LECTURE SLIDES ON

CONVEX ANALYSIS AND OPTIMIZATION

BASED ON LECTURES GIVEN AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

CAMBRIDGE, MASS

BY DIMITRI P. BERTSEKAS

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

Last updated: 10/12/2005

These lecture slides are based on the author's book: "Convex Analysis and Optimization," Athena Scientific, 2003; see

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

LECTURE 1
AN INTRODUCTION TO THE COURSE

LECTURE OUTLINE

- Convex and Nonconvex Optimization Problems
- Why is Convexity Important in Optimization
- Lagrange Multipliers and 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.,

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

- 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 IN OPTIMIZATION?

- 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 $g_1(x) \le 0, \ldots, g_r(x) \le 0$

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

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

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 $\mu \ge 0$

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).

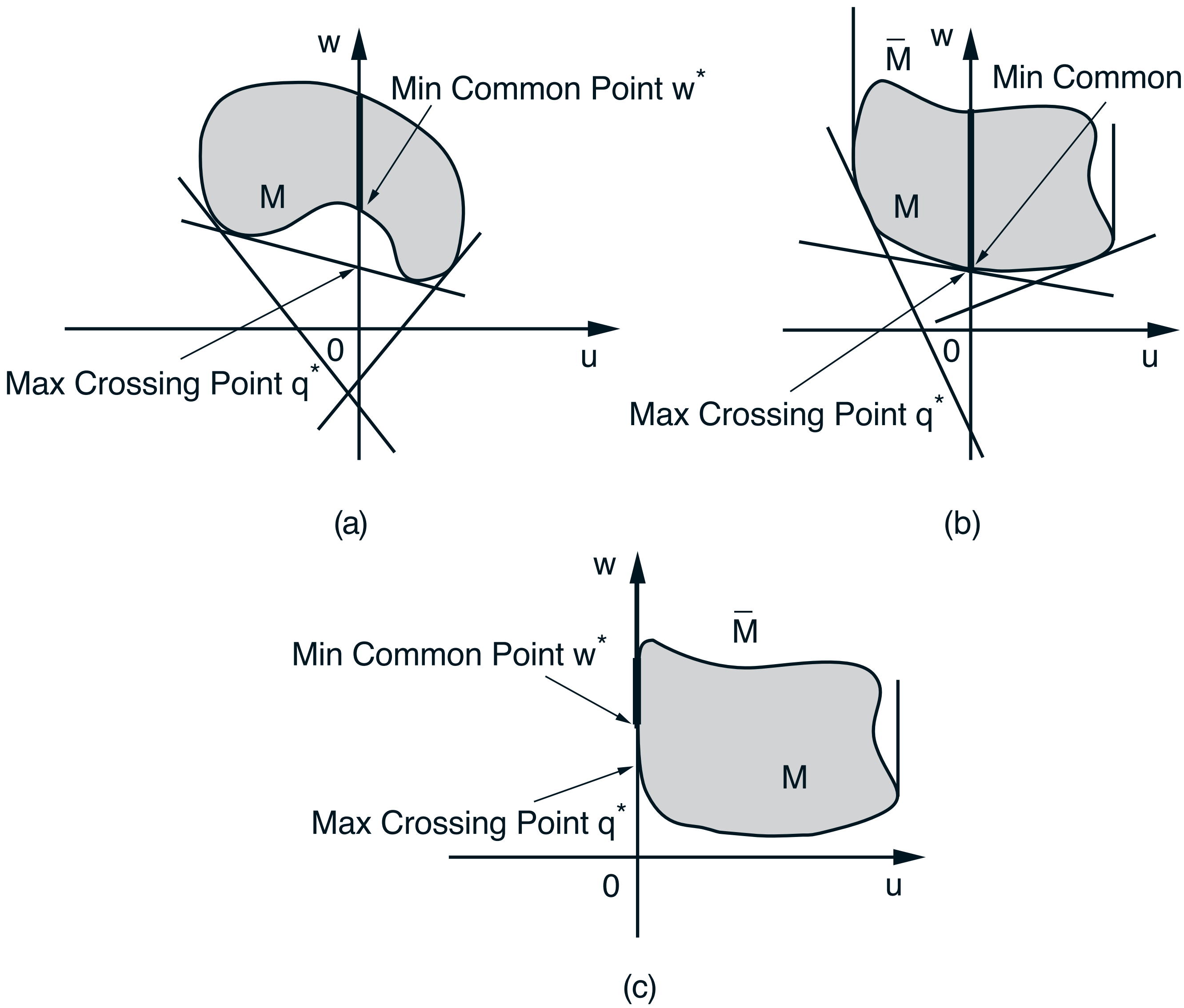
• Under favorable circumstances (convexity in the primal problem, plus ...):

 – 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

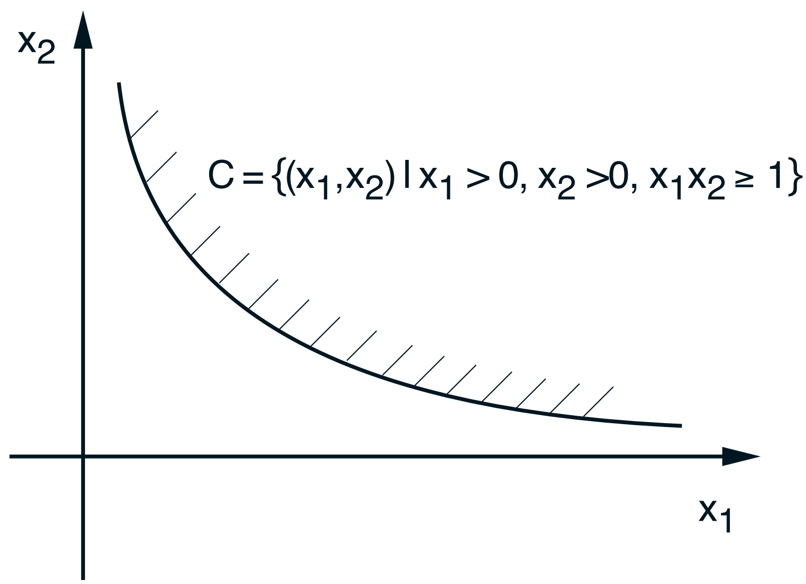


- All of duality theory and all of (convex/concave) minimax theory can be developed/explained in terms of this one figure.

- 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. Recession cones.

2) **Convexity and Optimization (4):** Directions of recession and existence of optimal solutions. Hyperplanes. Min common/max crossing duality. Saddle points and minimax theory.

3) **Polyhedral Convexity (3):** Polyhedral sets. Extreme points. Polyhedral aspects of optimization. Polyhedral aspects of duality.

4) **Subgradients (3):** Subgradients. Conical approximations. Optimality conditions.

5) **Lagrange Multipliers (3):** Fritz John theory. Pseudonormality and constraint qualifications.

6) **Lagrangian Duality (3):** Constrained optimization duality. Linear and quadratic programming duality. Duality theorems.

7) **Conjugate Duality (3):** Fenchel duality theorem. Conic and semidefinite programming. Exact penalty functions.

8) **Dual Computational Methods (3):** Classical subgradient and cutting plane methods. Application in Lagrangian relaxation and combinatorial optimization.

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 (see 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 Analysis" textbook

LECTURE 2

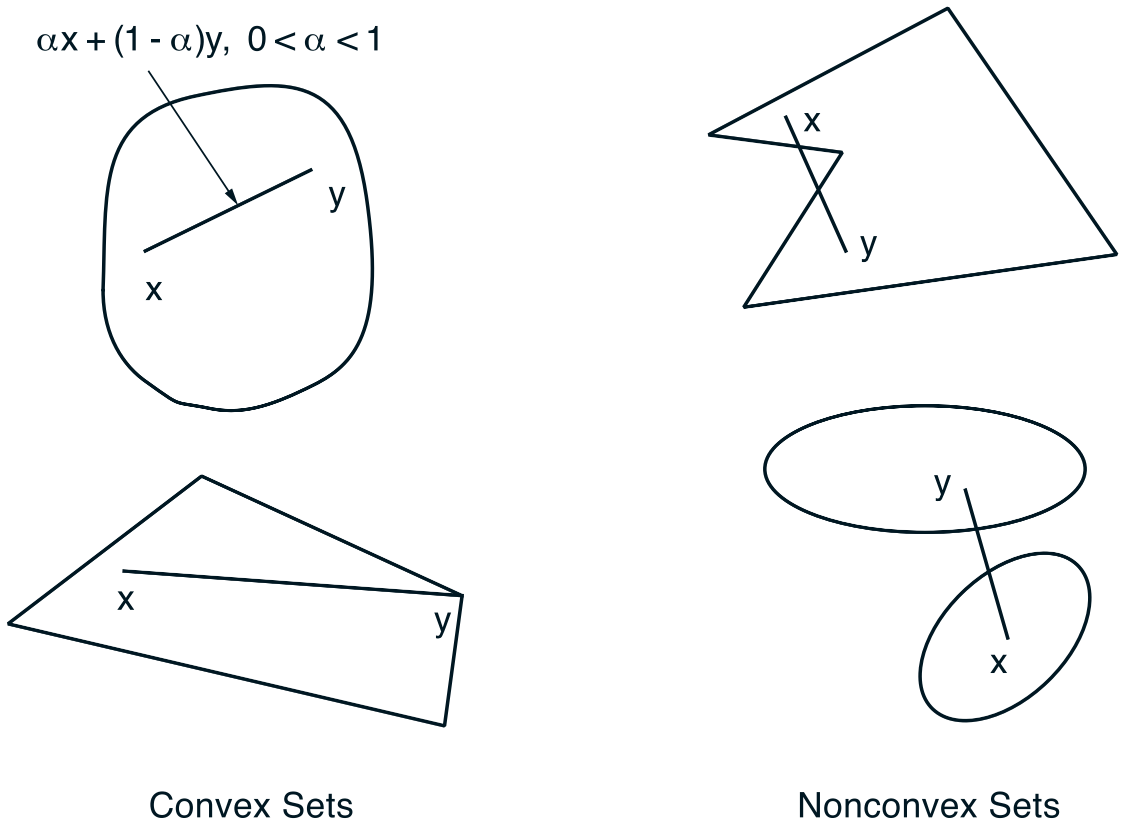
LECTURE OUTLINE

- 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 Section 1.1 of the textbook for an overview of the linear algebra and real analysis background that we will use

CONVEX SETS



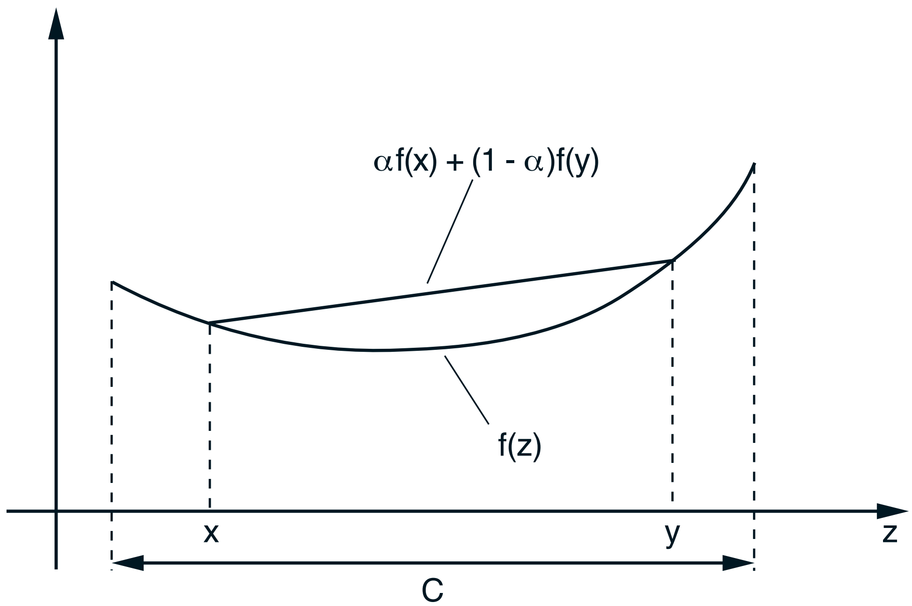
- A subset C of $\Re^n$ is called *convex* if

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

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

- 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

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

• The *effective domain* of f is the set

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

• 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 $\mathrm{epi}(f)$ is a closed set.

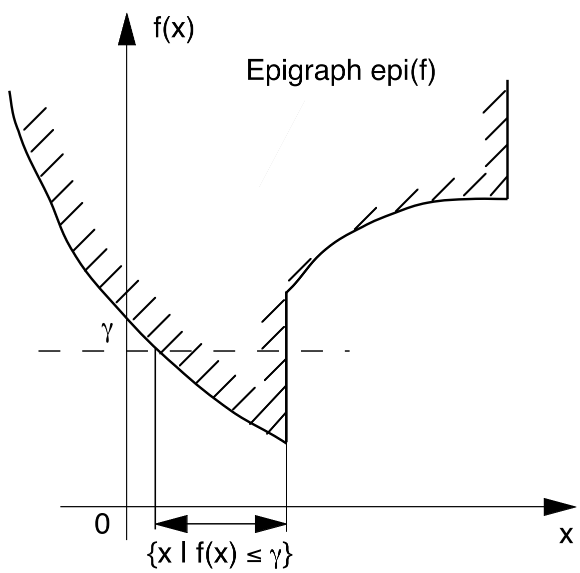
CLOSEDNESS AND SEMICONTINUITY

• *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.



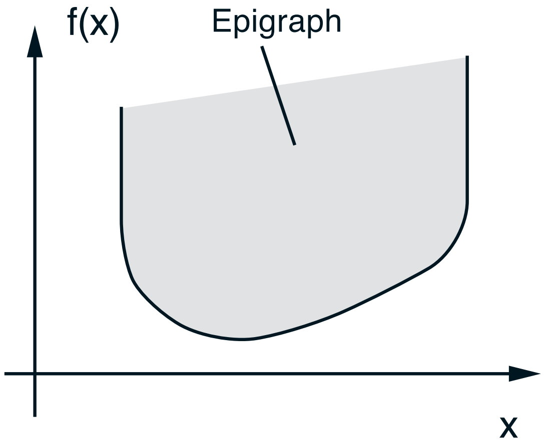
• Note that:

– If f is lower semicontinuous at all $x \in \mathrm{dom}(f)$, it is not necessarily closed

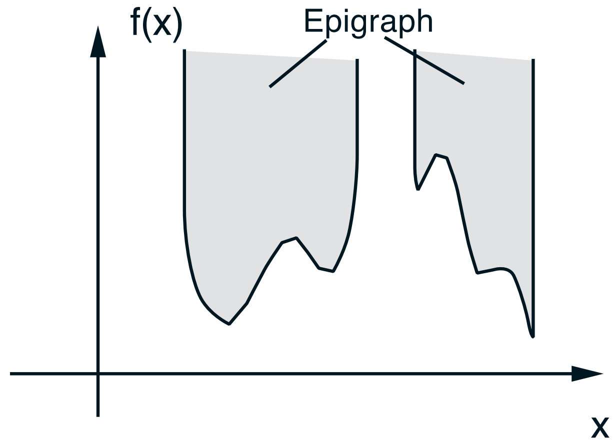
– If f is closed, $\mathrm{dom}(f)$ is not necessarily closed

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

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

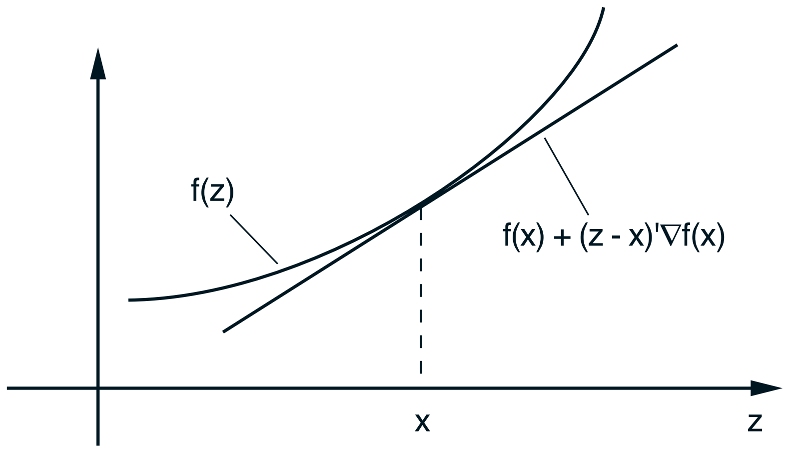
is convex (or closed) if the f_i are convex (respectively, closed).

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 if and only if

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

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

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

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 $\sum_{i=1}^{m} \alpha_i = 1$.

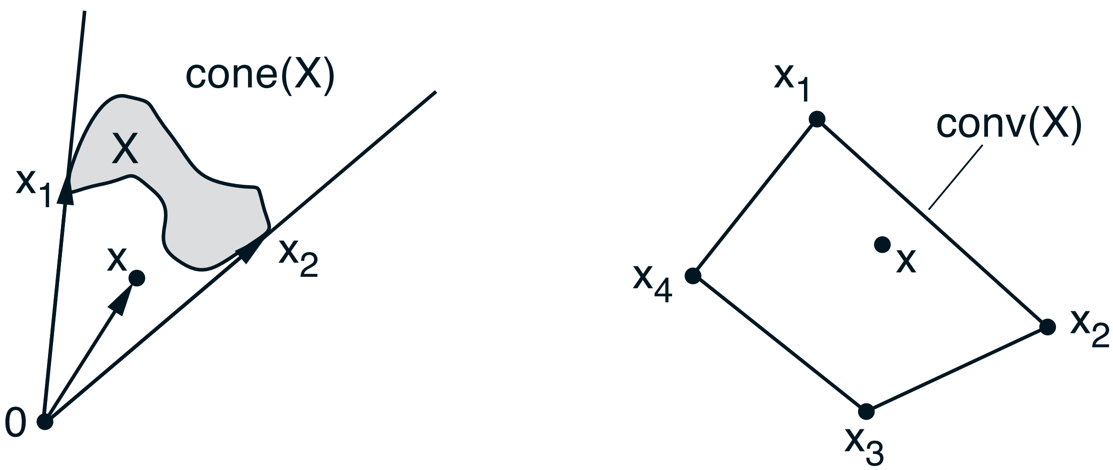
- The *convex hull* of X, denoted $\text{conv}(X)$, is the intersection of all convex sets containing X (also the set of all convex combinations from X).

- The *affine hull* of X, denoted $\text{aff}(X)$, is the intersection of all affine sets containing X (an affine set is a set of the form $\overline{x} + S$, where S is a subspace). Note that $\text{aff}(X)$ 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 $\text{cone}(X)$, is the set of all nonnegative combinations from X:
  - It is a convex cone containing the origin.
  - It need not be closed.
  - If X is a finite set, $\text{cone}(X)$ is closed (nontrivial to show!)

CARATHEODORY'S THEOREM



(a) (b)

- Let X be a nonempty subset of $\Re^n$.

  (a) Every $x \neq 0$ in $\mathrm{cone}(X)$ 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 $\mathrm{conv}(X)$ 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 $\mathrm{cone}(X)$, 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 $\Re^{n+1}$

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

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

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

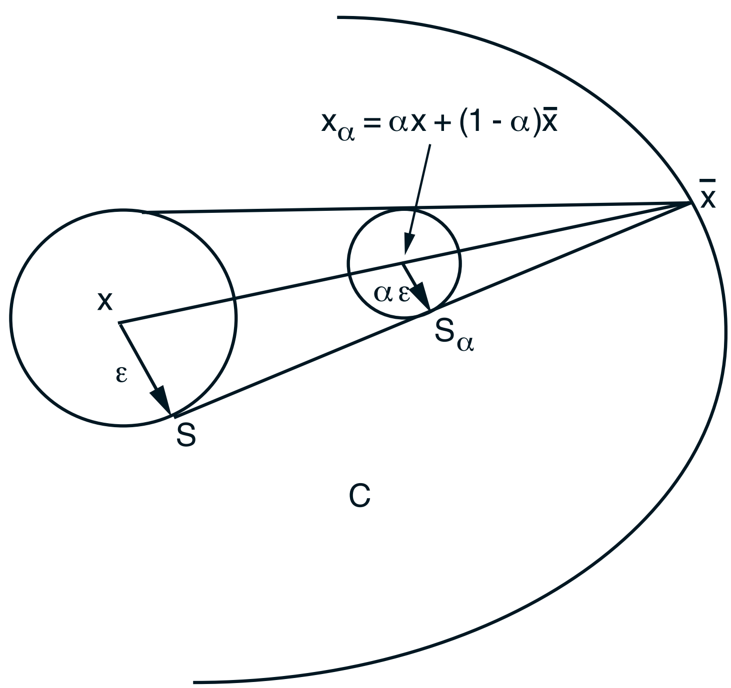
which must satisfy $\sum_{i=1}^{n+1} \alpha_i = 1$, and $\alpha_i \geq 0$, $x_i \in X$ for all i. Thus, the vector $\sum_{i=1}^{n+1} \alpha_i x_i$, which belongs to $\mathrm{conv}(X)$, is a limit point of the sequence $\left\{\sum_{i=1}^{n+1} \alpha_i^k x_i^k\right\}$, showing that $\mathrm{conv}(X)$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

- $\mathrm{ri}(C)$ 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, $x \in \mathrm{ri}(C)$ and $\overline{x} \in \mathrm{cl}(C)$, then all points on the line segment connecting x and $\overline{x}$, except possibly $\overline{x}$, belong to $\mathrm{ri}(C)$.

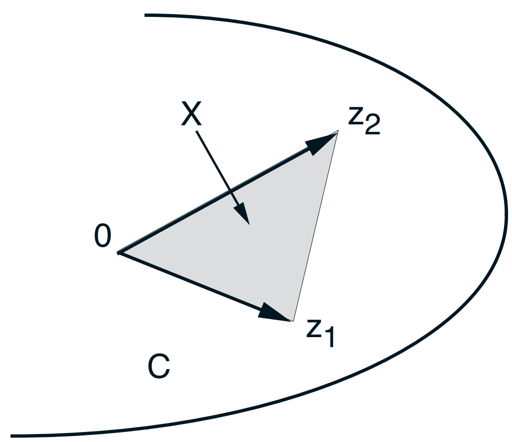


ADDITIONAL MAJOR RESULTS

- 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.



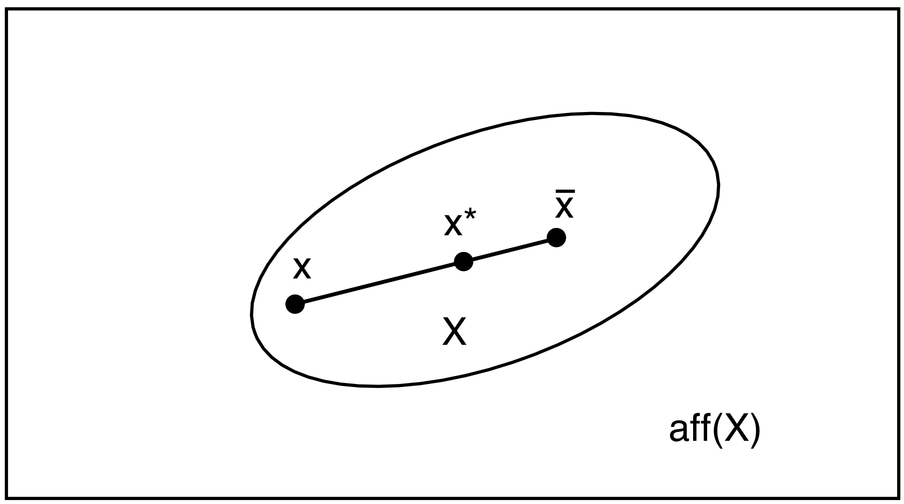
Proof: (a) Assume that $0 \in C$. We choose m linearly independent vectors $z_1, \ldots, z_m \in C$, where m is the dimension of $\mathrm{aff}(C)$, and we let

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

(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
- Recession cones

****

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

- Three important properties of $\mathrm{ri}(C)$ of a convex set C:
  - $\mathrm{ri}(C)$ is nonempty
  - *Line Segment Principle*: If $x \in \mathrm{ri}(C)$ and $\overline{x} \in \mathrm{cl}(C)$, then all points on the line segment connecting x and $\overline{x}$, except possibly $\overline{x}$, belong to $\mathrm{ri}(C)$
  - *Prolongation Principle*: If $x \in \mathrm{ri}(C)$ and $\overline{x} \in C$, the line segment connecting $\overline{x}$ and x can be prolonged beyond x without leaving C

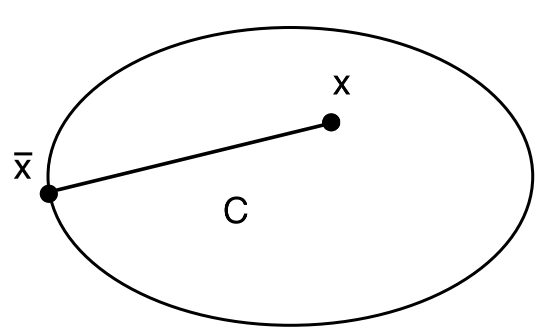
A SUMMARY OF FACTS

- The closure of a convex set is equal to the closure of its relative interior.
- The relative interior of a convex set is equal to the relative interior of 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 closure nor relative interior 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 $A \cdot \text{ri}(C) = \text{ri}(A \cdot C)$.

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

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

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

To show the converse, assuming that C is bounded, choose any $z \in \text{cl}(A \cdot C)$. 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 $x \in \text{cl}(C)$, and we must have $Ax = z$. It follows that $z \in A \cdot \text{cl}(C)$. **Q.E.D.**

Note that in general, we may have

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

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

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

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

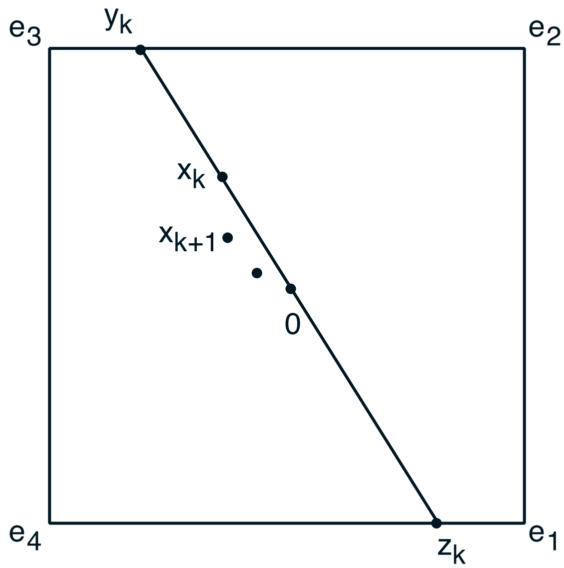
Proof of (a): $C_1 + C_2$ is the result of the linear transformation $(x_1, x_2) \mapsto x_1 + x_2$.

- 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

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

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

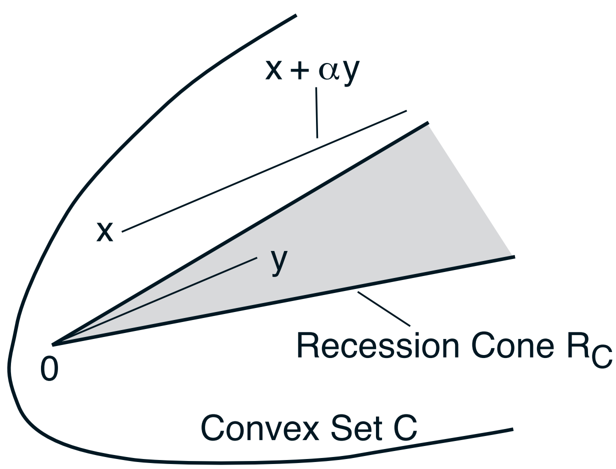
Since $\|x_k\|_\infty \to 0$, $f(x_k) \to f(0)$. **Q.E.D.**

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

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.

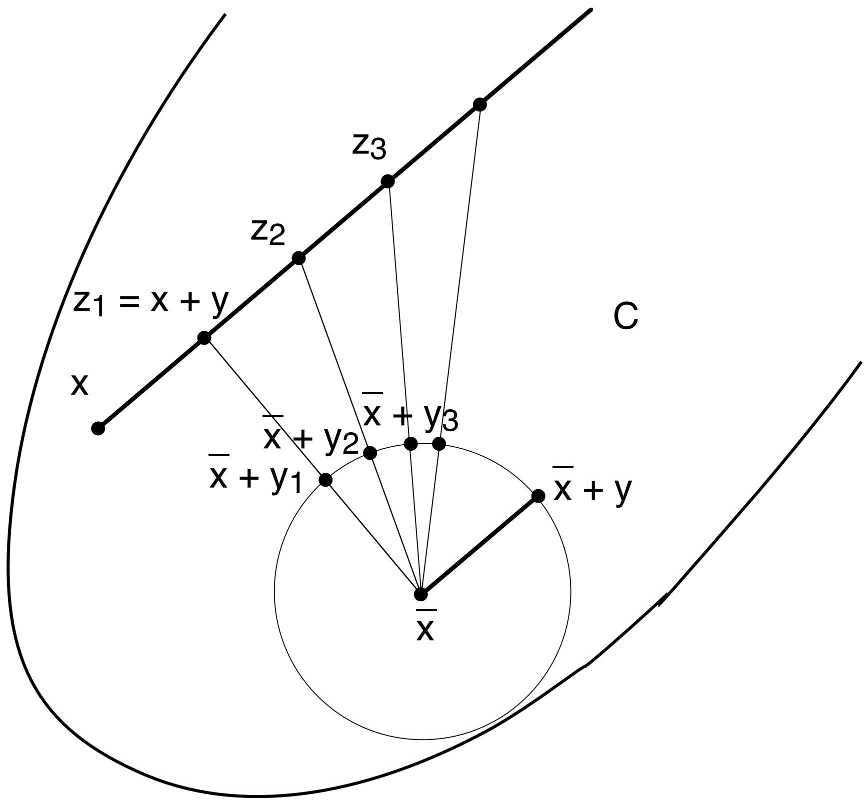
  (e) If D is another closed convex set such that $C \cap D \neq \emptyset$, we have

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

  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

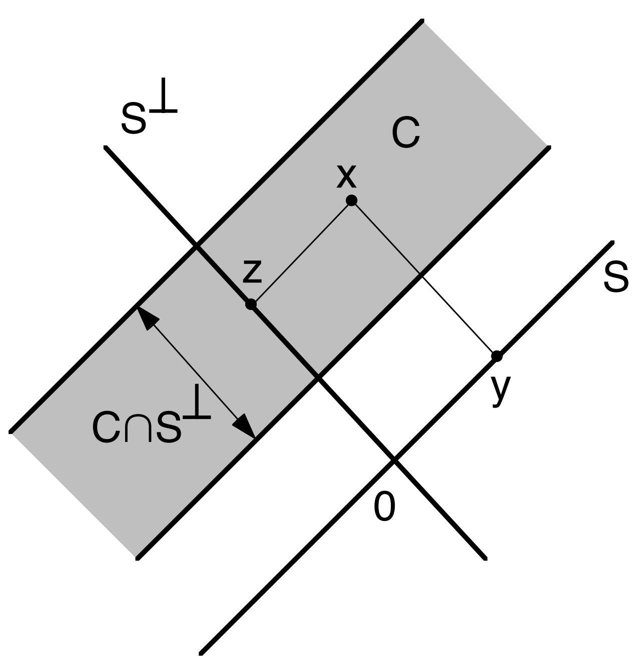
- The *lineality space* of a convex set C, denoted by L_C, is the subspace of vectors y such that $y \in R_C$ and $-y \in R_C$:

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

- *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).



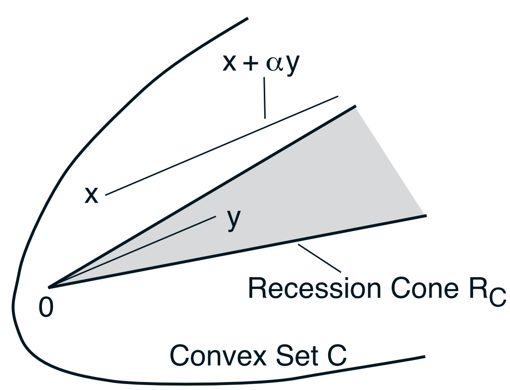
LECTURE 5

LECTURE OUTLINE

- Directions of recession of convex functions
- Existence of optimal solutions - Weierstrass' theorem
- Intersection of nested sequences of closed sets
- Asymptotic directions

– –

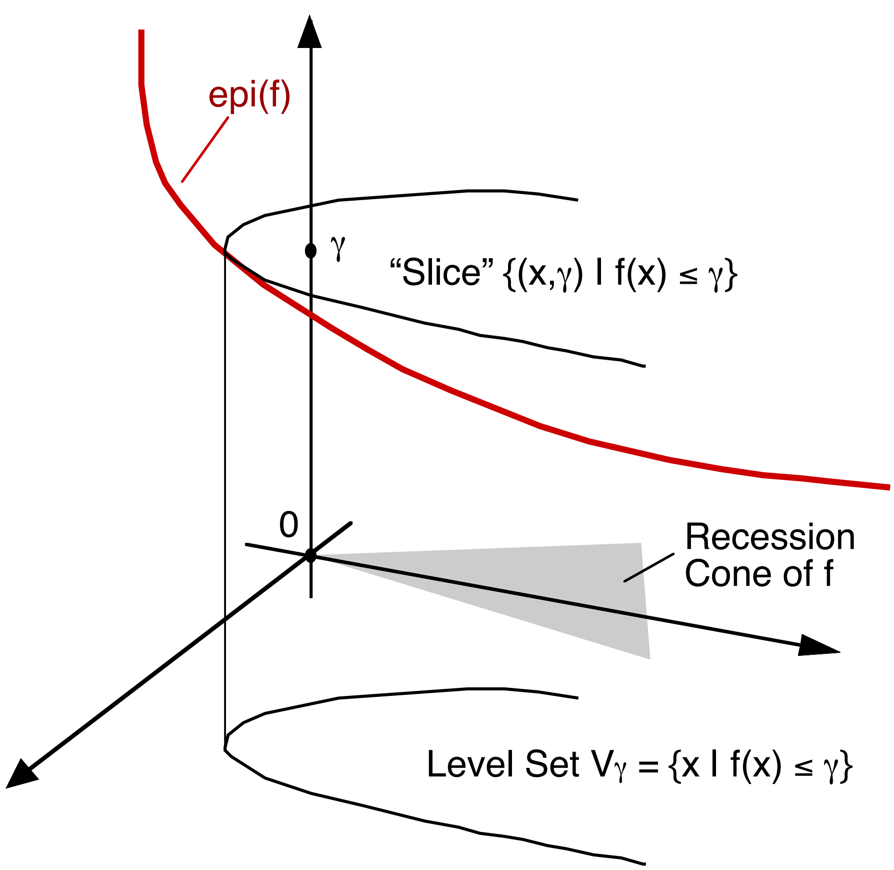
- For a closed convex set C, recall that y is a $direction\ of\ recession$ if $x + \alpha y \in C$, for all $x \in C$ and $\alpha \geq 0$.



- Recession cone theorem: If this property is true for one $x \in C$, it is true for all $x \in C$; also C is compact iff $R_C = \{0\}$.

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 $V_\gamma = \{x \mid f(x) \leq \gamma\}$, where γ is a scalar. Then:

  (a) All the nonempty level sets V_γ have the same recession cone, given by

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

  (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,

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

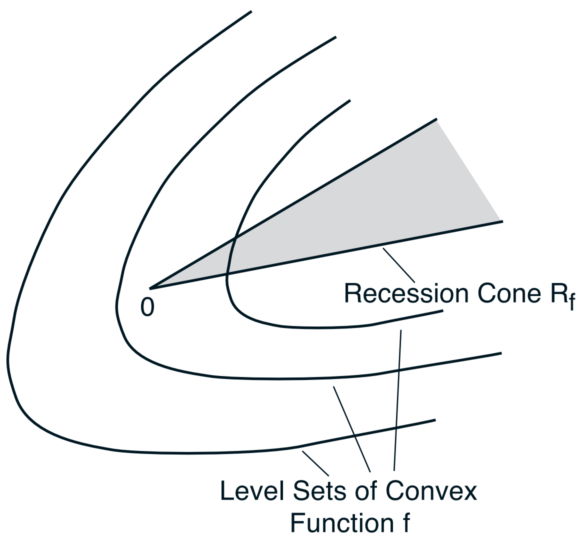
The recession cone of the set on the left is $\{(y, 0) \mid y \in R_{V_\gamma}\}$. The recession cone of the set on the right is the intersection of $R_{\mathrm{epi}}(f)$ and the recession cone of $\{(x, \gamma) \mid x \in \Re^n\}$. Thus we have

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

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 $V_\gamma = \{x \mid f(x) \leq \gamma\}$, $\gamma \in \Re$, is the *recession cone of f*, and is denoted by R_f.



• Terminology:

– $y \in R_f$: a *direction of recession of f*.

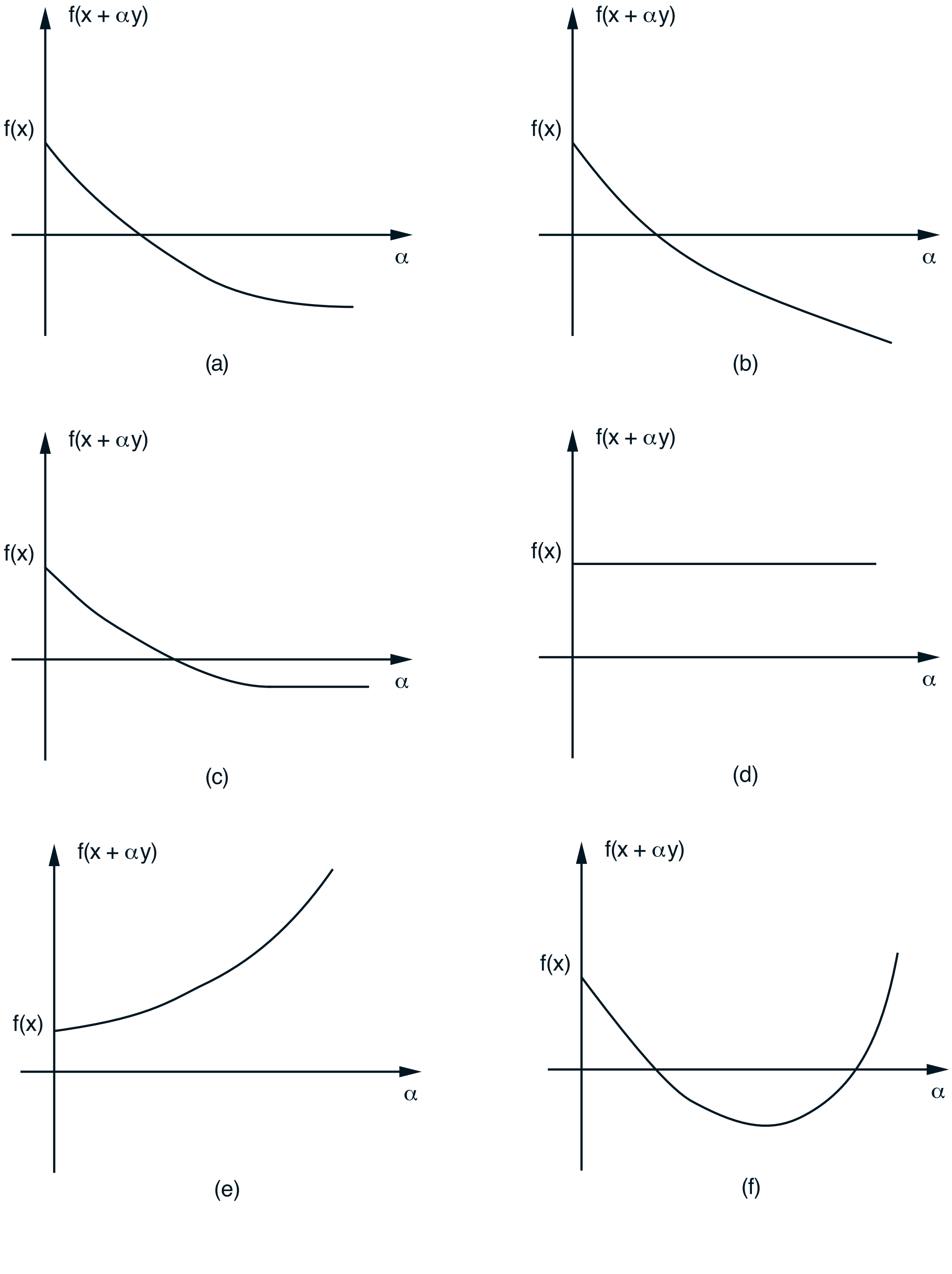
– $L_f = R_f \cap (-R_f)$: the *lineality space of f*.

– $y \in L_f$: 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 $r_f(y) \leq 0$.

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

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

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

Conversely, we have

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

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 OF THE IDEA

- **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

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

- The set intersection/compactness argument generalizes to nonconvex.

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

(1) X is bounded.

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

(3) $\tilde{f}$ is coercive, i.e., for every sequence $\{x_k\} \subset X$ s. t. $\|x_k\| \to \infty$, we have $\lim_{k\to\infty} f(x_k) = \infty$.

Proof: In all cases the level sets of $\tilde{f}$ are compact. **Q.E.D.**

THE ROLE OF CLOSED SET INTERSECTIONS

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

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

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

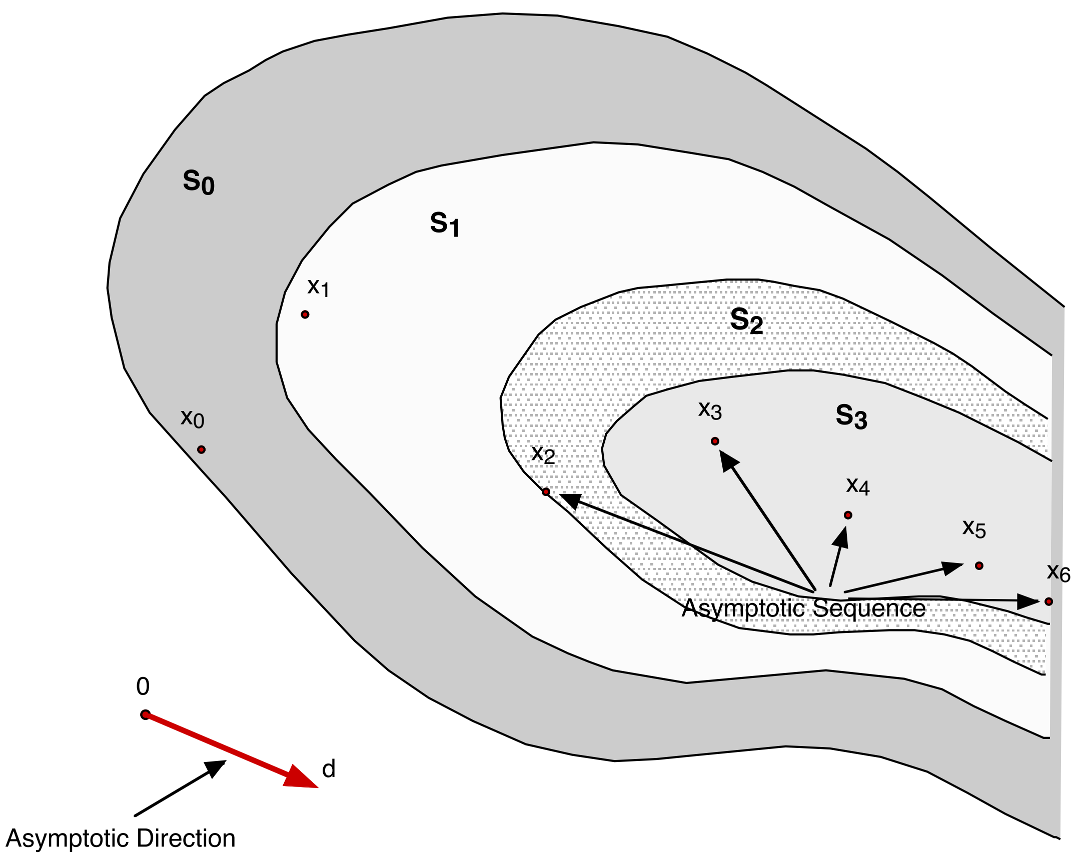
ASYMPTOTIC DIRECTIONS

- Given a sequence of nonempty nested closed sets $\{S_k\}$, we say that a vector $d \neq 0$ is an *asymptotic direction of* $\{S_k\}$ if there exists $\{x_k\}$ s. t.

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

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

- A sequence $\{x_k\}$ associated with an asymptotic direction d as above is called an *asymptotic sequence corresponding to* d.



CONNECTION WITH RECESSION CONES

- We say that d is an *asymptotic direction of a nonempty closed set* S if it is an asymptotic direction of the sequence $\{S_k\}$, where $S_k = S$ for all k.

- **Notation:** The set of asymptotic directions of S is denoted A_S.

- **Important facts:**
  - The set of asymptotic directions of a closed set sequence $\{S_k\}$ is

$$\cap_{k=0}^{\infty} A_{S_k}$$

  - For a closed *convex* set S

$$A_S = R_S \setminus \{0\}$$

  - The set of asymptotic directions of a closed *convex* set sequence $\{S_k\}$ is

$$\cap_{k=0}^{\infty} R_{S_k} \setminus \{0\}$$

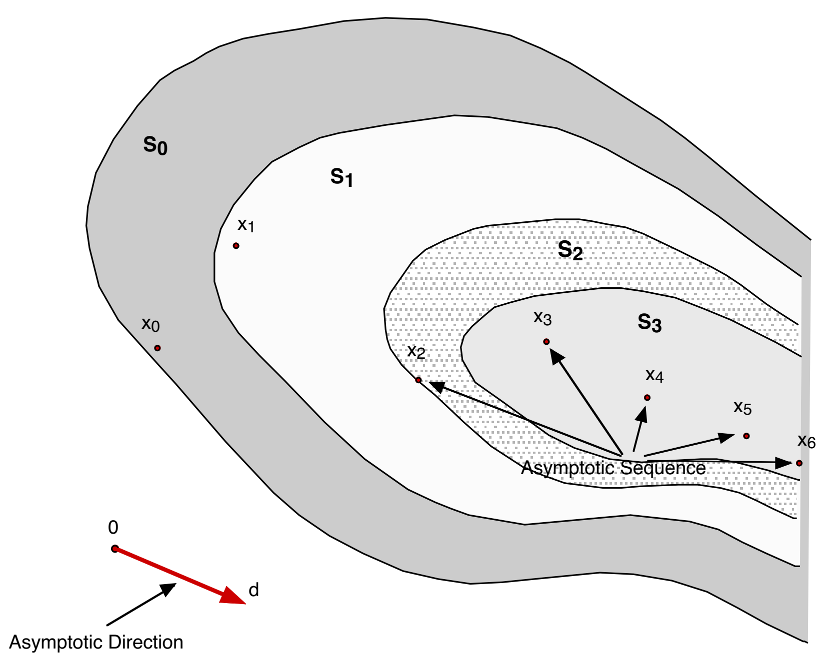
LECTURE 6

LECTURE OUTLINE

- Asymptotic directions that are retractive
- Nonemptiness of closed set intersections
- Frank-Wolfe Theorem
- Horizon directions
- Existence of optimal solutions
- Preservation of closure under linear transformation and partial minimization

---

Asymptotic directions of a closed set sequence

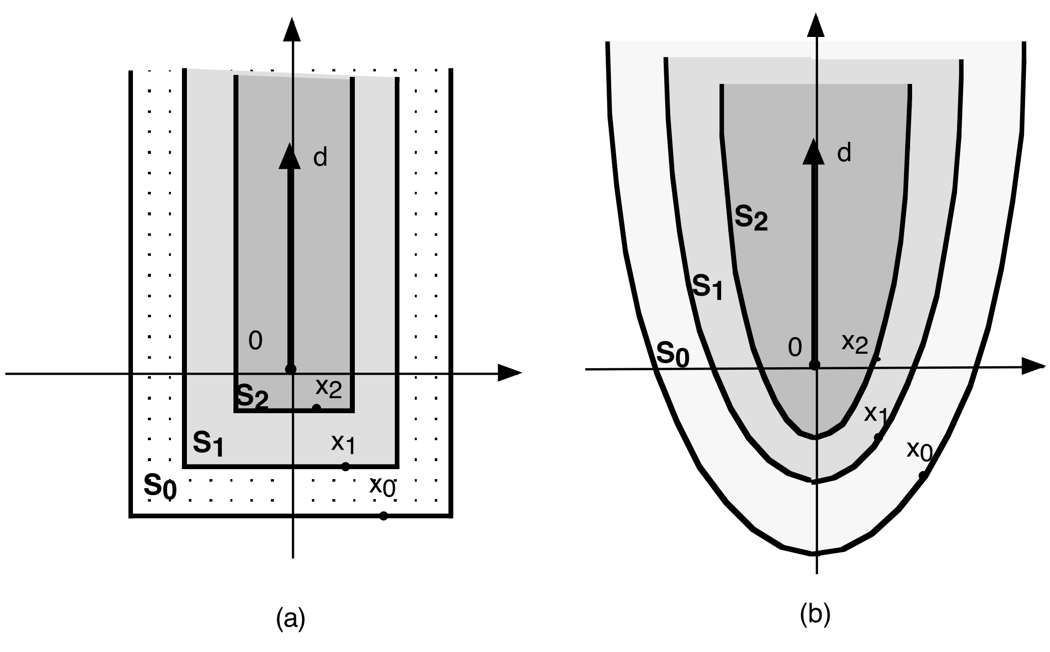


RETRACTIVE ASYMPTOTIC DIRECTIONS

- Consider a nested closed set sequence $\{S_k\}$.
- An asymptotic direction d is called *retractive* if for every asymptotic sequence $\{x_k\}$ there exists an index $\overline{k}$ such that

$$x_k - d \in S_k, \qquad \forall\, k \geq \overline{k}.$$

- $\{S_k\}$ is called *retractive* if all its asymptotic directions are retractive.
- These definitions specialize to closed convex sets S by taking $S_k \equiv S$.

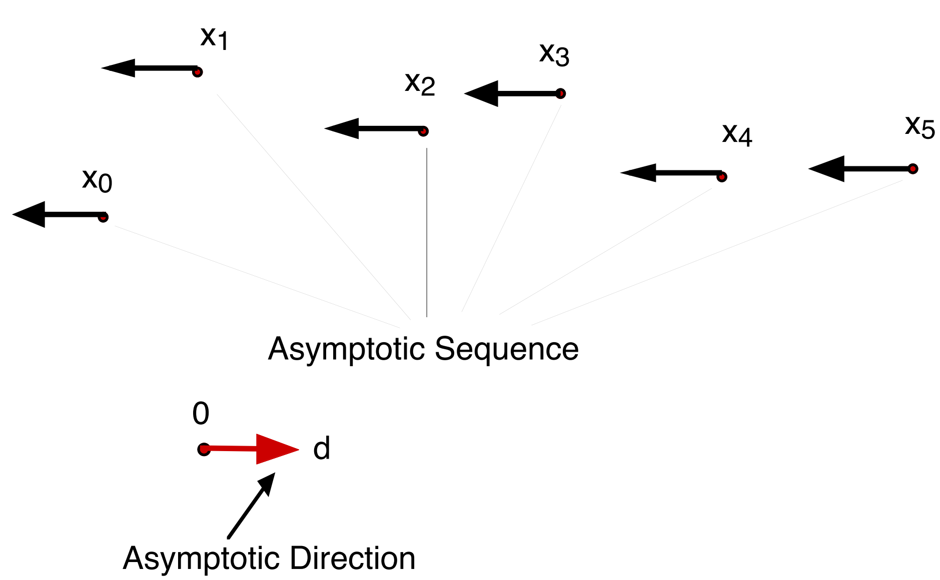


SET INTERSECTION THEOREM

- If $\{S_k\}$ is retractive, then $\cap_{k=0}^{\infty} S_k$ is nonempty.
- Key proof ideas:

  (a) The intersection $\cap_{k=0}^{\infty} S_k$ is empty iff there is an unbounded sequence $\{x_k\}$ consisting of minimum norm vectors from the S_k.

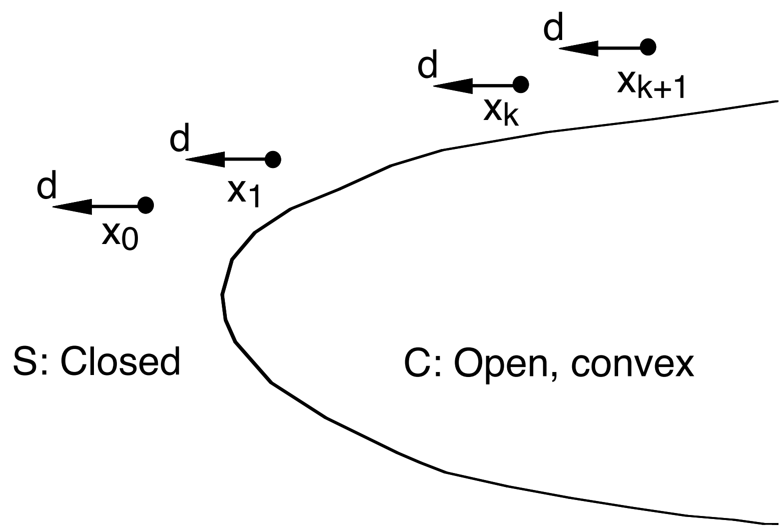
  (b) An asymptotic sequence $\{x_k\}$ consisting of minimum norm vectors from the S_k cannot be retractive, because such a sequence eventually gets closer to 0 when shifted opposite to the asymptotic direction.



RECOGNIZING RETRACTIVE SETS

- Unions, intersections, and Cartesian produsts of retractive sets are retractive.

- The complement of an open convex set is retractive.



- Closed halfspaces are retractive.

- Polyhedral sets are retractive.

- Sets of the form $\{x \mid f_j(x) \geq 0,\, j = 1, \ldots, r\}$, where $f_j : \Re^n \mapsto \Re$ is convex, are retractive.

- Vector sum of a compact set and a retractive set is retractive.

- Nonpolyhedral cones are not retractive, level sets of quadratic functions are not retractive.

LINEAR AND QUADRATIC PROGRAMMING

- **Frank-Wolfe Theorem:** Let

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

where Q is symmetric (not necessarily positive semidefinite). If the minimal value of f over X is finite, there exists a minimum of f of over X.

- **Proof (outline):** Choose $\{\gamma_k\}$ s.t. $\gamma_k \downarrow f^*$, where f^* is the optimal value, and let

$$S_k = \{x \in X \mid x'Qx + c'x \leq \gamma_k\}$$

The set of optimal solutions is $\cap_{k=0}^{\infty} S_k$, so it will suffice to show that for each asymptotic direction of $\{S_k\}$, each corresponding asymptotic sequence is retractive.

Choose an asymptotic direction d and a corresponding asymptotic sequence. Note that X is retractive, so for k sufficiently large, we have $x_k - d \in X$.

PROOF OUTLINE – CONTINUED

- We use the relation $x_k' Q x_k + c' x_k \leq \gamma_k$ to show that

$$d' Q d \leq 0, \qquad a_j' d \leq 0, \quad j = 1, \ldots, r$$

- Then show, using the finiteness of f^* [which implies $f(x + \alpha d) \geq f^*$ for all $x \in X$], that

$$(c + 2Qx)' d \geq 0, \qquad \forall\, x \in X$$

- Thus,

$$\begin{aligned} f(x_k - d) &= (x_k - d)' Q (x_k - d) + c'(x_k - d) \\ &= {x_k}' Q x_k + c' x_k - (c + 2Q x_k)' d + d' Q d \\ &\leq {x_k}' Q x_k + c' x_k \\ &\leq \gamma_k, \end{aligned}$$

so $x_k - d \in S_k$. **Q.E.D.**

INTERSECTION THEOREM FOR CONVEX SETS

Let $\{C_k\}$ be a nested sequence of nonempty closed convex sets. Denote

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

(a) If $R = L$, then $\{C_k\}$ is retractive, and $\cap_{k=0}^{\infty} C_k$ is nonempty. Furthermore, we have

$$\cap_{k=0}^{\infty} C_k = L + \tilde{C},$$

where $\tilde{C}$ is some nonempty and compact set.

(b) Let X be a retractive closed set. Assume that all the sets $S_k = X \cap C_k$ are nonempty, and that

$$A_X \cap R \subset L.$$

Then, $\{S_k\}$ is retractive, and $\cap_{k=0}^{\infty} S_k$ is nonempty.

CRITICAL ASYMPTOTES

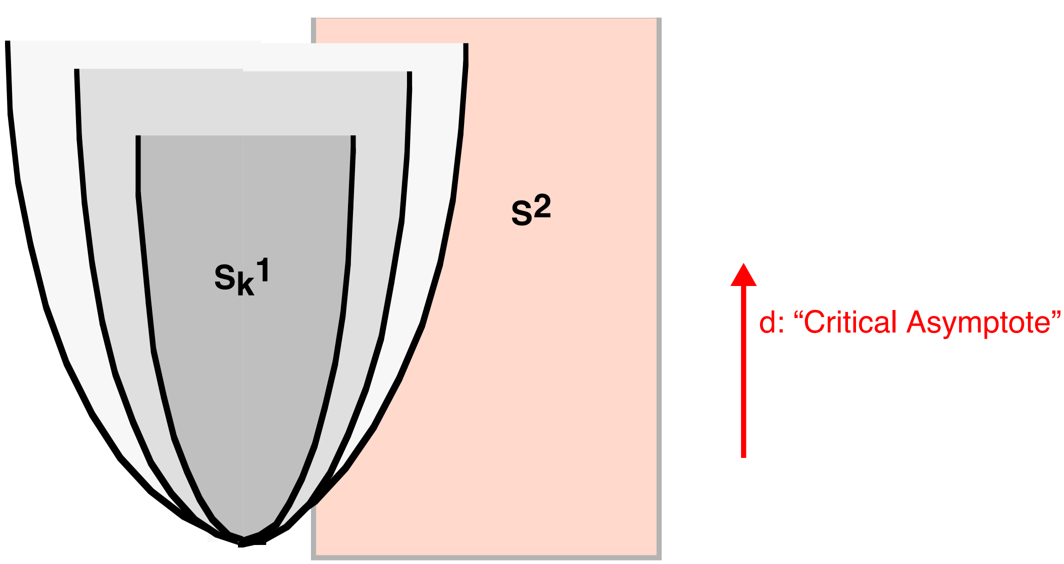
- Retractiveness works well for sets with a polyhedral structure, but not for sets specified by convex quadratic inequalities.

- **Key question:** Given nested sequences $\{S_k^1\}$ and $\{S_k^2\}$ each with nonempty intersection by itself, and with

$$S_k^1 \cap S_k^2 \neq \emptyset, \qquad k = 0, 1, \dots,$$

what causes the intersection sequence $\{S_k^1 \cap S_k^2\}$ to have an empty intersection?

- The trouble lies with the existence of some "critical asymptotes."



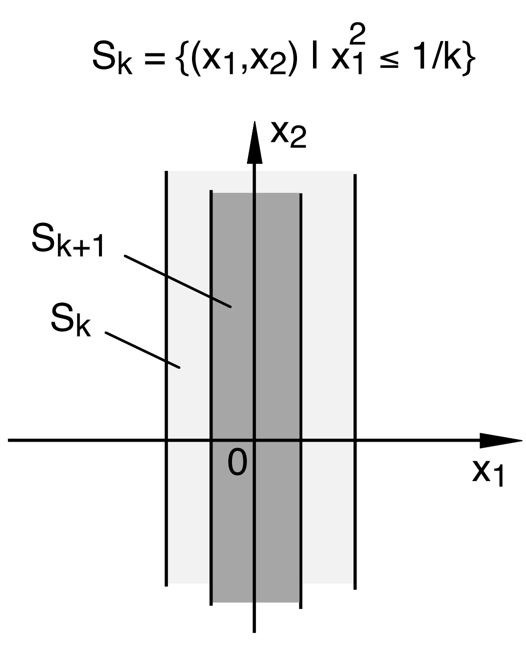
HORIZON DIRECTIONS

- Consider $\{S_k\}$ with $\cap_{k=0}^{\infty} S_k \neq \emptyset$. An asymptotic direction d of $\{S_k\}$ is:

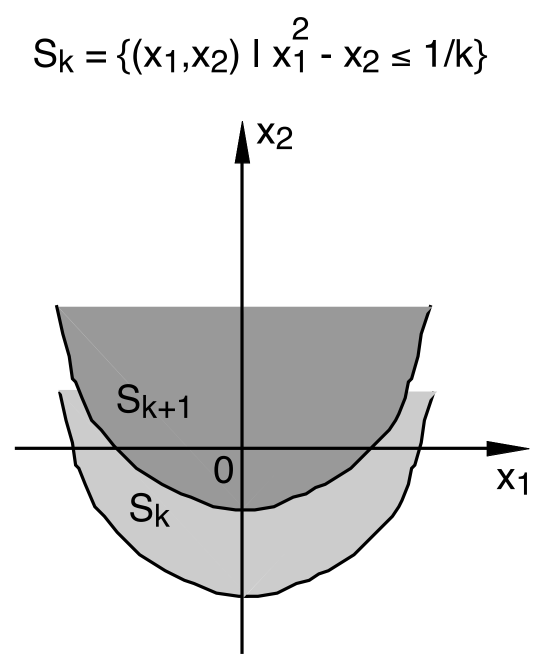
  (a) A *local horizon direction* if, for every $x \in \cap_{k=0}^{\infty} S_k$, there exists a scalar $\overline{\alpha} \geq 0$ such that $x + \alpha d \in \cap_{k=0}^{\infty} S_k$ for all $\alpha \geq \overline{\alpha}$.

  (b) A *global horizon direction* if for every $x \in \Re^n$ there exists a scalar $\overline{\alpha} \geq 0$ such that $x + \alpha d \in \cap_{k=0}^{\infty} S_k$ for all $\alpha \geq \overline{\alpha}$.

- **Example: (2-D Convex Quadratic Set Sequences)**



Directions $(0,\gamma)$, $\gamma \neq 0$, are local horizon directions that are retractive



Directions $(0,\gamma)$, $\gamma > 0$, are global horizon directions

GENERAL CONVEX QUADRATIC SETS

- Let $S_k = \big\{x \mid x'Qx + a'x + b \le \gamma_k\big\}$, where $\gamma_k \downarrow 0$. Then, if all the sets S_k are nonempty, $\cap_{k=0}^{\infty} S_k \neq \emptyset$.

- Asymptotic directions: $d \neq 0$ such that $Qd = 0$ and $a'd \le 0$. There are two possibilities:

  (a) $Qd = 0$ and $a'd < 0$, in which case d is a global horizon direction.

  (b) $Qd = 0$ and $a'd = 0$, in which case d is a direction of constancy of f, and it follows that d is a retractive local horizon direction.

- Drawing some 2-dimensional pictures and using the structure of asymptotic directions demonstrated above, we conjecture that there are no "critical asymptotes" for set sequences of the form $\{S_k^1 \cap S_k^2\}$ when S_k^1 and S_k^2 are convex quadratic sets.

- This motivates a general definition of noncritical asymptotic direction.

CRITICAL DIRECTIONS

- Given a nested closed set sequence $\{S_k\}$ with nonempty intersection, we say that an asymptotic direction d of $\{S_k\}$ is *noncritical* if d is either a global horizon direction of $\{S_k\}$, or a retractive local horizon direction of $\{S_k\}$.

- **Proposition:** Let $S_k = S_k^1 \cap S_k^2 \cap \cdots \cap S_k^r$, where $\{S_k^j\}$ are nested sequence such that

$$S_k \neq \emptyset, \quad \forall\, k, \qquad \cap_{k=0}^{\infty} S_k^j \neq \emptyset, \qquad \forall\, j.$$

Assume that all the asymptotic directions of all $\{S_k^j\}$ are noncritical. Then $\cap_{k=0}^{\infty} S_k \neq \emptyset$.

- **Special case: (Convex Quadratic Inequalities)** Let

$$S_k = \left\{ x \mid x'Q_j x + a_j' x + b_j \leq \gamma_k^j,\ j = 1, \ldots, r \right\}$$

where $\{\gamma_k^j\}$ are scalar sequences with $\gamma_k^j \downarrow 0$. Assume that $S_k \neq \emptyset$ is nonempty for all k. Then, $\cap_{k=0}^{\infty} S_k \neq \emptyset$.

APPLICATION TO QUADRATIC MINIMIZATION

- Let

$$f(x) = x'Qx + c'x,$$

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

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: Let f^* be the minimal value, and let $\gamma_k \downarrow f^*$. The set of optimal solutions is

$$X^* = \cap_{k=0}^{\infty}\big(X \cap \{x \mid x'Qx + c'x \leq \gamma_k\}\big).$$

All the set sequences involved in the intersection are convex quadratic and hence have no critical directions. By the preceding proposition, X^* is nonenpty. **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) $A\,C$ is closed if $R_C \cap N(A) \subset L_C$.

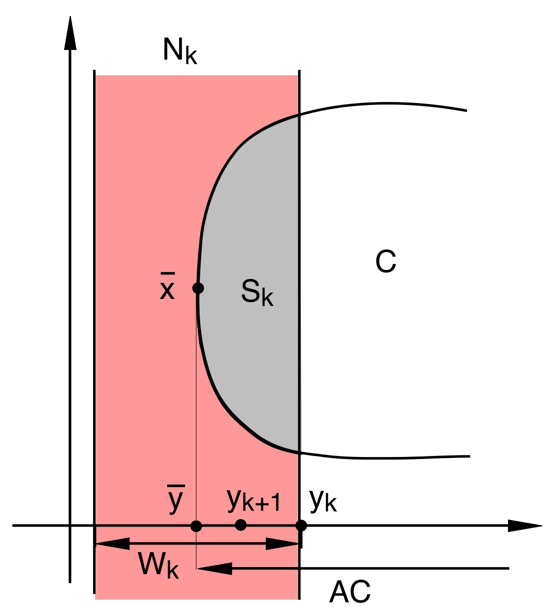
  (b) $A(X \cap C)$ is closed if X is a polyhedral set and

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

  (c) AC is closed if $C = \{x \mid f_j(x) \le 0,\, j = 1, \ldots, r\}$ with f_j: convex quadratic functions.

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

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



LECTURE 7

LECTURE OUTLINE

- Existence of optimal solutions
- Preservation of closure under partial minimization
- Hyperplane separation
- Nonvertical hyperplanes
- Min common and max crossing problems

- We have talked so far about set intersection theorems that use two types of asymptotic directions:
  - Retractive directions (mostly for polyhedral-type sets)
  - Horizon directions (for special types of sets - e.g., quadratic)
- We now apply these theorems to issues of existence of optimal solutions, and preservation of closedness under linear transformation, vector sum, and partial minimization.

PROJECTION THEOREM

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

(a) For every $x \in \Re^n$, there exists a unique vector $P_C(x)$ that minimizes $\|z - x\|$ over all $z \in C$ (called the *projection of* x *on* C).

(b) For every $x \in \Re^n$, a vector $z \in C$ is equal to $P_C(x)$ if and only if

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

In the case where C is an affine set, the above condition is equivalent to

$$x - z \in S^\perp,$$

where S is the subspace that is parallel to C.

(c) The function $f : \Re^n \mapsto C$ defined by $f(x) = P_C(x)$ is continuous and nonexpansive, i.e.,

$$\big\|P_C(x) - P_C(y)\big\| \leq \|x - y\|, \qquad \forall\, x, y \in \Re^n$$

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 three 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.

(3) $f^* > -\infty$, and f and X are specified by convex quadratic functions:

$$f(x) = x'Qx + c'x,$$

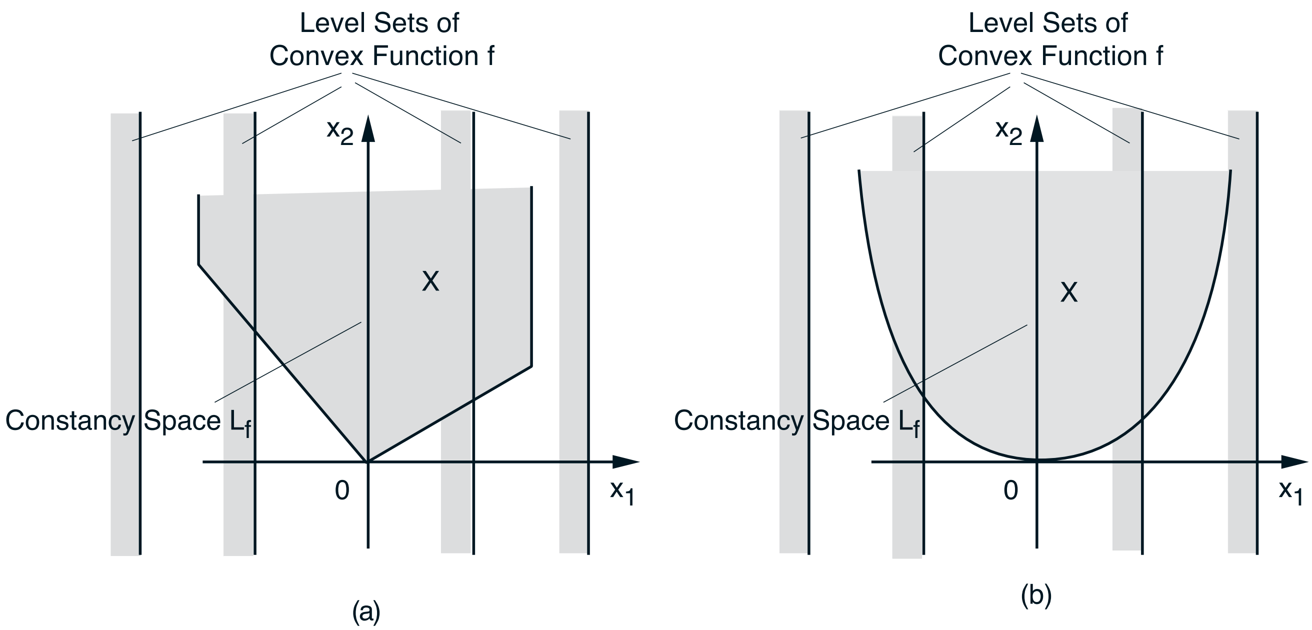
$$X = \{x \mid x'Q_jx + a_j'x + b_j \leq 0,\ j = 1, \ldots, r\}.$$

Proof: Follows by writing

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

and by applying the corresponding set intersection theorems. **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),

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

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

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

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

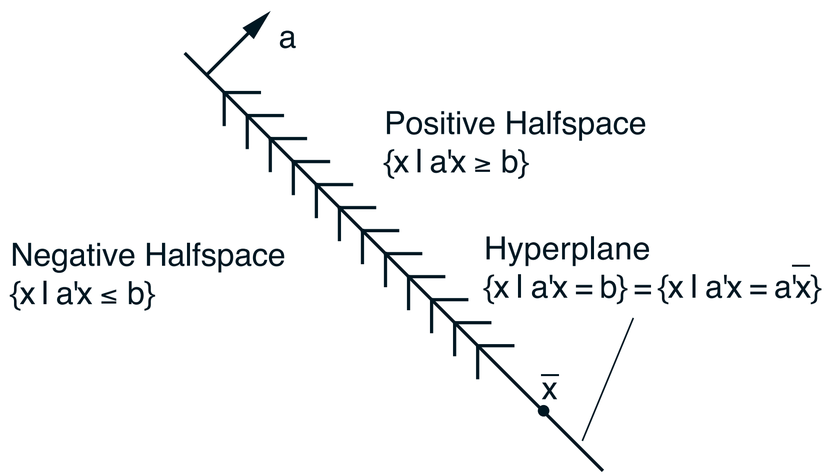
is nonempty and compact, then f is convex, closed, and proper. Also, for each $x \in \mathrm{dom}(f)$, the set of minima of $F(x, \cdot)$ is nonempty and compact.

Proof: (Outline) By the hypothesis, there is no nonzero y such that $(0, y, 0) \in R_{\mathrm{epi}(F)}$. Also, all the nonempty level sets

$$\{z \mid F(x, z) \leq \gamma\}, \qquad x \in \Re^n, \;\; \gamma \in \Re,$$

have the same recession cone, which by hypothesis, is equal to $\{0\}$.

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.,

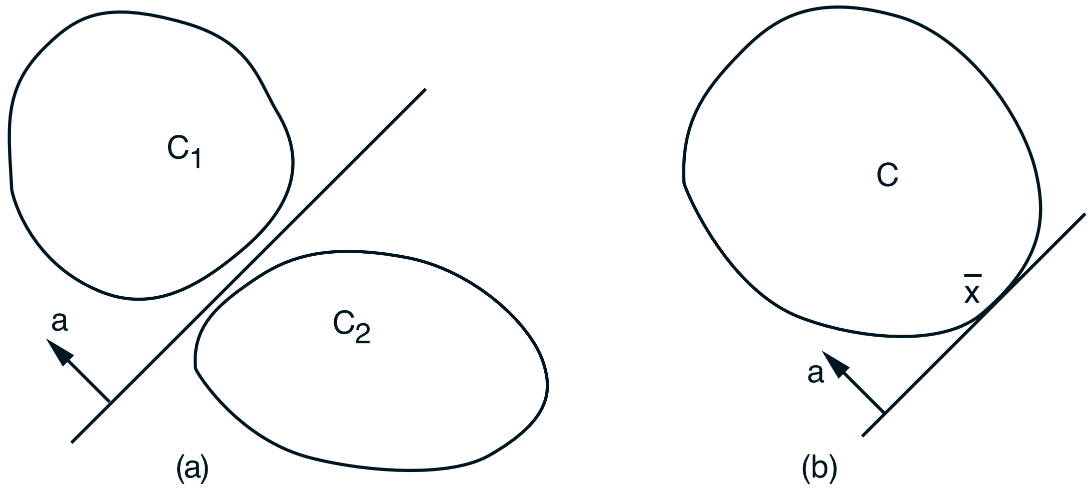
$$\text{either} \quad a'x_1 \le b \le a'x_2, \qquad \forall\, x_1 \in C_1, \forall\, x_2 \in C_2,$$

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

- 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* $\overline{x}$.

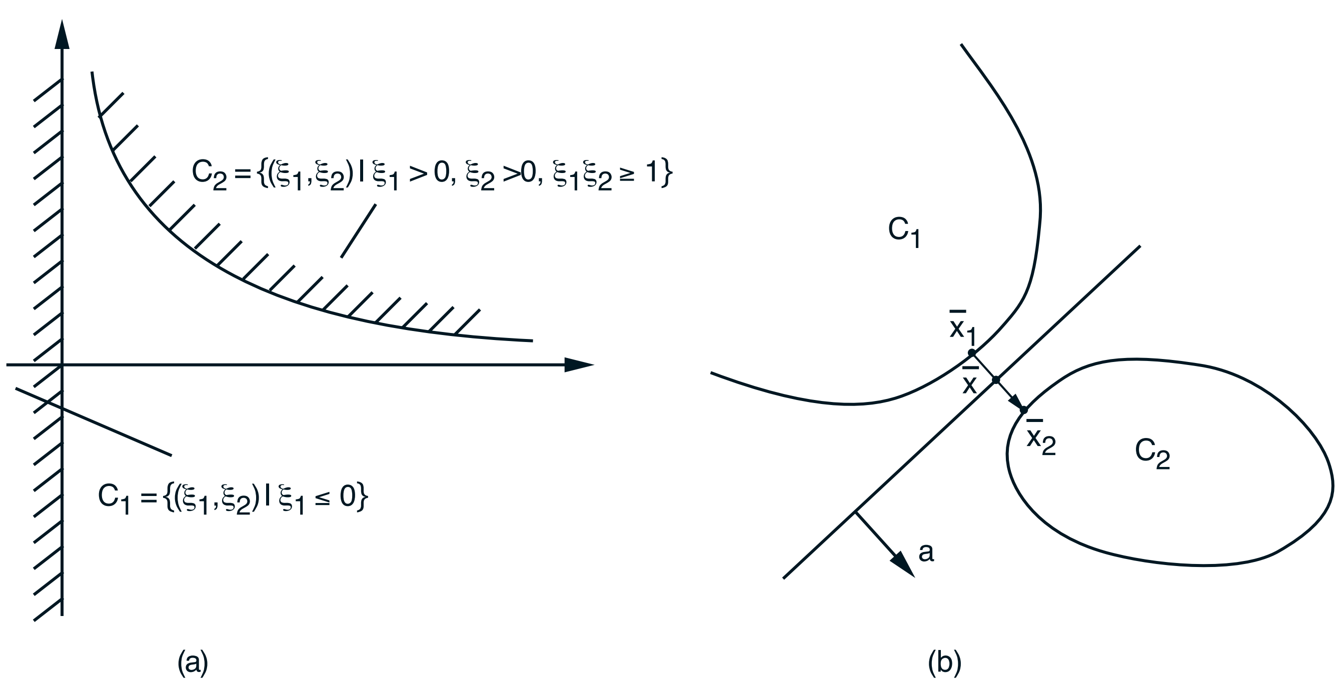
VISUALIZATION

- Separating and supporting hyperplanes:



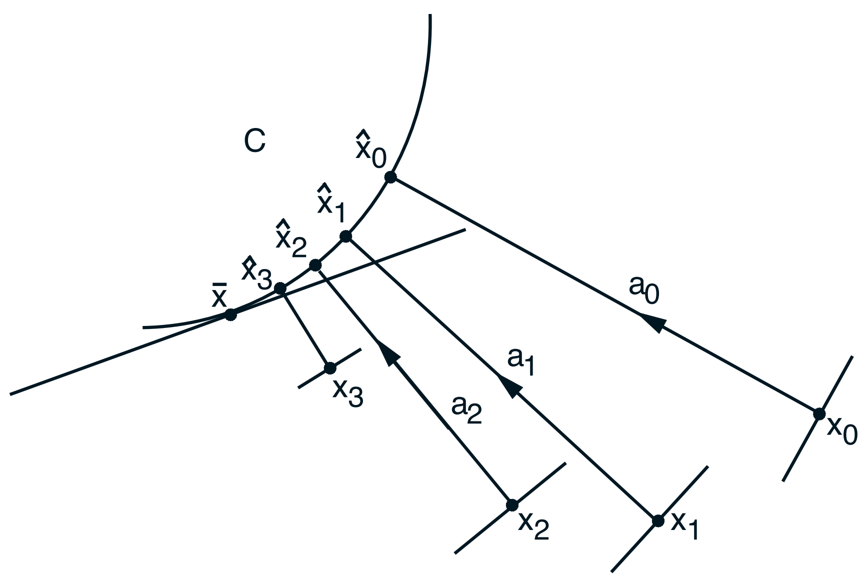
- A separating $\{x \mid a'x = b\}$ that is disjoint from C_1 and C_2 is called *strictly* separating:

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



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

SEPARATING HYPERPLANE THEOREM

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

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

Proof: Consider the convex set

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

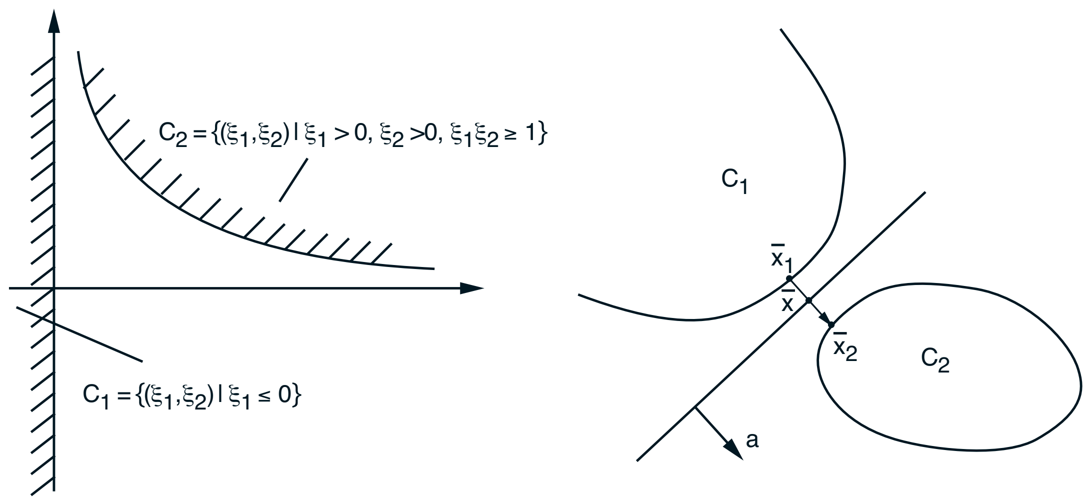
Since C_1 and C_2 are disjoint, the origin does not belong to $C_1 - C_2$, so by the Supporting Hyperplane Theorem, there exists a vector $a \neq 0$ such that

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

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

STRICT SEPARATION THEOREM

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



(a) (b)

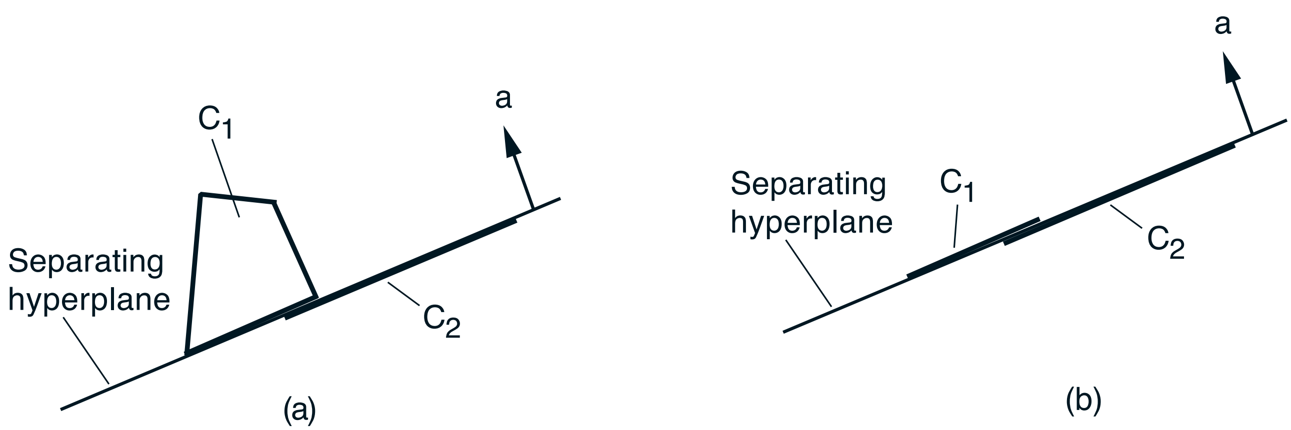
Proof: (Outline) Consider the set $C_1 - C_2$. Since C_1 is closed and C_2 is compact, $C_1 - C_2$ is closed. Since $C_1 \cap C_2 = \emptyset$, $0 \notin C_1 - C_2$. Let $\overline{x}_1 - \overline{x}_2$ be the projection of 0 onto $C_1 - C_2$. The strictly separating hyperplane is constructed as in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

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

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

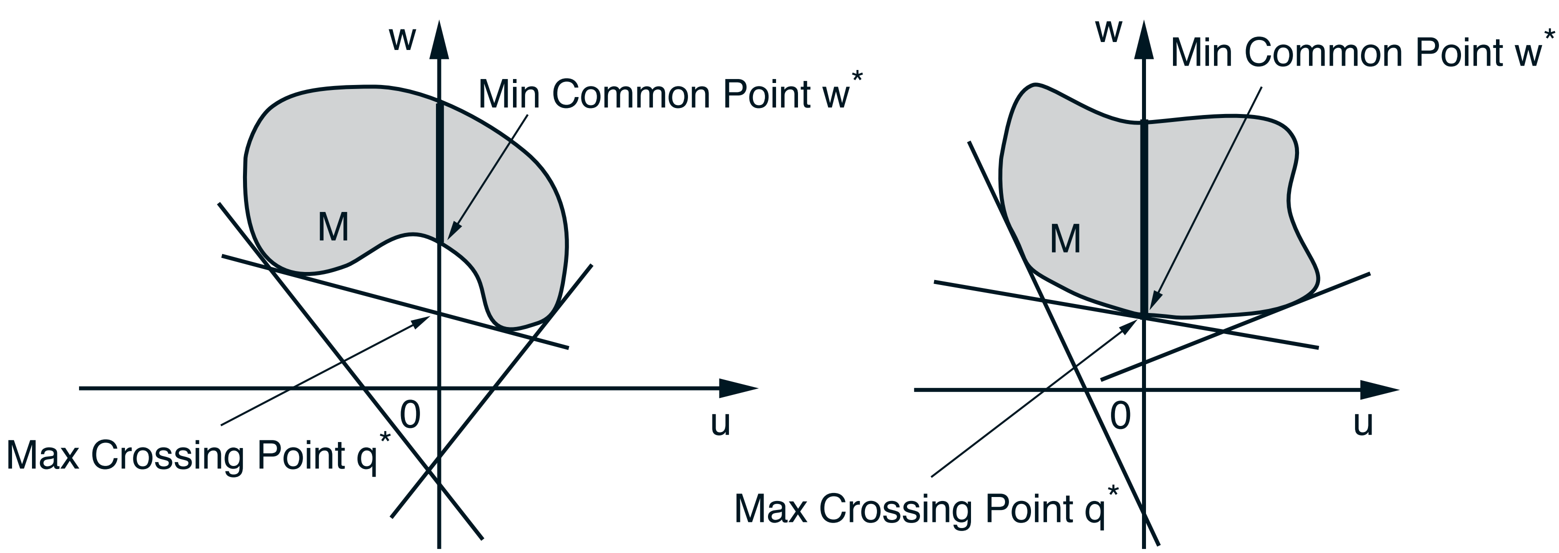
$$\text{ri}(C_1) \cap \text{ri}(C_2) = \emptyset$$

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.

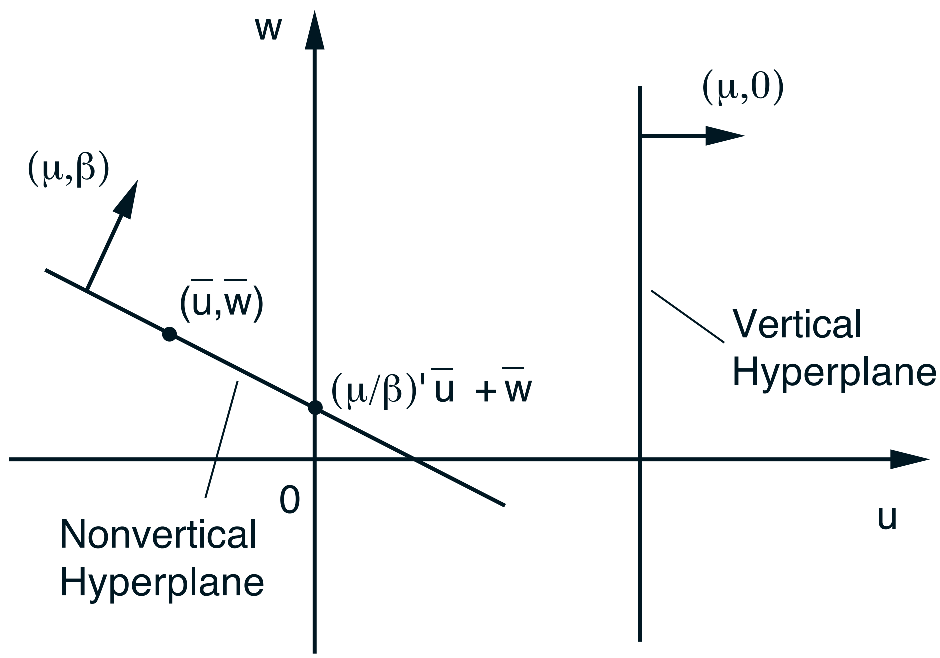


- We first need to study "nonvertical" hyperplanes.

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.

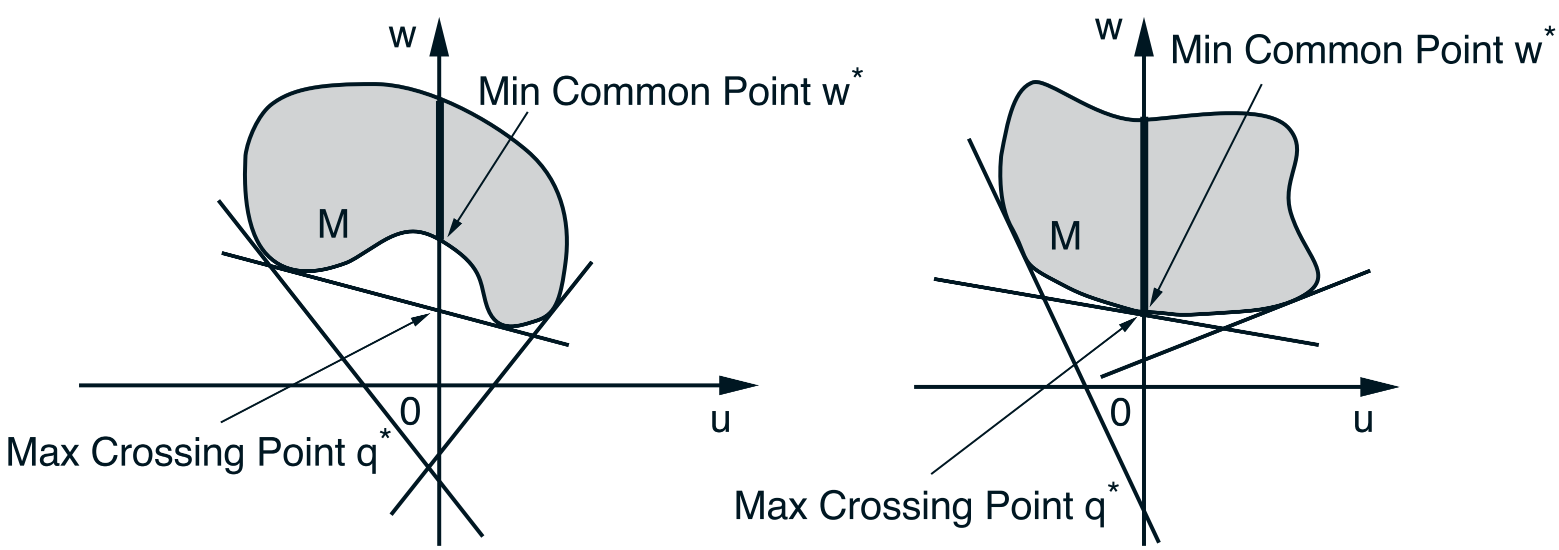
(a) C is the intersection of the closed halfspaces containing C. If all these corresponded to vertical hyperplanes, C would contain a vertical line.

(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

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

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

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

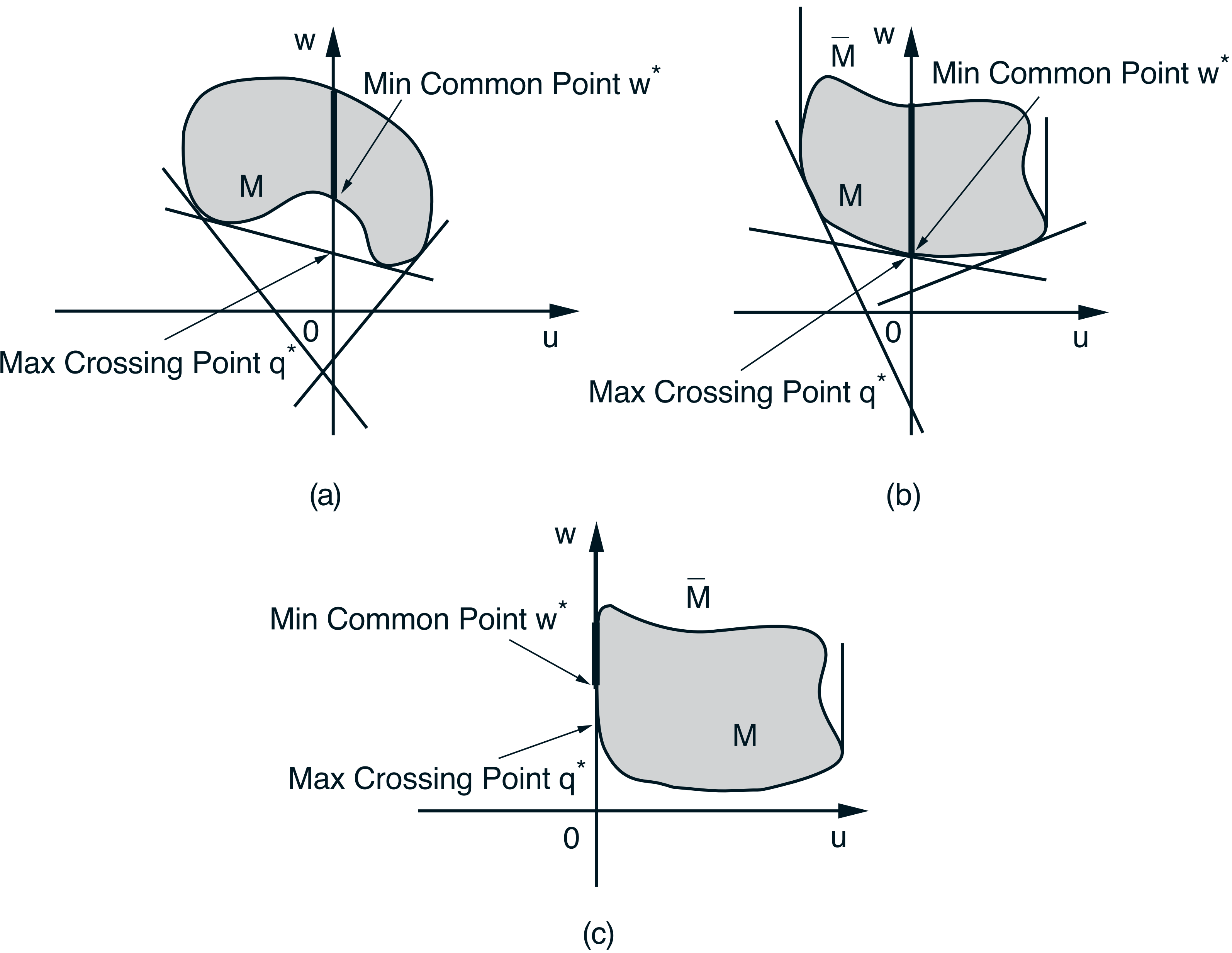
$$q(\mu) = \inf_{(u,w)\in M}\{w + \mu' u\} \le \inf_{(0,w)\in M} w = w^*,$$

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?



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 a vector $\mu \in \Re^n$ such that $q(\mu) = q^*$. If D contains the origin in its interior, the set of all $\mu \in \Re^n$ such that $q(\mu) = q^*$ is compact.

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

PROOF OF THEOREM I

- Assume that for every sequence $\{(u_k, w_k)\} \subset M$ with $u_k \to 0$, there holds $w^* \leq \liminf_{k\to\infty} w_k$. 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 $w^* - \epsilon \leq \xi \leq w^*$, so $w^* - \epsilon \leq q^* \leq w^*$. Since ϵ can be arbitrarily small, it follows that $q^* = w^*$.

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

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

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

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

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

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

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

Since for any $(\overline{u}, \overline{w}) \in M$, the set $\overline{M}$ contains the halfline $\big\{(\overline{u}, w) \mid \overline{w} \le w\big\}$, it follows that $\beta \ge 0$. If $\beta = 0$, then $0 \le \mu' u$ 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

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

Since the inequality $q^* \le w^*$ 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

$$\text{minimize } \sup_{z \in Z} \phi(x, z)$$
$$\text{subject to } x \in X$$

and

$$\text{maximize } \inf_{x \in X} \phi(x, z)$$
$$\text{subject to } z \in Z.$$

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

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

  - 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

minimize $f(x)$

subject to $x \in X, \qquad g_j(x) \le 0, \quad j = 1, \ldots, r$

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

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

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 $\max_z x' A z$
- 2nd player maximizes $\min_x x' A z$

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

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

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

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

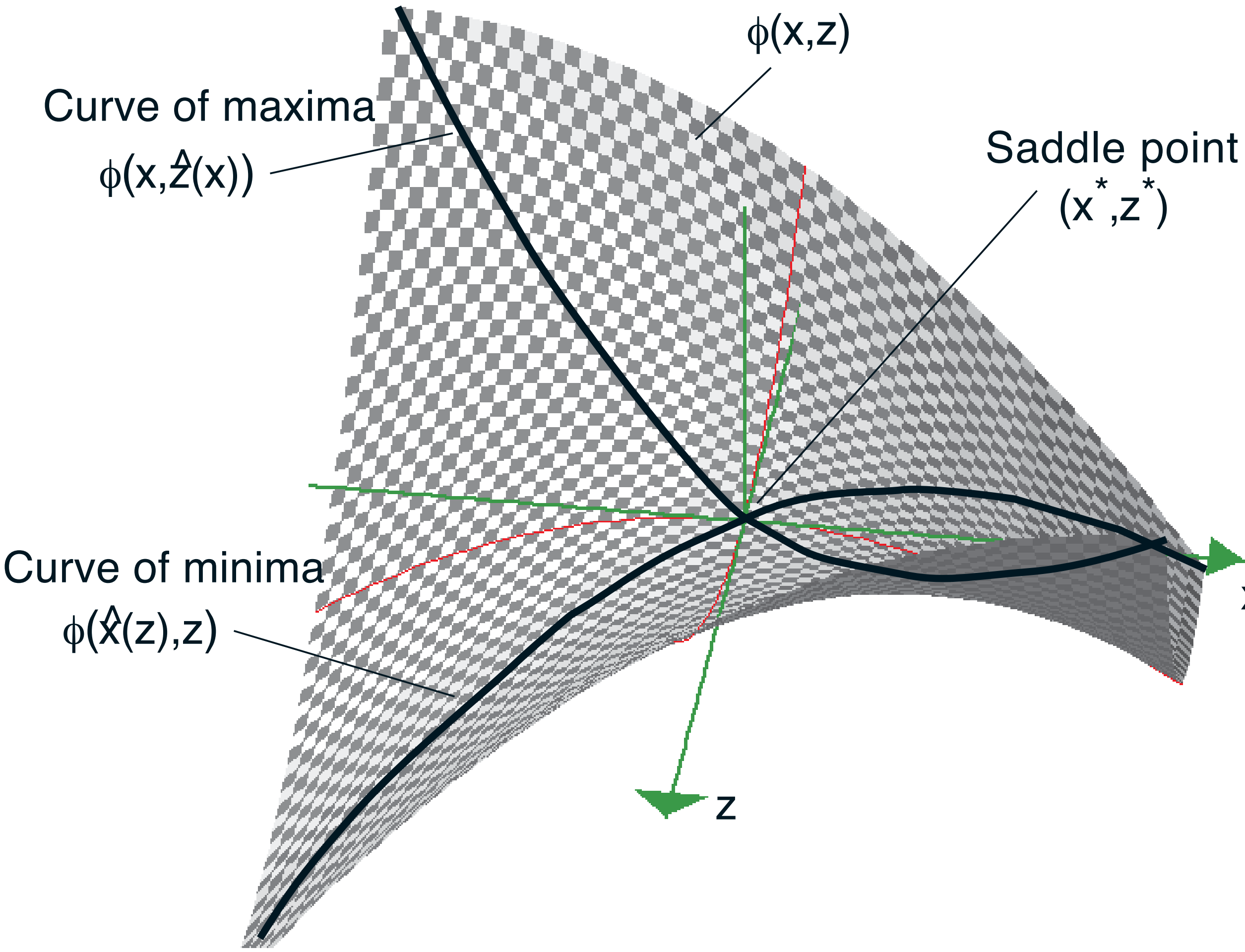
Conversely, if Eq. (*) holds, then

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

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

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

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

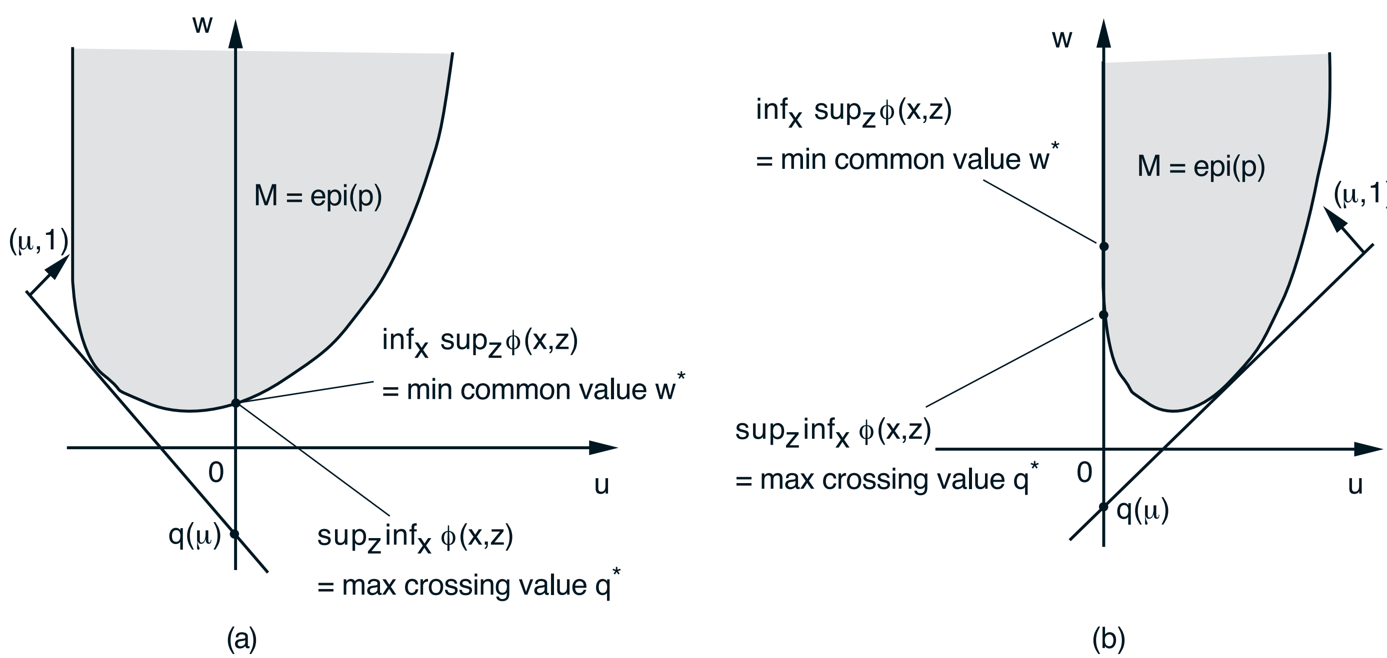
Saddle points correspond to points where these two curves meet.

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 the set M equal to the epigraph of p.
- Application of a more general idea: To evaluate a quantity of interest w^*, introduce a suitable perturbation u and function p, with $p(0) = w^*$.
- 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

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

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 $\mathrm{epi}(p)$] 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 $\mathrm{dom}(f)$.]

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

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

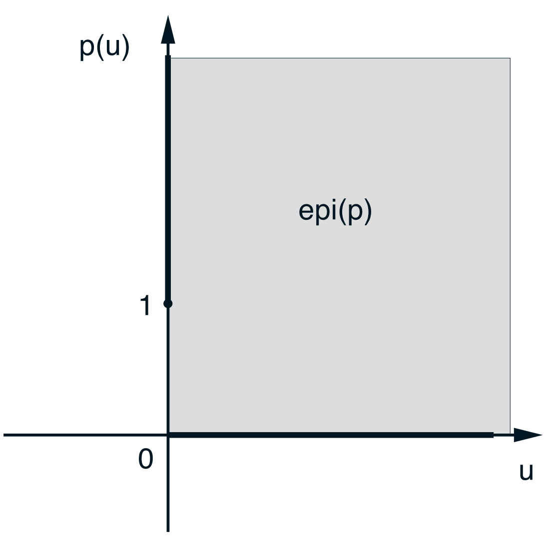
which satisfy the convexity and closedness assumptions. For all $z \geq 0$,

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

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

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

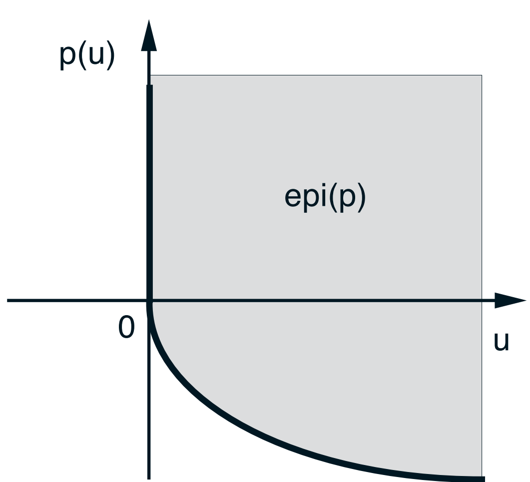
which satisfy the convexity and closedness assumptions. For all $z \geq 0$,

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

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 has underscored 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:

- X and Z are convex and compact.
- $\phi(\cdot, z)$: convex for each $z \in Z$, and $\phi(x, \cdot)$ is concave and upper semicontinuous over Z for each $x \in X$, so that the min common/max crossing framework is applicable.
- $\phi(\cdot, z)$ is lower semicontinuous over X, so that F is convex and closed (it is the pointwise supremum over $z \in Z$ of closed convex functions).

(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.

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 $\{x \in X \mid \phi(x, \overline{z}) \leq \gamma\}$ 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 $\{z \in Z \mid \phi(\overline{x}, z) \geq \gamma\}$ 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

$$\{x \in X \mid \phi(x, \overline{z}) \leq \gamma\}, \quad \{z \in Z \mid \phi(\overline{x}, z) \geq \gamma\},$$

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

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

- The main convexity concepts so far have been:
  - Closure, convex hull, affine hull, relative interior, directions of recession
  - 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

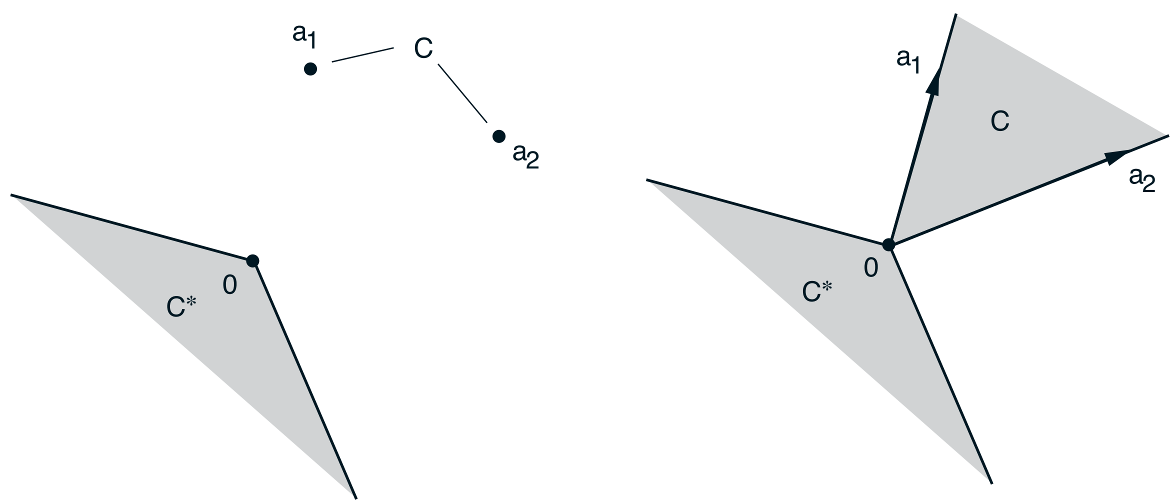
- We now introduce new concepts with important theoretical and algorithmic implications: polyhedral convexity, extreme points, and related issues.

POLAR CONES

- Given a set C, the cone given by

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

is called the *polar cone* of C.



(a) (b)

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

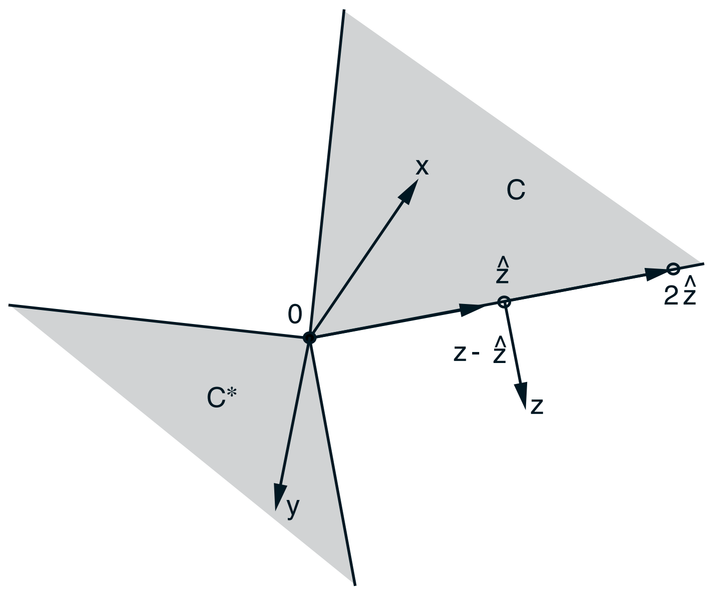
- Note that

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

- Important example: 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 $x'y \leq 0$ for all $y \in C^*$, so that $x \in (C^*)^*$, and $C \subset (C^*)^*$.

To prove the reverse inclusion, take $z \in (C^*)^*$, and let $\hat{z}$ be the projection of z on C, so that $(z - \hat{z})'(x - \hat{z}) \leq 0$, for all $x \in C$. Taking $x = 0$ and $x = 2\hat{z}$, we obtain $(z - \hat{z})'\hat{z} = 0$, so that $(z - \hat{z})'x \leq 0$ for all $x \in C$. Therefore, $(z - \hat{z}) \in C^*$, and since $z \in (C^*)^*$, we have $(z - \hat{z})'z \leq 0$. Subtracting $(z - \hat{z})'\hat{z} = 0$ yields $\|z - \hat{z}\|^2 \leq 0$. It follows that $z = \hat{z}$ and $z \in C$, implying that $(C^*)^* \subset C$.

POLYHEDRAL AND FINITELY GENERATED CONES

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

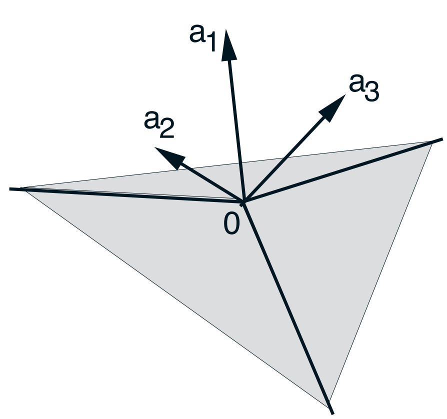
$$C = \{x \mid a_j'x \leq 0,\ j = 1, \ldots, r\},$$

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

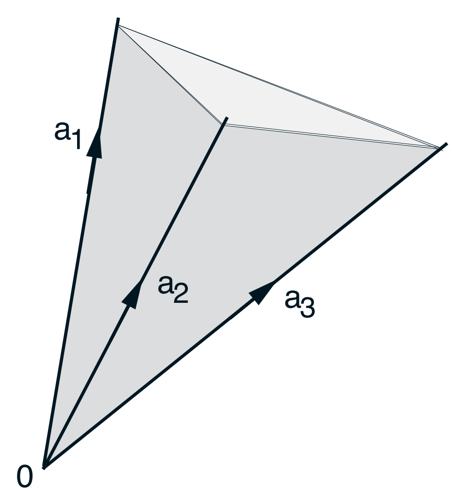
- A cone $C \subset \Re^n$ is *finitely generated*, if

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

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



(a)



(b)

FARKAS-MINKOWSKI-WEYL THEOREMS

Let $a_1, \ldots, a_r$ be given vectors in $\Re^n$, and let

$$C = \text{cone}\big(\{a_1, \ldots, a_r\}\big)$$

(a) C is closed and

$$C^* = \big\{y \mid a_j' y \le 0,\ j = 1, \ldots, r\big\}$$

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

$$\big\{y \mid a_j' y \le 0,\ j = 1, \ldots, r\big\}^* = C$$

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

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

PROOF OUTLINE

(a) First show that for $C = \text{cone}(\{a_1, \ldots, a_r\})$,

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

If $y'a_j \le 0$ for all j, then $y'x \le 0$ for all $x \in C$, so $C^* \supset \{y \mid a_j'y \le 0,\ j = 1, \ldots, r\}$. Conversely, if $y \in C^*$, i.e., if $y'x \le 0$ for all $x \in C$, then, since $a_j \in C$, we have $y'a_j \le 0$, for all j. Thus, $C^* \subset \{y \mid a_j'y \le 0,\ j = 1, \ldots, r\}$.

• Showing that $C = \text{cone}(\{a_1, \ldots, a_r\})$ is closed is nontrivial! Follows from Prop. 1.5.8(b), which shows (as a special case where $C = \Re^n$) that closedness of polyhedral sets is preserved by linear transformations. (The text has two other lines of proof.)

(b) Assume no equalities. Farkas' Lemma says:

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

Since by part (a), $C^* = \{y \mid a_j'y \le 0,\ j = 1, \ldots, r\}$ and C is closed and convex, the result follows by the Polar Cone Theorem.

(c) See the text.

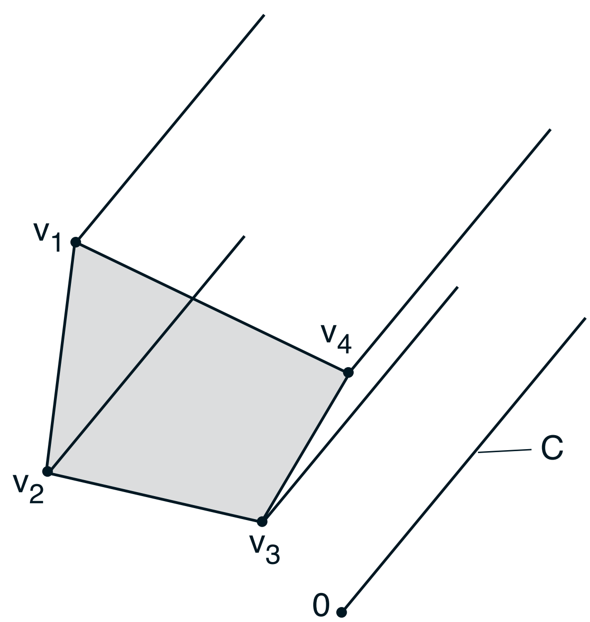
POLYHEDRAL SETS

- A set $P \subset \Re^n$ is said to be *polyhedral* if it is nonempty and

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

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

- A polyhedral set may involve affine equalities (convert each into two affine inequalities).



Theorem: A set P is polyhedral if and only if

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

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

PROOF OUTLINE

Proof: Assume that P is polyhedral. Then,

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

for some a_j and b_j. Consider the polyhedral cone

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

and note that $P = \{x \mid (x,1) \in \hat{P}\}$. By Minkowski-Weyl, $\hat{P}$ is finitely generated, so it has the form

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

for some v_j and d_j. Since $w \ge 0$ for all vectors $(x,w) \in \hat{P}$, we see that $d_j \ge 0$ for all j. Let

$$J^+ = \{j \mid d_j > 0\}, \qquad J^0 = \{j \mid d_j = 0\}$$

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 $P = \{x \mid (x,1) \in \hat{P}\}$, 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 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 can be written as

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

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 a polyhedral and can be represented as the intersection of a finite collection of closed halfspaces of the form $\big\{(x,w) \mid a_j'x + b_j \le c_j w\big\}$, $j = 1,\ldots,r$, 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 $\gamma \ge 0$, it follows that $c_j \ge 0$, 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,

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

Thus

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

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

Q.E.D.

LECTURE 11

LECTURE OUTLINE

- Extreme points
- Extreme points of polyhedral sets
- Extreme points and linear/integer programming

Recall some of the facts of polyhedral convexity:

- Polarity relation between polyhedral and finitely generated cones

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

- Farkas' Lemma

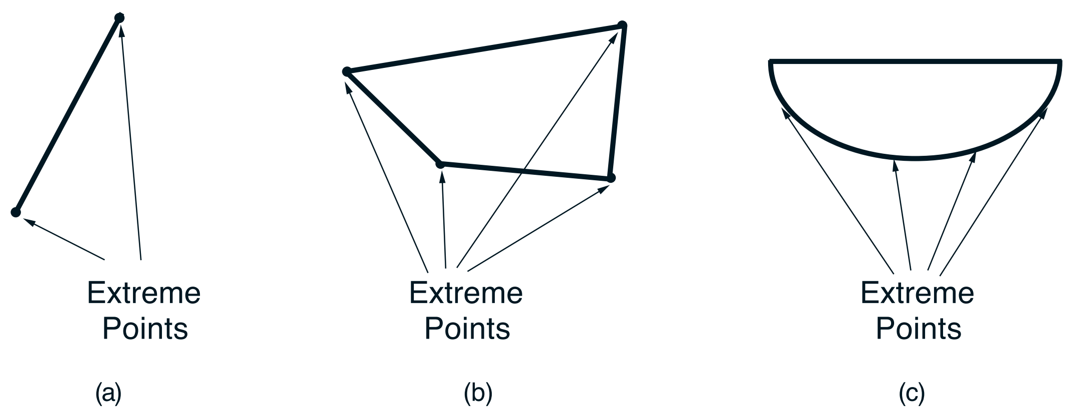
$$\{x \mid a_j'x \le 0,\ j = 1, \ldots, r\}^* = \text{cone}\big(\{a_1, \ldots, a_r\}\big)$$

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

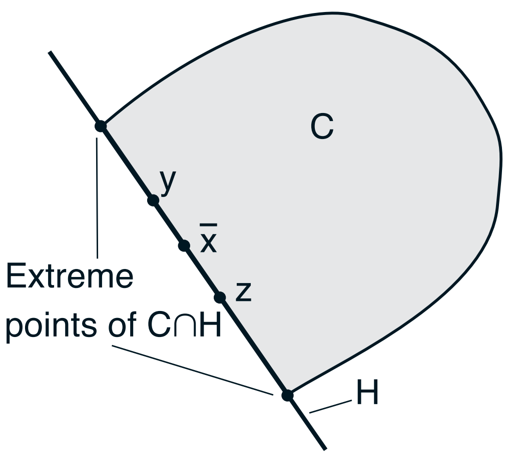
Polyhedral set $P = \text{conv}\big(\{v_1, \ldots, v_m\}\big) + R_P$

EXTREME POINTS

- A vector x is an *extreme point* of a convex set C if $x \in C$ and x cannot be expressed as a convex combination of two vectors of C, both of which are different from x.



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: Let $\overline{x} \in C \cap H$ be a nonextreme point of C. Then $\overline{x} = \alpha y + (1-\alpha)z$ for some $\alpha \in (0,1)$, $y, z \in C$, with $y \neq x$ and $z \neq x$. Since $\overline{x} \in H$, the closed halfspace containing C is of the form $\{x \mid a'x \geq a'\overline{x}\}$. Then $a'y \geq a'\overline{x}$ and $a'z \geq a'\overline{x}$, which in view of $\overline{x} = \alpha y + (1-\alpha)z$, implies that $a'y = a'\overline{x}$ and $a'z = a'\overline{x}$. Thus, $y, z \in C \cap H$, showing that $\overline{x}$ is not an extreme point of $C \cap H$.

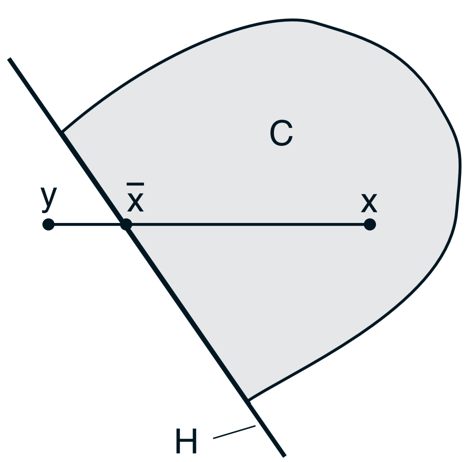
PROPERTIES OF EXTREME POINTS I

Proposition: A closed and 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 - 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.

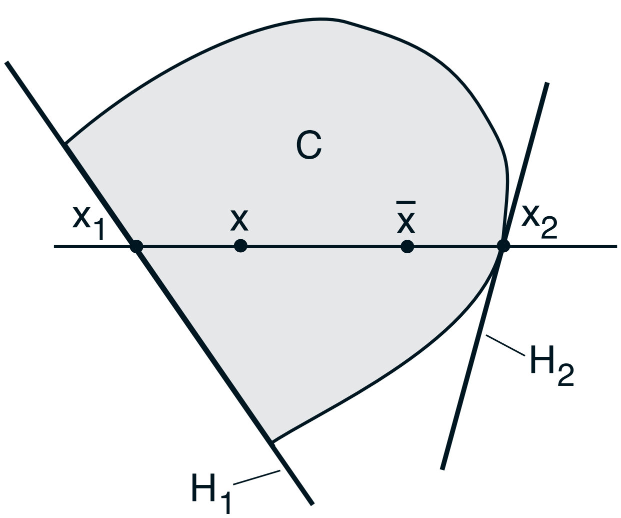
PROPERTIES OF EXTREME POINTS II

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.

EXTREME POINTS OF POLYHEDRAL SETS

- Let P be a polyhedral subset of $\Re^n$. If the set of extreme points of P is nonempty, then it is finite.

Proof: Consider the representation $P = \hat{P} + C$, where

$$\hat{P} = \text{conv}\big(\{v_1, \ldots, v_m\}\big)$$

and C is a finitely generated cone.

- An extreme point $\overline{x}$ of P cannot be of the form $\overline{x} = \hat{x} + y$, where $\hat{x} \in \hat{P}$ and $y \neq 0$, $y \in C$, since in this case $\overline{x}$ would be the midpoint of the line segment connecting the distinct vectors $\hat{x}$ and $\hat{x} + 2y$. Therefore, an extreme point of P must belong to $\hat{P}$, and since $\hat{P} \subset P$, it must also be an extreme point of $\hat{P}$.

- An extreme point of $\hat{P}$ must be one of the vectors $v_1, \ldots, v_m$, since otherwise this point would be expressible as a convex combination of $v_1, \ldots, v_m$. Thus the extreme points of P belong to the finite set $\{v_1, \ldots, v_m\}$. **Q.E.D.**

CHARACTERIZATION OF EXTREME POINTS

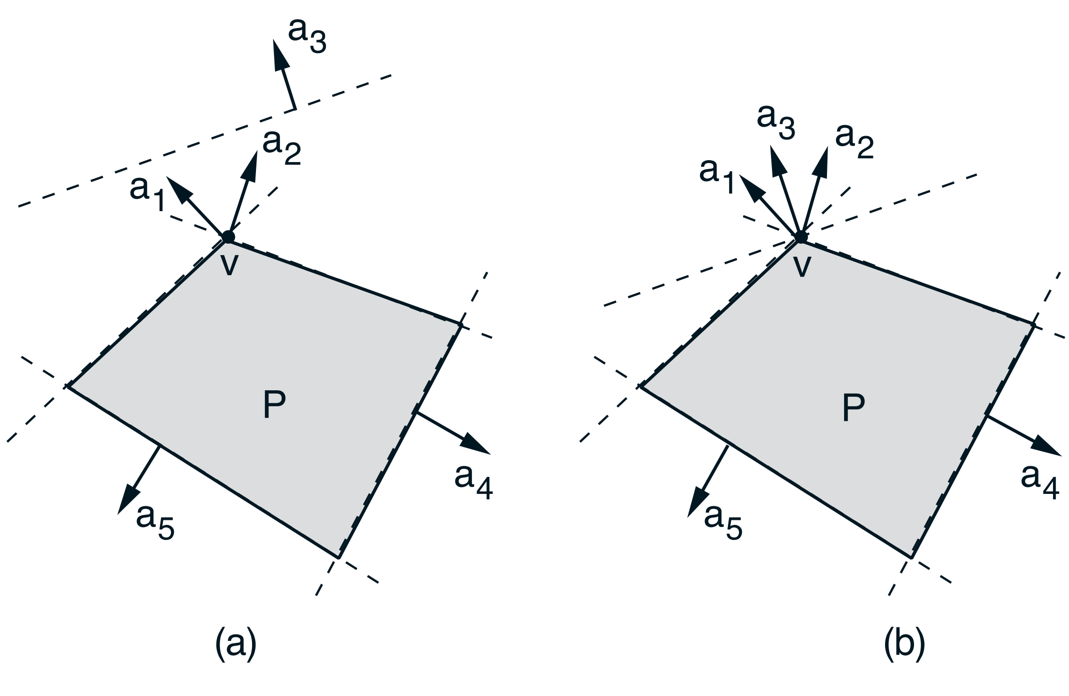
Proposition: Let P be a polyhedral subset of $\Re^n$. If P has the form

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

where a_j and b_j are given vectors and scalars, respectively, then 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.



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 $a_j \in \bar{A}_v$,

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

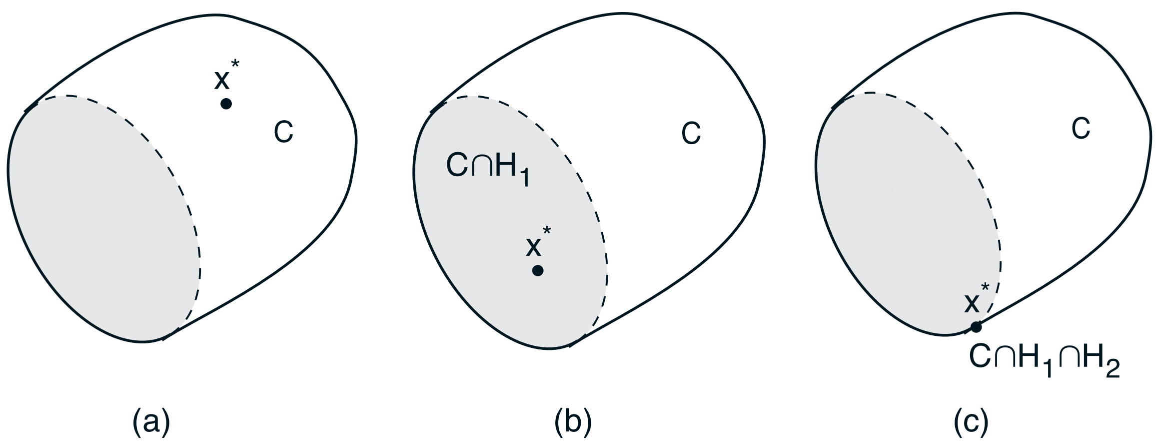
Thus, v, y, and z are all solutions of the system of n linearly independent equations

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

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

EXTREME POINTS AND CONCAVE MINIMIZATION

- 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.



Proof (abbreviated): If $x^* \in \mathrm{ri}(C)$ [see (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^*.

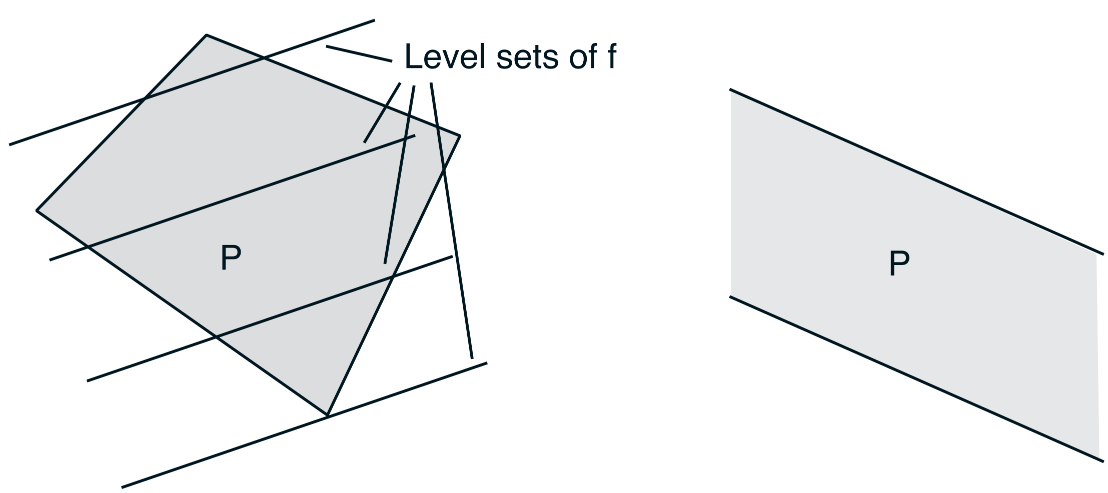
If $x^* \in \mathrm{ri}(C \cap H_1)$ [see (b)], then f must be constant over $C \cap H_1$, so it attains a minimum at an extreme point $C \cap H_1$. This optimal extreme point is also an extreme point of C. If $x^* \notin \mathrm{ri}(C \cap H_1)$, there is a hyperplane H_2 supporting $C \cap H_1$ through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



(a) (b)

EXTREME POINTS AND 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

- Polyhedral aspects of duality
- Hyperplane proper polyhedral separation
- Min Common/Max Crossing Theorem under polyhedral assumptions
- Nonlinear Farkas Lemma
- Application to convex programming

HYPERPLANE PROPER POLYHEDRAL SEPARATION

- Recall that two convex sets C and P such that

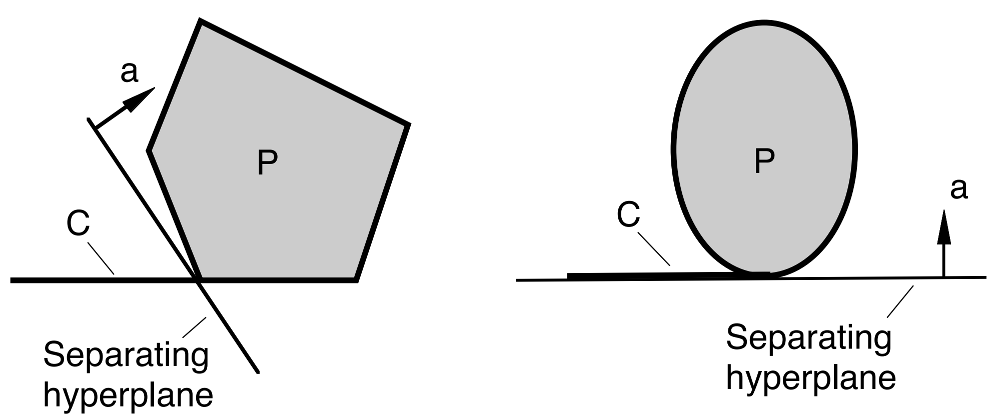
$$\mathrm{ri}(C) \cap \mathrm{ri}(P) = \emptyset$$

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 non-polyhedral 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 COMMON/MAX CROSSING TH. - SIMPLE

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

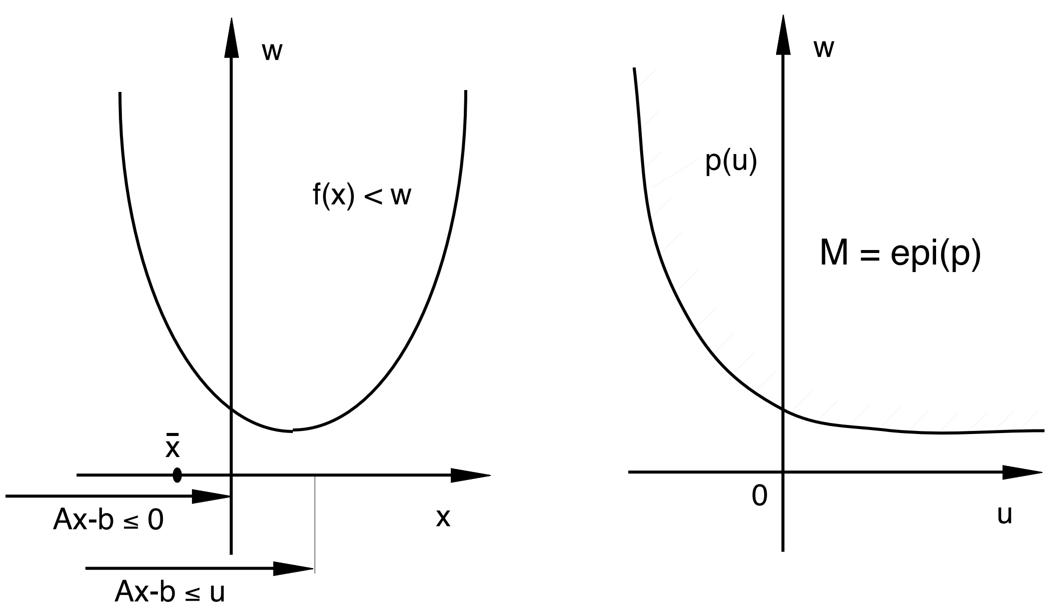
  (1) The set M is defined in terms of a convex function $f : \Re^m \mapsto (-\infty, \infty]$, an $r \times m$ matrix A, and a vector $b \in \Re^r$:

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

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

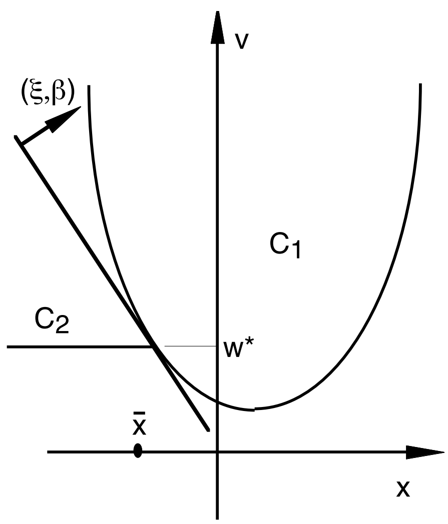
Then $q^* = w^*$ and there is a $\mu \ge 0$ with $q(\mu) = q^*$.

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



PROOF

- Consider the disjoint convex sets



$$C_1 = \Big\{(x, v) \mid f(x) < v\Big\}$$

$$C_2 = \Big\{(x, w^*) \mid Ax - b \le 0\Big\}$$

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

$$\beta w^* + \xi' z \le \beta v + \xi' x, \ \forall\, (x, v) \in C_1, \ \forall\, z \text{ with } Az - b \le 0,$$

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

Because of the relative interior point, $\beta \neq 0$, so we may assume that $\beta = 1$. Hence

$$\sup_{Az-b\le 0} \big\{w^* + \xi' z\big\} \le \inf_{(x,w)\in \text{epi}(f)} \big\{w + \xi' x\big\}$$

The LP on the left has an optimal solution z^*.

PROOF (CONTINUED)

- Let a_j' be the rows of A, and $\overline{J} = \{j \mid a_j' z^* = b_j\}$. We have

$$\xi' y \le 0, \qquad \forall\, y \text{ with } a_j' y \le 0,\ \forall\, j \in \overline{J},$$

so by Farkas' Lemma, there exist $\mu_j \ge 0$, $i \in \overline{J}$, such that $\xi = \sum_{j \in \overline{J}} \mu_j a_j$. Defining $\mu_j = 0$ for $j \notin \overline{J}$, we have

$$\xi = A'\mu \text{ and } \mu'(Az^* - b) = 0, \quad \text{so } \xi' z^* = \mu' b$$

- Hence from $w^* + \xi' z^* \le \inf_{(x,w)\in \mathrm{epi}(f)} \big\{ w + \xi' x \big\}$,

$$\begin{aligned} w^* &\le \inf_{(x,w)\in \mathrm{epi}(f)} \Big\{ w + \mu'(Ax - b) \Big\} \\ &\le \inf_{\substack{(x,w)\in \mathrm{epi}(f),\\ Ax - b \le u}} \{ w + \mu'(Ax - b) \} \\ &\le \inf_{\substack{(x,w)\in \mathrm{epi}(f),\ u \in \Re^n\\ Ax - b \le u}} \{ w + \mu' u \} \\ &= \inf_{(u,w)\in M} \{ w + \mu' u \} = q(\mu) \le q^*. \end{aligned}$$

Since generically $q^* \le w^*$, it follows that $q(\mu) = q^* = w^*$. **Q.E.D.**

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

$$f(x) \geq 0, \qquad \forall\, x \in F = \big\{x \in C \mid g(x) \leq 0\big\},$$

and one of the following two conditions holds:

(1) 0 is in the relative interior of the set $D = \big\{u \mid g(x) \leq u \text{ for some } x \in C\big\}$.

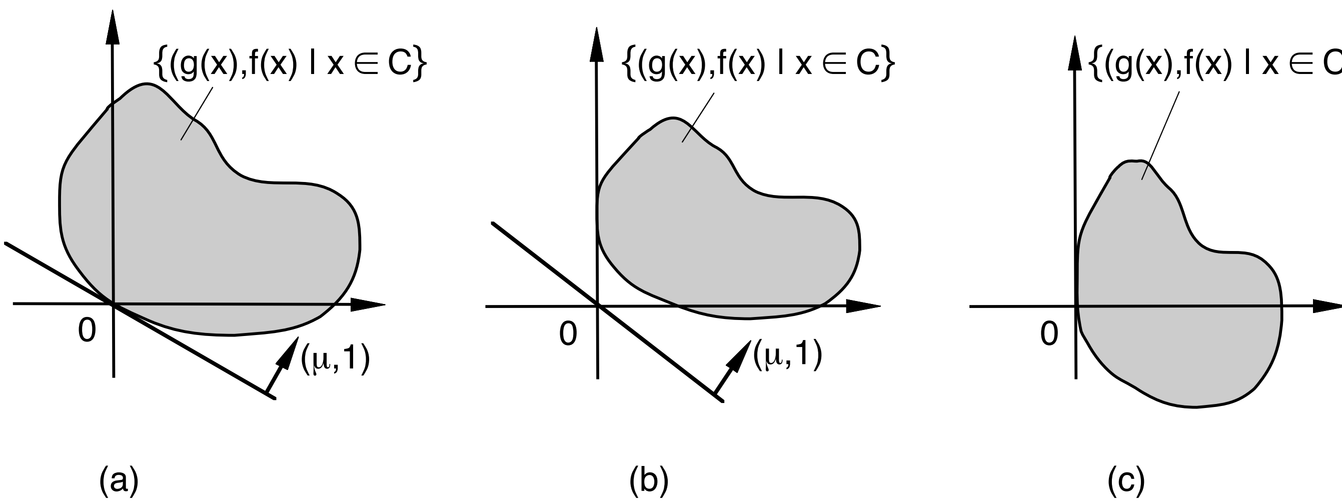
(2) The functions g_j, $j = 1, \ldots, r$, are affine, and F contains a relative interior point of C.

Then, there exist scalars $\mu_j^* \geq 0$, $j = 1, \ldots, r$, s. t.

$$f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \geq 0, \qquad \forall\, x \in C$$

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

VISUALIZATION OF NONLINEAR FARKAS' LEMMA



- Assuming that for all $x \in C$ with $g(x) \leq 0$, we have $f(x) \geq 0$, etc.

- 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.

PROOF OF NONLINEAR FARKAS' LEMMA

- Apply Min Common/Max Crossing to

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

- Under condition (1), Min Common/Max Crossing Theorem II applies: $0 \in \mathrm{ri}(D)$, where

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

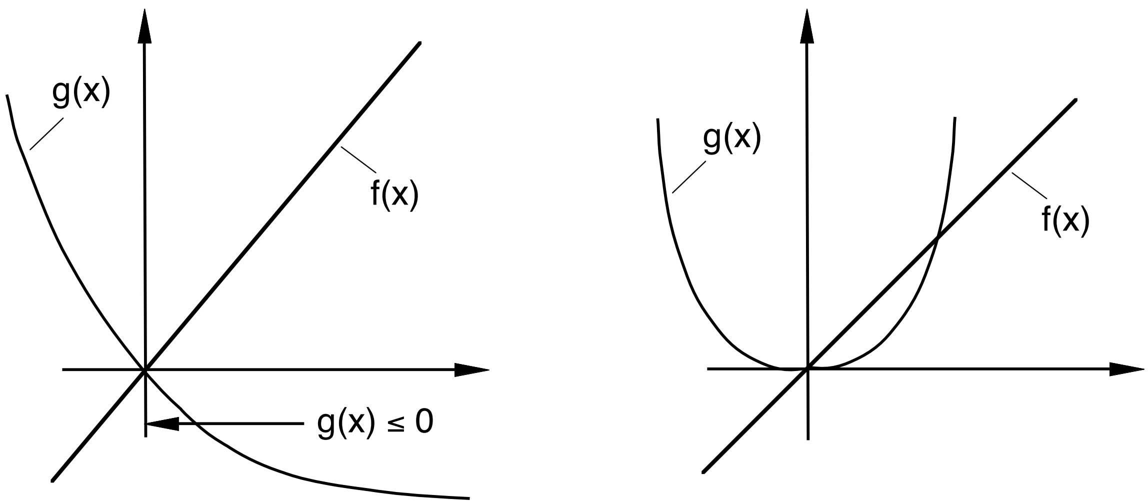
- Under condition (2), Min Common/Max Crossing Theorem III applies: $g(x) \le 0$ can be written as $Ax - b \le 0$.

- Hence for some μ^*, we have $w^* = \sup_\mu q(\mu) = q(\mu^*)$, where $q(\mu) = \inf_{(u,w)\in M}\{w + \mu' u\}$. Using the definition of M,

$$q(\mu) = \begin{cases} \inf_{x\in C}\big\{f(x) + \sum_{j=1}^r \mu_j g_j(x)\big\} & \text{if } \mu \ge 0, \\ -\infty & \text{otherwise,} \end{cases}$$

so $\mu^* \ge 0$ and $\inf_{x\in C}\Big\{f(x) + \sum_{j=1}^r \mu_j^* g_j(x)\Big\} = w^* \ge 0.$

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.

APPLICATION TO CONVEX PROGRAMMING

Consider the problem

minimize $f(x)$

subject to $x \in F = \{x \in C \mid g_j(x) \leq 0,\ j = 1, \ldots, r\}$

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

- 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

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

Since $F \subset C$ and $\mu_j^* g_j(x) \leq 0$ for all $x \in F$,

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

Thus equality holds throughout, and we have

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

CONVEX PROGRAMMING DUALITY - OUTLINE

- Define the dual function

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

and the dual problem $\max_{\mu \geq 0} \; q(\mu)$.

- Note that for all $\mu \geq 0$ and $x \in C$ with $g(x) \leq 0$

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

Therefore, we have the $weak \; duality$ relation

$$q^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in C,\, g(x) \leq 0} f(x) = f^*$$

- If we can use Farkas' Lemma, there exists $\mu^* \geq 0$ that solves the dual problem and $q^* = f^*$.

- This is so if (1) there exists $\overline{x} \in C$ with $g_j(\overline{x}) < 0$ for all j, or (2) the constraint functions g_j are affine and there is a feasible point in $\mathrm{ri}(C)$.

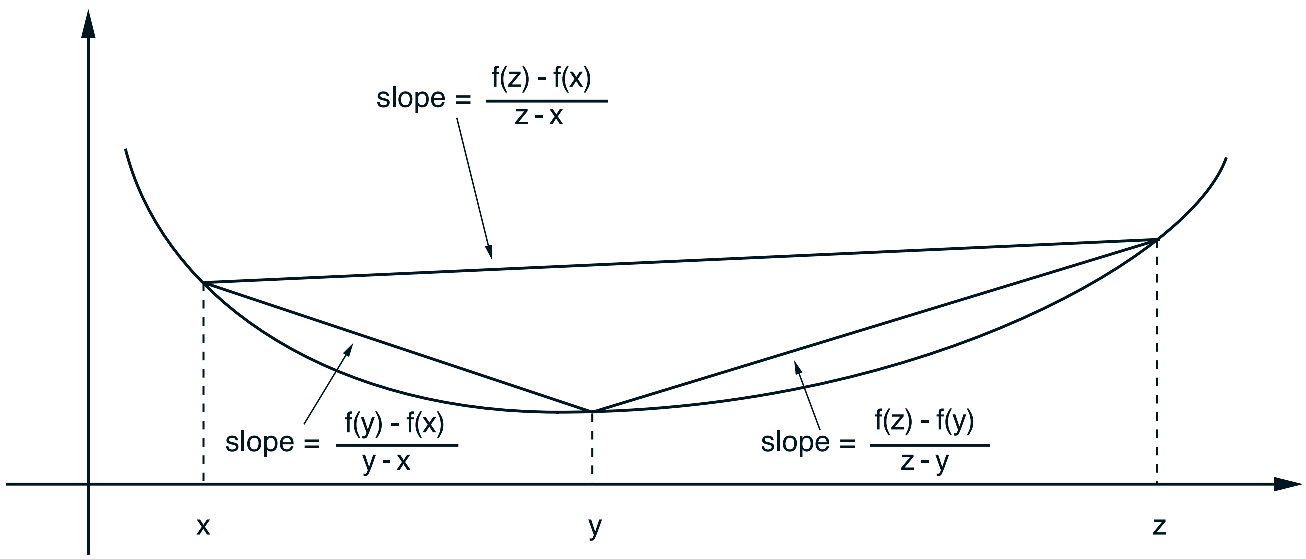
LECTURE 13

LECTURE OUTLINE

- Directional derivatives of one-dimensional convex functions
- Directional derivatives of multi-dimensional convex functions
- Subgradients and subdifferentials
- Properties of subgradients

ONE-DIMENSIONAL DIRECTIONAL DERIVATIVES

- Three slopes relation for a convex $f : \Re \mapsto \Re$:



$$\frac{f(y)-f(x)}{y-x} \leq \frac{f(z)-f(x)}{z-x} \leq \frac{f(z)-f(y)}{z-y}$$

- Right and left directional derivatives exist

$$f^+(x) = \lim_{\alpha \downarrow 0} \frac{f(x+\alpha)-f(x)}{\alpha}$$

$$f^-(x) = \lim_{\alpha \downarrow 0} \frac{f(x)-f(x-\alpha)}{\alpha}$$

MULTI-DIMENSIONAL DIRECTIONAL DERIVATIVES

- For a convex $f : \Re^n \mapsto \Re$

$$f'(x; y) = \lim_{\alpha \downarrow 0} \frac{f(x + \alpha y) - f(x)}{\alpha},$$

is the directional derivative at x in the direction y.

- Exists for all x and all directions.

- f is differentiable at x if $f'(x; y)$ is a linear function of y denoted by

$$f'(x; y) = \nabla f(x)' y,$$

where $\nabla f(x)$ is the gradient of f at x.

- Directional derivatives can be defined for extended real-valued convex functions, but we will not pursue this topic (see the textbook).

SUBGRADIENTS

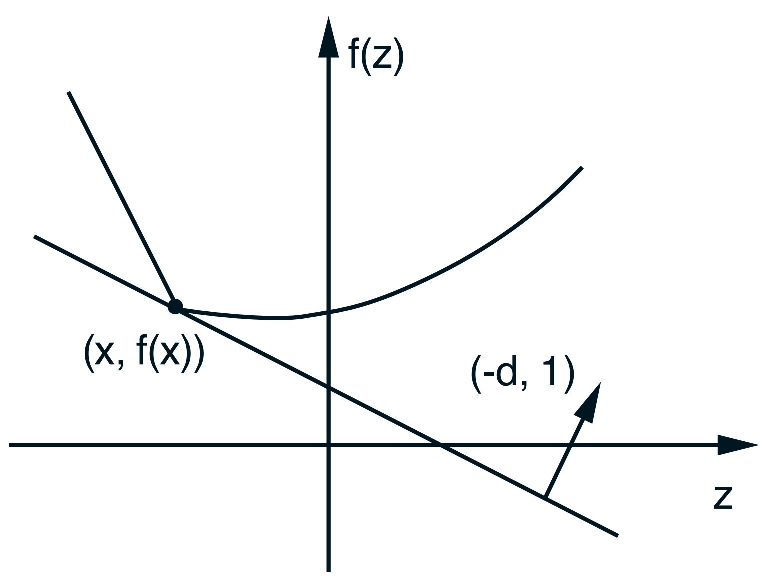
- Let $f : \Re^n \mapsto \Re$ be a convex function. A vector $d \in \Re^n$ is a *subgradient* of f at a point $x \in \Re^n$ if

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

- d is a subgradient if and only if

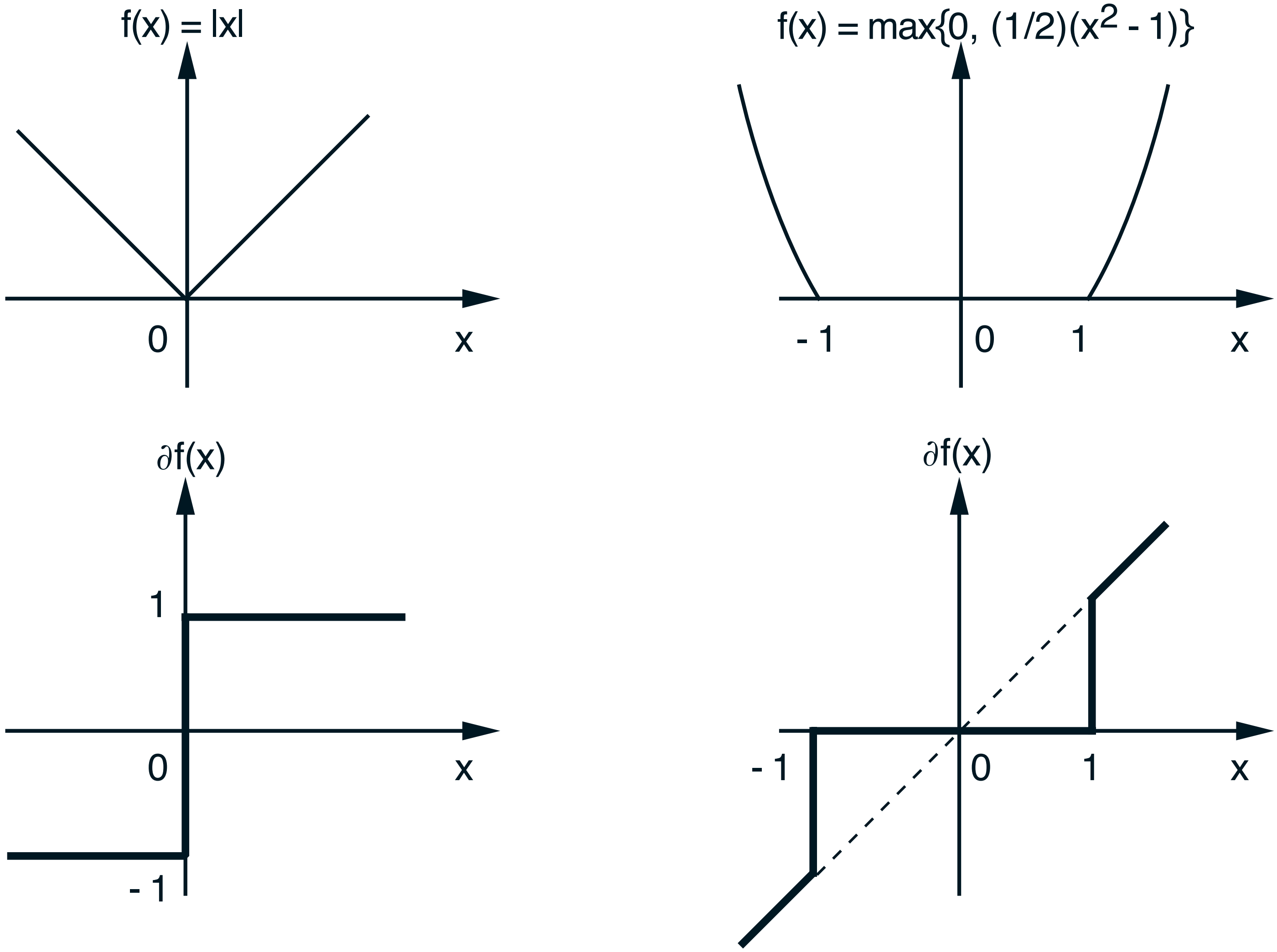
$$f(z) - z'd \geq f(x) - x'd, \qquad \forall\, z \in \Re^n$$

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



SUBDIFFERENTIAL

- The set of all subgradients of a convex function f at x is called the *subdifferential* of f at x, and is denoted by $\partial f(x)$.
- Examples of subdifferentials:



PROPERTIES OF SUBGRADIENTS I

- $\partial f(x)$ is nonempty, convex, and compact.

Proof: Consider the min common/max crossing framework with

$$M = \big\{(u, w) \mid u \in \Re^n,\ f(x+u) \leq w\big\}$$

Min common value: $w^* = f(x)$. Crossing value function is $q(\mu) = \inf_{(u,w)\in M}\{w + \mu' u\}$. We have $w^* = q^* = q(\mu)$ iff $f(x) = \inf_{(u,w)\in M}\{w + \mu' u\}$, or

$$f(x) \leq f(x+u) + \mu' u, \qquad \forall\, u \in \Re^n$$

Thus, the set of optimal solutions of the max crossing problem is precisely $-\partial f(x)$. Use the Min Common/Max Crossing Theorem II: since the set

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

contains the origin in its interior, the set of optimal solutions of the max crossing problem is nonempty, convex, and compact. **Q.E.D.**

PROPERTIES OF SUBGRADIENTS II

- For every $x \in \Re^n$, we have

$$f'(x;y) = \max_{d \in \partial f(x)} y'd, \qquad \forall\, y \in \Re^n$$

- f is differentiable at x with gradient $\nabla f(x)$, if and only if it has $\nabla f(x)$ as its unique subgradient at x.

- If $f = \alpha_1 f_1 + \cdots + \alpha_m f_m$, where the $f_j : \Re^n \mapsto \Re$ are convex and $\alpha_j > 0$,

$$\partial f(x) = \alpha_1 \partial f_1(x) + \cdots + \alpha_m \partial f_m(x)$$

- Chain Rule: If $F(x) = f(Ax)$, where A is a matrix,

$$\partial F(x) = A' \partial f(Ax) = \big\{ A'g \mid g \in \partial f(Ax) \big\}$$

- Generalizes to functions $F(x) = g\big(f(x)\big)$, where g is smooth.

ADDITIONAL RESULTS ON SUBGRADIENTS

- **Danskin's Theorem**: Let Z be compact, and $\phi : \Re^n \times Z \mapsto \Re$ be continuous. Assume that $\phi(\cdot, z)$ is convex and differentiable for all $z \in Z$. Then the function $f : \Re^n \mapsto \Re$ given by

$$f(x) = \max_{z \in Z} \phi(x, z)$$

is convex and for all x

$$\partial f(x) = \mathrm{conv}\big\{\nabla_x \phi(x, z) \mid z \in Z(x)\big\}$$

- The subdifferential of an extended real valued convex function $f : \Re^n \mapsto (-\infty, \infty]$ is defined by

$$\partial f(x) = \big\{d \mid f(z) \geq f(x) + (z - x)'d,\ \forall\ z \in \Re^n\big\}$$

- $\partial f(x)$, is closed but may be empty at relative boundary points of $\mathrm{dom}(f)$, and may be unbounded.

- $\partial f(x)$ is nonempty at all $x \in \mathrm{ri}\big(\mathrm{dom}(f)\big)$, and it is compact if and only if $x \in \mathrm{int}\big(\mathrm{dom}(f)\big)$. The proof again is by Min Common/Max Crossing II.

LECTURE 14

LECTURE OUTLINE

- Conical approximations
- Cone of feasible directions
- Tangent and normal cones
- Conditions for optimality

- A basic necessary condition:
  - If x^* minimizes a function $f(x)$ over $x \in X$, then for every $y \in \Re^n$, $\alpha^* = 0$ minimizes $g(\alpha) \equiv f(x + \alpha y)$ over the line subset

$$\{\alpha \mid x + \alpha y \in X\}$$

- Special cases of this condition (f: differentiable):
  - $X = \Re^n$: $\nabla f(x^*) = 0$.
  - X is convex: $\nabla f(x^*)'(x - x^*) \geq 0$, $\forall\, x \in X$.
- We will aim for more general conditions.

CONE OF FEASIBLE DIRECTIONS

- Consider a subset X of $\Re^n$ and a vector $x \in X$.

- A vector $y \in \Re^n$ is a *feasible direction* of X at x if there exists an $\overline{\alpha} > 0$ such that $x + \alpha y \in X$ for all $\alpha \in [0, \overline{\alpha}]$.

- The set of all feasible directions of X at x is denoted by $F_X(x)$.

- $F_X(x)$ is a cone containing the origin. It need not be closed or convex.

- If X is convex, $F_X(x)$ consists of the vectors of the form $\alpha(\overline{x} - x)$ with $\alpha > 0$ and $\overline{x} \in X$.

- Easy optimality condition: If x^* minimizes a differentiable function $f(x)$ over $x \in X$, then

$$\nabla f(x^*)' y \geq 0, \qquad \forall\, y \in F_X(x^*)$$

- Difficulty: The condition may be vacuous because there may be no feasible directions (other than 0), e.g., take X to be the boundary of a circle.

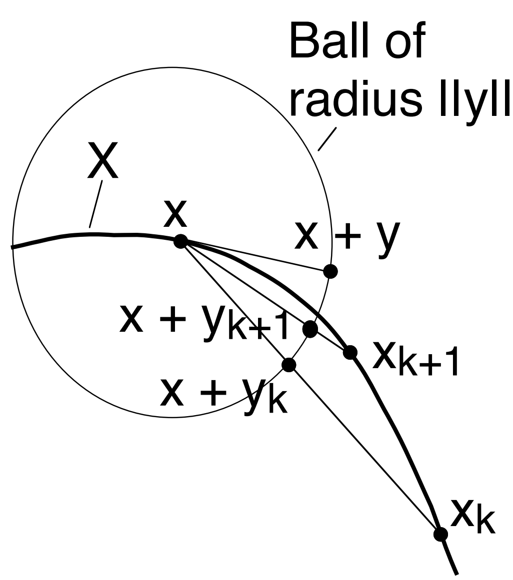
TANGENT CONE

- Consider a subset X of $\Re^n$ and a vector $x \in X$.

- A vector $y \in \Re^n$ is said to be a *tangent* of X at x if either $y = 0$ or there exists a sequence $\{x_k\} \subset X$ such that $x_k \neq x$ for all k and

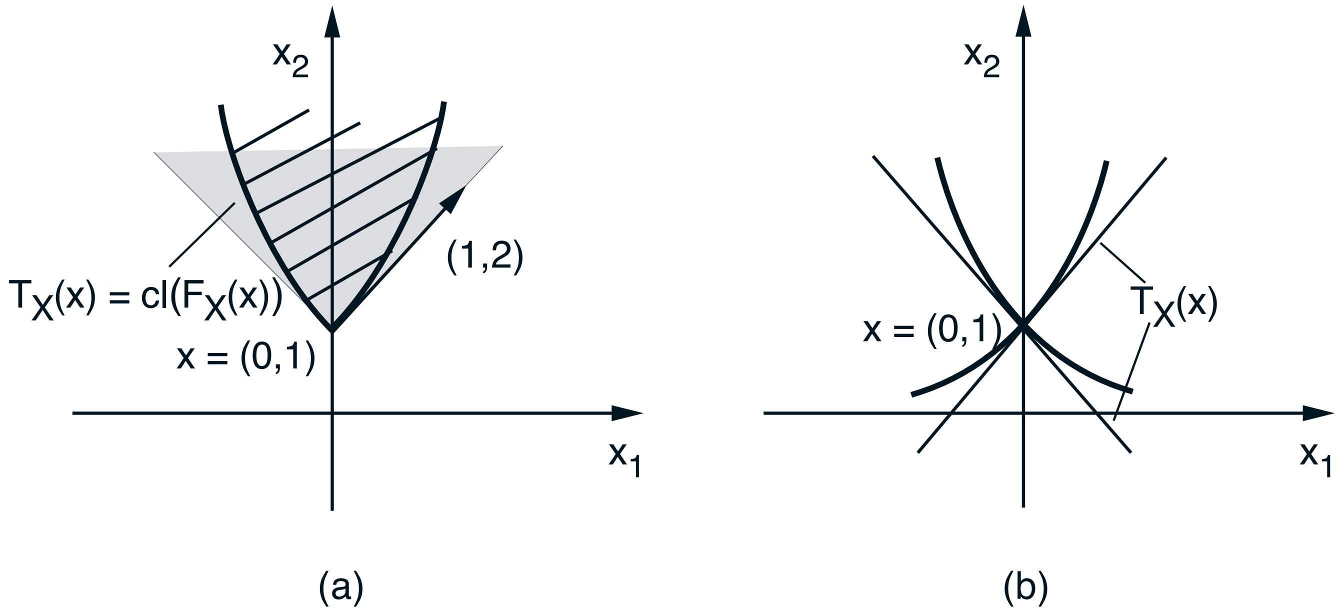
$$x_k \to x, \qquad \frac{x_k - x}{\|x_k - x\|} \to \frac{y}{\|y\|}$$

- The set of all tangents of X at x is called the *tangent cone* of X at x, and is denoted by $T_X(x)$.



- y is a tangent of X at x iff there exists $\{x_k\} \subset X$ with $x_k \to x$, and a positive scalar sequence $\{\alpha_k\}$ such that $\alpha_k \to 0$ and $(x_k - x)/\alpha_k \to y$.

EXAMPLES



- In (a), X is convex: The tangent cone $T_X(x)$ is equal to the closure of the cone of feas. directions $F_X(x)$.
- In (b), X is nonconvex: $T_X(x)$ is closed but not convex, while $F_X(x)$ consists of just the zero vector.
- In general, $F_X(x) \subset T_X(x)$.
- For X: polyhedral, $F_X(x) = T_X(x)$.

RELATION OF CONES

- Let X be a subset of $\Re^n$ and let x be a vector in X. The following hold.

  (a) $T_X(x)$ is a closed cone.

  (b) $\mathrm{cl}\big(F_X(x)\big) \subset T_X(x)$.

  (c) If X is convex, then $F_X(x)$ and $T_X(x)$ are convex, and we have

$$\mathrm{cl}\big(F_X(x)\big) = T_X(x)$$

Proof: (a) Let $\{y_k\}$ be a sequence in $T_X(x)$ that converges to some $y \in \Re^n$. We show that $y \in T_X(x)$ ...

(b) Every feasible direction is a tangent, so $F_X(x) \subset T_X(x)$. Since by part (a), $T_X(x)$ is closed, the result follows.

(c) Since X is convex, the set $F_X(x)$ consists of the vectors of the form $\alpha(\overline{x} - x)$ with $\alpha > 0$ and $\overline{x} \in X$. Verify definition of convexity ...

NORMAL CONE

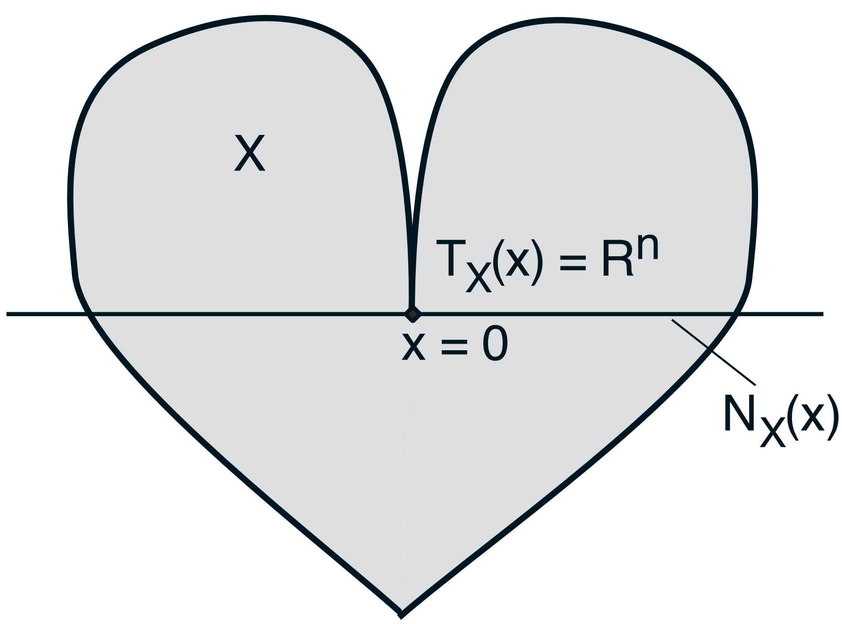
- Consider subset X of $\Re^n$ and a vector $x \in X$.

- A vector $z \in \Re^n$ is said to be a $normal$ of X at x if there exist sequences $\{x_k\} \subset X$ and $\{z_k\}$ with

$$x_k \to x, \qquad z_k \to z, \qquad z_k \in T_X(x_k)^*, \quad \forall\, k$$

- The set of all normals of X at x is called the $normal\ cone$ of X at x and is denoted by $N_X(x)$.

- Example:



- $N_X(x)$ is "usually equal" to the polar $T_X(x)^*$, but may differ at points of "discontinuity" of $T_X(x)$.

RELATION OF NORMAL AND POLAR CONES

- We have $T_X(x)^* \subset N_X(x)$.

- When $N_X(x) = T_X(x)^*$, we say that X is $regular$ at x.

- If X is convex, then for all $x \in X$, we have

$$z \in T_X(x)^* \text{ if and only if } z'(\overline{x}-x) \leq 0, \quad \forall\, \overline{x} \in X$$

Furthermore, X is regular at all $x \in X$. In particular, we have

$$T_X(x)^* = N_X(x), \qquad T_X(x) = N_X(x)^*$$

- Note that convexity of $T_X(x)$ does not imply regularity of X at x.

- Important fact in nonsmooth analysis: If X is closed and regular at x, then

$$T_X(x) = N_X(x)^*.$$

In particular, $T_X(x)$ is convex.

OPTIMALITY CONDITIONS I

- Let $f : \Re^n \mapsto \Re$ be a smooth function. If x^* is a local minimum of f over a set $X \subset \Re^n$, then

$$\nabla f(x^*)'y \geq 0, \qquad \forall\, y \in T_X(x^*)$$

Proof: Let $y \in T_X(x^*)$ with $y \neq 0$. Then, there exist $\{\xi_k\} \subset \Re$ and $\{x_k\} \subset X$ such that $x_k \neq x^*$ for all k, $\xi_k \to 0$, $x_k \to x^*$, and

$$(x_k - x^*)/\|x_k - x^*\| = y/\|y\| + \xi_k$$

By the Mean Value Theorem, we have for all k

$$f(x_k) = f(x^*) + \nabla f(\tilde{x}_k)'(x_k - x^*),$$

where $\tilde{x}_k$ is a vector that lies on the line segment joining x_k and x^*. Combining these equations,

$$f(x_k) = f(x^*) + (\|x_k - x^*\|/\|y\|)\nabla f(\tilde{x}_k)'y_k,$$

where $y_k = y + \|y\|\xi_k$. If $\nabla f(x^*)'y < 0$, since $\tilde{x}_k \to x^*$ and $y_k \to y$, for sufficiently large k, $\nabla f(\tilde{x}_k)'y_k < 0$ and $f(x_k) < f(x^*)$. This contradicts the local optimality of x^*.

OPTIMALITY CONDITIONS II

• Let $f : \Re^n \mapsto \Re$ be a convex function. A vector x^* minimizes f over a convex set X if and only if there exists a subgradient $d \in \partial f(x^*)$ such that

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

Proof: If for some $d \in \partial f(x^*)$ and all $x \in X$, we have $d'(x - x^*) \geq 0$, then, from the definition of a subgradient we have $f(x) - f(x^*) \geq d'(x - x^*)$ for all $x \in X$. Hence $f(x) - f(x^*) \geq 0$ for all $x \in X$.

Conversely, suppose that x^* minimizes f over X. Then, x^* minimizes f over the closure of X, and we have

$$f'(x^*; x - x^*) = \sup_{d \in \partial f(x^*)} d'(x - x^*) \geq 0, \; \forall\, x \in \mathrm{cl}(X)$$

Therefore,

$$\inf_{x \in \mathrm{cl}(X) \cap \{z \mid \|z - x^*\| \leq 1\}} \; \sup_{d \in \partial f(x^*)} d'(x - x^*) = 0$$

Apply the saddle point theorem to conclude that "infsup=supinf" and that the supremum is attained by some $d \in \partial f(x^*)$.

OPTIMALITY CONDITIONS III

- Let x^* be a local minimum of a function $f : \Re^n \mapsto \Re$ over a subset X of $\Re^n$. Assume that the tangent cone $T_X(x^*)$ is convex, and that f has the form

$$f(x) = f_1(x) + f_2(x),$$

where f_1 is convex and f_2 is smooth. Then

$$-\nabla f_2(x^*) \in \partial f_1(x^*) + T_X(x^*)^*$$

- The convexity assumption on $T_X(x^*)$ (which is implied by regularity) is essential in general.

- **Example:** Consider the subset of $\Re^2$

$$X = \big\{(x_1, x_2) \mid x_1 x_2 = 0\big\}$$

Then $T_X(0)^* = \{0\}$. Take f to be any convex non-differentiable function for which $x^* = 0$ is a global minimum over X, but $x^* = 0$ is not an unconstrained global minimum. Such a function violates the necessary condition.

LECTURE 15

LECTURE OUTLINE

- Intro to Lagrange multipliers
- Enhanced Fritz John Theory

- Problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad h_1(x) = 0, \ldots, h_m(x) = 0 \\ &\qquad\qquad\qquad\quad g_1(x) \le 0, \ldots, g_r(x) \le 0 \end{aligned}$$

where f, h_i, $g_j : \Re^n \mapsto \Re$ are smooth functions, and X is a nonempty closed set

- Main issue: What is the structure of the constraint set that guarantees the existence of Lagrange multipliers?

DEFINITION OF LAGRANGE MULTIPLIER

- Let x^* be a local minimum. Then $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ are Lagrange multipliers if

$$\mu_j^* \geq 0, \qquad \forall\, j = 1, \ldots, r,$$

$$\mu_j^* = 0, \qquad \forall\, j \text{ with } g_j(x^*) < 0,$$

$$\nabla_x L(x^*, \lambda^*, \mu^*)' y \geq 0, \quad \forall\, y \in T_X(x^*),$$

where L is the Lagrangian function

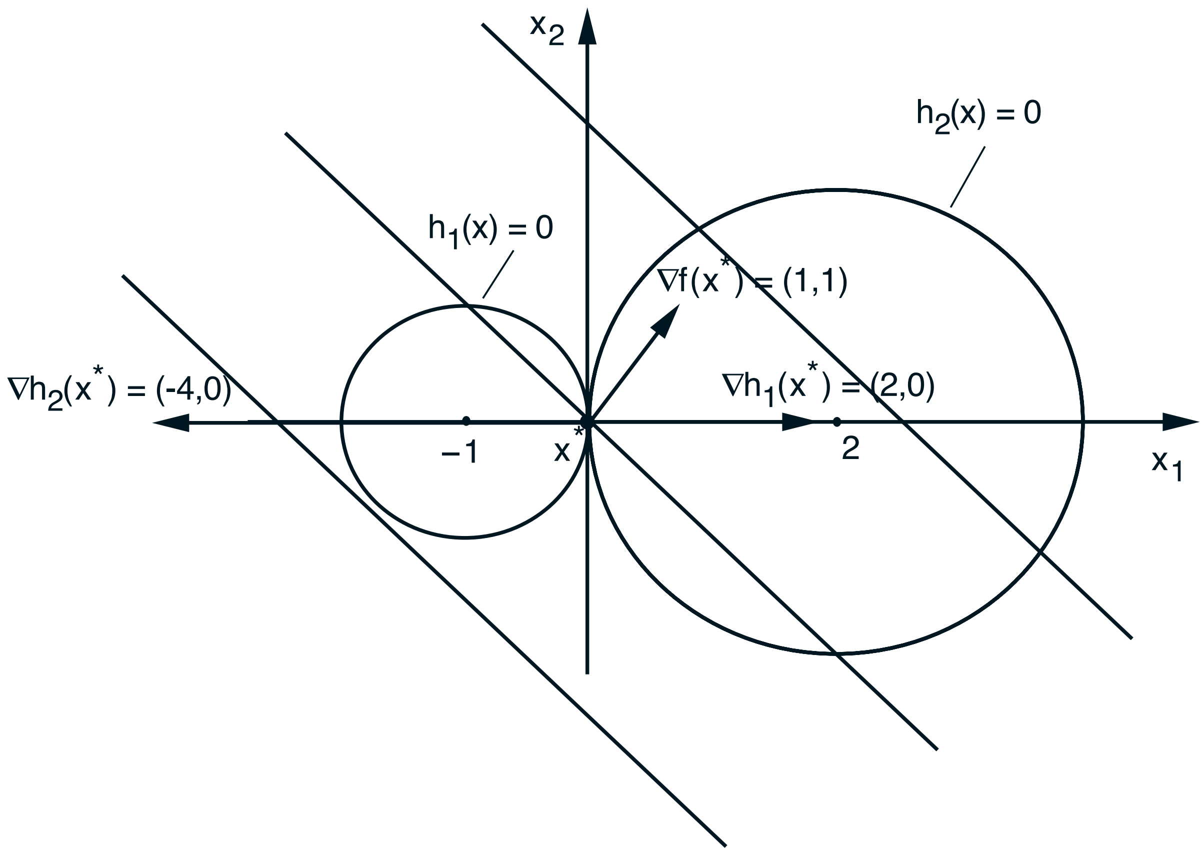
$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{j=1}^{r} \mu_j g_j(x)$$

- Note: When $X = \Re^n$, then $T_X(x^*) = \Re^n$ and the Lagrangian stationarity condition becomes

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0$$

EXAMPLE OF NONEXISTENCE

OF A LAGRANGE MULTIPLIER



Minimize

$$f(x) = x_1 + x_2$$

subject to the two constraints

$$h_1(x) = (x_1 + 1)^2 + x_2^2 - 1 = 0,$$

$$h_2(x) = (x_1 - 2)^2 + x_2^2 - 4 = 0$$

CLASSICAL ANALYSIS

- Necessary condition at a local minimum x^*:

$$-\nabla f(x^*) \in T(x^*)^*$$

- Assume linear equality constraints only

$$h_i(x) = a_i' x - b_i, \qquad i = 1, \ldots, m,$$

- The tangent cone is

$$T(x^*) = \{y \mid a_i' y = 0,\ i = 1, \ldots, m\}$$

and its polar, $T(x^*)^*$, is the range space of the matrix having as columns the a_i, so for some scalars λ_i^*

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* a_i = 0$$

QUASIREGULARITY

- If the h_i are nonlinear AND

$$T(x^*) = \{y \mid \nabla h_i(x^*)'y = 0,\ i = 1, \ldots, m\} \quad (*)$$

similarly, for some scalars λ_i^*, we have

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0$$

- Eq. $(*)$ (called *quasiregularity*) can be shown to hold if the $\nabla h_i(x^*)$ are linearly independent
- Extension to inequality constraints: If quasiregularity holds, i.e.,

$$T(x^*) = \{y \mid \nabla h_i(x^*)'y = 0,\ \nabla g_j(x^*)'y \le 0,\ \forall j \in A(x^*)\}$$

where $A(x^*) = \{j \mid g_j(x^*) = 0\}$, the condition $-\nabla f(x^*) \in T(x^*)^*$, by Farkas' lemma, implies $\mu_j^* = 0\ \forall\ j \notin A(x^*)$ and

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*) = 0$$

FRITZ JOHN THEORY

- Back to equality constraints. There are two possibilities:
  - Either $\nabla h_i(x^*)$ are linearly independent and

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0$$

  - or for some λ_i^* (not all 0)

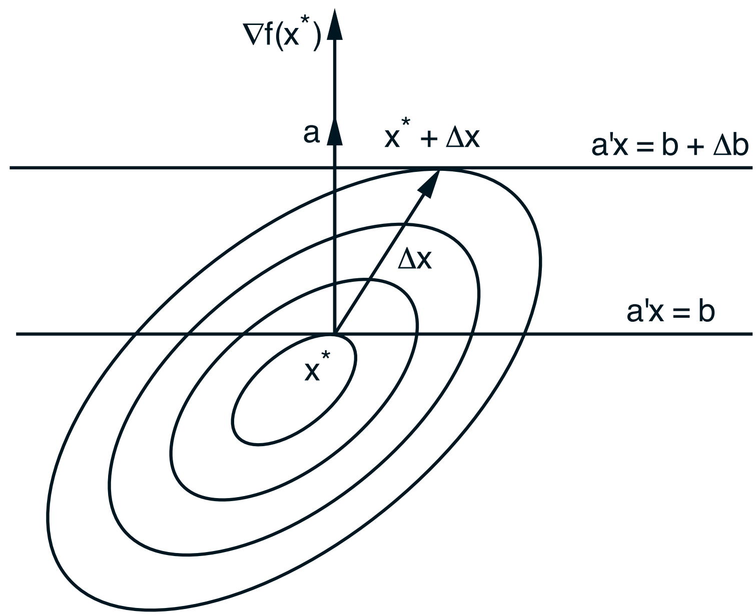
$$\sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0$$

- Combination of the two: There exist $\mu_0^* \geq 0$ and $\lambda_1^*, \ldots, \lambda_m^*$ (not all 0) such that

$$\mu_0^* \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) = 0$$

- Question now becomes: When is $\mu_0^* \neq 0$?

SENSITIVITY (SINGLE LINEAR CONSTRAINT)



- Perturb RHS of the constraint by Δb. The minimum is perturbed by Δx, where $a'\Delta x = \Delta b$.
- If λ^* is Lagrange multiplier, $\nabla f(x^*) = -\lambda^* a$,

$$\Delta \text{cost} = \nabla f(x^*)' \Delta x + o(\|\Delta x\|) = -\lambda^* a' \Delta x + o(\|\Delta x\|$$

- So $\Delta \text{cost} = -\lambda^* \Delta b + o(\|\Delta x\|)$, and up to first order we have

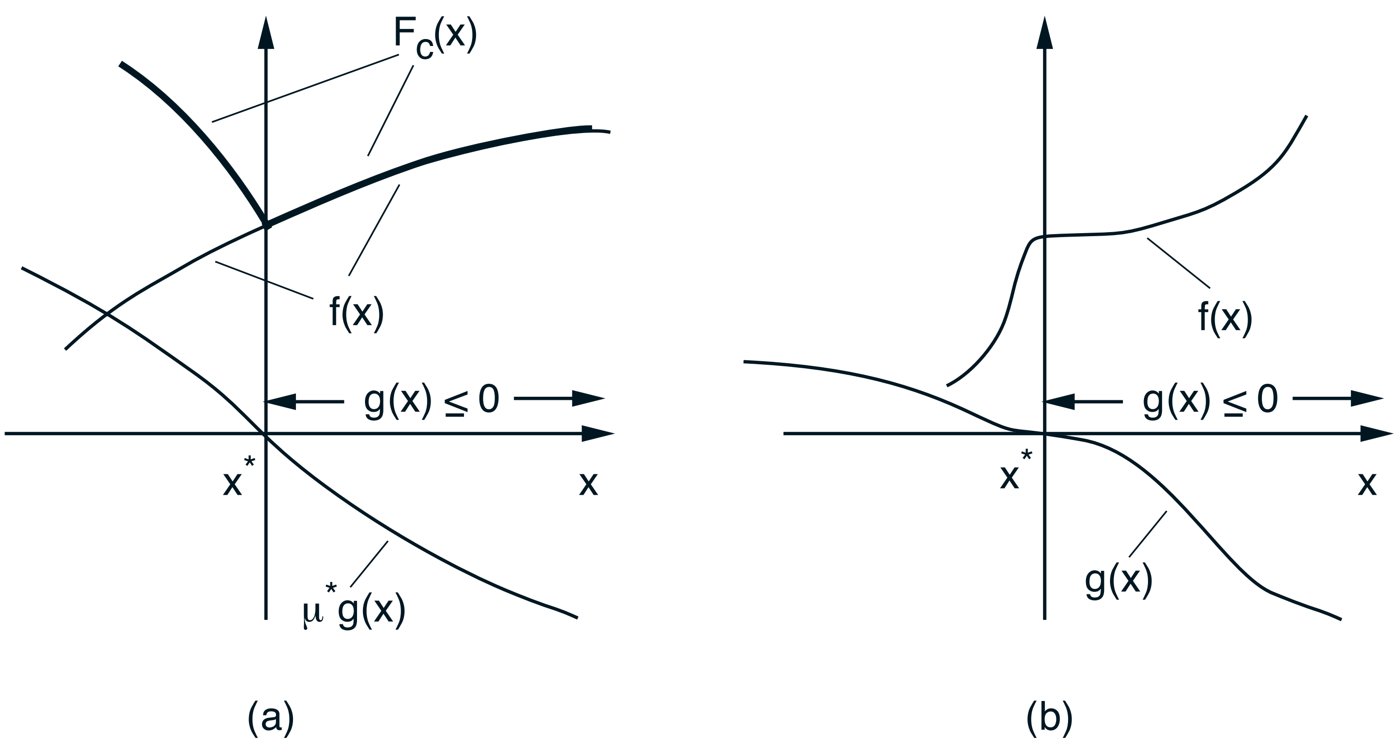
$$\lambda^* = -\frac{\Delta \text{cost}}{\Delta b}$$

EXACT PENALTY FUNCTIONS

- Consider

$$F_c(x) = f(x) + c\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right)$$

- A local min x^* of the constrained opt. problem is typically a local minimum of F_c, provided c is larger than some threshold value.



- Connection with Lagrange multipliers.

OUR APPROACH

- Abandon the classical approach – it does not work when $X \neq \Re^n$.

- Enhance the Fritz John conditions so that they become really useful.

- Show (under minimal assumptions) that when Lagrange multipliers exist, there exist some that are $informative$ in the sense that pick out the "important constraints" and have meaningful sensitivity interpretation.

- Use the notion of $constraint\ pseudonormality$ as the linchpin of a theory of constraint qualifications, and the connection with exact penalty functions.

- Make the connection with nonsmooth analysis notions such as regularity and the normal cone.

ENHANCED FRITZ JOHN CONDITIONS

If x^* is a local minimum, there exist $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$, satisfying the following:

(i) $$-\left(\mu_0^* \nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$$

(ii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^* \geq 0$ and not all 0

(iii) If

$$J = \{j \neq 0 \mid \mu_j^* > 0\}$$

is nonempty, there exists a sequence $\{x^k\} \subset X$ converging to x^* and such that for all k,

$$f(x^k) < f(x^*), \qquad g_j(x^k) > 0, \quad \forall\, j \in J,$$

$$g_j^+(x^k) = o\left(\min_{j \in J} g_j(x^k)\right), \quad \forall\, j \notin J$$

- The last condition is stronger than the classical

$$g_j(x^*) = 0, \qquad \forall\, j \in J$$

LECTURE 16

LECTURE OUTLINE

- Enhanced Fritz John Conditions
- Pseudonormality
- Constraint qualifications

- Problem

$$\begin{aligned} & \text{minimize} \;\; f(x) \\ & \text{subject to} \;\; x \in X, \;\; h_1(x) = 0, \ldots, h_m(x) = 0 \\ & \qquad\qquad\qquad\qquad g_1(x) \le 0, \ldots, g_r(x) \le 0 \end{aligned}$$

where f, h_i, $g_j : \Re^n \mapsto \Re$ are smooth functions, and X is a nonempty closed set.

- To simplify notation, we will often assume no equality constraints.

DEFINITION OF LAGRANGE MULTIPLIER

- Consider the Lagrangian function

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

Let x^* be a local minimum. Then λ^* and μ^* are Lagrange multipliers if for all j,

$$\mu_j^* \geq 0, \qquad \mu_j^* = 0 \text{ if } g_j(x^*) < 0,$$

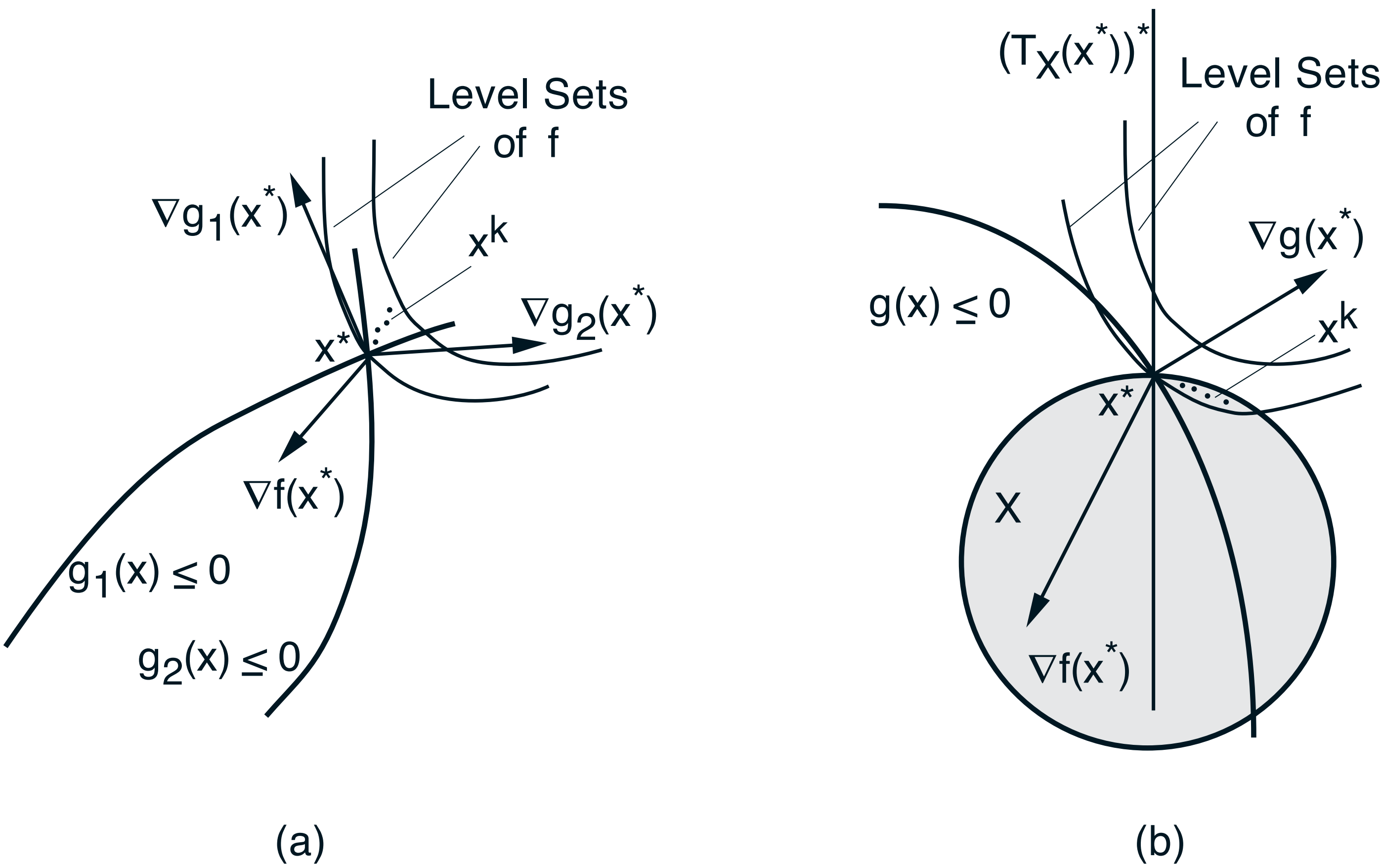
and the Lagrangian is stationary at x^*, i.e., has ≥ 0 slope along the tangent directions of X at x^* (feasible directions in case where X is convex):

$$\nabla_x L(x^*, \lambda^*, \mu^*)' y \geq 0, \quad \forall\, y \in T_X(x^*)$$

- **Note 1:** If $X = \Re^n$, Lagrangian stationarity means $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$.

- **Note 2:** If X is convex and the Lagrangian is convex in x for $\mu \geq 0$, Lagrangian stationarity means that $L(\cdot, \lambda^*, \mu^*)$ is minimized over $x \in X$ at x^*.

ILLUSTRATION OF LAGRANGE MULTIPLIERS



- (a) Case where $X = \Re^n$: $-\nabla f(x^*)$ is in the cone generated by the gradients $\nabla g_j(x^*)$ of the active constraints.
- (b) Case where $X \neq \Re^n$: $-\nabla f(x^*)$ is in the cone generated by the gradients $\nabla g_j(x^*)$ of the active constraints and the polar cone $T_X(x^*)^*$.

ENHANCED FRITZ JOHN NECESSARY CONDITIONS

If x^* is a local minimum, there exist $\mu_0^*, \mu_1^*, \ldots, \mu_r^*$, satisfying the following:

(i) $$-\left(\mu_0^* \nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$$

(ii) $\mu_0^*, \mu_1^*, \ldots, \mu_r^* \geq 0$ and not all 0

(iii) If

$$J = \{j \neq 0 \mid \mu_j^* > 0\}$$

is nonempty, there exists a sequence $\{x^k\} \subset X$ converging to x^* and such that for all k,

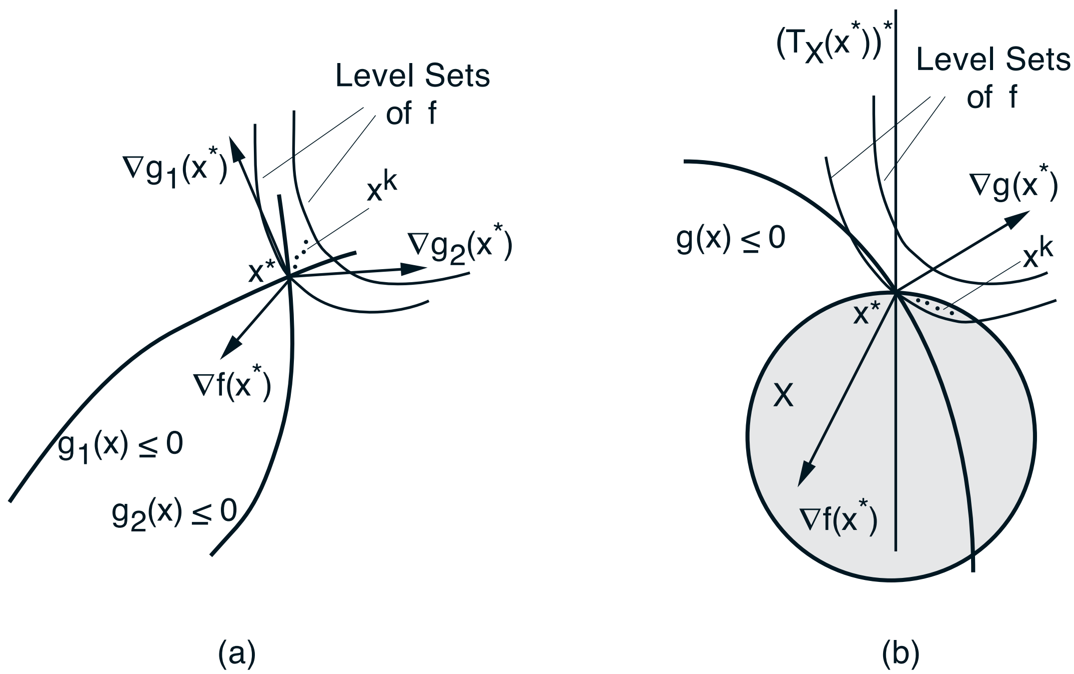
$$f(x^k) < f(x^*), \qquad g_j(x^k) > 0, \quad \forall\, j \in J,$$

$$g_j^+(x^k) = o\left(\min_{j \in J} g_j(x^k)\right), \quad \forall\, j \notin J$$

• **Note:** In the classical Fritz John theorem, condition (iii) is replaced by the weaker condition that

$$\mu_j^* = 0, \qquad \forall\, j \text{ with } g_j(x^*) < 0$$

GEOM. INTERPRETATION OF LAST CONDITION



- **Note:** Multipliers satisfying the classical Fritz John conditions may not satisfy condition (iii).

- **Example:** Start with any problem $\min_{h(x)=0} f(x)$ that has a local min-Lagrange multiplier pair (x^*, λ^*) with $\nabla f(x^*) \neq 0$ and $\nabla h(x^*) \neq 0$. Convert it to the problem $\min_{h(x)\leq 0,\, -h(x)\leq 0} f(x)$. The (μ_0^*, μ^*) satisfying the classical FJ conditions:

$$\mu_0^* = 0,\ \mu_1^* = \mu_2^* \neq 0 \text{ or } \mu_0^* > 0,\ (\mu_0^*)^{-1}(\mu_1^* - \mu_2^*) = \lambda^*$$

The enhanced FJ conditions are satisfied only for

$$\mu_0^* > 0,\ \mu_1^* = \lambda^*/\mu_0^*,\ \mu_2^* = 0 \text{ or } \mu_0^* > 0,\ \mu_1^* = 0,\ \mu_2^* = -\lambda^*/\mu_0^*$$

PROOF OF ENHANCED FJ THEOREM

• We use a quadratic penalty function approach. Let $g_j^+(x) = \max\{0, g_j(x)\}$, and for each k, consider

$$\min_{X \cap S} F^k(x) \equiv f(x) + \frac{k}{2}\sum_{j=1}^{r}\big(g_j^+(x)\big)^2 + \frac{1}{2}||x - x^*||^2$$

where $S = \{x \mid ||x - x^*|| \leq \epsilon\}$, and $\epsilon > 0$ is such that $f(x^*) \leq f(x)$ for all feasible x with $x \in S$. Using Weierstrass' theorem, we select an optimal solution x^k. For all k, $F^k(x^k) \leq F^k(x^*)$, or

$$f(x^k) + \frac{k}{2}\sum_{j=1}^{r}\big(g_j^+(x^k)\big)^2 + \frac{1}{2}||x^k - x^*||^2 \leq f(x^*)$$

Since $f(x^k)$ is bounded over $X \cap S$, $g_j^+(x^k) \to 0$, and every limit point $\overline{x}$ of $\{x^k\}$ is feasible. Also, $f(x^k) + (1/2)||x^k - x^*||^2 \leq f(x^*)$ for all k, so

$$f(\overline{x}) + \frac{1}{2}||\overline{x} - x^*||^2 \leq f(x^*)$$

• Since $\overline{x} \in S$ and $\overline{x}$ is feasible, we have $f(x^*) \leq f(\overline{x})$, so $\overline{x} = x^*$. Thus $x^k \to x^*$, and x^k is an interior point of the closed sphere S for all large k.

PROOF (CONTINUED)

- For k large, we have the necessary condition $-\nabla F^k(x^k) \in T_X(x^k)^*$, which is written as

$$-\left(\nabla f(x^k) + \sum_{j=1}^{r} \zeta_j^k \nabla g_j(x^k) + (x^k - x^*)\right) \in T_X(x^k)^*,$$

where $\zeta_j^k = k g_j^+(x^k)$. Denote

$$\delta^k = \sqrt{1 + \sum_{j=1}^{r} (\zeta_j^k)^2}, \quad \mu_0^k = \frac{1}{\delta^k}, \quad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j > 0$$

Dividing with δ^k,

$$-\left(\mu_0^k \nabla f(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k) + \frac{1}{\delta^k}(x^k - x^*)\right) \in T_X(x^k)^*$$

Since by construction $(\mu_0^k)^2 + \sum_{j=1}^{r} (\mu_j^k)^2 = 1$, the sequence $\{\mu_0^k, \mu_1^k, \ldots, \mu_r^k\}$ is bounded and must contain a subsequence that converges to some limit $\{\mu_0^*, \mu_1^*, \ldots, \mu_r^*\}$. This limit has the required properties ...

CONSTRAINT QUALIFICATIONS

Suppose there do NOT exist $\mu_1, \ldots, \mu_r$, satisfying:

(i) $-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \in N_X(x^*)$.

(ii) $\mu_1, \ldots, \mu_r \geq 0$ and not all 0.

• Then we must have $\mu_0^* > 0$ in FJ, and can take $\mu_0^* = 1$. So there exist $\mu_1^*, \ldots, \mu_r^*$, satisfying all the Lagrange multiplier conditions except that:

$$-\left(\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*)$$

rather than $-(\cdot) \in T_X(x^*)^*$ (such multipliers are called R-*multipliers*).

• If X is regular at x^*, R-multipliers are Lagrange multipliers.

• **LICQ (Lin. Independence Constr. Qual.):** There exists a unique Lagrange multiplier vector if $X = \Re^n$ and x^* is a *regular point*, i.e.,

$$\{\nabla g_j(x^*) \mid j \text{ with } g_j(x^*) = 0\}$$

are linearly independent.

PSEUDONORMALITY

A feasible vector x^* is $pseudonormal$ if there are NO scalars $\mu_1, \ldots, \mu_r$, and a sequence $\{x^k\} \subset X$ such that:

(i) $-\left(\sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.

(ii) $\mu_j \geq 0$, for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.

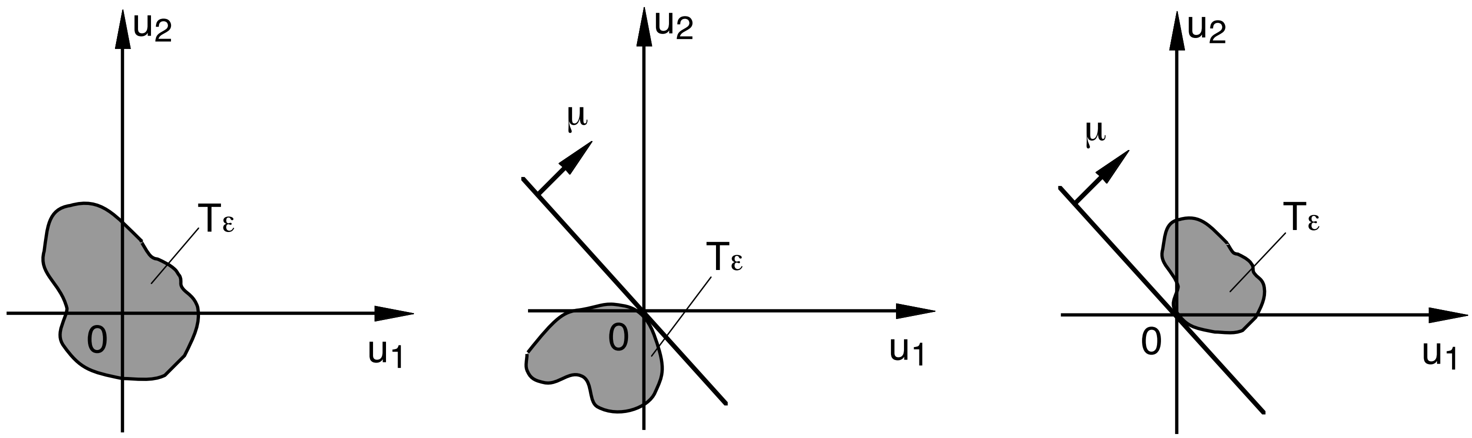
(iii) $\{x^k\}$ converges to x^* and

$$\sum_{j=1}^{r} \mu_j g_j(x^k) > 0, \qquad \forall\ k$$

- From Enhanced FJ conditions:
  - If x^* is pseudonormal there exists an R-multiplier vector.
  - If in addition X is regular at x^*, there exists a Lagrange multiplier vector.

GEOM. INTERPRETATION OF PSEUDONORMALITY I

- Assume that $X = \Re^n$



- Consider, for a small positive scalar ϵ, the set

$$T_\epsilon = \big\{g(x) \mid \|x - x^*\| < \epsilon\big\}$$

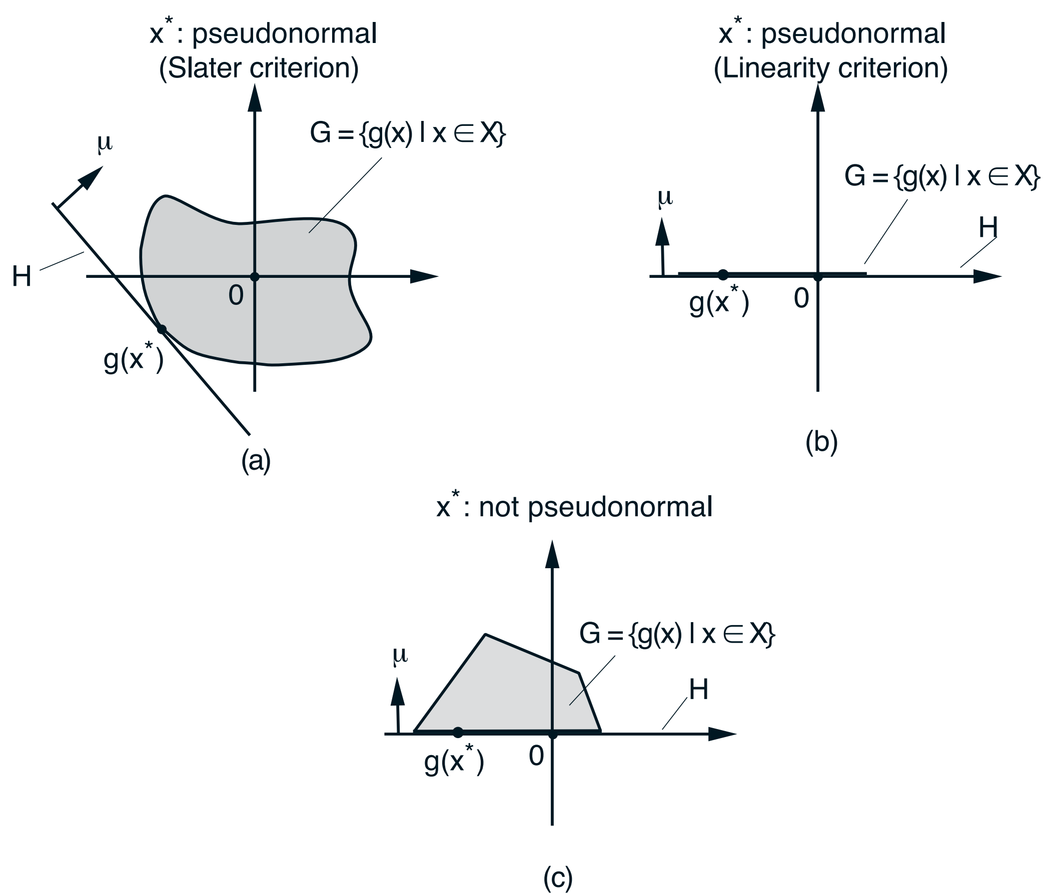
- x^* is pseudonormal if and only if either
  - (1) the gradients $\nabla g_j(x^*)$, $j = 1, \ldots, r$, are linearly independent, or
  - (2) for every $\mu \geq 0$ with $\mu \neq 0$ and such that $\sum_{j=1}^r \mu_j \nabla g_j(x^*) = 0$, there is a small enough ϵ, such that the set T_ϵ does not cross into the positive open halfspace of the hyperplane through 0 whose normal is μ. This is true if the g_j are concave [then $\mu' g(x)$ is maximized at x^* so $\mu' g(x) \leq 0$ for all $x \in \Re^n$].

GEOM. INTERPRETATION OF PSEUDONORMALITY II

- Assume that X and the g_j are convex, so that

$$-\left(\sum_{j=1}^{r}\mu_j\nabla g_j(x^*)\right)\in N_X(x^*)$$

if and only if $x^*\in\arg\min_{x\in X}\sum_{j=1}^{r}\mu_j g_j(x)$. Pseudonormality holds if and only if for every hyperplane with normal $\mu\geq 0$ that passes through the origin and supports the set $G=\{g(x)\mid x\in X\}$, contains G in its negative halfspace.



SOME MAJOR CONSTRAINT QUALIFICATIONS

CQ1: $X = \Re^n$, and the functions g_j are concave.

CQ2: There exists a $y \in N_X(x^*)^*$ such that

$$\nabla g_j(x^*)'y < 0, \quad \forall\, j \in A(x^*)$$

- Special case of CQ2: The Slater condition (X is convex, g_j are convex, and there exists $\overline{x} \in X$ s.t. $g_j(\overline{x}) < 0$ for all j).
- CQ2 is known as the (generalized) Mangasarian-Fromowitz CQ. The version with equality constraints:

  (a) There does not exist a nonzero vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ such that

  $$\sum_{i=1}^{m} \lambda_i \nabla h_i(x^*) \in N_X(x^*)$$

  (b) There exists a $y \in N_X(x^*)^*$ such that

  $$\nabla h_i(x^*)'y = 0,\ \forall\, i,\ \ \nabla g_j(x^*)'y < 0,\ \forall\, j \in A(x^*)$$

CONSTRAINT QUALIFICATION THEOREM

- If CQ1 or CQ2 holds, then x^* is pseudonormal.

Proof: Assume that there are scalars μ_j, $j = 1, \ldots, r$, satisfying conditions (i)-(iii) of the definition of pseudonormality. Then assume that each of the constraint qualifications is in turn also satisfied, and in each case arrive at a contradiction.

Case of CQ1: By the concavity of g_j, the condition $\sum_{j=1}^r \mu_j \nabla g_j(x^*) = 0$, implies that x^* maximizes $\mu' g(x)$ over $x \in \Re^n$, so

$$\mu' g(x) \leq \mu' g(x^*) = 0, \qquad \forall\, x \in \Re^n$$

This contradicts condition (iii) [arbitrarily close to x^*, there is an x satisfying $\sum_{j=1}^r \mu_j g_j(x) > 0$].

Case of CQ2: We must have $\mu_j > 0$ for at least one j, and since $\mu_j \geq 0$ for all j with $\mu_j = 0$ for $j \notin A(x^*)$, we obtain

$$\sum_{j=1}^r \mu_j \nabla g_j(x^*)' y < 0,$$

for the vector y of $N_X(x^*)^*$ that appears in CQ2.

PROOF (CONTINUED)

Thus,

$$-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \notin \big(N_X(x^*)^*\big)^*$$

Since $N_X(x^*) \subset \big(N_X(x^*)^*\big)^*$,

$$-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \notin N_X(x^*)$$

a contradiction of conditions (i) and (ii). **Q.E.D.**

- If $X = \Re^n$, CQ2 is equivalent to the cone $\{y \mid \nabla g_j(x^*)'y \leq 0,\ j \in A(x^*)\}$ having nonempty interior, which (by Gordan's theorem) is equivalent to conditions (i) and (ii) of pseudonormality.

- Note that CQ2 can also be shown to be equivalent to conditions (i) and (ii) of pseudonormality, even when $X \neq \Re^n$, as long as X is regular at x^*. These conditions can in turn can be shown in turn to be equivalent to nonemptiness and compactness of the set of Lagrange multipliers (which is always closed and convex as the intersection of a collection of halfspaces).

LECTURE 17

LECTURE OUTLINE

- Sensitivity Issues
- Exact penalty functions
- Extended representations

Review of Lagrange Multipliers

- Problem: $\min f(x)$ subject to $x \in X$, and $g_j(x) \leq 0$, $j = 1, \ldots, r$.
- Key issue is the existence of Lagrange multipliers for a given local min x^*.
- Existence is guaranteed if X is regular at x^* and we can choose $\mu_0^* = 1$ in the FJ conditions.
- Pseudonormality of x^* guarantees that we can take $\mu_0^* = 1$ in the FJ conditions.
- We derived several constraint qualifications on X and g_j that imply pseudonormality.

PSEUDONORMALITY

A feasible vector x^* is $pseudonormal$ if there are NO scalars $\mu_1, \ldots, \mu_r$, and a sequence $\{x^k\} \subset X$ such that:

(i) $-\left(\sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.

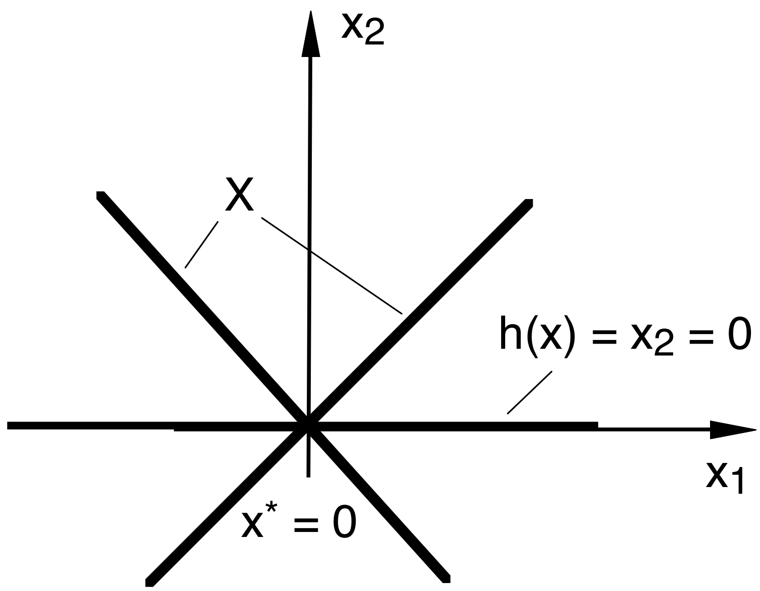
(ii) $\mu_j \geq 0$, for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*) = \{j \mid g_j(x^*) = 0\}$.

(iii) $\{x^k\}$ converges to x^* and

$$\sum_{j=1}^{r} \mu_j g_j(x^k) > 0, \qquad \forall\, k$$

- From Enhanced FJ conditions:
  - If x^* is pseudonormal, there exists an R-multiplier vector.
  - If in addition X is regular at x^*, there exists a Lagrange multiplier vector.

EXAMPLE WHERE X IS NOT REGULAR



- We have

$$T_X(x^*) = X, \quad T_X(x^*)^* = \{0\}, \quad N_X(x^*) = X$$

Let $h(x) = x_2 = 0$ be a single equality constraint. The only feasible point $x^* = (0,0)$ is pseudonormal (satisfies CQ2).

- There exists no Lagrange multiplier for some choices of f.

- For each f, there exists an R-multiplier, i.e., a λ^* such that $-(\nabla f(x^*) + \lambda^* \nabla h(x^*)) \in N_X(x^*)$ ... BUT for f such that there is no L-multiplier, the Lagrangian has negative slope along a tangent direction of X at x^*.

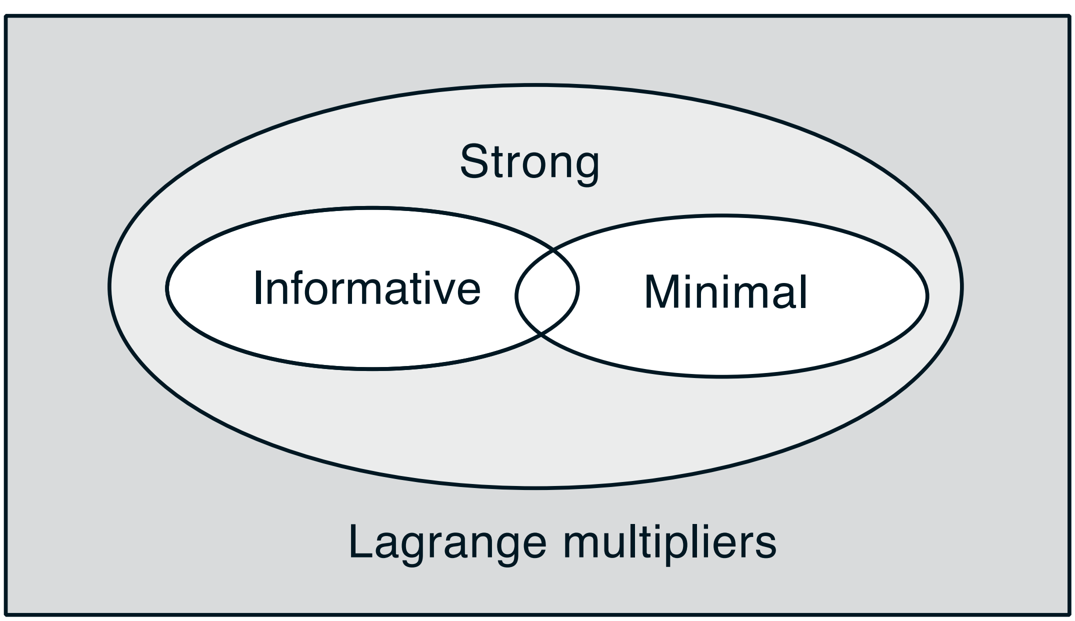
TYPES OF LAGRANGE MULTIPLIERS

• **Informative:** Those that satisfy condition (iii) of the FJ Theorem

• **Strong:** Those that are informative if the constraints with $\mu_j^* = 0$ are neglected

• **Minimal:** Those that have a minimum number of positive components

• **Proposition:** Assume that $T_X(x^*)$ is convex. Then the inclusion properties illustrated in the following figure hold. Furthermore, if there exists at least one Lagrange multiplier, there exists one that is informative (the multiplier of min norm is informative - among possibly others).



SENSITIVITY

• Informative multipliers provide a certain amount of sensitivity.

• They indicate the constraints that need to be violated [those with $\mu_j^* > 0$ and $g_j(x^k) > 0$] in order to be able to reduce the cost from the optimal value [$f(x^k) < f(x^*)$].

• The L-multiplier μ^* of minimum norm is informative, but it is also special; it provides quantitative sensitivity information.

• More precisely, let $d^* \in T_X(x^*)$ be the direction of maximum cost improvement for a given value of norm of constraint violation (up to 1st order; see the text for precise definition). Then for $\{x^k\} \subset X$ converging to x^* along d^*,we have

$$f(x^k) = f(x^*) - \sum_{j=1}^{r} \mu_j^* g_j(x^k) + o(\|x^k - x^*\|)$$

• In the case where there is a unique L-multiplier and $X = \Re^n$, this reduces to the classical interpretation of L-multiplier.

EXACT PENALTY FUNCTIONS

- Exact penalty function

$$F_c(x) = f(x) + c\left(\sum_{i=1}^{m} |h_i(x)| + \sum_{j=1}^{r} g_j^+(x)\right),$$

where c is a positive scalar, and

$$g_j^+(x) = \max\{0, g_j(x)\}$$

- We say that the constraint set C *admits an exact penalty* at a feasible point x^* if for every smooth f for which x^* is a strict local minimum of f over C, there is a $c > 0$ such that x^* is also a local minimum of F_c over X.

- The strictness condition in the definition is essential.

Main Result: If $x^* \in C$ is pseudonormal, the constraint set admits an exact penalty at x^*.

PROOF NEEDS AN INTERMEDIATE RESULT

- First use the (generalized) Mangasarian-Fromovitz CQ to obtain a necessary condition for a local minimum of the exact penalty function.

Proposition: Let x^* be a local minimum of $F_c = f + c\sum_{j=1}^r g_j^+$ over X. Then there exist $\mu_1^*, \ldots, \mu_r^*$ such that

$$-\left(\nabla f(x^*) + c\sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*),$$

$$\mu_j^* = 1 \quad \text{if } g_j(x^*) > 0, \qquad \mu_j^* = 0 \quad \text{if } g_j(x^*) < 0,$$

$$\mu_j^* \in [0,1] \quad \text{if } g_j(x^*) = 0$$

Proof: Convert minimization of $F_c(x)$ over X to minimizing $f(x) + c\sum_{j=1}^r v_j$ subject to

$$x \in X, \qquad g_j(x) \le v_j, \qquad 0 \le v_j, \quad j = 1, \ldots, r$$

PROOF THAT PN IMPLIES EXACT PENALTY

- Assume PN holds and that there exists a smooth f such that x^* is a strict local minimum of f over C, while x^* is not a local minimum over $x \in X$ of $F_k = f + k \sum_{j=1}^{r} g_j^+$ for all $k = 1, 2, \ldots$

- Let x^k minimize F_k over all $x \in X$ satisfying $\|x - x^*\| \leq \epsilon$ (where ϵ is s.t. $f(x^*) < f(x)$ for all $x \in X$ with $x \neq 0$ and $\|x - x^*\| < \epsilon$). Then $x^k \neq x^*$, x^k is infeasible, and

$$F_k(x^k) = f(x^k) + k \sum_{j=1}^{r} g_j^+(x^k) \leq f(x^*)$$

so $g_j^+(x^k) \to 0$ and limit points of x^k are feasible.

- Can assume $x^k \to x^*$, so $\|x^k - x^*\| < \epsilon$ for large k, and we have the necessary conditions

$$-\left(\frac{1}{k}\nabla f(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in N_X(x^k)$$

where $\mu_j^k = 1$ if $g_j(x^k) > 0$, $\mu_j^k = 0$ if $g_j(x^k) < 0$, and $\mu_j^k \in [0, 1]$ if $g_j(x^k) = 0$.

PROOF CONTINUED

- We can find a subsequence $\{\mu^k\}_{k\in\mathcal{K}}$ such that for some j we have $\mu_j^k = 1$ and $g_j(x^k) > 0$ for all $k \in \mathcal{K}$. Let μ be a limit point of this subsequence. Then $\mu \neq 0$, $\mu \geq 0$, and

$$-\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) \in N_X(x^*)$$

[using the closure of the mapping $x \mapsto N_X(x)$].

- Finally, for all $k \in \mathcal{K}$, we have $\mu_j^k g_j(x^k) \geq 0$ for all j, so that, for all $k \in \mathcal{K}$, $\mu_j g_j(x^k) \geq 0$ for all j. Since by construction of the subsequence $\{\mu^k\}_{k\in\mathcal{K}}$, we have for some j and all $k \in \mathcal{K}$, $\mu_j^k = 1$ and $g_j(x^k) > 0$, it follows that for all $k \in \mathcal{K}$,

$$\sum_{j=1}^{r} \mu_j g_j(x^k) > 0$$

This contradicts the pseudonormality of x^*. **Q.E.D.**

EXTENDED REPRESENTATION

- X can often be described as

$$X = \left\{x \mid g_j(x) \leq 0,\, j = r+1, \ldots, \overline{r}\right\}$$

- Then C can alternatively be described without an abstract set constraint,

$$C = \{x \mid g_j(x) \leq 0, \;\; j = 1, \ldots, \overline{r}\}$$

We call this the *extended representation* of C.

Proposition:

(a) If the constraint set admits Lagrange multipliers in the extended representation, it admits Lagrange multipliers in the original representation.

(b) If the constraint set admits an exact penalty in the extended representation, it admits an exact penalty in the original representation.

PROOF OF (A)

- By conditions for case $X = \Re^n$ there exist $\mu_1^*, \ldots, \mu_{\overline{r}}^*$ satisfying

$$\nabla f(x^*) + \sum_{j=1}^{\overline{r}} \mu_j^* \nabla g_j(x^*) = 0,$$

$$\mu_j^* \geq 0, \ \forall\, j = 0, 1, \ldots, \overline{r}, \quad \mu_j^* = 0, \ \forall\, j \notin \overline{A}(x^*),$$

where

$$\overline{A}(x^*) = \{j \mid g_j(x^*) = 0, \ j = 1, \ldots, \overline{r}\}$$

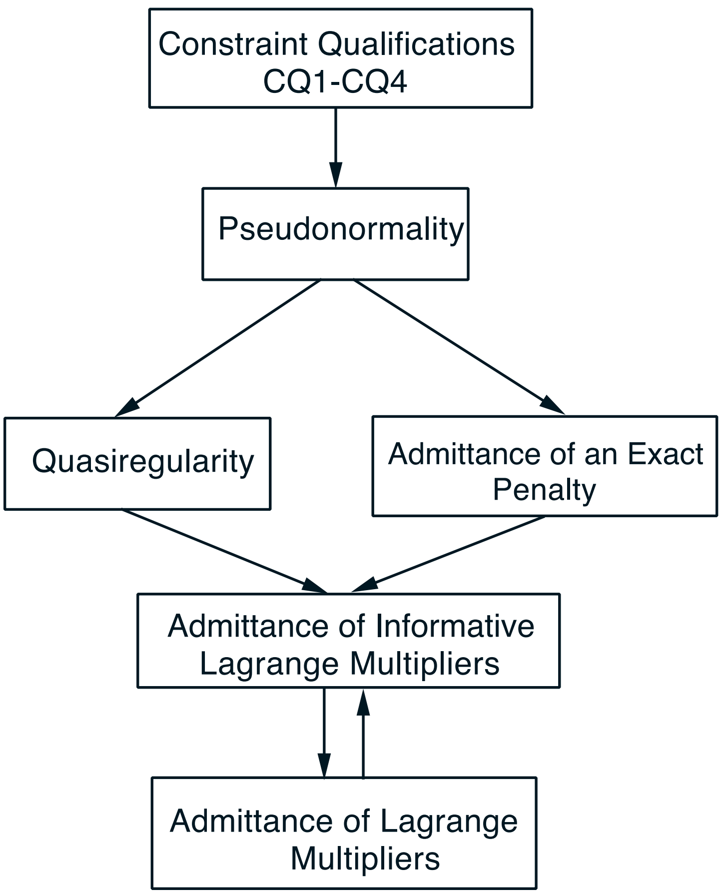
For $y \in T_X(x^*)$, we have $\nabla g_j(x^*)'y \leq 0$ for all $j = r+1, \ldots, \overline{r}$ with $j \in \overline{A}(x^*)$. Hence

$$\left(\nabla f(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\, y \in T_X(x^*),$$

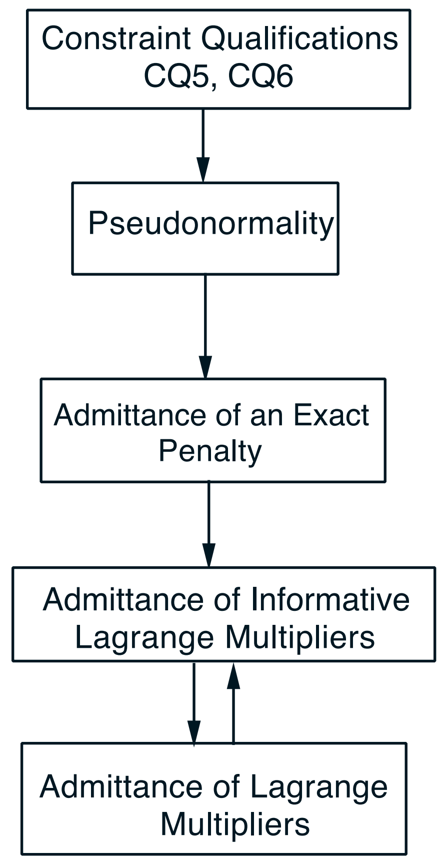
and the μ_j^*, $j = 1, \ldots, r$, are Lagrange multipliers for the original representation.

THE BIG PICTURE

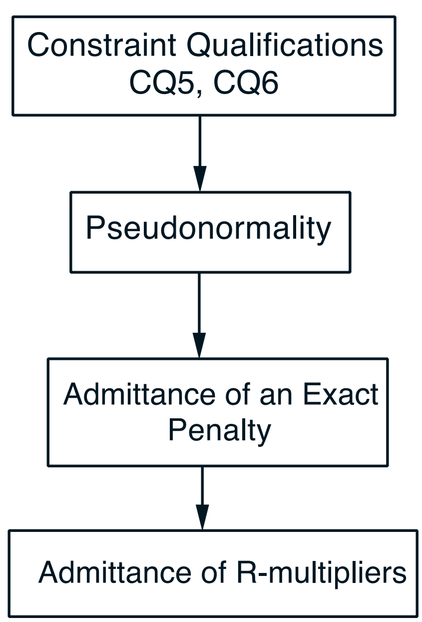
$X = R^n$



$X \neq R^n$ and Regular



$X \neq R^n$



LECTURE 18

LECTURE OUTLINE

- Convexity, geometric multipliers, and duality
- Relation of geometric and Lagrange multipliers
- The dual function and the dual problem
- Weak and strong duality
- Duality and geometric multipliers

GEOMETRICAL FRAMEWORK FOR MULTIPLIERS

- We start an alternative geometric approach to Lagrange multipliers and duality for the problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in X,\ g_1(x) \le 0, \ldots, g_r(x) \le 0 \end{aligned}$$

- We assume nothing on X, f, and g_j, except that

$$-\infty < f^* = \inf_{\substack{x \in X \\ g_j(x) \le 0,\, j=1,\ldots,r}} f(x) < \infty$$

- A vector $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ is said to be a *geometric multiplier* if $\mu^* \ge 0$ and

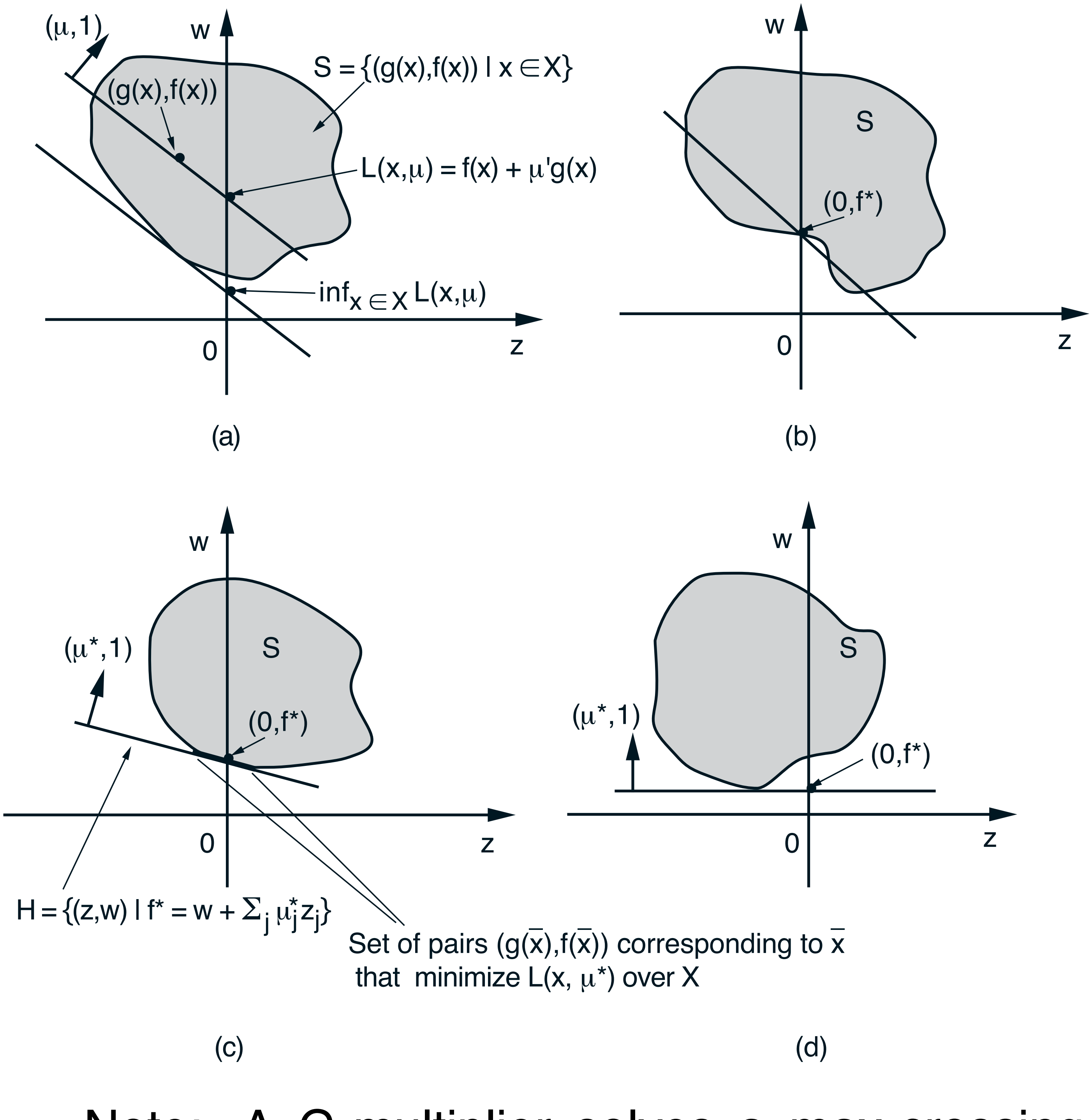
$$f^* = \inf_{x \in X} L(x, \mu^*),$$

where

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

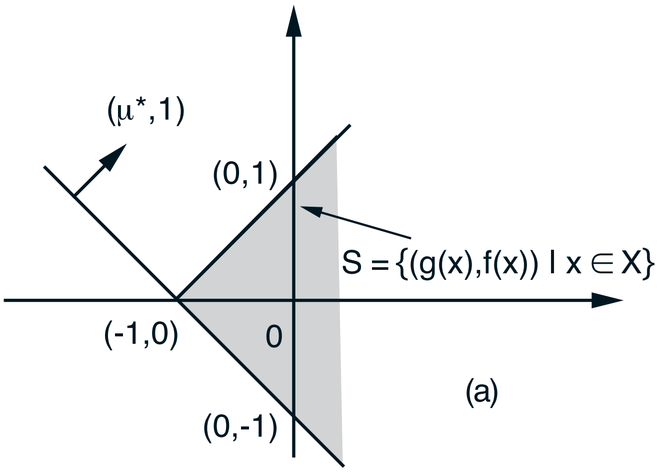
- Note that *a G-multiplier is associated with the problem and not with a specific local minimum.*

VISUALIZATION

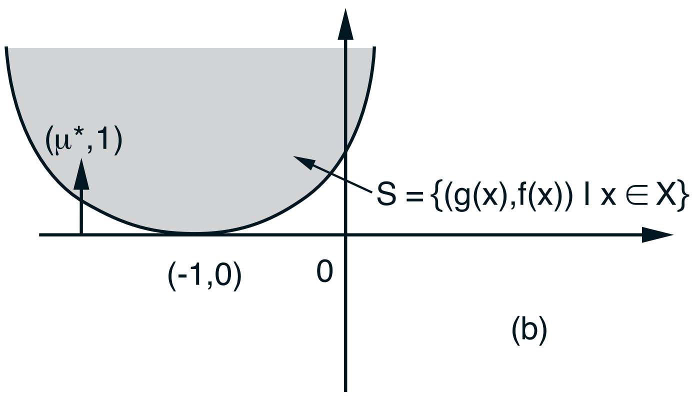


- Note: A G-multiplier solves a max-crossing problem whose min common problem has optimal value f^*.

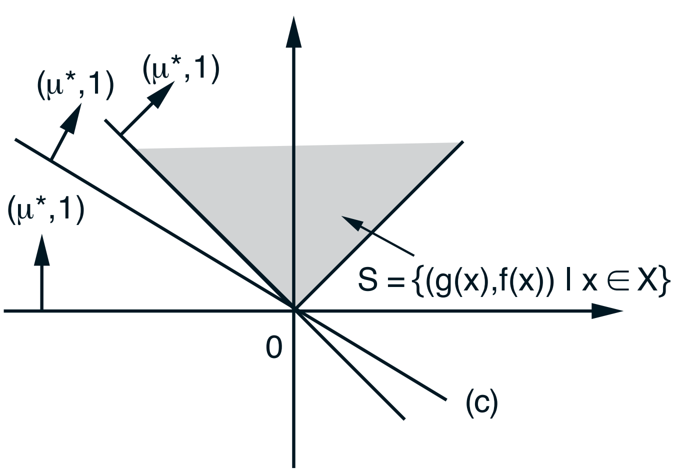
EXAMPLES: A G-MULTIPLIER EXISTS



$$\min\ f(x) = x_1 - x_2$$
$$\text{s.t. } g(x) = x_1 + x_2 - 1 \le 0$$
$$x \in X = \{(x_1,x_2) \mid x_1 \ge 0,\ x_2 \ge 0\}$$

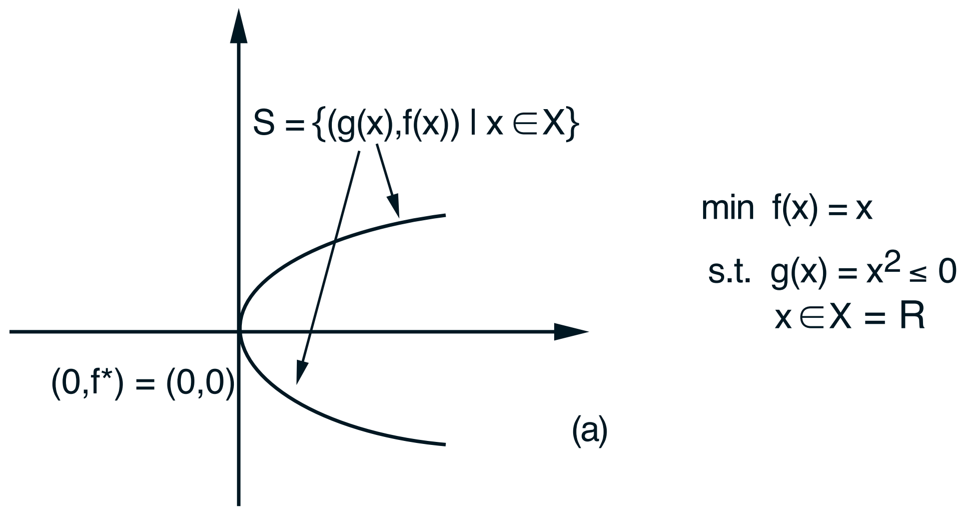


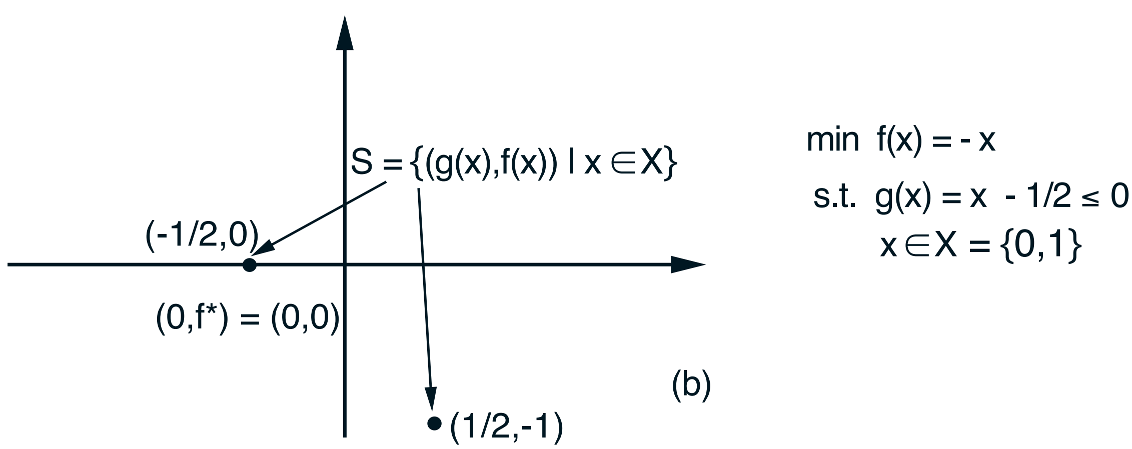
$$\min\ f(x) = (1/2)\,(x_1^2 + x_2^2)$$
$$\text{s.t. } g(x) = x_1 - 1 \le 0$$
$$x \in X = R^2$$



$$\min\ f(x) = |x_1| + x_2$$
$$\text{s.t. } g(x) = x_1 \le 0$$
$$x \in X = \{(x_1,x_2) \mid x_2 \ge 0\}$$

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

$$x^* = \arg\min_{x \in X} L(x, \mu^*), \quad \mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r$$

RELATION BETWEEN G- AND L- MULTIPLIERS

- Assume the problem is convex (X closed and convex, and f and g_j are convex and differentiable over $\Re^n$), and x^* is a global minimum. Then the set of L-multipliers concides with the set of G-multipliers.

- For convex problems, the set of G-multipliers does not depend on the optimal solution x^* (it is the same for all x^*, and may be nonempty even if the problem has no optimal solution x^*).

- In general (for nonconvex problems):
  - Set of G-multipliers may be empty even if the set of L-multipliers is nonempty. [Example problem: $\min_{x=0}(-x^2)$]
  - "Typically" there is no G-multiplier if the set

    $$\big\{(u, w) \mid \text{for some } x \in X,\ g(x) \le u,\ f(x) \le w\big\}$$

    is nonconvex, which often happens if the problem is nonconvex.
  - The G-multiplier idea underlies duality even if the problem is nonconvex.

THE DUAL FUNCTION AND THE DUAL PROBLEM

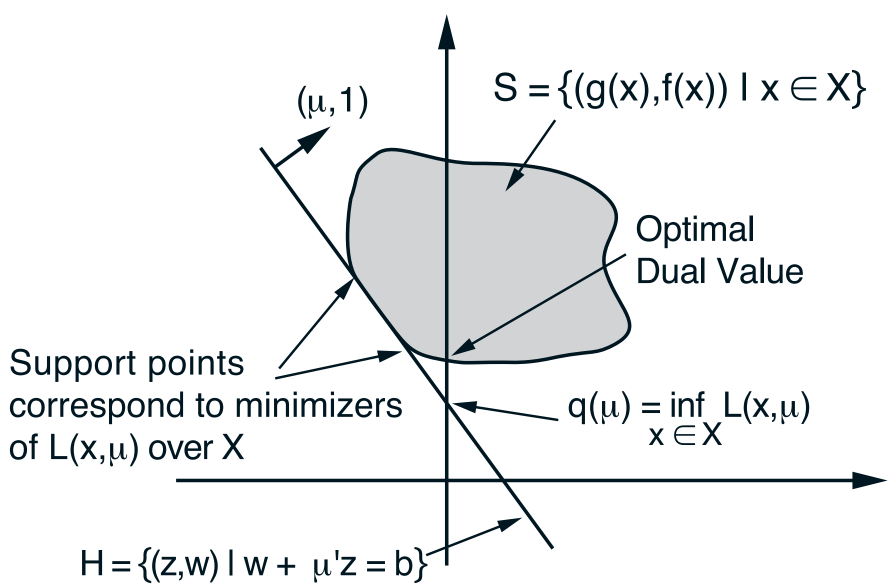
- The *dual problem* is

$$\begin{aligned} &\text{maximize} \ \ q(\mu) \\ &\text{subject to} \ \ \mu \geq 0, \end{aligned}$$

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.



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

$$q^* = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in X,\, g(x) \leq 0} f(x) = f^*$$

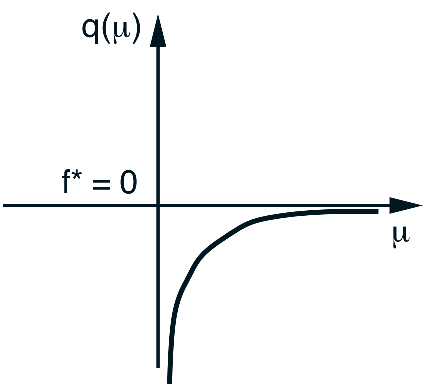
DUAL OPTIMAL SOLUTIONS AND G-MULTIPLIERS

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):



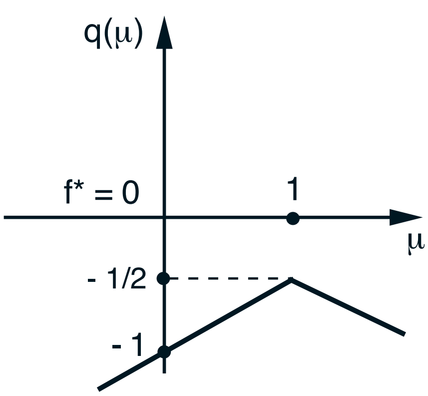
$\min \ f(x) = x$

s.t. $g(x) = x^2 \leq 0$

$x \in X = R$

$$q(\mu) = \min_{x \in R} \{x + \mu x^2\} = \begin{cases} -1/(4\mu) & \text{if } \mu > 0 \\ -\infty & \text{if } \mu \leq 0 \end{cases}$$

(a)



$\min \ f(x) = -x$

s.t. $g(x) = x - 1/2 \leq 0$

$x \in X = \{0,1\}$

$$q(\mu) = \min_{x \in \{0,1\}} \{-x + \mu(x - 1/2)\} = \min\{-\mu/2, \mu/2 - 1\}$$

(b)

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

$$L(x^*, \mu) \leq L(x^*, \mu^*) \leq L(x, \mu^*), \ \ \forall\, x \in X, \ \mu \geq 0$$

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 $x \notin X$.

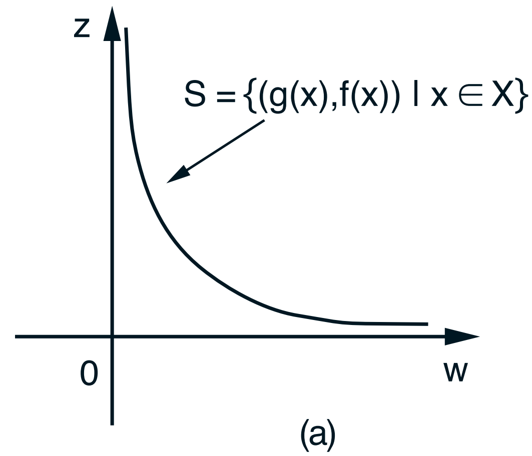
- f is convex over X (its Hessian is positive definite in the interior of X), and $f^* = 1$.

- 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

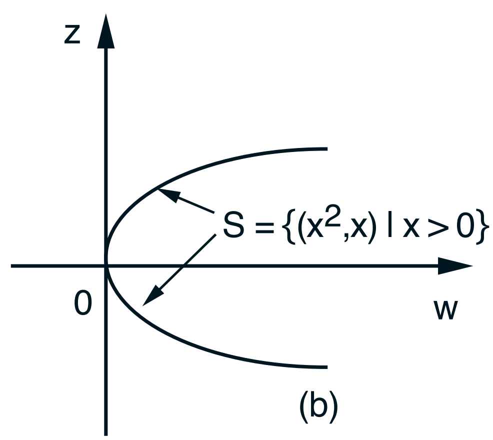


$\min\ f(x) = 1/x$

s.t. $g(x) = x \le 0$

$x \in X = \{x \mid x > 0\}$

$f^* = \infty, \quad q^* = \infty$

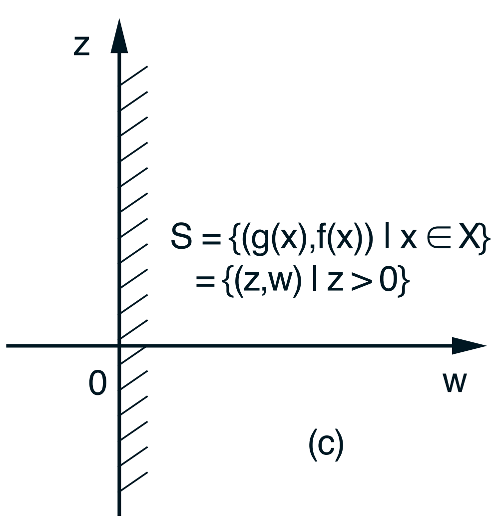


$\min\ f(x) = x$

s.t. $g(x) = x^2 \le 0$

$x \in X = \{x \mid x > 0\}$

$f^* = \infty, \quad q^* = 0$



$\min\ f(x) = x_1 + x_2$

s.t. $g(x) = x_1 \le 0$

$x \in X = \{(x_1, x_2) \mid x_1 > 0\}$

$f^* = \infty, \quad q^* = -\infty$

LECTURE 19

LECTURE OUTLINE

- Linear and quadratic programming duality
- Conditions for existence of geometric multipliers
- Conditions for strong duality

- Primal problem: Minimize $f(x)$ subject to $x \in X$, and $g_1(x) \le 0, \ldots, g_r(x) \le 0$ (assuming $-\infty < f^* < \infty$). It is equivalent to $\inf_{x \in X} \sup_{\mu \ge 0} L(x, \mu)$.
- Dual problem: Maximize $q(\mu)$ subject to $\mu \ge 0$, where $q(\mu) = \inf_{x \in X} L(x, \mu)$. It is equivalent to $\sup_{\mu \ge 0} \inf_{x \in X} L(x, \mu)$.
- 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 geometric multiplier?

LINEAR AND QUADRATIC PROGRAMMING DUALITY

- Consider a LP or positive semidefinite QP under the assumption

$$-\infty < f^* < \infty$$

- We know from Chapter 2 that

$$-\infty < f^* < \infty \quad \Rightarrow \quad \text{there is an optimal solution } x^*$$

- Since the constraints are linear, there exist L-multipliers corresponding to x^*, so we can use Lagrange multiplier theory.
- Since the problem is convex, the L-multipliers coincide with the G-multipliers.
- Hence there exists a G-multiplier, $f^* = q^*$ and the optimal solutions of the dual problem coincide with the Lagrange multipliers.

THE DUAL OF A LINEAR PROGRAM

- Consider the linear program

minimize $c'x$

subject to $e_i'x = d_i, \quad i = 1, \ldots, m, \qquad x \geq 0$

- 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

$$\text{maximize} \quad \sum_{i=1}^{m} \lambda_i d_i$$

$$\text{subject to} \quad \sum_{i=1}^{m} \lambda_i e_{ij} \leq c_j, \qquad j = 1, \ldots, n.$$

THE DUAL OF A QUADRATIC PROGRAM

- Consider the quadratic program

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

where Q is a given $n \times n$ positive definite symmetric matrix, A is a given $r \times n$ matrix, and $b \in \Re^r$ and $c \in \Re^n$ are given vectors.

- Dual function:

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

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

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

- The dual problem, after a sign change, is

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

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

RECALL 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

$$f(x) \geq 0, \qquad \forall\, x \in F = \big\{x \in C \mid g(x) \leq 0\big\},$$

and one of the following two conditions holds:

(1) 0 is in the relative interior of the set $D = \big\{u \mid g(x) \leq u \text{ for some } x \in C\big\}$.

(2) The functions g_j, $j = 1, \ldots, r$, are affine, and F contains a relative interior point of C.

Then, there exist scalars $\mu_j^* \geq 0$, $j = 1, \ldots, r$, s. t.

$$f(x) + \sum_{j=1}^{r} \mu_j^* g_j(x) \geq 0, \qquad \forall\, x \in C$$

APPLICATION TO CONVEX PROGRAMMING

Consider the problem

minimize $f(x)$

subject to $x \in C, \quad g_j(x) \leq 0,\ j = 1, \ldots, r,$

where C, $f : C \mapsto \Re$, and $g_j : C \mapsto \Re$ 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 $\mu_j^* \geq 0$ such that

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

Since $F \subset C$ and $\mu_j^* g_j(x) \leq 0$ for all $x \in F$,

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

Thus equality holds throughout, we have

$$f^* = \inf_{x \in C} \left\{ f(x) + \mu^{*\prime} g(x) \right\},$$

and μ^* is a geometric multiplier.

STRONG DUALITY THEOREM I

Assumption : (Convexity and Linear Constraints) f^* is finite, and the following hold:

(1) $X = P \cap C$, where P is polyhedral and C is convex.

(2) The cost function f is convex over C and the functions g_j are affine.

(3) There exists a feasible solution of the problem that belongs to the relative interior of C.

Proposition : Under the above assumption, there exists at least one geometric multiplier.

Proof: If $P = \Re^n$ the result holds by Farkas. If $P \neq \Re^n$, express P as

$$P = \{x \mid a_j' x - b_j \leq 0,\ j = r+1, \ldots, p\}$$

Apply Farkas to the extended representation, with

$$F = \{x \in C \mid a_j' x - b_j \leq 0,\ j = 1, \ldots, p\}$$

Assert the existence of geometric multipliers in the extended representation, and pass back to the original representation. **Q.E.D.**

STRONG DUALITY THEOREM II

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

$$\sup_{\mu \geq 0} \big\{ L(x, \mu) - \mu' u \big\} = \sup_{\mu \geq 0} \big\{ f(x) + \mu' \big(g(x) - u \big) \big\} = \begin{cases} f(x) & \text{if } g(x) \leq u, \\ \infty & \text{otherwise.} \end{cases}$$

- 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 $p(0) = f^*$].

CONDITIONS FOR NO DUALITY GAP

- Apply the minimax theory specialized to $L(x, \mu)$.
- Assume that $f^* < \infty$, and that 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.

LECTURE 20

LECTURE OUTLINE

- The primal function
- Conditions for strong duality
- Sensitivity
- Fritz John conditions for convex programming

$-----------------------------$

- 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 $\inf_{x \in X} \sup_{\mu \geq 0} L(x, \mu)$.

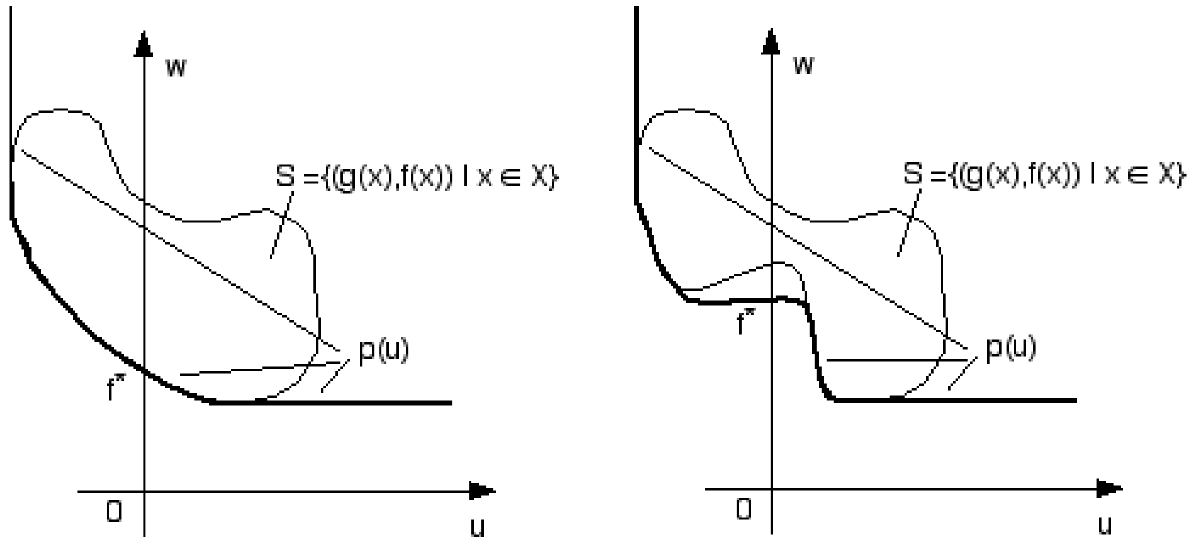
- The primal function is the *perturbed optimal value*

$$p(u) = \inf_{x \in X} \sup_{\mu \geq 0} \big\{ L(x, \mu) - \mu' u \big\} = \inf_{\substack{x \in X \\ g(x) \leq u}} f(x)$$

- Note that $p(u)$ is the result of partial minimization over X of the function $F(x, u)$ given by

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

PRIMAL FUNCTION AND STRONG DUALITY



- Apply min common-max crossing framework with set $M = \text{epi}(p)$, assuming p is convex and $-\infty < p(0) < \infty$.

- 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 the partial minimization $p(u) = \inf_{\substack{x \in X \\ g(x) \leq u}} f(x)$, e.g.:
  - X is convex and compact, f, g_j: convex.
  - Extensions involving the recession cones of X, f, g_j.
  - $X = \Re^n$, f, g_j: convex quadratic.

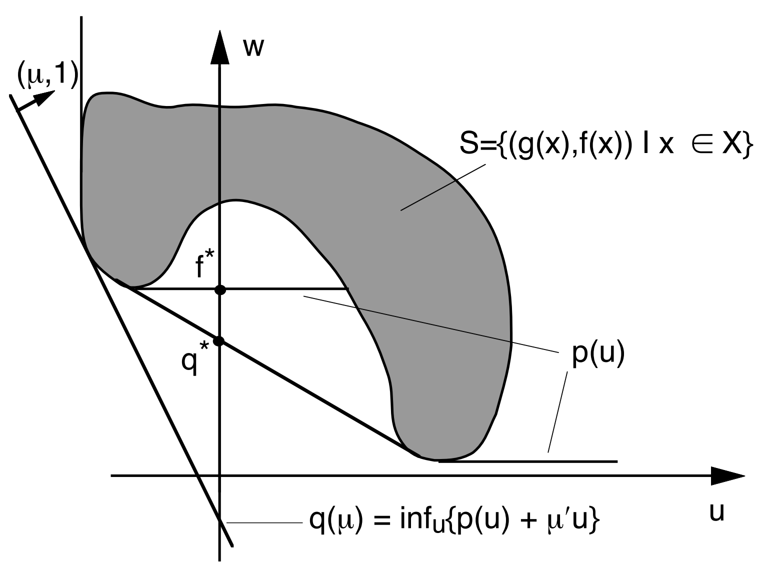
RELATION OF PRIMAL AND DUAL FUNCTIONS

- Consider the dual function q. For every $\mu \geq 0$, we have

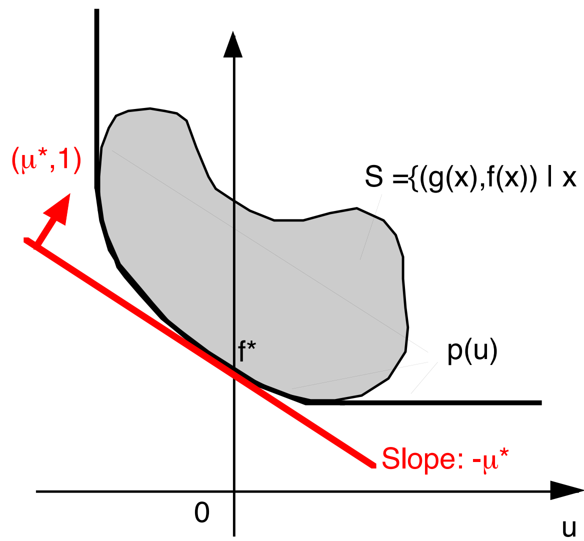
$$
\begin{aligned}
q(\mu) &= \inf_{x \in X} \{f(x) + \mu' g(x)\} \\
&= \inf_{\{(u,x) \mid x \in X,\, g(x) \leq u,\, j=1,\ldots,r\}} \{f(x) + \mu' g(x)\} \\
&= \inf_{\{(u,x) \mid x \in X,\, g(x) \leq u\}} \{f(x) + \mu' u\} \\
&= \inf_{u \in \Re^r} \inf_{x \in X,\, g(x) \leq u} \{f(x) + \mu' u\}.
\end{aligned}
$$

- Thus

$$
q(\mu) = \inf_{u \in \Re^r} \big\{p(u) + \mu' u\big\}, \qquad \forall\ \mu \geq 0
$$



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 $-\partial p(0)$ (negative subdifferential of p at $u = 0$). This follows from the relation

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

  - If the origin lies in the relative interior of the effective domain of p, then there exists a G-multiplier.
  - If the origin lies in the interior of the effective domain of p, the set of G-multipliers is nonempty and compact.

SENSITIVITY ANALYSIS I

- Assume that p is convex and differentiable. Then $-\nabla p(0)$ is the unique G-multiplier μ^*, and we have

$$\mu_j^* = -\frac{\partial p(0)}{\partial u_j}, \qquad \forall\ j$$

- Let μ^* be a G-multiplier, and consider a vector u_j^γ of the form

$$u_j^\gamma = (0, \ldots, 0, \gamma, 0, \ldots, 0)$$

where γ is a scalar in the jth position. Then

$$\lim_{\gamma \uparrow 0} \frac{p(u_j^\gamma) - p(0)}{\gamma} \leq -\mu_j^* \leq \lim_{\gamma \downarrow 0} \frac{p(u_j^\gamma) - p(0)}{\gamma}$$

Thus $-\mu_j^*$ lies between the left and the right slope of p in the direction of the jth axis starting at $u = 0$.

SENSITIVITY ANALYSIS II

- Assume that p is convex and finite in a neighborhood of 0. Then, from the theory of subgradients:
  - $\partial p(0)$ is nonempty and compact.
  - The directional derivative $p'(0; y)$ is a real-valued convex function of y satisfying

$$p'(0; y) = \max_{g \in \partial p(0)} y'g$$

- Consider the direction of steepest descent of p at 0, i.e., the $\overline{y}$ that minimizes $p'(0; y)$ over $\|y\| \leq 1$. Using the Saddle Point Theorem,

$$p'(0; \overline{y}) = \min_{\|y\| \leq 1} p'(0; y) = \min_{\|y\| \leq 1} \max_{g \in \partial p(0)} y'g = \max_{g \in \partial p(0)} \min_{\|y\| \leq 1} y'g$$

- The saddle point is $(g^*, \overline{y})$, where g^* is the subgradient of minimum norm in $\partial p(0)$ and $\overline{y} = -g^*/\|g^*\|$. The min-max value is $-\|g^*\|$.

- **Conclusion:** If μ^* is the G-multiplier of minimum norm and $\mu^* \neq 0$, the direction of steepest descent of p at 0 is $\overline{y} = \mu^*/\|\mu^*\|$, while the rate of steepest descent (per unit norm of constraint violation) is $\|\mu^*\|$.

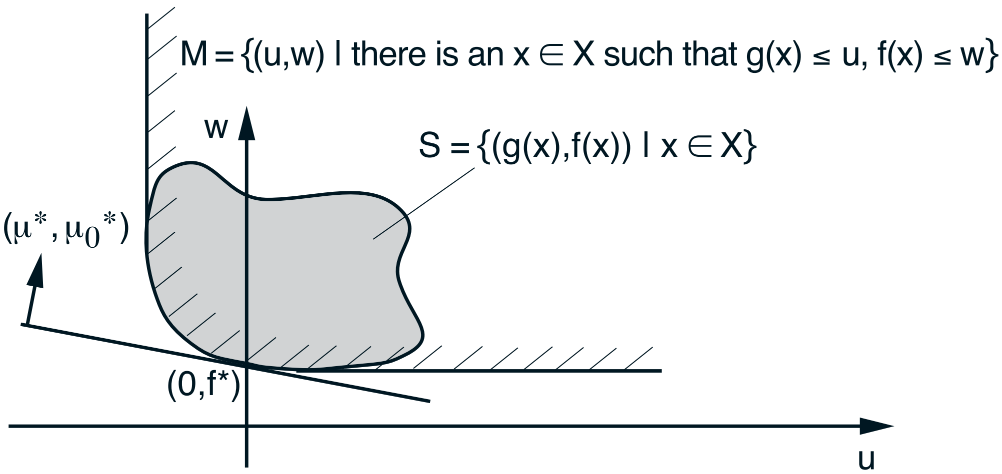
FRITZ JOHN THEORY FOR CONVEX PROBLEMS

- 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) $\mu_0^* f^* = \inf_{x \in X} \big\{ \mu_0^* f(x) + \mu^{*\prime} g(x) \big\}$.

  (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.



- 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 $\mu^{*\prime} g(\overline{x}) < 0$.

FRITZ JOHN 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.

- The text has more material on pseudonormality, informative geometric multipliers, etc.

LECTURE 21

LECTURE OUTLINE

- Fenchel Duality
- Conjugate Convex Functions
- Relation of Primal and Dual Functions
- Fenchel Duality Theorems

FENCHEL DUALITY FRAMEWORK

- Consider the problem

$$\begin{aligned} &\text{minimize} \quad f_1(x) - f_2(x) \\ &\text{subject to} \quad x \in X_1 \cap X_2, \end{aligned}$$

where f_1 and f_2 are real-valued functions on $\Re^n$, and X_1 and X_2 are subsets of $\Re^n$.

- Assume that $f^* < \infty$.
- Convert problem to

$$\begin{aligned} &\text{minimize} \quad f_1(y) - f_2(z) \\ &\text{subject to} \quad z = y, \qquad y \in X_1, \qquad z \in X_2, \end{aligned}$$

and dualize the constraint $z = y$:

$$\begin{aligned} q(\lambda) &= \inf_{y \in X_1,\, z \in X_2} \big\{ f_1(y) - f_2(z) + (z - y)'\lambda \big\} \\ &= \inf_{z \in X_2} \big\{ z'\lambda - f_2(z) \big\} - \sup_{y \in X_1} \big\{ y'\lambda - f_1(y) \big\} \\ &= g_2(\lambda) - g_1(\lambda) \end{aligned}$$

CONJUGATE FUNCTIONS

- The functions $g_1(\lambda)$ and $g_2(\lambda)$ are called the *conjugate convex* and *conjugate concave* functions corresponding to the pairs (f_1, X_1) and (f_2, X_2).
- An equivalent definition of g_1 is

$$g_1(\lambda) = \sup_{x \in \Re^n} \big\{x'\lambda - \tilde{f}_1(x)\big\},$$

where $\tilde{f}_1$ is the extended real-valed function

$$\tilde{f}_1(x) = \begin{cases} f_1(x) & \text{if } x \in X_1, \\ \infty & \text{if } x \notin X_1. \end{cases}$$

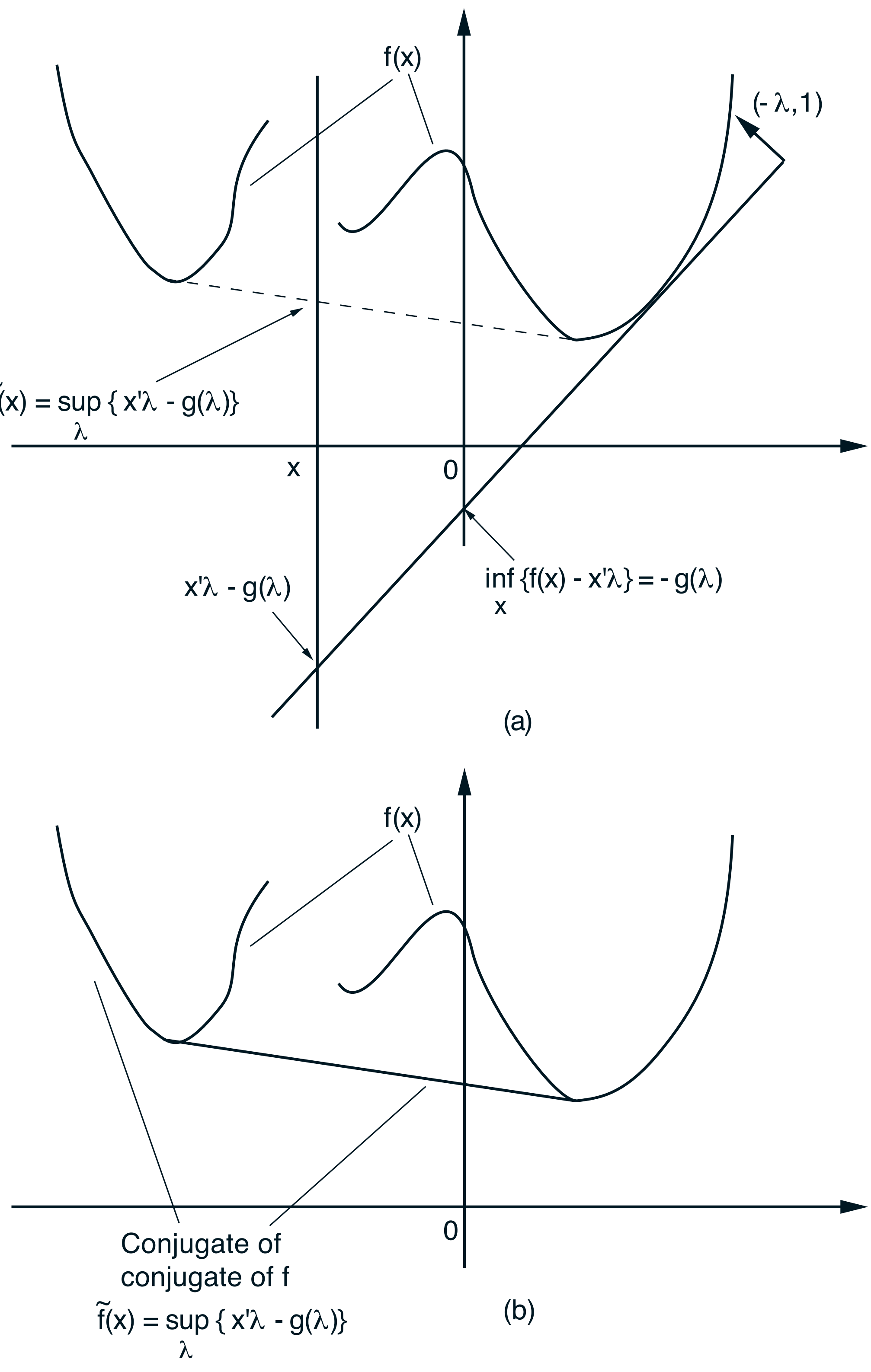
- We are led to consider the conjugate convex function of a general extended real-valued proper function $f : \Re^n \mapsto (-\infty, \infty]$:

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

- Conjugate concave functions are defined through conjugate convex functions after appropriate sign reversals.

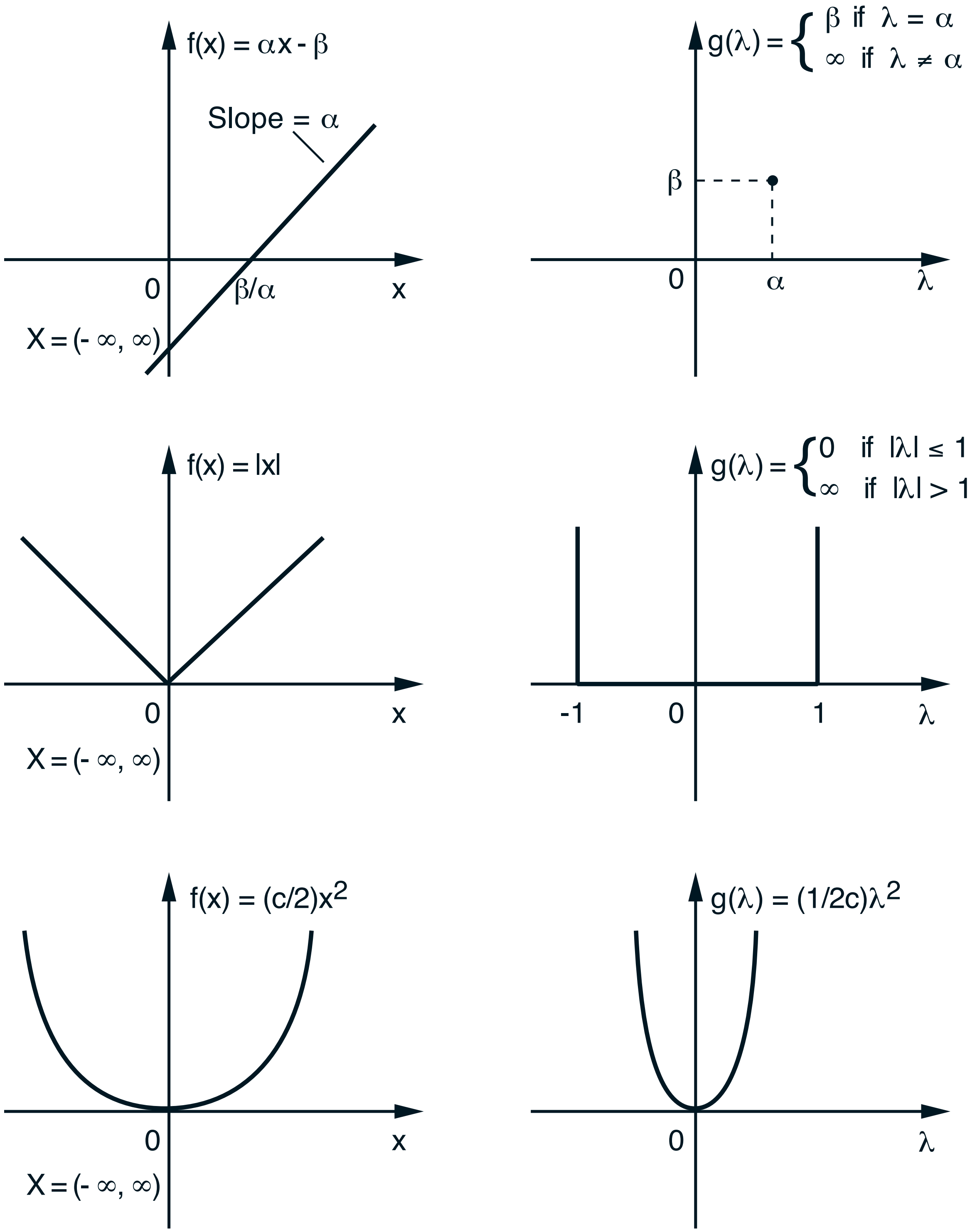
VISUALIZATION

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



EXAMPLES OF CONJUGATE PAIRS

$$g(\lambda) = \sup_{x \in \Re^n} \big\{x'\lambda - f(x)\big\}, \quad \tilde{f}(x) = \sup_{\lambda \in \Re^n} \big\{x'\lambda - g(\lambda)\big\}$$



CONJUGATE OF THE CONJUGATE FUNCTION

- Two cases to consider:
  - f is a closed proper convex function.
  - f is a general extended real-valued proper function.

- We will see that for closed proper convex functions, the conjugacy operation is symmetric, i.e., *the congugate of f is a closed proper convex function, and the conjugate of the conjugate is f.*

- Leads to a symmetric/dual Fenchel duality theorem for the case where the functions involved are closed convex/concave.

- The result can be generalized:
  - The *convex closure of f*, is the function that has as epigraph the closure of the convex hull if epi(f) [also the smallest closed and convex set containing epi(f)].
  - The epigraph of the convex closure of f is the intersection of all closed halfspaces of $\Re^{n+1}$ that contain the epigraph of f.

CONJUGATE FUNCTION THEOREM

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

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

(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, g, 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

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

CONJUGACY OF PRIMAL AND DUAL FUNCTIONS

- Consider the problem

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

- We showed in the previous lecture the following relation between primal and dual functions:

$$q(\mu) = \inf_{u \in \Re^r} \big\{ p(u) + \mu' u \big\}, \qquad \forall\, \mu \geq 0.$$

- Thus, $q(\mu) = -\sup_{u \in \Re^r} \big\{ -\mu' u - p(u) \big\}$ or

$$q(\mu) = -h(-\mu), \qquad \forall\, \mu \geq 0,$$

where h is the conjugate convex function of p:

$$h(\nu) = \sup_{u \in \Re^r} \big\{ \nu' u - p(u) \big\}$$

INDICATOR AND SUPPORT FUNCTIONS

- The *indicator function* of a nonempty set is

$$\delta_X(x) = \begin{cases} 0 & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

- The conjugate of δ_X, given by

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

is called the *support function of X*.

- X has the same support function as $\mathrm{cl}\big(\mathrm{conv}(X)\big)$ (by the Conjugacy Theorem).

- If X is closed and convex, δ_X is closed and convex, and by the Conjugacy Theorem the conjugate of its support function is its indicator function.

- The support function satisfies

$$\sigma_X(\alpha\lambda) = \alpha\sigma_X(\lambda), \qquad \forall\, \alpha > 0,\ \forall\, \lambda \in \Re^n$$

so its epigraph is a cone. Functions with this property are called *positively homogeneous*.

MORE ON SUPPORT FUNCTIONS

• For a cone C, we have

$$\sigma_C(\lambda) = \sup_{x \in C} \lambda' x = \begin{cases} 0 & \text{if } \lambda \in C^*, \\ \infty & \text{otherwise,} \end{cases}$$

i.e., the support function of a cone is the indicator function of its polar.

• The support function of a polyhedral set is a polyhedral function that is pos. homogeneous. The conjugate of a pos. homogeneous polyhedral function is the support function of some polyhedral set.

• A function can be equivalently specified in terms of its epigraph. As a consequence, we will see that the conjugate of a function can be specified in terms of the support function of its epigraph.

• The conjugate of f, can equivalently be written as $g(\lambda) = \sup_{(x,w) \in \text{epi}(f)} \{x'\lambda - w\}$, so

$$g(\lambda) = \sigma_{\text{epi}(f)}(\lambda, -1), \qquad \forall\, \lambda \in \Re^n$$

• From this formula, we also obtain that the conjugate of a polyhedral function is polyhedral.

LECTURE 22

LECTURE OUTLINE

- Fenchel Duality
- Fenchel Duality Theorems
- Cone Programming
- Semidefinite Programming

−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−

- Recall the conjugate convex function of a general extended real-valued proper function $f : \Re^n \mapsto (-\infty, \infty]$:

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

- **Conjugacy Theorem:** If f is closed and convex, then f is equal to the 2nd conjugate (the conjugate of the conjugate).

FENCHEL DUALITY FRAMEWORK

- Consider the problem

$$\begin{array}{ll} \text{minimize} & f_1(x) - f_2(x) \\ \text{subject to} & x \in X_1 \cap X_2, \end{array}$$

where $f_1 : \Re^n \mapsto (-\infty, \infty]$ and $f_2 : \Re^n \mapsto [-\infty, \infty)$.

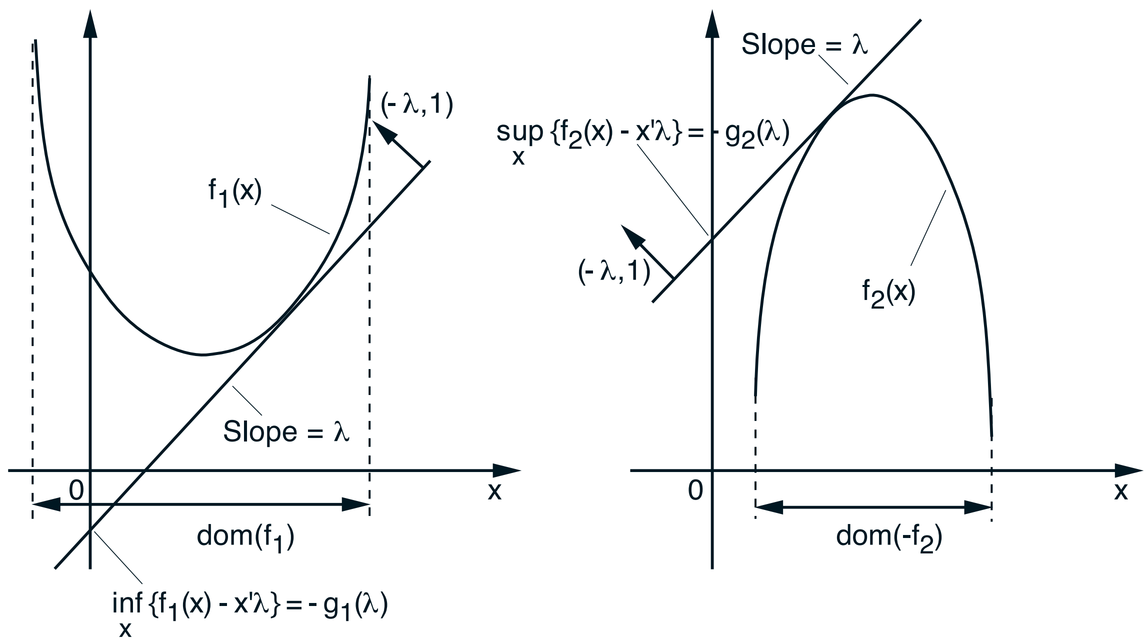
- Assume that $f^* < \infty$.
- Convert problem to

$$\begin{array}{ll} \text{minimize} & f_1(y) - f_2(z) \\ \text{subject to} & z = y, \quad y \in \text{dom}(f_1), \quad z \in \text{dom}(-f_2), \end{array}$$

and dualize the constraint $z = y$:

$$\begin{aligned} q(\lambda) &= \inf_{y \in \Re^n,\, z \in \Re^n} \big\{ f_1(y) - f_2(z) + (z - y)'\lambda \big\} \\ &= \inf_{z \in \Re^n} \big\{ z'\lambda - f_2(z) \big\} - \sup_{y \in \Re^n} \big\{ y'\lambda - f_1(y) \big\} \\ &= g_2(\lambda) - g_1(\lambda) \end{aligned}$$

FENCHEL DUALITY THEOREM



- Assume that f_1 and f_2 are convex and concave, respectively. If either
  - The relative interiors of $\mathrm{dom}(f_1)$ and $\mathrm{dom}(-f_2)$ intersect, or
  - $\mathrm{dom}(f_1)$ and $\mathrm{dom}(-f_2)$ are polyhedral, and f_1 and f_2 can be extended to real-valued convex and concave functions over $\Re^n$.

Then the geometric multipliers existence theorem applies and we have

$$f^* = \max_{\lambda \in \Re^n} \big\{ g_2(\lambda) - g_1(\lambda) \big\},$$

while the maximum above is attained.

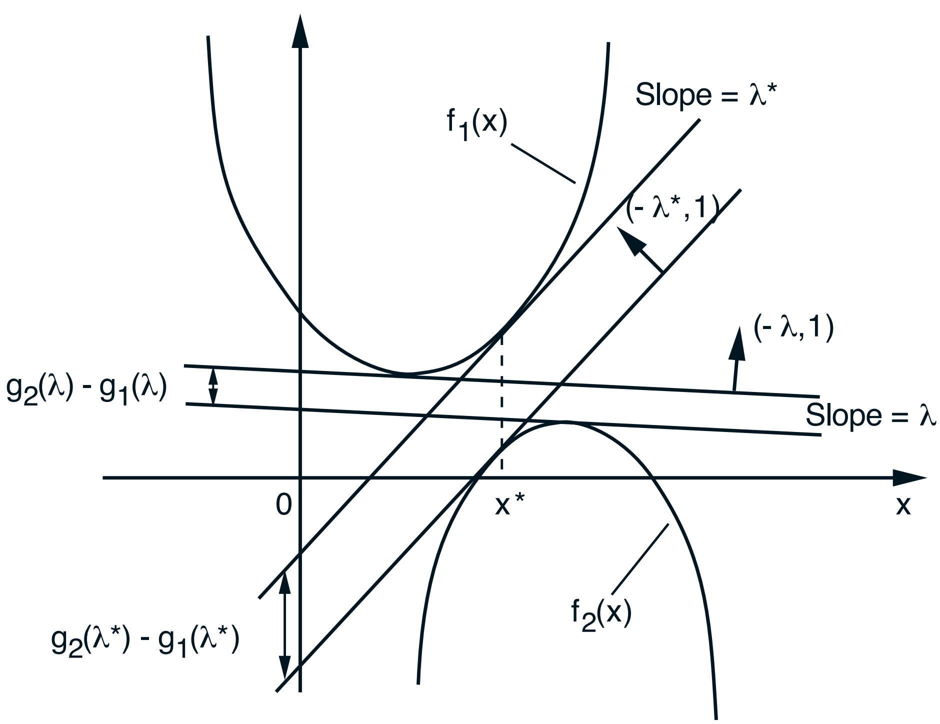
OPTIMALITY CONDITIONS

- There is no duality gap, while (x^*, λ^*) is an optimal primal and dual solution pair, if and only if

$$x^* \in \mathrm{dom}(f_1) \cap \mathrm{dom}(-f_2), \qquad \text{(primal feasibility)},$$

$$\lambda^* \in \mathrm{dom}(g_1) \cap \mathrm{dom}(-g_2), \qquad \text{(dual feasibility)},$$

$$x^* = \arg\max_{x \in \Re^n} \{x'\lambda^* - f_1(x)\}$$
$$= \arg\min_{x \in \Re^n} \{x'\lambda^* - f_2(x)\}, \quad \text{(Lagr. optimality)}.$$



- Note: The Lagrangian optimality condition is equivalent to $\lambda^* \in \partial f_1(x^*) \cap \partial f_1(x^*)$.

DUAL FENCHEL DUALITY THEOREM

- The dual problem

$$\max_{\lambda \in \Re^n} \big\{ g_2(\lambda) - g_1(\lambda) \big\}$$

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 g_1 and g_2.

- **Conclusion:** The primal problem has an optimal solution, there is no duality gap, and we have

$$\min_{x \in \Re^n} \big\{ f_1(x) - f_2(x) \big\} = \sup_{\lambda \in \Re^n} \big\{ g_2(\lambda) - g_1(\lambda) \big\},$$

if either

– The relative interiors of $\mathrm{dom}(g_1)$ and $\mathrm{dom}(-g_2)$ intersect, or

– $\mathrm{dom}(g_1)$ and $\mathrm{dom}(-g_2)$ are polyhedral, and g_1 and g_2 can be extended to real-valued convex and concave functions over $\Re^n$.

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

$$g_1(\lambda) = \sup_{x \in \Re^n} \big\{ \lambda' x - f(x) \big\}, \; g_2(\lambda) = \inf_{x \in C} x' \lambda = \begin{cases} 0 & \text{if } \lambda \in \hat{C}, \\ -\infty & \text{if } \lambda \notin \hat{C}, \end{cases}$$

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

$$\inf_{x \in C} f(x) = \sup_{\lambda \in \hat{C}} -g(\lambda),$$

where $g(\lambda)$ 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) $\mathrm{ri}(\mathrm{dom}(g)) \cap \mathrm{ri}(\hat{C}) \neq \emptyset$.

(b) g can be extended to a real-valued convex function over $\Re^n$, and $\mathrm{dom}(g)$ and $\hat{C}$ are polyhedral.

THE AFFINE COST CASE OF CONIC DUALITY

- Let f be affine, $f(x) = c'x$, with $\mathrm{dom}(f)$ being an affine set, $\mathrm{dom}(f) = b + S$, 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

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

so the dual problem is

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

- The primal and dual have the same form.
- If C is closed, the dual of the dual yields the primal.

SEMIDEFINITE PROGRAMMING: A SPECIAL CASE

• Consider the symmetric $n \times n$ matrices. Inner product $< X, Y >=$ 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.

LECTURE 23

LECTURE OUTLINE

- Overview of Dual Methods
- Nondifferentiable Optimization

****

- Consider the primal problem

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

assuming $-\infty < f^* < \infty$.

- Dual problem: Maximize

$$q(\mu) = \inf_{x \in X} L(x, \mu) = \inf_{x \in X} \{f(x) + \mu' g(x)\}$$

subject to $\mu \geq 0$.

PROS AND CONS FOR SOLVING THE DUAL

- The dual is concave.
- The dual may have smaller dimension and/or simpler constraints.
- If there is no duality gap and the dual is solved exactly for a geometric multiplier μ^*, all optimal primal solutions can be obtained by minimizing the Lagrangian $L(x, \mu^*)$ over $x \in X$.
- Even if there is a duality gap, $q(\mu)$ is a lower bound to the optimal primal value for every $\mu \geq 0$.
- Evaluating $q(\mu)$ requires minimization of $L(x, \mu)$ over $x \in X$.
- The dual function is often nondifferentiable.
- Even if we find an optimal dual solution μ^*, it may be difficult to obtain a primal optimal solution.

FAVORABLE STRUCTURE

- Separability: Classical duality structure (Lagrangian relaxation).
- Partitioning: The problem

$$\begin{aligned}&\text{minimize } F(x) + G(y)\\&\text{subject to } Ax + By = c, \quad x \in X, \ y \in Y\end{aligned}$$

can be written as

$$\begin{aligned}&\text{minimize } F(x) + \inf_{By=c-Ax,\, y\in Y} G(y)\\&\text{subject to } x \in X.\end{aligned}$$

With no duality gap, this problem is written as

$$\begin{aligned}&\text{minimize } F(x) + Q(Ax)\\&\text{subject to } x \in X,\end{aligned}$$

where

$$Q(Ax) = \max_{\lambda} q(\lambda, Ax)$$

$$q(\lambda, Ax) = \inf_{y\in Y} \big\{G(y) + \lambda'(Ax + By - c)\big\}$$

DUAL DERIVATIVES

- Let

$$x_\mu = \arg\min_{x\in X} L(x,\mu) = \arg\min_{x\in X}\big\{f(x)+\mu' g(x)\big\}$$

Then for all $\overline{\mu} \in \Re^r$,

$$\begin{aligned} q(\overline{\mu}) &= \inf_{x\in X}\big\{f(x)+\overline{\mu}' g(x)\big\} \\ &\le f(x_\mu)+\overline{\mu}' g(x_\mu) \\ &= f(x_\mu)+\mu' g(x_\mu)+(\overline{\mu}-\mu)' g(x_\mu) \\ &= q(\mu)+(\overline{\mu}-\mu)' g(x_\mu). \end{aligned}$$

- Thus $g(x_\mu)$ *is a subgradient of* q *at* μ.
- **Proposition:** Let X be compact, and let f and g be continuous over X. Assume also that for every μ, $L(x,\mu)$ is minimized over $x \in X$ at a unique point x_μ. Then, q is everywhere continuously differentiable and

$$\nabla q(\mu) = g(x_\mu), \qquad \forall\, \mu \in \Re^r$$

NONDIFFERENTIABILITY OF THE DUAL

- If there exists a duality gap, the dual function is nondifferentiable at every dual optimal solution (see the textbook).

- Important nondifferentiable case: When q is polyhedral, that is,

$$q(\mu) = \min_{i \in I} \big\{ a_i' \mu + b_i \big\},$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given (arises when X is a discrete set, as in integer programming).

- **Proposition:** Let q be polyhedral as above, and let I_μ be the set of indices attaining the minimum

$$I_\mu = \big\{ i \in I \mid a_i' \mu + b_i = q(\mu) \big\}$$

The set of all subgradients of q at μ is

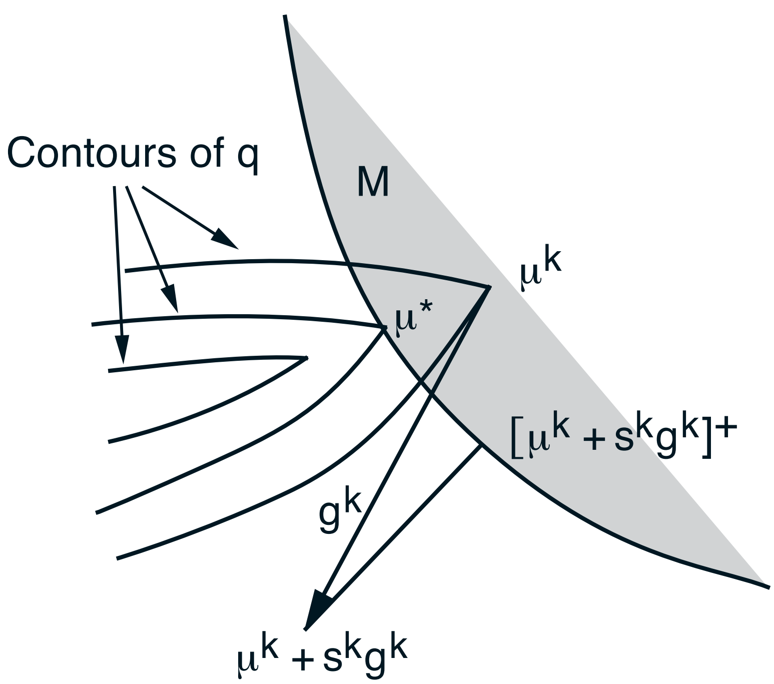
$$\partial q(\mu) = \left\{ g \;\middle|\; g = \sum_{i \in I_\mu} \xi_i a_i,\ \xi_i \geq 0,\ \sum_{i \in I_\mu} \xi_i = 1 \right\}$$

NONDIFFERENTIABLE OPTIMIZATION

- Consider maximization of $q(\mu)$ over $M = \{\mu \geq 0 \mid q(\mu) > -\infty\}$
- Subgradient method:

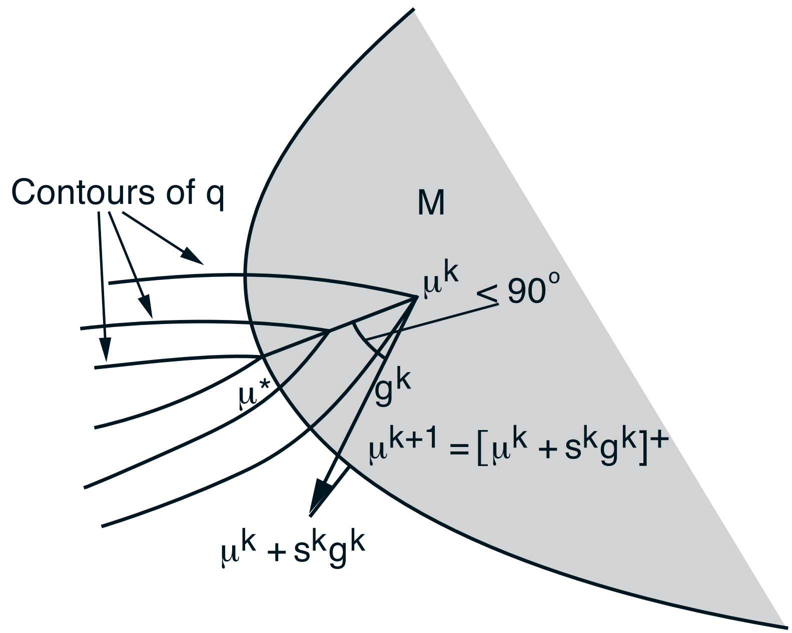
$$\mu^{k+1} = \left[\mu^k + s^k g^k\right]^+,$$

where g^k is the subgradient $g(x_{\mu^k})$, $[\cdot]^+$ denotes projection on the closed convex set M, and s^k is a positive scalar stepsize.



KEY SUBGRADIENT METHOD PROPERTY

- For a small stepsize it reduces the Euclidean distance to the optimum.



- **Proposition:** For any dual optimal solution μ^*, we have

$$\|\mu^{k+1} - \mu^*\| < \|\mu^k - \mu^*\|,$$

for all stepsizes s^k such that

$$0 < s^k < \frac{2\big(q(\mu^*) - q(\mu^k)\big)}{\|g^k\|^2}$$

STEPSIZE RULES

- Constant stepsize: $s^k \equiv s$ for some $s > 0$.
- If $\|g^k\| \le C$ for some constant C and all k,

$$\|\mu^{k+1}-\mu^*\|^2 \le \|\mu^k-\mu^*\|^2-2s\big(q(\mu^*)-q(\mu^k)\big)+s^2C^2,$$

so the distance to the optimum decreases if

$$0 < s < \frac{2\big(q(\mu^*) - q(\mu^k)\big)}{C^2}$$

or equivalently, if μ^k belongs to the level set

$$\left\{\mu \;\middle|\; q(\mu) < q(\mu^*) - \frac{sC^2}{2}\right\}$$

- With a little further analysis, it can be shown that the method, at least asymptotically, reaches this level set, i.e.

$$\limsup_{k\to\infty} q(\mu^k) \ge q(\mu^*) - \frac{sC^2}{2}$$

OTHER STEPSIZE RULES

• Diminishing stepsize: $s^k \to 0$ with some restrictions.

• Dynamic stepsize rule (involves a scalar sequence $\{q^k\}$):

$$s^k = \frac{\alpha^k \big(q^k - q(\mu^k)\big)}{\|g^k\|^2},$$

where $q^k \approx q^*$ and $0 < \alpha^k < 2$.

• Some possibilities:

– q^k is the best known upper bound to q^*: start with $\alpha^0 = 1$ and decrease α^k by a certain factor every few iterations.

– $\alpha^k = 1$ for all k and

$$q^k = \big(1 + \beta(k)\big)\hat{q}^k,$$

where $\hat{q}^k = \max_{0 \le i \le k} q(\mu^i)$, and $\beta(k) > 0$ is adjusted depending on algorithmic progress of the algorithm.

LECTURE 24

LECTURE OUTLINE

- Subgradient Methods
- Stepsize Rules and Convergence Analysis

****

- Consider a generic convex problem $\min_{x \in X} f(x)$, where $f : \Re^n \mapsto \Re$ is a convex function and X is a closed convex set, and the subgradient method

$$x_{k+1} = [x_k - \alpha_k g_k]^+ ,$$

where g_k is a subgradient of f at x_k, α_k is a positive stepsize, and $[\cdot]^+$ denotes projection on the set X.

- Incremental version for problem $\min_{x \in X} \sum_{i=1}^m f_i(x)$

$$x_{k+1} = \psi_{m,k}, \quad \psi_{i,k} = [\psi_{i-1,k} - \alpha_k g_{i,k}]^+ , \; i = 1, \ldots, m$$

starting with $\psi_{0,k} = x_k$, where $g_{i,k}$ is a subgradient of f_i at $\psi_{i-1,k}$.

ASSUMPTIONS AND KEY INEQUALITY

- **Assumption:** (Subgradient Boundedness)

$$||g|| \le C_i, \qquad \forall\, g \in \partial f_i(x_k) \cup \partial f_i(\psi_{i-1,k}), \quad \forall\, i,\, k,$$

for some scalars $C_1, \ldots, C_m$. (Satisfied when the f_i are polyhedral as in integer programming.)

- **Key Lemma:** For all $y \in X$ and k,

$$||x_{k+1}-y||^2 \le ||x_k-y||^2 - 2\alpha_k\big(f(x_k)-f(y)\big) + \alpha_k^2 C^2,$$

where

$$C = \sum_{i=1}^{m} C_i$$

and C_i is as in the boundedness assumption.

- Note: For any y that is better than x_k, the distance to y is improved if α_k is small enough:

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

PROOF OF KEY LEMMA

- For each f_i and all $y \in X$, and i, k

$$
\begin{aligned}
||\psi_{i,k} - y||^2 &= ||[\psi_{i-1,k} - \alpha_k g_{i,k}]^+ - y||^2 \\
&\le ||\psi_{i-1,k} - \alpha_k g_{i,k} - y||^2 \\
&\le ||\psi_{i-1,k} - y||^2 - 2\alpha_k g_{i,k}'(\psi_{i-1,k} - y) + \alpha_k^2 C_i^2 \\
&\le ||\psi_{i-1,k} - y||^2 - 2\alpha_k \big(f_i(\psi_{i-1,k}) - f_i(y)\big) + \alpha_k^2 C_i^2
\end{aligned}
$$

By adding over i, and strengthening,

$$
\begin{aligned}
||x_{k+1} - y||^2 &\le ||x_k - y||^2 - 2\alpha_k \big(f(x_k) - f(y)\big) \\
&\quad + 2\alpha_k \sum_{i=1}^{m} C_i ||\psi_{i-1,k} - x_k|| + \alpha_k^2 \sum_{i=1}^{m} C_i^2 \\
&\le ||x_k - y||^2 - 2\alpha_k \big(f(x_k) - f(y)\big) \\
&\quad + \alpha_k^2 \left(2 \sum_{i=2}^{m} C_i \left(\sum_{j=1}^{i-1} C_j \right) + \sum_{i=1}^{m} C_i^2 \right) \\
&= ||x_k - y||^2 - 2\alpha_k \big(f(x_k) - f(y)\big) + \alpha_k^2 \left(\sum_{i=1}^{m} C_i \right)^2 \\
&= ||x_k - y||^2 - 2\alpha_k \big(f(x_k) - f(y)\big) + \alpha_k^2 C^2.
\end{aligned}
$$

STEPSIZE RULES

- Constant Stepsize $\alpha_k \equiv \alpha$:
  - By key lemma with $f(y) \approx f^*$, it makes progress to the optimal if $0 < \alpha < \frac{2\left(f(x_k)-f^*\right)}{C^2}$, i.e., if

$$f(x_k) > f^* + \frac{\alpha C^2}{2}$$

- Diminishing Stepsize $\alpha_k \to 0$, $\sum_k \alpha_k = \infty$:
  - Eventually makes progress (once α_k becomes small enough). Can show that

$$\liminf_{k\to\infty} f(x_k) = f^*$$

- Dynamic Stepsize $\alpha_k = \frac{f(x_k)-f_k}{C^2}$ where $f_k = f^*$ or (more practically) 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 $\{x \mid f(x) \leq f_k\}$.

CONSTANT STEPSIZE ANALYSIS

- **Proposition:** For $\alpha_k \equiv \alpha$, we have

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \frac{\alpha C^2}{2},$$

where $C = \sum_{i=1}^m C_i$ (in the case where $f^* = -\infty$, we have $\liminf_{k\to\infty} f(x_k) = -\infty$.)

- Proof by contradiction. Let $\epsilon > 0$ be s.t.

$$\liminf_{k\to\infty} f(x_k) > f^* + \frac{\alpha C^2}{2} + 2\epsilon,$$

and let $\hat{y} \in X$ be such that

$$\liminf_{k\to\infty} f(x_k) \geq f(\hat{y}) + \frac{\alpha C^2}{2} + 2\epsilon$$

For all k large enough, we have

$$f(x_k) \geq \liminf_{k\to\infty} f(x_k) - \epsilon$$

Add to get $f(x_k) - f(\hat{y}) \geq \alpha C^2/2 + \epsilon$. Use the key lemma for $y = \hat{y}$ to obtain a contradiction.

COMPLEXITY ESTIMATE FOR CONSTANT STEP

- For any $\epsilon > 0$, we have

$$\min_{0\le k\le K} f(x_k) \le f^* + \frac{\alpha C^2 + \epsilon}{2}$$

where

$$K = \left\lfloor \frac{\big(d(x_0, X^*)\big)^2}{\alpha\epsilon} \right\rfloor$$

- By contradiction. Assume that for $0 \le k \le K$

$$f(x_k) > f^* + \frac{\alpha C^2 + \epsilon}{2}$$

Using this relation in the key lemma,

$$\big(d(x_{k+1}, X^*)\big)^2 \le \Big(d(x_k, X^*)\Big)^2 - 2\alpha\big(f(x_k) - f^*\big) + \alpha^2 C^2$$
$$\le \big(d(x_k, X^*)\big)^2 - (\alpha^2 C^2 + \alpha\epsilon) + \alpha^2 C^2$$
$$= \big(d(x_k, X^*)\big)^2 - \alpha\epsilon.$$

Sum over k to get $\big(d(x_0, X^*)\big)^2 - (K + 1)\alpha\epsilon \ge 0$.

CONVERGENCE FOR OTHER STEPSIZE RULES

- (Diminishing Step): Assume that

$$\alpha_k > 0, \qquad \lim_{k \to \infty} \alpha_k = 0, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty$$

Then,

$$\liminf_{k \to \infty} f(x_k) = f^*$$

If the set of optimal solutions X^* is nonempty and compact,

$$\lim_{k \to \infty} d(x_k, X^*) = 0, \qquad \lim_{k \to \infty} f(x_k) = f^*$$

- (Dynamic Stepsize with $f_k = f^*$): If X^* is nonempty, x_k converges to some optimal solution.

DYNAMIC STEPSIZE WITH ESTIMATE

- Estimation method:

$$f_k^{\text{lev}} = \min_{0 \le j \le k} f(x_j) - \delta_k,$$

and δ_k is updated according to

$$\delta_{k+1} = \begin{cases} \rho\delta_k & \text{if } f(x_{k+1}) \le f_k^{\text{lev}}, \\ \max\{\beta\delta_k, \delta\} & \text{if } f(x_{k+1}) > f_k^{\text{lev}}, \end{cases}$$

where δ, β, and ρ are fixed positive constants with $\beta < 1$ and $\rho \ge 1$.

- Here we essentially "aspire" to reach a target level that is smaller by δ_k over the best value achieved thus far.

- We can show that

$$\inf_{k \ge 0} f(x_k) \le f^* + \delta$$

(or $\inf_{k \ge 0} f(x_k) = f^*$ if $f^* = -\infty$).

LECTURE 25

LECTURE OUTLINE

- Incremental Subgradient Methods
- Convergence Rate Analysis and Randomized Methods

****

- Incremental subgradient method for problem $\min_{x\in X} \sum_{i=1}^{m} f_i(x)$

$$x_{k+1} = \psi_{m,k}, \quad \psi_{i,k} = \left[\psi_{i-1,k} - \alpha_k g_{i,k}\right]^+, \, i = 1, \ldots, m$$

starting with $\psi_{0,k} = x_k$, where $g_{i,k}$ is a subgradient of f_i at $\psi_{i-1,k}$.

- **Key Lemma:** For all $y \in X$ and k,

$$||x_{k+1}-y||^2 \le ||x_k-y||^2 - 2\alpha_k\big(f(x_k)-f(y)\big) + \alpha_k^2 C^2,$$

where $C = \sum_{i=1}^{m} C_i$ and

$$C_i = \sup_k \big\{ ||g|| \mid g \in \partial f_i(x_k) \cup \partial f_i(\psi_{i-1,k}) \big\}$$

CONSTANT STEPSIZE

- For $\alpha_k \equiv \alpha$, we have

$$\liminf_{k\to\infty} f(x_k) \leq f^* + \frac{\alpha C^2}{2}$$

- Sharpness of the estimate:
  - Consider the problem

$$\min_x \sum_{i=1}^{M} C_0\big(|x+1| + 2|x| + |x-1|\big)$$

with the worst component processing order

- Lower bound on the error. There is a problem, where even with best processing order,

$$f^* + \frac{\alpha m C_0^2}{2} \leq \liminf_{k\to\infty} f(x_k)$$

where

$$C_0 = \max\{C_1, \ldots, C_m\}$$

COMPLEXITY ESTIMATE FOR CONSTANT STEP

- For any $\epsilon > 0$, we have

$$\min_{0 \le k \le K} f(x_k) \le f^* + \frac{\alpha C^2 + \epsilon}{2}$$

where

$$K = \left\lfloor \frac{\big(d(x_0, X^*)\big)^2}{\alpha \epsilon} \right\rfloor$$

RANDOMIZED ORDER METHODS

$$x_{k+1} = \left[x_k - \alpha_k g(\omega_k, x_k)\right]^+$$

where ω_k is a random variable taking equiprobable values from the set $\{1, \ldots, m\}$, and $g(\omega_k, x_k)$ is a subgradient of the component f_{ω_k} at x_k.

- Assumptions:
  (a) $\{\omega_k\}$ is a sequence of independent random variables. Furthermore, the sequence $\{\omega_k\}$ is independent of the sequence $\{x_k\}$.
  (b) The set of subgradients $\{g(\omega_k, x_k) \mid k = 0, 1, \ldots\}$ is bounded, i.e., there exists a positive constant C_0 such that with prob. 1

$$||g(\omega_k, x_k)|| \leq C_0, \qquad \forall\, k \geq 0$$

- Stepsize Rules:
  - Constant: $\alpha_k \equiv \alpha$
  - Diminishing: $\sum_k \alpha_k = \infty$, $\sum_k (\alpha_k)^2 < \infty$
  - Dynamic

RANDOMIZED METHOD W/ CONSTANT STEP

- With probability 1

$$\inf_{k\geq 0} f(x_k) \leq f^* + \frac{\alpha m C_0^2}{2}$$

(with $\inf_{k\geq 0} f(x_k) = -\infty$ when $f^* = -\infty$).

Proof: By adapting key lemma, for all $y \in X,\ k$

$$\|x_{k+1}-y\|^2 \leq \|x_k-y\|^2 - 2\alpha\big(f_{\omega_k}(x_k) - f_{\omega_k}(y)\big) + \alpha^2 C_0^2$$

Take conditional expectation with $\mathcal{F}_k = \{x_0, \ldots, x_k\}$

$$\begin{aligned} E\big\{\|x_{k+1}-y\|^2 \mid \mathcal{F}_k\big\} &\leq \|x_k - y\|^2 \\ &\quad - 2\alpha E\big\{f_{\omega_k}(x_k) - f_{\omega_k}(y) \mid \mathcal{F}_k\big\} + \alpha^2 C_0^2 \\ &= \|x_k - y\|^2 - 2\alpha \sum_{i=1}^{m} \frac{1}{m}\big(f_i(x_k) - f_i(y)\big) + \alpha^2 C_0^2 \\ &= \|x_k - y\|^2 - \frac{2\alpha}{m}\big(f(x_k) - f(y)\big) + \alpha^2 C_0^2, \end{aligned}$$

where the first equality follows since ω_k takes the values $1, \ldots, m$ with equal probability $1/m$.

PROOF CONTINUED I

- Fix $\gamma > 0$, consider the level set L_γ defined by

$$L_\gamma = \left\{ x \in X \mid f(x) < f^* + \frac{2}{\gamma} + \frac{\alpha m C_0^2}{2} \right\}$$

and let $y_\gamma \in L_\gamma$ be such that $f(y_\gamma) = f^* + \frac{1}{\gamma}$. Define a new process $\{\hat{x}_k\}$ as follows

$$\hat{x}_{k+1} = \begin{cases} \left[\hat{x}_k - \alpha g(\omega_k, \hat{x}_k)\right]^+ & \text{if } \hat{x}_k \notin L_\gamma, \\ y_\gamma & \text{otherwise,} \end{cases}$$

where $\hat{x}_0 = x_0$. We argue that $\{\hat{x}_k\}$ (and hence also $\{x_k\}$) will eventually enter each of the sets L_γ.

Using key lemma with $y = y_\gamma$, we have

$$E\left\{ ||\hat{x}_{k+1} - y_\gamma||^2 \mid \mathcal{F}_k \right\} \leq ||\hat{x}_k - y_\gamma||^2 - z_k,$$

where

$$z_k = \begin{cases} \frac{2\alpha}{m}\big(f(\hat{x}_k) - f(y_\gamma)\big) - \alpha^2 C_0^2 & \text{if } \hat{x}_k \notin L_\gamma, \\ 0 & \text{if } \hat{x}_k = y_\gamma. \end{cases}$$

PROOF CONTINUED II

- If $\hat{x}_k \notin L_\gamma$, we have

$$
\begin{aligned}
z_k &= \frac{2\alpha}{m}\big(f(\hat{x}_k) - f(y_\gamma)\big) - \alpha^2 C_0^2 \\
&\geq \frac{2\alpha}{m}\left(f^* + \frac{2}{\gamma} + \frac{\alpha m C_0^2}{2} - f^* - \frac{1}{\gamma}\right) - \alpha^2 C_0^2 \\
&= \frac{2\alpha}{m\gamma}.
\end{aligned}
$$

Hence, as long as $\hat{x}_k \notin L_\gamma$, we have

$$
E\big\{||\hat{x}_{k+1} - y_\gamma||^2 \mid \mathcal{F}_k\big\} \leq ||\hat{x}_k - y_\gamma||^2 - \frac{2\alpha}{m\gamma}
$$

This, cannot happen for an infinite number of iterations, so that $\hat{x}_k \in L_\gamma$ for sufficiently large k. Hence, in the original process we have

$$
\inf_{k\geq 0} f(x_k) \leq f^* + \frac{2}{\gamma} + \frac{\alpha m C_0^2}{2}
$$

with probability 1. Letting $\gamma \to \infty$, we obtain $\inf_{k\geq 0} f(x_k) \leq f^* + \alpha m C_0^2/2$. **Q.E.D.**

CONVERGENCE RATE

- Let $\alpha_k \equiv \alpha$ in the randomized method. Then, for any positive scalar ϵ, we have with prob. 1

$$\min_{0 \le k \le N} f(x_k) \le f^* + \frac{\alpha m C_0^2 + \epsilon}{2},$$

where N is a random variable with

$$E\{N\} \le \frac{m\big(d(x_0, X^*)\big)^2}{\alpha \epsilon}$$

- Compare w/ the deterministic method. It is guaranteed to reach after processing no more than

$$K = \frac{m\big(d(x_0, X^*)\big)^2}{\alpha \epsilon}$$

components the level set

$$\left\{ x \;\middle|\; f(x) \le f^* + \frac{\alpha m^2 C_0^2 + \epsilon}{2} \right\}$$

BASIC TOOL FOR PROVING CONVERGENCE

- **Supermartingale Convergence Theorem:** Let Y_k, Z_k, and W_k, $k = 0, 1, 2, \ldots$, be three sequences of random variables and let $\mathcal{F}_k$, $k = 0, 1, 2, \ldots$, be sets of random variables such that $\mathcal{F}_k \subset \mathcal{F}_{k+1}$ for all k. Suppose that:

  (a) The random variables Y_k, Z_k, and W_k are nonnegative, and are functions of the random variables in $\mathcal{F}_k$.

  (b) For each k, we have

  $$E\big\{Y_{k+1} \mid \mathcal{F}_k\big\} \leq Y_k - Z_k + W_k$$

  (c) There holds $\sum_{k=0}^{\infty} W_k < \infty$.

  Then, $\sum_{k=0}^{\infty} Z_k < \infty$, and the sequence Y_k converges to a nonnegative random variable Y, with prob. 1.

- Can be used to show convergence of randomized subgradient methods with diminishing and dynamic stepsize rules.

LECTURE 26

LECTURE OUTLINE

- Additional Dual Methods
- Cutting Plane Methods
- Decomposition

****

- Consider the primal problem

minimize $f(x)$
subject to $x \in X, \qquad g_j(x) \le 0, \qquad j = 1, \ldots, r,$

assuming $-\infty < f^* < \infty$.

- Dual problem: Maximize

$$q(\mu) = \inf_{x \in X} L(x, \mu) = \inf_{x \in X} \{f(x) + \mu' g(x)\}$$

subject to $\mu \in M = \{\mu \mid \mu \ge 0,\ q(\mu) > -\infty\}$.

CUTTING PLANE METHOD

- kth iteration, after μ^i and $g^i = g\big(x_{\mu^i}\big)$ have been generated for $i = 0, \ldots, k-1$: Solve

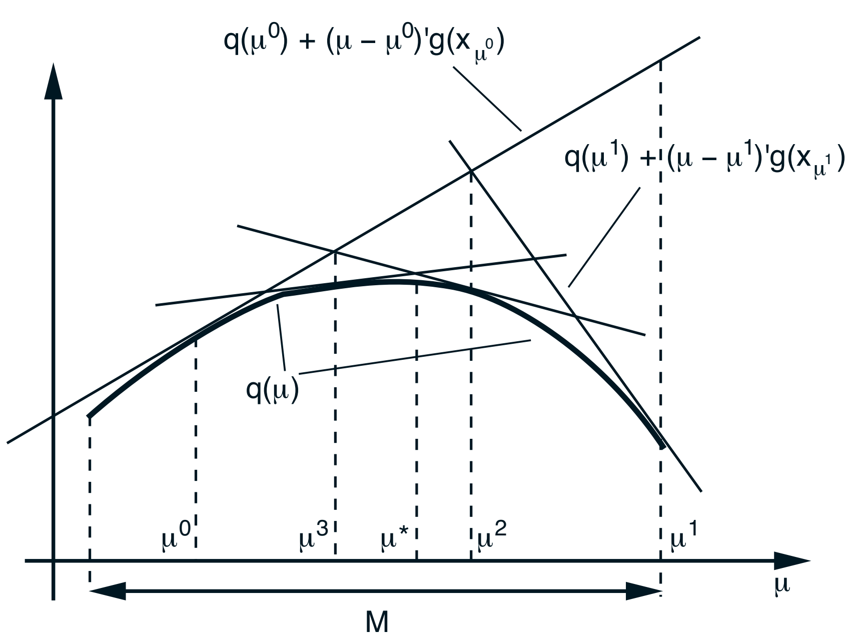
$$\max_{\mu \in M} Q^k(\mu)$$

where

$$Q^k(\mu) = \min_{i=0,\ldots,k-1} \big\{ q(\mu^i) + (\mu - \mu^i)' g^i \big\}$$

Set

$$\mu^k = \arg\max_{\mu \in M} Q^k(\mu)$$



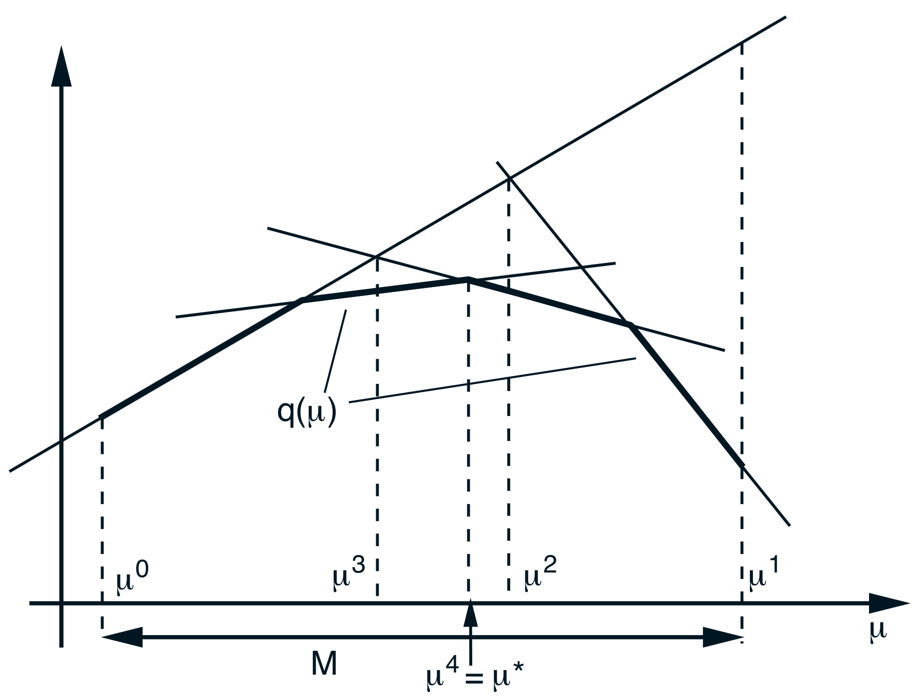
POLYHEDRAL CASE

$$q(\mu) = \min_{i \in I} \{ a_i' \mu + b_i \}$$

where I is a finite index set, and $a_i \in \Re^r$ and b_i are given.

- Then subgradient g^k in the cutting plane method is a vector a_{i^k} for which the minimum is attained.

- Finite termination expected.



CONVERGENCE

- **Proposition:** Assume that the min of Q_k over M is attained and that the sequence g^k is bounded. Then every limit point of a sequence $\{\mu^k\}$ generated by the cutting plane method is a dual optimal solution.

Proof: g^i is a subgradient of q at μ^i, so

$$q(\mu^i) + (\mu - \mu^i)' g^i \geq q(\mu), \qquad \forall\, \mu \in M,$$

$$Q^k(\mu^k) \geq Q^k(\mu) \geq q(\mu), \qquad \forall\, \mu \in M. \qquad (1)$$

- Suppose $\{\mu^k\}_K$ converges to $\bar{\mu}$. Then, $\bar{\mu} \in M$, and from (1), we obtain for all k and $i < k$,

$$q(\mu^i) + (\mu^k - \mu^i)' g^i \geq Q^k(\mu^k) \geq Q^k(\bar{\mu}) \geq q(\bar{\mu})$$

- Take the limit as $i \to \infty$, $k \to \infty$, $i \in K$, $k \in K$,

$$\lim_{k \to \infty,\, k \in K} Q^k(\mu^k) = q(\bar{\mu})$$

Combining with (1), $q(\bar{\mu}) = \max_{\mu \in M} q(\mu)$.

LAGRANGIAN RELAXATION

- Solving the dual of the separable problem

$$\text{minimize} \quad \sum_{j=1}^{J} f_j(x_j)$$

$$\text{subject to} \quad x_j \in X_j, \quad j = 1, \ldots, J, \qquad \sum_{j=1}^{J} A_j x_j = b.$$

- Dual function is

$$q(\lambda) = \sum_{j=1}^{J} \min_{x_j \in X_j} \big\{ f_j(x_j) + \lambda' A_j x_j \big\} - \lambda' b$$

$$= \sum_{j=1}^{J} \Big\{ f_j\big(x_j(\lambda)\big) + \lambda' A_j x_j(\lambda) \Big\} - \lambda' b$$

where $x_j(\lambda)$ attains the min. A subgradient at λ is

$$g_\lambda = \sum_{j=1}^{J} A_j x_j(\lambda) - b$$

DANTSIG-WOLFE DECOMPOSITION

- D-W decomposition method is just the cutting plane applied to the dual problem $\max_\lambda q(\lambda)$.

- At the kth iteration, we solve the "approximate dual"

$$\lambda^k = \arg\max_{\lambda \in \Re^r} Q^k(\lambda) \equiv \min_{i=0,\ldots,k-1} \left\{ q(\lambda^i) + (\lambda - \lambda^i)' g^i \right\}$$

- Equivalent linear program in v and λ

maximize v

subject to $v \le q(\lambda^i) + (\lambda - \lambda^i)' g^i, \quad i = 0, \ldots, k-1$

The dual of this (called *master problem*) is

$$\text{minimize} \quad \sum_{i=0}^{k-1} \xi^i \big(q(\lambda^i) - \lambda^{i\prime} g^i \big)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \qquad \sum_{i=0}^{k-1} \xi^i g^i = 0,$$

$$\xi^i \ge 0, \quad i = 0, \ldots, k-1,$$

DANTSIG-WOLFE DECOMPOSITION (CONT.)

- The master problem is written as

$$\text{minimize} \quad \sum_{j=1}^{J} \left(\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big) \right)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \qquad \sum_{j=1}^{J} A_j \left(\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i) \right) = b,$$

$$\xi^i \geq 0, \quad i = 0, \ldots, k-1.$$

- The primal cost function terms $f_j(x_j)$ are approximated by

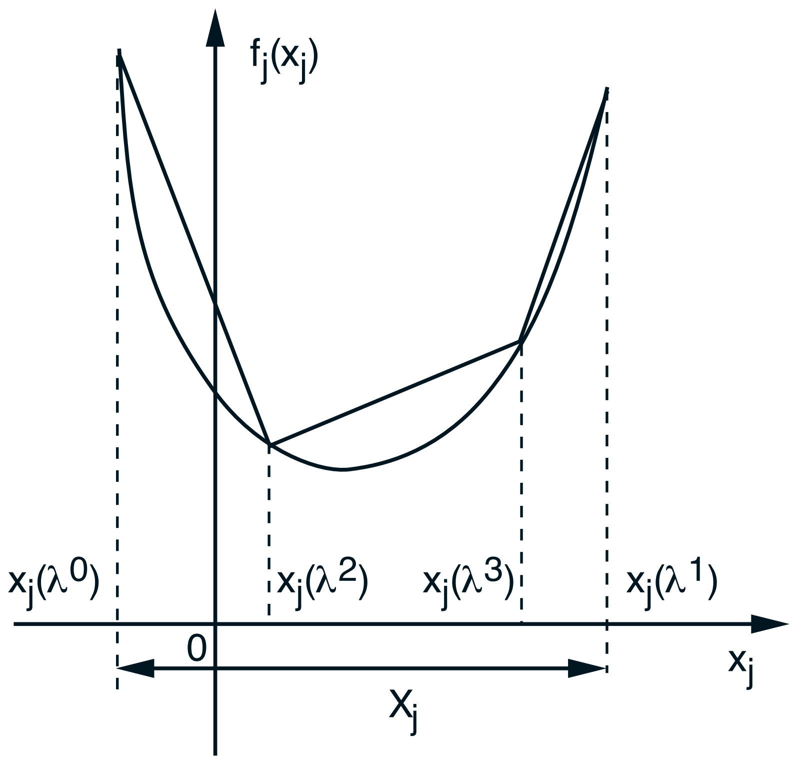
$$\sum_{i=0}^{k-1} \xi^i f_j\big(x_j(\lambda^i)\big)$$

- Vectors x_j are expressed as

$$\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i)$$

GEOMETRICAL INTERPRETATION

- Geometric interpretation of the master problem (the dual of the approximate dual solved in the cutting plane method) is *inner linearization*.



- This is a "dual" operation to the one involved in the cutting plane approximation, which can be viewed as *outer linearization*.