Sufficient descrease lemma of proximal gradient operator

Andersen Ang

Mathématique et recherche opérationnelle
UMONS
Belgium

Email: manshun.ang@umons.ac.be

Homepage: angms.science

August 23, 2018

Overview

Proximal mapping

The definition of proximal mapping : given a function f,

$$\text{prox}_f(\mathbf{x}) = \arg\min_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 \right\}$$

- input : $\mathbf{x}$, a vector
- function $f : \Omega \to (-\infty \;\; +\infty]$, meaning that
  - dom f is a set $\Omega \subset \mathbb{R}^n$
  - range is lower bounded
  - range is not upper bounded, so that means f can gives value $+\infty$
- output : $\mathbf{u}$, a vector in the subset of dom f, i.e. $\mathbf{u} \in \Omega' \subset \Omega$
- Ω' can be empty, singleton, or a set with multiple vectors
- If f is proper, closed and convex, then Ω' is singleton, meaning that $\text{prox}_f(\mathbf{x})$ exists (therefore non-empty) and unique. (In fact this is called the first prox theorem.)

Theorems on proximal mapping

For all vectors $\mathbf{x}, \mathbf{u} \in$dom$f$ where $f : \Omega \to (-\infty \ \ +\infty]$ is proper, closed and convex function, then the following are equivalent.

$$
\begin{aligned}
&\text{(i)} && \mathbf{u} = \text{prox}_f(\mathbf{x}) \\
&\text{(ii)} && \mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u}) \\
&\text{(iii)} && \langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \le f(\mathbf{y}) - f(\mathbf{u}) \ \forall \mathbf{y} \in \text{dom} f
\end{aligned}
$$

Proof (Equivalence of (i) and (ii))
As f is convex, so $prox_f(\mathbf{x})$ is unique. By definition of prox, $\mathbf{u} = \text{prox}_f(\mathbf{x})$ iff $\mathbf{u}$ is the minmizer of the problem

$$\arg\min_{\mathbf{v}} \left\{ f(\mathbf{v}) + \frac{1}{2}\|\mathbf{v} - \mathbf{x}\|_2^2 \right\}$$

i.e. $\mathbf{v} = \mathbf{u}$.
The first order optimality condition (also called Fermat's rule) is then

$$0 \in \partial f(\mathbf{u}) + \mathbf{u} - \mathbf{x}$$

Rearrange we get (ii).

Theorems on proximal mapping

For all vectors $\mathbf{x}, \mathbf{u} \in$dom$f$ where $f : \Omega \to (-\infty \;\; +\infty]$ is proper, closed and convex function, then the following are equivalent.

$$
\begin{aligned}
&\text{(i)} && \mathbf{u} = \text{prox}_f(\mathbf{x}) \\
&\text{(ii)} && \mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u}) \\
&\text{(iii)} && \langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \leq f(\mathbf{y}) - f(\mathbf{u}) \; \forall \mathbf{y} \in \text{dom} f
\end{aligned}
$$

Proof (Equivalence of (ii) and (iii))
Based on the definition of sub-gradient, we have

$$f(\mathbf{y}) \geq f(\mathbf{u}) + \langle \partial f(\mathbf{u}), \mathbf{y} - \mathbf{y} \rangle$$

Rearrange

$$f(\mathbf{y}) - f(\mathbf{u}) \geq \langle \partial f(\mathbf{u}), \mathbf{y} - \mathbf{y} \rangle$$

Using (ii), $\mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u})$

$$f(\mathbf{y}) - f(\mathbf{u}) \geq \langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{y} \rangle$$

Proximal gradient operator

Consider composite functon

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

where

- f is smooth but not necessarily convex
    - f is proper, closed
    - dom$f$ is convex but f is not necessarily convex
    - f is L_f-smooth over the iterior of dom$f$
- g is convex but not necessarily smooth
    - g is proper, closed, convex
    - dom$g \subset$ dom$f$
    - g is not necessarily smooth
- f, g are both extended valued function : their rages are $(-\infty + \infty]$

Proximal gradient operator

Let x^+ be the output of the proximal gradient operator

$$\mathbf{x}^+ = \mathsf{prox}_t^g\big(\mathbf{x} - t\nabla f(\mathbf{x})\big)$$

which is

$$\mathbf{x}^+ = \arg\min_{\mathbf{v}} \left\{ tg(\mathbf{v}) + \frac{1}{2}\|\mathbf{v} - \big(\mathbf{x} - t\nabla f(\mathbf{x})\big)\|_2^2 \right\}$$

Interpretation : $\mathbf{x}^+$ is the minimizer of $g(\mathbf{x})$ scaled by t plus a local quadratic model of f at the point $\mathbf{x}$

Sufficient descrease lemma of proximal gradient operator

$\forall \mathbf{x} \in \text{int}(\text{dom} f)$, let $L \in \left(\frac{L_f}{2}, \infty\right)$. Then

$$f(\mathbf{x}) + g(\mathbf{x}) \geq f(\mathbf{x}^+) + g(\mathbf{x}^+) + \left(L - \frac{L_f}{2}\right)\|\mathbf{x} - \mathbf{x}^+\|_2^2$$

- $L \in \left(\frac{L_f}{2}, \infty\right)$, not $L \in \left[\frac{L_f}{2}, \infty\right)$, meaning $L = \gamma \frac{L_f}{2}$, $\gamma > 1$, which means $\left(L - \frac{L_f}{2}\right) > 0$
- Why it is called decrease : the above inequality is

  $$F(\mathbf{x}) \geq F(\mathbf{x}^+) + \text{something positive},$$

  meaning that $F(\mathbf{x}^+)$ is less than $F(\mathbf{x})$ so the function value decreased after proximal gradient operator
- How to prove it : use the statement (iii) of the theorems on proximal mapping, and the smoothness of f

Proving the sufficient descrease lemma

Lemma $\forall \mathbf{x} \in \text{int}(\text{dom} f)$, let $L \in \left(\frac{L_f}{2}, \infty\right)$. Then

$$f(\mathbf{x}) + g(\mathbf{x}) \geq f(\mathbf{x}^+) + g(\mathbf{x}^+) + \left(L - \frac{L_f}{2}\right)\|\mathbf{x} - \mathbf{x}^+\|_2^2$$

Proof. As f is L_f-smooth, so we have

$$f(\mathbf{x}^+) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{x}^+ - \mathbf{x} \rangle + \frac{L_f}{2}\|\mathbf{x}^+ - \mathbf{x}\|_2^2 \tag{1}$$

Recall the statement (iii) of the theorems on proximal mapping

$$\langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \leq f(\mathbf{y}) - f(\mathbf{u})$$

As proximal is designed for non-smooth function so we put $f = \frac{1}{L}g$, and $\mathbf{x} = \mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x})$, $\mathbf{y} = \mathbf{x}$, $\mathbf{u} = \mathbf{x}^+$ and get

$$\langle \mathbf{x} - \frac{1}{L}\nabla f(\mathbf{x}) - \mathbf{x}^+,\ \mathbf{x} - \mathbf{x}^+ \rangle \leq \frac{1}{L}g(\mathbf{x}) - \frac{1}{L}g(\mathbf{x}^+)$$

Proving the sufficient descrease lemma

Rearrange we have $\langle (\mathbf{x}-\mathbf{x}^+) - \frac{1}{L}\nabla f(\mathbf{x}),\ \mathbf{x}-\mathbf{x}^+\rangle \leq \frac{1}{L}\Big(g(\mathbf{x}) - g(\mathbf{x}^+)\Big)$.
Expand we get

$$\|\mathbf{x}-\mathbf{x}^+\|_2^2 - \langle (\mathbf{x}-\mathbf{x}^+), \frac{1}{L}\nabla f(\mathbf{x})\rangle \leq \frac{1}{L}\Big(g(\mathbf{x}) - g(\mathbf{x}^+)\Big) \qquad (2)$$

Together with (1) :

$$f(\mathbf{x}^+) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{x}^+ - \mathbf{x}\rangle + \frac{L_f}{2}\|\mathbf{x}^+ - \mathbf{x}\|_2^2$$

We can see we have to cancel the term $\langle \nabla f(\mathbf{x}), \mathbf{x}^+ - \mathbf{x}\rangle$ by summing the two equations. First (2) gives

$$L\|\mathbf{x}-\mathbf{x}^+\|_2^2 + \langle (\mathbf{x}^+ - \mathbf{x}), \nabla f(\mathbf{x})\rangle \leq g(\mathbf{x}) - g(\mathbf{x}^+)$$

Add to (1)

$$L\|\mathbf{x}-\mathbf{x}^+\|_2^2 + f(\mathbf{x}^+) \leq f(\mathbf{x}) + \frac{L_f}{2}\|\mathbf{x}^+ - \mathbf{x}\|_2^2 + g(\mathbf{x}) - g(\mathbf{x}^+)$$

Rearrange we get the result and we finish the proof.

Last page - summary

- $\mathbf{u} = \text{prox}_f(\mathbf{x}) \iff \mathbf{x}-\mathbf{u} \in \partial f(\mathbf{u}) \iff \langle \mathbf{x}-\mathbf{u}, \mathbf{y}-\mathbf{y} \rangle \leq f(\mathbf{y}) - f(\mathbf{u})$
- Sufficient decrease lemma of proximal gradient operator

Reference

Bolte Jerome, Shoham Sabach, and Marc Teboulle.
"Proximal alternating linearized minimization or nonconvex and nonsmooth problems." Mathematical Programming 146.1-2 (2014): 459-494.

Beck Amir
First-Order Methods in Optimization. Vol. 25. SIAM, 2017.

End of document