Monotonictiy of gradient of α-strongly convex β-smooth 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 25, 2019

Convex, β-smooth and α-strongly convex function

A function $f(x)$ is **convex** if dom$f$ is convex and for all x, $y \in$ dom$f$:

- $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$
- $\langle x - y, \nabla f(x) - \nabla f(y) \rangle \geq 0$
- $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle$

A function $f(x)$ is β **-smooth** if for any two points x, $y \in$ dom f:

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

where $\beta > 0$

A function $f(x)$ is α**-strongly convex** if dom$f$ is convex and for all x, $y \in$ dom$f$:

- $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{\alpha}{2} \lambda (1-\lambda) \|x - y\|_2^2$
- $\langle x - y, \nabla f(x) - \nabla f(y) \rangle \geq \alpha \|x - y\|_2^2$
- $f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\alpha}{2} \|x - y\|_2^2$
- $f(x) - \frac{\alpha}{2} \|x\|_2^2$ is convex

where $\alpha > 0$

Monotonictiy of gradient of α-strongly convex β-smooth function

Theorem. If a function f is α-strongly convex and β-smooth, then for any two points x, $y \in \mathrm{dom} f$,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \frac{\alpha\beta}{\alpha+\beta}\|x-y\|_2^2 + \frac{1}{\alpha+\beta}\|\nabla f(x) - \nabla f(y)\|_2^2$$

Proof : Consider $\phi(x) = f(x) - \frac{\alpha}{2}\|x\|_2^2$. The goal is to show

- $\phi(x)$ is convex
- $\phi(x)$ is $(\beta - \alpha)$-smooth and thus having coerive gradient

For point 1, it is trivial : as f is α-strongly convex, hence $f(x) - \frac{\alpha}{2}\|x\|_2^2$ is convex. Done.

To show $\phi(x)$ is $(\beta-\alpha)$-smooth

Goal: show $\phi(y)-\phi(x)-\nabla\phi(x)^T(y-x)-\dfrac{\beta-\alpha}{2}\|y-x\|_2^2 \le 0$

$$
\begin{aligned}
& \phi(y)-\phi(x)-\nabla\phi(x)^T(y-x)-\frac{\beta-\alpha}{2}\|y-x\|_2^2 \\
= \quad & f(y)-\frac{\alpha}{2}\|y\|_2^2-f(x)+\frac{\alpha}{2}\|x\|_2^2-\nabla\phi(x)^T(y-x)-\frac{\beta-\alpha}{2}\|y-x\|_2^2 \\
= \quad & f(y)-f(x)+\frac{\alpha}{2}(\|x\|_2^2-\|y\|_2^2)-\nabla\phi(x)^T(y-x)-\frac{\beta-\alpha}{2}\|y-x\|_2^2 \\
\le \quad & \nabla f(x)^T(y-x)+\frac{\alpha}{2}(\|x\|_2^2-\|y\|_2^2)-\nabla\phi(x)^T(y-x)+\frac{\alpha}{2}\|y-x\|_2^2
\end{aligned}
$$

Remark. f is β-smooth so $f(y)-f(x)\le\nabla f(x)^T(y-x)+\dfrac{\beta}{2}\|y-x\|_2^2$ and the term $\dfrac{\beta}{2}\|y-x\|_2^2$ cancels with $\dfrac{\beta-\alpha}{2}\|y-x\|_2^2$.

To show $\phi(x)$ is $(\beta - \alpha)$-smooth ... 2

$\phi(x) = f(x) - \frac{\alpha}{2}\|x\|_2^2 \implies \nabla\phi(x) = \nabla f(x) - \alpha x$, thus

$$
\begin{aligned}
& \nabla f(x)^T(y-x) + \frac{\alpha}{2}(\|x\|_2^2 - \|y\|_2^2) - \nabla\phi(x)^T(y-x) + \frac{\alpha}{2}\|y-x\|_2^2 \\
= \; & \nabla f(x)^T(y-x) + \frac{\alpha}{2}(\|x\|_2^2 - \|y\|_2^2) - (\nabla f(x) - \alpha x)^T(y-x) + \frac{\alpha}{2}\|y-x\|_2^2 \\
= \; & \frac{\alpha}{2}(\|x\|_2^2 - \|y\|_2^2) + \alpha x^T(y-x) + \frac{\alpha}{2}\|y-x\|_2^2
\end{aligned}
$$

expand everything

$$\frac{\alpha}{2}(\|x\|_2^2 - \|y\|_2^2) + \alpha x^T y - \alpha\|x\|_2^2 + \frac{\alpha}{2}(\|y\|_2^2 - 2x^T y + \|x\|_2^2) = 0$$

So $\phi(y) - \phi(x) - \nabla\phi(x)^T(y-x) - \frac{\beta-\alpha}{2}\|y-x\|_2^2 \leq 0$ $\square$

Monotonictiy of gradient of α-strongly convex β-smooth function

(1) ϕ is convex, (2) ϕ is $(\beta-\alpha)$-smooth, by theorem of convex-smooth function, $\phi(x)$ has coerive gradient:

$$(\nabla\phi(x)-\nabla\phi(y))^T(x-y) \geq \frac{1}{\beta-\alpha}\|\nabla\phi(x)-\nabla\phi(y)\|_2^2$$

Now put $\nabla\phi(x)=\nabla f(x)-\alpha x$

$$\left(\nabla f(x)-\nabla f(y)-\alpha(x-y)\right)^T(x-y) \geq \frac{1}{\beta-\alpha}\|\nabla f(x)-\nabla f(y)-\alpha(x-y)\|_2^2$$

The right hand side can be expanded as

$$\frac{\|\nabla f(x)-\nabla f(y)\|_2^2-2\alpha(\nabla f(x)-\nabla f(y))^T(x-y)+\alpha^2\|x-y\|_2^2}{\beta-\alpha}$$

Left hand side can be expanded as

$$\left(\nabla f(x)-\nabla f(y)\right)^T(x-y)-\alpha\|x-y\|_2^2$$

Grouping the similar terms yields

$$\left(1+\frac{2\alpha}{\beta-\alpha}\right)\left(\nabla f(x)-\nabla f(y)\right)^T(x-y) \geq \frac{\|\nabla f(x)-\nabla f(y)\|_2^2}{\beta-\alpha}+\left(\frac{\alpha^2}{\beta-\alpha}+\alpha\right)\|x-y\|_2^2$$

Divide the whole equation by $\left(1+\frac{2\alpha}{\beta-\alpha}\right)$ gives the result. $\square$