



UMA MODIFICAÇÃO NO MÉTODO DO PONTO PROXIMAL PARA
FUNÇÕES DC EM VARIEDADES DE HADAMARD

Yldenilson Torres Almeida

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Rio de Janeiro
Junho de 2021

UMA MODIFICAÇÃO NO MÉTODO DO PONTO PROXIMAL PARA
FUNÇÕES DC EM VARIEDADES DE HADAMARD

Yldenilson Torres Almeida

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA 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.

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Aprovada por: Prof. Paulo Roberto Oliveira
Prof. João Xavier da Cruz Neto
Prof. Márcia Helena Costa Fampa
Prof. Glaydston de Carvalho Bento
Prof. João Carlos de Oliveira Souza
Prof. Erik Alex Papa Quiroz

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2021

Almeida, Yldenilson Torres

Uma modificação no método do ponto proximal para funções DC em variedades de Hadamard/Yldenilson Torres Almeida. – Rio de Janeiro: UFRJ/COPPE, 2021.

XI, 58 p.: il.; 29,7cm.

Orientadores: Paulo Roberto Oliveira

João Xavier da Cruz Neto

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2021.

Referências Bibliográficas: p. 53 – 58.

1. Método ponto proximal. 2. Função DC. 3. Variedades de Hadamard. I. Oliveira, Paulo Roberto *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Ao meu filho, João Lucca.

Agradecimentos

Agradeço primeiramente a DEUS, pelo dom da vida e por ter me dado a capacidade (e a oportunidade) de chegar até aqui.

Agradeço aos meus pais, que sempre me apoiaram e lutaram para que eu chegasse até aqui. Nunca mediram esforços para isso.

Agradeço ao Prof. Dr. Paulo Roberto, pela grande oportunidade que me proporcionou. Uma experiência única! Agradeço pela orientação, pela recepção quando cheguei ao Rio de Janeiro. Mais do que um orientador, foi um amigo e sempre foi bem compreensível quando precisei.

Agradeço ao Prof. Dr. João Xavier, que acreditou em mim desde a graduação. Sempre foi uma inspiração, um exemplo a ser seguido, tanto profissionalmente quanto pessoalmente.

Agradeço ao Prof. Dr. João Carlos, que me acompanha (e me orienta) desde o mestrado. Foi quem mais me encorajou a encarar esse desafio e uma das pessoas a quem mais sou grato.

Agradeço aos professores Glaydston Bento (UFG), Márcia Fampa (COPPE-UFRJ) e Erik Papa Quiroz (Universidad Mayor de San Marcos - Peru), por terem aceitado participar da banca e pelas valiosas observações e sugestões feitas para a melhoria da tese.

Agradeço aos vários amigos que fiz nessa jornada. Pelo apoio, pelo incentivo, as conversas, as motivações. De certa forma eles tornaram a jornada mais divertida. Não posso deixar de citar: Armando, (Alcendino) Neto, Tanilson e Orlando.

Agradeço aos professores e funcionários do PESC, pelo apoio prestado.

Agradeço a Capes, pelo importantíssimo apoio financeiro.

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

UMA MODIFICAÇÃO NO MÉTODO DO PONTO PROXIMAL PARA
FUNÇÕES DC EM VARIEDADES DE HADAMARD

Yldenilson Torres Almeida

Junho/2021

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Programa: Engenharia de Sistemas e Computação

Nesta tese, apresentamos uma modificação no método do ponto proximal para funções DC em variedades de Hadamard. A modificação consiste em usar a iterada calculada pelo método do ponto proximal para funções DC para gerar uma direção de descida. Nesse sentido, fazemos a análise de convergência e mostramos que com a ajuda dessa direção de descida conseguimos melhorar os resultados obtidos pelo método do ponto proximal para funções DC. Mostramos ainda que nosso método também melhora os resultados obtidos pelo método do ponto proximal clássico para funções convexas e calculamos a complexidade do nosso método aplicado à funções convexas. Fazemos ainda alguns experimentos numéricos para ilustrar, na prática, a melhora nos resultados.

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

A MODIFICATION IN THE PROXIMAL POINT METHOD FOR DC
FUNCTIONS ON HADAMARD MANIFOLDS

Yldenilson Torres Almeida

June/2021

Advisors: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Department: Systems Engineering and Computer Science

In this thesis, we present a modification in the proximal point method for DC functions in Hadamard manifolds. The modification consists of using the iteration computed by the proximal point method for DC functions to generate a descent direction. In this sense, we do the convergence analysis and show that with the help of this descent direction we were able to improve the results obtained by the proximal point method for DC functions. We also show that our method improves the results obtained by the classical proximal point method for convex functions and we compute the iteration-complexity of our method applied to convex functions. We still do some numerical experiments to illustrate, in practice, the improvement in results.

Sumário

Lista de Figuras	ix
Lista de Tabelas	x
Lista de Símbolos	xi
Introdução	1
1 Preliminares	4
1.1 Variedades diferenciáveis: conceitos, notações e resultados	4
1.1.1 Espaço tangente a uma variedade	5
1.1.2 Métrica Riemanniana	6
1.1.3 Conexão Riemanniana	7
1.1.4 Gradiente e Hessiana	8
1.1.5 Geodésicas e aplicação exponencial	9
1.1.6 Variedades de Hadamard	11
1.2 Otimização em variedades Riemannianas	15
1.2.1 Condições de otimalidade	16
1.2.2 Convexidade	17
2 O método e análise de convergência	20
2.1 O MDCPPA	20
2.2 Análise de convergência	22
2.2.1 Caso não convexo	22
2.2.2 Caso convexo	27
3 Experimentos numéricos	39
3.1 As matrizes simétricas positivas definidas	39
3.2 Aplicando o método em uma função convexa	40
3.3 Aplicando o método em funções DC	45
3.4 Conclusão	52
Referências Bibliográficas	53

Lista de Figuras

1.1	Espaço tangente a \mathcal{M} em x	5
1.2	Triângulo geodésico em uma variedade de Hadamard	14
1.3	Articulação geodésica em \mathcal{M}	15
3.1	Caso convexo com matrizes 5×5	42
3.2	Caso convexo com matrizes 10×10	43
3.3	Caso convexo com matrizes 20×20	43
3.4	Caso convexo com matrizes 50×50	44
3.5	Caso DC não convexo com matrizes 5×5	46
3.6	Caso DC não convexo com matrizes 10×10	47
3.7	Caso DC não convexo com matrizes 20×20	47
3.8	Caso DC não convexo com matrizes 50×50	48
3.9	Caso DC convexo com matrizes 5×5	50
3.10	Caso DC convexo com matrizes 10×10	50
3.11	Caso DC convexo com matrizes 20×20	51
3.12	Caso DC convexo com matrizes 50×50	51

Lista de Tabelas

3.1	Média do número de iteradas para cada matriz $n \times n$ no caso convexo	44
3.2	Média do tempo gasto (em seg) para cada cada matriz $n \times n$ no caso convexo	45
3.3	Média do número de iteradas para cada matriz $n \times n$ no caso DC não convexo	48
3.4	Média do tempo gasto (em seg) para cada cada matriz $n \times n$ no caso DC não convexo	49
3.5	Média do número de iteradas para cada matriz $n \times n$ no caso DC convexo	52
3.6	Média do tempo gasto (em seg) para cada cada matriz $n \times n$ no caso DC convexo	52

Lista de Símbolos

\emptyset	Conjunto vazio, p. 18
\mathbb{H}^n	Espaço hiperbólico de dimensão n , p. 11
\mathbb{N}	Conjunto dos números naturais, p. 11
\mathbb{R}	Conjunto dos números reais, p. 7
\mathbb{R}^n	Espaço Euclidiano de dimensão n , p. 4
\mathbb{R}_+	Conjunto dos números reais não negativos, p. 7
\mathbb{R}_{++}	Conjunto dos números reais positivos, p. 10
\mathbb{S}^n	Conjunto das matrizes simétricas $n \times n$, p. 37
\mathbb{S}_{++}^n	Conjunto das matrizes simétricas positivas definidas $n \times n$, p. 37
\mathbb{S}_+^n	Conjunto das matrizes simétricas positivas semidefinidas $n \times n$, p. 37

Introdução

Na grande maioria dos casos, problemas de otimização envolvem minimizar (ou maximizar) funções de uma ou mais variáveis em um determinado domínio. Visando resolver esse tipo de problema de forma cada vez mais rápida e eficaz são sugeridos na literatura diversos métodos computacionais (algoritmos). Entre os mais clássicos, destacam-se o *Método do Ponto Proximal*, o *Método Gradiente* e o *Método de Newton*. Para este trabalho, vamos focar no Método do Ponto Proximal, que será abreviado aqui por MPP.

O MPP apareceu inicialmente em MARTINET [35] e é baseado na noção de aplicação proximal $J_\lambda^f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, dada por

$$J_\lambda^f(x) := \arg \min_{y \in \mathbb{R}^n} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}, \quad (1)$$

que havia sido introduzida anteriormente por MOREAU [37]. Porém, o MPP só viria a se tornar popular na literatura a partir da abordagem apresentada em ROCKAFELLAR [41], onde foi provada a convergência da sequência $\{x^k\}$, gerada iterativamente por

$$0 \in c_k T(x^{k+1}) + x^{k+1} - x^k,$$

onde $\{c_k\}$ é uma sequência de números reais estritamente positivos e $T : H \rightarrow H$ é um operador monótono (maximal) definido em um espaço de Hilbert H . Rockafellar mostrou que a sequência $\{x^k\}$ converge para um $x^* \in H$ tal que $0 \in T(x^*)$. Como caso particular, podemos tomar $T = \partial f$, onde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função convexa e ∂f é o *subdiferencial de f* .

Existem na literatura vários trabalhos que focam na extensão de métodos para contextos mais gerais. Em FERREIRA e OLIVEIRA [18], por exemplo, é apresentado uma extensão do MPP, onde o conjunto das restrições é uma variedade de Hadamard e a função objetivo é convexa nessa variedade. Nessas condições, é provado que qualquer sequência gerada iterativamente por

$$x^{k+1} := \arg \min_{x \in \mathcal{M}} \left\{ f(x) + \frac{1}{2c_k} d^2(x, x^k) \right\},$$

está bem definida e converge para um ponto minimizador de f . Extensões como essa, tem suas vantagens. A princípio, destacamos a vantagem de transformar problemas de otimização com restrições em problemas irrestritos, desde que introduzida uma métrica Riemanniana apropriada. Uma segunda vantagem é que problemas não convexos podem ser visto como problemas convexos (veremos isso nos exemplos dos testes numéricos), o que na prática é o melhor dos cenários em otimização. Alguns outros trabalhos que abordam o MPP em variedades Riemannianas são [2, 7, 8, 32, 33, 47, 50, 56, 57].

Para este trabalho em particular, focamos em uma classe especial de funções que podem ser escritas como a diferença de duas funções convexas. Essas funções são chamadas de *funções DC*. Denotamos o problema de minimizar uma função DC por

$$\min_{x \in \mathcal{M}} f(x) := f_1(x) - f_2(x), \quad (2)$$

onde $f_1, f_2 : \mathcal{M} \rightarrow \mathbb{R}$ são funções convexas. Os casos mais interessantes para esse tipo de problema são os que a função f não é convexa. O interesse nesse tipo de função tem aumentado bastante nos últimos anos. No contexto Riemanniano ainda são poucos os métodos dedicados a teoria de funções DC. Já no contexto Euclidiano essa teoria já é bastante avançada e aparece em diferentes cenários. Por exemplo, em condições de dualidade e otimalidade (HARTMAN [26], HIRIART-URRUTY [27]), métodos subgradiente (TAO e SOUAD [52]), métodos *cutting angle* e *branch-and-bound* (FERRER *et al.* [21], TUY e HORST [53]), otimização vetorial (FLORES-BAZÁN e OETTLI [22], GUO e LI [25]), teoria dos jogos e problemas de desigualdade variacional (MUU e QUOC [38]), problemas de equilíbrio (DINH *et al.* [15]), problemas *trust-region* (TAO e AN [51]), problemas de transporte (HOLMBERG e TUY [28]). É possível encontrar ainda aplicações da teoria de funções DC em mecânica dos fluidos (FERNÁNDEZ CARA e MORENO [17]), rede de reações bioquímicas (ARTACHO *et al.* [3]), problemas de grupos dinâmicos (ciências comportamentais) (BENTO *et al.* [10]). Além disso, problemas de funções DC aparecem constantemente em várias partes da análise (BAČÁK e BORWEIN [5]).

O primeiro método a lidar com funções DC foi proposto por TAO e SOUAD [52]. Abreviado por DCA, ele é um método tipo subgradiente e, dado $x^0 \in \mathbb{R}^n$, gera uma sequência $\{x^k\}$ iterativamente por

$$y^k \in \partial f_2(x^k), \quad x^{k+1} \in \partial f_1^*(y^k).$$

onde f_1^* denota a função conjugado de f_1 . Nessas condições, é provado que se a sequência $\{x^k\}$ possui pontos de acumulação, então eles são soluções de (2).

Em SUN *et al.* [49] é proposto uma versão do MPP para minimização de funções DC (chamado DCPA). A idéia chave desse método é usar as propriedades con-

vexas das duas funções separadamente. Ele compartilha da mesma idéia do DCA, mas de um ponto de vista computacional é mais simples pois evita o uso de funções conjugadas e, além disso, é bem conhecido que métodos subgradientes podem não gerar sequências monótonas decrescentes, mesmo no caso que envolvam funções convexas (ver POLYAK [40]). Um exemplo comparando a performance do MPP aplicado em funções DC e o DCA pode ser encontrada em MAINGÉ e MOUDAFI [34]. Uma aplicação interessante do MPP em funções DC pode ser encontrado em CRUZ NETO *et al.* [14] onde os autores mostram como ele pode ser uma boa ferramenta para resolução do problema de produção ótimo de uma companhia usando a abordagem *Variational Rationality* (VR) de dinâmicas humanas (ver SOUBEYRAN [44, 45, 46]). Uma extensão do DCPA para o contexto Riemanniano foi proposto por SOUZA e OLIVEIRA [48].

A seguir, apresentamos a forma na qual esse trabalho está dividido. No capítulo 1 relembramos alguns conceitos e resultados que julgamos ser essenciais para o bom entendimento do trabalho. A maioria dos resultados são apresentados sem prova, podendo ser facilmente encontrados em livros introdutórios sobre o assunto. O capítulo 2 é dedicado a apresentação dos resultados publicados em ALMEIDA *et al.* [1]. Nele apresentamos uma modificação no MPP que consiste em combinar a iterada gerada pelo DCPA com uma busca linear para definir uma direção de busca. Provamos que a direção gerada não só é uma direção de descida para a função objetivo, como também é um direção de descida no ponto gerado pelo DCPA. Assim, com um leve esforço computacional adicional de uma busca linear, pode-se obter uma diminuição no valor funcional em cada iterada. Provamos que se a sequência gerada pelo algoritmo tem pontos de acumulação então esses são pontos críticos da função. Provamos ainda a convergência do algoritmo para o caso convexo, isto é, quando $f_2 \equiv 0$ em (2). Consequentemente, mostramos que o algoritmo “melhora” também o MPP clássico para funções convexas, no sentido de que o nosso método atinge o critério de parada usando menos iteradas e gastando menos tempo de processamento. No capítulo 3 apresentamos alguns exemplos e fazemos alguns testes numéricos. Tabelas e gráficos ilustram os resultados obtidos e a performance do método em relação a outros.

Nosso método é baseado na idéia de aceleração do DCA proposto em ARTACHO *et al.* [3].

Capítulo 1

Preliminares

Neste capítulo, relembramos alguns conceitos e apresentamos alguns resultados que usaremos no decorrer do trabalho. Na primeira parte fazemos uma breve revisão de algumas definições recorrentes em variedades diferenciáveis, além de convencionar as notações adotadas e enunciar os resultados clássicos que usaremos. Na segunda parte apresentamos alguns resultados da teoria de otimização em variedades Riemannianas que serão essenciais na análise de convergência do método.

1.1 Variedades diferenciáveis: conceitos, notações e resultados

Em resumo, uma *variedade diferenciável n -dimensional* é um conjunto em \mathbb{R}^n formado pela reunião de um número finito ou enumerável de domínios \mathcal{M}_i , onde cada um desses domínios possui um sistema de coordenadas (locais) que determinam de modo único qualquer ponto de \mathcal{M}_i por, digamos (x_1^i, \dots, x_n^i) . Se dois domínios \mathcal{M}_i e \mathcal{M}_j se interceptam, então os pontos em comum tem coordenadas de \mathcal{M}_i (resp. \mathcal{M}_j) que podem ser expressas por uma função diferenciável das coordenadas de \mathcal{M}_j (resp. \mathcal{M}_i). Para esta seção, \mathcal{M} denotará sempre a variedade diferenciável de dimensão n . Aqui, dizer que uma aplicação é diferenciável será equivalente a dizer que a mesma é de classe C^∞ . Os resultados apresentados aqui serão enunciados sem demonstração pois podem ser encontrados sem muita dificuldade em livros introdutórios de geometria Riemanniana, tais como DO CARMO [16] (de onde adotamos a maior parte das notações) e SAKAI [43].

1.1.1 Espaço tangente a uma variedade

Seja \mathcal{M} uma variedade diferenciável n -dimensional e considere

$$\begin{aligned}\alpha &: (-\epsilon, \epsilon) \rightarrow \mathcal{M} \\ t &\mapsto (\alpha_1(t), \dots, \alpha_n(t)),\end{aligned}$$

uma *curva diferenciável* em \mathcal{M} . Se t representa o tempo, então a derivada de α em t representa o vetor velocidade tangente a curva em $\alpha(t)$. Supondo então que $\alpha(0) = x$, temos que um *vetor tangente* a $x \in \mathcal{M}$ é o vetor tangente em $t = 0$ da curva α . O conjunto de todos os vetores tangentes a \mathcal{M} em x será denotado por $T_x\mathcal{M}$ (ver fig. 1.1). Esse conjunto, munido das operações usuais de funções, forma um espaço vetorial de dimensão n . A união de todos os espaços tangentes a \mathcal{M} é chamado de *fibrado tangente de \mathcal{M}* e denotado por $T\mathcal{M}$, isto é,

$$T\mathcal{M} := \bigcup_{x \in \mathcal{M}} T_x\mathcal{M}.$$

Chamamos de *campo de vetores* em \mathcal{M} toda aplicação $X : \mathcal{M} \rightarrow T\mathcal{M}$ que associa a cada $x \in \mathcal{M}$ um $X(x)$ em $T_x\mathcal{M}$. Denotamos por $\mathfrak{X}^k(\mathcal{M})$ o espaço dos campos de vetores em \mathcal{M} de classe C^k , $k \geq 0$. Quando $k = +\infty$ denotamos apenas por $\mathfrak{X}(\mathcal{M})$.

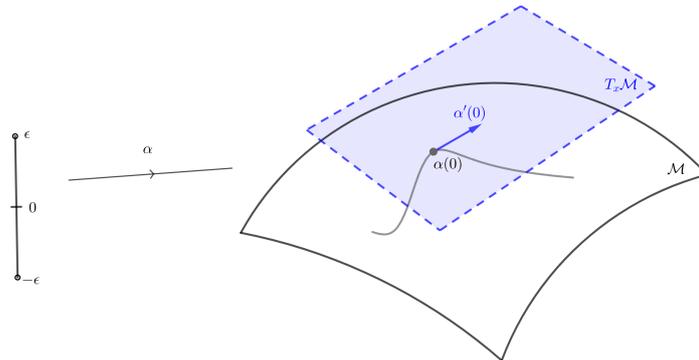


Figura 1.1: Espaço tangente a \mathcal{M} em x

A partir da idéia de espaço tangente é possível estender a noção de diferencial de uma aplicação diferenciável em variedades. Para isso, considere a seguinte

Proposição 1.1. *Sejam \mathcal{M}_1 e \mathcal{M}_2 variedades diferenciáveis de dimensões m e n , respectivamente. Considere $\varphi : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ uma aplicação diferenciável. Para todo $x \in \mathcal{M}_1$ e todo $v \in T_x\mathcal{M}_1$, escolha uma curva diferenciável $\alpha : (-\epsilon, \epsilon) \rightarrow \mathcal{M}_1$ tal*

que $\alpha(0) = x$ e $\alpha'(0) = v$. Tome $\beta := \varphi \circ \alpha$. A aplicação $d\varphi_x : T_x\mathcal{M}_1 \rightarrow T_{\varphi(x)}\mathcal{M}_2$ dada por $d\varphi_x(v) = \beta'(0)$ é uma aplicação linear que não depende da escolha de α .

Demonstração. Ver [16, Proposition 2.7, p. 9] □

A aplicação linear $d\varphi_x$ definida na proposição acima é chamada de *diferencial* de φ em x .

1.1.2 Métrica Riemanniana

Uma *métrica Riemanniana* em uma variedade diferenciável \mathcal{M} é uma correspondência que associa a cada ponto $x \in \mathcal{M}$, um produto interno $\langle \cdot, \cdot \rangle_x$ no espaço tangente $T_x\mathcal{M}$. Esse produto interno varia diferenciavelmente com x . Toda variedade diferenciável dotada de um métrica Riemanniana é chamada de *variedade Riemanniana*.

Proposição 1.2. *Toda variedade diferenciável (Hausdorff com base enumerável) tem uma métrica Riemanniana.*

Demonstração. Ver [16, Proposition 2.10, p. 43]. □

Dada uma métrica Riemanniana em \mathcal{M} , é possível definir a *norma* de um vetor $v \in T_x\mathcal{M}$, dada por

$$\|v\|_x = \langle v, v \rangle_x^{1/2}.$$

Além disso, se v e w são dois vetores não nulos em $T_x\mathcal{M}$ e $\theta \in [0, \pi]$ é o *ângulo* formado por esses dois vetores, então

$$\cos \theta = \frac{\langle v, w \rangle_x}{\|v\|_x \|w\|_x}.$$

A partir de agora, sempre quando não houver risco de confusão, o índice x será omitido do produto interno.

Dados $x, y \in \mathcal{M}$, denotamos por C_{xy} o conjunto de todas as curvas diferenciáveis por partes ligando os pontos x e y em \mathcal{M} . Em outras palavras, $\alpha \in C_{xy}$ se $\alpha : [a, b] \rightarrow \mathcal{M}$ é uma aplicação diferenciável por partes com $\alpha(a) = x$ e $\alpha(b) = y$. O *comprimento da curva* α em C_{xy} é dado por

$$l(\alpha) = \int_a^b \|\alpha'(t)\| dt,$$

onde

$$\|\alpha'(t)\| = \left\langle \frac{d}{dt}(\alpha(t)), \frac{d}{dt}(\alpha(t)) \right\rangle^{\frac{1}{2}}.$$

A *distância Riemanniana* entre dois pontos x e y em \mathcal{M} é dada por

$$d(x, y) = \inf_{\alpha \in C_{xy}} l(\alpha).$$

A função $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}_+$ definida acima é contínua e induz a topologia original em \mathcal{M} .

1.1.3 Conexão Riemanniana

Para o que vem a seguir, precisamos relembrar o conceito de derivada covariante em um espaço Euclidiano. Seja Y um campo de vetores diferenciável em \mathbb{R}^n , isto é, $Y \in \mathfrak{X}(\mathbb{R}^n)$. Dado X_x um vetor tangente a \mathbb{R}^n no ponto x , o vetor

$$\nabla_{X_x} Y := \left. \frac{d}{dt} Y(x + tX_x) \right|_{t=0},$$

tangente a \mathbb{R}^n no ponto x é chamado de *derivada covariante* de Y com respeito a X_x . Em termos de diferencial de uma aplicação, a igualdade acima é equivalente a

$$\nabla_{X_x} Y := dY(X_x).$$

Se $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é uma função real diferenciável, então $\nabla_{X_x} f = df(X_x)$ é a *derivada direcional* de f com respeito a X_x . Denotando $\nabla_{X_x} f$ por $X_x(f)$, temos que

$$\nabla_{X_x}(fY) = d(fY)(X_x) = df(X_x)Y + f dY(X_x) = X_x(f)Y + f \nabla_{X_x} Y.$$

Para estender a noção de derivada covariante para variedades Riemannianas faz-se necessário a seguinte

Definição 1. *Uma conexão afim em uma variedade Riemanniana \mathcal{M} é uma aplicação*

$$\nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \rightarrow \mathfrak{X}(\mathcal{M}),$$

denotada por $(X, Y) \mapsto \nabla_X Y$, que possui as seguintes propriedades:

- $\nabla_{fX+gY} Z = f \nabla_X Z + g \nabla_Y Z$
- $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$
- $\nabla_X (fY) = X(f)Y + f \nabla_X Y$,

onde $X, Y, Z \in \mathfrak{X}(\mathcal{M})$ e $f, g : \mathcal{M} \rightarrow \mathbb{R}$ são funções reais diferenciáveis.

Uma conexão afim ∇ induz uma *derivada covariante* com respeito ao campo de vetores X em \mathcal{M} , denotada por ∇_X . Além disso, dada uma conexão afim, podemos

determinar o *tensor curvatura* da variedade \mathcal{M} , dada por

$$R : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \rightarrow \mathfrak{X}(\mathcal{M}),$$

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X, Y]} Z, \quad (1.1)$$

onde $[X, Y] = \nabla_X Y - \nabla_Y X \in \mathfrak{X}(\mathcal{M})$.

Teorema 1.1 (Levi-Civita). *Dada uma variedade Riemanniana \mathcal{M} , existe uma única conexão (simétrica) afim ∇ em \mathcal{M} com a propriedade*

$$\nabla_X \langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle, \quad \forall X, Y, Z \in \mathfrak{X}(\mathcal{M}).$$

Demonstração. Ver [54, Theorem 2.2, p. 7] □

O teorema acima garante a existência de uma conexão afim em qualquer variedade Riemanniana. Essa conexão recebe o nome de *conexão Riemanniana* (ou *conexão de Levi-Civita*).

Seja \mathcal{M} uma variedade Riemanniana com tensor curvatura dado por (1.1). Seja x um ponto em \mathcal{M} e tome σ um subespaço vetorial bidimensional do espaço tangente $T_x \mathcal{M}$. Suponha que $\{X_x, Y_x\}$ é uma base de σ . Então, o número real

$$K_x(\sigma) = \frac{\langle R(X_x, Y_x)Y_x, X_x \rangle}{\|X_x\|^2 \|Y_x\|^2 - \langle X_x, Y_x \rangle^2},$$

onde $\|X\|^2 = \langle X, X \rangle$, é chamado de *curvatura seccional de \mathcal{M} ao longo da seção σ* .

1.1.4 Gradiente e Hessiana

Sejam \mathcal{M} uma variedade Riemanniana e $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função real de classe C^1 . Dado $x \in \mathcal{M}$, chamamos de *gradiente* de f em x , e denotamos por $\text{grad}f(x)$, o único campo de vetores tal que

$$\langle \text{grad}f(x), v \rangle = \nabla_v f = df_x(v), \quad \forall v \in T_x \mathcal{M}.$$

Os pontos que zeram o $\text{grad}f$ são chamados de *pontos críticos de f* .

Teorema 1.2. *Seja \mathcal{M} um variedade Riemanniana e suponha que $f : \mathcal{M} \rightarrow \mathbb{R}$ é uma função real de classe C^1 que não possui pontos críticos. Então:*

1. *O campo de vetores $\text{grad}f$ é ortogonal às hipersuperfícies das curvas de nível de f .*
2. *$\text{grad}f(x)$ mostra, em todo $x \in \mathcal{M}$, a direção e o sentido de maior crescimento da função f .*

Demonstração. Ver [54, Theorem 3.1, p. 9] □

Se $f : \mathcal{M} \rightarrow \mathbb{R}$ é uma função de classe C^2 , então a segunda derivada covariante de f recebe o nome de *Hessiana* de f e é denotada por $\text{Hess}f$. Em outros termos, dado $x \in \mathcal{M}$, $\text{Hess}f(x) : T_x\mathcal{M} \rightarrow T_x\mathcal{M}$ é uma aplicação linear definida por

$$\text{Hess}f(x)v = \nabla_v \text{grad}f = d(\text{grad}f)_x(v) \quad \forall v \in T_x\mathcal{M}.$$

1.1.5 Geodésicas e aplicação exponencial

Seja $\gamma : I \subset \mathbb{R} \rightarrow \mathcal{M}$ uma curva diferenciável. Tome $X : I \rightarrow T\mathcal{M}$ um campo de vetores diferenciáveis ao longo de γ , isto é,

$$X(t) = X(\gamma(t)) \in T_{\gamma(t)}\mathcal{M}, \quad \forall t \in I.$$

O campo de vetores X definido acima é dito *paralelo ao longo de γ* se

$$\nabla_{\gamma'(t)}X(t) = 0, \quad \forall t \in I.$$

Dizemos que a curva diferenciável γ é uma *geodésica* em \mathcal{M} quando

$$\nabla_{\gamma'(t)}\gamma'(t) = 0, \quad \forall t \in I,$$

isto é, γ é uma geodésica se, somente se, o vetor velocidade γ' é um campo paralelo ao longo de γ .

Note que a equação que define uma geodésica é uma equação diferencial ordinária não linear de segunda ordem. Então a geodésica γ pode ser determinada por sua posição x e sua velocidade v em x , fazendo sentido então denotarmos tal geodésica por γ_v . A restrição de uma geodésica a um intervalo fechado e limitado é chamado de *segmento geodésico*. O conjunto de todos os segmentos geodésicos ligando x a y em \mathcal{M} será denotado por Γ_{xy} . Um segmento geodésico ligando x a y em \mathcal{M} é dito *minimal* se seu comprimento é igual a $d(x, y)$.

Chamamos de *variedade Riemanniana completa* as variedades Riemannianas cujas geodésicas estão definidas para todo $t \in \mathbb{R}$. Se \mathcal{M} é uma variedade Riemanniana completa e γ_v é uma geodésica em \mathcal{M} então a igualdade

$$\gamma_{\lambda v}(t) = \gamma_v(\lambda t), \tag{1.2}$$

vale para todo $\lambda \in \mathbb{R}_{++}$ e todo $t \in \mathbb{R}$. Além disso, se γ_v é uma geodésica em \mathcal{M} tal

que $\gamma_v(0) = x$ então, podemos definir a *aplicação exponencial* em x por

$$\begin{aligned}\exp_x : T_x\mathcal{M} &\rightarrow \mathcal{M} \\ \exp_x v &= \gamma_v(1).\end{aligned}$$

Segue de (1.2) que

$$\gamma_v(t) = \gamma_{tv}(1) = \exp_x(tv), \quad \forall t \in \mathbb{R}.$$

A aplicação exponencial definida acima é diferenciável. Mais do que isso, para cada $x \in \mathcal{M}$, existe uma vizinhança da origem em $T_x\mathcal{M}$, digamos U , tal que

$$\exp_x : U \subset T_x\mathcal{M} \rightarrow \mathcal{M}$$

aplica difeomorficamente U em um subconjunto aberto de \mathcal{M} (ver [16, Proposition 2.9, p. 65]).

Seja $\gamma : I \subset \mathbb{R} \rightarrow \mathcal{M}$ uma geodésica em \mathcal{M} . O *transporte paralelo* ao longo de γ é uma aplicação

$$P_{\gamma, t_0, t} : T_{\gamma(t_0)}\mathcal{M} \rightarrow T_{\gamma(t)}\mathcal{M},$$

que identifica o espaço tangente a \mathcal{M} em $\gamma(t_0)$ com o espaço tangente em $\gamma(t)$. Essa identificação ocorre da seguinte forma: dado $v \in T_{\gamma(t_0)}\mathcal{M}$, o vetor $P_{\gamma, t_0, t}(v) \in T_{\gamma(t)}\mathcal{M}$ é obtido transportando-se *paralelamente* o vetor v ao longo da curva γ . Quando a geodésica ligando dois pontos é única então a aplicação transporte paralelo será identificada apenas por

$$P_{x^0, x} : T_{x^0}\mathcal{M} \rightarrow T_x\mathcal{M}.$$

Uma função $f : \mathcal{M} \rightarrow \mathbb{R}$ de classe C^1 é dita de *gradiente Lipschitz* com constante $L \geq 0$ se, para todo $x, y \in \mathcal{M}$ e $\gamma : [0, a] \rightarrow \mathcal{M}$ com $\gamma(0) = x$ e $\gamma(a) = y$, se verifica

$$\|\text{grad}f(\gamma(t)) - P_{\gamma, 0, t}\text{grad}f(x)\| \leq Ll(t), \quad \forall t \in [0, a], \quad (1.3)$$

onde $l(t)$ denota o comprimento da geodésica γ entre os pontos $\gamma(0)$ e $\gamma(t)$.

Teorema 1.3 (Hopf-Rinow). *Se \mathcal{M} uma variedade Riemanniana, então as seguintes afirmações são equivalentes:*

1. Para cada $x \in \mathcal{M}$, \exp_x está definida em todo $T_x\mathcal{M}$, ou seja, \mathcal{M} é uma variedade Riemanniana completa.
2. (\mathcal{M}, d) é um espaço métrico completo, onde d é a distância Riemanniana.
3. Os subconjuntos limitados e fechados de \mathcal{M} são compactos.

Além disso, cada uma das afirmações acima implica em

4. Para cada dois pontos $x, y \in \mathcal{M}$ existe uma geodésica minimizante γ ligando x a y .

Demonstração. Ver [16, Teorema 2.8, p. 146] □

Uma sequência $\{x^k\} \subset \mathcal{M}$ é dita ser *quasi-Fejér convergente* a um subconjunto não vazio $U \subset \mathcal{M}$ se existe uma sequência $\{\epsilon_k\}$ de números reais não negativos tal que $\sum_{k=0}^{+\infty} \epsilon_k < +\infty$ e

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + \epsilon_k, \quad \forall k \in \mathbb{N} \text{ e } \forall x \in U. \quad (1.4)$$

Quando $\epsilon_k = 0$ para todo $k \in \mathbb{N}$ e vale a desigualdade (1.4), dizemos que $\{x^k\}$ é *Fejér convergente* a U .

Lema 1.4. *Seja $\{x^k\}$ uma sequência no espaço métrico completo (\mathcal{M}, d) . Se $\{x^k\}$ é quasi-Fejér convergente a um subconjunto não vazio $U \subset \mathcal{M}$, então $\{x^k\}$ é uma sequência limitada. Além disso, se um ponto de acumulação \bar{x} pertence a U , então*

$$\lim_{k \rightarrow +\infty} x^k = \bar{x}.$$

Demonstração. Ver [13, Lemma 5.2]. □

Observação 1. *Um resultado análogo é obtido em [18, Lemma 6.1] para sequências Fejér convergente.*

1.1.6 Variedades de Hadamard

Seja \mathcal{M} uma variedade Riemanniana completa e simplesmente conexa. Se a curvatura seccional de \mathcal{M} é sempre não positiva dizemos que \mathcal{M} é uma *variedade de Hadamard*. Um importante exemplo de variedade de Hadamard que usaremos neste trabalho é o *espaço hiperbólico n -dimensional* \mathbb{H}^n , definido por

$$\mathbb{H}^n := \{x \in \mathbb{R}^{n+1} : \langle x, x \rangle_{\mathbb{H}} = -1, x_{n+1} > 0\},$$

onde $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ denota a *métrica de Lorentz em \mathbb{R}^{n+1}* , dada por

$$\langle x, y \rangle_{\mathbb{H}} := \sum_{i=1}^n x_i y_i - x_{n+1} y_{n+1}, \quad \forall x, y \in \mathbb{R}^{n+1}.$$

Nessa configuração, \mathbb{H}^n é uma variedade de Hadamard com curvatura seccional $K \equiv -1$ (ver [19] e [12]).

Teorema 1.5 (Cartan-Hadamard). *Se \mathcal{M} é uma variedade de Hadamard de dimensão n , então \mathcal{M} é difeomorfa ao espaço Euclidiano \mathbb{R}^n . Mais precisamente, para cada $x \in \mathcal{M}$, a aplicação $\exp_x : T_x \mathcal{M} \rightarrow \mathcal{M}$ é um difeomorfismo.*

Demonstração. Ver [16, Theorem 3.1, p. 149] □

O Teorema 1.5 garante que uma variedade de Hadamard tem a mesma topologia e estrutura diferenciável do \mathbb{R}^n . Além disso, ele explicita a existência da inversa da aplicação exponencial, dada por

$$\exp_x^{-1} : \mathcal{M} \rightarrow T_x\mathcal{M}.$$

Dados $x, y \in \mathcal{M}$, tome $v = \exp_x^{-1} y \in T_x\mathcal{M}$. Então, a geodésica que liga x a y em \mathcal{M} , está bem definida por

$$\gamma : [0, 1] \rightarrow \mathcal{M}, \quad \gamma(t) = \exp_x(tv). \quad (1.5)$$

Além disso, como $d(\exp_x)_0$ coincide com a aplicação *identidade* em $T_x\mathcal{M}$ (ver demonstração de [16, Proposition 2.9, p. 65]) temos que

$$\dot{\gamma}(0) = d(\exp_x)_0(v) = v.$$

Vale ainda a seguinte relação entre distância Riemanniana e aplicação exponencial:

$$d(x, y) = \|\exp_x^{-1} y\|. \quad (1.6)$$

Observe que $\exp_x^{-1} : \mathcal{M} \rightarrow T_x\mathcal{M}$ é uma aplicação diferenciável (consequência do Teorema 1.5). Então, fixando $\bar{x} \in \mathcal{M}$ e tomando a função distância como definido em (1.6), a função

$$\phi_{\bar{x}} : \mathcal{M} \rightarrow \mathbb{R}_+, \quad \phi_{\bar{x}}(x) = \frac{1}{2}d^2(x, \bar{x})$$

é diferenciável para todo $x \in \mathcal{M}$ com

$$\text{grad}\phi_{\bar{x}}(x) = -\exp_x^{-1}\bar{x}.$$

(Ver [43, Proposition 4.8, p. 108])

Proposição 1.3. *Seja \mathcal{M} uma variedade de Hadamard. Dado $\bar{x} \in \mathcal{M}$, considere $\{x^k\} \subset \mathcal{M}$ uma sequência tal que $x^k \rightarrow \bar{x}$ quando $k \rightarrow +\infty$. Então são válidas as seguintes afirmações:*

1. *Para todo $y \in \mathcal{M}$ vale:*

$$\lim_{k \rightarrow +\infty} \exp_{x^k}^{-1} y = \exp_{\bar{x}}^{-1} y \quad e \quad \lim_{k \rightarrow +\infty} \exp_{y^k}^{-1} x^k = \exp_y^{-1} \bar{x}.$$

2. *Se $v^k \in T_{x^k}\mathcal{M}$ para todo $k \in \mathbb{N}$ e $v^k \rightarrow \bar{v}$, então $\bar{v} \in T_{\bar{x}}\mathcal{M}$.*

3. Para qualquer $\bar{v} \in T_{\bar{x}}\mathcal{M}$, a função $F : \mathcal{M} \rightarrow T\mathcal{M}$ dada por $F(x) = P_{x,\bar{x}}\bar{v}$ para cada $x \in \mathcal{M}$ é contínua em \mathcal{M} .

Demonstração. Ver [33, Lemma 2.4, p. 666]. \square

Os dois lemas a seguir são consequências da Proposição 1.3. As demonstrações de ambos podem ser encontradas em [6].

Lema 1.6. *Sejam $\bar{x}, \bar{y} \in \mathcal{M}$ e considere $\{x^k\}, \{y^k\}$ sequências em \mathcal{M} tais que $\lim_{k \rightarrow +\infty} x^k = \bar{x}$ e $\lim_{k \rightarrow +\infty} y^k = \bar{y}$. Então,*

$$\lim_{k \rightarrow +\infty} \exp_{x^k}^{-1} y^k = \exp_{\bar{x}}^{-1} \bar{y}.$$

Demonstração. Ver [6, Lemma 1.1]. \square

Lema 1.7. *Sejam $\bar{x} \in \mathcal{M}$, $\bar{v} \in T_{\bar{x}}\mathcal{M}$, $\{x^k\} \subset \mathcal{M}$, $\{v^k\} \subset T_{x^k}\mathcal{M}$ tais que $\lim_{k \rightarrow +\infty} x^k = \bar{x}$ e $\lim_{k \rightarrow +\infty} v^k = \bar{v}$. Tome $\bar{\lambda} \in [0, 1)$ e considere $\{\lambda_k\} \subset (0, 1)$ uma sequência tal que $\lim_{k \rightarrow +\infty} \lambda_k = \bar{\lambda}$. Seja $\{y^k\}$ a sequência definida por $y^k := \exp_{x^k}(\lambda_k v^k)$ para todo $k \in \mathbb{N}$. Então,*

$$\lim_{k \rightarrow +\infty} P_{x^k, y^k} v^k = P_{\bar{x}, \bar{y}} \bar{v}, \quad \text{onde } \bar{y} := \exp_{\bar{x}}(\bar{\lambda} \bar{v}) = \lim_{k \rightarrow +\infty} y^k.$$

Demonstração. Ver [6, Lemma 1.2]. \square

As variedades de Hadamard possuem algumas propriedades geométricas similares às dos espaços Euclidianos. Apresentamos a seguir uma dessas propriedades. Mas antes, precisamos formalizar a definição de triângulos em uma variedade. Um *triângulo geodésico* em uma variedade Riemanniana \mathcal{M} é o conjunto formado por três pontos x^1, x^2, x^3 em \mathcal{M} (chamados de *vértices*) e três segmentos geodésicos minimizantes γ_{i+1} ligando os pontos x^{i+1} a x^{i+2} , $i \equiv 1, 2, 3 \pmod{3}$ (chamados de *lados*) (ver fig. 1.2). Esse conjunto será denotado por $\Delta(x^1, x^2, x^3)$.

Teorema 1.8 (Lei dos cossenos). *Seja \mathcal{M} uma variedade de Hadamard. Tome em \mathcal{M} o triângulo geodésico $\Delta(x^1, x^2, x^3)$. Denote por $\gamma_{i+1} : [0, l_{i+1}] \rightarrow \mathcal{M}$ o segmento geodésico que liga x^{i+1} a x^{i+2} e defina $l_{i+1} := l(\gamma_{i+1})$, $\theta_{i+1} = \angle(\gamma'_{i+1}(0), -\gamma'_i(l_i))$, onde $i \equiv 1, 2, 3 \pmod{3}$. Então*

$$\theta_1 + \theta_2 + \theta_3 \leq \pi, \tag{1.7}$$

$$l_{i+1}^2 + l_{i+2}^2 - 2l_{i+1}l_{i+2} \cos \theta_{i+2} \leq l_i^2. \tag{1.8}$$

Se tivermos $K < 0$ as desigualdades acima são estritas.

Demonstração. Esses resultados são consequências do Teorema de Topogonov (ver [43, Teorema 4.2]). \square

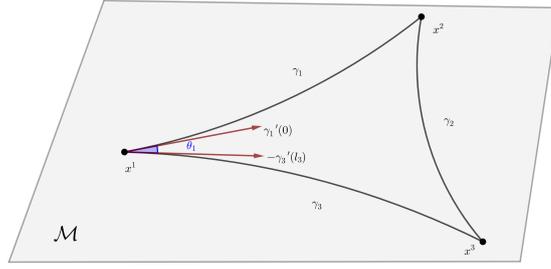


Figura 1.2: Triângulo geodésico em uma variedade de Hadamard

Observação 2. Se \mathcal{M} é uma variedade Riemanniana completa, então podemos usar a relação entre distância e aplicação exponencial dada em (1.6) combinada com a relação

$$\langle \exp_{x^{i+2}}^{-1} x^{i+1}, \exp_{x^{i+2}}^{-1} x^i \rangle = d(x^{i+1}, x^{i+2})d(x^i, x^{i+2}) \cos \theta_{i+2}$$

para reescrever a inequação (1.8) como

$$d^2(x^{i+1}, x^{i+2}) + d^2(x^{i+2}, x^i) - 2\langle \exp_{x^{i+2}}^{-1} x^{i+1}, \exp_{x^{i+2}}^{-1} x^i \rangle \leq d^2(x^i, x^{i+1}).$$

Para finalizar essa seção, apresentaremos alguns resultados assumindo que \mathcal{M} é uma variedade de Hadamard com curvatura seccional K limitada inferiormente por uma constante negativa κ . Enunciaremos alguns resultados geométricos que serão cruciais na análise de convergência do caso convexo do nosso método. Para começar, considere \mathcal{M}_κ^n a variedade de Hadamard obtida do espaço hiperbólico \mathbb{H}^n multiplicando a função distância pela constante $1/\sqrt{|\kappa|}$. Para simplificar as notações, denotaremos $\hat{\kappa} = \sqrt{|\kappa|}$.

Chamamos de *articulação geodésica*¹ em \mathcal{M} a figura geométrica composta por um ponto $x \in \mathcal{M}$ (chamado de vértice da articulação) e dois segmentos geodésicos γ_1, γ_2 (chamados de arestas da articulação) partindo de x (ver fig. 1.3). Essa articulação será denotada por $\Lambda(x; \gamma_1, \gamma_2)$. A articulação $\Lambda(\bar{x}; \bar{\gamma}_1, \bar{\gamma}_2)$ em \mathcal{M}_κ^2 é dita ser uma *articulação de comparação* de $\Lambda(x; \gamma_1, \gamma_2)$ se ela satisfaz

$$l(\bar{\gamma}_1) = l(\gamma_1), \quad l(\bar{\gamma}_2) = l(\gamma_2), \quad \text{e} \quad \angle_{\bar{x}}(\bar{\gamma}_1, \bar{\gamma}_2) = \angle_x(\gamma_1, \gamma_2),$$

onde $\angle_x(\gamma_1, \gamma_2)$ denota o ângulo interno formado pelos segmentos geodésicos γ_1 e γ_2 em x .

¹em inglês: *generalized geodesic hinge*

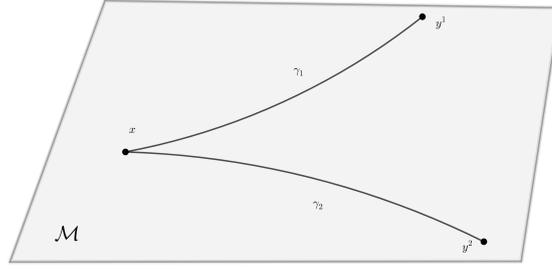


Figura 1.3: Articulação geodésica em \mathcal{M}

Proposição 1.4. *Seja $\Delta(x^1, x^2, x^3)$ um triângulo geodésico em \mathcal{M}_κ^2 . Então*

$$\begin{aligned} \cosh(\hat{\kappa} d(x^1, x^3)) &= \cosh(\hat{\kappa} d(x^3, x^2)) \cosh(\hat{\kappa} d(x^2, x^1)) \\ &\quad - \sinh(\hat{\kappa} d(x^3, x^2)) \sinh(\hat{\kappa} d(x^2, x^1)) \cos \theta, \end{aligned}$$

onde θ é o ângulo interno do $\Delta(x^1, x^2, x^3)$ em x^2 .

Demonstração. Ver [43, p. 138]. □

Proposição 1.5. *Seja \mathcal{M} uma variedade Riemanniana completa de curvatura limitada inferiormente por $\kappa < 0$. Considere $\Lambda(x; \gamma_1, \gamma_2)$ uma articulação geodésica em \mathcal{M} e seja $\Lambda(\bar{x}; \bar{\gamma}_1, \bar{\gamma}_2)$ a sua articulação de comparação em \mathcal{M}_κ^2 . Então*

$$d(y_{\gamma_1}, y_{\gamma_2}) \leq d(y_{\bar{\gamma}_1}, y_{\bar{\gamma}_2}),$$

onde $y_{\gamma_1}, y_{\gamma_2}, y_{\bar{\gamma}_1}$ e $y_{\bar{\gamma}_2}$ denotam os pontos finais que determinam as geodésicas $\gamma_1, \gamma_2, \bar{\gamma}_1$ e $\bar{\gamma}_2$, respectivamente.

Demonstração. Ver [43, Theorem 4.2(2), p. 161]. □

1.2 Otimização em variedades Riemannianas

Nesta seção, apresentamos algumas condições de otimalidade em variedades Riemannianas e resultados básicos de análise convexa, também no contexto das variedades Riemannianas. Optamos por demonstrar alguns resultados desta seção. Isso porque alguns elementos que aparecem nas demonstrações serão usados no decorrer do trabalho. Contudo, todos os resultados apresentados aqui podem ser encontrados em UDRISTE [54].

1.2.1 Condições de otimalidade

Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função de classe C^k , $k \geq 1$. Nosso interesse aqui, é obter condições de otimalidade para o problema

$$\min_{x \in \mathcal{M}} f(x) \quad (\mathcal{P})$$

Teorema 1.9 (Condição necessária de primeira ordem). *Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função de classe C^1 . Se x^* é solução de (\mathcal{P}) então $\text{grad}f(x^*) = 0$.*

Demonstração. Dado $v \in T_{x^*}\mathcal{M}$, considere $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ uma geodésica tal que $\gamma(0) = x^*$ e $\dot{\gamma}(0) = v$. Defina $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ por $\varphi(t) = f(\gamma(t))$. Como $t = 0$ é um ponto de mínimo de φ temos que

$$0 = \varphi'(0) = \langle \text{grad}f(x^*), v \rangle.$$

Como $v \in T_{x^*}\mathcal{M}$ foi tomado de forma arbitrária, segue o resultado. \square

Teorema 1.10 (Condição necessária de segunda ordem). *Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função de classe C^2 . Se x^* é solução de (\mathcal{P}) então*

$$\langle \text{Hess}f(x^*)v, v \rangle \geq 0, \quad \forall v \in T_{x^*}\mathcal{M}.$$

Demonstração. Tome $v \in T_{x^*}\mathcal{M}$ e seja $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ uma geodésica tal que $\gamma(0) = x^*$ e $\dot{\gamma}(0) = v$. Defina $\varphi(t) = f(\gamma(t))$. Então

$$\begin{aligned} \varphi'(t) &= \langle \text{grad}f(\gamma(t)), \dot{\gamma}(t) \rangle, \\ \varphi''(t) &= \langle \text{Hess}f(\gamma(t))\dot{\gamma}(t), \dot{\gamma}(t) \rangle + \langle \text{grad}f(\gamma(t)), \ddot{\gamma}(t) \rangle. \end{aligned}$$

Como γ é uma geodésica, $\ddot{\gamma}(t) = 0$ para todo $t \in \mathbb{R}$. Assim, na segunda igualdade acima, temos que

$$\varphi''(t) = \langle \text{Hess}f(\gamma(t))\dot{\gamma}(t), \dot{\gamma}(t) \rangle, \quad \forall t \in \mathbb{R}.$$

Como $t = 0$ minimiza φ , temos que $\varphi''(0) \geq 0$. Portanto,

$$\langle \text{Hess}f(x^*)v, v \rangle = \varphi''(0) \geq 0,$$

para qualquer que seja $v \in T_{x^*}\mathcal{M}$. \square

Teorema 1.11 (Condição suficiente de segunda ordem). *Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função de classe C^2 . Se x^* é um ponto crítico de f e $\text{Hess}f(x^*)$ é uma matriz*

definida positiva, isto é,

$$\langle \text{Hess}f(x^*)v, v \rangle > 0, \quad \forall v \in T_{x^*}\mathcal{M},$$

então x^* é ponto de mínimo local de f .

Demonstração. Tome $v \in T_{x^*}\mathcal{M}$ com $\|v\| = 1$. Seja η o menor auto-valor da matriz $\text{Hess}f(x^*)$. Então,

$$0 < \eta \leq \langle \text{Hess}f(x^*)v, v \rangle.$$

Agora, seja $\gamma : (-\epsilon, \epsilon) \rightarrow \mathcal{M}$ uma curva diferenciável em \mathcal{M} tal que $\gamma(0) = x^*$ e $\dot{\gamma}(0) = v$. Por Taylor, temos que

$$f(\gamma(t)) = f(x^*) + \frac{t^2}{2} \langle \text{Hess}f(x^*)v, v \rangle + o(t^2),$$

o que implica em

$$f(\gamma(t)) \geq f(x^*) + \frac{\eta}{2}t^2 + o(t^2).$$

Daí, existe t_0 (que não depende de v) tal que, para $0 \leq t \leq t_0$ tem-se que

$$f(\gamma(t)) \geq f(x^*).$$

□

1.2.2 Convexidade

Seja \mathcal{M} uma variedade Riemanniana. Um subconjunto $\Omega \subset \mathcal{M}$ é dito ser *convexo* se, para todo par de pontos x e y em Ω , existe uma geodésica minimal, contida em Ω , ligando x a y . Isso significa que se $\gamma : [a, b] \rightarrow \mathcal{M}$ é o segmento geodésico tal que $\gamma(a) = x$ e $\gamma(b) = y$, com $x, y \in \Omega$, então $\gamma((1-t)a + tb) \in \Omega$ para todo $t \in [0, 1]$. A partir daqui, vamos considerar que a variedade \mathcal{M} é sempre um conjunto convexo.

Uma função $f : \mathcal{M} \rightarrow \mathbb{R}$ é dita ser *função convexa* se a sua restrição a qualquer geodésica de \mathcal{M} é uma função convexa da reta, isto é, se $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ é uma geodésica em \mathcal{M} , então $f \circ \gamma : \mathbb{R} \rightarrow \mathbb{R}$ é uma função convexa. Em outros termos, $f : \mathcal{M} \rightarrow \mathbb{R}$ é uma função convexa se, somente se, para todo segmento geodésico $\gamma : [a, b] \rightarrow \mathcal{M}$ e todo $\lambda \in [0, 1]$, verifica-se

$$f(\gamma((1-\lambda)a + \lambda b)) \leq (1-\lambda)f(\gamma(a)) + \lambda f(\gamma(b)).$$

Teorema 1.12. *Se $f : \mathcal{M} \rightarrow \mathbb{R}$ é uma função convexa, então f é contínua.*

Demonstração. Ver [54, Corollaries 3.6 (1), p. 69].

□

Para os resultados a seguir vamos sempre supor que $f : \mathcal{M} \rightarrow \mathbb{R}$ é uma função de classe C^k , $k \geq 1$.

Teorema 1.13. *A função $f : \mathcal{M} \rightarrow \mathbb{R}$ é convexa se, e somente se, para qualquer ponto $x \in \mathcal{M}$ e geodésica $\gamma : \mathbb{R} \rightarrow \mathcal{M}$, com $\gamma(0) = x$, vale a desigualdade*

$$f(\gamma(t)) - f(x) \geq t \langle \text{grad} f(x), \dot{\gamma}(0) \rangle.$$

Demonstração. Ver [13, Theorem 3.1]. □

Corolário 1.13.1. *Se $f : \mathcal{M} \rightarrow \mathbb{R}$ é uma função convexa então todo ponto crítico de f é ponto de mínimo global.*

Demonstração. Basta tomar $x = x^*$ ponto crítico de f no Teorema 1.13. Assim,

$$f(\gamma(t)) \geq f(x^*), \quad \forall t \in \mathbb{R}.$$

□

Teorema 1.14. *A função $f : \mathcal{M} \rightarrow \mathbb{R}$ é convexa se, e somente se, para todo ponto $x \in \mathcal{M}$, a Hessiana de f em x é semi-definida positiva, isto é,*

$$\langle \text{Hess} f(x)v, v \rangle \geq 0, \quad \forall v \in T_x \mathcal{M}.$$

Demonstração. Ver [13, Theorem 3.2]. □

Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função convexa. Dado $\lambda > 0$, a *regularização de Moreau-Yosida* de f é a função $f_\lambda : \mathcal{M} \rightarrow \mathbb{R}$ dada por

$$f_\lambda(x) := \inf_{y \in \mathcal{M}} \left\{ f(y) + \frac{\lambda}{2} d^2(y, x) \right\}. \quad (1.9)$$

Proposição 1.6. *Se $f : \mathcal{M} \rightarrow \mathbb{R}$ é convexa então, para todo $x \in \mathcal{M}$ e $\lambda > 0$ existe um único ponto x^* tal que*

$$f(x^*) + \frac{\lambda}{2} d^2(x^*, x) = f_\lambda(x),$$

caracterizado ainda por

$$\text{grad} f(x^*) = \lambda \exp_{x^*}^{-1} x.$$

Demonstração. Ver [18, Lemma 4.2]. □

Um campo de vetores $X \in \mathfrak{X}(M)$ é dito ser *monótono* se

$$\langle v, \exp_x^{-1} y \rangle \leq \langle w, -\exp_y^{-1} x \rangle, \quad \forall x, y \in \text{dom}(X), \forall v \in X(x) \text{ e } \forall w \in X(y),$$

onde

$$\text{dom}(X) := \{x \in \mathcal{M} : X(x) \neq \emptyset\}$$

Quando a desigualdade acima é estrita dizemos que X é um *campo de vetores estritamente monótono*.

Proposição 1.7. *Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função de classe C^1 . Então, f é função convexa se, e somente se, $\text{grad}f$ é um campo de vetores monótono.*

Demonstração. Ver [54, Theorem 5.3, p. 80]. □

Capítulo 2

O método e análise de convergência

Seja \mathcal{M} uma variedade de Hadamard e considere $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função DC em \mathcal{M} , isto é, $f = f_1 - f_2$ onde $f_1, f_2 : \mathcal{M} \rightarrow \mathbb{R}$ são funções convexas. O nosso interesse, em um contexto geral, é resolver o problema

$$\min_{x \in \mathcal{M}} f_1(x) - f_2(x). \quad (\mathcal{P}')$$

Como dito antes, o interesse nesse tipo de problema tem aumentado bastante nos últimos anos devido as suas várias aplicações (ver Introdução). Na mesma proporção têm-se intensificado a busca por métodos computacionais que resolvam esse tipo de problema de forma cada vez mais eficiente. Pensando nisso, apresentamos neste capítulo uma modificação no MPP para funções DC. A priori, o método que apresentamos segue a idéia de ARTACHO *et al.* [3] que, em resumo, usa a iterada obtida pelo DCA para definir uma direção de busca, mostra que essa direção obtida é uma direção de descida e, assim, “andando nessa direção” em cada iterada, o método consegue melhorar o desempenho em relação ao DCA. Para o nosso método, a iterada usada para definir a direção de busca é gerada pelo MPP. Mais precisamente, o MPP para funções DC em variedades de Hadamard proposto por SOUZA e OLIVEIRA [48]. Fazemos a análise de convergência e apresentamos uma versão do nosso método para funções convexas que, por sua vez, também melhora os resultados em relação ao MPP clássico para funções convexas.

2.1 O MDCPPA

O método e os resultados de convergência apresentados no decorrer deste capítulo foram publicados em ALMEIDA *et al.* [1]. Para facilitar a identificação do nosso método, que será apresentado a seguir, nos referiremos ao mesmo por *MDCPPA*, abreviando *Modified DCPA*.

Seja $f : \mathcal{M} \rightarrow \mathbb{R}$ um função DC, com $f = f_1 - f_2$, onde $f_1, f_2 : \mathcal{M} \rightarrow \mathbb{R}$ são funções convexas de classe C^1 . Suponha que

$$\inf_{x \in \mathcal{M}} f(x) = f_1(x) - f_2(x) > -\infty.$$

Vamos supor ainda que f tem o gradiente Lipschitz com constante L , ou seja, existe $L > 0$ tal que $\text{grad}f$ satisfaz a desigualdade (1.3).

Algoritmo 1: MDCPPA

Passo 1: Tome um ponto inicial $x^0 \in \mathcal{M}$ e $\{c_k\}$ uma sequência limitada em \mathbb{R}_{++} com $c' \leq c_k \leq c''$, para todo $k \in \mathbb{N}$. Seja $k = 0$.

Passo 2: Defina

$$w^k := \text{grad} f_2(x^k) \quad \text{e} \quad z^k := \exp_{x^k}(c_k w^k). \quad (2.1)$$

Passo 3: Calcule

$$y^k := \arg \min_{x \in \mathcal{M}} \left\{ f_1(x) + \frac{1}{2c_k} d^2(x, z^k) \right\}. \quad (2.2)$$

Passo 4: Defina

$$d^k := -\exp_{y^k}^{-1} x^k.$$

Se $d^k = 0$ então PARE e RETORNE x^k . Caso contrário, vá para o Passo 5.

Passo 5: Calcule um comprimento de passo λ_k e defina

$$x^{k+1} := \exp_{y^k}(\lambda_k d^k).$$

Passo 6: Se $x^{k+1} = x^k$ então PARE e RETORNE x^k . Caso contrário, faça $k = k + 1$ e vá para o Passo 2.

Para o cálculo do comprimento de passo, consideramos as duas possibilidades a seguir:

1. **Busca de Armijo:** Fixe $\lambda > 0$ e $0 < \alpha < 1$. Escolha $\lambda_k := 2^{-i_k} \lambda$, onde i_k é o menor número inteiro positivo satisfazendo

$$f(\exp_{y^k}(\lambda_k d^k)) \leq f(y^k) - \alpha \lambda_k^2 d^2(x^k, y^k). \quad (2.3)$$

2. **Passo Fixo:** Fixe $\delta_1, \delta_2 > 0$ tal que $L\delta_1 + \delta_2 < 1$, onde L é a constante de Lipschitz associada ao gradiente de f . Tome $\{\lambda_k\}$ uma sequência de números

reais de modo que

$$\lambda_k \in \left(\delta_1, \frac{2}{L}(1 - \delta_2) \right), \quad \forall k \in \mathbb{N}.$$

Observação 3. *Note que se $\lambda_k = 0$ para todo k , então as iteradas do MDCPPA coincidem com as do DCPA em [48]. Além disso, se $f_2 \equiv 0$ então o MDCPPA “recupera” o método do ponto proximal clássico considerado em [18].*

Mais adiante, mostraremos que a direção d^k (definida no Passo 4 do MDCPPA) é uma direção de descida para f em y^k . Daí, a boa definição da Busca de Armijo, usada para calcular o comprimento de passo, é garantida em [23, Theorem 4.3], mais precisamente na Remark 4.5 feita após o teorema. Além disso, se λ_k é o comprimento de passo calculado usando a Busca de Armijo, então

$$0 < \lambda_k \leq \lambda, \quad \forall k \in \mathbb{N}.$$

Desde que estamos supondo que o gradiente de f é lipschitz contínuo, conseguimos garantir que

$$\liminf_{k \rightarrow +\infty} \lambda_k > 0.$$

Por outro lado, se o comprimento de passo for calculado usando a regra do Passo Fixo, então o [13, Theorem 5.1] garante que existe uma constante $\beta > 0$ tal que

$$f(\exp_{y^k}(\lambda_k d^k)) \leq f(y^k) - \beta \lambda_k^2 d^2(x^k, y^k), \quad \forall k \in \mathbb{N}. \quad (2.4)$$

Note que com a desigualdade definida acima, a regra do Passo Fixo pode ser visto como um caso particular da Busca de Armijo.

2.2 Análise de convergência

Apresentamos agora uma série de resultados que garantem a convergência do MDCPPA. Todos os resultados apresentados aqui podem ser encontrados em [1].

2.2.1 Caso não convexo

Proposição 2.1. *Para todo $k \in \mathbb{N}$, vale*

$$f(y^k) \leq f(x^k) - \frac{1}{c_k} d^2(x^k, y^k).$$

Demonstração. De (2.1) temos que

$$\operatorname{grad} f_2(x^k) = w^k = \frac{1}{c_k} \exp_{x^k}^{-1} z^k. \quad (2.5)$$

Olhando agora para (2.2), temos que y^k é a única solução da regularização de Moreau-Yosida da função f_1 com $\lambda = 1/c_k$ (ver Proposição 1.6). Assim,

$$\operatorname{grad} f_1(y^k) = \frac{1}{c_k} \exp_{y^k}^{-1} z^k. \quad (2.6)$$

Sendo as geodésicas como em (1.5) e usando o Teorema 1.13 temos que

$$\begin{aligned} f_1(x) &\geq f_1(y^k) + \frac{1}{c_k} \langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x \rangle \\ &\quad \text{e} \\ f_2(x) &\geq f_2(x^k) + \frac{1}{c_k} \langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} x \rangle, \end{aligned}$$

para qualquer que seja $x \in \mathcal{M}$. Fazendo então $x = x^k$ na primeira desigualdade, $x = y^k$ na segunda e somando as duas desigualdades, temos que

$$f(x^k) \geq f(y^k) + \frac{1}{c_k} \left(\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle + \langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k \rangle \right). \quad (2.7)$$

Seja agora o triângulo geodésico $\Delta(x^k, y^k, z^k)$. Defina por θ_1 o ângulo interno formado pelos segmentos $\exp_{y^k}^{-1} z^k$ e $\exp_{y^k}^{-1} x^k$, isto é, $\theta_1 = \angle(\exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k)$. Pelo Teorema 1.8 (Observação 2), temos que

$$d^2(z^k, y^k) + d^2(x^k, y^k) - 2\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle \leq d^2(z^k, x^k),$$

isto é,

$$\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle \geq \frac{1}{2} (d^2(z^k, y^k) + d^2(x^k, y^k) - d^2(z^k, x^k)).$$

No mesmo triângulo geodésico, defina $\theta_2 = \angle(\exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k)$. Usando novamente o Teorema 1.8, temos que

$$\langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k \rangle \geq \frac{1}{2} (d^2(z^k, x^k) + d^2(y^k, x^k) - d^2(z^k, y^k)).$$

Somando as duas desigualdades obtidas acima, temos que

$$\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle + \langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k \rangle \geq d^2(x^k, y^k). \quad (2.8)$$

Combinando com (2.7), temos o resultado desejado. \square

Proposição 2.2. Para todo $k \in \mathbb{N}$, temos que

$$\langle \text{grad} f(y^k), d^k \rangle \leq -\frac{1}{c_k} \|d^k\|^2,$$

isto é, d^k é uma direção de descida de f em y^k .

Demonstração. Como f_2 é uma função convexa, a Proposição 1.7 garante que

$$\langle \text{grad} f_2(y^k), \exp_{y^k}^{-1} x^k \rangle \leq \langle \text{grad} f_2(x^k), -\exp_{x^k}^{-1} y^k \rangle. \quad (2.9)$$

Daí,

$$\begin{aligned} \langle \text{grad} f(y^k), d^k \rangle &= \langle \text{grad} f_1(y^k) - \text{grad} f_2(y^k), -\exp_{y^k}^{-1} x^k \rangle \\ &= \langle \text{grad} f_1(y^k), -\exp_{y^k}^{-1} x^k \rangle + \langle \text{grad} f_2(y^k), \exp_{y^k}^{-1} x^k \rangle \\ &\leq \langle \text{grad} f_1(y^k), -\exp_{y^k}^{-1} x^k \rangle + \langle \text{grad} f_2(x^k), -\exp_{x^k}^{-1} y^k \rangle, \end{aligned}$$

onde a segunda igualdade vem da linearidade do produto interno e a desigualdade na terceira linha segue de (2.9). Substituindo $\text{grad} f_2(x^k)$ e $\text{grad} f_1(y^k)$ (já obtidos respectivamente em (2.5) e (2.6)) na desigualdade acima, obtemos:

$$\begin{aligned} \langle \text{grad} f(y^k), d^k \rangle &\leq \frac{1}{c_k} \langle \exp_{y^k}^{-1} z^k, -\exp_{y^k}^{-1} x^k \rangle + \frac{1}{c_k} \langle \exp_{x^k}^{-1} z^k, -\exp_{x^k}^{-1} y^k \rangle \\ &= -\frac{1}{c_k} \left(\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle + \langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k \rangle \right). \end{aligned} \quad (2.10)$$

Agora, considere o triângulo geodésico $\triangle(x^k, y^k, z^k)$. De (2.8) temos que

$$\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle + \langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k \rangle \geq \|d^k\|^2.$$

Multiplicando ambos os lados dessa desigualdade por $-1/c_k < 0$, chegamos a

$$-\frac{1}{c_k} \left(\langle \exp_{y^k}^{-1} z^k, \exp_{y^k}^{-1} x^k \rangle + \langle \exp_{x^k}^{-1} z^k, \exp_{x^k}^{-1} y^k \rangle \right) \leq -\frac{1}{c_k} \|d^k\|^2.$$

Substituindo em (2.10) obtemos

$$\langle \text{grad} f(y^k), d^k \rangle \leq -\frac{1}{c_k} \|d^k\|^2, \quad \forall k \in \mathbb{N},$$

o que conclui a prova. □

Proposição 2.3. Se $\{x^k\}$ é a sequência gerada pelo MDCPPA, então a sequência $\{f(x^k)\}$ é monótona decrescente e converge para algum $f^* \in \mathbb{R}$.

Demonstração. A princípio, faremos a demonstração usando a Busca de Armijo.

Assim, da desigualdade (2.3) temos que

$$\begin{aligned}
f(x^{k+1}) &= f(\exp_{y^k}(\lambda_k d^k)) \leq f(y^k) - \alpha \lambda_k^2 d^2(x^k, y^k) \\
&\leq f(x^k) - \frac{1}{c_k} d^2(x^k, y^k) - \alpha \lambda_k^2 d^2(x^k, y^k) \\
&= f(x^k) - \left(\frac{1}{c_k} + \alpha \lambda_k^2 \right) d^2(x^k, y^k),
\end{aligned} \tag{2.11}$$

onde a segunda desigualdade vem da Proposição 2.1. Assim,

$$f(x^{k+1}) < f(x^k), \quad \forall k \in \mathbb{N}.$$

Isso significa que $\{f(x^k)\}$ é uma sequência monótona decrescente em \mathbb{R} . Como f é limitada inferiormente, segue que $\{f(x^k)\}$ também é limitada inferiormente. Daí, existe $f^* \in \mathbb{R}$ tal que $f(x^k) \rightarrow f^*$ quando $k \rightarrow +\infty$.

Agora, se o comprimento de passo for calculado usando a regra do Passo Fixo, podemos usar a desigualdade (2.4) para garantir a existência de um $\beta > 0$ tal que

$$f(x^{k+1}) \leq f(x^k) - \left(\frac{1}{c_k} + \beta \lambda_k^2 \right) d^2(x^k, y^k), \quad \forall k \in \mathbb{N}. \tag{2.12}$$

Usando o mesmo argumento anterior, obtemos que a sequência $\{f(x^k)\}$ é monótona decrescente e limitada inferiormente. Logo, existe f^* tal que $f(x^k) \rightarrow f^*$ quando $k \rightarrow +\infty$, o que conclui a prova. \square

Proposição 2.4. *Seja $\{x^k\}$ uma sequência gerada pelo MDCPPA. Então,*

$$\lim_{k \rightarrow +\infty} d(x^k, y^k) = 0.$$

Consequentemente,

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

Demonstração. Faremos a demonstração usando a Busca de Armijo. Pela desigualdade (2.11), vale que

$$0 \leq \left(\frac{1}{c_k} + \alpha \lambda_k^2 \right) d^2(x^k, y^k) \leq f(x^k) - f(x^{k+1}).$$

Fazendo $k \rightarrow +\infty$ e tendo em mente a Proposição 2.3 temos que

$$\left(\frac{1}{c_k} + \alpha \lambda_k^2 \right) d^2(x^k, y^k) \rightarrow 0. \tag{2.13}$$

Observe que $\{c_k\}$ e $\{\lambda_k\}$ são sequências limitadas de números reais estritamente

positivos e que

$$\liminf_{k \rightarrow +\infty} \left(\frac{1}{c_k} + \alpha \lambda_k \right) > 0.$$

Usando isso em (2.13), temos que $d^2(x^k, y^k) \rightarrow 0$, quando $k \rightarrow +\infty$. Equivalentemente,

$$\lim_{k \rightarrow +\infty} d(x^k, y^k) = 0.$$

Para a segunda parte da proposição, note que

$$x^{k+1} = \exp_{y^k}(\lambda_k d^k) \quad \text{e} \quad d^k = -\exp_{y^k}^{-1} x^k$$

implicam em

$$d(x^{k+1}, y^k) = \lambda_k d(x^k, y^k).$$

Como $\{\lambda_k\} \subset \mathbb{R}_{++}$ é uma sequência limitada e já mostramos que $d(x^k, y^k) \rightarrow 0$ quando $k \rightarrow +\infty$, temos que $\lim_{k \rightarrow +\infty} d(x^{k+1}, y^k) = 0$. Agora, usando a desigualdade triangular,

$$0 \leq d(x^{k+1}, x^k) \leq d(x^{k+1}, y^k) + d(y^k, x^k).$$

Fazendo $k \rightarrow +\infty$ na desigualdade acima, temos que a soma do lado direito converge para 0. Portanto,

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

Assim como na demonstração da Proposição 2.3, uma prova análoga pode ser obtida aqui usando a regra do Passo Fixo. Basta usarmos a desigualdade obtida em (2.12) ao invés da (2.11). \square

Teorema 2.1. *Seja $\{x^k\}$ uma sequência gerada pelo MDCPPA. Então, todo ponto de acumulação de $\{x^k\}$, se existir, é um ponto crítico da função f .*

Demonstração. Seja \bar{x} um ponto de acumulação de $\{x^k\}$ e tome $\{x^{k_i}\}$ uma subsequência de $\{x^k\}$ que converge pra \bar{x} . Pela Proposição 2.4, y^{k_i} converge para \bar{x} . Como c_k é uma sequência limitada, todas as suas subsequências são limitadas. Em particular, a subsequência $\{c_{k_i}\}$ é limitada. Daí, sem perda de generalidade, podemos assumir que $\{c_{k_i}\}$ converge para um ponto, digamos c . Defina agora $\bar{w} = \text{grad} f_2(\bar{x})$. Como f_2 é uma função de classe C^1 , temos que $w^{k_i} = \text{grad} f_2(x^{k_i}) \rightarrow \bar{w}$, quando $i \rightarrow +\infty$. Então a subsequência $\{z^{k_i}\}$ converge para $\bar{z} = \exp_{\bar{x}}(c\bar{w})$ e

$$\text{grad} f_2(\bar{x}) = \bar{w} = \frac{1}{c} \exp_{\bar{x}}^{-1} \bar{z}.$$

Agora, desde que y^{k_i} é solução da regularização de Moreau-Yosida de f_1 em (2.2), temos que

$$\text{grad} f_1(y^{k_i}) = \frac{1}{c_{k_i}} \exp_{y^{k_i}}^{-1} z^{k_i}.$$

Sabendo que f_1 também é de classe C^1 , podemos fazer i tender a $+\infty$ na igualdade acima, obtendo que

$$\text{grad}f_1(\bar{x}) = \frac{1}{c} \exp_{\bar{x}}^{-1} \bar{z}.$$

Daí,

$$\text{grad}f(\bar{x}) = \text{grad}f_1(\bar{x}) - \text{grad}f_2(\bar{x}) = 0,$$

isto é, \bar{x} é ponto crítico de f . □

2.2.2 Caso convexo

Como já observado antes, se no MDCPPA temos $\lambda_k = 0$ para todo k e $f_2(x) = 0$ para todo x então recuperamos o MPP clássico (Observação 3). Já que a abordagem usada no nosso método é nova até mesmo para este caso, apresentamos aqui a versão do MDCPPA para funções convexas. Neste caso em particular, além de fazer a análise de convergência, calculamos a complexidade do método. Sabe-se que a complexidade do MPP e do método gradiente já foram calculadas no contexto Riemanniano e podem ser encontradas em BENTO *et al.* [9]. Lá é assumido que a curvatura seccional é não positiva para o MPP e não negativa para o método gradiente. Desde que o nosso método é uma combinação do MPP com um método de busca, não podemos usar essa abordagem. Porém, recentemente, WANG *et al.* [58] propôs uma nova abordagem para convergência de métodos subgradientes que lida com variedades de curvatura seccional positivas e negativas. Em contrapartida, é exigido que a variedade tenha curvatura seccional limitada inferiormente (ver também WANG *et al.* [59]). Esse tipo de abordagem foi usada por FERREIRA *et al.* [20] para obter a complexidade do método gradiente em variedades com curvatura limitada inferiormente e então generalizam a abordagem apresentada em CRUZ NETO *et al.* [13]. Com a proposta de calcular a complexidade do nosso método (que é nova, inclusive no contexto Euclideano), iremos usar a mesma abordagem que [20, 59].

Seja \mathcal{M} uma variedade de Hadamard com curvatura seccional K limitada inferiormente por $\kappa < 0$. Considere $f : \mathcal{M} \rightarrow \mathbb{R}$ uma função convexa de classe C^1 com

$$\inf_{x \in \mathcal{M}} f(x) > -\infty.$$

Suponha que o gradiente de f é lipschitz contínuo, com constante $L > 0$.

Apresentamos a versão do MDCPPA para funções convexas. Ele também encontra-se em [1], bem como a sua análise de convergência e cálculo da complexidade. Assim como o MDCPPA, mais adiante, iremos nos referir a este método como *MPPA (Modified PPA)*.

Algoritmo 2: MPPA

Passo 1: Tome um ponto inicial $x^0 \in \mathcal{M}$ e $\{c_k\}$ uma sequência limitada em \mathbb{R}_{++} com $c' \leq c_k \leq c''$, para todo $k \in \mathbb{N}$. Seja $k = 0$.

Passo 2: Calcule

$$y^k := \arg \min_{x \in \mathcal{M}} \left\{ f(x) + \frac{1}{2c_k} d^2(x, x^k) \right\}. \quad (2.14)$$

Passo 3: Defina

$$d^k := -\exp_{y^k}^{-1} x^k.$$

Se $d^k = 0$ então PARE e RETORNE x^k . Caso contrário, vá para o Passo 4.

Passo 4: Calcule um comprimento de passo λ_k assim como no MDCPPA e defina

$$x^{k+1} := \exp_{y^k}(\lambda_k d^k).$$

Passo 5: Se $x^{k+1} = x^k$ então PARE e RETORNE x^k . Caso contrário, faça $k = k + 1$ e vá para o Passo 2.

Denote $\hat{\kappa} = \sqrt{|\kappa|}$ e considere o seguinte conjunto:

$$S = \{x \in \mathcal{M} : f(x) \leq f(y), \quad \forall y \in \mathcal{M}\},$$

isto é, S é o conjunto dos pontos que minimizam f em \mathcal{M} . Vamos sempre supor que $S \neq \emptyset$.

Proposição 2.5. *Seja $\{x^k\}$ a sequência gerada pelo MPPA. Então, dado $x^* \in S$, temos que*

$$d(y^k, x^*) \leq d(x^k, x^*), \quad \forall k \in \mathbb{N}.$$

Demonstração. Para todo $k \in \mathbb{N}$ temos em (2.14) que y^k é solução da regularização de Moreau-Yosida de f com $\lambda = 1/c_k$. Então,

$$\text{grad}f(y^k) = \frac{1}{c_k} \exp_{y^k}^{-1} x^k.$$

Como f é convexa, vale o Teorema 1.13, isto é,

$$f(x) - f(y^k) \geq \frac{1}{c_k} \langle \exp_{y^k}^{-1} x^k \exp_{y^k}^{-1} x \rangle, \quad \forall x \in \mathcal{M}.$$

Em particular, tomando $x = x^*$, temos que

$$\frac{1}{c_k} \langle \exp_{y^k}^{-1} x^k \exp_{y^k}^{-1} x^* \rangle \leq f(x^*) - f(y^k) \leq 0.$$

Como $1/c_k > 0$, podemos concluir que

$$\langle \exp_{y^k}^{-1} x^k, \exp_{y^k}^{-1} x^* \rangle \leq 0. \quad (2.15)$$

Agora, considere o triângulo geodésico $\triangle(x^k, y^k, x^*)$. Pelo Teorema 1.8, vale que

$$d^2(x^k, y^k) + d^2(y^k, x^*) - 2\langle \exp_{y^k}^{-1} x^k, \exp_{y^k}^{-1} x^* \rangle \leq d^2(x^k, x^*),$$

isto é,

$$d^2(y^k, x^*) \leq d^2(x^k, x^*) - d^2(x^k, y^k) + 2\langle \exp_{y^k}^{-1} x^k, \exp_{y^k}^{-1} x^* \rangle.$$

Sabendo que a função distância é não negativa e usando a desigualdade (2.15), temos que

$$d^2(y^k, x^*) \leq d^2(x^k, x^*), \quad \forall k \in \mathbb{N}.$$

Usando novamente o fato da função distância ser não negativa, segue o resultado. \square

Enunciamos agora uma lema que será usado na proposição a seguir. Sua demonstração pode ser encontrada em [58].

Lema 2.2.

$$\cosh s - \cosh t \geq \frac{(s^2 - t^2) \sinh t}{2t}, \quad \forall s, t \geq 0.$$

Demonstração. Ver [58, Lemma 3.1]. \square

Nosso próximo resultado é de fundamental importância para o desenvolvimento do que vem a seguir. Ele consiste na adaptação do [20, Lemma 2.1] (que por sua vez é uma variante do [58, Lemma 3.2]) para o nosso caso.

Proposição 2.6. *Seja \mathcal{M} uma variedade Riemanniana completa de curvatura limitada inferiormente por $\kappa < 0$ e considere $\{x^k\}$ a sequência gerada pelo MPPA. Então, para todo $k \in \mathbb{N}$, vale*

$$\begin{aligned} & \cosh(\hat{\kappa} d(x^{k+1}, x)) \leq \cosh(\hat{\kappa} d(y^k, x)) \\ & + \hat{\kappa} \cosh(\hat{\kappa} d(y^k, x)) \sinh(\hat{\kappa} \lambda_k \|d^k\|) \left(\frac{\lambda_k \|d^k\|}{2} - \frac{c_k \tanh(\hat{\kappa} d(y^k, x)) (f(y^k) - f(x))}{\hat{\kappa} d(y^k, x) \|d^k\|} \right), \end{aligned}$$

para qualquer que seja $x \in \mathcal{M}$. Consequentemente,

$$\begin{aligned} d^2(x^{k+1}, x) & \leq d^2(y^k, x) \\ & + \frac{\sinh(\hat{\kappa} \lambda_k \|d^k\|)}{\hat{\kappa}} \left(\lambda_k \|d^k\| \frac{\hat{\kappa} d(y^k, x)}{\tanh(\hat{\kappa} d(y^k, x))} - \frac{2c_k}{\|d^k\|} (f(y^k) - f(x)) \right), \end{aligned}$$

para todo $k \in \mathbb{N}$ e todo x em \mathcal{M} .

Demonstração. Seja $\gamma : [0, +\infty) \rightarrow \mathcal{M}$ a geodésica dada por $\gamma(t) = \exp_{y^k}(td^k)$. Dado $x \in \mathcal{M}$, escolha uma geodésica τ tal que $l(\tau) = d(y^k, x)$. Considere a articulação geodésica $\Lambda(y^k, \gamma, \tau)$ em \mathcal{M} e defina θ como sendo o ângulo formado por γ e τ em y^k , isto é, $\theta := \angle_{y^k}(\gamma, \tau)$. Da Proposição 1.5 temos que existe uma articulação de comparação $\Lambda(\overline{y^k}, \overline{\gamma}, \overline{\tau})$ de $\Lambda(y^k, \gamma, \tau)$ em \mathcal{M}_κ^2 tal que

$$d(x^{k+1}, x) = d(\gamma(\lambda_k), x) \leq d(\overline{x^{k+1}}, \overline{x}),$$

onde $\overline{x^{k+1}}$ e \overline{x} denotam os pontos que delimitam as geodésicas $\overline{\gamma}$ e $\overline{\tau}$, respectivamente. Da definição de articulação de comparação temos que

$$\begin{aligned} d(\overline{y^k}, \overline{x^{k+1}}) &= d(y^k, \gamma(\lambda_k)) = d(y^k, x^{k+1}) = \lambda_k \|d^k\|, \\ d(\overline{y^k}, \overline{x}) &= d(y^k, x), \end{aligned}$$

com $\overline{\theta} = \angle_{\overline{y^k}}(\overline{\gamma}, \overline{\tau}) = \theta$. Seja $\Delta(\overline{x^{k+1}}, \overline{y^k}, \overline{x})$ o triângulo geodésico em \mathcal{M}_κ^2 . Pela Proposição 1.4, temos que

$$\begin{aligned} \cosh(\hat{\kappa}d(\overline{x^{k+1}}, \overline{x})) &= \cosh(\hat{\kappa}d(\overline{x}, \overline{y^k})) \cosh(\hat{\kappa}d(\overline{y^k}, \overline{x^{k+1}})) \\ &\quad - \sinh(\hat{\kappa}d(\overline{x}, \overline{y^k})) \sinh(\hat{\kappa}d(\overline{y^k}, \overline{x^{k+1}})) \cos \overline{\theta}. \end{aligned} \quad (2.16)$$

A função $\cosh : (0, +\infty] \rightarrow \mathbb{R}, t \mapsto \cosh(t)$, é uma função crescente. Além disso, usando ferramentas de Cálculo Elementar, prova-se que

$$\cosh(t) \leq 1 + \frac{t}{2} \sinh(t), \quad \forall t \geq 0.$$

Combinando esses fatos com a igualdade (2.16), temos que

$$\begin{aligned} \cosh(\hat{\kappa}d(x^{k+1}, x)) &\leq \cosh(\hat{\kappa}d(x, y^k)) \left(1 + \frac{\hat{\kappa}\lambda_k \|d^k\|}{2} \sinh(\hat{\kappa}\lambda_k \|d^k\|) \right) \\ &\quad - \sinh(\hat{\kappa}d(x, y^k)) \sinh(\hat{\kappa}\lambda_k \|d^k\|) \cos \theta, \end{aligned}$$

isto é,

$$\begin{aligned} \cosh(\hat{\kappa}d(x^{k+1}, x)) &\leq \cosh(\hat{\kappa}d(x, y^k)) \\ &\quad + \cosh(\hat{\kappa}d(x, y^k)) \sinh(\hat{\kappa}\lambda_k \|d^k\|) \left(\frac{\hat{\kappa}\lambda_k \|d^k\|}{2} - \tanh(\hat{\kappa}d(x, y^k)) \cos \theta \right). \end{aligned} \quad (2.17)$$

Agora, em (2.14), temos que y^k é solução da regularização de Moreau-Yosida de f

com $\lambda = 1/c_k$. Então,

$$\text{grad}f(y^k) = \frac{1}{c_k} \exp_{y^k}^{-1} x^k = -\frac{1}{c_k} d^k.$$

Desde que f é uma função convexa, podemos aplicar o Teorema 1.13, obtendo

$$f(x) - f(y^k) \geq -\frac{1}{c_k} \langle d^k, \exp_{y^k}^{-1} x \rangle. \quad (2.18)$$

Observe que $\theta = \angle_{y^k}(\gamma, \tau)$ implica que

$$\cos \theta = \frac{\langle d^k, \exp_{y^k}^{-1} x \rangle}{\|d^k\| \|\exp_{y^k}^{-1} x\|},$$

isto é,

$$\langle d^k, \exp_{y^k}^{-1} x \rangle = \|d^k\| d(x, y^k) \cos \theta.$$

Substituindo em (2.18) chegamos à seguinte desigualdade

$$-\cos \theta \leq -\frac{c_k(f(y^k) - f(x))}{\|d^k\| d(x, y^k)}.$$

Substituindo agora em (2.17), temos

$$\begin{aligned} \cosh(\hat{\kappa}d(x^{k+1}, x)) &\leq \cosh(\hat{\kappa}d(x, y^k)) \\ + \hat{\kappa} \cosh(\hat{\kappa}d(x, y^k)) \sinh(\hat{\kappa}\lambda_k \|d^k\|) &\left(\frac{\lambda_k \|d^k\|}{2} - \frac{c_k \tanh(\hat{\kappa}d(x, y^k)) (f(y^k) - f(x))}{\hat{\kappa}d(x, y^k) \|d^k\|} \right). \end{aligned} \quad (2.19)$$

Como $x \in \mathcal{M}$ foi tomado arbitrariamente, provamos a primeira parte da proposição.

Para a segunda parte, usaremos o Lema 2.2 com

$$s = \hat{\kappa}d(x^{k+1}, x) \quad \text{e} \quad t = \hat{\kappa}d(y^k, x).$$

Assim,

$$\begin{aligned} \cosh(\hat{\kappa}d(x^{k+1}, x)) - \cosh(\hat{\kappa}d(y^k, x)) &\geq \frac{((\hat{\kappa}d(x^{k+1}, x))^2 - (\hat{\kappa}d(y^k, x))^2) \sinh(\hat{\kappa}d(y^k, x))}{2\hat{\kappa}d(y^k, x)} \\ &= \frac{(d^2(x^{k+1}, x) - d^2(y^k, x)) \hat{\kappa} \sinh(\hat{\kappa}d(y^k, x))}{2d(y^k, x)}. \end{aligned}$$

Combinando com (2.19), temos que

$$\frac{(\mathrm{d}^2(x^{k+1}, x) - \mathrm{d}^2(y^k, x)) \hat{\kappa} \sinh(\hat{\kappa} \mathrm{d}(y^k, x))}{2\mathrm{d}(y^k, x)} \leq \hat{\kappa} \cosh(\hat{\kappa} \mathrm{d}(x, y^k)) \sinh(\hat{\kappa} \lambda_k \|d^k\|) \left(\frac{\lambda_k \|d^k\|}{2} - \frac{c_k \tanh(\hat{\kappa} \mathrm{d}(x, y^k)) (f(y^k) - f(x))}{\hat{\kappa} \mathrm{d}(x, y^k) \|d^k\|} \right),$$

ou seja,

$$\mathrm{d}^2(x^{k+1}, x) - \mathrm{d}^2(y^k, x) \leq \frac{\sinh(\hat{\kappa} \lambda_k \|d^k\|)}{\hat{\kappa}} \left(\lambda_k \|d^k\| \frac{\hat{\kappa} \mathrm{d}(x, y^k)}{\tanh(\hat{\kappa} \mathrm{d}(x, y^k))} - \frac{2c_k}{\|d^k\|} (f(y^k) - f(x)) \right),$$

para todo $x \in \mathcal{M}$. □

Para o próximo resultado, precisamos ter em mente o seguinte

Lema 2.3. *Seja $\{x^k\}$ a sequência gerada pelo MPPA. Então existe uma constante $\rho > 0$ tal que*

$$\sum_{k=0}^{+\infty} \lambda_k^2 \|d^k\|^2 \leq \rho.$$

Demonstração. Se usamos a Busca de Armijo para calcular o comprimento de passo, então, por definição, a desigualdade

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

é válida para todo $k \in \mathbb{N}$. Somando esta desigualdade de 0 a N temos que

$$\alpha \sum_{k=0}^N \lambda_k^2 \|d^k\|^2 \leq f(x^0) - f(x^{N+1}) \leq f(x^0) - f^*,$$

onde $f^* = \inf_{x \in \mathcal{M}} f(x)$. Desde que $\alpha > 0$, segue que

$$\sum_{k=0}^N \lambda_k^2 \|d^k\|^2 \leq \frac{f(x^0) - f^*}{\alpha}.$$

Assim, tomando $\rho = (f(x^0) - f^*)/\alpha > 0$ e fazendo N tender a $+\infty$ temos que

$$\sum_{k=0}^{+\infty} \lambda_k^2 \|d^k\|^2 \leq \rho,$$

como queríamos.

Por outro lado, se o comprimento de passo for calculado pela regra do Passo Fixo, então usamos a desigualdade (2.4) e, procedendo de forma análoga, garantimos a

existência de $\beta > 0$ tal que

$$\sum_{k=0}^N \lambda_k^2 \|d^k\|^2 \leq \frac{f(x^0) - f^*}{\beta},$$

o que implica a existência de $\rho = (f(x^0) - f^*)/\beta > 0$ satisfazendo a hipótese. Segue o resultado. \square

Uma consequência imediata do lema acima é que

$$\lambda_k \|d^k\| \leq \sqrt{\rho}, \quad \forall k \in \mathbb{N}, \quad (2.20)$$

onde $\rho > 0$ denotará, a partir de agora, a constante assumida no lema.

Assim como na Proposição 2.6, nosso próximo resultado é a ‘variação’ de um resultado já obtido em [20, Lemma 3.9]. Aqui, adaptamos novamente o lema em questão para o nosso caso.

Proposição 2.7. *Seja $\{x^k\}$ uma sequência gerada pelo MPPA. Se $x^* \in S$ então*

$$d(x^{k+1}, x^*) \leq \frac{1}{\hat{\kappa}} \cosh^{-1}(\cosh(\hat{\kappa} d(x^0, x^*)) e^{\frac{1}{2}\hat{\kappa}\sqrt{\rho}\sinh(\hat{\kappa}\sqrt{\rho})}), \quad \forall k \in \mathbb{N}.$$

Demonstração. Fazendo $x = x^*$ na primeira desigualdade da Proposição 2.6, temos válida a seguinte desigualdade

$$\cosh(\hat{\kappa} d(x^{k+1}, x^*)) \leq \cosh(\hat{\kappa} d(y^k, x^*)) + \frac{\hat{\kappa}\lambda_k \|d^k\|}{2} \cosh(\hat{\kappa} d(y^k, x^*)) \sinh(\hat{\kappa}\lambda_k \|d^k\|),$$

isto é,

$$\cosh(\hat{\kappa} d(x^{k+1}, x^*)) \leq \cosh(\hat{\kappa} d(y^k, x^*)) \left(1 + (\hat{\kappa}\lambda_k \|d^k\|)^2 \frac{\sinh(\hat{\kappa}\lambda_k \|d^k\|)}{2\hat{\kappa}\lambda_k \|d^k\|} \right), \quad \forall k \in \mathbb{N}.$$

Sabendo que a função

$$t \mapsto \frac{\sinh(t)}{t}$$

é uma função crescente para todo $t > 0$ e tendo em mente que vale (2.20), chegamos em

$$\begin{aligned} \cosh(\hat{\kappa} d(x^{k+1}, x^*)) &\leq \cosh(\hat{\kappa} d(y^k, x^*)) \left(1 + (\hat{\kappa}\lambda_k \|d^k\|)^2 \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{2\hat{\kappa}\sqrt{\rho}} \right) \\ &= \cosh(\hat{\kappa} d(y^k, x^*)) \left(1 + (\lambda_k \|d^k\|)^2 \frac{\hat{\kappa} \sinh(\hat{\kappa}\sqrt{\rho})}{2\sqrt{\rho}} \right). \end{aligned}$$

Defina $r := \hat{\kappa} \sinh(\hat{\kappa}\sqrt{\rho})/(2\sqrt{\rho})$. Então

$$\cosh(\hat{\kappa}d(x^{k+1}, x^*)) \leq \cosh(\hat{\kappa}d(y^k, x^*))(1 + r(\lambda_k \|d^k\|)^2).$$

Sabendo que

$$e^t \geq 1 + t, \quad \forall t \geq 0,$$

temos que

$$\begin{aligned} \cosh(\hat{\kappa}d(x^{k+1}, x^*)) &\leq \cosh(\hat{\kappa}d(y^k, x^*))e^{r(\lambda_k \|d^k\|)^2} \\ &\leq \cosh(\hat{\kappa}d(x^k, x^*))e^{r(\lambda_k \|d^k\|)^2}, \end{aligned}$$

onde a última desigualdade ocorre devido a Proposição 2.5. Usando esta desigualdade recursivamente, temos que

$$\begin{aligned} \cosh(\hat{\kappa}d(x^{k+1}, x^*)) &\leq \cosh(\hat{\kappa}d(x^k, x^*))e^{r(\lambda_k \|d^k\|)^2} \\ &\leq \cosh(\hat{\kappa}d(x^{k-1}, x^*))e^{r(\lambda_{k-1} \|d^{k-1}\|)^2} e^{r(\lambda_k \|d^k\|)^2} \\ &\quad \vdots \\ &\leq \cosh(\hat{\kappa}d(x^0, x^*))e^{r \sum_{i=0}^k (\lambda_i \|d^i\|)^2} \\ &\leq \cosh(\hat{\kappa}d(x^0, x^*))e^{r\rho}. \end{aligned}$$

Então,

$$\begin{aligned} d(x^{k+1}, x^*) &\leq \frac{1}{\hat{\kappa}} \cosh^{-1}(\cosh(\hat{\kappa}d(x^0, x^*))e^{r\rho}) \\ &= \frac{1}{\hat{\kappa}} \cosh^{-1}(\cosh(\hat{\kappa}d(x^0, x^*))e^{\frac{1}{2}\hat{\kappa}\sqrt{\rho}\sinh(\hat{\kappa}\sqrt{\rho})}), \end{aligned}$$

para todo $k \in \mathbb{N}$. □

O resultado a seguir é uma adaptação do [20, Lemma 3.10].

Proposição 2.8. *Seja $\{x^k\}$ a sequência gerada pelo MPPA. Se $x^* \in S$ então,*

$$d^2(x^{k+1}, x^*) \leq d^2(x^k, x^*) + \frac{\mathcal{C}^{x^*}_{\kappa, \rho}}{\nu} (f(x^k) - f(x^{k+1})) + 2\lambda_k c_k (f(x^*) - f(x^{k+1})),$$

para todo $k \in \mathbb{N}$, onde

$$\mathcal{C}^{x^*}_{\kappa, \rho} = \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{\hat{\kappa}\sqrt{\rho}} \left(1 + \hat{\kappa} \cosh^{-1}(\cosh(\hat{\kappa}d(x^0, x^*))e^{\frac{1}{2}\hat{\kappa}\sqrt{\rho}\sinh(\hat{\kappa}\sqrt{\rho})}) \right). \quad (2.21)$$

Demonstração. Tome $x = x^*$ na segunda desigualdade da Proposição 2.6. Fazendo

algumas manipulações algébricas obtemos que

$$\begin{aligned} d^2(x^{k+1}, x^*) &\leq d^2(y^k, x^*) \\ &+ \frac{\sinh(\hat{\kappa}\lambda_k\|d^k\|)}{\hat{\kappa}\lambda_k\|d^k\|} \left((\lambda_k\|d^k\|)^2 \frac{\hat{\kappa}d(y^k, x^*)}{\tanh(\hat{\kappa}d(y^k, x^*))} + 2c_k\lambda_k(f(x^*) - f(y^k)) \right). \end{aligned}$$

Sabendo que

$$\frac{t}{\tanh(t)} \leq 1 + t, \quad \forall t \geq 0,$$

e que $\sinh(t)/t$ é uma função crescente para $t > 0$ e limitada inferiormente por 1 temos que

$$\begin{aligned} d^2(x^{k+1}, x^*) &\leq d^2(y^k, x^*) \\ &+ \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{\hat{\kappa}\sqrt{\rho}} (\lambda_k\|d^k\|)^2 (1 + \hat{\kappa}d(y^k, x^*)) + 2c_k\lambda_k(f(x^*) - f(y^k)). \end{aligned}$$

Observe que

$$\begin{aligned} d^2(x^{k+1}, x^*) &\leq d^2(y^k, x^*) \\ &+ \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{\hat{\kappa}\sqrt{\rho}} \frac{(f(y^k) - f(x^{k+1}))}{\nu} (1 + \hat{\kappa}d(y^k, x^*)) + 2c_k\lambda_k(f(x^*) - f(y^k)), \end{aligned}$$

onde $\nu = \alpha$ se usamos a Busca de Armijo ou $\nu = \beta$ se usamos a regra do Passo Fixo. Usando agora a Proposição 2.7, temos

$$\begin{aligned} d^2(x^{k+1}, x^*) &\leq d^2(y^k, x^*) \\ &+ \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{\hat{\kappa}\sqrt{\rho}} \frac{(f(y^k) - f(x^{k+1}))}{\nu} \left(1 + \hat{\kappa} \cosh^{-1}(\cosh(\hat{\kappa}d(x^0, x^*))e^{\frac{1}{2}\hat{\kappa}\sqrt{\rho}\sinh(\hat{\kappa}\sqrt{\rho})}) \right) \\ &+ 2c_k\lambda_k(f(x^*) - f(y^k)). \quad (2.22) \end{aligned}$$

Observe que usando desigualdade na definição da Busca de Armijo ou a desigualdade (2.4) da regra do Passo Fixo temos que $f(x^{k+1}) \leq f(y^k)$ para todo k . Além disso, da Proposição 2.1, temos que $f(y^k) \leq f(x^k)$. Portanto, a desigualdade

$$f(x^{k+1}) \leq f(y^k) \leq f(x^k),$$

é válida para todo $k \in \mathbb{N}$. Assim, definindo

$$C_{\kappa, \rho}^{x^*} := \frac{\sinh(\hat{\kappa}\sqrt{\rho})}{\hat{\kappa}\sqrt{\rho}} \left(1 + \hat{\kappa} \cosh^{-1}(\cosh(\hat{\kappa}d(x^0, x^*))e^{\frac{1}{2}\hat{\kappa}\sqrt{\rho}\sinh(\hat{\kappa}\sqrt{\rho})}) \right),$$

e tendo em mente a Proposição 2.5, concluímos que a desigualdade (2.22) é equiva-

lente a

$$d^2(x^{k+1}, x^*) \leq d^2(x^k, x^*) + \frac{\mathcal{C}^{x^*}}{\nu}(f(x^k) - f(x^{k+1})) + 2c_k \lambda_k (f(x^*) - f(x^{k+1})),$$

para todo $k \in \mathbb{N}$, concluindo assim a demonstração. \square

O teorema a seguir é o principal resultado dessa seção. Nele garantimos a convergência da sequência $\{x^k\}$ para um minimizador de f .

Teorema 2.4. *Seja $\{x^k\}$ a sequência gerada pelo MPPA. Então $\{x^k\}$ é quasi-Fejér convergente a S . Consequentemente, $\{x^k\}$ converge para um ponto de S .*

Demonstração. Pela Proposição 2.8, dado $x^* \in S$,

$$d^2(x^{k+1}, x^*) \leq d^2(x^k, x^*) + \frac{\mathcal{C}^{x^*}}{\nu}(f(x^k) - f(x^{k+1})), \quad \forall k \in \mathbb{N},$$

onde \mathcal{C}^{x^*} é a constante positiva definida em (2.21). Tomando

$$\epsilon_k = \frac{\mathcal{C}^{x^*}}{\nu}(f(x^k) - f(x^{k+1})),$$

e somando de 0 a N , temos que

$$\sum_{k=0}^N \epsilon_k = \sum_{k=0}^N \frac{\mathcal{C}^{x^*}}{\nu}(f(x^k) - f(x^{k+1})) = \frac{\mathcal{C}^{x^*}}{\nu}(f(x^0) - f(x^{N+1})).$$

Sendo $f^* = \inf_{x \in \mathcal{M}} f(x)$, podemos tomar N ao $+\infty$ para obter que

$$\sum_{k=0}^{+\infty} \epsilon_k \leq \frac{\mathcal{C}^{x^*}}{\nu}(f(x^0) - f^*) < +\infty.$$

Isso implica, por definição, que a sequência $\{x^k\}$ é quasi-Fejér convergente ao conjunto S e, pelo Lema 1.4, $\{x^k\}$ é uma sequência limitada.

Se $\{x^k\}$ possui pontos de acumulação, então esses são pontos críticos de f (ver Teorema 2.1). Como f é convexa (e limitada inferiormente), pontos críticos são pontos de mínimo. Daí, o subconjunto S coincide com o conjunto dos pontos críticos de f . Portanto, se existe \bar{x} ponto de acumulação de $\{x^k\}$, então $\bar{x} \in S$, ou seja, $\{x^k\}$ converge para um ponto de S . \square

Teorema 2.5. *Seja $\{x^k\}$ a sequência gerada pelo MPPA. Dado $x^* \in S$, defina $f^* = f(x^*)$. Então*

$$f(x^{N+1}) - f^* \leq \frac{(\lambda' + 1)[\nu d^2(x^0, x^*) + \mathcal{C}^{x^*}(f(x^0) - f^*)]}{4c'\nu\lambda'N}, \quad \forall N \in \mathbb{N}.$$

Demonstração. Usando o Teorema 1.8 no triângulo geodésico $\Delta(x^k, y^k, x^*)$, temos que

$$d^2(x^k, y^k) + d^2(y^k, x^*) - 2\langle \exp_{y^k}^{-1} x^k, \exp_{y^k}^{-1} x^* \rangle \leq d^2(x^k, x^*), \quad \forall k \in \mathbb{N}.$$

Como f é uma função convexa de classe C^1 com $\text{grad}f(y^k) = (1/c_k) \exp_{y^k}^{-1} x^k$, podemos aplicar o Teorema 1.13 para obter que

$$f(y^k) - f^* \leq \frac{1}{2c_k} (d^2(x^k, x^*) - d^2(y^k, x^*) - d^2(x^k, y^k)).$$

Usando a desigualdade (2.22), temos que

$$f(y^k) - f^* \leq \frac{1}{2c_k} \left(d^2(x^k, x^*) - d^2(x^{k+1}, x^*) + \frac{\mathcal{C}_{\kappa, \rho}^{x^*}}{\nu} (f(x^k) - f(x^{k+1})) \right),$$

onde $\mathcal{C}_{\kappa, \rho}^{x^*}$ e ν foram definidos na Proposição 2.8. Somando esta desigualdade de 0 a N , obtemos

$$\begin{aligned} \sum_{k=0}^N (f(y^k) - f^*) &\leq \frac{1}{2c'} \left(d^2(x^0, x^*) - d^2(x^{N+1}, x^*) + \frac{\mathcal{C}_{\kappa, \rho}^{x^*}}{\nu} (f(x^0) - f(x^{N+1})) \right) \\ &\leq \frac{\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*} (f(x^0) - f^*)}{2c'\nu}. \end{aligned}$$

Observe que sempre que $N > k$ temos $f(y^N) \leq f(y^k)$. Daí,

$$N(f(y^N) - f^*) \leq \sum_{k=0}^N (f(y^k) - f^*).$$

Segue que

$$f(y^N) - f^* \leq \frac{\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*} (f(x^0) - f^*)}{2c'\nu N}, \quad \forall N \in \mathbb{N}. \quad (2.23)$$

Agora, da Proposição 2.8 temos que

$$2\lambda'c'(f(x^{k+1}) - f^*) \leq d^2(x^k, x^*) - d^2(x^{k+1}, x^*) + \frac{\mathcal{C}_{\kappa, \rho}^{x^*}}{\nu} (f(x^k) - f(x^{k+1})),$$

onde $\lambda' = \min_{k \in \mathbb{N}} \lambda_k > 0$. Somando esta desigualdade de 0 a N temos

$$\begin{aligned} 2\lambda'c' \sum_{k=0}^N (f(x^{k+1}) - f^*) &\leq d^2(x^0, x^*) - d^2(x^{N+1}, x^*) + \frac{\mathcal{C}_{\kappa, \rho}^{x^*}}{\nu} (f(x^0) - f(x^{N+1})) \\ &\leq d^2(x^0, x^*) + \frac{\mathcal{C}_{\kappa, \rho}^{x^*}}{\nu} (f(x^0) - f(x^{N+1})). \end{aligned}$$

Como $\{f(x^k)\}$ é uma sequência monótona decrescente, segue que

$$f(x^N) - f^* \leq \frac{\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*}(f(x^0) - f^*)}{2c'\nu\lambda'N}, \quad \forall N \in \mathbb{N}. \quad (2.24)$$

Observe agora que a complexidade do nosso método é a soma das complexidades nas suas duas fases, isto é, a soma das complexidades (2.23) e (2.24). Então,

$$\begin{aligned} 2(f(x^{N+1}) - f^*) &\leq (f(y^N) - f^*) + (f(x^N) - f^*) \\ &\leq \frac{\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*}(f(x^0) - f^*)}{2c'\nu N} + \frac{\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*}(f(x^0) - f^*)}{2c'\nu\lambda'N} \\ &= \frac{(\lambda' + 1)(\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*}(f(x^0) - f^*))}{2c'\nu\lambda'N}, \end{aligned}$$

isto é,

$$f(x^{N+1}) - f^* \leq \frac{(\lambda' + 1)(\nu d^2(x^0, x^*) + \mathcal{C}_{\kappa, \rho}^{x^*}(f(x^0) - f^*))}{4c'\nu\lambda'N}, \quad \forall N \in \mathbb{N}.$$

□

Capítulo 3

Experimentos numéricos

Neste capítulo, apresentamos alguns gráficos e tabelas buscando ilustrar o desempenho e eficiência do nosso método em relação a outros.

3.1 As matrizes simétricas positivas definidas

Seja \mathbb{S}^n o conjunto das matrizes simétricas $n \times n$. Denotamos por \mathbb{S}_+^n o conjunto das matrizes simétricas *positivas semidefinidas* e por \mathbb{S}_{++}^n o das *positivas definidas*. O conjunto \mathbb{S}_{++}^n é um *cone convexo*, isto é,

$$X, Y \in \mathbb{S}_{++}^n \quad \Rightarrow \quad X + tY \in \mathbb{S}_{++}^n, \forall t \in \mathbb{R}_{++}.$$

Além disso, \mathbb{S}_{++}^n é um subconjunto aberto de \mathbb{S}^n . Então, para cada $X \in \mathbb{S}_{++}^n$, o espaço tangente a \mathbb{S}_{++}^n em X é identificado por \mathbb{S}^n (ver [36]). Neste capítulo, consideramos a variedade Riemanniana $\mathcal{M} := (\mathbb{S}_{++}^n, \langle \cdot, \cdot \rangle)$ onde

$$\langle V, W \rangle := \text{tr}(X^{-1}VX^{-1}W), \quad X \in \mathcal{M}, \quad V, W \in T_X\mathcal{M} \approx \mathbb{S}^n, \quad (3.1)$$

com $\text{tr}(X)$ denotando o *traço* de $X \in \mathbb{S}^n$ (ver [42]). A norma correspondente é denotada por $\|\cdot\|$.

Afirmamos que \mathcal{M} é uma variedade de Hadamard (ver [30, Theorem 1.2, p. 325]) com curvatura seccional limitada inferiormente (ver [31, p. 428]). Dados $X, Y \in \mathcal{M}$, a única geodésica que liga esses dois pontos em \mathcal{M} é dada por

$$\gamma(t) = X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2}, \quad t \in [0, 1]$$

(ver [39, Theorem 6.1]). A aplicação exponencial nessa variedade é dada por

$$\exp_X V = X^{1/2}e^{X^{-1/2}VX^{-1/2}}X^{1/2}.$$

Definição 2. Uma retração em uma variedade \mathcal{M} é uma aplicação diferenciável R que vai de $T\mathcal{M}$ em \mathcal{M} satisfazendo as seguintes propriedades. Denote por R_x a restrição de R a $T_x\mathcal{M}$:

1. $R_x(0_x) = x$, onde 0_x denota o zero de $T_x\mathcal{M}$.
2. $\frac{d}{dt}R_x(t\xi) = \xi$ para todo $\xi \in T_x\mathcal{M}$.

A aplicação exponencial definida em uma variedade é um exemplo especial de retração. Dizemos que ela é uma retração de *primeira ordem*. Porém, aplicações exponenciais tem um certo “custo” computacional. Esse custo pode ser otimizado através da escolha de uma retração adequada. Pensando nisso, para efeito de comparação, incluímos testes substituindo a aplicação exponencial dada anteriormente pela seguinte retração:

$$R_X : T_X\mathcal{M} \rightarrow \mathcal{M},$$

$$R_X(V) = X + V + \frac{1}{2}VX^{-1}V.$$

Dada $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ uma função de classe C^1 , temos que

$$\text{grad}f(X) = Xf'(X)X,$$

onde $f'(X)$ é o gradiente Euclidiano de f em X . Se f é de classe C^2 , então para cada $V \in T_X\mathcal{M}$,

$$\text{Hess}f(X)V = Xf''(X)VX + \frac{1}{2}(Vf'(X)X + Xf'(X)V),$$

onde $f''(X)$ é a Hessiana Euclidiana de f em X .

3.2 Aplicando o método em uma função convexa

Seja $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ a função real dada por

$$f(X) = a \ln(\det(X))^2, \quad a \in \mathbb{R}_{++}.$$

A função f definida acima é um exemplo que reforça a importância da generalização de métodos para espaços mais gerais, como o proposto nesta tese. Mostraremos que ela não é convexa no sentido Euclidiano, mas se adotamos a métrica dada em (3.1) ela passa a ser convexa e, assim, podemos aplicar o nosso método para encontrar o seu minimizador.

O gradiente e a Hessiana de f no contexto Euclidiano são dados, respectivamente, por

$$\begin{aligned} f'(X) &= 2a \ln(\det(X))X^{-1} \\ f''(X)V &= 2a \operatorname{atr}(X^{-1}V)X^{-1} - 2a \ln(\det(X))X^{-1}VX^{-1}. \end{aligned}$$

Denote por \mathbf{I}_n a matriz *identidade* de ordem $n \times n$. Dado $t \in \mathbb{R}_{++}$, tome $X = t\mathbf{I}_n$ e $V = \mathbf{I}_n$. Então,

$$f''(t\mathbf{I}_n)\mathbf{I}_n = \frac{2an}{t^2}(1 - \ln(t))\mathbf{I}_n.$$

Isso implica que, se t é tomado suficientemente grande, $f''(t\mathbf{I}_n)$ não é definida positiva e, portanto, f não é convexa no sentido Euclidiano. Por outro lado, considerando a métrica Riemanniana definida em (3.1), temos que

$$\begin{aligned} \operatorname{grad}f(X) &= 2a \ln(\det(X))X \\ \operatorname{Hess}f(X)V &= 2a \operatorname{atr}(X^{-1}V)X. \end{aligned}$$

Observe que para cada $X \in \mathcal{M}$,

$$\langle \operatorname{Hess}f(X)V, V \rangle = 2a \operatorname{atr}(X^{-1}V)\langle X, V \rangle = 2a \operatorname{atr}(X^{-1}V)^2 > 0, \quad \forall V \in T_X\mathcal{M}.$$

Portanto, f é convexa em \mathcal{M} .

Para os testes nesta seção, consideramos $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$, dada por

$$f(X) = \frac{1}{2} \ln(\det(X))^2.$$

Note que

$$\operatorname{grad}f(X) = 0 \Leftrightarrow \ln(\det(X)) = 0 \Leftrightarrow \det(X) = 1.$$

Então os pontos críticos de f são todas as matrizes $X \in \mathbb{S}_{++}^n$ tais que $\det(X) = 1$.

Para os testes, comparamos nosso método com o MPP clássico proposto em [18], que será referido aqui como *Classic PPA*. Por conveniência, chamaremos de *MPPA* o método proposto neste trabalho para o caso convexo e de *MPPA retr* a sua versão usando retração para definir x^{k+1} . Implementamos esses métodos no Matlab R2017a em uma máquina com Intel(R) Core(TM) i3-6006U, CPU 2.00GHz e 4,00GB de memória ram. Para resolver os subproblemas e manusear variedades no Matlab foi utilizado a toolbox *Manopt* (ver BOUMAL *et al.* [11]). O 'solver' do Manopt utilizado para resolver os subproblemas foi o *SteepestDescent*. Os métodos foram inicializados 15 vezes para cada tamanho de matriz, isto é, cada método partiu 15 vezes uma matriz 5×5 , depois mais 15 vezes de uma matriz 10×10 , com o mesmo acontecendo com matrizes 20×20 e 50×50 . Em cada uma das

rodadas, uma matriz $X \in \mathbb{S}_{++}^n$ era gerada aleatoriamente e cada método tinha como ponto inicial a mesma matriz. Essas matrizes eram geradas de modo que seus autovalores estavam distribuídos (aleatoriamente) no intervalo $[1, 5]$. A sequência $\{c_k\}$ foi tomada como sendo $c_k = 1$ para todo k . A busca linear adotada foi a busca de Armijo com parâmetros $\lambda = 1$ e $\alpha = 0.5$. Para o critério de parada, utilizamos $d(x^{k+1}, x^k) < \epsilon$, com $\epsilon = 10^{-5}$. Os resultados obtidos nos testes são apresentados a seguir e encontram-se divididos em 4 grupos, de acordo com o tamanho da matriz adotada no teste, que foram matrizes 5×5 , 10×10 , 20×20 e 50×50 . No primeiros 2 quadros de cada grupo podemos comparar a quantidade de iteradas e o tempo gastos, respectivamente, em cada uma das 15 rodadas dos métodos. Já no terceiro quadro do grupo, temos o gráfico ‘ $\|\text{grad}f(x^k)\| \times k$ ’, obtido da última rodada do respectivo método. Nele podemos conferir o comportamento da norma do gradiente de f obtido em cada iterada.

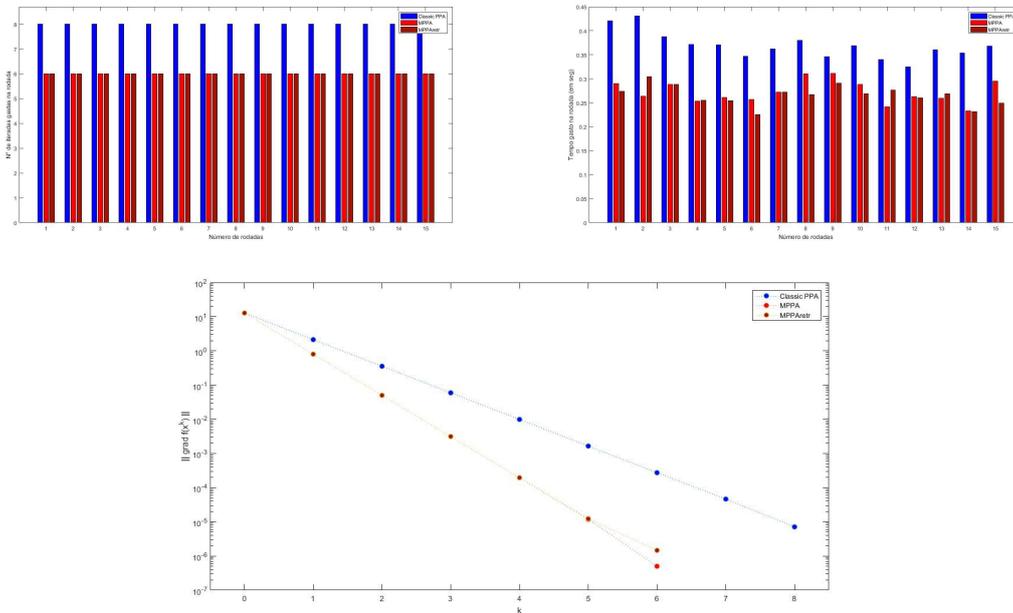


Figura 3.1: Caso convexo com matrizes 5×5

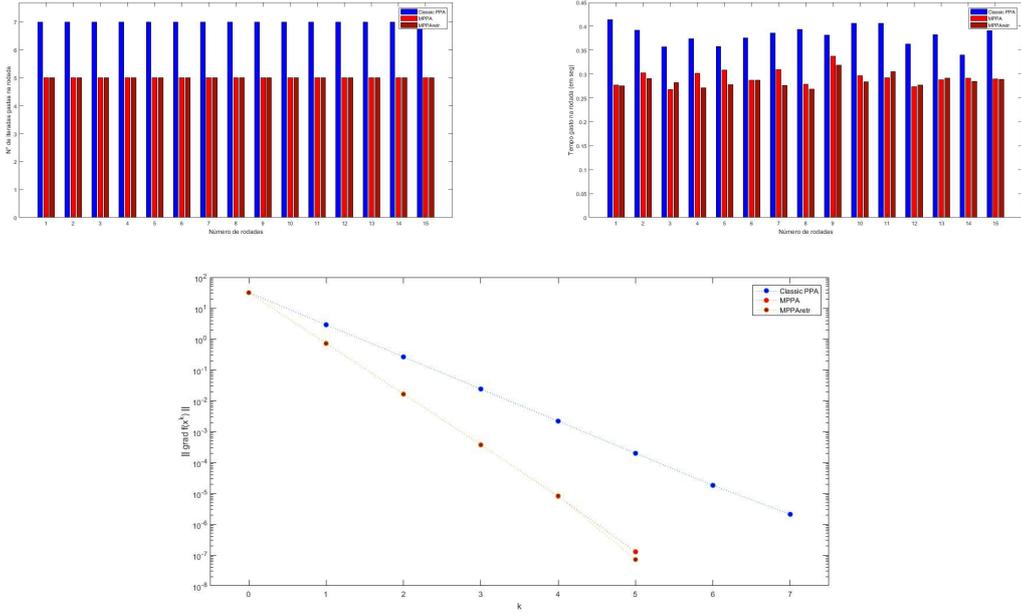


Figura 3.2: Caso convexo com matrizes 10×10

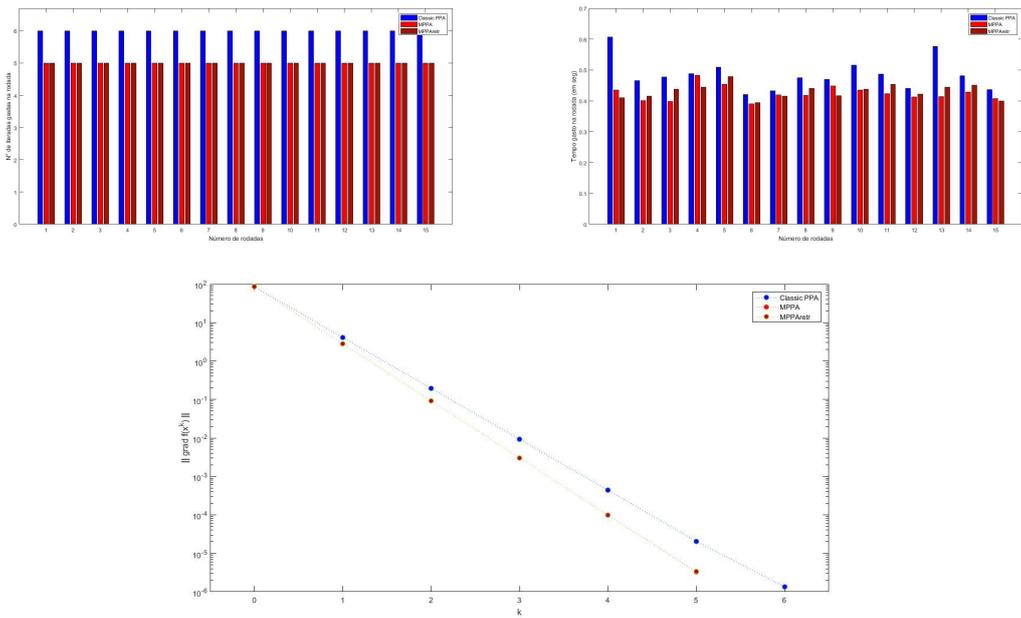


Figura 3.3: Caso convexo com matrizes 20×20

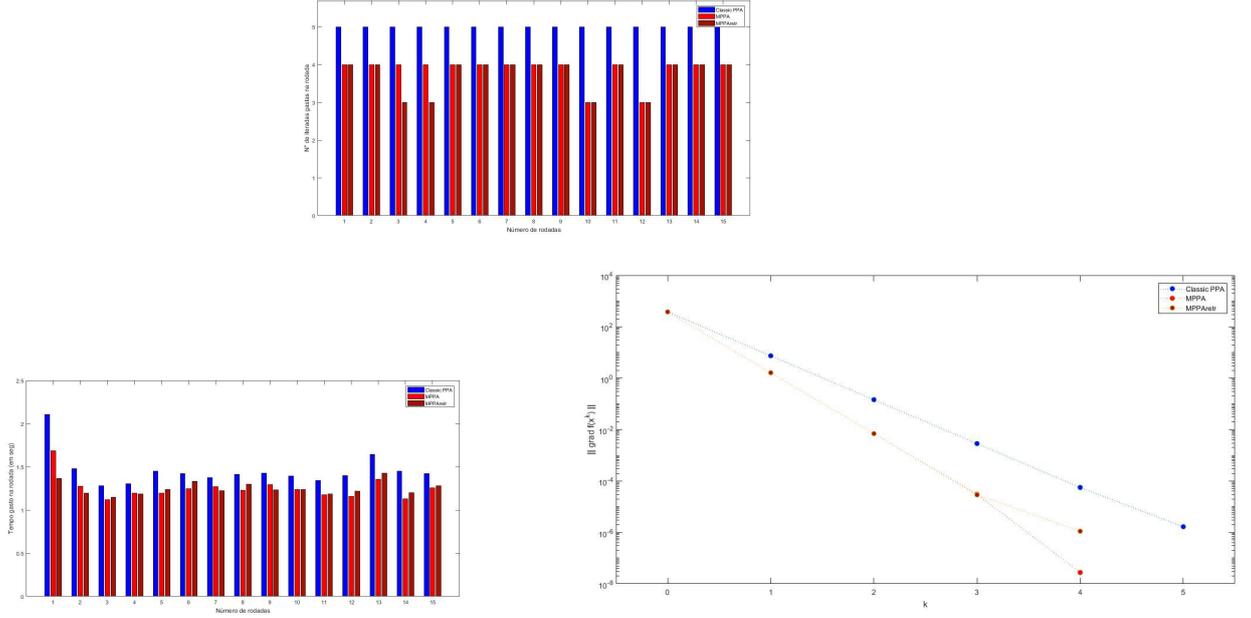


Figura 3.4: Caso convexo com matrizes 50×50

Nos gráficos apresentados acima podemos ver que nosso método melhora o Classic PPA. Melhora no sentido de que conseguimos uma aproximação ao ponto desejado, no caso minimizador da função, gastando menos iteradas e, conseqüentemente, menos tempo de processamento da CPU. Um dos motivos dessa melhora pode ser explicado pelo fato do valor funcional no ponto obtido pelo nosso método em cada iterada ser sempre menor que o obtido pelo Classic PPA.

As tabelas a seguir reforçam esse panorama de melhora do nosso método. Na primeira tabela, *min* indica o menor número de iteradas gastas contando todas as rodadas do respectivo método e *med* indica a média de iteradas gastas. Na segunda tabela, *min* e *med* seguem a mesma idéia, mas em relação ao tempo gasto.

n	PPA		MPPA			
	min	med	exp		retr	
			min	med	min	med
5	8	8	6	6	6	6
10	7	7	5	5	5	5
20	6	6	5	5	5	5
50	5	5	3	4	3	4

Tabela 3.1: Média do número de iteradas para cada matriz $n \times n$ no caso convexo

n	PPA		MPPA			
	min	med	exp		retr	
			min	med	min	med
5	0.3251	0.3678	0.2334	0.2634	0.2254	0.2688
10	0.3398	0.3827	0.2678	0.2913	0.2685	0.2836
20	0.4209	0.4773	0.3902	0.4189	0.3945	0.4383
50	1.2816	1.4203	1.1198	1.2396	1.1486	1.2317

Tabela 3.2: Média do tempo gasto (em seg) para cada cada matriz $n \times n$ no caso convexo

3.3 Aplicando o método em funções DC

Nesta seção vamos aplicar o nosso método a duas funções DC. A primeira função, apresentada a seguir, é uma função DC que não é convexa (nem no contexto Euclidiano nem no Riemanniano), enquanto que a segunda será uma função DC convexa, que faremos testes usando o MDCPPA.

Seja $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ dada por

$$f(X) = a \ln(\det(X))^4 - b \ln(\det(X))^2, \quad a, b \in \mathbb{R}_{++}.$$

Já vimos na seção anterior que $f_2(X) = b \ln(\det(X))^2$ é uma função convexa em \mathcal{M} . Afirmamos que $f_1(X) = a \ln(\det(X))^4$ é também convexa em \mathcal{M} . De fato, f_1 é a composição de uma função convexa em \mathcal{M} com um função crescente (convexa) em \mathbb{R} . Logo, pelo [54, Theorem 3.2, p.66], temos que f_1 é uma função convexa em \mathcal{M} .

Para cada $X \in \mathcal{M}$, o gradiente e a Hessiana Riemanniana de f_1 em X são dados, respectivamente, por

$$\begin{aligned} \text{grad} f_1(X) &= 4a \ln(\det(X))^3 X \\ \text{Hess} f_1(X)V &= 12a \ln(\det(X))^2 \text{tr}(X^{-1}V)X. \end{aligned}$$

Como já observado antes, a função f aqui não é convexa nem no sentido Euclidiano, nem no Riemanniano. De fato, para cada $X \in \mathcal{M}$ tal que $\det(X) \in [e^{-\sqrt{b/2a}}, e^{\sqrt{b/2a}}]$, temos que

$$\langle \text{Hess} f(X)V, V \rangle = 2\text{tr}(X^{-1}V)^2(6a \ln(\det(X))^2 - b) < 0, \quad \forall V \in T_X \mathcal{M}.$$

Daí, f não é convexa nesse intervalo. Em particular, f não é convexa em \mathcal{M} .

Para os nossos testes vamos considerar a função $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ dada por

$$f(X) = \ln(\det(X))^4 - 2\ln(\det(X))^2.$$

Os pontos críticos de f são caracterizados por

$$\text{grad}f(X) = \text{grad}f_1(X) - \text{grad}f_2(X) = 0 \Leftrightarrow 4\ln(\det(X))(\ln(\det(X))^2 - 1)X = 0,$$

isto é, $X \in \mathbb{S}_{++}^n$ é ponto crítico de f se $\det(X) = 1$, $\det(X) = e$ ou $\det(X) = e^{-1}$.

Aqui, vamos comparar o nosso método com o *DCPPA* considerado em [48]. Novamente para referência, vamos chamar nosso método de *MDCPPA* a versão usando exponencial, e de *MDCPPA_{retr}* a versão usando retração na definição de x^{k+1} . Os critérios e parâmetros adotados aqui são os mesmo que na seção anterior, inclusive ainda usando a busca de Armijo. Os resultados são apresentados a seguir.

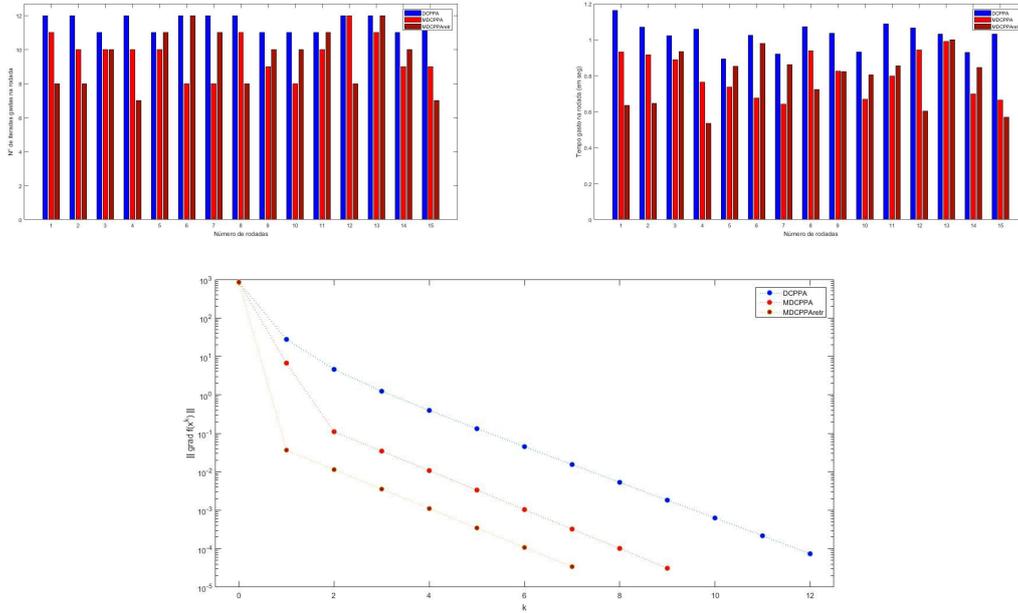


Figura 3.5: Caso DC não convexo com matrizes 5×5

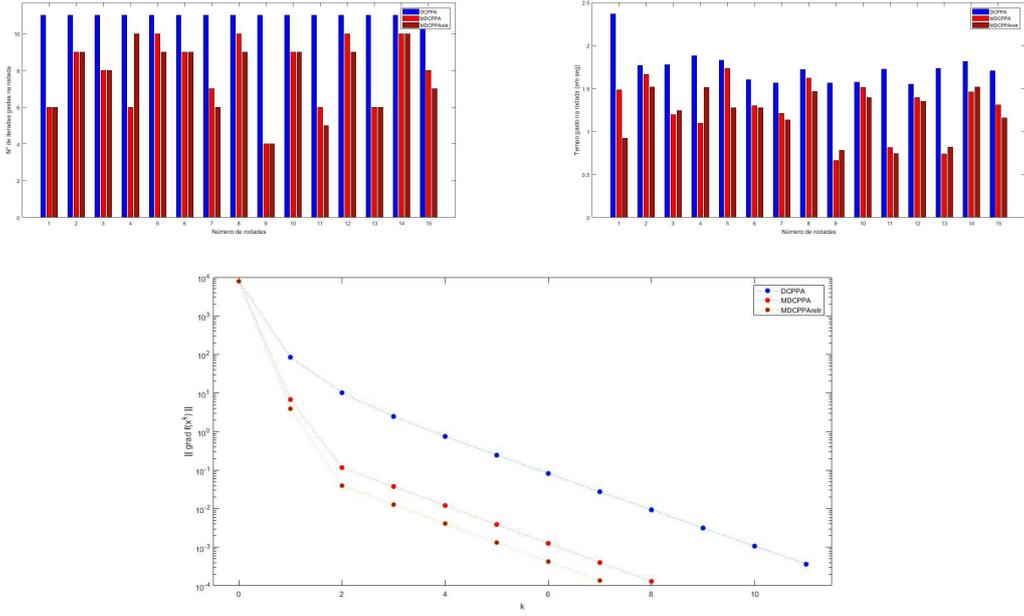


Figura 3.6: Caso DC não convexo com matrizes 10×10

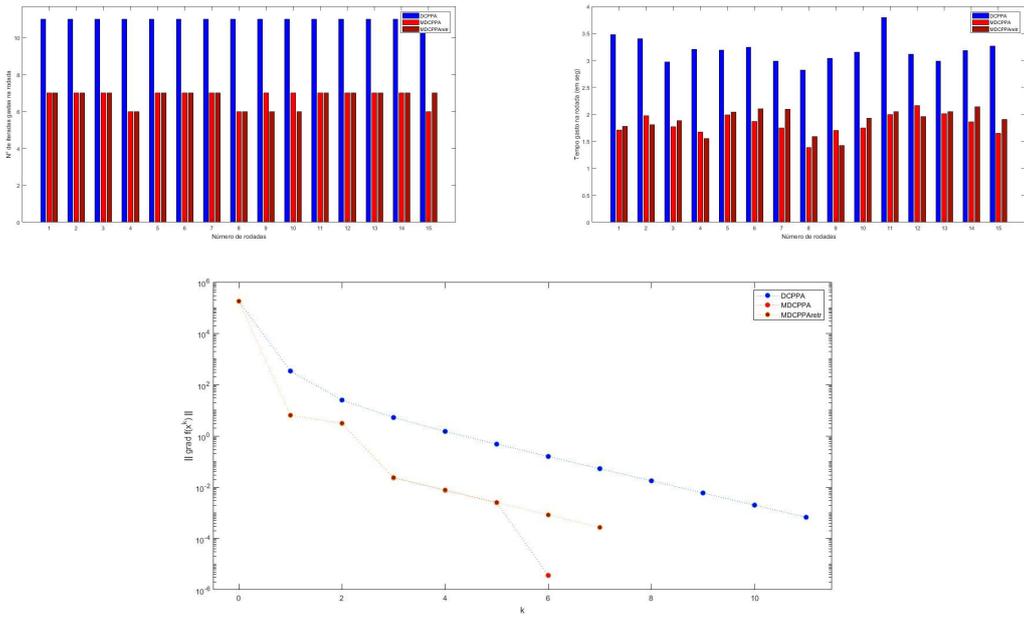


Figura 3.7: Caso DC não convexo com matrizes 20×20

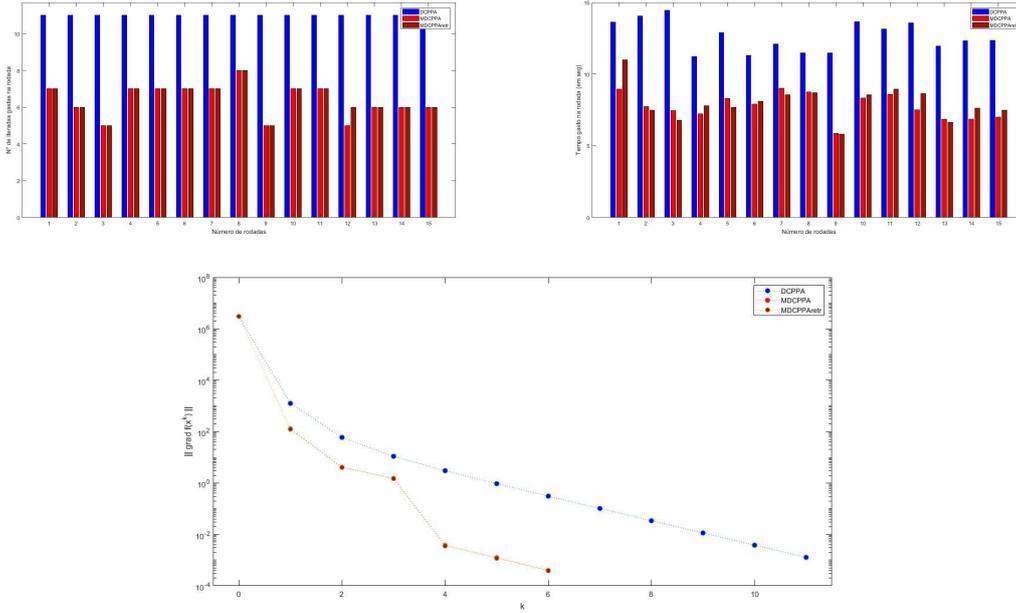


Figura 3.8: Caso DC não convexo com matrizes 50×50

Nos gráficos acima, assim como no caso convexo, podemos ver que nosso método mostrou-se mais eficaz em relação ao MDCPPA, atingindo o critério de parada usando menos tempo de processamento e iteradas. No caso particular para funções DC, observe que o nosso método obteve resultados bem satisfatórios de acordo que as dimensões das matrizes foram ficando cada vez maiores.

Assim como na seção anterior, as tabelas a seguir nos dão o panorama em relação ao desempenho do nosso método. Nelas podemos conferir quantas iteradas gastou cada método em seu melhor desempenho, bem como a média de iteradas que cada um usou. O mesmo é feito em relação ao tempo de processamento gasto.

n	DCPPA		MDCPPA			
	min	med	exp		retr	
			min	med	min	med
5	11	12	8	10	7	10
10	11	11	4	8	4	9
20	11	11	6	7	6	7
50	11	11	5	7	5	7

Tabela 3.3: Média do número de iteradas para cada matriz $n \times n$ no caso DC não convexo

n	DCPPA		MDCPPA			
			exp		retr	
	min	med	min	med	min	med
5	0.8923	1.0312	0.6430	0.7989	0.5363	0.8235
10	1.5519	1.7229	0.6605	1.3086	0.7436	1.276
20	2.8197	3.1798	1.3851	1.7722	1.4214	1.931
50	11.2220	12.3586	5.8724	7.7455	5.8272	7.787

Tabela 3.4: Média do tempo gasto (em seg) para cada cada matriz $n \times n$ no caso DC não convexo

Já comentamos antes que o MPPA é um caso particular do MDCPPA com $f_2 \equiv 0$. Então, para finalizar esta seção, vamos apresentar testes em uma função DC convexa usando o MDCPPA. Para isso, considere $g : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ uma função dada por

$$g(X) = a \ln(\det(X))^2 - b \ln(\det(X)), \quad a, b \in \mathbb{R}_{++}. \quad (3.2)$$

Dado $X \in \mathcal{M}$, o gradiente e a Hessiana da função $g_2(X) = b \ln(\det(X))$ são dados, respectivamente, por

$$\begin{aligned} \text{grad}g_2(X) &= bX \\ \text{Hess}g_2(X)V &= 0_X, \quad \forall V \in T_X\mathcal{M}, \end{aligned} \quad (3.3)$$

onde 0_X denota o *zero* em $T_X\mathcal{M}$. De (3.3), temos que g_2 é uma função *linear* definida em \mathcal{M} . Isso significa que g_2 e $-g_2$ são funções convexas em \mathcal{M} . Desse modo, a função g , definida em (3.2), pode ser vista tanto como uma função DC quanto como uma soma de funções convexas (que é função convexa). Logo, g é uma função DC convexa em \mathcal{M} .

Rodamos o MDCPPA na função $g : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ dada por

$$g(X) = \frac{1}{2} \ln(\det(X))^2 - \ln(\det(X)),$$

e os resultados encontram-se a seguir. Os parâmetros de entrada utilizados são os mesmos usados nos exemplos anteriores.

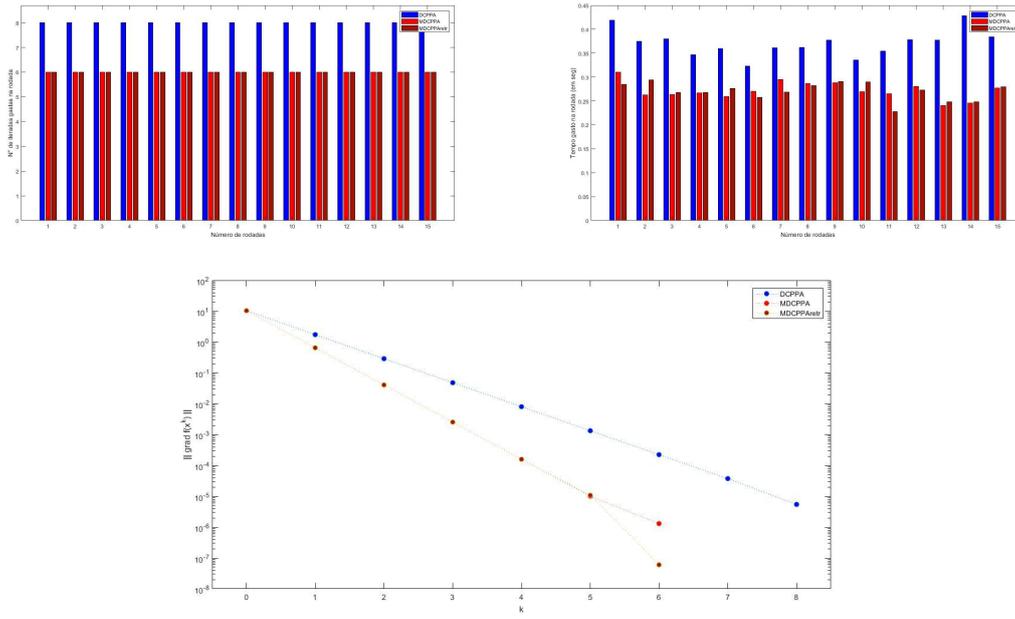


Figura 3.9: Caso DC convexo com matrizes 5×5

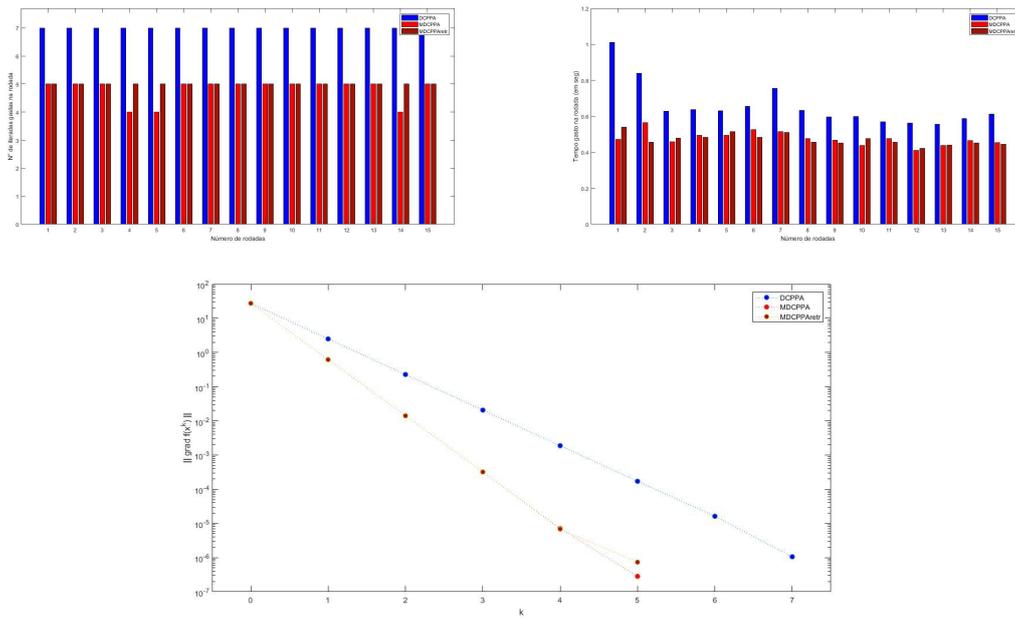


Figura 3.10: Caso DC convexo com matrizes 10×10

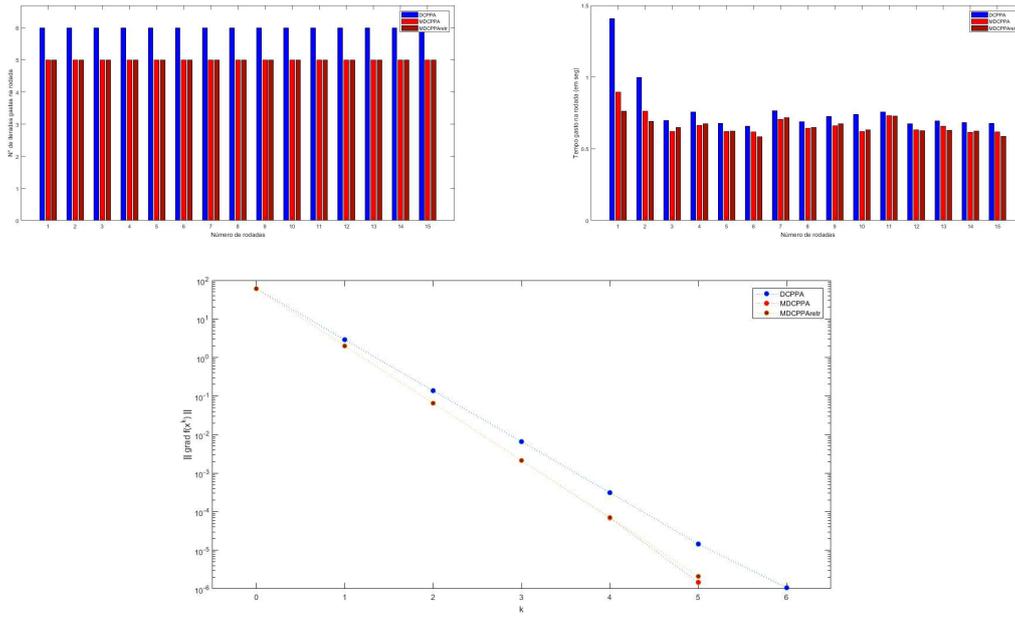


Figura 3.11: Caso DC convexo com matrizes 20×20

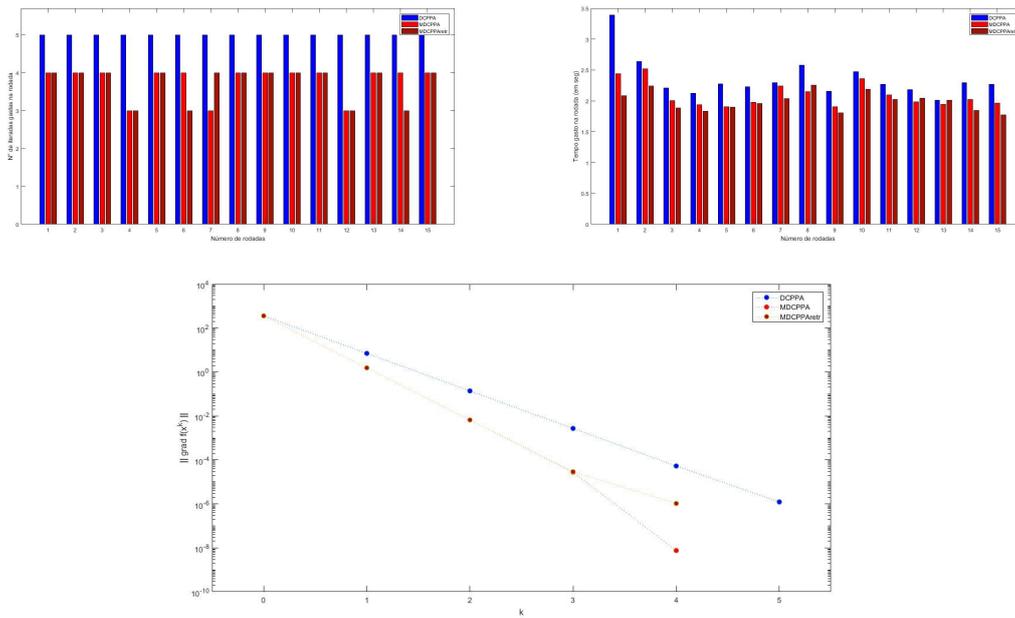


Figura 3.12: Caso DC convexo com matrizes 50×50

n	DCPPA		MDCPPA			
			exp		retr	
	min	med	min	med	min	med
5	8	8	6	6	6	6
10	7	7	4	5	5	5
20	6	6	5	5	5	5
50	5	5	3	4	3	4

Tabela 3.5: Média do número de iteradas para cada matriz $n \times n$ no caso DC convexo

n	DCPPA		MDCPPA			
			exp		retr	
	min	med	min	med	min	med
5	0.3227	0.3748	0.2405	0.2697	0.2280	0.2732
10	0.5558	0.6285	0.4119	0.4724	0.4217	0.4568
20	0.657	0.6967	0.6166	0.6451	0.5832	0.6482
50	2.0106	2.2695	1.9063	2.0019	1.7713	2.0106

Tabela 3.6: Média do tempo gasto (em seg) para cada cada matriz $n \times n$ no caso DC convexo

Os gráficos e tabelas obtidos acima mostram resultados bem próximos aos obtidos usando o MPPA na função convexa usada na seção anterior. Isso acontece porque, como já comentamos antes, o MPPA pode ser visto como um caso particular do MDCPPA.

3.4 Conclusão

Após todos os gráficos e tabelas apresentados nesse capítulo, podemos constatar que o nosso método, de fato, melhora os resultados do Classic PPA e do DCPPA. No caso do MDCPPA, essa diferença é mais nítida ainda se trabalhamos em dimensões maiores.

Referências Bibliográficas

- [1] ALMEIDA, Y. T., CRUZ NETO, J. X., OLIVEIRA, P. R., et al., 2020, “A modified proximal point method for DC functions on Hadamard manifolds”, *Computational Optimization and Applications*, v. 76, pp. 649–673.
- [2] ANSARI, H. Q., BABU, F., YAO, J.-C., 2019, “Regularization of proximal point algorithms in Hadamard manifolds”, *Journal of Fixed Point Theory and Applications*, v. 21, n. 1, pp. 25.
- [3] ARTACHO, F. J. A., FLEMING, R. M. T., VUONG, P. T., 2018, “Accelerating the DC algorithm for smooth functions”, *Mathematical Programming*, v. 169, n. 1, pp. 95–118.
- [4] BAČÁK, M., 2013, “The proximal point algorithm in metric spaces”, *Israel Journal of Mathematics*, v. 194, n. 2, pp. 689–701.
- [5] BAČÁK, M., BORWEIN, J. M., 2011, “On difference convexity of locally Lipschitz functions”, *Optimization*, v. 60, n. 8-9, pp. 961–978.
- [6] BATISTA, E. E. A., BENTO, G. C., FERREIRA, O. P., 2020, “An extragradient-type algorithm for variational inequality on Hadamard manifolds”, *ESAIM: Control, Optimisation and Calculus of Variations*, v. 26, pp. 63.
- [7] BENTO, G. C., FERREIRA, O. P., OLIVEIRA, P. R., 2010, “Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds”, *Nonlinear Analysis: Theory, Methods & Applications*, v. 73, n. 2, pp. 564–572.
- [8] BENTO, G. C., FERREIRA, O. P., OLIVEIRA, P. R., 2015, “Proximal point method for a special class of nonconvex functions on Hadamard manifolds”, *Optimization*, v. 64, n. 2, pp. 289–319.
- [9] BENTO, G. C., FERREIRA, O. P., MELO, J. G., 2017, “Iteration-complexity of gradient, subgradient and proximal point methods on Riemannian ma-

- nifolds”, *Journal of Optimization Theory and Applications*, v. 173, n. 2, pp. 548–562.
- [10] BENTO, G. C., BITAR, S. D. B., CRUZ NETO, J. X., et al., 2020, “A proximal point method for difference of convex functions in multi-objective optimization with application to group dynamic problems”, *Computational Optimization and Applications*, v. 75, n. 1, pp. 263–290.
- [11] BOUMAL, N., MISHRA, B., ABSIL, P.-A., et al., 2014, “Manopt, a Matlab Toolbox for Optimization on Manifolds”, *Journal of Machine Learning Research*, v. 15, n. 42, pp. 1455–1459. Disponível em: <<https://www.manopt.org>>.
- [12] BRIDSON, M., HAEFLIGER, A., 1999, *Metric spaces of non-positive curvature*. Springer-Verlag.
- [13] 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.
- [14] CRUZ NETO, J. X., OLIVEIRA, P. R., SOUBEYRAN, A., et al., 2020, “A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem”, *Annals of Operations Research*, v. 289, n. 2, pp. 313–339.
- [15] DINH, N., STRODIOT, J.-J., NGUYEN, V. H., 2010, “Duality and optimality conditions for generalized equilibrium problems involving DC functions”, *Journal of Global Optimization*, v. 48, n. 2, pp. 183–208.
- [16] DO CARMO, M. P., 1992, *Riemannian geometry*. Birkhäuser.
- [17] FERNÁNDEZ CARA, E., MORENO, C., 1988, “Critical point approximation through exact regularization”, *Mathematics of Computation*, v. 50, n. 181, pp. 139–153.
- [18] FERREIRA, O. P., OLIVEIRA, P. R., 2002, “Proximal point algorithms on Riemannian manifolds”, *Optimization*, v. 51, n. 2, pp. 257–270.
- [19] FERREIRA, O. P., PÉREZ, L. R. L., NÉMETH, S. Z., 2005, “Singularities of monotone vector fields and an extragradient-type algorithm”, *J. Global Optim.*, v. 31, pp. 133–151.
- [20] FERREIRA, O. P., LOUZEIRO, M. S., PRUDENTE, L. F., 2019, “Gradient method for optimization on Riemannian manifolds with lower bounded curvature”, *SIAM Journal on Optimization*, v. 29, n. 4, pp. 2517–2541.

- [21] FERRER, A., BAGIROV, A., BELIAKOV, G., 2015, “Solving DC programs using the cutting angle method”, *Journal of Global Optimization*, v. 61, n. 1, pp. 71–89.
- [22] FLORES-BAZÁN, F., OETTLI, W., 2001, “Simplified optimality conditions for minimizing the difference of vector-valued functions”, *Journal of Optimization Theory and Applications*, v. 108, n. 3, pp. 571–586.
- [23] GABAY, D., 1982, “Minimizing a differentiable function over a differential manifold”, *Journal of Optimization Theory and Applications*, v. 37, n. 2, pp. 177–219.
- [24] GÜLLER, O., 1991, “On the convergence of the proximal point algorithm for convex minimization”, *SIAM Journal on Control and Optimization*, v. 29, n. 2, pp. 403–419.
- [25] GUO, X. L., LI, S. J., 2014, “Optimality conditions for vector optimization problems with difference of convex maps”, *Journal of Optimization Theory and Applications*, v. 162, n. 3, pp. 821–844.
- [26] HARTMAN, P., 1959, “On functions representable as a difference of convex functions”, *Pacific Journal of Mathematics*, v. 9, n. 3, pp. 707–713.
- [27] HIRIART-URRUTY, J.-B., 1985, “Generalized differentiability/duality and optimization for problems dealing with differences of convex functions”. In: *Convexity and duality in optimization*, Springer, pp. 37–70.
- [28] HOLMBERG, K., TUY, H., 1999, “A production-transportation problem with stochastic demand and concave production costs”, *Mathematical Programming*, v. 85, n. 1, pp. 157–179.
- [29] IUSEM, A., 1995, “Métodos de Ponto Proximal em Otimização”, *Colóquio Brasileiro de Matemática, IMPA, R.J., Brasil*.
- [30] LANG, S., 2012, *Fundamentals of differential geometry*, v. 191. Springer Science & Business Media.
- [31] LENGLET, C., ROUSSON, M., DERICHE, R., et al., 2006, “Statistics on the manifold of multivariate normal distributions: Theory and application to diffusion tensor MRI processing”, *Journal of Mathematical Imaging and Vision*, v. 25, n. 3, pp. 423–444.
- [32] LI, C., YAO, J.-C., 2012, “Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal

- point algorithm”, *SIAM Journal on Control and Optimization*, v. 50, n. 4, pp. 2486–2514.
- [33] LI, C., LÓPEZ, G., MARTÍN-MÁRQUEZ, V., 2009, “Monotone vector fields and the proximal point algorithm on Hadamard manifolds”, *Journal of the London Mathematical Society*, v. 79, n. 3, pp. 663–683.
- [34] MAINGÉ, P.-E., MOUDAFI, A., 2008, “Convergence of new inertial proximal methods for DC programming”, *SIAM Journal on Optimization*, v. 19, n. 1, pp. 397–413.
- [35] MARTINET, B., 1970, “Brève communication. Régularisation d’inéquations variationnelles par approximations successives”, *Revue française d’informatique et de recherche opérationnelle. Série rouge*, v. 4, n. R3, pp. 154–158.
- [36] MOAKHER, M., 2005, “A differential geometric approach to the geometric mean of symmetric positive-definite matrices”, *SIAM Journal on Matrix Analysis and Applications*, v. 26, n. 3, pp. 735–747.
- [37] MOREAU, J.-J., 1965, “Proximité et dualité dans un espace hilbertien”, *Bulletin de la Société mathématique de France*, v. 93, pp. 273–299.
- [38] MUU, L. D., QUOC, T. D., 2010, “One step from DC optimization to DC mixed variational inequalities”, *Optimization*, v. 59, n. 1, pp. 63–76.
- [39] NESTEROV, Y. E., 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, pp. 333–361.
- [40] POLYAK, B. T., 1977, “Subgradient methods: a survey of Soviet research”. In: *Nonsmooth optimization: Proceedings of the IIASA workshop March*, pp. 5–30, mar.
- [41] ROCKAFELLAR, R. T., 1976, “Monotone operators and the proximal point algorithm”, *SIAM Journal on Control and Optimization*, v. 14, n. 5, pp. 877–898.
- [42] ROTHHAUS, O. S., 1960, “Domains of positivity”. In: *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, pp. 189–235, Hamburg, abr.
- [43] SAKAI, T., 1996, *Riemannian geometry*, v. 149. American Mathematical Soc.

- [44] SOUBEYRAN, A., 2009, “Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors”, *GREQAM, Aix Marseille University*.
- [45] SOUBEYRAN, A., 2010, “Variational rationality and the unsatisfied man: routines and the course pursuit between aspirations, capabilities and beliefs”, *GREQAM, Aix Marseille University*.
- [46] SOUBEYRAN, A., 2016, “Variational rationality: a theory of worthwhile stay and change approach-avoidance transitions ending in traps”, *GREQAM, Aix Marseille University*.
- [47] SOUZA, J. C. O., 2018, “Proximal point methods for Lipschitz functions on Hadamard manifolds: scalar and vectorial cases”, *Journal of Optimization Theory and Applications*, v. 179, n. 3, pp. 745–760.
- [48] SOUZA, J. C. O., OLIVEIRA, P. R., 2015, “A proximal point algorithm for DC functions on Hadamard manifolds”, *Journal of Global Optimization*, v. 63, n. 4, pp. 797–810.
- [49] SUN, W.-Y., SAMPAIO, R. J. B., CANDIDO, M. A. B., 2003, “Proximal point algorithm for minimization of DC function”, *Journal of Computational Mathematics*, v. 21, pp. 451–462.
- [50] TANG, G.-J., HUANG, N.-J., 2013, “An inexact proximal point algorithm for maximal monotone vector fields on Hadamard manifolds”, *Operations Research Letters*, v. 41, n. 6, pp. 586–591.
- [51] TAO, P. D., AN, L. T. H., 1998, “A DC optimization algorithm for solving the trust-region subproblem”, *SIAM Journal on Optimization*, v. 8, n. 2, pp. 476–505.
- [52] TAO, P. D., SOUAD, E. B., 1986, “Algorithms for solving a class of nonconvex optimization problems. Methods of subgradients”. In: *North-Holland Mathematics Studies*, v. 129, Elsevier, pp. 249–271.
- [53] TUY, H., HORST, R., 1988, “Convergence and restart in branch-and-bound algorithms for global optimization. Application to concave minimization and DC optimization problems”, *Mathematical Programming*, v. 41, n. 1, pp. 161–183.
- [54] UDRISTE, C., 1994, *Convex functions and optimization methods on Riemannian manifolds*, v. 297. Springer Science & Business Media.

- [55] UGWUNNADI, G. C., KHAN, A. R., ABBAS, M., 2018, “A hybrid proximal point algorithm for finding minimizers and fixed points in CAT(0) spaces”, *Journal of Fixed Point Theory and Applications*, v. 20, n. 2, pp. 1–19.
- [56] WANG, J., LI, C., LÓPEZ, G., et al., 2015, “Convergence analysis of inexact proximal point algorithms on Hadamard manifolds”, *Journal of Global Optimization*, v. 61, n. 3, pp. 553–573.
- [57] WANG, J., LI, C., LÓPEZ, G., et al., 2016, “Proximal point algorithms on Hadamard manifolds: linear convergence and finite termination”, *SIAM Journal on Optimization*, v. 26, n. 4, pp. 2696–2729.
- [58] WANG, X., LI, C., WANG, J., et al., 2015, “Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds”, *SIAM J. Optim.*, v. 25, n. 4, pp. 2334–2358.
- [59] WANG, X., LI, C., YAO, J. C., 2015, “Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures”, *Journal of Optimization Theory and Applications*, v. 164, n. 1, pp. 202–217.