LECTURE 25: REVIEW/EPILOGUE

LECTURE OUTLINE

CONVEX ANALYSIS AND DUALITY

- Basic concepts of convex analysis
- Basic concepts of convex optimization
- Geometric duality framework - MC/MC
- Constrained optimization duality
- Subgradients - Optimality conditions

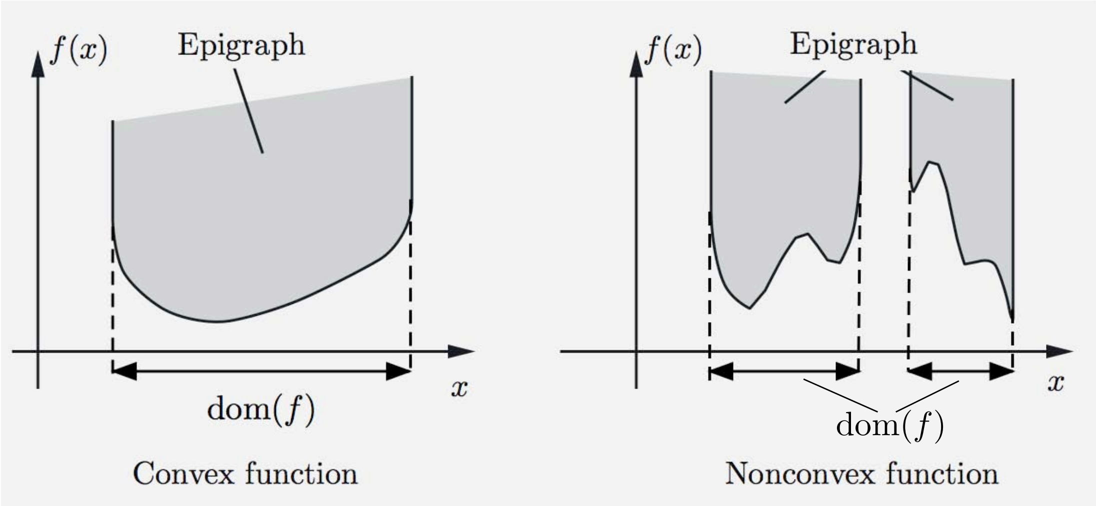
****

CONVEX OPTIMIZATION ALGORITHMS

- Special problem classes
- Subgradient methods
- Polyhedral approximation methods
- Proximal methods
- Dual proximal methods - Augmented Lagrangeans
- Interior point methods
- Incremental methods
- Optimal complexity methods
- Various combinations around proximal idea and generalizations

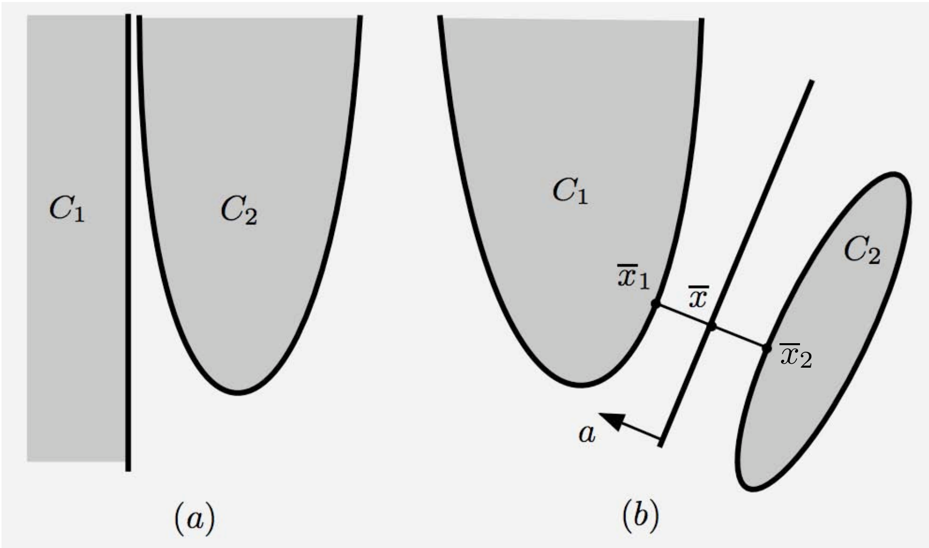
BASIC CONCEPTS OF CONVEX ANALYSIS

- Epigraphs, level sets, closedness, semicontinuity

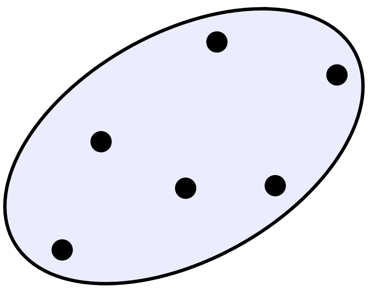


- Finite representations of generated cones and convex hulls - Caratheodory's Theorem.
- Relative interior:
  - Nonemptiness for a convex set
  - Line segment principle
  - Calculus of relative interiors
- Continuity of convex functions
- Nonemptiness of intersections of nested sequences of closed sets.
- Closure operations and their calculus.
- Recession cones and their calculus.
- Preservation of closedness by linear transformations and vector sums.

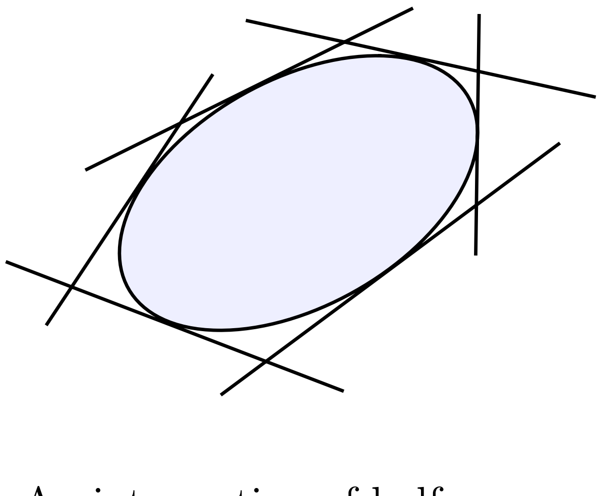
HYPERPLANE SEPARATION



- Separating/supporting hyperplane theorem.
- Strict and proper separation theorems.
- Dual representation of closed convex sets as unions of points and intersection of halfspaces.



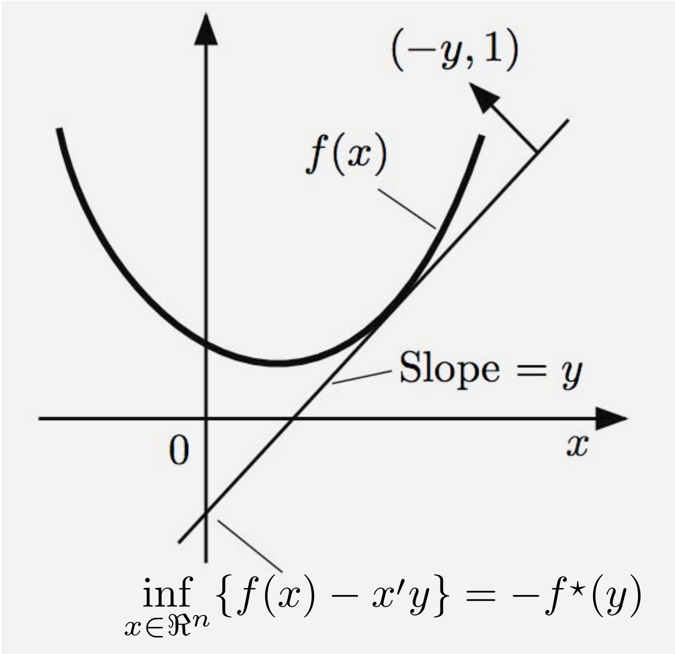
A union of points



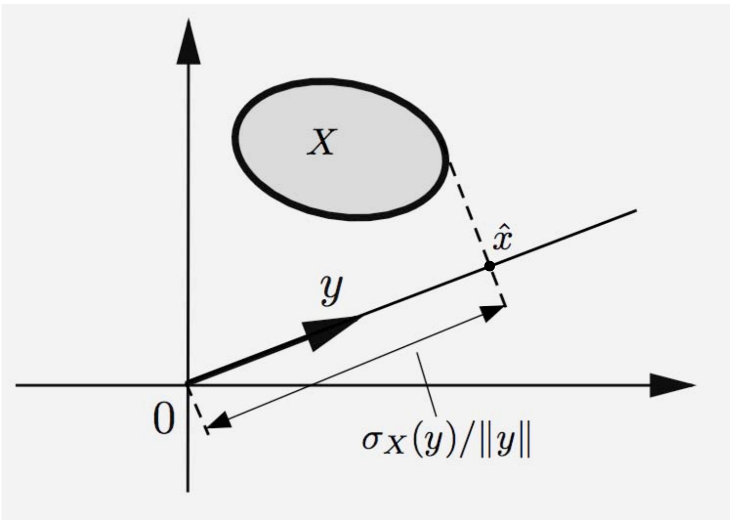
An intersection of halfspaces

- Nonvertical separating hyperplanes.

CONJUGATE FUNCTIONS



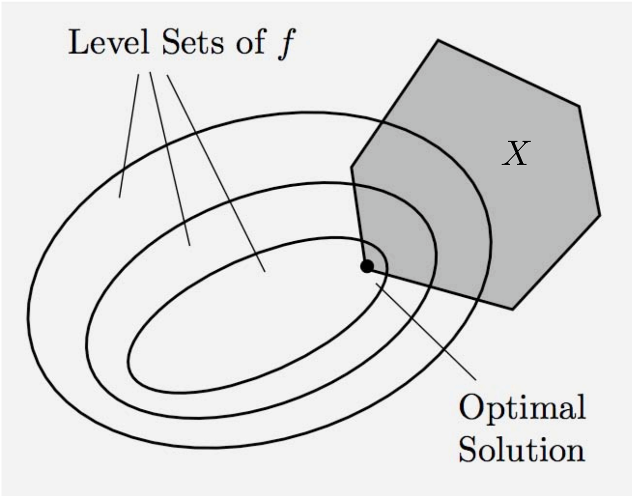
- Conjugacy theorem: $f = f^{\star\star}$
- Support functions



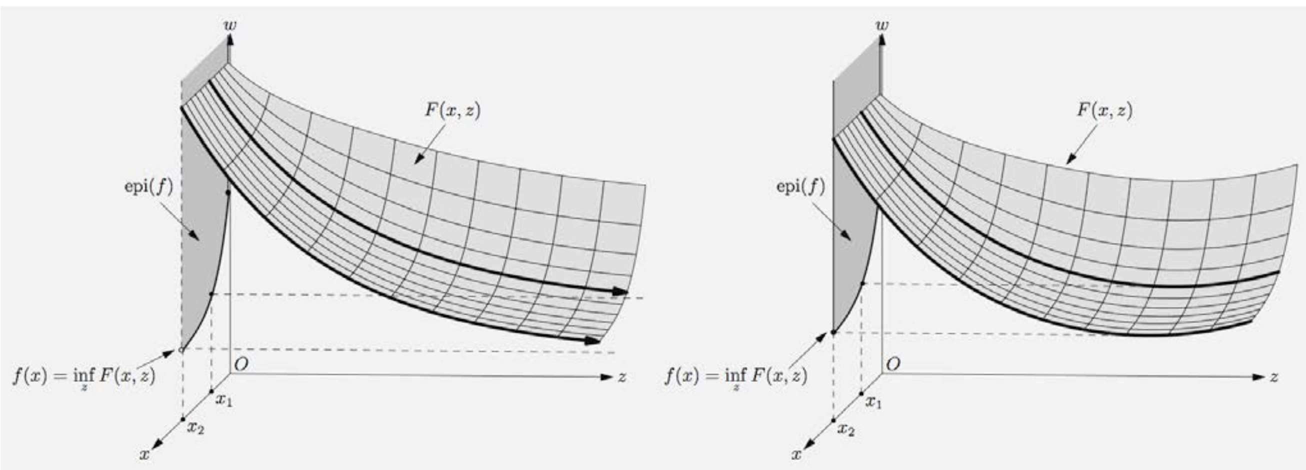
- Polar cone theorem: $C = C^{\star\star}$
  - Special case: Linear Farkas' lemma

BASIC CONCEPTS OF CONVEX OPTIMIZATION

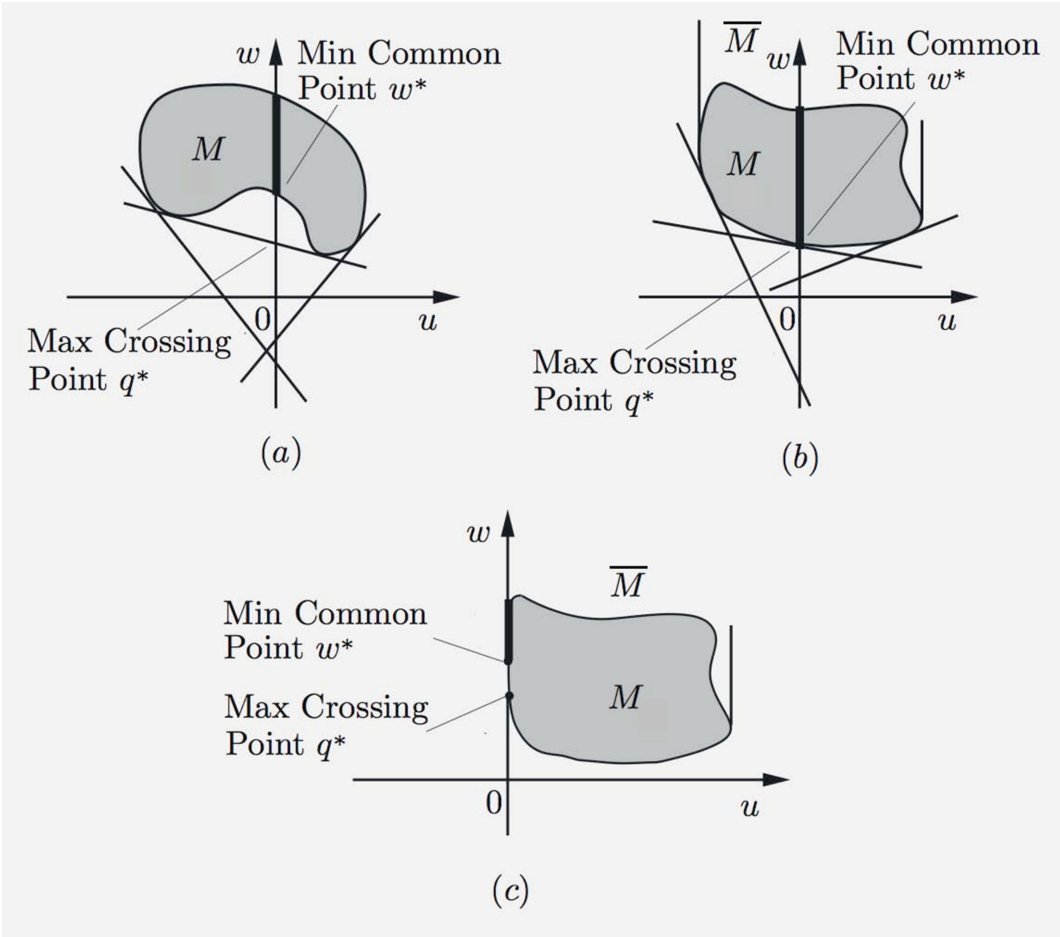
- **Weierstrass Theorem** and extensions.
- Characterization of existence of solutions in terms of nonemptiness of nested set intersections.



- Role of recession cone and lineality space.
- **Partial Minimization Theorems:** Characterization of closedness of $f(x) = \inf_{z\in\Re^m} F(x,z)$ in terms of closedness of F.



MIN COMMON/MAX CROSSING DUALITY



- Defined by a single set $M \subset \Re^{n+1}$.
- $w^* = \inf_{(0,w)\in M} w$
- $q^* = \sup_{\mu\in\Re^n} q(\mu) \stackrel{\triangle}{=} \inf_{(u,w)\in M}\{w + \mu' u\}$
- Weak duality: $q^* \leq w^*$
- Two key questions:
  - When does strong duality $q^* = w^*$ hold?
  - When do there exist optimal primal and dual solutions?

MC/MC THEOREMS ($\overline{M}$ CONVEX, $W^* < \infty$)

- **MC/MC Theorem I:** We have $q^* = w^*$ if and only if for every sequence $\big\{(u_k, w_k)\big\} \subset M$ with $u_k \to 0$, there holds

$$w^* \le \liminf_{k\to\infty} w_k.$$

- **MC/MC Theorem II:** Assume in addition that $-\infty < w^*$ and that

$$D = \big\{u \mid \text{there exists } w \in \Re \text{ with } (u, w) \in \overline{M}\}$$

contains the origin in its relative interior. Then $q^* = w^*$ and there exists μ such that $q(\mu) = q^*$.

- **MC/MC Theorem III:** Similar to II but involves special polyhedral assumptions.

  (1) $\overline{M}$ is a "horizontal translation" of $\tilde{M}$ by $-P$,

$$\overline{M} = \tilde{M} - \big\{(u, 0) \mid u \in P\big\},$$

where P: polyhedral and $\tilde{M}$: convex.

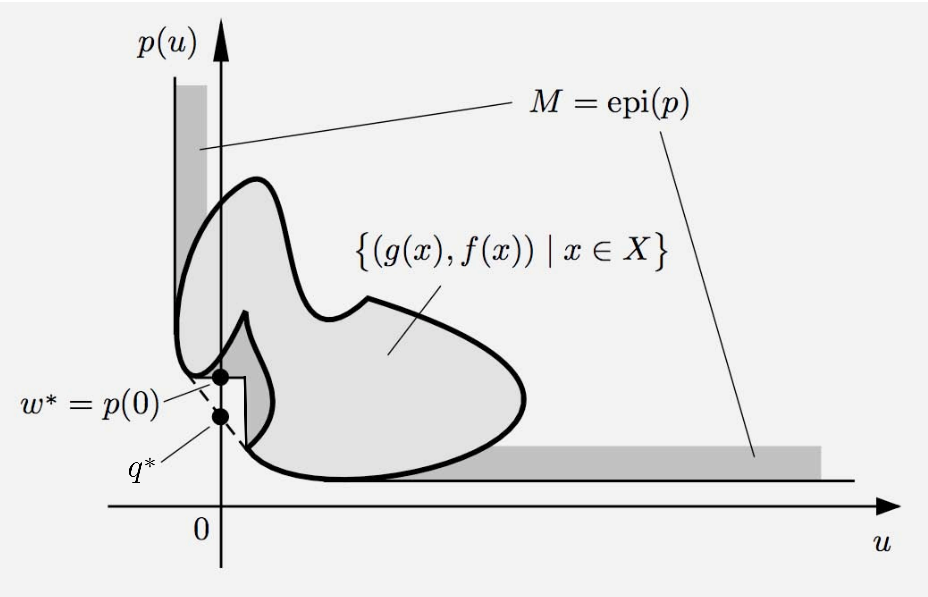
  (2) We have $\text{ri}(\tilde{D}) \cap P \neq \emptyset$, where

$$\tilde{D} = \big\{u \mid \text{there exists } w \in \Re \text{ with } (u, w) \in \tilde{M}\}$$

IMPORTANT SPECIAL CASE

- **Constrained optimization:** $\inf_{x\in X,\, g(x)\le 0} f(x)$
- Perturbation function (or *primal function*)

$$p(u) = \inf_{x\in X,\, g(x)\le u} f(x),$$



- Introduce $L(x,\mu) = f(x) + \mu' g(x)$. Then

$$
\begin{aligned}
q(\mu) &= \inf_{u\in\Re^r} \big\{p(u) + \mu' u\big\} \\
&= \inf_{u\in\Re^r,\, x\in X,\, g(x)\le u} \big\{f(x) + \mu' u\big\} \\
&= \begin{cases} \inf_{x\in X} L(x,\mu) & \text{if } \mu \ge 0, \\ -\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

NONLINEAR FARKAS' LEMMA

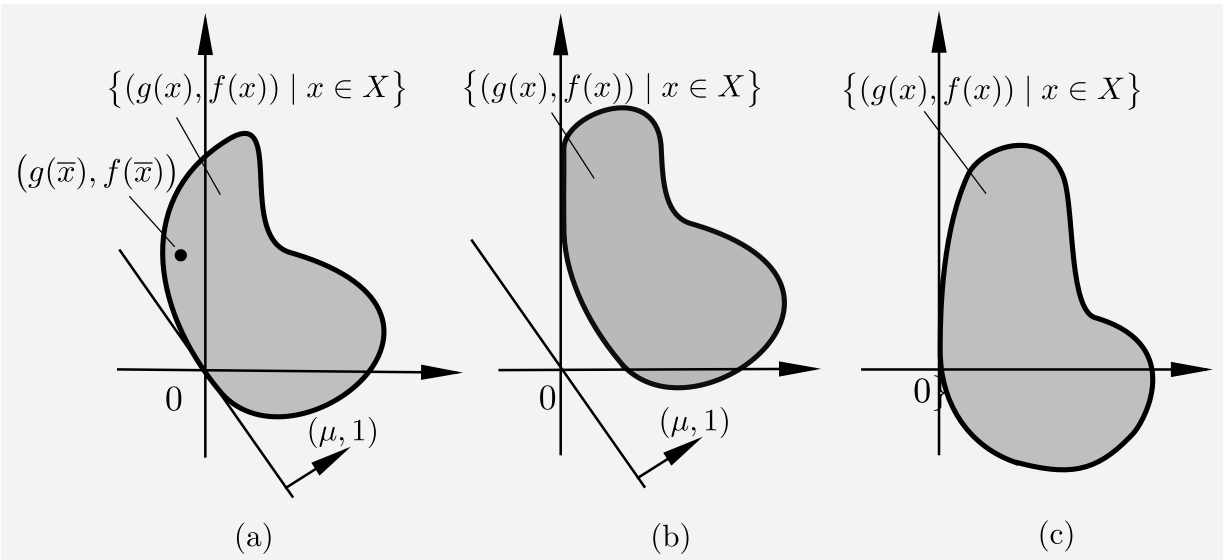
• Let $X \subset \Re^n$, $f : X \mapsto \Re$, and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, be convex. Assume that

$$f(x) \geq 0, \qquad \forall\, x \in X \text{ with } g(x) \leq 0$$

Let

$$Q^* = \big\{\mu \mid \mu \geq 0,\ f(x) + \mu' g(x) \geq 0, \forall\, x \in X\big\}.$$

• **Nonlinear version:** Then Q^* is nonempty and compact if and only if there exists a vector $\overline{x} \in X$ such that $g_j(\overline{x}) < 0$ for all $j = 1, \ldots, r$.



• **Polyhedral version:** Q^* is nonempty if g is linear $[g(x) = Ax - b]$ and there exists a vector $\overline{x} \in \mathrm{ri}(X)$ such that $A\overline{x} - b \leq 0$.

CONSTRAINED OPTIMIZATION DUALITY

$$\text{minimize } \; f(x)$$
$$\text{subject to } \; x \in X, \; g_j(x) \leq 0, \; j = 1, \ldots, r,$$

where $X \subset \Re^n$, $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$ are convex. Assume f^*: finite.

- **Connection with MC/MC:** $M = \text{epi}(p)$ with $p(u) = \inf_{x \in X,\, g(x) \leq u} f(x)$
- **Dual function:**

$$q(\mu) = \begin{cases} \inf_{x \in X} L(x, \mu) & \text{if } \mu \geq 0, \\ -\infty & \text{otherwise} \end{cases}$$

where $L(x, \mu) = f(x) + \mu' g(x)$ is the Lagrangian function.

- **Dual problem** of maximizing $q(\mu)$ over $\mu \geq 0$.
- **Strong Duality Theorem:** $q^* = f^*$ and there exists dual optimal solution if one of the following two conditions holds:

(1) There exists $\overline{x} \in X$ such that $g(\overline{x}) < 0$.

(2) The functions g_j, $j = 1, \ldots, r$, are affine, and there exists $\overline{x} \in \text{ri}(X)$ such that $g(\overline{x}) \leq 0$.

OPTIMALITY CONDITIONS

• We have $q^* = f^*$, and the vectors x^* and μ^* are optimal solutions of the primal and dual problems, respectively, iff x^* is feasible, $\mu^* \geq 0$, and

$$x^* \in \arg\min_{x \in X} L(x, \mu^*), \qquad \mu_j^* g_j(x^*) = 0, \quad \forall\ j.$$

• For the linear/quadratic program

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2} x' Q x + c' x \\ &\text{subject to} \quad Ax \leq b, \end{aligned}$$

where Q is positive semidefinite, (x^*, μ^*) is a primal and dual optimal solution pair if and only if:

(a) Primal and dual feasibility holds:

$$Ax^* \leq b, \qquad \mu^* \geq 0$$

(b) Lagrangian optimality holds [x^* minimizes $L(x, \mu^*)$ over $x \in \Re^n$]. (Unnecessary for LP.)

(c) Complementary slackness holds:

$$(Ax^* - b)' \mu^* = 0,$$

i.e., $\mu_j^* > 0$ implies that the jth constraint is tight. (Applies to inequality constraints only.)

FENCHEL DUALITY

- **Primal problem:**

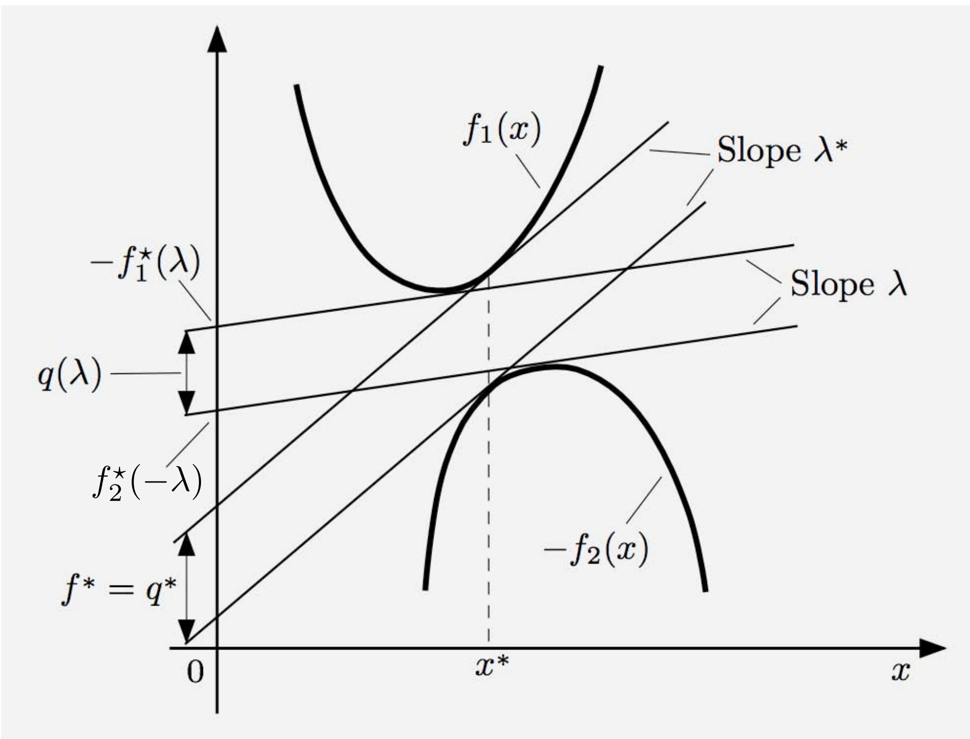
$$\begin{aligned} &\text{minimize} \quad f_1(x) + f_2(x) \\ &\text{subject to} \ \ x \in \Re^n, \end{aligned}$$

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

- **Dual problem:**

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

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



CONIC DUALITY

- Consider minimizing $f(x)$ over $x \in C$, where $f : \Re^n \mapsto (-\infty, \infty]$ is a closed proper convex function and C is a closed convex cone in $\Re^n$.
- We apply Fenchel duality with the definitions

$$f_1(x) = f(x), \qquad f_2(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

- **Linear Conic Programming:**

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \ \ x - b \in S, \quad x \in C. \end{aligned}$$

- The **dual linear conic** problem is equivalent to

$$\begin{aligned} &\text{minimize} \quad b'\lambda \\ &\text{subject to} \ \ \lambda - c \in S^{\perp}, \quad \lambda \in \hat{C}. \end{aligned}$$

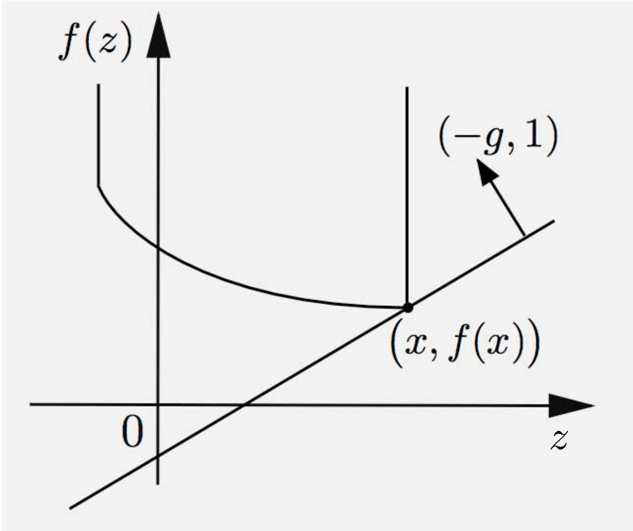
- **Special Linear-Conic Forms:**

$$\min_{Ax=b,\ x \in C} c'x \qquad \Longleftrightarrow \qquad \max_{c - A'\lambda \in \hat{C}} b'\lambda,$$

$$\min_{Ax-b \in C} c'x \qquad \Longleftrightarrow \qquad \max_{A'\lambda = c,\ \lambda \in \hat{C}} b'\lambda,$$

where $x \in \Re^n$, $\lambda \in \Re^m$, $c \in \Re^n$, $b \in \Re^m$, $A : m \times n$.

SUBGRADIENTS



- $\partial f(x) \neq \emptyset$ for $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$.

- **Conjugate Subgradient Theorem:** If f is closed proper convex, the following are equivalent for a pair of vectors (x, y):

  (i) $x'y = f(x) + f^\star(y)$.

  (ii) $y \in \partial f(x)$.

  (iii) $x \in \partial f^\star(y)$.

- **Characterization of optimal solution set** $X^* = \arg\min_{x \in \Re^n} f(x)$ of closed proper convex f:

  (a) $X^* = \partial f^\star(0)$.

  (b) X^* is nonempty if $0 \in \mathrm{ri}\big(\mathrm{dom}(f^\star)\big)$.

  (c) X^* is nonempty and compact if and only if $0 \in \mathrm{int}\big(\mathrm{dom}(f^\star)\big)$.

CONSTRAINED OPTIMALITY CONDITION

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be proper convex, let X be a convex subset of $\Re^n$, and assume that one of the following four conditions holds:

  (i) $\text{ri}\big(\text{dom}(f)\big) \cap \text{ri}(X) \neq \emptyset$.

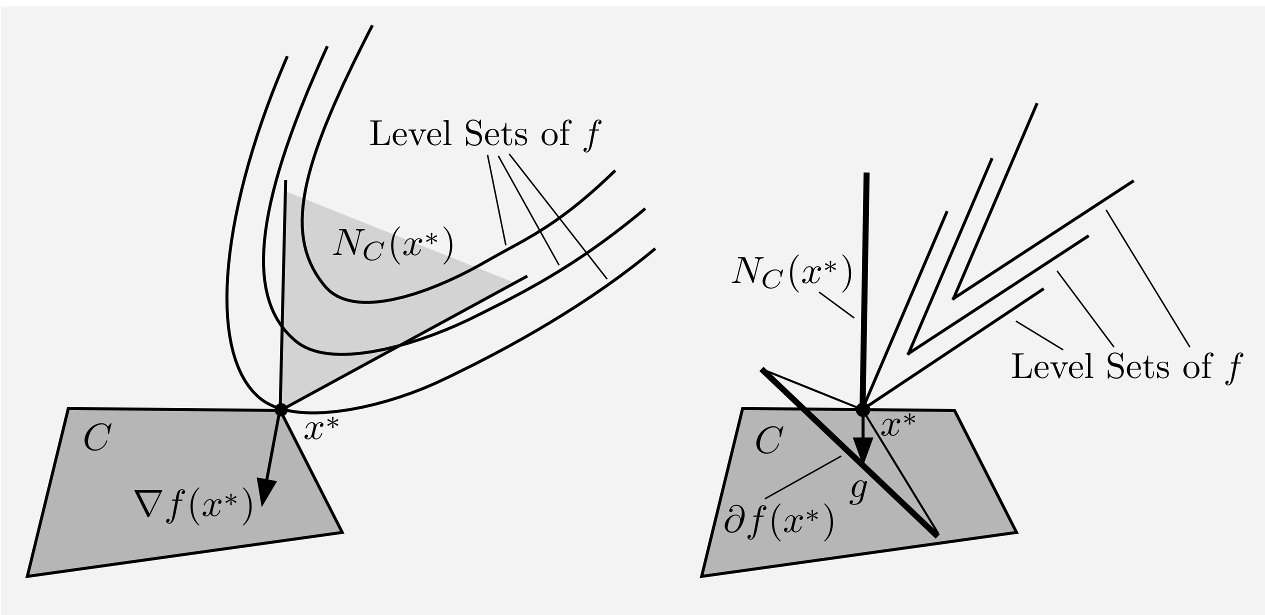
  (ii) f is polyhedral and $\text{dom}(f) \cap \text{ri}(X) \neq \emptyset$.

  (iii) X is polyhedral and $\text{ri}\big(\text{dom}(f)\big) \cap X \neq \emptyset$.

  (iv) f and X are polyhedral, and $\text{dom}(f) \cap X \neq \emptyset$.

Then, a vector x^* minimizes f over X iff there exists $g \in \partial f(x^*)$ such that $-g$ belongs to the normal cone $N_X(x^*)$, i.e.,

$$g'(x - x^*) \geq 0, \qquad \forall\, x \in X.$$

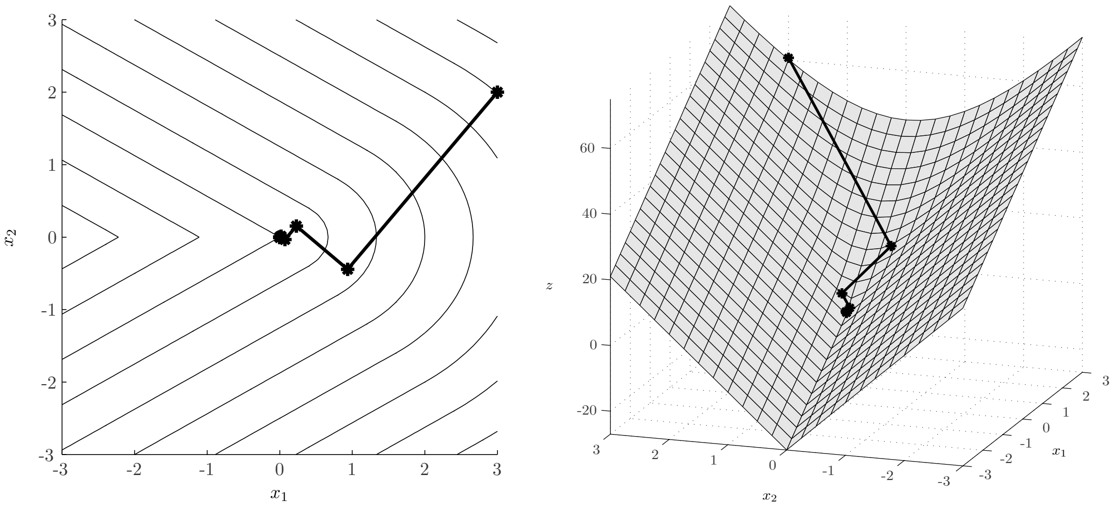


COMPUTATION: PROBLEM RANKING IN INCREASING COMPUTATIONAL DIFFICULTY

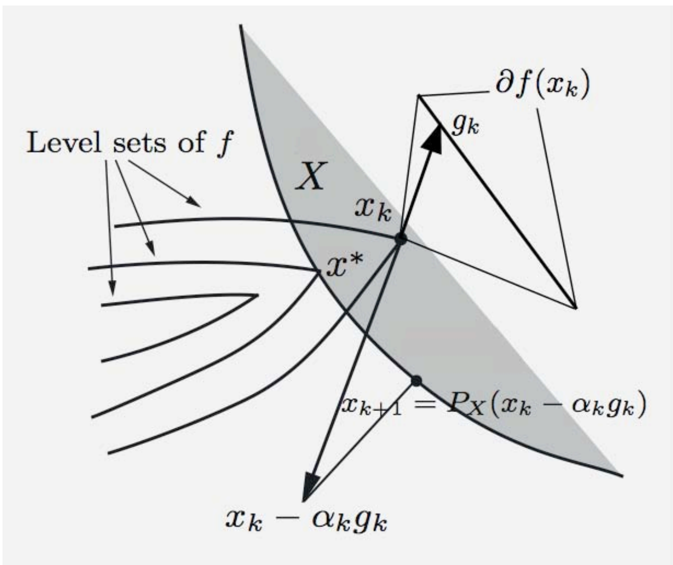
- Linear and (convex) quadratic programming.
  - Favorable special cases.
- Second order cone programming.
- Semidefinite programming.
- Convex programming.
  - Favorable cases, e.g., separable, large sum.
  - Geometric programming.
- Nonlinear/nonconvex/continuous programming.
  - Favorable special cases.
  - Unconstrained.
  - Constrained.
- Discrete optimization/Integer programming
  - Favorable special cases.
- Caveats/questions:
  - Important role of special structures.
  - What is the role of “optimal algorithms”?
  - Is complexity the right philosophical view to convex optimization?

DESCENT METHODS

- **Steepest descent method:** Use vector of min norm on $-\partial f(x)$; has convergence problems.



- **Subgradient method:**

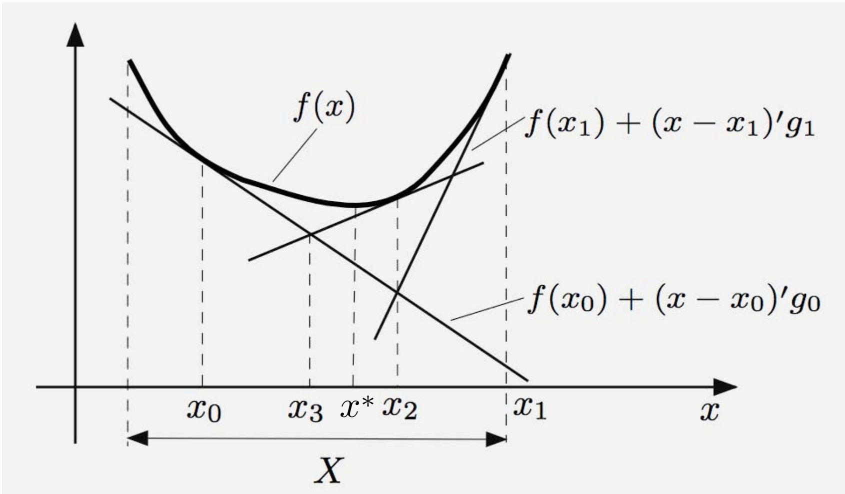


- **ϵ-subgradient method** (approx. subgradient)

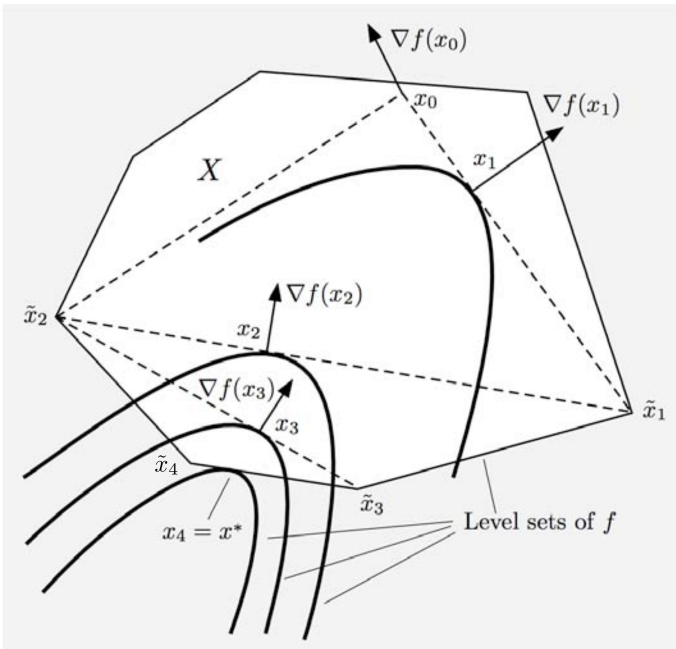
- **Incremental** (possibly randomized) variants for minimizing large sums (can be viewed as an approximate subgradient method).

OUTER AND INNER LINEARIZATION

- **Outer linearization:** Cutting plane



- **Inner linearization:** Simplicial decomposition



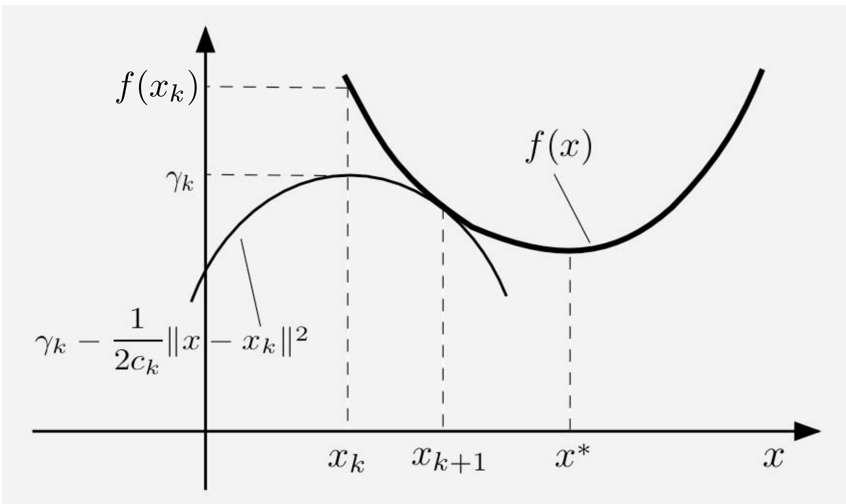
- Duality between outer and inner linearization.
  - *Extended monotropic programming* framework
  - Fenchel-like duality theory

PROXIMAL MINIMIZATION ALGORITHM

- A general algorithm for convex fn minimization

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

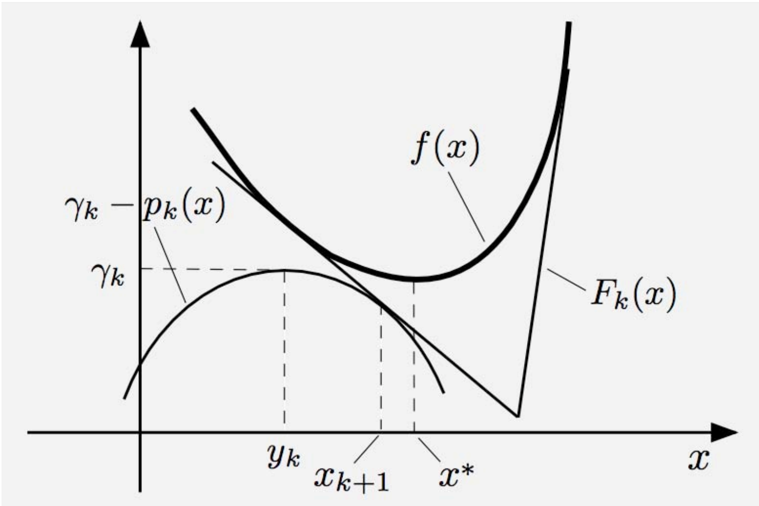
  - $f : \Re^n \mapsto (-\infty, \infty]$ is closed proper convex
  - c_k is a positive scalar parameter
  - x_0 is arbitrary starting point



- x_{k+1} exists because of the quadratic.
- Strong convergence properties
- Starting point for extensions (e.g., nonquadratic regularization) and combinations (e.g., with linearization)

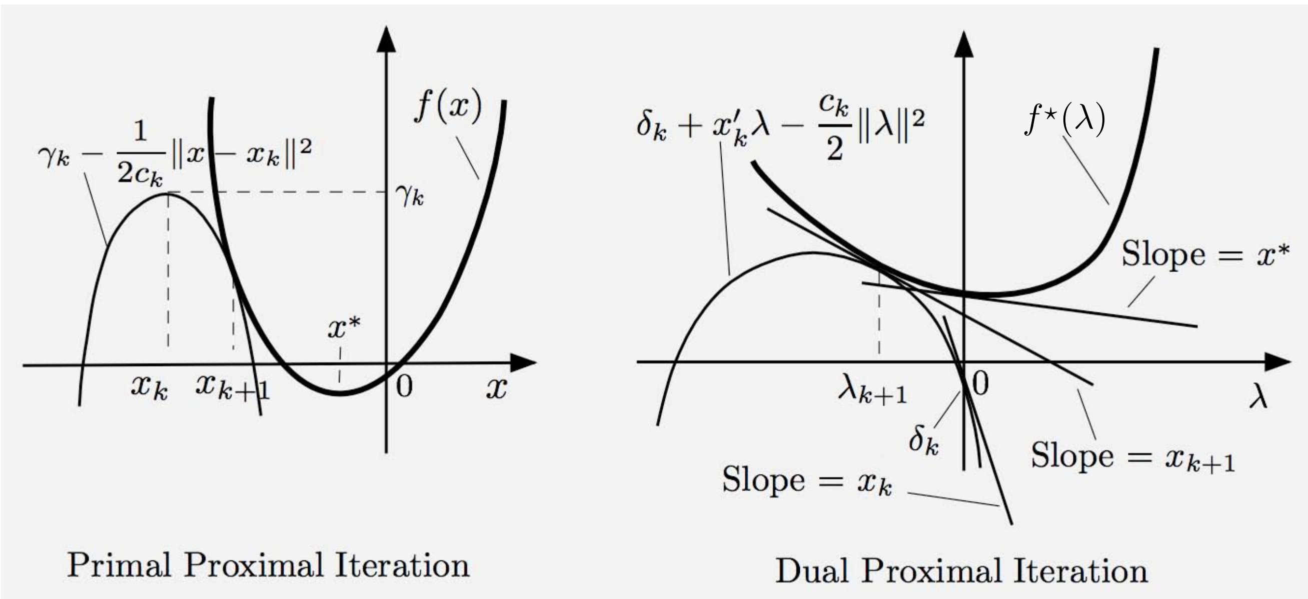
PROXIMAL-POLYHEDRAL METHODS

- **Proximal-cutting plane method**

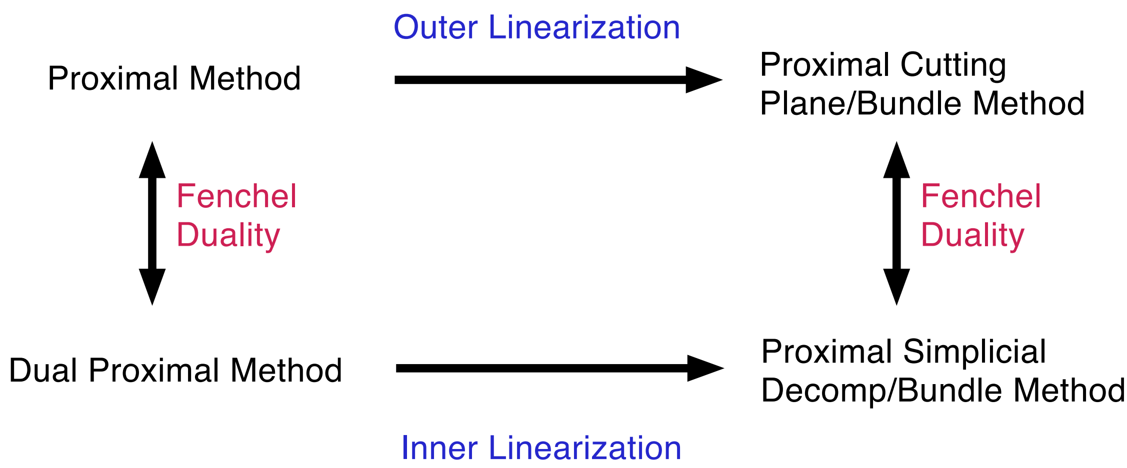


- **Proximal-cutting plane-bundle methods:** Replace f with a cutting plane approx. and/or change quadratic regularization more conservatively.

- **Dual Proximal - Augmented Lagrangian methods:** Proximal method applied to the dual problem of a constrained optimization problem.



DUALITY VIEW OF PROXIMAL METHODS



- Applies also to cost functions that are sums of convex functions

$$f(x) = \sum_{i=1}^{m} f_i(x)$$

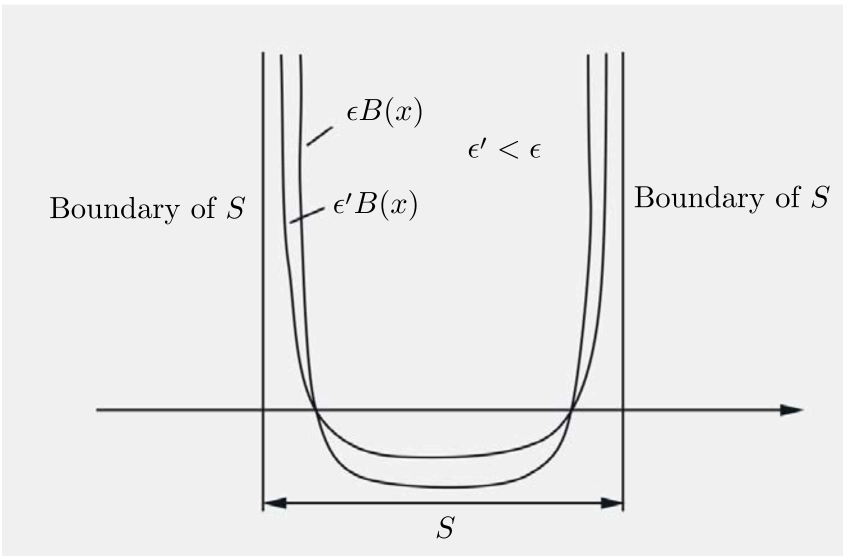
in the context of extended monotropic programming

INTERIOR POINT METHODS

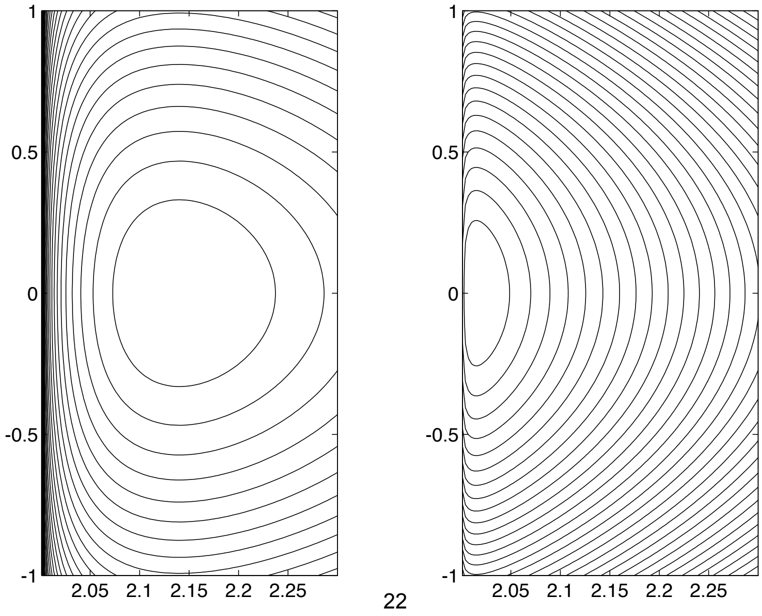
- **Barrier method:** Let

$$x_k = \arg\min_{x \in S} \big\{ f(x) + \epsilon_k B(x) \big\}, \qquad k = 0, 1, \ldots,$$

where $S = \{x \mid g_j(x) < 0,\ j = 1, \ldots, r\}$ and the parameter sequence $\{\epsilon_k\}$ satisfies $0 < \epsilon_{k+1} < \epsilon_k$ for all k and $\epsilon_k \to 0$.



- Ill-conditioning. Need for Newton's method



ADVANCED TOPICS

- Incremental subgradient-proximal methods
- Complexity view of first order algorithms
  - Gradient-projection for differentiable problems
  - Gradient-projection with extrapolation
  - Optimal iteration complexity version (Nesterov)
  - Extension to nondifferentiable problems by smoothing
- Gradient-proximal method
- Useful extension of proximal. General (nonquadratic) regularization - Bregman distance functions
  - Entropy-like regularization
  - Corresponding augmented Lagrangean method (exponential)
  - Corresponding gradient-proximal method
  - Nonlinear gradient/subgradient projection (entropic minimization methods)