

ALGUMAS APLICAÇÕES DA GEOMETRIA RIEMANNIANA À OTIMIZAÇÃO

Erik Alex Papa Quiroz

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

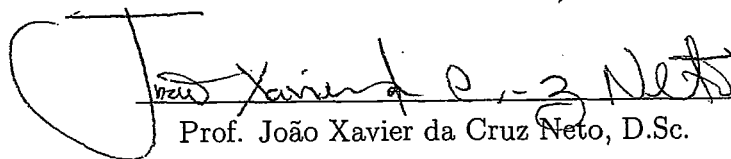
Aprovada por:



Prof. Paulo Roberto Oliveira, Dr. Ing.



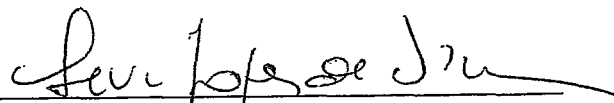
Prof. Susana Scheimberg de Makler, D.Sc



Prof. João Xavier da Cruz Neto, D.Sc.



Prof. Gregorio Malajovich Munoz, D.Sc



Prof. Levi Lopes de Lima, D.Sc

RIO DE JANEIRO, RJ-BRASIL

MARÇO DE 2007

PAPA QUIROZ, ERIK ALEX

Algumas aplicações da geometria Riemanniana à otimização [Rio de Janeiro] 2007

X, 139 p. 29,7 cm (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação, 2007)

Tese - Universidade Federal do Rio de Janeiro, COPPE

- 1 - Variedades Riemannianas
- 2 - Otimização Quase-Convexa
- 3 - Métodos de Ponto Proximal
- 4 - Método de Máxima Descida
- 5 - Barreiras Auto-concordantes
- 6 - Métricas Diagonais
- 7 - Otimização Semidefinida

I. COPPE/UFRJ II. Título (série)

Ao meus pais Maria e Alejandro, com carinho.
Aos meus irmãos Charo, Rosa, Fernando e Isaias.
Aos meus sobrinhos Luis Carlos, Jesús, Melissa, Muriel e Zayuri.

Agradecimentos

Em primeiro lugar, agradeço aos professores e amigos Angel Coca Balta e Edinson Montoro Alegre, docentes da Universidade Nacional del Callao, Lima-Perú, pelas cartas de recomendação para me candidatar ao programa de Engenharia de Sistemas e Computação da Pós-Graduação da COPPE. Em particular, gostaria de agradecer ao Angel Coca pela ajuda recebida antes e depois de minha chegada ao Rio de Janeiro, obrigado Angel por sua grande qualidade de ajuda ao estudante peruano.

Agradeço aos professores do programa pela formação acadêmica, em particular aos professores da otimização: Susana Scheimberg, Regina Burachik, Marcia Fampa, Paulo Roberto Oliveira e Adilson Xavier. Agradeço também aos professores do IM-UFRJ e do IMPA, onde recebi aulas de diversas disciplinas. Em especial agradeço aos professores Maria Fernanda Elbert e Manfredo Perdigão do Carmo com quenes aprendi a geometria Diferencial e Riemanniana.

Agradeço ao meu orientador Paulo Roberto Oliveira pelo convite para estudar aplicações da geometria Riemanniana à otimização e pela motivação para continuar estudando esta linha de pesquisa. Agradeço, a ele também, o convite para fazer o doutorado direto e principalmente pela liberdade de pesquisa com que sempre trabalhei.

Agradeço aos professores da minha banca de doutorado pela leitura desta tese e por suas valiosas sugestões. Um agradecimento especial ao professor Luis Mauricio Graña Drummond, que infelizmente não pudesse estar presente na data da defesa, pelas multiples correções de uma versão preliminar desta tese (exame de qualificação). Asi também agradeço aos colegas Francisco Gêvane Muniz Cunha e Sissy da Silva Souza pelas correções iniciais do português.

Agradeço a todos meus colegas, coterraneos e amigos com quenes compartilhe momentos de estudo e entretenimento. Em particular agradeço as grandes amizades do Alvaro Cuno Parari (quem sempre me tira de apuros tanto acadêmicos como pessoais, obrigado alvarito pela sua

amizade e pelos bons momentos vividos no ano 2002) e da Marisa Quispe Cárdenas (sempre presente em todos os momentos de minha carreira profissional). Obrigado também aos amigos Mariella Berrocal e Raúl Carita por toda a ajuda recibida em especial no ano 2003.

Finalmente, agradeço ao povo brasileiro pela hospitalidade e em especial à gente da Vila do Pinheiro onde passei os momentos mais felices de minha estadia no Rio.

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

ALGUMAS APLICAÇÕES DA GEOMETRIA RIEMANNIANA À OTIMIZAÇÃO

Erik Alex Papa Quiroz

Março/2007

Orientador: Paulo Roberto Oliveira

Programa: Engenharia de Sistemas e Computação

Esta tese está dividida em 5 partes. Na primeira estudamos as propriedades geométricas de uma classe de métricas Riemannianas diagonais sobre o ortante positivo \mathbb{R}_{++}^n e o hipercubo $(0, 1)^n$, espaços naturais onde se definem as restrições de alguns problemas de otimização. Obtemos alguns resultados úteis visando novos métodos de pontos interio-res na perspectiva das aplicações da geometria Riemanniana à otimização. Na segunda parte estendemos os resultados de convergência global do método de máxima descida com busca de Armijo generalizada e uma regularização proximal para resolver problemas de minimização em variedades Riemannianas quando a função objetivo é quase-convexa. Na terceira parte, generalizamos o método de ponto proximal com distâncias de Bregman para resolver problemas de otimização sobre variedades de Hadamard para funções convexas e quase-convexas. Na quarta introduzimos uma nova barreira auto-concordante para o hipercubo, estudamos propriedades de convergência da sua trajetória primal-dual, e, para problemas de otimização linear com variáveis limitadas, apresentamos também alguns novos algoritmos primais. Finalmente, na quinta parte, estendemos os resultados obtidos na parte anterior para resolver uma classe de problemas de otimização semidefinida.

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

SOME APPLICATIONS OF THE RIEMANNIAN GEOMETRY FOR OPTIMIZATION

Erik Alex Papa Quiroz

March/2007

Adviser: Paulo Roberto Oliveira

Department: Systems and Computation Engineering

This thesis is divided in five parts. In the first part, we study the geometric properties of a class of diagonal Riemannian metrics on the positive orthant \mathbb{R}_{++}^n and hypercube $(0, 1)^n$, natural spaces where are defined some optimization problems. We obtain some useful results to obtain new interior point methods in the point of view of applications of Riemannian geometry tools. In the second part, we extend the global convergence result of the steepest descent method with a generalized Armijo search and a proximal regularization for solving minimization problems on Riemannian manifolds when the objective function is quasiconvex. In the third part, we generalize the proximal point method with Bregman distances to solve minimization problems defined on Hadamard manifolds for quasiconvex and convex functions. In the fourth part, we introduce a new self-concordant barrier for the hypercube and we study convergence properties of the primal-dual central path and, for solving linear optimization problems with bounded variables, we also introduce some new primal algorithms. Finally, in the fifth part, we extend our results to solve a class of semidefinite optimization problems.

Índice

1	Introdução	1
2	Preliminares	6
2.1	Símbolos e Notações	6
2.2	Elementos Básicos de Geometria Riemanniana	7
2.3	Teoremas da Função Inversa e da Função Implícita	10
2.4	Funções de matrizes	11
2.5	Conjuntos Semi-Analíticos	12
3	Métricas Riemannianas Diagonais no Ortante Positivo e no Hipercubo	13
3.1	Introdução	13
3.2	Métricas Riemannianas Invariantes	15
3.3	Propriedades Geométricas de uma Classe de Métrica Diagonal	21
3.4	Exemplos de Geodésicas Explícitas e Variedades Completas em \mathbb{R}_{++}^n e $(0, 1)^n$:	30
3.5	Aplicações	32
3.5.1	Métodos Geodésicos	33
3.5.2	Método de Ponto Proximal	37
3.5.3	Barreiras para Problemas Canalizados	38
4	Método de Máxima Descida em Variedades Riemannianas para Funções Quase-Convexas	40
4.1	Introdução	40
4.2	Elementos Básicos	41

4.3	O Método de Máxima Descida	42
4.3.1	Método com uma busca de Armijo generalizada	43
4.3.2	Método com uma regularização proximal	46
4.4	Alguns Exemplos	47
5	Método de Ponto Proximal com Distâncias de Bregman em Variedades de Hadamard	50
5.1	Introdução	50
5.2	Análise Convexa em Variedades de Hadamard	52
5.3	Distâncias e Funções de Bregman em Variedades de Hadamard	55
5.4	Regularização	59
5.5	Método de Ponto Proximal	61
5.6	Resultados de Convergência	63
5.6.1	O caso quase-convexo	63
5.6.2	O caso convexo	65
5.7	Métodos de Ponto Proximal com distâncias Riemannianas	67
5.7.1	O caso quase-convexo	68
5.7.2	O caso convexo	69
5.8	Alguns Exemplos de Distâncias de Bregman	70
6	Uma Nova Barreira Auto-concordante para o Hipercubo	75
6.1	Introdução	75
6.2	A Nova Barreira	76
6.3	Uma Nova Trajetória Central para Problemas Convexos Canalizados com Restrições Lineares	78
6.3.1	A Trajetória Central	79
6.3.2	A Trajetória Central Primal	82
6.3.3	A Trajetória Central Primal-Dual	87
6.4	Aplicação para Otimização Linear	92
6.4.1	Propriedades em Vizinhanças da Trajetória Central	95
6.4.2	Um Algoritmo de Passos Longos Seguidor da Trajetória Central Primal	97
6.4.3	Análise da Complexidade	99

6.5	Um Método de Ponto Proximal	102
7	Nova Barreira para uma Classe de Problemas de Otimização Semidefinida	105
7.1	Introdução	105
7.2	Definição do Problema	105
7.3	Exemplos	106
7.4	Uma Nova Barreira e sua Trajetória Central	108
7.4.1	A Nova Barreira	108
7.4.2	Definição da Trajetória Central	109
7.4.3	Propriedades da Trajetória Central Primal	112
7.4.4	Propriedades da Trajetória Central Primal-Dual	117
7.5	Um Método de Ponto Proximal	123
8	Conclusões	125
8.1	Conclusões e Futuras Pesquisas do Capítulo 3	125
8.2	Conclusões e Futuras Pesquisas do Capítulo 4	126
8.3	Conclusões e Futuras Pesquisas do Capítulo 5	126
8.4	Conclusões e Futuras Pesquisas do Capítulo 6	126
8.5	Conclusões e Futuras Pesquisas do Capítulo 7	127
	Referências Bibliográficas	129

Capítulo 1

Introdução

As aplicações da geometria Riemanniana à Otimização datam pelo menos do ano 1972, com o trabalho desenvolvido por Luenberger [51], onde, usando o método de descida geodésica obteve a taxa de convergência do método de gradiente projetado para o problema de $\min f(x)$, sujeito a $h(x) = 0$, sendo $f : \mathbb{R}^n \rightarrow \mathbb{R}$, e $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, funções que definem o problema, e $n > m$. Esta linha de pesquisa teve continuidade com Gabay [28] no ano de 1982 em que, do ponto de vista desta teoria, estuda o método de gradiente reduzido, generaliza os métodos de quase-Newton e obtém convergência superlinear. Também apresentou uma análise computacional mostrando que a teoria e a prática interrelacionadas podem dar bons resultados. Em seguida, os principais métodos da otimização não linear tais como o método de gradiente, subgradiente, proximal e Newton se generalizaram para problemas mais gerais definidos agora em espaços Riemannianos, ver por exemplo [1, 16, 21, 23, 25, 24, 55, 75, 79, 82]. Em 1995, Oliveira e Cruz Neto [60] mostraram que a maioria dos métodos primais clássicos são métodos de gradiente em alguma métrica Riemanniana apropriadamente escolhida. Em particular, mostraram que os métodos de Cauchy e Newton são métodos de gradiente sob este ponto de vista.

Uma motivação para desenvolver estas generalizações vem do fato de que considerando a geometria intrínseca da variedade Riemanniana os problemas com restrições podem ser vistos como irrestritos. Uma outra motivação é que certos problemas de otimização não convexos podem-se transformar em convexos mediante a introdução de uma adequada métrica Riemanniana sobre a variedade, por isso podemos usar técnicas de otimização mais eficientes, ver

Gabay [28], Cruz Neto et al. [16], Ferreira e Oliveira [23], Luenberger [51], Rapcsák [75], Smith [79], Udriste [82] e suas referências. Uma outra motivação para desenvolver esta abordagem é que podemos empregar as métricas Riemannianas para introduzir novos algoritmos em métodos de pontos interiores, ver por exemplo Saigal [77], Cunha et al. [19], den Hertog [40], Pereira e Oliveira [72].

Como acontece com os métodos de otimização em espaços Euclidianos, as extensões dos métodos em variedades Riemannianas para problemas de otimização convexa gozam de boas propriedades de convergência, isto é, pode ser provado que a seqüência de pontos gerada por cada método converge a uma solução ótima do problema, quando ela existe. Para o método de máxima descida veja Cruz Neto et al. [16] e [18], para o método de Subgradiente ver Ferreira e Oliveira [23], para o método de ponto proximal ver Ferreira e Oliveira [25] e para os métodos de Newton veja Adler et al. [1], Dedieu et al. [21] e Ferreira e Svaiter [24].

Atualmente, existe uma extensa classe de problemas de otimização que tem tomado muito interesse pelos pesquisadores por suas múltiplas aplicações em teoria económica [81], teoria de localização [36], teoria de controle [4] e sistemas dinâmicos [33]. Ela é a classe de problemas de otimização quando a função objetivo é quase-convexa. Surge então uma pergunta natural: podemos estender os resultados de convergência global do caso convexo para o caso quase-convexo?

Por outro lado, os métodos de pontos interiores, chamados também métodos de penalidade interior, são usados para resolver problemas de otimização que contém restrições de desigualdades. Estes métodos foram introduzidos nos anos cinquenta (ver Frisch [27]) e profundamente estudados durante os anos sessenta por vários pesquisadores, entre eles Fiacco e McCormick [26] no contexto da otimização não linear. São os seguintes os elementos essenciais do método: Para um dado problema geral de otimização

$$\begin{aligned}
 (P) \quad & \min f(x) \\
 & h(x) = 0 \\
 & g(x) \leq 0,
 \end{aligned}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ e $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ são funções que definem o problema, os

métodos de pontos interiores resolvem, para cada valor de $\mu > 0$, o problema auxiliar

$$(PA) \quad \min f(x) + \mu B(g(x)) \\ h(x) = 0$$

onde μ é um parâmetro que em cada iteração aproxima-se de zero e B é uma função barreira, isto é, B é uma função que fornece valores muito grandes quando o valor da função g se aproxima de zero.

Na década de sessenta, propriedades de convergência foram provadas para duas principais barreiras, a saber, a barreira logarítmica $-\sum_{i=1}^m \ln(g_i(x))$ e barreira inversa $\sum_{i=1}^m (1/g_i(x))$. Entretanto terminou-se descobrindo que estes métodos tinham sérias deficiências numéricas no limite, devido ao mau condicionamento da hessiana quando o valor do parâmetro μ fica bem perto de zero. Além disso, sua convergência à solução era lenta, comparada com outros métodos existentes, assim eles perderam interesse para a comunidade científica nos anos setenta. O interesse pelo método renasceu depois do trabalho de Karmarkar [45], quando foi descoberta uma relação muito estreita entre o algoritmo de Karmarkar e uma nova versão dos métodos de pontos interiores com a barreira logarítmica (ver por exemplo Gill. et al. [31] e Renegar [74]). Esta nova versão dos métodos deslocam os pontos iterados ao longo de uma faixa que garante teoricamente a complexidade polinomial e o bom desempenho computacional.

O algoritmo de Karmarkar foi um aporte de muito sucesso na linha de pesquisa da otimização linear, já que, pela primeira vez, obteve-se um algoritmo teoricamente polinomial e que na prática, para alguns problemas, era superior ao método simplex. Ele obteve $O(nL)$ iterações e $O(n^{3.5}L)$ operações binárias para resolver um problema de otimização linear, onde n é o número de variáveis e L representa a dimensão de representação dos dados do problema ingressados no computador em código binário. A partir daí gerou-se uma linha de pesquisa de métodos de pontos interiores e suas aplicações ramificaram-se a quase todas as áreas da ciência e da engenharia.

As propriedades de polinomialidade para os algoritmos de pontos interiores na otimização linear foram estendidas para uma ampla classe de problemas de otimização convexa por Nesterov e Nemirovskii [58]. Estes pesquisadores desenvolveram uma teoria geral para resolver este tipo de problemas por métodos de pontos interiores usando uma classe de barreiras chamadas barreiras auto-concordantes. Nesterov e Nemirovskii provaram que o método de Newton apli-

cado na resolução de cada problema auxiliar (PA), usando barreiras auto-concordantes, permite desenvolver algoritmos eficientes do ponto de vista teórico. Apesar da importância desta abordagem, a única barreira conhecida é a barreira logarítmica (as outras são obtidas pela composição desta com uma função apropriada). Torna-se, portanto, natural a questão sobre a possibilidade de obtenção de uma outra barreira auto-concordante que não seja obtida via composição da barreira logarítmica.

Um ponto de interseção entre os métodos de pontos interiores e a geometria Riemanniana é a análise das propriedades das trajetórias contínuas dos métodos de pontos interiores na otimização linear e otimização convexa; entre outros temos Bayer e Lagarias [5, 6], Megiddo [52], Megiddo e Shub [53], o mesmo Karmarkar [46] e, recentemente, Nesterov e Todd [59]. Karmarkar analisou por elementos de geometria Riemanniana a complexidade do seu algoritmo provando que o número de iterações está relacionado à curvatura da trajetória. Nesterov e Todd analisaram também através de ferramentas de geometria Riemanniana a relação entre as trajetórias contínuas dos algoritmos de pontos interiores primais-duais usando barreiras auto-concordantes e as trajetórias das geodésicas geradas por métricas dadas pela Hessiana destas barreiras. Eles provaram que estas trajetórias estão bem próximas, pelo que concluíram que algoritmos geodésicos ou próximos das geodésicas são presumivelmente eficientes.

Com respeito à otimização linear, recentemente muitos esforços foram feitos para provar a complexidade assintótica de $\mathcal{O}(\sqrt{n}L)$ iterações. Os resultados de Rapcsák e Thang [76] no ano de 1996, abriram uma interrogação: será possível obter uma complexidade polinomial igual ou melhor com alguma métrica Riemanniana?

Motivados pelas questões anteriores, desenvolvemos esta tese, dividida em 7 capítulos. No capítulo 2 apresentamos as preliminares, notações e aspectos básicos, que serão úteis no decorrer do trabalho. No capítulo 3 estudamos as propriedades geométricas de uma classe especial de métricas Riemannianas diagonais no ortante positivo \mathbb{R}_{++}^n e no hipercubo $(0, 1)^n$. Estas propriedades são importantes para garantir as condições de convergência de algoritmos não Euclidianos, por exemplo, algoritmos de máxima descida, Newton, subgradientes e ponto proximal. O principal aporte deste capítulo é a introdução de novos algoritmos explícitos para resolver problemas de otimização irrestrita com variáveis limitadas. Este capítulo gerou o artigo [65], publicado pela *SOBRAPO* no ano 2004. No capítulo 4 estudamos o método de

máxima descida com uma busca de Armijo generalizada e uma regularização proximal em variedades Riemannianas para resolver problemas de minimização quando a função objetivo é quase-convexa. Com a hipótese de que o conjunto de soluções ótimas do problema é não vazio, obtemos a convergência global a um ponto crítico do problema. Este capítulo, gerou o artigo [69], aceito para publicação pelo *Journal of Mathematical Analysis and Applications*. No capítulo 5 é generalizado o algoritmo de ponto proximal com distâncias de Bregman para resolver problemas de otimização para funções quase-convexas e convexas em variedades de Hadamard. Este capítulo gerou o artigo [68] aceito para publicação pelo *Journal of Convex Analysis*. No capítulo 6 apresentamos uma nova barreira auto-concordante para resolver problemas de otimização convexa com restrições lineares canalizadas, obtemos boa definição da trajetória central primal-dual e a convergência da trajetória central primal. No caso linear, provamos a convergência da trajetória primal-dual e apresentamos uma aplicação onde introduzimos um algoritmo polinomial seguidor da trajetória central primal, obtendo limites superiores para o número de iterações para encontrar uma ϵ -solução. Além disso introduzimos também um novo algoritmo de ponto proximal com uma distância de Bregman induzida pela nova barreira e provamos a sua convergência. Este capítulo gerou o artigo [66] aceito para publicação pelo *Journal of Optimization Theory and Applications JOTA*. No capítulo 7 generalizamos os resultados do capítulo anterior para resolver uma classe de problemas de otimização semidefinida. Estudamos a trajetória central primal-dual, e, analogamente ao capítulo anterior, no caso linear semidefinido, provamos a convergência da trajetória primal-dual como também a convergência do algoritmo de ponto proximal com uma distância de Bregman induzida pela barreira generalizada. Este capítulo gerou o artigo [67], publicado pelo *RAIRO Operations Research* no ano 2006.

Nota Bibliográfica

Parte desta introdução foi desenvolvida (mais extensamente) em [61], texto didático que foi ampliado em [64].

Capítulo 2

Preliminares

Neste capítulo apresentamos a simbologia utilizada e faremos um resumo dos resultados básicos necessários ao desenvolvimento dos capítulos seguintes.

2.1 Símbolos e Notações

Ao longo desta tese, adotaremos a seguinte simbologia:

$\mathbb{R}^n = \{x = (x_1, x_2, \dots, x_n) : x_i \in \mathbb{R}, \forall i = 1, \dots, n\}$: o espaço vetorial euclidiano.

$\mathbb{R}_+^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_i \geq 0 \forall i = 1, \dots, n\}$: o ortante não negativo.

$\mathbb{R}_{++}^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_i > 0\}$: o ortante positivo.

$\mathcal{S}^n = \{X \in \mathbb{R}^{n \times n} : X = X^T\}$: o espaço das matrizes simétricas.

$[0, 1]^n = [0, 1] \times [0, 1] \times \dots \times [0, 1]$ (n vezes): o hipercubo fechado.

$(0, 1)^n = (0, 1) \times (0, 1) \times \dots \times (0, 1)$ (n vezes): o hipercubo aberto.

Dado $X \in \mathbb{R}^{n \times n}$, $0 \prec X$ significa X definida positiva.

$0 \prec X \prec I$ significa $0 \prec X$ e $0 \prec I - X$.

$\mathcal{S}_{[0,I]}^n := \{X \in \mathcal{S}^n : 0 \preceq X \preceq I\}$.

$\mathcal{S}_{(0,I)}^n := \{X \in \mathcal{S}^n : 0 \prec X \prec I\}$.

Dado $x \in \mathbb{R}^n$, $x \geq 0$ significa $x_i \geq 0, \forall i = 1, \dots, n$.

Dado $x \in \mathbb{R}^n$, $x > 0$ significa $x_i > 0, \forall i = 1, \dots, n$.

Dados $x, y \in \mathbb{R}^n$, $x^T y = \sum_{i=1}^n x_i y_i$: produto interno euclidiano em \mathbb{R}^n .

Dado $x \in \mathbb{R}^n$, $\|x\| := \sqrt{\sum_{i=1}^n x_i^2}$: a norma euclidiana.

Dado $x \in \mathbb{R}^n$, $\|x\|_1 := \sum_{i=1}^n |x_i|$: a norma da soma ou norma 1.

Dado $x \in \mathbb{R}^n$, $\|x\|_\infty := \max_{i=1, \dots, n} \{|x_i|\}$: a norma do máximo.

Dado $x \in \mathbb{R}^n$, $\|x\|_G = \langle Gx, x \rangle^{\frac{1}{2}}$: a norma induzida pela matriz definida positiva G .

Dada $X \in \mathcal{S}^n$, $\text{tr}(X)$ denota o traço da matriz X .

Dada $X \in \mathcal{S}^n$, $X \bullet Y := \text{tr}(XY)$: produto interno no espaço das matrizes simétricas.

$\|X\|_F = \sqrt{X \bullet X}$: a norma de Frobenius.

Dado um aberto Ω em \mathbb{R}^n , $C^p(\Omega) = \{f : \Omega \rightarrow \mathbb{R} \text{ tal que } f \text{ é uma função diferenciável de ordem } p\}$.

$C^\infty(\Omega) = \{f : \Omega \rightarrow \mathbb{R} \text{ tal que } f \text{ é uma função infinitamente diferenciável}\}$.

$e = (1, 1, \dots, 1) \in \mathbb{R}^n$: vetor de todas as componentes 1.

Dado um subconjunto $M \subset \mathbb{R}^n$, $\text{ri}(M)$ denota o interior relativo de M .

$\text{front}[0, 1]^n$: a fronteira de $[0, 1]^n$.

$\text{front}\mathcal{S}_{(0,I)}^n$: a fronteira de $\mathcal{S}_{(0,I)}^n$.

$\text{range}(f)$: a imagem da função f .

$\text{ker}(f) := \{x \in \mathbb{R}^n, f(x) = 0\}$: o núcleo de f .

$\text{rank}(A)$: o posto de uma matriz $A \in \mathbb{R}^{m \times n}$.

$b_k \rightarrow x, k \rightarrow \infty: \lim_{k \rightarrow \infty} b_k = x$.

$f(b_k) \rightarrow f(x^*), k \rightarrow \infty: \lim_{k \rightarrow \infty} f(b_k) = f(x)$.

Dados $H \in \mathcal{S}^n$ e $X \in \mathcal{S}^n$ e invertível, então: $(X^{-1} \circ X^{-1})H = X^{-1}HX^{-1}$.

2.2 Elementos Básicos de Geometria Riemanniana

Nesta seção introduzimos algumas propriedades fundamentais e notações de geometria Riemanniana, que usaremos nos próximos capítulos. Estes fatos básicos podem ser encontrados, por exemplo, em Manfredo do Carmo [11] e nos textos didáticos [61] e [64].

Seja S uma variedade diferenciável. Denotamos por $T_x S$ o espaço tangente a S no ponto $x \in S$ e $TS = \bigcup_{x \in S} T_x S$. $T_x S$ é um espaço vetorial e assim, definimos $\dim S = \dim T_x S$. Devido a que nosso estudo é feito em variedades reais, $T_x S$ é isomorfo a \mathbb{R}^n . Se S é dotado com uma métrica Riemanniana g então S é uma variedade Riemanniana e a denotamos por

(S, g) . Observemos que g pode ser sempre representado por uma matriz invertível. O produto interno de dois vetores $u, v \in T_x S$ é $\langle u, v \rangle_x := g_x(u, v)$, onde g_x é uma métrica no ponto x . A norma de um vetor $v \in T_x S$ é definida por $\|v\|_x := \sqrt{\langle v, v \rangle_x}$. Um grupo de Lie é uma variedade diferenciável S com uma estrutura de grupo \bullet tal que a aplicação $\rho : S \times S \rightarrow S$ com $\rho(x, y) = x \bullet y^{-1}$ é diferenciável. Para um elemento $y \in S$ a translação pela esquerda por x é a aplicação $L_x : S \rightarrow S$ definida por $L_x(y) = x \bullet y$. Note que a translação pela esquerda L_x é um difeomorfismo diferenciável. Seja S um grupo de Lie, a métrica Riemanniana g sobre S é dita *invariante pela esquerda* se para cada $x \in S$ a translação pela esquerda L_x é uma isometria, isto é,

$$\langle u, v \rangle_y = \langle d(L_x)_y u, d(L_x)_y v \rangle_{L_x y}, \forall y \in S, \quad (2.1)$$

onde $d(L_x)_y$ é a aplicação diferencial de L_x no ponto y .

A métrica pode ser usada para definir o comprimento de uma curva regular por pedaços $\alpha : [a, b] \rightarrow S$ ligando p' e p por $L(\alpha) = \int_a^b \|\alpha'(t)\| dt$, onde $\alpha(a) = p'$ e $\alpha(b) = p$. Minimizando este funcional sobre o conjunto de todas estas curvas obtemos uma distância Riemanniana $d(p', p)$ que induz a topologia original sobre S .

Dados dois campos de vetores $X, Y : S \rightarrow TS$, ao longo de uma curva $\alpha : [a, b] \rightarrow S$, a derivada covariante de Y na direção X é $\nabla_X Y$. O teorema de Levi-Civita garante que, dada uma variedade Riemanniana (S, g) , existe uma única conexão ∇ , a qual define a derivada covariante D/dt , simétrica e compatível com a métrica (o que significa que:

$\frac{d}{dt} \langle X, Y \rangle = \langle \frac{DX}{dt}, Y \rangle + \langle X, \frac{DY}{dt} \rangle$). Esta é chamada uma conexão Riemanniana. Uma curva $\alpha(t)$ é uma geodésica partindo do ponto p com direção v , se $\alpha(0) = p, \alpha'(0) = v$ e

$$\frac{d^2 \alpha_k}{dt^2} + \sum_{i,j} \Gamma_{ij}^k \frac{d\alpha_i}{dt} \frac{d\alpha_j}{dt} = 0, \quad k = 1, \dots, n, \quad (2.2)$$

onde Γ_{ij}^k são os símbolos de Christoffel expressados por

$$\Gamma_{ij}^m = \frac{1}{2} \sum_{k=1}^n \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km}, \quad (2.3)$$

(g^{ij}) denota a matriz inversa da métrica $g = (g_{ij})$, onde x_i são as coordenadas de x . Agora, supondo que X e Y são representados por $X = \sum_{i=1}^n u^i X_i, Y = \sum_{i=1}^n v^i X_i$, para alguma base local $\{X_i\}$ de $T_x S$, então $\nabla_{X_i} X_k = \sum_{j=1}^n \Gamma_{ik}^j X_j$. Uma variedade Riemanniana é *completa* se suas

curvas geodésicas estão definidas para qualquer valor de $t \in \mathbb{R}$. Denotamos por R o tensor curvatura definido por

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X, Y]} Z, \quad (2.4)$$

onde X, Y, Z são campos de vetores de S e $[X, Y] := YX - XY$ é o colchete de Lie. É fácil ver que esta fórmula pode ser simplificada quando $[X_i, X_j] = 0$. Claramente, a curvatura em uma variedade Riemanniana é tri-linear. Definamos agora a curvatura seccional com respeito a X e Y por

$$K(X, Y) = \frac{\langle R(X, Y)Y, X \rangle}{\|X\|^2 \|Y\|^2 - \langle X, Y \rangle^2}.$$

Se $K(X, Y) = 0$, S é uma variedade Riemanniana de curvatura zero. Uma variedade Riemanniana simplesmente conexa com curvatura seccional não-positiva ($K(X, Y) \leq 0$) é chamada *variedade de Hadamard*

Teorema 2.2.1 *Seja M uma variedade de Hadamard. Então M é difeomorfa ao espaço Euclidiano \mathbb{R}^n , $n = \dim M$. Assim, para cada ponto $x \in M$, a aplicação exponencial $\exp_x : T_x M \rightarrow M$ é um difeomorfismo global.*

Demonstração. Ver Sakai [78], Teorema 4.1, página 221. ■

Uma consequência do teorema anterior é que em variedades de Hadamard temos a propriedade da unicidade das curvas geodésicas que ligam qualquer par de pontos. Outra útil propriedade é a seguinte: seja $[x, y, z]$ o triângulo geodésico, que consiste de *vértices* e geodésicas ligando estes pontos. Temos o seguinte resultado.

Teorema 2.2.2 *Dado um triângulo geodésico $[x, y, z]$ em uma variedade de Hadamard, então:*

$$d^2(x, z) + d^2(z, y) - 2\langle \exp_z^{-1} x, \exp_z^{-1} y \rangle_z \leq d^2(x, y) \quad (2.5)$$

Demonstração. Ver Sakai [78], Proposição 4.5, página 223. ■

Existe também um resultado interessante para variedades Riemannianas de curvatura seccional não negativa ($K(X, Y) \geq 0$) conhecido como a lei de cosenos.

Uma geodésica “hinge” em M é um par de geodésicas normalizadas γ_1 e γ_2 tais que $\gamma_1(0) = \gamma_2(0)$ e pelo menos uma delas, digamos γ_1 , é minimal.

Teorema 2.2.3 *Em uma variedade Riemanniana M completa, de dimensão finita e com curvatura seccional não negativa é satisfeita a seguinte propriedade:*

$$l_3^2 \leq l_1^2 + l_2^2 - 2l_1l_2 \cos \alpha.$$

onde l_1 e l_2 denotam os comprimentos das curvas γ_1 e γ_2 , respectivamente, que formam a geodésica “hinge” em M , $l_3 = d(\gamma_1(l_1), \gamma_2(l_2))$ e $\alpha = \angle(\gamma_1'(0), \gamma_2'(0))$ é o ângulo entre os vetores $\gamma_1'(0)$ e $\gamma_2'(0)$.

Demonstração. Ver [16], Teorema 2.1. ■

O gradiente de uma função diferenciável $f : S \rightarrow \mathbb{R}$, denotado por $\text{grad}f$, é um campo vetorial sobre S definido por $df(X) = \langle \text{grad}f, X \rangle = X(f)$, onde X é também um campo vetorial sobre S . Agora, se f é duas vezes continuamente diferenciável, podemos definir a Hessiana H^f , como a derivada covariante do campo vetorial gradiente, isto é, $H^f = D(\text{grad}f)/dt$. Assim, a Hessiana de f em um ponto $x \in S$ na direção $v \in T_xS$ é $H_x^f(v) = (D(\text{grad}f)/dt)(x) = \nabla_v(\text{grad}f)(x)$.

2.3 Teoremas da Função Inversa e da Função Implícita

Teorema 2.3.1 *(Teorema da função inversa). Seja $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ uma aplicação diferenciável e suponhamos que em $q \in U$ a diferencial $df_q : \mathbb{R}^n \rightarrow \mathbb{R}^n$ seja um isomorfismo, então existe uma vizinhança $V \subset U$ de q e uma vizinhança $W \subset \mathbb{R}^n$ de $f(q)$ tais que $f : V \rightarrow W$ tem uma inversa $f^{-1} : W \rightarrow V$ diferenciável, isto é, f é um difeomorfismo de V em W .*

Demonstração. Ver [49], página 283. ■

Teorema 2.3.2 *(Teorema da função implícita). Suponhamos dado um conjunto de m equações e n variáveis, $m < n$:*

$$h_i(x) = 0, \forall i = 1, 2, \dots, m.$$

Seja $x^0 = (x_1^0, x_2^0, \dots, x_m^0, x_{m+1}^0, x_{m+2}^0, \dots, x_n^0) \in \mathbb{R}^n$ satisfazendo as propriedades:

i) Existe uma vizinhança V_0 de x^0 e um $p \geq 1$ tais que as funções $h_i \in C^p(V_0)$

ii) $h_i(x^0) = 0, \forall i = 1, \dots, m$

iii) A matriz Jacobiana de ordem m :

$$J_h(x^0) = \begin{bmatrix} \frac{\partial h_1}{\partial x_1}(x^0) & \frac{\partial h_1}{\partial x_2}(x^0) & \dots & \frac{\partial h_1}{\partial x_m}(x^0) \\ \frac{\partial h_2}{\partial x_1}(x^0) & \frac{\partial h_2}{\partial x_2}(x^0) & \dots & \frac{\partial h_2}{\partial x_m}(x^0) \\ \dots & \dots & \dots & \dots \\ \frac{\partial h_m}{\partial x_1}(x^0) & \frac{\partial h_m}{\partial x_2}(x^0) & \dots & \frac{\partial h_m}{\partial x_m}(x^0) \end{bmatrix}$$

é não singular. Então existe uma vizinhança \mathcal{U}^0 de $(x_{m+1}^0, x_{m+2}^0, \dots, x_n^0) \in \mathbb{R}^{n-m}$ tal que para cada $\hat{x} = (x_{m+1}, x_{m+2}, \dots, x_n) \in \mathcal{U}^0$ existem funções $\phi_i : \mathcal{U}_{\hat{x}} \subset \mathbb{R}^{n-m} \rightarrow \mathbb{R}^m$, onde $\mathcal{U}_{\hat{x}}$ é uma vizinhança de \hat{x} tal que:

a) $\phi_i \in C^p(U_{\hat{x}})$

b) $x_i^0 = \phi_i(x_{m+1}^0, x_{m+2}^0, \dots, x_n^0), \forall i = 1, 2, \dots, m$

c) $h_i(\phi_1(\hat{x}), \phi_2(\hat{x}), \dots, \phi_m(\hat{x}), \hat{x}) = 0, \forall i = 1, 2, \dots, m$

Demonstração. Ver [49], página 164. ■

2.4 Funções de matrizes

No capítulo 7, o problema de interesse será minimizar uma função onde a variável é uma matriz simétrica semidefinida positiva. Por isso, daremos uma pequena introdução a esta teoria. Estes fatos básicos podem ser encontrados em [32] e [41].

Toda matriz $X \in \mathcal{S}^n$ pode ser expressa na seguinte forma:

$$X = Q^T D Q$$

onde $D = \text{diag}(\lambda_1(X), \dots, \lambda_n(X))$, $\{\lambda_i(X)\}_{i=1, \dots, n}$ são os autovalores de X e Q é a correspondente matriz ortogonal dos autovetores de X . Daqui, para qualquer função real analítica g , podemos definir uma função em $X \in \mathcal{S}^n$ como sendo:

$$g(X) = Q^T g(D) Q$$

sempre que as funções $g(\lambda_i(X))$ sejam bem definidas (ver Horn e Johnson [41], Seção 6.2).

Em particular, obtemos que para qualquer $X \in \mathcal{S}_{++}^n$:

$$\ln X = Q^T D_{\ln} Q$$

onde $D_{\ln} = \text{diag}(\ln \lambda_1(X), \dots, \ln \lambda_n(X))$. Esta matriz é a inversa da matriz exponencial \exp . Assim, para qualquer $B \in \mathcal{S}_{++}^n$, $\log B = A$ se, e somente se, $B = \exp(A)$.

2.5 Conjuntos Semi-Analíticos

Nesta seção introduzimos a definição de curva (s)-analítica e o lema da curva selecionada. Estes fatos serão usados na prova de convergência da trajetória central primal-dual, definida por uma nova barreira, nos capítulos 6 e 7. Para maiores detalhes referenciamos a Lojasiewicz [50].

Definição 2.5.1 *Um subconjunto M de \mathbb{R}^n é um conjunto semi-analítico se M é determinado por um número finito de igualdades e desigualdades de funções analíticas. Isto é, se M pode ser escrito como*

$$\bigcup_{j=1}^k \bigcap_{i=1}^l \{x : f_{ij}(x) \sigma_{ij} \leq 0\},$$

onde $k, l \in \mathbb{N}$, f_{ij} são funções analíticas e σ_{ij} corresponde a um dos sinais $\{<\}, \{>\}$ ou $\{=\}$. Se M tem somente igualdades, ($\sigma_{ij} = 0$, para todo i, j) então M é chamado conjunto analítico.

Definição 2.5.2 *Seja M um conjunto analítico de \mathbb{R}^n . Uma curva $\gamma \subset M$ é uma curva (s)-analítica quando é a imagem de um mergulho analítico de $(0, 1]$ sobre um conjunto semi-analítico relativamente compacto de M .*

Lemma 2.5.1 *(Lema da curva selecionada) Se M é um conjunto semi-analítico de \mathbb{R}^n e $a \in \overline{M}$ (o fecho de M) não é um ponto isolado em M , então M contém uma curva (s)-analítica que converge para o ponto a . Equivalentemente, existe $\epsilon > 0$ e uma curva analítica $\gamma : [0, \epsilon) \rightarrow \mathbb{R}^n$ com $\gamma(0) = a$ e $\gamma(t) \in M$ para $t > 0$.*

Demonstração. Ver [50], Proposição 2, página 103. ■

Capítulo 3

Métricas Riemannianas Diagonais no Ortante Positivo e no Hipercubo

3.1 Introdução

A perspectiva das métricas na otimização contínua é evidente em muitos algoritmos e na sua análise de convergência, veja Nazareth [55] para minimização irrestrita, Karmarkar [46], Rapsáck e Thang [76] e Oliveira e Cruz Neto [60] para métodos de pontos interiores. Em um contexto mais geral temos Gabay [28], Rapcsák [75], Smith [79] e Udriste [82].

Em otimização linear Karmarkar [46], usando ferramentas de geometria Riemanniana, provou que a complexidade do seu algoritmo está relacionada à curvatura da trajetória contínua. Em otimização convexa, temos um interessante resultado de Nesterov e Todd [59] sobre a geometria Riemanniana definida por Hessianas de barreiras auto-concordantes: a trajetória contínua descrita pelos algoritmos primais-duais está bem próxima à de suas trajetórias geodésicas.

Também a classe de métricas $(\mathbb{R}_{++}^n, X^{-r})$, para $r \geq 1$, foi utilizada para criar novas famílias de algoritmos de pontos interiores, ver Saigal [77] para um algoritmo primal aplicado à otimização linear, Cunha et al. [19] com um método primal e dual affim escala para problemas lineares convexos, den Hertog [39] para um método de barreira para otimização linear, Oliveira e Oliveira [62] com um algoritmo proximal para problemas convexos diferenciáveis no ortante positivo, Souza e Oliveira [80] também no ortante positivo, para funções convexas não suaves e Pereira e Oliveira [72] para o problema de complementaridade não linear.

Neste capítulo, avançamos nestas idéias, em relação a dois espaços naturais em problemas de otimização: o ortante positivo e o hipercubo. Primeiro, observemos que do ponto de vista das aplicações da geometria Riemanniana, são requisitos necessários conhecer propriedades de curvatura, ter curvas geodésicas e distâncias Riemannianas explícitas, como também ter a propriedade de completude da variedade Riemanniana, isto é, que exista uma geodésica minimal ligando qualquer par de pontos, o que nos permite medir distâncias de uma maneira apropriada. Nosso objetivo é estudar que tipo de métricas diagonais cumprem estes requisitos. Segundo, observando a relação natural entre métricas Riemannianas e funções auto-concordantes introduzimos, mediante escolhas particulares da métrica diagonal, duas funções auto-concordantes. Como veremos depois, uma destas funções será de grande importância para desenvolver um novo algoritmo em métodos de pontos interiores com propriedades polinomiais.

O capítulo é organizado como segue. Na Seção 3.2 apresentamos uma generalização do método de obtenção de métricas Riemannianas invariantes por translações e apresentamos algumas métricas diagonais definidas sobre o ortante positivo \mathbb{R}_{++}^n , o hipercubo $(0, 1)^n$ e o produto de ortantes positivos $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$, considerados como variedades Riemannianas. Na Seção 3.3 estudamos as propriedades geométricas de uma especial classe de métricas diagonais, obtendo simples equações para obter curvas geodésicas, expressões explícitas do transporte paralelo, curvatura zero da variedade Riemanniana, condições suficientes para garantir completude, expressões explícitas para o gradiente e a Hessiana de uma função suficientemente diferenciável. Na Seção 3.4 fornecemos exemplos de curvas geodésicas para algumas métricas das variedades Riemannianas \mathbb{R}_{++}^n e $(0, 1)^n$, duas delas, a nosso entender, são originais. Na Seção 3.5 apresentamos algumas aplicações. Primeiro, introduzindo as métricas $\pi^2 \csc^4(\pi x) = \text{diag}(\pi^2 \csc^4(\pi x_1), \pi^2 \csc^4(\pi x_2), \dots, \pi^2 \csc^4(\pi x_n))$ e $X^{-2}(I - X)^{-2}$ na variedade diferenciável $(0, 1)^n$, apresentamos novos exemplos de algoritmos explícitos para resolver problemas de otimização irrestritas com variáveis limitadas. Estes algoritmos são, em particular, importantes, quando as funções objetivos não são convexas com respeito à métrica Euclidiana mas tornam-se convexas com respeito à métrica introduzida. Segundo, obtemos que as funções que geram as métricas anteriores são:

$$-\frac{1}{6} \sum_{i=1}^n \{4 \ln(\sin \pi x_i) - \cot^2(\pi x_i)\} \quad e$$

$$\sum_{i=1}^n (2x_i - 1) [\ln x_i - \ln(1 - x_i)]$$

e provamos que elas são funções barreiras para $[0, 1]^n$.

3.2 Métricas Riemannianas Invariantes

Karmarkar [46] utilizou a propriedade de invariância da métrica Riemanniana sob translações associada ao seu algoritmo projetivo, com o objetivo de estudar sua respectiva trajetória contínua. Esta propriedade, que essencialmente significa independência por mudança de coordenadas, é também útil para construir métricas em métodos de pontos interiores para otimização linear. Nesta seção generalizamos o método clássico para obter uma métrica Riemanniana invariante, e chamamos esta generalização de métrica Riemanniana invariante por H - translação.

Construção de métricas invariantes por H -translação

Seja S uma variedade diferenciável, $x \in S$ e $H : S \rightarrow S$ uma aplicação diferenciável. Para construir uma métrica invariante por H -translação considere os seguintes passos:

1. Definir em S uma estrutura de grupo de Lie.
2. Dado $x \in S$, considerar o elemento $L_{H(x)^{-1}}x$ e definir um produto interno

$$\langle \cdot, \cdot \rangle_{L_{H(x)^{-1}}x} : T_{L_{H(x)^{-1}}x}S \times T_{L_{H(x)^{-1}}x}S \rightarrow \mathbb{R}$$

isto é, definir uma aplicação bilinear simétrica e definida positiva.

3. $\forall u, v \in T_xS$ definir a métrica

$$\langle u, v \rangle_x := \langle d(L_{H(x)^{-1}})_x u, d(L_{H(x)^{-1}})_x v \rangle_{L_{H(x)^{-1}}x}.$$

A seguir damos algumas aplicações deste método para gerar métricas Riemannianas diagonais associadas com alguns conjuntos viáveis naturais em problemas de otimização.

Exemplo 3.1 *Métricas diagonais no ortante positivo.*

Consideremos a variedade diferenciável \mathbb{R}_{++}^n e seja a função $H : \mathbb{R}_{++}^n \rightarrow \mathbb{R}_{++}^n$ tal que $H(x) = (h_1(x_1), h_2(x_2), \dots, h_n(x_n))$ para funções $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ diferenciáveis, onde $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}_{++}^n$.

Passo 1. Definiremos uma estrutura de grupo abeliano de Lie.

Seja $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}_{++}^n$, definindo:

$$x \bullet y = (x_1 \cdot y_1, x_2 \cdot y_2, \dots, x_n \cdot y_n)$$

obtemos uma estrutura de grupo abeliano de Lie com elemento neutro $\bar{e} = (1, 1, 1, \dots, 1)$ e com inversa para $x = (x_1, x_2, \dots, x_n)$ dada por $x^{-1} = (\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n})$.

Passo 2. Com $x \in \mathbb{R}_{++}^n$ definiremos o produto interno no ponto $L_{H(x)^{-1}}x$. Sendo $x = (x_1, x_2, \dots, x_n)$, obtemos que $L_{H(x)^{-1}}x = H(x)^{-1} \bullet x = (\frac{x_1}{h_1(x)}, \frac{x_2}{h_2(x)}, \dots, \frac{x_n}{h_n(x)})$ e como $T_{H(x)^{-1} \bullet x} \mathbb{R}_{++}^n = \mathbb{R}^n$ podemos definir o produto interno neste ponto como o produto interno euclidiano:

$$\langle v, w \rangle_{L_{H(x)^{-1}}x} = (v, w) = v^T w.$$

Passo 3. Primeiramente achemos $d(L_{H(x)^{-1}})$. A aplicação $L_{H(x)} : \mathbb{R}_{++}^n \rightarrow \mathbb{R}_{++}^n$ é tal que: $L_{H(x)}y = H(x) \bullet y = (h_1(x_1)y_1, h_2(x_2)y_2, \dots, h_n(x_n)y_n)$. Então $d(L_{H(x)})_y : \mathbb{R}^n \rightarrow \mathbb{R}^n$ fica representada pela seguinte matriz

$$d(L_{H(x)})_y = \text{diag}(h_1(x_1), h_2(x_2), \dots, h_n(x_n)),$$

logo

$$d(L_{H^{-1}(x)})_y = \text{diag}\left(\frac{1}{h_1(x_1)}, \frac{1}{h_2(x_2)}, \dots, \frac{1}{h_n(x_n)}\right).$$

Agora, $\forall u, v \in T_x \mathbb{R}_{++}^n = \mathbb{R}^n$, definimos:

$$\langle u, v \rangle_x = (d(L_{H^{-1}(x)})_x u, d(L_{H^{-1}(x)})_x v) = \sum_{i=1}^n \frac{u_i \cdot v_i}{h_i(x_i)^2} = u^T G(x) v, \text{ onde}$$

$$G(x) = \text{diag}\left(\frac{1}{h_1^2(x_1)}, \frac{1}{h_2^2(x_2)}, \dots, \frac{1}{h_n^2(x_n)}\right).$$

Assim

$$g_{ij}(x) = \langle e_i, e_j \rangle_x = \frac{\delta_{ij}}{h_i(x_i)h_j(x_j)}.$$

Em particular:

- Se $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}, h_i(x_i) = x_i^{\frac{r}{2}}$, onde $r \in \mathbb{R}$, então obtemos a métrica diagonal primal generalizada:

$$G(x) = \text{diag} \left(\frac{1}{x_1^r}, \frac{1}{x_2^r}, \dots, \frac{1}{x_n^r} \right) = X^{-r},$$

onde $X = \text{diag}(x_1, x_2, \dots, x_n)$. Esta métrica foi usada para obter novas classes de métodos na otimização contínua, ver por exemplo [19], [62], [72], [77], [80]. Observe-mos que quando $r = 2$, obtemos a métrica de Dikin X^{-2} .

- Se $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}, h_i(x_i) = s_i^{-\frac{r}{2}} x_i^{\frac{r}{2}}$, onde $s_i \in \mathbb{R}_{++}$ é um ponto fixado e $r \in \mathbb{R}$, então:

$$G(x) = \text{diag} \left(\frac{s_1^r}{x_1^r}, \frac{s_2^r}{x_2^r}, \dots, \frac{s_n^r}{x_n^r} \right) = S^r X^{-r}$$

(métrica primal-dual generalizada).

- Se $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}, h_i(s_i) = s_i^{-\frac{r}{2}}$, onde $r \in \mathbb{R}$, então,

$$G(x) = \text{diag} \left(\frac{1}{s_1^r}, \frac{1}{s_2^r}, \dots, \frac{1}{s_n^r} \right) = S^{-r}$$

(métrica dual generalizada).

Exemplo 3.2 Métricas diagonais no hipercubo.

Considerando $(0, 1)^n$ como uma variedade Riemanniana podemos definir a métrica induzida de \mathbb{R}_{++}^n do exemplo 3.1, isto é, $\forall u, v \in T_x((0, 1)^n) = \mathbb{R}^n$:

$$\langle u, v \rangle_x = \sum_{i=1}^n \frac{u_i \cdot v_i}{h_i(x_i)^2} = u^T G(x) v,$$

onde $h_i : (0, 1) \rightarrow \mathbb{R}_{++}$ são funções diferenciáveis.

Em particular:

- Se $h_i(x_i) = \sin(\pi x_i)$, com $x_i \in (0, 1)$, então teríamos a métrica:

$$\csc^2(\pi x) = \text{diag}(\csc^2(\pi x_1), \csc^2(\pi x_2), \dots, \csc^2(\pi x_n)).$$

Esta métrica foi introduzida por Nesterov e Todd [59].

- Se $h_i(x_i) = \left(\frac{x_i^r(1-x_i)^r}{x_i^r+(1-x_i)^r} \right)^{\frac{1}{2}} \forall x_i \in (0, 1), i = 1, \dots, n$ e $r \in \mathbb{R}$ então:

$$G(x) = X^{-r} + (I - X)^{-r}.$$

- Se $h_i(x_i) = \sin^2(\pi x_i)/\pi$, com $x_i \in (0, 1)$ então temos a métrica:

$$\pi^2 \csc^4(\pi x) = \text{diag}(\pi^2 \csc^4(\pi x_1), \pi^2 \csc^4(\pi x_2), \dots, \pi^2 \csc^4(\pi x_n)).$$

- Se $h_i(x_i) = x_i^{\frac{r}{2}}(1-x_i)^{\frac{r}{2}}$ com $x_i \in (0, 1)$ e $r \in \mathbb{R}$ então obtemos a métrica:

$$X^{-r}(I - X)^{-r}.$$

Exemplo 3.3 Métricas diagonais no produto de ortantes positivos.

Sejam S_1^n e S_2^m duas variedades Riemannianas com métricas $\langle \cdot, \cdot \rangle_1$ e $\langle \cdot, \cdot \rangle_2$ respectivamente, e consideremos a variedade produto $S_1 \times S_2$ de dimensão $n + m$. Sejam também as projeções naturais: $\pi^1 : S_1 \times S_2 \rightarrow S_1$ e $\pi^2 : S_1 \times S_2 \rightarrow S_2$ definidos por $\pi^1(p, q) = p$ e $\pi^2(p, q) = q$, onde $(p, q) \in S_1 \times S_2$. Sabemos que o espaço vetorial tangente $T_{(p,q)}(S_1 \times S_2)$ é de dimensão $n + m$, assim o vetor $v \in T_{(p,q)}(S_1 \times S_2)$ tem $n + m$ componentes. As diferenciais destas aplicações, $P = d\pi_{(p,q)}^1$ e $Q = d\pi_{(p,q)}^2$ permitem definir uma métrica na variedade $S_1 \times S_2$. De fato, dados $u, v \in T_{(p,q)}(S_1 \times S_2)$ arbitrários, definindo:

$$\langle u, v \rangle_{(p,q)} = \langle P(u), P(v) \rangle_p + \langle Q(u), Q(v) \rangle_q,$$

temos que $\langle \cdot, \cdot \rangle_{(p,q)}$ é um produto interno em $T_{(p,q)}(S_1 \times S_2)$.

Agora consideremos as variedades Riemannianas $(\mathbb{R}_{++}^n, G(x)), (\mathbb{R}_{++}^m, G(s))$, invariantes por H -translação e T -translação respectivamente, onde $H(x) = (h_1(x_1), h_2(x_2), \dots, h_n(x_n))$ e $T(s) = (T_1(s_1), T_2(s_2), \dots, T_m(s_m))$ para funções T_i e $H_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$, com métricas expressas por

$$G(x) = \text{diag} \left(\frac{1}{h_1(x_1)^2}, \frac{1}{h_2(x_2)^2}, \dots, \frac{1}{h_n(x_n)^2} \right)$$

e

$$G(s) = \text{diag} \left(\frac{1}{T_1(s_1)^2}, \frac{1}{T_2(s_2)^2}, \dots, \frac{1}{T_m(s_m)^2} \right).$$

Consideremos as projeções

$$\pi^1 : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}_{++}^n, \pi^1(x, s) = x,$$

$$\pi^2 : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}_{++}^n, \pi^2(x, s) = s.$$

Usando o fato de que $T_{(x,s)}(\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n) = T_x \mathbb{R}_{++}^n \times T_s \mathbb{R}_{++}^n = \mathbb{R}^n \times \mathbb{R}^n$, as diferenciais das projeções no ponto $(x, s) \in \mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$ são, respectivamente,

$$d\pi_{(x,s)}^1 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n, d\pi_{(x,s)}^1(u_x, u_s) = [I_{n \times n} \ \theta_{n \times n}] \begin{bmatrix} u_x \\ u_s \end{bmatrix} = u_x,$$

$$d\pi_{(x,s)}^2 : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n, d\pi_{(x,s)}^2(u_x, u_s) = [\theta_{n \times n} \ I_{n \times n}] \begin{bmatrix} u_x \\ u_s \end{bmatrix} = u_s,$$

onde $u = (u_x, u_s)$, com $u_x \in \mathbb{R}^n$ e $u_s \in \mathbb{R}^n$.

Definindo para $u = (u_x, u_s), v = (v_x, v_s) \in T_{(x,s)}(\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n) = \mathbb{R}^{2n}$

$$\langle u, v \rangle_{(x,s)} = \langle u_x, v_x \rangle_x + \langle u_s, v_s \rangle_s$$

obtemos

$$\langle u, v \rangle_{(x,s)} = u^T G_1(x)v + v^T G_2(s)v.$$

Assim, este produto interno define uma métrica invariante por $H \times T$ -translação na variedade produto $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$. Tomando a base canônica $\left\{ \frac{\partial}{\partial x_i} \right\}_{i=1, \dots, n}$ em $\mathbb{R}^n \times \mathbb{R}^n$, temos que a matriz que representa a métrica da variedade produto $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$ é:

$$G(x, s) = \begin{bmatrix} G_1(x)_{n \times n} & 0_{n \times n} \\ 0_{n \times n} & G_2(s)_{n \times n} \end{bmatrix}_{(2n \times 2n)}$$

Em particular, se $G_1(x) = X^{-r}$ e $G_2(s) = S^{-r}$, obtemos a métrica produto invariante

$$G(x, s) = \begin{bmatrix} X_{n \times n}^{-r} & 0_{n \times n} \\ 0_{n \times n} & S_{n \times n}^{-r} \end{bmatrix}_{(2n \times 2n)}.$$

Observação 3.2.1 *A maneira como construímos métricas invariantes por H -translação, inclui um grande número de possibilidades para gerar métricas. Em particular, esta inclui a classe de métricas que são matrizes Hessianas diagonais definidas por $h_i(x_i) = \left(\frac{1}{p_i''(x_i)}\right)^{\frac{1}{2}}$ onde $p_i : \mathbb{R}_{++} \rightarrow \mathbb{R}$ são funções reais tais que $p_i''(x_i) > 0$. A motivação para definir este tipo de funções é justamente dada pelas funções barreira na otimização contínua que se apresentam na forma*

$$p(x) = \sum_{i=1}^n p_i(x_i).$$

(Para maiores detalhes ver [60], página 15).

Exemplo 3.4 *Métrica projetiva generalizada no simplex unitário.*

Consideremos a variedade $\Delta^n = \{x \in \mathbb{R}_{++}^n : e^T x = 1\}$, $e := (1, 1, \dots, 1) \in \mathbb{R}^n$, e seja $H : \Delta^n \rightarrow \Delta^n$ uma função tal que: $H(x) = (h_1(x_1), h_2(x_2), \dots, h_n(x_n))$, onde $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ são funções diferenciáveis satisfazendo $\sum_{i=1}^n h_i(x_i) = 1$.

Passo 1. Definimos uma estrutura de grupo de Lie.

Sejam $x = (x_1, x_2, \dots, x_n)$, $y = (y_1, y_2, \dots, y_n) \in \Delta^n$, definimos $x \bullet y = \frac{1}{x^T y} (x_1 \cdot y_1, x_2 \cdot y_2, \dots, x_n \cdot y_n)$.

Esta operação define uma estrutura de grupo com elemento identidade $\bar{e} = (1/n, 1/n, 1/n, \dots, 1/n)$

e inversa de $x = (x_1, x_2, \dots, x_n)$ dada por $x^{-1} = \frac{1}{\sum_{i=1}^n (1/x_i)} \left(\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n}\right)$.

Passo 2. Definimos um produto interno no ponto $L_{H(x)^{-1}x}$.

Seja $x = (x_1, x_2, \dots, x_n) \in \Delta^n$, então

$$L_{H(x)^{-1}x} = H(x)^{-1} \bullet x = \frac{1}{\sum_{i=1}^n (x_i/h_i(x_i))} \left(\frac{x_1}{h_1(x)}, \frac{x_2}{h_2(x)}, \dots, \frac{x_n}{h_n(x)}\right).$$

Devido a que $T_{H(x)^{-1} \bullet x} \Delta^n = \{e\}^\perp$, podemos definir uma base e um produto interno em $T_{H(x)^{-1} \bullet x} \Delta^n$ como sendo $\{\alpha_i := u_i - H(x)^{-1} \bullet x\}$ e $\langle v, w \rangle_{L_{H(x)^{-1}x}} = (v, w) = v^T w$ respectivamente, onde u_i é o vetor canônico em \mathbb{R}^n .

Step 3. Definimos a métrica para todo $x \in \Delta^n$.

Considerando a base $\{v_1, \dots, v_{n-1}\}$ em $T_x \Delta^n$, onde $v_i = u_i - x$, obtemos

$$d(L_{H(x)^{-1}x})_x v_i = \frac{\alpha_i}{h_i(x_i) \sum_{i=1}^n (x_i/h_i(x_i))}.$$

Agora, $\forall u, v \in T_x(\Delta^n) = \{e\}^\perp$ temos:

$$g_{ij}(x) = \langle v_i, v_j \rangle_x =$$

$$\frac{\left(\sum_{i=1}^n \frac{x_i}{h_i(x_i)}\right)^{-2}}{h_i(x_i)h_j(x_j)} \left[\delta_{ij} - \frac{1}{\sum_{i=1}^n x_i/h(x_i)} \left(\frac{x_j}{h_j(x_j)} + \frac{x_i}{h_j(x_i)} \right) + \frac{1}{\left(\sum_{i=1}^n x_i/h(x_i)\right)^2} \sum_{i=1}^n x_i^2/h(x_i)^2 \right].$$

Em particular, quando $h_i(x_i) = x_i$, temos a métrica canônica de Karmarkar:

$$g_{ij}(x) = \langle v_i, v_j \rangle_x = \frac{1}{n^2 x_i x_j} [\delta_{i,j} - 1/n].$$

Na próxima seção estaremos interessados em estudar uma classe especial de métricas Riemannianas diagonais definidas no ortante positivo e no hipercubo.

3.3 Propriedades Geométricas de uma Classe de Métrica Diagonal

Nesta seção, derivamos propriedades geométricas da variedade Riemanniana $(\mathbb{R}_{++}^n, G(x))$ onde:

$$G(x) = \text{diag} \left(\frac{1}{h_1(x_1)^2}, \frac{1}{h_2(x_2)^2}, \dots, \frac{1}{h_n(x_n)^2} \right).$$

Neste sentido, obtemos expressões simples dos símbolos de Christoffel e assim a equação para obter curvas geodésicas explícitas é simplificada a resolver uma integral de uma certa função que depende das funções h_i . Provamos a curvatura zero da variedade, apresentamos condições suficientes para garantir que a variedade seja completa, fornecemos expressões explícitas para o transporte paralelo, para o gradiente e Hessiana de uma função de classe C^2 definida na variedade. Devido a que a métrica no hipercubo $(0, 1)^n$ é induzida pela métrica de \mathbb{R}_{++}^n , os resultados podem ser estendidos para o hipercubo. Finalmente, usando resultados de geometria Riemanniana, nossos resultados também podem ser estendidos para $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$. Para finalizar a seção, damos alguns exemplos de geodésicas e distâncias Riemannianas explícitas introduzindo algumas métricas particulares.

1. Símbolos de Christoffel.

Vimos que a relação da métrica com os símbolos de Christoffel é dada pela equação

(2.3). Agora, com respeito à métrica diagonal, quando $k \neq m$ tem-se que $g^{mk} = 0$, assim a expressão é reduzida a:

$$\Gamma_{ij}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{mi} - \frac{\partial}{\partial x_m} g_{ij} \right\} g^{mm}.$$

Primeiro caso: $i = j$

$$\Gamma_{ii}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{im} + \frac{\partial}{\partial x_i} g_{mi} - \frac{\partial}{\partial x_m} g_{ii} \right\} g^{mm}.$$

Se $m = i$

$$\Gamma_{ii}^m = -\frac{1}{h_i(x_i)} \frac{\partial}{\partial x_i} (h_i(x_i)),$$

se $m \neq i$

$$\Gamma_{ii}^m = 0.$$

Segundo caso: $i \neq j$

$$\Gamma_{ij}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{mi} \right\} g^{mm}.$$

Se $m = i$ então $m \neq j$ e:

$$\Gamma_{ij}^i = 0.$$

Se $m = j$ então $m \neq i$ e:

$$\Gamma_{ij}^j = 0.$$

Se $m \neq i$ e $m \neq j$ então :

$$\Gamma_{ij}^m = 0.$$

Em conclusão:

$$\Gamma_{ij}^m = -\frac{1}{h_i(x_i)} \frac{\partial (h_i(x_i))}{\partial x_i} \delta_{im} \delta_{ij}. \quad (3.1)$$

2. Expressão da derivada covariante e transporte paralelo.

Seja dado um campo vetorial $V = \sum_{i=1}^n v_i X_i$ sobre a curva $x(t) = (x_1(t), x_2(t), \dots, x_n(t)) \in \mathbb{R}_{++}^n$, onde $\{X_i\}_{i=1, \dots, n}$ é uma base para $T_p(\mathbb{R}_{++}^n) = \mathbb{R}^n$. É bem conhecido que

$$\frac{DV}{dt} = \sum_{i=1}^n \left(\frac{dv_i}{dt} + \sum_{j=1}^n v_j \frac{dx_j}{dt} \Gamma_{ij}^k \right) X_k \quad (3.2)$$

Substituindo os símbolos de Christoffel (3.1) em (3.2) temos:

$$\frac{DV}{dt} = \sum_{i=1}^n \left(\frac{dv_i}{dt} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} v_i \frac{dx_i}{dt} \right) X_i.$$

Agora, para $V_0 = (V_1^0, V_2^0, \dots, V_n^0) \in T_p(\mathbb{R}_{++}^n)$, a equação diferencial do transporte paralelo $V(t) = (V_1(t), \dots, V_n(t))$ ao longo da curva $x(t)$ é:

$$\frac{dV_i}{dt} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} V_i \frac{dx_i}{dt} = 0, \quad \forall i = 1, \dots, n,$$

com a condição

$$V(0) = V_0.$$

É fácil checar que esta equação é resolvida por:

$$V(t) = V_i^0 \frac{h_i(x_i(t))}{h_i(p_i)}, \quad \forall i = 1, 2, \dots, n.$$

Portanto, podemos definir, em forma explícita, o transporte paralelo ao longo da curva $x(t)$ como sendo a aplicação $P_{x(t)} : T_p(\mathbb{R}_{++}^n) \rightarrow T_{x(t)}\mathbb{R}_{++}^n$ tal que

$$P_{x(t)}(V) = \left(v_1 \frac{h_1(x_1(t))}{h_1(p_1)}, v_2 \frac{h_2(x_2(t))}{h_2(p_2)}, \dots, v_n \frac{h_n(x_n(t))}{h_n(p_n)} \right).$$

Em particular:

Se $h_i(x_i) = x_i^{r/2}$ ($G(x) = X^{-r}$, $r \in \mathbb{R}$ não nulo) então

$$P_{x(t)}(V)_i = \frac{v_i x_i(t)^{r/2}}{p_i^{r/2}}.$$

Se $h_i(x_i) = x_i^{r/2}(1-x_i)^{r/2}$ ($G(x) = X^{-r}(I-X)^{-r}$, $r \in \mathbb{R}$ não nulo) com $x_i \in (0, 1)$ então

$$P_{x(t)}(V)_i = \frac{v_i x_i^{r/2}(1-x_i(t))^{r/2}}{p_i^{r/2}(1-p_i)^{r/2}}. \quad (3.3)$$

Se $h_i(x_i) = \sin^2(\pi x_i)/\pi$, para $x_i \in (0, 1)$ ($G(x) = \pi^2 \csc^4(\pi x)$) temos

$$P_{x(t)}(V)_i = \frac{v_i \sin^2(\pi x_i(t))}{\sin^2(\pi p_i)}. \quad (3.4)$$

3. Equação da geodésica.

Seja $p = (p_1, p_2, \dots, p_n) \in \mathbb{R}_{++}^n$ e $v = (v_1, v_2, \dots, v_n) \in T_p(\mathbb{R}_{++}^n) = \mathbb{R}^n$ com

$$x : I \rightarrow \mathbb{R}_{++}^n ; x(t) = (x_1(t), x_2(t), \dots, x_n(t))$$

tal que $x(0) = p$ e $\frac{dx(0)}{dt} = v$.

Substituindo os símbolos de Christoffel (3.1) na equação (2.2) obtemos:

$$\frac{d^2x_i}{dt^2} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \left(\frac{dx_i}{dt}\right)^2 = 0, \forall i = 1, \dots, n. \quad (3.5)$$

A equação diferencial (3.5) é equivalente a resolver:

$$\frac{dx_i}{dt} = h_i(x_i)a_i$$

para alguma constante a_i , o que também é equivalente a resolver a integral:

$$\int \frac{dx_i}{h_i(x_i)} = a_i t + b_i, \quad i = 1, 2, \dots, n$$

para certas constantes a_i e b_i em \mathbb{R} .

Assim, a única geodésica em $(\mathbb{R}_{++}^n, G(x))$ partindo do ponto p com direção v é uma função $\gamma(t)$ que resolve o seguinte problema:

$$\int \frac{d\gamma_k}{h_k(\gamma_k)} = a_k t + b_k \quad k = 1, \dots, n, \quad (3.6)$$

onde a_i e b_i são constantes reais tais que:

$$\begin{aligned} \gamma_i(0) &= p_i, & i &= 1, \dots, n, \\ \gamma'_i(0) &= v_i, & i &= 1, \dots, n. \end{aligned}$$

4. Distância Riemanniana.

Do teorema de Hopf-Rinow existe uma curva geodésica γ ligando os pontos $\gamma(0) = x$ e $\gamma(1) = y$ tal que

$$\begin{aligned} d(x, y) &= \int_0^1 \|\gamma'(t)\| dt \\ &= \int_0^1 \left(\langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} \right)^{1/2} dt \\ &= \int_0^1 \left(\sum_{i=1}^n \left(\frac{\gamma'_i(t)}{h_i(\gamma_i(t))} \right)^2 \right)^{1/2} dt. \end{aligned}$$

Como $\frac{\gamma'_i(t)}{h_i(\gamma_i(t))} = \frac{\gamma'_i(0)}{h_i(\gamma_i(0))}$ então

$$d(x, y) = \left(\sum_{i=1}^n \left(\frac{\gamma'_i(0)}{h_i(\gamma_i(0))} \right)^2 \right)^{1/2}. \quad (3.7)$$

De (3.6) temos que

$$\int \frac{1}{h_i(\gamma_i)} d\gamma_i = \frac{\gamma_i'(0)}{h_i(\gamma_i(0))} t + \left(\int \frac{d\gamma_i}{h_i(\gamma_i)} \right) \Big|_{t=0}.$$

Daqui,

$$\frac{\gamma_i'(0)}{h_i(\gamma_i(0))} = \left(\int \frac{d\gamma_i}{h_i(\gamma_i)} \right) \Big|_{t=1} - \left(\int \frac{d\gamma_i}{h_i(\gamma_i)} \right) \Big|_{t=0}$$

Substituindo esta igualdade em (3.7) obtemos finalmente

$$d(x, y) = \left(\sum_{i=1}^n \left(\left(\int \frac{d\gamma_i}{h_i(\gamma_i)} \right) \Big|_{t=1} - \left(\int \frac{d\gamma_i}{h_i(\gamma_i)} \right) \Big|_{t=0} \right)^2 \right)^{1/2}.$$

5. Curvatura Zero e Isometria Local com \mathbb{R}^n .

5.1 Curvatura Zero. Dado $p \in \mathbb{R}_{++}^n$, seja $\{X_i\}_{i=1, \dots, n}$ uma base de $T_p \mathbb{R}_{++}^n = \mathbb{R}^n$.

Podemos escrever $X = \sum_{i=1}^n u^i X_i$, $Y = \sum_{j=1}^n v^j X_j$, $Z = \sum_{k=1}^n w^k X_k$. Como o tensor curvatura R é tri-linear temos:

$$R(X, Y)Z = \sum_{i,j,k} u^i v^j w^k R(X_i, X_j)X_k.$$

Da definição de R (ver (2.4) do Capítulo 2, Seção 2.2) obtemos:

$$R(X_i, X_j)X_k = \nabla_{X_j}(\nabla_{X_i}X_k) - \nabla_{X_i}(\nabla_{X_j}X_k) + \nabla_{[X_i, X_j]}X_k.$$

Agora, para a conexão Riemanniana, temos $[X_i, X_j] = 0$, e

$$R(X_i, X_j)X_k = \nabla_{X_j}(\nabla_{X_i}X_k) - \nabla_{X_i}(\nabla_{X_j}X_k).$$

Se $i = j$, então $R(X_i, X_j)X_k = 0$.

Se $i \neq j$, da Seção 2.2 temos que $\nabla_{X_i}X_k = \sum_{j=1}^n \Gamma_{ik}^j X_j$. Substituindo os símbolos de Christoffel (3.1), obtemos

$$\nabla_{X_i}X_k = \sum_{j=1}^n \left(-\frac{1}{h_i(x_i)} \frac{\partial h_i(x_i)}{\partial x_i} \delta_{ij} \delta_{ik} \right) X_j = -\frac{1}{h_i(x_i)} \frac{\partial h_i(x_i)}{\partial x_i} \delta_{ik} X_i. \quad (3.8)$$

Portanto:

$$\nabla_{X_j}(\nabla_{X_i}X_k) = \nabla_{X_j} \left(-\frac{1}{h_i(x_i)} \frac{\partial h_i(x_i)}{\partial x_i} \delta_{ik} X_i \right).$$

Temos os seguintes casos para analisar. Se $i \neq k$, é imediato que $\nabla_{X_j}(\nabla_{X_i}X_k) = 0$, $i = 1, \dots, n$. Caso contrário, se $i = k$, e $j \neq k$, obtemos, aplicando (3.8)

$$\nabla_{X_j}(\nabla_{X_k}X_k) = \nabla_{X_j} \left(-\frac{1}{h_k(x_k)} \frac{\partial h_k(x_k)}{\partial x_k} X_k \right)$$

$$= \frac{1}{h_k(x_k)} \frac{\partial h_k(x_k)}{\partial x_k} \delta_{jk} \frac{1}{h_j(x_j)} \frac{\partial h_j(x_j)}{\partial x_j} X_j = 0.$$

Assim

$$\nabla_{X_j}(\nabla_{X_i} X_k) = 0.$$

De forma análoga,

$$\nabla_{X_i}(\nabla_{X_j} X_k) = 0.$$

Estes resultados levam a:

$$R(X_i, X_j)X_k = 0, \quad i, j, k, l = 1, 2, \dots, n.$$

Daqui,

$$R(X, Y)Z = 0.$$

Assim, a variedade Riemanniana $(\mathbb{R}_{++}^n, G(x))$ tem curvatura nula.

5.2 Isometria local com \mathbb{R}^n . Além da curvatura nula, podemos provar que $(\mathbb{R}_{++}^n, G(x))$ é localmente isométrico ao espaço Euclidiano \mathbb{R}^n . De fato, seja

$$\pi(x) = \left(\int_{x_1}^1 1/h_1(s) ds, \dots, \int_{x_n}^1 1/h_n(s) ds \right).$$

π é claramente diferenciável e $d(\pi)_x = J_\pi(x) = -\text{diag}(1/h_1(x_1), \dots, 1/h_n(x_n))$, $\forall x \in \mathbb{R}_{++}^n$, onde J_π denota a matriz Jacobiana de π . Como $h_i(x_i) > 0$ então $\det(d(\pi)_x) \neq 0$. Agora, aplicando o teorema da função inversa (ver Capítulo 2, Seção 2.3), garantimos a existência de duas vizinhanças, V_x de x e $W_{\pi(x)}$ de $\pi(x)$, tais que $\pi : V_x \rightarrow W_{\pi(x)}$ é um difeomorfismo. Por outro lado, dado $x \in \mathbb{R}_{++}^n$, sejam $u, v \in T_x \mathbb{R}_{++}^n = \mathbb{R}^n$, então

$$\langle u, v \rangle_x = u^T J_\pi(x) J_\pi(x) v = (J_\pi(x) u)^T (J_\pi(x) v) = \langle d\pi_x u, d\pi_x v \rangle_{\pi(x)}.$$

Assim, $(\mathbb{R}_{++}^n, G(x))$ é localmente isométrico a \mathbb{R}^n .

6. Condições suficientes para a completude da variedade.

Propriedades essenciais como convergência dos algoritmos, precisam, do ponto de vista das aplicações da geometria Riemanniana, a completude da variedade. Oliveira e Cruz Neto [60] provaram uma condição suficiente para a completude de \mathbb{R}_{++}^n , dotado com uma métrica diagonal, dada pela Hessiana de certa função barreira. Tal prova pode ser estendida para a métrica diagonal $G(x)$.

Teorema 3.3.1 *Seja \mathbb{R}_{++}^n a variedade com métrica invariante por H -translação $G(x)$, então*

(a) *Se $x_i^\alpha \geq \beta h_i(x_i)^2$, com $\alpha \geq 2$ para $x_i \in (0, 1]$ e $\alpha \leq 2$ para $x_i \in (1, \infty) \forall i = 1, \dots, n$ e algum $\beta > 0$, então \mathbb{R}_{++}^n é isométrico a \mathbb{R}^n . Em particular $(\mathbb{R}_{++}^n, G(x))$ é completo.*

(b) *Se $x_i^\alpha \leq \beta h_i(x_i)^2$, com $\alpha < 2$ para $x_i \in (0, 1] \forall i = 1, \dots, n$ e algum $\beta > 0$ ou se $x_i^\alpha \geq \beta h_i(x_i)^2$, com $\alpha > 2$ para $x_i \in [1, +\infty)$ para algum $\beta > 0$, então $(\mathbb{R}_{++}^n, G(x))$ é incompleto.*

7. A variedade produto.

Já sabemos que as variedades Riemannianas $(\mathbb{R}_{++}^n, G_1(x))$, $(\mathbb{R}_{++}^n, G_2(s))$ têm curvatura nula. Então, por um resultado conhecido de Geometria Riemanniana, a variedade produto $(\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n, G(x, s))$ tem curvatura nula. Além disso, se as variedades $(\mathbb{R}_{++}^n, G_1(x))$ e $(\mathbb{R}_{++}^n, G_2(s))$ são completas com geodésicas γ_1 e γ_2 , então a variedade produto $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$ é completa com geodésicas $\gamma_1 \times \gamma_2$.

8. Gradiente e Hessiana.

Seja a variedade Riemanniana \mathbb{R}_{++}^n com a métrica $G(x) = \text{diag}\left(\frac{1}{h_1(x_1)^2}, \dots, \frac{1}{h_n(x_n)^2}\right)$, para funções $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$. Seja $f : \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ diferenciável em \mathbb{R}_{++}^n .

(a) O gradiente e a Hessiana de f na variedade Riemanniana $(\mathbb{R}_{++}^n, G(x))$, denotada por $\text{grad}f$ e H_x^f respectivamente, são:

$$\text{grad}f(x) = G(x)^{-1} \nabla f(x) = (h_1(x_1)^2 \frac{\partial f}{\partial x_1}(x), \dots, h_n(x_n)^2 \frac{\partial f}{\partial x_n}(x)) \quad (3.9)$$

$$H_x^f = \nabla^2 f(x) + G(x)^{\frac{1}{2}} (G(x)^{-\frac{1}{2}})' \mathcal{F}'(x), \quad (3.10)$$

onde:

$$\mathcal{F}'(x) = \text{diag}\left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \dots, \frac{\partial f(x)}{\partial x_n}\right) \text{ e } (G(x)^{-\frac{1}{2}})' = \text{diag}\left(\frac{\partial h_1(x_1)}{\partial x_1}, \frac{\partial h_2(x_2)}{\partial x_2}, \dots, \frac{\partial h_n(x_n)}{\partial x_n}\right).$$

Em particular, se $h_i(x_i) = x_i^r$, $r \in \mathbb{R}$, então

$$\text{grad}f(x) = \sum_{i=1}^n x_i^r \frac{\partial f}{\partial x_i}(x) = X^r \nabla f(x),$$

$$H_x^f = \nabla^2 f(x) + \frac{r}{2} X^{-1} \mathcal{F}'(x).$$

De fato, dado $v \in T_x M$ temos:

$$df_x(v) = f'(x)^T v = f'(x)^T (G(x)^{-1})^T G(x)v = (G(x)^{-1} f'(x))^T G(x)v = \langle G(x)^{-1} f'(x), v \rangle_x.$$

Desta última igualdade (3.9) é obtida.

Agora provaremos a expressão (3.10). A Hessiana de f é definida por

$$H^f = \frac{D}{dt} (\text{grad } f).$$

Calculando no ponto x na direção $v \in T_x M$ temos

$$H_x^f(v) = \frac{D}{dt} (\text{grad } f)(x) = \nabla_v \text{grad } f(x).$$

Pode ser provado que o operador $H_x^f : T_x M \rightarrow T_x M$ é linear e auto-adjunto, por isso podemos introduzir a forma quadrática $q_x^f : T_x M \times T_x M \rightarrow \mathbb{R}$ definida por

$$q_x^f(v, w) = \langle H_x^f v, w \rangle_x.$$

Em termos de campos vetoriais em M , podemos então definir a seguinte aplicação

$$q^f(X, Y) = \langle \nabla_X \text{grad } f, Y \rangle.$$

É fácil verificar que

$$q^f(X, Y) = (XY - \nabla_X Y)f = (YX - \nabla_Y X)f.$$

Daqui, tomando uma base $\{X_i\}_{i=1, \dots, n}$ temos

$$q^f(X_i, X_j) = \langle H_p^f X_i, X_j \rangle = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} - \sum_{m=1}^n \Gamma_{ij}^m \frac{\partial f}{\partial x_m} \right)$$

Substituindo os símbolos de Christoffel (3.1) na última fórmula obtemos

$$q^f(X_i, X_j) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{1}{h_i(x_i)} \frac{\partial h_i(x_i)}{\partial x_i} \delta_{ij} \frac{\partial f}{\partial x_i} \right).$$

Portanto, a matriz Hessiana, denotada por simplicidade por H_x^f , é dada pela fórmula (3.10).

- (b) Analogamente, o gradiente e a Hessiana de f na variedade $((0, 1)^n, G(x))$ têm expressões semelhantes a (3.9) e (3.10).

Em particular:

- i. Seja a métrica $X^{-r} + (I - X)^{-r}$, então

$$\text{grad}f(x) = (X^{-r} + (I - X)^{-r})^{-1} \nabla f(x),$$

$$H_x^f = \nabla^2 f(x) + \frac{r}{2} [X^{-r} + (I - X)^{-r}]^{-1} [X^{-r-1} - (I - X)^{-r-1}] \mathcal{F}'(x).$$

- ii. Seja $G(x) = \pi^2 \csc^4(\pi x)$, então

$$\text{grad}f(x) = \frac{1}{\pi^2} \sin^4(\pi x) \nabla f(x),$$

$$H_x^f = \nabla^2 f(x) + 2\pi^3 X \cot(\pi x) \mathcal{F}'(x),$$

onde $\cot(\pi x) = \text{diag}(\cot(\pi x_1), \dots, \cot(\pi x_n))$.

- iii. Seja a métrica dada por $X^{-r}(I - X)^{-r}$, então

$$\text{grad}f(x) = X^r (I - X)^r \nabla f(x),$$

$$H_x^f = \nabla^2 f(x) + \frac{r}{2} [X^{-1} - (I - X)^{-1}] \mathcal{F}'(x).$$

- (c) Seja a variedade Riemanniana produto $\mathbb{R}_{++}^n \times \mathbb{R}_{++}^n$ com a métrica natural induzida, então

$$\text{grad}f(x, s) = \begin{bmatrix} G_1(x)_{n \times n} & O \\ 0 & G_2(s)_{n \times n} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial s} \end{bmatrix}.$$

- (d) Consideremos a subvariedade $M = \{x \in \mathbb{R}_{++}^n : Ax = b\}$ com a métrica $G(x)$ induzida de \mathbb{R}_{++}^n , então

$$\text{grad}f(x) = P_M(x) G(x)^{-1} \nabla f(x),$$

onde $P_M(x) = (I - G(x)^{-1} A^T (A G(x)^{-1} A^T)^{-1} A)$.

- (e) Mais geralmente, seja $l : \mathbb{R}^n \rightarrow \mathbb{R}^m$ uma função arbitrária e consideremos a subvariedade $M = \{x \in \mathbb{R}^n : l(x) = 0\}$ com a métrica $G(x)$ induzida de \mathbb{R}^n então

$$\text{grad}f(x) = P_M(x) G(x)^{-1} \nabla f(x),$$

onde $P_M(x) = (I - G(x)^{-1}J_l(x)^T(J_l(x)G^{-1}(x)J_l(x)^T)^{-1}J_l(x))$ e J_l denota a matriz Jacobiana de l .

3.4 Exemplos de Geodésicas Explícitas e Variedades Completas em \mathbb{R}_{++}^n e $(0, 1)^n$:

A seguir damos alguns exemplos de curvas geodésicas explícitas, distâncias Riemannianas e métricas que tornam as variedades \mathbb{R}_{++}^n e $(0, 1)^n$ completas. Os exemplos 2, 3, 4 e 5 são, a nosso entender, novos.

1. Considerando a variedade $(\mathbb{R}_{++}^n, X^{-r})$, a geodésica $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$ definida nesta variedade tal que $x(0) = p = (p_1, p_2, \dots, p_n)$ e $x'(0) = v = (v_1, v_2, \dots, v_n)$ é:

- (a) Para $r \neq 2$,

$$x_i(t) = \left(\frac{2-r}{2}\right)^{\frac{2}{2-r}} \left(\frac{v_i}{p_i^{\frac{r}{2}}}t + \frac{2}{2-r}p_i^{1-\frac{r}{2}}\right)^{\frac{2}{2-r}} \quad \forall i = 1, 2, \dots, n.$$

- (b) Para $r = 2$,

$$x_i(t) = p_i \exp\left(\frac{v_i}{p_i}t\right) \quad \forall i = 1, 2, \dots, n.$$

Observemos que quando $r = 2$, a geodésica $x(t)$ está definida para todo $t \in \mathbb{R}$. Além disso, a distância geodésica de um ponto p ao ponto $q = x(t_0)$, $t_0 > 0$, é dado por:

$$d(p, q) = \int_0^{t_0} \|x'(t)\| dt = \left\{ \sum_{i=1}^n \left[\ln\left(\frac{q_i}{p_i}\right) \right]^2 \right\}^{\frac{1}{2}}.$$

Assim, $(\mathbb{R}_{++}^n, X^{-2})$ é variedade completa com curvatura nula.

2. Considerando a variedade $((0, 1)^n, \pi^2 \csc^4(\pi x))$, onde $h_i(x_i) = \frac{1}{\pi} \sin^2(\pi x_i)$, a curva geodésica $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$ definida nesta variedade tal que $x(0) = p = (p_1, p_2, \dots, p_n)$ e $x'(0) = v = (v_1, v_2, \dots, v_n)$ é:

$$x_i(t) = \frac{1}{\pi} \operatorname{arccot} \cot\left(-\pi \csc^2(\pi p_i) v_i t + \cot(\pi p_i)\right), \quad i = 1, 2, \dots, n. \quad (3.11)$$

A geodésica está definida para todo $t \in \mathbb{R}$. Sua distância geodésica de um ponto p ao ponto $q = x(t_0)$, $t_0 > 0$, é dado por:

$$d(p, q) = \int_0^{t_0} \|x'(t)\| dt = \left\{ \sum_{i=1}^n [\cot(\pi q_i) - \cot(\pi p_i)]^2 \right\}^{\frac{1}{2}}. \quad (3.12)$$

Assim, esta variedade é completa com curvatura nula.

3. Considerando a variedade $((0, 1)^n, X^{-2}(I - X)^{-2})$ com $h_i(x_i) = x_i(1 - x_i)$. A curva geodésica $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$ definida nesta variedade tal que $x(0) = p = (p_1, p_2, \dots, p_n)$ e $x'(0) = v = (v_1, v_2, \dots, v_n)$ é:

$$x_i(t) = \frac{1}{2} \left\{ 1 + \tanh \left(\frac{1}{2} \frac{v_i}{p_i(1-p_i)} t + \frac{1}{2} \ln \frac{p_i}{1-p_i} \right) \right\} \quad i = 1, 2, \dots, n. \quad (3.13)$$

onde $\tanh(z) = (e^z - e^{-z})/(e^z + e^{-z})$ é a função tangente hiperbólica. Esta geodésica está definida para todo $t \in \mathbb{R}$, e a distância geodésica de um ponto p ao ponto $q = x(t_0)$, $t_0 > 0$, é dado por:

$$d(p, q) = \int_0^{t_0} \|x'(t)\| dt = \left\{ \sum_{i=1}^n \left[\ln \left(\frac{q_i}{1-q_i} \right) - \ln \left(\frac{p_i}{1-p_i} \right) \right]^2 \right\}^{\frac{1}{2}}. \quad (3.14)$$

Da mesma maneira, esta variedade é completa com curvatura nula.

4. Considere a variedade Riemanniana $((0, 1)^n, X^{-r}(I - X)^{-r})$, com $r > 2$, e $h_i(x_i) = x_i^{r/2}(1 - x_i)^{r/2}$. As curvas geodésicas não têm expressões explícitas (ver (3.6)), mas podemos provar que esta variedade é completa. De fato, usando o resultado da Seção 3.3, item 5.2, mas agora adaptado para o hipercubo $(0, 1)^n$, temos que $(0, 1)^n$ é localmente isométrico ao espaço Euclidiano \mathbb{R}^n . Por isso, para provar nosso objetivo é suficiente mostrar que π é uma função bijetiva.

Como $d\pi_i(x_i)/dx_i = -1/h_i(x_i) < 0$, π_i é uma função diferenciável estritamente decrescente em $(0, 1)$, obtendo-se que π é injetiva.

Agora, provaremos que π_i é sobrejetiva $\forall i = 1, 2, \dots, n$.

Como $r > 2$, $\pi_i(x_i) = \int_{x_i}^1 s^{-r/2}(1-s)^{-r/2} \geq \int_{x_i}^1 s^{-r/2} ds = (2/(2-r))(1-x_i^{(2-r)/2})$ então:

$$\lim_{x_i \rightarrow 0} \pi_i(x_i) \geq +\infty. \quad (3.15)$$

Agora, se $s \in (1/2, 1)$ então $s^{-r}(1-s)^{-r} < (1-s)^{-2r}$ e

$$\pi_i(x_i) \leq \int_{x_i}^1 (1-s)^{-r} ds = (1-x_i)^{1-r}/(1-r).$$

Tomando limite quando $x_i \rightarrow 1$ temos

$$\lim_{x_i \rightarrow 1} \pi_i(x_i) \leq -\infty. \quad (3.16)$$

De (3.15) e (3.16) podemos concluir que π_i é sobrejetiva. Devido a que π é injetiva e sobrejetiva então ela é bijetiva, e assim a variedade $((0, 1)^n, X^{-r}(I-X)^{-r})$ é isométrica a \mathbb{R}^n . Em particular, $((0, 1)^n, X^{-r}(I-X)^{-r})$ é uma variedade completa.

5. De forma análoga, podemos provar que a variedade Riemanniana $((0, 1)^n, X^{-r} + (I-X)^{-r})$, com $r > 2$, é isométrica a \mathbb{R}^n . Assim, esta variedade é completa.

3.5 Aplicações

Nesta seção, estamos interessados em resolver o seguinte problema de otimização: $\min f(x)$ s.a $0 \leq x_i \leq 1$, $i = 1, 2, \dots, n$, onde f é uma função diferenciável ou não diferenciável. Introduzimos alguns novos exemplos de métodos explícitos de descida geodésica e ponto proximal para resolver o problema dado. Os resultados de convergência para estes métodos precisam de hipóteses de convexidade e foram provados para métricas Riemannianas gerais. Devemos observar que estes métodos são importantes no caso em que a função objetivo não é convexa no sentido clássico (com a métrica Euclidiana) mas tornam-se convexas com respeito à métrica Riemanniana introduzida.

Para resultados de convergência e taxa de convergência do método de descida geodésica, para o caso diferenciável, ver Cruz Neto et al. [16] e para o caso não diferenciável ver Ferreira e Oliveira [23]. Já para o método de ponto proximal ver Ferreira e Oliveira [25]. Finalmente, considerando as métricas $\pi^2 \csc^4(\pi x)$ e $X^{-2}(I-X)^{-2}$, introduzimos duas novas funções barreira para o hipercubo $[0, 1]^n$.

Consideremos inicialmente o espaço $(0, 1)^n$ e duas variedades Riemannianas: a primeira dotada com a métrica $X^{-2}(I-X)^{-2}$ e a segunda com $\pi^2 \csc^4(\pi x)$. Antes de apresentar os métodos e seus respectivos resultados de convergência, precisamos deduzir algumas propriedades geométricas que serão úteis para garantir estes fatos.

- As variedades $((0, 1)^n, X^{-2}(I - X)^{-2})$ e $((0, 1)^n, \pi^2 \csc^4(\pi x))$ são geodesicamente convexas. De fato, dado dois pontos arbitrários $p, q \in (0, 1)^n$ então, para a métrica $X^{-2}(I - X)^{-2}$ a única geodésica $x : [0, 1] \rightarrow (0, 1)^n$ ligando p e q , onde $x(0) = p$ e $x(1) = q$ é:

$$x_i(t) = \frac{1}{2} \left[1 + \tanh \left(\frac{1}{2} \left\{ \left[\ln \left(\frac{q_i}{1 - q_i} \right) - \ln \left(\frac{p_i}{1 - p_i} \right) \right] t + \ln \left(\frac{p_i}{1 - p_i} \right) \right\} \right) \right],$$

e para a métrica $\pi^2 \csc^4(\pi x)$ é:

$$x_i(t) = \frac{1}{\pi} \arccot \cot [(\cot(\pi q_i) - \cot(\pi p_i))t + \cot(\pi p_i)].$$

É fácil verificar, para ambos os casos, que $0 < x_i(t) < 1, \forall i = 1, 2, \dots, n$. Assim, $(0, 1)^n$ é geodesicamente convexa com estas métricas.

- As variedades $((0, 1)^n, X^{-2}(I - X)^{-2})$ e $((0, 1)^n, \pi^2 \csc^4(\pi x))$ são conexas (trivial).
- As variedades $((0, 1)^n, X^{-2}(I - X)^{-2})$ e $((0, 1)^n, \pi^2 \csc^4(\pi x))$ são completas e de curvatura zero, com geodésicas explícitas (ver (3.11) e (3.13)).
- Gradientes e as Hessianas das funções diferenciáveis definidas nas variedades $((0, 1)^n, X^{-2}(I - X)^{-2})$ e $((0, 1)^n, \pi^2 \csc^4(\pi x))$ são explícitas (ver (3.9) e (3.10)).

3.5.1 Métodos Geodésicos

Alguns exemplos de métodos geodésicos para resolver problemas de minimização no ortante positivo e no simplex unitário foram apresentados por Cruz Neto e Oliveira [15]. Nesta seção acrescentamos métodos geodésicos para minimização no hipercubo.

Consideremos o problema de Otimização :

$$\min_{x \in \mathbb{R}^n} \{f(x) : 0 \leq x_i \leq 1, i = 1, 2, \dots, n\} \quad (3.17)$$

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

Caso diferenciável: método de Cauchy

O método proposto é de descida geodésica e consta essencialmente de 3 passos. O primeiro considera a direção do gradiente negativo da função f , $d = -\text{grad}f(x)$ (gradiente no ponto

de vista da variedade Riemanniana). O segundo passo considera a geodésica partindo de um ponto p com direção dada por d , e o último passo faz uma busca unidimensional ao longo da geodésica. Neste passo consideramos dois métodos: busca unidimensional de Armijo e o outro de passos fixos. Garante-se convergência fraca (isto é, a distância geodésica entre duas iterações consecutivas converge para zero) para funções arbitrárias diferenciáveis e convergência global com velocidade de convergência linear sob algumas condições de convexidade impostas à função f .

Método A: Método de Cauchy com busca linear de Armijo

1. Dado $x^k = (x_1^k, x_2^k, \dots, x_n^k)$, $k \geq 0$, viável, computar a direção de descida $d^k = -X_k^2(I - X_k)^2 \nabla f(x^k)$ ($d^k = -(1/\pi^2) \sin^4(\pi x^k) \nabla f(x^k)$, respectivamente) onde $X_k = \text{diag}(x_1^k, x_2^k, \dots, x_n^k)$.
2. Determinar a única geodésica $x(t)$ em $(0, 1)^n$ tal que $x(0) = x^k$ e $x'(0) = d^k$, cuja expressão explícita é:

$$x_i^k(t) = \frac{1}{2} \left\{ 1 + \tanh \left(\frac{1}{2} \frac{d_i^k}{x_i^k(1-x_i^k)} t + \frac{1}{2} \ln \frac{x_i^k}{1-x_i^k} \right) \right\}, \quad i = 1, 2, \dots, n.$$

$$\left(x_i^k(t) = \frac{1}{\pi} \arccot \left(-\pi \csc^2(\pi x_i^k) d_i^k t + \cot(\pi x_i^k) \right), \quad i = 1, 2, \dots, n, \text{ respectivamente} \right).$$

3. Escolher $t_k = 2^{-i_k} \bar{t}$, onde $\bar{t} > 0$ é dado e i_k é o menor número natural positivo tal que:

$$f(x(t_k)) \leq f(x^k) - \beta t_k^2 \|d^k\|^2,$$

com $\beta \in (0, 1)$.

4. Fazer $x^{k+1} = x^k(t_k)$ e obter a distância geodésica entre os pontos x^k e x^{k+1} dada por:

$$d(x^k, x^{k+1}) = \left\{ \sum_{i=1}^n \left[\ln \left(\frac{x_i^{k+1}}{1-x_i^{k+1}} \right) - \ln \left(\frac{x_i^k}{1-x_i^k} \right) \right]^2 \right\}^{\frac{1}{2}}$$

$$\left(d(x^k, x^{k+1}) = \left\{ \sum_{i=1}^n \left[\cot(\pi x_i^{k+1}) - \cot(\pi x_i^k) \right]^2 \right\}^{\frac{1}{2}}, \text{ respectivamente} \right).$$

5. Testar o critério de parada $\|d(x^k, x^{k+1})\| < \epsilon$. Se cumpre a condição, parar. Senão, fazer $x^k \leftarrow x^{k+1}$ e retornar ao passo 1.

Método B: Método de Cauchy com passos fixos

Este método é similar ao anterior, exceto pelo passo 4, o qual é substituído por:

Dados $\delta_1 > 0$ e $\delta_2 > 0$ tal que $\delta_1\Gamma + \delta_2 < 1$, escolher

$$t_k \in \left(\delta_1, \frac{2}{\Gamma}(1 - \delta_2) \right),$$

onde Γ é a constante de Lipschitz associada a $\text{grad}f$, isto é, Γ satisfaz a seguinte propriedade: para quaisquer $p, q \in (0, 1)^n$ e qualquer segmento geodésico $\alpha : [0, a] \rightarrow (0, 1)^n$ ligando p e q :

$$\|\text{grad}f(\alpha(t)) - P_{\alpha,0,t}(\text{grad}f(p))\| \leq \Gamma d(\alpha(0), \alpha(t)) \quad (3.18)$$

para $t \in [0, a]$ arbitrário, onde $P_{\alpha,0,t}$ é o transporte paralelo, ver (3.3) e (3.4).

Resultados de Convergência

Temos os seguintes resultados, todos demonstrados para métricas gerais por Cruz Neto et al. [16]. Devemos notar que para o método B, a função objetivo deve satisfazer a propriedade (3.18).

Teorema 3.5.1 (Teorema 5.1 [16]). *Seja $f \in C^1$ e $\{x^k\}$ uma seqüência de pontos gerados por um dos métodos acima, então*

1. *Existe uma constante $\beta \in (0, 1)$, tal que*

$$f(x^{k+1}) \leq f(x^k) - \beta t_k^2 \|d^k\|^2 \quad (3.19)$$

Em particular, $\{f(x^k)\}$ é não crescente.

2. *A seqüência $\{x^k\}$ é fracamente convergente, em relação à métrica, no seguinte sentido:*

(a) *$\{x^k\}$ é limitado.*

(b) *$\lim_{k \rightarrow \infty} d(x^{k+1}, x^k) = 0$*

(c) *Qualquer ponto de acumulação de $\{x^k\}$ é um ponto crítico de f .*

Teorema 3.5.2 (Teorema 5.3 [16]). *Seja $f \in C^1$ geodesicamente convexa, então a seqüência $\{x^k\}$ converge globalmente para um ponto de mínimo.*

Teorema 3.5.3 (Teorema 5.4 [16]). *Seja $f \in C^2$ e (geodesicamente) fortemente convexa, isto é, $a \leq H^f \leq b$, onde H^f é hessiana de f e $0 < a < b$. Então a velocidade de convergência do algoritmo geodésico de Cauchy com busca de Armijo é dada pelo raio de Kantorovich.*

Caso não diferenciável: método subgradiente

O método subgradiente em variedades Riemannianas é uma extensão natural do algoritmo subgradiente introduzido por Shor [83] e generaliza o algoritmo gradiente, estudado na subseção anterior.

Considere que a função f do problema (3.17) é geodesicamente convexa na variedade Riemanniana $((0, 1)^n, X^{-2}(I-X)^{-2})$ ou respectivamente na variedade $((0, 1)^n, \pi^2 \csc^4(\pi x))$. Seja O^* , o conjunto de minimizadores de f , $f^* = \inf f(x)$, o valor ínfimo de f e $\partial f(x)$ o subdiferencial de f no ponto $x \in (0, 1)^n$. O problema é estimar f^* e também encontrar um ponto de O^* se este conjunto não é vazio.

Método C

Dada uma seqüência de números reais $\{t_k\}$ tal que $t_k > 0, \forall k = 1, 2, \dots$

1. Inicializar. Escolha $x_1 \in (0, 1)^n$ e obtenha $s_1 \in \partial f(x_1)$. Fazer $k = 1$.
2. Se $s_k = 0$, parar. Senão, determinar a única geodésica $x(t)$ de $(0, 1)^n$ tal que $x(0) = x^k$ e $x'(0) = d^k = -\frac{s_k}{\|s_k\|}$, cuja expressão explícita é:

$$x_i^k(t) = \frac{1}{2} \left\{ 1 + \tanh \left(\frac{1}{2} \frac{d_i^k}{x_i^k(1-x_i^k)} t + \frac{1}{2} \ln \frac{x_i^k}{1-x_i^k} \right) \right\}, \quad i = 1, 2, \dots, n$$

$$\left(x_i^k(t) = \frac{1}{\pi} \arccot \left(-\pi \csc^2(\pi x_i^k) d_i^k t + \cot(\pi x_i^k) \right), \quad i = 1, 2, \dots, n, \text{ respectivamente} \right).$$

3. Fazer $x^{k+1} = x_k(t_k)$.
4. Obter $s_{k+1} \in \partial f(x^{k+1})$, e voltar ao passo 2.

Resultados de Convergência

O seguinte resultado é uma adaptação para nossas variedades $((0, 1)^n, X^{-2}(I-X)^{-2})$ e $((0, 1)^n, \pi^2 \csc^4(\pi x))$ do resultado geral para curvatura não negativa demonstrado por Ferreira e Oliveira [23].

Teorema 3.5.4 (Teorema 5.1 [23]). *Seja f uma função convexa na variedade $(0, 1)^n$ e $\{x^k\}$ uma seqüência de pontos gerados pelo método acima. Se a seqüência de números reais $\{t_k\}$,*

com $t_k > 0, \forall k = 1, 2, \dots$, é escolhida tal que

$$\sum_{k=0}^{\infty} t_k = \infty,$$

$$\sum_{k=0}^{\infty} t_k^2 < \infty,$$

então $\liminf f(x^k) \rightarrow f^*$, quando $k \rightarrow \infty$. Além disso, se $O^* \neq \emptyset$, então a seqüência $\{x^k\}$ converge a um ponto $x^* \in O^*$.

3.5.2 Método de Ponto Proximal

Consideremos o problema (3.17). Para cada $x \in (0, 1)^n$, a regularização de f (introduzido por Moreau-Yosida), com $\beta > 0$, é dada por:

$$f_{\beta}(x) = \min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{\beta}{2} d^2(x, y) : y \in (0, 1)^n \right\}$$

Nesta aplicação, a distância d pode ser escolhida como (3.12), ou (3.14).

Definição 3.5.1 O ponto $x = \bar{x}(x) = \arg \min_{x \in (0, 1)^n} f_{\beta}(x)$, é chamado ponto proximal de x com respeito a β, f e d^2 .

Baseado nas propriedades geométricas de $((0, 1)^n, X^{-2}(I - X)^{-2})$ e $((0, 1)^n, \pi^2 \csc^4(\pi x))$ (curvatura zero, completude, geodésicas e distâncias Riemannianas explícitas), apresentamos o método de ponto proximal estudado no contexto das variedades Riemannianas com curvatura seccional não positiva por Ferreira e Oliveira [25].

Método D: Ponto proximal com distâncias Riemannianas.

Dados um ponto inicial $x^0 \in (0, 1)^n$ e $\beta^0 > 0$. Escolha a seqüência $\{\beta^k\}$ tal que

$$\sum_{k=0}^{\infty} \frac{1}{\beta^k} = +\infty.$$

1. Fazer $k = 0$.
2. Testar o critério de parada para o ponto x^k ($0 \in \partial f(x^k)$).

3. Calcular

$$x^{k+1} = \arg \min \left\{ f(y) + \frac{\beta^k}{2} d^2(x^k, y) : y \in (0, 1)^n \right\},$$

onde d é dada por (3.12) ou (3.14).

4. Atualizar β^k , k , x^k , e voltar ao passo 2.

Um importante fato deste método, a diferença dos algoritmos apresentados na seção anterior, é que ele não precisa da busca geodésica.

Resultado de Convergência

Teorema 3.5.5 *Se f é geodesicamente convexa na variedade $(0, 1)^n$ e existe uma solução ótima, então a seqüência gerada pelo método converge globalmente para um mínimo do problema.*

3.5.3 Barreiras para Problemas Canalizados

Devemos notar que as duas métricas Riemannianas estudadas, a saber, $X^{-2}(I - X)^{-2}$ e $\pi^2 \csc^4(\pi x)$ são, respectivamente, as Hessianas das seguintes funções:

$$b_1(x) = -\frac{1}{6} \sum_{i=1}^n [4 \ln(\sin \pi x_i) - \cot^2(\pi x_i)],$$

$$b_2(x) = \sum_{i=1}^n (2x_i - 1) [\ln x_i - \ln(1 - x_i)].$$

De fato, as derivadas de primeira e segunda ordem são:

$$b'_1(x)_i = -\pi \left(\cot(\pi x_i) + \frac{1}{3} \cot^3(\pi x_i) \right), \quad i = 1, \dots, n.$$

$$b''_1(x) = \pi^2 \csc^4(\pi x).$$

$$b'_2(x)_i = 2 [\ln x_i - \ln(1 - x_i)] + (2x_i - 1) \left[\frac{1}{x_i(1 - x_i)} \right], \quad i = 1, \dots, n.$$

$$b''_2(x) = X^{-2}(I - X)^{-2}.$$

Claramente, ambas Hessianas são definidas positivas, por isso, as funções b_1 e b_2 são estritamente convexas. Além disso, se $x \rightarrow \partial((0, 1)^n)$ (x se aproxima da fronteira de $(0, 1)^n$) então

$b_1(x) \rightarrow \infty$ e $b_2(x) \rightarrow \infty$, por isso elas são funções barreira para o hipercubo $[0, 1]^n$. A segunda função, b_2 , será estudada no Capítulo 6 em um contexto mais específico para algoritmos de pontos interiores seguidores da trajetória central.

Nota Bibliográfica

Parte deste capítulo se encontra em [61] e [64].

Capítulo 4

Método de Máxima Descida em Variedades Riemannianas para Funções Quase-Convexas

4.1 Introdução

O método de máxima descida é um dos mais antigos e mais conhecidos na literatura para resolver problemas de otimização diferenciáveis. Porém os resultados de convergência para uma função objetivo arbitrária são fracos devido à convergência global do método não estar garantida. Somente podemos provar que todo ponto de acumulação, se ele existe, é um ponto crítico do problema. A situação é muito diferente quando a função objetivo é convexa, já que com a única hipótese de existência das soluções ótimas o método de máxima descida com busca de Armijo e uma regularização proximal converge a um ponto ótimo, ver Burachik et al. [9] e Iusem e Svaiter [42] respectivamente. O primeiro resultado foi recentemente estendido para otimização vetorial por Graña Drummond e Svaiter [35].

O método de máxima descida para resolver problemas de minimização para funções arbitrárias em variedades Riemannianas de dimensão finita foi estudado por Udriste [82], Smith [79] and Rapcsák [75], onde eles obtiveram os mesmos resultados de convergência do método em \mathbb{R}^n . Para o caso convexo em uma variedade Riemanniana com curvatura seccional não neg-

ativa, a convergência global do método usando busca de Armijo, passos fixos e uma regularização proximal foi desenvolvido por Cruz Neto et al. [16] e [18].

Neste capítulo estendemos a convergência global do método de máxima descida sobre estas variedades para o caso quase-convexo seguindo as ideias de Kiwiel e Murty [48] mas usando a teoria de quase-Fejér convergência. De fato, com a hipótese que o conjunto dos pontos ótimos é não vazio provamos a convergência global deste método para um ponto crítico do problema.

O esquema deste capítulo é a seguinte. Na Seção 4.2 apresentamos alguns resultados básicos de quase-Fejér convergência e análise convexa em variedades Riemannianas. Na Seção 4.3 estudamos o método de máxima descida em variedades Riemannianas completas de dimensão finita e com curvatura seccional não negativa para problemas de minimização quando a função objetivo é quase-convexa. Finalmente, na Seção 4.4 apresentamos alguns exemplos do método de máxima descida.

4.2 Elementos Básicos

Definição 4.2.1 *Seja (X, d) um espaço métrico completo. Uma seqüência $\{y^k\}$, $k \geq 0$, de X , é quase-Fejér convergente ao conjunto $U \subset X$, se para cada $u \in U$ existe uma seqüência $\{\epsilon_k\} \subseteq \mathbb{R}$ tal que $\epsilon_k \geq 0$, $\sum_{k=0}^{+\infty} \epsilon_k < +\infty$ e*

$$d^2(y^{k+1}, u) \leq d^2(y^k, u) + \epsilon_k.$$

Teorema 4.2.1 *Em um espaço métrico completo (X, d) , se $\{y^k\}$ é quase-Fejér convergente para um conjunto $U \subseteq X$, então $\{y^k\}$ é limitado. Se, além disso, um ponto de acumulação \bar{y} de $\{y^k\}$ pertence a U , então $\{y^k\}$ converge e $\lim_{k \rightarrow +\infty} y^k = \bar{y}$.*

Demonstração. Analoga à prova de Burachik et al. [9] substituindo a norma Euclidiana pela distância d . ■

Definição 4.2.2 *Seja M uma variedade Riemanniana e $f : M \rightarrow \mathbb{R}$ uma função real. f é chamada quase-convexa em M se para todo $x, y \in M$, $t \in [0, 1]$, se cumpre*

$$f(\gamma(t)) \leq \max\{f(x), f(y)\},$$

para toda curva geodésica $\gamma : [0, 1] \rightarrow M$, tal que $\gamma(0) = x$ e $\gamma(1) = y$.

Teorema 4.2.2 *Seja $f : M \rightarrow \mathbb{R}$ uma função diferenciável e quase-convexa em uma variedade Riemanniana completa M e seja $x, y \in M$. Se $f(x) \leq f(y)$ então*

$$\langle \text{grad}f(y), \gamma'(0) \rangle \leq 0,$$

onde $\text{grad}f$ é o gradiente de f e γ é a curva geodésica tal que $\gamma(0) = y$ e $\gamma(1) = x$.

Demonstração. Ver Németh [56], Proposição 3.1. ■

Definição 4.2.3 *Uma função diferenciável $f : M \rightarrow \mathbb{R}$ é pseudoconvexa se, para todo par de pontos distintos $x, y \in M$ e toda curva geodésica ligando x a y ($\gamma(0) = x$ e $\gamma(1) = y$) temos*

$$\langle \text{grad}f(x), \gamma'(0) \rangle \geq 0, \quad \text{então} \quad f(y) \geq f(x).$$

Teorema 4.2.3 *Seja $f : M \rightarrow \mathbb{R}$ uma função diferenciável e pseudoconvexa. Então, x^* é um mínimo global de f se, somente se, $\text{grad}f(x^*) = 0$.*

Demonstração. Imediato. ■

4.3 O Método de Máxima Descida

Estamos interessados em resolver o seguinte problema de otimização:

$$(p) \min_{x \in M} f(x)$$

onde M é uma variedade Riemanniana conexa, completa de dimensão finita e $f : M \rightarrow \mathbb{R}$ é uma função continuamente diferenciável e quase-convexa. O método de máxima descida gera uma seqüência de pontos $\{x^k\}$ dados por:

$$x^0 \in M, \tag{4.1}$$

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad}f(x^k)) \tag{4.2}$$

onde \exp é a aplicação exponencial e t_k é um parâmetro positivo.

Hipótese A1. O conjunto de pontos ótimos globais do problema (p), denotado por X^* , é não

vazio.

Denotamos o valor ótimo de (p) por f^* . Agora, definamos o seguinte conjunto

$$U := \{x \in M : f(x) \leq \inf_k f(x^k)\}.$$

O lema seguinte é a chave de nosso trabalho já que ele será usado para provar que a seqüência, definida pelo método de máxima descida, é quasi-Fejér convergente a U .

Lemma 4.3.1 *Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e quase-convexa em uma variedade Riemanniana conexa, completa e de dimensão finita com curvatura seccional não negativa, então*

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + t_k^2 \|\text{grad}f(x^k)\|^2$$

para todo $x \in U$ e todo $t_k > 0$.

Demonstração. Seja $x \in U$ arbitrário. Seja também $\gamma_1 : [0, 1] \rightarrow M$ a geodésica minimal ligando x^k e x e $\gamma_2 : [0, 1] \rightarrow M$ uma geodésica ligando x^k e x^{k+1} com $\gamma_2'(0) = -t_k \text{grad}f(x^k)$.

Do Teorema 2.2.3 temos:

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + t_k^2 \|\text{grad}f(x^k)\|^2 + 2t_k d(x^k, x) \langle \text{grad}f(x^k), \gamma_1'(0) \rangle.$$

Como f é quase-convexa e $f(x) \leq f(x^k)$, do Teorema 4.2.2 obtemos que

$$\langle \text{grad}f(x^k), \gamma_1'(0) \rangle \leq 0.$$

Usando este resultado na desigualdade anterior obtemos o resultado desejado. ■

A partir de agora M denotará uma variedade Riemanniana *conexa, completa e de dimensão finita com curvatura seccional não negativa*.

4.3.1 Método com uma busca de Armijo generalizada

O método de máxima descida com busca de Armijo gera uma seqüência de pontos $\{x^k\}$ dados por (4.1)-(4.2) onde

$$t_k = \arg \max \{t : f(\exp_{x^k}(-t \text{grad}f(x^k))) \leq f(x^k) - \alpha t \|\text{grad}f(x^k)\|^2, t = 2^{-i}, i = 0, 1, \dots\} \quad (4.3)$$

com $\alpha \in (0, 1)$.

Nesta subseção provamos a convergência global deste método para o caso quase-convexo. Nossos resultados são uma generalização de Kiwiel e Murty [48] para variedades Riemannianas e estende prévios resultados de convergência obtidos, para o caso convexo, por Burachik et al. [9] e Cruz Neto et al. [16]. Como em [48], considere a seguinte hipótese:

Hipótese A2. Seja $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ uma função tal que:

A2.1 Existe $\alpha \in (0, 1)$, $\tau_\alpha > 0$, tal que $\forall t \in (0, \tau_\alpha] : \phi(t) \leq \alpha t$,

A2.2 Existe $\beta > 0$, $\tau_\beta \in (0, +\infty]$, tal que $\forall t \in (0, \tau_\beta) \cap \mathbb{R} : \phi(t) \geq \beta t^2$,

A2.3 Para todo k , $f(x^{k+1}) \leq f(x^k) - \phi(t_k) \|\text{grad}f(x^k)\|^2$ e $0 < t_k \leq \tau_\beta$ em (4.2),

A2.4 Existe $\gamma > 1$, $\tau_\gamma > 0$, tal que $\forall k : t_k \geq \tau_\gamma$ ou

$$\left[\text{existe } \bar{t}_k \in [t_k, \gamma t_k] : f(\exp_{x^k}(-\bar{t}_k \text{grad}f(x^k))) \geq f(x^k) - \phi(\bar{t}_k) \|\text{grad}f(x^k)\|^2 \right].$$

Observação 4.3.1 Observemos que a hipótese A2 é satisfeita pela regra de Armijo (4.3) para $\phi(t) = \alpha t$, $\beta = \alpha$, $\gamma = 2$ e $\tau_\alpha = \tau_\beta = \tau_\gamma = 1$.

Observação 4.3.2 A hipótese A2 também é satisfeita pelo método de máxima descida com passos fixos introduzida em [9] e generalizada para variedades Riemannianas em [16]. De fato, em [9] e [16] a regra para obter t_k é a seguinte:

Dados δ_1 e δ_2 tal que $\delta_1 \Gamma + \delta_2 < 1$, onde Γ é a constante de Lipschitz associada a $\text{grad}f$, escolher

$$t_k \in \left(\delta_1, \frac{2}{\Gamma}(1 - \delta_2) \right).$$

Definindo $\phi(t) = \beta t^2$, com $\beta = \frac{\Gamma \delta_2}{2(1 - \delta_2)}$, $\tau_\gamma = \delta_1$, $\tau_\beta = (2/\Gamma)(1 - \delta_2)$, $\alpha \in (0, 1)$ arbitrário e $\tau_\alpha = \alpha/\beta$, garantimos a hipótese A2.

Proposição 4.3.1 Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e quase-convexa. Suponha que as hipóteses A1 e A2 são satisfeitas. Então a seqüência $\{x^k\}$ gerada pelo método de máxima descida com busca de Armijo generalizada é quase-Fejér convergente a U .

Demonstração. Das hipóteses A2.2 e A2.3 temos

$$\beta t_k^2 \|\text{grad}f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}) \quad (4.4)$$

Isto implica que

$$\sum_{k=0}^{+\infty} t_k^2 \|\text{grad}f(x^k)\|^2 \leq \frac{f(x^0) - f^*}{\beta} < +\infty.$$

Do Lemma 4.3.1 e Definição 4.2.1 temos o resultado. ■

Teorema 4.3.1 *Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e quase-convexa. Suponha que as hipóteses A1 e A2 são satisfeitas. Então a seqüência $\{x^k\}$ gerada pelo método de máxima descida com busca de Armijo generalizada converge. Além disso, ela converge para um ponto estacionário (um ponto \bar{x} tal que $\text{grad}f(\bar{x}) = 0$).*

Demonstração. Da Proposição anterior, $\{x^k\}$ é quase-Fejér convergente a U , assim $\{x^k\}$ é limitado (ver Teorema 4.2.1). Então existem \bar{x} e uma subsequência $\{x^{k_j}\}$ de $\{x^k\}$ convergindo para \bar{x} . Da continuidade de f obtemos

$$\lim_{j \rightarrow +\infty} f(x^{k_j}) = f(\bar{x}).$$

Devido a que $\{f(x^k)\}$ é uma seqüência não crescente, ver (4.4), com uma subsequência convergindo para $f(\bar{x})$, toda a seqüência converge para $f(\bar{x})$ e assim

$$f(\bar{x}) \leq f(x^k), \forall k \in \mathbb{N}.$$

Isto implica que $\bar{x} \in U$. Agora, do Teorema 4.2.1, concluímos que $\{x^k\}$ converge para \bar{x} .

Finalmente, provaremos que $\text{grad}f(\bar{x}) = 0$. Por contradição, suponhamos que $\text{grad}f(\bar{x}) \neq 0$.

Claramente, temos que $\text{grad}f(x^k) \rightarrow \text{grad}f(\bar{x}) \neq 0$ e $f(x^k) \rightarrow f(\bar{x})$. Agora, de (4.4), cumpre-se que

$$\lim_{k \rightarrow +\infty} t_k = 0. \quad (4.5)$$

Por outro lado, usando A2.4 e A2.1, temos, para k suficientemente grande,

$$f(\exp_{x^k}(-\bar{t}_k \text{grad}f(x^k))) - f(x^k) \geq -\alpha \bar{t}_k \|\text{grad}f(x^k)\|^2. \quad (4.6)$$

Além disso, do teorema do valor medio, para cada k , existe $t_k^* \in [0, \bar{t}_k]$ tal que

$$-\langle \text{grad}f(\exp_{x^k}(-t_k^* \text{grad}f(x^k))), P_{\gamma_k, 0, t_k^*} \text{grad}f(x^k) \rangle \geq -\alpha \|\text{grad}f(x^k)\|^2$$

onde $P_{\gamma_k, 0, t_k^*}$ é o transporte paralelo ao longo da geodésica γ_k tal que $\gamma_k(0) = x^k$ e $\gamma_k'(0) = -\text{grad}f(x^k)$. Agora, (4.5) e A2.4 implicam que $\lim_{k \rightarrow +\infty} t_k^* = 0$. Fazendo $k \rightarrow +\infty$ na desigualdade acima e tomando em conta a continuidade de $\text{grad}f$, \exp e o transporte paralelo, temos que $1 \leq \alpha$, o que contradiz A2.1. Portanto, $\text{grad}f(\bar{x}) = 0$. ■

Como consequência do teorema anterior e do Teorema 4.2.3 temos o seguinte resultado.

Corolário 4.3.1 *Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e pseudoconvexa. Então, com as hipóteses A1 e A2, a seqüência $\{x^k\}$ converge para um ponto de mínimo global do problema (p).*

4.3.2 Método com uma regularização proximal

Seja $\{\lambda_k\}$ uma seqüência de números reais tal que

$$\lambda' \leq \lambda_k \leq \lambda'',$$

onde $0 < \lambda' \leq \lambda''$. O método de máxima descida com uma regularização proximal gera uma seqüência $\{x^k\}$ definida por (4.1)-(4.2) onde

$$t_k = \arg \min \{f(\exp_{x^k}(-t \text{grad}f(x^k))) + t^2 \lambda_k \|\text{grad}f(x^k)\|^2 : t \geq 0\}. \quad (4.7)$$

Este método foi introduzido por Iusem e Svaiter em [42] para resolver problemas de otimização convexa em espaços Euclidianos e logo generalizado para variedades Riemannianas em Cruz Neto et al. [18]. Nesta subseção, estendemos os resultados de convergência global destes trabalhos para o caso quase-convexo.

Proposição 4.3.2 *Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e quase-convexa. Suponha que a hipótese A1 é satisfeita. Então, a seqüência $\{x^k\}$, gerada por (4.1), (4.2) e (4.7), é quase-Fejér convergente ao conjunto U .*

Demonstração. De (5.11) e (4.7) :

$$f(x^{k+1}) + t_k^2 \lambda_k \|\text{grad}f(x^k)\|^2 \leq f(x^k). \quad (4.8)$$

Daqui, é fácil verificar que

$$\sum_{k=0}^{+\infty} t_k^2 \|\text{grad}f(x^k)\|^2 \leq (1/\lambda')(f(x^0) - f^*) < +\infty.$$

Do Lema 4.3.1 e Definição 4.2.1, obtemos o resultado desejado. ■

Teorema 4.3.2 *Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e quase-convexa. Suponha que a hipótese A1 é satisfeita. Então, a seqüência $\{x^k\}$, gerada por (4.1),(4.2) e (4.7), converge a um ponto estacionário.*

Demonstração. De (4.8) temos que $\{f(x^k)\}$ é uma seqüência não crescente. Usando os mesmos argumentos da demonstração do Teorema 4.3.1, podemos mostrar que $\{x^k\}$ converge a um ponto $x^* \in U$. Finalmente, temos $\text{grad}f(x^*) = 0$, como uma aplicação do Teorema 4.1, *iiii*, em [18], onde isto foi provado para uma função arbitária. ■

Similar ao Corolário 4.3.1 temos o seguinte resultado

Corolário 4.3.2 *Seja $f : M \rightarrow \mathbb{R}$ uma função continuamente diferenciável e pseudoconvexa. Então, com a hipótese A1, a seqüência $\{x^k\}$ converge a um ponto de mínimo global de (p) .*

4.4 Alguns Exemplos

Os seguintes algoritmos resolvem, em particular, problemas de minimização não convexa em espaços Euclidianos.

4.1 Um algoritmo de máxima descida para \mathbb{R}^n .

Considere o problema

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

Considere \mathbb{R}^n associado à métrica

$$G(x) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & 0 & 0 & 1 + 4x_{n-1}^2 & -2x_{n-1} \\ 0 & \dots & 0 & 0 & 0 & -2x_{n-1} & 1 \end{bmatrix}.$$

Assim, $(\mathbb{R}^n, G(x))$ é uma variedade Riemanniana conexa, completa, de dimensão finita e com curvatura seccional nula, ver [17]. O gradiente de f é dado por $\text{grad}f(x) = G^{-1}(x)\nabla f(x)$ e a iteração do método é

$$\begin{aligned} x_i^{k+1} &= x_i^k - t_k \frac{\partial f(x^k)}{\partial x_i}, \forall i = 1, \dots, n-2, \\ x_{n-1}^{k+1} &= x_{n-1}^k - t_k \left(\frac{\partial f(x^k)}{\partial x_{n-1}} + 2x_{n-1}^k \frac{\partial f(x^k)}{\partial x_n} \right), \\ x_n^{k+1} &= x_n^k - t_k \left(2x_{n-1}^k \frac{\partial f(x^k)}{\partial x_{n-1}} + (1 + 4(x_{n-1}^k)^2) \frac{\partial f(x^k)}{\partial x_n} \right). \end{aligned}$$

4.2 Um algoritmo de máxima descida para \mathbb{R}_{++}^n .

Considere o problema

$$\min\{f(x) : x \geq 0\}.$$

Tome o ortante positivo \mathbb{R}_{++}^n e considere a variedade Riemanniana $(\mathbb{R}_{++}^n, X^{-2})$ (X^{-2} é a Hessiana da barreira logarítmica). Este espaço é conexo, completo e de dimensão finita com curvatura seccional nula, ver Capítulo 3, Seção 3.4. O gradiente de f é dado por $\text{grad}f(x) = X^2\nabla f(x)$ (o oposto da direção affim escala) e a iteração é

$$x_i^{k+1} = x_i^k \exp\left(-x_i^k \frac{\partial f(x^k)}{\partial x_i} t_k\right), \quad i = 1, 2, \dots, n.$$

4.3 Algoritmo de máxima descida para o hipercubo.

Considere o problema

$$\min\{f(x) : 0 \leq x \leq e\}.$$

Consideremos a variedade $(0, 1)^n$. Introduziremos três variedades Riemannianas conexas, completas com curvatura seccional nula, ver Capítulo 3, Seção 3.4.

- a.** $((0, 1)^n, X^{-2}(I - X)^{-2})$. Então, $\text{grad}f(x) = X^2(I - X)^2\nabla f(x)$. A iteração de máxima decida é

$$x_i^{k+1} = \frac{1}{2} \left\{ 1 + \tanh \left(-\frac{1}{2} x_i^k (1 - x_i^k) \frac{\partial f(x^k)}{\partial x_i} t_k + \frac{1}{2} \ln \frac{x_i^k}{1 - x_i^k} \right) \right\}, \quad i = 1, 2, \dots, n.$$

- b.** $((0, 1)^n, \csc^4(\pi x))$. Então, $\text{grad}f(x) = \sin^4(\pi x)\nabla f(x)$. A iteração de máxima decida é:

$$x_i^{k+1} = \frac{1}{\pi} \text{arc cot} \left(\pi \sin^2(\pi x_i^k) \frac{\partial f(x^k)}{\partial x_i} t_k + \cot(\pi x_i^k) \right), \quad i = 1, 2, \dots, n.$$

- c.** $((0, 1)^n, \csc^2(\pi x))$, ver Nesterov e Todd [59]. O gradiente de f é dado por $\text{grad}f(x) = \sin^2(\pi x)\nabla f(x)$ e a iteração de máxima decida é:

$$x_i^{k+1} = \psi^{-1} \left(-\sin(\pi x_i^k) \frac{\partial f(x^k)}{\partial x_i} t_k + \psi(x_i^k) \right), \quad i = 1, 2, \dots, n,$$

onde

$$\psi(z) := (-1/\pi) \ln(\csc(\pi z) + \cot(\pi z)).$$

Capítulo 5

Método de Ponto Proximal com Distâncias de Bregman em Variedades de Hadamard

5.1 Introdução

Considere o seguinte problema de otimização:

$$\min_{x \in X} f(x),$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa sobre um conjunto convexo fechado X de \mathbb{R}^n .

O método de ponto proximal com distância de Bregman, abreviado a partir de agora como método PDB, gera uma seqüência $\{x^k\}$ definida por:

Dado $x^0 \in S$,

$$x^k = \arg \min_{x \in X \cap \bar{S}} \{f(x) + \lambda_k D_h(x, x^{k-1})\},$$

onde h é uma certa função definida em \bar{S} , tal que, $X \cap \bar{S} \neq \emptyset$, λ_k é um parâmetro positivo e D_h é uma distância de Bregman definido como:

$$D_h(x, y) = h(x) - h(y) - \langle \nabla f(y), x - y \rangle,$$

onde $\langle \cdot, \cdot \rangle$ denota o produto interno usual em \mathbb{R}^n .

Resultados de convergência e taxa de convergência, com hipóteses apropriadas sobre o problema, foram provadas por varios pesquisadores para certas escolhas do parâmetro regularizador

λ_k (ver por exemplo [12, 14, 44, 47]). Tal algoritmo também foi generalizado para resolver problemas de Desigualdades Variacionais em espaços mais gerais como espaços de Hilbert e Banach, ver [8, 10, 29]. Lembremos que problemas de Desigualdades Variacionais surgem naturalmente em varias aplicações da Engenharia e recobrem problemas de otimização como um caso particular.

Neste capítulo generalizamos o método PDB para resolver problemas de otimização para funções convexas e quase-convexas em variedades de Hadamard. Nossa abordagem é nova, mas ela está relacionada ao trabalho desenvolvido por Ferreira e Oliveira [25]. Naquele artigo os autores generalizaram o método de ponto proximal usando a distância Riemanniana em variedades de Hadamard. Neste trabalho consideramos distâncias de Bregman e usamos os seguintes parâmetros de regularização

$$\lim_{k \rightarrow +\infty} \lambda_k = 0, \quad \text{com } \lambda_k > 0. \quad (5.1)$$

$$0 < \lambda_k < \bar{\lambda}, \quad (5.2)$$

Para λ_k satisfazendo (5.1), obtemos a convergência do algoritmo PDB para problemas de otimização com funções quase-convexas. Para λ_k satisfazendo (5.2), obtemos a convergência do algoritmo PDB para problemas de otimização com funções convexas. A noção de quase-convexidade aparece em aplicações em Economia [3, 30, 81] e recentemente também na teoria de controle [4] e em sistemas dinâmicos [33].

Este capítulo está dividido como segue. Na Seção 5.2 lembramos alguns fatos de análise convexa em variedades de Hadamard. Na Seção 5.3 é introduzida a definição de funções de Bregman em variedades de Hadamard, além disso, fornecemos algumas propriedades destas funções. Na Seção 5.4 apresentamos a regularização de Moreau-Yosida considerando distâncias e funções de Bregman. Na Seção 5.5 introduzimos o método PDB e as hipóteses para resolver problemas de otimização em variedades de Hadamard. Logo, na Seção 5.6, provamos a convergência da seqüência gerada pelo método, para funções quase-convexas e contínuas usando a condição (5.1) e para funções convexas, quando o parâmetro regularizador verifica (5.2). A Seção 5.7 é uma aplicação de nossa abordagem para o método de ponto proximal com distâncias Riemannianas para resolver problemas de otimização com funções quase-convexas, estendendo assim os resultados obtidos em [25]. Finalmente, na Seção 5.8 apresentamos alguns

exemplos de distâncias de Bregman em algumas variedades de Hadamard.

5.2 Análise Convexa em Variedades de Hadamard

Nesta Seção estabelecemos algumas definições e resultados básicos de Análise Convexa em variedades de Hadamard. As referências, para leitores interessados, são Ferreira e Oliveira [25] e Udriste [82].

Definição 5.2.1 *Seja M uma variedade de Hadamard. Um subconjunto A é dito convexo em M se para qualquer par de pontos, a (única) geodésica ligando estes pontos está contida em A , isto é, dados $x, y \in A$ e $\gamma : [0, 1] \rightarrow M$ a curva geodésica tal que $\gamma(0) = x$ e $\gamma(1) = y$, temos que $\gamma(t) \in A$, para todo $t \in [0, 1]$.*

Definição 5.2.2 *Seja A um conjunto convexo em uma variedade de Hadamard M e $f : A \rightarrow \mathbb{R}$ uma função de valor real. f é chamada uma função convexa em A se para todo $x, y \in A$ e $t \in [0, 1]$*

$$f(\gamma(t)) \leq tf(y) + (1 - t)f(x),$$

onde $\gamma : [0, 1] \rightarrow M$ é a curva geodésica tal que $\gamma(0) = x$ e $\gamma(1) = y$. Quando a desigualdade anterior é estrita para $x \neq y$ e $t \in (0, 1)$, a função f é chamada estritamente convexa.

Teorema 5.2.1 *Seja M uma variedade de Hadamard e A um conjunto convexo de M . A função $f : A \rightarrow \mathbb{R}$ é convexa se, e somente se, $\forall x, y \in A$ e $\gamma : [0, 1] \rightarrow M$ (a geodésica ligando x e y), a função $f(\gamma(t))$ é convexa em $[0, 1]$.*

Demonstração. Ver Udriste [82], página 61, Teorema 2.2. ■

Uma função $f : A \rightarrow \mathbb{R}$ é chamada côncava se $-f$ é convexa. Além disso, se f é convexa e côncava então f é dita afim sobre A . Pode-se provar que uma função f duas vezes diferenciável em um conjunto convexo aberto A é afim se, e somente se, $\langle H_x^f(v), v \rangle_x = 0$, para todo $x \in A$ e $v \in T_x M$. De fato, $\langle H_x^f(v), v \rangle_x = 0$, se e somente se $\langle H_x^f(v), v \rangle_x \geq 0$ e $\langle H_x^f(v), v \rangle_x \leq 0$. Em outras palavras, f é afim se, e somente se, o campo $\text{grad} f$ é paralelo.

Proposição 5.2.1 *Seja M uma variedade de Hadamard e $h : M \rightarrow \mathbb{R}$ uma função diferenciável. Seja $y \in M$ e defina $g : M \rightarrow \mathbb{R}$ tal que*

$$g(x) = \langle \text{grad} h(y), \exp_y^{-1} x \rangle_y,$$

para $x \in M$. Então, os seguintes fatos se cumprem:

i. $\text{grad} g(x) = P_{\gamma,0,1} \text{grad} h(y)$, onde $\gamma : [0, 1] \rightarrow M$ é a geodésica tal que $\gamma(0) = y$ e $\gamma(1) = x$.

ii. g é afim sobre M .

Demonstração.

i. Seja $v \in T_x M$ (arbitrário). Considere a variação da geodésica $\alpha : [0, 1] \times (-\epsilon, \epsilon) \rightarrow M$ tal que

$$\alpha(t, s) = \exp_y(tu(s)),$$

onde $u(s) = \exp_y^{-1} x + sP_{\gamma,1,0} v$. Então

$$\begin{aligned} (dg)_x v &= \frac{d}{ds} (g(\alpha(1, s))) \Big|_{s=0} \\ &= \frac{d}{ds} \langle \text{grad} h(y), u(s) \rangle \Big|_{s=0} \\ &= \langle \text{grad} h(y), u'(0) \rangle \\ &= \langle \text{grad} h(y), P_{\gamma,1,0} v \rangle \\ &= \langle P_{\gamma,0,1} \text{grad} h(y), v \rangle. \end{aligned}$$

Portanto,

$$\text{grad} g(x) = P_{\gamma,0,1} \text{grad} h(y).$$

ii. O resultado segue de i. ■

Definição 5.2.3 *Seja M uma variedade de Hadamard e seja $f : M \rightarrow \mathbb{R}$ uma função convexa. Dado $y \in M$, o vetor $s \in T_y M$ é dito subgradiente de f em y se*

$$f(x) \geq f(y) + \langle s, \exp_y^{-1} x \rangle_y, \tag{5.3}$$

para todo $x \in M$. O conjunto de todos os subgradientes de f em y é chamado subdiferencial de f em y e é denotado por $\partial f(y)$.

Teorema 5.2.2 *Seja M uma variedade de Hadamard e seja $f : M \rightarrow \mathbb{R}$ uma função convexa. Para cada $y \in M$, existe $s \in T_y M$ tal que $\forall x \in M$ é satisfeita (5.3).*

Demonstração. Ver Ferreira e Oliveira [25], Teorema 3.3. ■

Do Teorema anterior o subdiferencial $\partial f(x)$ de uma função convexa f em $x \in M$ é não vazio.

Teorema 5.2.3 *Seja M uma variedade de Hadamard e $f : M \rightarrow \mathbb{R}$ uma função convexa. $0 \in \partial f(x)$ se e somente se x é um ponto de mínimo de f em M .*

Demonstração. Imediato. ■

Definição 5.2.4 *Seja A um conjunto convexo em uma variedade de Hadamard M e $f : A \rightarrow \mathbb{R}$ uma função real. f é chamada quase-convexa em A se para todo $x, y \in A$, $t \in [0, 1]$, se cumpre que*

$$f(\gamma(t)) \leq \max\{f(x), f(y)\},$$

onde $\gamma : [0, 1] \rightarrow \mathbb{R}$ é a geodésica tal que $\gamma(0) = x$ e $\gamma(1) = y$.

Teorema 5.2.4 *Seja A um conjunto convexo em uma variedade de Hadamard M . A função $f : A \rightarrow \mathbb{R}$ é quase-convexa se, e somente se, o conjunto $\{x \in A : f(x) \leq c\}$ é convexo para cada $c \in \mathbb{R}$.*

Demonstração. Ver [82], página 98, Teorema 10.2. ■

Definição 5.2.5 *Seja A um conjunto convexo em uma variedade de Hadamard M e $f : A \rightarrow \mathbb{R}$ uma função real. A função f é dita fortemente quase-convexa se para cada $x, y \in A$, com $x \neq y$ temos*

$$f(\gamma(t)) < \max\{f(x), f(y)\},$$

$\forall t \in (0, 1)$, onde $\gamma : [0, 1] \rightarrow \mathbb{R}$ é a curva geodésica tal que $\gamma(0) = x$ e $\gamma(1) = y$.

Observação 5.2.1 *Observemos, da prévia definição, que toda função estritamente convexa é fortemente quase-convexa.*

Teorema 5.2.5 *Seja $f : A \rightarrow \mathbb{R}$ uma função fortemente quase-convexa. Considere o problema $\min f(x), x \in A$, onde A é um conjunto convexo não vazio em M . Se \bar{x} é um ótimo local, então \bar{x} é o único mínimo global do problema.*

Proof. Desde que \bar{x} é uma solução ótima local, então existe $B(\bar{x}, \epsilon) := \{z \in M : d(z, \bar{x}) < \epsilon\}$ tal que $f(\bar{x}) \leq f(x)$, para todo $x \in A \cap B(\bar{x}, \epsilon)$. Suponha, por contradição, que existe um ponto $x^* \in A$ tal que $x^* \neq \bar{x}$ e $f(x^*) \leq f(\bar{x})$. Pela forte quase-convexidade de f , segue-se que

$$f(\gamma(t)) < \max\{f(x^*), f(\bar{x})\} = f(\bar{x}),$$

$\forall t \in (0, 1)$, onde $\gamma : [0, 1] \rightarrow \mathbb{R}$ é a curva geodésica tal que $\gamma(0) = \bar{x}$ e $\gamma(1) = x^*$.

Para t suficientemente pequeno, $\gamma(t) \in A \cap B(\bar{x}, \epsilon)$, por isso, a desigualdade estrita acima contradiz a otimalidade local de \bar{x} . Isto completa a demonstração. ■

Teorema 5.2.6 *Seja C um conjunto convexo fechado em uma variedade de Hadamard M . Seja $y \in M$ um ponto fixo arbitrário, então existe uma única projeção $z = P_C y$. Além disso, a seguinte desigualdade é satisfeita*

$$\langle \exp_z^{-1} y, \exp_z^{-1} x \rangle \leq 0, \tag{5.4}$$

para todo $x \in C$.

Demonstração. Ver em [25], Proposições 3.1 e 3.2. ■

5.3 Distâncias e Funções de Bregman em Variedades de Hadamard

Para construir algoritmos de ponto proximal com distâncias de Bregman para resolver problemas de otimização em variedades de Hadamard, é necessário estender as definições de distâncias e funções de Bregman nesta estrutura. Estenderemos a definição dada por Censor e Lent [13].

Seja M uma variedade de Hadamard e S um conjunto não vazio, aberto e convexo em M com fecho topológico denotado por \bar{S} . Seja $h : M \rightarrow \mathbb{R}$ uma função estritamente convexa sobre

\bar{S} e diferenciável em S . A *distância de Bregman* associada a h , denotada por D_h , é definida como uma função $D_h(\cdot, \cdot) : \bar{S} \times S \rightarrow \mathbb{R}$ tal que

$$D_h(x, y) := h(x) - h(y) - \langle \text{grad}h(y), \exp_y^{-1} x \rangle_y. \quad (5.5)$$

Observemos que D_h não é uma distância no sentido usual da definição. Em geral, a desigualdade triangular assim como a simetria não são válidos. Notemos também que a expressão da distância de Bregman depende da definição da métrica Riemanniana. Alguns exemplos destas distâncias em variedades de Hadamard serão apresentados na Seção 5.8. Daremos a seguir a seguinte notação para os conjuntos (parciais) de nível de D_h .

$$\Gamma_1(\alpha, y) := \{x \in \bar{S} : D_h(x, y) \leq \alpha\},$$

$$\Gamma_2(x, \alpha) := \{y \in S : D_h(x, y) \leq \alpha\},$$

onde $\alpha \in \mathbb{R}$.

Definição 5.3.1 *Seja M uma variedade de Hadamard. Uma função $h : M \rightarrow \mathbb{R}$ é chamada uma função de Bregman se existe um conjunto S não vazio, aberto e convexo tal que:*

- a.** h é contínua sobre \bar{S} ;
- b.** h é estritamente convexa sobre \bar{S} ;
- c.** h é continuamente diferenciável em S ;
- d.** Para todo $\alpha \in \mathbb{R}$ os conjuntos de nível $\Gamma_1(\alpha, y)$ e $\Gamma_2(x, \alpha)$ são limitados para todo $y \in S$ e $x \in \bar{S}$, respectivamente;
- e.** Se $\lim_{k \rightarrow \infty} y^k = y^* \in \bar{S}$, então $\lim_{k \rightarrow \infty} D_h(y^*, y^k) = 0$;
- f.** Se $\lim_{k \rightarrow \infty} D_h(z^k, y^k) = 0$, $\lim_{k \rightarrow \infty} y^k = y^* \in \bar{S}$ e $\{z^k\}$ é limitada, então $\lim_{k \rightarrow \infty} z^k = y^*$.

Denotaremos a família de funções de Bregman por \mathcal{B} e chamaremos ao conjunto S como a *zona da função h* .

Lemma 5.3.1 *Seja $h \in \mathcal{B}$ com zona S . Então,*

- i. $\text{grad } D_h(\cdot, y)(x) = \text{grad } h(x) - P_{\gamma,0,1} \text{grad } h(y)$, para todo $x, y \in S$, onde $\gamma : [0, 1] \rightarrow M$ é uma curva geodésica tal que $\gamma(0) = y$ e $\gamma(1) = x$.
- ii. Para cada $y \in S$, $D_h(\cdot, y)$ é estritamente convexa sobre \bar{S} .
- iii. Para todo $x \in \bar{S}$ e $y \in S$, $D_h(x, y) \geq 0$ e $D_h(x, y) = 0$ se, e somente se, $x = y$.

Demonstração.

- i. Da Proposição 5.2.1, i, obtemos o resultado.
- ii Como h é estritamente convexa em \bar{S} e $\langle \text{grad } h(y), \exp_y^{-1} x \rangle_y$, é uma função linear afim (Proposition 5.2.1, ii) o resultado é estabelecido.
- iii. Use, de novo, a estrita convexidade de h . ■

A partir de agora, usaremos $\text{grad } D_h(x, y)$ para denotar $\text{grad } D_h(\cdot, y)(x)$. Por isso, usando o Lemma 5.3.1, i, temos que

$$\text{grad } D_h(x, y) = \text{grad } h(x) - P_{\gamma,0,1} \text{grad } h(y).$$

onde γ é a curva geodésica tal que $\gamma(0) = y$ e $\gamma(1) = x$.

Definição 5.3.2 Seja $\Omega \subset M$ e S um conjunto convexo aberto e seja $y \in S$. Um ponto $Py \in \Omega \cap \bar{S}$ que satisfaz

$$D_h(Py, y) = \min_{x \in \Omega \cap \bar{S}} D_h(x, y), \quad (5.6)$$

é chamado D_h -projeção do ponto y sobre o conjunto Ω .

O seguinte Lema estabelece a existência e unicidade da D_h -projeção para uma função de Bregman considerando uma hipótese sobre Ω .

Lemma 5.3.2 Seja Ω um conjunto convexo fechado em M e $h \in \mathcal{B}$ com zona S . Se $\Omega \cap \bar{S} \neq \emptyset$, então para cada $y \in S$, existe uma única D_h -projeção Py sobre Ω .

Demonstração. Para cada $x \in \Omega \cap \bar{S}$, o conjunto

$$B := \{z \in \bar{S} : D_h(z, y) \leq D_h(x, y)\}$$

é limitado (da Definição 5.3.1, d) e fechado (porque $D_h(\cdot, y)$ é contínua em \bar{S} , devido à Definição 5.3.1, a). Portanto, o conjunto

$$T := (\Omega \cap \bar{S}) \cap B$$

é não vazio, já que $x \in B \cap \Omega$, e limitado. Como a interseção de conjuntos fechados é fechado, então T é também fechado, daqui compacto. Conseqüentemente, $D_h(\cdot, y)$ atinge um valor de mínimo sobre o conjunto compacto T . Seja $x^* \in T$ um ponto de mínimo de $D_h(\cdot, y)$. Para todo $z \in \Omega \cap \bar{S}$ tal que $z \notin B$ temos

$$D_h(x^*, y) < D_h(z, y);$$

daqui, x^* satisfaz (5.6) e devido à estrita convexidade de $D_h(\cdot, y)$ temos que x^* é único e portanto,

$$x^* = Py. \quad \blacksquare$$

Lemma 5.3.3 *Seja $h \in \mathcal{B}$ com zona S e $y \in S$. Suponha que $Py \in S$, onde Py é a D_h -projeção sobre algum conjunto convexo e fechado Ω tal que $\Omega \cap \bar{S} \neq \emptyset$. Então, a função*

$$g(x) := D_h(x, y) - D_h(x, Py)$$

é linear afim sobre \bar{S} .

Demonstração. De (5.5)

$$D_h(x, y) - D_h(x, Py) = h(Py) - h(y) + \langle \text{grad}h(Py), \exp_{Py}^{-1}x \rangle_{Py} - \langle \text{grad}h(y), \exp_y^{-1}x \rangle_y.$$

Devido à linearidade afim das funções $\langle \text{grad}h(Py), \exp_{Py}^{-1}x \rangle_{Py}$ e $\langle \text{grad}h(y), \exp_y^{-1}x \rangle_y$ em x , o resultado é obtido. \blacksquare

Proposição 5.3.1 *Seja $h \in \mathcal{B}$ com zona S , $\Omega \subset M$ é algum conjunto convexo fechado dado tal que $\Omega \cap \bar{S} \neq \emptyset$. Seja $y \in S$ e assume-se que $Py \in S$, onde Py denota a D_h -projeção de y sobre Ω . Para qualquer $x \in \Omega \cap \bar{S}$ a seguinte desigualdade*

$$D_h(Py, y) \leq D_h(x, y) - D_h(x, Py)$$

é satisfeita.

Demonstração. Seja $\gamma : [0, 1] \rightarrow M$ uma curva geodésica tal que $\gamma(0) = Py$ e $\gamma(1) = x$. Devido ao Lema 5.3.3 a função

$$G(x) = D_h(x, y) - D_h(x, Py)$$

é convexa sobre \bar{S} . Então, em particular $G(\gamma(t))$ é convexa em $t \in (0, 1)$ (ver Teorema 5.2.1).

Assim,

$$G(\gamma(t)) \leq tG(x) + (1-t)G(Py),$$

daqui,

$$D_h(\gamma(t), y) - D_h(\gamma(t), Py) \leq t(D_h(x, y) - D_h(x, Py)) + D_h(Py, y) - tD_h(Py, y),$$

onde usamos o fato de que $D_h(Py, Py) = 0$. A desigualdade acima é equivalente a

$$(1/t)(D_h(\gamma(t), y) - D_h(Py, y)) - (1/t)D_h(\gamma(t), Py) \leq D_h(x, y) - D_h(x, Py) - D_h(Py, y). \quad (5.7)$$

Como $\Omega \cap \bar{S}$ é convexo e $x, Py \in \Omega \cap \bar{S}$ então $\gamma(t) \in \Omega \cap \bar{S}$ para todo $t \in (0, 1)$. Usando o fato de que Py é a projeção temos que

$$(1/t)(D_h(\gamma(t), y) - D_h(Py, y)) \geq 0.$$

Usando esta desigualdade em (5.7) obtemos

$$-(1/t)D_h(\gamma(t), Py) \leq D_h(x, y) - D_h(x, Py) - D_h(Py, y).$$

Agora, como $D_h(\cdot, z)$ é diferenciável para todo $z \in S$, podemos tomar limite em t obtendo

$$-\langle \text{grad}D_h(Py, Py), \exp_{Py}^{-1} x \rangle_{Py} \leq D_h(x, y) - D_h(x, Py) - D_h(Py, y).$$

O lado direito da desigualdade é zero, obtendo assim o resultado desejado. ■

5.4 Regularização

Seja M uma variedade de Hadamard e $f : X \subset M \rightarrow \mathbb{R}$ uma função de valor real. Seja S um conjunto aberto e convexo e $h : \bar{S} \rightarrow \mathbb{R}$ uma função diferenciável em S . Para $\lambda > 0$, a regularização de Moreau-Yosida $f_\lambda : S \rightarrow \mathbb{R}$ de f é definida por

$$f_\lambda(y) = \inf_{x \in X \cap \bar{S}} \{f(x) + \lambda D_h(x, y)\} \quad (5.8)$$

onde $D_h(x, y)$ é dada em (5.5). Para que a função (5.8) esteja bem definida, h and f devem satisfazer algumas condições.

Proposição 5.4.1 *Se $f : X \subset M \rightarrow \mathbb{R}$ é uma função contínua e limitada inferiormente sobre um conjunto convexo fechado X e $h \in \mathcal{B}$ com zona S tal que $X \cap \bar{S} \neq \emptyset$, então para todo $y \in S$ e $\lambda > 0$ existe um ponto, denotado por $x_f(y, \lambda)$, tal que*

$$f_\lambda(y) = f(x_f(y, \lambda)) + D_h(x_f(y, \lambda), y). \quad (5.9)$$

Além disso, a unicidade de $x_f(y, \lambda)$ é garantida se f é tal que $f(\cdot) + \lambda D_h(\cdot, y)$ é fortemente quase-convexa.

Demonstração. Seja β uma cota inferior para f sobre X , então para todo $x \in X \cap \bar{S}$:

$$f(x) + \lambda D_h(x, y) \geq \beta + \lambda D_h(x, y),$$

Segue-se da Definição 5.3.1, d, que os conjuntos de nível da função $(f(\cdot) + \lambda D_h(\cdot, y))$ são limitados. Além disto, esta função é contínua em $X \cap \bar{S}$, devido à Definição 5.3.1, a. Por isso os conjuntos de nível de $(f(\cdot) + \lambda D_h(\cdot, y))$ são fechados, logo compactos. Agora, usando argumentos de continuidade e compacidade, $(f(\cdot) + \lambda D_h(\cdot, y))$ atinge um mínimo sobre $X \cap \bar{S}$. Finalmente se f é tal que $(f(\cdot) + \lambda D_h(\cdot, y))$ é fortemente quase-convexa, para $\lambda > 0$, então do Teorema 5.2.5, $x_f(y, \lambda)$ único. ■

Para estudar a boa definição da função (5.8) sobre outras condições para h , consideremos o caso irrestrito $X = S = M$, isto é, quando (5.8) é reduzido para

$$f_\lambda(y) = \inf_{x \in M} \{f(x) + \lambda D_h(x, y)\}.$$

Para isso introduzimos a seguinte definição.

Definição 5.4.1 *Uma função $g : M \rightarrow \mathbb{R}$ é dita 1-coerciva em $y \in M$ se*

$$\lim_{d(x,y) \rightarrow \infty} \frac{g(x)}{d(x,y)} = +\infty.$$

Notemos que, se $g : M \rightarrow \mathbb{R}$ é uma função contínua e 1-coerciva em $y \in M$, então é fácil verificar que o conjunto de minimizadores de g em M é não vazio.

Lemma 5.4.1 Se $f : M \rightarrow \mathbb{R}$ é limitada inferiormente, $\lambda > 0$ e $h : M \rightarrow \mathbb{R}$ é uma função diferenciável e 1-coerciva em $y \in M$, então a função $(f(\cdot) + \lambda D_h(\cdot, y)) : M \rightarrow \mathbb{R}$ é 1-coerciva em $y \in M$.

Demonstração. Seja β um limite inferior para f . Então:

$$\begin{aligned} \frac{f(x) + \lambda D_h(x, y)}{d(x, y)} &\geq \frac{\beta}{d(x, y)} + \lambda \frac{D_h(x, y)}{d(x, y)} \\ &= \frac{\beta}{d(x, y)} + \lambda \frac{h(x)}{d(x, y)} - \lambda \frac{h(y)}{d(x, y)} - \lambda \left\langle \text{grad}h(y), \frac{\exp_y^{-1} x}{d(x, y)} \right\rangle_y \\ &\geq \frac{\beta}{d(x, y)} + \lambda \frac{h(x)}{d(x, y)} - \lambda \frac{h(y)}{d(x, y)} - \lambda \|\text{grad}h(y)\|, \end{aligned}$$

onde a igualdade vem da definição de D_h , e a última desigualdade é obtida aplicando a desigualdade de Cauchy, e do fato de que $\|\exp_y^{-1} x\| = d(x, y)$. Tomando $d(x, y) \rightarrow +\infty$ e usando a 1-coercividade de h em y obtemos

$$\lim_{d(x, y) \rightarrow +\infty} \frac{(f(\cdot) + \lambda D_h(\cdot, y))(x)}{d(x, y)} = +\infty. \quad \blacksquare$$

Proposição 5.4.2 Seja $h : M \rightarrow \mathbb{R}$ uma função diferenciável e 1-coerciva em $y \in M$, $f : M \rightarrow \mathbb{R}$ uma função contínua e limitada inferiormente. Então existe um ponto $x_f(y, \lambda)$ tal que

$$f_\lambda(y) = f(x_f(y, \lambda)) + \lambda D_h(x_f(y, \lambda), y).$$

Além disso, a unicidade de $x_f(y, \lambda)$ é garantida se f é tal que $f(\cdot) + \lambda D_h(\cdot, y)$ é fortemente quase-convexa.

Demonstração. O resultado segue do Lema 5.4.1 e da observação da Definição 5.4.1. \blacksquare

5.5 Método de Ponto Proximal

Consideremos o seguinte problema de otimização

$$(p) \min_{x \in M} f(x)$$

onde M é uma variedade de Hadamard. Os principais resultados serão dados quando f é uma função contínua quase-convexa ou convexa em M . O método PDB gera uma seqüência de pontos $\{x^k\}$ definidos por

$$x^0 \in M, \quad (5.10)$$

$$x^k \in \arg \min_{x \in M} \{f(x) + \lambda_k D_h(x, x^{k-1})\}, \quad (5.11)$$

onde h é uma função de Bregman com zona M , D_h é dado por (5.5) e λ_k é um parâmetro positivo.

Observe que si f é contínua e limitada inferiormente, a iteração acima existe, ver Proposição 5.4.1.

No caso particular em que M é o espaço Euclidiano \mathbb{R}^n e $h(x) = (1/2)x^T x$ temos

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

Portanto, o método PDB é uma natural generalização do método de ponto proximal em \mathbb{R}^n .

Usaremos os seguintes parâmetros de regularização para o método PDB:

$$0 < \lambda_k < \bar{\lambda}, \quad \text{or} \quad (5.12)$$

$$\lim_{k \rightarrow +\infty} \lambda_k = 0, \quad \text{with } \lambda_k > 0. \quad (5.13)$$

Note que (5.12) implica que $\sum_{k=1}^{\infty} (1/\lambda_k) = +\infty$. Observe que esta expressão é a condição mínima em métodos de ponto proximal em espaços Euclidianos e espaços de Banach.

A hipótese que usaremos para resolver o problema (p) é a seguinte:

Hipótese A. O conjunto de soluções ótimas do problema (p) , denotado por X^* , é não vazio.

Observação 5.5.1 Para verificar que $h : M \rightarrow \mathbb{R}$ é uma função de Bregman, com zona M , é suficiente mostrar que as condições **a** – **d** na Definição 5.3.1 sejam satisfeitas, já que **e** e **f** são conseqüências imediatas de **a**, **b**, **c** e **d**.

5.6 Resultados de Convergência

Nesta seção, com as hipóteses dadas na seção anterior, provaremos a convergência do método proposto. Nossos resultados são motivados pelos trabalhos de Iusem [44], Chen e Teboulle [14] e Censor e Zenios [12].

5.6.1 O caso quase-convexo

Teorema 5.6.1 *Se a hipóteses A é satisfeita e f é uma função quase-convexa e contínua então, a seqüência $\{x^k\}$, gerada pelo método PDB, é limitada.*

Demonstração. Desde que x^k satisfaz (5.11) temos

$$f(x^k) + \lambda_k D_h(x^k, x^{k-1}) \leq f(x) + \lambda_k D_h(x, x^{k-1}), \quad \forall x \in M. \quad (5.14)$$

Daqui, $\forall x \in M$ tal que $f(x) \leq f(x^k)$ cumpre-se

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

Potanto, x^k é a única D_h -projecção de x^{k-1} sobre o conjunto convexo fechado

$$\Omega := \{x \in M : f(x) \leq f(x^k)\}.$$

Usando a Proposição 5.3.1 e o fato de que $X^* \subset \Omega$ obtemos

$$0 \leq D_h(x^k, x^{k-1}) \leq D_h(x^*, x^{k-1}) - D_h(x^*, x^k), \quad (5.15)$$

para todo $x^* \in X^*$. Assim,

$$D_h(x^*, x^k) \leq D_h(x^*, x^{k-1}). \quad (5.16)$$

Isto significa que $\{x^k\}$ é D_h -Fejér monótono relativamente ao conjunto X^* . Finalmente, aplicando a definição 5.3.1, d, temos que $\{x^k\}$ é limitado, já que

$$x^k \in \Gamma_2(x^*, \alpha),$$

onde $\alpha = D_h(x^*, x^0)$. ■

Proposição 5.6.1 *Com as mesmas hipóteses do teorema anterior, os seguintes fatos são verdadeiros:*

a. Para todo $x^* \in X^*$ a seqüência $\{D_h(x^*, x^k)\}$ é convergente;

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

c. $\{f(x^k)\}$ é não crescente;

d. Se $\lim_{j \rightarrow +\infty} x^{k_j} = \bar{x}$ então, $\lim_{j \rightarrow +\infty} x^{k_j+1} = \bar{x}$.

Demonstração.

a. De (5.16), $\{D_h(x^*, x^k)\}$ é limitada inferiormente e não crescente, portanto convergente.

b. Considerando o limite quando k vai para infinito em (5.15) e usando o resultado anterior obtemos $\lim_{k \rightarrow \infty} D_h(x^k, x^{k-1}) = 0$.

c. Considerando $x = x^{k-1}$ em (5.14) segue-se que

$$0 \leq D_h(x^k, x^{k-1}) \leq (1/\lambda_k)(f(x^{k-1}) - f(x^k)), \quad (5.17)$$

desde que $D_h(x^{k-1}, x^{k-1}) = 0$. Assim, $\{f(x^k)\}$ é não crescente.

d. Tomando $z^k = x^{k_j+1}$ e $y^k = x^{k_j}$ na Definição 5.3.1, f, obtemos o resultado. ■

Teorema 5.6.2 *Se a hipótese A é satisfeita e f é uma função quase-convexa e contínua, então qualquer ponto de acumulação de $\{x^k\}$, gerado pelo método PDB com λ_k satisfazendo (5.13) é uma solução ótima do problema (p).*

Demonstração. Sejam $x^* \in X^*$ uma solução ótima de (p) e $\bar{x} \in M$ um ponto de acumulação de $\{x^k\}$; então existe uma subsequência $\{x^{k_j}\}$ tal que

$$\lim_{j \rightarrow \infty} x^{k_j} = \bar{x}.$$

Devido a que x^{k_j} é solução de (5.11) temos

$$f(x^{k_j}) + \lambda_{k_j} D_h(x^{k_j}, x^{k_j-1}) \leq f(x^*) + \lambda_{k_j} D_h(x^*, x^{k_j-1}).$$

Isto implica

$$\lambda_{k_j} (D_h(x^{k_j}, x^{k_j-1}) - D_h(x^*, x^{k_j-1})) \leq f(x^*) - f(x^{k_j}).$$

Usando a caracterização de funções convexas diferenciáveis para $D_h(\cdot, x^{k_j-1})$ obtemos:

$$\lambda_{k_j} \langle \text{grad} D_h(x^*, x^{k_j-1}), \exp_{x^*}^{-1} x^{k_j} \rangle_{x^*} \leq f(x^*) - f(x^{k_j}).$$

Fazendo $j \rightarrow \infty$, usando a continuidade das funções $\text{grad}D_h(x^*, \cdot)$ e $\exp_{x^*}^{-1}$ e como λ_k satisfaz (5.13) obtemos

$$f(\bar{x}) \leq f(x^*).$$

Assim, qualquer ponto de acumulação de $\{x^k\}$ é uma solução ótima do problema (p). ■

Teorema 5.6.3 *Se a hipótese A é satisfeita e f é uma função quase-convexa e contínua, então a seqüência $\{x^k\}$ gerada pelo método PDB, com λ_k satisfazendo (5.13), converge a uma solução ótima do problema (p).*

Demonstração. Do Teorema 5.6.1 $\{x^k\}$ é limitada, por isso existe uma subsequência convergente. Seja $\{x^{k_j}\}$ uma subsequência de $\{x^k\}$ tal que $\lim_{j \rightarrow \infty} x^{k_j} = x^*$. Da Definição 5.3.1, e, é verdade que

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

Do Teorema 5.6.2 temos que x^* é uma solução ótima de (p), por isso, da Proposição 5.6.1, a, $D_h(x^*, x^k)$ é uma seqüência convergente, com uma subsequência convergindo para 0. Consequentemente toda a seqüência converge para 0, isto é,

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

Para provar que $\{x^k\}$ têm um único ponto de acumulação, seja $\bar{x} \in X^*$ um outro ponto de acumulação de $\{x^k\}$. Então $\lim_{l \rightarrow \infty} D_h(x^*, x^{k_l}) = 0$ com $\lim_{l \rightarrow \infty} x^{k_l} = \bar{x}$. Logo, da Definição 5.3.1, f, $x^* = \bar{x}$. Segue-se que $\{x^k\}$ não pode ter mais de um ponto de acumulação e portanto,

$$\lim_{k \rightarrow +\infty} x^k = x^* \in X^*. \quad \blacksquare$$

5.6.2 O caso convexo

Teorema 5.6.4 *Suponhamos que a hipótese A é satisfeita e que $f : M \rightarrow \mathbb{R}$ é uma função convexa. Se λ_k satisfaz (5.12), então qualquer ponto de acumulação de $\{x^k\}$ é uma solução ótima do problema (p).*

Demonstração. Seja $\bar{x} \in M$ um ponto de acumulação de $\{x^k\}$. Logo existe uma subsequência $\{x^{k_j}\}$ tal que

$$\lim_{j \rightarrow \infty} x^{k_j} = \bar{x}.$$

De (5.11) e Teorema 5.2.3 temos

$$0 \in \partial[f(\cdot) + \lambda_{k_j} D_h(\cdot, x^{k_j})](x^{k_j+1}),$$

equivalentemente,

$$-\lambda_{k_j+1} \text{grad} D_h(x^{k_j+1}, x^{k_j}) \in \partial f(x^{k_j+1}).$$

Seja γ_{k_j} a curva geodésica tal que $\gamma_{k_j}(0) = x^{k_j}$ and $\gamma_{k_j}(1) = x^{k_j+1}$. Pelo Lema 5.3.1, i, obtemos

$$\lambda_{k_j+1} [P_{\gamma_{k_j}, 0, 1} \text{grad} h(x^{k_j}) - \text{grad} h(x^{k_j+1})] \in \partial f(x^{k_j+1}).$$

Seja x^* uma solução ótima do problema (p). Usando (5.3) para $x = x^*$ e $y = x^{k_j+1}$ temos

$$f(x^*) - f(x^{k_j+1}) \geq \langle y^{k_j}, \exp_{x^{k_j+1}}^{-1} x^* \rangle_{x^{k_j+1}} \quad (5.18)$$

onde,

$$y^{k_j} := \lambda_{k_j+1} [P_{\gamma_{k_j}, 0, 1} \text{grad} h(x^{k_j}) - \text{grad} h(x^{k_j+1})].$$

Por outro lado, da desigualdade de Cauchy

$$|\langle y^{k_j}, \exp_{x^{k_j+1}}^{-1} x^* \rangle_{x^{k_j+1}}| \leq \|y^{k_j}\|_{x^{k_j+1}} \|\exp_{x^{k_j+1}}^{-1} x^*\|_{x^{k_j+1}}.$$

Como $\|\exp_{x^{k_j+1}}^{-1} x^*\|_{x^{k_j+1}} = d(x^*, x^{k_j+1})$, e do Teorema 5.6.1, existe $M > 0$ tal que

$$|\langle y^{k_j}, \exp_{x^{k_j+1}}^{-1} x^* \rangle_{x^{k_j+1}}| \leq M \|y^{k_j}\|_{x^{k_j+1}}.$$

Usando este fato na desigualdade (5.18) obtemos

$$f(x^*) - f(x^{k_j+1}) \geq -M \|y^{k_j}\|_{x^{k_j+1}}. \quad (5.19)$$

Para concluir a demonstração provaremos que

$$\lim_{j \rightarrow \infty} \|y^{k_j}\|_{x^{k_j+1}} = 0.$$

De fato, usando a continuidade do transporte paralelo, continuidade do campo gradiente, Proposição 5.6.1, d, e a limitação de λ_k obtemos que $\lim_{j \rightarrow \infty} \|y^{k_j}\|_{x^{k_j+1}} = 0$, como desejado.

Finalmente, fazendo $j \rightarrow +\infty$ em (5.19) e usando a continuidade de f obtemos que

$$f(x^*) \geq f(\bar{x}).$$

Portanto, qualquer ponto de acumulação de $\{x^k\}$ é uma solução ótima do problema (p). ■

Teorema 5.6.5 *Com a Hipótese A e a convexidade de f , a seqüência $\{x^k\}$ gerada pelo método PDB, com λ_k satisfazendo (5.13), converge para uma solução ótima do problema (p).*

Demonstração. A prova é análoga ao Teorema 5.6.3. ■

5.7 Métodos de Ponto Proximal com distâncias Riemannianas

Nesta seção adaptamos os resultados da seção anterior para o clássico método de ponto proximal em variedades de Hadamard, baseada na distância Riemanniana. Observe que nestas variedades a função distância é convexa (ver Sakai [78]). Nossos resultados são uma extensão do trabalho de Ferreira e Oliveira [25], onde consideraram o caso convexo.

Consideremos inicialmente a regularização de Moreau-Yosida. Para $\lambda > 0$, seja:

$$\varphi_\lambda(y) = \inf_{x \in M} \{f(x) + \lambda d^2(x, y)\}$$

temos o seguinte resultado:

Proposição 5.7.1 *Se $f : M \rightarrow \mathbb{R}$ é uma função contínua e limitada inferiormente então, para cada $y \in M$ e $\lambda > 0$, existe um ponto, denotado por $x(y, \lambda)$, tal que*

$$\varphi_\lambda(y) = f(x(y, \lambda)) + \lambda d^2(x(y, \lambda), y).$$

Demonstração. Claramente, a função $d^2(., .)$ é 1-coerciva. Portanto o Lema 5.4.1 e a Proposição 5.4.2 são facilmente adaptáveis. ■

Agora apresentamos os resultados de convergência para o método PPA, definido por

$$x^0 \in M, \tag{5.20}$$

$$x^k \in \arg \min_{x \in M} \{f(x) + (\lambda_k/2)d^2(x, x^{k-1})\}, \tag{5.21}$$

Observe, analogamente ao método PBD, que se f é contínua e limitada inferiormente, a iteração acima existe.

5.7.1 O caso quase-convexo

Teorema 5.7.1 *Se a hipótese A é satisfeita e f é uma função contínua e quase-convexa, então a seqüência $\{x^k\}$ gerada pelo método PPA, (5.20)-(5.21), é limitada.*

Demonstração. Considerando o Teorema 5.6.1, e substituindo $D_h(\cdot, \cdot)$ por $d^2(\cdot, \cdot)$, observamos que todos os passos da prova podem ser obtidos, o único ponto que falta justificar é a desigualdade (5.15), o qual é para o nosso caso

$$0 \leq d^2(x^k, x^{k-1}) \leq d^2(x^*, x^{k-1}) - d^2(x^*, x^k) \quad (5.22)$$

para todo $x^* \in X^*$. De fato, isto é uma conseqüência do Teorema 2.2.2 e do Teorema 5.2.6: seja em (2.5), $x = x^*$, $y = x^{k-1}$ e $z = x^k$, e usando (5.4) obtemos (5.22).

Proposição 5.7.2 *Com as hipóteses do teorema anterior, os seguintes fatos são verdadeiros*

- a. *Para todo $x^* \in X^*$ a seqüência $\{d^2(x^*, x^k)\}$ é convergente;*
- b. $\lim_{k \rightarrow +\infty} d^2(x^k, x^{k-1}) = 0$;
- c. $\{f(x^k)\}$ *é não crescente;*
- d. *Se $\lim_{j \rightarrow +\infty} x^{k_j} = \bar{x}$, então $\lim_{j \rightarrow +\infty} x^{k_j+1} = \bar{x}$.*

Demonstração. Para **a**, **b** e **c**, ver, respectivamente, a demonstração de **a**, **b** e **c**, da Proposição 5.6.1, com a óbvia substituição de $D_h(\cdot, \cdot)$ por $d^2(\cdot, \cdot)$. Para **d**, considere a propriedade da desigualdade triangular, aplicada à distância Riemanniana d , a qual dá, particularmente

$$d(x^{k_j+1}, \bar{x}) \leq d(x^{k_j+1}, x^{k_j}) + d(x^{k_j}, \bar{x}).$$

Tomando $j \rightarrow \infty$ e usando **b**, obtemos o resultado. ■

Teorema 5.7.2 *Se a hipótese A é satisfeita e f é uma função contínua e quase-convexa, então a seqüência $\{x^k\}$ gerada pelo método PPA, (5.20)-(5.21), com λ_k satisfazendo (5.13), é uma solução ótima de (p).*

Demonstração. A adaptação do Teorema 5.6.2 é imediata. ■

Teorema 5.7.3 *Se a hipótese A é satisfeita e f é uma função contínua e quase-convexa, então a seqüência $\{x^k\}$ gerada pelo método PPA, (5.20)-(5.21), com λ_k satisfazendo (5.13), converge a uma solução ótima de (p).*

Demonstração. Como no Teorema 5.6.3. ■

5.7.2 O caso convexo

Resultados de convergência para o caso convexo, usando λ_k tal que $\sum_{k=1}^n (1/\lambda_k) = +\infty$, foi provado em [25]. Aqui apresentamos uma estimativa para o resíduo $f(x^k) - f(x^*)$, onde x^* é uma solução ótima de (p).

Teorema 5.7.4 *Seja f uma função convexa, com a hipótese A e λ_k tal que $\sum_{k=1}^n (1/\lambda_k) = +\infty$. Então a seqüência $\{x^k\}$, gerada pelo método PPA, converge para uma solução ótima de (p), e satisfaz*

$$f(x^n) - f(x^*) \leq \frac{d(x^*, x^0)}{2 \sum_{k=1}^n (1/\lambda_k)}.$$

Demonstração. A prova de convergência foi apresentada por Ferreira e Oliveira [25]. Por isso, provaremos a segunda parte. De (5.21) temos

$$f(x^k) + (\lambda_k/2)d^2(x^k, x^{k-1}) \leq f(x) + (\lambda_k/2)d^2(x, x^{k-1}),$$

para todo $x \in M$. Fixando $x = x^{k-1}$ obtemos

$$(\lambda_k/2)d^2(x^k, x^{k-1}) \leq f(x^{k-1}) - f(x^k).$$

Defina $\sigma_k = (1/\lambda_k) + \sigma_{k-1}$ com $\sigma_0 = 0$. Da última desigualdade temos

$$\begin{aligned} (\lambda_k/2)\sigma_{k-1}d^2(x^k, x^{k-1}) &\leq \sigma_{k-1}(f(x^{k-1}) - f(x^k)) \\ &= \sigma_{k-1}f(x^{k-1}) - (\sigma_k - (1/\lambda_k))f(x^k) \\ &= \sigma_{k-1}f(x^{k-1}) - \sigma_k f(x^k) + (1/\lambda_k)f(x^k). \end{aligned}$$

Somando a soma de $k = 1$ até n e multiplicando por 2 temos

$$\sum_{k=1}^n \lambda_k \sigma_{k-1} d^2(x^k, x^{k-1}) \leq -2\sigma_n f(x^n) + \sum_{k=1}^n (2/\lambda_k) f(x^k). \quad (5.23)$$

Por outro lado, pode ser provado (ver [25], Lema 6.2) que

$$(2/\lambda_k)(f(x^k) - f(x)) \leq d^2(x, x^{k-1}) - d^2(x^k, x^{k-1}) - d^2(x, x^k).$$

Daqui,

$$\sum_{k=1}^n (2/\lambda_k)(f(x^k) - f(x)) \leq \sum_{k=1}^n (d^2(x, x^{k-1}) - d^2(x^k, x^{k-1}) - d^2(x, x^k)).$$

Isto implica que

$$\sum_{k=1}^n (2/\lambda_k)f(x^k) \leq 2\sigma_n f(x) + d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^n d^2(x^k, x^{k-1}).$$

De (5.23) temos

$$\sum_{k=1}^n \lambda_k \sigma_{k-1} d^2(x^k, x^{k-1}) + 2\sigma_n f(x^n) \leq 2\sigma_n f(x) + d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^n d^2(x^k, x^{k-1}).$$

Assim

$$2\sigma_n(f(x^n) - f(x)) \leq d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^n (1 + \lambda_k \sigma_{k-1}) d^2(x^k, x^{k-1}).$$

Como $1 + \lambda_k \sigma_{k-1} = \lambda_k \sigma_k$ então

$$2\sigma_n(f(x^n) - f(x)) \leq d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^n \lambda_k \sigma_k d^2(x^k, x^{k-1}).$$

Portanto

$$f(x^n) - f(x) \leq \frac{d^2(x, x^0)}{2\sigma_n}.$$

Fazendo $x^* \in X^*$ na desigualdade anterior, concluímos a demonstração. ■

5.8 Alguns Exemplos de Distâncias de Bregman

Os exemplos 5.1 até 5.4 são variedades Riemannianas com curvatura seccional nula. No exemplo 5.5, é negativa. Todas estas variedades são de Hadamard. Em todos os exemplos apresentamos a formulação geral da distância de Bregman, dependendo da função h , e, devido ao possível uso do método de ponto proximal clássico, as distâncias Riemannianas.

Exemplo 5.1 O espaço Euclidiano é uma variedade de Hadamard com a métrica $G(x) = I$.

As curvas geodésicas são as retas e as distâncias de Bregman têm a forma

$$D_h(x, y) = h(x) - h(y) - \sum_{k=1}^n (x_k - y_k) \frac{\partial h(y)}{\partial y_k}$$

A distância é dada por

$$d(x, y) = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2}.$$

Exemplo 5.2 Seja \mathbb{R}^n com a métrica

$$G(x) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & \dots & 1 & 0 & 0 & 0 \\ 0 & \dots & & 1 + 4x_{n-1}^2 & -2x_{n-1} & \\ 0 & 0 & & -2x_{n-1} & 1 & \end{bmatrix}$$

Assim, $(\mathbb{R}^n, G(x))$ é uma variedade de Hadamard isométrica a (\mathbb{R}^n, I) pela aplicação $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ definida por $\Phi(x) = (x_1, x_2, \dots, x_{n-1}, x_{n-1}^2 - x_n)$, ver [17]. A única curva geodésica, ligando os pontos $\gamma(0) = y$ e $\gamma(1) = x$ é $\gamma(t) = (\gamma_1, \dots, \gamma_n)$, onde $\gamma_i(t) = y_i + t(x_i - y_i), \forall i = 1, \dots, n-1$ e $\gamma_n(t) = y_n + t((x_n - y_n) - 2(x_{n-1} - y_{n-1})^2) + 2t^2(x_{n-1} - y_{n-1})^2$.

Então, a distância de Bregman é:

$$D_h(x, y) = h(x) - h(y) - \sum_{k=1}^n (x_k - y_k) \frac{\partial h(y)}{\partial y_k} + 2 \frac{\partial h(y)}{\partial y_n} (x_n - y_n).$$

A distância Riemanniana, ver [17], é:

$$d(x, y) = \left[\sum_{i=1}^{n-1} (x_i - y_i)^2 + (x_{n-1}^2 - x_n - y_{n-1}^2 + y_n)^2 \right]^{1/2}$$

Exemplo 5.3 $M = \mathbb{R}_{++}^n$ com a métrica de Dikin X^{-2} , é uma variedade de Hadamard.

Definindo $\pi : M \rightarrow \mathbb{R}^n$ tal que $\pi(x) = (-\ln x_1, \dots, -\ln x_n)$, podemos verificar que π é uma isometria. É bem conhecido, ver por exemplo [75], que a única curva geodésica ligando os pontos $\gamma(0) = y$ e $\gamma(1) = x$ é:

$$\gamma(t) = (x_1^t y_1^{1-t}, \dots, x_n^t y_n^{1-t}),$$

onde

$$\gamma'(t) = (x_1^t y_1^{1-t} (\ln x_1 - \ln y_1), \dots, x_n^t y_n^{1-t} (\ln x_n - \ln y_n)).$$

A distância de Bregman é:

$$D_h(x, y) = h(x) - h(y) - \sum_{i=1}^n y_i \ln(x_i/y_i) \frac{\partial h(y)}{\partial y_i}.$$

A distância Riemanniana é:

$$d(x, y) = \left[\sum_{i=1}^n \left(\ln \frac{y_i}{x_i} \right)^2 \right]^{1/2}$$

Exemplo 5.4 Seja $M = (0, 1)^n$. Consideremos três métricas Riemannianas.

1. $(M, X^{-2}(I - X)^{-2})$ é uma variedade de Hadamard; esta variedade é isométrica a \mathbb{R}^n mediante a função $\pi(x) = \left(\ln \left(\frac{x_1}{1-x_1} \right), \dots, \ln \left(\frac{x_n}{1-x_n} \right) \right)$. A única curva geodésica, ligando os pontos $\gamma(0) = y$ e $\gamma(1) = x$, é: $\gamma(t) = (\gamma_1, \dots, \gamma_n)$ onde

$$\gamma_i(t) = \frac{1}{2} + \frac{1}{2} \tanh \left[(1/2) \left\{ \ln \left(\frac{x_i}{1-x_i} \right) - \ln \left(\frac{y_i}{1-y_i} \right) \right\} t + (1/2) \ln \left(\frac{y_i}{1-y_i} \right) \right],$$

com derivada

$$\gamma'_i(t) = \frac{\ln(x_i/(1-x_i)) - \ln(y_i/(1-y_i))}{4 \cosh((1/2)t + (1/2) \ln(y_i/(1-y_i)))}.$$

A distância de Bregman é:

$$D_h(x, y) = h(x) - h(y) - \sum_{i=1}^n \frac{(1-y_i)^2}{4y_i^2 \cosh^2(1/2)} \left\{ \ln \left(\frac{x_i}{1-x_i} \right) - \ln \left(\frac{y_i}{1-y_i} \right) \right\} \frac{\partial h(y)}{\partial y_i}.$$

A distância Riemanniana é

$$d(x, y) = \left[\sum_{i=1}^n \left(\ln \left(\frac{y_i}{1-y_i} \right) - \ln \left(\frac{x_i}{1-x_i} \right) \right)^2 \right]^{1/2}.$$

2. $(M, \csc^4(\pi x))$ é uma variedade de Hadamard, isométrica a \mathbb{R}^n , mediante a função $\pi(x) = \frac{1}{\pi} (\cot(\pi x_1), \dots, \cot(\pi x_n))$. A única curva geodésica ligando os pontos $\gamma(0) = y$ e $\gamma(1) = x$ é $\gamma(t) = (\gamma_1, \dots, \gamma_n)$ onde

$$\gamma_i(t) = \frac{1}{\pi} \arg \cot[(\cot \pi x_i - \cot \pi y_i)t + \cot(\pi y_i)],$$

com derivada

$$\gamma'_i(t) = (1/\pi) (\cot(\pi y_i) - \cot(\pi x_i)) \sin^2(\pi \gamma_i(t)),$$

e a distância de Bregman é:

$$D_h(x, y) = h(x) - h(y) - \sum_{i=1}^n \frac{1}{\pi} (\cot(\pi y_i) - \cot(\pi x_i)) \sin^2(\pi y_i) \frac{\partial h(y)}{\partial y_i}.$$

A distância Riemanniana é:

$$d(x, y) = \left[\sum_{i=1}^n [\cot(\pi y_i) - \cot(\pi x_i)]^2 \right]^{1/2}.$$

3. Finalmente, consideremos a métrica $\csc^2(\pi x)$. Esta variedade é de Hadamard isométrica a \mathbb{R}^n , veja [59]. A única curva geodésica ligando os pontos $\gamma(0) = y$ e $\gamma(1) = x$ é $\gamma(t) = (\gamma_1, \dots, \gamma_n)$ onde

$$\gamma_i(t) = \psi^{-1}(\psi(y_i) + t(\psi(x_i) - \psi(y_i))),$$

e

$$\psi(\tau) := \ln(\csc(\pi\tau) - \cot(\pi\tau)).$$

Por isso,

$$\gamma'_i(t) = (1/\pi) \ln \left(\frac{\csc(\pi x_i) - \cot(\pi x_i)}{\csc(\pi y_i) - \cot(\pi y_i)} \right) \sin(\pi \gamma_i(t)).$$

Portanto, a distância de Bregman é:

$$D_h(x, y) = h(x) - h(y) - \frac{1}{\pi} \sum_{i=1}^n \ln \left(\frac{\csc(\pi x_i) - \cot(\pi x_i)}{\csc(\pi y_i) - \cot(\pi y_i)} \right) \sin(\pi y_i) \frac{\partial h(y)}{\partial y_i}.$$

A distância Riemanniana é:

$$d(x, y) = \left[\sum_{i=1}^n (\psi(y_i) - \psi(x_i))^2 \right]^{1/2}.$$

Exemplo 5.5 $M = \mathcal{S}_{++}^n$, o conjunto das matrizes $n \times n$ simétricas e definidas positivas, com a métrica dada pela Hessiana de $-\ln \det(X)$, é uma variedade de Hadamard com curvatura não positiva. A curva geodésica ligando os pontos $\gamma(0) = Y$ e $\gamma(1) = X$, see [59], é dado por

$$\gamma(t) = Y^{1/2} (Y^{-1/2} X Y^{-1/2})^t Y^{1/2},$$

com

$$\gamma'(t) = Y^{1/2} \ln(Y^{-1/2} X Y^{-1/2}) (Y^{-1/2} X Y^{-1/2})^t Y^{1/2}.$$

Logo, a distância de Bregman é dada por:

$$D_h(X, Y) = h(X) - h(Y) - \text{tr}[\nabla h(Y)Y^{1/2} \ln(Y^{-1/2}XY^{-1/2})Y^{1/2}].$$

A distância Riemanniana é:

$$d^2(X, Y) = \sum_{i=1}^n \ln^2 \lambda_i(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}),$$

onde $\lambda(A)$ denota os autovalores da matriz simétrica A .

Capítulo 6

Uma Nova Barreira Auto-concordante para o Hipercubo

6.1 Introdução

A função barreira logarítmica, proposta por Frisch [27] no ano de 1955, e extensamente estudada por Fiacco e McCormick [26], cumpre um papel fundamental na obtenção de métodos de pontos interiores polinomiais para certas classes de problemas de otimização convexa, por exemplo, programas lineares, programas quadráticos convexos e programas semidefinidos positivos. Outras funções barreiras foram propostas, garantindo propriedades de convergência para seus algoritmos mas eles não são polinomiais (ver [40], [22], [73]). Observamos que todas estas barreiras não são barreiras auto-concordantes, propriedade que foi introduzida por Nesterov e Nemirovskii [58]. Estes autores desenvolveram uma abordagem geral que permite o uso de métodos de pontos interiores para resolver programas convexos baseados nestas classes de barreiras. Eles provaram que barreiras auto-concordantes associadas com métodos de Newton permitem desenvolver algoritmos com complexidade polinomial na análise de pior caso. Embora a grande importância teórica desta abordagem, atualmente a única barreira auto-concordante conhecida é a função logarítmica. Por isso, surge naturalmente uma pergunta: é possível obter uma outra barreira auto-concordante com propriedades similares à barreira logarítmica? Ela foi a motivação deste trabalho.

Neste capítulo respondemos parcialmente a esta pergunta introduzindo uma nova barreira

auto-concordante para o hipercubo $[0, 1]^n$, com parâmetro de auto-concordância $(3/2)n$. De fato, esta barreira é diferente da logarítmica e embora o parâmetro de auto-concordância seja pior (a função logaritmo tem parâmetro n), construiremos novos métodos de pontos interiores: um algoritmo polinomial de passos longos seguidor da trajetória central e um método de ponto proximal com distância de Bregman.

O esquema deste capítulo é a seguinte. Na seção 6.2, introduzimos a função $B(x) = \sum_{i=1}^n (2x_i - 1)[\ln x_i - \ln(1 - x_i)]$ e provamos que ela é $(3/2)n$ barreira-auto-concordante para $[0, 1]^n$. Na Seção 6.3 apresentamos um problema de otimização convexa com restrições lineares e variáveis limitadas, impomos algumas hipóteses sobre este problema e definimos uma nova trajetória central primal-dual. Provamos que a trajetória primal converge e o ponto de convergência é o centro analítico do conjunto de soluções ótimas do problema primal. Obtemos também convergência da trajetória primal-dual no caso linear. Na Seção 6.4 apresentamos uma aplicação para otimização linear. Para esta classe de problemas, introduzimos um novo algoritmo primal polinomial de passos longos, derivando limites superiores para obter uma ϵ -solução. A complexidade de nosso algoritmo é $O(n \ln(n\mu_0/\epsilon))$, para o número de iterações, idêntica ao algoritmo seguidor da trajetória central por passos longos que utiliza a função barreira logarítmica. Finalmente, na Seção 6.5 introduzimos um método de ponto proximal com distância tipo Bregman obtendo resultados globais de convergência.

6.2 A Nova Barreira

Nesta seção, introduzimos uma nova barreira para o hipercubo $[0, 1]^n$ e mostramos que ela é $(3/2)$ -barreira auto-concordante.

Definimos em $(0, 1)^n$ a função:

$$B(x) = \sum_{i=1}^n (2x_i - 1)[\ln x_i - \ln(1 - x_i)]. \quad (6.1)$$

Temos imediatamente as seguintes propriedades:

- Se $x \rightarrow \text{front}[0, 1]^n$ (x se aproxima da fronteira) então $B(x) \rightarrow \infty$.
- $B(x) \geq 0$, para todo $x \in (0, 1)^n$.

- $B \in C^\infty((0, 1)^n)$, e a derivada de segunda ordem de B é $X^{-2}(I - X)^{-2}$. Assim, B é uma função estritamente convexa em $(0, 1)^n$.

Lembremos inicialmente a definição dada por Nesterov e Nemirovskii, [58].

Definição 6.2.1 *Seja E um espaço vectorial real, $Q \subset E$ um subconjunto convexo aberto de E e $g : Q \rightarrow \mathbb{R}$ uma função. g é chamada barreira auto-concordante para \bar{Q} com parâmetro $c > 0$ (c barreira auto-concordante) se:*

1. $g \in C^3$ é uma função barreira para \bar{Q} , isto é, $g(x) \rightarrow +\infty$ quando $x \rightarrow \text{front}Q$.
2. g é uma função convexa sobre Q .
3. Para qualquer $x \in Q$ e $h \in E$:

$$|\nabla^3 g(x)[h, h, h]| \leq 2(\nabla^2 g(x)[h, h])^{\frac{3}{2}}.$$

4. Existe $c > 0$ tal que g satisfaz:

$$|\nabla g(x)[h]| \leq c^{\frac{1}{2}}(\nabla^2 g(x)[h, h])^{\frac{1}{2}}.$$

Teorema 6.2.1 *A função $B(x) = \sum_{i=1}^n (2x_i - 1)[\ln x_i - \ln(1 - x_i)]$ é $(3/2)n$ barreira auto-concordante para $[0, 1]^n$.*

Demonstração. Primeiro, consideremos o termo geral

$$b(z) = (2z - 1)[\ln z - \ln(1 - z)], \quad z \in \langle 0, 1 \rangle$$

A primeira, segunda e terceira derivada são, respectivamente:

$$b'(z) = 2[\ln z - \ln(1 - z)] - \frac{1}{z} + \frac{1}{1 - z},$$

$$b''(z) = \frac{1}{z^2(1 - z)^2},$$

$$b'''(z) = \frac{2(2z - 1)}{z^3(1 - z)^3}.$$

É fácil checar que $b(z)$ satisfaz as duas primeiras condições, por isso provaremos somente as restantes. Como

$$\frac{|b'''(z)|}{2(b''(z))^{\frac{3}{2}}} = |2z - 1| \leq 1,$$

obtemos que $b(z)$ satisfaz a terceira condição (freqüentemente chamada condição de 1-auto-concordância). Finalmente, provaremos a última condição:

$$\frac{|b'(z)|}{(b''(z))^{\frac{1}{2}}} = \frac{|2 \ln(\frac{z}{1-z}) + (2z-1)\frac{1}{z(1-z)}|}{\frac{1}{z(1-z)}} = \left| 2z(1-z) \ln\left(\frac{z}{1-z}\right) + 2z-1 \right| < 1.2,$$

por isso que c pode ser tomado como $(1.2)^2 = 1.44 < (3/2)$. Devido à propriedade de estabilidade com respeito ao produto direto, ver Proposições 2.1.1 e 2.3.1 de [58], obtemos que a função B é $(3/2)n$ barreira auto-concordante para $[0, 1]^n$ e, portanto, a prova está estabelecida.

■

6.3 Uma Nova Trajetória Central para Problemas Convexos Canalizados com Restrições Lineares

Neste capítulo estamos interessados em resolver o seguinte problema de otimização:

$$\min\{f(x) : Ax = b, 0 \leq x \leq e\} \tag{6.1}$$

onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa continuamente diferenciável, A é uma matriz $m \times n$ com $m < n$, e é o vetor n -dimensional onde todas as componentes são 1, b e c são vetores m - e n -dimensional, respectivamente; o vetor n -dimensional x é a variável. Observamos que todos os resultados obtidos neste capítulo podem ser facilmente estendidos para hipercubos mais gerais, dados por $\alpha \leq x \leq \beta$, para quaisquer vetores n -dimensionais α e β tais que $\alpha < \beta$.

Denotamos por

$$F = \{x \in \mathbb{R}^n : Ax = b, 0 \leq x \leq e\}$$

o conjunto viável do problema (6.1) e

$$F^0 = \{x \in \mathbb{R}^n : Ax = b, 0 < x < e\}$$

seu interior relativo. Impomos as seguintes hipóteses sobre o problema (6.1)

Hipóteses:

1. O conjunto F^0 é não vazio.
2. A função objetivo satisfaz a seguinte condição: existe um subespaço vetorial W de \mathbb{R}^n , tal que $\ker(\nabla^2 f(x)) = W$, para todo $x \in (0, 1)^n$.
3. A matriz A tem posto completo, isto é, $\text{rank}(A) = m$.

Observemos que 1 é uma hipótese necessária em métodos de pontos interiores. A hipótese 2, introduzida por Isem et al. [43] em um contexto mais geral, será importante para provar a convergência da trajetória central para o centro analítico do conjunto de soluções ótimas do problema (6.1). Esta hipótese é satisfeita pelas funções auto-concordantes (ver Nesterov and Nemirovskii [58], Corollary 2.1.1), em particular, as funções objetivo lineares e quadrático convexas. A hipótese 3 pode sempre ser obtida em qualquer problema do tipo (6.1), retirando linhas no caso em que estas sejam redundantes.

Como f é uma função contínua sobre o conjunto compacto F , o ponto de mínimo global é atingido em F . Além disso, a convexidade de f implica que todo mínimo local é mínimo global. Assim, o conjunto de soluções ótimas do problema (6.1), que denotaremos por $\text{sol}(P)$, é um conjunto não-vazio convexo e limitado.

A formulação dual de Wolfe para o problema (6.1) é:

$$\begin{aligned}
 \max \quad & d(x, y, s, w) = f(x) - y^T(Ax - b) + w^T(x - e) - s^T x \\
 & A^T y + s - w = \nabla f(x) \\
 & w, s \geq 0 \\
 & y \in \mathbb{R}^m; x, s, w \in \mathbb{R}^n
 \end{aligned} \tag{6.2}$$

Um ponto viável (x, y, s, w) para (6.2), com $(s, w) > 0$, é chamado solução dual viável interior.

6.3.1 A Trajetória Central

Nesta subseção propomos um novo problema auxiliar de penalidade interior para resolver (6.1) e estudamos a trajetória central (primal-dual), obtida pelas condições de otimalidade de KKT. Veremos que uma propriedade desta trajetória é que é primal viável, com respeito a (6.1), e dual inviável, com respeito ao problema dual (6.2), pelo termo $\mu[\ln X(\mu) - \ln(I - X(\mu))]e$, e

que converge para zero quando μ converge a zero (ver Corolário 6.3.1).

Para resolver (6.1) propomos o seguinte problema penalizado:

$$\begin{aligned} \min \quad & \phi_B(x, \mu) = f(x) + \mu \sum_{i=1}^n (2x_i - 1) [\ln x_i - \ln(1 - x_i)] \\ & Ax = b \\ & (0 < x < e) \end{aligned} \tag{6.3}$$

onde $\mu > 0$ é um parâmetro positivo. As derivadas de primeira e segunda ordem de ϕ_B são, respectivamente:

$$g(x, \mu) := \nabla f(x) + 2\mu[\ln X - \ln(I - X)]e - \mu X^{-1}e + \mu(I - X)^{-1}e \tag{6.4}$$

$$H(x, \mu) := \nabla^2 f(x) + \mu X^{-2}(I - X)^{-2}. \tag{6.5}$$

Simplificamos a notação, denotando $g := g(x, \mu)$ e $H := H(x, \mu)$.

Como $\phi_B(x, \mu)$ é estritamente convexa (porque f é uma função convexa e $X^{-2}(I - X)^{-2}$ é uma matriz simétrica definida positiva) no interior relativo do conjunto viável, e tende a infinito na fronteira de F , esta função atinge o valor mínimo no seu domínio (para μ fixo) em um único ponto. Seja $x(\mu)$ este ponto. As condições de otimalidade de KKT, necessárias e suficientes, para $x(\mu)$ são:

$$A^T y + s - w = \nabla f(x) + 2\mu[\ln X - \ln(I - X)]e \tag{6.6}$$

$$Ax = b \tag{6.7}$$

$$Xs = \mu e \tag{6.8}$$

$$(I - X)w = \mu e \tag{6.9}$$

$$(s, w) \geq 0 \tag{6.10}$$

$$0 < x < e \tag{6.11}$$

A única solução deste sistema, denotada por $(x(\mu), y(\mu), s(\mu), w(\mu))$, $\mu > 0$, chamaremos de trajetória central primal-dual. Claramente, esta trajetória é primal viável e dual inviável em relação aos problemas (6.1) e (6.2), respectivamente.

A seguir, provamos que a trajetória central primal-dual é uma curva regular.

Teorema 6.3.1 *A aplicação $\mu \rightarrow (x(\mu), y(\mu), s(\mu), w(\mu))$, $\mu > 0$, é continuamente diferenciável.*

Demonstração. Usaremos o Teorema da Função Implícita. Seja a função $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R} \longrightarrow \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ definida por:

$$F(x, y, s, w, \mu) = \begin{bmatrix} \nabla f(x) + 2\mu[\ln X - \ln(I - X)]e - A^T y - s + w \\ Ax - b \\ XSe - \mu e \\ (I - X)w - \mu e \end{bmatrix} \quad (6.12)$$

e consideremos a equação não linear:

$$F(x, y, w, s, \mu) = 0. \quad (6.13)$$

Fixe $\mu_0 > 0$. Pela definição da trajetória central, $(x(\mu_0), y(\mu_0), s(\mu_0), w(\mu_0), \mu_0)$ cumpre (6.13). Também tem-se que F é continuamente diferenciável, logo para utilizar o Teorema da Função Implícita (ver capítulo 2, Teorema 2.3.2) deve-se provar que o Jacobiano $J_F(\bar{z})$, onde $\bar{z} = (x(\mu_0), y(\mu_0), s(\mu_0), w(\mu_0), \mu_0)$, é não singular.

Derivando F em relação a (x, y, s, w) , obtemos:

$$J_F(\bar{z}) = \begin{bmatrix} \nabla^2 f(x(\mu_0)) + 2\mu_0(X^{-1}(\mu_0) + (I - X(\mu_0))^{-1}) & -A_{n \times m}^T & -I_{n \times n} & I_{n \times n} \\ A_{m \times n} & 0_{m \times m} & 0_{m \times n} & 0_{m \times n} \\ S_{n \times n}(\mu_0) & 0_{n \times m} & X_{n \times n}(\mu_0) & 0_{n \times n} \\ -W_{n \times n}(\mu_0) & 0_{n \times m} & 0_{n \times n} & I_{n \times n} - X_{n \times n}(\mu_0) \end{bmatrix}.$$

Seja $(p, q, r, t) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n$ tal que

$$\nabla^2 f(x(\mu_0))p + 2\mu_0(X^{-1}(\mu_0) + (I - X(\mu_0))^{-1})p - A^T q - r + t = 0 \quad (6.14)$$

$$Ap = 0 \quad (6.15)$$

$$S(\mu_0)p + X(\mu_0)r = 0 \quad (6.16)$$

$$-W(\mu_0)p + (I - X(\mu_0))t = 0 \quad (6.17)$$

Das equações (6.16) e (6.17) tem-se que:

$$r = -X(\mu_0)^{-1}S(\mu_0)p \quad (6.18)$$

$$t = (I - X(\mu_0))^{-1}W(\mu_0)p \quad (6.19)$$

Multiplicando a equação (6.14) por p^T e usando (6.15) obtemos:

$$p^T \nabla^2 f(x(\mu_0))p + 2\mu_0 p^T (X^{-1}(\mu_0) + (I - X(\mu_0))^{-1})p - p^T r + p^T t = 0 \quad (6.20)$$

Substituindo (6.18) e (6.19) em (6.20) obtemos:

$$p^T \left(\nabla^2 f(x(\mu_0)) + 2\mu_0 (X^{-1}(\mu_0) + (I - X(\mu_0))^{-1}) \right) p + p^T \left(X^{-1}(\mu_0) S(\mu_0) + (I - X(\mu_0))^{-1} W \right) p = 0$$

Devido a que $\nabla^2 f(x(\mu_0)) + 2\mu_0 (X^{-1}(\mu_0) + (I - X(\mu_0))^{-1})$ é definida positiva, deduz-se que $p = 0$. Daqui é imediato que $r = t = 0$. Finalmente, substituindo $p = r = t = 0$ em (6.14) obtemos

$$A^T q = 0.$$

Multiplicando por A , e lembrando que AA^T é invertível, tem-se que $q = 0$. Portanto $J_F(\bar{z})$ é não singular. Finalmente, aplicando o Teorema da Função Implícita, obtemos o resultado desejado. ■

6.3.2 A Trajetória Central Primal

Lemma 6.3.1 *Seja $x(\mu_i) := \arg \min\{\phi_B(x, \mu_i) : Ax = b, 0 < x < e\}, i = 1, 2$, com $0 < \mu_2 < \mu_1$. Os seguintes enunciados são verdadeiros:*

1. $\phi_B(x(\mu_2), \mu_2) \leq \phi_B(x(\mu_1), \mu_1)$, onde ϕ_B vem dada por (6.3).
2. $\sum_{i=1}^n (2x_i(\mu_1) - 1) [\ln x_i(\mu_1) - \ln(1 - x_i(\mu_1))] \leq \sum_{i=1}^n (2x_i(\mu_2) - 1) [\ln x_i(\mu_2) - \ln(1 - x_i(\mu_2))]$.
3. $f(x(\mu_2)) \leq f(x(\mu_1))$.

Demonstração. Desde que $0 < \mu_2 < \mu_1$, $x(\mu_2)$ minimiza $\phi_B(x, \mu_2)$ e $B(x) \geq 0, \forall x \in F$ temos:

$$\begin{aligned} \phi_B(x(\mu_2), \mu_2) &= f(x(\mu_2)) + \mu_2 B(x(\mu_2)) \\ &\leq f(x(\mu_1)) + \mu_2 B(x(\mu_1)) \\ &\leq f(x(\mu_1)) + \mu_1 B(x(\mu_1)) \\ &= \phi_B(x(\mu_1), \mu_1). \end{aligned}$$

Isto prova 1. Para provar 2 e 3, usar o Teorema 8 de [26]. ■

Lemma 6.3.2 *Se x^* é uma solução ótima do problema (6.1) e $\mu > 0$, então*

$$f(x^*) \leq f(x(\mu)) \leq \phi_B(x(\mu), \mu).$$

Demonstração. Usando o fato que $f(x^*) \leq f(x(\mu))$, para todo $\mu > 0$, e $B \geq 0$, obtemos o resultado desejado. ■

A partir deste ponto usaremos a notação:

$$(x^k, y^k, s^k, w^k) = (x(\mu_k), y(\mu_k), s(\mu_k), w(\mu_k)),$$

onde $\{\mu_k\} \subset \mathbb{R}_{++}$.

O seguinte teorema é um resultado bem conhecido para funções barreiras gerais não negativas.

Teorema 6.3.2 *Seja (μ_k) uma seqüência de números reais positivos tal que $\mu_k \rightarrow 0$ quando $k \rightarrow \infty$. Então*

1. $\mu_k B(x^k, \mu_k) \rightarrow 0$, quando $k \rightarrow \infty$.
2. *Todo ponto de acumulação da seqüência $\{x^k\}$ é uma solução ótima do problema (6.1).*

Demonstração. Ver Bazaara [7], Teorema 9.4.3. ■

Corolário 6.3.1 *Seja (x^k) a seqüência da trajetória central primal. Temos que:*

$$\mu_k [\ln x_i^k - \ln(1 - x_i^k)] \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n.$$

Demonstração. Pelo teorema anterior

$$\mu_k \sum_{i=1}^n [(2x_i^k - 1)(\ln x_i^k - \ln(1 - x_i^k))] \rightarrow 0, \quad k \rightarrow \infty.$$

Como $b(x_i^k) := (2x_i^k - 1)(\ln x_i^k - \ln(1 - x_i^k)) \geq 0$ então

$$\mu_k b(x_i^k) \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n,$$

ou equivalentemente:

$$2\mu_k [x_i^k \ln x_i^k + (1 - x_i^k) \ln(1 - x_i^k)] - \mu_k \ln x_i^k - \mu_k \ln(1 - x_i^k) \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n. \quad (6.21)$$

A seqüência $\gamma(x_i^k) := x_i^k \ln x_i^k + (1 - x_i^k) \ln(1 - x_i^k)$ é limitada. Agora, como $\mu_k \rightarrow 0$, $k \rightarrow \infty$, $i = 1, \dots, n$, temos

$$2\mu_k \gamma(x_i^k) \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n. \quad (6.22)$$

Subtraindo a expressão (6.22) de (6.21), e levando em conta que as duas são convergentes, temos

$$-\mu_k \ln x_i^k - \mu_k \ln(1 - x_i^k) \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n.$$

Além disso, como $x_i^k \in (0, 1)^n$, então $-\mu_k \ln x_i^k > 0$ e $-\mu_k \ln(1 - x_i^k) > 0$, $i = 1, \dots, n$. Por isso,

$$\mu_k \ln x_i^k \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n, \quad (6.23)$$

e

$$-\mu_k \ln(1 - x_i^k) \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n. \quad (6.24)$$

Somando (6.23) e (6.24) concluímos que

$$\mu_k [\ln x_i^k - \ln(1 - x_i^k)] \rightarrow 0, \quad k \rightarrow \infty, \quad i = 1, \dots, n. \quad \blacksquare$$

Provaremos agora que a trajetória central primal converge, isto é, existe limite de $x(\mu)$ quando $\mu \rightarrow 0$. Além disso, podemos caracterizar tal ponto como o centro analítico do conjunto de soluções ótimas $\text{sol}(P)$. Para provar este fato, usaremos a hipótese 2 e adaptaremos os resultados obtidos para desigualdades variacionais por Iusen et al. [43].

O centro analítico do conjunto de soluções ótimas é definido como a solução de

$$\begin{aligned} \min \quad & \sum_{j \in J} (2x_j - 1) [\ln x_j - \ln(1 - x_j)] \\ & x \in \text{ri}(\text{sol}(P)), \end{aligned} \quad (6.25)$$

onde $\text{ri}(\text{sol}(P))$ é o interior relativo do conjunto de soluções ótimas do problema (6.1), e $J = \{j \in \{1, 2, \dots, n\} : \exists x \in \text{sol}(P), 0 < x_j < 1\}$. Observe que, da convexidade de $\text{sol}(P)$, segue-se que

$$\text{ri}(\text{sol}(P)) = \{x \in \text{sol}(P) : 0 < x_j < 1\}.$$

Se $\text{sol}(P)$ é um conjunto unitário, a única solução é um ponto extremo de $[0, 1]^n$ então $J = \emptyset$, e, por convenção, $\sum_{j \in J} (2x_j - 1) [\ln x_j - \ln(1 - x_j)] := 0$. Note que a função objetivo de (6.25)

é estritamente convexa sobre $\text{ri}(\text{sol}(P))$, e diverge na fronteira relativa de $\text{sol}(P)$. Logo, devido à propriedade que $\text{sol}(P)$ é limitado, o centro analítico está bem definido.

Provaremos que a trajetória central primal converge ao centro analítico do conjunto de soluções ótimas. Os argumentos são similares aos usados para a barreira logarítmica, em Graña Drummond e Svaiter [34].

Teorema 6.3.3 *A trajetória central primal $x(\mu)$ converge ao centro analítico do conjunto de soluções ótimas $\text{sol}(P)$.*

Demonstração. Estamos interessados no caso em que $J \neq \emptyset$ (Se $J = \emptyset$, então $\text{sol}(P)$ é um conjunto unitário e a única solução é um ponto extremo, por isso, usando o Teorema 6.3.2, parte 2, $x(\mu)$ converge para esse ponto). Como a trajetória central primal $\{x(\mu)\}$ é limitada, existe um ponto de acumulação x^* . Seja (μ_k) uma seqüência de números positivos tal que

$$\mu_k \rightarrow 0, \quad \text{e} \quad x(\mu_k) \rightarrow x^* \text{ quando } k \rightarrow \infty.$$

Denote $x^k = x(\mu_k)$. Seja \bar{x} a solução de (6.25). Em vista do Teorema 6.3.2 provaremos que $0 < x_j^* < 1$ e

$$\sum_{j \in J} (2x_j^* - 1) [\ln x_j^* - \ln(1 - x_j^*)] \leq \sum_{j \in J} (2\bar{x}_j - 1) [\ln \bar{x}_j - \ln(1 - \bar{x}_j)].$$

A primeira condição implica que x^* é viável para (6.25), a segunda implica que x^* é um ponto de mínimo para (6.25). Definindo a seqüência

$$p^k = \bar{x} + x^k - x^*,$$

obtemos imediatamente que $p^k \rightarrow \bar{x}$ quando $k \rightarrow \infty$. Primeiro, provaremos que para k suficientemente grande, p^k é um ponto estritamente viável para o problema (6.25). Claramente, sendo x^k viável e x^* uma solução ótima, (ver Teorema 6.3.2), temos que $Ap^k = A(\bar{x} + x^k - x^*) = b$, para todo k . Agora, suponha-se que $j \notin J$. Mostraremos que $\bar{x}_j = x_j^* = 0$ ou $\bar{x}_j = x_j^* = 1$. De fato, suponha $\bar{x}_j \neq x_j^*$, para algum $j \notin J$. Claramente, $z(t) = x^* + t(\bar{x} - x^*)$ resolve (P). Da convexidade de f e da otimalidade de \bar{x} e x^* temos que,

$$f(z(t)) = f((1-t)x^* + t\bar{x}) \leq (1-t)f(x^*) + tf(\bar{x}) = f(x^*).$$

Para o índice j acima, temos $z_j(t) = x_j^* + t(\bar{x}_j - x_j^*)$. Se $\bar{x}_j = 0$, então, da hipótese, $x_j^* = 1$, e $z_j(t) = 1 - t$, para qualquer $t \in [0, 1]$. Usando $t = 1/2$, conseguimos que $z(1/2) = x^* + (\bar{x} - x^*)/2$ é uma solução para (P) . Agora, observe que $0 < z_j(1/2) < 1$, o qual significa que $j \in J$, uma contradição. A prova é análoga se temos $\bar{x}_j = 1$ e $x_j^* = 0$.

Assim, $x_j^k = p_j^k$ para $j \notin J$ e $0 < p_j^k < 1$, pois $0 < x_j^k < 1$. Por outro lado, para $j \in J$, desde que $x_j^k \rightarrow x_j^*$ quando $k \rightarrow \infty$, e $0 < \bar{x}_j < 1$, então $0 < p_j^k < 1$ para k suficientemente grande. Concluimos que $0 < p^k < e$ para k suficientemente grande.

Agora, provaremos que

$$f(p^k) = f(x^k).$$

Considerando $f(x^* + t(\bar{x} - x^*))$ como uma função de $t \in [0, 1]$. Desde que x^* e \bar{x} pertence ao conjunto convexo $\text{sol}(P)$ e usando a convexidade de f , segue que $f(x^* + t(\bar{x} - x^*))$ é constante. Conseqüentemente,

$$\nabla f(x^*)^T(\bar{x} - x^*) = 0,$$

$$(\bar{x} - x^*)^T \nabla^2 f(x^*)(\bar{x} - x^*) = 0.$$

Desde que a Hessiana de f é uma matriz simétrica semidefinida positiva, segue-se que

$$\nabla^2 f(x^*)(\bar{x} - x^*) = 0,$$

assim $\bar{x} - x^* \in \ker(\nabla^2 f(x^*))$. Em vista da hipótese 2, é verdade que $\bar{x} - x^* \in W$. Além disso, pela fórmula da integral de Taylor, temos

$$\nabla f(x^k) = \nabla f(x^*) + \int_0^1 \nabla^2 f(x^* + t(x^k - x^*))(x^k - x^*) dt.$$

Temos também que $\nabla^2 f(x^* + t(x^k - x^*))(x^k - x^*) \in \text{range}(\nabla^2 f(x^* + t(x^k - x^*))) = (\ker(\nabla^2 f(x^* + t(x^k - x^*))))^{\perp} = (\ker(\nabla^2 f(x^* + t(x^k - x^*))))^{\perp} = W^{\perp}$, novamente, pela hipótese 2, segue-se que $\int_0^1 \nabla^2 f(x^* + t(x^k - x^*))(x^k - x^*) dt \in W^{\perp}$. Assim

$$\nabla f(x^k)^T(\bar{x} - x^*) = 0. \tag{6.26}$$

Finalmente, usando a caracterização das funções convexas, a definição de p^k , e (6.26), obtemos

$$f(p^k) \geq f(x^k) + \nabla f(x^k)^T(p^k - x^k)$$

$$\begin{aligned}
&= f(x^k) + \nabla f(x^k)^T(\bar{x} - x^*) \\
&= f(x^k).
\end{aligned}$$

Assim

$$f(p^k) \geq f(x^k). \quad (6.27)$$

Similarmente, trocando x^k por p^k , temos

$$f(x^k) \geq f(p^k),$$

o qual, com (6.27), leva para

$$f(x^k) = f(p^k).$$

Agora, desde que $Ap^k = b$ e $0 < p^k < e$, para k suficientemente grande, a propriedade de otimalidade de x^k fornece, para k suficientemente grande,

$$f(x^k) + \mu_k \sum_{i=1}^n (2x_i^k - 1) [\ln x_i^k - \ln(1 - x_i^k)] \leq f(p^k) + \mu_k \sum_{i=1}^n (2p_i^k - 1) [\ln p_i^k - \ln(1 - p_i^k)].$$

Como $p_j^k = x_j^k$, para todo $j \notin J$, e $f(x^k) = f(p^k)$, temos também, para k suficientemente grande:

$$\sum_{j \in J} (2x_j^k - 1) [\ln x_j^k - \ln(1 - x_j^k)] \leq \sum_{j \in J} (2p_j^k - 1) [\ln p_j^k - \ln(1 - p_j^k)].$$

Como o limite em k obtemos para $j \in J$ que

$$0 < x_j^* < 1, \quad e$$

$$\sum_{j \in J} (2x_j^* - 1) [\ln x_j^* - \ln(1 - x_j^*)] \leq \sum_{j \in J} (2\bar{x}_j - 1) [\ln \bar{x}_j - \ln(1 - \bar{x}_j)].$$

Desde que \bar{x} é única, pela estrita convexidade da função objetivo do problema (6.25), concluímos que $x^* = \bar{x}$, portanto a trajetória central primal $\{x(\mu), \mu > 0\}$ converge para o centro analítico de $\text{sol}(P)$. ■

6.3.3 A Trajetória Central Primal-Dual

Proposição 6.3.1 Para todo $r > 0$ o conjunto $\{(x(\mu), s(\mu), w(\mu)) : \mu \leq r\}$ é limitado.

Prova. Como $x(\mu) \in (0, 1)^n$ então o conjunto $\{x(\mu)\}$ é limitado e, em particular, quando $\mu \leq r$. Agora, provaremos que $\{(s(\mu), w(\mu)) : \mu \leq r\}$ é também limitado.

Sabemos que $(x(\mu), y(\mu), s(\mu), w(\mu))$ resolve a equação

$$A^T y + s - w = \nabla f(x) + 2\mu[\ln X - \ln(I - X)]e,$$

por isso, definindo a Lagrangeana

$$L(x, y, s, w) = f(x) - y^T(Ax - b) - s^T x + w^T(x - e) + 2\mu e^T[X \ln X + (I - X) \ln(I - X)]e$$

temos

$$\nabla_x L(x(\mu), y(\mu), s(\mu), w(\mu)) = 0 \quad (6.28)$$

onde ∇_x denota o gradiente de L com respeito à variável x . Agora, observando que $L(\cdot, y(\mu), s(\mu), w(\mu))$ é uma função estritamente convexa sobre $(0, 1)^n$, (6.28) implica

$$x(\mu) \in \arg \min\{L(x, y(\mu), s(\mu), w(\mu)), x \in (0, 1)^n\}.$$

Por outro lado, temos que: $f(x(\mu)) - L(x(\mu), y(\mu), s(\mu), w(\mu))$

$$\begin{aligned} &= s(\mu)^T x(\mu) - w(\mu)^T(x(\mu) - e) - 2\mu e^T[X(\mu) \ln X(\mu) + (I - X(\mu)) \ln(I - X(\mu))]e \\ &= n\mu + n\mu - 2\mu e^T[X(\mu) \ln X(\mu) + (I - X(\mu)) \ln(I - X(\mu))]e \\ &\leq 2n\mu(1 + \ln 2). \end{aligned}$$

Logo, pela hipótese 1, podemos tomar um ponto x^0 primal viável, obtendo

$$\begin{aligned} f(x(\mu)) - 2n\mu(1 + \ln 2) &\leq L(x(\mu), y(\mu), s(\mu), w(\mu)) \\ &\leq L(x^0, y(\mu), s(\mu), w(\mu)) \\ &\leq f(x^0) - s(\mu)^T x^0 + w(\mu)^T(x^0 - e), \end{aligned}$$

onde a última desigualdade é uma consequência da viabilidade de x^0 , e que $\sum_{i=1}^n [x_i^0 \ln x_i^0 + (1 - x_i^0) \ln(1 - x_i^0)] \leq 0$. Assim,

$$f(x(\mu)) - 2n\mu(1 + \ln 2) \leq f(x^0) - s(\mu)^T x^0 + w(\mu)^T(x^0 - e). \quad (6.29)$$

Desde que $0 < x_i^0 < 1$, $i = 1, \dots, n$, o valor ξ definido por

$$\xi = \min\{x_i^0, 1 - x_i^0, i = 1, \dots, n\}$$

é positivo. Agora, seja f^* o valor ótimo do problema (6.1). De (6.29), e usando $\mu \leq r$, obtemos que:

$$\begin{aligned} \|s(\mu)\|_1 + \|w(\mu)\|_1 &\leq \frac{1}{\xi}(f(x^0) + 2n\mu(1 + \ln 2) - f(x(\mu))) \\ &\leq \frac{1}{\xi}(f(x^0) + 2n(1 + \ln 2)r - f^*), \end{aligned}$$

onde $\|\cdot\|_1$ é a norma da soma (ver notação do capítulo 1). Portanto, o conjunto $\{(x(\mu), s(\mu), w(\mu)) : \mu \leq r\}$ é limitado. ■

A seguir provamos a existência de pontos de acumulação da trajetória central primal-dual.

Lemma 6.3.3 . *O conjunto dos pontos de acumulação da trajetória central primal-dual $\{(x(\mu), y(\mu), s(\mu), w(\mu))\}$ é não vazio.*

Prova. Devido a que $\{(x(\mu), s(\mu), w(\mu)) : \mu \leq r\}$ é limitado (ver proposição anterior), existe uma subsequência $\{(x(\mu_j), s(\mu_j), w(\mu_j))\}$ e um ponto $(\bar{x}, \bar{s}, \bar{w})$ tal que:

$$\begin{aligned} (x(\mu_j), s(\mu_j), w(\mu_j)) &\rightarrow (\bar{x}, \bar{s}, \bar{w}), j \rightarrow \infty, \quad e \\ \mu_j &\rightarrow 0, j \rightarrow \infty. \end{aligned}$$

Da equação da trajetória central,

$$y(\mu_j) = (AA^T)^{-1}A(\nabla f(x(\mu_j)) + 2\mu_j[\ln(X(\mu_j)) - \ln(I - X(\mu_j))] - s(\mu_j) + w(\mu_j)).$$

Esta expressão levada ao limite em j , usando a continuidade de ∇f e o resultado do Corolário 6.3.1 fornece

$$y(\mu_j) \rightarrow (AA^T)^{-1}A(\nabla f(\bar{x}) - \bar{s} + \bar{w}), j \rightarrow \infty.$$

Definindo $\bar{y} := (AA^T)^{-1}A(\nabla f(\bar{x}) - \bar{s} + \bar{w})$, obtemos que $(\bar{x}, \bar{y}, \bar{s}, \bar{w})$ é um ponto de acumulação da trajetória primal-dual. Portanto, o conjunto de pontos de acumulação é não vazio. ■

Proposição 6.3.2 *Todo ponto de acumulação da trajetória central primal-dual é solução ótima do par de problemas primal (6.1) e dual (6.2).*

Prova. Seja $(\bar{x}, \bar{y}, \bar{s}, \bar{w})$ um ponto de acumulação da trajetória central primal-dual. Então existem seqüências $\{\mu_k\}$ e $\{(x(\mu_k), y(\mu_k), s(\mu_k), w(\mu_k))\}$ tais que

$$\mu_k \rightarrow 0, k \rightarrow \infty, \quad e$$

$$(x(\mu_k), y(\mu_k), s(\mu_k), w(\mu_k)) \rightarrow (\bar{x}, \bar{y}, \bar{s}, \bar{w}), k \rightarrow \infty.$$

Pelo Teorema 6.3.2, \bar{x} é solução ótima do problema primal (6.1). Logo, resta provar que $(\bar{y}, \bar{s}, \bar{w})$ é solução do problema dual (6.2). Das condições de KKT:

$$A^T y^k + s^k - w^k = \nabla f(x^k) + 2\mu[\ln X^k - \ln(I - X^k)]e,$$

logo, considerando o limite em k , usando a propriedade de continuidade e o Corolário 6.3.1, tem-se:

$$A^T \bar{y} + \bar{s} - \bar{w} = \nabla f(\bar{x}). \quad (6.30)$$

Também, como $s^k > 0$ e $w^k > 0$ obtemos:

$$\bar{s} \geq 0 \quad \text{e} \quad \bar{w} \geq 0. \quad (6.31)$$

De (6.30) e (6.31) concluímos que $(\bar{y}, \bar{s}, \bar{w})$ é um ponto viável do problema (6.2). Para verificar que $(\bar{y}, \bar{s}, \bar{w})$ é uma solução ótima, é suficiente provar a propriedade de complementaridade, isto é, que $\bar{w}_i(\bar{x}_i - 1) = 0$ e $\bar{s}_i \bar{x}_i = 0$, para todo $i = 1, 2, \dots, n$.

Das condições de KKT, $x_i^k s_i^k = \mu_k$, logo no limite

$$\bar{x}_i \bar{s}_i = 0, \quad \forall i = 1, 2, \dots, n.$$

Da mesma forma pode-se mostrar que

$$\bar{w}_i(1 - \bar{x}_i) = 0, \quad \forall i = 1, 2, \dots, n.$$

Portanto, a prova está concluída. ■

Finalmente, provaremos que no caso linear a trajetória central primal-dual converge. Para obter este resultado usaremos o Lema 2.5.1 do Capítulo 2.

Teorema 6.3.4 *No caso linear, a trajetória central primal-dual $(x(\mu), y(\mu), s(\mu), w(\mu))$ converge a um ponto $(x^*, y^*, s^*, w^*) \in \text{sol}(P) \times \text{sol}(D)$.*

Demonstração. Como a seqüência $\{(x(\mu), y(\mu), s(\mu), w(\mu))\}$ é limitada, existe um ponto $(x^*, y^*, s^*, w^*) \in \text{sol}(P) \times \text{sol}(D)$ e uma subseqüência $\{\mu_k\}$ tal que:

$$(x^k, y^k, s^k, w^k) \rightarrow (x^*, y^*, s^*, w^*) \quad \text{quando} \quad k \rightarrow \infty \quad \text{e} \quad (6.32)$$

$\mu_k \rightarrow 0$, quando $k \rightarrow \infty$.

Definamos o seguinte conjunto:

$$M = \left\{ (\bar{x}, \bar{y}, \bar{s}, \bar{w}, \mu) : \begin{array}{l} A\bar{x} = 0 \\ A^T\bar{y} + \bar{s} - \bar{w} = 2\mu[\ln(\bar{X} + X^*) - \ln(I - (\bar{X} + X^*))]e \\ (\bar{X} + X^*)(\bar{S} + S^*)e = \mu e \\ (I - (\bar{X} + X^*))(\bar{W} + W^*)e = \mu e \\ (\bar{x} + x^*) > 0, (e - (\bar{x} + x^*)) > 0, (\bar{s} + s^*) > 0, (\bar{w} + w^*) > 0 \text{ e } \mu > 0. \end{array} \right\}.$$

É fácil verificar que se existe um ponto $(\bar{x}, \bar{y}, \bar{s}, \bar{w}, \mu) \in M$ então $(\bar{x} + x^*, \bar{y} + y^*, \bar{s} + s^*, \bar{w} + w^*)$ é um ponto da trajetória central primal-dual. Temos também que o elemento zero pertence ao fecho de M . De fato, definamos a seqüência

$$(\bar{x}_k, \bar{y}_k, \bar{s}_k, \bar{w}_k, \bar{\mu}_k) := (x^k - x^*, y^k - y^*, s^k - s^*, w^k - w^*, \mu_k).$$

Claramente, $(\bar{x}_k, \bar{y}_k, \bar{s}_k, \bar{w}_k, \bar{\mu}_k) \in M$, e de (6.32) obtemos

$$(\bar{x}_k, \bar{y}_k, \bar{s}_k, \bar{w}_k, \bar{\mu}_k) \rightarrow (0, 0, 0, 0, 0) \text{ quando } k \rightarrow \infty.$$

Portanto, $0 \in \overline{M}$. Agora, com vistas ao resultado de convergência, aplicaremos o Lema 2.5.1 da curva selecionada. Para isso, observemos que este lema implica a existência de um $\epsilon > 0$ e uma função analítica $\gamma: [0, \epsilon) \rightarrow \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ tais que

$$\gamma(t) = (\bar{x}(t), \bar{y}(t), \bar{s}(t), \bar{w}(t), \mu(t)) \rightarrow (0, 0, 0, 0, 0) \text{ quando } t \rightarrow 0, \quad (6.33)$$

e se $t > 0$, $(\bar{x}(t), \bar{y}(t), \bar{s}(t), \bar{w}(t), \mu(t)) \in M$, isto é,

$$\begin{array}{rcl} A\bar{x}(t) & = & 0, \\ A^T\bar{y}(t) + \bar{s}(t) - \bar{w}(t) & = & 2\mu[\ln(\bar{X}(t) + X^*) - \ln(I - (\bar{X}(t) + X^*))]e, \\ (\bar{X}(t) + X^*)(\bar{S}(t) + S^*)e & = & \mu(t)e, \\ (I - (\bar{X}(t) + X^*))(\bar{W}(t) + W^*)e & = & \mu(t)e, \\ (\bar{x}(t) + x^*) & > & 0, \\ (e - (\bar{x}(t) + x^*)) & > & 0, \\ (\bar{s}(t) + s^*) & > & 0, \\ (\bar{w}(t) + w^*) & > & 0, \\ \mu(t) & > & 0. \end{array}$$

(6.34)

Desde que a trajetória central é única, o sistema (6.34) também tem uma única solução dada por:

$$x(\mu(t)) = \bar{x}(t) + x^*, \quad y(\mu(t)) = \bar{y} + y^*, \quad s(\mu(t)) = \bar{s}(t) + s^*, \quad w(\mu(t)) = \bar{w}(t) + w^*,$$

se $t > 0$. Apliquemos agora o limite quando $t \rightarrow 0$ e usemos (6.33), obtendo

$$\lim_{t \rightarrow 0} x(\mu(t)) = x^*, \quad \lim_{t \rightarrow 0} y(\mu(t)) = y^*, \quad \lim_{t \rightarrow 0} s(\mu(t)) = s^*, \quad \lim_{t \rightarrow 0} w(\mu(t)) = w^*, \quad \lim_{t \rightarrow 0} \mu(t) = 0.$$

Desde que $\mu(t) > 0$ em $(0, \epsilon)$, $\mu(0) = 0$, e $\mu(t)$ é analítica em $[0, \epsilon)$, existe um intervalo, $(0, \epsilon')$ onde $\mu'(t) > 0$. Por isso, a função inversa $\mu^{-1}: \mu(t) \rightarrow t$ existe num intervalo $(0, \mu(\epsilon'))$. Além disso, $\mu^{-1}(t) > 0$ para todo $t \in (0, \mu(\epsilon'))$, e $\lim_{t \rightarrow 0} \mu^{-1}(t) = 0$. Portanto, segue-se que

$$\lim_{t \rightarrow 0} x(t) = \lim_{t \rightarrow 0} x(\mu(\mu^{-1}(t))) = \lim_{t \rightarrow 0} \bar{x}(\mu^{-1}(t)) + x^* = x^*,$$

e, de uma maneira similar

$$\lim_{t \rightarrow 0} y(t) = y^* \quad \lim_{t \rightarrow 0} s(t) = s^*, \quad \lim_{t \rightarrow 0} w(t) = w^*.$$

Desde que (x^*, y^*, s^*, w^*) é arbitrário, ele deve ser único, concluindo assim a demonstração.

■

6.4 Aplicação para Otimização Linear

Nesta seção aplicamos a nova barreira para resolver problemas de otimização linear. Apresentamos um algoritmo primal seguidor da trajetória central e fornecemos um limite superior para o número total de iterações aplicando o método de Newton, para obter uma ϵ -solução ótima, alcançamos a complexidade de $O(n \ln(n\mu_0/\epsilon))$ iterações, onde μ_0 é o parâmetro inicial do algoritmo.

Consideremos o problema (6.1) quando a função objetivo é linear:

$$\begin{aligned} \min \quad & c^T x \\ & Ax = b \\ & 0 \leq x \leq e \end{aligned} \tag{6.35}$$

As hipóteses 1 e 3 da Seção 6.3 são mantidas, enquanto a hipótese 2 é imediatamente satisfeita, devido à linearidade da função objetivo. Além disso, a função $c^T x$ é contínua sobre o conjunto compacto F ; devido a isto existe um ponto de mínimo para o problema. Além do mais, o conjunto de soluções ótimas $\text{sol}(P)$ é não vazio e limitado. A formulação dual de (6.1) é:

$$\begin{aligned} \max \quad & d(y, s, w) = b^T y - w^T e \\ & A^T y + s - w = c \\ & w, s \geq 0 \\ & y \in \mathbb{R}^m \end{aligned} \tag{6.36}$$

onde y, s são vetores m - e n - dimensionais, respectivamente. O salto de dualidade, de (6.35) e (6.36), é dado por:

$$\text{gap} := c^T x - (b^T y - w^T e) = e^T (Xs + (I - X)w). \tag{6.37}$$

Segue-se que

$$d(y, s, w) \leq z^* \leq c^T x,$$

onde z^* denota o valor objetivo ótimo de (6.35). É também bem conhecido que, se x^* é um ponto ótimo de (6.35), então existe (y^*, s^*, w^*) que é ponto ótimo para (6.36), com $c^T x^* = d(y^*, s^*, w^*) = z^*$. Estas soluções são caracterizadas pelas condições de Karush-Kuhn-Tucker (KKT):

$$A^T y + s - w = c \tag{6.38}$$

$$Ax = b \tag{6.39}$$

$$Xs = 0 \tag{6.40}$$

$$(I - X)w = 0 \tag{6.41}$$

$$(s, w) \geq 0 \tag{6.42}$$

$$0 \leq x \leq e \tag{6.43}$$

A seguir consideramos a função barreira para o problema primal (6.35):

$$\phi_B(x, \mu) = \frac{1}{\mu} (c^T x) + \sum_{i=1}^n (2x_i - 1) [\ln x_i - \ln(1 - x_i)] \tag{6.44}$$

onde μ é um parâmetro positivo. As derivadas de primeira e segunda ordem de ϕ_B são:

$$g = g(x, \mu) := \frac{c}{\mu} + 2[\ln X - \ln(I - X)]e - X^{-1}e + (I - X)^{-1}e \quad (6.45)$$

$$H = H(x, \mu) := X^{-2}(I - X)^{-2}. \quad (6.46)$$

Como $\phi_B(x, \mu)$ é estritamente convexa no interior relativo do conjunto de soluções viáveis e tende a infinito em pontos que se aproximam da fronteira de F , o valor mínimo desta função é atingido no seu domínio (para μ fixado) em um único ponto $x(\mu)$. As condições de otimalidade de primeira ordem (condições de KKT) necessárias e suficientes para $x(\mu)$ são:

$$A^T y + s - w = c + 2\mu[\ln X - \ln(I - X)]e \quad (6.47)$$

$$Ax = b \quad (6.48)$$

$$Xs = \mu e \quad (6.49)$$

$$(I - X)w = \mu e \quad (6.50)$$

$$(s, w) \geq 0 \quad (6.51)$$

$$(0 < x < e) \quad (6.52)$$

A única solução deste sistema é de novo denotada por $(x(\mu), y(\mu), s(\mu), w(\mu))$. O salto de dualidade nesta solução satisfaz:

$$\text{gap}(\mu) := c^T x(\mu) - d(x(\mu), y(\mu), s(\mu), w(\mu)) = 2\mu(n - e^T[\ln X(\mu) - \ln(I - X(\mu))]x(\mu)).$$

Já provamos que o termo $\mu[\ln X - \ln(I - X)]e$ converge para zero e que a trajetória central primal converge ao centro analítico do conjunto de soluções ótimas $\text{sol}(P)$ (ver Corolário 6.3.1 e Teorema 6.3.3), logo o salto de dualidade converge para zero quando $\mu \rightarrow 0$.

No algoritmo que apresentaremos na Subsecção 6.4.2, dado um ponto estritamente viável $x \in F^0$, a direção de descida p será obtida resolvendo o problema:

$$\begin{aligned} \min \quad & \phi_B(x, \mu) + g(x, \mu)d + \frac{1}{2}p^T H(x, \mu)p \\ \text{s.a :} \quad & Ap = 0 \end{aligned} \quad (6.53)$$

As condições de otimalidade são dadas pelo seguinte sistema:

$$g + Hp = A^T \lambda \quad (6.54)$$

$$Ap = 0. \quad (6.55)$$

Esta direção é chamada **direção projetada de Newton** e sua expressão explícita é:

$$p = -H^{-1/2} \left(I - H^{-1/2} A^T (A H^{-1} A^T)^{-1} A H^{-1/2} \right) H^{-1/2} g. \quad (6.56)$$

A estimativa dual é:

$$\lambda = (A H^{-1} A^T)^{-1} A H^{-1} g \quad (6.57)$$

De forma usual, escrevemos $\|q\|_H = \sqrt{q^T H q}$, $q \in \mathbb{R}^n$, para medir a distância entre pontos.

Claramente $\|\cdot\|_H$ define uma norma pois H é definida positiva.

6.4.1 Propriedades em Vizinhanças da Trajetória Central

Nesta subseção apresentamos alguns lemas que serão necessários para obter um limite superior para o número total de iterações internas e externas. Os seguintes dois lemas são essenciais e foram introduzidas por Nesterov and Nemirovskii [58] (ver também Den Hertog [39], para uma versão mais simples das provas).

Lema 6.4.1 *Sejam $x \in (0, 1)^n$ e $d \in \mathbb{R}^n$ arbitrários. Se $\|d\|_{H(x, \mu)} < 1$, então $x + d \in (0, 1)^n$.*

Demonstração. Ver [39], Lema 2.20. ■

Lema 6.4.2 *Sejam $x \in F^0$, p uma direção projetada de Newton e $x^+ = x + p$. Se $\|p\|_H < 1$ então $x^+ \in F^0$ e*

$$\|p\|_{H(x^+, \mu)} \leq \frac{1}{(1 - \|p\|_H)^2} \|p\|_H^2.$$

Para $\|p\|_H < \frac{3-\sqrt{5}}{2}$, isto implica que $\|p(x^+, \mu)\| < \|p\|_H$, assim asseguramos a convergência do método de Newton. Para $\|p\|_H < \frac{1}{3}$, cumpre-se que

$$\|p\|_{H(x^+, \mu)} \leq \frac{9}{4} \|p\|_H^2. \quad (6.58)$$

Demonstração. Ver [39], Lema 2.21. ■

Seja um ponto da trajetória central primal $x(\mu)$ (chamado centro com respeito a μ). Dizemos que **um ponto estritamente viável $x \in F^0$ está centrado com respeito a $x(\mu)$ se $\|p\|_H < \frac{1}{3}$.**

O seguinte lema fornece um limite superior para a diferença dos valores da barreira B entre um ponto centrado x e seu centro $x(\mu)$.

Lema 6.4.3 *Se a direção projetada de Newton p satisfaz $\|p\|_H < \frac{1}{3}$, então*

$$\phi_B(x, \mu) - \phi_B(x(\mu), \mu) \leq \frac{\|p\|_H^2}{1 - (\frac{9}{4})^2 \|p\|_H^2}. \quad (6.59)$$

Demonstração. A função barreira é convexa em x , donde

$$p^T g + \phi_B(x, \mu) \leq \phi_B(x + p, \mu).$$

Usando agora o fato que $g = A^T \lambda - Hp$ e $Ap = 0$ temos

$$p^T g = -p^T Hp = -\|p\|_H^2.$$

Substituindo esta expressão na última desigualdade:

$$\phi_B(x, \mu) - \phi_B(x + p, \mu) \leq -p^T g = \|p\|_H^2. \quad (6.60)$$

Sejam $x^0 := x$, $p^0 := p$; se x^1, x^2, \dots denota a seqüência de pontos obtidos pela repetição de passos de Newton $x^k = x^{k-1} + p^{k-1}$, começando em x^0 , com $k = 0, 1, 2, \dots$, então pelo Lema 6.4.1,

$$\|p^k\|_H \leq \frac{4}{9} \left(\frac{9}{4} \|p\|_H \right)^{2^k}.$$

Retomando a (6.60) podemos escrever

$$\begin{aligned} \phi_B(x, \mu) - \phi_B(x(\mu), \mu) &= \sum_{k=0}^{\infty} \left(\phi_B(x^k, \mu) - \phi_B(x^{k+1}, \mu) \right) \\ &\leq \sum_{k=0}^{\infty} \|p^k\|_H^2 \\ &\leq \sum_{k=0}^{\infty} \left(\frac{4}{9} \left(\frac{9}{4} \|p\|_H \right)^{2^k} \right)^2 \\ &= \left(\frac{4}{9} \right)^2 \sum_{k=0}^{\infty} \left(\frac{9}{4} \|p\|_H \right)^{2^{k+1}} \\ &\leq \frac{\|p\|_H^2}{1 - (\frac{9}{4})^2 \|p\|_H^2} \quad \blacksquare \end{aligned}$$

O seguinte Lema fornece um limite superior para a diferença de valores da função objetivo de um ponto centrado x e seu centro $x(\mu)$.

Lema 6.4.4 *Se a direção projetada de Newton p satisfaz $\|p\|_H < \frac{1}{3}$ então*

$$|c^T x - c^T x(\mu)| \leq \frac{1 + \frac{9}{4} \|p\|_H}{1 - \frac{9}{4} \|p\|_H} \mu \sqrt{\frac{3n}{2}} \|p\|_H. \quad (6.61)$$

Demonstração. Como $g = c/\mu + \nabla B(x)$, temos

$$c^T(x+p) - c^T x = c^T p = \mu(g^T p - \nabla B(x)^T p) = \mu(-\|p\|_H^2 - \nabla B(x)^T p).$$

Tomando valor absoluto e usando a desigualdade triangular:

$$|c^T(x+p) - c^T x| \leq \mu(\|p\|_H^2 + |\nabla B(x)^T p|).$$

Da propriedade de auto-concordância da barreira B ($|\nabla B(x)^T p| \leq \sqrt{\frac{3}{2}n}\|p\|_H$), segue-se:

$$\begin{aligned} |c^T(x+p) - c^T x| &\leq \mu(\|p\|_H^2 + \sqrt{\frac{3}{2}n}\|p\|_H) \\ &\leq \sqrt{\frac{3}{2}n}\mu\|p\|_H(1 + \|p\|_H). \end{aligned}$$

De novo, sejam $x := x^0$, $d := d^0$ e x^1, x^2, \dots, x^k a seqüência de pontos obtidos pela repetição de passos do método de Newton, começando por x^0 . Temos

$$\begin{aligned} |c^T x(\mu) - c^T x| &\leq \sum_{k=0}^{\infty} |c^T x^k - c^T x^{k+1}| \\ &\leq \sqrt{\frac{3}{2}n}\mu \sum_{k=0}^{\infty} \|p^k\|_H(1 + \|p^k\|_H). \end{aligned}$$

Usando a propriedade

$$\|p^k\|_H \leq \frac{4}{9} \left(\frac{9}{4}\|p\|_H\right)^{2^k} < \frac{9}{4}\|p\|_H,$$

obtemos

$$\begin{aligned} |c^T x(\mu) - c^T x| &\leq \sqrt{\frac{3}{2}n}\mu(1 + \frac{9}{4}\|p\|_H) \sum_{k=0}^{\infty} \|p^k\|_H \\ &\leq \frac{1+\frac{9}{4}\|p\|_H}{1-\frac{9}{4}\|p\|_H} \sqrt{\frac{3}{2}n}\mu\|p\|_H \quad \blacksquare \end{aligned}$$

Notamos que, se $\|p\|_H = 0$, então os lemas 6.4.3 e 6.4.4 implicam $\phi_B(y, \mu) = \phi(x(\mu), \mu)$ e $c^T x = c^T x(\mu)$.

6.4.2 Um Algoritmo de Passos Longos Seguidor da Trajetória Central Primal

A seguir apresentamos um algoritmo de passos longos seguidor da trajetória central o qual analisamos na próxima seção. Os escalares τ, ϵ e θ são parâmetros que devem ser especificados. Em nosso caso, usamos $\tau = 1/3$, condição que assegura a convergência polinomial de

nosso algoritmo. Este algoritmo também requer uma solução inicial interior x^0 de (6.35), e um valor inicial μ_0 , do parâmetro da barreira. Descrevemos agora o algoritmo STPL (Seguidor da Trajetória Central por Passos Longos) para encontrar uma ϵ -solução.

Algoritmo (STPL):

Dados de Entrada

$0 < \epsilon$ é o parâmetro de precisão;

$\tau = \frac{1}{3}$ é o parâmetro de proximidade da trajetória central;

θ é o parâmetro de redução ($0 < \theta < 1$);

$0 < \mu_0$ é valor inicial do parâmetro da barreira;

x^0 é o ponto interior viável tal que $\|p(x^0, \mu_0)\|_{H(x^0, \mu_0)} < \tau$

Início

$x := x_0; \mu := \mu_0;$

enquanto $\mu > \frac{\epsilon}{6n}$ **fazer**

início (passos externos)

$\mu := (1 - \theta)\mu;$

enquanto $\|p\|_H \geq \tau$ **fazer**

início (passos internos)

$\bar{\alpha} := \arg \min_{\alpha > 0} \{\phi_B(x + \alpha p, \mu) : x + \alpha p \in F^0\}$

$x := x + \bar{\alpha} p$

fim(passos intenos)

fim(passos externos)

Fim.

O comprimento de passo α usado nas iterações internas pode ser qualquer valor tal que $x + \alpha p$ seja uma solução interior de (6.35). O mínimo decréscimo de $\phi_B(x, \mu)$ é garantido no Lema 6.4.8, que será enunciado e provado depois, assumindo que o valor inicial (x^0, μ_0) satisfaz o critério de proximidade $\|p\|_H \leq \tau$. Para encontrar um ponto inicial que satisfaz tal critério, referimos o leitor Renegar [74], Monteiro e [54] e Güler et al. [37].

6.4.3 Análise da Complexidade

Nesta subseção obtemos a complexidade do algoritmo STPL. No algoritmo usaremos $\tau = 1/3$. Começamos com uma propriedade de auto-concordância da barreira.

Lema 6.4.5 *Seja F uma ν -barreira auto-concordante. Então para qualquer $x \in \text{dom}F$, $y \in c1(\text{dom}F)$ temos*

$$(y - x)^T \nabla F(x) \leq \nu. \quad (6.62)$$

Demonstração. Ver Nesterov [57], Teorema 4.2.4. ■

Lema 6.4.6 *Seja $\mu > 0$, e z^* o custo ótimo para o problema (6.35). Temos que*

$$c^T x(\mu) - z^* \leq \frac{3}{2} n \mu. \quad (6.63)$$

Demonstração. Devido ao Lema anterior, a propriedade de B ser $(3n/2)$ -barreira auto-concordante, e a expressão $g = (c/\mu) + \nabla B(x)$, temos o resultado. ■

O próximo teorema fornece um limite superior para o número de iterações externas.

Teorema 6.4.7 *No pior caso, após*

$$\frac{1}{\theta} \ln \left(\frac{6n\mu_0}{\epsilon} \right) \quad (6.64)$$

iterações externas, o algoritmo (STPL) finaliza com uma solução $x \in F^0$ verificando $c^T x - z^ \leq \epsilon$.*

Demonstração. O algoritmo para quando $\mu_K = (1 - \theta)^K \mu_0 \leq \epsilon/6n$. Equivalentemente, tomando logaritmos,

$$-K \ln(1 - \theta) \geq \ln \frac{6n\mu_0}{\epsilon}.$$

Desde que $\theta \leq -\ln(1 - \theta)$, cumpre-se

$$K \geq \frac{1}{\theta} \ln \frac{6n\mu_0}{\epsilon}.$$

Mostraremos agora que as diferenças entre $c^T x$ e z^* é menor do que ϵ .

$$c^T x - z^* = (c^T x - c^T x(\mu_k)) + (c^T x(\mu_k) - z^*)$$

$$\begin{aligned}
&\leq \frac{1+\frac{9}{4}\|p\|_H}{1-\frac{9}{4}\|p\|_H} \mu_K \sqrt{\frac{3n}{2}} \|p\|_H + \frac{3n}{2} \mu_K \\
&\leq \frac{3}{2} \mu_K \left(\frac{7}{3} \sqrt{n} + n \right) \\
&\leq \frac{\epsilon}{4n} \left(\frac{7}{3} \sqrt{n} + n \right) \\
&\leq \epsilon,
\end{aligned}$$

onde a primeira desigualdade é devida ao Lema 6.4.3 e Lema 6.4.4. ■

O lema que se segue é uma adaptação para nossa barreira de uma propriedade geral de barreiras auto-concordantes.

Lema 6.4.8 *Seja $\bar{\alpha} = (1 - \|p\|_H)^{-1}$, então*

$$\Delta = \phi_B(x, \mu) - \phi_B(x + \bar{\alpha}, \mu) \geq \|p\|_H - \ln(1 + \|p\|_H) > 0 \quad (6.65)$$

Demonstração. Ver den Hertog [39], Lema 2.24. ■

O teorema a seguir fornece um limite superior para o número total de iterações internas em cada iteração externa.

Teorema 6.4.9 *Cada iteração externa requer, no pior caso,*

$$\frac{22\theta}{(1-\theta)^2} \left(\frac{7}{3} \sqrt{\frac{3n}{2}} + \frac{3n}{2} \right) + \frac{22}{3} \quad (6.66)$$

iterações internas.

Demonstração. Denotamos o valor do parâmetro da barreira numa iteração externa arbitrária por $\bar{\bar{\mu}}$, enquanto o valor do parâmetro na iteração prévia (externa) é denotado por $\bar{\mu}$. O iterado ao início da iteração externa é denotado por x . Vê-se que x é centrado com respeito a $x(\bar{\mu})$, e $\bar{\bar{\mu}} = (1 - \theta) \bar{\mu}$. Pelo Lema 6.4.8, durante cada iteração interna o decréscimo do valor da função barreira é pelo menos

$$\Delta = \frac{1}{3} - \ln\left(1 + \frac{1}{3}\right) > \frac{1}{22}.$$

Seja N o número de iterações internas durante uma iteração externa. Temos

$$\frac{N}{22} < N\Delta \leq \phi_B(x, \bar{\bar{\mu}}) - \phi_B(x(\bar{\mu}), \bar{\mu}). \quad (6.67)$$

Agora definimos a função

$$\Psi_B(x, \bar{\mu}) := \phi_B(x, \bar{\mu}) - \phi_B(x(\bar{\mu}), \bar{\mu}).$$

De acordo com o Teorema do Valor Médio existe $\tilde{\mu} \in (\bar{\mu}, \bar{\mu})$ tal que

$$\Psi_B(x, \bar{\mu}) = \Psi_B(x, \bar{\mu}) - \frac{d}{d\mu} \Psi_B(x, \mu) \Big|_{\mu=\tilde{\mu}} (\bar{\mu} - \bar{\mu}). \quad (6.68)$$

Lembrando que $\phi_B(x, \mu) = (c^T x / \mu) + B(x)$, é imediato que

$$\frac{d}{d\mu} \phi_B(x, \mu) = -\frac{c^T x}{\mu^2},$$

e, denotando as derivadas de $x(\mu)$ com respeito a μ por x' ,

$$\frac{d}{d\mu} \phi_B(x(\mu), \mu) = -(c^T x(\mu)) / \mu^2 + (c^T x'(\mu)) / \mu + \nabla B(x(\mu))^T x'(\mu).$$

Como $\nabla B(x(\mu)) = 2[\ln X(\mu) - \ln(I - X(\mu))]e - (1/\mu)s(\mu) + (1/\mu)w(\mu)$ então

$$\begin{aligned} \frac{d}{d\mu} \phi_B(x(\mu), \mu) &= -(1/\mu^2)c^T x(\mu) + (1/\mu)(c + 2\mu[\ln X(\mu) - \ln(I - X(\mu))]e - s(\mu) + w(\mu))^T x'(\mu) \\ &= -(1/\mu^2)c^T x(\mu) + (1/\mu)(A^T y(\mu))^T x'(\mu) \\ &= -(1/\mu^2)c^T x(\mu) + (1/\mu)y(\mu)^T Ax'(\mu). \end{aligned}$$

Devido a que $Ax'(\mu) = 0$ obtemos

$$\frac{d}{d\mu} \phi_B(x(\mu), \mu) = -\frac{c^T x(\mu)}{\mu^2}.$$

Portanto

$$\begin{aligned} -\frac{d}{d\mu} \Psi_B(x, \mu) \Big|_{\mu=\tilde{\mu}} &= \left(\frac{c^T x - c^T x(\mu)}{\mu^2} \right) \Big|_{\mu=\tilde{\mu}} \\ &\leq \left| \frac{c^T x - c^T x(\bar{\mu})}{\bar{\mu}^2} \right|, \end{aligned}$$

onde a última desigualdade segue do fato que $\bar{\mu} < \tilde{\mu}$ e $c^T x(\bar{\mu}) \leq c^T x(\tilde{\mu})$.

Substituindo isto em (6.68) dá

$$\begin{aligned} \Psi_B(x, \bar{\mu}) &\leq \Psi_B(x, \bar{\mu}) + \frac{c^T x - c^T x(\bar{\mu})}{\bar{\mu}^2} (\bar{\mu} - \bar{\mu}) \\ &\leq \Psi_B(x, \bar{\mu}) + \left(\frac{|c^T x(\bar{\mu}) - c^T x|}{\bar{\mu}} + \frac{c^T x(\bar{\mu}) - c^T x(\bar{\mu})}{\bar{\mu}} \right) \frac{\bar{\mu} - \bar{\mu}}{\bar{\mu}}. \end{aligned} \quad (6.69)$$

Devido a que $\|p(x, \bar{\mu})\|_{H(x, \bar{\mu})} < \frac{1}{3}$, pelo Lema 6.4.3 obtemos

$$\Psi_B(x, \bar{\mu}) \leq \frac{1}{3}.$$

Agora, pelo Lema 6.4.4 e pelo fato de que $\|p\|_H < \frac{1}{3}$:

$$|c^T x - c^T x(\bar{\mu})| \leq \frac{7}{3} \sqrt{\frac{3}{2}n} \bar{\mu},$$

temos também

$$\frac{c^T(\bar{\mu}) - c^T x(\bar{\mu})}{\bar{\mu}} \leq \frac{c^T x(\bar{\mu}) - z^*}{\bar{\mu}} \leq \frac{\frac{3}{2}n \bar{\mu}}{\bar{\mu}}.$$

Juntando todos estes limites superiores em (6.69) obtemos

$$\Psi_B(x, \bar{\mu}) \leq \frac{1}{3} + \left(\frac{7}{3} \sqrt{\frac{3}{2}n} + \frac{3}{2}n \right) \frac{\theta}{(1-\theta)^2}. \quad (6.70)$$

Substituindo o último limite em (6.67) temos

$$\begin{aligned} N &\leq 22 \left(\frac{1}{3} + \left(\frac{7}{3} \sqrt{\frac{3}{2}n} + \frac{3}{2}n \right) \frac{\theta}{(1-\theta)^2} \right) \\ &\leq \frac{22\theta}{(1-\theta)^2} \left(\frac{7}{3} \sqrt{\frac{3}{2}n} + \frac{3}{2}n \right) + \frac{22}{3} \quad \blacksquare \end{aligned}$$

Finalmente, podemos agregar os teoremas 6.4.7 e 6.4.9, obtendo o número total de iterações de Newton requerida para nosso algoritmo:

Teorema 6.4.10 *O limite superior para o número total de iterações de Newton é dado por*

$$\left[\frac{22}{(1-\theta)^2} \left(\frac{7}{3} \sqrt{\frac{3}{2}n} + \frac{3}{2}n \right) + \frac{22}{3\theta} \right] \ln \left(\frac{6n\mu_0}{\epsilon} \right). \quad (6.71)$$

É claro que para obter uma ϵ -solução o algoritmo precisa de $O(n \ln(n\mu_0/\epsilon))$ iterações de Newton para variantes de passos longos ($0 < \theta < 1$).

6.5 Um Método de Ponto Proximal

Nesta Seção introduzimos um método de ponto proximal para resolver o problema (6.35) baseada na distância de Bregman induzida pela barreira B , isto é,

$$D_B(z, y) = B(z) - B(y) - \nabla B(y)^T (y - z).$$

Este método gera uma seqüência $\{z^k\} \subset (0, 1)^n$ definido por

$$z^0 \in (0, 1)^n \text{ tal que } \nabla B(z^0) \in \text{Im}(A^T), \quad (6.72)$$

$$z^{k+1} = \arg \min_{z \in \mathbb{R}^n} \{c^T z + \lambda_k D_B(z, z^k) : Az = b, 0 \leq z \leq e\} \quad (6.73)$$

onde $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfaz

$$\sum_{k=0}^{\infty} (1/\lambda_k) = \infty \quad (6.74)$$

e

$$D_B(z, z^k) = \sum_{i=1}^n \left[(2z - 1) \left(\ln(z/z^k) + \ln((1-z)/(1-z^k)) \right) + (z/z^k) + (1-z)/(1-z^k) - 2 \right].$$

Desde que a função objetivo de (6.73) é estritamente convexa sobre F^0 , e é ilimitada $(+\infty)$ na fronteira de F , a seqüência $\{z^k\}$ gerada por (6.72)-(6.74) é bem definida e única (para cada k). Potanto, $z^k \in F^0$.

Como z^{k+1} é a solução do problema (6.73) então existe $y^k \in \mathbb{R}^m$ tal que

$$c + \lambda_k (\nabla B(z^{k+1}) - \nabla B(z^k)) = A^T y^k \quad (6.75)$$

É facil provar, usando a hipótese de que $\nabla B(z^0) \in \text{Im}(A^T)$, que a trajetória central primal $X(\mu)$, como foi definido na seção anterior, é a única solução do seguinte problema

$$\min_{z \in \mathbb{R}^n} \{c^T z + \mu D_B(z, z^0) : Az = b, 0 \leq z \leq e.\}$$

Assim $x(\mu)$ satisfaz

$$c + \mu (\nabla B(x(\mu)) - \nabla B(z^0)) = A^T w(\mu) \quad (6.76)$$

para algum $w(\mu) \in \mathbb{R}^m$.

O próximo resultado é uma conseqüência da convergência da trajetória central.

Teorema 6.5.1 *A seqüência $\{z^k\}$ gerada por (6.72)-(6.74) converge ao centro analítico do conjunto de soluções ótimas do problema (6.35).*

Demonstração. Seja $\mu_k = 1/(\sum_{j=0}^{k-1} (1/\lambda_j))$. Obviamente $\{\mu_k\}$ é uma seqüência decrescente para cada $k \geq 1$ e converge para zero quando k vai para infinito. De (7.46) temos

$$c + \mu_k (\nabla B(x(\mu_k)) - \nabla B(z^0)) = A^T w(\mu_k),$$

$$c + \mu_{k+1}(\nabla B(x(\mu_{k+1})) - \nabla B(z^0)) = A^T w(\mu_{k+1}),$$

para alguma seqüência $\{w(\mu_k)\} \subset \mathbb{R}^m$. Estas equações implicam que

$$c + \lambda_k(\nabla B(x(\mu_{k+1})) - \nabla B(x(\mu_k))) = A^T y^k,$$

onde $y^k = \lambda_k(\mu_{k+1}^{-1}w(\mu_{k+1}) - \mu_k^{-1}w(\mu_k))$. Devido à unicidade do mínimo de (7.45) podemos garantir que $z^k = x(\mu_k)$.

Assim,

$$\lim_{k \rightarrow \infty} z^k = \lim_{k \rightarrow \infty} x(\mu_k) = x^*$$

onde x^* é o centro analítico do problema (6.35) e a última igualdade é consequência do Teorema 6.3.3. ■

Capítulo 7

Nova Barreira para uma Classe de Problemas de Otimização Semidefinida

7.1 Introdução

Neste capítulo introduziremos uma nova barreira para resolver uma classe de problemas de otimização semidefinida, proporemos um problema auxiliar penalizado associado à barreira e definiremos a trajetória central obtida a partir das suas condições de otimalidade. A seguir estudaremos as propriedades de convergência da trajetória central primal e primal-dual e apresentaremos um resultado de convergência global quando a função objetivo é linear. Finalmente introduzimos um método de ponto proximal com uma distância de Bregman gerada pela nova barreira para resolver problemas lineares semidefinidos e provamos sua convergência.

7.2 Definição do Problema

Neste capítulo estamos interessados em resolver o seguinte problema de otimização:

$$\min_{X \in \mathcal{S}^n} \{f(X) : \mathcal{A}X = b, 0 \preceq X \preceq I\} \quad (7.1)$$

onde $f : \mathcal{S}^n \rightarrow \mathbb{R}$ é uma função real, \mathcal{S}^n é o espaço vetorial das matrizes simétricas, $I \in \mathcal{S}^n$ é a matriz identidade, o operador $\mathcal{A} : \mathcal{S}^n \rightarrow \mathbb{R}^m$ é definido por $\mathcal{A}X := (A_i \bullet X)_{i=1}^m \in \mathbb{R}^m$,

onde $A_i \in \mathcal{S}^n$; $X \in \mathcal{S}^n$, $X \succeq 0$ significa que X é semidefinida positiva, e $X \preceq I$ significa que $I - X \succeq 0$. Denotamos por

$$F = \{X \in \mathcal{S}^n : \mathcal{A}X = b, 0 \preceq X \preceq I\}$$

o conjunto viável de (7.1) e

$$F^0 = \{X \in \mathcal{S}^n : \mathcal{A}X = b, 0 \prec X \prec I\}$$

seu interior relativo.

Impomos as seguintes hipóteses ao problema (7.1):

1. O conjunto F^0 é não vazio.
2. f é uma função convexa sobre \mathcal{S}^n .
3. As matrizes A_i são linearmente independentes.

Devido a que f é uma função contínua sobre o conjunto compacto F , o ponto de mínimo global é atingido em F . Além disso, a convexidade de f implica que todo mínimo local é mínimo global. Assim, o conjunto de soluções ótimas do problema (7.1), que denotaremos por $\text{Sol}(P)$, é um conjunto não-vazio convexo e limitado.

A formulação dual de Wolfe para o problema (7.1) é:

$$\begin{aligned} \max d(X, y, S, W) &= f(X) - y^T(\mathcal{A}X - b) + W \bullet (X - I) - S \bullet X \\ \mathcal{A}^*y + S - W &= \nabla f(X) \\ W, S &\succeq 0 \\ y \in \mathbb{R}^m; X, S, W &\in \mathcal{S}^n \end{aligned} \tag{7.2}$$

Um ponto viável (X, y, S, W) para (7.2), com $S, W \succ 0$ é chamada uma solução dual viável interior.

7.3 Exemplos

Nesta seção apresentamos alguns exemplos de problemas encontrados na literatura que podem ser escritos na forma (7.1) para o caso linear.

Exemplo 7.1 *Soma dos maiores autovalores.*

Overton e Womersley [63] (Teorema 3.4, pagina 329), apresentaram a seguinte caracterização para a soma dos k primeiros maiores autovalores de uma matriz simétrica:

$$\lambda_1(A) + \dots + \lambda_k(A) = \max\{A \bullet X : \text{Tr}X = k, 0 \preceq X \preceq I\}, \quad (7.3)$$

onde $\lambda_j(A), j = 1, \dots, k$, denota o j -ésimo maior autovalor de A . Podemos escrever (7.3) como $\max\{A \bullet X : \mathcal{A}X = b, 0 \preceq X \preceq I\}$, onde $\mathcal{A}X = I \bullet X$ e $b = k$. Portanto, (7.3) é um caso particular de (7.1).

Exemplo 7.2 *Mínimo da soma dos maiores autovalores.*

Consideremos o problema de minimizar a soma dos primeiros k maiores autovalores de uma matriz simétrica:

$$\min \lambda_1(A(y)) + \dots + \lambda_k(A(y)) \quad \text{onde} \quad A(y) = A_0 + \sum_{i=1}^m y_i A_i, \quad y \in \mathbb{R}^m. \quad (7.4)$$

Pode ser provado (ver Alizadeh [2], Teorema 4.3), que o dual deste problema é:

$$\max\{A_0 \bullet X : \text{Tr}X = k, \quad A_j \bullet X = 0 \quad \text{para} \quad j = 1, \dots, m, \quad 0 \preceq X \preceq I\}. \quad (7.5)$$

Note que podemos escrever (7.5) como $\max\{A_0 \bullet X : \mathcal{A}X = b, \quad 0 \preceq X \preceq I\}$, onde $\mathcal{A}X = (I \bullet X, A_1 \bullet X, \dots, A_m \bullet X)^T$ e $b = (k, 0, 0, \dots, 0) \in \mathbb{R}^{m+1}$. Portanto, isto é uma aplicação de (7.1).

Exemplo 7.3 *Minimização da soma ponderada dos autovalores.* Considere o seguinte problema:

$$\min m_1 \lambda_1(A) + \dots + m_k \lambda_k(A) \quad \text{onde} \quad m_1 \geq \dots \geq m_k > 0. \quad (7.6)$$

Observe que devido à condição $m_1 \geq \dots \geq m_k > 0$, este problema é convexo. Donath e Hoffman em [20] reformularam esta soma como segue:

$$m_1 \lambda_1 + \dots + m_k \lambda_k = (m_1 - m_2) \lambda_1 + (m_2 - m_3) (\lambda_1 + \lambda_2) + \dots + (m_{k-1} - m_k) (\lambda_1 + \dots + \lambda_{k-1}) + m_k (\lambda_1 + \dots + \lambda_k).$$

Para cada soma parcial de autovalores no lado direito da igualdade podemos usar a formulação do exemplo 2, obtendo:

$$\begin{aligned} \min & (m_1 - m_2) X_1 \bullet A + (m_2 - m_3) X_2 \bullet A + \dots + m_k X_k \bullet A \\ \text{s.a:} & \quad \text{Tr}(X_i) = i \quad \text{for} \quad i = 1, \dots, k \\ & \quad 0 \preceq X_i \preceq I. \end{aligned}$$

Agora observe que este problema pode ser posto na forma (7.1), onde

$C = \text{diag}((m_1 - m_2)A_1, (m_2 - m_3)A_2, \dots, m_k A_k)$, $X = \text{diag}(X_1, X_2, \dots, X_n)$, $\mathcal{A}(X) = I \bullet X$ and $b = (1, 2, \dots, k)^T$. Assim, (7.6) é um caso particular de (7.1).

Exemplo 7.4 O problema de partição de um grafo. Um importante caso do problema de partição de um grafo, ver [2], pode ser formulado como

$$\min\{C \bullet X: X_{ii} = k/n, \quad 0 \preceq X \preceq I\}, \quad (7.7)$$

Podemos escrever (7.7) como (7.1) com $A_1 = \text{diag}(1, 0, \dots, 0)$, $A_2 = \text{diag}(0, 1, 0, \dots, 0), \dots$, $A_n = (0, 0, \dots, 0, 1)$ e $b = (k/n, k/n, \dots, k/n) \in \mathbb{R}^n$.

7.4 Uma Nova Barreira e sua Trajetória Central

Nesta seção introduzimos uma nova barreira e propomos um novo problema auxiliar de penalidade interior para resolver (7.1) e estudamos a conduta da trajetória central (primal-dual), obtida pelas suas condições de otimalidade (KKT). Observaremos que uma propriedade desta trajetória é que é primal viável, com respeito a (7.1), e dual inviável, com respeito ao problema dual (7.2), pelo termo $\mu[\ln X(\mu) - \ln(I - X(\mu))]$, que converge para zero quando μ converge a zero (ver Corolário 7.4.1). Provaremos também, usando algumas ideias de geometria semi-analítica, que no caso linear esta trajetória converge.

7.4.1 A Nova Barreira

Começamos denotando o *cubo matricial* como $\mathcal{S}_{[0,I]}^n = \{X \in \mathcal{S}^n : 0 \preceq X \preceq I\}$ e seu interior relativo como $\mathcal{S}_{(0,I)}^n$. Definimos sobre $\mathcal{S}_{(0,I)}^n = \{X \in \mathcal{S}^n : 0 \prec X \prec I\}$ a seguinte função:

$$B(X) = \text{tr} [(2X - I)(\ln X - \ln(I - X))]. \quad (7.8)$$

Esta função generaliza para $\mathcal{S}_{(0,I)}^n$ a função barreira auto-concordante para o hipercubo introduzida no capítulo anterior. Observemos que $B(X)$ é invariante sob decomposição da matriz X . Além disso, podemos escrever $B(X)$ como:

$$B(X) = \sum_{i=1}^n (2\lambda_i(X) - 1)(\ln \lambda_i(X) - \ln(1 - \lambda_i(X))), \quad (7.9)$$

onde $0 < \lambda_i(X) < 1$, $i = 1, \dots, n$, são os autovalores de X .

A seguir apresentamos algumas propriedades desta função:

- B pode ser expressada em função de seus autovalores:

$$B(X) = \sum_{i=1}^n (2\lambda_i(X) - 1)(\ln \lambda_i(X) - \ln(1 - \lambda_i(X))), \quad (7.10)$$

onde $0 \leq \lambda_i(X) \leq 1$ são os autovalores de X .

- $B(X) \geq 0$, para todo $X \in \mathcal{S}_{(0,I)}^n$
- $B(X)$ é estritamente convexa e infinitamente diferenciável sobre $\mathcal{S}_{(0,I)}^n$.
- Se $X \rightarrow \text{front}\mathcal{S}_{(0,I)}^n$ (X se aproxima da fronteira, isto é, $\lambda_i(X) \rightarrow 0$ ou $\lambda_i(X) \rightarrow 1$), então $B(X) \rightarrow \infty$. Portanto, B é uma função barreira sobre $\mathcal{S}_{(0,I)}^n$.
- $B'(X)[H] = [2(\ln X - \ln(I - X)) - X^{-1} + (I - X)^{-1}] \bullet H$ (Derivada de Gateaux na direção H).
- $B''(X)[J, H] = 2[X^{-1} + (I - X)^{-1}]J \bullet H + X^{-1}JX^{-1} \bullet H + (I - X)^{-1}J(I - X)^{-1} \bullet H$ (Hessiana de Gateaux nas direções de J e H).

7.4.2 Definição da Trajetória Central

Para resolver o problema (7.1) propomos o seguinte problema penalizado:

$$\begin{aligned} \min \quad & \phi_B(X, \mu) = f(X) + \mu \text{tr}[(2X - I)(\ln X - \ln(I - X))] \\ & \mathcal{A}X = b \\ & (0 \prec X \prec I) \end{aligned} \quad (7.11)$$

onde $\mu > 0$ é um parâmetro. As derivadas de primeira e segunda ordem de ϕ_B são, respectivamente:

$$g(x, \mu) := \nabla f(X) + 2\mu[(\ln X - \ln(I - X)) - X^{-1} + (I - X)^{-1}] \quad (7.12)$$

$$\nabla^2 \phi_B(X, \mu) := \nabla^2 f(X) + 2\mu[X^{-1} + (I - X)^{-1}] + \mu(X^{-1} \circ X^{-1}) + \mu((I - X)^{-1} \circ (I - X)^{-1}). \quad (7.13)$$

onde $X^{-1} \circ X^{-1}$, $[X^{-1} + (I - X)^{-1}]$ e $(I - X)^{-1} \circ (I - X)^{-1}$ são operadores definidos respectivamente por

$$(X^{-1} \circ X^{-1})Z = X^{-1}ZX^{-1}$$

$$[X^{-1} + (I - X)^{-1}]Z = X^{-1}Z + (I - X)^{-1}Z$$

$$((I - X)^{-1} \circ (I - X)^{-1})Z = (I - X)^{-1}Z(I - X)^{-1}.$$

Desde que $\phi_B(x, \mu)$ é estritamente convexa (porque f é uma função convexa e B é estritamente convexa) sobre o interior relativo do conjunto viável, e é ilimitada $(+\infty)$ na fronteira de F , esta função atinge o valor mínimo no seu domínio (para cada valor de μ fixo) em um único ponto. Seja $X(\mu)$ este ponto. As condições de otimalidade KKT, necessárias e suficientes de primeira ordem, para $X(\mu)$ são:

$$\mathcal{A}^*y + S - W = \nabla f(X) + 2\mu[\ln X - \ln(I - X)] \quad (7.14)$$

$$\mathcal{A}X = b \quad (7.15)$$

$$S - \mu X^{-1} = 0 \quad (7.16)$$

$$W - \mu(I - X)^{-1} = 0 \quad (7.17)$$

$$S, W \succ 0 \quad (7.18)$$

$$(0 \prec X \prec I) \quad (7.19)$$

A única solução deste sistema, denotada por $(X(\mu), y(\mu), S(\mu), W(\mu)), \mu > 0$, chamaremos de **trajetória central primal-dual**. Claramente esta trajetória é primal viável e dual inviável com respeito aos problemas (7.1) e (7.2) respectivamente.

A seguir provamos que a trajetória contínua descrita pela trajetória central primal-dual é uma curva regular.

Teorema 7.4.1 *A aplicação $\mu \rightarrow (x(\mu), y(\mu), s(\mu), w(\mu)), \mu > 0$, é continuamente diferenciável.*

Demonstração. A prova segue do Teorema da Função Implícita, apresentado no Capítulo 2, sempre que provemos que as equações que definem a trajetória central sejam diferenciáveis e sua derivada (com respeito a (X, y, S, W)) seja não singular em todo ponto da trajetória.

Obviamente as equações da trajetória central primal-dual são diferenciáveis faltando portanto provar a segunda parte.

Seja $F : \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R} \longrightarrow \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n \times \mathcal{S}^n$ definido por:

$$F(X, y, S, W, \mu) = \begin{bmatrix} \nabla f(X) + 2\mu[\ln X - \ln(I - X)] - \mathcal{A}^*y - S + W \\ \mathcal{A}X - b \\ S - \mu X^{-1} \\ W - \mu(I - X)^{-1} \end{bmatrix} \quad (7.20)$$

Derivando F com respeito a (X, y, S, W) obtemos

$$J_F((X, y, S, W)) = \begin{bmatrix} \nabla^2 f(X) + 2\mu[X^{-1} + (I - X)^{-1}] & -\mathcal{A}^* & -\mathcal{I}_n & \mathcal{I}_n \\ \mathcal{A} & 0_{m \times m} & 0_{m \times n} & 0_{m \times n} \\ \mu(X^{-1} \circ X^{-1}) & 0_{n \times m} & \mathcal{I}_n & 0_{n \times n} \\ -\mu((I - X)^{-1} \circ (I - X)^{-1}) & 0_{n \times m} & 0_{n \times n} & \mathcal{I}_n \end{bmatrix}$$

Desejamos mostrar que J_F é não singular, assim é suficiente verificar que o espaço nulo tem um único elemento (que é o zero).

Sejam $U, T, P \in \mathbb{R}^{n \times n}$ e $v \in \mathbb{R}^m$. Considere a equação:

$$\nabla^2 f(X)U + 2\mu[X^{-1} + (I - X)^{-1}]U - \mathcal{A}^*v - T + P = 0 \quad (7.21)$$

$$\mathcal{A}U = 0 \quad (7.22)$$

$$\mu(X^{-1} \circ X^{-1})U + T = 0 \quad (7.23)$$

$$-\mu[(I - X)^{-1} \circ (I - X)^{-1}]U + P = 0 \quad (7.24)$$

Da equação (7.23) e (7.24) tem-se:

$$T = -\mu(X^{-1} \circ X^{-1})U \quad (7.25)$$

$$P = \mu[(I - X)^{-1} \circ (I - X)^{-1}]U \quad (7.26)$$

Substituindo (7.25) e (7.26) em (7.21):

$$\nabla^2 f(X)U + 2\mu[(X^{-1} + (I - X)^{-1}) + \mu(X^{-1} \circ X^{-1}) + \mu((I - X)^{-1} \circ (I - X)^{-1})] - \mathcal{A}^*v = 0,$$

isto é (ver (7.13)),

$$\nabla^2 \phi_B(X, \mu)U - \mathcal{A}^*v = 0. \quad (7.27)$$

Aplicando $\nabla^2 \phi_B(X, \mu)^{-1}$ em (7.27):

$$U + \nabla^2 \phi_B(X, \mu)^{-1} \mathcal{A}^*v = 0,$$

e multiplicando por \mathcal{A} temos

$$\mathcal{A} \nabla^2 \phi_B(X, \mu)^{-1} \mathcal{A}^*v = 0.$$

Desde que as matrizes A_i são linearmente independentes $\mathcal{A} \nabla^2 \phi_B(X, \mu)^{-1} \mathcal{A}^*$ é invertível, por conseguinte

$$v = 0. \quad (7.28)$$

Substituindo (7.28) em (7.27) temos $\nabla^2 \phi_B(X, \mu)U = 0$, e

$$U = 0. \quad (7.29)$$

Substituindo este valor em (7.25) e (7.26):

$$T = 0 \quad e \quad P = 0. \quad (7.30)$$

De (7.28), (7.29) e (7.30), concluímos que J_F é não singular sobre a trajetória central. Portanto a trajetória central primal-dual $(x(\mu), y(\mu), s(\mu), w(\mu))$ é continuamente diferenciável. ■

Observação 7.4.1 *No caso linear, seguindo os mesmos passos, podemos provar que a aplicação $\mu \rightarrow (x(\mu), y(\mu), s(\mu), w(\mu))$, $\mu > 0$, é analítica, veja [67].*

7.4.3 Propriedades da Trajetória Central Primal

Apresentamos nesta subseção algumas propriedades da trajetória central primal. Estes resultados são extensões fáceis de barreiras gerais não-negativas em \mathbb{R}^n para \mathcal{S}^n .

Lemma 7.4.1 *A função $0 < \mu \rightarrow B(X(\mu))$, onde B é definido por (7.8), é não-crescente.*

Demonstração. Seja $\mu_2 < \mu_1$. Provaremos que $B(X(\mu_1)) \leq B(X(\mu_2))$. Como $X(\mu_1)$ minimiza $\phi_B(X, \mu_1)$ e $X(\mu_2)$ minimiza $\phi_B(X, \mu_2)$, ver (7.11), temos:

$$f(X(\mu_1)) + \mu_1 B(X(\mu_1)) \leq f(X(\mu_2)) + \mu_1 B(X(\mu_2))$$

e

$$f(X(\mu_2)) + \mu_2 B(X(\mu_2)) \leq f(X(\mu_1)) + \mu_2 B(X(\mu_1)).$$

Somando estas desigualdades

$$\mu_2 B(X(\mu_2)) + \mu_1 B(X(\mu_1)) \leq \mu_2 B(X(\mu_1)) + \mu_1 B(X(\mu_2)),$$

isto implica

$$(\mu_1 - \mu_2)B(X(\mu_1)) \leq (\mu_1 - \mu_2)B(X(\mu_2)).$$

Como $\mu_2 < \mu_1$ temos

$$B(X(\mu_1)) \leq B(X(\mu_2)). \quad \blacksquare$$

Lemma 7.4.2 *Se $\mu_2 < \mu_1$ então*

i. $\phi_B(X(\mu_2), \mu_2) \leq \phi_B(X(\mu_1), \mu_1);$

ii. $f(X(\mu_2)) \leq f(X(\mu_1)).$

Demonstração. Desde que $\mu_2 < \mu_1$, $X(\mu_2)$ minimiza $\phi_B(X, \mu_2)$ e $B \geq 0$:

$$\begin{aligned} \phi_B(X(\mu_2), \mu_2) &= f(X(\mu_2)) + \mu_2 B(X(\mu_2)) \\ &\leq f(X(\mu_1)) + \mu_2 B(X(\mu_1)) \\ &\leq f(X(\mu_1)) + \mu_1 B(X(\mu_1)) \\ &= \phi_B(X(\mu_1), \mu_1). \end{aligned}$$

Isto prova *i*. Para provar *ii*, consideremos a seguinte desigualdade

$$f(X(\mu_2)) + \mu_2 B(X(\mu_2)) \leq f(X(\mu_1)) + \mu_2 B(X(\mu_1)).$$

Desde que $B(X(\mu_2)) \leq B(X(\mu_1))$, a desigualdade acima implica

$$f(X(\mu_2)) \leq f(X(\mu_1)). \quad \blacksquare$$

Lemma 7.4.3 *Se X^* é uma solução ótima do problema (7.1), e $\mu > 0$, então*

$$f(X^*) \leq f(X(\mu)) \leq \phi_B(X(\mu), \mu).$$

Demonstração. Desde que X^* é uma solução ótima do problema primal, temos que $f(X^*) \leq f(X(\mu))$. Como $B \geq 0$,

$$f(X(\mu)) \leq f(X(\mu)) + \mu B(X(\mu)),$$

isto implica

$$f(X(\mu)) \leq \phi_B(X(\mu), \mu).$$

Então

$$f(X^*) \leq C \bullet X(\mu) \leq \phi_B(X(\mu), \mu). \quad \blacksquare$$

Proposição 7.4.1 *Qualquer ponto de acumulação da trajetória central primal $\{X(\mu)\}$ é uma solução ótima do problema (7.1).*

Demonstração. Seja \bar{X} um ponto de acumulação de $\{X(\mu)\}$. Note que $\mathcal{A}\bar{X} = b$ e $0 \preceq \bar{X} \preceq I$. Seja $\{\mu_k\}$ uma seqüência positiva de números tal que

$$\lim_{k \rightarrow \infty} \mu_k = 0$$

e

$$\lim_{k \rightarrow \infty} X(\mu_k) = \bar{X}.$$

Fixe $X \in \mathcal{P}$, uma solução viável arbitraria do problema (7.1). Pela hipótese 1, podemos tomar um ponto X^0 primal viável. Consequentemente, para todo $\epsilon \in (0, 1)$ se tem

$$X(\epsilon) = (1 - \epsilon)X + \epsilon X^0 \in \mathcal{P}^0.$$

Da condição de otimalidade para $X(\mu_k)$:

$$f(X(\mu_k)) + \mu_k B(X(\mu_k)) \leq f(X(\epsilon)) + \mu_k B(X(\epsilon)).$$

Isto implica

$$\mu_k [B(X(\mu_k)) - B(X(\epsilon))] \leq f(X(\epsilon)) - f(X(\mu_k)).$$

Como B é convexa, temos que

$$\mu_k [\nabla B(X(\epsilon))(X(\mu_k) - X(\epsilon))] \leq f(X(\epsilon)) - f(X(\mu_k)).$$

Levando ao limite quando $k \rightarrow \infty$, obtemos

$$0 \cdot [\nabla B(X(\epsilon)) \bullet (\bar{X} - X(\epsilon))] \leq f(X(\epsilon)) - f(\bar{X}),$$

isto é,

$$0 \leq f(X(\epsilon)) - f(\bar{X}).$$

Agora, para $\epsilon \rightarrow 0$ e usando a continuidade de f obtemos

$$f(\bar{X}) \leq f(X).$$

Desde que X é uma solução viável arbitrária, a última desigualdade implica que \bar{X} é uma solução ótima de (7.1). ■

Proposição 7.4.2 *Seja $\{\mu_k\}$ uma seqüência de números reais positivos tal que $\mu_k \rightarrow 0$ quando $k \rightarrow \infty$. Então:*

$$\mu_k B(X(\mu_k)) \rightarrow 0, \text{ quando } k \rightarrow \infty.$$

Demonstração. Por simplicidade denotaremos $X(\mu_k)$ por X^k . Seja \bar{X} uma solução ótima do problema primal (7.1). Pelo Lema 7.4.2, **ii**, $\{f(X^k)\}$ é uma seqüência não crescente e limitada inferiormente por $f(\bar{X})$, daqui $\{f(X^k)\}$ converge, isto é, existe $\alpha \in \mathbb{R}$ tal que

$$f(X^k) \rightarrow \alpha, \text{ quando } k \rightarrow \infty. \quad (7.31)$$

Do Lema 7.4.3, temos

$$f(\bar{X}) \leq f(X^k) \leq \phi_B(X^k, \mu_k)$$

e portanto a seqüência $\{\phi_B(X^k, \mu_k)\}$ é limitada inferiormente. Além disso, pelo Lema 7.4.2, **i**, sabemos que ϕ_B é não-crescente. Logo, existe $\zeta^* \in \mathbb{R}$ tal que

$$\phi_B(X^k, \mu_k) \rightarrow \zeta^* \text{ quando } k \rightarrow \infty, \quad (7.32)$$

onde $f(\bar{X}) \leq \zeta^*$.

De (7.31) e (7.32) temos

$$f(X^k) - \phi_B(X^k, \mu_k) \rightarrow \alpha - \zeta^*, \text{ quando } k \rightarrow \infty.$$

Desde que

$$\phi_B(X^k, \mu_k) = f(X^k) + \mu_k B(X^k),$$

obtemos

$$\mu_k B(X^k) \rightarrow \zeta^* - \alpha, \text{ quando } k \rightarrow \infty.$$

Como $B \geq 0$ e ela é não-crescente, é claro que

$$\mu_k B(X^k) \rightarrow 0, \text{ quando } k \rightarrow \infty,$$

já que $\mu_k \rightarrow 0$, quando $k \rightarrow \infty$. ■

Considere agora a seguinte notação:

$$(X^k, y^k, S^k, W^k) = (X(\mu_k), y(\mu_k), S(\mu_k), W(\mu_k)).$$

Lembramos que nos teoremas de convergência que envolvem matrizes, usaremos a topologia dada pela norma de Frobenius, isto é, dizemos que $Z^k \rightarrow Z$, $k \rightarrow \infty$ quando $\|Z^k - Z\|_F \rightarrow 0$, $k \rightarrow \infty$ (onde $\|A\|_F = \text{tr}(AA^T)$).

Corolário 7.4.1 *Seja $\{X^k\}$ a seqüência da trajetória central primal. Temos que:*

$$\mu_k [\ln X^k - \ln(I - X^k)] \rightarrow 0, k \rightarrow \infty.$$

Demonstração. Pela Proposição anterior

$$\mu_k \text{tr}[(2X^k - I)(\ln X^k - \ln(I - X^k))] \rightarrow 0, k \rightarrow \infty.$$

De (7.10), isto é equivalente a:

$$\mu_k \sum_{i=1}^n (2\lambda_i(X^k) - 1) [\ln \lambda_i(X^k) - \ln(1 - \lambda_i(X^k))] \rightarrow 0, k \rightarrow \infty, i = 1, \dots, n,$$

onde $\lambda_i(X^k)$, $i = 1, \dots, n$, são os autovalores de X^k .

Usando o Corolário 6.3.1, do Capítulo 6, temos:

$$\mu_k [\ln \lambda_i(X^k) - \ln(1 - \lambda_i(X^k))] \rightarrow 0, k \rightarrow \infty, i = 1, \dots, n.$$

elevando ao quadrado e aplicando somatório:

$$\mu_k^2 \sum_{i=1}^n [\ln \lambda_i(X^k) - \ln(1 - \lambda_i(X^k))]^2 \rightarrow 0, \quad k \rightarrow \infty.$$

Isto implica que

$$\mu_k \|(\ln X^k - \ln(I - X^k))\|_F \rightarrow 0, \quad k \rightarrow \infty.$$

Portanto, concluímos que:

$$\mu_k [\ln X^k - \ln(I - X^k)] \rightarrow 0, \quad k \rightarrow \infty. \quad \blacksquare$$

7.4.4 Propriedades da Trajetória Central Primal-Dual

Proposição 7.4.3 Para todo $r > 0$, o conjunto $\{(X(\mu), S(\mu), W(\mu)) : \mu \leq r\}$ é limitado.

Prova. Como $X(\mu) \in \mathcal{S}_{(0,I)}^n$, o conjunto $\{X(\mu)\}$ é limitado, em particular, quando $\mu \leq r$. Agora, provaremos que $\{(S(\mu), W(\mu)) : \mu \leq r\}$ é também limitado. Sabemos que $(X(\mu), y(\mu), S(\mu), W(\mu))$ resolve as equações

$$\mathcal{A}^*y + S - W = \nabla f(X) + 2\mu[\ln X - \ln(I - X)],$$

$$W = \mu(I - X)^{-1},$$

$$S = \mu X^{-1}.$$

Por isso, definindo a função Lagrangeana sobre $\mathcal{S}_{(0,I)}^n \times \mathbb{R}^m \times \mathcal{S}^n \times \mathcal{S}^n$

$$L(X, y, S, W) = f(X) - y^T(\mathcal{A}X - b) - S \bullet X - W \bullet (I - X) + 2\mu \operatorname{tr}[X \ln X + (I - X) \ln(I - X)]$$

temos

$$\nabla_X L(X(\mu), y(\mu), S(\mu), W(\mu)) = 0, \quad (7.33)$$

onde ∇_X denota o gradiente de L relativo à variável X . Agora, observando que $L(\cdot, y(\mu), S(\mu), W(\mu))$ é uma função estritamente convexa sobre $\mathcal{S}_{(0,I)}^n$, (7.33) implica que $X(\mu)$ é o único mínimo da função $L(X, y(\mu), S(\mu), W(\mu))$ sobre $\mathcal{S}_{(0,I)}^n$. Por outro lado, temos que:

$$f(X(\mu)) - L(X(\mu), y(\mu), S(\mu), W(\mu))$$

$$= S(\mu) \bullet X(\mu) - W(\mu)(I - X(\mu)) - 2\mu \operatorname{tr}[X \ln X + (I - X) \ln(I - X)]$$

$$\begin{aligned}
&= n\mu + n\mu - 2\mu \operatorname{tr}[X(\mu) \ln X(\mu) + (I - X(\mu)) \ln(I - X(\mu))] \\
&\leq 2n\mu(1 + \ln 2).
\end{aligned}$$

Logo, pela hipótese 1, podemos tomar um ponto X^0 primal viável, obtendo

$$\begin{aligned}
f(X(\mu)) - 2n\mu(1 + \ln 2) &\leq L(X(\mu), y(\mu), S(\mu), W(\mu)) \\
&\leq L(X^0, y(\mu), S(\mu), W(\mu)) \\
&\leq f(X^0) - S(\mu) \bullet X^0 - W(\mu) \bullet (I - X^0) \tag{7.34}
\end{aligned}$$

onde a última desigualdade é consequência da viabilidade de X^0 , e de que

$$\sum_{i=1}^n [\lambda_i(X^0) \ln \lambda_i(X^0) + (1 - \lambda_i(X^0)) \ln(1 - \lambda_i(X^0))] \leq 0.$$

Usando $\lambda_{\min}(X) \operatorname{tr} S \leq X \bullet S$ e $(1 - \lambda_{\min}(X)) \operatorname{tr} W \leq (I - X) \bullet W$ em (7.34), obtemos:

$$f(X(\mu)) - 2n\mu(1 + \ln 2) \leq f(X^0) - \lambda_{\min}(X^0) \operatorname{tr} S(\mu) - (1 - \lambda_{\min}(X^0)) \operatorname{tr} W(\mu),$$

e assim

$$\lambda_{\min}(X^0) \operatorname{tr} S(\mu) + (1 - \lambda_{\min}(X^0)) \operatorname{tr} W(\mu) \leq f(X^0) + 2n\mu(1 + \ln 2) - f(X(\mu)). \tag{7.35}$$

Desde que $0 < \lambda_i(X^0) < 1$ para todo $i = 1, \dots, n$, temos, em particular, $0 < \lambda_{\min}(X^0) < 1$, e assim, o valor ϵ definido por

$$\epsilon = \min\{\lambda_{\min}(X^0), 1 - \lambda_{\min}(X^0)\}$$

é positivo.

Agora, seja f^* o valor ótimo do problema (7.1). De (7.35), e usando $\mu \leq r$, obtemos que:

$$\begin{aligned}
\operatorname{tr} S(\mu) + \operatorname{tr} W(\mu) &\leq \frac{1}{\epsilon} [f(X^0) + 2n\mu(1 + \ln 2) - f(X(\mu))] \\
&\leq \frac{1}{\epsilon} (f(X^0) + 2n(1 + \ln 2)r - f^*).
\end{aligned}$$

Como $\|S\|_F \leq \operatorname{tr} S$ e $\|W\|_F \leq \operatorname{tr} W$ temos:

$$\|S(\mu)\|_F + \|W(\mu)\|_F \leq \frac{1}{\epsilon} [f(X^0) + 2n\mu(1 + \ln 2) - f(X(\mu))].$$

Portanto o conjunto $\{(X(\mu), S(\mu), W(\mu)) : \mu \leq r\}$ é limitado. ■

Embora a função $\ln(X(\mu)) - \ln(I - X(\mu))$ não seja limitada em $\mathcal{S}_{(0,I)}^n$, é possível provar a existência de pontos de acumulação da trajetória central primal-dual.

Lemma 7.4.4 *O conjunto dos pontos de acumulação da trajetória central primal-dual $\{(X(\mu), y(\mu), S(\mu), W(\mu))\}$ é não vazio.*

Prova. Devido que $\{(X(\mu), S(\mu), W(\mu)) : \mu \leq r\}$ é limitado (ver proposição anterior) existe uma subsequência $\{(X(\mu_j), S(\mu_j), W(\mu_j))\}$ e um ponto $(\bar{X}, \bar{S}, \bar{W})$ tal que:

$$(X(\mu_j), S(\mu_j), W(\mu_j)) \rightarrow (\bar{X}, \bar{S}, \bar{W}), j \rightarrow \infty, \text{ e}$$

$$\mu_j \rightarrow 0, j \rightarrow \infty.$$

Da equação da trajetória central:

$$y(\mu_j) = (\mathcal{A}\mathcal{A}^*)^{-1} \mathcal{A} (\nabla f(X(\mu_j)) + 2\mu_j [\ln(X(\mu_j)) - \ln(I - X(\mu_j))] - S(\mu_j) + W(\mu_j)).$$

Aplicando o limite em j , usando a continuidade de ∇f e o resultado do Corolário 7.4.1 temos

$$y(\mu_j) \rightarrow (\mathcal{A}\mathcal{A}^*)^{-1} \mathcal{A} (\nabla f(\bar{X}) - \bar{S} + \bar{W}), j \rightarrow \infty.$$

Definindo $\bar{y} := (\mathcal{A}\mathcal{A}^*)^{-1} \mathcal{A} (\nabla f(\bar{X}) - \bar{S} + \bar{W})$, obtemos que $(\bar{X}, \bar{y}, \bar{S}, \bar{W})$ é um ponto de acumulação da trajetória primal-dual. Portanto o conjunto de pontos de acumulação é não vazio. ■

Proposição 7.4.4 *Todo ponto de acumulação da trajetória central primal-dual é solução ótima do par de problemas primal (7.1) e dual (7.2).*

Prova. Seja $(\bar{X}, \bar{y}, \bar{S}, \bar{W})$ um ponto de acumulação da trajetória central primal-dual. Então existem seqüências $\{\mu_k\}$ e $(X(\mu_k), y(\mu_k), S(\mu_k), W(\mu_k))$ tais que

$$\mu_k \rightarrow 0, k \rightarrow \infty, \text{ e}$$

$$(X(\mu_k), y(\mu_k), S(\mu_k), W(\mu_k)) \rightarrow (\bar{X}, \bar{y}, \bar{S}, \bar{W}), k \rightarrow \infty.$$

Pela Proposição 7.4.1 \bar{X} é solução do problema primal (7.1). Resta então provar que $(\bar{X}, \bar{y}, \bar{S}, \bar{W})$ é solução do problema dual (7.2). Das condições de KKT do problema penalizado:

$$\mathcal{A}^*y^k + S^k - W^k = \nabla f(X^k) + 2\mu[\ln X^k - \ln(I - X^k)],$$

levando ao limite em k , usando a propriedade de continuidade de ∇f e o Corolário 7.4.1, tem-se:

$$\mathcal{A}^*\bar{y} + \bar{S} - \bar{W} = \nabla f(\bar{X}) \quad (7.36)$$

Também como $S^k \succ 0$ e $W^k \succ 0$ obtemos:

$$\bar{S} \succeq 0 \quad \text{e} \quad \bar{W} \succeq 0 \quad (7.37)$$

De (7.36) e (7.37) concluímos que $(\bar{X}, \bar{y}, \bar{S}, \bar{W})$ é um ponto viável do problema (7.2). Para verificar que $(\bar{X}, \bar{y}, \bar{S}, \bar{W})$ é solução ótima, é suficiente provar a propriedade de complementaridade, isto é, $\bar{W} \bullet (\bar{X} - I) = 0$ e $\bar{S} \bullet \bar{X} = 0$.

Das condições de KKT do problema penalizado, $X(\mu_k) \bullet S(\mu_k) = \mu_k$, ou, no limite,

$$\bar{X} \bullet \bar{S} = 0.$$

Da mesma forma pode-se mostrar que

$$\bar{W} \bullet (I - \bar{X}) = 0.$$

Portanto, a prova está concluída. ■

Mostraremos que no caso de ser a função objetivo linear a trajetória central primal-dual converge. Para obter este resultado usaremos o Lemma 2.5.1. Os argumentos são similares aos usados para a barreira logarítmica, em Halická et al. [38].

Teorema 7.4.2 *Se $f(X) = C \bullet X$, então a trajetória $(X(\mu), y(\mu), S(\mu), W(\mu))$ converge para um ponto $(X^*, y^*, S^*, W^*) \in P^* \times D^*$.*

Proof. Usando o resultado de existência dos pontos de acumulação da seqüência

$(X(\mu), y(\mu), S(\mu), W(\mu))$, existe um ponto $(X^*, y^*, S^*, W^*) \in P^* \times D^*$ e uma seqüência $\{\mu_j\}$ tal que

$$\lim_{\mu_j \rightarrow 0} (X(\mu_j), y(\mu_j), S(\mu_j), W(\mu_j)) = (X^*, y^*, S^*, W^*).$$

Definamos o seguinte conjunto:

$$M :=$$

$$\left\{ \begin{array}{l} \mathcal{A}\bar{X} = 0 \\ \mathcal{A}^*\bar{y} + \bar{S} - \bar{W} = 2\mu[\ln(\bar{X} + X^*) - \ln(I - (\bar{X} + X^*))] \\ (\bar{X} + X^*)(\bar{S} + S^*) = \mu I \\ (I - (\bar{X} + X^*))(\bar{W} + W^*) = \mu I \\ (\bar{X} + X^*) \succ 0, (I - (\bar{X} + X^*)) \succ 0, (\bar{S} + S^*) \succ 0, (\bar{W} + W^*) \succ 0 \text{ and } \mu > 0. \end{array} \right\}.$$

É simples mostrar que se existe $(\bar{X}, \bar{y}, \bar{S}, \bar{W}, \mu) \in M$, então $(\bar{X} + X^*, \bar{y} + y^*, \bar{S} + S^*, \bar{W} + W^*)$ é um ponto da trajetória central primal-dual. Temos também que o elemento zero está contido no fecho de M . De fato, como

$$\lim_{\mu_j \rightarrow 0} (X(\mu_j), y(\mu_j), S(\mu_j), W(\mu_j)) = (X^*, y^*, S^*, W^*), \quad (7.38)$$

podemos definir a seqüência

$$(\bar{X}_j, \bar{y}_j, \bar{S}_j, \bar{W}_j, \bar{\mu}_j) := (X(\mu_j) - X^*, y(\mu_j) - y^*, S(\mu_j) - S^*, W(\mu_j) - W^*, \mu_j).$$

De (7.38) temos que $(\bar{X}_j, \bar{y}_j, \bar{S}_j, \bar{W}_j, \bar{\mu}_j) \in M$ e

$$\lim_{j \rightarrow \infty} (\bar{X}_j, \bar{y}_j, \bar{S}_j, \bar{W}_j, \bar{\mu}_j) = (0_{n \times n}, 0_m, 0_{n \times n}, 0_{n \times n}, 0).$$

Portanto, $0 \in \overline{M}$.

Neste ponto aplicaremos o Lema 2.5.1 da curva Seleccionada. Observe que este Lema implica a existência de um $\epsilon > 0$ e uma função analítica $\gamma: [0, \epsilon) \rightarrow \mathcal{S}^n \times \mathbb{R}^m \times \mathcal{S}^n \times \mathcal{S}^n \times \mathbb{R}$ tal que

$$\gamma(t) = (\bar{X}(t), \bar{y}(t), \bar{S}(t), \bar{W}(t), \mu(t)) \rightarrow (0_{n \times n}, 0_m, 0_{n \times n}, 0_{n \times n}, 0) \text{ quando } t \rightarrow 0, \quad (7.39)$$

e se $t > 0$, $(\bar{X}(t), \bar{y}(t), \bar{S}(t), \bar{W}(t), \mu(t)) \in M$, isto é,

$$\begin{aligned}
\mathcal{A}\bar{X}(t) &= 0, \\
\mathcal{A}^*\bar{y}(t) + \bar{S}(t) - \bar{W}(t) &= 2\mu[\ln(\bar{X}(t) + X^*) - \ln(I - (\bar{X}(t) + X^*))], \\
(\bar{X}(t) + X^*)(\bar{S}(t) + S^*) &= \mu I, \\
(I - (\bar{X}(t) + X^*))(\bar{W}(t) + W^*) &= \mu I, \\
\bar{X}(t) + X^* &\gamma 0, \\
I - (\bar{X}(t) + X^*) &\gamma 0, \\
\bar{S}(t) + S^* &\gamma 0, \\
\bar{W}(t) + W^* &\gamma 0, \\
\mu(t) &> 0.
\end{aligned} \tag{7.40}$$

Observando que a trajetória central é única, então o sistema (7.40) também tem uma única solução, dada por

$$X(\mu(t)) = \bar{X}(t) + X^*, \quad y(\mu(t)) = \bar{y}(t) + y^*, \quad S(\mu(t)) = \bar{S}(t) + S^*, \quad W(\mu(t)) = \bar{W}(t) + W^*,$$

para $t > 0$. Agora, aplicando limite quando $t \rightarrow 0$ e usando (7.39), obtemos, para $\lim_{t \rightarrow 0} \mu(t) = 0$,

$$\lim_{t \rightarrow 0} X(\mu(t)) = X^*, \quad \lim_{t \rightarrow 0} y(\mu(t)) = y^*, \quad \lim_{t \rightarrow 0} S(\mu(t)) = S^*, \quad \lim_{t \rightarrow 0} W(\mu(t)) = W^* \quad .$$

Desde que $\mu(t) > 0$ em $(0, \epsilon)$, $\mu(0) = 0$, e $\mu(t)$ é analítica em $[0, \epsilon)$, existe um intervalo, $(0, \epsilon')$ onde $\mu'(t) > 0$. Devido a isto a função inversa $\mu^{-1}: \mu(t) \rightarrow t$ existe no intervalo $(0, \mu(\epsilon'))$. Além disso, $\mu^{-1}(t) > 0$ para todo $t \in (0, \mu(\epsilon'))$, e $\lim_{t \rightarrow 0} \mu^{-1}(t) = 0$. Portanto, segue-se que

$$\lim_{t \rightarrow 0} X(t) = \lim_{t \rightarrow 0} X(\mu(\mu^{-1}(t))) = \lim_{t \rightarrow 0} \bar{X}(\mu^{-1}(t)) + X^* = X^*.$$

de maneira similar, obtemos

$$\lim_{t \rightarrow 0} y(t) = y^*, \quad \lim_{t \rightarrow 0} S(t) = S^*, \quad \lim_{t \rightarrow 0} W(t) = W^*.$$

Desde que (X^*, y^*, S^*, W^*) foi arbitrário, ele deve ser único, terminando assim a demonstração.

■

7.5 Um Método de Ponto Proximal

Nesta seção introduzimos um método de ponto proximal para resolver o problema (7.1) baseada em uma distância de Bregman induzida pela barreira B , isto é,

$$D_B(Z, Y) = B(Z) - B(Y) - \text{tr}[\nabla B(Y)(Y - Z)]. \quad (7.41)$$

O método gera uma seqüência $\{Z^k\} \subset \mathcal{S}_{(0,I)}^n$ definida como

$$Z^0 \in \mathcal{S}_{(0,I)}^n \text{ tal que } \nabla B(Z^0) \in \text{Im}(A^*), \quad (7.42)$$

$$Z^{k+1} = \arg \min_{Z \in \mathcal{S}^n} \{C \bullet Z + \lambda_k D_B(Z, Z^k) : AZ = b, 0 \preceq Z \preceq I\} \quad (7.43)$$

onde $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfaz

$$\sum_{k=0}^{\infty} \lambda_k^{-1} = \infty. \quad (7.44)$$

e

$$D_B(Z, Z^k) = 2\text{tr}[Z \ln Z - Z \ln Z^k - Z \ln(I - Z) + Z \ln(I - Z^k)] - \ln \det(ZZ^{k-1}) + \ln \det(I - Z)(I - Z^k)^{-1} + \text{tr}(ZZ^{k-1}) + \text{tr}((I - Z)(I - Z^k)^{-1}) - 2n.$$

A seqüência $\{Z^k\}$, $k = 1, 2, \dots$, gerada pelo método, é bem definida e única (para cada k) devido à estrita convexidade da função objetivo sobre \mathcal{P}^0 e toma valores infinitos sobre a fronteira de \mathcal{P} . Assim, para todo $k \geq 1$, $Z^k \in \mathcal{P}^0$.

Como Z^{k+1} é a solução do problema (7.43), existe $v^k \in \mathbb{R}^m$ tal que

$$C + \lambda_k(\nabla B(Z^{k+1}) - \nabla B(Z^k)) = A^*v^k \quad (7.45)$$

Agora, é fácil verificar, usando a hipótese $\nabla B(Z^0) \in \text{Im}(A^*)$, que a trajetória central primal $X(\mu)$, como foi definida na seção anterior, é a única solução do seguinte problema

$$\min_{Z \in \mathcal{S}^n} \{C \bullet Z + \mu D_B(Z, Z^0) : AZ = b, 0 \preceq Z \preceq I\}.$$

Assim, $X(\mu)$ satisfaz

$$C + \mu(\nabla B(X(\mu)) - \nabla B(Z^0)) = A^*w(\mu) \quad (7.46)$$

para algum $w(\mu) \in \mathbb{R}^m$.

O resultado que mostraremos a seguir é uma natural extensão para programação semidefinida do resultado obtido por Iusem et al. [43].

Teorema 7.5.1 A sequência $\{X^k\}$ gerada por (7.42)-(7.43) e a trajetória central convergem para o mesmo ponto.

Demonstração. Seja $\mu_k = \left(\sum_{j=0}^{k-1} \lambda_j^{-1}\right)^{-1}$. Obviamente $\{\mu_k\}$ é uma sequência decrescente para cada $k \geq 1$ e converge para zero quando k vai para infinito. De (7.46) temos

$$C + \mu_k(\nabla B(X(\mu_k)) - \nabla B(Z^0)) = \mathcal{A}^* w(\mu_k)$$

$$C + \mu_{k+1}(\nabla B(X(\mu_{k+1})) - \nabla B(Z^0)) = \mathcal{A}^* w(\mu_{k+1}),$$

para alguma sequência $\{w(\mu_k)\}$.

Usando o seguinte fato $\mu_{k+1}^{-1} - \mu_k^{-1} = \lambda_k^{-1}$, a desigualdade acima implica que

$$C + \lambda_k(\nabla B(X(\mu_{k+1})) - \nabla B(X(\mu_k))) = \mathcal{A}^* v^k,$$

onde

$$v^k = \lambda_k(\mu_{k+1}^{-1} w(\mu_{k+1}) - \mu_k^{-1} w(\mu_k)).$$

Agora, devido à unicidade do ponto de mínimo de (7.43), temos, da equação anterior e de (7.45), que $Z^k = X(\mu_k)$. Finalmente

$$\lim_{k \rightarrow \infty} Z^k = \lim_{k \rightarrow \infty} X(\mu_k) = X^*$$

onde a última igualdade é consequência do Teorema 7.4.2. ■

Capítulo 8

Conclusões

8.1 Conclusões e Futuras Pesquisas do Capítulo 3

No Capítulo 3, estudamos as propriedades geométricas das métricas diagonais no ortante positivo \mathbb{R}_{++}^n e o hipercubo $(0, 1)^n$, vistos como variedades Riemannianas. Obtemos alguns resultados importantes no contexto das aplicações da geometria Riemanniana à otimização, por exemplo, curvatura zero, equações bem mais simples para obter geodésicas e distâncias Riemannianas, condições suficientes para garantir completude da variedade Riemanniana, fórmulas explícitas para o gradiente e a Hessiana de uma função diferenciável. Este estudo permitiu descobrir duas métricas diagonais, $\pi^2 \csc^4(\pi x)$ e $X^{-2}(I - X)^{-2}$, com curvas e distâncias geodésicas explícitas, obtendo assim métodos de máxima descida, subgradientes e ponto proximal, cujos resultados de convergência são conhecidos na literatura de aplicações da geometria Riemanniana à otimização. Outro fato interessante é que estas métricas são as Hessianas das funções barreiras

$$b_1(x) = -\frac{1}{6} \sum_{i=1}^n \{4 \ln(\sin \pi x_i) - \cot^2(\pi x_i)\},$$

$$b_2(x) = \sum_{i=1}^n (2x_i - 1) [\ln x_i - \ln(1 - x_i)]$$

para o hipercubo $[0, 1]^n$. Uma futura pesquisa é o estudo das propriedades geométricas de uma outra classe de métricas, não necessariamente diagonais, no ortante positivo e no hipercubo.

8.2 Conclusões e Futuras Pesquisas do Capítulo 4

No Capítulo 4 estendemos os resultados de convergência global do método de máxima descida com uma busca de Armijo generalizada e uma regularização proximal para resolver problemas de minimização para funções quase-convexas em variedades Riemannianas. Em particular, este método resolve problemas de minimização restrita quando as restrições constituem uma variedade Riemanniana conexa, completa, de dimensão finita com curvatura sectional não negativa e a função objetivo é quase-convexa na variedade. Extensões de métodos subgradientes para minimizar problemas não diferenciáveis e quase-convexas de \mathbb{R}^n para variedades Riemannianas constitui uma futura pesquisa.

8.3 Conclusões e Futuras Pesquisas do Capítulo 5

No Capítulo 5 generalizamos o método de ponto proximal, PBD, para resolver problemas de otimização definidas em variedades de Hadamard. Observamos que em nenhuma de nossas provas usamos o fato da curvatura da variedade ser não positiva, isto é, somente precisamos a unicidade da curva geodésica mínima. Por isso, nossa abordagem pode ser estendida para variedades mais gerais, especificamente para variedades Riemannianas sem pontos focais.

O método de ponto proximal com a distância Logaritmo-Quadrática para funções quase-convexas no ortante não negativo foi desenvolvido em [70] e para problemas restritos em variedades de Hadamard com distâncias Riemannianas e de Bregman está sendo desenvolvido em [71]. Uma futura pesquisa é a generalização deste método para resolver problemas de desigualdades variacionais e de equilíbrio.

8.4 Conclusões e Futuras Pesquisas do Capítulo 6

No Capítulo 6, introduzimos a função $B(x) = \sum_{i=1}^n (2x_i - 1)[\ln x_i - \ln(1 - x_i)]$, e provamos que ela é $(3/2)n$ barreira auto-concordante para $[0, 1]^n$. Esta barreira é de fato nova em métodos de pontos interiores, devido a que ela não pode ser obtida através da regra dada por Nesterov and Nemirovskii [58]. Logo, usamos B para resolver problemas convexos com restrições

lineares limitadas. Provamos que a trajetória central converge para o centro analítico do conjunto das soluções ótimas e no caso linear obtemos a convergência da trajetória primal-dual. Fornecemos também um novo algoritmo primal polinomial para este tipo de problemas com limites superiores para o número de iterações para obter uma ϵ -solução, como também introduzimos um novo método proximal com convergência para o mesmo ponto da trajetória central no caso linear. Como futura pesquisa, propomos o estudo da trajetória dual, além disso, se tal limite existe, obter uma caracterização deste ponto, como foi feito para o caso primal.

Por outro lado, devido a que a constante de auto-concordância obtida para B não é ótima (no sentido de que ela é maior do que n) nossa barreira, pelo menos para algoritmos primais, tem uma pequena desvantagem comparados com algoritmos dados pela barreira logarítmica tradicional. Nossa esperança é que ela tenha melhores propriedades quando trabalhemos com algoritmos primais-duais, devido a que nestes métodos o parâmetro de auto-concordância não é empregado.

8.5 Conclusões e Futuras Pesquisas do Capítulo 7

No Capítulo 7, introduzimos a função

$$B(X) = \text{tr} [(2X - I)(\ln X - \ln(I - X))]$$

para resolver a classe de problemas de otimização semidefinida: $\min \{f(X) \text{ s.a: } A_i \bullet X = b_i, 0 \preceq X \preceq I\}$. Esta família é motivada por alguns problemas, como minimização de somas dos maiores autovalores de matrizes simétricas, ver Alizadeh [2]. Provamos que a trajetória central primal-dual, obtida pelas condições de KKT do problema auxiliar associado à barreira é continuamente diferenciável e que todo ponto de acumulação, o qual existe, é solução do par de problemas primal e dual. No caso linear provamos a convergência de toda a seqüência para uma solução ótima. Também apresentamos um novo método proximal com uma distância de Bregman que converge para a mesma solução ótima (no caso linear) da trajetória central.

Um questão em aberto é a auto-concordância da barreira B , que se põe naturalmente por ser ela uma generalização para $\mathcal{S}_{[0,I]}^n$ de uma barreira auto-concordante em $[0, 1]^n$. Além desta, uma futura pesquisa é a construção de um algoritmo primal-dual (com possível complexidade

polinomial) que melhore as propriedades da barreira logarítmica tradicional.

Por outro lado, do ponto de vista das aplicações da geometria Riemanniana à otimização, uma futura pesquisa seria estudar as propriedades geométricas, como curvatura, geodésicas explícitas e distâncias Riemannianas, da variedade Riemanniana

$$\mathcal{S}_{(0,I)}^n = \{X \in \mathcal{S}^n : 0 \prec X \prec I\}$$

com métrica induzida pela Hessiana da barreira B .

Bibliografia

- [1] ADLER, R., DEDIEU, J.P, MARGULIES, J., MARTENS, M., SHUB, M., “Newton Method on Riemannian Manifold and a Geometric Model for the Human Spine”, *IMA J. of Numerical Analysis*, v. 22, pp.1-32, 2002.
- [2] ALIZADEH, F. “Interior Point Methods in Semidefinite Programming with Applications to Combinatorial Optimization”, *SIAM J. Optimization*, v. 5, n. 1, pp.13-51, 1995.
- [3] ARROW, K.J., DEBREU, G. “Existence of an Equilibrium for a Competitive Economy”, *Econometrica*, v. 22, pp. 265-290, 1954.
- [4] BARRON, N., LIU, W. “Calculus of Variation l^∞ ”, *Applied Math. Optim.*, v. 35, pp. 237-263, 1997.
- [5] BAYER, K., LAGARIAS, J.C. “The Nonlinear Geometry of Linear Programming I: Affine and Projective Scaling Trajectories”, *Transaction of the American Mathematical Society*, v. 314, n. 2, pp. 499-526, 1989.
- [6] BAYER, K., LAGARIAS, J.C. “The Nonlinear Geometry of Linear Programming II: Legendre Transform Coordinates and Central Trajectories”. *Transaction of the American Mathematical Society*, v. 314, n. 2, pp. 527-581, 1989.
- [7] BAZAARA, M.S, SHERALI, H.D., SHETTY, C. M., *Nonlinear Programming: Theory and Algorithms*, 2 ed. New York, John Wiley and Sons, 1993.
- [8] BURACHIK, R.S., *Generalized Proximal Point Algorithm for the Variational Inequality Problem*. Ph.D. dissertation, IMPA-Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brasil, 1995.

- [9] BURACHIK, R.S., GRAÑA DRUMMOND, L.M, IUSEM, A.N., SVAITER, B.F., Full convergence of the steepest descent method with inexact line searches, *Optimization*, v. 32, 137-146, 1995.
- [10] BURACHIK, R.S., SCHEIMBERG, S. “A Proximal Point Method for the Variational Inequality Problem in Banach Spaces”, *SIAM J. Control Optim.*, v. 39, n. 5, pp. 1633-1649, 2000.
- [11] do CARMO, M.P. *Riemannian Geometry*, Birkhausen, Boston, 1992.
- [12] CENSOR, Y., ZENIOS, A. “Proximal minimization algorithms with D-functions”, *Journal of Optimization Theory and Application*, v. 73, n. 3, pp. 451-464, 1992.
- [13] CENSOR, Y., LENT, A. “An Iterative Row-Action Method for Interval Convex Programming”, *Journal of Optimization Theory and applications*, v. 34, n. 3, pp. 321-353, Jul. 1981.
- [14] CHEN, G., TEBOULLE, M. “Convergence Analysis of the Proximal-Like Minimization Algorithm Using Bregman Functions”, *SIAM J. Optimization*, v. 3, pp. 538-543, August 1993.
- [15] da CRUZ NETO, J.X., OLIVEIRA, P.R., *Geodesic Methods in Riemannian Manifolds*, Technical Report ES-352/95, Systems Engineering and Computer Sciences, PESC/COPPE, Federal University of Rio de Janeiro, 1995.
- [16] da CRUZ NETO, J. X., LIMA, L.L., OLIVEIRA, P.R., “Geodesic Algorithms in Riemannian Geometry”, *Balkan Journal of Geometry and its Applications*, v. 3, n. 2, pp. 89-100, 1998.
- [17] da CRUZ NETO, J.X., FERREIRA, O.P., LUCAMBIO PEREZ, L., NÉMETH, S.Z., “Convex-and Monotone-Transformable Mathematical Programming and a proximal-Like Point Method”, aceito para publicação em *Journal Of Global Optimization*.
- [18] da CRUZ NETO, J.X., FERREIRA, O.P., LUCAMBIO PEREZ, A proximal regularization of the steepest descent method in Riemannian manifolds, *Balkan Journal of Geometry and its Applications*, v. 4, n. 2, pp. 1-18, 1999.

- [19] CUNHA, G.F.M., PINTO, A.M., OLIVEIRA, P.R., da CRUZ NETO, J.X., *Generalization of the Primal and Dual Affine Scaling Algorithms*, Technical Report ES 675/05, PESC/COPPE, Federal University of Rio de Janeiro, 2005.
- [20] DONATH, W.E., HOFFMAN, A.J., “Lower Bounds for the Partitioning of Graphs”, *IBM J. Res. and Devel.*, v. 17, pp. 420-425, 1973.
- [21] DEDIEU, J.P., PRIOURET, P., MALAJOVICH, G., Newton’s Method on Riemannian Manifolds: Covariant Alpha-Theory, *IMA J. Num. Anal.*, v. 23, n.3, pp. 395-419, 2003.
- [22] ERIKSSON, J.R., *An Iterative Primal-Dual Algorithm for Linear Programming*, Report LiTH-MAT-R-1985-10, Department of Mathematics, Linköping University, Linköping, Sweden, 1985.
- [23] FERREIRA, O.P., OLIVEIRA, P.R., “Sub gradient Algorithm on Riemannian Manifolds”, *Journal of Optimization Theory and Application*, v. 97, n.1, pp. 93-104, April, 1998.
- [24] FERREIRA, O.P., SVAITER, B.F., “Kantorovich’s Theorem on Newton’s Method in Riemannian Manifold”, *Journal of Complexity*, v. 18, n.1, pp. 304-329, 2002.
- [25] FERREIRA, O.P., OLIVEIRA, P.R., “Proximal Point Algorithm on Riemannian Manifolds”, *Optimization*, v. 51, n. 2, pp. 257-270, 2002.
- [26] FIACCO, A.V., MCCORMICK, G. P., *Non Linear Programming: Sequential Unconstrained Minimization Techniques*, 1 ed. New York, Wiley, 1968, Reprinted by SIAM Publication, 1990.
- [27] FRISCH, K.R., *The Logarithm Potential Method of Convex Programming*, Technical Report, University Institute of Economics, Oslo, Norway, 1955.
- [28] GABAY, D., “Minimizing a Differentiable Function over a Differentiable Manifold”, *Journal of Optimization Theory and Application*, v. 37, pp. 177-219, 1982.
- [29] GARCIGA, O. IUSEM, A.N, “Proximal Methods with Penalization Effects in Banach Spaces”, *Numer. Funct. Anal. Optim.* v. 25, n. 1-2, pp. 69-91, 2004.

- [30] GINSBERG, W., “Concavity and Quasiconcavity in Economics”, *Journal of Economic Theory*, v. 6, pp. 596-605, 1973.
- [31] GILL, P. E., MURRAY, W., SAUNDERS, M.A., TOMLIN, J. A., WRIGHT M. H., “On Projected Newton Barrier Methods for Linear Programming and an Equivalence to Karmarkar’s Projected Method”, *Mathematical Programming*, v. 36, pp. 183-209, 1986.
- [32] GOLUB, G., VAN LOAN, C. *Matrix Computations*, Third Edition, The Johns Hopkins University Press, 1996.
- [33] GOUDOU, X., MUNIER, J., “The Heavy Ball with Friction Method: the quasiconvex case”, 2005, Submitted.
- [34] GRAÑA DRUMMOND, L.M., SVAITER, B.F., “On Well Definedness of the Central Path”, *Journal of Optimization Theory and Applications*, v. 102, n. 2, pp. 223-237, 1999.
- [35] GRAÑA DRUMMOND, L.M., SVAITER, B.F., A steepest descent method for vector optimization, *J. Comp. Appl. Math.*, v. 175, pp. 5-29, 2005.
- [36] GROMICHO, J., *Quasiconvex optimization and location theory*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998.
- [37] GÜLER, O., ROOS, C., TERLAKY, T., VIAL, J.Ph., *Interior Point Approach to the Theory of Linear Programming*, Technical Report 1992.3, Université de Geneve, Département d’Economie Commerciale et Industrielle, Geneve, Switzerland, 1992.
- [38] HALICKÁ, M., de KLERK, E., ROSS, C., “On the Convergence of the Central Path in Semidefinite Optimization”, *SIAM J. Optimization*, v. 12, n. 4, pp. 1090-1099, 2002.
- [39] den HERTOOG, D., *Interior Point Approach to Linear, Quadratic and Convex Programming*, 1 ed. Kluwer Academic Publishers, 1992.
- [40] den HERTOOG, D., ROOS, C., TERLAKY, T., *Inverse Barrier Methods for Linear Programming*, Report 91-27, Faculty of Technical Mathematics and Informatics, Delf University of Technology, The Netherlands, 1991.

- [41] HORN, R.A., JHONSON, C.R., *Topics in Matrix Analysis*, Cambridge University Press, 1991.
- [42] IUSEM, A.N., SVAITER, B.F., A proximal regularization of the steepest descent method, *RAIRO Recherche Opérationnelle/Operation Reseach*, v. 29, n. 2, pp. 123-130, 1995.
- [43] IUSEM, A.N., SVAITER, B.S., da CRUZ NETO, J.X., “Central Paths, Generalized Proximal Point Methods, and Cauchy Trajectories in Riemannian Manifolds”, *SIAM J. Control Optim.*, v. 37, n. 2, pp. 566-588, 1999.
- [44] IUSEM A.N., *Métodos de Ponto Proximal em Otimização*, Colóquio Brasileiro de Matemática, IMPA, Brazil, 1995.
- [45] KARMARKAR, N., “A New Polynomial-Time Algorithm for Linear Programming”, *Combinatorica*, v. 4, pp. 373-395, 1984.
- [46] KARMARKAR, N., “Riemannian Geometry Underlying Interior-Point Methods for Linear Programming”, *Contemporary Mathematics*, v. 114, pp. 51-75, 1990.
- [47] KIWIEL, K.C., “Proximal Minimization Methods with Generalized Bregman Functions”, *SIAM Journal of Control and Optimization*, v. 35, pp. 1142-1168, 1997.
- [48] KIWIEL, K., MURTY, K., Convergence of the steepest descent method for minimization quasiconvex functions, *JOTA Journal of Optimization Theory and Applications*, v. 89, n. 1, pp. 221-226, 1996.
- [49] LIMA, E.L., *Curso de Análise*, v. 2, IMPA, Rio de Janeiro, 1971.
- [50] LOJASIEWICZ, S., *Ensembles Semi-Analitiques*, I.H.E.S., Bures-sur-Yvette, 1965.
- [51] LUENBERGER, D.G., “The Gradient Projection Method Along Geodesics”, *Management Science*, v. 18, n. 11, pp. 620-631, 1972.
- [52] MEGIDDO, N., “Pathways to the Optimal Set in Linear Programming”. In: *Progress in Mathematical Programming, Interior-Point and Related Methods (N. Megiddo, Ed.)*, Springer-Verlag, New York, pp. 131-158, 1989.

- [53] MEGIDDO, N., SHUB, M., Boundary Behavior of Interior Point Algorithms in Linear Programming. *Math. Oper. Res.*, 14, pp. 97-146, 1989.
- [54] MONTEIRO, R.D.C., ADLER, I., “Interior Path Following Primal-Dual Algorithms, Part I: Linear Programming”, *Mathematical Programming*, v. 44, pp. 43-66, 1989.
- [55] NAZARETH, J.L., *The Newton-Cauchy Framework: A Unified Approach to Unconstrained Nonlinear Minimization*, Springer-Verlag, Berlin, Heidelberg, 1994.
- [56] NÉMETH S., *Five kinds of monotone vector fields*, *Pure Mathematics and Applications*, v. 9, n. 3-4, pp. 417-428, 1999.
- [57] NESTEROV. Y., *Introductory Lectures on Convex Programming, Volume I: Basic Course*, Published by Kluwer Academic Publishers, 2001.
- [58] NESTEROV., Y.E., NEMIROVSKII, A.N., *Interior-Point Polynomial Algorithms in Convex Programming*, Philadelphia, Society for Industrial and Applied Mathematics, 1994.
- [59] NESTEROV, Y.E., TODD, M.J., “On the Riemannian Geometry Defined by Self-Concordant Barrier and Interior-Point Methods”, *Foundations of Computational Mathematics*, v. 2, pp. 333-361, 2002.
- [60] OLIVEIRA, P.R., da CRUZ NETO, J.X., *A Unified View of Primal Methods Through Riemannian Metrics*, Technical Report ES-363/95, Systems Engineering and Computer Science PESC/COPPE, Federal University of Rio de Janeiro, 1995.
- [61] OLIVEIRA, P.R., da CRUZ NETO, J.X., *Elementos de Geometria Riemanniana*, Technical Report, Systems Engineering and Computer Science, PESC/COPPE, Federal University of Rio de Janeiro Technical, 1995.
- [62] OLIVEIRA, G.L., OLIVEIRA, P.R., *A New Class of Interior Point Methods for Optimization Under Positivity Constraints*. Technical Report, Systems Engineering and Computer Science, PESC/COPPE, Federal University of Rio de Janeiro, 2002.

- [63] OVERTON, M.L., WOMERSLEY, R.S., “Optimality Conditions and Duality Theory for Minimizing Sums of the Largest Eigenvalues of Symmetric Matrices”, *Math. Programming*, v. 62, pp. 321-357, 1993.
- [64] PAPA QUIROZ, E. A., *Elementos de Geometria Riemanniana para Métodos de Pontos Interiores Primais-Duais na Otimização Linear*, Monografia de Passagem ao Doutorado Direto, PESC/COPPE, Federal University of Rio de Janeiro, Decembre 2003.
- [65] PAPA QUIROZ, E. A., OLIVEIRA, P. R. “New Results on Linear Optimization Through Diagonal Metrics and Riemannian Geometry Tools”, *SOBRAPO, RJ-BRAZIL*, pp. 1938-1949, 2004.
- [66] PAPA QUIROZ, E. A., OLIVEIRA, P. R., “A New Self-concordant Barrier for the Hypercube”, to appear in *Journal of Optimization Theory and Applications JOTA*, August 2007.
- [67] PAPA QUIROZ, E. A., OLIVEIRA, P. R., “A New Barrier for a Class of Semidefinite Problems”, *RAIRO Operations Research*, v. 40, pp. 303-323, 2006.
- [68] PAPA QUIROZ, E. A., OLIVEIRA, P. R., “Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard Manifolds”, to appear in *Journal of Convex Analysis*, 2008.
- [69] PAPA QUIROZ, E. A., QUISPE, E. M., OLIVEIRA, P. R., “Steepest Descent Method with a Generalized Armijo Search for quasiconvex Functions on Riemannian Manifolds”, to appear in *Journal of Mathematical Analysis and Applications*, 2007.
- [70] PAPA QUIROZ, E. A., OLIVEIRA, P. R., *Classic and Logarithmic-Quadratic Proximal Point Method for Quasiconvex Minimization*, Optimization Online Preprint, 2006.
- [71] PAPA QUIROZ, E. A., OLIVEIRA, P. R., *Proximal Point Methods for Functions Involving Lojasiewicz, Quasiconvex and Convex Features on Hadamard Manifolds*, Working Paper, 2007.

- [72] PEREIRA, G.J., OLIVEIRA, P.R., “A New Class of Proximal Algorithms for the Non-linear Complementary Problems”, In: *Optimization and Control with Applications (Qi, Liqun; Teo Koklay; Yang, Xiaoqi, Eds)*, Springer, Florida, pp. 549-561, 2005.
- [73] POLYAK, R.A., *Modified Barrier Function*, Mathematical Sciences Department, IBM Research Division, T.J. Watson Research Center, Yorktown Heights, New York, 1989.
- [74] RENEGAR, J., “A Polynomial-Time Algorithm, Based on Newton’s Method, for Linear Programming”, *Mathematical Programming*, v. 40, pp. 59-93, 1988.
- [75] RAPCSÁK, T., *Smooth Nonlinear Optimization in \mathbb{R}^n* , Kluwer Academic Publishers, 1997.
- [76] RAPCSÁK, T., THANG, T.T, “A Class of Polynomial Variable Metric Algorithms for Linear Optimization”, *Mathematical Programming*, v. 74, pp. 319-331, 1996.
- [77] SAIGAL, R., *The Primal Power Affine Scaling Method*, Tech. Report, No 93-21, Dep. Ind. and Oper. Eng., University of Michigan, 1993.
- [78] SAKAI, T., *Riemannian Geometry*, American Mathematical Society, Providence, RI, 1996.
- [79] SMITH, S.T., “Optimization Techniques on Riemannian Manifolds”, v. 3 of Fields Institute Communications, American Mathematical Society, Providence, R.I., pp. 113-146, 1993.
- [80] SOUZA, S., OLIVEIRA, P. R., *A New Class of Interior Proximal Methods with Variable Metric for Optimization over the Positive Orthant*, Optimization Online Preprint 2006.
- [81] TAKAYAMA, A., *Mathematical Economics*, 2nd Edition, Cambridge Univ. Press, 1995.
- [82] UDRISTE, C. *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers, 1997. Philadelphia, SIAM, 1997.
- [83] ZHOR, N.Z., *Minimization Methods for Nondifferentiable Functions*, Springer Verlag, Berlin, Germany, 1985.