LECTURE 22

LECTURE OUTLINE

- Review of Fenchel Duality
- Review of Proximal Minimization
- Dual Proximal Minimization Algorithm
- Augmented Lagrangian Methods

All figures are courtesy of Athena Scientific, and are used with permission.

FENCHEL DUALITY FRAMEWORK

- Consider the problem

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

where $f_1 : \Re^n \mapsto (-\infty, \infty]$ and $f_2 : \Re^n \mapsto (-\infty, \infty]$ are closed proper convex functions.

- **Line of Analysis:** Convert to the equivalent problem

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) \\ \text{subject to} & x_1 = x_2, \quad x_1 \in \text{dom}(f_1),\ x_2 \in \text{dom}(f_2) \end{array}$$

- Apply convex programming duality for equality constraints and obtain the dual problem

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

where $f_1^\star$ and $f_2^\star$ are the conjugates.

- Complete symmetry of primal and dual (after a sign change to convert the dual to minimization).

FENCHEL DUALITY THEOREM

Consider the Fenchel framework:

(a) If f^* is finite and $\text{ri dom}(f_1) \cap \text{ri dom}(f_2) \neq \emptyset$, 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

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

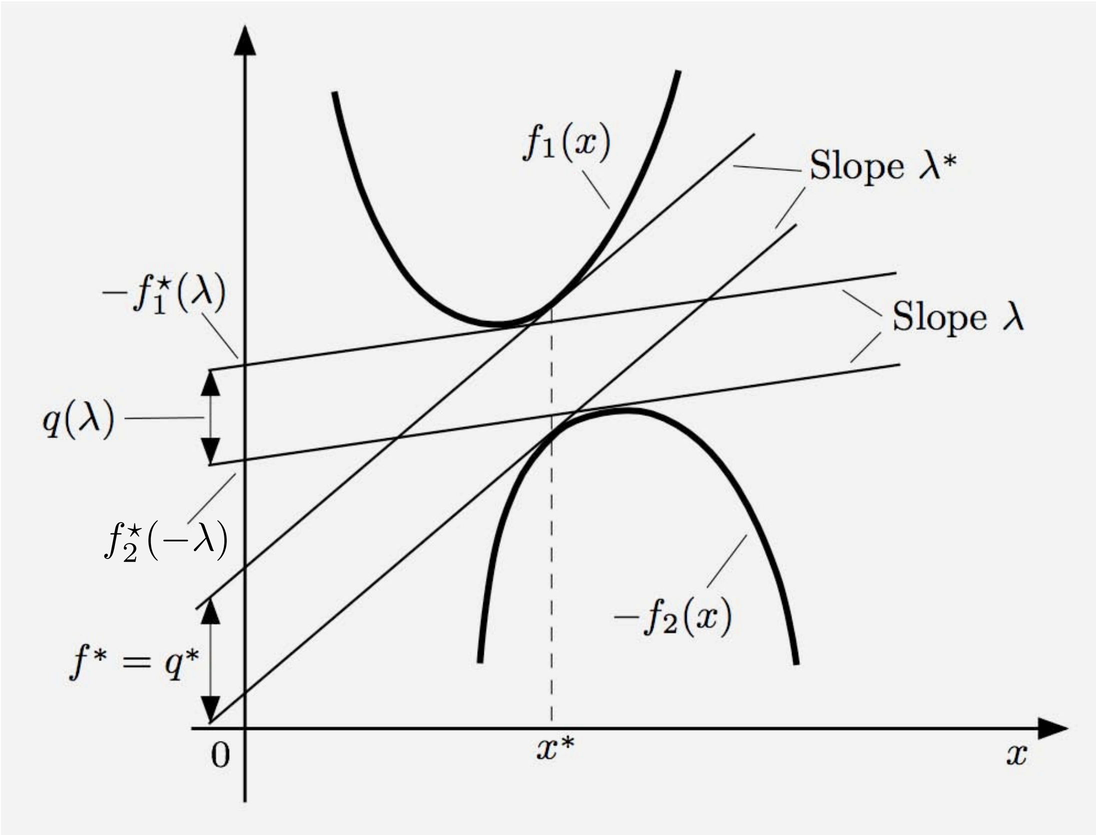
• By Fenchel inequality, the last condition is equivalent to

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

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

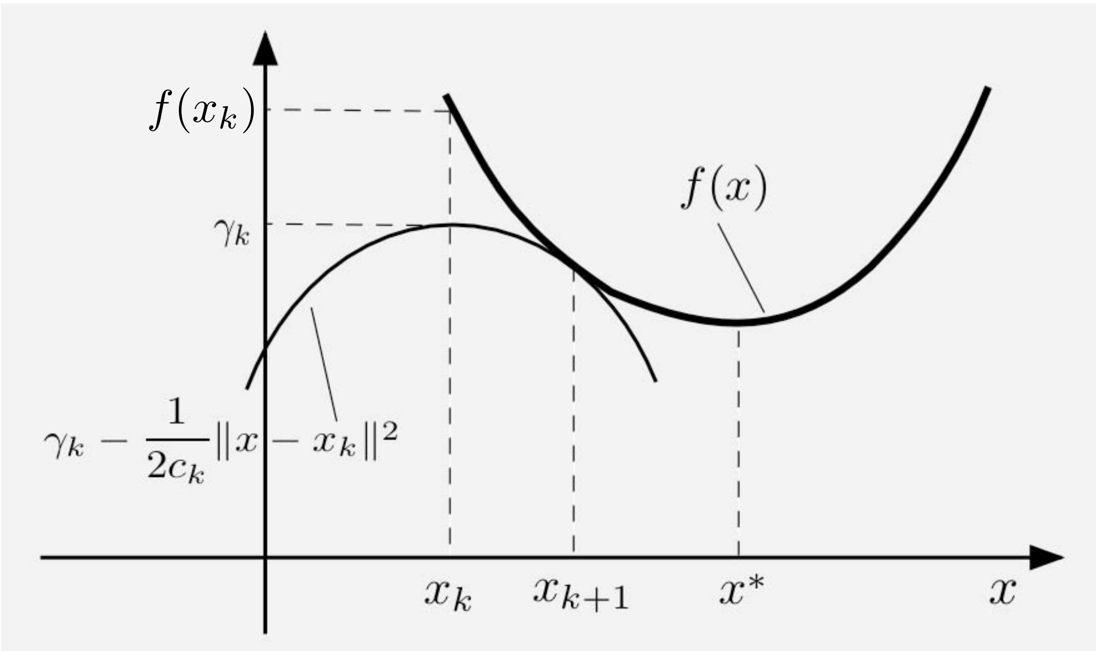
$$\lambda^* = \nabla f_1(x^*) \quad \text{and/or} \quad \lambda^* = -\nabla f_2(x^*)$$

RECALL PROXIMAL MINIMIZATION

- Applies to minimization of closed convex proper f:

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

where $f : \Re^n \mapsto (-\infty, \infty]$, x_0 is an arbitrary starting point, and $\{c_k\}$ is a positive scalar parameter sequence with $\inf_{k \geq 0} c_k > 0$.



- We have $f(x_k) \to f^*$. Also $x_k \to$ some minimizer of f, provided one exists.

- Finite convergence for polyhedral f.

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

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

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

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

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:

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

- Lagragian optimality conditions:

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

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

or equivalently,

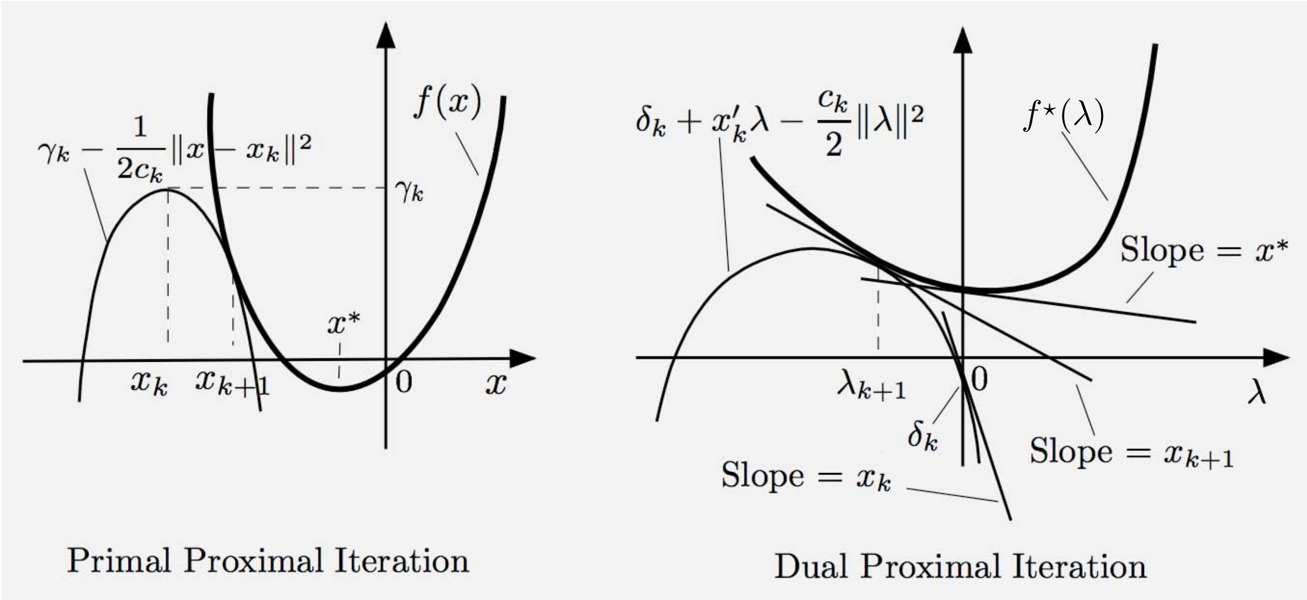
$$\lambda_{k+1} \in \partial f(x_{k+1}), \qquad \lambda_{k+1} = \frac{x_k - x_{k+1}}{c_k}$$

- **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 $\min_{g(x)\leq 0} F(x)$, the dual algorithm is equivalent to an important general purpose algorithm: the Augmented Lagrangian method.

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

AUGMENTED LAGRANGIAN METHOD

- Consider the convex constrained problem

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

- Primal and dual functions:

$$p(v) = \inf_{\substack{x \in X, \\ Ex - d = v}} f(x), \ \ q(\lambda) = \inf_{x \in X} f(x) + \lambda'(Ex - d)$$

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

$$\lambda_{k+1} = \arg \max_{\mu \in \Re^m} \ q(\lambda) - \frac{1}{2c_k} \|\lambda - \lambda_k\|^2$$

$$v_{k+1} = \arg \min_{v \in \Re^m} \ p(v) + \lambda_k' v + \frac{c_k}{2} \|v\|^2$$

Dual update: $\lambda_{k+1} = \lambda_k + c_k v_{k+1}$

- Implementation:

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

where L_c is the *Augmented Lagrangian* function

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

GRADIENT INTERPRETATION

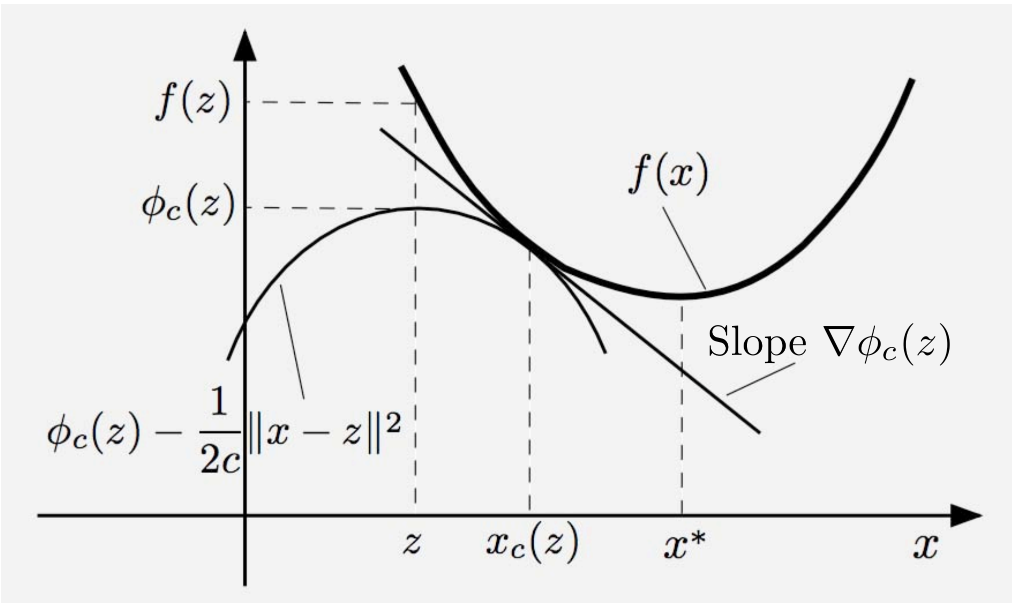
- λ_{k+1} can be viewed as a gradient:

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

where

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

(For geometrical insight, consider the case where f is linear in the following figure.)



- 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:

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

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

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

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

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

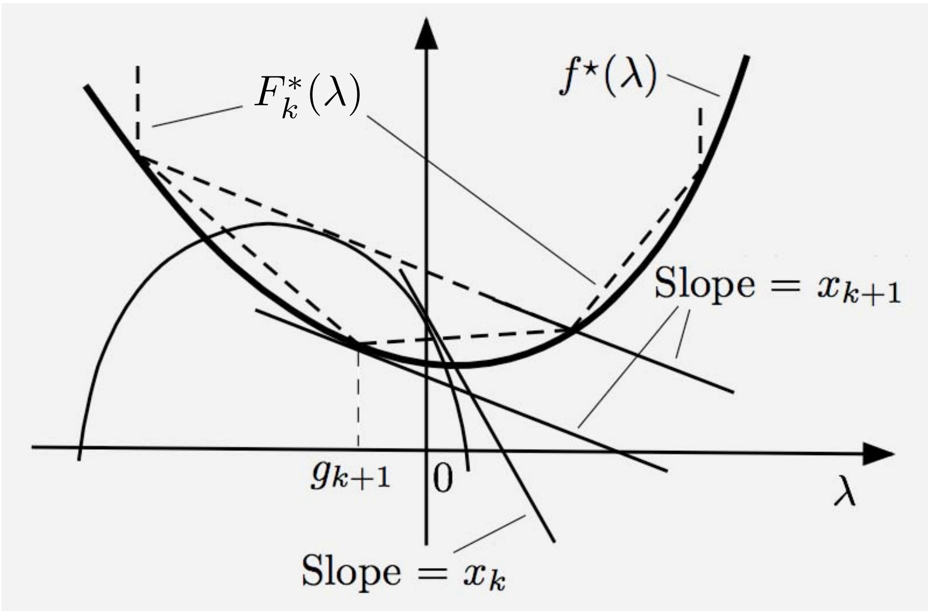
Obtain $g_{k+1} \in \partial f(x_{k+1})$, either directly or via

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

- 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**.

MIT OpenCourseWare
http://ocw.mit.edu

6.253 Convex Analysis and Optimization
Spring 2010

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.