If f has M-Lipschitz Hessian, then $\left| f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2} \big\langle \nabla^2 f(x)(y-x), (y-x) \big\rangle \right| \leq \frac{M}{6} \|y - x\|^3$

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft : December 16, 2018
Last update : December 16, 2018

The statement

For a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, if it has M-Lipschitz continuous Hessian, then

$$\|\nabla^2 f(y) - \nabla^2 f(x)\| \leq M\|y - x\|,$$

where $M > 0$. Here x, y are vectors in $\mathbb{R}^n$ and $\| \cot \|$ is L_2 norm.

If f is twice differentiable and has M-Lipschitz continuous Hessian, then we have the following inequality

$$\left| f(y) - f(x) - \langle \nabla f(x), y - x \rangle - \frac{1}{2} \langle \nabla^2 f(x)(y - x), (y - x) \rangle \right| \leq \frac{M}{6} \|y - x\|^3$$

This document : give the proof of this inequality.

This inequality come from Lemma 1.2.4 of : Yu. Nesterov, Introductory Lectures on Convex Optimization - A basic course. This document is to show the second part of the lemma.

The proof ... 1/6

As f has M-Lipschitz Hessian, we have

$$\|\nabla^2 f(a) - \nabla^2 f(b)\| \leq M\|a - b\|.$$

Put $a = x + \alpha(x - y)$, $b = x$ with $\alpha \in [0\ 1]$,

$$\begin{aligned} \|\nabla^2 f(x + \alpha(x - y)) - \nabla^2 f(x)\| &\leq M\|\alpha(x - y)\| \\ &\leq M|\alpha| \cdot \|x - y\| \\ &= M\alpha\|x - y\|, \end{aligned}$$

where in the last step we used the fact that $\alpha \geq 0$.

That is, we have

$$\|\nabla^2 f(x + \alpha(x - y)) - \nabla^2 f(x)\| \leq M\alpha\|x - y\|. \tag{1}$$

We will use (1) later.

The proof ... 2/6

As f is twice differentiable, we have

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \int_{\tau=0}^{\tau=1} \int_{\alpha=0}^{\alpha=\tau} \Big\langle \nabla^2 f\big(x + \alpha(y-x)\big)(y-x), y-x \Big\rangle d\alpha d\tau$$

Standard mathematics trick

$$\begin{aligned} f(y) = {} & f(x) + \langle \nabla f(x), y - x \rangle \\ & + \int_0^1 \int_0^\tau \Big\langle \Big(\nabla^2 f\big(x + \alpha(y-x)\big) + \nabla^2 f(x) - \nabla^2 f(x) \Big)(y-x), y-x \Big\rangle d\alpha d\tau \\ = {} & f(x) + \langle \nabla f(x), y - x \rangle + \int_0^1 \int_0^\tau \Big\langle \nabla^2 f(x)(y-x), y-x \Big\rangle d\alpha d\tau \\ & + \int_0^1 \int_0^\tau \Big\langle \Big(\nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \Big)(y-x), y-x \Big\rangle d\alpha d\tau \end{aligned}$$

Continue from last page :

$$
\begin{aligned}
f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \Big\langle \nabla^2 f(x)(y-x), y-x \Big\rangle \underbrace{\int_0^1 \int_0^\tau d\alpha d\tau}_{\frac{1}{2}} \\
&\quad + \int_0^1 \int_0^\tau \Big\langle \Big(\nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \Big)(y-x), y-x \Big\rangle d\alpha d\tau \\
&= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \Big\langle \nabla^2 f(x)(y-x), y-x \Big\rangle \\
&\quad + \int_0^1 \int_0^\tau \Big\langle \Big(\nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \Big)(y-x), y-x \Big\rangle d\alpha d\tau
\end{aligned}
$$

The proof ... 4/6

Continue from last page, we have

$$f(y) - f(x) - \langle \nabla f(x), y - x \rangle + \frac{1}{2}\Big\langle \nabla^2 f(x)(y-x), y-x \Big\rangle$$
$$= \int_0^1 \int_0^\tau \Big\langle \Big(\nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \Big)(y-x), y-x \Big\rangle d\alpha d\tau$$

Then we have

$$\Big| f(y) - f(x) - \langle \nabla f(x), y - x \rangle + \frac{1}{2}\Big\langle \nabla^2 f(x)(y-x), y-x \Big\rangle \Big|$$
$$= \left| \int_0^1 \int_0^\tau \Big\langle \Big(\nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \Big)(y-x), y-x \Big\rangle d\alpha d\tau \right|$$

That is, to prove the inequality, we can show the absolute value of the integral is upper bounded by $\dfrac{M}{6}\|y-x\|^3$, which is the main focus of the next few slides.

The proof ... 5/6

Using the inequality of absolute value of integral, we have

$$\left|\int_0^1\int_0^\tau \left\langle \left(\nabla^2 f\big(x+\alpha(y-x)\big)-\nabla^2 f(x)\right)(y-x), y-x\right\rangle d\alpha d\tau\right|$$
$$\leq \int_0^1\int_0^\tau \left|\left\langle \left(\nabla^2 f\big(x+\alpha(y-x)\big)-\nabla^2 f(x)\right)(y-x), y-x\right\rangle\right| d\alpha d\tau$$
$$\leq \int_0^1\int_0^\tau \left\|\nabla^2 f\big(x+\alpha(y-x)\big)-\nabla^2 f(x)\right\| d\alpha d\tau \cdot \|y-x\|^2$$

We can now use (1) to bound $\left\|\nabla^2 f\big(x+\alpha(y-x)\big)-\nabla^2 f(x)\right\|$ inside the integral

The proof ... 6/6

Apply (1) we have

$$\int_0^1 \int_0^\tau \left\| \nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \right\| d\alpha d\tau \cdot \|y-x\|^2$$

$$\leq \int_0^1 \int_0^\tau \left| M\alpha \|x-y\| \right| d\alpha d\tau \cdot \|y-x\|^2$$

As $M, \alpha > 0$, we can remove the absolute value sign. Using $\|x-y\| = \|y-x\|$, we have

$$\int_0^1 \int_0^\tau \left\| \nabla^2 f\big(x + \alpha(y-x)\big) - \nabla^2 f(x) \right\| d\alpha d\tau \cdot \|y-x\|^2$$

$$\leq \underbrace{\int_0^1 \int_0^\tau \alpha d\alpha d\tau}_{\frac{1}{6}} \cdot M\|y-x\|^3 = \frac{M}{6}\|y-x\|^3 \quad \square$$

End of document