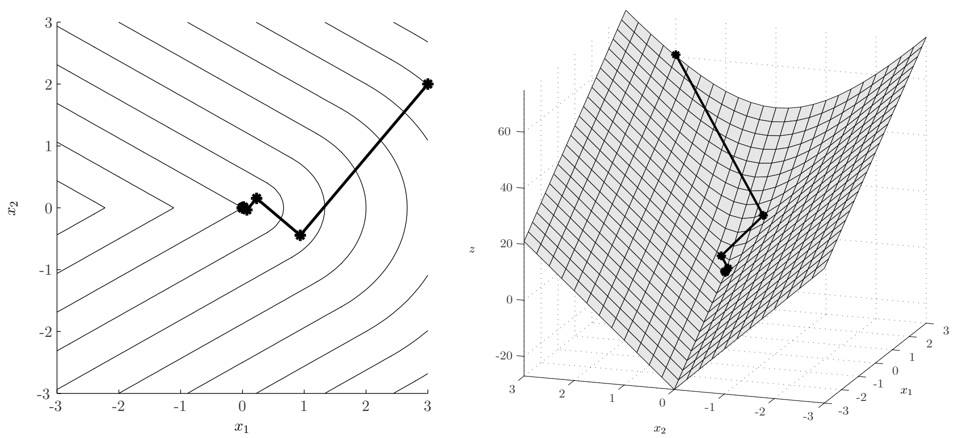
LECTURE 15

LECTURE OUTLINE

- Subgradient methods
- Calculation of subgradients
- Convergence

****

- Steepest descent at a point requires knowledge of the entire subdifferential at a point
- Convergence failure of steepest descent



- Subgradient methods abandon the idea of computing the full subdifferential to effect cost function descent ...
- Move instead along the direction of a single arbitrary subgradient

SINGLE SUBGRADIENT CALCULATION

- **Key special case:** Minimax

$$f(x) = \sup_{z \in Z} \phi(x, z)$$

where $Z \subset \Re^m$ and $\phi(\cdot, z)$ is convex for all $z \in Z$.

- For fixed $x \in \mathrm{dom}(f)$, assume that $z_x \in Z$ attains the supremum above. Then

$$g_x \in \partial \phi(x, z_x) \qquad \Rightarrow \qquad g_x \in \partial f(x)$$

- **Proof:** From subgradient inequality, for all y,

$$\begin{aligned} f(y) &= \sup_{z \in Z} \phi(y, z) \geq \phi(y, z_x) \geq \phi(x, z_x) + g_x'(y - x) \\ &= f(x) + g_x'(y - x) \end{aligned}$$

- **Special case:** Dual problem of $\min_{x \in X,\, g(x) \leq 0} f(x)$:

$$\max_{\mu \geq 0} q(\mu) \equiv \inf_{x \in X} L(x, \mu) = \inf_{x \in X} \big\{ f(x) + \mu' g(x) \big\}$$

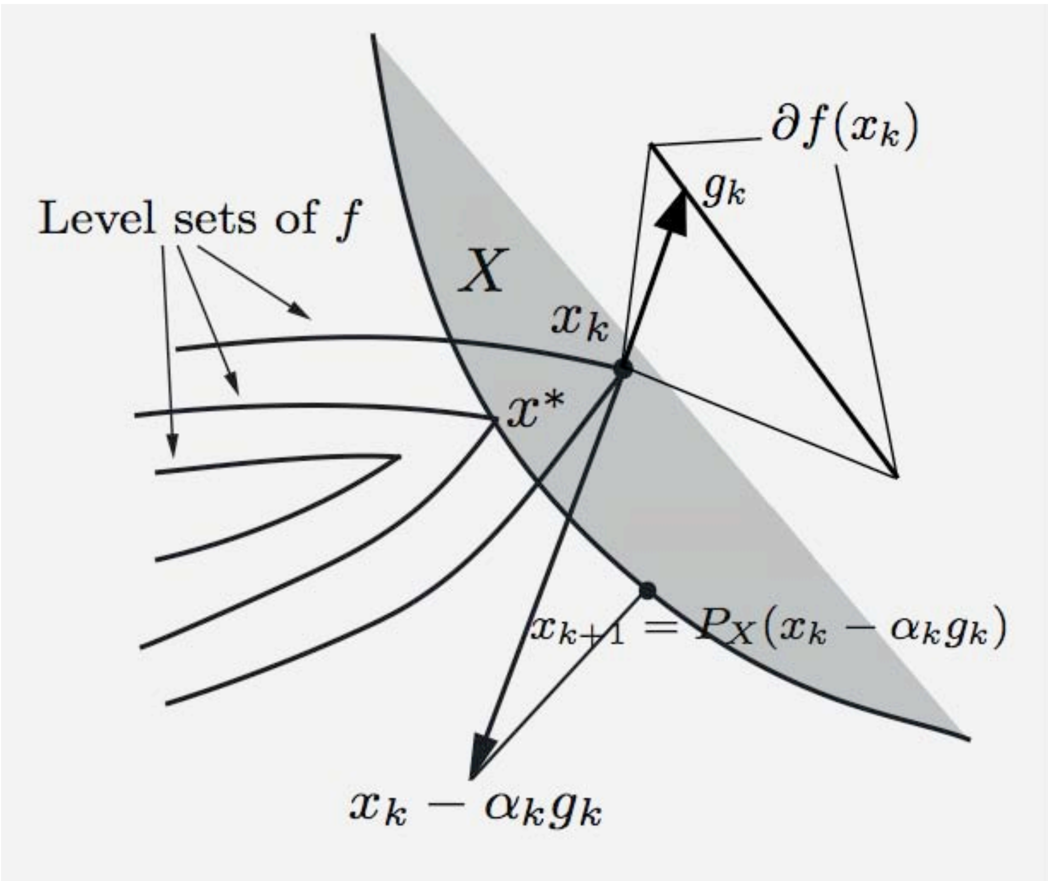
or $\min_{\mu \geq 0} F(\mu)$, where $F(-\mu) \equiv -q(\mu)$.

ALGORITHMS: SUBGRADIENT METHOD

- **Problem:** Minimize convex function $f : \Re^n \mapsto \Re$ over a closed convex set X.
- **Subgradient method**:

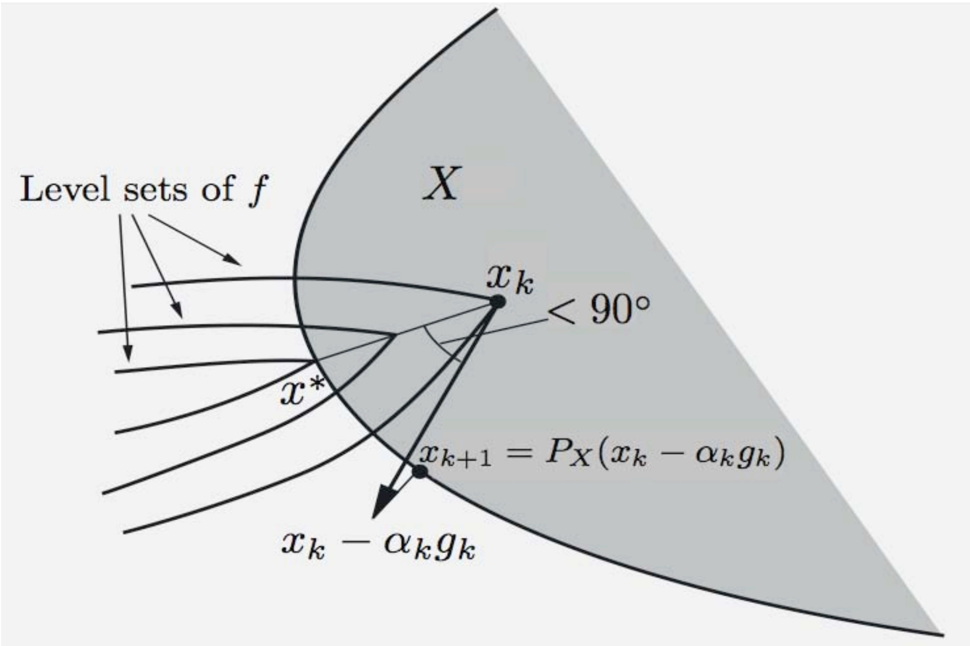
$$x_{k+1} = P_X(x_k - \alpha_k g_k),$$

where g_k is **any** subgradient of f at x_k, α_k is a positive stepsize, and $P_X(\cdot)$ is projection on X.



KEY PROPERTY OF SUBGRADIENT METHOD

- For a small enough stepsize α_k, it reduces the Euclidean distance to the optimum.



- **Proposition:** Let $\{x_k\}$ be generated by the subgradient method. Then, for all $y \in X$ and k:

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

and if $f(y) < f(x_k)$,

$$\|x_{k+1} - y\| < \|x_k - y\|,$$

for all α_k such that

$$0 < \alpha_k < \frac{2\big(f(x_k) - f(y)\big)}{\|g_k\|^2}.$$

PROOF

- **Proof of nonexpansive property**

$$\|P_X(x) - P_X(y)\| \le \|x - y\|, \qquad \forall\, x, y \in \Re^n.$$

Use the projection theorem to write

$$\big(z - P_X(x)\big)'\big(x - P_X(x)\big) \le 0, \qquad \forall\, z \in X$$

from which $\big(P_X(y) - P_X(x)\big)'\big(x - P_X(x)\big) \le 0$. Similarly, $\big(P_X(x) - P_X(y)\big)'\big(y - P_X(y)\big) \le 0$.

Adding and using the Schwarz inequality,

$$\begin{aligned} \big\|P_X(y) - P_X(x)\big\|^2 &\le \big(P_X(y) - P_X(x)\big)'(y - x) \\ &\le \big\|P_X(y) - P_X(x)\big\| \cdot \|y - x\| \end{aligned}$$

Q.E.D.

- **Proof of proposition:** Since projection is nonexpansive, we obtain for all $y \in X$ and k,

$$\begin{aligned} \|x_{k+1} - y\|^2 &= \big\|P_X\left(x_k - \alpha_k g_k\right) - y\big\|^2 \\ &\le \|x_k - \alpha_k g_k - y\|^2 \\ &= \|x_k - y\|^2 - 2\alpha_k g_k'(x_k - y) + \alpha_k^2\|g_k\|^2 \\ &\le \|x_k - y\|^2 - 2\alpha_k\big(f(x_k) - f(y)\big) + \alpha_k^2\|g_k\|^2, \end{aligned}$$

where the last inequality follows from the subgradient inequality. **Q.E.D.**

CONVERGENCE MECHANISM

- Assume constant stepsize: $\alpha_k \equiv \alpha$
- If $\|g_k\| \le c$ for some constant c and all k,

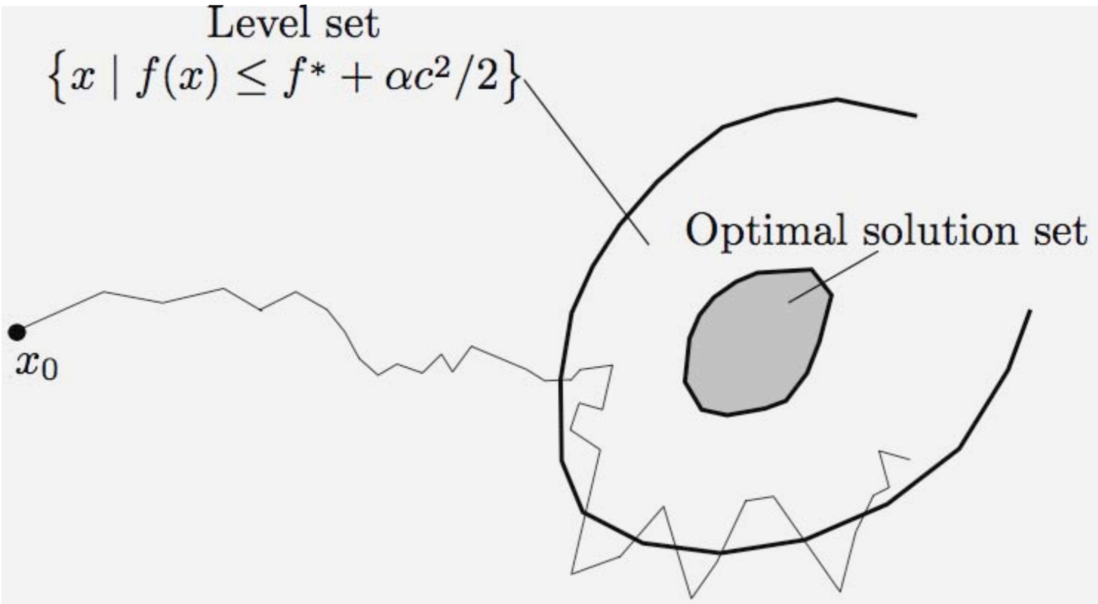
$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2 - 2\alpha\big(f(x_k)-f(x^*)\big) + \alpha^2 c^2$$

so the distance to the optimum decreases if

$$0 < \alpha < \frac{2\big(f(x_k)-f(x^*)\big)}{c^2}$$

or equivalently, if x_k does not belong to the level set

$$\left\{x \;\middle|\; f(x) < f(x^*) + \frac{\alpha c^2}{2}\right\}$$



STEPSIZE RULES

- **Constant Stepsize:** $\alpha_k \equiv \alpha$.
- **Diminishing Stepsize:** $\alpha_k \to 0$, $\sum_k \alpha_k = \infty$
- **Dynamic Stepsize:**

$$\alpha_k = \frac{f(x_k) - f_k}{c^2}$$

where f_k is 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 $\{x \mid f(x) \leq f_k\}$.
  - f_k can be adjusted based on the progress of the method.

- **Example of dynamic stepsize rule:**

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

and δ_k (the "aspiration level of cost reduction") is updated according to

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

where $\delta > 0$, $\beta < 1$, and $\rho \geq 1$ are fixed constants.

SAMPLE CONVERGENCE RESULTS

- Let $\overline{f} = \inf_{k\geq 0} f(x_k)$, and assume that for some c, we have

$$c \geq \sup_{k\geq 0} \big\{ \|g\| \mid g \in \partial f(x_k) \big\}.$$

- **Proposition:** Assume that α_k is fixed at some positive scalar α. Then:

  (a) If $f^* = -\infty$, then $\overline{f} = f^*$.

  (b) If $f^* > -\infty$, then

$$\overline{f} \leq f^* + \frac{\alpha c^2}{2}.$$

- **Proposition:** If α_k satisfies

$$\lim_{k\to\infty} \alpha_k = 0, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty,$$

then $\overline{f} = f^*$.

- Similar propositions for dynamic stepsize rules.
- Many variants ...