Monotonictiy of gradient of strongly convex smooth function

Andersen Ang

ECS, Uni. Southampton, UK
andersen.ang@soton.ac.uk

Homepage angms.science

Version: April 6, 2023
First draft: June 6, 2017

Content

$\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2$ is $(L-m)$-smooth

$$\Big\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \Big\rangle \;\geq\; \frac{mL}{m+L}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{1}{m+L}\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2.$$

Recall of convex, smooth and strongly convex function (details here)

- Given $f : \mathbb{R}^n \to \mathbb{R}^n$ with $\mathrm{dom} f = \mathbb{R}^n$ is a convex set
- f is **convex** if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom} f$
  - $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y})$ — Jansen
  - $\langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \rangle \geq 0$ — grad is monotone
  - $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle$ — linear global under-estimator
- f is L-**smooth** $(L > 0)$ if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom} f$
  - $\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\| \leq L \|\boldsymbol{x} - \boldsymbol{y}\|$ — ∇f is L-Lipschitz
  - $f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \dfrac{L}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ — global quadratic over-estimator
- f is m-**strongly convex** $(m > 0)$ if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathrm{dom} f$
  - $f(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda f(\boldsymbol{x}) + (1-\lambda) f(\boldsymbol{y}) - \dfrac{m}{2} \lambda (1-\lambda) \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ — Jansen with a quadratic term
  - $\langle \boldsymbol{x} - \boldsymbol{y}, \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}) \rangle \geq m \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ — grad is strongly monotone
  - $f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle + \dfrac{m}{2} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$ — quadratic global under-estimator
  - $f(\boldsymbol{x}) - \dfrac{m}{2} \|\boldsymbol{x}\|_2^2$ is convex
- This PDF: prove if f is m-strongly convex and L-smooth, then

$$\Big\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \Big\rangle \;\geq\; \frac{mL}{m+L} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \frac{1}{m+L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2.$$

Lemma: $\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2$ is $(L-m)$-smooth

- Goal: show $\phi(\boldsymbol{y}) - \phi(\boldsymbol{x}) - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) - \frac{L-m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \le 0$
- By $\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2$:

$$
\begin{aligned}
& \phi(\boldsymbol{y}) - \phi(\boldsymbol{x}) - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) - \frac{L-m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \\
\overset{\phi = f - \frac{m}{2}\|\cdot\|_2^2}{=} \quad & f(\boldsymbol{y}) - \frac{m}{2}\|\boldsymbol{y}\|_2^2 - f(\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{x}\|_2^2 - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) - \frac{L-m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \\
= \quad & f(\boldsymbol{y}) - f(\boldsymbol{x}) + \frac{m}{2}\left(\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}\|_2^2\right) - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) - \frac{L-m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \\
\overset{(1)}{\le} \quad & \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\left(\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}\|_2^2\right) - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2
\end{aligned}
$$

(1): f is L-smooth so $f(\boldsymbol{y}) - f(\boldsymbol{x}) \le \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2$ and $\frac{L}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2$ cancels with $\frac{L-m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2$.

- Our goal is now showing

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

To show $\phi(x)$ is $(L-m)$-smooth ... 2

- Goal: show $\nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\Big(\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}\|_2^2\Big) - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \ \leq \ 0$

- By $\phi(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{m}{2}\|\boldsymbol{x}\|_2^2 \implies \nabla\phi(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) - m\boldsymbol{x}$

$$\begin{aligned} & \nabla f(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\Big(\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}\|_2^2\Big) - \Big(\nabla f(\boldsymbol{x}) - m\boldsymbol{x}\Big)^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \\ = \ & \underbrace{\frac{m}{2}\Big(\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{y}\|_2^2\Big) + m\boldsymbol{x}^\top(\boldsymbol{y}-\boldsymbol{x}) + \frac{m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2}_{=0} \end{aligned}$$

So $\phi(\boldsymbol{y}) - \phi(\boldsymbol{x}) - \nabla\phi(\boldsymbol{x})^\top(\boldsymbol{y}-\boldsymbol{x}) - \frac{L-m}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 \leq 0$ $\square$

We need: L-smooth function has co-coercive gradient

- If f is L-smooth, then

$$\left\langle \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle \;\geq\; \frac{1}{\beta} \|\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x})\|_2^2$$

  Details p.10 here

- ϕ is $(L-m)$-smooth, so

$$\left\langle \nabla \phi(\boldsymbol{y}) - \nabla \phi(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle \;\geq\; \frac{1}{L-m} \|\nabla \phi(\boldsymbol{y}) - \nabla \phi(\boldsymbol{x})\|_2^2 \qquad (\#)$$

  What's next: use $(\#)$ to prove

$$\left\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \right\rangle \;\geq\; \frac{mL}{m+L} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \frac{1}{m+L} \|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2.$$

Monotonicity of gradient of m-strongly convex L-smooth function

$$\left\langle \nabla\phi(\boldsymbol{y}) - \nabla\phi(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle \;\geq\; \frac{1}{L-m}\|\nabla\phi(\boldsymbol{y}) - \nabla\phi(\boldsymbol{x})\|_2^2 \qquad (\#)$$

- $\nabla\phi(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) - m\boldsymbol{x}$

$$\left\langle \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}) - m(\boldsymbol{y}-\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle \;\geq\; \frac{1}{L-m}\|\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}) - m(\boldsymbol{y}-\boldsymbol{x})\|_2^2$$

  - Left hand side $= \left\langle \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle - m\|\boldsymbol{y}-\boldsymbol{x}\|_2^2$
  - Right hand size $= \dfrac{\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 - 2m\left\langle \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle + m^2\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}{L-m}$

- Grouping the similar terms yields

$$\left(1 + \frac{2m}{L-m}\right)\left\langle \nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \right\rangle \;\;\geq\;\; \frac{\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2}{L-m} + \left(\frac{m^2}{L-m} + m\right)\|\boldsymbol{x}-\boldsymbol{y}\|_2^2$$

  Divide the whole equation by $\left(1 + \dfrac{2m}{L-m}\right)$ gives

$$\left\langle \nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y} \right\rangle \;\geq\; \frac{mL}{m+L}\|\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \frac{1}{m+L}\|\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{y})\|_2^2 \quad \square$$