 ISSN 1984-820X

Um Algoritmo Proximal Associado a uma Quase-métrica

Felipe A. G. Moreno*, **Paulo R. Oliveira**,
Programa de Engenharia de Sistemas e Computação, COPPE, UFRJ,
21941-972, Rio de Janeiro, RJ
E-mail: fgarcia@cos.ufrj.br, poliveir@cos.ufrj.br.

Resumo: *Neste trabalho estudamos a resolução do problema de minimização irrestrita, com a função objetivo sendo não convexa, não diferenciável e com caracteríticas analíticas. Desenvolvemos para este problema um algoritmo de ponto proximal, em que o termo proximal que consideramos é uma função que provém de uma quase-métrica. Se a sequência gerada pelo nosso algoritmo é limitada, então establecemos a sua convergência para um ponto crítico da função.*
Palavras-chaves: *Métodos proximais, quase-métrica, subdiferenciais Fréchet e limite.*

1 Introdução

Neste trabalho consideramos o seguinte problema de otimização:

$$\min\{f(x) : x \in \mathbb{R}^n\}, \tag{1}$$

onde $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é uma função própria e semicontínua inferior.

O algoritmo de ponto proximal foi introduzido como um método de aproximação-regularização para a resolução de problemas de otimização convexa e desigualdades variacionais associadas a operadores monótonos maximais (veja [10, 12]). Assim daremos uma breve descrição deste método. Considerando (1), o método de ponto proximal gera uma sequência $\{x^k\} \subset \mathbb{R}^n$ correspondente à recursão

$$x^{k+1} \in \arg\min\left\{f(u) + \frac{1}{\lambda_k}\|u - x^k\|^2 \ : \ u \in \mathbb{R}^n\right\}, \tag{2}$$

onde x^0 é um ponto dado, $\{\lambda_k\}_{k\in\mathbb{N}}$ é uma seqüencia positiva e $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$ denota a norma usual.

Nas últimas décadas foram desenvolvidas extensões do esquema (2) utilizando termos de regularização mais gerais, por exemplo, trocando a função quadrática de (2) por uma distância de Bregman, $D_h(u, x^k)$, induzida por uma função h (veja [6]). Observemos que D_h não verifica a definição de distância, porém preserva as propriedades de convexidade, continuidade e coercividade como no caso quadrático.

Por outro lado, o problema (1) já tem sido tema de estudo para uma classe de funções não-lineares que verificam a desigualdade de Łojasiewicz (veja [1, 3, 2]). Em [2], os autores usaram a desigualdade de Łojasiewicz como ferramenta para a obtenção da convergência do método proximal (2) quando f não é necessariamente convexa e diferenciável.

Lembramos que no contexto topologico uma possível generalização da métrica usual é uma aplicação quase-métrica, já que na definição desta não é requerida a verificação da simetria (veja [7, 5, 15]). Algumas das aplicações das quase-métricas se encontram na teoria da complexidade (veja [4, 8]), na economia, por exemplo nas funções de utilidade e escolhas dos consumidores (veja [14]).

*Partially supported by CAPES/CNPq - IEL Nacional - Brasil

A essência deste trabalho é apresentar um método proximal generalizado para minimizar uma função não convexa, não diferenciável e que verifica a desigualdade de Łojasiewicz. Este método gera uma sequência $\{x^k\} \subset \mathbb{R}^n$ por meio da recursão:

$$x^{k+1} \in \arg\min \left\{ f(u) + \frac{1}{\lambda_k} q^2(u, x^k) \ : \ u \in \mathbb{R}^n \right\}, \tag{3}$$

onde $q(\cdot,\cdot)$ denota uma quase-métrica.

Observemos também que em geral uma aplicação quase-métrica não necessariamente é convexa, diferenciável e coerciva; assim a análise da convergência do nosso método (3) não é semelhante à dos já existentes na literatura para problemas convexos. Estabelecemos a convergência da sequência gerada por (3) para um ponto crítico generalizado, usando a idéia apresentada em [2]. Portanto, nosso resultado é uma extensão dos resultados obtidos em [2].

Nosso trabalho é organizado como segue: Na seção 2 lembramos propriedades das funções quase-métricas, da teoria subdiferencial e da desigualdade de Łojasiewicz. Na seção 3 é apresentado nosso algoritmo de ponto proximal com quase-métrica. A análise de convergência é estabelecida na seção 4 e na seção 5 as nossas conclusões.

2 Preliminares

2.1 Quase-métrica

Definição 2.1 *Seja X um conjunto. A aplicação $q : X \times X \to \mathbb{R}_+$ é dita quase-métrica se*

1. $q(x,y) = q(y,x) = 0 \Leftrightarrow x = y \quad \forall\, x, y \in X.$,
2. $q(x,z) \leq q(x,y) + q(y,z) \quad \forall\, x, y, z \in X.$

O par (X, q) é chamado espaço quase-métrico.

Observemos que, se q goza da propriedade de simetria, isto é, $\forall\, x, y \in X$, $q(x,y) := q(y,x)$, então q é uma métrica. Mais ainda, dada uma quase-métrica q temos:

- $\overline{q}(x,y) := q(y,x)$, denota a quase-métrica conjugada de q;
- $\hat{q}(x,y) = \max\{q(x,y), \overline{q}(x,y)\}$, denota a métrica associada a q.

Exemplo 2.1.1 *[5] Seja $q : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+$ definida por*

$$q(x,y) = \begin{cases} 0 & \textit{if } x \leq y, \\ 1 & \textit{caso contrário} \end{cases}$$

então $(\mathbb{R}, q)$ é um espaço quase-métrico.

Exemplo 2.1.2 *[8] Seja $g : X \to \mathbb{R}_+$ uma função injetiva. Então para x, $y \in X$ definamos $q_g(x,y) = \max\{g(x) - g(y), 0\}$, assim (X, q_g) é um espaço quase-métrico.*

O resultado seguinte é obtido diretamente da definição.

Proposição 2.1 *Sejam $q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ uma quase-métrica e $z \in \mathbb{R}^n$. Então $q(\,\cdot\,, z)$ é Lipschitz e $q^2(\,\cdot\,, z)$ é localmente Lipschitz.* ■

2.2 Teoria Subdiferencial

Seja $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. O dominio de h é o conjunto $\mathrm{dom}(h) = \{x \in \mathbb{R}^n : h(x) < +\infty\}$. Assim h é própria se $\mathrm{dom}(h) \neq \emptyset$, e h é coerciva se $\lim_{\|x\| \to \infty} h(x) = +\infty$.

Lema 2.1 *Dados $z \in \mathbb{R}^n$ e $\lambda > 0$. Se $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é limitada inferiormente e $q^2(\,\cdot\,, z)$ é uma função coerciva, então a função $f + \lambda q^2(\,\cdot\,, z) : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é coerciva.* ■

O próximo resultado é bem conhecido na literatura (veja por exemplo [13, Teorema 1.9]).

Teorema 2.1 *Suponhamos que $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ é uma função semicontínua inferior, coerciva e própria. Então o valor $\inf h$ é finito e o conjunto $\arg\min h$ é não vazio e compacto.* ■

Definição 2.2 *Seja $h : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferior e própria.*

1. *Para cada $x \in \mathrm{dom}(h)$, o subdiferencial Fréchet de h em x, denotado por $\widehat{\partial} h(x)$, é o conjunto de vetores $x^* \in \mathbb{R}^n$ tais que*

$$\liminf_{\substack{y \neq x \\ y \to x}} \frac{h(y) - h(x) - \langle x^*, y - x\rangle}{\|x - y\|} \geq 0.$$

Se $x \neq \mathrm{dom}(h)$, então $\widehat{\partial} h(x) = \emptyset$.

2. *O subdiferencial limite de h em $x \in \mathbb{R}^n$, denotado por ∂h, é definido como segue*

$$\partial h(x) := \left\{x^* \in \mathbb{R}^n : \exists\, x_n \to x,\ h(x_n) \to h(x),\ x_n^* \in \widehat{\partial} h(x_n) \to x^*\right\}$$

3. *O ponto $\overline{x}$ é dito ponto crítico-limite se $0 \in \partial h(\overline{x})$. Denotamos por $\mathrm{crit}(h)$ o conjunto dos pontos críticos de h.*

Os próximos resultados serão usados para analisar as operações envolvendo o subdiferencial limite e o subdiferencial Fréchet (veja por exemplo [13], [11, Teorema 7.1, Proposição 2.7]).

Teorema 2.2 *Se $F = f_1 + f_2$ com f_1 localmente Lipschitz em $\overline{x}$ enquanto f_2 é semicontínua inferior e própria com $f_2(\overline{x})$ finito, então $\partial F(\overline{x}) \subset \partial f_1(\overline{x}) + \partial f_2(\overline{x})$.* ■

Teorema 2.3 *Sejam $\varphi_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2$, localmente Lipschitz em $\overline{x}$. Então*

$$\partial(\varphi_1.\varphi_2)(\overline{x}) = \partial\left(\varphi_2(\overline{x})\varphi_1 + \varphi_1(\overline{x})\varphi_2\right)(\overline{x}) \subset \partial\left(\varphi_2(\overline{x})\varphi_1\right)(\overline{x}) + \partial\left(\varphi_1(\overline{x})\varphi_2\right)(\overline{x}). \tag{4}$$

Além disso, se $\varphi_i \geq 0$, $i = 1, 2$, o lado direito de (4) é igual a $\varphi_2(\overline{x})\,\partial(\varphi_1)(\overline{x}) + \varphi_1(\overline{x})\,\partial(\varphi_2)(\overline{x})$. ■

Teorema 2.4 *Sejam $\varphi_i : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $i = 1, 2$, funções semicontínuas inferiores com alguma sendo localmente Lipschitz em $\overline{x} \in \mathrm{dom}(\varphi_1) \cap \mathrm{dom}(\varphi_2)$. Então para qualquer $\delta > 0$, e $\gamma > 0$ temos*

$$\widehat{\partial}\left(\varphi_1 + \varphi_2\right)(\overline{x}) \subset \bigcup \left\{\widehat{\partial}\varphi_1(x_1) + \widehat{\partial}\varphi_2(x_2)\ :\ x_i \in B_\delta(\overline{x}), |\varphi_i(x_i) - \varphi_i(\overline{x})| \leq \delta,\ i = 1, 2\right\} + \gamma B_1(0),$$

onde $B_\delta(\bar{x})$ denota a bola fechada centrada em $\bar{x}$ e raio δ. ■

2.3 Desigualdade de Łojasiewicz

Nesta seção lembramos a generalização da desigualdade de Łojasiewicz no contexto do subdiferencial limite usada em [1, 2]. Seja $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ uma função semicontínua inferior tal que a restrição da f para seu domínio é uma função contínua.

Dizemos que f satisfaz a **desigualdade de Łojasiewicz** se: para qualquer $\widehat{x} \in \mathrm{crit}(f)$, existem C, $\epsilon > 0$ e $\theta \in [0, 1)$ tais que

$$|f(x) - f(\widehat{x})|^\theta \leq C\|x^*\|, \quad \forall\, x \in B_\epsilon(\widehat{x}), \quad \forall\, x^* \in \partial f(x). \tag{5}$$

Quando $\theta = 0$ adotamos a convenção $0^0 = 0$, e portanto se $|f(x) - f(\widehat{x})|^0 = 0$ temos $f(x) = f(\widehat{x})$. Para a prova do próximo resultado ver [2, Lema 1].

Lema 2.2 *Suponhamos que f satisfaz a desigualdade de Łojasiewicz.*

1. *Se K é um subconjunto conexo de $\text{crit}(f)$ então f é constante sobre K.*
2. *Além disto, se K é compacto, então existem $C,\, \epsilon > 0$ e $\theta \in [0, 1)$ tais que*

$$\forall\, x \in \mathbb{R}^n,\; d(x, K) \leq \epsilon,\; \forall\, x^* \in \partial f(x), \;\; |f(x) - f(\widehat{x})|^\theta \leq C\|x^*\|. \tag{6}$$

■

Exemplo 2.3.1 *[9]As funções seguintes satisfazem a desigualdade de Łojasiewicz:*

1. *Funções Reais-analíticas;*
2. *Funções Semi-algébricas, isto é, funções cujos gráficos podem ser expressos como*

$$\bigcup_{i=1}^{p} \bigcap_{j=1}^{q} \left\{ x \in \mathbb{R}^n \,:\, P_{ij}(x) = 0,\, Q_{ij}(x) > 0 \right\},$$

onde $\forall\; 1 \leq i \leq p$, $\forall\; 1 \leq j \leq q$ as P_{ij}, $Q_{ij} : \mathbb{R}^n \to \mathbb{R}$ são funções polinomiais.

Exemplo 2.3.2 *[2]As funções convexas que verificam a seguinte condição de crescimento: $\forall\, \widehat{x} \in \arg\min(f)$, $\exists\, C > 0$, $r \geq 1$, $\epsilon > 0$ tais que*

$$f(x) \geq f(\widehat{x}) + C d(x, \arg\min(f))^r \qquad \forall\, x \in B_\epsilon(\widehat{x}),$$

satisfazem a desigualdade de Łojasiewicz, em particular as funções fortemente convexas.

Para mais informações acerca da propriedade de Łojasiewicz e caracterizações ver [9, 1, 3].

3 O algoritmo de ponto proximal

A hipótese seguinte será assumida em todo nosso estudo:

$$(\mathcal{H}_1) \quad -\infty < \inf_{x \in \mathbb{R}^n} f(x).$$

O algoritmo de ponto proximal com quase-métrica é definido como:

1. Considere $x^0 \in \text{dom}(f)$.

2. Dado x^k, encontrar x^{k+1} tal que

$$x^{k+1} \in \arg\min \left\{ f(u) + \frac{1}{\lambda_k} q^2(u, x^k) \;:\; u \in \mathbb{R}^n \right\}, \tag{7}$$

3. Se $x^{k+1} = x^k$, **STOP**.

onde $\{\lambda_k\}_{k \in \mathbb{N}}$ é uma sequência positiva tal que $0 < \lambda_- < \lambda_k < \lambda_+$.

A próxima hipótese será necessária para garantir a existência dos iterandos:

$$(\mathcal{H}_2) \quad \text{Dado } z \in \mathbb{R}^n \text{ a função } q^2(\,\cdot\,, z) \text{ é coerciva.}$$

Da Proposição 2.1, temos que para qualquer $z \in \mathbb{R}^n$, $q^2(\,\cdot\,, z)$ é localmente Lipschitz. Assim, usando [13, Teorema 10.1] e o Teorema 2.2, para $f_1 := \frac{1}{\lambda} q^2(\,\cdot\,, u)$ com $u \in \mathbb{R}^n$, $\lambda > 0$ e $f_2 := f$, obtemos

$$0 \in \partial \left(f(\cdot) + \frac{1}{\lambda_k} q^2(\,\cdot\,, x^k) \right) (x^{k+1}) \subset \partial f(x^{k+1}) + \frac{1}{\lambda_k} \partial \left(q^2(\cdot, x^k) \right) (x^{k+1}).$$

Portanto existem $\xi^{k+1} \in \partial f(x^{k+1})$ e $\zeta^{k+1} \in \partial \left(q^2(\,\cdot\,, x^k)\right)(x^{k+1})$ tais que

$$0 = \xi^{k+1} + \frac{1}{\lambda_k}\zeta^{k+1}. \tag{8}$$

Considerando a seguinte hipótese:

$(\mathcal{H}_3)$ A restrição de f no seu domínio é uma função contínua;

obtemos os seguintes resultados, análogos a [2].

Proposição 3.1 *Seja $\{x^k\}_{k\in\mathbb{N}}$ uma sequência gerada por (7). Então*

1. *A sequência $\{f(x^k)\}_{k\in\mathbb{N}}$ é não crescente,*
2. $\sum\limits_{k=0}^{+\infty} q^2(x^{k+1}, x^k) < +\infty.$ ∎

Proposição 3.2 *Seja $\{x^k\}_{k\in\mathbb{N}}$ uma sequência gerada por (7) e denotando por $\omega(x^0)$ o conjunto dos pontos de acumulação da $\{x^k\}_{k\in\mathbb{N}}$. Se f satisfaz $(\mathcal{H}_3)$ então*

1. *f é finito e constante sobre $\omega(x^0)$,*
2. *Se $\{x^k\}_{k\in\mathbb{N}}$ é limitada então $\omega(x^0)$ é um conjunto não vazio, compacto e conexo, e* $\lim\limits_{k\to+\infty} d\left(x^k, \omega(x^0)\right) = 0.$ ∎

Observação 3.1 *Ainda quando $\{x^k\}_{k\in\mathbb{N}}$ é limitada, a convergência da $\{x^k\}_{k\in\mathbb{N}}$ pode falhar para uma função finita e suave f, ver [1].*

4 Resultado de Convergência

O próximo resultado é uma consequência de [13, Teorema 9.13, Proposição 5.15].

Lema 4.1 *Seja $\{x^k\}_{k\in\mathbb{N}}$ uma sequência gerada pelo algoritmo de ponto proximal. Se $\{x^k\}_{k\in\mathbb{N}}$ é limitada, então o conjunto $\partial\left(q(\,\cdot\,, x^{k-1})\right)(x^k)$ é limitado para todo k.* ∎

Finalmente, apresentamos o principal resultado deste trabalho.

Teorema 4.1 *Supondo que f satisfaz $(\mathcal{H}_1)$-$(\mathcal{H}_3)$ e a desigualdade de Łojasiewicz (6) e seja $\{x^k\}_{k\in\mathbb{N}}$ uma sequência gerada pelo algoritmo de ponto proximal. Se $\{x^k\}_{k\in\mathbb{N}}$ é limitada, então*

$$\sum_{k=0}^{+\infty} q(x^{k+1}, x^k) < +\infty,$$

em particular $\{x^k\}_{k\in\mathbb{N}}$ converge para algum ponto crítico-limite de f.

Esboço da prova: Usando a Proposição 3.1, Lema 2.2, [13, Teorema 10.1] e o Lema 4.1, obtemos

$$\lim_{k\to\infty} q(x^{k1}, x^k) = 0.$$

Assim, existem as sequências $\{x_1^{k_j}\}_{j\in\mathbb{N}}$, $\{f(x_1^{k_j})\}_{j\in\mathbb{N}}$ e $\{\xi^{k_j}\}_{j\in\mathbb{N}}$ com $\xi^{k_j} \in \widehat{\partial} f(x_1^{k_j})$ tais que

$$\lim_{j\to+\infty} x_1^{k_j} = \overline{x}, \quad \lim_{j\to+\infty} f(x_1^{k_j}) = f(\overline{x}) \quad \text{e} \quad \lim_{j\to+\infty} \xi^{k_j} = 0.$$

Portanto $0 \in \partial f(\overline{x})$. ∎

5 Conclusão

Neste trabalho apresentamos um método proximal, com o termo proximal sendo uma quase-métrica, para funções não convexas e não diferenciáveis que verificam a desigualdade de Łojasiewicz. Mostramos que a sequência gerada por (7) converge para algum ponto crítico-limite, sem fazer uso de argumentos baseados na convexidade e diferenciabilidade da função regularizadora, como em [12].

Referências

[1] P. A. Absil, R. Mahony e B. Andrews, Convergence of the iterates of descent methods for analytic cost functions, *SIAM J. Optim.*, 16 (2005), 531-547.

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

[3] J. Bolte, A. Daniilidis, O. Ley e L. Mazet, Characterizations of Łojasiewicz inequalities and applications, *Preprint* (2009), 1-47.

[4] V. Brattka, Recursive quasi-metric spaces, *Theoretical Computer Science*, 305 (2003), 17-42.

[5] A. Di Concilio e G. Gerla, Quasi-metric spaces and point-free geometry, *Math. Struct. in Comp. Science*, 16 (2006), 115-137.

[6] K. C. Kiwiel, Proximal minimization methods with generalized Bregman functions, *SIAM J. Control Optimization*, 35 (1997), 1142-1168.

[7] H.-P. A. Künzi, Nonsymmetric Topology, *Bolyai Society in Mathematical Studies 4, Topology with Applications*, (1993), 303-338.

[8] H.-P. A. Künzi, H. Pajoohesh e M. P. Schellekens, Partial quasi-metrics, *Theoretical Computer Science*, 365 (2006), 237-246.

[9] S. Łojasiewicz, Ensembles Semi-Analytiques, *Institut des Hautes Etudes Scientifiques*, Bures-sur-Yvette (Seine-et-Oise) France, (1965).

[10] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Française Informat. Recherche Opérationnelle*, 4 (1970), 154-158.

[11] B. S. Mordukhovich e Y. Shao, Nonsmooth Sequential Analysis in Asplund Spaces, *Transactions of the American Mathematical Society*, 348 (1996), 1235-1280.

[12] R. T. Rockafellar, Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14 (1976), 877-898.

[13] R. T. Rockafellar e R. J-B. Wets, "Variational Analysis", Springer, Berlin, 1998.

[14] S. Romaguera e M. Sanchis, Applications of utility functions defined on quasi-metric spaces, *J. Math. Anal. Appl.*, 283 (2003), 219-235.

[15] A. Stojmirović, Quasi-metric Spaces with Measure, *Topology Proceedings*, 28 (2004), 655-671.