Sandwich theorem for β-smooth convex function

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be   Homepage: angms.science

First draft : June 6, 2017
Last update : July 27, 2019

Overview

Convex β-smooth function

A function $f(x)$ is convex and β-smooth if :

- (Convex) $f : \mathbb{R}^n \to \mathbb{R}$ and $\text{dom} f$ are convex
- (β-smooth) ∇f is Lipschitz continous with Lipschitz constant β. i.e., for any two points a, $b \in \text{dom} f$, we have a constant $\beta \in \mathbb{R}$ such that:
$$\|\nabla f(a) - \nabla f(b)\| \leq \frac{\beta}{2}\|a - b\|$$
- Using the notation from Nesterov, a convex β-smooth function f is denoted as $f \in C_\beta^{1,1}(\mathbb{R}^n)$

Sandwich theorem for β-smooth convex function

Theorem. If function $f \in C_\beta^{1,1}(\mathbb{R}^n)$, then $\forall x,\, y \in \mathbb{R}^n$

$$\frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|_2^2 \le |f(y) - f(x) - \nabla f(x)^\top (y-x)| \le \frac{\beta}{2}\|y-x\|_2^2$$

Interpretation

- Recall $f(x) + \nabla f(x)^\top (y-x) =$ 1st-order Taylor approximation of f at x, and $f(y) - f(x) - \nabla f(x)^\top (y-x)$ is the Bregman divergence (error made by the approximation)
- The theorem means: for a convex β-smooth f, the approximation error of it's 1st-order Taylor apporimxation at any point is bounded above and below (i.e. such function's 1st-order Taylor approximation will not be "too bad")
- In short: "for a convex β-smooth f, the Bregman divergence is bounded"

The idea of proving upper bound 1/2

The difference between $f(y)$ and $f(x)$ can be expressed as an integral:

$$f(y) - f(x) = \int_0^1 \nabla f\big(x + t(y-x)\big)^\top (y-x)dt$$

Explanation By the definition of *directional derivative*: for a scalar function $f(\mathbf{x}) = f(x_1, x_2, ..., x_n)$, the directional derivative, along a vector $\mathbf{v} = (v_1, v_2, ..., v_n)$, denoted as $\nabla_{\mathbf{v}} f(\mathbf{x})$ is

$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \lim_{t \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{t}$$

which can be also expressed as $\nabla_{\mathbf{v}} f(\mathbf{x}) = \nabla f(\mathbf{x})^\top \mathbf{v}$, where ∇ is gradient operator. (Note. $\mathbf{v}$ specifies the direction but not necessarily a unit vector). So the term $\nabla f\big(x + t(y-x)\big)^\top (y-x)$ represents the directional derivative of f at point x along the vector $y - x$.

The idea of proving upper bound - 2/2

Note that the value of $t \in [0, 1]$ in the integral, thus the expression

$$\int_0^1 \nabla f\big(x + t(y - x)\big)^\top (y - x) dt$$

means take the integral of gradient of f

- starting at point x (correspond to $t = 0$)
- integrates (moves) along the direction $y - x$
- until it reaches point y (where $t = 1$)

i.e., it moves along the close line segment $[x\ y]$.

Proving the upper bound - 1/2

Proof. Consider the integral with two new terms

$$f(y) - f(x) = \int_0^1 \Big[\nabla f\big(x + t(y-x)\big) + \nabla f(x) - \nabla f(x)\Big]^\top (y-x)dt$$

Take $\nabla f(x)$ out from the integral and move it to the left hand size:

$$f(y) - f(x) - \nabla f(x)^\top (y-x) = \int_0^1 \Big[\nabla f\big(x + t(y-x)\big) - \nabla f(x)\Big]^\top (y-x)dt$$

Take absolute value on both side

$$\begin{aligned} 0 &\leq \Big| f(y) - f(x) - \nabla f(x)^\top (y-x) \Big| \\ &= \left| \int_0^1 \Big[\nabla f\big(x + t(y-x)\big) - \nabla f(x)\Big]^\top (y-x)dt \right| \end{aligned}$$

By $|\int_a^b f(x)dx| \leq \int_a^b |f(x)|dx$ and $|ab| \leq |a||b|$ we have

$$\begin{aligned} 0 &\leq \Big| f(y) - f(x) - \nabla f(x)^\top (y-x) \Big| \\ &\leq \int_0^1 \|\nabla f\big(x + t(y-x)\big) - \nabla f(x)\| \cdot \|y - x\| dt \end{aligned}$$

Proving the upper bound - 2/2

f is β-smooth $\iff \|\nabla f\big(x + t(y-x)\big) - \nabla f(x)\| \leq \frac{\beta}{2}\|t(y-x)\|$, so

$$0 \leq \left|f(y) - f(x) - \nabla f(x)^\top (y-x)\right| \leq \int_0^1 \frac{\beta}{2}|t|\|y-x\|^2 dt = \frac{\beta}{2}\|y-x\|^2$$

f is convex $\iff$ the Bregman divergence $f(y) - f(x) - \nabla f(x)^\top (y-x)$ is non-negative so absolute sign can be removed

$$0 \leq f(y) - f(x) - \nabla f(x)^\top (y-x) \leq \frac{\beta}{2}\|y-x\|_2^2 \quad \square$$

The upper bound inequality is also called the Quadratic Upper bound.

Note. If f is only β-smooth but non-convex, the lower bound cannot be guaranteed without the absolute sign.

The idea of proving the lower bound

Now consider the lower bound. i.e., for a function f that is convex and β-smooth, we have

$$\frac{1}{2\beta}\|\nabla f(y) - \nabla f(x)\| \leq f(y) - f(x) - \nabla f(x)^\top (y - x)$$

for any two points x, $y \in \mathrm{dom} f$.

To prove this we need the following:

- For a convex f, we have $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$, for any two points x,$y \in \mathrm{dom} f$
- For a convex and β-smooth f, we have $f(y) - f(x) - \nabla f(x)^\top (y - x) \leq \dfrac{\beta}{2}\|y - x\|_2^2$ (the quadratic upper bound we just derived)

Proving the lower bound - 1/2

$$
\begin{aligned}
f(y) - f(x) &= f(z) - f(x) - [f(z) - f(y)] \\
&\overset{(1,2)}{\geq} \nabla f(x)^\top (z - x) - \left[\nabla f(y)^\top (z - y) + \frac{\beta}{2}\|z - y\|_2^2\right] \\
&= \nabla f(x)^\top (z - x) - \nabla f(y)^\top (z - y) - \frac{\beta}{2}\|z - y\|_2^2 \\
&\overset{(3)}{=} \nabla f(x)^\top (y - x + z - y) - \nabla f(y)^\top (z - y) - \frac{\beta}{2}\|z - y\|_2^2 \\
&= \nabla f(x)^\top (y - x) + \big(\nabla f(x) - \nabla f(y)\big)^\top (z - y) - \frac{\beta}{2}\|z - y\|_2^2
\end{aligned}
$$

(1) f is convex $f(z) \geq f(x) + \nabla f(x)^\top (z - x)$
(2) Quadratic upper bound inequality
(3) just a trick
Rearrange the last inequality we get

$$
\big(\nabla f(x) - \nabla f(y)\big)^\top (z - y) - \frac{\beta}{2}\|z - y\|_2^2 \leq f(y) - f(x) - \nabla f(x)^\top (y - x)
$$

Proving the lower bound - 2/2

The next step is very tricky: pick z as $z = y - \frac{1}{\beta}\big(\nabla f(y) - \nabla f(x)\big)$

we have

$$
\begin{aligned}
z - y &= -\frac{1}{\beta}\big(\nabla f(y) - \nabla f(x)\big) \\
\big(\nabla f(x) - \nabla f(y)\big)^\top (z - y) &= \frac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|_2^2 \\
-\frac{\beta}{2}\|z - y\|_2^2 &= -\frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|_2^2 \\
\big(\nabla f(x) - \nabla f(y)\big)^\top (z - y) - \frac{\beta}{2}\|z - y\|_2^2 &= \frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|^2
\end{aligned}
$$

Thus

$$\frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|^2 \le f(y) - f(x) - \nabla f(x)^\top (y - x)$$

$\square$

An application - Proving Coercivity of gradient

Consider the lower bound of the theorem on x, y and y, x

$$\frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|^2 \leq f(y) - f(x) - \nabla f(x)^\top (y - x)$$

$$\frac{1}{2\beta}\|\nabla f(y) - \nabla f(x)\|^2 \leq f(x) - f(y) - \nabla f(y)^\top (x - y)$$

Sum them up

$$\frac{1}{\beta}\|\nabla f(y) - \nabla f(x)\|^2 \leq \big(\nabla f(y) - \nabla f(x)\big)^\top (y - x)$$

this inequality is called coericivity of gradient.

Remark: apply Cauchy-Schwarz inequality $|ab| \leq |a||b|$ on coericivity gives the Lipschitz continuity of ∇f: $\|\nabla f(y) - \nabla f(x)\| \leq \beta \|y - x\|$.
So f is β-smooth $\iff$ ∇f is Lipschitz continuous $\iff$ ∇f is coercive.

Last page - summary

1. For a β-smooth f, for any two points x, $y \in \mathrm{dom} f$, we have

$$0 \le \left| f(y) - f(x) - \nabla f(x)^\top (y-x) \right| \le \frac{\beta}{2} \|y-x\|_2^2$$

2. Sandwich theorem for β-smooth convex function. For a convex and β-smooth function f, for any two points x, $y \in \mathrm{dom} f$, we have

$$\frac{1}{2\beta} \|\nabla f(x) - \nabla f(y)\|_2^2 \le f(y) - f(x) - \nabla f(x)^\top (y-x) \le \frac{\beta}{2} \|y-x\|_2^2$$

Intrepretation: the error of first order Taylor approximation is bounded.

3. Application: proving Coericivity of gradient

$$\frac{1}{\beta} \|\nabla f(y) - \nabla f(x)\|^2 \le \big(\nabla f(y) - \nabla f(x)\big)^\top (y-x)$$

End of document