LECTURE 18

LECTURE OUTLINE

- Proximal algorithm
- Convergence
- Rate of convergence
- Extensions

****

Consider minimization of closed proper convex $f : \Re^n \mapsto (-\infty, +\infty]$ using a different type of approximation:

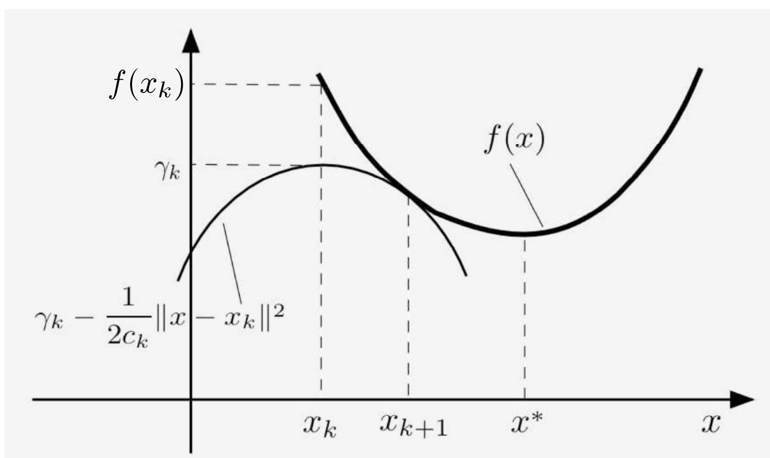
- Regularization in place of linearization
- Add a quadratic term to f to make it strictly convex and "well-behaved"
- Refine the approximation at each iteration by changing the quadratic term

PROXIMAL MINIMIZATION ALGORITHM

- A general algorithm for convex fn minimization

$$x_{k+1} \in \arg\min_{x \in \Re^n} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

  - $f : \Re^n \mapsto (-\infty, \infty]$ is closed proper convex
  - c_k is a positive scalar parameter
  - x_0 is arbitrary starting point



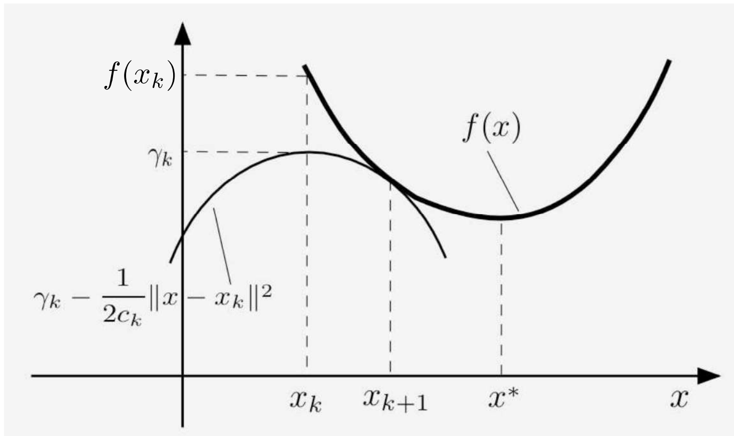
- x_{k+1} exists because of the quadratic.

- Note it does not have the instability problem of cutting plane method

- If x_k is optimal, $x_{k+1} = x_k$.

- Main Convergence Theorem: If $\sum_k c_k = \infty$, $f(x_k) \to f^*$. Moreover $\{x_k\}$ converges to an optimal solution if one exists.

CONVERGENCE: SOME BASIC PROPERTIES



- Note the connection with Fenchel framework

- From subdifferential of sum formula (or Fenchel duality theorem)

$$(x_k - x_{k+1})/c_k \in \partial f(x_{k+1})$$

Note the similarity with the subgradient method $(x_k - x_{k+1})/c_k \in \partial f(x_k)$

- Cost improves:

$$f(x_{k+1}) + \frac{1}{2c_k}\|x_{k+1} - x_k\|^2 \leq f(x_k)$$

- Distance to the optimum improves:

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2c_k\big(f(x_{k+1}) - f(y)\big) - \|x_k - x_{k+1}\|^2$$

for all k and $y \in \Re^n$.

CONVERGENCE PROOF I

- Main Convergence Theorem: If $\sum_k c_k = \infty$, $f(x_k) \downarrow f^*$. Moreover $\{x_k\}$ converges to an optimal solution if one exists.

Proof: Have $f(x_k) \downarrow f_\infty \geq f^*$. For all y and k,

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2c_k\big(f(x_{k+1}) - f(y)\big)$$

By adding over $k = 0, \ldots, N$,

$$\|x_{N+1} - y\|^2 + 2\sum_{k=0}^{N} c_k\big(f(x_{k+1}) - f(y)\big) \leq \|x_0 - y\|^2,$$

so taking the limit as $N \to \infty$,

$$2\sum_{k=0}^{\infty} c_k\big(f(x_{k+1}) - f(y)\big) \leq \|x_0 - y\|^2 \qquad (*)$$

- Argue by contradiction: Assume $f_\infty > f^*$, and let $\hat{y}$ be such that $f_\infty > f(\hat{y}) > f^*$. Then

$$f(x_{k+1}) - f(\hat{y}) \geq f_\infty - f(\hat{y}) > 0.$$

Since $\sum_{k=0}^{\infty} c_k = \infty$, (*) leads to a contradiction. Thus $f_\infty = f^*$.

CONVERGENCE PROOF II

- Assume $X^* \neq \emptyset$. We will show convergence to some $x^* \in X^*$. Applying

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - 2c_k\big(f(x_{k+1}) - f(y)\big)$$

with $y = x^* \in X^*$,

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - 2c_k\big(f(x_{k+1}) - f(x^*)\big), \tag{**}$$

Thus $\|x_k - x^*\|^2$ is monotonically nonincreasing, so $\{x_k\}$ is bounded.

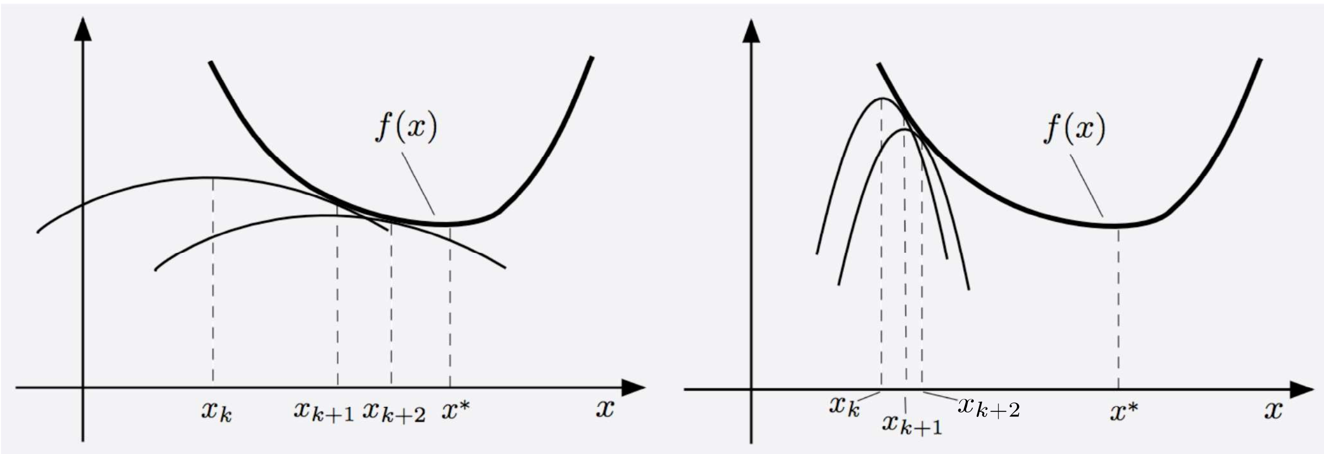
- If $\{x_k\}_\mathcal{K} \to \overline{z}$, the limit point $\overline{z}$ must belong to X^*, since $f(x_k) \downarrow f^*$, and f is closed, so

$$f(\overline{z}) \leq \liminf_{k\to\infty,\, k\in\mathcal{K}} f(x_k) = f^*$$

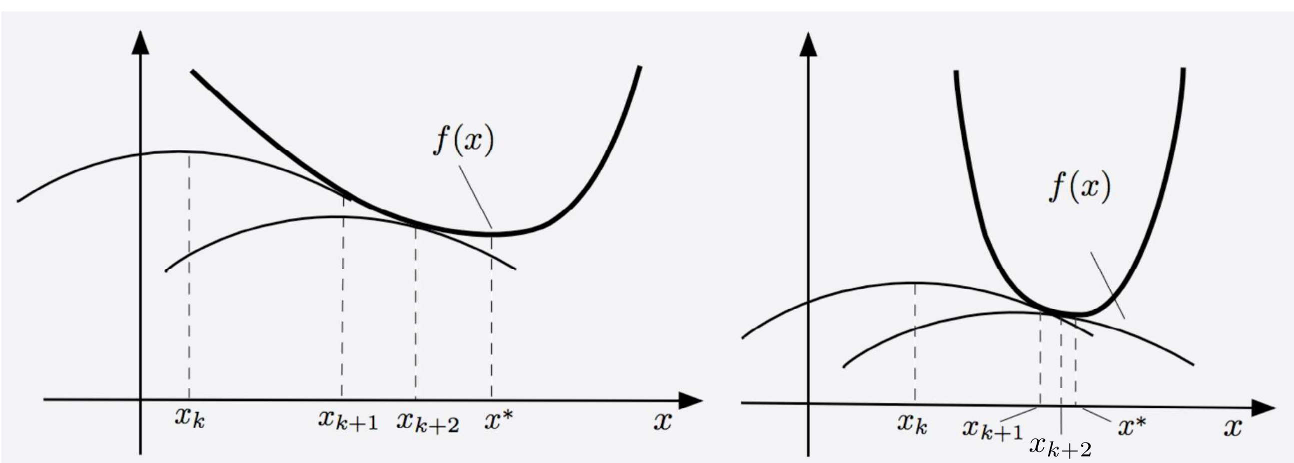
- By (**), the distance of x_k to each limit point is monotonically nonincreasing, so $\{x_k\}$ must converge to a unique limit, which must be an element of X^*. **Q.E.D.**

RATE OF CONVERGENCE I

- Role of penalty parameter c_k:



- Role of growth properties of f near optimal solution set:



RATE OF CONVERGENCE II

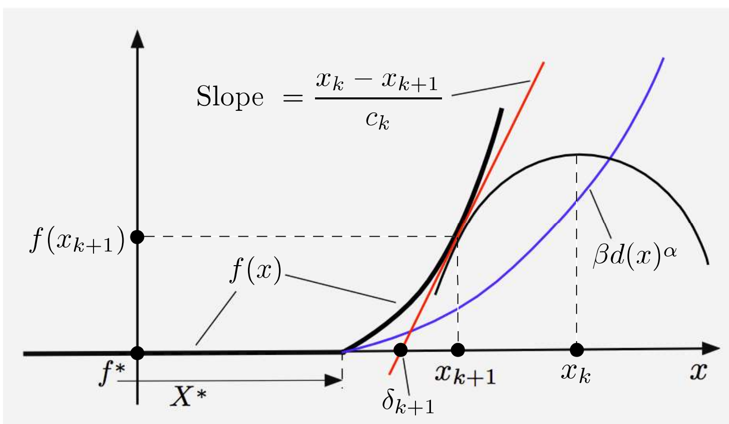
- Assume growth of order α from optimal solution set X^*, i.e., that for some $\beta > 0$, $\delta > 0$, and $\alpha \geq 1$,

$$f^* + \beta\big(d(x)\big)^\alpha \leq f(x), \quad \forall\, x \in \Re^n \ \text{ with } d(x) \leq \delta$$

where $d(x) = \min_{x^* \in X^*} \|x - x^*\|$

- Key property: For all k sufficiently large,

$$d(x_{k+1}) + \beta c_k \big(d(x_{k+1})\big)^{\alpha-1} \leq d(x_k)$$



- We have (in one dimension)

$$\begin{aligned}
\beta\big(d(x_{k+1})\big)^\alpha &\leq f(x_{k+1}) - f^* \\
&= \frac{x_k - x_{k+1}}{c_k} \cdot (x_{k+1} - \delta_{k+1}) \\
&\leq \frac{d(x_k) - d(x_{k+1})}{c_k} \cdot d(x_{k+1})
\end{aligned}$$

LINEAR AND SUPERLINEAR CONVERGENCE

- Use the key relation

$$d(x_{k+1}) + \beta c_k \big(d(x_{k+1})\big)^{\alpha-1} \leq d(x_k)$$

for various values of order of growth $\alpha \geq 1$.

- If $\alpha = 2$ and $\lim_{k\to\infty} c_k = \bar{c}$, then

$$\limsup_{k\to\infty} \frac{d(x_{k+1})}{d(x_k)} \leq \frac{1}{1+\beta\bar{c}}$$

linear convergence.

- If $1 < \alpha < 2$, then

$$\limsup_{k\to\infty} \frac{d(x_{k+1})}{\big(d(x_k)\big)^{1/(\alpha-1)}} < \infty$$

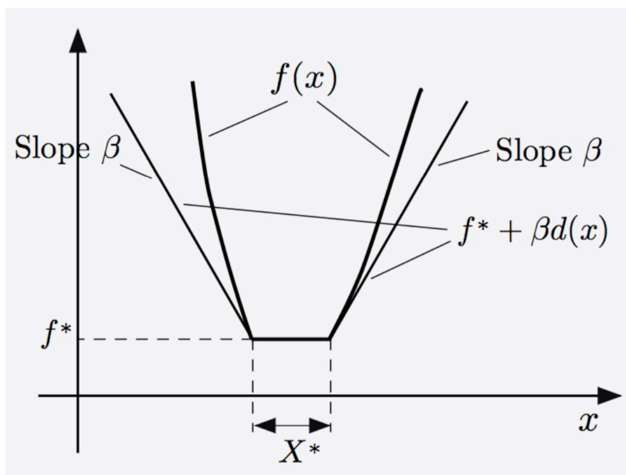
superlinear convergence.

FINITE CONVERGENCE

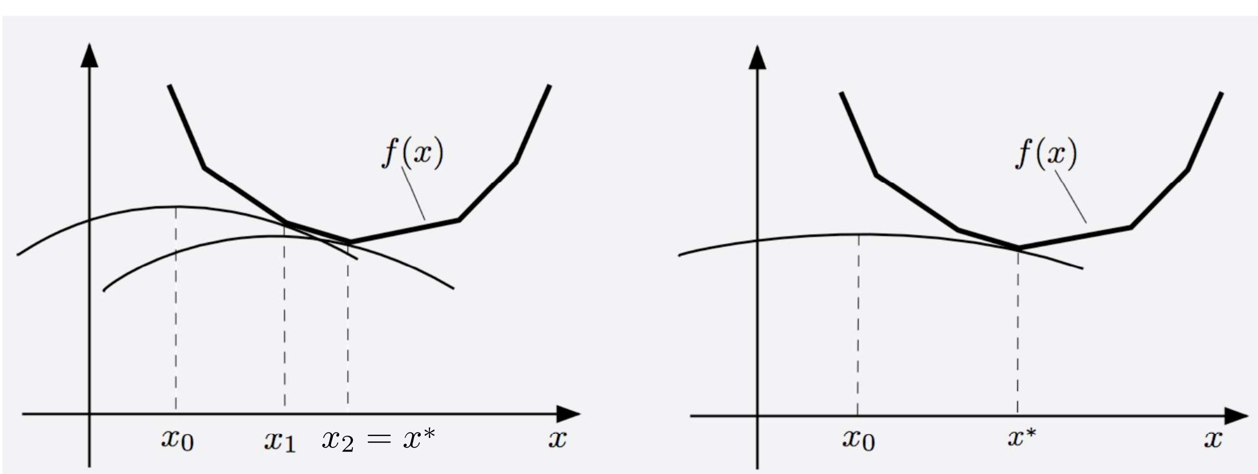
- Assume growth order $\alpha = 1$:

$$f^* + \beta d(x) \leq f(x), \qquad \forall\ x \in \Re^n$$

Can be shown to hold if f is polyhedral.



- Method converges finitely (in a single step for c_0 sufficiently large).



EXTENSIONS

- Combine with polyhedral approximation of f, to take advantage of finite convergence property.
  - Leads to bundle methods, which involve a mechanism to prevent the inherent instability of cutting plane method.
- Extension to more general problems:
  - Application to variational inequalities and games.
  - Application to finding a zero of a "maximally monotone multi-valued" mapping.
  - Allow nonconvex f (the theory is not clean and complete).
- Replace quadratic regularization by more general proximal term.