Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia

COPPE UFRJ

MÉTODO DO PONTO PROXIMAL INEXATO PARA MINIMIZAÇÃO QUASE-CONVEXA EM VARIEDADES DE HADAMARD

Nancy Baygorrea Cusihuallpa

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.

Orientador: Nelson Maculan Filho

Rio de Janeiro
Fevereiro de 2017

MÉTODO DO PONTO PROXIMAL INEXATO PARA MINIMIZAÇÃO QUASE-CONVEXA EM VARIEDADES DE HADAMARD

Nancy Baygorrea Cusihuallpa

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

Examinada por:

Prof. Nelson Maculan Filho, D.Sc.

Prof. Erik Alex Papa Quiroz, D.Sc.

Prof. Luidi Gelabert Simonetti, D.Sc.

Prof. José Mario Martínez Pérez, D.Sc.

Prof. Susana Scheimberg de Makler, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
FEVEREIRO DE 2017

Cusihuallpa, Nancy Baygorrea

Método do ponto proximal inexato para minimização quase-convexa em variedades de Hadamard/Nancy Baygorrea Cusihuallpa. – Rio de Janeiro: UFRJ/COPPE, 2017.

VIII, 73 p. 29, 7cm.

Orientador: Nelson Maculan Filho

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2017.

Referências Bibliográficas: p. 69 – 73.

1. Método do ponto proximal. 2. Função quase-convexa. 3. Variedades de Hadamard. 4. Otimização não diferenciável. I. Maculan Filho, Nelson. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

A minha família, pelo apoio e carinho mesmo de tão longe.

Agradecimentos

Ao Prof. Maculan, por ter aceitado me orientar, pela confiança com que pude contar na realização deste trabalho e sobretudo pelo incentivo, apoio, amizade durante a minha estadia no PESC.

Ao Prof. Erik Papa, por me motivar e introduzir na área de otimização com ferramentas de geometria Riemanniana desde a graduação . Pelos ensinamentos acadêmicos e de vida.

Aos meus pais que me deram o suporte necessário para o início da minha carreira escolar, me estimulando e apoiando sempre. Ao meu filho Lucas, por ser o meu motivo e motor de vida.

Aos meus amigos, que me acompanharam durante todo este processo: Carol do LAND, Orlando Sarmiento, Hugo Cesar da inteligência artificial, João Carlos e Lara de Piauí, Daniela, Hugo e Renan do Labotim, o Alexandre que apareceu surpresivamente no final desta etapa, e outros que não menciono aqui e que apareceram sem deixar de ser menos importantes.

Aos funcionários administrativos do PESC/COPPE pelo empenho e dedicação em todo o processo, sempre atentos a atender as necessidades imediatas dos alunos e professores.

Ao Capes e CNPq pela bolsa proporcionada durante todo nesse periodo.

Ao povo brasileiro que sempre me abriu as portas com enorme carinho. Amo o Brasil como se fosse meu país, minha segunda terrinha.

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

MÉTODO DO PONTO PROXIMAL INEXATO PARA MINIMIZAÇÃO QUASE-CONVEXA EM VARIEDADES DE HADAMARD

Nancy Baygorrea Cusihuallpa

Fevereiro/2017

Orientador: Nelson Maculan Filho

Programa: Engenharia de Sistemas e Computação

Nesta tese apresentamos um algoritmo inexato de ponto proximal para resolver problemas de otimização quase-convexa em variedades Riemannianas com curvatura seccional não positiva, chamada de variedades de Hadamard. Considerando hipóteses naturais no problema, provamos a convergência da sequência gerada pelo método para um ponto crítico do problema. Além disso, provamos que a taxa de convergência do método é linear e superlinear em alguns casos.

Objetivando a importância das aplicações tanto na economia quanto na teoria de localização, estendemos o algoritmo para resolver problemas de otimização irrestrita, quase-convexa e multiobjetivo onde provamos, supondo hipóteses razoáveis, a convergência da sequência para um ponto crítico Pareto-Clarke. Finalmente, realizamos alguns experimentos computacionais para validar o método proposto e resultados encontrados.

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

INEXACT PROXIMAL POINT METHODS FOR QUASICONVEX MINIMIZATION ON HADAMARD MANIFOLDS

Nancy Baygorrea Cusihuallpa

February/2017

Advisor: Nelson Maculan Filho

Department: Systems Engineering and Computer Science

In this thesis, we present an inexact proximal point algorithm to solve quasiconvex optimization problems in Riemannian manifolds with non positive sectional curvature, called Hadamard manifolds. Then, we show that under mild hypotheses on the optimization problem, the sequence generated by the proposed method are well defined and converge to critical points of the problem. We also prove that the convergence rate of the ones is linear and superlinear in some cases.

Furthermore, by focusing on the importance of applications in economics and localization theory, we extend the proposed algorithm for solving multiobjective quasiconvex optimization problem. Moreover, convergencia of the sequence to a Pareto-Clarke critical point is obtained assuming reasonable hypotheses. Finally, computational experiments were done to validate the proposed model and results found.

Sumário

1 Introdução **1**

2 Preliminares **6**

2.1 Alguns resultados de geometria riemanniana 6

2.1.1 Alguns exemplos de variedades de Hadamard 9

2.2 Alguns resultados de análise variacional 16

2.3 Cálculo não diferenciável 21

2.4 Comentários. 33

3 Versões inexatas do método do ponto proximal **35**

3.1 Algumas propriedades do algoritmo HMIP 38

3.2 Algumas variantes do algoritmo do ponto proximal 40

3.2.1 Algoritmo HMIP1 e resultados de convergência 41

3.2.2 Algoritmo HMIP2 e resultados de convergência 44

3.3 Comentários. 46

4 Taxa de convergência do algoritmo do ponto proximal **48**

4.1 Análise da taxa de convergência 49

4.2 Experimentos numéricos 52

4.2.1 Exemplo: 53

4.3 Comentários 56

5 Algoritmo do ponto proximal inexato para problemas de minimização multiobjetivo **58**

5.1 Definição do problema multiobjetivo e do algoritmo HMIP 58

5.2 Variantes inexatas do algoritmo HISPP para minimização multiobjetivo 63

5.2.1 Algoritmo HISPP1 e resultados de convergência. 63

5.2.2 Algoritmo HISPP2 e resultados de convergência. 67

5.3 Comentários. 68

Referências Bibliográficas **69**

Capítulo 1

Introdução

Muitos dos algoritmos usados para resolver problemas de otimização tem sido generalizados dos espaços lineares (por exemplo os espaços euclideanos, Hilbert, Banach) para variedades diferenciais, particularmente, no contexto de variedades Riemannianas. Isto foi motivado por problemas onde os algoritmos clássicos não podem ser utilizados em sua forma original devido à estrutura natural do problema, por exemplo, problemas de autovalores, cálculo de subespaço invariante, problemas de minimização restrita, problemas de valor fronteira [21], métodos de regressão discreta no cone das matrices definidas positivas[16], média geométrica para matrices semidefinidas positivas[15], modelo geométrico para a coluna vertebral humana[1], detecção de humanos em imagens usando matrizes de covariância como descritores de objeto[57].

Uma outra motivação do ponto de vista da otimização é o fato que alguns problemas de otimização não convexa podem ser transformados em problemas convexos mediante uma escolha adequada de uma métrica Riemanniana, ver [24, 25, 31, 49, 58]. Por exemplo, considere a função clássica banana de Rosenbrock de dimensão 2, dado por $f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$. Essa função torna-se convexa, no contexto Riemanniano, quando consideramos em $I\!R^2$ a métrica Riemanniana dada por

$$G(x) = \begin{pmatrix} 1 + 4x_1^2 & -2x_1 \\ -2x_1 & 1 \end{pmatrix}.$$

De fato, $F(x_1, x_2) = (x_1, -x_2 + x_1^2)$ é um difeomorfismo entre $(I\!R^2, I)$ e $(I\!R^2, G(x))$

sendo bem conhecido que a função $g(x_1, x_2) = (1-x_1)^2+100x_2^2$ é convexa em $(I\!R^2, I)$. Como $f = g \circ F^{-1}$, e segundo Papa Quiroz and Oliveira[44, Teorema 4.5], temos que f é convexa na variedade $(I\!R^2, G(x))$. Considerando outro exemplo, considere a função $f(x_1, x_2) = x_1^2 + (e^{x_1} - x_2)^2$, o qual não é convexa em $(I\!R^2, I)$ mas torna-se convexa, ver [44, Exemplo 4.1], sobre a variedade $(I\!R^2, G_2(x))$, onde

$$G_2(x) = \begin{pmatrix} 1+e^{2x_1} & -e^{x_1} \\ -e^{x_1} & 1 \end{pmatrix}.$$

Nesta tese, estamos interesados em resolver problemas de otimização quase-convexas em variedades Riemannianas completas, conexas e de curvatura não positiva (variedades de Hadamard). Nossa motivação são os problemas de decisão em economia. De fato, considere um agente econômico que vai tomar uma decisão sobre o espaço que representa um plano de produção, por exemplo pode ser o cone de segunda ordem $k := \{z = (\tau, x) \in I\!R^{n+1} \mid \tau > \|x\|_2\}$ dotada pela métrica dada pela hesssiana de $-\ln(\tau^2 - \|x\|_2)$ que é uma variedade de Hadamard de curvatura negativa. Se esta escolha está baseado numa preferência que pode ser representada por uma função de utilidade convexa (que está relacionada a diversificação da escolha e que é natural em economia) então temos um problema de otimização quase-convexa numa variedade de Hadamard.

A seguir, consideremos o seguinte problema de otimização

$$\min_{x \in M} f(x), \tag{1.1}$$

onde M é uma variedade de Hadamard e $f : M \to I\!R \cup \{+\infty\}$ é uma função semicontínua inferior e própria.

Um dos métodos mais populares, para resolver o problema (1.1), utilizado pela comunidade de otimização é o algoritmo do ponto proximal. Esse algoritmo, o qual é uma generalização do algoritmo do ponto proximal clássico de Rocafellar[50], foi introduzido, no contexto Riemanniano, por Ferreira e Oliveira[32] quando f é uma função convexa. O método gera uma sequência $\{x^k\} \subset M$, através do seguinte procedimento iterativo: Dado um ponto inicial $x_0 \in M$, então

$$x^{k+1} \in \arg\min_{x \in M} \left\{ f(x) + \frac{1}{2\lambda_k} d^2(x, x^k) \right\}, \tag{1.2}$$

onde d é a distância Riemanniana definido sobre M e $\{\lambda_k\}$ é uma sequência de números positivos. Nessas condições, foi provado em [32] que a sequência $\{x^k\} \subset M$ gerada pela iteração (1.2) está bem definida, com x^k unicamente determinado. Isto ocorre pois a função $f(\cdot) + \frac{\lambda_k}{2} d^2(\cdot, x^{k-1}) : M \to I\!R$ é 1-coerciva para cada $k \in I\!N$. Além disso, supondo $\sum_{k=0}^{+\infty} \frac{1}{\lambda_k} = +\infty$ e que f tem um minimizador, em [32] também provaram que a sequência $\{f(x^k)\}$ converge para o valor mínimo e a sequência $\{x^k\}$ converge para um ponto minimizador. A partir disso, muitas contribuições foram feitas nesse contexto no intuito de estender o problema (1.1) e a gama de aplicações desse problema. Ver por exemplo Papa Quiroz e Oliveira[41, 42, 44, 45], Bento et al[11, 12], Tang e Huang[55], da Cruz Neto et al.[26], Bento e Cruz Neto[14] e Attouch[4, 5].

Observe que quando f (definido no problema 1.1) é uma função não convexa, resolver o subproblema (1.2) poderia ser tão dificil como resolver o problema original. No contexto quase-convexo, Papa Quiroz e Oliveira[44], em vez de considerar a iteração (1.2), propuseram a seguinte iteração dado por:

$$0 \in \partial(f(\cdot) + (\lambda_k/2)d^2(\cdot, x^{k-1}))(x^k), \tag{1.3}$$

onde $\partial = \partial^F$ é o subdiferencial de Fréchet em variedades de Hadamard. Eles provaram convergência global da sequência para um ponto minimizador.

Além disso, Papa Quiroz e Oliveira[45] também consideraram a iteração (1.3), quando ∂ é o subdiferencial de Clarke em M, para resolver problemas de otimização localmente Lipschitz quase-convexo, e eles provaram que a sequência $\{x^k\}$ gerado por (1.3) existe e converge para um ponto crítico de f.

No intuito de construir algoritmos de pontos proximais computacionalmente mais implementáveis para resolver problemas quase-convexos, nesta tese foram considerados uma sequência de erros. De fato, geralmente é difícil achar um valor exato para x^k que resolva o subproblema (1.3). Além disso, claramente é ineficiente demais gastar esforço computacional tentando estimar os valores de x^k quando talvez seja só necessário do limite dessa sequência que tenha propriedades desejadas. Então, é razoável pensar que (1.3) deveria ser resolvido aproximadamente. Uma forma de obter essa aproximação é substituir (1.3) por a seguinte iteração

$$\epsilon^k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1},$$

onde ∂ é um subdiferencial abstracto, definido na Seção 3 do Capítulo 2 desta monografía.

Então, neste trabalho vamos considerar dois criterios para a sequência de erro do algoritmo inexato do ponto proximal: O primeiro algoritmo inexato, motivado pelo trabalho de Rockafellar[50], está baseado sob a seguinte condição do erro dado por:

$$\sum_{k=1}^{+\infty} \|\epsilon^k\| < +\infty.$$

A segunda versão inexata do algoritmo, motivado pelos trabalhos recentes de

Tang e Huang[55] para campos vetoriais monótonos maximais em variedades de Hadamard, considera os critérios a seguirem:

$$\|\epsilon^k\| < \eta_k d(x^k, x^{k-1}), \qquad \sum_{k=1}^{+\infty} \eta_k^2 < +\infty.$$

Por outro lado, para resolver o problema (1.1) é desejável que a sequência $f(x^k)$ seja decrescente. Para garantir essa propriedade impomos o seguinte criterio:

$$d(\exp_{x^k}\epsilon^k, x^{k-1}) \leq \max\left\{\|\epsilon^k\|, d(x^k, x^{k-1})\right\}.$$

Essa condição pode ser considerada como uma geralização do algoritmo do ponto proximal para funções quase-convexas em espaços euclideanos os quais foram estudados por Papa Quiroz et al.[43].

Outro propósito desta parte da tese é estender o conceito de subdiferencial abstracto, introducido por Aussel[6] no contexto euclidiano, para variedades Riemannianas, mais precisamente em variedades de Hadamard, o qual será utilizado para resolver problemas de otimização não convexa e não suave do tipo (1.1).

Vale ressaltar que os resultados obtidos nos capítulos 3 e 4 desta tese foram publicados na revista Journal of the Operations Research Society of China, ver [7] e [8], respetivamente.

Na segunda etapa deste estudo, estenderemos o algoritmo do pronto proximal para problemas do tipo multiobjetivo. De fato, existe uma classe mais geral de problemas conhecida como otimização multiobjetivo (contida no campo da otimização vetorial). A importância da otimização multiobjetivo pode ser conferida em uma grande variedade de aplicações presentes na literatura. Ver por exemplo [37, 52, 60].

Nesse contexto, o problema (1.1) pode ser reformulado a seguir:

$$\min_{x \in M} F(x) \tag{1.4}$$

onde M é uma variedade de Hadamard e $F := (F_1, \ldots, F_m)$, para cada $F_i : M \to I\!R \cup \{+\infty\}$.

Nos últimos anos, foram vários os pesquisadores que começaram o estudo de problemas de otimização vetorial no contexto Riemanniano. Ver, por exemplo [9, 10, 13, 18]. Além disso, Papa Quiroz e Oliveira[39] apresentaram um método de ponto proximal com distância de Bregman em variedades de Hadamard. Em [40] os autores trabalharam em uma extensão do algoritmo do ponto proximal para minimização quase-convexa irrestrita e restrita ao ortante não negativo. Nesse novo enfoque, faremos uma re-adaptação dos dois algoritmos do ponto proximal inexatos, propostos anteriormente para o caso de funções reais valoradas, para resolver o problema (1.4).

Além disso, sob algumas condições razoáveis, provamos que qualquer ponto de acumulação da sequência gerada pelo algoritmo, para resolver problemas do tipo 1.4, com qualquer dessas duas versões é um ponto crítico Pareto-Clarke.

No entanto, ao nosso conhecimento, esse é o primeiro trabalho que extende o algoritmo do ponto proximal para otimização multiobjetivo quase-convexo no contexto de variedades Riemannianas.

Segue uma breve descrição de cada capítulo. No capítulo 2 apresentamos as definições básicas e notações, como também alguns conceitos e resultados a respeito da teoria do subdiferencial abstracto e da teoria de resultados necessários e suficientes para atingir mínimo em variedades Riemannianas. No capítulo 3, desenvolvemos o método do ponto proximal em variedades de Hadamard (HMIP), apresentando duas versões inexatas do algoritmo (HMIP1 e HMIP2). Além disso, garantimos a boa definição da sequência gerada pelo dois algoritmos propostos e os seus respectivos resultados de convergência. No Capítulo 4, analisamos a taxa de convergência do algoritmo do ponto proximal proposto no Capítulo 2 e testamos o método proposto apresentando resultados computacionais. Finalmente, no Capítulo 5, inspirados no Capítulo 2, estendemos o método do algoritmo do ponto proximal para programação multiobjetivo.

Capítulo 2

Preliminares

Neste capítulo apresentamos os fundamentos teóricos que usaremos no desenvolvimento desta tese. Enunciamos algumas definições e resultados que são provenientes da geometria Riemanniana, de análise variacional e da teoria de subdiferenciais, o quais serão necessários a fim de facilitar a compreensão de nosso trabalho.

2.1 Alguns resultados de geometria riemanniana

Nesta seção vamos fazer uma revisão de algumas propriedades fundamentais, resultados, definições e notações em variedades Riemannianas. Todos estes fatos podem ser encontrados em muitas bibliografias tais como do Carmo [27], Sakai [51], Udriste [58] and Rapcsák [49].

Uma variedade n-dimensional M é um espaço topológico onde cada ponto tem uma vizinhança que é homeomorfa ao espaço euclidiano n-dimensional. Em qualquer ponto $x \in M$, os vetores tangentes são definidos como as tangentes de curvas parametrizadas que passam por x. Denotemos por T_xM o espaço tangente de M em x e $TM = \bigcup_{x \in M} T_xM$, onde T_xM é um espaço linear e tem a mesma dimensão de M.métrica Riemanniana g, então M é uma variedade Riemanniana e denotaremos por (M, G) ou M quando não tem confusão, onde G denota a representação matricial da métrica g. O produto interno de dois vetores $u, v \in T_xM$ é dado por $\langle u, v \rangle_x := g_x(u, v)$, onde g_x é a métrica no ponto x. A norma de um vetor $v \in T_xM$ é dado por $\|v\|_x := \langle v, v \rangle_x^{1/2}$. Se não tem confusão, denotemos $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_x$ and $\| \cdot \| = \| \cdot \|_x$. As métricas podem ser utilizados para definir o comprimento de uma curva suave a troços $\alpha : [t_0, t_1] \to M$ ligando os pontos $\alpha(t_0) = p'$ e $\alpha(t_1) = p$ through $L(\alpha) = \int_{t_0}^{t_1} \|\alpha'(t)\|_{\alpha(t)} dt$. Minimizando o funcional de comprimento sobre o conjunto de todas as curvas obtemos uma distância Riemannian $d(p', p)$ o qual inclui a topologia original em M.

Sabemos que dada uma variedade diferenciável M, um campo de vetores X em M

é uma aplicação que a cada ponto $p \in M$ associa um vetor $X(p)$ pertence no espaço tangente T_pM. Dado dois campos vetoriais V e W em M, a derivada covariante de W na direção V é denotado por $\nabla_V W$. Nesta tese ∇ é a conexão de Levi-Civita associado a (M, G). Essa conexão define uma única derivada covariante D/dt, onde, para cada campo vetorial V, ao longo de uma curva suave $\alpha : [t_0, t_1] \to M$, outro campo vetorial é obtido, denotado por DV/dt e dado por

$$\frac{DV}{dt} = \sum_{j=1}^{n} \frac{dv^j}{dt} X_j + \sum_{i,j=1}^{n} v^j \frac{d\alpha_i}{dt} \nabla_{X_i} X_j, \tag{2.1}$$

onde os coeficientes da conexão afím são as funções $\Gamma_{i,j}^m : U \subset M \to I\!R$ caracterizadas por

$$\Gamma_{i,j}^m = \frac{1}{2} \sum_{k=1}^{n} \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km}. \tag{2.2}$$

O transporte paralelo ao longo de α de $\alpha(t_0)$ para $\alpha(t_1)$, denotado por P_{α,t_0,t_1}, é uma aplicação $P_{\alpha,t_0,t_1} : T_{\alpha(t_0)}M \to T_{\alpha(t_1)}M$ definido por $P_{\alpha,t_0,t_1}(v) = V(t_1)$ onde V é o único campo vetorial ao longo de α tal que $DV/dt = 0$ e $V(t_0) = v$. Desde que ∇ é uma conexão Riemanniana, P_{α,t_0,t_1} é uma isometria linear. Além disso, $P_{\alpha,t_0,t_1}^{-1} = P_{\alpha,t_1,t_0}$ e $P_{\alpha,t_0,t_1} = P_{\alpha,t,t_1} P_{\alpha,t_0,t}$, para todo $t \in [t_0, t_1]$. Uma curva $\gamma : I \to M$ é chamado uma geodésica se $D\gamma'/dt = 0$, o qual é dado pelo seguinte sistema

$$\frac{d^2\alpha_k}{dt^2} + \sum_{i,j=1}^{n} \Gamma_{i,j}^k \frac{d\alpha_j}{dt} . \frac{d\alpha_i}{dt} = 0, \quad k = 1, 2, \ldots, n. \tag{2.3}$$

Uma variedade Riemanniana é completa se as suas geodésicas são todas definidas para qualquer valor de $t \in I\!R$. Seja $x \in M$, a aplicação exponencial $\exp_x : T_xM \to M$ é definido por $\exp_x(v) = \gamma(1, x, v)$, para cada $x \in M$ e a aplicação logarítmica $\log_x := \exp_x^{-1} : M \to T_xM$ definidos sobre variedades diferenciais. Se M é completa, então $\exp_x$ é definido para todo $v \in T_xM$. Além disso, existe uma geodésica minimal (o comprimento é a distância entre os pontos extremos).

Expomos, a seguir, algumas propriedades da função exponencial. Ver por exemplo [34], para uma demostração do seguinte teorema.

Teorema 2.1 *Seja (M, g) uma variedade riemanniana. Consideramos a derivação de Levi-Civita. Então para cada $x \in M$ existe um número $r > 0$ tal que a aplicação $\exp_x : B(0_x, r) \subset TM_x \to M$ está definida por:*

i. $\exp_x : B(0_x, \delta) \to B_M(x, \delta)$ *é um difeomorfismo bi-Lipschitz C^∞, para toda $\delta \in (0, r]$.*

ii. $\exp_x$ *transforma segmentos que passam por 0_x contidos na vizinhança $B(0_x, r)$*

em geodésicas em $B_M(x,r)$*.*

iii. $\mathrm{dexp}_x(0_x) = id_{TM_x}$.
Em particular, tendo em consideração a condição 3 desse teorema, para cada $C > 1$*, o raio* r *pode ser escolhido suficientemente pequeno para que as aplicações* $\exp_x : B(0_x, r) \to B_M(x,r)$ *e* $\exp_x^{-1} : B_M(x,r) \to B(0_x, r)$ *sejam* C*-Lipschitz.*

A seguir, definimos o produto direto de duas variedades Riemannianas (M_1, g_1) e (M_2, g_2). Em cada ponto (p_1, p_2) da variedade produto $M = M_1 \times M_2$, o espaço tangente em M é escrito como $M_{(p_1,p_2)} = (M_1)_{p_1} + (M_2)_{p_2}$ (com a soma direta), no qual definimos um produto interno $g_{(p_1,p_2)}$ por $g_{(p_1,p_2)} = g_{p_1} \oplus g_{p_2}$ (a soma direta ortogonal). Nesse contexto, obtemos uma métrica Riemanniana g, o qual chamamos o produto direto de g_1 e g_2 e é escrito como $g_1 \times g_2$. A variedade Riemannianna (M, g) é chamado o produto direto de (M_1, g_1) e (M_2, g_2) e é denotado como $(M_1, g_1) \times (M_2, g_2)$. Um vetor u tangente a M em $p = (p_1, p_2)$, pode ser representado como $u = (u_1, u_2)$ com $u_1 \in T_{p_1}M_1$ e $u_2 \in T_{p_2}M_2$ e com essas identificações a aplicação exponencial em p é dado por $\exp_p u = (\exp_{p_1} u_1, \exp_{p_2} u_2)$, assim também $\exp_p^{-1} u = (\exp_{p_1}^{-1} u_1, \exp_{p_2}^{-1} u_2)$.

Dados os campos vetoriais X, Y, Z em M, denotemos por R como o tensor de curvatura definido por $R(X,Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X,Y]} Z$, onde $[X,Y] := XY - YX$ é o produto de Lie. Logo, a curvatura seccional quando relacionado com X e Y é definido por

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

Cabe lembrar que as variedades Riemannianas completas, simplemente conexas e com curvatura não positiva são chamadas de *variedades de Hadamard.* Alguns exemplos de variedades de Hadamard podem ser achados na Seção 4 of Papa Quiroz and Oliveira [44].

Teorema 2.2 *Seja* M *uma variedade de Hadamard. Então* M *é difeomorfo para o espaço euclidiano* $I\!R^n$, $n = \dim M$*. Mais precisamente, em qualquer ponto* $x \in M$*, a aplicação exponencial* $\exp_x : T_x M \to M$ *é um difeomorfismo global.*

Proof. Ver Sakai, [51], Teorema 4.1, página 221. ∎

Uma consequência do teorema anterior é o fato que as variedades de Hadamard tem a propriedade da unicidade das geodésicas ligando qualquer dois pontos. Outra propriedade útil é a seguinte: Seja $[x, y, z]$ um triângulo geodésico, o qual consiste de *vértices* e as geodésicas ligando aqueles pontos, temos

Teorema 2.3 *Dado um triângulo geodésico* $[x, y, z]$ *numa variedade de Hadamard. Se cumpre que*

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

onde $\exp_z^{-1}$ *denota a inversa de* $\exp_z$ *.*

Prova. Ver Sakai, [51], Proposição 4.5, página 223. ■

Teorema 2.4 *Seja M uma variedade de Hadamard e $y \in M$ um ponto fixo. Então , a função $g(x) = d^2(x, y)$ é estritamente convexa e* $\operatorname{grad} g(x) = -2 \exp_x^{-1} y$.

Prova. Ver Ferreira e Oliveira, [32], Proposição II.8.3. ■

Proposição 2.1 *Tome $x \in M$. A aplicação $d^2(\cdot, x)/2$ é fortemente convexa.*

Prova. Ver Da Cruz Neto et al. [24].

2.1.1 Alguns exemplos de variedades de Hadamard

Como foi mencionado no começo deste capítulo, variedades de Hadamard são variedades Riemannianas, completas e simplemente conexas com curvatura não positiva. A importância e interesse de estudarmos, nesta tese, algoritmos de otimização sobre esse tipo de variedades baseia-se num resultado bem conhecido na geometria Riemanniana o qual diz que uma variedade de Hadamard de dimensão n é issomorfa ao espaço euclidiano $I\!R^n$, ver [51, Teorema 4.3], o qual seria uma motivação e bom sinal de que algoritmos de otimização estudados em espaços euclideanos também funcionam nesse tipo de espaços.

A seguir, a classe de variedades de Hadamard mais utilizadas podem incluir os seguintes espaços com suas respetivas métricas:

Espaços hiperbólicos.

Dotamos $I\!R^{n+1}$ com o $(-1, n)$-produto interno

$$\langle x, y\rangle_{(-1,n)} := -x_0 y_0 + \sum_{i=1}^{n} x_i y_i$$

para $x = (x_0, \ldots, x_n)$ e $y = (y_0, \ldots, y_n)$.
Definimos

$$\mathbb{H}^n := \{x \in I\!R^{n+1} \mid \langle x, x\rangle_{(-1,n)} = -1, \quad x_0 > 0\}.$$

Então $\langle\cdot,\cdot\rangle$ induce uma métrica Riemanniana g no espaço tangente $T_p\mathbb{H}^n \subset T_p I\!R^{n+1}$ para $p \in \mathbb{H}^n$.

A seguir, vamos demonstrar que o espaço hiperbólico H^n, para $n = 2$, é una variedade de Hadamard, isto é, uma variedade Riemanniana, simplesmente conexa, completa e de curvatura seccional não positiva. Com essa finalidade é necessário calcular as componentes da conexão Riemanniana: os símbolos de Christoffel, as equações das geodésicas, a distância Riemanniana e a curvatura seccional.

a. Símbolos de Christoffel.

As componentes da métrica estão dados como

$$\begin{aligned} g_{11} = g_{22} &= \frac{1}{x_2^2} \\ g_{12} = g_{21} &= 0 \end{aligned} \tag{2.4}$$

Se $k \neq m$, entonces $g_{km} = 0$ então, da equação (2.2), se tem

$$\Gamma_{ij}^m = \frac{1}{2} g^{mm} \left(\frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right). \tag{2.5}$$

Substituindo (2.4) em (2.5), a conexão Riemanniana em H^2 tem as seguintes componentes

$$\begin{aligned} \Gamma_{11}^1 = \Gamma_{22}^1 = \Gamma_{12}^2 = \Gamma_{21}^2 = 0 \\ \Gamma_{12}^1 = \Gamma_{21}^1 = \Gamma_{22}^2 = -\Gamma_{21}^1 = -\frac{1}{x_2} \end{aligned} \tag{2.6}$$

b. Ecuação da geodésica.

Seja uma curva $\alpha(t) = (\alpha_1(t), \alpha_2(t)) \in H$ tal que $\alpha(0) = (y_1, y_2)$ y $\alpha'(0) = (v_1, v_2)$. Então substituindo as componentes da conexão afím em (2.3), tem-se o seguinte sistema:

$$\frac{d^2\alpha_1}{dt^2} - \frac{2}{\alpha_2}\frac{d\alpha_1}{dt}\frac{d\alpha_2}{dt} = 0 \tag{2.7}$$

$$\frac{d^2\alpha_2}{dt^2} + \frac{1}{\alpha_2}\left(\frac{d\alpha_1}{dt}\right)^2 - \frac{1}{\alpha_2}\left(\frac{d\alpha_2}{dt}\right)^2 = 0. \tag{2.8}$$

Substituindo $p(t) = \dfrac{d\alpha_1}{dt}$ em (2.7), tem-se

$$\frac{dp}{dt} - \frac{2}{\alpha_2}\frac{d\alpha_2}{dt} p = 0,$$

como consequencia de um resultado de cálculo diferencial obtem-se a solução :

$$p(t) = k\alpha_2^2(t), \qquad k > 0. \tag{2.9}$$

Substituindo (2.9) em (2.8),

$$\frac{d^2\alpha_2}{dt} + k^2\alpha_2^3 - \frac{1}{\alpha_2}\left(\frac{d\alpha_2}{dt}\right)^2 = 0. \tag{2.10}$$

Definimos

$$\frac{d\alpha_2}{dt} := z(t)\,\alpha_2(t), \tag{2.11}$$

então,

$$\frac{d^2\alpha_2}{dt^2} = \alpha_2(t)\left(\frac{dz}{dt} + z(t)^2\right). \tag{2.12}$$

Substituindo (2.11) e (2.12) em (2.10) e como $\alpha_2(t) \in H$, isto é, $\alpha_2(t) > 0$ tem-se

$$z(t) = r - k\alpha_1(t), \qquad r \in I\!R. \tag{2.13}$$

Substituindo (2.13) em (2.11),

$$\alpha_2(t) = s\, e^{rt - k\int \alpha_1(t)dt}, \qquad s > 0. \tag{2.14}$$

Se $\alpha_1(t) = c$ com c uma constante arbitrária, então $\alpha_2(t) = se^{ht}$, com $h \in I\!R$. Portanto,

$$(\alpha_1(t), \alpha_2(t)) = \left(c, se^{ht}\right). \tag{2.15}$$

Se $\alpha_1(t)$ não é constante em (2.14). Definimos:

$$\alpha_2(t) = \frac{s}{k}\ sech\, pt, \qquad s > 0. \tag{2.16}$$

Igualando (2.16) em (2.14) e derivando com respeito de t, tem-se

$$\alpha_1(t) - b = \frac{s}{k} tanh\, pt, \qquad b \in I\!R \tag{2.17}$$

Portanto,

$$(\alpha_1(t), \alpha_2(t)) = \left(b + \frac{s}{k}\, tanh\, pt, \frac{s}{k}\, sech\, pt\right). \tag{2.18}$$

Além disso, de (2.16) e (2.17), tem-se

$$(\alpha_1(t) - b)^2 + \alpha_2(t)^2 = R^2,$$

com $0 < R = s/k$, isto é, $\alpha_1, \alpha_2 \in C_{b,R}$ com $\alpha_2(t) > 0$ onde $C_{b,R}$ é a semicircunsferencia de centro b e raio r.

Evaluando $\alpha(t) \in H$ em (2.15) e (2.18) tal que $\alpha(0) = (y_1, y_2)$ e $\alpha'(0) = (v_1, v_2)$, tem-se

$$
\begin{aligned}
(\alpha_1(t),\alpha_2(t)) &= \left(y_1, y_2 e^{v_2/y_2\, t}\right), \quad v_2 = 0 && (2.19)\\
(\alpha_1(t),\alpha_2(t)) &= \left(y_1 + y_2\, tanh\, \frac{v_1}{y_2}t, y_2\, sech\, \frac{v_1}{y_2}t\right), \quad v_1 = 0 && (2.20)
\end{aligned}
$$

os quais são as geodésicas do plano Hiperbólico H^2.

c. Distância Riemanniana.

Seja $\alpha(y,v,1) = (z_1,z_2)$, isto é, $(\alpha_1(t),\alpha_2(t))\mid_{t=1} = (z_1,z_2)$.

c.1) Se $\alpha(t)$ é a geodésica dada em (2.20), obtem-se que $z_1 = y_1$ e $v_2 = y_2 \ln \dfrac{z_2}{y_2}$.
Logo, $\exp_y v = x$, se e somente se, $v = \exp_y^{-1} x$.
Então,

$$
\begin{aligned}
d^2(z,y) &= \left\|\exp_y^{-1} z\right\|_G^2\\
&= \langle v, G(y)v\rangle\\
&= \frac{v_2^2}{y_2^2}\\
&= \ln^2 \frac{z_2}{y_2}.
\end{aligned}
$$

c.2) Se $\alpha(t)$ é a geodésica dada em (2.19), tem-se que

$$
\begin{aligned}
z_1 &= y_1 + y_2\, tanh\, \frac{v_1}{y_2}\\
z_2 &= y_2\, sech\, \frac{v_1}{y_2}.
\end{aligned}
$$

Calculando o sistema anterior, obtem-se

$$
v_1 = \ln\left(\frac{z_1 y_2 + \sqrt{z_1^2 y_2^2 - z_2^2 y_1^2 + z_2^2 y_2^2}}{z_2(y_1+y_2)}\right) y_2,
$$

$$
\begin{aligned}
d^2(z,y) &= \left\|\, \exp_y^{-1} z\,\right\|_G^2\\
&= \langle v, G(y)v\rangle\\
&= \frac{v_1^2}{y_2^2}\\
&= \ln^2\left(\frac{z_1 y_2 + \sqrt{z_1^2 y_2^2 - z_2^2 y_1^2 + z_2^2 y_2^2}}{z_2(y_1+y_2)}\right).
\end{aligned}
$$

Portanto, a distância hiperbólica entre os pontos $y = (y_1,y_2)$ y $z = (z_1,z_2)$ estão dado como:

$$d(z,y) = \begin{cases} \left|\ln \frac{z_2}{y_2}\right|, & x_1 = x_2 \\ \\ \left|\ln\left(\dfrac{z_1y_2 + \sqrt{z_1^2y_2^2 - z_2^2y_1^2 + z_2^2y_2^2}}{z_2(y_1+y_2)}\right)\right|, & y, z \in C_{b,R}. \end{cases}$$

d. Curvatura seccional.

Seja $\{X_1,\, X_2\}$ uma base local ortonormal do espacio hiperbólico H^2, o qual denotamos por $X_i = \frac{\partial}{\partial x_i}$ con $i = 1, 2$.

Logo,

$$\begin{aligned} \nabla_{X_1}X_1 &= -\nabla_{X_2}X_2 &= \frac{1}{x_2}X_2, \\ \nabla_{X_1}X_2 &= -\nabla_{X_2}X_1 &= -\frac{1}{x_2}X_1, \end{aligned}$$

e,

$$\nabla_{X_2}\left(\nabla_{X_1}X_2\right) = \nabla_{X_2}\left(\nabla_{X_2}X_1\right) = \nabla_{X_2}\left(\nabla_{X_2}X_2\right) = 2\,\nabla_{X_1}\left(\nabla_{X_2}X_2\right) = -2\,\nabla_{X_1}\left(\nabla_{X_1}X_1\right) = \frac{2}{x_2^2}X_1$$

$$\nabla_{X_2}\left(\nabla_{X_1}X_1\right) = 2\,\nabla_{X_1}\left(\nabla_{X_1}X_2\right) = -\frac{2}{x_2^2}X_2.$$

Além disso,

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

isto é,

$$R\left(X_1, X_1\right)X_1 = R\left(X_2, X_2\right)X_1 = R\left(X_1, X_1\right)X_2 = R\left(X_2, X_2\right)X_2 = 0$$

$$\begin{aligned} R\left(X_2, X_1\right)X_1 &= -R\left(X_1, X_2\right)X_1 = \frac{1}{x_2^2}X_2 \\ R\left(X_2, X_1\right)X_2 &= -R\left(X_1, X_2\right)X_2 = -\frac{1}{x_2^2}X_1. \end{aligned}$$

Como $R_{ijkl} = \langle R\left(X_i, X_j\right)X_k, X_l\rangle$ e pela ortonormalidade de $\{X_1,\, X_2\}$, tem-se que

$$\begin{aligned} R_{1212} &= -R_{1221} = -1, \\ R_{2121} &= -R_{2112} = -1. \end{aligned}$$

Logo, a curvatura seccional é -1.

Além disso, H^2 está dotada de uma métrica Riemanniana $\langle\cdot,\cdot\rangle$ con $G(x) = \left(\frac{1}{x_2}\delta_{ij}\right)$ com $i, j = 1, 2$, para dois pontos quaisquer dados en H^2 tem um único caminho ligando-os e que as geodésicas dadas em (2.19) y (2.20) estão definidas para todo $t \in I\!R$, então a variedade H^2 é uma variedade Riemanniana, simplesmente conexa e completa respectivamente. Portanto, o espaço hiperbólico H^2 e é uma variedade de Hadamard.

Cone das matrizes simétricas definida positivas (variedades SPD)

Seja $\mathcal{M}$ ser o conjunto de matrizes reais simétricas $n \times n$. Denotamos por:

$$\mathcal{S}(n) := \{A \in \mathcal{M},\, A^T = A\}$$

e por

$$\mathcal{S}(n)_+ := \{A \in \mathcal{S}(n),\, A > 0\}$$

o conjunto de todas as matrizes simétricas definidas positivas. $A > 0$ significa que a forma quadrática $x^T A x > 0$ para todo $x \in I\!R^n \backslash \{0\}$. É simples ver que $\mathcal{S}(n)_+$ é um cone aberto e convexo.

Como $A^{-1/2} X A^{-1/2} \subset \mathcal{S}(n)$, para cada $A \in \mathcal{S}(n)_+$ identificamos o conjunto $T_A\mathcal{S}(n)_+$ de vetores tangentes a $\mathcal{S}(n)_+$ em A com $\mathcal{S}(n)$. A métrica riemanniana invariante afim sobre $\mathcal{S}(n)_+$, introduzido por [46, 47], é definido como

$$\langle X, Y\rangle_A := tr(A^{-1}XA^{-1}Y), \qquad X, Y \in T_A\mathcal{S}(n)_+, \quad A \in \mathcal{S}(n)_+ \tag{2.21}$$

onde $tr(A)$ denota o traço de A, para $A \in \mathcal{S}(n)_+$ e $X, Y \in T_A\mathcal{S}(n)_+$ Observe que, para $X \in \mathcal{S}(n)_+$, o fato de ser $\langle \cdot, \cdot \rangle_A$ definida positiva é consequência da propriedade do produto interno de Frobenius ser definida positiva, isto é, tendo em conta que o traço é invariante sob permutações cíclicas, temos que:

$$\begin{aligned}
\langle X, X\rangle_A &= tr(A^{-1}XA^{-1}X) \\
&= tr(A^{-1/2}A^{-1/2}XA^{-1/2}A^{-1/2}X) \\
&= tr(A^{-1/2}XA^{-1/2}A^{-1/2}XA^{-1/2}) \\
&= \langle A^{-1/2}XA^{-1/2}, A^{-1/2}XA^{-1/2}\rangle_F \\
&= \|A^{-1/2}XA^{-1/2}\|_F > 0.
\end{aligned}$$

A última desigualdade é consequência de $X \in \mathcal{S}(n)_+$.

A geodésica que passa por $I \in \mathcal{S}(n)_+$ na direção de V, uma (simétrica) matriz no espaço tangente em I, é explícitamente dado por e^{tV}, ver [22]. Usando invariança sob transformações congruentes, a geodésica $A(t)$ tal que $A(0) = A$ e $\dot{A}(0) = V$ é dado por

$$A(t) := A(A,t,V) = A^{1/2} e^{tA^{-1/2}VA^{-1/2}} A^{1/2}.$$

Além disso, a aplicação exponencial no ponto $A \in \mathcal{S}(n)_+$ é dado por

$$\exp_A V := A(A,1,V) = A^{1/2} e^{A^{-1/2}VA^{-1/2}} A^{1/2}. \tag{2.22}$$

Teorema 2.5 *Seja* $A \in \mathcal{S}(n)_+$ *e* $\log_A : \mathcal{S}(n)_+ \to \mathcal{S}(n)$. *A aplicação* $\exp_A$ *e o logaritmo matricial* $\log_A$ *são difeomorfos.*

Prova. Ver [2, Teorema 2.8]. ∎

Seja $B = \exp_A V$. Logo, de (2.22) temos $e^{A^{-1/2}VA^{-1/2}} = A^{-1/2}BA^{-1/2}$, o que implica que $A^{-1/2}VA^{-1/2} = \log(A^{-1/2}BA^{-1/2})$. Logo,

$$\log_A(B) = V = A^{1/2} \log(A^{-1/2}BA^{-1/2}) A^{1/2}.$$

Para matrizes $X \in \mathcal{S}(n)_+$, a aplicação exponencial e logaritmo matricial podem ser estimadas via SVD. Seja $X = Udiag(\lambda_i)U^T$ ser SVD de X,:então

$$\begin{aligned} \log(X) &= \sum_{r=1}^{+\infty} \frac{(-1)^{r-1}}{r}(X-I) = Udiag(\ln(\lambda_i))U^T; \\ \exp_A(X) &= \sum_{r=0}^{+\infty} \frac{1}{r!} X^r = Udiag(e^{\lambda_i})U^T \end{aligned}$$

Segue que a distância Riemanniana de A e B em $\mathcal{S}(n)_+$ é

$$d_{\mathcal{S}(n)_+}(A,B) := \| \ln(A^{-1/2}BA^{-1/2}) \|_F = \| \ln(B^{-1/2}AB^{-1/2}) \|_F.$$

$\mathcal{S}(n)_+$ forma uma variedade Riemanniana com curvatura negativa quando adotada a métrica invariante afim, ver [47], Esta geometria é estudado em detalhes no paper de Arsigny et al.[2].

O ortante positivo

O espaço $\mathbb{R}^n_{++} = \{x \in \mathbb{R}^n \mid x_i > 0,\ i = 1, \ldots, n\}$ dotado com a métrica afim, também chamado de métrica de Dikin, o qual é definido como,

$$G(x) := \begin{pmatrix} 1/x_1^2 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1/x_n^2 \end{pmatrix}, \tag{2.23}$$

isto é, para qualquer $x \in I\!R^n_{++}$ e $u, v \in T_x I\!R^n_{++}$,

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

É bem conhecido que $(I\!R^n_{++}, G(x))$ é uma variedade de Hadamard com curvatura seccional nula e cujo espaço tangente no ponto $x \in M$ é $I\!R^n$. A geodésica ligando os pontos x a y é a curva $\gamma : [0,1] \to I\!R^n_{++}$ definido por

$$\gamma(t) = (x_1^{1-t} y_1^t, \ldots, x_n^{1-t} y^t).$$

Portanto, a distância entre x y y está dado por:

$$d(x,y) = \left(\sum_{i=1}^{n} (\ln \frac{y_i}{x_i})^2 \right)^{\frac{1}{2}}.$$

O hipercubo $(0,1)^n$

Esse espaço foi estudado estensivamente na tese de Papa Quiroz[38] A geodésica $\gamma(t) = (\gamma_1(t), \ldots, \gamma_n(t))$ definida nessa variedade tal que $\gamma(0) = x = (x_1, \ldots, x_n)$ e $\gamma'(0) = v = (v_1, \ldots, v_n)$ é

$$\gamma_i(t) = \frac{1}{2}\left\{ 1 + \tanh(\frac{1}{2}\frac{v_i}{x_i(1-x_i)} t + \frac{1}{2}\ln\frac{x_i}{1-x_i}) \right\}, \quad i = 1, \ldots, n$$

onde $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ é a função tangente hiperbólica. Essa geodésica está definida para todo $t \in I\!R$ e a distância geodésica de um ponto x ao ponto $y = \gamma(t_0)$, $t_0 > 0$ é dado por

$$d(p,q) = \int_0^{t_0} \|\gamma'(t)\| dt = \left\{ \sum_{i=1}^{n} \left(\ln(\frac{y_i}{1-y_i}) - \ln(\frac{x_i}{1-x_i}) \right)^2 \right\}^{1/2}$$

2.2 Alguns resultados de análise variacional

Seja M uma variedade Riemanniana. Dada uma função real valorada $f : M \to I\!R \cup \{+\infty\}$, denotemos o seu dominio por $dom(f) := \{x \in M : f(x) < +\infty\}$ e o seu epigrafo por $epi(f) := \{(x, \beta) \in M \times I\!R : f(x) \leq \beta\}$. Além disso, f é dito ser própria se $dom(f) \neq \emptyset$ e para todo $x \in \mathtt{domf}$, temos $f(x) > -\infty$. A função f é dita semicontínua inferior se $epi(f)$ é um subconjunto fechado de $M \times I\!R$.

Uma função $f : M \to I\!R \cup \{+\infty\}$ é convexa em M se e somente se para qualquer geodésica $\gamma : I\!R \to M$, a composição $f \circ \gamma : I\!R \to I\!R \cup \{+\infty\}$ é convexa.

A gradiente de uma função diferenciável $f : M \to I\!R$, denotado por $grad(f)$, é um campo vetorial em M definido por $df(X) = \langle grad(f), X\rangle = X(f)$, onde X é tambem um campo vetorial em M. Seja $f : M \to I\!R \cup \{+\infty\}$ uma função convexa, $x \in M$ e um vetor $v \in T_xM$. A derivada direccional (clássica) de f em $x \in M$ na direção de um vetor v é definido por:

$$f'(x, v) := \lim_{t\to 0^+} \frac{f(\gamma(t)) - f(x)}{t}, \tag{2.24}$$

onde $\gamma : I\!R \to M$ é uma geodésica tal que $\gamma(0) = x$ and $\gamma'(0) = v$.

Dado $x \in M$, um vetor $s \in T_xM$ é dito subgradiente de uma função convexa f em x se e somente se para qualquer geodésica $\gamma : I\!R \to M$ com $\gamma(0) = x$,

$$(f \circ \gamma)(t) \geq f(x) + t\langle s, \gamma'(0)\rangle.$$

O conjunto de todos os subgradientes de f em x, $\partial^{FM} f(x)$, é chamado de subdiferencial de Fenchel-Moreau de f em x.

Seja $f : M \to I\!R \cup \{+\infty\}$ uma função propria convexa, o qual é contínua no ponto x onde f é finito. Isto segue que

$$f'(x, v) = \max_{s\in\partial f(x)} \langle s, v\rangle, \tag{2.25}$$

para todo $v \in T_xM$. Esta prova pode ser encontrada na Proposição 3.2 de da Cruz Neto et al.[23].

Seja $f : M \to I\!R\cup\{+\infty\}$ ser uma função propria. Então, f é dita quase-convexa se para todo $x, y \in M$, $t \in [0, 1]$, se cumpre

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

para a geodésica $\gamma : [0, 1] \to I\!R$, tal que $\gamma(0) = x$ e $\gamma(1) = y$.
Um resultado bem conhecido na análise variacional estabelece que a composição de uma função não decrescente com outra função convexa gera uma função quase-convexa. De fato, seja $f := g\circ h : M \to I\!R$ com g sendo uma função não decrescente e h sendo uma convexa. Logo, como h é convexa temos que para toda curva geodésica $\gamma : I\!R \to M$ e para todo $t \in [0, 1]$ temos

$$h(\gamma(t)) \leq th(\gamma(0)) + (1 - t)h(\gamma(1)). \tag{2.26}$$

Além disso, como h é quase-convexo, temos que

$$th(\gamma(0)) + (1 - t)h(\gamma(1)) \leq \max\{h(x), h(y)\}. \tag{2.27}$$

De (2.26) e (2.27) e desde que g é não decrescente, temos que

$$\begin{aligned} g(h(\gamma(t))) &\leq g(th(\gamma(0)) + (1-t)h(\gamma(1))) \\ &\leq g(\max\{h(x), h(y)\}) \\ &\leq \max\{(g \circ h)(x), (g \circ h)(y)\}. \end{aligned}$$

Então, f é quase-convexa.

Exemplo 2.1 *Considere a variedade $((0,1)^2, X^{-2}(I-X)^{-2})$. A função $f(x_1, x_2) = \sqrt{-\log(x_1(1-x_1)x_2(1-x_2))}$ é quase-convexa nessa variedade. De fato, pode se observar que f é a composição de duas funções: g sendo a função raiz quadrada, o qual é não decrescente, e $h := -\log(x_1(1-x_1)x_2(1-x_2))$. Agora só resta provar que h é uma função convexa nessa variedade, isto é que o hessiano de h seja uma matriz semidefinida positiva. Ver [58, Teorema 6.2]. Então,*

$$\begin{aligned} H_x^h(x_1,x_2) &= \nabla^2 f(x_1,x_2) + \left(X^{-1} - (I-X)^{-1}\right) f'(x_1,x_2) \\ &= \begin{pmatrix} \frac{2x_1^2-2x_1+1}{x_1^2(1-x_1)^2} & 0 \\ 0 & \frac{2x_2^2-2x_2+1}{x_2^2(1-x_2)^2} \end{pmatrix} - \begin{pmatrix} \frac{1-2x_1}{x_1(1-x_1)} & 0 \\ 0 & \frac{1-2x_2}{x_2(1-x_2)} \end{pmatrix} \begin{pmatrix} -\frac{(1-2x_1)}{x_1(1-x_1)} & 0 \\ 0 & -\frac{(1-2x_2)}{x_2(1-x_2)} \end{pmatrix} \\ &= \begin{pmatrix} \frac{6((x_1-1/2)^2+1/12)}{x_1^2(1-x_1)^2} & 0 \\ 0 & \frac{6((x_2-1/2)^2+1/12)}{x_2^2(1-x_2)^2} \end{pmatrix} \end{aligned}$$

Logo, H_x^h é uma matriz semidefinida positiva. Logo, h é uma função convexa sobre $((0,1)^2, X^{-2}(I-X)^{-2})$. Portanto, f é uma função quase-convexa sobre essa variedade.

Dizemos que $f : M \to I\!R \cup \{+\infty\}$ é dito semicontínua inferior no ponto $\bar{x} \in M$ se, para qualquer sequência $\{x^k\}$ que converge para $\bar{x}$, temos que

$$\liminf_{k\to\infty} f(x^k) \geq f(\bar{x}).$$

A função f é semicontínua inferior em M se ela é semicontínua inferior em cada ponto de M.

Além disso, a função $f : M \to I\!R \cup \{+\infty\}$ é dito localmente Lipschitz se para cada $z \in M$ existe uma constante $L := L_z > 0$ tal que para cada x, y dentre uma vizinhança de $z \in M$ temos

$$|f(x) - f(y)| \leq Ld(x,y). \tag{2.28}$$

A funcão $f : M \to I\!R \cup \{+\infty\}$ é dito ser coerciva se para cada sequência $\{x^k\} \subset$

M para o qual $\lim_{k\to+\infty} d(x^k, x) = +\infty$, para algum $x \in M$, temos que

$$\limsup_{k\to+\infty} f(x^k) = +\infty.$$

Exemplo 2.2 *Seja M uma variedade de Hadamard e $x \in M$ um ponto fixo. A função $f : M \to I\!R \cup \{+\infty\}$ tal que $f(\cdot) = d^2(\cdot, x)$ é uma função coerciva. De fato, seja $\{y^k\} \subset M$ tal que $\lim_{k\to+\infty} d(y^k, w) = +\infty$ para algum $w \in M$. Aplicando o Teorema 2.3 no triângulo geodésico $[y^k, x, w]$, temos*

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

Como $-d(y^k, w)d(x, w) \leq -\langle \exp_w^{-1} y^k, \exp_w^{-1} x\rangle$, temos

$$(d(y^k, w) - d(x, w))^2 \leq d^2(y^k, x) := f(y^k).$$

Agora, suponha que $\limsup_{k\to+\infty} f(y^k) < +\infty$. Isto é, existe $M > 0$ tal que $|d(y^k, w) - d(x, w)| \leq M$. Então, $\lim_{k\to+\infty} d(y^k, w) \leq M + d(x, w)$. Como $d(x, w)$ é finito, segue que $\lim_{k\to+\infty} d(y^k, w) < +\infty$, o qual é uma contradição

A seguir, os seguintes resultados garantem a existencia de mínimos globais para uma função própria, semicontínua inferior e simples-valorada com domínio compato.

Teorema 2.6 *Seja $f : M \longrightarrow I\!R$ ser uma função própria e semicontínua inferior. Se $\mathrm{dom}(f)$ é compacto, então f atinge seu mínimo global em M.*

Prova. Primeiro, vamos mostrar que f é limitado inferiormente. Para isso, suponha por contradição que existe uma sequência $\{x^k\} \subset \mathrm{dom}(f)$ tal que

$$\lim_{k\to+\infty} f(x^k) = -\infty. \tag{2.29}$$

Como $\mathrm{dom}(f)$ é compacto, existe $\hat{x} \in dom(f)$ e uma subsequência $\{x^{k_j}\} \subset \{x^k\}$ tal que $\lim_{j\to+\infty} x^{k_j} = \widehat{x}$. Desde que f é semicontínua inferior, temos

$$\liminf_{j\to+\infty} f(x^{k_j}) \geq f(\widehat{x}),$$

o qual contradiz a relação (2.29). Portanto, f é limitado inferiormente. Isso implica que existe $f^* \in I\!R$ tal que $f^* := inf\{f(x) \mid x \in M\}$. Segundo a propriedade do ínfimo, existe uma subsequência $\{x^k\} \subset \mathrm{dom}(f)$ tal que $\lim_{k\to\infty} f(x^k) = f^*$.

Desde que $\mathrm{dom}(f)$ é compacto, existe $\bar{x} \in dom(f)$ e $\{x^{k_j}\} \subseteq \{x^k\}$ tal que

$\lim_{j\to\infty} x^{k_j} = \bar{x}$. Além disso, desde que f é semicontínua inferior, temos

$$\liminf_{j\to\infty} f(x^{k_j}) \geq f(\bar{x}). \tag{2.30}$$

Como $\{f(x^k)\}$ converge para f^*, então a subsequência $\{f(x^{k_j})\}$ converge para f^*, e de (2.30) temos $f^* \geq f(\bar{x})$. Portanto, $\bar{x}$ é um mínimo global de f em M. ■

Observação 2.1 *Observe que a compacidade é uma condição muito forte para garantir a existência de mínimos globais.*

Com o intuito de não exigir condições muito fortes ao Teorema 2.6, definimos, para $\alpha \in I\!R$, o seguinte conjunto

$$L_f(\alpha) := \{x \in M \mid f(x) \leq \alpha\},$$

o qual é chamado *conjunto de nível inferior* de f.

Corolario 2.1 *Seja $\alpha \in I\!R$ e $f : M \to I\!R \cup \{+\infty\}$ uma função própria e semicontínua inferior em $L_f(\alpha)$. Se $L_f(\alpha)$ é compacto e não vazio, então existe um mínimo global de f em M.*

Prova. Seja um ponto arbitrário $x \in dom(f)$. Se $x \in L_f(\alpha)$, por Teorema 2.6 aplicado a $L_f(\alpha)$, segue que existe um minimizador global de f em $L_f(\alpha)$. Isso pode ser tomado como $\bar{x} \in L_f(\alpha)$ tal que $f(\bar{x}) \leq f(x)$. Caso contrário, para todo $x \notin L_f(\alpha)$ temos que $f(x) > \alpha \geq f(\bar{x})$. Logo, $f(\bar{x}) < f(x)$. Portanto, em ambos os casos, concluimos que $\bar{x} \in dom(f)$ é um minimizador global de f em M. ■

Corolario 2.2 *Seja $f : M \to I\!R \cup \{+\infty\}$ uma função coerciva e semicontínua inferior, então existe um minimizador global de f.*

Prova. Desde que $M \neq \emptyset$, existe $\tilde{x} \in M$ tal que $L_f(\alpha) \neq \emptyset$ com $\alpha = f(\tilde{x})$. Agora, mostraremos que $L_f(\alpha)$ é limitado. Por contradição, suponha que $L_f(\alpha)$ é não limitado. Então, existe uma sequência $\{x^k\} \subset L_f(\alpha)$ tal que

$$\lim_{k\to+\infty} d(x^k, \bar{x}) = +\infty,$$

para algum $\bar{x} \in L_f(\alpha)$. Desde que $x^k \in L_f(\alpha)$, temos que $f(x^k) \leq \alpha$. Tomando o limite superior, quando k converge para $+\infty$, e usando a propriedade de coercividade de f obtemos $+\infty = \limsup_{k\to+\infty} f(x^k) \leq \alpha$, o qual é uma contradição . Portanto, o conjunto $L_f(\alpha)$ é limitado.

Agora provaremos que $L_f(\alpha)$ é um conjunto fechado. Seja $\{x^k\} \subset L_f(\alpha)$ uma sequência tal que $\{x^k\}$ converge para $\bar{x}$ e suponha que $\bar{x} \notin L_f(\alpha)$, isto é, $f(\bar{x}) > \alpha$.

Como consequência da semicontinuidade de f, segue-se que

$$\alpha < f(\bar{x}) \leq \liminf_{k \to +\infty} f(x^k) \leq \alpha,$$

o qual é uma contradição. Portanto, $\bar{x} \in L_f(\alpha)$ e então $L_f(\alpha)$ é um conjunto fechado.

Para finalizar, como M é uma variedade Riemanniana completa e usando o conhecido Teorema Hopf-Rinow, ver por exemplo Do Carmo[27, Teorema 2.8], temos que $L_f(\alpha)$ é um conjunto compacto. Logo, o resultado segue como consequência do Corolario 2.1. ■

Seja M uma variedade de Hadamard. Seja o conjunto $D \subset M$ convexo e seja $x \in D$. O cone normal da análise convexa de D em x, denotado por $\mathcal{N}_D(x)$, é definido como

$$\mathcal{N}_D(x) := \{v \in T_xM \mid \langle v, \exp_x^{-1} y \rangle \leq 0, \quad \forall\, y \in D\}. \tag{2.31}$$

Se $D = \{x\}$, então $\mathcal{N}_D(x) = T_xM$.

Além disso, seja $D \subset M$ definimos o cone normal de Fréchet de D em $x \in M$ como segue

$$\hat{N}_D(x) := \left\{ v \in T_xM \mid \limsup_{\substack{D \\ u \to x}} \frac{\langle v, \exp_x^{-1} u \rangle}{d(u,x)} \leq 0 \right\}, \tag{2.32}$$

onde a notação $\overset{D}{u \to x}$ significa que $u \to x$ e $u \in D$.

Por outro lado, o cone normal do limite, denotado por $N_D(x)$, está definido como

$$N_D(x) = \limsup_{u \to x} \hat{N}_D(u), \tag{2.33}$$

o qual é geralmente não convexo e não fechado.
De (2.32) e (2.33), segue que $\hat{N}_D(x) \subset N_D(x)$. Se $D \subset M$ é convexo então

$$\hat{N}_D(x) = N_D(x) = \mathcal{N}_D(x). \tag{2.34}$$

2.3 Cálculo não diferenciável

Com o objetivo de resolver problemas de otimização não convexas e não suaves, tem sidos generalizados em espaços vetoriais muitos conceitos de subgradiente e subdiferencial de diferentes formas, como consequência, existem diversos subdiferenciais tais como Fréchet, Clarke, Clarke-Rockafellar, Hadamard, Dinni, subdiferencial abstrato, entre outros.

Nesta seção, vamos estender a definição de subdiferencial abstracto de espaços de Banach, introduzido por Aussel et al.[6], para variedades de Hadamard. Além disso, provaremos algumas propriedades o qual serão utilizadas ao longo desta tese. Note que a generalização para variedades Riemannianas é natural devido que não se precisa da informação da curvatura do espaço.

Definição 2.1 *Seja M uma variedade de Hadamard. Chamamos subdiferencial abstracto, denotado por ∂, qualquer operador associado ao conjunto $\partial f(x)$ de T_xM para qualquer função semicontínua inferior $f : M \to I\!R \cup \{+\infty\}$ e qualquer $x \in M$, satisfazendo as seguintes propriedades:*

(P1) *Se f é convexa, então $\partial f = \partial^{FM} f$ onde*

$$\partial^{FM} f(x) := \{g \in T_xM \mid \langle g, \exp_x^{-1} z\rangle + f(x) \leq f(z), \quad \forall\, z \in M\}$$

é chamado o subdiferencial de Fenchel-Moreau;

(P2) $0 \in \partial f(x)$, *se $x \in M$ é um ponto mínimo local de f;*

(P3) $\partial(f+g)(x) \subset \partial f(x) + \partial g(x)$, *sempre que $g : M \to I\!R \cup \{+\infty\}$ é uma função convexa contínua, o qual é ∂-diferenciável em $x \in M$,*

onde a função g é ∂-diferenciável em x significa que $\partial g(x)$ e $\partial(-g)(x)$ são não vazios. Dizemos que uma função f é ∂-subdiferenciável em x quando $\partial f(x)$ é não vazio.

Exemplo 2.3 *O subdiferencial abstrato abrange uma grande classe de subdiferenciais clássicos conhecidos na literatura de análise variacional. Dentre aqueles subdiferenciais que satisfazem a definição temos*

i). *Seja $f : M \to I\!R \cup \{+\infty\}$ uma função semicontínua inferior com $dom(f) \neq \emptyset$. O subdiferencial de Fréchet de f no ponto $x \in dom(f)$ é definido como o conjunto $\partial^F f(x)$ de todo $s \in T_xM$ com a propriedade que*

$$\liminf_{\substack{u \to x \\ u \neq x}} \frac{1}{d(x,u)} \left[f(u) - f(x) - \langle s, \exp_x^{-1} u\rangle_x \right] \geq 0.$$

Claramente, $s \in \partial^F f(x)$ se e somente se para cada $\eta > 0$, existe $\epsilon > 0$ tal que

$$\langle s, \exp_x^{-1} u\rangle_x \leq f(u) - f(x) + \eta d(x,u), \quad \text{para todo } u \in B(x,\epsilon). \tag{2.35}$$

Se $x \notin dom(f)$ então estabelecemos $\partial^F f(x) = \emptyset$.

ii). *Seja $f : M \to I\!R \cup \{+\infty\}$ uma função própria localmente Lipschitz, com $dom(f) \neq \emptyset$, em $x \in dom(f)$ e $d \in T_xM$. A derivada direcional generalizada de Clarke de f em $x \in M$ na direção $d \in T_xM$, denotado como $f^\circ(x,d)$, é definido como segue*

$$f^\circ(x,d) := \limsup_{\substack{u \to x \\ t \searrow 0}} \frac{f(\exp_u t(D\exp_x)_{\exp_x^{-1}u} d) - f(u)}{t},$$

onde $(D\exp_x)_{\exp_x^{-1}u} : T_{\exp_x^{-1}u}(T_xM) \simeq T_xM \to T_uM$ é o diferencial da aplicação exponencial em $\exp_x^{-1}u$. O subdiferencial generalizado de Clarke de f em $x \in M$, é o conjunto $\partial^\circ f(x)$ de $T_xM^* \simeq T_xM$ definido como

$$\partial^\circ f(x) := \{s \in T_xM \mid \langle s, d\rangle_x \leq f^\circ(x,d),\ \forall\, d \in T_xM\}.$$

Note que para uma função localmente Lipschitz $f : M \to I\!R \cup \{+\infty\}$, o conjunto $\partial^\circ f(x)$ é um subconjunto convexo fechado e não vazio de T_xM para cada $x \in M$.

iii). *Seja $f : M \to I\!R \cup \{+\infty\}$ uma função semicontínua inferior. Então ao conjunto*

$$\partial^p f(x) := \{d\psi(x) : \psi \in C^2(M),\ f - \psi \text{ alcança um mínimo local em } x\}$$

chamamos de subdiferencial proximal de f em $x \in M$.

Outra forma de caracterizar os $s \in \partial^p f(x)$ é da seguinte forma: $s \in \partial^p f(x)$ se e somente se, existe um número positivo ρ tal que

$$f(y) \geq f(x) + \langle s, \exp_x^{-1} y\rangle - \rho d(x,y)^2$$

iv). *O subdiferencial de Hadamard de $f : M \to I\!R \cup \{+\infty\}$, no ponto $x \in M$, é o conjunto definido como*

$$\partial^H f(x) := \{s \in T_xM \mid \langle s, d\rangle \leq \liminf_{\substack{t \searrow 0 \\ w \to d}} \frac{f(\exp_x tw) - f(x)}{t},\ \forall\, d \in T_xM\}$$

v). *A derivada de Clarke-Rockafellar de $f : M \to I\!R \cup \{+\infty\}$ no ponto $x \in M$, na direção $v \in T_xM$, é dado por*

$$f^\uparrow(x,v) = \sup_{\varepsilon > 0} \limsup_{\substack{u \xrightarrow{f} x \\ t \searrow 0}} \inf_{d \in B_\epsilon(v)} \frac{f(\exp_u t(D\exp_x)_{\exp_x^{-1}u} d) - f(u)}{t},$$

onde $(D\exp_x)_{\exp_x^{-1}u}$ *é o diferencial da função exponencial,* $\exp_x$*, em* $\exp_x^{-1}u$*,* $B_\epsilon(v) = \{d \mid \|d-v\| < \varepsilon\}$*,* $t \searrow 0$ *representa o fato que* $t > 0$ *e* $t \to 0$*, e* $u \xrightarrow{f} x$ *significa que* $u \to x$ *e* $f(u) \to f(x)$*. O subdiferencial de Clarke-Rockafellar de* f *no ponto* $x \in M$ *é definido como*

$$\partial^{CR} f(x) = \left\{ s \in T_xM \mid \langle s, v\rangle \leq f^{\uparrow}(x,v), \ \ \forall\, v \in T_xM \right\}.$$

vi). A derivada de Dinni superior em $x \in M$ *na direção* $v \in T_xM$ *é definido por*

$$f^{D^+}(x,v) = \limsup_{t\searrow 0} \frac{f(\exp_x(tv)) - f(x)}{t},$$

onde o subdiferencial associado com essa derivada, chamado de subdiferencial de Dini superior, é dado por

$$\partial^{D^+} f(x) = \left\{ s \in T_xM \mid \langle s, v\rangle \leq f^{D^+}(x,v), \ \ \forall\, v \in T_xM \right\}.$$

Na seção anterior, revisamos alguns conceitos de alguns tipos de cones. O seguinte resultado mostra que o cone normal de Fréchet pode ser caracterizado em termos de subdiferenciabilidade. Consideremos uma função indicadora δ_D de $D \subset M$: $\delta_D(u) = 0$ para $u \in D$ e $\delta_D(u) = +\infty$ no caso contrário.

Lema 2.1 *Seja* $f : M \to I\!R \cup \{+\infty\}$ *uma função convexa então* $\hat{N}_{epi(f)}(\bar{x}, f(\bar{x})) = \partial^F \delta_{epi(f)}(\bar{x}, f(\bar{x}))$.

Prova. Seja $w \in \partial^F \delta_{epi(f)}(\bar{x}, f(\bar{x}))$. Como $(\bar{x}, f(\bar{x})) \in epi(f)$ temos que $\delta_{epi(f)}(\bar{x}, f(\bar{x})) = 0$. De (2.35), para cada $\eta > 0$ existe $\epsilon > 0$ tal que

$$\langle w, \exp^{-1}_{(\bar{x},f(\bar{x}))}(x, f(x))\rangle \leq \delta_{epi(f)}(x, f(x)) + \eta d((x, f(x)), (\bar{x}, f(\bar{x}))), \tag{2.36}$$

para $(x, f(x)) \in B((\bar{x}, f(\bar{x})), \epsilon)$.

Se $(x, f(x)) \notin epi(f)$, então $\delta_{epi(f)}(x, f(x)) = +\infty$. Logo, de (2.36) temos que

$$\frac{\langle w, \exp^{-1}_{(\bar{x},f(\bar{x}))}(x, f(x))\rangle}{d((x, f(x)), (\bar{x}, f(\bar{x})))} \leq +\infty. \tag{2.37}$$

Se $(x, f(x)) \in epi(f)$, temos que $\delta_{epi(f)}(x, f(x)) = 0$. Logo, de (2.36) temos que

$$\frac{\langle w, \exp^{-1}_{(\bar{x},f(\bar{x}))}(x, f(x))\rangle}{d((x, f(x)), (\bar{x}, f(\bar{x})))} \leq \eta. \tag{2.38}$$

Tomando limite superior na relação (2.38) e desde que $\eta > 0$ foi tomado arbitrariamente, temos que

$$\limsup_{(x,f(x))\to(\bar{x},f(\bar{x}))} \frac{\langle w, \exp^{-1}_{(\bar{x},f(\bar{x}))}(x,f(x))\rangle}{d((x,f(x)),(\bar{x},f(\bar{x})))} \leq 0. \tag{2.39}$$

Portanto, de (2.37) e (2.39) obtemos o resultado. ∎

Lema 2.2 *Seja* $f : M \to I\!R \cup \{+\infty\}$ *uma função semicontínua inferior. Então* $s \in \partial^F f(x)$ *se e somente se* $(s,-1) \in \hat{N}_{epi(f)}(x, f(x))$.

Prova. A prova é simples. De fato, seja $(s,-1) \in \hat{N}_{epi(f)}(x, f(x))$, então

$$\begin{aligned}
0 &\geq \limsup_{(u,f(u))\to(x,f(x))} \frac{\langle (s,-1), \exp^{-1}_{(x,f(x))}(u,f(u))\rangle}{d((u,f(u)),(x,f(x)))} \leq 0 \\
&= \limsup_{(u,f(u))\to(x,f(x))} \frac{\langle (s,-1), (\exp^{-1}_x u, f(u)-f(x))\rangle}{d((u,f(u)),(x,f(x)))} \\
&= \limsup_{(u,f(u))\to(x,f(x))} \frac{\langle s, \exp^{-1}_x u\rangle - (f(u)-f(x))}{d((u,f(u)),(x,f(x)))} \\
&= -\liminf_{(u,f(u))\to(x,f(x))} \frac{f(u)-f(x)-\langle s, \exp^{-1}_x u\rangle}{d((u,f(u)),(x,f(x)))}
\end{aligned}$$

Logo, $s \in \partial^F f(x)$.

Suponha agora que $s \in \partial^F f(x)$. De (2.35), para cada $\eta > 0$ existe $\epsilon > 0$ tal que

$$\begin{aligned}
\eta d(x,u) &\geq \langle s, \exp^{-1}_x u\rangle - (f(u)-f(x)) \\
&= \langle (s,-1), (\exp^{-1}_x u, f(u)-f(x))\rangle \\
&= \langle (s,-1), (\exp^{-1}_{(x,f(x))}(u,f(u)))\rangle.
\end{aligned}$$

Então, dividindo ambos os membros por $d(x,u)$ e tomando limite superior, quando $d(x,u) \to 0$, temos que

$$\limsup_{d(x,u)\to 0} \frac{\langle (s,-1), ((u-x), f(u)-f(x))\rangle}{d(x,u)} \leq \eta.$$

O resultado segue do fato que $\eta > 0$ é arbitrário. ∎

Lema 2.3 *Seja* $f : M \to I\!R \cup \{+\infty\}$. *Então* $s \in \partial^{FM} f(x)$ *se e somente se* $(s,-1) \in N_{epi(f)}(x, f(x))$.

Prova. Imediato.

Proposição 2.2 *Os subdiferenciais apresentados no Exemplo 2.3 satisfazem as propriedades da Definição 2.1.*

Prova.

i). Caso $\partial := \partial^F$. (P1) De relação (2.34) e Lemas 2.2 e 2.3 temos que $\partial^{FM} = \partial^F$.
(P2) Imediato.
(P3) Seja $s \in \partial^F(f+g)(\bar{x})$ e usando a propriedade da regra da suma para funções semicontínuas inferiores, ver por exemplo [35, Theorem 4.4], para qualquer $\epsilon > 0$, existe x_1, $x_2 \in B(\bar{x}, \epsilon)$, $s_1 \in \partial^F f(x_1)$ tal que $|f(x_1) - f(\bar{x})| < \epsilon$ e $s_2 \in \partial^F g(x_2)$ tal que $|g(x_2) - g(\bar{x})| < \epsilon$ satisfazendo

$$|\langle s, v\rangle_{\bar{x}} - \sum_{i=1}^{L} \langle s_i, v\rangle_{x_i}| < \epsilon, \tag{2.40}$$

para qualquer $v \in V^\infty(M)$.

Além disso, como $s_2 \in \partial^F g(x_2)$ e para qualquer $\epsilon > 0$ encontrar $\tau = \epsilon - d(x_2, \bar{x})$ tal que

$$\langle s_2, \exp_{x_2}^{-1} u\rangle_{x_2} \leq g(u) - g(x_2) + \epsilon d(u, x_2), \tag{2.41}$$

sempre que $d(x_2, u) < \tau$.
Por outro lado, podemos tomar $v := \exp_{x_2}^{-1} u \in T_{x_2}M$. Então, relação (2.40) satisfaz

$$\begin{aligned} \epsilon + \langle s_2, v\rangle_{x_2} &> \langle s, \mathcal{P}_{x_2,\bar{x}} v\rangle_{\bar{x}} - \langle s_1, \mathcal{P}_{x_2,x_1} v\rangle_{x_1} \\ &= \langle s, \mathcal{P}_{x_2,\bar{x}} v\rangle_{\bar{x}} - \langle \mathcal{P}_{x_1,\bar{x}} s_1, \mathcal{P}_{x_1,\bar{x}} \mathcal{P}_{x_2,x_1} v\rangle_{x_1} \\ &= \langle s - \mathcal{P}_{x_1,\bar{x}} s_1, \mathcal{P}_{x_2,\bar{x}} v\rangle_{\bar{x}}. \end{aligned} \tag{2.42}$$

Denote $z := \exp_{\bar{x}} \mathcal{P}_{x_2,\bar{x}} v$. Logo, segue de (2.41) e (2.42), para todo $z \in B(\bar{x}, \tau)$, que

$$\begin{aligned} \langle s - \mathcal{P}_{x_1,\bar{x}}\, s_1, \exp_{\bar{x}}^{-1} z\rangle_{\bar{x}} &< g(u) - g(x_2) + \epsilon d(u, x_2) + \epsilon \\ &= g(z) - g(\bar{x}) + g(\bar{x}) - g(z) + g(u) - g(x_2) + \epsilon d(u, x_2) + \epsilon \end{aligned} \tag{2.43}$$

Desde que g é localmente Lipschitz, existe uma constante positiva $L_{\bar{x}}$ e L_{x_2} tal que relação (2.43) satisfaz

$$\begin{aligned} \langle s - \mathcal{P}_{x_1,\bar{x}}\, s_1, \exp_{\bar{x}}^{-1} z\rangle_{\bar{x}} &< g(z) - g(\bar{x}) + L_{\bar{x}} d(\bar{x}, z) + (L_{x_2} + \epsilon) d(u, x_2) + \epsilon \\ &= g(z) - g(\bar{x}) + (L_{\bar{x}} + L_{x_2} + \epsilon) d(\bar{x}, z) + \epsilon. \end{aligned} \tag{2.44}$$

A última desigualdade é obtida do fato que $d(u, x_2) = d(\bar{x}, z)$.

Portanto, para $\epsilon > 0$ suficientemente pequeno, relação (2.44) implica

$$\langle s - \mathcal{P}_{x_1,\bar{x}} s_1, z - \bar{x} \rangle \le g(z) - g(\bar{x}) + (L_{\bar{x}} + L_{x_2}) d(\bar{x}, z),$$

sempre que $z \in B(\bar{x}, \tau)$. Logo, $s - \mathcal{P}_{x_1,\bar{x}} s_1 \in \partial^F g(\bar{x})$. Isso implica que $s \in \mathcal{P}_{x_1,\bar{x}} s_1 + \partial^F g(\bar{x})$. Além disso, nessas condições pode ser provado que se $s_1 \in \partial^F f(x_1)$ então $\mathcal{P}_{x_1,\bar{x}} s_1 \in \partial^F f(\bar{x})$. Portanto, $s \in \partial^F f(\bar{x}) + \partial^F g(\bar{x})$ e obtemos o resultado.

iv). Seja $\partial := \partial^{CR}$. $(P1)$ é imediato.
$(P2)$ Seja $x \in M$ um ponto mínimo local de f. Suponha que $0 \notin \partial^{CR} f(x)$.Isso implica que

$$\sup_{\varepsilon > 0} \limsup_{\substack{u \xrightarrow{f} x \\ t \searrow 0}} \inf_{d \in B_\varepsilon(x)} \frac{f(\exp_u t (D\exp_x)_{\exp_x^{-1} u} d) - f(u)}{t},$$

$(P3)$ Seja $s \in \partial^{CR}(f+g)(x)$. Para todo $v \in T_x M$, temos

$$\langle s, v \rangle \le \sup_{\epsilon > 0} \limsup_{\substack{u \to x \\ t \searrow 0}} \inf_{d \in B_\epsilon(v)} \frac{(f+g)(\exp_u t (D\exp_x)_{\exp_x^{-1} u} v) - (f+g)(u)}{t}.$$

v). Seja $\partial := \partial^{D^+}$. $(P1)$ Seja $s \in \partial^{D^+} f(x)$ e $f : M \to I\!R \cup \{+\infty\}$ uma função convexa. Mostraremos que $\partial^{D^+} f(x) = \{s \in T_x M \mid \langle s, \exp_x^{-1} z \rangle + f(x) \le f(y),\ \forall z \in M\}$. Toma qualquer $v \in T_x M$. Denotemos $\psi(t) := \exp_x tv$ uma curva geodésica tal que $\psi(0) = x$ e $\psi(1) = \exp_x v$. Para todo $t \in [0,1]$ e da convexidade de f, temos

$$\frac{f(\psi(t)) - f(x)}{t} \le \frac{(1-t) f(x) + t f(\exp_x v) - f(x)}{t} = f(\exp_x v) - f(x).$$

Logo, para t suficientemente pequeno, segue que

$$\limsup_{t \searrow 0} \frac{f(\psi(t)) - f(x)}{t} \le f(\exp_x v) - f(x),$$

e consequentemente,

$$\langle s, v \rangle \le f(\exp_x v) - f(x). \tag{2.45}$$

Portanto, o resultado segue tomando $z := \exp_x v$ in (2.45).

$(P2)$ Seja $x \in M$ um ponto de mínimo local de f, para todo $v \in T_x M$ existe $\delta > 0$ tal que

$$f(x) \le f(\exp_x tv), \qquad \forall\, t \in (0, \delta)$$

consequentemente, para t suficientemente pequeno, obtemos

$$0 \leq \liminf_{t \searrow 0} \frac{f(\exp_x tv) - f(x)}{t}.$$

Portanto, $0 \in \partial^{D^+} f(x)$.

$(P3)$ Seja $s \in \partial^{D^+}(f+g)(x)$. Da continuidade de g, para todo $v \in T_xM$ temos

$$\begin{aligned}
\langle s, v\rangle &\leq \limsup_{t \searrow 0} \left(\frac{(f+g)(\exp_x(tv)) - (f+g)(x)}{t} \right) \\
&\leq \limsup_{t \searrow 0} \frac{f(\exp_x(tv)) - f(x)}{t} + \lim_{t \searrow 0} \frac{g(\exp_x(tv)) - g(x)}{t} \\
&= \limsup_{t \searrow 0} \frac{f(\exp_x(tv)) - f(x)}{t} + g'(x,v), \qquad (2.46)
\end{aligned}$$

onde $g'(x,v)$ é a derivada direccional na direção de g em x na direção v. De (2.25) e item a. (desde que g é convexa) existe $\bar{r} \in \partial g(x)$ tal que $g^{D^+}(x,v) = g'(x,v) = \langle \bar{r}, s\rangle$. Portanto, (2.46) fica como

$$\langle s - \bar{r}, v\rangle \leq \limsup_{t \searrow 0} \frac{f(\exp_x(tv)) - f(x)}{t}.$$

Isso implica que $s - \bar{r} \in \partial^{D^+} f(x)$ e, portanto, obtemos que $s \in \partial^{D^+} f(x) + \partial^{D^+} g(x)$. ■

Lema 2.4 *Seja $f : M \to I\!R \cup \{+\infty\}$ é uma função própria localmente Lipschitz. Então $\partial^F f(x) \subset \partial^{CR} f(x)$ para todo $x \in dom(f)$.*

Prova. Primeiro vamos provar que $\partial^F f(x) \subset \partial^\circ f(x)$. De fato, isso é consequência de $\partial^F f(x) = \partial^F(f \circ \exp_x)(0_x)$, $\partial^\circ f(x) = \partial^\circ(f \circ \exp_x)(0_x)$ e por Correa et al.[20, Propriedade 2.7]. Logo, é imediato observar que $\partial^\circ f(x) \subset \partial^{CR} f(x)$. Portanto, conseguimos o resultado. ■

Repare que, no cálculo da regra da soma de subdiferenciais, por exemplo de Fréchet, uma consequência da definição desse subdiferencial é

$$\partial^F f_1(x) + \partial^F f_2(x) \subseteq \partial^F(f_1 + f_2)(x),$$

mas a outra inclusão

$$\partial^F(f_1 + f_2)(x) \subseteq \partial^F f_1(x) + \partial^F f_2(x)$$

não toma em geral. Mas a regra da soma de subdiferenciais é um resultado central de qualquer cálculo subdiferencial. Então, pode se observar que a teoria subdiferencial tem alguns detalhes a serem consideradas na regras de cálculo dos mesmos. Essas dificuldades muitas vezes começam com cálculos básicos quando tratamos com operações básicas sobre funções, tais como soma de número finito de funções ou ainda quando tratados com produtos por uma constante negativa. Alguns subdiferenciais, por exemplo, o subdiferencial de Clarke de uma função Lipschitz, a multiplicação por uma constante negativa ainda preserva a regra da igualdade, isso é, $\partial^\circ(-\alpha f)(x) = -\alpha\partial^\circ f(x)$, para $\alpha > 0$.

A seguir, veremos algumas regras que cumprem os diferenciais que formam parte do subdiferencial abstrato e que serão importantes do decorrer da tese.

Lema 2.5 *Sejam $f_i : M \to I\!R \cup \{+\infty\}$, $i \in \mathcal{I} \subset I\!N$, funções próprias localmente Lipschitz. Então, para todo $d \in T_xM$,*

$$\begin{aligned} (\sum_{i\in\mathcal{I}} f_i)^\circ(x,d) &\leq \sum_{i\in\mathcal{I}} f_i^\circ(x,d) && (2.47)\\ (\alpha f_i)^\circ(x,d) &= \alpha(f_i^\circ(x,d)), \quad \forall\, \alpha \geq 0. && (2.48)\\ f_i^\circ(x,\alpha d) &= \alpha f_i^\circ(x,d), \quad \forall\, \alpha \geq 0. && (2.49) \end{aligned}$$

Prova. A relação (2.47) é consequência de combinar as propriedades de limite superior com a definição de derivada direcional de Clarke. A relação (2.48) é imedita. Por último, a relação (2.49) (o que significa que a função $d \mapsto f^\circ(x,d)$ é positivamente homogéneo) é simples de provar para cada $d \in T_xM$. De fato, para $\alpha > 0$ temos

$$\begin{aligned} \alpha f^\circ(x,d) &= \limsup_{\substack{u\to x\\ t\searrow 0}} \frac{\alpha f(\exp_u t(D\exp_x)_{\exp_x^{-1}u}d) - \alpha f(u)}{t}\\ &= \limsup_{\substack{u\to x\\ (t/\alpha)\searrow 0}} \frac{f(\exp_u (t/\alpha)(D\exp_x)_{\exp_x^{-1}u}\alpha d) - f(u)}{t/\alpha}\\ &= f^\circ(x,\alpha d). \end{aligned}$$

onde a segunda igualdade é devido á linearidade da aplicação diferencial. Para o caso $\alpha = 0$ o caso é imediato. ∎

Lema 2.6 *Sejam $f_i : M \to I\!R$, $i \in \mathcal{I} \subset I\!N$, funções localmente Lipschitz. Então,*

$$\begin{aligned} \partial^\circ(\sum_{i\in\mathcal{I}} f_i)(x) &\subset \sum_{i\in\mathcal{I}} \partial^\circ f_i(x), && (2.50)\\ \partial^\circ(\alpha f_i)(x) &= \alpha\partial^\circ f_i(x) && (2.51) \end{aligned}$$

Prova. Do Lema 2.5, a relação (2.50) segue imediatamente. Para provar a relação (2.51), considere $s \in \partial^\circ(\alpha f_i)(x)$. Da relação (2.48), para todo $d \in T_xM$ temos que

$\langle s, d\rangle \leq \alpha f_i^\circ(x, d)$. Logo, $s/\alpha \in \partial^\circ f(x)$ desde que α é um número não negativo. Portanto, $s \in \alpha\, \partial^\circ f(x)$. O contrário pode ser provado de forma análoga. ■

Lema 2.7 *Seja $\Omega \subset M$ um conjunto aberto. Se $f : M \to I\!R$ é localmente Lipschitz e $g : M \to I\!R$ é convexa e diferenciável sobre Ω, então*

$$\partial^\circ(f+g)(x) = \partial^\circ f(x) + grad\, g(x), \qquad \forall\, x \in \Omega$$

Prova. Ver Bento et al.[12, Lema 3.1]. ■

Proposição 2.3 *Seja $f : M \to I\!R \cup \{+\infty\}$ uma função localmente Lipschitz em M e $g : I\!R \to I\!R$ ser uma função continuamente diferenciável em $I\!R$. Então, para qualquer $\bar{x} \in dom(f)$, temos que*

$$(g \circ f)^\circ(\bar{x}, d) = g'(f(\bar{x}))f^\circ(\bar{x}, d), \qquad \forall\; d \in T_{\bar{x}}M$$

Prova. Seja $x \in dom(f)$ tal que x converge para $\bar{x}$. Seja $\tilde{d} := (D\exp_{\bar{x}})_{\exp_{\bar{x}}^{-1}x} d \in T_xM$, para qualquer $d \in T_xM$. Considere $\gamma : I\!R \to M$ uma geodésica tal que $\gamma(x, 1, t\tilde{d}) := \exp_x t\tilde{d} \in M$ com $t > 0$. Como $t \searrow 0$, $\gamma(x, 1, t\tilde{d})$ converge para x. Como x converge para $\bar{x}$, $\lim_{x\to\bar{x}}(D\exp_{\bar{x}})_{\exp_{\bar{x}}^{-1}x} d = (D\exp_{\bar{x}})_0 d = d$, então $\gamma(x, 1, t\tilde{d})$ converge para $\bar{x}$.

Definimos

$$r(y) := \begin{cases} \frac{g(y)-g(f(x))}{y-f(\bar{x})} - g'(f(x)), & y \neq f(x) \\ 0, & y = f(x) \end{cases}$$

Como g é continuamente diferenciável, segue que

$$\lim_{y\to f(x)} r(y) = 0.$$

Portanto, temos

$$g(y) - g(f(x)) = g'(f(x))(y - f(x)) + r(y)(y - f(x)). \tag{2.52}$$

Tome $y := f(\gamma(x, 1, t\tilde{d}))$. Desde que f é uma função localmente Lipschitz, y converge para $f(\bar{x})$ quando $x \to \bar{x}$ e $t \searrow 0$. Além disso, $\lim r(y) = 0$ quando $y \to f(\bar{x})$.

Aplicando limite superior à relação (2.52), temos

$$
\begin{aligned}
(g\circ f)^{\circ}(\bar{x},d) &= \limsup_{\substack{x\to\bar{x}\\ t\searrow 0}} \frac{(g\circ f)(\gamma(x,1,t\tilde{d}))-(g\circ f)(x)}{t} && (2.53)\\
&= g'(f(\bar{x}))\limsup_{\substack{x\to\bar{x}\\ t\searrow 0}} \frac{f(\gamma(x,1,t\tilde{d}))-f(x)}{t}\\
&= g'(f(\bar{x}))f^{\circ}(\bar{x},d). && (2.54)
\end{aligned}
$$

Logo, o resultado é consequência imediata da definição de subdiferencial generalizado de Clarke do lado esquerdo da relação (2.54). ∎

Cabe ressaltar que subdiferenciais podem ser usados para caracterizar funções semicontínuas inferiores quase-convexas, o qual pode ser observado no seguinte resultado. Esse resultado terá um papel muito importante no decorrer desta tese.

Lema 2.8 *Seja M uma variedade de Hadamard. Suponha que $f : M \to I\!R\cup\{+\infty\}$ seja uma função própria, semicontínua inferior e quase-convexa e que existe $g \in \partial f(x)$ tal que $\langle g, \exp_x^{-1}z\rangle > 0$. Então, temos as seguintes propriedades:*

i. *Se $\partial f \subset \partial^{CR} f$ e suponha que f seja contínua, então $f(x) \leq f(z)$.*

ii. *Se $\partial f \subset \partial^{D^+} f$, então $f(x) < f(z)$.*

Prova. i. No caso que $\partial f \subset \partial^{CR} f$. Seja $x, z \in M$ e $g \in \partial f(x)$ tal que

$$0 < \langle g, \exp_x^{-1}z\rangle \leq f^{\uparrow}(x, \exp_x^{-1}z).$$

Então, existe $\varepsilon' > 0$ tal que

$$0 \;<\; \inf_{\eta>0}\; \sup_{d(u,x)+t+|f(u)-f(x)|<\eta}\; \inf_{d\in B_{\epsilon'}(\exp_x^{-1}z)} \frac{f(\exp_u t(D\exp_x)_{\exp_x^{-1}u}d)-f(u)}{t}.$$

Logo, para todo $\eta > 0$ existe $u(\eta) \in M$ e $t(\eta) \in I\!R_{++}$ tal que $d(u(\eta), x) + t(\eta) + |f(u(\eta)) - f(x)| < \eta$ para todo $d \in B_{\varepsilon'}(\exp_x^{-1}z)$, obtemos

$$0 < \frac{f(\exp_{u(\eta)} t(\eta)(D\exp_x)_{\exp_x^{-1}u(\eta)}d) - f(u(\eta))}{t(\eta)}.$$

Considerando $\eta = 1/k$, da propriedade do ínfimo, existe sequências $\{u^k\} \in M$ e $\{t^k\} \in I\!R_{++}$ tal que $d(u^k, x) + t^k + |f(u^k) - f(x)| < 1/k$ e

$$f(\exp_{u^k} t^k(D\exp_x)_{\exp_x^{-1}u^k}(\exp_x^{-1}z)) > f(u^k). \tag{2.55}$$

Denotemos $r^k = (D\exp_x)_{\exp_x^{-1}u^k}(\exp_x^{-1}z)$, o qual resulta

$$\lim_{k\to+\infty} r^k = (D\exp_x)_0(\exp_x^{-1}z) = (\exp_x^{-1}z) \tag{2.56}$$

Seja $\alpha : [0,1] \to M$ uma geodésica minimizante tal que $\alpha(0) = u^k$ e $\alpha'(0) = r^k$. Portanto, equação (2.55) pode ser escrita como

$$f(u^k) < f(\alpha(u^k, 1, t^k r^k)) = f(\alpha(u^k, t^k, r^k)), \tag{2.57}$$

onde a última igualdade é obtida pela bem-conhecida propriedade de homogeneidade das geodésicas.

Por outro lado, desde que $t^k < 1$, segue que $0 < t^k < 1/k - d(u^k, x) < 1$, e portanto, usando a propriedade de quase-convexidade da função f, temos

$$\begin{aligned}\max\left\{f(\alpha(u^k,0,r^k), f(\alpha(u^k,1,r^k))\right\} &\geq f(\alpha(u^k,t^k,r^k)) \\ &> f(u^k),\end{aligned}$$

onde a última desigualdade é obtida pela equação (2.57). Desde que $f(\alpha(u^k,0,r^k) = f(u^k)$, temos que $f(\alpha(u^k,1,r^k) > f(u^k)$. Supondo continuidade, tomando limite para $k \to +\infty$ e considerando a equação (2.56) concluimos a prova do resultado no item i.

ii. No caso que $\partial f \subset \partial^{D^+}$, a prova é simples. De fato, se $x, z \in M$ e $g \in T_xM$ tal que

$$0 < \langle g, \exp_x^{-1}z\rangle \leq f^{D+}(x, \exp_x^{-1}z).$$

Logo, existe $\bar{t} \in]0,1[$ tal que

$$f(x) < f(\exp_x(\bar{t}\exp_x^{-1}z)). \tag{2.58}$$

Seja

$$\gamma(t) = \exp_x(t\exp_x^{-1}z) = \gamma(1, x, t\exp_x^{-1}z) = \gamma(t, x, \exp_x^{-1}z),$$

onde a última igualdade é obtida como consequência da propriedade homogênea da geodésica γ, ver [27, Lema 2.6]. Portanto, $\gamma(0) = x$ e $\gamma(1) = z$, então da propriedade de quase-convexidade de f e como $\bar{t} \in]0,1[$, temos

$$f(\exp_x(\bar{t}\exp_x^{-1}z)) = f(\gamma(\bar{t})) \leq \max\{f(x), f(z)\}$$

Da equação (3.23), concluimos que $f(x) < f(z)$. ■

Nesta tese também precisaremos do seguinte conceito: O subdiferencial limite de f em $x \in M$, denotado por $\partial^{Lim} f(x)$, é definido como segue

$$\partial^{Lim} f(x) := \{s \in T_x M \mid \exists\, x^k \to x,\ f(x^k) \to f(x), \exists\, s^k \in \partial f(x^k) : P_{\gamma_k,0,1} s^k \to s\}.$$

Observa que como consequência imediata temos que

$$\partial f(x) \subseteq \partial^{Lim} f(x), \qquad \forall\, x \in M \tag{2.59}$$

De fato, seja $g \in \partial f(x)$. Tomando $\{x^k\} := \{x\}$ e $\{g^k\} := \{g\}$ com $g^k \in \partial f(x^k)$, segue que g^k converge para g. Portanto, $g \in \partial^{Lim} f(x)$.

2.4 Comentários.

A utilização das ferramentas e técnicas da geometria Riemanniana no campo da otimização, é muito interessante e atraente, cada vez mais, na comunidade de otimização. Uma das vantagens bem conhecidas é que alguns problemas de otimização não convexa podem ser expressos como um problema de otimização convexa assim que dotamos alguma métrica Riemanniana apropriada naquele espaço. Por exemplo, consideremos o problema de minimizar a função $f : \mathbb{R}^2_{++} \to \mathbb{R}$ tal que $f(x_1, \ldots, x_n) = \sum_{i=1}^m c_i \prod_{j=1}^n x_j^{b_{ij}}$, onde $c_i \in \mathbb{R}^n_{++}$ e $b_{ij} \in \mathbb{R}$ para qualquer $i, j \in \mathbb{N}$. Então, é simples verificar que a função f é convexa sobre $\mathbb{R}^n_{++}$ com respeito à métrica afim definida em relação (2.23). Por tanto, resolvermos o problema de otimização não convexa sobre $\mathbb{R}^n$ com a métrica Euclideana é equivalente resolver o problema de minimização irrestrita sobre o ortante positivo com respeito à métrica afim. Para mais detalhes, ver Rapcsák [49].

Além disso, sobre as variedades em dimensões infinitas, por exemplo aquelas modeladas em espaços de Hilbert, o Teorema de Hop-Rinow não é mais válido. Isso implica que não podemos garantir a existencia de geodésica. Não obstante, se a variedade em questão é de Hadamard (variedades Riemannianas completas conexas e de curvatura não positiva) obtemos a existência das geodésicas como consequência da curvatura não positiva. Nesse contexto, Ekeland[28] provou, utilizando o célebre princípio variacional (ver [29]), que em dimensões infinitas, o conjunto de pontos podem ser ligados por uma geodésica minimal em M é denso.

Por outro lado, nesse capítulo também estudamos um tipo de subdiferencial abstrato. Esta classe abstracta de subdiferenciais foram considerados por Ioffe[33] num outro contexto. Ver também [20, 56]. A definição apresentada por Aussel [6] é mais geral porque tem a propriedade aditiva menos restritiva. Como foi visto em Aussel et al.[6], os subdiferenciais de Clarke-Rockafellar e Dinni são são os subconjuntos maiores dentre todos os subdiferenciais clássicos. Nesse capítulo podemos conferir que essa mesma ideia pode ser extendido a espaços topológicos mais gerais como as variedades de Hadamard.

Capítulo 3

Versões inexatas do método do ponto proximal

Neste capítulo temos interesse em resolver problemas de otimização em variedades de Hadamard definidos como:

$$\min_{x \in M} f(x) \tag{3.1}$$

onde M uma variedade de Hadamard (variedade Riemanniana simplesmente conexa e completa com curvatura seccional não positiva) e $f : M \to I\!R \cup \{+\infty\}$ é uma função pópria, real valorada, limitada inferiormente, semicontínua inferior e quase-convexa.

O algoritmo do ponto proximal para resolver problemas de minimização convexa, proposto por Ferreira e Oliveira[32] no contexto Riemanniano, consiste em resolver a seguinte inclusão

$$\frac{1}{\lambda_k} \exp_x^{-1} x^{k-1} \in \partial f(x), \qquad x \in M \tag{3.2}$$

Logo, por questões computacionais e como geralmente é impossível calcular um valor exato para $x \in M$ dado por relação (3.2), resolver (3.2) de forma aproximada se torna viável. Então, a relação (3.2) será substituido pela seguinte inclusão:

$$\epsilon^k \in \lambda_k \partial f(x) - \exp_x^{-1} x^{k-1}; \tag{3.3}$$

onde $\epsilon^k \in T_x M$ representa o vetor de erro ao procurar uma estimativa de $x \in M$. Daí, é natural considerar $x^{k+1} := x$.

Por outro lado, o algoritmo do ponto proximal para o caso convexo, como muitos outros algoritmos de minimização para resolver (3.1), tem a propriedade de gerar uma sequência minimizante, isto é, uma sequência $\{x^k\}$ tal que $\{f(x^k)\}$ seja decrescente em cada iteração. Observe que o problema dado por (3.1) é quase-convexo, então uma forma de ainda garantir tal propriedade mencionada está focado num resultado para funções quase-convexas:

$$\text{Se existir } g^k \in \partial f(x^k) \; : \; \langle g^k, \exp^{-1}_{x^k} x^{k-1} \rangle > 0 \implies f(x^k) \leq f(x^{k-1}),$$

ver Lema 2.8. Então, de (3.3) e supondo que $\langle \epsilon^k + \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x^{k-1} \rangle > 0$, isso implicaria que

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1} \rangle > -d(x^{k-1}, x^k).$$

Portanto, uma condição suficiente para obter $f(x^k) \leq f(x^{k-1})$ é dado por

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1} \rangle \geq 0;$$

ou equivalentemente,

$$0 \leq \|\epsilon^k\| d(x^k, x^{k-1}) \cos \alpha,$$

onde α é o ângulo entre os vetores ϵ^k e $\exp^{-1}_{x^k} x^{k-1}$. Então, x^k é melhor que x^{k-1}, no sentido que cumpre a propriedade de descida da função, sempre que α seja um ângulo agudo.

Consideremos o triângulo geodésico $[x^{k-1}, x^k, \exp_{x^k} \epsilon^k]$. É sabido que o ângulo α será agudo se os lados opostos não é mais longo do que os outros lados do triângulo, ou equivalentemente, é suficiente ter o seguinte critério:

$$d(\exp_{x^k} \epsilon^k, x^{k-1}) \leq \max\{\|\epsilon^k\|, d(x^k, x^{k-1})\}. \tag{3.4}$$

Logo, considerando as equações (3.3) e (3.4), apresentaremos uma versão inexata do algoritmo do ponto proximal em variedades de Hadamard para resolver o problema (3.1), denotado como algoritmo HMIP:

Algoritmo HMIP .

Passo inicial: Pega $x^0 \in M$. Estabelecer $k = 0$.

Passo iterativo: Dado um $x^{k-1} \in M$, achar $x^k \in M$ e $\epsilon^k \in T_{x^k} M$ satisfazendo

$$\epsilon^k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1}, \tag{3.5}$$

onde ∂ é o subdiferencial abstrato, tal que

$$d(\exp_{x^k} \epsilon^k, x^{k-1}) \leq \max \left\{ ||\epsilon^k||, d(x^k, x^{k-1}) \right\}. \tag{3.6}$$

Criterio de parada: Se $x^{k-1} = x^k$ ou $0 \in \partial f(x^k)$, parar. Caso contrario, $k \leftarrow k-1$ e ir para o *Passo iterativo*.

Observação 3.1 *Seja $g^k \in \partial f(x^k)$, de relação (3.5) tem-se*

$$\lambda_k g^k = \exp^{-1}_{x^k} x^{k-1} + \epsilon^k. \tag{3.7}$$

Observação 3.2 *Algumas considerações podem ser extraídas deste algoritmo:*

1. *Se $x^k = x^{k+1}$, segue que $\epsilon^k \in \lambda_k \partial f(x^k)$, o qual implica que x^k é uma solução aproximada de um ponto crítico sempre que $\frac{\varepsilon_k}{\lambda_k}$ seja suficientemente pequeno.*

2. *Se $M = I\!R^n$, então* (3.5) *e* (3.6) *ficam simplificados, respectivamente, para*

$$\epsilon^k \in \lambda_k \partial f(x^k) + x^k - x^{k-1},$$

e

$$\|x^{k-1} - x^k - \epsilon^k\| \leq \max\{\|\epsilon^k\|, \|x^k - x^{k-1}\|\}$$

Então, o algoritmo HMIP estende o algoritmo do ponto proximal inexato proposto por Papa Quiroz et al. [43] dos espaços $I\!R^n$ para variedades de Hadamard.

3. *Se $\epsilon^k = 0$ em* (3.5)*, para cada $k \geq 0$, obtemos*

$$\frac{1}{\lambda_k} \exp^{-1}_{x^k} x^{k-1} \in \partial f(x^k).$$

Portanto, se f é convexo e $\varepsilon_k = 0$, então o algoritmo se reduz para o algoritmo do ponto proximal em variedades de Hadamard estudado por Ferreira e Oliveira [32].

Teorema 3.1 *Seja M uma variedade de Hadamard. Se $f : M \to I\!R \cup \{+\infty\}$ é uma função própria, limitado inferiomente e semicontínua inferior. Então, a sequência $\{x^k\}$, gerada pelo método do ponto proximal, está bem definida.*

Prova. Definimos $\phi : M \to I\!R \cup \{+\infty\}$ como

$$\phi(z) = f(z) + \frac{1}{2\lambda_k} d^2(z, x^{k-1}),$$

cujo conjunto de nível inferior correspondente é dado por

$$\Gamma^{\phi}_{\alpha} := \{x \mid \phi(z) \leq \alpha, \quad \text{para algum } \alpha \in I\!R\}.$$

Observe que desde que $f(\cdot)$ e $d^2(\cdot, x^{k-1})$ são funções semicontínuas inferiores e limitadas inferiormente, isto implica que $\phi(\cdot)$ é único. Procedemos por indução. A existência de x^0 é válido para $k = 0$ pelo passo inicial do algoritmo. Assuma que

x^k existe. Logo, provamos que x^{k+1} e ε^k existem e satisfazem (3.5) e (3.6). Para atingir o objetivo, dividimos a prova em 3 passos:

Passo 1. Compacidade do conjunto de nível Γ_α^ϕ. Logo, $\phi(z^k) = \infty$ sempre que $\{z^k\} \subset M$ e $\lim\limits_{k\to+\infty} \|z^k\| = \infty$. Segundo Udriste [58, Teorema 8.6], Γ_α^ϕ é limitada. Além disso, desde que $\phi(z)$ é semicontínua inferior, obtemos o fecho desse conjunto. Portanto, Γ_α^ϕ é um conjunto compacto.

Passo 2. O conjunto $\phi(z)$ atinge o seu valor mínimo. De fato, como $\phi(\cdot)$ é limitada inferiormente, obtemos que $a = \inf \phi(z)$ com $a \in I\!R$ para todo $z \in \Gamma_\alpha^\phi$. Isto implica que

$$a \leq \phi(\tilde{z}), \tag{3.8}$$

para algum $\tilde{z} \in \Gamma_\alpha^\phi$.

Por outro lado, consideremos $\{z^l\} \subset \Gamma_\alpha^\phi$ tal que $f(z^l)$ converge para a. Desde que Γ_α^ϕ é um conjunto compacto, existe uma subsequência $\{z^{l_k}\} \subset \{z^l\}$ tal que $\{z^{l_k}\}$ converge para $\tilde{z}$, com $\tilde{z} \subset \Gamma_\alpha^\phi$. Logo, desde que $\phi(\cdot)$ é semicontínua inferior, isto segue que

$$\phi(\tilde{z}) \leq \liminf_{k\to+\infty} \phi(z^{l_k}) = \lim_{l\to+\infty} \phi(z^l) = a. \tag{3.9}$$

Combinando as equações (3.8) e (3.9), concluimos que $\phi(\cdot)$ atinge o seu valor mínimo, como queriamos provar.

Passo 3. Existência de iterações satisfazendo algoritmo HMIP. Em vista de $\tilde{z} \in M$ ser um valor mínimo, temos

$$\begin{aligned} 0 &\in \partial\left(f(\cdot) + \frac{1}{2\lambda_k} d^2(\cdot, x^{k-1})\right)(\tilde{z}). \\ &\subset \lambda_k \partial f(\tilde{z}) - \exp_{\tilde{z}}^{-1} x^{k-1}, \end{aligned} \tag{3.10}$$

onde a última formulação é obtida do fato que $d^2(\cdot, x^{k-1})$ é uma função contínua e convexa, o qual é também ∂-diferenciável em $\tilde{z} \in M$. De fato, pelo Teorema 2.4 e desde que $\exp_{\tilde{z}} x^{k-1}$ seja uma aplicação bem definida, é fácil observar que $d^2(\cdot, x^{k-1})$ é ∂-diferenciável. Portanto, tal ponto minimizador $\tilde{z} \in M$ satisfaz (3.5) e pode ser escolhido como x^k. Segundo (3.10) e considerando $\epsilon^k = 0$, isto satisfaz equação (3.6). Além disso, concluimos a existência de iterações $x^k \in M$ e $\epsilon^k \in T_{x^k}M$ satisfazendo equações (3.5) e (3.6). ■

3.1 Algumas propriedades do algoritmo HMIP

Os dois resultados seguintes são usados para mostrar que o algoritmo HMIP é um método de descida, o que é uma propriedade bastante comun em vários algoritmos de minimização.

Lema 3.1 *Seja M uma variedade de Hadamard. Suponha que se cumpre a equação* (3.6). *Então,*

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1} \rangle \geq 0, \qquad \forall\, k \in I\!N$$

Prova. Seja o triângulo geodésico $[\exp_{x^k}\epsilon^k, x^{k-1}, x^k]$. Do Teorema 2.3, temos

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

Como $d(\exp_{x^k}\epsilon^k, x^k) = \|\exp^{-1}_{x^k}\exp_{x^k}\epsilon^k\| = \|\epsilon^k\|$, obtemos

$$\|\epsilon^k\|^2 + d^2(x^k, x^{k-1}) - d^2(\exp_{x^k}\epsilon^k, x^{k-1}) \leq 2\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1}\rangle.$$

Isso implica que

$$\max\{\|\epsilon^k\|^2, d^2(x^k, x^{k-1})\} - d^2(\exp_{x^k}\epsilon^k, x^{k-1}) \leq 2\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1}\rangle.$$

Logo, da relação (3.6), temos que o lado esquerdo da última desigualdade é não negativo. Portanto, obtemos o resultado. ■

Consideramos a seguinte hipótese sobre o subdiferencial, definido na Seção 2.3 desta tese, para estudar o algoritmo HMIP:

(H2) $\qquad \big(\partial \subset \partial^{D+}\big)$ ou $\big(\ \partial \subset \partial^{CR}$ e f é contínua em $M.\big)$

A seguir, o seguinte resultado estabelece a propriedade de descida do algoritmo HMIP.

Proposição 3.1 *Suponha que as hipóteses* (H1) *e* (H2) *são satisfeitas e seja $\{x^k\} \subset M$ uma sequência gerada pelo algoritmo HMIP, então*

a. Se $\partial \subset \partial^{CR}$ e f é contínua em M então $f(x^k) \leq f(x^{k-1})$

b. Se $\partial \subset \partial^{D^+}$ então $f(x^k) < f(x^{k-1})$.

Portanto, temos que $\{f(x^k)\}$ é não crescente (decrescente no caso b) e convergente.

Prova. Seja $g^k \in \partial f(x^k)$. De (3.5) e Teorema 2.4 tem-se que

$$\begin{aligned} \epsilon^k &= \lambda_k g^k - \exp^{-1}_{x^k} x^{k-1} && (3.11)\\ &= \lambda_k g^k + \operatorname{grad}\frac{1}{2}d^2(\cdot, x^{k-1})(x^k). && (3.12) \end{aligned}$$

Por outro lado, como consequência da Proposição 2.1, $d^2(\cdot, x^{k-1})$ é fortemente convexa e consequentemente $\operatorname{grad}\frac{d^2}{2}(\cdot, x^{k-1})$ é fortemente monótona e do fato que

$x^k \neq x^{k-1}$, temos

$$\langle \mathcal{P}_{x^k,x^{k-1}} \mathrm{grad}\frac{d^2}{2}(\cdot, x^{k-1})(x^k) - \mathrm{grad}\frac{d^2}{2}(\cdot, x^{k-1})(x^{k-1}), \exp^{-1}_{x^{k-1}} x^k\rangle > 0, \tag{3.13}$$

onde $\mathcal{P}_{x,y}$ é o transporte paralelo de x para y ao longo da (mínima) geodésica.
Do fato que $\mathrm{grad}\frac{d^2}{2}(\cdot, y)(y) = 0$, $\mathcal{P}_{x,y}\exp^{-1}_x y = -\exp^{-1}_y x$ e utilizando a propriedade isométrica do transporte paralelo $\mathcal{P}_{x,y}$, (3.13) se torna

$$\langle \mathrm{grad}\frac{d^2}{2}(\cdot, x^{k-1})(x^k), \exp^{-1}_{x^k} x^{k-1}\rangle < 0. \tag{3.14}$$

Combinando (3.14) e (3.12), obtemos

$$\langle \epsilon^k - \lambda_k g^k, \exp^{-1}_{x^k} x^{k-1}\rangle < 0,$$

e consequentemente,

$$\langle g^k, \exp^{-1}_{x^k} x^{k-1}\rangle > \frac{1}{\lambda_k}\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1}\rangle.$$

Logo, do Lemma 3.1 temos

$$\langle g^k, \exp^{-1}_{x^k} x^{k-1}\rangle > 0.$$

Portanto, o resultado segue do Lema 2.8. ■

Definimos

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

Note que se $U = \emptyset$, então para todo $x \in M$ segue que $f(x) > \inf_k f(x^k) \geq \inf_{x\in M} f(x)$, e consequentemente, isto fica claro que $\lim_{k\to\infty} f(x^k) = \inf_{x\in \mathbb{N}} f(x^k) = \inf_{x\in M} f(x)$.

No decorrer desta tese, supomos que U é não vazio.

3.2 Algumas variantes do algoritmo do ponto proximal

Nesta seção apresentamos e analisamos dois variantes do algoitmo HMIP baseado em dois critérios de erro, o qual garante que, com certas condições sobre a estimação do erro, os resultados de convergência dos dois algoritmos estão garantidos.

3.2.1 Algoritmo HMIP1 e resultados de convergência

Nessa subseção estabeleceremos um algoritmo para resolver o problema (3.1), denotado por HMIP1, o qual consiste do algoritmo HMIP (estudado nesse capítulo) com um critério de erro classicamente estudado e motivado por Rockafellar [50].

Algoritmo HMIP1.
Considere o *Passo inicial* do algoritmo HMIP. Seja $\{x^k\} \subset M$ e $\{\epsilon^k\} \subset T_{x^k}M$ serem sequências geradas por (3.5)-(3.6) tal que

(B1) $\sum_{k=1}^{+\infty} \|\epsilon^k\| < +\infty$.

Observação 3.3 *Na hipótese* (B1)*, é importante observar que há um certo grau de liberdade em escolher o vetor de erro* $\epsilon^k \in T_{x^k}M$*. De fato, essas hipóteses foram também consideradas naturalmente no algoritmo do ponto proximal clássico, ver Rockafellar* [50]*.*

O seguinte resultado é usado para mostrar que as sequências $\{x^k\}$, geradas pelo algoritmo HMIP1, é limitada.

Proposição 3.2 *Se todas as hipóteses do problema* (H1) *e* (H2) *são satisfeitas então para todo* $k \in N$*, segue que*

$$d^2(x, x^k) \leq d^2(x, x^{k-1}) - d^2(x^{k-1}, x^k) - 2\langle \epsilon^k, \exp^{-1}_{x^k} x \rangle, \quad \forall\, x \in U.$$

Prova. Consideremos dois casos:

i. Se $\partial f \subset \partial^{CR} f$. Seja $x \in U$. Desde que $U \subset U_k := \{x \in M : f(x) \leq f(x^k)\}$ e $int(U_k) = \{x \in M : f(x) < f(x^k)\}$ é não vazio (se esse conjunto for vazio então para todo $x \in M$, $f(x) \geq f(x^k)$. Portanto, x^k é um ponto mínimo de f e $0 \in \partial f(x^k)$, então o algoritmo HMIP para) existe $\{x^l\} \subset M$ com $f(x^l) < f(x^k)$ tal que x^l converge para x. Segue do Lema 2.8 e (3.11) que

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^l \rangle \leq -\langle \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x^l \rangle \tag{3.15}$$

Tomando $z = x^k, x = x^{k-1}$ e $y = x^l$ no Teorema 2.3, obtemos

$$-2\langle \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x^l \rangle \leq d^2(x^{k-1}, x^l) - d^2(x^{k-1}, x^k) - d^2(x^k, x^l). \tag{3.16}$$

Combinando as relações (3.15) e (3.16), obtemos

$$2\langle \epsilon^k, \exp^{-1}_{x^k} x^l \rangle \leq d^2(x^{k-1}, x^l) - d^2(x^{k-1}, x^k) - d^2(x^k, x^l).$$

Portanto, o resultado segue tomando o limite quando $l \to +\infty$.

ii. Se $\partial f \subset \partial^{D^+} f$. Seja $x \in U$ e $g^k \in T_{x^k}M$. Segue de Lema 4.2 e relação (4.12) que

$$\langle \epsilon^k, \exp_{x^k}^{-1} x\rangle \leq -\langle \exp_{x^k}^{-1} x^{k-1}, \exp_{x^k}^{-1} x\rangle \tag{3.17}$$

Tomando $z = x^k, x = x^{k-1}$ e $y = x$ no Teorema 2.3, obtemos

$$-2\langle \exp_{x^k}^{-1} x^{k-1}, \exp_{x^k}^{-1} x\rangle \leq d^2(x^{k-1}, x) - d^2(x^{k-1}, x^k) - d^2(x^k, x). \tag{3.18}$$

Combinando as relações (3.17) e (3.18), obtemos

$$2\langle \epsilon^k, \exp_{x^k}^{-1} x\rangle \leq d^2(x^{k-1}, x) - d^2(x^{k-1}, x^k) - d^2(x^k, x). \quad \blacksquare$$

O seguinte resultado será importante para obter a convergência de $\{d(x, x^k)\}$ para todo $x \in U$.

Lema 3.2 *Sejam* $\{v_k\}$,$\{\gamma_k\}$ *e* $\{\beta_k\}$ *sequências não negativas de números reais satisfazendo* $v_{k+1} \leq (1 + \gamma_k)\, v_k + \beta_k$ *e tal que* $\sum_{k=1}^{\infty} \beta_k < \infty$, $\sum_{k=1}^{\infty} \gamma_k < \infty$. *Então , a sequência* $\{v_k\}$ *converge.*

Proof. Ver Polyak [48], Lema 2, pp. 44. $\blacksquare$

A análise de convergência do algoritmo HMIP1 será muito simplificada pelo fato dos seguintes resultados:

Lema 3.3 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo* HMIP1. *Suponha que hipóteses* $(H1), (H2)$ *e* $(B1)$ *são satisfeitas. Então, se cumpre que*

(i) *para todo* $x \in U$, *a sequência* $\{d(x^k, x)\}$ *é convergente;*

(ii) *a sequência* $\{x^k\}$ *é limitada;*

(iii) $\lim_{k\to\infty} d(x^k, x^{k-1}) = 0$*;*

(iv) *se* $\lim_{j\to+\infty} x^{k_j} = \bar{x}$ *então* $\lim_{k\to\infty} x^{k_j-1} = \bar{x}$.

Prova. (i) De Proposição 3.2 e usando $d(x, x^k) = \|\exp_{x^k}^{-1} x\|$, obtemos

$$d^2(x, x^k) \leq d^2(x, x^{k-1}) - d^2(x^{k-1}, x^k) + 2\|\epsilon^k\| d(x, x^k), \tag{3.19}$$

para todo $x \in U$. Da última desigualdade $(a-1/2)^2 \geq 0$, para todo $a \geq 0$, obtemos que $d(x,x^k) \leq d^2(x,x^k) + 1/4$ e assim a desigualdade acima implica que

$$d^2(x,x^k) \leq d^2(x,x^{k-1}) - d^2(x^{k-1},x^k) + 2\|\epsilon^k\| \left(d^2(x,x^k) + \frac{1}{4} \right), \quad \forall\, x \in U.$$

Portanto, obtemos

$$d^2(x,x^k) \leq \left(\frac{1}{1-2\|\epsilon^k\|} \right) d^2(x,x^{k-1}) + \frac{1}{2} \left(\frac{\|\epsilon^k\|}{1-2\|\epsilon^k\|} \right), \quad \forall k \geq k_0$$

Como $\lim_{k\to+\infty} \|\epsilon^k\| = 0$, então existe $k_0 \in I\!N$ tal que $\|\epsilon^k\| \leq 1/4$, para todo $k \geq k_0$. Isso implica que $\frac{1}{1-2\|\epsilon^k\|} \leq 1 + 4\|\epsilon^k\|$ e $\frac{1}{1-2\|\epsilon^k\|} \leq 2$, para todo $k \geq k_0$. Usando esses resultados na desigualdade acima obtemos

$$d^2(x,x^k) \leq \left(1 + 4\|\epsilon^k\|\right) d^2(x,x^{k-1}) + \|\epsilon^k\|,$$

para todo $k \geq k_0$. Finalmente, usando a hipótese (B1) e Lema 3.2 obtemos que $\{d(x,x^k)\}$ converge.

(ii) Como $\{d(x,x^k)\}$ converge, obtemos imediatamente que $\{x^k\}$ é limitado.

(iii) De (3.19) e do item (i) temos que $\lim_{k\to\infty} d(x^k,x^{k-1}) = 0$.

(iv) Da desigualdade triangular e do item (iii) obtemos o resultado. ∎

A seguir, vamos provar dois resultados principais sobre a teoria de convergência do algoritmo HMIP1.

Teorema 3.2 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo* HMIP1. *Considere as hipóteses* (H1), (H2) *e* (B1). *Então, a sequência gerada pelo algoritmo converge para um ponto de* U.

Prova. Esta prova é bastante similar à prova do Teorema 4.3 de [44]. ∎

Teorema 3.3 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo* HMIP1. *Considere as hipóteses* (H1), (H2) *e* (B1) *com* $\tilde{\lambda}$ *tal que* $\tilde{\lambda} < \lambda_k$ *e* f *seja uma função contínua no* $dom(f)$. *Então,* $\{x^k\}$ *converge para algum* $\bar{x}$ *com* $0 \in \partial^{Lim} f(\bar{x})$.

Prova. Como uma consequência direta do Teorema 3.2, existe um ponto $\bar{x} \in U$ tal que $\lim_{k\to+\infty} x^k = \bar{x}$. Desde que f é contínua, temos que $\lim_{k\to+\infty} f(x^k) = f(\bar{x})$. De (3.11), seja $g^k \in \partial f(x^k)$ tal que

$$g^k = \frac{1}{\lambda_k}(\epsilon^k + \exp^{-1}_{x^k} x^{k-1}).$$

Logo, só resta provar que $\mathcal{P}_{\psi_k,0,1} g^k$ converge para zero, para a geodésica ψ_k ligando os pontos x^k e $\bar{x}$. De fato, pela hipótese (B1), a continuidade de $\exp^{-1}_{x^k}(\cdot)$ e desde que $\mathcal{P}_{\psi_k,0,1}$ é um operador linear, consequentemente contínuo, temos que

$$\lim_{k\to\infty} \mathcal{P}_{\psi_k,0,1}(\epsilon^k + \exp_{x^k}^{-1} x^{k-1}) = \lim_{k\to\infty} \mathcal{P}_{\psi_k,0,1}\epsilon^k + \lim_{k\to\infty} \mathcal{P}_{\psi_k,0,1}\exp_{x^k}^{-1} x^{k-1} = 0 + 0,$$

Por conseguinte, o resultado segue do fato que $1/\lambda_k$ é limitado. ∎

Teorema 3.4 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo* HMIP1. *Se as hipóteses* (H1),(H2), (B1) *e* (B2) *são satisfeitas com* $\tilde{\lambda}$ *tal que* $\tilde{\lambda} < \lambda_k$ *e* f *é localmente Lipschitz, então* $\{x^k\}$ *converge para algum* $\bar{x}$ *com* $0 \in \partial^o f(\bar{x})$.

Prova. De (3.11), $(1/\lambda_k)(\epsilon_k + \exp_{x^k} x^{k-1}) \in \partial^\circ f(x^k)$, então

$$f^\circ(x^k, v) \geq (1/\lambda_k)\langle \epsilon^k + \exp_{x^k} x^{k-1}, v\rangle, \ \forall v \in T_{x^k}M.$$

Seja $\bar{x}$ um ponto limite da sequência $\{x^k\}$ e $\bar{v} \in T_{\bar{x}}M$, então

$$f^\circ(x^k, v^k) \geq (1/\lambda_k)\langle \epsilon^k + \exp_{x^k} x^{k-1}, v^k\rangle,$$

onde $v^k = D(\exp_{\bar{x}})_{\exp_{\bar{x}}^{-1} x^k}\bar{v}$. Tomando o limite superior na desigualdade acima e usando Proposição 4.1 de [11], obtemos que

$$f^\circ(\bar{x}, \bar{v}) \geq \limsup_{k\to+\infty} f^\circ(x^k, v^k) \geq 0$$

Portanto, segue que $0 \in \partial^\circ f(\bar{x})$. ∎

3.2.2 Algoritmo HMIP2 e resultados de convergência

Nesta subseção estabeleceremos uma classe do algoritmo HMIP para resolver (3.1). Além disso, vale a pena mencionar que resultados de convergência do seguinte algoritmo não precisam da teoria Quasi-Fejér convergente como foi para o algoritmo HMIP1 na Subseção 3.2.1.

Algoritmo HMIP2

Considere o passo inicial do algoritmo HMIP. Sejam $\{x^k\} \subset M$ e $\{\epsilon^k\} \subset T_{x^k}M$ sequências geradas pelas relações (3.5)-(3.6) tal que

(C1) $\|\epsilon^k\| \leq \eta_k d(x^k, x^{k-1})$.

(C2) $\sum_{k=1}^{+\infty} \eta_k^2 < +\infty$.

Lema 3.4 *Se todas as hipóteses* (H1),(H2), (C1) *e* (C2) *do problema são satisfeitas, então existe um inteiro $k_0 \geq 0$ tal que para todo $k \geq k_0$ temos*

$$d^2(x^k, x) \leq \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right) d^2(x^{k-1}, x) - \frac{1}{2} d^2(x^k, x^{k-1}), \quad \forall \in x \in U \tag{3.20}$$

Além disso, $\{x^k\}$ é uma sequência limitada e $\lim_{k \to +\infty} d(x^k, x^{k-1}) = 0$.

Prova. A prova é obtida usando argumentos similares ao Tang e Huang [55, Lema 3.2] De fato, para $\eta_k > 0$ e como η_k converge para zero, existe $k_0 \geq 0$ tal que para todo $k \geq k_0$, $1 - 2\eta_k^2 > 0$. Logo, expressão (3.20) pode ser fácilmente obtida.

Para provar que $\{x^k\}$ é uma sequência limitada, podemos usar a desigualdade dada na relação (3.20). Portanto,

$$d^2(x^k, x) \leq \prod_{i=k_0}^{k} \left(1 + \frac{2\eta_i^2}{1 - 2\eta_i^2}\right) d^2(x^{k_0}, x). \tag{3.21}$$

Como $\{\eta_k^2\}$ é uma sequência convergente, para todo $\varepsilon > 0$ existe $\tilde{k}_0$ tal que $\eta_k^2 < \varepsilon$, para $k \geq \tilde{k}_0$. Desta forma, para todo $\tilde{k}_0 > k_0$ obtemos

$$0 < 1 - 2\varepsilon < 1 - 2\eta_k^2 < 1.$$

Portanto, para todo $\varepsilon < 1/2$ temos

$$\frac{2\eta_k^2}{1 - 2\eta_k^2} < \frac{2\eta_k^2}{1 - 2\varepsilon}, \qquad \forall\ k > k_0. \tag{3.22}$$

Somando todos os términos na desigualdade (3.22) e considerando (C2),

$$\sum_{k=1}^{+\infty} \frac{2\eta_k^2}{1 - 2\eta_k^2} < +\infty. \tag{3.23}$$

Por outro lado, das propriedades da desigualdade aritmética e geométrica, obtemos que

$$\begin{aligned} \prod_{k=k_0}^{n} \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right) &\leq \left(\frac{1}{n - k_0 + 1} \sum_{k=k_0}^{n} \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right)\right)^{n-k_0+1} \\ &= \left(1 + \frac{1}{n - k_0 + 1} \sum_{k=k_0}^{n} \frac{2\eta_k^2}{1 - 2\eta_k^2}\right)^{n-k_0+1}. \end{aligned}$$

Tomando limite, quando $n \to +\infty$, temos

$$\lim_{n\to+\infty}\prod_{k=k_0}^{n}\left(1+\frac{2\eta_k^2}{1-2\eta_k^2}\right)\leq e^{\sum_{k=k_0}^{n}\frac{2\eta_k^2}{1-2\eta_k^2}},$$

o qual, combinado com (3.23), dá $\prod_{k=k_0}^{+\infty}\left(1+\frac{2\eta_k^2}{1-2\eta_k^2}\right) < +\infty$. Portanto, Por (3.21) concluimos que $\{x^k\}$ é uma sequência limitada.

Para finalizar a prova, mostraremos o último resultado. Somando os términos de (3.20), temos que para todo $k \geq k_0$

$$\begin{aligned}\frac{1}{2}\sum_{k=k_0}^{n} d^2(x^{k+1},x^k) &\leq \sum_{k=k_0}^{n}(d^2(x^k,x)-d^2(x^{k+1},x))+\sum_{k=k_0}^{n}\frac{2\eta_k^2}{1-2\eta_k^2}d^2(x^k,x)\\ &\leq d^2(x^{k_0},x)-d^2(x^n,x)+\max_{n\geq k\geq k_0}\{d^2(x^k,x)\}\sum_{k=k_0}^{n}\frac{2\eta_k^2}{1-2\eta_k^2}.\end{aligned}$$

Como $\{x^k\}$ é uma sequência limitada, então temos que $d(x^k,x)$ também é limitada para todo $k \geq 0$. Portanto, tomando limite quando $n \to +\infty$ e considerando a relação (3.23), podemos concluir que $d(x^{k+1},x^k)$ converge para zero. ■

Teorema 3.5 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo* HMIP2. *Se as hipóteses* (H1),(H2), (C1) *e* (C2) *são satisfeitas com* $\tilde{\lambda} > 0$ *tal que* $\tilde{\lambda} < \lambda_k$ *e* f *seja uma função contínua, então a sequência* $\{x^k\}$ *converge para algum* $\bar{x}$ *com* $0 \in \partial^{Lim} f(\bar{x})$.

Prova. O resultado se obtem de forma análoga à prova do Teorema 3.3 combinado com [55, Teorema 3.1] ■

Teorema 3.6 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo* HMIP2. *Se as hipóteses* (H1),(H2), (C1) *e* (C2) *são satisfeitas com* $\tilde{\lambda} > 0$ *tal que* $\tilde{\lambda} < \lambda_k$ *e* f *é uma função localmente Lipschitz, então* $\{x^k\}$ *converge para algum* $\bar{x} \in M$ *com* $0 \in \partial^o f(\bar{x})$.

Prova. Análogo à prova do Teorema 3.4. ■

3.3 Comentários.

Neste capítulo foi introduzido duas versões inexatas do algoritmo de ponto proximal para resolver o problema dado em (3.1) quando a função objetivo é quase-convexa e considerando um subdiferencial abstrato mais geral, o qual inclui os clássicos subdiferenciais de Clarke-Rockafellar e de Fréchet. Vale ressaltar que a segunda versão inexata do algoritmo apresentado nesse capítulo, denotado por HMIP2, foi estudado por Tang e Huang[55] para campos vetoriais monótonos em variedades de Hadamard. Portanto, o algoritmo apresentado nesta abordagem é uma extensão para resolver problemas de otimização quase-convexa.

Finalmente, baseado no artigo de Li e Yao[36] para o caso convexo, uma questao interessante seria saber se será possível estender o método proximal para o caso quase-convexo em variedades Riemannianas arbitrárias.

Capítulo 4

Taxa de convergência do algoritmo do ponto proximal

Como foi visto no capítulo anterior, estamos interessados no estudo do algoritmo do ponto proximal para resolver problemas de minimização, no contexto de variedades de Hadamard, cuja função objetivo $f : M \to I\!R \cup \{+\infty\}$ satisfaz as seguintes hipóteses:

(H1) f é uma função própria, limitada inferiormente, semicontínua inferior e quase-convexa.

(H2) $(\partial \subset \partial^{D+})$ ou $(\partial \subset \partial^{CR}$ e f é uma função contínua em $M)$.

Neste capítulo, nosso objetivo é analisar a taxa de convergência do algoritmo do ponto proximal para funções quase-convexas em variedades de Hadamard, o qual foi estudado no Capítulo 3 desta tese. O algoritmo é descrito a seguir: e é descrito a seguir:

Algoritmo HMIP2.

Inicialização: Seja $x^0 \in M$. Estabelece $k = 0$.
Passo iterativo: Dado $x^{k-1} \in M$, encontrar $x^k \in M$ e $\epsilon^k \in T_{x^k}M$ tal que

$$\epsilon^k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1} \tag{4.1}$$

onde

$$d(\exp_{x^k}\epsilon^k, x^{k-1}) \leq \max\left\{||\epsilon^k||, d(x^k, x^{k-1})\right\}. \tag{4.2}$$

$$\|\epsilon^k\| \leq \eta_k d(x^k, x^{k-1}) \tag{4.3}$$

$$\sum_{k=1}^{+\infty} \eta_k^2 < +\infty. \tag{4.4}$$

Critério de parada: Se $x^{k-1} = x^k$ ou $0 \in \partial f(x^k)$ parar. Caso contrário, $k \leftarrow k-1$ e ir para o *Passo iterativo*.

Com esse objetivo, vamos denotar o seguinte conjunto

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

Repare que se existirem pontos minimizadores, U é o conjunto que contém esses pontos.

A seguir, denotamos o seguinte conjunto:

$$Z := U \cap \{x \in M \mid 0 \in \partial^{Lim} f(x)\}.$$

4.1 Análise da taxa de convergência

No intuito de estudar a taxa de convergência, vamos considerar a seguinte hipótese:

(H3) Para $\bar{x} \in Z$ tal que $\lim x^k = \bar{x}$, existe $\delta := \delta(\bar{x}) > 0$ e $\tau := \tau(\bar{x}) > 0$ tais que para todo $w \in B(0, \delta) \subset T_{\bar{x}}M$ e todo x^k tal que $P_{\psi,1,0}(w) \in \partial^{Lim} f(x^k)$ com curva geodésica ψ_k ligando $\psi_k(0) = x$ e $\psi_k(1) = \bar{x}^k$, temos

$$d(x^k, \bar{x}) \leq \tau \|w\|_{T_{\bar{x}}M}.$$

Lema 4.1 *Sejam $\{x^k\} \subset M$ e $\{\epsilon^k\} \subset T_{x^k}M$ sequências geradas pelo algoritmo HMIP2 (estudado na Subseção 3.2.2). Suponha que elas satisfaçam as hipóteses* (H1), (H2) *e* (H3). *Então,*

(i) existe $\bar{k} \in I\!N$ tal que

$$\|g^k\|_{T_{x^k}M} < \delta, \tag{4.5}$$

para todo $k \geq \bar{k}$, onde g^k é dado por (3.7);

(ii) se cumpre que

$$d(x^k, \bar{x}) \leq \tau \frac{(\eta_k + 1)}{\lambda_k} d(x^k, x^{k-1}), \tag{4.6}$$

para todo $k \geq \bar{k}$.

Prova. (i) Seja $\bar{x} = \lim_{k\to\infty} x^k$ e $g^k \in \partial f(x^k)$ dado pela relação (3.7). Logo, de (4.3) e uma vez que $\tilde{\lambda} < \lambda_k$ temos

$$\begin{aligned}
\|g^k\|_{T_{x^k}M} &= \frac{1}{\lambda_k}\|\exp^{-1}_{x^k} x^{k-1} + \epsilon^k\|_{T_{x^k}M} \\
&\leq \frac{1}{\lambda_k}\left(\|\exp^{-1}_{x^k} x^{k-1}\|_{T_{x^k}M} + \|\epsilon^k\|_{T_{x^k}M}\right) \\
&\leq \left(\frac{\eta_k + 1}{\lambda_k}\right) d(x^k, x^{k-1}), \qquad (4.7)\\
&\leq \left(\frac{\eta_k + 1}{\tilde{\lambda}}\right) d(x^k, x^{k-1}), \qquad k \geq 1. \qquad (4.8)
\end{aligned}$$

Além disso, como $\eta_k \to 0$, $d(x^k, x^{k-1}) \to 0$ (ver Lema 3.4) e fazendo $\varepsilon = \delta$ temos que existe $\bar{k} \in I\!N$ tal que $\|g^k\|_{T_{x^k}M} < \delta$ para todo $k \geq \bar{k}$.
Agora provaremos o item (ii). Seja $w = P_{\psi_k,0,1} g^k$ na hipótese (H3). Devido à isometria do transporte paralelo $P_{\psi_k,0,1}$, para todo $k \geq \bar{k}$, temos

$$\begin{aligned}
d(x^k, \bar{x}) &\leq \tau \|w\|_{T_{\bar{x}}M} \\
&= \tau \|P_{\psi_k,0,1} g^k\|_{T_{\bar{x}}M} \\
&= \tau \|g^k\|_{T_{x^k}M}
\end{aligned}$$

Portanto, a relação (4.6) é obtida combinando a última desigualdade e (4.7). ■

A seguir, vamos apresentar o teorema relacionado à taxa de convergência para o algoritmo HMIP2, completando assim o resultado de convergência dado pelo Teorema 3.5.

Teorema 4.1 *Sejam $\{x^k\} \subset M$ e $\{e^k\} \subset T_{x^k}M$ duas sequências geradas pelo algoritmo HMIP2. Suponha que as hipóteses* (H1), (H2) *e* (H3) *tal que $\lambda_k \in [\tilde{\lambda}, +\infty)$ com $\tilde{\lambda} > 0$ são satisfeitas. Suponha também que f é uma função contínua. Então, $\{x^k\}$ converge linearmente para $\bar{x} \in Z$. Além disso, se $\lambda_k \nearrow +\infty$, então a convergência é superlinear.*

Prova. Seja $\bar{x} \in Z$ um ponto limite da sequência $\{x^k\}$ e $g^k \in \partial f(x^k)$ dado por (3.7).
Definimos

$$w^k := P_{\psi_k,0,1} g^k,$$

onde ψ_k é a curva geodésica ligando os pontos x^k e $\bar{x}$. Devido à propriedade isométrica do transporte paralelo $P_{\psi_k,0,1}$ e a relação (4.5), temos que

$$\|w^k\|_{T_{\bar{x}}M} = \|P_{\psi_k,0,1} g^k\|_{T_{\bar{x}}M} = \|g^k\|_{T_{x^k}M} < \delta,$$

para $k \geq \bar{k}$. Isto é, $w^k \in B(0,\delta) \subset T_{\bar{x}}M$, para $k \geq \bar{k}$.

Por outro lado,

$$P_{\psi_k,1,0}w^k = P_{\psi_k,1,0}(P_{\psi_k,0,1}g^k) = g^k \in \partial f(x^k). \tag{4.9}$$

Então, de (2.59) e (4.9) temos que $P_{\psi_k,1,0}w^k \in \partial^{Lim} f(x^k)$, para todo $k \geq \bar{k}$. Além disso, aplicando (4.6) na equação (3.20), para todo $k \geq \max\{k_0, \bar{k}\}$, temos que

$$d^2(x^k,\bar{x}) \leq \left(1 + \frac{2\eta_k^2}{1-2\eta_k^2}\right) d^2(x^{k-1},\bar{x}) - \frac{1}{2}\left(\frac{\lambda_k}{\tau(\eta_k+1)}\right)^2 d^2(x^k,\bar{x}).$$

Logo,

$$\left(1 + \frac{\lambda_k^2}{2\tau^2(\eta_k+1)^2}\right) d^2(x^k,\bar{x}) \leq \frac{1}{1-2\eta_k^2} d^2(x^{k-1},\bar{x}).$$

Isto implica que

$$d^2(x^k,\bar{x}) \leq \alpha_k^2 d^2(x^{k-1},\bar{x}), \qquad k \geq \max\{k_0, \bar{k}\} \tag{4.10}$$

onde

$$\alpha_k^2 = \frac{1}{1-2\eta_k^2}\left(\frac{2\tau^2(\eta_k+1)^2}{2\tau^2(\eta_k+1)^2 + \lambda_k^2}\right). \tag{4.11}$$

Uma vez que $0 < \tilde{\lambda} < \lambda_k$, para todo $k \in I\!N$, mostra-se que

$$\alpha_k^2 \leq r_k, \tag{4.12}$$

onde

$$r_k = \frac{1}{1-2\eta_k^2}\left(\frac{2\tau^2(\eta_k+1)^2}{2\tau^2(\eta_k+1)^2 + \tilde{\lambda}^2}\right).$$

Além disso, considerando que $\{\eta_k\}$ converge para zero, temos que

$$r_k \to \frac{2\tau^2}{2\tau^2 + \tilde{\lambda}^2}.$$

Portanto, existe um número positivo $k_1 \in I\!N$ com $k \geq k_1$ tal que

$$r_k < \frac{1}{2}\left(1 + \frac{2\tau^2}{2\tau^2 + \tilde{\lambda}^2}\right), \qquad \forall\ k \geq k_1 \tag{4.13}$$

Daí, combinando as relações (4.12) e (4.13), obtemos

$$\alpha_k^2 < \frac{1}{2}\left(1 + \frac{2\tau^2}{2\tau^2 + \tilde{\lambda}^2}\right) := \theta < 1, \qquad \forall\ k \geq k_1 \tag{4.14}$$

De (4.10) e (4.14), temos que

$$d(x^k, \bar{x}) \leq \theta^{1/2} d(x^{k-1}, \bar{x}),$$

para todo $k \geq \max\{\bar{k}, k_0, k_1\}$. Portanto, a sequência $\{x^k\}$ converge linearmente para $\bar{x}$.

Por outro lado, com o objetivo de obter convergência superlinear da sequência gerada pelo algoritmo HMIP2, vamos considerar $\lambda_k \nearrow +\infty$. Uma vez que $\eta_k \to 0$, da relação (4.11) temos que $\alpha_k \to 0$, completando assim a prova. ■

Para finalizar, o seguinte teorema mostra a taxa de convergência do algoritmo HMIP2.

Teorema 4.2 *Sejam as sequências* $\{x^k\} \subset M$ *e* $\{e^k\} \subset T_{x^k}M$ *geradas pelo algoritmo HMIP2. Suponha que as hipóteses* (H1), (H2) *e* (H3)*, tal que* $\lambda_k \in [\tilde{\lambda}, +\infty)$ *com* $\tilde{\lambda} > 0$ *são satisfeitas. Suponha também que* f *é uma função localmente Lipschitz. Então a sequência* $\{x^k\}$ *converge linearmente para* $\bar{x} \in Z$*. Além disso, se* $\lambda_k \nearrow +\infty$*, então a sequência* $\{x^k\}$ *converge superlinearmente.*

Prova. Observe que a condição localmente Lipschitz implica a continuidade da função de f. Tomando $\partial = \partial^\circ$ e da relação (2.59), segue que $\partial^\circ f(\bar{x}) \subseteq \partial^{Lim} f(\bar{x})$. Portanto, os resultados da taxa de convergência segundo o resultado do Teorema 3.6 é um caso particular do Teorema 4.1 . ■

4.2 Experimentos numéricos

Para avaliar nosso algoritmo HMIP, experimentos numéricos foram realizados para resolver problemas de minimização, com funções objetivos satisfazendo as hipóteses (H1) e (H2) sobre algumas variedades de Hadamard estudadas na Seção 2.2. Realizamos nossas experiências em um computador Intel Core 7 Quad CPU Q9550 8 GHz, 16GB de RAM e SO Linux 64 bits. O algoritmo HMIP foi codificado em Matlab 2013a.

i). **HMIP no $\mathbb{H}^n$:**
Considerando a variedade de Hadamard $(\mathbb{H}^n, \langle , \rangle_{(-1,n)})$, o HMIP trabalha essencialmente como descrito a seguir:

Dado um ponto $x^{k-1} \in \mathbb{H}^n$, $k \geq 0$, achar $x^k \in \mathbb{H}^n$ e $\epsilon^k \in T_{x^k}\mathbb{H}^n$ tais que

$$\epsilon^k \in \lambda_k \partial f(x^k) - \operatorname{arcCosh}(-\langle x^k, x^{k-1} \rangle_{(-1,n)}) \left(\frac{x^{k-1} + \langle x^k, x^{k-1} \rangle_{(-1,n)} x^k}{\sqrt{\langle x^k, x^{k-1} \rangle^2_{(-1,n)} - 1}} \right),$$

onde

$$\operatorname{arcCosh}\left(-\langle x^k \cosh(\|\epsilon^k\|) - \frac{\epsilon^k}{\|\epsilon^k\|}\sinh(\|\epsilon^k\|), x^{k-1}\rangle_{(-1,n)}\right) \leq \max\left\{\|\epsilon^k\|, \operatorname{arcCosh}(-\langle x^k, x^{k-1}\rangle_{(-

$$
\begin{aligned}
gradf(x) &= X^2(I-X)^2\nabla f(x) \\
&= \begin{pmatrix} x_1^2(1-x_1)^2 & 0 \\ 0 & x_2^2(1-x_2)^2 \end{pmatrix} \begin{pmatrix} \frac{1-2x_1}{2x_1(1-x_1)\sqrt{-\log(x_1(1-x_1)x_2(1-x_2))}} \\ \frac{1-2x_2}{2x_2(1-x_2)\sqrt{-\log(x_1(1-x_1)x_2(1-x_2))}} \end{pmatrix} \\
&= \begin{pmatrix} \frac{(2x_1-1)x_1(1-x_1)}{2\sqrt{-\log(x_1(1-x_1)x_2(1-x_2))}} \\ \frac{(2x_2-1)x_2(1-x_2)}{2\sqrt{-\log(x_1(1-x_1)x_2(1-x_2))}} \end{pmatrix} \qquad (4.15)
\end{aligned}
$$

Roteiro Matlab

```
%%% fichero onde sera armazenada a funcao:
function y =f(x)
y = sqrt(-log(x(1)*(1-x(1))*x(2)*(1-x(2))));
```

```
%%% fichero onde sera armazenada a gradiente da funcao:
function y = df(x)

y=[(2*x(1)-1)*x(1)*(1-x(1))/2* sqrt(-log(x(1)*(1-x(1))*x(2)*(1-x(2)))) ;
(2*x(2)-1)*x(2)*(1-x(2))/2* sqrt(-log(x(1)*(1-x(1))*x(2)*(1-x(2))))
];
```

```
function [x,err] = HMIPhypercube(x0)

%parameters
itermax = 200;
tolDist = 1e-6;
tolGrad = 1e-6;

% Testing optimality of x0
gradnorm = norm(df(x0));
if gradnorm < tolGrad
    x = x0;
    err = [];
    return
end

% initializations
iter=0;
dist = 1;
gradnorm = 1;
```

```
x = x0;

% Main Loop
while ( iter<itermax && dist > tolDist && gradnorm > tolGrad )


    iter=iter +1
    lambda=1/(iter +1);
    z=rand(2,1);
%  Parametros para hacer una aleatoriedad con distribuicao normal
bivariante
   % r=0.1;
  %  sigma=r/3; %distribuicao normal bivariada
   % z = x + sigma*randn(); %N(x,sigma^2)
   % z(z<=0)=1e-6;
    %z(z>=1)=1-1e-6;

    w = (x.*(1-z))./(z.*(1-x));
    err = lambda*df(z)-(z.*(1-z)).*log(w);
    v=log(w) - err./(z.*(1-z));
    f1 = norm(v);

    f2=norm(log(w));


    while (f1>f2)
        z=rand(2,1);
       % r=0.1;
        %sigma=r/3; %distribuicao normal bivariada
        %z = x + sigma*randn(); %N(x,sigma^2)
        %z(z<=0)=1e-6;
        %z(z>=1)=1-1e-6;

        w = (x.*(1-z))./(z.*(1-x));
        err = lambda*df(z)-(z.*(1-z)).*log(w);
        v=log(w) - err./(z.*(1-z));
        f1 = norm(v);

        f2=norm(log(w));
    end

    x = z

end

end
```

Tabela 4.1: Resultados do algoritmo HMIP2.

$X0$	$Iter.$	$Opt.point$	$Opt.value$	$Riem.distance$
(0.480000,0.550000)	120	(0.499999,0.500003)	1.6666666	9.27 e-005

Na tabela acima, $X0$ denota o ponto inicial do algoritmo, $Iter$ denota o número de iterações que o algoritmo realizou até cumprir o critério de parada estabelecido, $Opt.point$ denota o ponto ótimo aproximado, $Opt.value$ denota o valor ótimo aproximado e $Riem.distance$ denota a distância Riemanniana entre dois pontos iterativos adjacentes.

Além disso, é necessário fazer algumas observações: O código acima foi pensado para algoritmos do tipo HMIP2, considerando sequências $\eta_k = 1/k$, para todo $k \in I\!N$, de modo a facilitar a comparação da distância de dois pontos iterativos adjacentes e o erro ϵ^k. Por outro lado, é bem sabido que resolver desigualdades, no contexto numérico, é um procedimento computacional quase falido.
Para tentar resolver esse inconveniente no código, consideramos a estrategia de gerar pontos aleatórios para obter qualquer ponto que satisfaça a desigualdade referida. Como consequência da escolha dessa estrategia, o tempo computacional, como foi observado quando o código foi simulado, é muito caro. Nessa situação propoe-se tentar simular o código usando o método de Newton para resolver o problema $\lambda_k grad(f(x^k)) - \exp_{x^k}^{-1} x^{k-1} = 0$ aproximadamente, com uma tolerância de erro de $\|\epsilon^k\| < 10^{-1}$. Logo depois, testar se satisfaz a desigualdade $d(\exp_{x^k}\epsilon^k, x^{k-1}) \leq d(x^k, x^{k-1})$. No caso de falhar o teste, diminuimos ainda mais a tolerância do erro até satisfazer o teste.

4.3 Comentários

A hipótese (H3) pode ser também chamada de condição de crescimento no ponto de convergência $\bar{x} \in M$. Repare que essa condição é diferente da condição dada por Tang e Huang[55] a qual foi dada para resolver problemas de singularidade de campos vetoriais monótonos maximais, particularmente para problemas de minimização convexa em variedades de Hadamard. De fato, uma vez que as iterações $\{x^k\}$ são obtidas, a hipótese (H3) estabelece uma propriedade de crescimento do operador subdiferencial no ponto $x^k \in M$, longe do ponto solução $\bar{x}$.

Por outro lado, na taxa de convergência do algoritmo HMIP2 proposto nesta tese, definimos o conjunto $Z := U \cap \{x \in M \mid 0 \in \partial^{Lim} f(x)\}$, o qual é um conjunto não vazio sempre que U for não vazio. Se Z for convexo, então supondo a condição (H2) de Tang e Huang[55] e seguindo as mesmas ideias dessa tese, é possível obter a mesma taxa de convergência do algoritmo proposto neste trabalho. Nesse sentido, estaríamos generalizando a taxa de convergência de Tang e Huang[55] para problemas de minimização quase-convexa em variedades de Hadamard.

Para finalizar, é preciso fazer uma comparação entre os resultados da taxa de convergência de Tang e Huang[55] e dos nossos. De fato, os autores mencionados obtiveram a convergência linear/superlinear da sequência gerada pelo algoritmo do ponto proximal com relação ao conjunto solução do problema. Na nossa abordagem, obtivemos a convergência linear/superlinear da sequência gerada pelo algoritmo proposto nesta tese para um ponto crítico do problema sob a hipótese (H3).

Capítulo 5

Algoritmo do ponto proximal inexato para problemas de minimização multiobjetivo

5.1 Definição do problema multiobjetivo e do algoritmo HMIP

Seja M uma variedade de Hadamard. Nesta tese estamos interessados em resolver o problema:

$$\min_{x\in M} F(x) \tag{5.1}$$

onde $F := (F_1, \ldots, F_m)$, para cada $F_i : M \to I\!R$, é uma função vetorial que satisfaz as seguintes hipóteses:

(H1) F é uma função localmente Lipschitz.

(H2) F é uma função quase-convexa.

Definimos o conjunto

$$\Omega_k := \{x \in M \mid F(x) \preceq F(x^k)\}. \tag{5.2}$$

A versão exata do método do ponto proximal escalarizado no contexto de espaços euclidianos, proposto por Apolinario et al.[3], para resolver (5.1), é dado como segue: dado $x^k \in I\!R^n$, encontrar $x^{k+1} \in \Omega_k$ tal que

$$0 \in \partial^\circ \left(\langle F(\cdot), z^k\rangle + \frac{\lambda_k}{2}\| \cdot - x^k\|^2 \right)(x^{k+1}) + \mathcal{N}_{\Omega_k}(x^{k+1}),$$

onde ∂° é o subdiferencial de Clarke, λ_k é um parâmetro positivo, $\{z^k\} \subset \mathbb{R}^m_+ \backslash \{0\}$ com $\|z^k\| = 1$ e $\mathcal{N}_{\Omega_k}(x^{k+1})$ é o cone normal definido na relação (2.31).
O objetivo desse capítulo é estender o algoritmo do ponto proximal escalarizado de Apolinario et al.[3] para o caso inexato e para variedades de Hadamard, o qual será denotado por HISPP e descrito como segue:

Algoritmo HISPP.

Inicio: Seja um ponto inicial arbitrário $x^0 \in M$. Faça $k = 0$.

Passo iterativo: Dado $x^k \in M$, encontrar $x^{k+1} \in \Omega_k$ tal que

$$\epsilon^{k+1} \in \partial^\circ(\langle F(\cdot), z^k \rangle)(x^{k+1}) - \lambda_k \exp^{-1}_{x^{k+1}} x^k + \mathcal{N}_{\Omega_k}(x^{k+1}), \tag{5.3}$$

onde λ_k é um parâmetro positivo e $\{z^k\} \subset \mathbb{R}^m_+ \setminus \{0\}$com $\|z^k\| = 1$

Critério de parada: Se $x^{k+1} = x^k$ or x^{k+1} é um ponto crítico Pareto-Clarke, parar. Caso contrário, $k + 1 \leftarrow k$ e ir para o Passo iterativo.

Observação 5.1 1. *Se $M = \mathbb{R}^n$, então a equação* (5.3) *vira*

$$\epsilon^{k+1} \in \partial^\circ(\langle F(\cdot), z^k \rangle)(x^{k+1}) - \lambda_k(x^k - x^{k+1}) + \mathcal{N}_{\Omega_k}(x^{k+1}).$$

De Lema 2.7*, temos*

$$\epsilon^{k+1} \in \partial^\circ \left(\langle F(\cdot), z^k \rangle + \frac{\lambda_k}{2} \| \cdot - x^k \|^2 \right) (x^{k+1}) + \mathcal{N}_{\Omega_k}(x^{k+1}).$$

Portanto, o algoritmo HISPP é uma extensão da escalarização do algoritmo do ponto proximal proposto por Apolinario et al.[3].

2. *Se F é uma função simples valorada, $z^k = 1$ e $x^{k+1} \in int(\Omega_k)$, então* (5.3) *se reduz a*

$$\epsilon^{k+1} \in \partial^\circ F(x^k) - \lambda_k \exp^{-1}_{x^{k+1}} x^k, \tag{5.4}$$

que é um caso particular do algoritmo do ponto proximal inexato para minimização quase-convexa estudada no Capítulo 3 desta tese, no sentido que o subdiferencial abstrato discutido aqui é reduzido ao subdiferencial generalizado de Clarke.

Além disso, se $\epsilon^{k+1} = 0$, então como consequência do Lema 2.7, a equação (5.4) fica como

$$0 \in \partial^{\circ}(F(\cdot) + \frac{\lambda_k}{2} d^2(\cdot, x^k))(x^{k+1}).$$

Portanto, o algoritmo HISPP também extende o algoritmo do ponto proximal proposto por Papa Quiroz and Oliveira[45].

Por outro lado, segundo a Observação 3.0.1 feito em Apolinario et al.[3] (ver também Huang e Yang[54]) podemos também considerar, sem perda de generalidade, a seguinte condição:

(C3) $F_i(x) \geq 0$ para todo $i \in \mathcal{I} = \{1, 2, \ldots, m\}$, ou equivalentemente, $0 \preceq F$.

Teorema 5.1 *Considerando as hipóteses* (H1) *e* (H2)*, a sequência* $\{x^k\}$ *gerada pelo algoritmo HISPP está bem definida.*

Prova. Vamos provar por indução. Segundo o passo inicial cumpre-se para $k = 0$. Logo, suponha que $x^k \in M$ existe. Definimos

$$\psi_k(x) = \langle F(x), z^k \rangle + \frac{\lambda_k}{2} d^2(x, x^k) + \delta_{\Omega_k}(x).$$

Observe que $0 \preceq F$ e $\{z^k\} \subset I\!R^m_+ \backslash \{0\}$, então $\psi_k : M \to I\!R \cup \{+\infty\}$ é limitado inferiormente. Lembramos que Ω_k é fechado se e somente se $epi(\Omega_k) = \Omega_k \times I\!R_+$ é fechado se e somente se δ_{Ω_k} é semicontínua inferior. Então, como F é localmente Lipschitz temos que $\psi_k(\cdot)$ é também semicontínua inferior. Além disso, como $d^2(\cdot, x^k)$ é coerciva, ver Exemplo 2.2, segue que $\psi_k(\cdot)$ é também coerciva. Portanto, do Corolario 2.2, existe um mínimo global e pode ser considerado como x^{k+1}. Além disso, como o subdiferencial de Fréchet satisfaz a propriedade (P3) da Definição 2.1, temos que $0 \in \partial^F(\langle F(\cdot), z^k \rangle)(x^{k+1}) + \frac{\lambda_k}{2} \operatorname{grad} d^2(x^{k+1}, x^k) + \mathcal{N}_{\Omega_k}(x^{k+1})$, obtendo assim o resultado.

A condição a seguir, considerada em vários trabalhos na literatura dos algoritmos de pontos proximais, por exemplo em [17, 19, 59], é uma condição necessária e suficiente para a iteração ter um ponto limite e assim garantir a convergência forte do algoritmo HISPP:

(H4) $$(F(x^0) - I\!R^m_+) \cup F(M)$$

é $I\!R^m_+$-completa, isso quer dizer que para todas as sequências $\{a^k\} \subset M$ com $a^0 = x^0$, tal que $F(a^{k+1}) \preceq F(a^k)$ para todo $k \in I\!N$, existe $a \in M$ tal que $F(a) \preceq F(a^k)$ para todo $k \in I\!N$.

Observa-se que quando o critério de parada é atingido em alguma iteração , a sequência permanece constante depois disso e assim é convergente à iteração de parada. Portanto, assumiremos ao longo deste capítulo que a regra do critério de parada não é aplicável.
Denotemos o seguinte conjunto:

$$E := \{x \in M \mid F(x) \preceq F(x^k), \forall\, k \in I\!N\} \tag{5.5}$$

Repare que $E \subseteq \Omega_k$, para qualquer $k \in I\!N$.

Observação 5.2 *Pela definição do algoritmo HISPP, $\{x^k\}$ satisfaz $F(x^{k+1}) \preceq F(x^k)$. Então, pela hipótese* (H4)*, existe $a \in M$ tal que $F(a) \preceq F(x^k)$, para todo $k \in I\!N$. Portanto, o conjunto E, dado por (5.5), é não vazio. Além disso, pelas hipóteses* (H1) *e* (H2)*, E é também um conjunto convexo e fechado, respectivamente. É simples verificar também que $int(\Omega_{k+1})$ é não vazio.*

Os seguintes resultados desempenham um papel central no decorrer desta pesquisa. Mais especificamente, eles serão úteis na análise de convergência das sequências geradas pelo algoritmo HISPP.

Lema 5.1 *Seja $F := (F_1, \ldots, F_m)$, para cada $F_i : M \to I\!R$. Seja $\{x^l\} \subset M$ uma sequência tal que $x^{l+1} \in \Omega_l$. Suponha que as hipóteses* (H1), (C3) *e* (H4) *são satisfeitas. Se $\hat{x} \in M$ é um ponto de acumulação de $\{x^l\}$, então $\hat{x} \in E$.*

Prova. Seja $\hat{x} \in M$ um ponto de acumulação de $\{x^k\}$, então existe uma subsequência $\{x^{l_j}\} \subset \{x^l\}$ tal que $x^{l_j} \to \hat{x}$. Seja $p \in M$ qualquer ponto arbitrário. Uma vez que F é localmente Lipschitz, existe $L_p > 0$ e $\delta_p > 0$ tal que para qualquer $x,\, y \in B(p, \delta_p)$, temos $|F_i(x) - F_i(y)| \leq L_p\, d(x, y)$, para todo $i \in \mathcal{I}$.

Seja $w \in I\!R^m_+ \backslash \{0\}$ um vetor arbitrário. Uma vez que F_i é localmente Lipschitz e como consequência da desigualdade de Cauchy-Schwarz, temos que $|\langle F(x), w\rangle - \langle F(y), w\rangle| \leq L_p \|w\| d(x, y)$. Portanto, $\langle F(\cdot), w\rangle$ é localmente Lipschitz porém contínua em M. Então, segue que $\lim_{j\to+\infty}\langle F(x^{l_j}), w\rangle = \langle F(\hat{x}), w\rangle$.

Por outro lado, como $x^{l+1} \in \Omega_l$ e pela relação (5.2), segue que $F_i(x^{l+1}) \leq F_i(x^l)$ para todo $i \in \mathcal{I}$. Isto é claro que $\langle F(x^{l+1}), w\rangle \leq \langle F(x^l), w\rangle$, para qualquer $w \in I\!R^m_+ \backslash \{0\}$. Portanto, $\langle F(\cdot), w\rangle$ é uma função decrescente. Além disso, pela hipótese (C3) e para qualquer $w \in I\!R^m_+ \backslash \{0\}$, $0 \leq \langle F(x^l), w\rangle$. Então, como $\langle F(\hat{x}), w\rangle$ é um ponto de acumulação de $\{\langle F(x^l), w\rangle\}$, segue que $\langle F(x^l), w\rangle$ é convergente, produzindo assim

$$\langle F(\hat{x}), w\rangle = \lim_{l\to+\infty} \langle F(x^l), w\rangle = \inf_{l\in I\!N} \{\langle F(x^l), w\rangle\} \leq \langle F(x^l), w\rangle, \tag{5.6}$$

para todo $l \in I\!N$.

A relação (5.6) implica que

$$0 \leq \langle F(x^l) - F(\hat{x}), w\rangle, \qquad \forall\, l \in I\!N.$$

e para qualquer vetor $w \in I\!R^m_+$ não nulo. T Seja $w = \{e_1, \ldots, e_m\}$ onde $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$. Isto implica que $F(\hat{x}) \preceq F(x^l)$, para todo $l \in I\!N$. Pela relação (5.5), temos que $\hat{x} \in E$. ■

Proposição 5.1 *Seja* $\{x^k\} \subset M$ *a sequência gerada pelo algoritmo* HISPP *e* $\{\epsilon^k\} \subset T_{x^k}M$ *uma sequência de erro definido em* (5.3). *Se as hipóteses* (H1), (H2), (C3) *e* (H4) *são satisfeitas com* $\tilde{\lambda}$ *tal que* $\tilde{\lambda} < \lambda_k$, *então para qualquer* $x \in E$ *e todo* $k \in I\!N$ *temos que*

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) - d^2(x^{k+1}, x^k) - \frac{2}{\tilde{\lambda}} \langle \epsilon^{k+1}, \exp^{-1}_{x^{k+1}} x\rangle.$$

Prova. Pelo Lema 2.6, a relação (5.3) fica como

$$\epsilon^{k+1} = \sum_{i\in\mathcal{I}} z_i^k g_i^k - \lambda_k \exp^{-1}_{x^{k+1}} x^k + v^k,$$

onde $v^k \in \mathcal{N}_{\Omega_k}(x^{k+1})$ e $g_i^k \in \partial^\circ F_i(x^{k+1})$. Portanto,

$$\exp^{-1}_{x^{k+1}} x^k = \frac{1}{\lambda_k}(\sum_{i\in\mathcal{I}} z_i^k g_i^k + v^k - \epsilon^{k+1}). \tag{5.7}$$

Seja $x \in E \subseteq \Omega_{k+1}$ um ponto arbitrário para qualquer $k \in I\!N$. Uma vez que E é um conjunto fechado, ver Observação 5.2, existe $\{x^l\} \subset int(\Omega_{k+1})$ tal que $\{x^l\}$ converge para x.

Como $x^l \in int(\Omega_{k+1})$, então $F_i(x^{k+1}) > F_i(x^l)$ para todo $i \in \mathcal{I}$. Logo, para $g_i^k \in \partial^\circ F_i(x^{k+1})$ e pelo Lema 2.8, obtém-se que $\langle g_i^k, \exp^{-1}_{x^{k+1}} x^l\rangle \leq 0$, para todo $i \in \mathcal{I}$. Isso implica que

$$\langle \sum_{i\in\mathcal{I}} z_i^k g_i^k, \exp^{-1}_{x^{k+1}} x^l\rangle \leq 0 \tag{5.8}$$

com $z_i^k \in I\!R_+$.

Logo, como $v^k \in \mathcal{N}_{\Omega_k}(x^{k+1}) \subseteq \mathcal{N}_{\Omega_{k+1}}(x^{k+1})$ e $x^l \in \Omega_{k+1}$, segue que $\langle v^k, \exp^{-1}_{x^{k+1}} x^l\rangle \leq 0$. Então a relação (5.8) fica como

$$\begin{aligned} 0 &\geq \langle \sum_{i\in\mathcal{I}} z_i^k g_i^k, \exp^{-1}_{x^{k+1}} x^l\rangle + \langle v^k, \exp^{-1}_{x^{k+1}} x^l\rangle \\ &= \langle \lambda_k \exp^{-1}_{x^{k+1}} x^k + \epsilon^{k+1}, \exp^{-1}_{x^{k+1}} x^l\rangle. \end{aligned} \tag{5.9}$$

A última igualdade obtida na expressão acima é consequência da relação (5.7). Portanto, como $\tilde{\lambda} < \lambda_k$, temos que a relação (5.9) produz

$$\langle \exp^{-1}_{x^{k+1}} x^k, \exp^{-1}_{x^{k+1}} x^l\rangle + \frac{1}{\tilde{\lambda}}\langle \epsilon^{k+1}, \exp^{-1}_{x^{k+1}} x^l\rangle < 0.$$

Do Teorema 2.3, para o triângulo geodésico $[x^k, x^l, x^{k+1}]$, segue que

$$d^2(x^k, x^l) \geq d^2(x^k, x^{k+1}) + d^2(x^{k+1}, x^l) + \frac{2}{\tilde{\lambda}}\langle \epsilon^{k+1}, \exp^{-1}_{x^{k+1}} x^l\rangle, \tag{5.10}$$

e tomando o limite, quando $l \to +\infty$, obtemos o resultado. ■

5.2 Variantes inexatas do algoritmo HISPP para minimização multiobjetivo

Nesta seção, continuando com a mesma abordagem feita no Capítulo 3 desta tese para funções simples-valoradas, discutiremos os resultados de convergência de dois algoritmos proximais inexatos para minimização multiobjetivo. Nessa cena, duas questões devem ser consideradas: primeiro, sobre uma versão clássica quando o critério de erro é considerado com hipóteses naturais; Segundo, sobre o critério de erro de Tang e Huang[55].

5.2.1 Algoritmo HISPP1 e resultados de convergência.

Nesta subseção, discutiremos o algoritmo HISPP, estudado nesse capítulo, considerando a versão clássica do critério do erro de estimação , o qual será chamado de algoritmo HISPP1, descrito a seguir:

Seja $\{x^k\} \in M$ uma sequência gerada pelo algoritmo HISPP. Consideremos a seguinte hipótese:

(A1) $\sum_{k=0}^{+\infty} \|\epsilon^k\| < +\infty$.

A análise de convergência do algoritmo HISPP1 está baseada na teoria de p-Quase-Fejér convergencia, o qual foi apresentado em Ermol'ev [30] no contexto de sequências de variáveis aleatórias. Lembremos que uma sequência $\{x^k\} \subset M$ é dita p-Quase-Fejér convergente para o conjunto $U \subset M$, com $U \neq \emptyset$, se para cada $u \in U$ existe uma sequência não negativa somável $\{\rho_k\}$ tal que

$$d^p(x^{k+1}, u) \leq d^p(x^k, u) + \rho_k \qquad k = 0, 1 \ldots$$

O seguinte resultado sobre p-Quase-Fejér convergencia é bem conhecido.

Lema 5.2 *Seja (M, d) um espaço métrico completo, $p \in (0, +\infty)$ e $\{x^k\} \subset M$ uma sequência p-Quase-Fejér convergente para um conjunto não vazio $U \subset M$, então a sequência $\{x^k\}$ é limitada. Além disso, se um ponto de acumulação $x \in M$ de $\{x^k\}$ pertence a U, então converge para tal ponto.*

Prova. Ver Svaiter[53], Proposição 2.1 ■

Lema 5.3 *Sejam $\{x^k\} \subset M$ e $\{\epsilon^k\} \subset T_{x^k}M$ sequências geradas pelo algoritmo HISPP1. Considere as hipóteses* (H1), (H2), (C3) *e* (H4). *Então , cumpre-se que*

(i) *a sequência $\{x^k\}$ é limitada;*

(ii) $\lim_{k \to +\infty} d(x^k, x^{k+1}) = 0$;

(iii) *se $\lim_{j \to +\infty} x^{k_j} = \hat{x}$ então $\lim_{j \to +\infty} x^{k_j+1} = \hat{x}$.*

Prova. (i) Seja $z \in E$ um ponto fixo. Da Proposição 5.1, temos que

$$\begin{aligned} d^2(x^{k+1}, z) &\leq d^2(x^k, z) - \frac{2}{\tilde{\lambda}} \langle \epsilon^{k+1}, \exp^{-1}_{x^{k+1}} z \rangle \\ &\leq d^2(x^k, z) + \frac{2}{\tilde{\lambda}} \|\epsilon^{k+1}\| d(x^{k+1}, z). \end{aligned}$$

Isto implica que

$$\left(d(x^{k+1}, z) - \frac{1}{\tilde{\lambda}} \|\epsilon^{k+1}\| \right)^2 \leq d^2(x^k, z) + \frac{1}{\tilde{\lambda}^2} \|\epsilon^{k+1}\|^2 \tag{5.11}$$

Isto é claro que se $d(x^{k+1}, z) - \frac{1}{\tilde{\lambda}} \|\epsilon^{k+1}\| \leq 0$, obtem-se $d(x^{k+1}, z) \leq \frac{1}{\tilde{\lambda}} \|\epsilon^{k+1}\| < \delta$, para algum $\delta > 0$. Logo, segue que $\{x^k\}$ é limitado. Nesse caso, repare que para k suficientemente grande, obtemos que $\{x^k\}$ converge para z. Como z foi escolhido arbitrariamente, temos que $E = \{z\}$. Caso contrário, pela relação (5.11) obtemos que

$$\begin{aligned} d(x^{k+1}, z) - \frac{1}{\tilde{\lambda}} \|\epsilon^{k+1}\| &\leq \sqrt{d^2(x^k, z) + \frac{1}{\tilde{\lambda}^2} \|\epsilon^{k+1}\|^2} \\ &\leq d(x^k, z) + \frac{1}{\tilde{\lambda}} \|\epsilon^{k+1}\| \end{aligned}$$

Logo, $d(x^{k+1},z) \le d(x^k,z) + \frac{2}{\tilde{\lambda}}\|\epsilon^{k+1}\|$. Portanto, como consequência da somabilidade de $\|\epsilon^{k+1}\|$ tem-se que $\{x^k\}$ é 1-Quase-Fejér convergente para E. Logo, pelo Lema 5.2, $\{x^k\}$ é limitada.

(ii) Usando de novo a Proposição 5.1, tem-se que

$$\begin{aligned} d^2(x^{k+1},x^k) &\le d^2(x^k,z) - d^2(x^{k+1},z) - \frac{2}{\tilde{\lambda}}\langle \epsilon^{k+1}, \exp^{-1}_{x^{k+1}} z\rangle \\ &\le d^2(x^k,z) - d^2(x^{k+1},z) + \frac{2}{\tilde{\lambda}}\|\epsilon^{k+1}\| d(x^{k+1},z). \end{aligned}$$

Fazendo o somatório, temos

$$\sum_{k=0}^{n} d^2(x^{k+1},x^k) \le d^2(x^0,z) - d^2(x^{k+1},z) + \frac{2}{\tilde{\lambda}} \max\{d(x^{k+1},z)\} \sum_{k=0}^{n} \|\epsilon^{k+1}\|.$$

Pela hipótese (A1), tem-se que $\sum_{k=0}^{n} d^2(x^{k+1},x^k) < +\infty$, obtendo assim o resultado.

(iii) Uma vez que $d(x^{k_j+1},\hat{x}) \le d(x^{k_j+1},x^{k_j}) + d(x^{k_j},\hat{x})$ e $\hat{x}$ é um ponto limite da sequência $\{x^{k_j}\}$, o resultado é obtido pelo item (ii). ∎

Proposição 5.2 *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}M$ *sequências geradas pelo algoritmo HISPP1. Se as hipóteses* (H1), (H2), (C3) *e* (H4) *são satisfeitas com* $\tilde{\lambda} > 0$ *tal que* $0 < \lambda_k < \tilde{\lambda}$, *então a sequência* $\{x^k\}$ *converge para um ponto de* E.

Prova. Pelo Lema 5.3(i), existe $\{x^{k_j}\} \subset \{x^k\}$ e $\hat{x} \in M$ tal que x^{k_j} converge para $\hat{x}$. Além disso, pelo Lema 5.1, $\hat{x} \in E$. Portanto, do Lema 5.2 podemos concluir que a sequência dada é convergente para $\hat{x} \in E$. ∎

Teorema 5.2 *Seja* $F := (F_1, \ldots, F_m)$, *para cada* $F_i : M \to I\!R \cup \{+\infty\}$. *Sejam* $\{x^k\} \subset M$ *e* $\{\epsilon^k\} \subset T_{x^k}$ *sequências geradas pelo algoritmo HISPP1. Se as hipóteses* (H1), (H2), (C3) *e* (H4) *são satisfeitas com* $\tilde{\lambda} > 0$ *tal que* $0 < \lambda_k < \tilde{\lambda}$, *então* $\{x^k\}$ *converge para um ponto crítico Pareto-Clarke em* M.

Prova. Pela Proposição 5.2, existe $\hat{x} \in E$ tal que $\{x^k\}$ converge para $\hat{x}$. Agora, provaremos que $\hat{x}$ é um ponto crítico Pareto-Clarke. Por contradição, suponha que existe $\tilde{d} \in T_{\hat{x}}M$ satisfazendo

$$F_i^{\circ}(\hat{x},\tilde{d}) < 0, \tag{5.12}$$

para todo $i \in \mathcal{I}$. Isto é, existe $\delta > 0$ tal que $F(\exp_{\hat{x}}(t\tilde{d})) \prec F(\hat{x})$, para todo $t \in (0,\delta]$. Denotemos $\gamma(t) := \exp_{\hat{x}}(t\tilde{d})$. Então , $\gamma(t) \in E$

Por outro lado, lembremos que $\{x^{k+1}\} \in \Omega_k$ é uma sequência gerada pela relação

(5.3),

$$\epsilon^{k+1} + \lambda_k \exp^{-1}_{x^{k+1}} x^k - v^k \in \partial^\circ(\langle F(\cdot), z^k\rangle)(x^{k+1}),$$

onde $v^k \in \mathcal{N}_{\Omega_k}(x^{k+1})$. Pela definição de subdiferencial generalizado de Clarke, para todo $d \in T_{x^{k+1}}M$, obtemos

$$\langle \epsilon^{k+1}, d\rangle + \lambda_k \langle \exp^{-1}_{x^{k+1}} x^k, d\rangle - \langle v^k, d\rangle \leq \langle F(\cdot), z^k\rangle^\circ(x^{k+1}, d). \tag{5.13}$$

Seja $d := \exp^{-1}_{x^{k+1}}\gamma(t)$. Uma vez que $\gamma(t) \in \Omega_k$, segue que $\langle v^k, \exp^{-1}_{x^{k+1}}\gamma(t)\rangle \leq 0$, para todo $t \in (0, \delta]$.

Então, a relação (5.13) produz

$$\langle \epsilon^{k+1}, \exp^{-1}_{x^{k+1}}\gamma(t)\rangle + \lambda_k\langle \exp^{-1}_{x^{k+1}} x^k, \exp^{-1}_{x^{k+1}}\gamma(t)\rangle \leq \langle F(\cdot), z^k\rangle^\circ(x^{k+1}, \exp^{-1}_{x^{k+1}}\gamma(t)) \tag{5.14}$$

Pela propriedade da desigualdade triangular e tendo-se em conta que $\|\exp^{-1}_w s\| = d(s, w)$, para todo $w, s \in M$, a relação (5.14) produz

$$-\|\epsilon^{k+1}\| d(\gamma(t), x^{k+1}) - \lambda_k d(x^k, x^{k+1}) d(\gamma(t), x^{k+1}) \leq \langle F(\cdot), z^k\rangle^\circ(x^{k+1}, \exp^{-1}_{x^{k+1}}\gamma(t)). \tag{5.15}$$

Como $\{z^k\}$ é uma sequência de vetor unitário, existe $\{z^{k_j}\} \subseteq \{z^k\}$ tal que $z^{k_j} \to \bar{z}$ com $\bar{z} \in I\!R^m \backslash \{0\}$. Considere uma subsequência, se necessário, sem perda de generalidade, que $\{x^{k_j}\} \subseteq \{x^k\}$. Pelo Lema 2.5, e pelo fato que $\lambda_k < \tilde{\lambda}$, (5.15) tem-se

$$-\|\epsilon^{k_j+1}\| d(\gamma(t), x^{k_j+1}) - \tilde{\lambda} d(x^{k_j}, x^{k_j+1}) d(\gamma(t), x^{k_j+1}) \leq \sum_{i\in\mathcal{I}} z_i^{k_j} F_i^\circ(x^{k_j+1}, \exp^{-1}_{x^{k_j+1}}\gamma(t)). \tag{5.16}$$

Além disso, tem-se que $\lim_{j\to+\infty} \exp^{-1}_{x^{k_j+1}}\gamma(t) = \exp^{-1}_{\hat{x}}(\exp_{\hat{x}} t\tilde{d}) = t\tilde{d}$, with $t \in (0, \delta]$. Repare que pelo Lema 5.3 (ii)-(iii), o lado esquerdo da relação (5.16) tende para zero quando $j \to +\infty$.

Então, aplicando lim sup à relação (5.16) e do fato que $\lim_{j\to+\infty} z_i^{k_j} = \bar{z}_i$, segue da relação (5.16) que

$$\begin{aligned} 0 &\leq \limsup_{j\to+\infty}\Big(\sum_{i\in\mathcal{I}} z_i^{k_j} F_i^\circ(x^{k_j+1}, \exp^{-1}_{x^{k_j+1}}\gamma(t))\Big) \\ &\leq \sum_{i\in\mathcal{I}} \limsup_{j\to+\infty} z_i^{k_j} F_i^\circ(x^{k_j+1}, \exp^{-1}_{x^{k_j+1}}\gamma(t)) \\ &\leq \sum_{i\in\mathcal{I}} \bar{z}_i t F_i^\circ(\hat{x}, \tilde{d}), \end{aligned}$$

com $\bar{z}_i \in I\!R_+$ e $t \in (0, \delta]$. Isso implica que existe pelo menos algum $i_0 \in \mathcal{I}$ tal que $0 \leq F_{i_0}(\hat{x}, \tilde{d})$, o que contradiz a relação (5.12). Portanto, $\hat{x}$ é um ponto crítico Pareto-Clarke. $\blacksquare$

5.2.2 Algoritmo HISPP2 e resultados de convergência.

Nesta subseção, discutiremos o algoritmo HISPP com um critério de erro estudado na Subseção 3.1 para resolver minimização quase-convexa para funções simples valoradas em variedades de Hadamard. Esse algoritmo será chamado de algoritmo HISPP2 e é dado como segue:

Seja $\{x^k\} \subset M$ uma sequência gerada pelo algoritmo HISPP. Considere as seguintes hipóteses sobre o critério de erro:

(B1) $\|\epsilon^k\| \leq \eta_k d(x^k, x^{k-1})$;

(B2) $\sum_{k=1}^{+\infty} \eta_k^2 < +\infty$.

Proposição 5.3 *Seja $F := (F_1, \ldots, F_m)$, para cada $F_i : M \to I\!R \cup \{+\infty\}$. Sejam $\{x^k\} \subset M$ e $\{\epsilon^k\} \subset T_{x^k}M$ sequências geradas pelo algoritmo* HISPP2. *Se as hipóteses* (H1), (H2), (C3) *e* (H4) *são satisfeitas com $\tilde{\lambda} > 0$ tal que $0 < \lambda_k < \tilde{\lambda}$, então a sequência $\{x^k\}$ converge para um ponto de E.*

Prova. O Lema 3.4 implica que existe $\{x^{k_j}\} \subset \{x^k\}$ e $\hat{x} \in M$ tal que x^{k_j} converge para $\hat{x}$. Analogamente ao feito na última parte da prova do Teorema 3.1 em [55], a convergência de toda a sequência $\{x^k\}$ para um ponto $\hat{x} \in M$ é obtida. Aliás, pelo Lema 5.1, $\hat{x} \in E$ $\blacksquare$

Teorema 5.3 *Seja $F := (F_1, \ldots, F_m)$, para cada $F_i : M \to I\!R \cup \{+\infty\}$. Sejam $\{x^k\} \subset M$ e $\{\epsilon^k\} \subset T_{x^k}M$ sequências geradas pelo algoritmo* HISPP2. *Se as hipóteses* (H1), (H2), (C3) *e* (H4) *são satisfeitas com $\tilde{\lambda} > 0$ tal que $0 < \lambda_k < \tilde{\lambda}$, então $\{x^k\}$ converge para um ponto crítico Pareto-Clarke.*

Prova. Consideremos o primeiro parágrafo da prova do Teorema 5.2. Seja $\{x^k\}$ uma sequência gerada pelo algoritmo HISPP2 e $\hat{x} \in M$ tal que x^k converge para $\hat{x}$. Suponha-se que $\hat{x}$ não é ponto crítico Pareto-Clarke. Então , pela hipóteses (B1) e (B2), da relação (5.16) obtém-se que

$$\begin{aligned} \sum_{i \in \mathcal{I}} z_i^{k_j} F_i^{\circ}(x^{k_j+1}, \exp^{-1}_{x^{k_j+1}} \gamma(t)) &\geq -\eta_{k_j} d(x^{k_j+1}, x^{k_j}) d(\gamma(t), x^{k_j+1}) - \tilde{\lambda} d(x^{k_j}, x^{k_j+1}) d(\gamma(t), x^{k_j+1}) \\ &\geq -(M + \tilde{\lambda}) d(x^{k_j+1}, x^{k_j}) d(\gamma(t), x^{k_j+1}), \end{aligned} \tag{5.17}$$

para algum $M > 0$. Repare que $d(\gamma(t), x^{k_j+1}) = \|\exp^{-1}_{x^{k_j+1}} \gamma(t)\|$. Então , $\lim_{j \to +\infty} d(\gamma(t), x^{k_j+1}) = t\|\tilde{d}\|$ para todo $t \in (0, \delta]$. Logo, do fato que $\{x^k\}$ e $\{z^k\}$

são convergentes, o lado direito da relação (5.17) tende para zero quando $j \to +\infty$. Portanto, da relação (5.17) tem-se que

$$\liminf_{j \to +\infty} \sum_{i \in \mathcal{I}} \bar{z}_i F_i^{\circ}(\hat{x}, t\tilde{d}) \geq 0,$$

onde $\bar{z}_i > 0$ para todo $i \in \mathcal{I}$, $t \in (0, \delta]$ e algum $\tilde{d} \in T_{\hat{x}}M$. Isto implica que $F_i^{\circ}(\hat{x}, \tilde{d}) \geq 0$ para algum $i \in \mathcal{I}$, o que contradiz a relação (5.12). ■

5.3 Comentários.

Nesse capítulo, nós geralizamos o algoritmo do ponto proximal escalarizado para problemas de otimização multiobjetivo irrestrito com funções objetivos quase-convexas e localmente Lipschitz de espaços euclidianos, estudados por Apolinario et al. [3], para variedades de Hadamard.

Além disso, sobre hipóteses razoáveis, provamos que qualquer ponto de acumulação é um ponto crítico Pareto-Clarke para quaisquer dessas duas versões inexatas.

Além disso, para calcular as iterações do algoritmo HSIPP precisamos ter uma estimativa dos vetores do cone normal, o que pode ser difícil de obter. Então, superar esse obstáculo requer de algumas técnicas para ser estudadas.

Referências Bibliográficas

[1] Adler, L.R., Dedieu, J.P., Margulies, J.Y., Martens, M. and Shub, M., Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA J. Numer. Anal., 22 (2002).

[2] Arsigny, V., Fillard, P., Pennec, X. and Ayache N., Geometric means in a novel vector space structure on symmetric positive-definite matrices, SIAM Journal on Matrix Analysis and Applications, 29(1), 328–347 (2007)

[3] Apolinario, H.C.F., Papa Quiroz, E.A. and Oliveira, P.R., A scalarization proximal point point method for quasiconvex multiobjective minimization, Journal of Global Opt. (64)1: 79-96 (2016)

[4] Attouch, H., Bolte, J., On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Math. Program. 116, 5-16 (2009)

[5] Attouch, H., Bolte, J., Redont, P. and Soubeyran, A., Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality, Math. Oper. Res. 35(2) 438-457 (2010)

[6] Aussel, D., Corvellec, J.N. and Lassonde, M., Mean-value Property and Subdifferential Criteria for Lower Semicontinuous Functions, Trans. of the American Math. Society, 347, 4147-4161 (1995)

[7] Baygorrea N., Papa Quiroz E.A. and Maculan N., Inexact proximal point methods for quasiconvex minimization on Hadamard manifolds, Journal of the Operations Research Society of China, 4(4), 397-424 (2016)

[8] Baygorrea N., Papa Quiroz E.A. and Maculan N., On the convergence rate of an inexact proximal point algorithm for quasiconvex minimization on Hadamard manifolds, Journal of the Operations Research Society of China, 5(1), 1-11 (2016)

[9] Bento, G.C., Cruz Neto, J.X.: A subgradient method for multiobjective optimization on Riemannian manifolds. J. Optim. Theory Appl. 159(1), 125-137 (2013)

[10] Bento, G.C., da Cruz Neto, J.X., Santos, P.S.M., An inexact steepest descent method for multicriteria optimization on Riemannian manifolds, J. Optim. Theory Appl. 159(1), 108-124 (2013)

[11] Bento, G.C., Ferreira, O.P. and Oliveira, P.R., Proximal point method for a special class of nonconvex functions on Hadamard manifolds, Optimization, 64(2), 289-319 (2015).

[12] Bento, G.C., Ferreira, O.P. and Oliveira, P.R., Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds, Nonlinear Anal. Theory Methods Appl. 73(2), 564-572 (2010)

[13] Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds, J. Optim. Theory Appl. 154(1), 88-107 (2012)

[14] Bento, G.C., da Cruz Neto, J.X., Finite termination of the proximal point method for convex functions on Hadamard manifolds. Optimization 63(9) 1281-1288 (2014)

[15] Bonnabel, S. and Sepulchre, R., Geometric distance and mean for positive semidefinite matrices of fixed rank, SIAM J. Matrix Anal. Appl. 31, 1055-1070 (2009)

[16] Boumal, N. and Absil, P.A., Discrete regression methods on the cone of positive-definite matrices, Pages 4232-4235 of: Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on IEEE (2011)

[17] Bonnel, H., Iusem, A.N. and Svaiter, B.F., Proximal methods in vector optimization, SIAM J. Optim. 15(4), 953-970 (2005)

[18] Bonnel, H., Todjihounde, L., Udriste, C., Semivectorial bilevel optimization on Riemannian manifolds, J. Opt. Theory and App., 167(2), 464–486 (2015)

[19] Ceng, L.C. and Yao, J.C., Approximate proximal methods in vector optimization. European Journal of Operational Research. 183,1-19 (2007)

[20] Correa R., Jofré A., and Thibault L., Subdifferential monotonicity as characterization of convex functions, Numer. Funct. Anal. Optim. 15, 531-535, (1994)

[21] Edelman, A., Arias, T.A. and Smith, S.T., The geometry of algorithms with orthogonality constraints, SIAM J. Matrix Anal. Appl. 20, 303-353 (1998).

[22] Maab, H., Siegel's Modular forms and Dirichlet series, Lecture Notes in Math. 216, Springer-Verlag, Heidelberg (1971)

[23] da Cruz Neto, J.X., Ferreira, O.P. and Lucâmbio Pérez L.R., Monotone point-to-set vector fields, Balkan J. Geom. Appl. 5, 69 -79 (2000)

[24] da Cruz Neto, J.X., Ferreira, O.P. and Lucâmbio Pérez L.R., Contribution to the study of monotone vector fields, Acta Mathematica Hungarica, 94(4), 307-320 (2002)

[25] da Cruz Neto, J.X., Ferreira, O.P., Lucâmbio Perez, L.R., Németh, S.Z., Convex and monotone transformable mathematical programming and a proximal-like point method, J. Glob. Optim. 35, 53-69 (2006)

[26] da Cruz Neto, J.X., Oliveira, P.R., Soares, P.A., Soubeyran, A., Learning how to play Nash, potential games and alternating minimization method for structure nonconvex problems on Riemannian manifolds, J. Convex Anal. 20(2) 395-438 (2013)

[27] do Carmo, M. P., Riemannian geometry, Bikhausen, Boston (1992)

[28] Ekeland, I., The Hopf-Rinow theorem in infinite dimension, J. Differential Geom. 13(2), 287-301 (1978)

[29] Ekeland, I., Nonconvex minimization problems, Bull. Amer. Math. Soc. 1 (3), 443-474 (1979)

[30] Ermol′ev J.M., The method of generalized stochastic gradients and quasi-Fejér sequences, Cybernetics and Systems Analysis, 5(2), 208-220 (1969)

[31] Ferreira, O.P., Lucâmbio Perez, L.R., Németh, S.Z., Singularities of monotone vector fields and a extragradient-type algorithm, J. Glob. Optim. 31, 133-151 (2005)

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

[33] Ioffe A.D., Approximate subdiferentials and applications. I: The finite dimensional theory, Trans. Amer. Math. Soc. 281, 389-416 (1984)

[34] Klingenberg, W., Riemannian geometry, Walter de Gruyter Studies in Mathematics 1, Berlin-New York (1982)

[35] Ledyaev Y. and Zhu Q., Nonsmooth analysis on smooth manifolds, Transactions of the American Mathematical Society, 359(8), 3687-3732 (2007)

[36] Li, C. and Yao, J.C., Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm, SIAM J. Control Optim. 50(4), 2486-2514 (2012)

[37] Miettinen, K.M., Nonlinear multiobjective optimization, Kluwer, Boston (1999)

[38] Papa Quiroz, E.A.,Algumas aplicações da geometria Riemanniana à otimizaçãọ Tese de doutorado. UFRJ. RJ. Brazil (2007)

[39] Papa Quiroz, E.A. and Oliveira, P.R., Proximal point methods for quasiconvex and convex function with Bregman distances on Hadamard manifolds, J. Convex Anal., 355, 469-478 (2009)

[40] Papa Quiroz, E.A. and Oliveira, P.R., An extension of proximal methods for quasiconvex minimization on the nonnegative orthant, Eur. J. Oper. Res. 216 26-32 (2012)

[41] Papa Quiroz E.A., Oliveira P.R., Proximal point method for functions involving Lojasiewicz, quasiconvex and convex properties on Hadamard manifolds, http://www.optimization-online.org/DB_FILE/2008/06/1996.pdf

[42] Papa Quiroz, E. A., O. P. and Oliveira, P. R., Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. J. Convex Anal. 16(1) 49-69 (2009)

[43] Papa Quiroz, E.A., Mallma Ramirez L. and Oliveira, P.R., An inexact proximal method for quasiconvex minimizations, European J. of Operational Research 246(3), 721-729 (2015)

[44] Papa Quiroz, E.A. and Oliveira, P.R., Full convergence of the proximal point method for quasiconvex function on Hadamard manifolds, ESAIM: Control, Optimisation and Calculus of Variations, 18, 483-500 (2012)

[45] Papa Quiroz, E.A. and Oliveira, P.R., Proximal point methods for minimizing quasiconvex locally Lipschitz functions on Hadamard manifolds, Nonlinnear Analysis. 75, 5924-5932, (2012)

[46] Pennec, X., Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. J. of Math. Imaging and Vision,25(1), 127-154 (2006)

[47] Pennec, X., Fillard, P. and Ayache, N.. A Riemannian framework for tensor computing, International J. of Computer Vision 66(1), 41-66 (2006)

[48] Polyak, B.T., Introduction to optimization, Optimization Software, NY (1987)

[49] Rapcsák, T., Smooth nonlinear optimization in $\mathbb{R}^n$, Kluwer Academic Publishers (1997)

[50] Rockafellar, R.T., Monotone operators and the proximal point algorithm, SIAM Journal of control and Opt., 14, 877-898 (1976)

[51] Sakai, T. Riemannian geometry, American Math. Society, Providence (1996)

[52] Stewart, T.J., Bandte, O., Braun, H., Chakraborti, N, Ehrgott, M., Gobelt, M., Jin, Y., Nakayama, H., Poles, S., Distefano, D., Real-World Applications of Multiobjective Optimization, In: J. Branke, K. Deb, K. Miettinen, R. Slowinski (Eds.), Multiobjective Optimization, Lecture Notes in Computer Science, vol. 5252, 285-327 (chapter 11), Springer, Berlin, (2008)

[53] Svaiter, B.F., A class of Fejér convergent algorithms, approximate resolvents and the Hybrid Proximal-Extragradient method, J. Optim. Theory App., 162(1), 133-153 (2014)

[54] Huang, X.X. and Yang, X.Q., Duality for multiobjective optimization via nonlinear Lagrangian functions, Journal of Optimization, theory and App., 120, 111-127 (2004)

[55] Tang G.J. and Huang N.J., An inexact proximal point algorithm for maximal monotone vector fields on Hadamard manifolds. Operations Research Letters, (41)6, 586-591 (2013)

[56] Thibault L. and Zagrodny D., Integration of subdifferentials of lower semicontinuous functions on Banach spaces, J. Math, Anal. Appl. 189, 33-58 (1995)

[57] Tuzel, O., Porikli, F., Meer, P., Human detection via classification on Riemannian manifolds. In: CVPR (2007)

[58] Udriste, C., Convex function and optimization methods on Riemannian manifolds, Kluwer Academic Publishers (1994)

[59] Villacorta, K.D.V and Oliveira, P.R., An interior proximal method in vector optimization. European J. of Operational Research, 214, 485-492 (2011)

[60] White, D.J., A bibliography on the applications of mathematical programming multiple-objective methods, J. of the Operational Research Society 41(8), 669-691 (1990)