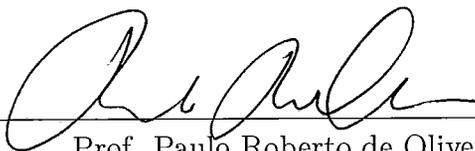


UM ALGORITMO DE PONTO PROXIMAL EM PROGRAMAÇÃO
SEMIDEFINIDA

Ronaldo Gregório

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
MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E
COMPUTAÇÃO.

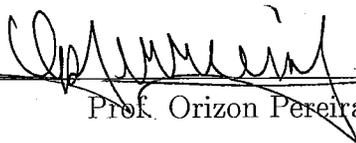
Aprovada por:



Prof. Paulo Roberto de Oliveira, Dr. Ing.



Prof. Nair Maria Maia de Abreu, D.Sc.



Prof. Orizon Pereira Oliveira, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

AGOSTO DE 2004

GREGÓRIO, RONALDO

Um Algoritmo de Ponto Proximal em Programação Semidefinida [Rio de Janeiro] 2004

VIII, 62 p. 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia de Sistemas e Computação, 2004)

Tese – Universidade Federal do Rio de Janeiro, COPPE

1 - Algoritmo de Ponto Proximal

2 - Variedades Riemannianas

3 - Programação Semidefinida

I. COPPE/UFRJ II. Título (série)

Agradecimentos

A Deus.

A meus familiares, por manterem nossos laços afetivos tão vivos.

A meu orientador Paulo Roberto, pela atenção e apoio dispensados durante todo este período.

À professora Nair Maria Maia de Abreu e ao professor Orizon Pereira Ferreira, por fazerem parte da banca examinadora de minha dissertação.

Aos professores da Linha de Otimização que direta ou indiretamente colaboraram para minha formação.

Aos meus professores da graduação. De maneira especial o professor Valdomiro Neves Lima (orientador da graduação) que me fez acreditar que era possível atingir esse objetivo, apesar de todas as dificuldades.

A todos os amigos do Labotim que tornaram menos árduo e mais descontraído o caminho trilhado.

A todos os colegas de estudos e funcionários do Programa.

À CAPES pela bolsa de estudos concedida durante o período de mestrado.

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

UM ALGORITMO DE PONTO PROXIMAL EM PROGRAMAÇÃO SEMIDEFINIDA

Ronaldo Gregório

Agosto/2004

Orientador: Paulo Roberto Oliveira

Programa: Engenharia de Sistemas e Computação

Neste Trabalho, estamos apresentando uma proposta algorítmica para Programação Semidefinida (*PSD*). Esta aparenta ser a primeira na classe de métodos de ponto proximal, entre os algoritmos que transformam Programação Semidefinida em Programação Não-linear no ortante positivo do espaço euclidiano de dimensão n . Isso representa uma redução significativa no número de variáveis do problema original (um (*PSD*), em geral, possui dimensão $\frac{n(n+1)}{2}$). A proposta é implementável e a iteração principal do algoritmo de ponto proximal conceitual é substituída por uma sucessão de problemas Não-lineares em R^n . Estamos motivados por resultados de convergência do algoritmo proximal clássico estendido a Variedades Riemannianas com curvatura seccional não-positiva. Um importante exemplo de tal variedade é o espaço de matrizes simétricas definidas positivas, onde a métrica é dada pela hessiana da barreira padrão $-\ln \det(X)$. Portanto, observando que o método proximal clássico não depende das geodésicas deste espaço, aplicamos estas idéias para desenvolver um algoritmo de ponto proximal para (*PSD*).

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

A PROXIMAL POINT ALGORITHM IN SEMIDEFINITE
PROGRAMMING

Ronaldo Gregório

August/2004

Advisor: Paulo Roberto Oliveira

Department: Computing and Systems Engineering

In this work, we propose a new (proximal) algorithm for Semidefinite Programming (*SDP*). It appears to be the first in the proximal class on the set of methods that convert Semidefinite Programming in Nonlinear Programming. It is implementable, and each main iteration of the conceptual proximal point algorithm is replaced by a sequence of non-linear programming problems in the positive octant of the real vector space R^n . We are motivated by results of the classical proximal algorithm extended to Riemannian manifolds with non positive sectional curvature. An important example of such manifold is the space of symmetric definite matrices, where the metrics is given by the Hessian of the standard barrier function $-\ln \det(X)$. Then, observing the obvious fact that proximal algorithms do not depend on the geodesics, we apply those ideas to develop a proximal point algorithm for (*SDP*).

Conteúdo

Introdução	1
1 Elementos de Geometria Riemaniana e Barreiras Autoconcordantes	4
1.1 Introdução	4
1.2 Barreiras autoconcordantes	5
1.3 Conceitos preliminares de geometria riemaniana	7
1.4 Geometria riemaniana gerada por hessiana de barreiras autoconcordantes logaritmicamente homogêneas	13
1.5 Propriedades da métrica riemaniana	18
2 O Cone das Matrizes Simétricas Semidefinidas Positivas	21
2.1 Introdução	21
2.2 Alguns resultados de Álgebra Linear	23
2.3 Uma barreira autoconcordante em S_{++}^n	29
2.4 Distância em S_{++}^n	33
3 Algoritmo de Ponto Proximal em Programação Semidefinida	36
3.1 Introdução	36
3.2 Algoritmo de ponto proximal em S_{++}^n	36

3.3	Um algoritmo em R_{++}^n	39
3.4	Um problema bem posto	45
3.5	Aplicando o método de Newton	49
3.6	Aplicação à (<i>PSD</i>)	50
4	Conclusões, Trabalhos Futuros e Extensões	54
	Apêndice	57
	Bibliografia	59

Lista de Figuras

3.1	Gráfico da função $\ln^2 t$, para $t \in (0, 20]$	47
-----	--	----

Introdução

O modelo de Programação Semidefinida tem-se mostrado uma poderosa ferramenta em Otimização com aplicações em sistemas de controle e teoria dos grafos [1]. No entanto, tal modelo somente começou a ganhar destaque após o desenvolvimento de algoritmos para resolvê-lo.

Essencialmente, dois pesquisadores, Nesterov e Nemirovsk [2], [3] e [4], através do conceito de autoconcordância juntamente com a teoria dos métodos de pontos-interiores para programação convexa, deram a partida para o grande número de trabalhos voltados para aquela perspectiva. Estes autores e Alizadeh [5] mostraram que uma extensa classe de algoritmos de pontos-interiores para Programação Linear pode ser estendida à programação semidefinida. Evidentemente, estamos falando dos algoritmos primais-duais que representam o maior percentual de pesquisas nessa linha (ver, por exemplo, [6]).

Os algoritmos primais-duais mostram-se particularmente eficientes quando aplicados a problemas de pequeno e médio porte. Para grande porte, os métodos baseados em *low-rank factorization* têm mostrado excelente desempenho computacional quando aplicados a determinadas classes de problemas (ver [7], [8] e [9] para formulação primal e [10] para dual). Estes métodos estão baseados na substituição da variável (matriz) por uma decomposição do

tipo VDV , para escolhas particulares de V e D , transformando o problema de programação semidefinida em um problema não-linear em R^l para algum inteiro l . Vale lembrar que a convergência global é assegurada apenas para algoritmos baseados na formulação dual (ver [10]). Uma outra perspectiva que trabalha com informações de segunda ordem dos problemas de grande porte em programação semidefinida é dada através da noção de complemento matricial. Ver [11] para maiores detalhes.

Finalmente, Helmberg e Rendl [12] propõem um método de feixes espectral aplicado a determinadas transformações do problema primal, que geram um problema convexo não-diferenciável.

Neste trabalho estamos propondo um novo algoritmo para programação semidefinida. Tal proposta parece ser a primeira em ponto proximal na classe de algoritmos que transformam Programação Semidefinida em Programação Não-linear. Este algoritmo é implementável e reduz cada iteração à resolução de uma sucessão de problemas de programação não-linear no ortante estritamente positivo de R^n . A motivação inicial vem de resultados preliminares de convergência do algoritmo clássico de ponto proximal estendido a variedades riemanianas de curvatura seccional não-positiva [13]. Um importante exemplo de tal variedade é o espaço das matrizes simétricas definidas positivas, onde a métrica é dada pela hessiana da barreira $-\ln \det(X)$ (ver [14] e [15]). Observando que o algoritmo clássico não depende das geodésicas deste espaço, aplicamos estes resultados para obter um algoritmo proximal para programação semidefinida.

Nos capítulos 2 e 3 que seguem, estaremos apresentando a teoria que dá sustentação a nossa proposta. No quarto, destacaremos a proposta e encer-

raremos este trabalho com algumas considerações a respeito do algoritmo proposto, bem como algumas sugestões de trabalhos futuros e possíveis extensões.

Capítulo 1

Elementos de Geometria Riemaniana e Barreiras Autoconcordantes

1.1 Introdução

O estudo de funções barreiras representa um importante papel no desenvolvimento de algoritmos primais, duais e primais-duais de pontos interiores. O conceito de autoconcordância apresentado por Nemirovsk e Nesterov [4] juntamente com o método de Newton contribuíram de forma ostensiva para obtenção de resultados relevantes sobre a complexidade desses algoritmos. Nesse trabalho, eles mostram que o método de Newton (com busca linear aproximada) apresenta ótimos resultados na minimização de funções autoconcordantes e que em alguns casos determina a convergência polinomial.

A escolha adequada de uma barreira pode representar, além do sucesso dos algoritmos de pontos interiores, a determinação de uma estrutura métrica diferenciável denominada Variedade Riemaniana, veja [14]. Um levantamento mais detalhado sobre barreiras autoconcordantes para cones convexos em pontos interiores pode ser encontrado em [16].

Neste capítulo destacamos alguns resultados sobre estruturas riemania-
nas geradas por barreiras autoconcordantes logaritmicamente homogêneas,
caracterizando as geodésicas e apontando algumas propriedades da métrica
obtida. Os resultados apresentados podem ser encontrados em [14] e [4].
Em alguns casos, apenas acrescentamos a demonstração em outros, deta-
lhamos peculiaridades importantes para o trabalho. Apresentamos também
uma pequena revisão de elementos clássicos de geometria riemaniana. De-
monstrações e detalhes mais específicos relacionados ao assunto podem ser
encontrados em do Carmo[18].

1.2 Barreiras autoconcordantes

Definição 1.1 *Seja $Q \subseteq E$ um subconjunto convexo e fechado de um espaço
de vetores reais E , de dimensão n com $\text{int } Q \neq \emptyset$. Uma função $F : Q \rightarrow \mathbb{R}$
é a -autoconcordante ($a > 0$) em Q , se:*

(i) $F \in C^3$;

(ii) $|D^3F(x)[h, h, h]| \leq 2a^{-\frac{1}{2}} (D^2F(x)[h, h])^{\frac{3}{2}}, \forall x \in Q \text{ e } h \in E$;

e ainda, se F é 1-autoconcordante e satisfaz

(iii) $|DF(x)[h]| \leq (\nu D^2F(x)[h, h])^{\frac{1}{2}}, \forall x \in \text{int } Q \text{ e } h \in E$;

(iv) $F(x) \rightarrow \infty$, quando $x \rightarrow \partial Q$,

onde $D^kF(x)[h_1, \dots, h_n]$ representa a k -ésima diferencial de F na direção
 h_1, \dots, h_n e ∂Q , a fronteira de Q , então F é uma barreira
 ν -autoconcordante em Q .

As notações $DF(x)[h]$, $D^2F(x)[h_1, h_2]$ e $D^3F(x)[h_1, h_2, h_3]$ representam, respectivamente, $\langle F'(x), h \rangle$, $\langle F''(x)h_1, h_2 \rangle$ e $\langle (F'''(x)h_1)h_2, h_3 \rangle$. ν é o parâmetro da função barreira. F'' e F''' são formas bilinear e trilinear simétricas, respectivamente, em E .

A segunda desigualdade, item (iii) da Definição (1.1), mostra que uma barreira autoconcordante F tem sua diferencial de 1ª ordem limitada em qualquer direção pela raiz quadrada do parâmetro ν multiplicado pela norma induzida por F'' dessa direção, ou ainda, a derivada de F é limitada.

Nos casos em que Q é um cone ($x \in Q \Rightarrow tx \in Q, \forall t \geq 0$), F é chamada ν -logaritmicamente homogênea se satisfaz

$$F(\tau x) = F(x) - \nu \ln \tau, \quad (1.1)$$

para todo $x \in \text{int } Q$ e $\tau > 0$.

As propriedades que seguem decorrem naturalmente de (1.1).

Proposição 1.1 *Uma barreira $F : \text{int } Q \rightarrow \mathbb{R}$, ν -logaritmicamente homogênea, definida em Q satisfaz as seguintes propriedades:*

$$(i) \quad F'(\tau x) = \tau^{-1} F'(x), \quad \langle F'(x), x \rangle = -\nu;$$

$$(ii) \quad F''(x)x = -F'(x), \quad \langle F''(x)x, x \rangle = \nu;$$

$$(iii) \quad F'''(x)[x] = -2F''(x), \quad \text{para todo } x \in \text{int } Q \text{ e } \tau > 0.$$

Demonstração: Aplicando a regra da cadeia para diferenciar (1.1) em relação a x , vem que $(F(\tau x))' = \tau F'(\tau x) = F'(x)$, e $\langle F'(x), x \rangle = \frac{d}{dt} \Big|_{t=0} F(x + tx) = \frac{d}{dt} \Big|_{t=0} (F(x) - \nu \ln(1 + t)) = -\nu$. Diferenciando agora a segunda parte

de (i), com respeito a x , obtém-se $\langle F''(x), x \rangle' = F'''(x)x + F''(x) = 0$, donde segue a primeira parte de (ii). Aplicando a segunda parte de (i) e (ii), conclui-se a segunda parte. Por último, diferenciando-se a segunda parte de (ii) vem que $\langle F'''(x)x, x \rangle' = (F''''(x)x + F'''(x))x + F'''(x)x = 0$. ■

1.3 Conceitos preliminares de geometria riemanniana

Definição 1.2 *Uma variedade diferenciável de dimensão n é um conjunto M e uma família de aplicações biunívocas $y_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow M$ de abertos U_α de \mathbb{R}^n em M , tais que:*

- (i) $\bigcup_\alpha y_\alpha(U_\alpha) = M$.
- (ii) *Para todo par α, β , com $y_\alpha(U_\alpha) \cap y_\beta(U_\beta) = W \neq \emptyset$, os conjuntos $y_\alpha^{-1}(W)$ e $y_\beta^{-1}(W)$ são abertos em \mathbb{R}^n e as aplicações $y_\beta^{-1} \circ y_\alpha$ são diferenciáveis.*
- (iii) *A família $\{(U_\alpha, y_\alpha)\}$ é máxima relativamente às condições (i) e (ii).*

Uma estrutura diferenciável em um conjunto M induz uma topologia em M . Basta definir que $A \subset M$ é um aberto de M se $y_\alpha^{-1}(A \cap y_\alpha(U_\alpha))$ é um aberto de \mathbb{R}^n , para todo α .

Uma curva diferenciável em M é uma aplicação diferenciável $\varphi : (-\epsilon, \epsilon) \rightarrow M$. Sejam $x \in M$ tal que $\varphi(0) = x$ e D o conjunto das funções de M diferenciáveis em x . O vetor tangente à φ em $t = 0$ é a função $\varphi'(0) : D \rightarrow \mathbb{R}$, dada por

$$\varphi'(0)f = \frac{d(f \circ \varphi)}{dt} \Big|_{t=0}, \quad f \in D.$$

Um vetor tangente a M em x é o vetor tangente em $(t = 0)$ de alguma curva diferenciável φ , com $\varphi(0) = x$. O conjunto de todos os vetores tangentes a M em x é chamado de espaço tangente a M em x e é representado por $T_x M$. $TM = \bigcup_{x \in M} T_x M$ é o *fibrado tangente* de M .

Definição 1.3 *Uma métrica riemaniana (estrutura riemaniana) em uma variedade diferenciável M é uma correspondência que associa a cada ponto $x \in M$ um produto interno $\langle \cdot, \cdot \rangle_x$ diferenciável em $T_x M$. $\|\bullet\|_x$ é a norma correspondente à $\langle \cdot, \cdot \rangle_x$, dada por $\|\nu\|_x = \langle \nu, \nu \rangle_x^{\frac{1}{2}}$, $\nu \in T_x M$.*

Uma variedade diferenciável M munida de uma métrica riemaniana é chamada *variedade riemaniana*.

Conhecida a métrica riemaniana, podemos calcular o comprimento de uma curva diferenciável φ ligando dois pontos x_1, x_2 . Com efeito, considere $\varphi : [a, b] \rightarrow M$ satisfazendo $\varphi(a) = x_1$ e $\varphi(b) = x_2$. O comprimento riemaniano de φ é definido por

$$L(\varphi) = \int_a^b \|\varphi'(t)\|_{\varphi(t)} dt, \quad (1.2)$$

e a distância riemaniana entre x_1 e x_2 define-se como

$$d(x_1, x_2) = \inf_{\varphi \in C_{x_1, x_2}} L(\varphi), \quad (1.3)$$

onde C_{x_1, x_2} é o conjunto de todas as curvas diferenciáveis ligando x_1 à x_2 .

Definição 1.4 *Uma curva diferenciável φ é dita normal se $\|\varphi'(t)\|_{\varphi(t)}$ é constante, para todo $t \in (a, b)$.*

Decorre imediatamente de (1.2) que uma curva normal tem seu comprimento dado por $\|\varphi'(t)\|_{\varphi(t)}$.

O conceito de curva normal é de fundamental importância em geometria riemanniana, pois simplifica o cálculo da distância entre dois pontos dados. O próximo resultado evidencia ainda mais a relevância deste conceito.

Lema 1.1 *Toda curva diferenciável pode ser reparametrizada na forma normal.*

Demonstração: Considere uma curva diferenciável $\varphi(t)$, $t \in (a, b)$, satisfazendo $\varphi(a) = x$ e $\varphi(b) = y$. Defina uma nova curva $\alpha(s(t)) := \varphi(t)$, com s dada por

$$s(t) := \frac{1}{L(\varphi)} \int_a^t \|\varphi'(r)\|_{\varphi(r)} dr. \quad 1$$

Portanto, $[\alpha(s(t))]' = \alpha'(s) \cdot s'(t) = \alpha'(s) \cdot \frac{\|\varphi'(t)\|_{\varphi(t)}}{L(\varphi)} = \varphi'(t)$. Logo, $\|\alpha'(s)\|_{\alpha(s)} = L(\varphi)$, para todo $s \in (0, 1)$. Tem-se ainda que $s(a) = 0$ e $s(b) = 1$, donde se conclui que $\alpha(0) = x$ e $\alpha(1) = y$. ■

Um *campo de vetores* X em uma variedade diferenciável M é uma correspondência que a cada ponto $x \in M$ associa um vetor $X(x) \in T_x M$. $\chi(M)$ representa o conjunto dos campos de vetores de classe C^∞ em M e $D(M)$ o conjunto das funções reais de classe C^∞ definidas em M . Uma conexão afim ∇ em M é uma aplicação $\nabla : \chi(M) \times \chi(M) \rightarrow \chi(M)$, indicada por $(X, Y) \rightarrow \nabla_X Y$, e que satisfaz as seguintes propriedades:

$$(i) \quad \nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z.$$

$$(ii) \quad \nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z.$$

¹Como $\|\varphi'(t)\|_{\varphi(t)}$ é contínua em $[a, b]$, s pode ser vista como a primitiva de $\|\varphi'(t)\|_{\varphi(t)}$ a menos da constante $\frac{1}{L(\varphi)}$ (toda função contínua definida em intervalo compacto possui primitiva. Ver [17]).

(iii) $\nabla_X(fY) = f\nabla_X Y + X(f)Y$, onde $X, Y, Z \in \chi(M)$ e $f, g \in D(M)$.

Em geral, a derivada de um campo de vetores X , ao longo de uma curva diferenciável φ em M , representada por $\frac{dX}{dt}(t)$, não pertence a $T_{\varphi(t)}M$ (por exemplo, considere o caso em que M é uma superfície regular em R^3). Torna-se necessário o conceito de *derivada covariante* de X ($\frac{DX}{dt}(t)$).

Proposição 1.2 *Seja M uma variedade diferenciável com uma conexão afim ∇ . Existe uma única correspondência que associa a um campo vetorial X ao longo de uma curva diferenciável φ um outro campo vetorial ao longo de φ , denominado derivada covariante de X , tal que:*

$$(i) \quad \frac{D}{dt}(X + Y) = \frac{DX}{dt} + \frac{DY}{dt}.$$

(ii) $\frac{D}{dt}(fX) = \frac{df}{dt}X + f\frac{DX}{dt}$, onde Y é um campo de vetores ao longo de φ e f restrita a φ é diferenciável.

(iii) Se X é induzido por um campo de vetores $Z \in \chi(M)$, isto é, $X(t) = Z(\varphi(t))$, então $\frac{DX}{dt} = \nabla_{\frac{d\varphi}{dt}}Z$.

Uma das exigências para extensão dos conceitos estudados no Cálculo Diferencial às variedades riemannianas é que o produto interno definido no espaço tangente $T_x M$ a uma variedade diferenciável M em um ponto x seja simétrico. Como nos espaços euclidianos o produto escalar é definido por matrizes simétricas torna-se natural a escolha de conexões afins simétricas.

Definição 1.5 *Uma conexão afim ∇ é dita simétrica se satisfaz*

$$\nabla_X Y - \nabla_Y X = [X, Y], \quad \forall X, Y \in \chi(M),$$

onde o campo vetorial $[X, Y] = XY - YX$ é chamado de colchete de X, Y .

Outra questão básica é manter a regra usual de derivada do produto. Conexões que preservam tal propriedade são denominadas *compatíveis com a métrica*.

Definição 1.6 *Seja M uma variedade diferenciável com uma conexão afim ∇ e uma métrica riemanniana $\langle \cdot, \cdot \rangle_{(\bullet)}$. Dizemos que ∇ é compatível com a métrica $\langle \cdot, \cdot \rangle_{(\bullet)}$ se para todo par de campo de vetores X, Y ao longo de uma curva diferenciável $\varphi : [a, b] \subset \mathbb{R} \rightarrow M$,*

$$\frac{d}{dt} \langle X, Y \rangle_{\varphi(t)} = \langle \frac{DX}{dt}, Y \rangle_{\varphi(t)} + \langle X, \frac{DY}{dt} \rangle_{\varphi(t)}$$

Teorema 1.1 (Levi-Civita) *Dada uma variedade riemanniana M , existe uma única conexão afim ∇ associada a M satisfazendo as condições:*

- (i) ∇ é simétrica;
- (ii) ∇ é compatível com a métrica riemanniana.

Esta conexão é denominada conexão de Levi-Civita ou Riemanniana.

Um campo de vetores X ao longo de uma curva φ é dito paralelo se $\frac{D}{dt}X = 0$, para todo $t \in (a, b)$.

Definição 1.7 *Uma curva diferenciável $\gamma : [a, b] \rightarrow M$ é uma geodésica se γ' é paralelo.*

Decorre da definição que

$$\begin{aligned} \frac{d}{dt} \langle \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} &= \\ &= \langle \frac{D}{dt} \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} + \langle \gamma'(t), \frac{D}{dt} \gamma'(t) \rangle_{\gamma(t)} = \langle 2 \frac{D}{dt} \gamma'(t), \gamma'(t) \rangle_{\gamma(t)} = 0, \end{aligned} \quad (1.4)$$

ou ainda, γ é normal.

Seja \exp_x a aplicação de $T_x M$ em M que associa a cada vetor $\nu \in T_x M$ o ponto $\gamma_\nu(1)$, onde γ_ν é a geodésica que passa por x com velocidade ν . Se \exp_x é um difeomorfismo em uma vizinhança V da origem em $T_x M$, $\exp_x V = U$ é chamada uma *vizinhança normal* de x . Se uma bola aberta em torno da origem de $T_x M$ ($B_\epsilon(0)$) é tal que $B_\epsilon(0) \subset V$, chamamos $\exp_x B_\epsilon(0) = B_\epsilon(x)$ a *bola normal* ou *bola geodésica* de centro x e raio ϵ .

Definição 1.8 *Uma variedade riemanniana M é geodésicamente completa se para todo $x \in M$, a aplicação exponencial, \exp_x , está definida para todo $\nu \in T_x M$, isto é, se as geodésicas $\gamma(t)$ que partem de x estão definidas para todos os valores do parâmetro $t \in \mathbb{R}$.*

Teorema 1.2 (Hopf e Rinow) *Seja M uma variedade riemanniana e $x \in M$. As seguintes afirmações são equivalentes:*

- (i) \exp_x está definida em todo $T_x M$;
- (ii) os limitados e fechados de M são compactos;
- (iii) M é completa como espaço métrico;
- (iv) M é geodesicamente completa.
- (v) existe uma sucessão de compactos $K_n \subset M$, $K_n \subset K_{n+1}$ e $\bigcup_n K_n = M$, tais que se $y_n \notin K_n$ então $d(x, y_n) \rightarrow \infty$.

Além disso, cada uma das afirmações acima implica que

- (vi) Para todo $y \in M$ existe uma geodésica γ ligando x a y com $L(\gamma) = d(x, y)$.

1.4 Geometria riemaniana gerada por hessiana de barreiras autoconcordantes logaritmicamente homogêneas

Sejam $x \in \text{int } Q$ e F uma barreira convexa e autoconcordante definida em Q , tal que F é não-degenerada em x , ou seja, o subespaço E_F de E , dado por

$$E_F = \{h \in E; \langle F''(x)h, h \rangle = 0\},$$

contém apenas o elemento neutro de E . A hessiana de F induz uma norma local $\|\bullet\|_x$, definida por

$$\|h\|_x = (\langle F''(x)h, h \rangle)^{\frac{1}{2}}.$$

A necessidade da não-degenerescência de F vem de que a convexidade garante apenas a semidefinição positiva da hessiana. Em alguns casos (como o da barreira padrão para o problema de Programação Semidefinida que estudaremos nos próximos capítulos) esta hipótese é garantida pela própria construção da barreira.

Considere agora dois pontos $x, y \in \text{int } Q$ e uma curva suave $c(t)$, $t \in (0, 1)$, satisfazendo as condições de fronteira $c(0) = x$, $c(1) = y$ e $c(t) \in \text{int } Q$, $\forall t \in (0, 1)$. O comprimento de c medido através da norma induzida por F'' é dado por

$$L(c) = \int_0^1 \|c'(t)\|_{c(t)} dt,$$

e a distância riemaniana entre x e y é

$$d(x, y) = \inf_{c \in C_{x,y}} L(c).$$

O resultado que segue reúne uma série de propriedades de geometria riemaniana. Algumas delas são extensões do Teorema de Hopf e Rinow para variedades completas [18].

Teorema 1.3 *Seja Q um subconjunto convexo e fechado de E tal que $\text{int } Q \neq \emptyset$ e $F : \text{int } Q \rightarrow \mathbb{R}$ uma barreira convexa e autoconcordante e não-degenerada. As seguintes afirmações são verdadeiras:*

(i) *$(\text{int } Q, d)$ é um espaço métrico completo;*

(ii) *para quaisquer $x, y \in \text{int } Q$, existe uma geodésica γ ligando x a y com*

$$L(\gamma) = d(x, y);$$

(iii) *Seja γ uma geodésica dada como em (ii). Então*

$$F''(\gamma(t))\gamma''(t) + \frac{1}{2}F'''(\gamma(t))[\gamma'(t), \gamma'(t)] = 0, \quad \forall t \in (0, 1); \quad (1.5)$$

(iv) *Considere γ como em (ii). Se Q é um cone e F é logaritmicamente homogênea em Q então a função $\langle F'(\gamma(t)), \gamma'(t) \rangle$ é constante em t .*

Demonstração: (i) A afirmação que d é uma métrica pode ser facilmente comprovada através da definição de ínfimo e de conceitos como bola normal. Vários textos clássicos de geometria riemaniana fornecem tal demonstração. Ver, por exemplo, [18]. O que diz respeito a completeza de $(\text{int } Q, d)$ deixaremos para o final da seção por não depender dos demais resultados. (ii) segue imeditamente do Teorema de Hopf e Rinow. (iii): Seja γ uma geodésica satisfazendo (ii). Temos que,

$$\begin{aligned}
\frac{d}{dt}\langle F'''(\gamma(t))\gamma'(t), \gamma'(t) \rangle &= \langle (F''''(\gamma(t))\gamma'(t))\gamma'(t) + 2F'''(\gamma(t))\gamma''(t), \gamma'(t) \rangle. \text{ Identificando os termos correspondentes em (1.4) e nesta última expressão, temos a derivada covariante de } \gamma' \text{ dada por } \frac{D}{dt}\gamma'(t) = \frac{1}{2}F''''(\gamma(t))\gamma'(t)\gamma'(t) + F'''(\gamma(t))\gamma''(t), \text{ de onde conclui-se pela Definição (1.7) a terceira parte do teorema. (iv): Utilizando a Proposição 1.1 e aplicando (iii), temos} \\
\frac{d}{dt}\langle F'(\gamma(t)), \gamma'(t) \rangle &= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle + \langle F'(\gamma(t)), \gamma''(t) \rangle \\
&= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle - \langle F''(\gamma(t))\gamma(t), \gamma''(t) \rangle \quad (\text{Prop. 1.1 (ii)}) \\
&= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle - \langle F''(\gamma(t))\gamma''(t), \gamma(t) \rangle \quad (\text{simetria de } F'') \\
&= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle + \langle \frac{1}{2}(F''''(\gamma(t))\gamma'(t))\gamma'(t), \gamma(t) \rangle \quad (\text{por (1.5)}) \\
&= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle + \langle \frac{1}{2}(F''''(\gamma(t))\gamma(t))\gamma'(t), \gamma'(t) \rangle \quad (F'''' \text{ é trilinear simétrica}) \\
&= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle - \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle = 0. \quad (\text{Prop. 1.1 (iii)}) \quad \blacksquare
\end{aligned}$$

A partir deste momento, destacaremos alguns resultados de pontos interiores que possibilitaram verificar a completeza descrita no item (i) do Teorema (1.3).

Um importante conceito no estudo de métodos de pontos interiores é o elipsóide de Dikin, que apresentamos a seguir.

Definição 1.9 *Seja $x \in \text{int } Q$. Uma r -vizinhança de x em relação à norma induzida por $F''(x)$ ($W_r(x) = \{y : \|x - y\|_x \leq r\}$) é chamada de elipsóide de Dikin de centro x e raio r .*

Proposição 1.3 *Sejam F uma função a -autoconcordante em um convexo $Q \subset E$ e $x \in Q$. Então*

(i) *para cada $y \in Q$ tal que $r \equiv \|x - y\|_x \leq 1$,*

$$(1-r)^2 D^2 F(x)[h, h] \leq D^2 F(y)[h, h] \leq \frac{1}{(1-r)^2} D^2 F(x)[h, h], \forall h \in E;$$

(ii) se F é uma barreira autoconcordante para Q , então todo elipsóide de Dikin de centro x e raio $r < 1$ está contido em Q .

Demonstração: Ver [4], páginas 13, 14 e 15. ■

Lema 1.2 Dados F uma barreira autoconcordante em Q e $x, y \in \text{int } Q$ tal que $r := \|x - y\|_x < 1$. Então

$$r - \frac{1}{2}r^2 \leq d(x, y) \leq -\ln(1-r). \quad (1.6)$$

E ainda, se $d(x, y) \leq k - \frac{1}{2}k^2$ para algum $k \in [0, 1)$, então $r \leq k < 1$ e portanto, a desigualdade (1.6) é assegurada.

Demonstração: Seja $\eta(t) = ty + (1-t)x$, $t \in (0, 1)$.

$$\begin{aligned} d(x, y) &\leq \int_0^1 \|\eta'(t)\|_{\eta(t)} dt \quad (\text{por 1.3}) \\ &= \int_0^1 \langle F''(\eta(t))(y-x), (y-x) \rangle^{\frac{1}{2}} dt \leq \int_0^1 \left(\frac{\langle F''(x)(y-x), (y-x) \rangle}{(1-\|x-\eta(t)\|_x)^2} \right)^{\frac{1}{2}} dt \\ &(\text{Prop. 1.3 (i), segunda desigualdade}) \\ &= \int_0^1 \frac{\langle F''(x)(y-x), (y-x) \rangle^{\frac{1}{2}}}{1-t\|y-x\|_x} dt = \int_0^1 \frac{r}{1-tr} dt = -\ln(1-r). \end{aligned}$$

Para o limite inferior, considere uma geodésica γ ligando x a y e tome \bar{t} igual ao ponto em que γ atinge a fronteira do elipsóide $\{z : \|z - x\|_x \leq r\}$ pela primeira vez. Note que tal ponto existe e está bem definido (no pior caso $\bar{t} = 1$), pois $\gamma(1) = y$ e por definição $\|y - x\|_x = r$,

$$\|\gamma(1) - x\|_x = \|y - x\|_x = r.$$

Defina agora uma nova função δ , dada por $\delta(t) = \|\gamma(t) - x\|_x = \langle F''(x)(\gamma(t) - x), (\gamma(t) - x) \rangle^{\frac{1}{2}}$, $t \in (0, 1)$. Temos $\delta'(t) = \frac{1}{2} \langle F''(x)(\gamma(t) -$

$x), (\gamma(t) - x)\rangle^{-\frac{1}{2}} (\langle F''(x)\gamma'(t), (\gamma(t) - x)\rangle + \langle F''(x)(\gamma(t) - x), \gamma'(t)\rangle) =$
 $\frac{1}{\delta(\bar{t})}\langle F''(x)(\gamma(t) - x), \gamma'(t)\rangle = \frac{1}{\delta(\bar{t})}\langle \gamma(t) - x, \gamma'(t)\rangle_x$. Aplicando a desigualdade
 de Cauchy-Schwarz, vem que $\delta'(t) \leq \frac{1}{\delta(t)}\|\gamma(t) - x\|_x\|\gamma'(t)\|_x = \|\gamma'(t)\|_x$.

Considere

$$\begin{aligned}
 d(x, y) &= \\
 &= \int_0^{\bar{t}} \|\gamma'(t)\|_{\gamma(t)} dt \geq \int_0^{\bar{t}} \|\gamma'(t)\|_{\gamma(t)} dt = \int_0^{\bar{t}} \langle F''(\gamma(t))\gamma'(t), \gamma'(t)\rangle^{\frac{1}{2}} dt \\
 &\geq \int_0^{\bar{t}} ((1 - \|\gamma(t) - x\|_x)^2 \langle F''(x)\gamma'(t), \gamma'(t)\rangle)^{\frac{1}{2}} dt \quad (\text{Prop. 1.3 (i), primeira} \\
 &\text{desigualdade}) \\
 &= \int_0^{\bar{t}} (1 - \|\gamma(t) - x\|_x)\|\gamma'(t)\|_x dt \geq \int_0^{\bar{t}} (1 - \delta(t))\delta'(t) dt = \delta(\bar{t}) - \frac{1}{2}\delta^2(\bar{t}) = \\
 &r - \frac{1}{2}r^2. \quad \text{Por outro lado, assumamos que } r = \|x - y\|_x > k. \quad \text{Escolha}
 \end{aligned}$$

$k' \in (k, \min\{1, r\})$ e construa o elipsóide de Dikin com centro em x e raio k' .

Então, qualquer geodésica ligando x a y deve atingir a fronteira do elipsóide pela primeira vez em algum ponto y' . Temos então que $\|y' - x\|_x \leq k' < 1$.

Podemos portanto afirmar que $d(x, y) \geq d(x, y') \geq k' - \frac{1}{2}(k')^2 > k - \frac{1}{2}k^2$, chegamos a uma contradição, onde a segunda desigualdade provém da aplicação do limite inferior da primeira parte deste lema.

■

Encerraremos esta seção com um corolário do Lema (1.2) que completa o Teorema (1.3).

Corolário 1.1 *(int Q, d) é completo.*

Demonstração: Seja (x_n) uma sequência de Cauchy e escolha N natural tal que $d(x_l, x_m) \leq \frac{3}{8} = \frac{1}{2} - \frac{1}{2}(\frac{1}{2})^2$ para todo $l, m \geq N$. Pela segunda parte do Lema(1.2), $\|x_m - x_l\|_{x_l} \leq \frac{1}{2}$, para todo $l, m \geq N$. Em particular, $\|x_n - x_N\|_{x_N} \leq \frac{1}{2}$, para todo $n \geq N$. Defina $\delta' := \max_{i=1, \dots, N} \|x_i - x_N\|_{x_N}$

e escolha $\delta = \max\{\delta', \frac{1}{2}\}$. A seqüência (s_n) dada por, $s_n = \|x_n - x_N\|_{x_N}$, é limitada por 0 e δ . Isto implica que (s_n) deve possuir uma subsequência convergente e portanto (x_n) também. Denote por (x_{n_k}) e x a subsequência convergente de (x_n) e seu limite, respectivamente. É fácil verificar que $x_n \rightarrow x$. Com efeito, dado $\epsilon > 0$, existe n' natural, tal que $n_k \geq n'$ implica em $d(x_{n_k}, x) < \frac{\epsilon}{2}$. Existe também n'' natural, tal que $d(x_m, x_n) < \frac{\epsilon}{2}$ sempre que $m, n \geq n''$. Defina $n_0 = \max\{n', n''\}$. Para $n \geq n_0$, pode-se também escolher $n_k \geq n_0$ e obter $d(x_n, x) \leq d(x_n, x_{n_k}) + d(x_{n_k}, x) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$. ■

1.5 Propriedades da métrica riemaniana

A métrica d definida em (1.3) satisfaz algumas propriedades importantes para determinação de curvas geodésicas em alguns cones. Dentre elas, destacam-se quatro: *desigualdade triangular*, *simetria*, *invariância por transformações lineares não-singulares* e *homogeneidade*.

As duas primeiras compõem o corpo de axiomas de uma métrica e sua verificação pode ser vista em do Carmo [18]. A seguir mostraremos as duas últimas propriedades.

Agora, sejam Q um cone convexo fechado em um espaço vetorial E (real, de dimensão n), $A_{n \times n} : E \rightarrow E$ uma transformação linear não-singular e F uma barreira autoconcordante e logaritmicamente homogênea em Q . Considere o conjunto $\bar{Q} = \{y = Ax : x \in Q\} = \{y : A^{-1}y \in Q\}$. Podemos estabelecer uma barreira \bar{F} em \bar{Q} induzida por F e conseqüentemente as noções de fronteira e interior em \bar{Q} dependem de Q . De fato, defina \bar{F} por

$$\bar{F}(y) := F(A^{-1}(y)).$$

Definição 1.10 *Sejam Q , \bar{Q} , F e \bar{F} descritos como no parágrafo anterior.*

Diremos que um ponto $y \in \partial\bar{Q}$ se $A^{-1}y = x \in \partial Q$. De maneira análoga, $y \in \text{int } \bar{Q}$ se $A^{-1}y = x \in \text{int } Q$.

Proposição 1.4 *Sejam $x_1, x_2 \in \text{int } Q$ e γ uma geodésica satisfazendo $\gamma(0) = x_1$, $\gamma(1) = x_2$ e $\gamma(t) \in \text{int } Q, \forall t \in (0, 1)$. Então*

$$L(\gamma) = L(\eta), \quad (1.7)$$

para η definida por $\eta(t) := A\gamma(t)$.

Demonstração: Com efeito, dados $y \in \text{int } \bar{Q}$ e $\bar{h}_1, \bar{h}_2 \in E_A$,

$$\begin{aligned} D\bar{F}(y)[\bar{h}_1] &= \\ &= \left. \frac{d}{dt} \right|_{t=0} \bar{F}(y + t\bar{h}_1) = \left. \frac{d}{dt} \right|_{t=0} F(A^{-1}(y + t\bar{h}_1)) = \left. \frac{d}{dt} \right|_{t=0} F(A^{-1}y + tA^{-1}\bar{h}_1) \\ &= \langle DF(A^{-1}y + tA^{-1}\bar{h}_1), A^{-1}\bar{h}_1 \rangle \Big|_{t=0} = DF(A^{-1}y)[A^{-1}\bar{h}_1]. \end{aligned}$$

$$\begin{aligned} D^2\bar{F}(y)[\bar{h}_1, \bar{h}_2] &= \\ &= \left. \frac{d}{dt} \right|_{t=0} D\bar{F}(y + t\bar{h}_2)[\bar{h}_1] = \left. \frac{d}{dt} \right|_{t=0} DF(A^{-1}(y + t\bar{h}_2))[A^{-1}\bar{h}_1] \\ &= \langle D^2F(A^{-1}y + tA^{-1}\bar{h}_2)A^{-1}\bar{h}_1, A^{-1}\bar{h}_2 \rangle \Big|_{t=0} = D^2F(A^{-1}y)[A^{-1}\bar{h}_1, A^{-1}\bar{h}_2]. \end{aligned}$$

Por outro lado, $\eta'(t) = A\gamma'(t)$. Logo,

$$\begin{aligned} D^2\bar{F}(\eta(t))[\eta'(t), \eta'(t)] &= D^2F(A^{-1}\eta(t))[A^{-1}\eta'(t), A^{-1}\eta'(t)] \\ &= D^2F(\gamma(t))[\gamma'(t), \gamma'(t)]. \end{aligned}$$

Logo, $\|\eta'(t)\|_{\eta(t)} = \|\gamma'(t)\|_{\gamma(t)}$. O que implica (1.7). ■

Considerando que em alguns espaços uma transformação linear pode ser aplicada à direita de x (por exemplo, x pode ser uma variável no cone das matrizes simétricas definidas positivas), podemos mostrar de maneira semelhante que a Proposição 1.4 também é válida neste caso.

A propriedade de homogeneidade assegura que a distância entre $x, y \in \text{int } Q$ permanece inalterada quando multiplicamos x e y por um escalar

estritamente positivo. De fato, dado $\tau > 0$, $\tau\gamma$ é uma geodésica ligando τx a τy . Diferenciando a equação (1.1) duas vezes, com respeito à x , vem que

$$\tau^2 F''(\tau x) = F''(x). \quad (1.8)$$

Logo,

$$\begin{aligned} d^2(\tau x, \tau y) &= \langle F''(\tau\gamma(t))\tau\gamma'(t), \tau\gamma'(t) \rangle = \langle \frac{1}{\tau^2} F''(\gamma(t))\tau\gamma'(t), \tau\gamma'(t) \rangle && \text{(por 1.8)} \\ &= \langle F''(\gamma(t))\gamma'(t), \gamma'(t) \rangle = d^2(x, y). \end{aligned}$$

Donde se conclui que,

$$d(\tau x, \tau y) = d(x, y), \quad \forall \tau > 0.$$

Capítulo 2

O Cone das Matrizes Simétricas Semidefinidas Positivas

2.1 Introdução

Denote por S^n , S_+^n e S_{++}^n os conjuntos das matrizes simétricas, simétricas semidefinidas positivas e simétricas definidas positivas, respectivamente. O segundo e terceiro conjunto têm sido objeto de estudo de diversos pesquisadores nas últimas décadas. Isto se deve à descoberta de uma classe de problemas de otimização conhecida como Programação Semidefinida (*PSD*).

Os problemas de Programação Semidefinida aparecem em diferentes campos do conhecimento científico, por exemplo, em Engenharia e Otimização Combinatória. Sua formulação pode ser vista ainda como extensão da Programação Linear (*PL*), com as restrições de não-negatividade vetorial repassadas pela semidefinição positiva matricial.

Podemos destacar o grande número de trabalhos voltados a relaxações de problemas de Combinatória que possibilitam a aproximação do problema original como um (*PSD*) (ver, por exemplo, a reformulação do problema de Max-Cut como um (*PSD*) em [19]). Devem ser citadas ainda as extensões

de algoritmos primais-duais de pontos interiores em (PL) para Programação Semidefinida. Maiores detalhes sobre teoria e propostas algorítmicas podem ser encontradas em [1].

Na verdade um (PSD) consiste em um problema de Programação Convexa podendo envolver funções não-lineares. Dessa forma, seu universo de aplicação é muito mais amplo que a Programação Linear. Essa classe representa um caso particular de Programação Não-linear.

O primal de Programação Semidefinida consiste em

$$\{\min C \bullet X; \text{ s.a. } A_i \bullet X = b_i, i = 1, \dots, m, X \succeq 0\}, \quad (2.1)$$

onde as matrizes simétricas $C, A_i, i = 1 \dots m$ e $b \in R^m$ são dados do problema. A operação \bullet representa o produto interno $X \bullet Y = \sum_{i,j=1}^n x_{ij}y_{ij}$ definido para matrizes simétricas X, Y e a restrição $X \succeq 0$ indica que a matriz simétrica X é semidefinida positiva.

Neste capítulo apresentaremos inicialmente alguns resultados de Álgebra Linear que podem ser encontrados em [20] e [21], a seguir destacaremos algumas propriedades da barreira padrão para o cone das matrizes simétricas semidefinidas positivas (as propriedades podem ser vistas em [2] e [3])

$$F(X) = -\ln \det(X). \quad (2.2)$$

Encerraremos o capítulo com a expressão da distância riemanniana entre dois pontos quaisquer de S_{++}^n (ver [14] e [15]). Apenas acrescentamos algumas demonstrações omitidas em alguns resultados e detalhamos os cálculos de algumas derivadas em S_{++}^n para facilitar a compreensão do leitor.

2.2 Alguns resultados de Álgebra Linear

O produto escalar do espaço euclidiano ($x^T y$; $x, y \in R^n$) pode ser facilmente estendido ao espaço de matrizes retangulares levando-se em consideração a aplicação $\text{vec} : M_{m \times n} \rightarrow R^{n \cdot m}$ que transforma matrizes em vetores. Tal aplicação pode ser descrita da seguinte maneira:

- alocar primeira coluna da matriz;
- alocar coluna posterior imediatamente abaixo da anterior;
- repetir procedimento até que a última coluna seja alocada.

Com isso, fica definido o seguinte produto escalar entre matrizes

$$A \bullet B = \text{vec}(B)^T \text{vec}(A) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} \cdot b_{ij} = \text{Tr } B^T A.$$

A norma resultante é

$$\|A\|_{\bullet} = (A \bullet A)^{\frac{1}{2}} = (\text{Tr } A^T A)^{\frac{1}{2}} = \|A\|_F,$$

onde $\|A\|_F$ é a norma de Frobenius de A .

Se A e B são matrizes simétricas de ordem n então

$$A \bullet B = \text{Tr } B^T A = \text{Tr } (BA)^T = \text{Tr } A^T B^T = \text{Tr } AB.$$

Lema 2.1 *O polinômio característico associado a um operador simétrico $A_{n \times n}$ possui n raízes reais.*

Demonstração: Ver [20]. ■

Lema 2.2 *Os autovetores de uma matriz simétrica $A_{n \times n}$ associados a autovalores distintos são dois a dois ortogonais.*

Demonstração: Sejam x_i, x_j autovetores de A associados aos autovalores λ_i, λ_j distintos. Devido a simetria de A ,

$$(Ax_i)^T x_j = x_i^T Ax_j$$

$$(\lambda_i x_i)^T x_j = x_i^T (\lambda_j x_j)$$

$$(\lambda_i - \lambda_j)x_i^T x_j = 0.$$

Mas $\lambda_i - \lambda_j \neq 0$. Logo, $x_i^T x_j = 0$, donde se conclui o lema. ■

Teorema 2.1 (Teorema espectral para matrizes simétricas) *Uma matriz $A_{n \times n}$ é simétrica se, e somente se, existe uma matriz $Q_{n \times n}$ tal que $Q^T Q = I$ e $Q^T A Q = \Lambda$, onde Λ é real e diagonal.*

Demonstração: Suponha que A é simétrica. Tomando Q de forma que Q_i seja igual ao i -ésimo autovetor normalizado de A , onde Q_i é a i -ésima coluna de Q , vem, pelo Lema 2.2, que Q é ortogonal e

$$Q_i^T A Q_i = Q_i^T (\lambda_i Q_i) = \lambda_i Q_i^T Q_i = \lambda_i,$$

com $\lambda_i \in \mathbb{R}$ (Lema 2.1). A implicação no sentido contrário é trivial, donde se conclui o teorema. ■

$\Lambda = Q^T A Q$ é a decomposição de Schur ou decomposição em autovalores da matriz simétrica A .

É importante reafirmar que matrizes simétricas possuem todos os seus autovalores reais (Lema 2.1).

Lema 2.3 (Weyl) *Sejam A e B matrizes simétricas de ordem n e um inteiro k , $1 \leq k \leq n$ fixo. Então*

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A + B) \leq \lambda_k(A) + \lambda_n(B),$$

onde $\lambda_1(B)$ e $\lambda_n(B)$ representam o menor e o maior autovalor de B , respectivamente.

Demonstração: Ver [25]. ■

Teorema 2.2 *Dada uma matriz $A \in S^n$, um vetor $x \in R^n$ e uma escalar $c \in R$, defina $\bar{A} \in S^{n+1}$ por*

$$\bar{A} = \begin{bmatrix} A & x \\ x^T & c \end{bmatrix}.$$

Denote por $\{\lambda_i ; i = 1, \dots, n\}$ e $\{\bar{\lambda}_i ; i = 1, \dots, n+1\}$ os autovalores de A e \bar{A} , respectivamente. Assuma também que os autovalores estão arranjados em ordem não-decrescente, isto é, $\lambda_1 \leq \dots \leq \lambda_n$ e $\bar{\lambda}_1 \leq \dots \leq \bar{\lambda}_{n+1}$. Então

$$\bar{\lambda}_1 \leq \lambda_1 \leq \bar{\lambda}_2 \leq \lambda_2 \leq \dots \leq \lambda_{n-1} \leq \bar{\lambda}_n \leq \lambda_n \leq \bar{\lambda}_{n+1}$$

Demonstração: Ver [20]. ■

Este resultado mostra que os autovalores de A e \bar{A} estão “interligados” entre si, ou seja, entre dois autovalores consecutivos de \bar{A} existe um único autovalor de A .

Para os próximos resultados é importante lembrar o conceito de matriz semidefinida positiva (definida positiva).

Definição 2.1 Uma matriz quadrada A de ordem n é *semidefinida positiva* se $x^T Ax \geq 0$, para todo $x \in R^n$ e *definida positiva* se $x^T Ax > 0$, para todo $x \neq 0$.

Proposição 2.1 Seja B uma matriz não-singular de ordem n . $A \in S^n$ é *semidefinida positiva* se, e somente se $B^T AB$ é *semidefinida positiva*.

O resultado se mantém para matrizes definidas positivas. Demonstraremos somente a semidefinição positiva. A definição positiva segue da mesma maneira considerando $x \neq 0$.

Demonstração: Com efeito, A é semidefinida positiva se, e somente se $x^T Ax \geq 0$, para todo $x \in R^n$. Portanto, $[x^T (B^{-1})^T] B^T AB [B^{-1}x] \geq 0$, fazendo $y = B^{-1}x$, vem que $y^T B^T AB y \geq 0$. Mas, a correspondência entre x e y é biunívoca. Logo, $B^T AB$ é semidefinida positiva. ■

O resultado a seguir reúne diversas caracterizações para S_{++}^n .

Teorema 2.3 Seja A uma matriz simétrica de ordem n . São equivalentes as seguintes afirmações:

- (i) A é *definida positiva*;
- (ii) $\lambda_i(A) > 0$, $i = 1 \cdots n$, onde $\lambda_i(A)$ é o i -ésimo autovalor de A ;
- (iii) Existe uma matriz quadrada C de ordem n , com $\text{posto}(C) = n$, tal que $A = C^T C$;
- (iv) $\det(A_i) > 0$, onde A_i , $i = 1 \cdots n$ são as submatrizes principais de A .

Demonstração: A demonstração seguirá os seguintes passos $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$ e $(i) \Leftrightarrow (iv)$.

$(i) \Rightarrow (ii)$ Com efeito, se $A \in S^n$ é definida positiva, tomando x^i igual ao i -ésimo autovetor (suposto normalizado) de A , vem que $(x^i)^T A x^i = \lambda_i > 0$, $i = 1, \dots, n$.

$(ii) \Rightarrow (iii)$ Seja $\Lambda = Q^T A Q$, a decomposição de Schur de A . Como Λ é diagonal com os elementos da diagonal estritamente positivos, $\Lambda^{-\frac{1}{2}} Q^T A Q \Lambda^{-\frac{1}{2}} = I$ ($\Lambda^{-\frac{1}{2}}$ é a matriz diagonal onde cada elemento da diagonal é o inverso da raiz quadrada de um autovalor de A). Fazendo $\bar{C} = Q \Lambda^{-\frac{1}{2}}$, temos que \bar{C} é invertível (e portanto, seu posto é n). Dessa forma, $A = (\bar{C}^{-1})^T \bar{C}^{-1}$. Logo, basta escolher $C = \bar{C}^{-1}$.

$(iii) \Rightarrow (i)$ De fato, $x^T A x = x^T C^T C x = (C x)^T C x = \|C x\|_2^2 > 0, \forall x \neq 0$ (C é não-singular, pois $\text{posto}(C) = n$).

$(i) \Rightarrow (iv)$ se A é definida positiva então $x^T A x > 0$ para todo $x \in R^n, x \neq 0$. Em particular, para qualquer vetor da forma $x = (x_1, \dots, x_i, 0, \dots, 0)^T$, onde as componentes com índices maiores que i são todas nulas,

$$0 < x^T A x = (x^i)^T A_i x^i,$$

onde A_i é a k -ésima submatriz menor principal de A . Donde se conclui que $A_i \succ 0$. Sabendo que $\det A_i$ é igual ao produto dos autovalores de A_i e aplicando o item (ii) vem que $\det A_i > 0$. Quando $i = n$, $A_n = A$ e por hipótese A é definida positiva. Portanto, $\det A_n = \det A > 0$ (pelo item (ii)).

$(iv) \Rightarrow (i)$ Por indução sobre i ($i = 1 \dots n$). Temos que $\det(A_1) > 0$ e A_1 é um escalar, portanto A_1 é definida positiva. Suponha agora que A_i é definida positiva. Pelo Teorema (2.2), podemos assegurar que os autovalores de A_{i+1} são estritamente positivos, excetuando o menor deles. Suponha que este seja

não-positivo. Como o determinante de uma matriz é igual ao produto de seus autovalores, $\det(A_{i+1}) \leq 0$. Temos uma Contradição. Logo $\lambda_i(A) > 0$, $i = 1 \cdots n$. ■

Um resultado semelhante pode ser enunciado para matrizes simétricas semidefinidas positivas.

Teorema 2.4 *Seja $A \in S^n$. São equivalentes as seguintes afirmações:*

- (i) *A é semidefinida positiva;*
- (ii) *$\lambda_i(A) \geq 0$, $i = 1, \dots, n$;*
- (iii) *existe $C \in M_{m \times n}$, tal que $A = C^T C$, com $\text{posto}(C) = \text{posto}(A)$;*
- (iv) *$A \bullet B \geq 0$, para toda matriz B semidefinida positiva.*

Demonstração: Ver [20]. ■

Em relação ao Teorema (2.3), as modificações nos itens (ii), (iii) e (iv) devem-se ao fato de que A pode ser singular.

Os autovalores de uma matriz quadrada A são as raízes do polinômio característico

$$P(\lambda) = \det(A - \lambda I),$$

que depende continuamente dos elementos de A . Portanto, para cada matriz simétrica A definida positiva, deve existir $\epsilon > 0$, tal que a bola aberta de centro A e raio ϵ contém apenas matrizes cujos autovalores são estritamente

positivos. Logo, S_{++}^n constitui o interior de S_+^n , e sua fronteira é formada pelas matrizes que possuem pelo menos um autovalor nulo.

Definição 2.2 *Um cone $K \subset S^n$ é dito autopolar (ou autodual) se o seu cone polar (ou dual), definido por*

$$K^* := \{Y \in S^n : X \bullet Y \geq 0, \forall X \in K\},$$

é o próprio K .

É fácil perceber, pelo item (iv) do Teorema (2.4), que S_+^n é autopolar ou autodual.

Uma discussão mais profunda destes resultados distancia o objetivo principal deste trabalho, por isso preferimos apresentá-los informalmente. Para maiores detalhes ver [21].

2.3 Uma barreira autoconcordante em S_{++}^n

O sucesso dos algoritmos primais-duais em Programação Semidefinida depende da escolha de uma barreira para o interior de S_+^n . Existem diferentes propostas de barreira para este cone (por exemplo, a função traço [22]). No entanto, determinadas escolhas podem omitir propriedades relevantes a respeito do problema.

Neste trabalho, utilizaremos a barreira padrão (2.2), apresentada em Rothaus[15], por se tratar de uma função *analítica, estritamente convexa e autoconcordante* (ver [4]).

A priori, (2.2) parece artificialmente construída. Mas, com base no Teorema (2.3), pode-se verificar que essa talvez represente a escolha mais natural possível de barreira em S_+^n .

Com efeito, uma das caracterizações de S_{++}^n é

$$\{X \in S^n : \lambda_i(X) > 0, i = 1, \dots, n\}.$$

Para este cone (a comparar com o ortante estritamente positivo de R^n), podemos aplicar a barreira logarítmica

$$F(X) = - \sum_{i=1}^n \ln \lambda_i(X).$$

Verifica-se que tal função é estritamente convexa (ver [23]) e $F \rightarrow \infty$ quando pelo menos um autovalor de X se aproxima de zero.

Por outro lado,

$$F(X) = - \sum_{i=1}^n \ln \lambda_i(X) = -\ln \prod_{i=1}^n \lambda_i(X) = -\ln \det(X).$$

Torna-se evidente a existência de uma relação estrita entre os métodos primais-duais de barreira logarítmica em R^n e os métodos primais-duais para (PSD) baseados na barreira padrão (2.2).

Lema 2.4 *Para quaisquer $a, b \geq 0$, temos*

$$a^{\frac{3}{2}} + b^{\frac{3}{2}} \leq (a + b)^{\frac{3}{2}}.$$

Demonstração: Sejam $a, b \geq 0$, temos

$$a^{\frac{3}{2}} + b^{\frac{3}{2}} = \sqrt{(\sqrt{a^3} + \sqrt{b^3})^2} = \sqrt{a^3 + 2ab\sqrt{ab} + b^3}. \quad (2.3)$$

Por outro lado, podemos verificar que $\sqrt{ab} \leq a + b$. De fato, $(\sqrt{ab})^2 = ab \leq a^2 + 2ab + b^2 = (a + b)^2$, de onde se conclui que $\sqrt{ab} \leq a + b$. Aplicando este fato em (2.3), vem que

$$\begin{aligned} a^{\frac{3}{2}} + b^{\frac{3}{2}} &= \sqrt{a^3 + 2ab\sqrt{ab} + b^3} \leq \sqrt{a^3 + 2ab(a + b) + b^3} \leq \\ &\sqrt{a^3 + 3ab(a + b) + b^3} = \sqrt{(a + b)^3} = (a + b)^{\frac{3}{2}}. \quad \blacksquare \end{aligned}$$

F é analítica, pois é de classe C^∞ (o logarítmo é uma função de classe C^∞) e autoconcordante como veremos através do próximo resultado.

Proposição 2.2 *A função (2.2) é n -logaritmicamente homogênea e 1-autoconcordante em S_{++}^n .*

Demonstração: Dados $X \in S_{++}^n$ e $\tau > 0$,

$$F(\tau X) = -\ln \det(\tau X) = -\ln(\tau^n \det(X)) = F(x) - n \ln \tau,$$

que é o primeiro resultado. Seja agora $H \in S^n$.

$$DF(X)[H] =$$

$$\begin{aligned} &-\frac{d}{dt} \Big|_{t=0} \ln \det(X + tH) = -\frac{d}{dt} \Big|_{t=0} \ln \det[X(I + tX^{-1}H)] \\ &= -\frac{d}{dt} \Big|_{t=0} \ln \det(I + tX^{-1}H). \end{aligned}$$

Usando que $(1 + t\lambda_i)$ é o i -ésimo autovalor de $(I + tX^{-1}H)$ dado que λ_i é o i -ésimo autovalor de $X^{-1}H$, vem que

$$DF(X)[H] = -\frac{d}{dt} \Big|_{t=0} \sum_{i=1}^n \ln(1 + t\lambda_i) = -\sum_{i=1}^n \lambda_i = -\text{Tr } X^{-1}H.$$

$$D^2F(X)[H, H] =$$

$$\begin{aligned} &-\frac{d}{dt} \Big|_{t=0} \text{Tr } (X + tH)^{-1} H = -\frac{d}{dt} \Big|_{t=0} \text{Tr } [X(I + tX^{-1}H)]^{-1} H \\ &= -\frac{d}{dt} \Big|_{t=0} \text{Tr } (I + tX^{-1}H)^{-1} X^{-1}H = -\frac{d}{dt} \Big|_{t=0} \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i} \\ &= \sum_{i=1}^n \lambda_i^2 = \text{Tr } X^{-1}HX^{-1}H. \end{aligned}$$

$$\begin{aligned}
D^3F(X)[H, H, H] &= \\
&\frac{d}{dt} \Big|_{t=0} \text{Tr} (X + tH)^{-1} H (X + tH)^{-1} H \\
&= \frac{d}{dt} \Big|_{t=0} \text{Tr} (I + tX^{-1}H)^{-1} X^{-1}H (I + tX^{-1}H)^{-1} X^{-1}H \\
&= \frac{d}{dt} \Big|_{t=0} \sum_{i=1}^n \left(\frac{\lambda_i}{1 + t\lambda_i} \right)^2 = 2 \sum_{i=1}^n \lambda_i (-\lambda_i^2) \\
&= -2 \sum_{i=1}^n \lambda_i^3 = -2 \text{Tr} X^{-1}HX^{-1}HX^{-1}H.
\end{aligned}$$

Como X é definida positiva, $X^{-\frac{1}{2}}$ é não-singular. Portanto, $X^{-1}H$ é similar à matriz simétrica $X^{-\frac{1}{2}}HX^{-\frac{1}{2}}$. Logo,

$$\begin{aligned}
|D^3F(X)[H, H, H]| &= \\
2 \left| \text{Tr} \left(X^{-\frac{1}{2}}HX^{-\frac{1}{2}} \right)^3 \right| &= 2 \left| \sum_{i=1}^n \lambda_i^3 \right| \leq 2 \sum_{i=1}^n |\lambda_i|^3 \\
&= 2 \sum_{i=1}^n (\lambda_i^2)^{\frac{3}{2}} \leq 2 \left(\sum_{i=1}^n \lambda_i^2 \right)^{\frac{3}{2}} \quad (\text{Lema 2.4}) \\
&= 2 (D^2F(X)[H, H])^{\frac{3}{2}}. \quad \blacksquare
\end{aligned}$$

A segunda desigualdade da definição de autoconcordância (Definição 1.1) também é assegurada para barreira (2.2) devido a mesma ser logaritmicamente homogênea (ver [4]).

Analisando as diferenciais de 1ª, 2ª e 3ª ordem de F conclui-se que $F'(X) = -X^{-1}$, F'' é a transformação linear que satisfaz

$$F''(X)H = X^{-1}HX^{-1}. \quad (2.4)$$

e F''' é o operador bilinear simétrico dado por

$$(F'''(X)H_1)H_2 = -2X^{-1}H_1X^{-1}H_2X^{-1}, \quad (2.5)$$

com $H_1, H_2 \in S^n$.

De posse das derivadas de F podemos verificar sua convexidade estrita.

Com efeito, dados $X \in S_{++}^n$ e $H \in S^n$,

$$H \bullet F''(X)H = H \bullet (X^{-1}HX^{-1}) = \text{Tr}(X^{-1}HX^{-1}H) = \|X^{-1}H\|_F^2,$$

onde $\|\bullet\|_F$ é a norma de Frobenius que é estritamente positiva para toda matriz simétrica $H \neq 0$, visto que X^{-1} é invertível (e portanto, injetiva), o que mostra que $F''(X)$ é definida positiva e, por conseqüência, F é estritamente convexa.

2.4 Distância em S_{++}^n

Um dos resultados de maior utilidade neste trabalho encontra-se nesta seção.

Aqui, será apresentada a expressão da distância riemaniana em S_{++}^n .

Teorema 2.5 *Seja F a barreira dada em (2.2). Dados $X, Y \in S_{++}^n$, defina por $\lambda_i \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)$ o i -ésimo autovalor de $X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}$. Então*

$$d(X, Y) = \left[\sum_{i=1}^n \ln^2 \lambda_i \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) \right]^{\frac{1}{2}} = \|\ln \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)\| = \left(\text{Tr} \ln^2 X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)^{\frac{1}{2}}, \quad (2.6)$$

Demonstração: Como vimos no capítulo anterior, a métrica riemaniana é invariante segundo transformações não-singulares sobre as variáveis X e Y (propriedade (1.7)). Dessa forma, para quaisquer $X, Y \in S_{++}^n$,

$$d(X, Y) = d\left(I, X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right).$$

Veremos agora que existe uma expressão fechada para os segmentos geodésicos S_{++}^n que permite obter a expressão da distância (2.6). Seja Q a matriz ortogonal cujas colunas são os autovetores de $X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}$, isto é,

$$Q^T \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right) Q = Z,$$

onde Z é diagonal com os autovalores de $X^{-\frac{1}{2}}YX^{-\frac{1}{2}}$ na diagonal. É claro que Z é definida positiva. O problema agora consiste em determinar a distância entre uma matriz diagonal arbitrária definida positiva Z e a identidade. Por outro lado, segundo o Teorema (1.3), uma curva ζ é uma geodésica se satisfaz (1.5). Considerando as expressões das derivadas de 2ª (2.4) e de 3ª (2.5) ordem de F , a equação (1.5) pode ser rescrita como

$$\begin{aligned} 0 &= F''(\zeta(t))\zeta''(t) + \frac{1}{2}F'''(\zeta(t))[\zeta'(t), \zeta'(t)] \\ &= \zeta^{-1}(t)\zeta''(t)\zeta^{-1}(t) - \zeta^{-1}(t)\zeta'(t)\zeta^{-1}(t)\zeta'(t)\zeta^{-1}(t), \end{aligned}$$

que se resume a

$$\zeta''(t) = \zeta'(t)\zeta^{-1}(t)\zeta'(t). \quad (2.7)$$

É fácil verificar que $\bar{\zeta}(t) = Z^t$ é solução da equação diferencial (2.7). De fato, $\bar{\zeta}'(t) = \frac{d}{dt}(Z^t)$. Os elementos não nulos de Z^t pertencem á diagonal e têm a forma Z_{ii}^t . Logo derivando cada elemento da diagonal, em relação a t , obtemos $\ln Z_{ii} Z_{ii}^t$, de que resulta $\bar{\zeta}'(t) = \frac{d}{dt}(Z^t) = \ln(Z)Z^t$. Temos ainda que $\bar{\zeta}''(t) = \frac{d}{dt}(\ln(Z)Z^t) = \ln(Z)\ln(Z)Z^t$. onde $\ln(Z)$ é a matriz diagonal formada pelos logaritmos dos elementos da diagonal de Z . Substituindo na equação (2.7), temos

$$\zeta'(t)\zeta^{-1}(t)\zeta'(t) = \ln(Z)Z^tZ^{-t}\ln(Z)Z^t = \ln(Z)\ln(Z)Z^t = \zeta''(t).$$

Por outro lado, $\bar{\zeta}$ é também a única geodésica ligando I a Z (ver [15]), portanto,

$$\begin{aligned} d(I, Z) &= \\ &= \left(\int_0^1 \langle F''(\bar{\zeta}(t))\bar{\zeta}'(t), \bar{\zeta}'(t) \rangle dt \right)^{\frac{1}{2}} = \left(\int_0^1 \langle F''(Z^t)\ln(Z)Z^t, \ln(Z)Z^t \rangle dt \right)^{\frac{1}{2}} \\ &= \left(\int_0^1 \langle Z^{-t}\ln(Z)Z^tZ^{-t}, \ln(Z)Z^t \rangle dt \right)^{\frac{1}{2}} = \left(\int_0^1 \langle Z^{-t}\ln(Z), \ln(Z)Z^t \rangle dt \right)^{\frac{1}{2}} \\ &= \left(\int_0^1 \text{Tr} \{ \ln(Z)Z^tZ^{-t}\ln(Z) \} dt \right)^{\frac{1}{2}} = \text{Tr} \{ \ln(Z)\ln(Z) \}^{\frac{1}{2}} \end{aligned}$$

¹ Z^t é a matriz diagonal cujas entradas são iguais a t -ésima potência dos elementos da diagonal de Z .

$$= \left(\sum_{i=1}^n \ln^2 \lambda_i(Z) \right)^{\frac{1}{2}}.$$

Mas, os autovalores de Z são os autovalores de $(X^{-\frac{1}{2}}YX^{-\frac{1}{2}})^2$. Logo,

$$d(X, Y) = d\left(I, \left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)\right) = \left(\sum_{i=1}^n \ln^2 \lambda_i \left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right) \right)^{\frac{1}{2}}.$$

■

É importante ressaltar também que S_{++}^n munido da métrica definida pela Hessiana de (2.2) constitui uma variedade riemanniana de curvatura seccional negativa. Tal resultado pode ser encontrado em [15]. Esta hipótese representa o ponto chave do que estaremos propondo no próximo capítulo.

² Z é obtida de $(X^{-\frac{1}{2}}YX^{-\frac{1}{2}})$ por transformações similares.

Capítulo 3

Algoritmo de Ponto Proximal em Programação Semidefinida

3.1 Introdução

Este capítulo apresenta uma nova proposta algorítmica para o problema primal de Programação Semidefinida (2.1) baseada em resultados de convergência pré-estabelecidos em [13]. O método é implementável e reduz potencialmente os custos computacionais envolvidos em cada iteração do algoritmo de ponto proximal clássico estendido ao conjunto das matrizes simétricas semidefinidas positivas munido da métrica induzida pela hessiana de (2.2), substituindo a iteração principal por uma sucessão de problemas de Programação Não-linear no ortante estritamente positivo de R^n . Estamos também aproveitando o fato do método proximal depender apenas da distância dada em (2.6) não sendo necessário conhecer as geodésicas deste espaço.

3.2 Algoritmo de ponto proximal em S_{++}^n

Definição 3.1 Uma função f de uma variedade riemanniana M em R é dita geodesicamente convexa (convexa estrita) se a composição $f \circ \gamma : R \rightarrow R$ é convexa (convexa estrita), para toda geodesica γ em M .

Considere o problema:

$$\{ \min f(X); \text{ s.a } X \succeq 0 \}, \quad (3.1)$$

onde $f : S_+^n \rightarrow R$ é uma função de classe C^1 e geodesicamente convexa em S_{++}^n . Com base na distância (2.6), defina, para cada $X \in S_{++}^n$, a função $\rho_X : S_{++}^n \rightarrow R$ por

$$\rho_X(Y) = \frac{1}{2}d^2(X, Y).$$

Para cada $\beta > 0$, A regularização de Moreau-Yosida $f_\beta : S_{++}^n \rightarrow R$ de f é definida por

$$f_\beta(X) = \min_{Y \succ 0} \{ f(Y) + \beta \rho_X(Y) \}. \quad (3.2)$$

Definição 3.2 $X_\beta = \arg f_\beta$ é chamado ponto proximal de X associado a β, f e ρ_X .

Como f é geodesicamente convexa (limitada inferiormente) e ρ_X é geodesicamente convexa estrita (ver [13]), a função $g_\beta(X, \bullet)$, definida por

$$g_\beta(X, Y) = f(Y) + \beta \rho_X(Y), \quad (3.3)$$

tem no máximo um minimizador. Por outro lado, g_β é 1-coerciva (ver [13]).

Dessa forma, para cada $\beta > 0$ existe um único X_β , tal que

$$f(X_\beta) + \beta \rho_X(X_\beta) = \min_{Y \succ 0} g_\beta(X, Y),$$

caracterizado por

$$\beta \nabla \rho_X(X_\beta) = -\nabla f(X_\beta). \quad (3.4)$$

Quando f é não-diferenciável, a condição (3.4) pode ser rescrita como

$$\beta \nabla \rho_X(X_\beta) \in -\partial f(f_\beta(X)),$$

onde $\partial f(Y)$ representa o *subdiferencial* de f em Y .

A metodologia de ponto proximal consiste em gerar, a partir de um ponto inicial $X^0 \in S_{++}^n$, uma seqüência de minimizadores (X^k) definida por

$$X^{k+1} = \arg \min_{Y \succ 0} f_{\beta^k}(X^k),$$

com β^k satisfazendo

$$\sum_{k=0}^{\infty} \frac{1}{\beta^k} \rightarrow \infty. \quad (3.5)$$

Algoritmo 3.1 (Algoritmo conceitual de ponto proximal em S_{++}^n)

Dados $X^0 \succ 0$, $\beta^0 > 0$ e uma tolerância $\delta > 0$:

passo 1. $k \rightarrow 0$;

passo 2. *repetir*

$$X^{k+1} = \operatorname{argmin}_{Y \succ 0} \left\{ f(Y) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i \left((X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}} \right) \right\}, \quad (3.6)$$

enquanto $\|X^{k+1} - X^k\| > \delta$

atualizar β , k e retornar ao passo 2.

Como consequência dos resultados obtidos em [13], a convergência global do algoritmo é garantida para seqüências (β^k) satisfazendo (3.5) e funções geodesicamente convexas relativamente à métrica induzida pela hessiana de (2.2).

3.3 Um algoritmo em R_{++}^n

Com a finalidade de obter um algoritmo implementável utilizaremos uma importante propriedade da função distância, a *invariância por transformações não singulares*.

Como visto na seção anterior, a função a ser minimizada a cada iteração do Algoritmo (3.1) é dada por

$$g_{\beta^k}(Y) = f(Y) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i \left((X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}} \right). \quad (3.7)$$

Lembremos que a distância (2.6) é obtida através de transformações não-singulares sobre X^k e Y . Levando-se em consideração que X^k está fixada na k -ésima iteração de (3.1), defina a transformação $T_{X^k} : S_{++}^n \rightarrow S_{++}^n$, por

$$T_{X^k}(Y) = (X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}}.$$

É evidente que T_{X^k} é linear e sobrejetiva. Com efeito, a linearidade é trivial. Por outro lado, cada $\bar{Z} \in S_{++}^n$ é imagem de $(X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}}$ por T_{X^k} .

Podemos verificar também que T_{X^k} é uma isometria em relação à métrica (2.6). De fato, T_{X^k} é uma isometria se, e somente se $\langle d_Y T_{X^k} \cdot V, d_Y T_{X^k} \cdot U \rangle_{T_{X^k}(Y)} = \langle V, U \rangle_Y$, para vetores $U, V \in T_Y S_{++}^n = S^n$, onde $d_Y T_{X^k} \cdot V$ é a diferencial de T_{X^k} na direção V . Note que,

$d_Y T_{X^k} \cdot V = (X^k)^{-\frac{1}{2}} V (X^k)^{-\frac{1}{2}}$. Assim,

$$\begin{aligned}
& \langle d_Y T_{X^k} \cdot V, d_Y T_{X^k} \cdot U \rangle_{T_{X^k}(Y)} = \langle F''(T_{X^k}(Y)) d_Y T_{X^k} \cdot V, d_Y T_{X^k} \cdot U \rangle \\
& = \langle (T_{X^k}(Y))^{-1} d_Y T_{X^k} \cdot V (T_{X^k}(Y))^{-1}, d_Y T_{X^k} \cdot U \rangle \\
& = \langle \left((X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}} \right)^{-1} \left((X^k)^{-\frac{1}{2}} V (X^k)^{-\frac{1}{2}} \right) \left((X^k)^{-\frac{1}{2}} Y (X^k)^{-\frac{1}{2}} \right)^{-1}, (X^k)^{-\frac{1}{2}} U (X^k)^{-\frac{1}{2}} \rangle \\
& = \langle (X^k)^{\frac{1}{2}} Y^{-1} (X^k)^{\frac{1}{2}} (X^k)^{-\frac{1}{2}} V (X^k)^{-\frac{1}{2}} (X^k)^{\frac{1}{2}} Y^{-1} (X^k)^{\frac{1}{2}}, (X^k)^{-\frac{1}{2}} U (X^k)^{-\frac{1}{2}} \rangle \\
& = \langle (X^k)^{\frac{1}{2}} Y^{-1} V Y^{-1} (X^k)^{\frac{1}{2}}, (X^k)^{-\frac{1}{2}} U (X^k)^{-\frac{1}{2}} \rangle \quad (\text{note que } \text{Tr}(AB) = \text{Tr}(BA)) \\
& = \langle Y^{-1} V Y^{-1}, U \rangle = \langle F''(Y) V, U \rangle = \langle V, U \rangle_Y.
\end{aligned}$$

Em particular, as isometrias preservam distâncias.

Lema 3.1 *Sejam $Y \in S_{++}^n$ e $\delta > 0$. Então*

$$T_{X^k}(B_\delta(Y)) = B_\delta(T_{X^k}(Y))^1.$$

Demonstração: Como toda isometria é um homeomorfismo, a inclusão $T_{X^k}(B_\delta(Y)) \subset B_\delta(T_{X^k}(Y))$ é imediata. Agora, seja $\bar{Z} \in B_\delta(T_{X^k}(Y))$, então $\delta > d(\bar{Z}, T_{X^k}(Y)) = d\left((X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}}, Y\right)$. De onde se conclui que $(X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}} \in B_\delta(Y)$. Para todo $\bar{Z} \in T_{X^k}(B_\delta(Y))$. ■

Defina agora, a função $\phi : T_{X^k}(S_{++}^n) = S_{++}^n \rightarrow R$ por

$$\phi(\bar{Z}) = f\left((X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}}\right)^2. \quad (3.8)$$

Proposição 3.1 *Se $\bar{\bar{Z}}$ é um minimizador local de ϕ então $\bar{Y} = (X^k)^{\frac{1}{2}} \bar{\bar{Z}} (X^k)^{\frac{1}{2}}$ é minimizador local de f .*

Demonstração: De fato, existe $\epsilon > 0$ tal que $\phi(\bar{\bar{Z}}) \leq \phi(\bar{Z})$ para toda matriz $\bar{Z} \in B_\epsilon(\bar{\bar{Z}})$. Aplicando a definição de ϕ vem que

¹Para $Y \in S_{++}^n$ e $\delta > 0$, $B_\delta(Y)$ é a bola aberta de centro Y e raio δ .

²Aplicamos f sobre a imagem da inversa de T_{X^k} , que é dada por $T_{X^k}^{-1}(\bar{Z}) = (X^k)^{\frac{1}{2}} \bar{Z} (X^k)^{\frac{1}{2}}$.

$$f(\bar{Y}) = f\left((X^k)^{\frac{1}{2}}\bar{Z}(X^k)^{\frac{1}{2}}\right) = \phi(\bar{Z}) \leq \phi(\bar{Z}) = f\left((X^k)^{\frac{1}{2}}\bar{Z}(X^k)^{\frac{1}{2}}\right) = f(Y).$$

Como $T_{X^k}^{-1}\left(B_\epsilon\left(\bar{Z}\right)\right) = B_\epsilon(\bar{Y})$ (Lema (3.1) aplicado à $T_{X^k}^{-1}$), a desigualdade é válida para toda matriz $Y \in B_\epsilon(\bar{Y})$ ³. ■

Portanto, a iteração (3.6) do Algoritmo (3.1) pode ser substituída por

$$\bar{Z}^{k+1} = \operatorname{argmin}_{\bar{Z} \succ 0} \left\{ \phi(\bar{Z}) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(\bar{Z}) \right\}, \quad (3.9)$$

e

$$X^{k+1} = (X^k)^{\frac{1}{2}}\bar{Z}(X^k)^{\frac{1}{2}}. \quad (3.10)$$

O espaço de busca da nova variável \bar{Z} é o espaço transformado de Y por T_{X^k} . Como (3.9) e (3.10) representam transformações genéricas de similaridade (em particular, os autovalores permanecem inalterados), podemos escolhê-las de modo que \bar{Z} seja uma matriz diagonal.

Seja então Q_j uma matriz ortogonal que diagonaliza $\bar{Z} = (X^k)^{-\frac{1}{2}}Y(X^k)^{-\frac{1}{2}}$, ou seja, $(Q_j)^T(X^k)^{-\frac{1}{2}}Y(X^k)^{-\frac{1}{2}}Q_j = Z$ (Z é diagonal com elementos diagonais estritamente positivos). É fácil verificar que o operador $T_{(X^k)(Q_j)} : S_{++}^n \rightarrow S_{++}^n$ dada por

$$T_{(X^k)(Q_j)}(Y) = Q_j^T(X^k)^{-\frac{1}{2}}Y(X^k)^{-\frac{1}{2}}Q_j$$

também é uma isometria em relação a métrica (2.6) (argumentação análoga a empregada em T_{X^k}).

Restringindo o domínio de $T_{(X^k)(Q_j)}$ ao conjunto de matrizes diagonais definidas positivas, o problema local passa a ser definido em R_{++}^n . Contudo, este novo problema deixa de ser equivalente ao problema original.

³ $T_{X^k}^{-1}$ também é uma isometria.

Nossa meta então é substituir (3.9) por uma sucessão de problemas em

R_{++}^n , dados por

$$Z_{j+1} = \operatorname{argmin}_{Z \succ 0} \left\{ \phi(Z) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(Z) \right\}, \quad (3.11)$$

onde $\phi(Z) = f \left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} \right)$ e

$$Y_{j+1} = (X^k)^{\frac{1}{2}} Q_j Z_{j+1} Q_j^T (X^k)^{\frac{1}{2}}.$$

Algoritmo 3.2 (Algoritmo Proximal Modificado) Dados $X^0 \succ 0$,

$\beta^0, \delta_1, \delta_2 > 0$:

$k \leftarrow 0$;

repetir

. $Y_0 \leftarrow I$;

. $j \leftarrow 0$;

. *repetir*

. *obter* Q_j *ortogonal, tal que*

. $Q_j^T (X^k)^{-\frac{1}{2}} Y_j (X^k)^{-\frac{1}{2}} Q_j = \bar{Z}_j$, *onde* $\bar{Z}_j \succ 0$ *é diagonal (Schur)*;

. *definir* \bar{Z}_j *como ponto inicial e calcular*

$$\cdot Z_{j+1} = \operatorname{argmin}_{Z \succ 0} \left\{ f \left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} \right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(Z) \right\}; \quad (3.12)$$

. *se* $Z_{j+1} = \bar{Z}_j$, *então*

. $X^{k+1} = Y_j$;

. *break*;

. *fim-se*;

. $Y_{j+1} = (X^k)^{\frac{1}{2}} Q_j Z_{j+1} Q_j^T (X^k)^{\frac{1}{2}}$;

. $j \leftarrow j + 1$

. *enquanto* $d(Y_{j+1}, Y_j) > \delta_2$

. $X^{k+1} = Y_{j+1}$;

. $\beta^k \leftarrow \beta^{k+1}$;

. $k \leftarrow k + 1$;

enquanto $d(X^{k+1}, X^k) > \delta_1$.

Lema 3.2 *Os iterados* Z_j *e* \bar{Z}_j *do Algoritmo (3.2) são similares.*

Demonstração: Com efeito,

$$\begin{aligned}
\bar{Z}_j &= \\
&= Q_j^T (X^k)^{-\frac{1}{2}} Y_j (X^k)^{-\frac{1}{2}} Q_j = Q_j^T (X^k)^{-\frac{1}{2}} (X^k)^{\frac{1}{2}} Q_{j-1} Z_j Q_{j-1}^T (X^k)^{\frac{1}{2}} (X^k)^{-\frac{1}{2}} Q_j \\
&= Q_j^T Q_{j-1} Z_j Q_{j-1}^T Q_j. \quad \blacksquare
\end{aligned}$$

Como consequência imediata do Lema (3.2), temos que

$$\begin{aligned}
&f\left((X^k)^{\frac{1}{2}} Q_{j-1} Z_j Q_{j-1}^T (X^k)^{\frac{1}{2}}\right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(Z_j) = \\
&= f(Y_j) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(Z_j) = \\
&= f\left((X^k)^{\frac{1}{2}} Q_j \bar{Z}_j Q_j^T (X^k)^{\frac{1}{2}}\right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(\bar{Z}_j).
\end{aligned}$$

Proposição 3.2 *Se $Z_{j+1} \neq \bar{Z}_j$, então $g_{\beta^k}(Y_{j+1}) < g_{\beta^k}(Y_j)$, isto é, a seqüência (Y_j) gerada pelo Algoritmo (3.2) é monótona decrescente.*

Demonstração: Os passos da demonstração levam em consideração a definição da seqüência $(Y_j)_{j \in \mathbb{N}}$ e a propriedade de similaridade entre matrizes. Considere a seqüência $(g_{\beta^k}(Y_j))_{j \in \mathbb{N}}$, temos que:

$$\begin{aligned}
g_{\beta^k}(Y_{j+1}) &= \\
&= f(Y_{j+1}) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i\left((X^k)^{-\frac{1}{2}} Y_{j+1} (X^k)^{-\frac{1}{2}}\right) \\
&= f\left((X^k)^{\frac{1}{2}} Q_j Z_{j+1} Q_j^T (X^k)^{\frac{1}{2}}\right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i\left((X^k)^{-\frac{1}{2}} (X^k)^{\frac{1}{2}} Q_j Z_{j+1} Q_j^T (X^k)^{\frac{1}{2}} (X^k)^{-\frac{1}{2}}\right) \\
&= f\left((X^k)^{\frac{1}{2}} Q_j Z_{j+1} Q_j^T (X^k)^{\frac{1}{2}}\right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(Q_j Z_{j+1} Q_j^T) \\
&= f\left((X^k)^{\frac{1}{2}} Q_j Z_{j+1} Q_j^T (X^k)^{\frac{1}{2}}\right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(Z_{j+1}) \\
&< f\left((X^k)^{\frac{1}{2}} Q_j \bar{Z}_j Q_j^T (X^k)^{\frac{1}{2}}\right) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i(\bar{Z}_j) \quad (\text{por construção}) \\
&= f(Y_j) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i\left(Q_j^T (X^k)^{-\frac{1}{2}} Y_j (X^k)^{-\frac{1}{2}} Q_j\right) \\
&= f(Y_j) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 \lambda_i\left((X^k)^{-\frac{1}{2}} Y_j (X^k)^{-\frac{1}{2}}\right) \\
&= g_{\beta^k}(Y_j) \quad \blacksquare
\end{aligned}$$

Para obter a convergência do laço interno do Algoritmo (3.2) é necessário mostrar que se $Z_{j+1} = \bar{Z}_j$ então Y_j é solução de 3.6, isto é, $Y_j = X^{k+1}$. Este constitui um dos pontos em aberto do nosso trabalho. A convergência do laço externo do algoritmo é assegurada através dos resultados apresentados em [13].

3.4 Um problema bem posto

A iteração (3.12) do Algoritmo (3.2) é um problema de minimização no espaço de matrizes diagonais definidas positivas que é equivalente a um problema de programação não-linear no ortante estritamente positivo de R^n , e, claramente, este possui complexidade significativamente menor do que o original definido em S_{++}^n .

Defina $\varphi_j : R_{++}^n \rightarrow R$ por

$$\varphi_j(z) = f \left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} \right), \quad (3.13)$$

onde Z é a matriz diagonal cuja diagonal é formada pelas componentes de z , isto é, $Z = \text{diag}(z)$.

A regularização $\sum_{i=1}^n \ln^2 \lambda_i(Z)$ pode ser vista também como uma função de R_{++}^n em R , pois os autovalores de uma matriz diagonal Z são os próprios elementos da diagonal. É importante ressaltar que $\sum_{i=1}^n \ln^2 z_i$ é uma barreira para R_{++}^n .

Sem qualquer abuso de notação, (3.12) pode ser reescrito como

$$z_{j+1} = \operatorname{argmin}_{z>0} \left\{ \varphi_j(z) + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 z_i \right\}. \quad (3.14)$$

Como f é diferenciável, para um dado vetor $v \in R^n$ e t suficientemente

pequeno, a derivada direcional de φ_j na direção v é:

$$\begin{aligned} \frac{d\varphi_j}{dv}(z) &= \\ \frac{d}{dt} \Big|_{t=0} \varphi_j(z + tv) &= \frac{d}{dt} \Big|_{t=0} f \left((X^k)^{\frac{1}{2}} Q_j [Z + tV] Q_j^T (X^k)^{\frac{1}{2}} \right) \\ &= Df \left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} \right) \bullet \left((X^k)^{\frac{1}{2}} Q_j V Q_j^T (X^k)^{\frac{1}{2}} \right), \end{aligned}$$

onde $V = \text{diag}(v)$. Portanto,

$$\nabla \varphi_j(z)_i = Df \left((X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} \right) \bullet \left((X^k)^{\frac{1}{2}} Q^k E_i (Q^k)^T (X^k)^{\frac{1}{2}} \right),$$

onde $E_i = \text{diag}(e_i)$ (i -ésimo vetor da base canônica de R^n).

Apesar de não ser estritamente convexa em R_{++}^n , $\sum_{i=1}^n \ln^2 z_i$ é coerciva. Note que $\|z\| \rightarrow \infty$ implica que $z_k \rightarrow \infty$, para algum $1 \leq k \leq n$, que é suficiente para que $\sum_{i=1}^n \ln^2 z_i \rightarrow \infty$.

Definição 3.3 *Seja $S \subseteq R^n$ convexo e fechado e $f : S \rightarrow R$. f é dita fortemente quasiconvexa se para cada $x_1, x_2 \in S$, com $x_1 \neq x_2$, $f(tx_1 + (1-t)x_2) < \max \{f(x_1), f(x_2)\}$, para todo $t \in (0, 1)$.*

O próximo resultado é uma extensão do teorema de unicidade de solução para funções estritamente convexas.

Teorema 3.1 *Seja $f : S \subseteq R^n \rightarrow R$ fortemente quasiconvexa. Considere o problema*

$$\{\min f(x); x \in S\}. \quad (3.15)$$

Se \bar{x} é uma solução local de (3.15) então \bar{x} é a única solução global.

Demonstração: ver [24] ■

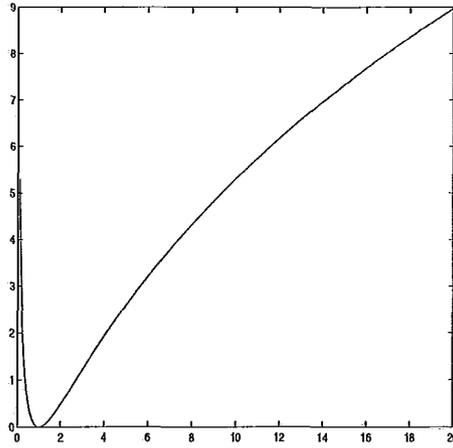


Figura 3.1: Gráfico da função $\ln^2 t$, para $t \in (0, 20]$.

Indicamos na Figura 3.1 uma parcela do gráfico de $\sum_{i=1}^n \ln^2 z_i$. A figura sugere que tal parcela satisfaz a Definição (3.3).

Lema 3.3 A função $f : R_{++} \rightarrow R_+$, definida por

$$f(t) = \ln^2 t,$$

é fortemente quasiconvexa em R_{++} .

Demonstração: De fato, dados $t_1, t_2 \in R_{++}$ com $t_1 \neq t_2$, defina $h(\alpha) = f(\alpha t_1 + (1 - \alpha)t_2)$, com $\alpha \in [0, 1]$. h é uma função contínua, diferenciável no aberto $(0, 1)$ e possui máximo e mínimo em $[0, 1]$ ($[0, 1]$ é compacto). Um extremo interior de h deve satisfazer $h'(\bar{\alpha}) = 0$. Por outro lado,

$$h'(\alpha) = \frac{2\ln(\alpha t_1 + (1-\alpha)t_2)}{\alpha t_1 + (1-\alpha)t_2} (t_1 - t_2),$$

donde vem que

$$h'(\bar{\alpha}) = 0 \iff \frac{2\ln(\bar{\alpha}t_1 + (1-\bar{\alpha})t_2)}{\bar{\alpha}t_1 + (1-\bar{\alpha})t_2} (t_1 - t_2) = 0 \iff \bar{\alpha}t_1 + (1 - \bar{\alpha})t_2 = 1.$$

resolvendo a equação em $\bar{\alpha}$ temos,

$$\bar{\alpha} = \frac{1-t_2}{t_1-t_2} \quad \text{e ainda } h(\bar{\alpha}) = 0.$$

Logo,

$h(\bar{\alpha}) < \max\{h(0), h(1)\} = \max\{\ln^2 t_1, \ln^2 t_2\}$. Lembremos que $t_1 \neq t_2$ implica que $\ln^2 t_1$ e $\ln^2 t_2$ não podem ser simultâneamente nulos.

Se $\bar{\alpha}$ for ponto de mínimo ou inflexão então o máximo de h deverá ser atingido em um dos extremos do intervalo $[0, 1]$. Portanto,

$$(i) \quad \alpha = 0 \implies h(0) = \ln^2 t_2 \leq \max\{\ln^2 t_1, \ln^2 t_2\}.$$

$$(ii) \quad \alpha = 1 \implies h(1) = \ln^2 t_1 \leq \max\{\ln^2 t_1, \ln^2 t_2\}.$$

■

O fato da desigualdade não ser estrita nos itens (i) e (ii) da demonstração não contradiz a Definição 3.3, uma vez que nestes casos $\alpha \notin (0, 1)$.

A barreira $\sum_{i=1}^n \ln^2 z_i$ é uma soma de funções fortemente quasiconvexas e consequentemente é fortemente quasiconvexa.

Quando uma função fortemente quasiconvexa é adicionada a uma convexa preserva-se sua característica inicial. Esta propriedade é fácil de ser verificada. De fato, dadas funções f_1 convexa e f_2 fortemente quasiconvexa, respectivamente em $S \subseteq R^n$ e $x_1, x_2 \in S (x_1 \neq x_2)$, $t \in (0, 1)$,

$$(f_1 + f_2)(tx_1 + (1 - t)x_2) = f_1(tx_1 + (1 - t)x_2) + f_2(tx_1 + (1 - t)x_2) < tf_1(x_1) + (1 - t)f_1(x_2) + \max\{f_2(x_1), f_2(x_2)\} < t \max\{f_1(x_1), f_1(x_2)\} + (1 - t) \max\{f_1(x_1), f_1(x_2)\} + \max\{f_2(x_1), f_2(x_2)\} = \max\{(f_1 + f_2)(x_1), (f_1 + f_2)(x_2)\}.$$

Teorema 3.2 *A função objetivo do subproblema (3.14) possui um único minimizador.*

Demonstração: O resultado é consequência imediata da coercividade de $\sum_{i=1}^n \ln^2 z_i$, do Lema 3.3 e da definição de φ em (3.13). ■

3.5 Aplicando o método de Newton

Suponha φ_j fortemente convexa e de classe C^2 . Baseado nessa hipótese é possível estabelecer condições necessárias e suficientes à aplicação do método de Newton, isto é, estabelecer condições para que a hessiana da função objetivo em (3.14) seja definida positiva.

Seja $\alpha > 0$, tal que $\alpha < \mu$, onde $\mu = \lambda_{\min}(\varphi_j''(z))$, $z \in R_{++}^n$. temos

$$\lambda_{\min} \left(\varphi_j''(z) + \beta \operatorname{diag} \left(\frac{1 - \ln z_i}{z_i^2} \right) \right) \geq \mu + \beta \min_{1 \leq i \leq n} \left(\frac{1 - \ln z_i}{z_i^2} \right) \geq \alpha. \quad (3.16)$$

A desigualdade é consequência do Lema (2.3). Isto implica que

$$\beta \min_{1 \leq i \leq n} \left(\frac{1 - \ln z_i}{z_i^2} \right) \geq \alpha - \mu. \quad (3.17)$$

Como $\alpha - \mu < 0$, a parte interessante dessa análise restringe-se ao caso em que $\min_{1 \leq i \leq n} (1 - \ln z_i) < 0$, isto é quando $z_i > e$, para algum $1 \leq i \leq n$. Neste caso, vem de (3.17) que

$$\beta \leq \min_{z_i > e, 1 \leq i \leq n} \left(\frac{(\mu - \alpha)z_i^2}{\ln z_i - 1} \right).$$

Analisando a função $h(t) = \frac{t^2}{\ln t - 1}$ para $t > e$, conclui-se que $\bar{t} = e^{\frac{3}{2}} = \operatorname{argmin}_{t > e} h(t)$ e que o valor ótimo associado a \bar{t} é dado por $h(\bar{t}) = 2e^3$.

Logo, a escolha

$$\beta \leq (\mu - \alpha)2e^3,$$

em particular,

$$\beta \leq \mu e^3 \quad (3.18)$$

satisfaz (3.16). Observe que esta condição é coerente com a hipótese de $\beta^k \rightarrow 0$.

Vale observar que no conjunto $(0, e)^n$ a função barreira $\sum_{i=1}^n \ln^2 z_i$ é estritamente convexa. Isto é importante para problemas cujos autovalores da variável X são limitados, como o de Programação semidefinida com traço constante (por exemplo, acrescente à (2.1) a restrição $\text{Tr}X = 1$), pois após as transformações propostas pelo Algoritmo 3.2 obtemos um problema limitado em R_{++}^n (como $\text{Tr}X = 1$, devemos ter $\lambda_i(X) \leq 1, \forall i = 1, \dots, n$).

3.6 Aplicação à (PSD)

Retomemos a formulação primal (2.1) apresentada no início do capítulo:

(PSD)

$$\{\min C \bullet X; \text{ s.a. } A_i \bullet X = b_i, i = 1, \dots, m, X \succeq 0\}.$$

Com base na proposta de ponto proximal, a iteração (3.12) consiste em determinar o único vetor z^{k+1} solução do problema

(PP)_{PSD}

$$\begin{aligned} \min C \bullet (X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 z_i \\ \text{s.a } A_i \bullet (X^k)^{\frac{1}{2}} Q_j Z Q_j^T (X^k)^{\frac{1}{2}} = b_i, i = 1, \dots, m \\ z > 0, \end{aligned}$$

onde $Z = \text{diag}(z)$.

Como a função custo e as restrições são lineares, os termos podem ser agregados de forma que o problema $(PP)_{PSD}$ possa ser reescrito no seguinte formato

$(PP)_{PSD}$

$$\begin{aligned} \min (c^j)^T z + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 z_i \\ \text{s.a } A^j z = b \\ z > 0, \end{aligned}$$

para algum vetor $c^j \in R^n$ e alguma matriz $A_{m \times n}^j$.

As hipóteses necessárias a aplicação do método de Newton podem, como veremos, ser garantidas através do acréscimo de uma regularização $-\mu^k \sum_{i=1}^n \ln z_i$. Consideremos, em consequência, o seguinte problema regularizado:

$(PPR)_{PSD}$

$$\begin{aligned} \min (c^j)^T z - \mu^k \sum_{i=1}^n \ln z_i + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 z_i \\ \text{s.a } A^j z = b \\ z > 0. \end{aligned}$$

A função de Lagrange associada ao $(PPR)_{PSD}$ é dada por

$$L(z, y) = (c^j)^T z - \mu^k \sum_{i=1}^n \ln z_i + \frac{\beta^k}{2} \sum_{i=1}^n \ln^2 z_i - y^T (A^j z - b).$$

As condições de *KKT* para o $(PPR)_{PSD}$ são:

$$\begin{aligned} c^j - \mu^j z^{-1} + \beta^k [\text{diag}(z)]^{-1} \ln z - (A^j)^T y &= 0 \\ b - A^j z &= 0 \\ z &> 0. \end{aligned}$$

Como o objetivo é determinar a direção de Newton, seja $\bar{z} > 0$ um ponto viável para o $(PPR)_{PSD}$ e $\bar{y} \in R^n$. O passo de Newton satisfaz

$$\left[\mu^k \bar{Z}^{-2} + \beta^k \text{diag} \left(\frac{\bar{e} - \ln \bar{z}}{\bar{z}_i^2} \right) \right] \Delta z - (A^j)^T \Delta y = -c^j + \mu^k \bar{z}^{-1} - \beta^k [\text{diag}(\bar{z})]^{-1} \ln \bar{z} + (A^j)^T \bar{y} \quad (3.19)$$

$$-A^j \Delta z = 0, \quad (3.20)$$

onde \bar{e} é o vetor de R^n cujas coordenadas são todas iguais a 1. As equações (3.19) e (3.20) resultam da definição do passo de Newton para sistemas não-lineares ⁴.

Denote por $U = \text{diag} \left(\frac{\mu^k + \beta^k (1 - \ln \bar{z}_i)}{\bar{z}_i^2} \right)$ o coeficiente de Δz . Através da expressão da inversa da matriz dos coeficientes do sistema dado pelas equações (3.19) e (3.20), vê-se, claramente, a necessidade da regularização:

$$\begin{bmatrix} U^{-1} (I + (A^j)^T P^{-1} A^j U^{-1}) & -U^{-1} (A^j)^T P^{-1} \\ -P^{-1} A^j U^{-1} & P^{-1} \end{bmatrix},$$

onde $P = -A^j U (A^j)^T$.

Lema 3.4 *Suponha que a matriz A^j do $(PPR)_{PSD}$ tenha posto completo.*

Então, para todo $\mu^k > 0$, satisfazendo

$$\frac{\mu^k}{\beta^k} > \min_{\bar{z}_i > e, 1 \leq i \leq n} (\ln \bar{z}_i - 1).$$

, o algoritmo de Newton pode ser aplicado ao $(PPR)_{PSD}$.

Demonstração: Aplicando a mesma análise empregada na seção anterior, vem que

⁴Lembrando que para um sistema de equações não-lineares da forma $F(x) = 0$, o passo de Newton é aquele que satisfaz $F'(x) \Delta x = -F(x)$.

$$\mu^k + \min_{1 \leq i \leq n} \beta^k (1 - \ln \bar{z}_i) > 0.$$

Se $\min_{1 \leq i \leq n} (1 - \ln \bar{z}_i) \geq 0$ então para qualquer $\mu > 0$ a aplicação do método de Newton está assegurada. Caso $\min_{1 \leq i \leq n} (1 - \ln \bar{z}_i) < 0$, isto é, quando $\bar{z}_i > e$, para algum $1 \leq i \leq n$ devemos ter

$$\frac{\mu^k}{\beta^k} > \min_{z_i > e, 1 \leq i \leq n} (\ln \bar{z}_i - 1) > 0.$$

■

Capítulo 4

Conclusões, Trabalhos Futuros e Extensões

Neste trabalho estamos propondo um novo algoritmo implementável de ponto proximal para programação semidefinida baseado na transformação do problema original em uma sucessão de problemas de programação não-linear em R_{++}^n .

Podemos observar que o procedimento de diagonalização aplicado no terceiro passo do Algoritmo (3.2) não aumenta demasiadamente os custos computacionais do método, visto que existem rotinas implementadas em bibliotecas de diferentes linguagens, como a decomposição de Schur para matrizes simétricas, que determinam Q^k de forma rápida e eficiente.

Como a pesquisa foi motivada pela convergência global do algoritmo de ponto proximal original estabelecido em [13], acreditamos que a mesma possa ser obtido para o algoritmo proposto, apesar não termos asseguradas as hipóteses iniciais tal qual a convexidade da função objetivo do $(PPR)_{PSD}$ relativa a métrica riemanniana (2.6).

Vimos que o Algoritmo proximal modificado requer a inicialização de um

ponto Y_0^k viável para o (PSD) a cada iteração. Verificada esta exigência, temos garantida, pela Proposição (2.1) e pelo passo de Newton, a viabilidade do iterado X^{k+1} . Inicialmente, não existe problemas em se tomar o mesmo ponto Y_0^k para cada iteração.

Contudo, na prática, existe uma grande dificuldade em se determinar um ponto inicial viável. por isto, devemos estar combinando o algoritmo com formulações alternativas do (PSD), como o big-M em [26] ou a formulação apresentada em [19], que permitam superar este obstáculo.

Como trabalhos futuros, estamos propondo uma implementação do algoritmo para problemas de grande porte em programação semidefinida que possibilite a realização de testes numéricos.

Por fim, o algoritmo proximal modificado pode ser naturalmente estendido para o caso não-suave através da noção de feixes híbrido.

Esta metodologia consiste em aplicar o método de feixes clássico somente à função objetivo do problema original na iteração principal do algoritmo de ponto proximal. A regularização é mantida.

Isso é feito da seguinte forma: inicialmente, escolhe-se um modelo funcional que aproxima inferiormente a função objetivo do problema original. Logo após, minimiza-se este modelo somado a regularização da metodologia proximal (ao contrário da função objetivo do problema original somada a regularização como descreve a teoria). Se o valor da função objetivo do problema original for igual ao valor do modelo após o processo de minimização, então o ponto obtido é a solução da iteração corrente. Caso contrário, o modelo é substituído por uma aproximação de 1ª ordem, dada por um de

seus subgradientes neste ponto, e minimiza-se esta aproximação somada a regularização da metodologia proximal. Os passos são repetidos até que se obtenha a solução da iteração corrente. Maiores detalhes a respeito desta metodologia, bem como a análise de convergência podem ser encontradas em (ver [27]).

Apêndice

Lista de Notações

Neste apêndice destacamos uma lista com algumas notações usuais de Programação Semidefinida, bem como outras específicas deste trabalho, com o objetivo de facilitar a compreensão e a leitura. As demais notações são apresentadas no decorrer do texto.

- R^n - espaço euclidiano de dimensão n .
- R_+^n - ortante não-negativo de R^n .
- R_{++}^n - ortante positivo de R^n .
- S^n - conjunto das matrizes simétricas de ordem n .
- S_+^n - conjunto das matrizes simétricas semidefinidas positivas.
- S_{++}^n - conjunto das matrizes simétricas definidas positivas.
- $Z = \text{diag}(z)$ - matriz diagonal cujos elementos da diagonal são as componentes do vetor z .
- d - distância riemanniana entre dois pontos.
- $B_\delta(Y)$ - bola aberta de centro Y e raio δ .

- $X \bullet Y$ - produto escalar do espaço de matrizes definido por $X \bullet Y = \text{Tr}(Y^T X)$.
- (PSD) - Problema primal padrão de Programação Semidefinida.
- $(PP)_{PSD}$ - Problema proximal associado ao problema de Programação Semidefinida.
- $(PPR)_{PSD}$ - Problema proximal regularizado.
- $X \succeq 0$ - X é semidefinida positiva.
- $X \succ 0$ - X é definida positiva.

Bibliografia

- [1] Saigal, R., Vandenberghe, L., Wolkowicz, H., *Handbook of Semidefinite Programming*. 1nd ed. Boston-Dordrecht-London, Kluwer Academic Publishers, 2000.
- [2] Nesterov, Y.E., Nemirovskii, A.S., “Polynomial barrier methods in convex programming”, *Ekonomika i Matem. Metody*, v. 24, pp. 1084-1091, 1988.
- [3] Nesterov, Y.E., Nemirovskii, A.S., *Optimization over positive semidefinite matrices: Mathematical background and users manual*. Technical report, Central Economic and Mathematical Institute, USSR Acad. Sci. Moscow, USSR, 1990.
- [4] Nesterov, Y.E., Nemirovskii, A.S., *Interior-Point polynomial algorithms in convex Programming, vol. 13 of Studies in Applied Mathematics*. 1nd ed. Philadelphia, SIAM, 1994.
- [5] Alizadeh, F., “Interior point methods in Semidefinite programming with applications to combinatorial optimization”, *SIAM Journal on Optimization*, v. 5, n. 1, pp. 13-51, 1995.
- [6] Monteiro, R.D.C., “First and second-order methods for semidefinite programming”, *Mathematical Programming, Series B*, n. 97, pp. 209-244, 2003.

- [7] Horner, S., Peinado, M., “Design and performance of parallel and distributed approximation algorithms for max cut”, *Journal of Parallel and Distributed computing*, v. 46, pp. 48-61, 1997.
- [8] Burer, S., Monteiro, R.D.C., “A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization”, *Mathematical Programming*, Series B, n. 95, pp. 329-357, 2003.
- [9] Vanderbei, R.J., Benson, Y., *On formulating semidefinite programming problems as smooth nonlinear optimization problems. Report orfe 99-01*, Princeton, Operations Research and Financial engineering, 1999.
- [10] Burer, S., Monteiro, R.D.C., Zhang, Y., “Solving a class of semidefinite programs via nonlinear programming”, *Mathematical Programming*, v. 93, pp. 97-123, 2002.
- [11] Fukuda, M., Kojima, M., Murota, K., Nakata, K., “Exploiting sparsity in semidefinite programming via matrix completion I: General framework”, *SIAM Journal on Optimization*, v. 11, pp. 647-674, 2001.
- [12] Helmberg, C., Rendl, F., “A spectral bundle method for semidefinite programming”, *SIAM Journal on Optimization*, v. 10, pp. 673-696, 2000.
- [13] Ferreira, O.P., Oliveira, P.R., “Proximal point algorithm on Riemannian manifolds”, *Optimization*, v. 51, n. 2, pp. 257-270, 2002.
- [14] Nesterov, Y.E., Todd, M.J., “On the Riemannian geometry defined by self-concordant barriers and interior-point methods”, *Foundations of Computational Mathematics*, v. 2, pp. 333-361, 2002.
- [15] Rothaus, O.S., “Domains of positivity”, *Abh. Math. Sem. Univ. Hamburg*, v. 24, pp. 189-235, 1960.

- [16] Güler, O., “Barrier functions in interior-point methods”, *Mathematics of Operations Research*, v. 21, pp. 860-885, 1996.
- [17] Lima, E.L., *Curso de análise, vol.1*. 1ª ed. Rio de Janeiro, impa, 1995.
- [18] do Carmo, M.P., *Geometria riemanniana*. 2ª ed. Rio de Janeiro, impa, 1988.
- [19] Helmberg, C., *An Interior Point Method for Semidefinite Programming and Max-Cut Bounds*. Ph.D. dissertation, Technischen Universität Graz, Graz, 1994.
- [20] Horn, R.A., Johnson, C.R., *Matrix Analysis*. 1st ed. Cambridge, Cambridge University Press, 1985.
- [21] Horn, R.A., Johnson, C.R., *Topics in Matrix Analysis*. 1st ed. Cambridge, Cambridge University Press, 1991.
- [22] Mosheye, L., Zibulevsk, M., *Penalty/Barrier Multiplier Algorithm for Semidefinite Programming*. Technical report, 1999.
- [23] Roos, T., Terlaky, T., Vial, J.-P., *Theory and Algorithms for Linear Optimization, An Interior Point Approach*. 1st ed. New York, John Wiley and Sons, 1997.
- [24] Bazaraa, M.S., Sherali, H.D., Shetty, C.M, *Nonlinear Programming. Theory and Algorithms*. 2nd ed. New York, Wiley, 1993.
- [25] Golub, G.H., Van Loan, C.F., *Matrix computations*. 3rd ed. Baltimore and London, The Johns Hopkins University Press, 1996.
- [26] Vandenberghe, L., Boyd, S., “Semidefinite Programming”, *SIAM Journal on Optimization*, v. 38, n. 1, pp. 49-95, 1996.

- [27] Hiriart-Urruty, J-B., Lemaréchal, C., *Convex analysis and minimization algorithms II, advanced theory and bundle methods*. 1nd ed. Heidelberg, springer-verlag, 1993.