Convergence rate of gradient descent algorithm on β-smooth α-strongly convex function

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

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

First draft: August 2, 2017
Last update : July 27, 2019

Overview

1. β-smooth α-strongly convex function
2. Convergence rate of gradient descent algorithm for β-smooth α-strongly convex function
3. Convergence rate of GD for β-smooth α-strongly convex function with constant step size
4. Convergence rate of GD for β-smooth α-strongly convex function with constant step size with $\beta \gg \alpha$
5. Summary

Convex, β-smooth and α-strongly convex function

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

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

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

- $f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y)$
- $(x - y)^T\big(\nabla f(x) - \nabla f(y)\big) \geq 0$
- $f(y) \geq f(x) + \nabla f(x)^T(y - x)$

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

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

For details : see here.

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

Theorem 1. If a function f is α-strongly convex and β-smooth, then for any two points $x, y \in \text{dom} f$,

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

For details : see here.

Convergence rate of gradient descent algorithm for β-smooth α-strongly convex function

Theorem 1. If f is β-smooth and α-strongly convex, GD algorithm $x_{k+1} = x_k - t_k \nabla f(x_k)$ with step sizes $t_k \leq \frac{2}{\alpha + \beta}$ satisfies

$$f(x_k) - f^* \leq \frac{\beta}{2} \prod_{i=0}^{k-1} \left(1 - 2t_i \frac{\alpha\beta}{\alpha + \beta}\right) \|x_0 - x^*\|_2^2$$

Proof: f is β-smooth so $f(x_k) - f^* \leq \nabla f(x^*)(x_k - x^*) + \frac{\beta}{2}\|x_k - x^*\|_2^2$.
By 1st-order optimality condition $\nabla f(x^*) = 0$, so

$$f(x_k) - f^* \leq \frac{\beta}{2}\|x_k - x^*\|_2^2 \qquad (*)$$

Apply the GD update formula $x_k = x_{k-1} - t_{k-1}\nabla f(x_{k-1})$

$$\begin{aligned} \|x_k - x^*\|_2^2 &= \|x_{k-1} - t_{k-1}\nabla f(x_{k-1}) - x^*\|_2^2 \\ &= \|(x_{k-1} - x^*) - t_{k-1}\nabla f(x_{k-1})\|_2^2 \\ &= \|x_{k-1} - x^*\|_2^2 - 2t_{k-1}\nabla f(x_{k-1})^T(x_{k-1} - x^*) + t_{k-1}^2\|\nabla f(x_{k-1})\|_2^2 \quad (1) \end{aligned}$$

f is β-smooth and α-strongly convex, by theorem 1, $\forall x, y \in \mathrm{dom} f$:

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

Put $x = x_{k-1}$ and $y = x^*$ with $\nabla f(x^*) = 0$

$$\nabla f(x_{k-1})^T (x_{k-1} - x^*) \geq \frac{\alpha\beta}{\alpha+\beta}\|x_{k-1} - x^*\|_2^2 + \frac{1}{\alpha+\beta}\|\nabla f(x_{k-1})\|_2^2$$

Thus

$$-\nabla f(x_{k-1})^T (x_{k-1} - x^*) \leq -\frac{\alpha\beta}{\alpha+\beta}\|x_{k-1} - x^*\|_2^2 - \frac{1}{\alpha+\beta}\|\nabla f(x_{k-1})\|_2^2$$

$$-2t_{k-1}\nabla f(x_{k-1})^T (x_{k-1} - x^*) \leq 2t_{k-1}\left[\frac{\alpha\beta}{\alpha+\beta}\|x_{k-1} - x^*\|_2^2 - \frac{1}{\alpha+\beta}\|\nabla f(x_{k-1})\|_2^2\right]$$

(1) becomes

$$\begin{aligned} \|x_k - x^*\|_2^2 &\leq \|x_{k-1} - x^*\|_2^2 - 2t_{k-1}\nabla f(x_{k-1})^T (x_{k-1} - x^*) + t_{k-1}^2\|\nabla f(x_{k-1})\|_2^2 \\ &\leq \left[1 - 2t_{k-1}\frac{\alpha\beta}{\alpha+\beta}\right]\|x_{k-1} - x^*\|_2^2 + \left[t_{k-1}^2 - \frac{2t_{k-1}}{\alpha+\beta}\right]\|\nabla f(x_{k-1})\|_2^2 \end{aligned}$$

As $t_k \leq \frac{2}{\alpha+\beta}$ $\forall k$ so $t_{k-1}^2 - \frac{2t_{k-1}}{\alpha+\beta} = t_{k-1}\left(t_{k-1} - \frac{2}{\alpha+\beta}\right) \leq 0$ and

$$\left[t_{k-1}^2 - \frac{2t_{k-1}}{\alpha+\beta}\right] \|\nabla f(x_{k-1})\|_2^2 \leq 0$$

or equivalently

$$0 \leq -\left[t_{k-1}^2 - \frac{2t_{k-1}}{\alpha+\beta}\right] \|\nabla f(x_{k-1})\|_2^2$$

Add this to

$$\|x_k - x^*\|_2^2 \leq \left[1 - 2t_{k-1}\frac{\alpha\beta}{\alpha+\beta}\right] \|x_{k-1} - x^*\|_2^2 + \left[t_{k-1}^2 - \frac{2t_{k-1}}{\alpha+\beta}\right] \|\nabla f(x_{k-1})\|_2^2$$

gives

$$\|x_k - x^*\|_2^2 \leq \left[1 - 2t_{k-1}\frac{\alpha\beta}{\alpha+\beta}\right] \|x_{k-1} - x^*\|_2^2$$

$$
\begin{aligned}
\|x_k - x^*\|_2^2 &\le \left[1 - 2t_{k-1}\frac{\alpha\beta}{\alpha+\beta}\right]\|x_{k-1} - x^*\|_2^2 \\
&\le \left[1 - 2t_{k-1}\frac{\alpha\beta}{\alpha+\beta}\right]\left[1 - 2t_{k-2}\frac{\alpha\beta}{\alpha+\beta}\right]\|x_{k-2} - x^*\|_2^2 \\
&\le \vdots \\
&\le \prod_{i=0}^{k-1}\left[1 - 2t_i\frac{\alpha\beta}{\alpha+\beta}\right]\|x_0 - x^*\|_2^2
\end{aligned}
$$

Put this back to (*) in the beginning

$$
f(x_k) - f(x^*) \le \frac{\beta}{2}\prod_{i=0}^{k-1}\left[1 - 2t_i\frac{\alpha\beta}{\alpha+\beta}\right]\|x_0 - x^*\|_2^2 \quad \square
$$

Note. The requirement on step sizes is $t_k \le \dfrac{2}{\alpha+\beta}$, hence the theorem holds for any step size sequence fulfilling this condition.

Convergence rate of GD for β-smooth α-strongly convex function with constant step size

Theorem 2. If f is β-smooth and α-strongly convex, GD algorithm $x_{k+1} = x_k - t_k \nabla f(x_k)$ with constant step size $t_k = \frac{2}{\alpha+\beta}$ satisfies

$$f(x_k) - f^* \leq \frac{\beta}{2}\left(\frac{Q-1}{Q+1}\right)^{2k} \|x_0 - x^*\|_2^2$$

where $Q = \frac{\beta}{\alpha} \geq 1$ is the *conditional number*

Proof : Consider theorem 1 with $t_k = t$ for all k,

$$f(x_k) - f^* \quad \leq \quad \frac{\beta}{2}\left[1 - 2t\frac{\alpha\beta}{\alpha+\beta}\right]^k \|x_0 - x^*\|_2^2$$

Put $t = \frac{2}{\alpha+\beta}$ and consider the bracket

$$\begin{aligned} 1 - 2t\frac{\alpha\beta}{\alpha+\beta} &= 1 - \frac{4\alpha\beta}{(\alpha+\beta)^2} = \frac{(\alpha+\beta)^2}{(\alpha+\beta)^2} - \frac{4\alpha\beta}{(\alpha+\beta)^2} = \frac{\alpha^2 - 2\alpha\beta + \beta^2}{(\alpha+\beta)^2} \\ &= \left[\frac{\alpha-\beta}{\alpha+\beta}\right]^2 = \left[\frac{Q-1}{Q+1}\right]^2 \quad \square \end{aligned}$$

Convergence rate of GD for β-smooth α-strongly convex function with constant step size with $\beta \gg \alpha$

Theorem 3. If f is β-smooth and α-strongly convex, and $\beta \gg \alpha$ (so $Q = \dfrac{\beta}{\alpha} \gg 1$), GD with constant step size $t_k = \dfrac{2}{\alpha + \beta}$ satisfies

$$f(x_k) - f^* \leq \frac{\beta}{2} \exp\left(-\frac{4k}{Q+1}\right) \|x_0 - x^*\|_2^2$$

Proof: from theorem 2, we have

$$f(x_k) - f^* \quad \leq \quad \frac{\beta}{2}\left(\frac{Q-1}{Q+1}\right)^{2k} \|x_0 - x^*\|_2^2$$

Now $\frac{Q-1}{Q+1} = \frac{Q+1-2}{Q+1} = 1 - \frac{2}{Q+1}$. As $\beta \gg \alpha \implies Q \gg 1 \implies \frac{2}{Q+1}$ very small $\implies 1 - \frac{2}{Q+1} \leq \exp\left(-\frac{2}{Q+1}\right)$ $(1 - x \leq e^{-x}$ for small $x)$.
Hence consider the bracket,

$$\left(\frac{Q-1}{Q+1}\right)^{2k} \leq \exp\left(-\frac{2}{Q+1}\right)^{2k} = \exp\left(-\frac{4k}{Q+1}\right) \quad \square$$

Last page - summary

Theorem 1 If f is β-smooth α-strongly convex, GD update $x_{k+1} = x_k - t_k \nabla f(x_k)$ with $t_k \leq \frac{2}{\alpha+\beta}$ satisfies

$$f(x_k) - f^* \leq \frac{\beta}{2} \prod_{i=0}^{k-1} \left[1 - 2t_i \frac{\alpha\beta}{\alpha+\beta}\right] \|x_0 - x^*\|_2^2$$

Theorem 2 If constant step size $t_k = \frac{2}{\alpha+\beta}$ then

$$f(x_k) - f^* \leq \frac{\beta}{2} \left(\frac{Q-1}{Q+1}\right)^{2k} \|x_0 - x^*\|_2^2$$

where $Q = \frac{\beta}{\alpha} \geq 1$ is the *conditional number*
Theorem 3 If constant step size $t_k = \frac{2}{\alpha+\beta}$ and $\beta \gg \alpha$ then

$$f(x_k) - f^* \leq \frac{\beta}{2} \exp\left(-\frac{4k}{Q+1}\right) \|x_0 - x^*\|_2^2$$

End of document