Método proximal com fatorações de Schur para determinação da média riemanniana de matrizes simétricas definidas positivas[1]

Ronaldo Gregório
Universidade Federal Rural do Rio de Janeiro
Departamento de Tecnologia e Linguagens
Av. Gov. Roberto Silveira, s/n, Moquetá, Nova Iguaçu - CEP 26020-740, RJ, Brazil
rgregor@ufrrj.br

Paulo Roberto Oliveira
Universidade Federal do Rio de Janeiro
Departamento de Engenharia de Sistemas e Computação, Caixa Postal 68511
CEP 21945-970, Rio de Janeiro, Brazil
poliveir@cos.ufrj.br

Resumo

Neste trabalho apresentamos um método proximal inexato para determinar uma ϵ-solução para o problema da média riemanniana de matrizes simétricas definidas positivas. A proposta consiste em decompor o domínio do problema no produto cartesiano do conjunto das matrizes diagonais definidas positivas com o grupo ortogonal. A viabilidade da decomposição é assegurada pelo Teorema de Schur aplicado a matrizes simétricas definidas positivas. A aproximação para solução da iteração principal do método é calculada em duas etapas repetidas iterativamente. Primeiro, resolvemos um problema de programação não-linear, riemanniano, no conjunto das matrizes diagonais definidas positivas e depois, um problema de programação não-linear, riemanniano, no grupo ortogonal. Nossa proposta pertence a classe de métodos denominados *preditores-corretores*.

Palavras Chave: Média riemanniana, matriz definida positiva, método proximal.

Área principal: Programação matemática.

Abstract

In this work we present an inexact proximal point algorithm, based in the proximal point methodology on Hadamard manifolds, to determine an ϵ-solution for the riemannian mean problem on the cone of symmetric definite positive matrices. The method decomposes the domain of the problem in the cartesian product of the cone of diagonal positive definite matrices with the orthogonal group. The viability of the method is assured by the Schur factorization for symmetric definite positive matrices. The approched solution of the main iteration of our inexact proximal method is computed in two steps repeated iteractively. First, we compute the approched global solution of a riemannian optimization problem on the cone of diagonal definite positive matrices and after, an approched local (global) solution of a riemannian optimization problem on the orthogonal group. Our method belongs on the class of the *predictor–corrector methods*.

Keywords: Riemannian mean, definite positive matrices, proximal point method.

Main area: Mathematical programming.

[1]Projeto de pesquisa *Algoritmo de ponto proximal com decomposições de Schur em domínios de positividade*, edital APQ1 2011/2 (auxílio financeiro). Projeto subsidiado pela FAPERJ.

1 Introdução

A otimização em variedades riemannianas tem sua origem no estudo de problemas de programação não-linear com restrições de igualdade. Em geral, um problema não-linear com restrições de igualdade em um espaço euclideano de dimensão n é dado por

$$\begin{array}{rl} \min & f(x) \\ \text{s. a} & h_i(x) = 0,\ i = 1, \cdots, m \\ & x \in \mathbb{R}^n, \end{array}$$

onde $f, h_i : \mathbb{R}^n \longrightarrow \mathbb{R},\ i = 1, \cdots, m$, são funções não-lineares. Denote por $H : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ a aplicação definida por $H(x) = (h_1(x), \cdots, h_m(x))^T$. Dado $c \in \mathrm{Im(H)} \subset \mathbb{R}^m$, o *teorema da função implícita* garante que se H é de classe C^k e os gradientes $\nabla h_i, i = 1, \cdots, m$, das funções coordenadas de H são linearmente independentes na imagem inversa de c, representada por $[H(c)]^{-1}$ (condição de regularidade de c), então $[H(c)]^{-1}$ é uma hipersuperfície de classe C^k e dimensão $n - m$. Além disso, para cada ponto $x \in [H(c)]^{-1}$, o hiperplano tangente à hipersuperfície $[H(c)]^{-1}$, em x, é o núcleo da transformação $H'(x) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

As condições necessárias e suficientes de otimalidade para problemas com restrições de igualdade se dão no plano tangente à hipersuperfície gerada pelas restrições. Segundo Luenberger (2010), em face da hipótese de regularidade das restrições, o hiperplano tangente à hipersuperfície $[H(0)]^{-1}$, em um ponto $x^* \in [H(0)]^{-1}$, é caracterizado como o complemento ortogonal do espaço gerado pelos gradientes das restrições calculados em x^*, e x^* é uma solução local para o problema não-linear com restrições de igualdade se, e somente se o gradiente da função objetivo f, calculado em x^* ($\nabla f(x^*)$), é ortogonal ao hiperplano tangente à $[H(0)]^{-1}$, em x^*, ou equivalentemente, $\nabla f(x^*)$ pertence ao espaço gerado pelos gradientes das restrições, calculados em x^*.

Pesquisadores, tal como Smith(1994), Ferreira and Oliveira (1998) e Da Cruz Neto et al (1998) percebendo que hipersuperfícies diferenciáveis definidas por restrições de igualdade representam casos particulares de variedades Riemannianas, propõem extensões de algoritmos clássicos de programação não-linear, tais como gradiente, Newton e gradientes conjugados.

Extensões visando implementações práticas também foram propostas. Dentre elas, podemos destacar o método de direções descentes com busca linear de Armijo, denominado *método de Armijo generalizado* em Yang (1999) e Papa Quiroz et al(2008).

No entanto, com a evolução da otimização surgiram problemas cujo domínio apresenta uma estrutura riemanniana bem definida, sendo eles não necessariamente gerados por restrições de igualdade. Dentre tais problemas, podemos destacar os com restrições de ortogonalidade, em Nishimori and Akaho (2005) e a média riemanniana de matrizes simétricas definidas positivas, em Moakher (2005).

Este último, em particular, tem sua origem histórica em 1960, com a introdução, por Rothaus (1960), do conceito de *domínio de positividade* que estende o cone das matrizes simétricas definidas positivas e suas propriedades. Vale destacar que Rothaus (1960) estabelece que, para escolhas adequadas de funções barreiras, todo domínio de positividade (munido da métrica definida pela hessiana da barreira) é uma variedade de Hadamard, ou seja, uma variedade riemanniana de curvatura seccional não-positiva.

Denote por S^n_{++} o cone das matrizes simétricas definidas positivas. Quando munido da métrica definida pela hessiana da função barreira logarítmica

$$F(X) = -\ln \det(X), \tag{1}$$

S^n_{++} representa um exemplo de variedade de Hadamard.

O problema da *média riemanninana* em S^n_{++}, segundo Moakher (2005), consiste em

$$\begin{array}{rl} \text{minimizar} & \dfrac{1}{2}\displaystyle\sum_{i=1}^{m} d^2(X^i, Y), \\ \text{sujeito a} & Y \in S^n_{++}, \end{array} \tag{2}$$

onde $X^i \in S^n_{++}, \;\; i = 1, \cdots, m$, são dados de entrada do problema. A distancia riemanniana d em S^n_{++}, com respeito à métrica definida pela hessiana de (1), apresentada em Rothaus (1960), é dada por

$$d(A, B) = \sqrt{\sum_{l=1}^{n} \ln^2 \, \mu_l(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})}, \tag{3}$$

onde $A, B \in S^n_{++}$ e $\mu_l(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})$ é o l–ésimo autovalor de $A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$, $l = 1, \cdots, n$.

De acordo com Fiori (2009), a média riemanniana de matrizes simétricas definidas positivas pode ser aplicada em áreas como Inteligência Artificial e Cognição, com destaque em análise de deformações, análise de imagens, análise estatística de difusão de dados tensoriais em medicina, controle automático e cognitivo, detecção humana via classificação no espaço de matrizes simétricas definidas positivas, reconhecimento de padrões, modelagem de evolução cognitiva, modelagem de funções cerebrais através de imagens, dentre outras áreas de pesquisa.

A hipótese de convexidade geodésica em variedades riemannianas, retratada com detalhes em Sakai (1996), implica que soluções locais para problemas de otimização são também soluções globais. Em variedades de Hadamard, em particular, para cada $X \in S^n_{++}$, tem-se assegurada a convexidade geodésica estrita da função $f_X : S^n_{++} \longrightarrow \mathbb{R}$, dada por

$$f_X(Y) = \frac{1}{2} d^2(X, Y). \tag{4}$$

Isso implica na unicidade de solução global para o problema da média riemanniana, visto que a adição de funções geodesicamente convexas estritas é geodesicamente convexa estrita. A existência de minimizador para (4) é consequência de sua coercividade. De fato, de acordo com a definição 4.1 em Ferreira e Oliveira (2002), uma função com valores reais h, definida em uma variedade de Hadamard M, é dita coerciva se

$$\lim_{d(x,y) \to +\infty} \frac{h(y)}{d(x, y)} = +\infty. \tag{5}$$

Vê-se claramente que f_X satisfaz a condição (5).

Paralelamente, o *método de ponto proximal*, introduzido por Martinet (1970) e estendido para determinação de zeros de operadores monótonos, por Rockafellar (1976), em suas diferentes propostas de regularização (quadrática, Bregman, φ-divergência em, por exemplo, Iusem (1995) e quase-distâncias, em Moreno et al (2011) apresenta-se como uma das mais elegantes técnicas de otimização do ponto de vista teórico, com fundamentação matemática sólida e consistente.

Uma versão do método proximal para variedades de Hadamard é apresentada por Ferreira and Oliveira (2002). Fundamentado nesse último artigo, destaca-se o algoritmo de ponto proximal com fatorações de Schur para matrizes simétricas definidas positivas apresentado por Gregório and Oliveira (2009). O foco desse algoritmo está voltado para problemas geodesicamente convexos e irrestritos em S^n_{++}, com respeito à métrica dada pela hessiana de (1).

A metodologia do algoritmo de ponto proximal com fatorações de Schur consiste em subdividir o espaço de busca (S^n_{++}) em dois espaços, a saber, o subespaço das matrizes diagonais definidas positivas e o grupo ortogonal, ambos vistos sob a ótica riemanianna. O primeiro é isomorfo ao octante positivo de um espaço euclideano de dimensão n (R^n_{++}) e a métrica em vigor resulta da restrição da métrica definida pela hessiana de (1) ao conjunto das matrizes diagonais definidas positivas.

O algoritmo de ponto proximal com fatorações de Schur possui estrutura semelhante à dos métodos preditores-corretores, onde, primeiramente, fixa-se uma matriz ortogonal e determina-se a solução do subproblema definido no conjunto das matrizes diagonais definidas positivas e em seguida, fixa-se a solução obtida e dá-se um passo corretor no grupo ortogonal.

A viabilidade do método é garantida pelo *teorema de Schur* para matrizes simétricas e o *teorema espectral*, para matrizes simétricas definidas positivas (ver Horn and Johnson (2006)). Combinados, eles asseguram que toda matriz simétrica definida positiva X poder ser escrita sob a forma $Q\Lambda Q^T$, com Q ortogonal, Λ diagonal e os elementos da diagonal de Λ estritamente positivos.

Contudo, em Gregório and Oliveira (2009) não foi apresentada proposta algorítmica para resolução dos subproblemas definidos no conjunto das matrizes diagonais definidas positivas. Além disso, pode-se observar que a proposta de atualização para matrizes ortogonais não leva em consideração informações da função objetivo do problema regularizado. Estas constatações derivam do fato do algoritmo ter sido desenvolvido para resolver problemas irrestritos que, teoricamente, atendem à hipótese de convexidade geodésica em S^n_{++}. Dessa forma, o algoritmo pode ser aplicado a uma classe ampla de problemas e não apenas a um problema em particular.

Por outro lado, o método proximal descrito em Gregório and Oliveira(2009) apresenta a vantagem de possuir uma versão inexata, com convergência assegurada, o que viabiliza sua implementação e aplicação a problemas reais.

Neste trabalho propomos a implementação da versão inexata do algoritmo introduzido por Gregório and Oliveira (2009) para obter ϵ-soluções para o problema da média riemanniana em S^n_{++}. Baseando-se em informações da função objetivo do problema da média riemanniana, aplicamos metodologias iterativas para computar as soluções dos subproblemas definidos no conjunto das matrizes diagonais definidas positivas, além das atualizações para as matrizes ortogonais, fundamentadas no método de Armijo generalizado apresentado por Yang (1999).

Simulações numéricas com matrizes simétricas definidas positivas, geradas aleatoriamente são apresentadas ao final.

2 Preliminares

2.1 O cone das matrizes simétricas definidas positivas

Matrizes simétricas definidas positivas aparecem naturalmente na descrição de várias teorias matemáticas, dentre as quais podemos citar a caracterização de minimizadores locais para funções diferenciáveis, a várias variáveis. Com efeito, seja $f : R^n \to R$ uma função duas vezes diferenciável. Um ponto crítico $x^* \in R^n$ de f é um minimizador local se, e somente se a hessiana de f é definida positiva em x^*. Outros exemplos de aplicação desse conceito podem ser encontrados em Horn and Johnson (2006).

Definição 2.1.1 *Seja $S \in R^{n\times n}$. Dizemos que S é simétrica se $S^T = S$.*

Um dos resultados mais relevantes para este trabalho é apresentado a seguir.

Teorema 2.1.1 (Teorema de Schur caso simétrico) *Seja $S \in R^{n\times n}$, tal que $S^T = S$. Então existem $Q, \Lambda \in R^{n\times n}$, com $Q^TQ = QQ^T = I$, e Λ, diagonal, tal que $S = Q\Lambda Q^T$.*

Prova: Teorema 8.1.1 em Golub and Van Loan (1996). ■

Na demonstração do Teorema de Schur, Q é a matriz cujas colunas são os autovetores de S e Λ, a matriz diagonal cujos elementos da diagonal são seus autovalores.

Definição 2.1.2 *Seja $S \in R^{n\times n}$, tal que $S^T = S$. S é dita semidefinida (definida) positiva se $x^TSx \geq (>)0$, para todo $x \in R^n$ $(x \neq 0)$. Se $x^TSx \leq (<)0$, para todo $x \in R^n$ $(x \neq 0)$, então S é dita semidefinida (definida) negativa.*

Decorre do Teorema de Schur que se S é simétrica então $x^TSx = x^T\left(Q\Lambda Q^T\right)x = y^T\Lambda y = \sum_{i=1}^{n}\lambda_{ii}y_i^2$, onde $y = Q^Tx$. Isto implica que os conjuntos das matrizes simétricas semidefinidas e definidas positivas, denotados, respectivamente, por S_+^n e S_{++}^n, assumem as seguintes caracterizações

$$S_+^n = \{S \in R^{n\times n} | \lambda_i(S) \geq 0, 1 \leq i \leq n\}, \quad S_{++}^n = \{S \in R^{n\times n} | \lambda_i(S) > 0, 1 \leq i \leq n\}. \quad (6)$$

Conclui-se, de (6), que a estrutura topológica de S_{++}^n é semelhante a do octante positivo de R^n. S_{++}^n é um conjunto aberto cujo fecho é S_+^n.

Equivalentemente a (6), o Teorema 7.2.5 em Horn and Johnson (2006) estabelece que uma matriz S é definida positiva se, e somente se os determinantes das submatrizes menores principais S_i, $1 \leq i \leq n$, de S são todos positivos. Além disso, o conjunto das matrizes simétricas definidas positivas é um cone convexo aberto.

Seja $S = Q\Lambda Q^T \in S_{++}^n$. Se f é uma função real, analítica em cada vizinhança contendo um autovalor λ de S então $f(S) = Qf(\Lambda)Q^T$, onde $f(\Lambda)$ é a matriz diagonal, com elementos da diagonal da forma $[f(\Lambda)]_{ii} = f(\lambda_{ii})$. Isso simplifica o cômputo de funções matriciais conhecidas, como por exemplo, $e^S = Qe^{\Lambda}Q^T$ e $\ln(S) = Q\ln(\Lambda)Q^T$. Maiores detalhes sobre funções matriciais podem ser encontrados em Golub and Van Loan (1996), capítulo 11.

2.2 A estrutura riemanniana de S_{++}^n

O conjunto das matrizes simétricas definidas positivas S_{++}^n contém em seu corpo estruturas matemáticas importantes, como por exemplo o conceito de variedade de Hadamard. De fato, Seja $S(t)$, $t \in R$ uma curva diferenciável em S_{++}^n. Cada elemento s_{ij} de S é uma função diferenciável, com respeito a t, de maneira que $s_{ij}(t) = s_{ji}(t)$ para todo $i, j = 1, \cdots, n$ com $i \neq j$. Dessa forma $S'(t)$, para cada $t \in R$, é uma matriz simétrica. No entanto, $S'(t)$ não é necessariamente definida positiva.

Exemplo 2.2.1 *Seja $S : R \to S_{++}^n$, a curva definida por*

$$S(t) = \begin{bmatrix} t^2+1 & t^2 \\ t^2 & t^2+1 \end{bmatrix}.$$

Note que, para todo $t \in R$, $\det(S_1(t)) = \det(S_2(t)) = t^2 + 1 > 0$, ou seja, $S(t)$ é definida positiva. Por outro lado,

$$S'(t) = \begin{bmatrix} 2t & 2t \\ 2t & 2t \end{bmatrix}.$$

é simétrica, contudo, indefinida (semidefininida positiva para $t \geq 0$ e definida negativa para $t < 0$).

O espaço tangente a S^n_{++}, em um ponto S, representado em textos clássicos de geometria riemanniana por $T_S S^n_{++}$, é o conjunto das matrizes simétricas. Para efeito de simplificação, $T_S S^n_{++}$ será denotado por S^n. Note que o *fibrado tangente*, definido como $TS^n_{++} = \cup_{S \in S^n_{++}} T_S S^n_{++} = S^n$.

S^n é um subespaço vetorial de $R^{m\times n}$, com respeito as operações de adição de matrizes e multiplicação por escalar usuais. O produto escalar usual em $R^{m\times n}$ dado por $\langle A, B\rangle_F = \mathrm{Tr}\left(B^T A\right)$, onde $\mathrm{Tr}\,(X)$ representa o *traço* de X, definido como a soma dos elementos da diagonal principal da matriz quadrada X. O sub-índice F é uma homenagem ao matemático Ferdinand Georg Frobenius. A norma associada ao produto escalar usual é dada por $\|A\|_F = \sqrt{\mathrm{Tr}\,(A^T A)}$.

Como $\mathrm{Tr}\left(B^T A\right) = \mathrm{Tr}\left(A^T B\right)$, para quaisquer matrizes $A, B \in R^{m\times n}$, $\langle A, B\rangle_F = \mathrm{Tr}\,(AB)$, para quaisquer matrizes $A, B \in S^n$. A norma de Frobenius de uma matriz $A \in S^n$ é dada por $\|A\|_F = \sqrt{\mathrm{Tr}\,(A^2)}$.

De acordo com Lema 1.4 em Rothaus (1960), a heissiana de (1), em um ponto $X \in S^n_{++}$, é a transformação linear $F'' : S^n \to S^n$, que satisfaz $F''(X)H = X^{-1}HX^{-1}$. Note que, para $H \in S^n$, $H \neq 0$,

$$\begin{aligned}\langle F''(X)H, H\rangle_F &= \langle X^{-1}HX^{-1}, H\rangle_F = \mathrm{Tr}\left(X^{-1}HX^{-1}H\right)\\ &= \mathrm{Tr}\left([X^{-1}H]^2\right) = \|X^{-1}H\|_F^2 > 0.\end{aligned}$$

Donde se conclui que, para cada $X \in S^n_{++}$, $F''(X)$ define um produto interno em S^n. A notação $\langle H_1, H_2\rangle_X$ será empregada para denotar

$$\langle H_1, H_2\rangle_X = \langle X^{-1}H_1X^{-1}, H_2\rangle_F = \mathrm{Tr}\ \left\{X^{-1}H_1X^{-1}H_2\right\} \tag{7}$$

e a norma associada a (7) é dada por $\|H\|_X = \sqrt{\mathrm{Tr}\ \left\{\left(X^{-1}H\right)^2\right\}}$.

O Corolário 5.10 em Rothaus (1960) estabelece que S^n_{++}, munido da métrica definida por F'', é uma variedade riemanniana de curvatura seccional não-positiva ou variedade de Hadamard. O Teorema de Hadamard garante que as curvas que minimizam distâncias entre dois pontos quaisquer em S^n_{++}, denominadas *geodésicas*, são obtidas de maneira única.

A seção 2.3 em Moakher (2005) estabelece que a expressão da geodésica ξ, em S^n_{++}, que satisfaz $\xi(0) = X$, $\xi'(0) = V$, é dada por

$$\xi(t) = X^{\frac{1}{2}} e^{tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}} X^{\frac{1}{2}}, \quad t \in R. \tag{8}$$

O Teorema 6.1 em Nesterov and Todd (2002) apresenta ainda a expressão do segmento geodésico γ em S^n_{++}, satisfazendo $\gamma(0) = X, \gamma(1) = Y$,

$$\gamma(t) = X^{\frac{1}{2}} \left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)^t X^{\frac{1}{2}}, \quad t \in [0, 1], \tag{9}$$

e a expressão (3) resulta do comprimento de γ calculado com base no produto interno definido em (7).

Decorre imediatamente de (9) que se $\Lambda, \bar{\Lambda} \in S^n_{++}$ são diagonais então o segmento geodésico conectando Λ e $\bar{\Lambda}$ contém apenas matrizes diagonais. Em outras palavras, o conjunto das matrizes diagonais definidas positivas, denotado por Ω^n_{++}, é um subconjunto convexo de S^n_{++}.

Definição 2.2.1 *Seja $f : S^n_{++} \to R$ uma função de classe C^1. O gradiente natural de f, denotado por* grad $f : S^n_{++} \to S^n$*, é a aplicação que satisfaz* $\langle \mathrm{grad}\ f(S), \bar{S}\rangle_S = \langle \nabla f(S), \bar{S}\rangle_F$, $\quad \forall \bar{S} \in S^n$*, onde $\nabla f(S)$ é o gradiente euclidiano de f em S.*

A definição apresentada é uma adaptação da Definição 1.5 em Sakai (1996) a S^n_{++}. Como $\langle \mathrm{grad}\ f(S), \bar{S}\rangle_S = \langle S^{-1}\mathrm{grad}\ f(S)S^{-1}, \bar{S}\rangle_F$, segue imediatamente que

$$\mathrm{grad}\ f(S) = S\nabla f(S)S. \tag{10}$$

2.3 A estrutura riemanniana do grupo ortogonal

Definição 2.3.1 *Uma matriz* $Q \in R^{n\times n}$ *é dita ortogonal se* $Q^TQ = QQ^T = I$.

Denote por $\mathcal{O}_n$ o conjunto das matrizes ortogonais, de ordem n. Alguns fatos sobre $\mathcal{O}_n$ decorrem naturalmente da definição. Note que se $Q \in \mathcal{O}_n$ então suas colunas formam uma base ortonormal de R^n, $|\det\,(Q)| = 1$ e que $\mathcal{O}_n$ é um grupo, em relação a multiplicação de matrizes.

Por outro lado, a convergência e o limite de sequências de matrizes quadradas, de ordem n, pode ser interpretada como a convergência e o limite de n^2 sequências de números reais. O próximo resultado segue dessa observação.

Lema 2.3.1 $\mathcal{O}_n$ *é compacto.*

Prova: A demonstração será dividida em duas etapas: (i) $\mathcal{O}_n$ é limitado e (ii) $\mathcal{O}_n$ é fechado. (i) Seja $Q \in \mathcal{O}_n$. Note que $\|Q\|_F = \sqrt{\mathrm{Tr}\,(Q^TQ)} = \sqrt{\mathrm{Tr}\,(I)} = \sqrt{n}$. (ii) Seja $\{Q_k\}_{k\in N} \subset \mathcal{O}_n$. Sem perda de generalidade, assuma que $Q_k \to W$ (como $R^{n\times n}$ é um espaço euclidiano, o Teorema de Bolzano-Weierstrass garante que $\{Q_k\}_{k\in N}$ possui uma subsequência convergente. Dessa forma, passe a uma subsequência convergente, caso $\{Q_k\}_{k\in N}$ não seja). Verifica-se facilmente que $W \in \mathcal{O}_n$. De fato, $Q_k^TQ_k = I$ para todo $k \in N$. Então, $\lim\limits_{k\to+\infty} Q_k^TQ_k = W^TW = I$. A etapa (i) mostra que $\mathcal{O}_n$ é limitado e (ii), que é fechado. Como os subconjuntos limitados e fechados de espaços euclidianos são compactos, o lema segue. ■

A seção 3 em Nishimori e Akaho (2005) discorre sobre fatos acerca da estrutura riemanniana de $\mathcal{O}_n$. As mesmas considerações sobre $\mathcal{O}_n$ podem ser encontradas também na seção 2.2 de Fiori (2005). De fato, Seja $Q(t)$ uma curva diferenciável em $\mathcal{O}_n$, tal que $Q(0) = Q$. Segue da definição (2.3.1) que $Q(t)^TQ(t) = I$, $\forall t \in R$. Deriva desse fato que $\frac{dQ(t)}{dt}|_{t=0}^{T}Q(0) + Q(0)^T\frac{dQ(t)}{dt}|_{t=0} = (Q'(0))^TQ + Q^TQ'(0) = 0$, isto é, $T_Q\mathcal{O}_n = \{X \in R^{n\times n} | XQ + Q^TX = 0\}$. Além disso, o produto interno, em $T_Q\mathcal{O}_n$, coincide com o produto interno de Frobenius, as geodésicas são dadas, equivalentemente, pelas expressões $\Gamma(t) = \exp(tVQ^T)Q$ ou $\Gamma(t) = Q\exp(tQ^TV)$, com $\exp(tX) = \sum_{k=0}^{\infty}\frac{(tX)^k}{k!}$, e o gradiente natural de uma função $f : \mathcal{O}_n \to \mathbb{R}$, de classe C^1, em $T_Q\mathcal{O}_n$, pela expressão

$$\mathrm{grad}\ f(Q) = \frac{1}{2}(\nabla f(Q) - Q\nabla f(Q)^TQ), \tag{11}$$

onde $\nabla f(Q)$ é o gradiente euclidiano de f, em Q.

3 Algoritmo de ponto proximal com decomposições de Schur aplicado a média riemanniana em S^n_{++}

3.1 Algoritmo de ponto proximal aplicado à média riemanniana em S^n_{++}

Denote por f a função objetivo (2) do problema da média riemanniana de matrizes simétricas definidas positivas. Dados $X^0 \in S^n_{++}, \beta^0 > 0$, a iteração principal do algoritmo de ponto proximal proposto por Ferreira e Oliveira and Oliveira (2002) para determinação de um minimizador de f consiste em calcular X^{k+1}, definido por

$$X^{k+1} = \mathrm{argmin}\{f(Y) + \frac{1}{2}\beta^kd^2(X^k, Y)|Y \in S^n_{++}\}, \quad k = 0, 1, \cdots. \tag{12}$$

Note que f é dada pela soma de m funções f_i, da forma (4), com $X = X^i$. A regularização acrescentada na iteração principal (12) também possui a mesma forma de (4), com $X = X^k$. A igualdade 3.4 na demosntração da proposição 3.4 em Moakher (2005) estabelece que o gradiente euclidiano de (4) é dada por $\nabla f_X(Y) = \ln\left(X^{-1}Y\right)Y^{-1}$. Segue portanto, da relação (10), que grad $f_X(Y) = Y\ln\left(X^{-1}Y\right)Y^{-1}Y = Y\ln\left(X^{-1}Y\right)$.

Teorema 3.1.1 *A sequência $\{X^k\}_{k\in N}$ gerada por (12) está bem definida e é caracterizada por*

$$\sum_{i=1}^{m}\ln\left(X^{i^{-1}}X^{k+1}\right) = -\beta^k\ln\left(X^{k^{-1}}X^{k+1}\right). \tag{13}$$

Prova: A boa definição da sequência é garantida pela coercividade (condição (5)) e pela estrita convexidade geodésica (Teorema 3.3 em Ferreira e Oliveira (2002)) de $f(Y) + \frac{1}{2}\beta^k d^2(X^k, Y)$ e sua diferenciabilidade implica que X^{k+1} satisfaz $X^{k+1}\sum_{i=1}^{m}\ln\left(X^{i^{-1}}X^{k+1}\right) + \beta^k X^{k+1}\ln\left(X^{k^{-1}}X^{k+1}\right) = 0$. Pré-multiplicando a igualdade pela inversa de X^{k+1} o teorema segue. ∎

Teorema 3.1.2 *Seja $\{X^k\}_{k\in N}$ a sequência gerada por (12). Se $\{\beta^k\}_{k\in N}$ é tal que $\sum_{k=0}^{\infty}(1/\beta^k) = +\infty$, então $\lim_{k\to+\infty} X^k = X^*$, onde $X^* = \operatorname{argmin}\{f(X)|X \in S^n_{++}\}$.*

Prova: Teorema 6.1 em Ferreira e Oliveira (2002). ∎

3.2 Algoritmo de ponto proximal com decomposições de Schur inexato

Com base no Teorema 3.1.2, Gregório and Oliveira proporam uma metodologia, pertencente a classe de algoritmos preditores-corretores, denominado algoritmo de ponto proximal com decomposições de Schur, que determina X^{k+1}, a cada iteração k, realizando buscas alternadas no espaço de matrizes diagonais definidas positivas e no espaço de matrizes ortogonais, respectivamente. Uma das vantagens desse método é a presença de uma versão inexata, onde o iterado X^{k+1} é obtido de maneira aproximada.

Admita que k é a iteração corrente do algoritmo de ponto proximal. Dados $\Lambda_0^k \in \Omega^n_{++}, Q_0^k \in \mathcal{O}_n, \beta^0, \epsilon^0 > 0$, pela versão inexata do algoritmo de ponto proximal com decomposições de Schur, inicie $Y_0^k = \left(X^k\right)^{\frac{1}{2}} Q_0^k\Lambda_0^k Q_0^{k^T}\left(X^k\right)^{\frac{1}{2}}$ e $j = 0$. Se $\|\sum_{i=1}^{m}\ln\left(X^{i^{-1}}Y_j^k\right) + \beta^k\ln\left(X^{k^{-1}}Y_j^k\right)\|_F = 0$ então, pela caracterização (13), $Y_j^k = X^{k+1}$. Na versão inexata, essa condição é relaxada para $\|\sum_{i=1}^{m}\ln\left(X^{i^{-1}}Y_j^k\right) + \beta^k\ln\left(X^{k^{-1}}Y_j^k\right)\|_F < \epsilon^k$. Caso contrário, calcula-se $\bar{\Lambda}_{j+1}^k$ solução do problema

$$\min\left\{f(X^{k^{\frac{1}{2}}}Q_j^k\Lambda Q_j^{k^T}X^{k^{\frac{1}{2}}}) + \frac{1}{2}\beta^k f_{X^k}(X^{k^{\frac{1}{2}}}Q_j^k\Lambda Q_j^{k^T}X^{k^{\frac{1}{2}}})|\Lambda \in \Omega^n_{++}\right\}. \tag{14}$$

Para atualizar Q_j^k, determina-se $Q_{j+1}^k \in \mathcal{O}_n$ e $\Lambda_{j+1}^k \in \Omega^n_{++}$, tal que

$$\Lambda_{j+1}^k = Q_{j+1}^{k^T}[X^{k^{-\frac{1}{2}}}\left(X^{k^{\frac{1}{2}}}Q_j^k\bar{\Lambda}_{j+1}^k Q_j^{k^T}X^{k^{\frac{1}{2}}}\right)X^{k^{-\frac{1}{2}}}]Q_{j+1}^k. \tag{15}$$

O Lema 5 em Gregório and Oliveira (2009) garante que Λ^k_{j+1} e $\bar{\Lambda}^k_{j+1}$ são similares. Segue ainda que $Y^k_{j+1} = Q^k_{j+1}\Lambda^k_{j+1}Q^{k^T}_{j+1} = Q^k_j\bar{\Lambda}^k_{j+1}Q^{k^T}_j = \bar{Y}^k_{j+1}$. Isso mostra que a forma de atualização de Q^k_j como prosposta originalmente não altera o valor da função objetivo regularizada.

Teorema 3.2.1 *Seja $\{X^k\}_{k\in N}$ a sequência gerada por (12). Para cada $k \in N$, admita que X^{k+1} satisfaz*

$$\| \sum_{i=1}^{m} \ln\left(X^{i^{-1}}X^{k+1}\right) + \beta^k \ln\left(X^{k^{-1}}X^{k+1}\right) \|_F < \epsilon^k. \tag{16}$$

Se $\{\beta^k\}_{k\in N}$ e $\{\epsilon^k\}_{k\in N}$ são sequências de números reais positivos, tais que $\sum_{k=0}^{\infty}\frac{1}{\beta^k} = +\infty$, $\sum_{k=0}^{\infty}\epsilon^k < +\infty$ e $\sum_{k=0}^{\infty}\frac{\epsilon^k}{\beta^k} < +\infty$ então $\lim_{k\to+\infty} X^k = X^*$, onde $X^* = \operatorname{argmin}\{f(X)|X \in S^n_{++}\}$.

Prova: De fato, a condição (16) pode ser reescrita como

$$\sum_{i=1}^{m} \ln\left(X^{i^{-1}}X^{k+1}\right) = -\beta^k \ln\left(X^{k^{-1}}X^{k+1}\right) + \epsilon^k E,$$

onde $E = \frac{1}{\sqrt{n}}I$. Em outras palavras, $-\beta^k\ln\left(X^{k^{-1}}X^{k+1}\right)$ é um ϵ^k-gradiente de f em X^{k+1}. O resultado segue como caso particular do Teorema 2 em Gregório and Oliveira (2009), adaptado ao caso em que f é diferenciável. ■

Neste trabalho, é proposta uma nova atualização para Q^k_j que substitui a etapa (15). De fato, admiti-se como Q^k_{j+1} qualquer solução local (global) do problema

$$\min\ \{f(X^{k^{\frac{1}{2}}}Q\Lambda^k_{j+1}Q^TX^{k^{\frac{1}{2}}}) + \frac{1}{2}\beta^k f_{X^k}(X^{k^{\frac{1}{2}}}Q\Lambda^k_{j+1}Q^TX^{k^{\frac{1}{2}}})|Q \in \mathcal{O}_n\} \tag{17}$$

Lema 3.2.1 *Existe solução Q^k_{j+1} para o problema (17).*

Prova: Teorema de Weierstrass. ■

O Algoritmo de ponto proximal com decomposições de Schur aplicado ao problema da média riemanniana, com atualização das matrizes ortogonais Q^k_j dada por (17), pode ser escrito esquematicamente como segue abaixo.

Algoritmo 3.2.1 **Dados** *$\epsilon, \beta^0, \epsilon^0 > 0$ and $X^0 \succ 0$;*

1. $k \leftarrow 0$;

2. **Enquanto** $\| \sum_{i=1}^{m} \mathrm{Ln}\left(X^{i^{-1}}X^k\right) \|_F > \epsilon$

3. **Escolha** $\Lambda^k_0 \in \Omega^n_{++}$, $Q^k_0 \in \mathcal{O}_n$;

4. $Y^k_0 \leftarrow (X^k)^{\frac{1}{2}} Q^k_0\Lambda^k_0Q^{k^T}_0 (X^k)^{\frac{1}{2}}$;

5. $j \leftarrow 0$;

6. **Enquanto** $\| \sum_{i=1}^{m} \mathrm{Ln}\left(X^{i^{-1}}Y^k_j\right) + \beta^k\mathrm{Ln}\left(X^{k^{-1}}Y^k_j\right) \|_F > \epsilon^k$

7. **Calcule** Λ^k_{j+1} *em (14);*

8. **Calcule** Q^k_{j+1} *em (17);*

9. $\quad Y^k_{j+1} \leftarrow \left(X^k\right)^{\frac{1}{2}} Q^k_{j+1}\Lambda^k_{j+1}Q^{k^T}_{j+1}\left(X^k\right)^{\frac{1}{2}}$;
10. $\quad j \leftarrow j+1$;
11. $\quad$ **retorne** *ao passo 6;*
12. $\quad$ **Fim**;
13. $\quad X^{k+1} \leftarrow Y^k_j$;
14. $\quad$ **Atualize** β^k, ϵ^k;
15. $\quad k \leftarrow k+1$
16. $\quad$ **retorne** *ao passo 2;*
17. **Fim**.

3.3 Implementação e resultados computacionais

A função objetivo de (14) pode ser reescrita como $\phi(\lambda)+\beta\rho(\lambda)$, com $\lambda \in R^n, \lambda_l > 0,\ \ i = 1,\cdots,n$, onde $\phi(\lambda) = f(X^{k^{\frac{1}{2}}}Q^k_j\mathrm{diag}(\lambda)Q^{k^T}_j X^{k^{\frac{1}{2}}})$, $\rho(\lambda) = f_{X^k}(X^{k^{\frac{1}{2}}}Q^k_j\mathrm{diag}(\lambda)Q^{k^T}_j X^{k^{\frac{1}{2}}})$ e $\mathrm{diag}(\lambda)$, a matriz diagonal cujos elementos da diagonal são as componentes de λ, ou seja, $[\mathrm{diag}(\lambda)]_{ll} = \lambda_l,\ \ l = 1,\cdots,n$.

Como $f(Y) = \frac{1}{2}\sum_{i=1}^{m}\sum_{l=1}^{n}\ln^2\theta_l(X^{i^{-\frac{1}{2}}}YX^{i^{-\frac{1}{2}}})$, ϕ é dada explicitamente por

$$\phi(\lambda) = \frac{1}{2}\sum_{i=1}^{m}\sum_{l=1}^{n}\ln^2\theta_l(W^{k_{ij}}\mathrm{diag}(\lambda)W^{k^T_{ij}}), \tag{18}$$

onde $W^{k_{ij}} = X^{i^{-\frac{1}{2}}}X^{k^{\frac{1}{2}}}Q^k_j$. Como $W^{k^T_{ij}}$ é inversível, $W^{k_{ij}}\mathrm{diag}(\lambda)W^{k^T_{ij}}$ possui os mesmos autovalores de $W^{k^T_{ij}}W^{k_{ij}}\mathrm{diag}(\lambda)$. Mas, $A^{k_{ij}} = W^{k^T_{ij}}W^{k_{ij}} \in S^n_{++}$. Com efeito, dado $x \in R^n$, $x \neq 0$, $x^TA^{k_{ij}}x = x^T[W^{k^T_{ij}}W^{k_{ij}}]x = \|W^{k_{ij}}x\|^2_2 > 0$, uma vez que $W^{k_{ij}}$ é não-singular e $x \neq 0$. ϕ pode ser reescrita como

$$\phi(\lambda) = \frac{1}{2}\sum_{i=1}^{m}\sum_{l=1}^{n}\ln^2\theta_l(A^{k_{ij}}\mathrm{diag}(\lambda)) = \frac{1}{2}\sum_{i=1}^{m}\sum_{l=1}^{n}\ln^2\theta_l(A^{k^{\frac{1}{2}}_{ij}}\mathrm{diag}(\lambda)A^{k^{\frac{1}{2}}_{ij}}).$$

Das considerações anteriores, conclui-se que $\phi(\lambda) = \frac{1}{2}\sum_{i=1}^{m}d^2(A^{k^{-1}_{ij}},\mathrm{diag}(\lambda))$. Isso mostra que ϕ é geodesicamente convexa estrita e diferenciável. Do ponto de vista prático, uma boa aproximação $\widetilde{\mathrm{grad}}\ \phi$ pode ser computada calculando-se, primeiramente, de maneira aproximada as componentes de $\nabla\phi(\lambda)$ (por exemplo, segundo a seção 2.5 em Stoer and Burlish (2010), a i-ésima componente de $\nabla\phi(\lambda)$ pode ser calculada numericamente através da expressão $\dfrac{\phi(\lambda+he_i)-\phi(\lambda-he_i)}{h}$, onde e_i é o i-ésimo vetor da base canônica de R^n, com h suficientemente pequeno). O gradiente natural de (18), em λ, é obtido então através da relação (10).

O gradiente da função objetivo em (17) pode ser obtido, de maneira similar, calculando-se primeiro seu gradiente euclidiano e depois, aplicando a relação (11). Note que em (17)

$$f_{X^k}(X^{k^{\frac{1}{2}}}Q\Lambda^k_{j+1}Q^TX^{k^{\frac{1}{2}}}) =$$
$$= \sum_{l=1}^{n}\theta_l(X^{k^{-\frac{1}{2}}}X^{k^{\frac{1}{2}}}Q\Lambda^k_{j+1}Q^TX^{k^{\frac{1}{2}}}X^{k^{-\frac{1}{2}}}) = \sum_{l=1}^{n}\theta_l(Q\Lambda^k_{j+1}Q^T) = \sum_{l=1}^{n}\theta_l(\Lambda^k_{j+1}).$$

Isso mostra que $f_{X^k}(X^{k^{\frac{1}{2}}} Q \Lambda^k_{j+1} Q^T X^{k^{\frac{1}{2}}})$ não depende de Q. Q^k_{j+1} é definida como uma solução local do problema

$$\min \{ f(X^{k^{\frac{1}{2}}} Q \Lambda^k_{j+1} Q^T X^{k^{\frac{1}{2}}}) | Q \in \mathcal{O}_n \} \tag{19}$$

Denote por $\bar{X}$ as soluções aproximadas obtidas pelo método. A tabela 1 apresenta alguns resultados para simulações numéricas realizadas com conjuntos de matrizes simétricas definidas positivas geradas aleatoriamente. Para todas as simulações foram escolhidos $X^0 = \Lambda^k_0 = Q^k_0 = I$, $\beta^k = \beta$ e $\epsilon^k = 10^{-3} \cdot \tau^k$, $k = 0, 1, \cdots$, $\tau \in (0, 1)$.

n	β	$\Vert \sum_{i=1}^{m} \ln \left(X^{i^{-1}} X^0 \right) \Vert_F$	$\Vert \sum_{i=1}^{m} \ln \left(X^{i^{-1}} \bar{X} \right) \Vert_F$	k	$\bar{j}/k$
2	1	5.5227	1.9859e-004	3	11
3	1	6.4821	4.7244e-004	3	10.3
4	1	6.9875	6.4427e-004	3	15.7
7	0.5	15.3113	6.1639e-004	3	18.3
10	0.5	20.8471	6.0303e-004	3	24.3

Tabela 1: $m = 50$, $\tau = 0.8$, k (n^0. it. externas), $\bar{j}/k$ (média do n^0. it. internas/it. externa).

4 Considerações finais

Neste trabalho, apresentamos uma variante do método de ponto proximal com decomposições de Schur, apresentado por Gregório and Oliveira (2009), onde propomos uma nova atualização para as matrizes ortogonais que leva em consideração informações da função objetivo do problema da média riemanniana. Empregamos o método de Armijo generalizado, descrito em Yang (1999), para computar Λ^k_{j+1} e Q^k_{j+1}, a cada iteração interna do algoritmo, onde as direções de descida empregadas foram aquelas dadas pelos anti-gradientes naturais das funções objetivos em (14) e (19). Em todos os testes, a precisão $\epsilon = 10^{-3}$ foi utilizada como critério de parada para o algoritmo. Nas simulações apresentadas na tabela 1, foi possível observar que aumentos na dimensão das matrizes implicam também em aumento na dificuldade de se computar Λ^k_{j+1} e Q^k_{j+1} empregando do método de Armijo generalizado. Particularmente, para $n \geq 3$, quando a norma dos gradientes das funções (18) e (19), calculados segundo as expressões (10) e (11), respectivamente, é menor que 10^{-5}, os comprimentos dos passos determinados pela busca de Armijo sobre as geodésicas tornam-se relativamente pequenos e o método estaciona antes de obter as respectivas soluções Λ^k_{j+1} e Q^k_{j+1}. Para contornar essa dificuldade, diminuimos o valor absoluto de β que implica na redução do peso da regularização proximal. Como trabalho futuro, sugerimos a adaptação do método dos gradientes conjugados, discutido em Eldeman et al (1998), para computar Q^k_{j+1}. Além disso, propomos a análise de complexidade do método, sua aplicação a outros problemas geodesicamente convexos em S^n_{++} e a extensão a outros domínios de positividade.

Referências

[1] **Da Cruz Neto, J. X., De Lima, L. L., Oliveira, P. R.** (1998), Geodesic algorithms in Riemannian geometry, *Balkan J. Geom. Appl.*, v. 3, n. 2, pp. 89-100.

[2] **Eldeman, A., Arias, A. T., Smith, T. S.,** (1998), The geometry of algorithms with orthogonality constraints, *SIAM J. Matrix Anal. Appl.*, v. 20, n. 2, pp. 303-353.

[3] **Ferreira, O. P., Oliveira, P. R.** (1998), Subgradient algorithm on Riemannian manifolds, *J. Optim. Theory Appl.*, v. 97, n. 1, pp. 93-104.

[4] **Ferreira, O.P.and Oliveira, P.R.** (2002), Proximal point algorithm on Riemannian manifolds, *Optimization*, v. 51, n. 2, pp. 257-270.

[5] **Fletcher, P. T. and Joshi, S.** (2007), Riemannian geometry for the statistical analysis of diffusion tensor data, *Signal Processing*, v.1, n. 87, p. 250-262.

[6] **Fiori, S.** (2005), Quasi-geodesic neural learning algorithms over the orthogonal group: A tutorial, *Journal of Machine Learning Research*, v. 6, p. 743-781.

[7] **Fiori, S.** (2009), Learning the Fréchet mean over the manifold of symmetric positive-definite matrices, *Journal of Cognitive Computation*, v.1, n.4, p. 279-291.

[8] **Golub, G. H. and Van Loan, C.** (1996), Matrix Computations, 3^{th} edition, The Johns Hopkins University Press, USA.

[9] **Gregório, R. and Oliveira,P.R.** (2009), Proximal point algorithm with Schur decomposition on the cone of symmetric semidefinite positive matrices, *Journal of Mathematical Analysis and Applications*, v. 2, n. 355, pp. 469-478.

[10] **Horn, R.A. and Johnson, C.R.** (2006) , Matrix Analysis,. 1^{st} ed., Cambridge University Press, USA.

[11] **Iusem, A.** (1995), Métodos de ponto proximal em Otimização, In: 20^0 *Colóquio brasileiro de matemática, IMPA*, Rio de Janeiro, Brasil.

[12] **Luenberger ,D.G. and Ye, Y.** (2010), Linear and nonlinear programming, *Springer*, 3^{th} edition, New York.

[13] **Martinet, B.** (1970), Regularisation d'inéquations variationnelles par approximations successives, *Rev. Francaise Informat. Recherche Operationnelle*, vol. 4 , Ser. R-3, p. 154-158.

[14] **Moakher, M.** (2005), A differential geometry approach to the geometric mean of symmetric positive-definite matrices, *SIAM Journal of Matrix Analysis and Applications*, v.26, n.3, 735-747.

[15] **Moreno F.G., Oliveira, P.R. and Soubeyran, A.** (2011), A proximal Algorithm with Quasi Dis- tance. Application to Habit's Formation, *Optimization: A Journal of Mathematical Programming and Operations Research*, v. 1, p. 1-21.

[16] **Nesterov, Y. E. and Todd, M. J.** (2002), On the Riemannian geometry defined by self-concordant barriers and interior-point methods, *Foundations of Computational Mathematics*, v.2, n.4, p. 333-361.

[17] **Nishimori, Y. and Akaho, S.** (2005), Learning algorithms utilizing quasi–geodesic flows on the Stiefel manifold, *Journal of Neurocomputing*, Vol. 67, N.1, p. 106-135.

[18] **Papa Quiroz, E. A., Quispe, E. M., Oliveira, P. R.** (2008), Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds, *J. Math. Anal. Appl.*, v. 341, n. 1, pp. 467-477.

[19] **Rockafellar, R.T.** (1976), Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, v. 14, n. 5, pp. 877-898.

[20] **Rothaus, O. S.** (1960) , Domains of positivity, *Abh. Math. Sem. Univ.*, v.24, n.3, p. 189-235.

[21] **Sakai, T.** (1996), Riemannian geometry - translations of mathematical monographs, *American Mathematical Society*, v. 149, Providence, R.I.

[22] **Smith, S. T.** (1994), Optimization techniques on riemannian manifolds - Hamiltonian and gradient flows, algorithms and control, Fields Int. Commun., *Amer. Math Soc.*, , vol. 3, p. 113-136.

[23] **Stoer, J. and Bulirsch, R.** (2010), Introduction to numerical analysis, 3^{th} edition, Springer-Verlag, New York.

[24] **Yang, Y.** (1999), Optimization on riemannian manifolds, In: *Proceendings of the* 38^{th} *conference on decision & control*, Phoenix, Arizona, USA, p.888-893, .