COPPE/UFRJ

MÉTODOS PARA OTIMIZAÇÃO EM VARIEDADES RIEMANNIANAS: GRADIENTE PARA FUNÇÕES VETORIAIS E PROXIMAL LOCAL PARA FUNÇÕES REAIS

Glaydston de Carvalho Bento

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

Orientadores: Paulo Roberto Oliveira
Orizon Pereira Ferreira

Rio de Janeiro
Janeiro de 2010

MÉTODOS PARA OTIMIZAÇÃO EM VARIEDADES RIEMANNIANAS: GRADIENTE PARA FUNÇÕES VETORIAIS E PROXIMAL LOCAL PARA FUNÇÕES REAIS

Glaydston de Carvalho Bento

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. Paulo Roberto Oliveira, D.Ing.

Prof. Orizon Pereira Ferreira, D.Sc.

Prof. João Xavier da Cruz Neto, D.Sc.

Prof. Rolando Garciga Otero, Dr.

Prof. Susana Scheimberg de Makler, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JANEIRO DE 2010

Bento, Glaydston de Carvalho

Métodos para otimização em variedades Riemannianas: gradiente para funções vetoriais e proximal local para funções reais/Glaydston de Carvalho Bento. – Rio de Janeiro: UFRJ/COPPE, 2010.

IX, 80 p. $29,7$cm.

Orientadores: Paulo Roberto Oliveira

Orizon Pereira Ferreira

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2010.

Referências Bibliográficas: p. 75 – 80.

1. método de ponto proximal. 2. método de descida. 3. funções não convexas. 4. variedades Riemannianas. 5. Pareto otimalidade. 6. quasi convexidade. 7. otimização multicritério. 8. quasi-Fejér convergência. I. Oliveira, Paulo Roberto *et al.*. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Agradecimentos

Primeiramente a Deus pela oportunidade e força concedida para atravessar os momentos difíceis e acreditar que seria possível continuar mesmo que tudo indicasse que a única alternativa seria retroceder.

Ao Prof. Paulo Roberto Oliveira por ter aceitado me orientar e pelo incentivo, apoio, amizade e confiança com que pude contar na realização deste trabalho.

Ao Prof. Orizon Pereira Ferreira por ter aceitado me orientar e pelo apoio, amizade, confiança e incentivo, com que pude contar em todos momentos da realização deste trabalho.

Aos Professores João Xavier da Cruz Neto, Rolando Garciga Otero e Susana Scheimberg de Makler por terem aceito participar da banca de defesa desta tese de doutorado, pelo tempo que disponibilizaram à leitura da mesma e pelas consideráveis observações e sugestões.

Ao Prof. Geci José Pereira da Silva através do qual tive meu primeiro contato com a otimização (durante a orientação de minha dissertação de mestrado) sendo uma pessoa com a qual pude contar com o apoio, amizade e compreensão.

À Prof. Rosely Maria Barbosa Goes pelo apoio e por ter acreditado em mim num momento em que muito precisei.

Ao Prof. Ronaldo Alves Garcia que foi responsável por uma iniciativa que sem dúvida foi fundamental para que a tese fosse concluída dentro do prazo estabelecido.

Aos Professores e colegas do IME/UFG pelo apoio, tendo assumido minhas aulas durante o período que se fez necessário, e à direção do IME que, na figura da Prof. Gisele de Araújo Prateado Gusmão, me prestou considerável apoio.

Aos meus pais que me deram o suporte necessário para o início da minha carreira escolar, em especial à minha mãe que sempre se fez presente me estimulando e apoiando; pela paciência e grande amizade com que sempre me ouviu e ajudou.

À minha esposa pelo inestimável apoio familiar que preencheu as diversas falhas que fui tendo por força das circunstâncias, e pela paciência e compreensão reveladas ao longo destes anos.

À minha irmã Valéria pelo apoio, amizade e carinho.

Ao Felipe Antonio Garcia Moreno e à Kely Diana V. Villacorta com os quais foi estabelecida uma boa amizade e pude contar com o apoio em diversos momentos que precisei.

A todos os colegas que tive a oportunidade de conhecer durante o curso de doutorado no PESC/COPPE-UFRJ.

Ao João Eduardo Reis pelo apoio e amizade.

A todos os Professores e Funcionários do PESC/COPPE-UFRJ e também aos funcionários do IME/UFG que de alguma forma contribuíram para a realização deste trabalho.

Aos demais amigos e familiares, pelo apoio direto e indireto durante toda a realização do curso.

A todas as pessoas que não constam nesta lista, mas que de alguma forma contribuíram para a realização deste trabalho.

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

MÉTODOS PARA OTIMIZAÇÃO EM VARIEDADES RIEMANNIANAS: GRADIENTE PARA FUNÇÕES VETORIAIS E PROXIMAL LOCAL PARA FUNÇÕES REAIS

Glaydston de Carvalho Bento

Janeiro/2010

Orientadores: Paulo Roberto Oliveira
Orizon Pereira Ferreira

Programa: Engenharia de Sistemas e Computação

Apresentamos uma análise local do método de ponto proximal para uma classe particular de funções não convexas em variedades de Hadamard. A boa definição da seqüência gerada pelo método é garantida. Além disso, provamos que cada ponto de acumulação desta seqüência (caso existam) satisfaz a condição de otimalidade de primeira ordem e, com respeito a hipóteses adicionais, convergência total a um minimizador é obtida.

Sob uma outra perspectiva, apresentamos também um método de máxima descida com regra de Armijo para otimização multicritério no contexto Riemanniano. A boa definição da seqüência gerada pelo método é garantida e, com respeito a hipóteses moderadas sobre a função multicritério, provamos que os pontos de acumulação desta seqüência (caso existam) satisfazem a condição necessária de otimalidade segundo Pareto. Além disso, assumindo quasi-convexidade da função multicritério e curvatura não negativa da variedade Riemanniana, provamos convergência total da seqüência a um Pareto crítico.

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

OPTIMIZATION METHODS ON RIEMANNIAN MANIFOLDS: GRADIENT FOR VECTOR FUNCTIONS AND LOCAL PROXIMAL FOR REAL FUNCTIONS

Glaydston de Carvalho Bento

January/2010

Advisors: Paulo Roberto Oliveira
Orizon Pereira Ferreira

Department: Systems Engineering and Computer Science

We presented a local analysis of the proximal point method for a particular class of nonconvex functions on Hadamard manifolds. The well definedness of the sequence generated by the method is guaranteed. Moreover, we proved that each accumulation point of this sequence (if they exist) satisfies the necessary optimality condition and, under additional assumptions, full convergence to a minimizer is obtained.

Under another perspective, we also presented a steepest descent method with Armijo's rule for multicriteria optimization in the Riemannian context. The well definedness of the sequence generated by the method is guaranteed and, under mild assumptions on the multicriteria function, we proved that the accumulation points of this sequence (if they exist) satisfy the Pareto's necessary optimality condition. Moreover, assuming quasi-convexity of the multicriteria function and non-negative curvature of the Riemannian manifold, we prove full convergence of the sequence to a Pareto critical point.

Sumário

Introdução

A extensão de conceitos, técnicas bem como métodos da Programação Matemática, desenvolvido no espaço Euclideano $\mathbb{R}^n$, a variedades Riemannianas é natural e, em geral, não trivial. Nos últimos anos tais extensões, com propósitos teórico e também prático, tem sido assunto de muitas recentes pesquisas. Recentes trabalhos tratando sobre este assunto incluem [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. A generalização de métodos da Programação matemática desenvolvidos no espaço Euclideano ao contexto Riemanniano garante certas vantagens importantes. Por exemplo, problemas de otimização restritos podem ser considerados como irrestritos do ponto de vista da geometria Riemanniana e, neste caso, temos uma possibilidade alternativa além da idéia de projeção para resolver o problema. Além disso, problemas não convexos no sentido clássico podem tornar-se convexos através da introdução de uma métrica Riemanniana apropriada (ver, por exemplo, [6]).

Considere o seguinte problema de minimização

$$\begin{aligned} &\min F(p) \\ &\text{s.t. } \ p \in X. \end{aligned} \tag{1}$$

Se $F : X \to \mathbb{R}$ e $X = \mathbb{R}^n$, o método de ponto proximal, introduzido por MARTINET [15] e ROCKAFELLAR [16], gera uma seqüência $\{p^k\} \subset X$ do seguinte modo: dado $p^0 \in X$ e algum $\tilde{\lambda} > 0$,

$$p^{k+1} := \operatorname{argmin}_{p\in X} \left\{ F(p) + \frac{\lambda_k}{2}\|p - p^k\|^2 \right\}, \tag{2}$$

onde a seqüência de números reais $\{\lambda_k\}$ satisfaz

$$0 < \lambda_k \leq \tilde{\lambda}, \tag{3}$$

e $\| \, . \, \|$ é a norma Euclideana, ver também IUSEM [17]. A boa definição da seqüência $\{p^k\}$ é garantida sob a hipótese de convexidade da função objetivo F. Além disso, se o conjunto solução do Problema (1), U^*, é não vazio, a convergência da seqüência a um minimizador $p^* \in U^*$ é garantida. A convergência do método baseia-se na chamada Féjer convergência da seqüência $\{p^k\}$ a U^*, que ocorre em virtude da convexidade de F junto com uma propriedade geométrica característica do espaço Euclideano, a saber, a chamada lei dos cossenos. A primeira proposta de extensão deste método ao contexto Riemanniano, isto é, quando $X = M$ (M é uma variedade Riemanniana) foi feita por FERREIRA E OLIVEIRA [18], onde os autores propuseram uma generalização natural de (2) como segue

$$p^{k+1} := \operatorname{argmin}_{p \in M} \left\{ F(p) + \frac{\lambda_k}{2} d^2(p, p^k) \right\}, \tag{4}$$

com $p^\circ \in M$ um ponto arbitrário, d a distância Riemanniana intrinseca e $\{\lambda_k\}$ uma seqüência de números positivos. A boa definição e a convergência total foram generalizados sob a hipótese restritiva de a variedade Riemanniana ser do tipo Hadamard. Esta classe de variedades, além de contar com a mesma topologia e estrutura diferenciável do espaço Euclideano, também goza de propriedades geométricas importantes à análise de convergência do método, entre elas uma lei dos cossenos com desigualdade satisfatória à caracterização da Féjer convergência da seqüência $\{p^k\}$ ao conjunto solução U^*. Observe que no método proposto os autores descartaram a hipótese de existência de um parâmetro $\tilde{\lambda}$ tal que a seqüência $\{\lambda_k\}$ satisfizesse (3). Porém, na prova de convergência foi fundamental a hipótese adicional de que $\sum_{k=0}^{+\infty} 1/\lambda_k = +\infty$. Seguindo a mesma idéia de [18], PAPA QUIROZ E OLIVEIRA [12] generalizaram o método de ponto proximal com distância de Bregman para resolver problemas de otimização convexo e quasiconvexo também em variedades de Hadamard. LI et al. [9] estenderam a importante noção de monotonicidade maximal de um operador ponto-a-conjunto, definido em um espaço de Banach, para campos de vetores ponto-a-conjunto em uma variedade de Hadamard. Além disso os autores apresentaram um método de ponto proximal geral para encontrar singularidades de um campo de vetores ponto-a-conjunto. Em particular, como uma aplicação do resultado de convergência obtido para o método proposto, problemas de otimização restritos têm sido resolvidos.

Um objeto de pesquisa interessante, no contexto Riemanniano, seria estender o campo de aplicação do método de ponto proximal para resolver problemas de otimização não convexos em variedades de Hadarmad, a saber, quando a função objetivo F não é necessariamente convexa no sentido Riemanniano. No caso do problema de encontrar singularidades de operadores ponto-a-conjunto essa situação é similar à ausência de monotonicidade do operador (por exemplo que o operador seja hypomonótono), ver por exemplo [19, 20, 21, 22]. SPINGARN [19] tem, em particular, desenvolvido o método de ponto proximal para a minimização de uma certa classe de funções não convexas e não diferenciáveis, a saber, as funções lower-C^2 definidas no espaço Euclideano, ver também [23]. É válido ressaltar que, no caso Euclideano, uma função (localmente Lipschitz) é lower-C^2 se, e somente se, seu subdiferencial generalizado de Clarke é estritamente hypomonótono, ver [24]. KAPLAN E TICHATSCHKE [25] também aplicaram o método de ponto proximal para a minimização de uma classe similar àquelas de [19], a saber, funções definidas como o máximo de uma certa coleção (finita/infinita) de funções continuamente diferenciáveis. Sob essa perspectiva, seguindo as idéias de KAPLAN E TICHATSCHKE [25], neste trabalho estudamos a mesma classe de funções estudada pelos autores, agora no contexto Riemanniano, e aplicamos o método de ponto proximal (4) para resolver o Problema (1) com a função objetivo naquela classe e $X = M$ (M variedade de Hadamard). Para isto, se fez necessário estudar a derivada direcional e o subdiferencial generalizados no contexto Riemanniano. Vários trabalhos têm estudado tais conceitos e apresentado muitos resultados úteis à otimização Riemanniana, ver por exemplo [3], [8], [26] e [27].

Considerando novamente o Problema (1), no caso que $F : X \to \mathbb{R}$ é uma função continuamente diferenciável e $X = \mathbb{R}^n$, o método de máxima descida com regra de Armijo gera uma seqüência $\{p^k\}$ como segue

$$p^{k+1} = p^k + t_k v^k, \qquad v^k = -F'(p^k), \qquad k = 0, 1, \ldots,$$

onde

$$t_k = \max\{2^{-j} : F(p^k + tv^k) \leq F(p^k) + \beta 2^{-j} v^k, j = 0, 1 \ldots\},$$

$\beta \in (0, 1)$. Se F é continuamente diferenciável, resultados clássicos asseguram somente que os pontos de acumulação de $\{p^k\}$, caso existam, são críticos de F. Este fato foi generalizado para otimização multicritério por FLIEGE E SVAITER [28], i.e., quando a função objetivo é uma função vetorial $F : \mathbb{R}^n \to \mathbb{R}^m$ e a ordem parcial em $\mathbb{R}^m$ é a usual (ou seja, a ordem componente a componente). A convergência total é assegurada com respeito à hipótese de o conjunto solução U^* ser não vazio e $F : \mathbb{R}^n \to \mathbb{R}$ ser uma função convexa, ver BURACHIK et al. [29] (ou, mais geralmente, quasi-convexa, ver KIWIEL E MURTY [30]), o que tem sido generalizado para otimização vetorial por GRAÑA DRUMMOND E SVAITER [31] (ver também, GRAÑA DRUMMOND E IUSEM [32]).

O método de máxima descida para o Problema (1), no caso que $X = M$ (M uma variedade Riemanniana completa e conexa) e $F : M \to \mathbb{R}$ é continuamente diferenciável tem sido estudado por UDRISTE [33], SMITH [34] e RAPCSÁK [35] e resultados de convergência parcial foram obtidos. Para o caso convexo o resultado de convergência total, usando regra de Armijo, tem sido generalizado por DA CRUZ NETO et al. [36], no caso particular que M tem curvatura não negativa. Com respeito à mesma hipótese restritiva sobre a variedade M, PAPA QUIROZ et al. [11] generalizaram o resultado de convergência total usando regra de Armijo generalizado para o caso quasi-convexo.

Ressaltamos que no caso Euclideano (real ou vetorial) a análise de convergência total baseia-se na chamada quasi-Féjer convegência da seqüência $\{p^k\}$ a um certo conjunto não vazio, a qual foi provada ocorrer graças à convexidade da função F junto com a lei dos cossenos. O mesmo ocorre no caso Riemanniano mas com uma hipótese restritiva sobre a variedade M, a saber, que a mesma tenha curvatura não negativa. Para esta classe de variedades a chamada lei dos cossenos ocorre com desigualdade satisfatória à caracterização da quasi-Féjer convergência da seqüência $\{p^k\}$.

Também como nossa contribuição, seguindo as idéias de FLIEGE E SVAITER [28], apresentamos uma generalização do resultado de convergência parcial para otimização multicritério ao contexto Riemanniano. Além disso, seguindo as idéias de GRAÑA DRUMMOND E SVAITER [31], generalizamos o resultado de convergência total para otimização multicritério no caso que a função multicritério é quasi-convexa e levando em conta que a variedade Riemanniana M tem curvatura não negativa.

A organização deste trabalho é como segue. No Capítulo 1 definimos as notações e apresetamos um breve resumo dos conceitos e resultados básicos de geometria Riemanniana úteis ao longo do trabalho. No Capítulo 2 recordamos alguns fatos de análise convexa em variedades de Hadamard. No capítulo 3 apresentamos algumas propriedades da derivada direcional de uma função convexa definida sobre uma variedade de Hadamard e também uma definição para a derivada direcional (respectivamente, subdiferencial) generalizada de uma função localmente Lipschitz (não necessariamente convexa), incluindo algumas propriedades desses conceitos. No capítulo 4 estudamos o método de ponto proximal (4) para resolver o Problema (1) com a função objetivo sendo dada pelo máximo de uma certa classe de funções continuamente diferenciáveis e $X = M$ (M variedade de Hadamard). Além disso apresentamos dois exemplos onde o método em questão se aplica. No Capítulo 5 apresentamos o problema multicritério, condições de otimalidade de primeira ordem para ele e algumas definições básicas relacionadas. Propomos um método de descida para minimizar uma função vetorial em uma variedade Riemanniana completa e apresentamos uma completa análise de convergência. Além disso apresentamos alguns exemplos de variedades Riemannianas completas com curvas geodésicas explícitas e a iteração de descida da seqüência gerada pelo método proposto. No Capítulo 6 fazemos as considerações finais e apresentamos algumas possibilidades de pesquisa futura. Finalmente no Capítulo 7 apresentamos, numa situação particular, uma caracterização do subdiferencial generalizado de Clarke de uma função dada pelo máximo de uma coleção finita de funções continuamente diferenciáveis, fundamental na exposição do Exemplo 4.1.2.

Capítulo 1

Preliminares sobre geometria Riemanniana

Neste capítulo introduzimos algumas notações e propriedades fundamentais sobre geometria Riemanniana. Esses fatos básicos podem ser encontrados em qualquer livro introdutório de geometria Riemanniana, tais como em [37] e [38]. Por questões de uniformidade vamos unificar a notação de acordo com [37].

Seja M uma variedade conexa n-dimencional. Denotamos por T_pM o *espaço tangente* n-dimencional de M em p, por $TM = \cup_{p\in M} T_pM$ o *fibrado tangente* de M e por $\mathcal{X}(M)$ o conjunto dos campos de vetores de classe C^∞ em M. Quando M é munida com uma métrica Riemanniana $\langle\, ,\, \rangle$, com a correspondente norma denotada por $\|\ \|$, então M é uma variedade Riemanniana. Recorde que a métrica pode ser usada para definir o comprimento de curvas diferenciáveis por partes $\gamma : [a,b] \to M$ ligando p a q, i.e., tais que $\gamma(a) = p$ e $\gamma(b) = q$, por

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

e, além disso, minimizando esse funcional comprimento sobre o conjunto de todas tais curvas, obtemos a distância Riemanniana $d(p,q)$ que induz a topologia original em M. A métrica induz uma aplicação $f \mapsto \operatorname{grad} f \in \mathcal{X}(M)$ que associa a cada função diferenciável em M seu gradiente via a regra $\langle \operatorname{grad} f, X\rangle = Df(X)$, $X \in \mathcal{X}(M)$, onde Df representa a diferencial de f. Seja ∇ a convexão Levi-Civita associada a $(M, \langle\, ,\, \rangle)$. Em cada ponto $p \in M$, temos uma aplicação linear $A_X(p)\colon T_pM \to T_pM$ definida por

$$A_X(p)v = \nabla_v X.$$

Se $X = \operatorname{grad} f$, onde $f\colon M \to \mathbb{R}$ é uma função duas vezes diferenciável, então $A_X(p)$ é a *Hessiana* de f em p e é denotado por Hess ${}_p f$. Um campo de vetores V ao longo de γ é dito ser *paralelo* se $\nabla_{\gamma'} V = 0$. Se o campo tangente γ' é paralelo ao longo de γ dizemos que γ é uma *geodésica.* Dado que a equação geodésica $\nabla\ _{\gamma'}\gamma' = 0$ é uma equação diferencial ordinária não linear de segunda ordem, então a geodésica $\gamma = \gamma_v(.,p)$ é determinada por sua posição p e velocidade v em p. É fácil ver que $\|\gamma'\|$ é constante. Dizemos que γ é *normalizada* se $\|\gamma'\| = 1$. A restrição de uma geodésica a um intervalo fechado e limitado é chamado um *segmento geodésico.* Um segmento geodésico ligando p a q em M é dito ser *minimal* se seu comprimento é igual a $d(p,q)$ e essa geodésica é chamada uma *geodésica minimizante.* Se $\gamma : [a,b] \to M$ é um segmento geodésico ligando os pontos $p := \gamma(a)$ e $q := \gamma(b)$ em M então, para cada $t \in [a,b]$, ∇ induz uma isometria linear, relativa a $\langle\ ,\ \rangle$, $P_{\gamma(a)\gamma(t)} : T_{\gamma(a)}M \to T_{\gamma(t)}M$, chamada *transporte paralelo* ao longo de γ de $\gamma(a)$ a $\gamma(t)$. A aplicação inversa de $P_{\gamma(a)\gamma(t)}$ é denotada por $P^{-1}_{\gamma(a)\gamma(t)} : T_{\gamma(t)}M \to T_{\gamma(a)}M$. No caso particular que γ é o único segmento geodésico ligando os pontos p e q em M, então o transporte paralelo ao longo de γ de p a q é denotado por $P_{pq} : T_pM \to T_qM$.

Uma variedade Riemanniana é *completa* se as geodésicas são definidas para qualquer valor de t. O Teorema de Hopf-Rinow assegura que se este é o caso, então qualquer par de pontos, digamos p e q, em M pode ser ligado por um (não necessariamente único) segmento geodésico minimal. Além disso, (M,d) é um espaço métrico completo e, subconjuntos fechados e limitados são compactos. Neste caso, dado $p \in M$, a *aplicação exponencial* $exp_p : T_pM \to M$ é definida por $exp_p v = \gamma_v(1,p)$.

Denotemos por R *o tensor curvatura* definido por

$$R(X,Y) = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z, \qquad X,Y,Z \in \mathcal{X}(M),$$

onde $[X,Y] = YX - XY$. Então, a *curvatura seccional* com respeito a X e Y é dado por $K(X,Y) = \langle R(X,Y)Y, X\rangle/(||X||^2||Y||^2 - \langle X\ ,\ Y\rangle^2)$, onde $||X|| = \langle X, X\rangle^{\frac{1}{2}}$. Se $K(X,Y) \leqslant 0$ para todo X e Y, então M é chamada *variedade Riemanniana de curvatura não positiva* e usamos a seguinte notação $K \leqslant 0$. Se $K(X,Y) \geqslant 0$ para todo X e Y, então M é chamada *variedade Riemanniana de curvatura não negativa* e usamos a seguinte notação $K \geqslant 0$.

Na Seção 5.3.2 deste trabalho nosso interesse foi as variedades Riemannianas de curvatura não negativa. Uma propriedade geométrica fundamental desta classe de variedades é a *lei dos cossenos* que agora passamos a descrever. Um *ângulo geodésico* em M é um par de segmentos geodésicos normalizados γ_1 e γ_2 tais que $\gamma_1(0) = \gamma_2(0)$ com pelo menos uma delas, digamos γ_1, sendo minimal. De agora em diante $l_1 = l(\gamma_1)$, $l_2 = l(\gamma_2)$, $l_3 = d(\gamma_1(l_1), \gamma_2(l_2))$ e $\alpha = \sphericalangle(\gamma_1'(0), \gamma_2'(0))$.

Teorema 1.1 *(Lei dos cossenos) Seja M uma variedade Rieamanniana completa com curvatura não negativa. Com respeito à notação introduzida acima, temos*

$$l_3^2 \leq l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha. \tag{1.1}$$

Prova. Este resultado é uma conseqüência do Teorema de Toponogov. Ver Teorema 4.2, p. 161 de [38]. **c.q.d**

Teorema 1.2 *Seja M uma variedade Riemanniana completa e simplesmente conexa com curvatura seccional não positiva. Então, M é difeomorfa ao espaço Euclideano $\mathbb{R}^n$, $n = dimM$. Mais precisamente, em qualquer ponto $p \in M$, a aplicação exponencial $exp_p : T_pM \to M$ é um difeomorfismo.*

Prova. Ver Lema 3.2 de [37], p. 149, ou Teorema 4.1 de [38], p. 221. **c.q.d**

Uma variedade Riemanniana completa e simplismente conexa com curvatura seccional não positiva é chamada uma *variedade de Hadamard.* O Teorema 1.2 diz que se M é uma variedade de Hadamard, então M tem a mesma topologia e estrutura diferenciável do espaço Euclideana $\mathbb{R}^n$. Além disso, são conhecidas algumas propriedades geométricas similares às do espaço Euclideano $\mathbb{R}^n$, tais como, dados dois pontos existe um único segmento geodésico os ligando. *Com excessão do Capítulo 5, onde as variedades M são assumidas serem apenas conexas e completas com dimensão finita, em todo o trabalho as variedades M são assumidas serem de Hadamard com dimensão finita.*

Capítulo 2

Convexidade em variedades de Hadamard

Neste capítulo introduzimos algumas propriedades e notações fundamentais de análise convexa, em variedade de Hadamard, que serão usadas nos próximos capítulos. Veremos que essas propriedades são similares àquelas obtidas em análise convexa no espaço Euclideano $\mathbb{R}^n$. Para referências sobre análise convexa no espaço Euclideano citamos [39] e [40], e em variedade Riemanniana [41], [18], [35], [38], [34] e [33].

O conjunto $\Omega \subset M$ é dito ser *convexo* se qualquer segmento geodésico com pontos finais em Ω está contido em Ω. Seja $\Omega \subset M$ um conjunto aberto e convexo. Uma função $f : M \to \mathbb{R}$ é dita ser *convexa* (respectivamente, *estritamente convexa*) em Ω se, para qualquer segmento geodésico $\gamma : [-\delta, \delta] \to \Omega$, $\delta > 0$, a composição $f \circ \gamma : [-\delta, \delta] \to \mathbb{R}$ é convexa (respectivamente, estritamente convexa). Agora, uma função $f : M \to \mathbb{R}$ é dita ser *fortemente convexa* em Ω com constante $L > 0$ se, para qualquer segmento geodésico $\gamma : [-\delta, \delta] \to \Omega$, a composição $f \circ \gamma : [-\delta, \delta] \to \mathbb{R}$ é fortemente convexa com constante $L\|\gamma'(0)\|^2$. Tome $p \in M$. Um vetor $s \in T_pM$ é dito ser um *subgradiente* de f em p se,

$$f(q) \geq f(p) + \langle s, \exp_p^{-1} q \rangle,$$

para todo $q \in M$. O conjunto de todos os subgradientes de f em p, denotado por $\partial f(p)$, é chamado o *subdiferencial* de f em p.

O próximo resultado provê uma caracterização para convexidade no caso de funções diferenciáveis.

Proposição 2.1 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $f : M \to \mathbb{R}$ uma função diferenciável em Ω. Dizemos que f é convexa em Ω se, e somente se, para qualquer $p \in \Omega$*

$$f(q) - f(p) \geq \langle \operatorname{grad} f(p), \exp_p^{-1} q \rangle,$$

para todo $q \in \Omega$.

Prova. Ver Teorema 5.1 de [33], página 78. **c.q.d**

A conseqüência mais importante da proposição anterior é que, se f é convexa, então qualquer de seus pontos críticos são pontos de mínimo global. Além disso, $0 \in \partial f(p)$ se, e somente se, p é um ponto de mínimo de f em M.

Tome $p \in M$. Seja $exp_p^{-1} : M \to T_pM$ a inversa da aplicação exponencial que é também C^∞. Note que $d(q\,,\,p) = ||exp_p^{-1}q||$, a aplicação $d^2(\,.\,,p)\colon M \to \mathbb{R}$ é C^∞ e

$$\operatorname{grad} \frac{1}{2} d^2(q,p) = -exp_q^{-1} p,$$

(recorde que M é uma variedade de Hadamard) ver, por exemplo, [38].

Definição 2.1 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e X um campo de vetores definido em M. X é dito ser monótono em Ω, se*

$$\left\langle \exp_q^{-1} p\,,\, P_{qp}^{-1} X(p) - X(q) \right\rangle \geq 0, \qquad p,\, q \in \Omega, \tag{2.1}$$

onde P_{qp} é o transporte paralelo ao longo da geodésica ligando q a p. Se (2.1) é satisfeita com desigualdade estrita para todo $p, q \in \Omega$, $p \neq q$, então X é dito ser estritamente monótono. Além disso, X é fortemente monótono se existe $\lambda > 0$ tal que

$$\left\langle \exp_q^{-1} p\,,\, P_{qp}^{-1} X(p) - X(q) \right\rangle \geq \lambda d^2(p,q), \qquad p,\, q \in \Omega. \tag{2.2}$$

Observação 2.1 *No caso particular que $M = \mathbb{R}^n$ com a métrica usual, as desigualdades* (2.1) *e* (2.2) *são reduzidas a*

$$\langle p - q\,,\, X(p) - X(q) \rangle \geq 0, \qquad \langle p - q\,,\, X(p) - X(q) \rangle \geq \lambda \|p - q\|^2,$$

visto que $\exp_q^{-1} p = p - q$ e $P_{qp}^{-1} = I$. Portanto a definição 2.1 estende, à variedade Riemanniana, o conceito de operadores monótonos em $\mathbb{R}^n$.

O seguinte resultado nos provê um importante exemplo de campo de vetores monótono que será útil neste trabalho.

Proposição 2.2 *Seja $p \in M$. O campo gradiente* $\operatorname{grad}(d^2(\,.\,,p)/2)$ *é fortemente monótono com $\lambda =\ 1$.*

Prova. Ver Proposição 3.2 de [41]. **c.q.d**

Proposição 2.3 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $f : M \to \mathbb{R}$ uma função diferenciável em Ω.*

(i) f é convexa em Ω se, e somente se, o campo de vetores $\operatorname{grad} f$ *é monótono em Ω;*

(ii) f é estritamente convexa em Ω se, e somente se, o campo de vetores $\operatorname{grad} f$ *é estritamente monótono em Ω;*

(iii) f é fortemente convexa em Ω se, e somente se, o campo de vetores $\operatorname{grad} f$ *é fortemente monótono em Ω.*

Prova. Ver Proposição 3.4 de [41]. **c.q.d**

Como uma conseqüência imediata das duas últimas proposições, temos a seguinte

Proposição 2.4 *Seja $p \in M$. A aplicação $d^2(\,.\,,p)/2$ é fortemente convexa.*

Proposição 2.5 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, $T \subset \mathbb{R}$ um conjunto compacto e $\psi : M \times T \to \mathbb{R}$ uma função contínua em $\Omega \times T$ tal que $\psi_\tau := \psi(.,\tau) : M \to \mathbb{R}$ é fortemente convexa em Ω com constante $L > 0$ para todo $\tau \in T$. Então, $\phi : M \to \mathbb{R}$, definida por*

$$\phi(p) := \max\nolimits_{\tau \in T} \psi(p,\tau),$$

é fortemente convexa em Ω com constante L. Em particular, se ψ_τ é convexa para todo $\tau \in T$, então ϕ é convexa em Ω.

Prova. Visto que T é compacto e ψ é contínua, a função ϕ está bem definida. Seja $\gamma : [-\delta,\delta] \to \Omega$, $\delta > 0$, um segmento geodésico. Dado que ψ_τ é fortemente convexa com constante L para cada $\tau \in T$, temos

$$(\psi_\tau\circ\gamma)(\alpha t_1+(1-\alpha)t_2) \leq \alpha(\psi_\tau\circ\gamma)(t_1)+(1-\alpha)(\psi_\tau\circ\gamma)(t_2)-\frac{1}{2}(1-\alpha)\alpha L\|\gamma'(0)\|^2(t_1-t_2)^2,$$

para todo $t_1, t_2 \in [-\delta, \delta]$ e $\alpha \in [0,1]$. Assim, tomando o máximo em τ em ambos os lados da desigualdade acima, obtemos

$$(\phi\circ\gamma)(\alpha t_1+(1-\alpha)t_2) \leq \alpha(\phi\circ\gamma)(t_1)+(1-\alpha)(\phi\circ\gamma)(t_2)-\frac{1}{2}(1-\alpha)\alpha L\|\gamma'(0)\|^2(t_1-t_2)^2,$$

o que implica que $\phi\circ\gamma : [-\delta, \delta] \to \mathbb{R}$ é fortemente convexa com constante $L\|\gamma'(0)\|^2$. Assim, ϕ é fortemente convexa em Ω com constante L. A prova da segunda parte é imediata. **c.q.d**

Definição 2.2 *Seja $\Omega \subset M$ um conjunto aberto e convexo. Uma função $f : M \to \mathbb{R}$ é dita ser Lipschitz em Ω se existe uma constante $L := L(\Omega) \geq 0$ tal que*

$$|f(p) - f(q)| \leq Ld(p,q), \qquad p, q \in \Omega. \tag{2.3}$$

Além disso, se é estabelecido que para todo $p_0 \in \Omega$ existe $L(p_0) \geq 0$ e $\delta = \delta(p_0) > 0$ tal que a desigualdade (2.3) ocorre com $L = L(p_0)$ para todo $p, q \in B_\delta(p_0) := \{p \in \Omega : d(p, p_0) < \delta\}$, então f é chamada localmente Lipschitz em Ω.

Observação 2.2 *Como uma conseqüência imediata da desigualdade triangular, obtemos que $|d(p,p_0) - d(q,p_0)| \leq d(p,q)$ para todo p, q e $p_0 \in M$. Então, da Definição 2.2, adquirimos que a função distância Riemanniana a um ponto fixo, $d(\cdot, q)$ é Lipschitz e portanto localmente Lipschitz. De fato, conhecemos bem que cada função convexa é localmente Lipschitz e conseqüentemente contínua. Ver [33].*

Proposição 2.6 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, $f : M \to \mathbb{R}$ e $p \in M$. Se existe $\lambda > 0$ tal que $f + (\lambda/2)\, d^2(.\,, p) : M \to \mathbb{R}$ é convexa em Ω, então f é localmente Lipschitz em Ω.*

Prova. Visto que $f + (\lambda/2)\, d^2(.\,, p)$ é convexa, segue da Observação 2.2 que para qualquer $\tilde{p} \in \Omega$ existe $L_1, \delta_1 > 0$ tal que

$$\big|[f(q_1) + (\lambda/2)\, d^2(q_1\,, p)] - [f(q_2) + (\lambda/2)\, d^2(q_2\,, p)]\big| \leq L_1 d(q_1, q_2), \quad q_1, q_2 \in B(\tilde{p}, \delta_1). \tag{2.4}$$

Além disso, Proposição 2.4 junto com Observação 2.2 implicam que existem $L_2, \delta_2 > 0$ tal que

$$|(1/2)d^2(q_1\,, p) - (1/2)d^2(q_2\,, p)| \leq L_2 d(q_1, q_2), \quad q_1, q_2 \in B(\tilde{p}, \delta_1). \tag{2.5}$$

Simples manipulações algébricas nos permite obter que

$$|f(q_1) - f(q_2)| \leq \left|[f(q_1) + \lambda/2)\, d^2(q_1\,, p)] - [f(q_2) + (\lambda/2)\, d^2(q_2\,, p)]\right| + \\ + \left|(\lambda/2)\, d^2(q_2\,, p) - (\lambda/2)\, d^2(q_1\,, p)\right|.$$

Portanto, tomando $\delta = \min\{\delta_1, \delta_2\}$ e usando as desigualdades (2.4) e (2.5), concluímos da última desigualdade que

$$|f(q_1) - f(q_2)| \leq (L_1 + \lambda L_2) d(q_1, q_2), \qquad q_1, q_2 \in B(\tilde{p}, \delta),$$

e a prova está terminada. **c.q.d**

A seguinte definição tem aparecido em [36].

Definição 2.3 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $f : M \to \mathbb{R}$ uma função continuamente diferenciável em Ω. O campo gradiente* grad f *é dito ser Lipschitz com constante $\Gamma \geq 0$ em Ω sempre que*

$$\| \operatorname{grad} f(q) - P_{pq} \operatorname{grad} f(p) \| \leq \Gamma d(p, q), \qquad p, q \in \Omega,$$

onde P_{pq} é o transporte paralelo ao longo do segmento geodésico ligando p a q.

Proposição 2.7 *Seja $\Omega \subset M$ um conjunto aberto e convexo e $f : M \to \mathbb{R}$ uma função duas vezes continuamente diferenciável em Ω. Se Hess $_p f$ é limitado em Ω, então o campo de vetor gradiente* grad f *é Lipschitz em Ω.*

Prova. A prova é uma consequência imediata do teorema fundamental do cálculo para campos de vetores. Ver, por exemplo, [5]. **c.q.d**

Capítulo 3

Derivada direcional

Neste capítulo apresentamos algumas propriedades da derivada direcional de uma função convexa definida sobre uma variedade de Hadamard, incluindo uma caracterização da derivada direcional e do subdiferencial do máximo de uma certa coleção de funções convexas. Apresentamos também uma definição para a derivada direcional (respectivamente, subdiferencial) generalizada de uma função localmente Lipschitz (não necessariamente convexa) que, no caso Euclideano, coincide com a derivada direcional (respectivamente, subdiferencial) generalizada de Clarke. Além disso, algumas propriedades desses conceitos são apresentadas, entre elas, a semicontinuidade superior da derivada direcional, uma relação entre o subdiferencial da soma de duas funções localmente Lipschitz (no caso particular que uma delas é diferenciável) e seus subdiferenciais, e uma importante propriedade do subdiferencial do máximo de uma coleção finita de funções continuamente diferenciáveis.

3.1 Derivada direcional de funções convexas

Nesta seção apresentamos a definição de derivada direcional de uma função convexa em variedade de Hadamard e algumas propriedades envolvendo seu subdiferencial (ver, por exemplo, [33]), as quais nos permitiram obter uma importante caracterização para o subdiferencial do máximo de uma certa coleção de funções convexas.

Sejam $\Omega \subset M$ aberto e convexo, e $f : M \to \mathbb{R}$ uma função convexa em Ω. Tome $p \in \Omega$, $v \in T_pM$ e $\delta > 0$ e seja $\gamma : [-\delta\,,\,\delta] \to \Omega$ o segmento geodésico tal que $\gamma(0) = p$ e $\gamma'(0) = v$. Devido à convexidade de $f \circ \gamma : [-\delta\,,\,\delta] \to \mathbb{R}$, temos que a função $q_\gamma : (0\,,\, \delta] \to \mathbb{R}$, dada por

$$q_\gamma(t) := \frac{f(\gamma(t)) - f(p)}{t},$$

é não decrescente. Além disso, visto que f é localmente Lipschitziana, segue que q_γ é limitada próximo de zero. Então, a seguinte definição faz sentido.

Definição 3.1 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $f : M \to \mathbb{R}$ uma função convexa em Ω. Então, a derivada direcional de f em $p \in \Omega$ na direção de $v \in T_pM$ é definida por*

$$f'(p, v) := \lim_{t\to 0^+} q_\gamma(t) = \inf_{t>0} q_\gamma(t), \tag{3.1}$$

onde $\delta > 0$ e $\gamma : [-\delta\,,\, \delta] \to \Omega$ é o segmento geodésico tal que $\gamma(0) = p$ e $\gamma'(0) = v$.

Proposição 3.1 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $f : M \to \mathbb{R}$ uma função convexa em Ω. Então, para cada ponto $p \in \Omega$ fixado, o subdiferencial de f em p, $\partial f(p)$, é convexo.*

Prova. Ver Teorema 4.6 de [33], p. 74. **c.q.d**

Proposição 3.2 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $f : M \to \mathbb{R}$ uma função convexa em Ω. Então, para cada ponto $p \in \Omega$ fixado,*

i) $f'(p, v) = \max_{s\in\partial f(p)}\langle s, v\rangle$, para todo $v \in T_pM$;

ii) $\partial f(p) = \{s \in T_pM : f'(p, v) \geq \langle s, v\rangle, v \in T_pM\}$.

Prova Ver Proposição 3.2 de [42]. **c.q.d**

O seguinte resultado estende, ao contexto Riemanniano, o Teorema 3.4.14 de [40].

Proposição 3.3 *Sejam T um conjunto compacto qualquer, $\Omega \subset M$ um conjunto aberto e convexo, e $h : M \times T \to \mathbb{R}$ uma função contínua em $\Omega \times T$ tal que $h(.\,, \tau) : M \to \mathbb{R}$ é convexa em Ω para todo $\tau \in T$. Se $f : M \to \mathbb{R}$ é dada por $f(p) = \max_{\tau\in T} h(p, \tau)$, então f é convexa em Ω e*

$$f'(p, v) = \max_{\tau\in T(p)} h'(p, \tau, v), \qquad p \in \Omega, \quad v \in T_pM,$$

onde $T(p) = \{\tau \in T : f(p) = h(p, \tau)\}$. Além disso, se $h(., \tau)$ é diferenciável em Ω para todo $\tau \in T$ e $\operatorname{grad}_p h(p, .)$ é contínua para todo $p \in \Omega$, então

$$\partial f(p) = \operatorname{conv}\left\{\operatorname{grad}_p h(p, \tau) : \tau \in T(p)\right\}.$$

Prova. Visto que T é compacto, f está bem definida. Sua convexidade segue da Proposition 2.5. Agora, tome $p \in \Omega$, $v \in T_pM$ e o segmento geodésico $\gamma : [-\delta, \delta] \to \Omega$, $\delta > 0$, tal que $\gamma(0) = p$ e $\gamma'(0) = v$. Usando que T é compacto, temos que $T(p) \neq \emptyset$. Portanto, tomando $\tau \in T(p)$, da definição de f e $T(p)$ obtemos

$$\frac{f(\gamma(t)) - f(p)}{t} \geq \frac{h(\gamma(t), \tau) - h(p, \tau)}{t}, \qquad t \in (0, \delta).$$

Dado que f e $h(., \tau)$ são convexas, fazendo t tender a 0, a desigualdade acima torna-se

$$f'(p, v) \geq h'(p, \tau, v), \qquad p \in \Omega,\ v \in T_pM,\ \ \tau \in T(p).$$

Portanto,

$$f'(p, v) \geq \sup_{\tau \in T(p)} h'(p, \tau, v), \qquad p \in \Omega,\ v \in T_pM. \tag{3.2}$$

Agora iremos provar a igualdade na equação acima. Seja $\{t_k\} \subset (0, \delta)$ tal que t_k converge a 0 quando k tendo a $+\infty$. Definamos

$$p^k := \gamma(t_k), \qquad \tau_k \in T(p^k). \tag{3.3}$$

A última equação implica que $f(\gamma(t_k)) = h(\gamma(t_k), \tau_k)$. Portanto, como f é convexa, combinando (3.1) e definição de f, obtemos

$$f'(p, v) \leq \frac{f(\gamma(t_k)) - f(p)}{t_k} \leq \frac{h(\gamma(t_k), \tau_k) - h(p, \tau_k)}{t_k}.$$

Visto que $\{\tau_k\} \subset T$ e T é compacto, podemos supor (tomando uma subseqüência, se necessário) que a seqüência converge a $\bar{\tau} \in T$ quando k vai a $+\infty$. Assim, fazendo k tender a $+\infty$ na última desigualdade, temos

$$f'(p, v) \leq h'(p, \bar{\tau}, v),$$

dado que h é contínua e $x \mapsto h(x, \tau)$ é convexa para cada $\tau \in T$. Note que, se $\bar{\tau} \in T(p)$, então a última desigualdade implica que (3.2) ocorre com igualdade. Assim, para concluirmos a primeira parte, é suficiente provar que $\bar{\tau} \in T(p)$. Primeiro note que, usando (3.3), definições de $T(p^k)$ e f, concluímos que

$$h(p^k, \tau_k) = \max_{\tau \in T} h(p^k, \tau) \geq h(p^k, \tau), \qquad \tau \in T.$$

Portanto, fazendo k tender a $+\infty$, temos que $h(p, \bar{\tau}) \geq h(p, \tau)$, para todo $\tau \in T$, que, junto com a definição de f, dá $f(p) = h(p, \bar{\tau})$, o que conclui a primeira parte.

Para provar a segunda parte, tome $p \in \Omega$ e $\tau \in T(p)$. Da Proposição 2.1, a convexidade de $h(.,\tau)$ implica que

$$h(q,\tau) \geq h(p,\tau) + \langle \operatorname{grad}_p h(p,\tau), \exp_p^{-1} q \rangle, \qquad q \in \Omega.$$

Visto que $\tau \in T(p)$, $h(p,\tau) = f(p)$, o que junto com definição de f e última equação torna

$$f(q) \geq f(p) + \langle \operatorname{grad}_p h(p,\tau), \exp_p^{-1} q \rangle.$$

Assim, $\operatorname{grad}_p h(p,\tau) \in \partial f(p)$. Como Proposição 3.1 implica que $\partial f(p)$ é convexa, concluímos que

$$\operatorname{conv}\{\operatorname{grad}_p h(p,\tau) : \tau \in T(p)\} \subseteq \partial f(p).$$

Afirmamos que a inclusão acima ocorre com igualdade. De fato, assuma, por contradição, que

$$\exists\, y \in \partial f(p), \quad y \notin \operatorname{conv}\{\operatorname{grad}_p h(p,\tau) : \tau \in T(p)\}.$$

Devido o fato que $\operatorname{grad}_p h(p,.)$ é contínua e $T(p)$ é um conjunto compacto, concluímos que o conjunto $\operatorname{conv}\{\operatorname{grad}_p h(p,\tau) : \tau \in T(p)\}$ é compacto. Assim, pelo Teorema de separação em T_pM, existe $v \in T_pM - \{0\}$ e $a \in \mathbb{R}$ tal que

$$\langle y, v \rangle > a > \langle \operatorname{grad}_p h(p,\tau), v \rangle, \quad \forall\, \tau \in T(p).$$

Como $h'(p,\tau,v) = \langle \operatorname{grad}_p h(p,\tau), v) \rangle$, segue da última desigualdade e primeira parte da proposição que

$$\langle y, v \rangle > \max_{\tau \in T(p)} h'(p,\tau,v) = f'(p,v).$$

Visto que $y \in \partial f(p)$, obtemos uma contradição com a Proposição 3.2 i. Portanto a afirmação está provada, o que conclui a prova. **c.q.d**

Corolário 3.1 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, e $h_i : M \to \mathbb{R}$ uma função convexa diferenciável em Ω para $i = 1, ..., m$. Se $h : M \to \mathbb{R}$ é definida por $h(p) := \max_{i \in I} h_i(p)$, onde $I = \{1, ..., m\}$, então*

$$\partial h(p) = \operatorname{conv}\{\operatorname{grad} h_i : i \in I(p)\} =$$
$$\left\{ y \in T_pM : y = \sum_{i \in I(p)} \alpha_i \operatorname{grad} h_i(p), \sum_{i \in I(p)} \alpha_i = 1, \alpha_i \geq 0 \right\},$$

onde $I(p) := \{i : h(p) = h_i(p), i = 1, ..., m\}$. *Em particular,* p *minimiza* h *em* Ω *se, e somente se, existem* $\alpha_i \geq 0$, $i \in I(p)$, *tais que*

$$0 = \sum_{i \in I(p)} \alpha_i \operatorname{grad} h_i(p), \qquad \sum_{i \in I(p)} \alpha_i = 1.$$

Prova. Segue da Proposição 3.3. **c.q.d**

3.2 Derivada direcional de funções localmente Lipschitz

Nesta seção apresentamos definições para a derivada direcional e o subdiferencial generalizado de uma função localmente Lipschitz (não necessariamente convexa). Além disso, algumas propriedades desses conceitos são apresentadas, entre elas, a semicontinuidade superior da derivada direcional, uma relação entre o subdiferencial da soma de duas funções localmente Lipschitz (no caso particular que uma delas é diferenciável) e seus subdiferenciais, e uma importante propriedade do subdiferencial do máximo de uma coleção finita de funções continuamente diferenciáveis.

Definição 3.2 *Sejam* $\Omega \subset M$ *um conjunto aberto, e* $f : M \to \mathbb{R}$ *uma função localmente Lipschitz em* Ω. *A derivada direcional generalizada* $f^\circ : T\Omega \to \mathbb{R}$ *de* f *é definida por*

$$f^\circ(p, v) := \limsup_{t \downarrow 0 \ q \to p} \frac{f\left(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v\right) - f(q)}{t}, \tag{3.4}$$

onde $(D\exp_p)_{\exp_p^{-1} q}$ *denota a diferencial de* $\exp_p$ *em* $\exp_p^{-1} q$.

É válido ressaltar que uma definição equivalente tem aparecido em [3].

Observação 3.1 *A derivada direcional generalizada está bem definida. De fato, seja* $L_p > 0$ *a constante de Lipschitz de* f *em* p *e* $\delta = \delta(p) > 0$ *tal que*

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \leq L_p\, d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v,\, q),$$

para todo $q \in B_\delta(p)$ *e* $t \in [0, \delta)$. *Dado que*

$$d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v,\, q) = t\|(D\exp_p)_{\exp_p^{-1} q} v\|,$$

a desigualdade acima torna-se

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q}v) - f(q)| \leq L_p\, t\|(D\exp_p)_{\exp_p^{-1} q}v\|, \quad q \in B_\delta(p), \quad t \in [0,\delta).$$

Visto que $\lim_{q\to p} (D\exp_p)_{\exp_p^{-1} q}v = v$*, nossa afirmação segue da última desigualdade.*

Observação 3.2 *Note que, se* $M = \mathbb{R}^n$ *então* $\exp_p w = p+w$ *e* $(D\exp_p)_{\exp_p^{-1} q}v = v$*. Neste caso,* (3.4) *torna-se*

$$f_E^\circ(p,v) = \limsup_{t\downarrow 0\ q\to p} \frac{f(q+tv) - f(q)}{t},$$

que é a derivada direcional generalizada de Clarke no caso Euclideano, ver [43] p. 25. Portanto, a derivada direcional generalizada em variedade de Hadamard é uma extensão natural da derivada direcional generalizada de Clarke.

Agora iremos provar a semicontinuidade superior da derivada direcional generalizada.

Proposição 3.4 *Sejam* $\Omega \subset M$ *um conjunto aberto e* $f : M \to \mathbb{R}$ *uma função localmente Lipschitz em* Ω*. Então,* f° *é semicontínua superior em* $T\Omega$*, i.e., se* $(p,v) \in T\Omega$ *e* $\{(p^k,v^k)\}$ *é uma seqüência em* $T\Omega$ *tal que* $\lim_{k\to+\infty}(p^k,v^k) = (p,v)$*, então*

$$\limsup_{k\to+\infty} f^\circ(p^k,v^k) \leq f^\circ(p,v). \tag{3.5}$$

Prova. Sejam $(p,v) \in T\Omega$ e $\{(p^k,v^k)\} \subset T\Omega$ tal que $\lim_{k\to+\infty}(p^k,v^k) = (p,v)$. Para provar a desigualdade (3.5) primeiro note que, para cada k

$$f^\circ(p^k,v^k) \leq \limsup_{t\downarrow 0\ (q,w)\to(p^k,v^k)} \frac{f(\exp_q tw) - f(q)}{t}, \qquad (q,w) \in T\Omega.$$

Assim, pela definição de limite superior, existem $(q^k,w^k) \in T\Omega - \{(p^k,v^k)\}$ e $t_k > 0$ tais que

$$f^\circ(p^k,v^k) - \frac{1}{k} < \frac{f(\exp_{q^k} t_k w^k) - f(q^k)}{t_k}, \qquad \tilde{d}((q^k,w^k),(p^k,v^k)) + t_k < \frac{1}{k}, \tag{3.6}$$

com $\tilde{d}$ sendo a distância Riemanniana em TM. Seja $U_p \subset \Omega$ uma vizinhança de p tal que $TU_p \approx U_p \times \mathbb{R}^n$, f é Lipschitz em U_p com constante L_p and exp é Lipschitz em TU_p com constante K. Da primeira desigualdade em (3.6), obtemos

$$f^\circ(p^k,v^k) - \frac{1}{k} < \frac{f\left(\exp_{q^k} t_k(D\exp_p)_{\exp_p^{-1} q^k}v\right) - f(q^k)}{t_k} + \frac{f(\exp_{q^k} t_k w^k) - f\left(\exp_{q^k} t_k(D\exp_p)_{\exp_p^{-1} q^k}v\right)}{t_k}. \tag{3.7}$$

Por outro lado, como $\lim_{k\to+\infty}(p^k, v^k) = (p, v)$, concluímos, da segunda desigualdade em (3.6) que

$$\exp_{q^k} t_k w^k \in U_p, \qquad \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v) \in U_p, \qquad k > k_0,$$

para k_0 suficientemente grande. Assim, como f é Lipschitz em U_p, para $k > k_0$, temos

$$\left| f(\exp_{q^k} t_k w^k) - f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right)\right| \leq \\ L_p\, d\left(\exp_{q^k} t_k w^k,\, \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right). \quad (3.8)$$

Agora, levando em conta que exp é Lipschitz em TU_p, no caso particular que $k > k_0$

$$d\left(\exp_{q^k} t_k w^k,\, \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) \leq k\| t_k w^k - t_k (D\exp_p)_{\exp_p^{-1} q^k} v\|.$$

Visto que $\lim_{k\to+\infty} p^k = p$, a segunda equação em (3.6) implica que $\lim_{k\to+\infty} q^k = p$. Consequentemente $\lim_{k\to+\infty} (D\exp_p)_{\exp_p^{-1} q_k} v = v$ o que, junto com última desigualdade, implica

$$\lim_{k\to+\infty} d\left(\exp_{q^k} t_k w^k,\, \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right)/t_k = 0.$$

Portanto, combinando a última equação, (3.7), (3.8) e Definição 3.2, o resultado segue. **c.q.d**

A seguir generalizamos a definição de subdiferencial para funções localmente Lipschitz definidas em variedade de Hadamard, ver Proposição 3.2 item ii.

Definição 3.3 *Sejam $\Omega \subset M$ um conjunto aberto e $f : M \to \mathbb{R}$ uma função localmente Lipschitz em Ω. O subdiferencial generalizado de f em $p \in \Omega$, denotado por $\partial^\circ f(p)$, é definido por*

$$\partial^\circ f(p) := \left\{ w \in T_pM : f^\circ(p, v) \geq \langle w, v\rangle, \forall\, v \in T_pM \right\}.$$

Observação 3.3 *Se Ω é convexo e a função $f : M \to \mathbb{R}$ é convexa em Ω, então $f^\circ(p, v) = f'(p, v)$ (respectivamente, $\partial^\circ f(p) = \partial f(p)$) para todo $p \in \Omega$, i.e., a derivada direcional (respectivamente, subdiferencial) para funções Lipschitz é uma generalização da derivada direcional (respectivamente, subdiferencial) para funções convexas. Ver [3] Afirmação* 5.4 *na prova do Teorema* 5.3.

Definição 3.4 *Seja $f : M \to \mathbb{R}$ uma função localmente Lipschitz. Um ponto $p \in \Omega$ é dito ser um ponto estacionário de f sempre que $0 \in \partial^\circ f(p)$.*

Lema 3.1 *Seja $\Omega \subset M$ um conjunto aberto. Se $f : M \to \mathbb{R}$ é localmente Lipschitz em Ω e $g : M \to \mathbb{R}$ é convexa em Ω, então*

$$(f+g)^\circ(p,v) = f^\circ(p,v) + g'(p,v) \qquad p \in \Omega, \quad v \in T_pM. \tag{3.9}$$

Além disso, se em adição g é também diferenciável, temos

$$\partial^\circ(f+g)(p) = \partial^\circ f(p) + \operatorname{grad} g(p), \qquad p \in \Omega. \tag{3.10}$$

Prova. Usando a definição da derivada direcional generalizada e simples manipulações algébricas, obtemos

$$(f+g)^\circ(p,v) = \limsup_{t\downarrow 0 \;\, q\to p} \left[\frac{f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)}{t} + \frac{g(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - g(q)}{t} \right].$$

Propriedade básica do limite superior junto com a definição da derivada direcional generalizada e Observação 3.3, segue que

$$(f+g)^\circ(p,v) \le f^\circ(p,v) + g'(p,v). \tag{3.11}$$

Agora, como $f^\circ(p,v) = ((f+g)+(-g))^\circ(p,v)$, desigualdade acima implica, em particular

$$f^\circ(p,v) \le (f+g)^\circ(p,v) + (-g)'(p,v),$$

que é equivalente a

$$(f+g)^\circ(p,v) \ge f^\circ(p,v) + g'(p,v).$$

Assim, combinando última desigualdade com desigualdade (3.11), a igualdade (3.9) é obtida.

No caso que g é também diferenciável, em particular $g'(p,v) = \langle \operatorname{grad} g(p), v\rangle$. Portanto, a prova da igualdade (3.10) é uma consequência imediata da igualdade (3.9) e da definição do subdiferencial generalizado. **c.q.d**

Corolário 3.2 *Sejam $\Omega \subset M$ um conjunto convexo, $\tilde{p} \in M$ e $\lambda > 0$ tal que $f + (\lambda/2)\, d^2(.,\tilde{p}) : M \to \mathbb{R}$ é convexa em Ω. Se $p \in \Omega$ é um minimizador de $f + (\lambda/2)d^2(.,\tilde{p})$, então*

$$\lambda \exp_p^{-1} \tilde{p} \in \partial^\circ f(p).$$

Prova. Visto que p é um minimizador de $f + (\lambda/2)d^2(.,\tilde{p})$, obtemos

$$0 \in \partial \left(f + \frac{\lambda}{2} d^2(\,.\,,\tilde{p}) \right)(p). \tag{3.12}$$

Por outro lado, como $f+(\lambda/2)\,d^2(\,.\,,\tilde{p})$ é convexa em Ω e $(\lambda/2)\,d^2(\,.\,,\tilde{p})$ é diferenciável com $\operatorname{grad}(\lambda/2)\,d^2(q,p) = -\lambda \exp_q^{-1} p$, usando Observação 3.3, Proposição 2.6 e aplicando Lema 3.1 com $g = (\lambda/2)\,d^2(\,.\,,\tilde{p})$, temos

$$\partial \left(f + \frac{\lambda}{2} d^2(.,\tilde{p}) \right)(p) = \partial^\circ \left(f + \frac{\lambda}{2} d^2(.,\tilde{p}) \right)(p) = \partial^\circ f(p) - \lambda \exp_p^{-1} \tilde{p}. \tag{3.13}$$

Portanto, o resultado segue pela combinação de (3.12) com (3.13). **c.q.d**

Observação 3.4 *O resultado do Lema 3.1 foi utilizado apenas no caso que a função g é a função convexa e diferenciável $(\lambda/2)\,d^2(\,.\,,\tilde{p})$ com $\lambda > 0$ e $\tilde{p} \in M$. Por isso no referido lema admitimos a hipótese de convexidade sobre a função g. Provaremos logo depois que se g é continuamente diferenciável então g é localmente Lipschitz e existe $g^\circ = g'$. Neste caso, notamos que o resultado do Lemma 3.1 pode ser obtido considerando g apenas continuamente diferenciável.*

Os dois próximos teoremas além de servirem como resultados preliminares úteis na prova do Lema 3.1 no caso que g é apenas continuamente diferenciável, mostram que o sudiferencial é de fato uma generalização da derivada clássica.

Teorema 3.1 *Sejam $\Omega \subset M$ um conjunto aberto e $f : M \to \mathbb{R}$ uma função localmente Lipschitz em Ω. Se f é diferenciável em Ω então*

$$\operatorname{grad} f(p) \in \partial^\circ f(p), \qquad p \in \Omega.$$

Prova. Visto que f é diferenciável em p, em particular temos que a derivada direcional $f'(p,v)$ exite e

$$f'(p,v) = \langle \operatorname{grad} f(p), v \rangle, p \in \Omega, \qquad v \in T_pM.$$

Como $f'(p,v) \leq f^\circ(p,v)$, para todo $p \in \Omega$ e $v \in T_pM$, da última igualdade obtemos

$$f^\circ(p,v) \geq \langle \operatorname{grad} f(p), v \rangle, \qquad p \in \Omega, \qquad v \in T_pM,$$

e o resultado segue da definição do subdiferencial generalizado. **c.q.d**

Teorema 3.2 *Sejam $\Omega \subset M$ um conjunto aberto e $f : M \to \mathbb{R}$ uma função localmente Lipschitz em Ω. Se f é continuamente diferenciável em Ω então f é localmente Lipschitz em Ω e $\partial f(p) = \{\operatorname{grad} f(p)\}$, para todo $p \in \Omega$.*

Prova. Tome $p \in \Omega$. Dado que f é continuamente diferenciável em Ω, em particular, f é contiamente diferenciável em p e, neste caso, existem $\delta, M > 0$ tais que

$$\|\operatorname{grad} f(q)\| \leq M, \qquad q \in B_\delta(p). \tag{3.14}$$

Dados $p_1, p_2 \in B_\delta(p)$ e $v \in T_{p_1}M$ tal que $\exp_{p_1} v = p_2$, definamos $h : [0,1] \to \mathbb{R}$ por $h(t) = f(\exp_{p_1} tv)$. Pelo teorema do valor médio para o caso escalar existe $\tilde{t} \in (0,1)$ tal que

$$h(1) - h(0) = h'(\tilde{t}).$$

Assim, visto que $h(0) = f(p_1)$, $h(1) = f(p_2)$ e $h'(\tilde{t}) = \langle \operatorname{grad} f(\exp_{p_1} \tilde{t}v), \tilde{t}v \rangle$, da última igualdade segue que

$$f(p_2) - f(p_1) = \langle \operatorname{grad} f(\exp_{p_1} \tilde{t}v), \tilde{t}v \rangle.$$

Como $|\langle \operatorname{grad} f(\exp_{p_1} \tilde{t}v), \tilde{t}v \rangle| \leq \tilde{t}\|\operatorname{grad} f(\exp_{p_1} \tilde{t}v)\|\|v\|$, levando em conta que $\exp_{p_1} \tilde{t}v \in B_\delta(p)$, (3.14) e $\|v\| = \|\exp_{p_1}^{-1} p_2\| = d(p_2, p_1)$, da última igualdade concluímos que

$$|f(p_2) - f(p_1)| \leq \tilde{t} M d(p_2, p_1),$$

o que prova a primeira parte do teorema.

Do Teorema 3.1, para provar a segunda parte é suficiente mostrar que para qualquer $w \in \partial^\circ f(p)$, $w = \operatorname{grad} f(p)$. Para isto tome uma seqüência $\{p^j\} \subset \Omega$ convergindo a p e $v \in T_pM$. Visto que f é diferenciável e $(D\exp_p)_{\exp_p^{-1} p^j} v \in T_{p^j}M$, temos

$$f'\left(p^j, (D\exp_p)_{\exp_p^{-1} p^j} v\right) = \left\langle \operatorname{grad} f(p^j), (D\exp_p)_{\exp_p^{-1} p^j} v \right\rangle, \qquad j = 0, 1, \ldots.$$

Assim, da continuidade de $\operatorname{grad} f$ em Ω junto com o fato que $\lim_{p^j \to p} (D\exp_p)_{\exp_p^{-1} p^j} v = v$, segue que

$$\lim_{p^j \to p} f'(p^j, (D\exp_p)_{\exp_p^{-1} p^j} v) = \langle \operatorname{grad} f(p), v \rangle = f'(p, v).$$

Por outro lado,

$$\lim_{p^j \to p} f'(p^j, (D\exp_p)_{\exp_p^{-1} p^j} v) = \lim_{p^j \to p} \lim_{t \downarrow 0} \frac{f\left(\exp_{p^j} t(D\exp_p)_{\exp_p^{-1} p^j} v\right) - f(p^j)}{t},$$

e, da definição de derivada direcional generalizada junto com a última igualdade, concluímos que

$$f^{\circ}(p,v) = f'(p,v) = \langle \operatorname{grad} f(p), v\rangle.$$

Agora seja $w \in \partial^{\circ} f(p)$. Então, da defenição do subdiferencial generalizado junto com a última igualdade, temos que $\langle \operatorname{grad} f(p), v\rangle \geq \langle w, v\rangle$. Portanto, da arbitrariedade de v concluímos que $w = \operatorname{grad} f(p)$ e a prova está concluída. **c.q.d**

Proposição 3.5 *Sejam $\Omega \subset M$ um conjunto aberto e convexo, $I = \{1, ..., m\}$, $f_i : M \to \mathbb{R}$ uma função continuamente diferenciável em Ω para todo $i \in I$ e $f : M \to \mathbb{R}$ definida por*

$$f(p) := \max_{i \in I} f_i(p).$$

Então, f é localmente Lipschitz em Ω e, para cada $p \in \Omega$, ocorre

$$\operatorname{conv}\{\operatorname{grad} f_i(p) : i \in I(p)\} \subset \partial^{\circ} f(p),$$

onde $I(p) := \{i : f_i(p) = f(p),\ i = 1, ..., m\}$.

Prova. Dado que f_i é continuamente diferenciável em Ω, concluímos que f_i é localmente Lipschitz em Ω, para todo $i \in I$. Assim, para cada $\tilde{p} \in \Omega$ e $i \in I$, existem $\delta_i, L_i > 0$ tais que

$$|f_i(p) - f_i(q)| \leq L_i d(p,q), \qquad p\,, q \in B_{\delta_i}(\tilde{p}).$$

Por outro lado,

$$|\max_{i \in I} f_i(p) - \max_{i \in I} f_i(q)| \leq \max_{i \in I} |f_i(p) - f_i(q)|.$$

Combinando as duas últimas expressões com definição de f, obtemos

$$|f(p) - f(q)| \leq L d(p,q) \qquad p, q \in B_{\delta}(\tilde{p}),$$

onde $\delta = \min_{i \in I} \delta_i$ e $L = \max_{i \in I} L_i$, o que prova primeira parte.

Para provar a segunda parte, tome $p \in \Omega$, $u \in \operatorname{conv}\{\operatorname{grad} f_i(p) : i \in I(p)\}$ e $v \in T_pM$. Então existem constantes, $\alpha_i \geq 0$ para $i \in I(p)$ com $\sum_{i \in I(p)} \alpha_i = 1$, tais que

$$u = \sum_{i \in I(p)} \alpha_i \operatorname{grad} f_i(p).$$

Como f_i é diferenciável em Ω, para todo $i \in I$, simples manipulações algébricas nos permitem obter

$$\langle u, v \rangle = \sum_{i \in I(p)} \alpha_i \langle \operatorname{grad} f_i(p), v \rangle = \sum_{i \in I(p)} \alpha_i f_i'(p, v).$$

Dado que f é localmente Lipschitz em p, as definições de f, $I(p)$ e derivada direcional generalizada implicam

$$f_i'(p, v) \leq f^\circ(p, v),$$

que, junto com a última equação, nos fornece $\langle u, v \rangle \leq f^\circ(p, v)$, e a prova segue da definição de $\partial^\circ f(p)$. **c.q.d**

Capítulo 4

Método de ponto proximal local em variedades Riemannianas

Seguindo as idéias de KAPLAN E TICHATSCHKE [25], neste capítulo estendemos o campo de aplicação do método de ponto proximal para resolver o problema

$$\begin{aligned} &\min f(p) \\ &\text{s.t. } \ p \in M, \end{aligned} \tag{4.1}$$

no caso que a função objetivo é uma função (não necessariamente convexa) dada pelo máximo de uma certa coleção de funções continuamente diferenciáveis variando sobre um conjunto de índices compacto (finito ou infinito). Provamos o seguinte teorema:

Teorema 4.1 *Seja $\Omega \subset M$ um conjunto aberto e convexo, $q \in M$ e $T \subset \mathbb{R}$ um conjunto compacto. Seja $\varphi : M \times T \to \mathbb{R}$ uma função contínua em $\Omega \times T$ tal que $\varphi(., \tau) : M \to \mathbb{R}$ é continuamente diferenciável em Ω e contínua em $\bar{\Omega}$ (fecho de Ω), para todo $\tau \in T$, e $f : M \to \mathbb{R}$ definida por*

$$f(p) := \max_{\tau \in T} \varphi(p, \tau).$$

Assuma que $-\infty < \inf_{p \in M} f(p)$, $\operatorname{grad}_p \varphi(., \tau)$ é Lipschitz sobre Ω com constante L_τ para cada $\tau \in T$ tal que $\sup_{\tau \in T} L_\tau < +\infty$ e

$$L_f(f(q)) = \{p \in M : f(p) \leq f(q)\} \subset \Omega, \qquad \inf_{p \in M} f(p) < f(q).$$

Tome $0 < \bar{\lambda}$ e uma sequência $\{\lambda_k\}$ satisfazendo $\sup_{\tau \in T} L_\tau < \lambda_k \leq \bar{\lambda}$ e $\hat{p} \in L_f(f(q))$.

Então o método de ponto proximal

$$p^{k+1} := \text{argmin}_{p\in M}\left\{f(p) + \frac{\lambda_k}{2}d^2(p,p^k)\right\}, \qquad k = 0,1,\ldots, \tag{4.2}$$

com ponto inicial $p^0 = \hat{p}$, está bem definido, a sequência gerada $\{p^k\}$ permanece em $L_f(f(q))$ e satisfaz somente uma das seguintes afirmações

i) $\{p^k\}$ *é finita, i.e., $p^{k+1} = p^k$ para algum k e, neste caso, p^k é um ponto estacionário de f,*

ii) $\{p^k\}$ *é infinito e, neste caso, qualquer ponto de acumulação de $\{p^k\}$ é um ponto estacionário de f.*

Além disso, assuma que o conjunto dos minimizadores de f é não vazio, i. e.,

h1) $U^* = \{p : f(p) = \inf_{\tilde{p}\in M} f(\tilde{p})\} \neq \emptyset$.

Seja $c \in (\inf_{p\in M} f(p),\ f(q))$. Se adicionalmente as seguintes hipóteses ocorrem:

h2) *$L_f(c)$ é convexo e f é convexa sobre $L_f(c)$;*

h3) *Para todo $p \in L_f(f(q)) \setminus L_f(c)$ e $y(p) \in \partial^\circ f(p)$ temos $\|y(p)\| > \delta > 0$,*

então a sequência $\{p^k\}$, gerada por (4.2) com

$$\sup_{\tau\in T} L_\tau < \lambda_k \leq \bar{\lambda}, \qquad k = 0,1,\ldots \tag{4.3}$$

converge para um ponto $p^* \in U^*$.

Observação 4.1 *A continuidade de cada função $\varphi(.,\tau)$ sobre $\bar{\Omega}$ garante que os conjuntos de nível da função f, em particular, o conjunto solução U^*, são fechados na topologia da variedade M.*

Na próxima observação mostramos que se Ω é limitado e $\varphi(.,\tau)$ é convexa sobre Ω então f satisfaz as hipóteses **h2** e **h3**.

Observação 4.2 *Se $\varphi(.,\tau)$ é uma função convexa em Ω e contínua em $\bar{\Omega}$ para todo $\tau \in T$ então, pela Proposição 2.5, a função f é convexa em Ω e a hipótese* **h2** *é satisfeita para todo $c \leq f(q)$. Além disso, da Observação 3.3,*

$$\partial^\circ f(p) = \partial f(p), \qquad p \in \Omega. \tag{4.4}$$

Tome $c \in (\inf_{p\in M} f(p),\ f(q))$*, suponha que* **h1** *ocorre e* Ω *é limitado. Então, temos*

$$0 < \sup\{d(p^*, p) : p^* \in U^*, p \in L_f(f(q)) \setminus L_f(c)\} = \epsilon < +\infty. \qquad (4.5)$$

Seja $p^* \in U^*$ *fixado,* $p \in L_f(f(q)) \setminus L_f(c)$ *e* $y(p) \in \partial f(p)$*. A convexidade de* f *em* Ω *implica que*

$$\langle y(p)\ ,\ -\exp_p^{-1} p^*\rangle \geq f(p) - f(p^*).$$

Visto que $\|y(p)\|\|\exp_p^{-1} p^*\| \geq \langle y(p), -\exp_p^{-1} p^*\rangle$*,* $d(p^*, p) = \|\exp_p^{-1} p^*\|$*,* $p \in L_f(f(q)) \setminus L_f(c)$ *e* U^* *é um subconjunto próprio de* $L_f(c)$*, da desigualdade acima obtemos*

$$\|y(p)\| d(p^*, p) > c - f(p^*) > 0.$$

Assim, de (4.5) *junto com última desigualdade,* $\|y(p)\|\epsilon > c - f(p^*) > 0$*. Portanto, fixando* $\delta = (c - f(p^*))/\epsilon$*, temos*

$$\|y(p)\| > \delta > 0,$$

que junto com (4.4) *mostra que* f *satisfaz* **h3***.*

Para provar o teorema acima precisamos de alguns resultados preliminares. De agora em diante assumimos que cada uma das hipóteses no Teorema 4.1 ocorre, com exceção de **h1**, **h2** e **h3**, que serão assumidas válidas somente quando explicitamente mencionado.

Lema 4.1 *Para todo* $\tilde{p} \in M$ *e* λ *satisfazendo*

$$\sup_{\tau\in T} L_\tau < \lambda,$$

a função $\varphi(.,\tau) + (\lambda/2)d^2(.\,,\tilde{p})$*, é fortemente convexa em* Ω *com constante* $\lambda - \sup_{\tau\in T} L_\tau$*. Como uma consequência,* $f + (\lambda/2)d^2(.\,,\tilde{p})$ *é fortemente convexa em* Ω *com constante* $\lambda - \sup_{\tau\in T} L_\tau$*.*

Prova. Visto que T é compacto e φ é contínua, a função f está bem definida. Tome $\tau \in T$, $\tilde{p} \in \Omega$ e defina $\psi_\tau := \varphi_\tau + (\lambda/2)d^2(.\,,\tilde{p})$, onde $\varphi_\tau := \varphi(.,\tau)$. Note que $\operatorname{grad}\psi_\tau(p) = \operatorname{grad}\varphi_\tau(p) - \lambda \exp_p^{-1}\tilde{p}$. Assim, para todo $p, q \in \Omega$

$$\begin{aligned}\langle P_{qp}^{-1} \operatorname{grad}\psi_\tau(p) - \operatorname{grad}\psi_\tau(q), \exp_q^{-1} p\rangle &= \langle P_{qp}^{-1} \operatorname{grad}\varphi_\tau(p) - \operatorname{grad}\varphi_\tau(q), \exp_q^{-1} p\rangle \\ &\quad - \lambda\langle P_{qp}^{-1} \exp_p^{-1}\tilde{p} - \exp_q^{-1}\tilde{p}, \exp_q^{-1} p\rangle.\end{aligned}$$

Como

$$\langle P_{qp}^{-1}\operatorname{grad}\varphi_\tau(p)-\operatorname{grad}\varphi_\tau(q),\exp_q^{-1}p\rangle \geq -\|P_{qp}^{-1}\operatorname{grad}\varphi_\tau(p)-\operatorname{grad}\varphi_\tau(q)\|\|\exp_q^{-1}p\|,$$

usando a igualdade $d(q,p)=\|\exp_q^{-1}p\|$, Proposição 2.2 e igualdade acima, obtemos

$$\begin{aligned}\langle P_{qp}^{-1}\operatorname{grad}\psi_\tau(p)-\operatorname{grad}\psi_\tau(q),\exp_q^{-1}p\rangle \geq \\ -\|P_{qp}^{-1}\operatorname{grad}\varphi_\tau(p)-\operatorname{grad}\varphi_\tau(q)\|d(q,p)+\lambda d^2(q,p).\end{aligned}$$

Agora, como $\operatorname{grad}\varphi_\tau$ é Lipschitz em Ω com constante L_τ e o transporte paralelo é uma isometria, última desigualdade torna-se

$$\langle P_{qp}^{-1}\operatorname{grad}\psi_\tau(p)-\operatorname{grad}\psi_\tau(q),\exp_q^{-1}p\rangle \geq (\lambda-L_\tau)d^2(q,p).$$

Por hipótese, $\lambda>\sup_{\tau\in I}L_\tau$. Assim, a desigualdade acima e a Definição 2.1 implicam que $\operatorname{grad}\psi_\tau$ é fortemente monótono com constante $\lambda-\sup_{\tau\in I}L_\tau$. Com isso, da Proposição 2.3, concluímos que ψ_τ é fortemente monótono com constante $\lambda-\sup_{\tau\in I}L_\tau$. É fácil ver que

$$\max_{\tau\in I}\psi_\tau=f+(\lambda/2)d^2(.\,,\tilde{p}).$$

Portanto, usando Proposição 2.5, a afirmação da proposição segue. **c.q.d**

Corolário 4.1 *O método de ponto proximal* (4.2) *aplicado a f com ponto inicial $p^0=\hat{p}$ está bem definido.*

Prova. Assuma que $p^k\in L_f(f(q))$ para algum k. Note que os minimizadores de $\psi_k:=f+(\lambda_k/2)d^2(.,p^k)$, caso existam, estão em $L_{\psi_k}(\psi_k(p^k))\subset L_f(f(q))\subset\Omega$, mais precisamente,

$$\operatorname{argmin}_{p\in M}\psi_k(p)=\operatorname{argmin}_{p\in L_{\psi_k}(\psi_k(p^k))}\psi_k(p)).$$

Como f é contínua em $\bar{\Omega}$, $L_{\psi_k}(\psi_k(p^k))$ é fechado com respeito à topologia da variedade M. Além disso, visto que $\sup_{\tau\in T}L_\tau<\lambda_k$, concluímos do Lema 4.1 que a aplicação ψ_k é fortemente convexa em Ω com constante $\beta=\lambda-\sup_{\tau\in T}L_\tau$. Nestas condições, afim de que ψ_k tenha um único minimizador em M é suficiente mostrarmos que $L_{\psi_k}(\psi_k(p^k))$ é um conjunto limitado. Da convexidade ψ_k em Ω temos, em particular, que $L_{\psi_k}(\psi_k(p^k))$ é um conjunto convexo. Suponhamos por absurdo que $L_{\psi_k}(\psi_k(p^k))$ é ilimitado. Então, existe $v\in T_{p^k}M$ e uma geodésica

$\gamma : [0, +\infty] \to L_{\psi_k}(\psi_k(p^k))$ tal que $\gamma(0) = p^k$ e $\gamma'(0) = v$. Fixemos $t_0 > 0$. Para todo $t > t_0$, temos que

$$\psi_k(\gamma(t_0)) = \psi_k(\gamma(\frac{t_0}{t}t + (1 - \frac{t_0}{t})0)).$$

Agora, levando em conta que ψ_k é fortemente convexa com constante β, da última igualdade decorre que

$$\psi_k(\gamma(t_0)) \leq \frac{t_0}{t}\psi_k(\gamma(t)) + (1 - \frac{t_0}{t})\psi_k(\gamma(0)) - \frac{1}{2}(1 - \frac{t_0}{t})\frac{t_0}{t}\beta\|v\|^2 t^2.$$

Como, $\psi_k(\gamma(t)), \psi_k(\gamma(0)) \in L_{\psi_k}(\psi_k(p^k)$, da última desigualdade obtemos

$$\psi_k(\gamma(t_0)) \leq \psi_k(p^k) - \frac{1}{2}(t - t_0)t_0\beta\|v\|^2,$$

que é um absurdo, pois $\psi_k(\gamma(t_0))$ é finito e $\psi_k(p^k) - \frac{1}{2}(t - t_0)t_0\beta\|v\|^2$ tende a $-\infty$ quando t vai para $+\infty$. Portanto $L_{\psi_k}(\psi_k(p^k))$ é limitado e, consequentemente, p^{k+1} está bem definido. Dado que $p^0 = \hat{p} \in L_f(f(q))$, a prova segue de um simples argumento de indução. **c.q.d**

Lema 4.2 *Seja* $\{p^k\}$ *a seqüência gerada pelo método de ponto proximal* (4). *Então,*

i) $f(p^{k+1}) + (\lambda_k/2)d^2(p^{k+1}, p^k) \leq f(p^k), \quad k = 0, 1, \ldots;$

ii) $\{p^k\} \subset L_f(f(q));$

iii) $0 \in \partial\left(f + \frac{\lambda_k}{2}d^2(.\,, p^k)\right)(p^{k+1}), \quad k = 0, 1, \ldots;$

iv) $-\infty < \bar{f} = \lim_{k\to\infty} f(p^k);$

v) $\lim_{k\to\infty} d(p^{k+1}, p^k) = 0.$

Prova. O item i é uma conseqüência imediata de (4.2), o que implica que $\{f(p^k)\}$ é monótona não crescente e, portanto, que o item ii ocorre. Visto que $\sup_{\tau\in T} L_\tau < \lambda_k$, Lema 4.1 implica que $f + (\lambda_k/2)d^2(.\,, p^k)$ é convexa em Ω, que, junto com (4.2) prova item iii. Novamente usando que $\{f(p^k)\}$ é uma seqüência monótona não crescente e que $-\infty < \inf_{p\in M} f(p)$, item iv segue. Finalmente, o item v é conseqüência dos items i e iv. **c.q.d**

Lema 4.3 *Seja* $\{p^k\}$ *a seqüência gerada pelo método de ponto proximal* (4.2), *com* λ_k *satisfazendo* (4.3). *Assuma que* **h1** *e* **h2** *ocorrem. Se* $p^k \in L_f(c)$ *para algum* k *então* $\{p^k\}$ *converge a um ponto* $p^* \in U^* \subset \Omega$.

Prova. Por hipótese, $p^k \in L_f(c)$ para algum k, i.e., existe k_0 tal que $f(p^{k_0}) \leq c$. Então, do Lemma 4.2 item i, $\{p^k\} \subset L_f(c)$ para todo $k \geq k_0$. Por outro lado, de (4.3), temos

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

Portanto, usando **h1** e **h2**, o resultado segue de argumentos similares usados na prova do Teorema 6.1 of [18]. **c.q.d**

Lema 4.4 *Seja* $\{p^k\}$ *a seqüência gerada pelo método de ponto proximal* (4.2) *com* λ_k *satisfazendo* (4.3). *Se* **h3** *ocorre, então depois de um número finito de passos as iteradas do método entram no conjunto* $L_f(c)$.

Prova. Primeiro note que, de $\inf_{p\in M} f(p) < c$, temos $L_f(c) \neq \emptyset$. Suponha por absurdo que $p^k \in L_f(f(q)) \setminus L_f(c)$ para todo k. De (4.3), temos que $\sup_{\tau\in T} L_\tau < \lambda_k$. Além disso, do Lema 4.1 com $\lambda = \lambda_k$ e $\tilde{p} = p^k$, segue que

$$f + (\lambda_k/2)\, d^2(., p^k)$$

é convexa em Ω. Visto que o ponto $p^{k+1} \in \Omega$ (ver item ii do Lema 4.2) e, de (4.2), o mesmo é um minimizador da função acima, aplicando Corolário 3.2 com $p = p^{k+1}$, obtemos

$$\lambda_k \exp^{-1}_{p^{k+1}} p^k \in \partial^\circ f(p^{k+1}), \qquad k = 0, 1, \ldots.$$

Dado que $p^{k+1} \in L_f(f(q)) \setminus L_f(c)$, hipótese **h3** e inclusão acima acarretam

$$\|\lambda_k \exp^{-1}_{p^{k+1}} p^k\| > \delta, \qquad k = 0, 1, \ldots.$$

De (4.3) temos $\lambda_k \leq \bar{\lambda}$. Como $d(p^k, p^{k+1}) = \|\exp^{-1}_{p^{k+1}} p^k\|$, a última desigualdade implica que

$$d(p^k, p^{k+1}) > \frac{\delta}{\bar{\lambda}}, \qquad k = 0, 1, \ldots.$$

Assim, do item v do Lema 4.2, chegamos em uma contradição. Portanto existe k_0 tal que $f(p^{k_0}) \leq c$ e o resultado segue do item i do Lema 4.2. **c.q.d**

Prova do Teorema 4.1

A boa definição do método de ponto proximal (4.2) é uma conseqüência do Corolário 4.1. Seja $\{p^k\}$ a seqüência gerada pelo método (4.2). Dado que $p^0 =$

$\hat{p} \in L_f(f(q))$, item ii do Lema 4.2 nos diz que toda a seqüência está em $L_f(f(q))$. Do item iii do Lema 4.2, temos

$$0 \in \partial\left(f + \frac{\lambda_k}{2} d^2(\,.\,,p^k)\right)(p^{k+1}), \quad k = 0, 1, \ldots.$$

Visto que $\sup_{\tau \in T} L_\tau < \lambda_k$, Lema 4.1 implica que $f + (\lambda_k/2)d^2(\,.\,,p^k)$ é fortemente convexa em Ω, o que, junto com a Proposição 2.6, nos dá que f é localmente Lipschitz em Ω. Assim, usando a definição de p^{k+1}, concluímos, do Corolário 3.2 com $\lambda = \lambda_k$, $\tilde{p} = p^k$ e $p = p^{k+1}$, que

$$\lambda_k \exp^{-1}_{p^{k+1}} p^k \in \partial^\circ f(p^{k+1}). \tag{4.6}$$

Se $\{p^k\}$ é uma seqüência finita, então $p^{k+1} = p^k$ para algum k e a última inclusão implica que $0 \in \partial^\circ f(p^{k+1})$, i.e., p^k é um ponto estacionário de f. Agora assuma que $\{p^k\}$ é uma seqüência infinita. Se $\bar{p}$ é um ponto de acumulação de $\{p^k\}$, então existe uma subseqüência $\{p^{k_s}\}$ de $\{p^k\}$ tal que $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$, e item v do Lemma 4.2 implica que

$$\lim_{s\to\infty} \| \exp^{-1}_{p^{k_s+1}} p^{k_s} \| = \lim_{s\to\infty} d(p^{k_s+1}, p^{k_s}) = 0. \tag{4.7}$$

Agora, da relação (4.6), temos que

$$f^\circ(p^{k_s+1}, v) \geq \lambda_{k_s} \langle \exp^{-1}_{p^{k_s+1}} p^{k_s}, v \rangle, \qquad \forall\, v \in T_{p^{k_s+1}}M.$$

Seja $\bar{v} \in T_{\bar{p}}M$. Portanto, da última desigualdade, obtemos

$$f^\circ(p^{k_s+1}, v^{k_s+1}) \geq \lambda_{k_s} \langle \exp^{-1}_{p^{k_s+1}} p^{k_s}, v^{k_s+1} \rangle, \qquad v^{k_s+1} = D(\exp_{\bar{p}})_{\exp^{-1}_{\bar{p}} p^{k_s+1}} \bar{v}.$$

Note que $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ implica que $\lim_{s\to+\infty} v^{k_s+1} = \bar{v}$. Visto que $\{\lambda_{k_s}\}$ é limitada, deixando s ir a $+\infty$ na última desigualdade, Proposition 3.4 junto com (4.7) nos dá

$$f^\circ(\bar{p}, \bar{v}) \geq \lim_{s\to+\infty} \sup f^\circ(p^{k_s+1}, v^{k_s+1}) \geq 0,$$

o que acarreta que $0 \in \partial^\circ f(\bar{p})$, i.e., $\bar{p}$ é um ponto estacionário de f e a primeira parte do teorema está concluída.

A segunda parte segue dos Lemas 4.3 e 4.4 e a prova do teorema está terminada.
c.q.d

Observação 4.3 *Na análise de convergência apresentada neste trabalho foram fundamentais algumas propriedades da derivada direcional e subdiferencial generalizados, como por exemplo a semicontinuidade superior da primeira, caracterizada na*

Proposição 3.4. Esta propriedade foi fundamental na prova da afirmação ii do Teorema 4.1. É possível provar a afirmação ii do Teorema 4.1 utilizando uma outra técnica, válida somente no caso que o conjunto de índices T é finito, que não necessita da semicontinuidade superior da derivada direcional generalizada. Tal resultado é uma generalização do Teorema 1 de [25] à estrutura Riemanniana. A seguir apresentamos:

Prova do Teorema 4.1 i-ii no caso que T é finito

Suponhamos que $T = \{1, 2, ..., m\}$ e $\varphi_\tau = \varphi(., \tau)$. Visto que $\max_{\tau \in T} L_\tau < \lambda_k$, Lema 4.1 implica que $\varphi_\tau + (\lambda_k/2)d^2(\, . \,, p^k)$ e $f + (\lambda_k/2)d^2(\, . \,, p^k)$ são fortemente convexas. Assim, do item iii do Lema 4.2, temos

$$0 \in \partial \left(f + \frac{\lambda_k}{2} d^2(\, . \,, p^k) \right) (p^{k+1}), \quad k = 0, 1, \ldots .$$

Além disso, aplicando Corolário 3.1 com $h_\tau = \varphi_\tau + (\lambda_k/2)d^2(\, . \,, p^k)$ e $h = f + (\lambda_k/2)d^2(\, . \,, p^k)$, concluímos que existem constantes $\alpha_\tau^{k+1} \geq 0$ com $\tau \in T(p^{k+1})$ tais que

$$0 = \sum_{\tau \in T(p^{k+1})} \alpha_\tau^{k+1} \operatorname{grad} \left(\varphi_\tau + \frac{\lambda_k}{2} d^2(\, . \,, p^k) \right) (p^{k+1}), \qquad \sum_{\tau \in T(p^{k+1})} \alpha_\tau^{k+1} = 1,$$

onde $T(p^{k+1}) = \{\tau \in T : f(p^{k+1}) = \varphi(p^{k+1}, \tau)\}$. Mas isto nos diz que

$$0 = \sum_{\tau \in T(p^{k+1})} \alpha_\tau^{k+1} \operatorname{grad} \varphi_\tau(p^{k+1}) - \lambda_k \exp^{-1}_{p^{k+1}} p^k, \qquad \sum_{\tau \in T(p^{k+1})} \alpha_\tau^{k+1} = 1, \quad k = 0, 1, \ldots . \tag{4.8}$$

Se a seqüência $\{p^k\}$ é finita, então existe k tal que $p^{k+1} = p^k$. Neste caso, $\exp^{-1}_{p^{k+1}} p^k = 0$ e a primeira igualdade em (4.8) torna

$$0 = \sum_{\tau \in T(p^{k+1})} \alpha_\tau^{k+1} \operatorname{grad} f_\tau(p^{k+1}),$$

que, junto com a Proposição 3.5, implica que $0 \in \partial^\circ f(p^k)$. Assim, p^k é um ponto estacionário de f e o item i está provado.

Agora, assuma que a seqüência $\{p^k\}$ é infinita e $\bar{p}$ é um ponto de acumulação dela. Seja $\{\alpha_\tau^{k+1}\} \subset \mathbb{R}^m$ a seqüência definida por

$$\alpha^{k+1} = (\alpha_1^{k+1}, \ldots, \alpha_m^{k+1}), \qquad \alpha_\tau^{k+1} = 0, \quad \tau \in T \backslash T(p^{k+1}).$$

Visto que $\sum_{\tau \in T(p^{k+1})} \alpha_i^{k+1} = 1$ temos que $\|\alpha^{k+1}\|_1 = 1$ para todo k, onde $\| \ \|_1$ denota a norma da soma em $\mathbb{R}^m$. Assim, $\{\alpha^{k+1}\}$ é limitada. Sejam $\{p^{k_s+1}\}$ e $\{\alpha^{k_s+1}\}$ subseqüências de $\{p^{k+1}\}$ e $\{\alpha^{k+1}\}$, respectivamente, tais que, $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ e $\lim_{s\to+\infty} \alpha^{k_s+1} = \bar{\alpha}$. Como f é contínua em Ω, item ii do Lema 4.2 implica que $\bar{p} \in L_f(f(q)) \subset \Omega$. Dado que T é finito, podemos assumir, sem perda de generalidade, que

$$T(p^{k_1+1}) = T(p^{k_2+1}) = ... = \bar{T}, \tag{4.9}$$

e equação (4.8) torna

$$0 = \sum_{i\in\bar{T}} \alpha_i^{k_s+1} \operatorname{grad} f_i(p^{k_s+1}) - \lambda_{k_s} \exp^{-1}_{p^{k_s+1}} p^{k_s}, \qquad \sum_{i\in\bar{T}} \alpha_i^{k_s+1} = 1, \quad s = 1, 2, \ldots.$$

Por outro lado, item v do Lema 4.2 implica que

$$\lim_{s\to\infty} \| \exp^{-1}_{p^{k_s+1}} p^{k_s} \| = \lim_{s\to\infty} d(p^{k_s+1}, p^{k_s}) = 0.$$

Dado que λ_{k_s} é limitada, $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ e $\lim_{s\to+\infty} \alpha^{k_s+1} = \bar{\alpha}$, fazendo s tender a $+\infty$ nas igualdades acima, temos

$$0 = \sum_{i\in\bar{T}} \bar{\alpha}_i \operatorname{grad} f_i(\bar{p}), \qquad \sum_{i\in\bar{T}} \bar{\alpha}_i = 1.$$

Usando definição de $T(\bar{p})$, relação (4.9) e continuidade de f, obtemos $\bar{T} \subset T(\bar{p})$. Portanto, como $\bar{p} \in \Omega$, segue da Proposição 3.5 que

$$0 \in \partial^\circ f(\bar{p}),$$

i.e., $\bar{p}$ é um ponto estacionário de f e o item ii está provado. **c.q.d**

4.1 Exemplos

Nesta seção apresentamos dois exemplos. No primeiro consideramos um problema de minimização não convexo onde a função objetivo é definida em uma variedade de Hadamard com curvatura identicamente zero. No próximo exemplo "generalizamos" o primeiro a um caso onde a variedade de Hadamard tem curvatura não identicamente zero. Em ambos os exemplos, tanto o método de ponto proximal local clássico (ver [25]) bem como o método de ponto proximal Riemanniano (ver [18]) não podem ser aplicados. Porém, o método proposto neste trabalho pode ser aplicado.

4.1.1 Método de ponto proximal para semi-reta positiva

Seja $(\mathbb{R}_{++}, \langle\, , \,\rangle)$ a variedade Riemanniana, onde $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$ e $\langle\, , \,\rangle$ é a métrica Riemanniana $\langle u, v\rangle = g(x)uv$ com $g : \mathbb{R}_{++} \to (0, +\infty)$. Assim, os símbolos de Christoffel e a equação geodésica são dados por

$$\Gamma(x) = \frac{1}{2}g^{-1}(x)\frac{dg(x)}{dx} = \frac{d}{dx}\ln\sqrt{g(x)}, \qquad \frac{d^2x}{dt^2} + \Gamma(x)\left(\frac{dx}{dt}\right)^2 = 0,$$

respectivamente. Além disso, em relação a uma função duas vezes diferenciável $h : \mathbb{R}_{++} \to \mathbb{R}$, o Gradiente e a Hessiana de h são dados por

$$\operatorname{grad} h = g^{-1}h', \qquad \text{hess } h = h'' - \Gamma h',$$

respectivamente, onde h' e h'' denotam a primeira e a segunda derivada de h no sentido Euclideano. Para mais detalhes ver [33]. Assim, no caso particular que $g(x) = x^{-2}$,

$$\Gamma(x) = -x^{-1}, \quad \operatorname{grad} h(x) = x^2h'(x), \quad \text{hess } h(x) = h''(x) + x^{-1}h'(x). \tag{4.10}$$

Além disso, a aplicação $\psi : \mathbb{R} \to \mathbb{R}_{++}$, definida por $\psi(x) = \mathrm{e}^x$, é uma isometria entre o espaço Euclideano $\mathbb{R}$ e a variedade $(\mathbb{R}_{++}, \langle\, , \,\rangle)$ e a distância Riemanniana, $d : \mathbb{R}_{++} \times \mathbb{R}_{++} \to \mathbb{R}_{+}$, é dada por

$$d(x, y) = |\psi^{-1}(x) - \psi^{-1}(y)| = |\ln(x/y)|, \tag{4.11}$$

ver, Exemplo 3.1.2 de [6]. Portanto, $(\mathbb{R}_{++}, \langle\, , \,\rangle)$ é uma variedade de Hadamard e a única geodésica $x : \mathbb{R} \to \mathbb{R}_{++}$, com condições iniciais $x(0) = x_0$ e $x'(0) = v$, é dada por

$$x(t) = x_0\mathrm{e}^{(v/x_0)t}.$$

Da expressão acima é fácil ver que qualquer intervalo $I \subset \mathbb{R}_{++}$ é um conjunto convexo da variedade $(\mathbb{R}_{++}, \langle\, , \,\rangle)$.

Agora sejam $f_1, f_2, f : \mathbb{R}_{++} \to \mathbb{R}$ e $\varphi : \mathbb{R}_{++} \times [0, 1] \to \mathbb{R}$ funções reais tais que

$$\varphi(x, \tau) = f_1(x) + \tau(f_2(x) - f_1(x)), \quad f(x) = \max_{\tau\in[0,1]} \varphi(x, \tau), \tag{4.12}$$

e considere o problema

$$\begin{aligned} &\min f(x) \\ &\text{s.a. } x \in \mathbb{R}_{++}. \end{aligned}$$

Tome uma seqüência $\{\lambda_k\}$ satisfazendo $0 < \lambda_k$. De (4.11), o método de ponto proximal (4.2) torna

$$x^{k+1} := \mathrm{argmin}_{x\in\mathbb{R}_{++}} \left\{ f(x) + \frac{\lambda_k}{2} \ln^2\left(\frac{x}{x^k}\right) \right\}, \qquad k = 0, 1, \ldots.$$

Se f_1 e f_2 são dadas, respectivamente, por

$$f_1(x) = \ln(x), \qquad f_2(x) = -\ln(x) + \mathrm{e}^{-2x} - \mathrm{e}^{-2},$$

então φ é contínua e $\varphi(., \tau)$ é continuamente diferenciável para cada $\tau \in [0, 1]$. A última expressão em (4.10) implica que

$$\text{hess } f_1(x) = 0 \quad , \quad \text{hess } f_2(x) = (4 - 2/x)\mathrm{e}^{-2x}, \qquad x \in \mathbb{R}_{++}, \tag{4.13}$$

e, como uma conseqüência, da primeira expressão em (4.12), obtemos

$$\text{hess}_x\, \varphi(x, \tau) = \tau \text{hess } f_2(x), \quad \forall\, x \in \mathbb{R}_{++} \quad \forall\, \tau \in [0, 1].$$

Note que, para $0 < \epsilon < 1/4$ e $\Omega = (\epsilon, +\infty)$, hess f_2 é limitada em Ω e, portanto, grad f_2 é Lipschitz em Ω. Denotamos por L a constante de Lipschitz of grad f_2. Da última igualdade, $\text{hess}_x\, \varphi(., \tau)$ é também limitada em Ω e $\text{grad}_x\, \varphi(., \tau)$ é Lipschitz em Ω com constante $L_\tau = \tau L$ para todo $\tau \in [0, 1]$. Além disso, $\sup_{\tau\in[0,1]} L_\tau = L < +\infty$.

Afirmamos que $f(x) = \max_{j=1,2} f_j(x)$. De fato, note que $f_2(x) - f_1(x) > 0$ para $x \in (0, 1)$, $f_2(x) - f_1(x) < 0$ para $x \in (1, +\infty)$ e $f_1(1) = f_2(1)$. Assim, a função afim $[0, 1] \ni \tau \mapsto \varphi(x, \tau)$ satisfaz

$$\max_{\tau\in[0,1]} \varphi(x, \tau) = \begin{cases} f_1(x), & x \in (0, 1), \\ f_2(x), & x \in (1, +\infty) \end{cases}$$

e a afirmação segue. Tome $q = 5/16$. Assim, $0 = \inf_{x\in\mathbb{R}_{++}} f(x) < f(q)$ e $L_f(f(q)) \subset \Omega$. Claramente a hipótese **h1** do Teorema 4.1 é verificada com $U^* = \{1\}$.

Afirmamos que existe $c \in (0, f(1/2)) = (0, f_1(1/2))$ tal que $L_f(c)$ é convexa e f é convexa em $L_f(c)$ (no sentido Riemanniano). De fato, visto que hess $f_1 \geq 0$ em $\mathbb{R}_{++}$ e hess $f_2 \geq 0$ em $[1/2, +\infty)$, Teorema 6.2 de [6] implica que f_1 é convexa em $\mathbb{R}_{++}$ e f_2 é convexa em $[1/2, +\infty)$. Assim, segue da Proposição 2.5 que f é convexa em $[1/2, +\infty)$. Note que para todo $c \in (0, f(1/2))$, temos $L_f(c) \cap [1/2, +\infty) = L_f(c)$. Portanto, da convexidade de f em $(1/2, +\infty)$ concluímos que $L_f(c)$ é convexo e a afirmação está provada. Assim, $L_f(c)$ e f satisfazem a hipótese **h2** do Teorema 4.1, por exemplo com $c = f(3/4)$.

Agora, note que $L_f(f(q)) \backslash L_f(c) = [5/16, 3/4) \cup (a, b]$, onde $a = (4/3)\mathrm{e}^{(\mathrm{e}^{-3/2} - \mathrm{e}^{-2})}$ e $b = (16/5)\mathrm{e}^{(\mathrm{e}^{-5/8} - \mathrm{e}^{-2})}$. Além disso, f é diferenciável em $L_f(f(q)) \backslash L_f(c)$ com $\operatorname{grad} f(x) = \operatorname{grad} f_1(x)$ para $x \in [a, b]$ e $\operatorname{grad} f(x) = \operatorname{grad} f_2(x)$ para $x \in [5/16, 3/4]$. Da segunda expressão em (4.10), temos

$$\operatorname{grad} f(x) = x, \quad x \in (a, b] \quad \text{e} \quad \operatorname{grad} f(x) = -x - 2x^2 \mathrm{e}^{-2x}, \quad x \in [5/16, 3/4).$$

Assim, temos $\| \operatorname{grad} f(x) \| \geq \| \operatorname{grad} f(5/16) \| > 2/5$ e f satisfaz a hipótese **h3** do Teorema 4.1.

Resumindo, todas as hipóteses do Teorema 4.1 são satisfeitas com $\Omega = (\epsilon, +\infty)$, $q = 5/16$, $c = f(3/4)$ e $\delta = 2/5$. Portanto, se $x^0 \in \mathbb{R}_{++}$ e $\bar{\lambda} > 0$ são tais que $x^0 \in L_f(f(q))$ e $\max_{i \in I} L_i < \lambda_k \leq \bar{\lambda}$, o método de ponto proximal, caracterizado no Teorema 4.1, pode ser aplicado para resolver o problema não convexo deste exemplo.

Observação 4.4 *A função $f(x) = \max_\tau \varphi(x, \tau)$, no exemplo acima, é não convexa (no sentido Euclideano) quando restrita a qualquer vizinhança aberta contendo seu minimizador $x^* = 1$. Portanto, o método de ponto proximal local clássico (ver [25]) não pode ser aplicado para minimizar essa função. Além disso, como f é não convexa no sentido Riemanniano, o método de ponto proximal Riemanniano (ver [18]) também não pode ser aplicado para minimizar essa função.*

4.1.2 Método de ponto proximal para o cone das matrizes simétricas definidas positivas

Seja $\mathbb{S}^n$ o conjunto das matrizes simétricas, $\mathbb{S}^n_+$ o cone das matrizes simétricas semi-definidas positivas e $\mathbb{S}^n_{++}$ o cone das matrizes simétricas definidas positivas, ambas $n \times n$. Para $X, Y \in \mathbb{S}^n_+$, $Y \succeq X$ (ou $X \preceq Y$) diz que $Y - X \in \mathbb{S}^m_+$ e $Y \succ X$ (ou $X \prec Y$) diz que $Y - X \in \mathbb{S}^n_{++}$. Denotaremos a norma de Frobenius por $\| \, . \, \|_F$.

Seguindo Rothaus [44], seja $M := (\mathbb{S}^n_{++}, \langle \, , \, \rangle)$ a variedade Riemanniana munida com a métrica induzida pela Hessiana Euclideana de $\Psi(X) = -\ln \det X$,

$$\langle U, V \rangle = \operatorname{tr}(V \Psi''(X) U) = \operatorname{tr}(V X^{-1} U X^{-1}), \qquad X \in M, \qquad U, V \in T_X M, \quad (4.14)$$

onde $\operatorname{tr}(A)$ denota o traço da matriz $A \in \mathbb{S}^n$ e $T_X M \approx \mathbb{S}^n$, com a correspondente norma denotada por $\| \, . \, \|$. Neste caso o único segmento geodésico conectando quaisquer $X, Y \in M$ é dado por

$$\gamma(t) = X^{1/2} \left(X^{-1/2} Y X^{-1/2} \right)^t X^{1/2}, \qquad t \in [0,1],$$

ver, por exemplo, [45]. Mais precisamente, M é uma variedade de Hadamard, ver por exemplo [46], Teorema 1.2. pág. 325. Da igualdade acima é imediato que

$$\gamma'(0) = X^{1/2} \ln \left(X^{-1/2} Y X^{-1/2} \right) X^{1/2}.$$

Assim, para cada $X \in M$, $\exp_X^{-1} : M \to T_X M$ e $\exp_X : T_X M \to M$ são dadas, respectivamente, por

$$\exp_X^{-1} Y = X^{1/2} \ln \left(X^{-1/2} Y X^{-1/2} \right) X^{1/2}, \quad \exp_X V = X^{1/2} e^{\left(X^{-1/2} V X^{-1/2} \right)} X^{1/2}. \tag{4.15}$$

Agora, visto que a distância Riemannaina d é dada por $d(X\,,\,Y) = ||exp_X^{-1} Y||$, de (4.14) junto com primeira expressão em (4.15), concluímos que

$$d^2(X,Y) = \operatorname{tr}\left(\ln^2 X^{-1/2} Y X^{-1/2} \right) = \sum_{i=1}^{n} \ln^2 \lambda_i \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right), \tag{4.16}$$

onde $\lambda_i(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})$ denota o i-ésimo autovalor da matriz $X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}$. O gradiente e a Hessiana de uma função duas vezes diferenciável $F : \mathbb{S}^n_{++} \to \mathbb{R}$ são dados, respectivamente, por

$$\operatorname{grad} F(X) = X F'(X) X, \ \operatorname{hess} F(X)\big(V,V\big) = \operatorname{tr}\big(V F''(X) V\big) + \operatorname{tr}\big(F'(X) V X^{-1} V\big),$$

para todo $V \in T_X M$, onde $F'(X)$ e $F''(X)$ são o gradiente e a Hessiana Euclideana, respectivamente. Recordamos que uma função duas vezes difereciável $F : \mathbb{S}^n_{++} \to \mathbb{R}$ é convexa em M se ela satisfaz

$$\operatorname{hess} F(X)\big(V,V\big) \geq 0, \qquad\qquad X \in M, \quad V \in T_X M.$$

Seja $F_1, F_2, F_3 : \mathbb{S}^n_{++} \to \mathbb{R}$ dadas, respectivamente, por

$$F_1(X) = \ln \det X, \quad F_2(X) = -4 \ln \det X + e^{-2 \operatorname{tr} X} - e^{-2n}, \quad F_3(X) = \operatorname{tr} X^{-1} - n.$$

Note que F_1, F_2 e F_3 são funções duas vezes diferenciáveis em $\mathbb{S}^n_{++}$. Tomando $X \in T_X M$ e $V \in T_X M$, os gradientes Euclideanos são dados por

$$F_1'(X) = X^{-1}, \quad F_2'(X) = -4X^{-1} - 2e^{-2 \operatorname{tr} X}, \quad F_3'(X) = -X^{-2},$$

e as Hessianas Euclideanas são dadas por

$$
\begin{aligned}
F_1''(X)V &= -X^{-1}VX^{-1},\\
F_2''(X)V &= 4X^{-1}VX^{-1} + 4e^{-2\operatorname{tr} X}V,\\
F_3''(X)V &= X^{-1}VX^{-2} + X^{-2}VX^{-1}.
\end{aligned}
$$

Assim os gradientes Riemannianos de F_1, F_2 e F_3 são dados por

$$
\operatorname{grad} F_1(X) = X, \quad \operatorname{grad} F_2(X) = -4X - 2e^{-2\operatorname{tr} X}X^2, \quad \operatorname{grad} F_3(X) = -I, \tag{4.17}
$$

onde I denota a matriz identidade, e as Hessianas Riemannainas de F_1, F_2 e F_3 são dadas por

$$
\begin{aligned}
&\operatorname{hess} F_1(X)\,(V,V) = 0,\\
&\operatorname{hess} F_2(X)\,(V,V) = 2e^{-2\operatorname{tr} X}\operatorname{tr}\left[(2I - X^{-1})V^2\right],\\
&\operatorname{hess} F_3(X)\,(V,V) = \operatorname{tr}\left(X^{-1}VX^{-2}V\right) = \|X^{-1}VX^{-1/2}\|_F^2.
\end{aligned} \tag{4.18}
$$

É fácil ver que as funções F_1 e F_3 são convexas em toda a variedade e a função F_2 é convexa em qualquer subconjunto do conjunto

$$
C := \left\{X \in \mathbb{S}^n_{++} : 2I - X^{-1} \succ 0\right\} = \left\{X \in \mathbb{S}^n_{++} : \lambda_{\min}(X) > 1/2\right\}, \tag{4.19}
$$

onde $\lambda_{\min}(A)$ denota o menor autovalor da matriz A.

Seja $F : \mathbb{S}^n_{++} \to \mathbb{R}$ dada por

$$
F(X) = \max\nolimits_{j=1,2,3} F_j(X).
$$

Da Proposição 2.5 F é convexa em qualquer subconjunto convexo de C. Considere o seguinte problema de otimização

$$
\begin{aligned}
&\min F(X)\\
&\text{s.a. } \ X \in \mathbb{S}^n_{++}.
\end{aligned} \tag{4.20}
$$

Se $\{\eta_k\}$ é uma sequência satisfazendo $0 < \eta_k$ então, de (4.16) o método de ponto proximal (4.2) torna

$$
X_{k+1} = \arg\min\nolimits_{X \in \mathbb{S}^n_{++}} \left\{ f(X) + \frac{\eta_k}{2}\sum_{i=1}^n \ln^2 \lambda_i\left(X_k^{-\frac{1}{2}} X X_k^{-\frac{1}{2}}\right)\right\}, \qquad k = 0, 1, \ldots.
$$

Defina os conjuntos U_1, U_2 e U_3 como

$$U_1 := \{X \in \mathbb{S}^n_{++} : \det X = 1\},$$
$$U_2 := \{X \in \mathbb{S}^n_{++} : \det X > 1\},$$
$$U_3 := \{X \in \mathbb{S}^n_{++} : \det X < 1\},$$

e note que $\mathbb{S}^n_{++} = U_1 \cup U_2 \cup U_3$ e $U_i \cap U_j = \emptyset$, para $i, j = 1, 2, 3$, $i \neq j$.

Afirmação 4.1

$$F(x) = \begin{cases} F_3(X) \geq 0, & X \in U_1; \\ \max\{F_1(X), F_3(X)\} > 0, & X \in U_2; \\ \max\{F_2(X), F_3(X)\} > 0, & X \in U_3. \end{cases} \tag{4.21}$$

Em particular, $\inf_{\mathbb{S}^n_{++}} F(X) = 0$ *e* $U^* = \{I\}$*, onde* U^* *é o conjunto solução do problema* (4.20) *e* I *é matriz identidade* $n \times n$*.*

Prova. Primeiramente note que, para cada $X \in \mathbb{S}^n_{++}$, temos

$$\sqrt[n]{\det X} \leq \frac{\operatorname{tr} X}{n}. \tag{4.22}$$

Se $X \in U_1$ então $\det X = 1$. Usando a desigualdade acima obtemos $\operatorname{tr} X \geq n$, o que implica

$$F_2(X) = e^{-2 \operatorname{tr} X} - e^{-2n} \leq 0 = F_1(X).$$

Por outro lado, $\det X^{-1} = 1$. Consequentemente, de (4.22) temos $\operatorname{tr} X^{-1} - n \geq 0$ e, portanto, $F_3(X) \geq 0$. Assim, da definição de F e levando em conta a desigualdade acima, concluímos que

$$F(X) = F_3(X) \geq 0, \qquad X \in U_1. \tag{4.23}$$

Se $X \in U_2$ então $\det X > 1$. Então, de (4.22) também obtemos $\operatorname{tr} X > n$. Portanto,

$$F_2(X) = -4 \ln \det X + e^{-2 \operatorname{tr} X} - e^{-2n} < -\ln \det X < 0.$$

Visto que $\det X > 1$ temos $F_1(X) = \ln \det X > 0$. Portanto,

$$F(X) = \max\{F_1(X), F_3(X)\} > 0, \qquad X \in U_2. \tag{4.24}$$

Finalmente, se $X \in U_3$ então $\det X^{-1} > 1$. Assim, a desigualdade (4.22) implica que $\operatorname{tr} X^{-1} > n$ e, consequentemente, $F_3(X) = \operatorname{tr} X^{-1} - n > 0$. Como $\det X < 1$ temos $F_1(X) = \ln\det(X) < 0$. Logo,

$$F(X) = \max\{F_2(X), F_3(X)\} > 0, \qquad X \in U_3, \tag{4.25}$$

e a primeira parte da afirmação está provada.

Da primeira parte concluímos que $F(X) \geq 0$, para todo $X \in \mathbb{S}^n_{++}$. Visto que $F(I) = 0$, temos $\inf_{\mathbb{S}^n_{++}} F(X) = 0$. Para provar a última afirmação note que, se $F(X) = 0$ então $X \in U_1$. Definição de U_1 e (4.21) nos permite concluir que $\det X^{-1} = 1$ e $F(X) = \operatorname{tr}(X^{-1}) - n$, para todo $X \in U_1$. Assim, se $X \in U^*$ então $\operatorname{tr}(X^{-1}) = n$ e $\det X^{-1} = 1$, o que implica que (4.22) ocorre com igualdade. Por outro lado, (4.22) ocorre com igualdade se, e somente se, $\lambda_1(X) = ... = \lambda_n(X) > 0$. Como $\operatorname{tr}(X^{-1}) = n$ concluímos que $\lambda_i(X) = 1$, para $i = 1, \ldots, n$ e o resultado segue. **c.q.d**

Afirmação 4.2 $F_1(X) > F_3(X)$ *para* $X \in A$, *onde*

$$A := \left\{X \in \mathbb{S}^n_{++} : X - I \succ 0\right\} = \left\{X \in \mathbb{S}^n_{++} : \lambda_{\min}(X) > 1\right\} \subset U_2.$$

Consequentemente, $F(X) = F_1(X)$ *para* $X \in A$.

Prova. Defina a seguinte função

$$\psi(X) := F_1(X) - F_3(X).$$

Do teorema do valor médio no espaço Euclideano $\mathbb{S}^n_{++}$, para $X \in U_2$ existe $\tilde{X} \in U_2$ tal que

$$\psi(X) = \psi(I) + \operatorname{tr}\left(\psi'(\tilde{X})(X - I)\right) = \operatorname{tr}\left((\tilde{X}^{-1} + \tilde{X}^{-2})(X - I)\right).$$

Se $X \in A$ então $X - I \in \mathbb{S}^n_{++}$. Visto que $X - I$ e $\tilde{X}^{-1} + \tilde{X}^{-2}$ pertencem a $\mathbb{S}^n_{++}$ e o traço do produto de matrizes definidas positivas é positivo, concluímos da igualdade acima que $\psi(X) = F_1(X) - F_3(X) > 0$, para todo $X \in A$, o que prova a primeira parte da afirmação. A segunda parte segue da Afirmação 4.1. **c.q.d**

Afirmação 4.3 $F(X) = F_2(X)$ *para* $X \in B$, *onde*

$$B := \left\{X \in \mathbb{S}^n_{++} : (1/4)I \prec X \prec I\right\} = \left\{X \in \mathbb{S}^n_{++} : 1/4 < \lambda_{\min}(X),\ \lambda_{\max}(X) < 1\right\}.$$

Prova. Note que $B \subset U_3 \cap \{X \in \mathbb{S}^n_{++} : \operatorname{tr} X \leq n\}$. Assim, de (4.25), para provar a afirmação é suficiente verificar que $F_2(X) > F_3(X)$ para todo $X \in B$. Seja $\phi : \mathbb{R}_{++} \to \mathbb{R}$ a função dada por $\phi(t) := -4\ln t - 1/t + 1$. Da convexidade de $-4\ln t$, temos

$$\phi(t) \geq -4\ln 1 - 4(t-1) - \frac{1}{t} + 1 > 0, \qquad t \in (1/4\,,1). \tag{4.26}$$

Por outro lado, das definições de F_2, F_4 e ϕ, obtemos

$$F_2(X) - F_4(X) = e^{-2\operatorname{tr} X} - e^{-2n} + \sum_{j=1}^{n} \phi(\lambda_j(X)).$$

Visto que $e^{-2\operatorname{tr} X} - e^{-2n} > 0$ para $X \in B$, combinando (4.26) com a última igualdade concluímos que $F_2(X) > F_3(X)$ para $X \in B$. Portanto, a igualdade desejada segue. **c.q.d**

É fácil verificar que F é uma função coerciva. Assim, $L_F(a)$ é limitada para todo $a \in \mathbb{R}$, i.e., todos os conjuntos de nível de F são limitados. Seja

$$Q := \operatorname{diag}(1/4, \ldots, 1/4). \tag{4.27}$$

Como $L_F(F(Q)) \subset \mathbb{S}^n_{++}$ é um conjunto limitado, então existe $\kappa > 4$ suficientemente grande tal que

$$\lambda_{\min}(X) > 1/\kappa, \quad \lambda_{\max}(X) < \kappa, \qquad X \in L_F(F(Q)). \tag{4.28}$$

Agora definamos

$$\Omega := \{X \in \mathbb{S}^n_{++} : \operatorname{tr} X^{-1} < \kappa n\}.$$

Visto que a função $\mathbb{S}^n_{++} \ni X \mapsto \operatorname{tr} X^{-1}$ é convexa, o conjunto Ω é convexo. Da Afirmação 4.1 $0 = \inf_{\mathbb{S}^n_{++}} F(X) < F(Q)$ e, consequentemente, $\operatorname{int} L_F(F(Q)) \neq \emptyset$, onde $\operatorname{int} A$ representa o interior do conjunto A. Além disso, é imediato verificar que

$$C \subset \Omega, \qquad I \in \operatorname{int} L_F(F(Q)) \subset L_F(F(Q)) \subset \Omega.$$

Afirmação 4.4 *Os campos de vetores gradientes* $\operatorname{grad} F_1, \operatorname{grad} F_2$ *e* $\operatorname{grad} F_3$ *são Lipschitz em* Ω.

Prova. Da primeira igualdade em (4.18) e Proposição 2.7 é imediato que o campo de vetor gradiente grad F_1 é Lipschitz. Agora, a segunda igualdade em (4.18) implica que

$$|\operatorname{hess} F_2(X)\,(V,V)\,| = 2e^{-2\operatorname{tr} X}|\operatorname{tr}\left[(2I - X^{-1})V^2\right]|, \qquad X \in M,\ V \in T_X M.$$

Como $\operatorname{tr} X > 0$ e usando a desigualdade de Cauchy-Schwartz, obtemos da igualdade acima

$$|\operatorname{hess} F_2(X)\,(V,V)\,| \leq 2\|2I - X^{-1}\|_F \|V^2\|_F, \qquad X \in M,\ V \in T_X M,$$

e, consequentemente, usando a definição da métrica, concluímos que

$$|\operatorname{hess} F_2(X)\,(V,V)\,| \leq 2\left(2\sqrt{n} + \|X^{-1}\|_F\right)\|V^2\|_F, \qquad X \in M,\ V \in T_X M.$$

Assim, como $\|X^{-1}\|_F \leq \operatorname{tr} X^{-1}$, da última desigualdade junto com a definição de Ω, temos

$$|\operatorname{hess} F_2(X)\,(V,V)\,| \leq 4\sqrt{n} + 2\kappa n, \qquad X \in \Omega,\ \|V\|_F = 1. \tag{4.29}$$

Por outro lado, usando a terceira igualdade em (4.18) e a desigualdade de Cauchy-Schwarz, obtemos

$$|\operatorname{hess} F_3(X)\,(V,V)\,| = |\operatorname{tr}\left(X^{-1}VX^{-2}V\right)| \leq \|X^{-1}VX^{-1}\|_F \|VX^{-1}\|_F,$$
$$X \in M,\ V \in T_X M.$$

Visto que $\|.\|_F$ é submultiplicativa e usando novamente que $\|X^{-1}\|_F \leq \operatorname{tr} X^{-1}$, desigualdade acima que

$$|\operatorname{hess} F_3(X)\,(V,V)\,| \leq \left(\operatorname{tr} X^{-1}\right)^3 \|V\|_F^2, \quad X \in M,\ V \in T_X M.$$

Portanto,

$$|\operatorname{hess} F_3(X)\,(V,V)\,| \leq \kappa^3 n^3, \qquad X \in \Omega,\ \|V\|_F = 1. \tag{4.30}$$

De (4.29) e (4.30) concluímos que *hess* F_2 e *hess* F_3 são também operadores limitados quando restrito a Ω. Logo, da Proposição 2.7, grad F_1, grad F_2 e grad F_3 são Lipschitz em Ω. **c.q.d**

Denotamos por L_i as constantes de Lipschitz de grad f_i, $i = 1, 2, 3$.

Da Afirmação 4.1 a hipótese **h1** do Teorema 4.1 é satisfeita com $U^* = \{I\}$. O próximo resultado será útil para assegurar que a hipótese **h2** do Teorema 4.1 é satisfeita.

Afirmação 4.5 *Existe* $\sigma > 0$ *e* $\hat{X} \in \mathbb{S}^n_{++} \backslash \{I\}$ *tal que,*

$$L_F(F(\hat{X})) \subset B_\sigma(I) \subset \operatorname{int} L_F(F(Q)) \cap C,$$

onde $B_\sigma(I) := \{X \in \mathbb{S}^n_{++} : d(X, I) < \sigma\}$, C *é definida em* (4.19) *e* Q *em* (4.27).

Prova. Visto que $U^* = \{I\}$ e $I \neq Q$, concluímos que $I \in \operatorname{int} L_F(F(Q)) \cap C$. Assim, é fácil ver que existe $\sigma > 0$ tal que

$$B_\sigma(I) \subset \operatorname{int} L_F(F(Q)) \cap C,$$

provando a última inclusão. Para provar a primeira inclusão assuma, por contradição, que para cada $X \in B_\sigma(I)$ temos $L_F(F(X)) \cap (\mathbb{S}^n_{++} \backslash B_\sigma(I)) \neq \emptyset$. Portanto, existem seqüências $\{X^k\}$, $\{Y^k\} \subset \mathbb{S}^n_{++}$ tal que

$$X^k \in B_\sigma(I), \quad \lim_{k \to \infty} X^k = I, \qquad Y^k \in L_F(F(X^k)) \backslash B_\sigma(I), \qquad k = 0, 1, \ldots. \tag{4.31}$$

Visto que $\{X^k\} \subset B_\sigma(I) \subset L_F(F(Q))$, temos que $L_F(F(X^k)) \subset L_F(F(Q))$, para $k = 0, 1, \ldots$. Portanto, $\{Y^k\} \subset L_F(F(Q))$. Como $L_F(F(Q))$ é um conjunto limitado, assumimos (tomando uma subseqüência, se necessário) que $\{Y^k\}$ converge a algum $\bar{Y} \in L_F(F(Q))$. Assim, continuidade de F junto com as duas últimas equações em (4.31) implicam que

$$F(\bar{Y}) \leq F(I), \qquad \bar{Y} \neq I,$$

e, como $U^* = \{I\}$, obtemos uma contradição, o que prova a afirmação. **c.q.d**

Da afirmação acima junto com a convexidade da bola $B_\sigma(I)$, concluímos que a hipótese **h2** é satisfeita com $c = F(\hat{X})$.

Afirmação 4.6 *Existe* $\tilde{\delta} > 0$ *tal que*

$$\| \operatorname{grad} F_j(X) \| > \tilde{\delta}, \qquad X \in L_F(F(Q)), \qquad j = 1, 2, 3, \tag{4.32}$$

onde Q *é definida em* (4.27).

Prova. Seja $X \in M$. Usando (4.17) junto com a definição da métrica Riemanniana, temos

$$\begin{aligned}
\| \operatorname{grad} F_1(X) \| &= \sqrt{n}, \\
\| \operatorname{grad} F_2(X) \| &= \sqrt{\operatorname{tr}(I + 4e^{-2 \operatorname{tr} X} \operatorname{tr} X + 4e^{-4 \operatorname{tr} X} \operatorname{tr} X^2)}, \\
\| \operatorname{grad} F_3(X) \| &= \sqrt{\operatorname{tr} X^{-2}}.
\end{aligned}$$

Agora, se $X \in L_F(F(Q))$ então da segunda desigualdade em (4.28) concluímos que

$$\| \operatorname{grad} F_3(X)\| > \sqrt{n}/\kappa.$$

É fácil ver que $\| \operatorname{grad} F_2(X)\| > \sqrt{n}$. Assim, tomando $\tilde{\delta} = \sqrt{n}/\kappa$, o resultado segue.
c.q.d

Para verificar que F satisfaz a hipótese **h3** usaremos a seguinte notação

$$B_r^E(I) := \{X \in S_{++}^n : \|X - I\|_F < r\}, \tag{4.33}$$

par a bola Euclideana com centro em I e raio r. Visto que $I \in \operatorname{int} L_F(F(\hat{X}))$, onde $\hat{X}$ é como definido na Afirmação 4.5, tome $r > 0$ tal que

$$B_r^E(I) \subset L_F(F(\hat{X})).$$

Usando Afirmação 4.5 temos $L_F(F(\hat{X})) \subset L_F(F(Q))$. Então, para provar que F satisfaz a hipótese **h3** é suficiente provar:

Afirmação 4.7 *Existe $\delta > 0$ tal que*

$$\|y(X)\| > \delta, \qquad X \in L_F(F(Q))\backslash B_r^E(I), \quad y(X) \in \partial^\circ F(X). \tag{4.34}$$

Prova. Seja $B_r^E(I)$ a bola Euclideana como definida em (4.33). Tome $X \in L_F(F(Q))\backslash B_r^E(I)$ e considere o seguinte conjunto

$$I(X) := \{i : F(X) = F_i(X), i = 1, 2, 3\}.$$

Então, Afirmação 4.1 implica que

$$I(X) = \begin{cases} \{3\}, & X \in U_1; \\ \{1\},\ \{3\},\ or\ \{1,3\}, & X \in U_2; \\ \{2\},\ \{3\},\ or\ \{2,3\}, & X \in U_3. \end{cases} \tag{4.35}$$

Assim, concluímos da Afirmação 4.6 que, para provar a afirmação em questão, é suficiente considerar os seguintes casos:

a) $X \in D_1 = \{X \in U_2 : I(X) = \{1, 3\}\}$;

b) $X \in D_2 = \{X \in U_3 : I(X) = \{2, 3\}\}$.

Suponhamos inicialmente que $a)$ ocorre, i.e., $X \in D_1$ e tome $y(X) \in \partial^\circ F(X)$. Lema 7.3 garante que existe $\alpha \in [0,1]$ tal que

$$y(X) = \alpha \operatorname{grad} F_1(X) + (1-\alpha) \operatorname{grad} F_3(X).$$

Agora, usando (4.17), definição da métrica Riemanniana e última expressão, temos

$$\|y(X)\|^2 = \operatorname{tr}\left(\alpha I - (1-\alpha)X^{-1}\right)^2,$$

que, depois de simples manipulções algébricas, torna

$$\|y(X)\|^2 = \sum_{j=1}^{n} \left(\alpha - (1-\alpha)\lambda_j^{-1}(X)\right)^2, \tag{4.36}$$

onde $\lambda_j(X)$ denota o j^{th} autovalor da matriz X. De (4.36) é fácil ver que

$$\|y(X)\|^2 \geq \left(\alpha - (1-\alpha)\lambda_{\min}^{-1}(X)\right)^2 + \left(\alpha - (1-\alpha)\lambda_{\max}^{-1}(X)\right)^2.$$

Note que, para cada X fixado, o lado direito na desigualdade acima é um polinômio de grau dois na variável α. Assim, minimizando tal polinômio, obtemos

$$\|y(X)\|^2 \geq \frac{\left(\lambda_{\min}^{-1}(X) - \lambda_{\max}^{-1}(X)\right)^2}{\left(1+\lambda_{\min}^{-1}(X)\right)^2 + (1+\lambda_{\max}^{-1}(X))^2}. \tag{4.37}$$

Visto que $X \in D_1 \subset U_2$, combinando Afirmação 4.1 com Afirmação 4.2, temos $\lambda_{\min}(X) \leq 1$. Além disso, neste caso $\det X > 1$ que, combinando com $\lambda_{\min}(X) \leq 1$, implica que

$$\lambda_{\max}(X) > 1.$$

Por outro lado, como $\det X > 1$, equação (4.22) permite concluir que $\operatorname{tr} X > n$. Como $X \notin B_r^E(I)$ e $\operatorname{tr} X > n$, última desigualmente e algumas manipulações algébricas implicam

$$r^2 \leq \operatorname{tr}(X-I)^2 = \operatorname{tr} X^2 - 2\operatorname{tr} X + n \leq n\lambda_{\max}^2(X) - 2n + n\lambda_{\max}^2(X),$$

de onde segue que

$$\lambda_{\max}(X) > \sqrt{1 + r^2/2n}.$$

Combinando a primeira desigualdade em (4.28) com as desigualdade (4.37) e a última desigualdade, obtemos

$$\|y(X)\| > \delta_1 = \frac{1 - 1/\sqrt{1+r/2n}}{\sqrt{(1+\kappa)^2 + (1+1/\sqrt{1+r/2n})^2}} > 0, \quad X \in D_1, \quad y(X) \in \partial^\circ F(X). \tag{4.38}$$

Agora, suponhamos que o item *b*) ocorre, i.e., $X \in D_1$. Tome $y(X) \in \partial^\circ F(X)$. Devido ao Lema 7.3 existe $\alpha \in [0,1]$ tal que

$$y(X) = \alpha \operatorname{grad} F_1(X) + (1-\alpha) \operatorname{grad} F_3(X).$$

Assim, (4.17), definição da métrica Riemanniana e última expressão implicam que

$$\|y(X)\|^2 = \operatorname{tr}\left(\alpha(-X - 2e^{-2\operatorname{tr} X}X^2) - (1-\alpha)I\right)^2.$$

Algumas manipulações algébricas na última igualdade e levando em conta que $\lambda_j(X) > 0$, para todo $j = 1, \ldots, n$, nos permite obter

$$\|y(X)\|^2 > \alpha^2 n + 2\alpha(1-\alpha)\operatorname{tr} X^{-1} + (1-\alpha)\operatorname{tr} X^{-2}. \tag{4.39}$$

Devido $X \in D_2 \subset U_3$, temos $\det X^{-1} > 0$. Portanto, usando (4.22), obtemos

$$\operatorname{tr} X^{-1} > n, \qquad \operatorname{tr} X^{-2} > n.$$

Visto que $\alpha \in [0,1]$, substituindo duas desigualdades acima em (4.39), obtemos

$$\|y(X)\|^2 > n, \qquad X \in D_2, \quad y(X) \in \partial^\circ F(X). \tag{4.40}$$

Portanto, tamando $\delta = \min\{\tilde{\delta}, \delta_1, n\} > 0$, onde $\tilde{\delta}$ é dado na última afirmação, a afirmação segue de (4.35), (4.38), (4.40) e Afirmação 4.6. **c.q.d**

Logo, como $L_F(F(Q)) \backslash L_F(F(\hat{X})) \subset L_F(F(Q)) \backslash B_r^E(I)$, a hipótese **h3** é satisfeita com $\delta = \min\{\tilde{\delta}, \delta_1, \delta_2\} > 0$.

Resumindo, todas as hipóteses do Teorema 4.1 são satisfeitas com $\Omega = \{X \in \mathbb{S}^n_{++} : \operatorname{tr} X^{-1} < \kappa n\}$, $q = \operatorname{diag}(1/4, \ldots, 1/4)$, $c = F(\hat{X})$, onde $\hat{X}$ é como definido na Afirmação 4.5, e $\delta > 0$ como definido acima. Se $X^0 \in \mathbb{S}^n_{++}$ e $\bar{\lambda} > 0$ são tais que $X^0 \in L_f(f(q))$ e $\max_{i \in I} L_i < \lambda_k \leq \bar{\lambda}$, o método de ponto proximal, caracterizado no Teorema 4.1, pode ser aplicado para resolver o problema não convexo (4.20).

Observação 4.5 *De acordo com as afirmações 4.1 e 4.2, a função*

$$F(X) = \max\{\ln \det X, -4 \ln \det X + e^{-2\operatorname{tr} X} - e^{-2n}, \operatorname{tr} X^{-1} - n\},$$

no exemplo acima, é não convexa (no sentido Euclideano) quando restrita a qualquer vizinhança aberta e convexa contendo seu minimizador $X^* = I$. Portanto, o método de ponto proximal local clássico não pode ser aplicado para minimizar está função. Da segunda relação em (4.18) *junto com a Afirmação 4.3, concluímos que a função F também é não convexa no sentido Riemanniano. Portanto, o método de ponto proximal Riemanniano também não pode ser aplicado para minimizar está função.*

Capítulo 5

Método de máxima descida para otimização multicritério em variedades Riemannianas

Neste capítulo apresentamos o problema multicritério, condições de otimalidade de primeira ordem para ele e algumas definições básicas relacionadas. Propomos um método de descida para minimizar uma função vetorial em uma variedade Riemanniana completa e, seguindo as idéias de FLIEGE E SVAITER [28] e GRAÑA DRUMMOND E SVAITER [31], uma completa análise de convergência é apresentada. Além disso apresentamos alguns exemplos de variedades Riemannianas completas com curvas geodésicas explicitas e a iteração de descida da seqüência gerada pelo método proposto.

5.1 O problema multicritério

Nesta seção apresetamos o problema multicritério, a condição de otimalidade de primeira ordem e algumas definições relacionadas.

Seja $I := \{1, \ldots, m\}$, $\mathbb{R}^m_+ = \{x \in \mathbb{R}^m : x_i \geq 0, j \in I\}$ e $\mathbb{R}^m_{++} = \{x \in \mathbb{R}^m : x_j > 0, j \in I\}$. Para x, $y \in \mathbb{R}^m_+$, $y \succeq x$ (or $x \preceq y$) diz que $y - x \in \mathbb{R}^m_+$ e $y \succ x$ (ou $x \prec y$) diz que $y - x \in \mathbb{R}^m_{++}$.

Dada uma função vetorial continuamente diferenciável $F : M \to \mathbb{R}^m$, consideramos o problema de encontrar um *ponto Pareto ótimo* de F, i.e., um ponto $p^* \in M$

tal que não existe outro $p \in M$ com $F(p) \preceq F(p^*)$ e $F(p) \neq F(p^*)$. Denotamos este problema irrestrito no contexto Riemanniano como

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

Seja F dada por $F(p) := (f_1(p), \ldots, f_m(p))$. Denotamos a jacobiana Riemanniana de F por

$$\nabla F(p) := (\operatorname{grad} f_1(p), \ldots, \operatorname{grad} f_m(p)), \qquad p \in M,$$

e a imagem da jacobiana Riemanniana de F no ponto $p \in M$ por

$$\operatorname{Im}(\nabla F(p)) := \{\nabla F(p)v = (\langle \operatorname{grad} f_1(p), v\rangle, \ldots, \langle \operatorname{grad} f_m(p), v\rangle) : v \in T_pM\}, \qquad p \in M.$$

Usando a igualdade acima, a condição de otimalidade de primeira ordem para o problema (5.1) é declarado como

$$p \in M, \qquad \operatorname{Im}(\nabla F(p)) \cap (-\mathbb{R}^m_{++}) = \emptyset. \tag{5.2}$$

Observação 5.1 *Note que a condição em* (5.2) *generaliza, para otimização vetorial, a condição clássica* $\nabla F(p) = 0$ *para o caso escalar, i.e.,* $m = 1$.

Em geral, (5.2) é necessária, mas não suficiente para otimalidade. Assim, um ponto $p \in M$ satisfazendo (5.2) é chamado *Pareto crítico.*

5.2 O método

Nesta seção apresentamos o método de máxima descida Riemanniano para resolver problemas multicritérios e estabelecemos a boa definição da seqüência por ele gerada.

Seja $p \in M$ um ponto que não é Pareto crítico. Então, existe uma direção $v \in T_pM$ satisfazendo

$$\nabla F(p)v \in -\mathbb{R}^m_{++},$$

isto é, $\nabla F(p)v \prec 0$. Neste caso, v é chamado uma *direção de descida* para F em p.

Para cada $p \in M$, consideramos o seguinte problema de otimização irrestrito no plano tangente T_pM,

$$\min_{v\in T_pM} \left\{\max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2\right\}, \qquad I = \{1, \ldots, m\}. \tag{5.3}$$

Lema 5.1 *O problema de otimização irrestrito em* (5.3) *tem uma única solução. Além disso, o vetor v é a solução do problema em* (5.3) *se, e somente se, existe $\alpha_i \geq 0$, $i \in I(p,v)$, tal que*

$$v = -\sum_{i\in I(p,v)} \alpha_i \operatorname{grad} f_i(p), \qquad \sum_{i\in I(p,v)} \alpha_i = 1,$$

onde $I(p,v) := \{i \in I : \langle \operatorname{grad} f_i(p), v\rangle = \max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle\}$.

Prova. Visto que a função

$$T_pM \ni v \mapsto \max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle,$$

é o máximo de funções lineares no espaço linear T_pM, ela é convexa. Assim, é fácil ver que a função

$$T_pM \ni v \longmapsto \max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2, \tag{5.4}$$

é fortemente convexa, o que implica que o problema em (5.3) tem uma única solução em T_pM e a primeira afirmação está provada.

Da convexidade da função em (5.4), é conhecido que v é solução do problema em (5.3) se, e somente se,

$$0 \in \partial\left(\max_{i\in I}\langle \operatorname{grad} f_i(p), \,.\,\rangle + (1/2)\|\,.\,\|^2\right)(v),$$

ou equivalentemente,

$$-v \in \partial\left(\max_{i\in I}\langle \operatorname{grad} f_i(p), \,.\,\rangle\right)(v).$$

Portanto, a segunda afirmação segue da fórmula para o subdferencial do máximo de funções convexas (ver [39], Corolário VI.4.3.2). **c.q.d**

Lema 5.2 *Se $p \in M$ não é Pareto crítico de F e v é a solução do problema em* (5.3)*, então*

$$\max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2 < 0.$$

Em particular, v é uma direção de descida.

Prova. Visto que p não é Pareto crítico, existe $0 \neq \hat{v} \in T_pM$ tal que $\nabla F(p)\hat{v} \prec 0$. Em particular,

$$\beta = \max_{i\in I}\langle \operatorname{grad} f_i(p), \hat{v}\rangle < 0.$$

Como $-\beta/\|\hat{v}\|^2 > 0$, fazendo $\bar{v} = (-\beta/\|\hat{v}\|^2)\hat{v}$, obtemos

$$\max_{i\in I}\langle \operatorname{grad} f_i(p), \bar{v}\rangle + (1/2)\|\bar{v}\|^2 = -\frac{\beta^2}{2\|\hat{v}\|^2} < 0,$$

Usando que v é a solução do problema em (5.3), a primeira parte do lema segue da última desigualdade. A segunda parte do lema é uma conseqüência imediata da primeira. **c.q.d**

Em vista dos dois lemas anteriores e (5.3), definimos a função direção de máxima descida para F como segue:

Definição 5.1 *A função direção de máxima descida para F é definida como*

$$M \ni p \longmapsto v(p) := \operatorname{argmin}_{v\in T_pM}\left\{\max_{i\in I}\langle \operatorname{grad} f_i(p), v\rangle + (1/2)\|v\|^2\right\} \in T_pM.$$

Observação 5.2 *Como uma conseqüência imediata do Lema 5.1 segue que a direção de máxima descida para funções vetoriais tornam a direção de máxima descida quando $m = 1$, ver [47], [35], [34] e [33]. No caso $M = \mathbb{R}^n$ recuperamos a direção de máxima descida proposta em [28].*

O *método de máxima descida com regra de Armijo* para resolver o problema de otimização irrestrito (5.1) é como segue:

Método 5.1 (Método de máxima descida com regra de Armijo)

INICIALIZAÇÃO. *Tome $\beta \in (0, 1)$ e $p_0 \in M$. Fixe $k = 0$.*
CRITÉRIO DE PARADA. *Se p^k é Pareto crítico, PARE. Caso contrário.*
PASSO ITERATIVO. *Calcule a direção de máxima descida v^k para F em p^k, i.e.,*

$$v^k := v(p^k), \tag{5.5}$$

e o comprimento de passo $t_k \in (0, 1]$ do seguinte modo:

$$t_k := \max\left\{2^{-j} : j \in \mathbb{N},\, F\left(exp_{p^k}(2^{-j}v^k)\right) \preceq F(p^k) + \beta 2^{-j}\,\nabla F(p^k)v^k\right\}, \tag{5.6}$$

estabeleça

$$p^{k+1} := exp_{p^k}(t_k v^k), \tag{5.7}$$

e volte ao CRITÉRIO DE PARADA.

Observação 5.3 *O método proposto acima recupera o método de máxima descida Riemanniano escalar quando $m = 1$, o qual tem aparecido, por exemplo, em [36], [35], [34], [33].*

Proposição 5.1 *A seqüência $\{p^k\}$ gerada pelo método de máxima descida com regra de Armijo está bem definido.*

Prova. Assuma que p^k não é Pareto crítico. Da Definição 5.1 e Lema 5.1, $v^k = v(p^k)$ está bem definida. Assim, para provar a boa definição do método proposto é suficiente provar a boa definição do comprimento de passo. Para isto, primeiro note que da Definição 5.1 e Lema 5.2

$$\nabla F(p^k)v^k \prec 0.$$

Visto que $F : M \to \mathbb{R}^m$ é uma função vetorial continuamente diferenciável, $\nabla F(p^k)v^k \prec 0$ e $\beta \in (0,1)$, temos

$$\lim_{t \to 0^+} \frac{F\left(exp_{p^k}(tv^k)\right) - F(p^k)}{t} = \nabla F(p^k)v^k \prec \beta \nabla F(p^k)v^k \prec 0.$$

Portanto, é direto mostrar que existe $\delta \in (0,1]$ tal que

$$F\left(exp_{p^k}(tv^k)\right) \prec F(p^k) + \beta t\, \nabla F(p^k)v^k, \qquad t \in (0,\, \delta).$$

Como $\lim_{j\to\infty} 2^{-j} = 0$, última desigualdade vetorial implica que o comprimento de passo (5.6) está bem definido. Consequentemente, p^{k+1} está bem definido e a proprosição está concluída. **c.q.d**

5.3 Análise de convergência

É imediato que se a seqüência gerada pelo Método 5.1 termina depois de um número finito de iterações, ela termina em um Pareto crítico de F. No restante desta seção assumimos que $\{p^k\}$, $\{v^k\}$ e $\{t_k\}$ são seqüências infinitas geradas pelo Método 5.1. Nesta seção, seguindo as idéias de [28], provamos convergência parcial da seqüência $\{p^k\}$ no sentido que cada ponto de acumulação dela é Pareto crítico de F sem qualquer hipótese adicional sobre F além da continuidade diferenciável. Na seqüência, seguindo [31], assumindo quasi-convexidade de F e curvatura não-negativa para M, estendemos para otimização de funções vetoriais o resultado de convergência total apresentado em [36] e [11].

5.3.1 Convergência parcial

Nesta seção provamos que cada ponto de acumulação da seqüência $\{p^k\}$ é Pareto crítico de F. Antes provamos o seguinte fato preliminar importante na prova do referido resultado.

Lema 5.3 *A função direção de máxima descida para F, $M \ni p \mapsto v(p) \in T_pM$, é contínua.*

Prova. Seja $\{q^k\} \subset M$ uma seqüência que converge a $\bar{q}$ e $U_{\bar{q}} \subset M$ uma vizinhança de $\bar{q}$ tal que $TU_{\bar{q}} \approx U_{\bar{q}} \times \mathbb{R}^n$. Visto que $\{q^k\}$ converge a $\bar{q}$ e $TU_{\bar{q}} \subset TM$ é um conjunto aberto, assumimos que a seqüência inteira $\{(q^k, v(q^k))\}$ está em $TU_{\bar{q}}$. Defina $v^k := v(q^k)$. Combinando Definição 5.1 com Lema 5.2 é fácil ver que

$$\|v^k\| \leq 2\max_{i\in I} \|\operatorname{grad} f_j(q^k)\|.$$

Como F é continuamente diferenciável e $\{q^k\}$ é convergente, a desigualdade acima implica que a seqüência $\{v^k\}$ é limitada. Seja $\bar{v}$ um ponto de acumulação da seqüência $\{v^k\}$. Da Definição 5.1 e Lema 5.1, concluímos que existe $\alpha_i^k \geq 0$, $i \in I(q^k, v^k)$, tais que

$$v^k = -\sum_{i\in I(q^k,v^k)} \alpha_i \operatorname{grad} f_i(q^k), \qquad \sum_{i\in I(q^k,v^k)} \alpha_i^k = 1, \qquad k = 0, 1, \ldots. \tag{5.8}$$

onde $I(q^k, v^k) := \{i \in I : \langle \operatorname{grad} f_i(q^k), v^k\rangle = \max_{i\in I}\langle \operatorname{grad} f_i(q^k), v^k\rangle\}$. Usando constantes acima e índices associados, defina a seqüência $\{\alpha^k\}$ como

$$\alpha^k := (\alpha_1^k, \ldots, \alpha_m^k), \qquad \alpha_i^k = 0, \quad i \in I \setminus I(q^k, v^k), \qquad k = 0, 1, \ldots.$$

Seja $\|\,.\,\|_1$ a norma da soma em $\mathbb{R}^m$. Visto que $\sum_{i\in I(q^k,v^k)} \alpha_i^k = 1$, temos $\|\alpha^k\|_1 = 1$ para todo k, o que implica que a seqüência $\{\alpha^k\}$ é limitada. Seja $\bar{\alpha}$ um ponto de acumulação da seqüência $\{\alpha^k\}$ e $\{v^{k_s}\}$, $\{\alpha^{k_s}\}$ subseqüências de $\{v^k\}$ e $\{\alpha^k\}$ respectivamente, tal que

$$\lim_{s\to+\infty} v^{k_s} = \bar{v}, \qquad \lim_{s\to+\infty} \alpha^{k_s} = \bar{\alpha}.$$

Como o conjunto de índices I é finito e $I(q^{k_s}, v^{k_s}) \subset I$ para todo s, assumimos, sem perda de generalidade, que

$$I(q^{k_1}, v^{k_1}) = I(q^{k_2}, v^{k_2}) = \ldots = \bar{I}. \tag{5.9}$$

Assim, concluímos de (5.8) e da última igualdade que

$$v^{k_s} = -\sum_{i\in\bar{I}} \alpha_i^{k_s} \operatorname{grad} f_i(q^{k_s}), \qquad \sum_{i\in\bar{I}} \alpha_i^{k_s} = 1, \qquad s = 0, 1, \ldots.$$

Fazendo s ir a $+\infty$ na igualdade acima, obtemos

$$\bar{v} = \sum_{i\in\bar{I}} \bar{\alpha}_i \operatorname{grad} f_i(\bar{q}), \qquad \sum_{i\in\bar{I}} \bar{\alpha}_i = 1. \tag{5.10}$$

Por outro lado, $I(q^{k_s}, v^{k_s}) = \{i \in I : \langle \operatorname{grad} f_i(q^{k_s}), v^{k_s}\rangle = \max_{i\in I}\langle \operatorname{grad} f_i(q^{k_s}), v^{k_s}\rangle\}$. Assim, equação (5.9) implica que

$$\langle \operatorname{grad} f_i(q^{k_s}), v^{k_s}\rangle = \max_{i\in I}\langle \operatorname{grad} f_i(q^{k_s}), v^{k_s}\rangle, \qquad i \in \bar{I}, \qquad s = 0, 1, \ldots.$$

Usando continuidade de ∇F e última igualdade, temos

$$\langle \operatorname{grad} f_i(\bar{q}), \bar{v}\rangle = \max_{i\in I}\langle \operatorname{grad} f_i(\bar{q}), \bar{v}\rangle, \qquad i \in \bar{I}.$$

Da definição de $I(\bar{q}, \bar{v})$, obtemos $\bar{I} \subset I(\bar{q}, \bar{v})$. Portanto, combinando novamente Definição 5.1 com Lema 5.1 e (5.10), concluímos que $\bar{v} = v(\bar{q})$ e o resultado desejado está provado. **c.q.d**

No próximo resultado usamos apenas que F é continuamente diferenciável para assegurar que a seqüência dos valores funcionais da seqüência $\{p^k\}$, $\{F(p^k)\}$, é monótona decrescente e que seus pontos de acumulação são Pareto críticos.

Teorema 5.1 *As seguintes afirmações ocorrem:*

i) $\{F(p^k)\}$ *é monótona decrescente;*

ii) Cada ponto de acumulação da seqüência $\{p^k\}$ *é um ponto Pareto crítico.*

Prova. O passo iterativo no Método 5.1 implica que

$$F(p^{k+1}) \preceq F(p^k) + \beta t_k \nabla F(p^k) v^k, \qquad p^{k+1} = \exp_{p^k} t_k v^k, \qquad k = 0, 1, \ldots. \tag{5.11}$$

Visto que $\{p^k\}$ é uma seqüência infinita, para todo k, p^k não é Pareto crítico de F. Assim, item i segue da definição de v^k junto com a Definição 5.1, Lema 5.2 e última desigualdade vetorial.

Seja $\bar{p} \in M$ um ponto de acumulação da seqüência $\{p^k\}$ e $\{p^{k_s}\}$ uma subseqüência de $\{p^k\}$ tal que $\lim_{s\to+\infty} p^{k_s} = \bar{p}$. Visto que F é contínua e $\lim_{s\to+\infty} p^{k_s} = \bar{p}$, temos

$\lim_{s\to+\infty} F(p^{k_s}) = F(\bar{p})$. Assim, levando em conta que $\{F(p^k)\}$ é uma seqüência decrescente e tem $F(\bar{p})$ como um ponto de acumulação, é fácil concluir que a seqüência inteira $\{F(p^k)\}$ converge a $F(\bar{p})$. Usando a equação (5.11), Definição 5.1 e Lema 5.2, concluímos que

$$F(p^{k+1}) - F(p^k) \preceq \beta t_k \nabla F(p^k) v^k \preceq 0, \qquad k = 0, 1, \ldots.$$

Visto que $\lim_{s\to+\infty} F(p^k) = F(\bar{p})$, a última desigualdade implica que

$$\lim_{k\to+\infty} \beta t_k \nabla F(p^k) v^k = 0. \tag{5.12}$$

Como $\{p^{k_s}\}$ converge a $\bar{p}$, assumimos que $\{(p^{k_s}, v^{k_s})\} \subset TU_{\bar{p}}$, onde $U_{\bar{p}}$ é uma vizinhança de $\bar{p}$ tal que $TU_{\bar{p}} \approx U_{\bar{p}} \times \mathbb{R}^n$. Além disso, como a seqüência, $\{t_k\} \subset (0,1]$ tem um ponto de acumulação $\bar{t} \in [0,1]$, assumimos, sem perda de generalidade, que $\{t_{k_s}\}$ converge a $\bar{t}$. Temos duas possibilidades a considerar:

a) $\bar{t} > 0$;

b) $\bar{t} = 0$.

Assuma que o item **a** ocorre. Neste caso, de (5.12), continuidade de ∇F, (5.5) e Lema 5.3, obtemos

$$\nabla F(\bar{p}) v(\bar{p}) = 0,$$

o que implica que

$$\max_{i\in I} \langle \operatorname{grad} f_i(\bar{p}), v(\bar{p}) \rangle = 0. \tag{5.13}$$

Por outro lado, da Definição 5.1 junto com Lema 5.2,

$$\max_{i\in I} \langle \operatorname{grad} f_i(p^{k_s}), v^{k_s} \rangle + (1/2)\|v^{k_s}\|^2 < 0.$$

Fazendo s ir a $+\infty$ na desigualdade acima e usando Lema 5.3 combinado com a continuidade de ∇F e igualdade (5.13), concluímos que

$$\max_{i\in I} \langle \operatorname{grad} f_i(\bar{p}), v(\bar{p}) \rangle + (1/2)\|v(\bar{p})\|^2 = 0.$$

Consequentemente, segue da última igualdade, Definição 5.1 e Lema 5.2 que $\bar{p}$ é um Pareto crítico.

Agora, assuma que o item **b** ocorre. Visto que para todo s p^{k_s} não é um Pareto crítico, temos

$$\max_{i\in I} \langle \operatorname{grad} f_i(p^{k_s}), v^{k_s} \rangle \le \max_{i\in I} \langle \operatorname{grad} f_i(p^{k_s}), v^{k_s} \rangle + (1/2)\|v^{k_s}\|^2 < 0,$$

onde a última desigualdade é conseqüência da Definição 5.1 junto com Lemma 5.2. Portanto, fazendo s ir a $+\infty$ na última desigualdade, usando (5.5) e Lema 5.3, obtemos

$$\max_{i\in I}\langle \text{grad}\, f_i(\bar{p}), v(\bar{p})\rangle \leq \max_{i\in I}\langle \text{grad}\, f_i(\bar{p}), v(\bar{p})\rangle + (1/2)\|v(\bar{p})\|^2 \leq 0. \qquad (5.14)$$

Tome $r \in \mathbb{N}$. Visto que $\{t_{k_s}\}$ converge a $\bar{t} = 0$, concluímos que, para s suficientemente grande,

$$t_{k_s} < 2^{-r}.$$

De (5.6) isto diz que a condição de Armijo (5.11) não é satisfeita para $t = 2^{-r}$, i.e.,

$$F(\exp_{p^k}(2^{-j}v^{k_s})) \npreceq F(p^{k_s}) + \beta 2^{-r}\nabla F(p^{k_s})v^{k_s}.$$

Mas isto nos diz que existe pelo menos um $i_0 \in I$ tal que

$$f_{i_0}(\exp_{p^{k_s}}(2^{-r}v^{k_s})) > f_{i_0}(p^{k_s}) + \beta 2^{-r}\langle \text{grad}\, f_{i_0}(p^{k_s}), v^{k_s}\rangle.$$

Fazendo s ir a $+\infty$ na desigualdade acima, levando em conta que ∇F e exp são contínuas e usando Lema 5.3, obtemos

$$f_{i_0}(\exp_{\bar{p}}(2^{-r}v(\bar{p}))) \geq f_{i_0}(\bar{p}) + \beta 2^{-r}\langle \text{grad}\, f_{i_0}(\bar{p}), v(\bar{p})\rangle.$$

A última desigualdade é equivalente a

$$\frac{f_{i_0}(\exp_{\bar{p}}(2^{-r}v(\bar{p}))) - f_{i_0}(\bar{p})}{2^{-r}} \geq \beta\langle \text{grad}\, f_{i_0}(\bar{p}), v(\bar{p})\rangle,$$

que, fazendo r ir a $+\infty$ e usando que $0 < \beta < 1$, fornece $\langle \text{grad}\, f_{i_0}(\bar{p}), v(\bar{p})\rangle \geq 0$. Assim,

$$\max_{i\in I}\langle \text{grad}\, f_i(\bar{p}), v(\bar{p})\rangle \geq 0.$$

Combinando a última desigualdade com (5.14), temos

$$\max_{i\in I}\langle \text{grad}\, f_i(\bar{p}), v(\bar{p})\rangle + (1/2)\|v(\bar{p})\|^2 = 0.$$

Logo, novamente da Definition 5.1 e Lema 5.2, segue que $\bar{p}$ é um Pareto crítico e a prova está concluída. **c.q.d**

Observação 5.4 *Se a seqüência $\{p^k\}$ inicia em um conjunto de nível limitado, por exemplo, se*

$$L_F(F(p_0)) := \{p \in M : F(p) \preceq F(p_0)\},$$

é um conjunto limitado, e sendo F uma função contínua, o teorema de Hopf-Rinow assegura que $L_F(F(p_0))$ é um conjunto compacto. Assim, item i do Teorema 5.1 implica que $\{p^k\} \subset L_F(F(p_0))$ e, consequentemente, que $\{p^k\}$ é limitada. Em particular, $\{p^k\}$ tem pelo menos um ponto de acumulação. Portanto, Teorema 5.1 estende para otimização vetorial o resultado do Teorema 5.1 *de [36], ver também Remark* 4.5 *de [48].*

5.3.2 Convergência total

Nesta seção, com respeito às hipóteses de quasi-convexidade sobre a função F e curvatura não-negativa sobre M, convergência total do método de máxima descida é obtida.

Definição 5.2 *Seja $H : M \to \mathbb{R}^m$ uma função vetorial.*

i) H é chamada convexa em M se, para cada $p, q \in M$ e cada segmento geodésico $\gamma : [0,1] \to M$ ligando p a q (i.e., $\gamma(0) = p$ e $\gamma(1) = q$), ocorre

$$H(\gamma(t)) \preceq (1-t)H(p) + tH(q), \qquad t \in [0,1].$$

ii) H é chamada quasi-convexa em M se, para cada $p, q \in M$ e cada segmento geodésico $\gamma : [0,1] \to M$ ligando p a q, ocorre

$$H(\gamma(t)) \preceq \max\{H(p), H(q)\}, \qquad t \in [0,1],$$

onde o máximo é considerado coordenada a coordenada.

Observação 5.5 *A primeira definição acima é uma extensão natural da definição de convexidade enquanto que a segunda é uma extensão de uma caracterização da definição de quasi-convexidade, do espaço Euclideano ao contexto Riemanniano. Ver Definição* 6.2 *e Corolário* 6.6 *de [49], pag.* 29 *e* 31, *respectivamente. Note que as definições acima são equivalentes, respectivamente, a H ser convexa e quasi-convexa ao longo de cada segmento geodésico $\gamma : [a,b] \to M$, $a < b$. Assim, quando $m = 1$ essas definições coincidem com as definições de convexidade e quasi-convexidade em [33], respectivamente. Além disso, é imediato das definições acima que se H é convexa então ela é quasi-convexa. No caso que H é diferenciável, convexidade de*

H implica que para cada $p, q \in M$ e cada segmento geodésico $\gamma : [0,1] \to M$ tal que $\gamma(0) = p$ e $\gamma(1) = q$,

$$\nabla H(p)\gamma'(0) \preceq H(q) - H(p).$$

Proposição 5.2 *Seja $H : M \to \mathbb{R}^m$ uma função quasi-convexa diferenciável. Então, para cada $p, q \in M$ e cada segmento geodésico $\gamma : [0,1] \to M$ ligando p a q, ocorre*

$$H(q) \preceq H(p) \quad \Rightarrow \quad \nabla H(p)\gamma'(0) \preceq 0.$$

Prova. Tome $p, q \in M$ tais que $H(q) \preceq H(p)$ e um segmento geodésico $\gamma : [0,1] \to M$ tal que $\gamma(0) = p$ e $\gamma(1) = q$. Visto que H é quasi-convexa, temos

$$H(\gamma(t)) \preceq H(p), \qquad t \in [0,1].$$

Usando a última desigualdade o resultado é uma conseqüência imediata da diferenciabilidade de H. **c.q.d**

Conhecemos que criticalidade é condição necessária mas não suficiente para otimalidade. Porém, com respeito à convexidade da função vetorial F, provamos que criticalidade é equivalente à otimalidade fraca.

Definição 5.3 *Um ponto $p^* \in M$ é Pareto ótimo fraco de F se não existe $p \in M$ com $F(p) \prec F(p^*)$.*

Proposição 5.3 *Seja $H : M \to \mathbb{R}^m$ uma função convexa diferenciável. Então, $p \in M$ é um Pareto crítico de H, i.e.,*

$$Im(\nabla H(p)) \cap (-\mathbb{R}^m_{++}) = \emptyset,$$

se, e somente se, p é Pareto ótimo fraco de H.

Prova. Suponhamos que p é Pareto crítico de H e assumamos, por contradição, que p não é Pareto ótimo fraco de H. Visto que p não é Pareto ótimo fraco de H, existe $\tilde{p} \in M$ tal que

$$H(\tilde{p}) \prec H(p). \tag{5.15}$$

Seja $\gamma : [0,1] \to M$ segmento geodésico ligando p a $\tilde{p}$ (i.e., $\gamma(0) = p$ e $\gamma(1) = \tilde{p}$). Como H é diferenciável e convexa, a última parte da Observação 5.5 e (5.15) implicam que

$$\nabla H(p)\gamma'(0) \preceq H(\tilde{p}) - H(p) \prec 0.$$

Mas isto contradiz o fato de p ser Pareto crítico de H, e a primeira parte está concluída.

Agora, suponhamos que p é Pareto ótimo fraco de H e assumamos, por contradição, que p não é Pareto crítico de H. Visto que p não é Pareto crítico, então $\mathrm{Im}(\nabla H(p)) \cap (-\mathbb{R}^m_{++}) \neq \emptyset$, i.e., existe $v \in T_pM$ uma direção de descida para F em p. Assim, da diferenciabilidade de H, temos

$$\lim_{t \to 0^+} \frac{H\left(exp_p(tv)\right) - H(p)}{t} = \nabla H(p)v \prec 0,$$

o que implica que existe $\delta > 0$ tal que

$$H\left(exp_p(tv)\right) \prec H(p) + t\nabla H(p)v, \qquad t \in (0,\, \delta).$$

Como v é uma direção de descida para F em p e $t \in (0,\, \delta)$, temos $t\nabla H(p)v \prec 0$. Portanto, a última desigualdade vetorial fornece

$$H\left(exp_p(tv)\right) \prec H(p), \qquad t \in (0,\, \delta),$$

contradizendo o fato de p ser Pareto ótimo fraco de H, o que conclui a prova. **c.q.d**

Definição 5.4 *Uma seqüência $\{q^k\} \subset M$ é quasi-Fejér convergente a um conjunto não vazio U se, para todo $p \in U$, existe uma seqüência $\{\epsilon_k\} \subset \mathbb{R}_+$ tal que*

$$\sum_{k=0}^{+\infty} \epsilon_k < +\infty, \qquad d^2(q^{k+1}, q) \leq d^2(q^k, q) + \epsilon_k, \qquad k = 0, 1, \ldots .$$

No próximo lema recordamos o chamado teorema da convergência quasi-Fejér.

Lema 5.4 *Seja $U \subset M$ um conjunto não vazio e $\{q^k\} \subset M$ uma seqüência quasi-Fejér convergente. Então, $\{q^k\}$ é limitada. Além disso, se um ponto de acumulação $\bar{q}$ de $\{q^k\}$ pertence a U, então a seqüência inteira $\{q^k\}$ converge a $\bar{q}$.*

Prova. Análoga à prova do Teorema 1 em Burachik et al. [29], trocando a distância Euclideana pela distância Riemanniana d. **c.q.d**

Considere o seguinte conjunto

$$U := \{p \in M : F(p) \preceq F(p^k), \;\; k = 0, 1, \ldots\}. \tag{5.16}$$

Em geral, o conjunto acima pode ser vazio. Para garantir que U é não vazio, uma hipótese adicional sobre a seqüência $\{p^k\}$ é necessária. Na próxima observação damos uma tal condição.

Observação 5.6 *Se a seqüência $\{p^k\}$ tem um ponto de acumulação, então U é não vazio. De fato, seja $\bar{p}$ um ponto de acumulação da seqüência $\{p^k\}$. Então, existe uma subseqüência $\{p^{k_j}\}$ de $\{p^k\}$ que converge a $\bar{p}$. Visto que F é contínua $\{F(p^k)\}$ tem $F(\bar{p})$ como um ponto de acumulação. Assim, usando que $\{F(p^k)\}$ é uma seqüência decrescente (ver item i do Teorema 5.1) argumentos usuais mostram facilmente que a seqüência inteira $\{F(p^k)\}$ converge a $F(\bar{p})$ e ocorre*

$$F(\bar{p}) \preceq F(p^k), \qquad k = 0, 1, \ldots,$$

o que implica que $\bar{p} \in U$, i.e., $U \neq \emptyset$.

No próximo lema apresentamos o resultado chave desta seção. Ele é fundamental na prova da convergência global da seqüência $\{p^k\}$.

Lema 5.5 *Suponha que F é quasi-convexa, M tem curvatura não negativa e U, definido em* (5.16)*, é não vazio. Então, para todo $\tilde{p} \in U$, a seguinte desigualdade ocorre:*

$$d^2(p^{k+1}, \tilde{p}) \leq d^2(p^k, \tilde{p}) + t_k^2 \|v^k\|^2.$$

Prova. Considere ângulo geodésico $(\gamma_1, \gamma_2, \alpha)$, onde γ_1 é um segmento geodésico minimal normalizado ligando p^k a $\tilde{p}$, γ_2 é o segmento geodésico ligando p^k to p^{k+1} tal que $\gamma_2'(0) = t_k v^k$ e $\alpha = \angle(\gamma_1'(0), v^k)$. Pela lei dos cossenos (Teorema 1.1), temos

$$d^2(p^{k+1}, \tilde{p}) \leq d^2(p^k, \tilde{p}) + t_k^2 \|v^k\|^2 - 2d(p^k, p) t_k \|v^k\| \cos\alpha, \qquad k = 0, 1, \ldots.$$

Assim, levando em conta que $\cos(\pi - \alpha) = -\cos\alpha$ e $\langle -v^k, \gamma_1'(0) \rangle = \|v^k\| \cos(\pi - \alpha)$, desigualdade acima se torna

$$d^2(p^{k+1}, \tilde{p}) \leq d^2(p^k, \tilde{p}) + t_k^2 \|v^k\|^2 + 2d(p^k, \tilde{p}) t_k \langle -v^k, \gamma_1'(0) \rangle, \qquad k = 0, 1, \ldots.$$

Por outro lado, de (5.5), da Definição 5.1 e do Lema 5.1, existe $\alpha_i^k \geq 0$, com $i \in I_k := I(p^k, v^k)$, tal que

$$v^k = -\sum_{i \in I_k} \alpha_i^k \operatorname{grad} f_i(p^k), \qquad \sum_{i \in I_k} \alpha_i^k = 1, \qquad k = 0, 1, \ldots.$$

Assim, última desigualdade rende

$$d^2(p^{k+1}, \tilde{p}) \leq d^2(p^k, \tilde{p}) + t_k^2 \|v^k\|^2 + 2d(p^k, \tilde{p}) t_k \sum_{i \in I_k} \alpha_i^k \langle \operatorname{grad} f_i(p^k), \gamma_1'(0) \rangle,$$

$$k = 0, 1, \ldots. \quad (5.17)$$

Visto que F é quasi-convexa e $\tilde{p} \in U$, da Proposição 5.2 com $H = F$, $p = p^k$, $q = \tilde{p}$ e $\gamma = \gamma_1$, temos

$$\nabla F(p^k)\gamma_1'(0) \preceq 0, \qquad k = 0, 1, \dots,$$

ou equivalentemente,

$$\langle \operatorname{grad} f_i(p^k), \gamma_1'(0) \rangle \leq 0, \qquad i = 1, \dots, m, \quad k = 0, 1, \dots. \tag{5.18}$$

Portanto, combinando (5.17) com (5.18), o lema segue. **c.q.d**

Proposição 5.4 *Se F é quasi-convexa, M tem curvatura não negativa e U, definido em* (5.16)*, é um conjunto não vazio, então a seqüência $\{p^k\}$ é quasi-Fejér convergente a U.*

Prova. Para simplificar notações, defina a função escalar $\varphi : \mathbb{R}^m \to \mathbb{R}$ como segue

$$\varphi(y) = \max_{i \in I} \langle y, e_i \rangle, \qquad I = \{1, \dots, m\},$$

onde $\{e_i\} \subset \mathbb{R}^m$ é a base canônica do espaço $\mathbb{R}^m$. É fácil ver que as seguintes propriedades sobre a função φ ocorrem:

$$\varphi(x + y) \leq \varphi(x) + \varphi(y), \qquad \varphi(tx) = t\varphi(x), \qquad x, y \in \mathbb{R}^m, \quad t \geq 0. \tag{5.19}$$

$$x \preceq y \quad \Rightarrow \quad \varphi(x) \leq \varphi(y), \qquad x, y \in \mathbb{R}^m. \tag{5.20}$$

Da definição de t_k em (5.6) e p^{k+1} em (5.7), temos

$$F(p^{k+1}) \preceq F(p^k) + \beta t_k \nabla F(p^k) v^k, \qquad k = 0, 1 \dots.$$

Assim, usando (5.19), (5.20) e a última desigualdade, obtemos

$$\varphi(F(p^{k+1})) \leq \varphi(F(p^k)) + \beta t_k \varphi(\nabla F(p^k) v^k), \qquad k = 0, 1 \dots. \tag{5.21}$$

Por outro lado, combinando definição de v^k em (5.5), Definição 5.1, Lema 5.2 e definição de φ, concluímos que

$$\varphi(\nabla F(p^k) v^k) + (1/2)\|v^k\|^2 < 0, \qquad k = 0, 1 \dots,$$

que, junto com (5.21), implica que

$$\varphi(F(p^{k+1})) < \varphi(F(p^k)) - (\beta t_k/2)\|v^k\|^2, \qquad k = 0, 1 \dots. \tag{5.22}$$

Mas isto nos diz que

$$t_k\|v^k\|^2 < 2[\varphi(F(p^k)) - \varphi(F(p^{k+1}))]/\beta, \qquad k = 0, 1 \ldots.$$

Como $t_k \in (0, 1]$, é imediato que

$$t_k^2\|v^k\|^2 < 2[\varphi(F(p^k)) - \varphi(F(p^{k+1}))]/\beta, \qquad k = 0, 1 \ldots.$$

Assim, da última desigualdade obtemos

$$\sum_{k=0}^{n} t_k^2\|v^k\|^2 < 2\left[\varphi(F(p^0) - \varphi(F(p^{n+1}))\right]/\beta, \quad n > 0.$$

Tome $\bar{p} \in U$. Então, $F(\bar{p}) \preceq F(p^{n+1})$. Portanto, de (5.20), $\varphi(F(\bar{p})) \leq \varphi(F(p^{n+1}))$ e a última desigualdade rende

$$\sum_{k=0}^{n} t_k^2\|v^k\|^2 < 2\left[\varphi(F(p^0) - \varphi(F(p))\right]/\beta, \quad n > 0.$$

Mas isto implica que $\{t_k^2\|v^k\|^2\}$ é uma seqüência somável. Portanto, o resultado desejado segue do Lema 5.5 combinado com Definição 5.4. **c.q.d**

Teorema 5.2 *Se F é quasi-convexa, M tem curvatura não negativa e U, definido em* (5.16)*, é um conjunto não vazio, então a seqüência $\{p^k\}$ converge a um Pareto crítico de F.*

Prova. Da Proposição 5.4, $\{p^k\}$ é Fejér convergente a U. Assim, Lema 5.4 garante que $\{p^k\}$ é limitada e, do teorema de Hopf-Rinow, existe $\{p^{k_s}\}$, subseqüência de $\{p^k\}$, que converge a $\bar{p} \in M$. Visto que F é contínua e $\{F(p^k)\}$ é uma seqüência decrescente (ver item i do Teorema 5.1), concluímos que $F(p^k)$ converge a $F(\bar{p})$ quando k tende a $+\infty$, o que implica que

$$F(\bar{p}) \preceq F(p^k), \qquad k = 0, 1, \ldots,$$

i.e., $\bar{p} \in U$. Portanto, do Lema 5.4, concluímos que a seqüência inteira $\{p^k\}$ converge a $\bar{p}$, e a conclusão da prova é conseqüência do item ii do Teorema 5.1. **c.q.d**

Corolário 5.1 *Se F é convexa, M tem curvatura não negativa e U, definido em* (5.16)*, é um conjunto não vazio, então a seqüência $\{p^k\}$ converge a um Pareto ótimo fraco de F.*

Prova. Visto que F é convexa, em particular, ela é quasi-convexa (ver Observação 5.5). Portanto, o corolário é uma conseqüência do prévio teorema combinado com a Proposição 5.3. **c.q.d**

5.4 Exemplos

Nesta seção apresentamos alguns exemplos de variedades Riemannianas completas com curvas geodésicas explicitas e a iteração de descida da seqüência gerada pelo Método 5.1. Recordamos que $F : M \to \mathbb{R}^m$, $F(p) = (f_1(p), \ldots, f_m(p))$, é uma função diferenciável. Se (M, G) é uma variedade Riemanniana então o gradiente Riemanniano de f_j é dado por $\operatorname{grad} f_i(p) = G(p)^{-1} f_i'(p)$, $i \in I := \{1, \ldots, n\}$. Portanto, se $v(p)$ é a direção de máxima descida para F em p (ver Definição 5.1), do Lema 5.1, existem constantes $\alpha_i \geq 0$, $i \in I(p, v)$, tais que

$$v = -\sum_{i \in I(p,v)} \alpha_i G(p)^{-1} f_i'(p), \qquad \sum_{i \in I(p,v)} \alpha_i = 1, \tag{5.23}$$

onde $I(p, v) := \{i \in I : \langle G(p)^{-1} f_i'(p), v\rangle = \max_{i \in I} \langle G(p)^{-1} f_i'(p), v\rangle\}$.

5.4.1 Método de máxima descida para o octante positivo

Seja M o octante positivo, $\mathbb{R}^n_{++}$, munido com a métrica Riemanniana

$$M \ni v \mapsto G(p) = P^{-2} := \operatorname{diag}\left(p_1^{-2}, \ldots, p_n^{-2}\right),$$

(métrica induzida pela Hessiana da barreira logarítmica). Visto que (M, G) é isométrica ao espaço Euclideano munido com a métrica usual (ver, Da Cruz Neto et al. [6]) segue que M tem curvatura constante e igual a zero. Por outro lado, é fácil ver que a única geodésica $p = p(t)$ tal que $p(0) = p^0 = (p_1^0, \ldots, p_n^0)$ e $p'(0) = v^0 = (v_1^0, \ldots, v_n^0)$ é dada por $p(t) = (p_1(t), \ldots, p_n(t))$, onde

$$p_j(t) = p_j^0 \mathrm{e}^{\left(v_j^0/p_j^0\right)t}, \qquad j = 1, \ldots, n. \tag{5.24}$$

Assim, concluímos que (M, G) é também completa. Neste caso, de (5.24) e (5.23), existem $\alpha_i^k \geq 0$, $i \in I(p^k, v^k)$, tais que a iteração de máxima descida da seqüência generada pelo Método 5.1 é dada por

$$p_j^{k+1} = p_j^k \mathrm{e}^{\left(v_j^k/p_j^k\right)t_k}, \quad v_j^k = -\sum_{i \in I(p^k, v^k)} \alpha_i^k (p_j^k)^2 \frac{\partial f_i}{\partial p_j}(p^k), \quad \sum_{i \in I(p^k, v^k)} \alpha_i^k = 1, \; i = 1, \ldots, n.$$

5.4.2 Método de máxima descida para o hipercubo

Seja M o hipercubo $(0,1) \times \ldots \times (0,1)$ munido com a métrica Riemanniana

$$M \ni v \mapsto G(p) = P^{-2}(I - P)^{-2} := \operatorname{diag}\left((p_1)^2(1 - p_1)^2, \ldots, (p_n)^2(1 - p_n)^2\right),$$

(métrica induzida pela Hessiana da barreira $b(p) = \sum_{i=1}^{n}(2p_i - 1)\big(\ln p_i - \ln(1-p_i)\big)$. A variedade Riemanniana (M, G) é completa e a geodésica $p = p(t)$, satisfazendo $p(0) = p^0 = (p_1^0, \ldots, p_n^0)$ e $p'(0) = v^0 0 = (v_1^0, \ldots, v_n^0)$, é dada por $p(t) = (p_1(t), \ldots, p_n(t))$,

$$p_j(t) = (1/2)\left[1 + \tanh\left((1/2)\,\frac{v_j}{p_j(1-p_j)}t + (1/2)\ln\left(\frac{p_j}{1-p_j}\right)\right)\right], \qquad j = 1, ..., n, \tag{5.25}$$

onde $\tanh(z) := (e^z - e^{-z})/(e^z + e^{-z})$. Além disso, (M, G) tem curvatura constante e igual zero, ver Teorema 3.1 e 3.2 de [50]. Neste caso, de (5.25) e (5.23), existem $\alpha_i^k \geq 0$, $i \in I(p^k, v^k)$, tais que a iteração de máxima descida da seqüência generada pelo Método 5.1 é dada por

$$p_j^{k+1} = (1/2)\left[1 + \tanh\left((1/2)\,\frac{v_j^k}{p_j^k(1-p_j^k)}t_k + (1/2)\ln\left(\frac{p_j^k}{1-p_j^k}\right)\right)\right], \qquad j = 1, ..., n,$$

com,

$$v_j^k = -\sum_{i \in I(p^k, v^k)} \alpha_i^k (p_j^k)^2 (1-p_j^k)^2 \frac{\partial f_i}{\partial p_j}(p^k), \qquad \sum_{i \in I(p^k, v^k)} \alpha_i^k = 1, \qquad j = 1, \ldots, n.$$

5.4.3 Método de máxima descida para o cone das matrizes simétricas definidas positivas

Seja $M = \mathbb{S}_{++}^n$ o cone das matrizes simétricas definidas positivas munida com a métrica Riemanniana induzida pela Hessiana Euclideana de $\Psi(X) = -\ln\det X$, i.e., $G(X) := \Psi''(X)$. Conforme mensionado no Exemplo 4.1.2, (M, G) é uma variedade de Hadamard (com curvatura não identicamente zero) e o único segmento geodésico conectando quaisquer $X, Y \in M$ é dado por

$$X(t) = X^{1/2}\left(X^{-1/2} Y X^{-1/2}\right)^t X^{1/2}, \qquad t \in [0, 1],$$

Em particular, a única geodésica $X = X(t)$ tal que $X(0) = X$ e $X'(0) = V$ é dada por

$$X(t) = X^{1/2} \mathrm{e}^{tX^{-1/2} V X^{-1/2}} X^{1/2}. \tag{5.26}$$

Assim, de (5.26) e (5.23), existem $\alpha_i^k \geq 0$, $i \in I(p^k, v^k)$, tais que a iteração de máxima descida da seqüência generada pelo Método 5.1 é dada por

$$X^{k+1} = (X^k)^{1/2} \mathrm{e}^{t_k (X^k)^{-1/2} V^k (X^k)^{-1/2}} (X^k)^{1/2},$$

com,

$$V^k = -\sum_{i \in I(X^k, V^k)} \alpha_i^k X^k f_i'(X^k) X^k, \qquad \sum_{i \in I(X^k, V^k)} \alpha_i^k = 1.$$

Observação 5.7 *Com respeito à hipótese de convexidade sobre a função vetorial F, se (M, G) é a variedade Riemanniana do primeiro ou do segundo exemplo, então Corolário 5.1 assegura a convergência total da seqüência generada pelo Método 5.1. Este fato não necessariamente ocorre se (M, G) é a variedade Riemanniana do último exemplo, visto que neste caso (M, G) tem curvatura não positiva, i.e., $K \leq 0$. Porém, o Teorema 5.1 assegura pelo menos convergência parcial.*

Capítulo 6

Considerações Finais

Neste trabalho temos estendido o campo de aplicação do método de ponto proximal para resolver uma classe particular de problemas de otimização não convexos em variedades de Hadamard, a saber, no caso que a função objetivo é dada pelo máximo de uma certa classe de funções continuamente diferenciáveis. Temos certificado, através de exemplos, que a classe de problemas de minimização para o qual o método de ponto proximal local pode ser aplicado é diferente daquela para a qual tanto o método de ponto proximal clássico quanto o método de ponto proximal Riemanniano são aplicados. Sobre uma outra perspectiva temos também extendido o método de máxima descida com regra de Armijo para otimização multicritério ao contexto Riemanniano. A convergência total é obtida com respeito à hipótese de quasi-convexidade da função multicritério e curvatura não-negativa da variedade Riemanniana.

Possibilidades de pesquisa futura:

1. A derivada direcional e o subdiferencial generalizados

No Capítulo 3 apresentamos uma generalização da derivada direcional e do subdiferencial de uma função localmente Lipschitz não necessariamente convexa. Algumas propriedades e conceitos foram obtidas, porém nos limitamos àquelas que seriam úteis no restante do trabalho desenvolvido até o momento. Uma possibilidade de pesquisa futura seria a extensão de alguns resultados da análise não diferenciável desenvolvida sobre o ambiente Euclideano ao ambiente Riemanniano, ou mesmo de resultados já obtidos no ambiente Riemanniano que dependem de certas hipóteses restritivas sobre os elementos envolvidos. Um exemplo seria a Proposição 3.2 i, a qual caracteriza a derivada direcional de uma função convexa como a função suporte de seu subdiferencial. Um resultado que precede este é o seguinte: o subdiferencial de uma função convexa é não vazio, convexo e compacto. Propomos então provar o seguinte resultado:

Teorema 6.1 *Se $f : M \to \mathbb{R}$ é uma função localmente Lipschitz em um dado ponto $p \in M$, então*

i) $\partial^\circ f(p)$ é não vazio, convexo e compacto;

ii) $f^\circ(p, v) = \max_{s \in \partial^\circ f(p)} \langle s, v \rangle$, para todo $v \in T_pM$;

No caso particular que f é como no Teorema 4.1, da Proposição 2.6 f é localmente Lipschitz e, com uma técnica similar à utilizada na prova da segunda parte da Proposição 3.5, temos que

$$\operatorname{conv}\{\operatorname{grad} \varphi_\tau(p) : \tau \in T(p)\} \subset \partial^\circ f(p),$$

o que prova que $\partial^\circ f(p)$ é não vazio. Os demais fatos do item i seguem como feito no caso Euclideano para f localmente Lipschitz em p arbitrária. No caso geral, a prova tanto da primeira parte do item i como do item ii, é uma simples aplicação do Teorema de Hahn Banach (como ocorre no caso Euclideano), desde que a aplicação $v \mapsto f^\circ(p, v)$ seja positivamente homogênea e subaditiva sobre T_pM. A primeira propriedade é direta, porém a segunda está em aberto. Assim, afim de garantirmos a validade do resultado apresentado no Teorema 6.1, devemos provar o seguinte:

Teorema 6.2 *Se $f : M \to \mathbb{R}$ é uma função localmente Lipschitz em um dado ponto $p \in M$, então a aplicação $v \mapsto f^\circ(p, v)$ é subaditiva sobre T_pM.*

Como uma aplicação do Teorema 6.1, poderíamos provar o seguinte resultado:

Teorema 6.3 *Se $f_i : M \to \mathbb{R}$ é uma função localmente Lipschitz em um dado ponto $p \in M$ para $i = 1, 2, ..., m$, então para escalares $\lambda_i \in \mathbb{R}_+$,*

$$\partial^\circ(\sum_{i=1}^{m} \lambda_i f_i)(x) \subset \sum_{i=1}^{m} \lambda_i \partial^\circ f_i(x). \tag{6.1}$$

Seja $\Omega \subset M$ e $f : M \to \mathbb{R}$ uma função convexa sobre Ω. Recordemos que, conforme Definição 3.1, a derivada direcional de f em um dado ponto $p \in \Omega$ na direção $v \in T_pM$ é dada por

$$f'(p,v) = \lim_{t \to 0^+} \frac{f(\exp_p tv) - f(p)}{t} = \inf_{t>0} \frac{f(\exp_p tv) - f(p)}{t}.$$

De um modo mais geral, no caso que f é não necessariamente convexa, a derivada direcional de f em p na direção v pode ser definida simplesmente como

$$f'(p,v) = \lim_{t \to 0^+} \frac{f(\exp_p tv) - f(p)}{t},$$

quando o limite existe.

Definição 6.1 *Seja $f : M \to \mathbb{R}$ uma função e $p \in M$. Dizemos que f é regular em p se para todo $v \in T_pM$ a derivada direcional $f'(p,v)$ existe e $f'(p,v) = f^\circ(p,v)$.*

A inclusão (6.1) ocorre com igualdade se, em adição, cada f_i é regular em p. Neste caso teríamos de fato uma generalização da segunda parte do resultado do Lema 3.1.

2. Aplicação do método de ponto proximal no caso que a função objetivo é semicontínua inferior

Observamos que, na linha do que foi apresentado no Capítulo 5, se M é o espaço Euclideano, a hipótese **h3** é trocada por

- **h4** Existe $\tilde{p} \in M$ e $0 < \mu < \bar{\lambda}$ tal que $f + (\mu/2)d^2(.\,,\tilde{p})$ é convexa e

$$\|y(p)\| > \delta > 0,\ p \in L_f(f(q)) \backslash L_f(c),\ y(p) \in \partial\left(f + \frac{\mu}{2}d^2(.\,,\tilde{p})\right)(p) + \mu \exp_p^{-1} \tilde{p},$$

e a seqüência gerada pelo Método (4.2) está bem definida, é possível garantir sua convergência no caso que f é considerada ser apenas semicontínua inferior, ver [25]. A base deste resultado está no seguinte teorema válido no espaço Euclideano:

Teorema 6.4 *Se existem $\tilde{p} \in M$ e $\mu > 0$, tais que $f + (\mu/2)d^2(.\,,\tilde{p})$ é convexa em Ω, então $f + (\lambda/2)d^2(.\,,\bar{p})$ é convexa em Ω, para todo $\bar{p} \in M$ e $\lambda > \mu$.*

Ainda não sabemos se este teorema é válido no contexto Riemanniano. Porém, se isto ocorre o resultado de convergência se estende de forma natural. Propomos então estudar o método de ponto proximal em variedades de Hadamard no caso que a função objetivo é semicontínua inferior e determinar sob quais condições a seqüência gerada está bem definida e investigar se o resultado do Teorema 6.4 se estende ao contexto Riemanniano.

3. Aplicação do método de ponto proximal no caso que a função objetivo é lower-C^2

Nos últimos anos alguns pesquisadores estudaram o método de ponto proximal para encontrar zeros de um operador não necessariamente monótono. O primeiro resultado nessa direção foi dado por Spingarn em [19], onde o autor provou convergência local do método de ponto proximal para operadores ponto-conjunto sob a hipótese de o operador em questão ser estritamente hypomonótono, ver também [20, 21, 22]. Além disso, como uma aplicação, o autor também adaptou o método ao problema de minimização local no caso que a função objetivo é lower-C^2. É válido ressaltar que, no caso Euclideano, uma função (localmente Lipschitz) é lower-C^2 se, e somente se, seu subdiferencial generalizado de Clarke é estritamente hypomonótono, ver [24]. Em [23] os autores apresentaram um algoritmo implementável para calcular a iterada do método de ponto proximal de funções não convexas, o qual foi provado convergir para funções lower-C^2. Nossa proposta é aplicar o método de ponto proximal ao problema de minimização (4.1) quando a função objetivo é do tipo lower-C^2 continuando na mesma linha do que foi apresentado no Capítulo 5.

A seguir apresentamos a definição de função lower-C^2 e um resultado que certamente será útil à realização de nossa proposta. Antes apresentamos um resultado que caracteriza a monotonicidade de um determinado campo de vetores a partir do fato de sua diferencial ser positiva definida.

Lema 6.1 *Seja X um campo de vetores sobre M.*

(i) X é monótono se, e somente se, $\langle A_X(p)v, v\rangle \geq 0$ para todo $p \in M$ e $v \in T_pM$.

(ii) X é fortemente monótono se, e somente se, existe $\lambda > 0$ tal que $\langle A_X(p)v, v\rangle \geq \lambda\|v\|^2$, para todo $p \in M$ e $v \in T_pM$.

Prova. Ver Proposição 3.3 de [41]. **c.q.d**

Definição 6.2 *Seja $V \subset M$ um conjunto aberto. Dizemos que $f : M \to \mathbb{R}$ é lower-C^2 em V se, para cada ponto $\bar{p} \in V$, existem uma vizinhança $W \subset V$ de $\bar{p}$, um conjunto compacto $T \subset \mathbb{R}^m$ e uma função $\varphi : W \times T \to \mathbb{R}$ contínua tal que*

$$f(p) = \max_{\tau \in T} \varphi(p, \tau), \qquad \forall\, p \in W, \tag{6.2}$$

com $\operatorname{grad}_p \varphi$ e $\operatorname{hess}_p \varphi$ contínuas em $W \times T$.

Proposição 6.1 *Sejam $f : M \to \mathbb{R}$ uma função lower-C^2 sobre um conjunto aberto $V \subset M$ e $\bar{p} \in V$. Então existem $\epsilon > 0$, $K > 0$ e $\rho > 0$ tais que*

i) para todo ponto p^0 e parâmetro $R \geq \rho$ a função $f + (R/2)d^2(., p^0)$ é convexa sobre a bola fechada $\bar{B}_\epsilon(\bar{p})$;

ii) a função f é Lipschitz contínua com constante K sobre a bola fechada $\bar{B}_\epsilon(\bar{p})$.

Prova. Dado que f é lower-C^2, então existe uma vizinhaça W de $\bar{p}$ onde f tem uma representação (6.2). Seja $\epsilon > 0$ tal que $\bar{B}_\epsilon(\bar{p}) \subset W$. Como a matriz Hessiana $\operatorname{hess}_p \varphi$ depende continuamente de (p, τ) no compacto $\bar{B}_\epsilon(\bar{p}) \times T \subset W \times T$, então

$$\min\{\langle \operatorname{hess}_p \varphi(p, \tau).v, v\rangle : (p, \tau) \in \bar{B}_\epsilon(\bar{p}) \times T, \|v\| = 1\} > -\infty.$$

Denotemos este mínimo por $\tilde{\rho}$. Sejam $R \geq \rho := |\tilde{\rho}|$ e

$$F(p, \tau) := \varphi(p, \tau) + (R/2)d^2(p, p^0), \qquad p^0 \in M. \tag{6.3}$$

Definição de F e propriedades básicas da Hessiana e do produto interno implicam que

$$\langle \operatorname{hess}_p F(p, \tau).v, v\rangle = \langle \operatorname{hess}_p \varphi(p, \tau).v, v\rangle + R \left\langle \operatorname{hess}_p (1/2)d^2(p, p^0).v, v\right\rangle.$$

Por outro lado, da Proposição 2.2 junto com o Lema 6.1 item ii, temos

$$\left\langle \operatorname{hess}_p (1/2)d^2(p, p^0).v, v\right\rangle \geq \|v\|^2,$$

que, junto com a igualdade acima, nos permite obter a seguinte desigualdade

$$\langle \operatorname{hess}_p F(p, \tau).v, v\rangle \geq \langle \operatorname{hess}_p \varphi(p, \tau).v, v\rangle + R\|v\|^2.$$

Assim, visto que $R > |\tilde{\rho}|$, da última desigualdade adquirimos que

$$\langle \operatorname{hess}_p F(p, \tau).v, v\rangle \geq \langle \operatorname{hess}_p \varphi(p, \tau).v, v\rangle - \tilde{\rho}\|v\|^2. \tag{6.4}$$

Pela definição de $\tilde{\rho}$ temos então que $\langle \text{hess}_p F(p,\tau).v, v\rangle \geq 0$, para todo $(p,\tau) \in \bar{B}_\epsilon(\bar{p})\times T$ quando $\|v\| = 1$ e portanto para todo $v \in T_pM$, visto que ambos os lados de (6.4) são positivamente homogêneos de grau 2 com respeito a v. Com isso mostramos que a matriz Hessiana $\text{hess}_p F(p,\tau)$ é uma matriz semidefinida positiva para cada $(p,\tau) \in \bar{B}_\epsilon(\bar{p}) \times T$. Assim, do Lema 6.1 item i junto com Proposição 2.3 item i, a função $p \mapsto F(p,\tau)$ é convexa sobre $\bar{B}_\epsilon(\bar{p})$ para cada $\tau \in T$. Portanto, tomando máximo em $\tau \in T$ na igualdade (6.3), Proposição 2.5 implica que $f + (\rho/2)d^2(.,p^0)$ é convexa sobre a bola fechada $\bar{B}_\epsilon(\bar{p})$ e o item i segue. O item ii é uma consequência imediata da Proposição 2.6. **c.q.d**

Dado um parâmetro positivo R e $p \in M$, definimos a aplicação ponto proximal de uma função $f : M \to \mathbb{R}$ em p por

$$P_R f(q) := \text{argmin}_{p\in M} \left\{ f(p) + (R/2)d^2(p\,,q) \right\}.$$

Se f é lower-C^2 em um conjunto aberto $V \subset M$, $\bar{p} \in V$ e $\epsilon, \rho > 0$ são tomados como na última proposição, então, para $R \geq \rho$ tal que $P_R f(q) \neq \emptyset$, é fácil ver que para todo $p \in P_R f(q) \cap B_\epsilon(\bar{p})$, $R \exp_p^{-1} q \in \partial^\circ f(p)$. O próximo passo agora é determinar condições sobre as quais a aplicação de ponto proximal está localmente bem definida (ver [23], Teorema 1).

4. Métodos de otimização multicritério no contexto Riemanniano

De forma bem suscinta, seguindo a mesma linha do que foi feito no Capítulo 5, propomos a extenção, ao contexto Riemanniano, do método de ponto proximal (ver BONNEL et al. [51]) e do método de Newton (ver FLIEGE et al. [52]), ambos para otimização multicritério.

Capítulo 7

Apêndice

Nesta seção consideramos as notações do Exemplo 4.1.2.

Lema 7.1 *Se $X \in M$ e $V \in T_X M$, então*

$$\lim_{t\to 0\ Y\to X} \frac{d\left(\exp_Y t(D\exp_X)_{\exp_X^{-1} Y} V,\ Y + tV\right)}{t} = 0.$$

Prova. Preliminarmente, note que

$$d\left(\exp_Y t(D\exp_X)_{\exp_X^{-1} Y} V,\ Y + tV\right) = \\ d\left(\exp_Y t(D\exp_X)_{\exp_X^{-1} Y} V, \exp_Y\left(\exp_Y^{-1}(Y + tV)\right)\right).$$

Visto que a aplicação exponencial $\exp : TM \to M$ é diferenciável, ela é, em particular, localmente Lipschitz. Seja $U_X \subset M$ uma vizinhança de X tal que $TU_X \approx U_X \times T_X M$ e a aplicação exp é Lipschitz em TU_X com constante K. Tome $\delta = \delta(X) > 0$ tal que, para todo $Y \in B_\delta(X)$ e $t \in (0, \delta)$,

$$d\left(\exp_Y t(D\exp_X)_{\exp_X^{-1} Y} V, \exp_Y\left(\exp_Y^{-1}(Y + tV)\right)\right) \leq K\|\mathcal{H}(t, Y)\|, \tag{7.1}$$

onde $\mathcal{H}(t, Y) = tD\exp_X)_{\exp_X^{-1} Y} V - \exp_Y^{-1}(Y + tV)$. Por outro lado, pela definição da métrica (4.14)

$$\|\mathcal{H}(t, Y)\|^2 = \operatorname{tr}\left(\left(t(D\exp_X)_{\exp_X^{-1} Y} V - \exp_Y^{-1}(Y + tV)\right) Y^{-1}\right)^2.$$

A definição de $\exp_Y^{-1}$ em (4.15) e igualdade acima rende

$$\|\mathcal{H}(t, Y)\|^2 \leq \operatorname{tr}\left[\left(t(D\exp_X)_{\exp_X^{-1} Y} V - Y^{1/2} \ln\left(I + tY^{-1/2} V Y^{-1/2}\right) Y^{1/2}\right) Y^{-1}\right]^2,$$

que, depois de simples manipulações algébricas, implica

$$\left\|\frac{\mathcal{H}(t,Y)}{t}\right\|^2 \leq \operatorname{tr}\left[\left((D\exp_X)_{\exp_X^{-1}Y}V - Y^{1/2}\ln\left(I + tY^{-1/2}VY^{-1/2}\right)^{1/t}Y^{1/2}\right)Y^{-1}\right]^2. \tag{7.2}$$

É fácil ver, que

$$\lim_{Y\to X} D\exp_X)_{\exp_X^{-1}Y}V = V, \qquad \lim_{t\to 0\, Y\to X} \ln\left(I + tY^{-1/2}VY^{-1/2}\right)^{1/t} = X^{-1/2}VX^{-1/2}.$$

Portanto, combinando desigualdade (7.2) com as duas últimas igualdades, concluímos

$$\lim_{t\to 0\, Y\to X}\left\|\frac{\mathcal{H}(t,Y)}{t}\right\| = 0.$$

A última igualdade junto com (7.1) implam o resultado desejado. **c.q.d**

Lema 7.2 *Seja $\Omega \subset \mathbb{S}^n_{++}$ um conjunto aberto e convexo. Se F é uma função localmente Lipschitz em Ω, então*

$$F^\circ(X,V) = F^\circ_E(X,V), \qquad X \in \Omega, \qquad V \in T_XM.$$

Prova. Tome $X \in \Omega$ e $V \in T_XM$. Visto que F é localmente Lipschitz em Ω e TM é localmente um produto, existe $\delta = \delta(X) > 0$ tal que $TB_\delta(X) \approx B_\delta(X) \times \mathbb{R}^n$ e

$$\left|\frac{F(\mathcal{G}(t,Y)) - F(Y+tV)}{t}\right| \leq L_X\,\frac{d(\mathcal{G}(t,Y),\, Y+tV)}{t}, \qquad Y \in B_\delta(X),\, t \in (0,\delta),$$

onde $\mathcal{G}(t,Y) = \exp_Y t(D\exp_X)_{\exp_X^{-1}Y}V$. Note que a desigualdade acima é equivalente a

$$\left|\frac{F(\mathcal{G}(t,Y)) - F(Y)}{t} - \frac{F(Y+tV) - F(Y)}{t}\right| \leq L_X\,\frac{d(\mathcal{G}(t,Y),\, Y+tV)}{t}.$$

Por outro lado, Lema 7.1 implica que

$$\lim_{t\to 0\ q\to p}\frac{d(\mathcal{G}(t,Y),\, Y+tV)}{t} = 0,$$

que, combinado com a última desigualdade e definições das derivadas direcionais generalizadas nos permite concluir o lema. **c.q.d**

Lema 7.3 *Seja $\Omega \subset \mathbb{S}^n_{++}$ um conjunto aberto e convexo, $I = \{1,...,m\}$, $F_i : M \to \mathbb{R}$ uma função continuamente diferenciável Ω para cada $i \in I$ e $F : M \to \mathbb{R}$ definida por*

$$F(X) := \max_{i\in I} F_i(X).$$

Então F é localmente Lipschitz em Ω e ocorre

$$\partial^\circ F(X) = \mathrm{conv}\{\mathrm{grad}\, F_i(X) : i \in I(X)\}, \qquad X \in \Omega,$$

onde $I(X) := \{i : F_i(X) = F(X),\ i = 1, ..., m\}$.

Prova. Tome $X \in \Omega$. Da Proposição 3.5 é suficiente provar que

$$\partial^\circ F(X) \subset \mathrm{conv}\{\mathrm{grad}\, F_i(X) : i \in I(X)\}.$$

Seja $W \in \partial^\circ F(X)$. Então, definição do conjunto $\partial^\circ F(X)$ e (4.14) implicam que

$$F^\circ(X, V) \geq \langle W, V \rangle = tr\left(V \Psi''(X) W\right), \qquad V \in \mathbb{S}^n.$$

onde $\Psi(X) = -\ln \det X$. Combinando Lemma 7.2 com a definição do subdiferencial generalizado Euclideano de F em X, $\partial^\circ_E F(X)$, concluímos que

$$\Psi''(X) W \in \partial^\circ_E F(X),$$

Por outro lado,

$$\partial^\circ_E F(X) \subset \mathrm{conv}\{\nabla F_i(X) : i \in I(X)\},$$

onde $I(X) := \{i \in I : F_i(X) = F(X)\}$, ver [43] Proposição 2.3.12. Então, existem constantes $\alpha_i \geq 0$ para $i \in I(X)$ com $\sum_{i \in I(X)} \alpha_i = 1$, tal que

$$\Psi''(X) W = \sum_{i \in I(X)} \alpha_i \nabla F_i(X).$$

Assim, visto que $\Psi''(X)$ é invertível e $\mathrm{grad}\, F_i(X) = \Psi''(X)^{-1} \nabla F_i(X)$, igualdade acima rende

$$W = \sum_{i \in I(X)} \alpha_i \,\mathrm{grad}\, F_i(X).$$

Consequentemente $W \in \mathrm{conv}\{\mathrm{grad}\, F_i(X) : i \in I(X)\}$ e o resultado está provado. **c.q.d**

Referências Bibliográficas

[1] ABSIL, P.-A., BAKER, C. G., GALLIVAN, K. A., "Trust-region methods on Riemannian manifolds", *Found. Comput. Math.*, v. 7, n. 3, pp. 303–330, 2007.

[2] ATTOUCH, H., BOLTE, J., REDONT, P., et al., "Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization", *Optimization*, v. 53, n. 5-6, pp. 435–454, 2004.

[3] AZAGRA, D., FERRERA, J., LÓPEZ-MESAS, F., "Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds", *J. Funct. Anal.*, v. 220, n. 2, pp. 304–361, 2005.

[4] BARANI, A., POURYAYEVALI, M. R., "Invariant monotone vector fields on Riemannian manifolds", *Nonlinear Anal.*, v. 70, n. 5, pp. 1850–1861, 2009.

[5] FERREIRA, O. P., SVAITER, B. F., "Kantorovich's theorem on Newton's method in Riemannian manifolds", *J. Complexity*, v. 18, n. 1, pp. 304–329, 2002.

[6] DA CRUZ NETO, J. X., FERREIRA, O. P., PÉREZ, L. R. L., et al., "Convex- and monotone-transformable mathematical programming problems and a proximal-like point method", *J. Global Optim.*, v. 35, n. 1, pp. 53–69, 2006.

[7] FERREIRA, O. P., OLIVEIRA, P. R., "Subgradient algorithm on Riemannian manifolds", *J. Optim. Theory Appl.*, v. 97, n. 1, pp. 93–104, 1998.

[8] LEDYAEV, Y. S., ZHU, Q. J., "Nonsmooth analysis on smooth manifolds", *Trans. Amer. Math. Soc.*, v. 359, n. 8, pp. 3687–3732 (electronic), 2007.

[9] LI, C., LÓPEZ, G., MARTÍN-MÁRQUEZ, V., "Monotone vector fields and the proximal point algorithm on Hadamard manifolds", *J. Lond. Math. Soc. (2)*, v. 79, n. 3, pp. 663–683, 2009.

[10] LI, S.-L., LI, C., LIOU, Y.-C., et al., "Existence of solutions for variational inequalities on Riemannian manifolds", *Nonlinear Anal.*, v. 71, n. 11, pp. 5695–5706, 2009.

[11] PAPA QUIROZ, E. A., QUISPE, E. M., OLIVEIRA, P. R., "Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds", *J. Math. Anal. Appl.*, v. 341, n. 1, pp. 467–477, 2008.

[12] PAPA QUIROZ, E. A., OLIVEIRA, P. R., "Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds", *J. Convex Anal.*, v. 16, n. 1, pp. 49–69, 2009.

[13] WANG, J.-H., LI, C., "Convergence of the family of Euler-Halley type methods on Riemannian manifolds under the γ-condition", *Taiwanese J. Math.*, v. 13, n. 2A, pp. 585–606, 2009.

[14] WANG, J.-H., HUANG, S., LI, C., "Extended Newton's method for mappings on Riemannian manifolds with values in a cone", *Taiwanese J. Math.*, v. 13, n. 2B, pp. 633–656, 2009.

[15] MARTINET, B., "Régularisation d'inéquations variationnelles par approximations successives", *Rev. Française Informat. Recherche Opérationnelle*, v. 4, n. Ser. R-3, pp. 154–158, 1970.

[16] ROCKAFELLAR, R. T., "Monotone operators and the proximal point algorithm", *SIAM J. Control Optimization*, v. 14, n. 5, pp. 877–898, 1976.

[17] IUSEM, A. N., *Proximal point method in optimization*. v. 20. *Colóquio Brasileiro de Matemática*, IMPA: Rio de Janeiro, RJ, 1995.

[18] FERREIRA, O. P., OLIVEIRA, P. R., "Proximal point algorithm on Riemannian manifolds", *Optimization*, v. 51, n. 2, pp. 257–270, 2002.

[19] SPINGARN, J. E., "Submonotone mappings and the proximal point algorithm", *Numer. Funct. Anal. Optim.*, v. 4, n. 2, pp. 123–150, 1981/82.

[20] PENNANEN, T., "Local convergence of the proximal point algorithm and multiplier methods without monotonicity", *Math. Oper. Res.*, v. 27, n. 1, pp. 170–191, 2002.

[21] IUSEM, A. N., PENNANEN, T., SVAITER, B. F., "Inexact variants of the proximal point algorithm without monotonicity", *SIAM J. Optim.*, v. 13, n. 4, pp. 1080–1097 (electronic), 2003.

[22] GÁRCIGA OTERO, R., IUSEM, A. N., "Proximal methods in reflexive Banach spaces without monotonicity", *J. Math. Anal. Appl.*, v. 330, n. 1, pp. 433–450, 2007.

[23] HARE, W., SAGASTIZÁBAL, C., "Computing proximal points of nonconvex functions", *Math. Program.*, v. 116, n. 1-2, Ser. B, pp. 221–258, 2009.

[24] ROCKAFELLAR, R. T., "Favorable classes of Lipschitz-continuous functions in subgradient optimization", In: *Progress in nondifferentiable optimization*, v. 8, pp. 125–143, *IIASA Collaborative Proc. Ser. CP-82*, Internat. Inst. Appl. Systems Anal.: Laxenburg, 1982.

[25] KAPLAN, A., TICHATSCHKE, R., "Proximal point methods and nonconvex optimization", *J. Global Optim.*, v. 13, n. 4, pp. 389–406, 1998, Workshop on Global Optimization (Trier, 1997).

[26] MOTREANU, D., PAVEL, N. H., "Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds", *J. Math. Anal. Appl.*, v. 88, n. 1, pp. 116–132, 1982.

[27] THÄMELT, W., "Directional derivatives and generalized gradients on manifolds", *Optimization*, v. 25, n. 2-3, pp. 97–115, 1992.

[28] FLIEGE, J., SVAITER, B. F., "Steepest descent methods for multicriteria optimization", *Math. Methods Oper. Res.*, v. 51, n. 3, pp. 479–494, 2000.

[29] BURACHIK, R., DRUMMOND, L. M. G., IUSEM, A. N., et al., "Full convergence of the steepest descent method with inexact line searches", *Optimization*, v. 32, n. 2, pp. 137–146, 1995.

[30] KIWIEL, K. C., MURTY, K., "Convergence of the steepest descent method for minimizing quasiconvex functions", *J. Optim. Theory Appl.*, v. 89, n. 1, pp. 221–226, 1996.

[31] GRAÑA DRUMMOND, L. M., SVAITER, B. F., "A steepest descent method for vector optimization", *J. Comput. Appl. Math.*, v. 175, n. 2, pp. 395–414, 2005.

[32] DRUMMOND, L. M. G., IUSEM, A. N., "A projected gradient method for vector optimization problems", *Comput. Optim. Appl.*, v. 28, n. 1, pp. 5–29, 2004.

[33] UDRIŞTE, C., *Convex functions and optimization methods on Riemannian manifolds*. v. 297. *Mathematics and its Applications*, Kluwer Academic Publishers Group: Dordrecht, 1994.

[34] SMITH, S. T., "Optimization techniques on Riemannian manifolds", In: *Hamiltonian and gradient flows, algorithms and control*, v. 3, pp. 113–136, *Fields Inst. Commun.*, Amer. Math. Soc.: Providence, RI, 1994.

[35] RAPCSÁK, T., *Smooth nonlinear optimization in* $\mathbb{R}^n$. v. 19. *Nonconvex Optimization and its Applications*, Kluwer Academic Publishers: Dordrecht, 1997.

[36] DA CRUZ NETO, J. X., DE LIMA, L. L., OLIVEIRA, P. R., "Geodesic algorithms in Riemannian geometry", *Balkan J. Geom. Appl.*, v. 3, n. 2, pp. 89–100, 1998.

[37] DO CARMO, M. P., *Riemannian geometry*. *Mathematics: Theory & Applications*, Birkhäuser Boston Inc.: Boston, MA, 1992, Translated from the second Portuguese edition by Francis Flaherty.

[38] SAKAI, T., *Riemannian geometry*. v. 149. *Translations of Mathematical Monographs*, American Mathematical Society: Providence, RI, 1996, Translated from the 1992 Japanese original by the author.

[39] HIRIART-URRUTY, J.-B., LEMARÉCHAL, C., *Convex analysis and minimization algorithms. I*. v. 305. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, Springer-Verlag: Berlin, 1993, Fundamentals.

[40] IZMAILOV, A., SOLODOV, M., *Otimizaçã: Condições de Otimalidade, Elementos de Análise Convexa e de Dualidade*. v. 1. IMPA: Rio de Janeiro, RJ, 2005.

[41] DA CRUZ NETO, J. X., FERREIRA, O. P., LUCAMBIO PÉREZ, L. R., "Contributions to the study of monotone vector fields", *Acta Math. Hungar.*, v. 94, n. 4, pp. 307–320, 2002.

[42] DA CRUZ NETO, J. X., FERREIRA, O. P., LUCAMBIO PÉREZ, L. R., "Monotone point-to-set vector fields", *Balkan J. Geom. Appl.*, v. 5, n. 1, pp. 69–79, 2000, Dedicated to Professor Constantin Udrişte.

[43] CLARKE, F. H., *Optimization and Nonsmooth Analysis*. John Wiley Sons: New York, EUA, 1983.

[44] ROTHAUS, O. S., "Domains of positivity", *Abh. Math. Sem. Univ. Hamburg*, v. 24, pp. 189–235, 1960.

[45] NESTEROV, Y. E., TODD, M. J., "On the Riemannian geometry defined by self-concordant barriers and interior-point methods", *Found. Comput. Math.*, v. 2, n. 4, pp. 333–361, 2002.

[46] LANG, S., *Fundamentals of differential geometry*. v. 191. *Graduate Texts in Mathematics*, Springer-Verlag: New York, 1999.

[47] MUNIER, J., "Steepest descent method on a Riemannian manifold: the convex case", *Balkan J. Geom. Appl.*, v. 12, n. 2, pp. 98–106, 2007.

[48] GABAY, D., "Minimizing a differentiable function over a differential manifold", *J. Optim. Theory Appl.*, v. 37, n. 2, pp. 177–219, 1982.

[49] LUC, D. T., *Theory of vector optimization*. v. 319. *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag: Berlin, 1989.

[50] PAPA QUIROZ, E. A., OLIVEIRA, P. R., "New self-concordant barrier for the hypercube", *J. Optim. Theory Appl.*, v. 135, n. 3, pp. 475–490, 2007.

[51] BONNEL, H., IUSEM, A. N., SVAITER, B. F., "Proximal methods in vector optimization", *SIAM J. Optim.*, v. 15, n. 4, pp. 953–970 (electronic), 2005.

[52] FLIEGE, J., GRAÑA DRUMMOND, L. M., SVAITER, B. F., "Newton's method for multiobjective optimization", *SIAM J. Optim.*, v. 20, n. 2, pp. 602–626, 2009.

[53] FERREIRA, O. P., "Proximal subgradient and a characterization of Lipschitz function on Riemannian manifolds", *J. Math. Anal. Appl.*, v. 313, n. 2, pp. 587–597, 2006.