SYS 6003: Optimization	Fall 2016

Lecture 15

Instructor: Quanquan Gu — *Date: October 19^{th}, 2016*

In last lecture, we introduced L-smooth functions and proved an important property of those functions. We summarize and show more properties of L-smooth function as follows.

Lemma 1 (Properties of Smooth Functions) *If f is L-smooth, then for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have*

$$
\begin{aligned}
&(1) \quad f(\mathbf{x}) \leq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top(\mathbf{x}-\mathbf{y}) + \frac{L}{2}\|\mathbf{x}-\mathbf{y}\|_2^2;\\
&(2) \quad (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top(\mathbf{x}-\mathbf{y}) \leq L\|\mathbf{x}-\mathbf{y}\|_2^2.
\end{aligned}
$$

If f is in addition convex, we have

$$
\begin{aligned}
&(3) \quad f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top(\mathbf{x}-\mathbf{y}) + \frac{1}{2L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2;\\
&(4) \quad (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top(\mathbf{x}-\mathbf{y}) \geq \frac{1}{L}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2.
\end{aligned}
$$

Remark 1 *If f is twice differentiable, then there is an equivalent, and perhaps easier, definition of smoothness: f is smooth if there exists a constant $L > 0$ such that $\nabla^2 f(x) \preceq L\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. In other words, the largest eigenvalue of the Hessian of f is uniformly upper bounded by L everywhere.*

Now we introduce strongly convex functions, which basically are convex functions that subtracting a quadratic function from it remains convex.

Definition 1 (Strongly Convex Function) *$f(\mathbf{x})$ is μ-strongly convex if $f(\mathbf{x}) - \mu\|\mathbf{x}\|_2^2/2$ is convex, where $\mu > 0$. The positive constant μ is called the modulus of strong convexity of f.*

Strongly convex functions also have several properties. We summarize them in the following lemma.

Lemma 2 (Properties of Strongly Convex Functions) *If f is a μ-strongly convex function, then for any $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\, f$, we have*

$$
\begin{aligned}
&(1) \quad f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x}) + (1-\alpha)\mathbf{y} - \frac{\mu}{2}\alpha(1-\alpha)\|\mathbf{x}-\mathbf{y}\|_2^2, \text{ for any } \alpha \in [0,1];\\
&(2) \quad f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top(\mathbf{x}-\mathbf{y}) + \frac{\mu}{2}\|\mathbf{x}-\mathbf{y}\|_2^2;\\
&(3) \quad (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top(\mathbf{x}-\mathbf{y}) \geq \mu\|\mathbf{x}-\mathbf{y}\|_2^2;\\
&(4) \quad f(\mathbf{x}) \leq f(\mathbf{y}) + \nabla f(\mathbf{y})^\top(\mathbf{x}-\mathbf{y}) + \frac{1}{2\mu}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2;\\
&(5) \quad (\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^\top(\mathbf{x}-\mathbf{y}) \leq \frac{1}{\mu}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2.
\end{aligned}
$$

As we can see, these properties are analogous to those of smooth functions, hence we omit the proof here.

Remark 2 *If f is twice differentiable, then there is an equivalent, and perhaps easier, definition of strong convexity: f is strongly convex if there exists a constant $\mu > 0$ such that $\nabla^2 f(x) \succeq \mu\mathbf{I}$, where $\mathbf{I}$ is an identity matrix. In other words, the smallest eigenvalue of the Hessian of f is uniformly lower bounded by μ everywhere.*

Remark 3 *Recall that if a function is L-smooth, the largest eigenvalue of the Hessian of f is uniformly upper bounded by L everywhere. Thus, strongly convexity can be deemed as an analogy to smoothness. By imposing strongly convexity and smoothness together, we basically require that the smallest eigenvalue of the Hessian of f is bounded away from zero, and the largest eigenvalue of the Hessian of f is bounded from above, i.e., $\mu\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq L\mathbf{I}$. In particular, we must have $\mu \leq L$ for the same function f.*

Recall that we have proved the following theorem regarding the convergence rate of gradient descent method for L-smooth functions.

Theorem 1 *If f is convex and L-smooth, then the gradient descent with stepsize $\eta = 1/L$ satisfies*

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}^*) \leq \frac{L\|\mathbf{x}_1 - \mathbf{x}^*\|}{2t}.$$

In the proof of this theorem, we have also showed that the function value is monotonically decreasing. We summarize it in the following lemma.

Lemma 3 *Under the same conditions of Theorem 1, we have*

$$f(\mathbf{x}_{t+1}) - f(\mathbf{x}_t) \leq -\frac{1}{2L}\|\nabla f(\mathbf{x}_t)\|_2^2.$$

In fact, we can show that not only the function value is monotonically decreasing, but also the estimation error of the iterate is monotonically decreasing.

Lemma 4 *Under the same conditions of Theorem 1, we have*

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|_2 \leq \|\mathbf{x}_t - \mathbf{x}^*\|_2 - 1/L^2 \cdot \|\nabla f(\mathbf{x}_t)\|_2^2$$

We are going to prove it next time.