

TRAJETÓRIA CENTRAL EM PROGRAMAÇÃO SEMIDEFINIDA, MÉTODO
DO PONTO PROXIMAL GENERALIZADO E TRAJETÓRIA DE CAUCHY
EM VARIEDADE RIEMANNIANA

Roberto Cristóvão Mesquita Silva

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. Orizon Pereira Ferreira, D. Sc.

Prof. Carlos Humes Junior, Ph.D.

Prof. Gregorio Malajovich Munoz, Ph.D.

Prof. Márcia Helena Costa Fampa, D. Sc.

Prof. Paulo José da Silva e Silva, D. Sc.

RIO DE JANEIRO, RJ - BRASIL
OUTUBRO DE 2006

SILVA, ROBERTO CRISTÓVÃO
MESQUITA

Trajetória Central em Programação
Semidefinida, Método do Ponto Proximal
Generalizado e Trajetória de Cauchy em Va-
riedade Riemanniana [Rio de Janeiro] 2006

VI, 51 p. 29,7 cm (COPPE/UFRJ, D.Sc.,
Engenharia de Sistemas e Computação, 2006)

Tese – Universidade Federal do Rio de
Janeiro, COPPE

- 1 - Trajetória Central
- 2 - Método do Ponto Proximal Generalizado
- 3 - Trajetória de Cauchy
- 4 - Programação Semidefinida
- 5 - Variedade Riemanniana

I. COPPE/UFRJ II. Título (série)

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

TRAJETÓRIA CENTRAL EM PROGRAMAÇÃO SEMIDEFINIDA, MÉTODO DO PONTO PROXIMAL GENERALIZADO E TRAJETÓRIA DE CAUCHY EM VARIEDADE RIEMANNIANA

Roberto Cristóvão Mesquita Silva

Outubro/2006

Orientadores: Paulo Roberto Oliveira

Orizon Pereira Ferreira

Programa: Engenharia de Sistemas e Computação

Estudamos a convergência das trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, para o problema de programação semidefinida. Analisamos também a convergência da trajetória central primal associada a uma classe de funções estritamente convexas, incluindo a entropia e a barreira logarítmica. Como uma aplicação obtemos resultados de convergência para o método do ponto proximal generalizado. Em particular, estudamos o método do ponto proximal generalizado com a distância Kullback-Leibler. Finalmente provamos que a trajetória central primal, induzida por uma função estritamente convexa, coincide com a trajetória de Cauchy na variedade Riemanniana munida com a métrica dada pela Hessiana desta função.

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

CENTRAL PATHS IN SEMIDEFINITE PROGRAMMING, GENERALIZED
PROXIMAL POINT METHOD AND CAUCHY TRAJECTORIES IN
RIEMANNIAN MANIFOLDS

Roberto Cristóvão Mesquita Silva

October/2006

Advisors: Paulo Roberto Oliveira
Orizon Pereira Ferreira

Department: Systems and Computation Engineering

We study the convergence of primal and dual central paths associated to entropy and exponential functions, respectively, for semidefinite programming problem. We also analyze the convergence of primal central path associated to a class of strictly convex functions, including the entropy and the logarithmic barrier. As an application we obtain convergence results for the generalized proximal point method. In particular, we study the generalized proximal point method with the Kullback-Leibler distance. Finally we prove that the primal central path, induced by a strictly convex function, coincides with the Cauchy trajectory on the Riemannian manifold endowed with the metric given by the Hessian of this function.

Sumário

1	Introdução	1
2	Preliminares	7
2.1	Introdução	7
2.2	Notações e Terminologia	7
2.3	Derivada em Espaços de Matrizes	8
2.4	Alguns Elementos da Análise Matricial	12
2.5	Elementos da Geometria Riemanniana	15
2.6	Resultados Básicos de Programação Semidefinida	16
3	Trajetoórias Central em Programação Semidefinida	18
3.1	Introdução	18
3.2	Trajetoória Central Primal-Dual	18
3.3	Analiticidade da Trajetória Central Primal-Dual	20
3.4	Convergência da Trajetória Central Primal-Dual	22
3.5	Trajetoória Central Primal	29
3.6	Convergência da Trajetória Central Primal	34
4	Método do ponto proximal generalizado	38
4.1	Introdução	38
4.2	Método do Ponto Proximal Generalizado com a Distância Kullback-Leibler	39
4.3	Método do Ponto Proximal Generalizado Associado a uma Distância Generalizada	51
5	Trajetoória de Cauchy na variedade Riemanniana	60
5.1	Introdução	60
5.2	Trajetoória Central Primal e Trajetória de Cauchy	61

6	Considerações Finais	64
	Referências Bibliográficas	66

Capítulo 1

Introdução

A extensão dos conceitos e das técnicas da programação linear para a programação semidefinida tornou-se atrativa após os primeiros trabalhos devidos a Alizadeh [1] e Nesterov e Nemirovskii [42]. É bem conhecido que o conceito de trajetória central, com respeito à função barreira logarítmica, desempenha papel importante em programação linear e programação semidefinida, veja por exemplo, Güler [22] e Todd [57]. Halická et al. [24] mostraram que a trajetória central para problemas de programação semidefinida converge. Em um contexto mais geral, Graña Drummond e Peterzil [21] estabeleceram sua convergência para uma certa classe de problemas de programação semidefinida convexos e analíticos. No que concerne à trajetória central na programação linear, tem-se a convergência para o centro analítico do conjunto solução, o que não ocorre na programação semidefinida, veja Halická et al. [24]. Caracterizações parciais do ponto limite foram dadas por Luo et al. [37], Sporre e Forsgren [55] e Halická et al. [23]. Ainda em relação à programação linear, barreiras diferentes da função barreira logarítmica foram bastante estudadas, veja por exemplo, Cominetti e San Martin [7], Iusem et al. [29], Iusem e Monteiro [28] e da Cruz Neto et al. [10]. Entretanto o estudo da trajetória central para problemas de programação semidefinida, associada a barreiras diferentes da função barreira logarítmica, é pouco estudado. Isto foi uma das motivações para o desenvolvimento desta tese.

O primeiro resultado obtido foi prova da convergência das trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, para o problema de programação semidefinida (PSD). Aqui destacamos nossa primeira contribuição. Para sermos mais precisos, consideremos \mathbb{R}^n o espaço Euclidiano n -dimensional, S^n o conjunto de todas as matrizes simétricas $n \times n$, S_+^n (resp.,

S_{++}^n) o cone das matrizes simétricas $n \times n$ semidefinidas (resp., definidas) positivas. Denotemos $X \succeq 0$ (resp., $X \succ 0$) para matrizes $X \in S_+^n$ (resp., $X \in S_{++}^n$), tr representa o traço de uma matriz $n \times n$ e façamos $X \bullet Y = \text{tr} XY$ para matrizes $X, Y \in S^n$. O problema PSD primal se escreve

$$(P) \quad \min \{C \bullet X : \mathcal{A}X = b, X \succeq 0\}, \quad (1.1)$$

onde $C \in S^n$, $b \in \mathbb{R}^m$ e a aplicação linear $\mathcal{A} : S^n \rightarrow \mathbb{R}^m$ são dados e a variável primal é $X \in S^n$. Adicionando a função penalidade entropia à função objetivo de (P), obtemos sua versão penalizada

$$(P_\mu) \quad \min \{C \bullet X + \mu X \bullet \ln(X) : \mathcal{A}X = b, X \succ 0\}, \quad \mu > 0. \quad (1.2)$$

Para maiores detalhes sobre $\ln(X)$ veja seção 2.4. O problema dual associado a (P) é

$$(D) \quad \max \{b^T y : \mathcal{A}^* y + S = C, S \succeq 0\},$$

onde $\mathcal{A}^* : \mathbb{R}^m \rightarrow S^n$ denota a aplicação adjunta de \mathcal{A} e $(y, S) \in \mathbb{R}^m \times S^n$ são as variáveis duais. Adicionando a função penalidade exponencial à função objetivo de (D), obtemos sua versão penalizada

$$(D_\mu) \quad \max \{b^T y - \mu \text{tr} e^{-S/\mu - I} : \mathcal{A}^* y + S = C\}, \quad \mu > 0. \quad (1.3)$$

O conjunto das viáveis primal do problema (P) é denotado por

$$\mathcal{F}(P) = \{X \in S^n : \mathcal{A}X = b, X \succeq 0\},$$

e o conjunto das variáveis dual associado a (P) denotamos por

$$\mathcal{F}(D) = \{(y, S) \in \mathbb{R}^m \times S^n : \mathcal{A}^* y + S = C, S \succeq 0\}.$$

Os interiores dos conjuntos viáveis primal e dual são denotados por $\mathcal{F}^0(P)$ e $\mathcal{F}^0(D)$, respectivamente. Denotamos também por $\mathcal{F}^*(P)$ e $\mathcal{F}^*(D)$, os conjuntos de soluções ótimas primal e dual, respectivamente. Suponhamos ainda que as duas seguintes condições acontecem:

- A aplicação linear $\mathcal{A} : S^n \rightarrow \mathbb{R}^m$ é sobrejetiva;
- $\mathcal{F}^0(P) \neq \emptyset$ e $\mathcal{F}^0(D) \neq \emptyset$.

Um dos nossos primeiros objetivos é provar que a segunda hipótese implica que os problemas (P_μ) e (D_μ) têm única solução $X(\mu)$ e $(S(\mu), y(\mu))$, respectivamente, ou, equivalentemente, que as trajetórias central primal e dual denotadas por

$$\{X(\mu) : \mu > 0\}, \quad \{S(\mu) : \mu > 0\},$$

respectivamente, estão bem definidas. Como consequência $X(\mu)$ e $S(\mu)$ satisfazem a igualdade

$$S(\mu) = -\mu \ln(X(\mu)) - \mu I, \quad \mu > 0.$$

Também provaremos que as trajetórias central primal e dual convergem para uma solução de (P) e (D) , respectivamente, quando μ vai para 0. Desta forma, podemos pensar em (P_μ) e (D_μ) como métodos de penalidade entropia e exponencial, respectivamente, para resolver problemas PSD. Vale mencionar que Cominetti e San Martín [7] investigaram o comportamento assintótico das trajetórias central primal e dual associadas às funções penalidades entropia e exponencial, respectivamente, para problemas de programação linear. Em particular, eles obtiveram uma caracterização de seus pontos limites. Por outro lado, Iusem e Monteiro [28] forneceram uma caracterização do ponto limite da trajetória central dual associada a uma ampla classe de funções penalidades, incluindo a exponencial, para problemas de programação convexa com restrições lineares.

Outro objetivo é estudar a trajetória central primal associada a uma classe de funções estritamente convexas e de classe C^2 que inclui a entropia e a barreira logarítmica. Recentemente, Auslender e Héctor Ramírez [3] ao estudarem a trajetória central associada à problemas PSD convexas com restrições convexas mais gerais, provaram a limitação daquela trajetória e que todo ponto de acumulação dela é solução. Observamos que a classe de problemas PSD abordada por aqueles autores inclui o problema que tratamos aqui. Mas, aqui neste trabalho conseguimos obter a convergência da trajetória central primal e dual, além da caracterização do ponto limite da trajetória central primal, propriedades relevantes que não são obtidas em trabalhos anteriores. Mais precisamente, provaremos que a trajetória central primal associada a esta classe é bem definida e converge. Além disso, desde que as funções desta classe se estendam continuamente a S_+^n provaremos que a trajetória central primal converge para o centro analítico do conjunto solução do problema (P) .

Muitas generalizações do método do ponto proximal clássico estudado em Rockafellar [52] foram consideradas, devido ao importante papel que ele tem no desen-

volvimento de algoritmos do Lagrangeano aumentado, por exemplo, veja Censor e Zenios [8], Chen e Teboulle [9], Eckstein [18] e Iusem [30]. Em particular, Doljansky e Teboule [17] introduziram um método do ponto proximal generalizado para problemas PSD convexos e estabeleceram suas propriedades de convergência. Adicionalmente, eles estudaram o método do Lagrangeano aumentado dual correspondente. Outros trabalhos referentes aos métodos do ponto proximal e Lagrangeano aumentado em PSD incluem Auslender e Teboulle [4] e Mosheyev e Zibulebsky [40]. Isto nos motivou a estudar o método do ponto proximal generalizado para resolver problemas PSD com restrições lineares, pois a convergência completa deste método ainda não está estabelecida.

Um outro objetivo é utilizar os resultados obtidos, com respeito às trajetórias central primal e dual, para obter resultados de convergência para o método do ponto proximal generalizado aplicado a problemas PSD. O método do ponto proximal generalizado descrito neste trabalho está incluído no trabalho de Doljansky e Teboule [17]. Mas ressaltamos que aqui obtemos a convergência da seqüência de ponto proximal primal e dual e a caracterização do limite da seqüência primal, propriedades ainda não obtidas em outros trabalhos. Particularmente, analisaremos o método do ponto proximal generalizado com a distância Kullback-Leibler. Este método gera uma seqüência $\{X_k\} \subset S_{++}^n$ com ponto inicial $X_0 \in \mathcal{F}^0(P)$ de acordo com a iteração

$$X_{k+1} = \arg \min_{X \in S_{++}^n} \{C \bullet X + \lambda_k D(X, X_k) : \mathcal{A}X = b\}, \quad (1.4)$$

onde a seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfaz $\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty$ e $D : S_{++}^n \times S_{++}^n \rightarrow \mathbb{R}$ é a distância Kullback-Leibler definida por

$$D(X, Y) = X \bullet \ln(X) - X \bullet \ln(Y) + \text{tr } Y - \text{tr } X.$$

Demonstraremos que a seqüência $\{X_k\}$ está contida na trajetória central primal. Como conseqüência, ambas convergem para uma mesma solução ótima, a saber, o centro analítico do conjunto ótimo relativo à distância Kullback-Leibler. Esta idéia foi primeiramente utilizada por Iusem et al. [29]. Estes autores provaram aquela relação entre a trajetória central e a seqüência de ponto proximal generalizada para alguns casos especiais, incluindo a programação linear, mas não incluindo problemas PSD. Por outro lado, sabe-se que a trajetória central dual não contém a seqüência de ponto proximal dual, mas sim uma certa seqüência construída a partir dela, veja Iusem e Monteiro [28]. Mostraremos que o mesmo acontece no contexto da programação semidefinida, isto é, a trajetória central dual contém uma certa seqüência

construída a partir da seqüência de ponto proximal dual. Mais especificamente, desde que as condições de otimalidade para (1.4) determinam a seqüência dual $\{S_k\}$ por

$$S_k = \lambda_k(\ln(X_k) - \ln(X_{k+1})), \quad k = 0, 1, 2, \dots,$$

podemos, a partir desta, definir a seqüência dual ponderada $\{\bar{S}_k\}$ como

$$\bar{S}_k = \sum_{j=0}^k \lambda_j^{-1} \mu_k S_j, \quad \mu_k = \left(\sum_{j=0}^k \lambda_j^{-1} \right)^{-1}, \quad k = 0, 1, 2, \dots.$$

e provar que ela está contida na trajetória central dual. Como a trajetória central dual converge para uma solução do problema (D), deduz-se que $\{\bar{S}_k\}$ também convergirá. Resultados parciais a respeito do comportamento da seqüência dual ponderada em programação linear foram obtidos em vários artigos incluindo Jensen e Polyak [31], Polyak e Teboulle [45], Powell [46] e Tseng e Bertsekas [56]. A convergência completa da seqüência dual ponderada, para distâncias de Bregman incluindo a distância Kullback-Leibler, para problemas de programação convexa com restrições lineares foi obtida por Iusem e Monteiro [28].

Nosso próximo objetivo será estudar o método do ponto proximal generalizado associado a uma distância generalizada $D_\varphi : S_{++}^n \times S_{++}^n \rightarrow \mathbb{R}$, definida como

$$D_\varphi(X, Y) = \varphi(X) - \varphi(Y) - \nabla\varphi(Y) \bullet (X - Y).$$

onde $\varphi : S_{++}^n \rightarrow \mathbb{R}$ é uma função estritamente convexa e de classe C^2 . Neste caso também obteremos os mesmos resultados obtidos para a distância Kullback-Leibler, relativos à seqüência de ponto proximal primal.

Finalmente, analisaremos a trajetória de Cauchy em variedade Riemanniana. Consideraremos funções φ estritamente convexas e de classe C^2 definidas numa variedade Riemanniana M , e seu gradiente $\text{grad } \varphi$. Dessa maneira, definimos a trajetória de Cauchy $Z(t)$ como uma curva contida em M tal que $Z'(t) = -\text{grad } \varphi(Z(t))$. Veremos que a trajetória central primal induzida pela função φ coincide com a trajetória de Cauchy (ou trajetória do gradiente) na variedade Riemanniana munida com a métrica dada pela Hessiana de φ . Este resultado foi obtido por Iusem et al. [29], no contexto da programação linear. Nesta tese generalizamos esta propriedade para a programação semidefinida.

Recentemente a Geometria Riemanniana tem desempenhado papel importante na otimização. Trabalhos que usam recursos daquelas áreas para estudar a trajetória central em programação linear incluem em Nesterov e Todd [43] e Dedieu et

al. [14]. Extensões dos conceitos e das técnicas do espaço Euclidiano para variedades Riemannianas são naturais. Isto foi feito freqüentemente nos últimos anos, com objetivos teóricos e a fim de também obter algoritmos eficazes da otimização naquelas variedades. Trabalhos referentes incluem Karmarkar [32], Smith [54], Udriste [58], Rapcsák e Thang [49], Rapcsák [48], da Cruz Neto et al. [12], Ferreira e Oliveira [20], Ferreira e Oliveira [19], Nesterov e Todd [43] e Nesterov e Nemirovski [41]. Alguns artigos que tratam do comportamento da trajetória de Cauchy em variedade Riemanniana incluem Karmarkar [32], Helmke e Moore [25], Balogh et al. [5] e Alvarez et al. [2].

A organização desta tese é descrita como segue. No capítulo 2 listamos algumas notações básicas e terminologia usadas durante o desenvolvimento da tese. No capítulo 3 apresentamos o estudo sobre a convergência das trajetórias central primal e dual em programação semidefinida. No capítulo 4 descrevemos o método do ponto proximal generalizado associado ao problema PSD. No capítulo 5 apresentamos os elementos da Geometria Riemanniana necessários para relacionarmos a trajetória de Cauchy com a trajetória central primal. Finalmente, no capítulo 6, destacamos as considerações finais e pesquisas futuras.

Capítulo 2

Preliminares

2.1 Introdução

Neste capítulo fixamos notações, revisamos alguns conceitos e resultados teóricos. Incluiremos as demonstrações da maioria dos resultados que desejamos enfatizar, e os demais serão referenciados.

2.2 Notações e Terminologia

As seguintes notações e resultados serão usados ao longo desta tese.

\mathbb{R}^n é o espaço Euclidiano n -dimensional.

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

$\mathbb{R}_{++}^n = \{(x_1, \dots, x_n) \in \mathbb{R}^n; x_i > 0 \forall i = 1, \dots, n\}$ é o ortante positivo.

$\mathbb{R}^{n \times m}$ é o conjunto de todas as matrizes $n \times m$.

I denota a matriz identidade cuja dimensão será dada pelo contexto.

X_{ij} denota o (i, j) -ésimo elemento de uma matriz $X \in \mathbb{R}^{n \times m}$.

X^T denota a transposta de $X \in \mathbb{R}^{n \times m}$.

S^n denota o conjunto de todas as matrizes simétricas, $n \times n$.

S_+^n denota o cone das matrizes simétricas, $n \times n$, semidefinidas positivas.

S_{++}^n denota o cone das matrizes simétricas, $n \times n$, definidas positivas.

∂S_+^n denota a fronteira de S_+^n .

$X \succeq 0$ significa que X é uma matriz simétrica semidefinida positiva, ou seja, $X \in S_+^n$.

$X \succ 0$ significa que X é uma matriz simétrica definida positiva, ou seja, $X \in S_{++}^n$.

$\text{tr } X \equiv \sum_{i=1}^n X_{ii}$ denota o traço de uma matriz $X \in \mathbb{R}^{n \times n}$.

$X \bullet Y \equiv \langle X, Y \rangle \equiv \text{tr } X^T Y = \sum_{i=1, j=1}^{n, m} X_{ij} Y_{ij}$ denota o produto interno entre as matrizes $X, Y \in \mathbb{R}^{n \times m}$.

$\|X\| \equiv (X \bullet X)^{1/2}$ denota a norma de Frobenius da matriz X .

X_{JK} denota a submatriz de X cujas linhas são indexadas pelo conjunto J , e as colunas indexadas pelo conjunto K , onde J e K são dois subconjuntos de $\{1, \dots, n\}$. Para matrizes quadradas X , X_{JJ} é chamada submatriz principal de X , que é denotada simplesmente por X_J .

Dados $A_1, \dots, A_m \in S^n$, definamos a aplicação linear $\mathcal{A} : S^n \rightarrow \mathbb{R}^m$ por

$$\mathcal{A}X = (A_1 \bullet X, \dots, A_m \bullet X)^T. \quad (2.1)$$

Observe que a aplicação adjunta de \mathcal{A} é a única aplicação linear $\mathcal{A}^* : \mathbb{R}^m \rightarrow S^n$ dada por $\mathcal{A}^*v = \sum_{i=1}^m v_i A_i$ satisfazendo $\langle \mathcal{A}X, y \rangle = X \bullet \mathcal{A}^*y$. Os espaços imagem e nulo do operador linear \mathcal{A} são denotados por $\text{Im}(\mathcal{A})$ e $\text{Null}(\mathcal{A})$, respectivamente.

Proposição 2.2.1. *A aplicação linear $\mathcal{A} : S^n \rightarrow \mathbb{R}^m$ definida em (2.1) é sobrejetiva se e somente se, A_1, \dots, A_m são Linearmente Independentes (L.I.).*

Demonstração. Seja $v \in \mathbb{R}^m$ tal que $\mathcal{A}^*v = \sum_{i=1}^m v_i A_i = 0$. Mostraremos que $v = 0$. Denotando por $e_j \in \mathbb{R}^m$ o vetor cuja j -ésima coordenada é igual a um e as outras são nulas, temos pela sobrejetividade de \mathcal{A} que existe um $X_j \in S^n$ tal que $\mathcal{A}X_j = e_j$. Observe que $\langle e_j, v \rangle = \langle \mathcal{A}X_j, v \rangle = X_j \bullet \mathcal{A}^*v \quad \forall j = 1, \dots, m$. Então $\langle e_j, v \rangle = 0 \quad \forall j = 1, \dots, m$, o que resulta em $v = 0$. Portanto A_1, \dots, A_m são L.I.. Reciprocamente, se A_1, \dots, A_m são L.I. temos que \mathcal{A}^* é injetiva. Note que uma aplicação linear e sua adjunta possuem o mesmo posto. Segue que $m = \dim \text{Im}(\mathcal{A}^*) = \dim \text{Im}(\mathcal{A})$, portanto \mathcal{A} é sobrejetiva. \square

2.3 Derivada em Espaços de Matrizes

Apresentamos a seguir alguns exemplos, onde calculamos a diferencial e o gradiente de algumas funções matriciais, úteis em nosso contexto de PSD. Tais resultados podem ser encontrados em [35] e Dieudonné [16].

Exemplo 2.3.1. *Seja $\text{tr} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ a função traço. A sua diferencial, para toda matriz $H \in \mathbb{R}^{n \times n}$, e o vetor gradiente no ponto X são dados, respectivamente, por*

$$d \text{tr}(X)H = \text{tr}H = I \bullet H \quad e \quad \nabla \text{tr}(X) = I,$$

onde I é a matriz identidade.

Exemplo 2.3.2. Seja $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ definida por $f(X) = \text{tr}(X^k)$. A derivada parcial com respeito ao (i, j) -ésimo elemento é descrita como

$$\frac{\partial \text{tr}(X^k)}{\partial X_{ij}} = \lim_{\lambda \rightarrow 0} \frac{\text{tr}((X + \lambda E_{ij})^k) - \text{tr}(X^k)}{\lambda},$$

onde E_{ij} é a matriz elementar cujo (i, j) -ésimo elemento é um e todos os outros são zero. Por exemplo, quando $k = 3$,

$$(X + \lambda E_{ij})^3 = (X^3 + \lambda E_{ij} X X + \lambda X E_{ij} X + \lambda X X E_{ij}) + O(\lambda^2).$$

Note que o traço desta igualdade é simplesmente descrito por

$$\text{tr}((X + \lambda E_{ij})^3) = \text{tr}(X^3) + 3\lambda \text{tr}(E_{ij} X^2) + O(\lambda^2) = \text{tr}(X^3) + 3\lambda X_{j,i}^2 + O(\lambda^2).$$

Assim, $\nabla \text{tr}(X^3) = 3(X^2)^T$. Podemos provar que para qualquer k , vale

$$\nabla \text{tr}(X^k) = k(X^{k-1})^T. \quad (2.2)$$

Portanto, a diferencial de $f(X) = \text{tr}(X^k)$ no ponto X é dada por

$$d \text{tr}(X) H = \nabla \text{tr}(X^k) \bullet H = k(X^{k-1})^T \bullet H.$$

Exemplo 2.3.3. Seja $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ a função que associa a cada matriz $X \in \mathbb{R}^{n \times n}$ seu determinante. O desenvolvimento de Laplace segunda a i -ésima linha de X nos dá

$$\det X = \sum_{j=1}^n (-1)^{i+j} X_{ij} \det X_{\bar{i}\bar{j}},$$

onde $X_{\bar{i}\bar{j}}$ é a matriz de ordem $(n-1) \times (n-1)$ obtida de X omitindo a i -ésima linha e a j -ésima coluna. Segue-se da igualdade acima que

$$\frac{\partial \det}{\partial X_{ij}}(X) = (-1)^{i+j} \det X_{\bar{i}\bar{j}}.$$

Assim a diferencial e o vetor gradiente no ponto X são dados, respectivamente, por

$$d \det(X) H = (\text{adj} X)^T \bullet H \quad e \quad \nabla \det(X) = (\text{adj} X)^T,$$

onde $H \in \mathbb{R}^{n \times n}$ e $\text{adj} X$ denota a adjunta clássica de X que é dada por

$$\text{adj} X = ((-1)^{i+j} \det X_{\bar{i}\bar{j}})^T.$$

Exemplo 2.3.4. Sejam $\det : S_{++}^n \rightarrow \mathbb{R}$ e $\ln : (0, +\infty) \rightarrow \mathbb{R}$. Consideremos a composição $\varphi : S_{++}^n \rightarrow \mathbb{R}$ dada por $\varphi(X) = \ln \det(X)$. Então pela regra da cadeia e Exemplo 2.3.3 temos

$$d\varphi(X)H = d(\ln \det)(X)H = \frac{1}{\det X} d\det(X)H = \frac{1}{\det X} (\text{adj}X)^T \bullet H,$$

onde $H \in S^n$. Como $X \in S_{++}^n$, temos que $\text{adj}X = (\det X)X^{-1}$ e assim segue da igualdade anterior que

$$d\varphi(X)H = X^{-1} \bullet H \quad e \quad \nabla\varphi(X) = X^{-1}.$$

Exemplo 2.3.5. Sejam $\det : S_{++}^n \rightarrow \mathbb{R}$ e $g : \mathbb{R} \rightarrow \mathbb{R}$ dada por $g(y) = y^{-\alpha}$, onde $\alpha > 0$. Consideremos a composição $\varphi : S_{++}^n \rightarrow \mathbb{R}$ dada por $\varphi(X) = g(\det(X)) = \det(X)^{-\alpha}$. Então pela regra da cadeia e Exemplo 2.3.3, temos, para toda matriz H ,

$$d(g \circ \det)(X)H = -\alpha \det(X)^{-\alpha-1} d\det(X)H = -\alpha \det(X)^{-\alpha-1} (\text{adj}X)^T \bullet H.$$

Como $X \in S_{++}^n$, temos que $\text{adj}X = (\det X)X^{-1}$ e assim, em consequência da igualdade anterior, tem-se

$$d\varphi(X)H = -\alpha \det(X)^{-\alpha} X^{-1} \bullet H \quad e \quad \nabla\varphi(X) = -\alpha \det(X)^{-\alpha} X^{-1}.$$

Exemplo 2.3.6. Seja $F : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ definida por $F(X) = X^2$. Note que

$$F(X + H) = (X + H)(X + H) = F(X) + XH + HX + \|H\|\eta(H),$$

onde $\eta(H) = H^2/\|H\|$. Agora note que $\|\eta(H)\| \leq \|H\|$, assim $\lim_{v \rightarrow 0} \eta(v) = 0$. Portanto a derivada é

$$DF(X)H = XH + HX.$$

Exemplo 2.3.7. Seja $F : S_{++}^n \rightarrow \mathbb{R}^{n \times n}$ definida por $F(X) = X^{-1}$. Como $\det X \neq 0$ e $\text{adj}X = (\det X)X^{-1}$ segue que F é diferenciável. Note que $XF(X) = Id_n$. Então temos

$$0 = D(XF(X))H = D(X)HF(X) + XDF(X)H = HF(X) + XDF(X)H,$$

o que implica que $DF(X)H = -X^{-1}HF(X)$ e assim obtemos

$$DF(X)H = -X^{-1}HX^{-1}.$$

Exemplo 2.3.8. Seja $F : S_{++}^n \rightarrow \mathbb{R}^{n \times n}$ definida por $F(X) = X^{-2}$. Como $\det X \neq 0$ e $\text{adj}X = (\det X)X^{-1}$ decorre que F é diferenciável. Note que $X^2F(X) = Id_n$. Então, usando o Exemplo 2.3.6 temos

$$0 = D(X^2F(X))H = (HX + XH)F(X) + X^2DF(X)H,$$

Como $F(X) = X^{-2}$, após algumas manipulações algébricas simples na equação acima obtemos que

$$DF(X)H = -X^{-2}HX^{-1} - X^{-1}HX^{-2}.$$

Exemplo 2.3.9. Sejam $\varphi : S_{++}^n \rightarrow \mathbb{R}$ dada por

$$\varphi(X) = \text{tr}(F(X)) = \text{tr} X^{-1}.$$

Então pela regra da cadeia e Exemplo 2.3.7 temos

$$D\varphi(X) = \text{tr}(-X^{-1}HX^{-1}) = -X^{-2} \bullet H.$$

Assim obtemos

$$D\varphi(X)H = -X^{-2} \bullet H, \quad \nabla\varphi(X) = -X^{-2}.$$

Existem muitas variações do teorema da função implícita e uma de suas versões é válida para espaços de Banach, veja por exemplo [16], Teorema 10.2.1, página 265. Como o espaço das matrizes simétricas S^n é um espaço vetorial de dimensão $n(n+1)/2$ ele é naturalmente identificado com o espaço $\mathbb{R}^{n(n+1)/2}$. Sendo assim, para maior comodidade de uso, enunciamos a seguir uma versão do teorema da função implícita que trata das funções analíticas. Faremos uso desta versão para provarmos a analiticidade da trajetória central primal-dual.

Teorema 2.3.1. ([16], Teorema 10.2.4, página 268) Seja $f : U \subset \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ analítica no aberto $U \subset \mathbb{R}^n \times \mathbb{R}^m$. Suponhamos que no ponto (x_0, y_0) , com $f(x_0, y_0) = 0$, a matriz $n \times n$

$$\left[\frac{\partial f}{\partial y}(x_0, y_0) \right] (i, j = 1, \dots, n)$$

seja não singular. Então existem $Z \subset U$, aberto contendo (x_0, y_0) , $V \subset \mathbb{R}^m$, aberto contendo x_0 , e $\xi : V \rightarrow \mathbb{R}^n$ analítica, com $y_0 = \xi(x_0)$ e $f(x, \xi(x)) = 0$ para todo $x_0 \in V$. Além disso,

$$\xi'(x) = - \left[\frac{\partial f}{\partial y}(x_0, y_0) \right]^{-1} \cdot \frac{\partial f}{\partial x}(x_0, y_0).$$

2.4 Alguns Elementos da Análise Matricial

Os resultados a seguir podem ser encontrados em [13], [26], [47] e [50]. Começamos com um resultado sobre o traço do produto de duas matrizes devido a Von Neumann (veja, por exemplo, Marshall e Olkin [39], pág. 514 ou Prasolov [47], Teorema 34.4.2, pág. 148).

Lema 2.4.1. ([47], Teorema 34.4.2, pág. 148) Dadas as matrizes $X, Y \in S^n$. Sejam $\lambda_1(X) \geq \dots \geq \lambda_n(X)$ os autovalores de X e $\lambda_1(Y) \geq \dots \geq \lambda_n(Y)$ os autovalores de Y . Então

$$X \bullet Y \leq \sum_{i=1}^n \lambda_i(X) \lambda_i(Y).$$

Lema 2.4.2. Para todo $X, Y \in S_+^n$ vale $X \bullet Y \geq 0$.

Demonstração. Sejam $X, Y \in S_+^n$, então

$$X \bullet Y = \text{tr}(XY) = \text{tr}(X^{\frac{1}{2}} X^{\frac{1}{2}} Y^{\frac{1}{2}} Y^{\frac{1}{2}}) = \text{tr}(X^{\frac{1}{2}} Y^{\frac{1}{2}} Y^{\frac{1}{2}} X^{\frac{1}{2}}) = \|X^{\frac{1}{2}} Y^{\frac{1}{2}}\|^2 \geq 0.$$

□

Lema 2.4.3. Para todo $X, Y \in S_+^n$, $\lambda_{\min}(X) \text{tr} Y \leq X \bullet Y$, onde $\lambda_{\min}(X)$ é o menor autovalor da matriz X .

Demonstração. Seja v um autovetor de X e λ o seu autovalor associado. Então

$$(X - \lambda_{\min}(X)I)v = Xv - \lambda_{\min}(X)Iv = \lambda v - \lambda_{\min}(X)v = (\lambda - \lambda_{\min}(X))v,$$

onde I é a matriz identidade. Portanto $\lambda - \lambda_{\min}(X)$ é autovalor de $X - \lambda_{\min}(X)I$. Agora, note que $\lambda - \lambda_{\min}(X) \geq 0$, visto que $\lambda_{\min}(X)$ é o menor autovalor da matriz X . Logo, $X - \lambda_{\min}(X)I \succeq 0$. Como $Y \succeq 0$, segue do Lema 2.4.2 que

$$(X - \lambda_{\min}(X)I) \bullet Y \geq 0,$$

o que implica a desigualdade desejada. □

Lema 2.4.4. Para todo $X \in S_+^n$ vale $\|X\| \leq \text{tr} X$.

Demonstração. Sabemos que

$$\sqrt{\sum_{i=1}^n x_i^2} \leq \sum_{i=1}^n x_i, \quad \forall x \in \mathbb{R}_+^n. \quad (2.3)$$

Seja $\lambda_i(X)$, $i = 1, \dots, n$ os autovalores de X . Então a desigualdade anterior implica

$$\|X\|^2 = X \bullet X = \text{tr}(X^2) = \sum_{i=1}^n \lambda_i(X)^2 \leq \left(\sum_{i=1}^n \lambda_i(X) \right)^2 = (\text{tr } X)^2.$$

Desta desigualdade segue que $\|X\| \leq \text{tr } X$. □

Dada uma matriz $X \in \mathbb{R}^{n \times n}$, $\exp(X)$ e $\ln(X)$ denotam a exponencial e o logaritmo matricial, respectivamente. A exponencial matricial é definida como a seguinte série de potências

$$\exp(X) := \sum_{j=0}^{\infty} \frac{1}{j!} X^j = I + X + \frac{1}{2!} X^2 + \frac{1}{3!} X^3 + \dots + \frac{1}{j!} X^j + \dots.$$

No caso de matrizes simétricas, a exponencial de matriz pode ser calculada usando a decomposição de autovalores $X = Q^T \Lambda Q$, onde Q é uma matriz ortogonal cujas colunas são os autovetores de X e Λ é uma matriz diagonal cujos elementos da diagonal são os autovalores de X . Assim, $\exp(X) = Q^T(\exp(\Lambda))Q$, onde $(\exp(\Lambda))_{i,i} = \exp(\Lambda_{i,i})$. O logaritmo de matriz, $\ln(X)$, é definido como a função inversa de $\exp(X)$, o qual nem sempre existe para uma matriz arbitrária X . Porém, quando X é uma matriz simétrica definida positiva, $\ln(X)$ é calculado através de $\ln(X) := Q^T(\ln(\Lambda))Q$, onde $(\ln(\Lambda))_{i,i} = \ln(\Lambda_{i,i})$. Assim, para qualquer $Y \in S_{++}^n$, $\ln(Y) = X$ se, e somente se, $Y = \exp(X)$.

Exemplo 2.4.1. *Seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ definida por $\varphi(X) = X \bullet \ln(X)$. Consideremos a expansão*

$$\ln(X) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} (X - I)^i,$$

ou equivalentemente,

$$\ln(W + I) = \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} W^i,$$

com $X = W + I$, onde I é a matriz identidade. Da última expressão obtemos,

$$\begin{aligned}
(W + I) \ln(W + I) - (W + I) &= (W + I) \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} W^i - (W + I) \\
&= \sum_{i=1}^{\infty} \frac{(-1)^{i-1}}{i} (W^{i+1} + W^i) - (W + I) \\
&= W^2 + W - \frac{1}{2} W^3 - \frac{1}{2} W^2 + \dots - W - I \\
&= \frac{1}{2} W^2 - \frac{1}{6} W^3 + \frac{1}{12} W^4 - \frac{1}{20} W^5 + \dots - I \\
&= \sum_{i=2}^{\infty} \frac{(-1)^i}{i(i-1)} W^i - I.
\end{aligned}$$

Como $X = W + I$, da expressão anterior temos que

$$X \ln(X) - X = \sum_{i=2}^{\infty} \frac{(-1)^i}{i(i-1)} (X - I)^i - I.$$

Visto que $\ln(X)$ é analítico, aplicando uma versão modificada de (2.2), isto é,

$$\nabla \operatorname{tr}((X - I)^k) = k(X - I)^{k-1},$$

a todos os termos da última igualdade acima, obtemos $\nabla \operatorname{tr}(X \ln(X) - X) = \ln(X)$, isto é,

$$\nabla \varphi(X) = \nabla \operatorname{tr}(X \ln(X)) = \ln(X) + I.$$

Mostraremos agora que φ é estritamente convexa. Primeiro, como $X, Y \in S_{++}^n$, existem matrizes ortonormais Q e R tais que $X = Q^T \Lambda Q$ e $Y = R^T \Omega R$, onde Λ e Ω são matrizes diagonais cujos elementos da diagonal são, respectivamente, os autovalores λ_i de X e α_i de Y . Consideremos a seguinte expressão

$$\begin{aligned}
d(X, Y) &= \varphi(X) - \varphi(Y) - \nabla \varphi(Y) \bullet (X - Y) \\
&= X \bullet \ln(X) - Y \bullet \ln(Y) - (\ln Y + I) \bullet (X - Y) \\
&= X \bullet \ln(X) - X \bullet \ln(Y) + I \bullet Y - I \bullet X \\
&\geq \sum_{i=1}^n \lambda_i \ln(\lambda_i) - \sum_{i=1}^n \lambda_i \ln(\alpha_i) + \sum_{i=1}^n \alpha_i - \sum_{i=1}^n \lambda_i \\
&= \sum_{i=1}^n \left(\lambda_i \ln\left(\frac{\lambda_i}{\alpha_i}\right) + \alpha_i - \lambda_i \right). \tag{2.4}
\end{aligned}$$

onde na primeira desigualdade usamos o Lema 2.4.1. Observe que o lado direito da última igualdade acima é a expressão da divergência Kullback-Leibler associada

à função entropia $\mathbb{R}_{++}^n \ni \lambda \mapsto \sum_{i=1}^n \lambda_i \ln(\lambda_i)$. Dessa forma a garantimos a desigualdade estrita, ou seja, $d(X, Y) > 0$ se os autovalores de X e Y são diferentes. Portanto,

$$\varphi(X) > \varphi(Y) + \nabla\varphi(Y) \bullet (X - Y), \quad X, Y \in S_{++}^n.$$

e assim φ é estritamente convexa.

2.5 Elementos da Geometria Riemanniana

A seguir apresentamos um breve resumo de conceitos básicos da Geometria Riemanniana, os quais podem ser encontrados em do Carmo [15].

Definição 2.5.1. (*Variedade Diferenciável*). Uma variedade diferenciável de dimensão n é um conjunto M e uma família de aplicações injetivas $\mathcal{X} : U_\alpha \rightarrow M$, $\alpha \in I$ (conjunto de parâmetros), definidos em abertos U_α de \mathbb{R}^n em M tais que se cumprem as seguintes condições:

1. $M = \bigcup_{\alpha \in I} \mathcal{X}_\alpha(U_\alpha)$

2. Para todo par $\mathcal{X}_\alpha, \mathcal{X}_\beta$ com $\mathcal{X}_\alpha(U_\alpha) \cap \mathcal{X}_\beta(U_\beta) = W \neq \emptyset$ os conjuntos $\mathcal{X}_\alpha^{-1}(W)$ e $\mathcal{X}_\beta^{-1}(W)$ são abertos em \mathbb{R}^n e as aplicações $\mathcal{X}_\beta^{-1} \circ \mathcal{X}_\alpha : \mathcal{X}_\alpha^{-1}(W) \rightarrow \mathcal{X}_\beta^{-1}(W)$ são diferenciáveis. O par $(U_\alpha, \mathcal{X}_\alpha)$ com $p \in \mathcal{X}_\alpha(U_\alpha)$ é chamado uma parametrização ou sistema de coordenadas locais de M em p . Uma família satisfazendo 1) e 2) é chamada estrutura diferenciável de M .

Definição 2.5.2. (*Espaço Tangente a uma Variedade*). O espaço tangente a uma variedade M em um ponto p representado por T_pM , é o conjunto de todos os vetores tangentes à M em p . Assim $T_pM = \{v \in \mathbb{R}^n / v \text{ é um vetor tangente à } M \text{ em } p\}$. A união de espaqs tangentes

$$\bigcup_{p \in M} T_pM$$

é chamado fibrado tangente de M . Um campo de vetores \mathcal{P} em M de classe C^r , $r \geq 0$, é uma aplicação $\mathcal{P} : M \rightarrow TM$, dada por $p \mapsto X_p \in T_pM$, de classe C^r . Denotamos o espaço dos campos de vetores em M de classe C^r por $\mathcal{X}^r(M)$.

Definição 2.5.3. Dada uma variedade diferenciável de dimensão n , uma métrica Riemanniana em M é uma correspondência que associa a cada ponto $p \in M$ um produto interno $\langle \cdot, \cdot \rangle_p$ no espaço tangente T_pM (isto é, uma forma bilinear simétrica, positiva definida em T_pM) e que varia diferenciavelmente com p .

Definição 2.5.4. (*Variedade Riemanniana*) Uma variedade diferenciável M munida de uma métrica Riemanniana $\langle \cdot, \cdot \rangle$ é chamada uma variedade Riemanniana. Denotamos uma variedade Riemanniana M e sua métrica por $(M, \langle \cdot, \cdot \rangle)$.

Definição 2.5.5. Sejam M uma variedade Riemanniana e $f : M \rightarrow \mathbb{R}$ uma função de classe C^1 . O gradiente de f é o único campo $\text{grad } f \in \mathcal{X}^0(M)$ definido por

$$df_p v = \langle \text{grad } f_p, v \rangle$$

para todo $p \in M$ e $v \in T_p M$.

2.6 Resultados Básicos de Programação Semidefinida

Consideremos o problema PSD primal

$$(P) \quad \min \{C \bullet X : \mathcal{A}X = b, X \succeq 0\}$$

e o seu problema dual associado

$$(D) \quad \max \{b^T y : \mathcal{A}^* y + S = C, S \succeq 0\}.$$

Neste trabalho suporemos as duas condições seguintes, sem mencioná-las explicitamente em nossos enunciados.

A1) A aplicação linear $\mathcal{A} : S^n \rightarrow \mathbb{R}^m$ definida em (2.1) é sobrejetiva;

A2) $\mathcal{F}^0(P) \neq \emptyset$ e $\mathcal{F}^0(D) \neq \emptyset$.

A hipótese **A1** não é essencial para nossa análise, mas é conveniente para assegurar a correspondência biunívoca entre as variáveis duais S e y . A hipótese **A2** garante que ambos os problemas (P) e (D) têm soluções ótimas, os valores ótimos de (P) e (D) são iguais e os conjuntos de soluções ótimas $F^*(P)$ e $F^*(D)$ são limitados (veja por exemplo Todd [57]). Ela também é importante para assegurar a existência da trajetória central.

A seguir mostraremos algumas relações entre os problemas (P) e (D) .

Proposição 2.6.1. (*Dualidade fraca*) Suponhamos que ambos os problemas (P) e (D) sejam viáveis. Então $C \bullet X \geq b^T y$, para quaisquer $X \in F(P)$ e $(y, S) \in F(D)$.

Demonstração. Para quaisquer $X \in F(P)$ e $(y, S) \in F(D)$, temos

$$C \bullet X - b^T y = \left(\sum_{i=1}^m y_i A_i + S \right) \bullet X - b^T y = \sum_{i=1}^m (A_i \bullet X) y_i + S \bullet X - b^T y = S \bullet X = X \bullet S.$$

Então pelo Lema 2.4.2 temos $X \bullet S \geq 0$, o que implica $C \bullet X \geq b^T y$. \square

Observação 2.6.1. *Denominamos Gap de Dualidade ao número real*

$$X \bullet S = C \bullet X - b^T y.$$

O próximo resultado pode ser encontrado em vários artigos que tratam sobre otimização semidefinida. Esse resultado apresenta condições que asseguram quando a dualidade forte acontece. Em contra partida podemos encontrar muitos exemplos na literatura que mostram como a dualidade forte pode não acontecer caso não tenhamos satisfeita a hipótese $\mathcal{F}^0(P) \neq \emptyset$ ou $\mathcal{F}^0(D) \neq \emptyset$, veja por exemplo, Vanderberghe e Boyd [59] e Luo et al. [38].

Teorema 2.6.1. *([13, 57], Dualidade Forte) Suponhamos que $F(P)$ e $F^0(D)$ sejam não vazios. Então o conjunto de soluções ótimas $F^*(P)$ é não vazio e compacto e os valores ótimos de (P) e (D) são iguais.*

Observação 2.6.2. *Se tiver a dualidade forte, então temos como condições de optimalidade para o par de problemas (P) e (D) , o seguinte sistema:*

$$\begin{aligned} \mathcal{A}X &= b, \quad X \succeq 0, \\ \mathcal{A}^*y + S &= C, \quad S \succeq 0, \\ XS &= 0. \end{aligned} \tag{2.5}$$

Lema 2.6.1. *(Ortogonalidade) Sejam $(X, (y, S)) \in F(P) \times F(D)$ e $(X^0, (y^0, S^0)) \in F(P) \times F(D)$. Então vale a seguinte relação de ortogonalidade:*

$$(X - X^0) \bullet (S - S^0) = 0.$$

Demonstração. Pela viabilidade primal de X e X^0 temos que $\mathcal{A}X = b$ e $\mathcal{A}X^0 = b$, que leva a

$$\mathcal{A}(X - X^0) = \mathcal{A}X - \mathcal{A}X^0 = 0.$$

Daí deduz-se que $(X - X^0) \in \text{Null } \mathcal{A}$. Similarmente, pela viabilidade dual de (y, S) e (y^0, S^0) obtemos $\mathcal{A}^*y + S = C$ e $\mathcal{A}^*y^0 + S^0 = C$. Resulta das duas igualdades anteriores que

$$\mathcal{A}^*(y - y^0) + (S - S^0) = C - C = 0,$$

o que implica que $S - S^0 \in \text{Im } \mathcal{A}^*$. Como $\text{Null } \mathcal{A}$ é complemento ortogonal de $\text{Im } \mathcal{A}^*$, tem-se o resultado. \square

Capítulo 3

Trajетórias Central em Programação Semidefinida

3.1 Introdução

Neste capítulo estudamos as trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, para problemas PSD. Provaremos que estas trajetórias estão bem definidas, formam uma curva analítica e convergem para uma solução dos problemas primal e dual, respectivamente. Em um contexto mais amplo, estudamos a trajetória central primal associada a uma classe de funções estritamente convexas e de classe C^2 que inclui a entropia e a barreira logarítmica. Provaremos que a trajetória central primal associada a esta classe está bem definida e converge. Além disso, desde que as funções desta classe se estendam continuamente a S_+^n provaremos que a trajetória central primal converge para o centro analítico do conjunto solução do Problema Primal (P).

3.2 Trajetória Central Primal-Dual

A *trajetória central primal* para o Problema (P), relativamente à função penalidade entropia

$$S_{++}^n \ni X \mapsto X \bullet \ln(X),$$

é o conjunto de pontos $\{X(\mu) : \mu > 0\}$ onde $X(\mu)$ é definido como

$$X(\mu) = \operatorname{argmin} \{C \bullet X + \mu X \bullet \ln(X) : \mathcal{A}X = b, X \succ 0\}, \quad \mu > 0, \quad (3.1)$$

isto é, $X(\mu)$ é a única solução do problem (P_μ) definido em (1.2).

No próximo resultado provamos a boa definição da trajetória central associada a função penalidade entropia.

Teorema 3.2.1. *A trajetória central primal está bem definida e contida em $\mathcal{F}^0(P)$.*

Demonstração. Para cada $\mu > 0$ definimos $\phi_\mu : S_{++}^n \rightarrow \mathbb{R}$ por

$$\phi_\mu(X) = C \bullet X + \mu X \bullet \ln(X).$$

A função $\phi_\mu(\cdot)$ é estritamente convexa e contínua em S_+^n com a convenção de que $0 \ln 0 = 0$. O seu gradiente é dado por

$$\nabla \phi_\mu(X) = C + \mu \ln(X) + \mu I,$$

e $e^{-(C+\mu I)/\mu} \in S_{++}^n$ é o único minimizador. Considere $\tilde{X} \in \mathcal{F}^0(P)$, assim

$$\mathcal{L} = \{X \in S_+^n : \phi_\mu(X) \leq \phi_\mu(\tilde{X})\},$$

é limitado e não vazio e como $\phi_\mu(\cdot)$ é contínua em S_+^n concluímos que \mathcal{L} é compacto e não vazio. Como $\mathcal{F}(P)$ é fechado e não-vazio temos que $\mathcal{L} \cap \mathcal{F}(P)$ é também compacto e não vazio. Portanto, a convexidade estrita de $\phi_\mu(\cdot)$ implica que ela tem um único minimizador $X(\mu) \in \mathcal{F}(P)$, e portanto, a trajetória central primal está bem definida.

Resta mostrar que $X(\mu) \in \mathcal{F}^0(P)$. Suponhamos por contradição que $X(\mu)$ pertença a fronteira do conjunto viável primal, isto é,

$$X(\mu) \in \partial \mathcal{F}(P) = \{X \in S^n : AX = b, X \succeq 0, \det X = 0\},$$

onde $\det X$ denota o determinante da matriz X . Definimos

$$Z_\varepsilon = (1 - \varepsilon)X(\mu) + \varepsilon \tilde{X},$$

onde $\varepsilon \in (0, 1)$. Então, como $\tilde{X} \in \mathcal{F}^0(P)$, $X(\mu) \in \partial \mathcal{F}(P)$, $\varepsilon \in (0, 1)$ e $\mathcal{F}^0(P)$ é convexo, concluímos que $Z_\varepsilon \in \mathcal{F}^0(P)$ para todo $\varepsilon \in (0, 1)$. A seguir, combinando as definições de $X(\mu)$ e Z_ε com a convexidade de $\phi_\mu(\cdot)$, após algumas manipulações algébricas chegamos a

$$0 \leq \phi_\mu(Z_\varepsilon) - \phi_\mu(X(\mu)) \leq \nabla \phi_\mu(Z_\varepsilon) \bullet (Z_\varepsilon - X(\mu)) = \frac{\varepsilon}{1 - \varepsilon} \nabla \phi_\mu(Z_\varepsilon) \bullet (\tilde{X} - Z_\varepsilon),$$

que implica $0 \leq \nabla \phi_\mu(Z_\varepsilon) \bullet (\tilde{X} - Z_\varepsilon)$. Assim, do Lemma 2.4.1

$$\begin{aligned} 0 &\leq \nabla \phi_\mu(Z_\varepsilon) \bullet (\tilde{X} - Z_\varepsilon) = (C + \mu \ln(Z_\varepsilon) + \mu I) \bullet (\tilde{X} - Z_\varepsilon) \\ &= \mu \ln(Z_\varepsilon) \bullet \tilde{X} - C \bullet Z_\varepsilon - \mu \ln(Z_\varepsilon) \bullet Z_\varepsilon - \mu I \bullet Z_\varepsilon + (C + \mu I) \bullet \tilde{X} \\ &\leq \mu \sum_{i=1}^n \lambda_i(\tilde{X}) \lambda_i(\ln(Z_\varepsilon)) - \mu \phi_\mu(Z_\varepsilon) - \mu I \bullet Z_\varepsilon + (C + \mu I) \bullet \tilde{X}. \end{aligned}$$

Dado que a desigualdade acima é válida para todo $\varepsilon \in (0, 1)$, fazendo ε ir para 0 obtemos um absurdo. De fato, como estamos sob a hipótese de que $X(\mu) \in \partial\mathcal{F}(P)$, usando também o fato que Z_ε tende para $X(\mu)$, $\mu > 0$, $\tilde{X} \succ 0$ e a função ϕ_μ é contínua, o lado direito da desigualdade acima converge para $-\infty$. Portanto, este absurdo resulta no resultado desejado. \square

A *trajetória central dual* associada a (P) é o conjunto de pontos $\{S(\mu) : \mu > 0\}$, onde $S(\mu)$ satisfaz

$$S(\mu) = -\mu \ln(X(\mu)) - \mu I, \quad \mu > 0, \quad (3.2)$$

ou, equivalentemente, $(S(\mu), y(\mu))$ é a única solução do problema (D_μ) , definido em (1.2), para algum $y(\mu) \in \mathbb{R}^m$. O conjunto de pontos $\{(X(\mu), y(\mu), S(\mu)) : \mu > 0\}$ denota a *trajetória central primal-dual* e é a única solução do seguinte sistema de equações não-lineares

$$\begin{aligned} \mathcal{A}X &= b, \quad X \succ 0, \\ \mathcal{A}^*y + S &= C, \\ S + \mu \ln(X) + \mu I &= 0, \quad \mu > 0. \end{aligned} \quad (3.3)$$

3.3 Analiticidade da Trajetória Central Primal-Dual

Provaremos que a trajetória central primal-dual é uma curva analítica contida no conjunto $F^0(P) \times F(D)$. Isto será conseqüência do teorema da função implícita Teorema 2.3.1.

Teorema 3.3.1. *A trajetória central primal-dual é uma curva analítica contida em $S_{++}^n \times \mathbb{R}^m \times S^n$.*

Demonstração. Para simplificar a notação, seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ dada por $\varphi(X) = X \bullet \ln(X)$. Assim, (3.2) é equivalente a

$$S(\mu) = -\mu \nabla \varphi(X(\mu)), \quad \mu > 0.$$

Seja $(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0) \in S_{++}^n \times \mathbb{R}^m \times S^n \times \mathbb{R}_{++}$. Introduzimos a aplicação $\Psi : S_{++}^n \times \mathbb{R}^m \times S^n \times \mathbb{R}_{++} \rightarrow \mathbb{R}^m \times S^n \times S^n$ dada por

$$\Psi(X, y, S, \mu) = \begin{pmatrix} \mathcal{A}X - b \\ \mathcal{A}^*y + S - C \\ \mu \nabla \varphi(X) + S \end{pmatrix}.$$

Observe que $\Psi(X, y, S, \mu) = 0$ é equivalente ao sistema (3.3). Dado que a trajetória central primal-dual é a única solução do sistema (3.3) temos $\Psi(X(\mu), y(\mu), S(\mu), \mu) = 0$, para todo $\mu > 0$. Assim, como Ψ é uma função analítica a afirmação é conseqüência do teorema da função implícita desde que mostremos que a sua derivada relativamente a (X, y, S) é não-singular em $(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0)$. Com esta finalidade, basta provar que seu espaço nulo é o trivial. Suponhamos então que

$$\nabla \Psi_{(X, y, S)}(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0)(U, v, W) = 0,$$

ou equivalentemente,

$$\begin{aligned} \mathcal{A}U &= 0 \\ \mathcal{A}^*v + W &= 0 \\ \mu_0 \nabla^2 \varphi(X(\mu_0))U + W &= 0 \end{aligned} \tag{3.4}$$

A última equação de (4.4) se escreve $W = -\mu_0 \nabla^2 \varphi(X(\mu_0))U$. Visto que φ é estritamente convexa (veja Exemplo 2.4.1), substituindo a última expressão na segunda equação de (4.4), temos, de imediato,

$$\mu_0 U = (\nabla^2 \varphi(X(\mu_0)))^{-1} \mathcal{A}^*v.$$

Usando esta equação juntamente primeira equação do sistema acima obtemos

$$\mathcal{A}(\nabla^2 \varphi(X(\mu_0)))^{-1} \mathcal{A}^*v = 0.$$

Como $\mathcal{A}(\nabla^2 \varphi(X(\mu_0)))^{-1} \mathcal{A}$ é não-singular, assim a última equação implica que $v = 0$, conseqüentemente $W = U = 0$. Logo, a derivada de Ψ com respeito a (X, y, S) é não-singular em $(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0)$. Então, pelo teorema da função implícita, existe uma função analítica $\gamma : \mu \mapsto S^n \times \mathbb{R}^m \times S^n$ definida em algum intervalo aberto contendo μ_0 tal que $\Psi(\gamma(\mu), \mu) = 0$ se μ pertence a este intervalo. Além disso, $\gamma(\mu_0) = (X(\mu_0), y(\mu_0), S(\mu_0))$ e daí $\gamma(\mu) \in S_{++}^n \times \mathbb{R}^m \times S^n$ para μ suficientemente próximo de μ_0 , pois γ é contínua. Assim, encontramos um intervalo aberto contendo μ_0 tal que a condição $\Psi(\gamma(\mu), \mu) = 0$ é equivalente ao sistema (3.3) para valores de μ neste intervalo. Lembrando que o sistema (3.3) tem uma única solução, tem-se $\gamma(\mu) = (X(\mu), y(\mu), S(\mu))$ neste intervalo, e a analiticidade da trajetória como conseqüência. \square

De acordo com o Teorema 3.2.1 e (3.1), temos para todo $\mu > 0$ que

$$\mu \ln(X(\mu)) + \mu I = -C + \mathcal{A}^*y(\mu), \tag{3.5}$$

para algum $y(\mu) \in \mathbb{R}^m$.

3.4 Convergência da Trajetória Central Primal-Dual

A seguir provaremos que a trajetória central primal é limitada e que todos os seus pontos de acumulação são soluções do problema (P) , quando μ vai para 0.

Proposição 3.4.1. *As seguintes afirmações são verdadeiras:*

- (i) a função $0 < \mu \mapsto X(\mu) \bullet \ln(X(\mu))$ é não-crescente;
- (ii) o conjunto $\{X(\mu) : 0 < \mu \leq \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;
- (iii) todos os pontos de acumulação da trajetória central primal são soluções do Problema (P) , quando μ vai para 0.

Demonstração. Para simplificar a notação, seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ dada por $\varphi(X) = X \bullet \ln(X)$. Assim, (3.5) é equivalente a

$$\mu \nabla \varphi(X(\mu)) = -C + \mathcal{A}^* y(\mu).$$

Sejam $\mu_1, \mu_2 > 0$ com $\mu_1 < \mu_2$. Como φ é convexa e $(X(\mu_1) - X(\mu_2)) \in \text{Null } \mathcal{A}$, temos, da igualdade anterior, que

$$\mu_1(\varphi(X(\mu_1)) - \varphi(X(\mu_2))) \leq \mu_1 \nabla \varphi(X(\mu_1)) \bullet (X(\mu_1) - X(\mu_2)) = -C \bullet (X(\mu_1) - X(\mu_2))$$

e

$$\mu_2(\varphi(X(\mu_2)) - \varphi(X(\mu_1))) \leq \mu_2 \nabla \varphi(X(\mu_2)) \bullet (X(\mu_2) - X(\mu_1)) = -C \bullet (X(\mu_2) - X(\mu_1)).$$

Agora, somando termo a termo as duas últimas desigualdades, obtemos

$$(\mu_1 - \mu_2)(\varphi(X(\mu_1)) - \varphi(X(\mu_2))) \leq 0,$$

e, como $\mu_1 < \mu_2$, temos $\varphi(X(\mu_2)) \leq \varphi(X(\mu_1))$. Desta forma, a afirmação (i) é estabelecida.

Agora, fixe $\bar{\mu} > 0$. Com argumento similar ao usado para provar o item (i) deduz-se que

$$\mu(\varphi(X(\mu)) - \varphi(X(\bar{\mu}))) \leq -C \bullet (X(\mu) - X(\bar{\mu})),$$

para todo $0 < \mu < \bar{\mu}$. Do item (i) sabemos que $0 \leq \varphi(X(\mu)) - \varphi(X(\bar{\mu}))$, para todo $0 < \mu < \bar{\mu}$, então a última desigualdade leva diretamente a $C \bullet X(\mu) \leq C \bullet X(\bar{\mu})$, para todo $0 < \mu < \bar{\mu}$. Com isto,

$$\{X(\mu) : 0 < \mu < \bar{\mu}\} \subset \{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}.$$

Como $\mathcal{F}^*(P)$ é limitado, também o conjunto de nível $\{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}$ verifica esta propriedade. Logo, a afirmação (ii) segue da última inclusão.

Suponha agora \bar{X} um ponto de acumulação de $\{X(\mu) : \mu > 0\}$. Primeiro, note que $\mathcal{A}\bar{X} = b$ e $\bar{X} \succeq 0$, isto é, $\bar{X} \in \mathcal{F}(P)$. Considere uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Seja X^* uma solução do Problema (P) e $X \in \mathcal{F}^0(P)$. Para cada $\epsilon > 0$, definimos

$$Y(\epsilon) = (1 - \epsilon)X^* + \epsilon X.$$

Devido ao fato de que $X^* \in \partial\mathcal{F}^0(P)$, $X \in \mathcal{F}^0(P)$ e $\mathcal{F}^0(P)$ é convexo temos $Y(\epsilon) \in \mathcal{F}^0(P)$, para $\epsilon \in (0, 1]$. De (3.1) temos

$$C \bullet X(\mu_k) + \mu_k \varphi(X(\mu_k)) \leq C \bullet Y(\epsilon) + \mu_k \varphi(Y(\epsilon)),$$

ou,

$$\mu_k (\varphi(X(\mu_k)) - \varphi(Y(\epsilon))) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Agora, porquanto φ é convexa e $Y(\epsilon) \in \mathcal{F}^0(P)$, é fácil concluir da desigualdade acima que

$$\mu_k \nabla \varphi(Y(\epsilon)) \bullet (X(\mu_k) - Y(\epsilon)) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Assim, levando ao limite, para $k \rightarrow +\infty$, ambos lados da última desigualdade, obtemos $0 \leq C \bullet (Y(\epsilon) - \bar{X})$. Nesta, se ϵ tende para 0, temos

$$0 \leq C \bullet (X^* - \bar{X}), \quad \text{ou,} \quad C \bullet \bar{X} \leq C \bullet X^*.$$

Portanto, como X^* é uma solução do Problema (P) e $\bar{X} \in \mathcal{F}(P)$, temos da desigualdade anterior que \bar{X} também é solução do Problema (P), e a prova da afirmação (iii) está concluída. \square

A seguir provaremos a convergência da trajetória central primal e a caracterização do seu ponto limite, quando μ vai para 0.

Teorema 3.4.1. *Seja $X^c \in S_+^n$ o centro analítico de $\mathcal{F}^*(P)$, i.e., o único ponto satisfazendo*

$$X^c = \operatorname{argmin} \{X \bullet \ln(X) : X \in \mathcal{F}^*(P)\}. \quad (3.6)$$

Então $\lim_{\mu \rightarrow 0} X(\mu) = X^c$.

Demonstração. Para simplificar as notações, seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ a função definida na prova da proposição anterior. Usando a convenção de que $0 \ln 0 = 0$, é fácil verificar que φ é contínua em S_+^n . Seja \bar{X} um ponto de acumulação da trajetória central primal e uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Por outro lado, de (3.1) temos $C + \mu_k \nabla \varphi(X(\mu_k)) = \mathcal{A}^* y(\mu_k)$, para algum $y(\mu_k) \in \mathbb{R}^m$. Assim,

$$\mu_k \nabla \varphi(X(\mu_k)) \bullet (X - X(\mu_k)) = (\mathcal{A}^* y(\mu_k) - C) \bullet (X - X(\mu_k)),$$

para todo $X \in \mathcal{F}^*(P)$. Usando a convexidade de φ e o fato que $(X - X(\mu_k)) \in \text{Null}(\mathcal{A})$ a última equação se torna

$$\mu_k (\varphi(X(\mu_k)) - \varphi(X)) \leq C \bullet X - C \bullet X(\mu_k).$$

Como $X \in \mathcal{F}^*(P)$ e $\mu_k > 0$, segue da última desigualdade que $\varphi(X(\mu_k)) \leq \varphi(X)$. Agora, porque φ é contínua podemos considerar o limite, quando k vai para $+\infty$, nesta desigualdade, para concluir que $\varphi(\bar{X}) \leq \varphi(X)$, i.e., $\bar{X} \bullet \ln(\bar{X}) \leq X \bullet \ln(X)$, para todo $X \in \mathcal{F}^*(P)$. Assim, qualquer ponto de acumulação da trajetória central primal satisfaz (3.6). Por conseguinte, como X^c é o único ponto satisfazendo (3.6), a trajetória central primal converge para ele e o teorema está provado. \square

Na próxima proposição nosso objetivo é provar que a trajetória central dual é limitada, quando μ vai para 0, e que todos os seus pontos de acumulação são soluções do Problema (D).

Proposição 3.4.2. *As seguintes afirmações são verdadeiras:*

- (i) o conjunto $\{S(\mu) : 0 < \mu \leq \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;
- (ii) todos os pontos de acumulação da trajetória central dual são soluções do Problema (D), quando μ vai para 0.

Demonstração. Para provar o item (i), partamos de X^0 e S^0 estritamente viáveis para (P) e (D), respectivamente. Pelo Lema 2.6.1 (ortogonalidade) segue que

$$(X(\mu) - X^0) \bullet (S(\mu) - S^0) = 0, \quad \mu > 0.$$

Como $X(\mu) \succ 0$ e $S^0 \succ 0$, manipulações algébricas simples na equação acima conduzem a $X^0 \bullet S(\mu) \leq X(\mu) \bullet S(\mu) + X^0 \bullet S^0$. Agora, substituindo $S(\mu)$, dado por (3.2), no lado direito desta desigualdade obtemos

$$X^0 \bullet S(\mu) \leq -\mu X(\mu) \bullet (\ln(X(\mu)) + I) + X^0 \bullet S^0.$$

Então, como $X(\mu) \succ 0$, use a Proposição 3.4.1(i) e $\mu > 0$, na última desigualdade para obter

$$X^0 \bullet S(\mu) \leq -\mu X(\bar{\mu}) \bullet \ln(X(\bar{\mu})) + X^0 \bullet S^0, \quad (3.7)$$

para todo $0 < \mu \leq \bar{\mu}$. Observe que já teríamos provado a afirmação (i), usando os Lemas 2.4.3 e 2.4.4, caso $S(\mu)$ fosse definida positiva. Não podemos, entretanto, assegurar esta propriedade. Voltemos então ao item (i). Seja $X(\mu) \in S_{++}^n$, logo existe uma matriz ortogonal $Q(\mu)$ tal que

$$X(\mu) = Q^T(\mu)\Lambda(\mu)Q(\mu),$$

onde $\Lambda(\mu) \in S_{++}^n$ é uma matriz diagonal cujos elementos da diagonal são os autovalores de $X(\mu)$. De (3.2), obtemos

$$S(\mu) = Q^T(\mu)(-\mu(\ln(\Lambda(\mu)) + I))Q(\mu), \quad (3.8)$$

onde $-\mu(\ln(\Lambda(\mu)) + I) = \text{diag}(-\mu(\ln(\lambda_1(X(\mu)) + 1)), \dots, -\mu(\ln(\lambda_n(X(\mu)) + 1)))$.

Considere $X^c \in S_+^n$ o centro analítico de $\mathcal{F}^*(P)$,

$$B := \{j : \lambda_j(X^c) > 0\} \quad \text{e} \quad \tilde{B} := \{j : \lambda_j(X^c) = 0\}.$$

Do Teorema 3.4.1 temos que $X^c = \lim_{\mu \rightarrow 0} X(\mu)$. Assim, é fácil verificar que

$$\lim_{\mu \rightarrow 0} -\mu(\ln(\Lambda(\mu)_B) + I_B) = 0, \quad (3.9)$$

e que existe $\tilde{\mu} > 0$ tal que para todo $0 < \mu < \tilde{\mu} \leq \bar{\mu}$ se tem

$$-\mu(\ln(\Lambda(\mu)_{\tilde{B}}) + I_{\tilde{B}}) \succ 0. \quad (3.10)$$

Denote $\Sigma(\mu) := -\mu(\ln(\Lambda(\mu)) + I)$. Daí, de (3.9) resulta que $\Sigma(\mu)_B$ tende para a matriz nula, quando μ vai para 0, e de (3.10) vemos que $\Sigma(\mu)_{\tilde{B}}$ é definida positiva. Segue de (3.8) que $S(\mu) = Q^T(\mu)\Sigma(\mu)Q(\mu)$, logo

$$X^0 \bullet S(\mu) = (Q(\mu)X^0Q^T(\mu))_B \bullet \Sigma(\mu)_B + (Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \bullet \Sigma(\mu)_{\tilde{B}}.$$

Como $\Sigma(\mu)_{\tilde{B}} \succ 0$ e $(Q^T(\mu)X^0Q(\mu))_{\tilde{B}} \succ 0$ é evidente que

$$\lambda_{\min}((Q(\mu)X^0Q^T(\mu))_{\tilde{B}}) \|\Sigma(\mu)_{\tilde{B}}\| \leq (Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \bullet \Sigma(\mu)_{\tilde{B}}.$$

Assim, combinando a equação e desigualdades acima com $\lambda_{\min}(X^0) \leq \lambda_{\min}((Q(\mu)X^0Q^T(\mu))_{\tilde{B}})$ (veja em Horn e Johnson [27], Teorema 4.3.15, página 189) obtemos que

$$\|\Sigma(\mu)_{\tilde{B}}\| \leq (X^0 \bullet S(\mu) - (Q(\mu)X^0Q^T(\mu))_B \bullet \Sigma(\mu)_B) / \lambda_{\min}(X^0).$$

Usando agora o fato de que $\lim_{\mu \rightarrow 0} \Sigma(\mu)_B = 0$, e que $Q(\mu)$ é uma matriz ortogonal, a última equação junto com (3.7) levam à limitação de $\Sigma(\mu)_{\bar{B}}$ quando μ vai para 0. Conseqüentemente, como

$$\|S(\mu)\|^2 = \|\Sigma(\mu)_B\|^2 + \|\Sigma(\mu)_{\bar{B}}\|^2,$$

$\Sigma(\mu)_B$ e $\Sigma(\mu)_{\bar{B}}$ são limitadas quando μ vai para 0, concluímos que $\{S(\mu) : 0 < \mu \leq \bar{\mu}\}$ satisfaz a mesma propriedade. Desta forma, a afirmação (i) está estabelecida.

Para provar o item (ii), seja \bar{S} um ponto de acumulação da trajetória central dual. Note que devemos mostrar que

$$\mathcal{A}^* \bar{y} + \bar{S} = C, \quad X^* \bar{S} = 0, \quad \bar{S} \succeq 0, \quad (3.11)$$

para algum $\bar{y} \in \mathbb{R}^m$ e $X^* \in \mathcal{F}^*(P)$. Como a trajetória central dual satisfaz à segunda equação em (3.3) é suficiente verificar as duas últimas equações em (3.11). Seja $\{\mu_k\}$ uma seqüência tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\bar{S} = \lim_{k \rightarrow +\infty} S(\mu_k)$. Primeiro, note que $X(\mu) \ln(X(\mu))$ é limitada quando μ vai para 0, e do Teorema 3.4.1 temos que $X^c = \lim_{\mu \rightarrow 0} X(\mu)$. Logo, segue de (3.2) que

$$X^c \bar{S} = \lim_{k \rightarrow +\infty} X(\mu_k) S(\mu_k) = - \lim_{k \rightarrow +\infty} (\mu_k X(\mu_k) \ln(X(\mu_k)) + \mu_k X(\mu_k)) = 0.$$

Como $X^c \in \mathcal{F}^*(P)$ chegamos então à segunda relação em (3.11). Resta mostrar a terceira relação em (3.11). Usando a mesma notação como na prova do item (i), temos de (3.8) que

$$\bar{S} = \lim_{k \rightarrow +\infty} S(\mu_k) = \lim_{k \rightarrow +\infty} (Q^T(\mu_k)(-\mu_k(\ln(\Lambda(\mu_k)) + I)Q(\mu_k)). \quad (3.12)$$

Visto que $Q(\mu_k)$ é ortogonal para todo k , podemos supor sem perda de generalidade que $\lim_{k \rightarrow +\infty} Q(\mu_k) = Q$. Por outro lado, da convergência de $-\mu_k(\ln(\Lambda(\mu_k)) + I)$ quando k vai para $+\infty$, de (3.9) e (3.10) obtemos que

$$\bar{S}_B = \lim_{k \rightarrow +\infty} -\mu_k(\ln(\Lambda(\mu_k)_B) + I_B) = 0, \quad \bar{S}_{\bar{B}} = \lim_{k \rightarrow +\infty} -\mu_k(\ln(\Lambda(\mu_k)_{\bar{B}}) + I_{\bar{B}}) \succeq 0.$$

Assim, concluímos de (3.12) que $\bar{S} \succeq 0$. Portanto, a terceira relação em (3.11) está provada e a afirmação (ii) é estabelecida. \square

A Proposição 3.4.2 estende para a programação semidefinida a Proposição 3.1 de Cominetti e San Martín [7]. Agora provaremos a convergência da trajetória central primal-dual usando resultados da teoria dos conjuntos semianalíticos devidos a Lojasiewicz [36]. Vale ressaltar que os principais argumentos em nossa prova são os mesmos de Halická et al. [24].

Definição 3.4.1. Um subconjunto $W \subseteq \mathbb{R}^n$ é chamado um conjunto semianalítico se ele é descrito por uma união finita de conjuntos

$$\{x \in \mathbb{R}^n : f_1(x) = 0, \dots, f_m(x) = 0, g_1(x) > 0, \dots, g_l(x) > 0\},$$

onde $f_1, \dots, f_m, g_1, \dots, g_l$ são funções analíticas reais.

A seguir enunciamos o lema de seleção de curva que será útil na prova de convergência da trajetória central primal-dual.

Lema 3.4.1. ([36], Proposição 2, página 103) Seja $W \subseteq \mathbb{R}^n$ um conjunto semianalítico. Se $0 \in \overline{W} - W$, onde \overline{W} é o fecho de W , então existe algum $\varepsilon > 0$ e uma curva analítica real $\alpha : [0, \varepsilon) \rightarrow \overline{W}$ com $\alpha(0) = 0$ e $\alpha(t) \in W$ para $t \in (0, \varepsilon)$.

Uma versão particular deste lema foi usada por Kojima et al. [33] e Halická et al. [24], para provar a convergência da trajetória central, em um contexto diferente. Outras aplicações deste lema em programação matemática podem ser encontradas em Bolte et al. [6] e Papa Quiroz e Oliveira [44]. Este lema foi usado com outros objetivos, veja por exemplo Kurdyka et al. [34] e referências que tratam sobre ele. Para uma versão mais geral deste lema, veja Shiota [53], propriedade I.2.1.7, página 42.

Lema 3.4.2. ([13], Teorema D.1, página 248) Seja $f : I \rightarrow \mathbb{R}$ uma função analítica tal que $f(x) = 0$ para todo $x \in J$, onde $J \subset I$ é um conjunto com um ponto de acumulação $x_0 \in I$. Então $f(x) = 0$ para todo $x \in I$.

Teorema 3.4.2. A trajetória central primal-dual converge.

Demonstração. Seja (X^*, y^*, S^*) um ponto de acumulação da trajetória central primal-dual, o qual existe devido à Proposição 3.4.1 e Proposição 3.4.2, e seja $\{\mu_k\}$ uma seqüência de números positivos tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} (X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$. Considere W um conjunto semianalítico definido por

$$W = \left\{ (\bar{X}, \bar{y}, \bar{S}, \mu) \in S_{++}^n \times \mathbb{R}^m \times S_{++}^n \times \mathbb{R}_{++} : \begin{array}{l} \mathcal{A}\bar{X} = 0 \\ \mathcal{A}^*\bar{y} + \bar{S} = 0 \\ (\bar{S} + S^*) + \mu \ln(\bar{X} + X^*) + \mu I = 0 \\ \bar{X} + X^* \succ 0 \\ \mu > 0 \end{array} \right\}.$$

Note que o elemento zero pertence a $\overline{W} - W$. De fato, defina a seqüência

$$(\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) := (X(\mu_k) - X^*, y(\mu_k) - y^*, S(\mu_k) - S^*, \mu_k).$$

Obviamente, $(\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) \in W$. Logo, como $\lim_{k \rightarrow +\infty} (X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$ temos que

$$\lim_{k \rightarrow +\infty} (\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) = (0_{n \times n}, 0_m, 0_{n \times n}, 0).$$

Assim, o Lema 3.4.1 determina a existência de um $\varepsilon > 0$ e de uma função analítica $\alpha : [0, \varepsilon) \mapsto \overline{W}$ com $\alpha(0) = 0$ e $\alpha(t) = (\bar{X}(t), \bar{y}(t), \bar{S}(t), \mu(t)) \in W$ para $t \in (0, \varepsilon)$. Diante disto, visto também que o sistema que define a trajetória central primal-dual tem uma única solução, é fácil ver que o sistema que define W também tem uma única solução, que é dada por

$$\bar{X}(t) = X(\mu(t)) - X^*, \quad \bar{y}(t) = y(\mu(t)) - y^*, \quad \bar{S}(t) = S(\mu(t)) - S^*, \quad \mu(t) > 0,$$

para $t > 0$. Como $\mu(0) = 0$, $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} (X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$ as igualdades acima implicam

$$\lim_{t \downarrow 0} X(\mu(t)) = X^*, \quad \lim_{t \downarrow 0} y(\mu(t)) = y^*, \quad \lim_{t \downarrow 0} S(\mu(t)) = S^*, \quad \lim_{t \downarrow 0} \mu(t) = 0.$$

Como $\mu : [0, \varepsilon) \mapsto \mathbb{R}$ é uma função analítica real satisfazendo $\mu(t) > 0$ sobre $(0, \varepsilon)$ e $\mu(0) = 0$, devemos ter $\mu'(0) \geq 0$. Assim, temos duas possibilidades:

- i) $\mu'(0) > 0$;
- ii) $\mu'(0) = 0$.

Se $\mu'(0) > 0$, existe um intervalo $(0, \delta)$ onde $\mu'(t) > 0$. Logo, μ é crescente, e, por conseguinte, ela é invertível neste intervalo. No segundo caso, se $\mu'(0) = 0$ afirmamos que existe um intervalo $(0, \delta)$ onde $\mu'(t) > 0$. Caso contrário, existiria uma seqüência $\{t_k\}$ em $(0, \varepsilon)$ tal que $\lim_{k \rightarrow \infty} t_k = 0$ e $\mu'(t_k) = 0$. Como μ é uma função analítica obtemos do Lema 3.4.2 que $\mu'(t) = 0$ para todo $t \in [0, \varepsilon)$, isto é, μ é constante em $[0, \varepsilon)$. Mas $\mu(0) = 0$, donde concluímos que $\mu(t) = 0$ para todo $t \in [0, \varepsilon)$, que é um absurdo. Assim, a afirmação é estabelecida e como consequência μ é invertível neste intervalo. Portanto, em qualquer das duas possibilidades, existe a função inversa $\mu^{-1} : [0, \mu(\delta)) \rightarrow [0, \delta)$ com $\mu^{-1}(0) = 0$. Isto importa

$$\lim_{s \rightarrow 0^+} X(s) = \lim_{s \rightarrow 0^+} X(\mu(\mu^{-1}(s))) = \lim_{s \rightarrow 0^+} \bar{X}(\mu^{-1}(s)) + X^* = X^*.$$

Similarmente, $\lim_{s \rightarrow 0^+} y(s) = y^*$, $\lim_{s \rightarrow 0^+} S(s) = S^*$ e o resultado segue. \square

Cominetti e San Martin [7] obtiveram a caracterização do ponto limite da trajetória central primal-dual associada à penalidade entropia-exponencial em programação linear. O teorema acima garante a convergência da trajetória central primal-dual para PSD. No Teorema 3.4.1 acima a caracterização do ponto limite é obtida somente com respeito à trajetória central primal. A caracterização do ponto limite para a trajetória central dual é um problema em aberto.

3.5 Trajetória Central Primal

A seguir, estudamos a trajetória central primal, associada a uma classe de funções estritamente convexas e de classe C^2 que inclui a entropia e a barreira logarítmica. Desde que as funções desta classe cumpram determinadas hipóteses, provaremos que a trajetória central primal associada a esta classe, está bem definida e converge. Além disso, se as funções desta classe se estendam continuamente a S_+^n provaremos que a trajetória central primal converge para o centro analítico do conjunto solução do Problema (P).

Seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ uma função estritamente convexa e de classe C^2 . Em todos os resultados desta seção suporemos esta hipótese sem mencioná-la explicitamente. A *trajetória central primal* para o Problema (P) com respeito à função φ é o conjunto de pontos $\{X(\mu) : \mu > 0\}$ definido por

$$X(\mu) = \operatorname{argmin}_{X \succ 0} \{C \bullet X + \mu\varphi(X) : \mathcal{A}X = b\}, \mu \in \mathbb{R}_{++}. \quad (3.13)$$

Alguns de nossos resultados requerem uma das seguintes hipóteses sobre a φ :

A3) i) A função φ é contínua em S_+^n e para todo $\alpha \in \mathbb{R}$ o conjunto de nível

$$L_\alpha = \{X \in S_+^n : \varphi(X) \leq \alpha\},$$

é limitado;

ii) Para toda seqüência $\{X_k\} \subset S_{++}^n$ tal que $\lim_{k \rightarrow \infty} X_k = X \in \partial S_+^n$, temos que

$$\lim_{k \rightarrow \infty} \nabla \varphi(X_k) \bullet (\tilde{X} - X_k) = -\infty,$$

para todo $\tilde{X} \in S_{++}^n$.

A4) i) A função φ tende para $+\infty$ quando X se aproxima da fronteira ∂S_+^n de S_+^n , i.e.,

$$\lim_{X \rightarrow \partial S_+^n} \varphi(X) = +\infty.$$

ii) Para cada $V \in S_{++}^n$ e $\mu > 0$ a função $\phi_{V,\mu} : S_{++}^n \rightarrow \mathbb{R}$ definida por $\phi_{V,\mu}(X) = V \bullet X + \mu\varphi(X)$ satisfaz

$$\lim_{\|X\| \rightarrow +\infty} \phi_{V,\mu}(X) = +\infty.$$

As hipóteses **A3** e **A4** são importantes para assegurar a boa definição da trajetória central primal. A seguir, mostraremos que a hipótese **A4** implica que $\phi_{V,\mu}$ tem um conjunto de nível compacto.

Lema 3.5.1. *Suponhamos que a hipótese **A4** é satisfeita. Então o conjunto de nível*

$$K_{\alpha,\mu}(V) = \{X \in S_{++}^n : \phi_{V,\mu}(X) \leq \alpha\},$$

é compacto para cada $\alpha \in \mathbb{R}$. Consequentemente, $\phi_{V,\mu}$ tem um minimizador em S_{++}^n .

Demonstração. Seja $\alpha \in \mathbb{R}$. Afirmamos que $K_{\alpha,\mu}(V)$ é limitado. De fato, suponhamos por contradição que $K_{\alpha,\mu}(V)$ é ilimitado. Então existe uma seqüência $\{X_k\} \subset K_{\alpha,\mu}(V)$ tal que $\lim_{k \rightarrow +\infty} \|X_k\| = +\infty$. Mas, usando a hipótese **A4.ii**, tem-se $\lim_{\|X_k\| \rightarrow +\infty} \phi_{V,\mu}(X_k) = +\infty$ o que é um absurdo, pois $\phi_{V,\mu}(X_k) \leq \alpha$, para todo k . Assim, obtivemos a limitação. Mostraremos agora que $K_{\alpha,\mu}(V)$ é fechado. Seja $\{X_k\} \subset K_{\alpha,\mu}(V)$ tal que $\lim_{k \rightarrow +\infty} X_k = \bar{X}$. Dado que $\{X_k\} \subset S_{++}^n$, temos duas possibilidades: $\bar{X} \in S_{++}^n$ ou $\bar{X} \in \partial S_{++}^n$, onde ∂S_{++}^n denota a fronteira de S_{++}^n . Como $\mu > 0$, a hipótese **A4.i** implica que $\bar{X} \notin \partial S_{++}^n$. Assim, $\bar{X} \in S_{++}^n$. Por outro lado, temos, da continuidade de $\phi_{V,\mu}$ em S_{++}^n que $\alpha \geq \lim_{k \rightarrow +\infty} \phi_{V,\mu}(X_k) = \phi_{V,\mu}(\bar{X})$, i.e., $\bar{X} \in K_{\alpha,\mu}$. Portanto, $K_{\alpha,\mu}(V)$ é fechado. Logo $K_{\alpha,\mu}(V)$ é compacto, e é fácil concluir que $\phi_{V,\mu}$ tem um minimizador em S_{++}^n . \square

As hipóteses **A3** e **A4** serão aplicadas no exemplo a seguir. Em todos os casos, se considera $\varphi : S_{++}^n \rightarrow \mathbb{R}$.

Exemplo 3.5.1. *i) $\varphi(X) = X \bullet \ln(X)$;*

ii) $\varphi(X) = -\ln \det(X)$;

iii) $\varphi(X) = \det X^{-\alpha}$, onde $\alpha > 0$.

iv) $\varphi(X) = \text{tr } X^{-1}$;

Veremos a seguir que a função dada em i) satisfaz a hipótese **A3** e as funções dadas em ii), iii) e iv) satisfazem a hipótese **A4**.

i) Primeiramente, observe que $\varphi(X) = X \bullet \ln(X)$ é contínua em S_+^n com a convenção de que $0 \ln 0 = 0$. Como vimos no Exemplo 2.4.1, φ é estritamente convexa e o seu gradiente é dado por $\nabla\varphi(X) = \ln(X) + I$. É fácil verificar que φ tem um único minimizador $X^* = e^{-I}$. Portanto, $L_\alpha = \{X \in S_+^n : \varphi(X) \leq \alpha\}$ é limitado. Suponha agora $\{X_k\} \subset S_{++}^n$ tal que $\lim_{k \rightarrow \infty} X_k = X \in \partial S_+^n$ e $\tilde{X} \in S_{++}^n$. Podemos escrever,

$$\begin{aligned} \nabla\varphi(X_K) \bullet (\tilde{X} - X_k) &= (\ln(X_K) + I) \bullet (\tilde{X} - X_k) \\ &= \ln(X_K) \bullet \tilde{X} - \ln(X_K) \bullet X_k + I \bullet \tilde{X} - I \bullet X_k \\ &\leq \sum_{i=1}^n \lambda_i(\tilde{X}) \lambda_i(\ln(X_k)) - \varphi(X_k) + \sum_{i=1}^n \lambda_i(\tilde{X}) - \sum_{i=1}^n \lambda_i(X_k), \end{aligned}$$

onde a desigualdade é uma aplicação do Lema 2.4.1. Como $\tilde{X} \in S_{++}^n$ e $\lim_{k \rightarrow \infty} X_k = X \in \partial S_+^n$, o primeiro termo do lado direito da última desigualdade tende para $-\infty$ quando k tende para ∞ , e devido ao fato de que os outros termos têm limites finitos, deduz-se que

$$\lim_{k \rightarrow \infty} \nabla\varphi(X_k) \bullet (\tilde{X} - X_k) = -\infty.$$

Portanto, φ satisfaz **A3**.

ii) Nos Exemplos 2.3.4 e 2.3.7 obtivemos o gradiente e a hessiana de $\varphi(X) = -\ln \det(X)$ dados respectivamente por $\nabla\varphi(X) = -X^{-1}$ e $\nabla^2\varphi(X)H = X^{-1}HX^{-1}$, onde $H \in S^n$. Daí, é imediato que

$$\begin{aligned} (\nabla^2\varphi(X)H) \bullet H &= (X^{-1}HX^{-1}) \bullet H \\ &= (X^{-1/2}HX^{-1/2}) \bullet (X^{-1/2}HX^{-1/2}) \\ &= \|X^{-1/2}HX^{-1/2}\|^2 > 0, \end{aligned}$$

se $H \neq 0$. Logo $\varphi(X) = -\ln \det(X)$ é estritamente convexa. Assim, para cada $V \in S_{++}^n$ e $\mu > 0$ a função $\phi_{V,\mu}(X) = V \bullet X - \mu \ln \det(X)$ também é estritamente convexa. Visto que $\phi_{V,\mu}$ é convexa, obtemos

$$\phi_{V,\mu}(X) \geq \phi_{V,\mu}(2\mu V^{-1}) + \nabla\phi_{V,\mu}(2\mu V^{-1}) \bullet (X - 2\mu V^{-1}),$$

ou equivalentemente

$$\phi_{V,\mu}(X) \geq \frac{1}{2}V \bullet X + \mu n - \mu \ln \det(2\mu V^{-1}).$$

Como $V \in S_{++}^n$, e por conseguinte $\lim_{\|X\| \rightarrow +\infty} V \bullet X = +\infty$, obtemos da última desigualdade que $\lim_{\|X\| \rightarrow +\infty} \phi_{V,\mu}(X) = +\infty$. Por outro lado,

$$\lim_{X \rightarrow \partial S_+^n} -\ln \det(X) = +\infty.$$

Portanto, φ satisfaz **A4**.

- iii) Dos Exemplos 2.3.5 e 2.3.7, sabemos que o gradiente e a hessiana de $\varphi(X) = \det X^{-\alpha}$ são dados, respectivamente, por $\nabla\varphi(X) = -\alpha \det X^{-\alpha} X^{-1}$ e

$$\nabla^2\varphi(X)H = \alpha \det X^{-\alpha} (\alpha(X^{-1} \bullet H)X^{-1} + X^{-1}HX^{-1}),$$

onde $H \in S^n$. Dessa forma obtemos

$$(\nabla^2\varphi(X)H) \bullet H = \alpha \det X^{-\alpha} (\alpha(X^{-1} \bullet H)^2 + \|X^{-1/2}HX^{-1/2}\|^2) > 0,$$

para todo $H \neq 0$. Assim, φ é estritamente convexa, conseqüentemente $\phi_{V,\mu}(X) = V \bullet X + \mu \det X^{-\alpha}$ também o é, para todo $V \in S_{++}^n$ e $\mu > 0$. Agora, dado que $\mu > 0$, $\det X^{-\alpha} > 0$ e $V \in S_{++}^n$ temos

$$\lim_{\|X\| \rightarrow +\infty} \phi_{V,\mu}(X) \geq \lim_{\|X\| \rightarrow +\infty} V \bullet X = +\infty.$$

Finalmente, como $\lim_{X \rightarrow \partial S_{++}^n} \det X^{-\alpha} = +\infty$, temos φ satisfazendo **A4**.

- iv) Conforme obtidos nos Exemplos 2.3.9 e 2.3.8, o gradiente e a hessiana de $\varphi(X) = \text{tr} X^{-1}$ são dados, respectivamente, por

$$\nabla\varphi(X) = -X^{-2}, \quad \nabla^2\varphi(X)H = X^{-2}HX^{-1} + X^{-1}HX^{-2},$$

onde $H \in S^n$. Então deduz-se última igualdade que

$$(\nabla^2\varphi(X)H) \bullet H = \|X^{-1}HX^{-1/2}\|^2 + \|X^{-1/2}HX^{-1}\|^2 > 0,$$

para todo $H \neq 0$, o qual implica na estrita convexidade de φ . Portanto, $\phi_{V,\mu}(X) = V \bullet X + \mu \text{tr} X^{-1}$ também é estritamente convexa, para todo $V \in S_{++}^n$ e $\mu > 0$. Como $\text{tr} X^{-1} > 0$, $\mu > 0$ e $V \in S_{++}^n$ temos

$$\lim_{\|X\| \rightarrow +\infty} \phi_{V,\mu}(X) \geq \lim_{\|X\| \rightarrow +\infty} V \bullet X = +\infty.$$

Finalmente, como $\lim_{X \rightarrow \partial S_+^n} \text{tr} X^{-1} = +\infty$, concluímos que φ satisfaz **A4**.

A seguir provaremos a boa definição da trajetória central primal usando as hipóteses **A3** e **A4** separadamente.

Teorema 3.5.1. *Se φ satisfaz a hipótese **A3** ou **A4**, então a trajetória central primal $\{X(\mu) : \mu > 0\}$ está bem definida e contida em $\mathcal{F}^0(P)$.*

Demonstração. Sejam $X_0 \in \mathcal{F}^0(P)$ e $(y_0, S_0) \in \mathcal{F}^0(D)$. Para $\mu > 0$, definamos $\phi_{(S_0, \mu)} : S_{++}^n \rightarrow \mathbb{R}$ por

$$\phi_{(S_0, \mu)}(X) = S_0 \bullet X + \mu\varphi(X).$$

Primeiramente, suponhamos que φ satisfaz a hipótese **A3**. É fácil verificar que $\langle C, X \rangle = \langle S_0, X \rangle + b^T y_0$, para todo $X \in \mathcal{F}(P)$, daí (3.13) é equivalente a

$$X(\mu) = \operatorname{argmin}_{X \succ 0} \{ \phi_{(S_0, \mu)}(X) : \mathcal{A}X = b, \phi_{(S_0, \mu)}(X) \leq \phi_{(S_0, \mu)}(X_0) \}. \quad (3.14)$$

Consideremos o conjunto de nível $L_\beta = \{X \in S_+^n : \phi_{(S_0, \mu)}(X) \leq \beta\}$, onde $\beta = \phi_{(S_0, \mu)}(X_0)$. Observe que $L_\beta \subset L_\alpha = \{X \in S_+^n : \varphi(X) \leq \alpha\}$, onde $\alpha = \beta/\mu$. Por **A3.i**, L_α é limitado. Assim, L_β também é limitado. Como $\phi_{(S_0, \mu)}$ é contínua em S_+^n temos L_β compacto, o qual implica na compacidade de $L_\beta \cap \{X \in S_+^n : \mathcal{A}X = b\}$. Como $\phi_{(S_0, \mu)}$ é estritamente convexa, existe um único minimizador $X(\mu) \in \mathcal{F}(P)$ e portanto (4.21) está bem definida. Logo a trajetória central primal $\{X(\mu) : \mu > 0\}$ também está bem definida. Agora mostraremos que $X(\mu) \in \mathcal{F}^0(P)$. Suponhamos por contradição que $X(\mu) \in \partial\mathcal{F}(P) = \{X \in \partial S_+^n : \mathcal{A}X = b\}$ e definamos a seqüência

$$Z_k = (1 - \varepsilon_k)X(\mu) + \varepsilon_k X_0,$$

onde $\{\varepsilon_k\}$ é uma seqüência satisfazendo $\varepsilon_k \in (0, 1)$ e $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Então, dado que $X_0 \in \mathcal{F}^0(P)$, $X(\mu) \in \partial\mathcal{F}(P)$, $\varepsilon_k \in (0, 1)$ e $\mathcal{F}^0(P)$ é convexo, concluímos que $Z_k \in \mathcal{F}^0(P)$ para todo $\varepsilon_k \in (0, 1)$. A seguir combinando as definições de $X(\mu)$ e da seqüência $\{Z_k\}$ com a convexidade de φ , obtemos

$$\begin{aligned} 0 &\leq C \bullet Z_k + \mu\varphi(Z_k) - C \bullet X(\mu) - \mu\varphi(X(\mu)) \\ &\leq C \bullet (Z_k - X(\mu)) + \mu\nabla\varphi(Z_k) \bullet (Z_k - X(\mu)) \\ &= \varepsilon_k C \bullet (X_0 - X(\mu)) + \mu \frac{\varepsilon_k}{1 - \varepsilon_k} \nabla\varphi(Z_k) \bullet (X_0 - Z_k). \end{aligned}$$

A última desigualdade se reescreve,

$$\frac{(1 - \varepsilon_k)}{\mu} C \bullet (X(\mu) - X_0) \leq \nabla\varphi(Z_k) \bullet (X_0 - Z_k).$$

Como $\lim_{k \rightarrow \infty} Z_k = X(\mu) \in \partial\mathcal{F}(P)$, temos o termo do lado direito da desigualdade acima tendendo para $-\infty$, quando k tende para ∞ , devido a **A3**.ii, porém o termo do lado esquerdo desta desigualdade tem um limite finito, contradição que determina que $X(\mu) \in \mathcal{F}^0(P)$.

Finalmente, suponhamos que φ satisfaz a hipótese **A4**. Consideremos o conjunto de nível

$$K_{\alpha,\mu}(S_0) = \{X \in S_{++}^n : \phi_{(S_0,\mu)}(X) \leq \alpha\},$$

onde $\alpha = \phi_{(S_0,\mu)}(X_0)$. Aplicando o Lema 3.5.1 verifica-se facilmente que $K_{\alpha,\mu}(S_0) \cap \{X \in S_+^n : \mathcal{A}X = b\}$ é compacto. Dessa forma, um argumento similar ao usado na primeira parte da prova permite concluir que a trajetória central primal $\{X(\mu) : \mu > 0\}$ está bem definida e a definição de $K_{\alpha,\mu}(S_0)$ garante que ela está contida em $\mathcal{F}^0(P)$. \square

De acordo com o Teorema que acabamos de mostrar, podemos escrever, para todo $\mu > 0$, com (3.13), que

$$\mu \nabla \varphi(X(\mu)) = -C + \mathcal{A}^*y(\mu), \quad (3.15)$$

para algum $y(\mu) \in \mathbb{R}^m$.

3.6 Convergência da Trajetória Central Primal

A seguir provaremos que a trajetória central primal é limitada e que todos os seus pontos de acumulação são soluções do problema (P) , quando μ vai para 0. Os argumentos são semelhantes aos usados na prova da Proposição 3.4.1.

Proposição 3.6.1. *Suponhamos que φ satisfaz a hipótese **A3** ou **A4**. Então, as seguintes afirmações são verdadeiras:*

- i) a função $0 < \mu \mapsto \varphi(X(\mu))$ é não-crescente;*
- ii) o conjunto $\{X(\mu) : 0 < \mu < \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;*
- iii) todos os pontos de acumulação da trajetória central primal são soluções do Problema (P) , quando μ vai para 0.*

Demonstração. Sejam $\mu_1, \mu_2 > 0$ com $\mu_1 < \mu_2$. Como φ é convexa e $(X(\mu_1) - X(\mu_2)) \in \text{Null } \mathcal{A}$, temos, de (3.15), que

$$\mu_1(\varphi(X(\mu_1)) - \varphi(X(\mu_2))) \leq \mu_1 \nabla \varphi(X(\mu_1)) \bullet (X(\mu_1) - X(\mu_2)) = -C \bullet (X(\mu_1) - X(\mu_2))$$

e

$$\mu_2(\varphi(X(\mu_2)) - \varphi(X(\mu_1))) \leq \mu_2 \nabla \varphi(X(\mu_2)) \bullet (X(\mu_2) - X(\mu_1)) = -C \bullet (X(\mu_2) - X(\mu_1)).$$

Agora, somando termo a termo as duas últimas desigualdades, obtemos

$$(\mu_1 - \mu_2)(\varphi(X(\mu_1)) - \varphi(X(\mu_2))) \leq 0,$$

e, como $\mu_1 < \mu_2$, temos $\varphi(X(\mu_2)) \leq \varphi(X(\mu_1))$. Desta forma, a afirmação (i) é estabelecida.

Agora, fixe $\bar{\mu} > 0$. Com argumento similar ao usado para provar o item (i) deduz-se que

$$\mu(\varphi(X(\mu)) - \varphi(X(\bar{\mu}))) \leq -C \bullet (X(\mu) - X(\bar{\mu})),$$

para todo $0 < \mu < \bar{\mu}$. Do item (i) sabemos que $0 \leq \varphi(X(\mu)) - \varphi(X(\bar{\mu}))$, para todo $0 < \mu < \bar{\mu}$, então a última desigualdade leva diretamente a $C \bullet X(\mu) \leq C \bullet X(\bar{\mu})$, para todo $0 < \mu < \bar{\mu}$. Com isto,

$$\{X(\mu) : 0 < \mu < \bar{\mu}\} \subset \{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}.$$

Como $\mathcal{F}^*(P)$ é limitado, também o conjunto de nível $\{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}$ verifica esta propriedade. Logo, a afirmação (ii) segue da última inclusão.

Suponha agora \bar{X} um ponto de acumulação de $\{X(\mu) : \mu > 0\}$. Primeiro, note que $\mathcal{A}\bar{X} = b$ e $\bar{X} \succeq 0$, isto é, $\bar{X} \in \mathcal{F}(P)$. Considere uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Seja X^* uma solução do Problema (P) e $X \in \mathcal{F}^0(P)$. Para cada $\epsilon > 0$, definimos

$$Y(\epsilon) = (1 - \epsilon)X^* + \epsilon X.$$

Devido ao fato de que $X^* \in \partial \mathcal{F}^0(P)$, $X \in \mathcal{F}^0(P)$ e $\mathcal{F}^0(P)$ é convexo temos $Y(\epsilon) \in \mathcal{F}^0(P)$, para $\epsilon \in (0, 1]$. De (3.1) temos

$$C \bullet X(\mu_k) + \mu_k \varphi(X(\mu_k)) \leq C \bullet Y(\epsilon) + \mu_k \varphi(Y(\epsilon)).$$

Com alguma simples manipulação algébrica nesta equação obtemos

$$\mu_k(\varphi(X(\mu_k)) - \varphi(Y(\epsilon))) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Agora, porquanto φ é convexa e $Y(\epsilon) \in \mathcal{F}^0(P)$, é fácil concluir da desigualdade acima que

$$\mu_k \nabla \varphi(Y(\epsilon)) \bullet (X(\mu_k) - Y(\epsilon)) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Assim, levando ao limite, para $k \rightarrow +\infty$, ambos lados da última desigualdade, obtemos $0 \leq C \bullet (Y(\epsilon) - \bar{X})$. Nesta, se ϵ tende para 0, temos

$$0 \leq C \bullet (X^* - \bar{X}), \quad \text{ou,} \quad C \bullet \bar{X} \leq C \bullet X^*.$$

Portanto, como X^* é uma solução do Problema (P) e $\bar{X} \in \mathcal{F}(P)$, temos da desigualdade anterior que \bar{X} também é solução do Problema (P) , e a prova da afirmação (iii) está concluída. \square

Desde que a hipótese **A3** é satisfeita provaremos a convergência da trajetória central primal, quando μ vai para 0, e a caracterização do seu ponto limite de maneira similar, a prova do Teorema 3.4.1.

Teorema 3.6.1. *Suponhamos que φ satisfaz a hipótese **A3**. Seja $X^c \in S_+^n$ o centro analítico de $\mathcal{F}^*(P)$, i.e., o único ponto satisfazendo*

$$X^c = \operatorname{argmin} \{ \varphi(X) : X \in \mathcal{F}^*(P) \}. \quad (3.16)$$

Então $\lim_{\mu \rightarrow 0} X(\mu) = X^c$.

Demonstração. Seja \bar{X} um ponto de acumulação da trajetória central primal e uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Por outro lado, de (3.15) temos $C + \mu_k \nabla \varphi(X(\mu_k)) = \mathcal{A}^* y(\mu_k)$, para algum $y(\mu_k) \in \mathbb{R}^m$. Assim,

$$\mu_k \nabla \varphi(X(\mu_k)) \bullet (X - X(\mu_k)) = (\mathcal{A}^* y(\mu_k) - C) \bullet (X - X(\mu_k)),$$

para todo $X \in \mathcal{F}^*(P)$. Usando a convexidade de φ e o fato que $(X - X(\mu_k)) \in \operatorname{Null}(\mathcal{A})$ a última equação se torna

$$\mu_k (\varphi(X(\mu_k)) - \varphi(X)) \leq C \bullet X - C \bullet X(\mu_k).$$

Como $X \in \mathcal{F}^*(P)$ e $\mu_k > 0$, segue da última desigualdade que $\varphi(X(\mu_k)) \leq \varphi(X)$. Agora, porque φ é contínua podemos considerar o limite, quando k vai para $+\infty$, nesta desigualdade, para concluir que $\varphi(\bar{X}) \leq \varphi(X)$, para todo $X \in \mathcal{F}^*(P)$. Assim, qualquer ponto de acumulação da trajetória central primal satisfaz (3.16). Por conseguinte, como X^c é o único ponto satisfazendo (3.16), a trajetória central primal converge para ele e o teorema está provado. \square

A seguir enunciamos o resultado de convergência da trajetória central primal com respeito à barreira logarítmica, bastante conhecido em otimização semidefinida.

Teorema 3.6.2. [21, 24] *A trajetória central primal $\{X(\mu) : \mu > 0\}$ com respeito à função $\varphi(X) = -\ln \det(X)$ converge, quando μ vai para 0.*

Encerramos este capítulo observando que estudo sobre a convergência da trajetória central dual associada ao problema (P) , está em aberto, face às dificuldades de provar a sua limitação ou exibir um contra-exemplo.

Capítulo 4

Método do ponto proximal generalizado

4.1 Introdução

Neste capítulo usamos os resultados obtidos relativos às trajetórias central primal e dual, para obter convergência do método do ponto proximal generalizado no contexto da programação semidefinida. Em particular, estudamos o método do ponto proximal generalizado com a distância Kullback-Leibler. Provaremos que a seqüência de ponto proximal está contida na trajetória central primal. Como conseqüência, ambas convergem para a mesma solução ótima, a saber, o centro analítico do conjunto ótimo com respeito à distância Kullback-Leibler. Por outro lado, provaremos que a trajetória central dual contém uma seqüência dual ponderada construída a partir da seqüência de ponto proximal dual. Conseqüentemente, a seqüência dual ponderada convergirá para uma solução do Problema (D). Em um contexto mais amplo, estudaremos o método do ponto proximal generalizado associado a uma distância generalizada. Como no caso particular, mostraremos que a seqüência gerada pelo método está contida na trajetória central primal. Conseqüentemente ambas convergem para a mesma solução ótima, a saber, o centro analítico do conjunto ótimo definido pela distância generalizada. É importante mencionar que nosso objetivo é trazer para a programação semidefinida as idéias de Iusem et al. [29] e Iusem e Monteiro [28].

4.2 Método do Ponto Proximal Generalizado com a Distância Kullback-Leibler

Definimos a *distância Kullback-Leibler* $D : S_{++}^n \times S_{++}^n \rightarrow \mathbb{R}$ por

$$D(X, Y) = X \bullet \ln(X) - X \bullet \ln(Y) + \text{tr } Y - \text{tr } X.$$

Esta função pode ser vista também como uma distância de Bregman associada à barreira entropia $\varphi(X) = X \bullet \ln(X)$ considerada em Doljansky e Teboule [17].

Observação 4.2.1. *Para cada $Y \in S_{++}^n$ fixado é fácil verificar que $D(\cdot, Y)$ é de classe C^2 , estritamente convexa e se estende continuamente a S_+^n com a convenção que $0 \ln 0 = 0$.*

A trajetória central primal para o Problema (P), induzida pela função $D(\cdot, X_0)$, é o conjunto de pontos $\{X(\mu) : \mu > 0\}$, onde $X(\mu)$ é definido como

$$X(\mu) = \operatorname{argmin}_{X \in S_{++}^n} \{C \bullet X + \mu D(X, X_0) : \mathcal{A}X = b\}, \quad \mu > 0. \quad (4.1)$$

A seguir provaremos a boa definição da trajetória central primal, induzida pela função $D(\cdot, X_0)$ usando a Observação 4.2.1 e argumentos similares aos usados na prova do Teorema 3.2.1.

Teorema 4.2.1. *A trajetória central primal induzida pela função $D(\cdot, X_0)$ está bem definida e contida em $\mathcal{F}^0(P)$.*

Demonstração. Dados $\mu > 0$ e $X_0 \in S_{++}^n$. Definimos $\phi_\mu : S_{++}^n \times S_{++}^n \rightarrow \mathbb{R}$ por

$$\phi_\mu(X, X_0) = C \bullet X + \mu D(X, X_0).$$

Segue da Observação 4.2.1 que a função $\phi_\mu(\cdot, X_0)$ é estritamente convexa e contínua em S_+^n com a convenção de que $0 \ln 0 = 0$. O seu gradiente é dado por

$$\nabla \phi_\mu(X, X_0) = C + \mu \ln(X) - \mu \ln(X_0),$$

e $e^{-(C - \mu \ln(X_0))/\mu} \in S_{++}^n$ é o único minimizador. Considere $\tilde{X} \in \mathcal{F}^0(P)$, assim

$$\mathcal{L} = \{X \in S_+^n : \phi_\mu(X, X_0) \leq \phi_\mu(\tilde{X}, X_0)\},$$

é limitado e não vazio e como $\phi_\mu(\cdot, X_0)$ é contínua em S_+^n concluímos que \mathcal{L} é compacto e não vazio. Como $\mathcal{F}(P)$ é fechado e não vazio temos que $\mathcal{L} \cap \mathcal{F}(P)$ é

também compacto e não vazio. Portanto, a convexidade estrita de $\phi_\mu(\cdot, X_0)$ implica que ela tem um único minimizador $X(\mu) \in \mathcal{F}(P)$, e portanto, a trajetória central primal está bem definida.

Resta mostrar que $X(\mu) \in \mathcal{F}^0(P)$. Suponhamos por contradição que

$$X(\mu) \in \partial\mathcal{F}(P) = \{X \in S^n : AX = b, X \succeq 0, \det X = 0\},$$

onde $\det X$ denota o determinante da matriz X . Definimos

$$Z_\varepsilon = (1 - \varepsilon)X(\mu) + \varepsilon\tilde{X},$$

onde $\varepsilon \in (0, 1)$. Então, como $\tilde{X} \in \mathcal{F}^0(P)$, $X(\mu) \in \partial\mathcal{F}(P)$, $\varepsilon \in (0, 1)$ and $F^0(P)$ é convexo, concluimos que $Z_\varepsilon \in F^0(P)$ para todo $\varepsilon \in (0, 1)$. A seguir, combinando as definições de $X(\mu)$ e Z_ε com a convexidade de $\phi_\mu(\cdot, X_0)$, após algumas manipulações algébricas chegamos a

$$\begin{aligned} 0 &\leq \phi_\mu(Z_\varepsilon, X_0) - \phi_\mu(X(\mu), X_0) \\ &\leq \nabla\phi_\mu(Z_\varepsilon, X_0) \bullet (Z_\varepsilon - X(\mu)) \\ &= \frac{\varepsilon}{1 - \varepsilon} \nabla\phi_\mu(Z_\varepsilon, X_0) \bullet (\tilde{X} - Z_\varepsilon). \end{aligned}$$

Logo $0 \leq \nabla\phi_\mu(Z_\varepsilon, X_0) \bullet (\tilde{X} - Z_\varepsilon)$. Assim, desta inequação e do Lema 2.4.1 temos

$$\begin{aligned} 0 &\leq \nabla\phi_\mu(Z_\varepsilon, X_0) \bullet (\tilde{X} - Z_\varepsilon) \\ &= (C + \mu \ln(Z_\varepsilon) - \mu \ln(X_0)) \bullet (\tilde{X} - Z_\varepsilon) \\ &= \mu \ln(Z_\varepsilon) \bullet \tilde{X} - \mu \ln(Z_\varepsilon) \bullet Z_\varepsilon + (C - \mu \ln(X_0)) \bullet (\tilde{X} - Z_\varepsilon) \\ &\leq \mu \sum_{i=1}^n \lambda_i(\tilde{X}) \lambda_i(\ln(Z_\varepsilon)) - \mu \ln(Z_\varepsilon) \bullet Z_\varepsilon + (C - \mu \ln(X_0)) \bullet (\tilde{X} - Z_\varepsilon) \end{aligned}$$

Dado que a desigualdade acima é válida para todo $\varepsilon \in (0, 1)$, fazendo ε ir para 0 obtemos um absurdo. De fato, como estamos sob a hipótese de que $X(\mu) \in \partial\mathcal{F}(P)$, usando também o fato que Z_ε tende para $X(\mu)$, $\mu > 0$, $\tilde{X} \succ 0$ e que a função

$$S_{++}^n \ni X \mapsto X \bullet \ln(X)$$

é contínua, o lado direito da desigualdade acima converge para $-\infty$. Portanto, este absurdo resulta no resultado desejado. \square

A trajetória central dual associada ao Problema (P) , relativo à função $D(\cdot, X_0)$, é o conjunto de pontos $\{S(\mu) : \mu > 0\}$, onde $S(\mu)$ satisfaz

$$S(\mu) = -\mu(\ln(X(\mu)) - \ln(X_0)), \quad \mu > 0, \quad (4.2)$$

ou equivalentemente, $(y(\mu), S(\mu))$ é a única solução do problema de otimização

$$\max \{ b^T y - \mu \operatorname{tr} e^{-S/\mu + \ln(X_0)} : \mathcal{A}^* y + S = C \}, \quad \mu > 0.$$

O conjunto $\{(X(\mu), y(\mu), S(\mu)) : \mu > 0\}$ denota a trajetória central primal-dual gerada por $D(\cdot, X_0)$, e é a única solução do seguinte sistema de equações não-lineares

$$\begin{aligned} \mathcal{A}X &= b, \quad X \succ 0, \\ \mathcal{A}^* y + S &= C, \\ S + \mu \ln(X) - \mu \ln(X_0) &= 0, \quad \mu > 0. \end{aligned} \tag{4.3}$$

Similarmente à prova do Teorema 3.3.1 usando a Observação 4.2.1 mostraremos a seguir que a trajetória central primal-dual, associada à função $D(\cdot, X_0)$, é uma curva analítica contida em $S_{++}^n \times \mathbb{R}^m \times S^n$.

Teorema 4.2.2. *A trajetória central primal-dual, associada à função $D(\cdot, X_0)$, é uma curva analítica contida em $S_{++}^n \times \mathbb{R}^m \times S^n$.*

Demonstração. Primeiro observe que (4.2) é equivalente a

$$S(\mu) = -\mu \nabla D(X(\mu), X_0), \quad \mu > 0.$$

Seja $(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0) \in S_{++}^n \times \mathbb{R}^m \times S^n \times \mathbb{R}_{++}$. Introduzimos a aplicação $\Psi : S_{++}^n \times \mathbb{R}^m \times S^n \times \mathbb{R}_{++} \rightarrow \mathbb{R}^m \times S^n \times S^n$ dada por

$$\Psi(X, y, S, \mu) = \begin{pmatrix} \mathcal{A}X - b \\ \mathcal{A}^* y + S - C \\ \mu \nabla D(X, X_0) + S \end{pmatrix}.$$

Observe que $\Psi(X, y, S, \mu) = 0$ é equivalente ao sistema (4.3). Dado que a trajetória central primal-dual é a única solução do sistema (4.3) temos $\Psi(X(\mu), y(\mu), S(\mu), \mu) = 0$, para todo $\mu > 0$. Assim, como Ψ é uma função analítica a afirmação é consequência do teorema da função implícita desde que mostremos que a sua derivada relativamente a (X, y, S) é não-singular em $(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0)$. Com esta finalidade, basta provar que seu espaço nulo é o trivial. Suponhamos então que

$$\nabla \Psi_{(X, y, S)}(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0)(U, v, W) = 0,$$

ou

$$\begin{aligned} \mathcal{A}U &= 0 \\ \mathcal{A}^* v + W &= 0 \\ \mu_0 \nabla^2 D(X(\mu_0), X_0)U + W &= 0 \end{aligned} \tag{4.4}$$

A última equação de (4.4) se escreve $W = -\mu_0 \nabla^2 D(X(\mu_0), X_0)U$. Visto que $D(\cdot, X_0)$ é estritamente convexa (veja Observação 4.2.1), substituindo a última expressão na segunda equação de (4.4), temos, de imediato,

$$\mu_0 U = (\nabla^2 D(X(\mu_0), X_0))^{-1} \mathcal{A}^* v,$$

e, em vista da primeira equação no sistema (4.4) obtemos a igualdade

$$\mathcal{A}(\nabla^2 D(X(\mu_0), X_0))^{-1} \mathcal{A}^* v = 0.$$

Como a matriz deste sistema é não-singular obtemos que $v = 0$. A segunda equação do sistema (4.4) obtemos $W = 0$. Como a função $D(\cdot, X_0)$ é estritamente convexa segue da última equação do sistema (4.4) que $U = 0$. Logo, a derivada de Ψ com respeito a (X, y, S) é não-singular em $(X(\mu_0), y(\mu_0), S(\mu_0), \mu_0)$. Então, pelo teorema da função implícita, existe uma função analítica $\gamma : \mu \mapsto S^n \times \mathbb{R}^m \times S^n$ definida em algum intervalo aberto contendo μ_0 tal que $\Psi(\gamma(\mu), \mu) = 0$ se μ pertence a este intervalo. Além disso, $\gamma(\mu_0) = (X(\mu_0), y(\mu_0), S(\mu_0))$ e daí $\gamma(\mu) \in S_{++}^n \times \mathbb{R}^m \times S^n$ para μ suficientemente próximo de μ_0 , pois γ é contínua. Assim, encontramos um intervalo aberto contendo μ_0 tal que a condição $\Psi(\gamma(\mu), \mu) = 0$ é equivalente ao sistema (4.3) para valores de μ neste intervalo. Lembrando que o sistema (4.3) tem uma única solução, tem-se $\gamma(\mu) = (X(\mu), y(\mu), S(\mu))$ neste intervalo, e a analiticidade da trajetória como consequência. \square

Do Teorema 4.2.1, podemos escrever, usando (4.1) que, para todo $\mu > 0$,

$$\mu \nabla D(X(\mu), X_0) = -C + \mathcal{A}^* y(\mu), \quad (4.5)$$

para algum $y(\mu) \in \mathbb{R}^m$.

De maneira similar à prova da Proposição 3.4.1 usando a Observação 4.2.1 provaremos a seguir que a trajetória central primal, induzida pela função $D(\cdot, X_0)$, é limitada e que todos os seus pontos de acumulação são soluções para o problema (P) , quando μ vai para 0.

Proposição 4.2.1. *As seguintes afirmações são verdadeiras:*

- (i) a função $0 < \mu \mapsto D(X(\mu), X_0)$ é não-crescente;
- (ii) o conjunto $\{X(\mu) : 0 < \mu \leq \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;

(iii) todos os pontos de acumulação da trajetória central primal induzida pela função $D(\cdot, X_0)$, são soluções do Problema (P), quando μ vai para 0.

Demonstração. Sejam $\mu_1, \mu_2 > 0$ com $\mu_1 < \mu_2$. Como $D(\cdot, X_0)$ é convexa (veja Observação 4.2.1) e $(X(\mu_1) - X(\mu_2)) \in \text{Null } \mathcal{A}$, temos, de (4.5), que

$$\begin{aligned} \mu_1(D(X(\mu_1), X_0) - D(X(\mu_2), X_0)) &\leq \mu_1 \nabla D(X(\mu_1), X_0) \bullet (X(\mu_1) - X(\mu_2)) \\ &= -C \bullet (X(\mu_1) - X(\mu_2)), \end{aligned}$$

Da mesma maneira vale a desigualdade a seguir

$$\begin{aligned} \mu_2(D(X(\mu_2), X_0) - D(X(\mu_1), X_0)) &\leq \mu_2 \nabla D(X(\mu_2), X_0) \bullet (X(\mu_2) - X(\mu_1)) \\ &= -C \bullet (X(\mu_2) - X(\mu_1)). \end{aligned}$$

Agora, somando termo a termo as duas últimas desigualdades, obtemos que

$$(\mu_1 - \mu_2)(D(X(\mu_1), X_0) - D(X(\mu_2), X_0)) \leq 0$$

e, como $\mu_1 < \mu_2$ temos $D(X(\mu_2), X_0) \leq D(X(\mu_1), X_0)$. Desta forma, a afirmação (i) é estabelecida.

Agora, fixe $\bar{\mu} > 0$. Com argumento similar ao usado para provar o item (i) deduz-se que

$$\mu(D(X(\mu), X_0) - D(X(\bar{\mu}), X_0)) \leq -C \bullet (X(\mu) - X(\bar{\mu})),$$

para todo $0 < \mu < \bar{\mu}$. Do item (i) sabemos que $0 \leq D(X(\mu), X_0) - D(X(\bar{\mu}), X_0)$, para todo $0 < \mu < \bar{\mu}$, então a última desigualdade leva diretamente a $C \bullet X(\mu) \leq C \bullet X(\bar{\mu})$, para todo $0 < \mu < \bar{\mu}$. Com isto,

$$\{X(\mu) : 0 < \mu < \bar{\mu}\} \subset \{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}.$$

Como $\mathcal{F}^*(P)$ é limitado, também o conjunto de nível $\{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}$ verifica esta propriedade. Logo, a afirmação (ii) segue da última inclusão.

Suponha agora \bar{X} um ponto de acumulação de $\{X(\mu) : \mu > 0\}$. Primeiro, note que $\mathcal{A}\bar{X} = b$ e $\bar{X} \succeq 0$, isto é, $\bar{X} \in \mathcal{F}(P)$. Considere uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Seja X^* uma solução do Problema (P) e $X \in \mathcal{F}^0(P)$. Para cada $\epsilon > 0$, definimos

$$Y(\epsilon) = (1 - \epsilon)X^* + \epsilon X.$$

Devido ao fato que $X^* \in \partial\mathcal{F}^0(P)$, $X \in \mathcal{F}^0(P)$ e $\mathcal{F}^0(P)$ é convexo temos $Y(\epsilon) \in \mathcal{F}^0(P)$, para $\epsilon \in (0, 1]$. De (3.1) temos

$$C \bullet X(\mu_k) + \mu_k D(X(\mu_k), X_0) \leq C \bullet Y(\epsilon) + \mu_k D(Y(\epsilon), X_0),$$

ou equivalentemente

$$\mu_k (D(X(\mu_k), X_0) - D(Y(\epsilon), X_0)) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Agora, porquanto $D(\cdot, X_0)$ é convexa (veja Observação 4.2.1) e $Y(\epsilon) \in \mathcal{F}^0(P)$, é fácil concluir da desigualdade acima que

$$\mu_k \nabla D(Y(\epsilon), X_0) \bullet (X(\mu_k) - Y(\epsilon)) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Assim, levando ao limite, para k goes to $+\infty$, ambos lados da última desigualdade, obtemos $0 \leq C \bullet (Y(\epsilon) - \bar{X})$. Nesta, se ϵ tende para 0, temos

$$0 \leq C \bullet (X^* - \bar{X}), \quad \text{ou,} \quad C \bullet \bar{X} \leq C \bullet X^*.$$

Portanto, como X^* é uma solução do Problema (P) e $\bar{X} \in \mathcal{F}(P)$, temos da desigualdade anterior que \bar{X} também é solução do Problema (P) , e a prova da afirmação (iii) está concluída. \square

A seguir usando a Observação 4.2.1 provaremos a convergência da trajetória central primal, induzida pela função $D(\cdot, X_0)$, quando μ vai para 0, e a caracterização do seu ponto limite de maneira similar, a prova do Teorema 3.4.1.

Teorema 4.2.3. *Seja $X^c \in S_+^n$ o centro analítico de $\mathcal{F}^*(P)$, i.e., o único ponto satisfazendo*

$$X^c = \operatorname{argmin} \{D(X, X_0) : X \in \mathcal{F}^*(P)\}. \quad (4.6)$$

Então $\lim_{\mu \rightarrow 0} X(\mu) = X^c$.

Demonstração. Seja \bar{X} um ponto de acumulação da trajetória central primal e uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Por outro lado, de (4.1) temos $C + \mu_k \nabla D(X(\mu_k), X_0) = \mathcal{A}^* y(\mu_k)$, para algum $y(\mu_k) \in \mathbb{R}^m$. Assim,

$$\mu_k \nabla D(X(\mu_k), X_0) \bullet (X - X(\mu_k)) = (\mathcal{A}^* y(\mu_k) - C) \bullet (X - X(\mu_k)),$$

para todo $X \in \mathcal{F}^*(P)$. Usando a convexidade de $D(\cdot, X_0)$ (veja Observação 4.2.1) e o fato que $X - X(\mu_k) \in \text{Null}(\mathcal{A})$ a última equação se torna

$$\mu_k(D(X(\mu_k), X_0) - D(X, X_0)) \leq C \bullet X - C \bullet X(\mu_k).$$

Como $X \in \mathcal{F}^*(P)$ e $\mu_k > 0$, segue da última desigualdade que $D(X(\mu_k), X_0) \leq D(X, X_0)$. Agora, porque $D(\cdot, X_0)$ é contínua (veja Observação 4.2.1) podemos considerar o limite, quando k vai para $+\infty$, nesta desigualdade para concluir que $D(\bar{X}, X_0) \leq D(X, X_0)$, para todo $X \in \mathcal{F}^*(P)$. Assim, qualquer ponto de acumulação da trajetória central primal satisfaz (4.6). Por conseguinte, como X^c é o único ponto satisfazendo (4.6) a trajetória central primal converge para ele e o teorema está provado. \square

Na próxima proposição nosso objetivo é provar que a trajetória central dual associada à função $D(\cdot, X_0)$ é limitada, quando μ vai para 0, e que todos os seus pontos de acumulação são soluções para o problema (D) , usando a Observação 4.2.1 e argumentos similares aos usados para provar a Proposição 3.4.2.

Proposição 4.2.2. *As seguintes afirmações são verdadeiras:*

- (i) o conjunto $\{S(\mu) : 0 < \mu \leq \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;
- (ii) todos os pontos de acumulação da trajetória central dual associada à função $D(\cdot, X_0)$ são soluções do Problema (D) , quando μ vai para 0.

Demonstração. Para provar o item (i), partamos de X^0 e S^0 estritamente viáveis para (P) e (D) , respectivamente. Pelo Lema 2.6.1 (Ortogonalidade) segue que

$$(X(\mu) - X^0) \bullet (S(\mu) - S^0) = 0, \quad \mu > 0.$$

Como $X(\mu) \succ 0$ e $S^0 \succ 0$, manipulações algébricas simples na equação acima conduzem a $X^0 \bullet S(\mu) \leq X(\mu) \bullet S(\mu) + X^0 \bullet S^0$. Agora, substituindo $S(\mu)$, dado por (4.2), no lado direito desta desigualdade obtemos

$$X^0 \bullet S(\mu) \leq -\mu X(\mu) \bullet \ln(X(\mu)) + \mu X(\mu) \bullet \ln(X_0) + X^0 \bullet S^0.$$

Então, como $X(\mu) \succ 0$, use a Proposição 3.4.1(i) e $\mu > 0$, na última desigualdade para obter

$$X^0 \bullet S(\mu) \leq -\mu X(\bar{\mu}) \bullet \ln(X(\bar{\mu})) + \mu X(\mu) \bullet \ln(X_0) + X^0 \bullet S^0, \quad (4.7)$$

para todo $0 < \mu \leq \bar{\mu}$. Observe que já teríamos provado a afirmação (i), usando os Lemas 2.4.3 e 2.4.4, caso $S(\mu)$ fosse definida positiva. Não podemos, entretanto, assegurar esta propriedade. Voltemos então ao item (i). Seja $X(\mu) \in S_{++}^n$, logo existe uma matriz ortogonal $Q(\mu)$ tal que

$$X(\mu) = Q^T(\mu)\Lambda(\mu)Q(\mu),$$

onde $\Lambda(\mu) \in S_{++}^n$ é uma matriz diagonal cujos elementos da diagonal são os autovalores de $X(\mu)$. De (4.2), obtemos

$$S(\mu) = Q^T(\mu) \left(-\mu(\ln(\Lambda(\mu)) - Q(\mu) \ln(X_0)Q^T(\mu)) \right) Q(\mu), \quad (4.8)$$

Considere $X^c \in S_+^n$ o centro analítico de $\mathcal{F}^*(P)$,

$$B := \{j : \lambda_j(X^c) > 0\} \quad \text{e} \quad \tilde{B} := \{j : \lambda_j(X^c) = 0\}.$$

Do Teorema 4.2.3 temos que $X^c = \lim_{\mu \rightarrow 0} X(\mu)$. Visto que as entradas da matriz $Q(\mu) \ln(X_0)Q^T(\mu)$ são limitadas, é fácil verificar que

$$\lim_{\mu \rightarrow 0} -\mu \left(\ln(\Lambda(\mu)_B) - (Q(\mu) \ln(X_0)Q^T(\mu))_B \right) = 0, \quad (4.9)$$

e que existe $\tilde{\mu} > 0$ tal que para todo $0 < \mu < \tilde{\mu} \leq \bar{\mu}$ se tem

$$-\mu \left(\ln(\Lambda(\mu)_{\tilde{B}}) - (Q(\mu) \ln(X_0)Q^T(\mu))_{\tilde{B}} \right) \succ 0. \quad (4.10)$$

Denote $\Sigma(\mu) := -\mu \left(\ln(\Lambda(\mu)) - Q(\mu) \ln(X_0)Q^T(\mu) \right)$. Daí, de (4.9) resulta que $\Sigma(\mu)_B$ tende para a matriz nula, quando μ vai para 0, e de (4.10) vemos que $\Sigma(\mu)_{\tilde{B}}$ é definida positiva. Segue de (4.8) que $S(\mu) = Q^T(\mu)\Sigma(\mu)Q(\mu)$, logo

$$X^0 \bullet S(\mu) = (Q(\mu)X^0Q^T(\mu))_B \bullet \Sigma(\mu)_B + (Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \bullet \Sigma(\mu)_{\tilde{B}}.$$

Como $\Sigma(\mu)_{\tilde{B}} \succ 0$ e $(Q^T(\mu)X^0Q(\mu))_{\tilde{B}} \succ 0$ é evidente que

$$\lambda_{\min} \left((Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \right) \|\Sigma(\mu)_{\tilde{B}}\| \leq (Q(\mu)X^0Q^T(\mu))_{\tilde{B}} \bullet \Sigma(\mu)_{\tilde{B}}.$$

Assim, combinando a equação e desigualdades acima com $\lambda_{\min}(X^0) \leq \lambda_{\min}((Q(\mu)X^0Q^T(\mu))_{\tilde{B}})$ (veja em Horn e Johnson (1985), Teorema 4.3.15, página 189) obtemos que

$$\|\Sigma(\mu)_{\tilde{B}}\| \leq (X^0 \bullet S(\mu) - (Q(\mu)X^0Q^T(\mu))_B \bullet \Sigma(\mu)_B) / \lambda_{\min}(X^0).$$

Usando agora o fato de que $\lim_{\mu \rightarrow 0} \Sigma(\mu)_B = 0$, e que $Q(\mu)$ é uma matriz ortogonal, a última equação junto com (4.7) levam à limitação de $\Sigma(\mu)_{\bar{B}}$ quando μ vai para 0. Conseqüentemente, como

$$\|S(\mu)\|^2 = \|\Sigma(\mu)_B\|^2 + \|\Sigma(\mu)_{\bar{B}}\|^2,$$

$\Sigma(\mu)_B$ e $\Sigma(\mu)_{\bar{B}}$ são limitadas quando μ vai para 0, concluímos que $\{S(\mu) : 0 < \mu \leq \bar{\mu}\}$ satisfaz a mesma propriedade. Desta forma, a afirmação (i) está estabelecida.

Para provar o item (ii), seja \bar{S} um ponto de acumulação da trajetória central dual. Note que devemos mostrar que

$$\mathcal{A}^* \bar{y} + \bar{S} = C, \quad X^* \bar{S} = 0, \quad \bar{S} \succeq 0, \quad (4.11)$$

para algum $\bar{y} \in \mathbb{R}^m$ and $X^* \in \mathcal{F}^*(P)$. Como a trajetória central dual satisfaz a segunda equação em (4.3) é suficiente verificar as duas equações em (4.11). Seja $\{\mu_k\}$ uma seqüência tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\bar{S} = \lim_{k \rightarrow +\infty} S(\mu_k)$. Primeiro, note que $X(\mu) \ln(X(\mu))$ é limitado quando μ vai para 0, e do Teorema 4.2.3 temos que $X^c = \lim_{\mu \rightarrow 0} X(\mu)$. Logo, segue de (4.2) que

$$X^c \bar{S} = \lim_{k \rightarrow +\infty} X(\mu_k) S(\mu_k) = - \lim_{k \rightarrow +\infty} (\mu_k X(\mu_k) \ln(X(\mu_k)) - \mu_k X(\mu_k) \ln(X_0)) = 0.$$

Como $X^c \in \mathcal{F}^*(P)$ chegamos então à segunda relação em (4.11). Resta mostrar a terceira relação em (4.11). Usando a mesma notação como na prova do item (i), temos de (4.8) que

$$\begin{aligned} \bar{S} &= \lim_{k \rightarrow +\infty} S(\mu_k) \\ &= \lim_{k \rightarrow +\infty} (Q^T(\mu_k) (-\mu_k (\ln(\Lambda(\mu_k)) - Q(\mu_k) \ln(X_0) Q^T(\mu_k))) Q(\mu_k)). \end{aligned} \quad (4.12)$$

Visto que $Q(\mu_k)$ é ortogonal para todo k , podemos supor sem perda de generalidade que $\lim_{k \rightarrow +\infty} Q(\mu_k) = Q$. Por outro lado, da convergência de $-\mu_k (\ln(\Lambda(\mu_k)) - Q(\mu_k) \ln(x_0) Q^T(\mu_k))$ quando k vai para $+\infty$, de (4.9) e (4.10) obtemos que

$$\bar{S}_B = \lim_{k \rightarrow +\infty} -\mu_k (\ln(\Lambda(\mu_k)_B) - (Q(\mu_k) \ln(x_0) Q^T(\mu_k))_B) = 0$$

and

$$\bar{S}_{\bar{B}} = \lim_{k \rightarrow +\infty} -\mu_k (\ln(\Lambda(\mu_k)_{\bar{B}}) - (Q(\mu_k) \ln(x_0) Q^T(\mu_k))_{\bar{B}}) \geq 0.$$

Assim, concluímos de (3.12) que $\bar{S} \succeq 0$. Portanto, a terceira relação em (4.11) está provada e a afirmação (ii) é estabelecida. \square

A seguir provaremos a convergência da trajetória central primal-dual associada à função $D(\cdot, X_0)$, usando a Observação 4.2.1 e argumentos similares aos usados para provar o Teorema 3.4.2.

Teorema 4.2.4. *A trajetória central primal-dual associada à função $D(\cdot, X_0)$ converge.*

Demonstração. Seja (X^*, y^*, S^*) um ponto de acumulação da trajetória central primal-dual, o qual existe devido à Proposição 4.3.1 e Proposição 3.4.2, e seja $\{\mu_k\}$ uma seqüência de números positivos tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} (X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$. Considere W um conjunto semi-analítico definido por

$$W = \left\{ (\bar{X}, \bar{y}, \bar{S}, \mu) \in \Omega : \begin{array}{l} \mathcal{A}\bar{X} = 0 \\ \mathcal{A}^*\bar{y} + \bar{S} = 0 \\ (\bar{S} + S^*) + \mu \ln(\bar{X} + X^*) - \mu \ln(X_0) = 0 \\ \bar{X} + X^* \succ 0 \\ \mu > 0 \end{array} \right\}.$$

onde $\Omega := S_{++}^n \times \mathbb{R}^m \times S^n \times \mathbb{R}_{++}$. Note que o elemento zero pertence a $\overline{W} - W$. De fato, defina a seqüência

$$(\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) := (X(\mu_k) - X^*, y(\mu_k) - y^*, S(\mu_k) - S^*, \mu_k).$$

Obviamente, $(\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) \in W$. Logo, como $\lim_{k \rightarrow +\infty} (X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$ temos que

$$\lim_{k \rightarrow +\infty} (\bar{X}_k, \bar{y}_k, \bar{S}_k, \mu_k) = (0_{n \times n}, 0_m, 0_{n \times n}, 0).$$

Assim, o Lema 3.4.1 determina a existência de um $\varepsilon > 0$ e de uma função analítica $\alpha : [0, \varepsilon) \mapsto \overline{W}$ com $\alpha(0) = 0$ e $\alpha(t) = (\bar{X}(t), \bar{y}(t), \bar{S}(t), \mu(t)) \in W$ para $t \in (0, \varepsilon)$. Diante disto, visto também que o sistema que define a trajetória central primal-dual tem uma única solução, é fácil ver que o sistema que define W também tem uma única solução, que é dada por

$$\bar{X}(t) = X(\mu(t)) - X^*, \quad \bar{y}(t) = y(\mu(t)) - y^*, \quad \bar{S}(t) = S(\mu(t)) - S^*, \quad \mu(t) > 0,$$

para $t > 0$. Como $\mu(0) = 0$, $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} (X(\mu_k), y(\mu_k), S(\mu_k)) = (X^*, y^*, S^*)$ as igualdades acima implicam

$$\lim_{t \downarrow 0} X(\mu(t)) = X^*, \quad \lim_{t \downarrow 0} y(\mu(t)) = y^*, \quad \lim_{t \downarrow 0} S(\mu(t)) = S^*, \quad \lim_{t \downarrow 0} \mu(t) = 0.$$

Como $\mu : [0, \varepsilon) \mapsto \mathbb{R}$ é uma função analítica real satisfazendo $\mu(t) > 0$ sobre $(0, \varepsilon)$ e $\mu(0) = 0$, devemos ter $\mu'(0) \geq 0$. Assim, temos duas possibilidades:

i) $\mu'(0) > 0$;

ii) $\mu'(0) = 0$.

Se $\mu'(0) > 0$, existe um intervalo $(0, \delta)$ onde $\mu'(t) > 0$. Logo, μ é crescente, e, por conseguinte, ela é invertível neste intervalo. No segundo caso, se $\mu'(0) = 0$ afirmamos que existe um intervalo $(0, \delta)$ onde $\mu'(t) > 0$. Caso contrário, existiria uma seqüência $\{t_k\}$ em $(0, \varepsilon)$ tal que $\lim_{k \rightarrow \infty} t_k = 0$ e $\mu'(t_k) = 0$. Como μ é uma função analítica obtemos do Lema 3.4.2 que $\mu'(t) = 0$ para todo $t \in [0, \varepsilon)$, isto é, μ é constante em $[0, \varepsilon)$. Mas $\mu(0) = 0$, donde concluímos que $\mu(t) = 0$ para todo $t \in [0, \varepsilon)$, que é um absurdo. Assim, a afirmação é estabelecida e como consequência μ é invertível neste intervalo. Portanto, em qualquer das duas possibilidades, existe a função inversa $\mu^{-1} : [0, \mu(\delta)) \rightarrow [0, \delta)$ com $\mu^{-1}(0) = 0$. Isto importa

$$\lim_{s \rightarrow 0^+} X(s) = \lim_{s \rightarrow 0^+} X(\mu(\mu^{-1}(s))) = \lim_{s \rightarrow 0^+} \bar{X}(\mu^{-1}(s)) + X^* = X^*.$$

Similarmente, $\lim_{s \rightarrow 0^+} y(s) = y^*$, $\lim_{s \rightarrow 0^+} S(s) = S^*$ e o resultado segue. \square

Seja $X_0 \in \mathcal{F}^0(P)$. O método do ponto proximal com a distância Kullback-Leibler gera uma seqüência $\{X_k\} \subset S_{++}^n$ com ponto inicial $X_0 \in \mathcal{F}^0(P)$ e

$$X_{k+1} = \arg \min_{X \in S_{++}^n} \{C \bullet X + \lambda_k D(X, X_k) : \mathcal{A}X = b\}, \quad (4.13)$$

onde a seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfaz

$$\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty. \quad (4.14)$$

De agora em diante referiremos a seqüência acima $\{X_k\}$ como a seqüência de ponto proximal primal com respeito a D , associada a $\{\lambda_k\}$ e ponto inicial X_0 . A observação 4.2.1 e um argumento similar ao usado na prova do Teorema 3.2.1 permite provar a boa definição da seqüência de ponto proximal primal. Além disso, (4.13) implica que $\{X_k\}$ satisfaz

$$C + \lambda_k (\ln(X_{k+1}) - \ln(X_k)) = \mathcal{A}^* z_k, \quad (4.15)$$

para alguma seqüência $\{z_k\}$ em \mathbb{R}^m e $k = 0, 1, 2, \dots$. Também, as condições de otimalidade para (4.13) determinam a seqüência dual $\{S_k\}$ como

$$S_k = \lambda_k (\ln(X_k) - \ln(X_{k+1})), \quad k = 0, 1, 2, \dots \quad (4.16)$$

Da seqüência dual $\{S_k\}$ definimos a seqüência proximal dual ponderada $\{\bar{S}_k\}$ como

$$\bar{S}_k = \sum_{j=0}^k \lambda_j^{-1} \mu_k S_j, \quad \mu_k = \left(\sum_{j=0}^k \lambda_j^{-1} \right)^{-1}, \quad k = 0, 1, 2, \dots \quad (4.17)$$

Teorema 4.2.5. *Sejam $\{X(\mu) : \mu > 0\}$ e $\{S(\mu) : \mu > 0\}$ as trajetórias central primal e dual associadas a $D(\cdot, X_0)$, respectivamente. Suponhamos dada uma seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfazendo (4.14), e a seqüência $\{\mu_k\}$ definida como*

$$\mu_k = \left(\sum_{j=0}^k \lambda_j^{-1} \right)^{-1}, \quad \text{para } k = 0, 1, 2, \dots \quad (4.18)$$

Então $X_{k+1} = X(\mu_k)$ e $\bar{S}_k = S(\mu_k)$ para $k = 0, 1, 2, \dots$, onde $\{X_k\}$ e $\{\bar{S}_k\}$ são as seqüências de ponto proximal primal e dual ponderada associadas a $\{\lambda_k\}$, respectivamente. Em conseqüência,

$$\lim_{k \rightarrow +\infty} (X_k, \bar{S}_k) = (X^*, S^*),$$

onde $(X^*, S^*) = \lim_{\mu \rightarrow 0} (X(\mu), S(\mu))$.

Demonstração. Sejam $\{X_k\}$ e $\{S_k\}$ as seqüências de ponto proximal primal e dual, respectivamente. De (4.13), (4.15) e (4.16), temos X_k e S_k satisfazendo

$$\begin{aligned} \mathcal{A}X_{k+1} &= b, \quad X_{k+1} \succ 0, \\ \mathcal{A}^* z_k + S_k &= C, \\ S_k &= \lambda_k (\ln(X_k) - \ln(X_{k+1})), \quad \lambda_k > 0 \end{aligned}$$

para alguma seqüência $\{z_k\}$ em \mathbb{R}^m e $k = 0, 1, 2, \dots$. Da última equação do sistema anterior, segue que $\sum_{j=0}^k (1/\lambda_j) S_j = \ln(X_0) - \ln(X_{k+1})$. Deduz-se da última expressão, juntamente com (4.17) e (4.18), que

$$\bar{S}_k = -\mu_k (\ln(X_{k+1}) - \ln(X_0)).$$

Assim, é fácil concluir que $\{X_k\}$ e $\{\bar{S}_k\}$ satisfazem

$$\begin{aligned} \mathcal{A}X_{k+1} &= b, \quad X_{k+1} \succ 0, \\ \mathcal{A}^* \bar{y}_k + \bar{S}_k &= C, \\ \bar{S}_k + \mu_k \ln(X_{k+1}) - \mu_k \ln(X_0) &= 0, \quad \mu_k > 0. \end{aligned} \quad (4.19)$$

para $\bar{y}_k = \mu_k \sum_{j=0}^k (1/\lambda_j) z_j$, $k = 0, 1, 2, \dots$. Dessa forma, dos sistemas (4.3) e (4.19) obtemos $X_{k+1} = X(\mu_k)$, $\bar{y}_k = y(\mu_k)$ e $\bar{S}_k = S(\mu_k)$. Como $\{\lambda_k\}$ satisfaz (4.14), é claro que $\lim_{k \rightarrow +\infty} \mu_k = 0$. Agora, use o fato que $\lim_{\mu \rightarrow 0} (X(\mu), S(\mu)) = (X^*, S^*)$ para concluir que $\lim_{k \rightarrow +\infty} (X_k, \bar{S}_k) = (X^*, S^*)$, e a prova está completa. \square

Com argumentos similares aos usados na demonstração do Teorema 3 de Iusem et al. [29] podemos provar que, para cada seqüência decrescente positiva $\{\mu_k\}$, existe uma seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfazendo (4.14) tal que a seqüência primal $\{X_k\}$ e a seqüência dual ponderada $\{\bar{S}_k\}$ a ela associadas satisfazem $X_{k+1} = X(\mu_k)$ e $\bar{S}_k = S(\mu_k)$, onde $\{X(\mu) : \mu > 0\}$ e $\{S(\mu) : \mu > 0\}$ são as trajetórias central primal e dual associadas a $D(\cdot, X_0)$, respectivamente.

4.3 Método do Ponto Proximal Generalizado Associado a uma Distância Generalizada

Definimos inicialmente uma distância generalizada. Seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ uma função estritamente convexa e de classe C^2 . A *distância* generalizada $D_\varphi : S_{++}^n \times S_{++}^n \rightarrow \mathbb{R}$, induzida por φ , é definida por

$$D_\varphi(X, Y) = \varphi(X) - \varphi(Y) - \nabla\varphi(Y) \bullet (X - Y).$$

Exemplo 4.3.1. Para cada uma das funções do Exemplo 3.5.1, as distâncias generalizadas correspondentes são dadas, respectivamente, por:

- i) $D_\varphi(X, Y) = X \bullet \ln X - X \bullet \ln Y + \text{tr } Y - \text{tr } X$;
- ii) $D_\varphi(X, Y) = X \bullet Y^{-1} - \ln \det (XY^{-1}) - n$;
- iii) $D_\varphi(X, Y) = \det X^{-\alpha} + \det Y^{-\alpha}(\alpha X \bullet Y^{-1} - \alpha n - 1)$, onde $\alpha > 0$;
- iv) $D_\varphi(X, Y) = \text{tr } X^{-1} - 2 \text{tr } Y^{-1} + X \bullet Y^{-2}$.

Note que a distância dada em i) é a distância Kullback-Leibler.

Observação 4.3.1. Note que se φ é estritamente convexa e de classe C^2 então para cada $Y \in S_{++}^n$ fixado, a função $D_\varphi(\cdot, Y)$ também é estritamente convexa e de classe C^2 . Além disso, as seguintes afirmações são verdadeiras:

- 1) Se φ satisfaz a hipótese **A3** então $D_\varphi(\cdot, Y)$ satisfaz a hipótese **A3**;
- 2) Se φ satisfaz a hipótese **A4** então $D_\varphi(\cdot, Y)$ satisfaz a hipótese **A4**.

De fato, suponhamos que φ satisfaz **A3**, com isso é imediato concluir que $D_\varphi(\cdot, Y)$ também é contínua em S_+^n . Então, visto que $D_\varphi(\cdot, Y)$ tem um minimizador e é estritamente convexa em S_{++}^n , temos a compacidade do conjunto de nível $L_\alpha = \{X \in$

$S_+^n : D_\varphi(X, Y) \leq \alpha$, para todo $\alpha \in \mathbb{R}$. Logo, $D_\varphi(\cdot, Y)$ satisfaz **A3.i**. Consideremos $\{X_k\} \subset S_{++}^n$ tal que $\lim_{k \rightarrow \infty} X_k = X \in \partial S_+^n$. Assim,

$$\nabla D_\varphi(X_k, Y) \bullet (\tilde{X} - X_k) = \nabla \varphi(X_k) \bullet (\tilde{X} - X_k) - \nabla \varphi(Y) \bullet (\tilde{X} - X_k),$$

para todo $\tilde{X} \in S_{++}^n$. Por **A3.ii** temos que o primeiro termo do lado direito da última desigualdade tende para $-\infty$ e o segundo termo converge, quando k tende para ∞ . Portanto, $D_\varphi(\cdot, Y)$ satisfaz **A3**.

Suponhamos a seguir que φ satisfaz **A4**. Sejam $V \in S_{++}^n$ e $\mu > 0$. Dado que $D_\varphi(X, Y) \geq 0$, para $X, Y \in S_{++}^n$, e $V \in S_{++}^n$ é fácil verificar que

$$\lim_{\|X\| \rightarrow +\infty} (V \bullet X + \mu D_\varphi(X, Y)) \geq \lim_{\|X\| \rightarrow +\infty} V \bullet X = +\infty.$$

Por outro lado, se φ satisfaz **A4.i** então temos

$$\lim_{X \rightarrow \partial S_{++}^n} D_\varphi(X, Y) = \lim_{X \rightarrow \partial S_{++}^n} (\varphi(X) - \varphi(Y) - \nabla \varphi(Y) \bullet (X - Y)) = +\infty.$$

Portanto $D_\varphi(\cdot, Y)$ satisfaz **A4**, e as afirmações estão provadas.

A trajetória central primal para o Problema (P), relativamente à função $D_\varphi(\cdot, X_0)$, é o conjunto de pontos $\{X(\mu) : \mu > 0\}$, onde $X(\mu)$ é definido como

$$X(\mu) = \operatorname{argmin}_{X \in S_{++}^n} \{C \bullet X + \mu D_\varphi(X, X_0) : \mathcal{A}X = b\}, \quad \mu > 0. \quad (4.20)$$

A seguir provaremos a boa definição da trajetória central primal relativamente à função $D_\varphi(\cdot, X_0)$ usando a Observação 4.3.1 e argumentos similares aos usados para provar o Teorema 3.5.1.

Teorema 4.3.1. *Se φ satisfaz a hipótese **A3** ou **A4**, então a trajetória central primal relativamente à função $D_\varphi(\cdot, X_0)$ está bem definida e contida em $\mathcal{F}^0(P)$.*

Demonstração. Sejam $\hat{X}_0 \in \mathcal{F}^0(P)$ e $(y_0, S_0) \in \mathcal{F}^0(D)$. Para $\mu > 0$, definamos $\phi_{(\mu, S_0)} : S_{++}^n \times S_{++}^n \rightarrow \mathbb{R}$ por

$$\phi_{(\mu, S_0)}(X, X_0) = S_0 \bullet X + \mu D_\varphi(X, X_0).$$

Primeiramente, suponhamos que φ satisfaz **A3**. Fácilmente podemos verificar que

$$C \bullet X = S_0 \bullet X + b^T y_0,$$

para todo $X \in \mathcal{F}(P)$. Assim a equação (4.20) pode ser reescrita na seguinte forma

$$X(\mu) = \operatorname{argmin}_{X \succ 0} \{\phi_{(\mu, S_0)}(X, X_0) : \mathcal{A}X = b, \phi_{(\mu, S_0)}(X, X_0) \leq \beta\}, \quad (4.21)$$

onde $\beta = \phi_{(\mu, S_0)}(\hat{X}_0, X_0)$. Agora, consideremos o seguinte conjunto de nível

$$L_\beta = \{X \in S_+^n : \phi_{(\mu, S_0)}(X, X_0) \leq \beta\}.$$

Observe que $L_\beta \subset L_\alpha = \{X \in S_+^n : D_\varphi(X, X_0) \leq \alpha\}$, onde $\alpha = \beta/\mu$. Pela Observação 4.3.1 item 1), L_α é limitado. Assim, L_β também é limitado. Como $\phi_{(\mu, S_0)}$ é contínua em S_+^n temos L_β compacto, o qual implica na compacidade de

$$L_\beta \cap \{X \in S_+^n : AX = b\}.$$

Como $\phi_{(\mu, S_0)}$ é estritamente convexa, existe um único minimizador $X(\mu) \in \mathcal{F}(P)$ e portanto (4.21) está bem definida. Logo a trajetória central primal $\{X(\mu) : \mu > 0\}$ também está bem definida. Agora mostraremos que $X(\mu) \in \mathcal{F}^0(P)$. Suponhamos por contradição que $X(\mu) \in \partial\mathcal{F}(P) = \{X \in \partial S_+^n : AX = b\}$ e definimos a seqüência

$$Z_k = (1 - \varepsilon_k)X(\mu) + \varepsilon_k\hat{X}_0,$$

onde $\{\varepsilon_k\}$ é uma seqüência satisfazendo $\varepsilon_k \in (0, 1)$ e $\lim_{k \rightarrow \infty} \varepsilon_k = 0$. Então, como $\hat{X}_0 \in \mathcal{F}^0(P)$, $X(\mu) \in \partial\mathcal{F}(P)$, $\varepsilon_k \in (0, 1)$ e $F^0(P)$ é convexo, concluímos que $Z_k \in F^0(P)$ para todo $\varepsilon_k \in (0, 1)$. A seguir, combinando as definições de $X(\mu)$ e da seqüência $\{Z_k\}$ com a convexidade de $D_\varphi(\cdot, X_0)$, obtemos

$$\begin{aligned} 0 &\leq C \bullet Z_k + \mu D_\varphi(Z_k, X_0) - C \bullet X(\mu) - \mu D_\varphi(X(\mu), X_0) \\ &\leq C \bullet (Z_k - X(\mu)) + \mu \nabla D_\varphi(Z_k, X_0) \bullet (Z_k - X(\mu)) \\ &= \varepsilon_k C \bullet (\hat{X}_0 - X(\mu)) + \mu \frac{\varepsilon_k}{1 - \varepsilon_k} \nabla D_\varphi(Z_k, X_0) \bullet (\hat{X}_0 - Z_k). \end{aligned}$$

A última desigualdade se reescreve,

$$\frac{(1 - \varepsilon_k)}{\mu} C \bullet (X(\mu) - \hat{X}_0) \leq \nabla D_\varphi(Z_k, X_0) \bullet (\hat{X}_0 - Z_k).$$

Como $\lim_{k \rightarrow \infty} Z_k = X(\mu) \in \partial\mathcal{F}(P)$, temos o termo do lado direito da desigualdade acima tendendo para $-\infty$, quando k tende para ∞ , devido a Observação 4.3.1 item 1), porém o termo do lado esquerdo desta desigualdade tem um limite finito, contradição que determina que $X(\mu) \in \mathcal{F}^0(P)$.

Finalmente, suponhamos que φ satisfaz a hipótese **A4**. Note que pela Observação 4.3.1 item 2), $D_\varphi(\cdot, X_0)$ cumpre a mesma propriedade. Agora, consideremos o conjunto de nível

$$K_{\alpha, \mu}(S_0) = \{X \in S_{++}^n : \phi_{(\mu, S_0)}(X, X_0) \leq \alpha\},$$

onde $\alpha = \phi_{(\mu, S_0)}(\hat{X}_0, X_0)$. Aplicando o Lema 3.5.1 verifica-se facilmente que $K_{\alpha, \mu}(S_0) \cap \{X \in S_+^n \mid \mathcal{A}X = b\}$ é compacto. Dessa forma, um argumento similar ao usado na primeira parte da prova permite concluir que a trajetória central primal $\{X(\mu) : \mu > 0\}$ está bem definida e a definição de $K_{\alpha, \mu}(S_0)$ garante que ela está contida em $\mathcal{F}^0(P)$. \square

O Teorema 4.3.1 assegura que a trajetória central primal para o Problema (P) , induzida por $D_\varphi(X, X_0)$, está bem definida e contida em \mathcal{F}^0 . Assim, para todo $\mu > 0$, temos de (4.20) que

$$\mu \nabla D_\varphi(X(\mu), X_0) = -C + \mathcal{A}^*y(\mu), \quad (4.22)$$

ou equivalentemente,

$$C + \mu (\nabla \varphi(X(\mu)) - \nabla \varphi(X_0)) = \mathcal{A}^*y(\mu), \quad (4.23)$$

para alguma trajetória $\{y(\mu) : \mu > 0\}$ em \mathbb{R}^m .

De maneira similar à prova da Proposição 3.6.1 utilizando a Observação 4.3.1 mostraremos, a seguir, que a trajetória central primal, induzida pela função $D(\cdot, X_0)$, é limitada e que todos os seus pontos de acumulação são soluções para o problema (P) , quando μ vai para 0.

Proposição 4.3.1. *Suponhamos que φ satisfaz a hipótese A3 ou A4. Então, as seguintes afirmações são verdadeiras:*

- (i) a função $0 < \mu \mapsto D_\varphi(X(\mu), X_0)$ é não-crescente;
- (ii) o conjunto $\{X(\mu) : 0 < \mu \leq \bar{\mu}\}$ é limitado, para cada $\bar{\mu} > 0$;
- (iii) todos os pontos de acumulação da trajetória central primal induzida pela função $D_\varphi(\cdot, X_0)$, são soluções do Problema (P) , quando μ vai para 0.

Demonstração. Sejam $\mu_1, \mu_2 > 0$ com $\mu_1 < \mu_2$. Como $D_\varphi(\cdot, X_0)$ é convexa (veja Observação 4.3.1) e $(X(\mu_1) - X(\mu_2)) \in \text{Null } \mathcal{A}$, temos, de (4.22), que

$$\begin{aligned} \mu_1 (D_\varphi(X(\mu_1), X_0) - D_\varphi(X(\mu_2), X_0)) &\leq \mu_1 \nabla D_\varphi(X(\mu_1), X_0) \bullet (X(\mu_1) - X(\mu_2)) \\ &= -C \bullet (X(\mu_1) - X(\mu_2)). \end{aligned}$$

De maneira análoga vale a seguinte desigualdade

$$\begin{aligned} \mu_2 (D_\varphi(X(\mu_2), X_0) - D_\varphi(X(\mu_1), X_0)) &\leq \mu_2 \nabla D_\varphi(X(\mu_2), X_0) \bullet (X(\mu_2) - X(\mu_1)) \\ &= -C \bullet (X(\mu_2) - X(\mu_1)). \end{aligned}$$

Agora, somando termo a termo as duas últimas desigualdades, obtemos que

$$(\mu_1 - \mu_2)(D_\varphi(X(\mu_1), X_0) - D_\varphi(X(\mu_2), X_0)) \leq 0$$

e, como $\mu_1 < \mu_2$ temos $D_\varphi(X(\mu_2), X_0) \leq D_\varphi(X(\mu_1), X_0)$. Desta forma, a afirmação (i) é estabelecida.

Agora, fixe $\bar{\mu} > 0$. Com argumento similar ao usado para provar o item (i) deduz-se que

$$\mu(D_\varphi(X(\mu), X_0) - D_\varphi(X(\bar{\mu}), X_0)) \leq -C \bullet (X(\mu) - X(\bar{\mu})),$$

para todo $0 < \mu < \bar{\mu}$. Do item (i) sabemos que $0 \leq D_\varphi(X(\mu), X_0) - D_\varphi(X(\bar{\mu}), X_0)$, para todo $0 < \mu < \bar{\mu}$, então a última desigualdade leva diretamente a $C \bullet X(\mu) \leq C \bullet X(\bar{\mu})$, para todo $0 < \mu < \bar{\mu}$. Com isto,

$$\{X(\mu) : 0 < \mu < \bar{\mu}\} \subset \{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}.$$

Como $\mathcal{F}^*(P)$ é limitado, também o conjunto de nível $\{X \in \mathcal{F}(P) : C \bullet X \leq C \bullet X(\bar{\mu})\}$ verifica esta propriedade. Logo, a afirmação (ii) segue da última inclusão.

Suponha agora \bar{X} um ponto de acumulação de $\{X(\mu) : \mu > 0\}$. Primeiro, note que $\mathcal{A}\bar{X} = b$ e $\bar{X} \succeq 0$, isto é, $\bar{X} \in \mathcal{F}(P)$. Considere uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Seja X^* uma solução do Problema (P) e $X \in \mathcal{F}^0(P)$. Para cada $\epsilon > 0$, definimos

$$Y(\epsilon) = (1 - \epsilon)X^* + \epsilon X.$$

Devido ao fato que $X^* \in \partial\mathcal{F}^0(P)$, $X \in \mathcal{F}^0(P)$ e $\mathcal{F}^0(P)$ é convexo temos $Y(\epsilon) \in \mathcal{F}^0(P)$, para $\epsilon \in (0, 1]$. De (4.20) temos

$$C \bullet X(\mu_k) + \mu_k D_\varphi(X(\mu_k), X_0) \leq C \bullet Y(\epsilon) + \mu_k D_\varphi(Y(\epsilon), X_0),$$

ou equivalentemente,

$$\mu_k(D_\varphi(X(\mu_k), X_0) - D_\varphi(Y(\epsilon), X_0)) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Agora, porquanto $D_\varphi(\cdot, X_0)$ é convexa (veja Observação 4.3.1) e $Y(\epsilon) \in \mathcal{F}^0(P)$, é fácil concluir da desigualdade acima que

$$\mu_k \nabla D_\varphi(Y(\epsilon), X_0) \bullet (X(\mu_k) - Y(\epsilon)) \leq C \bullet (Y(\epsilon) - X(\mu_k)).$$

Assim, levando ao limite, para k goes to $+\infty$, ambos lados da última desigualdade, obtemos $0 \leq C \bullet (Y(\epsilon) - \bar{X})$. Nesta, se ϵ tende para 0, temos

$$0 \leq C \bullet (X^* - \bar{X}), \quad \text{ou,} \quad C \bullet \bar{X} \leq C \bullet X^*.$$

Portanto, como X^* é uma solução do Problema (P) e $\bar{X} \in \mathcal{F}(P)$, temos da desigualdade anterior que \bar{X} também é solução do Problema (P) , e a prova da afirmação (iii) está concluída. \square

A seguir provaremos a convergência da trajetória central primal, induzida pela função $D_\varphi(\cdot, X_0)$, quando μ vai para 0, e a caracterização do seu ponto limite utilizando a Observação 4.3.1 e argumentos similares aos usados para provar o Teorema 3.6.1.

Teorema 4.3.2. *Suponhamos que φ satisfaz a hipótese **A3**. Se $X^c \in S_+^n$ é o centro analítico de $\mathcal{F}^*(P)$, i.e., o único ponto satisfazendo*

$$X^c = \operatorname{argmin} \{D_\varphi(X, X_0) : X \in \mathcal{F}^*(P)\}, \quad (4.24)$$

então $\lim_{\mu \rightarrow 0} X(\mu) = X^c$.

Demonstração. Seja \bar{X} um ponto de acumulação da trajetória central primal e uma seqüência de números positivos $\{\mu_k\}$ tal que $\lim_{k \rightarrow +\infty} \mu_k = 0$ e $\lim_{k \rightarrow +\infty} X(\mu_k) = \bar{X}$. Por outro lado, de (4.22) temos

$$C + \mu_k \nabla D_\varphi(X(\mu_k), X_0) = \mathcal{A}^* y(\mu_k),$$

para algum $y(\mu_k) \in \mathbb{R}^m$. Desta equação segue imediatamente que

$$\mu_k \nabla D_\varphi(X(\mu_k), X_0) \bullet (X - X(\mu_k)) = (\mathcal{A}^* y(\mu_k) - C) \bullet (X - X(\mu_k)),$$

para todo $X \in \mathcal{F}^*(P)$. Usando a convexidade de $D_\varphi(\cdot, X_0)$ (veja Observação 4.3.1) e o fato que $X - X(\mu_k) \in \operatorname{Null}(\mathcal{A})$ a última equação se torna

$$\mu_k (D_\varphi(X(\mu_k), X_0) - D_\varphi(X, X_0)) \leq C \bullet X - C \bullet X(\mu_k).$$

Como $X \in \mathcal{F}^*(P)$ e $\mu_k > 0$, segue da última desigualdade que

$$D_\varphi(X(\mu_k), X_0) \leq D_\varphi(X, X_0).$$

Agora, porque $D_\varphi(\cdot, X_0)$ é contínua (veja Observação 4.3.1) podemos considerar o limite, quando k vai para $+\infty$, nesta desigualdade para concluir que $D_\varphi(\bar{X}, X_0) \leq$

$D_\varphi(X, X_0)$, para todo $X \in \mathcal{F}^*(P)$. Assim, qualquer ponto de acumulação da trajetória central primal satisfaz (4.24). Por conseguinte, como X^c é o único ponto satisfazendo (4.24) a trajetória central primal converge para ele e o teorema está provado. \square

Seja $X_0 \in F^0(P)$. O método de ponto proximal com a distância generalizada D_φ , para resolver o problema (P), gera uma seqüência $\{X_k\} \subset S_{++}^n$ com ponto inicial $X_0 \in F^0(P)$ e

$$X_{k+1} = \arg \min_{X \in S_{++}^n} \{C \bullet X + \lambda_k D_\varphi(X, X_k) : \mathcal{A}X = b\},$$

onde a seqüência $\{\lambda_k\} \subset \mathbb{R}_{++}$ satisfaz

$$\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty. \quad (4.25)$$

A partir de agora chamaremos a seqüência $\{X_k\}$ acima de seqüência de ponto proximal primal relativa a D_φ , associada a $\{\lambda_k\}$ e ponto inicial X_0 . Se φ satisfaz **A3** ou **A4** então a Observação 4.3.1 permite-nos usar um argumento similar ao utilizado na demonstração do Teorema 3.5.1 para provar a boa definição da seqüência de ponto proximal $\{X_k\}$. Além disso, $\{X_k\}$ satisfaz

$$C + \lambda_k (\nabla \varphi(X_{k+1}) - \nabla \varphi(X_k)) = \mathcal{A}^* z_k, \quad (4.26)$$

para alguma seqüência $\{z_k\}$ em \mathbb{R}^m e $k = 0, 1, 2, \dots$.

Lema 4.3.1. *Seja $\{\mu_k\}$ uma seqüência decrescente positiva com $\lim_{k \rightarrow +\infty} \mu_k = 0$. Se $\lambda_k := ((\mu_k)^{-1} - (\mu_{k-1})^{-1})^{-1}$ então*

$$\sum_{k=0}^{\infty} \lambda_k^{-1} = +\infty.$$

Demonstração. Basta notar que da definição de λ_k temos que

$$\sum_{k=0}^{\ell} \lambda_k^{-1} = \mu_\ell^{-1} - \mu_0^{-1}, \quad \forall \ell.$$

Como $\lim_{k \rightarrow +\infty} \mu_k = 0$, fazendo $\ell \rightarrow \infty$ na igualdade acima obtemos o resultado. \square

Teorema 4.3.3. *Suponhamos que φ satisfaz a hipótese **A3** ou **A4**. Seja $\{X(\mu) : \mu > 0\}$ a trajetória central primal relativo a $D_\varphi(\cdot, X_0)$ e seja $\{X_k\}$ a seqüência de ponto proximal primal. Se a seqüência $\{\mu_k\}$ é definida como*

$$\mu_k = \left(\sum_{j=0}^k \lambda_j^{-1} \right)^{-1}, \quad \text{para } k = 0, 1, 2, \dots, \quad (4.27)$$

então $X_{k+1} = X(\mu_k)$, para $k = 0, 1, 2, \dots$. Por outro lado, para cada seqüência decrescente positiva $\{\mu_k\}$ convergindo para zero, existe uma seqüência de números positivos $\{\lambda_k\}$ satisfazendo (4.25) tal que a seqüência de ponto proximal $\{X_k\}$ a ela associada satisfaz $X_{k+1} = X(\mu_k)$. Além disso, se φ satisfaz a hipótese **A3** então $\{X_k\}$ converge para o centro analítico do conjunto solução do Problema (P) definido através de $D_\varphi(\cdot, X_0)$.

Demonstração. (4.26) pode ser reescrito como

$$\frac{C}{\lambda_j} + \nabla\varphi(X_{j+1}) - \nabla\varphi(X_j) = \mathcal{A}^* \left(\frac{z_j}{\lambda_j} \right), \quad \text{for } j = 0, 1, 2, \dots$$

Somando esta igualdade de $j = 0$ até k , definindo μ_k como em (4.27) e $v_k = \mu_k \sum_{j=0}^k \lambda_j^{-1} z_j$ obtemos

$$C + \mu_k (\nabla\varphi(X_k) - \nabla\varphi(X_0)) = \mathcal{A}^* v_k,$$

para todo $k \geq 0$. Assim, esta igualdade e (4.23) implicam que $X_{k+1} = X(\mu_k)$, para todo $k \geq 0$, e a primeira parte está provada.

Para a segunda parte, seja $\{X(\mu) : \mu > 0\}$ a trajetória central primal e seja $\{y(\mu) : \mu > 0\}$ dado por (4.23). Consideremos uma seqüência decrescente positiva $\{\mu_k\}$ e definamos as seqüências $X_{k+1} = X(\mu_k)$ e $y_{k+1} = y(\mu_k)$. Segue de (4.23) que

$$C + \mu_{k-1} (\nabla\varphi(X_k) - \nabla\varphi(X_0)) = \mathcal{A}^* (y_k), \quad C + \mu_k (\nabla\varphi(X_{k+1}) - \nabla\varphi(X_0)) = \mathcal{A}^* (y_{k+1}).$$

As duas equações anteriores implicam que

$$((\mu_k)^{-1} - (\mu_{k-1})^{-1})C + \nabla\varphi(X_{k+1}) - \nabla\varphi(X_k) = \mathcal{A}^* ((\mu_k)^{-1} y_{k+1} - (\mu_{k-1})^{-1} y_k).$$

Agora, definindo as seguintes quantidades

$$\lambda_k = ((\mu_k)^{-1} - (\mu_{k-1})^{-1})^{-1}, \quad z_k = \lambda_k ((\mu_k)^{-1} y_{k+1} - (\mu_{k-1})^{-1} y_k),$$

é substituindo na igualdade acima obtemos que

$$C + \lambda_k (\nabla\varphi(X_{k+1}) - \nabla\varphi(X_k)) = \mathcal{A}^* z_k.$$

Portanto, pelo Lema 4.3.1 temos que $\lambda_k = ((\mu_k)^{-1} - (\mu_{k-1})^{-1})^{-1}$ satisfaz (4.25), tem-se em consequência da última igualdade e de (4.26) que $X_{k+1} = X(\mu_k)$. Finalmente, se φ satisfaz a hipótese **A3** já havíamos provado que $X_{k+1} = X(\mu_k)$. Como $\lim_{k \rightarrow +\infty} \mu_k = 0$, usamos o fato que $\lim_{\mu \rightarrow 0} X(\mu) = X^c$ (veja Teorema 4.3.2) para concluir que $\lim_{\mu \rightarrow 0} X_k = X^c$, e a prova está concluída. \square

É importante ressaltar que o Teorema 4.3.3 é uma extensão natural para programação semidefinida do Teorema 3 de Iusem et al. [29].

Lema 4.3.2. *Suponhamos que φ satisfaz a hipótese **A3** ou **A4**. Seja $\{X(\mu) : \mu > 0\}$ a trajetória central primal para o Problema (P) com respeito a $D_\varphi(\cdot, X_0)$. Se $\nabla\varphi(X_0) \in \text{Im } \mathcal{A}^*$ então $\{X(\mu) : \mu > 0\}$ também é a trajetória central primal para o Problema (P) com respeito a φ .*

Demonstração. Suponha que φ satisfaça a hipótese **A3** ou **A4**. Da Observação 4.3.1 temos que $D_\varphi(\cdot, X_0)$, também satisfaz a hipótese **A3** ou **A4**. Assim, do Teorema 4.3.1 a trajetória central primal $\{X(\mu) : \mu > 0\}$ está bem definida e verifica

$$C + \mu(\nabla\varphi(X(\mu)) - \nabla\varphi(X_0)) = \mathcal{A}^*(y(\mu)),$$

para alguma trajetória $\{y(\mu) : \mu > 0\}$ em \mathbb{R}^m . Desde que $\nabla\varphi(X_0) \in \text{Im } \mathcal{A}^*$, existe $y_0 \in \mathbb{R}^m$ tal que $\mathcal{A}^*y_0 = \nabla\varphi(X_0)$, dessa forma, combinando as duas últimas igualdades, obtemos

$$C + \mu\nabla\varphi(X(\mu)) = \mathcal{A}^*(y(\mu) + \mu y_0).$$

Assim, $\{X(\mu) : \mu > 0\}$ é uma solução para (3.15), no entanto (3.15) tem uma única solução, a saber, a trajetória central primal para o Problema (P) induzida por φ . \square

Lema 4.3.3. *Seja $\{X(\mu) : \mu > 0\}$ a trajetória central primal para o Problema (P) com respeito a $D_\varphi(\cdot, X_0)$, onde $\varphi(X) = -\ln \det(X)$. Se $X_0^{-1} \in \text{Im } \mathcal{A}^*$, então $X(\mu)$ converge quando μ vai para 0.*

Demonstração. No Exemplo 3.5.1 *ii*) mostramos que $\varphi(X) = -\ln \det(X)$ satisfaz a hipótese **A4**. Como $\nabla\varphi(X_0) = -X_0^{-1} \in \text{Im } \mathcal{A}^*$, a aplicação do Lema 4.3.2 nos leva à igualdade das trajetórias central primal geradas através de $D_\varphi(\cdot, X_0)$ e φ . A afirmação é então conseqüência do Teorema 3.6.2. \square

Teorema 4.3.4. *Seja $\{X(\mu) : \mu > 0\}$ a trajetória central primal com respeito a $D_\varphi(\cdot, X_0)$, onde $\varphi(X) = -\ln \det(X)$. Suponhamos que $X_0^{-1} \in \text{Im } \mathcal{A}^*$. Então a seqüência de ponto proximal $\{X_k\}$ com ponto inicial X_0 , converge.*

Demonstração. Se $\{X_k\}$ é a seqüência de ponto proximal com ponto inicial X_0 , então fixando μ_k , como em (4.27), obtemos do Teorema 4.3.3 que $X_{k+1} = X(\mu_k)$, para $k \geq 0$. Como $\lim_{k \rightarrow +\infty} \mu_k = 0$, o resultado provém do Lema 4.3.3. \square

Capítulo 5

Trajetoária de Cauchy na variedade Riemanniana

5.1 Introdução

Neste capítulo usamos o conceito da Hessiana de uma função para obtermos uma nova métrica para uma determinada variedade Riemanniana. Uma abordagem similar deste conceito pode ser observada no estudo de métodos de pontos interiores para otimização convexa devido a Renegar [51]. Provaremos que a trajetória central primal, para o Problema (P) , induzida por uma função estritamente convexa e de classe C^2 coincide com a trajetória de Cauchy na variedade Riemanniana munida com a métrica dada pela Hessiana desta função. Este resultado estende para o contexto da programação semidefinida o resultado correspondente para programação linear, veja Seção 4 de Iusem et al. [29].

Consideremos o conjunto das matrizes simétricas, $n \times n$, definidas positivas S_{++}^n com sua estrutura diferenciável usual e munido com a métrica Euclidiana $\langle \cdot, \cdot \rangle$. O espaço tangente a S_{++}^n em X é dado por $T_X S_{++}^n = \{Y - X; Y \in S^n\} = S^n$. Seja $\varphi : S_{++}^n \rightarrow \mathbb{R}$ estritamente convexa e de classe C^2 . Definamos uma nova métrica em S_{++}^n como

$$\langle U, V \rangle_{\nabla^2 \varphi(X)} = (U \bullet V)_{\nabla^2 \varphi(X)} = \nabla^2 \varphi(X) U \bullet V$$

de modo que $M := (S_{++}^n, \langle \cdot, \cdot \rangle_{\nabla^2 \varphi(X)})$ é agora uma variedade Riemanniana. A métrica de M induz uma aplicação $\psi \mapsto \text{grad } \psi$ que associa a cada $\psi \in C^1(S_{++}^n)$ seu *gradiente* $\text{grad } \psi \in S^n$ por

$$d\psi_X(V) = (\text{grad } \psi(X) \bullet V)_{\nabla^2 \varphi(X)},$$

onde $X \in S_{++}^n$ e $V \in S^n$. É fácil verificar que o *campo de vetor de gradiente* em M é dado por

$$\text{grad } \psi(X) = (\nabla^2 \varphi(X))^{-1} \nabla \psi(X),$$

onde $\nabla \psi(X)$ é campo de vetor de gradiente Euclidiano em X , i.e., $\nabla \psi(X)$ é o gradiente induzida pela métrica Euclidiana.

Seja $\mathcal{A} : S^n \rightarrow \mathbb{R}^m$ a aplicação linear definida em (2.1). Das hipóteses **A1** e **A2** decorre que o conjunto $\mathcal{F}^0(P) = \{X \in S^n : \mathcal{A}X = b, X \in S_{++}^n\}$ é uma subvariedade Riemanniana de M com a métrica induzida e espaço tangente em X dado por

$$T_X \mathcal{F}^0(P) = \{V \in S^n; \mathcal{A}V = 0\}.$$

A aplicação adjunta de \mathcal{A} , sob esta métrica de M é $(\nabla^2 \varphi(X))^{-1} \mathcal{A}^*$, onde $\mathcal{A}^* : \mathbb{R}^m \rightarrow S^n$ é a aplicação adjunta usual de \mathcal{A} . Neste caso, a projeção ortogonal $\Pi_X : S^n \rightarrow T_X \mathcal{F}^0(P)$ sempre relativamente à métrica de M é

$$\Pi_X = I - (\nabla^2 \varphi(X))^{-1} \mathcal{A}^* (\mathcal{A}(\nabla^2 \varphi(X))^{-1} \mathcal{A}^*)^{-1} \mathcal{A}.$$

O campo de vetor de gradiente para a função $\psi|_{\mathcal{F}^0(P)} : \mathcal{F}^0(P) \rightarrow \mathbb{R}$, sob a métrica de M , é dado por $\text{grad } \psi|_{\mathcal{F}^0(P)} = \Pi \text{grad } \psi$, i.e.,

$$\text{grad } \psi|_{\mathcal{F}^0(P)} = \left(I - (\nabla^2 \varphi(X))^{-1} \mathcal{A}^* (\mathcal{A}(\nabla^2 \varphi(X))^{-1} \mathcal{A}^*)^{-1} \mathcal{A} \right) (\nabla^2 \varphi(X))^{-1} \nabla \psi. \quad (5.1)$$

5.2 Trajetória Central Primal e Trajetória de Cauchy

A *trajetória de Cauchy* para a função $\psi|_{\mathcal{F}^0(P)}$, determinada pela métrica induzida por φ , é a curva diferenciável $Z : [0, \beta) \rightarrow \mathcal{F}^0(P)$ dada por

$$\begin{cases} Z'(t) = -\text{grad } \psi|_{\mathcal{F}^0(P)}(Z(t)) \\ Z(0) = Z_0, \end{cases} \quad (5.2)$$

para o ponto inicial Z_0 e algum $\beta > 0$.

Observação 5.2.1. *É bem conhecido que para cada $Z_0 \in \mathcal{F}^0(P)$, existe $\beta > 0$ tal que (5.2) tem uma única solução $Z(t)$ definida em $[0, \beta)$.*

Consideremos a seguinte parametrização da trajetória central primal $\{X(t) : t \geq 0\}$, onde

$$X(t) = \text{argmin}_{X \succ 0} \{tC \bullet X + \varphi(X) : \mathcal{A}X = b\}. \quad (5.3)$$

O próximo resultado estende para programação semidefinida o resultado similar em Iusem et al. [29] obtido para programação linear.

Teorema 5.2.1. *Seja $\{X(t) : t \geq 0\}$ a trajetória central primal definida em (5.3), e seja $\psi(X) = C \bullet X$. Se $Z_0 \in \mathcal{F}^0(P)$ satisfaz $\nabla\varphi(Z_0) = \mathcal{A}^*z_0$, para algum $z_0 \in \mathbb{R}^m$, então a trajetória central primal $\{X(t) : t \geq 0\}$ é uma solução para (5.2) com $X(0) = Z_0$.*

Demonstração. Primeiramente observe que das condições de otimalidade para (5.3) temos

$$tC + \nabla\varphi(X(t)) = \mathcal{A}^*y(t), \quad (5.4)$$

para todo $t \geq 0$ e algum $y(t) \in \mathbb{R}^m$. Assim, $\nabla\varphi(X(0)) = \mathcal{A}^*y(0)$. Como $\nabla\varphi(Z_0) = \mathcal{A}^*z_0$ e φ é estritamente convexa temos $X(0) = Z_0$ e $y(0) = z_0$. A seguir, derivemos em relação a t a expressão em (5.4), obtendo $C + \nabla^2\varphi(X(t))X'(t) = \mathcal{A}^*y'(t)$, ou

$$(\nabla^2\varphi(X(t)))^{-1}C + X'(t) = (\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*y'(t). \quad (5.5)$$

Aplicando \mathcal{A} nesta igualdade, temos

$$\mathcal{A}(\nabla^2\varphi(X(t)))^{-1}C + \mathcal{A}X'(t) = \mathcal{A}(\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*y'(t).$$

Como $X'(t) \in T_X\mathcal{F}^0(P)$, provém desta que

$$\mathcal{A}(\nabla^2\varphi(X(t)))^{-1}C = \mathcal{A}(\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*y'(t).$$

Observe agora que $\mathcal{A}(\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*$ é não-singular, logo decorre da última igualdade que

$$y'(t) = \left(\mathcal{A}(\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*\right)^{-1}\mathcal{A}(\nabla^2\varphi(X(t)))^{-1}C.$$

Esta expressão, substituída em (5.5), leva a

$$X'(t) = -\left(I - (\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*\left(\mathcal{A}(\nabla^2\varphi(X(t)))^{-1}\mathcal{A}^*\right)^{-1}\mathcal{A}\right)(\nabla^2\varphi(X(t)))^{-1}C.$$

Finalmente, como $\psi(X) = \langle C, X \rangle$ segue da última equação e de (5.1) que $X(t)$ satisfaz (5.2) e a afirmação do teorema está provada. \square

Como conseqüência do Teorema anterior apresentamos uma prova alternativa à Proposição 3.4.1(ii), para a limitação da trajetória central primal.

Corolário 5.2.1. *A trajetória central primal $\{X(t) : t \geq 0\}$ para o Problema (P) com ponto inicial $X_0 \in \mathcal{F}^0(P)$, onde X_0 satisfaz $\nabla\varphi(X_0) = \mathcal{A}^*y_0$ para algum $y_0 \in \mathbb{R}^m$, é limitada.*

Demonstração. Seja $\psi(X) = C \bullet X$ e definamos $\xi(t) = \psi(X(t))$. Segue do Teorema 5.2.1 que

$$\begin{aligned}
\xi'(t) &= \left(\text{grad } \psi|_{\mathcal{F}^0(P)}(X(t)) \bullet X' \right)_{\nabla^2 \varphi(X)} \\
&= - \left(\text{grad } \psi|_{\mathcal{F}^0(P)}(X(t)) \bullet \text{grad } \psi|_{\mathcal{F}^0(P)}(X(t)) \right)_{\nabla^2 \varphi(X)} \\
&= - \left(\nabla^2 \varphi(X(t)) \text{grad } \psi|_{\mathcal{F}^0(P)}(X(t)) \bullet \text{grad } \psi|_{\mathcal{F}^0(P)}(X(t)) \right) \\
&= - \left\| (\nabla^2 \varphi(X(t)))^{1/2} \text{grad } \psi|_{\mathcal{F}^0(P)}(X(t)) \right\|^2 < 0, \forall t \in (0, +\infty).
\end{aligned}$$

Então ψ é decrescente ao longo da trajetória central primal, o qual implica que $\{X(t) : t > 0\} \subset \{X \in S_+^n; \psi(X) \leq \psi(X_0)\}$. Visto que o conjunto de soluções ótimas de (P) é compacto, segue da convexidade de ψ que $\{X \in S_+^n; \psi(X) \leq \psi(X_0), \mathcal{A}X = b\}$ também é compacto. Portanto, $\{X(t) : t > 0\}$ é limitado. \square

Capítulo 6

Considerações Finais

Nesta tese, estudamos a convergência das trajetórias central primal e dual associadas às funções entropia e exponencial, respectivamente, para problemas PSD. Descrevemos ainda um estudo exclusivo sobre as propriedades e convergência da trajetória central primal associada a uma função satisfazendo algumas hipóteses específicas. Desde que esta função seja contínua na fronteira provamos que a trajetória central primal converge para o centro analítico do conjunto solução do Problema (P). Para uma classe de funções mais gerais, incluindo as funções apresentadas no Exemplo 3.5.1 iii), iv) a convergência e caracterização do ponto limite da trajetória central associada a elas é um problema em aberto. Vale ressaltar que se a trajetória central dual associada ao problema (P) é limitada e o sistema que determina a trajetória central primal-dual é analítico podemos provar a convergência da trajetória central primal-dual usando argumentos similares aos usados na prova do Teorema 3.4.2. Mas infelizmente, não conseguimos provar a veracidade daquele fato ou exibir um contra-exemplo.

Como aplicação do estudo das trajetórias central primal e dual, mostramos a convergência das seqüências proximal primal e dual ponderada associadas à distância Kullback-Leibler em PSD. Embora tenhamos obtido a convergência da seqüência dual ponderada, a convergência completa da seqüência proximal dual é um problema em aberto. Em adição, apresentamos algumas propriedades de convergência do método do ponto proximal generalizada associado a uma distância generalizada. Por outro lado, os resultados de convergência para o método do ponto proximal generalizado associado a uma classe de funções incluindo as funções apresentadas no Exemplo 3.5.1 iii),iv está em aberto.

Finalmente, provamos que a trajetória central primal coincide com a trajetória

de Cauchy na variedade Riemanniana.

Como pesquisa futura, propomos desenvolver o seguinte:

- (i) Provar a convergência e caracterização do ponto limite da trajetória central associada a uma classe de funções mais gerais, incluindo as funções apresentadas no Exemplo 3.5.1 iii),iv).
- ii) Mostrar a convergência completa da seqüência proximal dual.
- iii) Estudar o problema PSD no caso em que a função objetivo seja uma função convexa e diferenciável qualquer.
- iv) Estudar a trajetória central no contexto de variedades Riemanniannas relacionando-a com a trajetória de Cauchy na variedade produto $S_{++}^m \times S_{++}^m$.

Referências Bibliográficas

- [1] 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.
- [2] Alvarez, F., Bolte, J. and Brahic, O. "Hessian Riemannian Gradient Flows in Convex Programming", *SIAM Journal on Control and Optimization*, v. 43, n. 2, pp. 477-501, 2004.
- [3] Auslender, A. and Héctor Ramírez C. "Penalty and Barrier Methods for Convex Semidefinite Programming", *Mathematical Methods of Operations Research*, v. 63, pp. 195-219, 2006.
- [4] Auslender, A. and Teboulle, M. "Interior Gradient and Proximal Methods for Convex and Conic Optimization", *SIAM Journal on Optimization*, v. 16, n. 3, pp. 697-725, 2006.
- [5] Balogh, J., Csendes, T. and Rapcsák, T. "Some Global Optimization Problems on Stiefel Manifolds", *Journal of Global Optimization*, v. 30, n. 1, pp. 91-101, 2004.
- [6] Bolte, J., Daniilidis, A. and Lewis, A. "The Lojasiewicz Inequality for Non-smooth Subanalytic Functions with Applications to Subgradient Dynamical Systems", to appear in *SIAM Optimization*, 2005.
- [7] Cominetti, R. and San Martín, J. "Asymptotic Analysis of the Exponential Penalty Trajectory in Linear Programming", *Mathematical Programming*, v. 67, n. 2, pp. 169-187, 1994.
- [8] Censor, Y. and Zenios, S. "The Proximal Minimization Algorithm with D-Functions", *Journal of Optimization Theory and Applications*, v. 73, n. 3, pp. 451-464, 1992.

- [9] Chen, G. and Teboulle, M. "Convergence Analysis of a Proximal-Like Minimization Algorithm using Bregman Functions", *SIAM Journal on Optimization*, v. 3, n. 3, pp. 538-543, 1993.
- [10] da Cruz Neto, J.X., Ferreira, O. P., Iusem, A.N., et. al. "Dual convergence of the proximal point method with Bregman distances for linear programming", *Optimization Methods and Software*, England, v. Online, pp. 1-23, 2006.
- [11] da Cruz Neto, J.X., Ferreira, O.P. and Monteiro, R.D.C. "Asymptotic Behavior of the Central Path for a Special Class of Degenerate SDP Problems", *Mathematical Programming*, v. 103, n. 3, pp. 487-514, 2005.
- [12] da Cruz Neto, J. X., de Lima, L. L. and Oliveira, P. R. "Geodesic Algorithms in Riemannian Geometry", *Balkan Journal of Geometry and Its Applications*, v. 3, n. 2, pp. 89-100, 1998.
- [13] de Klerk, E. *Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications*, Applied Optimization, v. 65, Kluwer Academic Publishers, Dordrecht, 2002.
- [14] Dedieu, J-P., Malajovich, G. and Shub, Mike. "On the Curvature of the Central Path of Linear Programming Theory", *Foundations of computational mathematics*, New York, v. 5, n. 2, pp. 145-171, 2005.
- [15] do Carmo, M.P. *Geometria Riemanniana*. 2 ed. Rio de Janeiro, Instituto de Matemática Pura e Aplicada, Projeto Euclides, 1988.
- [16] Dieudonné, J.A. *Foundations of Modern Analysis*, New York, Academic Press, 1960.
- [17] Doljansky, M. and Teboulle, M. "An Interior Proximal Algorithm and the Exponential Multiplier Method for Semidefinite Programming", *SIAM Journal on Optimization*, v. 9, n. 1, pp. 1-13, 1998.
- [18] Eckstein, J. "Nonlinear Proximal Point Algorithm using Bregman Functions, with Applications to Convex Programming", *Mathematics of Operations Research*, v. 18, n. 1, pp. 202-226, 1993.
- [19] Ferreira, O. P. and Oliveira, P. R. "Proximal Point Algorithm on Riemannian Manifolds", *Optimization*, v. 51, n. 2, pp. 257-270, 2002.

- [20] Ferreira, O. P. and Oliveira, P. R. "Subgradient Algorithm on Riemannian Manifolds", *Journal of Optimization Theory and Applications*, v. 97, n. 1, pp. 93-104, 1998.
- [21] Graña Drummond, L.M. and Peterzil, H.Y. "The Central Path in Smooth Convex Semidefinite Programs", *Optimization* v. 51, n. 2, pp. 207-233, 2002.
- [22] Güler, O. "Limiting Behavior of Weighted Central Paths in Linear Programming", *Mathematical Programming*, v. 65, pp. 347-363, 1994.
- [23] Halická, M., de Klerk, E. and Roos, C. "Limiting Behavior of the Central Path in Semidefinite Optimization", *Optimization Methods Software*, v. 20, n. 1, pp. 99-113, 2005.
- [24] Halická, M., de Klerk, E. and Roos, C. "On the Convergence of the Central Path in Semidefinite Optimization", *SIAM Journal on Optimization* v. 12, n. 4, pp. 1090-1099, 2002.
- [25] Helmke, U. and Moore, J.B. *Optimization and Dynamical Systems*, New York, Springer Verlag, 1994.
- [26] Horn, R.A., and Johnson, C.R. *Topics in Matrix Analysis*, New York, Cambridge University Press, 1991.
- [27] Horn, R.A., and Johnson, C.R. *Matrix Analysis*, New York, Cambridge University Press, 1985.
- [28] Iusem, A.N., and Monteiro, R.D.C. "On Dual Convergence of the Generalized Proximal Point Method with Bregman Distances", *Mathematics of Operations Research*, v. 25, n. 4, pp. 606-624, 2000.
- [29] Iusem, A.N., Svaiter, B.F. and da Cruz Neto, J.X. "Central Paths, Generalized Proximal Point Methods and Cauchy Trajectories in Riemannian Manifolds", *SIAM Journal on Control and Optimization*, v. 37, n. 2, pp. 566-588, 1999.
- [30] Iusem, A.N. "On Some Properties of Generalized Proximal Point Methods for Variational Inequalities", *Journal of Optimization Theory and Applications*, v. 96, n. 2, pp. 337-362, 1998.

- [31] Jensen, D.L. and R.A.Polyak. "The Convergence of a Modified Barrier Method for Convex Programming", *IBM Journal of Research and Development*, v. 38, n. 3, pp. 307-321, 1994.
- [32] Karmarkar, N. "Riemannian Geometry Underlying Interior Point Methods for Linear Programming", In: *Mathematical Developments Arising from Linear Programming* (J.C. Lagarias and M.J. Todd, eds.) v. 114, *Contemporary Mathematics*, American Mathematical Society, Providence, RI, pp. 51-75, 1990.
- [33] Kojima, M., Meggido, N., Noma, T., et. al. "A Unified Approach to Interior Point Algorithms for Linear Complementarity Problems", In: *Computer Science*, v. 538, *Lecture Notes*, Springer Verlag, Berlin, Germany, 1991.
- [34] Kurdyka, K., Mostowski, T., and Parusinski, A. "Proof of the Gradient Conjecture of R. Thom", In: *Annals of Mathematics*, v. 152, n. 3, pp. 763-792, II Series, 2000.
- [35] Lima, E.L. *Análise Real*, v. 2, IMPA, Coleção Matemática Universitária, 2004.
- [36] Lojasiewicz, S., *Ensembles semi-analitiques*, I.H.E.S., Bures-sur-Yvette, 1965.
- [37] Luo, Z-Q., Sturm, J,F., Zhang, S. "Superlinear Convergence of a Symmetric Primal-Dual Path Following Algorithm for Semidefinite Programming", *SIAM Journal on Optimization*, v. 8, n. 1, pp. 59-81, 1998.
- [38] Luo, Z-Q., Sturm, J,F. and Zhang, S. "Conic Convex Programming and Self Embedding", *Optimization Methods and Software*, v. 14, pp. 169-218, 2000.
- [39] Marshall, A.W. and Olkin, I. *Inequalities: Theory of Majorization and its Applications*, New York, Academic Press, 1979.
- [40] Mosheyev, L. and Zibulevski, M. "Penalty/Barrier Multiplier Algorithm for Semidefinite Programming", *Optimization Methods and Software*, v. 13, pp. 235-261, 2000.
- [41] Nesterov, Yu. and Nemirovski, A., *Central Paths and Riemannian Distances*, In: Technical report, CORE Discussion Paper, CORE, Université Catholique de Louvain, Belgium, 2003.

- [42] Nesterov, Yu. and Nemirovski, A. *Interior Point Polynomial Methods in Convex Programming*, SIAM, Philadelphia, PA, 1994.
- [43] Nesterov, Y. and Todd, M.J. "On the Riemannian Geometry Defined by Self-Concordant Barriers and Interior-Point Methods", *Foundations of Computational Mathematics*, v. 2, n. 4, pp. 333-361, 2002.
- [44] Papa Quiroz, E. A. and Roberto Oliveira, P. "A New Barrier for a Class of Semidefinite Problems", *RAIRO-Operations Research*, to appear.
- [45] Polyak, R., and Teboulle, M. "Nonlinear Rescaling and Proximal-Like Methods in Convex Optimization", *Mathematical Programming*, v. 76, n. 2, pp. 265-284, 1997.
- [46] Powell, M.J.D. "Some Convergence Properties of the Modified Log Barrier Method for Linear Programming", *SIAM Journal on Optimization*, v. 5, n. 4, pp. 695-739, 1995.
- [47] Prasolov, V. V. "Problems and Theorems in Linear Algebra", In: *Translations of Mathematical Monographs* v. 134, AMS, Providence, Rhode Island, 1994.
- [48] Rapcsák, T. *Smooth Nonlinear Optimization in R^n* , Dordrecht, Holland, Kluwer Academic Publishers, 1997.
- [49] Rapcsák, T. and Thang, T. T. "A Class of Polynomial Variable-Metric Algorithms for Linear Optimization", *Mathematical Programming*, v. 74, pp. 319-331, 1996.
- [50] Rätsch, G., Tsuda, K. and Warmuth, M.K. "Matrix Exponential Gradient Updates for On-line Learning and Bregman Projection", *Journal of Machine Learning Research*, v. 6, pp. 995-1018, 2005.
- [51] Renegar. J. *A Mathematical View of Interior-Point Methods in Convex Optimization*, Philadelphia, MPS-SIAM Series on Optimization, 2001.
- [52] Rockafellar, R.T. "Monotone Operators and the Proximal Point Algorithm", *SIAM Journal on Control and Optimization*, v. 14, n. 5, pp. 877-898, 1976.
- [53] Shiota, Masahiro. *Geometry of Subanalytic and Semialgebraic Sets*, Boston, Birkhäuser, 1997.

- [54] Smith, S.T. "Optimization Techniques on Riemannian Manifolds", In: *Fields Institute Communications*, v. 3, *American Mathematical Society*, Providence, Rhode Island, pp. 113-136, 1994.
- [55] Sporre, G., Forsgren, A. *Characterization of the Limit Point of the Central Path in Semidefinite Programming*, In: Technical Report TRITA-MAT-2002-OS12, Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, June 2002.
- [56] Tseng, P. and Bertsekas, D.P. "On the Convergence of the Exponential Multiplier Method for Convex Programming", *Mathematical Programming*, v. 60, n. 1, pp. 1-19, 1993.
- [57] Todd, M.J. "Semidefinite Optimization", *Acta Numerica*, v. 10, pp. 515-560, 2001.
- [58] Udriste, C. *Convex Functions and Optimization Methods on Riemannian Manifolds*, Dordrecht, Holland, Kluwer Academic Publishers, 1994.
- [59] Vandenberghe, L. and Boyd, S. "Semidefinite Programming", *SIAM Rev.* v. 38, pp. 49-95, 1996.