LECTURE 25

LECTURE OUTLINE

- Incremental Subgradient Methods
- Convergence Rate Analysis and Randomized Methods

****

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

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

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

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

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

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

$$C_i = \sup_k \big\{ \|g\| \mid g \in \partial f_i(x_k) \cup \partial f_i(\psi_{i-1,k}) \big\}$$

CONSTANT STEPSIZE

- For $\alpha_k \equiv \alpha$, we have

$$\liminf_{k\to\infty} f(x_k) \le f^* + \frac{\alpha C^2}{2}.$$

- Sharpness of the estimate:
  - Consider the problem

$$\min_x \sum_{i=1}^{M} C_0\big(|x+1| + 2|x| + |x-1|\big)$$

with the worst component processing order

- Lower bound on the error. There is a problem, where even with best processing order,

$$f^* + \frac{\alpha m C_0^2}{2} \le \liminf_{k\to\infty} f(x_k)$$

where

$$C_0 = \max\{C_1, \ldots, C_m\}$$

COMPLEXITY ESTIMATE FOR CONSTANT STEP

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

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

where

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

RANDOMIZED ORDER METHODS

$$x_{k+1} = \left[x_k - \alpha_k g(\omega_k, x_k)\right]^+$$

where ω_k is a random variable taking equiprobable values from the set $\{1, \ldots, m\}$, and $g(\omega_k, x_k)$ is a subgradient of the component f_{ω_k} at x_k.

- Assumptions:

  (a) $\{\omega_k\}$ is a sequence of independent random variables. Furthermore, the sequence $\{\omega_k\}$ is independent of the sequence $\{x_k\}$.

  (b) The set of subgradients $\{g(\omega_k, x_k) \mid k = 0, 1, \ldots\}$ is bounded, i.e., there exists a positive constant C_0 such that with prob. 1

$$||g(\omega_k, x_k)|| \le C_0, \qquad \forall\, k \ge 0.$$

- Stepsize Rules:
  - Constant: $\alpha_k \equiv \alpha$
  - Diminishing: $\sum_k \alpha_k = \infty$, $\sum_k (\alpha_k)^2 < \infty$
  - Dynamic

RANDOMIZED METHOD W/ CONSTANT STEP

- With probability 1

$$\inf_{k \geq 0} f(x_k) \leq f^* + \frac{\alpha m C_0^2}{2}$$

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

Proof: By adapting key lemma, for all $y \in X,\ k$

$$||x_{k+1}-y||^2 \leq ||x_k-y||^2 - 2\alpha\big(f_{\omega_k}(x_k) - f_{\omega_k}(y)\big) + \alpha^2 C_0^2$$

Take conditional expectation with $\mathcal{F}_k = \{x_0, \ldots, x_k\}$

$$\begin{aligned} E\big\{||x_{k+1}-y||^2 \mid \mathcal{F}_k\big\} &\leq ||x_k - y||^2 \\ &\quad - 2\alpha E\big\{f_{\omega_k}(x_k) - f_{\omega_k}(y) \mid \mathcal{F}_k\big\} + \alpha^2 C_0^2 \\ &= ||x_k - y||^2 - 2\alpha \sum_{i=1}^{m} \frac{1}{m}\big(f_i(x_k) - f_i(y)\big) + \alpha^2 C_0^2 \\ &= ||x_k - y||^2 - \frac{2\alpha}{m}\big(f(x_k) - f(y)\big) + \alpha^2 C_0^2, \end{aligned}$$

where the first equality follows since ω_k takes the values $1, \ldots, m$ with equal probability $1/m$.

PROOF CONTINUED I

- Fix $\gamma > 0$, consider the level set L_γ defined by

$$L_\gamma = \left\{ x \in X \mid f(x) < f^* + \frac{2}{\gamma} + \frac{\alpha m C_0^2}{2} \right\}$$

and let $y_\gamma \in L_\gamma$ be such that $f(y_\gamma) = f^* + \frac{1}{\gamma}$. Define a new process $\{\hat{x}_k\}$ as follows

$$\hat{x}_{k+1} = \begin{cases} \left[\hat{x}_k - \alpha g(\omega_k, \hat{x}_k)\right]^+ & \text{if } \hat{x}_k \notin L_\gamma, \\ y_\gamma & \text{otherwise,} \end{cases}$$

where $\hat{x}_0 = x_0$. We argue that $\{\hat{x}_k\}$ (and hence also $\{x_k\}$) will eventually enter each of the sets L_γ.

Using key lemma with $y = y_\gamma$, we have

$$E\left\{ ||\hat{x}_{k+1} - y_\gamma||^2 \mid \mathcal{F}_k \right\} \leq ||\hat{x}_k - y_\gamma||^2 - z_k,$$

where

$$z_k = \begin{cases} \frac{2\alpha}{m}\big(f(\hat{x}_k) - f(y_\gamma)\big) - \alpha^2 C_0^2 & \text{if } \hat{x}_k \notin L_\gamma, \\ 0 & \text{if } \hat{x}_k = y_\gamma. \end{cases}$$

PROOF CONTINUED II

- If $\hat{x}_k \notin L_\gamma$, we have

$$
\begin{aligned}
z_k &= \frac{2\alpha}{m}\big(f(\hat{x}_k) - f(y_\gamma)\big) - \alpha^2 C_0^2 \\
&\geq \frac{2\alpha}{m}\left(f^* + \frac{2}{\gamma} + \frac{\alpha m C_0^2}{2} - f^* - \frac{1}{\gamma}\right) - \alpha^2 C_0^2 \\
&= \frac{2\alpha}{m\gamma}.
\end{aligned}
$$

Hence, as long as $\hat{x}_k \notin L_\gamma$, we have

$$
E\big\{||\hat{x}_{k+1} - y_\gamma||^2 \mid \mathcal{F}_k\big\} \leq ||\hat{x}_k - y_\gamma||^2 - \frac{2\alpha}{m\gamma}.
$$

This, cannot happen for an infinite number of iterations, so that $\hat{x}_k \in L_\gamma$ for sufficiently large k. Hence, in the original process we have

$$
\inf_{k\geq 0} f(x_k) \leq f^* + \frac{2}{\gamma} + \frac{\alpha m C_0^2}{2}
$$

with probability 1. Letting $\gamma \to \infty$, we obtain $\inf_{k\geq 0} f(x_k) \leq f^* + \alpha m C_0^2/2$. **Q.E.D.**

CONVERGENCE RATE

- Let $\alpha_k \equiv \alpha$ in the randomized method. Then, for any positive scalar ϵ, we have with prob. 1

$$\min_{0 \le k \le N} f(x_k) \le f^* + \frac{\alpha m C_0^2 + \epsilon}{2},$$

where N is a random variable with

$$E\{N\} \le \frac{m\big(d(x_0, X^*)\big)^2}{\alpha \epsilon}$$

- Compare w/ the deterministic method. It is guaranteed to reach after processing no more than

$$K = \frac{m\big(d(x_0, X^*)\big)^2}{\alpha \epsilon}$$

components the level set

$$\left\{ x \;\middle|\; f(x) \le f^* + \frac{\alpha m^2 C_0^2 + \epsilon}{2} \right\}$$

BASIC TOOL FOR PROVING CONVERGENCE

- **Supermartingale Convergence Theorem:** Let Y_k, Z_k, and W_k, $k = 0, 1, 2, \ldots$, be three sequences of random variables and let $\mathcal{F}_k$, $k = 0, 1, 2, \ldots$, be sets of random variables such that $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ for all k. Suppose that:

  (a) The random variables Y_k, Z_k, and W_k are nonnegative, and are functions of the random variables in $\mathcal{F}_k$.

  (b) For each k, we have

$$E\big\{Y_{k+1} \mid \mathcal{F}_k\big\} \le Y_k - Z_k + W_k$$

  (c) There holds $\sum_{k=0}^{\infty} W_k < \infty$.

Then, $\sum_{k=0}^{\infty} Z_k < \infty$, and the sequence Y_k converges to a nonnegative random variable Y, with prob. 1.

- Can be used to show convergence of randomized subgradient methods with diminishing and dynamic stepsize rules.