

Método del Punto Proximal

Dr. Erik Alex Papa Quiroz

Universidad Nacional Mayor de San Marcos, UNMSM-PERÚ

Optimización

Facultad de Ciencias Matemáticas

Universidad Nacional Mayor de San Marcos

23/09/2025

Resumen

- 1 Introducción
- 2 Extensión para el caso irrestricto
- 3 Extensión para el ortante no negativo
- 4 Referencias

1. Problema Irrestricto

Considere el problema:

$$\begin{cases} \min f(x) \\ \text{s.a. :} \\ x \in \mathbb{R}^n. \end{cases}$$

El Método de Punto Proximal (MPP) genera una sucesión $\{x^k\}$ definido por:

$$x^0 \in \mathbb{R}^n,$$

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

donde $\lambda_k > 0$.

Datos históricos

- ▶ El MPP fue introducido por Martinet (1970).
- ▶ Rockafellar (1976) extendió el MPP para resolver ceros de operadores monótonos maximales.
- ▶ En los últimos años se estudia la extensión del método para problemas no convexos y problemas variacionales donde el operador no es monótono.

Importancia

- ▶ Se ha convertido en la base teórica para justificar la convergencia de los métodos de multiplicadores para resolver problemas de optimización. En particular, se ha demostrado que detrás de cada método multiplicador existe un método proximal que lo genera (Rockafellar, 1976).
- ▶ Se ha probado también que una gran clase de métodos de descomposición de problemas convexos son casos particulares del método del punto proximal para encontrar un cero de un operador monótono maximal, específicamente el método de Douglas-Rachford estudiado por Lions y Mercier (que abarca una gran clase de métodos de optimización) es una versión particular del método proximal, (Eckstein y Bertsekas, 1992).

Algoritmo

Dado x^0 y $\{\lambda_k\}$ sucesión de parámetros positivos.

Para $k = 0, 1, \dots$

Si: $0 \in \partial f(x^k)$, parar !!

Caso contrario:

Resolver el problema:

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

para obtener x^{k+1} y asignar $k + 1 \mapsto k$ y volver a la condición **Si**.

Resultados de Convergencia

- Si f es convexa y $\{\lambda_k\}$ satisface

$$\sum_{k=1}^{+\infty} (1/\lambda_k) = +\infty.$$

entonces,

$$\lim_{k \rightarrow \infty} f(x^k) = \inf \{f(x) : x \in \mathbb{R}^n\}.$$

- Además, si el conjunto óptimo X^* es no vacío, entonces $\{x^k\}$ converge a un punto óptimo del problema.

El caso cuadrático convexo

Si $f(x) = \frac{1}{2}x^T Ax - b^T x + c$, $A = A^T$ y $A \succeq 0$, entonces,

1. Las iteraciones son explícitas:

$$x^k = (A + \lambda_k I)^{-1} (b + \lambda_k x^{k-1}) .$$

2. En general si f es difernciable se tiene que

$$x^k = x^{k-1} - \frac{1}{\lambda_k} \nabla f(x^k).$$

El método, salvo el caso cuadrático, es implícito en el sentido tradicional de los métodos iterativos.

Ejercicio: Implementar

$$\left\{ \begin{array}{l} \min f(x) = \frac{1}{2}x^T \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} x - (2 \quad 1) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + c \\ \text{s.a :} \\ x \in \mathbb{R}^n \end{array} \right.$$

Usaremos el punto inicial $x^0 = (0, 0)$ y realizaremos pasos del método proximal para encontrar la solución óptima $x^* = (1, 1)$.

MPP para problemas Restringidos

Considere el problema:

$$\begin{cases} \min f(x) \\ \text{s.t. :} \\ x \in C, \end{cases}$$

donde f es una función convexa y C es un conjunto convexo cerrado con interior no vacío.

- ▶ El MPP para resolver problemas restringidos introduce una función llamada distancia que sustituye a $\|\cdot\|^2$ por $d(\cdot, \cdot)$ en la regularización proximal de tal manera que los puntos obtenidos sean viables y en la mayoría de los casos interiores.

MPP con distancias de Bregman

Dados x^0 y $\{\lambda_k\}$.

Para $k = 0, 1, \dots$

Si $0 \in \partial f(x^k)$ finalizar !!!!!

Caso contrario, resolver el problema:

$$\begin{cases} \min f(x) + \lambda_k D_h(x, x^k) \\ \text{s.a :} \\ x \in C \end{cases}$$

Obteniendo $x^{k+1} \in C$, donde,

$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle$ y h es una función de Bregman.

Hacer $k = k + 1$ y volver al test de parada **Si**.

Fin Para

Función de Bregman

La función $h : \bar{S} \rightarrow \mathbb{R}$ es llamada función de Bregman con zona S , si se verifican las siguientes condiciones:

- B1). h es estrictamente convexa y continua en $\bar{S}$.
- B2). h es continuamente diferenciable en S .
- B3). $\forall x \in \bar{S}, \forall \delta \in \mathbb{R}$, el conjunto de nivel respecto a la segunda variable
$$L_{D_h}(x, \delta) = \{y \in S \mid D_h(x, y) \leq \delta\}$$
es acotada.
- B4). Si $\{y^k\} \subset S$ converge a y , entonces $\{D_h(y, y^k)\}$ converge a cero.
- B5). Si $\{x^k\} \subset \bar{S}$ y $\{y^k\} \subset S$ son sucesiones tal que $\{y^k\}$ es acotado, $\lim x^k = x^*$ y $\lim D_h(y^k, x^k) = 0$, entonces $\lim y^k = x^*$.

Ejemplos de Función de Bregman

1. Si $S = \mathbb{R}_{++}^n$, $h(x) = \sum_{i=1}^n x_i \ln x_i$, con la convención $0 \ln 0 = 0$, es una función de Bregman y

$$D_h(x, y) = \sum_{i=1}^n (x_i \ln(x_i/y_i) + y_i - x_i)$$

llamada divergencia de Kullback-Leibler, que es usada en estadística.

2. Si $S = \mathbb{R}_{++}^n$, $h(x) = \sum_{i=1}^n (x_i^\alpha - x_i^\beta)$ con $\alpha \geq 1$, $0 < \beta < 1$.

Para $\alpha = 2$, $\beta = 1/2$ obtenemos:

$$D_h(x, y) = \|x - y\|^2 + \sum_{i=1}^n \frac{1}{2\sqrt{y_i}} (\sqrt{x_i} - \sqrt{y_i})^2.$$

Para $\alpha = 1, \beta = 1/2$ tenemos:

$$D_h(x, y) = \sum_{i=1}^n \frac{1}{2\sqrt{y_i}} (\sqrt{x_i} - \sqrt{y_i})^2.$$

Proposición:

Si h es una función de Bregman con zona S entonces:

- i) $\forall x \in \bar{S}, \forall z, y \in S,$
 $D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle .$
- ii) $\forall x, y \in S,$
 $\nabla_x D_h(x, y) = \nabla h(x) - \nabla h(y).$
- iii) $D_h(., y)$ es estrictamente convexa $\forall y \in S.$

Resultados de convergencia

- ▶ Si f es convexa y $\{\lambda_k\}$ satisface

$$\sum_{k=1}^{+\infty} (1/\lambda_k) = +\infty.$$

entonces,

$$\lim_{k \rightarrow \infty} f(x^k) = \inf \{f(x) : x \in \mathbb{R}_+^n\}.$$

- ▶ Además, si el conjunto óptimo X^* es no vacío, entonces $\{x^k\}$ converge a un punto óptimo del problema.

Ejemplo PMKullbackLeibler

Considere el problema

$$\min\{f(x) = -x_1^{0.5}x_2^{0.5} : 2x_1 + x_2 \leq 6, x_1 \geq 0, x_2 \geq 0\}.$$

La solución óptima es: $x_1^* = 3/2, x_2^* = 3$.

Tomamos la distancia:

$$d(x, y) = \sum_{i=1}^n (x_i \ln(x_i/y_i) + y_i - x_i)$$

y definir

$$D_h(x, z) = d(y(x), y(z)), \text{ con } y(x) = Ax - b$$

Obteniendo:

$$D_h(x, z) = (6 - 2x_1 - x_2) \ln\left(\frac{6 - 2x_1 - x_2}{6 - 2z_1 - z_2}\right) + (-2z_1 - z_2 + 2x_1 + x_2) + \\ + (x_1 \ln(x_1/z_1) + z_1 - x_1) + (x_2 \ln(x_2/z_2) + z_2 - x_2).$$

Problemas en el ortante no negativo
Considere el problema

$$\min\{f(x) : x \in \mathbb{R}_+^n\}$$

donde f es una función convexa no necesariamente diferenciable.

MPP con distancias ϕ -divergencia

Dado x^0 y $\{\lambda_k\}$.

Para $k = 0, 1, \dots, n$.

Si $0 \in \partial f(x^k)$ parar !!!!!

Caso contrario, resolver el problema:

$$\begin{cases} \min f(x) + \lambda_k d_\varphi(x, x^k) \\ \text{s.a} \\ x \in \mathbb{R}_+^n \end{cases}$$

obteniendo $x^{k+1} \in C$, donde,

$$d_\varphi(x, y) = \sum_{j=1}^p y_j \varphi\left(\frac{x_j}{y_j}\right),$$

Asignar $k \rightarrow k + 1$ y volver al test **Si**.

$\varphi : \mathbb{R} \rightarrow (-\infty, +\infty]$ es una función convexa propia y cerrada, tal que, $\text{dom}\varphi \subseteq [0, +\infty)$, y que cumple las siguientes propiedades:

- (i) φ es dos veces continuamente diferenciable en $\text{int}(\text{dom}\varphi) = (0, +\infty)$.
- (ii) φ es estrictamente convexa en su dominio.
- (iii) $\lim_{t \rightarrow 0} \varphi'(t) = -\infty$.
- (iv) $\varphi(1) = \varphi'(1) = 0$ y $\varphi''(1) > 0$.

Denotamos por Φ las clases de funciones que satisfacen (i)-(iv).
Introduciremos dos sub clases de núcleo φ :

$$\Phi_1 := \left\{ \varphi \in \Phi, \text{ existe } M > 0 \text{ tal que } \varphi''(t) \leq M, \forall t \geq 1 \right\}.$$

$$\Phi_2 := \left\{ \varphi \in \Phi; \varphi''(1)\left(1 - \frac{1}{t}\right) \leq \varphi'(t) \leq \varphi''(1)(t-1), \forall t > 0 \right\}.$$

Existen funciones $\varphi \in \Phi_1 \cap \Phi_2$ por ejemplo:

1. $\varphi_1(t) = t \ln t + 1$, $\text{dom} \varphi = [0, +\infty)$, obteniendo

$$d_{\varphi_1(t)}(x, y) = \sum_{i=1}^p (x_i \ln \frac{x_i}{y_i} + y_i - x_i).$$

2. $\varphi_2(t) = -\ln t + t - 1$, $\text{dom} \varphi = (0, +\infty)$, obteniendo

$$d_{\varphi_2(t)}(x, y) = \sum_{j=1}^n y_j \ln(y_j/x_j) + x_j - y_j$$

3. $\varphi_3(t) = 2(\sqrt{t} - 1)^2$, $\text{dom} \varphi = [0, +\infty)$, obteniendo

$$d_{\varphi_3(t)}(x, y) = 2 \sum_{j=1}^n (\sqrt{x_j} - \sqrt{y_j})^2$$

Resultados de convergencia

- ▶ Si f es convexa y $\{\lambda_k\}$ satisface

$$\sum_{k=1}^{+\infty} (1/\lambda_k) = +\infty.$$

entonces,

$$\lim_{k \rightarrow \infty} f(x^k) = \inf \{f(x) : x \in \mathbb{R}_+^n\}.$$

- ▶ Además, si el conjunto óptimo X^* es no vacío, entonces $\{x^k\}$ converge a un punto óptimo del problema.

Ejemplo PMPhi

Considere el problema

$$\min\{f(x) = -x_1^{0.5}x_2^{0.5} : 2x_1 + x_2 \leq 6, x_1 \geq 0, x_2 \geq 0\}.$$

La solución óptima es: $x_1^* = 3/2, x_2^* = 3$.

Tomamos la distancia:

$$d(x, y) = \sum_{i=1}^n (y_i \ln(y_i/x_i) + x_i - y_i)$$

y definir

$$D_\phi(x, z) = d(y(x), y(z)), \text{ con } y(x) = Ax - b$$

obteniendo:

$$D_\phi(x, z) = (6 - 2z_1 - z_2) \ln\left(\frac{6 - 2z_1 - z_2}{6 - 2x_1 - x_2}\right) + (-2x_1 - x_2 + 2z_1 + z_2) + \\ + (z_1 \ln(z_1/x_1) + x_1 - z_1) + (z_2 \ln(z_2/x_2) + x_2 - z_2).$$

MPP con distancias Log-cuadráticas

Dado x^0 y $\{\lambda_k\}$.

Para $k = 0, 1, \dots, n$.

Si $0 \in \partial f(x^k)$ parar !!!!!

Caso contrario, resolver el problema:

$$\begin{cases} \min f(x) + \lambda_k d_\varphi(x, x^k) \\ s.a \\ x \in \mathbb{R}_+^n \end{cases}$$

obteniendo $x^{k+1} \in C$, donde,

$$d_\varphi(x, y) = \sum_{j=1}^p y_j^2 \varphi\left(\frac{x_j}{y_j}\right),$$

Asignar $k \rightarrow k + 1$ y volver al test **Si**.

Usando la distancia: $d_0(u, v) := \sum_{i=1}^2 v_i^2 \varphi(\frac{u_i}{v_i})$, con

$\varphi(t) := \sigma(-\log t + t - 1) + \frac{\nu}{2}(t - 1)^2$, $\nu > \sigma > 0$, obtenemos:

$$d_0(u, v) := \sum_{i=1}^2 \frac{\nu}{2} (u_i - v_i)^2 + \sigma (v_i^2 \log \frac{v_i}{u_i} + u_i v_i - v_i^2).$$

Ejemplo PMLog

Considere el problema

$$\min\{f(x) = -x_1^{0.5}x_2^{0.5} : 2x_1 + x_2 \leq 6, x_1 \geq 0, x_2 \geq 0\}.$$

La solución óptima es: $x_1^* = 3/2, x_2^* = 3$.

Tomamos la distancia:

$$d_0(u, v) := \sum_{i=1}^2 \frac{\nu}{2} (u_i - v_i)^2 + \sigma(v_i^2 \log \frac{v_i}{u_i} + u_i v_i - v_i^2).$$

4. Referencias

ABB04 Alvarez, F., Bolte, J., and Brahic, O., Hessian Riemannian Gradient Flows in Convex Programming, SIAM J. Optim., 43, 2, 477-501, 2004.

Attouch, H., and Bolte, J. On the Convergence of the Proximal Algorithm for Nonsmooth Functions Involving Analytic Features, Mathematical Programming, 116, 1-2, B, 5-16, 2009.

Attouch, H., and Teboulle, M. Regularized Lotka-Volterra Dynamical System as Continuous Proximal-Like Method in Optimization, Journal Optimization Theory and Applications, 121, 541-580, 2004.

Chen, J. S., and S. Pan, S. H. Proximal-Like Algorithm for a Class of Nonconvex Programming, Pacific Journal of Optimization, 4, 319-333, 2008.

Cunha, G. F. M., da Cruz Neto, J. X., and Oliveira, P. R. A Proximal Point Algorithm with ϕ -Divergence to Quasiconvex Programming, Optimization, 59, 777-792, 2010.

Eckstein J. y Bertsekas D.P. On the Douglas-Rachford splitting method and the proximal algorithm for maximal monotone operators. Mathematical Programming Vol 55, N°3: 293-318, 1992.

[Gromicho J.](#) Quasiconvex optimization and location theory. Kluwer Academic Publishers, Dordrecht, the Netherlands, (1998).

[Guler O.](#) On the Convergence of the Proximal Point Algorithm for Convex Minimization. SIAM J. Control and Optimization, 29 (2), 403-4019, 1991.

[Kaplan, A., and Tichatschke, R.](#) Proximal Point Methods and Nonconvex Optimization, Journal of global Optimization, 13, 389-406, 1998.

[Martinet B.](#) Régularisation, D'inéquations Variationelles par Approximations Successives, Revue Française D'Informatique et de Recherche Operationelle, 154-159, 1970.

[Mas-Colell A, Whinston MD, Green JR.](#) Microeconomic theory, Oxford University Press: New York, NY, USA; 1995.

[Pan, S. H., and Chen, J. S.](#) Entropy-Like Proximal Algorithms Based on a Second-Order Homogeneous Distances Function for Quasiconvex Programming, Journal of Global Optimization, 39, 555-575, 2007.

[Papa Quiroz, E.A., Oliveira, P.R.](#) Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard Manifolds. Journal of Convex Analysis, 16,1, pp 49-69, 2009.

[Papa Quiroz, E.A., Oliveira, P.R.](#) Full Convergence of the Proximal Point Method for Quasiconvex Functions on Hadamard Manifolds. ESAIM: Control, Optimization and Calculus of Variations, v.18(2), pp. 483-500, 2012.

Papa Quiroz, E.A. An Extension of the Proximal Point Algorithm with Bregman Distances on Hadamard Manifolds, Sometido a JOGO, 2011.

Papa Quiroz, E.A., Mallma Ramirez, L. Proximal Point Algorithm with Generalized Distances for Quasiconvex functions with linear inequalities, Sometido a EJOR, 2012.

Papa Quiroz, E.A., Oliveira, P.R. Proximal Point Method for Minimizing Quasiconvex Locally Lipschitz Functions on Hadamard Manifolds, Nonlinear Analysis, 75, pp. 5924-5932, 2012.

Papa Quiroz, E.A., Oliveira, P.R. An extension of proximal methods for quasiconvex minimization on the nonnegative orthant, European Journal of Operational Research, 2016, pp. 26-32, 2012.

Rockafellar, R. T. *Augmented Lagrangians and Application of Proximal Point Algorithm in Convex Programming*, Mathematics of Operation Research, Vol.1, pp. 97-116, 1976.

Takayama A. Mathematical economics, 2nd Edition, Cambridge Univ. Press, 1995.

Tseng P. Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, Journal of Optimization Theory and Applications, 109, 3, 475-494, 2001.

Souza, S. S., Oliveira, P. R., da Cruz Neto, J. X., and Soubeyran, A. A Proximal Method with Separable Bregman Distance for Quasiconvex Minimization on the Nonnegative Orthant, European Journal of Operational Research, 201, 365-376, 2010.