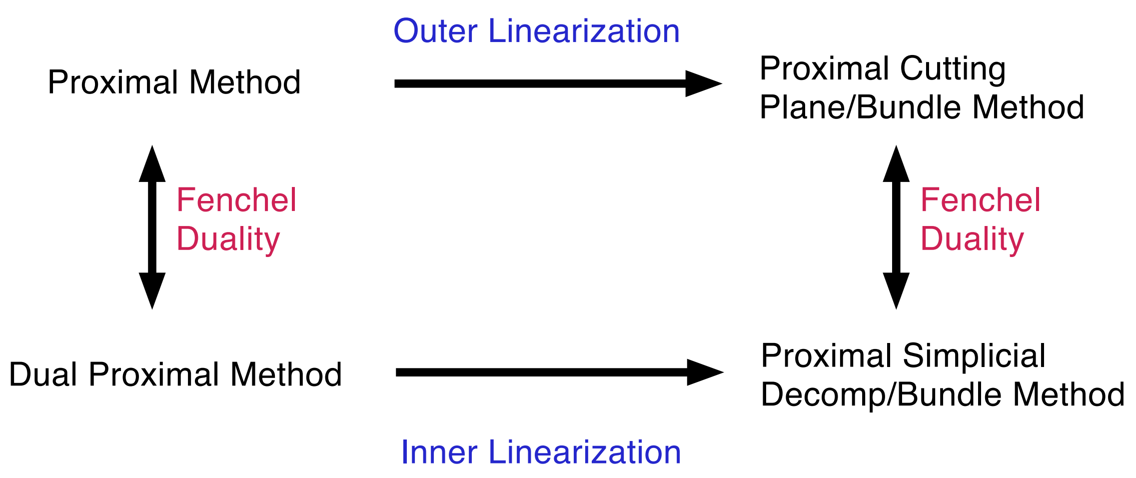
LECTURE 20

LECTURE OUTLINE

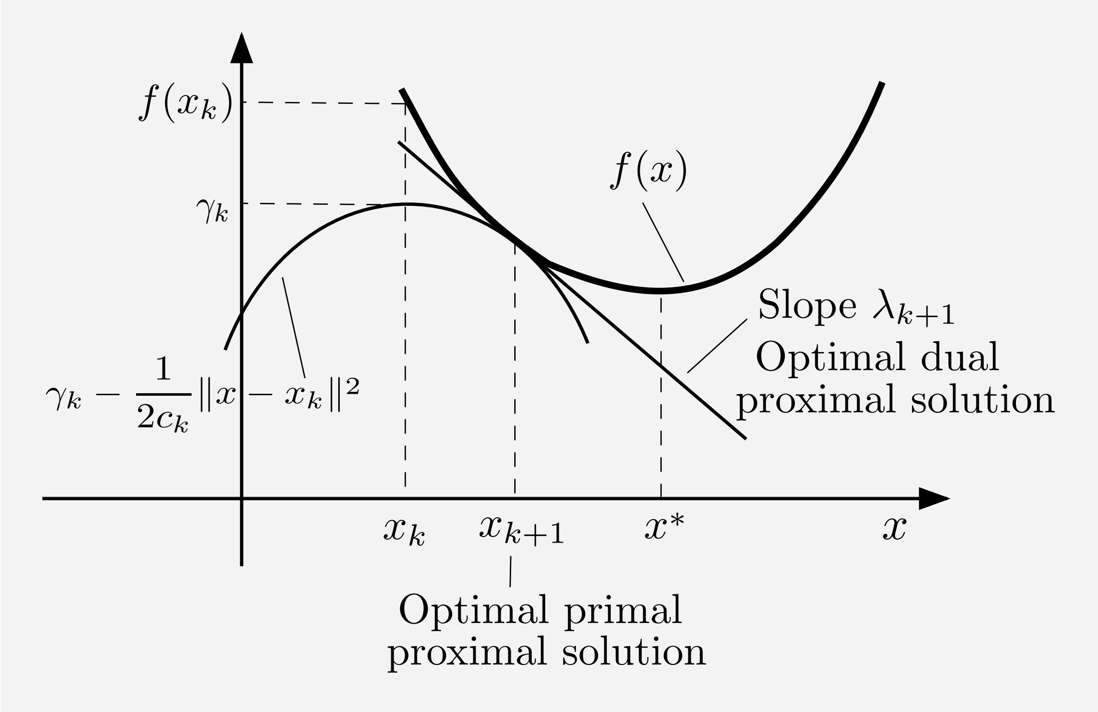
- Proximal methods
- Review of Proximal Minimization
- Proximal cutting plane algorithm
- Bundle methods
- Augmented Lagrangian Methods
- Dual Proximal Minimization Algorithm

****

- Method relationships to be established:



RECALL PROXIMAL MINIMIZATION



- Minimizes closed convex proper f:

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

where x_0 is an arbitrary starting point, and $\{c_k\}$ is a positive parameter sequence.

- We have $f(x_k) \to f^*$. Also $x_k \to$ some minimizer of f, provided one exists.
- Finite convergence for polyhedral f.
- Each iteration can be viewed in terms of Fenchel duality.

PROXIMAL/BUNDLE METHODS

- Replace f with a cutting plane approx. and/or change quadratic regularization more conservatively.
- A general form:

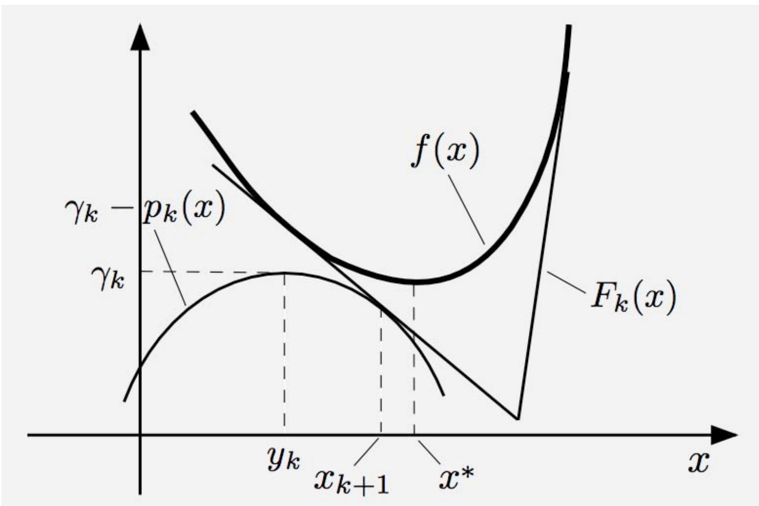
$$x_{k+1} \in \arg\min_{x \in X} \left\{ F_k(x) + p_k(x) \right\}$$

$$F_k(x) = \max\left\{ f(x_0) + (x - x_0)' g_0, \ldots, f(x_k) + (x - x_k)' g_k \right\}$$

$$p_k(x) = \frac{1}{2c_k} \|x - y_k\|^2$$

where c_k is a positive scalar parameter.

- We refer to $p_k(x)$ as the *proximal term*, and to its center y_k as the *proximal center*.



Change y_k in different ways => different methods.

PROXIMAL CUTTING PLANE METHODS

- Keeps moving the proximal center at each iteration ($y_k = x_k$)

- Same as proximal minimization algorithm, but f **is replaced by a cutting plane approximation** F_k**:**

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

where

$$F_k(x) = \max\big\{ f(x_0) + (x - x_0)' g_0, \ldots, f(x_k) + (x - x_k)' g_k \big\}$$

- Drawbacks:

  (a) **Stability issue:** For large enough c_k and polyhedral X, x_{k+1} is the exact minimum of F_k over X in a single minimization, so it is identical to the ordinary cutting plane method. For small c_k convergence is slow.

  (b) **The number of subgradients used in** F_k **may become very large;** the quadratic program may become very time-consuming.

- These drawbacks motivate algorithmic variants, called *bundle methods*.

BUNDLE METHODS

- Allow a proximal center $y_k \neq x_k$:

$$x_{k+1} \in \arg\min_{x \in X} \left\{ F_k(x) + p_k(x) \right\}$$

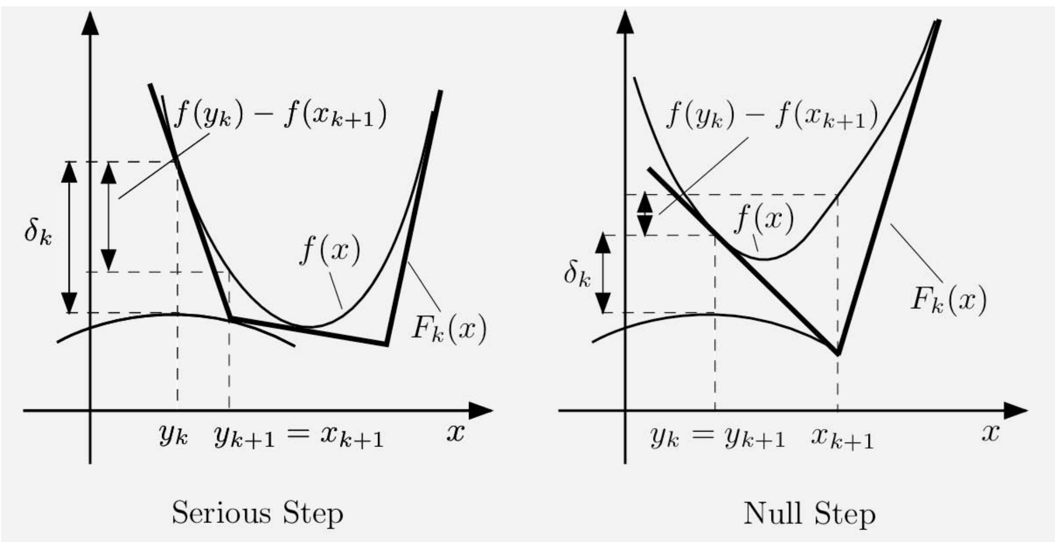
$$F_k(x) = \max\left\{ f(x_0)+(x-x_0)'g_0, \ldots, f(x_k)+(x-x_k)'g_k \right\}$$

$$p_k(x) = \frac{1}{2c_k}\|x - y_k\|^2$$

- **Null/Serious test** for changing y_k: For some fixed $\beta \in (0,1)$

$$y_{k+1} = \begin{cases} x_{k+1} & \text{if } f(y_k) - f(x_{k+1}) \geq \beta\delta_k, \\ y_k & \text{if } f(y_k) - f(x_{k+1}) < \beta\delta_k, \end{cases}$$

$$\delta_k = f(y_k) - \big(F_k(x_{k+1}) + p_k(x_{k+1})\big) > 0$$



REVIEW OF FENCHEL DUALITY

- Consider the problem

$$\begin{aligned} &\text{minimize} \quad f_1(x) + f_2(x) \\ &\text{subject to} \;\; x \in \Re^n, \end{aligned}$$

where f_1 and f_2 are closed proper convex.

- **Duality Theorem:**

  (a) If f^* is finite and $\mathrm{ri}\big(\mathrm{dom}(f_1)\big) \cap \mathrm{ri}\big(\mathrm{dom}(f_2)\big) \neq \emptyset$, then strong duality holds and there exists at least one dual optimal solution.

  (b) Strong duality holds, and (x^*, λ^*) is a primal and dual optimal solution pair if and only if

$$x^* \in \arg\min_{x \in \Re^n} \big\{ f_1(x) - x'\lambda^* \big\}, \;\; x^* \in \arg\min_{x \in \Re^n} \big\{ f_2(x) + x'\lambda^* \big\}$$

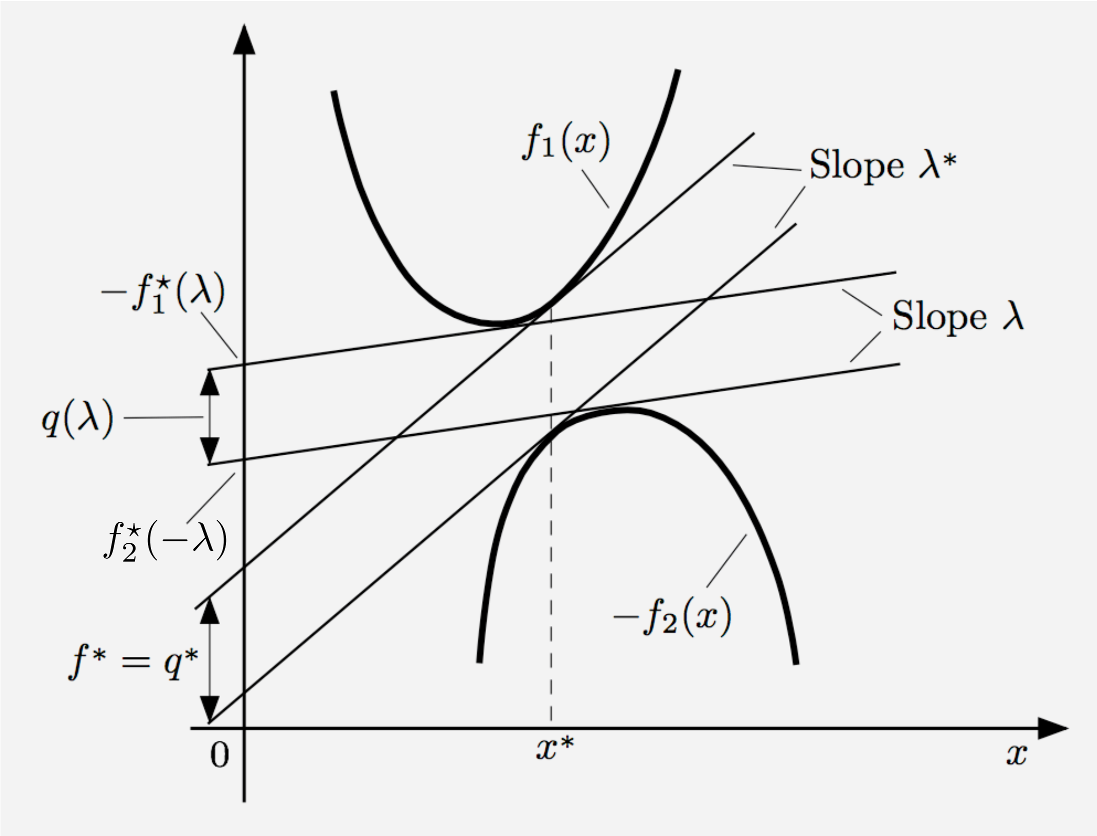
- By Fenchel inequality, the last condition is equivalent to

$$\lambda^* \in \partial f_1(x^*) \qquad [\text{or equivalently } x^* \in \partial f_1^\star(\lambda^*)]$$

and

$$-\lambda^* \in \partial f_2(x^*) \qquad [\text{or equivalently } x^* \in \partial f_2^\star(-\lambda^*)]$$

GEOMETRIC INTERPRETATION



- When f_1 and/or f_2 are differentiable, the optimality condition is equivalent to

$$\lambda^* = \nabla f_1(x^*) \quad \text{and/or} \quad \lambda^* = -\nabla f_2(x^*)$$

DUAL PROXIMAL MINIMIZATION

- The proximal iteration can be written in the Fenchel form: $\min_x \{f_1(x) + f_2(x)\}$ with

$$f_1(x) = f(x), \qquad f_2(x) = \frac{1}{2c_k}\|x - x_k\|^2$$

- The Fenchel dual is

$$\begin{aligned} &\text{minimize} \quad f_1^\star(\lambda) + f_2^\star(-\lambda) \\ &\text{subject to} \ \ \lambda \in \Re^n \end{aligned}$$

- We have $f_2^\star(-\lambda) = -x_k'\lambda + \frac{c_k}{2}\|\lambda\|^2$, so the dual problem is

$$\begin{aligned} &\text{minimize} \quad f^\star(\lambda) - x_k'\lambda + \frac{c_k}{2}\|\lambda\|^2 \\ &\text{subject to} \ \ \lambda \in \Re^n \end{aligned}$$

where $f^\star$ is the conjugate of f.

- f_2 is real-valued, so no duality gap.
- Both primal and dual problems have a unique solution, since they involve a closed, strictly convex, and coercive cost function.

DUAL PROXIMAL ALGORITHM

- Can solve the Fenchel-dual problem instead of the primal at each iteration:

$$\lambda_{k+1} = \arg \min_{\lambda \in \Re^n} \left\{ f^\star(\lambda) - x_k'\lambda + \frac{c_k}{2}\|\lambda\|^2 \right\} \qquad (1)$$

- Lagragian optimality conditions:

$$x_{k+1} \in \arg \max_{x \in \Re^n} \left\{ x'\lambda_{k+1} - f(x) \right\}$$

$$x_{k+1} = \arg \min_{x \in \Re^n} \left\{ x'\lambda_{k+1} + \frac{1}{2c_k}\|x - x_k\|^2 \right\}$$

or equivalently,

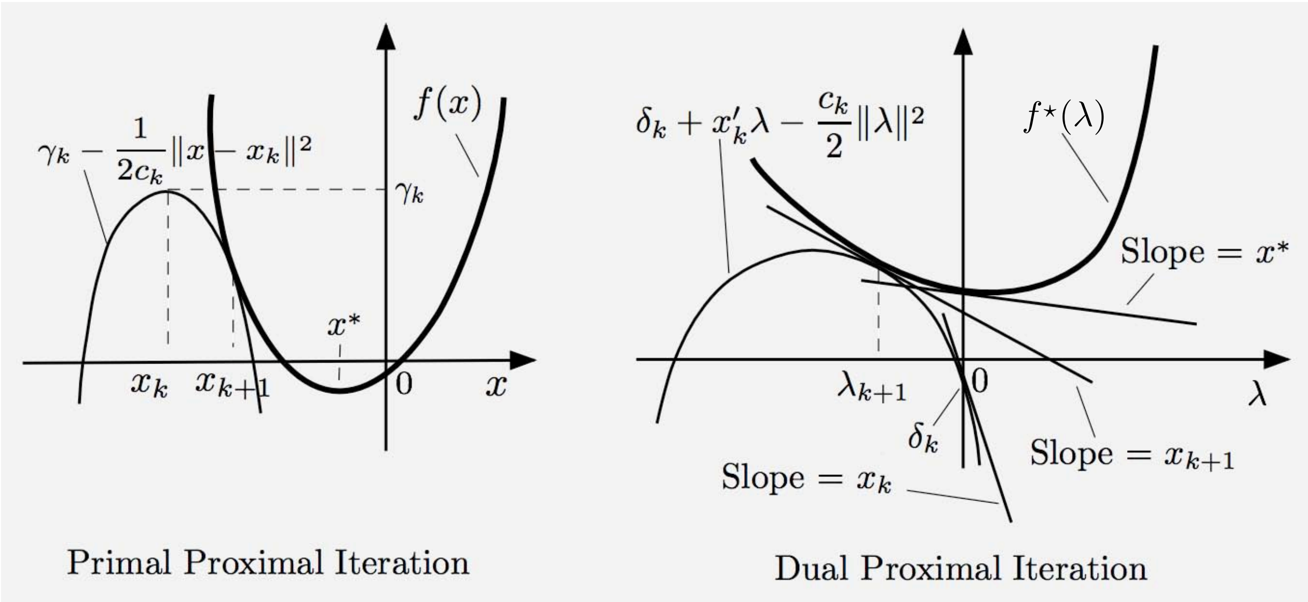
$$\lambda_{k+1} \in \partial f(x_{k+1}), \qquad \lambda_{k+1} = \frac{x_k - x_{k+1}}{c_k}$$

- **Dual algorithm:** At iteration k, obtain λ_{k+1} from the dual proximal minimization (1) and set

$$x_{k+1} = x_k - c_k \lambda_{k+1}$$

- As x_k converges to a primal optimal solution x^*, the dual sequence λ_k converges to 0 (a subgradient of f at x^*).

VISUALIZATION



- **The primal and dual implementations are mathematically equivalent and generate identical sequences $\{x_k\}$.**

- Which one is preferable depends on whether f or its conjugate $f^\star$ has more convenient structure.

- **Special case:** When $-f$ is the dual function of the constrained minimization $\min_{g(x)\le 0} F(x)$, the dual algorithm is equivalent to an important general purpose algorithm: the Augmented Lagrangian method.

- This method aims to find a subgradient of the primal function $p(u) = \min_{g(x)\le u} F(x)$ at $u = 0$ (i.e., a dual optimal solution).

AUGMENTED LAGRANGIAN METHOD

- Consider the convex constrained problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \ \ x \in X, \quad Ex = d \end{aligned}$$

- Primal and dual functions:

$$p(u) = \inf_{\substack{x \in X \\ Ex - d = u}} f(x), \quad q(\mu) = \inf_{x \in X} \big\{ f(x) + \mu'(Ex - d) \big\}$$

- Assume p: closed, so (q, p) are "conjugate" pair.
- Proximal algorithms for maximizing q:

$$\mu_{k+1} = \arg \max_{\mu \in \Re^m} \left\{ q(\mu) - \frac{1}{2c_k} \|\mu - \mu_k\|^2 \right\}$$

$$u_{k+1} = \arg \min_{u \in \Re^m} \left\{ p(u) + \mu_k' u + \frac{c_k}{2} \|u\|^2 \right\}$$

Dual update: $\mu_{k+1} = \mu_k + c_k u_{k+1}$

- Implementation:

$$u_{k+1} = Ex_{k+1} - d, \qquad x_{k+1} \in \arg \min_{x \in X} L_{c_k}(x, \mu_k)$$

where L_c is the *Augmented Lagrangian* function

$$L_c(x, \mu) = f(x) + \mu'(Ex - d) + \frac{c}{2} \|Ex - d\|^2$$

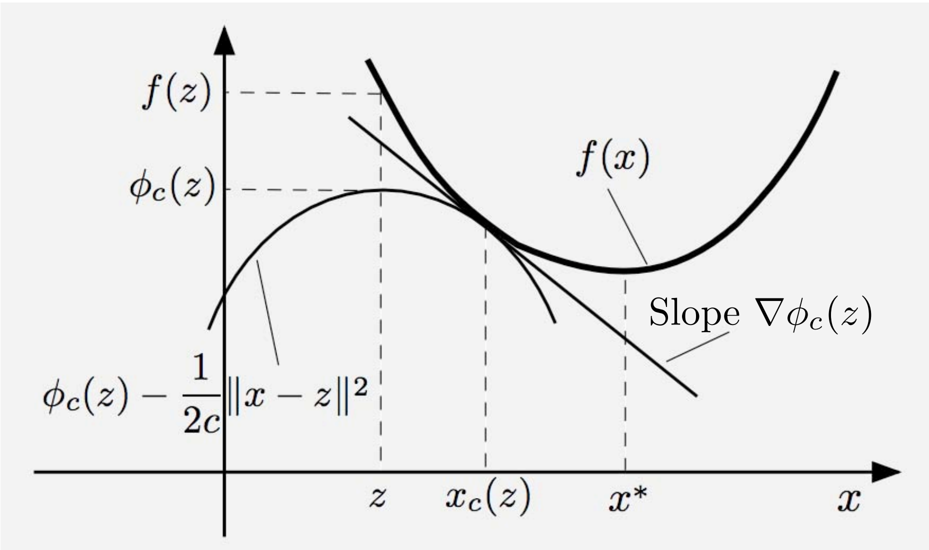
GRADIENT INTERPRETATION

- Back to the dual proximal algorithm and the dual update $\lambda_{k+1} = \frac{x_k - x_{k+1}}{c_k}$
- **Proposition:** λ_{k+1} can be viewed as a gradient:

$$\lambda_{k+1} = \frac{x_k - x_{k+1}}{c_k} = \nabla \phi_{c_k}(x_k),$$

where

$$\phi_c(z) = \inf_{x \in \Re^n} \left\{ f(x) + \frac{1}{2c} \|x - z\|^2 \right\}$$



- So the dual update $x_{k+1} = x_k - c_k \lambda_{k+1}$ can be viewed as a gradient iteration for minimizing $\phi_c(z)$ (which has the same minima as f).
- The gradient is calculated by the dual proximal minimization. Possibilities for faster methods (e.g., Newton, Quasi-Newton). Useful in augmented Lagrangian methods.

PROXIMAL LINEAR APPROXIMATION

- **Convex problem:** Min $f : \Re^n \mapsto \Re$ over X.

- **Proximal outer linearization method:** Same as proximal minimization algorithm, but f is replaced by a cutting plane approximation F_k:

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

$$\lambda_{k+1} = \frac{x_k - x_{k+1}}{c_k}$$

where $g_i \in \partial f(x_i)$ for $i \leq k$ and

$$F_k(x) = \max\Big\{ f(x_0)+(x-x_0)'g_0, \ldots, f(x_k)+(x-x_k)'g_k \Big\} + \delta_X(x)$$

- **Proximal Inner Linearization Method (Dual proximal implementation):** Let $F_k^\star$ be the conjugate of F_k. Set

$$\lambda_{k+1} \in \arg \min_{\lambda \in \Re^n} \left\{ F_k^\star(\lambda) - x_k'\lambda + \frac{c_k}{2} \|\lambda\|^2 \right\}$$

$$x_{k+1} = x_k - c_k \lambda_{k+1}$$

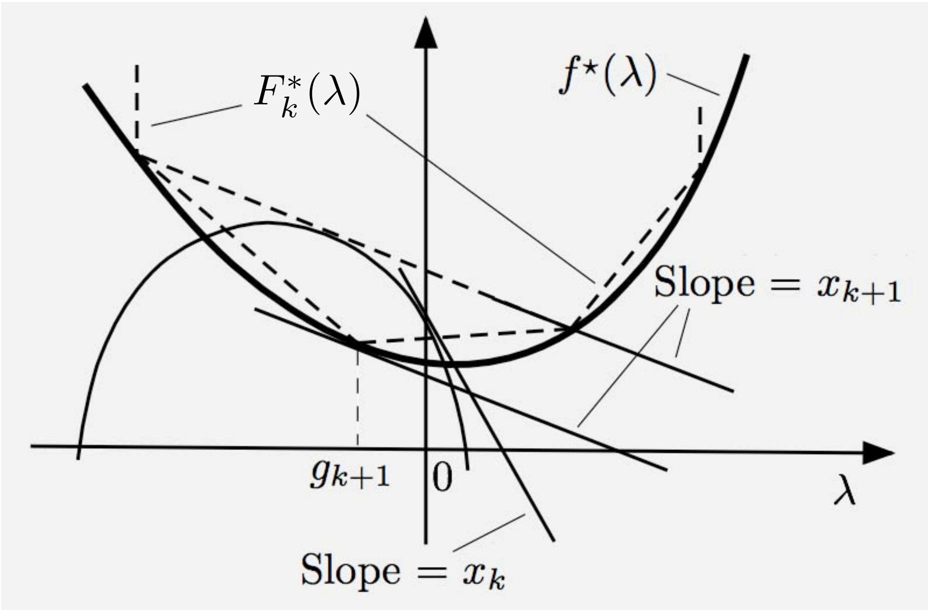
Obtain $g_{k+1} \in \partial f(x_{k+1})$, either directly or via

$$g_{k+1} \in \arg \max_{\lambda \in \Re^n} \left\{ x_{k+1}'\lambda - f^\star(\lambda) \right\}$$

- Add g_{k+1} to the outer linearization, or x_{k+1} to the inner linearization, and continue.

PROXIMAL INNER LINEARIZATION

- It is a mathematical equivalent dual to the outer linearization method.



- Here we use the conjugacy relation between outer and inner linearization.

- Versions of these methods where the proximal center is changed only after some "algorithmic progress" is made:
  - The outer linearization version is the (standard) bundle method.
  - The inner linearization version is an **inner approximation version of a bundle method**.

6.253 Convex Analysis and Optimization
Spring 2012