The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Convex Optimization — Fall 2019

Lecture 24

Aryan Mokhtari Monday, November 25, 2019.

Goal: In this lecture we talk about strong convexity and smoothness and their side results.

1 Unconstrained optimization

In this lecture (and the following lectures), we focus on studying iterative methods for solving unconstrained convex problems. An important property of this class of problems is that

$$\hat{\mathbf{x}} \text{ is a global minimum if and only if } \nabla f(\hat{\mathbf{x}}) = 0.$$

In this lecture, we focus on minimizing a strongly convex function and characterize the number of iterations for solving such problems.

2 Strong convexity

Definition 1. *If there exists a constant $m > 0$ such that*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{m}{2} \|\mathbf{y} - \mathbf{x}\|^2 \tag{1}$$

*for all $\mathbf{x}, \mathbf{y} \in S$, then the function f is m-*strongly convex *on S.*

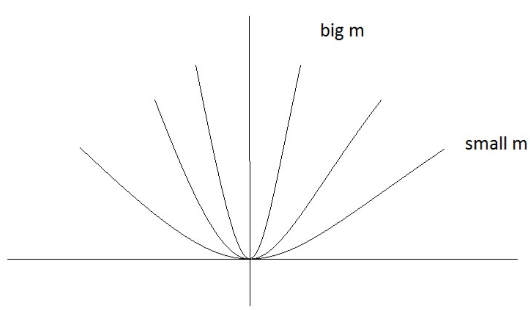

Figure 1: A strongly convex function with different parameter m. The larger m is, the steeper the function looks like.

When $m = 0$, we recover the basic inequality characterizing convexity; for $m > 0$, we obtain a better lower bound on $f(\mathbf{y})$ than that from convexity alone. The value of m reflects the shape of convex functions.

Typically as shown in Figure (1) , a small m corresponds to a 'flat' convex function while a large m corresponds to a 'steep' convex function.

2.1 Side results of strong convexity

Strong convexity has several interesting consequences. We will see that we can bound both $f^* - f(\mathbf{x})$ and $||\mathbf{x} - \mathbf{x}^*||_2$ in this section.

Lemma 1. *Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ that is m-strongly convex. Then, the objective function sub-optimality $f(\mathbf{x}) - f^*$ is bounded above by*

$$f(\mathbf{x}) - f(\mathbf{x}^*) \leq \frac{1}{2m}\|\nabla f(\mathbf{x})\|_2^2$$

Proof. The righthand side of the strong convexity inequality is a convex quadratic function of y (for fixed x). Setting the gradient with respect to y equal to zero, we can find the $\tilde{y}$ that minimizes the right hand side.

$$\begin{aligned}
\frac{\partial}{\partial x}(f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2) &= 0 \\
\nabla f(x) - m(y - x) &= 0 \\
y = x - \frac{1}{m}\nabla f(x)
\end{aligned}$$

So $\tilde{y} = x - (1/m)\nabla f(x)$ minimizes the righthand side. Plug this into the righthand side, we can derive the lower bound of $f(y)$, for arbitrary y.

$$\begin{aligned}
f(y) &\geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2 \\
&\geq f(x) + \langle \nabla f(x), \tilde{y} - x\rangle + \frac{m}{2}\|\tilde{y} - x\|^2 \\
&= f(x) - \frac{1}{2m}\|\nabla f(x)\|^2
\end{aligned}$$

By substituting y with x^*,

$$f^* \geq f(x) - \frac{1}{2m}\|\nabla f(x)\|^2 \tag{2}$$

and therefore

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

□

This result allows us to realize how fast you get to a minimum as a function of gradient. If the gradient is small at a point, then the point is nearly optimal. This upper bound also implies that if we find a point $\hat{\mathbf{x}}$ such that, $\|\nabla f(\hat{\mathbf{x}})\|_2 \leq \sqrt{2m\epsilon}$, then we can conclude that $\hat{\mathbf{x}}$ is ϵ-suboptimal, i.e., $f(\hat{\mathbf{x}}) - f(\mathbf{x}^*) \leq \epsilon$.

Similarly, we can also derive an upper bound on $\|\mathbf{x} - \mathbf{x}^*\|_2$, the distance between x and any optimal point x^*, in terms of $\|\nabla f(\mathbf{x})\|_2$:

Lemma 2. *Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ that is m-strongly convex. Then, the optimality distance $\|\mathbf{x} - \mathbf{x}^*\|_2$ is bounded above by*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \;\; \leq \;\; \frac{2}{m}\|\nabla f(\mathbf{x})\|_2 \tag{3}$$

where $\mathbf{x}^* = \arg\min_{\mathbf{x}} f(\mathbf{x})$ is the unique minimizer of f.

Proof. We apply (1) with $y = x^*$ to obtain:

$$\begin{aligned} f^* = f(x^*) &\geq f(x) + \langle \nabla f(x), x^* - x \rangle + \frac{m}{2}\|x^* - x\|_2^2 \\ &\geq f(x) - \|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2, \end{aligned}$$

Since $f^* \leq f(x)$, the terms following $f(x)$ on the righthand side must be negative. We have

$$\begin{aligned} -\|\nabla f(x)\|_2 \|x^* - x\|_2 + \frac{m}{2}\|x^* - x\|_2^2 &\leq 0 \\ \|x - x^*\|_2 &\leq \tfrac{2}{m}\|\nabla f(x)\|_2 \end{aligned}$$

from which (3) follows. $\square$

One consequence of (3) is the solution locates within a ball of radius of $\frac{2}{m}\|\nabla f(x)\|_2$ around the optimal solution.

Lemma 3. *Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ that is m-strongly convex. Then, for any $\mathbf{x}$ and $\mathbf{y}$ and $\alpha \in [0, 1]$ we have*

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) - \frac{\alpha(1-\alpha)m}{2}\|\mathbf{x} - \mathbf{y}\|^2$$

and

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top (\mathbf{x} - \mathbf{y}) \geq m\|\mathbf{x} - \mathbf{y}\|^2$$

Proof. Homework. $\square$

3 Smoothness (Lipschitz continuous gradients)

Definition 2. *A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is called M-smooth or has M-Lipschitz continuous gradients if for some $M > 0$ we have*

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\|.$$

This condition implies that the derivatives of the function do not change rapidly when two points are close to each other.

Lemma 4. *If f is M-smooth, then the following condition holds*

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) + \frac{M}{2}\|\mathbf{y} - \mathbf{x}\|^2 \tag{4}$$

Proof. We know that

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{x}) + \int_0^1 \nabla f(s\mathbf{x} + (1-s)\mathbf{y})^\top (\mathbf{y} - \mathbf{x}) ds \\ &= f(\mathbf{x}) + \nabla f^\top (\mathbf{y} - \mathbf{x}) + \int_0^1 \big(\nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\big)^\top (\mathbf{y} - \mathbf{x}) ds \end{aligned}$$

Hence,

$$
\begin{aligned}
|f(\mathbf{y}) - f(\mathbf{x}) - \nabla f^\top(\mathbf{y}-\mathbf{x})| &\le \left| \int_0^1 \big(\nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\big)^\top (\mathbf{y}-\mathbf{x}) ds \right| \\
&\le \int_0^1 \left\| \nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\big)^\top (\mathbf{y}-\mathbf{x}) \right\| ds \\
&\le \int_0^1 \left\| \nabla f(s\mathbf{x} + (1-s)\mathbf{y}) - \nabla f(\mathbf{x})\big) \right\| \|\mathbf{y}-\mathbf{x}\| ds \\
&\le \|\mathbf{y}-\mathbf{x}\| \int_0^1 M(1-s)\|\mathbf{y}-\mathbf{x}\| ds \\
&= \frac{M}{2}\|\mathbf{y}-\mathbf{x}\|^2
\end{aligned}
$$

□

3.1 Side results of smoothness

We can use M-smoothness to derive a lower bound on the suboptimality:

Lemma 5. *If the function f is smooth, then*

$$f(\mathbf{x}) - f^* \ge \frac{1}{2M}\|\nabla f(\mathbf{x})\|_2^2$$

Proof. We can minimize both sides of the following inequality with respect to $\mathbf{y}$

$$\min_{\mathbf{y}} f(\mathbf{y}) \le \min_{\mathbf{y}} \left(f(\mathbf{x}) + \nabla f(\mathbf{x})^\top (\mathbf{y}-\mathbf{x}) + \frac{M}{2}\|\mathbf{y}-\mathbf{x}\|^2 \right)$$

which implies that the optimal solution for the right hand side $\tilde{\mathbf{y}}$ is $\tilde{\mathbf{y}} = \mathbf{x} - \frac{1}{M}\nabla f(\mathbf{x})$. The optimal value of the left hand side is f^* and therefore we have

$$f^* \le f(\mathbf{x}) - \frac{1}{2M}\|\nabla f(\mathbf{x})\|_2^2$$

□

Lemma 6. *Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ that is M-smooth. Then, for any $\mathbf{x}$ and $\mathbf{y}$ and $\alpha \in [0,1]$ we have*

$$f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \ge \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) - \frac{\alpha(1-\alpha)M}{2}\|\mathbf{x}-\mathbf{y}\|^2$$

and

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top (\mathbf{x}-\mathbf{y}) \le M\|\mathbf{x}-\mathbf{y}\|^2$$

Proof. Homework. □

4 Strong convexity and smoothness for twice differentiable functions

Definition 3. *A twice differentiable function is m-strongly convex if*

$$\nabla^2 f(\mathbf{x}) \succeq m\mathbf{I}$$

Definition 4. *A twice differentiable function is M-smooth if*

$$-M\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq M\mathbf{I} \quad \Leftrightarrow \quad \|\nabla^2 f(\mathbf{x})\| \leq M$$

Next we show that when a function is twice differentiable, then the definitions of smoothness are equivalent.

For any $\mathbf{x}$ and $\mathbf{y}$ we have

$$\begin{aligned}
\nabla f(\mathbf{x}) &= \nabla f(\mathbf{y}) + \int_0^1 \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x}))(\mathbf{y} - \mathbf{x}) d\tau \\
&= \nabla f(\mathbf{y}) + \left(\int_0^1 \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) d\tau \right) (\mathbf{y} - \mathbf{x})
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| &= \left\| \left(\int_0^1 \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) d\tau \right) (\mathbf{y} - \mathbf{x}) \right\| \\
&\leq \left\| \int_0^1 \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) d\tau \right\| \|\mathbf{y} - \mathbf{x}\| \\
&\leq \left(\int_0^1 \left\| \nabla^2 f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) \right\| d\tau \right) \|\mathbf{y} - \mathbf{x}\| \\
&\leq M \|\mathbf{y} - \mathbf{x}\|
\end{aligned}$$

On the other hand, if function f satisfies $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\|$ then we can show that for any $\mathbf{s} \in \mathbb{R}^n$ we have

$$\left\| \left(\int_0^\alpha \nabla^2 f(\mathbf{x} + \tau\mathbf{s}) d\tau \right) .\mathbf{s} \right\| = \|\nabla f(\mathbf{x} + \alpha\mathbf{s}) - \nabla f(\mathbf{x})\| \leq \alpha M \|\mathbf{s}\|$$

If we divide this inequality by $\alpha\|\mathbf{s}\|$ and send $\alpha \to 0$, then we obtain

$$\left\| \nabla^2 f(\mathbf{x}) \right\| \leq M$$

5 Condition Number

From the strong convexity inequality and the smoothness inequality, we have:

$$mI \preceq \nabla^2 f(\mathbf{x}) \preceq MI \tag{5}$$

Definition 5. *If $mI \preceq \nabla^2 f(\mathbf{x}) \preceq MI$ for all $\mathbf{x} \in S$, then the* **condition number** *of f is $\kappa = \frac{M}{m}$.*

The condition number is thus a uniform (and hence upper) bound on the condition number of the matrix $\nabla^2 f(\mathbf{x})$ at any given $\mathbf{x}$.

Definition 6. *When the ratio is close to 1, we call it* **well-conditioned**. *When the ratio is very large, we call it* **ill-conditioned**.