LECTURE SLIDES ON

CONVEX ANALYSIS AND OPTIMIZATION

BASED ON 6.253 CLASS LECTURES AT THE

MASS. INSTITUTE OF TECHNOLOGY

CAMBRIDGE, MASS

SPRING 2012

BY DIMITRI P. BERTSEKAS

http://web.mit.edu/dimitrib/www/home.html

Based on the book

"Convex Optimization Theory," Athena Scientific, 2009, including the on-line Chapter 6 and supplementary material at

http://www.athenasc.com/convexduality.html

LECTURE 1
AN INTRODUCTION TO THE COURSE

LECTURE OUTLINE

- The Role of Convexity in Optimization
- Duality Theory
- Algorithms and Duality
- Course Organization

HISTORY AND PREHISTORY

- Prehistory: Early 1900s - 1949.
  - Caratheodory, Minkowski, Steinitz, Farkas.
  - Properties of convex sets and functions.

- Fenchel - Rockafellar era: 1949 - mid 1980s.
  - Duality theory.
  - Minimax/game theory (von Neumann).
  - (Sub)differentiability, optimality conditions, sensitivity.

- Modern era - Paradigm shift: Mid 1980s - present.
  - Nonsmooth analysis (a theoretical/esoteric direction).
  - Algorithms (a practical/high impact direction).
  - A change in the assumptions underlying the field.

OPTIMIZATION PROBLEMS

- Generic form:

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x \in C \end{aligned}$$

Cost function $f : \Re^n \mapsto \Re$, constraint set C, e.g.,

$$\begin{aligned} C = X &\cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \\ &\cap \big\{x \mid g_1(x) \le 0, \ldots, g_r(x) \le 0\big\} \end{aligned}$$

- Continuous vs discrete problem distinction

- Convex programming problems are those for which f and C are convex
  - They are continuous problems
  - They are nice, and have beautiful and intuitive structure

- However, convexity permeates all of optimization, including discrete problems

- Principal vehicle for continuous-discrete connection is duality:
  - The dual problem of a discrete problem is continuous/convex
  - The dual problem provides important information for the solution of the discrete primal (e.g., lower bounds, etc)

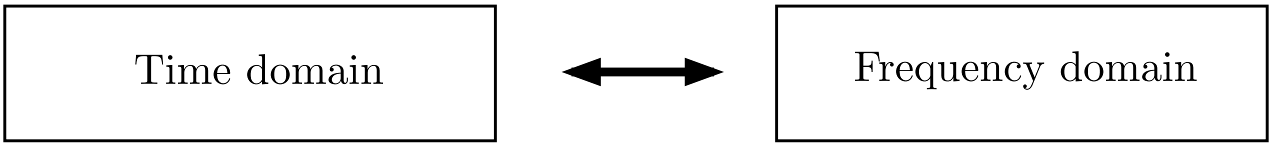
WHY IS CONVEXITY SO SPECIAL?

- A convex function has no local minima that are not global
- A nonconvex function can be "convexified" while maintaining the optimality of its global minima
- A convex set has a nonempty relative interior
- A convex set is connected and has feasible directions at any point
- The existence of a global minimum of a convex function over a convex set is conveniently characterized in terms of directions of recession
- A polyhedral convex set is characterized in terms of a finite set of extreme points and extreme directions
- A real-valued convex function is continuous and has nice differentiability properties
- Closed convex cones are self-dual with respect to polarity
- Convex, lower semicontinuous functions are self-dual with respect to conjugacy

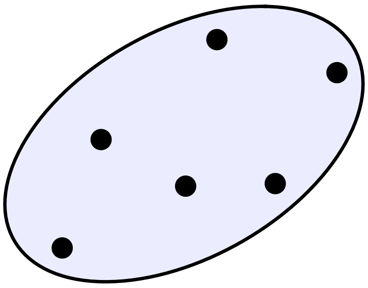
DUALITY

- Two different views of the same object.

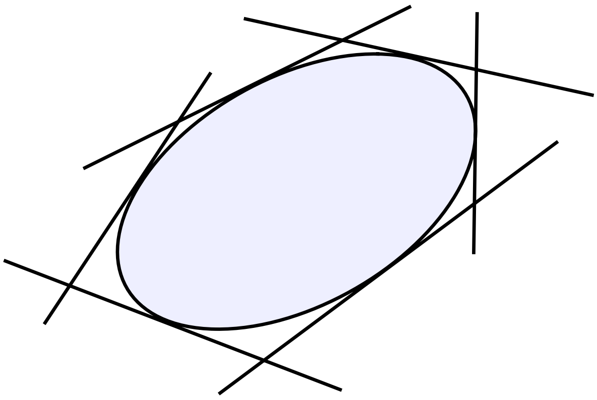
- Example: Dual description of signals.



- Dual description of **closed** convex sets



A union of points



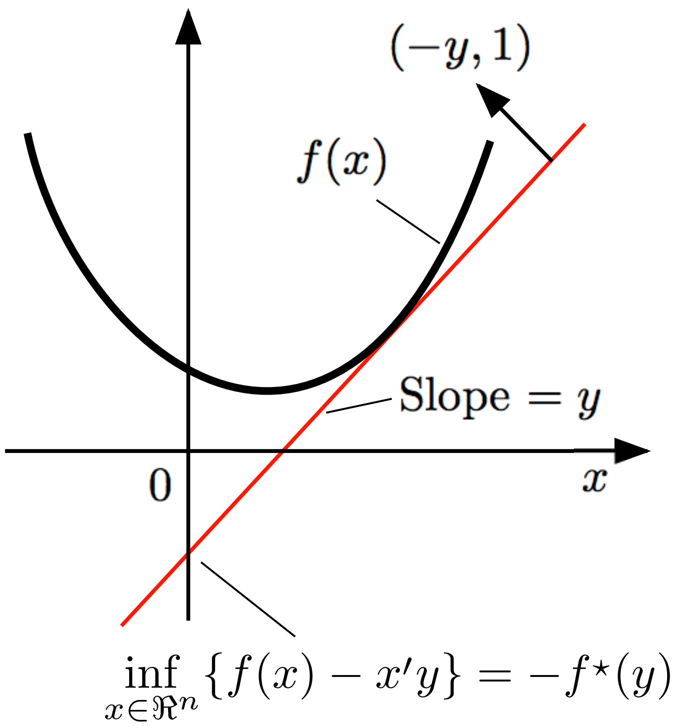
An intersection of halfspaces

DUAL DESCRIPTION OF CONVEX FUNCTIONS

- Define a closed convex function by its epigraph.

- Describe the epigraph by hyperplanes.

- Associate hyperplanes with crossing points (the conjugate function).



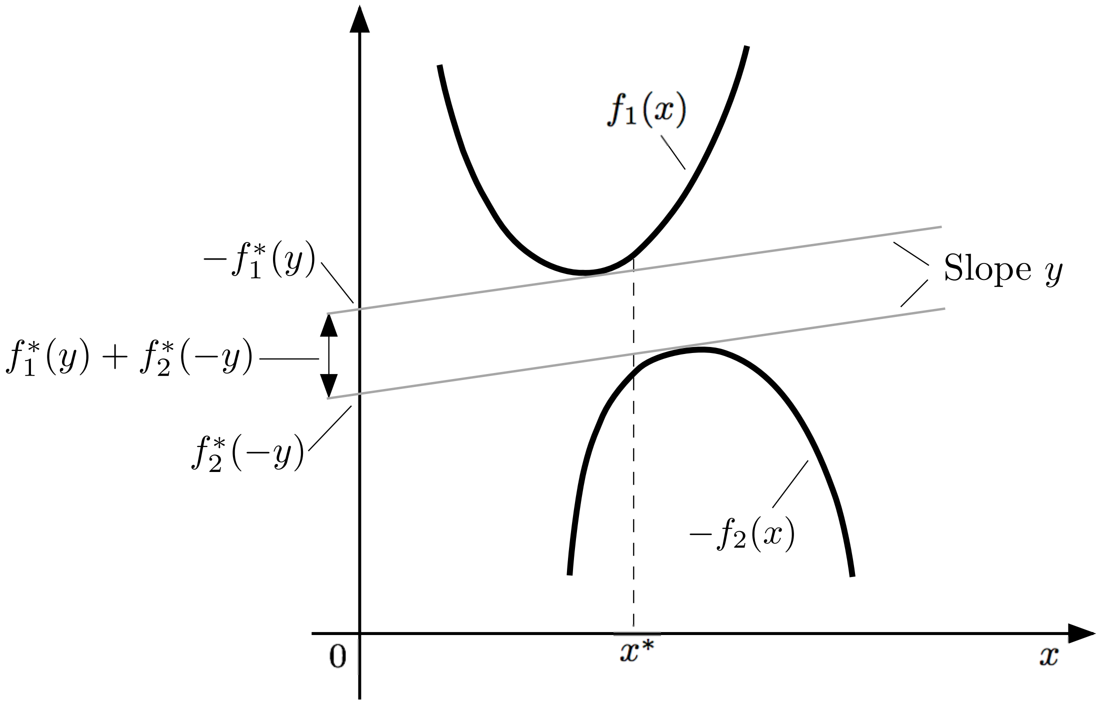
Primal Description

Values $f(x)$

Dual Description

Crossing points $f^*(y)$

FENCHEL PRIMAL AND DUAL PROBLEMS



Primal Problem Description
Vertical Distances

Dual Problem Description
Crossing Point Differentials

- Primal problem:

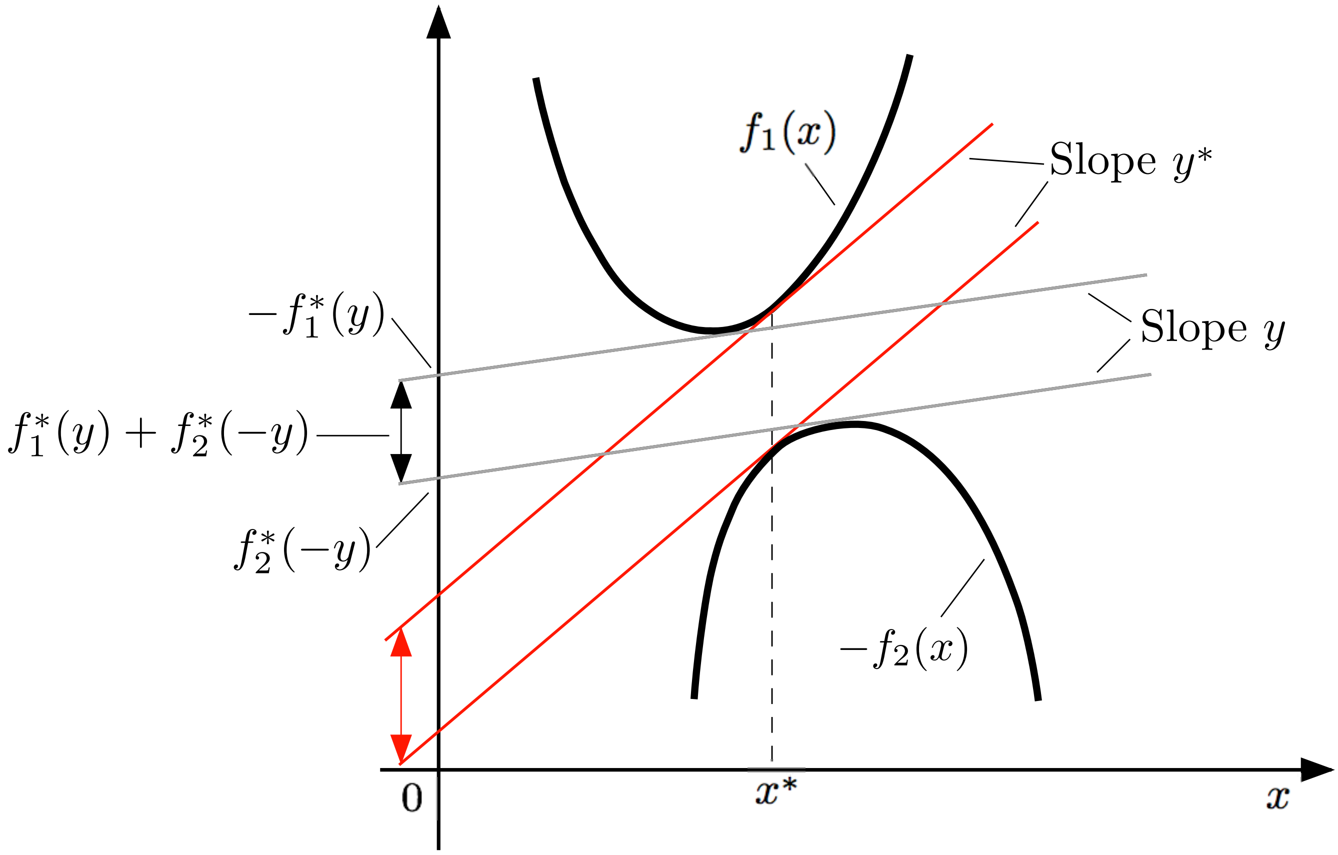
$$\min_x \big\{ f_1(x) + f_2(x) \big\}$$

- Dual problem:

$$\max_y \big\{ - f_1^*(y) - f_2^*(-y) \big\}$$

where f_1^* and f_2^* are the conjugates

FENCHEL DUALITY



$$\min_x \left\{ f_1(x) + f_2(x) \right\} = \max_y \left\{ -f_1^\star(y) - f_2^\star(-y) \right\}$$

- Under favorable conditions (convexity):
  - The optimal primal and dual values are equal
  - The optimal primal and dual solutions are related

A MORE ABSTRACT VIEW OF DUALITY

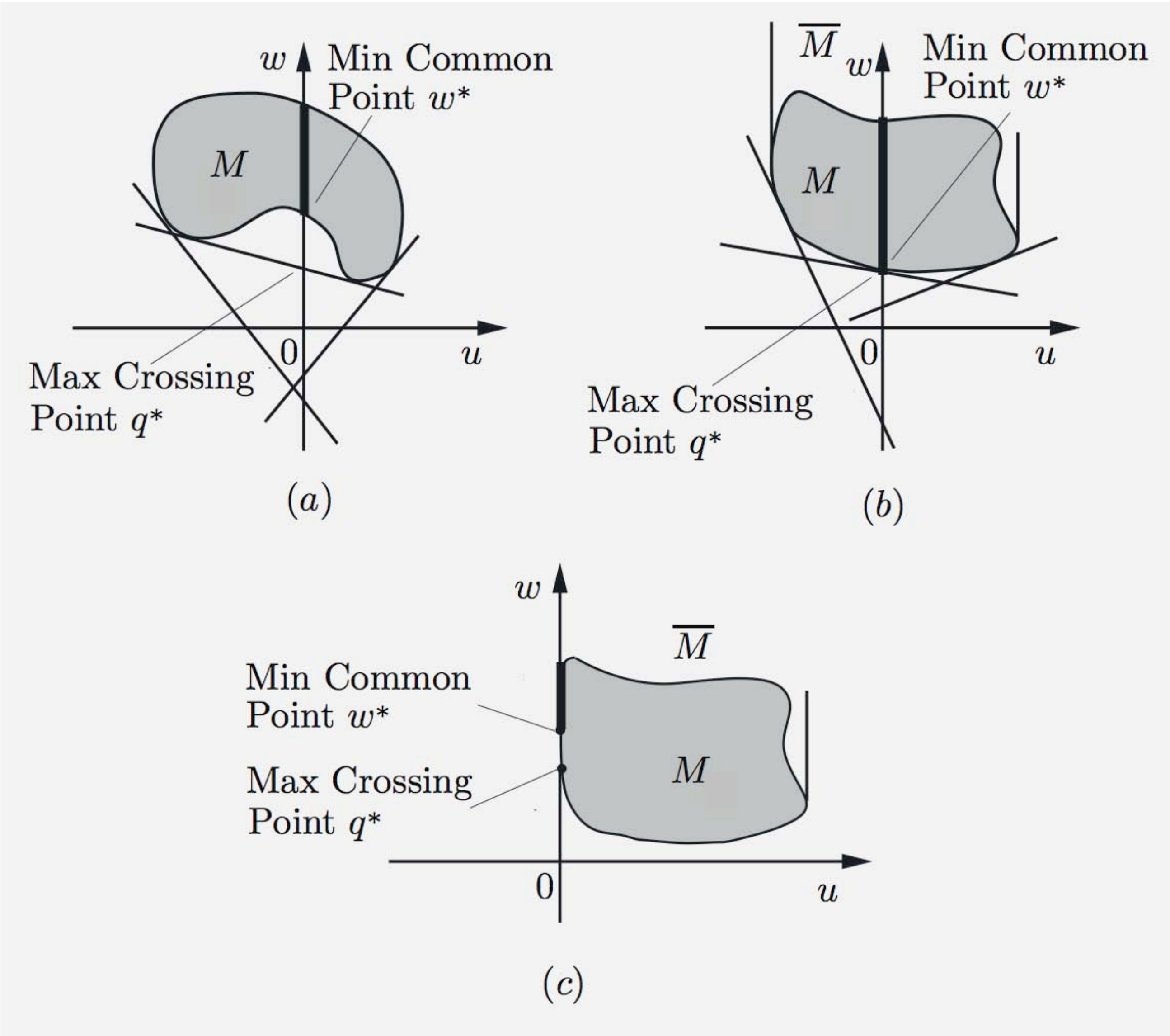
- Despite its elegance, the Fenchel framework is somewhat indirect.

- From duality of set descriptions, to
  - duality of functional descriptions, to
  - duality of problem descriptions.

- A more direct approach:
  - Start with a set, then
  - Define two simple prototype problems dual to each other.

- Avoid functional descriptions (a simpler, less constrained framework).

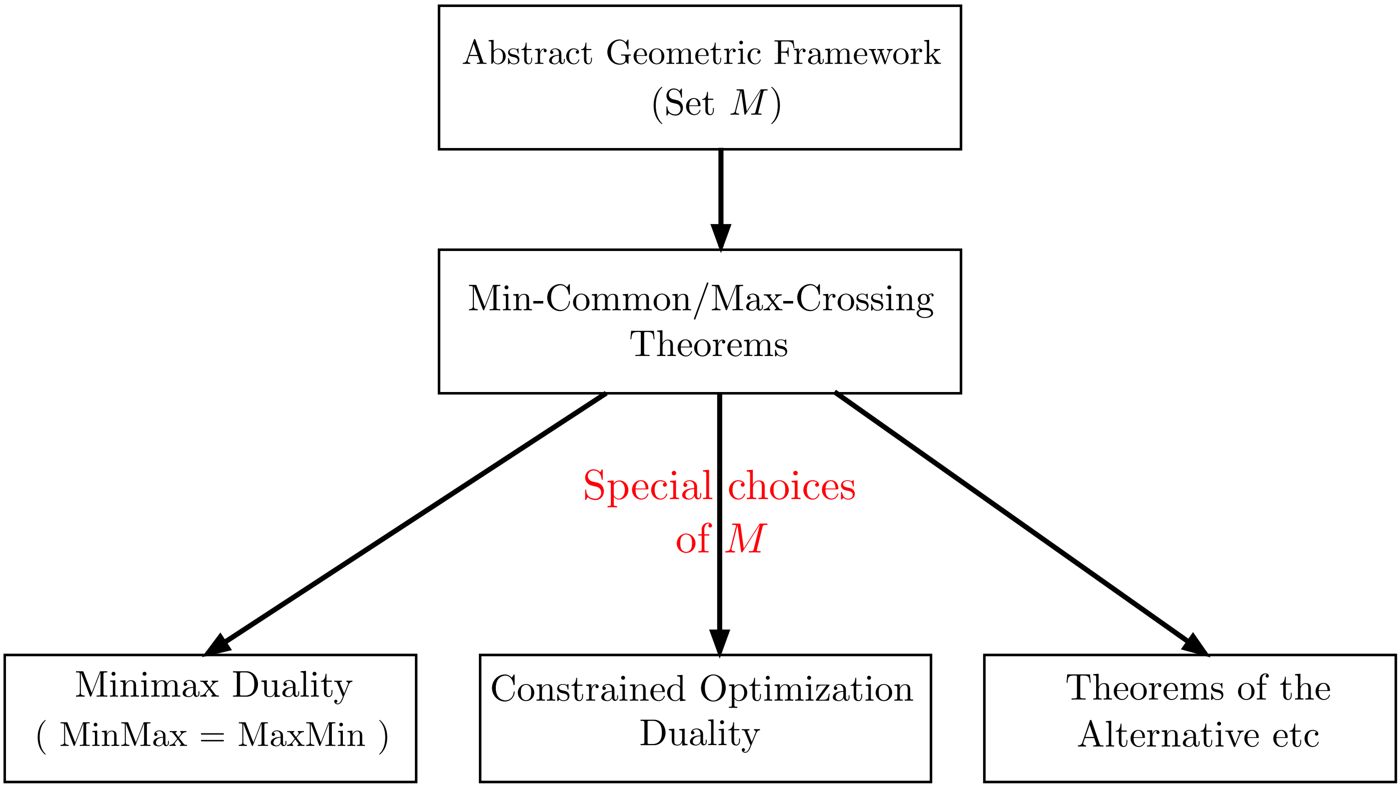
MIN COMMON/MAX CROSSING DUALITY



- All of duality theory and all of (convex/concave) minimax theory can be developed/explained in terms of this one figure.

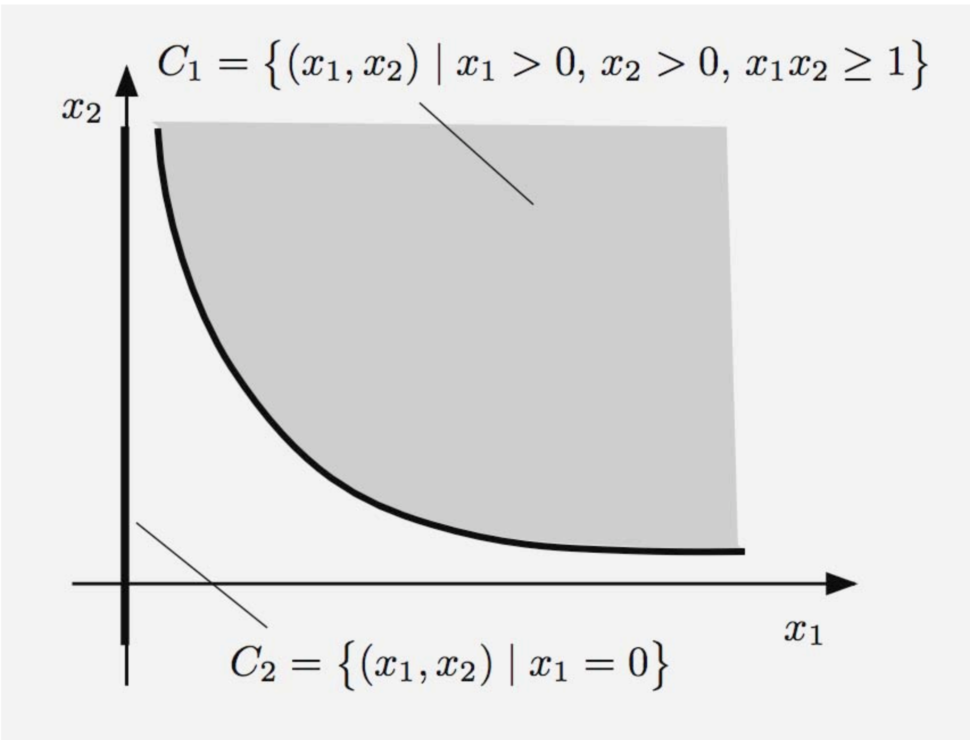
- The machinery of convex analysis is needed to flesh out this figure, and to rule out the exceptional/pathological behavior shown in (c).

ABSTRACT/GENERAL DUALITY ANALYSIS



EXCEPTIONAL BEHAVIOR

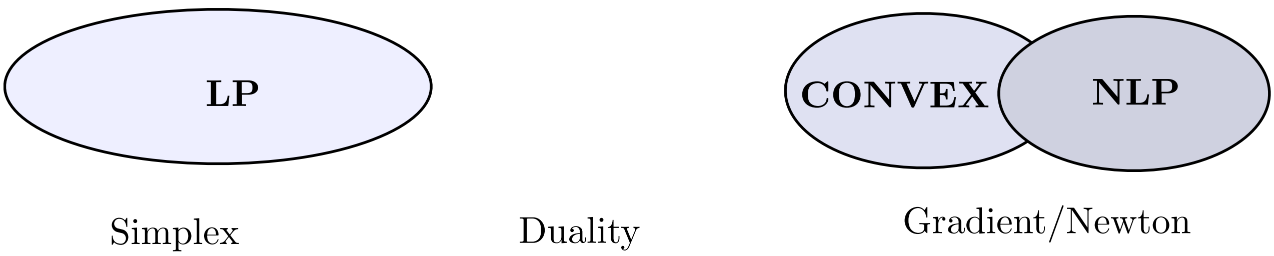
- If convex structure is so favorable, what is the source of exceptional/pathological behavior?
- **Answer:** Some common operations on convex sets do not preserve some basic properties.
- **Example:** A linearly transformed closed convex set need not be closed (contrary to compact and polyhedral sets).
  - Also the vector sum of two closed convex sets need not be closed.



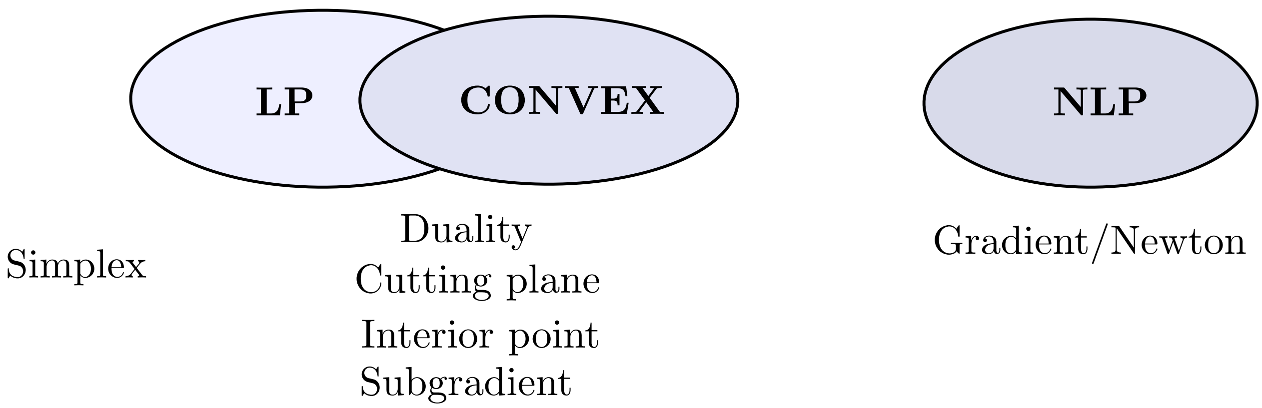
- This is a major reason for the analytical difficulties in convex analysis and pathological behavior in convex optimization (and the favorable character of polyhedral sets).

MODERN VIEW OF CONVEX OPTIMIZATION

- Traditional view: Pre 1990s
  - LPs are solved by simplex method
  - NLPs are solved by gradient/Newton methods
  - Convex programs are special cases of NLPs



- Modern view: Post 1990s
  - LPs are often solved by nonsimplex/convex methods
  - Convex problems are often solved by the same methods as LPs
  - "Key distinction is not Linear-Nonlinear but Convex-Nonconvex" (Rockafellar)



THE RISE OF THE ALGORITHMIC ERA

- Convex programs and LPs connect around
  - Duality
  - Large-scale piecewise linear problems
- Synergy of:
  - Duality
  - Algorithms
  - Applications
- New problem paradigms with rich applications
- Duality-based decomposition
  - Large-scale resource allocation
  - Lagrangian relaxation, discrete optimization
  - Stochastic programming
- Conic programming
  - Robust optimization
  - Semidefinite programming
- Machine learning
  - Support vector machines
  - l_1 regularization/Robust regression/Compressed sensing

METHODOLOGICAL TRENDS

- New methods, renewed interest in old methods.
  - Interior point methods
  - Subgradient/incremental methods
  - Polyhedral approximation/cutting plane methods
  - Regularization/proximal methods
  - Incremental methods

- Renewed emphasis on complexity analysis
  - Nesterov, Nemirovski, and others ...
  - "Optimal algorithms" (e.g., extrapolated gradient methods)

- Emphasis on interesting (often duality-related) large-scale special structures

COURSE OUTLINE

- We will follow closely the textbook
  - Bertsekas, "Convex Optimization Theory," Athena Scientific, 2009, including the on-line Chapter 6 and supplementary material at http://www.athenasc.com/convexduality.html
- Additional book references:
  - Rockafellar, "Convex Analysis," 1970.
  - Boyd and Vanderbergue, "Convex Optimization," Cambridge U. Press, 2004. (On-line at http://www.stanford.edu/~boyd/cvxbook/)
  - Bertsekas, Nedic, and Ozdaglar, "Convex Analysis and Optimization," Ath. Scientific, 2003.
- Topics (the text's design is modular, and the following sequence involves no loss of continuity):
  - **Basic Convexity Concepts:** Sect. 1.1-1.4.
  - **Convexity and Optimization:** Ch. 3.
  - **Hyperplanes & Conjugacy:** Sect. 1.5, 1.6.
  - **Polyhedral Convexity:** Ch. 2.
  - **Geometric Duality Framework:** Ch. 4.
  - **Duality Theory:** Sect. 5.1-5.3.
  - **Subgradients:** Sect. 5.4.
  - **Algorithms:** Ch. 6.

WHAT TO EXPECT FROM THIS COURSE

- Requirements: Homework (25%), midterm (25%), and a term paper (50%)
- We aim:
  - To develop insight and deep understanding of a fundamental optimization topic
  - To treat with mathematical rigor an important branch of methodological research, and to provide an account of the state of the art in the field
  - To get an understanding of the merits, limitations, and characteristics of the rich set of available algorithms
- Mathematical level:
  - Prerequisites are linear algebra (preferably abstract) and real analysis (a course in each)
  - Proofs will matter ... but the rich geometry of the subject helps guide the mathematics
- Applications:
  - They are many and pervasive ... but don't expect much in this course. The book by Boyd and Vandenberghe describes a lot of practical convex optimization models
  - You can do your term paper on an application area

A NOTE ON THESE SLIDES

- These slides are a teaching aid, not a text
- Don't expect a rigorous mathematical development
- The statements of theorems are fairly precise, but the proofs are not
- Many proofs have been omitted or greatly abbreviated
- Figures are meant to convey and enhance understanding of ideas, not to express them precisely
- The omitted proofs and a fuller discussion can be found in the "Convex Optimization Theory" textbook and its supplementary material

LECTURE 2

LECTURE OUTLINE

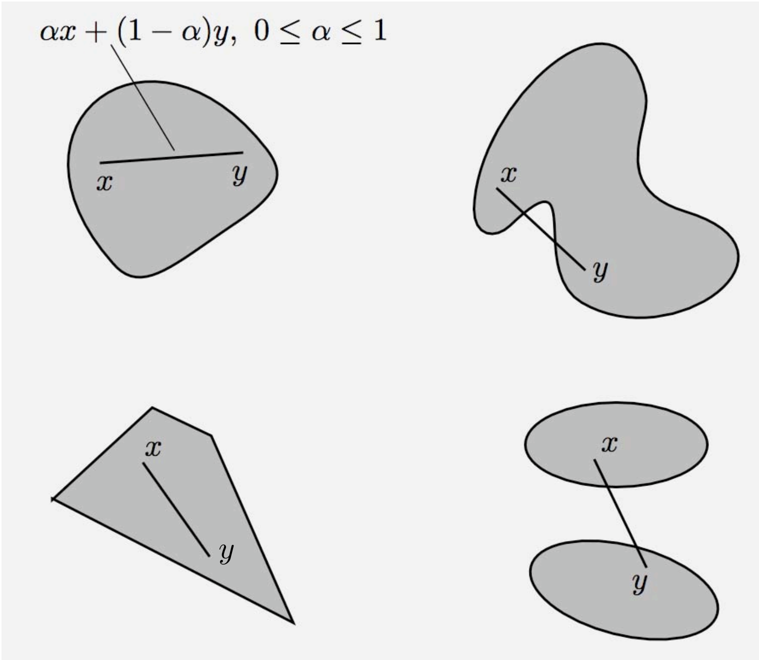
- Convex sets and functions
- Epigraphs
- Closed convex functions
- Recognizing convex functions

Reading: Section 1.1

SOME MATH CONVENTIONS

- All of our work is done in $\Re^n$: space of n-tuples $x = (x_1, \ldots, x_n)$
- All vectors are assumed column vectors
- "$'$" denotes transpose, so we use x' to denote a row vector
- $x'y$ is the inner product $\sum_{i=1}^n x_i y_i$ of vectors x and y
- $\|x\| = \sqrt{x'x}$ is the (Euclidean) norm of x. We use this norm almost exclusively
- See the textbook for an overview of the linear algebra and real analysis background that we will use. Particularly the following:
  - Definition of sup and inf of a set of real numbers
  - Convergence of sequences (definitions of lim inf, lim sup of a sequence of real numbers, and definition of lim of a sequence of vectors)
  - Open, closed, and compact sets and their properties
  - Definition and properties of differentiation

CONVEX SETS



- A subset C of $\Re^n$ is called *convex* if

$$\alpha x + (1-\alpha)y \in C, \qquad \forall\, x, y \in C,\ \forall\, \alpha \in [0,1]$$

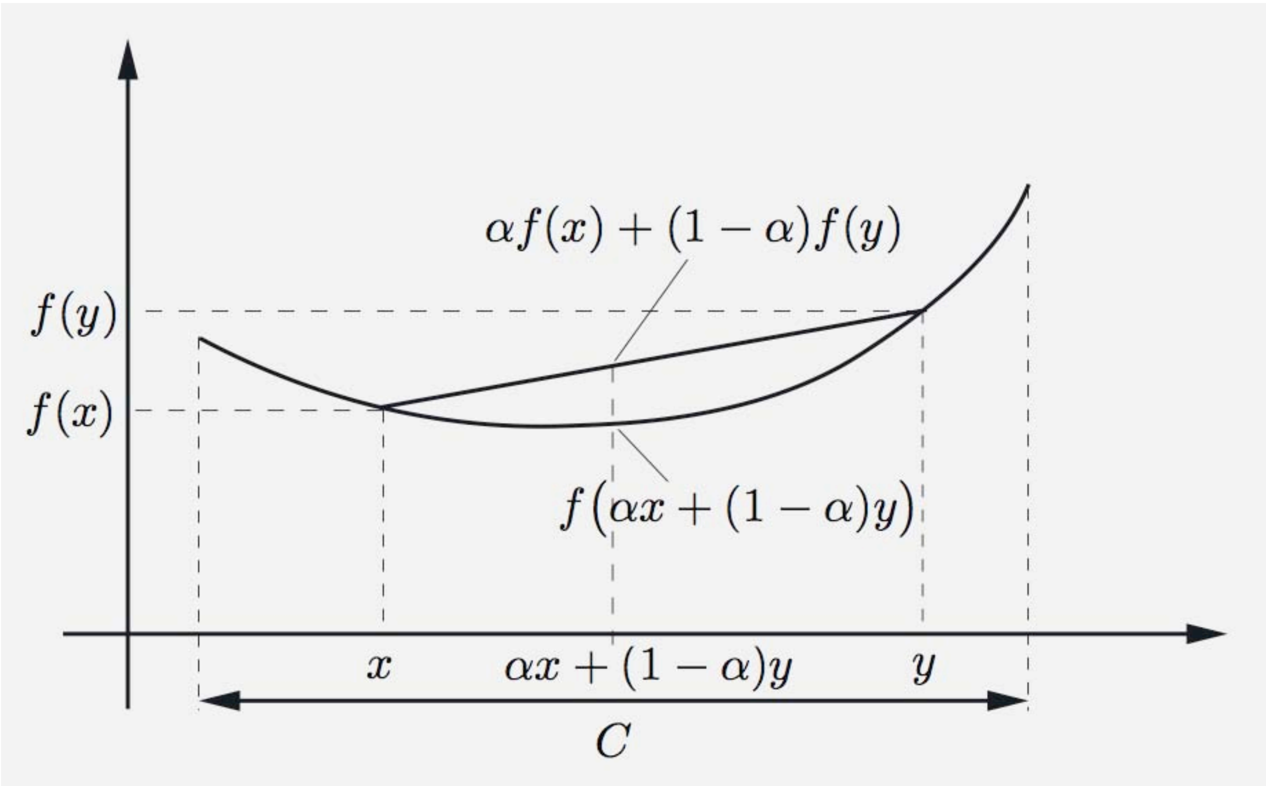
- Operations that preserve convexity
  - Intersection, scalar multiplication, vector sum, closure, interior, linear transformations

- Special convex sets:
  - **Polyhedral sets:** Nonempty sets of the form

$$\{x \mid a_j'x \le b_j,\ j = 1, \ldots, r\}$$

    (always convex, closed, not always bounded)
  - **Cones:** Sets C such that $\lambda x \in C$ for all $\lambda > 0$ and $x \in C$ (not always convex or closed)

CONVEX FUNCTIONS



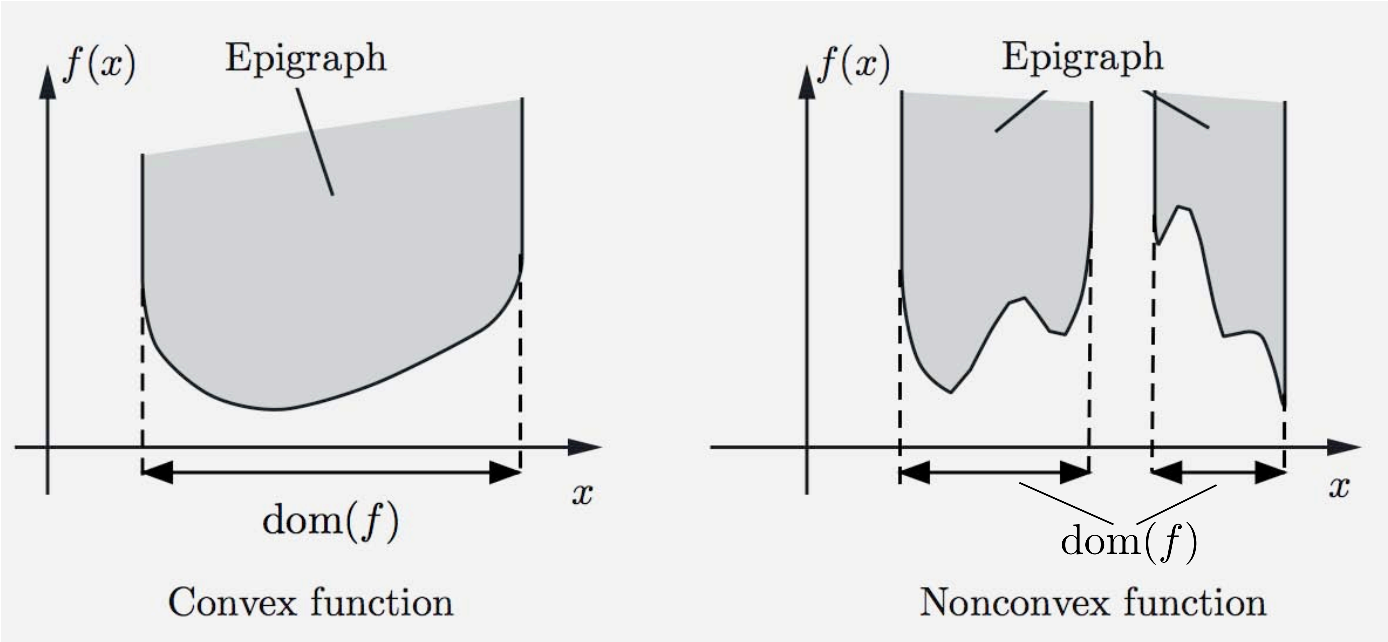
- Let C be a convex subset of $\Re^n$. A function $f : C \mapsto \Re$ is called *convex* if for all $\alpha \in [0,1]$

$$f\big(\alpha x+(1-\alpha)y\big) \leq \alpha f(x)+(1-\alpha)f(y), \quad \forall\, x,y \in C$$

If the inequality is strict whenever $a \in (0,1)$ and $x \neq y$, then f is called *strictly convex* over C.

- If f is a convex function, then all its level sets $\{x \in C \mid f(x) \leq \gamma\}$ and $\{x \in C \mid f(x) < \gamma\}$, where γ is a scalar, are convex.

EXTENDED REAL-VALUED FUNCTIONS



• The *epigraph* of a function $f : X \mapsto [-\infty, \infty]$ is the subset of $\Re^{n+1}$ given by

$$\text{epi}(f) = \big\{(x, w) \mid x \in X,\ w \in \Re,\ f(x) \le w\big\}$$

• The *effective domain* of f is the set

$$\text{dom}(f) = \big\{x \in X \mid f(x) < \infty\big\}$$

• We say that f is *convex* if epi(f) is a convex set. If $f(x) \in \Re$ for all $x \in X$ and X is convex, the definition "coincides" with the earlier one.

• We say that f is *closed* if epi(f) is a closed set.

• We say that f is *lower semicontinuous* at a vector $x \in X$ if $f(x) \le \liminf_{k\to\infty} f(x_k)$ for every sequence $\{x_k\} \subset X$ with $x_k \to x$.

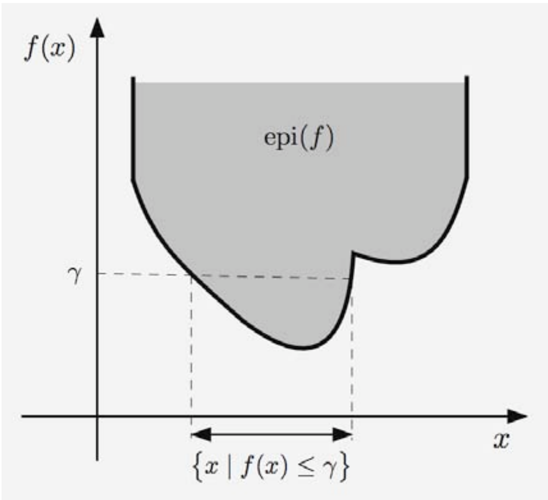
CLOSEDNESS AND SEMICONTINUITY I

- *Proposition*: For a function $f : \Re^n \mapsto [-\infty, \infty]$, the following are equivalent:

  (i) $V_\gamma = \{x \mid f(x) \leq \gamma\}$ is closed for all $\gamma \in \Re$.

  (ii) f is lower semicontinuous at all $x \in \Re^n$.

  (iii) f is closed.



- (ii) $\Rightarrow$ (iii): Let $\{(x_k, w_k)\} \subset \text{epi}(f)$ with $(x_k, w_k) \to (\overline{x}, \overline{w})$. Then $f(x_k) \leq w_k$, and

$$f(\overline{x}) \leq \liminf_{k \to \infty} f(x_k) \leq \overline{w} \quad \text{so } (\overline{x}, \overline{w}) \in \text{epi}(f)$$

- (iii) $\Rightarrow$ (i): Let $\{x_k\} \subset V_\gamma$ and $x_k \to \overline{x}$. Then $(x_k, \gamma) \in \text{epi}(f)$ and $(x_k, \gamma) \to (\overline{x}, \gamma)$, so $(\overline{x}, \gamma) \in \text{epi}(f)$, and $\overline{x} \in V_\gamma$.

- (i) $\Rightarrow$ (ii): If $x_k \to \overline{x}$ and $f(\overline{x}) > \gamma > \liminf_{k \to \infty} f(x_k)$ consider subsequence $\{x_k\}_\mathcal{K} \to \overline{x}$ with $f(x_k) \leq \gamma$ - contradicts closedness of V_γ.

CLOSEDNESS AND SEMICONTINUITY II

- Lower semicontinuity of a function is a "domain-specific" property, but closeness is not:
  - If we change the domain of the function without changing its epigraph, its lower semicontinuity properties may be affected.
  - **Example:** Define $f : (0,1) \to [-\infty,\infty]$ and $\hat{f} : [0,1] \to [-\infty,\infty]$ by

$$f(x) = 0, \qquad \forall\, x \in (0,1),$$

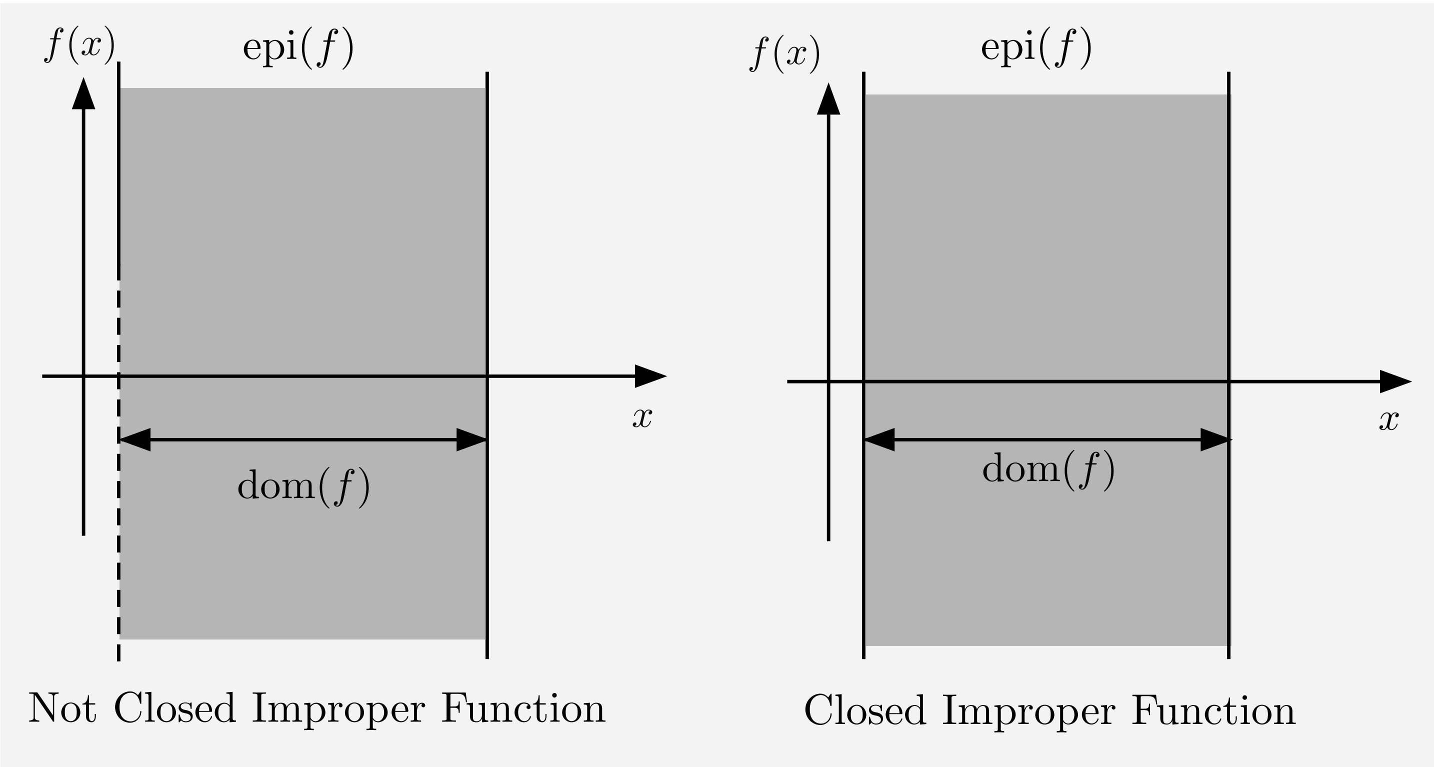
$$\hat{f}(x) = \begin{cases} 0 & \text{if } x \in (0,1), \\ \infty & \text{if } x = 0 \text{ or } x = 1. \end{cases}$$

    Then f and $\hat{f}$ have the same epigraph, and both are not closed. But f is lower-semicontinuous while $\hat{f}$ is not.

- Note that:
  - If f is lower semicontinuous at all $x \in \text{dom}(f)$, it is not necessarily closed
  - If f is closed, $\text{dom}(f)$ is not necessarily closed

- *Proposition*: Let $f : X \mapsto [-\infty,\infty]$ be a function. If $\text{dom}(f)$ is closed and f is lower semicontinuous at all $x \in \text{dom}(f)$, then f is closed.

PROPER AND IMPROPER CONVEX FUNCTION



- We say that f is *proper* if $f(x) < \infty$ for at least one $x \in X$ and $f(x) > -\infty$ for all $x \in X$, and we will call f *improper* if it is not proper.

- Note that f is proper if and only if its epigraph is nonempty and does not contain a "vertical line."

- An improper *closed* convex function is very peculiar: it takes an infinite value (∞ or $-\infty$) at every point.

RECOGNIZING CONVEX FUNCTIONS

- Some important classes of elementary convex functions: Affine functions, positive semidefinite quadratic functions, norm functions, etc.

- *Proposition*: (a) The function $g : \Re^n \mapsto (-\infty, \infty]$ given by

$$g(x) = \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x), \qquad \lambda_i > 0$$

is convex (or closed) if $f_1, \ldots, f_m$ are convex (respectively, closed).

(b) The function $g : \Re^n \mapsto (-\infty, \infty]$ given by

$$g(x) = f(Ax)$$

where A is an $m \times n$ matrix is convex (or closed) if f is convex (respectively, closed).

(c) Consider $f_i : \Re^n \mapsto (-\infty, \infty]$, $i \in I$, where I is any index set. The function $g : \Re^n \mapsto (-\infty, \infty]$ given by

$$g(x) = \sup_{i \in I} f_i(x)$$

is convex (or closed) if the f_i are convex (respectively, closed).

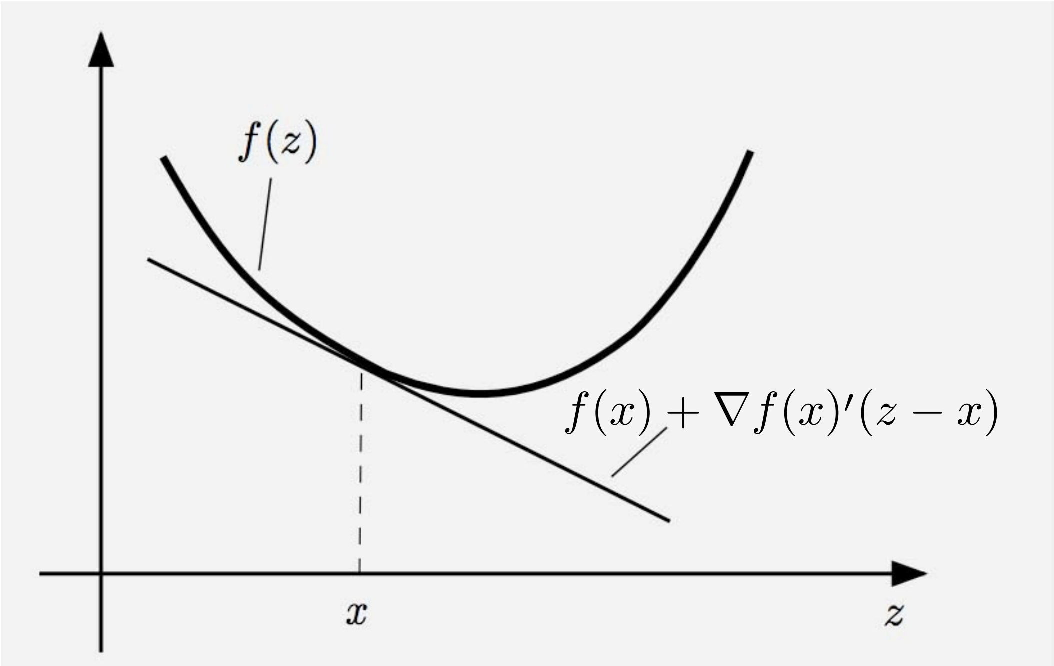
LECTURE 3

LECTURE OUTLINE

- Differentiable Convex Functions
- Convex and Affine Hulls
- Caratheodory's Theorem

Reading: Sections 1.1, 1.2

DIFFERENTIABLE CONVEX FUNCTIONS



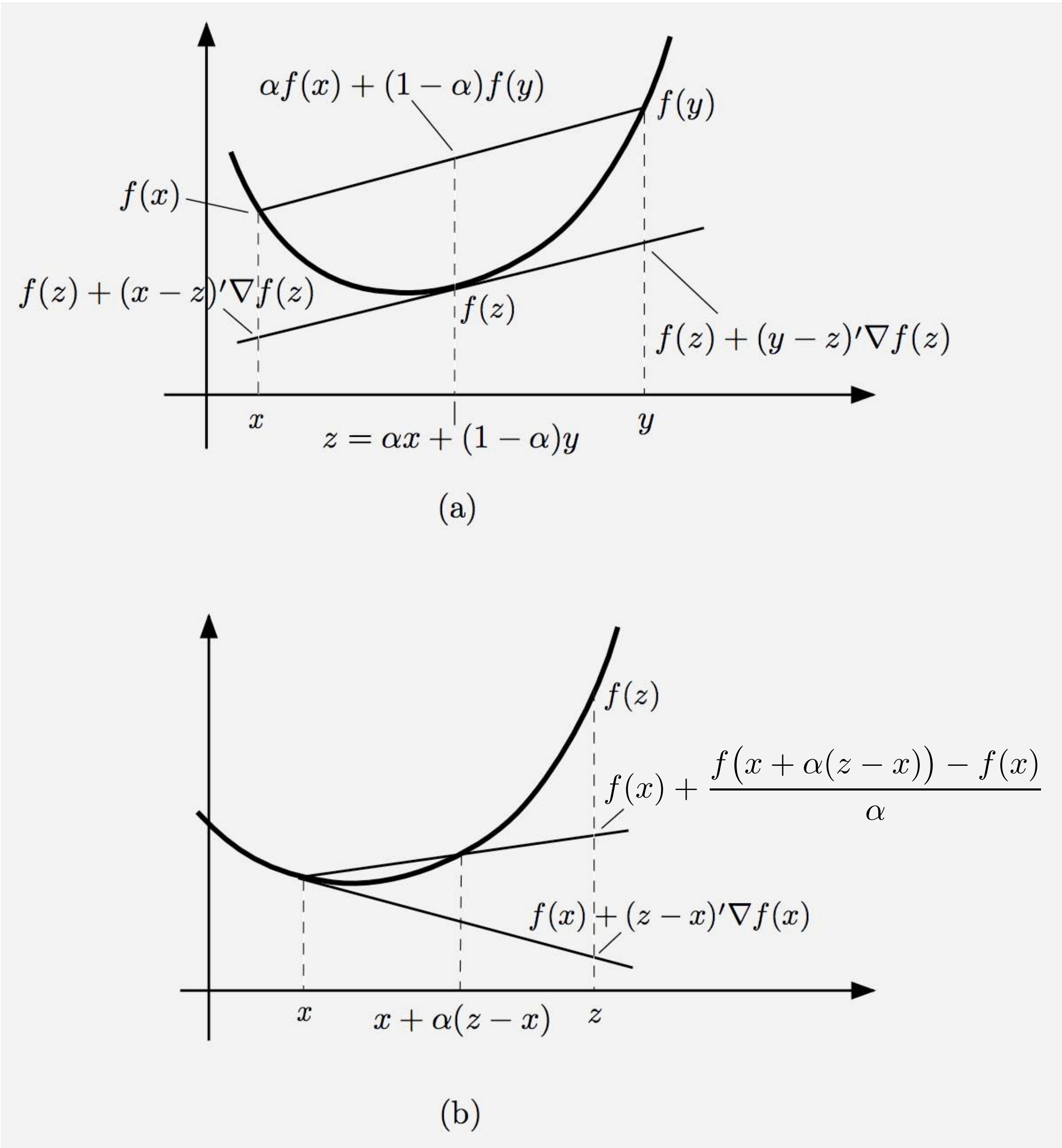
- Let $C \subset \Re^n$ be a convex set and let $f : \Re^n \mapsto \Re$ be differentiable over $\Re^n$.

  (a) The function f is convex over C iff

  $$f(z) \geq f(x) + (z - x)'\nabla f(x), \qquad \forall\, x, z \in C$$

  (b) If the inequality is strict whenever $x \neq z$, then f is strictly convex over C.

PROOF IDEAS


$\alpha f(x) + (1-\alpha)f(y)$
$f(y)$
$f(x)$
$f(z) + (x-z)'\nabla f(z)$
$f(z)$
$f(z) + (y-z)'\nabla f(z)$
x
y
$z = \alpha x + (1-\alpha)y$
(a)
$f(z)$
$f(x) + \frac{f\big(x+\alpha(z-x)\big) - f(x)}{\alpha}$
$f(x) + (z-x)'\nabla f(x)$
x
$x + \alpha(z-x)$
z
(b)

OPTIMALITY CONDITION

• Let C be a nonempty convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be convex and differentiable over an open set that contains C. Then a vector $x^* \in C$ minimizes f over C if and only if

$$\nabla f(x^*)'(x - x^*) \geq 0, \qquad \forall\, x \in C.$$

Proof: If the condition holds, then

$$f(x) \geq f(x^*) + (x - x^*)'\nabla f(x^*) \geq f(x^*), \quad \forall\, x \in C,$$

so x^* minimizes f over C.

Converse: Assume the contrary, i.e., x^* minimizes f over C and $\nabla f(x^*)'(x - x^*) < 0$ for some $x \in C$. By differentiation, we have

$$\lim_{\alpha \downarrow 0} \frac{f\big(x^* + \alpha(x - x^*)\big) - f(x^*)}{\alpha} = \nabla f(x^*)'(x - x^*) < 0$$

so $f\big(x^* + \alpha(x - x^*)\big)$ decreases strictly for sufficiently small $\alpha > 0$, contradicting the optimality of x^*. **Q.E.D.**

PROJECTION THEOREM

- Let C be a nonempty closed convex set in $\Re^n$.

  (a) For every $z \in \Re^n$, there exists a unique minimum of
  $$f(x) = \|z - x\|^2$$
  over all $x \in C$ (called the *projection of z on C*).

  (b) x^* is the projection of z if and only if
  $$(x - x^*)'(z - x^*) \leq 0, \qquad \forall\, x \in C$$

Proof: (a) f is strictly convex and has compact level sets.

(b) This is just the necessary and sufficient optimality condition

$$\nabla f(x^*)'(x - x^*) \geq 0, \qquad \forall\, x \in C.$$

TWICE DIFFERENTIABLE CONVEX FNS

- Let C be a convex subset of $\Re^n$ and let $f : \Re^n \mapsto \Re$ be twice continuously differentiable over $\Re^n$.

  (a) If $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$, then f is convex over C.

  (b) If $\nabla^2 f(x)$ is positive definite for all $x \in C$, then f is strictly convex over C.

  (c) If C is open and f is convex over C, then $\nabla^2 f(x)$ is positive semidefinite for all $x \in C$.

Proof: (a) By mean value theorem, for $x, y \in C$

$$f(y) = f(x) + (y-x)'\nabla f(x) + \tfrac{1}{2}(y-x)'\nabla^2 f\big(x+\alpha(y-x)\big)(y-x)$$

for some $\alpha \in [0,1]$. Using the positive semidefiniteness of $\nabla^2 f$, we obtain

$$f(y) \geq f(x) + (y-x)'\nabla f(x), \qquad \forall\, x, y \in C$$

From the preceding result, f is convex.

(b) Similar to (a), we have $f(y) > f(x) + (y - x)'\nabla f(x)$ for all $x, y \in C$ with $x \neq y$, and we use the preceding result.

(c) By contradiction ... similar.

CONVEX AND AFFINE HULLS

- Given a set $X \subset \Re^n$:

- A *convex combination* of elements of X is a vector of the form $\sum_{i=1}^m \alpha_i x_i$, where $x_i \in X$, $\alpha_i \geq 0$, and $\sum_{i=1}^m \alpha_i = 1$.

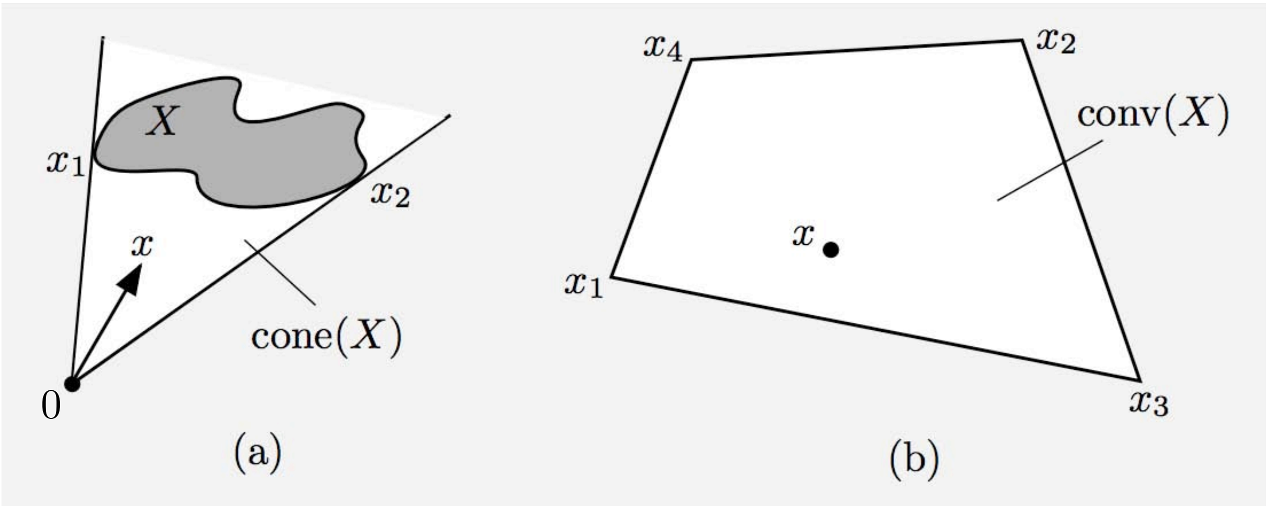
- The *convex hull* of X, denoted $\text{conv}(X)$, is the intersection of all convex sets containing X. (Can be shown to be equal to the set of all convex combinations from X).

- The *affine hull* of X, denoted $\text{aff}(X)$, is the intersection of all affine sets containing X (an affine set is a set of the form $\overline{x} + S$, where S is a subspace).

- A *nonnegative combination* of elements of X is a vector of the form $\sum_{i=1}^m \alpha_i x_i$, where $x_i \in X$ and $\alpha_i \geq 0$ for all i.

- The *cone generated by* X, denoted $\text{cone}(X)$, is the set of all nonnegative combinations from X:
  - It is a convex cone containing the origin.
  - It need not be closed!
  - If X is a finite set, $\text{cone}(X)$ is closed (nontrivial to show!)

CARATHEODORY'S THEOREM



- Let X be a nonempty subset of $\Re^n$.
  (a) Every $x \neq 0$ in cone(X) can be represented as a positive combination of vectors $x_1, \ldots, x_m$ from X that are linearly independent (so $m \leq n$).
  (b) Every $x \notin X$ that belongs to conv(X) can be represented as a convex combination of vectors $x_1, \ldots, x_m$ from X with $m \leq n + 1$.

PROOF OF CARATHEODORY'S THEOREM

(a) Let x be a nonzero vector in cone(X), and let m be the smallest integer such that x has the form $\sum_{i=1}^{m} \alpha_i x_i$, where $\alpha_i > 0$ and $x_i \in X$ for all $i = 1, \ldots, m$. If the vectors x_i were linearly dependent, there would exist $\lambda_1, \ldots, \lambda_m$, with

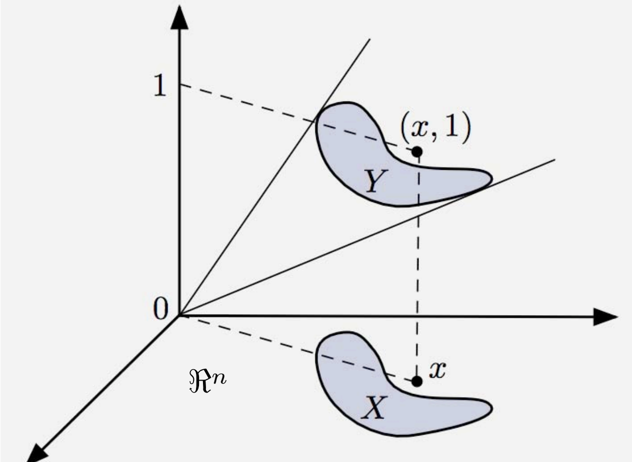
$$\sum_{i=1}^{m} \lambda_i x_i = 0$$

and at least one of the λ_i is positive. Consider

$$\sum_{i=1}^{m} (\alpha_i - \overline{\gamma}\lambda_i) x_i,$$

where $\overline{\gamma}$ is the largest γ such that $\alpha_i - \gamma\lambda_i \geq 0$ for all i. This combination provides a representation of x as a positive combination of fewer than m vectors of X – a contradiction. Therefore, $x_1, \ldots, x_m$, are linearly independent.

(b) Use "lifting" argument: apply part (a) to $Y = \{(x, 1) \mid x \in X\}$.



AN APPLICATION OF CARATHEODORY

- The convex hull of a compact set is compact.

Proof: Let X be compact. We take a sequence in $\text{conv}(X)$ and show that it has a convergent subsequence whose limit is in $\text{conv}(X)$.

By Caratheodory, a sequence in $\text{conv}(X)$ can be expressed as $\left\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\right\}$, where for all k and i, $\alpha_i^k \geq 0$, $x_i^k \in X$, and $\sum_{i=1}^{n+1} \alpha_i^k = 1$. Since the sequence

$$\left\{(\alpha_1^k, \ldots, \alpha_{n+1}^k, x_1^k, \ldots, x_{n+1}^k)\right\}$$

is bounded, it has a limit point

$$\left\{(\alpha_1, \ldots, \alpha_{n+1}, x_1, \ldots, x_{n+1})\right\},$$

which must satisfy $\sum_{i=1}^{n+1} \alpha_i = 1$, and $\alpha_i \geq 0$, $x_i \in X$ for all i.

The vector $\sum_{i=1}^{n+1} \alpha_i x_i$ belongs to $\text{conv}(X)$ and is a limit point of $\left\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\right\}$, showing that $\text{conv}(X)$ is compact. **Q.E.D.**

- Note that the convex hull of a closed set need not be closed!

LECTURE 4

LECTURE OUTLINE

- Relative interior and closure
- Algebra of relative interiors and closures
- Continuity of convex functions
- Closures of functions

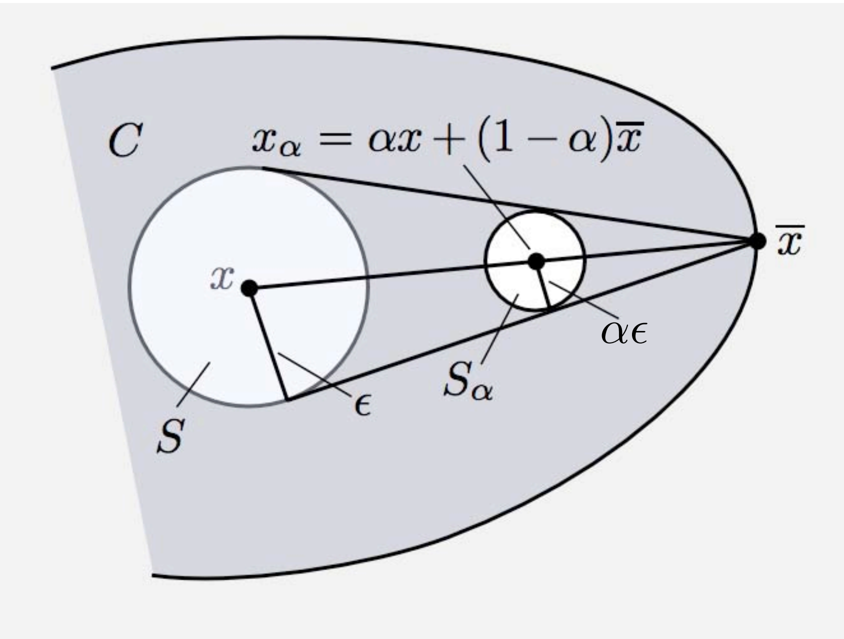
Reading: Section 1.3

RELATIVE INTERIOR

- x is a *relative interior point* of C, if x is an interior point of C relative to aff(C).

- ri(C) denotes the *relative interior of* C, i.e., the set of all relative interior points of C.

- **Line Segment Principle**: If C is a convex set, $x \in \text{ri}(C)$ and $\overline{x} \in \text{cl}(C)$, then all points on the line segment connecting x and $\overline{x}$, except possibly $\overline{x}$, belong to ri(C).

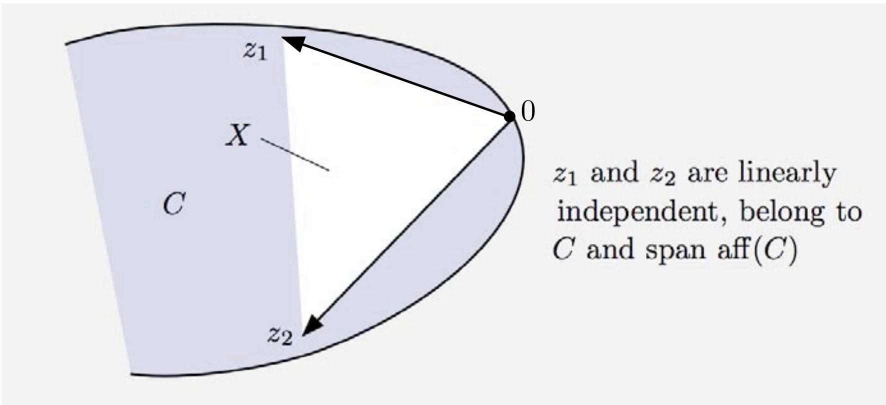


- Proof of case where $\overline{x} \in C$: See the figure.

- Proof of case where $\overline{x} \notin C$: Take sequence $\{x_k\} \subset C$ with $x_k \to \overline{x}$. Argue as in the figure.

ADDITIONAL MAJOR RESULTS

- Let C be a nonempty convex set.
  (a) $\mathrm{ri}(C)$ is a nonempty convex set, and has the same affine hull as C.
  (b) **Prolongation Lemma:** $x \in \mathrm{ri}(C)$ if and only if every line segment in C having x as one endpoint can be prolonged beyond x without leaving C.



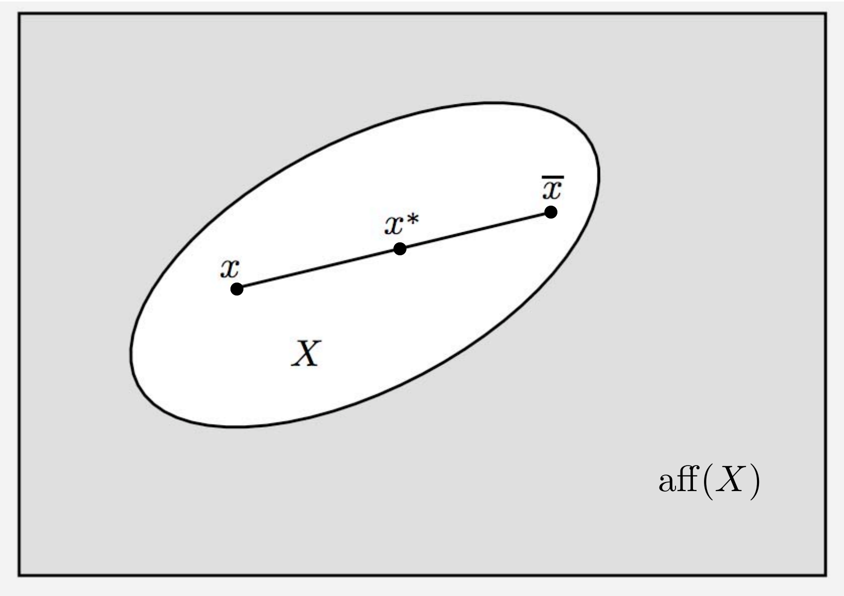
Proof: (a) Assume that $0 \in C$. We choose m linearly independent vectors $z_1, \ldots, z_m \in C$, where m is the dimension of $\mathrm{aff}(C)$, and we let

$$X = \left\{ \sum_{i=1}^{m} \alpha_i z_i \;\middle|\; \sum_{i=1}^{m} \alpha_i < 1,\ \alpha_i > 0,\ i = 1, \ldots, m \right\}$$

(b) => is clear by the def. of rel. interior. Reverse: take any $\overline{x} \in \mathrm{ri}(C)$; use Line Segment Principle.

OPTIMIZATION APPLICATION

• A concave function $f : \Re^n \mapsto \Re$ that attains its minimum over a convex set X at an $x^* \in \mathrm{ri}(X)$ must be constant over X.



Proof: (By contradiction) Let $x \in X$ be such that $f(x) > f(x^*)$. Prolong beyond x^* the line segment x-to-x^* to a point $\overline{x} \in X$. By concavity of f, we have for some $\alpha \in (0,1)$

$$f(x^*) \geq \alpha f(x) + (1 - \alpha) f(\overline{x}),$$

and since $f(x) > f(x^*)$, we must have $f(x^*) > f(\overline{x})$ - a contradiction. **Q.E.D.**

• **Corollary:** A nonconstant linear function cannot attain a minimum at an interior point of a convex set.

CALCULUS OF REL. INTERIORS: SUMMARY

- The ri(C) and cl(C) of a convex set C "differ very little."
  - Any set "between" ri(C) and cl(C) has the same relative interior and closure.
  - The relative interior of a convex set is equal to the relative interior of its closure.
  - The closure of the relative interior of a convex set is equal to its closure.

- Relative interior and closure commute with Cartesian product and inverse image under a linear transformation.

- Relative interior commutes with image under a linear transformation and vector sum, but closure does not.

- Neither relative interior nor closure commute with set intersection.

CLOSURE VS RELATIVE INTERIOR

- *Proposition*:

  (a) We have $\mathrm{cl}(C) = \mathrm{cl}\big(\mathrm{ri}(C)\big)$ and $\mathrm{ri}(C) = \mathrm{ri}\big(\mathrm{cl}(C)\big)$.

  (b) Let $\overline{C}$ be another nonempty convex set. Then the following three conditions are equivalent:

    (i) C and $\overline{C}$ have the same rel. interior.

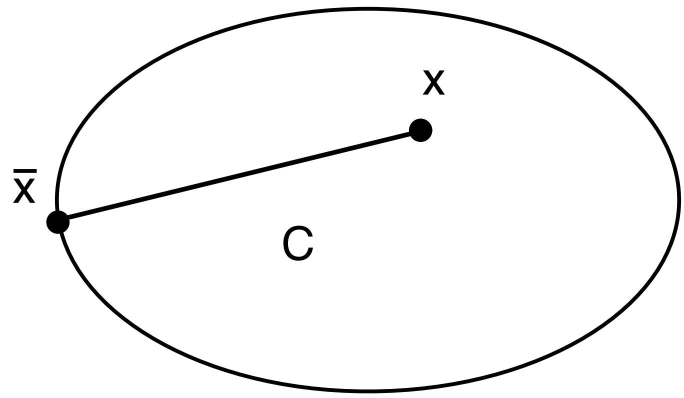
    (ii) C and $\overline{C}$ have the same closure.

    (iii) $\mathrm{ri}(C) \subset \overline{C} \subset \mathrm{cl}(C)$.

Proof: (a) Since $\mathrm{ri}(C) \subset C$, we have $\mathrm{cl}\big(\mathrm{ri}(C)\big) \subset \mathrm{cl}(C)$. Conversely, let $\overline{x} \in \mathrm{cl}(C)$. Let $x \in \mathrm{ri}(C)$. By the Line Segment Principle, we have

$$\alpha x + (1-\alpha)\overline{x} \in \mathrm{ri}(C), \qquad \forall\, \alpha \in (0,1].$$

Thus, $\overline{x}$ is the limit of a sequence that lies in $\mathrm{ri}(C)$, so $\overline{x} \in \mathrm{cl}\big(\mathrm{ri}(C)\big)$.



The proof of $\mathrm{ri}(C) = \mathrm{ri}\big(\mathrm{cl}(C)\big)$ is similar.

LINEAR TRANSFORMATIONS

- Let C be a nonempty convex subset of $\Re^n$ and let A be an $m \times n$ matrix.

  (a) We have $A \cdot \mathrm{ri}(C) = \mathrm{ri}(A \cdot C)$.

  (b) We have $A \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A \cdot C)$. Furthermore, if C is bounded, then $A \cdot \mathrm{cl}(C) = \mathrm{cl}(A \cdot C)$.

Proof: (a) Intuition: Spheres within C are mapped onto spheres within $A \cdot C$ (relative to the affine hull).

(b) We have $A \cdot \mathrm{cl}(C) \subset \mathrm{cl}(A \cdot C)$, since if a sequence $\{x_k\} \subset C$ converges to some $x \in \mathrm{cl}(C)$ then the sequence $\{Ax_k\}$, which belongs to $A \cdot C$, converges to Ax, implying that $Ax \in \mathrm{cl}(A \cdot C)$.

To show the converse, assuming that C is bounded, choose any $z \in \mathrm{cl}(A \cdot C)$. Then, there exists $\{x_k\} \subset C$ such that $Ax_k \to z$. Since C is bounded, $\{x_k\}$ has a subsequence that converges to some $x \in \mathrm{cl}(C)$, and we must have $Ax = z$. It follows that $z \in A \cdot \mathrm{cl}(C)$. **Q.E.D.**

Note that in general, we may have

$$A \cdot \mathrm{int}(C) \neq \mathrm{int}(A \cdot C), \qquad A \cdot \mathrm{cl}(C) \neq \mathrm{cl}(A \cdot C)$$

INTERSECTIONS AND VECTOR SUMS

- Let C_1 and C_2 be nonempty convex sets.

 (a) We have

$$\mathrm{ri}(C_1 + C_2) = \mathrm{ri}(C_1) + \mathrm{ri}(C_2),$$

$$\mathrm{cl}(C_1) + \mathrm{cl}(C_2) \subset \mathrm{cl}(C_1 + C_2)$$

 If one of C_1 and C_2 is bounded, then

$$\mathrm{cl}(C_1) + \mathrm{cl}(C_2) = \mathrm{cl}(C_1 + C_2)$$

 (b) We have

$$\mathrm{ri}(C_1)\cap\mathrm{ri}(C_2) \subset \mathrm{ri}(C_1\cap C_2), \ \ \mathrm{cl}(C_1\cap C_2) \subset \mathrm{cl}(C_1)\cap\mathrm{cl}(C_2)$$

 If $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$, then

$$\mathrm{ri}(C_1\cap C_2) = \mathrm{ri}(C_1)\cap\mathrm{ri}(C_2), \ \ \mathrm{cl}(C_1\cap C_2) = \mathrm{cl}(C_1)\cap\mathrm{cl}(C_2)$$

Proof of (a): $C_1 + C_2$ is the result of the linear transformation $(x_1, x_2) \mapsto x_1 + x_2$.

- Counterexample for (b):

$$C_1 = \{x \mid x \leq 0\}, \qquad C_2 = \{x \mid x \geq 0\}$$

$$C_1 = \{x \mid x < 0\}, \qquad C_2 = \{x \mid x > 0\}$$

CARTESIAN PRODUCT - GENERALIZATION

- Let C be convex set in $\Re^{n+m}$. For $x \in \Re^n$, let

$$C_x = \{y \mid (x, y) \in C\},$$

and let

$$D = \{x \mid C_x \neq \emptyset\}.$$

Then

$$\text{ri}(C) = \big\{(x, y) \mid x \in \text{ri}(D),\, y \in \text{ri}(C_x)\big\}.$$

Proof: Since D is projection of C on x-axis,

$$\text{ri}(D) = \big\{x \mid \text{there exists } y \in \Re^m \text{ with } (x, y) \in \text{ri}(C)\big\},$$

so that

$$\text{ri}(C) = \cup_{x \in \text{ri}(D)} \Big(M_x \cap \text{ri}(C)\Big),$$

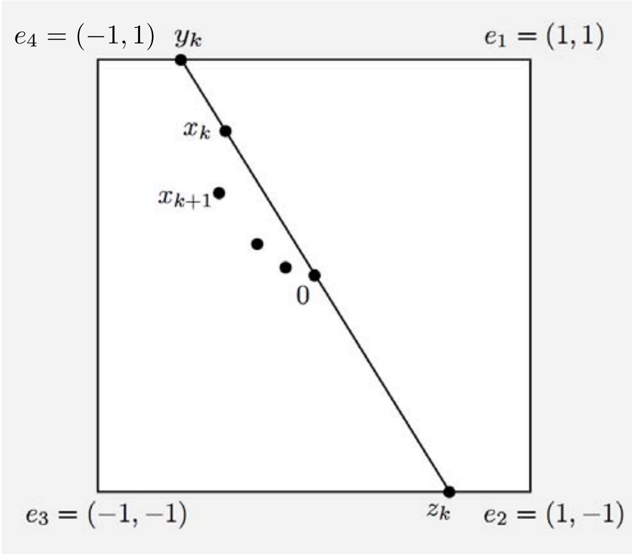
where $M_x = \big\{(x, y) \mid y \in \Re^m\big\}$. For every $x \in \text{ri}(D)$, we have

$$M_x \cap \text{ri}(C) = \text{ri}(M_x \cap C) = \big\{(x, y) \mid y \in \text{ri}(C_x)\big\}.$$

Combine the preceding two equations. **Q.E.D.**

CONTINUITY OF CONVEX FUNCTIONS

- If $f : \Re^n \mapsto \Re$ is convex, then it is continuous.



Proof: We will show that f is continuous at 0. By convexity, f is bounded within the unit cube by the max value of f over the corners of the cube.

Consider sequence $x_k \to 0$ and the sequences $y_k = x_k/\|x_k\|_\infty$, $z_k = -x_k/\|x_k\|_\infty$. Then

$$f(x_k) \leq \big(1 - \|x_k\|_\infty\big) f(0) + \|x_k\|_\infty f(y_k)$$

$$f(0) \leq \frac{\|x_k\|_\infty}{\|x_k\|_\infty + 1} f(z_k) + \frac{1}{\|x_k\|_\infty + 1} f(x_k)$$

Take limit as $k \to \infty$. Since $\|x_k\|_\infty \to 0$, we have

$$\limsup_{k\to\infty} \|x_k\|_\infty f(y_k) \leq 0, \quad \limsup_{k\to\infty} \frac{\|x_k\|_\infty}{\|x_k\|_\infty + 1} f(z_k) \leq 0$$

so $f(x_k) \to f(0)$. **Q.E.D.**

- Extension to continuity over ri(dom(f)).

CLOSURES OF FUNCTIONS

- The *closure* of a function $f : X \mapsto [-\infty, \infty]$ is the function $\mathrm{cl}\, f : \Re^n \mapsto [-\infty, \infty]$ with

$$\mathrm{epi}(\mathrm{cl}\, f) = \mathrm{cl}\big(\mathrm{epi}(f)\big)$$

- The *convex closure of* f is the function $\check{\mathrm{cl}}\, f$ with

$$\mathrm{epi}(\check{\mathrm{cl}}\, f) = \mathrm{cl}\Big(\mathrm{conv}\big(\mathrm{epi}(f)\big)\Big)$$

- *Proposition*: For any $f : X \mapsto [-\infty, \infty]$

$$\inf_{x \in X} f(x) = \inf_{x \in \Re^n} (\mathrm{cl}\, f)(x) = \inf_{x \in \Re^n} (\check{\mathrm{cl}}\, f)(x).$$

Also, any vector that attains the infimum of f over X also attains the infimum of $\mathrm{cl}\, f$ and $\check{\mathrm{cl}}\, f$.

- *Proposition*: For any $f : X \mapsto [-\infty, \infty]$:
  (a) $\mathrm{cl}\, f$ (or $\check{\mathrm{cl}}\, f$) is the greatest closed (or closed convex, resp.) function majorized by f.
  (b) If f is convex, then $\mathrm{cl}\, f$ is convex, and it is proper if and only if f is proper. Also,

$$(\mathrm{cl}\, f)(x) = f(x), \qquad \forall\, x \in \mathrm{ri}\big(\mathrm{dom}(f)\big),$$

and if $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$ and $y \in \mathrm{dom}(\mathrm{cl}\, f)$,

$$(\mathrm{cl}\, f)(y) = \lim_{\alpha \downarrow 0} f\big(y + \alpha(x - y)\big).$$

LECTURE 5

LECTURE OUTLINE

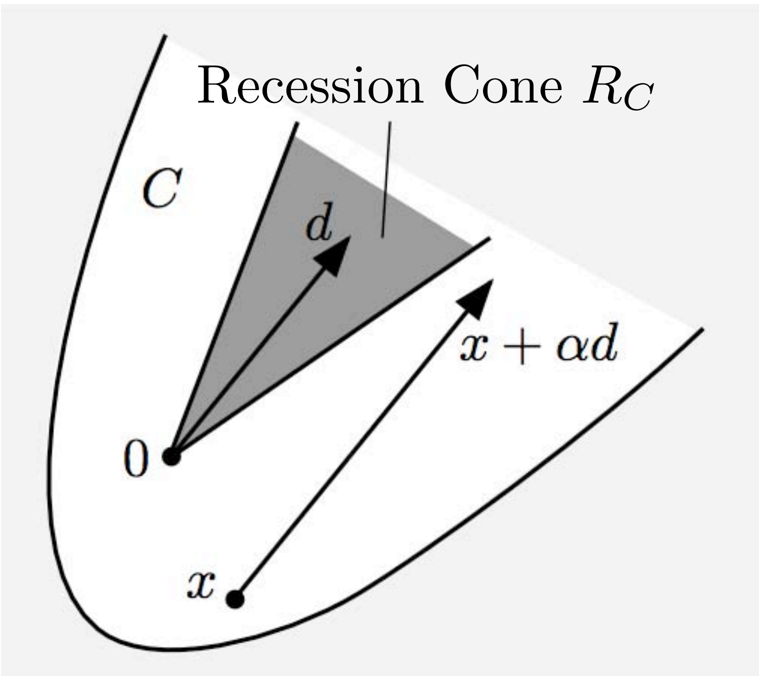
- Recession cones and lineality space
- Directions of recession of convex functions
- Local and global minima
- Existence of optimal solutions

Reading: Section 1.4, 3.1, 3.2

RECESSION CONE OF A CONVEX SET

• Given a nonempty convex set C, a vector d is a *direction of recession* if starting at **any** x in C and going indefinitely along d, we never cross the relative boundary of C to points outside C:

$$x + \alpha d \in C, \qquad \forall\, x \in C, \ \ \forall\, \alpha \geq 0$$



• *Recession cone* of C (denoted by R_C): The set of all directions of recession.

• R_C is a cone containing the origin.

RECESSION CONE THEOREM

- Let C be a nonempty closed convex set.

(a) The recession cone R_C is a closed convex cone.

(b) A vector d belongs to R_C if and only if there exists *some* vector $x \in C$ such that $x + \alpha d \in C$ for all $\alpha \geq 0$.

(c) R_C contains a nonzero direction if and only if C is unbounded.

(d) The recession cones of C and $\mathrm{ri}(C)$ are equal.

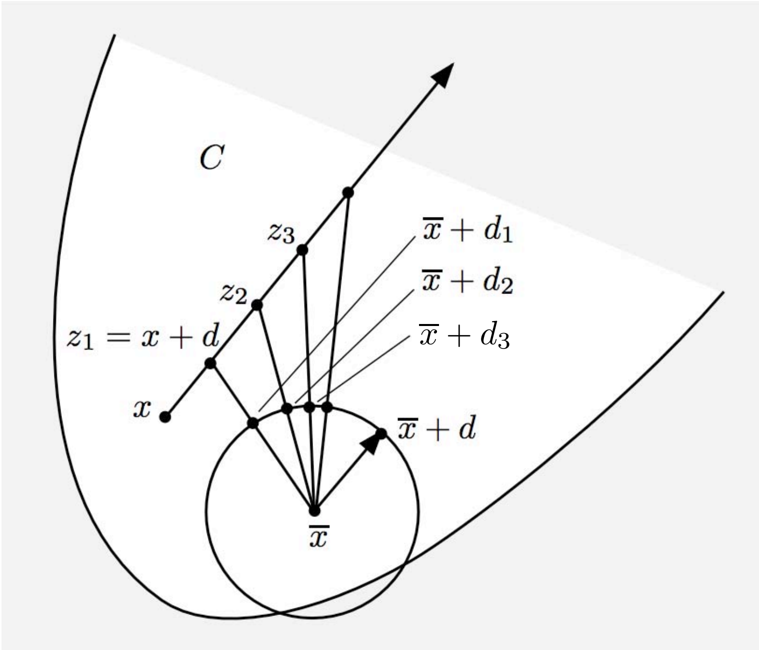
(e) If D is another closed convex set such that $C \cap D \neq \emptyset$, we have

$$R_{C \cap D} = R_C \cap R_D$$

More generally, for any collection of closed convex sets C_i, $i \in I$, where I is an arbitrary index set and $\cap_{i \in I} C_i$ is nonempty, we have

$$R_{\cap_{i \in I} C_i} = \cap_{i \in I} R_{C_i}$$

PROOF OF PART (B)



- Let $d \neq 0$ be such that there exists a vector $x \in C$ with $x + \alpha d \in C$ for all $\alpha \geq 0$. We fix $\overline{x} \in C$ and $\alpha > 0$, and we show that $\overline{x} + \alpha d \in C$. By scaling d, it is enough to show that $\overline{x} + d \in C$.

For $k = 1, 2, \ldots$, let

$$z_k = x + kd, \qquad d_k = \frac{(z_k - \overline{x})}{\|z_k - \overline{x}\|}\|d\|$$

We have

$$\frac{d_k}{\|d\|} = \frac{\|z_k - x\|}{\|z_k - \overline{x}\|}\frac{d}{\|d\|} + \frac{x - \overline{x}}{\|z_k - \overline{x}\|}, \quad \frac{\|z_k - x\|}{\|z_k - \overline{x}\|} \to 1, \quad \frac{x - \overline{x}}{\|z_k - \overline{x}\|} \to 0,$$

so $d_k \to d$ and $\overline{x} + d_k \to \overline{x} + d$. Use the convexity and closedness of C to conclude that $\overline{x} + d \in C$.

LINEALITY SPACE

• The *lineality space* of a convex set C, denoted by L_C, is the subspace of vectors d such that $d \in R_C$ and $-d \in R_C$:

$$L_C = R_C \cap (-R_C)$$

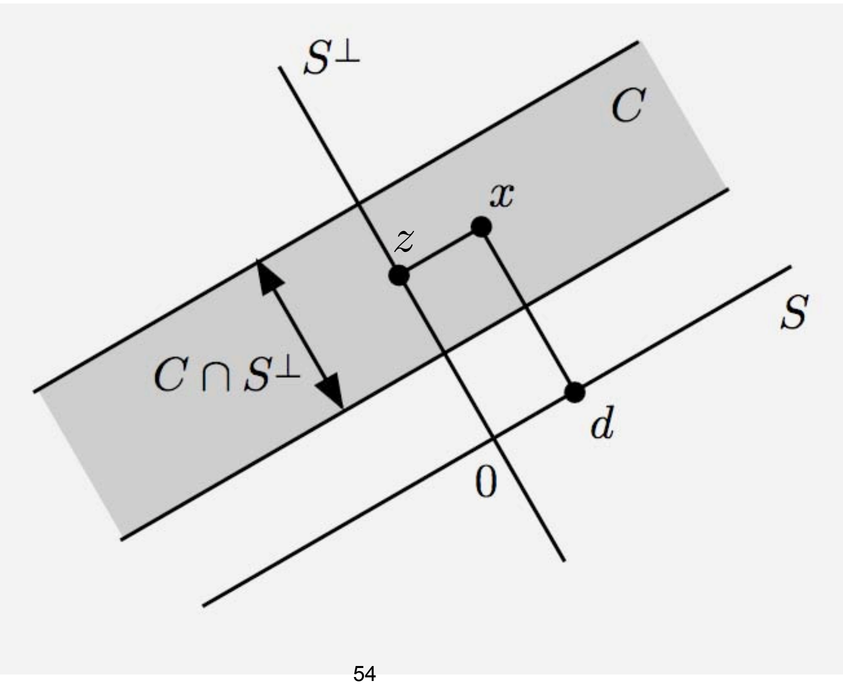
• If $d \in L_C$, the entire line defined by d is contained in C, starting at any point of C.

• *Decomposition of a Convex Set*: Let C be a nonempty convex subset of $\Re^n$. Then,

$$C = L_C + (C \cap L_C^{\perp}).$$

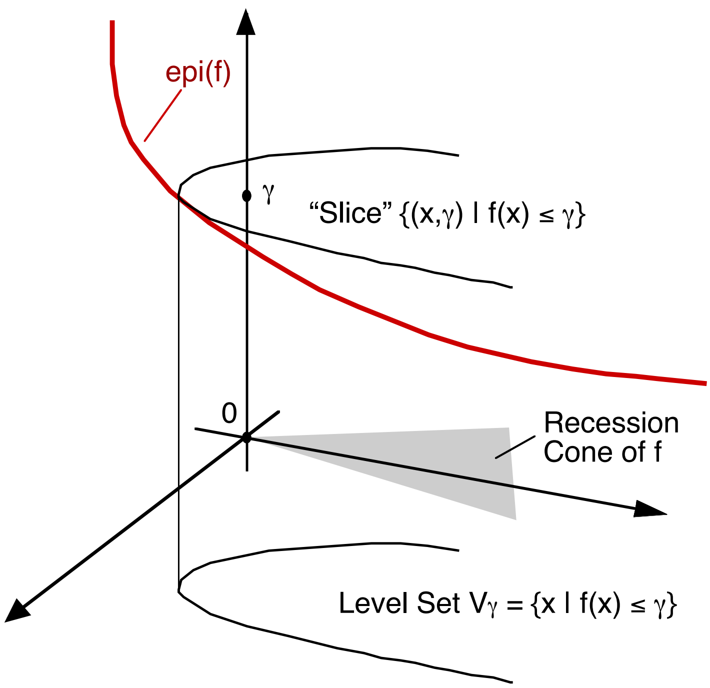
• Allows us to prove properties of C on $C \cap L_C^{\perp}$ and extend them to C.

• True also if L_C is replaced by a subspace $S \subset L_C$.



DIRECTIONS OF RECESSION OF A FN

- We aim to characterize directions of monotonic decrease of convex functions.
- Some basic geometric observations:
  - The "horizontal directions" in the recession cone of the epigraph of a convex function f are directions along which the level sets are unbounded.
  - Along these directions the level sets $\{x \mid f(x) \leq \gamma\}$ are unbounded and f is monotonically nondecreasing.
- These are the *directions of recession* of f.



RECESSION CONE OF LEVEL SETS

- *Proposition*: Let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed proper convex function and consider the level sets $V_\gamma = \big\{x \mid f(x) \leq \gamma\big\}$, where γ is a scalar. Then:

  (a) All the nonempty level sets V_γ have the same recession cone:

$$R_{V_\gamma} = \big\{d \mid (d, 0) \in R_{\text{epi}(f)}\big\}$$

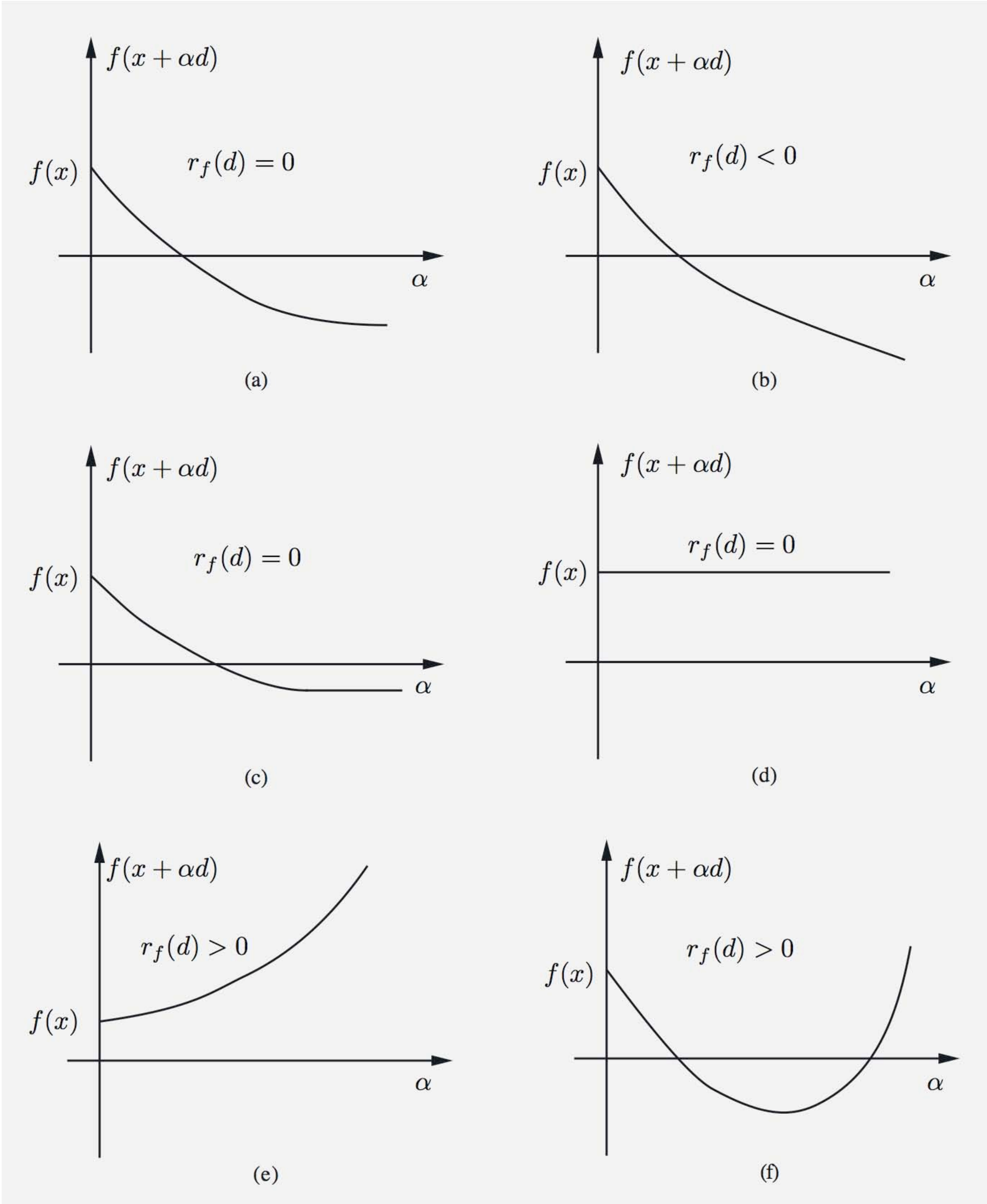
  (b) If one nonempty level set V_γ is compact, then all level sets are compact.

Proof: (a) Just translate to math the fact that

$R_{V_\gamma} =$ the "horizontal" directions of recession of epi(f)

(b) Follows from (a).

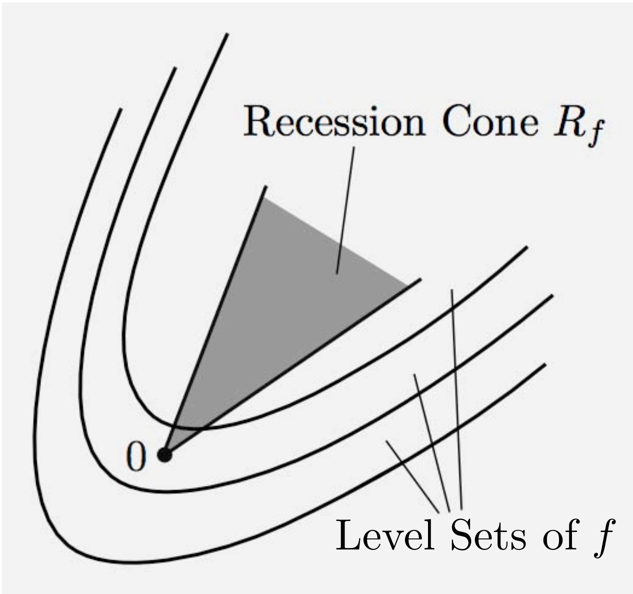
DESCENT BEHAVIOR OF A CONVEX FN



- y is a direction of recession in (a)-(d).
- This behavior is *independent of the starting point* x, as long as $x \in \text{dom}(f)$.

RECESSION CONE OF A CONVEX FUNCTION

- For a closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$, the (common) recession cone of the nonempty level sets $V_\gamma = \{x \mid f(x) \le \gamma\}$, $\gamma \in \Re$, is the *recession cone of* f, and is denoted by R_f.



- Terminology:
  - $d \in R_f$: a *direction of recession of* f.
  - $L_f = R_f \cap (-R_f)$: the *lineality space of* f.
  - $d \in L_f$: a *direction of constancy* of f.

- **Example:** For the pos. semidefinite quadratic

$$f(x) = x'Qx + a'x + b,$$

the recession cone and constancy space are

$$R_f = \{d \mid Qd = 0,\ a'd \le 0\}, \quad L_f = \{d \mid Qd = 0,\ a'd = 0\}$$

RECESSION FUNCTION

- Function $r_f : \Re^n \mapsto (-\infty, \infty]$ whose epigraph is $R_{\mathrm{epi}(f)}$ is the *recession function* of f.
- Characterizes the recession cone:

$$R_f = \big\{d \mid r_f(d) \le 0\big\}, \quad L_f = \big\{d \mid r_f(d) = r_f(-d) = 0\big\}$$

since $R_f = \{(d, 0) \in R_{\mathrm{epi}(f)}\}$.

- Can be shown that

$$r_f(d) = \sup_{\alpha > 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \lim_{\alpha \to \infty} \frac{f(x + \alpha d) - f(x)}{\alpha}$$

- Thus $r_f(d)$ is the "asymptotic slope" of f in the direction d. In fact,

$$r_f(d) = \lim_{\alpha \to \infty} \nabla f(x + \alpha d)' d, \qquad \forall\, x, d \in \Re^n$$

if f is differentiable.

- Calculus of recession functions:

$$r_{f_1 + \cdots + f_m}(d) = r_{f_1}(d) + \cdots + r_{f_m}(d),$$

$$r_{\sup_{i \in I} f_i}(d) = \sup_{i \in I} r_{f_i}(d)$$

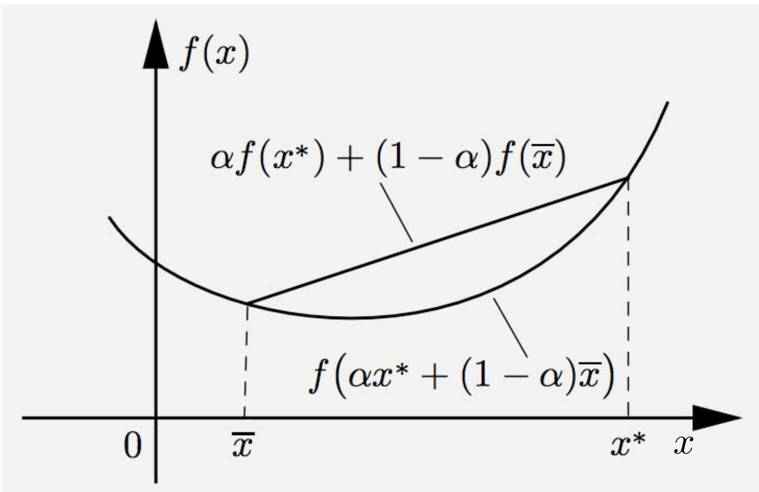
LOCAL AND GLOBAL MINIMA

- Consider minimizing $f : \Re^n \mapsto (-\infty, \infty]$ over a set $X \subset \Re^n$
- x is **feasible** if $x \in X \cap \text{dom}(f)$
- x^* is a (global) **minimum** of f over X if x^* is feasible and $f(x^*) = \inf_{x \in X} f(x)$
- x^* is a **local minimum** of f over X if x^* is a minimum of f over a set $X \cap \{x \mid \|x - x^*\| \leq \epsilon\}$

Proposition: If X is convex and f is convex, then:

(a) A local minimum of f over X is also a global minimum of f over X.

(b) If f is strictly convex, then there exists at most one global minimum of f over X.



EXISTENCE OF OPTIMAL SOLUTIONS

- The set of minima of a proper $f : \Re^n \mapsto (-\infty, \infty]$ is the intersection of its nonempty level sets.

- The set of minima of f is nonempty and compact if the level sets of f are compact.

- **(An Extension of the) Weierstrass' Theorem:** The set of minima of f over X is nonempty and compact if X is closed, f is lower semicontinuous over X, and one of the following conditions holds:

  (1) X is bounded.

  (2) Some set $\{x \in X \mid f(x) \leq \gamma\}$ is nonempty and bounded.

  (3) For every sequence $\{x_k\} \subset X$ s. t. $\|x_k\| \to \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$. (Coercivity property).

Proof: In all cases the level sets of $f \cap X$ are compact. **Q.E.D.**

EXISTENCE OF SOLUTIONS - CONVEX CASE

- **Weierstrass' Theorem specialized to convex functions:** Let X be a closed convex subset of $\Re^n$, and let $f : \Re^n \mapsto (-\infty, \infty]$ be closed convex with $X \cap \text{dom}(f) \neq \emptyset$. The set of minima of f over X is nonempty and compact if and only if X and f have no common nonzero direction of recession.

Proof: Let $f^* = \inf_{x \in X} f(x)$ and note that $f^* < \infty$ since $X \cap \text{dom}(f) \neq \emptyset$. Let $\{\gamma_k\}$ be a scalar sequence with $\gamma_k \downarrow f^*$, and consider the sets

$$V_k = \big\{x \mid f(x) \leq \gamma_k\big\}.$$

Then the set of minima of f over X is

$$X^* = \cap_{k=1}^{\infty} (X \cap V_k).$$

The sets $X \cap V_k$ are nonempty and have $R_X \cap R_f$ as their common recession cone, which is also the recession cone of X^*, when $X^* \neq \emptyset$. It follows that X^* is nonempty and compact if and only if $R_X \cap R_f = \{0\}$. **Q.E.D.**

EXISTENCE OF SOLUTION, SUM OF FNS

- Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be closed proper convex functions such that the function

$$f = f_1 + \cdots + f_m$$

is proper. Assume that a single function f_i satisfies $r_{f_i}(d) = \infty$ for all $d \neq 0$. Then the set of minima of f is nonempty and compact.

- **Proof:** We have $r_f(d) = \infty$ for all $d \neq 0$ since $r_f(d) = \sum_{i=1}^m r_{f_i}(d)$. Hence f has no nonzero directions of recession. **Q.E.D.**

- True also for $f = \max\{f_1, \ldots, f_m\}$.
- **Example of application:** If one of the f_i is positive definite quadratic, the set of minima of the sum f is nonempty and compact.
- Also f has a unique minimum because the positive definite quadratic is strictly convex, which makes f strictly convex.

LECTURE 6

LECTURE OUTLINE

- Nonemptiness of closed set intersections
  - Simple version
  - More complex version
- Existence of optimal solutions
- Preservation of closure under linear transformation
- Hyperplanes

ROLE OF CLOSED SET INTERSECTIONS I

- **A fundamental question:** Given a sequence of nonempty closed sets $\{C_k\}$ in $\Re^n$ with $C_{k+1} \subset C_k$ for all k, when is $\cap_{k=0}^{\infty} C_k$ nonempty?

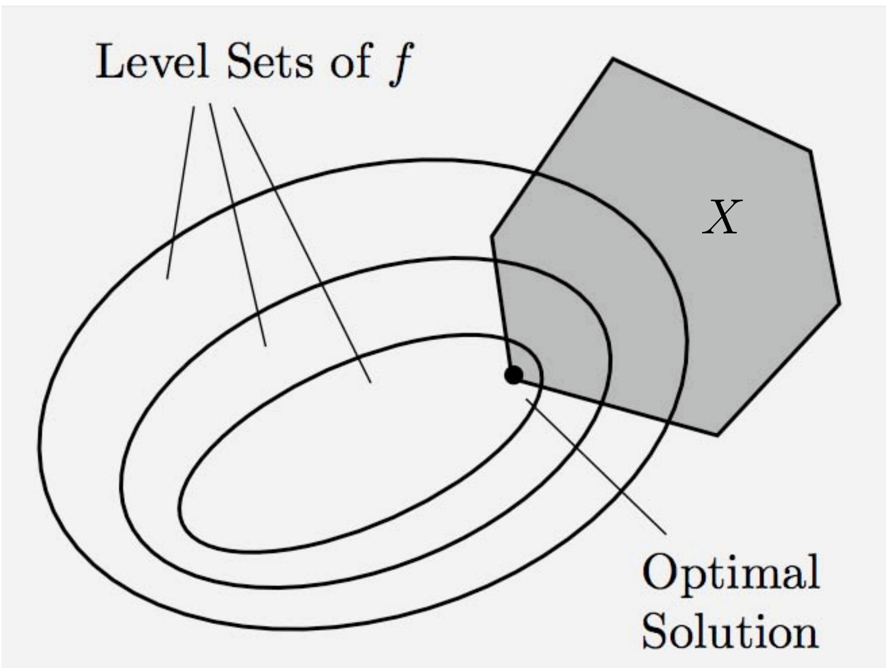
- Set intersection theorems are significant in at least three major contexts, which we will discuss in what follows:

Does a function $f : \Re^n \mapsto (-\infty, \infty]$ attain a minimum over a set X?

This is true if and only if

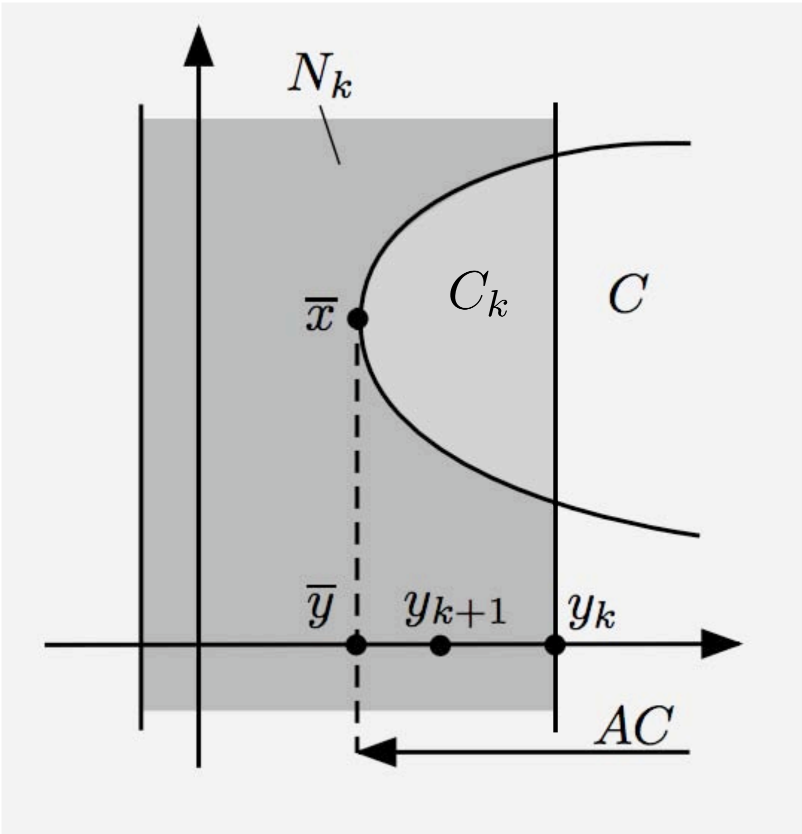
$$\text{Intersection of nonempty } \big\{x \in X \mid f(x) \leq \gamma_k\big\}$$

is nonempty.



ROLE OF CLOSED SET INTERSECTIONS II

If C is closed and A is a matrix, is AC closed?



- **If C_1 and C_2 are closed, is $C_1 + C_2$ closed?**
  - This is a special case.
  - Write

$$C_1 + C_2 = A(C_1 \times C_2),$$

where $A(x_1, x_2) = x_1 + x_2$.

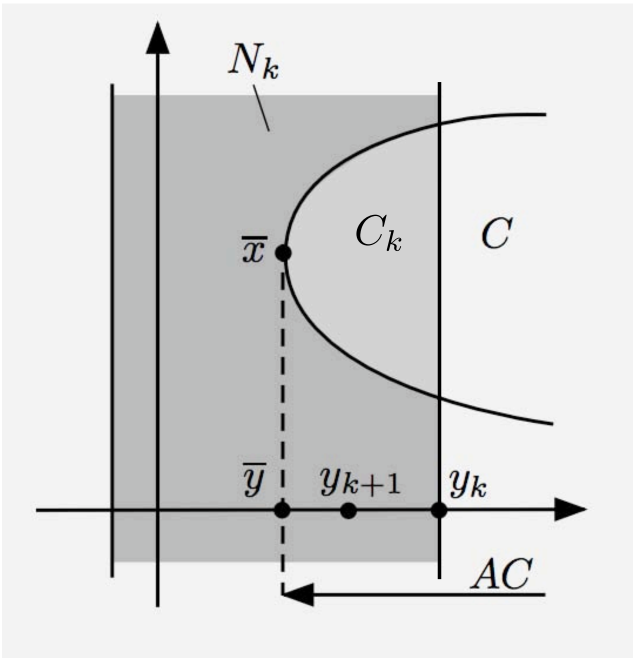
CLOSURE UNDER LINEAR TRANSFORMATION

- Let C be a nonempty closed convex, and let A be a matrix with nullspace $N(A)$. Then AC is closed if $R_C \cap N(A) = \{0\}$.

Proof: Let $\{y_k\} \subset AC$ with $y_k \to \overline{y}$. Define the nested sequence $C_k = C \cap N_k$, where

$$N_k = \{x \mid Ax \in W_k\}, \quad W_k = \big\{z \mid \|z-\overline{y}\| \leq \|y_k-\overline{y}\|\big\}$$

We have $R_{N_k} = N(A)$, so C_k is compact, and $\{C_k\}$ has nonempty intersection. **Q.E.D.**



- **A special case:** $C_1 + C_2$ is closed if C_1, C_2 are closed and one of the two is compact. [Write $C_1+C_2 = A(C_1 \times C_2)$, where $A(x_1, x_2) = x_1+x_2$.]

- **Related theorem:** AX is closed if X is polyhedral. To be shown later by a more refined method.

ROLE OF CLOSED SET INTERSECTIONS III

- Let $F : \Re^{n+m} \mapsto (-\infty, \infty]$ be a closed proper convex function, and consider

$$f(x) = \inf_{z \in \Re^m} F(x, z)$$

- **If $F(x, z)$ is closed, is $f(x)$ closed?**
  - Critical question in duality theory.
- **1st fact:** If F is convex, then f is also convex.
- **2nd fact:**

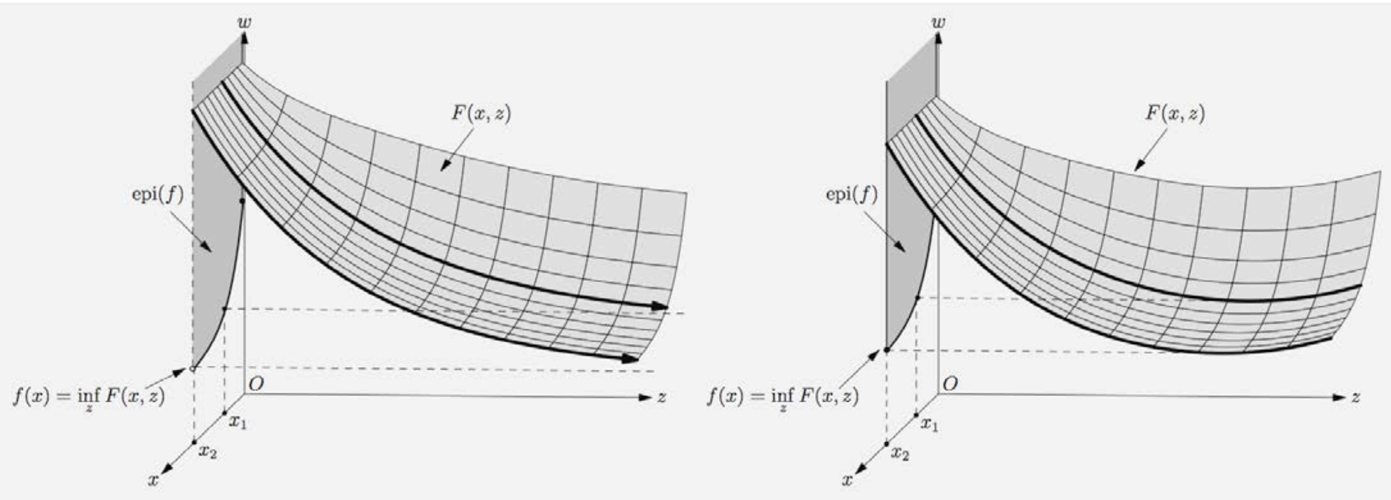
$$P\big(\text{epi}(F)\big) \subset \text{epi}(f) \subset \text{cl}\Big(P\big(\text{epi}(F)\big)\Big),$$

where $P(\cdot)$ denotes projection on the space of (x, w), i.e., for any subset S of $\Re^{n+m+1}$, $P(S) = \big\{(x, w) \mid (x, z, w) \in S\big\}$.

- Thus, if F is closed and there is structure guaranteeing that the projection preserves closedness, then f is closed.
- ... but convexity and closedness of F does not guarantee closedness of f.

PARTIAL MINIMIZATION: VISUALIZATION

- Connection of preservation of closedness under partial minimization and attainment of infimum over z for fixed x.



- **Counterexample:** Let

$$F(x,z) = \begin{cases} e^{-\sqrt{xz}} & \text{if } x \geq 0,\ z \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

- F convex and closed, but

$$f(x) = \inf_{z \in \Re} F(x,z) = \begin{cases} 0 & \text{if } x > 0, \\ 1 & \text{if } x = 0, \\ \infty & \text{if } x < 0, \end{cases}$$

is not closed.

PARTIAL MINIMIZATION THEOREM

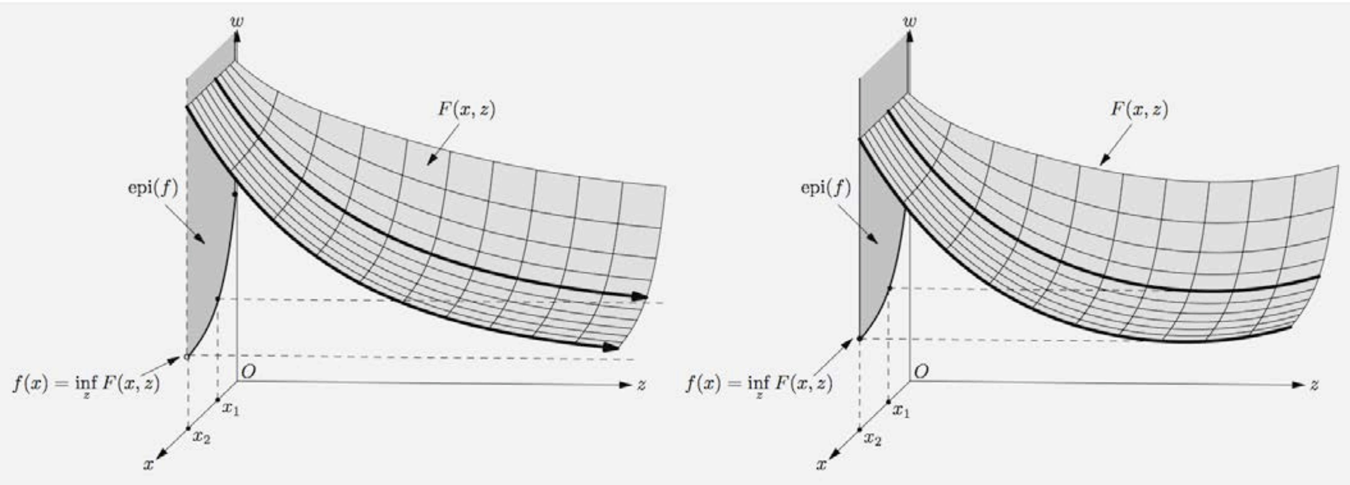
- Let $F : \Re^{n+m} \mapsto (-\infty, \infty]$ be a closed proper convex function, and consider $f(x) = \inf_{z \in \Re^m} F(x, z)$.

- Every set intersection theorem yields a closedness result. The simplest case is the following:

- **Preservation of Closedness Under Compactness:** If there exist $\overline{x} \in \Re^n$, $\overline{\gamma} \in \Re$ such that the set

$$\big\{z \mid F(\overline{x}, z) \leq \overline{\gamma}\big\}$$

is nonempty and compact, then f is convex, closed, and proper. Also, for each $x \in \text{dom}(f)$, the set of minima of $F(x, \cdot)$ is nonempty and compact.



MORE REFINED ANALYSIS - A SUMMARY

- We noted that there is a common mathematical root to three basic questions:
  - Existence of of solutions of convex optimization problems
  - Preservation of closedness of convex sets under a linear transformation
  - Preservation of closedness of convex functions under partial minimization
- The common root is the question of nonemptiness of intersection of a nested sequence of closed sets
- The preceding development in this lecture resolved this question by assuming that all the sets in the sequence are compact
- A more refined development makes instead various assumptions about the directions of recession and the lineality space of the sets in the sequence
- Once the appropriately refined set intersection theory is developed, sharper results relating to the three questions can be obtained
- The remaining slides up to hyperplanes summarize this development as an aid for self-study using Sections 1.4.2, 1.4.3, and Sections 3.2, 3.3

ASYMPTOTIC SEQUENCES

- Given nested sequence $\{C_k\}$ of closed convex sets, $\{x_k\}$ is an *asymptotic sequence* if

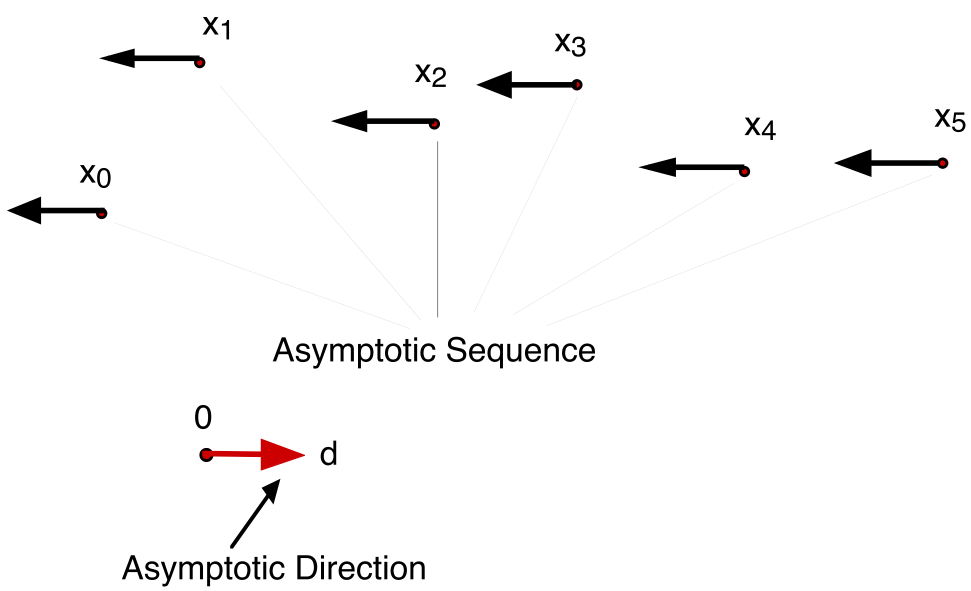
$$x_k \in C_k, \qquad x_k \neq 0, \qquad k = 0, 1, \ldots$$

$$\|x_k\| \to \infty, \qquad \frac{x_k}{\|x_k\|} \to \frac{d}{\|d\|}$$

where d is a nonzero common direction of recession of the sets C_k.

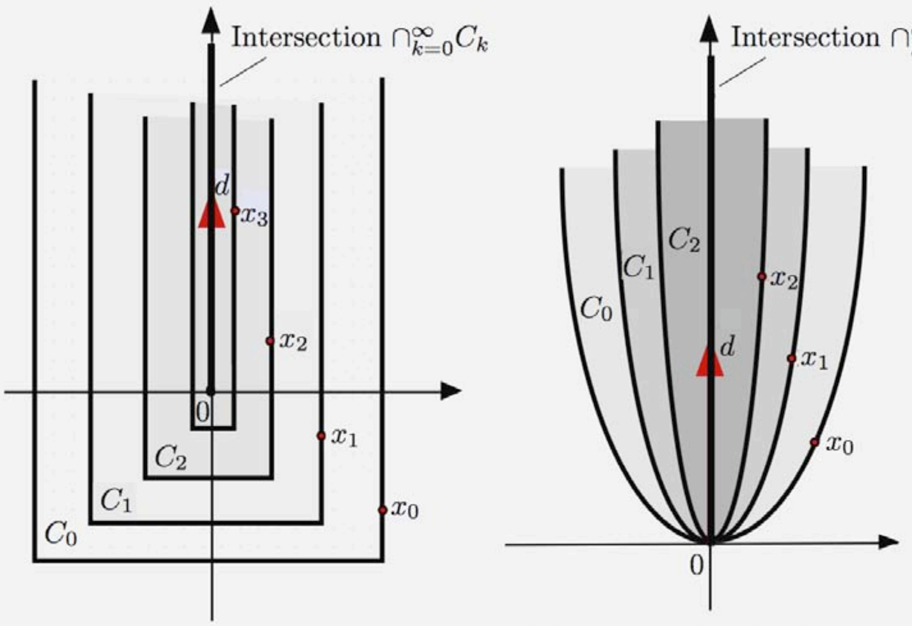
- As a special case we define asymptotic sequence of a closed convex set C (use $C_k \equiv C$).
- Every unbounded $\{x_k\}$ with $x_k \in C_k$ has an asymptotic subsequence.
- $\{x_k\}$ is called *retractive* if for some $\overline{k}$, we have

$$x_k - d \in C_k, \quad \forall\, k \geq \overline{k}.$$



RETRACTIVE SEQUENCES

• A nested sequence $\{C_k\}$ of closed convex sets is *retractive* if all its asymptotic sequences are retractive.



(a) Retractive Set Sequence (b) Nonretractive Set Sequence

• A closed halfspace (viewed as a sequence with identical components) is retractive.

• Intersections and Cartesian products of retractive set sequences are retractive.

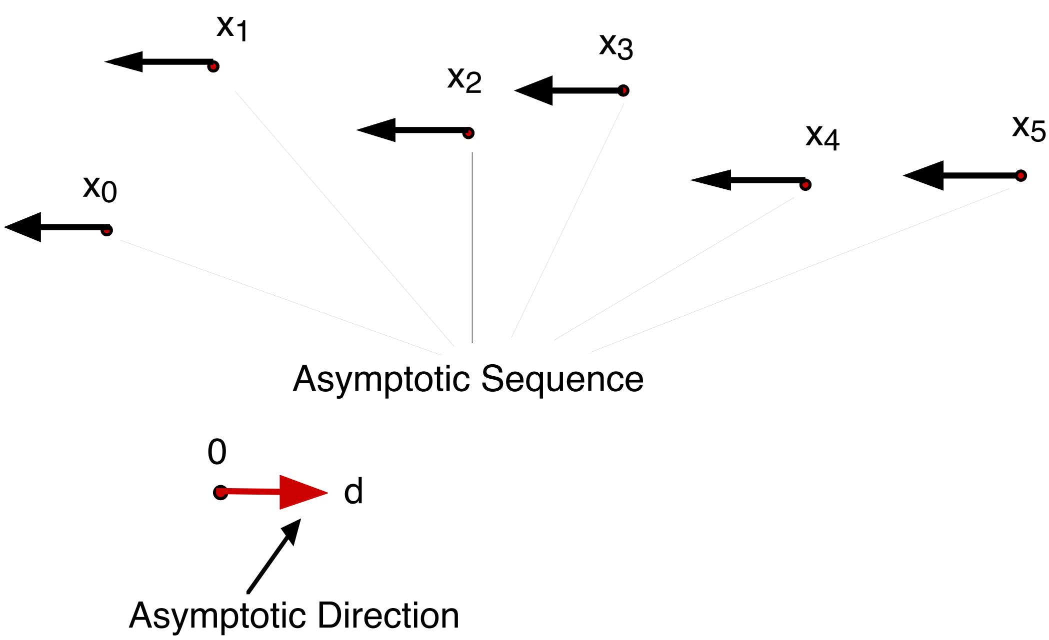
• A polyhedral set is retractive. Also the vector sum of a convex compact set and a retractive convex set is retractive.

• Nonpolyhedral cones and level sets of quadratic functions need not be retractive.

SET INTERSECTION THEOREM I

Proposition: If $\{C_k\}$ is retractive, then $\cap_{k=0}^{\infty} C_k$ is nonempty.

- Key proof ideas:
  (a) The intersection $\cap_{k=0}^{\infty} C_k$ is empty iff the sequence $\{x_k\}$ of minimum norm vectors of C_k is unbounded (so a subsequence is asymptotic).
  (b) An asymptotic sequence $\{x_k\}$ of minimum norm vectors cannot be retractive, because such a sequence eventually gets closer to 0 when shifted opposite to the asymptotic direction.



SET INTERSECTION THEOREM II

Proposition: Let $\{C_k\}$ be a nested sequence of nonempty closed convex sets, and X be a retractive set such that all the sets $\overline{C}_k = X \cap C_k$ are nonempty. Assume that

$$R_X \cap R \subset L,$$

where

$$R = \cap_{k=0}^{\infty} R_{C_k}, \qquad L = \cap_{k=0}^{\infty} L_{C_k}$$

Then $\{\overline{C}_k\}$ is retractive and $\cap_{k=0}^{\infty} \overline{C}_k$ is nonempty.

- Special cases:
  - $X = \Re^n$, $R = L$ ("cylindrical" sets C_k)
  - $R_X \cap R = \{0\}$ (no nonzero common recession direction of X and $\cap_k C_k$)

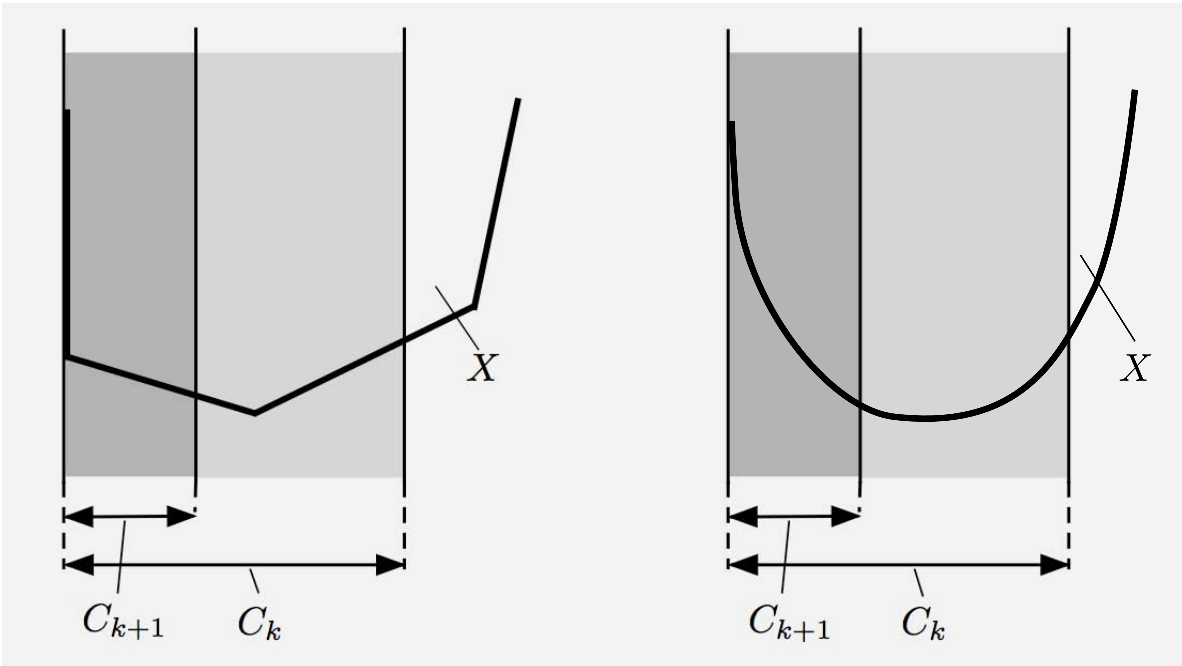
Proof: The set of common directions of recession of $\overline{C}_k$ is $R_X \cap R$. For any asymptotic sequence $\{x_k\}$ corresponding to $d \in R_X \cap R$:

(1) $x_k - d \in C_k$ (because $d \in L$)

(2) $x_k - d \in X$ (because X is retractive)

So $\{\overline{C}_k\}$ is retractive.

NEED TO ASSUME THAT X IS RETRACTIVE



Consider $\quad \cap_{k=0}^{\infty} \overline{C}_k, \quad$ with $\quad \overline{C}_k = X \cap C_k$

- The condition $R_X \cap R \subset L$ holds
- In the figure on the left, X is polyhedral.
- In the figure on the right, X is nonpolyhedral and nonretrative, and

$$\cap_{k=0}^{\infty} \overline{C}_k = \emptyset$$

LINEAR AND QUADRATIC PROGRAMMING

- **Theorem:** Let

$$f(x) = x'Qx + c'x, \quad X = \{x \mid a_j'x + b_j \leq 0,\ j = 1, \ldots, r\}$$

where Q is symmetric positive semidefinite. If the minimal value of f over X is finite, there exists a minimum of f over X.

Proof: (Outline) Write

$$\text{Set of Minima} = \cap_{k=0}^{\infty}\big(X \cap \{x \mid x'Qx + c'x \leq \gamma_k\}\big)$$

with

$$\gamma_k \downarrow f^* = \inf_{x \in X} f(x).$$

Verify the condition $R_X \cap R \subset L$ of the preceding set intersection theorem, where R and L are the sets of common recession and lineality directions of the sets

$$\{x \mid x'Qx + c'x \leq \gamma_k\}$$

Q.E.D.

CLOSURE UNDER LINEAR TRANSFORMATION

- Let C be a nonempty closed convex, and let A be a matrix with nullspace $N(A)$.

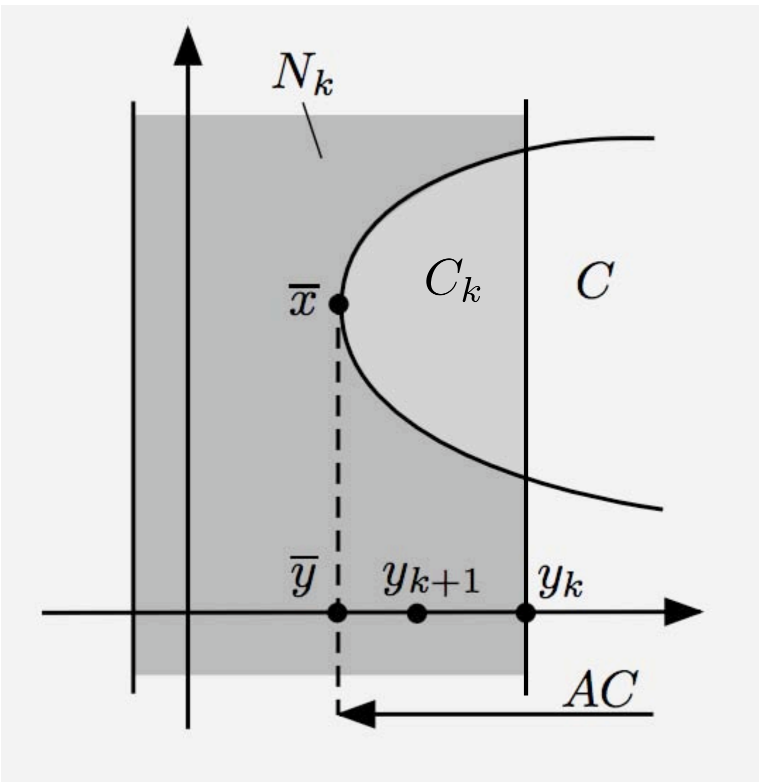
  (a) AC is closed if $R_C \cap N(A) \subset L_C$.

  (b) $A(X \cap C)$ is closed if X is a retractive set and

$$R_X \cap R_C \cap N(A) \subset L_C,$$

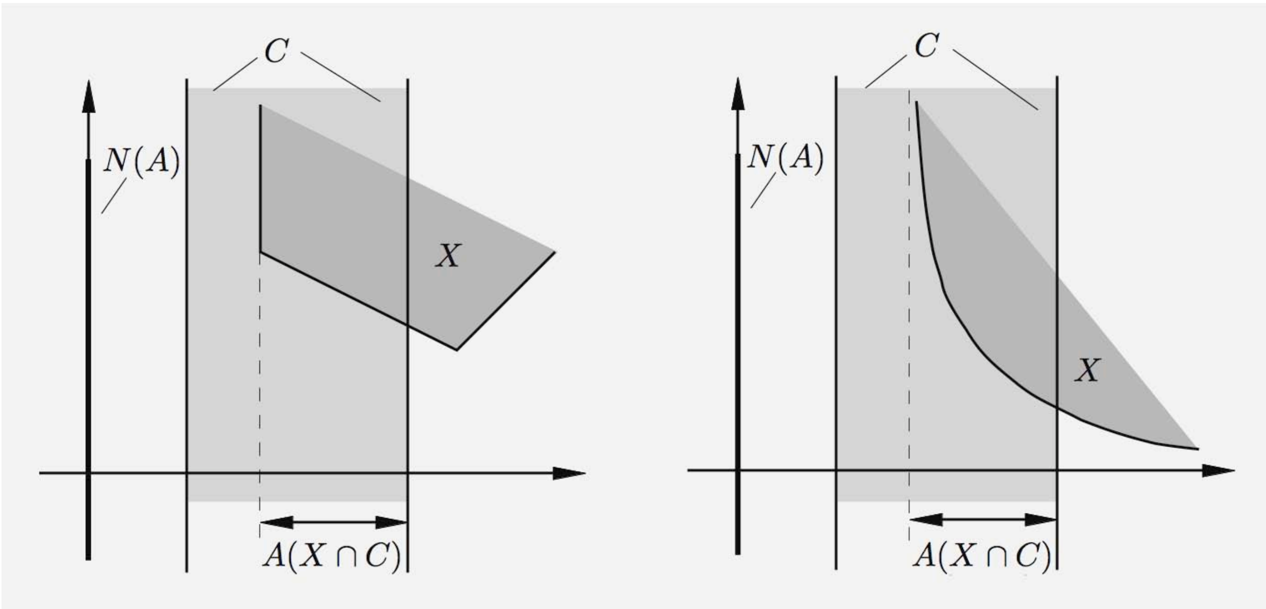
Proof: (Outline) Let $\{y_k\} \subset AC$ with $y_k \to \overline{y}$. We prove $\cap_{k=0}^{\infty} C_k \neq \emptyset$, where $C_k = C \cap N_k$, and

$$N_k = \{x \mid Ax \in W_k\}, \quad W_k = \big\{z \mid \|z-\overline{y}\| \leq \|y_k-\overline{y}\|\big\}$$



- **Special Case:** AX is closed if X is polyhedral.

NEED TO ASSUME THAT X IS RETRACTIVE



Consider closedness of $A(X \cap C)$

- In both examples the condition

$$R_X \cap R_C \cap N(A) \subset L_C$$

is satisfied.

- However, in the example on the right, X is not retractive, and the set $A(X \cap C)$ is not closed.

CLOSEDNESS OF VECTOR SUMS

- Let $C_1, \ldots, C_m$ be nonempty closed convex subsets of $\Re^n$ such that the equality $d_1 + \cdots + d_m = 0$ for some vectors $d_i \in R_{C_i}$ implies that $d_i = 0$ for all $i = 1, \ldots, m$. Then $C_1 + \cdots + C_m$ is a closed set.

- **Special Case:** If C_1 and $-C_2$ are closed convex sets, then $C_1 - C_2$ is closed if $R_{C_1} \cap R_{C_2} = \{0\}$.

Proof: The Cartesian product $C = C_1 \times \cdots \times C_m$ is closed convex, and its recession cone is $R_C = R_{C_1} \times \cdots \times R_{C_m}$. Let A be defined by

$$A(x_1, \ldots, x_m) = x_1 + \cdots + x_m$$

Then

$$A\,C = C_1 + \cdots + C_m,$$

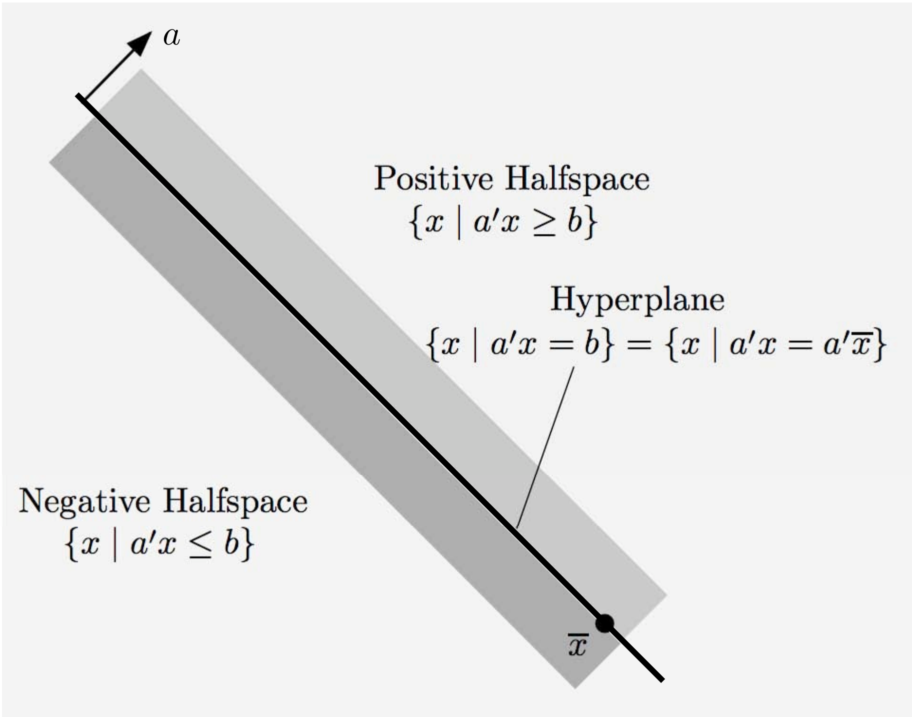
and

$$N(A) = \Big\{(d_1, \ldots, d_m) \mid d_1 + \cdots + d_m = 0\Big\}$$

$$R_C \cap N(A) = \Big\{(d_1, \ldots, d_m) \mid d_1 + \cdots + d_m = 0,\ d_i \in R_{C_i},\ \forall\, i\Big\}$$

By the given condition, $R_C \cap N(A) = \{0\}$, so $A\,C$ is closed. **Q.E.D.**

HYPERPLANES



- A *hyperplane* is a set of the form $\{x \mid a'x = b\}$, where a is nonzero vector in $\Re^n$ and b is a scalar.

- We say that two sets C_1 and C_2 are *separated by a hyperplane* $H = \{x \mid a'x = b\}$ if each lies in a different closed halfspace associated with H, i.e.,

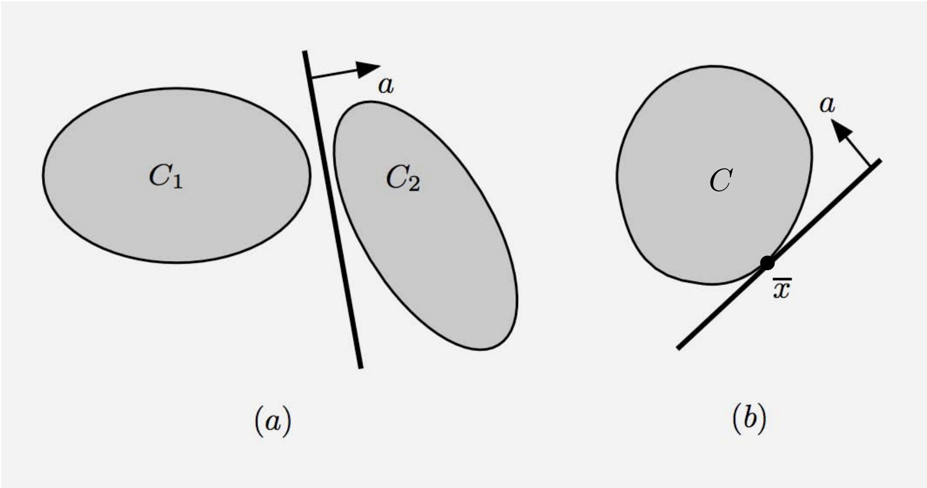
$$\text{either} \quad a'x_1 \leq b \leq a'x_2, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2,$$

$$\text{or} \quad a'x_2 \leq b \leq a'x_1, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2$$

- If $\overline{x}$ belongs to the closure of a set C, a hyperplane that separates C and the singleton set $\{\overline{x}\}$ is said be *supporting* C *at* $\overline{x}$.

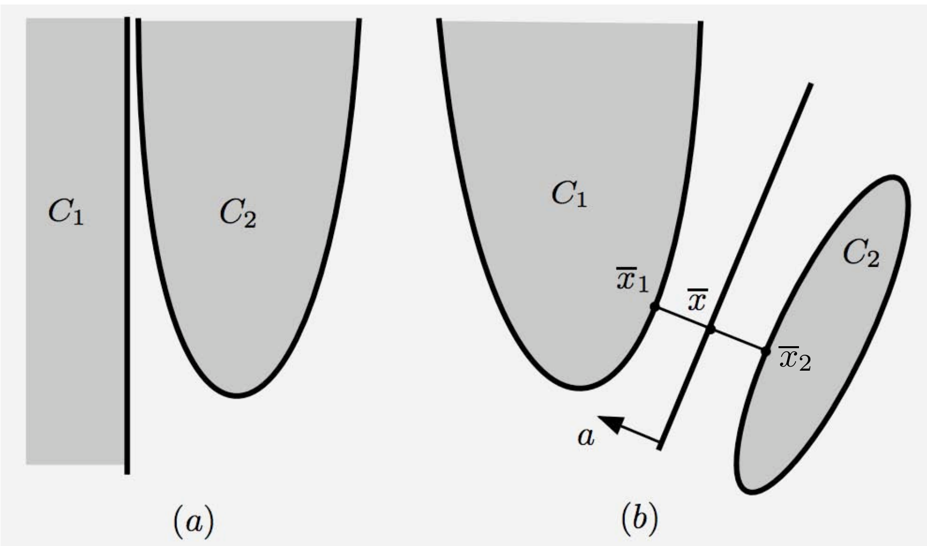
VISUALIZATION

- Separating and supporting hyperplanes:



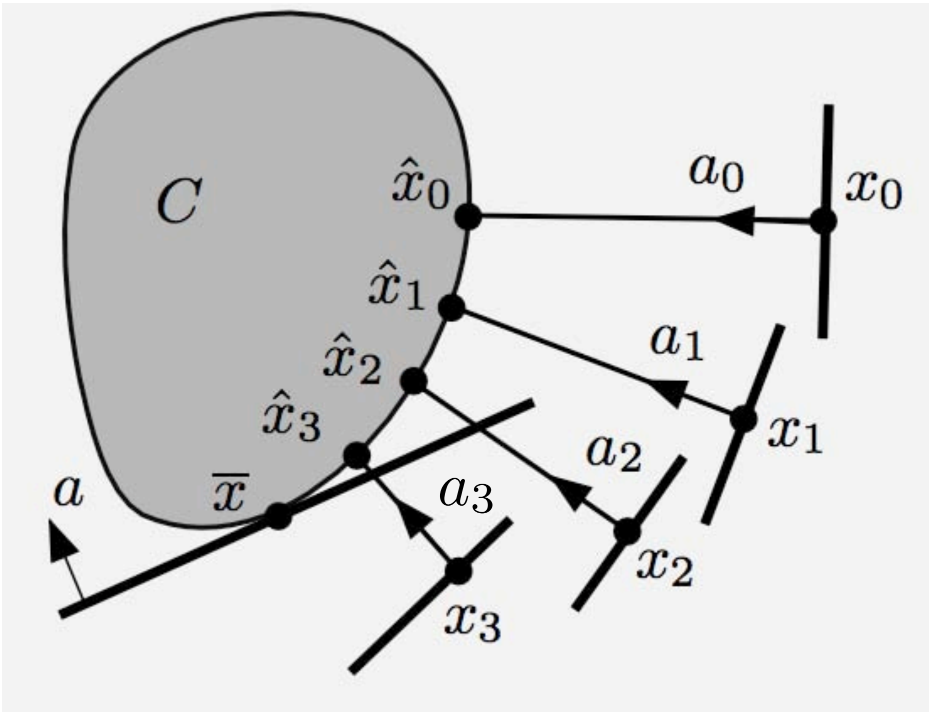
- A separating $\{x \mid a'x = b\}$ that is disjoint from C_1 and C_2 is called *strictly* separating:

$$a'x_1 < b < a'x_2, \qquad \forall\ x_1 \in C_1,\ \forall\ x_2 \in C_2$$



SUPPORTING HYPERPLANE THEOREM

- Let C be convex and let $\overline{x}$ be a vector that is not an interior point of C. Then, there exists a hyperplane that passes through $\overline{x}$ and contains C in one of its closed halfspaces.



Proof: Take a sequence $\{x_k\}$ that does not belong to $\mathrm{cl}(C)$ and converges to $\overline{x}$. Let $\hat{x}_k$ be the projection of x_k on $\mathrm{cl}(C)$. We have for all $x \in \mathrm{cl}(C)$

$$a_k' x \geq a_k' x_k, \qquad \forall\, x \in \mathrm{cl}(C),\ \forall\, k = 0, 1, \ldots,$$

where $a_k = (\hat{x}_k - x_k)/\|\hat{x}_k - x_k\|$. Let a be a limit point of $\{a_k\}$, and take limit as $k \to \infty$. **Q.E.D.**

SEPARATING HYPERPLANE THEOREM

- Let C_1 and C_2 be two nonempty convex subsets of $\Re^n$. If C_1 and C_2 are disjoint, there exists a hyperplane that separates them, i.e., there exists a vector $a \neq 0$ such that

$$a'x_1 \leq a'x_2, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2.$$

Proof: Consider the convex set

$$C_1 - C_2 = \{x_2 - x_1 \mid x_1 \in C_1,\, x_2 \in C_2\}$$

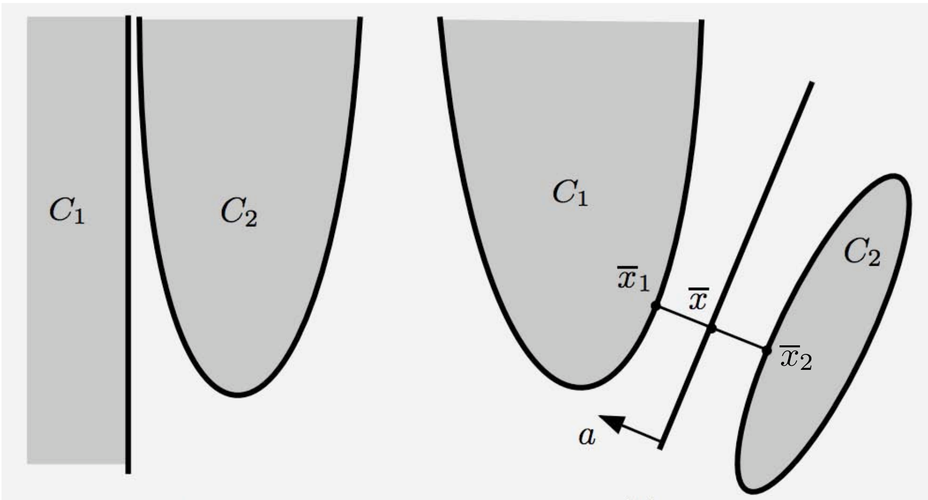
Since C_1 and C_2 are disjoint, the origin does not belong to $C_1 - C_2$, so by the Supporting Hyperplane Theorem, there exists a vector $a \neq 0$ such that

$$0 \leq a'x, \qquad \forall\, x \in C_1 - C_2,$$

which is equivalent to the desired relation. **Q.E.D.**

STRICT SEPARATION THEOREM

- **Strict Separation Theorem**: Let C_1 and C_2 be two disjoint nonempty convex sets. If C_1 is closed, and C_2 is compact, there exists a hyperplane that strictly separates them.



(a) (b)

Proof: (Outline) Consider the set $C_1 - C_2$. Since C_1 is closed and C_2 is compact, $C_1 - C_2$ is closed. Since $C_1 \cap C_2 = \emptyset$, $0 \notin C_1 - C_2$. Let $\overline{x}_1 - \overline{x}_2$ be the projection of 0 onto $C_1 - C_2$. The strictly separating hyperplane is constructed as in (b).

- **Note**: Any conditions that guarantee closedness of $C_1 - C_2$ guarantee existence of a strictly separating hyperplane. However, there may exist a strictly separating hyperplane without $C_1 - C_2$ being closed.

LECTURE 7

LECTURE OUTLINE

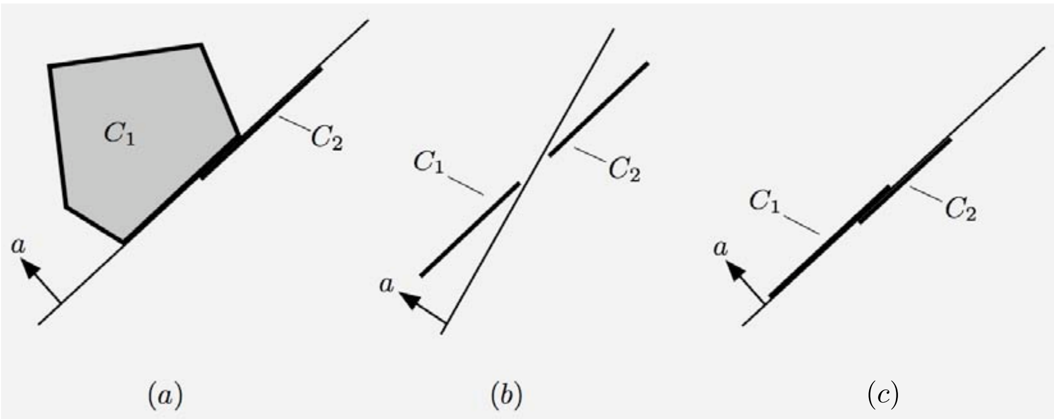
- Review of hyperplane separation
- Nonvertical hyperplanes
- Convex conjugate functions
- Conjugacy theorem
- Examples

Reading: Section 1.5, 1.6

ADDITIONAL THEOREMS

- **Fundamental Characterization**: The closure of the convex hull of a set $C \subset \Re^n$ is the intersection of the closed halfspaces that contain C. (Proof uses the strict separation theorem.)

- We say that a hyperplane *properly separates* C_1 *and* C_2 if it separates C_1 and C_2 and does not fully contain both C_1 and C_2.



- **Proper Separation Theorem**: Let C_1 and C_2 be two nonempty convex subsets of $\Re^n$. There exists a hyperplane that properly separates C_1 and C_2 if and only if

$$\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$$

PROPER POLYHEDRAL SEPARATION

- Recall that two convex sets C and P such that

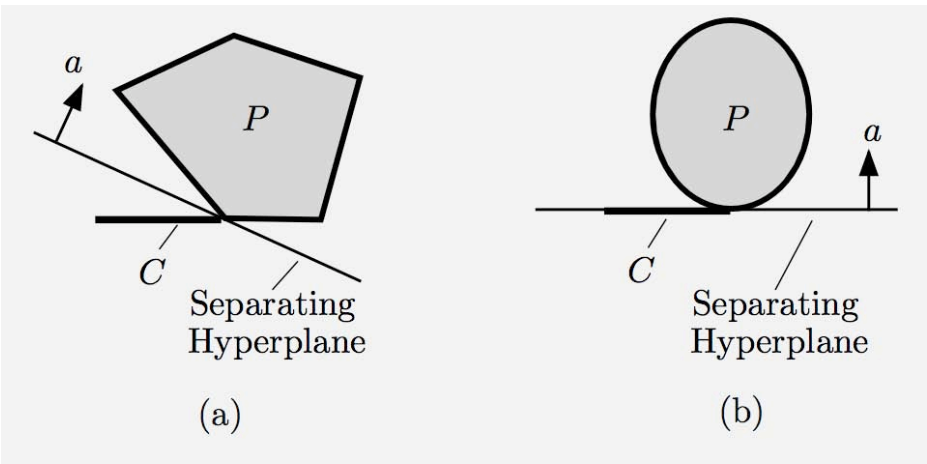
$$\mathrm{ri}(C) \cap \mathrm{ri}(P) = \emptyset$$

can be properly separated, i.e., by a hyperplane that does not contain both C and P.

- If P is polyhedral and the slightly stronger condition

$$\mathrm{ri}(C) \cap P = \emptyset$$

holds, then the properly separating hyperplane can be chosen so that it does not contain the nonpolyhedral set C while it may contain P.

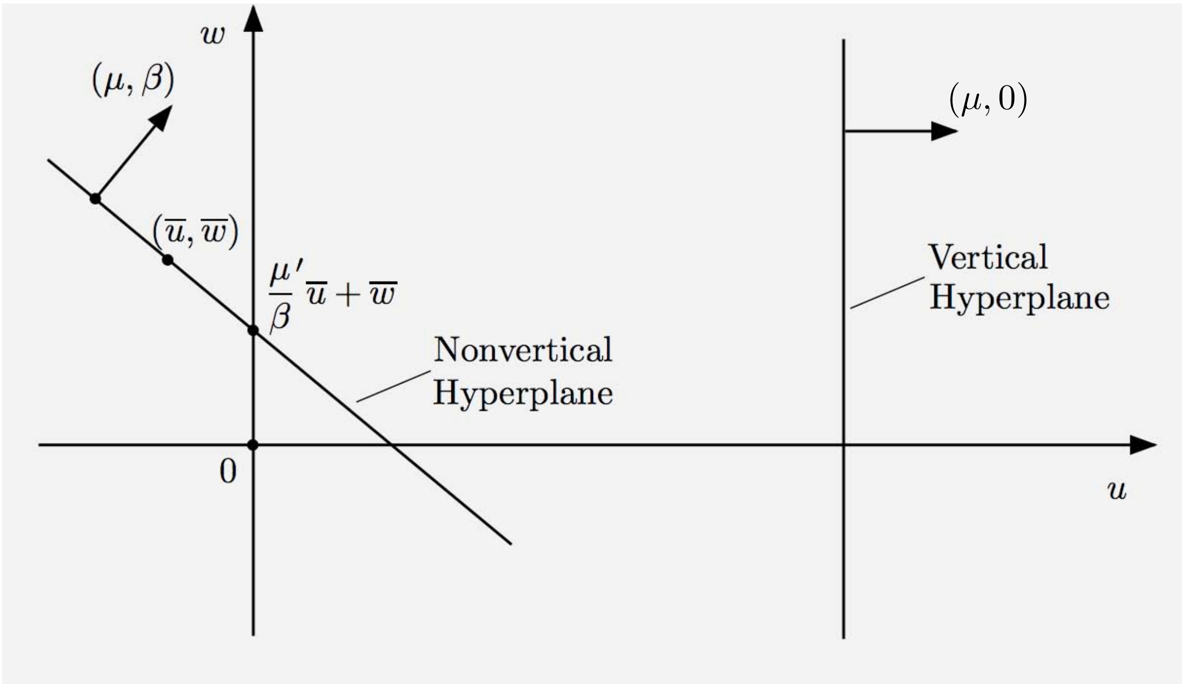


On the left, the separating hyperplane can be chosen so that it does not contain C. On the right where P is not polyhedral, this is not possible.

NONVERTICAL HYPERPLANES

- A hyperplane in $\Re^{n+1}$ with normal (μ, β) is nonvertical if $\beta \neq 0$.

- It intersects the $(n+1)$st axis at $\xi = (\mu/\beta)'\overline{u} + \overline{w}$, where $(\overline{u}, \overline{w})$ is any vector on the hyperplane.



- A nonvertical hyperplane that contains the epigraph of a function in its "upper" halfspace, provides lower bounds to the function values.

- The epigraph of a proper convex function does not contain a vertical line, so it appears plausible that it is contained in the "upper" halfspace of some nonvertical hyperplane.

NONVERTICAL HYPERPLANE THEOREM

- Let C be a nonempty convex subset of $\Re^{n+1}$ that contains no vertical lines. Then:

  (a) C is contained in a closed halfspace of a nonvertical hyperplane, i.e., there exist $\mu \in \Re^n$, $\beta \in \Re$ with $\beta \neq 0$, and $\gamma \in \Re$ such that $\mu' u + \beta w \geq \gamma$ for all $(u, w) \in C$.

  (b) If $(\overline{u}, \overline{w}) \notin \mathrm{cl}(C)$, there exists a nonvertical hyperplane strictly separating $(\overline{u}, \overline{w})$ and C.

Proof: Note that $\mathrm{cl}(C)$ contains no vert. line [since C contains no vert. line, $\mathrm{ri}(C)$ contains no vert. line, and $\mathrm{ri}(C)$ and $\mathrm{cl}(C)$ have the same recession cone]. So we just consider the case: C closed.

(a) C is the intersection of the closed halfspaces containing C. If all these corresponded to vertical hyperplanes, C would contain a vertical line.

(b) There is a hyperplane strictly separating $(\overline{u}, \overline{w})$ and C. If it is nonvertical, we are done, so assume it is vertical. "Add" to this vertical hyperplane a small ϵ-multiple of a nonvertical hyperplane containing C in one of its halfspaces as per (a).

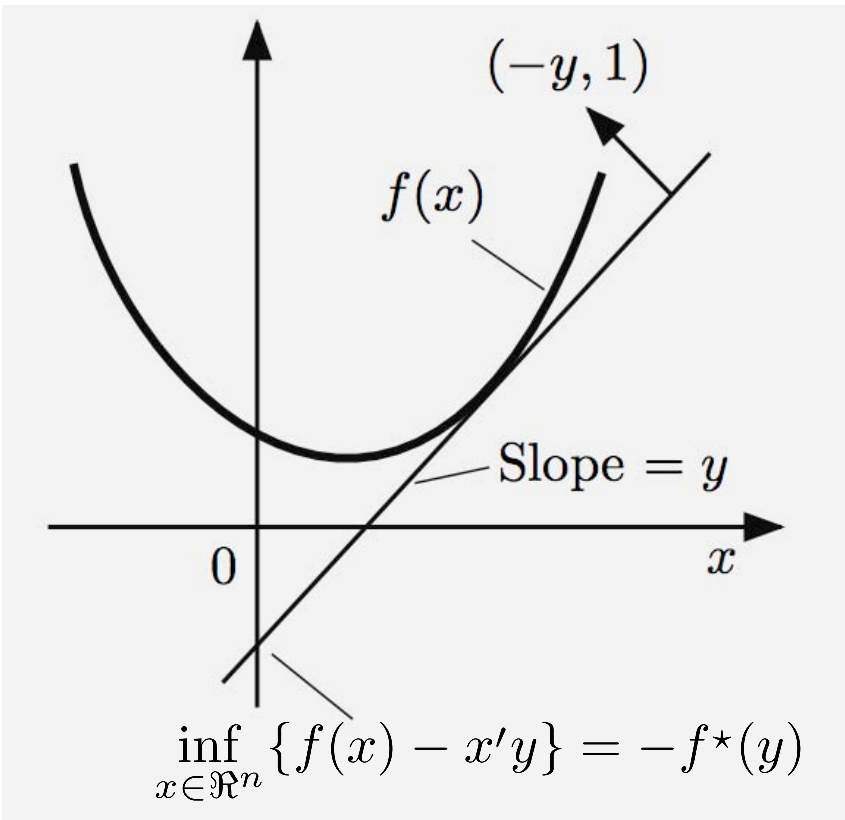
CONJUGATE CONVEX FUNCTIONS

- Consider a function f and its epigraph

Nonvertical hyperplanes supporting epi(f)

$\mapsto$ Crossing points of vertical axis

$$f^\star(y) = \sup_{x \in \Re^n} \big\{ x'y - f(x) \big\}, \qquad y \in \Re^n.$$

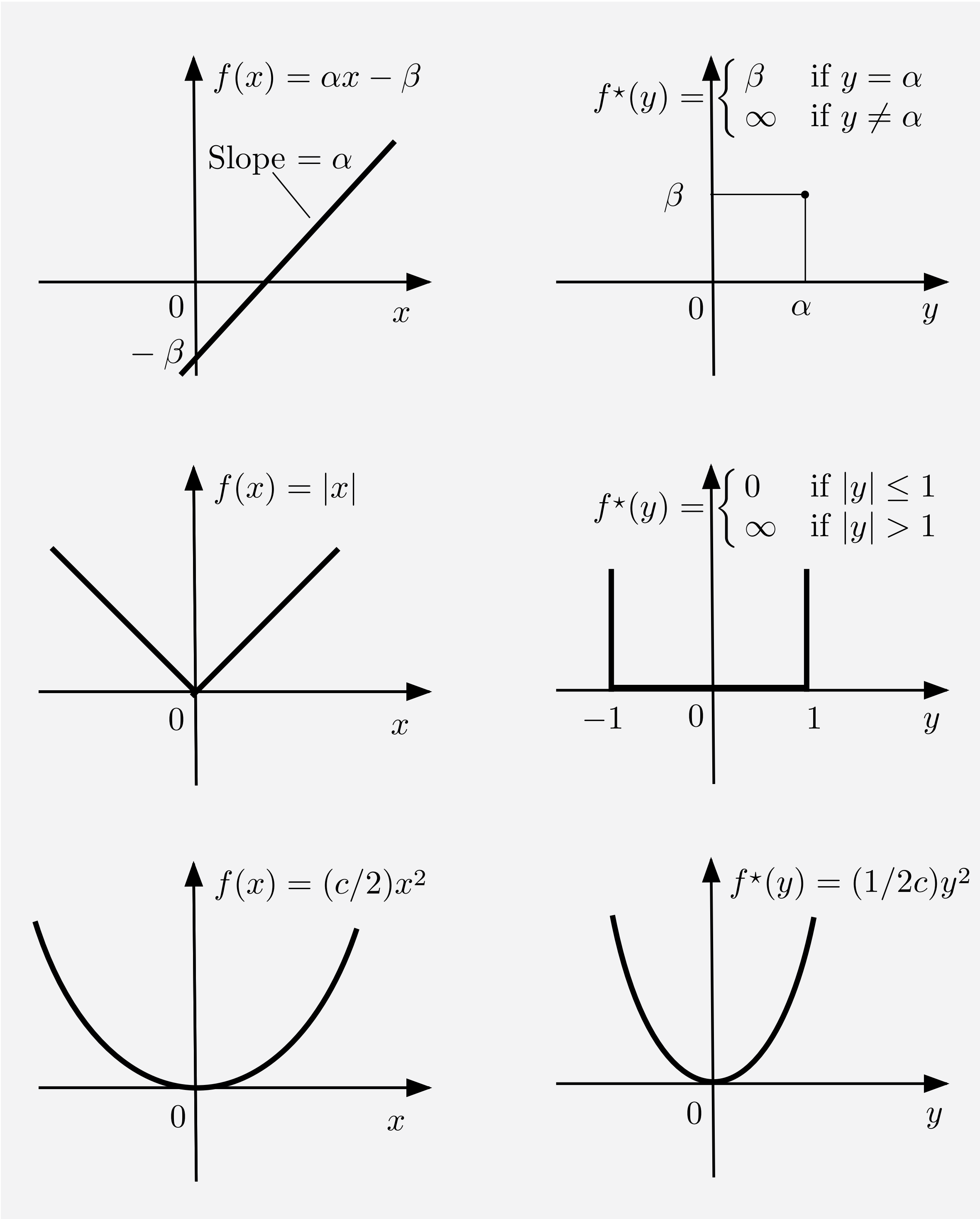


- For any $f : \Re^n \mapsto [-\infty, \infty]$, its *conjugate convex function* is defined by

$$f^\star(y) = \sup_{x \in \Re^n} \big\{ x'y - f(x) \big\}, \qquad y \in \Re^n$$

EXAMPLES

$$f^\star(y) = \sup_{x\in\Re^n} \big\{x'y - f(x)\big\}, \qquad y \in \Re^n$$



CONJUGATE OF CONJUGATE

- From the definition

$$f^\star(y) = \sup_{x \in \Re^n} \big\{x'y - f(x)\big\}, \qquad y \in \Re^n,$$

note that $f^\star$ *is convex and closed*.

- Reason: $\text{epi}(f^\star)$ is the intersection of the epigraphs of the linear functions of y

$$x'y - f(x)$$

as x ranges over $\Re^n$.

- Consider the conjugate of the conjugate:

$$f^{\star\star}(x) = \sup_{y \in \Re^n} \big\{y'x - f^\star(y)\big\}, \qquad x \in \Re^n.$$

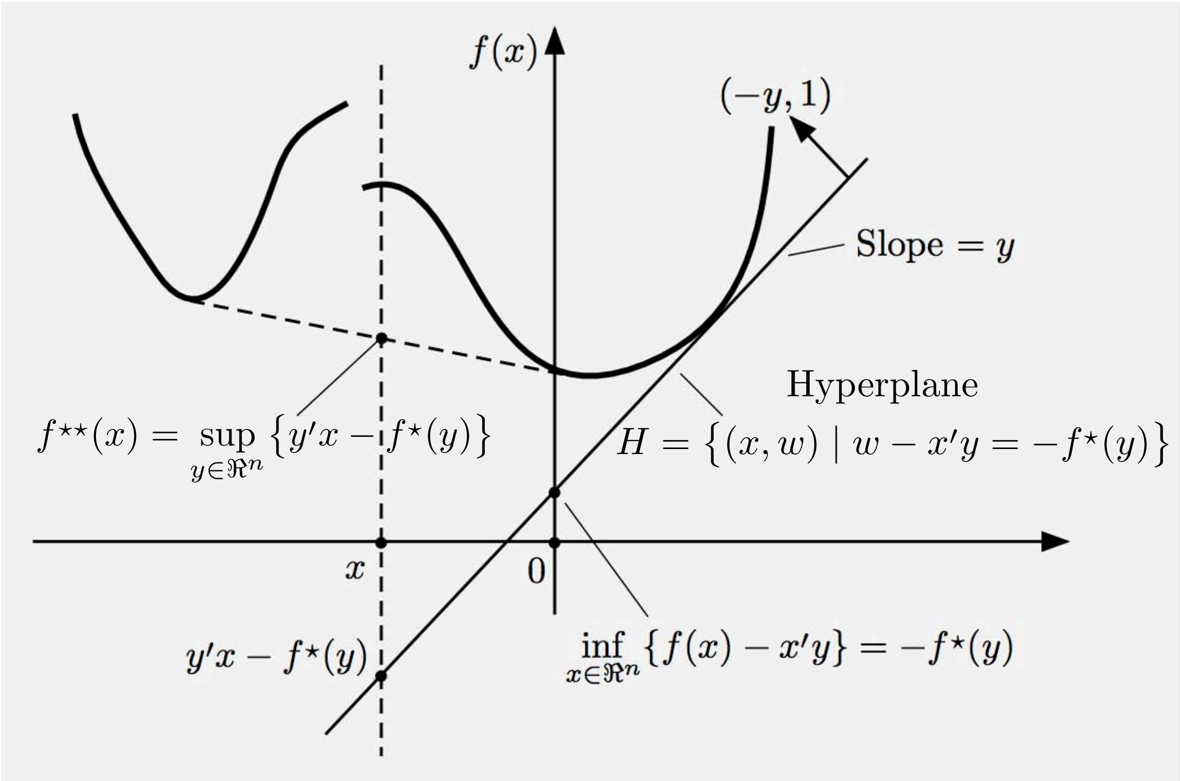
- $f^{\star\star}$ is convex and closed.
- **Important fact/Conjugacy theorem:** If f is closed proper convex, then $f^{\star\star} = f$.

CONJUGACY THEOREM - VISUALIZATION

$$f^\star(y) = \sup_{x\in\Re^n} \big\{x'y - f(x)\big\}, \qquad y \in \Re^n$$

$$f^{\star\star}(x) = \sup_{y\in\Re^n} \big\{y'x - f^\star(y)\big\}, \qquad x \in \Re^n$$

- If f is closed convex proper, then $f^{\star\star} = f$.



CONJUGACY THEOREM

• Let $f : \Re^n \mapsto (-\infty, \infty]$ be a function, let $\check{\mathrm{cl}}\, f$ be its convex closure, let $f^\star$ be its convex conjugate, and consider the conjugate of $f^\star$,

$$f^{\star\star}(x) = \sup_{y \in \Re^n} \big\{ y'x - f^\star(y) \big\}, \qquad x \in \Re^n$$

(a) We have

$$f(x) \geq f^{\star\star}(x), \qquad \forall\, x \in \Re^n$$

(b) If f is convex, then properness of any one of f, $f^\star$, and $f^{\star\star}$ implies properness of the other two.

(c) If f is closed proper and convex, then

$$f(x) = f^{\star\star}(x), \qquad \forall\, x \in \Re^n$$

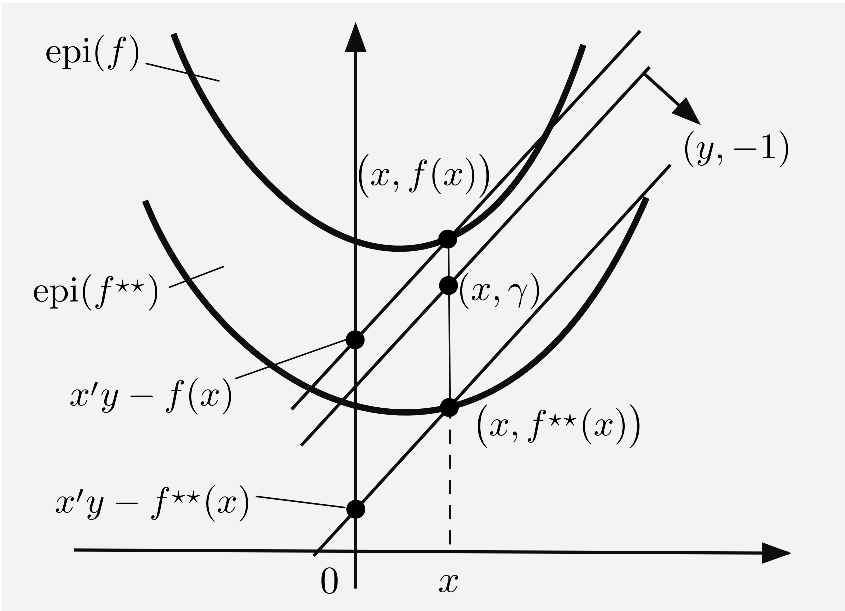
(d) If $\check{\mathrm{cl}}\, f(x) > -\infty$ for all $x \in \Re^n$, then

$$\check{\mathrm{cl}}\, f(x) = f^{\star\star}(x), \qquad \forall\, x \in \Re^n$$

PROOF OF CONJUGACY THEOREM (A), (C)

- (a) For all x, y, we have $f^\star(y) \geq y'x - f(x)$, implying that $f(x) \geq \sup_y\{y'x - f^\star(y)\} = f^{\star\star}(x)$.

- (c) By contradiction. Assume there is $(x, \gamma) \in \text{epi}(f^{\star\star})$ with $(x, \gamma) \notin \text{epi}(f)$. There exists a nonvertical hyperplane with normal $(y, -1)$ that strictly separates (x, γ) and $\text{epi}(f)$. (The vertical component of the normal vector is normalized to -1.)



- Consider two parallel hyperplanes, translated to pass through $\big(x, f(x)\big)$ and $\big(x, f^{\star\star}(x)\big)$. Their vertical crossing points are $x'y - f(x)$ and $x'y - f^{\star\star}(x)$, and lie strictly above and below the crossing point of the strictly sep. hyperplane. Hence

$$x'y - f(x) > x'y - f^{\star\star}(x)$$

the fact $f \geq f^{\star\star}$. **Q.E.D.**

A COUNTEREXAMPLE

- A counterexample (with closed convex but improper f) showing the need to assume properness in order for $f = f^{\star\star}$:

$$f(x) = \begin{cases} \infty & \text{if } x > 0, \\ -\infty & \text{if } x \le 0. \end{cases}$$

We have

$$f^{\star}(y) = \infty, \qquad \forall\, y \in \Re^n,$$

$$f^{\star\star}(x) = -\infty, \qquad \forall\, x \in \Re^n.$$

But

$$\check{\mathrm{cl}}\, f = f,$$

so $\check{\mathrm{cl}}\, f \neq f^{\star\star}$.

LECTURE 8

LECTURE OUTLINE

- Review of conjugate convex functions
- Min common/max crossing duality
- Weak duality
- Special cases

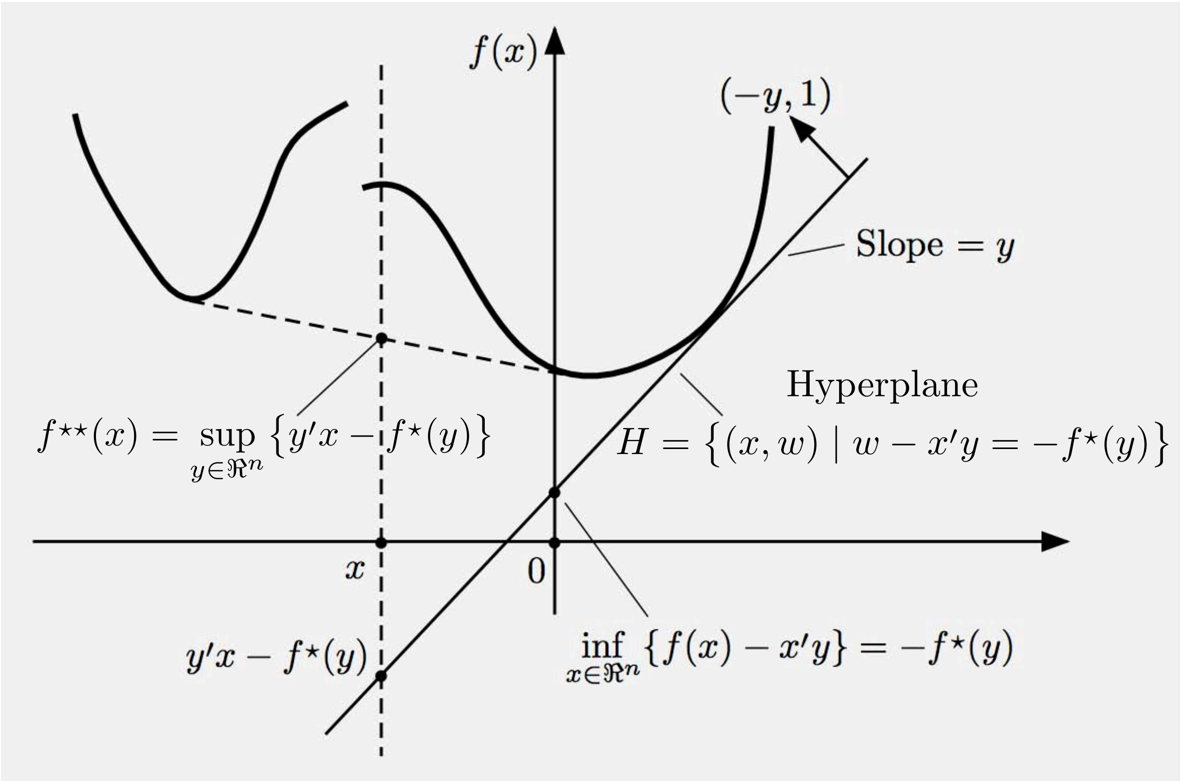
Reading: Sections 1.6, 4.1, 4.2

CONJUGACY THEOREM

$$f^\star(y) = \sup_{x \in \Re^n} \big\{x'y - f(x)\big\}, \qquad y \in \Re^n$$

$$f^{\star\star}(x) = \sup_{y \in \Re^n} \big\{y'x - f^\star(y)\big\}, \qquad x \in \Re^n$$

- If f is closed convex proper, then $f^{\star\star} = f$.



A FEW EXAMPLES

- l_p and l_q norm conjugacy, where $\frac{1}{p} + \frac{1}{q} = 1$

$$f(x) = \frac{1}{p}\sum_{i=1}^{n} |x_i|^p, \qquad f^\star(y) = \frac{1}{q}\sum_{i=1}^{n} |y_i|^q$$

- Conjugate of a strictly convex quadratic

$$f(x) = \frac{1}{2}x'Qx + a'x + b,$$

$$f^\star(y) = \frac{1}{2}(y-a)'Q^{-1}(y-a) - b.$$

- Conjugate of a function obtained by invertible linear transformation/translation of a function p

$$f(x) = p\big(A(x-c)\big) + a'x + b,$$

$$f^\star(y) = q\big((A')^{-1}(y-a)\big) + c'y + d,$$

where q is the conjugate of p and $d = -(c'a + b)$.

SUPPORT FUNCTIONS

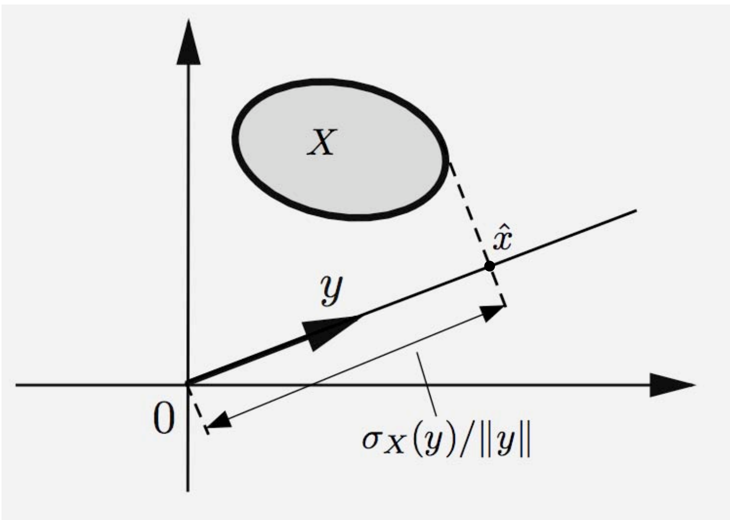
- Conjugate of indicator function δ_X of set X

$$\sigma_X(y) = \sup_{x \in X} y'x$$

is called the *support function of* X.

- To determine $\sigma_X(y)$ for a given vector y, we project the set X on the line determined by y, we find $\hat{x}$, the extreme point of projection in the direction y, and we scale by setting

$$\sigma_X(y) = \|\hat{x}\| \cdot \|y\|$$



- epi(σ_X) is a closed convex cone.

- The sets X, cl(X), conv(X), and cl$\big(\text{conv}(X)\big)$ all have the same support function (by the conjugacy theorem).

SUPPORT FN OF A CONE - POLAR CONE

- The conjugate of the indicator function δ_C is the support function, $\sigma_C(y) = \sup_{x \in C} y'x$.
- If C is a cone,

$$\sigma_C(y) = \begin{cases} 0 & \text{if } y'x \leq 0, \ \forall\ x \in C, \\ \infty & \text{otherwise} \end{cases}$$

i.e., σ_C is the indicator function δ_{C^*} of the cone

$$C^* = \{y \mid y'x \leq 0, \ \forall\ x \in C\}$$

This is called the *polar cone of* C.

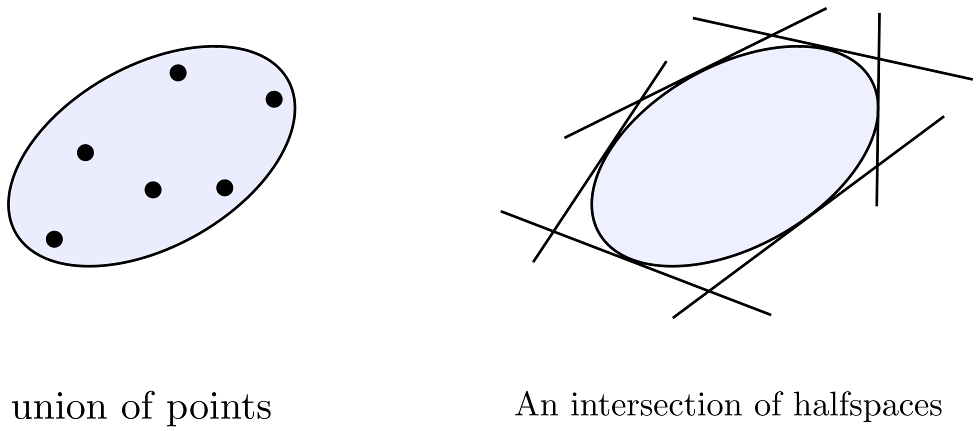
- By the Conjugacy Theorem the polar cone of C^* is $\text{cl}\big(\text{conv}(C)\big)$. This is the *Polar Cone Theorem.*
- **Special case:** If $C = \text{cone}\big(\{a_1, \ldots, a_r\}\big)$, then

$$C^* = \{x \mid a_j'x \leq 0, \ j = 1, \ldots, r\}$$

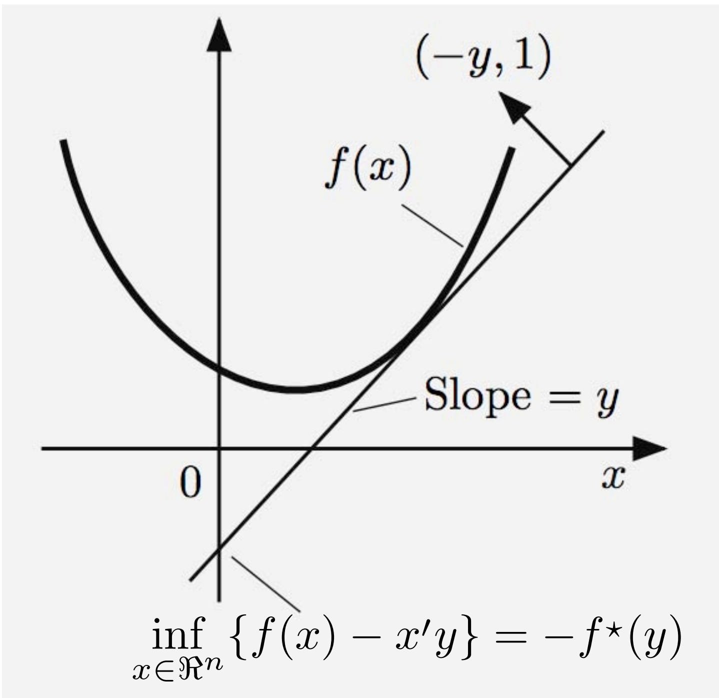
- **Farkas' Lemma:** $(C^*)^* = C$.
- True because C is a closed set $[\text{cone}\big(\{a_1, \ldots, a_r\}\big)$ is the image of the positive orthant $\{\alpha \mid \alpha \geq 0\}$ under the linear transformation that maps α to $\sum_{j=1}^r \alpha_j a_j]$, and the image of any polyhedral set under a linear transformation is a closed set.

EXTENDING DUALITY CONCEPTS

- From **dual descriptions of sets**



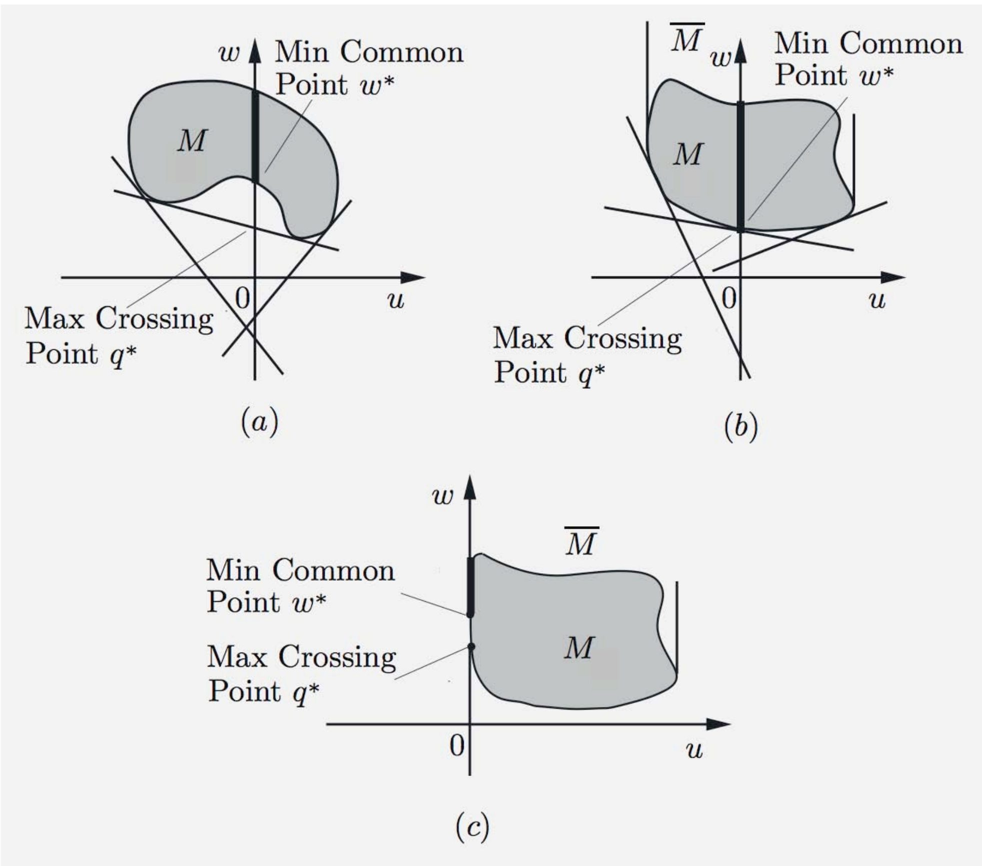
- To **dual descriptions of functions** (applying set duality to epigraphs)



- We now go to **dual descriptions of problems**, by applying conjugacy constructions to a simple generic geometric optimization problem

MIN COMMON / MAX CROSSING PROBLEMS

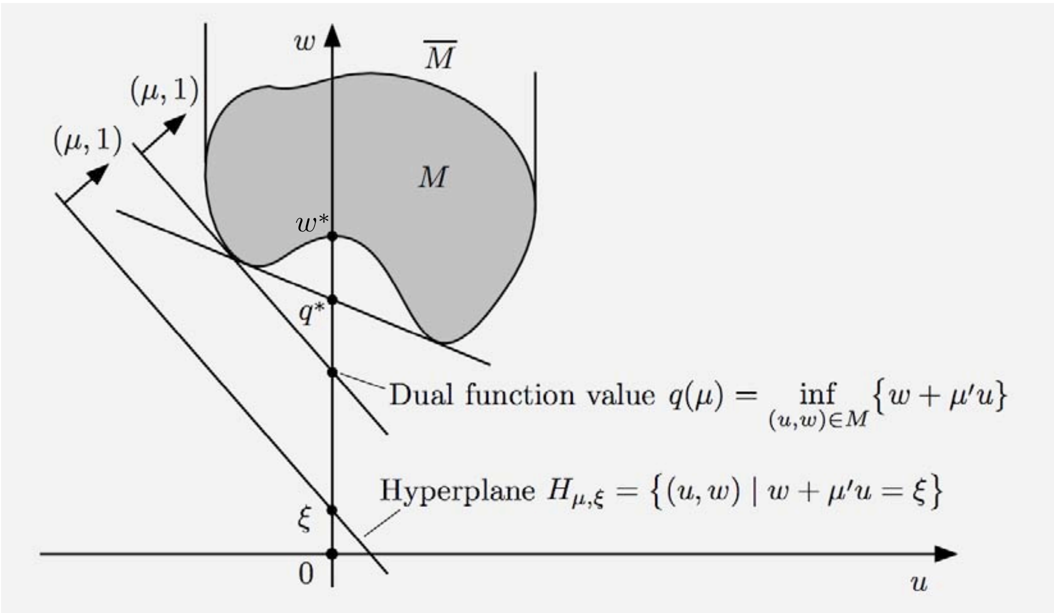
- We introduce a pair of fundamental problems:
- Let M be a nonempty subset of $\Re^{n+1}$
  (a) *Min Common Point Problem*: Consider all vectors that are common to M and the $(n+1)$st axis. Find one whose $(n+1)$st component is minimum.
  (b) *Max Crossing Point Problem*: Consider non-vertical hyperplanes that contain M in their "upper" closed halfspace. Find one whose crossing point of the $(n+1)$st axis is maximum.



MATHEMATICAL FORMULATIONS

- **Optimal value of min common problem:**

$$w^* = \inf_{(0,w)\in M} w$$



- **Math formulation of max crossing problem:** Focus on hyperplanes with normals $(\mu, 1)$ whose crossing point ξ satisfies

$$\xi \le w + \mu' u, \qquad \forall\ (u,w) \in M$$

Max crossing problem is to maximize ξ subject to $\xi \le \inf_{(u,w)\in M}\{w + \mu' u\}$, $\mu \in \Re^n$, or

$$\text{maximize} \quad q(\mu) \stackrel{\triangle}{=} \inf_{(u,w)\in M} \{w + \mu' u\}$$

$$\text{subject to} \quad \mu \in \Re^n.$$

GENERIC PROPERTIES – WEAK DUALITY

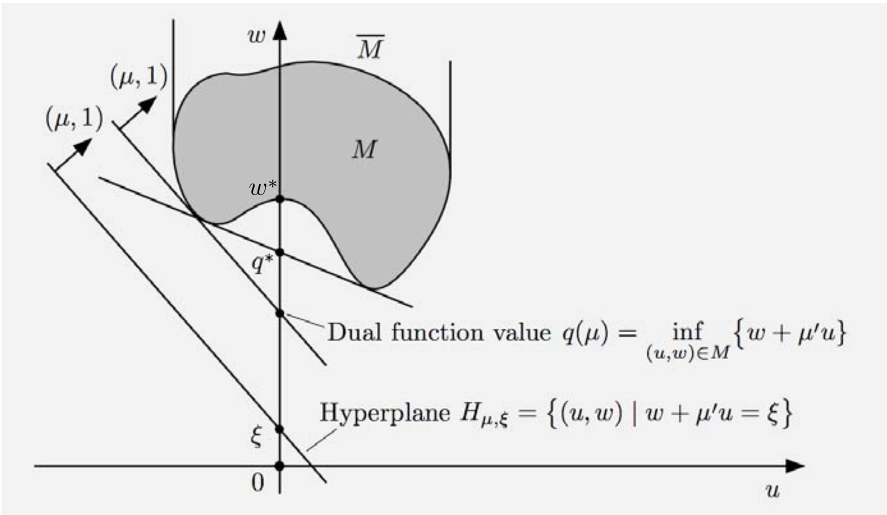
- Min common problem

$$\inf_{(0,w)\in M} w$$

- Max crossing problem

$$\text{maximize} \quad q(\mu) \stackrel{\triangle}{=} \inf_{(u,w)\in M} \{w + \mu' u\}$$

$$\text{subject to} \quad \mu \in \Re^n.$$



- Note that q is concave and upper-semicontinuous (inf of linear functions).

- **Weak Duality:** For all $\mu \in \Re^n$

$$q(\mu) = \inf_{(u,w)\in M} \{w + \mu' u\} \le \inf_{(0,w)\in M} w = w^*,$$

so maximizing over $\mu \in \Re^n$, we obtain $q^* \le w^*$.

- We say that **strong duality** holds if $q^* = w^*$.

CONNECTION TO CONJUGACY

- **An important special case:**

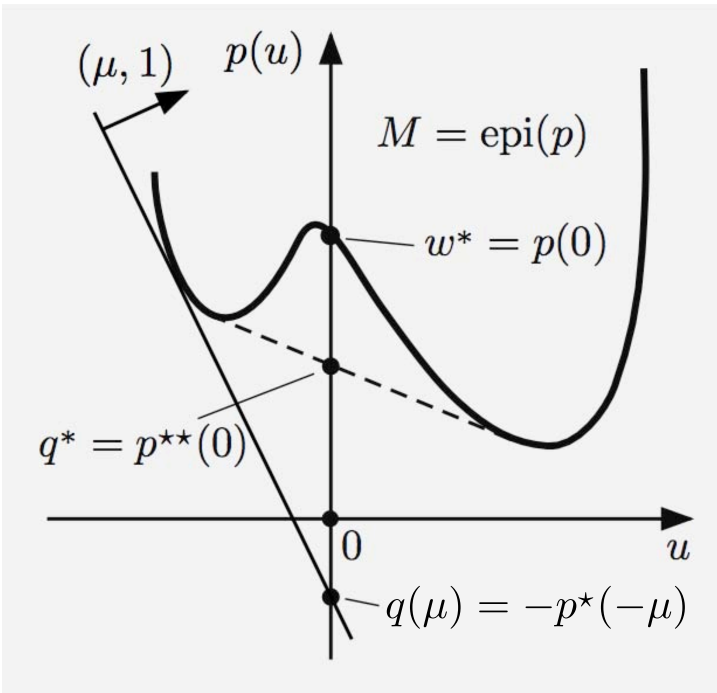
$$M = \text{epi}(p)$$

where $p : \Re^n \mapsto [-\infty, \infty]$. Then $w^* = p(0)$, and

$$q(\mu) = \inf_{(u,w)\in \text{epi}(p)} \{w + \mu' u\} = \inf_{\{(u,w)\mid p(u)\le w\}} \{w + \mu' u\},$$

and finally

$$q(\mu) = \inf_{u\in\Re^m} \big\{p(u) + \mu' u\big\}$$



- Thus, $q(\mu) = -p^\star(-\mu)$ and

$$q^* = \sup_{\mu\in\Re^n} q(\mu) = \sup_{\mu\in\Re^n} \big\{0\cdot(-\mu) - p^\star(-\mu)\big\} = p^{\star\star}(0)$$

GENERAL OPTIMIZATION DUALITY

- Consider minimizing a function $f : \Re^n \mapsto [-\infty, \infty]$.
- Let $F : \Re^{n+r} \mapsto [-\infty, \infty]$ be a function with

$$f(x) = F(x, 0), \qquad \forall\, x \in \Re^n$$

- Consider the *perturbation function*

$$p(u) = \inf_{x \in \Re^n} F(x, u)$$

and the MC/MC framework with $M = \text{epi}(p)$

- The min common value w^* is

$$w^* = p(0) = \inf_{x \in \Re^n} F(x, 0) = \inf_{x \in \Re^n} f(x)$$

- The dual function is

$$q(\mu) = \inf_{u \in \Re^r} \big\{p(u) + \mu' u\big\} = \inf_{(x,u) \in \Re^{n+r}} \big\{F(x, u) + \mu' u\big\}$$

so $q(\mu) = -F^\star(0, -\mu)$, where $F^\star$ is the conjugate of F, viewed as a function of (x, u)

- We have

$$q^* = \sup_{\mu \in \Re^r} q(\mu) = -\inf_{\mu \in \Re^r} F^\star(0, -\mu) = -\inf_{\mu \in \Re^r} F^\star(0, \mu),$$

and weak duality has the form

$$w^* = \inf_{x \in \Re^n} F(x, 0) \geq -\inf_{\mu \in \Re^r} F^\star(0, \mu) = q^*$$

CONSTRAINED OPTIMIZATION

- Minimize $f : \Re^n \mapsto \Re$ over the set

$$C = \{x \in X \mid g(x) \leq 0\},$$

where $X \subset \Re^n$ and $g : \Re^n \mapsto \Re^r$.

- Introduce a "perturbed constraint set"

$$C_u = \{x \in X \mid g(x) \leq u\}, \qquad u \in \Re^r,$$

and the function

$$F(x,u) = \begin{cases} f(x) & \text{if } x \in C_u, \\ \infty & \text{otherwise,} \end{cases}$$

which satisfies $F(x,0) = f(x)$ for all $x \in C$.

- Consider *perturbation function*

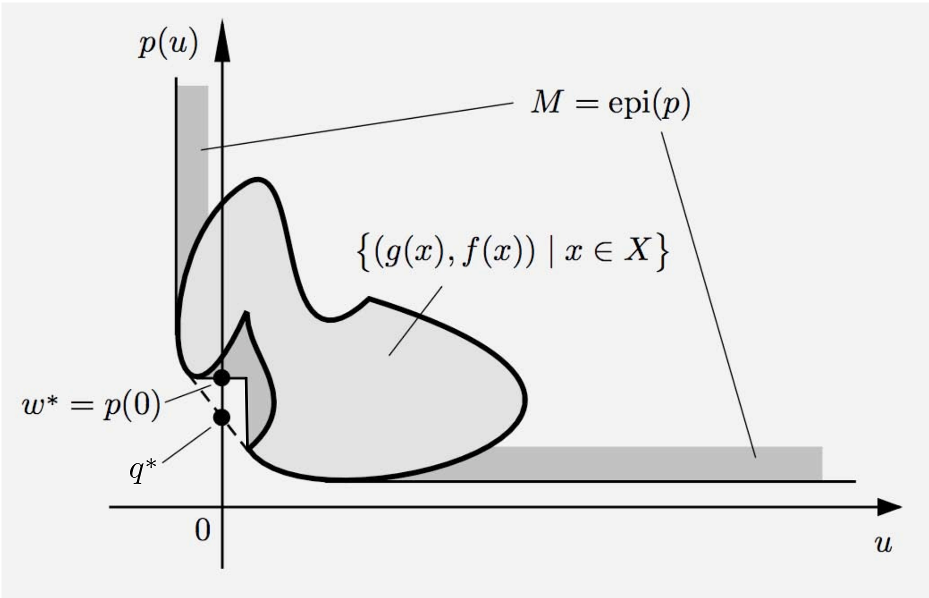
$$p(u) = \inf_{x \in \Re^n} F(x,u) = \inf_{x \in X,\, g(x) \leq u} f(x),$$

and the MC/MC framework with $M = \text{epi}(p)$.

CONSTR. OPT. - PRIMAL AND DUAL FNS

- 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

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

LINEAR PROGRAMMING DUALITY

- Consider the linear program

$$\begin{aligned} &\text{minimize } \; c'x \\ &\text{subject to } \; a_j'x \geq b_j, \quad j = 1, \ldots, r, \end{aligned}$$

where $c \in \Re^n$, $a_j \in \Re^n$, and $b_j \in \Re$, $j = 1, \ldots, r$.

- For $\mu \geq 0$, the dual function has the form

$$\begin{aligned} q(\mu) &= \inf_{x \in \Re^n} L(x, \mu) \\ &= \inf_{x \in \Re^n} \left\{ c'x + \sum_{j=1}^{r} \mu_j (b_j - a_j'x) \right\} \\ &= \begin{cases} b'\mu & \text{if } \sum_{j=1}^{r} a_j \mu_j = c, \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

- Thus the dual problem is

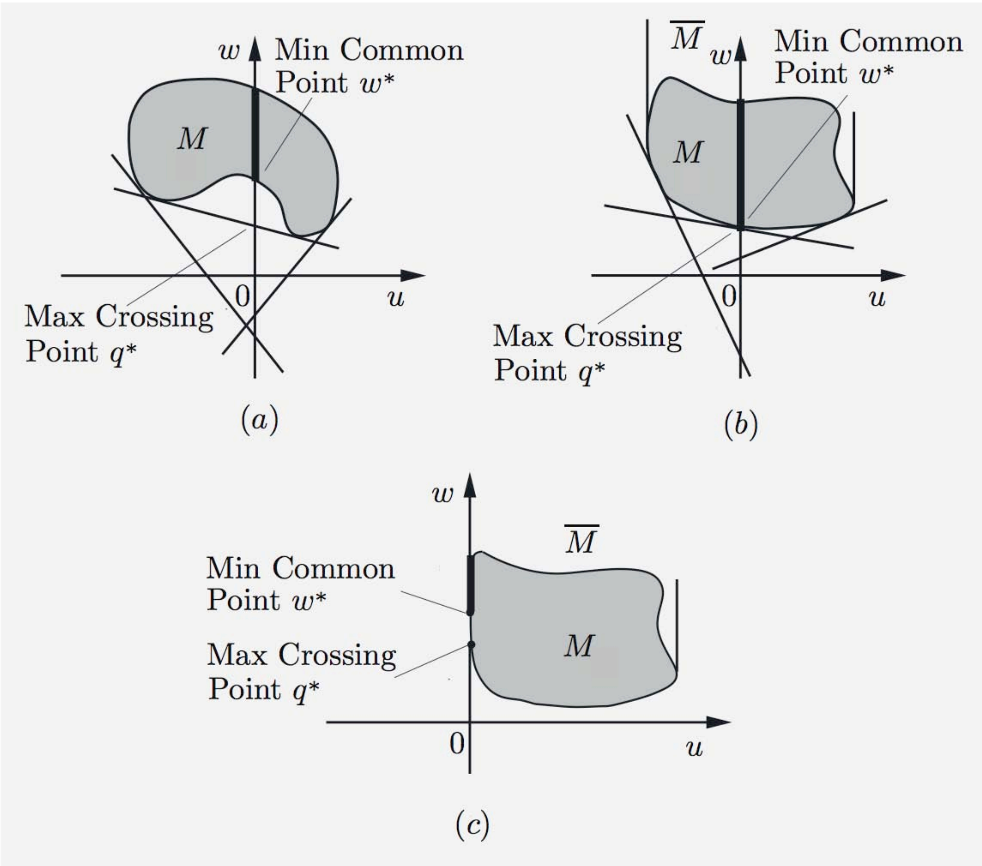
$$\begin{aligned} &\text{maximize } \; b'\mu \\ &\text{subject to } \; \sum_{j=1}^{r} a_j \mu_j = c, \quad \mu \geq 0. \end{aligned}$$

LECTURE 9

LECTURE OUTLINE

- Minimax problems and zero-sum games
- Min Common / Max Crossing duality for minimax and zero-sum games
- Min Common / Max Crossing duality theorems
- Strong duality conditions
- Existence of dual optimal solutions

Reading: Sections 3.4, 4.3, 4.4, 5.1



REVIEW OF THE MC/MC FRAMEWORK

- Given set $M \subset \Re^{n+1}$,

$$w^* = \inf_{(0,w)\in M} w, \quad q^* = \sup_{\mu\in\Re^n} q(\mu) \stackrel{\triangle}{=} \inf_{(u,w)\in M} \{w+\mu' u\}$$

- **Weak Duality:** $q^* \leq w^*$
- **Important special case:** $M = \text{epi}(p)$. Then $w^* = p(0)$, $q^* = p^{\star\star}(0)$, so we have $w^* = q^*$ if p is closed, proper, convex.
- Some applications:
  - Constrained optimization: $\min_{x\in X,\, g(x)\leq 0} f(x)$, with $p(u) = \inf_{x\in X,\, g(x)\leq u} f(x)$
  - Other optimization problems: Fenchel and conic optimization
  - Useful theorems related to optimization: Farkas' lemma, theorems of the alternative
  - Subgradient theory
  - Minimax problems, 0-sum games
- **Strong Duality:** $q^* = w^*$. Requires that M have some convexity structure, among other conditions

MINIMAX PROBLEMS

Given $\phi : X \times Z \mapsto \Re$, where $X \subset \Re^n$, $Z \subset \Re^m$ consider

$$\begin{array}{ll} \text{minimize} & \sup_{z \in Z} \phi(x,z) \\ \text{subject to} & x \in X \end{array}$$

or

$$\begin{array}{ll} \text{maximize} & \inf_{x \in X} \phi(x,z) \\ \text{subject to} & z \in Z. \end{array}$$

- Some important contexts:
  - Constrained optimization duality theory
  - Zero sum game theory
- We always have

$$\sup_{z \in Z} \inf_{x \in X} \phi(x,z) \le \inf_{x \in X} \sup_{z \in Z} \phi(x,z)$$

- **Key question:** When does equality hold?

CONSTRAINED OPTIMIZATION DUALITY

- For the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \qquad g(x) \le 0 \end{array}$$

introduce the Lagrangian function

$$L(x, \mu) = f(x) + \mu' g(x)$$

- Primal problem (equivalent to the original)

$$\min_{x \in X} \sup_{\mu \ge 0} L(x, \mu) = \begin{cases} f(x) & \text{if } g(x) \le 0, \\ \infty & \text{otherwise,} \end{cases}$$

- Dual problem

$$\max_{\mu \ge 0} \inf_{x \in X} L(x, \mu)$$

- Key duality question: Is it true that

$$\inf_{x \in \Re^n} \sup_{\mu \ge 0} L(x, \mu) = w^* \overset{?}{=} q^* = \sup_{\mu \ge 0} \inf_{x \in \Re^n} L(x, \mu)$$

ZERO SUM GAMES

- Two players: 1st chooses $i \in \{1, \ldots, n\}$, 2nd chooses $j \in \{1, \ldots, m\}$.
- If i and j are selected, the 1st player gives a_{ij} to the 2nd.
- Mixed strategies are allowed: The two players select probability distributions

$$x = (x_1, \ldots, x_n), \qquad z = (z_1, \ldots, z_m)$$

over their possible choices.

- Probability of (i, j) is $x_i z_j$, so the expected amount to be paid by the 1st player

$$x'Az = \sum_{i,j} a_{ij} x_i z_j$$

where A is the $n \times m$ matrix with elements a_{ij}.

- Each player optimizes his choice against the worst possible selection by the other player. So
  - 1st player minimizes $\max_z x'Az$
  - 2nd player maximizes $\min_x x'Az$

SADDLE POINTS

Definition: (x^*, z^*) is called a *saddle point* of ϕ if

$$\phi(x^*, z) \leq \phi(x^*, z^*) \leq \phi(x, z^*), \quad \forall\, x \in X, \forall\, z \in Z$$

Proposition: (x^*, z^*) is a saddle point if and only if the minimax equality holds and

$$x^* \in \arg\min_{x \in X} \sup_{z \in Z} \phi(x, z), \quad z^* \in \arg\max_{z \in Z} \inf_{x \in X} \phi(x, z) \quad (*)$$

Proof: If (x^*, z^*) is a saddle point, then

$$\begin{aligned} \inf_{x \in X} \sup_{z \in Z} \phi(x, z) &\leq \sup_{z \in Z} \phi(x^*, z) = \phi(x^*, z^*) \\ &= \inf_{x \in X} \phi(x, z^*) \leq \sup_{z \in Z} \inf_{x \in X} \phi(x, z) \end{aligned}$$

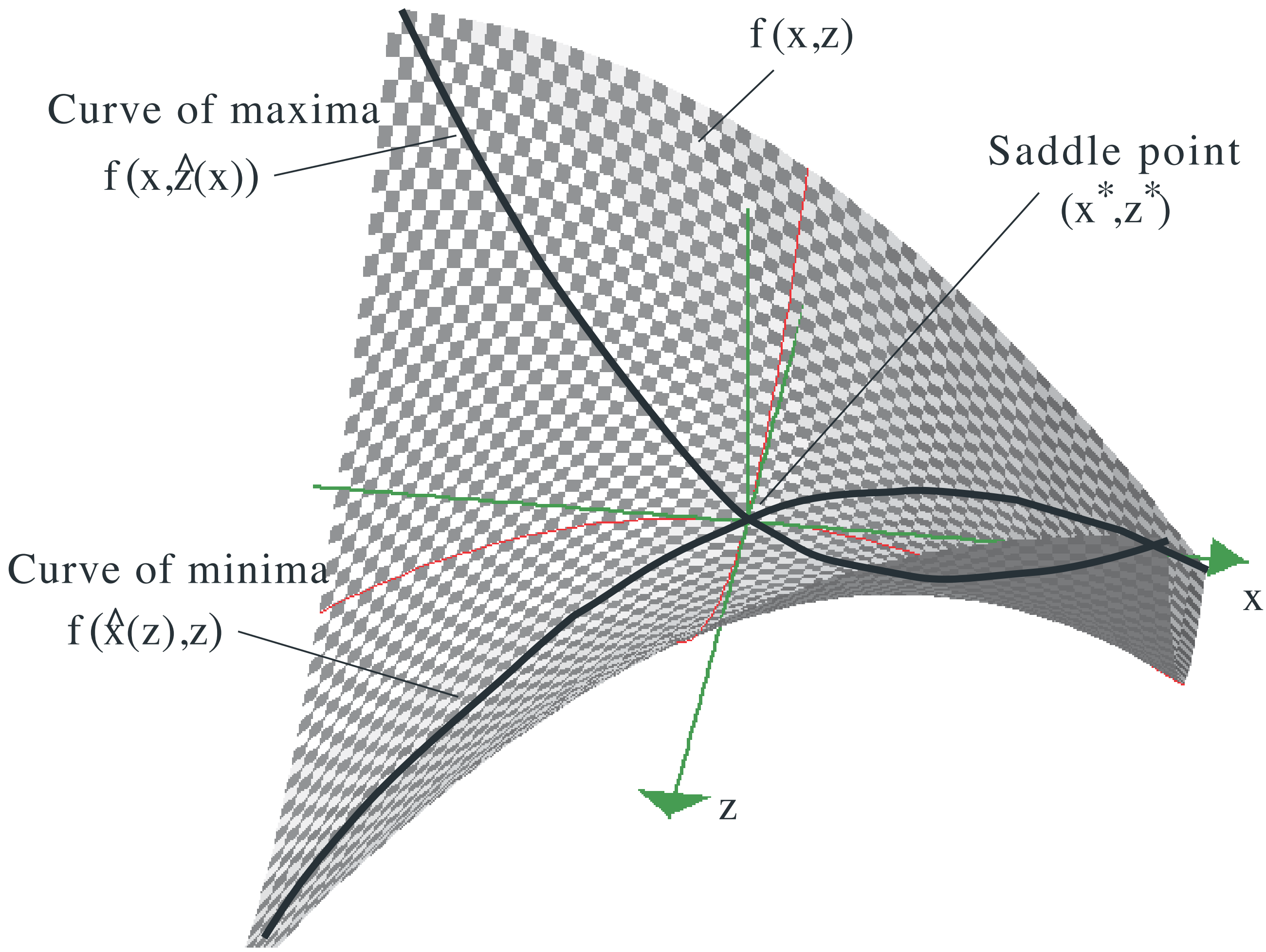
By the minimax inequality, the above holds as an equality throughout, so the minimax equality and Eq. (*) hold.

Conversely, if Eq. (*) holds, then

$$\begin{aligned} \sup_{z \in Z} \inf_{x \in X} \phi(x, z) &= \inf_{x \in X} \phi(x, z^*) \leq \phi(x^*, z^*) \\ &\leq \sup_{z \in Z} \phi(x^*, z) = \inf_{x \in X} \sup_{z \in Z} \phi(x, z) \end{aligned}$$

Using the minimax equ., (x^*, z^*) is a saddle point.

VISUALIZATION



The curve of maxima $f(x, \hat{z}(x))$ lies above the curve of minima $f(\hat{x}(z), z)$, where

$$\hat{z}(x) = \arg\max_z f(x, z), \qquad \hat{x}(z) = \arg\min_x f(x, z)$$

Saddle points correspond to points where these two curves meet.

MINIMAX MC/MC FRAMEWORK

- Introduce perturbation function $p : \Re^m \mapsto [-\infty, \infty]$

$$p(u) = \inf_{x \in X} \sup_{z \in Z} \big\{ \phi(x, z) - u'z \big\}, \qquad u \in \Re^m$$

- Apply the MC/MC framework with $M = \text{epi}(p)$. If p is convex, closed, and proper, no duality gap.
- Introduce $\hat{\text{cl}}\,\phi$, the *concave closure of* ϕ viewed as a function of z for fixed x
- We have

$$\sup_{z \in Z} \phi(x, z) = \sup_{z \in \Re^m} (\hat{\text{cl}}\,\phi)(x, z),$$

so

$$w^* = p(0) = \inf_{x \in X} \sup_{z \in \Re^m} (\hat{\text{cl}}\,\phi)(x, z).$$

- The dual function can be shown to be

$$q(\mu) = \inf_{x \in X} (\hat{\text{cl}}\,\phi)(x, \mu), \qquad \forall\, \mu \in \Re^m$$

so if $\phi(x, \cdot)$ is concave and closed,

$$w^* = \inf_{x \in X} \sup_{z \in \Re^m} \phi(x, z), \qquad q^* = \sup_{z \in \Re^m} \inf_{x \in X} \phi(x, z)$$

PROOF OF FORM OF DUAL FUNCTION

- Write $p(u) = \inf_{x\in X} p_x(u)$, where

$$p_x(u) = \sup_{z\in Z}\big\{\phi(x,z) - u'z\big\}, \qquad x \in X,$$

and note that

$$\inf_{u\in\Re^m}\big\{p_x(u)+u'\mu\big\} = -\sup_{u\in\Re^m}\big\{u'(-\mu)-p_x(u)\big\} = -p_x^\star(-\mu)$$

Except for a sign change, p_x is the conjugate of $(-\phi)(x,\cdot)$ [assuming $(-\hat{\text{cl}}\,\phi)(x,\cdot)$ is proper], so

$$p_x^\star(-\mu) = -(\hat{\text{cl}}\,\phi)(x,\mu).$$

Hence, for all $\mu \in \Re^m$,

$$\begin{aligned} q(\mu) &= \inf_{u\in\Re^m}\big\{p(u) + u'\mu\big\} \\ &= \inf_{u\in\Re^m}\inf_{x\in X}\big\{p_x(u) + u'\mu\big\} \\ &= \inf_{x\in X}\inf_{u\in\Re^m}\big\{p_x(u) + u'\mu\big\} \\ &= \inf_{x\in X}\big\{-p_x^\star(-\mu)\big\} \\ &= \inf_{x\in X}(\hat{\text{cl}}\,\phi)(x,\mu) \end{aligned}$$

DUALITY THEOREMS

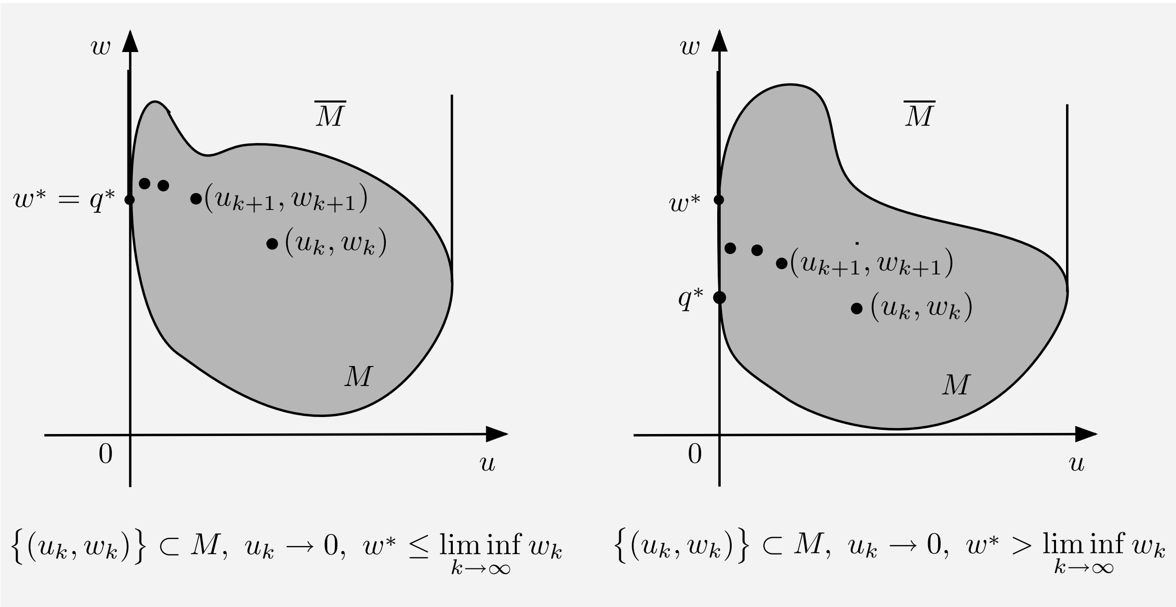
- Assume that $w^* < \infty$ and that the set

$$\overline{M} = \Big\{(u, w) \mid \text{there exists } \overline{w} \text{ with } \overline{w} \le w \text{ and } (u, \overline{w}) \in M\Big\}$$

is convex.

- **Min Common/Max Crossing 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

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



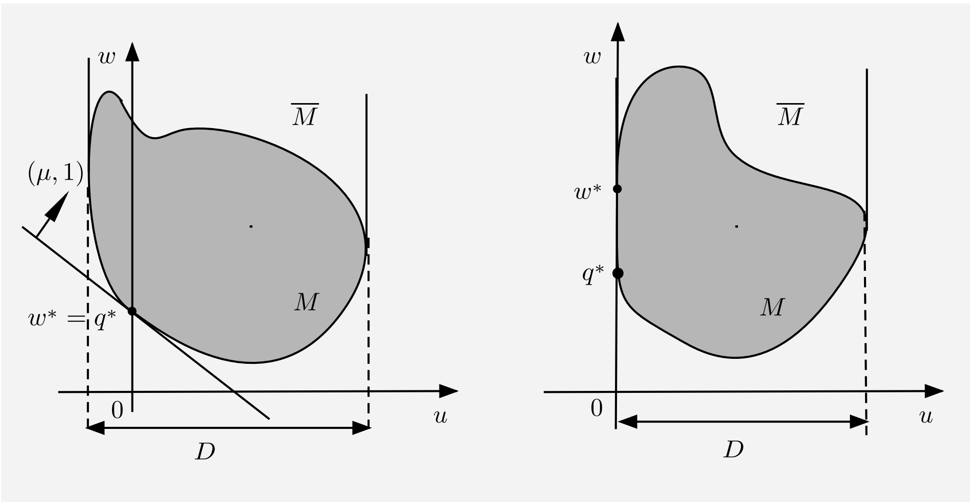
- **Corollary:** If $M = \text{epi}(p)$ where p is closed proper convex and $p(0) < \infty$, then $q^* = w^*$.

DUALITY THEOREMS (CONTINUED)

- **Min Common/Max Crossing Theorem II:** Assume in addition that $-\infty < w^*$ and that

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

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



- Furthermore, the set $\{\mu \mid q(\mu) = q^*\}$ is nonempty and compact if and only if D contains the origin in its interior.

- **Min Common/Max Crossing Theorem III:** Involves polyhedral assumptions, and will be developed later.

PROOF OF THEOREM I

- Assume that $q^* = w^*$. Let $\{(u_k, w_k)\} \subset M$ be such that $u_k \to 0$. Then,

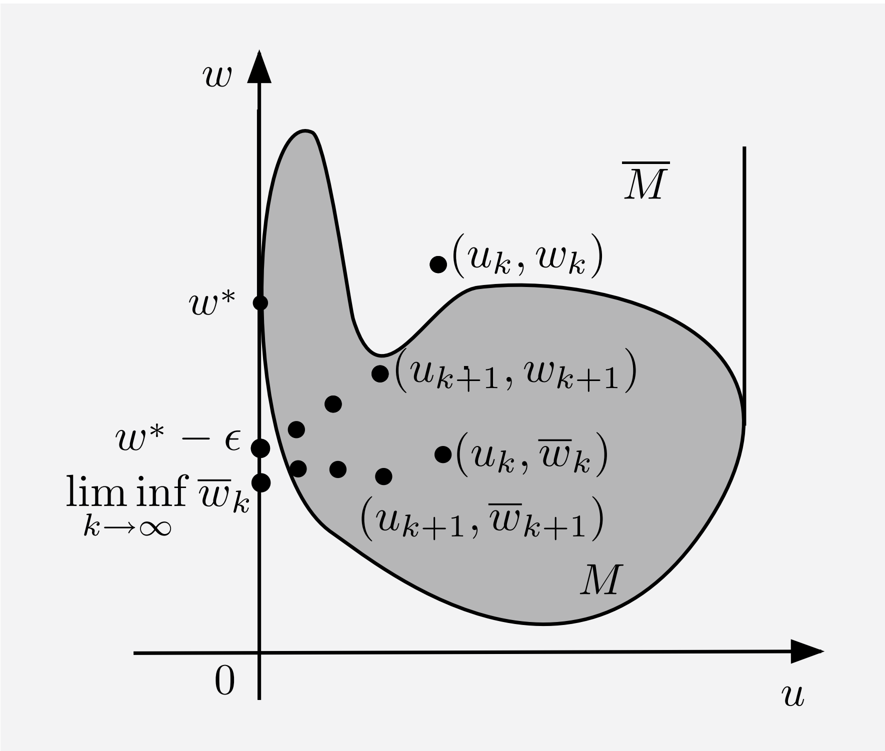
$$q(\mu) = \inf_{(u,w)\in M} \{w + \mu' u\} \le w_k + \mu' u_k, \quad \forall\, k,\ \forall\, \mu \in \Re^n$$

Taking the limit as $k \to \infty$, we obtain $q(\mu) \le \liminf_{k\to\infty} w_k$, for all $\mu \in \Re^n$, implying that

$$w^* = q^* = \sup_{\mu\in\Re^n} q(\mu) \le \liminf_{k\to\infty} w_k$$

Conversely, assume that for every sequence $\{(u_k, w_k)\} \subset M$ with $u_k \to 0$, there holds $w^* \le \liminf_{k\to\infty} w_k$. If $w^* = -\infty$, then $q^* = -\infty$, by weak duality, so assume that $-\infty < w^*$. Steps:

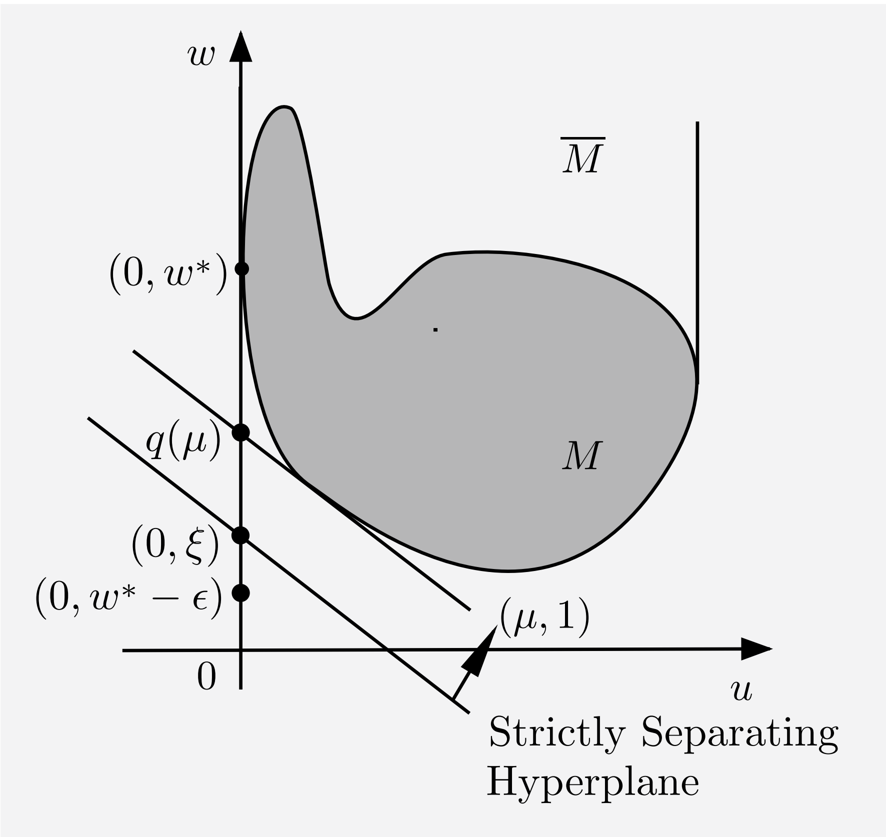
- **Step 1:** $(0, w^* - \epsilon) \notin \mathrm{cl}(\overline{M})$ for any $\epsilon > 0$.



PROOF OF THEOREM I (CONTINUED)

- **Step 2:** $\overline{M}$ does not contain any vertical lines. If this were not so, $(0,-1)$ would be a direction of recession of $\text{cl}(\overline{M})$. Because $(0,w^*) \in \text{cl}(\overline{M})$, the entire halfline $\big\{(0,w^*-\epsilon) \mid \epsilon \geq 0\big\}$ belongs to $\text{cl}(\overline{M})$, contradicting Step 1.

- **Step 3:** For any $\epsilon > 0$, since $(0,w^*-\epsilon) \notin \text{cl}(\overline{M})$, there exists a nonvertical hyperplane strictly separating $(0,w^*-\epsilon)$ and $\overline{M}$. This hyperplane crosses the $(n+1)$st axis at a vector $(0,\xi)$ with $w^*-\epsilon \leq \xi \leq w^*$, so $w^*-\epsilon \leq q^* \leq w^*$. Since ϵ can be arbitrarily small, it follows that $q^* = w^*$.



PROOF OF THEOREM II

- Note that $(0, w^*)$ is not a relative interior point of $\overline{M}$. Therefore, by the Proper Separation Theorem, there is a hyperplane that passes through $(0, w^*)$, contains $\overline{M}$ in one of its closed halfspaces, but does not fully contain $\overline{M}$, i.e., for some $(\mu, \beta) \neq (0, 0)$

$$\beta w^* \leq \mu' u + \beta w, \qquad \forall\ (u, w) \in \overline{M},$$

$$\beta w^* < \sup_{(u,w) \in \overline{M}} \{\mu' u + \beta w\}$$

Will show that the hyperplane is nonvertical.

- Since for any $(\overline{u}, \overline{w}) \in M$, the set $\overline{M}$ contains the halfline $\big\{(\overline{u}, w) \mid \overline{w} \leq w\big\}$, it follows that $\beta \geq 0$. If $\beta = 0$, then $0 \leq \mu' u$ for all $u \in D$. Since $0 \in \mathrm{ri}(D)$ by assumption, we must have $\mu' u = 0$ for all $u \in D$ a contradiction. Therefore, $\beta > 0$, and we can assume that $\beta = 1$. It follows that

$$w^* \leq \inf_{(u,w) \in \overline{M}} \{\mu' u + w\} = q(\mu) \leq q^*$$

Since the inequality $q^* \leq w^*$ holds always, we must have $q(\mu) = q^* = w^*$.

NONLINEAR FARKAS' LEMMA

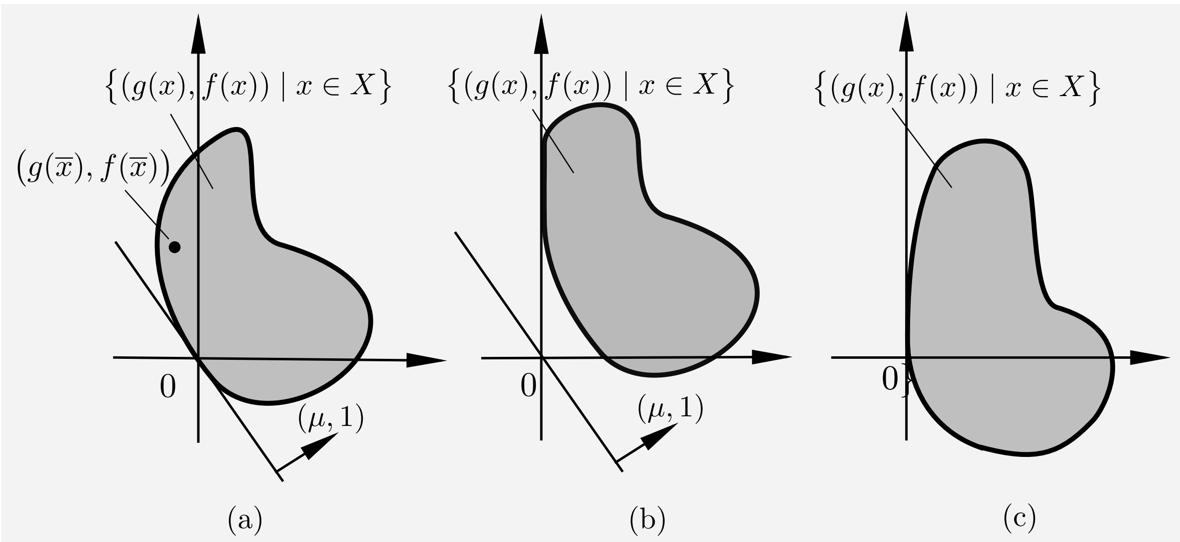
- Let $X \subset \Re^n$, $f : X \mapsto \Re$, and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, be convex. Assume that

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

Let

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

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 $j = 1, \ldots, r$.



- The lemma asserts the existence of a nonvertical hyperplane in $\Re^{r+1}$, with normal $(\mu, 1)$, that passes through the origin and contains the set

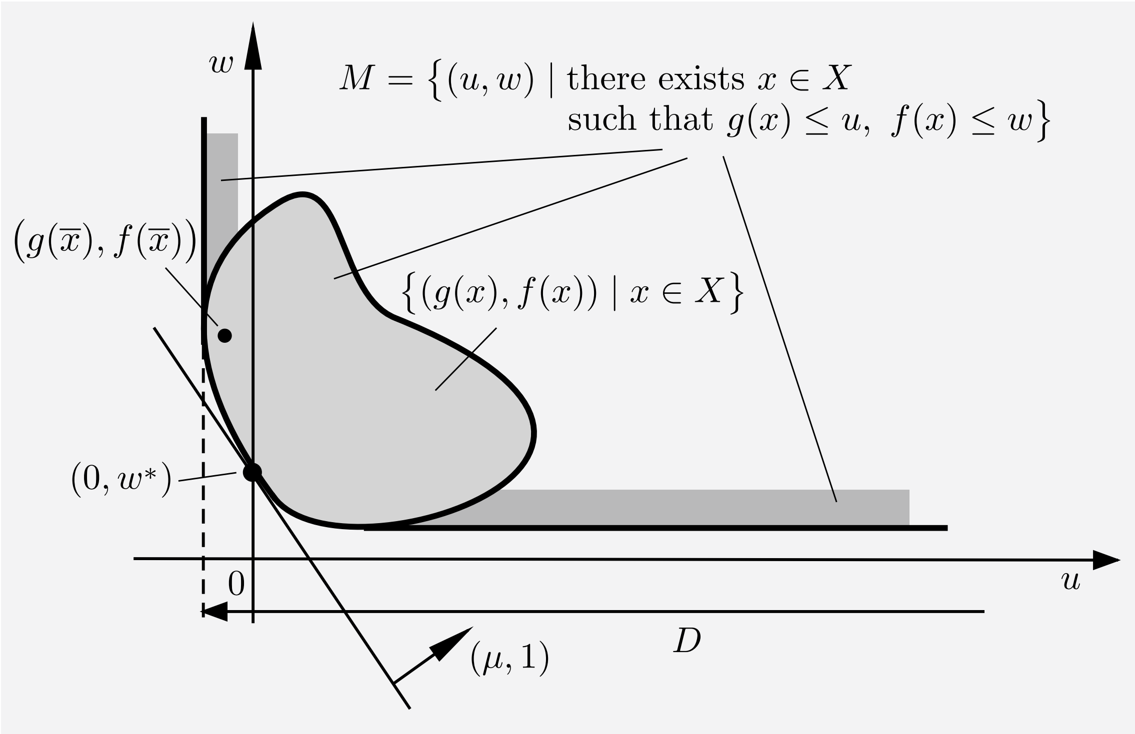
$$\big\{\big(g(x), f(x)\big) \mid x \in X\big\}$$

in its positive halfspace.

PROOF OF NONLINEAR FARKAS' LEMMA

- Apply MC/MC to

$$M = \big\{ (u,w) \mid \text{there is } x \in X \text{ s. t. } g(x) \le u,\ f(x) \le w \big\}$$



- M is equal to $\overline{M}$ and is formed as the union of positive orthants translated to points $\big(g(x), f(x)\big)$, $x \in X$.

- The convexity of X, f, and g_j implies convexity of M.

- MC/MC Theorem II applies: we have

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

and $0 \in \text{int}(D)$, because $\big(g(\overline{x}), f(\overline{x})\big) \in M$.

LECTURE 10

LECTURE OUTLINE

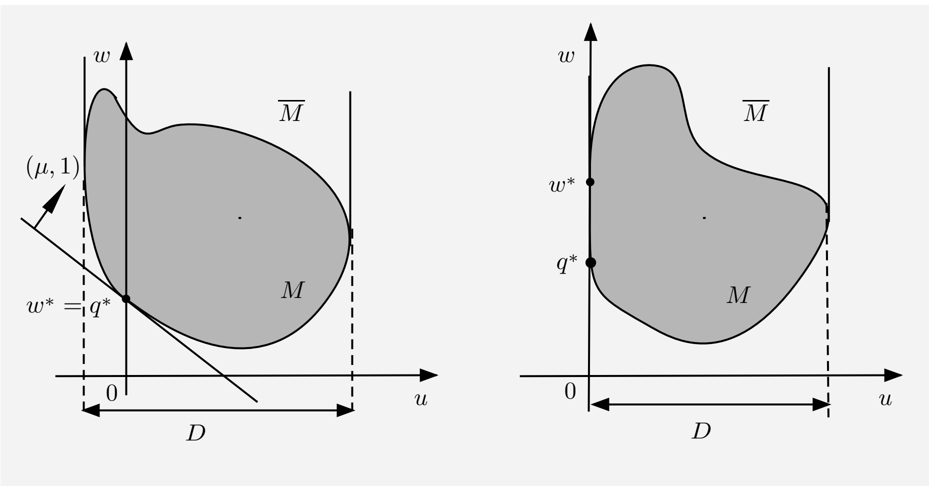
- Min Common/Max Crossing Th. III
- Nonlinear Farkas Lemma/Linear Constraints
- Linear Programming Duality
- Convex Programming Duality
- Optimality Conditions

Reading: Sections 4.5, 5.1,5.2, 5.3.1, 5.3.2

Recall the MC/MC Theorem II: If $-\infty < w^*$ and

$$0 \in \text{ri}(D) = \big\{u \mid \text{there exists } w \in \Re \text{ with } (u, w) \in \overline{M}\big\}$$

then $q^* = w^*$ and there exists μ s. t. $q(\mu) = q^*$.



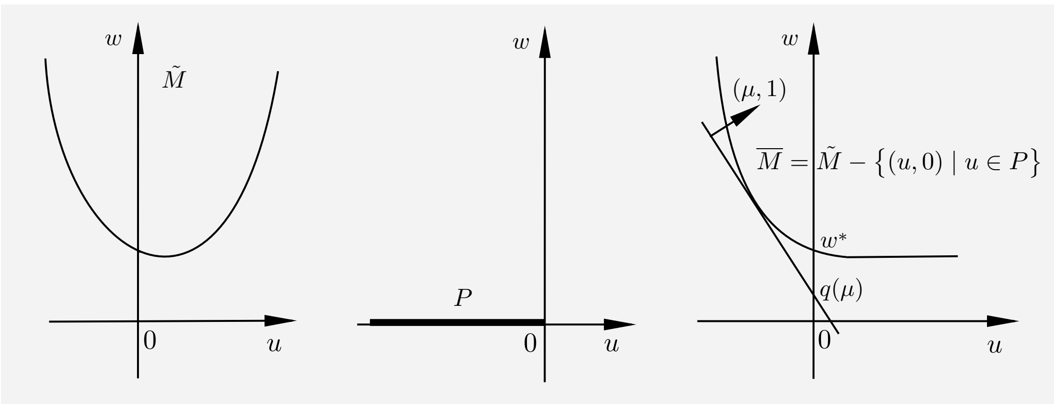
MC/MC TH. III - POLYHEDRAL

- Consider the MC/MC problems, and assume that $-\infty < w^*$ and:

(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

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

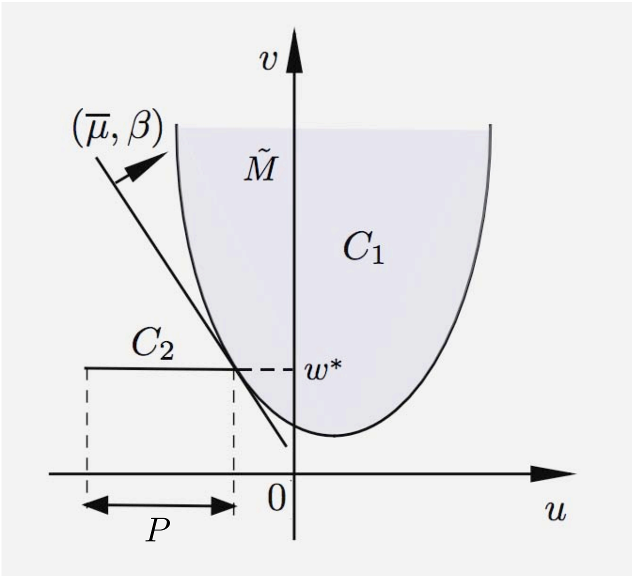
Then $q^* = w^*$, there is a max crossing solution, and all max crossing solutions $\overline{\mu}$ satisfy $\overline{\mu}'d \leq 0$ for all $d \in R_P$.

- **Comparison with Th. II:** Since $D = \tilde{D} - P$, the condition $0 \in \text{ri}(D)$ of Theorem II is

$$\text{ri}(\tilde{D}) \cap \text{ri}(P) \neq \emptyset$$

PROOF OF MC/MC TH. III

- Consider the *disjoint* convex sets $C_1 = \{(u, v) \mid v > w \text{ for some } (u, w) \in \tilde{M}\}$ and $C_2 = \{(u, w^*) \mid u \in P\}$ [$u \in P$ and $(u, w) \in \tilde{M}$ with $w^* > w$ contradicts the definition of w^*]



- Since C_2 is polyhedral, there exists a separating hyperplane not containing C_1, i.e., a $(\overline{\mu}, \beta) \neq (0, 0)$ such that

$$\beta w^* + \overline{\mu}' z \leq \beta v + \overline{\mu}' x, \quad \forall\ (x, v) \in C_1,\ \forall\ z \in P$$

$$\inf_{(x,v) \in C_1} \{\beta v + \overline{\mu}' x\} < \sup_{(x,v) \in C_1} \{\beta v + \overline{\mu}' x\}$$

Since $(0, 1)$ is a direction of recession of C_1, we see that $\beta \geq 0$. Because of the relative interior point assumption, $\beta \neq 0$, so we may assume that $\beta = 1$.

PROOF (CONTINUED)

- Hence,

$$w^* + \overline{\mu}'z \le \inf_{(u,v)\in C_1} \{v + \overline{\mu}'u\}, \qquad \forall\, z \in P,$$

so that

$$\begin{aligned} w^* &\le \inf_{(u,v)\in C_1,\, z\in P} \big\{v + \overline{\mu}'(u-z)\big\} \\ &= \inf_{(u,v)\in \tilde{M}-P} \{v + \overline{\mu}'u\} \\ &= \inf_{(u,v)\in \overline{M}} \{v + \overline{\mu}'u\} \\ &= q(\overline{\mu}) \end{aligned}$$

Using $q^* \le w^*$ (weak duality), we have $q(\overline{\mu}) = q^* = w^*$.

Proof that all max crossing solutions $\overline{\mu}$ satisfy $\overline{\mu}'d \le 0$ for all $d \in R_P$: follows from

$$q(\mu) = \inf_{(u,v)\in C_1,\, z\in P} \big\{v + \mu'(u-z)\big\}$$

so that $q(\mu) = -\infty$ if $\mu'd > 0$. **Q.E.D.**

- Geometrical intuition: every $(0, -d)$ with $d \in R_P$, is direction of recession of $\overline{M}$.

MC/MC TH. III - A SPECIAL CASE

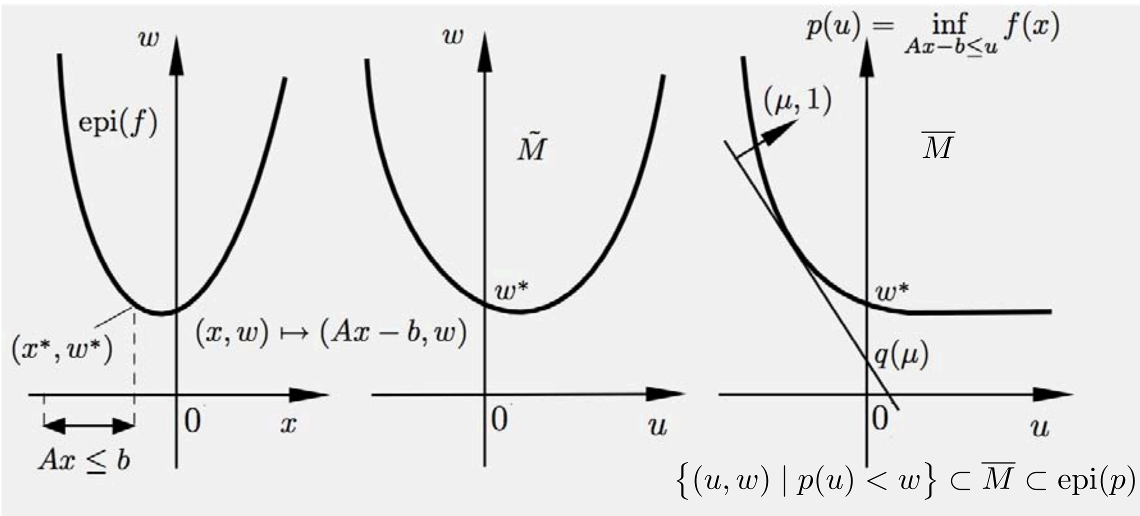
- Consider the MC/MC framework, and assume:

(1) For a convex function $f : \Re^m \mapsto (-\infty, \infty]$, an $r \times m$ matrix A, and a vector $b \in \Re^r$:

$$\overline{M} = \big\{(u, w) \mid \text{for some } (x, w) \in \text{epi}(f),\ Ax - b \leq u\big\}$$

so $\overline{M} = \tilde{M} + \text{Positive Orthant}$, where

$$\tilde{M} = \big\{(Ax - b, w) \mid (x, w) \in \text{epi}(f)\big\}$$



(2) There is an $\overline{x} \in \text{ri}(\text{dom}(f))$ s. t. $A\overline{x} - b \leq 0$.

Then $q^* = w^*$ and there is a $\mu \geq 0$ with $q(\mu) = q^*$.

- Also $\overline{M} = M \approx \text{epi}(p)$, where $p(u) = \inf_{Ax-b\leq u} f(x)$.
- We have $w^* = p(0) = \inf_{Ax-b\leq 0} f(x)$.

NONL. FARKAS' L. - POLYHEDRAL ASSUM.

- Let $X \subset \Re^n$ be convex, and $f : X \mapsto \Re$ and $g_j : \Re^n \mapsto \Re$, $j = 1, \ldots, r$, be linear so $g(x) = Ax - b$ for some A and b. Assume that

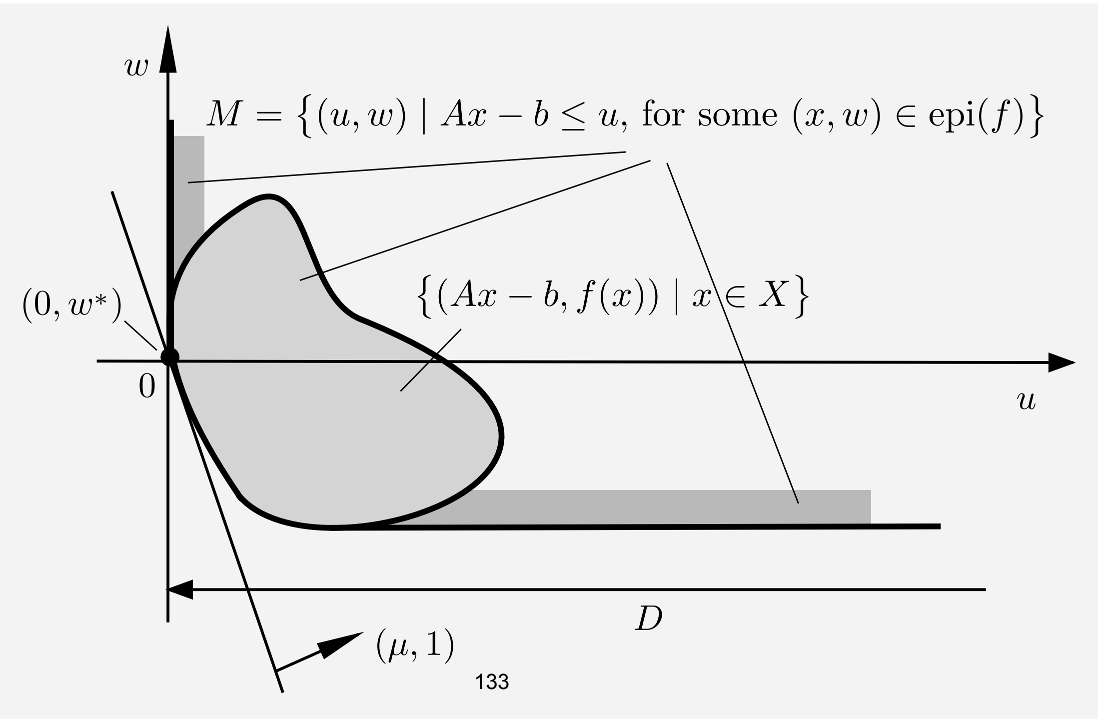
$$f(x) \geq 0, \qquad \forall\, x \in X \text{ with } Ax - b \leq 0$$

Let

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

Assume that there exists a vector $\overline{x} \in \mathrm{ri}(X)$ such that $A\overline{x} - b \leq 0$. Then Q^* is nonempty.

Proof: As before, apply special case of MC/MC Th. III of preceding slide, using the fact $w^* \geq 0$, implied by the assumption.



(LINEAR) FARKAS' LEMMA

- Let A be an $m \times n$ matrix and $c \in \Re^m$. The system $Ay = c$, $y \geq 0$ has a solution if and only if

$$A'x \leq 0 \qquad \Rightarrow \qquad c'x \leq 0. \qquad (*)$$

- **Alternative/Equivalent Statement:** If $P = \text{cone}\{a_1, \ldots, a_n\}$, where $a_1, \ldots, a_n$ are the columns of A, then $P = (P^*)^*$ (Polar Cone Theorem).

Proof: If $y \in \Re^n$ is such that $Ay = c$, $y \geq 0$, then $y'A'x = c'x$ for all $x \in \Re^m$, which implies Eq. (*).

Conversely, apply the Nonlinear Farkas' Lemma with $f(x) = -c'x$, $g(x) = A'x$, and $X = \Re^m$. Condition (*) implies the existence of $\mu \geq 0$ such that

$$-c'x + \mu'A'x \geq 0, \qquad \forall\, x \in \Re^m,$$

or equivalently

$$(A\mu - c)'x \geq 0, \qquad \forall\, x \in \Re^m,$$

or $A\mu = c$.

LINEAR PROGRAMMING DUALITY

- Consider the linear program

$$\begin{aligned} &\text{minimize } \ c'x \\ &\text{subject to } \ a_j'x \geq b_j, \quad j = 1, \ldots, r, \end{aligned}$$

where $c \in \Re^n$, $a_j \in \Re^n$, and $b_j \in \Re$, $j = 1, \ldots, r$.

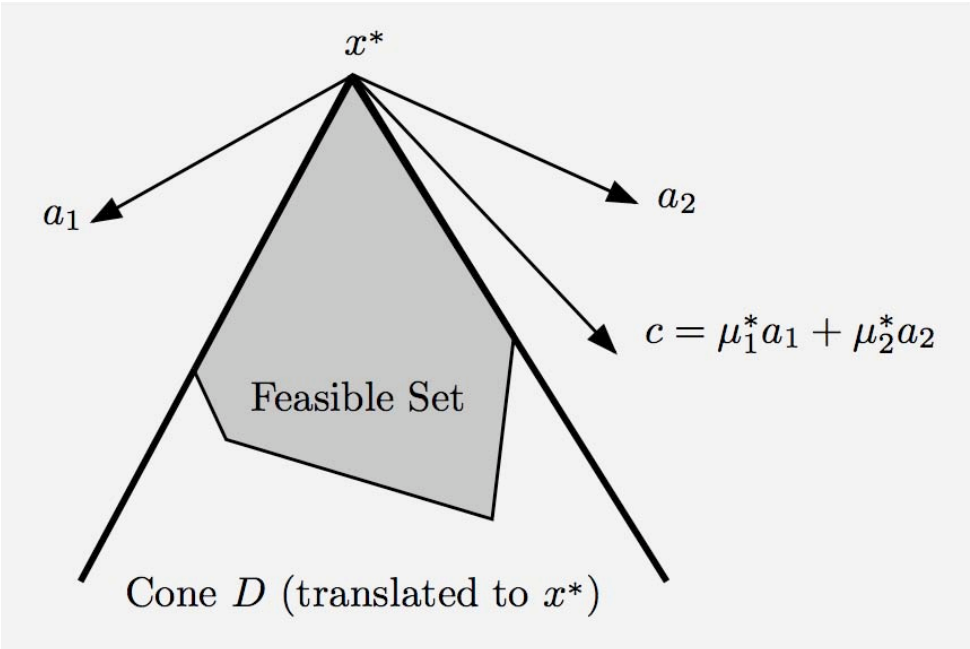
- The dual problem is

$$\begin{aligned} &\text{maximize } \ b'\mu \\ &\text{subject to } \ \sum_{j=1}^{r} a_j \mu_j = c, \quad \mu \geq 0. \end{aligned}$$

- **Linear Programming Duality Theorem:**
  (a) If either f^* or q^* is finite, then $f^* = q^*$ and both the primal and the dual problem have optimal solutions.
  (b) If $f^* = -\infty$, then $q^* = -\infty$.
  (c) If $q^* = \infty$, then $f^* = \infty$.

Proof: (b) and (c) follow from weak duality. For part (a): If f^* is finite, there is a primal optimal solution x^*, by existence of solutions of quadratic programs. Use Farkas' Lemma to construct a dual feasible μ^* such that $c'x^* = b'\mu^*$ (next slide).

PROOF OF LP DUALITY (CONTINUED)



- Let x^* be a primal optimal solution, and let $J = \{j \mid a_j' x^* = b_j\}$. Then, $c'y \geq 0$ for all y in the cone of "feasible directions"

$$D = \{y \mid a_j' y \geq 0,\ \forall\ j \in J\}$$

By Farkas' Lemma, for some scalars $\mu_j^* \geq 0$, c can be expressed as

$$c = \sum_{j=1}^{r} \mu_j^* a_j, \quad \mu_j^* \geq 0,\ \forall\ j \in J,\ \ \mu_j^* = 0,\ \forall\ j \notin J.$$

Taking inner product with x^*, we obtain $c'x^* = b'\mu^*$, which in view of $q^* \leq f^*$, shows that $q^* = f^*$ and that μ^* is optimal.

LINEAR PROGRAMMING OPT. CONDITIONS

A pair of vectors (x^*, μ^*) form a primal and dual optimal solution pair if and only if x^* is primal-feasible, μ^* is dual-feasible, and

$$\mu_j^*(b_j - a_j' x^*) = 0, \qquad \forall\, j = 1, \ldots, r. \qquad (*)$$

Proof: If x^* is primal-feasible and μ^* is dual-feasible, then

$$\begin{aligned} b'\mu^* &= \sum_{j=1}^{r} b_j \mu_j^* + \left(c - \sum_{j=1}^{r} a_j \mu_j^* \right)' x^* \\ &= c'x^* + \sum_{j=1}^{r} \mu_j^*(b_j - a_j' x^*) \end{aligned} \qquad (**)$$

So if Eq. (*) holds, we have $b'\mu^* = c'x^*$, and weak duality implies that x^* is primal optimal and μ^* is dual optimal.

Conversely, if (x^*, μ^*) form a primal and dual optimal solution pair, then x^* is primal-feasible, μ^* is dual-feasible, and by the duality theorem, we have $b'\mu^* = c'x^*$. From Eq. (**), we obtain Eq. (*).

CONVEX PROGRAMMING

Consider the problem

$$\begin{aligned}&\text{minimize } \ f(x)\\&\text{subject to } \ x \in X, \ g_j(x) \leq 0, \ j = 1, \dots, r,\end{aligned}$$

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

- Recall the connection with the max crossing problem in the MC/MC framework where $M = \text{epi}(p)$ with

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

- Consider the Lagrangian function

$$L(x, \mu) = f(x) + \mu' g(x),$$

the dual function

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

and the dual problem of maximizing $\inf_{x \in X} L(x, \mu)$ over $\mu \geq 0$.

STRONG DUALITY THEOREM

- Assume that f^* is finite, and that 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 $\overline{x} \in \mathrm{ri}(X)$ such that $g(\overline{x}) \leq 0$.

Then $q^* = f^*$ and the set of optimal solutions of the dual problem is nonempty. Under condition (1) this set is also compact.

- **Proof:** Replace $f(x)$ by $f(x) - f^*$ so that $f(x) - f^* \geq 0$ for all $x \in X$ w/ $g(x) \leq 0$. Apply Nonlinear Farkas' Lemma. Then, there exist $\mu_j^* \geq 0$, s.t.

$$f^* \leq f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x), \qquad \forall\, x \in X$$

- It follows that

$$f^* \leq \inf_{x \in X} \big\{ f(x) + \mu^{*\prime} g(x) \big\} \leq \inf_{x \in X,\, g(x) \leq 0} f(x) = f^*.$$

Thus equality holds throughout, and we have

$$f^* = \inf_{x \in X} \left\{ f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \right\} = q(\mu^*)$$

QUADRATIC PROGRAMMING DUALITY

- Consider the quadratic program

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

where Q is positive definite.

- If f^* is finite, then $f^* = q^*$ and there exist both primal and dual optimal solutions, since the constraints are linear.

- Calculation of dual function:

$$q(\mu) = \inf_{x \in \Re^n} \{ \tfrac{1}{2}x'Qx + c'x + \mu'(Ax - b) \}$$

The infimum is attained for $x = -Q^{-1}(c + A'\mu)$, and, after substitution and calculation,

$$q(\mu) = -\tfrac{1}{2}\mu' AQ^{-1}A'\mu - \mu'(b + AQ^{-1}c) - \tfrac{1}{2}c'Q^{-1}c$$

- The dual problem, after a sign change, is

$$\begin{aligned} &\text{minimize} \ \ \tfrac{1}{2}\mu'P\mu + t'\mu \\ &\text{subject to} \ \ \mu \ge 0, \end{aligned}$$

where $P = AQ^{-1}A'$ and $t = b + AQ^{-1}c$.

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

$$x^* \in \arg\min_{x \in X} L(x, \mu^*), \qquad \mu_j^* g_j(x^*) = 0, \quad \forall\, j. \tag{1}$$

Proof: If $q^* = f^*$, and x^*, μ^* are optimal, then

$$\begin{aligned} f^* = q^* = q(\mu^*) = \inf_{x \in X} L(x, \mu^*) &\leq L(x^*, \mu^*) \\ &= f(x^*) + \sum_{j=1}^{r} \mu_j^* g_j(x^*) \leq f(x^*), \end{aligned}$$

where the last inequality follows from $\mu_j^* \geq 0$ and $g_j(x^*) \leq 0$ for all j. Hence equality holds throughout above, and (1) holds.

Conversely, if x^*, μ^* are feasible, and (1) holds,

$$\begin{aligned} q(\mu^*) = \inf_{x \in X} L(x, \mu^*) &= L(x^*, \mu^*) \\ &= f(x^*) + \sum_{j=1}^{r} \mu_j^* g_j(x^*) = f(x^*), \end{aligned}$$

so $q^* = f^*$, and x^*, μ^* are optimal. **Q.E.D.**

QUADRATIC PROGRAMMING OPT. COND.

For the quadratic program

$$\text{minimize } \tfrac{1}{2}x'Qx + c'x$$
$$\text{subject to } Ax \le b,$$

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

- Primal and dual feasibility holds:

$$Ax^* \le b, \qquad \mu^* \ge 0$$

- Lagrangian optimality holds [x^* minimizes $L(x, \mu^*)$ over $x \in \Re^n$]. This yields

$$x^* = -Q^{-1}(c + A'\mu^*)$$

- Complementary slackness holds [$(Ax^* - b)'\mu^* = 0$]. It can be written as

$$\mu_j^* > 0 \qquad \Rightarrow \qquad a_j'x^* = b_j, \quad \forall\, j = 1, \ldots, r,$$

where a_j' is the jth row of A, and b_j is the jth component of b.

LINEAR EQUALITY CONSTRAINTS

- The problem is

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \;\; x \in X, \;\; g(x) \leq 0, \;\; Ax = b, \end{aligned}$$

where X is convex, $g(x) = \big(g_1(x), \ldots, g_r(x)\big)'$, $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, are convex.

- Convert the constraint $Ax = b$ to $Ax \leq b$ and $-Ax \leq -b$, with corresponding dual variables $\lambda^+ \geq 0$ and $\lambda^- \geq 0$.

- The Lagrangian function is

$$f(x) + \mu' g(x) + (\lambda^+ - \lambda^-)'(Ax - b),$$

and by introducing a dual variable $\lambda = \lambda^+ - \lambda^-$, with no sign restriction, it can be written as

$$L(x, \mu, \lambda) = f(x) + \mu' g(x) + \lambda'(Ax - b).$$

- The dual problem is

$$\begin{aligned} &\text{maximize} \quad q(\mu, \lambda) \equiv \inf_{x \in X} L(x, \mu, \lambda) \\ &\text{subject to} \quad \mu \geq 0, \;\; \lambda \in \Re^m. \end{aligned}$$

DUALITY AND OPTIMALITY COND.

- **Pure equality constraints:**

  (a) Assume that f^*: finite and there exists $\overline{x} \in \text{ri}(X)$ such that $A\overline{x} = b$. Then $f^* = q^*$ and there exists a dual optimal solution.

  (b) $f^* = q^*$, and (x^*, λ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, and

$$x^* \in \arg\min_{x \in X} L(x, \lambda^*)$$

Note: No complementary slackness for equality constraints.

- **Linear and nonlinear constraints:**

  (a) Assume f^*: finite, that there exists $\overline{x} \in X$ such that $A\overline{x} = b$ and $g(\overline{x}) < 0$, and that there exists $\tilde{x} \in \text{ri}(X)$ such that $A\tilde{x} = b$. Then $q^* = f^*$ and there exists a dual optimal solution.

  (b) $f^* = q^*$, and (x^*, μ^*, λ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, $\mu^* \geq 0$, and

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

LECTURE 11

LECTURE OUTLINE

- Review of convex progr. duality/counterexamples
- Fenchel Duality
- Conic Duality

Reading: Sections 5.3.1-5.3.6

Line of analysis so far:

- Convex analysis (rel. int., dir. of recession, hyperplanes, conjugacy)
- MC/MC - Three general theorems: Strong duality, existence of dual optimal solutions, polyhedral refinements
- Nonlinear Farkas' Lemma
- Linear programming (duality, opt. conditions)
- Convex programming

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \;\; x \in X, \quad g(x) \leq 0, \quad Ax = b, \end{aligned}$$

where X is convex, $g(x) = \big(g_1(x), \ldots, g_r(x)\big)'$, $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, are convex. (Nonlin. Farkas' Lemma, duality, opt. conditions)

DUALITY AND OPTIMALITY COND.

- **Pure equality constraints:**

  (a) Assume that f^*: finite and there exists $\overline{x} \in \text{ri}(X)$ such that $A\overline{x} = b$. Then $f^* = q^*$ and there exists a dual optimal solution.

  (b) $f^* = q^*$, and (x^*, λ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, and

$$x^* \in \arg\min_{x \in X} L(x, \lambda^*)$$

Note: No complementary slackness for equality constraints.

- **Linear and nonlinear constraints:**

  (a) Assume f^*: finite, that there exists $\overline{x} \in X$ such that $A\overline{x} = b$ and $g(\overline{x}) < 0$, and that there exists $\tilde{x} \in \text{ri}(X)$ such that $A\tilde{x} = b$. Then $q^* = f^*$ and there exists a dual optimal solution.

  (b) $f^* = q^*$, and (x^*, μ^*, λ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, $\mu^* \geq 0$, and

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

COUNTEREXAMPLE I

- **Strong Duality Counterexample:** Consider

$$\text{minimize } \; f(x) = e^{-\sqrt{x_1 x_2}}$$
$$\text{subject to } \; x_1 = 0, \qquad x \in X = \{x \mid x \geq 0\}$$

Here $f^* = 1$ and f is convex (its Hessian is > 0 in the interior of X). The dual function is

$$q(\lambda) = \inf_{x \geq 0} \left\{ e^{-\sqrt{x_1 x_2}} + \lambda x_1 \right\} = \begin{cases} 0 & \text{if } \lambda \geq 0, \\ -\infty & \text{otherwise,} \end{cases}$$

(when $\lambda \geq 0$, the expression in braces is nonnegative for $x \geq 0$ and can approach zero by taking $x_1 \to 0$ and $x_1 x_2 \to \infty$). Thus $q^* = 0$.

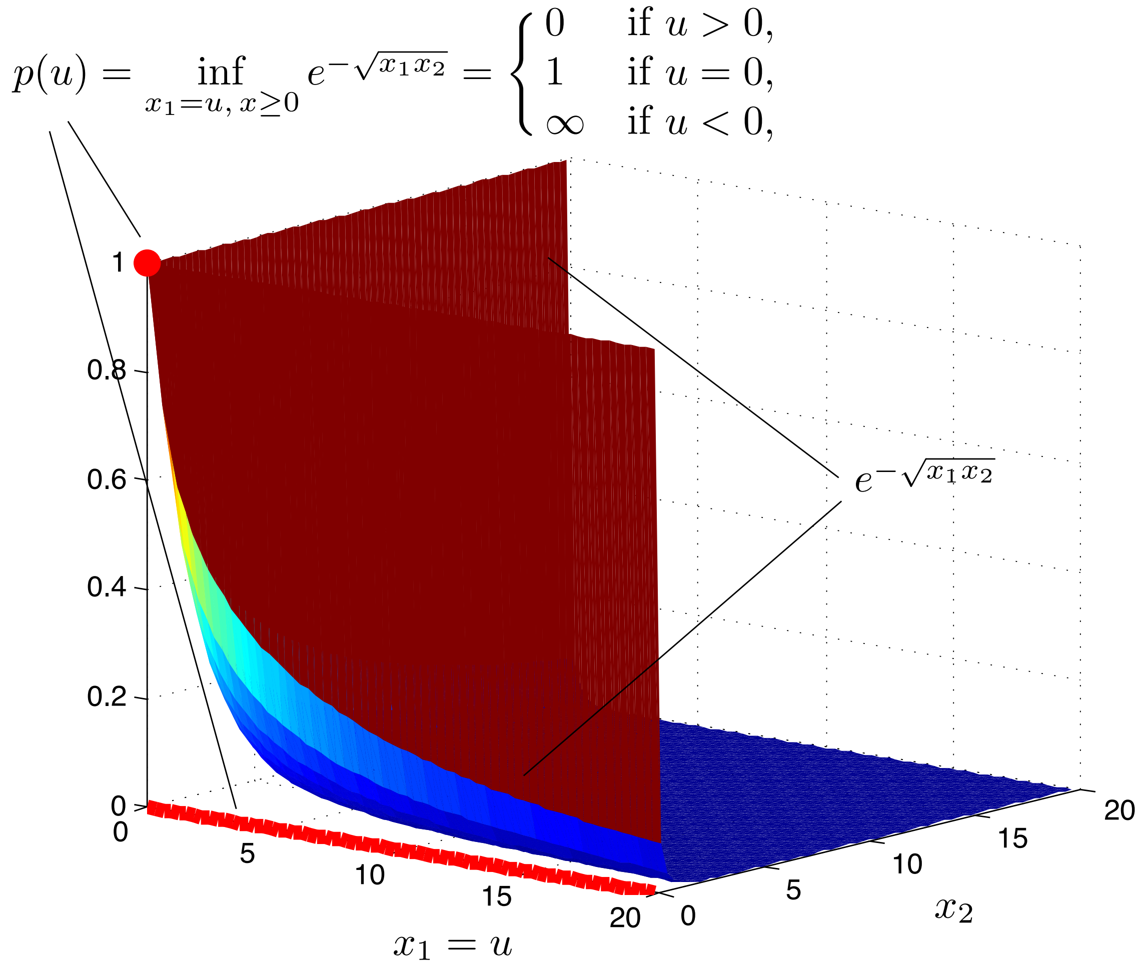
- The relative interior assumption is violated.

- As predicted by the corresponding MC/MC framework, the perturbation function

$$p(u) = \inf_{x_1 = u,\, x \geq 0} e^{-\sqrt{x_1 x_2}} = \begin{cases} 0 & \text{if } u > 0, \\ 1 & \text{if } u = 0, \\ \infty & \text{if } u < 0, \end{cases}$$

is not lower semicontinuous at $u = 0$.

COUNTEREXAMPLE VISUALIZATION



- Connection with counterexample for preservation of closedness under partial minimization.

COUNTEREXAMPLE II

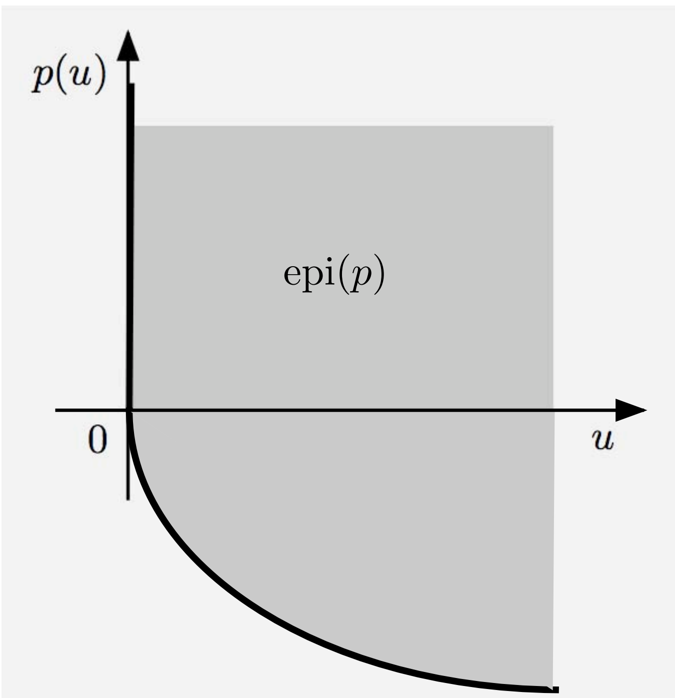
- **Existence of Solutions Counterexample:** Let $X = \Re$, $f(x) = x$, $g(x) = x^2$. Then $x^* = 0$ is the only feasible/optimal solution, and we have

$$q(\mu) = \inf_{x \in \Re} \{x + \mu x^2\} = -\frac{1}{4\mu}, \qquad \forall\ \mu > 0,$$

and $q(\mu) = -\infty$ for $\mu \leq 0$, so that $q^* = f^* = 0$. However, there is no $\mu^* \geq 0$ such that $q(\mu^*) = q^* = 0$.

- The perturbation function is

$$p(u) = \inf_{x^2 \leq u} x = \begin{cases} -\sqrt{u} & \text{if } u \geq 0, \\ \infty & \text{if } u < 0. \end{cases}$$



FENCHEL DUALITY FRAMEWORK

- Consider the problem

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

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

- Convert to the equivalent problem

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

- The dual function is

$$\begin{aligned} q(\lambda) &= \inf_{x_1 \in \text{dom}(f_1), \; x_2 \in \text{dom}(f_2)} \big\{ f_1(x_1) + f_2(x_2) + \lambda'(x_2 - x_1) \big\} \\ &= \inf_{x_1 \in \Re^n} \big\{ f_1(x_1) - \lambda' x_1 \big\} + \inf_{x_2 \in \Re^n} \big\{ f_2(x_2) + \lambda' x_2 \big\} \end{aligned}$$

- **Dual problem:** $\max_\lambda \{-f_1^\star(\lambda) - f_2^\star(-\lambda)\} = -\min_\lambda \{-q(\lambda)\}$ or

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

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

FENCHEL DUALITY THEOREM

- Consider the Fenchel framework:

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

  (b) There holds $f^* = q^*$, and (x^*, λ^*) is a primal and dual optimal solution pair if and only if

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

Proof: For strong duality use the equality constrained problem

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

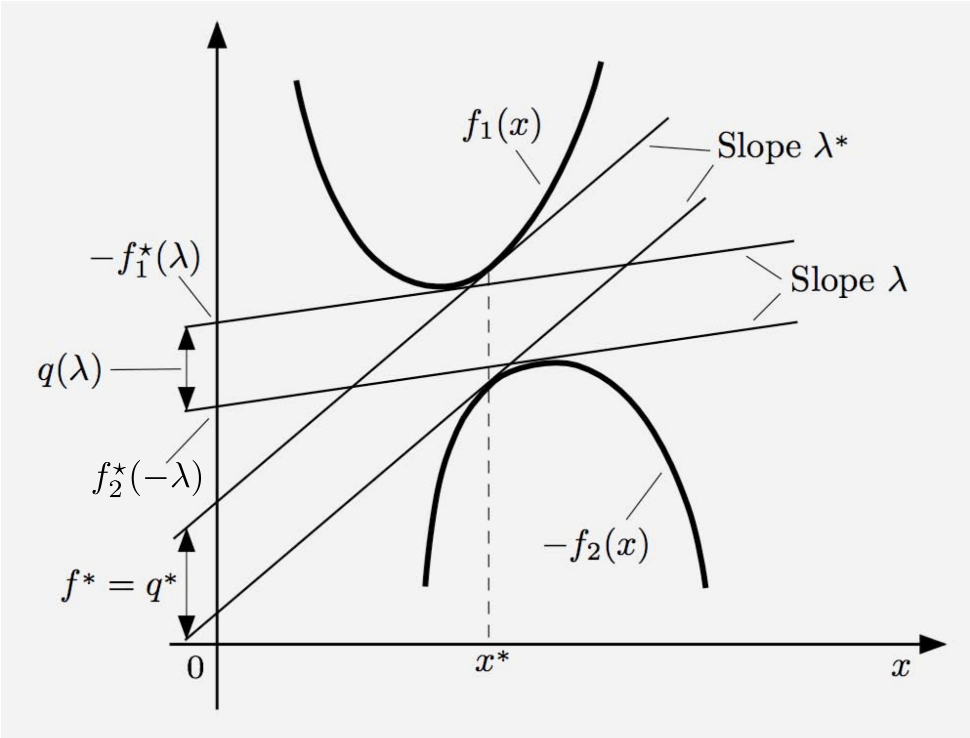
and the fact

$$\text{ri}\big(\text{dom}(f_1) \times \text{dom}(f_2)\big) = \text{ri}\big(\text{dom}(f_1)\big) \times \big(\text{dom}(f_2)\big)$$

to satisfy the relative interior condition.

For part (b), apply the optimality conditions (primal and dual feasibility, and Lagrangian optimality).

GEOMETRIC INTERPRETATION



• When $\mathrm{dom}(f_1) = \mathrm{dom}(f_2) = \Re^n$, and f_1 and f_2 are differentiable, the optimality condition is equivalent to

$$\lambda^* = \nabla f_1(x^*) = -\nabla f_2(x^*)$$

• By reversing the roles of the (symmetric) primal and dual problems, we obtain alternative criteria for strong duality: if q^* is finite and $\mathrm{ri}\big(\mathrm{dom}(f_1^\star)\big) \cap \mathrm{ri}\big(-\mathrm{dom}(f_2^\star)\big) \neq \emptyset$, then $f^* = q^*$ and there exists at least one primal optimal solution.

CONIC PROBLEMS

• A conic problem is to minimize a convex function $f : \Re^n \mapsto (-\infty, \infty]$ subject to a cone constraint.

• The most useful/popular special cases:

– Linear-conic programming

– Second order cone programming

– Semidefinite programming

involve minimization of a linear function over the intersection of an affine set and a cone.

• Can be analyzed as a special case of Fenchel duality.

• There are many interesting applications of conic problems, including in discrete optimization.

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

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

The conjugates are

$$f_1^\star(\lambda) = \sup_{x \in \Re^n} \big\{\lambda' x - f(x)\big\}, \; f_2^\star(\lambda) = \sup_{x \in C} \lambda' x = \begin{cases} 0 & \text{if } \lambda \in C^*, \\ \infty & \text{if } \lambda \notin C^*, \end{cases}$$

where $C^* = \{\lambda \mid \lambda' x \le 0, \forall\, x \in C\}$.

• The dual problem is

$$\begin{aligned} &\text{minimize} \quad f^\star(\lambda) \\ &\text{subject to} \;\; \lambda \in \hat{C}, \end{aligned}$$

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

$$\hat{C} = \{\lambda \mid \lambda' x \ge 0, \forall\, x \in C\}.$$

$\hat{C}$ and $-\hat{C}$ are called the *dual* and *polar* cones.

CONIC DUALITY THEOREM

- Assume that the optimal value of the primal conic problem is finite, and that

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

Then, there is no duality gap and the dual problem has an optimal solution.

- Using the symmetry of the primal and dual problems, we also obtain that there is no duality gap and the primal problem has an optimal solution if the optimal value of the dual conic problem is finite, and

$$\text{ri}\big(\text{dom}(f^\star)\big) \cap \text{ri}(\hat{C}) \neq \emptyset.$$

LINEAR CONIC PROGRAMMING

- Let f be linear over its domain, i.e.,

$$f(x) = \begin{cases} c'x & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

where c is a vector, and $X = b + S$ is an affine set.

- Primal problem is

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

- We have

$$\begin{aligned} f^\star(\lambda) &= \sup_{x-b\in S} (\lambda - c)'x = \sup_{y\in S} (\lambda - c)'(y + b) \\ &= \begin{cases} (\lambda - c)'b & \text{if } \lambda - c \in S^\perp, \\ \infty & \text{if } \lambda - c \notin S. \end{cases} \end{aligned}$$

- Dual problem is equivalent to

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

- If $X \cap \text{ri}(C) = \emptyset$, there is no duality gap and there exists a dual optimal solution.

ANOTHER APPROACH TO DUALITY

- Consider the problem

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

and perturbation fn $p(u) = \inf_{x \in X,\, g(x) \leq u} f(x)$

- Recall the MC/MC framework with $M = \text{epi}(p)$. Assuming that p is convex and $f^* < \infty$, by 1st MC/MC theorem, we have $f^* = q^*$ if and only if p is lower semicontinuous at 0.

- **Duality Theorem:** Assume that X, f, and g_j are closed convex, and the feasible set is nonempty and compact. Then $f^* = q^*$ and the set of optimal primal solutions is nonempty and compact.

Proof: Use partial minimization theory w/ the function

$$F(x, u) = \begin{cases} f(x) & \text{if } x \in X,\ g(x) \leq u, \\ \infty & \text{otherwise.} \end{cases}$$

p is obtained by the partial minimization:

$$p(u) = \inf_{x \in \Re^n} F(x, u).$$

Under the given assumption, p is closed convex.

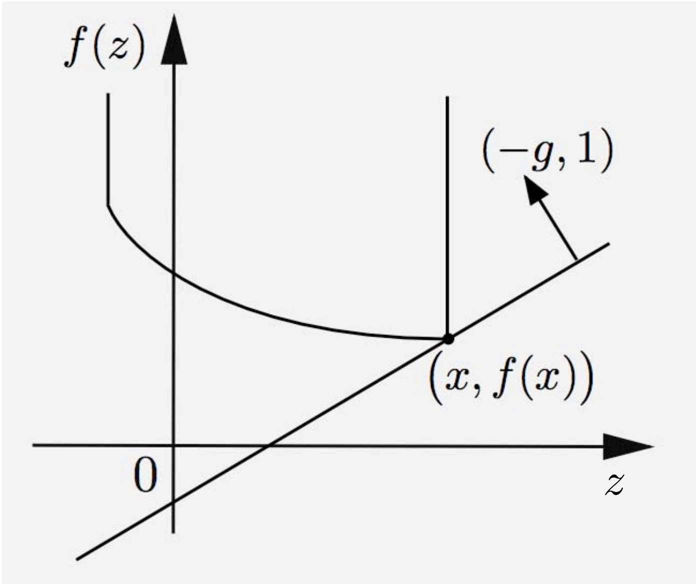
LECTURE 12

LECTURE OUTLINE

- Subgradients
- Fenchel inequality
- Sensitivity in constrained optimization
- Subdifferential calculus
- Optimality conditions

Reading: Section 5.4

SUBGRADIENTS



- Let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function. A vector $g \in \Re^n$ is a *subgradient* of f at a point $x \in \text{dom}(f)$ if

$$f(z) \geq f(x) + (z - x)'g, \qquad \forall\ z \in \Re^n$$

- **Support Hyperplane Interpretation:** g is a subgradient if and only if

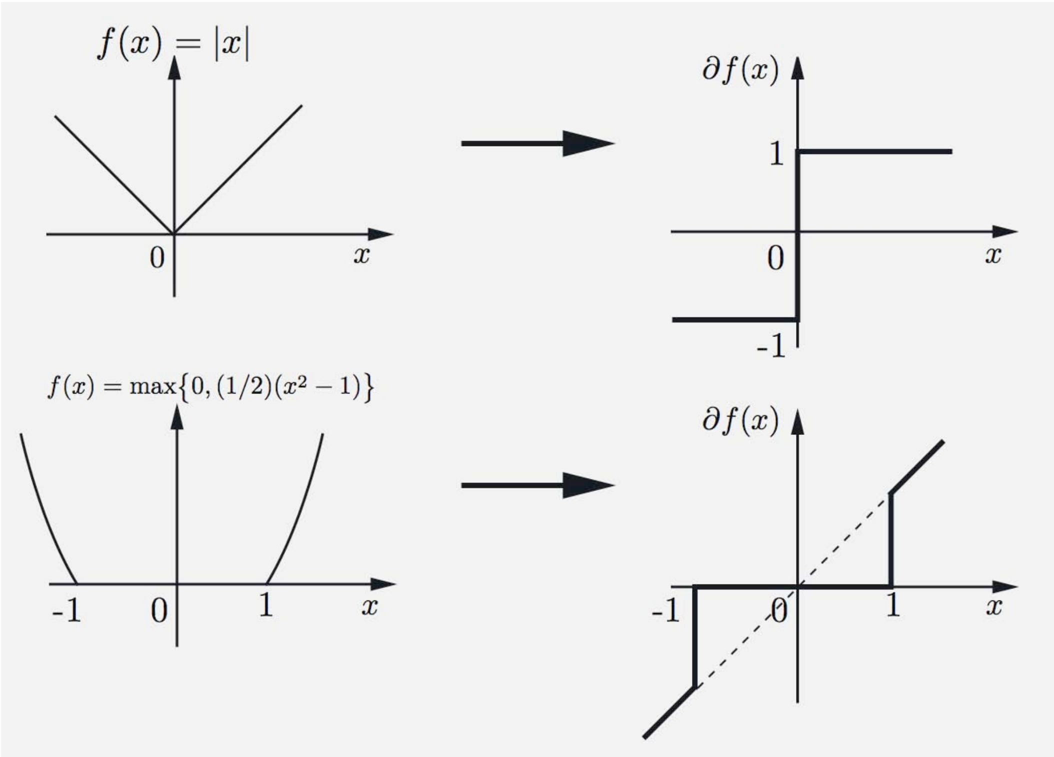
$$f(z) - z'g \geq f(x) - x'g, \qquad \forall\ z \in \Re^n$$

so g is a subgradient at x if and only if the hyperplane in $\Re^{n+1}$ that has normal $(-g, 1)$ and passes through $\big(x, f(x)\big)$ supports the epigraph of f.

- The set of all subgradients at x is the *subdifferential of f at x*, denoted $\partial f(x)$.
- By convention $\partial f(x) = \emptyset$ for $x \notin \text{dom}(f)$.

EXAMPLES OF SUBDIFFERENTIALS

- Some examples:



- If f is differentiable, then $\partial f(x) = \{\nabla f(x)\}$.

Proof: If $g \in \partial f(x)$, then

$$f(x + z) \geq f(x) + g'z, \qquad \forall\ z \in \Re^n.$$

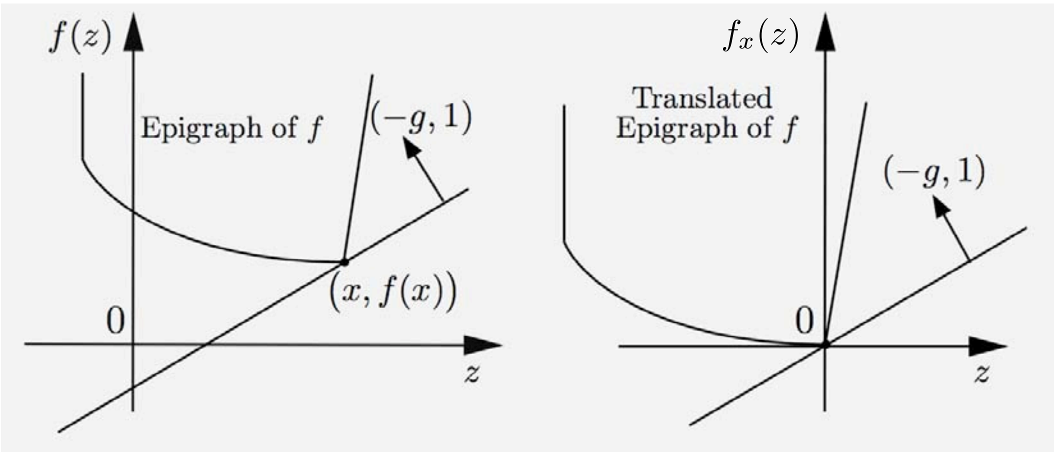
Apply this with $z = \gamma\big(\nabla f(x) - g\big)$, $\gamma \in \Re$, and use 1st order Taylor series expansion to obtain

$$\|\nabla f(x) - g\|^2 \leq -o(\gamma)/\gamma, \qquad \forall\ \gamma < 0$$

EXISTENCE OF SUBGRADIENTS

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be proper convex.
- Consider MC/MC with

$$M = \text{epi}(f_x), \qquad f_x(z) = f(x+z) - f(x)$$



- By 2nd MC/MC Duality Theorem, $\partial f(x)$ is nonempty and compact if and only if x is in the interior of dom(f).
- More generally: for every $x \in \text{ri}\big(\text{dom}(f)\big)$,

$$\partial f(x) = S^{\perp} + G,$$

where:

- S is the subspace that is parallel to the affine hull of dom(f)
- G is a nonempty and compact set.

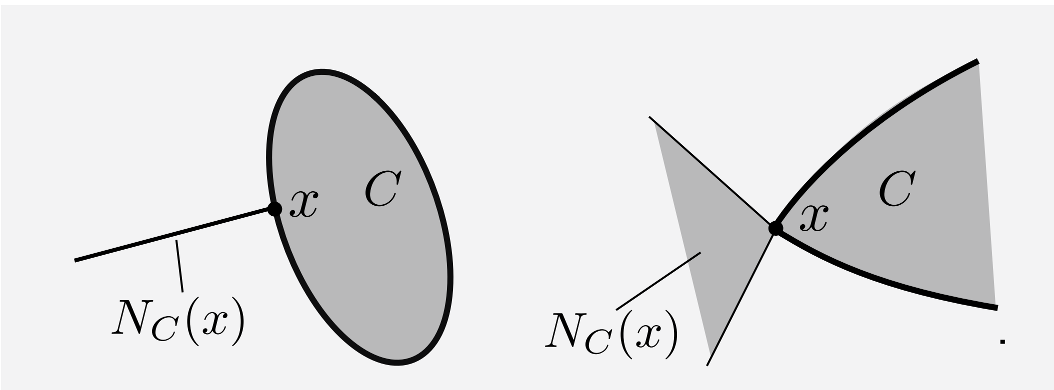
EXAMPLE: SUBDIFFERENTIAL OF INDICATOR

- Let C be a convex set, and δ_C be its indicator function.
- For $x \notin C$, $\partial\delta_C(x) = \emptyset$ (by convention).
- For $x \in C$, we have $g \in \partial\delta_C(x)$ iff

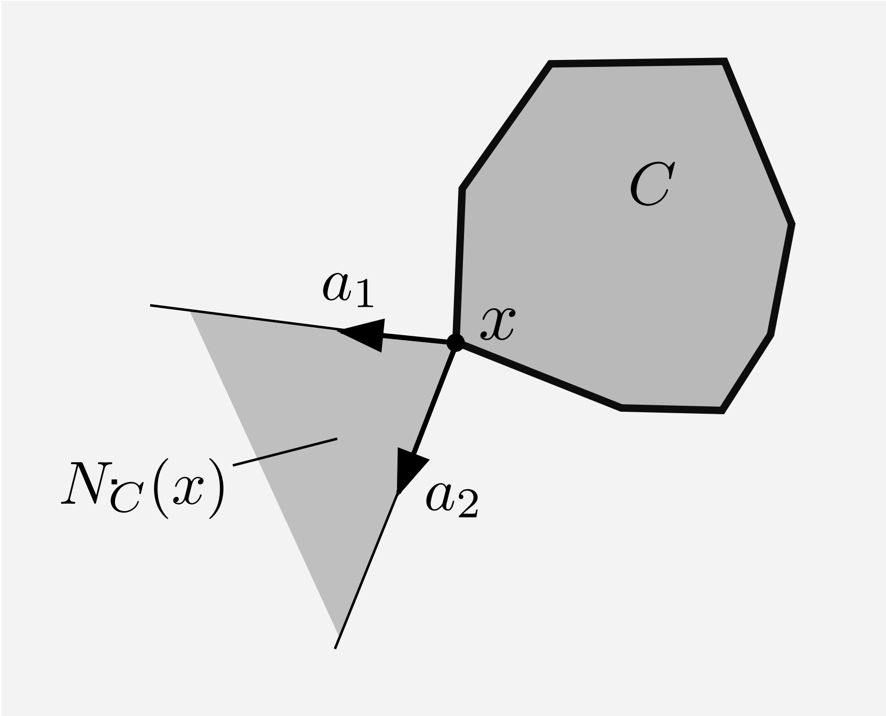
$$\delta_C(z) \geq \delta_C(x) + g'(z - x), \qquad \forall\ z \in C,$$

or equivalently $g'(z - x) \leq 0$ for all $z \in C$. Thus $\partial\delta_C(x)$ is the *normal cone of* C *at* x, denoted $N_C(x)$:

$$N_C(x) = \big\{ g \mid g'(z - x) \leq 0, \forall\ z \in C \big\}.$$



EXAMPLE: POLYHEDRAL CASE



- For the case of a polyhedral set

$$C = \{x \mid a_i'x \le b_i,\ i = 1, \ldots, m\},$$

we have

$$N_C(x) = \begin{cases} \{0\} & \text{if } x \in \text{int}(C), \\ \text{cone}\big(\{a_i \mid a_i'x = b_i\}\big) & \text{if } x \notin \text{int}(C). \end{cases}$$

- **Proof:** Given x, disregard inequalities with $a_i'x < b_i$, and translate C to move x to 0, so it becomes a cone. The polar cone is $N_C(x)$.

FENCHEL INEQUALITY

• Let $f : \Re^n \mapsto (-\infty, \infty]$ be proper convex and let $f^\star$ be its conjugate. Using the definition of conjugacy, we have *Fenchel's inequality*:

$$x'y \le f(x) + f^\star(y), \qquad \forall\ x \in \Re^n,\ y \in \Re^n.$$

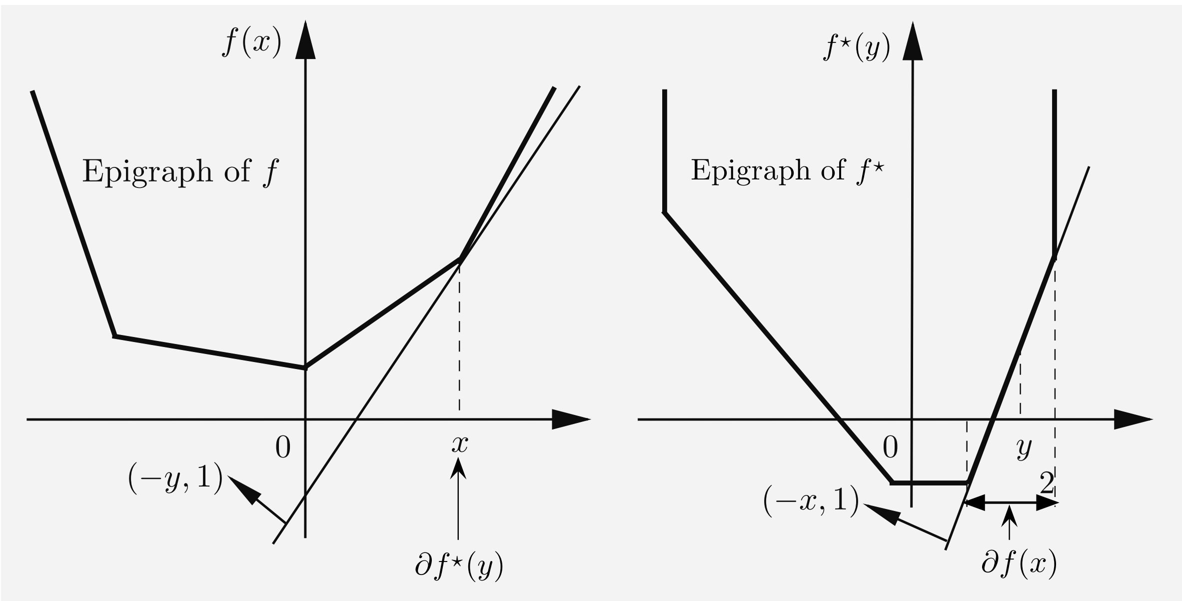
• **Conjugate Subgradient Theorem:** The following two relations are equivalent for a pair of vectors (x, y):

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

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

If f is closed, (i) and (ii) are equivalent to

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



MINIMA OF CONVEX FUNCTIONS

- **Application:** Let f be closed proper convex and let X^* be the set of minima of f over $\Re^n$. Then:

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

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

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

Proof: (a) We have $x^* \in X^*$ iff $f(x) \geq f(x^*)$ for all $x \in \Re^n$. So

$$x^* \in X^* \qquad \text{iff} \qquad 0 \in \partial f(x^*) \qquad \text{iff} \qquad x^* \in \partial f^\star(0)$$

where:

- 1st relation follows from the subgradient inequality
- 2nd relation follows from the conjugate subgradient theorem

(b) $\partial f^\star(0)$ is nonempty if $0 \in \text{ri}\big(\text{dom}(f^\star)\big)$.

(c) $\partial f^\star(0)$ is nonempty and compact if and only if $0 \in \text{int}\big(\text{dom}(f^\star)\big)$. **Q.E.D.**

SENSITIVITY INTERPRETATION

- Consider MC/MC for the case $M = \text{epi}(p)$.
- Dual function is

$$q(\mu) = \inf_{u \in \Re^m} \{p(u) + \mu' u\} = -p^\star(-\mu),$$

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

- Assume p is proper convex and strong duality holds, so $p(0) = w^* = q^* = \sup_{\mu \in \Re^m} \{-p^\star(-\mu)\}$. Let Q^* be the set of dual optimal solutions,

$$Q^* = \{\mu^* \mid p(0) + p^\star(-\mu^*) = 0\}.$$

From Conjugate Subgradient Theorem, $\mu^* \in Q^*$ if and only if $-\mu^* \in \partial p(0)$, i.e., $Q^* = -\partial p(0)$.

- If p is convex and differentiable at 0, $-\nabla p(0)$ is equal to the unique dual optimal solution μ^*.
- **Constrained optimization example:**

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

If p is convex and differentiable,

$$\mu_j^* = -\frac{\partial p(0)}{\partial u_j}, \qquad j = 1, \ldots, r.$$

EXAMPLE: SUBDIFF. OF SUPPORT FUNCTION

- Consider the support function $\sigma_X(y)$ of a set X. To calculate $\partial\sigma_X(\overline{y})$ at some $\overline{y}$, we introduce

$$r(y) = \sigma_X(y + \overline{y}), \qquad y \in \Re^n.$$

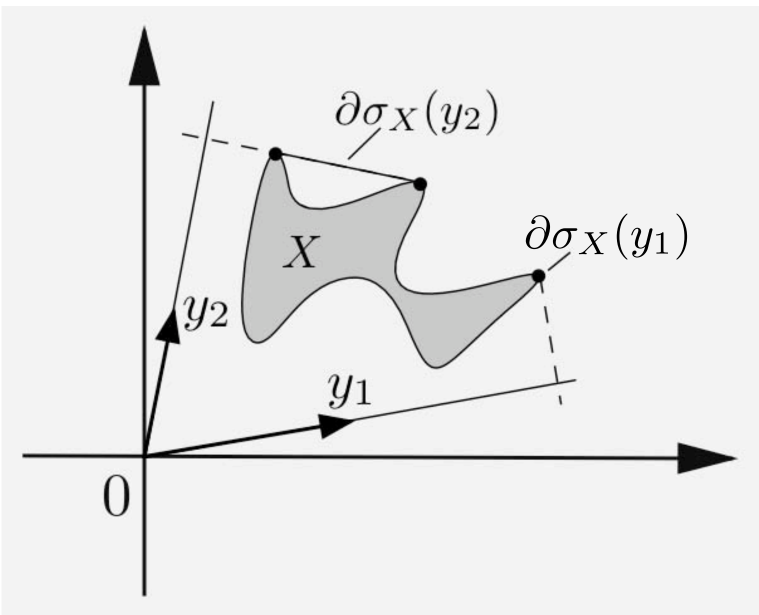
- We have $\partial\sigma_X(\overline{y}) = \partial r(0) = \arg\min_{x\in\Re^n} r^\star(x)$.
- We have $r^\star(x) = \sup_{y\in\Re^n}\{y'x - r(y)\}$, or

$$r^\star(x) = \sup_{y\in\Re^n}\{y'x - \sigma_X(y + \overline{y})\} = \delta(x) - \overline{y}'x,$$

where δ is the indicator function of $\mathrm{cl}\big(\mathrm{conv}(X)\big)$.

- Hence $\partial\sigma_X(\overline{y}) = \arg\min_{x\in\Re^n}\big\{\delta(x) - \overline{y}'x\big\}$, or

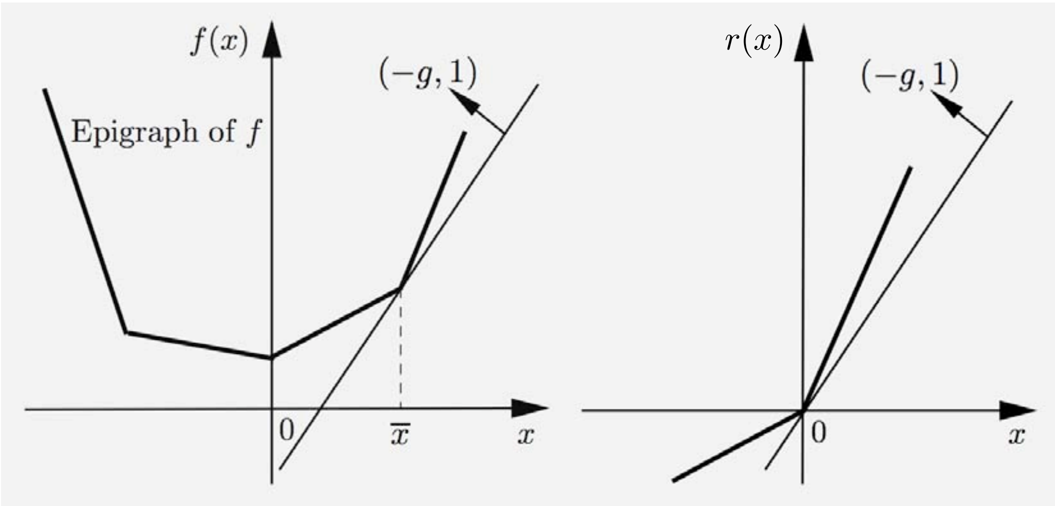
$$\partial\sigma_X(\overline{y}) = \arg\max_{x\in\mathrm{cl}(\mathrm{conv}(X))} \overline{y}'x$$



EXAMPLE: SUBDIFF. OF POLYHEDRAL FN

- Let

$$f(x) = \max\{a_1'x + b_1, \ldots, a_r'x + b_r\}.$$



- For a fixed $\overline{x} \in \Re^n$, consider

$$A_{\overline{x}} = \big\{ j \mid a_j'\overline{x} + b_j = f(\overline{x}) \big\}$$

and the function $r(x) = \max\big\{a_j'x \mid j \in A_{\overline{x}}\big\}$.

- It can be seen that $\partial f(\overline{x}) = \partial r(0)$.
- Since r is the support function of the finite set $\{a_j \mid j \in A_{\overline{x}}\}$, we see that

$$\partial f(\overline{x}) = \partial r(0) = \text{conv}\big(\{a_j \mid j \in A_{\overline{x}}\}\big)$$

CHAIN RULE

- Let $f : \Re^m \mapsto (-\infty, \infty]$ be convex, and A be a matrix. Consider $F(x) = f(Ax)$ and assume that F is proper. If either f is polyhedral or else $\text{Range}(A) \cap \text{ri}(\text{dom}(f)) \neq \emptyset$, then

$$\partial F(x) = A'\partial f(Ax), \qquad \forall\, x \in \Re^n.$$

Proof: Showing $\partial F(x) \supset A'\partial f(Ax)$ is simple and does not require the relative interior assumption. For the reverse inclusion, let $d \in \partial F(x)$ so $F(z) \geq F(x) + (z-x)'d \geq 0$ or $f(Az) - z'd \geq f(Ax) - x'd$ for all z, so (Ax, x) solves

$$\begin{aligned} &\text{minimize} \quad f(y) - z'd \\ &\text{subject to} \;\; y \in \text{dom}(f), \quad Az = y. \end{aligned}$$

If $R(A) \cap \text{ri}(\text{dom}(f)) \neq \emptyset$, by strong duality theorem, there is a dual optimal solution λ, such that

$$(Ax, x) \in \arg \min_{y \in \Re^m,\, z \in \Re^n} \big\{ f(y) - z'd + \lambda'(Az - y) \big\}$$

Since the min over z is unconstrained, we have $d = A'\lambda$, so $Ax \in \arg\min_{y \in \Re^m} \big\{ f(y) - \lambda' y \big\}$, or

$$f(y) \geq f(Ax) + \lambda'(y - Ax), \qquad \forall\, y \in \Re^m.$$

Hence $\lambda \in \partial f(Ax)$, so that $d = A'\lambda \in A'\partial f(Ax)$. It follows that $\partial F(x) \subset A'\partial f(Ax)$. In the polyhedral case, $\text{dom}(f)$ is polyhedral. **Q.E.D.**

SUM OF FUNCTIONS

- Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be proper convex functions, and let

$$F = f_1 + \cdots + f_m.$$

- Assume that $\cap_{1=1}^m \mathrm{ri}\big(\mathrm{dom}(f_i)\big) \neq \emptyset$.
- Then

$$\partial F(x) = \partial f_1(x) + \cdots + \partial f_m(x), \qquad \forall\, x \in \Re^n.$$

Proof: We can write F in the form $F(x) = f(Ax)$, where A is the matrix defined by $Ax = (x, \ldots, x)$, and $f : \Re^{mn} \mapsto (-\infty, \infty]$ is the function

$$f(x_1, \ldots, x_m) = f_1(x_1) + \cdots + f_m(x_m).$$

Use the proof of the chain rule.

- **Extension:** If for some k, the functions f_i, $i = 1, \ldots, k$, are polyhedral, it is sufficient to assume

$$\Big(\cap_{i=1}^k \mathrm{dom}(f_i) \Big) \cap \Big(\cap_{i=k+1}^m \mathrm{ri}\big(\mathrm{dom}(f_i)\big) \Big) \neq \emptyset.$$

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.,

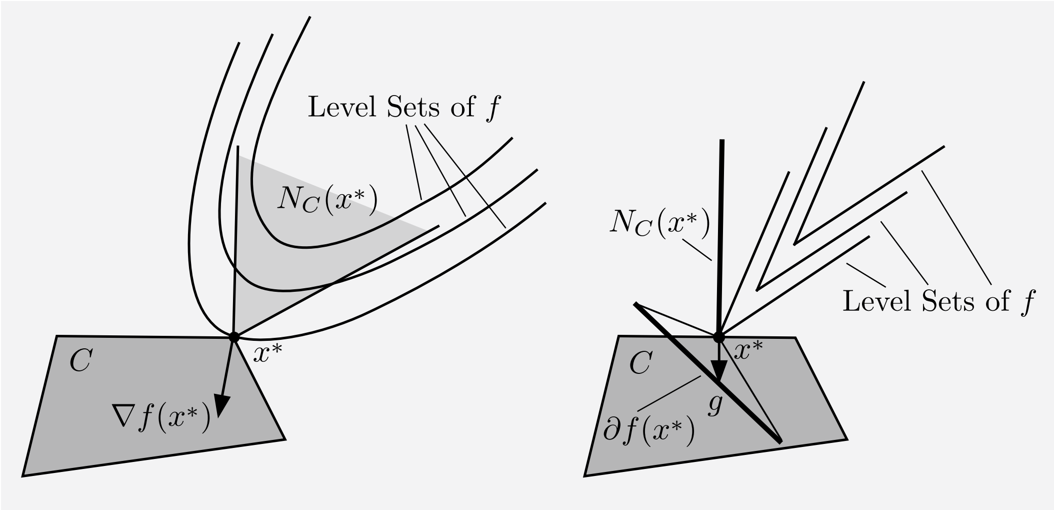
$$g'(x - x^*) \geq 0, \qquad \forall\, x \in X.$$

Proof: x^* minimizes

$$F(x) = f(x) + \delta_X(x)$$

if and only if $0 \in \partial F(x^*)$. Use the formula for subdifferential of sum. **Q.E.D.**

ILLUSTRATION OF OPTIMALITY CONDITION



• In the figure on the left, f is differentiable and the condition is that

$$-\nabla f(x^*) \in N_C(x^*),$$

which is equivalent to

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

• In the figure on the right, f is nondifferentiable, and the condition is that

$$-g \in N_C(x^*) \quad \text{for some } g \in \partial f(x^*).$$

LECTURE 13

LECTURE OUTLINE

- Problem Structures
  - Separable problems
  - Integer/discrete problems – Branch-and-bound
  - Large sum problems
  - Problems with many constraints
- Conic Programming
  - Second Order Cone Programming
  - Semidefinite Programming

SEPARABLE PROBLEMS

- Consider the problem

$$\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i)$$

$$\text{s. t.} \ \sum_{i=1}^{m} g_{ji}(x_i) \le 0, \quad j = 1, \ldots, r, \quad x_i \in X_i, \quad \forall\, i$$

where $f_i : \Re^{n_i} \mapsto \Re$ and $g_{ji} : \Re^{n_i} \mapsto \Re$ are given functions, and X_i are given subsets of $\Re^{n_i}$.

- Form the dual problem

$$\text{maximize} \ \sum_{i=1}^{m} q_i(\mu) \equiv \sum_{i=1}^{m} \inf_{x_i \in X_i} \left\{ f_i(x_i) + \sum_{j=1}^{r} \mu_j g_{ji}(x_i) \right\}$$

subject to $\mu \ge 0$

- **Important point:** The calculation of the dual function has been **decomposed** into n simpler minimizations. Moreover, the calculation of dual subgradients is **a byproduct of these minimizations** (this will be discussed later)

- **Another important point:** If X_i is a discrete set (e.g., $X_i = \{0, 1\}$), the dual optimal value is a lower bound to the optimal primal value. It is still useful in a branch-and-bound scheme.

LARGE SUM PROBLEMS

- Consider cost function of the form

$$f(x) = \sum_{i=1}^{m} f_i(x), \qquad m \text{ is very large,}$$

where $f_i : \Re^n \mapsto \Re$ are convex. Some examples:

- **Dual cost of a separable problem.**
- **Data analysis/machine learning:** x is parameter vector of a model; each f_i corresponds to error between data and output of the model.
  - Least squares problems (f_i quadratic).
  - ℓ_1-regularization (least squares plus ℓ_1 penalty):

$$\min_x \sum_{j=1}^{m} (a_j' x - b_j)^2 + \gamma \sum_{i=1}^{n} |x_i|$$

The nondifferentiable penalty tends to set a large number of components of x to 0.

- **Min of an expected value** $E\big\{F(x, w)\big\}$, where w is a random variable taking a finite but very large number of values w_i, $i = 1, \ldots, m$, with corresponding probabilities π_i.
- **Stochastic programming:**

$$\min_x \left[F_1(x) + E_w\{\min_y F_2(x, y, w)\} \right]$$

- Special methods, called **incremental** apply.

PROBLEMS WITH MANY CONSTRAINTS

- Problems of the form

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \ \ a_j'x \le b_j, \quad j = 1, \ldots, r, \end{aligned}$$

where r: very large.

- One possibility is a *penalty function approach*: Replace problem with

$$\min_{x \in \Re^n} f(x) + c \sum_{j=1}^{r} P(a_j'x - b_j)$$

where $P(\cdot)$ is a scalar penalty function satisfying $P(t) = 0$ if $t \le 0$, and $P(t) > 0$ if $t > 0$, and c is a positive penalty parameter.

- Examples:
  - The quadratic penalty $P(t) = \big(\max\{0, t\}\big)^2$.
  - The nondifferentiable penalty $P(t) = \max\{0, t\}$.

- Another possibility: Initially discard some of the constraints, solve a less constrained problem, and later reintroduce constraints that seem to be violated at the optimum (*outer approximation*).

- Also *inner approximation* of the constraint set.

CONIC PROBLEMS

- A conic problem is to minimize a convex function $f : \Re^n \mapsto (-\infty, \infty]$ subject to a cone constraint.

- The most useful/popular special cases:
  - Linear-conic programming
  - Second order cone programming
  - Semidefinite programming

involve minimization of a linear function over the intersection of an affine set and a cone.

- Can be analyzed as a special case of Fenchel duality.

- There are many interesting applications of conic problems, including in discrete optimization.

PROBLEM RANKING IN INCREASING PRACTICAL DIFFICULTY

- Linear and (convex) quadratic programming.
  - Favorable special cases (e.g., network flows).
- **Second order cone programming.**
- **Semidefinite programming.**
- Convex programming.
  - Favorable special cases (e.g., network flows, monotropic programming, geometric programming).
- Nonlinear/nonconvex/continuous programming.
  - Favorable special cases (e.g., twice differentiable, quasi-convex programming).
  - Unconstrained.
  - Constrained.
- Discrete optimization/Integer programming
  - Favorable special cases.

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

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

The conjugates are

$$f_1^\star(\lambda) = \sup_{x \in \Re^n} \big\{\lambda' x - f(x)\big\},\ f_2^\star(\lambda) = \sup_{x \in C} \lambda' x = \begin{cases} 0 & \text{if } \lambda \in C^*, \\ \infty & \text{if } \lambda \notin C^*, \end{cases}$$

where $C^* = \{\lambda \mid \lambda' x \leq 0,\ \forall\ x \in C\}$ is the polar cone of C.

- The dual problem is

$$\begin{aligned} &\text{minimize} \quad f^\star(\lambda) \\ &\text{subject to} \ \ \lambda \in \hat{C}, \end{aligned}$$

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

$$\hat{C} = \{\lambda \mid \lambda' x \geq 0,\ \forall\ x \in C\}.$$

$\hat{C} = -C^*$ is called the *dual* cone.

LINEAR-CONIC PROBLEMS

- Let f be affine, $f(x) = c'x$, with $\text{dom}(f)$ being an affine set, $\text{dom}(f) = b + S$, where S is a subspace.

- The primal problem is

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

- The conjugate is

$$\begin{aligned} f^\star(\lambda) &= \sup_{x-b\in S} (\lambda - c)'x = \sup_{y\in S} (\lambda - c)'(y + b) \\ &= \begin{cases} (\lambda - c)'b & \text{if } \lambda - c \in S^\perp, \\ \infty & \text{if } \lambda - c \notin S^\perp, \end{cases} \end{aligned}$$

so the dual problem can be written as

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

- The primal and dual have the same form.

- If C is closed, the dual of the dual yields the primal.

SPECIAL LINEAR-CONIC FORMS

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

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

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

- For the first relation, let $\overline{x}$ be such that $A\overline{x} = b$, and write the problem on the left as

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad x - \overline{x} \in \text{N}(A), \quad x \in C \end{aligned}$$

- The dual conic problem is

$$\begin{aligned} &\text{minimize} \quad \overline{x}'\mu \\ &\text{subject to} \quad \mu - c \in \text{N}(A)^\perp, \quad \mu \in \hat{C}. \end{aligned}$$

- Using $\text{N}(A)^\perp = \text{Ra}(A')$, write the constraints as $c - \mu \in -\text{Ra}(A') = \text{Ra}(A')$, $\mu \in \hat{C}$, or

$$c - \mu = A'\lambda, \qquad \mu \in \hat{C}, \qquad \text{for some } \lambda \in \Re^m.$$

- Change variables $\mu = c - A'\lambda$, write the dual as

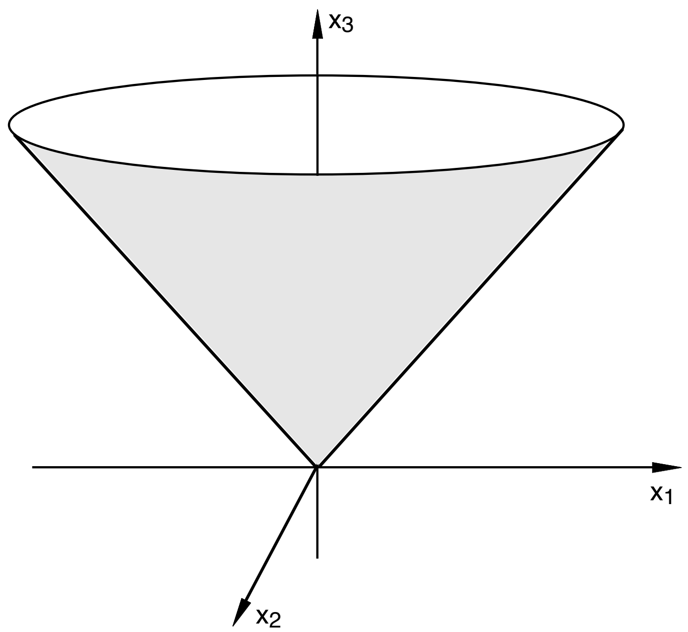
$$\begin{aligned} &\text{minimize} \quad \overline{x}'(c - A'\lambda) \\ &\text{subject to} \quad c - A'\lambda \in \hat{C} \end{aligned}$$

discard the constant $\overline{x}'c$, use the fact $A\overline{x} = b$, and change from min to max.

SOME EXAMPLES

- **Nonnegative Orthant:** $C = \{x \mid x \geq 0\}$.
- **The Second Order Cone:** Let

$$C = \left\{(x_1, \ldots, x_n) \mid x_n \geq \sqrt{x_1^2 + \cdots + x_{n-1}^2}\right\}$$



- **The Positive Semidefinite Cone:** Consider the space of symmetric $n \times n$ matrices, viewed as the space $\Re^{n^2}$ with the inner product

$$< X, Y >= \text{trace}(XY) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} y_{ij}$$

Let C be the cone of matrices that are positive semidefinite.

- All these are *self-dual*, i.e., $C = -C^* = \hat{C}$.

SECOND ORDER CONE PROGRAMMING

- Second order cone programming is the linear-conic problem

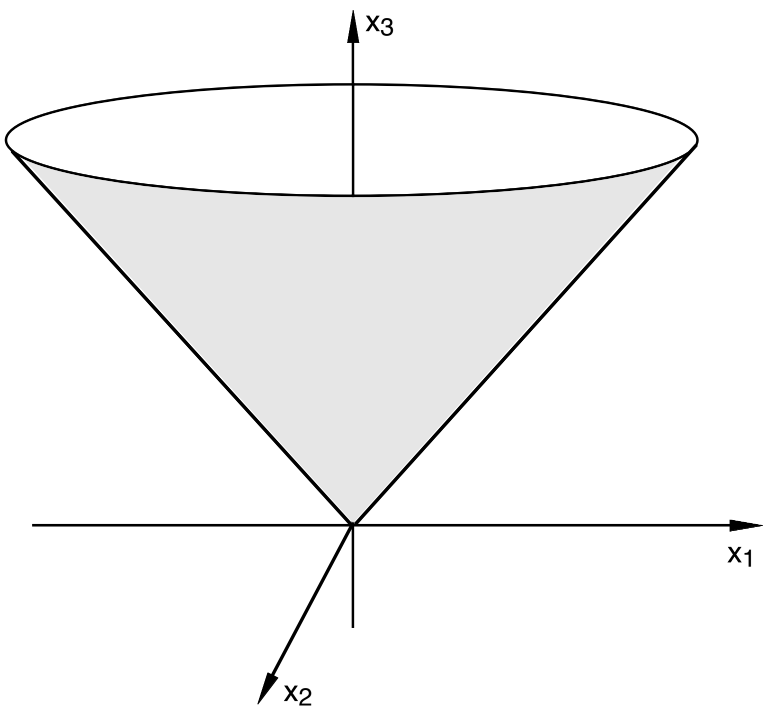
$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \;\; A_i x - b_i \in C_i, \;\; i = 1, \ldots, m, \end{aligned}$$

where c, b_i are vectors, A_i are matrices, b_i is a vector in $\Re^{n_i}$, and

$$C_i : \;\text{the second order cone of } \Re^{n_i}$$

- The cone here is

$$C = C_1 \times \cdots \times C_m$$



SECOND ORDER CONE DUALITY

- Using the generic special duality form

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

and self duality of C, the dual problem is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} b_i'\lambda_i \\ &\text{subject to} \quad \sum_{i=1}^{m} A_i'\lambda_i = c, \quad \lambda_i \in C_i,\ i = 1,\ldots,m, \end{aligned}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$.

- The duality theory is no more favorable than the one for linear-conic problems.

- There is no duality gap if there exists a feasible solution in the interior of the 2nd order cones C_i.

- Generally, 2nd order cone problems can be recognized from the presence of norm or convex quadratic functions in the cost or the constraint functions.

- There are many applications.

LECTURE 14

LECTURE OUTLINE

- Conic programming
- Semidefinite programming
- Exact penalty functions
- Descent methods for convex/nondifferentiable optimization
- Steepest descent method

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

- Second order cone programming:

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad A_i x - b_i \in C_i,\ i = 1, \ldots, m, \end{aligned}$$

where c, b_i are vectors, A_i are matrices, b_i is a vector in $\Re^{n_i}$, and

$$C_i : \text{ the second order cone of } \Re^{n_i}$$

- The cone here is $C = C_1 \times \cdots \times C_m$
- The dual problem is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^m b_i' \lambda_i \\ &\text{subject to} \quad \sum_{i=1}^m A_i' \lambda_i = c, \quad \lambda_i \in C_i,\ i = 1, \ldots, m, \end{aligned}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$.

EXAMPLE: ROBUST LINEAR PROGRAMMING

$$\begin{aligned} &\text{minimize } c'x \\ &\text{subject to } a_j'x \le b_j, \quad \forall\, (a_j, b_j) \in T_j, \quad j = 1, \ldots, r, \end{aligned}$$

where $c \in \Re^n$, and T_j is a given subset of $\Re^{n+1}$.

- We convert the problem to the equivalent form

$$\begin{aligned} &\text{minimize } c'x \\ &\text{subject to } g_j(x) \le 0, \qquad j = 1, \ldots, r, \end{aligned}$$

where $g_j(x) = \sup_{(a_j, b_j) \in T_j} \{a_j'x - b_j\}$.

- For special choice where T_j is an ellipsoid,

$$T_j = \big\{(\overline{a}_j + P_j u_j, \overline{b}_j + q_j' u_j) \mid \|u_j\| \le 1,\, u_j \in \Re^{n_j}\big\}$$

we can express $g_j(x) \le 0$ in terms of a SOC:

$$\begin{aligned} g_j(x) &= \sup_{\|u_j\| \le 1} \big\{(\overline{a}_j + P_j u_j)'x - (\overline{b}_j + q_j' u_j)\big\} \\ &= \sup_{\|u_j\| \le 1} (P_j'x - q_j)'u_j + \overline{a}_j'x - \overline{b}_j, \\ &= \|P_j'x - q_j\| + \overline{a}_j'x - \overline{b}_j. \end{aligned}$$

Thus, $g_j(x) \le 0$ iff $(P_j'x - q_j, \overline{b}_j - \overline{a}_j'x) \in C_j$, where C_j is the SOC of $\Re^{n_j+1}$.

SEMIDEFINITE PROGRAMMING

- Consider the symmetric $n \times n$ matrices. Inner product $< X, Y >= \text{trace}(XY) = \sum_{i,j=1}^{n} x_{ij} y_{ij}$.
- Let C be the cone of pos. semidefinite matrices.
- C is self-dual, and its interior is the set of positive definite matrices.
- Fix symmetric matrices D, $A_1, \ldots, A_m$, and vectors $b_1, \ldots, b_m$, and consider

$$\begin{array}{ll} \text{minimize} & < D, X > \\ \text{subject to} & < A_i, X >= b_i, \quad i = 1, \ldots, m, \quad X \in C \end{array}$$

- Viewing this as a linear-conic problem (the first special form), the dual problem (using also self-duality of C) is

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{m} b_i \lambda_i \\ \text{subject to} & D - (\lambda_1 A_1 + \cdots + \lambda_m A_m) \in C \end{array}$$

- There is no duality gap if there exists primal feasible solution that is pos. definite, or there exists $\overline{\lambda}$ such that $D - (\overline{\lambda}_1 A_1 + \cdots + \overline{\lambda}_m A_m)$ is pos. definite.

EXAMPLE: MINIMIZE THE MAXIMUM EIGENVALUE

- Given $n \times n$ symmetric matrix $M(\lambda)$, depending on a parameter vector λ, choose λ to minimize the maximum eigenvalue of $M(\lambda)$.
- We pose this problem as

$$\begin{array}{ll} \text{minimize} & z \\ \text{subject to} & \text{maximum eigenvalue of } M(\lambda) \leq z, \end{array}$$

or equivalently

$$\begin{array}{ll} \text{minimize} & z \\ \text{subject to} & zI - M(\lambda) \in C, \end{array}$$

where I is the $n \times n$ identity matrix, and C is the semidefinite cone.

- If $M(\lambda)$ is an affine function of λ,

$$M(\lambda) = D + \lambda_1 M_1 + \cdots + \lambda_m M_m,$$

the problem has the form of the dual semidefinite problem, with the optimization variables being $(z, \lambda_1, \ldots, \lambda_m)$.

EXAMPLE: LOWER BOUNDS FOR DISCRETE OPTIMIZATION

- Quadr. problem with quadr. equality constraints

$$\begin{array}{ll} \text{minimize} & x'Q_0x + a_0'x + b_0 \\ \text{subject to} & x'Q_ix + a_i'x + b_i = 0, \quad i = 1, \ldots, m, \end{array}$$

$Q_0, \ldots, Q_m$: symmetric (not necessarily ≥ 0).

- Can be used for discrete optimization. For example an integer constraint $x_i \in \{0, 1\}$ can be expressed by $x_i^2 - x_i = 0$.
- The dual function is

$$q(\lambda) = \inf_{x \in \Re^n} \big\{ x'Q(\lambda)x + a(\lambda)'x + b(\lambda) \big\},$$

where

$$Q(\lambda) = Q_0 + \sum_{i=1}^m \lambda_i Q_i,$$

$$a(\lambda) = a_0 + \sum_{i=1}^m \lambda_i a_i, \quad b(\lambda) = b_0 + \sum_{i=1}^m \lambda_i b_i$$

- It turns out that the dual problem is equivalent to a semidefinite program ...

EXACT PENALTY FUNCTIONS

- We use Fenchel duality to derive an equivalence between a constrained convex optimization problem, and a penalized problem that is less constrained or is entirely unconstrained.

- We consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X, \qquad g(x) \leq 0, \end{array}$$

where $g(x) = \big(g_1(x), \ldots, g_r(x)\big)$, X is a convex subset of $\Re^n$, and $f : \Re^n \to \Re$ and $g_j : \Re^n \to \Re$ are real-valued convex functions.

- We introduce a convex function $P : \Re^r \mapsto \Re$, called *penalty function*, which satisfies

$$P(u) = 0, \ \forall\, u \leq 0, \quad P(u) > 0, \ \text{if } u_i > 0 \text{ for some } i$$

- We consider solving, in place of the original, the "penalized" problem

$$\begin{array}{ll} \text{minimize} & f(x) + P\big(g(x)\big) \\ \text{subject to} & x \in X, \end{array}$$

FENCHEL DUALITY

- We have

$$\inf_{x\in X}\big\{f(x)+P\big(g(x)\big)\big\} = \inf_{u\in\Re^r}\big\{p(u)+P(u)\big\}$$

where $p(u) = \inf_{x\in X,\, g(x)\le u} f(x)$ is the primal function.

- Assume $-\infty < q^*$ and $f^* < \infty$ so that p is proper (in addition to being convex).

- By Fenchel duality

$$\inf_{u\in\Re^r}\big\{p(u)+P(u)\big\} = \sup_{\mu\ge 0}\big\{q(\mu)-Q(\mu)\big\},$$

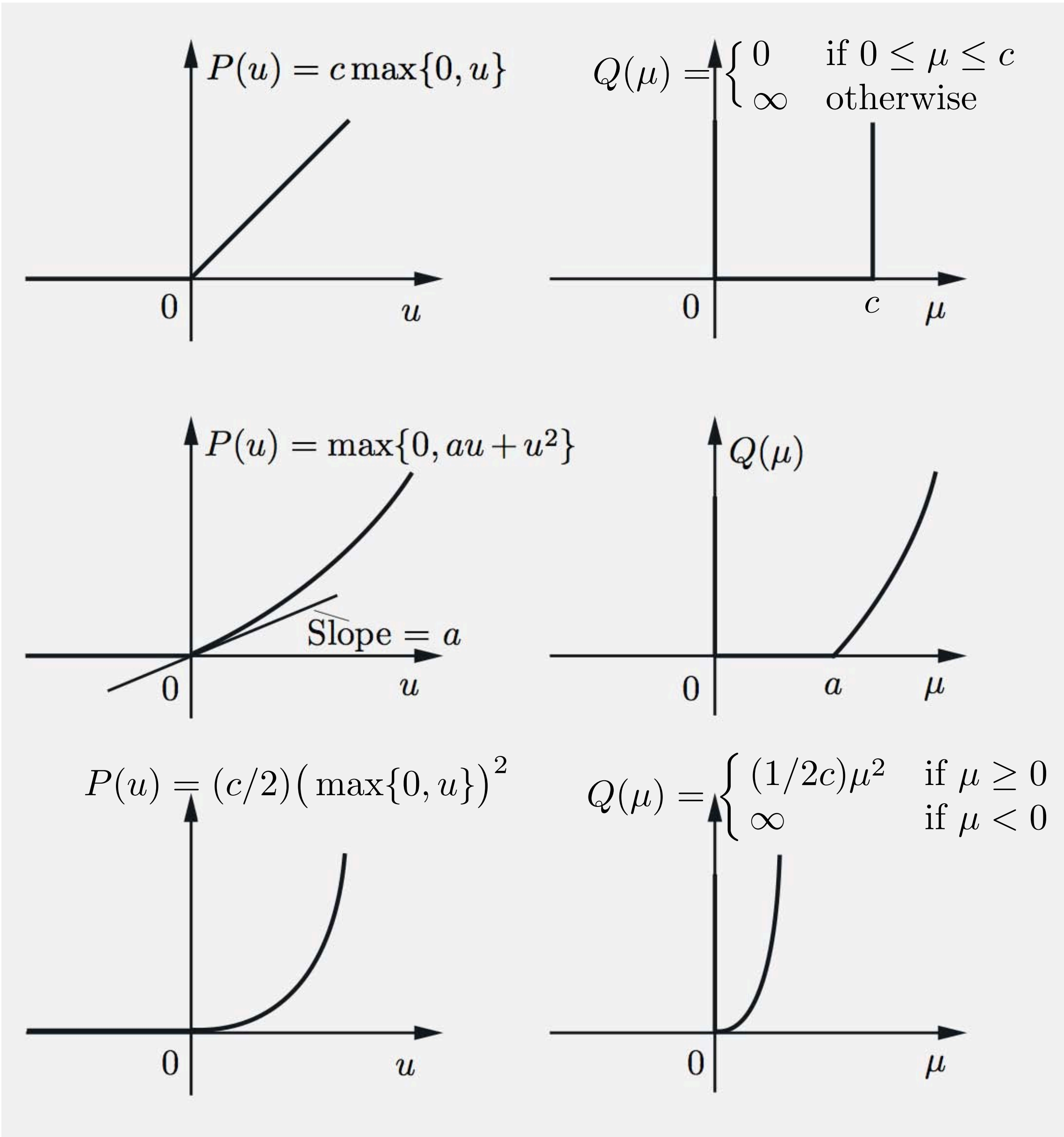
where for $\mu \ge 0$,

$$q(\mu) = \inf_{x\in X}\big\{f(x)+\mu' g(x)\big\}$$

is the dual function, and Q is the conjugate convex function of P:

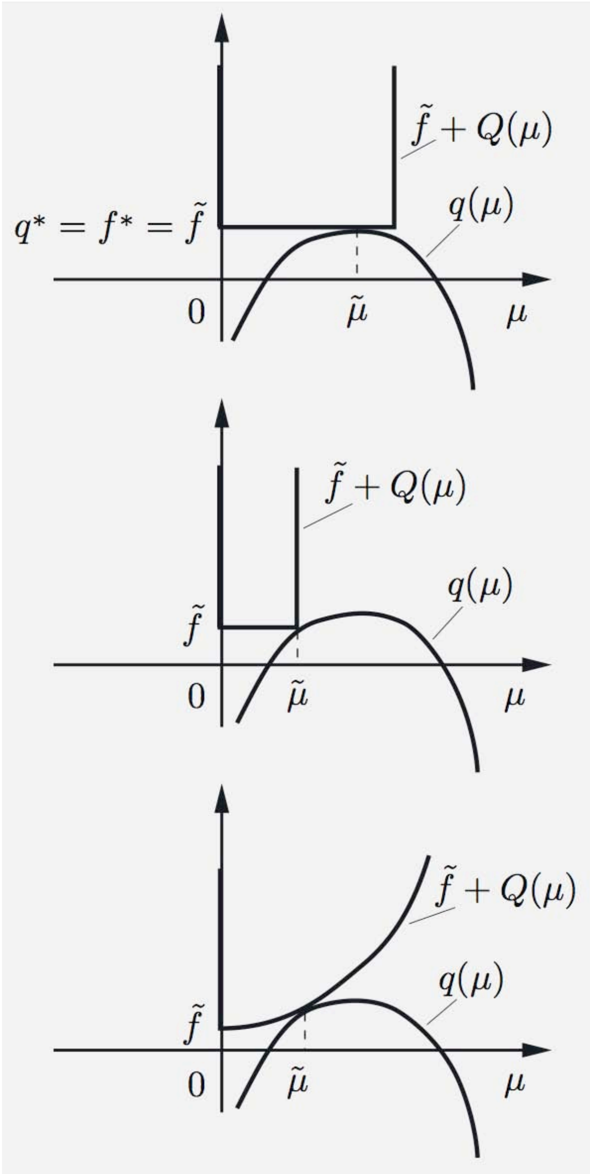
$$Q(\mu) = \sup_{u\in\Re^r}\big\{u'\mu - P(u)\big\}$$

PENALTY CONJUGATES



- **Important observation:** For Q to be flat for some $\mu > 0$, P must be nondifferentiable at 0.

FENCHEL DUALITY VIEW



- For the penalized and the original problem to have equal optimal values, Q must be "flat enough" so that some optimal dual solution μ^* minimizes Q, i.e., $0 \in \partial Q(\mu^*)$ or equivalently

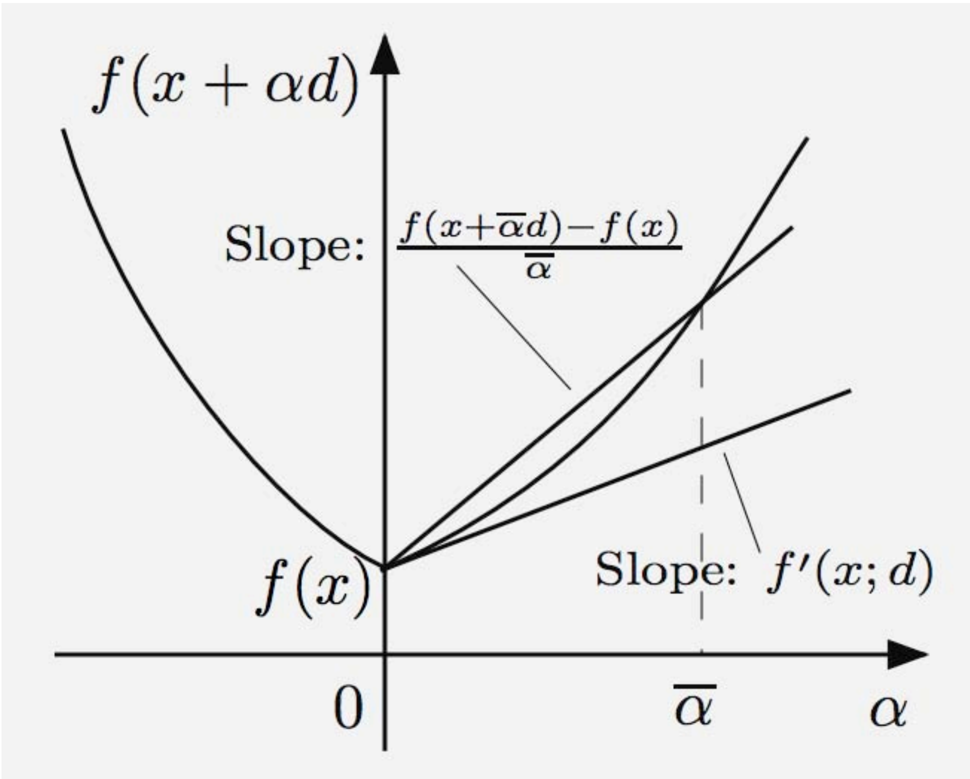
$$\mu^* \in \partial P(0)$$

- True if $P(u) = c\sum_{j=1}^{r} \max\{0, u_j\}$ with $c \geq \|\mu^*\|$ for some optimal dual solution μ^*.

DIRECTIONAL DERIVATIVES

- Directional derivative of a proper convex f:

$$f'(x;d) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha d) - f(x)}{\alpha}, \quad x \in \mathrm{dom}(f),\ d \in \Re^n$$



- The ratio

$$\frac{f(x+\alpha d) - f(x)}{\alpha}$$

is monotonically nonincreasing as $\alpha \downarrow 0$ and converges to $f'(x;d)$.

- For all $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, $f'(x;\cdot)$ is the support function of $\partial f(x)$.

STEEPEST DESCENT DIRECTION

- Consider unconstrained minimization of convex $f : \Re^n \mapsto \Re$.
- A descent direction d at x is one for which $f'(x; d) < 0$, where

$$f'(x; d) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha d) - f(x)}{\alpha} = \sup_{g \in \partial f(x)} d'g$$

is the directional derivative.

- Can decrease f by moving from x along descent direction d by small stepsize α.
- Direction of steepest descent solves the problem

$$\begin{aligned} &\text{minimize} \quad f'(x; d) \\ &\text{subject to} \quad \|d\| \le 1 \end{aligned}$$

- **Interesting fact:** The steepest descent direction is $-g^*$, where g^* is the vector of minimum norm in $\partial f(x)$:

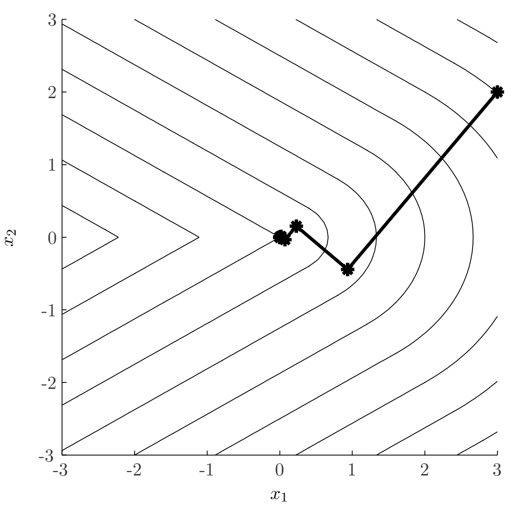
$$\begin{aligned} \min_{\|d\| \le 1} f'(x; d) &= \min_{\|d\| \le 1} \max_{g \in \partial f(x)} d'g = \max_{g \in \partial f(x)} \min_{\|d\| \le 1} d'g \\ &= \max_{g \in \partial f(x)} \big(-\|g\|\big) = -\min_{g \in \partial f(x)} \|g\| \end{aligned}$$

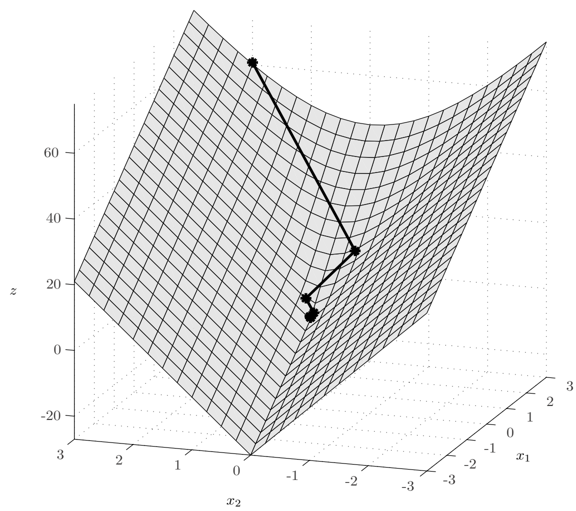
STEEPEST DESCENT METHOD

- Start with any $x_0 \in \Re^n$.
- For $k \geq 0$, calculate $-g_k$, the steepest descent direction at x_k and set

$$x_{k+1} = x_k - \alpha_k g_k$$

- **Difficulties:**
  - Need the entire $\partial f(x_k)$ to compute g_k.
  - Serious convergence issues due to discontinuity of $\partial f(x)$ (the method has no clue that $\partial f(x)$ may change drastically nearby).
- Example with α_k determined by minimization along $-g_k$: $\{x_k\}$ converges to nonoptimal point.





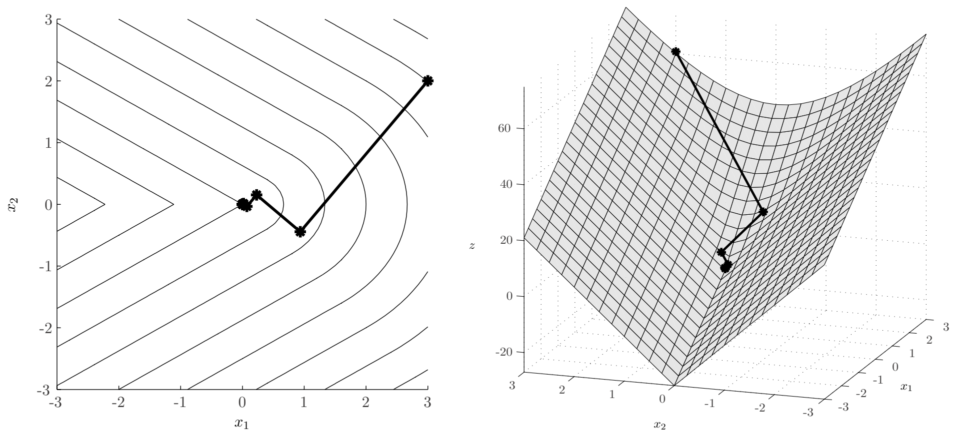
LECTURE 15

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 $x \in \text{dom}(f)$, 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 $\min_{x \in X,\, g(x) \leq 0} f(x)$:

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

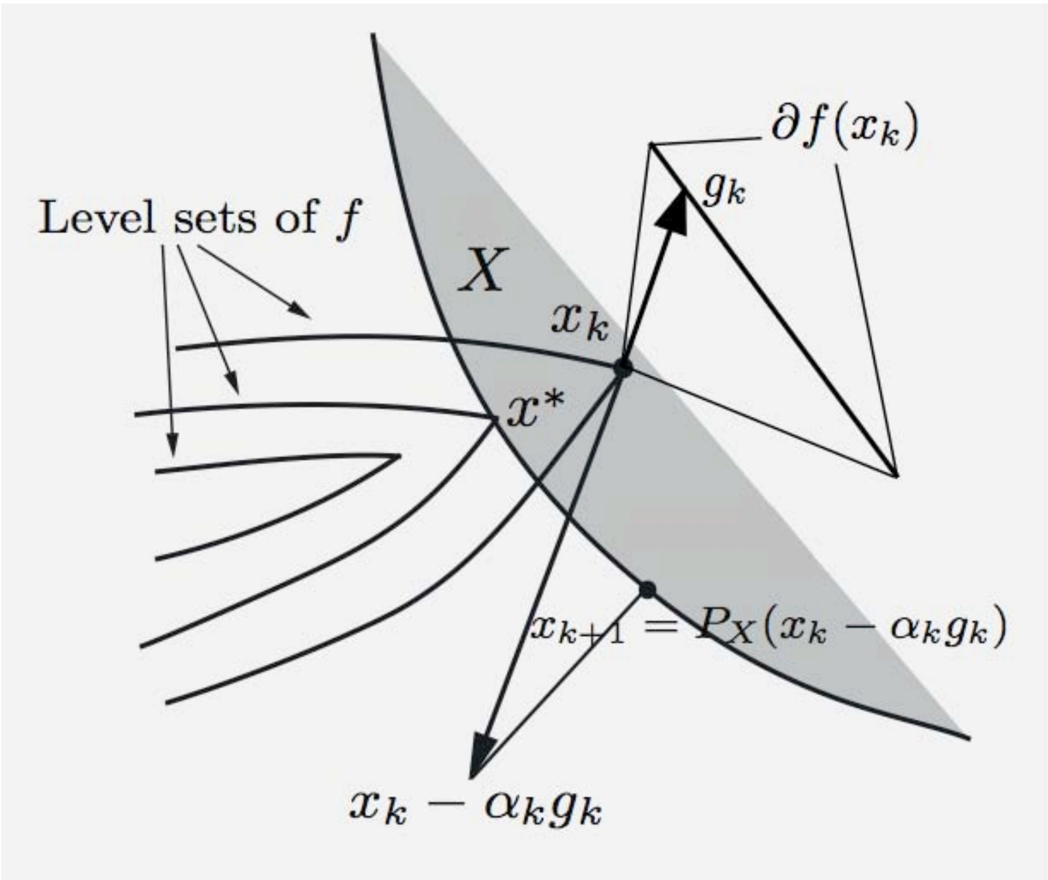
or $\min_{\mu \geq 0} F(\mu)$, where $F(-\mu) \equiv -q(\mu)$.

ALGORITHMS: SUBGRADIENT METHOD

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

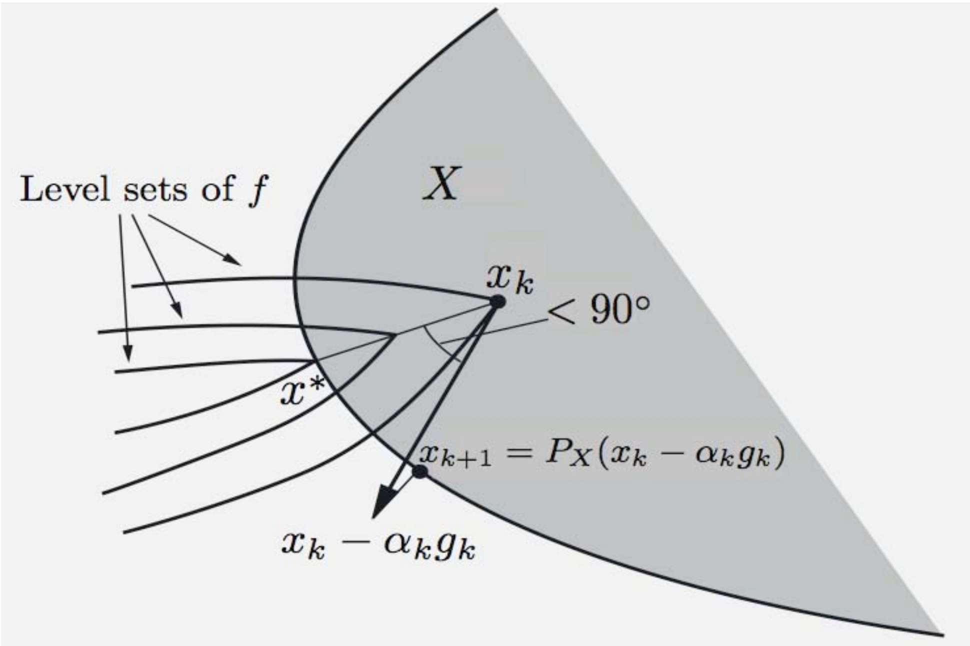
$$x_{k+1} = P_X(x_k - \alpha_k g_k),$$

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**

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

Use the projection theorem to write

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

from which $\big(P_X(y) - P_X(x)\big)'\big(x - P_X(x)\big) \leq 0$. Similarly, $\big(P_X(x) - P_X(y)\big)'\big(y - P_X(y)\big) \leq 0$.
Adding and using the Schwarz inequality,

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

Q.E.D.

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

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

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,

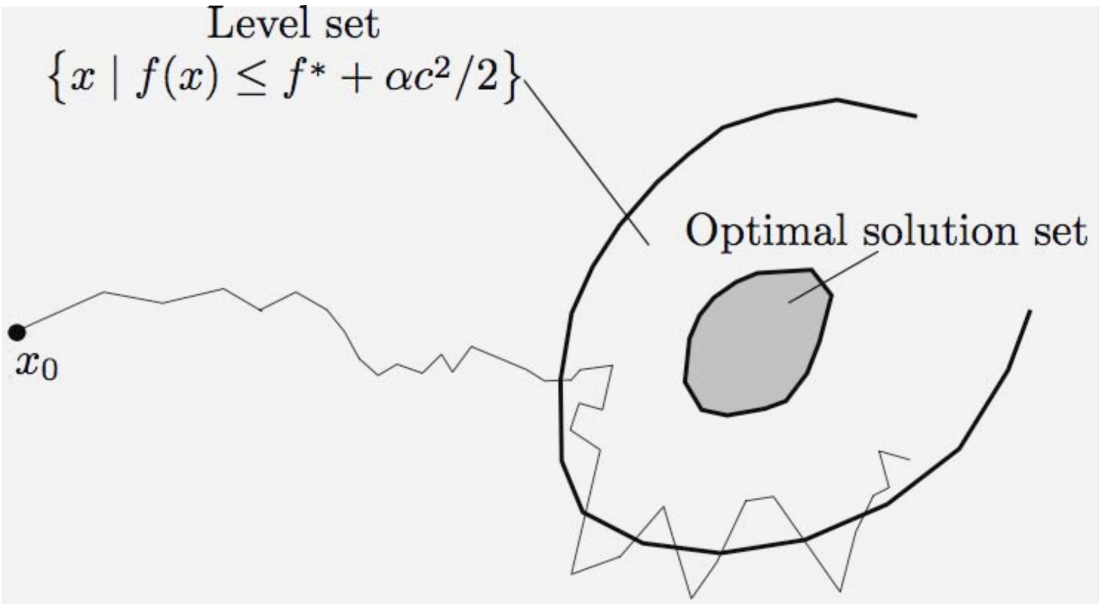
$$\|x_{k+1}-x^*\|^2 \le \|x_k-x^*\|^2-2\alpha\big(f(x_k)-f(x^*)\big)+\alpha^2c^2$$

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 ...

LECTURE 16

LECTURE OUTLINE

- Approximate subgradient methods
- Approximation methods
- Cutting plane methods

APPROXIMATE SUBGRADIENT METHODS

- Consider minimization of

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

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

- To compute subgradients of f at $x \in \text{dom}(f)$, we find $z_x \in Z$ attaining the supremum above. Then

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

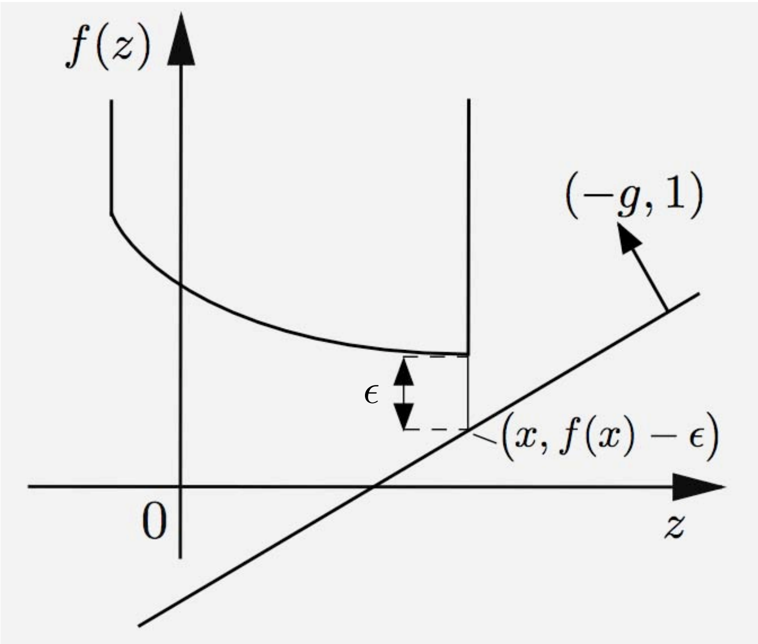
- **Potential difficulty:** For subgradient method, we need to solve exactly the above maximization over $z \in Z$.

- We consider methods that use "approximate" subgradients that can be computed more easily.

ϵ-SUBDIFFERENTIAL

- Fot a proper convex $f : \Re^n \mapsto (-\infty, \infty]$ and $\epsilon > 0$, we say that a vector g is an *ϵ-subgradient* of f at a point $x \in \text{dom}(f)$ if

$$f(z) \geq f(x) + (z - x)'g - \epsilon, \qquad \forall\, z \in \Re^n$$



- The *ϵ-subdifferential* $\partial_\epsilon f(x)$ is the set of all ϵ-subgradients of f at x. By convention, $\partial_\epsilon f(x) = \emptyset$ for $x \notin \text{dom}(f)$.

- We have $\cap_{\epsilon \downarrow 0} \partial_\epsilon f(x) = \partial f(x)$ and

$$\partial_{\epsilon_1} f(x) \subset \partial_{\epsilon_2} f(x) \quad \text{ if } 0 < \epsilon_1 < \epsilon_2$$

CALCULATION OF AN ϵ-SUBGRADIENT

- Consider minimization of

$$f(x) = \sup_{z \in Z} \phi(x, z), \qquad (1)$$

where $x \in \Re^n$, $z \in \Re^m$, Z is a subset of $\Re^m$, and $\phi : \Re^n \times \Re^m \mapsto (-\infty, \infty]$ is a function such that $\phi(\cdot, z)$ is convex and closed for each $z \in Z$.

- How to calculate ϵ-subgradient at $x \in \text{dom}(f)$?

- Let $z_x \in Z$ attain the supremum within $\epsilon \geq 0$ in Eq. (1), and let g_x be some subgradient of the convex function $\phi(\cdot, z_x)$.

- For all $y \in \Re^n$, using the subgradient inequality,

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

i.e., g_x is an ϵ-subgradient of f at x, so

$$\phi(x, z_x) \geq \sup_{z \in Z} \phi(x, z) - \epsilon \text{ and } g_x \in \partial \phi(x, z_x)$$

$$\Rightarrow \qquad g_x \in \partial_\epsilon f(x)$$

ϵ-SUBGRADIENT METHOD

- Uses an ϵ-subgradient in place of a subgradient.
- **Problem:** Minimize convex $f : \Re^n \mapsto \Re$ over a closed convex set X.
- **Method:**

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

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

- Can be viewed as subgradient method with "errors".

CONVERGENCE ANALYSIS

- **Basic inequality:** If $\{x_k\}$ is the ϵ-subgradient method sequence, for all $y \in X$ and $k \geq 0$

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

- Replicate the entire convergence analysis for subgradient methods, but carry along the ϵ_k terms.

- **Example:** Constant $\alpha_k \equiv \alpha$, constant $\epsilon_k \equiv \epsilon$. Assume $\|g_k\| \leq c$ for all k. For any optimal x^*,

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

so the distance to x^* decreases if

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

or equivalently, if x_k is outside the level set

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

- **Example:** If $\alpha_k \to 0$, $\sum_k \alpha_k \to \infty$, and $\epsilon_k \to \epsilon$, we get convergence to the ϵ-optimal set.

INCREMENTAL SUBGRADIENT METHODS

- Consider minimization of sum

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

- Often arises in duality contexts with m: **very large** (e.g., separable problems).

- Incremental method **moves x along a subgradient g_i of a component function f_i** NOT the (expensive) subgradient of f, which is $\sum_i g_i$.

- View an iteration as a cycle of m subiterations, one for each component f_i.

- Let x_k be obtained after k cycles. To obtain x_{k+1}, do one more cycle: Start with $\psi_0 = x_k$, and set $x_{k+1} = \psi_m$, after the m steps

$$\psi_i = P_X(\psi_{i-1} - \alpha_k g_i), \qquad i = 1, \ldots, m$$

with g_i being a subgradient of f_i at ψ_{i-1}.

- **Motivation is faster convergence.** A cycle can make much more progress than a subgradient iteration with essentially the same computation.

CONNECTION WITH ϵ-SUBGRADIENTS

- **Neighborhood property:** If x and $\overline{x}$ are "near" each other, then subgradients at $\overline{x}$ can be viewed as ϵ-subgradients at x, with ϵ "small."

- If $g \in \partial f(\overline{x})$, we have for all $z \in \Re^n$,

$$
\begin{aligned}
f(z) &\geq f(\overline{x}) + g'(z - \overline{x}) \\
&\geq f(x) + g'(z - x) + f(\overline{x}) - f(x) + g'(x - \overline{x}) \\
&\geq f(x) + g'(z - x) - \epsilon,
\end{aligned}
$$

where $\epsilon = |f(\overline{x}) - f(x)| + \|g\| \cdot \|\overline{x} - x\|$. Thus, $g \in \partial_\epsilon f(x)$, with ϵ: small when $\overline{x}$ is near x.

- The incremental subgradient iter. is an ϵ-subgradient iter. with $\epsilon = \epsilon_1 + \cdots + \epsilon_m$, where ϵ_i is the "error" in ith step in the cycle (ϵ_i: Proportional to α_k).

- Use

$$
\partial_{\epsilon_1} f_1(x) + \cdots + \partial_{\epsilon_m} f_m(x) \subset \partial_\epsilon f(x),
$$

where $\epsilon = \epsilon_1 + \cdots + \epsilon_m$, to approximate the ϵ-subdifferential of the sum $f = \sum_{i=1}^m f_i$.

- Convergence to optimal if $\alpha_k \to 0$, $\sum_k \alpha_k \to \infty$.

APPROXIMATION APPROACHES

- Approximation methods replace the original problem with an approximate problem.
- The approximation may be iteratively refined, for convergence to an exact optimum.
- A partial list of methods:
  - Cutting plane/outer approximation.
  - Simplicial decomposition/inner approximation.
  - Proximal methods (including Augmented Lagrangian methods for constrained minimization).
  - Interior point methods.
- A partial list of combination of methods:
  - Combined inner-outer approximation.
  - Bundle methods (proximal-cutting plane).
  - Combined proximal-subgradient (incremental option).

SUBGRADIENTS-OUTER APPROXIMATION

- Consider minimization of a convex function $f : \Re^n \mapsto \Re$, over a closed convex set X.

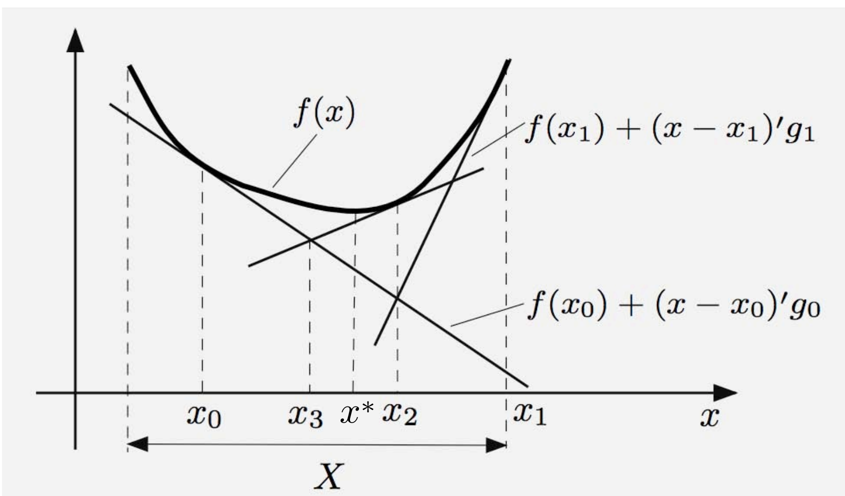
- We assume that at each $x \in X$, a subgradient g of f can be computed.

- We have

$$f(z) \geq f(x) + g'(z - x), \qquad \forall\, z \in \Re^n,$$

so each subgradient defines a plane (a linear function) that approximates f from below.

- The idea of the outer approximation/cutting plane approach is to build an ever more accurate approximation of f using such planes.



CUTTING PLANE METHOD

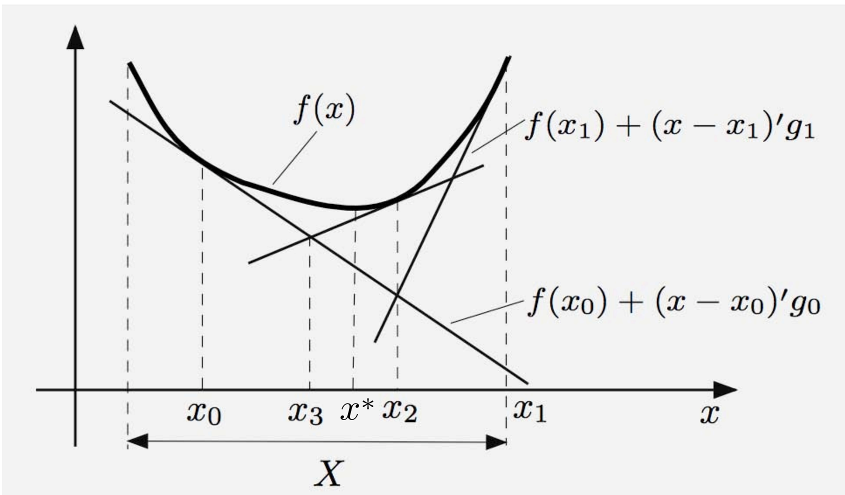
- Start with any $x_0 \in X$. For $k \geq 0$, set

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

where

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

and g_i is a subgradient of f at x_i.



- Note that $F_k(x) \leq f(x)$ for all x, and that $F_k(x_{k+1})$ increases monotonically with k. These imply that all limit points of x_k are optimal.

Proof: If $x_k \to \overline{x}$ then $F_k(x_k) \to f(\overline{x})$, [otherwise there would exist a hyperplane strictly separating epi(f) and $(\overline{x}, \lim_{k\to\infty} F_k(x_k))$]. This implies that $f(\overline{x}) \leq \lim_{k\to\infty} F_k(x) \leq f(x)$ for all x. **Q.E.D.**

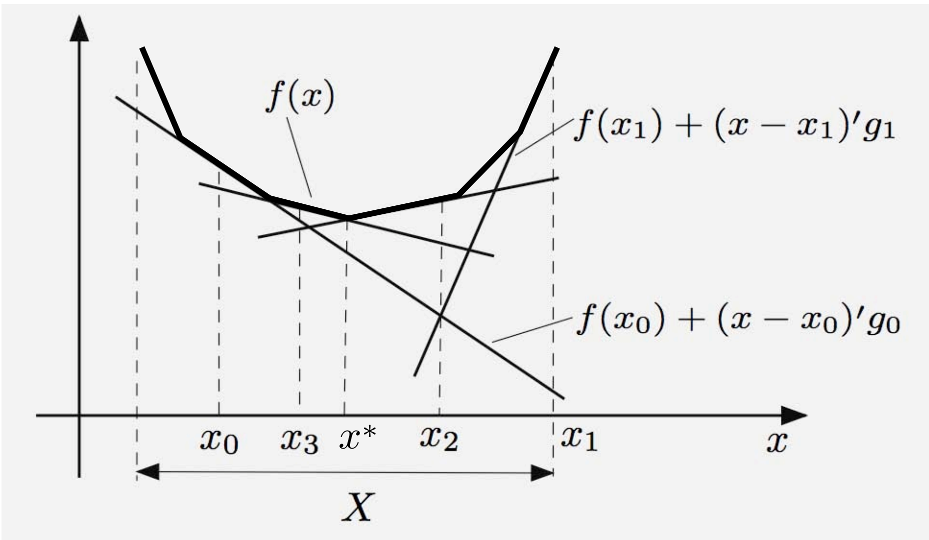
CONVERGENCE AND TERMINATION

- We have for all k

$$F_k(x_{k+1}) \leq f^* \leq \min_{i \leq k} f(x_i)$$

- Termination when $\min_{i \leq k} f(x_i) - F_k(x_{k+1})$ comes to within some small tolerance.

- For f polyhedral, we have finite termination with an exactly optimal solution.

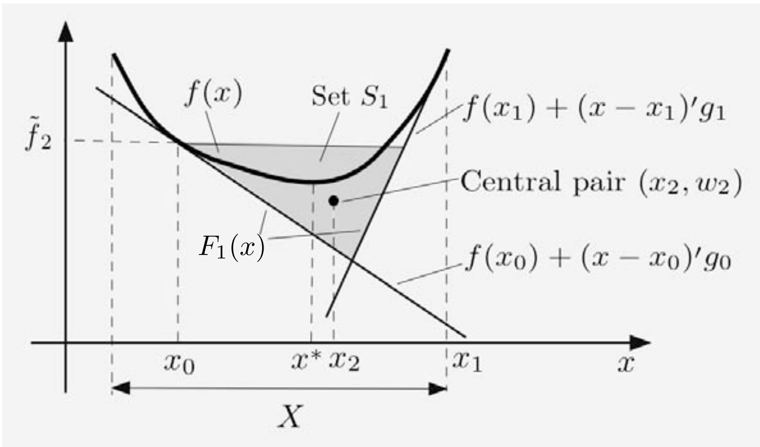


- **Instability problem:** The method can make large moves that deteriorate the value of f.

- Starting from the exact minimum it typically moves away from that minimum.

VARIANTS

- **Variant I:** Simultaneously with f, construct polyhedral approximations to X.
- **Variant II:** Central cutting plane methods



LECTURE 17

LECTURE OUTLINE

- Review of cutting plane method
- Simplicial decomposition
- Duality between cutting plane and simplicial decomposition

CUTTING PLANE METHOD

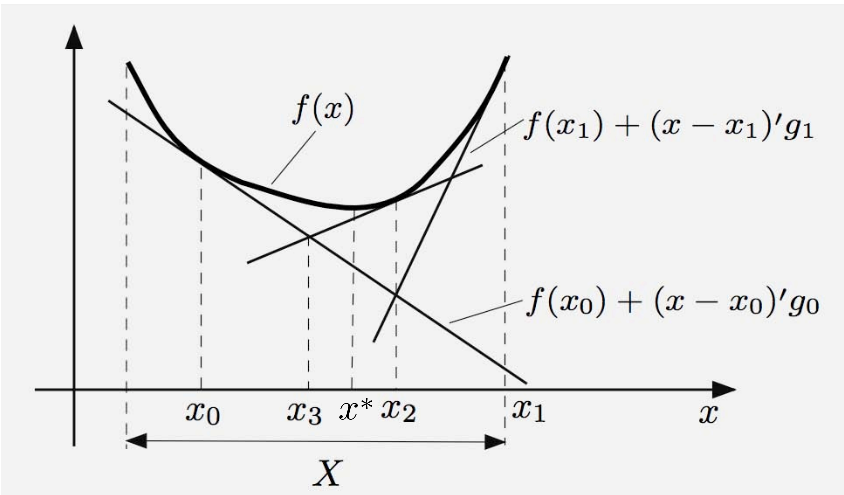
- Start with any $x_0 \in X$. For $k \geq 0$, set

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

where

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

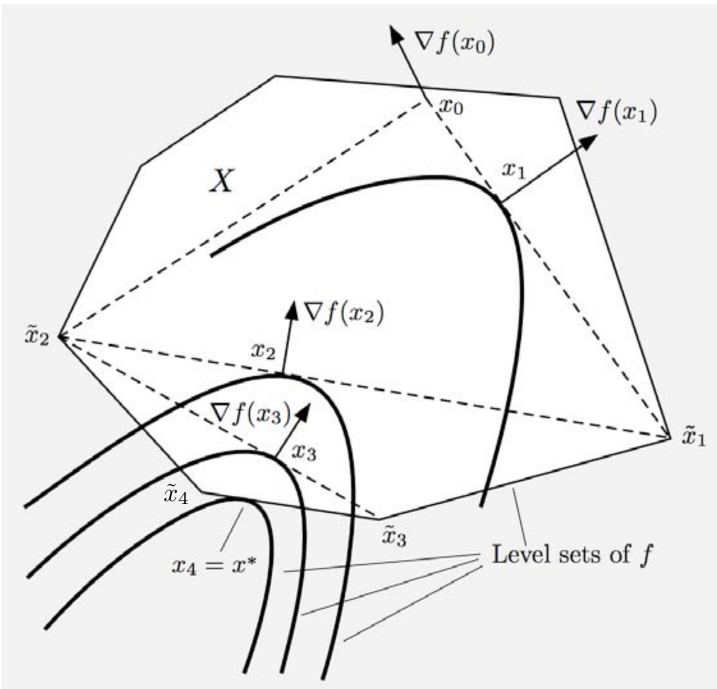
and g_i is a subgradient of f at x_i.



- We have $F_k(x) \leq f(x)$ for all x, and $F_k(x_{k+1})$ increases monotonically with k.
- These imply that all limit points of x_k are optimal.

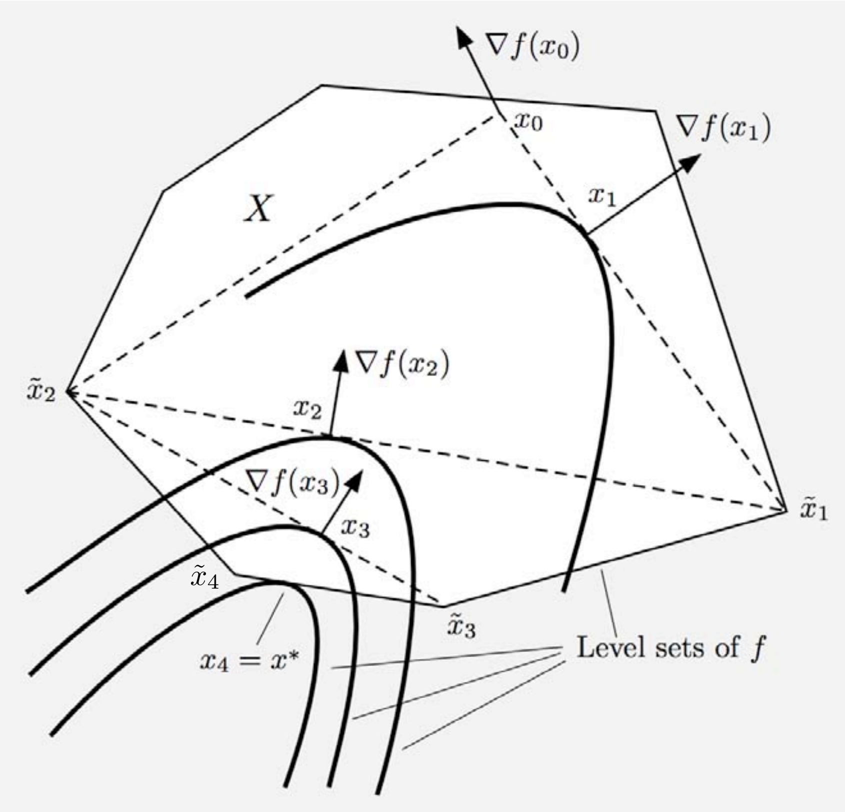
BASIC SIMPLICIAL DECOMPOSITION

- Minimize a *differentiable* convex $f : \Re^n \mapsto \Re$ over *bounded polyhedral constraint set* X.
- **Approximate** X with a simpler inner approximating polyhedral set.
- Construct a more refined problem by solving a **linear** minimization over the original constraint.



- The method is appealing under two conditions:
  - Minimizing f over the convex hull of a relative small number of extreme points is much simpler than minimizing f over X.
  - Minimizing a linear function over X is much simpler than minimizing f over X.

SIMPLICIAL DECOMPOSITION METHOD



- Given current iterate x_k, and finite set $X_k \subset X$ (initially $x_0 \in X$, $X_0 = \{x_0\}$).
- Let $\tilde{x}_{k+1}$ be extreme point of X that solves

$$\begin{aligned} &\text{minimize} \quad \nabla f(x_k)'(x - x_k) \\ &\text{subject to} \quad x \in X \end{aligned}$$

and add $\tilde{x}_{k+1}$ to X_k: $X_{k+1} = \{\tilde{x}_{k+1}\} \cup X_k$.

- Generate x_{k+1} as optimal solution of

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in \text{conv}(X_{k+1}). \end{aligned}$$

CONVERGENCE

- There are two possibilities for $\tilde{x}_{k+1}$:

  (a) We have

  $$0 \le \nabla f(x_k)'(\tilde{x}_{k+1}-x_k) = \min_{x \in X} \nabla f(x_k)'(x-x_k)$$

  Then x_k minimizes f over X (satisfies the optimality condition)

  (b) We have

  $$0 > \nabla f(x_k)'(\tilde{x}_{k+1} - x_k)$$

  Then $\tilde{x}_{k+1} \notin \text{conv}(X_k)$, since x_k minimizes f over $x \in \text{conv}(X_k)$, so that

  $$\nabla f(x_k)'(x - x_k) \ge 0, \qquad \forall\, x \in \text{conv}(X_k)$$

- Case (b) cannot occur an infinite number of times ($\tilde{x}_{k+1} \notin X_k$ and X has finitely many extreme points), so case (a) must eventually occur.

- The method will find a minimizer of f over X in a finite number of iterations.

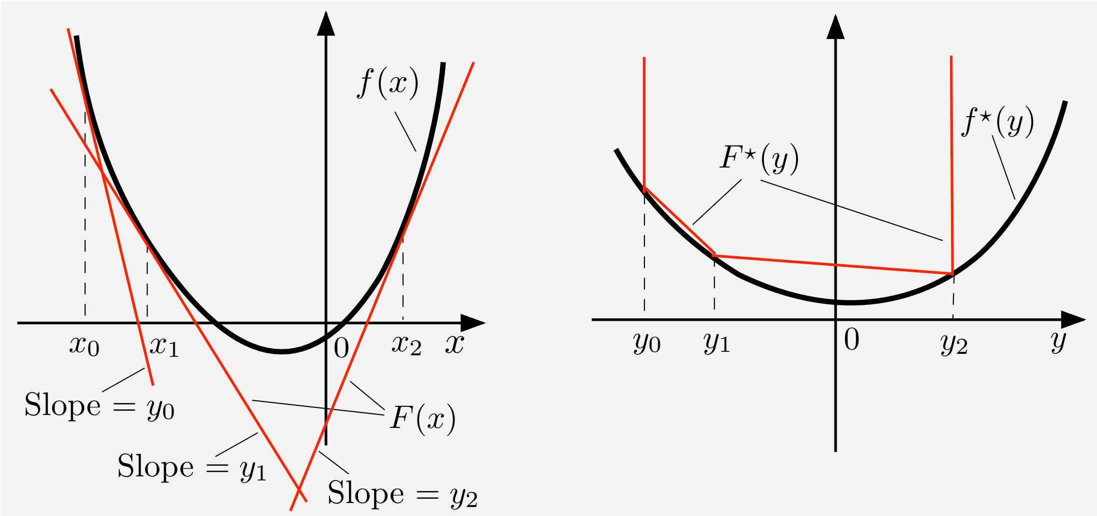
COMMENTS ON SIMPLICIAL DECOMP.

- Important specialized applications
- Variant to enhance efficiency. Discard some of the extreme points that seem unlikely to "participate" in the optimal solution, i.e., all $\tilde{x}$ such that

$$\nabla f(x_{k+1})'(\tilde{x} - x_{k+1}) > 0$$

- Variant to remove the boundedness assumption on X (impose artificial constraints)
- Extension to X nonpolyhedral (method remains unchanged, but convergence proof is more complex)
- Extension to f nondifferentiable (requires use of subgradients in place of gradients, and more sophistication)
- **Duality relation with cutting plane methods**
- We will view cutting plane and simplicial decomposition as special cases of two polyhedral approximation methods that are dual to each other

OUTER LINEARIZATION OF FNS



Outer Linearization of f Inner Linearization of Conjugate $f^\star$

- Outer linearization of closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$
- Defined by set of "slopes" $\{y_1, \ldots, y_\ell\}$, where $y_j \in \partial f(x_j)$ for some x_j
- Given by

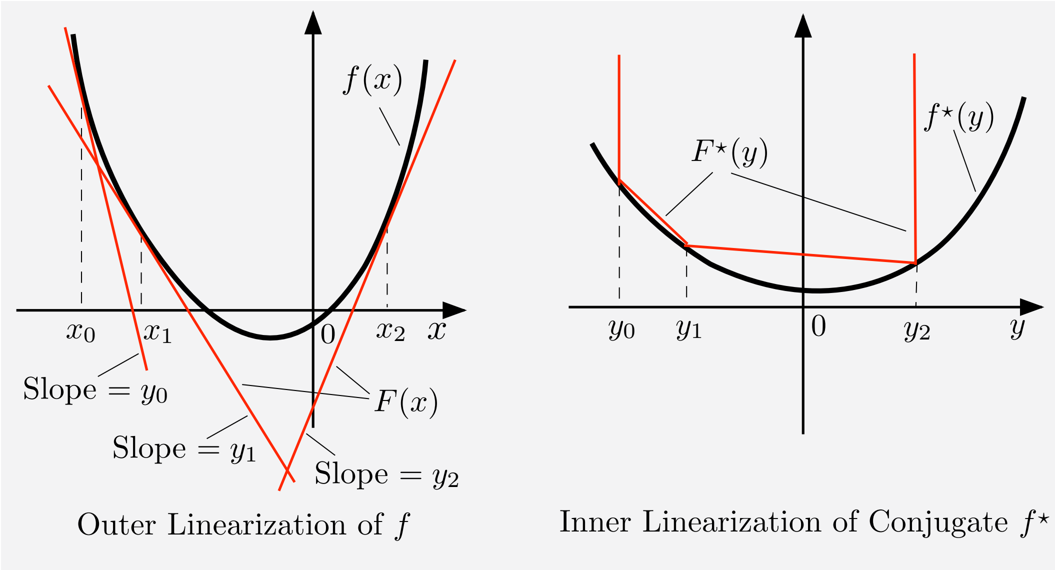
$$F(x) = \max_{j=1,\ldots,\ell} \big\{ f(x_j) + (x - x_j)' y_j \big\}, \qquad x \in \Re^n$$

or equivalently

$$F(x) = \max_{j=1,\ldots,\ell} \big\{ y_j' x - f^\star(y_j) \big\}$$

[this follows using $x_j' y_j = f(x_j) + f^\star(y_j)$, which is implied by $y_j \in \partial f(x_j)$ – the Conjugate Subgradient Theorem]

INNER LINEARIZATION OF FNS



- Consider conjugate $F^\star$ of outer linearization F
- After calculation using the formula

$$F(x) = \max_{j=1,\ldots,\ell} \big\{y_j'x - f^\star(y_j)\big\}$$

$F^\star$ is a piecewise linear approximation of $f^\star$ defined by "break points" at $y_1, \ldots, y_\ell$

- We have

$$\text{dom}(F^\star) = \text{conv}\big(\{y_1, \ldots, y_\ell\}\big),$$

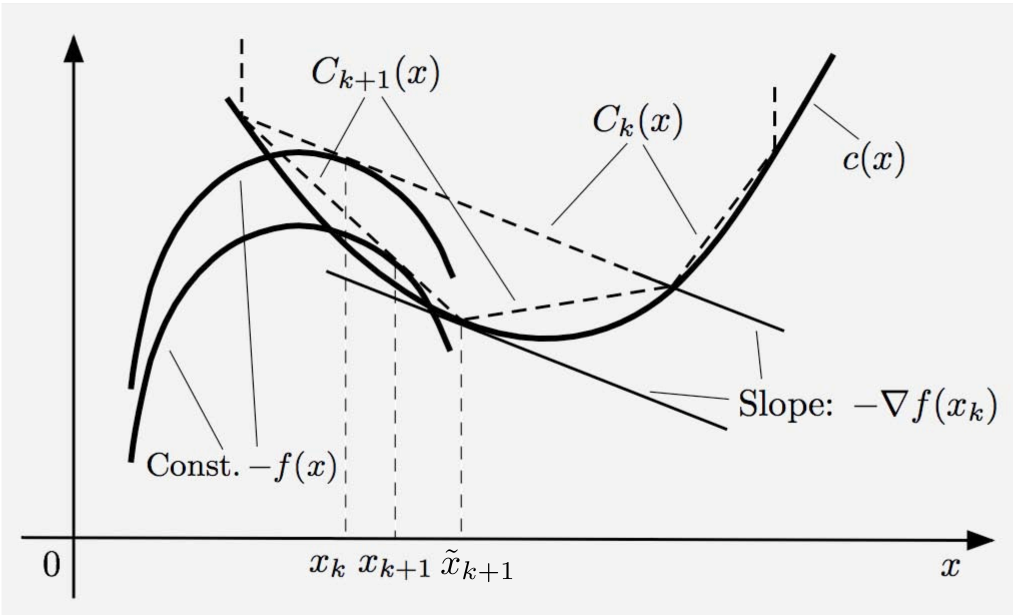
with values at $y_1, \ldots, y_\ell$ equal to the corresponding values of $f^\star$

- Epigraph of $F^\star$ is the convex hull of the union of the vertical halflines corresponding to $y_1, \ldots, y_\ell$:

$$\text{epi}(F^\star) = \text{conv}\Big(\cup_{j=1,\ldots,\ell} \big\{(y_j, w) \mid f^\star(y_j) \le w\big\}\Big)$$

GENERALIZED SIMPLICIAL DECOMPOSITION

- Consider minimization of $f(x) + c(x)$, over $x \in \Re^n$, where f and c are closed proper convex
- Case where f is differentiable



- Given C_k: inner linearization of c, obtain

$$x_k \in \arg\min_{x \in \Re^n} \{ f(x) + C_k(x) \}$$

- Obtain $\tilde{x}_{k+1}$ such that

$$-\nabla f(x_k) \in \partial c(\tilde{x}_{k+1}),$$

and form $X_{k+1} = X_k \cup \{\tilde{x}_{k+1}\}$

NONDIFFERENTIABLE CASE

- Given C_k: inner linearization of c, obtain

$$x_k \in \arg\min_{x \in \Re^n} \{ f(x) + C_k(x) \}$$

- Obtain a subgradient $g_k \in \partial f(x_k)$ such that

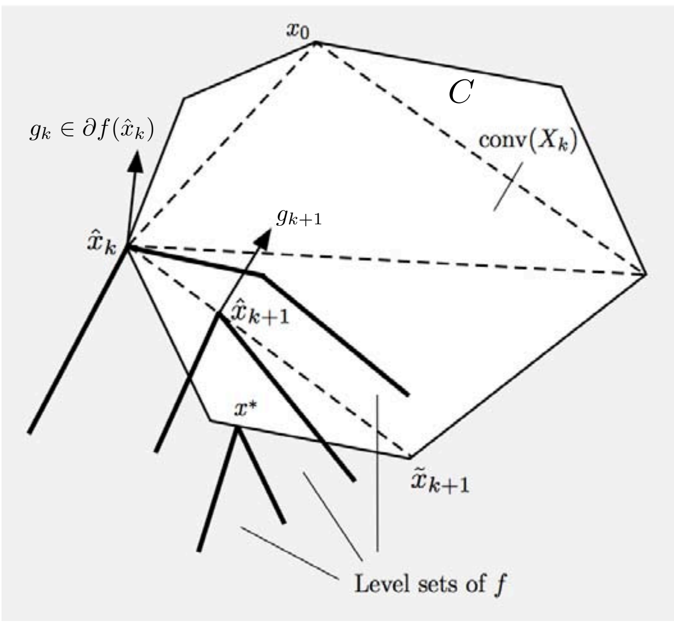
$$-g_k \in \partial C_k(x_k)$$

- Obtain $\tilde{x}_{k+1}$ such that

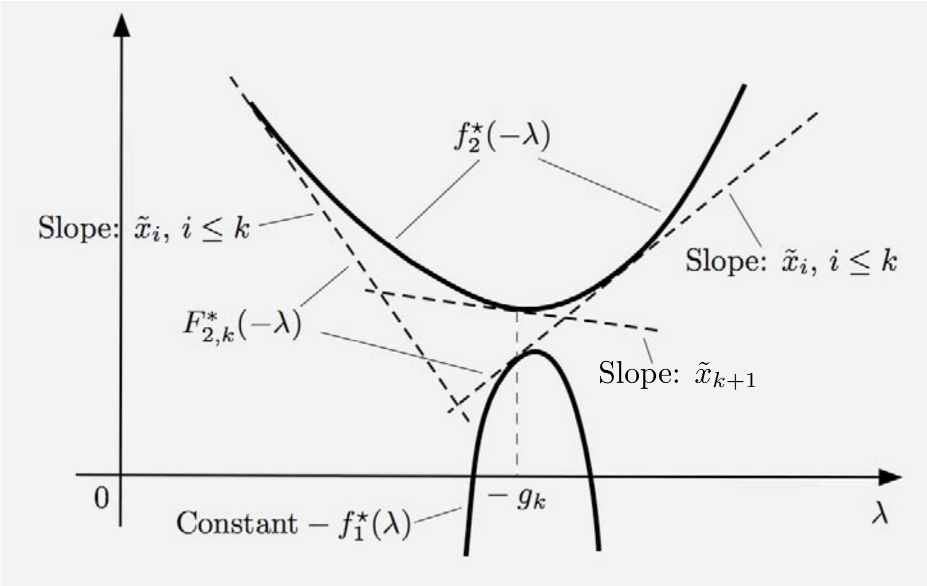
$$-g_k \in \partial c(\tilde{x}_{k+1}),$$

and form $X_{k+1} = X_k \cup \{\tilde{x}_{k+1}\}$

- **Example:** c is the indicator function of a polyhedral set



DUAL CUTTING PLANE IMPLEMENTATION



- Primal and dual Fenchel pair

$$\min_{x\in\Re^n} f_1(x) + f_2(x), \qquad \min_{\lambda\in\Re^n} f_1^\star(\lambda) + f_2^\star(-\lambda)$$

- Primal and dual approximations

$$\min_{x\in\Re^n} f_1(x) + F_{2,k}(x) \qquad \min_{\lambda\in\Re^n} f_1^\star(\lambda) + F_{2,k}^\star(-\lambda)$$

- $F_{2,k}$ and $F_{2,k}^\star$ are inner and outer approximations of f_2 and $f_2^\star$

- $\tilde{x}_{i+1}$ and g_i are solutions of the primal or the dual approximating problem (and corresponding subgradients)

LECTURE 18

LECTURE OUTLINE

- Generalized polyhedral approximation methods
- Combined cutting plane and simplicial decomposition methods
- Lecture based on the paper

D. P. Bertsekas and H. Yu, "A Unifying Polyhedral Approximation Framework for Convex Optimization," SIAM J. on Optimization, Vol. 21, 2011, pp. 333-360.

Generalized Polyhedral Approximations in Convex Optimization

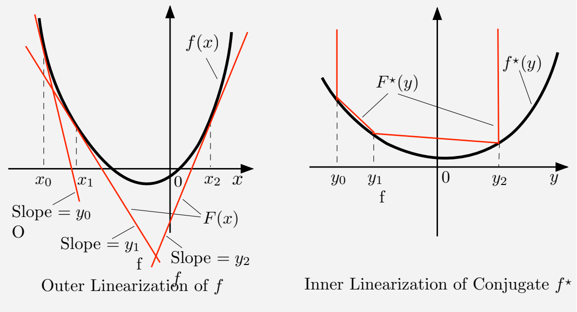
Dimitri P. Bertsekas

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Lecture 18, 6.253 Class

Lecture Summary

- Outer/inner linearization and their duality.



- A unifying framework for polyhedral approximation methods.
- Includes classical methods:
  - Cutting plane/Outer linearization
  - Simplicial decomposition/Inner linearization
- Includes new methods, and new versions/extensions of old methods.

Vehicle for Unification

- Extended monotropic programming (EMP)

$$\min_{(x_1,\ldots,x_m)\in S} \quad \sum_{i=1}^{m} f_i(x_i)$$

where $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ is a convex function and S is a subspace.

- The dual EMP is

$$\min_{(y_1,\ldots,y_m)\in S^\perp} \sum_{i=1}^{m} f_i^\star(y_i)$$

where $f_i^\star$ is the convex conjugate function of f_i.

- Algorithmic Ideas:
  - Outer or inner linearization for some of the f_i
  - Refinement of linearization using duality
- Features of outer or inner linearization use:
  - They are combined in the same algorithm
  - Their roles are reversed in the dual problem
  - Become two (mathematically equivalent dual) faces of the same coin

Advantage over Classical Cutting Plane Methods

- The refinement process is much faster.
  - Reason: At each iteration we add multiple cutting planes (as many as one per component function f_i).
  - By contrast a single cutting plane is added in classical methods.
- The refinement process may be more convenient.
  - For example, when f_i is a scalar function, adding a cutting plane to the polyhedral approximation of f_i can be very simple.
  - By contrast, adding a cutting plane may require solving a complicated optimization process in classical methods.

References

- D. P. Bertsekas, “Extended Monotropic Programming and Duality," Lab. for Information and Decision Systems Report 2692, MIT, Feb. 2010; a version appeared in JOTA, 2008, Vol. 139, pp. 209-225.
- D. P. Bertsekas, “Convex Optimization Theory," 2009, www-based “living chapter" on algorithms.
- D. P. Bertsekas and H. Yu, “A Unifying Polyhedral Approximation Framework for Convex Optimization," Lab. for Information and Decision Systems Report LIDS-P-2820, MIT, September 2009; SIAM J. on Optimization, Vol. 21, 2011, pp. 333-360.

Outline

1. Polyhedral Approximation
   - Outer and Inner Linearization
   - Cutting Plane and Simplicial Decomposition Methods

2. Extended Monotropic Programming
   - Duality Theory
   - General Approximation Algorithm

3. Special Cases
   - Cutting Plane Methods
   - Simplicial Decomposition for $\min_{x \in C} f(x)$

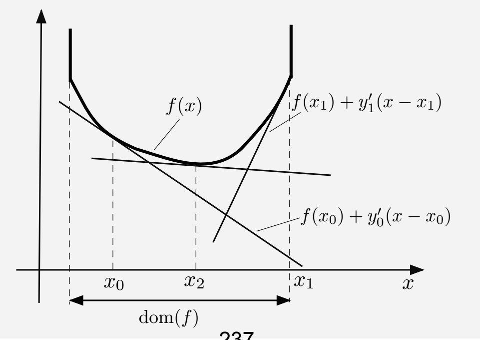
Outer Linearization - Epigraph Approximation by Halfspaces

- Given a convex function $f : \Re^n \mapsto (-\infty, \infty]$.
- Approximation using subgradients:

$$\max\left\{f(x_0) + y_0'(x - x_0), \ldots, f(x_k) + y_k'(x - x_k)\right\}$$

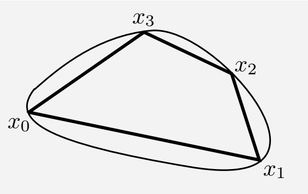
where

$$y_i \in \partial f(x_i), \qquad i = 0, \ldots, k$$

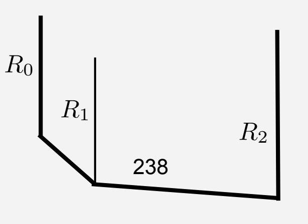


Convex Hulls

- Convex hull of a finite set of points x_i



- Convex hull of a union of a finite number of rays R_i

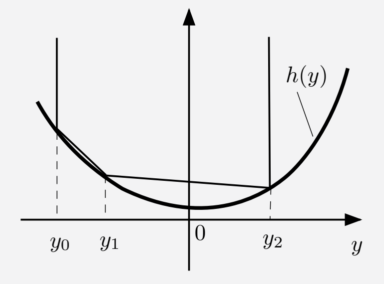


Inner Linearization - Epigraph Approximation by Convex Hulls

- Given a convex function $h : \Re^n \mapsto (-\infty, \infty]$ and a finite set of points

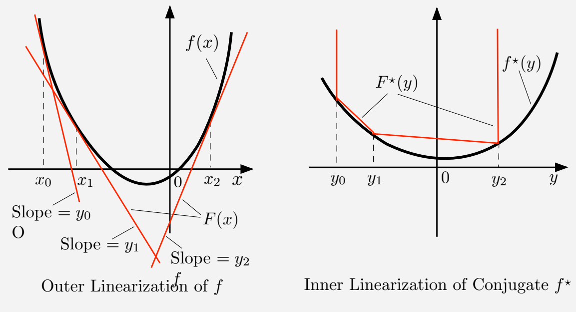
$$y_0, \ldots, y_k \in \mathrm{dom}(h)$$

- Epigraph approximation by convex hull of rays $\{(y_i, w) \mid w \geq h(y_i)\}$



Conjugacy of Outer/Inner Linearization

- Given a function $f : \Re^n \mapsto (-\infty, \infty]$ and its conjugate $f^\star$.
- The conjugate of an outer linearization of f is an inner linearization of $f^\star$.



- Subgradients in outer lin. <==> Break points in inner lin.

Cutting Plane Method for $\min_{x\in C} f(x)$ (C polyhedral)

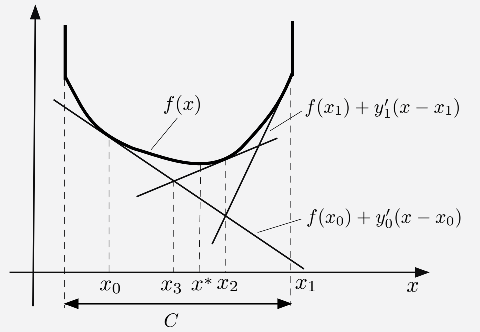
- Given $y_i \in \partial f(x_i)$ for $i = 0, \ldots, k$, form

$$F_k(x) = \max\left\{f(x_0) + y_0'(x - x_0), \ldots, f(x_k) + y_k'(x - x_k)\right\}$$

and let

$$x_{k+1} \in \arg\min_{x\in C} F_k(x)$$

- At each iteration solves LP of large dimension (which is simpler than the original problem).



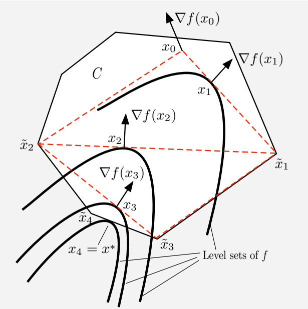
Simplicial Decomposition for $\min_{x\in C} f(x)$ (f smooth, C polyhedral)

- At the typical iteration we have x_k and $X_k = \{x_0, \tilde{x}_1, \ldots, \tilde{x}_k\}$, where $\tilde{x}_1, \ldots, \tilde{x}_k$ are extreme points of C.
- Solve LP of large dimension: Generate

$$\tilde{x}_{k+1} \in \arg\min_{x\in C} \{\nabla f(x_k)'(x - x_k)\}$$

- Solve NLP of small dimension: Set $X_{k+1} = \{\tilde{x}_{k+1}\} \cup X_k$, and generate x_{k+1} as

$$x_{k+1} \in \arg \min_{x\in \mathrm{conv}(X_{k+1})} f(x)$$



- Finite convergence if C is a bounded polyhedron.

Comparison: Cutting Plane - Simplicial Decomposition

- Cutting plane aims to use LP with same dimension and smaller number of constraints.
- Most useful when problem has small dimension and:
  - There are many linear constraints, or
  - The cost function is nonlinear and linear versions of the problem are much simpler
- Simplicial decomposition aims to use NLP over a simplex of small dimension [i.e., $conv(X_k)$].
- Most useful when problem has large dimension and:
  - Cost is nonlinear, and
  - Solving linear versions of the (large-dimensional) problem is much simpler (possibly due to decomposition)
- The two methods appear very different, with unclear connection, despite the general conjugacy relation between outer and inner linearization.
- We will see that they are special cases of two methods that are dual (and mathematically equivalent) to each other.

Extended Monotropic Programming (EMP)

$$\min_{(x_1,\ldots,x_m)\in S} \quad \sum_{i=1}^{m} f_i(x_i)$$

where $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ is a closed proper convex, S is subspace.

- Monotropic programming (Rockafellar, Minty), where f_i: scalar functions.
- Single commodity network flow (S: circulation subspace of a graph).
- Block separable problems with linear constraints.
- Fenchel duality framework: Let $m = 2$ and $S = \{(x, x) \mid x \in \Re^n\}$. Then the problem

  $$\min_{(x_1,x_2)\in S} f_1(x_1) + f_2(x_2)$$

  can be written in the Fenchel format

  $$\min_{x\in\Re^n} f_1(x) + f_2(x)$$

- Conic programs (second order, semidefinite - special case of Fenchel).
- Sum of functions (e.g., machine learning): For $S = \{(x, \ldots, x) \mid x \in \Re^n\}$

  $$\min_{x\in\Re^n} \sum_{i=1}^{m} f_i(x)$$

Dual EMP

- Derivation: Introduce $z_i \in \Re^{n_i}$ and convert EMP to an equivalent form

$$\min_{(x_1,\dots,x_m)\in S} \sum_{i=1}^{m} f_i(x_i) \qquad \equiv \qquad \min_{\substack{z_i = x_i,\, i=1,\dots,m,\\ (x_1,\dots,x_m)\in S}} \sum_{i=1}^{m} f_i(z_i)$$

- Assign multiplier $y_i \in \Re^{n_i}$ to constraint $z_i = x_i$, and form the Lagrangian

$$L(x, z, y) = \sum_{i=1}^{m} f_i(z_i) + y_i'(x_i - z_i)$$

where $y = (y_1, \dots, y_m)$.

- The dual problem is to maximize the dual function

$$q(y) = \inf_{(x_1,\dots,x_m)\in S,\; z_i\in\Re^{n_i}} L(x, z, y)$$

- Exploiting the separability of $L(x, z, y)$ and changing sign to convert maximization to minimization, we obtain the dual EMP in symmetric form

$$\min_{(y_1,\dots,y_m)\in S^\perp} \sum_{i=1}^{m} f_i^\star(y_i)$$

where $f_i^\star$ is the convex conjugate function of f_i.

Optimality Conditions

- There are powerful conditions for strong duality $q^* = f^*$ (generalizing classical monotropic programming results):
  - Vector Sum Condition for Strong Duality: Assume that for all feasible x, the set
  $$S^\perp + \partial_\epsilon(f_1 + \cdots + f_m)(x)$$
  is closed for all $\epsilon > 0$. Then $q^* = f^*$.
  - Special Case: Assume each f_i is finite, or is polyhedral, or is essentially one-dimensional, or is domain one-dimensional. Then $q^* = f^*$.
  - By considering the dual EMP, "finite" may be replaced by "co-finite" in the above statement.

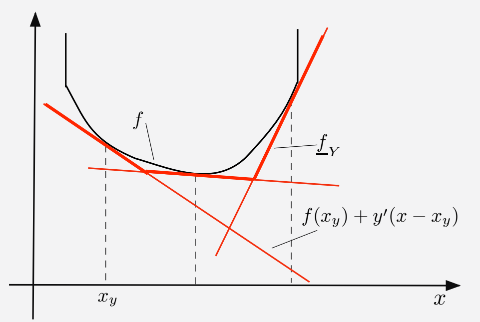
- Optimality conditions, assuming $-\infty < q^* = f^* < \infty$:
  - (x^*, y^*) is an optimal primal and dual solution pair if and only if
  $$x^* \in S, \qquad y^* \in S^\perp, \qquad y_i^* \in \partial f_i(x_i^*), \quad i = 1, \ldots, m$$
  - Symmetric conditions involving the dual EMP:
  $$x^* \in S, \qquad y^* \in S^\perp, \qquad x_i^* \in \partial f_i^\star(y_i^*), \quad i = 1, \ldots, m$$

Outer Linearization of a Convex Function: Definition

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be closed proper convex.
- Given a finite set $Y \subset \mathrm{dom}(f^\star)$, we define the outer linearization of f

$$\underline{f}_Y(x) = \max_{y \in Y} \left\{ f(x_y) + y'(x - x_y) \right\}$$

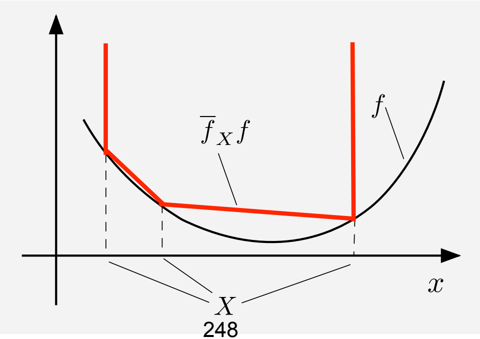
where x_y is such that $y \in \partial f(x_y)$.



Inner Linearization of a Convex Function: Definition

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be closed proper convex.
- Given a finite set $X \subset \mathrm{dom}(f)$, we define the inner linearization of f as the function $\bar{f}_X$ whose epigraph is the convex hull of the rays $\big\{(x, w) \mid w \geq f(x),\ x \in X\big\}$:

$$\bar{f}_X(z) = \begin{cases} \min_{\substack{\sum_{x \in X} \alpha_x x = z, \\ \sum_{x \in X} \alpha_x = 1,\ \alpha_x \geq 0,\ x \in X}} \sum_{x \in X} \alpha_x f(z) & \text{if } z \in \mathrm{conv}(X) \\ \infty & \text{otherwise} \end{cases}$$



Polyhedral Approximation Algorithm

- Let $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ be closed proper convex, with conjugates $f_i^\star$. Consider the EMP

$$\min_{(x_1,\ldots,x_m)\in S} \sum_{i=1}^{m} f_i(x_i)$$

- Introduce a fixed partition of the index set:

$$\{1,\ldots,m\} = I \cup \underline{I} \cup \bar{I}, \qquad \underline{I} : \text{Outer indices}, \quad \bar{I} : \text{Inner indices}$$

- **Typical Iteration:** We have finite subsets $Y_i \subset \mathrm{dom}(f_i^\star)$ for each $i \in \underline{I}$, and $X_i \subset \mathrm{dom}(f_i)$ for each $i \in \bar{I}$.

  Find primal-dual optimal pair $\hat{x} = (\hat{x}_1,\ldots,\hat{x}_m)$, and $\hat{y} = (\hat{y}_1,\ldots,\hat{y}_m)$ of the approximate EMP

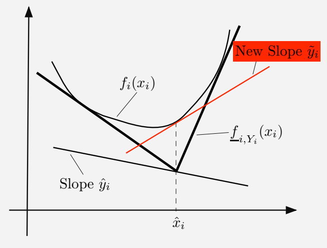
$$\min_{(x_1,\ldots,x_m)\in S} \sum_{i\in I} f_i(x_i) + \sum_{i\in \underline{I}} \underline{f}_{i,Y_i}(x_i) + \sum_{i\in \bar{I}} \bar{f}_{i,X_i}(x_i)$$

  Enlarge Y_i and X_i by differentiation:

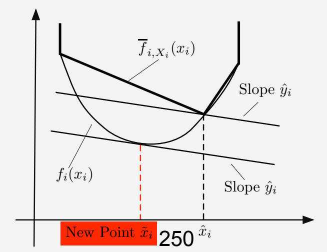
  - For each $i \in \underline{I}$, add $\tilde{y}_i$ to Y_i where $\tilde{y}_i \in \partial f_i(\hat{x}_i)$
  - For each $i \in \bar{I}$, add $\tilde{x}_i$ to X_i where $\tilde{x}_i \in \partial f_i^\star(\hat{y}_i)$.

Enlargement Step for ith Component Function

- Outer: For each $i \in \underline{I}$, add $\tilde{y}_i$ to Y_i where $\tilde{y}_i \in \partial f_i(\hat{x}_i)$.



- Inner: For each $i \in \bar{I}$, add $\tilde{x}_i$ to X_i where $\tilde{x}_i \in \partial f_i^\star(\hat{y}_i)$.



Mathematically Equivalent Dual Algorithm

- Instead of solving the primal approximate EMP

$$\min_{(x_1,\ldots,x_m)\in S} \sum_{i\in I} f_i(x_i) + \sum_{i\in \underline{I}} \underline{f}_{i,Y_i}(x_i) + \sum_{i\in \bar{I}} \bar{f}_{i,X_i}(x_i)$$

we may solve its dual

$$\min_{(y_1,\ldots,y_m)\in S^\perp} \sum_{i\in I} f_i^\star(y_i) + \sum_{i\in \underline{I}} \underline{f^\star}_{i,Y_i}(y_i) + \sum_{i\in \bar{I}} \bar{f^\star}_{i,X_i}(x_i)$$

where $\underline{f^\star}_{i,Y_i}$ and $\bar{f^\star}_{i,X_i}$ are the conjugates of $\underline{f}_{i,Y_i}$ and $\bar{f}_{i,X_i}$.

- Note that $\underline{f^\star}_{i,Y_i}$ is an inner linearization, and $\bar{f^\star}_{i,X_i}$ is an outer linearization (roles of inner/outer have been reversed).
- The choice of primal or dual is a matter of computational convenience, but does not affect the primal-dual sequences produced.

Comments on Polyhedral Approximation Algorithm

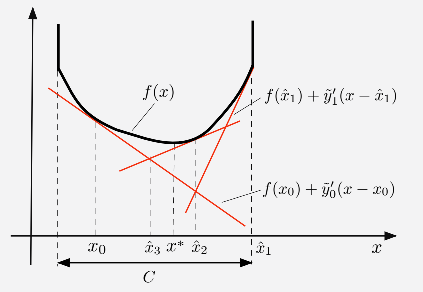
- In some cases we may use an algorithm that solves simultaneously the primal and the dual.
  - **Example:** Monotropic programming, where x_i is one-dimensional.
  - **Special case:** Convex separable network flow, where x_i is the one-dimensional flow of a directed arc of a graph, S is the circulation subspace of the graph.
- In other cases, it may be preferable to focus on solution of either the primal or the dual approximate EMP.
- After solving the primal, the refinement of the approximation ($\tilde{y}_i$ for $i \in \underline{I}$, and $\tilde{x}_i$ for $i \in \overline{I}$) may be found later by differentiation and/or some special procedure/optimization.
  - This may be easy, e.g., in the cutting plane method, or
  - This may be nontrivial, e.g., in the simplicial decomposition method.
- Subgradient duality [$y \in \partial f(x)$ iff $x \in \partial f^\star(y)$] may be useful.

Cutting Plane Method for $\min_{x \in C} f(x)$

- EMP equivalent: $\min_{x_1 = x_2} f(x_1) + \delta(x_2 \mid C)$, where $\delta(x_2 \mid C)$ is the indicator function of C.
- Classical cutting plane algorithm: Outer linearize f only, and solve the primal approximate EMP. It has the form

$$\min_{x \in C} \underline{f}_Y(x)$$

where Y is the set of subgradients of f obtained so far. If $\hat{x}$ is the solution, add to Y a subgradient $\tilde{y} \in \partial f(\hat{x})$.



Simplicial Decomposition Method for $\min_{x \in C} f(x)$

- EMP equivalent: $\min_{x_1 = x_2} f(x_1) + \delta(x_2 \mid C)$, where $\delta(x_2 \mid C)$ is the indicator function of C.
- Generalized Simplicial Decomposition: Inner linearize C only, and solve the primal approximate EMP. In has the form

$$\min_{x \in \bar{C}_X} f(x)$$

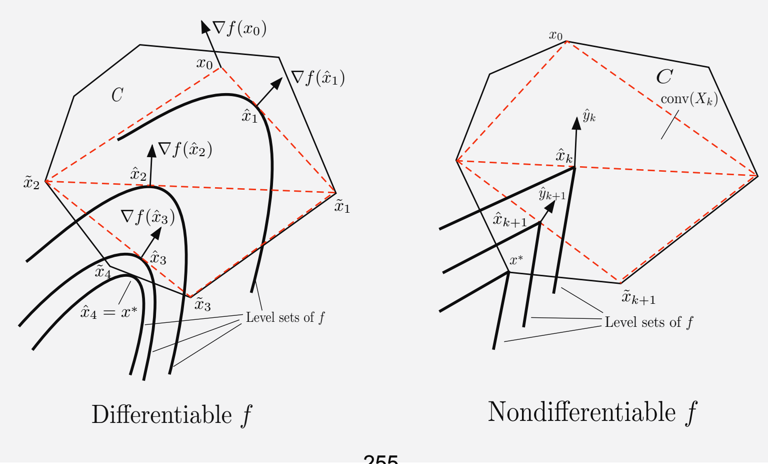
  where $\bar{C}_X$ is an inner approximation to C.
- Assume that $\hat{x}$ is the solution of the approximate EMP.
  - Dual approximate EMP solutions:

$$\{(\hat{y}, -\hat{y}) \mid \hat{y} \in \partial f(\hat{x}),\ -\hat{y} \in (\text{normal cone of } \bar{C}_X \text{ at } \hat{x})\}$$

  - In the classical case where f is differentiable, $\hat{y} = \nabla f(\hat{x})$.
  - Add to X a point $\tilde{x}$ such that $-\hat{y} \in \partial\delta(\tilde{x} \mid C)$, or

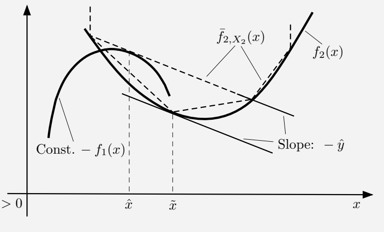
$$\tilde{x} \in \arg\min_{x \in C} \hat{y}' x$$

Illustration of Simplicial Decomposition for $\min_{x \in C} f(x)$

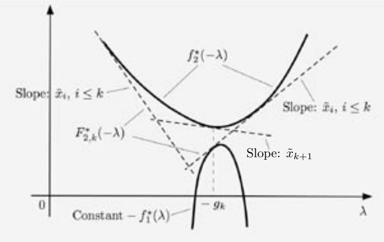


Dual Views for $\min_{x\in\Re^n}\{f_1(x)+f_2(x)\}$

- Inner linearize f_2



- Dual view: Outer linearize $f_2^\star$



Convergence - Polyhedral Case

- Assume that
  - All outer linearized functions f_i are finite polyhedral
  - All inner linearized functions f_i are co-finite polyhedral
  - The vectors $\tilde{y}_i$ and $\tilde{x}_i$ added to the polyhedral approximations are elements of the finite representations of the corresponding f_i
- Finite convergence: The algorithm terminates with an optimal primal-dual pair.
- Proof sketch: At each iteration two possibilities:
  - Either $(\hat{x}, \hat{y})$ is an optimal primal-dual pair for the original problem
  - Or the approximation of one of the f_i, $i \in \underline{I} \cup \bar{I}$, will be refined/improved
- By assumption there can be only a finite number of refinements. □

Convergence - Nonpolyhedral Case

- Convergence, pure outer linearization ($\bar{I}$: Empty). Assume that the sequence $\{\tilde{y}_i^k\}$ is bounded for every $i \in \underline{I}$. Then every limit point of $\{\hat{x}^k\}$ is primal optimal.
- Proof sketch: For all k, $\ell \leq k-1$, and $x \in S$, we have

$$\sum_{i \notin \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \underline{I}} \left(f_i(\hat{x}_i^\ell) + (\hat{x}_i^k - \hat{x}_i^\ell)' \tilde{y}_i^\ell \right) \leq \sum_{i \notin \underline{I}} f_i(\hat{x}_i^k) + \sum_{i \in \underline{I}} \underline{f}_{i, Y_i^{k-1}}(\hat{x}_i^k) \leq \sum_{i=1}^m f_i(x_i)$$

- Let $\{\hat{x}^k\}_{\mathcal{K}} \to \bar{x}$ and take limit as $\ell \to \infty$, $k \in \mathcal{K}$, $\ell \in \mathcal{K}$, $\ell < k$. □
- Exchanging roles of primal and dual, we obtain a convergence result for pure inner linearization case.
- Convergence, pure inner linearization ($\underline{I}$: Empty). Assume that the sequence $\{\tilde{x}_i^k\}$ is bounded for every $i \in \bar{I}$. Then every limit point of $\{\hat{y}^k\}$ is dual optimal.
- General mixed case: Convergence proof is more complicated (see the Bertsekas and Yu paper).

Concluding Remarks

- A unifying framework for polyhedral approximations based on EMP.
- Dual and symmetric roles for outer and inner approximations.
- There is option to solve the approximation using a primal method or a dual mathematical equivalent - whichever is more convenient/efficient.
- Several classical methods and some new methods are special cases.
- Proximal/bundle-like versions:
    - Convex proximal terms can be easily incorporated for stabilization and for improvement of rate of convergence.
    - Outer/inner approximations can be carried from one proximal iteration to the next.

LECTURE 19

LECTURE OUTLINE

- Proximal minimization algorithm
- Extensions

****

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

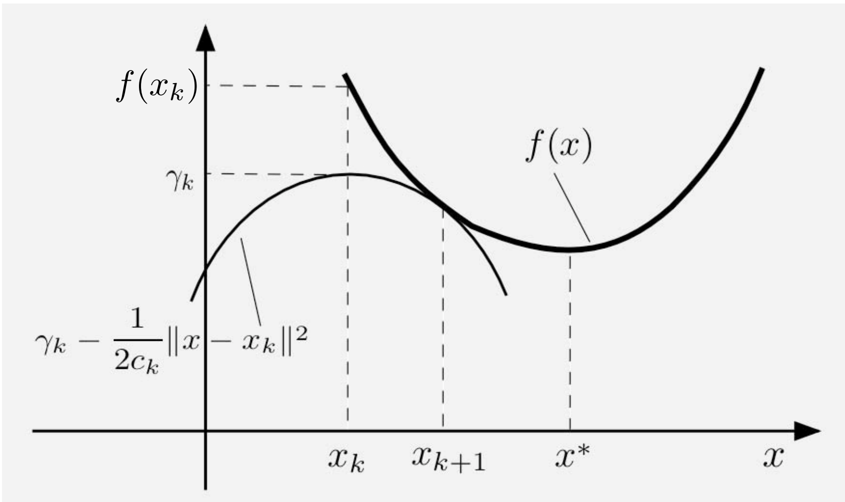
- Regularization in place of linearization
- Add a quadratic term to f to make it strictly convex and "well-behaved"
- Refine the approximation at each iteration by changing the quadratic term

PROXIMAL MINIMIZATION ALGORITHM

- A general algorithm for convex fn minimization

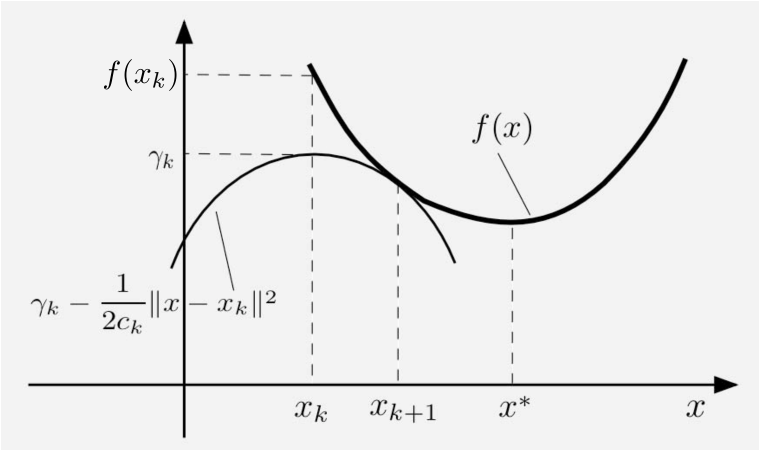
$$x_{k+1} \in \arg\min_{x \in \Re^n} \left\{ f(x) + \frac{1}{2c_k} \|x - x_k\|^2 \right\}$$

  - $f : \Re^n \mapsto (-\infty, \infty]$ 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, $x_{k+1} = x_k$.
- If $\sum_k c_k = \infty$, $f(x_k) \to f^*$ and $\{x_k\}$ converges to some optimal solution if one exists.

CONVERGENCE



- Some basic properties: For all k

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

so x_k to x_{k+1} move is "nearly" a subgradient step.

- For all k and $y \in \Re^n$

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

Distance to the optimum is improved.

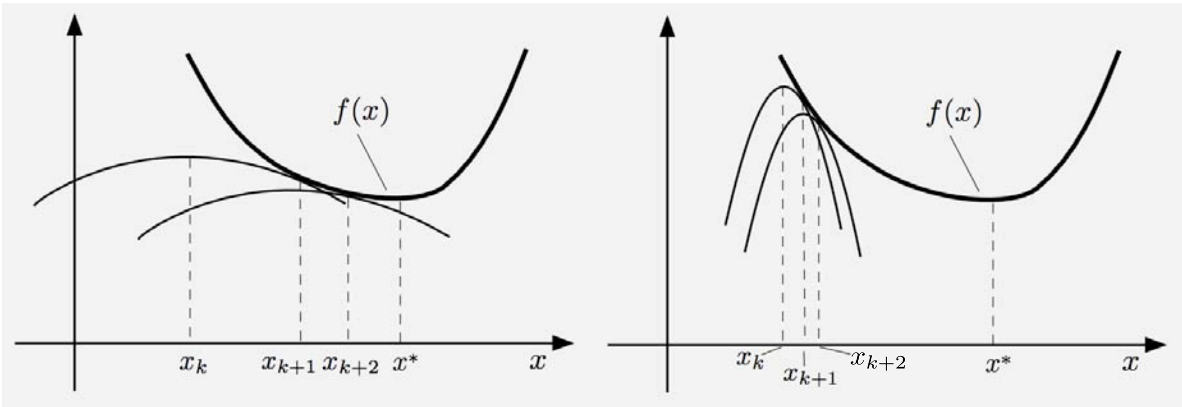
- Convergence mechanism:

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

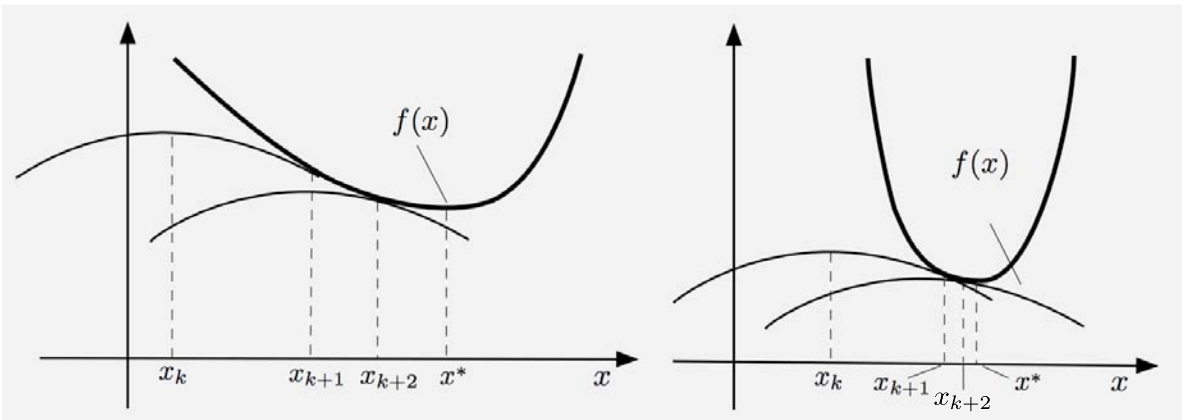
Cost improves by at least $\frac{1}{2c_k}\|x_{k+1} - x_k\|^2$, and this is sufficient to guarantee convergence.

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 that for some scalars $\beta > 0$, $\delta > 0$, 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

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

i.e., **growth of order α from optimal solution set X^*.**

- 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

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

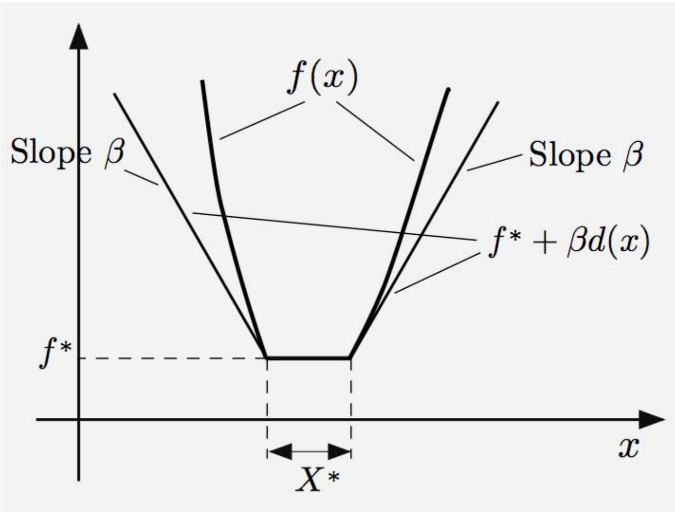
superlinear convergence.

FINITE CONVERGENCE

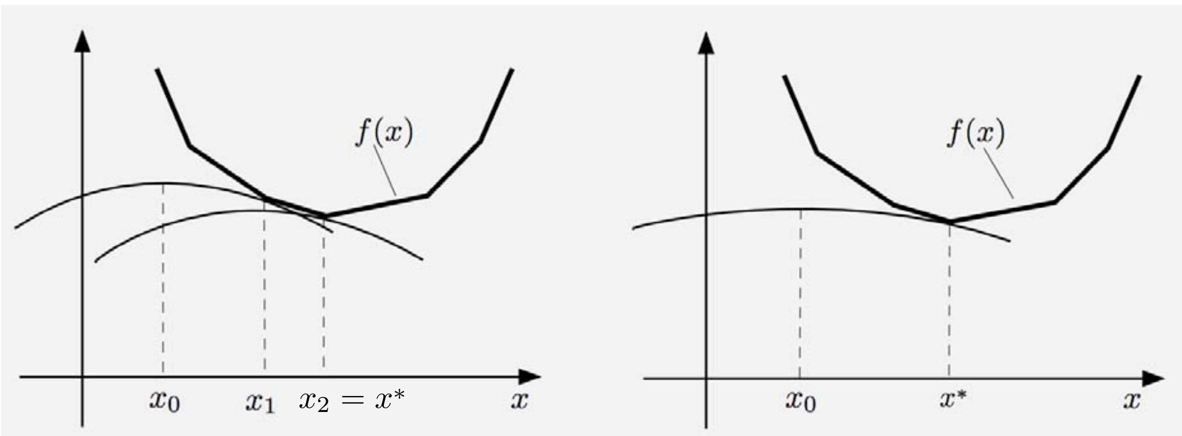
- Assume growth order $\alpha = 1$:

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

e.g., f is polyhedral.

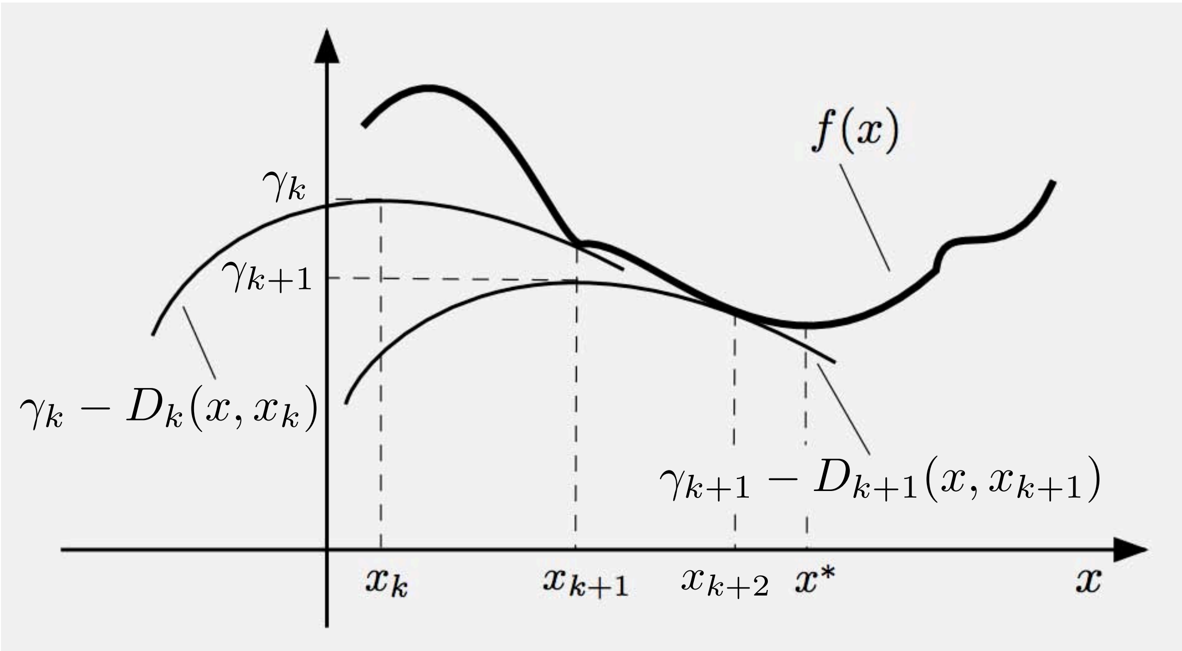


- Method converges finitely (in a single step for c_0 sufficiently large).



IMPORTANT EXTENSIONS

- Replace quadratic regularization by more general proximal term.
- Allow nonconvex f.



- Combine with linearization of f (we will focus on this first).

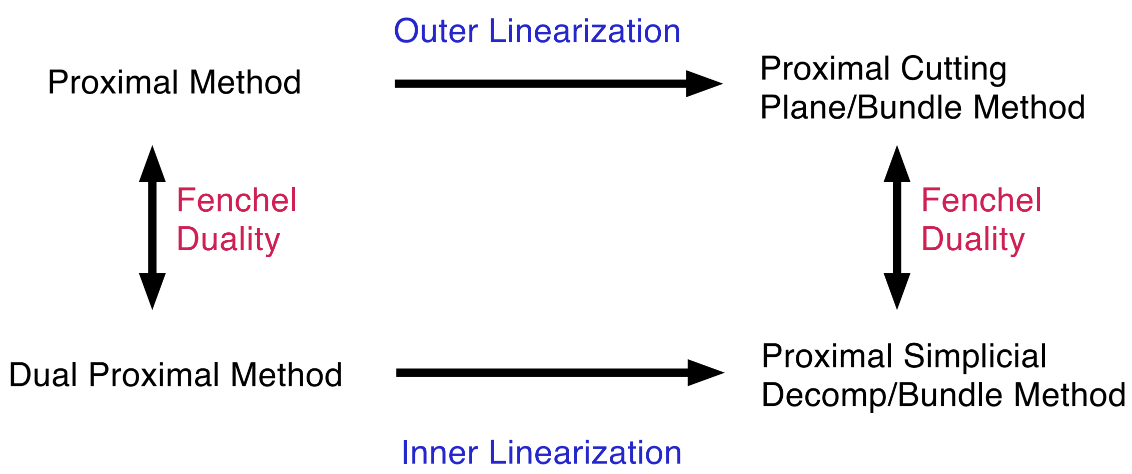
LECTURE 20

LECTURE OUTLINE

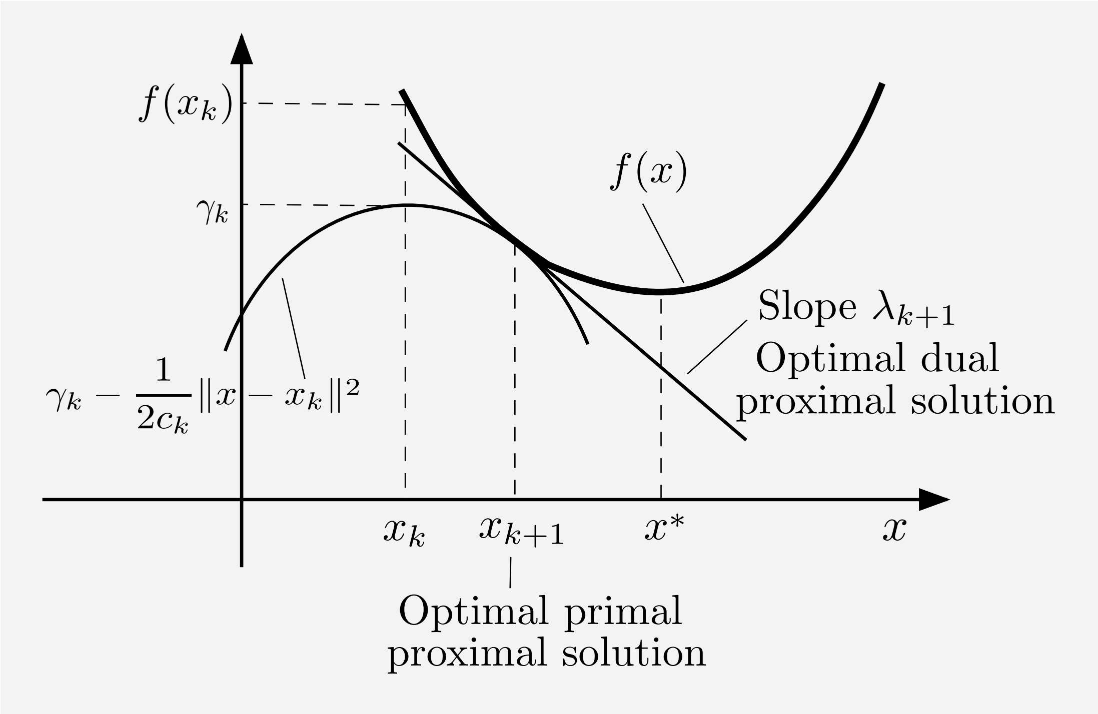
- Proximal methods
- Review of Proximal Minimization
- Proximal cutting plane algorithm
- Bundle methods
- Augmented Lagrangian Methods
- Dual Proximal Minimization Algorithm

****

- 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:

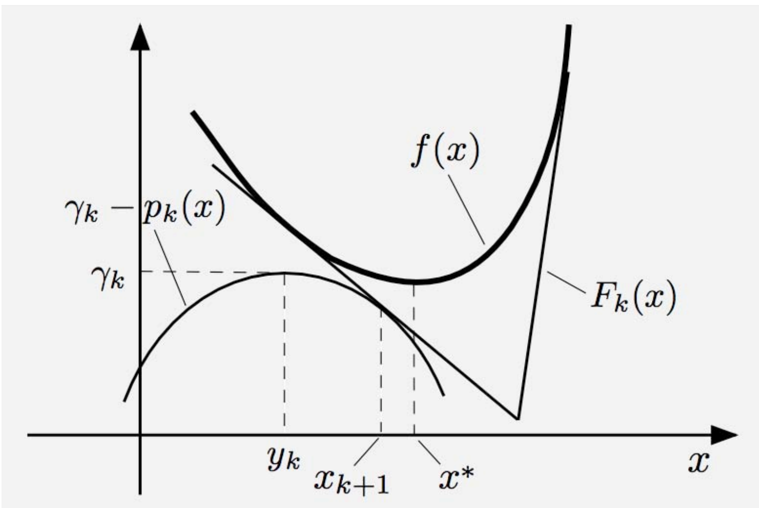
$$x_{k+1} \in \arg\min_{x \in X} \big\{ F_k(x) + p_k(x) \big\}$$

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

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

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

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

- 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 $y_k \neq x_k$:

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

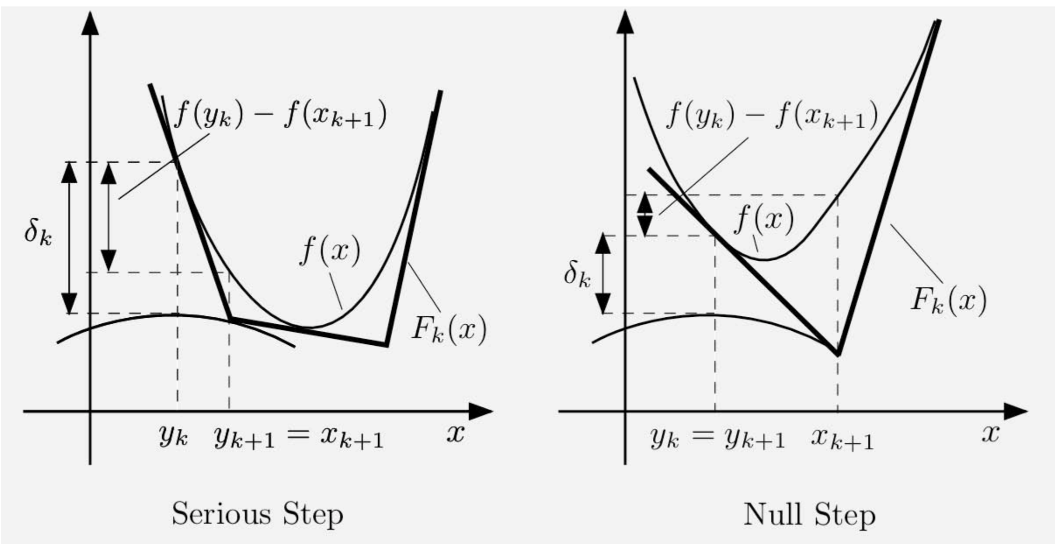
$$F_k(x) = \max\big\{f(x_0)+(x-x_0)'g_0,\ldots,f(x_k)+(x-x_k)'g_k\big\}$$

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

- **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 $\text{ri}\big(\text{dom}(f_1)\big) \cap \text{ri}\big(\text{dom}(f_2)\big) \neq \emptyset$, 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

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

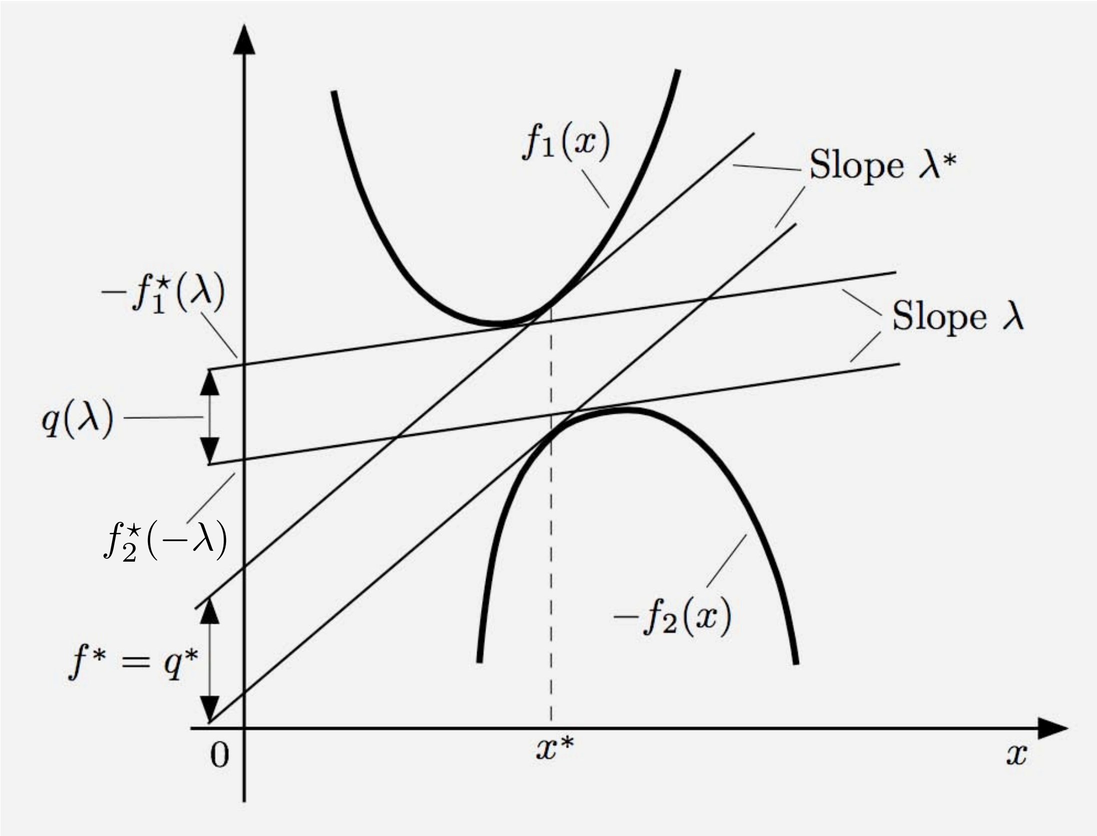
- By Fenchel inequality, the last condition is equivalent to

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

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

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

- 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:

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

- 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,

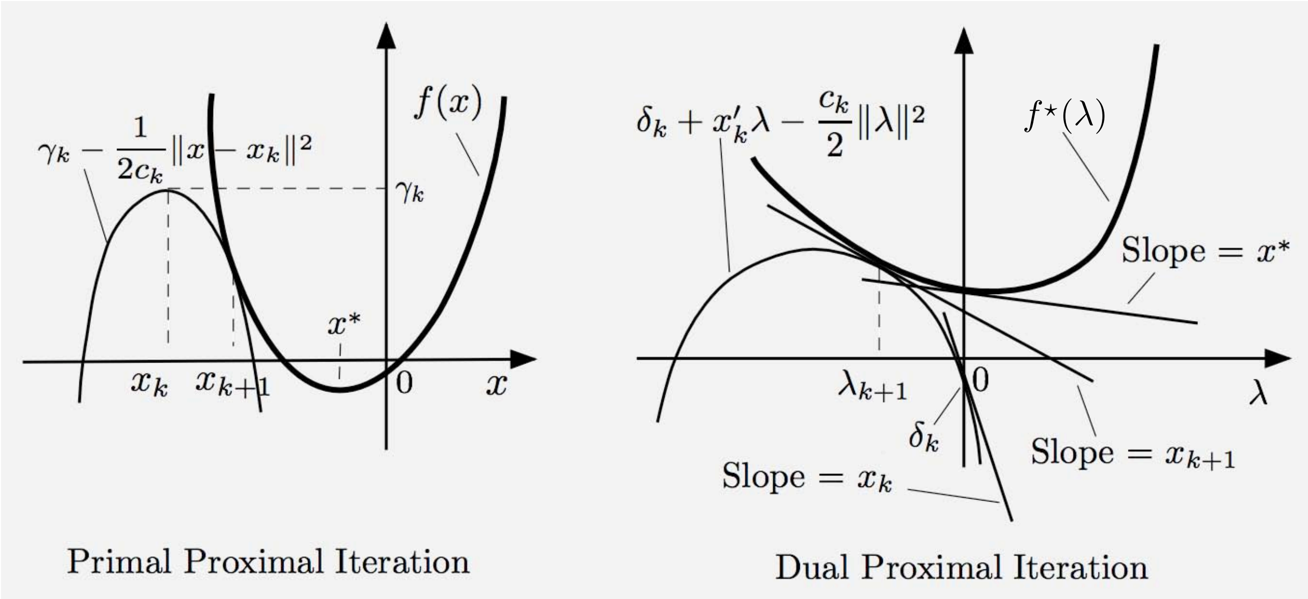
$$\lambda_{k+1} \in \partial f(x_{k+1}), \qquad \lambda_{k+1} = \frac{x_k - x_{k+1}}{c_k}$$

- **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 $\min_{g(x)\leq 0} F(x)$, 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 $p(u) = \min_{g(x)\leq u} F(x)$ at $u = 0$ (i.e., a dual optimal solution).

AUGMENTED LAGRANGIAN METHOD

- Consider the convex constrained problem

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

- 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

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

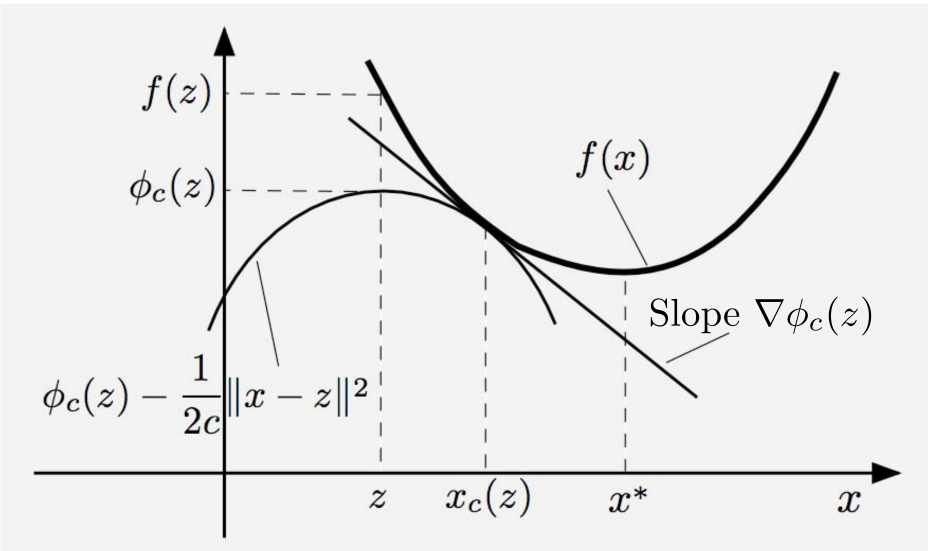
GRADIENT INTERPRETATION

- 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

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

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

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

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

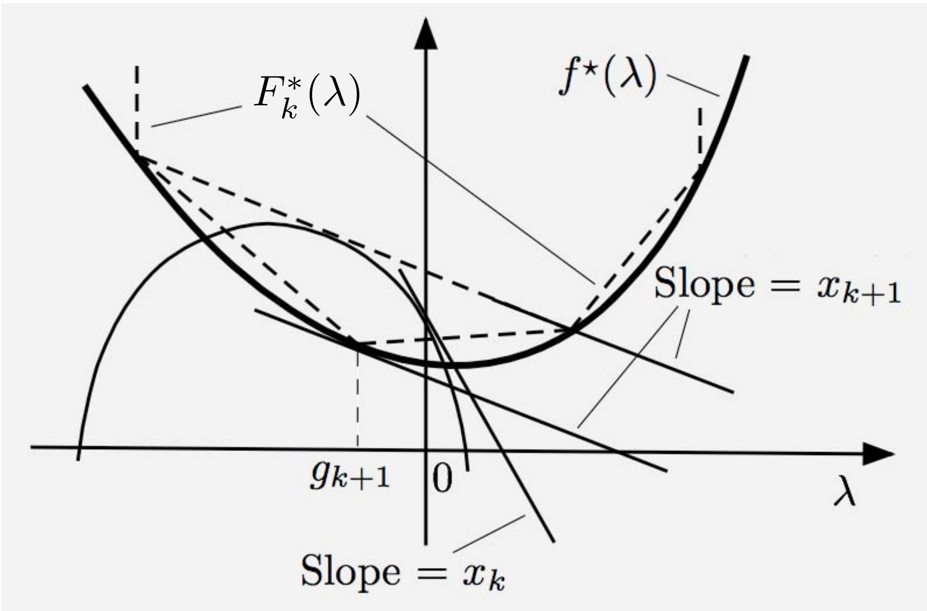
Obtain $g_{k+1} \in \partial f(x_{k+1})$, either directly or via

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

- 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**.

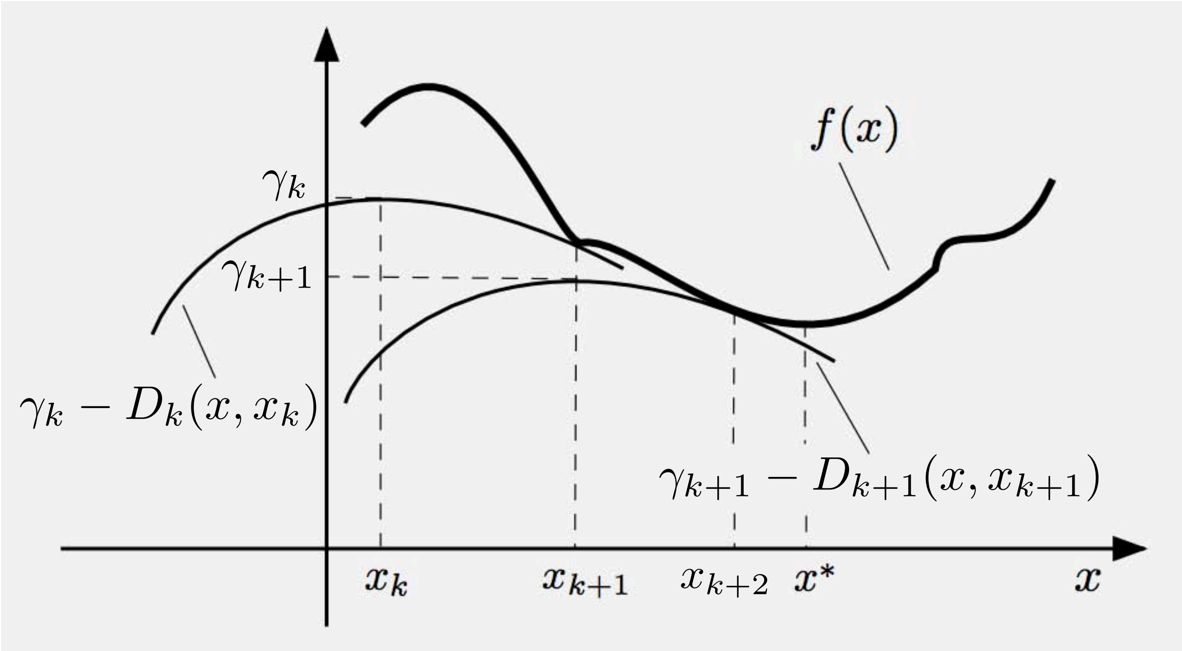
LECTURE 21

LECTURE OUTLINE

- Generalized forms of the proximal point algorithm
- Interior point methods
- Constrained optimization case - Barrier method
- Conic programming cases

GENERALIZED PROXIMAL ALGORITHM

- Replace quadratic regularization by more general proximal term.
- Minimize possibly nonconvex $f :\mapsto (-\infty, \infty]$.



- Introduce a general regularization term $D_k : \Re^{2n} \mapsto (-\infty, \infty]$:

$$x_{k+1} \in \arg \min_{x \in \Re^n} \left\{ f(x) + D_k(x, x_k) \right\}$$

- Assume attainment of min (but this is not automatically guaranteed)
- Complex/unreliable behavior when f is nonconvex

SOME GUARANTEES ON GOOD BEHAVIOR

- Assume

$$D_k(x, x_k) \geq D_k(x_k, x_k), \qquad \forall\, x \in \Re^n,\ k \qquad (1)$$

Then we have a cost improvement property:

$$\begin{aligned} f(x_{k+1}) &\leq f(x_{k+1}) + D_k(x_{k+1}, x_k) - D_k(x_k, x_k) \\ &\leq f(x_k) + D_k(x_k, x_k) - D_k(x_k, x_k) \\ &= f(x_k) \end{aligned}$$

- Assume algorithm stops only when x_k in optimal solution set X^*, i.e.,

$$x_k \in \arg\min_{x\in\Re^n} \big\{ f(x) + D_k(x, x_k) \big\} \quad \Rightarrow \quad x_k \in X^*$$

- Then strict cost improvement for $x_k \notin X^*$
- Guaranteed if f is convex and
  - (a) $D_k(\cdot, x_k)$ satisfies (1), and is convex and differentiable at x_k
  - (b) We have

$$\mathrm{ri}\big(\mathrm{dom}(f)\big) \cap \mathrm{ri}\big(\mathrm{dom}(D_k(\cdot, x_k))\big) \neq \emptyset$$

EXAMPLE METHODS

- Bregman distance function

$$D_k(x, y) = \frac{1}{c_k}\Big(\phi(x) - \phi(y) - \nabla\phi(y)'(x - y)\Big),$$

where $\phi : \Re^n \mapsto (-\infty, \infty]$ is a convex function, differentiable within an open set containing $\mathrm{dom}(f)$, and c_k is a positive penalty parameter.

- Majorization-Minimization algorithm:

$$D_k(x, y) = M_k(x, y) - M_k(y, y),$$

where M satisfies

$$M_k(y, y) = f(y), \qquad \forall\ y \in \Re^n,\ k = 0, 1,$$

$$M_k(x, x_k) \geq f(x_k), \qquad \forall\ x \in \Re^n,\ k = 0, 1, \ldots$$

- Example for case $f(x) = R(x) + \|Ax - b\|^2$, where R is a convex regularization function

$$M(x, y) = R(x) + \|Ax - b\|^2 - \|Ax - Ay\|^2 + \|x - y\|^2$$

- Expectation-Maximization (EM) algorithm (special context in inference, f nonconvex)

INTERIOR POINT METHODS

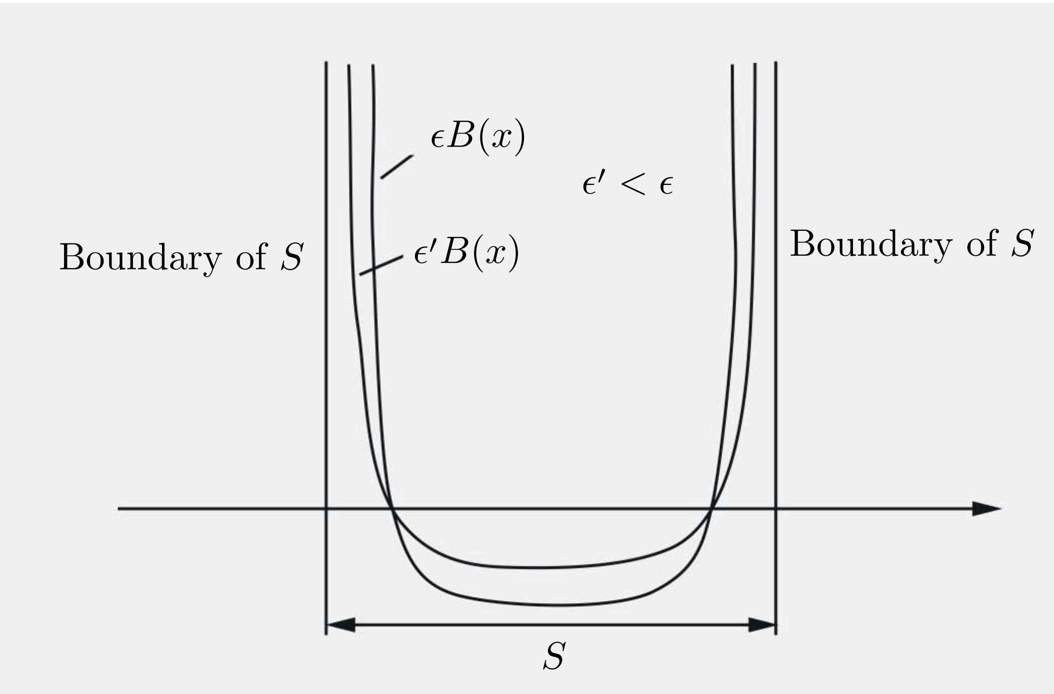
- Consider $\min f(x)$ s. t. $g_j(x) \le 0,\ j = 1, \ldots, r$
- A **barrier function**, that is continuous and goes to ∞ as any one of the constraints $g_j(x)$ approaches 0 from negative values; e.g.,

$$B(x) = -\sum_{j=1}^{r} \ln\{-g_j(x)\}, \qquad B(x) = -\sum_{j=1}^{r} \frac{1}{g_j(x)}.$$

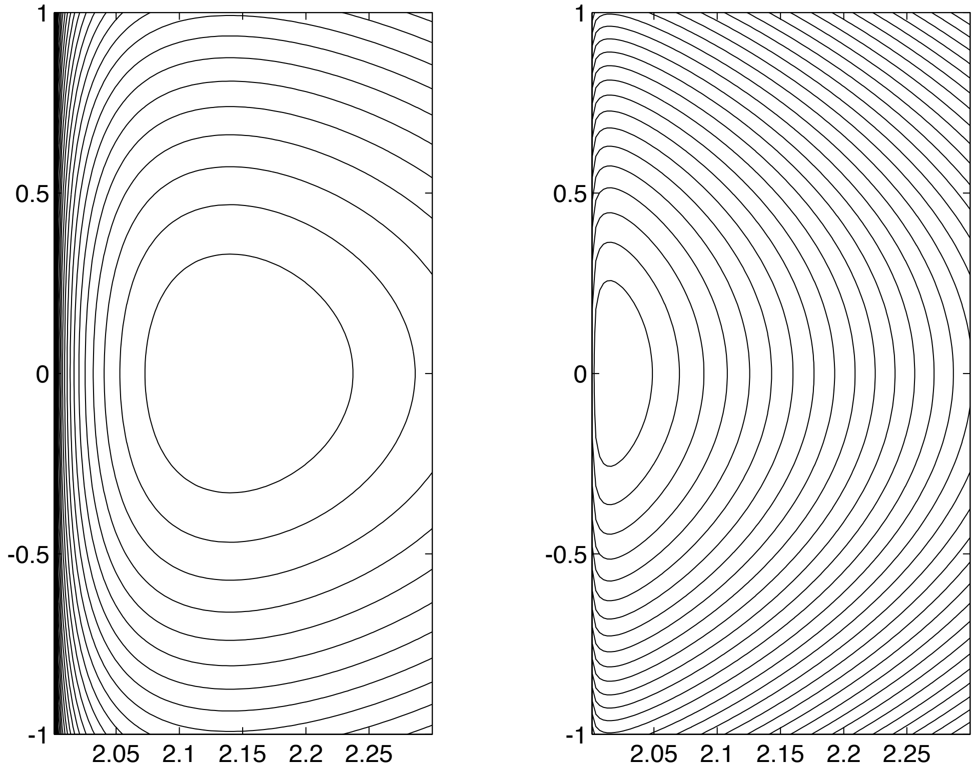
- **Barrier method:** Let

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

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 $\epsilon_k \to 0$.



BARRIER METHOD - EXAMPLE



$$\text{minimize} \quad f(x) = \tfrac{1}{2}\big((x^1)^2 + (x^2)^2\big)$$
$$\text{subject to} \quad 2 \le x^1,$$

with optimal solution $x^* = (2, 0)$.

- Logarithmic barrier: $B(x) = -\ln(x^1 - 2)$
- We have $x_k = \big(1 + \sqrt{1 + \epsilon_k}\,, 0\big)$ from

$$x_k \in \arg\min_{x^1 > 2} \left\{ \tfrac{1}{2}\big((x^1)^2 + (x^2)^2\big) - \epsilon_k \ln(x^1 - 2) \right\}$$

- As ϵ_k is decreased, the unconstrained minimum x_k approaches the constrained minimum $x^* = (2, 0)$.
- As $\epsilon_k \to 0$, computing x_k becomes more difficult because of ill-conditioning (a Newton-like method is essential for solving the approximate problems).

CONVERGENCE

- Every limit point of a sequence $\{x_k\}$ generated by a barrier method is a minimum of the original constrained problem.

Proof: Let $\{\overline{x}\}$ be the limit of a subsequence $\{x_k\}_{k\in K}$. Since $x_k \in S$ and X is closed, $\overline{x}$ is feasible for the original problem.

If $\overline{x}$ is not a minimum, there exists a feasible x^* such that $f(x^*) < f(\overline{x})$ and therefore also an interior point $\tilde{x} \in S$ such that $f(\tilde{x}) < f(\overline{x})$. By the definition of x_k,

$$f(x_k) + \epsilon_k B(x_k) \le f(\tilde{x}) + \epsilon_k B(\tilde{x}), \qquad \forall\ k,$$

so by taking limit

$$f(\overline{x}) + \liminf_{k\to\infty,\, k\in K} \epsilon_k B(x_k) \le f(\tilde{x}) < f(\overline{x})$$

Hence $\liminf_{k\to\infty,\, k\in K} \epsilon_k B(x_k) < 0$.

If $\overline{x} \in S$, we have $\lim_{k\to\infty,\, k\in K} \epsilon_k B(x_k) = 0$, while if $\overline{x}$ lies on the boundary of S, we have by assumption $\lim_{k\to\infty,\, k\in K} B(x_k) = \infty$. Thus

$$\liminf_{k\to\infty} \epsilon_k B(x_k) \ge 0,$$

– a contradiction.

SECOND ORDER CONE PROGRAMMING

- Consider the SOCP

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \;\; A_i x - b_i \in C_i,\;\; i = 1, \ldots, m, \end{aligned}$$

where $x \in \Re^n$, c is a vector in $\Re^n$, and for $i = 1, \ldots, m$, A_i is an $n_i \times n$ matrix, b_i is a vector in $\Re^{n_i}$, and C_i is the second order cone of $\Re^{n_i}$.

- We approximate this problem with

$$\begin{aligned} &\text{minimize} \quad c'x + \epsilon_k \sum_{i=1}^{m} B_i(A_i x - b_i) \\ &\text{subject to} \;\; x \in \Re^n, \quad A_i x - b_i \in \text{int}(C_i),\;\; i = 1, \ldots, m, \end{aligned}$$

where B_i is the logarithmic barrier function:

$$B_i(y) = -\ln\Big(y_{n_i}^2 - (y_1^2 + \cdots + y_{n_i-1}^2)\Big), \quad y \in \text{int}(C_i),$$

and $\{\epsilon_k\}$ is a positive sequence with $\epsilon_k \to 0$.

- Essential to use Newton's method to solve the approximating problems.
- Interesting complexity analysis

SEMIDEFINITE PROGRAMMING

- Consider the dual SDP

$$\begin{array}{ll} \text{maximize} & b'\lambda \\ \text{subject to} & D - (\lambda_1 A_1 + \cdots + \lambda_m A_m) \in C, \end{array}$$

where $b \in \Re^m$, $D, A_1, \ldots, A_m$ are symmetric matrices, and C is the cone of positive semidefinite matrices.

- The logarithmic barrier method uses approximating problems of the form

$$\text{maximize} \quad b'\lambda + \epsilon_k \ln\Big(\det(D - \lambda_1 A_1 - \cdots - \lambda_m A_m)\Big)$$

over all $\lambda \in \Re^m$ such that $D - (\lambda_1 A_1 + \cdots + \lambda_m A_m)$ is positive definite.

- Here $\epsilon_k > 0$ and $\epsilon_k \to 0$.

- Furthermore, we should use a starting point such that $D - \lambda_1 A_1 - \cdots - \lambda_m A_m$ is positive definite, and Newton's method should ensure that the iterates keep $D - \lambda_1 A_1 - \cdots - \lambda_m A_m$ within the positive definite cone.

LECTURE 22

LECTURE OUTLINE

- Incremental methods
- Review of large sum problems
- Review of incremental gradient and subgradient methods
- Combined incremental subgradient and proximal methods
- Convergence analysis
- Cyclic and randomized component selection
- References:

(1) D. P. Bertsekas, "Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey", Lab. for Information and Decision Systems Report LIDS-P-2848, MIT, August 2010

(2) Published versions in Math. Programming J., and the edited volume "Optimization for Machine Learning," by S. Sra, S. Nowozin, and S. J. Wright, MIT Press, Cambridge, MA, 2012.

LARGE SUM PROBLEMS

- Minimize over $X \subset \Re^n$

$$f(x) = \sum_{i=1}^{m} f_i(x), \qquad m \text{ is very large,}$$

where X, f_i are convex. Some examples:

- **Dual cost of a separable problem.**

- **Data analysis/machine learning:** x is parameter vector of a model; each f_i corresponds to error between data and output of the model.
  - Least squares problems (f_i quadratic).
  - ℓ_1-regularization (least squares plus ℓ_1 penalty):

$$\min_x \gamma \sum_{j=1}^{n} |x^j| + \sum_{i=1}^{m} (c_i'x - d_i)^2$$

The nondifferentiable penalty tends to set a large number of components of x to 0.

- **Min of an expected value** $\min_x E\big\{F(x,w)\big\}$ **- Stochastic programming:**

$$\min_x \left[F_1(x) + E_w\{\min_y F_2(x,y,w)\big\} \right]$$

- **More** (many constraint problems, distributed incremental optimization ...)

INCREMENTAL SUBGRADIENT METHODS

- The special structure of the sum

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

can be exploited by incremental methods.

- We first consider incremental subgradient methods which **move x along a subgradient $\tilde{\nabla} f_i$ of a component function f_i** NOT the (expensive) subgradient of f, which is $\sum_i \tilde{\nabla} f_i$.

- At iteration k select a component i_k and set

$$x_{k+1} = P_X\Big(x_k - \alpha_k \tilde{\nabla} f_{i_k}(x_k)\Big),$$

with $\tilde{\nabla} f_{i_k}(x_k)$ being a subgradient of f_{i_k} at x_k.

- **Motivation is faster convergence.** A cycle can make much more progress than a subgradient iteration with essentially the same computation.

CONVERGENCE PROCESS: AN EXAMPLE

- **Example 1:** Consider

$$\min_{x \in \Re} \frac{1}{2}\Big\{(1-x)^2 + (1+x)^2\Big\}$$

- Constant stepsize: Convergence to a limit cycle
- Diminishing stepsize: Convergence to the optimal solution
- **Example 2:** Consider

$$\min_{x \in \Re} \Big\{|1-x| + |1+x| + |x|\Big\}$$

- Constant stepsize: Convergence to a limit cycle that depends on the starting point
- Diminishing stepsize: Convergence to the optimal solution
- What is the effect of the order of component selection?

CONVERGENCE: CYCLIC ORDER

- Algorithm

$$x_{k+1} = P_X\Big(x_k - \alpha_k \tilde\nabla f_{i_k}(x_k)\Big)$$

- Assume all subgradients generated by the algorithm are bounded: $\|\tilde\nabla f_{i_k}(x_k)\| \le c$ for all k
- Assume components are chosen for iteration in cyclic order, and stepsize is constant within a cycle of iterations (for all k with $i_k = 1$ we have $\alpha_k = \alpha_{k+1} = \ldots = \alpha_{k+m-1}\big)$
- **Key inequality:** For all $y \in X$ and all k that mark the beginning of a cycle

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

- Result for a constant stepsize $\alpha_k \equiv \alpha$:

$$\liminf_{k\to\infty} f(x_k) \le f^* + \alpha\frac{m^2c^2}{2}$$

- Convergence for $\alpha_k \downarrow 0$ with $\sum_{k=0}^{\infty} \alpha_k = \infty$.

CONVERGENCE: RANDOMIZED ORDER

- Algorithm

$$x_{k+1} = P_X\left(x_k - \alpha_k \tilde{\nabla} f_{i_k}(x_k)\right)$$

- Assume component i_k chosen for iteration in randomized order (independently with equal probability)
- Assume all subgradients generated by the algorithm are bounded: $\|\tilde{\nabla} f_{i_k}(x_k)\| \leq c$ for all k
- Result for a constant stepsize $\alpha_k \equiv \alpha$:

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \alpha \frac{mc^2}{2}$$

(with probability 1)

- Convergence for $\alpha_k \downarrow 0$ with $\sum_{k=0}^{\infty} \alpha_k = \infty$. (with probability 1)
- In practice, randomized stepsize and variations (such as randomization of the order within a cycle at the start of a cycle) often work much faster

PROXIMAL-SUBGRADIENT CONNECTION

- **Key Connection:** The proximal iteration

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

can be written as

$$x_{k+1} = P_X\big(x_k - \alpha_k \tilde{\nabla} f(x_{k+1})\big)$$

where $\tilde{\nabla} f(x_{k+1})$ is *some* subgradient of f at x_{k+1}.

- Consider an incremental proximal iteration for $\min_{x \in X} \sum_{i=1}^m f_i(x)$

$$x_{k+1} = \arg\min_{x \in X} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\}$$

- **Motivation:** Proximal methods are more "stable" than subgradient methods
- **Drawback:** Proximal methods require special structure to avoid large overhead
- This motivates a combination of incremental subgradient and proximal

INCR. SUBGRADIENT-PROXIMAL METHODS

- Consider the problem

$$\min_{x \in X} F(x) \stackrel{\text{def}}{=} \sum_{i=1}^{m} F_i(x)$$

where for all i,

$$F_i(x) = f_i(x) + h_i(x)$$

X, f_i and h_i are convex.

- We consider combinations of subgradient and proximal incremental iterations

$$z_k = \arg\min_{x \in X} \left\{ f_{i_k}(x) + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\}$$

$$x_{k+1} = P_X \big(z_k - \alpha_k \tilde{\nabla} h_{i_k}(z_k) \big)$$

- Variations:
  - Min. over $\Re^n$ (rather than X) in proximal
  - Do the subgradient without projection first and then the proximal

- **Idea:** Handle "favorable" components f_i with the more stable proximal iteration; handle other components h_i with subgradient iteration.

CONVERGENCE: CYCLIC ORDER

- Assume all subgradients generated by the algorithm are bounded: $\|\tilde{\nabla} f_{i_k}(x_k)\| \leq c$, $\|\tilde{\nabla} h_{i_k}(x_k)\| \leq c$ for all k, plus mild additional conditions
- Assume components are chosen for iteration in cyclic order, and stepsize is constant within a cycle of iterations
- **Key inequality:** For all $y \in X$ and all k that mark the beginning of a cycle:

$$\|x_{k+m}-y\|^2 \leq \|x_k-y\|^2 - 2\alpha_k\big(F(x_k)-F(y)\big) + \beta\alpha_k^2 m^2 c^2$$

where β is a (small) constant

- Result for a constant stepsize $\alpha_k \equiv \alpha$:

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \alpha\beta\frac{m^2c^2}{2}$$

- Convergence for $\alpha_k \downarrow 0$ with $\sum_{k=0}^{\infty} \alpha_k = \infty$.

CONVERGENCE: RANDOMIZED ORDER

- Result for a constant stepsize $\alpha_k \equiv \alpha$:

$$\liminf_{k\to\infty} f(x_k) \le f^* + \alpha\beta\frac{mc^2}{2}$$

(with probability 1)

- Convergence for $\alpha_k \downarrow 0$ with $\sum_{k=0}^{\infty} \alpha_k = \infty$. (with probability 1)

EXAMPLE

- ℓ_1-Regularization for least squares with large number of terms

$$\min_{x \in \Re^n} \left\{ \gamma \|x\|_1 + \frac{1}{2} \sum_{i=1}^{m} (c_i' x - d_i)^2 \right\}$$

- Use incremental gradient or proximal on the quadratic terms
- Use proximal on the $\|x\|_1$ term:

$$z_k = \arg \min_{x \in \Re^n} \left\{ \gamma \|x\|_1 + \frac{1}{2\alpha_k} \|x - x_k\|^2 \right\}$$

- Decomposes into the n one-dimensional minimizations

$$z_k^j = \arg \min_{x^j \in \Re} \left\{ \gamma \, |x^j| + \frac{1}{2\alpha_k} |x^j - x_k^j|^2 \right\},$$

and can be done in closed form

$$z_k^j = \begin{cases} x_k^j - \gamma\alpha_k & \text{if } \gamma\alpha_k \le x_k^j, \\ 0 & \text{if } -\gamma\alpha_k < x_k^j < \gamma\alpha_k, \\ x_k^j + \gamma\alpha_k & \text{if } x_k^j \le -\gamma\alpha_k. \end{cases}$$

- Note that “small” coordinates x_k^j are set to 0.

LECTURE 23

LECTURE OUTLINE

- Review of subgradient methods
- Application to differentiable problems - Gradient projection
- Iteration complexity issues
- Complexity of gradient projection
- Projection method with extrapolation
- Optimal algorithms

****

- Reference: The on-line chapter of the textbook

SUBGRADIENT METHOD

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

- **Subgradient method - constant step α:**

$$x_{k+1} = P_X\Big(x_k - \alpha_k \tilde{\nabla} f(x_k)\Big),$$

where $\tilde{\nabla} f(x_k)$ is a subgradient of f at x_k, and $P_X(\cdot)$ is projection on X.

- Assume $\|\tilde{\nabla} f(x_k)\| \leq c$ for all k.

- **Key inequality:** For all optimal x^*

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

- Convergence to a neighborhood result:

$$\liminf_{k \to \infty} f(x_k) \leq f^* + \frac{\alpha c^2}{2}$$

- Iteration complexity result: For any $\epsilon > 0$,

$$\min_{0 \leq k \leq K} f(x_k) \leq f^* + \frac{\alpha c^2 + \epsilon}{2},$$

where $K = \left\lfloor \frac{\min_{x^* \in X^*} \|x_0 - x^*\|^2}{\alpha \epsilon} \right\rfloor$.

- For $\alpha = \epsilon / c^2$, we need $O(1/\epsilon^2)$ iterations to get within ϵ of the optimal value f^*.

GRADIENT PROJECTION METHOD

- Let f be differentiable and assume

$$\big\|\nabla f(x) - \nabla f(y)\big\| \leq L\,\|x-y\|, \qquad \forall\ x,y \in X$$

- **Gradient projection method:**

$$x_{k+1} = P_X\big(x_k - \alpha\nabla f(x_k)\big)$$

- Define the *linear approximation function at x*

$$\ell(y;x) = f(x) + \nabla f(x)'(y-x), \qquad y \in \Re^n$$

- **First key inequality:** For all $x, y \in X$

$$f(y) \leq \ell(y;x) + \frac{L}{2}\|y-x\|^2$$

- Using the projection theorem to write

$$\big(x_k - \alpha\nabla f(x_k) - x_{k+1}\big)'(x_k - x_{k+1}) \leq 0,$$

and then the 1st key inequality, we have

$$f(x_{k+1}) \leq f(x_k) - \left(\frac{1}{\alpha} - \frac{L}{2}\right)\|x_{k+1} - x_k\|^2$$

so there is cost reduction for $\alpha \in \big(0, \frac{2}{L}\big)$

ITERATION COMPLEXITY

- Connection with proximal algorithm

$$y = \arg\min_{z\in X}\left\{\ell(z;x) + \frac{1}{2\alpha}\|z-x\|^2\right\} = P_X\big(x-\alpha\nabla f(x)\big)$$

- **Second key inequality:** For any $x \in X$, if $y = P_X\big(x - \alpha\nabla f(x)\big)$, then for all $z \in X$, we have

$$\ell(y;x)+\frac{1}{2\alpha}\|y-x\|^2 \le \ell(z;x)+\frac{1}{2\alpha}\|z-x\|^2-\frac{1}{2\alpha}\|z-y\|^2$$

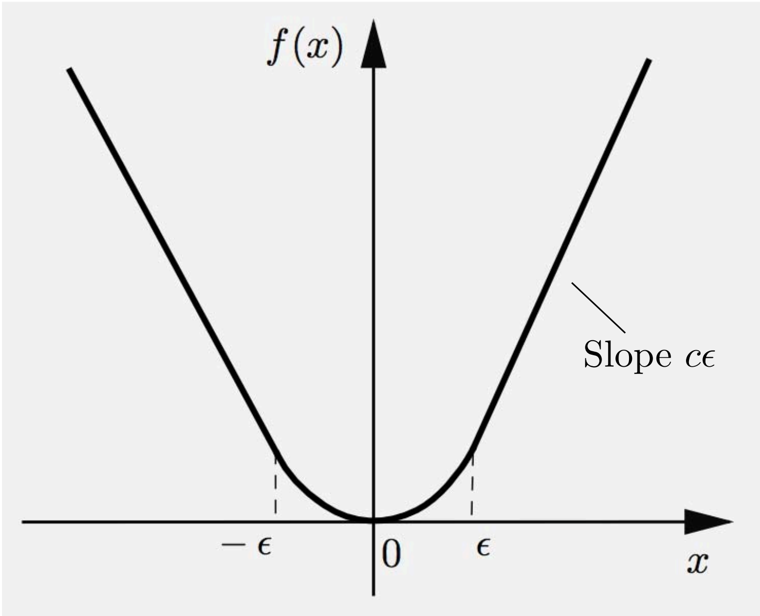
- **Complexity Estimate:** Let the stepsize of the method be $\alpha = 1/L$. Then for all k

$$f(x_k) - f^* \le \frac{L\min_{x^*\in X^*}\|x_0 - x^*\|^2}{2k}$$

- Thus, we need $O(1/\epsilon)$ iterations to get within ϵ of f^*. **Better than nondifferentiable case.**
- Practical implementation/same complexity: Start with some α and reduce it by some factor as many times as necessary to get

$$f(x_{k+1}) \le \ell(x_{k+1};x_k) + \frac{1}{2\alpha}\|x_{k+1} - x_k\|^2$$

SHARPNESS OF COMPLEXITY ESTIMATE



- Unconstrained minimization of

$$f(x) = \begin{cases} \frac{c}{2}|x|^2 & \text{if } |x| \le \epsilon, \\ c\epsilon|x| - \frac{c\epsilon^2}{2} & \text{if } |x| > \epsilon \end{cases}$$

- With stepsize $\alpha = 1/L = 1/c$ and any $x_k > \epsilon$,

$$x_{k+1} = x_k - \frac{1}{L}\nabla f(x_k) = x_k - \frac{1}{c}c\,\epsilon = x_k - \epsilon$$

- The number of iterations to get within an ϵ-neighborhood of $x^* = 0$ is $|x_0|/\epsilon$.
- The number of iterations to get to within ϵ of $f^* = 0$ is proportional to $1/\epsilon$ for large x_0.

EXTRAPOLATION VARIANTS

• An old method for unconstrained optimization, known as the *heavy-ball* method or gradient method with *momentum*:

$$x_{k+1} = x_k - \alpha \nabla f(x_k) + \beta(x_k - x_{k-1}),$$

where $x_{-1} = x_0$ and β is a scalar with $0 < \beta < 1$.

• A variant of this scheme for constrained problems separates the extrapolation and the gradient steps:

$$y_k = x_k + \beta(x_k - x_{k-1}), \quad \text{(extrapolation step)},$$
$$x_{k+1} = P_X\big(y_k - \alpha \nabla f(y_k)\big), \quad \text{(grad. projection step)}.$$

• When applied to the preceding example, the method converges to the optimum, and reaches a neighborhood of the optimum more quickly

• However, the method still has an $O(1/\epsilon)$ iteration complexity, since for $x_0 >> 1$, we have

$$x_{k+1} - x_k = \beta(x_k - x_{k-1}) - \epsilon$$

so $x_{k+1} - x_k \approx \epsilon/(1 - \beta)$, and the number of iterations needed to obtain $x_k < \epsilon$ is $O\big((1 - \beta)/\epsilon\big)$.

OPTIMAL COMPLEXITY ALGORITHM

- Surprisingly with a proper more vigorous extrapolation $\beta_k \to 1$ in the extrapolation scheme

$$
\begin{aligned}
y_k &= x_k + \beta_k(x_k - x_{k-1}), && \text{(extrapolation step)},\\
x_{k+1} &= P_X\big(y_k - \alpha \nabla f(y_k)\big), && \text{(grad. projection step)},
\end{aligned}
$$

the method has iteration complexity $O\big(1/\sqrt{\epsilon}\big)$.

- Choices that work

$$\beta_k = \frac{\theta_k(1-\theta_{k-1})}{\theta_{k-1}}$$

where the sequence $\{\theta_k\}$ satisfies $\theta_0 = \theta_1 \in (0,1]$, and

$$\frac{1-\theta_{k+1}}{\theta_{k+1}^2} \le \frac{1}{\theta_k^2}, \qquad \theta_k \le \frac{2}{k+2}$$

- One possible choice is

$$\beta_k = \begin{cases} 0 & \text{if } k = 0,\\ \frac{k-1}{k+2} & \text{if } k \ge 1, \end{cases} \qquad \theta_k = \begin{cases} 1 & \text{if } k = -1,\\ \frac{2}{k+2} & \text{if } k \ge 0. \end{cases}$$

- Highly unintuitive. Good performance reported.

EXTENSION TO NONDIFFERENTIABLE CASE

- Consider the nondifferentiable problem of minimizing convex function $f : \Re^n \mapsto \Re$ over a closed convex set X.

- Approach: "Smooth" f, i.e., approximate it with a differentiable function by using a proximal minimization scheme.

- Apply optimal complexity gradient projection method with extrapolation. Then an $O(1/\epsilon)$ iteration complexity algorithm is obtained.

- Can be shown that this complexity bound is sharp.

- Improves on the subgradient complexity bound by a an ϵ factor.

- Limited experience with such methods.

- Major disadvantage: Cannot take advantage of special structure, e.g., there are no incremental versions.

LECTURE 24

LECTURE OUTLINE

- Gradient proximal minimization method
- Nonquadratic proximal algorithms
- Entropy minimization algorithm
- Exponential augmented Lagrangian mehod
- Entropic descent algorithm

****

References:

- Beck, A., and Teboulle, M., 2010. "Gradient-Based Algorithms with Applications to Signal Recovery Problems, in Convex Optimization in Signal Processing and Communications (Y. Eldar and D. Palomar, eds.), Cambridge University Press, pp. 42-88.

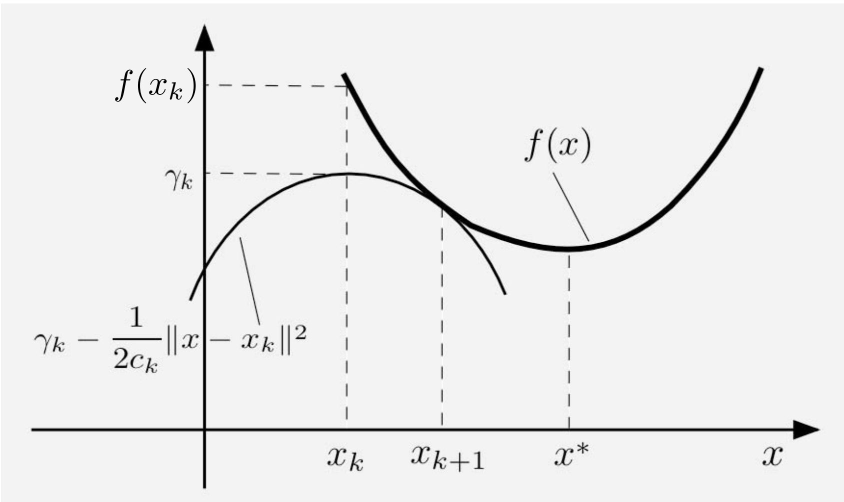
- Beck, A., and Teboulle, M., 2003. "Mirror Descent and Nonlinear Projected Subgradient Methods for Convex Optimization," Operations Research Letters, Vol. 31, pp. 167-175.

- Bertsekas, D. P., 1999. Nonlinear Programming, Athena Scientific, Belmont, MA.

PROXIMAL AND GRADIENT PROJECTION

- Proximal algorithm to minimize convex f over closed convex X

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



- Let f be differentiable and assume

$$\big\|\nabla f(x) - \nabla f(y)\big\| \leq L\,\|x - y\|, \qquad \forall\ x, y \in X$$

- Define the *linear approximation function at* x

$$\ell(y; x) = f(x) + \nabla f(x)'(y - x), \qquad y \in \Re^n$$

- Connection of proximal with gradient projection

$$y = \arg\min_{z \in X} \left\{ \ell(z; x) + \frac{1}{2\alpha} \|z - x\|^2 \right\} = P_X\big(x - \alpha \nabla f(x)\big)$$

GRADIENT-PROXIMAL METHOD I

- Minimize $f(x)+g(x)$ over $x \in X$, where X: closed convex, f, g: convex, f *is differentiable.*

- **Gradient-proximal method:**

$$x_{k+1} \in \arg\min_{x \in X} \left\{ \ell(x; x_k) + g(x) + \frac{1}{2\alpha} \|x - x_k\|^2 \right\}$$

- Recall key inequality: For all $x, y \in X$

$$f(y) \le \ell(y; x) + \frac{L}{2} \|y - x\|^2$$

- Cost reduction for $\alpha \le 1/L$:

$$\begin{aligned} f(x_{k+1}) + g(x_{k+1}) &\le \ell(x_{k+1}; x_k) + \frac{L}{2} \|x_{k+1} - x_k\|^2 + g(x_{k+1}) \\ &\le \ell(x_{k+1}; x_k) + g(x_{k+1}) + \frac{1}{2\alpha} \|x_{k+1} - x_k\|^2 \\ &\le \ell(x_k; x_k) + g(x_k) \\ &= f(x_k) + g(x_k) \end{aligned}$$

- This is a key insight for the convergence analysis.

GRADIENT-PROXIMAL METHOD II

- Equivalent definition of gradient-proximal:

$$z_k = x_k - \alpha \nabla f(x_k)$$

$$x_{k+1} \in \arg\min_{x \in X} \left\{ g(x) + \frac{1}{2\alpha} \|x - z_k\|^2 \right\}$$

- Simplifies the implementation of proximal, by using gradient iteration to deal with the case of an inconvenient component f
- This is similar to incremental subgradient-proximal method, but the gradient-proximal method does not extend to the case where the cost consists of the sum of multiple components.
- Allows a constant stepsize (under the restriction $\alpha \leq 1/L$). This does not extend to incremental methods.
- Like all gradient and subgradient methods, convergence can be slow.
- There are special cases where the method can be fruitfully applied (see the reference by Beck and Teboulle).

GENERALIZED PROXIMAL ALGORITHM

- Introduce a general regularization term D_k:

$$x_{k+1} \in \arg\min_{x \in X} \big\{ f(x) + D_k(x, x_k) \big\}$$

- **Example:** Bregman distance function

$$D_k(x, y) = \frac{1}{c_k} \Big(\phi(x) - \phi(y) - \nabla \phi(y)'(x - y) \Big),$$

where $\phi : \Re^n \mapsto (-\infty, \infty]$ is a convex function, differentiable within an open set containing $\text{dom}(f)$, and c_k is a positive penalty parameter.

- All the ideas for applications and connections of the quadratic form of the proximal algorithm extend to the nonquadratic case (although the analysis may not be trivial). In particular we have:
  - A dual proximal algorithm (based on Fenchel duality)
  - Equivalence with (nonquadratic) augmented Lagrangean method
  - Combinations with polyhedral approximations (bundle-type methods)
  - Incremental subgradient-proximal methods
  - Nonlinear gradient projection algorithms

ENTROPY MINIMIZATION ALGORITHM

- A special case involving entropy regularization:

$$x_{k+1} \in \arg\min_{x \in X} \left\{ f(x) + \frac{1}{c_k} \sum_{i=1}^{n} x^i \left(\ln\left(\frac{x^i}{x_k^i}\right) - 1 \right) \right\}$$

where x_0 and all subsequent x_k have positive components

- We use Fenchel duality to obtain a dual form of this minimization
- Note: The logarithmic function

$$p(x) = \begin{cases} x(\ln x - 1) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ \infty & \text{if } x < 0, \end{cases}$$

and the exponential function

$$p^\star(y) = e^y$$

are a conjugate pair.

- The dual problem is

$$y_{k+1} \in \arg\min_{y \in \Re^n} \left\{ f^\star(y) + \frac{1}{c_k} \sum_{i=1}^{n} x_k^i e^{c_k y^i} \right\}$$

EXPONENTIAL AUGMENTED LAGRANGIAN

- The dual proximal iteration is

$$x_{k+1}^i = x_k^i e^{c_k y_{k+1}^i}, \qquad i = 1, \ldots, n$$

where y_{k+1} is obtained from the dual proximal:

$$y_{k+1} \in \arg \min_{y \in \Re^n} \left\{ f^\star(y) + \frac{1}{c_k} \sum_{i=1}^{n} x_k^i e^{c_k y^i} \right\}$$

- A special case for the convex problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad g_1(x) \le 0, \ldots, g_r(x) \le 0, \quad x \in X \end{aligned}$$

is the **exponential augmented Lagrangean method**

- Consists of unconstrained minimizations

$$x_k \in \arg \min_{x \in X} \left\{ f(x) + \frac{1}{c_k} \sum_{j=1}^{r} \mu_k^j e^{c_k g_j(x)} \right\},$$

followed by the multiplier iterations

$$\mu_{k+1}^j = \mu_k^j e^{c_k^j g_j(x_k)}, \qquad j = 1, \ldots, r$$

NONLINEAR PROJECTION ALGORITHM

- Subgradient projection with general regularization term D_k:

$$x_{k+1} \in \arg\min_{x\in X} \Big\{ f(x_k) + \tilde{\nabla} f(x_k)'(x - x_k) + D_k(x, x_k) \Big\}$$

where $\tilde{\nabla} f(x_k)$ is a subgradient of f at x_k. Also called **mirror descent** method.

- Linearization of f simplifies the minimization
- The use of nonquadratic linearization is useful in problems with special structure
- **Entropic descent method:** Minimize $f(x)$ over the unit simplex $X = \big\{ x \geq 0 \mid \sum_{i=1}^n x^i = 1 \big\}$.
- Method:

$$x_{k+1} \in \arg\min_{x\in X} \sum_{i=1}^n x^i \left(g_k^i + \frac{1}{\alpha_k} \ln\left(\frac{x^i}{x_k^i}\right)\right)$$

where g_k^i are the components of $\tilde{\nabla} f(x_k)$.

- This minimization can be done in closed form:

$$x_{k+1}^i = \frac{x_k^i e^{-\alpha_k g_k^i}}{\sum_{j=1}^n x_k^j e^{-\alpha_k g_k^j}}, \qquad i = 1, \ldots, n$$

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

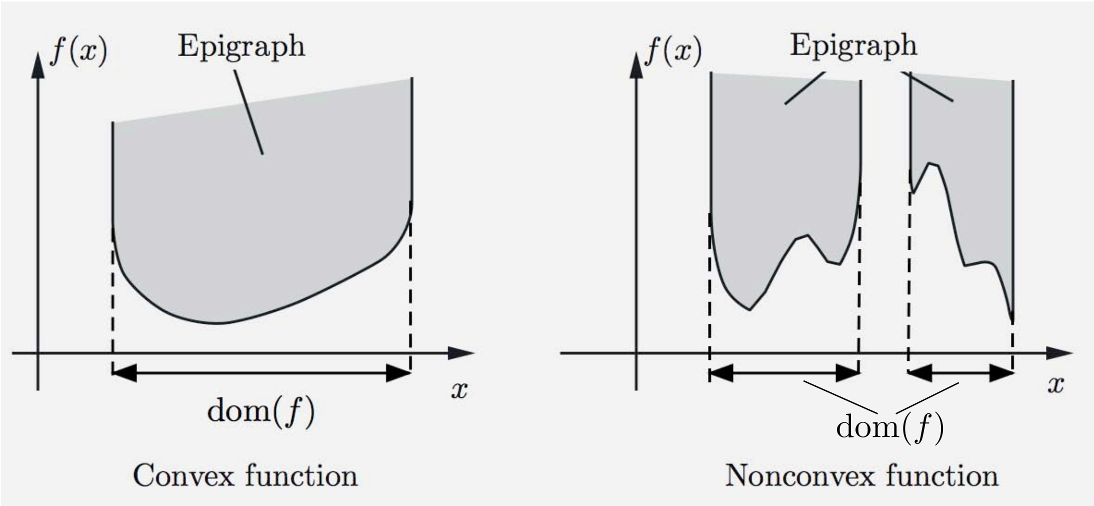
****

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

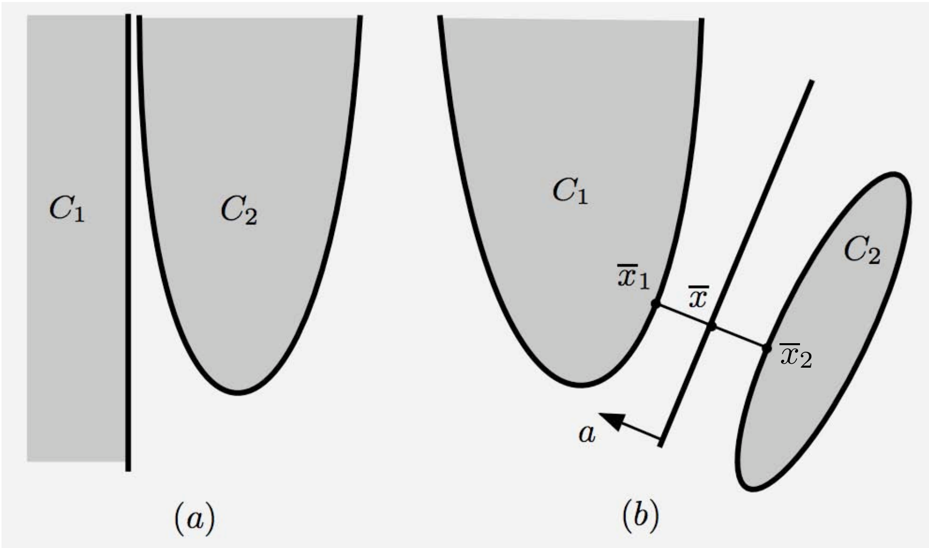
BASIC CONCEPTS OF CONVEX ANALYSIS

- Epigraphs, level sets, closedness, semicontinuity

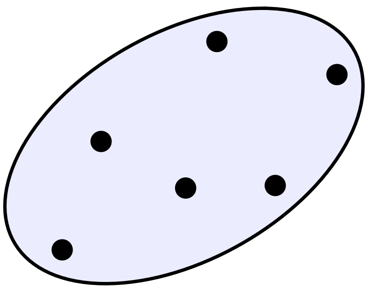


- Finite representations of generated cones and convex hulls - Caratheodory's Theorem.
- Relative interior:
  - Nonemptiness for a convex set
  - Line segment principle
  - Calculus of relative interiors
- Continuity of convex functions
- Nonemptiness of intersections of nested sequences of closed sets.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

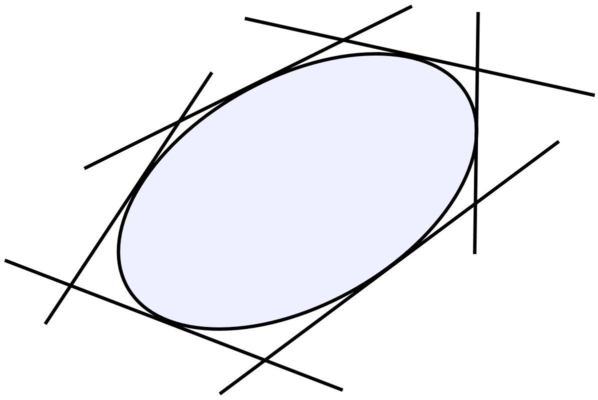
HYPERPLANE SEPARATION



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



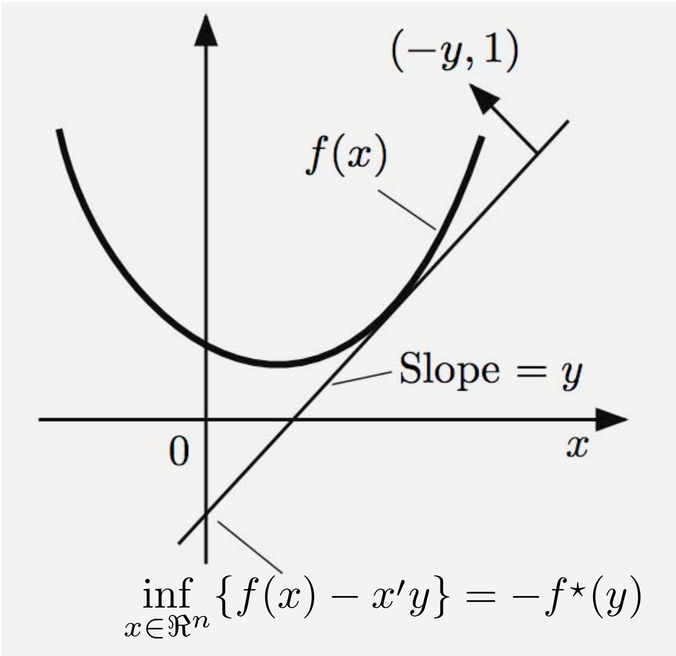
A union of points



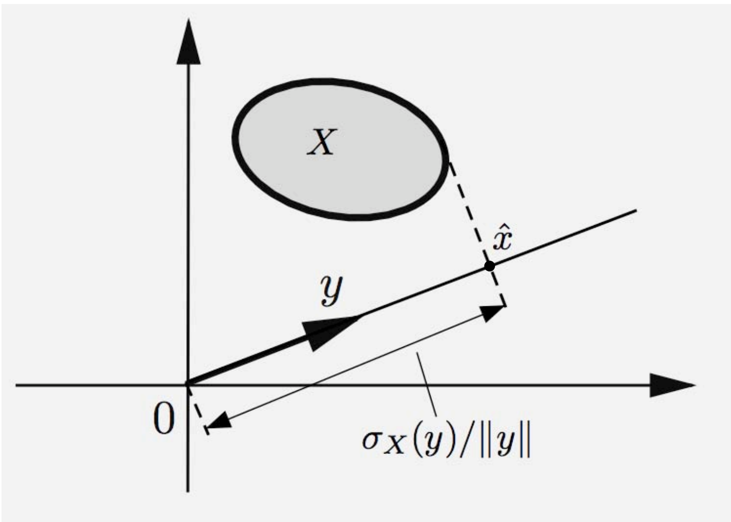
An intersection of halfspaces

- Nonvertical separating hyperplanes.

CONJUGATE FUNCTIONS



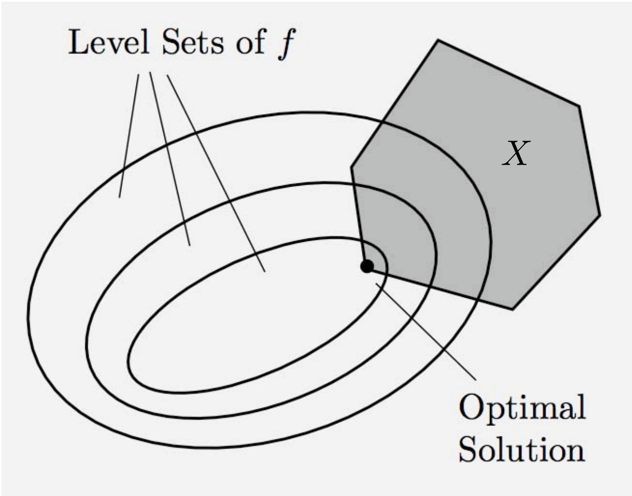
- Conjugacy theorem: $f = f^{\star\star}$
- Support functions



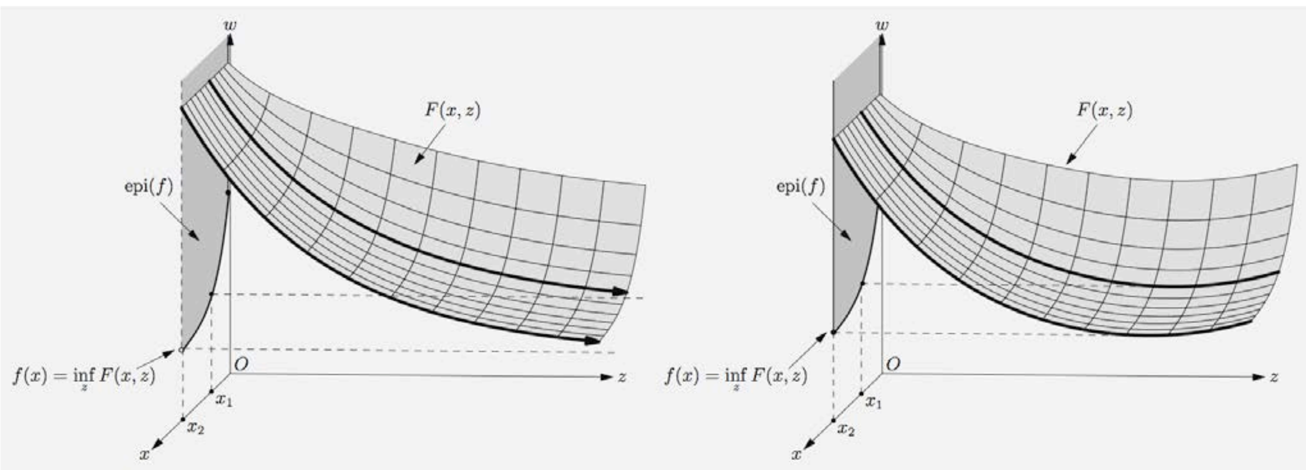
- Polar cone theorem: $C = C^{\star\star}$
  – Special case: Linear Farkas' lemma

BASIC CONCEPTS OF CONVEX OPTIMIZATION

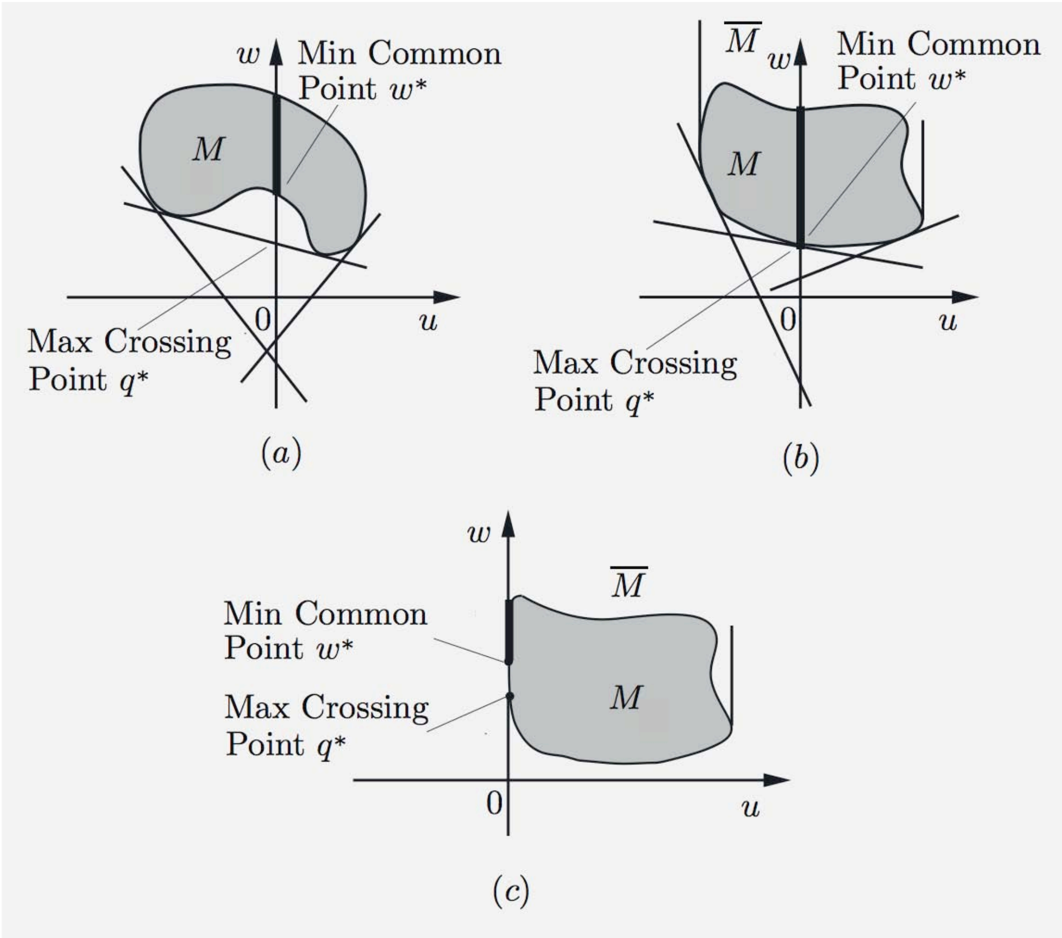
- **Weierstrass Theorem** and extensions.
- Characterization of existence of solutions in terms of nonemptiness of nested set intersections.



- 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: $q^* \le w^*$
- 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

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

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.

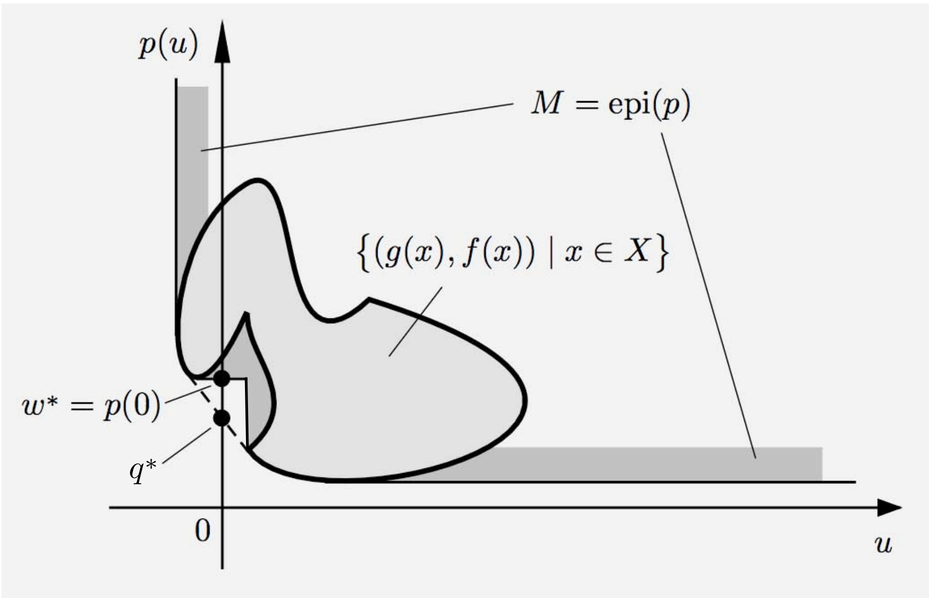
  (2) We have $\text{ri}(\tilde{D}) \cap P \neq \emptyset$, where

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

IMPORTANT SPECIAL CASE

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

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



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

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

NONLINEAR FARKAS' LEMMA

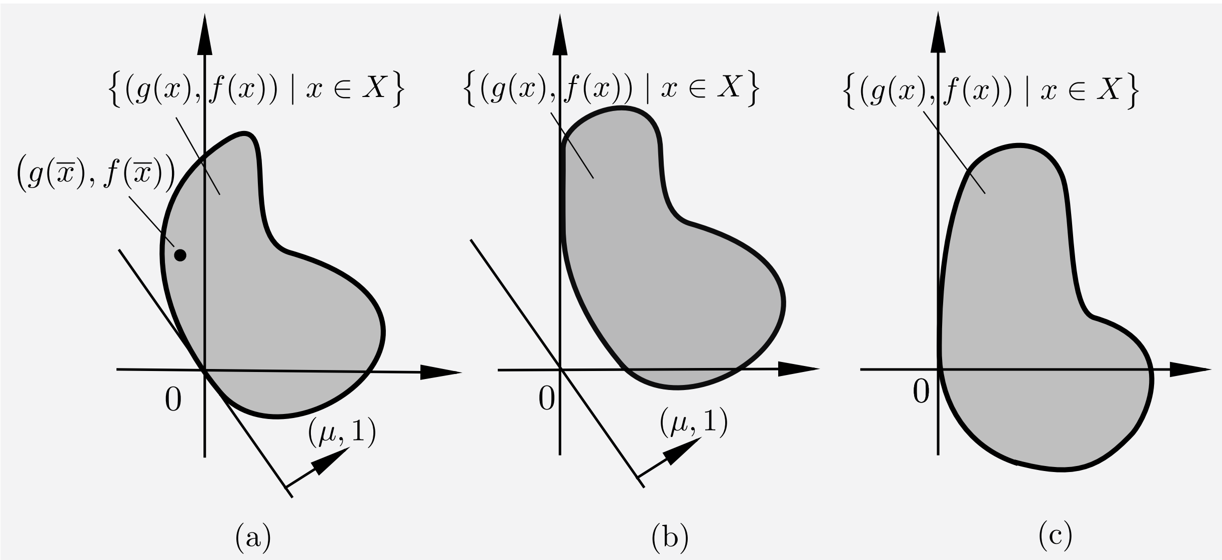
• Let $X \subset \Re^n$, $f : X \mapsto \Re$, and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, be convex. Assume that

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

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

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x \in X, \;\; g_j(x) \le 0, \;\; j = 1, \ldots, r, \end{aligned}$$

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 $p(u) = \inf_{x \in X,\, g(x) \le u} f(x)$
- **Dual function:**

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

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

- **Dual problem** of maximizing $q(\mu)$ over $\mu \ge 0$.
- **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 $g(\overline{x}) \le 0$.

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

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

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:**

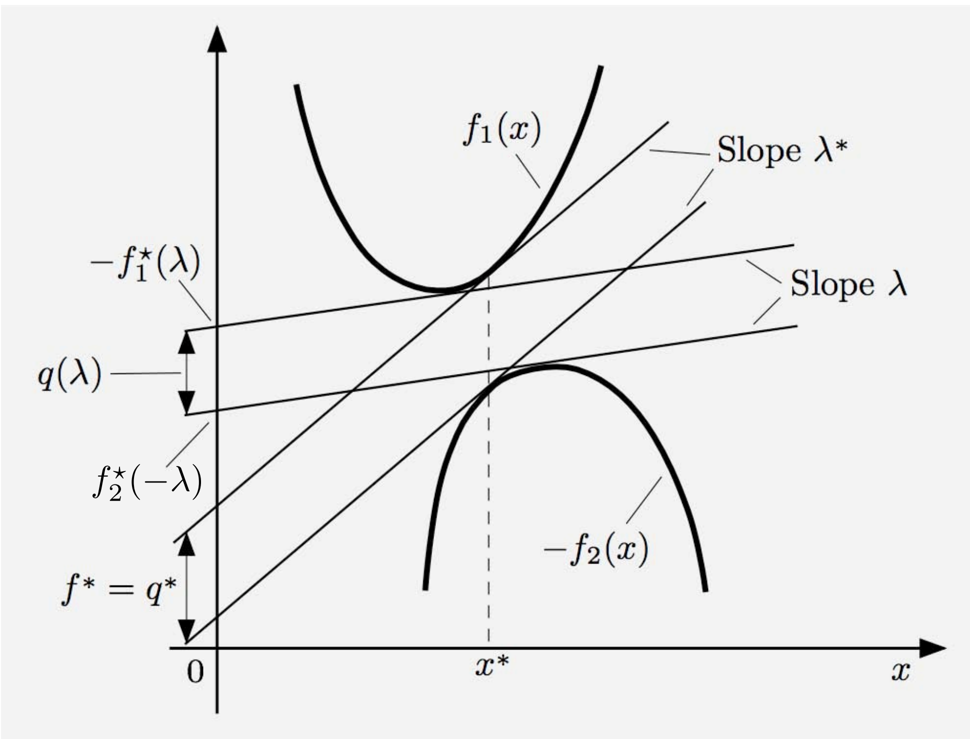
$$\begin{aligned} &\text{minimize} \quad f_1(x) + f_2(x) \\ &\text{subject to} \ \ x \in \Re^n, \end{aligned}$$

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:**

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

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

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

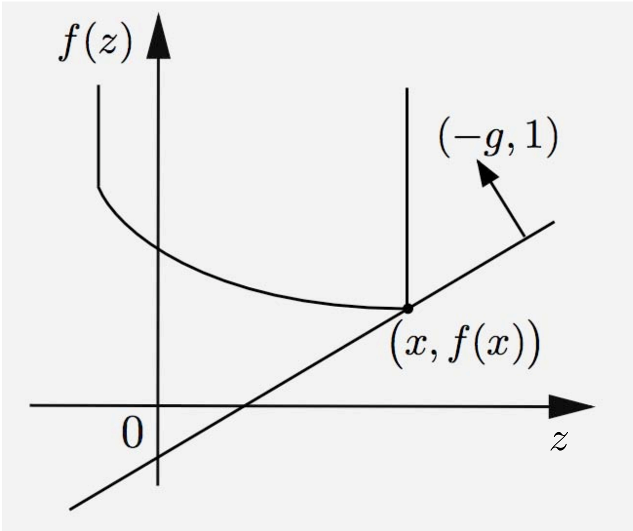
- **Special Linear-Conic Forms:**

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

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

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 $x \in \text{ri}\big(\text{dom}(f)\big)$.

- **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 $0 \in \text{ri}\big(\text{dom}(f^\star)\big)$.

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

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

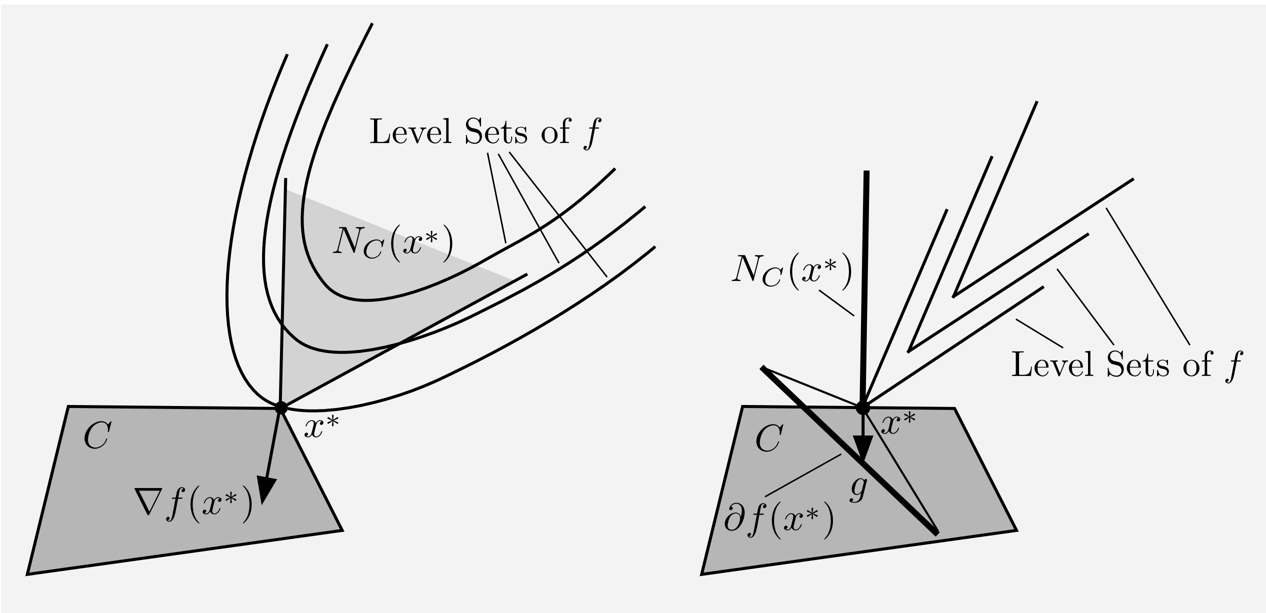
(ii) f is polyhedral and $\text{dom}(f) \cap \text{ri}(X) \neq \emptyset$.

(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.,

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

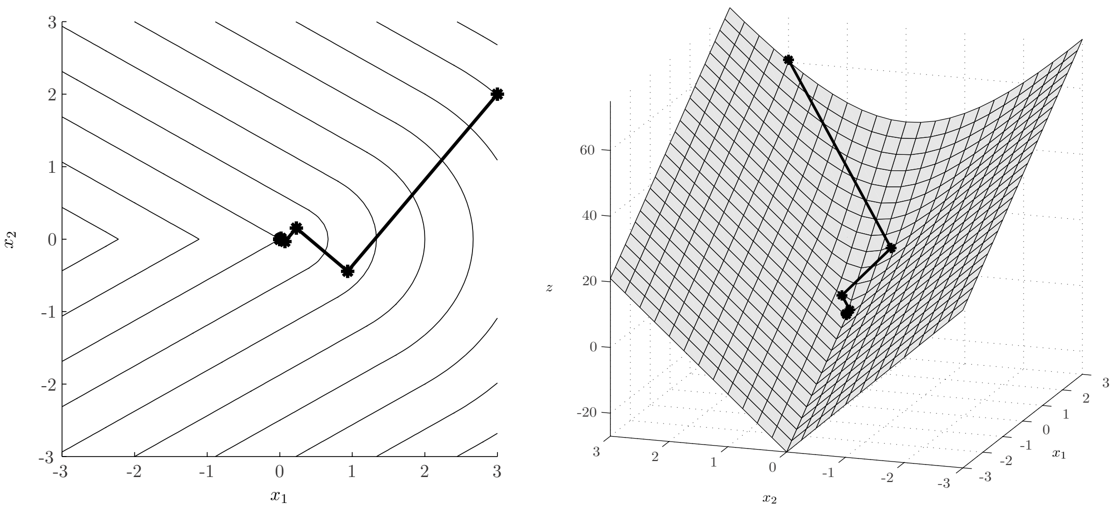


COMPUTATION: PROBLEM RANKING IN INCREASING COMPUTATIONAL DIFFICULTY

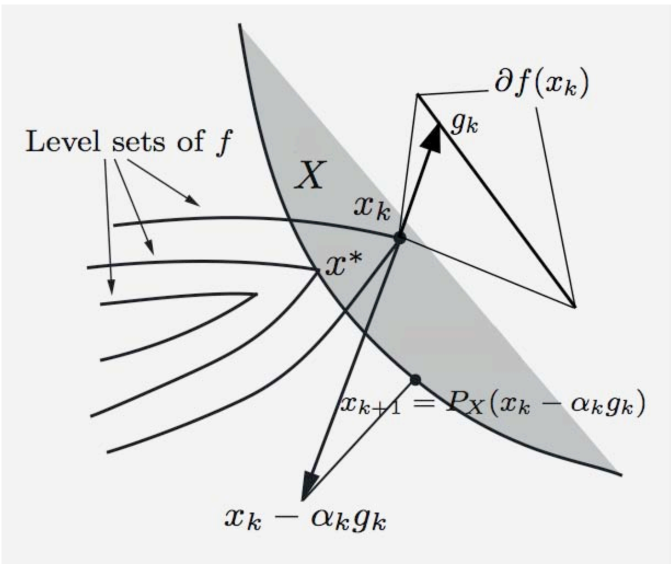
- Linear and (convex) quadratic programming.
  - Favorable special cases.
- Second order cone programming.
- Semidefinite programming.
- Convex programming.
  - Favorable cases, e.g., separable, large sum.
  - Geometric programming.
- Nonlinear/nonconvex/continuous programming.
  - Favorable special 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:**

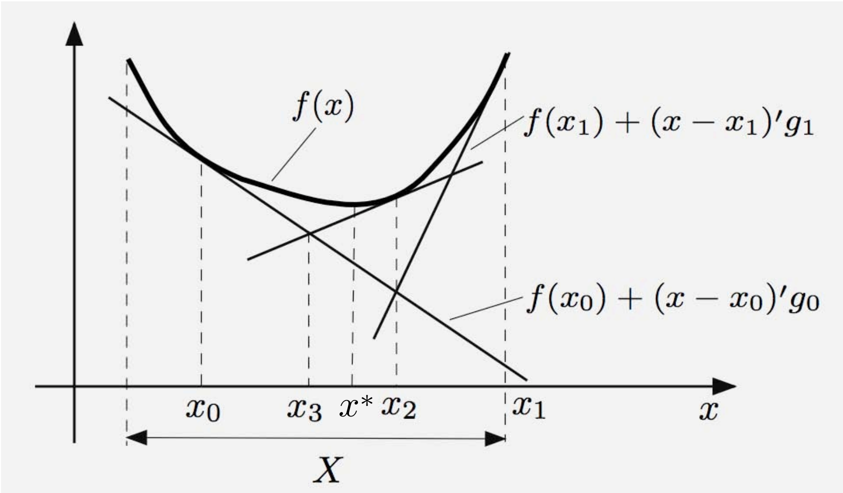


- **ϵ-subgradient method** (approx. subgradient)

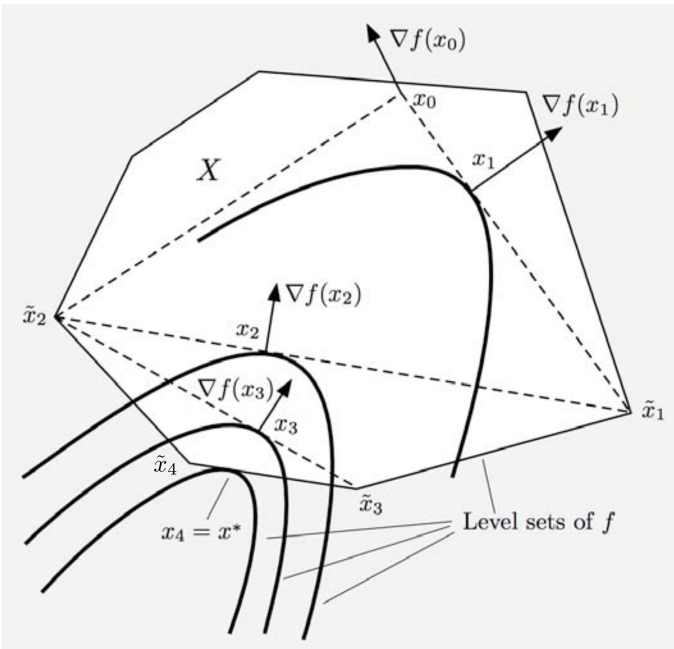
- **Incremental** (possibly randomized) variants for minimizing large sums (can be viewed as an approximate subgradient method).

OUTER AND INNER LINEARIZATION

- **Outer linearization:** Cutting plane



- **Inner linearization:** Simplicial decomposition



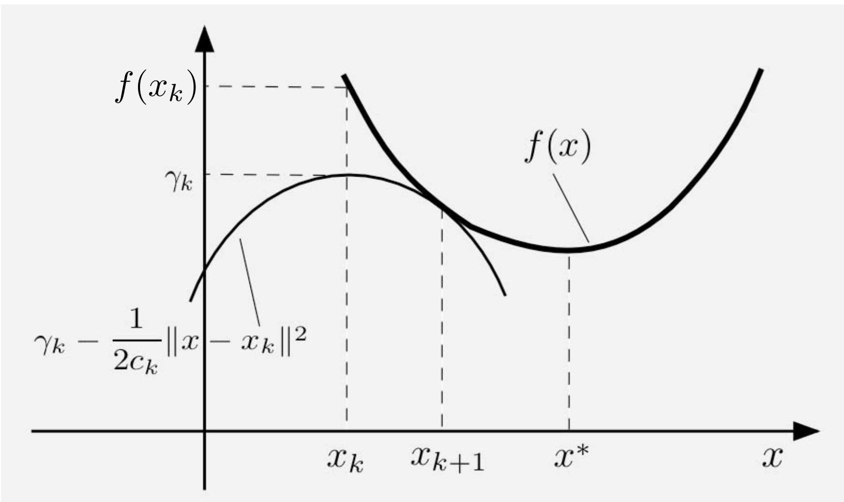
- Duality between outer and inner linearization.
  - *Extended monotropic programming* framework
  - Fenchel-like duality theory

PROXIMAL MINIMIZATION ALGORITHM

- A general algorithm for convex fn minimization

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

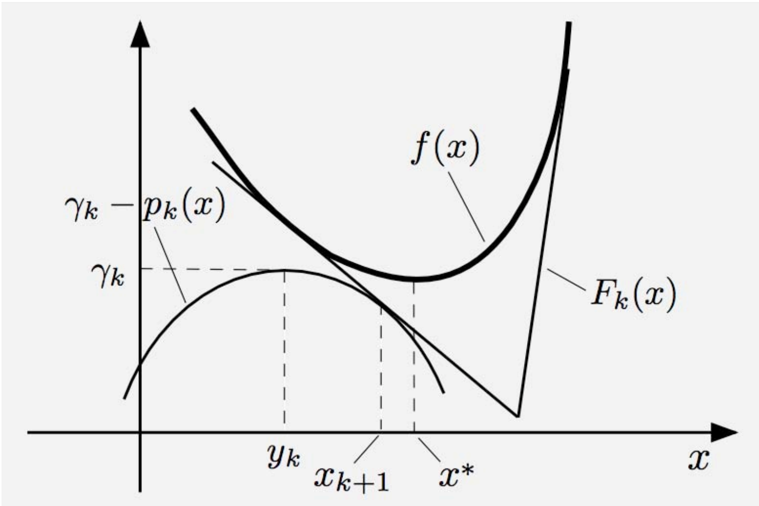
  - $f : \Re^n \mapsto (-\infty, \infty]$ is closed proper convex
  - c_k is a positive scalar parameter
  - x_0 is arbitrary starting point



- 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)

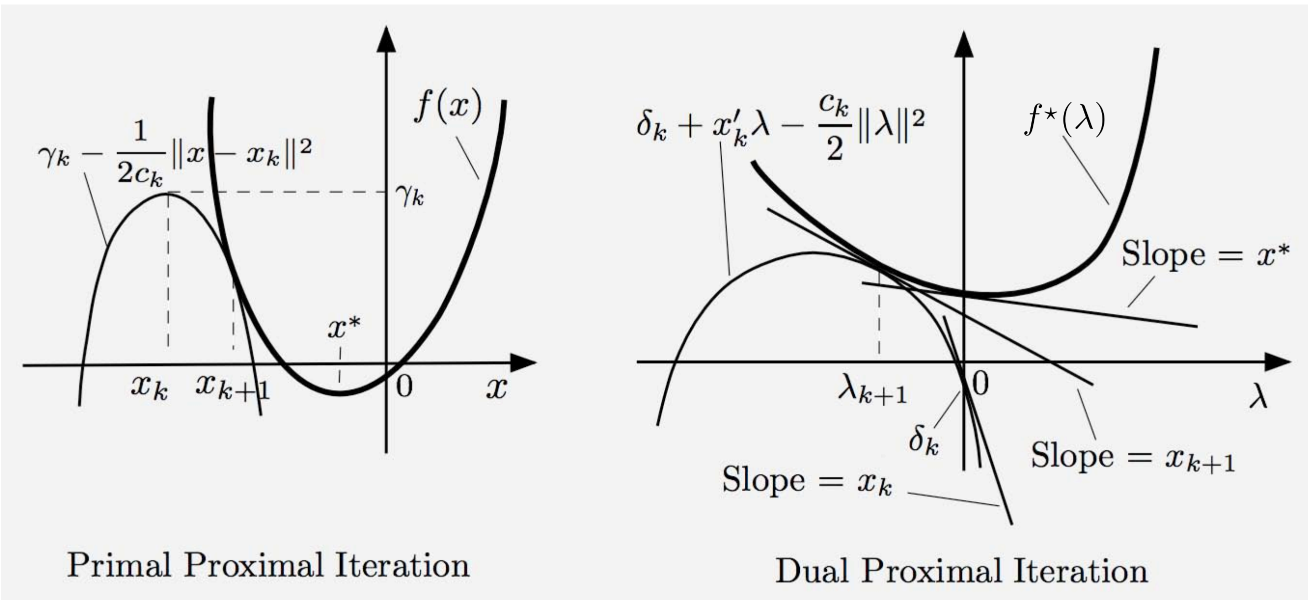
PROXIMAL-POLYHEDRAL METHODS

- **Proximal-cutting plane method**

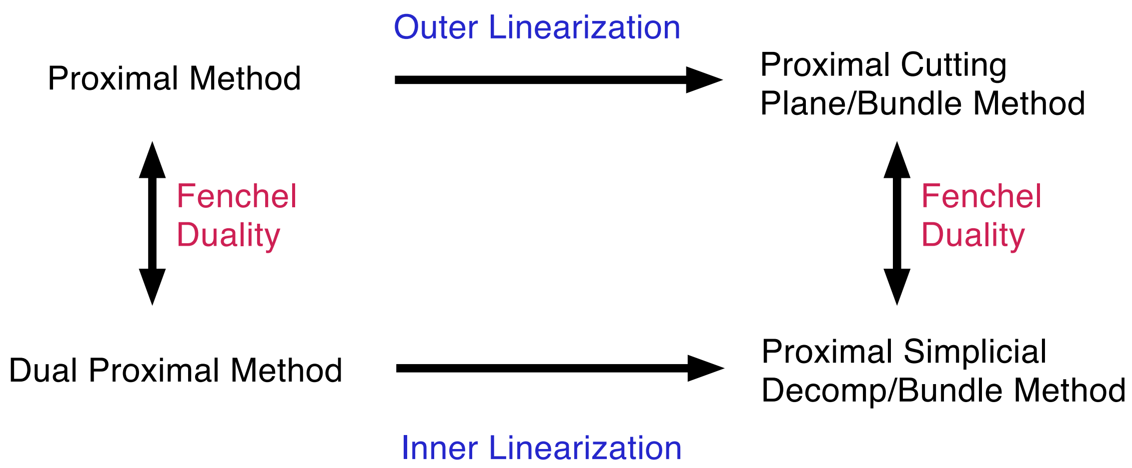


- **Proximal-cutting plane-bundle methods:** Replace f with a cutting plane approx. and/or change quadratic regularization more conservatively.

- **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

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

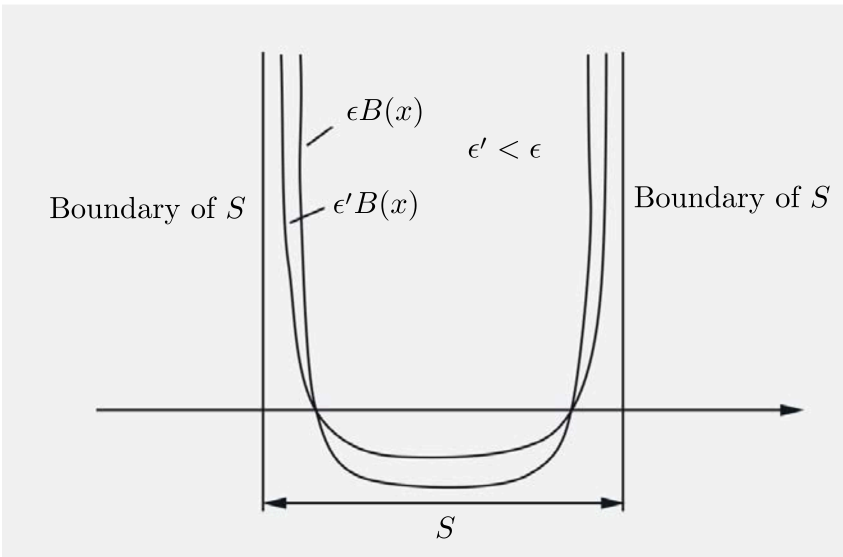
in the context of extended monotropic programming

INTERIOR POINT METHODS

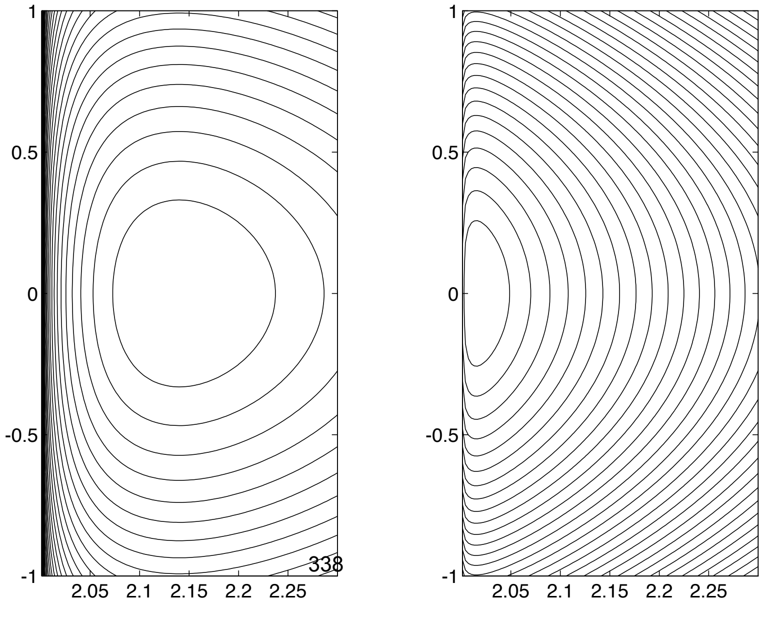
- **Barrier method:** Let

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

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 (non-quadratic) regularization - Bregman distance functions
  - Entropy-like regularization
  - Corresponding augmented Lagrangean method (exponential)
  - Corresponding gradient-proximal method
  - Nonlinear gradient/subgradient projection (entropic minimization methods)