O Método de Ponto Proximal para Otimização Convexa

José Henrique Salazar do **Amaral**[1] , Glaydston de Carvalho **Bento**[2]

Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus II- Caixa Postal 131, CEP 74001-970 - Goiânia, GO, Brasil.

E-mail: jusehenrique@hotmail.com[1]; glaydston@mat.ufg.br[2]

Palavras chaves: convergência,método do ponto proximal, otimização convexa.

1 Introdução

Considere o seguinte problema de otimização

$$\begin{aligned} &\min f(x) \\ &\text{t.q. } \; x \in \mathbb{R}^n, \end{aligned} \tag{1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ (função objetivo) é não necessariamente diferenciável. O método de ponto proximal, introduzido por Martinet [1] e Rockafellar [2] na década de 70, é um método iterativo utilizado para resolver o problema (1). O método de ponto proximal gera uma sequência $\{x^k\} \subset \mathbb{R}^n$ como segue: dado $x_0 \in \mathbb{R}^n$,

$$x^{k+1} := \operatorname{argmin}_{x\in \ \mathbb{R}^n} \left\{ f(x) + \lambda_k ||x - x^k||^2 \right\}, \tag{2}$$

onde $\{\lambda_k\}$ é uma sequência limitada de números positivos. Sendo $f_k := f(x) + \lambda_k||x - x^k||^2$ uma regularização de f, espera-se que f_k forneça algum tipo de aproximação do minimizador (caso exista) da função objetivo f.

2 Material e Métodos

Os materiais utilizados neste trabalho são basicamente livros e artigos. A metodologia utilizada foi discussões semanais com o orienador sobre os artigos e resultados relacionados ao tema de pesquisa.

[1]Orientando Mestrado IME

[2]Orientador

Revisado pelo orientador.

3 Resultados e Discussão

Nesta seção apresentamos algumas definições e resultados básicos de análise convexa alguns dos quais podem ser encontrados, por exemplo, em [3, 4]. Além disso, apresentamos uma análise de convergência do método de ponto proximal para resolver problemas de minimização convexa, ver [5, 6]. Daqui em diante, f sempre designará uma função definida em $\mathbb{R}^n$ e assumindo valores em $\mathbb{R}$, salvo menção explícita em contrário.

Definição 1. *Um conjunto $D \subset \mathbb{R}^n$ é conjunto convexo se, para quaisquer $x, y \in D$ e $\alpha \in [0,1]$, tem-se $\alpha x + (1-\alpha)y \in D$.*

Definição 2. *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \to \mathbb{R}$ é dita ser convexa em D quando, para quaisquer $x, y \in D$ e $\alpha \in [0,1]$, tem-se*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y).$$

Definição 3. *Seja f uma função convexa e ξ um vetor em $\mathbb{R}^n$. Dizemos que ξ é um subgradiente de f em x se*

$$f(y) \geq f(x) + \langle \xi, y - x \rangle, \qquad y \in \mathbb{R}^n.$$

O conjunto dos subgradientes de f em x é chamado subdiferencial de f em x, e é denotado por

$$\partial f(x) = \{\xi \in \mathbb{R}^n : \xi \text{ é subgradiente de } f \text{ em } x\}.$$

Teorema 1. *Seja f uma função convexa e $x \in \mathbb{R}^n$. O conjunto $\partial f(x)$ é não vazio.*

Demonstração. Ver, por exemplo, [4, Teorema 2.1.5, pág. 14]. □

Proposição 2. *Sejam f e g funções convexas de $\mathbb{R}^n$ em $\mathbb{R}$. Então, $\partial(f+g)(x) = \partial f(x) + \partial g(x)$, para todo $x \in \mathbb{R}^n$.*

Demonstração. Ver, por exemplo, [3, Proposição 3.4.8, pág. 172]. □

Proposição 3. *Seja f uma função convexa e $x \in \mathbb{R}^n$. Então, $x \in \mathbb{R}^n$ é minimizador de f se, e somente se, $0 \in \partial f(x)$.*

Demonstração. A prova segue imediatamente da definição de minimizador irrestrito combinada com a definição de subdiferencial de f em x. □

Definição 4. *Seja f uma função e $\lambda > 0$. O operador $P_\lambda : \mathbb{R}^n \to \mathbb{R}^n$, dada por*

$$P_\lambda(w) = argmin_{x\in\mathbb{R}^n}\{f(x) + \frac{\lambda}{2}\|x - w\|^2\},$$

é chamada aplicação proximal.

Proposição 4. *O operador aplicação proximal P_λ está bem definida. Além disso,*

$$\|P_\lambda(w_1) - P_\lambda(w_2)\|^2 \leq \langle w_1 - w_2, P_\lambda(w_1) - P_\lambda(w_2)\rangle, \qquad w_1, w_2 \in \mathbb{R}^n. \tag{3}$$

Demonstração. Para a prova da primeira parte, veja [6, Pág. 28]. Para provar a segunda parte, note que, pela definição do operador P_λ e da Proposição 3,

$$0 \in \partial(f(.) + \frac{\lambda}{2}\|. - w\|^2)(P_\lambda(w)), \qquad w \in \mathbb{R}^n.$$

Usando Proposição 2 com $g = \frac{\lambda}{2}\|. - w\|^2$, da última inclusão, obtemos

$$\lambda(w - P_\lambda(w)) \in \partial f(P_\lambda(w)).$$

Assim, dados $w_1, w_2 \in \mathbb{R}^n$, temos que

$$\lambda(w_1 - P_\lambda(w_1)) \in \partial f(P_\lambda(w_1)) \ \ e \ \ \lambda(w_2 - P_\lambda(w_2)) \in \partial f(P_\lambda(w_2)),$$

de onde obtemos a seguinte desigualdade

$$0 \leq \langle \lambda(w_1 - P_\lambda(w_1)) - \lambda(w_2 - P_\lambda(w_2)), P_\lambda(w_1) - P_\lambda(w_2)\rangle.$$

Levando em conta que $\lambda > 0$, a segunda parte da proposição segue de simples manipulações algébricas efetuadas na última desigualdade. □

Observação 1. *Em termos do operador aplicação proximal, a iterada de ordem $k+1$ do método de ponto proximal* (2) *pode ser reescrita como segue*

$$x^{k+1} = P_{\lambda_k}(x^k) := P_k(x^k). \tag{4}$$

Portanto, a boa definição da sequência gerada pelo método de ponto proximal (2) *segue da Proposição 4.*

De agora em diante o conjunto X^* denotará o conjunto solução do problema (1).

Lema 5. *Suponha que X^* é não vazio. Então, $w \in X^*$ se, e somente se, $P_\lambda(w) = w$.*

Demonstração. A prova segue imediatamente da definição de P_λ combinada com a Proposição 4. □

Observação 2. *Note que a desigualdade do item ii) da Proposição 3 é equivalente a*

$$\|P_\lambda(w_1) - P_\lambda(w_2)\|^2 \leq \|w_1 - w_2\|^2 - \|(I - P_\lambda)(w_1) - (I - P_\lambda)(w_2)\|^2.$$

Definição 5. *Dizemos que uma sequência $\{y^k\} \subset \mathbb{R}^n$ é Fejér convergente a um conjunto $U \subset \mathbb{R}^n$, com respeito a norma Euclidiana, se*

$$\|y^{k+1} - u\| \leq \|y^k - u\|, \qquad u \in U.$$

Proposição 6. *Se $\{y^k\}$ é Fejér convergente a um conjunto $U \neq \emptyset$, então $\{y^k\}$ é limitada. Se um ponto de acumulação y de $\{y^k\}$ pertence a U, então $y = \lim_{k\to\infty} y^k$.*

Demonstração. Ver, por exemplo, [5, Proposition 2.1, pág. 4]. □

Teorema 7. *Supondo $X^* \neq \emptyset$ temos que a sequência $\{x^k\}$ gerada por 2 converge a um ponto $x^* \in X^*$.*

Demonstração. Da segunda parte da Proposição 4 junto com a Observação 2, temos que

$$\|P_k(w_1) - P_k(w_2)\|^2 \leq \|w_1 - w_2\|^2 - \|(I - P_k)(w_1) - (I - P_k)(w_2)\|^2.$$

Fazendo $w_2 = x^*$, $w_1 = x^k$, $P_k(x^k) = x^{k+1}$ e usando que $P_k(x^*) = x^*$ (ver Lema 5), obtemos que

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - x^*\|^2 - \|x^k - x^{k+1}\|^2. \tag{5}$$

De (5) segue que $\{\|x^k - x^*\|\}$ é uma sequência decrescente e não negativa. Portanto, $\{\|x^k - x^*\|\}$ é uma sequência convergente e, novamente da desigualdade (5), podemos concluir que $\lim_{k\to\infty}(x^{k+1} - x^k) = 0$. Além disso, de (5), podemos concluir que a sequência $\{x^k\}$ é Fejér convergente a X^*. Assim, pela Proposição 6, tal sequência é limitada. Seja x^* um ponto de acumulação de $\{x^k\}$ e $\{x^{k_j}\}$ uma subsequência de x^k tal que $\lim_{k\to\infty} x^{k_j} = x^*$. Visto que $P_{\lambda_{k_j}}(x^{k_j}) = x^{k_j+1}$, temos que

$$\lambda_{k_j}(x^{k_j} - x^{k_j+1}) \in \partial f(x^{k_j+1}).$$

Assim, pela Definição 3,

$$f(x) - f(x^{k_j+1}) \geq \lambda_{k_j}\langle x^{k_j} - x^{k_j+1}, x - x^{k_j+1}\rangle,$$

e o resultado desejado segue por combinando os seguintes fatos: $\{\lambda_k\}$ é limitada, $\{x^{k_j}\}$ converge a x^* e f é contínua (f é convexa). □

4 Conclusões

Concluímos que a sequência gerada pelo método de ponto proximal converge a uma solução do problema de minimização (1), mesmo no caso que a função objetivo do problema de minimização é não necessariamente diferenciável.

Referências

[1] Martinet B., Régularization, d'inéqations variationelles par approximations successives, Revue Française d'Informatique et de Recherche Opertionelle, (1970), 154-159.

[2] Rockafellar, R. T., Monotone operators and the proximal point method, SIAM, J. Control. Optim., 14 (1970), 154-159.

[3] Izmailov, A., Solodov, M., Otimização Volume 1. Condições de otimalidade, elementos de análise conexa e de dualidade Rio de Janeiro: IMPA,2005.

[4] Mäkelä, M. M., Neitaanmäki, P., Nonsmooth optimization: analysis and algorithms with applications to optimal control Marko.

[5] Iusen, A.N., Proximal point methods in optimization, 20. Colóquio Brasileiro de Matemática, IMPA, (1995).

[6] Silva, S. R. P, Algoritmo de Ponto Proximal para Otimização em $\mathbb{R}^n$, UFG, Dissertação de mestrado, (2002).