LECTURE SLIDES ON

CONVEX ANALYSIS AND OPTIMIZATION

BASED ON 6.253 CLASS LECTURES AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

CAMBRIDGE, MASS

FALL 2007

BY DIMITRI P. BERTSEKAS

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

LECTURE 1
AN INTRODUCTION TO THE COURSE

LECTURE OUTLINE

- Convex and Nonconvex Optimization Problems
- Why is Convexity Important in Optimization
- Multipliers and Lagrangian Duality
- Min Common/Max Crossing Duality

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.,

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

- Examples of problem classifications:
  - Continuous vs discrete
  - Linear vs nonlinear
  - Deterministic vs stochastic
  - Static vs dynamic

- Convex programming problems are those for which f is convex and C is convex (they are continuous problems).

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

WHY IS CONVEXITY SO SPECIAL?

- A convex function has no local minima that are not global
- A convex set has a nonempty relative interior
- A convex set is connected and has feasible directions at any point
- A nonconvex function can be "convexified" while maintaining the optimality of its global minima
- 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

CONVEXITY AND DUALITY

- A multiplier vector for the problem

minimize $f(x)$ subject to $g_1(x) \leq 0, \ldots, g_r(x) \leq 0$

is a $\mu^* = (\mu_1^*, \ldots, \mu_r^*) \geq 0$ such that

$$\inf_{g_j(x) \leq 0,\, j=1,\ldots,r} f(x) = \inf_{x \in \Re^n} L(x, \mu^*)$$

where L is the Lagrangian function

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x), \qquad x \in \Re^n,\ \mu \in \Re^r.$$

- Dual function (always concave)

$$q(\mu) = \inf_{x \in \Re^n} L(x, \mu)$$

- Dual problem: Maximize $q(\mu)$ over $\mu \geq 0$

KEY DUALITY RELATIONS

- Optimal primal value

$$f^* = \inf_{g_j(x)\le 0,\, j=1,\ldots,r} f(x) = \inf_{x\in\Re^n} \sup_{\mu\ge 0} L(x,\mu)$$

- Optimal dual value

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

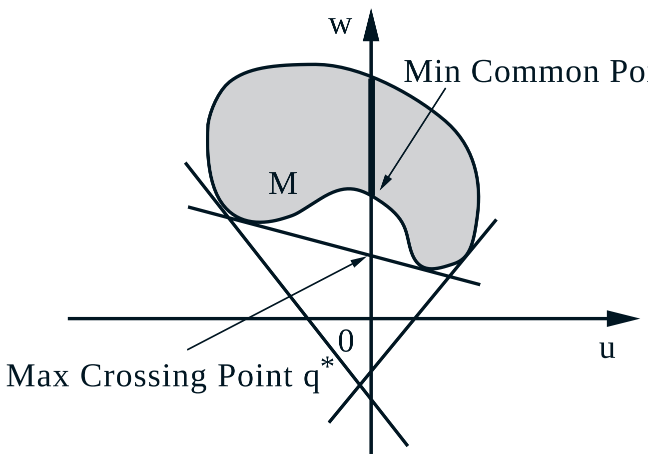
- We always have $q^* \le f^*$ (weak duality - important in discrete optimization problems).

- Under favorable circumstances (convexity in the primal problem, plus ...):
  - We have $q^* = f^*$
  - Optimal solutions of the dual problem are multipliers for the primal problem

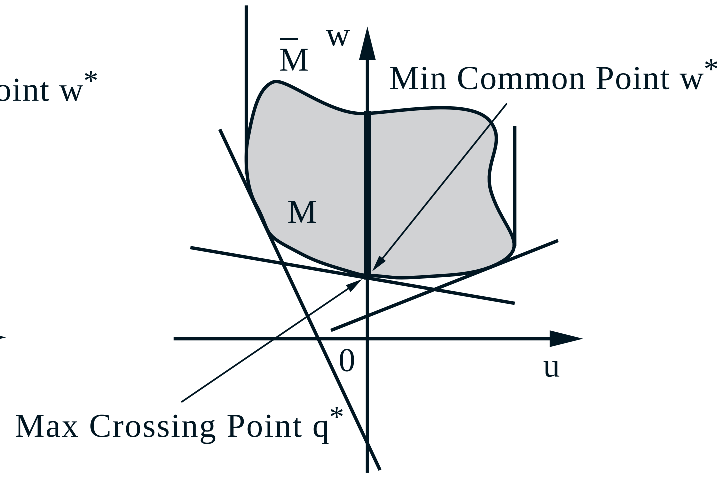
- This opens a wealth of analytical and computational possibilities, and insightful interpretations.

- Note that the equality of "sup inf" and "inf sup" is a key issue in minimax theory and game theory.

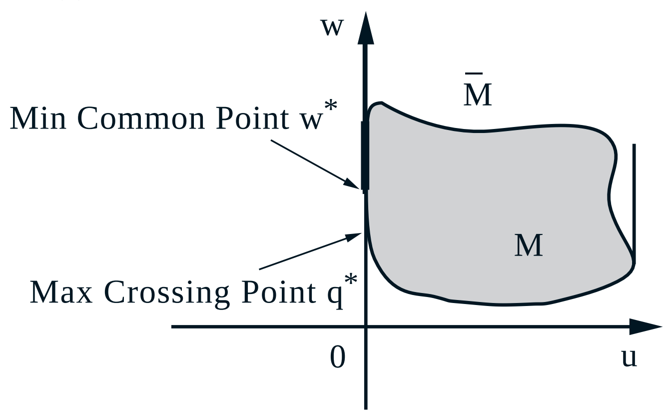
MIN COMMON/MAX CROSSING DUALITY



(a)



(b)



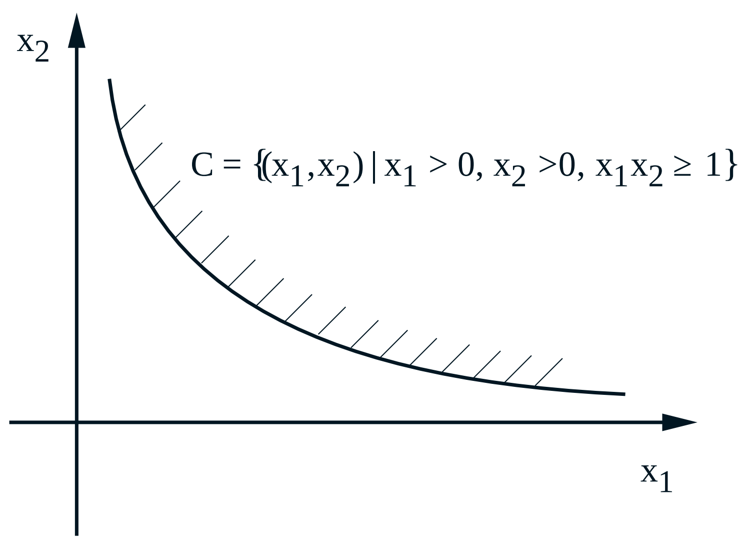
(c)

- 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).

EXCEPTIONAL BEHAVIOR

- If convex structure is so favorable, what is the source of exceptional/pathological behavior [like in (c) of the preceding slide]?
- **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).



- 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).

COURSE OUTLINE

1) **Basic Concepts (4):** Convex hulls. Closure, relative interior, and continuity.
2) **Convexity and Optimization (3):** Directions of recession and existence of optimal solutions.
3) **Hyperplanes, Duality, and Minimax (3):** Hyperplanes. Min common/max crossing duality. Saddle points and minimax theory.
4) **Polyhedral Convexity (4):** Polyhedral sets. Extreme points. Polyhedral aspects of optimization. Polyhedral aspects of duality. Linear programming. Introduction to convex programming.
5) **Conjugate Convex Functions (2):** Support functions. Conjugate operations.
6) **Subgradients and Algorithms (4):** Subgradients. Optimality conditions. Classical subgradient and cutting plane methods. Proximal algorithms. Bundle methods.
7) **Lagrangian Duality (2):** Constrained optimization duality. Separable problems. Conditions for existence of dual solution. Conditions for no duality gap.
8) **Conjugate Duality (3):** Fenchel duality theorem. Conic and semidefinite programming. Monotropic programming. Exact penalty functions.

WHAT TO EXPECT FROM THIS COURSE

- Requirements: Homework and a term paper
- We aim:
  - To develop insight and deep understanding of a fundamental optimization topic
  - To treat rigorously an important branch of applied math, and to provide some appreciation of the research in the field
- 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 (http://www.stanford.edu/ boyd/cvxbook.html)
  - 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 ideas, not to express them precisely
- The omitted proofs and a much fuller discussion can be found in the "Convex Optimization" textbook

LECTURE 2

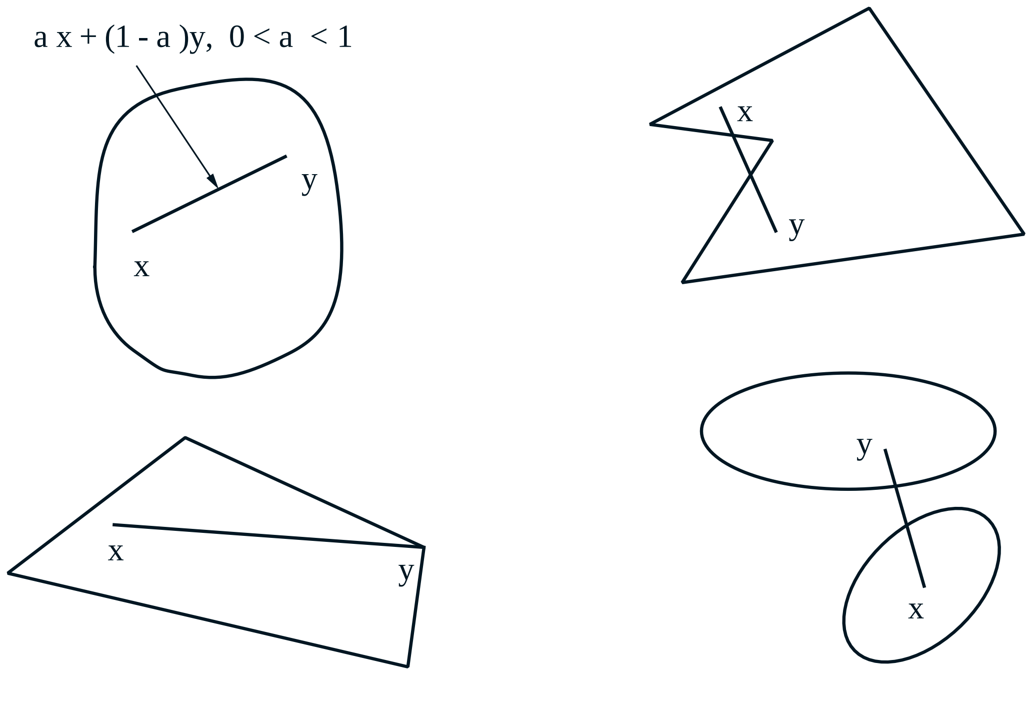
LECTURE OUTLINE

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

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

CONVEX SETS



Convex Sets   Nonconvex Sets

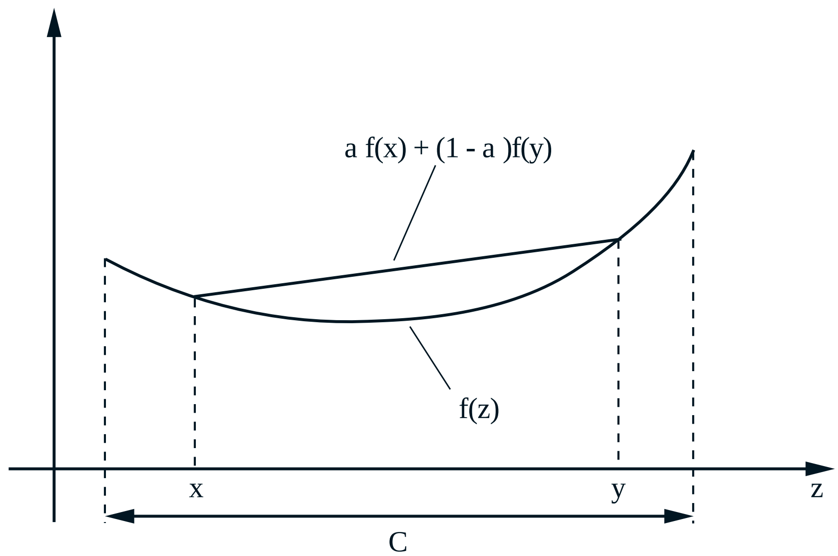
- A subset C of $\Re^n$ is called *convex* if

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

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

- 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

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

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

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

- 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 *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 extended real-valued function $f : X \mapsto [-\infty, \infty]$ is called *lower semicontinuous* at a vector $x \in X$ if $f(x) \leq \liminf_{k\to\infty} f(x_k)$ for every sequence $\{x_k\} \subset X$ with $x_k \to x$.

- We say that f is *closed* if $\text{epi}(f)$ is a closed set.

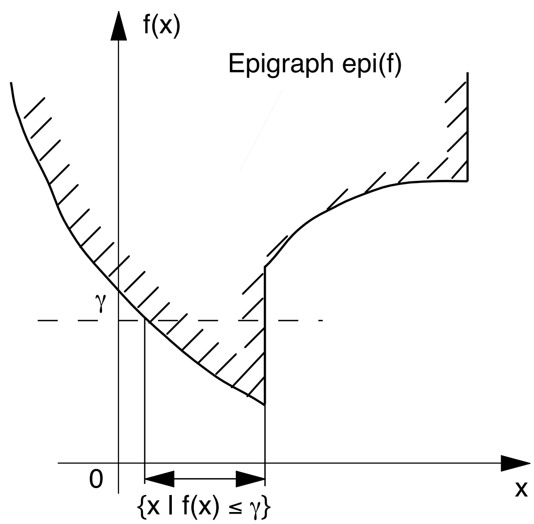
CLOSEDNESS AND SEMICONTINUITY

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

  (i) $\{x \mid f(x) \leq a\}$ is closed for every scalar a.

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

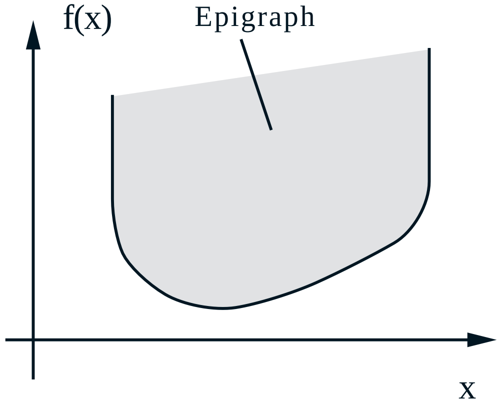
  (iii) f is closed.



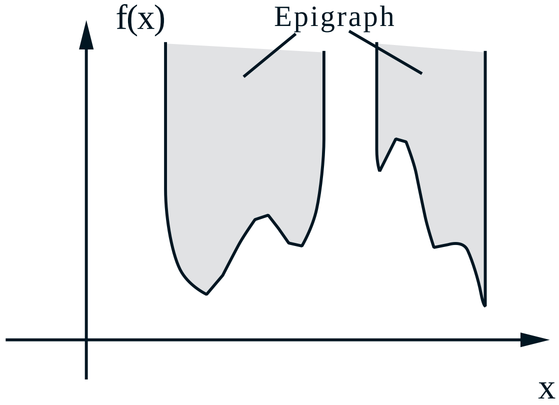
- Note that:
  - If f is lower semicontinuous at all $x \in \text{dom}(f)$, it is not necessarily closed
  - If f is closed, $\text{dom}(f)$ is not necessarily closed

- *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.

EXTENDED REAL-VALUED CONVEX FUNCTIONS



Convex function



Nonconvex function

- Let C be a convex subset of $\Re^n$. An extended real-valued function $f : C \mapsto [-\infty, \infty]$ is called *convex* if epi(f) is a convex subset of $\Re^{n+1}$.
- If f is proper, this definition is equivalent to

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

- 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*: Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i \in I$, be given functions (I is an arbitrary index set).

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

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

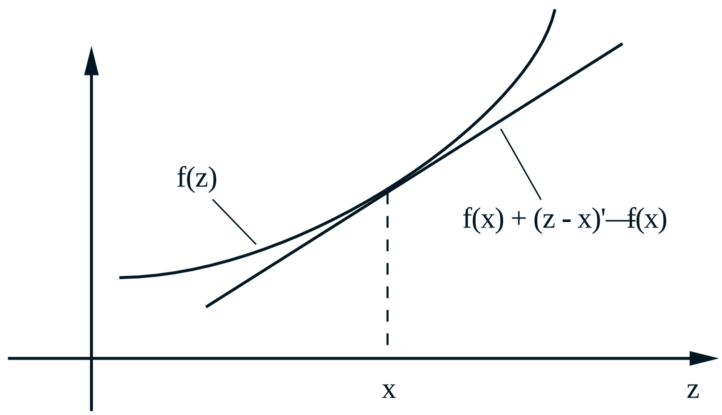
is convex (or closed) if the f_i are convex (respectively, closed).

LECTURE 3

LECTURE OUTLINE

- Differentiable Convex Functions
- Convex and Affine Hulls
- Caratheodory's Theorem
- Closure, Relative Interior, Continuity

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

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

  [Implies necessary and sufficient condition for x^* to minimize f over C: $\nabla f(x^*)'(x - x^*) \geq 0,\ \forall\, x \in C$.]

  (b) If the inequality is strict whenever $x \neq z$, then f is strictly convex over C, i.e., for all $\alpha \in (0, 1)$ and $x, y \in C$, with $x \neq y$

  $$f\big(\alpha x + (1 - \alpha)y\big) < \alpha f(x) + (1 - \alpha)f(y)$$

TWICE DIFFERENTIABLE CONVEX FUNCTIONS

- 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

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

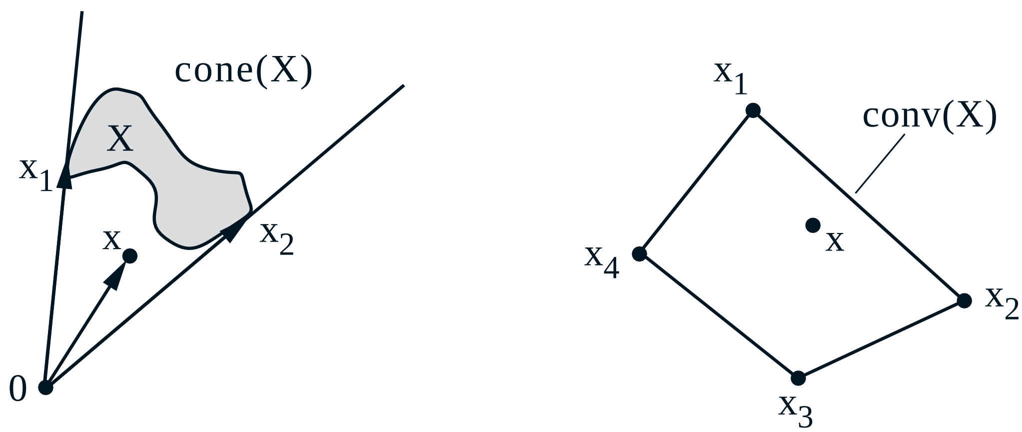
From the preceding result, f is convex.

(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.

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 $\mathrm{conv}(X)$, is the intersection of all convex sets containing X (also the set of all convex combinations from X).
- The *affine hull* of X, denoted $\mathrm{aff}(X)$, 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). Note that $\mathrm{aff}(X)$ is itself an affine set.
- 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 $\mathrm{cone}(X)$, 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, $\mathrm{cone}(X)$ is closed (nontrivial to show!)

CARATHEODORY'S THEOREM



(a) (b)

- 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.

  (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 such that $x_2 - x_1, \ldots, x_m - x_1$ are linearly independent.

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) Apply part (a) to the subset of $\Re^{n+1}$

$$Y = \{(x, 1) \mid x \in X\}$$

AN APPLICATION OF CARATHEODORY

- The convex hull of a compact set is compact.

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

By Caratheodory, a sequence in $\mathrm{conv}(X)$ 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

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

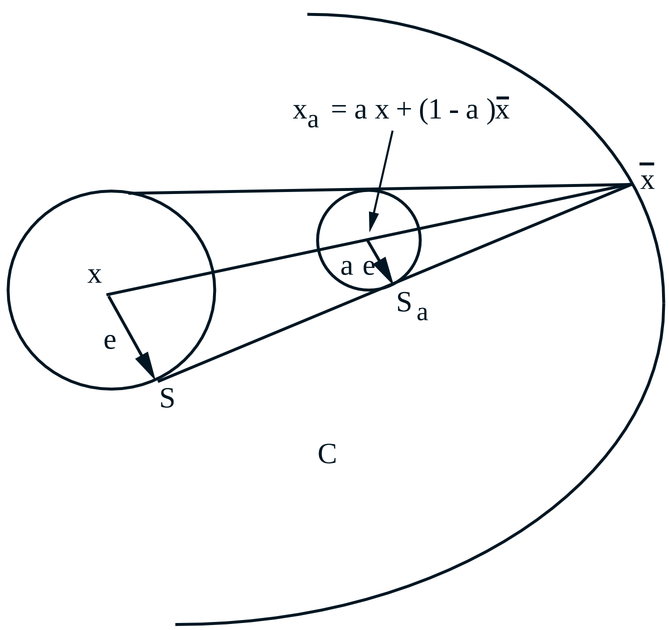
which must satisfy $\sum_{i=1}^{n+1} \alpha_i = 1$, and $\alpha_i \geq 0$, $x_i \in X$ for all i. Thus, the vector $\sum_{i=1}^{n+1} \alpha_i x_i$, which belongs to $\mathrm{conv}(X)$, is a limit point of the sequence $\left\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\right\}$, showing that $\mathrm{conv}(X)$ is compact. **Q.E.D.**

RELATIVE INTERIOR

- x is a *relative interior point* of C, if x is an interior point of C relative to $\text{aff}(C)$.

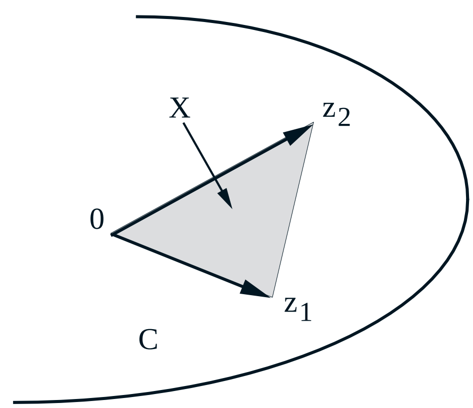
- $\text{ri}(C)$ denotes the *relative interior of* C, i.e., the set of all relative interior points of C.

- **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 $\text{ri}(C)$.



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) $x \in \mathrm{ri}(C)$ if and only if every line segment in C having x as one endpoint can be prolonged beyond x without leaving C.



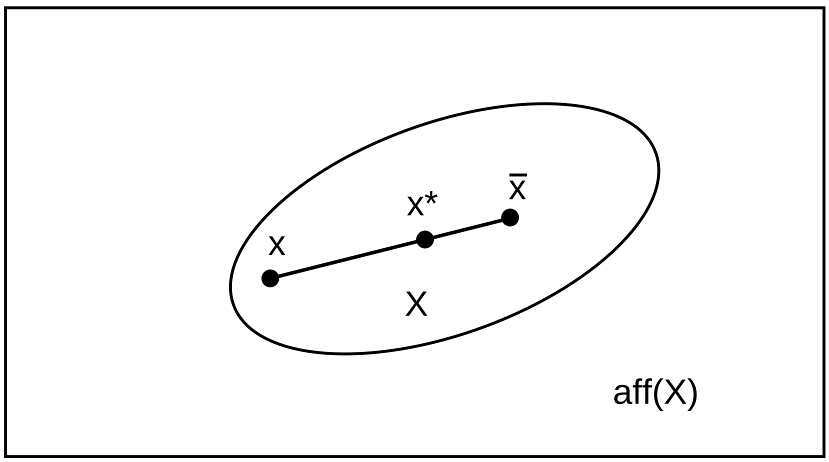
Proof: (a) Assume that $0 \in C$. We choose m linearly independent vectors $z_1, \ldots, z_m \in C$, where m is the dimension of $\mathrm{aff}(C)$, and we let

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

(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.**

LECTURE 4

LECTURE OUTLINE

- Review of relative interior
- Algebra of relative interiors and closures
- Continuity of convex functions
- Existence of optimal solutions - Weierstrass' theorem
- Projection Theorem

RELATIVE INTERIOR: REVIEW

- Recall: x is a *relative interior point* of C, if x is an interior point of C relative to $\text{aff}(C)$

- Three important properties of $\text{ri}(C)$ of a convex set C:
  - $\text{ri}(C)$ is nonempty
  - *Line Segment Principle*: If $x \in \text{ri}(C)$ 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 $\text{ri}(C)$
  - *Prolongation Lemma*: If $x \in \text{ri}(C)$ and $\overline{x} \in C$, the line segment connecting $\overline{x}$ and x can be prolonged beyond x without leaving C

CALCULUS OF RELATIVE INTERIORS: SUMMARY

- 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

- Let C be a nonempty convex set. Then $\mathrm{ri}(C)$ and $\mathrm{cl}(C)$ are "not too different for each other."

- *Proposition*:

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

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

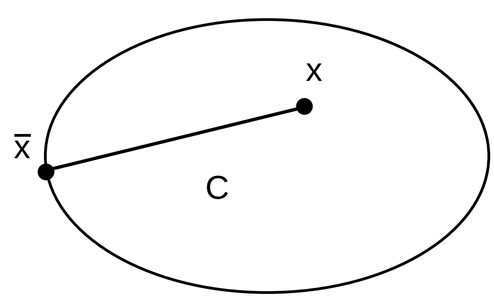
  (c) Let $\overline{C}$ be another nonempty convex set. Then the following three conditions are equivalent:

    (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 $\alpha x + (1-\alpha)\overline{x} \in \mathrm{ri}(C)$ for all $\alpha \in (0,1]$. Thus, $\overline{x}$ is the limit of a sequence that lies in $\mathrm{ri}(C)$, so $\overline{x} \in \mathrm{cl}\big(\mathrm{ri}(C)\big)$.



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 a sequence $\{x_k\} \subset C$ 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) If $\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) \neq \emptyset$, then

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

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

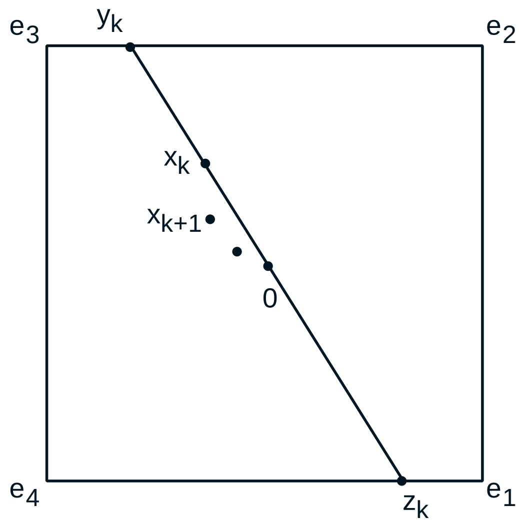
Proof of (a): $C_1 + C_2$ is the result of the linear transformation $(x_1, x_2) \mapsto x_1 + x_2$.

- Counterexample for (b):

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

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 maximum 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

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

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

Since $\|x_k\|_\infty \to 0$, $f(x_k) \to f(0)$. **Q.E.D.**

- Extension to continuity over $\mathrm{ri}(\mathrm{dom}(f))$.

PARTIAL MINIMIZATION

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

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

- **1st fact:** If F is convex, then f is also convex.
- **2nd fact:**

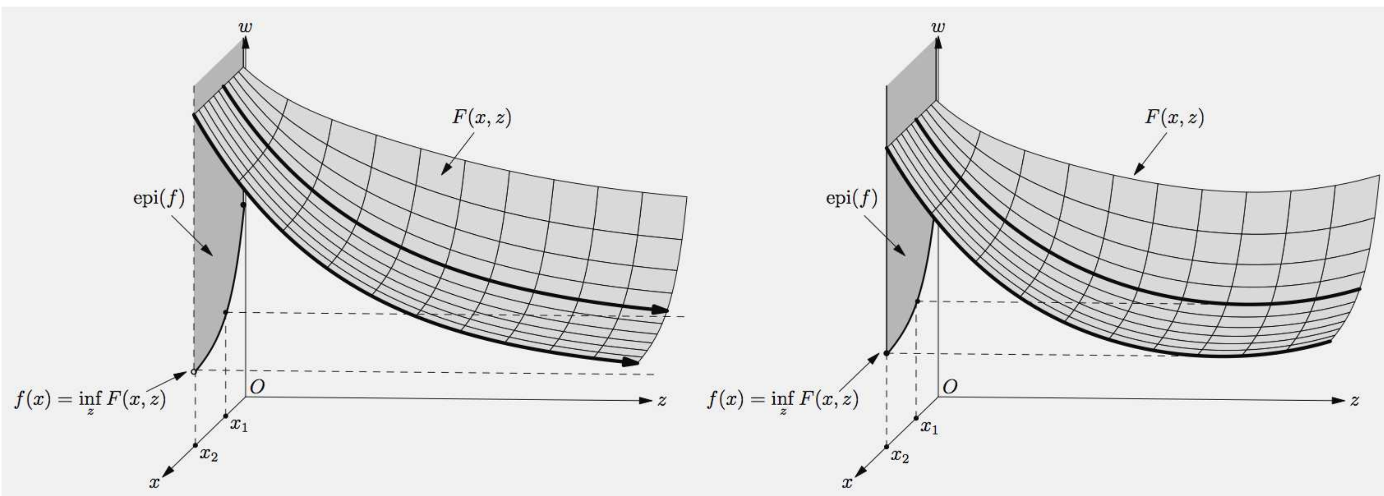
$$P\big(\text{epi}(F)\big) \subset \text{epi}(f) \subset \text{cl}\Big(P\big(\text{epi}(F)\big)\Big),$$

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.



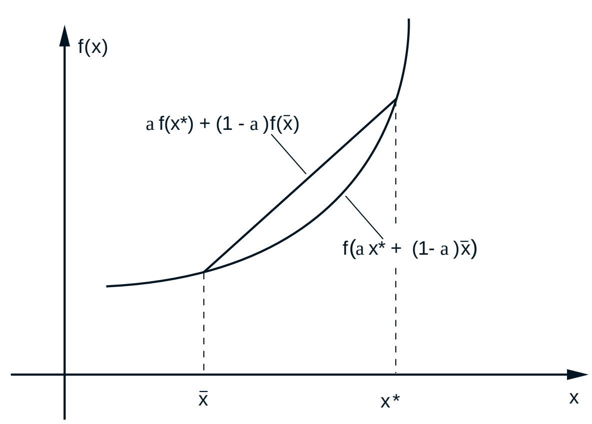
LOCAL AND GLOBAL MINIMA

- Consider minimizing $f : \Re^n \mapsto (-\infty, \infty]$ over a set $X \subset \Re^n$
- x is **feasible** if $x \in X \cap \mathrm{dom}(f)$
- 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

- **Note:** The intersection of a nested sequence of nonempty compact sets is compact

- **Conclusion:** The set of minima of f is nonempty and compact if the level sets of f are compact

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 $\big\{x \in X \mid f(x) \le \gamma\big\}$ 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.**

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 $\|z - x\|$ over all $x \in C$ (called the *projection of* z *on* C).

  (b) x^* is the projection of z if and only if

  $$(x - x^*)'(z - x^*) \leq 0, \qquad \forall\, x \in C$$

  (c) The projection operation is nonexpansive, i.e.,

  $$\|x_1^* - x_2^*\| \leq \|z_1 - z_2\|, \qquad \forall\, z_1, z_2 \in \Re^n,$$

  where x_1^* and x_2^* are the projections on C of z_1 and z_2, respectively.

LECTURE 5

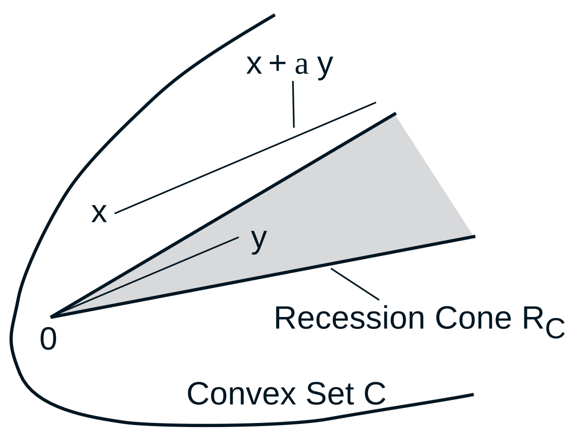
LECTURE OUTLINE

- Recession cones
- Directions of recession of convex functions
- Applications to existence of optimal solutions

RECESSION CONE OF A CONVEX SET

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

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



- *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 y belongs to R_C if and only if there exists a vector $x \in C$ such that $x + \alpha y \in C$ 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.

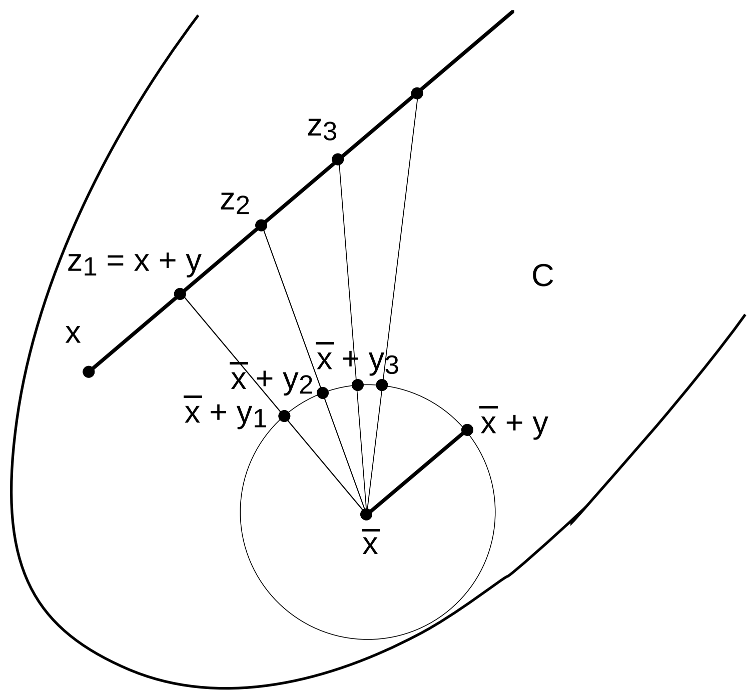
  (e) If D is another closed convex set such that $C \cap D \neq \emptyset$, we have

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

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

  Let $z_k = x + ky$ for $k = 1, 2, \ldots$, and $y_k = (z_k - \overline{x})\|y\|/\|z_k - \overline{x}\|$. We have

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

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

LINEALITY SPACE

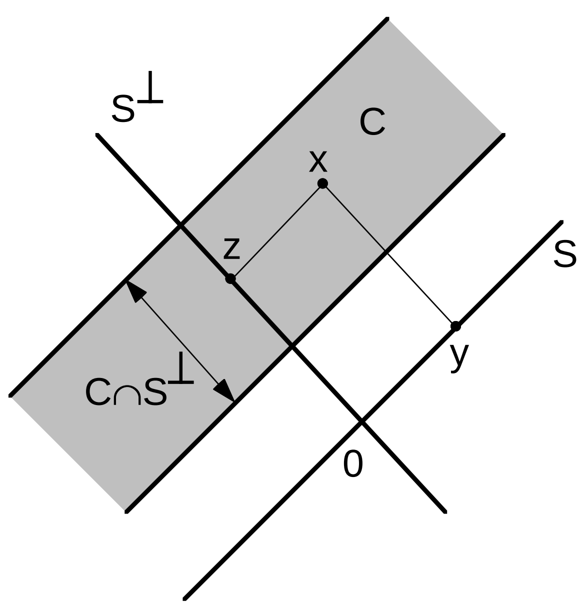
- The *lineality space* of a convex set C, denoted by L_C, is the subspace of vectors y such that $y \in R_C$ and $-y \in R_C$:

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

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

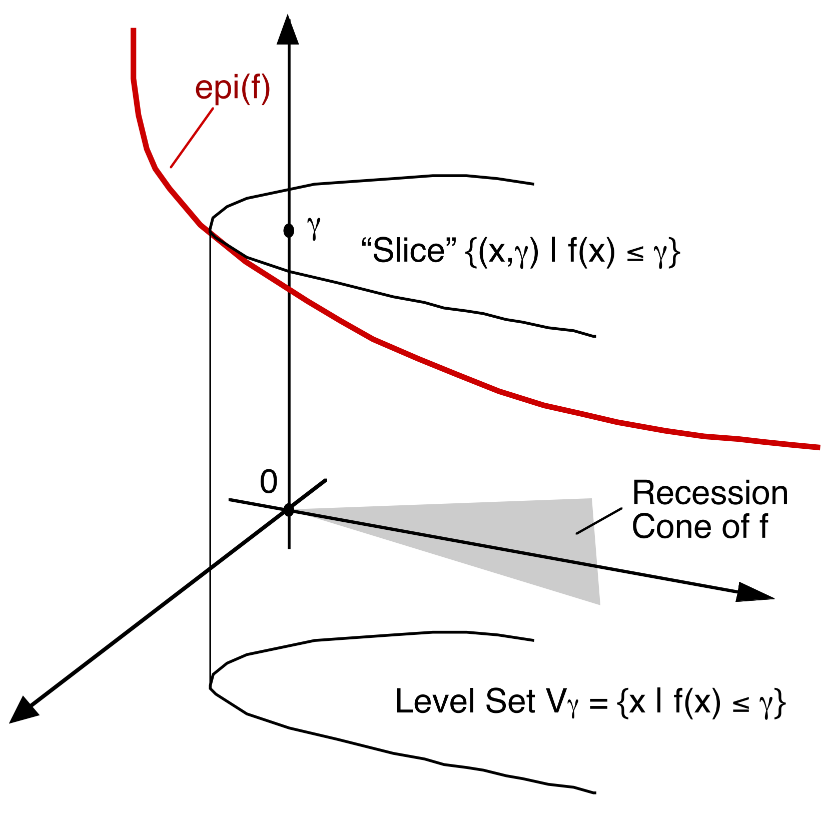
$$C = L_C + (C \cap L_C^{\perp}).$$

Also, if $L_C = R_C$, the component $C \cap L_C^{\perp}$ is compact (this will be shown later).



DIRECTIONS OF RECESSION OF A FUNCTION

- 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, given by

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

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

Proof: For all γ for which V_γ is nonempty,

$$\big\{(x, \gamma) \mid x \in V_\gamma\big\} = \mathrm{epi}(f) \cap \big\{(x, \gamma) \mid x \in \Re^n\big\}$$

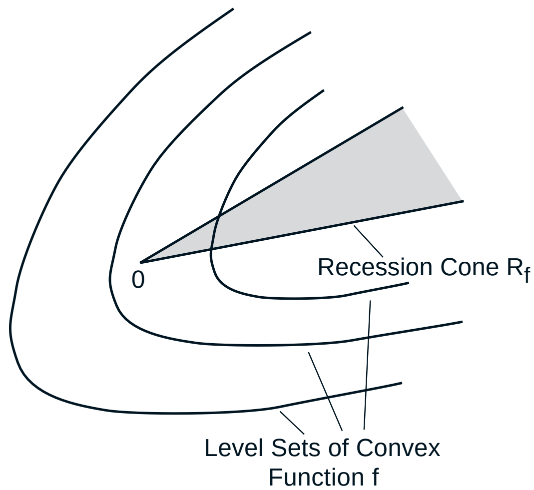
The recession cone of the set on the left is $\big\{(y, 0) \mid y \in R_{V_\gamma}\big\}$. The recession cone of the set on the right is the intersection of $R_{\mathrm{epi}}(f)$ and the recession cone of $\big\{(x, \gamma) \mid x \in \Re^n\big\}$. Thus we have

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

from which the result follows.

RECESSION CONE OF A CONVEX FUNCTION

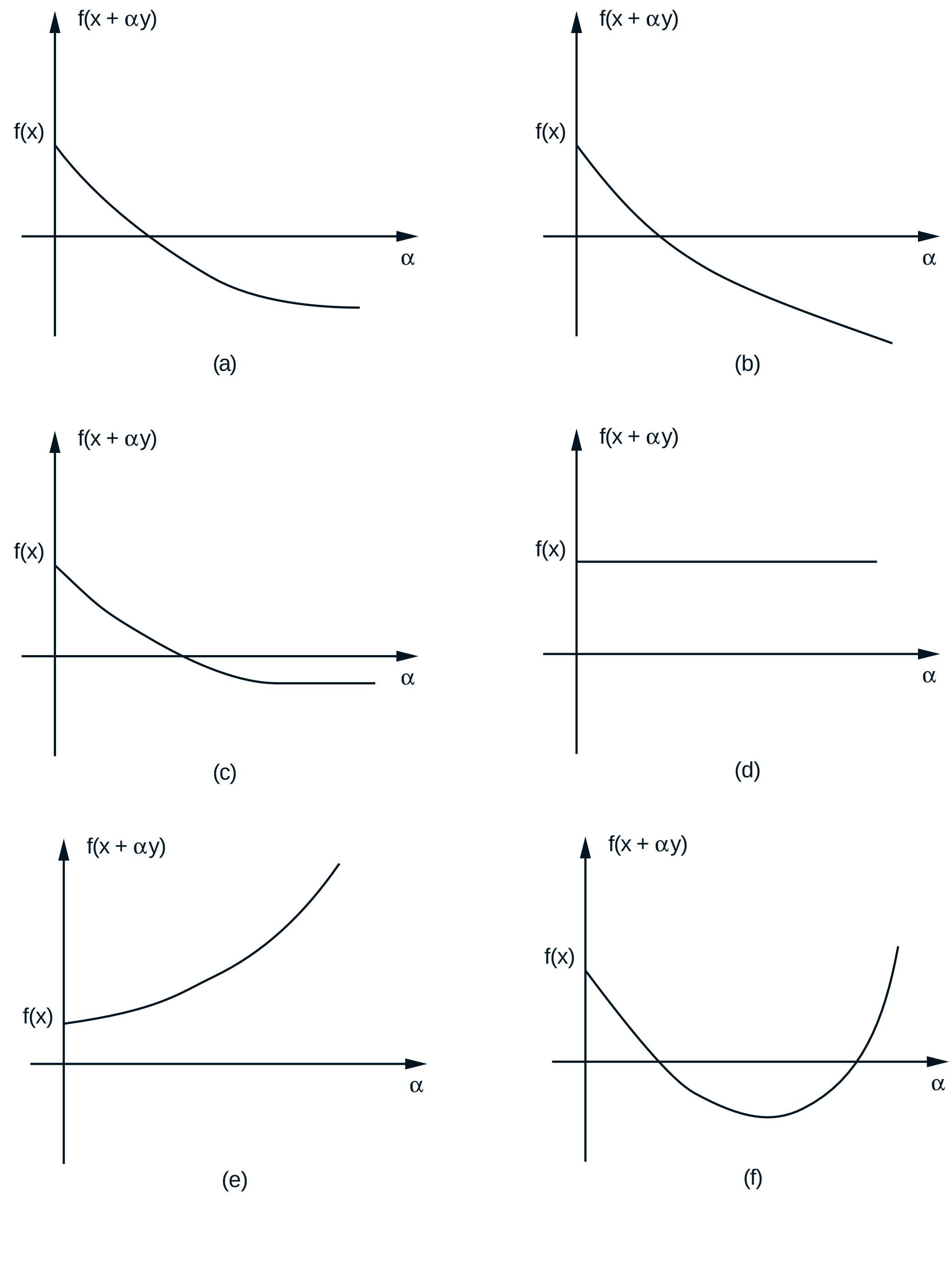
- For a closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$, the (common) recession cone of the nonempty level sets $V_\gamma = \big\{x \mid f(x) \le \gamma\big\}$, $\gamma \in \Re$, is the *recession cone of* f, and is denoted by R_f.



- Terminology:
  - $y \in R_f$: a *direction of recession of* f.
  - $L_f = R_f \cap (-R_f)$: the *lineality space of* f.
  - $y \in L_f$: a *direction of constancy* of f.
  - Function $r_f : \Re^n \mapsto (-\infty, \infty]$ whose epigraph is $R_{\text{epi}(f)}$: the *recession function* of f.

- Note: $r_f(y)$ is the "asymptotic slope" of f in the direction y. In fact, $r_f(y) = \lim_{\alpha\to\infty} \nabla f(x+\alpha y)'y$ if f is differentiable. Also, $y \in R_f$ iff $r_f(y) \le 0$.

DESCENT BEHAVIOR OF A CONVEX FUNCTION



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

EXISTENCE OF SOLUTIONS - BOUNDED CASE

Proposition: The set of minima of a closed proper convex function $f : \Re^n \mapsto (-\infty, \infty]$ is nonempty and compact if and only if f has no nonzero direction of recession.

Proof: Let X^* be the set of minima, let $f^* = \inf_{x \in \Re^n} f(x)$, and let $\{\gamma_k\}$ be a scalar sequence such that $\gamma_k \downarrow f^*$. Note that

$$X^* = \cap_{k=0}^{\infty} \big\{x \mid f(x) \leq \gamma_k\big\}$$

If f has no nonzero direction of recession, the sets $\big\{x \mid f(x) \leq \gamma_k\big\}$ are nonempty, compact, and nested, so X^* is nonempty and compact.

Conversely, we have

$$X^* = \big\{x \mid f(x) \leq f^*\big\},$$

so if X^* is nonempty and compact, all the level sets of f are compact and f has no nonzero direction of recession. **Q.E.D.**

SPECIALIZATION/GENERALIZATION

- **Important special case:** Minimize a real-valued function $f : \Re^n \mapsto \Re$ over a nonempty set X. Apply the preceding proposition to the extended real-valued function

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

- Optimal solution set is nonempty and compact iff X and f have no common nonzero direction of recession
- Set intersection issues are fundamental and play an important role in several seemingly unrelated optimization contexts
- Directions of recession play an important role in set intersection theory (see the next lecture)
- This theory generalizes to nonconvex sets (we will not cover this)

LECTURE 6

LECTURE OUTLINE

- Nonemptiness of closed set intersections
- Existence of optimal solutions
- Linear and Quadratic Programming
- Preservation of closure under linear transformation
- Preservation of closure under partial minimization

THE ROLE OF CLOSED SET INTERSECTIONS

- **A fundamental question:** Given a sequence of nonempty closed sets $\{C_k\}$ in $\Re^n$ with $C_{k+1} \subset S_k$ 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:

1. Does a function $f : \Re^n \mapsto (-\infty, \infty]$ attain a minimum over a set X? This is true iff the intersection of the nonempty level sets $\big\{x \in X \mid f(x) \leq \gamma_k\big\}$ is nonempty.

2. If C is closed and A is a matrix, is $A\,C$ closed? Special case:

– If C_1 and C_2 are closed, is $C_1 + C_2$ closed?

3. If $F(x, z)$ is closed, is $f(x) = \inf_z F(x, z)$ closed? (Critical question in duality theory.) Can be addressed by using the relation

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

where $P(\cdot)$ is projection on the space of (x, w).

ASYMPTOTIC DIRECTIONS

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

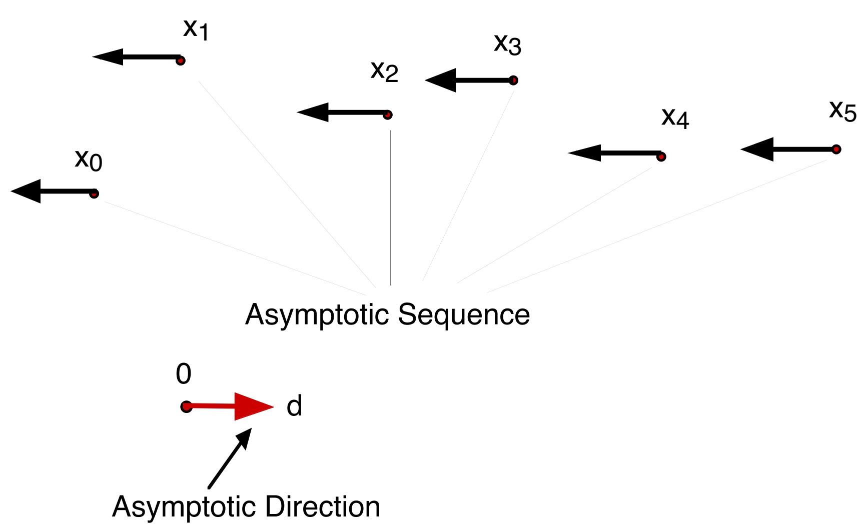
$$x_k \in C_k, \qquad x_k \neq 0, \qquad k = 0, 1, \ldots$$

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

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

- $\{x_k\}$ is called *retractive* if for some $\overline{k}$, we have

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



RETRACTIVE SEQUENCES

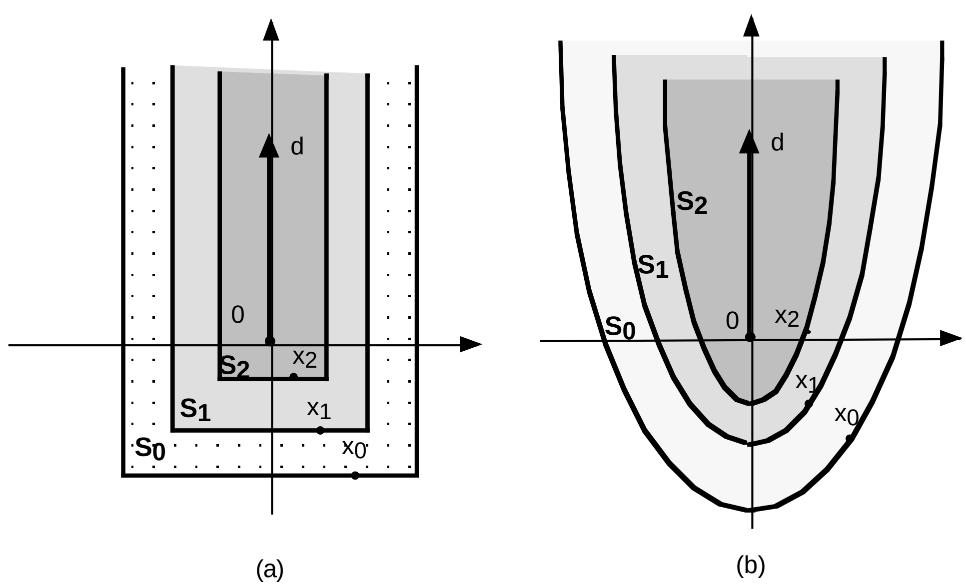
- A nested sequence $\{C_k\}$ of closed convex sets is *retractive* if all its asymptotic sequences are retractive.

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

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

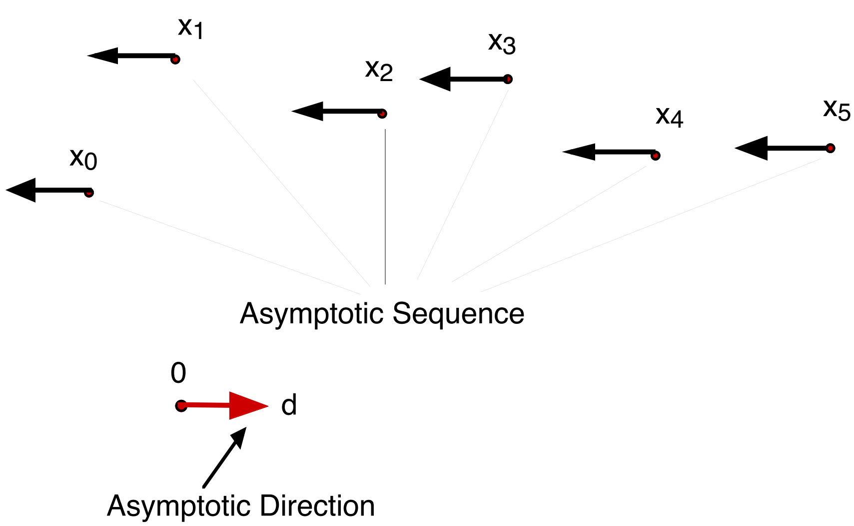


SET INTERSECTION THEOREM I

- 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

- Let $\{C_k\}$ be a nested sequence of nonempty closed convex sets, and X be a retractive set such that all the sets $S_k = X \cap C_k$ 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 $\{S_k\}$ is retractive and $\cap_{k=0}^{\infty} S_k$ is nonempty.

- Special case: $X = \Re^n$, $R = L$.

Proof: The set of common directions of recession of S_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 $\{S_k\}$ is retractive.

EXISTENCE OF OPTIMAL SOLUTIONS

- Let X and $f : \Re^n \mapsto (-\infty, \infty]$ be closed convex and such that $X \cap \mathrm{dom}(f) \neq \emptyset$. The set of minima of f over X is nonempty under any one of the following two conditions:

  (1) $R_X \cap R_f = L_X \cap L_f$.

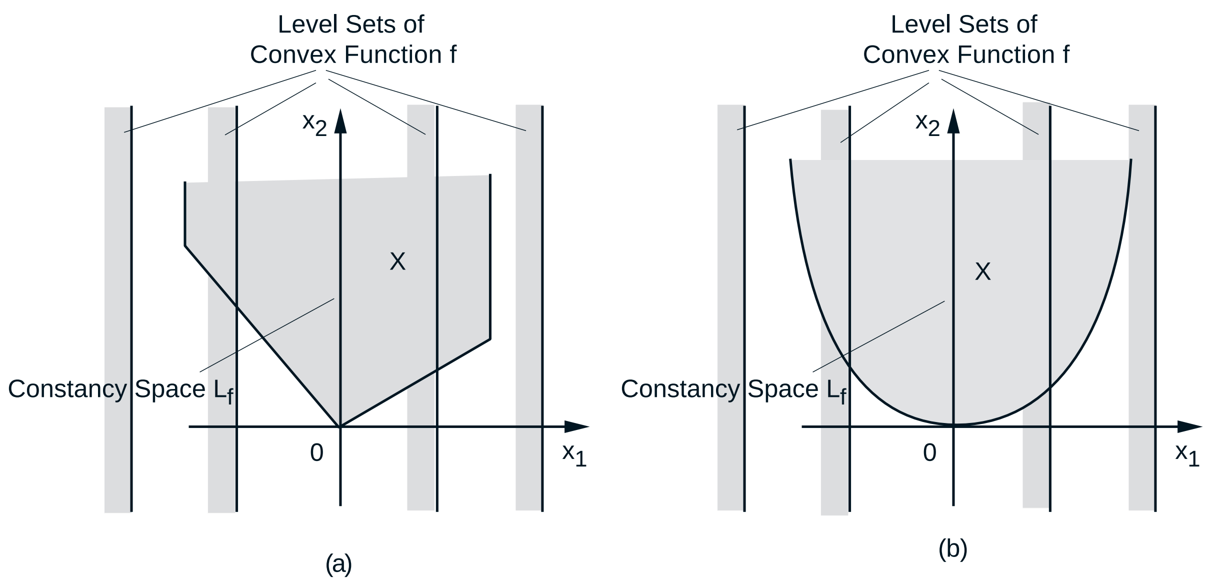
  (2) $R_X \cap R_f \subset L_f$, and X is polyhedral.

Proof: Follows by writing

Set of Minima $= X \cap$ (Nonempty Level Sets of f)

and by applying the preceding set intersection theorem. **Q.E.D.**

EXISTENCE OF OPTIMAL SOLUTIONS: EXAMPLE



- Here $f(x_1, x_2) = e^{x_1}$.
- In (a), X is polyhedral, and the minimum is attained.
- In (b),

$$X = \{(x_1, x_2) \mid x_1^2 \le x_2\}$$

We have $R_X \cap R_f \subset L_f$, but the minimum is not attained (X is not polyhedral).

LINEAR AND QUADRATIC PROGRAMMING

- **Theorem:** Let

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

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) Follows by writing

$$\text{Set of Minima} = X \cap (\text{Nonempty Level Sets of } f)$$

and by verifying 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

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

and

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

Q.E.D.

CLOSURE UNDER LINEAR TRANSFORMATIONS

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

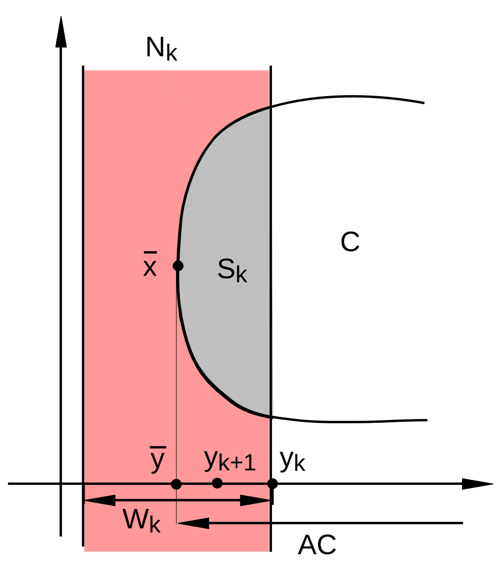
  (a) AC is closed if $R_C \cap N(A) \subset L_C$.

  (b) $A(X \cap C)$ is closed if X is a polyhedral 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 $\cap_{k=0}^{\infty} S_k \neq \emptyset$, where $S_k = C \cap N_k$, 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.

CONVEX "QUADRATIC" SET INTERSECTIONS

- Consider $\{C_k\}$ given by

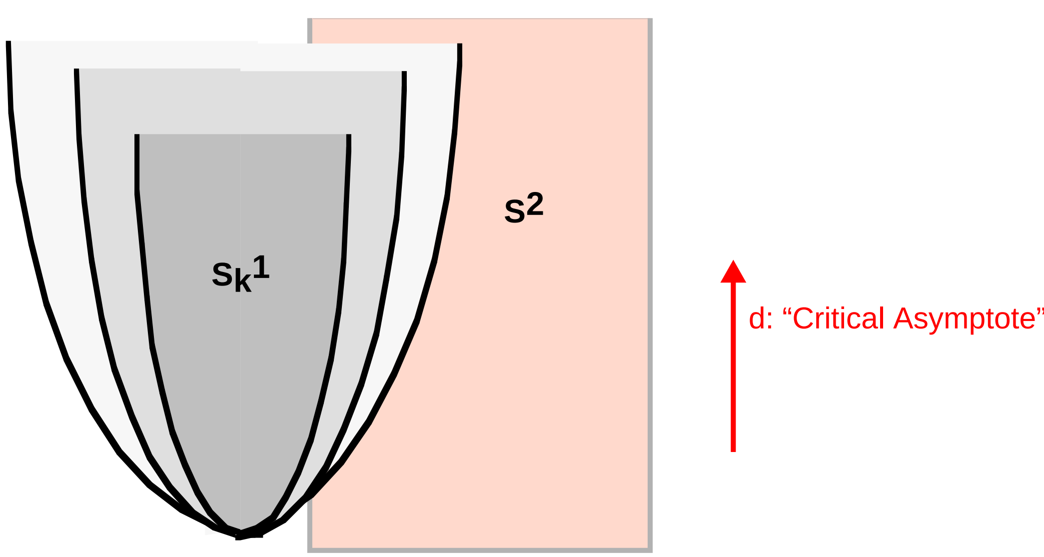
$$C_k = \big\{x \mid x'Qx + a'x + b \leq w_k\big\},$$

where $w_k \downarrow 0$. Let

$$X = \big\{x \mid x'Q_jx + a_j'x + b_j \leq 0,\ j = 1, \ldots, r\big\},$$

be such that $X \cap C_k$ is nonempty for all k. Then, the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ is nonempty.

- Key idea: For the intersection $X \cap \big(\cap_{k=0}^{\infty} C_k\big)$ to be empty, there must exist a "critical asymptote".



A RESULT ON QUADRATIC MINIMIZATION

- Let
$$f(x) = x'Qx + c'x,$$
$$X = \{x \mid x'R_j x + a_j'x + b_j \leq 0,\ j = 1, \ldots, r\},$$
where Q and R_j are positive semidefinite matrices. If the minimal value of f over X is finite, there exists a minimum of f of over X.

Proof: Follows by writing

Set of Minima $= X \cap$ (Nonempty Level Sets of f)

and by applying the "quadratic" set intersection theorem. **Q.E.D.**

- Transformations of "Quadratic" Sets: If C is specified by convex quadratic inequalities, the set AC is closed.

Proof: Follows by applying the "quadratic" set intersection theorem, similar to the earlier case. **Q.E.D.**

PARTIAL MINIMIZATION THEOREM

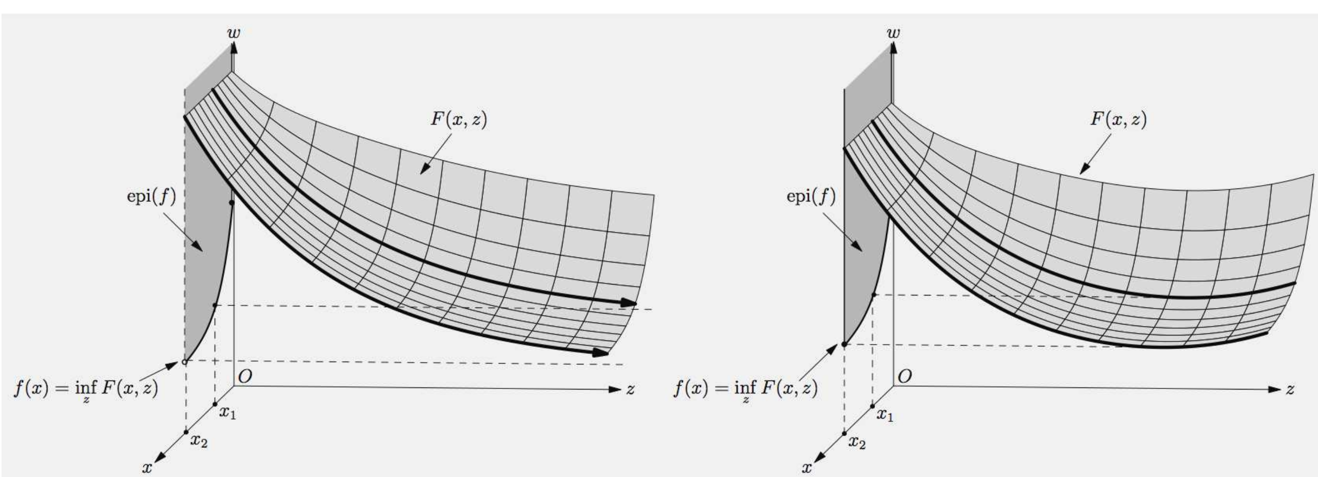
- Let $F : \Re^{n+m} \mapsto (-\infty, \infty]$ be a closed proper convex function, and consider $f(x) = \inf_{z \in \Re^m} F(x, z)$.

- Each of the major set intersection theorems 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 $x \in \mathrm{dom}(f)$, the set of minima of $F(x, \cdot)$ is nonempty and compact.

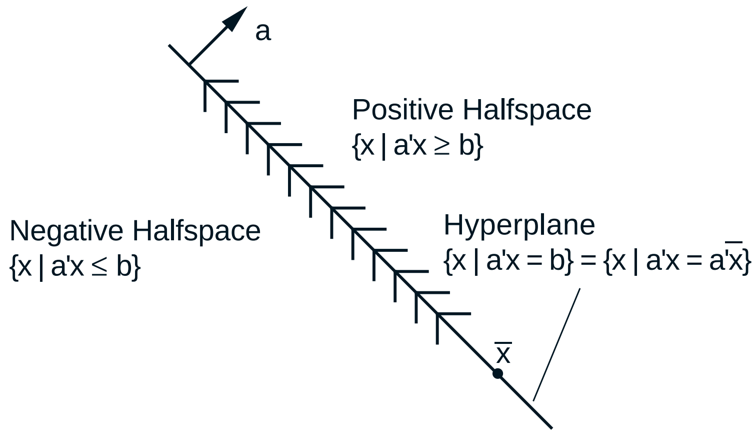


LECTURE 7

LECTURE OUTLINE

- Hyperplane separation
- Nonvertical hyperplanes
- Min common and max crossing problems

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.,

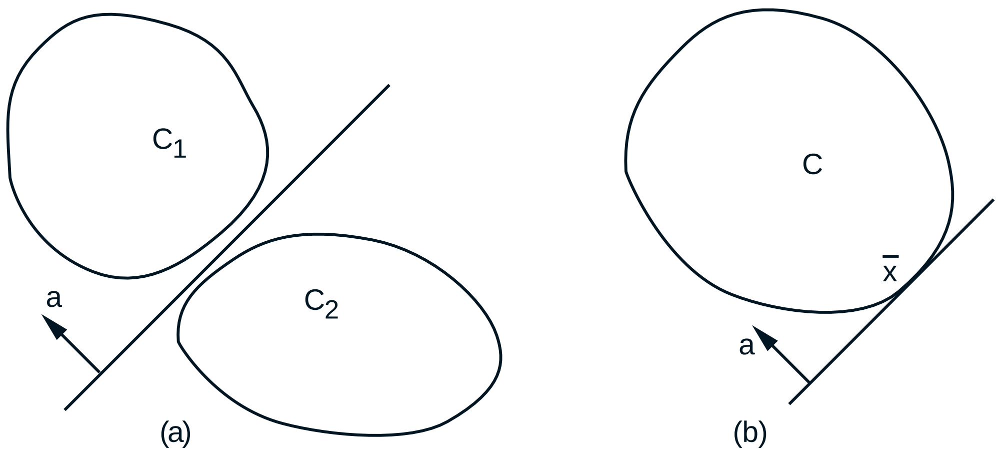
$$\text{either} \quad a'x_1 \leq b \leq a'x_2, \qquad \forall\, x_1 \in C_1,\ \forall\, x_2 \in C_2,$$

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

- 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* $\overline{x}$.

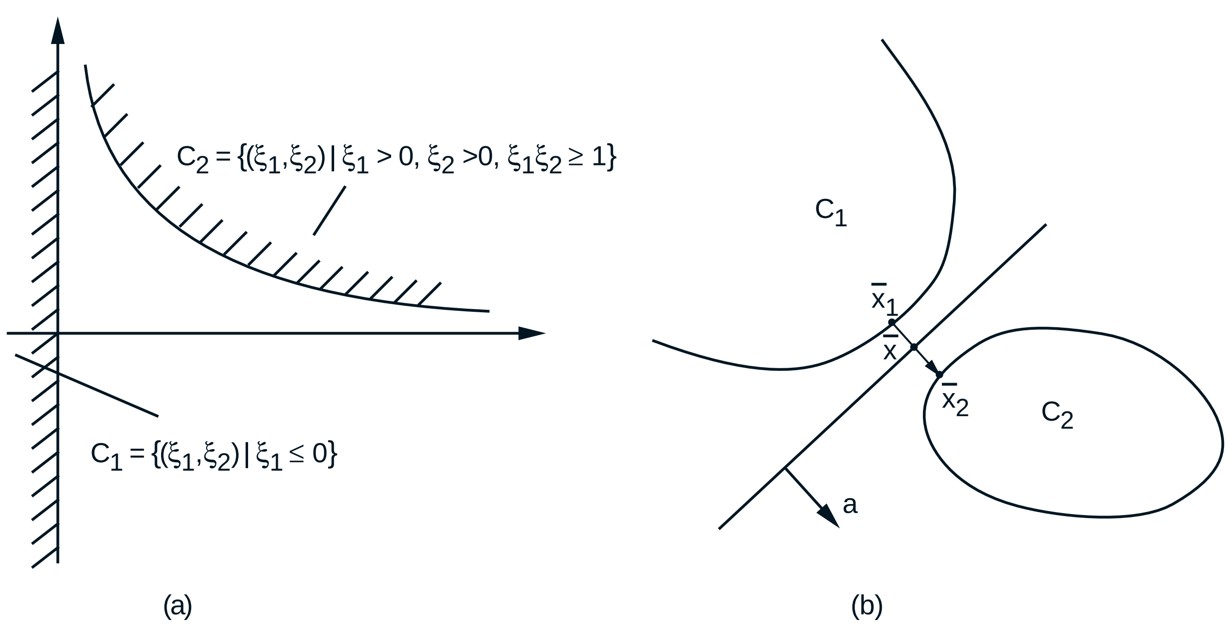
VISUALIZATION

- Separating and supporting hyperplanes:



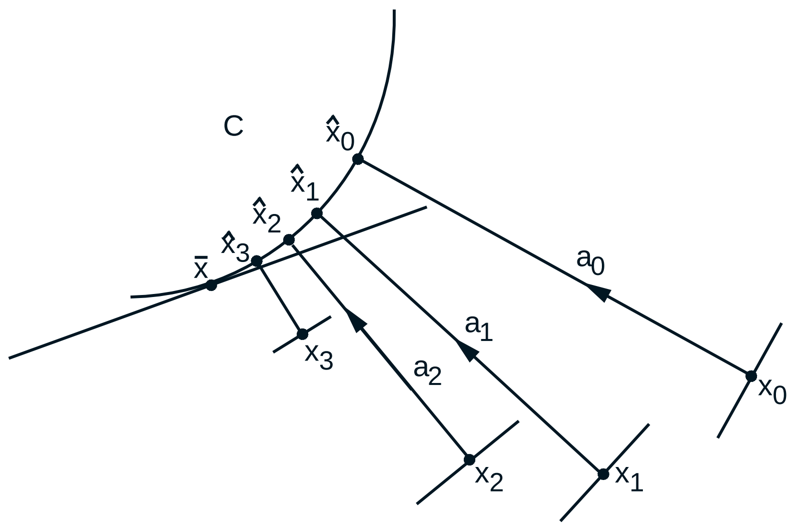
- A separating $\{x \mid a'x = b\}$ that is disjoint from C_1 and C_2 is called *strictly* separating:

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



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

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

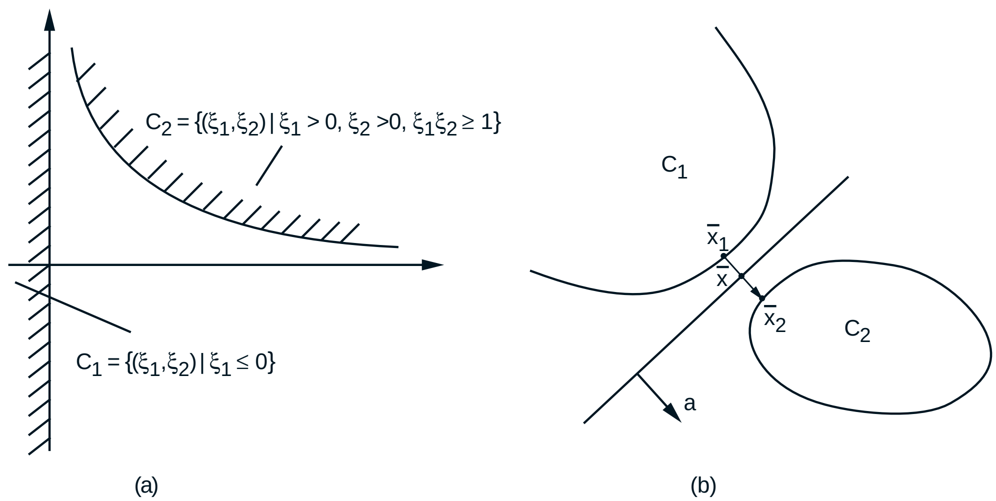
Since C_1 and C_2 are disjoint, the origin does not belong to $C_1 - C_2$, so by the Supporting Hyperplane Theorem, there exists a vector $a \neq 0$ such that

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

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.



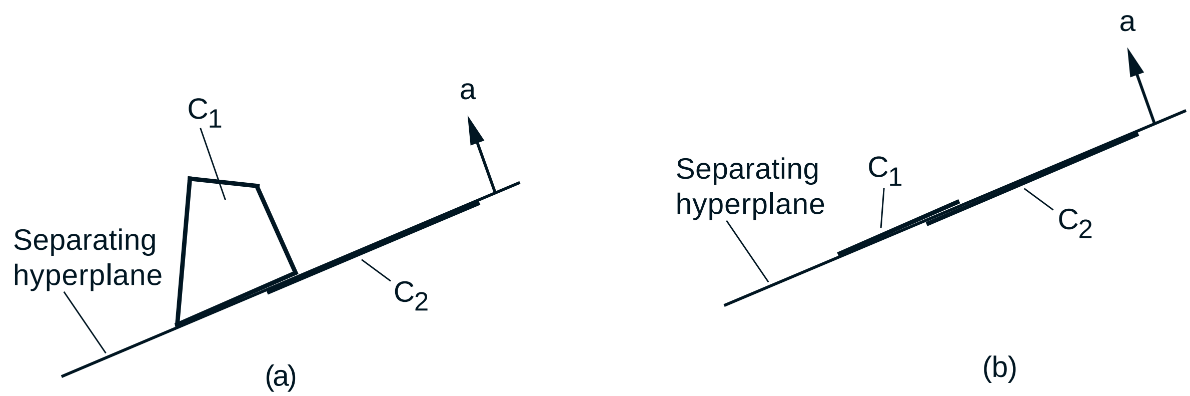
Proof: (Outline) Consider the set $C_1 - C_2$. Since C_1 is closed and C_2 is compact, $C_1 - C_2$ is closed. Since $C_1 \cap C_2 = \emptyset$, $0 \notin C_1 - C_2$. Let $\overline{x}_1 - \overline{x}_2$ be the projection of 0 onto $C_1 - C_2$. The strictly separating hyperplane is constructed as in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

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.

- 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

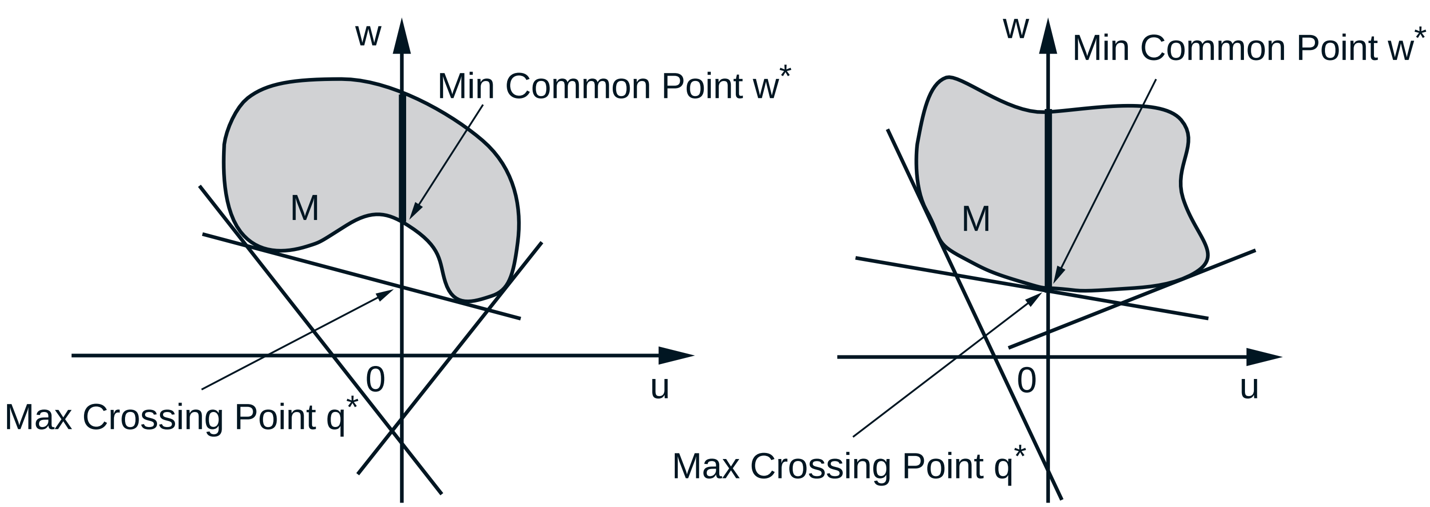
$$\mathrm{ri}(C_1) \cap \mathrm{ri}(C_2) = \emptyset$$

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.

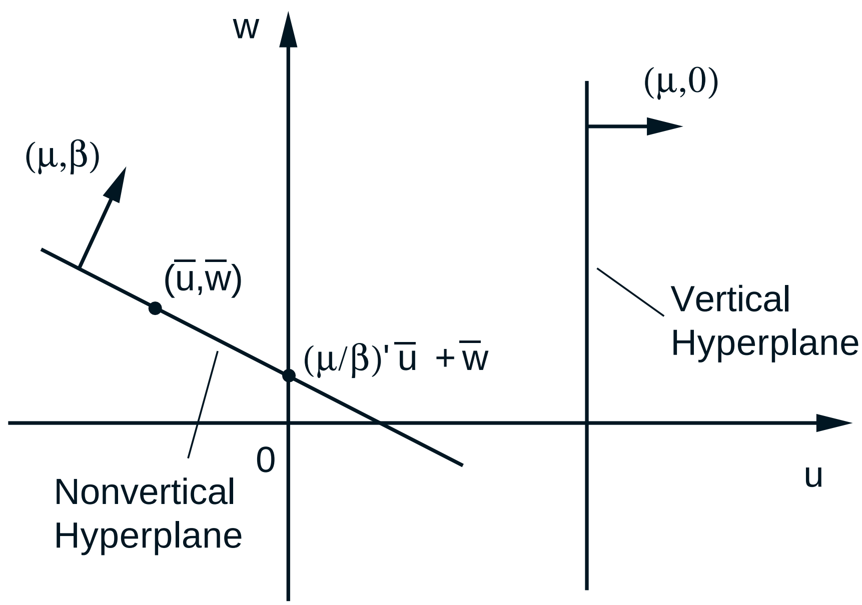
  (b) *Max Crossing Point Problem*: Consider "nonvertical" hyperplanes that contain M in their "upper" closed halfspace. Find one whose crossing point of the $(n+1)$st axis is maximum.



- We first need to study "nonvertical" hyperplanes.

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.

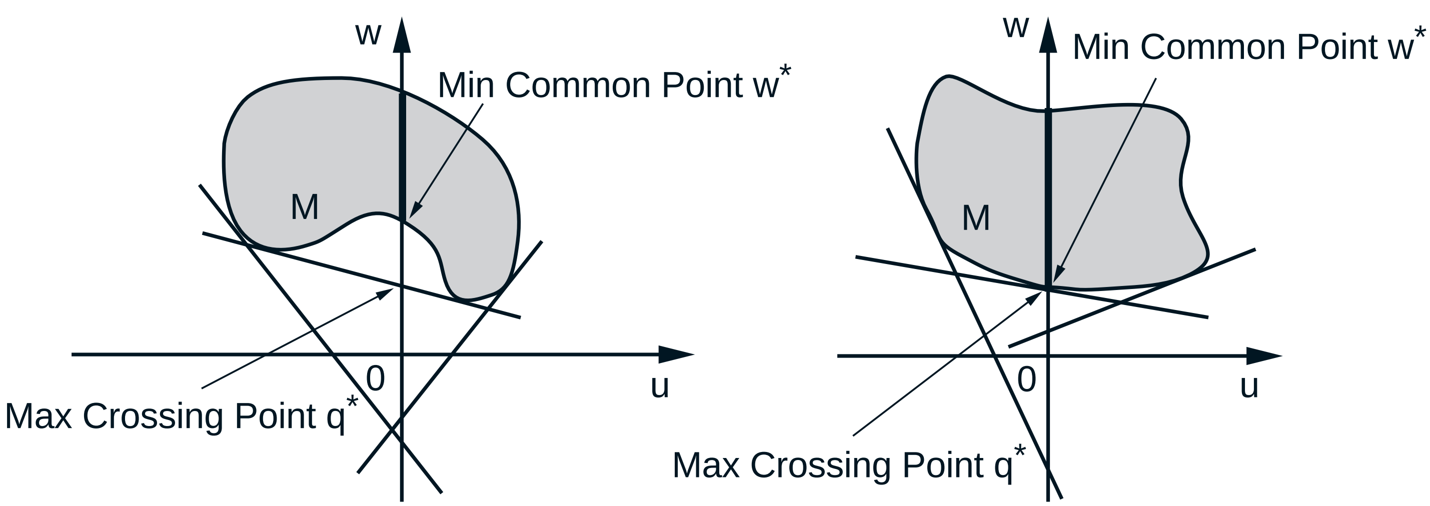
(a) C is the intersection of the closed halfspaces containing C. If all these corresponded to vertical hyperplanes, C would contain a vertical line.

(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).

LECTURE 8

LECTURE OUTLINE

- Min Common / Max Crossing problems
- Weak duality
- Strong duality
- Existence of optimal solutions
- Minimax problems



WEAK DUALITY

- Optimal value of the min common problem

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

- Math formulation of the 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

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

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

- Weak Duality: For all $(u,w) \in M$ and $\mu \in \Re^n$,

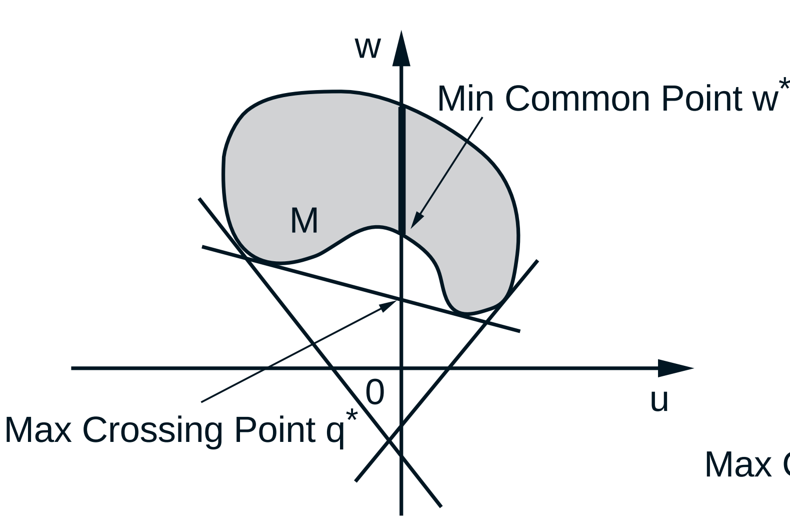
$$q(\mu) = \inf_{(u,w)\in M}\{w + \mu' u\} \le \inf_{(0,w)\in M} w = w^*,$$

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

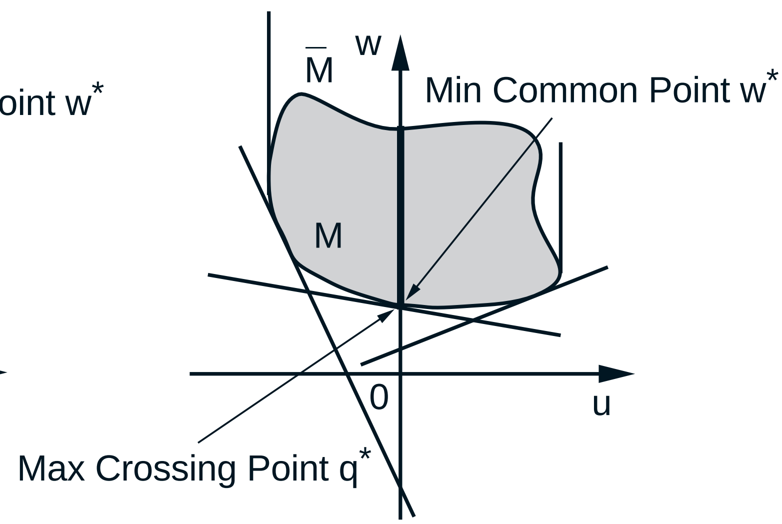
- Note that q is concave and upper-semicontinuous.

STRONG DUALITY

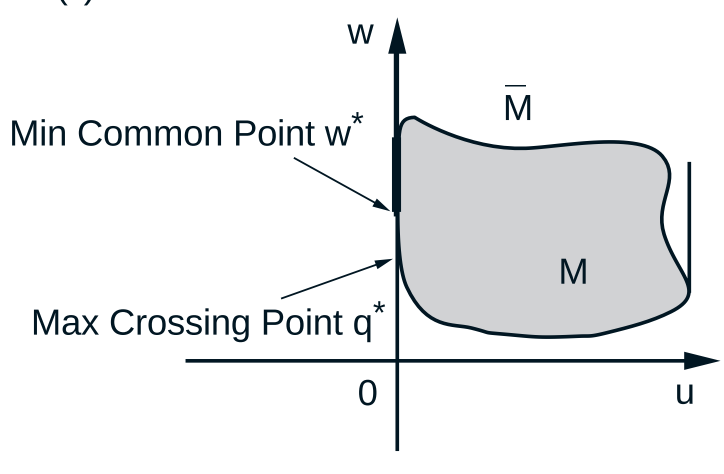
- Question: Under what conditions do we have $q^* = w^*$ and the supremum in the max crossing problem is attained?



(a)



(b)



(c)

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

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

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

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

(1) $\overline{M}$ does not contain any vertical lines.

(2) $(0, w^* - \epsilon) \notin \mathrm{cl}(\overline{M})$ for any $\epsilon > 0$.

(3) 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 $w^* - \epsilon \le \xi \le w^*$, so $w^* - \epsilon \le q^* \le w^*$. 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 exists 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., there exists (μ, β) such that

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

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

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 $q(\mu) = q^* = w^*$.

MINIMAX PROBLEMS

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

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

and

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

- Some important contexts:
  - Worst-case design. Special case: Minimize over $x \in X$

$$\max\big\{f_1(x), \ldots, f_m(x)\big\}$$

  - Duality theory and zero sum game theory (see the next two slides)

- We will study minimax problems using the min common/max crossing framework

CONSTRAINED OPTIMIZATION DUALITY

- For the problem

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

introduce the Lagrangian function

$$L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x)$$

- 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

$$\sup_{\mu \ge 0} \inf_{x \in \Re^n} L(x, \mu) = \inf_{x \in \Re^n} \sup_{\mu \ge 0} L(x, \mu)$$

ZERO SUM GAMES

- Two players: 1st chooses $i \in \{1, \ldots, n\}$, 2nd chooses $j \in \{1, \ldots, m\}$.

- If moves 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 moves.

- 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 $\min_x x'Az$

MINIMAX INEQUALITY

- We always have

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

[for every $\overline{z} \in Z$, write

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

and take the sup over $\overline{z} \in Z$ of the left-hand side].

- This is called the *minimax inequality*. When it holds as an equation, it is called the *minimax equality*.
- The minimax equality need not hold in general.
- When the minimax equality holds, it often leads to interesting interpretations and algorithms.
- The minimax inequality is often the basis for interesting bounding procedures.

LECTURE 9

LECTURE OUTLINE

- Min-Max Problems
- Saddle Points
- Min Common/Max Crossing for Min-Max

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

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

and

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

- Minimax inequality (holds always)

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

SADDLE POINTS

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

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

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

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

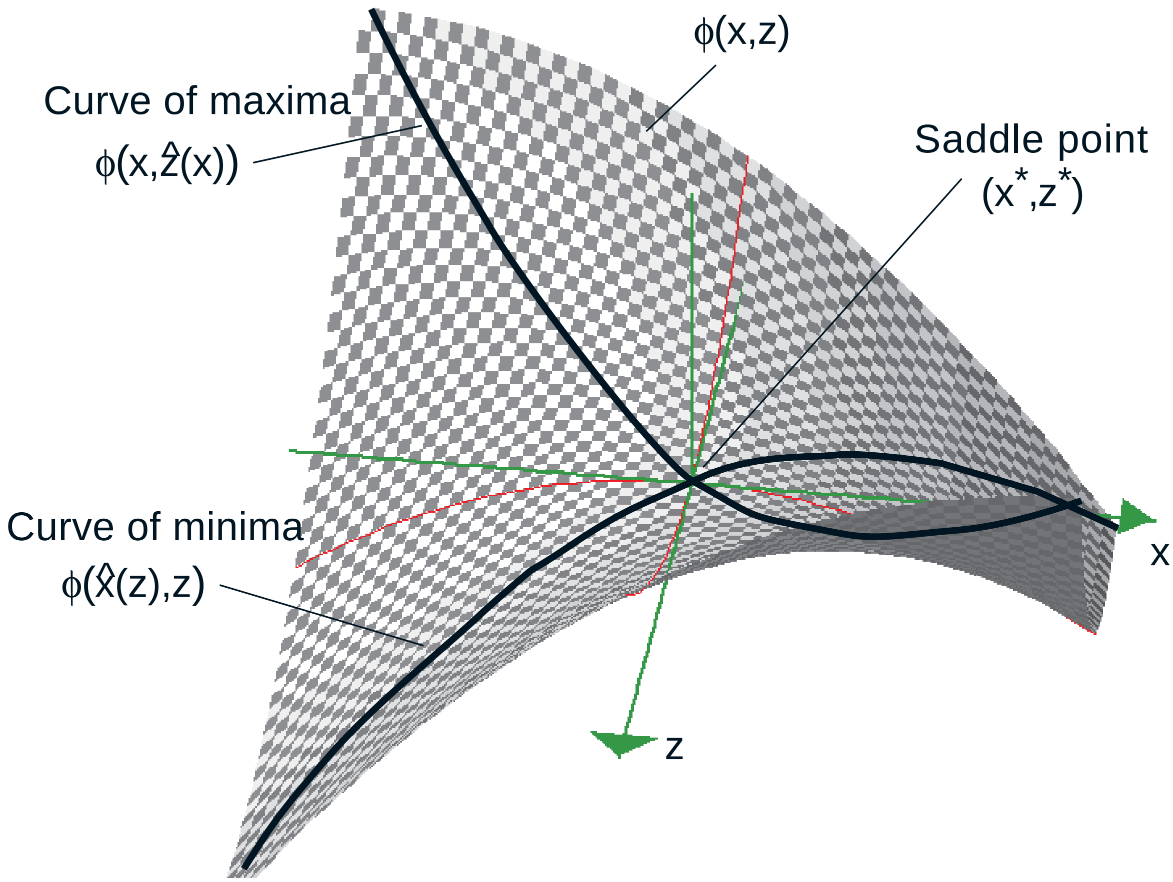
By the minimax inequality, the above holds as an equality throughout, so the minimax equality and Eq. (*) hold.

Conversely, if Eq. (*) holds, then

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

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

VISUALIZATION



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

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

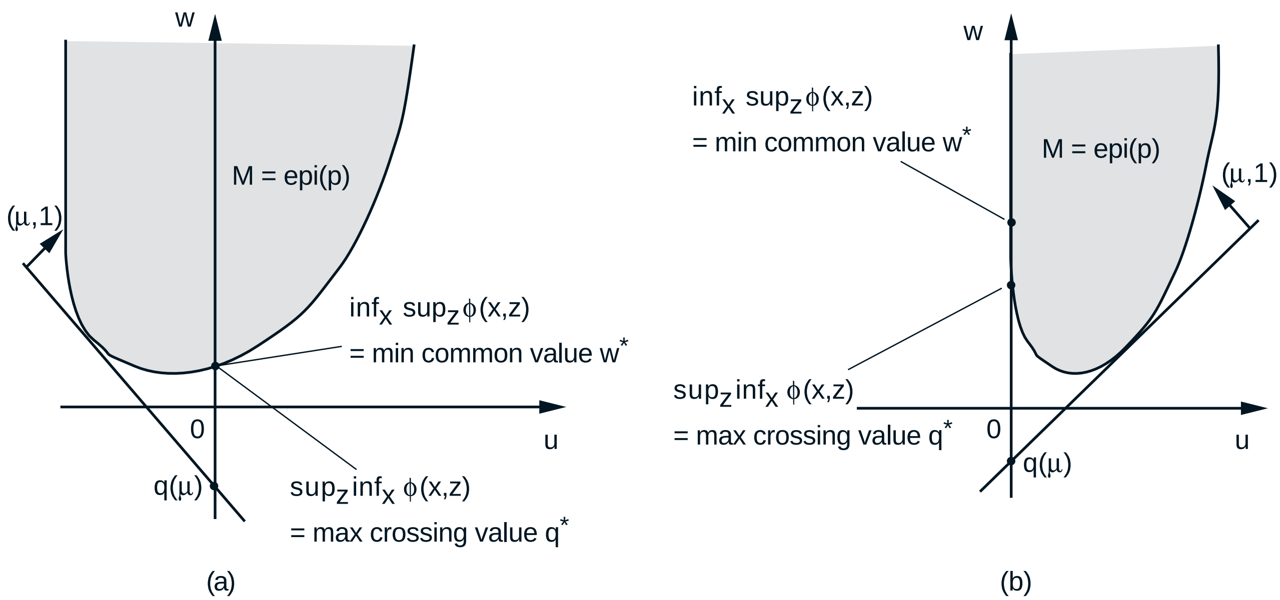
Saddle points correspond to points where these two curves meet.

MIN COMMON/MAX CROSSING FRAMEWORK

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

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

- Apply the min common/max crossing framework with $M = \text{epi}(p)$
- Note that $w^* = \inf \sup \phi$. We will show that:
  - Convexity in x implies that M is a convex set.
  - Concavity in z implies that $q^* = \sup \inf \phi$.



IMPLICATIONS OF CONVEXITY IN X

Lemma 1: Assume that X is convex and that for each $z \in Z$, the function $\phi(\cdot, z) : X \mapsto \Re$ is convex. Then p is a convex function.

Proof: Let

$$F(x, u) = \begin{cases} \sup_{z \in Z} \{\phi(x, z) - u'z\} & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

Since $\phi(\cdot, z)$ is convex, and taking pointwise supremum preserves convexity, F is convex. Since

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

and partial minimization preserves convexity, the convexity of p follows from the convexity of F. **Q.E.D.**

THE MAX CROSSING PROBLEM

- The max crossing problem is to maximize $q(\mu)$ over $\mu \in \Re^n$, where

$$q(\mu) = \inf_{(u,w)\in \mathrm{epi}(p)} \{w + \mu' u\} = \inf_{\{(u,w) \mid p(u) \le w\}} \{w + \mu' u\}$$
$$= \inf_{u \in \Re^m} \big\{p(u) + \mu' u\big\}$$

Using $p(u) = \inf_{x\in X} \sup_{z\in Z} \big\{\phi(x,z) - u'z\big\}$, we obtain

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

- By setting $z = \mu$ in the right-hand side,

$$\inf_{x\in X} \phi(x,\mu) \le q(\mu), \qquad \forall\ \mu \in Z$$

Hence, using also weak duality ($q^* \le w^*$),

$$\sup_{z\in Z} \inf_{x\in X} \phi(x,z) \le \sup_{\mu\in\Re^m} q(\mu) = q^*$$
$$\le w^* = p(0) = \inf_{x\in X} \sup_{z\in Z} \phi(x,z)$$

IMPLICATIONS OF CONCAVITY IN Z

Lemma 2: Assume that for each $x \in X$, the function $r_x : \Re^m \mapsto (-\infty, \infty]$ defined by

$$r_x(z) = \begin{cases} -\phi(x, z) & \text{if } z \in Z, \\ \infty & \text{otherwise,} \end{cases}$$

is closed and convex. Then

$$q(\mu) = \begin{cases} \inf_{x \in X} \phi(x, \mu) & \text{if } \mu \in Z, \\ -\infty & \text{if } \mu \notin Z. \end{cases}$$

Proof: (Outline) From the preceding slide,

$$\inf_{x \in X} \phi(x, \mu) \leq q(\mu), \qquad \forall\, \mu \in Z$$

We show that $q(\mu) \leq \inf_{x \in X} \phi(x, \mu)$ for all $\mu \in Z$ and $q(\mu) = -\infty$ for all $\mu \notin Z$, by considering separately the two cases where $\mu \in Z$ and $\mu \notin Z$.

First assume that $\mu \in Z$. Fix $x \in X$, and for $\epsilon > 0$, consider the point $\big(\mu, r_x(\mu) - \epsilon\big)$, which does not belong to epi(r_x). Since epi(r_x) does not contain any vertical lines, there exists a nonvertical strictly separating hyperplane ...

MINIMAX THEOREM I

Assume that:

(1) X and Z are convex.

(2) $p(0) = \inf_{x \in X} \sup_{z \in Z} \phi(x, z) < \infty$.

(3) For each $z \in Z$, the function $\phi(\cdot, z)$ is convex.

(4) For each $x \in X$, the function $-\phi(x, \cdot) : Z \mapsto \Re$ is closed and convex.

Then, the minimax equality holds if and only if the function p is lower semicontinuous at $u = 0$.

Proof: The convexity/concavity assumptions guarantee that the minimax equality is equivalent to $q^* = w^*$ in the min common/max crossing framework. Furthermore, $w^* < \infty$ by assumption, and the set $\overline{M}$ [equal to M and $\text{epi}(p)$] is convex.

By the 1st Min Common/Max Crossing Theorem, we have $w^* = q^*$ iff for every sequence $\big\{(u_k, w_k)\big\} \subset M$ with $u_k \to 0$, there holds $w^* \leq \liminf_{k \to \infty} w_k$. This is equivalent to the lower semicontinuity assumption on p:

$$p(0) \leq \liminf_{k \to \infty} p(u_k), \quad \text{for all } \{u_k\} \text{ with } u_k \to 0$$

MINIMAX THEOREM II

Assume that:

(1) X and Z are convex.

(2) $p(0) = \inf_{x \in X} \sup_{z \in Z} \phi(x, z) > -\infty$.

(3) For each $z \in Z$, the function $\phi(\cdot, z)$ is convex.

(4) For each $x \in X$, the function $-\phi(x, \cdot) : Z \mapsto \Re$ is closed and convex.

(5) 0 lies in the relative interior of $\mathrm{dom}(p)$.

Then, the minimax equality holds and the supremum in $\sup_{z \in Z} \inf_{x \in X} \phi(x, z)$ is attained by some $z \in Z$. [Also the set of z where the sup is attained is compact if 0 is in the interior of $\mathrm{dom}(p)$.]

Proof: Apply the 2nd Min Common/Max Crossing Theorem.

EXAMPLE I

- Let $X = \big\{(x_1, x_2) \mid x \geq 0\big\}$ and $Z = \{z \in \Re \mid z \geq 0\}$, and let

$$\phi(x, z) = e^{-\sqrt{x_1 x_2}} + z x_1,$$

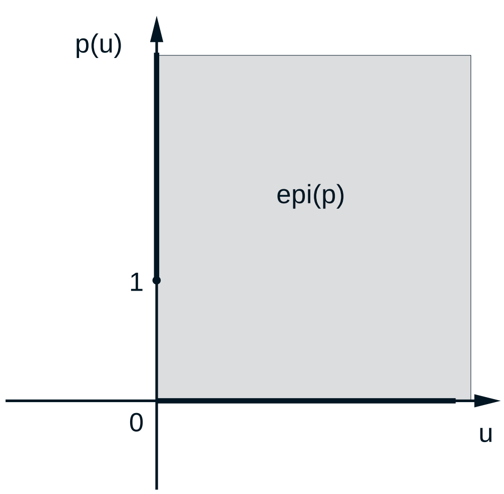
which satisfy the convexity and closedness assumptions. For all $z \geq 0$,

$$\inf_{x \geq 0} \big\{e^{-\sqrt{x_1 x_2}} + z x_1\big\} = 0,$$

so $\sup_{z \geq 0} \inf_{x \geq 0} \phi(x, z) = 0$. Also, for all $x \geq 0$,

$$\sup_{z \geq 0} \big\{e^{-\sqrt{x_1 x_2}} + z x_1\big\} = \begin{cases} 1 & \text{if } x_1 = 0, \\ \infty & \text{if } x_1 > 0, \end{cases}$$

so $\inf_{x \geq 0} \sup_{z \geq 0} \phi(x, z) = 1$.



$$p(u) = \inf_{x \geq 0} \sup_{z \geq 0} \big\{e^{-\sqrt{x_1 x_2}} + z(x_1 - u)\big\}$$

$$= \begin{cases} \infty & \text{if } u < 0, \\ 1 & \text{if } u = 0, \\ 0 & \text{if } u > 0, \end{cases}$$

EXAMPLE II

- Let $X = \Re$, $Z = \{z \in \Re \mid z \geq 0\}$, and let

$$\phi(x, z) = x + zx^2,$$

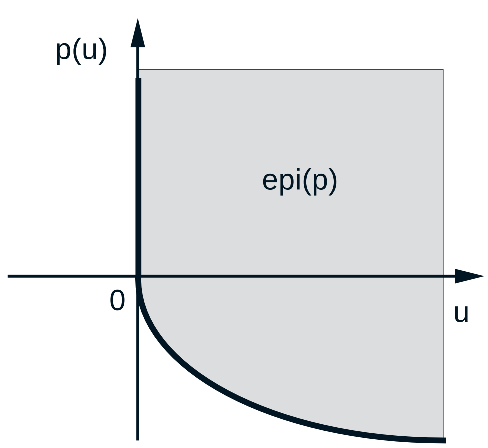
which satisfy the convexity and closedness assumptions. For all $z \geq 0$,

$$\inf_{x \in \Re} \{x + zx^2\} = \begin{cases} -1/(4z) & \text{if } z > 0, \\ -\infty & \text{if } z = 0, \end{cases}$$

so $\sup_{z \geq 0} \inf_{x \in \Re} \phi(x, z) = 0$. Also, for all $x \in \Re$,

$$\sup_{z \geq 0} \{x + zx^2\} = \begin{cases} 0 & \text{if } x = 0, \\ \infty & \text{otherwise}, \end{cases}$$

so $\inf_{x \in \Re} \sup_{z \geq 0} \phi(x, z) = 0$. However, the sup is not attained.



$$p(u) = \inf_{x \in \Re} \sup_{z \geq 0} \{x + zx^2 - uz\}$$

$$= \begin{cases} -\sqrt{u} & \text{if } u \geq 0, \\ \infty & \text{if } u < 0. \end{cases}$$

SADDLE POINT ANALYSIS

- The preceding analysis suggests the importance of the perturbation function

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

where

$$F(x, u) = \begin{cases} \sup_{z \in Z} \{\phi(x, z) - u'z\} & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

It suggests a two-step process to establish the minimax equality and the existence of a saddle point:

(1) Show that p is closed and convex, thereby showing that the minimax equality holds by using the first minimax theorem.

(2) Verify that the infimum of $\sup_{z \in Z} \phi(x, z)$ over $x \in X$, and the supremum of $\inf_{x \in X} \phi(x, z)$ over $z \in Z$ are attained, thereby showing that the set of saddle points is nonempty.

SADDLE POINT ANALYSIS (CONTINUED)

- Step (1) requires two types of assumptions:

  (a) Convexity/concavity/semicontinuity conditions of Minimax Theorem I (so the min common/max crossing framework applies).

  (b) Conditions for preservation of closedness by the partial minimization in

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

- Step (2) requires that either Weierstrass' Theorem can be applied, or else one of the conditions for existence of optimal solutions developed so far is satisfied.

SAMPLE THEOREM

- Assume convexity/concavity/semicontinuity of Φ. Consider the functions

$$t(x) = \begin{cases} \sup_{z \in Z} \phi(x, z) & \text{if } x \in X, \\ \infty & \text{if } x \notin X, \end{cases}$$

and

$$r(z) = \begin{cases} -\inf_{x \in X} \phi(x, z) & \text{if } z \in Z, \\ \infty & \text{if } z \notin Z. \end{cases}$$

Assume that they are proper.

- If the level sets of t are compact, the minimax equality holds, and the min over x of

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

[which is $t(x)$] is attained.

- If the level sets of t and r are compact, the set of saddle points is nonempty and compact.

SADDLE POINT THEOREM

Assume the convexity/concavity/semicontinuity conditions, and that any *one* of the following holds:

(1) X and Z are compact.

(2) Z is compact and there exists a vector $\overline{z} \in Z$ and a scalar γ such that the level set $\big\{x \in X \mid \phi(x, \overline{z}) \le \gamma\big\}$ is nonempty and compact.

(3) X is compact and there exists a vector $\overline{x} \in X$ and a scalar γ such that the level set $\big\{z \in Z \mid \phi(\overline{x}, z) \ge \gamma\big\}$ is nonempty and compact.

(4) There exist vectors $\overline{x} \in X$ and $\overline{z} \in Z$, and a scalar γ such that the level sets

$$\big\{x \in X \mid \phi(x, \overline{z}) \le \gamma\big\}, \quad \big\{z \in Z \mid \phi(\overline{x}, z) \ge \gamma\big\},$$

are nonempty and compact.

Then, the minimax equality holds, and the set of saddle points of ϕ is nonempty and compact.

LECTURE 10

LECTURE OUTLINE

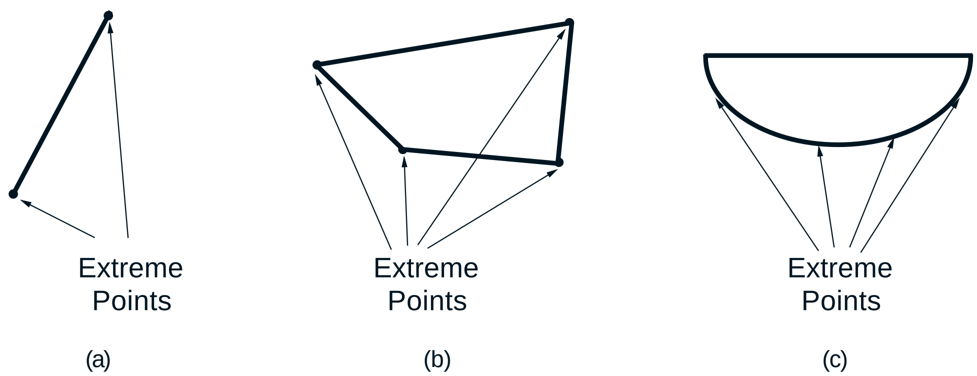
- Extreme points
- Polar cones and polar cone theorem
- Polyhedral and finitely generated cones
- Farkas Lemma, Minkowski-Weyl Theorem

- The main convexity concepts so far have been:
  - Closure, convex hull, affine hull, rel. interior
  - Directions of recession and set intersection theorems
  - Preservation of closure under linear transformation and partial minimization
  - Existence of optimal solutions
  - Hyperplanes, min common/max crossing duality, and application in minimax

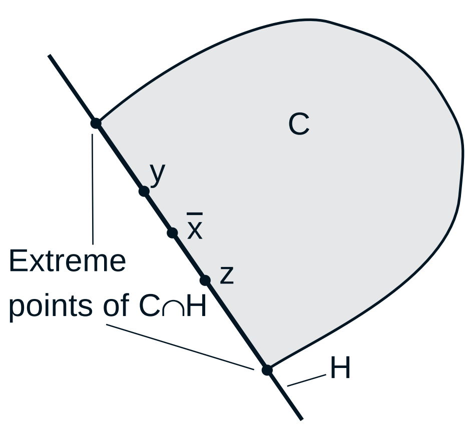
- We now introduce new concepts with important theoretical and algorithmic implications: extreme points, polyhedral convexity, and related issues.

EXTREME POINTS

- A vector x is an *extreme point* of a convex set C if $x \in C$ and x does not lie strictly within a line segment contained in C.



Proposition: Let C be closed and convex. If H is a hyperplane that contains C in one of its closed halfspaces, then every extreme point of $C \cap H$ is also an extreme point of C.



Proof: If $\overline{x} \in C \cap H$ is a nonextreme point of C, it lies strictly within a line segment $[y, z] \subset C$. If y belongs in the open upper halfspace of H, then z must belong to the open lower halfspace of H - contradiction since H supports C. Hence $y, z \in C \cap H$, implying that $\overline{x}$ is a nonextreme point of $C \cap H$.

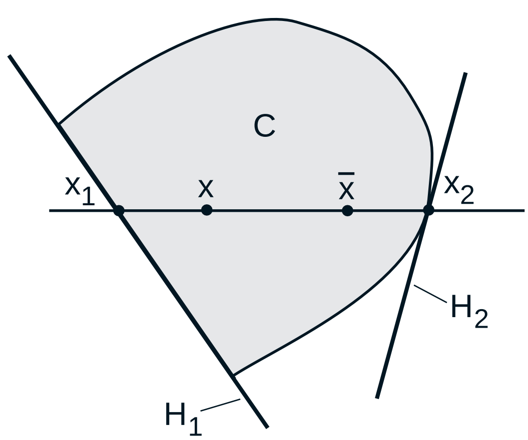
PROPERTIES OF EXTREME POINTS I

Krein-Milman Theorem: A convex and compact set is equal to the convex hull of its extreme points.

Proof: By convexity, the given set contains the convex hull of its extreme points.

Next show the reverse, i.e, every x in a compact and convex set C can be represented as a convex combination of extreme points of C.

Use induction on the dimension of the space. The result is true in $\Re$. Assume it is true for all convex and compact sets in $\Re^{n-1}$. Let $C \subset \Re^n$ and $x \in C$.



If $\overline{x}$ is another point in C, the points x_1 and x_2 shown can be represented as convex combinations of extreme points of the lower dimensional convex and compact sets $C \cap H_1$ and $C \cap H_2$, which are also extreme points of C, by the preceding theorem.

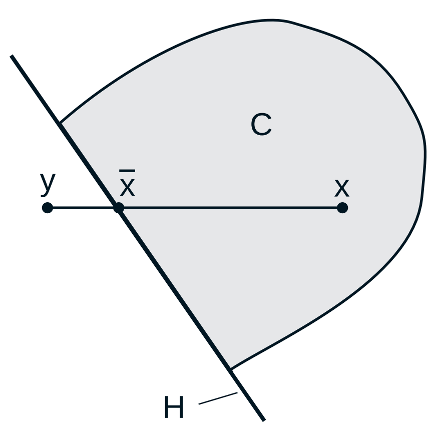
PROPERTIES OF EXTREME POINTS II

Proposition: A closed convex set has at least one extreme point if and only if it does not contain a line.

Proof: If C contains a line, then this line translated to pass through an extreme point is fully contained in C (use the Recession Cone Theorem) - impossible.

Conversely, we use induction on the dimension of the space to show that if C does not contain a line, it must have an extreme point. True in $\Re$, so assume it is true in $\Re^{n-1}$, where $n \geq 2$. We will show it is true in $\Re^n$.

Since C does not contain a line, there must exist points $x \in C$ and $y \notin C$. Consider the relative boundary point $\overline{x}$.



The set $C \cap H$ lies in an $(n-1)$-dimensional space and does not contain a line, so it contains an extreme point. By the preceding proposition, this extreme point must also be an extreme point of C.

CHARACTERIZATION OF EXTREME POINTS

Proposition: Consider a polyhedral set

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

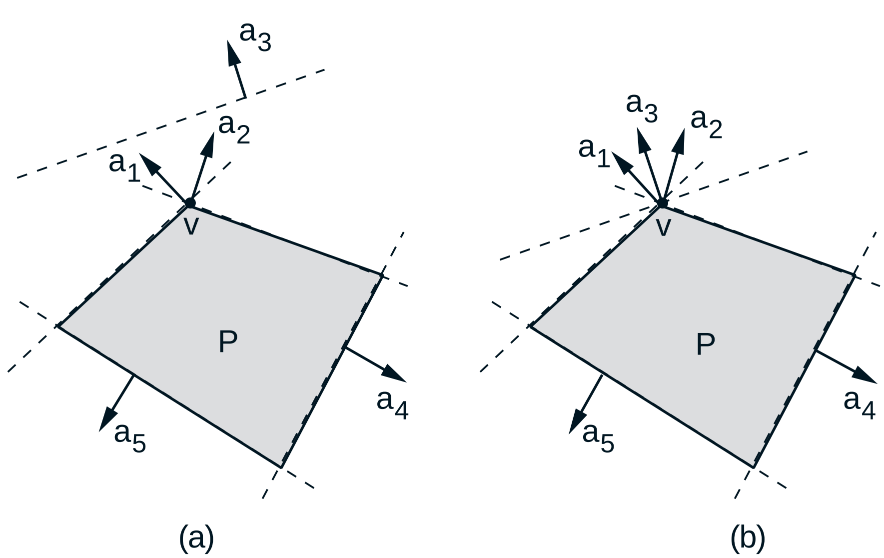
where a_j and b_j are given vectors and scalars.

(a) A vector $v \in P$ is an extreme point of P if and only if the set

$$A_v = \{a_j \mid a_j'v = b_j,\ j \in \{1, \ldots, r\}\}$$

contains n linearly independent vectors.

(b) P has an extreme point if and only if the set $\{a_j \mid j = 1, \ldots, r\}$ contains n linearly independent vectors.



PROOF OUTLINE

If the set A_v contains fewer than n linearly independent vectors, then the system of equations

$$a_j'w = 0, \qquad \forall\ a_j \in A_v$$

has a nonzero solution $\overline{w}$. For small $\gamma > 0$, we have $v + \gamma\overline{w} \in P$ and $v - \gamma\overline{w} \in P$, thus showing that v is not extreme. Thus, if v is extreme, A_v must contain n linearly independent vectors.

Conversely, assume that A_v contains a subset $\bar{A}_v$ of n linearly independent vectors. Suppose that for some $y \in P$, $z \in P$, and $\alpha \in (0, 1)$, we have $v = \alpha y + (1 - \alpha)z$. Then, for all $a_j \in \bar{A}_v$,

$$b_j = a_j'v = \alpha a_j'y + (1-\alpha)a_j'z \le \alpha b_j + (1-\alpha)b_j = b_j$$

Thus, v, y, and z are all solutions of the system of n linearly independent equations

$$a_j'w = b_j, \qquad \forall\ a_j \in \bar{A}_v$$

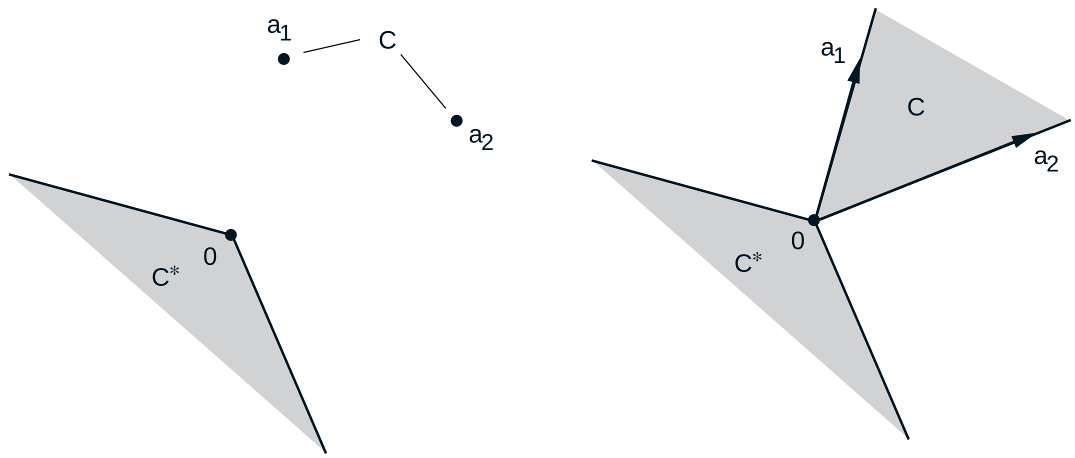
Hence, $v = y = z$, implying that v is an extreme point of P.

POLAR CONES

- Given a set C, the cone given by

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

is called the *polar cone* of C.



(a) (b)

- C^* is a closed convex cone, since it is the intersection of closed halfspaces.

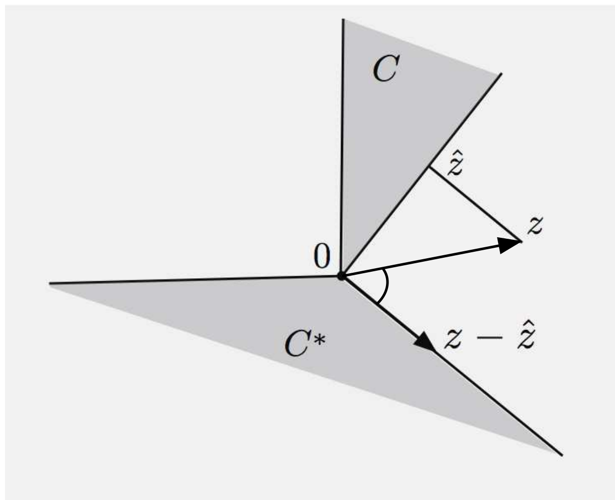
- Note that

$$C^* = \big(\mathrm{cl}(C)\big)^* = \big(\mathrm{conv}(C)\big)^* = \big(\mathrm{cone}(C)\big)^*$$

- Special case: If C is a subspace, $C^* = C^\perp$. In this case, we have $(C^*)^* = (C^\perp)^\perp = C$.

POLAR CONE THEOREM

- For any cone C, we have $(C^*)^* = \mathrm{cl}\big(\mathrm{conv}(C)\big)$. If C is closed and convex, we have $(C^*)^* = C$.



Proof: Consider the case where C is closed and convex. For any $x \in C$, we have $x'y \le 0$ for all $y \in C^*$, so that $x \in (C^*)^*$, and $C \subset (C^*)^*$.

To prove that $(C^*)^* \subset C$, we show that for any $z \in \Re^n$ and its projection on C, call it $\hat{z}$, we have $z - \hat{z} \in C^*$, so if $z \in (C^*)^*$, the geometry shown in the figure [(angle between z and $z - \hat{z}$) $< \pi/2$] is impossible, and we must have $z - \hat{z} = 0$, i.e., $z \in C$.

POLARS OF POLYHEDRAL CONES

- A cone $C \subset \Re^n$ is *polyhedral*, if

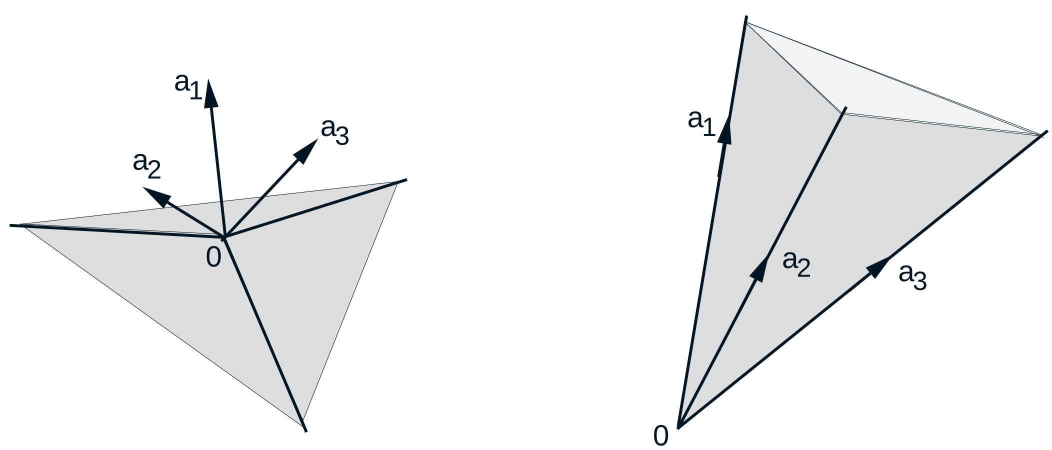
$$C = \{x \mid a_j' x \le 0,\ j = 1, \ldots, r\},$$

where $a_1, \ldots, a_r$ are some vectors in $\Re^n$.

- A cone $C \subset \Re^n$ is *finitely generated*, if

$$C = \left\{ x \;\middle|\; x = \sum_{j=1}^{r} \mu_j a_j,\ \mu_j \ge 0,\ j = 1, \ldots, r \right\} = \text{cone}\big(\{a_1, \ldots, a_r\}\big),$$

where $a_1, \ldots, a_r$ are some vectors in $\Re^n$.



(a) (b)

FARKAS-MINKOWSKI-WEYL THEOREMS

Let $a_1, \ldots, a_r \in \Re^n$.

(a) (*Farkas' Lemma*) We have

$$\big(\{y \mid a_j'y \le 0,\ j = 1, \ldots, r\}\big)^* = \mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)$$

(There is also a version of this involving sets described by linear equality as well as inequality constraints.)

(b) (*Minkowski-Weyl Theorem*) A cone is polyhedral if and only if it is finitely generated.

(c) (*Minkowski-Weyl Representation*) A set P is polyhedral if and only if

$$P = \mathrm{conv}\big(\{v_1, \ldots, v_m\}\big) + C,$$

for a nonempty finite set of vectors $\{v_1, \ldots, v_m\}$ and a finitely generated cone C.

PROOF OUTLINE

- $\{y \mid a_j'y \le 0,\ j = 1, \ldots, r\}$ is closed
- $\text{cone}(\{a_1, \ldots, a_r\})$ is closed,because it is the result of a linear transformation A applied to the polyhedral set $\{\mu \mid \mu \ge 0,\ \sum_{j=1}^{r} \mu_j = 1\}$, where A is the matrix with columns $a_1, \ldots, a_r$.
- By the definition of polar cone

$$\big(\text{cone}(\{a_1, \ldots, a_r\})\big)^* = \big\{y \mid a_j'y \le 0,\ j = 1, \ldots, r\big\}.$$

- By the Polar Cone Theorem

$$\Big(\big(\text{cone}(\{a_1, \ldots, a_r\})\big)^*\Big)^* = \Big(\big\{y \mid a_j'y \le 0,\ j = 1, \ldots, r\big\}\Big)^*$$

so by closedness

$$\text{cone}(\{a_1, \ldots, a_r\}) = \Big(\big\{y \mid a_j'y \le 0,\ j = 1, \ldots, r\big\}\Big)^*.$$

Q.E.D.

- Proofs of (b), (c) will be given in the next lecture.

LECTURE 11

LECTURE OUTLINE

- Proofs of Minkowski-Weyl Theorems
- Polyhedral aspects of optimization
- Linear programming and duality
- Integer programming

−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−

Recall some of the facts of polyhedral convexity:

- Polarity relation between polyhedral and finitely generated cones

$$\{x \mid a_j'x \leq 0,\ j = 1, \ldots, r\} = \text{cone}\big(\{a_1, \ldots, a_r\}\big)^*$$

- Farkas' Lemma

$$\{x \mid a_j'x \leq 0,\ j = 1, \ldots, r\}^* = \text{cone}\big(\{a_1, \ldots, a_r\}\big)$$

- Minkowski-Weyl Theorem: a cone is polyhedral iff it is finitely generated.
- A corollary (essentially) to be shown:

$$\text{Polyhedral set } P = \text{conv}\big(\{v_1, \ldots, v_m\}\big) + R_P$$

for some finite set of vectors $\{v_1, \ldots, v_m\}$.

MINKOWSKI-WEYL PROOF OUTLINE

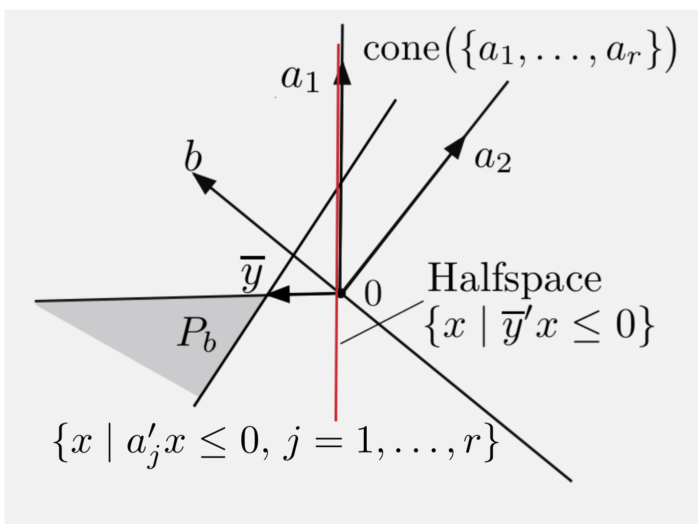
- **Step 1:** Show $\text{cone}(\{a_1,\ldots,a_r\})$ is polyhedral.

- **Step 2:** Use Step 1 and Farkas to show that $\{x \mid a_j'x \le 0,\ j = 1,\ldots,r\}$ is finitely generated.

- **Proof of Step 1:** Assume first that $a_1,\ldots,a_r$ span $\Re^n$. Given $b \notin \text{cone}(\{a_1,\ldots,a_r\})$,

$$P_b = \{y \mid b'y \ge 1,\ a_j'y \le 0,\ j = 1,\ldots,r\}$$

is nonempty and has at least one extreme point $\overline{y}$.



- Show that $b'\overline{y} = 1$ and $\{a_j \mid a_j'\overline{y} = 0\}$ contains $n-1$ linearly independent vectors. The halfspace $\{x \mid \overline{y}'x \le 0\}$, contains $\text{cone}(\{a_1,\ldots,a_r\})$, and does not contain b. Consider the intersection of all such halfspaces as b ranges over $\text{cone}(\{a_1,\ldots,a_r\})$.

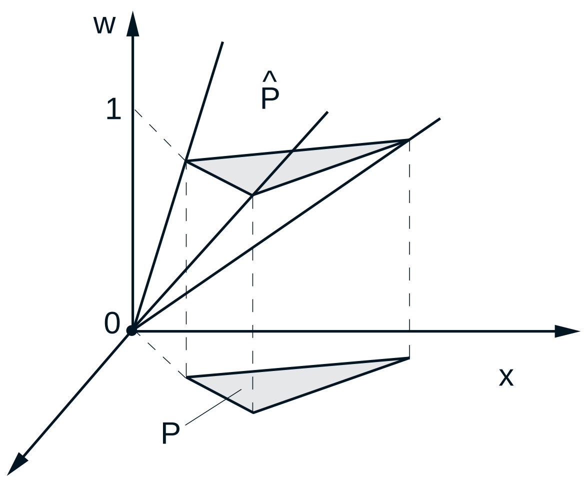
POLYHEDRAL REPRESENTATION PROOF

- We "lift the polyhedral set into a cone". Let

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

$$\hat{P} = \{(x, w) \mid 0 \le w,\ a_j'x \le b_j w,\ j = 1, \ldots, r\}$$

and note that $P = \{x \mid (x, 1) \in \hat{P}\}$.



- By Minkowski-Weyl, $\hat{P}$ is finitely generated, so

$$\hat{P} = \left\{(x, w) \;\middle|\; x = \sum_{j=1}^{m} \mu_j v_j,\ w = \sum_{j=1}^{m} \mu_j d_j,\ \mu_j \ge 0\right\}.$$

We have $d_j \ge 0$ for all j, since $w \ge 0$ for all $(x, w) \in \hat{P}$. Let $J^+ = \{j \mid d_j > 0\}$, $J^0 = \{j \mid d_j = 0\}$.

PROOF CONTINUED

- By replacing μ_j by μ_j/d_j for all $j \in J^+$,

$$\hat{P} = \left\{ (x, w) \;\middle|\; x = \sum_{j \in J^+ \cup J^0} \mu_j v_j,\; w = \sum_{j \in J^+} \mu_j,\; \mu_j \geq 0 \right\}$$

Since $P = \big\{x \mid (x, 1) \in \hat{P}\big\}$, we obtain

$$P = \left\{ x \;\middle|\; x = \sum_{j \in J^+ \cup J^0} \mu_j v_j,\; \sum_{j \in J^+} \mu_j = 1,\; \mu_j \geq 0 \right\}$$

Thus,

$$P = \operatorname{conv}\big(\{v_j \mid j \in J^+\}\big) + \left\{ \sum_{j \in J^0} \mu_j v_j \;\middle|\; \mu_j \geq 0,\; j \in J^0 \right\}$$

- To prove that the vector sum of $\operatorname{conv}\big(\{v_1, \ldots, v_m\}\big)$ and a finitely generated cone is a polyhedral set, we reverse the preceding argument. **Q.E.D.**

POLYHEDRAL CALCULUS

- The intersection and Cartesian product of polyhedral sets is polyhedral.

- The image of a polyhedral set under a linear transformation is polyhedral: To show this, let the polyhedral set P be represented as

$$P = \mathrm{conv}\big(\{v_1, \ldots, v_m\}\big) + \mathrm{cone}\big(\{a_1, \ldots, a_r\}\big),$$

and let A be a matrix. We have

$$AP = \mathrm{conv}\big(\{Av_1, \ldots, Av_m\}\big) + \mathrm{cone}\big(\{Aa_1, \ldots, Aa_r\}\big).$$

It follows that AP has a Minkowski-Weyl representation, and hence it is polyhedral.

- The vector sum of polyhedral sets is polyhedral (since vector sum operation is a special type of linear transformation).

POLYHEDRAL FUNCTIONS

- A function $f : \Re^n \mapsto (-\infty, \infty]$ is *polyhedral* if its epigraph is a polyhedral set in $\Re^{n+1}$.
- Note that every polyhedral function is closed, proper, and convex.

Theorem: Let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function. Then f is polyhedral if and only if $\mathrm{dom}(f)$ is a polyhedral set, and

$$f(x) = \max_{j=1,\ldots,m} \{a_j' x + b_j\}, \qquad \forall\, x \in \mathrm{dom}(f),$$

for some $a_j \in \Re^n$ and $b_j \in \Re$.

Proof: Assume that $\mathrm{dom}(f)$ is polyhedral and f has the above representation. We will show that f is polyhedral. The epigraph of f is

$$\begin{aligned} \mathrm{epi}(f) = &\big\{(x, w) \mid x \in \mathrm{dom}(f)\big\} \\ &\cap \big\{(x, w) \mid a_j' x + b_j \le w,\ j = 1, \ldots, m\big\}. \end{aligned}$$

Since the two sets on the right are polyhedral, $\mathrm{epi}(f)$ is also polyhedral. Hence f is polyhedral.

PROOF CONTINUED

- Conversely, if f is polyhedral, its epigraph is polyhedral and can be represented as the intersection of a finite collection of closed halfspaces of the form $\{(x,w) \mid a_j'x + b_j \leq c_j w\}$, $j = 1,\ldots,r$, where $a_j \in \Re^n$, and $b_j, c_j \in \Re$.

- Since for any $(x,w) \in \mathrm{epi}(f)$, we have $(x, w+\gamma) \in \mathrm{epi}(f)$ for all $\gamma \geq 0$, it follows that $c_j \geq 0$, so by normalizing if necessary, we may assume without loss of generality that either $c_j = 0$ or $c_j = 1$. Letting $c_j = 1$ for $j = 1,\ldots,m$, and $c_j = 0$ for $j = m+1,\ldots,r$, where m is some integer,

$$\mathrm{epi}(f) = \big\{(x,w) \mid a_j'x + b_j \leq w,\ j = 1,\ldots,m,\\ a_j'x + b_j \leq 0,\ j = m+1,\ldots,r\big\}.$$

Thus

$$\mathrm{dom}(f) = \big\{x \mid a_j'x + b_j \leq 0,\ j = m+1,\ldots,r\big\},$$

$$f(x) = \max_{j=1,\ldots,m} \{a_j'x + b_j\}, \qquad \forall\, x \in \mathrm{dom}(f)$$

Q.E.D.

OPERATIONS ON POLYHEDRAL FUNCTIONS

- The preceding representation of polyhedral functions can be used to derive various properties.

- The sum of polyhedral functions is polyhedral (provided their domains have a point in common).

- If g is polyhedral and A is a matrix, the function $f(x) = g(Ax)$ is polyhedral.

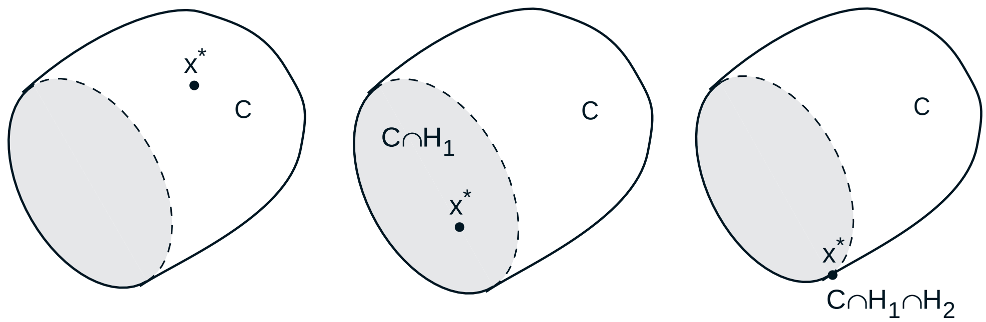
- Let F be a polyhedral function of (x, z). Then the function f obtained by the partial minimization

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

is polyhedral (assuming it is proper).

EXTREME POINTS AND CONCAVE MIN.

- Let C be a closed and convex set that has at least one extreme point. A concave function $f : C \mapsto \Re$ that attains a minimum over C attains the minimum at some extreme point of C.



(a) (b) (c)

Proof (abbreviated): If a minimum x^* belongs to $\mathrm{ri}(C)$ [see Fig. (a)], f must be constant over C, so it attains a minimum at an extreme point of C. If $x^* \notin \mathrm{ri}(C)$, there is a hyperplane H_1 that supports C and contains x^*.

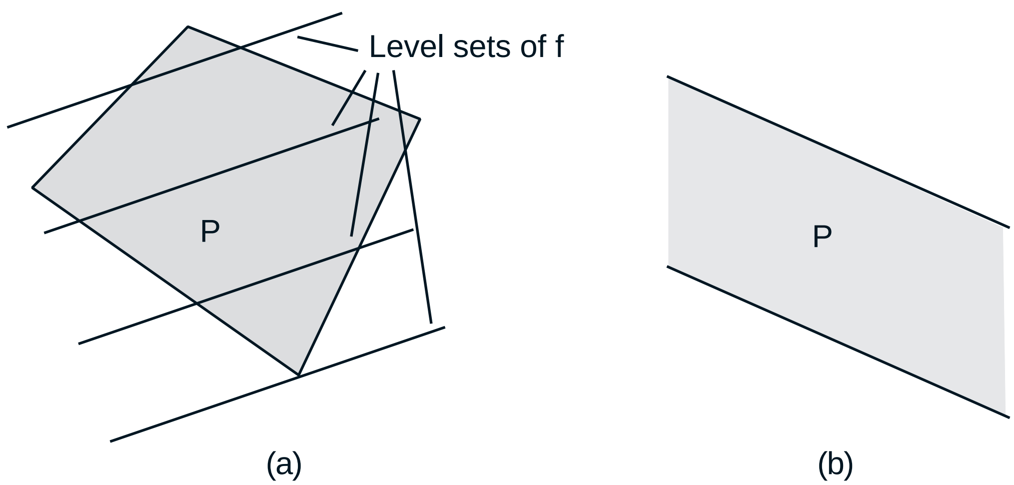
If $x^* \in \mathrm{ri}(C \cap H_1)$ [see (b)], then f must be constant over $C \cap H_1$, so it attains a minimum at an extreme point $C \cap H_1$. This optimal extreme point is also an extreme point of C. If $x^* \notin \mathrm{ri}(C \cap H_1)$, there is a hyperplane H_2 supporting $C \cap H_1$ through x^*. Continue until an optimal extreme point is obtained (which must also be an extreme point of C).

FUNDAMENTAL THEOREM OF LP

- Let P be a polyhedral set that has at least one extreme point. Then, if a linear function is bounded below over P, it attains a minimum at some extreme point of P.

Proof: Since the cost function is bounded below over P, it attains a minimum. The result now follows from the preceding theorem. **Q.E.D.**

- Two possible cases in LP: In (a) there is an extreme point; in (b) there is none.



LINEAR PROGRAMMING DUALITY

- Primal problem (optimal value $= f^*$):

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

where c and $a_1, \ldots, a_r$ are vectors in $\Re^n$.

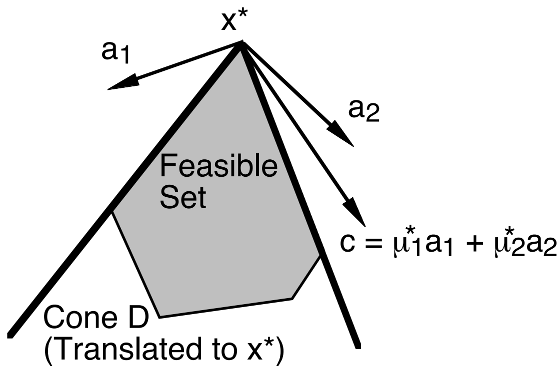
- Dual problem (optimal value $= q^*$):

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

- $f^* = \min_x \max_{\mu \geq 0} L$ and $q^* = \max_{\mu \geq 0} \min_x L$, where $L(x, \mu) = c'x + \sum_{j=1}^{r} \mu_j (b_j - a_j'x)$
- **Duality Theorem:**
  (a) If either f^* or q^* is finite, then $f^* = q^*$ and both problems have optimal solutions.
  (b) If $f^* = -\infty$, then $q^* = -\infty$.
  (c) If $q^* = \infty$, then $f^* = \infty$.

Proof: Use weak duality ($q^* \leq f^*$) and Farkas' Lemma (see next slide).

LINEAR PROGRAMMING DUALITY PROOF



Assume f^*: finite, and let x^* be a primal optimal solution (it exists because f^* is finite). Let J be the set of indices j with $a_j'x^* = b_j$. Then, $c'y \geq 0$ for all y in the cone $D = \{y \mid a_j'y \geq 0, \ \forall\, j \in J\}$. By Farkas',

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

Take inner product with x^*:

$$c'x^* = \sum_{j=1}^{r} \mu_j^* a_j{}' x^* = \sum_{j=1}^{r} \mu_j^* b_j = b'\mu^*.$$

This, together with $q^* \leq f^*$, implies that $q^* = f^*$ and that μ^* is optimal.

INTEGER PROGRAMMING

- Consider a polyhedral set

$$P = \{x \mid Ax = b,\ c \le x \le d\},$$

where A is $m \times n$, $b \in \Re^m$, and $c, d \in \Re^n$. Assume that all components of A and b, c, and d are integer.

- Question: Under what conditions do the extreme points of P have integer components?

Definition: A square matrix with integer components is *unimodular* if its determinant is 0, 1, or -1. A rectangular matrix with integer components is *totally unimodular* if each of its square submatrices is unimodular.

Theorem: If A is totally unimodular, all the extreme points of P have integer components.

- Most important special case: Linear network optimization problems (with "single commodity" and no "side constraints"), where A is the, so-called, *arc incidence matrix* of a given directed graph.

LECTURE 12

LECTURE OUTLINE

- Theorems of the Alternative - LP Applications
- Hyperplane proper polyhedral separation
- Min Common/Max Crossing Theorem under polyhedral assumptions

****

- Primal problem (optimal value $= f^*$):

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

where c and $a_1, \ldots, a_r$ are vectors in $\Re^n$.
- Dual problem (optimal value $= q^*$):

$$\text{maximize } b'\mu$$
$$\text{subject to } \sum_{j=1}^{r} a_j \mu_j = c, \qquad \mu_j \geq 0,\ j = 1, \ldots, r.$$

- Duality: $q^* = f^*$ (if finite) and solutions exist

LP OPTIMALITY CONDITIONS

Proposition: 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

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

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

$$\begin{aligned} b'\mu^* &= \sum_{j=1}^{r} b_j \mu_j^* + \left(c - \sum_{j=1}^{r} a_j \mu_j^* \right)' x^* \\ &= c'x^* + \sum_{j=1}^{r} \mu_j^*(b_j - a_j' x^*). \end{aligned} \tag{2}$$

Thus, if Eq. (1) holds, we have $b'\mu^* = c'x^*$, and weak duality implies optimality of x^* and μ^*.

Conversely, if (x^*, μ^*) are an optimal pair, then x^* is primal-feasible, μ^* is dual-feasible, and by the duality theorem, $b'\mu^* = c'x^*$. From Eq. (2), we obtain Eq. (1). **Q.E.D.**

THEOREMS OF THE ALTERNATIVE

- We consider conditions for feasibility, strict feasibility, and boundedness of systems of linear inequalities

- Example: Farkas' lemma which states that the system $Ax = c$, $x \geq 0$ has a solution if and only if

$$A'y \leq 0 \qquad \Rightarrow \qquad c'y \leq 0.$$

- Can be stated as a "theorem of the alternative", i.e., exactly one of the following two holds:

  (1) The system $Ax = c$, $x \geq 0$ has a solution

  (2) The system $A'y \leq 0$, $c'y > 0$ has no solution

- Another example: Gordan's Theorem which states that for any nonzero vectors $a_1, \ldots, a_r$, exactly one of the following two holds:

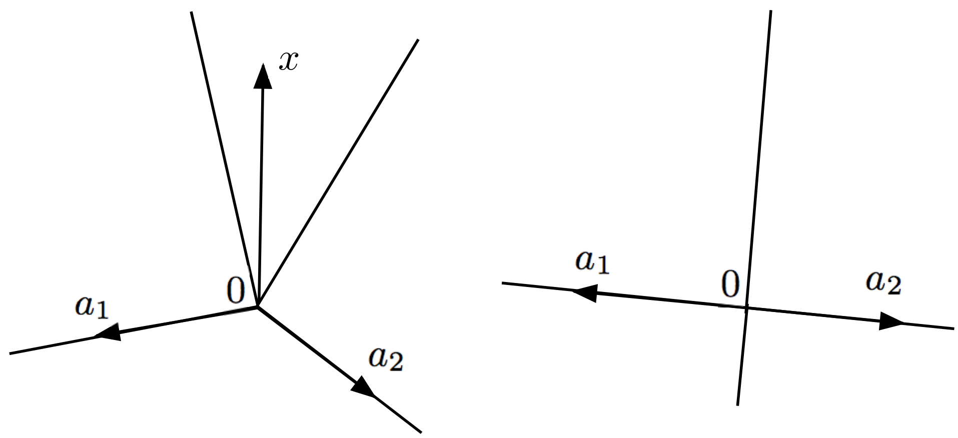
  (1) There exists x s.t. $a_1'x < 0, \ldots, a_r'x < 0$

  (2) There exists $\mu = (\mu_1, \ldots, \mu_r)$ s.t. $\mu \neq 0$, $\mu \geq 0$, and

$$\mu_1 a_1 + \cdots + \mu_r a_r = 0$$

GORDAN'S THEOREM

- Geometrically, $\big(\mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)\big)^*$ has nonempty interior iff $\mathrm{cone}\big(\{a_1, \ldots, a_r\}\big)$ contains a line



- **Gordan's Theorem - Generalized:** Let A be an $m \times n$ matrix and b be a vector in $\Re^m$. The following are equivalent:

  (i) There exists $x \in \Re^n$ such that $Ax < b$.

  (ii) For every $\mu \in \Re^m$,

$$\mu \geq 0, \quad A'\mu = 0, \quad \mu'b \leq 0 \qquad \Rightarrow \qquad \mu = 0$$

  (iii) Any polyhedral set of the form

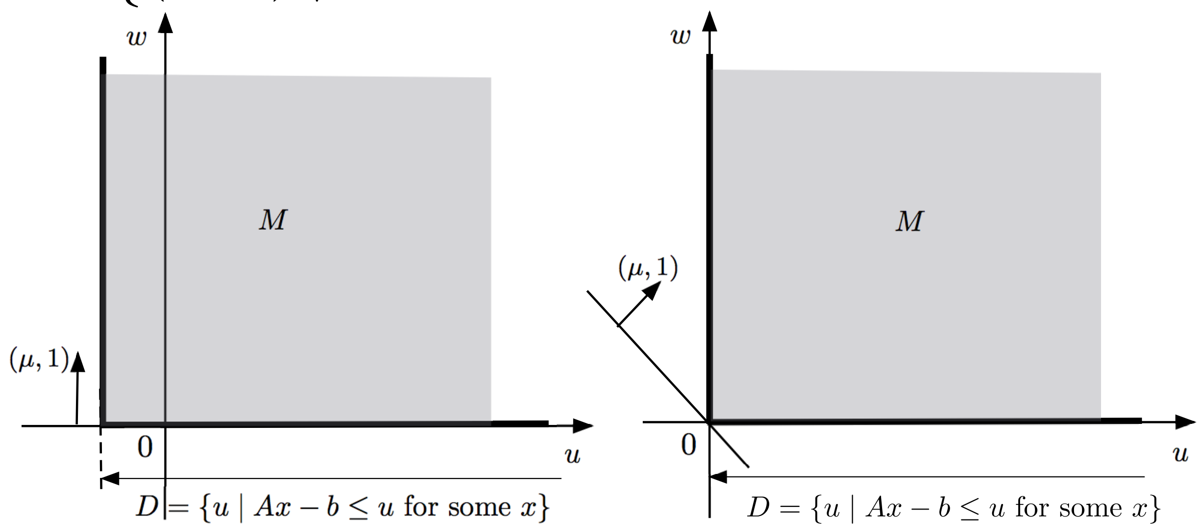
$$\{\mu \mid A'\mu = c,\, \mu'b \leq d,\, \mu \geq 0\},$$

  where $c \in \Re^n$ and $d \in \Re$, is compact.

PROOF OF GORDAN'S THEOREM

- Application of Min Common/Max Crossing with

$$M = \big\{(u, w) \mid w \geq 0,\ Ax - b \leq u \text{ for some } x \in \Re^n\big\}$$



- Condition (i) of G. Th. is equivalent to 0 being an interior point of the projection of M

$$D = \{u \mid Ax - b \leq u \text{ for some } x \in \Re^n\}$$

- Condition (ii) of G. Th. is equivalent to the max crossing solution set being nonempty and compact, or 0 being the only max crossing solution

- Condition (ii) of G. Th. is also equivalent to

Recession Cone of $\{\mu \mid A'\mu = c,\ \mu'b \leq d,\ \mu \geq 0\} = \{0\}$

which is equivalent to Condition (iii) of G. Th.

STIEMKE'S TRANSPOSITION THEOREM

- The most general theorem of the alternative for linear inequalities is Motzkin's Theorem (involves a mixture of equalities, inequalities, and strict inequalities).

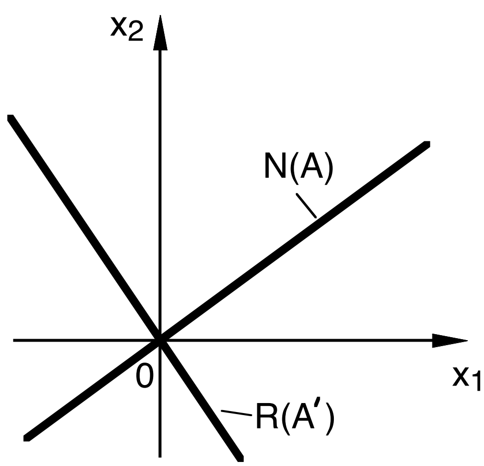
- It can be proved again using min common/max crossing. A special case is the following:

- **Stiemke's Transposition Theorem:** Let A be an $m \times n$ matrix, and let c be a vector in $\Re^m$. The system

$$Ax = c, \qquad x > 0$$

has a solution if and only if

$$A'\mu \geq 0 \text{ and } c'\mu \leq 0 \quad \Rightarrow \quad A'\mu = 0 \text{ and } c'\mu = 0$$



LP: STRICT FEASIBILITY - COMPACTNESS

- We say that *the primal linear program is strictly feasible* if there exists a primal-feasible vector x such that $a_j'x > b_j$ for all $j = 1, \ldots, r$.
- We say that *the dual linear program is strictly feasible* if there exists a dual-feasible vector μ with $\mu > 0$.

Proposition: Consider the primal and dual linear programs, and assume that their common optimal value is finite. Then:

(a) The dual optimal solution set is compact if and only if the primal problem is strictly feasible.

(b) Assuming that the set $\{a_1, \ldots, a_r\}$ contains n linearly independent vectors, the primal optimal solution set is compact if and only if the dual problem is strictly feasible.

Proof: (a) Apply Gordan's Theorem.

(b) Apply Stiemke's Transposition Theorem.

PROPER POLYHEDRAL SEPARATION

- Recall that two convex sets C and P such that

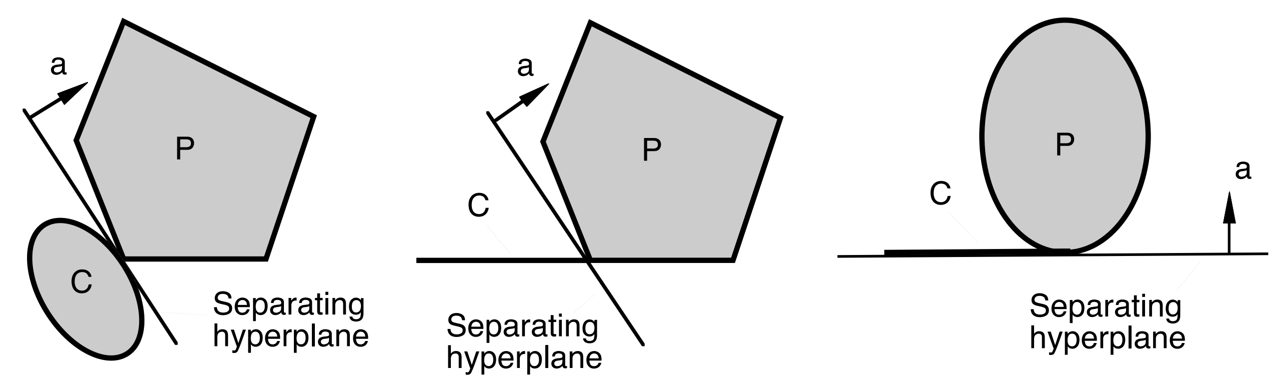
$$\mathrm{ri}(C) \cap \mathrm{ri}(P) = \emptyset$$

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.

MIN C/MAX C TH. III - POLYHEDRAL

- Consider the min common and max crossing problems, and assume the following:

  (1) $-\infty < w^*$.

  (2) The set $\overline{M}$ has the form

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

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

  (3) We have

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

  where

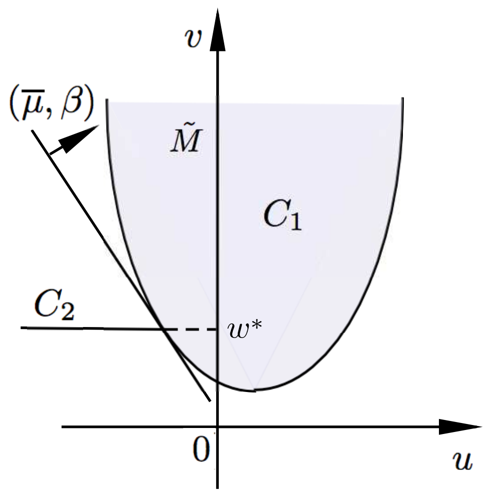
$$\tilde{D} = \{u \mid \text{there exists } w \in \Re \text{ with } (u, w) \in \tilde{M}\}$$

Then $q^* = w^*$, and Q^*, the set of optimal solutions of the max crossing problem, is a nonempty subset of R_P^*, the polar cone of the recession cone of P.

- Also, Q^* is compact if $\text{int}(\tilde{D}) \cap P \neq \emptyset$.

PROOF OF MIN C/MAX C TH. III

- Consider the disjoint convex sets



$$C_1 = \Big\{(u, v) \mid v > w \text{ for some } (u, w) \in \tilde{M}\Big\}$$

$$C_2 = \Big\{(u, w^*) \mid u \in P\Big\}$$

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

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

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

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,

$$w^* + \overline{\mu}' z \le \inf_{(u,v)\in C_1} \{v + \overline{\mu}' u\}, \qquad \forall\, z \in P, \quad (1)$$

which in particular implies that $\overline{\mu}' d \le 0$ for all d in the recession cone of P. Hence $\overline{\mu}$ belongs to the polar of this recession cone.

From Eq. (1), we also obtain

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

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

The proof of compactness of Q^* if $\mathrm{int}(\tilde{D}) \cap P \ne \emptyset$ is similar to the one of the nonpolyhedral MC/MC Theorem. **Q.E.D.**

MIN C/MAX C TH. III - A SPECIAL CASE

- Consider the min common and max crossing problems, and assume that:

  (1) The set $\overline{M}$ is defined in terms of a convex function $f : \Re^m \mapsto (-\infty, \infty]$, an $r \times m$ matrix A, and a vector $b \in \Re^r$:

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

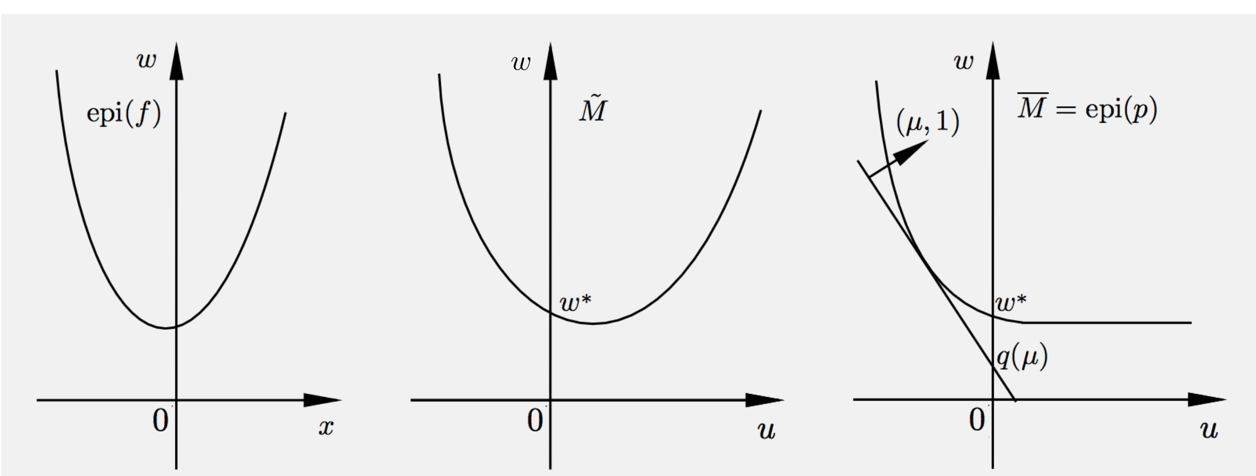
  (2) There is an $\overline{x} \in \text{ri}(\text{dom}(f))$ s. t. $A\overline{x} - b \leq 0$.

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

- We have $\overline{M} = \tilde{M} - \big\{(z, 0) \mid\ z \leq 0\big\}$, where

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

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



LECTURE 13

LECTURE OUTLINE

- Nonlinear Farkas Lemma
- Application to convex programming

We have now completed:

- The basic convexity theory, including hyperplane separation, and polyhedral convexity
- The basic theory of existence of optimal solutions, min common/max crossing duality, minimax theory, polyhedral/linear optimization
- There remain three major convex optimization topics in our course:
  - Convex/nonpolyhedral optimization
  - Conjugate convex functions (an algebraic form of min common/max crossing)
  - The theory of subgradients and associated convex optimization algorithms
- In this lecture, we overview the first topic (we will revisit it in more detail later)

MIN C/MAX C TH. III - A SPECIAL CASE

- Recall the linearly constrained optimization problem min common/max crossing framework:

  (1) The set $\overline{M}$ is defined in terms of a convex function $f : \Re^m \mapsto (-\infty, \infty]$, an $r \times m$ matrix A, and a vector $b \in \Re^r$:

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

  (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 $q(\mu) = q^*$.

- We have $\overline{M} = \text{epi}(p)$, where $p(u) = \inf_{Ax-b\leq u} f(x)$.
- We have $w^* = p(0) = \inf_{Ax-b\leq 0} f(x)$.
- The max crossing problem is to maximize over $\mu \in \Re^r$ the (dual) function q given by

$$\begin{aligned} q(\mu) &= \inf_{(u,w)\in \text{epi}(p)} \{w + \mu' u\} = \inf_{\{(u,w)\mid p(u)\leq w\}} \{w + \mu' u\} \\ &= \inf_{u\in\Re^m} \big\{p(u) + \mu' u\big\} = \inf_{u\in\Re^r} \inf_{Ax-b\leq u} \big\{f(x) + \mu' u\big\}, \end{aligned}$$

and finally

$$q(\mu) = \begin{cases} \inf_{x\in\Re^n} \big\{f(x) + \mu'(Ax - b)\big\} & \text{if } \mu \geq 0, \\ -\infty & \text{otherwise.} \end{cases}$$

NONLINEAR FARKAS' LEMMA

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

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

Let

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

Then:

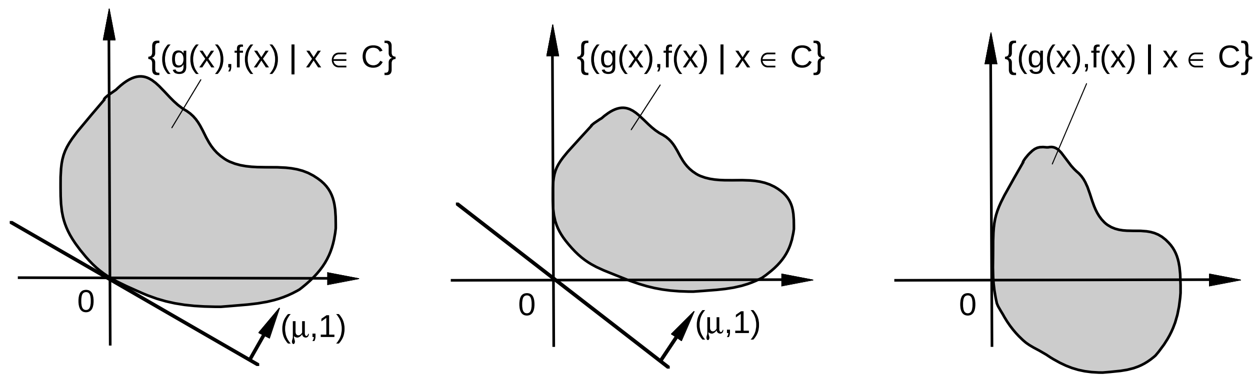
(a) Q^* is nonempty and compact if and only if there exists a vector $\overline{x} \in C$ such that $g_j(\overline{x}) < 0$ for all $j = 1, \ldots, r$.

(b) Q^* is nonempty if the functions g_j, $j = 1, \ldots, r$, are affine and there exists a vector $\overline{x} \in \mathrm{ri}(C)$ such that $g(\overline{x}) \leq 0$.

- Reduces to Farkas' Lemma if $C = \Re^n$, and f and g_j are linear.

- Part (b) follows from the preceding theorem.

VISUALIZATION OF NONLINEAR FARKAS' L.



(a) (b) (c)

- Assuming that for all $x \in C$ with $g(x) \leq 0$, we have $f(x) \geq 0$ (plus the other interior/rel. interior point condition).

- 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

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

in its positive halfspace.

- Figures (a) and (b) show examples where such a hyperplane exists, and figure (c) shows an example where it does not.

- In Fig. (a) there exists a point $\overline{x} \in C$ with $g(\overline{x}) < 0$.

PROOF OF NONLINEAR FARKAS' LEMMA

- Apply Min Common/Max Crossing to

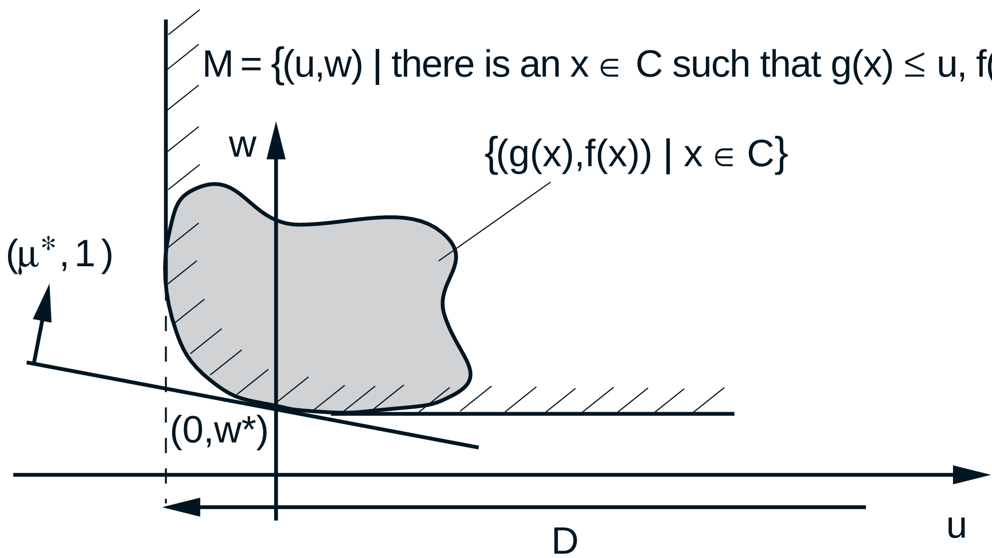
$$M = \big\{(u, w) \mid \text{there is } x \in C \text{ s. t. } g(x) \leq u,\ f(x) \leq w\big\}$$

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

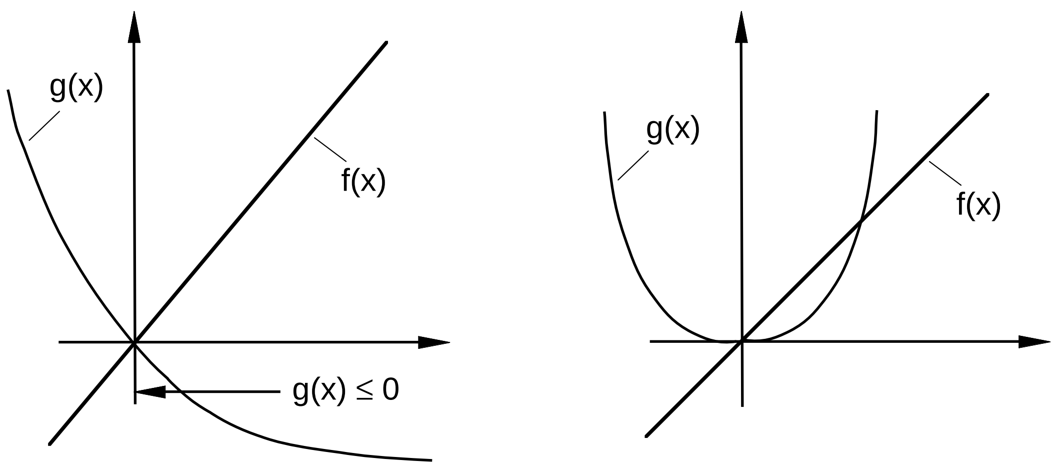
- Under condition (1), Min Common/Max Crossing Theorem II applies: we have

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

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



EXAMPLE



- Here $C = \Re$, $f(x) = x$. In the example on the left, g is given by $g(x) = e^{-x} - 1$, while in the example on the right, g is given by $g(x) = x^2$.

- In both examples, $f(x) \geq 0$ for all x such that $g(x) \leq 0$.

- On the left, condition (1) of the Nonlinear Farkas Lemma is satisfied, and for $\mu^* = 1$, we have

$$f(x) + \mu^* g(x) = x + e^{-x} - 1 \geq 0, \qquad \forall\, x \in \Re$$

- On the right, condition (1) is violated, and for every $\mu^* \geq 0$, the function $f(x) + \mu^* g(x) = x + \mu^* x^2$ takes negative values for x negative and sufficiently close to 0.

CONVEX PROGRAMMING

Consider the problem

$$\left.\begin{array}{ll}\text{minimize} & f(x)\\ \text{subject to} & x \in C,\ g_j(x) \le 0,\ j = 1, \ldots, r\end{array}\right\}$$

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

- Consider the Lagrangian function

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

and the minimax problem involving $L(x, \mu)$, over $x \in C$ and $\mu \ge 0$. Note $f^* = \inf_{x \in C} \sup_{\mu \ge 0} L(x, \mu)$.

- Consider the dual function

$$q(\mu) = \inf_{x \in C} L(x, \mu)$$

and the dual problem of maximizing $q(\mu)$ subject to $\mu \in \Re^r$.

- The dual optimal value, $q^* = \sup_{\mu \ge 0} q(\mu)$, satisfies $q^* \le f^*$ (this is just $\sup \inf L \le \inf \sup L$).

DUALITY THEOREM

- Assume that f and g_j are closed, and the function $t : C \mapsto (-\infty, \infty]$ given by

$$t(x) = \sup_{\mu \geq 0} L(x, \mu) = \begin{cases} f(x) & \text{if } g(x) \leq 0,\ x \in C, \\ \infty & \text{otherwise,} \end{cases}$$

has compact level sets. Then $f^* = q^*$ and the set of primal optimal solutions is nonempty and compact.

Proof: We have

$$\begin{aligned} f^* = \inf_{x \in C} t(x) &= \inf_{x \in C} \sup_{\mu \geq 0} L(x, \mu) \\ &= \sup_{\mu \geq 0} \inf_{x \in C} L(x, \mu) = \sup_{\mu \geq 0} q(\mu) = q^*, \end{aligned}$$

where inf and sup can be interchanged because a minimax theorem applies (t has compact level sets).

- The set of primal optimal solutions is the set of minima of t, and is nonempty and compact since t has compact level sets. **Q.E.D.**

EXISTENCE OF DUAL OPTIMAL SOLUTIONS

- Replace $f(x)$ by $f(x)-f^*$ and apply the Nonlinear Farkas' Lemma. Then, under the assumptions of the lemma, there exist $\mu_j^* \geq 0$, such that

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

- It follows that

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

Thus equality holds throughout, and we have

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

- Hence $f^* = q^*$ and μ^* is a dual optimal solution

- Note that we have use two different approaches to establish $q^* = f^*$:
  - Based on minimax theory (applies even if there is no dual optimal solution).
  - Based on the Nonlinear Farkas' Lemma (guarantees that there is a dual optimal solution).

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

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

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

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

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,

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

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

QUADRATIC PROGRAMMING 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 symmetric, and A, b, and c are given matrix/vectors.

- Dual function:

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

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

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

- 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 $t = b + AQ^{-1}c$.

- The dual has simpler constraints and perhaps smaller dimension.

LECTURE 14

LECTURE OUTLINE

- Convex conjugate functions
- Conjugacy theorem
- Examples
- Support functions

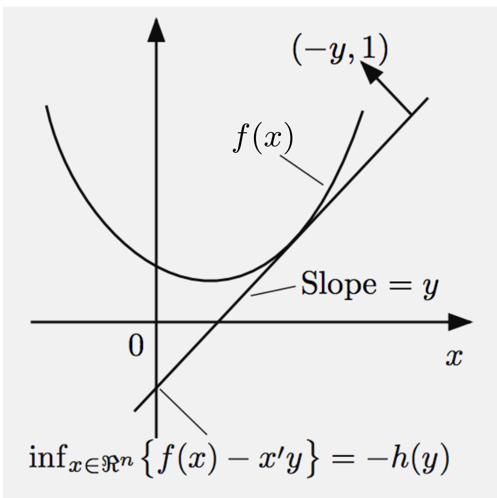
****

- Given f and its epigraph consider the function

Nonvertical hyperplanes supporting epi(f)

$\mapsto$ Crossing points of vertical axis

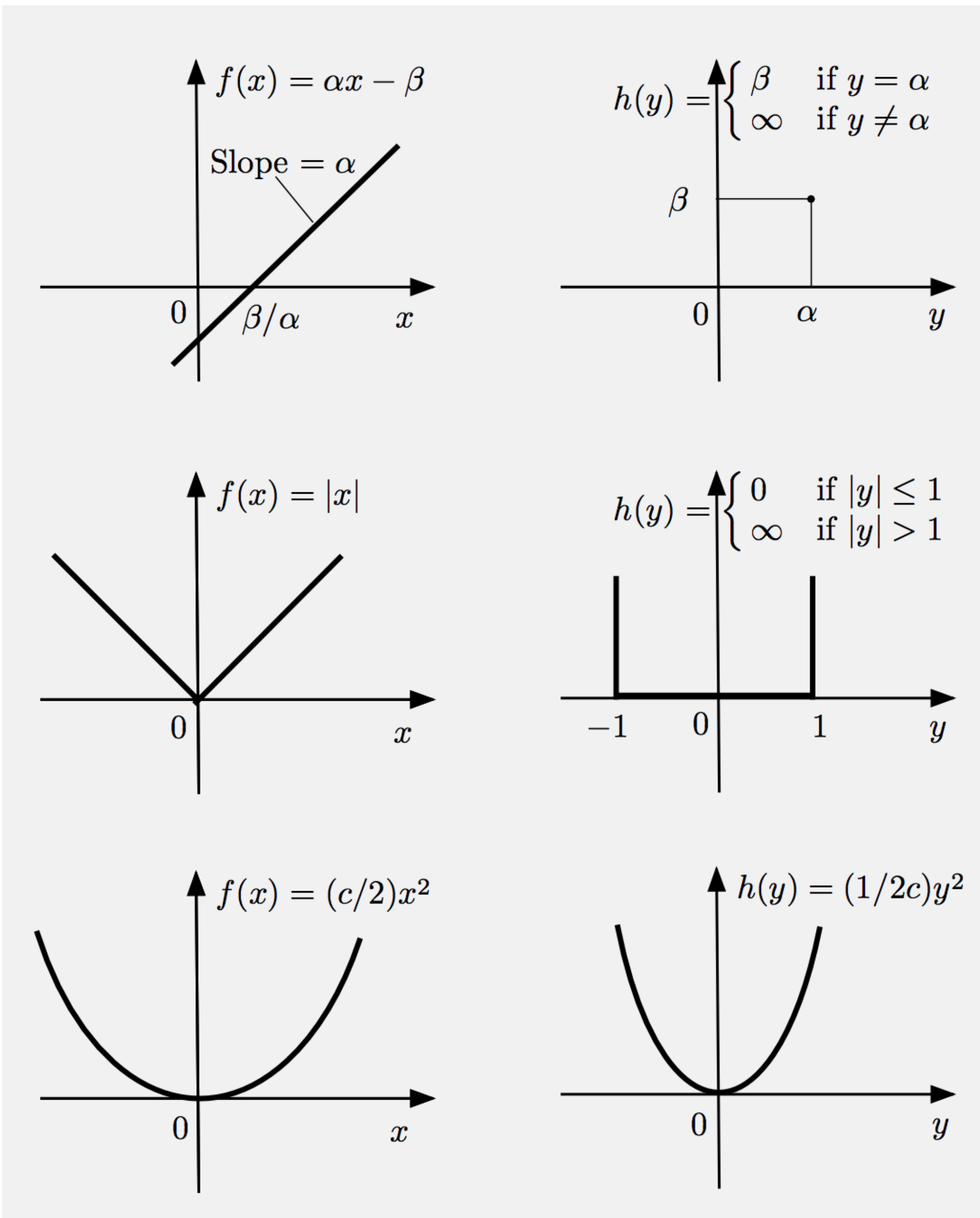
$$h(y) = \sup_{x \in \Re^n} \big\{x'y - f(x)\big\}, \qquad y \in \Re^n.$$



CONJUGATE FUNCTIONS

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

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



CONJUGATE OF CONJUGATE

- From the definition

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

note that h *is convex and closed*.

- Reason: $\text{epi}(h)$ is the intersection of the epigraphs of the convex and closed functions

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

as x ranges over $\Re^n$.

- Consider the conjugate of the conjugate:

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

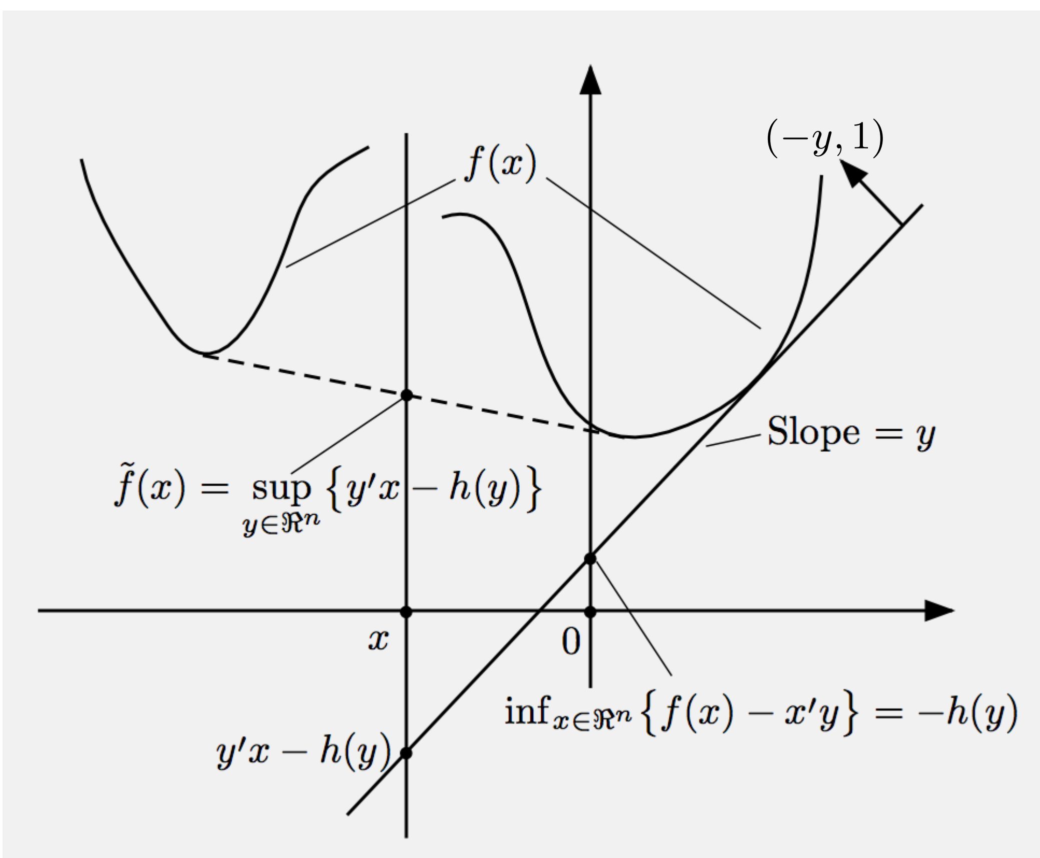
- $\tilde{f}$ is convex and closed.
- **Important fact/Conjugacy theorem:** If f is closed convex proper, then $\tilde{f} = f$.

CONJUGACY THEOREM - VISUALIZATION

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

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

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



EXTENSION TO NONCONVEX FUNCTIONS

- Let $f : \Re^n \mapsto [-\infty, \infty]$ be any function.

- Define $\hat{f} : \Re^n \mapsto [-\infty, \infty]$, the *convex closure of* f, as the function that has as epigraph the closure of the convex hull if $\text{epi}(f)$ [also the smallest closed and convex set containing $\text{epi}(f)$].

- The conjugate of the conjugate of f is $\hat{f}$, assuming $\hat{f}(x) > -\infty$ for all x.

- A counterexample (with closed convex but improper f) showing the need for the assumption:

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

We have

$$h(y) = \infty, \qquad \forall\ y \in \Re^n,$$

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

But the convex closure of f is $\hat{f} = f$ so $\hat{f} \neq \tilde{f}$.

CONJUGACY THEOREM

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

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

(a) We have

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

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

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

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

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

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

MIN COMMON/MAX CROSSING I

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be a function, and consider the min common/max crossing framework corresponding to

$$M = \overline{M} = \mathrm{epi}(f)$$

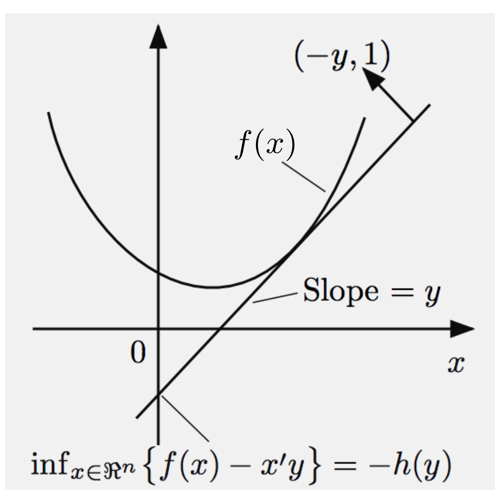
- From the figure it follows that the crossing value function is

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

and finally

$$q(\mu) = \inf_{u\in\Re^n} \big\{f(u) + \mu' u\big\} = -\sup_{u\in\Re^n} \big\{(-\mu)' u - f(u)\big\}.$$

- Thus $q(\mu) = -h(-\mu)$ where h: conjugate of f



MIN COMMON/MAX CROSSING I

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be a function, and consider the min common/max crossing framework corresponding to

$$M = \overline{M} = \mathrm{epi}(f)$$

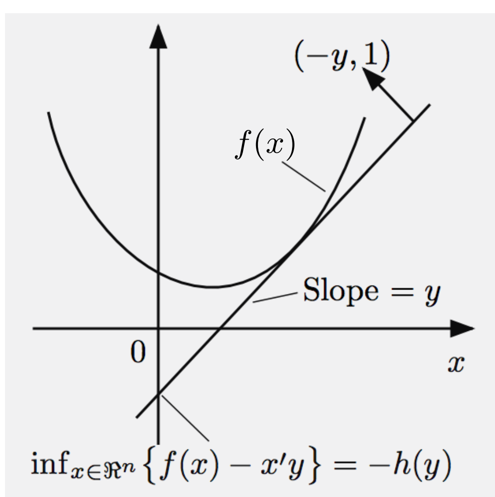
- From the figure it follows that the crossing value function is

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

and finally

$$q(\mu) = \inf_{u\in\Re^n} \big\{f(u)+\mu' u\big\} = -\sup_{u\in\Re^n} \big\{(-\mu)' u - f(u)\big\}.$$

- Thus $q(\mu) = -h(-\mu)$ where h: conjugate of f



MIN COMMON/MAX CROSSING II

- For $M = \text{epi}(f)$, we have

$$q^* = \tilde{f}(0) \leq f(0) = w^*,$$

where $\tilde{f}$ is the double conjugate of f.

- To see this, note that $w^* = f(0)$, and that by using the relation $h(y) = -q(-y)$ just shown, we have

$$\begin{aligned} \tilde{f}(0) &= \sup_{y \in \Re^n} \{-h(y)\} \\ &= \sup_{y \in \Re^n} q(-y) \\ &= \sup_{\mu \in \Re^n} q(\mu) \\ &= q^* \end{aligned}$$

- **Conclusion:** There is no duality gap ($q^* = w^*$) if and only if $f(0) = \tilde{f}(0)$, which is true if f is closed proper convex (Conjugacy Theorem).

- **Note:** Convexity of f plus $f(0) = \tilde{f}(0)$ is the essential assumption of Min Common/Max Crossing Theorem I.

CONJUGACY AND MINIMAX

- Consider the minimax problem involving $\phi : X \times Z \mapsto \Re$ with $x \in X$ and $z \in Z$.
- The min common/max crossing framework involves $M = \text{epi}(p)$, where

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

- We have in general

$$\begin{aligned} &\sup_{z \in Z} \inf_{x \in X} \phi(x, z) \le q^* \\ &\qquad = \tilde{p}(0) \le p(0) = w^* = \inf_{x \in X} \sup_{z \in Z} \phi(x, z), \end{aligned}$$

where $\tilde{p}$ is the double conjugate of p.

- The rightmost inequality holds as an equation if p is closed proper convex.
- The leftmost inequality holds as an equation if ϕ is concave and u.s.c. in z. It turns out that

$$\tilde{p}(0) = \sup_{z \in Z} \inf_{x \in X} \big\{ -\tilde{r}_x(z) \big\}$$

where $\tilde{r}_x$ is the double conjugate of $-\phi(x, \cdot)$.

CONJUGACY AND MINIMAX

- Consider the minimax problem involving $\phi : X \times Z \mapsto \Re$ with $x \in X$ and $z \in Z$.

- The min common/max crossing framework involves $M = \text{epi}(p)$, where

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

- We have in general

$$\sup_{z \in Z} \inf_{x \in X} \phi(x, z) \leq q^*$$
$$= \tilde{p}(0) \leq p(0) = w^* = \inf_{x \in X} \sup_{z \in Z} \phi(x, z),$$

where $\tilde{p}$ is the double conjugate of p.

- The rightmost inequality holds as an equation if p is closed proper convex.

- The leftmost inequality holds as an equation if ϕ is concave and u.s.c. in z. It turns out that

$$\tilde{p}(0) = \sup_{z \in \Re^m} \inf_{x \in X} \big\{ -\tilde{r}_x(z) \big\}$$

where $\tilde{r}_x$ is the double conjugate of $-\phi(x, \cdot)$.

A FEW EXAMPLES

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

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

- Conjugate of a strictly convex quadratic

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

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

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

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

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

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

SUPPORT FUNCTIONS

- Conjugate of indicator function δ_X of set X

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

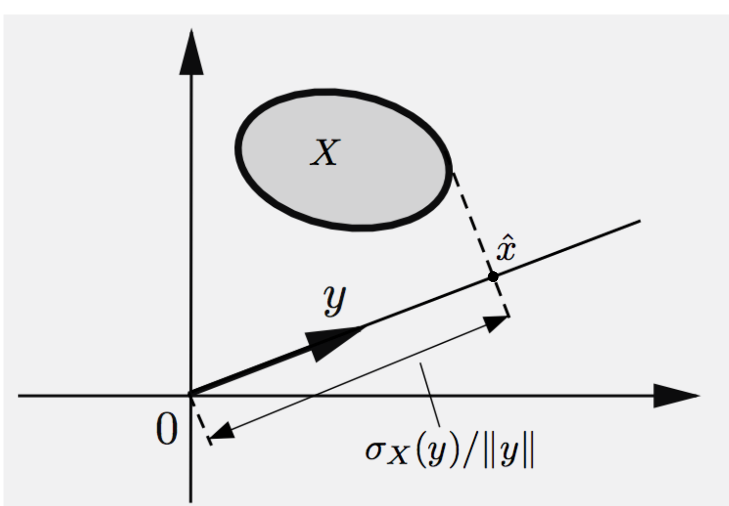
is called the *support function of* X.

- $\mathrm{epi}(\sigma_X)$ is a closed convex cone.

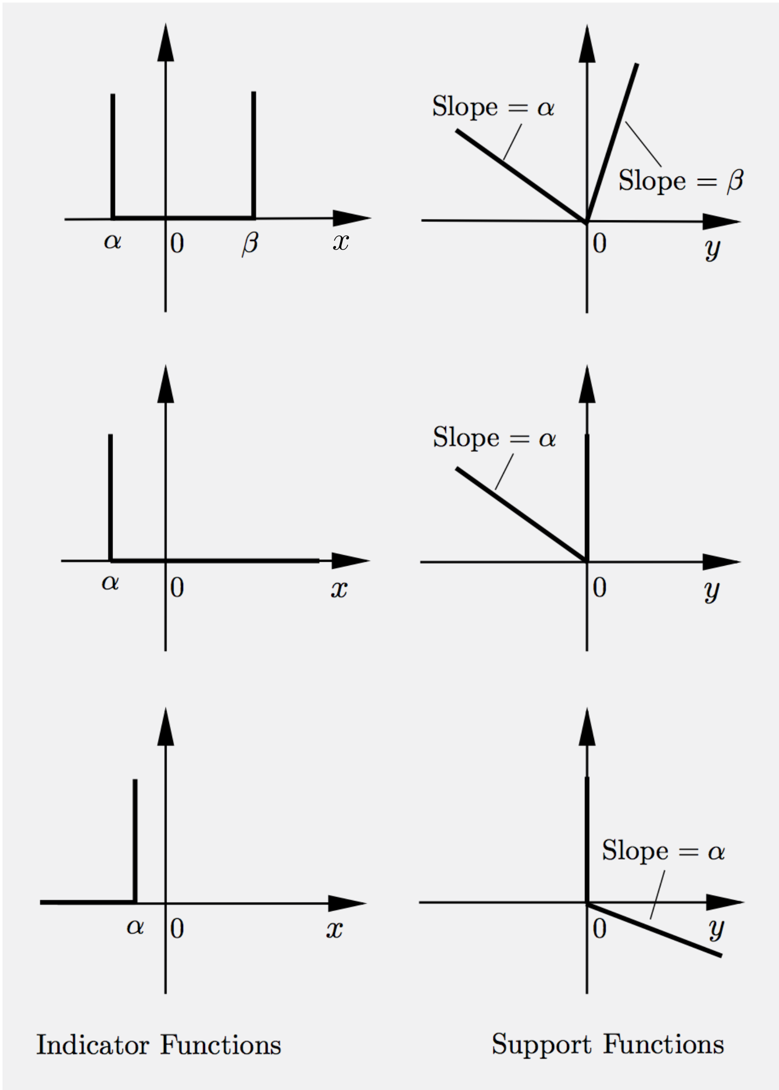
- The sets X, $\mathrm{cl}(X)$, $\mathrm{conv}(X)$, and $\mathrm{cl}\big(\mathrm{conv}(X)\big)$ all have the same support function (by the conjugacy theorem).

- 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

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



EXAMPLES OF SUPPORT FUNCTIONS I



- The support function of the union $X = \cup_{j=1}^r X_j$:

$$\sigma_X(y) = \sup_{x \in X} y'x = \max_{j=1,\ldots,r} \sup_{x \in X_i} y'x = \max_{j=1,\ldots,r} \sigma_{X_j}(y).$$

- The support function of the convex hull of $X = \cup_{j=1}^r X_j$ is the same.

EXAMPLES OF SUPPORT FUNCTIONS II

- The support function of a bounded ellipsoid $X = \{x \mid (x - \overline{x})'Q(x - \overline{x}) \leq b\}$:

$$\sigma_X(y) = y'\overline{x} + \left(b\, y'Q^{-1}y\right)^{1/2}, \qquad \forall\, y \in \Re^n$$

- The support function of a cone C: If $y'x \leq 0$ for all $x \in C$, i.e., $y \in C^*$, we have $\sigma_C(y) = 0$, since 0 is a closure point of C. On the other hand, if $y'x > 0$ for some $x \in C$, we have $\sigma_C(y) = \infty$, since C is a cone and therefore contains αx for all $\alpha > 0$. Thus,

$$\sigma_C(y) = \begin{cases} 0 & \text{if } y \in C^*, \\ \infty & \text{if } y \notin C^*, \end{cases}$$

i.e., the support function of C is equal to the indicator function of C^* ($\Rightarrow$ Polar Cone Theorem).

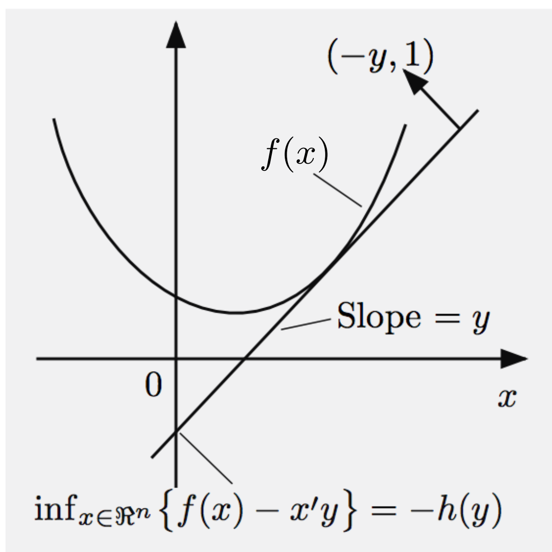
LECTURE 15

LECTURE OUTLINE

- Properties of convex conjugates and support functions

****

- Conjugate of f: $h(y) = \sup_{x\in\Re^n}\big\{x'y - f(x)\big\}$



- **Conjugacy Theorem:** The conjugate of the conjugate of a proper convex function f is the closure of f.

- Support function of set X = Conjugate of its indicator function

SUPPORT FUNCTIONS/POLYHEDRAL SETS I

- Consider the Minkowski-Weyl representation of a polyhedral set

$$X = \text{conv}\big(\{v_1, \ldots, v_m\}\big) + \text{cone}\big(\{d_1, \ldots, d_r\}\big)$$

- The support function is

$$\begin{aligned}
\sigma_X(y) &= \sup_{x \in X} y'x \\
&= \sup_{\substack{\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_r \geq 0 \\ \sum_{i=1}^m \alpha_i = 1}} \left\{ \sum_{i=1}^m \alpha_i v_i' y + \sum_{j=1}^r \beta_j d_j' y \right\} \\
&= \begin{cases} \max_{i=1,\ldots,m} v_i' y & \text{if } d_j' y \leq 0,\ j = 1, \ldots, r, \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}$$

- Hence, the support function of a polyhedral set is a polyhedral function.

SUPPORT FUNCTIONS/POLYHEDRAL SETS II

- Consider f, h, and $\mathrm{epi}(f)$. We have

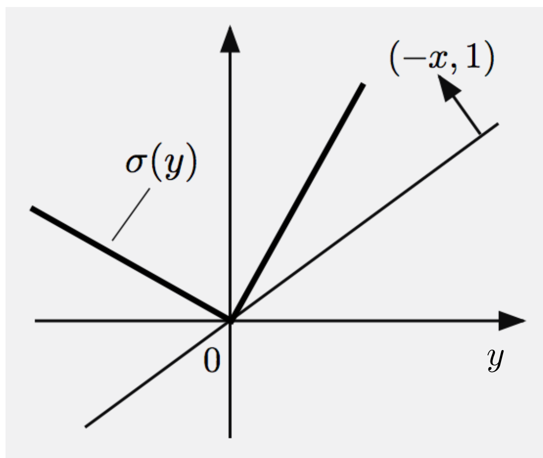
$$\begin{aligned} h(y) &= \sup_{x \in \Re^n} \big\{ x'y - f(x) \big\} \\ &= \sup_{(x,w) \in \mathrm{epi}(f)} \{ x'y - w \} \\ &= \sigma_{\mathrm{epi}(f)}(y, -1) \end{aligned}$$

- If f is polyhedral, $\mathrm{epi}(f)$ is a polyhedral set, so $\sigma_{\mathrm{epi}(f)}$ is a polyhedral function, so h is a polyhedral function.
- Conclusion: Conjugates of polyhedral functions are polyhedral.

POSITIVELY HOMOGENEOUS FUNCTIONS

- A function $f : \Re^n \mapsto [-\infty, \infty]$ is *positively homogeneous* if its epigraph is a cone, i.e.,

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



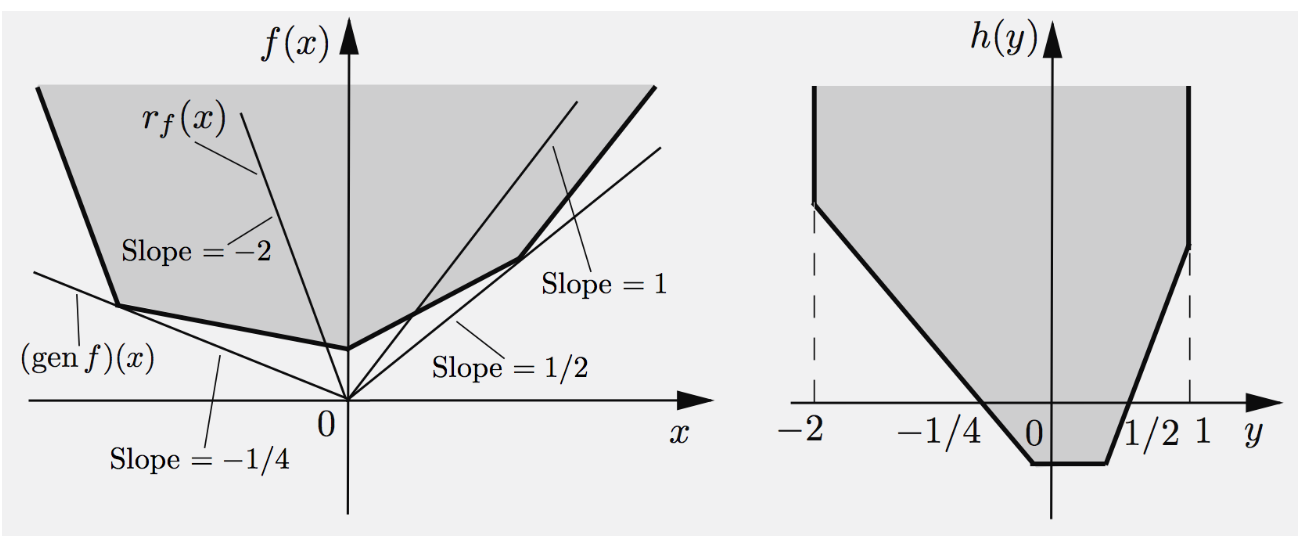
- A support function is closed, proper, convex, and positively homogeneous.

- *Converse Result:* The closure of a proper, convex, and positively homogeneous function σ is the support function of the closed convex set

$$X = \big\{x \mid y'x \le \sigma(y),\ \forall\, y \in \Re^n\big\}$$

CONES RELATING TO SETS AND FUNCTIONS

- Cones associated with a convex set C:
  - Polar cone, recession cone, generated cone, epigraph of support function
- Cones associated with a convex function f are the cones associated with its epigraph, which among others, give rise to:
  - The recession function of f and the closed function generated by f [function whose epigraph is the closure of the cone generated by epi(f)]



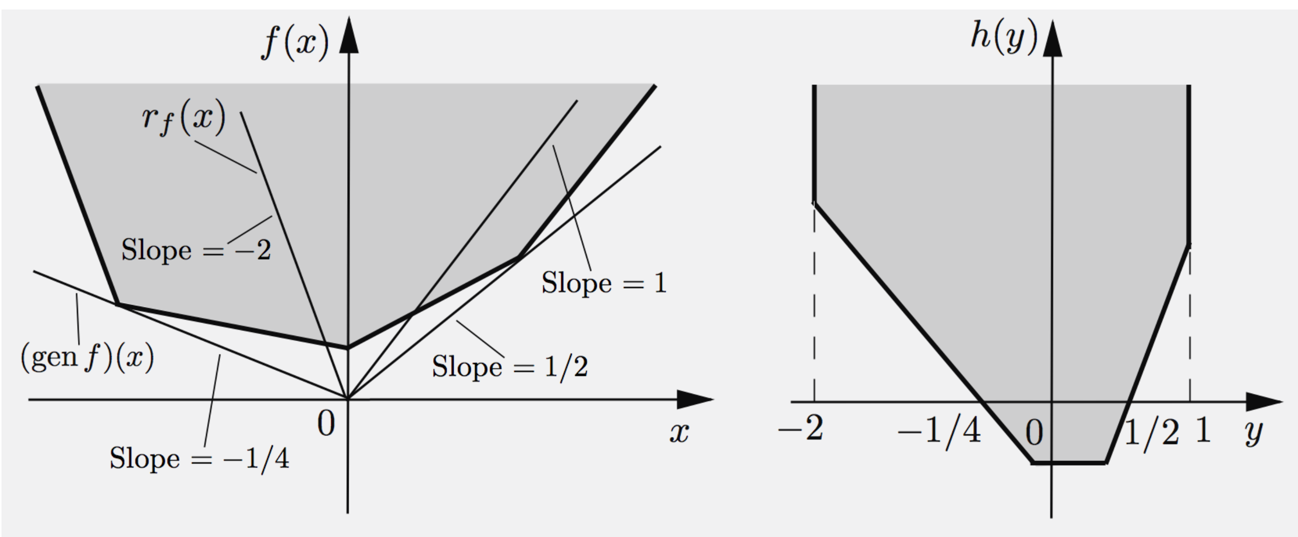
- The cones of a function f are epigraphs of support functions of sets associated with f.

FORMULAS FOR DOMAIN, LEVEL SETS, ETC I

- **Support Function of Domain:** Let $f : \Re^n \mapsto (-\infty, \infty]$ be a proper convex function, and let h be its conjugate.

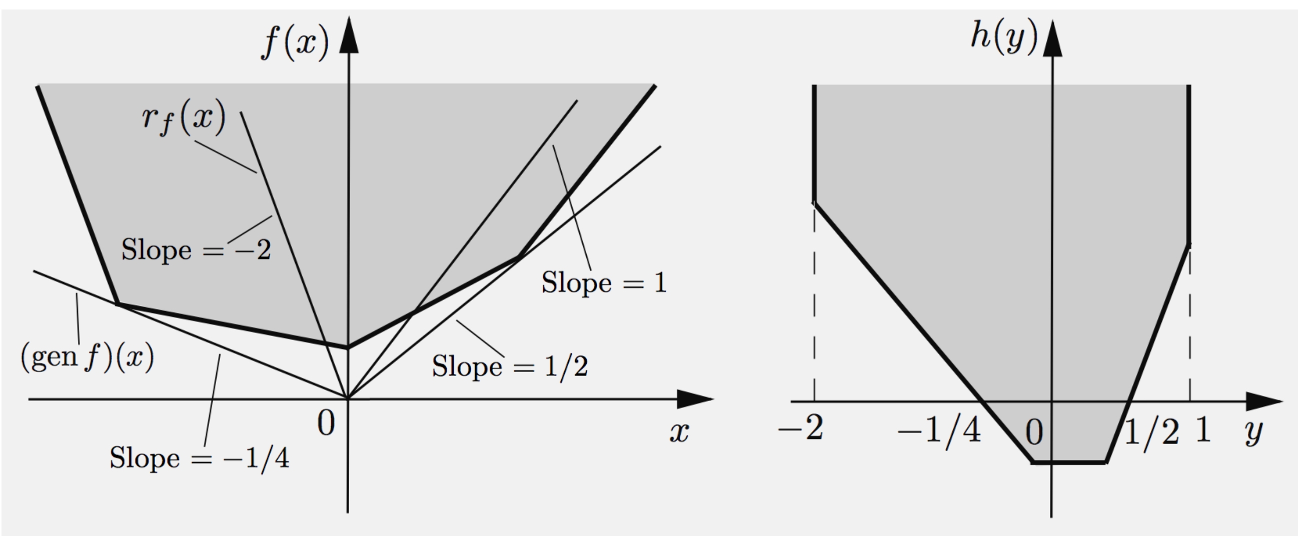
  (a) The support function of $\mathrm{dom}(f)$ is the recession function of h.

  (b) If f is closed, the support function of $\mathrm{dom}(h)$ is the recession function of f.



FORMULAS FOR DOMAIN, LEVEL SETS, ETC II

- **Support Function of 0-Level Set:** Let $f : \Re^n \mapsto (-\infty, \infty]$ be a closed proper convex function and let h be its conjugate.
  (a) If the level set $\{y \mid h(y) \leq 0\}$ is nonempty, its support function is the closed function generated by f.
  (b) If the level set $\{x \mid f(x) \leq 0\}$ is nonempty, its support function is the closed function generated by h.



- This can be used to characterize any nonempty level set of a closed convex function: add a constant to the function and convert the level set to a 0-level set.

RECESSION CONE/DOMAIN OF SUPPORT FN

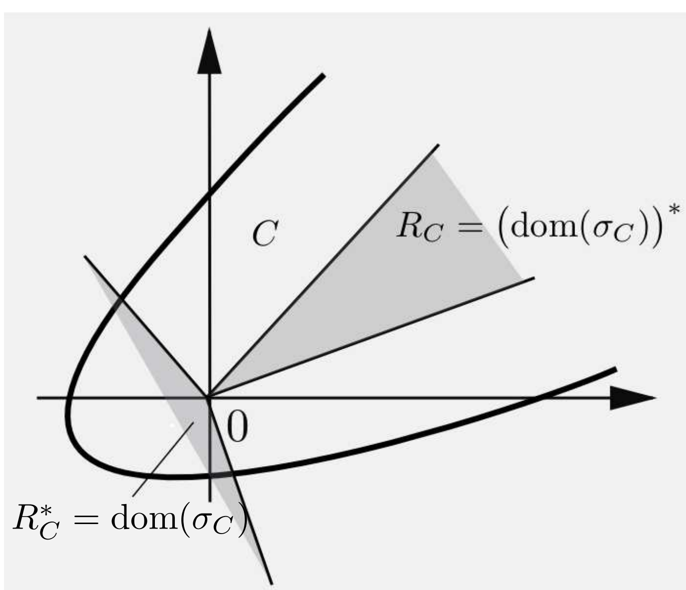
- Let C be a nonempty convex set in $\Re^n$.

(a) The polar cone of C is the 0-level set of σ_C:

$$C^* = \{y \mid \sigma_C(y) \leq 0\}.$$

(b) If C is closed, the recession cone of C is equal to the polar cone of the domain of σ_C:

$$R_C = \big(\mathrm{dom}(\sigma_C)\big)^*.$$



CALCULUS OF CONJUGATE FUNCTIONS

- **Example: (Linear Composition)** Consider $F(x) = f(Ax)$, where f is closed proper convex, and A is a matrix.

- If h is the conjugate of f, we have

$$
\begin{aligned}
f(Ax) &= \sup_y \big\{x'A'y - h(y)\big\} \\
&= \sup_{\{(y,z) \mid A'y = z\}} \big\{x'z - h(y)\big\} \\
&= \sup_z \left\{x'z - \inf_{A'y=z} h(y)\right\}
\end{aligned}
$$

so F is the conjugate of H given by

$$
H(z) = \inf_{A'y=z} h(y)
$$

called the *image function* of h under A'.

- Hence the conjugate of F is the *closure* of H, provided F is proper [true iff $R(A) \cap \mathrm{dom}(f) \neq \emptyset$].

- Issues of preservation of closedness under partial minimization [$N(A') \cap R_h \subset L_h \ \Rightarrow\ H$ is closed].

CONJUGATE OF A SUM OF FUNCTIONS

• Let $f_i : \Re^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, be closed proper convex functions, and let h_i be their conjugates. Let $F(x) = f_1(x) + \cdots + f_m(x)$. We have

$$
\begin{aligned}
F(x) &= \sum_{i=1}^{m} \sup_{y_i} \{x' y_i - h_i(y_i)\} \\
&= \sup_{y_1, \ldots, y_m} \left\{ x' \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} h_i(y_i) \right\} \\
&= \sup_{\{(y_1, \ldots, y_m, z) \mid \sum_{i=1}^{m} y_i = z\}} \left\{ x' z - \sum_{i=1}^{m} h_i(y_i) \right\} \\
&= \sup_{z} \left\{ x' z - \inf_{\sum_{i=1}^{m} y_i = z} \sum_{i=1}^{m} h_i(y_i) \right\}
\end{aligned}
$$

so F is the conjugate of H given by

$$
H(z) = \inf_{\sum_{i=1}^{m} y_i = z} \sum_{i=1}^{m} h_i(y_i)
$$

called the *infimal convolution* of $h_1, \ldots, h_m$.

• Hence the conjugate of F is the *closure* of H, provided F is proper [true iff $\cap_{i=1}^{m} \mathrm{dom}(f_i) \neq \emptyset$].

CLOSEDNESS OF IMAGE FUNCTION

- We view the image function

$$H(y) = \inf_{A'z=y} h(z)$$

as the result of partial minimization with respect to z of a function of (z, y).

- We use the results on preservation of closedness under partial minimization
  - The image function is closed and the infimum is attained for all $y \in \mathrm{dom}(H)$ if h is closed and every direction of recession of h that belongs to $N(A')$ is a direction along which h is constant.

- This condition can be translated to an alternative and more useful condition involving the relative interior of the domain of the conjugate of h. In particular, we can show that the condition is true if and only if

$$R(A) \cap \mathrm{ri}\big(\mathrm{dom}(f)\big) \neq \emptyset$$

- Similar analysis for infimal convolution.

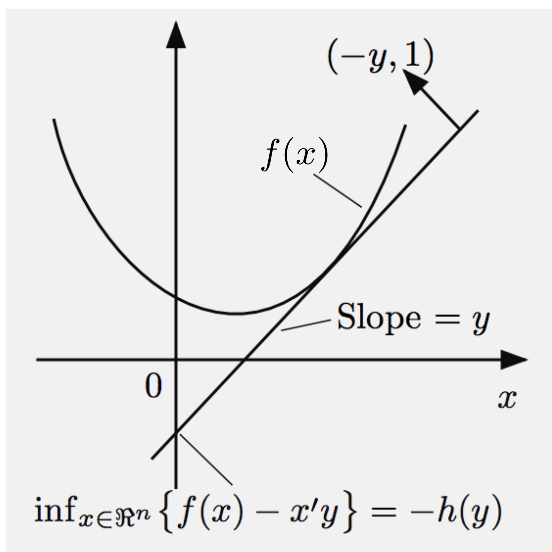
LECTURE 16

LECTURE OUTLINE

- Subgradients
- Calculus of subgradients

****

- Conjugate of f: $h(y) = \sup_{x\in\Re^n}\big\{x'y - f(x)\big\}$



- Conjugacy Theorem: If f is closed proper convex, it is equal to its double conjugate $\tilde{f}$.

SUBGRADIENTS

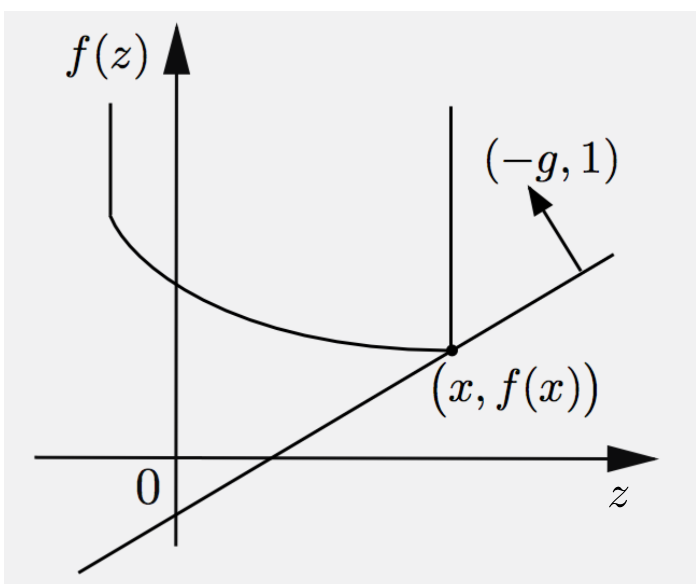
- Let $f : \Re^n \mapsto (-\infty, \infty]$ be a convex function. A vector $g \in \Re^n$ is a *subgradient* of f at a point $x \in \text{dom}(f)$ if

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

- 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 $\partial f(x)$.

EXAMPLES OF SUBDIFFERENTIALS

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

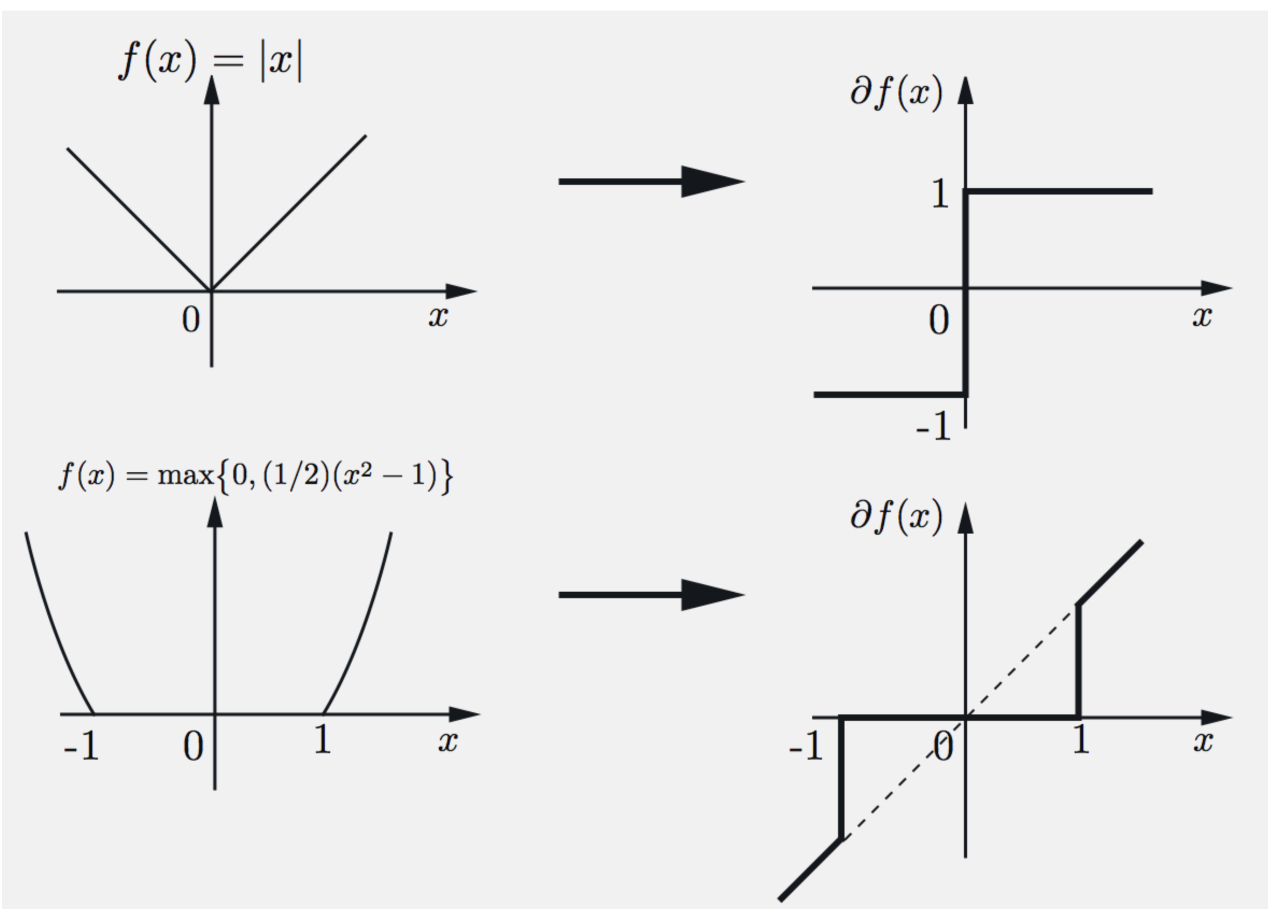
Proof: If $g \in \partial f(x)$, then

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

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

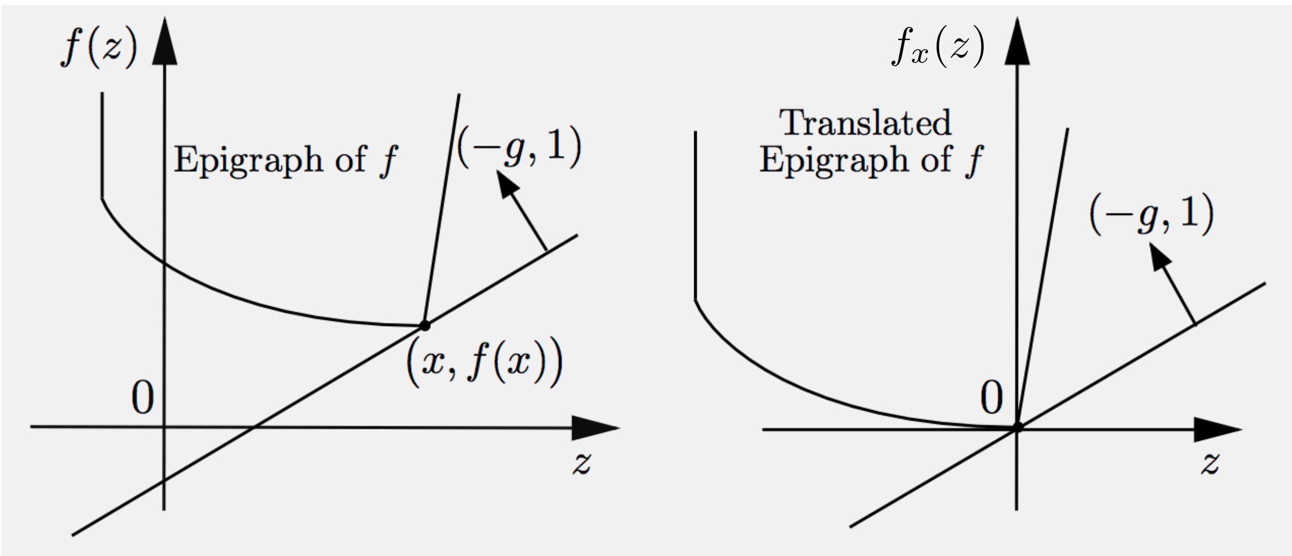
$$\gamma \|\nabla f(x) - g\|^2 \geq o(\gamma), \qquad \forall\, \gamma \in \Re$$

- Some examples:



EXISTENCE OF SUBGRADIENTS

• Note the connection with min common/max crossing [$M = \text{epi}(f_x)$, $f_x(z) = f(x+z) - f(x)$].



• Let $f : \Re^n \mapsto (-\infty, \infty]$ be a proper convex function. 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 $\text{dom}(f)$

– G is a nonempty and compact set.

• Furthermore, $\partial f(x)$ is nonempty and compact if and only if x is in the interior of $\text{dom}(f)$.

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:** For the case of a polyhedral set

$$P = \{x \mid a_i'x \leq b_i,\, i = 1, \ldots, m\},$$

we have

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

FENCHEL INEQUALITY

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

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

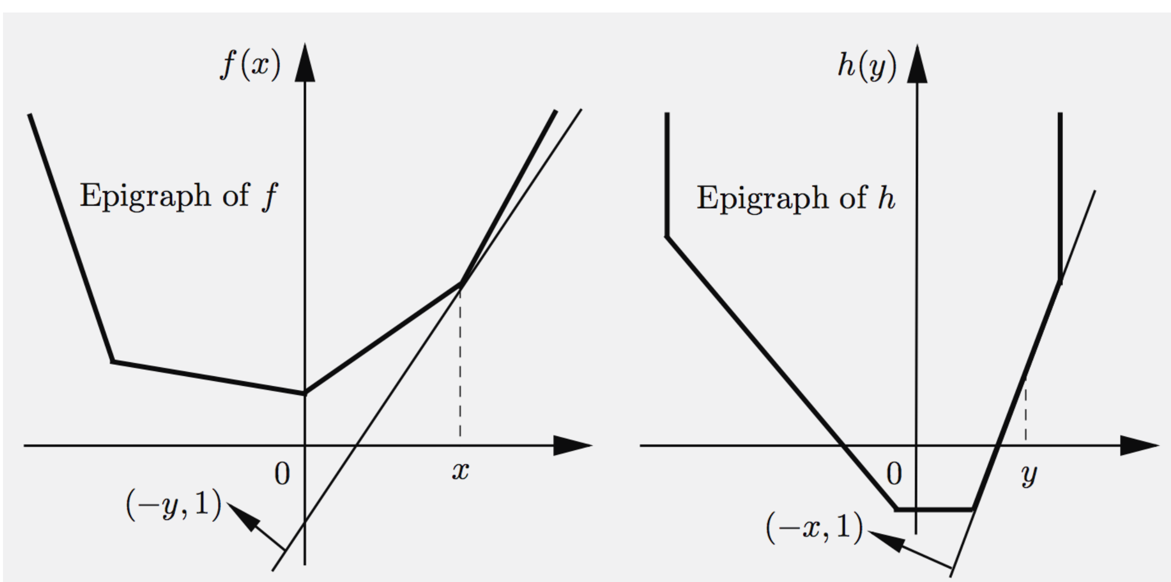
- **Proposition:** The following two relations are equivalent for a pair of vectors (x, y):

  (i) $x'y = f(x) + h(y)$.

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

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

(iii) $x \in \partial h(y)$.



MINIMA OF CONVEX FUNCTIONS

- **Application:** Let f be closed convex and let X^* be the set of minima of f over $\Re^n$. Then:
  (a) $X^* = \partial h(0)$.
  (b) X^* is nonempty if $0 \in \text{ri}\big(\text{dom}(h)\big)$.
  (c) X^* is nonempty and compact if and only if $0 \in \text{int}\big(\text{dom}(h)\big)$.
- **Proof:** (a) From the subgradient inequality,

$$x^* \text{ minimizes } f \quad \text{iff} \quad 0 \in \partial f(x^*),$$

which is true if and only if

$$x^* \in \partial h(0),$$

so $X^* = \partial h(0)$.

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

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

EXAMPLE: SUBDIFF. OF SUPPORT FUNCTION

- Consider the support function σ_C of a nonempty set C at a vector $\overline{y}$.
- To calculate $\partial\sigma_C(\overline{y})$, we introduce the function

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

- We have $\partial\sigma_C(\overline{y}) = \partial r(0)$, so $\partial\sigma_C(\overline{y})$ is equal to the set of minima over $\Re^n$ of the conjugate of r.
- The conjugate of r is $\sup_{y\in\Re^n}\{y'x - r(y)\}$, or

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

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

- Hence $\partial\sigma_C(\overline{y})$ is equal to the set of minima of $\delta(x) - \overline{y}'x$, or equivalently the set of maxima of $\overline{y}'x$ over $x \in \mathrm{cl}\big(\mathrm{conv}(C)\big)$.

EXAMPLE: SUBDIFF. OF POLYHEDRAL FN

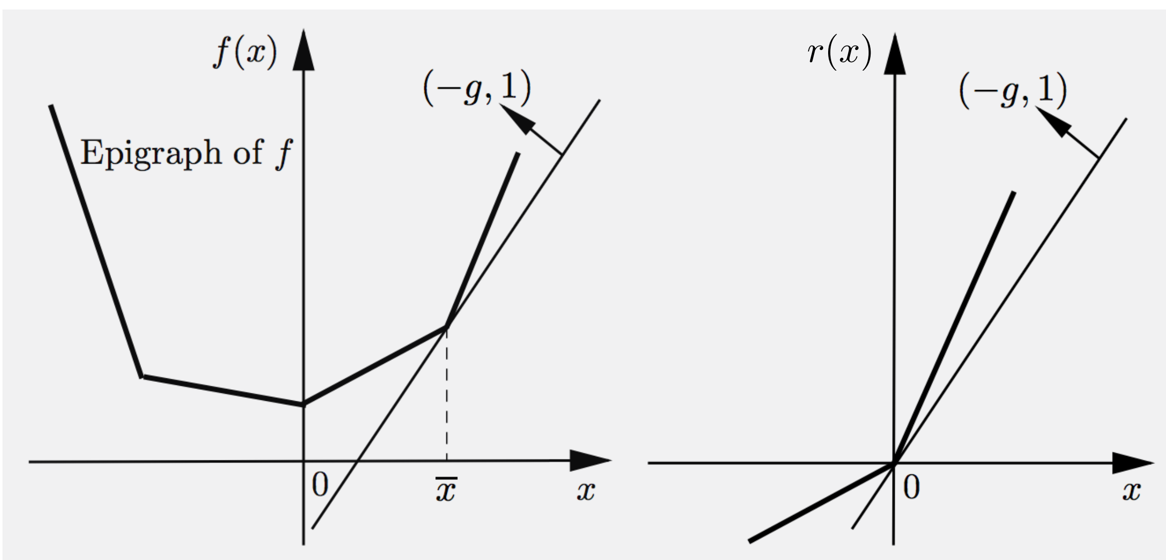
- Let

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

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

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

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



- It is easily shown 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

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

LECTURE 17

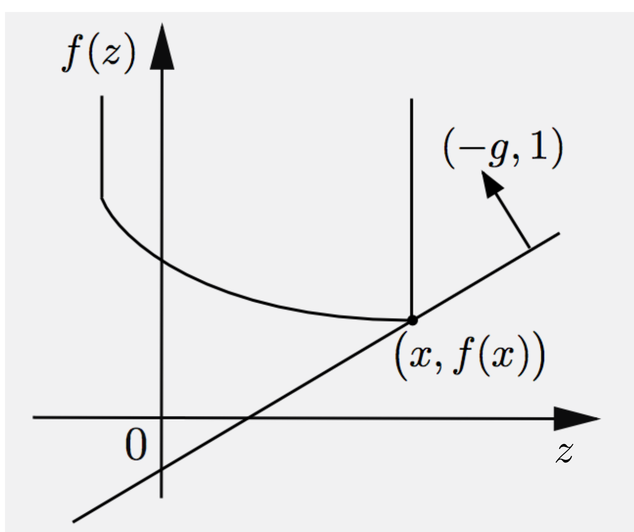
LECTURE OUTLINE

- Subdifferential of sum, chain rule
- Optimality conditions
- Directional derivatives
- Algorithms: Subgradient methods

****

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

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



- Recall: $y \in \partial f(x)$ iff $f(x) + h(y) = x'y$ (from Fenchel inequality)

CHAIN RULE

- Let $f : \Re^m \mapsto (-\infty, \infty]$ be proper convex, and A be a matrix. Consider $F(x) = f(Ax)$.
- **Claim:** If $R(A) \cap \mathrm{ri}\big(\mathrm{dom}(f)\big) \neq \emptyset$, then

$$\partial F(x) = A'\partial f(Ax).$$

- This condition guarantees that the conjugate of F is the image function

$$H(y) = \inf_{A'z=y} h(y)$$

where h is the conjugate of f, and the infimum is attained for all $y \in \mathrm{dom}(H)$.

Proof: We have $y \in \partial F(x)$ iff $F(x)+H(y) = x'y$, or iff there exists a vector z such that $A'z = y$ and $F(x) + h(y) = x'A'y$, or

$$f(Ax) + h(y) = x'A'y.$$

Therefore, $y \in \partial F(x)$ iff for some z such that $A'z = y$, we have $z \in \partial f(Ax)$. **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

$$\partial f(x) = \partial f_1(x) + \cdots + \partial f_m(x), \qquad \forall\, x \in \Re^n.$$

- **Extension:** If for some k, the functions f_i, $i = 1, \ldots, k$, are polyhedral, it is sufficient to assume

$$\Big(\cap_{i=1}^{k} \mathrm{dom}(f_i)\Big) \cap \Big(\cap_{i=k+1}^{m} \mathrm{ri}\big(\mathrm{dom}(f_i)\big)\Big) \neq \emptyset.$$

- Showing $\partial f(x) \supset \partial f_1(x) + \cdots + \partial f_m(x)$ is easy. For the reverse, we can use infimal convolution theory (as in the case of the chain rule).

EXAMPLE: SUBDIFF. OF POLYHEDRAL FN

- Let

$$f(x) = p(x) + \delta_P(x),$$

where P is a polyhedral set, δ_P is its indicator function, and p is the real-valued polyhedral function

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

with $a_1, \ldots, a_r \in \Re^n$ and $b_1, \ldots, b_r \in \Re$.

- We have

$$\partial f(x) = \partial p(x) + N_P(x),$$

so for $x \in P$, $\partial f(x)$ is a polyhedral set and the above is its Minkowski-Weyl representation.

- $\partial p(x)$ is the convex hull of the "active" a_j.
- $N_P(x)$ is the normal cone of P at x, (cone generated by normals to "active" halfspaces).

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

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

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

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

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

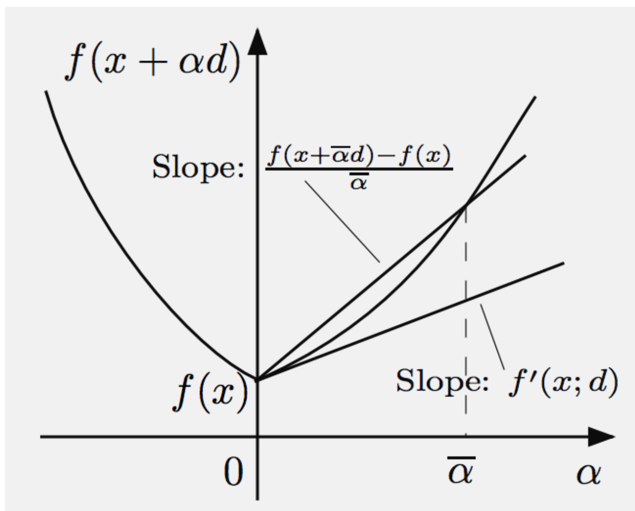
$$F(x) = f(x) + \delta_X(x)$$

if and only if $0 \in \partial F(x^*)$. Use the formula for subdifferential of sum. **Q.E.D.**

DIRECTIONAL DERIVATIVES

- Directional derivative of a proper convex f:

$$f'(x;d) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha d) - f(x)}{\alpha}, \;\; x \in \mathrm{dom}(f),\; d \in \Re^n$$



- 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 $\partial f(x)$.

ALGORITHMS: SUBGRADIENT METHOD

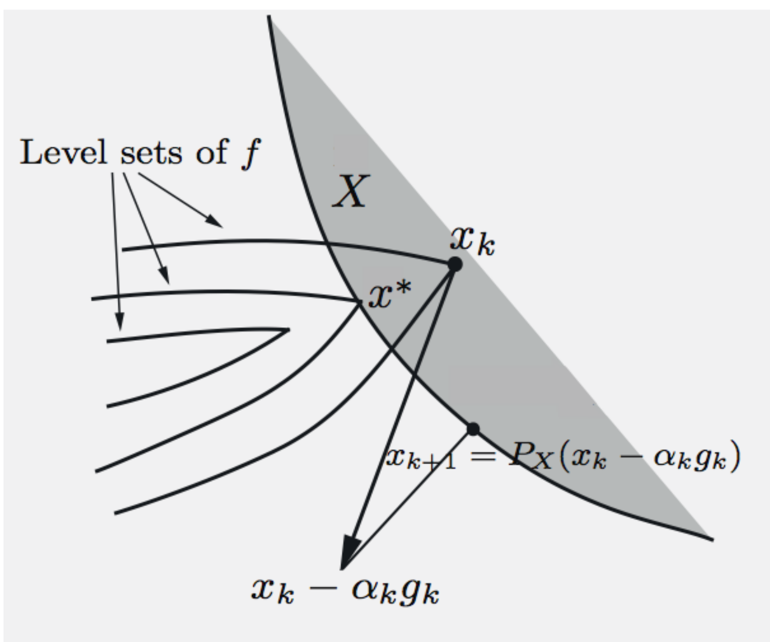
- **Problem:** Minimize convex function $f : \Re^n \mapsto \Re$ over a closed convex set X.

- Iterative descent idea has difficulties in the absence of differentiability of f.

- **Subgradient method**:

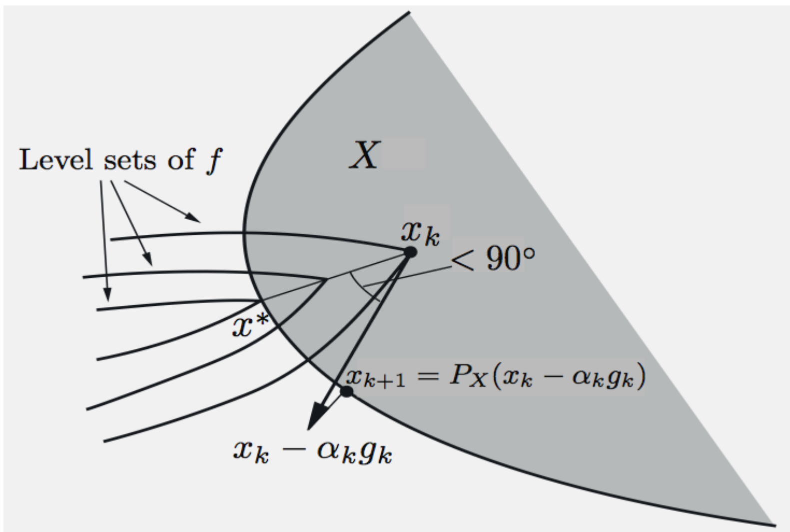
$$x_{k+1} = P_X(x_k - \alpha_k g_k),$$

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:

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

and if $f(y) < f(x_k)$,

$$\|x_{k+1} - y\| < \|x_k - y\|,$$

for all α_k such that

$$0 < \alpha_k < \frac{2\big(f(x_k) - f(y)\big)}{\|g_k\|^2}.$$

CONVERGENCE MECHANISM

- Assume constant stepsize: $\alpha_k \equiv \alpha$
- If $\|g_k\| \leq c$ for some constant c and all k,

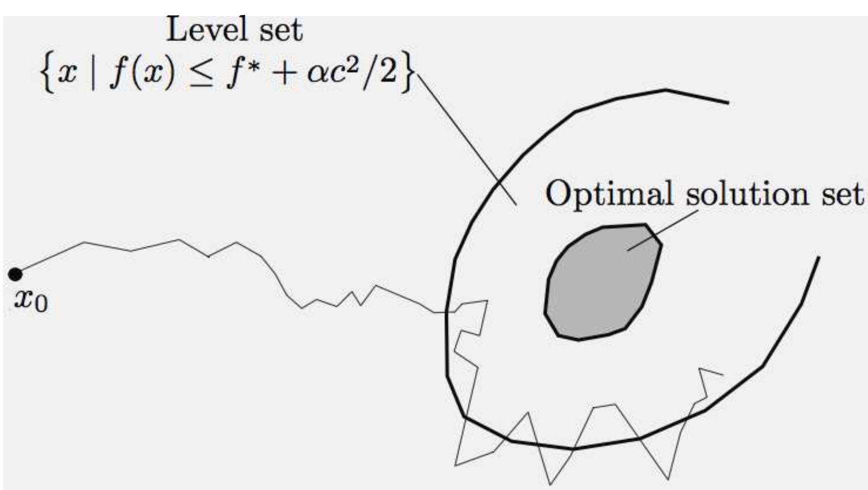
$$\|x_{k+1}-x^*\|^2 \leq \|x_k-x^*\|^2-2\alpha\big(f(x_k)-f(x^*)\big)+\alpha^2c^2$$

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 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 18

LECTURE OUTLINE

- Cutting plane methods
- Proximal minimization algorithm
- Proximal cutting plane algorithm
- Bundle methods

****

- 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 cutting plane method is to build an ever more accurate approximation of f using such planes.

CUTTING PLANE METHOD

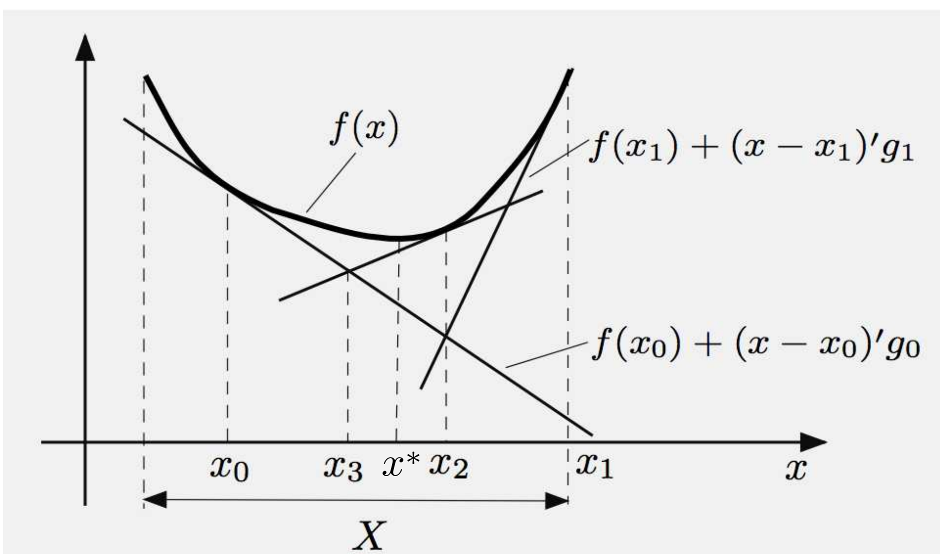
- Start with any $x_0 \in X$. For $k \geq 0$, set

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

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.

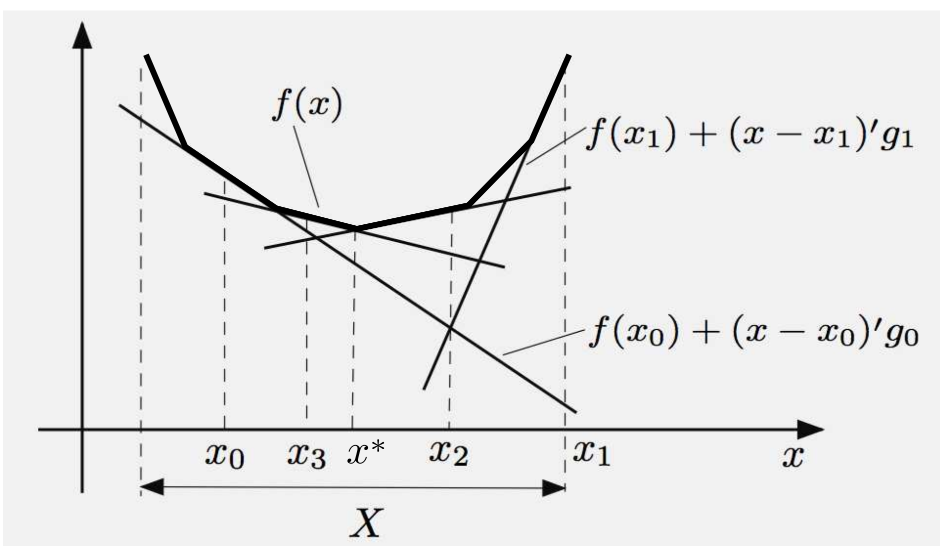
CONVERGENCE AND TERMINATION

- 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.

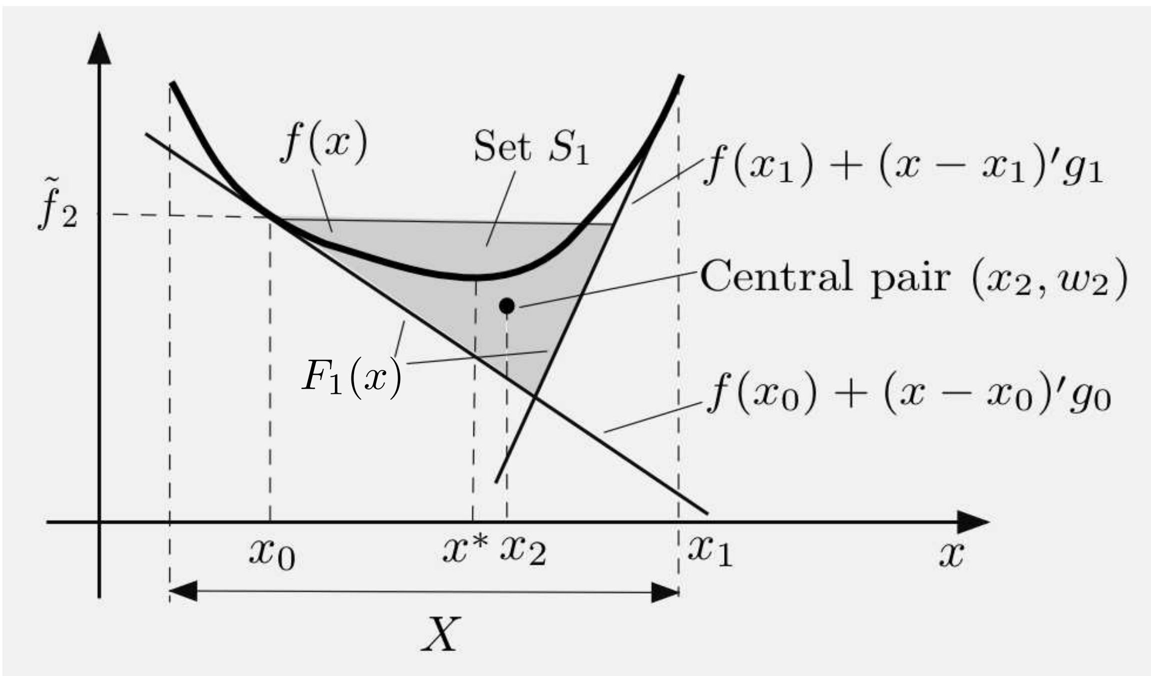
- For f polyhedral, we have finite termination with an exactly optimal solution.



- **Instability problem:** The method can make large moves that deteriorate the value of f.

VARIANTS

- **Variant I:** Simultaneously with f, construct polyhedral approximations to X.
- **Variant II:** Central cutting plane methods



- **Variant III:** Proximal methods - to be discussed next.

PROXIMAL/BUNDLE METHODS

- Aim to reduce the instability problem at the expense of solving a more difficult subproblem.
- A general form:

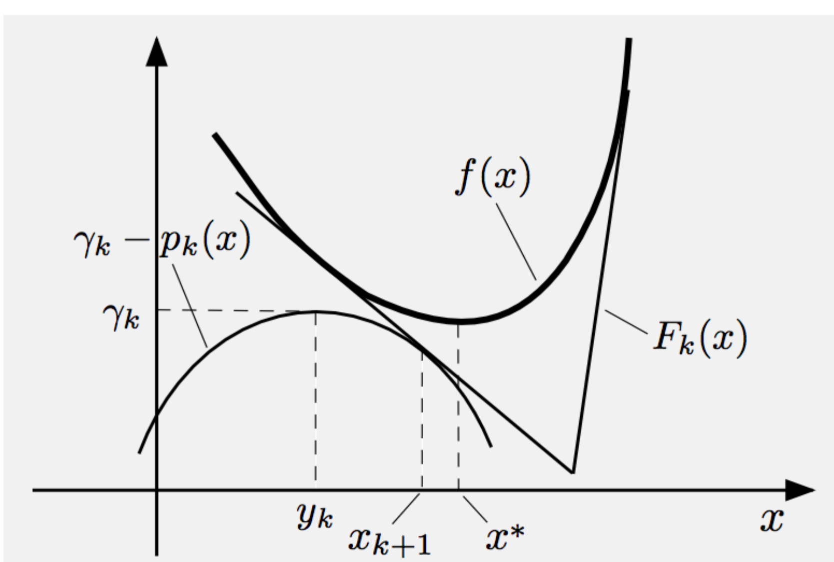
$$x_{k+1} \in \arg\min_{x \in X} \big\{ F_k(x) + p_k(x) \big\}$$

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

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

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*.

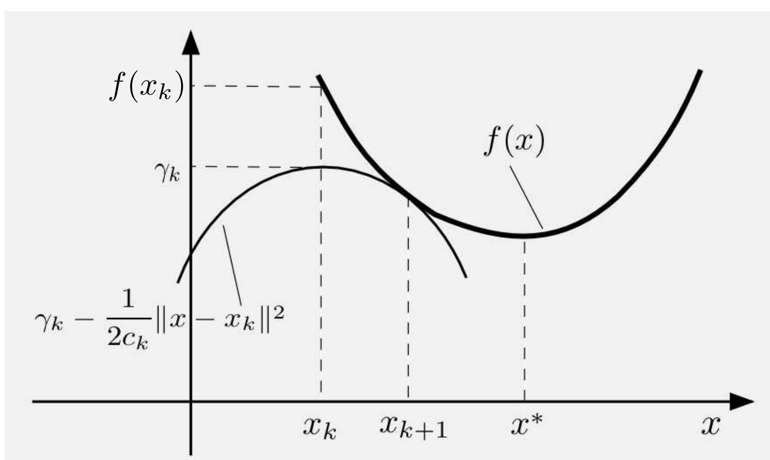


PROXIMAL MINIMIZATION ALGORITHM

- Starting point for analysis: A general algorithm for convex function 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



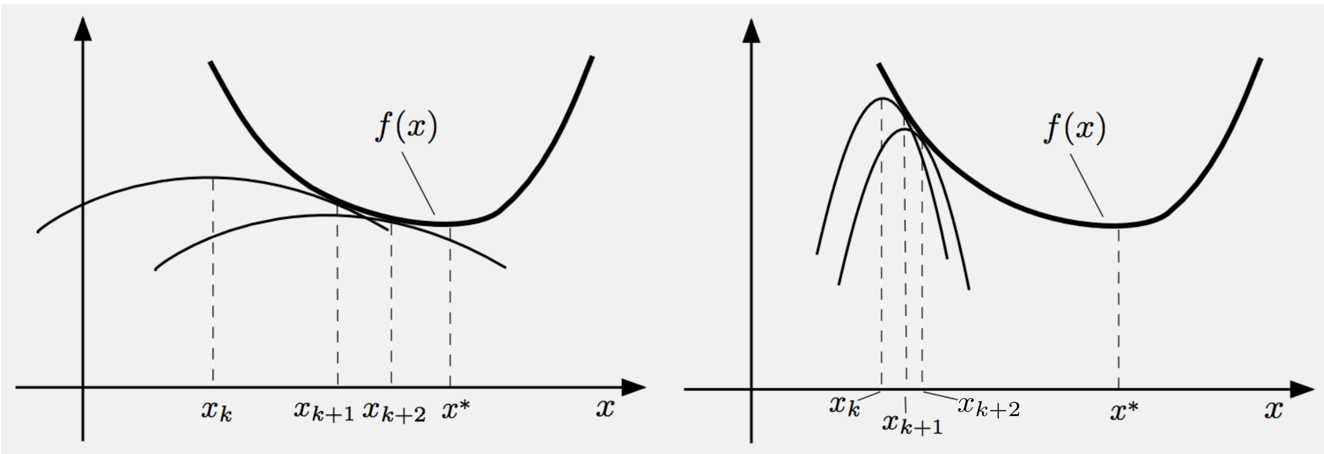
- Convergence mechanism:

$$\gamma_k = f(x_{k+1}) + \frac{1}{2c_k} \|x_{k+1} - x_k\|^2 < f(x_k).$$

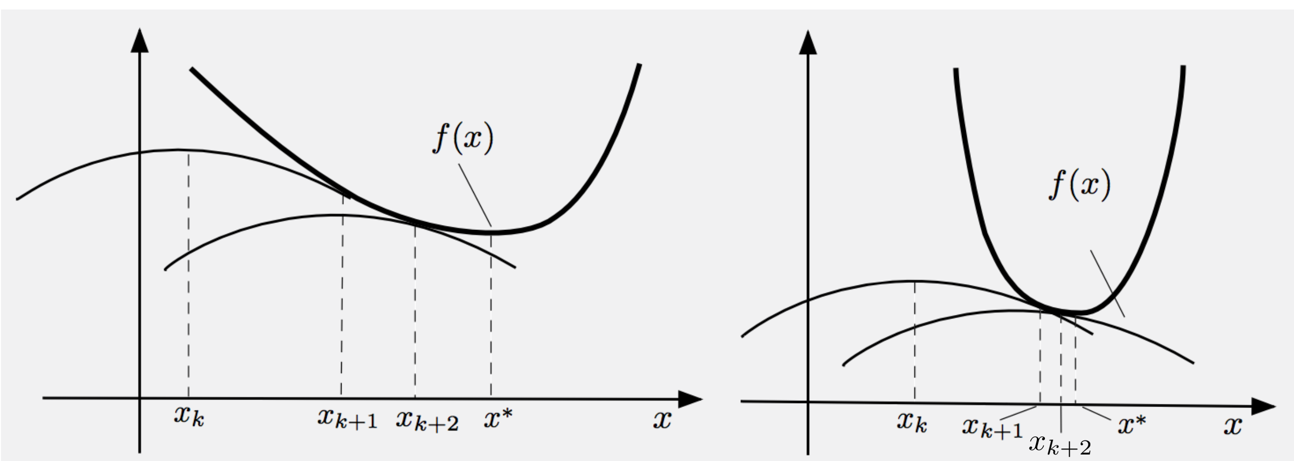
Cost improves by at least $\frac{1}{2c_k}\|x_{k+1} - x_k\|^2$, and this is sufficient to guarantee convergence.

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

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

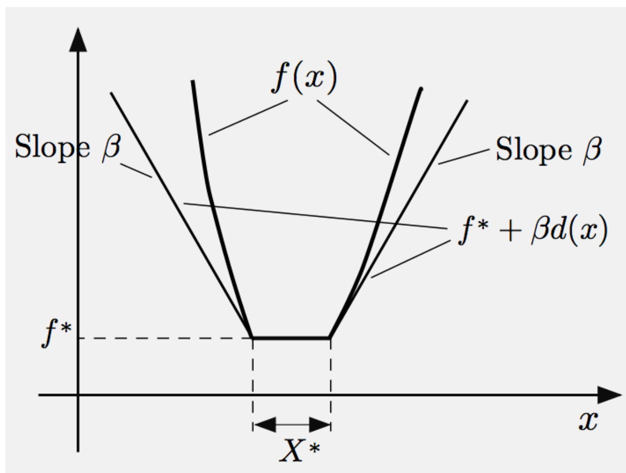
superlinear convergence.

FINITE CONVERGENCE

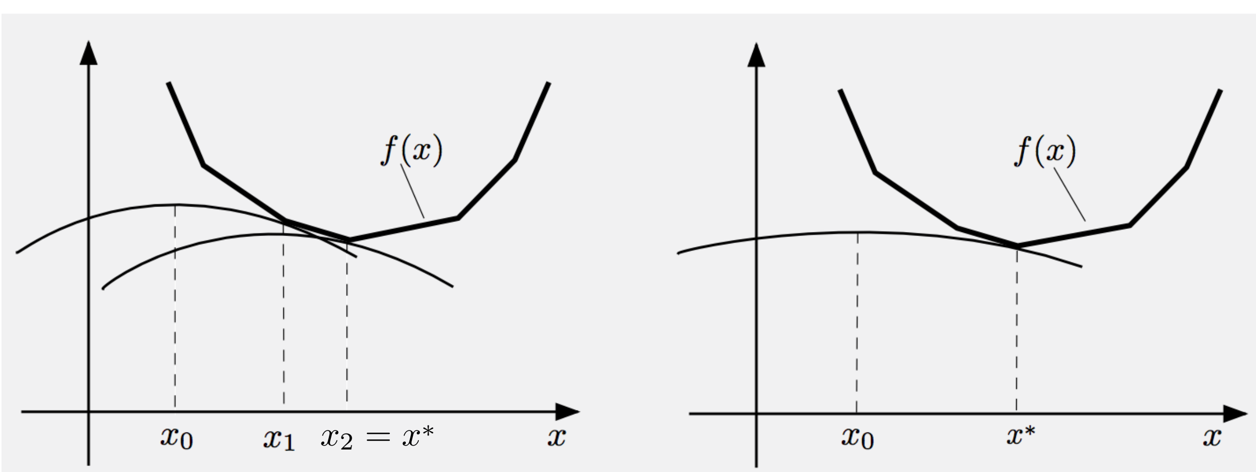
- Assume growth order $\alpha = 1$:

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

e.g., f is polyhedral.



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



PROXIMAL CUTTING PLANE METHODS

- 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

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

- Drawbacks:

  (a) **Hard stability tradeoff:** 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 $y_k \neq x_k$:

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

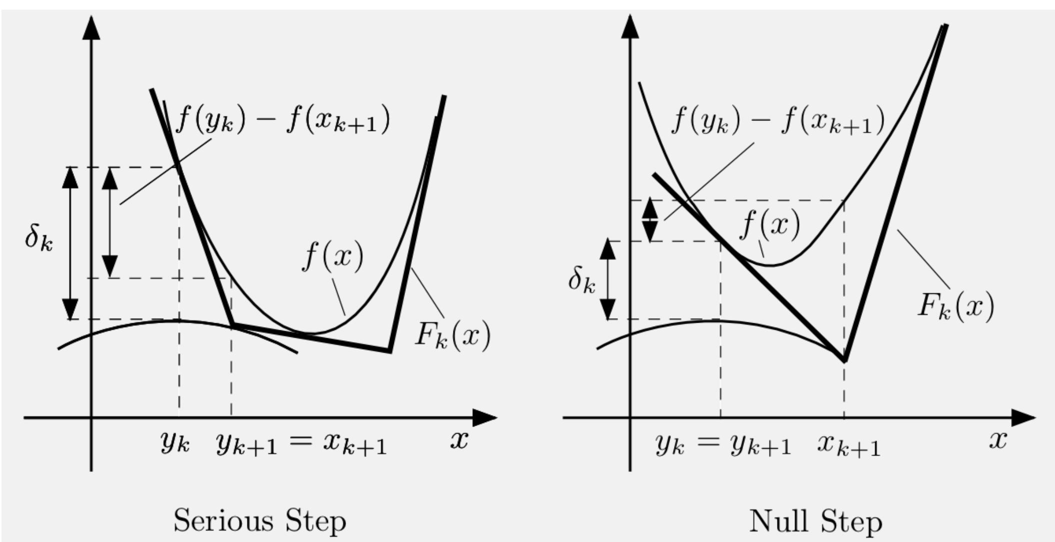
$$F_k(x) = \max\big\{ f(x_0)+(x-x_0)'g_0, \ldots, f(x_k)+(x-x_k)'g_k \big\}$$

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

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

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

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



Serious Step

Null Step

LECTURE 19

LECTURE OUTLINE

- Descent methods for convex/nondifferentiable optimization
- Steepest descent method
- ϵ-subdifferential
- ϵ-descent methods

****

- Consider minimization of a convex function $f : \Re^n \mapsto \Re$, over a closed convex set X.
- A basic iterative descent idea is to generate a sequence $\{x_k\}$ with

$$f(x_{k+1}) < f(x_k)$$

(unless x_k is optimal).

- If f is differentiable, we can use the gradient method

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

where α_k is a sufficiently small stepsize.

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

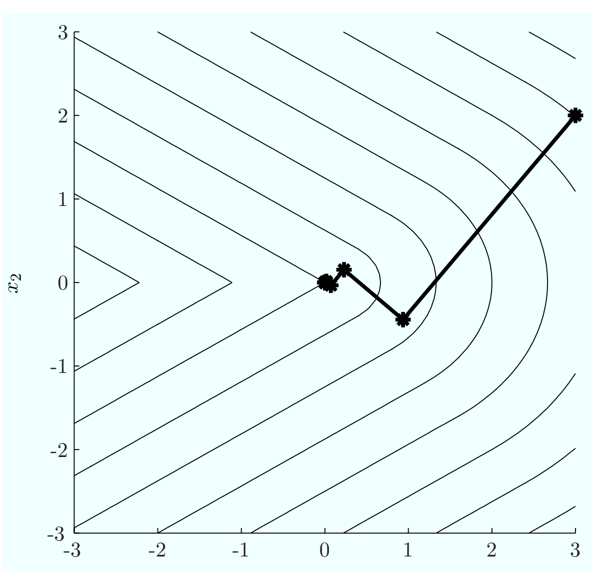
- Start with any $x_0 \in \Re^n$.

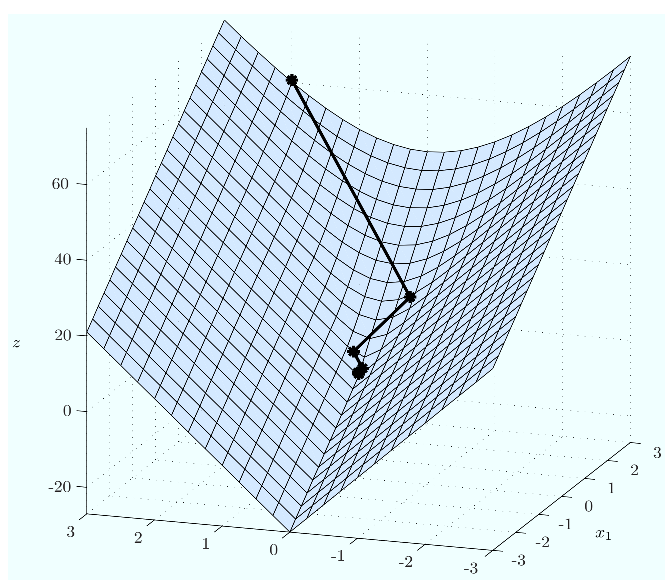
- For $k \geq 0$, calculate $-g_k$, the steepest descent direction at x_k and set

$$x_{k+1} = x_k - \alpha_k g_k$$

- **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.



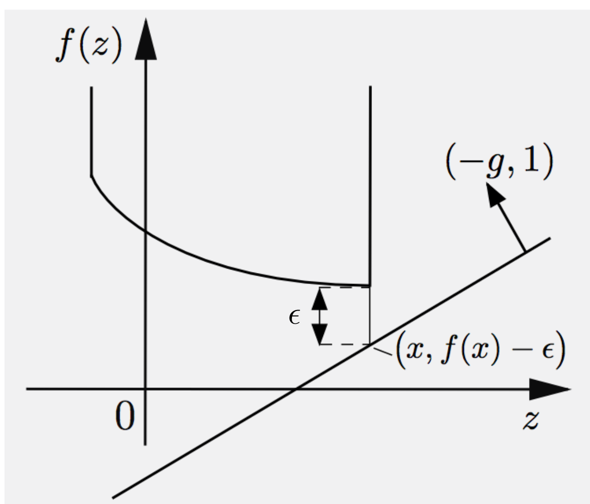


ϵ-SUBDIFFERENTIAL

- To correct the convergence deficiency of steepest descent, we may enlarge $\partial f(x)$ so that we take into account "nearby" subgradients.

- 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 $x \in \mathrm{dom}(f)$ 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, $\partial_\epsilon f(x) = \varnothing$ for $x \notin \mathrm{dom}(f)$.

- We have $\cap_{\epsilon \downarrow 0} \partial_\epsilon f(x) = \partial f(x)$ and

$$\partial_{\epsilon_1} f(x) \subset \partial_{\epsilon_2} f(x) \quad \text{if } 0 < \epsilon_1 < \epsilon_2$$

ϵ-SUBGRADIENTS AND CONJUGACY

- For any $x \in \mathrm{dom}(f)$, consider x-translation of f, i.e., the function f_x given by

$$f_x(d) = f(x+d) - f(x), \qquad \forall\, d \in \Re^n$$

and its conjugate

$$h_x(g) = \sup_{d \in \Re^n} \big\{ d'g - f(x+d) + f(x) \big\} = h(g) + f(x) - g'x$$

where h is the conjugate of f.

- We have

$$g \in \partial f(x) \qquad \text{iff} \qquad \sup_{d \in \Re^n} \big\{ g'd - f(x+d) + f(x) \big\} \le 0,$$

so $\partial f(x)$ can be characterized as a level set of h_x:

$$\partial f(x) = \big\{ g \mid h_x(g) \le 0 \big\}.$$

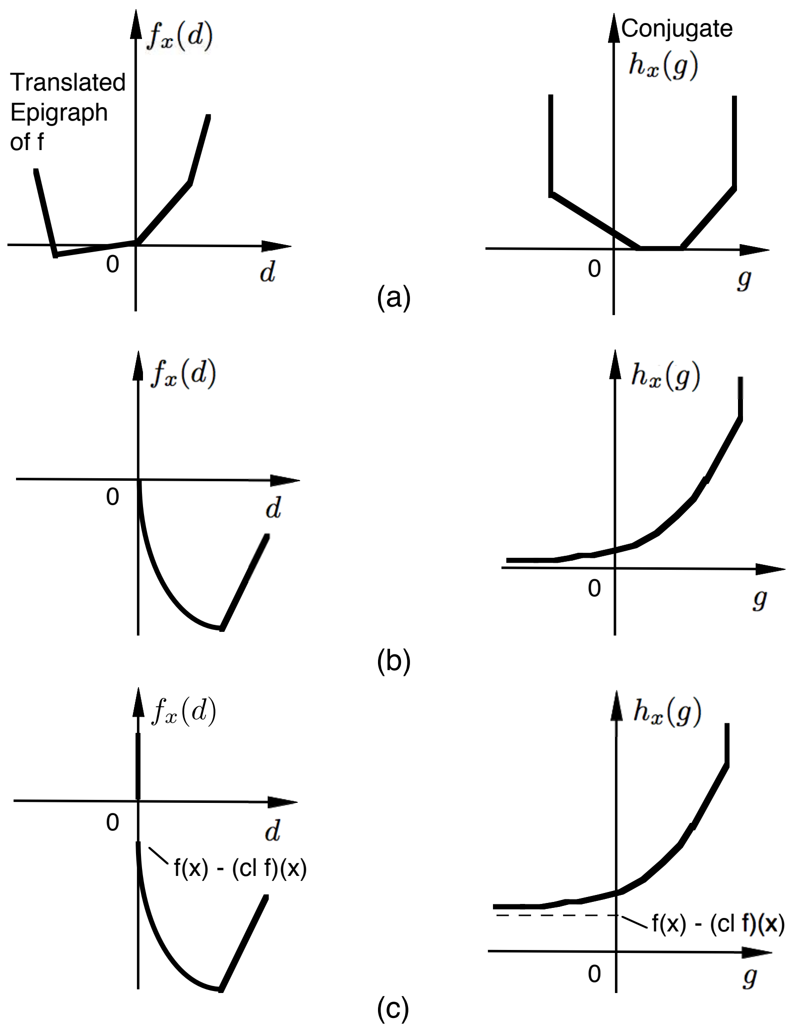
Similarly,

$$\partial_\epsilon f(x) = \big\{ g \mid h_x(g) \le \epsilon \big\}$$

ϵ-SUBDIFFERENTIALS AS LEVEL SETS

- For $h_x(g) = h(g) + f(x) - g'x$,

$$\partial_\epsilon f(x) = \{g \mid h_x(g) \leq \epsilon\}$$



- Since $(\mathrm{cl}\, f)(x) - f(x) = \sup_{g \in \Re^n} \{-h_x(g)\}$,

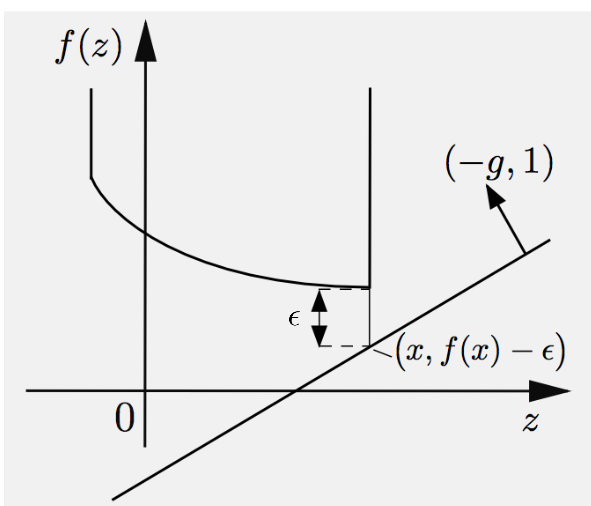
$$\inf_{g \in \Re^n} h_x(g) = 0 \qquad \text{if and only if} \qquad (\mathrm{cl}\, f)(x) = f(x),$$

so if f is closed, $\partial_\epsilon f(x) \neq \emptyset$ for every $x \in \mathrm{dom}(f)$.

PROPERTIES OF ϵ-SUBDIFFERENTIALS

- Assume that f is closed proper convex, $x \in \mathrm{dom}(f)$, and $\epsilon > 0$.
- $\partial_\epsilon f(x)$ is nonempty and closed.
- $\partial_\epsilon f(x)$ is compact iff h_x does no nonzero directions of recession. This is true in particular, if f is real-valued (support fn of dom is the recession fn of conjugate).
- The support function of $\partial_\epsilon f(x)$ is

$$\sigma_{\partial_\epsilon f(x)}(y) = \sup_{g \in \partial_\epsilon f(x)} y'g = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x) + \epsilon}{\alpha}$$



ϵ-DESCENT WITH ϵ-SUBDIFFERENTIALS

- We say that d is an *ϵ-descent direction at $x \in$* dom(f) if

$$\inf_{\alpha>0} f(x+\alpha d) < f(x) - \epsilon.$$

- Assuming f is closed proper convex, we have

$$\sigma_{\partial_\epsilon f(x)}(d) = \sup_{g\in\partial_\epsilon f(x)} d'g = \inf_{\alpha>0} \frac{f(x+\alpha d) - f(x) + \epsilon}{\alpha},$$

for all $d \in \Re^n$, so

$$d \text{ is an } \epsilon\text{-descent direction} \quad \text{iff} \quad \sup_{g\in\partial_\epsilon f(x)} d'g < 0$$

- If $0 \notin \partial_\epsilon f(x)$, the vector $-\overline{g}$, where

$$\overline{g} = \arg\min_{g\in\partial_\epsilon f(x)} \|g\|,$$

is an ϵ-descent direction.

- Also, from the definition, $0 \in \partial_\epsilon f(x)$ iff

$$f(x) \leq \inf_{z\in\Re^n} f(z) + \epsilon$$

ϵ-DESCENT METHOD

- The kth iteration is

$$x_{k+1} = x_k + \alpha_k d_k$$

where

$$-d_k = \arg \min_{g \in \partial_\epsilon f(x_k)} \|g\|,$$

and α_k is a positive stepsize.

- If $d_k = 0$, i.e., $0 \in \partial_\epsilon f(x_k)$, then x_k is an ϵ-optimal solution.

- If $d_k \neq 0$, choose α_k that reduces the cost function by at least ϵ, i.e.,

$$f(x_{k+1}) = f(x_k + \alpha_k d_k) \leq f(x_k) - \epsilon$$

- **Drawback:** Must know $\partial_\epsilon f(x_k)$.

- Motivation for a variant where $\partial_\epsilon f(x_k)$ is approximated by a set $A(x_k)$ that can be computed more easily than $\partial_\epsilon f(x_k)$.

- Then, $d_k = -g_k$, where

$$g_k = \arg \min_{g \in A(x_k)} \|g\|$$

ϵ-DESCENT METHOD - APPROXIMATIONS

- *Outer approximation methods*: Here $\partial_\epsilon f(x_k)$ is approximated by a set $A(x)$ such that

$$\partial_\epsilon f(x_k) \subset A(x_k) \subset \partial_{\gamma\epsilon} f(x_k),$$

where γ is a scalar with $\gamma > 1$.

- Example of outer approximation for case $f = f_1 + \cdots + f_m$:

$$A(x) = \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big),$$

based on the fact

$$\partial_\epsilon f(x) \subset \mathrm{cl}\big(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)\big) \subset \partial_{m\epsilon} f(x)$$

- Then the method terminates with an $m\epsilon$-optimal solution, and effects at least ϵ-reduction on f otherwise.

- Application to separable problems where each $\partial_\epsilon f_i(x)$ is a one-dimensional interval. Then to find an ϵ-descent direction, we must solve a quadratic program.

LECTURE 20

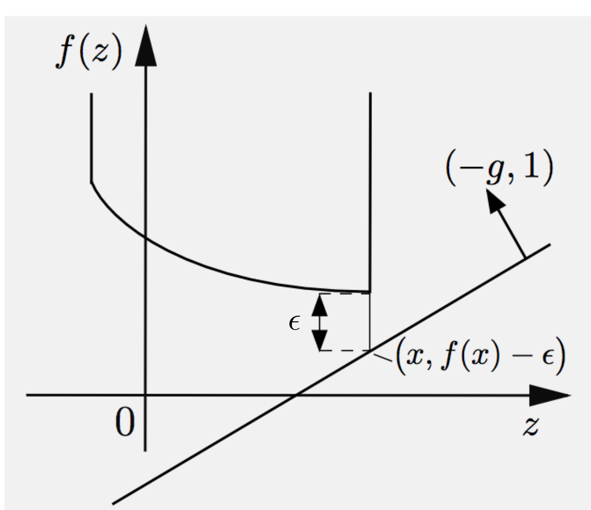
LECTURE OUTLINE

- Review of ϵ-subgradients
- ϵ-subgradient method
- Application to dual problems and minimax
- Incremental subgradient methods
- Connection with bundle methods

****

- For 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

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



ϵ-DESCENT WITH ϵ-SUBDIFFERENTIALS

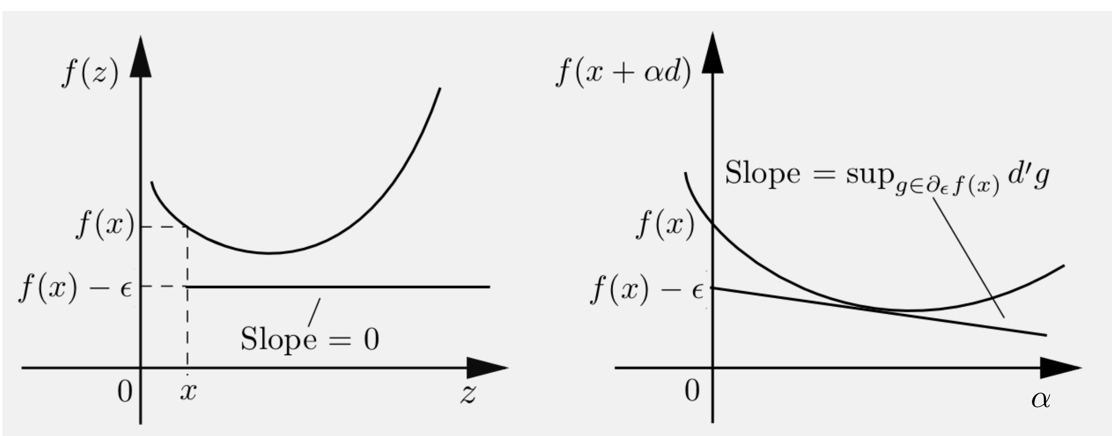
- Assume f is closed. We say that d is an *ϵ-descent direction at* $x \in \mathrm{dom}(f)$ if

$$\inf_{\alpha>0} f(x+\alpha d) < f(x) - \epsilon$$

Characterization:

$$d \text{ is an } \epsilon\text{-descent direction} \quad \text{iff} \quad \sup_{g\in\partial_\epsilon f(x)} d'g < 0$$

- Also, $0 \in \partial_\epsilon f(x)$ iff $f(x) \leq \inf_{z\in\Re^n} f(z) + \epsilon$



- If $0 \notin \partial_\epsilon f(x)$ and

$$\overline{g} = \arg \min_{g\in\partial_\epsilon f(x)} \|g\|$$

then $-\overline{g}$ is an ϵ-descent direction.

ϵ-DESCENT METHOD

- The kth iteration is

$$x_{k+1} = x_k + \alpha_k d_k$$

where

$$-d_k = \arg \min_{g \in \partial_\epsilon f(x_k)} \|g\|$$

and α_k is a positive stepsize.

- If $d_k = 0$, i.e., $0 \in \partial_\epsilon f(x_k)$, then x_k is an ϵ-optimal solution.

- If $d_k \neq 0$, choose α_k that reduces the cost function by at least ϵ, i.e.,

$$f(x_{k+1}) = f(x_k + \alpha_k d_k) \leq f(x_k) - \epsilon$$

- Drawback: Must know $\partial_\epsilon f(x_k)$.

- Need for variants.

ϵ-SUBGRADIENT METHOD

- This is an alternative/different type of method.
- Can be viewed as an approximate subgradient method, using 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.

- Fundamentally differs from ϵ-descent (it does not guarantee cost descent at each iteration).
- Can be viewed as subgradient method with "errors".
- Arises in several different contexts.

APPLICATION IN DUALITY AND MINIMAX

- Consider minimization of

$$f(x) = \sup_{z \in Z} \phi(x, z), \tag{1}$$

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

$$\phi(x, z_x) \geq \sup_{z \in Z} \phi(x, z) - \epsilon \text{ and } g_x \in \partial\phi(x, z_x)$$

$$\Rightarrow \qquad g_x \in \partial_\epsilon f(x)$$

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.

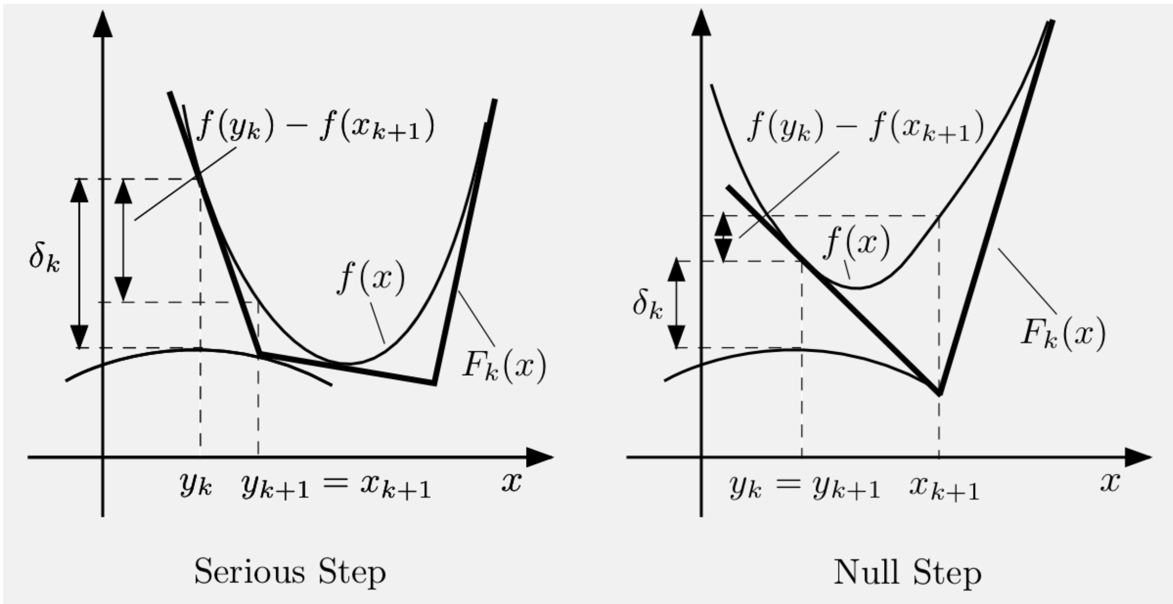
- The incremental subgradient iter. is an ϵ-subgradient iter. with $\epsilon = \epsilon_1 + \cdots + \epsilon_m$, where ϵ_i is the "error" in ith step in the cycle (ϵ_i: Proportional to α_k).
- Use

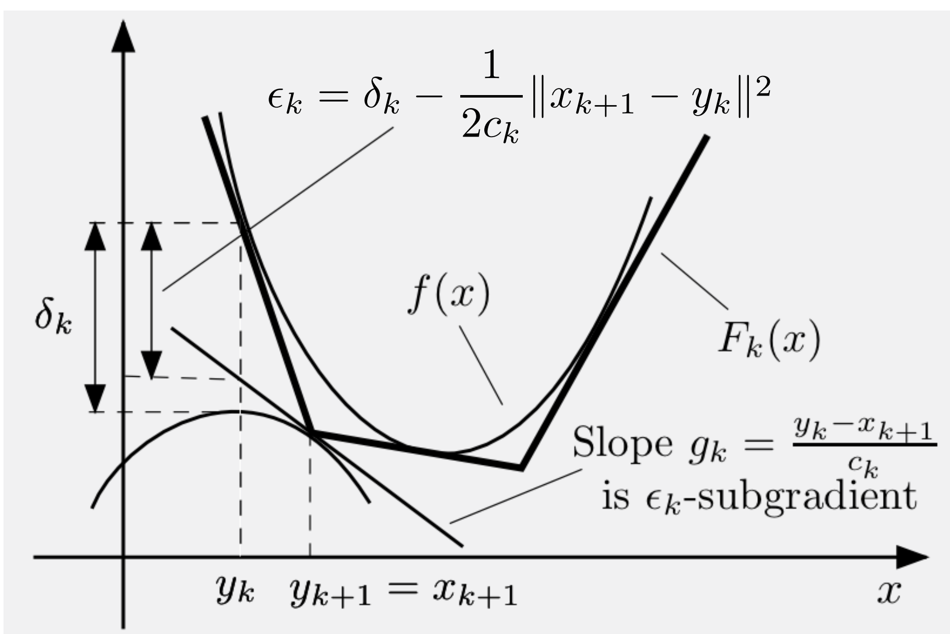
$$\partial_{\epsilon_1} f_1(x) + \cdots + \partial_{\epsilon_m} f_m(x) \subset \partial_\epsilon f(x),$$

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 $\alpha_k \to 0$, $\sum_k \alpha_k \to \infty$.

CONNECTION WITH BUNDLE METHOD





LECTURE 21

LECTURE OUTLINE

- Constrained minimization and duality
- Geometric Multipliers
- Dual problem - Weak duality
- Optimality Conditions
- Separable problems

****

- We consider the problem

$$\begin{array}{rl} & \text{minimize } \ f(x) \\ \text{subject to} & x \in X, \ \ g_1(x) \le 0, \ldots, g_r(x) \le 0 \end{array}$$

- We assume nothing on X, f, and g_j, except

$$-\infty < f^* = \inf_{\substack{x \in X \\ g_j(x) \le 0,\, j=1,\ldots,r}} f(x) < \infty$$

GEOMETRIC MULTIPLIERS

- A vector $\mu^* \geq 0$ is a *geometric multiplier* if

$$f^* = \inf_{x \in X} L(x, \mu^*),$$

where

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

- **Meaning of the definition:** μ^* is a G-multiplier if and only if $\mu^* \geq 0$ and the hyperplane of $\Re^{r+1}$ with normal $(\mu^*, 1)$ that passes through the point $(0, f^*)$ leaves every possible constraint-cost pair

$$\big(g(x), f(x)\big), \qquad x \in X,$$

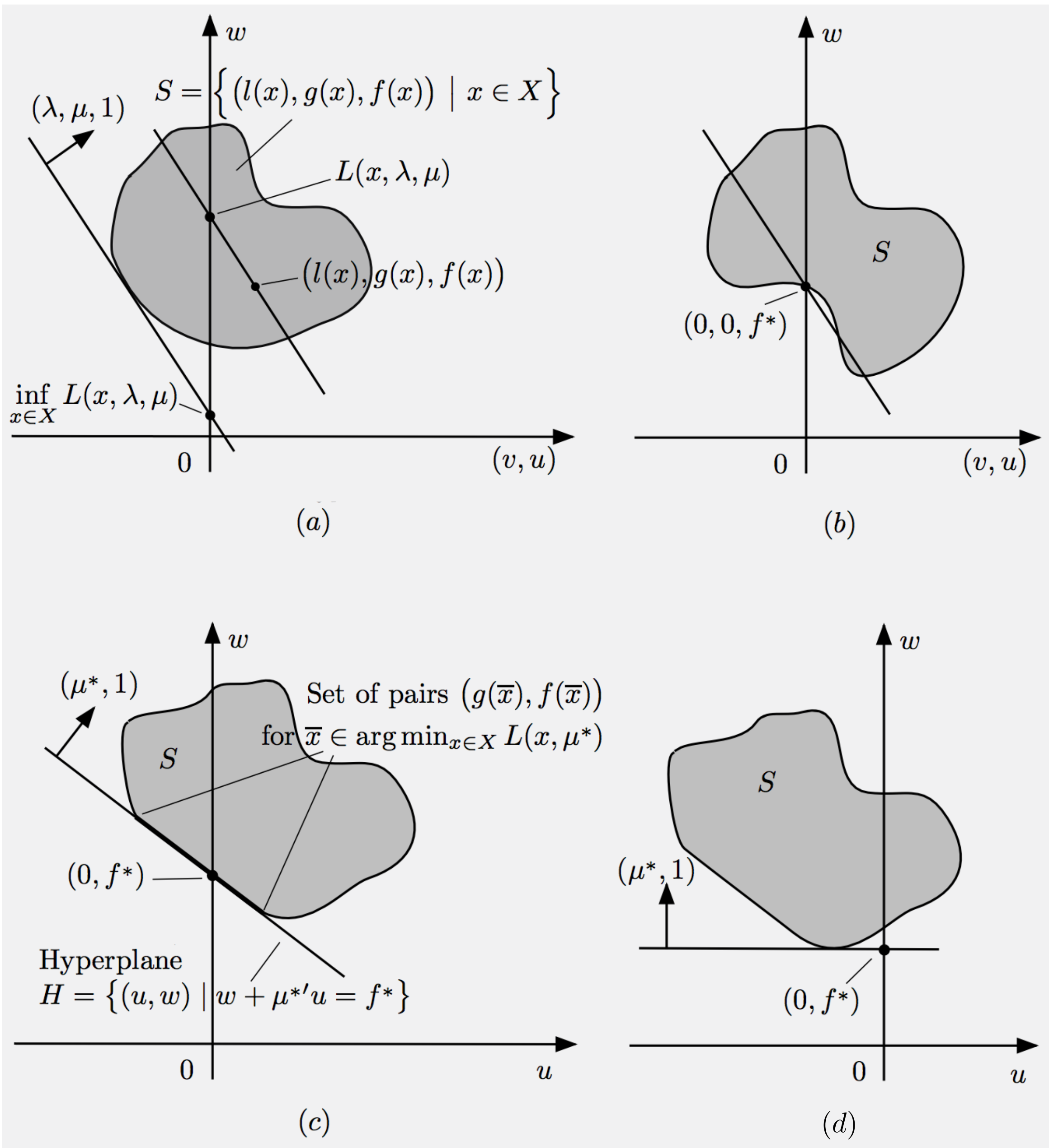
in its positive halfspace

$$\big\{(z, w) \in \Re^{r+1} \mid 1 \cdot w + \mu^{*\prime} \cdot z \geq 1 \cdot f^* + \mu^{*\prime} \cdot 0\big\}$$

- **Extension to equality constraints $l(x) = 0$:** A (λ^*, μ^*) is a geometric multiplier if $\mu^* \geq 0$ and

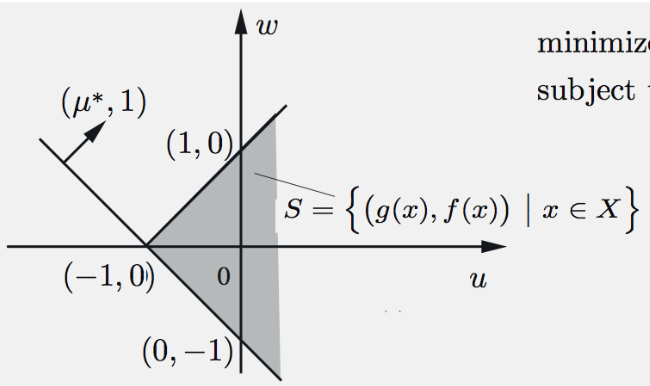
$$f^* = \inf_{x \in X} L(x, \lambda^*, \mu^*) = \inf_{x \in X} \big\{ f(x) + \lambda^{*\prime} l(x) + \mu^{*\prime} g(x) \big\}$$

VISUALIZATION



- Note: A G-multiplier solves a max-crossing problem whose min common problem has optimal value f^*.

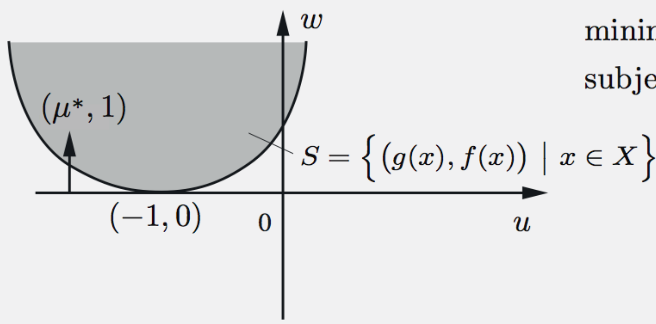
EXAMPLES: A G-MULTIPLIER EXISTS



minimize $\quad f(x) = x_1 - x_2$

subject to $\quad g(x) = x_1 + x_2 - 1 \leq 0$

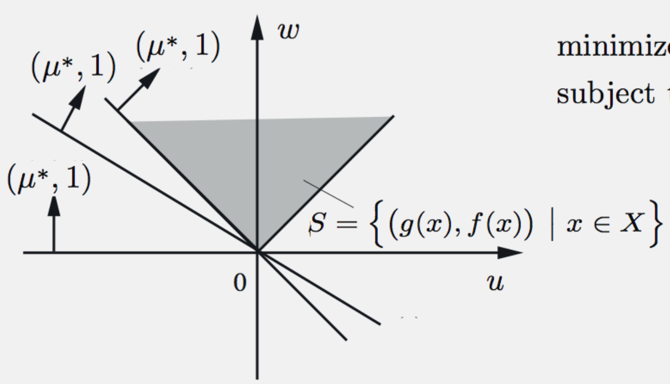
$\quad x \in X = \{(x_1, x_2) \mid x_1 \geq 0,\, x_2 \geq 0\}$



minimize $\quad f(x) = (1/2)(x_1^2 + x_2^2)$

subject to $\quad g(x) = x_1 - 1 \leq 0$

$\quad x \in X = \Re^2$

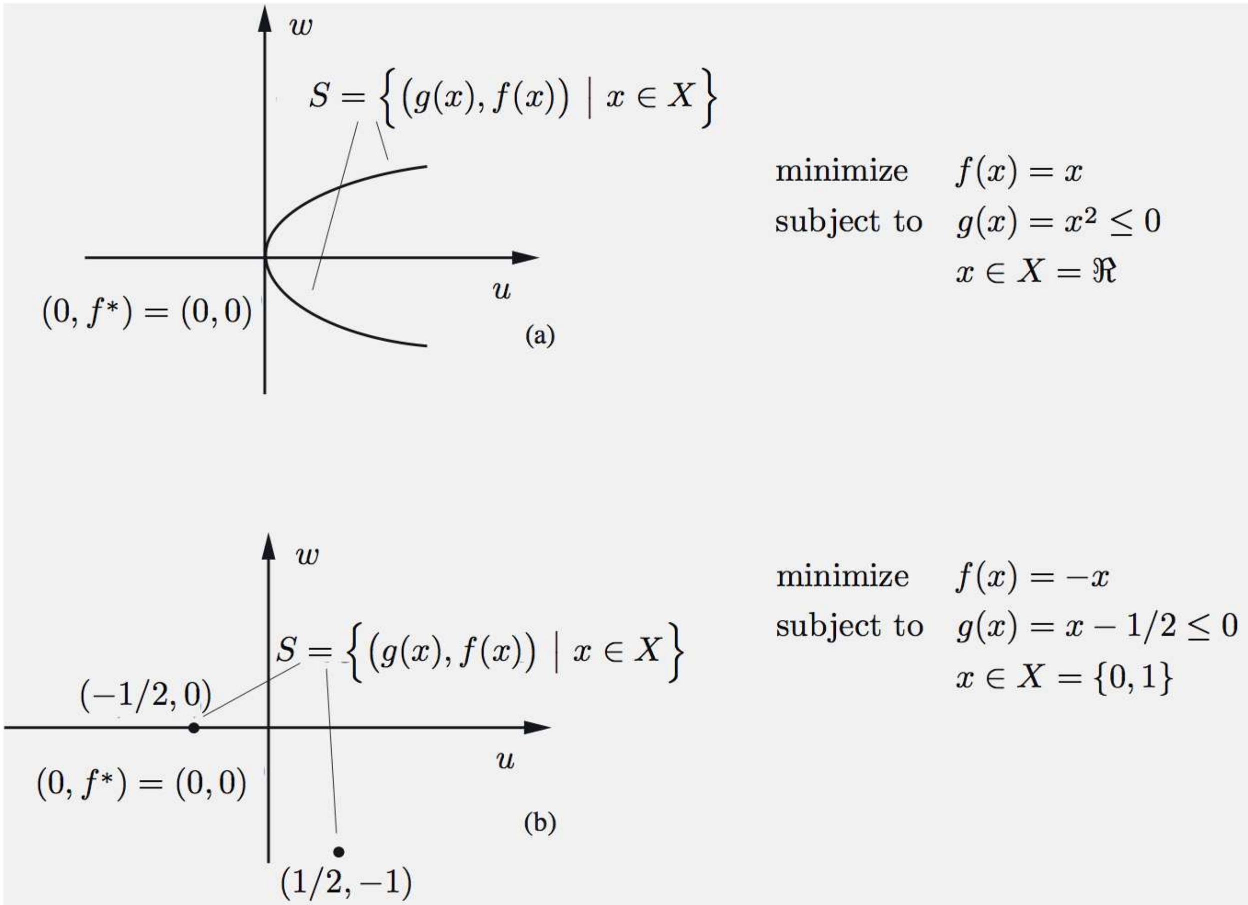


minimize $\quad f(x) = |x_1| + x_2$

subject to $\quad g(x) = x_1 \leq 0$

$\quad x \in X = \{(x_1, x_2) \mid x_2 \geq 0\}$

EXAMPLES: A G-MULTIPLIER DOESN'T EXIST



- **Proposition:** Let μ^* be a geometric multiplier. Then x^* is a global minimum of the primal problem if and only if x^* is feasible and

$$x^* = \arg\min_{x \in X} L(x, \mu^*), \quad \mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r$$

THE DUAL PROBLEM

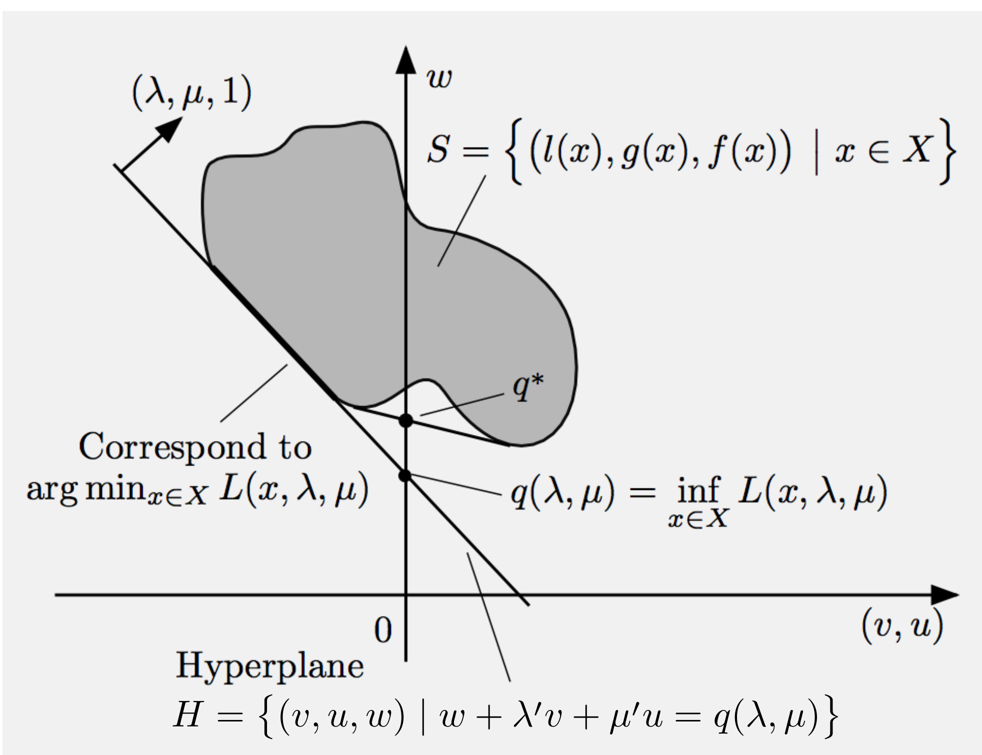
- The *dual problem* is

$$\begin{aligned} &\text{maximize} \;\; q(\mu) \\ &\text{subject to} \;\; \mu \geq 0, \end{aligned}$$

where q is the dual function

$$q(\mu) = \inf_{x \in X} L(x, \mu), \qquad \forall \; \mu \in \Re^r$$

- Note: The dual problem is equivalent to a max-crossing problem.



THE DUAL OF A LINEAR PROGRAM

- Consider the linear program

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad e_i'x = d_i, \quad i = 1, \ldots, m, \qquad x \geq 0 \end{aligned}$$

- Dual function

$$q(\lambda) = \inf_{x \geq 0} \left\{ \sum_{j=1}^{n} \left(c_j - \sum_{i=1}^{m} \lambda_i e_{ij} \right) x_j + \sum_{i=1}^{m} \lambda_i d_i \right\}$$

- If $c_j - \sum_{i=1}^{m} \lambda_i e_{ij} \geq 0$ for all j, the infimum is attained for $x = 0$, and $q(\lambda) = \sum_{i=1}^{m} \lambda_i d_i$. If $c_j - \sum_{i=1}^{m} \lambda_i e_{ij} < 0$ for some j, the expression in braces can be arbitrarily small by taking x_j suff. large, so $q(\lambda) = -\infty$. Thus, the dual is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} \lambda_i d_i \\ &\text{subject to} \quad \sum_{i=1}^{m} \lambda_i e_{ij} \leq c_j, \qquad j = 1, \ldots, n. \end{aligned}$$

WEAK DUALITY

- The *domain* of q is

$$D_q = \{\mu \mid q(\mu) > -\infty\}$$

- **Proposition:** q is concave, i.e., the domain D_q is a convex set and q is concave over D_q.

- **Proposition:** (Weak Duality Theorem) We have

$$q^* \leq f^*$$

Proof: For all $\mu \geq 0$, and $x \in X$ with $g(x) \leq 0$, we have

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

so

$$q^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in X,\, g(x) \leq 0} f(x) = f^*$$

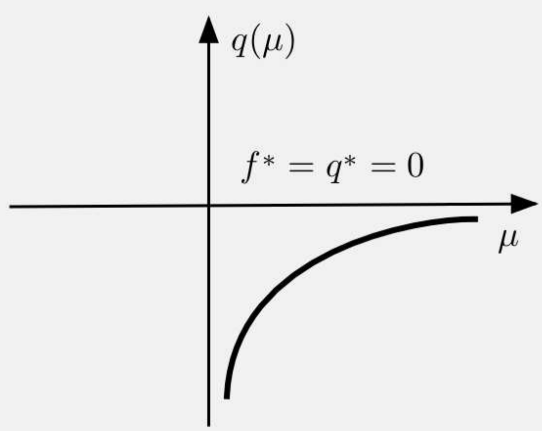
DUAL OPTIMAL SOLUTIONS

Proposition: (a) If $q^* = f^*$, the set of G-multipliers is equal to the set of optimal dual solutions.
(b) If $q^* < f^*$, the set of G-multipliers is empty (so if there exists a G-multiplier, $q^* = f^*$).

Proof: By definition, $\mu^* \geq 0$ is a G-multiplier if $f^* = q(\mu^*)$. Since $q(\mu^*) \leq q^*$ and $q^* \leq f^*$,

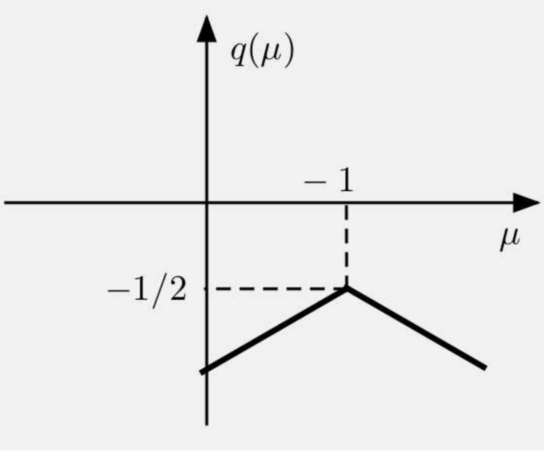
$$\mu^* \geq 0 \text{ is a G-multiplier} \quad \text{iff} \quad q(\mu^*) = q^* = f^*$$

- Examples (dual functions for the two problems with no G-multipliers, given earlier):



$$\begin{array}{ll} \text{minimize} & f(x) = x \\ \text{subject to} & g(x) = x^2 \leq 0 \\ & x \in X = \Re \end{array}$$

$$q(\mu) = \min_{x \in \Re} \{x + \mu x^2\} = \begin{cases} -1/(4\mu) & \text{if } \mu > 0 \\ -\infty & \text{if } \mu \leq 0 \end{cases}$$



$$\begin{array}{ll} \text{minimize} & f(x) = -x \\ \text{subject to} & g(x) = x - 1/2 \leq 0 \\ & x \in X = \{0, 1\} \end{array}$$

$$q(\mu) = \min_{x \in \{0,1\}} \{-x + \mu(x - 1/2)\} = \min\{-\mu/2, \mu/2 - 1\}$$

DUALITY AND MINIMAX THEORY

- The primal and dual problems can be viewed in terms of minimax theory:

$$\text{Primal Problem} \iff \inf_{x \in X} \sup_{\mu \geq 0} L(x, \mu)$$

$$\text{Dual Problem} \iff \sup_{\mu \geq 0} \inf_{x \in X} L(x, \mu)$$

- **Optimality Conditions:** (x^*, μ^*) is an optimal solution/G-multiplier pair if and only if

$$\begin{aligned} x^* \in X, \quad g(x^*) \leq 0, &\quad \text{(Primal Feasibility)}, \\ \mu^* \geq 0, &\quad \text{(Dual Feasibility)}, \\ x^* = \arg\min_{x \in X} L(x, \mu^*), &\quad \text{(Lagrangian Optimality)}, \\ \mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r, &\quad \text{(Compl. Slackness)}. \end{aligned}$$

- **Saddle Point Theorem:** (x^*, μ^*) is an optimal solution/G-multiplier pair if and only if $x^* \in X$, $\mu^* \geq 0$, and (x^*, μ^*) is a saddle point of the Lagrangian, in the sense that

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*), \quad \forall\, x \in X,\ \mu \geq 0$$

A CONVEX PROBLEM WITH A DUALITY GAP

- Consider the two-dimensional problem

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x_1 \le 0, \qquad x \in X = \{x \mid x \ge 0\}, \end{aligned}$$

where

$$f(x) = e^{-\sqrt{x_1 x_2}}, \qquad \forall\, x \in X,$$

and $f(x)$ is arbitrarily defined for $x \notin X$.

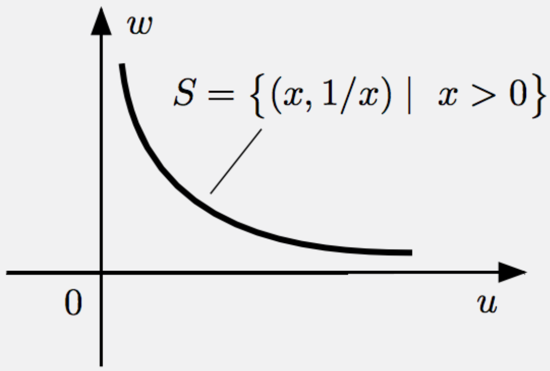
- f is convex over X (its Hessian is positive definite in the interior of X), and $f^* = 1$.

- Also, for all $\mu \ge 0$ we have

$$q(\mu) = \inf_{x \ge 0} \left\{ e^{-\sqrt{x_1 x_2}} + \mu x_1 \right\} = 0,$$

since the expression in braces is nonnegative for $x \ge 0$ and can approach zero by taking $x_1 \to 0$ and $x_1 x_2 \to \infty$. It follows that $q^* = 0$.

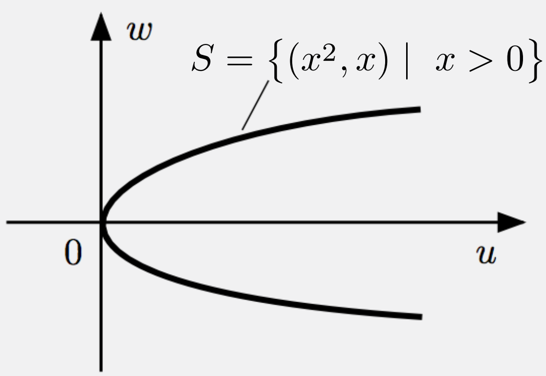
INFEASIBLE AND UNBOUNDED PROBLEMS



(a)

$$\begin{array}{ll} \text{minimize} & f(x) = 1/x \\ \text{subject to} & g(x) = x \le 0 \\ & x \in X = \{x \mid x > 0\} \end{array}$$

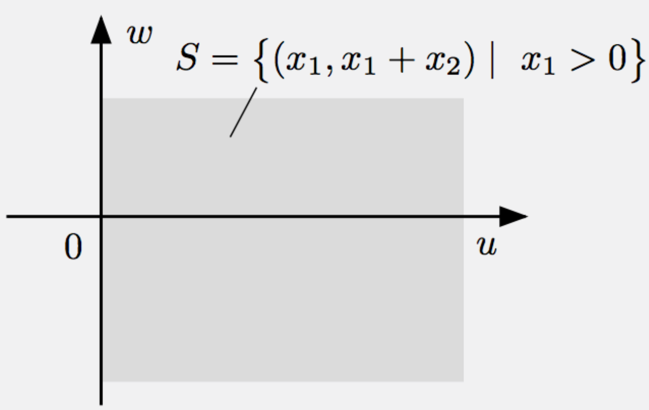
$f^* = \infty, \;\; q^* = \infty$



(b)

$$\begin{array}{ll} \text{minimize} & f(x) = x \\ \text{subject to} & g(x) = x^2 \le 0 \\ & x \in X = \{x \mid x > 0\} \end{array}$$

$f^* = \infty, \;\; q^* = 0$



(c)

$$\begin{array}{ll} \text{minimize} & f(x) = x_1 + x_2 \\ \text{subject to} & g(x) = x_1 \le 0 \\ & x \in X = \{(x_1, x_2) \mid x_1 > 0\} \end{array}$$

$f^* = \infty, \;\; q^* = -\infty$

SEPARABLE PROBLEMS I

- Suppose that $x = (x_1, \ldots, x_m)$, $x_i \in \Re^{n_i}$, and the problem is

$$
\begin{aligned}
&\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\
&\text{subject to} \ \sum_{i=1}^{m} g_{ij}(x_i) \le 0, \qquad j = 1, \ldots, r, \\
&\qquad\qquad\quad x_i \in X_i, \qquad i = 1, \ldots, m,
\end{aligned}
$$

where $f_i : \Re^{n_i} \mapsto \Re$ and $g_{ij} : \Re^{n_i} \mapsto \Re$, and $X_i \subset \Re^{n_i}$.

- Dual function:

$$
q(\mu) = \sum_{i=1}^{m} \inf_{x_i \in X_i} \left\{ f_i(x_i) + \sum_{j=1}^{r} \mu_j g_{ij}(x_i) \right\} = \sum_{i=1}^{m} q_i(\mu)
$$

- Set of constraint cost pairs $S = S_1 + \cdots + S_m$,

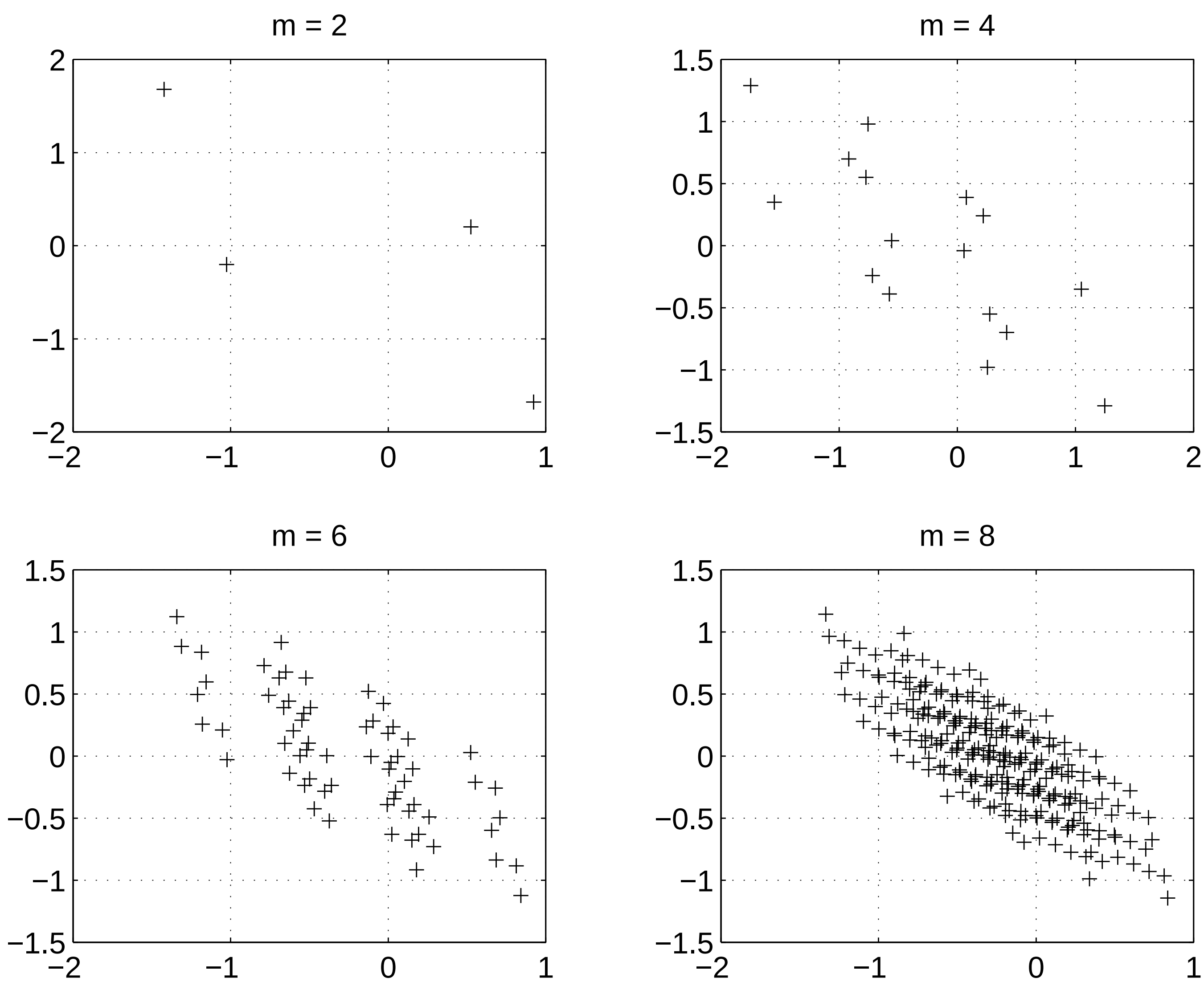
$$
S_i = \big\{ \big(g_i(x_i), f_i(x_i)\big) \mid x_i \in X_i \big\},
$$

and g_i is the function $g_i(x_i) = \big(g_{i1}(x_i), \ldots, g_{im}(x_i)\big)$.

SEPARABLE PROBLEMS II

- The sum of a large number of nonconvex sets is "almost" convex.

- **Shapley-Folkman Theorem:** Let X_i, $i = 1, \ldots, m$, be nonempty subsets of $\Re^n$ and let $X = X_1 + \cdots + X_m$. Then every vector $x \in \text{conv}(X)$ can be represented as $x = x_1 + \cdots + x_m$, where $x_i \in \text{conv}(X_i)$ for all $i = 1, \ldots, m$, and $x_i \in X_i$ for at least $m - n$ indices i.



LECTURE 22

LECTURE OUTLINE

- Conditions for existence of geometric multipliers
- Conditions for strong duality

- Primal problem: Minimize $f(x)$ subject to $x \in X$, and $g_1(x) \leq 0, \ldots, g_r(x) \leq 0$ (assuming $-\infty < f^* < \infty$). It is equivalent to $\inf_{x \in X} \sup_{\mu \geq 0} L(x, \mu)$.
- Dual problem: Maximize $q(\mu)$ subject to $\mu \geq 0$, where $q(\mu) = \inf_{x \in X} L(x, \mu)$. It is equivalent to $\sup_{\mu \geq 0} \inf_{x \in X} L(x, \mu)$.
- μ^* is a geometric multiplier if and only if $f^* = q^*$, and μ^* is an optimal solution of the dual problem.
- Question: Under what conditions $f^* = q^*$ and there exists a dual optimal solution?

RECALL NONLINEAR FARKAS' LEMMA

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

$$f(x) \geq 0, \qquad \forall\, x \in F = \big\{x \in X \mid g(x) \leq 0\big\},$$

and 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 F contains a relative interior point of X.

Then, there exists a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*) \geq 0$, such that

$$f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \geq 0, \qquad \forall\, x \in X$$

In case (1) the set of such μ^* is also compact.

APPLICATION TO CONVEX PROGRAMMING

Consider the problem

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

where X, $f : X \mapsto \Re$, and $g_j : X \mapsto \Re$ are convex. Assume that the optimal value f^* is finite.

- Replace $f(x)$ by $f(x) - f^*$ and assume that the conditions of Farkas' Lemma are satisfied. Then there exist $\mu_j^* \ge 0$ such that

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

Since $F \subset X$ and $\mu_j^* g_j(x) \le 0$ for all $x \in F$,

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

Thus equality holds throughout, we have

$$f^* = \inf_{x \in X} \left\{ f(x) + \mu^{*\prime} g(x) \right\},$$

and μ^* is a geometric multiplier.

STRONG DUALITY THEOREM I

Assumption : (Nonlinear Constraints - Slater Condition) f^* is finite, and the following hold:

(1) The functions f and g_j, $j = 1, \ldots, \overline{r}$, are convex over X.

(2) There exists a feasible vector $\bar{x}$ such that $g_j(\bar{x}) < 0$ for all $j = 1, \ldots, \overline{r}$.

Proposition : Under the above assumption, there exists at least one geometric multiplier.

Proof: Apply Farkas/condition(1).

STRONG DUALITY THEOREM II

Assumption : (Convexity and Linear Constraints) f^* is finite, and the following hold:

(1) The cost function f is convex over X and the functions g_j are affine.

(2) There exists a feasible solution of the problem that belongs to the relative interior of X.

Proposition : Under the above assumption, there exists at least one geometric multiplier.

Proof: Apply Farkas/condition(2).

- There is an extension to the case where $X = P \cap C$, where P is polyhedral and C is convex. Then f must be convex over C, and there must exist a feasible solution that belongs to the relative interior of C.

STRONG DUALITY THEOREM III

Assumption : (Linear and Nonlinear Constraints) f^* is finite, and the following hold:

(1) $X = P \cap C$, with P: polyhedral, C: convex.

(2) The functions f and g_j, $j = 1, \ldots, \overline{r}$, are convex over C, and the functions g_j, $j = \overline{r} + 1, \ldots, r$, are affine.

(3) There exists a feasible vector $\bar{x}$ such that $g_j(\bar{x}) < 0$ for all $j = 1, \ldots, \overline{r}$.

(4) There exists a vector that satisfies the linear constraints [but not necessarily the constraints $g_j(x) \leq 0$, $j = 1, \ldots, \overline{r}$] and belongs to the relative interior of C.

Proposition : Under the above assumption, there exists at least one geometric multiplier.

Proof: If $P = \Re^n$ and there are no linear constraints (the Slater condition), apply Farkas. Otherwise, lump the linear constraints within X, assert the existence of geometric multipliers for the nonlinear constraints, then use the preceding duality result for linear constraints. **Q.E.D.**

THE PRIMAL FUNCTION

- Minimax theory centered around the function

$$p(u) = \inf_{x \in X} \sup_{\mu \geq 0} \big\{ L(x, \mu) - \mu' u \big\}$$

- Properties of p around $u = 0$ are critical in analyzing the presence of a duality gap and the existence of primal and dual optimal solutions.

- p is known as the *primal function* of the constrained optimization problem.

- We have

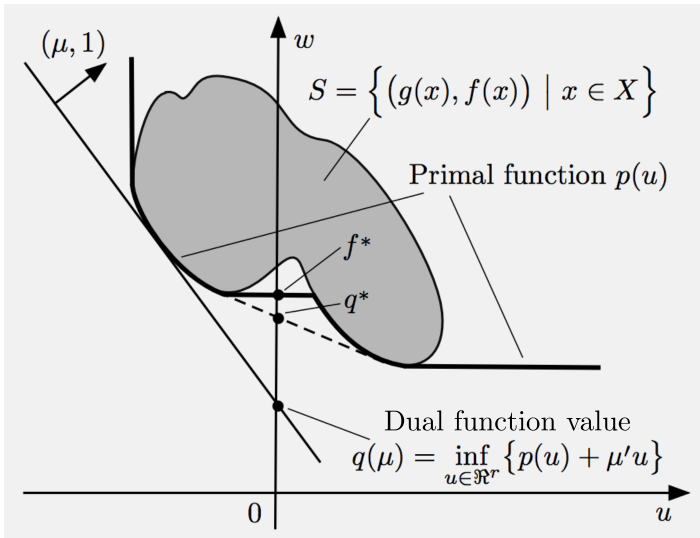
$$\begin{aligned} \sup_{\mu \geq 0} & \big\{ L(x, \mu) - \mu' u \big\} \\ & = \sup_{\mu \geq 0} \big\{ f(x) + \mu' \big(g(x) - u \big) \big\} \\ & = \begin{cases} f(x) & \text{if } g(x) \leq u, \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

- So

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

and $p(u)$ can be viewed as a *perturbed optimal value* [note that $p(0) = f^*$].

RELATION OF PRIMAL AND DUAL FUNCTIONS



- Consider the dual function q. For every $\mu \geq 0$, we have

$$
\begin{aligned}
q(\mu) &= \inf_{x \in X} \{f(x) + \mu' g(x)\} \\
&= \inf_{\{(u,x) \mid x \in X,\, g(x) \leq u\}} \{f(x) + \mu' g(x)\} \\
&= \inf_{\{(u,x) \mid x \in X,\, g(x) \leq u\}} \{f(x) + \mu' u\} \\
&= \inf_{u \in \Re^r} \inf_{x \in X,\, g(x) \leq u} \{f(x) + \mu' u\}.
\end{aligned}
$$

- Thus we have the conjugacy relation

$$
q(\mu) = \inf_{u \in \Re^r} \{p(u) + \mu' u\}, \qquad \forall\, \mu \geq 0
$$

CONDITIONS FOR NO DUALITY GAP

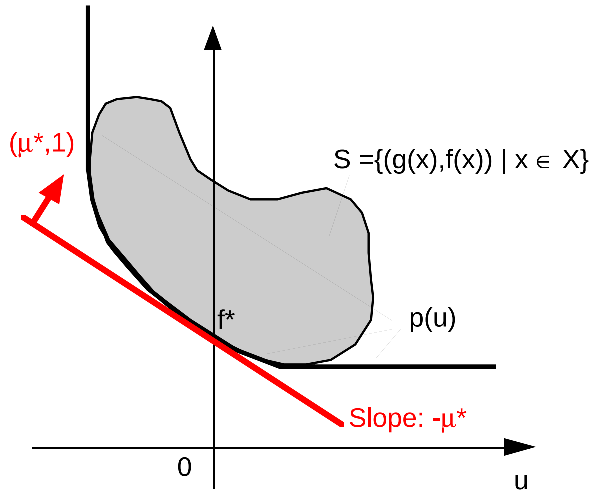
- Apply the minimax theory specialized to $L(x, \mu)$.

- Assume $f^* < \infty$, X is convex, and $L(\cdot, \mu)$ is convex over X for each $\mu \geq 0$. Then:
  - p is convex.
  - There is no duality gap if and only if p is lower semicontinuous at $u = 0$.

- Conditions that guarantee lower semicontinuity at $u = 0$, correspond to those for preservation of closure under partial minimization, e.g.:
  - $f^* < \infty$, X is convex and compact, and for each $\mu \geq 0$, the function $L(\cdot, \mu)$, restricted to have domain X, is closed and convex.
  - Extensions involving directions of recession of X, f, and g_j, and guaranteeing that the minimization in $p(u) = \inf_{\substack{x \in X \\ g(x) \leq u}} f(x)$ is (effectively) over a compact set.

- Under the above conditions, there is no duality gap, and the primal problem has a nonempty and compact optimal solution set. Furthermore, the primal function p is closed, proper, and convex.

SUBGRADIENTS OF THE PRIMAL FUNCTION



- Assume that p is convex, $p(0)$ is finite, and p is proper. Then:
  - The set of G-multipliers is $-\partial p(0)$. This follows from the relation

$$q(\mu) = \inf_{u \in \Re^r} \{p(u) + \mu' u\}, \qquad \forall\, \mu \geq 0$$

  - If p is differentiable at 0, there is a unique G-multiplier: $\mu^* = -\nabla p(0)$.
  - If the origin lies in the interior of $\text{dom}(p)$, the set of G-multipliers is nonempty and compact. (This is true iff the Slater condition holds.)

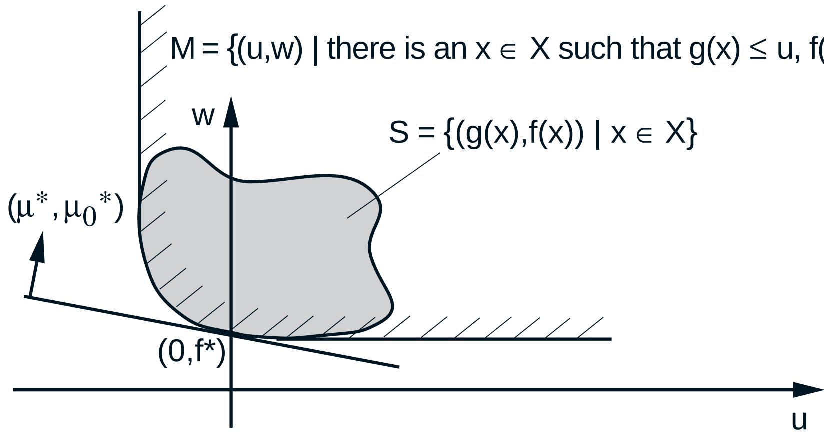
FRITZ JOHN THEORY

- Assume that X is convex, the functions f and g_j are convex over X, and $f^* < \infty$. Then there exist a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ satisfying the following conditions:

  (i) $\mu_0^* f^* = \inf_{x \in X} \{\mu_0^* f(x) + \mu^{*\prime} g(x)\}$.

  (ii) $\mu_j^* \geq 0$ for all $j = 0, 1, \ldots, r$.

  (iii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$ are not all equal to 0.



- If the multiplier μ_0^* can be proved positive, then μ^*/μ_0^* is a G-multiplier.

- Under the Slater condition (there exists $\overline{x} \in X$ s.t. $g(\overline{x}) < 0$), μ_0^* cannot be 0; if it were, then $0 = \inf_{x \in X} \mu^{*\prime} g(x)$ for some $\mu^* \geq 0$ with $\mu^* \neq 0$, while we would also have $\mu^{*\prime} g(\overline{x}) < 0$.

F-J THEORY FOR LINEAR CONSTRAINTS

- Assume that X is convex, f is convex over X, the g_j are affine, and $f^* < \infty$. Then there exist a scalar μ_0^* and a vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$, satisfying the following conditions:

  (i) $\mu_0^* f^* = \inf_{x \in X} \big\{ \mu_0^* f(x) + \mu^{*\prime} g(x) \big\}$.

  (ii) $\mu_j^* \geq 0$ for all $j = 0, 1, \ldots, r$.

  (iii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$ are not all equal to 0.

  (iv) If the index set $J = \{ j \neq 0 \mid \mu_j^* > 0 \}$ is nonempty, there exists a vector $\tilde{x} \in X$ such that $f(\tilde{x}) < f^*$ and $\mu^{*\prime} g(\tilde{x}) > 0$.

- Proof uses Polyhedral Proper Separation Th.

- Can be used to show that there exists a geometric multiplier if $X = P \cap C$, where P is polyhedral, and $\mathrm{ri}(C)$ contains a feasible solution.

- **Conclusion:** The Fritz John theory is sufficiently powerful to show the major constraint qualification theorems for convex programming.

LECTURE 23

LECTURE OUTLINE

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

−−

- We introduce another "standard" framework:

$$\begin{aligned} &\text{minimize} \;\; f_1(x) - f_2(x) \\ &\text{subject to} \;\; x \in X_1 \cap X_2, \end{aligned}$$

$f_1, f_2 : \Re^n \mapsto \Re$, and X_1, X_2 are subsets of $\Re^n$.

- It can be shown to be equivalent to the Lagrangian framework

$$\begin{aligned} &\text{minimize} \;\; f(x) \\ &\text{subject to} \;\; x \in X, \;\; g_1(x) \le 0, \ldots, g_r(x) \le 0 \end{aligned}$$

but it is more convenient for some applications, e.g., network flow, and conic/semidefinite programming.

FENCHEL DUALITY FRAMEWORK

- Consider the problem

$$\begin{aligned} &\text{minimize} \;\; f_1(x) - f_2(x) \\ &\text{subject to} \;\; x \in X_1 \cap X_2, \end{aligned}$$

where $f_1, f_2 : \Re^n \mapsto \Re$, and X_1, X_2 are subsets of $\Re^n$.

- Assume that $f^* < \infty$.
- Convert the problem to

$$\begin{aligned} &\text{minimize} \;\; f_1(y) - f_2(z) \\ &\text{subject to} \;\; z = y, \qquad y \in X_1, \quad z \in X_2, \end{aligned}$$

and dualize the constraint $z = y$:

$$\begin{aligned} q(\lambda) &= \inf_{y \in X_1,\, z \in X_2} \big\{ f_1(y) - f_2(z) + (z - y)'\lambda \big\} \\ &= \inf_{z \in X_2} \big\{ z'\lambda - f_2(z) \big\} - \sup_{y \in X_1} \big\{ y'\lambda - f_1(y) \big\} \\ &= h_2(\lambda) - h_1(\lambda) \end{aligned}$$

PRIMAL FENCHEL DUALITY THEOREM

- We view f_1 and $-f_2$ as extended real-valued with domains X_1 and X_2, and write the primal and dual problems as

$$\min_{x\in\Re^n}\big\{f_1(x)-f_2(x)\big\}, \qquad \max_{\lambda\in\Re^n}\big\{h_2(\lambda)-h_1(\lambda)\big\}$$

- Use strong duality theorems for the problem

$$\min_{z=y,\, y\in X_1,\, z\in X_2}\big\{f_1(y)-f_2(z)\big\}$$

- **Primal Fenchel Duality Theorem:** The dual problem has an optimal solution and we have

$$\inf_{x\in\Re^n}\big\{f_1(x)-f_2(x)\big\}=\max_{\lambda\in\Re^n}\big\{h_2(\lambda)-h_1(\lambda)\big\},$$

if f_1, $-f_2$, X_1, X_2 are convex, and *one* of the following two conditions holds:
  - The relative interiors of X_1 and X_2 intersect
  - X_1 and X_2 are polyhedral, and f_1 and f_2 can be extended to real-valued convex and concave functions over $\Re^n$.

OPTIMALITY CONDITIONS

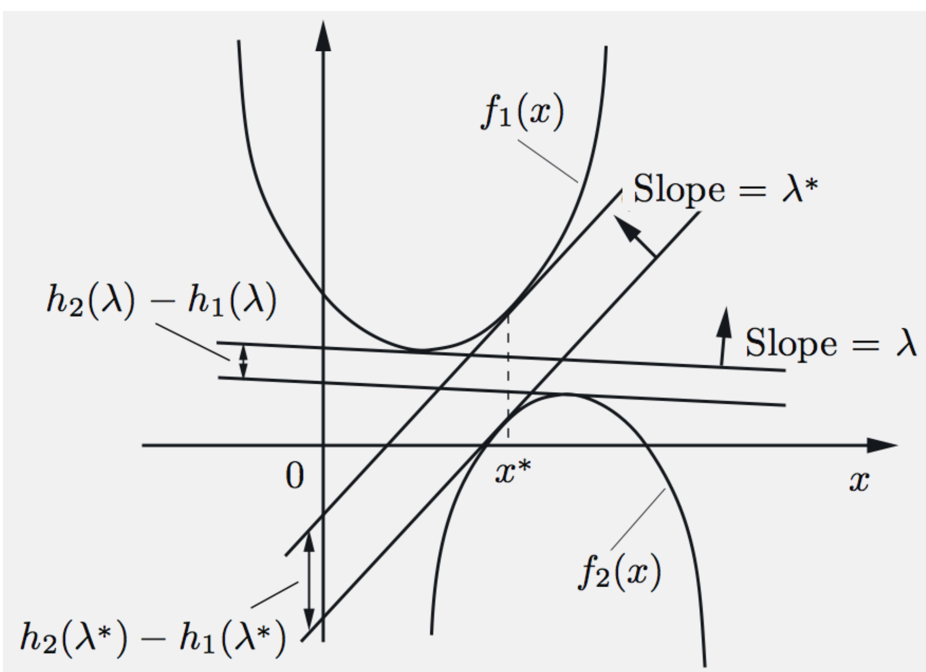
- Assume $-\infty < q^* = f^* < \infty$. Then (x^*, λ^*) is an optimal primal and dual solution pair if and only if

$$x^* \in \mathrm{dom}(f_1) \cap \mathrm{dom}(-f_2), \qquad \text{(primal feasibility)},$$

$$\lambda^* \in \mathrm{dom}(h_1) \cap \mathrm{dom}(-h_2), \qquad \text{(dual feasibility)},$$

$$x^* \in \arg\max_{y \in \Re^n} \big\{ y'\lambda^* - f_1(y) \big\}$$

$$x^* \in \arg\min_{z \in \Re^n} \big\{ z'\lambda^* - f_2(z) \big\}, \qquad \text{(Lagr. optimality)}.$$



- Note: The Lagrangian optimality condition is equivalent to $\lambda^* \in \partial f_1(x^*) \cap \partial f_2(x^*)$.

DUAL FENCHEL DUALITY THEOREM

- The dual problem

$$\max_{\lambda \in \Re^n} \big\{ h_2(\lambda) - h_1(\lambda) \big\}$$

is of the same form as the primal.

- By the conjugacy theorem, if the functions f_1 and f_2 are closed, in addition to being convex and concave, they are the conjugates of h_1 and h_2.

- **Conclusion:** The primal problem has an optimal solution and we have

$$\min_{x \in \Re^n} \big\{ f_1(x) - f_2(x) \big\} = \sup_{\lambda \in \Re^n} \big\{ h_2(\lambda) - h_1(\lambda) \big\}$$

if *one* of the following two conditions holds

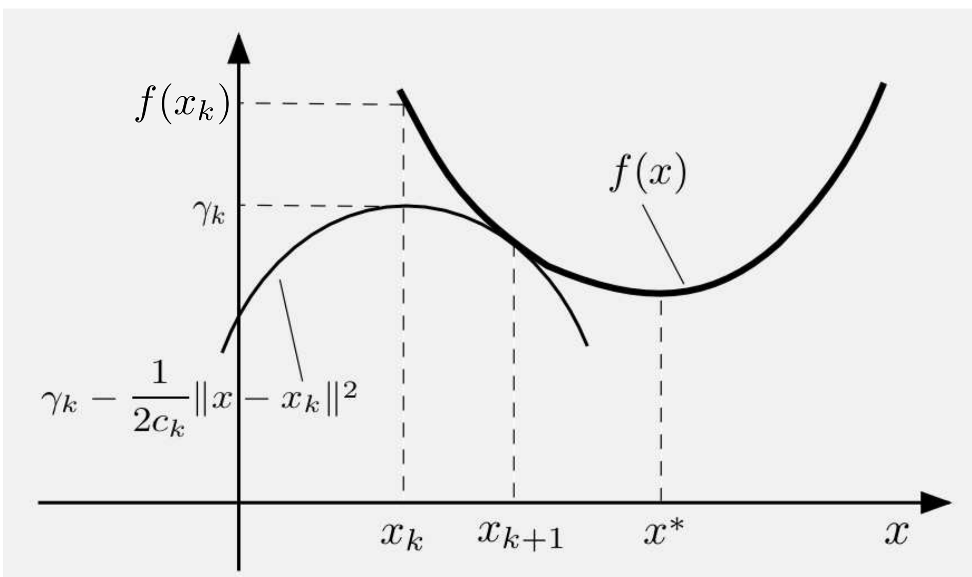
- The relative interiors of $\mathrm{dom}(h_1)$ and $\mathrm{dom}(-h_2)$ intersect.
- $\mathrm{dom}(h_1)$ and $\mathrm{dom}(-h_2)$ are polyhedral, and h_1 and h_2 can be extended to real-valued convex and concave functions over $\Re^n$.

RECALL PROXIMAL MINIMIZATION

- Applies to minimization of convex f:

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

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



- We have $f(x_k) \to f^*$ and $x_k \to$ some minimizer of f, provided one exists.
- Finite convergence for polyhedral f.

DUAL PROXIMAL MINIMIZATION

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

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

- The Fenchel dual is

$$\begin{aligned} &\text{maximize} \quad h_2(\lambda) - h_1(\lambda) \\ &\text{subject to} \quad \lambda \in \Re^n \end{aligned}$$

where h_1, h_2 are conjugates of f_1, f_2.

- After calculation, it becomes

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

where h is the convex conjugate of f.

- f_2 and h_2 are 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:

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

- Lagragian optimality conditions for primal:

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

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

or equivalently,

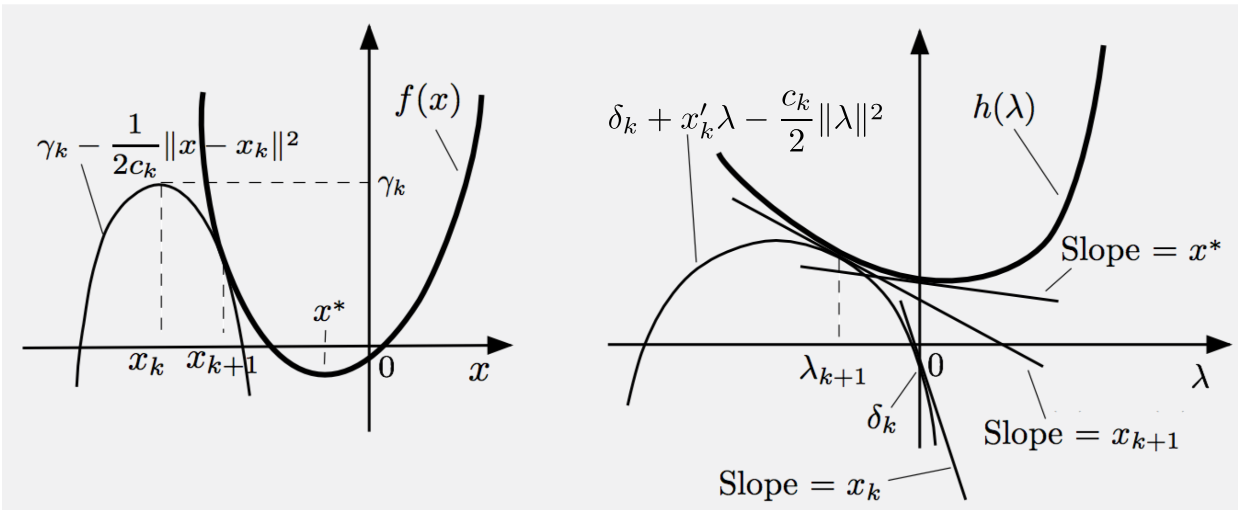
$$\lambda_{k+1} \in \partial f(x_{k+1}), \qquad x_{k+1} = x_k - c_k\lambda_{k+1}$$

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

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

- Aims to find a subgradient of h at 0: the limit of $\{x_k\}$.

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 h 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.

- Aims to find a subgradient of the primal function $p(u) = \min_{g(x)\leq u} f(x)$ at $u = 0$.

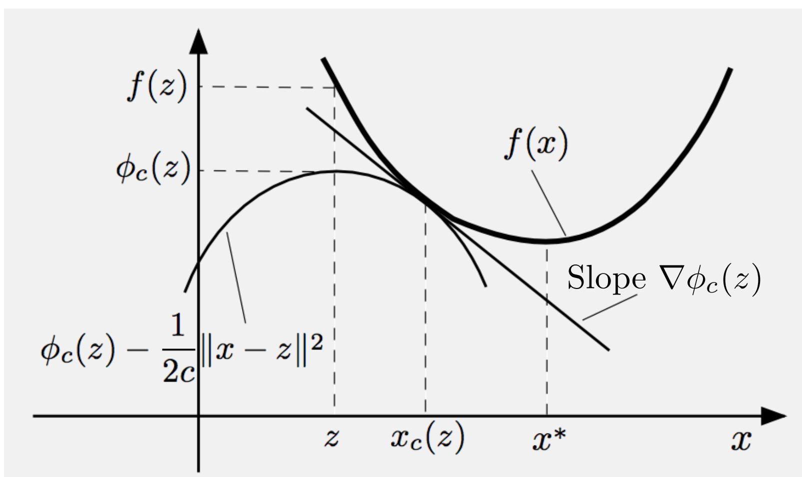
GRADIENT INTERPRETATION

- It can be shown that

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

where

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



- So the update $x_{k+1} = x_k - c_k\lambda_{k+1}$ can be viewed as a gradient iteration for minimizing $\phi_c(z)$ (it has the same minima as f).

- The gradient is calculated by the dual proximal minimization. Possibilities for faster methods (e.g., Newton, Quasi-Newton).

AUGMENTED LAGRANGIAN METHOD

- Consider the convex constrained problem

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

- Primal and dual functions:

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

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

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

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

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

- Implementation:

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

where L_c is the *Augmented Lagrangian* function

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

LECTURE 24

LECTURE OUTLINE

- Conic Programming
- Second Order Cone Programming

−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−

- Recall Fenchel duality framework:

$$\inf_{x\in\Re^n}\big\{f_1(x)-f_2(x)\big\}=\sup_{\lambda\in\Re^n}\big\{h_2(\lambda)-h_1(\lambda)\big\},$$

where

$$h_2(\lambda)=\inf_{z\in X_2}\big\{z'\lambda-f_2(z)\big\},$$

$$h_1(\lambda)=\sup_{y\in X_1}\big\{y'\lambda-f_1(y)\big\}.$$

- **Primal Fenchel Theorem**, under conditions on f_1, f_2, shows no duality gap, and existence of optimal solution of the dual problem.

- **Dual Fenchel Theorem**, under conditions on h_1, h_2, shows no duality gap, and existence of optimal solution of the primal problem.

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.
- **Second order cone programming.**
- **Semidefinite programming.**
- Convex programming.
  - Favorable special cases.
  - Quasi-convex programming.
  - Geometric programming.
- Nonlinear/nonconvex/continuous programming.
  - Favorable special cases.
  - Unconstrained.
  - Constrained.
- Discrete optimization/Integer programming
  - Favorable special cases.

CONIC DUALITY I

- Consider the problem

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

where C is a convex cone, and $f : \Re^n \mapsto (-\infty, \infty]$ is convex.

- Apply Fenchel duality with the definitions

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

We have

$$h_1(\lambda) = \sup_{x \in \Re^n} \big\{ \lambda' x - f(x) \big\},$$

$$h_2(\lambda) = \inf_{x \in C} x'\lambda = \begin{cases} 0 & \text{if } \lambda \in \hat{C}, \\ -\infty & \text{if } \lambda \notin \hat{C}, \end{cases}$$

where $\hat{C}$ is the negative polar cone (sometimes called the *dual cone* of C):

$$\hat{C} = -C^* = \{\lambda \mid x'\lambda \geq 0,\ \forall\ x \in C\}$$

CONIC DUALITY II

- Fenchel duality can be written as

$$\inf_{x \in C} f(x) = \sup_{\lambda \in \hat{C}} -h(\lambda),$$

where h is the conjugate of f.

- By the Primal Fenchel Theorem, there is no duality gap and the sup is attained if one of the following holds:

  (a) $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(C) \neq \emptyset$.

  (b) f can be extended to a real-valued convex function over $\Re^n$, and $\mathrm{dom}(f)$ and C are polyhedral.

- Similarly, by the Dual Fenchel Theorem, if f is closed and C is closed, there is no duality gap and the infimum in the primal problem is attained if one of the following two conditions holds:

  (a) $\mathrm{ri}(\mathrm{dom}(h)) \cap \mathrm{ri}(\hat{C}) \neq \emptyset$.

  (b) h can be extended to a real-valued convex function over $\Re^n$, and $\mathrm{dom}(h)$ and $\hat{C}$ are polyhedral.

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

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

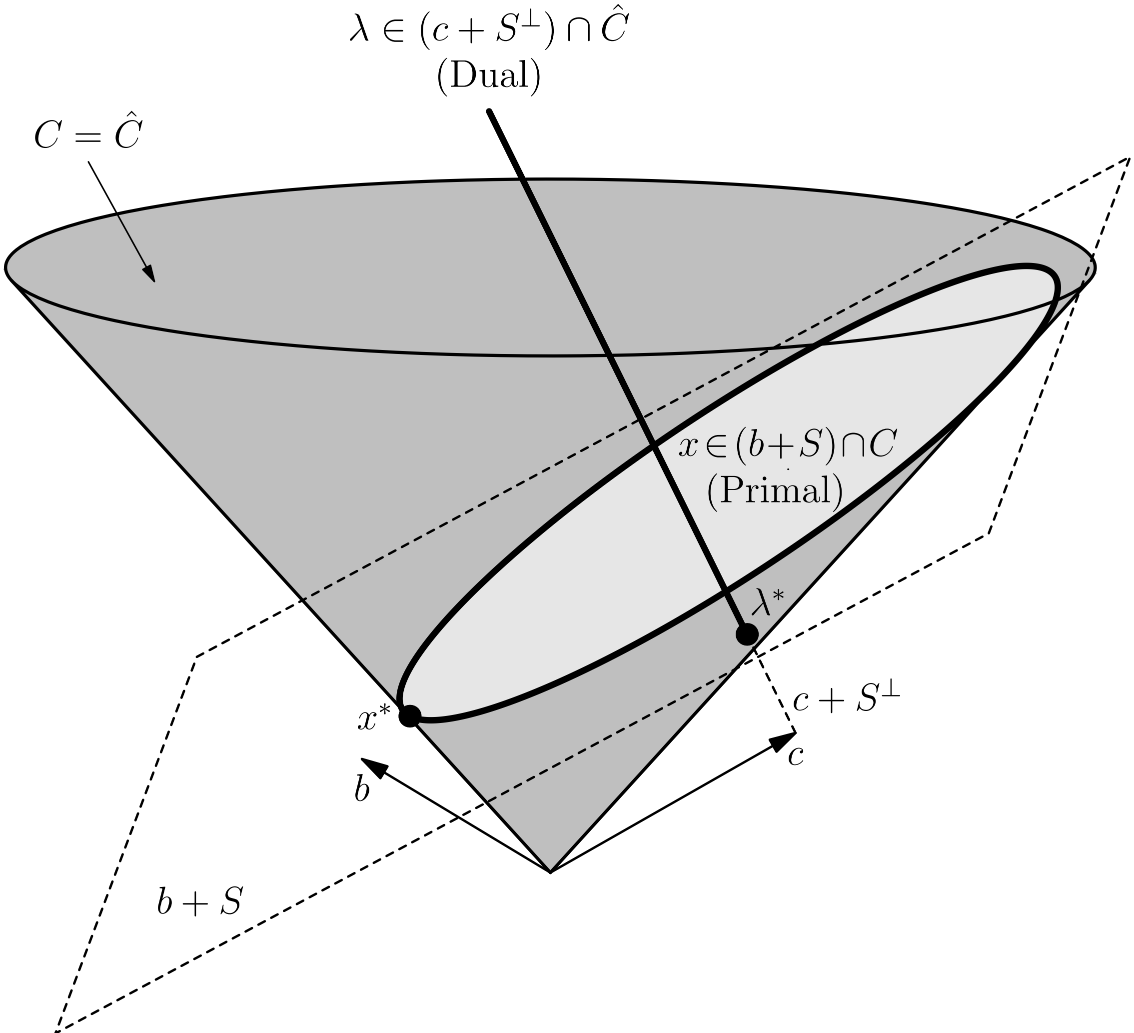
so the dual problem can be written as

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

- The primal and dual have the same form.

- If C is closed, the dual of the dual yields the primal.

VISUALIZATION OF LINEAR-CONIC PROBLEMS



Case where C is self-dual ($C = \hat{C}$).

CONES AND GENERALIZED INEQUALITIES

- Cones allow a shorthand expression of inequality constraints.

- **Example:** The constraint $Ax \geq b$ can be written as $z = Ax - b$ and $z \in C$, where C is the nonnegative orthant.

- **General Example:** For a closed convex cone C we have

$$x \in C \qquad \text{if and only if} \qquad y'x \leq 0, \ \forall\ y \in C^*$$

where C^* is the polar cone of C.

- **Generalized Inequalities:** Given a cone C, for two vectors $x, y \in \Re^n$, we write

$$x \succeq y \qquad \text{if} \qquad x - y \in C,$$

and for a function $g : \Re^m \mapsto \Re^n$, we write

$$g(x) \succeq 0 \qquad \text{if} \qquad g(x) \in C.$$

- **Desirable properties:** C closed, convex, and *pointed* in the sense that $C \cap (-C) = \{0\}$ (which implies that $x \succeq y,\ y \succeq x \ \Rightarrow\ x = y$).

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 corresponding generalized inequality is

$$x \succeq y \text{ if } x_n - y_n \geq \sqrt{(x_1 - y_1)^2 + \cdots + (x_{n-1} - y_{n-1})^2}.$$

- **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 D be the cone of matrices that are positive semidefinite. Then

$$X \succeq Y \qquad \text{if} \qquad X - Y \text{ is positive semidefinite.}$$

- All these cones are *self-dual*, i.e.,

$$C = -C^* = \hat{C}$$

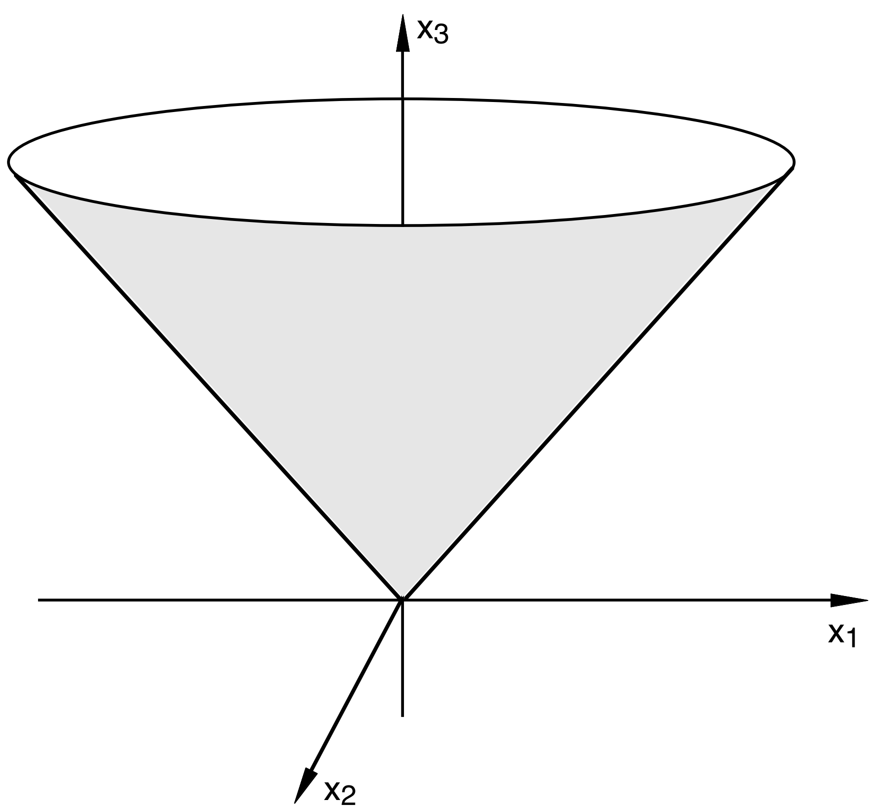
SECOND ORDER CONE PROGRAMMING

- 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

$$C_i: \;\text{the second order cone of } \Re^{n_i}$$



SECOND ORDER CONE DUALITY

- The dual of the second order cone problem (viewed as a special case of a linear-conic problem) is (after some manipulation)

$$\begin{array}{ll} \text{maximize} & \displaystyle\sum_{i=1}^{m} b_i'\lambda_i \\ \text{subject to} & \displaystyle\sum_{i=1}^{m} A_i'\lambda_i = c, \quad \lambda_i \in C_i,\ i = 1,\ldots,m, \end{array}$$

where $\lambda = (\lambda_1,\ldots,\lambda_m)$.

- The duality theory is derived from (and 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, second 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 25

LECTURE OUTLINE

- Special Cases of Fenchel Duality
- Semidefinite Programming
- Monotropic Programming

- Recall Fenchel duality framework:

$$\inf_{x\in\Re^n}\big\{f_1(x) - f_2(x)\big\} = \sup_{\lambda\in\Re^n}\big\{h_2(\lambda) - h_1(\lambda)\big\},$$

where

$$h_2(\lambda) = \inf_{z\in X_2}\big\{z'\lambda - f_2(z)\big\},$$

$$h_1(\lambda) = \sup_{y\in X_1}\big\{y'\lambda - f_1(y)\big\}.$$

- **Primal Fenchel Theorem**, under conditions on f_1, f_2, shows no duality gap, and existence of optimal solution of the dual problem.

- **Dual Fenchel Theorem**, under conditions on h_1, h_2, shows no duality gap, and existence of optimal solution of the primal problem.

LINEAR-CONIC PROBLEMS

- Let f_1 be affine, $f_1(x) = c'x$, with $\mathrm{dom}(f)$ being an affine set, $\mathrm{dom}(f) = b + S$, where S is a subspace. Let $-f_2$ be the indicator function of a cone C, with dual cone denoted $\hat{C}$.

- 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 of f_1 is

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

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.

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 D be the cone of pos. semidefinite matrices. Note that D is self-dual [$D = \hat{D}$, i.e., $< X, Y > \geq 0$ for all $Y \in D$ iff $X \in D$], and its interior is the set of pos. definite matrices.

- Fix symmetric matrices C, $A_1, \ldots, A_m$, and vectors $b_1, \ldots, b_m$, and consider

$$\begin{aligned} &\text{minimize} \quad < C, X > \\ &\text{subject to} \quad < A_i, X >= b_i, \quad i = 1, \ldots, m, \quad X \in D \end{aligned}$$

- Viewing this as an affine cost conic problem, the dual problem (after some manipulation) is

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} b_i \lambda_i \\ &\text{subject to} \quad C - (\lambda_1 A_1 + \cdots + \lambda_m A_m) \in D \end{aligned}$$

- There is no duality gap if there exists $\overline{\lambda}$ such that $C - (\overline{\lambda}_1 A_1 + \cdots + \overline{\lambda}_m A_m)$ is pos. definite.

EXAMPLE: MINIMIZE THE MAXIMUM EIGENVALUE

- Given $n \times n$ 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

$$\begin{array}{ll} \text{minimize} & z \\ \text{subject to} & zI - M(\lambda) \in D, \end{array}$$

where I is the $n \times n$ identity matrix, and D is the semidefinite cone.

- If $M(\lambda)$ is an affine function of λ,

$$M(\lambda) = C + \lambda_1 M_1 + \cdots + \lambda_m M_m,$$

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 ...

EXTENDED MONOTROPIC PROGRAMMING

- Let
  - $x = (x_1, \ldots, x_m)$ with $x_i \in \Re^{n_i}$
  - $f_i : \Re^{n_i} \mapsto (-\infty, \infty]$ is closed proper convex
  - S is a subspace of $\Re^{n_1+\cdots+n_m}$
- Extended monotropic programming problem:

$$
\begin{aligned}
&\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\
&\text{subject to} \;\; x \in S
\end{aligned}
$$

- Monotropic programming is the special case where each x_i is 1-dimensional.
- Models many important optimization problems (linear, quadratic, convex network, etc).
- Has a powerful symmetric duality theory.

DUALITY

- Convert to the equivalent form

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(z_i) \\ &\text{subject to} \;\; z_i = x_i, \quad i = 1, \ldots, m, \qquad x \in S \end{aligned}$$

- Assigning a multiplier vector $\lambda_i \in \Re^{n_i}$ to the constraint $z_i = x_i$, the dual function is

$$\begin{aligned} q(\lambda) &= \inf_{x \in S} \lambda' x + \sum_{i=1}^{m} \inf_{z_i \in \Re^{n_i}} \big\{ f_i(z_i) - \lambda_i' z_i \big\} \\ &= \begin{cases} \sum_{i=1}^{m} q_i(\lambda_i) & \text{if } \lambda \in S^\perp, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where $q_i(\lambda_i) = \inf_{z_i \in \Re} \big\{ f_i(z_i) - \lambda_i' z_i \big\}$.

- The dual problem is the (symmetric) monotropic program

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} q_i(\lambda_i) \\ &\text{subject to} \;\; \lambda \in S^\perp \end{aligned}$$

OPTIMALITY CONDITIONS

- Assume that $-\infty < q^* = f^* < \infty$. Then (x^*, λ^*) are optimal primal and dual solution pair if and only if

$$x^* \in S, \qquad \lambda^* \in S^\perp, \qquad \lambda_i^* \in \partial f_i(x_i^*), \quad \forall\ i$$

- **Specialization to the monotropic case ($n_i = 1$ for all i):** The vectors x^* and λ^* are optimal primal and dual solution pair if and only if

$$x^* \in S, \qquad \lambda^* \in S^\perp, \qquad (x_i^*, \lambda_i^*) \in \Gamma_i, \ \ \forall\ i$$

where

$$\Gamma_i = \big\{(x_i, \lambda_i) \mid x_i \in \mathrm{dom}(f_i),\ f_i^-(x_i) \le \lambda_i \le f_i^+(x_i)\big\}$$

- Interesting application of these conditions to electrical networks.

STRONG DUALITY THEOREM

- Assume that the extended monotropic programming problem is feasible, and that for all feasible solutions x, the set

$$S^\perp + \partial_\epsilon D_{1,\epsilon}(x) + \cdots + D_{m,\epsilon}(x)$$

is closed for all $\epsilon > 0$, where

$$D_{i,\epsilon}(x) = \big\{(0,\ldots,0,\lambda_i,0,\ldots,0) \mid \lambda_i \in \partial_\epsilon f_i(x_i)\big\}$$

Then $q^* = f^*$.

- An unusual duality condition. It is satisfied if each set $\partial_\epsilon f_i(x)$ is either compact or polyhedral. Proof is also unusual - uses the ϵ-descent method!

- **Monotropic programming case:** If $n_i = 1$, $D_{i,\epsilon}(x)$ is an interval, so it is polyhedral, and $q^* = f^*$.

- There are some other cases of interest. See Chapter 8.

- The monotropic duality result extends to convex separable problems with *nonlinear* constraints. (Hard to prove ...)

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

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \;\; x \in X, \qquad g(x) \leq 0, \quad i = 1, \ldots, r, \end{aligned}$$

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

$$\begin{aligned} &\text{minimize} \quad f(x) + P\big(g(x)\big) \\ &\text{subject to} \;\; x \in X, \end{aligned}$$

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

$$\inf_{u\in\Re^r}\big\{p(u)+P(u)\big\} = \sup_{\mu\ge 0}\big\{q(\mu)-Q(\mu)\big\},$$

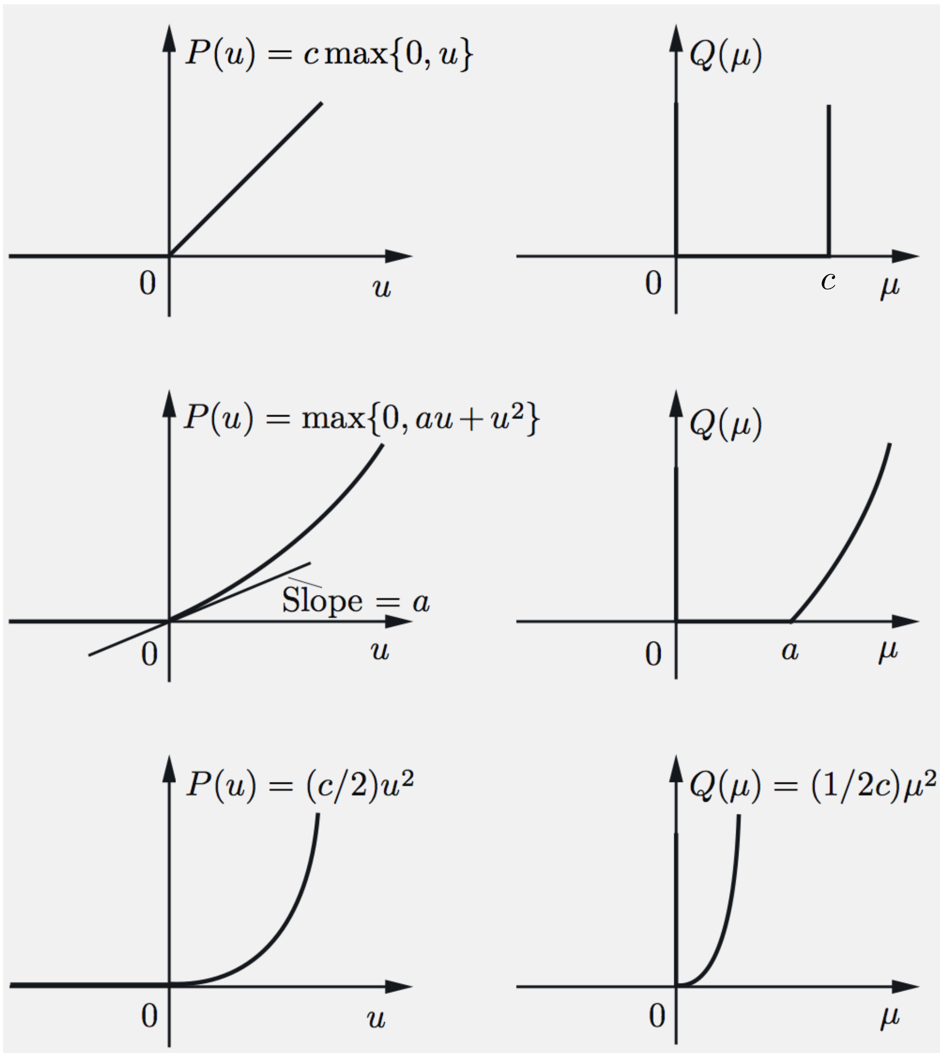
where

$$q(\mu) = \inf_{x\in X}\big\{f(x)+\mu' g(x)\big\}$$

is the dual function, and Q is the conjugate convex function of P:

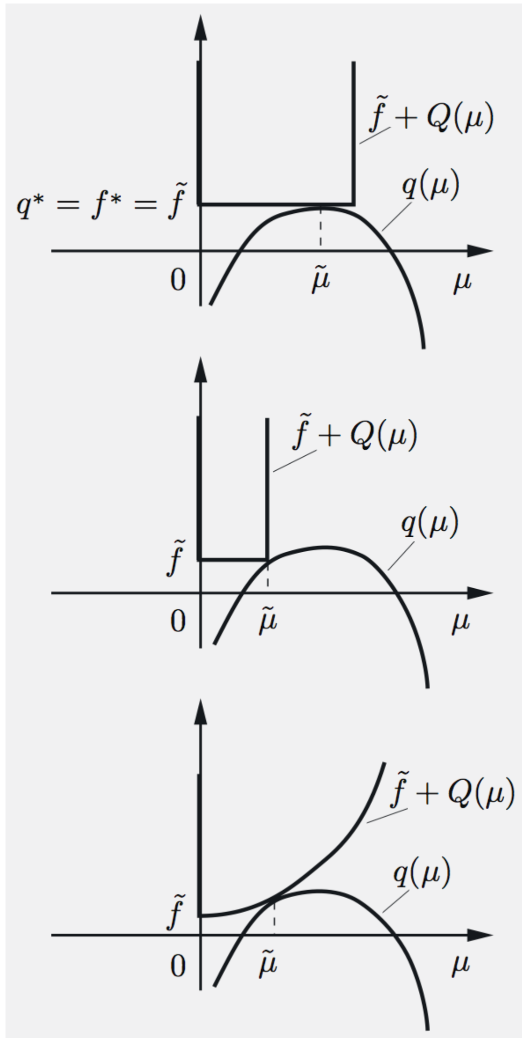
$$Q(\mu) = \sup_{u\in\Re^r}\big\{u'\mu - P(u)\big\}$$

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^*.