1 Outline

In this lecture, we study

- More properties of smooth and strongly convex functions,
- Convergence of gradient descent for functions that are smooth and strongly convex.

1.1 More properties of smooth and strongly convex functions

Another interesting result is that when f is smooth, we can measure the gap between the optimal value and $f(x)$ for any given solution x. More precisely, we prove the following result.

Theorem 11.1. *If $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth with respect to the ℓ_2 norm, then*

$$\frac{1}{2\beta}\|\nabla f(x)\|_2^2 \leq f(x) - f(x^*) \leq \frac{\beta}{2}\|x - x^*\|_2^2 \quad \forall x \in \mathbb{R}^d$$

where x^* is an optimal solution to $\min_{x\in\mathbb{R}^d} f(x)$.

Proof. Let us prove the upper bound on $f(x) - f(x^*)$ first. As f is β-smooth, we have

$$f(x) \leq f(x^*) + \nabla f(x^*)^\top (x - x^*) + \frac{\beta}{2}\|x - x^*\|_2^2,$$

which implies the upper bound as $\nabla f(x^*) = 0$. For the lower bound, note that for any $y \in \mathbb{R}^d$,

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

Here, we can take $y = x - (1/\beta)\nabla f(x)$, which makes the right-most side

$$f(x) - \frac{1}{2\beta}\|\nabla f(x)\|_2^2.$$

Then it follows that

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

as required. □

Based on Theorem 11.1, we can prove the following property of smooth functions.

Lemma 11.2. *If $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth with respect to the ℓ_2 norm, then*

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \frac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|_2^2$$

for any $x, y \in \mathbb{R}^d$.

Proof. Given $x, y \in \mathbb{R}^d$, we take the following two functions.

$$g(z) = f(z) - \nabla f(x)^\top z,$$
$$h(z) = f(z) - \nabla f(y)^\top z.$$

As $\nabla g(z) = \nabla f(z) - \nabla f(x)$ and $\nabla h(z) = \nabla f(z) - \nabla f(y)$, it follows that x and y minimize g and h, respectively. Moreover, g and h are both β-smooth. Note that

$$\begin{aligned} f(y) - f(x) - \nabla f(x)^\top (y - x) &= g(y) - g(x) \\ &\geq \frac{1}{2\beta}\|\nabla g(y)\|_2^2 \\ &= \frac{1}{2\beta}\|\nabla f(y) - \nabla f(x)\|_2^2. \end{aligned}$$

Similarly, we have

$$\begin{aligned} f(x) - f(y) - \nabla f(y)^\top (x - y) &= h(x) - h(y) \\ &\geq \frac{1}{2\beta}\|\nabla h(x)\|_2^2 \\ &= \frac{1}{2\beta}\|\nabla f(x) - \nabla f(y)\|_2^2. \end{aligned}$$

Adding these two inequalities, we obtain

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \frac{1}{\beta}\|\nabla f(x) - \nabla f(y)\|_2^2,$$

as required. □

Recall that Theorem 11.1 measures the optimality gap of any given solution x for a smooth function. We can provide a similar result for bounding the optimality gap for strongly convex functions.

Theorem 11.3. *If $f : \mathbb{R}^d \to \mathbb{R}$ is α-strongly convex with respect to the ℓ_2 norm, then*

$$\frac{\alpha}{2}\|x - x^*\|_2^2 \leq f(x) - f(x^*) \leq \frac{1}{2\alpha}\|\nabla f(x)\|_2^2 \quad \forall x \in \mathbb{R}^d$$

where x^* is an optimal solution to $\min_{x\in\mathbb{R}^d} f(x)$.

Proof. Let us prove the lower bound on $f(x) - f(x^*)$ first. As f is α-strongly convex, we have

$$f(x) \geq f(x^*) + \nabla f(x^*)^\top (x - x^*) + \frac{\alpha}{2}\|x - x^*\|_2^2,$$

which implies the lower bound as $\nabla f(x^*) = 0$. For the upper bound, note that

$$\begin{aligned} f(x^*) &\geq f(x) + \nabla f(x)^\top (x^* - x) + \frac{\alpha}{2}\|x^* - x\|_2^2 \\ &\geq \min_{y\in\mathbb{R}^d} f(x) + \nabla f(x)^\top (y - x) + \frac{\alpha}{2}\|y - x\|_2^2. \end{aligned}$$

The minimization term above is minimized when y satisfies $\nabla f(x) + \alpha(y - x) = 0$, which is equivalent to $y = x - (1/\alpha)\nabla f(x)$. Therefore,

$$f(x^*) \geq f(x) - \frac{1}{2\alpha}\|\nabla f(x)\|_2^2,$$

as required. □

Lastly, we show the following result for strongly convex functions, which holds because the monotonicity condition for $f(x) - (\alpha/2)\|x\|_2^2$.

Lemma 11.4. *If $f : \mathbb{R}^d \to \mathbb{R}$ is α-strongly convex with respect to the ℓ_2 norm, then*

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \alpha \|x - y\|_2^2$$

for any $x, y \in \mathbb{R}^d$.

Proof. As $g(x) = f(x) - (\alpha/2)\|x\|_2^2$ is convex, the monotonicity of the gradient of g implies that

$$(\nabla g(x) - \nabla g(y))^\top (x - y) \geq 0$$

for any $x, y \in \mathbb{R}^d$. Note that $\nabla g(x) = \nabla f(x) - \alpha x$ and $\nabla g(y) = \nabla f(y) - \alpha y$, which implies that

$$(\nabla g(x) - \nabla g(y))^\top (x - y) = (\nabla f(x) - \nabla f(y))^\top (x - y) - \alpha \|x - y\|_2^2.$$

Then we obtain $(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \alpha \|x - y\|_2^2$, as required. □

1.2 Convergence result for smooth and strongly convex functions

When f is both smooth and strongly convex, f satisfies the following property.

Lemma 11.5. *If $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth with respect to the ℓ_2 norm and $\beta \geq \alpha$, then $f(x) - (\alpha/2)\|x\|_2^2$ is $(\beta - \alpha)$-smooth.*

Proof. We learned from the previous lecture that $f(x) - (\alpha/2)\|x\|_2^2$ is $(\beta - \alpha)$-smooth if and only if

$$\frac{\beta - \alpha}{2}\|x\|_2^2 - \left(f(x) - \frac{\alpha}{2}\|x\|_2^2\right) = \frac{\beta}{2}\|x\|_2^2 - f(x)$$

is convex. Then, again, $(\beta/2)\|x\|_2^2 - f(x)$ is convex if and only if f is β-smooth. Since f is β-smooth, it follows that $f(x) - (\alpha/2)\|x\|_2^2$ is $(\beta - \alpha)$-smooth, as required. □

Based on Lemma 11.5, we can prove the following result on functions that are both smooth and strongly convex.

Lemma 11.6. *If $f : \mathbb{R}^d \to \mathbb{R}$ is β-smooth and α-strongly convex with respect to the ℓ_2 norm, then*

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

for any $x, y \in \mathbb{R}^d$.

Proof. Since f is α-strongly convex, $f(x) - (\alpha/2)\|x\|_2^2$ is convex. Moreover, $f(x) - (\alpha/2)\|x\|_2^2$ is $(\beta - \alpha)$-smooth by Lemma 11.5. Applying Lemma 11.2 to $f(x) - (\alpha/2)\|x\|_2^2$, it follows that

$$\begin{aligned}
&(\nabla f(x) - \nabla f(y))^\top (x - y) - \alpha\|x - y\|_2^2 \\
&\geq \frac{1}{\beta - \alpha}\|\nabla f(x) - \nabla f(y)\|_2^2 - \frac{2\alpha}{\beta - \alpha}(\nabla f(x) - \nabla f(y))^\top (x - y) + \frac{\alpha^2}{\beta - \alpha}\|x - y\|_2^2.
\end{aligned}$$

This implies that

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

as required. □

Obseve that Lemma 11.6 is a combination of Lemma 11.2 for smooth functions and Lemma 11.4 for strongly convex functions. This strong peroperty of functions that are both smooth and strongly convex leads to a linear convergence of gradient descent.

Theorem 11.7. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be β-smooth and α-strongly convex, and let $\{x_t : t = 1, \ldots, T+1\}$ be the sequence of iterates generated by gradient descent with sep size $\eta_t = 2/(\alpha + \beta)$ for each t. Then*

$$f(x_{T+1}) - f(x^*) \leq \frac{\beta}{2} \exp\left(-\frac{4T}{\kappa+1}\right) \|x_1 - x^*\|_2^2$$

where x^* is an optimal solution to $\min_{x \in \mathbb{R}^d} f(x)$.

Proof. Let $\eta_t = \eta$ for each $t \geq 1$. Note that

$$\begin{aligned}
\|x_{t+1} - x^*\|_2^2 &= \|x_t - \eta \nabla f(x_t) - x^*\|_2^2 \\
&= \|x_t - x^*\|_2^2 - 2\eta \nabla f(x_t)^\top (x_t - x^*) + \eta^2 \|\nabla f(x_t)\|_2^2 \\
&\leq \|x_t - x^*\|_2^2 - \frac{2\eta}{\alpha+\beta}\|\nabla f(x_t)\|_2^2 - \frac{2\eta\alpha\beta}{\alpha+\beta}\|x_t - x^*\|_2^2 + \eta^2 \|\nabla f(x_t)\|_2^2 \\
&= \left(1 - \frac{2\eta\alpha\beta}{\alpha+\beta}\right)\|x_t - x^*\|_2^2 + \left(\eta^2 - \frac{2\eta}{\alpha+\beta}\right)\|\nabla f(x_t)\|_2^2
\end{aligned}$$

where the inequality follows from Lemma 11.6. Setting $\eta = 2/(\alpha+\beta)$, we obtain

$$\|x_{t+1} - x^*\|_2^2 \leq \left(\frac{\beta-\alpha}{\beta+\alpha}\right)^2 \|x_t - x^*\|_2^2 = \left(\frac{\kappa-1}{\kappa+1}\right)^2 \|x_t - x^*\|_2^2,$$

which implies that

$$\|x_{t+1} - x^*\|_2 \leq \left(\frac{\kappa-1}{\kappa+1}\right)\|x_t - x^*\|_2 \leq \left(\frac{\kappa-1}{\kappa+1}\right)^t \|x_1 - x^*\|_2.$$

Since f is β-smooth, we have

$$f(x_{t+1}) - f(x^*) \leq \frac{\beta}{2}\|x_{t+1} - x^*\|_2^2 \leq \frac{\beta}{2}\left(\frac{\kappa-1}{\kappa+1}\right)^{2t} \|x_1 - x^*\|_2^2.$$

Lastly,

$$\left(\frac{\kappa-1}{\kappa+1}\right)^{2t} = \left(1 - \frac{2}{\kappa+1}\right)^{2t} \leq \exp\left(-\frac{4t}{\kappa+1}\right),$$

as required. ∎