Extensões do método ponto proximal para a minimização de composição de funções

Vitaliano S. Amaral – UFPI

XLIV Congresso Nacional de Matemática Aplicada e Computacional

Setembro de 2025

Método do Ponto Proximal

Dado o problema

$$\min f(x) \quad \text{sujeito a } x \in \mathbb{R}^n \tag{1}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$.

O **Método do Ponto Proximal - PPM** para resolver o Prblema (1) consiste em gerar uma sequência x^k tal que

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} f(x) + \frac{1}{2\alpha_k} \left\| x - x^k \right\|^2 .$$

Martinet[1] e Rockafellar[2] foram pioneiros em estudar esse tipo de métodos.

[1] B. Martinet, *Bréve communication. Régularisation d'inéquations variationnelles par approximations successives.* Revue française d'informatique et de recherche opérationnelle. Série rouge 4.R3 (1970), pp. 154–158.

[2] R. T. Rockafellar, *Monotone operator and the proximal point algorithm.* SIAM J. Control Optim. 14 (1976), pp. 877–898.

Extensões do PPM

O **PPM** foi estendido para composição de funções.

- No trabalho[3], foi proposto uma versão **exata** do PPM para

$$\min_{x\in\mathbb{R}^n} \; f(x) := g_1(x) + g_2(x) - h(x), \tag{2}$$

com g_1 semicontínua inferior, ∇g_2 Lipschitz e h convexa.

- Recentemente foi proposto estenções do PPM para o propblema

$$\min_{x\in\mathbb{R}^n} f(x) := g_1(x) + g_2(x) + h(x). \tag{3}$$

Em[4] foi estudado o caso em que g_1, g_2, h são convexas.
Em[5] o caso **não convexo**, com g_1 diferenciável. e g_2, h convexas.

---

[3] N.T. An, N.M. Nam, *Convergence analysis of a proximal point algorithm for minimizing differences of functions*, Optim., 66(1) (2017), 129–147.

[4] A. Yurtsever, V. Mangalick, and S. Sra, *Three operator splitting with subgradients, stochastic gradients, and adaptive learning rates.* NeurIPS (2021).

[5] A. Yurtsever, V. Mangalick, and S. Sra, *Three Operator Splitting with a Nonconvex Loss Function*, ICML (2021).

Extensões do PPM

Recentemente, (Amaral, V.S., Lopes, J.O., Santos, P.S.M., Silva, G.N.)[6], foi proposto uma versão **inexata** do PPM para o problema

$$\min_{x \in \mathbb{R}^n} \; f(x) := g_1(x) + g_2(x) - h(x),$$

assumindo:

- $g_1 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ própria e semicontínua inferior
- g_2 diferenciável com gradiente Hölder
- h convexa, possivelmente não diferenciável.

[6] Amaral, V.S., Lopes, J.O., Santos, P.S.M., Silva, G.N. (2024). *On the complexity of a quadratic regularization algorithm for minimizing nonsmooth and nonconvex functions*. OMS, 40(1), 1–23. `https://doi.org/10.1080/10556788.2024.2368578`.

Inexact proximal-type algorithm (IPTA)

O método IPTA: $x^0 \in \mathbb{R}^n$, $\alpha, \theta, \epsilon \in (0,1)$, $f_{\text{target}} \in \mathbb{R}$, $\rho_{\min} > 0$, $\eta > 0$.

Passo 1: Escolha $\epsilon_k \leqslant (\eta/\rho_k)^2$, calcule $w^k \in \partial_{\epsilon_k} h(x^k)$ e considere

$$M_{x^k,\rho_k}(x) = g_1(x) + g_2(x^k) - h(x^k) + \langle \nabla g_2(x^k) - w^k,\ x - x^k \rangle + \rho_k \|x - x^k\|^2.$$

Passo 2: Encontre $\bar{x}^{k+1}$ tal que $M_{x^k,\rho_k}(\bar{x}^{k+1}) \leqslant f(x^k)$ e

$$d\big(0,\ \partial^L g_1(\bar{x}^{k+1}) + \nabla g_2(x^k) + 2\rho_k(\bar{x}^{k+1} - x^k) - w^k\big) \leqslant \theta \|\bar{x}^{k+1} - x^k\|.$$

Passo 3: Se $d\big(0,\ \partial^L g_1(\bar{x}^{k+1}) + \nabla g_2(x^k) - w^k\big) < \epsilon$ ou $f(\bar{x}^{k+1}) \leqslant f_{\text{target}}$, **parar**.

Passo 4: Se

$$f(\bar{x}^{k+1}) \leqslant f(x^k) - \frac{\alpha}{36\rho_k}\epsilon^2,$$

fazer $x^{k+1} = \bar{x}^{k+1}$, $\rho_{k+1} = \rho_k$, $k \leftarrow k+1$ e voltar ao Passo 1; caso contrário $\rho_k \leftarrow 2\rho_k$ e voltar ao Passo 1.

Boa definição

- Coercividade de $M_{x^k,\rho_k}(x)$ garante a boa definição do Passo 2.
- Assumindo

$$g_2(y) \leqslant g_2(x) + \langle \nabla g_2(x), y - x \rangle + L\|y - x\|^{\beta+1}, \quad L > 0,\ \beta \in (0, 1],$$

provamos que

$$f(\bar{x}^{k+1}) \leqslant f(x^k) - \frac{\alpha}{36\rho}\, \epsilon^2.$$

sempre que

$$\rho \geqslant \max\left\{1, \frac{\theta}{4}, \left[\frac{L}{1-\alpha}\left(\frac{\epsilon}{6}\right)^{\beta-1} + \frac{\eta^2}{1-\alpha}\left(\frac{\epsilon}{6}\right)^{-2}\right]^{\frac{1}{\beta}}\right\}.$$

Complexidade do IPTA

- Números de Iterações é limitado superiormente por:

$$\frac{36\,\gamma\,(f(x^0) - f_{target})}{\alpha} \max\left\{\epsilon^{-\frac{\beta+1}{\beta}}, \eta^{\frac{2}{\beta}}\epsilon^{-\frac{2(\beta+1)}{\beta}}\right\}.$$

- Número de avaliações de f e seus subdiferenciais é limitado superiormente por:

$$\frac{36\,\gamma\,(f(x^0) - f_{target})}{\alpha} \max\{\epsilon^{-\frac{\beta+1}{\beta}}, \eta^{\frac{2}{\beta}}\epsilon^{-\frac{2(\beta+1)}{\beta}}\} + \log_2\left(\frac{\gamma}{\rho_{\mathbf{min}}}\right).$$

com

$$\gamma := 2\max\left\{1, \frac{\theta}{4}, \left[\frac{6^{1-\beta}L}{1-\alpha} + \frac{36}{1-\alpha}\right]^{1/\beta}, \rho_{min}\right\}.$$

Aplicações

O método IPTA pode resolver vários casos clássicos:

- **Caso diferenciável:** se $g_1 \equiv h \equiv 0$

$$\min_x g_2(x).$$

- **Função DC:** se $g_1 + g_2$ é convexa,

$$\min_x f(x) = g(x) - h(x), \quad g = g_1 + g_2.$$

- **Restrições convexas:** para C convexo,

$$\min_{x \in C} f(x) = g_2(x) - h(x),$$

  usando $g_1 = \delta_C$ (função indicadora de C).

Aplicações

- **o problema do lasso**

$$\min_{x\in\mathbb{R}^n} \left\{\frac{1}{2}\|Ax-b\|^2+\lambda\|x\|_1\right\},\ \lambda>0,\ \ A\in\mathbb{R}^{s\times n}, \tag{4}$$

considerando $g_1(x)=\lambda\|x\|_1,\ g_2(x)=\frac{1}{2}\|Ax-b\|^2$ e $h=0$.

- **o problema**

$$\min\left\{\|Ax-b\|^2 \colon 0\leqslant x_i\leqslant 1, i=1,2,\ldots,n\right\} \tag{5}$$

onde $A\in\mathbb{R}^{s\times n}$ e $b\in\mathbb{R}^s$.
Basta minimizar $f(x)=g_1(x)+g_2(x)$ com $g_1(x)=\delta_\Omega(x)$,
$\Omega=\{x\in\mathbb{R}^n : 0\leqslant x_i\leqslant 1, i=1,2,\ldots,n\},\ g_2(x)=\|Ax-b\|^2$.

- **Problema de mínimos quadrados penalizado por norma L_p ($1 < p < 2$):**

$$\min f(x) = \frac{1}{2}\|A(x) - b\|_2^2 + \frac{\lambda}{p}\|\Phi(x)\|_p^p,$$

onde $A : \mathbb{R}^n \to \mathbb{R}^m$ é diferenciável (não necessariamente linear), $b \in \mathbb{R}^m$, e Φ é operador linear.

Detalhes em:

A. Bernigaud, S. Gratton and E. Simon, A non-linear conjugate gradient in dual space for L_p-norm regularized non-linear least squares with application in data assimilation, *Numerical Algorithms*, **95**, 471–497 (2024).
https://doi.org/10.1007/s11075-023-01578-x

Testes Numéricos – Exemplo 2

Problema de regularização ℓ_1:

$$\min_x \ \tfrac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1,$$

com parâmetro $\lambda = 5 \times 10^{-5}$.

Configuração:

- Critério de parada: $\frac{\|x^k - x^{k-1}\|}{\|x^{k-1}\|} < \epsilon$.
- Métrica de avaliação: **PSNR**, definida por

$$\mathrm{PSNR}(x_k) = 10\log_{10}\left(\frac{255^2}{\mathrm{MSE}}\right), \quad \mathrm{MSE} = \frac{1}{mn}\|x_k - \overline{x}\|^2,$$

onde $\overline{x}$ é a imagem original.

Comparação com IMA e FISTA

Comparamos o IPTA com o FISTA(Fast iterative shrinkage-thresholding algorithm) seguindo[7]

Algoritmo	CPU (s)	Iterações	PSNR
FISTA	4.49	476	34.43
IPTA	0.89	435	28.63

(a) Original (b) Desfocada (c) FISTA (d) IPTA

- O IPTA tem bom tempo de CPU.
- O FISTA apresenta um PSNR mais alto do que o IMA e o IPTA.

[7] R. Wattanataweekul, and K. Janngam, *An accelerated common fixed point algorithm for a countable family of G-nonexpansive mappings with applications to image recovery*. J. Inequal. Appl. 2022 (2022), pp. 1–15.

Algumas outras extensões

Versões BCD[8]**:** Problema

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + g(x).$$

Para cada $i = 1, \cdots, q$ calcula $x^{k,i} = x^{k,i-1} + U_i s^k_{(i)}$ onde $s^k_{(i)} \in \mathbb{R}^{n_i}$ é uma solução do problema

$$\min_{s \in \mathbb{R}^{n_i}} \Psi_k(s) + \frac{\rho_k}{2}\|s\|^2,$$

onde $\Psi_{k,i-1}(x) =$
$\langle U_i^T \varphi_f(x^{k,i-1}, \lambda_k), s\rangle + \frac{1}{2}\langle B_{(i)}(x^{k,i-1})s, s\rangle + h(x^k + U_i s) - h(x^{k,i-1})$.

- **Complexidade:** $\mathcal{O}\left(\epsilon^{-\frac{\beta+1}{\beta}}\right)$.

[8] Amaral, V. (2025). - *A partially derivative-free cyclic block coordinate descent method for nonseparable composite optimization.* Mathematical Modelling and Analysis, 30(3), 535?552.

- **Problemas Multiobjetivo:** V.S. Amaral, P.B. Assunção, D.R. Souza[9] propuseram um método para

$$\min_{x\in\mathbb{R}^n} F(x), \quad F:\mathbb{R}^n \to \mathbb{R}^m,\ F_j = f_j + h_j,\ j = 1,\dots,m.$$

Os subproblemas aproximados têm a forma

$$\min_{x\in\mathbb{R}^n} \Phi_{x^k}(x) + \frac{\sigma_k}{2}\|x - x^k\|^2,$$

com

$$\Phi_{x^k}(x) = \max_j \left[\langle g_{f_j}(x^k,\lambda_k) + \tfrac{1}{2}B_j^k(x - x^k),\ x - x^k\rangle + h_j(x) - h_j(x^k)\right].$$

[9] V.S. Amaral, P.B. Assunção, D.R. Souza, *A partially derivative-free proximal method for composite multiobjective optimization in the Hölder setting*, arXiv:2508.20071 (2025).