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

MÉTODO DE MINIMIZAÇÃO ALTERNADO NO CONTEXTO DAS VARIEDADES RIEMANNIANAS E MÉTODO DO PONTO PROXIMAL NO CENÁRIO DAS VARIEDADES FINSLERIANAS

Pedro Antônio Soares Júnior

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

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Rio de Janeiro
Abril de 2012

MÉTODO DE MINIMIZAÇÃO ALTERNADO NO CONTEXTO DAS VARIEDADES RIEMANNIANAS E MÉTODO DO PONTO PROXIMAL NO CENÁRIO DAS VARIEDADES FINSLERIANAS

Pedro Antônio Soares Júnior

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. João Xavier da Cruz Neto, D.Sc.

Prof. Orizon Pereira Ferreira, D.Sc.

Prof. Juscelino Pereira Silva, D.Sc.

Prof. Susana Scheimberg de Makler, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
ABRIL DE 2012

Soares Júnior, Pedro Antônio

Método de minimização alternado no contexto das variedades Riemannianas e Método do ponto proximal no cenário das variedades Finslerianas/Pedro Antônio Soares Júnior. – Rio de Janeiro: UFRJ/COPPE, 2012.

XI, 61 p. $29,7$cm.

Orientadores: Paulo Roberto Oliveira

João Xavier da Cruz Neto

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2012.

Referências Bibliográficas: p. 58 – 61.

1. método de ponto proximal. 2. algoritmo alternado. 3. funções não convexas. 4. variedades Riemannianas. 5. variedade Finslerianas. 6. convergência. 7. propriedade de Kurdyka-Lojasiewicz. I. Oliveira, Paulo Roberto *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha esposa Silvânia;
Ao meu filho Pedro Neto;
À minha mãe Teresa Soares;
Ao meu pai Pedro Soares;
Aos irmãos Oscar e Paulo Jânio;
Às irmãs, Osenir e Osineide.

Agradecimentos

Primeiramente a Deus pela oportunidade e força concedida para atravessar as dificuldades que não foram poucas e acreditar que o êxito seria possível embora muitas vezes parecesse o contrário.

Ao Professor Paulo Roberto Oliveira pela viabilazação da oportunidade de realizar o doutorado, pela orientação, incentivo, amizade e confiança.

Ao Professor João Xavier da Cruz Neto pela viabilazação da oportunidade de realizar o doutorado, pela orientação, pelo apoio em todos os momentos, amizade, confiança e incentivo.

Aos Professores Orizon Pereira Ferreira, Juscelino Pereira Silva 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.

Aos Professores e colegas da UESPI/UFPI, pelo incentivo e torcida.

Aos meus pais que me deram o suporte necessário para o início da minha carreira escolar, estimulando e apoiando sempre que necessitei.

À minha esposa pela compreensão de suportar seis meses de ausência imediatamente após o casamento. Pelo apoio moral e cuidado com a relação nos momentos de dificuldades que passamos.

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

A todos os Professores e Funcionários do PESC/COPPE-UFRJ e também aos funcionários da UESPI 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.

Finalmente, a todas as pessoas não citadas anteriormente 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.)

MÉTODO DE MINIMIZAÇÃO ALTERNADO NO CONTEXTO DAS VARIEDADES RIEMANNIANAS E MÉTODO DO PONTO PROXIMAL NO CENÁRIO DAS VARIEDADES FINSLERIANAS

Pedro Antônio Soares Júnior

Abril/2012

Orientadores: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Programa: Engenharia de Sistemas e Computação

Nesta tese consideramos problemas de minimização com restrições. Estendemos o algoritmo proximal alternado ao contexto das variedades Riemannianas. Assumimos que a função objetivo, a ser minimizada, goza da propriedade de Kurdyka-Lojasiewicz e usamos quase-distância como regularização para obter a convergência da sequência gerada pelo algoritmo. Estendemos o método do ponto proximal ao cenário das variedades Finslerianas. Assumimos que a função objetivo é diferenciável e goza da propriedade de Kurdyka-Lojasiewicz para obter a convergência da sequência gerada pelo método.

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

ALTERNATING MINIMIZATION METHOD ON RIEMANNIAN MANIFOLDS AND PROXIMAL POINT METHOD ON FINSLERIAN MANIFOLD

Pedro Antônio Soares Júnior

April/2012

Advisors: Paulo Roberto Oliveira
João Xavier da Cruz Neto

Department: Systems Engineering and Computer Science

In this thesis we consider minimization problems with constraints. We extend the alternating minimization method to context of Riemannian manifolds. We assume that the objective function to be minimized, has the Kurdyka-Lojasiewicz property and used quasi-distance as a regularization function to derive the convergence of the sequence generated to a minimizer point. We extend the proximal point method to the setting of Finslerian manifolds. We assume that the objective function is differentiable and has the Kurdyka-Lojasiewicz property to derive the convergence of the sequence generated by the method to a minimizer point.

Sumário

Notações

$\mathbb{N}$	: conjunto dos números naturais;
$\mathbb{R}$	: conjunto dos números reais;
$\mathbb{R}_+$	: conjunto dos números reais não negativos;
$\mathbb{R} \cup \{+\infty\}$	: conjunto dos números reais estendidos;
$\mathbb{R}^n$	: espaço euclidiano de dimensão n;
M	: variedade diferenciável de dimensão m;
N	: variedade diferenciável de dimensão n;
T_xM	: espaço tangente de M em x;
TM	: fibrado tangente da variedade M:
$(\ \mathbb{R}, +\ , \cdot\ )$	: corpo dos números reais;
$< u, v >$	: produto interno euclidiano dos vetores $u \in T_xM$ e $v \in T_xM$;
$\|v\|$	: norma de um vetor v em T_xM;
$\nabla f(x)$	: gradiente no contexto euclidiano de f em x;
$\nabla_Y X$	: conexão numa variedade M;
$\mathrm{grad}\, f(x)$	: gradiente de f em $x \in M$;
$\hat{\partial} f(x)$	: subdiferencial de Fréchet de f em x;
$\partial f(x)$	: subdiferencial limite de f em x;
$\mathrm{dom}\, f$	: domínio efetivo;
$\mathrm{dist}(x, S)$	: distância de x ao conjunto S;
$B(x; \rho)$	: bola aberta de centro x e raio ρ;
$C^r(M, \mathbb{R})$	: conjunto das funções de classe C^r definidas de M em $\mathbb{R}$;
$\mathcal{O}$	: estrutura o-minimal em $(\ \mathbb{R}, +\ , \cdot\ )$;
$\mathbb{R}_{alg}$	: classe dos conjuntos semialgébricos;
$\mathbb{R}_{an}$	: classe das funções analíticas restritas;
$\mathcal{C}(M)$	: categoria geométrica analítica associada à variedade analítica real M.

Introdução

Neste trabalho propomos e analisamos o algoritmo proximal alternado no contexto das variedades Riemannianas, o qual descrevemos a seguir. Consideramos o problema

$$\begin{aligned} &\min H(x,y) \\ &s.t.\ \ (x,y) \in M \times N, \end{aligned} \tag{1}$$

em que M e N são variedades Riemannianas completas e $H : M \times N \to \mathbb{R} \cup \{+\infty\}$ é uma função semicontínua inferior, própria e limitada inferiormente. O algoritmo proximal alternado para resolver problemas de otimização da forma (1) gera, a partir de um ponto $z_0 = (x_0, y_0) \in M \times N$, uma sequência (z_k), com $z_k = (x_k, y_k) \in M \times N$, como segue:

$$(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$$

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{H(x, y_k) + \frac{1}{2\lambda_k} C_M^2(x_k, x), x \in M\} \\ \\ y_{k+1} \in \operatorname{argmin}\{H(x_{k+1}, y) + \frac{1}{2\mu_k} C_N^2(y_k, y), y \in N\} \end{cases}$$

em que C_M e C_N são quase-distâncias associadas às variedades M e N respectivamente, (λ_k) e (μ_k) são sequências de números positivos e a função H consiste de um termo separável $(x,y) \longmapsto f(x) + g(y)$ e um termo acoplador $(x,y) \longmapsto \Psi(x,y)$.

ATTOUCH et al. [1–3] desenvolveram trabalhos relacionados, onde $M = \mathbb{R}^m$, $N = \mathbb{R}^n$ e o termo acoplador Ψ é quadrático . LEVIS e MALICK [4] trataram do método de projeções alternadas, no contexto em que $M \subset \mathbb{R}^n$ e $N \subset \mathbb{R}^n$ são duas variedades que se interceptam tranversalmente.

ATTOUCH et al. [2] usaram a propriedade de Kurdyka-Lojasiewicz para obter a convergência do algoritmo proximal alternado para problemas não convexos em que a função regularizadora é simétrica. Neste trabalho, focamos no contexto das variedades de Hadamard, com funções de regularização não simétricas, que são mais adequadas para aplicações, por exemplo em teoria da decisão. Então, do ponto de vista matemático, nosso trabalho generaliza [2], pois usamos quase-distâncias em vez de distâncias.

Para gerar a sequência associada ao algoritmo alternado, em cada iteração, consideramos os seguintes subproblemas:

$$\begin{array}{c} \min H(x, y_k) + \frac{1}{2\lambda_k} C_M^2(x_k, x), \\ s.t.\ \ x \in M, \end{array} \tag{2}$$

$$\begin{array}{c} \min H(x_{k+1}, y) + \frac{1}{2\mu_k} C_N^2(y_k, y), \\ s.t.\ \ y \in N. \end{array} \tag{3}$$

Para resolver os subproblemas (2) ou (3), usamos o algoritmo do ponto proximal exato. Por exemplo o subproblema (2), a partir de um ponto $x_0 \in M$, gera a sequência (x_k), $x_k \in M$, como segue:

$$x_{k+1} \in \operatorname{argmin}_{x \in M} \left\{ f(x) + \frac{1}{2\lambda_k} C_M^2(x_k, x) \right\},$$

onde $f(x) = H(x, y_k)$, $\alpha d_M(x_k, x) \leq C_M(x_k, x) \leq \beta d_M(x, x_k)$, em que d_M é a distância Riemanniana (ver Seção 1.1) e α, β são números positivos. FERREIRA e OLIVEIRA [5] estudaram o algoritmo do ponto proximal no caso em que M é uma variedade de Hadamard (see Seção 1.1), $C_M(x_k, \cdot) = d_M(x_k, \cdot)$, $\operatorname{dom} f = M$ e f é convexa. Para obter a boa definição provaram que para cada $k \in \mathbb{N}$ a função $f(\cdot) + \frac{1}{2} d_M^2(x_k, \cdot) : M \to \mathbb{R}$ é 1-coerciva. A análise de convergência é baseada na hipótese de que $\sum_{k=0}^{+\infty} 1/\lambda_k = +\infty$ e f possui um minimizador, obtendo a convergência da sequência (x_k) para um minimizador. Em razão do resultado de convergência da sequência gerada pelo algoritmo proximal alternado (1) (ver Teorema 2.1), do ponto de vista matemático, generalizamos [5], pois usamos quase-distâncias em vez de distâncias e $f(x) = H(x, y) = H(x)$.

Nas últimas três décadas, vários autores propuseram o algoritmo do ponto proximal generalizado para minimização de alguns problemas convexos. Tanto quanto sabemos, FUKUSHIMA e MINE [6] realizaram a primeira generalização, no caso em que M é um espaço de Hilbert. No contexto Riemanniano, PAPA QUIROZ e OLIVEIRA [7] consideraram o algoritmo do ponto proximal para funções quaseconvexas (não necessariamente convexas) e provaram convergência total da seqüência (x_k) para um ponto minimizador, no caso em que M é uma variedade de Hadamard. BENTO et al. [8] consideraram o algoritmo do ponto proximal para funções do tipo C^1-inferior e obtiveram convergência local da seqüência gerada para um ponto minimizador, também considerando M uma variedade de Hadamard. Com o resultado de convergência da sequência gerada pelo nosso algoritmo (1) (ver Teoremas 2.1 e 2.3), nosso trabalho generaliza [7] e [8], usando quase-distâncias em vez de distâncias e $f(x) = H(x, y) = H(x)$.

Em geral, para analisar a convergência da sequência gerada pelo algoritmo do ponto proximal exato, para problemas de minimização convexa ou quaseconvexa, faz-se necessário considerar as variedades de Hadamard, já que a análise é baseada na Fejér convergência ao conjunto de minimizadores da função objetivo, como ocorre em [5]. Essas variedades, possuem topologia e estrutura diferenciável dos espaços Euclidianos. Para f não convexa, ainda no contexto de variedade de Hadamard, para obtermos a convergência da sequência é necessário supormos que f goza da propriedade de Kurdyka-Lojasiewicz:

Seja $U \subset \mathbb{R}^n$ um conjunto aberto, limitado e $g : U \to \mathbb{R}_+$ uma função diferenciável e definível numa estrutura o-minimal. Então existem constantes $c, \eta > 0$ e uma função positiva, estritamente crescente e definível numa estrutura o-minimal, $\varphi : \mathbb{R}_+ \to \mathbb{R}$ de classe C^1, tal que

$$\|\nabla (\varphi \circ g)(x)\| \geq c, \qquad x \in U \cap g^{-1}(0, \eta). \tag{4}$$

Observemos que se tomarmos $\varphi(t) = t^{1-\alpha}$, $\alpha \in [0, 1)$, na desigualdade (4), obtemos:

$$\|\nabla g(x)\| \geq c|g(x)|^{\alpha}, \tag{5}$$

em que $c = 1/(1-\alpha)$. Este conceito será apresentado com mais detalhes na Seção 1.3.

KURDYKA [9] introduziu essa propriedade para funções diferenciáveis definíveis numa estrutura o-minimal (ver seção 1.3). LOJASIEWICZ [10]) estabeleceu a expressão (5) inicialmente para g analítica (conhecida como desigualdade de Lojasiewicz). BOLTE et al. [11], ATTOUCH e BOLTE [12] obtiveram uma extensão da propriedade de Kurdyka-Lojasiewicz para funções subanalíticas (definidas em espaços Euclidianos). BOLTE et al. [13] estabeleceram uma extensão mais geral, no contexto dos espaços Euclidianos, para o subdiferencial de Clarke de funções semicontínuas inferiores definível em uma estrutura o-minimal. No contexto das variedades analíticas Riemannianas, LAGEMAN [14] obteve a propriedade de Kurdyka-Lojasiewicz (4), para $\mathcal{C}$-funções na categoria geométrica analítica (ver Seção 1.3), com o propósito de estabelecer um resultado abstrato de convergência da sequência gerada pelo método tipo gradiente, ver LAGEMAN [14, Teorema 2.1.22]. Ressaltamos que KURDYKA et al. [15] já haviam estabelecido uma extensão da desigualdade (5) para funções analíticas em variedades Riemannianas, com o propósito de resolver a conjectura do gradiente formulada por René Thom [15].

Nos últimos anos, conceitos e técnicas de programação matemática no espaço Euclidiano foram estendidos ao contexto Riemanniano, não somente com objetivos teóricos mas também para obter algoritmos efetivos ([5, 16–18]). Neste trabalho propomos e analisamos também o método do ponto proximal no contexto das variedades de Finsler. Consideramos o seguinte problema

$$\begin{aligned} &\min f(x) \\ &s.t.\ \ x \in M, \end{aligned} \tag{6}$$

onde $f : M \to R$ é uma função C^1 e M é uma variedade de Finsler (ver Capítulo 3).

Para resolver o problema (6), usamos o método do ponto proximal que, dado $x_0 \in M$, gera uma sequência $(x_k) \subset M$ de forma iterativa:

$$x_{k+1} = \operatorname{argmin}_{z\in M} \ \{f(z) + \frac{1}{\lambda_k} C_{x_k}(z)\}, \tag{7}$$

em que $C_{x_k} : M \to \mathbb{R}$ é definida por

$$C_{x_k}(z) = (1/2)d^2(x_k, z),$$

d é uma distância em M e λ_k é uma sequência de números positivos.

Até onde sabemos, KRYSTÁLY et al. [19] apresentaram o primeiro trabalho que modela problemas de otimização em ambiente de Finsler, mas a aplicação efetiva se deu em variedade Riemanniana.

Este trabalho está organizado em seis capítulos. No Capílulo 1, recordamos várias ferramentas e propriedades das variedades Riemannianas, elementos de análise não suave em variedades, propriedade de Kurdyka-Lojasiewicz no contexto Riemanniano e algumas noções de estruturas o-minimal e categorias geométricas analíticas em $(\mathbb{R}, +, \cdot)$, que constituem a parte teórica básica para utilização do método alternado. No Capítulo 2, apresentamos o algoritmo proximal alternado, estabelecemos as condições para assegurar a boa definição e efetuamos análise de convergência da sequência gerada pelo algoritmo alternado, usando quase-distância como função de regularização. No Capítulo 3, relembramos os elementos e propriedades fundamentais das variedades Finslerianas, apresentamos a propriedade de Kurdyka-Lojasiewicz no contexto Finsleriano e algumas noções de estruturas o-minimal e categorias geométricas analíticas em $(\mathbb{R}, +, \cdot)$. No Capítulo 4 , estendemos o método do ponto proximal ao contexto Finsleriano e efetuamos análise de convergência. No Capítulo 5, apresentamos duas interessantes aplicações, uma do método alternado à teoria dos jogos e a outra do método do ponto proximal à teoria da decisão. No Capítulo 6, expomos as considerações finais e apresentamos propostas de pesquisa no contexto das variedades Finslerianas.

Capítulo 1

Elementos de geometria Riemanniana, análise não suave e desigualdade de Kurdyka-Lojasiewicz

Na próxima seção apresentamos algumas propriedades fundamentais e notações da geometria Riemanniana, encontradas por exemplo em [20].

1.1 Geometria Riemanniana

Consideremos uma variedademapsto $\operatorname{grad} f \in \mathcal{X}(M)$ 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 conexão de Levi-Civita associada a (M, g). Um campo de vetores V ao longo de c é dito paralelo se $\nabla_{c'} V = 0$. Caso o campo tangente c' seja paralelo, dizemos que c é uma geodésica. $\nabla_{\gamma'}\gamma' = 0$ é uma equação diferencial não linear de segunda ordem, então $\gamma = \gamma_v(\cdot, x)$ é determinada pela sua posição x e velocidade v em x. Facilmente verificamos que $\|\gamma'\|$ é constante. Dizemos que γ é normalizada se $\|\gamma'\| = 1$. A restrição de uma geodésica a um intervalo fechado e limitado é chamada segmento geodésico. Um segmento geodésico ligando x' a x em M é chamado mínimo se seu comprimento é igual a $d(x, x')$ e a geodésica é chamada geodésica minimizante.

Uma variedade Riemanniana é *completa* se as geodésicas estão definidas para quaisquer valores de t. O teorema de Hopf-Rinow (ver por exemplo [20, Theorem 2.8, p. 146]) estabelece que se M é completa, então dados x_1, $x_2 \in M$ existe um segmento geodésico mínimo ligando x_1 a x_2. Ademais, (M, d) é um espaço métrico completo. Pelo fato de M ser completa a *aplicação exponencial* $\exp_p : T_pM \to M$

está definida para todo $v \in T_pM$ por $\exp_p v = \gamma_v(1,p)$, $p \in M$.

Agora apresentaremos o tensor curvatura em M, representado por R e definido por

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

em que X, Y e Z são campos de vetores de M e $[X,Y] = YX - XY$. Então a curvatura seccional com respeito a X e Y é dada por

$$K(X,Y) = (\langle R(X,Y)Y, X \rangle)/(\|X\|^2 \|Y\|^2 - \langle X, Y \rangle^2),$$

em que $\|X\|^2 = \langle X, X \rangle$. Caso $K(X,Y) \leq 0$ para todo X e Y, então M é denominada de variedade Riemanniana de curvatura não positiva e denotamos $K \leq 0$.

Uma variedade Riemanniana completa, simplesmente conexa, de curvatura seccional não positiva é chamada de variedade de Hadamard . Agora seja M uma variedade de Hadamard. Para cada $x \in M$ podemos definir a inversa da aplicação exponencial

$$exp_x^{-1} : M \to T_xM,$$

a qual é C^∞. Neste caso, $d(x;x') = \|exp_{x'}^{-1}(x)\|$, assim a aplicação $C_{x'} : M \to \mathbb{R}$ definida por $C_{x'}(x) = \frac{1}{2}d^2(x,x')$ também é C^∞. Lembremos que, FERREIRA e OLIVEIRA [5] estabeleceram que a função $C_{x'}$ é estritamente convexa, 1-coerciva e seu gradiente em x é dado por grad $C_{x'}(x) = -exp_x^{-1}(x')$.

Na seção seguinte apresentamos elementos de análise não suave no contexto das variedades, que podem ser encontrados, por exemplo, em [21].

1.2 Análise não suave em variedades

Seja $f : M \to \mathbb{R} \cup \{+\infty\}$ uma função que toma valores no conjunto dos números reais estendidos e denotemos por

$$\text{dom } f := \{x \in M : f(x) < +\infty\}$$

seu domínio. Lembramos que a função f é chamada própria quando dom $f \neq \emptyset$. O gráfico de uma função $f : M \to \mathbb{R} \cup \{+\infty\}$ é definido por

$$\text{Graf } f := \{(x,y) \in M \times \mathbb{R} : y = f(x)\}.$$

Semelhantemente se $F : M \rightrightarrows N$ é uma aplicação ponto-conjunto seu gráfico é definido por

$$\text{Graf } F := \{(x,y) \in M \times N : y \in F(x)\}.$$

Definição 1.1 *Seja f uma função semicontínua inferior. O subdiferencial de Fréchet de f no ponto $x \in M$ é definido por*

$$\hat{\partial} f(x) = \begin{cases} \{d\,h_x : h \in C^1(M) \text{ e } f - h \text{ atinge mínimo local em } x\}, & \text{se } x \in \operatorname{dom} f \\ \emptyset, & \text{se } \ x \notin \operatorname{dom} f, \end{cases}$$

em que $d\,h_x \in (T_xM)^*$ é dada por $d\,h_x(v) = \langle \operatorname{grad}\, h(x), v \rangle$, $v \in T_xM$.

Observemos que se f é diferenciável em x, então $\hat{\partial} f(x) = \{\operatorname{grad} f(x)\}$.

Definição 1.2 *Seja f uma função semicontínua inferior. O subdiferencial (limite) de f no ponto $x \in M$ é definido por*

$$\partial f(x) = \{v^* \in T_x^*M : \exists\ (x_n, v_n^*) \in Graf(\hat{\partial} f),\ (x_n, v_n^*) \to (x, v^*),\ f(x_n) \to f(x)\}.$$

em que $Graf(\hat{\partial} f) := \{(x, u) \in (TM)^* : u \in \hat{\partial} f(x)\}$.

Quando $M = \mathbb{R}^m$, então a função h na definição de subgradiente de Fréchet pode ser escolhida como sendo uma quadrática. Neste caso, a definição se reduz a de subgradiente proximal que tem interpretação geométrica natural em termos dos vetores normais ao epígrafo da função f (em se tratando de uma função suave f, o vetor $(f'(x), -1)$ é um vetor normal ao epígrafo de f).

Retornando ao caso das variedades, lembremos que x é ponto crítico de uma função f se $0 \in \partial f(x)$. Notemos que da definição de subdiferencial $\hat{\partial} f(x) \subset \partial f(x)$ e $\partial f(x)$ podem ser vazios. No entanto, se f atinge um mínimo local em x, então $0 \in \hat{\partial} f(x)$. Essas são propriedades usuais desejadas para um subdiferencial.

Agora, dado um subconjunto S de M, a distância entre um ponto $x \in M$ e o conjunto S é definida por

$$\operatorname{dist}(x, S) := \inf\{d(x, p) : p \in S\}.$$

Caso S seja vazio, definimos $\operatorname{dist}(x, S) := +\infty$ para todo $x \in M$. Quando o conjunto $S = \partial f(x)$ obtemos,

$$\operatorname{dist}(0, \partial f(x)) = \inf\{\|v\| : v \in \partial f(x)\},$$

onde f é uma função de M em $\mathbb{R} \cup \{+\infty\}$.

Proposição 1.1 *Seja $f : M \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferior. Suponhamos que (U, ϕ) é um sistema de coordenadas locais, com $x \in U$. Então,*

$$\partial f(x) = (\phi_x^*) \partial (f \circ \phi^{-1})(\phi(x)),$$

em que ϕ_x^* *denota a derivada adjunta de Fréchet da aplicação* ϕ.

Demonstração: Ver [21, Corollary 4.2]. □

Na próxima seção apresentamos a desigualdade de Kurdyka-Lojasiewicz no contexto Riemanniano e recordamos algumas noções básicas relativas às estruturas o-minimal em $(\mathbb{R}, +, \cdot)$ e categorias geométricas analíticas.

1.3 Desigualdade de Kurdyka-Lojasiewicz no contexto das variedades Riemannianas

Nosso principal interesse é observar em quais circuntâncias as funções semicontínuas inferiores verificam a desigualdade de Kurdyka-Lojasiewicz. No contexto Riemanniano, LAGEMAN [14, Corollary 1.1.25] apresentou o caso diferenciável. Devemos registrar que KURDYKA et al. [15] já haviam estabelecido a referida desigualdade para funções analíticas em variedades analíticas.

Observação 1.1 *Na próxima definição, consideramos* $f : M \to \mathbb{R} \cup \{+\infty\}$ *uma função semicontínua inferior e própria e ainda o conjunto:*

$$[\eta_1 < f < \eta_2] := \{x \in M : \eta_1 < f(x) < \eta_2\}, \quad -\infty < \eta_1 < \eta_2 < +\infty.$$

Definição 1.3 *Uma função* f*, conforme Observação 1.1, goza da propriedade de Kurdyka-Lojasiewicz em* $\bar{x} \in \operatorname{dom} \partial f$ *se existem* $\eta \in (0, +\infty]$*, uma vizinhaça* U *de* $\bar{x}$ *e uma função contínua e côncava* $\varphi : [0, \eta) \to \mathbb{R}_+$ *tal que:*

(i) $\varphi(0) = 0$, $\varphi \in C^1(0, \eta)$ *e, para todo* $s \in (0, \eta)$, $\varphi'(s) > 0$;

(ii) para todo $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$*, a desigualdade de Kurdyka-Lojasiewicz*

$$\varphi'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) \geq 1 \tag{1.1}$$

é verdadeira.

Uma função que verifica a desigualdade de Kurdyka-Lojasiewicz em cada ponto de $\operatorname{dom} \partial f$ *é chamada de função KL.*

No próximo lema, mostramos que se $\bar{x}$ é um ponto não crítico de uma função f semicontínua inferior, então f goza da propriedade de Kurdyka-Lojasiewicz no ponto $\bar{x}$.

Lema 1.1 *Dados uma função* $f : M \to \mathbb{R} \cup \{+\infty\}$ *semicontínua inferior, própria e* $\bar{x} \in \operatorname{dom} \partial f$ *tal que* $0 \notin \partial f(\bar{x})$*. Então* f *goza da propriedade de Kurdyka-Lojasiewicz em* $\bar{x}$.

Demonstração: Como $\bar{x}$ não é ponto crítico de f e $\partial f(\bar{x})$ é um conjunto fechado, temos

$$\delta := \operatorname{dist}(0, \partial f(\bar{x})) > 0.$$

Tomemos $\varphi(t) := t/\delta$, $U := B(\bar{x}, \delta/2)$, $\eta := \delta/2$ e notemos que, para cada $x \in M$,

$$\varphi'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) = \operatorname{dist}(0, \partial f(x))/\delta. \tag{1.2}$$

Agora, para cada $x \in U \cap [f(\bar{x}) - \eta < f < f(\bar{x}) + \eta]$ arbitrário, temos que

$$d(x, \bar{x}) + |f(x) - f(\bar{x})| < \delta.$$

Afirmamos que para cada x que verifica a última desigualdade temos

$$\operatorname{dist}(0, \partial f(x)) \geq \delta. \tag{1.3}$$

Suponhamos por contradição que 1.3 não é verdadeira. Neste caso, existiriam sequências $\{(y_k, v_k)\} \subset \text{Graf } \partial f$ e $\{\delta_k\} \subset \mathbb{R}_{++}$ tais que

$$d(y_k, \bar{x}) + |f(y_k) - f(\bar{x})| < \delta_k, \quad \text{e} \quad \|v_k\| \leq \delta_k$$

com (δ_k) convergindo para zero. Usando que $((y_k, v_k))$ e $(f(y_k))$ convergem para $(\bar{x}, 0)$ e $f(\bar{x})$ respectivamente, e ∂f é uma aplicação fechada, obtemos que $\bar{x}$ é um ponto crítico de f, o que é uma contradição. Logo, nossa afirmação é verdadeira. Portanto, o resultado do lema decorre da combinação de (1.2) com (1.3). $\square$

Uma função $f : M \to \mathbb{R}$, de classe C^2, em que seus pontos críticos $\bar{x}$ são não degenerados, i.e., os autovalores de $Hessf(\bar{x})$ são todos diferentes de zero, é denominada função de Morse. Decorre do teorema da função inversa que seus pontos críticos são isolados. Segundo um resultado de HIRSH [22, Theorem 1.2, page 147], as funções de Morse formam um conjunto aberto e denso no espaço das funções de classe C^2, mais precisamente:

Teorema 1.1 *Seja M uma variedade e denotemos $C^r(M, \mathbb{R})$ o conjunto de todas as funções $g : M \to \mathbb{R}$ de classe C^r. A coleção de todas as funções de Morse $f : M \to \mathbb{R}$ formam um conjunto aberto e denso no conjunto $C^r(M, \mathbb{R})$, $2 \leq r \leq +\infty$.*

Teorema 1.2 *Seja $f : M \to \mathbb{R}$ uma função de Morse. Então f é uma função KL.*

Demonstração: Seja $\bar{x} \in M$ um ponto crítico de f e seja $U = B(\bar{x}, \delta)$ uma vizinhança de $\bar{x}$ que não contenha outros pontos críticos de f. Usando a fórmula de Taylor para f e grad f, temos, para $x \in U$

$$f(x) - f(\bar{x}) = \frac{1}{2}\langle \operatorname{Hess} f(\bar{x}) \exp_{\bar{x}}^{-1} x, \exp_{\bar{x}}^{-1} x \rangle + o(d^2(x, \bar{x})),$$

$$\operatorname{grad} f(x) = \operatorname{Hess} f(\bar{x}) \exp_{\bar{x}}^{-1} x + o(d(x, \bar{x})).$$

Reduzindo, caso necessário, o tamanho do raio δ, podemos assegurar a existência de constantes positivas δ_1, δ_2 tais que

$$|f(x) - f(\bar{x})| \leq \delta_1^2 d^2(x, \bar{x}) \quad \text{e} \quad \delta_2 d(x, \bar{x}) \leq \| \operatorname{grad} f(x) \|.$$

Em decorrência das duas últimas desigualdades, podemos verificar a validade de (1.1) para $\varphi(s) = 2\delta_1 \sqrt{s}/\delta_2$, $U = B(\bar{x}, \delta)$ e $\eta = \delta$. Portanto, decorre do Lemma 1.1 que as funções de Morse são funções KL. $\square$

Observação 1.2 *Vale a pena salientar que, entre outros exemplos, este último foi tratado em [2], porém, no contexto Euclidiano.*

Recordamos algumas definições relativas às estruturas o-minimal no corpo dos números reais ($\mathbb{R}, +, \cdot$), para tanto seguimos as notações constantes de [13]. Uma detalhada discussão sobre estruturas o-minimal e categorias geométricas analíticas pode ser encontrada em [23] e suas referências.

Definição 1.4 *Dada uma sequência $\mathcal{O} = \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ em que cada $\mathcal{O}_n$ é uma coleção de subconjuntos de $\mathbb{R}^n$. $\mathcal{O}$ é chamada de estrutura o-minimal no corpo dos reais ($\mathbb{R}, +, \cdot$) quando, para cada $n \in \mathbb{N}$,:*

(i) $\mathcal{O}_n$ é uma álgebra booleana;

(ii) dado $A \in \mathcal{O}_n$, então $A \times \mathbb{R} \in \mathcal{O}_{n+1}$ e $\mathbb{R} \times A \in \mathcal{O}_{n+1}$;

(iii) dado $A \in \mathcal{O}_{n+1}$, então $\pi_n(A) \in \mathcal{O}_n$, onde $\pi_n : \mathbb{R}^{n+1} \to \mathbb{R}^n$ é a projeção nas n primeiras coordenadas;

(iv) $\mathcal{O}_n$ contém a família de subconjuntos algébricos de $\mathbb{R}^n$;

(v) $\mathcal{O}_1$ consiste de todas as uniões finitas de pontos e intervalos abertos.

Os elementos de $\mathcal{O}$ são chamados *definíveis* em $\mathcal{O}$. Uma função $f : \mathbb{R}^n \to \mathbb{R}$, cujo gráfico pertença a $\mathcal{O}_{n+1}$ é chamada *definível* em $\mathcal{O}$. Ademais, conforme COSTE [24], uma função $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, cuja imagem inversa $f^{-1}(+\infty)$ é um subconjuto definível em $\mathbb{R}^n$, é chamada definível em $\mathcal{O}$ e a restrição de f a $f^{-1}(\mathbb{R})$ é uma função definível com valores em $\mathbb{R}$.

Observação 1.3 *Ressaltamos que uma estrutura o-minimal no corpo dos reais ($\mathbb{R}, +, \cdot$) é uma generalização de um conjunto semialgébrico em $\mathbb{R}^n$, i.e., um conjunto que pode ser escrito como a união finita de conjuntos da forma*

$$\{x \in \mathbb{R}^n : p_i(x) = 0,\ q_i(x) < 0,\ i = 1, \ldots, r\, r \in \mathbb{N}\},$$

em que p_i, q_i, $i = 1, \ldots, r$, são funções polinomiais reais.

BOLTE et al. [13] apresentaram uma extensão da desigualdade de Kurdyka-Lojasiewicz para funções não suaves definíveis em estruturas o-minimal, no caso em que a função φ (da Definition 1.3) não necessariamente é côncava. ATTOUCH et al. [2] reconsideraram essa extensão, pois notaram que a função φ pode ser tomada côncava. Para uma extensa lista de exemplos de conjuntos e funções definíveis em estruturas o-minimal ver ([23] e [2]) e suas referências.

Limitamo-nos a apresentar apenas o material necessário para os nossos propósitos. A primeira classe elementar de exemplos de conjuntos definíveis é a dos conjuntos semialgébricos, a qual denotamos $\mathbb{R}_{alg}$. Uma outra classe de exemplos é das funções analíticas restritas $\mathbb{R}_{an}$, i.e., as menores estruturas que contém os gráficos de todas a funções analíticas $f \mid_{[0,1]^n}$ tais que $f : \mathbb{R}^n \to \mathbb{R}$ é uma função identicamente nula fora de $[0, 1]^n$.

Nas variedades analíticas, os conjuntos semianalíticos e subanalíticos desempenham papel semelhante ao dos conjuntos semi-algébricos em $\mathbb{R}^n$. Para mais detalhes ver [25] e [26].

Definição 1.5 *Um subconjunto de uma variedade analítica que é localmente descrito por um número finito de equações e inequações analíticas é denominado semianalítico.*

Definição 1.6 *Um subconjunto de uma variedade analítica que é localmente descrito por projeções de conjuntos semianalíticos relativamente compactos é denominado subanalítico.*

Uma generalização de conjuntos semianalíticos e subanalíticos em analogia à dada aos conjuntos semialgébricos em termos de estrutura o-minimal, são as categorias geométricas analíticas, cuja definição pode ser encontrada em [23].

Definição 1.7 *Uma categoria geométrica analítica $\mathcal{C}$ associa a cada variedade analítica real M uma coleção de conjuntos $\mathcal{C}(M)$ de tal forma que, para quaisquer variedades analíticas M e N:*

(i) $\mathcal{C}(M)$ *é uma álgebra booleana de subconjuntos de M, com $M \in \mathcal{C}(M)$;*

(ii) *se $A \in \mathcal{C}(M)$, então $A \times \mathbb{R} \in \mathcal{C}(A \times \mathbb{R})$;*

(iii) *se $f : M \to N$ é uma aplicação analítica própria e $A \in \mathcal{C}(M)$, então o conjunto $f(A) \in \mathcal{C}(N)$;*

(iv) *se $A \subset M$ e $\{U_i \mid i \in \Lambda\}$ é uma cobertura aberta de M, então $A \in \mathcal{C}(M)$ se, e somente se, $(A \cap U_i) \in \mathcal{C}(U_i)$, para todo $i \in \Lambda$;*

(v) todo conjunto limitado $A \in \mathcal{C}(\mathbb{R})$ possui fronteira finita, i.e., a fronteira topológica ∂A consiste de um número finito de pontos.

Os elementos de $\mathcal{C}(M)$ são chamados $\mathcal{C}$-conjuntos. Uma aplicação contínua $f : A \to B$, com $A \in \mathcal{C}(M), B \in \mathcal{C}(N)$, cuja gráfico está contido em $\mathcal{C}(M \times N)$, é denominada uma $\mathcal{C}$-função. Todos os subconjuntos subanalíticos e todas as aplicações contínuas subanalíticas de uma variedade M são, respectivamente, $\mathcal{C}$-conjuntos e $\mathcal{C}$-funções em M. Denotamos essa coleção $\mathcal{C}_{an}$ que representa a menor categoria geométrica analítica.

O próximo teorema estabelece uma correspondência biunívoca entre estruturas o-minimal contendo $\mathbb{R}_{an}$ e categorias geométricas analíticas.

Teorema 1.3 *Dada uma categoria geométrica analítica $\mathcal{C}$ existe uma estrutura o-minimal $\mathcal{O}(\mathcal{C})$ e analogamente, dada uma estrutura o-minimal $\mathcal{O}$ contendo $\mathbb{R}_{an}$ existe uma categoria geométrica analítica $\mathcal{C}(\mathcal{O})$, tal que*

(i) $A \in \mathcal{C}(\mathcal{O})$ se para todo $x \in M$ existe uma carta analítica $\phi : U \to \mathbb{R}^n$, $x \in U$, que aplica $A \cap U$ sobre um conjunto definível em $\mathcal{O}$.

(ii) $A \in \mathcal{O}(\mathcal{C})$ se é aplicado sobre um $\mathcal{C}$-conjunto limitado no espaço Euclidiano através de uma bijeção semialgébrica.

Ademais, a partir $\mathcal{C} = \mathcal{C}(\mathcal{O})$ obtemos a estrutura o-minimal $\mathcal{O}$ através da correspondência descrita, bem como a partir de $\mathcal{O} = \mathcal{O}(\mathcal{C})$ obtemos $\mathcal{C}$.

Demonstração: Ver [23] e [14, Teorema 1.1.3]. □

Como consequência da correspondência entre estruturas o-minimal, contendo $\mathbb{R}_{an}$, e categorias geométricas analíticas, os conjuntos definíveis nos possibilitam exibir exemplos de $\mathcal{C}$-conjuntos em $\mathcal{C}(\mathcal{O})$. Ademais, $\mathcal{C}$-conjuntos são localmente aplicados, através de cartas analíticas, em conjuntos definíveis de $\mathcal{O}(\mathcal{C})$.

Proposição 1.2 *Sejam $f : M \to \mathbb{R}$ uma $\mathcal{C}$-função e $\phi : U \to \mathbb{R}^n$, $U \subset M$ uma carta analítica local. Suponhamos que $V \subset M$ é um conjunto limitado, $\overline{V} \subset U$ e $U \subset \mathrm{dom}\, f$. Se f restrita a U é limitada, então*

$$f \circ \phi^{-1} : \phi(V) \to \mathbb{R} \tag{1.4}$$

é definível em $\mathcal{O}(\mathcal{C})$.

Demonstração: Ver LAGEMAN [14, Proposition 1.1.5.]. □

Teorema 1.4 *Sejam U uma subvariedade definível de $\mathbb{R}^n$, não necessariamente limitada, e $f : U \to R_+$ uma função diferenciável e definível. Então, existem funções $\psi : [0, \epsilon_0) \to \mathbb{R}_+$, contínua e definível, com $\psi(0) = 0$ e $\psi \in C^1$ em $(0, \epsilon_0)$, e $\chi : \mathbb{R}_+ \to (0, \epsilon_0)$, contínua e definível, tais que*

$$\|\nabla(\psi \circ f)(x)\| \geq 1, \quad \textit{para todo } x \textit{ tal que } \; 0 < f(x) \leq \chi(\|x\|).$$

Demonstração: Ver [13, Teorema 11]. □

Teorema 1.5 *Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferior, própria e definível. Então, f goza da propriedade de Kurdyka-Lojasiewicz em cada ponto de* dom f*. Ademais, a função φ de* (1.1) *é definível em $\mathcal{O}$.*

Demonstração: Ver [2, Teorema 4.1]. □

A seguir apresentamos uma extensão da propriedade de Kurdyka-Lojasiewicz para $\mathcal{C}$-funções não suaves definidas em variedades analíticas.

Teorema 1.6 *Sejam M uma variedade Riemanniana analítica e $f : M \to \mathbb{R}$ uma $\mathcal{C}$-função contínua. Então f é uma função KL e a função φ de* (1.1) *é definível em $\mathcal{O}$.*

Demonstração: Consideremos um ponto crítico $\bar{x} \in M$ de f e uma carta analítica local $\phi : V \to \mathbb{R}^n$, com a vizinhaça $V \subset M$ de $\bar{x}$ escolhida de forma que V e $f(V)$ são limitados. Assim, de acordo com a Proposição 1.2 a função $f \circ \phi^{-1} : \phi(V) \to \mathbb{R}$ é definível em $\mathcal{O}(\mathcal{C})$. Logo, como $\phi(V)$ é um conjunto aberto e limitado definível contendo $\bar{y} = \phi(\bar{x})$ e ϕ é definível, podemos aplicar o Teorema 1.4, com $U = \phi(V)$, e seguindo a demonstração do Teorema 1.5, obtemos que a função $h = f \circ \phi^{-1}$ goza da propriedade de Kurdyka-Lojasiewicz em $\bar{y} = \phi(\bar{x})$, i.e., existem $\eta \in (0, +\infty]$ e uma função côncava contínua $\Phi : [0, \eta) \to \mathbb{R}_+$ tais que:

(i) $\Phi(0) = 0$, $\Phi \in C^1(0, \eta)$ e para todo $s \in (0, \eta)$, $\Phi'(s) > 0$;

(ii) para todo $y \in U \cap [h(\bar{y}) < h < h(\bar{y}) + \eta]$,

$$\Phi'(h(y) - h(\bar{y}))\,\mathrm{dist}(0, \partial h(y)) \geq 1.$$

Como ϕ é um difeomorfismo e usando $y = \phi(x)$, $\bar{y} = \phi(\bar{x})$ e $h = f \circ \phi^{-1}$, a última desigualdade da Proposição 1.1 implica

$$\Phi'(f(x) - f(\bar{x}))\,\mathrm{dist}(0, (\phi_x^*)^{-1}\partial f(x)) \geq 1, \quad x \in V \cap [0 < f < f(\bar{x}) + \eta],$$

em que ϕ_x^* denota a derivada de Fréchet adjunta da função ϕ.

Tomemos um conjunto aberto $V' \subset V$ de forma que $K = \overline{V'}$ esteja contido no interior do conjunto V e $\bar{x} \in V'$. Logo, K é um conjunto compacto e para cada $x \in K$ existe $C_x > 0$ com

$$\|(\phi_x^*)^{-1} w\| \leq C_x \|w\|, \qquad w \in T_x M.$$

Novamente, usando que K é um conjunto compacto e $(\phi_x^*)^{-1}$ é um difeomorfismo, existe uma constante positiva $C := \sup\{C_x : x \in K\}$ tal que

$$\|(\phi_x^*)^{-1} w\| \leq C \|w\|, \qquad w \in T_x M, \ x \in K.$$

Portanto, para $x \in V' \cap [0 < f < f(\bar{x}) + \eta]$, segue-se que

$$1 \leq \Phi'(f(x) - f(\bar{x})) \operatorname{dist}(0, (\phi_x^*)^{-1} \partial f(x)) \leq C\, \Phi'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)).$$

Logo, f goza da propiedade de Kurdyka-Lojasiewicz em $\bar{x}$, com $\varphi = C\,\Phi$. Assim, como $\bar{x}$ é um ponto crítico arbitrário, concluimos do Lema 1.1 que f é uma função KL. A segunda parte decorre do Teorema 1.4 e assim encerramos a prova. $\square$

O resultado seguinte nos dá uma extensão da propriedade de Kurdyka-Lojasiewicz para funções definíveis não suaves e definidas em subvariedades do espaço Euclidiano, abordado em [24, Capítulo 6].

Teorema 1.7 *Seja $f : M \subset \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferior, própria e definível numa estrutura o-minimal $\mathcal{O}$. Se M é munida com a métrica induzida do espaço Euclidiano, então f é KL e a função φ de (1.1) é definível em $\mathcal{O}$.*

Demonstração: Tomemos um ponto crítico $\bar{x} \in M$ de f e um subconjunto limitado e definível W de $\mathbb{R}^n$ de forma que $\bar{x} \in W$. Como dom f e W são conjuntos definíveis em $\mathbb{R}^n$ e W é limitado, temos que dom $f \cap W$ é um conjunto limitado e definível em $\mathbb{R}^n$. Logo, conforme Teorema 1.4, com $U =$ dom $f \cap W$ e, usando o roteiro da prova do Teorema 1.5, concluimos que f goza da propriedade de Kurdyka-Lojasiewicz no ponto $\bar{x}$. Portanto, como $\bar{x}$ é um ponto crítico arbitrário, o Lema 1.1 nos permite concluir a primeira parte do teorema. A segunda parte decorre do Teorema 1.4. $\square$

Observação 1.4 *As subvariedades do espaço Euclidiano que são obtidas como imagem inversa de valor regular de uma função definível, i.e., se $F : \mathbb{R}^{n+k} \to \mathbb{R}^k$ é uma função de classe C^r e "0" é um valor regular de F, então $M = F^{-1}(0)$, são subvariedades definíveis de $\mathbb{R}^n$. Ademais, pelo Teorema de Nash ([27]), podemos mergulhar isometricamente em $\mathbb{R}^n$ uma pequena parte $\mathcal{Y}$ de M, a qual é uma* *subvariedade regular de $\mathbb{R}^n$. De fato, se $\epsilon > 0$ é pequeno o bastante, então o conjunto dos segmentos de raios normais ϵ, centrados nos pontos de $\mathcal{Y}$, determina uma vizinhança tubular $\mathcal{V}$ de $\mathcal{Y}$. Por outro lado, $\mathcal{V}$ possui um sistema de coordenadas naturais $y = (x,t) \in \mathcal{Y} \times B_\epsilon(0)$, em que $B_\epsilon(0) \subset \mathbb{R}^m$ é uma bola de raio ϵ (onde $n-m$, $m<n$, é a dimensão M). Identificamos $(x,0)$ com x e definimos $h(x,t)=t$. Concluímos que h é uma função definível e $\mathcal{Y} = \{y \in V; h(y)=0\}$ é uma subvariedade definível de $\mathbb{R}^n$.*

Observação 1.5 *Os resultados dos Teoremas 1.6 e 1.7 mostram a existência de funções KL no cenário das variedades Riemannianas, os quais serão importantes para se obter convergência da sequência gerada pelo método, pois supomos que a função a ser minimizada é KL, ver detalhes Seção 2.2.*

Capítulo 2

Algoritmo proximal alternado

Neste capítulo, fixamos nossos espaços de ações, mostramos a boa definição e a convergência da sequência gerada pelo algoritmo alternado. Consideramos duas variedades Riemannianas completas M e N, com $m = \dim(M)$ e $n = \dim(N)$, respectivamente, e uma função $H : M \times N \to \mathbb{R} \cup \{+\infty\}$.

Definição 2.1 *A função H possui a seguinte estrutura:*

1. $H(x, y) = f(x) + g(y) + \Psi(x, y)$, $\forall (x, y) \in M \times N$;
2. $f : M \to \mathbb{R} \cup \{+\infty\}$, $g : N \to \mathbb{R} \cup \{+\infty\}$ *são funções semicontinuas inferiores e próprias;*
3. $\Psi \in C^1$ e $\operatorname{grad} \Psi$ *é Lipschitz contínuo em subconjuntos limitados de $M \times N$.*

Observação 2.1 *A função objetivo H, com $M = \mathbb{R}^m$ e $N = \mathbb{R}^n$ foi tratada em [2].*

Observação 2.2 *Quando fixamos y em N, o subdiferencial da função $H(\cdot, y)$ em p é denotado por $\partial_x H(p, y)$. Similarmente, quando fixamos x em M, o subdiferencial da função $H(x, \cdot)$ em q é denotado por $\partial_y H(x, q)$.*

O próximo resultado é usado para provar a boa definição do método alternado. A demonstração é semelhante a que consta em .

Proposição 2.1 *Seja H como na Definição 2.1. Então,*

$$\partial H(x, y) = \partial(f(x) + g(x) + \Psi(x, y)) = \partial_x H(x, y) \times \partial_y H(x, y),$$

para todo $(x, y) \in \operatorname{dom} H = \operatorname{dom} f \times \operatorname{dom} g$.

Demonstração: Ver [2] e consultar também [21]. □

Definição 2.2 *Uma função $C_M : M \times M \to \mathbb{R}_+$ é denominada quase-distância quando goza das seguintes propriedades:*

(i) para todo $x_1, x_2 \in M,\ C_M(x_1, x_2) = C_M(x_2, x_1) = 0 \Leftrightarrow x_1 = x_2$*;*

(ii) para todo $x_1, x_2, x_3 \in M,\ C_M(x_1, x_3) \leq C_M(x_1, x_2) + C_M(x_2, x_3)$*.*

Observação 2.3 *Quando* C_M *é simétrica, ou seja, para todo* $x_1, x_2 \in M$*,* $C_M(x_1, x_2) = C_M(x_2, x_1)$*, então* C_M *é uma distância.*

Agora, definimos a distância que utilizamos na variedade produto $M \times N$.

Definição 2.3 *Uma função*

$$d : (M \times N) \times (M \times N) \to \mathbb{R}_+$$

dada por

$$d(z_1, z_2) = [d_M^2(x_1, x_2) + d_N^2(y_1, y_2)]^{1/2},$$

para todo $z_1 = (x_1, y_1),\ z_2 = (x_2, y_2)$ *em* $M \times N$*, em que* d_M *e* d_N *são distâncias em* M *e* N*, respectivamente, define uma distância* d *em* $M \times N$*.*

Dado $(x_0, y_0) \in M \times N$, o algoritmo alternado que estudamos é da forma:

$$(x_k, y_k) \to (x_{k+1}, y_k) \to (x_{k+1}, y_{k+1})$$

$$\begin{cases} x_{k+1} \in \operatorname{argmin}\{H(p, y_k) + \frac{1}{2\lambda_k} C_M^2(x_k, p), p \in M\} \\ \\ y_{k+1} \in \operatorname{argmin}\{H(x_{k+1}, q) + \frac{1}{2\mu_k} C_N^2(y_k, q), q \in N\} \end{cases} \tag{2.1}$$

em que C_M e C_N são quase-distâncias associadas às variedades M e N, respectivamente, e (λ_k), (μ_k) são sequências de números positivos.

2.1 Convergência

Agora estabelecemos as condições para assegurar a boa definição e convergência da sequência gerada pelo o algoritmo (2.1), quais sejam:

$$(\mathcal{A}) : \begin{cases} M \text{ e } N \text{ são variedades de Hadamard;} \\ C_M,\ C_N \text{ são } C^1 \text{ e existem } \alpha, \beta > 0 \text{ tais que} \\ \alpha d_M(x, p) \leq C_M(x, p) \leq \beta d_M(x, p),\ \alpha d_N(y, q) \leq C_N(y, q) \leq \beta d_N(y, q); \\ \inf_{M \times N} H > -\infty; \\ \text{a função } H(\cdot, y_0) \text{ é própria;} \\ \text{para números positivos } t_1 < t_2, \text{ as sequências } (\lambda_k),\ (\mu_k) \subset (t_1, t_2)\ . \end{cases}$$

Definição 2.4 *Uma função $h : M \to \mathbb{R}$ é denominada coerciva em $\bar{x} \in M$ quando*

$$\lim_{d(\bar{x},x)\to+\infty} h(x) = +\infty.$$

A próxima proposição é fundamental para este trabalho e será bastante utilizada. Assumimos que a função H é como na Definição 2.1 e consideramos os vetores

$$u_{k+1} := \operatorname{grad}_x \Psi(x_{k+1}, y_{k+1}) - \operatorname{grad}_x \Psi(x_{k+1}, y_k) - \lambda_k^{-1} C_M(x_k, x_{k+1}) \operatorname{grad} C_M(x_k, x_{k+1})$$

e

$$v_{k+1} := -\mu_k^{-1} C_N(y_k, y_{k+1}) \operatorname{grad} C_N(y_k, y_{k+1}).$$

Observação 2.4 *Notemos que:*
$\operatorname{grad} \Psi(x, y) = (\operatorname{grad}_x \Psi(x, y), \operatorname{grad}_y \Psi(x, y)) \in T_xM \times T_yN$ *e consequentemente* $(u_{k+1}, v_{k+1}) \in T_{x_{k+1}}M \times T_{y_{k+1}}N$.

Proposição 2.2 *Seja H conforme hipótese $(\mathcal{A})$. Então as sequências (x_k) e (y_k) estão bem definidas. Ademais,*

(i)

$$H(x_{k+1}, y_{k+1}) + \frac{1}{2\lambda_k} C_M^2(x_k, x_{k+1}) + \frac{1}{2\mu_k} C_N^2(y_k, y_{k+1}) \leq H(x_k, y_k); \qquad (2.2)$$

(ii)

$$\sum_{k=0}^{+\infty} [C_M^2(x_k, x_{k+1}) + C_N^2(y_k, y_{k+1})] < +\infty,$$

consequentemente, $lim_{k\to+\infty}[d_M(x_k, x_{k+1}) + d_N(y_k, y_{k+1})] = 0$;

(iii) para todo $k \geq 0$

$$(u_{k+1}, v_{k+1}) \in \partial H(x_{k+1}, y_{k+1}). \qquad (2.3)$$

Para toda subsequência limitada $(x_{k'}, y_{k'})$ de (x_k, y_k), obtemos que

$$(u_{k'}, v_{k'}) \to 0 \text{ e } \operatorname{dist}(0, \partial H(x_{k'}, y_{k'})) \to 0.$$

Demonstração: Decorre da hipótese $(\mathcal{A})$ que para todo número $t > 0$ e $(\bar{x}, \bar{y})$ em $M \times N$, as funções

$$x \to H(x, \bar{y}) + \frac{1}{2t} C_M^2(\bar{x}, x)$$

e

$$y \to H(\bar{x}, y) + \frac{1}{2t} C_N^2(\bar{y}, y)$$

são coercivas. Uma indução sobre k nos assegura que as sequências (x_k) e (y_k) estão bem definidas, bem como (i) e (ii) são verdadeiros para $k \geq 1$.

Provamos agora o item (iii). Como 0 pertence ao subdiferencial da função

$$p \to H(p, y_k) + \frac{1}{2\lambda_k} C_M^2(x_k, p)$$

no ponto x_{k+1}, então

$$0 \in (\frac{1}{\lambda_k} C_M(x_k, x_{k+1}) \operatorname{grad} C_M(x_k, x_{k+1}) + \partial_x H(x_{k+1}, y_k)), \forall k \geq 0.$$

Analogamente, 0 pertence ao subdiferential da função

$$q \to H(x_{k+1}, q) + \frac{1}{2\lambda_k} C_N^2(y_k, q),$$

no ponto y_{k+1}. O que implica

$$0 \in (\frac{1}{\mu_k} C_N(y_k, y_{k+1}) \operatorname{grad} C_N(y_k, y_{k+1}) + \partial_y H(x_{k+1}, y_k)), \forall k \geq 0.$$

Devido à estrutura de H, segue-se que

$$\begin{aligned} \partial_x H(x_{k+1}, y_k) &= \partial f(x_{k+1}) + \operatorname{grad}_x \Psi(x_{k+1}, y_k) \\ &+ \operatorname{grad}_x \Psi(x_{k+1}, y_{k+1}) - \operatorname{grad}_x \Psi(x_{k+1}, y_{k+1}) \\ &= \partial_x H(x_{k+1}, y_{k+1}) + \operatorname{grad}_x \Psi(x_{k+1}, y_k) - \operatorname{grad}_x \Psi(x_{k+1}, y_{k+1}). \end{aligned}$$

Assim,

$$\operatorname{grad}_x \Psi(x_{k+1}, y_{k+1}) - \operatorname{grad}_x \Psi(x_{k+1}, y_k) - \frac{1}{\lambda_k} C_M(x_k, x_{k+1}) \operatorname{grad} C_M(x_k, x_{k+1})$$

pertence a

$$\partial_x H(x_{k+1}, y_{k+1}),$$

e

$$-\frac{1}{\mu_k} C_N(y_k, y_{k+1}) \operatorname{grad} C_N(y_k, y_{k+1}) \in \partial_y H(x_{k+1}, y_{k+1}).$$

Pela Proposição 2.1 temos (2.3).

Seja (x', y') um ponto de acumulação da sequência $((x_k, y_k))$. Sem perda de generalidade podemos assumir que $(x_{k'}, y_{k'}) \to (x', y')$. Usando o item (ii), a desigualdade triangular , a continuidade uniforme do $\operatorname{grad}_x \Psi$ e o fato de que C_M, C_N

são C^1, temos das expressões de u_k e v_k que $(u'_k, v'_k) \to (0,0)$, assim

$$\text{dist}(0, \partial H(x_{k'+1}, y_{k'+1})) \to 0.$$

□

O Lema seguinte, especiamente os itens (ii) e (iii), mostra o primeiro resultado de convergência da sequência gerada por (2.1). Os Teoremas 2.1 e 2.3 apresentam propriedades de convergência mais precisas.

Lema 2.1 *Sejam H conforme hipótese $(\mathcal{A})$, $((x_k, y_k))$ a sequência dada por* (2.1) *e $\Gamma(x_0, y_0)$ o conjunto dos pontos de acumulação da sequência. Temos:*

(i) se $((x_k, y_k))$ é limitada, então $\Gamma(x_0, y_0)$ é não vazio, compacto, conexo e

$$\text{dist}((x_k, y_k), \Gamma(x_0, y_0)) \to 0, \text{ com } k \to +\infty;$$

(ii) $\Gamma(x_0, y_0) \subset critH$;

(iii) H é finita e constante em $\Gamma(x_0, y_0)$, com

$$inf_{k \in N} H(x_k, y_k) = \lim_{k \to +\infty} H(x_k, y_k).$$

Demonstração:
O item (i) segue do fato de que a sequência $(d_M(x_k, x_{k+1}) + d_N(y_k, y_{k+1}))$ converge para 0 e do Teorema de Hopf-Rinow;
(ii) pela boa definição de $((x_{k+1}, y_{k+1}))$, temos para $k \geq 0$

$$H(x_{k+1}, y_{k+1}) + \frac{1}{2\lambda_k} C_M^2(x_k, x_{k+1}) \leq H(p, y_{k+1}) + \frac{1}{2\lambda_k} C_M^2(x_k, p), \forall p \in M$$

e

$$H(x_{k+1}, y_{k+1}) + \frac{1}{2\mu_k} C_N^2(y_k, y_{k+1}) \leq H(x_{k+1}, q) + \frac{1}{2\mu_k} C_N^2(y_k, q), \forall q \in N.$$

Devido à específica forma de H, $0 < t_1 \leq \lambda_k \leq t_2$ e $0 < t_1 \leq \mu_k \leq t_2$, segue

$$f(x_{k+1}) + \Psi(x_{k+1}, y_k) + \frac{1}{2t_2} C_M^2(x_k, x_{k+1}) \leq f(p) + \Psi(p, y_k) + \frac{1}{2t_1} C_M^2(x_k, p), \forall p \in M \tag{2.4}$$

e

$$g(y_{k+1}) + \Psi(x_{k+1}, y_{k+1}) + \frac{1}{2t_2} C_N^2(y_k, y_{k+1})$$

$$\leq g(q) + \Psi(x_{k+1}, q) + \frac{1}{2t_1} C_N^2(y_k, q), \forall q \in N. \tag{2.5}$$

Seja $(\bar{x}, \bar{y})$ um ponto de $\Gamma(x_0, y_0)$, existe uma subsequência $((x_{k'}, y_{k'}))$ de $((x_k, y_k))$ que converge para $(\bar{x}, \bar{y})$. Como $C_M^2(x_k, x_{k+1}) \to 0$, deduzimos de (2.4) que

$$\liminf f(x_{k'}) + \Psi(\bar{x}, \bar{y}) \leq f(p) + \Psi(p, \bar{y}) + \frac{1}{2t_1} C_M^2(\bar{x}, p), \forall p \in M.$$

Para $p = \bar{x}$ temos

$$\liminf f(x_{k'}) \leq f(\bar{x}).$$

Como f é semicontínua inferior

$$\liminf f(x_k) = f(\bar{x}).$$

Sem perda de generalidade podemos assumir que toda a sequência $f(x_{k'})$ converge para $f(\bar{x})$, i.e.,

$$\lim f(x_k') = f(\bar{x}).$$

Analogamente, devido a (2.5), podemos assumir que $\lim g(y_{k'}) = g(\bar{y})$. Já que Ψ é contínua, temos

$$\lim \Psi(x_{k'}, y_{k'}) = \Psi(\bar{x}, \bar{y})$$

e portanto

$$\lim H(x_{k'}, y_{k'}) = H(\bar{x}, \bar{y}).$$

Pela Proposição 2.2, item (iii), com as mesmas notações,

$$(u_{k'}, v_{k'}) \in \partial H(x_{k'}, y_{k'}) \text{ e } (u_{k'}, v_{k'}) \to 0.$$

Sendo ∂H um conjunto fechado, finalmente obtemos

$$0 \in \partial H(\bar{x}, \bar{y}),$$

assim concluímos a prova de (ii);
(iii) seja $(\bar{x}, \bar{y})$ um ponto de $\Gamma(x_0, y_0)$, existe uma subsequência $((x_{k'}, y_{k'}))$ de $((x_k, y_k))$ que converge para $(\bar{x}, \bar{y})$, com $(H(x_k, y_k))$ convergindo para $H(\bar{x}, \bar{y})$. Sendo $(H(x_k, y_k))$ não crescente e $\inf H > -\infty$, segue-se que $H(\bar{x}, \bar{y}) = \inf H(x_k, y_k)$, para qualquer $(\bar{x}, \bar{y})$ em $\Gamma(x_0, y_0)$, logo H é finita e constante em $\Gamma(x_0, y_0)$. $\square$

A próxima seção é destinada à análise de convergência da sequência gerada pelo algoritmo (2.1). Apresentamos os principais resultados matemáticos deste trabalho, relativamente ao método alternado. Ressaltamos que ATTOUCH et al. [1, 3] trataram deste tema num cenário convexo, em que a função de acoplamento Ψ é quadrática.

2.2 Convergência para um ponto crítico

Inicialmente introduzimos algumas notações, quais sejam: $z_k = (x_k, y_k)$, $h_k = H(z_k)$, $\bar{z} = (\bar{x}, \bar{y})$ e $\bar{h} = H(\bar{z})$. Denotamos U, η e φ como na Definição (1.3), relativamente a H no ponto $\bar{z}$, e seja $\rho > 0$ tal que $B(\bar{z}, \rho) = \{z \in M \times N,\ d(\bar{z}, z) < \rho\} \subset U$. Assumimos as seguintes desigualdades,

$$\bar{h} < h_k < \bar{h} + \eta,\ k \geq 0 \tag{2.6}$$

e

$$\varphi(h_0 - \bar{h})E + \alpha^{-1}\sqrt{2t_2(h_0 - \bar{h})} + d(z_0, \bar{z}) < \rho, \tag{2.7}$$

em que $E = 2t_2(\delta + 2\tau\beta {t_1}^{-1})\alpha^{-2}$, δ é uma constante de Lipschitz para $\operatorname{grad}\Psi$, τ é uma cota superior para as normas $\operatorname{grad} C_M,\ \operatorname{grad} C_N$ na bola $B(\bar{z}, \sqrt{2}\rho)$ e α e β são constantes da hipótese $(\mathcal{A})$. Denotamos ainda, $D(z_k) := \sum_{i=k+1}^{+\infty} d(z_i, z_{i+1})$.

Observação 2.5 *No caso em que (z_k) é uma sequência limitada, podemos considerar um ponto de acumulação $\bar{z}$ de (z_k). Assim, $\varphi(h_k - \bar{h}) \to 0$. Logo a sequência (b_k),*

$$b_k := \varphi(h_k - \bar{h})E + \alpha^{-1}\sqrt{2t_2(h_k - \bar{h})} + d(\bar{z}, z_k)$$

admite 0 como ponto de acumulação. Assim dado $\rho > 0$, existe um número inteiro k_0 tal que $b_{k_0} < \rho$. Assim, tomando $z_0 = z_{k_0}$ obtemos (2.7).

Teorema 2.1 *Suponhamos que H é uma função KL em $\bar{z}$ e goza da hipótese $(\mathcal{A})$. Seja z_0 um ponto inicial da sequência (z_k) gerada por (2.1). Suponhamos também que (2.6) e (2.7) são verdadeiras. Então, a sequência (z_k) converge para o ponto crítico $\bar{z}$ de H, ademais, $\forall k \geq 0$:*

(i) $z_k \in B(\bar{z}, \rho)$;

(ii) $D(z_k) \leq \varphi(h_k - \bar{h})E + \alpha^{-1}\sqrt{2t_2(h_{k-1} - \bar{h})}$.

Demonstração: Suponhamos que $\bar{h} = 0$ (caso necessário podemos trocar H por $(H - \bar{h})$). Da Observação 2.5 assumimos que z_0 é tal que (2.6) é verdadeira. A partir de (2.2), temos, para todo inteiro $i \geq 0$:

$$\frac{1}{2\lambda_i} C_M^2(x_i, x_{i+1}) + \frac{1}{2\mu_i} C_N^2(y_i, y_{i+1}) \leq h_i - h_{i+1}.$$

Pela hipótese $\mathcal{A}$, temos

$$\frac{1}{2t_2}\alpha^2 d_M^2(x_i, x_{i+1}) + \frac{1}{2t_2}\alpha^2 d_N^2(y_i, y_{i+1}) \leq \frac{1}{2\lambda_i} C_M^2(x_i, x_{i+1}) + \frac{1}{2\mu_i} C_N^2(y_i, y_{i+1}).$$

Segue das duas últimas desigualdades que

$$\frac{1}{2t_2}\alpha^2 d^2(z_i, z_{i+1}) \leq h_i - h_{i+1}. \tag{2.8}$$

Como $\varphi'(h_i)$ está bem determinada em vista de (2.6) e $\varphi'(h_i) > 0$, obtemos

$$\frac{\varphi'(h_i)}{2t_2}\alpha^2 d^2(z_i, z_{i+1}) \leq \varphi'(h_i)(h_i - h_{i+1}).$$

A concavidade de φ implica:

$$\frac{\varphi'(h_i)}{2t_2}\alpha^2 d^2(z_i, z_{i+1}) \leq \varphi(h_i) - \varphi(h_{i+1}).$$

Vamos primeiramente checar o item (i) para $k = 0$ e $k = 1$. Devido a (2.7), z_0 pertence a $B(\bar{z}, \rho)$. Usando a desigualdade triangular, (2.7) e (2.8) segue que

$$\frac{\alpha^2}{2t_2} d^2(z_0, z_1) \leq h_0 - h_1 \leq h_0, \tag{2.9}$$

logo $z_1 \in B(\bar{z}, \rho)$. Vamos mostrar por indução em k que $z_k \in B(\bar{z}, \rho)$ para todo $k \geq 0$. Suponhamos que para algum $k \geq 2$ tenhamos $z_k \in B(\bar{z}, \rho)$. Agora, para $0 \leq i \leq k$, $z_i \in B(\bar{z}, \rho)$ e $0 < h_i < \eta$, podemos escrever a desigualdade de Kurdyka-Lojasiewicz em z_i:

$$1 \leq \varphi'(h_i)\,\mathrm{dist}(0, \partial H(z_i)).$$

Tomemos (u_i, v_i) como na Proposição 2.2 (iii) e recordemos que (u_i, v_i) é uma elemento de $\partial H(z_i)$, então, para $1 \leq i \leq k$:

$$1 \leq \varphi'(h_i)\|(u_i, v_i)\|. \tag{2.10}$$

Examinamos $\|(u_i, v_i)\|$, para $1 \leq i \leq k$. Como

$$\mathrm{grad}\, C_M^2(x_{i-1}, x_i) = 2C_M(x_{i-1}, x_i)\,\mathrm{grad}\, C_M(x_{i-1}, x_i)$$

e

$$\mathrm{grad}\, C_N^2(y_{i-1}, y_i) = 2C_N(y_{i-1}, y_i)\,\mathrm{grad}\, C_N(y_{i-1}, y_i),$$

temos

$$\begin{aligned}&\|(\frac{1}{2\lambda_{i-1}}\,\mathrm{grad}\, C_M^2(x_{i-1}, x_i), \frac{1}{2\mu_{i-1}}\,\mathrm{grad}\, C_N^2(y_{i-1}, y_i))\| \\ &\leq \frac{\tau\beta}{t_1}(d_M(x_{i-1}, x_i) + d_N(y_{i-1}, y_i)) \leq \frac{2\tau\beta}{t_1} d(z_{i-1}, z_i),\end{aligned}$$

em que usamos $d(z_{i-1}, z_i) = [d_M^2(x_{i-1}, x_i) + d_N^2(y_{i-1}, y_i)]^{1/2}$, além disso

$$d^2(\bar{z}, (x_i, y_{i-1})) \leq d^2(\bar{z}, z_i) + d^2(\bar{z}, z_{i-1}) \leq 2\rho^2.$$

Logo (x_i, y_{i-1}) e $z_i = (x_i, y_i)$ pertencem à $B(\bar{z}, \sqrt{2}\rho)$. Podemos então aplicar a desigualdade de Lipschitz

$$\|(\operatorname{grad}_x \Psi(x_i, y_i) - \operatorname{grad}_x \Psi(x_i, y_{i-1})\| \leq \delta d(z_{i-1}, z_i),$$

portanto, para $1 \leq i \leq k$

$$\|(u_i, v_i)\| \leq (\delta + 2\tau\beta {t_1}^{-1}) d(z_{i-1}, z_i), \tag{2.11}$$

usando (2.10)

$$[(\delta + 2\tau\beta t_1^{-1}) d(z_{i-1}, z_i)]^{-1} \leq \varphi'(h_i)$$

e usando (2.9)

$$\begin{aligned} \varphi(h_i) - \varphi(h_{i+1}) &\geq \frac{\varphi'(h_i)\alpha^2}{2t_2} d^2(z_i, z_{i+1}) \\ &\geq \frac{\alpha^2}{2t_2(\delta + 2\tau\beta {t_1}^{-1})} \frac{d^2(z_i, z_{i+1})}{d(z_{i-1}, z_i)}, \end{aligned}$$

a qual pode ser reescrita como

$$d(z_i, z_{i+1}) \leq ((\varphi(h_i) - \varphi(h_{i+1})) E d(z_{i-1}, z_i))^{1/2},$$

em que $E = 2t_2(\delta + 2\tau\beta {t_1}^{-1})\alpha^{-2}$. Para $a, b \in \mathbb{R}$, $2ab \leq a^2 + b^2$. Logo

$$2d(z_i, z_{i+1}) \leq (\varphi(h_i) - \varphi(h_{i+1}))E + d(z_{i-1}, z_i), \tag{2.12}$$

para todo $1 \leq i \leq k$. Somando-se a última desigualdade em i obtemos,

$$\sum_{i=1}^{k} d(z_i, z_{i+1}) + d(z_k, z_{k+1}) \leq (\varphi(h_1) - \varphi(h_{k+1}))E + d(z_0, z_1).$$

Logo, devido a monotonicidade de φ e h_k

$$\sum_{i=1}^{k} d(z_i, z_{i+1}) \leq \varphi(h_0)E + d(z_0, z_1). \tag{2.13}$$

De (2.9), temos

$$\begin{aligned} d(z_{k+1}, \bar{z}) &\leq \sum_{i=1}^{k} d(z_i, z_{i+1}) + d(z_1, \bar{z}) \\ &\leq \varphi(h_0)E + d(z_0, z_1) + d(z_0, \bar{z}) \\ &\leq \varphi(h_0)E + \alpha^{-1}\sqrt{2t_2 h_0} + d(\bar{z}, z_0). \end{aligned}$$

Assim, devido a (2.7) $z_{k+1} \in B(\bar{z}, \rho)$. Dessa forma, completamos a prova do item (i).

Agora, como a desigualdade (2.12) é verdadeira para $i \geq 1$, somamos em i, com $k \leq i \leq j$, para algum inteiro j,

$$\sum_{i=k}^{j} d(z_i, z_{i+1}) + d(z_j, z_{j+1}) \leq (\varphi(h_k) - \varphi(h_{j+1}))E + d(z_{k-1}, z_k).$$

Logo

$$\sum_{i=k}^{j} d(z_i, z_{i+1}) \leq \varphi(h_k)E + d(z_{k-1}, z_k).$$

Fazendo $j \to +\infty$

$$\sum_{i=k}^{+\infty} d(z_i, z_{i+1}) \leq \varphi(h_k)E + d(z_{k-1}, z_k). \tag{2.14}$$

Finalmente de (2.8) obtemos,

$$\sum_{i=k+1}^{+\infty} d(z_i, z_{i+1}) \leq \varphi(h_k)E + \alpha^{-1}\sqrt{2t_2(h_{k-1} - h_k)} \leq \varphi(h_k - \bar{h})E + \alpha^{-1}\sqrt{2t_2(h_{k-1} - \bar{h})},$$

assim,

$$\sum_{i=k+1}^{+\infty} d(z_i, z_{i+1}) < +\infty,$$

ou seja, (z_k) é uma sequência de Cauchy e portanto convergente. Pelo Lema 2.1 o limite de (z_k) é um ponto crítico de H. $\square$

Teorema 2.2 *Assumimos que a função H é KL e goza da hipótese $(\mathcal{A})$. Então, ou a sequência $(d(z_0, z_k))$ é ilimitada ou*

$$\sum_{i=1}^{+\infty} d(z_{k-1}, z_k) < +\infty,$$

fato que implica na convergência de (z_k) para um ponto crítico de H.

Demonstração: Suponhamos que $(d(z_0, z_k))$ não é ilimitada e que $\bar{z}$ é um ponto limite de (z_k), para o qual denotamos ρ, η e φ, os objetos associados à Definição 1.3.

Pelo Lema 2.1, $\bar{z}$ é um ponto crítico e (h_k) converge para 0.

Caso exista um inteiro k_0 tal que $H(z_{k_0}) = 0$, segue-se da Proposição 2.2 que $z_k = z_{k_0}$ para todo $k \geq k_0$, logo $z_{k_0} = \bar{z}$. Por outro lado, caso $h_k > 0$ e do fato de que o $\max\{\varphi(h_k), d(\bar{z}, z_k)\}$ admite 0 como um ponto de acumulação, obtemos $k_0 \geq 0$ tal que (2.7) é verificada para z_{k_0} como um novo ponto inicial. A conclusão da prova é uma consequência do Teorema 2.1. □

Teorema 2.3 *(Convergência local para um mínimo global). Assumimos que a função H é KL em $\bar{z}$ e goza da hipótese $(\mathcal{A})$. Sejam $\bar{z}$ um ponto de mínimo global de H, (z_k) a sequência gerada por* (2.1) *e z_0 um ponto inicial. Então existem ϵ e η tais que*

$$d(z_0, \bar{z}) < \epsilon, \quad \min H < H(z_0) < \min H + \eta.$$

Logo, a sequência (z_k) possui comprimento finito e converge para o ponto z^*, com $H(z^*) = \min H$.

Demonstração: Uma aplicação direta do Teorema 2.1 mostra que (z_k) converge para um ponto crítico z^* de H, com $H(z^*) \in [\min H, \min H + \eta)$. Agora, se $H(z^*)$ for diferente de $H(\bar{z})$, então a desigualdade de Kurdyka-Lojasiewicz implicaria

$$\varphi'(H(z^*) - H(\bar{z}))\,\mathrm{dist}(0, \partial H(z^*) \geq 1,$$

uma clara contradição, pois $0 \in \partial H(z^*)$. □

O próximo teorema apresenta um resultado de taxa de convergência, bastante desejado em um método iterativo.

Teorema 2.4 *Assumimos que H goza da hipótese $(\mathcal{A})$ e é uma função KL em z^*, com $\varphi(s) = cs^{1-\theta}$, onde $\theta \in [0,1)$ e $c > 0$. Se a sequência $(z_k = (x_k, y_k))$ converge para $z^* = (x^*, y^*)$, então:*

(i) se $\theta = 0$, a sequência (z_k) converge em um número finito de passos;

(ii) se $\theta \in (0, \frac{1}{2}]$, existem $c_0 > 0$ e $\varsigma \in [0,1)$ tais que

$$d(z_k, z^*) \leq c_0 \varsigma^k;$$

(iii) se $\theta \in (\frac{1}{2}, 1)$, existe $\xi > 0$ tal que

$$d(z_k, z^*) \leq \xi k^{-\frac{1-\theta}{2\theta-1}}.$$

Demonstração: Usamos as notações do Teorema 2.3. Sem perda de generalidade, podemos assumir que $H(z^*) = 0$.

(i) Se (h_k) é estacionária, então (z_k) também o é, conforme Proposição 2.2.

Se (h_k) não é estacionária, então da desigualdade de Kurdyka-Lojasiewicz, com $\varphi(s) = cs^{1-\theta}$, obtemos que $c \cdot \operatorname{dist}(0, \partial H(z_k)) \geq 1$, para todo k suficientemente grande. Por outro lado, como a sequência $z_k \to z^*$ e $\partial H(\cdot)$ é fechado, temos que $c \operatorname{dist}(0, \partial H(z_k)) \to 0$, o que é uma contradição. Portanto (z_k) converge em um número finito de passos.

Para completar a prova deste teorema consideramos $k \geq 0$ e

$$D(z_k) := \sum_{i=k+1}^{+\infty} [d_M^2(x_i, x_{i+1}) + d_N^2(y_i, y_{i+1})]^{1/2} = \sum_{i=k+1}^{+\infty} d(z_i, z_{i+1}),$$

o qual é finito pelo Teorema 2.2. Agora, para $k \in \mathbb{N}$ fixado, seja $j \in \mathbb{N}$, $j > k$. Logo, usando-se a desigualdade triangular,

$$d(z_k, z^*) \leq d(z_k, z_j) + d(z_j, z^*) \leq \sum_{i=k+1}^{j} d(z_i, z_{i+1}) + d(z_j, z^*),$$

fazendo $j \to +\infty$ na última desigualdade, temos

$$d(z_k, z^*) = [d_M^2(x_k, x^*) + d_N^2(y_k, y^*)]^{1/2} \leq D(z_k).$$

Com vistas às demonstrações dos itens (ii) e (iii) é suficiente estimar $D(z_k)$. Reescrevendo-se (2.14) obtemos,

$$D(z_k) \leq \varphi(h_k)E + D(z_{k-1}) - D(z_k). \tag{2.15}$$

Usando a desigualdade de Kurdyka-Lojasiewicz, segue-se que

$$1 \leq \varphi'(h_k) \operatorname{dist}(0, \partial H(z_k)) = c(1-\theta)h_k^{-\theta} \operatorname{dist}(0, \partial H(z_k)),$$

logo,

$$h_k^{\theta} \leq c(1-\theta) \operatorname{dist}(0, \partial H(z_k)).$$

Por outro lado, de (2.11) segue-se que

$$\operatorname{dist}(0, \partial H(z_k)) \leq \|(u_k, v_k)\| \leq (\delta + 2\tau\beta t_1^{-1})(D(z_{k-1}) - D(z_k)).$$

Assim,

$$h_k^{\theta} \leq c(1-\theta)(\delta + 2\tau\beta t_1^{-1})(D(z_{k-1}) - D(z_k))$$

e

$$h_k^{1-\theta} \leq [c(1-\theta)(\delta + 2\tau\beta t_1^{-1})]^{\frac{1-\theta}{\theta}} (D(z_{k-1}) - D(z_k))^{\frac{1-\theta}{\theta}}.$$

Usando-se as duas últimas desigualdades e fazendo $\Theta = c[c(1-\theta)(\delta + 2\tau\beta t_1^{-1})]^{\frac{1-\theta}{\theta}}$, obtemos

$$\varphi(h_k) = ch_k^{1-\theta} \leq \Theta(D(z_{k-1}) - D(z_k))^{\frac{1-\theta}{\theta}}.$$

Portanto, de (2.15)

$$D(z_k) \leq E\Theta(D(z_{k-1}) - D(z_k))^{\frac{1-\theta}{\theta}} + (D(z_{k-1}) - D(z_k)). \qquad (2.16)$$

(ii) De (2.16) e do fato de $\theta \in (0, 1/2]$, obtemos, para k suficientemente grande, uma constante positiva c_1 tal que

$$D(z_k) \leq c_1(D(z_{k-1}) - D(z_k)),$$

em que $c_1 = E\Theta + 1$. Logo

$$D(z_k) \leq \frac{c_1}{1+c_1} D(z_{k-1}).$$

Finalmente, usando recorrência em k, segue-se o resultado desejado

$$d(z_k, z^*) \leq D(z_k) \leq c_0 \varsigma^k,$$

em que $\varsigma = c_1/(1+c_1)$ e c_0 é uma constante positiva.
(iii) Para a prova desse item, utilizamos o roteiro abordado em [12, Theorem 2]. Primeiro, de (2.16) obtemos que existem inteiros $n_1 > n_0$ e uma constante $c_2 > 0$ tais que

$$D(z_k)^{\frac{\theta}{1-\theta}} \leq c_2(D(z_{k-1}) - D(z_k)),$$

para todo $k \geq n_1$, em que $c_2 = (E\Theta + 1)^{\frac{\theta}{1-\theta}}$. Segundo, usando argumentos técnicos de [12, Theorem 2] e o fato de que $\theta \in (1/2, 1)$, obtemos uma constante $r > 0$ tal que

$$0 < r \leq D(z_k)^{\frac{1-2\theta}{1-\theta}} - D(z_{k-1})^{\frac{1-2\theta}{1-\theta}}.$$

Portanto, para n suficientemente grande,

$$r(n - n_1) \leq D(z_n)^{\frac{1-2\theta}{1-\theta}} - D(z_{n_1})^{\frac{1-2\theta}{1-\theta}},$$

finalmente,

$$d(z_n, z^*) \leq D(z_n) \leq [r(n-n_1) + D(z_{n_1})^{\frac{1-2\theta}{1-\theta}}]^{\frac{1-\theta}{1-2\theta}} \leq \xi n^{-\frac{1-\theta}{2\theta-1}},$$

para algum $\xi > 0$. $\square$

Capítulo 3

Elementos de geometria Finsleriana e desigualdade de Kurdyka-Lojasiewicz

Na próxima seção introduzimos as notações, conceitos e propriedades fundamentais da geometria Finsleriana, os quais podem ser encontrados em [28] e [29].

3.1 Elementos de geometria Finsleriana

3.1.1 Conceitos básicos e exemplos

Para a próxima definição, consideramos uma variedade diferenciável (C^∞) M , de dimensão $m \in \mathbb{N}$ e também o fibrado tangente de M,

$$TM = \{(x, y) : x \in M, y \in T_xM\}.$$

Definição 3.1 *Uma função $F : TM \to \mathbb{R}_+$ tal que:*

(i) F é C^∞ em $TM \setminus \{0\}$;

(ii) para cada $x \in M$, $F_x := F_{|T_xM}$ é positivamente homogênea de grau 1, i.e., $F_x(ty) = tF_x(y)$ para todo $t > 0$ e $y \in T_xM$;

(iii) para cada $y \in T_xM \setminus \{0\}$, a forma bilinear simétrica g_y em T_xM é positiva definida, onde

$$g_y(u, v) = \frac{1}{2}\frac{\partial^2}{\partial s \partial t}[F_x^2(y + su + tv)]_{s=t=0};$$

é denominada métrica de Finsler. O par (M, F) é denominado variedade de Finsler.

Observação 3.1 *Em algumas situações, a métrica de Finsler F verifica a igualdade $F_x(-y) = F_x(y)$. Neste caso, diz-se que F é absolutamente homogênea.*

Observação 3.2 *Segue da Definição 3.1 que:*

$$\begin{aligned}
g_y(y,u) &:= \frac{1}{2}\frac{\partial^2}{\partial s\partial t}[F_x^2(y+ty+su)]_{s=t=0}, \\
&= \frac{1}{2}\frac{\partial}{\partial s}[F_x^2(y+su)]_{s=0},
\end{aligned} \tag{3.1}$$

$$\begin{aligned}
g_y(y,y) &:= \frac{1}{2}\frac{\partial^2}{\partial s\partial t}[F_x^2(y+sy+ty)]_{s=t=0} \\
&= \frac{1}{2}\frac{\partial^2}{\partial s\partial t}(1+s+t)^2[F_x^2(y)]_{s=t=0} \\
&= F_x^2(y)
\end{aligned} \tag{3.2}$$

e fixando uma base $\{b_i\}_{i=1}^m$ *para* T_xM *e expressando* $y = y^i b_i$,

$$g_{ij}(y) := g_y(b_i,b_j) = \left[\frac{1}{2}F^2(y)\right]_{y^iy^j}.$$

Mostramos a seguir alguns exemplos de métricas Finslerianas.

Exemplo 3.1 *(Métricas Riemannianas) Seja* $g = \{g_x\}_{x\in M}$ *, em que* g_x *é uma forma bilinear simétrica e positiva definida em* T_xM *tal que, em coordenadas locais* (x^i),

$$g_{ij}(x) := g_x\left(\frac{\partial}{\partial x^i},\frac{\partial}{\partial x^j}\right)$$

são funções C^∞. *Neste caso,* g *é chamada métrica Riemanniana. Definimos uma métrica de Finsler em* TM*, absolutamente homogênea, através do mecanismo*

$$F_x(y) := \sqrt{g_x(y,y)}.$$

Toda variedade Riemanniana (M,g) é portanto Finsleriana (M,F).

Exemplo 3.2 *(Métricas de Randers) Seja* $\alpha(y) = \sqrt{g_{ij}(x)y^iy^j}$ *uma métrica Riemanniana e* $\beta(y) = b_i(x)y^i$ *uma forma diferencial de grau um em* M*. Assumimos que para* $x \in M$ *e* $y \in T_xM$

$$\|\beta(y)\|_x = \sup_{\alpha(y)=1}\beta(y) < 1,$$

em que g_{ij} *são as componentes da métrica Riemanniana e* b_i *são as componetes da 1-forma. Uma métrica de Randers define uma métrica de Finsler em* TM *da seguinte forma*

$$F_x(y) := \alpha(y) + \beta(y).$$

Notemos que devido à presença do termo β, As métricas de Randers não verificam a igualdade $F_x(-y) = F_x(y)$ quando $b_i \neq 0$. De fato, $F_x(-y) = F_x(y)$ se, e somente se, M é uma variedade Riemanniana.

Uma interessante ilustração do Exemplo 3.2 em $\mathbb{R}^2$, tratado em [29, página 20]), é o seguinte:

Exemplo 3.3 *Seja* $M = \{x \in \mathbb{R}^2 : (x_1)^2 + (x_2)^2 < 2\}$ *o disco de Poincaré e*

$$F_x(y) = \frac{1}{1-\frac{r^2}{4}}\sqrt{y\cdot y} + \frac{r}{(1-\frac{r^2}{4})(1+\frac{r^2}{4})}dr(y),$$

em que $x = (x_1, x_2)$, $r^2 = (x_1)^2 + (x_2)^2$ *e* $y \in T_xM$.

3.1.2 Propriedades Fundamentais

Agora, apresentamos as definições de comprimento e distância na variedade que podem ser encontradas em [28, páginas 12 e 13]). Dada uma métrica de Finsler F numa variedade M e seja $\gamma : [0,1] \to M$ uma curva C^∞ por partes. Definimos o comprimento integral de γ por

$$L(\gamma) := \int_0^1 F_{\gamma(t)}(\dot{\gamma}(t))dt. \tag{3.3}$$

Para $x, z \in M$, denotamos $\Gamma(x,z)$ o conjunto de todas as curvas C^∞ por partes $\gamma : [0,1] \to M$ tais que $\gamma(0) = x$ e $\gamma(1) = z$. O comprimento L induz uma função

$$d_F : M \times M \to \mathbb{R}_+$$

definida por

$$d_F(x,z) := \inf_{\gamma\in\Gamma(x,z)} L(\gamma)$$

Pode-se mostrar que $d_F(x,z) \geq 0$, $d_F(x,z) = 0$ se, e somente se, $x = z$ e

$$d_F(x,z) \leq d_F(x,p) + d_F(p,z),$$

para todo $x, p, z \in M$.

Consideramos $\pi : TM \setminus \{0\} \to M$, $\pi(x,y) = x$, a projeção natural de TM e $\pi^*(TM)$ o pull-back do fibrado tangente TM. Diferente do caso da conexão de Levi-Civita em geometria Riemanniana, não existe uma única conexão no caso da geometria Finsleriana. Entre as conexões em $\pi^*(TM)$, escolhemos a de Chern, cujos coeficientes são denotados por Γ^k_{ij}, ver [29, página 38], a qual induz um tensor curvatura, denotado por R, ver [29, Chap. 3]. A conexão de Chern define uma derivada covariante $D_V U$ de um campo de vetor U, na direção $V \in T_xM$. Como em geral, os coeficientes da conexão de Chern Γ^k_{ij}, em coordenadas locais, dependem da direção, $D_V U$ é definida com um fixado vetor de referência. Em particular , seja $\sigma : [0, r] \to M$ uma curva suave com velocidade $T = T(t) = \dot{\sigma}(t)$. Sejam U e W campos de vetores ao longo de σ. Definimos $D_T U$ com respeito a W por

$$D_T U = \left[\frac{dU_i}{dt} + U^j T^k (\Gamma^k_{ij})_{(\sigma,W)}\right] \frac{\partial}{\partial x^i}_{|_{\sigma(t)}}, \tag{3.4}$$

em que $\{\frac{\partial}{\partial x^i}_{|_{\sigma(t)}}\}_{i=1\ldots m}$ é uma base de $T_{\sigma(t)}M$. Uma curva $\sigma : [0, r] \to M$, C^∞, com velocidade $T = \sigma'(t)$, tal que

$$D_T \left[\frac{T}{F(T)}\right] = 0$$

é uma geodésica. Se a velocidade (Finsleriana) da geodésica σ é constante, então

$$\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt}\frac{d\sigma^i}{dt}\Gamma^k_{ij}(\sigma, T) = 0 \qquad i = 1, \ldots, m = \dim M, \tag{3.5}$$

em que T é o campo velocidade associado a σ.

Uma variedade de Finsler (M, F), em que F é positivamente (mas, não necessariamente absolutamente) homogênea de grau 1 e toda geodésica $\sigma : [0, 1] \to M$, parametrizada de forma que sua velocidade finsleriana seja constante, pode ser extendida a $(0, \infty)$, é denominada geodesicamente completa (à frente). O Teorema de Hopf-Rinow (ver [29, página 168]) mostra várias caracterizações de variedades completas, uma delas é que qualquer par de pontos, x e z em M, pode ser ligado por um segmento geodésico minimizante (não necessariamente único). Doravante assumimos que todas as variedades são geodésicamente completas.

Dados $x \in M$ e $y \in T_xM$ definimos a aplicação exponencial

$$\exp_x : T_xM \to M,$$

$$\exp_x(y) = \sigma(1, x, y)$$

em que $\sigma(t, x, y)$ é única solução (geodésica) da equação diferencial (3.5), que passa por x em $t = 0$ com velocidade y. Ademais,

$$\mathrm{d}(\exp_x)(0) = id_{T_xM}. \tag{3.6}$$

Para U, V e W, campos de vetores ao longo de uma curva σ de velocidade $T = \sigma'$, temos a seguinte regra de derivação

$$\frac{d}{dt} g_W(U, V) = g_W(D_T U, V) + g_W(U, D_T V).$$

Seja (M, F) uma variedade de Finsler. Para um vetor $y \in T_xM \setminus \{0\}$, a curvatura Riemanniana $R_y : T_xM \to T_xM$ é uma tranformação linear auto adjunta com relação a g_y. Seja $P \subset T_xM$. Dado $y \in P \setminus \{0\}$, o número $K(P, y)$ dado por

$$K(P, y) = \frac{g_y(R_y(u), u)}{g_y(y, y)g_y(u, u) - g_y(y, u)g_y(y, u)} ,$$

em que $u \in P$ é tal que $\{y, u\}$ geram P, é denominado curvatura flag, relativa ao flag (P, y) de T_xM.

Quando F é Riemanniana, a curvatura flag coincide com a curvatura seccional. Dizemos que a curvatura flag é não positiva caso $K(P, y) \leq 0$, para todo flag (P, y).

Definição 3.2 *Uma variedade de Finsler (M, F), simplesmente conexa, geodésicamente completa e de curvatura flag não positiva, é denominada uma variedade de Hadamard.*

Teorema 3.1 *(Cartan-Hadamard) Seja (M, F) uma variedade de Finsler, simplesmente conexa, geodésicamente completa e de curvatura flag não positiva. Então, a aplicação exponencial $\exp_x$ é um difeomorfismo de classe C^1, de T_xM em M.*

Demonstração: Ver [29, página 238]. □

O próximo lema estabelece que a distância de um ponto $x_1 \in M$ a um ponto $x_2 \in M$ é relativamente "próxima"da distância de x_2 a x_1.

Lema 3.1 *Seja (M, F) uma variedade de Finsler. Então, fixado um ponto $x \in M$, existem uma vizinhaça coordenada U de x e uma constante $c_0 > 1$ tais que:*

(i) para todo $y \in T_xM$ e $x \in \overline{U}$,

$$F(-y) \leq c_0^2 F(y)$$

(ii) para todo $x_1, x_2 \in U$,

$$\frac{1}{c_0^2} d(x_1, x_2) \leq d(x_2, x_1) \leq c_0^2 d(x_1, x_2).$$

Demonstração: Ver prova em [29, página 146]. □

3.1.3 Gradiente em variedade de Finsler

Primeiro consideramos a relação entre uma norma de Minkowski e sua norma dual (ver [28, Definition 1.2.1]). Sejam $\overline{E}$ um espaço vetorial de dimensão finita e $\overline{E}^*$ o seu dual. Dada uma norma de Minkowski F em $\overline{E}$, onde F é uma norma no seguinte sentido: para todo $y, v \in \overline{E}$ e $t > 0$,

$$F(ty) = tF(y),$$

e

$$F(y+v) \leq F(y) + F(v).$$

Seja $\overline{E}^*$ o espaço dual $\overline{E}$. Definimos

$$F^*(\xi) := \sup_{F(y)=1} \xi(y),$$

para todo $\xi \in \overline{E}^*$. F^* deiniine uma norma de Minkowski em $\overline{E}^*$.

O resultado do próximo lema dá uma caracterização de gradiente no contexto das variedades de Finsler.

Lema 3.2 *Sejam F uma norma de Minkowski em $\overline{E}$ e F^* a norma dual em $\overline{E}^*$. Para $y \in \overline{E} \setminus \{0\}$ e $\xi = g_y(y, \cdot) \in \overline{E}^*$, temos*

$$F(y) = F^*(\xi) = \frac{\xi(y)}{F(y)}.$$

Ademais, dado $\xi \in \overline{E}^* \setminus \{0\}$, existe um único vetor $y \in \overline{E} \setminus \{0\}$ tal que $\xi = g_y(y, \cdot)$.

Demonstração: Ver [28, página 35]. $\square$

Dada uma função diferenciável $f : M \to \mathbb{R}$, a diferencial $\mathbf{d}f_x \in \bar{E}^*$, no ponto $x \in M$,

$$\mathbf{d}f_x = \frac{\partial f}{\partial x^i}(x) dx^i,$$

é um funcional linear em T_xM. Para associar $\mathbf{d}f_x$ a um vetor grad $f_x \in T_xM$, necessitamos de uma norma de Minkowski em T_xM. Dada uma métrica de Finsler F em M, por definição, F_x define uma norma de Minkowski em T_xM. Suponhamos que $\mathbf{d}f_x \neq 0$. Usando que a indicatriz $S := F^{-1}(1)$ é fortemente convexa, existe um único vetor unitário $s_x \in S_xM := F_x^{-1}(1)$ e um número positivo $t_x > 0$ tal que

$$W^{t_x} = \{v : \mathbf{d}f_x(v) = t_x\}$$

é tangente a S_xM em s_x. Do Lema 3.2

$$\mathbf{d}f_x(v) = t_x g_{s_x}(s_x, v)$$

em que

$$F_x^*(\mathbf{d}f_x) = \mathbf{d}f_x(s_x) = t_x g_{s_x}(s_x, s_x) = t_x.$$

Definimos

$$\text{grad } f_x := t_x s_x = F_x^*(\mathbf{d}f_x)s_x,$$

assim,

$$\mathbf{d}f_x(v) = g_{\text{grad } f_x}(\text{grad } f_x, v), v \in T_xM. \tag{3.7}$$

Observação 3.3 *Destacamos que o gradiente, em geral, é não linear, i.e.,* $\text{grad}\,(f_x + \bar{f}_x) \neq \text{grad } f_x + \text{grad } \bar{f}_x$*, pois a transformada de Legendre é não linear, ver [29, página 406].*

Para o próximo lema, sejam M uma variedade Finsler-Hadamard e $d = d_F$ a distância em M. Dado um compacto $S \subset M$, definimos as distâncias: $d_+(x) := d(S, x)$ e $d_-(x) := -d(x, S)$.

Lema 3.3 *Seja M uma variedade Finsler-Hadamard. Dado um compacto $S \subset M$ e sejam $d_+(x)$ e $d_-(x)$. Então*

$$F(\text{grad } d_+(x)) = 1 \text{ e } F(\text{grad } d_-(x)) = 1$$

Demonstração: Ver [28, página 44]. □

Dado $x \in M$ a inversa da aplicação exponencial

$$\exp_x^{-1} : M \to T_xM$$

é dada, no cenário das variedades de Finsler Hadamard, por

$$d(x, z) = F_x(\exp_x^{-1}(z)).$$

3.2 Desigualdade de Kurdyka-Lojasiewicz no contexto das variedades de Finsler

Na Seção 1.3 apresentamos a desigualdade de Kurdyka-Lojasiewicz no contexto Riemanniano. Nesta seção faremos as devidas adequações para o caso das variedades de Finsler.

Observação 3.4 *Para a definição seguinte, consideramos (M, F) uma variedade de Finsler, $f : M \to \mathbb{R}$ uma função de classe C^1 e o conjunto:*

$$[\eta_1 < f < \eta_2] := \{x \in M : \eta_1 < f(x) < \eta_2\}, \quad -\infty < \eta_1 < \eta_2 < +\infty.$$

Definição 3.3 *Uma função f, conforme Observação 3.4, goza da propriedade de Kurdyka-Lojasiewicz em $\bar{x} \in M$ se existem $\eta \in (0, +\infty]$, uma vizinhaça U de $\bar{x}$ e uma função contínua e côncava $\varphi : [0, \eta) \to \mathbb{R}_+$, tal que:*

(i) $\varphi(0) = 0$, $\varphi \in C^1(0, \eta)$ *e, para todo* $s \in (0, \eta)$, $\varphi'(s) > 0$;

(ii) *para todo* $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, *a desigualdade de Kurdyka-Lojasiewicz*

$$\varphi'(f(x) - f(\bar{x}))F(grad\, f(x)) \geq 1 \tag{3.8}$$

é verdadeira.

Observação 3.5 *Analogamente ao caso Riemanniano, uma função f que verifica a desigualdade de Kurdyka-Lojasiewicz em cada ponto de M é chamada de função KL.*

Lema 3.4 *Dados uma função* $f : M \to \mathbb{R}$, C^1, *e* $\bar{x} \in M$, *tais que* $\text{grad } f(\bar{x}) \neq 0$. *Então f goza da propriedade de Kurdyka-Lojasiewicz em $\bar{x}$.*

Demonstração: Seguimos o roteiro da prova do Lema 1.1. Como $\text{grad } f(\bar{x}) \neq 0$, seja

$$\delta := F(\text{grad } f(\bar{x})) > 0.$$

Tomemos $\varphi(t) := t/\delta$, $U := B(\bar{x}, \delta/2)$, $\eta := \delta/2$ e notemos que, para cada $x \in M$,

$$\varphi'(f(x) - f(\bar{x}))F(\text{grad } f(x)) = F(\text{grad } f(x))/\delta. \tag{3.9}$$

Agora, para cada $x \in U \cap [f(\bar{x}) - \eta < f < f(\bar{x}) + \eta]$ arbitrário, notemos que

$$d(x, \bar{x}) + |f(x) - f(\bar{x})| < \delta.$$

Afirmação: para cada x que torna a última desigualdade verdadeira, temos

$$F(\text{grad } f(x)) \geq \delta. \tag{3.10}$$

Suponhamos por contradição que 3.10 não é verdadeira. Neste caso, existiriam uma sequência $\{(z_k, \text{grad } f(z_k))\} \subset \text{graph}\, \text{grad } f$ e $\{\delta_k\} \subset \mathbb{R}_{++}$ tais que

$$d(z_k, \bar{x}) + |f(z_k) - f(\bar{x})| < \delta_k, \quad \text{e} \quad F(\text{grad } f(z_k)) \leq \delta_k,$$

com $\{\delta_k\}$ convergindo para zero. Assim , usando que $\{(z_k, \text{grad } f(z_k))\}$ e $\{f(z_k)\}$ convergem para $(\bar{x}, 0)$ e $f(\bar{x})$ respectivamente, e $F(\text{grad } f)$ é continua, concluimos que $\bar{x}$ é um ponto crítico de f, fato que prova nossa afirmação.

Portanto, o resultado do lema decorre da combinação de (3.9) com (3.10). $\square$

Doravante, assumimos sem maiores detalhes que alguns fatos tratados no Capítulo 1 sobre estruturas o-minimal e categorias geométricas analíticas se aplicam ao cenário das variedades de Finsler.

Observação 3.6 *O resultado do Teorema 1.3, que estabelece uma relação biunívoca entre estruturas o-minimal e categorias geométricas analíticas, continua válido no caso Finsleriano, assim como o da Proposição 1.2, já que usamos a hipótese de que a função f é C^1.*

Teorema 3.2 *Sejam (M, F) uma variedade Finsleriana analítica e $f : M \to \mathbb{R}$ uma $\mathcal{C}$-função C^1. Então f é uma função KL e a função φ de (3.8) é definível em $\mathcal{O}$.*

Demonstração: Consideremos um ponto crítico $\bar{x} \in M$ de f e uma carta analítica local $\phi : V \to \mathbb{R}^n$, com a vizinhaça $V \subset M$ de $\bar{x}$ escolhida de forma que V e $f(V)$ sejam conjuntos limitados. Assim, de acordo com Proposição 1.2, a função $f \circ \phi^{-1} : \phi(V) \to \mathbb{R}$ é definível em $\mathcal{O}(\mathcal{C})$. Logo, como $\phi(V)$ é um conjunto aberto e limitado definível contendo $\bar{z} = \phi(\bar{x})$ e ϕ é definível, podemos aplicar o resultado do Teorema 1.4 , com $U = \phi(V)$. Seguindo a demonstração do Teorema 1.5 obtemos que a função $h = f \circ \phi^{-1}$ goza da propriedade de Kurdyka-Lojasiewicz em $\bar{z} = \phi(\bar{x})$, i.e., existem $\eta \in (0, +\infty]$ e uma função côncava contínua $\Phi : [0, \eta) \to \mathbb{R}_+$ tais que:

(i) $\Phi(0) = 0$, $\Phi \in C^1(0, \eta)$ e para todo $s \in (0, \eta)$, $\Phi'(s) > 0$;

(ii) para todo $z \in U \cap [h(\bar{z}) < h < h(\bar{z}) + \eta]$,

$$\Phi'(h(z) - h(\bar{z}))F(grad\, h(z)) \geq 1.$$

Como ϕ é um difeomorfismo e usando $z = \phi(x)$, $\bar{z} = \phi(\bar{x})$ e $h = f \circ \phi^{-1}$, a última desigualdade da Proposição 1.1 implica que

$$\Phi'(f(x) - f(\bar{x}))F((\phi_x^*)^{-1} \operatorname{grad} f(x)) \geq 1, \quad x \in V \cap [0 < f < f(\bar{x}) + \eta],$$

em que ϕ_x^* denota a derivada adjunta da função ϕ.

Tomemos um conjunto aberto $V' \subset V$ de forma que $K = \overline{V'}$ esteja contido no interior do conjunto V e $\bar{x} \in V'$. Logo, K é um conjunto compacto e para cada $x \in K$ existe $C_x > 0$ com

$$F((\phi_x^*)^{-1}w) \leq C_x F(w), \qquad w \in T_xM.$$

Novamente, usando que K é um conjunto compacto e $(\phi_x^*)^{-1}$ é um difeomorfismo, existe uma constante positiva $C := \sup\{C_x : x \in K\}$ tal que

$$F((\phi_x^*)^{-1}w) \leq CF(w), \qquad w \in T_xM,\ x \in K.$$

Portanto, para $x \in V' \cap [0 < f < f(\bar{x}) + \eta]$, temos

$$1 \leq \Phi'(f(x) - f(\bar{x}))F((\phi_x^*)^{-1} \operatorname{grad} f(x)) \leq C\,\Phi'(f(x) - f(\bar{x}))F(\operatorname{grad} f(x)).$$

Logo, f goza da propiedade de Kurdyka-Lojasiewicz em $\bar{x}$, com $\varphi = C\,\Phi$. Como $\bar{x}$ é um ponto crítico arbitrário, concluimos do Lema 3.4 que f é uma função KL. A segunda parte também decorre do Teorema 1.4 e assim encerramos a prova. $\square$

Observação 3.7 *O resultado do Teorema 3.2 mostra a existência de funções KL no cenário das variedades Finslerianas, o qual é importante para se obter convergência da sequência gerada pelo método, pois supomos que a função a ser minimizada é KL, ver detalhes Seção 4.1.*

Capítulo 4

Método do ponto proximal no contexto de variedades de Finsler

Neste capítulo estendemos o método do ponto proximal para variedades de Finsler, vamos estabelecer as condições para garantir a boa definição e convergência da sequência gerada pelo método.

Sejam (M, F) uma variedade de Finsler e $f : M \to \mathbb{R}$ uma função C^1. Consideramos o problema

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

Para resolver o problema (4.1), usamos o método do ponto proximal que, para um dado $x_0 \in M$, gera de forma iterativa uma sequência $\{x_k\} \subset M$ tal que:

$$x_{k+1} = \operatorname{argmin}_{z \in M} \ \{f(z) + \frac{1}{\lambda_k} C_{x_k}(z)\} \tag{4.2}$$

em que $C_{x_k} : M \to \mathbb{R}$ é definida por

$$C_{x_k}(z) = (1/2) d^2(x_k, z),$$

onde d é a distância Finsleriana e (λ_k) é uma sequência de números positivos.

4.1 Boa definição e convergência

Doravante estabelecemos que M é uma variedade de Finsler Hadamard, f é uma função KL, $\inf f > -\infty$ e para $t_1 < t_2$ positivos, $\lambda_k \in (t_1, t_2)$, para todo $k \in \mathbb{N}$.

Proposição 4.1 *Seja (x_k) a sequência gerada por* (4.2). *Então (x_k) está bem definida, ademais:*

$$f(x_{k+1}) + \frac{1}{2\lambda_k} d^2(x_k, x_{k+1}) \leq f(x_k) \tag{4.3}$$

e consequentemente

$$\sum_{k=0}^{\infty} d^2(x_k, x_{k+1}) < \infty. \tag{4.4}$$

Demonstração: Como M é uma variedade de Hadamard e $\inf f > -\infty$, segue-se que, para $r > 0$ e $\bar{x} \in M$, a função

$$x \to f(x) + \frac{1}{2r} d^2(\bar{x}, x)$$

é um 1-coerciva e x_{k+1} é determinado de forma única, ver [5]. Uma indução sobre k nos dá que o método é de descida, ou seja, a desigualdade (4.3) ocorre. A partir de (4.3) e da soma telescópica obtemos (4.4). □

No próximo lema estimamos a norma do gradiente em termos da função distância.

Lema 4.1 *Seja (x_k) a sequência gerada por (4.2) e $x_{k_0} \in B(\tilde{x}, \rho) \subset U$, onde a vizinhança U é dada pelo Lema 3.1. Então*

$$F(\text{grad } f(x_{k_0})) \leq c_0^2 t_1^{-1} d(x_{k_0-1}, x_{k_0}),$$

em que $c_0 > 1$ e $t_1 < \lambda_{k_0}$.

Demonstração: Como x_{k_0} é um minimizador da função

$$z \to f(z) + \frac{1}{2\lambda_{k_0-1}} d^2(x_{k_0-1}, z),$$

temos

$$\text{grad} \left(f(x_{k_0}) + \frac{1}{2\lambda_{k_0-1}} d(x_{k_0-1}, x_{k_0}) \right) = 0,$$

consequentemente,

$$\mathbf{d} \left(f(x_{k_0}) + \frac{1}{2\lambda_{k_0-1}} d(x_{k_0-1}, x_{k_0}) \right) = 0.$$

Usando a linearidade da diferencial, segue-se que

$$\mathbf{d}f(x_{k_0}) = -\mathbf{d} \left(\frac{1}{2\lambda_{k_0-1}} d^2(x_{k_0-1}, x_{k_0} \right).$$

Denotamos $h(x) := (1/(2\lambda_{k-1})(d^2(x_{k-1}, x))$ e $F_x = F$. De (3.7)

$$\mathbf{d}f(x_{k_0})(\text{grad } f(x_{k_0})) = g_{\text{grad } f(x_{k_0})}(\text{grad } f(x_{k_0}), \text{grad } f(x_{k_0}))$$

e

$$\mathbf{d}h(x_{k_0})(\text{grad } f(x_{k_0})) = g_{\text{grad } h(x_{k_0})}(\text{grad } h(x_{k_0}), \text{grad } f(x_{k_0})).$$

Assim,

$$\begin{aligned} g_{\text{grad } f(x_{k_0})}(\text{grad } f(x_{k_0}), \text{grad } f(x_{k_0})) &= -g_{\text{grad } h(x_{k_0})}(\text{grad } h(x_{k_0}), \text{grad } f(x_{k_0})) \\ &= g_{\text{grad } h(x_{k_0})}(\text{grad } h(x_{k_0}), -\text{grad } f(x_{k_0})). \end{aligned}$$

Da desigualdade de Cauchy-Schwartz ($g_y(y,v) \leq F(y)F(v)$, ver [28, página 11]) e do Lemma 3.1, concluimos que

$$\begin{aligned} F^2(\text{grad } f(x_{k_0})) &\leq F(\text{grad } h(x_{k_0}))F(-\text{grad } f(x_{k_0})) \\ &\leq F(\text{grad } h(x_{k_0}))c_0^2 F(\text{grad } f(x_{k_0})). \end{aligned}$$

Por outro lado,

$$\text{grad } h(x_{k_0}) = \frac{1}{\lambda_{k_0-1}} d(x_{k_0-1}, x_{k_0}) \, \text{grad } \; d(x_{k_0-1}, x_{k_0}).$$

Como F é positivamente homogênea,

$$F(\text{grad } f(x_{k_0})) \leq c_0^2 F(\text{grad } h(x_{k_0})) = c_0^2 \frac{1}{\lambda_{k_0-1}} d(x_{k_0-1}, x_{k_0}) F(\text{grad } \; d(x_{k_0-1}, x_{k_0})).$$

Fazendo $S = \{x_{k_0-1}\}$ e $x = x_{k_0}$ no Lema 3.3, $F(\text{grad } \; d(x_{k_0-1}, x_{k_0})) = 1$. Logo,

$$F(\text{grad } f(x_{k_0})) \leq c_0^2 t_1^{-1} d(x_{k_0-1}, x_{k_0}).$$

□ A seguir, apresentamos um resultado que é usado na análise de convergência da sequência gerada pelo algoritmo 4.2.

Lema 4.2 *Seja (a_k) uma sequência de números positivos tal que*

$$\sum_{k=1}^{+\infty} a_k^2 / a_{k-1} < +\infty.$$

Então, $\sum_{k=1}^{+\infty} a_k < +\infty$.

Demonstração: Fixamos $j \in \mathbb{N}$. Observamos que,

$$\sum_{k=1}^{j} a_k = \sum_{k=1}^{j} \frac{a_k}{\sqrt{a_{k-1}}} \sqrt{a_{k-1}} \leq \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}} \right)^{1/2} \left(\sum_{k=1}^{j} a_{k-1} \right)^{1/2},$$

em que a última desigualdade segue da desigualdade de Cauchy-Schwartz em $\mathbb{R}^j$, com relação aos vetores $\left(a_1/\sqrt{a_0}, \ldots, a_j/\sqrt{a_{j-1}}\right)$ e $\left(\sqrt{a_0}, \ldots, \sqrt{a_{j-1}}\right)$. Logo,

$$\sum_{k=1}^{j} a_k \le \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2} \left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2}.$$

Adicionamos a_0 a ambos os lados da última desigualdade e observamos que $a_j > 0$, assim

$$\sum_{k=1}^{j} a_{k-1} \le a_0 + \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2} \left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2}.$$

Portanto, dividindo-se ambos os lados da última desigualdade por $\left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2}$ e observando que

$$a_0/\left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2} \le \sqrt{a_0} \qquad (a_k > 0, k = 0, 1, \ldots),$$

segue-se que

$$\left(\sum_{k=1}^{j} a_{k-1}\right)^{1/2} \le \sqrt{a_0} + \left(\sum_{k=1}^{j} \frac{a_k^2}{a_{k-1}}\right)^{1/2},$$

usando o teorema da comparação para séries de números reais, concluimos a prova.
□

Lema 4.3 *Sejam (x_k) a sequência gerada por (4.2), $f : M \to \mathbb{R}$ uma função C^1, $\tilde{x}$ um ponto de acumulação de (x_k), $a = 1/2t_2$, $b = c_0^2/t_1$ constantes e $\rho > 0$ tal que $B(\tilde{x}, \rho) \subset U$, onde a vizinhança U é dada pelo Lemma 3.1. Então existe $k_0 \in \mathbb{N}$ tal que*

$$f(\tilde{x}) < f(x_k) < f(\tilde{x}) + \eta, \qquad k \ge k_0, \tag{4.5}$$

$$d(\tilde{x}, x_{k_0}) + 2\sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}} + \frac{b}{a}\varphi(f(x_{k_0}) - f(\tilde{x})) < \rho. \tag{4.6}$$

Além disso,

$$\frac{b}{a}[\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x}))] \ge \frac{d^2(x_{k_0}, x_{k_0+1})}{d(x_{k_0-1}, x_{k_0})}. \tag{4.7}$$

Em particular, se $x_k \in B(\tilde{x}, \rho)$ para todo $k \ge k_0$, então $\sum_{k=k_0}^{+\infty} d(x_k, x_{k+1}) < +\infty$, logo a sequência (x_k) converge para $\tilde{x}$.

Demonstração: Seja (x_{k_j}) uma subsequência de (x_k), corvergente para $\tilde{x}$. De (4.3), e da continuidade de f, $(f(x_{k_j}))$ converge para $f(\tilde{x})$. Pela Proposição 4.1, $(f(x_k))$ é uma sequência decrescente, logo toda a sequência $(f(x_k))$ converge para

$f(\tilde{x})$, quando k tende $+\infty$, e portanto,

$$f(\tilde{x}) < f(x_k), \qquad k \in \mathbb{N}. \tag{4.8}$$

em particular, existe $N \in \mathbb{N}$ tal que

$$f(\tilde{x}) < f(x_k) < f(\tilde{x}) + \eta, \qquad k \geq N. \tag{4.9}$$

Devido a (4.8), definimos a sequência (b_k), por

$$b_k = d(\tilde{x}, x_k) + 2\sqrt{\frac{f(x_k) - f(\tilde{x})}{a}} + \frac{b}{a}\varphi(f(x_k) - f(\tilde{x})).$$

Como $d(\cdot, \tilde{x})$ e φ são contínuas, então 0 é um ponto de acumulação de $\{b_k\}$, e portanto, existe $k_0 := k_{j_0} > N$ tal que (4.6) ocorre. Em particular, como $k_0 > N$, de (4.9) temos (4.5).

De (4.5) e do fato de $x_{k_0} \in B(\tilde{x}, \rho)$ (que decorre de (4.6)), temos

$$x_{k_0} \in B(\tilde{x}, \rho) \cap [f(\tilde{x}) < f < f(\tilde{x}) + \eta].$$

Como f verifica a desigualdade de Kurdyka-Lojasiewicz em $\tilde{x}$, grad $f(x_{k_0}) \neq 0$. Do Lema 4.1

$$F(\text{grad } f(x_{k_0})) \leq c_0^2 t_1^{-1} d(x_{k_0 - 1}, x_{k_0},) = bd(x_{k_0-1}, x_{k_0}),$$

em que $b = c_0^2 t_1^{-1}$. Usando que f verifica a desigualdade de Kurdyka-Lojasiewicz em $\tilde{x}$, obtemos

$$\varphi'(f(x_{k_0}) - f(\tilde{x})) \geq \frac{1}{bd(x_{k_0-1}, x_{k_0})}. \tag{4.10}$$

Por outro lado, a concavidade de φ implica

$$\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x})) \geq \varphi'(f(x_{k_0}) - f(\tilde{x}))(f(x_{k_0}) - f(x_{k_0+1})).$$

A última desigualdade, combinada com o fato de $\varphi' > 0$ e (4.3), implica

$$\varphi(f(x_{k_0}) - f(\tilde{x})) - \varphi(f(x_{k_0+1}) - f(\tilde{x})) \geq \varphi'(f(x_{k_0}) - f(\tilde{x}))ad^2(x_{k_0}, x_{k_0+1}),$$

relembramos que $a = 1/2t_2 \leq 1/2\lambda_{k_0-1}$. Portanto, da última desigualdade e de (4.10), segue-se (4.7).

A prova da última parte deste lema segue de (4.7) e do Lemma 4.2. $\square$

Lema 4.4 *Seja* (x_k) *a sequência gerada por* (4.2) *e assumamos as hipóteses do*

Lema 4.3. Então, existe $k_0 \in \mathbb{N}$ tal que

$$x_k \in B(\tilde{x}, \rho), \qquad k > k_0. \tag{4.11}$$

Demonstração: A prova segue por indução sobre k. Observamos que desigualdade (4.3) implica que a sequência $(f(x_k))$ é não crescente e

$$d(x_k, x_{k+1}) \leq \sqrt{\frac{f(x_k) - f(x_{k+1})}{a}}, \qquad k \in \mathbb{N}. \tag{4.12}$$

Além disso, como a função f é contínua, do Lema 4.3, obtemos que existe $k_0 \in \mathbb{N}$ tal que (4.6) e (4.5) são verdadeiras, logo

$$x_{k_0} \in B(\tilde{x}, \rho), \qquad 0 < f(x_{k_0}) - f(x_{k_0+1}) < f(x_{k_0}) - f(\tilde{x}), \tag{4.13}$$

a última desigualdade e (4.12) $(k = k_0)$ implicam

$$d(x_{k_0}, x_{k_0+1}) \leq \sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}}. \tag{4.14}$$

Usando a desigualdade triangular, a última desigualdade e (4.6), temos

$$d(\tilde{x}, x_{k_0+1}) \leq \sqrt{\frac{f(x_{k_0}) - f(\tilde{x})}{a}} + d(\tilde{x}, x_{k_0}) < \rho,$$

logo, $x_{k_0+1} \in B(\tilde{x}, \rho)$.

Suponhamos que (4.11) ocorre para todo $k = k_0 + 1, \ldots, k_0 + j - 1$. Neste caso, para $k = k_0 + 1, \ldots, k_0 + j - 1$, temos (4.7) e, consequentemente

$$\sqrt{d(x_{k-1}, x_k)(b/a)[\varphi(f(x_k) - f(\tilde{x})) - \varphi(f(x_{k+1}) - f(\tilde{x}))]} \geq d(x_k, x_{k+1}). \tag{4.15}$$

Usando que $r + s \geq 2\sqrt{rs}$, para todo $r, s \geq 0$, e $k = k_0 + 1, \ldots, k_0 + j - 1$, ainda fazendo

$$r = d(x_{k-1}, x_k), \; s = (b/a)[\varphi(f(x_k) - f(\tilde{x})) - \varphi(f(x_{k+1}) - f(\tilde{x}))],$$

em (4.15), obtemos

$$2d(x_k, x_{k+1}) \leq d(x_{k-1}, x_k) + \frac{b}{a}[\varphi(f(x_k) - f(\tilde{x})) - \varphi(f(x_{k+1}) - f(\tilde{x}))].$$

Logo, somando mebro a membro, com $k = k_0 + 1, \ldots, k_0 + j - 1$, temos

$$\sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}) + d(x_{k_0+j-1}, x_{k_0+j}) \leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}[\varphi(f(x_{k_0+1}) - f(\tilde{x})) \\ - \varphi(f(x_{k_0+j}) - f(\tilde{x}))],$$

consequentemente,

$$\begin{aligned} \sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}) &\leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi(f(x_{k_0+1}) - f(\tilde{x})) \\ &\leq d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi(f(x_{k_0}) - f(\tilde{x})), \end{aligned} \tag{4.16}$$

em que a última desigualdade segue da segunda desigualdade de (4.13) e da monotonicidade da função φ. Usando a desigualdade triangular e o fato de que a distância $d(x, z) \geq 0$, para todo $x, z \in M$, obtemos a relação

$$d(\tilde{x}, x_{k_0+j}) \leq d(x_{k_0}, x_{k_0+j}) + d(\tilde{x}, x_{k_0}) \leq d(\tilde{x}, x_{k_0}) + d(x_{k_0}, x_{k_0+1}) + \sum_{i=k_0+1}^{k_0+j-1} d(x_i, x_{i+1}),$$

que, combinada com (4.16), implicam na desigualdade

$$d(\tilde{x}, x_{k_0+j}) \leq d(\tilde{x}, x_{k_0}) + 2d(x_{k_0}, x_{k_0+1}) + \frac{b}{a}\varphi(f(x_{k_0}) - f(\tilde{x})).$$

Finalmente, da última desigualdade, de (4.14) e (4.6), concluimos a prova, i.e., $x_{k_0+j} \in B(\tilde{x}, \rho)$. $\square$

No teorema seguinte provamos a convergência completa da sequência (x_k), gerada por (4.2), para um ponto crítico $\tilde{x}$ de uma função que verifica a desigualdade de Kurdyka-Lojasiewicz em $\tilde{x}$.

Teorema 4.1 *Sejam U, η e $\varphi : [0, \eta) \to \mathbb{R}_+$, os elementos dados na Definição 3.3. Se $\tilde{x} \in M$ é um ponto de acumulação da sequência (x_k), $\rho > 0$ tal que $B(\tilde{x}, \rho) \subset U$, f é C^1 e verifica a desigualdade de Kurdyka-Lojasiewicz em $\tilde{x}$. Então existe $k_0 \in \mathbb{N}$ tal que*

$$\sum_{k=k_0}^{+\infty} d(x_k, x_{k+1}) < +\infty. \tag{4.17}$$

Ademais, $f(x_k) \to f(\tilde{x})$, quando $k \to +\infty$, a sequência (x_k) converge para $\tilde{x}$ e $\tilde{x}$ é um ponto crítico de f.

Demonstração: O Lema 4.4 e o Lema 4.3 implicam (4.17), em particular a sequência (x_k) converge para $\tilde{x} \in M$. Logo, usando a continuidade da função f e (4.3),

$f(x_k) \to f(\tilde{x})$, quando $k \to +\infty$. De (4.17) e do Lemma 4.1,

$$F(\text{grad } f(x_k)) \leq bd(x_k, x_{k+1}) \to 0,$$

para $k = k_0$. Como F é uma norma de Minkowski, $F(gradf(\tilde{x})) = 0$ implica $gradf(\tilde{x}) = 0$. Portanto, $\tilde{x}$ é um ponto crítico de f. $\square$

Na próxima seção obtemos estimativas para velocidade de convergência da sequência gerada pelo algoritmo 4.2.

4.2 Taxa de Convergência

Teorema 4.2 *Assumindo as hipóteses do Lema 4.3, e ainda que (x_k) converge para x^* e f é uma função KL em x^*, com $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1)$ e $c > 0$, então:*

(i) se $\theta = 0$, a sequência (x_k) converge em um número finito de passos;

(ii) se $\theta \in (0, \frac{1}{2}]$, existem $b_0 > 0$ e $\varsigma \in [0, 1)$, tais que

$$d(x_k, x^*) \leq b_0 \varsigma^k;$$

(iii) se $\theta \in (\frac{1}{2}, 1)$, existe $\xi > 0$, tal que

$$d(x_k, x^*) \leq \xi k^{-\frac{1-\theta}{2\theta-1}}.$$

Demonstração: Usamos as notações do Lema 4.3 e assumimos que $f(x_k) \to 0$. Neste caso de acordo com a Proposição 4.1, tem-se $f(x^*) = 0$.
(i) Se a sequência $(f(x_k))$ é estacionária, então (x_k) também o é, conforme a Proposição 4.1. Entretanto, se $(f(x_k))$ não é estacionária, então da desigualdade de Kurdyka-Lojasiewicz, obtemos que $c \cdot F(\text{grad}\, f(x_k)) \geq 1$, para k suficientemente grande. Por outro lado, como $x_k \to x^*$, F e $\text{grad}\, f(x)$ são contínuas, segue-se que $c\, F(\text{grad}\, f(x_k)) \to 0$, o que é uma contradição. Portanto (x_k) converge em um número finito de passos.

Para completar a prova deste teorema consideramos $k \geq 0$ e

$$D(x_k) := \sum_{i=k+1}^{+\infty} d(x_i, x_{i+1}),$$

o qual é finito, conforme Teorema 4.1. Agora, para $k \in \mathbb{N}$ fixado, seja $j \in \mathbb{N}$, $j > k$. Logo, usando-se a desigualdade triangular,

$$d(x_k, x^*) \leq d(x_k, x_j) + d(x_j, x^*) \leq \sum_{i=k+1}^{j} d(x_i, x_{i+1}) + d(x_j, x^*),$$

fazendo $j \to +\infty$ na última desigualdade, temos

$$d(x_k, x^*) \leq \sum_{i=k+1}^{+\infty} d(x_i, x_{i+1}) = D(x_k).$$

Com vistas às demonstrações dos itens (ii) e (iii) é suficiente estimar $D(x_k)$. Reescrevendo-se (4.16), obtemos

$$D(x_k) \leq \varphi(f(x_k))E + D(x_{k-1}) - D(x_k), \tag{4.18}$$

em que $E = b/a$. Usando a desigualdade de Kurdyka-Lojasiewicz, segue-se que

$$1 \leq \varphi'(f(x_k))F(\operatorname{grad} f(x_k)) = c(1-\theta)f(x_k)^{-\theta}F(\operatorname{grad} f(x_k)),$$

logo,

$$(f(x_k))^{\theta} \leq c(1-\theta)F(\operatorname{grad} f(x_k)) \leq c(1-\theta)b(D(x_{k-1}) - D(x_k)),$$

portanto,

$$(f(x_k))^{1-\theta} \leq [c(1-\theta)b]^{\frac{1-\theta}{\theta}}(D(x_{k-1}) - D(x_k))^{\frac{1-\theta}{\theta}}.$$

Usando as duas últimas desigualdades e fazendo $\Theta = c[c(1-\theta)b]^{\frac{1-\theta}{\theta}}$, obtemos que

$$\varphi(f(x_k)) = c(f(x_k))^{1-\theta} \leq \Theta(D(x_{k-1}) - D(x_k))^{\frac{1-\theta}{\theta}}.$$

Portanto, de (4.18),

$$D(x_k) \leq E\Theta(D(x_{k-1}) - D(x_k))^{\frac{1-\theta}{\theta}} + (D(x_{k-1}) - D(x_k)). \tag{4.19}$$

(ii) De (4.19) e do fato de $\theta \in (0, 1/2]$, obtemos que existe uma constante c_1 tal que

$$D(x_k) \leq c_1(D(x_{k-1}) - D(x_k)),$$

para k suficientemente grande, em que $c_1 = E\Theta + 1$. Logo,

$$D(x_k) \leq \frac{c_1}{1+c_1}D(x_{k-1}).$$

Finalmente, usando recorrência em k, segue-se o resultado desejado,i.e.,

$$d(x_k, x^*) \leq D(x_k) \leq b_0 \varsigma^k,$$

em que $\varsigma = c_1/(1 + c_1)$ e $b_0 = D(x_{k_0})$.

(iii) Utilizamos roteiro semelhante abordado em [12, Theorem 2]. Primeiro, de (4.19) obtemos que existem inteiros $n_1 > n_0$ e uma constante $c_2 > 0$, tais que

$$D(x_k)^{\frac{\theta}{1-\theta}} \leq c_2(D(x_{k-1}) - D(x_k)),$$

para todo $k \geq n_1$, em que $c_2 = (E\Theta + 1)^{\frac{\theta}{1-\theta}}$. Segundo , usando argumentos técnicos de [12, Teorema 2] e o fato de $\theta \in (1/2, 1)$, obtemos uma constante $c_3 > 0$ tal que

$$0 < c_3 \leq D(x_k)^{\frac{1-2\theta}{1-\theta}} - D(x_{k-1})^{\frac{1-2\theta}{1-\theta}}.$$

Finalmente, para n suficientemente grande,

$$c_3(n - n_1) \leq D(x_n)^{\frac{1-2\theta}{1-\theta}} - D(x_{n_1})^{\frac{1-2\theta}{1-\theta}},$$

portanto

$$d(x_n, x^*) \leq D(x_n) \leq [c_3(n - n_1) + D(x_{n_1})^{\frac{1-2\theta}{1-\theta}}]^{\frac{1-\theta}{1-2\theta}} \leq \xi n^{-\frac{1-\theta}{2\theta-1}},$$

em que $\xi = (c_3 + 1)^{-\frac{1-\theta}{2\theta-1}} > 0$. $\square$

Capítulo 5

Aplicações

Na seção seguinte, descrevemos uma importante aplicação do método alternado. Detalhes podem ser encontrados no trabalho de (DA CRUZ NETO et al. [30]).

5.1 Aplicação do método alternado à teoria dos jogos

Na teoria dos jogos não cooperativos, um dos tópicos mais importantes é como os jogadores aprendem a jogar no equilíbrio de Nash, tema abordado em [31]. Dinâmicas de aprendizagem incluem a aprendizagem Bayesiana, jogador fictício, melhor resposta Cournot, a aprendizagem adaptativa, dinâmica evolutiva, aprendizado por reforço e outras. Os jogadores podem jogar em seqüência (como na melhor resposta de Cournot), movendo-se em alternância, um jogador em cada período. Neste caso eles seguem uma dinâmica regressiva se o jogador se desviar do que tem observado (conhecido) o que todos os outros jogadores fizeram no passado. A cada período o jogador escolhe uma melhor resposta, tomando-se como dadas, as ações jogadas um pouco antes pelo jogadores. Os jogadores podem jogar simultaneamente, cada período (como no jogo fictício, ver [32]). Primeiro, eles formam suas crenças sobre o que cada um dos outros jogadores irá fazer neste período. Essas crenças são geralmente a média da soma das ações desempenhadas no passado por cada um dos outros jogadores. Então eles dão uma melhor resposta a essas crenças. Neste caso eles seguem uma dinâmica progressiva. Crenças são atualizadas a cada período, como uma média das últimas ações de cada um dos outros jogadores e as suas crenças anteriores. Em ambos os casos, os jogos são dados na forma normal. Eles são definidos por suas funções de payoffs sobre suas estratégias (espaço de ações). Para dois jogadores com espaços de ações M e N suas funções payoffs são, respectivamente, $F(x, y) \in \mathbb{R}$ e $G(x, y) \in \mathbb{R}$, $x \in M$ e $y \in N$.

Neste contexto dinâmico do problema de (aprender a jogar Nash), quatro questões centrais são colocadas: i) como essas dinâmicas de aprendizagem convergem para o conjunto de equilíbrio de Nash (convergência positiva) ou convergem para um equilíbrio de Nash?; ii) o processo de convergência ocorre em tempo finito?; iii) qual é a velocidade de convergência, ela ocorre gradualmente ou abruptamente?; iv) como as restrições sobre os espaços de ações podem ser incluídas para cada jogador?

Variedades Riemannianas como espaços de ações. Sejam F e G dois jogadores, a forma normal usual do jogo é dada por $F(x,y) \in \mathbb{R}$ and $G(x,y) \in \mathbb{R}$, $x \in M$ e $y \in N$. Vamos mostrar através de exemplos como os conjuntos de ações possíveis M e N, associadas aos jogadores F e G, podem ser modelados como variedades Riemannianas, permitindo considerar o caso importante em que os jogadores fazem múltiplas atividades e os recursos e tempo são restrições materiais. Suponhamos que os jogadores F e G possam executar as listas de atividades $i \in I_F$ e $j \in I_G$. Sejam $l_F^i(x^i) \in \mathbb{R}_+$ e $l_G^j(y^j) \in \mathbb{R}_+$ os tempos ou recursos gastos para produzir as quantidades x^i e x^j, de insumos intermediários $i \in I_F$ e $j \in I_G$. Sejam $L_F \in \mathbb{R}_+$ e $L_G \in \mathbb{R}_+$ as quantidades de recursos disponíveis para os jogadores F e G a cada período. Então as limitações de recursos dos jogadores a cada período são $\Sigma_{i\in I_F} l_F^i(x^i) - L_F = 0$ e $\Sigma_{j\in I_G} l_G^j(y^j) - L_G = 0$.

Se $x = (x^i,\ i \in I_F) \in \mathbb{R}_+^{cardI_F} = X$ e $y = (y^j,\ \ j \in I_G) \in \mathbb{R}_+^{cardI_G} = Y$ são as quantidades de insumos intermediários produzidos por cada jogador (suas ações), as saídas (performances) produzidas por cada jogador são $\varphi_F(x) \in \Phi$ e $\varphi_G(y) \in \Phi$, onde Φ é o espaço de saída. Os rendimentos de cada jogador são $R_F\left[\varphi_F(x), \varphi_G(y)\right]$ e $R_G\left[\varphi_F(x), \varphi_G(y)\right]$, os custos para fazer essas ações são $K_F(x,y)$ e $K_G(x,y)$. Então, as funções explícitas de recompensa de cada jogador são

$$F(x,y) = R_F\left[\varphi_F(x), \varphi_G(y)\right] - K_F(x,y)$$

e

$$G(x,y) = R_G\left[\varphi_F(x), \varphi_G(y)\right] - K_G(x,y)$$

Os subconjuntos das possíveis ações dos dois jogadores são as variedades

$$M = \{x \in X : L_F(x) - L_F = 0\} \text{ e } N = \{y \in Y : L_G(y) - L_G = 0\},$$

onde $L_F(x) = \Sigma_{i\in I_F} l_F^i(x^i)$ e $L_G(y) = \Sigma_{j\in I_G} l_G^j(y^j)$. Cada jogador restringe a utilização de recursos de cada atividade.

Um exemplo relacionado à variedade de curvatura positiva é o seguinte:

$L_F(x) = \Sigma_{i \in I_F}(x^i)^2$ e $L_G(y) = \Sigma_{j \in I_G}(y^j)^2$, ou então,

$$L_F(x) = \sum_{i \in I_F} \left(\frac{x^i}{a_i}\right)^2 \quad \text{e} \quad L_G(y) = \sum_{j \in I_G} \left(\frac{y^j}{b_i}\right)^2.$$

Nestes casos as variedades são porções de esferas e elipsódes, respectivamente.

Outro exemplo relacionado à variedade de curvatura não positiva (caso das variedades de Hadamard) é o seguinte: consideramos um jogador fazendo o conjunto de atividades descritas por $x = ((x^i, i \in I), (x^j, j \in J))$. Todas atividades $i \in I$ produzem diariamente energia vital para o agente (como comer, descansar, curtir feriados, praticar esporte, realizar atividades saudáveis e de lazer, ...), as quais dão motivações adicionais para agir. Todas as atividades $j \in J$ consomem energia (trabalhar, pensar, ...). Uma restrição importante, quase sempre negligenciada na literatura econômica é que o agente pode conservar (regenerar) a sua energia com a evolução do tempo. Seja $e^i_+(x^i) \geq 0$ a energia produzida pela execução da ação $x^i \in \mathbb{R}$ e $e^j_-(x^j) \geq 0$ a energia consumida pela execução ação $x^j \in \mathbb{R}$. Então a regeneração de energia vital impõe a restrição $\Sigma_{i \in I} e^i_+(x^i) - \Sigma_{j \in J} e^j_-(x^j) = E > 0$. Funções de produção e consumo de energia podem ser quadráticas (quanto mais atividade de um agente, mais ele produz e consome energia, a uma taxa crescente). Neste caso, a expressão $\Sigma_{i \in I}(x^i)^2 - \Sigma_{j \in J}(x^j)^2 = E > 0$ define um hiperbolóide. Um exemplo mais real pode ser dado quando as funções de produção de energia são crescentes e côncavas, e as funções de consumo de energia são decrescentes e convexas.

Usando nosso principal resultado (Teorema 2.1) somos capazes de dar uma resposta às três primeiras questões sobre "como aprender a jogar no equilíbrio de Nash?"

Sejam M e N os espaços de ações, $x \in M$ e $y \in N$ as ações associadas a dois jogadores. As distâncias entre as ações x e x' do primeiro jogador e as ações y e y' do segundo jogador são dadas por $d_M(x, x')$ e $d_N(y, y')$. Por outro lado, a distância entre as ações de acoplamento $z = (x, y)$ e $z' = (x', y')$ é dada pela expressão $d(z, z') = \sqrt{d^2_M(x, x') + d^2_N(y, y')}$. A partir da desigualdade $a + b \leq 2\sqrt{a^2 + b^2}$, com $a \geq 0, b \geq 0$, segue-se que

$$d_M(x, x') + d_N(y, y') \leq 2\sqrt{d^2_M(x, x') + d^2_N(y, y')} = 2d(z, z').$$

Admitamos que as velocidades de movimento $\omega_F > 0$ e $\omega_G > 0$, de cada jogador, são constantes. Então, o tempo para cada jogador dar sua melhor resposta é proporcional à distância de movimentação, isto é,

$$(1/\omega_F)d_M(x_i, x_{i+1})\text{e}(1/\omega_G)d_N(y_i, y_{i+1}).$$

Para simplificar, tomamos a velocidade de movimentação $\omega_F = \omega_G = \omega > 0$. Assim, partindo de um ponto z_k, o tempo gasto para convergir é dado por

$$T(z_k) = (1/\omega)\Sigma_{i=k}^{+\infty}\left[d_M(x_i, x_{i+1}) + d_N(y_i, y_{i+1})\right]$$

e $T(z_k) \leq (2/\omega)D(z_k)$. Logo,

$$T(z_k) \leq (2/\omega)\left[E\varphi(h_k - \bar{h}) + \alpha^{-1}\sqrt{2t_2 h_{k-1} - \bar{h}}\right],$$

onde $h_k = H(z_k) = f(x_k) + \Psi(x_k, y_k) + g(y_k)$, $\bar{h} = H(\bar{z})$ e $E = 2t_2(\delta + 2\tau\beta {t_1}^{-1})\alpha^{-2}$, com $0 < t_1 \leq \lambda_k \leq t_2$ e $0 < t_1 \leq \mu_k \leq t_2$. A partir da última majoração decorrem várias conseqüências.

Primeira consequência: quanto maior for a velocidade de aprendizagem, ou seja, $\omega = \omega_F = \omega_G$., menor será o tempo de convergência

Segunda consequência: partimos do jogo na forma normalizada proximal,

$$x_{k+1} \in \text{argmin}\left\{F(p, y_k) + (1/2\lambda_k)C_M^2(x_k, p)\right\}$$

e

$$y_{k+1} \in \text{argmin}\left\{G(x_{k+1}, q) + (1/2\mu_k)C_N^2(y_k, q)\right\},$$

onde $2\lambda_k = \varphi_{k,F}/\delta_{k,F} > 0$ e $2\mu_k = \varphi_{k,G}/\delta_{k,G}$. Logo, quanto menor for o limite superior t_2, menores serão os coeficientes λ_k e μ_k. Ademais, menores comprimentos esperados de investigação $\varphi_{k,F}$ e $\varphi_{k,G}$ e maiores desutilidades de custos para mudar $\delta_{k,F}$ e $\delta_{k,G}$ implicam menor tempo de convergência. Destacamos que:
i) menores comprimentos esperados dos períodos de exploração $\varphi_{k,F}$ e $\varphi_{k,G}$ representam um processo de antecipação porque ao longo do caminho distâncias entre os estados convergem para zero. Então os comprimentos dos períodos de exploração (que também são os comprimentos dos períodos de investigação) convergem para zero. Assim os jogadores antecipam períodos de investigação e os tornam cada vez mais curtos. Isso faz com que nosso jogo seja convergente, não somente nos espaços das ações, mas também nos espaços das crenças (ver [32], para o jogo fictício com a convergência nos espaços de ações e crenças). Então o nosso jogo tem uma propriedade muito agradável em termos de estabilidade.
ii) maiores desutilidades dos custos para mudar $\delta_{k,F}$ e $\delta_{k,G}$, representam maiores

desutilidades para mudar

$$D_M(p, x_k) = (1/2\lambda_k)C_M^2(x_k, p) \text{ e } D_N(q, y_k) = (1/2\mu_k)C_N^2(y_k, q),$$

onde $C_M(x_k, p)$ e $C_N(y_k, q)$ são os custos para mudar. Consequentemente, maior inércia, isto é, maior resisitência para mudar.

Terceira consequência: o tempo de convergência é tanto menor:
i) quanto menor for a distância $\sqrt{d_M^2(x_k, \bar{x}) + d_N^2(y_k, \bar{y})}$ de um ponto inicial (x_k, y_k) ao ponto $(\bar{x}, \bar{y})$;
ii) quanto menor for a recompensa conjunta $h_k = H(x_k, y_k)$, ou seja, os ganhos de menor potencial podem ser obtidos movendo-se a partir do estado inicial para o equilíbrio de Nash;
iii) quanto menor for o coeficiente $E = 2t_2(\delta + 2\tau\beta {t_1}^{-1})\alpha^{-2}$, i.e., dado t_1, quanto menor é t_2 (assim, menor é λ_k, e menor é a constante de Lipschitz C do gradiente $grad\Psi$, i.e., menor é a recompensa conjunta $\Psi(x, y)$).

A interpretação do Teorema 2.4 é a seguinte: a desigualdade de Kurdyka-Lojasiewicz, com $\varphi(s) = cs^{1-\theta}$ e expoente $\theta \in [0, 1)$, $c > 0$, significa que a discrepância $|H(z) - H(z^*)|$ entre o valor presente de interesse de ganho comum $H(z)$ e o seu valor mínimo $H(z^*)$, a qual os jogadores querem preencher, é menor que alguma potência do subgradiente $H(z)$ no ponto z: $|H(z) - H(z^*)| \leq c\,|g|^{1/\theta}$ para todo $g \in \partial H(z)$ e todo $z \in B(z^*, \varepsilon)$,$\varepsilon > 0$. A razão $1/\theta \geq 1$ é tanto menor quanto maior for θ. Consideramos os dois primeiros casos (observações semelhantes podem ser dadas para o terceiro caso):
i) se $\theta = 0$, então, a convergência para um equilíbrio de Nash ocorre em um número finito de passos, um desejável resultado para teoria dos jogos;
ii) se $\theta \in (0, \frac{1}{2}]$ a razão $1/\theta$ é (≥ 2). O Teorema 2.4 nos diz que existem constantes $c_0 > 0$ e $\varsigma \in [0, 1)$ tais que $d(z_k, z^*) \leq D(z_0)\varsigma^k$, onde $D(z_0) = \Sigma_{i=1}^{+\infty} d(z_i, z_{i+1}) < +\infty$ representa o comprimento da trajetória que parte de z_k e converge para z^*, (ver Observação 5.1). Logo, partindo de qualquer ponto z_k, a distância de z_k ao seu ponto limite z^* é menor que uma fração ς^k do comprimento total $D(z_0)$ de sua trajetória. Então a distância restante $d(z_k, z^*)$ entre qualquer z_k da sequência e seu limite z^*, assim como o tempo restante para convergir (supondo-se a velocidade de mover constante ω) é tanto menor quanto menor for a constante $0 < \varsigma = c_1/(1 + c_1) < 1$, ou seja, quanto menor for a constante $c_1 = E\Theta + 1 > 0$, i.e., quanto menores forem as constantes E e Θ, onde $E = 2t_2(\delta + 2\tau\beta t_1^{-1})\alpha^{-2}$ e $\Theta = c\left[c(1 - \theta)(\delta + 2\tau\beta t_1^{-1})\right]^{(1-\theta)/\theta}$. Então, dados os valores t_1 e α, o tempo restante para convergir é tanto menor quanto menor for o coeficiente t_2, (o qual majora os pesos $0 < t_1 \leq \lambda_k, \mu_k \leq t_2$ dos custos para mudar), quanto menor for a constante de Lipschitz δ para variações do gradiente, quanto menor for a cota superior τ do

$gradC_M, gradC_N$, quanto menor for a constante de majoração do custo por unidade de distância β, $(0 < \alpha \leq c_M, c_N \leq \beta)$ e quanto menor for o raio de majoração c para discrepância $|H(z) - H(z^*)|$ que os jogadores querem preencher.

Observação 5.1 *Quando $\theta \in (0, \frac{1}{2}]$, a prova da parte ii) do Teorema 2.4 mostra que existe c_1 tal que $D(z_k) \leq c_1(D(z_{k-1}) - D(z_k))$ e para k suficientemente grande $D(z_k) \leq \varsigma D(z_{k-1})$, onde $\varsigma = c_1/(1 + c_1)$. Então, $d(z_k, z^*) \leq D(z_k) \leq c_0 \varsigma^k$, onde $c_0 = D(z_0) = \Sigma_{i=1}^{+\infty} d(z_i, z_{i+1}) < +\infty$. Precisões similares podem ser dadas no caso em que $\theta \in]1/2, 1[$.*

Na próxima seção, apresentamos uma aplicação do método do ponto proximal a um problema de teoria da decisão.

5.2 Aplicação do método do ponto proximal à Teoria da decisão

Nas ciências do comportamento, um problema de teoria da decisão consiste em examinar como um agente decide a qualidade de uma decisão, quão boa deverá sê-la, o tempo e o esforço perdido para encontrá-la. Este problema é mais conhecido na literatura como Effort-accuracy trad off. DA CRUZ NETO et al. [33] mostram como o algoritmo proximal numa variedade de Finsler pode modelar o problema "effort-accuracy trad off", usando a recente "abordagem variacional"das teorias da mudança, tratada em [34, 35].

Capítulo 6

Considerações Finais

Nos Capítulos 1 e 2 deste trabalho estendemos o algoritmo proximal alternado no contexto das variedades Riemanninas de Hadamard. Assumimos que a função objetivo, a ser minimizada, é KL e usamos quase-distâncias como regularização para obter a convergência da sequência gerada pelo algoritmo (2.1). Estudamos as relações intrínsecas entre estruturas o-minimal e categorias geométricas analíticas para assegurar a existência de funções KL no contexto Riemanniano. Obtivemos resultados teóricos de convergência (Teorema 2.1) e taxa de convegência (Teorema 2.4). Ressaltamos que o algoritmo alternado (2.1) é naturalmente aplicado na teoria dos jogos, especificamente, no modelo "exploration-exploitation", com interessantes interpretações, algumas delas abordadas no Capítulo 5, para detalhes (ver DA CRUZ NETO et al. [30]).

Nos Capítulos 3 e 4 deste trabalho estendemos o método do ponto proximal do contexto Riemanniano para o cenário Finsleriano. Assumimos que a função objetivo é KL e usamos distâncias Finslerianas como funções regularizadoras para obter a convergência da sequência gerada pelo algoritmo (4.2). Estudamos as relações intrínsecas entre estruturas o-minimal e categorias geométricas analíticas para assegurar a existência de funções KL no contexto Finsleriano. Obtivemos importantes resultados teóricos de convergência (Teorema 4.1) e taxa de convegência (Teorema 4.2). Destacamos que o algoritmo do ponto proximal (4.2) é naturalmente aplicado à teoria da decisão, quando as distâncias Finslerianas não são simétricas, (consultar DA CRUZ NETO et al. [33]).

Possibilidades de pesquisa futura:

1. Método gradiente no contexto Finsleriano

Inspirado em [16], propomos analizar o método gradiente, no contexto Finsleriano e utilizando a hipótese de que a função objetivo é KL em substituição às hipóteses de convexidade e não negatividade da curvatura. Objetivamos fazer análise de convergência da sequência gerada pelo algoritmo do método gradiente.

Sejam (M, F) uma variedade de Finsler e $f : M \to \mathbb{R}$ uma função C^1. Para resolver o problema:

$$\begin{aligned} &\min f(x) \\ &s.t. \;\; x \in M, \end{aligned} \tag{6.1}$$

consideramos o algoritmo:

Algoritmo 6.1 *(1) Dado $x_k \in M$, $k \geq 1$, calcular $p_k = -\operatorname{grad} f(x_k)$;*

(2) Determinar a geodésica $\gamma(t)$, $t \geq 0$ tal que $\gamma(0) = x_k$ e $\gamma'(0) = p_k$;

(3) Faça $x_{k+1} = \gamma(t_k)$, em que t_k é obtido usando um dos seguintes critérios:

- (Passo fixo) Dados δ_1, $\delta_2 > 0$ tais que $\delta_1\Gamma + \delta_2 < 1$, em que Γ é a constante de Lipschitz associada ao campo grad f*, escolhemos*

$$t_k \in (\delta_1, \frac{2}{\Gamma}(1 - \delta_2).$$

- (Busca de Armijo) Escolhemos $t_k = 2^{-i_k}\bar{t}$, em que $\bar{t} > 0$ é dado e i_k é o menor inteiro positivo tal que

$$f(\gamma(t_k)) \leq f(x_k) - \beta t_k^2 F^2(\operatorname{grad} f(x_k)),$$

com $\beta \in (0, 1)$.

DA CRUZ NETO et al [16] propuseram a seguinte definição no contexto Riemanniano, a qual estendemos ao contexto Finsleriano.

Definição 6.1 *Seja $f : M \to \mathbb{R}$ uma função C^1 e $L > 0$ um número real. Diz-se que f possui gradiente L-Lipschitziano, quando para todo $x, y \in M$ e para qualquer segmento geodésico $\gamma : [0, r] \to M$ ligando x e y, obtemos que*

$$F(\operatorname{grad} f(\gamma(t)) - P_{\gamma(a)\gamma(t)} \operatorname{grad} f(x)) \leq rl(t), \qquad t \in [0, r],$$

onde $l(t)$ denota o comprimento do segmento entre $\gamma(0) = x$ e $\gamma(t)$. Em particular, se M é uma variedade de Hadamard, então a última desigualdade reduz-se a

$$F(\operatorname{grad} f(\gamma(t)) - P_{x\gamma(t)} \operatorname{grad} f(x)) \leq rd(\gamma(t), x), \qquad t \in [0, r].$$

Vamos considerar a seguinte hipótese:

Hipótese 6.1 *Existe uma função $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ tal que:*

a) existem $\alpha \in (0, 1)$ e $\tau_\alpha > 0$, tais que $\phi(t) \leq \alpha t$, $\forall t \in (0, \tau_\alpha]$;

b) existem $\beta > 0$ e $\tau_\beta \in (0, +\infty]$, tais que $\phi(t) \geq \beta t^2$, $\forall t \in (0, \tau_\beta] \cap \mathbb{R}$;

c) para todo $k = 0, 1, \ldots$, $f(x_{k+1}) \leq f(x_k) - \phi(t_k)F^2(\operatorname{grad} f(x_k))$ e $0 < t_k \leq \tau_\beta$;

d) existem $\gamma > 1$, $\tau_\gamma > 0$, tais que $\forall k$, $t_k \geq \tau_\gamma$ ou

existe $\bar{t}_k \in [t_k, \gamma t_k] : \; f(\exp_{x_k}(-\bar{t}_k \operatorname{grad} f(x_k))) \geq f(x_k) - \phi(\bar{t}_k)F^2(\operatorname{grad} f(x_k))$.

Doravante, assumimos que f é uma função C^1 com gradiente L-Lipschitziano. Propomos provar, entre outros, o seguinte resultado.

Teorema 6.1 *Seja f uma função limitada inferiormente e KL. Então cada sequência limitada (x_k) gerado por 6.1 converge para algum ponto crítico $\bar{x}$ de f.*

2. Método subgradiente no contexto Finsleriano

Inspirado no trabalho de FERREIRA E OLIVEIRA [36], propomos analizar o método subgradiente no contexto Finsleriano, utilizando a hipótese de que a função objetivo é KL em substituição às hipóteses de convexidade e não positividade da curvatura.

Sejam (M, F) uma variedade de Finsler e $f : M \to \mathbb{R}$ semicontínua inferior. Para resolver o problema:

$$\begin{aligned} &\min f(x) \\ &s.t. \;\; x \in M, \end{aligned} \tag{6.2}$$

consideraremos o algoritmo:

1. Dado $x_0 \in M$,
2. defina $x_{k+1} = exp_{x_k}(-t_k \frac{s_k}{F(s_k)})$, onde $t_k > 0$, $s_k \in \partial f(x_k)$ e $s_k \neq 0$.

Objetivamos fazer análise de convergência da sequência gerada pelo algoritmo.

Referências Bibliográficas

[1] ATTOUCH, H., BOLTE, J., REDONT, P., et al., “Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE’s”, *J. Convex Anal*, v. 15, n. 3, pp. 485–506, 2008.

[2] ATTOUCH, H., BOLTE, J., REDONT, P., et al., “Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Lojasiewicz Inequality”, *Math. Oper. Res.*, v. 35, n. 2, pp. 438–457, 2010.

[3] ATTOUCH, H., REDONT, P., SOUBEYRAN, A., “A new class of alternating proximal minimization algorithms with costs to move”, *SIAM J. Optim.*, v. 18, n. 3, pp. 1061–1081, 2007.

[4] LEWIS, A. S., MALICK, J., “Alternating projection on manifolds”, *Math. Oper. Res.*, v. 33, n. 1, pp. 216–234, 2008.

[5] FERREIRA, O. P., OLIVEIRA, P. R., “Proximal point algorithm on Riemannian manifolds”, *Optimization*, v. 51, n. 2, pp. 257–270, 2002.

[6] FUKUSHIMA, M., MINE, H., “A generalized proximal point algorithm for certain nonconvex minimization problems”, *Int. J. Systems Sci.*, v. 12, n. 8, pp. 989–1000, 1981.

[7] 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.

[8] BENTO, G. C., FERREIRA, O. P., OLIVEIRA, P. R., “Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds”, *Nonlinear Analysis*, v. 73, pp. 564–572, 2010.

[9] KURDYKA, K., “On gradients of functions definable in o-minimal structures”, *Ann. Inst. Fourier*, v. 48, n. 3, pp. 769–783, 1998.

[10] LOJASIEWICZ, S., "Une propriété topologique des sous-ensembles analytiques réels, in Les Équations aux Dérivé es Partielles", *Éditions du centre National de la Recherche Scientifique. Paris*, pp. 87–89, 1963.

[11] BOLTE, J., DANIILIDIS, A., LEWIS, A., "The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems", *SIAM J. Optim.*, v. 17, n. 4, pp. 1205–1223, 2006.

[12] ATTOUCH, H., BOLTE, J., "On the convergence of the proximal algorithm for nonsmooth functions involving analytic features", *Math. Program. Ser. B*, v. 116, n. 1-2, pp. 5–16, 2009.

[13] BOLTE, J., DANIILIDIS, A., LEWIS, A., et al., "Clarke subgradients of stratifiable functions", *SIAM J. Optim.*, v. 18, n. 2, pp. 556–572, 2007.

[14] LAGEMAN, C., *Convergence of gradient-like dynamical systems and optimization algorithms*, Ph.D. Thesis, Julius-Maximilians-Universität, Würzburg, Bayerischen, Germany, 2007.

[15] KURDYKA, K., MOSTOWSKI, T., PARUSINSKI, A., "Proof of the gradient conjecture of R. Thom", *Annals of Mathematics*, v. 152, n. 3, pp. 763–792, 2000.

[16] DA CRUZ NETO, J. X., LIMA, L. L., OLIVEIRA, P. R., "Geodesic algorithm in Riemannian manifolds", *Balkan Journal of Geometry its Applications*, v. 3, n. 2, pp. 89–100, 1998.

[17] UDRISTE, C., *Convex functions and optimization methods on Riemannian manifolds*. *Mathematics and its Applications*, Kluwer Academic Publishers: Netherlands, 1994.

[18] ABSIL, P. A., MAHONY, R., SEPULCHRE, R., *Optimization Algorithms on Matrix Manifolds*. Princeton University Press: Princeton, New Jersey, 2008.

[19] KRISTÁLY, A., MOROSANU, G., RÓTH., "Optimal Placement of a Deposit between Markets: Riemann-Finsler Geometrical Approach", *J. Optim. Theory Appl.*, v. 139, n. 2, pp. 263–276, 2008.

[20] DO CARMO, M. P., *Riemannian geometry*. *Mathematics: Theory & Applications*, Birkhäuser Boston Inc.: Boston, MA, 1992, Translated from the second Portuguese edition by Francis Flaherty.

[21] LEDYAEV, Y. S., ZHU, Q. J., "Nonsmooth analysis on smooth manifolds", *Trans. Amer. Math. Soc.*, v. 359, n. 8, pp. 3687–3732 (electronic), 2007.

[22] HIRSCH, M. W., *Differential Topology*. Spring Verlag: New York, 1976.

[23] VAN DEN DRIES, L., MILLER, C., "Geometric categories and o-minimal structures", *Duke Mathematical Journal*, v. 84, n. 2, pp. 497–540, 1996.

[24] COSTE, M., *An introduction to O-minimal geometry*, Ph.D. Thesis, Univ. Pisa,, 2000.

[25] BIERSTONE, E., MILMAN, P. D., "Semianalytic and subanalytic sets", *Publicactions Mathematiques*, 1988.

[26] VAN DEN DRIES, L., "O-minimal structures and real analytic geometry", In Current developments in mathematics, 1998 (Cambridge, MA), pages 105-152. Int. Press, Somerville. MA.

[27] NASH, J., "The imbedding problem for Riemannian manifolds", *Annals of Mathematicsnal*, v. 63, n. 1, pp. 20–63, 1956.

[28] SHEN, Z., *Lectures on Finsler Geometry*. World Scientific Publishing Co. Pte. Ltd.: USA, 2001.

[29] BAO, D., CHERN, S., SHEN, Z., *An Introduction to Riemann-Finsler Geometry. Graduate Texts in Mathematics, vol. 200*, Springer: Berlim, 2000.

[30] DA CRUZ NETO, J. X., OLIVEIRA, P. R., SOARES JÚNIOR, P. A., et al., "Learning how to play Nash, potential games and Alternating minimization method for structured nonconvex problems on Riemannian manifolds", Disponível em: http://www.optimization-online.org/DB_HTML/2012/03/3405.html.

[31] CHEN, Y., GAZZALE, R., "When Does Learning in Games Generate Convergence to Nash Equilibria? The Role of Supermodularity in an Experimental Setting", *American Economic Review*, v. 9, n. 5, pp. 49–69, 2004.

[32] MONDERER, D., SHAPLEY, L., "Fictitious Play Property for Games with Identical Players", *Journal of Economic Theory*, v. 68, n. 14, pp. 258–265, 1996.

[33] DA CRUZ NETO, J. X., OLIVEIRA, P. R., SOARES JÚNIOR, P. A., et al., "Proximal point method on Finslerian manifolds and the Effort-Accuracy Trade off", Disponível em: http://www.optimization-online.org/DB_HTML/2011/08/3129.html.

[34] SOUBEYRAN, A., “Variational rationality, a theory of individual stability and change: worthwhile and ambidextry behaviors”, *Pre-print. GREQAM, Aix Marseillle University*, 2009.

[35] SOUBEYRAN, A., “Variational rationality and the ”unsatisfied man”: routines and the course pursuit between aspirations, capabilities and beliefs”, *Pre-print. GREQAM, Aix Marseillle University*, 2010.

[36] FERREIRA, O. P., OLIVEIRA, P. R., “Subgradient algorithm on Riemannian manifold”, *Journal of Optimization Theory and Applications*, v. 97, n. 1, pp. 93–104, 1998.

[37] BOLTE, J., DANIILIDIS, A., LEWIS, A., “A nonsmooth Morse-Sard theorem for subanalytic functions”, *J. Math. Anal. Appl.*, v. 321, n. 2, pp. 729–740, 2006.