Proximal point algorithm

Andersen Ang

Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Canada

ms$\mathbf{x}$ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: November 29, 2021 Last update: December 9, 2021

Setup

- Consider

$$\min_{\mathbf{x}} f(\mathbf{x})$$

where $f : \mathbb{R}^n \to \bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ is proper, convex, lower semi-continuous and possibly non-smooth.

- Moreau proximal operator

$$\mathbf{u}^* \in \operatorname*{argmin}_{\mathbf{u}} \; f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|_2^2.$$

- As f is convex, so $f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|_2^2$ is 1-strongly convex and thus the (global) minimizer is unique:

$$\mathbf{u}^* = \operatorname*{argmin}_{\mathbf{u}} \; f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x} - \mathbf{u}\|_2^2.$$

- Martinet's Proximal Point Method (PPM) is to iterate

$$\mathbf{x}_{k+1} = \operatorname{prox}_{\gamma f}(\mathbf{x}_k), \;\; \forall k.$$

What's the point of PPM

- Recall that the problem

$$\min_{\mathbf{x}} f(\mathbf{x})$$

  has a non-smooth f, i.e., $\nabla f(\mathbf{x})$ is not continuous for all $\mathbf{x}$. In other words, we are now trying to minimize a non-smooth function f.

- For non-smooth minimization, gradient descent cannot be used here because we do not have $\nabla f(\mathbf{x})$ for all $\mathbf{x}$.
  - One way to solve non-smooth minimization is to use the subgradient method, which is not the focus of this document.

- In this document, we consider solving non-smooth minimization by PPM.

Other points on PPM

$$\text{Problem}: \min_{\mathbf{x}} f(\mathbf{x}), \qquad \text{PPM}: \mathbf{x}_{k+1} = \text{prox}_{\gamma f}(\mathbf{x}_k) = \underset{\mathbf{u}}{\text{argmin}}\ \gamma f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2.$$

- PPM is one of the earliest proximal algorithm (since 1970[1]).
- PPM is a conceptual algorithm
  - Historically, PPM has not found many applications
  - Each PPM iteration requires us to minimize the function f plus a quadratic: in general, if f is already difficult to minimize, adding a quadratic makes it even more difficult to minimize.
  - Only in some special cases, solving the prox is easier than minimizing f directly
- PPM is the basis of augmented Lagrangian.

[1] B. Martinet, "Régularisation d'inéquations variationnelles par approximations successives," Revue Francaise de Informatique et Recherche Opérationelle, 1970.

Illustration ... (1/4)

Problem : $\min_{\mathbf{x}} f(\mathbf{x})$, PPM : $\mathbf{x}_{k+1} = \text{prox}_{\gamma f}(\mathbf{x}_k) = \underset{\mathbf{u}}{\text{argmin}}\ \gamma f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2.$

- Consider a simple scalar problem that $f(\mathbf{x}) = |x|$.
- Suppose we start with $x_0 = -3$

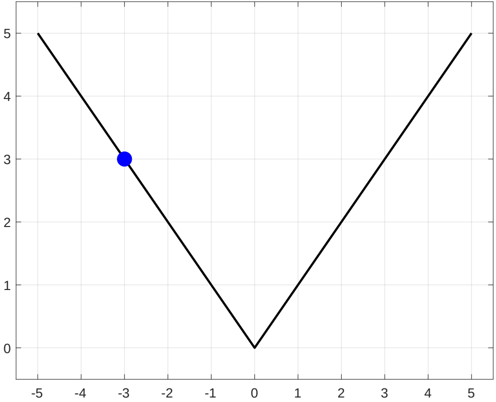


Illustration ... (2/4)

Problem : $\min_{\mathbf{x}} f(\mathbf{x})$, PPM : $\mathbf{x}_{k+1} = \text{prox}_{\gamma f}(\mathbf{x}_k) = \underset{\mathbf{u}}{\text{argmin}}\ \gamma f(\mathbf{u}) + \dfrac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2.$

- To find x_1, we first construct the function $\gamma f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2$.
- For simplicity let $\gamma = 1$, so the plot of $|u| + 0.5(x_0 - u)^2$ is

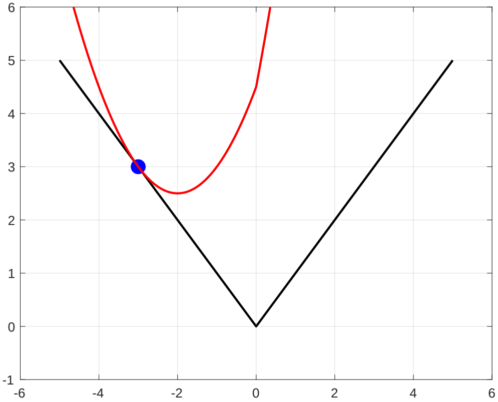


Illustration ... (3/4)

Problem : $\min_{\mathbf{x}} f(\mathbf{x})$, PPM : $\mathbf{x}_{k+1} = \mathrm{prox}_{\gamma f}(\mathbf{x}_k) = \underset{\mathbf{u}}{\mathrm{argmin}}\ \gamma f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2.$

- We set x_1 as the minimizer of $|u| + 0.5(x_0 - u)^2$, which is $x_1 = -2$ in this case

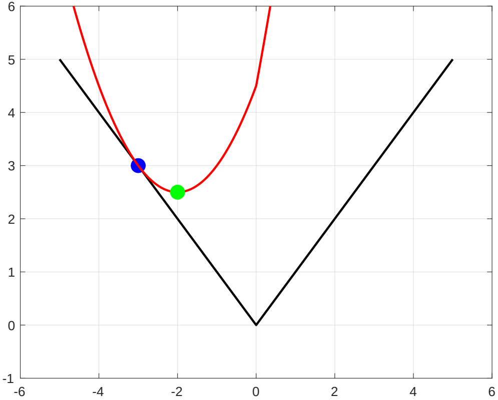
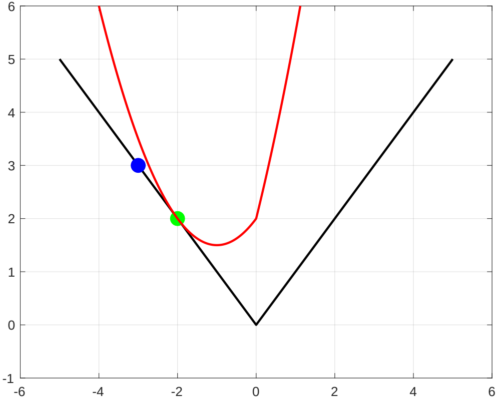


Illustration ... (4/4)

Problem : $\min_{\mathbf{x}} f(\mathbf{x})$, PPM : $\mathbf{x}_{k+1} = \text{prox}_{\gamma f}(\mathbf{x}_k) = \underset{\mathbf{u}}{\text{argmin}}\ \gamma f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2.$

- Now we construct the function $\gamma f(\mathbf{u}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{u}\|_2^2$ again on x_1, and the whole process repeats.



Convergence of PPM

- Now we recall proximal gradient method. For the problem

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}),$$

  where g is L_g-smooth and convex, the proximal gradient method is to iterate

$$\mathbf{x}_{k+1} = \operatorname{prox}_{\gamma f}\Big(\mathbf{x}_k - \alpha \nabla g(\mathbf{x}_k)\Big),$$

  with $\alpha \in]0, \frac{2}{L_g}[$.

- Now we see that PPM is the special case of proximal gradient method without the smooth part, therefore, the convergence of proximal gradient method applies to PPM.

- For completeness, we prove the convergence of PPM now.

Convergence proof of PPM

- By definition of PPM: $\mathbf{x}_{k+1} = \underset{\mathbf{u}}{\operatorname{argmin}} f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}_k\|_2^2$, so by subgradient first-order optimality,

$$\mathbf{0} \in \partial f(\mathbf{x}_{k+1}) + \mathbf{x}_{k+1} - \mathbf{x}_k \quad \Longrightarrow \quad -(\mathbf{x}_{k+1} - \mathbf{x}_k) \in \partial f(\mathbf{x}_{k+1})$$

  i.e. the vector $-(\mathbf{x}_{k+1} - \mathbf{x}_k)$ is a subgradient of f at $\mathbf{x}_{k+1}$

- Since f is convex,

$$f(\mathbf{z}) \geq f(\mathbf{x}_{k+1}) + \mathbf{q}^\top(\mathbf{z} - \mathbf{x}_{k+1}), \; \mathbf{q} \in \partial f(\mathbf{x}_{k+1})$$

  Using the fact that $-(\mathbf{x}_{k+1} - \mathbf{x}_k)$ is a subgradient of f at $\mathbf{x}_{k+1}$:

$$f(\mathbf{z}) \geq f(\mathbf{x}_{k+1}) - (\mathbf{x}_{k+1} - \mathbf{x}_k)^\top(\mathbf{z} - \mathbf{x}_{k+1}).$$

  Rearrange

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{z}) + (\mathbf{x}_{k+1} - \mathbf{x}_k)^\top(\mathbf{z} - \mathbf{x}_{k+1}) \\ &= f(\mathbf{z}) - (\mathbf{x}_k - \mathbf{x}_{k+1})^\top(\mathbf{z} - \mathbf{x}_{k+1}). \end{aligned}$$

- Put $\mathbf{z} = \mathbf{x}^*$

$$f(\mathbf{x}_{k+1}) \leq f^* - (\mathbf{x}_k - \mathbf{x}_{k+1})^\top(\mathbf{x}^* - \mathbf{x}_{k+1}).$$

- Now we have

$$f(\mathbf{x}_{k+1}) - f^* \le -(\mathbf{x}_k - \mathbf{x}_{k+1})^\top (\mathbf{x}^* - \mathbf{x}_{k+1}).$$

- A tricky step

$$f(\mathbf{x}_{k+1}) - f^* \le -(\mathbf{x}_k - \mathbf{x}_{k+1})^\top (\mathbf{x}^* - \mathbf{x}_{k+1}) + \frac{1}{2}\|\mathbf{x}_k - \mathbf{x}_{k+1}\|_2^2$$

- A very tricky step

$$\begin{aligned} -(\mathbf{x}_k - \mathbf{x}_{k+1})^\top (\mathbf{x}^* - \mathbf{x}_{k+1}) + \tfrac{1}{2}\|\mathbf{x}_k - \mathbf{x}_{k+1}\|_2^2 &= \tfrac{1}{2}\Big(\|\mathbf{x}_k - \mathbf{x}_{k+1} - (\mathbf{x}^* - \mathbf{x}_{k+1})\|_2^2 - \|\mathbf{x}^* - \mathbf{x}_{k+1}\|_2^2\Big) \\ &= \tfrac{1}{2}\Big(\|\mathbf{x}_k - \mathbf{x}^*\|_2^2 - \|\mathbf{x}^* - \mathbf{x}_{k+1}\|_2^2\Big) \end{aligned}$$

- We now have a telescoping sum

$$f(\mathbf{x}_{k+1}) - f^* \le \frac{1}{2}\Big(\|\mathbf{x}_k - \mathbf{x}^*\|_2^2 - \|\mathbf{x}^* - \mathbf{x}_{k+1}\|_2^2\Big)$$

- Sum from $k = 0$ to k

$$\sum_{i=0}^{k} f\Big((\mathbf{x}_i) - f^*\Big) \leq \frac{1}{2}\Big(\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 - \|\mathbf{x}^* - \mathbf{x}_{k+1}\|_2^2\Big) \leq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

- $f(\mathbf{x}_k)$ is non-increasing: $f(\mathbf{x}_k) \leq f(\mathbf{x}_{k-1}) \leq \cdots \leq f(\mathbf{x}_0)$

$$\sum_{i=0}^{k} \Big(f(\mathbf{x}_k) - f^*\Big) \leq \sum_{i=0}^{k} \Big(f(\mathbf{x}_i) - f^*\Big) \leq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

$\Longrightarrow$

$$\sum_{i=0}^{k} \Big(f(\mathbf{x}_k) - f^*\Big) \leq \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

$\Longrightarrow$

$$f(\mathbf{x}_k) - f^* \leq \frac{1}{2k}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

Last page - summary

- For the problem

$$\min_{\mathbf{x}} f(\mathbf{x})$$

where $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ that is proper, convex, lower semi-continuous and possibly non-smooth, the Proximal Point method is to iterate

$$\mathbf{x}_{k+1} = \text{prox}_{\gamma f}(\mathbf{x}_k)$$

- PPM is a special case of proximal gradient method
- Convergence rate of PPM

$$f(\mathbf{x}_k) - f^* \leq \frac{1}{2k}\|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$$

End of document