Lecture 9

Subgradient method
strongly convex case

f is μ-str. convex if

$$\alpha f(x) + (1-\alpha) f(y) \geq f(\alpha x + (1-\alpha) y) + \frac{\mu}{2} \alpha (1-\alpha) \|x-y\|^2$$

$\Leftrightarrow$ $f(x) - \frac{1}{2} \mu \|x\|^2$ is convex

if f is diff. then str. convexity $\Leftrightarrow$

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

In general,

$$f(y) - f(x) \geq \langle g_x, y-x \rangle + \frac{\mu}{2} \|y-x\|^2 \quad \forall g_x \in \partial f(x)$$

$x \neq y$ are minimizers

$$0 = f(y) - f(x) \geq 0 + \frac{\mu}{2} \|x-y\|^2$$

$$\begin{cases} g_k \in \partial f(x_k) \\ x_{k+1} = x_k - \alpha_k g_k \end{cases}$$

Subgrad. method

f is μ-str. convex

$\|g(x)\| \leq G \quad \forall x$

$$\alpha_k = \frac{2}{(k+1)\mu}$$

x^* is the solution.

$$\|x_{k+1} - x^*\|^2 = \|x_k - \alpha_k g_k - x^*\|^2$$

$$= \|x_k - x^*\|^2 - 2\alpha_k \langle g_k, x_k - x^* \rangle + \alpha_k^2 \|g_k\|^2 \leq$$

$$f(x^*) - f(x_k) \geq \langle g_k, x^* - x_k \rangle + \frac{\mu}{2} \|x_k - x^*\|^2$$

$$\leq \|x_k - x^*\|^2 + 2\alpha_k\left(f_* - f(x_k) - \frac{\mu}{2}\|x_k - x^*\|^2\right) + \alpha_k^2 G^2$$

$$f(x_k) - f_* \leq \cancel{\frac{1 - \alpha_k \mu}{2\alpha_k}} \|x_k - x^*\|^2 - \frac{1}{2\alpha_k}\|x_{k+1} - x^*\|^2 + \frac{G^2}{2}\alpha_k$$

$$\frac{1}{2}\left(\frac{1}{\alpha_k} - \mu\right)$$

$$\alpha_k = \frac{2}{(k+1)\mu} : \quad \frac{(k+1)\mu}{2} - \mu = \frac{(k-1)\mu}{2}$$

$$f(x_k) - f_* \leq \frac{1}{4}\mu(k-1)\|x_k - x^*\|^2 - \frac{1}{4}\mu(k+1)\|x_{k+1} - x^*\|^2 + \frac{G^2}{\mu(k+1)}$$

$$k\left(f(x_k) - f_*\right) \leq \frac{1}{4}k(k-1)\mu\|x_k - x^*\|^2 - \frac{1}{4}k(k+1)\mu\|x_{k+1} - x^*\|^2 + \frac{G^2}{\mu}\frac{k}{k+1}$$

$$\sum_{i=0}^{k} i\left(f(x_i) - f_*\right) \le -\frac{1}{4} - \dots + \frac{\sigma^2}{\mu} \sum_{i=0}^{k} \frac{i}{i+1}$$

$$\le \frac{\sigma^2 k}{\mu}$$

$$\text{LHS} \geq (1+2+\dots+k) \min_i \left(f(x_i) - f_*\right)$$

$$= \frac{k(k+1)}{2} \min f(x_i) - f_*$$

$$\min_{i=1,k} \left(f(x_i) - f_*\right) \le \frac{2\sigma^2}{\mu(k+1)} = O\left(\frac{1}{k}\right)$$

Rate $\|x_u - x^*\|^2$ is the best for

$$f(x) - f(x^*) \geq 0 + \frac{\mu}{2} \|x - x^*\|^2$$

For GD

$$f(x_{k+1}) - f_* \leq \left(1 - \frac{\mu}{L}\right)\left(f(x_k) - f_*\right)$$

$$\leq \left(1 - \frac{\mu}{L}\right)^k \left(f(x_1) - f_*\right)$$