdot)$ não se afasta infinitamente da função y^+ . Entre as famílias de penalidades consideradas no capítulo 4, a condição iii) é satisfeita somente por algumas funções de penalidades consideradas por Humes e Silva [31], mas com a desvantagem de não ser duas vezes diferenciáveis.

5.1.2 Família de penalidades coercivas pela direita

Consideremos a função $\theta : R \rightarrow R \cup \{+\infty\}$ convexa, própria, fechada, crescente e diferenciável com dom $\theta = (-\infty, b)$, $b \in R_{++}$ tal que:

- i) θ é estritamente convexa em $(-\infty, b)$.
- ii) θ é diferenciável em $(-\infty, b)$.
- iii) $\theta'_\infty(1) = +\infty$. (coercividade pela direita)
- iv) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$.
- v) $\theta(\cdot)$ limitada por baixo.

Note que a diferença principal entre estas duas famílias de penalidades está no comportamento no infinito pela direita, ambas correspondem basicamente às penalidades tipo I e tipo II definidas no capítulo II. Suas funções de recessão são respectivamente a função de penalidade exata e a função indicadora do ortante negativo. As penalidades tipo I e tipo II darão origem respectivamente às famílias de penalidades que denominaremos AL1 e AL2.

O desenvolvimento que apresentamos é válido para ambas famílias de penalidades incorporando um parâmetro positivo $\beta \geq 1$ no caso da família AL1, cuja função é a de aumentar gradualmente a penalidade quando for necessário e que para o caso da família AL2 não é utilizado. Note que com estas condições $\theta(\cdot)$ é um homeomorfismo sobre dom θ pois sua inversa existe e é contínua.

Mudança de variáveis.

Consideremos a função de penalidade θ do tipo I. Para $i = 1, \dots, m$, dado $\beta_i > 0$ e $\mu_i \in (0, \beta)$, defina a função $\tilde{y}_i(\cdot)$ como

$$\mu_i \in (0, \beta) \mapsto \tilde{y}_i(\mu_i) = (\theta')^{-1} \left(\frac{\mu_i}{\beta_i} \right) \quad \text{para } i = 1, \dots, m. \quad (5.3)$$

Onde não houver confusão possível faz-se o abuso de notação $\tilde{y}_i = \tilde{y}_i(\mu_i)$, assim escreveremos (5.3) equivalentemente como

$$\theta'(\tilde{y}_i) = \frac{\mu_i}{\beta_i} \quad \text{para } i = 1, \dots, m. \quad (5.4)$$

Assim, para cada valor de $\mu_i \in (0, \beta)$ e $\beta_i \geq 1$ está associado um valor de \tilde{y}_i que satisfaz (5.4). Isto determina uma importante mudança de variáveis em nossa

abordagem.

Similarmente, para o caso da penalidade do tipo II, dados $\beta_i \geq 1$ e $\mu_i \in R_{++}^m$ para $i = 1, \dots, m$, defina a função $\tilde{y}_i(\cdot)$ como

$$\mu_i \in R_{++} \mapsto \tilde{y}_i(\mu_i) = (\theta')^{-1}(\mu_i) \quad \text{para } i = 1, \dots, m. \quad (5.5)$$

Novamente, onde não houver confusão possível faz-se o abuso de notação $\tilde{y}_i = \tilde{y}_i(\mu_i)$, assim escreveremos (5.5) equivalentemente como

$$\theta'(\tilde{y}_i) = \mu_i \quad \text{para } i = 1, \dots, m. \quad (5.6)$$

É importante destacar que estas mudanças de variável ou *shifts*, permitem que cada função na respectiva família tenha na origem derivada igual a μ que é uma característica importante quando analisamos o que acontece na sua conjugada. Isto mostra-se na seqüência a seguir onde usamos a notação

$$(0, \beta) = \{z \in R^m \mid 0 < z < \beta\}$$

Família de penalidades AL1.

Dados $\beta \in R_{++}^m$, $r \in (0, 1]$ fixos, definimos uma penalidade $P_{\beta,r}$ pertencente à família de penalidades AL1 como

$$\begin{aligned} y \in R^m, \mu \in (0, \beta) \mapsto P_{\beta,r}(y, \mu) &= \sum_{i=1}^m p_{\beta_i,r}(y_i, \mu_i) \\ &= \sum_{i=1}^m \beta_i r \left[\theta\left(\frac{y_i}{r} + \tilde{y}_i(\mu_i)\right) - \theta(\tilde{y}_i(\mu_i)) \right], \end{aligned} \quad (5.7)$$

onde θ é uma penalidade do tipo I, e $\tilde{y} = \tilde{y}(\mu)$ satisfaz

$$\theta'(\tilde{y}_i) = \frac{\mu_i}{\beta_i} < 1$$

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

Observação:

Normalmente β , r e μ são fixados e conseqüentemente $\tilde{y}_i = \tilde{y}_i(\mu_i)$ também é fixo. Opera-se então com as funções $P_{\beta,r}(\cdot, \mu)$ e $p_{\beta_i,r}(\cdot, \mu_i)$. Nestes casos usaremos a notação simplificada $P_1(\cdot) = P_{\beta_i,r}(\cdot, \mu)$ e $p_{1_i}(\cdot) = p_{\beta_i,r}(\cdot, \mu_i)$.

Observe que para $i = 1, \dots, m$, β_i , μ_i e r fixos, temos

$$p_{1_i}'(y_i) = \beta_i \theta' \left(\frac{y_i}{r} + \tilde{y}_i \right),$$

assim, por (5.4)

$$p_{1_i}'(0) = \beta_i \theta'(\tilde{y}_i) = \mu_i \quad \text{para } i = 1, \dots, m$$

e também

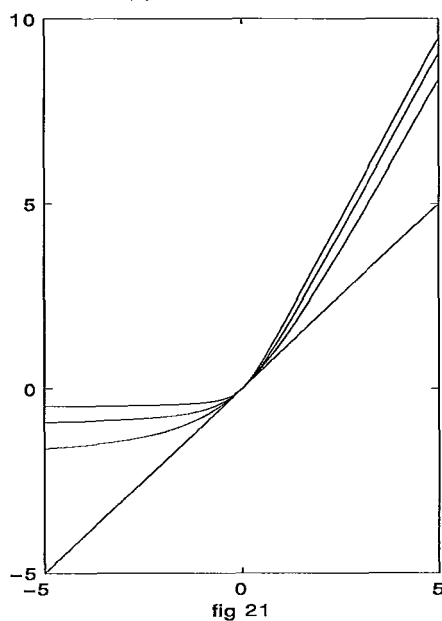
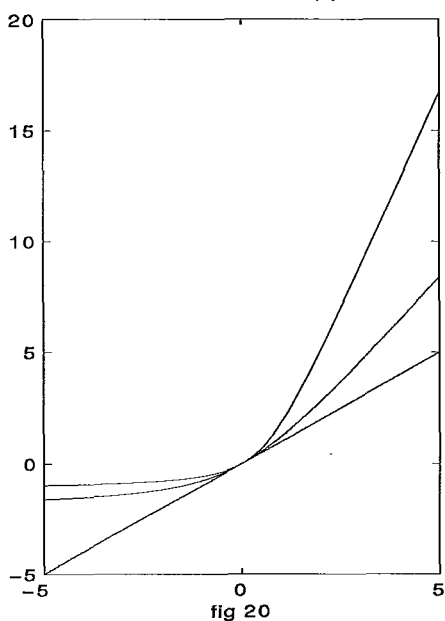
$$p_{1_i}(0, \mu_i) = 0 \quad \text{para } i = 1, \dots, m.$$

Note que se para $i = 1, \dots, m$ $y_i \leq 0 \Leftrightarrow p_{1_i}(y_i) \leq 0$.

De fato,

$$\begin{aligned} y_i \leq 0 &\Leftrightarrow \frac{y_i}{r} + \tilde{y}_i \leq \tilde{y}_i \\ &\Leftrightarrow \theta\left(\frac{y_i}{r} + \tilde{y}_i\right) \leq \theta(\tilde{y}_i) \\ &\Leftrightarrow \beta_i r [\theta\left(\frac{y_i}{r} + \tilde{y}_i\right) - \theta(\tilde{y}_i)] \leq 0 \end{aligned}$$

A fig.20 mostra a variação de $p_1(y)$ para μ, r fixos, e β aumentando, o efeito correspondente é penalizar pontos inviáveis. Na fig. 21 observamos que quando decrescemos o parâmetro r mantendo β e μ fixos estamos aproximando a função barreira exata (isto é, a função $\beta \max\{0, x\}$ para $x \in R$). A variação simultânea de ambos parâmetros produz ambos efeitos. Em ambas figuras fixamos como referência a reta $y = p'_{\beta,r}(0, 1)x$ e notamos que $p'_{\beta,r}(0, 1) = 1$.



Alguns exemplos de funções nesta família AL1 são:

a) Para $\theta(y) = \frac{1}{2}(y + \sqrt{y^2 + 4} - 2)$, temos

$$p_{\beta,r}(y, \mu) = \frac{1}{2}\beta r \left[\frac{y}{r} + \tilde{y} + \sqrt{\left(\frac{y}{r} + \tilde{y}\right)^2 + 4} - \left(\tilde{y} + \sqrt{\tilde{y}^2 + 4}\right) \right]$$

com $\tilde{y} = \frac{2\mu - b}{\sqrt{2\mu b - \mu^2}}$.

b) Para $\theta(y) = \ln(1 + e^y) - \ln 2$ temos

$$p_{\beta,r}(y, \mu) = \beta r \ln \left(\frac{1 + e^{\left(\frac{y}{r} + \tilde{y}\right)}}{1 + e^{\tilde{y}}} \right)$$

com $\tilde{y} = \ln\left(\frac{\mu}{\beta - \mu}\right)$.

Família de penalidades AL2.

De modo análogo, dado $r \in (0, 1]$, definimos uma penalidade P_r pertencente à família de penalidades AL2 como

$$\begin{aligned} y \in R^m, \mu \in R_{++}^m &\mapsto P_r(y, \mu) = \sum_{i=1}^m p_r(y_i, \mu_i) \\ &= \sum_{i=1}^m r \left[\theta\left(\frac{y_i}{r} + \tilde{y}_i(\mu_i)\right) - \theta(\tilde{y}_i(\mu_i)) \right], \end{aligned} \quad (5.8)$$

onde θ é penalidade do tipo II, e $\tilde{y}_i(\mu_i)$ satisfaz

$$\theta'(\tilde{y}_i(\mu_i)) = \mu_i$$

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

Similarmente, quando r e μ são fixados, $\tilde{y}_i = \tilde{y}_i(\mu_i)$ também é fixo. Opera-se então com as funções $P_r(\cdot, \mu)$ e $p_r(\cdot, \mu_i)$. Nestes casos usaremos a notação simplificada $P_2(\cdot) = P_r(\cdot, \mu)$ e $p_{2i}(\cdot) = p_r(\cdot, \mu_i)$.

Como exemplos nesta família de funções AL2 temos:

a) Para $\theta(y) = e^y$,

$$\begin{aligned} p_r(y, \mu) &= r[e^{\frac{y}{r} + \tilde{y}} - e^{\tilde{y}}] \\ &= r e^{\tilde{y}}(e^{\frac{y}{r}} - 1) \\ &= r\mu(e^{\frac{y}{r}} - 1) \end{aligned}$$

onde $\tilde{y} = \ln(\mu)$. Note que, se usarmos $\theta(y) = e^y + C$ com C constante, no lugar de $\theta(y) = e^y$, obtemos a mesma função de penalidade p .

b) Para $\theta(y) = \begin{cases} e^y & \text{se } y \leq 0 \\ 1 - \log(1 - y) & \text{se } y \geq 0, \end{cases}$ temos

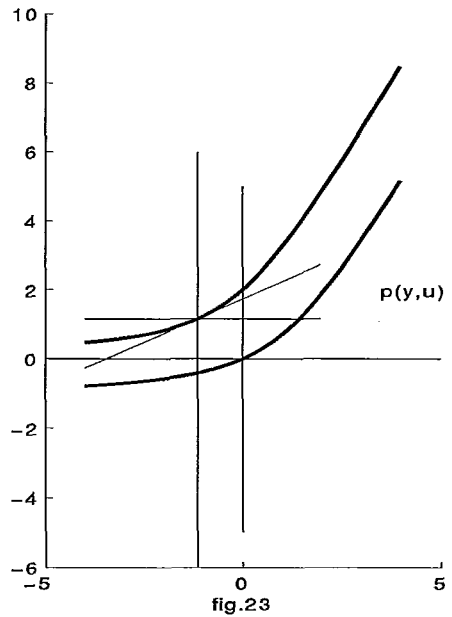
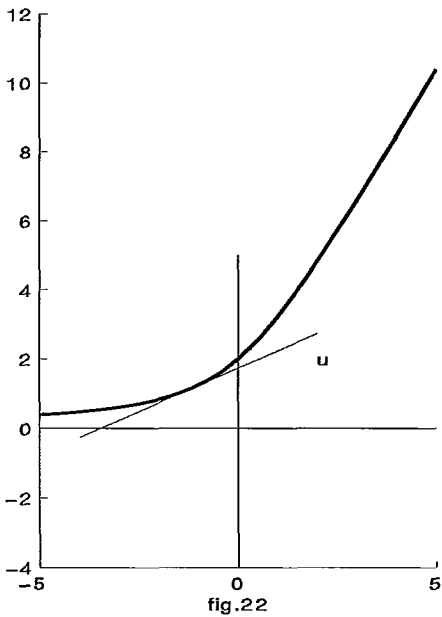
$$p_r(y, \mu) = \begin{cases} r\mu(e^{\frac{y}{r}} - 1) & \text{se } y \leq 0 \\ r(-\log(1 - y/r - \tilde{y}) + \log(1 - \tilde{y})) & \text{se } y \geq 0 \end{cases}$$

com $\tilde{y} = 1 - 1/\mu$.

Assim $p_r(y, \mu) = \begin{cases} r\mu(e^{\frac{y}{r}} - 1) & \text{se } y \leq 0 \\ -r\log(1 - \mu\frac{y}{r}) & \text{se } y \geq 0. \end{cases}$

Geometria dos shifts.

Na fig.22 mostra-se o gráfico de uma penalidade θ e uma reta tangente a θ com uma inclinação μ . O ponto de tangência corresponde a um ponto de abscissa \tilde{y} tal que $\theta'(\tilde{y}) = \mu$. Na fig.23 ilustramos o fato de que o shift equivale a transladar a origem ao ponto $(\tilde{y}, \theta(\tilde{y}))$. Finalmente, na mesma figura mostramos a penalidade $p(\tilde{y}, \mu)$ passando pela origem com derivada μ .



Observa-se que a diferença entre estas famílias de penalidades e as famílias consideradas no capítulo 4 consiste, basicamente, no comportamento pela direita das penalidades. Também pode-se notar diferenças específicas com cada família comparando o domínio de definição da penalidade θ , a limitação por baixo da mesma e a sua diferenciabilidade. É importante ressaltar que a introdução do

shift dá origem a distâncias de Bregman via dualidade de Fenchel.

5.2 Método de multiplicadores baseados em shifts

Uma das características principais presentes nos métodos de Lagrangeano aumentado definidos no capítulo anterior é que a derivada na origem das funções de penalidade envolvidas é igual a μ . Este fato é satisfeito pelas famílias $AL1$ e $AL2$ como consequência dos shifts, assim podemos definir para cada uma destas famílias, métodos de lagrangeano aumentado como métodos de multiplicadores baseados em shift. Note que em nenhuma das duas famílias de penalidades pedimos a condição $\theta'(0) = 1$.

5.2.1 Método de multiplicadores baseados em shifts para a família $AL1$

Dados $\beta \geq 1$, $r \in (0, 1]$, consideremos a função penalizada

$$x \in R^n, \mu \in (0, \beta) \mapsto L_{\beta,r}(x, \mu) = f_0(x) + \sum_{i=1}^m \beta_i r \left[\theta \left(\frac{f_i(x)}{r} + \tilde{y}_i(\mu_i) \right) - \theta(\tilde{y}_i(\mu_i)) \right],$$

com θ pertencente à família de penalidades tipo I, e \tilde{y} como definido em (5.3) satisfaz

$$\theta'(\tilde{y}_i(\mu_i)) = \frac{\mu_i}{\beta_i} < 1 \quad \text{para } i = 1, \dots, m.$$

Em geral essa função é utilizada com β, r e μ fixos, e denotaremos $L(\cdot) = L_{\beta,r}(\cdot, \mu)$. Destacamos que μ corresponde efetivamente aos multiplicadores de lagrange e aparece em forma implícita dentro da definição de \tilde{y} . Assim, o papel que β e r desempenham pode ser considerado como secundário.

O método é definido pelo algoritmo a seguir, que será comentado abaixo. Aqui suporemos que o cálculo em (5.9) sempre é bem definido. Mostraremos adiante que isto é verdade se β^0 for suficientemente grande.

Algoritmo 5.1 *Dados: $\beta^0 \geq e$, $r^0 = 1$, $\mu^0 \in (0, \frac{\beta^0}{2})$ e \tilde{y}^0 tal que $\theta'(\tilde{y}^0) = \mu^0/\beta^0$.*

Para $k=0, 1, \dots$

Encontre

$$x^{k+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \left\{ f_0(x) + \sum_{i=1}^m \theta_i^k (f_i(x)) \right\} \quad (5.9)$$

$$\text{com } y_i \in \mathbb{R} \mapsto \theta_i^k(y_i) = \beta_i^k r^k \left(\theta\left(\frac{y_i}{r^k} + \tilde{y}_i^k\right) - \theta(\tilde{y}_i^k) \right).$$

Atualize

$$\mu_i^{k+1} = \beta_i^k \theta' \left(\frac{f_i(x^{k+1})}{r^k} + \tilde{y}_i^k \right) \text{ para } i = 1, \dots, m. \quad (5.10)$$

Para $i = 1, \dots, m$: se $\mu_i^{k+1} > \frac{\beta_i^k}{2}$ então $\beta_i^{k+1} = 2\beta_i^k$ e

calcule \tilde{y}_i^{k+1} tal que $\theta'(\tilde{y}_i^{k+1}) = \frac{\mu_i^{k+1}}{\beta_i^{k+1}}$, senão

$$\beta_i^{k+1} = \beta_i^k \text{ e } \tilde{y}_i^{k+1} = \tilde{y}_i^k + y_i^{k+1}/r^k, \text{ onde } y_i^{k+1} = f_i(x^{k+1})$$

Escolha: $r^{k+1} \leq r^k$.

Comentários:

- Por definição de \tilde{y}_i^{k+1} , $\theta'(\tilde{y}_i^{k+1}) = \frac{\mu_i^{k+1}}{\beta_i^{k+1}}$,

da atualização do multiplicador temos, para $i = 1, \dots, m$

$$\frac{\mu_i^{k+1}}{\beta_i^k} = \theta' \left(\frac{y_i^{k+1}}{r^k} + \tilde{y}_i^k \right),$$

segue-se que se $\beta_i^{k+1} = \beta_i^k$ então $\tilde{y}_i^{k+1} = \tilde{y}_i^k + \frac{y_i^{k+1}}{r^k}$ onde $y_i^{k+1} = f_i(x^{k+1})$.

- Observe que pela definição de x^{k+1} , propriedades de subdiferencial e definição de L temos:

$$\begin{aligned} \partial_x L_{\beta,r}(x^{k+1}, \mu^k) &= \partial f_0(x^{k+1}) + \sum_{i=1}^m \beta_i^k \theta' \left(\frac{f_i(x^{k+1})}{r} + \tilde{y}_i^k \right) \partial f_i(x^{k+1}) \\ &= \partial f_0(x^{k+1}) + \sum_{i=1}^m \mu_i^{k+1} \partial f_i(x^{k+1}) \\ &= \partial_x l(x^{k+1}, \mu^{k+1}). \end{aligned}$$

Logo cumpre-se que

$$0 \in \partial_x L_{\beta,r}(x^{k+1}, \mu^k) \Leftrightarrow 0 \in \partial_x l(x^{k+1}, \mu^{k+1}), \quad (5.11)$$

o que constitui uma característica importante determinada pela atualização do multiplicador.

- Devido a que $\mu^0 > 0$ e θ é estritamente crescente temos que $\mu^k > 0$ para todo k .

- A atualização de β^k obedece a que o número de vezes em que aumenta é finito, devido a que mostramos adiante que a seqüência $\{\mu^k\}$ é limitada. Assim o cálculo do valor de \tilde{y}_i^{k+1} a partir de $\theta'(\cdot)$ só será necessário em algumas iterações. Nessas iterações, para encontrar o valor de \tilde{y}^{k+1} podemos resolver $\theta'(\tilde{y}_i^{k+1}) = \frac{\mu_i^{k+1}}{\beta_i^{k+1}}$ para $i = 1, \dots, m$ usando, por exemplo, o método de bissecção. Em ocasiões é possível encontrar uma expressão algébrica para \tilde{y}_i^{k+1} como $\tilde{y}_i^{k+1} = (\theta')^{-1}\left(\frac{\mu_i^{k+1}}{\beta_i^{k+1}}\right)$.
- Se $f_i(x^k) = 0$ para algum i e algum k , então $\mu_i^{k+1} = \beta_i^k \theta'(\tilde{y}_i^k) = \mu_i^k$.
- Se x^{k+1} é inviável para a restrição i , pelas características da penalidade p_1 inferimos que μ_i^{k+1} cresce e assim o efeito do shift é tal que traslada a origem a um ponto onde a penalidade aumenta.

5.2.2 Método de multiplicadores baseados em shifts para a família AL2

Diversamente da família AL1, chamamos a atenção que nesta família AL2 não existe o parâmetro β . Assim, para $r \in (0, 1]$ consideremos a função penalizada

$$x \in R^n, \mu \in R_{++}^m \mapsto L_r(x, \mu) = f_0(x) + \sum_{i=1}^m r \left[\theta \left(\frac{f_i(x)}{r} + \tilde{y}_i(\mu_i) \right) - \theta(\tilde{y}_i(\mu_i)) \right]$$

onde θ pertence à família de penalidades tipo II e \tilde{y} como definido em (5.3) satisfaz

$$\theta'(\tilde{y}_i(\mu_i)) = \mu_i \quad \text{para } i = 1, \dots, m.$$

O método é definido pelo algoritmo :

Algoritmo 5.2 Dados $r^0 = 1$, $\mu^0 = \theta'(0)$, $\tilde{y}^0 = 0$.

Para $k=0, 1, \dots$

Encontre

$$x^{k+1} \in \operatorname{argmin}_{x \in R^n} \left\{ f_0(x) + \sum_{i=1}^m \theta_i^k(f_i(x)) \right\} \quad (5.12)$$

$$\text{com } y_i \in R \mapsto \theta_i^k(y_i) = r^k \left(\theta \left(\frac{y_i}{r^k} + \tilde{y}_i^k \right) - \theta(\tilde{y}_i^k) \right) \text{ para } i = 1, \dots, m.$$

Calcule

$$\tilde{y}^{k+1} = \tilde{y}^k + y^{k+1}/r^k, \quad (5.13)$$

onde $y_i^{k+1} = f_i(x^{k+1})$ para $i = 1, \dots, m$.

Atualize

$$\mu_i^{k+1} = \theta'(\tilde{y}_i^{k+1}) \text{ para } i = 1, \dots, m. \quad (5.14)$$

Escolha $r^{k+1} \leq r^k$.

A diferença principal entre este algoritmo e o anterior consiste em que para a família *AL2* não existe o parâmetro β . Os comentários feitos para o algoritmo correspondente à família *AL1* também são válidos neste caso eliminando as observações que envolvem o parâmetro β . O parâmetro r é responsável pela precisão da aproximação mas pode se manter fixo em ambos casos. Neste caso, destacamos que o cálculo do valor \tilde{y}^k a partir de $\theta'(\cdot)$ nunca é necessário.

Mostraremos a seguir que as minimizações (5.9) e (5.12) estão bem definidas.

Lema 5.2 *Considere θ do tipo I satisfazendo i)-vi) . Seja $f : R^n \rightarrow R \cup \{+\infty\}$ uma função convexa fechada. Sejam $r \in (0, 1]$, $\beta \geq 1$, $\mu > 0$ fixos e $\tilde{y} \in R$ tal que $\theta'(\tilde{y}) = \frac{\mu}{\beta} < \frac{1}{2}$.*

Consideremos

$$p_1(x) = \begin{cases} \beta r \left[\theta\left(\frac{f_i(x)}{r} + \tilde{y}\right) - \theta(\tilde{y}) \right] & \text{se } x \in \text{dom}(f) \\ \infty & \text{caso contrário} \end{cases}$$

então $\forall d \in R^n$, $d \neq 0$ cumpre-se

$$(p_1)'_{\infty}(d) = \begin{cases} \beta \theta'_{\infty}(f'_{\infty}(d)) & \text{se } d \in \text{dom}(f'_{\infty}) \\ \infty & \text{caso contrário.} \end{cases}$$

Prova:

Para β, μ e r fixos, escrevemos p como

$$p_1(x) = \begin{cases} \beta r \theta(h(x)) + C & \text{se } x \in \text{dom } f \\ +\infty & \text{caso contrário} \end{cases}$$

onde $h(x) = \frac{f_i(x)}{r} + \tilde{y}$ e $C = \beta r \theta(\tilde{y})$.

Note que h é convexa, fechada e $\text{dom } h = \text{dom } f$. Usando a definição de função de recessão (5.1), temos $\forall d \in R^n, d \neq 0$

$$h'_\infty(d) = \frac{f'_\infty(d)}{r}.$$

Assim usando proposição 2.1 em [3] sobre a função de recessão de funções compostas e o fato que θ'_∞ é positivamente homogênea obtemos

$$\begin{aligned} (p_1)'_\infty(d) &= \begin{cases} \beta r \theta'_\infty(h'_\infty(d)) & \text{se } d \in \text{dom}(h'_\infty) \\ +\infty & \text{caso contrário} \end{cases} \\ &= \begin{cases} \beta \theta'_\infty(f'_\infty(d)) & \text{se } d \in \text{dom}(f'_\infty) \\ \infty & \text{caso contrário.} \end{cases} \end{aligned}$$

■

Lema 5.3 *Considere θ do tipo II e satisfazendo i)-v). Seja $f : R^n \rightarrow R \cup \{+\infty\}$ uma função convexa fechada. Sejam $r \in (0, 1]$, $\mu > 0$ fixos e $\tilde{y} \in \text{ri dom } \theta$ tal que $\theta'(\tilde{y}) = \mu$.*

Considere

$$p_2(x) = \begin{cases} r \left[\theta\left(\frac{f_i(x)}{r} + \tilde{y}\right) - \theta(\tilde{y}) \right] & \text{se } x \in \text{dom}(f) \\ \infty & \text{caso contrário} \end{cases}$$

Se $\text{dom } p \neq \emptyset$, então para $d \in R^n, d \neq 0$, temos

$$(p_2)'_\infty(d) = \begin{cases} \theta'_\infty(f'_\infty(d)) & \text{se } d \in \text{dom}(f'_\infty) \\ \infty & \text{caso contrário.} \end{cases}$$

Prova: Similar ao lema anterior omitindo o parâmetro β .

■

Proposição 5.4 *Suponha satisfeita a hipótese A1, θ penalidade do tipo II satisfazendo as condições i)-v), $r \in (0, 1]$ e para cada $i = 1, \dots, m$ sejam $\mu_i > 0$ e $\tilde{y}_i \in \text{ri dom } \theta$ tais que $\theta'(\tilde{y}_i) = \mu_i$. Então o conjunto solução do problema*

$$\text{argmin}_x L_r(x, \mu)$$

é não vazio e compacto.

Prova:

Considere $r \in (0, 1]$ e para cada $i = 1, \dots, m$ sejam $\mu_i > 0$ e $\tilde{y}_i \tilde{y}_i \in \text{ri dom } \theta$ tais que $\theta'(\tilde{y}_i) = \mu_i$.

Considere além disso

$$L_r(x, \mu) = f_0(x) + \sum_{i=1}^m r \left[\theta \left(\frac{f_i(x)}{r} + \tilde{y}_i \right) - \theta(\tilde{y}_i) \right].$$

Queremos mostrar que para $i = 1, \dots, m$, $L_r(\cdot, \mu)$ não tem direções de recessão, isto é,

$$\forall d \in R^n, \quad d \neq 0, \quad (L_r)'_{\infty}(d) > 0$$

ou equivalentemente o conjunto de minimizadores de $L_r(\cdot, \mu)$ é não vazio e compacto (ver [54] teorema 27.1.d).

Denotemos $\forall d \in R^n, \quad d \neq 0, \quad I_{\infty}^+(d)$ o conjunto de índices $i = 1, \dots, m$ tais que $(f_i)'_{\infty}(d) > 0$ para $d \neq 0$. Similarmente definamos $I_{\infty}^-(d) = \{i \mid (f_i)'_{\infty}(d) \leq 0\}$. Além disso, denotemos $\mathcal{K} = \{d \in R^n \mid d \in \cap \text{dom}(f_i)'_{\infty}, \quad \forall i = 0, 1, \dots, m\}$.

Por propriedades sobre função de recessão para adição de funções (ver teorema 9.3 em [54]), propriedade sobre função de recessão de funções compostas (ver proposição 2.1 em [3]) e lema 5.2 temos $\forall d \in \mathcal{K}, \quad d \neq 0$

$$(L_r)'_{\infty}(d) = \begin{cases} (f_0)'_{\infty}(d) + \sum_{i=1}^m \theta'_{\infty}((f_i)'_{\infty}(d)) & \text{se } d \in \mathcal{K} \\ \infty & \text{caso contrário} \end{cases}$$

Mas para $d \neq 0, \quad d \in \mathcal{K}$ temos

$$\sum_{i=1}^m \theta'_{\infty}((f_i)'_{\infty}(d)) = \sum_{i \in I_{\infty}^+(d)} (f_i)'_{\infty}(d) \theta'_{\infty}(1) + \sum_{i \in I_{\infty}^-(d)} (f_i)'_{\infty}(d) \theta'_{\infty}(-1)$$

Logo, usando as propriedades i),ii),iii) e iv) satisfeitas por θ concluímos

$$(L_r)'_{\infty}(d) = \begin{cases} (f_0)'_{\infty}(d) & \text{se } (f_i)'_{\infty}(d) \leq 0 \quad \forall i = 1, \dots, m \\ \infty & \text{caso contrário.} \end{cases}$$

Analisando esta expressão concluímos que se $(f_0)'_{\infty}(d) \leq 0$ contradiz a hipótese (A1) pois existiria $d \neq 0$ tal que $(f_i)'_{\infty}(d) \leq 0 \quad \forall i = 0, 1, \dots, m$.

Logo podemos afirmar $(f_0)'_{\infty}(d) > 0$ para todo $d \neq 0$, e conseqüentemente $(L_r)'_{\infty}(d) > 0 \quad \forall d \in \mathcal{K}, \quad d \neq 0$.

■

Proposição 5.5 *Considere θ do tipo I satisfazendo as condições i)-vi). Suponha satisfeita a hipótese A1 e $r \in (0, 1]$ fixo. Existe $\bar{\beta} \geq e$ tal para todo $\beta > \bar{\beta} \geq e$ e $\mu \in (0, \beta)$, o conjunto*

$$\operatorname{argmin}_x L_{\beta,r}(x, \mu)$$

é não vazio e compacto. Se $(f_0)_\infty$ é não negativo a conclusão é válida para todo $\beta \geq e$.

Prova:

Considere $r \in (0, 1]$ fixo, $\mu \in (0, \beta)$ arbitrário e para $i = 1, \dots, m$, \tilde{y}_i tais que $\theta'(\tilde{y}_i) = \mu_i/\beta_i \leq 1$.

Considere além disso

$$L(x) \equiv L_{\beta,r}(x, \mu) = f_0(x) + \sum_{i=1}^m \beta_i r \left[\theta \left(\frac{f_i(x)}{r} + \tilde{y}_i \right) - \theta(\tilde{y}_i) \right].$$

Novamente, por propriedades sobre função de recessão para adição de funções (ver teorema 9.3 em [54]), propriedade sobre função de recessão de funções compostas (ver proposição 2.1 em [3]) e lema 5.3 temos, sem perda de generalidade, $\forall d \in \mathcal{K}$ com $\|d\| = 1$ e \mathcal{K} definido como na proposicao anterior

$$L'_\infty(d) = \begin{cases} (f_0)'_\infty(d) + \sum_{i=1}^m \beta_i \theta'_\infty((f_i)'_\infty(d)) & \text{se } d \in \mathcal{K} \\ \infty & \text{caso contrário} \end{cases}$$

e assim, como na demonstração da proposição anterior, para $\|d\| = 1$, $d \in \mathcal{K}$, temos $\forall \beta \geq e$, $\mu \in (0, \beta)$,

$$\sum_{i=1}^m \beta_i \theta'_\infty((f_i)'_\infty(d)) = \sum_{i \in I_\infty^+(d)} \beta_i (f_i)'_\infty(d) \theta'_\infty(1) + \sum_{i \in I_\infty^-(d)} \beta_i (f_i)'_\infty(d) \theta'_\infty(-1)$$

Assim, usando as propriedades i),ii,iii) e iv) de θ concluímos para

$$\|d\| = 1, \quad d \in \mathcal{K}$$

$$(L)'_\infty(d) = \begin{cases} a) (f_0)'_\infty(d) & \text{se } (f_i)'_\infty(d) \leq 0 \quad \forall i \\ b) (f_0)'_\infty(d) + \sum_{i \in I_\infty^+(d)} \beta_i (f_i)'_\infty(d) & \text{se } I_\infty^+(d) \neq \emptyset \\ c) \infty & \text{se } d \notin \mathcal{K} \end{cases}$$

Analisando cada literal nesta última expressão temos:

1) Se acontece a) é claro que $(f_0)'_{\infty}(d) > 0$, caso contrário se contradiz a hipótese (A1) e assim

$$(L_{\beta,r})'_{\infty}(\cdot) > 0 \quad \forall \beta \geq e.$$

2) De b), se $(f_0)'_{\infty}(d) > 0$, então novamente

$$(L_{\beta,r})'_{\infty}(\cdot) > 0 \quad \forall \beta \geq e.$$

S não vazio e compacto é equivalente a

$$(f_i)'_{\infty}(d) \leq 0 \quad \forall i = 1, \dots, m \quad \Rightarrow d = 0.$$

Seja $(f_0)'_{\infty}(d) \leq 0$. Usando a continuidade da função de recessão, concluímos que existe $\delta > 0$ tal que

$$\forall d \in \mathcal{K}, \|d\| = 1, \quad \max_{i=1, \dots, m} (f_i)'_{\infty}(d) \geq \delta.$$

Assim temos para $\beta \geq e$, $\mu \in (0, \beta)$ e $\forall d \in \mathcal{K}, \|d\| = 1$,

$$\sum_{i=1}^m \beta_i \theta'_{\infty}((f_i)'_{\infty}(d)) \geq \beta_j \delta,$$

para algum $j \in \{1, \dots, m\}$. Portanto, $\forall d \in \mathcal{K}, \|d\| = 1$,

$$(L)'_{\infty}(d) \geq (f_0)'_{\infty}(d) + \beta_j \delta.$$

Tomando para todo $j = 1, \dots, m$

$$\bar{\beta}_j \geq \frac{-\inf\{(f_0)'_{\infty}(d) \mid d \in \mathcal{K}, \|d\| = 1\}}{\delta},$$

obtemos para todo $\beta > \bar{\beta}$, $\forall \mu \in (0, \beta)$, $\forall d \in \mathcal{K}, \|d\| = 1$,

$$(L_{\beta,r})'_{\infty}(d) > 0.$$

■

Comentário:

O teorema anterior garante que as minimizações consideradas no algoritmo 5.1 estão bem definidas desde que se inicialize o algoritmo com $\beta^0 \geq \bar{\beta}$.

5.3 Dualidade

O problema dual para (P) foi definido como

$$(D) \quad \begin{array}{ll} \text{maximizar} & d(\mu) \\ \text{sujeito a} & \mu \geq 0 \end{array}$$

onde

$$\begin{aligned} x \in R^n, \mu \in R^m &\mapsto L(x, \mu) = f_0(x) + \sum_{i=1}^m \mu_i f_i(x) \quad \text{e} \\ \mu \in R^m &\mapsto d(\mu) = \inf \{L(x, \mu) : x \in R^n\}. \end{aligned}$$

Escrevemos o lagrangeano aumentado correspondente à família de penalidades $AL1$ como

$$x \in R^n, \mu \in (0, \beta) \mapsto L_{\beta,r}(x, \mu) = f_0(x) + P_{\beta,r}(f(x), \mu)$$

onde P é não-decrescente e definida em (5.7) como

$$\begin{aligned} x \in R^n, \mu \in (0, \beta) \mapsto P_{\beta,r}(f(x), \mu) &= \sum_{i=1}^m p_{\beta,r}(f(x), \mu) \\ &= \sum_{i=1}^m \beta_i r \left[\theta \left(\frac{f_i(x)}{r} + \tilde{y}_i(\mu_i) \right) - \theta(\tilde{y}_i(\mu_i)) \right] \end{aligned}$$

onde para $i = 1, \dots, m$, $\tilde{y}_i(\mu_i) = \frac{\mu_i}{\beta_i} < 1$ como definido em (5.3). Para esta família de penalidades mostramos a forma da sua conjugada. Pode-se imediatamente reconhecer no lema a seguir o formato de uma quase-distância de Bregman.

Proposição 5.6 *Seja θ uma função de penalidade do tipo I que satisfaz as propriedades i)-vi). Dados $\beta \in R^m$, $r \in (0, 1]$ e $\mu \in (0, \beta)$, considere a função*

$$y \in R^m, \mu \in (0, \beta) \mapsto P_1(y) = \sum_{i=1}^m p_1(y_i) = \sum_{i=1}^m \beta_i r \left[\theta \left(\frac{y_i}{r} + \tilde{y}_i \right) - \theta(\tilde{y}_i) \right]$$

onde para $i = 1, \dots, m$, \tilde{y}_i satisfaz $\theta'(\tilde{y}_i) = \mu_i/\beta_i < 1$. Cumpre-se

$$P_1^*(s) = r \sum_{i=1}^m \left[\beta_i \theta^* \left(\frac{s_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i}{\beta_i} \right) (s_i - \mu_i) \right]. \quad (5.15)$$

Prova:

Observe que para cada $i = 1, \dots, m$, usando propriedades de funções conjugadas (ver proposição 4.3),

se $f(y_i) \equiv \theta(y_i + \tilde{y}_i)$ então $f^*(s_i) = \theta^*(s_i) - \tilde{y}_i s_i$, e

se $h(y_i) \equiv \beta_i r f\left(\frac{y_i}{r}\right)$ então

$$\begin{aligned} h^*(s_i) &= \beta_i r f^*\left(\frac{s_i}{\beta_i}\right) \\ &= \beta_i r \left(\theta^*\left(\frac{s_i}{\beta_i}\right) - \tilde{y}_i \frac{s_i}{\beta_i} \right) \\ &= \beta_i r \theta^*\left(\frac{s_i}{\beta_i}\right) - r \tilde{y}_i s_i. \end{aligned}$$

Assim, como $p(y_i) = h(y_i) - C_i$ onde $C_i = \beta_i r \theta(\tilde{y}_i)$ temos para $i = 1, \dots, m$

$$\begin{aligned} p_1^*(s) &= h^*(s_i) + C_i \\ &= r \left[\beta_i \theta^*\left(\frac{s_i}{\beta_i}\right) - \tilde{y}_i s_i + \beta_i \theta(\tilde{y}_i) \right]. \end{aligned}$$

Mas, usando a proposição(4.2) com $x = \tilde{y}_i$ e $s_i = \frac{\mu_i}{\beta_i}$ obtemos

$$\theta(\tilde{y}_i) = \frac{\mu_i}{\beta_i} \tilde{y}_i - \theta^*\left(\frac{\mu_i}{\beta_i}\right)$$

e assim para $i = 1, \dots, m$

$$\begin{aligned} p_1^*(s) &= r \left[\beta_i \theta^*\left(\frac{s_i}{\beta_i}\right) - \tilde{y}_i s_i + (\mu_i \tilde{y}_i - \beta_i \theta^*\left(\frac{\mu_i}{\beta_i}\right)) \right] \\ &= r \left[\beta_i \theta^*\left(\frac{s_i}{\beta_i}\right) - \beta_i \theta^*\left(\frac{\mu_i}{\beta_i}\right) - \tilde{y}_i (s_i - \mu_i) \right] \\ &= r \left[\beta_i \theta^*\left(\frac{s_i}{\beta_i}\right) - \beta_i \theta^*\left(\frac{\mu_i}{\beta_i}\right) - (\theta^*)'\left(\frac{\mu_i}{\beta_i}\right) (s_i - \mu_i) \right]. \end{aligned}$$

Finalmente obtemos

$$P_1^*(s) = r \sum_{i=1}^m \left[\beta_i \theta^*\left(\frac{s_i}{\beta_i}\right) - \beta_i \theta^*\left(\frac{\mu_i}{\beta_i}\right) - (\theta^*)'\left(\frac{\mu_i}{\beta_i}\right) (s_i - \mu_i) \right].$$

■

Em forma similar enunciamos a seguinte proposição para as penalidades AL2 cuja prova é análoga à anterior, basta omitir β ou fazer $\beta_i = 1$ para todo $i = 1, \dots, m$.

Proposição 5.7 *Seja θ uma função de penalidade do tipo II que satisfaz as propriedades i)-v), dados $r \in (0, 1]$ e $\mu \in R_{++}^m$, considere a função*

$$y \in R^m \mapsto P_2(y) = \sum_{i=1}^m p_2(y_i) = \sum_{i=1}^m r \left[\theta\left(\frac{y_i}{r} + \tilde{y}_i\right) - \theta(\tilde{y}_i) \right].$$

onde para $i = 1, \dots, m$ $\tilde{y}_i(\mu_i) = \tilde{y}_i$ satisfaz $\theta'(\tilde{y}_i) = \mu_i$. Cumpre-se

$$P_2^*(s) = r \sum_{i=1}^m \left[\theta^*(s_i) - \theta^*(\mu_i) - (\theta^*)'(\mu_i) (s_i - \mu_i) \right]. \quad (5.16)$$

Prova: Ver proposição anterior ■

As seguintes definições e teoremas serão de utilidade.

Definição 5.8 *Uma função $f : R^n \rightarrow R \cup \{+\infty\}$ é dita suave se e somente se f é finita e diferenciável em todo R^n .*

Definição 5.9 *Uma função $f : R^n \rightarrow R \cup \{+\infty\}$ própria convexa é dita essencialmente suave se satisfaz as três seguintes condições para $C = \text{int dom } f$:*

a) $C \neq \emptyset$,

b) f é diferenciável em todo C ,

c) $\lim_{k \rightarrow \infty} |\nabla f(x_k)| = +\infty$ sempre que x_1, x_2, \dots , é uma seqüência em C convergindo a um ponto x na fronteira de C .

Lema 5.10 *(ver lema 26.2 em [54]) Asumindo a) b) na definição anterior, a condição c) é equivalente a c') $f'(x + \lambda(a - x); a - x) \downarrow -\infty$ quando $\lambda \downarrow 0$ para todo $a \in C$ e todo ponto x na fronteira de C .*

Definição 5.11 *Uma função $f : R^n \rightarrow R \cup \{+\infty\}$ convexa própria é dita essencialmente estritamente convexa se f é estritamente convexa em cada subconjunto convexo de $\{x / \partial f(x) \neq \emptyset\} = \text{dom } f$.*

Teorema 5.12 *(Teorema 26.3 em [54]) Uma função $f : R^n \rightarrow R \cup \{+\infty\}$ própria convexa fechada é essencialmente estritamente convexa se e somente se sua conjugada é essencialmente suave.*

5.3.1 Quase-distâncias geradas por $P_{\beta, r, \mu}^*$ e $P_{r, \mu}^*$

Dado $\beta \geq e \in R^m$ e o intervalo $S_1 = (0, \beta)$, considere a função

$$s \in [0, \beta] \mapsto h_1(s) = \sum_{i=1}^m \beta_i \theta^* \left(\frac{s_i}{\beta_i} \right), \quad (5.17)$$

onde θ^* é a função conjugada de θ , penalidade do tipo I que define a família de penalidades AL1. Observe que v) e vi) garantem que $\text{dom } \theta^* = [0, 1]$ (ver teorema 27.1.a em [54]), segue-se que h_1 está bem definida em \bar{S}_1 .

Logo podemos escrever (5.15) como

$$\begin{aligned} P_1^*(s) &= r \sum_{i=1}^m \left[\beta_i \theta^* \left(\frac{s_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i}{\beta_i} \right) (s_i - \mu_i) \right] \\ &= r [h_1(s) - h_1(\mu) - \nabla h_1(\mu)^t (s - \mu)]. \end{aligned}$$

Definimos

$$s \in \bar{S}_1, \mu \in S_1 \mapsto D_{h_1}(s, \mu) = r [h_1(s) - h_1(\mu) - \nabla h_1(\mu)^t (s - \mu)]. \quad (5.18)$$

Da mesma forma, considerando o intervalo $S_2 = (0, +\infty)$ definimos

$$s \in \bar{S}_2 \mapsto h_2(s) = \sum_{i=1}^m \theta^*(s_i), \quad (5.19)$$

onde θ^* é a função conjugada de θ , penalidade do tipo II que define a família de penalidades *AL2* com $\text{dom } \theta^* = [0, +\infty)$. Obtemos de (5.16)

$$\begin{aligned} P_2^*(s) &= r \sum_{i=1}^m [\theta^*(s_i) - \theta^*(\mu_i) - (\theta^*)'(\mu_i)(s_i - \mu_i)] \\ &= r [h_2(s) - h_2(\mu) - \nabla h_2(\mu)^t (s - \mu)], \end{aligned}$$

assim definimos

$$s \in \bar{S}_2, \mu \in S \mapsto D_{h_2}(s, \mu) = r [h_2(s) - h_2(\mu) - \nabla h_2(\mu)^t (s - \mu)]. \quad (5.20)$$

Observe que em ambos casos, por ser θ estritamente crescente, θ^* é estritamente convexa (ver teorema 4.1.3 em [29]), assim para $h = h_1$ ou $h = h_2$ temos:

- i) $D_h(s, \mu) \geq 0$ e
- ii) $D_h(s, \mu) = 0$ se e somente se $s = \mu$.

e assim $D_h(s, \mu)$ define uma quase-distância.

Coercividade na Fronteira.

A penalidade θ do tipo I é essencialmente estritamente convexa em R por ser estritamente convexa em R (ver definições (5.9) e (5.11)), assim pelo teorema (5.12), θ^* é essencialmente suave em $S_1 = (0, 1)$. Além disso, θ^* é definida nos extremos 0 e 1, portanto coerciva na fronteira, daqui concluímos que h_1 é coerciva na fronteira no intervalo $\bar{S}_1 = [0, \beta]$.

De forma análoga, a penalidade θ do tipo II por ser essencialmente estritamente convexa em R , garante que θ^* seja essencialmente suave em $S = (0, +\infty)$. Além disso, θ^* é definida em 0. Logo, h_2 é coerciva em $\bar{S}_2 = [0, +\infty)$.

O seguintes teoremas (5.13) e (5.14) podem ser considerados como os principais resultados da presente tese. Garantem que as quase-distâncias D_{h_1} e D_{h_2} definidas em (5.18) e (5.20) são distâncias de Bregman.

Teorema 5.13 *Seja $\beta \geq e \in R^m$, $S_1 = (0, \beta)$ e considere a função*

$$s \in \bar{S}_1 \mapsto h_1(s) = \sum_{i=1}^m \beta_i \theta^* \left(\frac{s_i}{\beta_i} \right),$$

onde θ^* é a função conjugada de θ , penalidade do tipo I que define a família de penalidades AL1. Então D_{h_1} definida por

$$s \in \bar{S}_1, \mu \in S_1 \mapsto D_{h_1}(s, \mu) = r[h_1(s) - h_1(\mu) - \nabla h_1(\mu)^t(s - \mu)]$$

é uma distância de Bregman.

Prova:

Enunciamos e mostramos a validade de cada uma das condições, apontadas na seção (3.2.1), que são requeridas para que uma quase-distância possa ser considerada como uma distância de Bregman. Observe que h_1 é separável e assim a análise da prova baseia-se praticamente na função θ^* , dito de outra forma, podemos nos limitar ao caso unidimensional.

B1) h_1 é continuamente diferenciável em S_1 .

De fato, θ é estritamente convexa e fechada, pelo teorema 4.1.1 em [29] $\text{int dom } \theta^* \neq \emptyset$ e θ^* é continuamente diferenciável sobre $\text{int dom } \theta^* = (0, 1)$, segue-se que a condição é satisfeita para h_1 em $S_1 = (0, \beta)$.

B2) A função h_1 é estritamente convexa e contínua em $\bar{S}_1 = [0, \beta]$.

Devido a que θ é convexa, fechada e diferenciável sobre $\Omega = \text{int dom } \theta$ então, usando teorema 4.1.3 em [29] temos que θ^* é estritamente convexa em cada subconjunto $C \subset \nabla \theta(\Omega)$, e portanto contínua em $(0, 1)$, segue-se que h_1 é contínua em S_1 . Por outra parte, por ser θ^* convexa, própria e fechada, corolário 7.5.1 em [54] temos

$$\forall x, y \in \text{dom } \theta, \text{ e } 0 \leq \lambda < 1, \quad \theta^*(y) = \lim_{\lambda \uparrow 1} \theta^*((1 - \lambda)x + \lambda y)$$

assim, para $y = 0$

$$\theta^*(0) = \lim_{\lambda \uparrow 1} \theta^*((1 - \lambda)x)$$

fazendo $w = 1 - \lambda$ temos

$$\theta^*(0) = \lim_{w \downarrow 0} \theta^*(w)$$

e assim θ^* é contínua em 0. Em forma semelhante, para $y = 1$ e

fazendo $z = (1 - \lambda)x + \lambda$ temos

$$\begin{aligned} \theta^*(1) &= \lim_{\lambda \uparrow 1} \theta^*((1 - \lambda)x + \lambda) \\ &= \lim_{z \uparrow 1} \theta^*(z) \end{aligned}$$

Segue-se que h_1 é contínua em \bar{S} . Observe além disso que por ser θ^* uma função estritamente convexa e fechada nos extremos de S , é estritamente convexa em $[0, 1]$. Daí que h_1 é estritamente convexa em \bar{S}_1 .

B3) Para todo $\delta \in R$, os conjuntos de nível

$$\Gamma_1(\mu, \delta) = \{s \in \bar{S} = [0, \beta] \mid D_{h_1}(s, \mu) \leq \delta\} \text{ e}$$

$\Gamma_2(s, \delta) = \{\mu \in S = (0, \beta) \mid D_{h_1}(s, \mu) \leq \delta\}$ são limitados para todo $\mu \in S$ e para todo $s \in \bar{S}$ respectivamente.

Isto é imediato pois $\text{dom } \theta^*$ é um conjunto compacto.

B4) Seja $\{\mu^k\} \subset S$ uma seqüência tal que $\mu^k \rightarrow \hat{\mu}$ então $D_{h_1}(\hat{\mu}, \mu^k) \rightarrow 0$.

Sabemos que para todo k , se satisfaz

$$0 \leq D_{h_1}(\hat{\mu}, \mu^k) = r \sum_{i=1}^m \left[\beta_i \theta^* \left(\frac{\hat{\mu}_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (\hat{\mu}_i - \mu_i^k) \right]$$

Basta provar a propriedade para cada termo da somatória o que é equivalente a analisar o caso unidimensional, assim, para todo $i = 1, \dots, m$, suponhamos que $\mu_i^k \rightarrow \hat{\mu}_i$. Consideremos três possibilidades:

a) $0 < \hat{\mu}_i < \beta$.

Sabemos por (B4) que

$$\beta_i \theta^* \left(\frac{\hat{\mu}_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (\hat{\mu}_i - \mu_i^k) \text{ é limitado para todo } \hat{\mu}_i, \mu_i^k \in S$$

assim, pela continuidade de θ^* temos

$$\lim_{\mu^k \rightarrow \hat{\mu}} \left[\beta_i \theta^* \left(\frac{\hat{\mu}_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (\hat{\mu}_i - \mu_i^k) \right] = 0.$$

b) $\hat{\mu}_i = 0$.

Neste caso, observando que

$$\lim_{\mu^k \rightarrow 0} (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (\mu_i^k) \leq 0$$

e $\theta^*(0)$ é finito temos para $i = 1, \dots, m$

$$\begin{aligned} 0 &\leq \lim_{\mu^k \rightarrow 0} \left[\beta_i \theta^* \left(\frac{\hat{\mu}_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (\hat{\mu}_i - \mu_i^k) \right] \\ &= \lim_{\mu^k \rightarrow 0} \left[\beta_i \theta^*(0) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) + (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (\mu_i^k) \right] \\ &\leq \lim_{\mu_i^k \rightarrow 0} \left[\beta_i \theta^*(0) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) \right] \\ &= 0. \end{aligned}$$

c) $\hat{\mu}_i = \beta$

Similar ao caso anterior.

Logo concluímos que se $\mu^k \rightarrow \hat{\mu}$ então $D_{h_1}(\hat{\mu}, \mu^k) \rightarrow 0$ para toda seqüência $\{x^k\}$ em S , $\forall \hat{\mu} \in \bar{S}$.

B5) Se $\{s^k\} \subset \bar{S}$ e $\{\mu^k\} \subset S$ são seqüências tais que $\{s^k\}$ é limitada, $\lim_{k \rightarrow \infty} \mu^k = \hat{\mu}$ e $\lim_{k \rightarrow \infty} D_{h_1}(s^k, \mu^k) = 0$ então $\lim_{k \rightarrow \infty} s^k = \hat{\mu}$.

Prova: Observe que como a distância definida é soma de termos não negativos, converge a 0 se e somente se cada termo da soma converge a zero. Portanto podemos tratar a convergência em forma separada o que equivale a tratar o caso unidimensional.

Seja $\{s^k\}_{k \in K}$ uma subsequência de $\{s^k\}$ convergente a um ponto \hat{s} , devemos provar que $\hat{s} = \hat{\mu}$.

Consideremos três casos.

a) $0 < \hat{\mu} < \beta$.

Como por hipótese $\lim_{k \rightarrow \infty} D_{h_1}(s^k, \mu^k) = 0$ temos tomando limites para $i = 1, \dots, m$

$$\beta_i \theta^* \left(\frac{\hat{s}_i}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\hat{\mu}_i}{\beta_i} \right) - (\theta^*)' \left(\frac{\hat{\mu}_i}{\beta_i} \right) (\hat{s}_i - \hat{\mu}_i) = 0$$

e como θ^* é estritamente convexa em $\text{int } S$, verifica-se $\hat{s} = \hat{\mu}$.

b) $\hat{\mu} = 0$.

Suponha por contradição que $\hat{s} \neq 0$, então

$$\lim_{\mu^k \rightarrow 0} (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (s_i^k - \mu_i^k) = -\infty$$

como os dois primeiros termos são limitados e θ^* é coerciva na fronteira, temos

$$\beta_i \theta^* \left(\frac{s_i^k}{\beta_i} \right) - \beta_i \theta^* \left(\frac{\mu_i^k}{\beta_i} \right) - (\theta^*)' \left(\frac{\mu_i^k}{\beta_i} \right) (s_i - \mu_i^k) \not\rightarrow 0$$

Logo, $D_{h_1}(s^k, \mu^k) \not\rightarrow 0$ contradizendo a hipótese.

c) $\hat{\mu} = \beta$

Este caso é similar ao anterior. ■

Exemplo:

Dada $\theta(y) = \frac{1}{2}(y + \sqrt{y^2 + 4}) - 1$, $\beta = r = 1$ temos

$\theta^*(s) = -2\sqrt{s(1-s)} + 1$, se definimos

$$s \in [0, 1] \mapsto h_1(s) = \sum_{i=1}^m \theta^*(s_i) \quad \text{temos}$$

$$D_{h_1}(s, \mu) = 2\sqrt{\mu(1-\mu)} - 2\sqrt{s(1-s)} - \frac{2s-1}{\sqrt{\mu(1-\mu)}}(s-\mu).$$

Teorema 5.14 *Seja $S_2 = \mathbb{R}_{++}^n$ e considere a função*

$$s \in \bar{S}_2 \mapsto h_2(s) = \sum_{i=1}^m \theta^*(s_i),$$

onde θ^* é a função conjugada de θ , penalidade do tipo II que define a família de penalidades AL2. então D_{h_2} definida em (5.20) por

$$s \in \bar{S}_2, \mu \in S_2 \mapsto D_{h_2}(s, \mu) = r[h_2(s) - h_2(\mu) - \nabla h_2(\mu)^t(s - \mu)].$$

é uma distância de Bregman.

A prova é semelhante ao teorema anterior, mostraremos somente B3 pois B1, B2, B4 e B5 se satisfazem pela mesma justificativa dada no caso anterior.

Provemos que: a) $\Gamma_1(\mu, \delta) = \{s \in \bar{S}_2 = \mathbb{R}_+^n \mid D_{h_2}(s, \mu) \leq \delta\}$ é limitado para todo $\mu \in S_2$.

Sabemos que $D_{h_2}(\cdot, \mu)$ é estritamente convexa e não negativa para $\mu \in S_2$. Para $\delta = 0$ e $\mu \in S_2$ o conjunto $\Gamma_1(\mu, 0) = \{s \in \bar{S}_2 \mid D_{h_2}(s, \mu) \leq 0\} = \{0\}$ é limitado, segue-se que Γ_1 é limitado para todo $\delta \in \mathbb{R}$ e $\mu \in S_2$.

b) $\Gamma_2(s, \delta) = \{\mu \in S_2 = (0, +\infty) \mid D_{h_2}(s, \mu) < \delta\}$ é limitado para todo $\delta \in R$ e para todo $s \in \bar{S}_2$.

Seja $s \in \bar{S}_2$ fixo e arbitrário. Suponha que Γ_2 é ilimitado para algum $\delta \in R$. Então existe uma seqüência $\{\mu^k\}$ em $\Gamma_2(s, \delta)$ tal que $\mu^k \rightarrow \infty$.

Como D_{h_2} é soma de termos não negativos, podemos considerar somente um destes e restringirmos ao caso unidimensional.

Observe que, por ser θ^* convexa, temos

- 1) $\theta^*(s) \geq \theta^*(\mu^k) + (\theta^*)'(\mu^k)(s - \mu^k)$,
- 2) $\theta^*(s + 1) \geq \theta^*(\mu^k) + (\theta^*)'(\mu^k)(s - \mu^k) + (\theta^*)'(\mu^k)$,
- 3) $\theta^*(s) \geq \theta^*(s + 1) - (\theta^*)'(s + 1)$

somando 1), 2) e 3) obtemos

$$2\theta^*(s) \geq 2 \left[\theta^*(\mu^k) + (\theta^*)'(\mu^k)(s - \mu^k) \right] + (\theta^*)'(\mu^k) - (\theta^*)'(s + 1)$$

dividindo por 2 temos

$$\theta^*(s) - \theta^*(\mu^k) - (\theta^*)'(\mu^k)(s - \mu^k) \geq (\theta^*)'(\mu^k)/2 - (\theta^*)'(s + 1)/2$$

Como $s + 1$ é fixo em S_2 temos que $\theta^*(s + 1)$ é finito, mas como $(\theta^*)'(\cdot)$ é coerciva, $(\theta^*)'(\mu^k) \rightarrow +\infty$ e assim D_{h_2} não pode ser majorada por δ .

■

5.3.2 Métodos de ponto proximal usando distâncias de Bregman

Introduzimos dois métodos de ponto proximal baseados nas distâncias de Bregman definidas anteriormente e aplicadas diretamente para maximizar a função dual d num conjunto convexo e fechado \bar{S} . A idéia consiste em mostrar que as seqüências geradas por cada um destes métodos, coincidem respectivamente com as seqüências gerados pelos algoritmos (5.1) e (5.2) correspondentes aos métodos de multiplicadores baseados em *shifts* definidos na seção (5.2.1). Posteriormente estabelecemos sua convergência a uma solução ótima do problema dual considerado.

Consideremos o problema

$$\begin{aligned} & \text{maximizar } d(\mu) \\ (\hat{D}) \quad & \text{sujeito a } \mu \in [0, \beta] \end{aligned}$$

com $\beta \geq e$ fixo.

O método de ponto proximal para resolver o problema (\hat{D}) , usando a distância de Bregman definida em (5.18), gera uma seqüência tal que:

$$\begin{aligned} \mu^0 & \in S_1 = (0, \beta) \\ \mu^{k+1} & \in \operatorname{argmin}_{\mu \in \bar{S}_1} \{-d(\mu) + r^k D_{h_1}(\mu, \mu^k)\} \end{aligned} \quad (5.21)$$

ou equivalentemente

$$\mu^{k+1} \in \operatorname{argmax}_{\mu \in \bar{S}_1} \{d(\mu) - r^k D_{h_1}(\mu, \mu^k)\} \quad (5.22)$$

onde $r^k \in (0, 1]$ para todo k . Observe que estas expressões estão bem definidas pelo fato de ser $D_{h_1}(\cdot, \mu)$ estritamente convexa e coerciva.

Observe que

$$\nabla_s D_{h_1}(s, \mu) = \sum_{i=1}^m [(\theta^*)'(\frac{s_i}{\beta_i}) - (\theta^*)'(\frac{\mu_i}{\beta_i})]. \quad (5.23)$$

O teorema a seguir relaciona as seqüências geradas pelo método de multiplicadores com shifts usando a família de penalidades AL1 e o método de ponto proximal (5.21) para resolver o problema dual (\hat{D}) . Seguimos o esquema de prova mostrado em Iusem ([36]). Podemos considerar o vetor β fixo pois será mostrado na proposição 5.17 a partir de um certo valor de k .

Teorema 5.15 *Considere uma seqüência $\{r^k\}$ em $(0, 1]$. Seja $\{\mu^k\}$ a seqüência gerada por (5.21) aplicado ao problema (\hat{D}) e $\{\hat{x}^k\}$, $\{\hat{\mu}^k\}$ as seqüências (5.9), (5.10) geradas pelo algoritmo (5.1) correspondentes ao método de multiplicadores com shifts que usa a família de penalidades AL1. Se $\hat{\mu}^0 = \mu^0$ e $\beta^k = \beta$ para todo $k \in N$ onde $\beta \in R_{++}^m$, então $\forall k \in N$, $\hat{\mu}^k = \mu^k$.*

Prova:

Consideremos $\beta^k = \beta$ para todo $k \in N$ onde $\beta \in R_{++}^m$. Usemos indução. Suponhamos que para um dado $k \in N$, $\hat{\mu}^k = \mu^k$. Devemos provar que $\hat{\mu}^{k+1} = \mu^{k+1}$. Seja \tilde{y} tal que $\theta'(\tilde{y}_i) = \frac{\hat{\mu}_i^k}{\beta}$ e suponhamos que $\hat{\mu}^k = \mu^k$ com β, r^k fixos para todo k . Sabemos que se

$$\hat{x}^{k+1} \in \operatorname{argmin}_x L_{\beta, r^k}(\cdot, \hat{\mu}^k)$$

onde

$$L_{\beta, r^k}(\cdot, \hat{\mu}^k) = f_0(\cdot) + \sum_{i=1}^m \beta_i \left[r^k \theta \left(\frac{f_i(\cdot)}{r^k} + \tilde{y}_i \right) - \theta(\tilde{y}_i) \right]$$

então

$$\begin{aligned} 0 \in \partial_x L_{\beta, r^k}(\hat{x}^{k+1}, \hat{\mu}^k) &= \partial f_0(\hat{x}^{k+1}) + \sum_{i=1}^m \beta_i \theta' \left(\frac{f_i(\hat{x}^{k+1})}{r^k} + \tilde{y}_i \right) \partial f_i(\hat{x}^{k+1}) \\ &= \partial f_0(\hat{x}^{k+1}) + \sum_{i=1}^m \hat{\mu}_i^{k+1} \partial f_i(\hat{x}^{k+1}) \quad \text{por (5.10)} \\ &= \partial_x l(\hat{x}^{k+1}, \hat{\mu}^{k+1}), \end{aligned}$$

como $\hat{\mu}_i^{k+1} > 0$, e $l(\cdot, \hat{\mu}_i^{k+1})$ é convexo temos que (ver 5.11)

$$\hat{\mu}_i^{k+1} \text{ minimiza o lagrangeano } l(\cdot, \hat{\mu}_i^{k+1}) \quad (5.24)$$

Por outra parte, para todo $\mu \in R^m$

$$\begin{aligned} d(\mu) &= \inf \{ f_0(x) + \sum_{i=1}^m \mu_i f_i(x), x \in R^n \} \\ &\leq f_0(\hat{x}^{k+1}) + \sum_{i=1}^m \mu_i f_i(\hat{x}^{k+1}) \\ &= f_0(\hat{x}^{k+1}) + \sum_{i=1}^m \hat{\mu}_i^{k+1} f_i(\hat{x}^{k+1}) + \sum_{i=1}^m (\mu_i - \hat{\mu}_i^{k+1}) f_i(\hat{x}^{k+1}) \\ &= d(\hat{\mu}_i^{k+1}) + \sum_{i=1}^m (\mu_i - \hat{\mu}_i^{k+1}) f_i(\hat{x}^{k+1}) \end{aligned}$$

logo

$$f(\hat{x}^{k+1}) \in \partial d(\hat{\mu}^{k+1}) \quad (5.25)$$

onde $f(\cdot) = (f_1(\cdot), \dots, f_m(\cdot))$ e $\partial d(\mu)$ é o conjunto de subgradientes em μ da função côncava d .

Note que, de (5.10) temos $\forall i = 1, \dots, m$

$$\begin{aligned} \hat{\mu}_i^{k+1} &= \beta_i \theta' \left(\frac{f_i(\hat{x}^{k+1})}{r^k} + \tilde{y}_i \right) \\ \Leftrightarrow \frac{\hat{\mu}_i^{k+1}}{\beta_i} &= \theta' \left(\frac{f_i(\hat{x}^{k+1})}{r^k} + \tilde{y}_i \right) \\ \Leftrightarrow (\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - \tilde{y}_i &= \frac{f_i(\hat{x}^{k+1})}{r^k} \\ \Leftrightarrow f_i(\hat{x}^{k+1}) &= r^k \left[(\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - \tilde{y}_i \right] \\ \Leftrightarrow f_i(\hat{x}^{k+1}) &= r^k \left[(\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i} \right) - (\theta')^{-1} \left(\frac{\hat{\mu}_i^k}{\beta_i} \right) \right] \end{aligned}$$

De (5.25) e por hipótese indutiva, temos $\forall i = 1, \dots, m$

$$f_i(\hat{x}^{k+1}) = r^k \left[(\theta')^{-1} \left(\frac{\hat{\mu}_i^{k+1}}{\beta_i^k} \right) - (\theta')^{-1} \left(\frac{\mu_i^k}{\beta_i^k} \right) \right] \in \partial d(\hat{\mu}^{k+1}) \quad (5.26)$$

assim, observando (5.23) e o fato que $(\theta')^{-1} = (\theta^*)'$, afirmamos que $\hat{\mu}^{k+1}$ é solução de (5.21). Mas μ^{k+1} é solução única de (5.21) e obtido a partir do iterado anterior μ^k . Isto significa que $\hat{\mu}^{k+1} = \mu^{k+1}$. ■

De forma semelhante dada θ , função de penalidade do tipo II e o intervalo $S_2 = (0, +\infty)$, consideremos a sua função conjugada $\theta^* : \bar{S}_2 \rightarrow R$ e sua correspondente distância de Bregman $D_{h_2} : \bar{S}_2 \times S_2 \rightarrow R$ definida em (5.20) como

$$D_{h_2}(s, \mu) = h_2(s_i) - h_2(\mu_i) - \nabla h(\mu_i)^t (s_i - \mu_i),$$

com $h_2(s) = \sum_{i=1}^m \theta^*(s_i)$.

Consideremos o problema dual de (P) definido como

$$(D) \quad \begin{array}{l} \text{maximizar } d(\mu) \\ \text{sujeito a } \mu \geq 0. \end{array}$$

O método de ponto proximal para resolver o problema (D) usando a distância de Bregman definida em (5.20) gera uma seqüência tal que:

$$\mu^0 \in S_2$$

$$\mu^{k+1} \in \operatorname{argmin}_{\mu \in \bar{S}_2} \{-d(\mu) + r^k D_{h_2}(\mu, \mu^k)\} \quad (5.27)$$

ou equivalentemente

$$\mu^{k+1} \in \operatorname{argmax}_{\mu \in \bar{S}_2} \{d(\mu) - r^k D_{h_2}(\mu, \mu^k)\}, \quad (5.28)$$

onde $r^k \in (0, 1]$ para todo k .

Observe que

$$\nabla_s D_{h_2}(s, \mu) = \sum_{i=1}^m [(\theta^*)'(s_i) - (\theta^*)'(\mu_i)]. \quad (5.29)$$

Enunciamos sem prova o seguinte teorema por ser semelhante ao teorema (5.15) anterior. Relaciona as seqüências geradas pelo método de multiplicadores com shifts usando a família de penalidades AL2 e o método de ponto proximal definido em (5.27).

Teorema 5.16 *Considere uma seqüência $\{r^k\}$ com r^k em $(0, 1]$. Seja $\{\mu^k\}$ a seqüência gerada por (5.27) aplicado ao problema (D) e $\{\hat{x}^k\}$, $\{\hat{\mu}^k\}$ as seqüências (5.12), (5.14) geradas pelo algoritmo (5.2) correspondente ao método de multiplicadores com shifts que usa a família de penalidades AL2. Se $\hat{\mu}^0 = \mu^0$ então $\forall k \in N$, $\hat{\mu}^k = \mu^k$.*

Prova: Similar ao teorema anterior. ■

A seguinte proposição é importante pois mostra que a seqüência $\{\beta^k\}$ gerada pelo algoritmo (5.1) correspondente ao método de multiplicadores com shifts que usa a família de penalidades AL1 é limitada. Este fato garante o uso do método de ponto proximal com distância de Bregman D_{h_1} para um valor fixo de β .

Proposição 5.17 *Seja $\{\mu^k\}$ a seqüência (5.22) gerada pelo método de ponto proximal, então*

i) A seqüência $\{\mu^k\}$ é limitada.

ii) A seqüência $\{\beta^k\}$ gerada pelo algoritmo (5.1) correspondente ao método de multiplicadores com shifts que usa a família de penalidades AL1 é limitada, isto é, existe $k_1 \in N$ tal que $\forall k \geq k_1$, $\beta^k = \beta^{k_1}$.

Prova:

i) Inferimos de (A2) (condição de qualificação de Slater) que a função dual $d(\cdot)$ é limitada por cima pelo valor ótimo \hat{f} e, pela propriedade B3) o conjunto $D(\cdot, \mu^k)$ tem conjuntos de nível limitados.

Definimos os conjuntos

$L = \{\mu \in S \mid d(\mu) \geq d(\mu^0)\}$ e $Q = \max\{\|\mu\|_\infty \mid \mu \in L\}$. Assim, $\mu^k \in L$ para todo $k \in N$. Portanto $\|\mu^k\|_\infty \leq Q$, $\forall k \in N$.

ii) A atualização de β^k no algoritmo (5.1) estabelece que se para $i = 1, \dots, m$, $\mu_i^k > \frac{\beta_i^k}{2}$ então $\beta_i^{k+1} = 2\beta_i^k$, caso contrário $\beta_i^{k+1} = \beta_i^k$. Como $\forall k \in N$, $\|\mu\|_\infty \leq Q$, temos que se para $k_1 \in N$, $\beta^{k_1} \geq 2Q$ então, por construção $\beta^k = \beta^{k_1}$ para todo $k \geq k_1$, $k \in N$. ■

Comentamos que a existência do vetor β^0 que possibilita as minimizações correspondentes à seqüência $\{x^k\}$ no método de multiplicadores definido em (5.1)

está garantida pela proposição (5.5). Por outra parte, o método de ponto proximal definido em (5.22) precisa de um valor fixo de β : a existência deste valor está garantida pela proposição anterior após um número finito de iterações.

5.3.3 Convergência

Convergência das seqüência μ^k .

Teorema 5.18 *As seqüências $\{\mu^k\}$ geradas pelos algoritmos de ponto proximal com distâncias de Bregman (5.21) e (5.27) convergem respectivamente a soluções ótimas dos problemas (\hat{D}) e (D) .*

Prova: A convergência das seqüências $\{\mu^k\}$ geradas pelos algoritmos métodos de ponto proximal (5.21) e (5.27) a uma solução ótima dos problemas (\hat{D}) e (D) , está garantida pelo teoremas 4.1 que se mostra em Iusem [34]) (ver teorema 3.13 enunciado na seção (3.2.1) uma vez que suas hipóteses são satisfeitas. Neste caso, usamos distâncias de Bregman com funções de Bregman coercivas na fronteira, a função dual $-d(\cdot)$ é convexa com $\text{dom}(-d) \cap S \neq \emptyset$ (com S igual a S_1 ou S_2) e a seqüência $\{r^k\}$ é limitada superiormente. ■

Convergência das seqüência $\{x^k\}$

Uma dificuldade teórica presente nos métodos de lagrangeano aumentado está em estabelecer a convergência da seqüência de minimizadores $\{x^k\}$ a uma solução ótima do problema (P) , ou bem, na existência de pontos limites de $\{x^k\}$ soluções ótimas de (P) . O estudo da convergência tem sido associado juntamente com as φ -divergências e com as distâncias de Bregman sob diferentes hipóteses por diferentes autores, ver [21], [33], [41], [50], [11], [31], estabelecendo a convergência da seqüência $\{x^k\}$ num sentido ergódico, isto é, a convergência de médias de $\{x^k\}$ a soluções ótimas de (P) (isto foi comentado nas seções (4.2.1) e (4.2.1)). Um recente trabalho que aborda amplamente este tema é [36] e retoma novos

resultados apresentados em [35] válidos dentro de um enfoque que usa distâncias de Bregman. Comentamos os resultados que incluem o nosso caso.

Suponha alguma condição de qualificação de tal forma que o conjunto soluções \hat{S} dos problemas (P) e (D) satisfaça as condições de otimalidade.

Definição 5.19 *A condição de complementaridade estrita (CCE) se satisfaz para os problemas (P) e (D) se para todo $(\bar{x}, \bar{\mu}) \in \hat{S}$ e para $i = 1, \dots, m$ cumpre-se que ou $\bar{\mu}_i > 0$ ou $f_i(\bar{x}) < 0$.*

Teorema 5.20 *Suponha $\hat{S} \neq \emptyset$ e satisfeita a condição (CCE), então todo ponto limite da seqüência $\{x^k\}$ gerado pelo método lagrangeano aumentado, se existir, pertence ao conjunto de soluções ótimas do problema (P) .*

Comentários:

Uma desvantagem deste resultado é que a condição CCE de complementaridade estrita é uma hipótese forte, além disso, é suposta a existência de pontos de limites de $\{x^k\}$, em outras palavras, supõe-se que a seqüência $\{x^k\}$ é limitada. Kiwiel [41] considerou um lagrangeano duplamente aumentado, isto é, adiciona-se ao lagrangeano aumentado uma distância de Bregman que age como um regularizador tornando o lagrangeano resultante estritamente convexo. Define-se assim um método de lagrangeano aumentado cuja seqüência primal $\{x^k\}$ é limitada, contudo persiste a dificuldade em mostrar a viabilidade dos pontos limites de dita seqüência. Este duplo lagrangeano foi introduzido por Eckstein em [21]. Kiwiel define uma seqüência de médias de $\{x^k\}$ cujos pontos limites são soluções ótimas do problema (P) (convergência ergódica) sob hipótese de complementaridade no ponto (\bar{x}, μ^*) onde \bar{x} é um ponto limite da seqüência de medias e $\mu^* = \lim_{k \rightarrow +\infty} \mu^k$. Iusem e Burachick [35] mostram que esta complementaridade estrita não é necessária como hipótese para mostrar a convergência ergódica. Este último resultado também é válido para o lagrangeano aumentado não duplo e assim para o caso considerado nesta tese. Contudo, no seguinte teorema apresentamos uma prova de convergência ergódica sem hipótese de limitação para a seqüência primal e que aproveita o uso dos shifts. O resultado é válido quando usamos

indistintamente cada uma das famílias de penalidades $AL1$ ou $AL2$. Mostramos o resultado usando a família $AL2$.

Comentário:

Para o estudo da convergência de algoritmos usando barreiras do tipo ($AL1$) podemos considerar β constante, sem perda de generalidade, com $\beta > \tilde{\mu}$, onde $\tilde{\mu}$ é o maior valor de μ_i no conjunto de nível dual inicial (que é compacto). A função $\beta\theta(\cdot)$ pode ser considerada como função do tipo ($AL2$) com a condição iii) um pouco relaxada. Toda a análise de convergência será válida para a família ($AL1$). A seguir enunciamos um teorema que será usado como referência.

Corolário 5.21 (*Ver corolário 6.2-20 em Fiacco-McCormick [27]*) *Considere o problema (P) convexo satisfazendo a hipótese (A1). Então para k_0, k_1, \dots, k_m finitos, o conjunto*

$$S_k = \{x \in R^n \mid f_i(x) \leq k_i \quad i = 0, 1, \dots, m\}$$

é compacto.

Teorema 5.22 *Suponha válidas as hipóteses (A1) e (A2) com respeito ao problema (P). Considere as seqüências $\{x^k\}$, $\{\mu^k\}$, $\{r^k\}$ e $\{\beta^k\}$ geradas pelo método de multiplicadores com shifts definidas nos algoritmos (5.2) e (5.2) correspondentes às famílias de penalidades ($AL1$) e ($AL2$). Considere além destas a seqüência de médias primal $\{\bar{x}^k\}$ com $\bar{x}^k = \frac{\sum_{j=1}^k x^j / r^j}{\sum_{j=1}^k 1 / r^j}$. Então*

i) *Para $i = 1, \dots, m$, $\limsup_{k \rightarrow \infty} f_i(\bar{x}^k) \leq 0$ (Viabilidade ergódica).*

ii) *Para $i = 1, \dots, m$, $\lim_{k \rightarrow +\infty} \mu_i^k f_i(x^k) = 0$.*

iii) *Todo ponto limíte da seqüência de médias primal $\{\bar{x}^k\}$ é solução ótima do problema (P) (Convergência ergódica).*

iv) *Se além das condições anteriores tivermos adicionalmente a condição: se para todo $k \in N$, $i = 1, \dots, m$*

$$r^k \leq c \mu_i^k \text{ com } c > 0 \text{ constante ou}$$

$$|r^k \tilde{y}_i^k| \rightarrow 0.$$

Então para $i = 1, \dots, m$, $\lim_{k \rightarrow \infty} f_i(x^k) \leq 0$, $\{x^k\}$ é limitada e qualquer ponto limíte de $\{x^k\}$ é solução ótima de (P).

Prova:

Faremos a prova inicialmente usando a família *AL2*. Considere $\{\tilde{y}_i^k\}$ tal que $\theta'(\tilde{y}_i^k) = \mu_i^k$ para $i = 1, \dots, m$.

i) Mostremos que para $i = 1, \dots, m$, $\limsup_{k \rightarrow \infty} f_i(\bar{x}^k) \leq 0$.

Da iteração (5.13) do algoritmo 5.2 temos que para $i = 1, \dots, m$, $\forall k \in N$

$$\tilde{y}_i^{k+1} = \tilde{y}_i^k + \frac{y_i^{k+1}}{r^k} \quad (5.30)$$

onde $\theta'(\tilde{y}_i^{k+1}) = \mu_i^{k+1}$, $\theta'(\tilde{y}_i^k) = \mu_i^k$, $y_i^{k+1} = f_i(x^{k+1})$ e $r^k \in (0, 1]$.

Somando em ambos membros de (5.30) temos, para $i = 1, \dots, m$

$$\tilde{y}_i^k = \tilde{y}_i^0 + \sum_{j=1}^k \frac{y_i^j}{r^j}, \quad (5.31)$$

Por outra parte, como $f_i(\cdot)$ é convexa para $i = 1, \dots, m$, temos da definição de \bar{x}^k ,

$$f_i \left(\frac{\sum_{j=1}^k y_i^k / r^j}{\sum_{j=1}^k 1 / r^j} \right) \leq \frac{\sum_{j=1}^k f_i(x^j) / r^j}{\sum_{j=1}^k 1 / r^j} \quad (5.32)$$

Por teorema (5.18), sabemos que $\mu^k \rightarrow \mu^*$.

Para $i = 1, \dots, m$,

se $\mu_i^* > 0$ então $\tilde{y}_i^k \rightarrow (\theta')^{-1}(\mu_i^k)$, senão $\tilde{y}_i^k \rightarrow -\infty$.

Em ambos casos \tilde{y}_i^k é limitado, digamos $\tilde{y}_i^k \leq M \in \mathbb{R}$.

Segue-se de (5.31) que

$$\sum_{j=1}^k \frac{y_i^j}{r^j} \leq -\tilde{y}_i^0 + M, \quad (5.33)$$

assim, dividindo (5.33) por $\sum_{j=1}^k 1/r^j$ e usando (5.32) temos para $i = 1, \dots, m$

$$f_i \left(\frac{\sum_{j=1}^k y_i^k / r^j}{\sum_{j=1}^k 1 / r^j} \right) \leq \frac{\sum_{j=1}^k f_i(x^j) / r^j}{\sum_{j=1}^k 1 / r^j} \leq \frac{-\tilde{y}_i^0 + M}{\sum_{j=1}^k 1 / r^j}.$$

Como $r^k \leq 1$ para $k \in N$, o último denominador cresce ilimitadamente e portanto

$$\limsup_{k \rightarrow \infty} f_i(\bar{x}^k) \leq 0.$$

ii) A demonstração deste ítem foi inspirada em Polyak-Teboulle [50].

A função $d(\cdot)$ é côncava e por (4.4) $y^{k+1} \in \partial d(\mu^{k+1})$. Portanto

$$\sum_{i=1}^m (\mu_i^k - \mu_i^{k+1}) y_i^{k+1} \geq d(\mu^k) - d(\mu^{k+1}),$$

logo

$$\sum_{i=1}^m \mu_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) y_i^{k+1} \geq d(\mu^k) - d(\mu^{k+1}),$$

assim,

$$\sum_{k=1}^{\infty} \sum_{i=1}^m \mu_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) y_i^{k+1} \geq d(\mu^0) - d(\mu^k) > d(\mu^0) - \lim_{k \rightarrow \infty} d(\mu^k) > -\infty.$$

Sabemos de (5.30) que

$$\tilde{y}_i^{k+1} - \tilde{y}_i^k = \frac{y_i^{k+1}}{r^k},$$

$r^k \in (0, 1]$ e θ é convexa. Então a derivada $\mu_i = \theta'(\tilde{y}_i)$ é crescente. Assim, se $y_i^{k+1} \geq 0$ então $\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \leq 0$. Se $y_i^{k+1} < 0$ então $\frac{\mu_i^k}{\mu_i^{k+1}} - 1 > 0$, logo concluímos que todos os termos da soma acima são negativos e a soma é limitada por baixo, portanto

$$\sum_{i=1}^m \mu_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) y_i^{k+1} \rightarrow 0.$$

Logo, para $i = 1, \dots, m$

$$\mu_i^{k+1} y_i^{k+1} \left(\frac{\mu_i^k}{\mu_i^{k+1}} - 1 \right) \rightarrow 0. \quad (5.34)$$

Para $i = 1, \dots, m$, se $\mu_i^k \rightarrow \mu_i^* > 0$ então é fácil mostrar o resultado:

Temos $\mu_i^k = \theta'(\tilde{y}_i^k)$, $\frac{y_i^k}{r^k} = \tilde{y}_i^{k+1} - \tilde{y}_i^k$ e $\lim_{k \rightarrow \infty} \tilde{y}_i^k = (\theta')^{-1}(\mu_i^*)$

tomando limites, temos que

$$\frac{y_i^k}{r^k} \rightarrow 0 \text{ e como } r^k \in (0, 1], y_i^k \rightarrow 0.$$

Basta portanto estudar o caso em que $\mu_i^k \rightarrow 0$.

Suponha, tomando subsequências se for necessário, que para um dado $i = 1, \dots, m$,

$$\forall k \in N, \quad |\mu_i^{k+1} y_i^{k+1}| \geq \epsilon > 0.$$

Segue-se de (5.34)

$$\frac{\mu_i^k}{\mu_i^{k+1}} \rightarrow 1. \quad (5.35)$$

Como por hipótese, $\mu_i^k \rightarrow 0$, temos $|y_i^{k+1}| \geq \frac{\epsilon}{\mu_i^{k+1}}$. Defina $\bar{\theta} = \inf_{t \in R} \theta(t)$.

a) Se $y_i^{k+1} > 0$, usando $\tilde{y}_i^{k+1} - \tilde{y}_i^k = \frac{y_i^{k+1}}{r^k}$ temos pela convexidade de θ e $\mu_i^k = \theta'(\tilde{y}_i^k)$,

$$\begin{aligned}\theta(\tilde{y}_i^{k+1}) &\geq \theta(\tilde{y}_i^k) + \frac{\mu_i^k y_i^{k+1}}{r^k} \geq \frac{\epsilon}{r^k} \frac{\mu_i^k}{\mu_i^{k+1}} + \theta(\tilde{y}_i^k) \\ &\geq \bar{\theta} + \frac{\epsilon}{r^k} \frac{\mu_i^k}{\mu_i^{k+1}}\end{aligned}$$

Usando (5.35), para k grande e observando que $r^k \in (0, 1]$,

$$\theta(\tilde{y}_i^{k+1}) \geq \frac{\epsilon}{2} + \bar{\theta},$$

$$\tilde{y}_i^{k+1} > \theta^{-1}(\epsilon/2 + \bar{\theta}) \equiv \gamma \text{ porque a função } \theta \text{ é crescente,}$$

assim

$$\mu_i^{k+1} = \theta'(\tilde{y}_i^{k+1}) > \theta'(\gamma) > 0. \text{ Isto contradiz } \mu_i^k \rightarrow 0.$$

b) Se $y_i^{k+1} < 0$

$$\begin{aligned}\theta'(\tilde{y}_i^k) &\geq \theta'(\tilde{y}_i^{k+1}) - \mu_i^{k+1} \frac{y_i^{k+1}}{r^k} \\ &\geq \bar{\theta} + \frac{\epsilon}{r^k}\end{aligned}$$

e a contradição segue como no caso anterior.

iii) Sabemos x^k minimiza o lagrangeano $l(\cdot, \mu^k)$, logo usando ii) temos

$$\begin{aligned}\lim_{k \rightarrow +\infty} d(\mu^k) &= \lim_{k \rightarrow +\infty} [f_0(x^k) + \sum_{i=1}^m \mu_i^k f(x_i^k)] \\ &= \lim_{k \rightarrow +\infty} f_0(x^k)\end{aligned} \tag{5.36}$$

Usando o teorema de dualidade fraca temos

$$\forall k, \quad d(\mu^k) \leq d(\mu^*) = f(x^*)$$

Pela definição de $\{\bar{x}^k\}$ e convexidade de f_0 temos

$$f_0(\bar{x}^k) \leq \frac{\sum_{j=1}^k f_0(x^j)/r^j}{\sum_{j=1}^k r^j}$$

Assim

$$\limsup_{k \rightarrow \infty} f_0(\bar{x}^k) \leq \lim_{k \rightarrow \infty} \frac{\sum_{j=1}^k f_0(x^j)/r^j}{\sum_{j=1}^k r^j} = f_0^* \tag{5.37}$$

porque $f_0(x^k) \rightarrow f_0^*$ e $r^k \in (0, 1]$. Dado ϵ , para k suficientemente grande, de (5.37)

e ii) temos

$$\bar{x}^k \in \{x \in R^n \mid f_0(x) \leq f_0^* + \epsilon, \quad f_i(x) \leq \epsilon, \quad i = 1, \dots, m\},$$

por teorema 5.21, este conjunto é compacto e assim $\{\bar{x}^k\}$ é limitada.

Dado \hat{x} , ponto limite de $\{\bar{x}^k\}$, então $\bar{x}^k \xrightarrow{K} \hat{x}$, com $K \in N$.

Assim $f_0(\hat{x}) \leq f^*$ e \hat{x} é viável.

Segue-se \hat{x} é solução ótima de (P).

iv) Sabemos que $\mu^k \rightarrow \mu^*$. Tome $i = 1, \dots, m$. Se $\mu_i^k \rightarrow \mu_i^* > 0$, o resultado é imediato por ii).

Suponha que $\mu_i^k \rightarrow 0$, portanto $\tilde{y}_i^k \rightarrow -\infty$, $\tilde{y}_i^{k+1} \rightarrow -\infty$.

(1) Supondo $r^k \leq c\mu_i^k$, $\forall k \in N$.

Temos, pela convexidade de θ ,

$$\begin{aligned} \theta(\tilde{y}_i^{k+1}) &\geq \theta(\tilde{y}_i^k) + \mu_i^k \frac{y_i^{k+1}}{r^k} \\ &\geq \theta(\tilde{y}_i^k) \frac{y_i^{k+1}}{c} \end{aligned}$$

tomando limite temos

$$\bar{\theta} \geq \bar{\theta} + \frac{1}{c} \limsup_{k \rightarrow \infty} f_i(x^{k+1}),$$

Assim $\limsup_{k \rightarrow \infty} f_i(x^{k+1}) \leq 0$.

(2) Supondo $|r^k \tilde{y}_i^k| \rightarrow 0$, temos

$$\tilde{y}_i^{k+1} = \tilde{y}_i^k + \frac{y_i^{k+1}}{r^k}$$

logo

$$y_i^{k+1} = -r^k \tilde{y}_i^k + r^k \tilde{y}_i^{k+1}$$

Tomando limites, como $\tilde{y}_i^{k+1} \rightarrow -\infty$ e $r^k y_i^k \rightarrow 0$. temos

$$\limsup_{k \rightarrow \infty} f_i(x^{k+1}) \leq 0.$$

■

Exemplos

A primeira condição em iv) é idêntica à usada por BenTal e Zibulevsky. A segunda é geralmente mais fraca que a primeira. Por exemplo, para $\theta(y_i) = e^{y_i}$ temos em cada iteração

$$\mu_i = \theta'(\tilde{y}_i) = e^{\tilde{y}_i} \text{ e } \tilde{y}_i = \log \mu_i$$

basta tomar r tal que

$$r \log \mu_i \rightarrow 0$$

ou seja, $r = o(\frac{-1}{\log \mu_i})$, o que é muito maior que μ .

Para $\theta(y) = y + \sqrt{y^2 + 4}$,

desenvolvendo $\theta'(\tilde{y}_i) = \mu_i$ temos

$$\tilde{y}_i = \frac{2(\mu_i - 1)}{\sqrt{\mu_i} \sqrt{2 - \mu_i}},$$

temos então $-\tilde{y}_i = O(\frac{2}{\sqrt{\mu_i}})$.

Basta fazer $r^k \rightarrow 0$ mais rápido que $\sqrt{\mu_i}$, isto é, $r^k = o(\sqrt{\mu})$.

Para a função $\theta(y) = -\log(1 - y)$ ambas condições são similares.

5.4 Relaxação da família de penalidades AL2

Nesta seção consideraremos as funções de penalidade θ correspondentes às penalidades do tipo II eliminando a condição vi), isto é, prescindimos de que θ seja limitada por baixo. Isto conduz a que a sua conjugada θ^* , uma vez feita a mudança de variáveis, nãoesteja definida na origem e portanto não temos definida uma distância de Bregman em $S_2 = [0, +\infty)$ pois algumas das propriedades deixam de ser satisfeitas. Mas agora, nos restringimos naturalmente a seu interior $S_2 = (0, +\infty)$ onde ressaltamos se satisfazem as propriedades B1, B3, e B4 que principalmente usaremos. Embora não mostremos a convergência de toda a seqüência $\{\mu^k\}$ correspondente ao método de ponto proximal considerado, garantimos que todo ponto limite desta seqüência é solução ótima do problema dual.

Consideramos a função de penalidade do tipo II, $\theta : R \rightarrow R \cup \{+\infty\}$ convexa, própria, fechada, crescente e diferenciável com $\text{ridom } \theta = (-\infty, b)$, $b \in R_{++}$ tal que:

- i) θ é estritamente convexa em $(-\infty, b)$.
- ii) θ é diferenciável em $(-\infty, b)$.
- iii) $\theta'_\infty(1) = +\infty$. (coercividade pela direita)
- iv) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$.

v) $\theta(0) = 0$.

A forma da quase-distância correspondente é

$$s, \mu \in R_{++}^m \mapsto D_{h_2}(s, \mu) = h_2(s_i) - h_2(\mu_i) - \nabla h(\mu_i)^t (s_i - \mu_i), \quad (5.38)$$

com $h_2(s) = \sum_{i=1}^m \theta^*(s_i)$.

O método de ponto proximal para resolver o problema (D) usando a quase-distância (5.38) gera uma seqüência tal que:

$$\mu^0 \in S_2$$

$$\mu^{k+1} \in \operatorname{argmin}_{\mu \in \bar{S}_2} \{-d(\mu) + r^k D_{h_2}(\mu, \mu^k)\} \quad (5.39)$$

onde $r^k \in (0, 1]$.

Observe que da condição de otimalidade para este problema temos

$$r^k [\nabla h(\mu^k) - \nabla h(\mu^{k+1})] \in \partial(-d(\mu^{k+1})) \quad (5.40)$$

Teorema 5.23 *Se o problema (D) tem solução e θ^* é coerciva na fronteira, então com respeito à seqüência gerada por (5.39) temos*

i) *A seqüência $\{\mu^k\}$ está bem definida em S_2 .*

ii) *Seja $\bar{\mu} \in S_2$ tal que $d(\bar{\mu}) < d(\mu^{k+1})$, se satisfaz*

$$D_{h_2}(\bar{\mu}, \mu^{k+1}) \leq D_{h_2}(\bar{\mu}, \mu^k) - D_{h_2}(\mu^{k+1}, \mu^k) - \frac{1}{r^k} [d(\mu^{k+1}) - d(\bar{\mu})].$$

iii) *A seqüência de valores $d(\cdot)$ converge ao valor ótimo d^* de (D) e portanto todo ponto limite da seqüência gerada por (5.39) é solução do problema (D) .*

Prova:

i) Segue-se por passo 1 do teorema 4.1 em Iusem [34].

ii) Note que a expressão i) em 3.12 dada por

$$D_h(x, y) - D_h(x, z) - D_h(z, y) = (\nabla h(y) - \nabla h(z), z - x)$$

para todo $x \in \bar{S}$, $y, z \in S$, é válida neste caso para todo $x \in S_2$, $y, z \in S_2$ e $h = h_2$, assim substituindo x por $\bar{\mu}$, y por μ^k e z por μ^{k+1} temos de 5.40, fazendo

$y^k = r^k[\nabla h_2(\mu^k) - \nabla h_2(\mu^{k+1})]$ que

$$\begin{aligned} D_{h_2}(\bar{\mu}, \mu^k) - D_{h_2}(\bar{\mu}, \mu^{k+1}) - D_{h_2}(\mu^{k+1}, \mu^k) &= (\nabla h(\mu^k) - \nabla h(\mu^{k+1}), \mu^{k+1} - \bar{\mu}) \\ &\geq \frac{1}{r^k}(y^k, \mu^{k+1} - \mu^k) \\ &\geq [d(\mu^{k+1}) - d(\bar{\mu})]. \end{aligned}$$

O resultado segue-se transpondo termos.

iii) Suponha por absurdo que $\lim_{k \rightarrow +\infty} d(\mu^k) < d(\mu^*)$, onde μ^* é uma solução ótima de (D) . Logo, para $\epsilon > 0$ temos

$$\lim_{k \rightarrow +\infty} d(\mu^k) = d(\mu^*) - \epsilon.$$

Podemos escolher $\tilde{\mu} \in \text{int } S$ tal que $d(\tilde{\mu}) > d(\mu^*) - \epsilon/2$,

Temos assim que $(\forall k \in \mathbb{N})$, $d(\mu^k) < d(\tilde{\mu}) - \epsilon/2$, logo por ii), observando que $D_{h_2}(\mu^{k+1}, \mu^k) \geq 0$ se satisfaz

$$\begin{aligned} D_{h_2}(\tilde{\mu}, \mu^{k+1}) &\leq D_{h_2}(\tilde{\mu}, \mu^k) - \frac{1}{r^k}[d(\mu^{k+1}) - d(\tilde{\mu})] \\ &\leq D_{h_2}(\tilde{\mu}, \mu^k) - \frac{\epsilon}{2}. \end{aligned}$$

Daqui, $D_{h_2}(\tilde{\mu}, \mu^k) \rightarrow -\infty$ contradizendo B4.

Assim, finalmente por corolário 27.2.1 em [54], segue-se que todo ponto limite de $\{\mu^k\}$ é solução ótima de (D) . ■

Capítulo 6

Conclusões

Os resultados principais da presente tese podem-se separar naqueles correspondentes a métodos de penalidades, como foi observado no capítulo 1, e aqueles correspondentes a métodos de multiplicadores que foram abordados no capítulo 5, contudo que esta última metodologia pode ser considerada como um tipo particular de penalidade. Precisamente, este enfoque foi de utilidade na hora de obter resultados. Consideramos inicialmente uma família de funções que aproximam à função de penalidade exata seguindo um formato que define a esta última como sua função de recessão, mostramos que os métodos conhecidos e apresentados seguem basicamente o mesmo formato ainda que nos trabalhos originais isto não foi percebido com exceção do trabalho de Ben-Tal e Teboulle ([7]) que usa um específico tipo de função nesta família. Recentemente e paralelamente a nossa abordagem, Auslender ([3]) apresenta um trabalho mais elaborado abordando estas aproximações e aplicado a problemas de programação não-linear baseados em penalidades e abarcando problemas de programação semi-definida. Apresentamos um algoritmo para resolver o problema (P) mediante um esquema de penalidade onde intervêm as aproximações da penalidade exata parametrizadas por β e r . β é responsável pelo aumento da penalidade e r por a precisão da aproximação. Este enfoque, generaliza o método usado por Xavier ([62]) para um caso específico de função de penalidade e com diferentes hipóteses. Auslender apresenta um esquema de minimização similar usando a mesma família de penalidades para resolver

o problema (P) com hipótese de convexidade, nesse caso precisou incrementar a penalidade indefinidamente para mostrar convergência. Em nossa abordagem, consideramos o caso nãoconvexo diferenciável e sob a condição de Mangasarian-Fromovitz mostramos que a penalidade não precisa crescer indefinidamente. Isto abre a possibilidade de aplicar a mesma metodologia em diferentes problemas de otimização, podendo comparar o desempenho numérico para problemas de grande porte e aproveitando o uso de eficientes métodos computacionais uma vez que as penalidades envolvidas são duas vezes diferenciáveis.

Note que o método de lagrangeano aumentado surge inicialmente como uma ferramenta poderosa que melhora o desempenho computacional dos métodos de penalidade pois o parâmetro encarregado de aumentar a penalidade não precisa crescer indefinidamente e aproveita além disto a diferenciabilidade das funções de penalidade. Desta forma a idealizada função de penalidade exata é deslocada inicialmente como metodologia eficiente para resolver o problema (P) via penalização. Posteriormente observamos a tendência em usar funções de penalidade duas vezes diferenciáveis e coercivas pela direita, as quais garantem em forma natural um incremento na penalidade. Surge então um questionamento: É possível usar uma aproximação da penalidade exata como função de penalidade envolvida numa metodologia tipo lagrangeano aumentado?. Isto é respondido afirmativamente no capítulo 5 com hipóteses de convexidade através de uma novedosa mudança de variáveis que pode ser aplicada não somente às aproximações da função de penalidade exata (família ($AL1$)) senão também à família de penalidades ($AL2$) que inclui o uso da função exponencial.

Esta mudança, manipula de forma implícita a seqüência de multiplicadores e faz com que cada função de penalidade nas famílias consideradas tenha derivada na origem igual a $\mu > 0$. Para o caso da família ($AL1$), sob hipóteses razoáveis, mostramos a existência de um vetor $\bar{\beta}$ tal que para $\beta > \bar{\beta}$ o conjunto de soluções do problema primal é nãovazio e compacto. Definimos um algoritmo equivalente aos métodos de lagrangeano aumentado para cada uma das famílias de penalidades consideradas mostrando também que para o caso da família ($AL1$) o parâmetro

β envolvido é limitado. Um fato importante a destacar é que ambas famílias de penalidade conduzem a distâncias de Bregman via dualidade de Fenchel, assim podemos aplicar os resultados já conhecidos neste campo para o análise de convergência. Além disso, mostramos resultados de convergência ergódica primal sem precisar da hipótese de que a seqüência primal $\{x^k\}$ seja limitada, melhorando assim o resultado obtido em Iusem e Burachik [35]. Para o caso das penalidades da família $(AL1)$, precisamos que o parâmetro envolvido na conjuga seja fixo para poder aplicar o método de ponto proximal tal como é conhecido, esta limitação do parâmetro é mostrada e por tanto basta com usar uma limitante superior do mesmo. Além disto, o caso da família $(AL1)$ generaliza o método de lagrangeano aumentado usado por Xavier [62] usando a penalidade hiperbólica. Futuros testes computacionais podem ser feitos para avaliar o desempenho numérico dos métodos aqui propostos.

Bibliografia

- [1] Attouch H. e Wets R., “Epigraphical analysis”, Ann. Ins. Poincare: Anal. Nonlineaire 6, 1989.
- [2] Auslender A., “Penalty and Barrier Methods: Unified Framework”, Relatorio do Laboratoire D’econometrie de L’ecole Polytechnique, Paris, 1977.
- [3] A. Auslender, R. Cominetti and M. Haddou, “Asymptotic analysis for penalty and barrier methods in convex and linear programming ”, Mathematics of Operatons Research, vol 22, Nro. 1 (1997), pp.43-62.
- [4] Auslender A. e Haddou M., “An interior-proximal method for convex linearly constrained problems and its extension to variational inequalities”, Mathematical Programming, 71, pp. 77–100, 1.995.
- [5] Bazaraa M. e Goode J. “Sufficient conditions for a globally exact penalty function without convexity”, Mathematical Programming, 19, pp. 1–15, 1.982.
- [6] “ Ben-Tal A. e Zibulevsky M., “Penalty-Barrier multiplier methods for convex programming problems”, SIAM J. on Optimization, 7, Nro. 2, pp. 347–366, 1.997.
- [7] Ben-Tal A. e Teboulle M., “A smoothing technique for nondifferentiable optimization problems”, Optimization. Fifth French German Conference. Lecture Notes in Mathematics 1405, Spring Verlag, pp. 1–11, 1.989.
- [8] Bertsekas D. P., *Constrained Optimization and Lagrange Multiplier Méthods*, Academic Press, New York, 1982.

- [9] Bertsekas D. P., “Approximation procedures based on the method of multipliers”, JOTA, vol. 23, Nro. 4, pp. 487–510. 1.977.
- [10] Bregman L., “The relaxation method for finding the common point of convex sets and its application to the solution of problems in convex programming”, USSR Computational Mathematics and Mathematical Physics, vol 7, pp. 220–217, 1967.
- [11] Burachik R. S. e Iusem A., “A generalized proximal point algorithm for the variational inequality problem in a Hilbert space”, Work paper, IMPA, Rio de Janeiro, 1.991.
- [12] Censor Y. e Zenios S., “The proximal minimization algorithm with D-functions”, JOTA, vol 73, pp.451–464, 1992.
- [13] Chen G. e Teboulle M., “onvergence analysis of a proximal-like minimization algorithm using Bregman function”, SIAM J. Optimization, 3, pp. 538–543, 1993.
- [14] C. Chen and O. L. Mangasarian, “Smoothing methods for convex inequalities and linear complementary problems”, Mathematical Programming, vol 71 (1995), Nro 1, pp.1–112.
- [15] C. Chen and O. L. Mangasarian, *A class of smoothing fuctions for non-linear and mixed complementary problems*, Mathematical Programming Technical report 94-11, August 1.994.
- [16] Csiszar I., “Information-type measures of difference of probability distributions and indirect observations”, Studia Sci. Math. Hungar, 2, pp. 299–318, 1967.
- [17] R. Cominetti and J. San Martin, “Asymptotic analysis of the exponential penalty trajectory in linear programming”, Mathematical Programming, 67 (1994) pp. 169–187.

- [18] Conn A., Gould N. Toint Ph., *LANCELOT: a Fortran package for large-scale nonlinear optimization (release A)*, Springer Verlag, Berlin, 1992.
- [19] De Pierro A. e Iusem A., “A relaxed version of Bregman’s method for convex programming”, *JOTA*, 5, pp. 421–440, 1986.
- [20] Eaves B. C. e Zangwill W. I., “Generalized cutting plane algorithms”, *SIAM J. on Control* 9, pp. 529–542, 1971.
- [21] Eckstein J., “Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming”, *Mathematics of Operations Research*, 18, pp. 202–226, 1993.
- [22] Eggermont P., “Multiplicative Iterative Algorithms for convex programming”, *Linear Algebra Appl.*, pp. 25–42, 1990.
- [23] Eriksson J., *Algorithms for entropy and mathematical programming*, D.sc. Dissertation, Department of Mathematics, Linköping University, 1981.
- [24] Eriksson J., “An iterative primal-dual algorithm for linear programming”, Technical report 1985-10, Department of Mathematics, Linköping University, 1985.
- [25] Erlander S., “Entropy in linear programs”, *Mathematical Programming*, 21, pp. 137–151, 1.981.
- [26] Ermol’ev Y. M., “On the method of generalized stochastic gradients and quasi-Fejér sequences”, *Cybernetics* 5, pp. 208–220, 1.969.
- [27] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, SIAM, New York, 1.990.
- [28] Jean B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms I*, 1 ed. New York, Springer-Verlag, 1993.
- [29] Jean B. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms II*, 1 ed. New York, Springer-Verlag, 1993.

- [30] Hestenes M., “Multiplier and Gradient Methods”, JOTA, vol. 4, pp.303–320, 1.969.
- [31] Humes C. e da Silva P., *Proximal point, separators and augmented lagrangians*, Work paper, U.S.P. São Paulo, Brasil, 1997.
- [32] Iusem A., *Métodos de ponto proximal em Otimização*, 20mo Colóquio Brasileiro de matemática, IMPA, R.J. Brasil, 1995.
- [33] Iusem A., Teboulle M. e Svaiter B., “Entropy-like proximal methods in convex programming”, Mathematics of Operations Research, 19(4), pp. 790–814, 1994.
- [34] Iusem A., “Some properties of generalized proximal point methods for quadratic and linear programming”, JOTA , vol 85, Nro. 3, 1993.
- [35] Iusem A. Burachik R., “Generalized proximal point and augmented Lagrangian methods for convex optimization”, Work Paper, IMPA, Rio de Janeiro, 1.997.
- [36] Iusem A., *Augmented Lagrangian methods and proximal point methods for convex optimization*, Minicurso I workshop em otimização, Florianópolis, Brasil, Dezembro 1.997.
- [37] Iusem A. e Monteiro R., *On dual convergence of the generalized proximal point method with Bregman distances*, Work paper IMPA a ser publicado, Outubro 1.997.
- [38] Iusem A. e Teboulle M., “On the convergence rate of entropic proximal optimization methods”, Computational and Applied Mathematics, 12, pp. 153–168, 1.993.
- [39] Iusem A. e Teboulle M., “Convergence rate of nonquadratic proximal point methods for convex and linear programming”, Mathematics of Operations Research, 20, pp. 657–677, 1.995.

- [40] Iusem A., Svaiter B. e Xavier da Cruz Neto J., *Central Paths, generalized proximal point methods and cauchy trajectories in riemannian manifolds*, Work paper IMPA-UFRJ, Brasil, 1.994.
- [41] Kiwiel K., “Proximal minimization methods with generalized Bregman functions”, *SIAM J. on Control and Optimization*, 35, 1142-1168, 1997.
- [42] Lemaire B., *The proximal algorithm*, In International Series of Numerical Mathematics (J. P.Penot, editor). Birkhauser, Basel, 87, pp. 73–87 , 1989.
- [43] Luque F. J., “Asymptotic convergence analysis of the proximal point algorithm”, *SIAM J. and Optimization*, vol. 22, Nro. 2, pp. 277–293, 1.984.
- [44] Madsen K. e Nielsen H. B., “A finite algorithm for linear l_1 estimation”, *SIAM J. and Optimization*, vol. 3, Nro. 2, pp. 223–235, 1.993.
- [45] O. L.Mangasarian and S. Fromovitz, “The Fritz John necessary comditions in the presence of equality comstraints”, *Journal of Mathematical Analysis and Applications*, 17 (1967), pp34–47.
- [46] Martinet B., “ Regularisation D’inequations variationnelles par approximations successives”, *Revue Française de Informatiqué et Recherche Opérationelle* 2, pp. 154–159, 1970.
- [47] Minty G. , “Monotone nonlinear operators in Hilbert space”, *Duke Mathematical Journal* 29, pp. 341–346, 1978.
- [48] Moreau J. J., “Proximité et dualité dans un espace Hilbertien”, *Bull. Soc. Math. France* 93, pp. 273–299, 1965.
- [49] Pinar M. e Zenios S., “On smoothing exact penalty functions for convex constrained optimization”, *SIAM J. Optimization*, vol 4, Nro. 3, pp. 486–511, 1994.
- [50] Polyak R. e Teboulle M., “Nonlinear rescaling and proximal-like methods in convex programming”, 76, pp. 697–739, 1.997.

- [51] Polyak R., “Smooth Optimization methods for minimax problems”, SIAM J. Control and Optimization, vol. 26, Nro.6, pp. 1274–1286, 1.988.
- [52] Powell M., “A method for nonlinear constraints in minimizations problems”, in Optimization, R.Fletcher, Ed.,Academic Press, N.Y, pp.283–298, 1.969.
- [53] Powell M., “Some convergence properties of the shifted log barrier method for linear programming”, SIAM J. on Optimization, 5, pp. 697–739, 1995.
- [54] Rockafellar R. T. *Convex Analysis*,1 ed. Princeton University Press, Princeton, N.J.1970.
- [55] Rockafellar R. T., “Monotone operators and the proximal point algorithm in convex programming”, SIAM J. on Control and Optimization 14, pp. 877–898, 1976.
- [56] Rockafellar R. T., “Augmented Lagrangians and applications of the proximal point algorithm in convex programming”, Mathematics of Operations Researchs 1, pp. 97–116, 1976.
- [57] Rockafellar R. T., “The multiplier method of Hestenes and Powell applied to convex programming”, JOTA, vol. 12, nro. 6, pp. 555–562, 1.973.
- [58] Sosa W., “Regularizações para problemas de minimização convexos”, M.sc. dissertação, C.O.P.P.E. Universidade federal de Rio de Janeiro, Brasil, 1996.
- [59] Teboulle M., “Entropic proximal mappings with applications to nonlinear programming”, Mathematics of Operations Research 17, pp. 97–116, 1992.
- [60] Tishler A. e Zang I., “A new maximum likelihood method for piecewise regression”, Working Paper Nro.526/77/R. Faculty of Management. Tel Aviv University, Agosto 1977.
- [61] Tseng P. e Bertsekas D., “ On the convergence of the exponential multiplier method for convex programming”, Mathematical Programming, 60, pp. 1–19, 1.993.

- [62] A. E. Xavier, *Hyperbolic Penalization*, D.Sc Thesis, Coppe, U.F.R.J. Brazil. 1.992.
- [63] Zang I., "A Smoothing-out technique for min-max optimization", *Mathematical Programming*, 19, , 1980, pp. 61-77.
- [64] W. I. Zangwill, " Non-linear programming via penalty functions", *Management Science*, 13 Nro. 5 (1967), pp. 344–358.