

LECTURE 24

LECTURE OUTLINE

- Subgradient Methods
- Stepsize Rules and Convergence Analysis

- Consider a generic convex problem $\min_{x \in X} f(x)$, where $f : \mathcal{R}^n \mapsto \mathcal{R}$ is a convex function and X is a closed convex set, and the subgradient method

$$x_{k+1} = [x_k - \alpha_k g_k]^+,$$

where g_k is a subgradient of f at x_k , α_k is a positive stepsize, and $[\cdot]^+$ denotes projection on the set X .

- Incremental version for problem $\min_{x \in X} \sum_{i=1}^m f_i(x)$

$$x_{k+1} = \psi_{m,k}, \quad \psi_{i,k} = [\psi_{i-1,k} - \alpha_k g_{i,k}]^+, \quad i = 1, \dots, m$$

starting with $\psi_{0,k} = x_k$, where $g_{i,k}$ is a subgradient of f_i at $\psi_{i-1,k}$.

ASSUMPTIONS AND KEY INEQUALITY

- **Assumption: (Subgradient Boundedness)**

$$\|g\| \leq C_i, \quad \forall g \in \partial f_i(x_k) \cup \partial f_i(\psi_{i-1,k}), \quad \forall i, k,$$

for some scalars C_1, \dots, C_m . (Satisfied when the f_i are polyhedral as in integer programming.)

- **Key Lemma: For all $y \in X$ and k ,**

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2\alpha_k (f(x_k) - f(y)) + \alpha_k^2 C^2,$$

where

$$C = \sum_{i=1}^m C_i$$

and C_i is as in the boundedness assumption.

- *First Insight:* For any y that is better than x_k , the distance to y is improved if α_k is small enough:

$$0 < \alpha_k < \frac{2(f(x_k) - f(y))}{C^2}$$

PROOF OF KEY LEMMA

- For each f_i and all $y \in X$, and i, k

$$\begin{aligned}
 \|\psi_{i,k} - y\|^2 &= \|[\psi_{i-1,k} - \alpha_k g_{i,k}]^+ - y\|^2 \\
 &\leq \|\psi_{i-1,k} - \alpha_k g_{i,k} - y\|^2 \\
 &\leq \|\psi_{i-1,k} - y\|^2 - 2\alpha_k g'_{i,k}(\psi_{i-1,k} - y) + \alpha_k^2 C_i^2 \\
 &\leq \|\psi_{i-1,k} - y\|^2 - 2\alpha_k (f_i(\psi_{i-1,k}) - f_i(y)) + \alpha_k^2 C_i^2
 \end{aligned}$$

By adding over i , and strengthening,

$$\begin{aligned}
 \|x_{k+1} - y\|^2 &\leq \|x_k - y\|^2 - 2\alpha_k (f(x_k) - f(y)) \\
 &\quad + 2\alpha_k \sum_{i=1}^m C_i \|\psi_{i-1,k} - x_k\| + \alpha_k^2 \sum_{i=1}^m C_i^2 \\
 &\leq \|x_k - y\|^2 - 2\alpha_k (f(x_k) - f(y)) \\
 &\quad + \alpha_k^2 \left(2 \sum_{i=2}^m C_i \left(\sum_{j=1}^{i-1} C_j \right) + \sum_{i=1}^m C_i^2 \right) \\
 &= \|x_k - y\|^2 - 2\alpha_k (f(x_k) - f(y)) + \alpha_k^2 \left(\sum_{i=1}^m C_i \right)^2 \\
 &= \|x_k - y\|^2 - 2\alpha_k (f(x_k) - f(y)) + \alpha_k^2 C^2.
 \end{aligned}$$

STEP SIZE RULES

- Constant Stepsize $\alpha_k \equiv \alpha$:
 - By key lemma with $f(y) \approx f^*$, it makes progress to the optimal if $0 < \alpha < \frac{2(f(x_k) - f^*)}{C^2}$, i.e., if

$$f(x_k) > f^* + \frac{\alpha C^2}{2}$$

- Diminishing Stepsize $\alpha_k \rightarrow 0, \sum_k \alpha_k = \infty$:
 - Eventually makes progress (once α_k becomes small enough). Can show that

$$\liminf_{k \rightarrow \infty} f(x_k) = f^*.$$

- Dynamic Stepsize $\alpha_k = \frac{f(x_k) - f_k}{C^2}$ where $f_k = f^*$ or (more practically) an estimate of f^* :
 - If $f_k = f^*$, makes progress at every iteration. If $f_k < f^*$ it tends to oscillate around the optimum. If $f_k > f^*$ it tends towards the level set $\{x \mid f(x) \leq f_k\}$.

CONSTANT STEPSIZE ANALYSIS

- **Proposition:** For $\alpha_k \equiv \alpha$, we have

$$\liminf_{k \rightarrow \infty} f(x_k) \leq f^* + \frac{\alpha C^2}{2},$$

where $C = \sum_{i=1}^m C_i$ (in the case where $f^* = -\infty$, we have $\liminf_{k \rightarrow \infty} f(x_k) = -\infty$.)

- **Proof by contradiction.** Let $\epsilon > 0$ be s.t.

$$\liminf_{k \rightarrow \infty} f(x_k) > f^* + \frac{\alpha C^2}{2} + 2\epsilon,$$

and let $\hat{y} \in X$ be such that

$$\liminf_{k \rightarrow \infty} f(x_k) \geq f(\hat{y}) + \frac{\alpha C^2}{2} + 2\epsilon.$$

For all k large enough, we have

$$f(x_k) \geq \liminf_{k \rightarrow \infty} f(x_k) - \epsilon.$$

Add to get $f(x_k) - f(\hat{y}) \geq \alpha C^2 / 2 + \epsilon$. Use the key lemma for $y = \hat{y}$ to obtain a contradiction.

COMPLEXITY ESTIMATE FOR CONSTANT STEP

- For any $\epsilon > 0$, we have

$$\min_{0 \leq k \leq K} f(x_k) \leq f^* + \frac{\alpha C^2 + \epsilon}{2}$$

where

$$K = \left\lfloor \frac{(d(x_0, X^*))^2}{\alpha \epsilon} \right\rfloor.$$

- By contradiction. Assume that for $0 \leq k \leq K$

$$f(x_k) > f^* + \frac{\alpha C^2 + \epsilon}{2}.$$

Using this relation in the key lemma,

$$\begin{aligned} (d(x_{k+1}, X^*))^2 &\leq (d(x_k, X^*))^2 - 2\alpha(f(x_k) - f^*) + \alpha^2 C^2 \\ &\leq (d(x_k, X^*))^2 - (\alpha^2 C^2 + \alpha \epsilon) + \alpha^2 C^2 \\ &= (d(x_k, X^*))^2 - \alpha \epsilon. \end{aligned}$$

Sum over k to get $(d(x_0, X^*))^2 - (K + 1)\alpha \epsilon \geq 0$.

CONVERGENCE FOR OTHER STEPSIZE RULES

- (Diminishing Step): Assume that

$$\alpha_k > 0, \quad \lim_{k \rightarrow \infty} \alpha_k = 0, \quad \sum_{k=0}^{\infty} \alpha_k = \infty.$$

Then,

$$\liminf_{k \rightarrow \infty} f(x_k) = f^*.$$

If the set of optimal solutions X^* is nonempty and compact,

$$\lim_{k \rightarrow \infty} d(x_k, X^*) = 0, \quad \lim_{k \rightarrow \infty} f(x_k) = f^*.$$

- (Dynamic Stepsize with $f_k = f^*$): If X^* is nonempty, x_k converges to some optimal solution.

DYNAMIC STEPSIZE WITH ESTIMATE

- Estimation method:

$$f_k^{\text{lev}} = \min_{0 \leq j \leq k} f(x_j) - \delta_k,$$

and δ_k is updated according to

$$\delta_{k+1} = \begin{cases} \rho\delta_k & \text{if } f(x_{k+1}) \leq f_k^{\text{lev}}, \\ \max\{\beta\delta_k, \delta\} & \text{if } f(x_{k+1}) > f_k^{\text{lev}}, \end{cases}$$

where δ , β , and ρ are fixed positive constants with $\beta < 1$ and $\rho \geq 1$.

- Here we essentially “aspire” to reach a target level that is smaller by δ_k over the best value achieved thus far.
- We can show that

$$\inf_{k \geq 0} f(x_k) \leq f^* + \delta$$

(or $\inf_{k \geq 0} f(x_k) = f^*$ if $f^* = -\infty$).