Strong Convexity and Gradient Descent

Ruixin Guo

Department of Computer Science
Kent State University

May 23, 2023

Contents

Recall

Convex Function: Let $f : C \to \mathbb{R}$ be a convex function where $C \subseteq \mathbb{R}^d$ is a convex set. Then for any $x, y \in C$ and $\lambda \in [0, 1]$,
(1) $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$
(2) $f(x) + \nabla f(x)(y - x) \leq f(y)$

Lipschitz Smooth: Let $X \subseteq \mathbb{R}^d$. A function $f : X \to \mathbb{R}$ is L-smooth if for any $x, y \in X$,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

Especially, when X is convex, we have

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

Gradient Descent: Let $C \subseteq \mathbb{R}^d$ be a convex set and $f : C \to \mathbb{R}$ be a convex function. Let x^* be the minimizer of $\min_{x \in C} f(x)$. Let $x_0 \in C$ be a start point, the gradient descent algorithm makes iteration by $x_{k+1} = x_k - t\nabla f(x_k)$ for $k = 0, 1, 2 \ldots$.
When f is L-smooth and $t < 1/L$, the sequence $\{f(x_k)\}$ will converge to $f(x^*)$ by

$$f(x_k) - f(x^*) \leq \frac{\|x_0 - x^*\|^2}{2tk}$$

which is sublinear in the worst case.

A Lemma for L-Smooth function

Lemma 1: If $f : C \to \mathbb{R}$ is L-smooth and $\lambda > 0$ then for all $x, y \in C$,

$$f(x - \lambda \nabla f(x)) - f(x) \leq -\lambda \left(1 - \frac{\lambda L}{2}\right) \|\nabla f(x)\|^2$$

If moreover $\inf f > -\infty$, then for all $x \in C$,

$$\frac{1}{2L} \|\nabla f(x)\|^2 \leq f(x) - \inf f$$

Proof: Since f is L-smooth, for any $x, y \in C$,

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

Let $y = x - \lambda \nabla f(x)$, then

$$\begin{aligned} f(x - \lambda \nabla f(x)) &\leq f(x) - \lambda \langle \nabla f(x), \nabla f(x) \rangle + \frac{L\lambda^2}{2} \|\nabla f(x)\|^2 \\ &= f(x) - \lambda(1 - \frac{L\lambda}{2}) \|\nabla f(x)\|^2 \end{aligned}$$

A Lemma for L-Smooth function

Let $\lambda = 1/L$, we have

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

Assume $\inf f > -\infty$, then

$$\begin{aligned} f(x) - \inf f &\geq f(x) - f(x - \frac{1}{L}\nabla f(x)) \\ &\geq f(x) - \left(f(x) - \frac{1}{2L}\|\nabla f(x)\|^2\right) \\ &= \frac{1}{2L}\|\nabla f(x)\|^2 \end{aligned}$$

Contents

Strong Convexity

Definition (Strong Convexity): Let $C \subseteq \mathbb{R}^d$ be a convex set and $f : C \to \mathbb{R}$ be a convex function. We say that f is μ-strongly convex if for every $x, y \in C$ and any $t \in [0, 1]$ we have

$$\mu \frac{t(1-t)}{2} \|x - y\|^2 + f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

The above definition shows for $t \in (0, 1)$,

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

which means the segment $\lambda f(x) + (1-\lambda) f(y)$ is strictly above the curve $f(\lambda x + (1-\lambda)y)$ except on the endpoints.

Strong Convexity

Lemma 2: Let $C \subseteq \mathbb{R}^d$ be a convex set, $f : C \to \mathbb{R}$ be a convex function and $\mu > 0$. The function f is μ-strongly convex if and only if there exists a convex function $g : C \to \mathbb{R}$ such that $f(x) = g(x) + \frac{\mu}{2}\|x\|^2$.

Proof: Given f and μ, define $g(x) = f(x) - \frac{\mu}{2}\|x\|^2$. Note $z_t = (1-t)x + ty$. If f is μ-strongly convex, then for every $x, y \in C$ and $t \in [0,1]$,

$$f(z_t) + \frac{\mu}{2}t(1-t)\|x-y\|^2 \leq (1-t)f(x) + tf(y) \quad \Longleftrightarrow$$
$$g(z_t) + \frac{\mu}{2}\|z_t\|^2 + \frac{\mu}{2}t(1-t)\|x-y\|^2 \leq (1-t)g(x) + tg(y) + (1-t)\frac{\mu}{2}\|x\|^2 + t\frac{\mu}{2}\|y\|^2$$

The second inequality implies $g(z_t) \leq (1-t)g(x) + tg(y)$ which means g is convex. This is because all the terms containing $\frac{\mu}{2}$ can be cancelled:

$$\begin{aligned}
&\|z_t\|^2 + t(1-t)\|x-y\|^2 - (1-t)\|x\|^2 - t\|y\|^2 \\
&= (1-t)^2\|x\|^2 + t^2\|y\|^2 + 2t(1-t)\langle x, y\rangle + t(1-t)\|x\|^2 + t(1-t)\|y\|^2 \\
&\quad - 2t(1-t)\langle x, y\rangle - (1-t)\|x\|^2 - t\|y\|^2 \\
&= 0
\end{aligned}$$

Strong Convexity

Lemma 3: Let $C \subseteq \mathbb{R}^d$ be a convex set, $f : C \to \mathbb{R}$ be a convex function and $\mu > 0$. If f is μ-strongly convex, then for any $x, y \in C$,

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

Proof: Since f is μ-strongly convex, let $f(x) = g(x) + \frac{\mu}{2}\|x\|^2$, where g is a convex function by Lemma 2. Taking derivative with respect to x on both sides, we have $\nabla f(x) = \nabla g(x) + \mu x$. Since g is convex, for any $x, y \in C$ we have $g(y) - g(x) - \langle \nabla g(x), y - x \rangle \geq 0$. Thus,

$$\begin{aligned}
&f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\
&= g(y) + \frac{\mu}{2}\|y\|^2 - g(x) - \frac{\mu}{2}\|x\|^2 - \langle \nabla g(x) + \mu x, y - x \rangle \\
&= g(y) - g(x) - \langle \nabla g(x), y - x \rangle + \frac{\mu}{2}\|y\|^2 - \frac{\mu}{2}\|x\|^2 - \langle \mu x, y - x \rangle \\
&\geq \frac{\mu}{2}\|y\|^2 - \frac{\mu}{2}\|x\|^2 - \langle \mu x, y - x \rangle \\
&= \frac{\mu}{2}\|y\|^2 + \frac{\mu}{2}\|x\|^2 - \mu\langle x, y \rangle = \frac{\mu}{2}\|y - x\|^2
\end{aligned}$$

Contents

Gradient Descent Convergence Analysis

Theorem 1: Let $C \subseteq \mathbb{R}^d$ be a convex set, $f : C \to \mathbb{R}$ be a L-smooth and μ-strongly convex function and $x^* = \arg\min_x f(x)$. Then the Gradient Descent Iteration

$$x_{k+1} = x_k - t\nabla f(x_k)$$

with step size $t \leq 1/L$ satisfies the following:

$$\|x_{k+1} - x^*\|^2 \leq (1 - t\mu)^{k+1} \|x_0 - x^*\|^2$$

Proof:

$$\begin{aligned}
&\|x_{k+1} - x^*\|^2 = \|x_k - x^* - t\nabla f(x_k)\|^2 \\
&= \|x_k - x^*\|^2 - 2t\langle \nabla f(x_k), x_k - x^* \rangle + t^2 \|\nabla f(x_k)\|^2 \\
&\leq (1 - t\mu)\|x_k - x^*\|^2 - 2t(f(x_k) - f(x^*)) + t^2 \|\nabla f(x_k)\|^2 \qquad \text{[by Lemma 3]} \\
&\leq (1 - t\mu)\|x_k - x^*\|^2 - 2t(f(x_k) - f(x^*)) + 2Lt^2(f(x_k) - f(x^*)) \quad \text{[by Lemma 1]} \\
&= (1 - t\mu)\|x_k - x^*\|^2 - 2t(1 - Lt)(f(x_k) - f(x^*)) \\
&\leq (1 - t\mu)\|x_k - x^*\|^2 \qquad [Lt \leq 1]
\end{aligned}$$

Thus,

$$\|x_{k+1} - x^*\|^2 \leq (1 - t\mu)^{k+1} \|x_0 - x^*\|^2$$

Convergence Speed of Gradient Descent

When f is L-smooth and μ-strongly convex and $L \geq \mu > 0$, we have

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq (1 - t\mu)^{1/2} < 1$$

which means the gradient descent converges linearly in the worst case.

References

[1] Handbook of Convergence Theorems for (Stochastic) Gradient Methods. https://gowerrobert.github.io/pdf/M2_statistique_optimisation/grad_conv.pdf. (Important!)

[2] Gradient Descent: Convergence Analysis. https://www.stat.cmu.edu/~ryantibs/convexopt-F13/scribes/lec6.pdf.

[3] Rate of Convergence. https://en.wikipedia.org/wiki/Rate_of_convergence.

[4] Linear Convergence https://www.math.drexel.edu/~tolya/linearconvergence.pdf

[5] Proving that a Strongly Convex Function has at most 1 minimum. https://math.stackexchange.com/questions/337090/if-f-is-strictly-convex-in-a-convex-set-show-it-has-no-more-than-1-minimum