

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: 11/20/2005

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

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

The slides are copyrighted but may be freely reproduced and distributed for any noncommercial purpose.

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 $f : \mathbb{R}^n \mapsto \mathbb{R}$, constraint set C , e.g.,

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

- Examples of problem classifications:
 - Continuous vs discrete
 - Linear vs nonlinear
 - Deterministic vs stochastic
 - Static vs dynamic
- Convex programming problems are those for which f is convex and C is convex (they are continuous problems).
- However, convexity permeates all of optimization, including discrete problems.

WHY IS CONVEXITY SO SPECIAL 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 $g_1(x) \leq 0, \dots, g_r(x) \leq 0$

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

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

where L is the Lagrangian function

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

- Dual function (always concave)

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

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

KEY DUALITY RELATIONS

- Optimal primal value

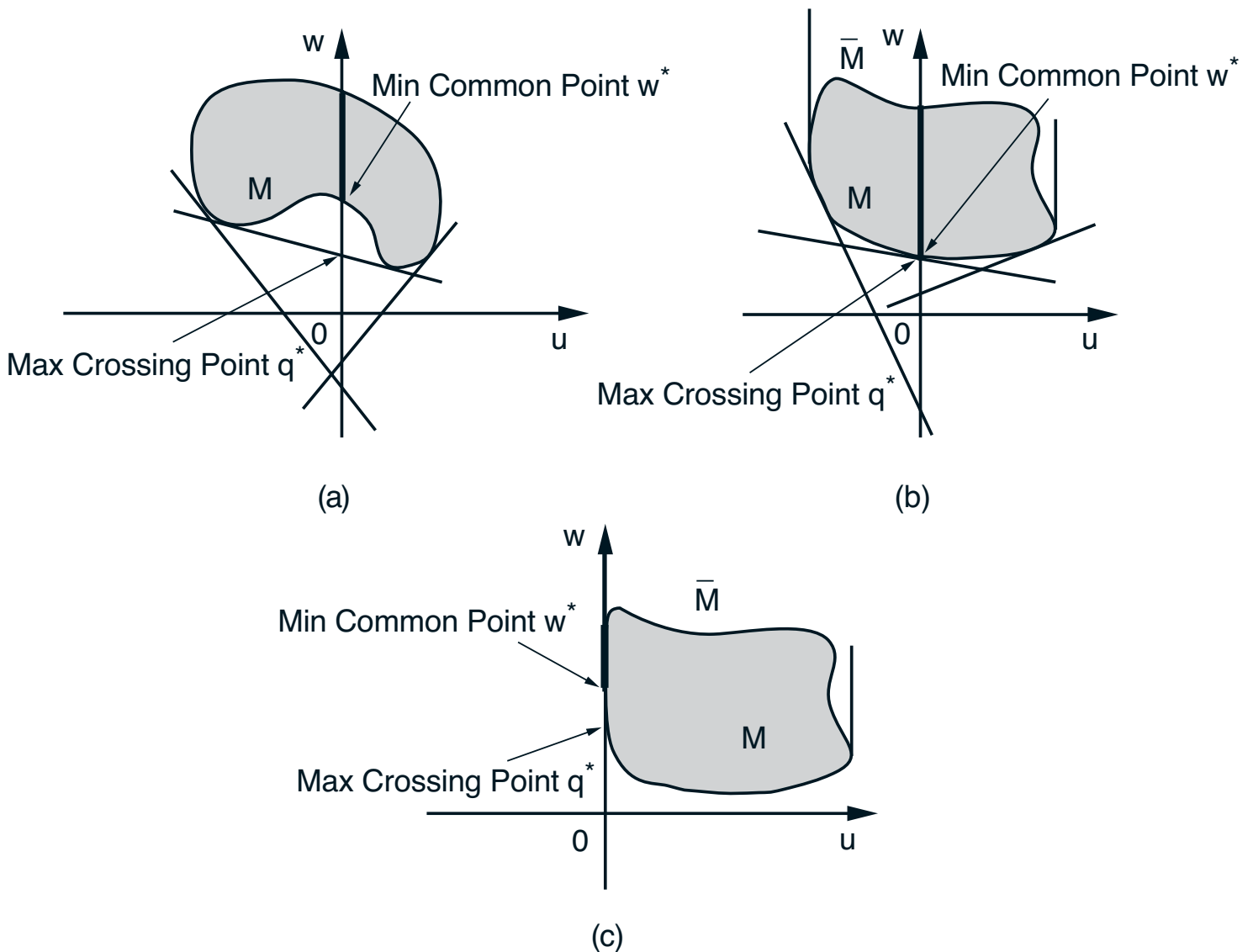
$$f^* = \inf_{g_j(x) \leq 0, j=1, \dots, r} f(x) = \inf_{x \in \mathcal{R}^n} \sup_{\mu \geq 0} L(x, \mu)$$

- Optimal dual value

$$q^* = \sup_{\mu \geq 0} q(\mu) = \sup_{\mu \geq 0} \inf_{x \in \mathcal{R}^n} L(x, \mu)$$

- We always have $q^* \leq f^*$ (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.

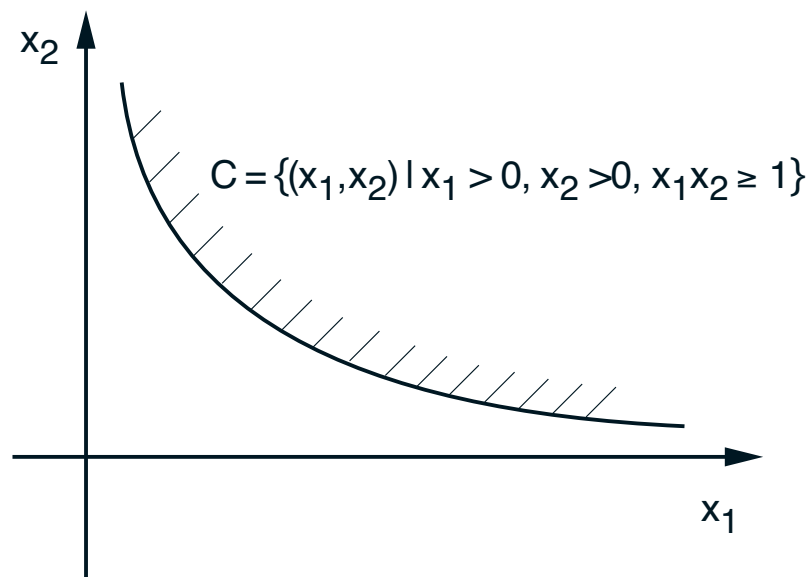
MIN COMMON/MAX CROSSING DUALITY



- 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

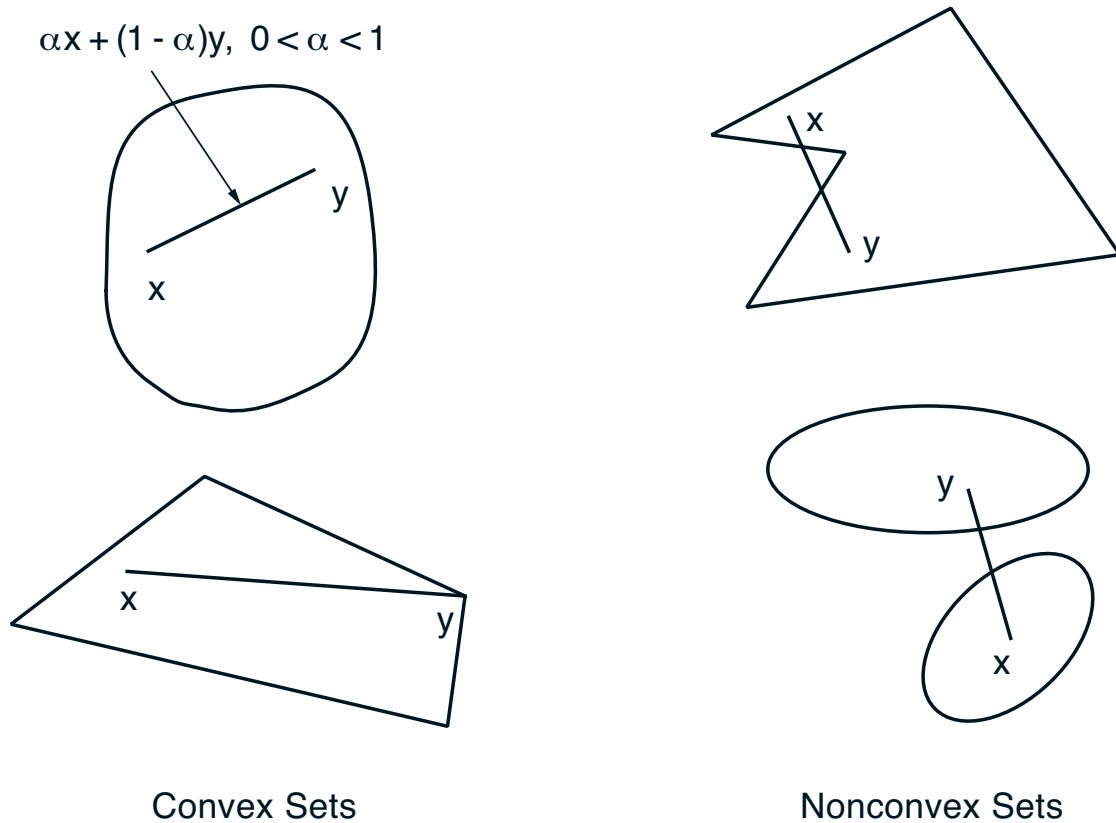
LECTURE OUTLINE

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

SOME MATH CONVENTIONS

- All of our work is done in \mathfrak{R}^n : space of n -tuples $x = (x_1, \dots, x_n)$
- 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

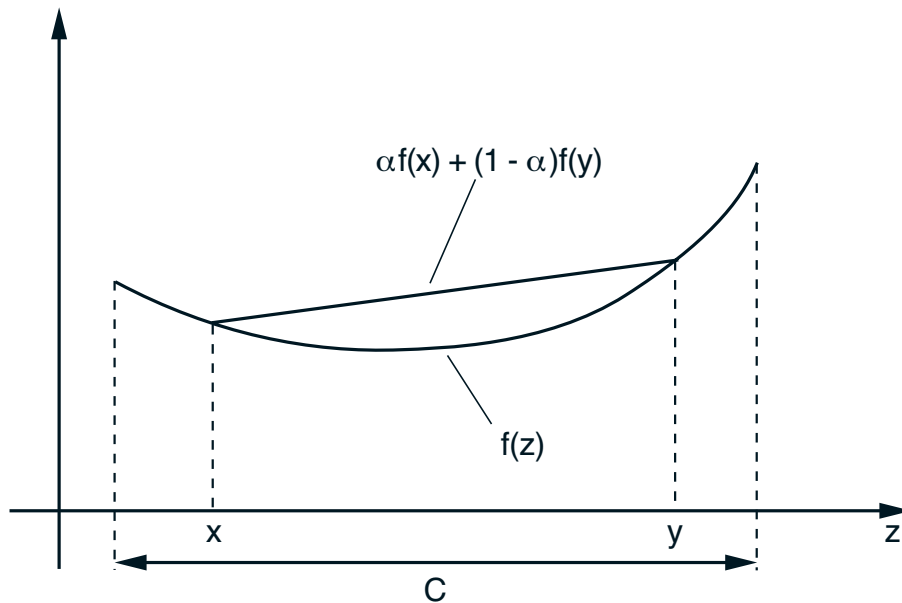


- A subset C of \mathbb{R}^n is called *convex* if

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

- 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 \mathbb{R}^n . A function $f : C \mapsto \mathbb{R}$ is called *convex* if

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

- 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 \mathfrak{R}^{n+1} given by

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

- The *effective domain* of f is the set

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

- 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.”

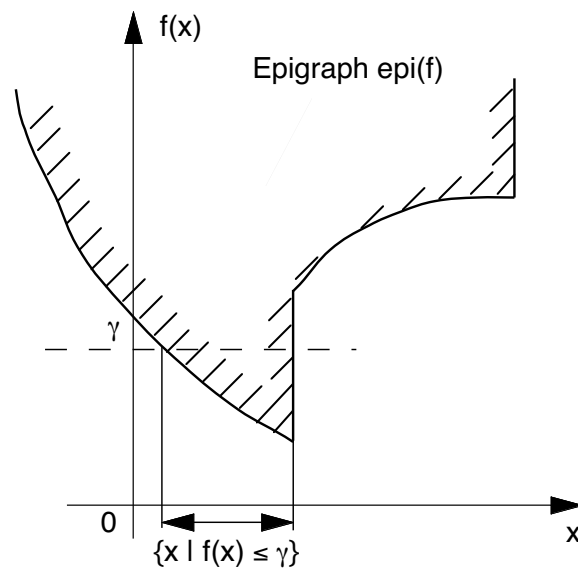
- An extended real-valued function $f : X \mapsto [-\infty, \infty]$ is called *lower semicontinuous* at a vector $x \in X$ if $f(x) \leq \liminf_{k \rightarrow \infty} f(x_k)$ for every sequence $\{x_k\} \subset X$ with $x_k \rightarrow x$.

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

CLOSEDNESS AND SEMICONTINUITY

• *Proposition:* For a function $f : \mathbb{R}^n \mapsto [-\infty, \infty]$, the following are equivalent:

- (i) $\{x \mid f(x) \leq a\}$ is closed for every scalar a .
- (ii) f is lower semicontinuous at all $x \in \mathbb{R}^n$.
- (iii) f is closed.

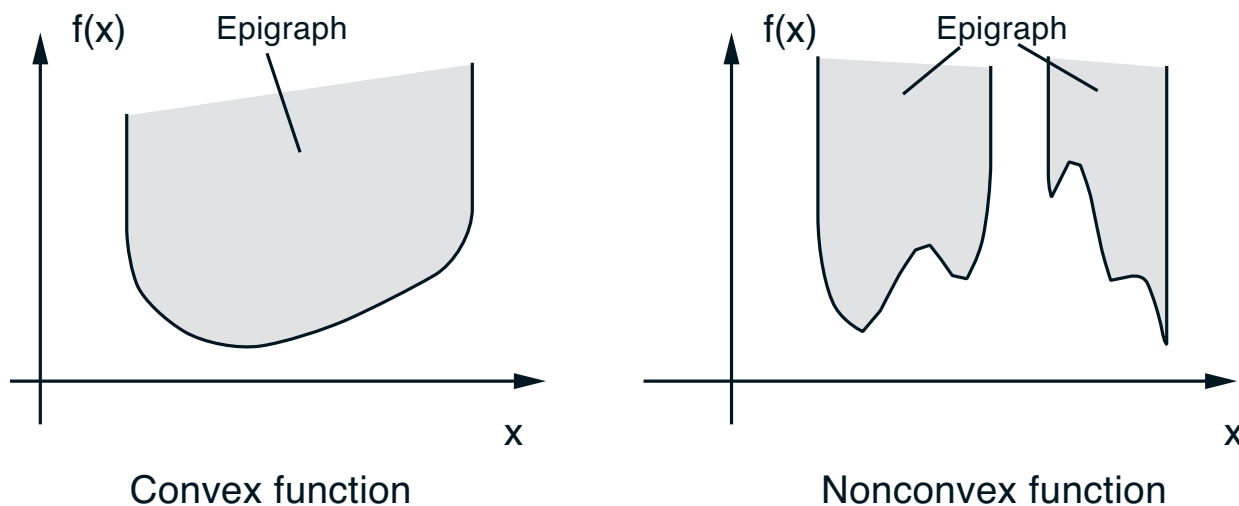


• Note that:

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

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

EXTENDED REAL-VALUED CONVEX FUNCTIONS



- Let C be a convex subset of \mathbb{R}^n . An extended real-valued function $f : C \mapsto [-\infty, \infty]$ is called *convex* if $\text{epi}(f)$ is a convex subset of \mathbb{R}^{n+1} .
- If f is proper, this definition is equivalent to
$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y), \quad \forall x, y \in C$$
- 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 $f_i : \mathbb{R}^n \mapsto (-\infty, \infty]$, $i \in I$, be given functions (I is an arbitrary index set).

(a) The function $g : \mathbb{R}^n \mapsto (-\infty, \infty]$ given by

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

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

(b) The function $g : \mathbb{R}^n \mapsto (-\infty, \infty]$ 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 $g : \mathbb{R}^n \mapsto (-\infty, \infty]$ given by

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

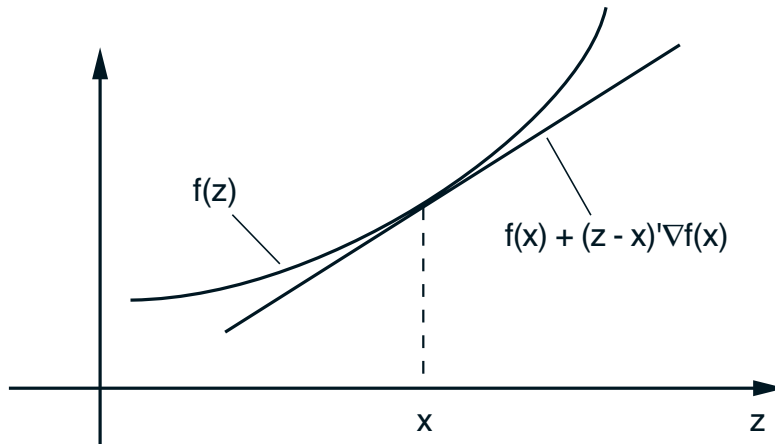
is convex (or closed) if the f_i are convex (respectively, closed).

LECTURE 3

LECTURE OUTLINE

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

DIFFERENTIABLE CONVEX FUNCTIONS



• Let $C \subset \mathbb{R}^n$ be a convex set and let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be differentiable over \mathbb{R}^n .

(a) The function f is convex over C if and only if

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

(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 $x \neq y$

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

TWICE DIFFERENTIABLE CONVEX FUNCTIONS

- Let C be a convex subset of \mathbb{R}^n and let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be twice continuously differentiable over \mathbb{R}^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 $x, y \in C$

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

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

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

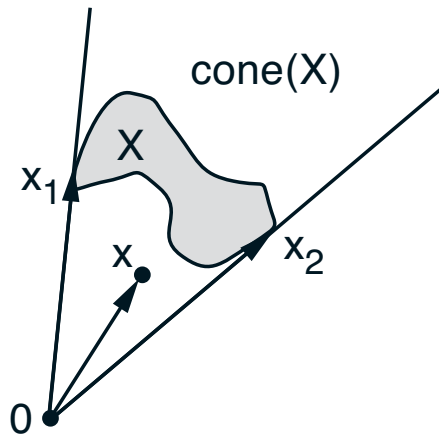
From the preceding result, f is convex.

(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.

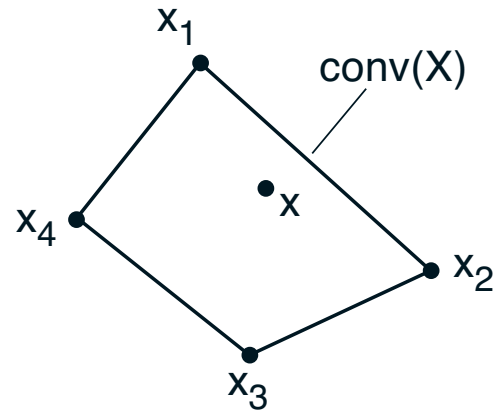
CONVEX AND AFFINE HULLS

- Given a set $X \subset \mathbb{R}^n$:
- 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 $\bar{x} + S$, 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 (non-trivial to show!)

CARATHEODORY'S THEOREM



(a)



(b)

- Let X be a nonempty subset of \mathbb{R}^n .
 - (a) Every $x \neq 0$ in $\text{cone}(X)$ can be represented as a positive combination of vectors x_1, \dots, x_m from X that are linearly independent.
 - (b) Every $x \notin X$ that belongs to $\text{conv}(X)$ can be represented as a convex combination of vectors x_1, \dots, x_m from X such that $x_2 - x_1, \dots, x_m - x_1$ are linearly independent.

PROOF OF CARATHEODORY'S THEOREM

(a) Let x be a nonzero vector in $\text{cone}(X)$, 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 $i = 1, \dots, m$. If the vectors x_i were linearly dependent, there would exist $\lambda_1, \dots, \lambda_m$, with

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

and at least one of the λ_i is positive. Consider

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

where $\bar{\gamma}$ 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, \dots, x_m , are linearly independent.

(b) Apply part (a) to the subset of \mathfrak{R}^{n+1}

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

AN APPLICATION OF CARATHEODORY

- The convex hull of a compact set is compact.

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

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

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

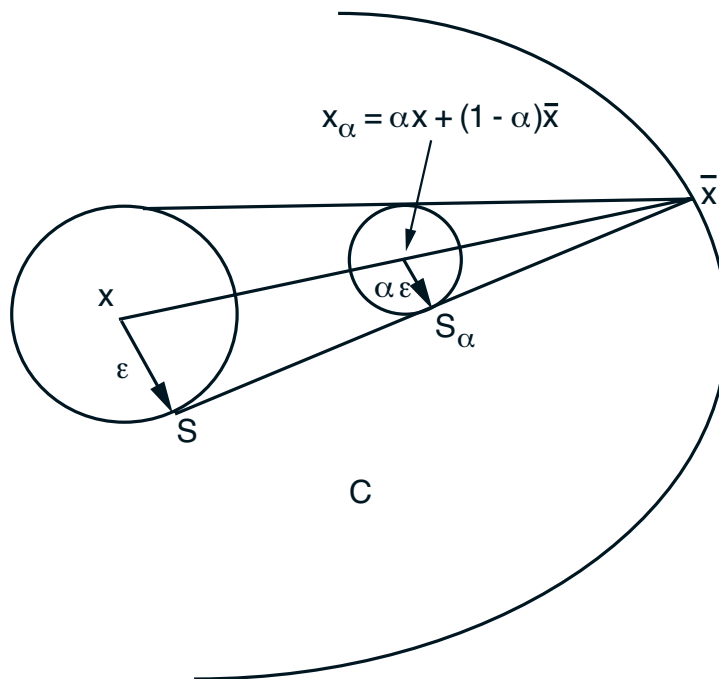
is bounded, it has a limit point

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

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

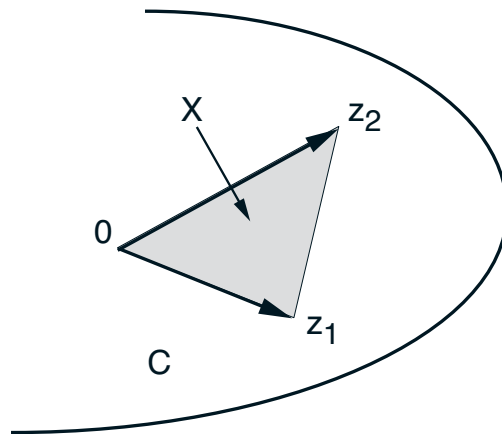
RELATIVE INTERIOR

- x is a *relative interior point* of C , if x is an interior point of C relative to $\text{aff}(C)$.
- $\text{ri}(C)$ denotes the *relative interior* of C , i.e., the set of all relative interior points of C .
- **Line Segment Principle:** If C is a convex set, $x \in \text{ri}(C)$ and $\bar{x} \in \text{cl}(C)$, then all points on the line segment connecting x and \bar{x} , except possibly \bar{x} , belong to $\text{ri}(C)$.



ADDITIONAL MAJOR RESULTS

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



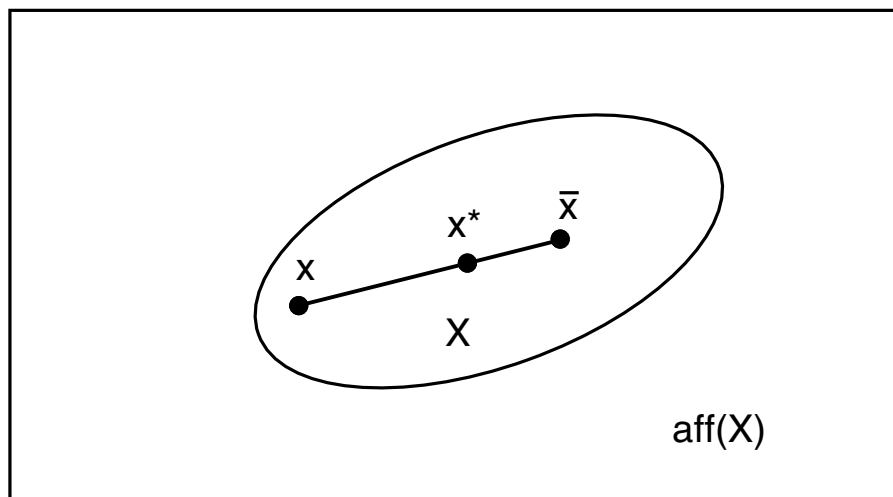
Proof: (a) Assume that $0 \in C$. We choose m linearly independent vectors $z_1, \dots, z_m \in C$, where m is the dimension of $\text{aff}(C)$, and we let

$$X = \left\{ \sum_{i=1}^m \alpha_i z_i \mid \sum_{i=1}^m \alpha_i < 1, \alpha_i > 0, i = 1, \dots, m \right\}$$

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

OPTIMIZATION APPLICATION

- A concave function $f : \mathbb{R}^n \mapsto \mathbb{R}$ that attains its minimum over a convex set X at an $x^* \in \text{ri}(X)$ 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 $\bar{x} \in X$. By concavity of f , we have for some $\alpha \in (0, 1)$

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

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

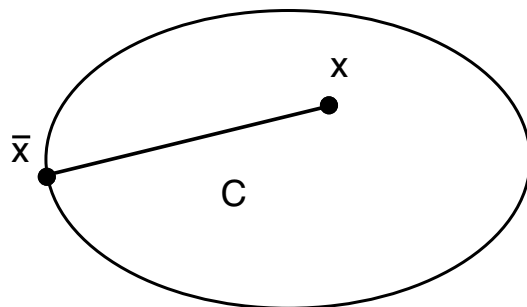
A SUMMARY OF FACTS

- 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 $\text{ri}(C)$ and $\text{cl}(C)$ are “not too different for each other.”
- *Proposition:*
 - (a) We have $\text{cl}(C) = \text{cl}(\text{ri}(C))$.
 - (b) We have $\text{ri}(C) = \text{ri}(\text{cl}(C))$.
 - (c) Let \bar{C} be another nonempty convex set. Then the following three conditions are equivalent:
 - (i) C and \bar{C} have the same rel. interior.
 - (ii) C and \bar{C} have the same closure.
 - (iii) $\text{ri}(C) \subset \bar{C} \subset \text{cl}(C)$.

Proof: (a) Since $\text{ri}(C) \subset C$, we have $\text{cl}(\text{ri}(C)) \subset \text{cl}(C)$. Conversely, let $\bar{x} \in \text{cl}(C)$. Let $x \in \text{ri}(C)$. By the Line Segment Principle, we have $\alpha x + (1 - \alpha)\bar{x} \in \text{ri}(C)$ for all $\alpha \in (0, 1]$. Thus, \bar{x} is the limit of a sequence that lies in $\text{ri}(C)$, so $\bar{x} \in \text{cl}(\text{ri}(C))$.



LINEAR TRANSFORMATIONS

• Let C be a nonempty convex subset of \mathfrak{R}^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 $Ax_k \rightarrow z$. 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

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

INTERSECTIONS AND VECTOR SUMS

- Let C_1 and C_2 be nonempty convex sets.
(a) We have

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

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

If one of C_1 and C_2 is bounded, then

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

- (b) If $\text{ri}(C_1) \cap \text{ri}(C_2) \neq \emptyset$, then

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

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

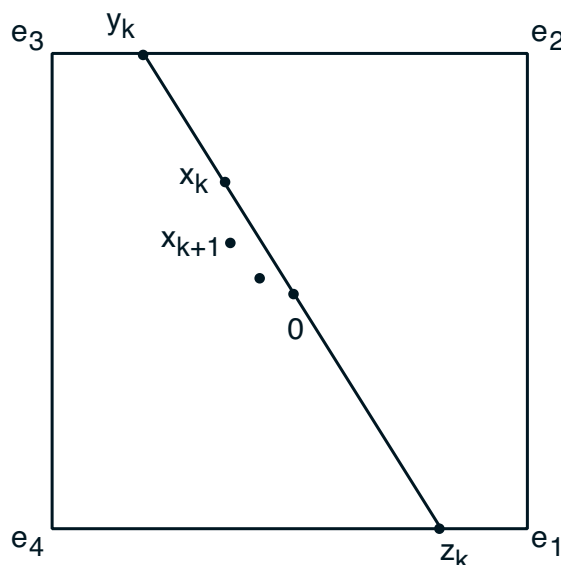
Proof of (a): $C_1 + C_2$ is the result of the linear transformation $(x_1, x_2) \mapsto x_1 + x_2$.

- Counterexample for (b):

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

CONTINUITY OF CONVEX FUNCTIONS

- If $f : \mathbb{R}^n \mapsto \mathbb{R}$ 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 $x_k \rightarrow 0$ and the sequences $y_k = x_k / \|x_k\|_\infty$, $z_k = -x_k / \|x_k\|_\infty$. Then

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

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

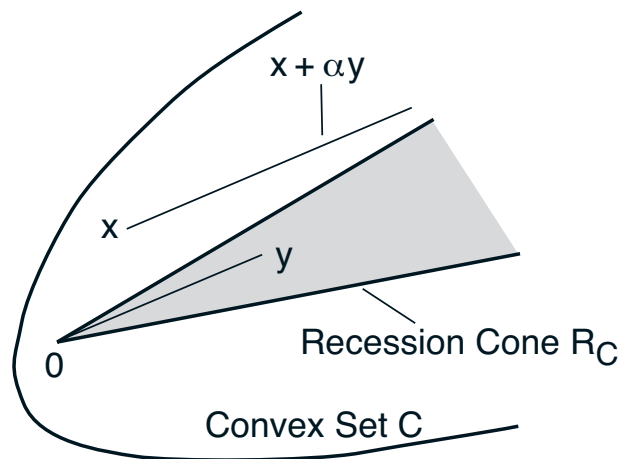
Since $\|x_k\|_\infty \rightarrow 0$, $f(x_k) \rightarrow f(0)$. **Q.E.D.**

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

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 :

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



- *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

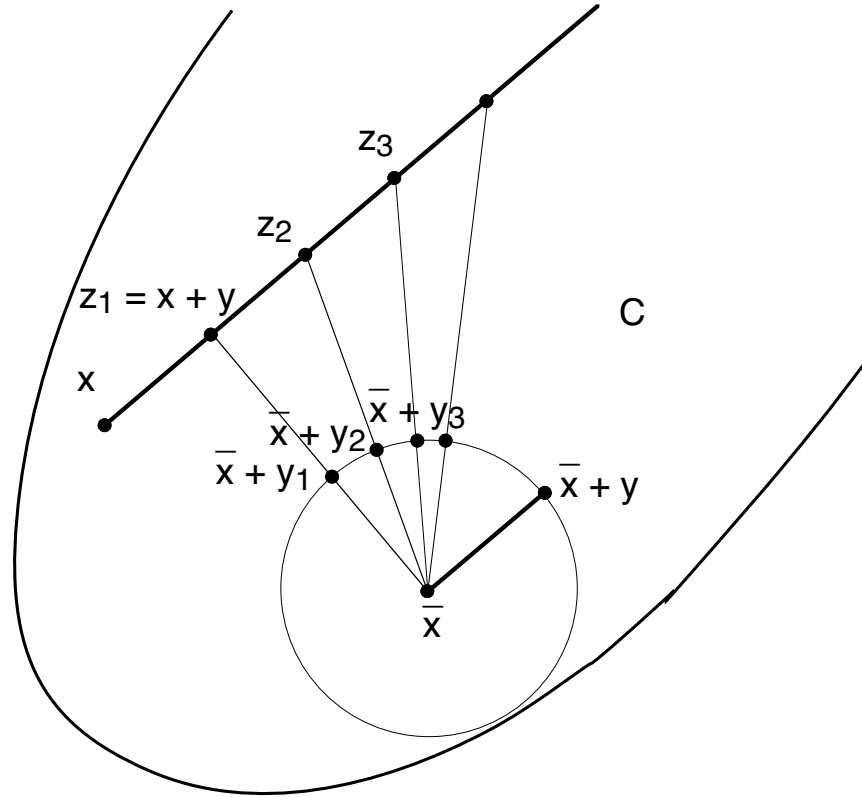
- Let C be a nonempty closed convex set.
 - (a) The recession cone R_C is a closed convex cone.
 - (b) A vector y belongs to R_C if and only if there exists a vector $x \in C$ such that $x + \alpha y \in C$ for all $\alpha \geq 0$.
 - (c) R_C contains a nonzero direction if and only if C is unbounded.
 - (d) The recession cones of C and $\text{ri}(C)$ 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 $\bigcap_{i \in I} C_i$ is nonempty, we have

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

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 $\bar{x} \in C$ and $\alpha > 0$, and we show that $\bar{x} + \alpha y \in C$. By scaling y , it is enough to show that $\bar{x} + y \in C$.

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

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

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

LINEALITY SPACE

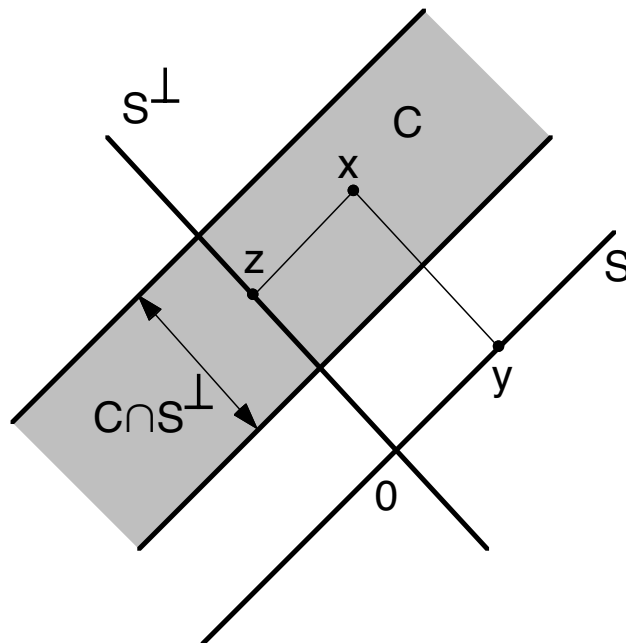
- The *lineality space* of a convex set C , denoted by L_C , is the subspace of vectors y such that $y \in R_C$ and $-y \in R_C$:

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

- *Decomposition of a Convex Set:* Let C be a nonempty convex subset of \mathbb{R}^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).

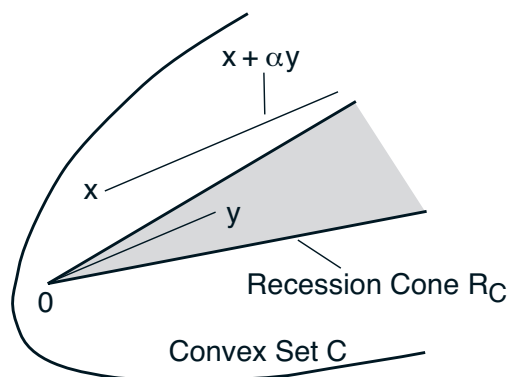


LECTURE 5

LECTURE OUTLINE

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

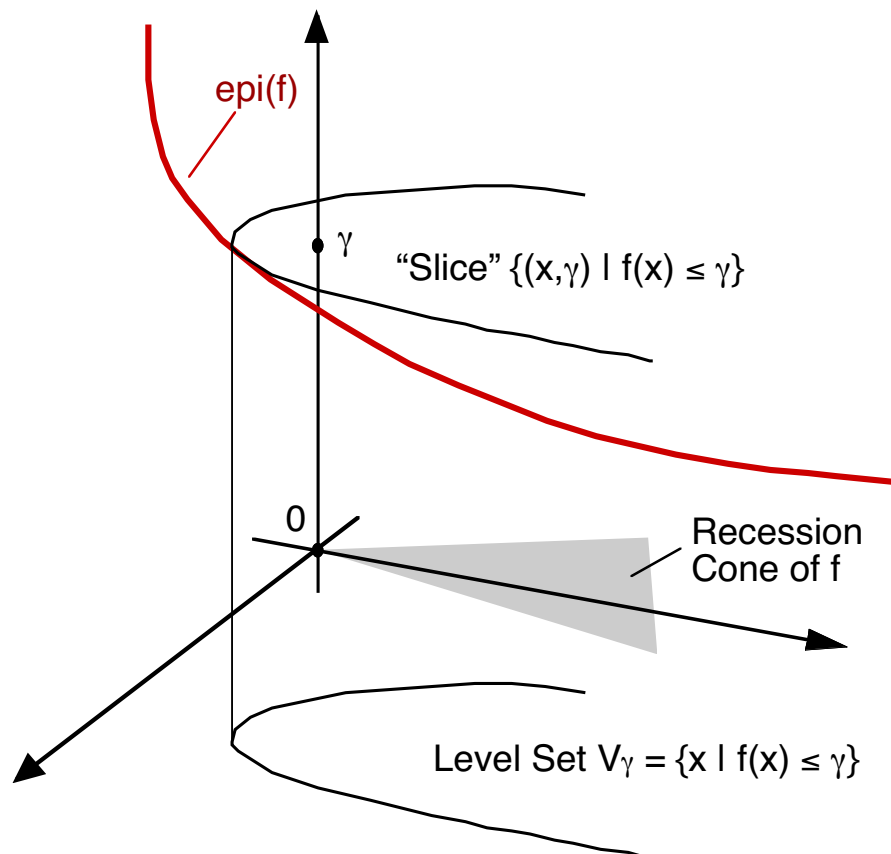
-
- For a closed convex set C , recall that y is a *direction of recession* if $x + \alpha y \in C$, for all $x \in C$ and $\alpha \geq 0$.



- 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 $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ 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

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

(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,

$$\{(x, \gamma) \mid x \in V_\gamma\} = \text{epi}(f) \cap \{(x, \gamma) \mid x \in \mathbb{R}^n\}$$

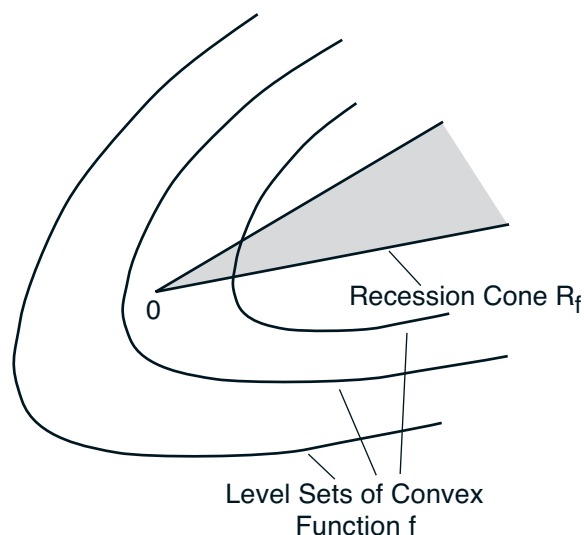
The recession cone of the set on the left is $\{(y, 0) \mid y \in R_{V_\gamma}\}$. The recession cone of the set on the right is the intersection of $R_{\text{epi}(f)}$ and the recession cone of $\{(x, \gamma) \mid x \in \mathbb{R}^n\}$. Thus we have

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

from which the result follows.

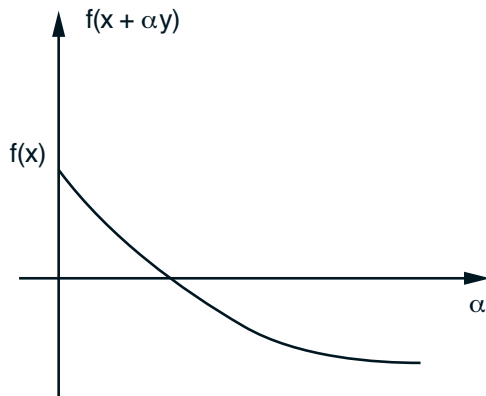
RECESSION CONE OF A CONVEX FUNCTION

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

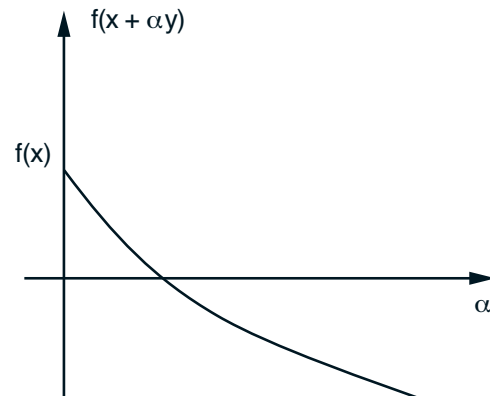


- Terminology:
 - $y \in R_f$: a *direction of recession* of f .
 - $L_f = R_f \cap (-R_f)$: the *lineality space* of f .
 - $y \in L_f$: a *direction of constancy* of f .
 - Function $r_f : \mathbb{R}^n \mapsto (-\infty, \infty]$ whose epigraph is $R_{\text{epi}(f)}$: the *recession function* of f .
- Note: $r_f(y)$ is the “asymptotic slope” of f in the direction y . In fact, $r_f(y) = \lim_{\alpha \rightarrow \infty} \nabla f(x + \alpha y)'y$ if f is differentiable. Also, $y \in R_f$ iff $r_f(y) \leq 0$.

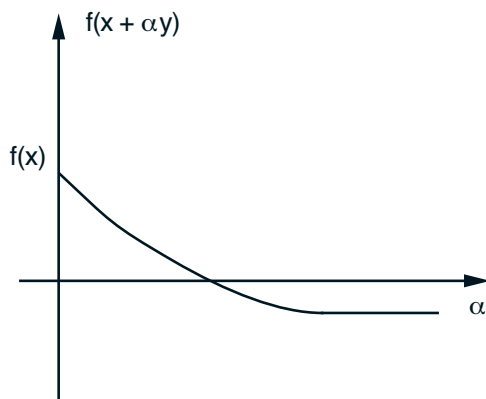
DESCENT BEHAVIOR OF A CONVEX FUNCTION



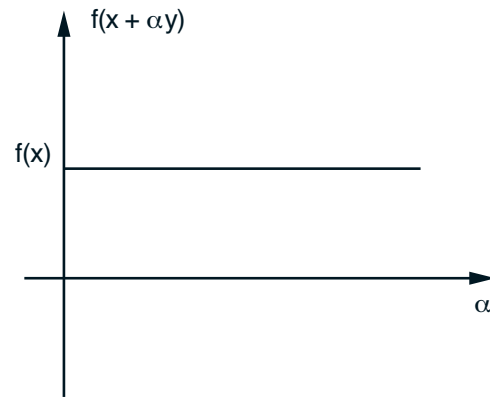
(a)



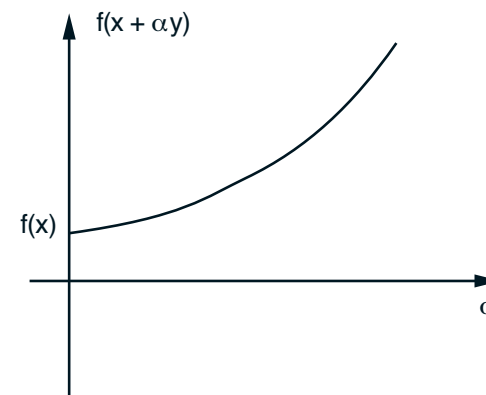
(b)



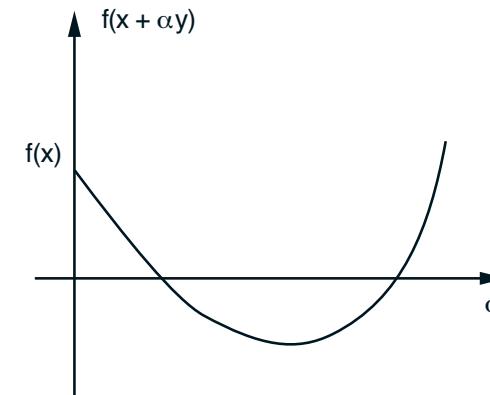
(c)



(d)



(e)



(f)

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

EXISTENCE OF SOLUTIONS - BOUNDED CASE

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

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

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

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 $f : \mathbb{R}^n \mapsto \mathbb{R}$ 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. $\|x_k\| \rightarrow \infty$, we have $\lim_{k \rightarrow \infty} f(x_k) = \infty$.

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 \mathfrak{R}^n with $S_{k+1} \subset S_k$ for all k , when is $\bigcap_{k=0}^{\infty} S_k$ nonempty?
- Set intersection theorems are significant in at least three major contexts, which we will discuss in what follows:
 1. Does a function $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ attain a minimum over a set X ? This is true iff the intersection of the nonempty level sets $\{x \in X \mid f(x) \leq \gamma_k\}$ is nonempty.
 2. If C is closed and A is a matrix, is AC 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

$$P(\text{epi}(F)) \subset \text{epi}(f) \subset \text{cl}\left(P(\text{epi}(F))\right)$$

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

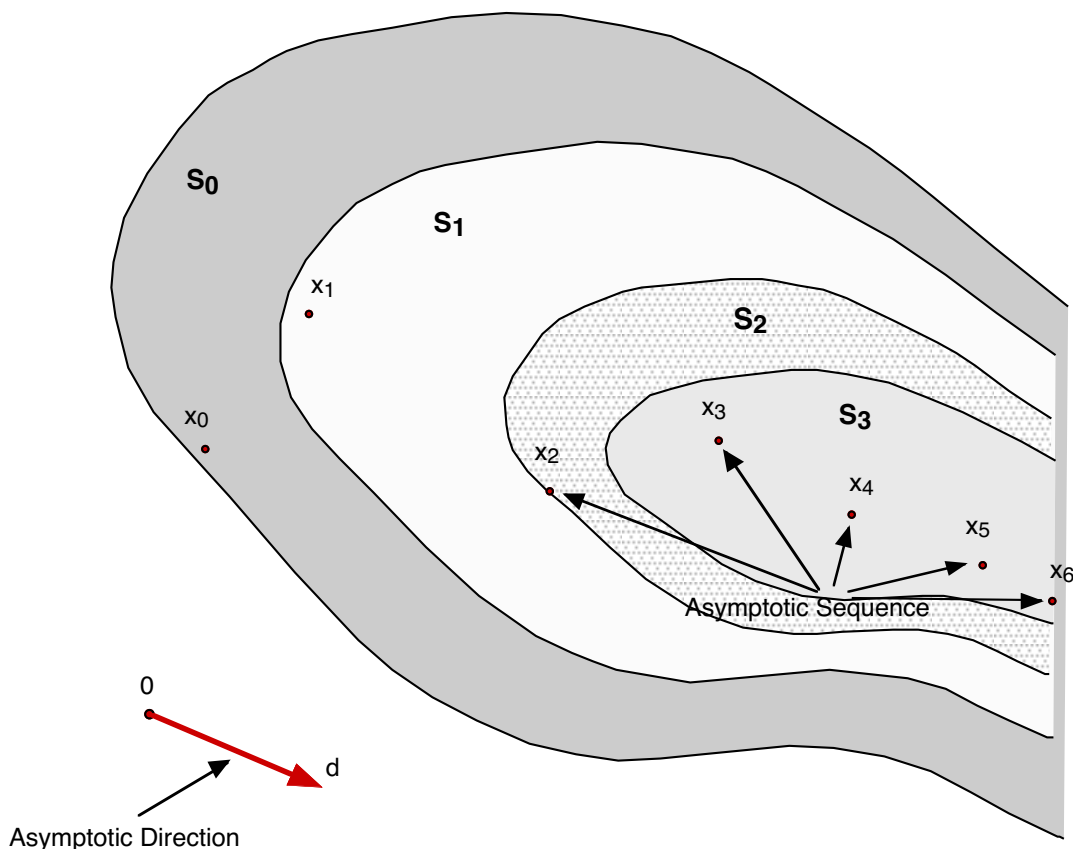
ASYMPTOTIC DIRECTIONS

- 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.

$$x_k \in S_k, \quad x_k \neq 0, \quad k = 0, 1, \dots$$

$$\|x_k\| \rightarrow \infty, \quad \frac{x_k}{\|x_k\|} \rightarrow \frac{d}{\|d\|}$$

- 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

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

- 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

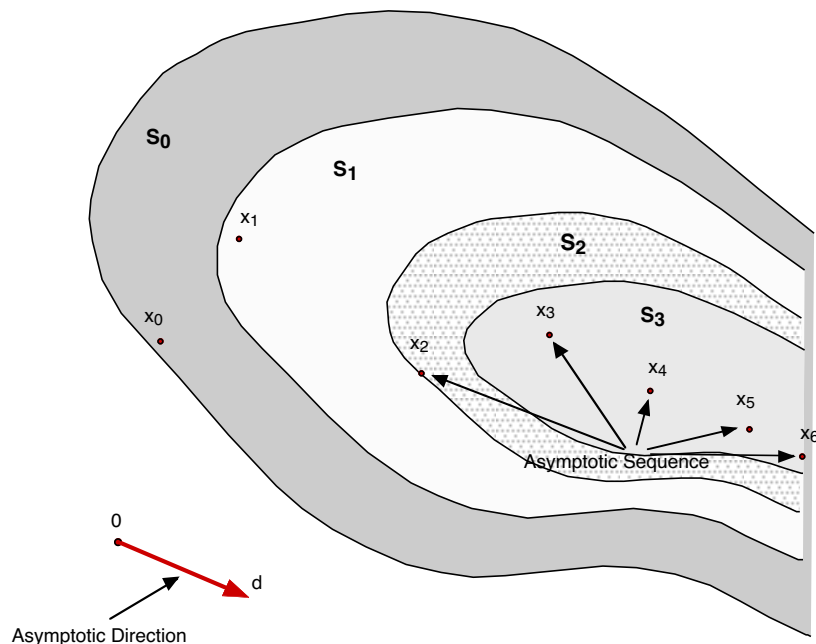
$$\bigcap_{k=0}^{\infty} R_{S_k} \setminus \{0\}$$

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

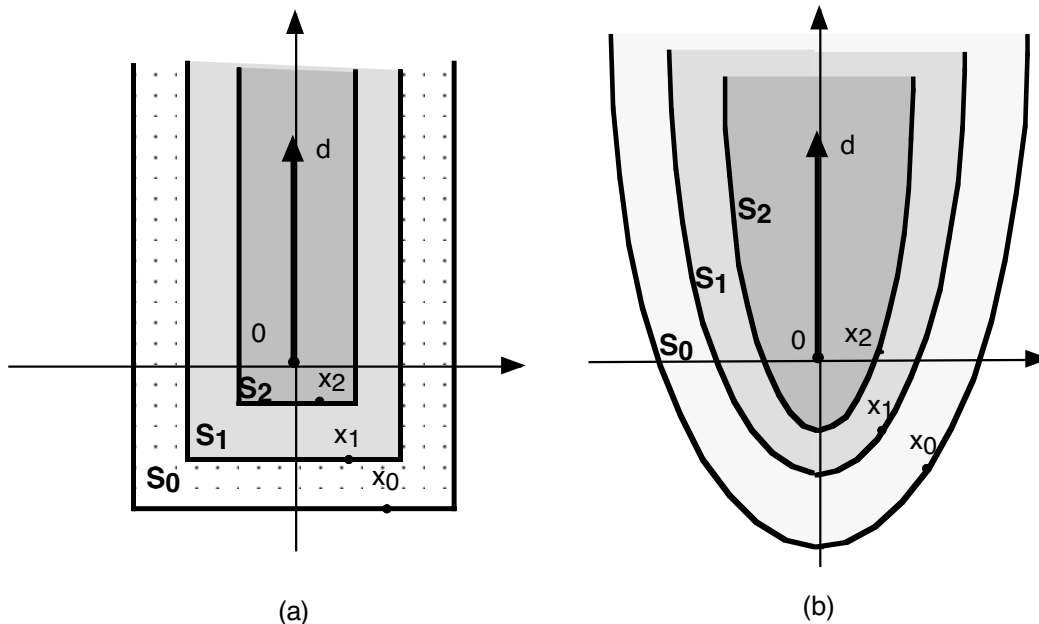


RETRACTIVE ASYMPTOTIC DIRECTIONS

- 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 \bar{k} such that

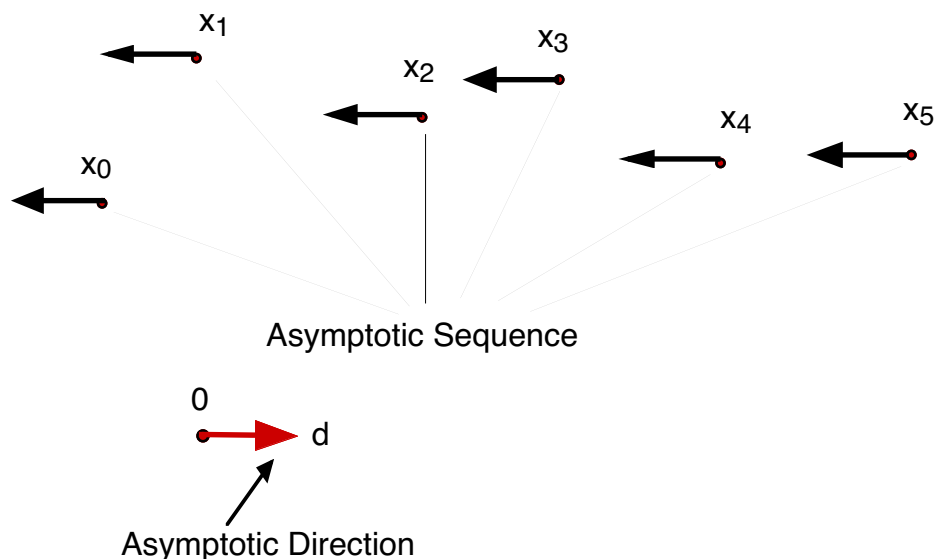
$$x_k - d \in S_k, \quad \forall k \geq \bar{k}.$$

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



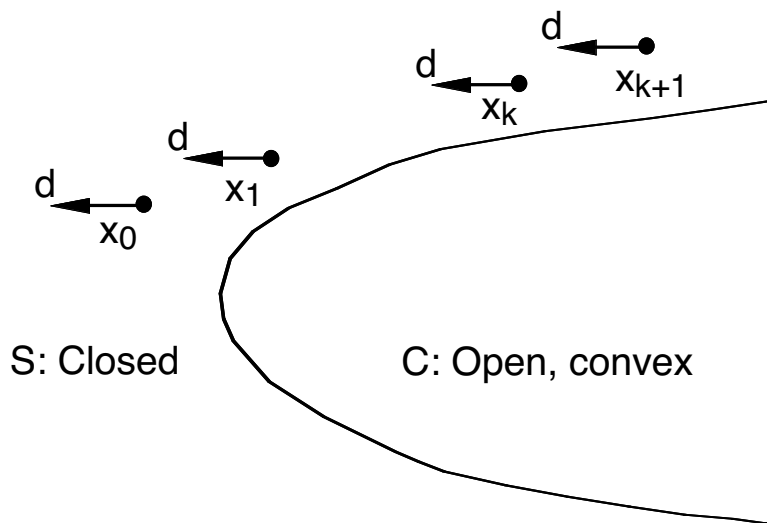
SET INTERSECTION THEOREM

- If $\{S_k\}$ is retractive, then $\bigcap_{k=0}^{\infty} S_k$ is nonempty.
- Key proof ideas:
 - (a) The intersection $\bigcap_{k=0}^{\infty} S_k$ 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 products 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 $\{x \mid f_j(x) \geq 0, j = 1, \dots, r\}$, where $f_j : \mathbb{R}^n \mapsto \mathbb{R}$ 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

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

where Q is symmetric (not necessarily positive semidefinite). If the minimal value of f over X is finite, there exists a minimum of f 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 $\bigcap_{k=0}^{\infty} S_k$, 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

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

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

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

- Thus,

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

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

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

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

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

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 $\bigcap_{k=0}^{\infty} S_k$ is nonempty.

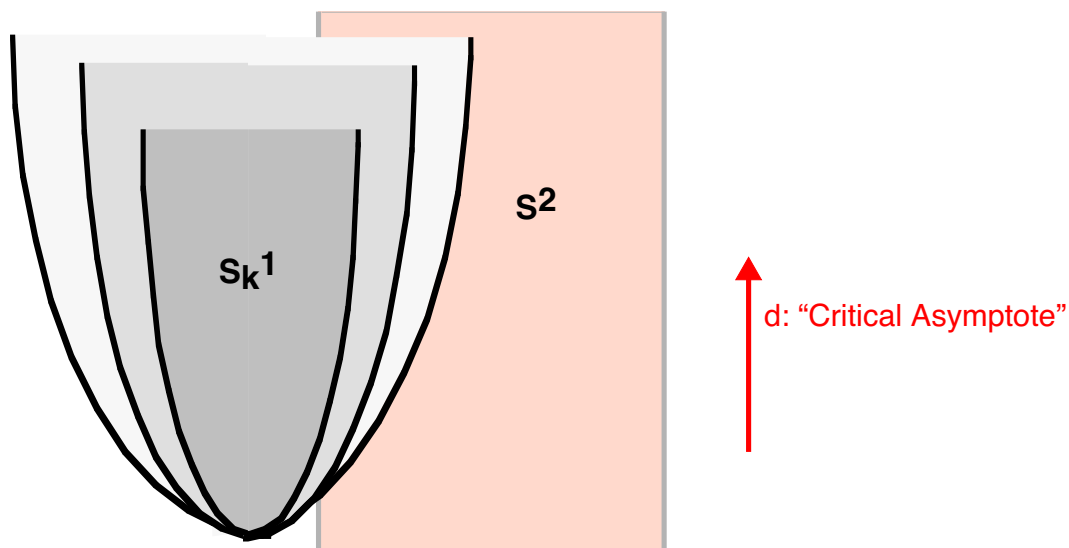
CRITICAL ASYMPTOTES

- 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

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

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

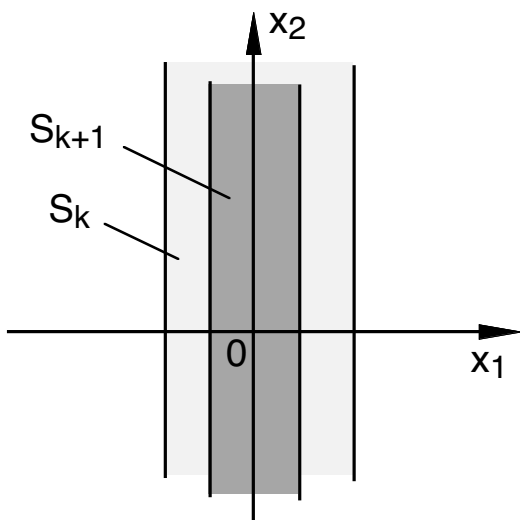
• Consider $\{S_k\}$ with $\bigcap_{k=0}^{\infty} S_k \neq \emptyset$. An asymptotic direction d of $\{S_k\}$ is:

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

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

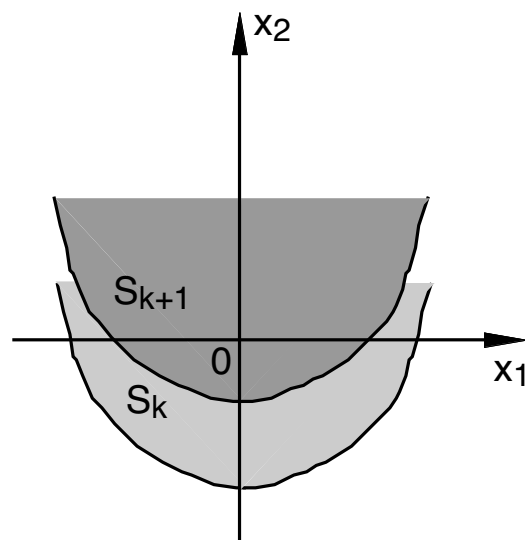
• **Example: (2-D Convex Quadratic Set Sequences)**

$$S_k = \{(x_1, x_2) \mid x_1^2 \leq 1/k\}$$



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

$$S_k = \{(x_1, x_2) \mid x_1^2 - x_2 \leq 1/k\}$$



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

GENERAL CONVEX QUADRATIC SETS

- Let $S_k = \{x \mid x'Qx + a'x + b \leq \gamma_k\}$, where $\gamma_k \downarrow 0$. Then, if all the sets S_k are nonempty, $\bigcap_{k=0}^{\infty} S_k \neq \emptyset$.
- Asymptotic directions: $d \neq 0$ such that $Qd = 0$ and $a'd \leq 0$. 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 $S_k = S_k^1 \cap S_k^2 \cap \dots \cap S_k^r$, where $\{S_k^j\}$ are nested sequence such that

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

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

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

$$S_k = \{x \mid x'Q_jx + a'_jx + b_j \leq \gamma_k^j, \quad j = 1, \dots, r\}$$

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, $\bigcap_{k=0}^{\infty} S_k \neq \emptyset$.

APPLICATION TO QUADRATIC MINIMIZATION

- Let

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

$$X = \{x \mid x'R_jx + a'_jx + b_j \leq 0, j = 1, \dots, r\},$$

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 over X .

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

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

All the set sequences involved in the intersection are convex quadratic and hence have no critical directions. By the preceding proposition, X^* is nonempty. **Q.E.D.**

CLOSURE UNDER LINEAR TRANSFORMATIONS

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

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

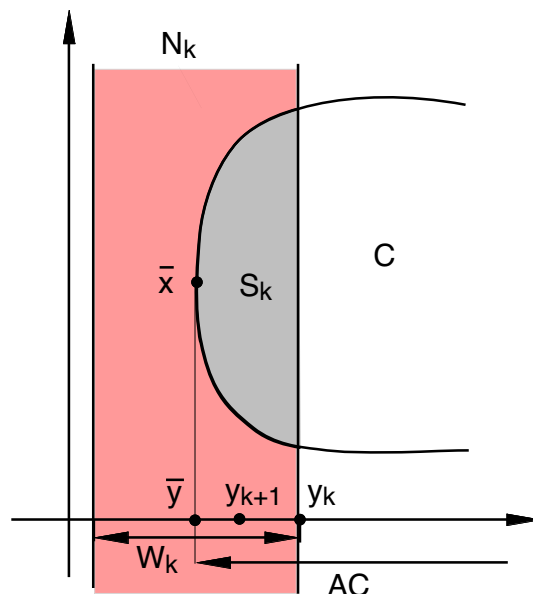
(b) $A(X \cap C)$ is closed if X is a polyhedral set and

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

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

Proof: (Outline) Let $\{y_k\} \subset AC$ with $y_k \rightarrow \bar{y}$. We prove $\bigcap_{k=0}^{\infty} S_k \neq \emptyset$, where $S_k = C \cap N_k$, and

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



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 \mathfrak{R}^n .
 - (a) For every $x \in \mathfrak{R}^n$, 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 $x \in \mathfrak{R}^n$, a vector $z \in C$ is equal to $P_C(x)$ if and only if

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

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 $f : \mathfrak{R}^n \mapsto C$ defined by $f(x) = P_C(x)$ is continuous and nonexpansive, i.e.,

$$\|P_C(x) - P_C(y)\| \leq \|x - y\|, \quad \forall x, y \in \mathfrak{R}^n$$

EXISTENCE OF OPTIMAL SOLUTIONS

• Let X and $f : \Re^n \mapsto (-\infty, \infty]$ be closed convex and such that $X \cap \text{dom}(f) \neq \emptyset$. 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:

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

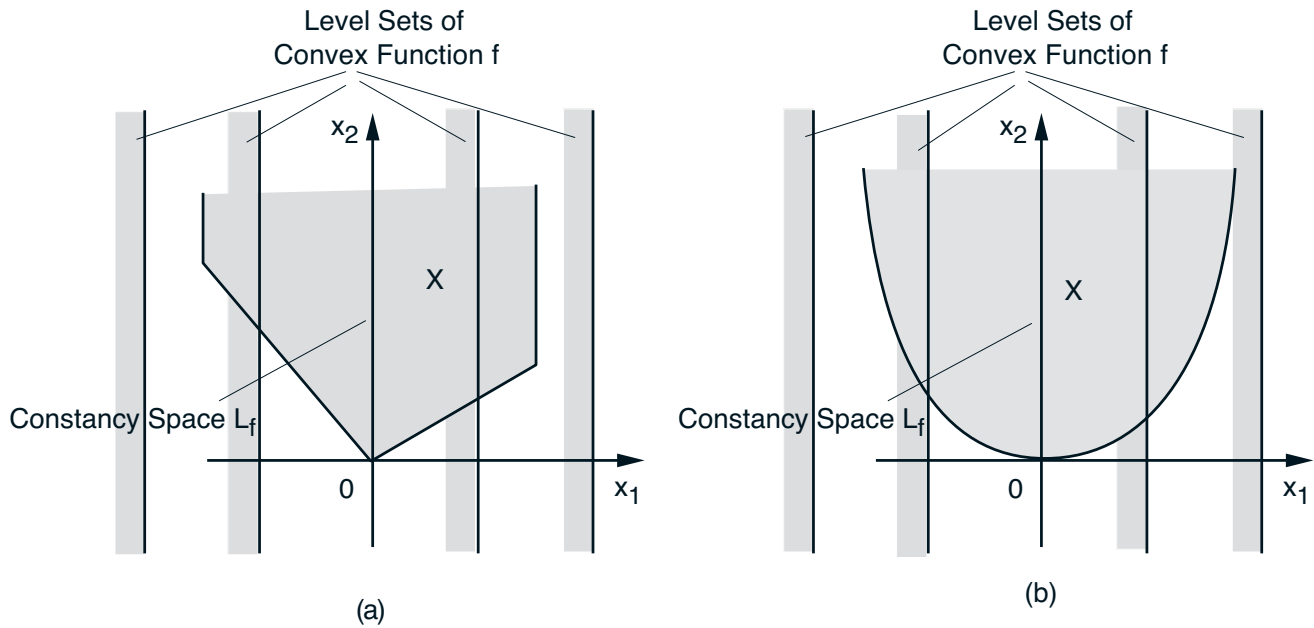
$$X = \{x \mid x'Q_jx + a'_jx + b_j \leq 0, j = 1, \dots, r\}.$$

Proof: Follows by writing

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

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),

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

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

PARTIAL MINIMIZATION THEOREM

- Let $F : \mathfrak{R}^{n+m} \mapsto (-\infty, \infty]$ be a closed proper convex function, and consider $f(x) = \inf_{z \in \mathfrak{R}^m} F(x, z)$.
- 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 $\bar{x} \in \mathfrak{R}^n$, $\bar{\gamma} \in \mathfrak{R}$ such that the set

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

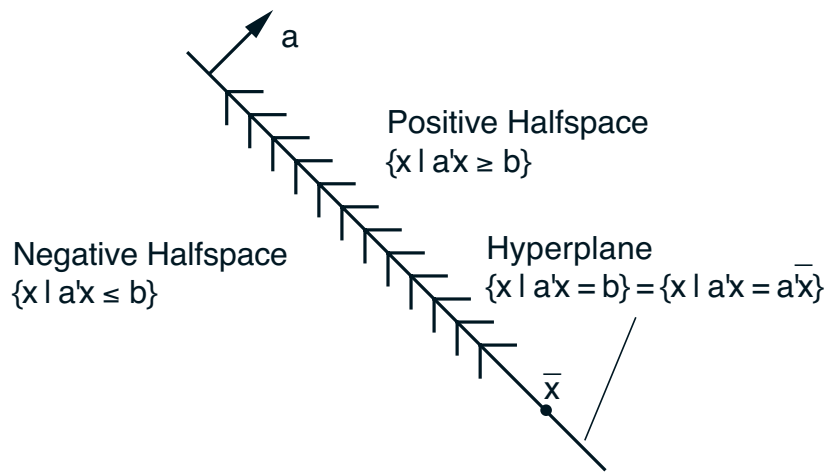
is nonempty and compact, then f is convex, closed, and proper. Also, for each $x \in \text{dom}(f)$, the set of minima of $F(x, \cdot)$ is nonempty and compact.

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

$$\{z \mid F(x, z) \leq \gamma\}, \quad x \in \mathfrak{R}^n, \quad \gamma \in \mathfrak{R},$$

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

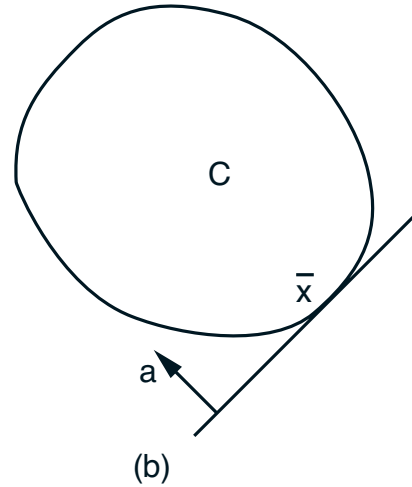
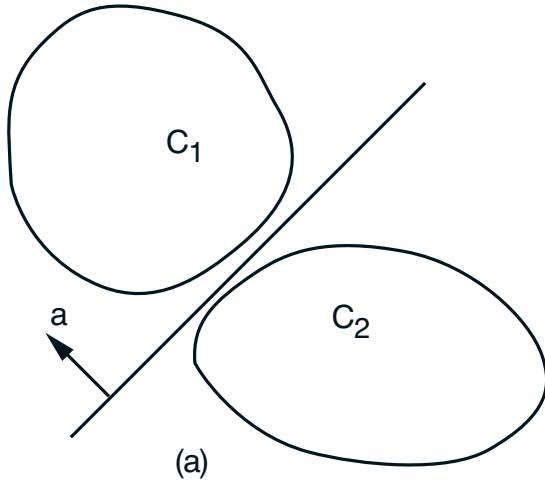
HYPERPLANES



- A *hyperplane* is a set of the form $\{x \mid a'x = b\}$, where a is nonzero vector in \mathbb{R}^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.,
 either $a'x_1 \leq b \leq a'x_2, \quad \forall x_1 \in C_1, \forall x_2 \in C_2,$
 or $a'x_2 \leq b \leq a'x_1, \quad \forall x_1 \in C_1, \forall x_2 \in C_2$
- If \bar{x} belongs to the closure of a set C , a hyperplane that separates C and the singleton set $\{\bar{x}\}$ is said be *supporting* C at \bar{x} .

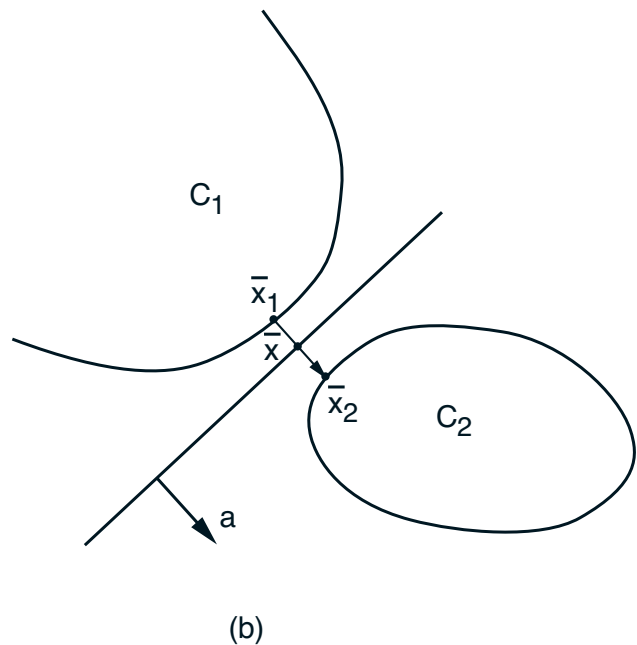
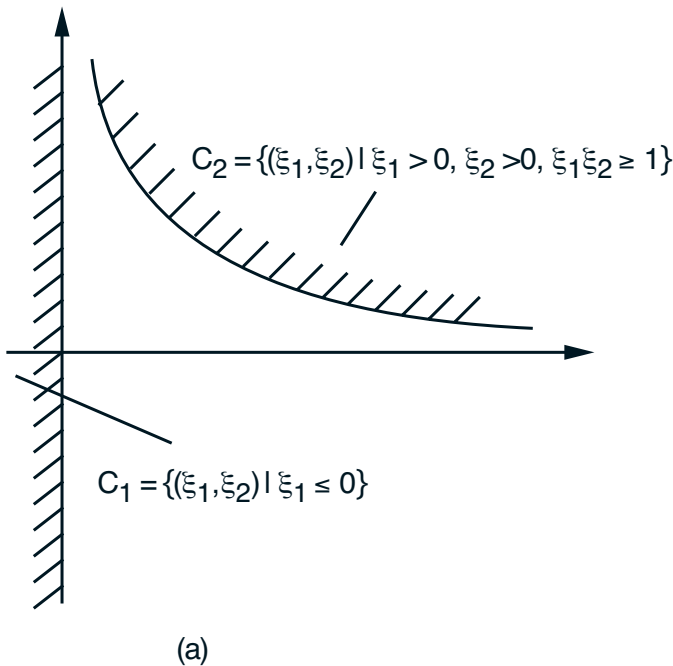
VISUALIZATION

- Separating and supporting hyperplanes:



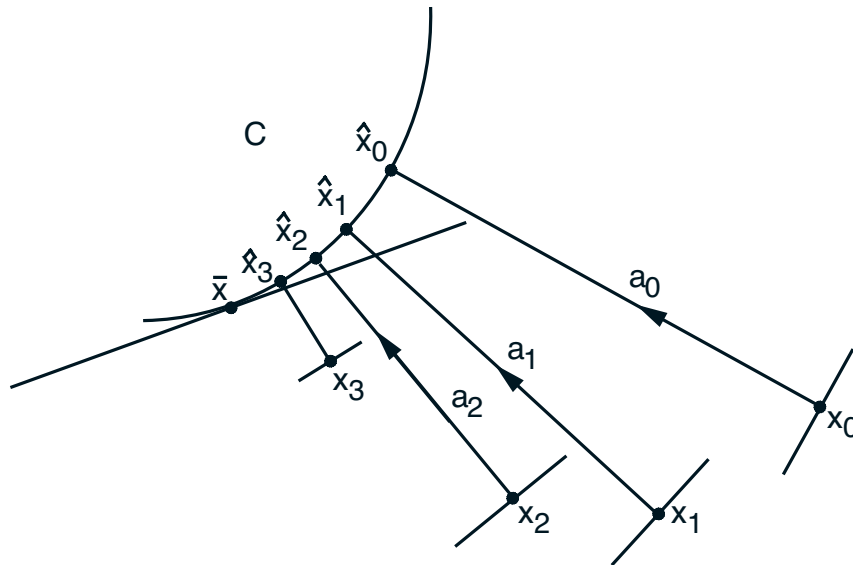
- A separating $\{x \mid a'x = b\}$ that is disjoint from C_1 and C_2 is called *strictly separating*:

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



SUPPORTING HYPERPLANE THEOREM

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



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

$$a'_k x \geq a'_k x_k, \quad \forall x \in \text{cl}(C), \quad \forall k = 0, 1, \dots,$$

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

SEPARATING HYPERPLANE THEOREM

• Let C_1 and C_2 be two nonempty convex subsets of \mathbb{R}^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

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

Proof: Consider the convex set

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

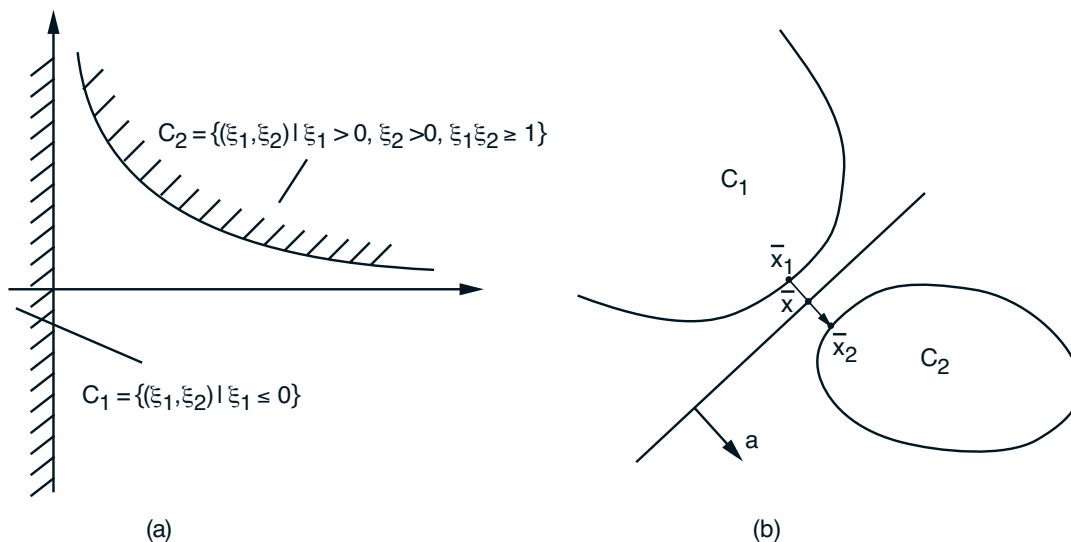
Since C_1 and C_2 are disjoint, the origin does not belong to $C_1 - C_2$, so by the Supporting Hyperplane Theorem, there exists a vector $a \neq 0$ such that

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

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

STRICT SEPARATION THEOREM

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

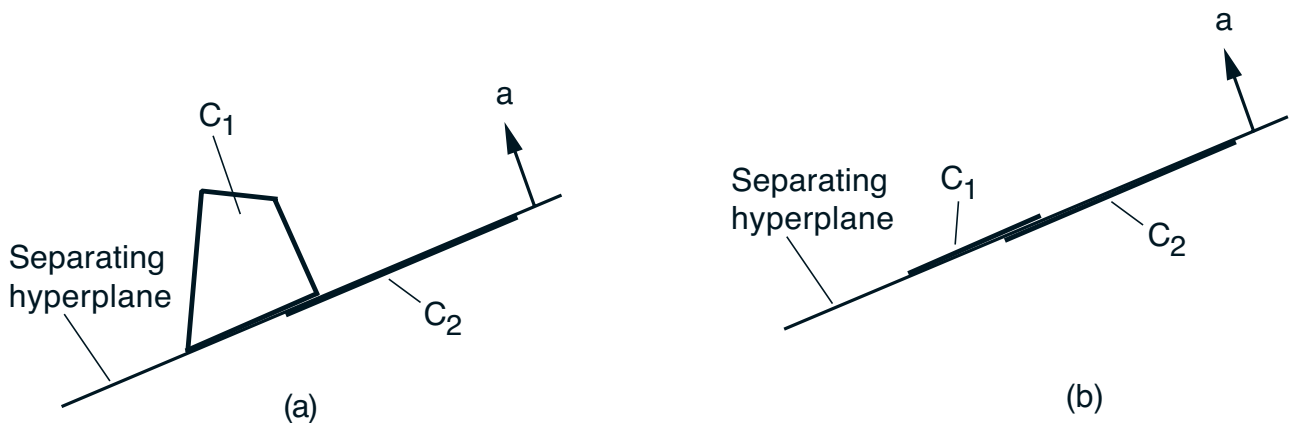


Proof: (Outline) Consider the set $C_1 - C_2$. Since C_1 is closed and C_2 is compact, $C_1 - C_2$ is closed. Since $C_1 \cap C_2 = \emptyset$, $0 \notin C_1 - C_2$. Let $\bar{x}_1 - \bar{x}_2$ be the projection of 0 onto $C_1 - C_2$. The strictly separating hyperplane is constructed as in (b).

- **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 $C \subset \mathbb{R}^n$ 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 .

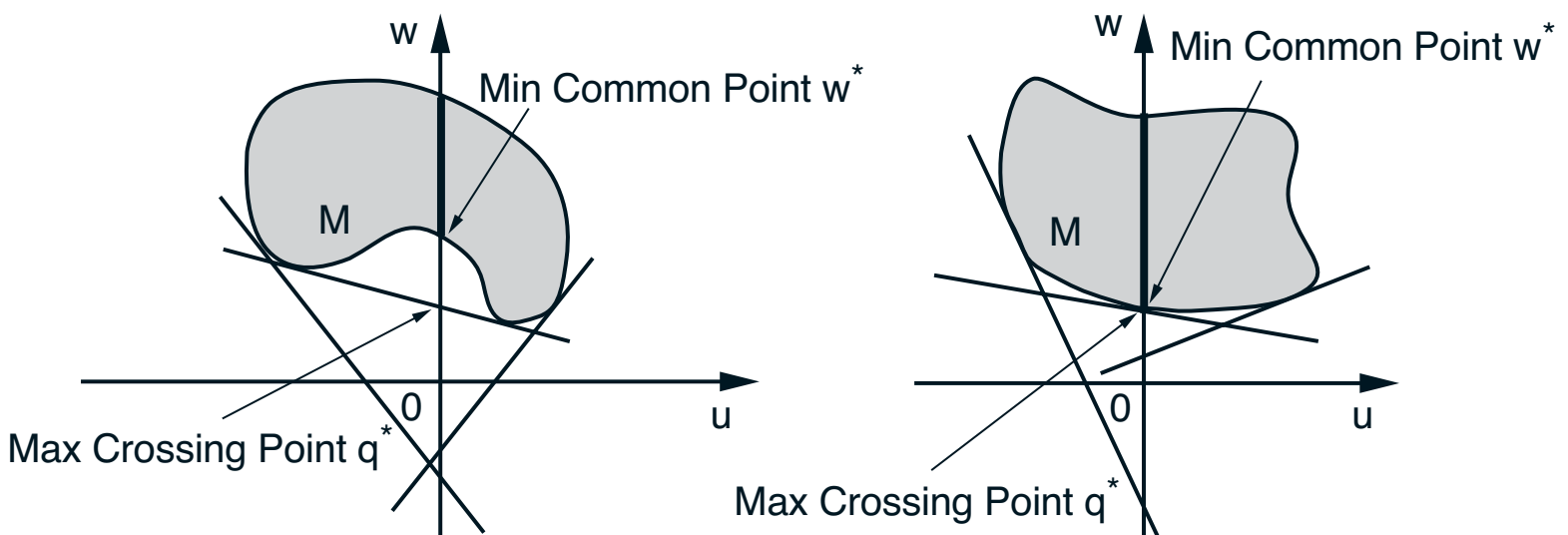


- **Proper Separation Theorem:** Let C_1 and C_2 be two nonempty convex subsets of \mathbb{R}^n . There exists a hyperplane that properly separates C_1 and C_2 if and only if

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

MIN COMMON / MAX CROSSING PROBLEMS

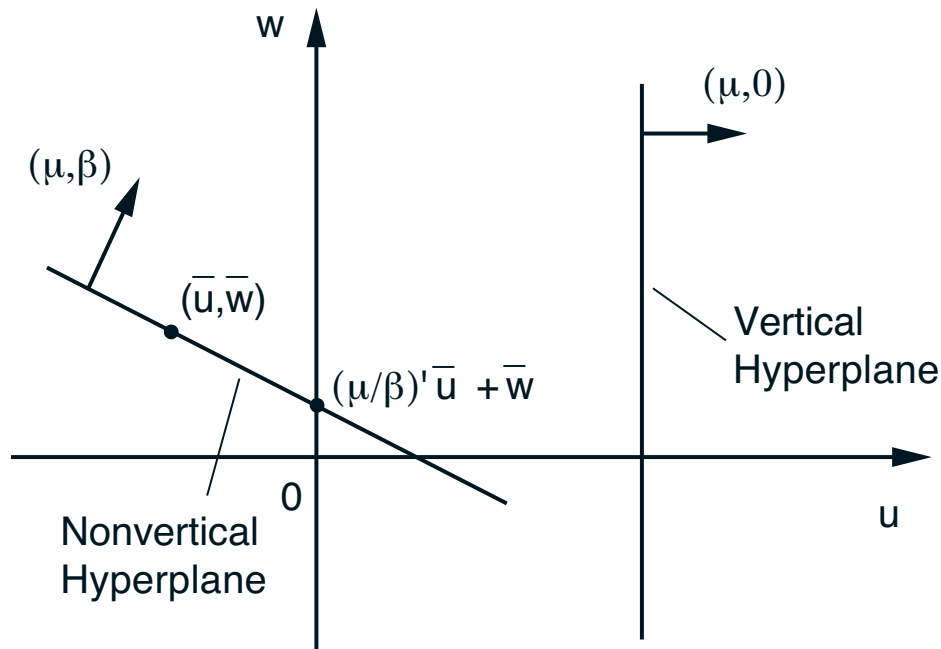
- We introduce a pair of fundamental problems:
- Let M be a nonempty subset of \mathbb{R}^{n+1}
 - (a) *Min Common Point Problem*: Consider all vectors that are common to M and the $(n + 1)$ st axis. Find one whose $(n + 1)$ st component is minimum.
 - (b) *Max Crossing Point Problem*: Consider “non-vertical”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



- We first need to study “nonvertical” hyperplanes.

NONVERTICAL HYPERPLANES

- A hyperplane in \Re^{n+1} with normal (μ, β) is non-vertical if $\beta \neq 0$.
- It intersects the $(n+1)$ st axis at $\xi = (\mu/\beta)' \bar{u} + \bar{w}$, where (\bar{u}, \bar{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 \mathbb{R}^{n+1} that contains no vertical lines. Then:
 - (a) C is contained in a closed halfspace of a nonvertical hyperplane, i.e., there exist $\mu \in \mathbb{R}^n$, $\beta \in \mathbb{R}$ with $\beta \neq 0$, and $\gamma \in \mathbb{R}$ such that $\mu'u + \beta w \geq \gamma$ for all $(u, w) \in C$.
 - (b) If $(\bar{u}, \bar{w}) \notin \text{cl}(C)$, there exists a nonvertical hyperplane strictly separating (\bar{u}, \bar{w}) and C .

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

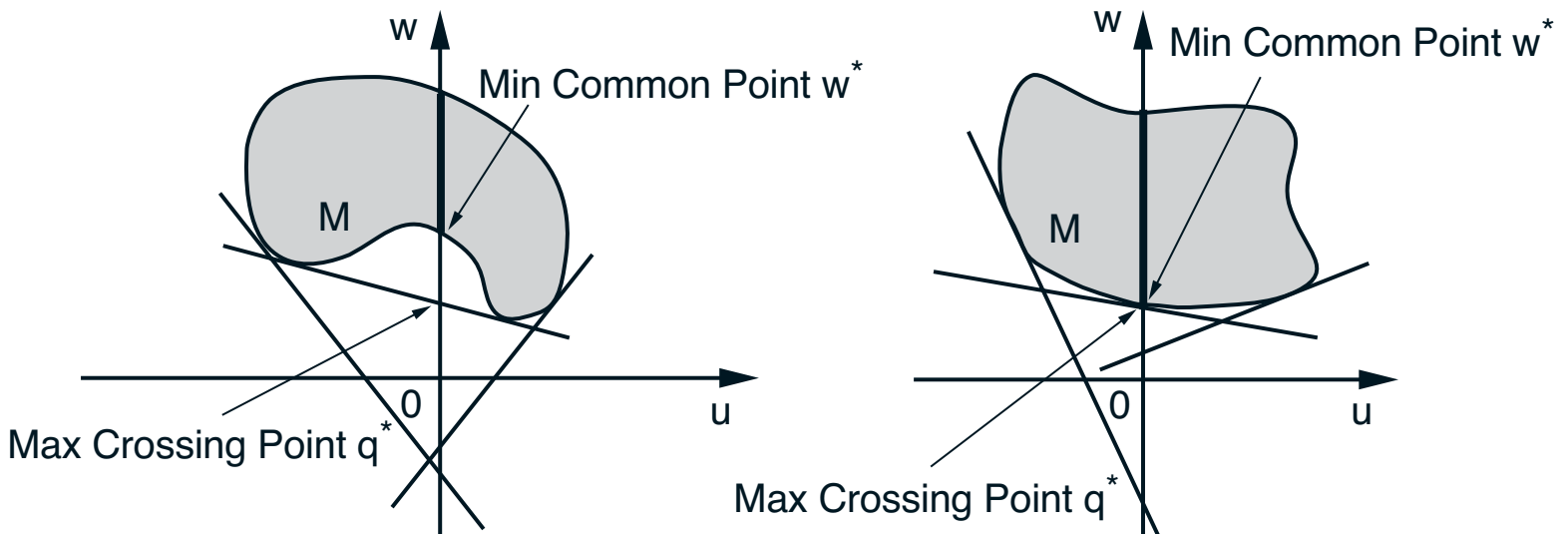
(a) C is the intersection of the closed halfspaces containing C . If all these corresponded to vertical hyperplanes, C would contain a vertical line.

(b) There is a hyperplane strictly separating (\bar{u}, \bar{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

$$\xi \leq w + \mu'u, \quad \forall (u, w) \in M$$

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

$$\text{maximize } q(\mu) \triangleq \inf_{(u,w) \in M} \{w + \mu'u\}$$

subject to $\mu \in \mathbb{R}^n$.

- For all $(u, w) \in M$ and $\mu \in \mathbb{R}^n$,

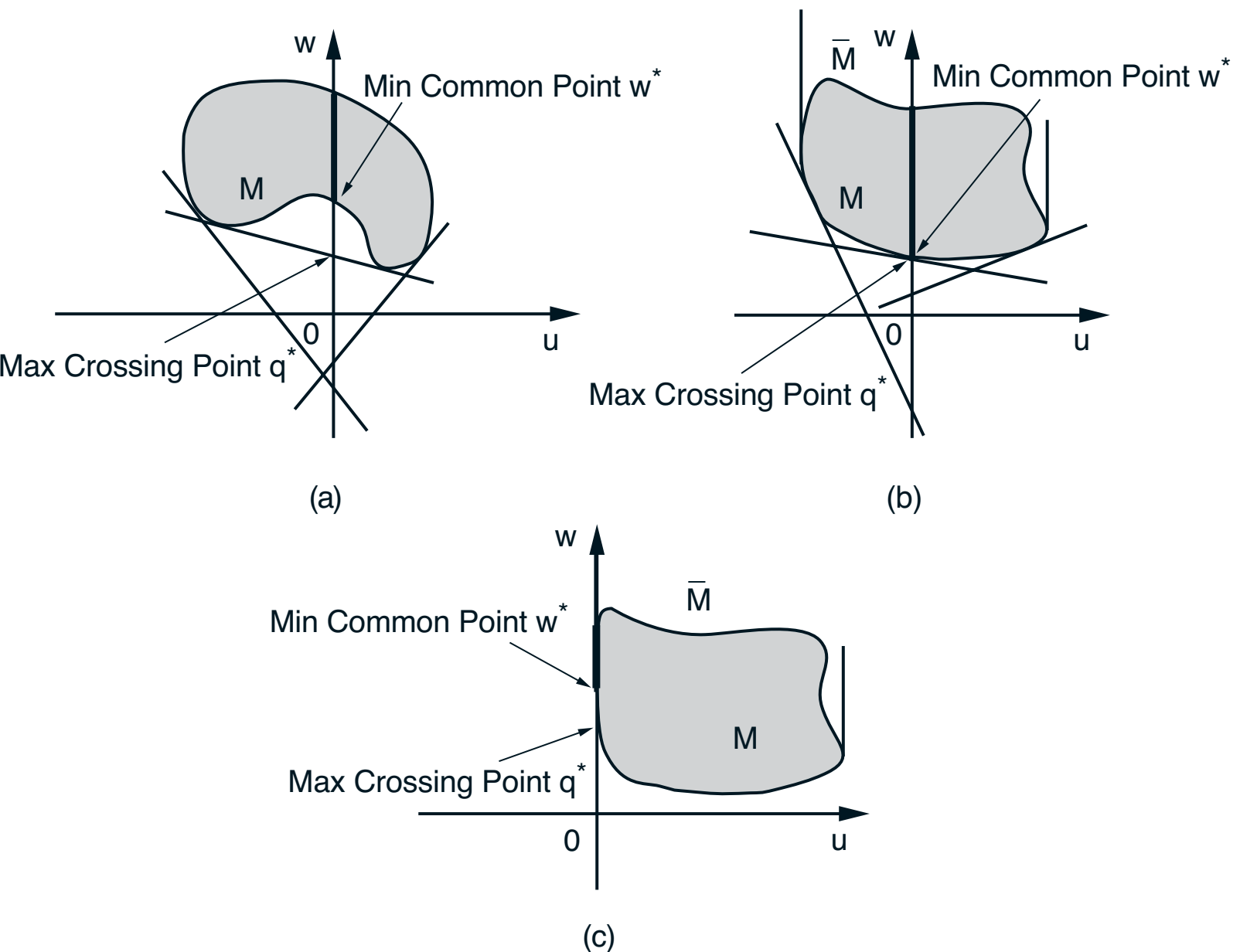
$$q(\mu) = \inf_{(u,w) \in M} \{w + \mu'u\} \leq \inf_{(0,w) \in M} w = w^*,$$

so maximizing over $\mu \in \mathbb{R}^n$, we obtain $q^* \leq w^*$.

- 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

$$\overline{M} = \left\{ (u, w) \mid \text{there exists } \overline{w} \text{ with } \overline{w} \leq w \text{ and } (u, \overline{w}) \in M \right\}$$

is convex.

- *Min Common/Max Crossing Theorem I:* We have $q^* = w^*$ if and only if for every sequence $\{(u_k, w_k)\} \subset M$ with $u_k \rightarrow 0$, there holds $w^* \leq \liminf_{k \rightarrow \infty} w_k$.

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

$$D = \left\{ u \mid \text{there exists } w \in \mathfrak{R} \text{ with } (u, w) \in \overline{M} \right\}$$

contains the origin in its relative interior. Then $q^* = w^*$ and there exists a vector $\mu \in \mathfrak{R}^n$ such that $q(\mu) = q^*$. If D contains the origin in its interior, the set of all $\mu \in \mathfrak{R}^n$ 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 $u_k \rightarrow 0$, there holds $w^* \leq \liminf_{k \rightarrow \infty} w_k$. 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) $(0, w^* - \epsilon) \notin \text{cl}(\overline{M})$ 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 $u_k \rightarrow 0$. Then,

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

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

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

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

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

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

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

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

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

MINIMAX PROBLEMS

Given $\phi : X \times Z \mapsto \mathbb{R}$, where $X \subset \mathbb{R}^n$, $Z \subset \mathbb{R}^m$
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 $x \in X$

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

- 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 $x \in X$, $g_j(x) \leq 0$, $j = 1, \dots, r$

introduce the Lagrangian function

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

- Primal problem (equivalent to the original)

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

- Dual problem

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

- Key duality question: Is it true that

$$\sup_{\mu \geq 0} \inf_{x \in \mathbb{R}^n} L(x, \mu) = \inf_{x \in \mathbb{R}^n} \sup_{\mu \geq 0} L(x, \mu)$$

ZERO SUM GAMES

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

$$x = (x_1, \dots, x_n), \quad z = (z_1, \dots, z_m)$$

over their possible moves.

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

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

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

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

MINIMAX INEQUALITY

- We always have

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

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

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

and take the sup over $\bar{z} \in Z$ 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 $\phi : X \times Z \mapsto \mathbb{R}$, where $X \subset \mathbb{R}^n$, $Z \subset \mathbb{R}^m$
consider

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

$$\text{subject to } x \in X$$

and

$$\text{maximize } \inf_{x \in X} \phi(x, z)$$

$$\text{subject to } z \in Z.$$

- 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

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

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

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

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

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

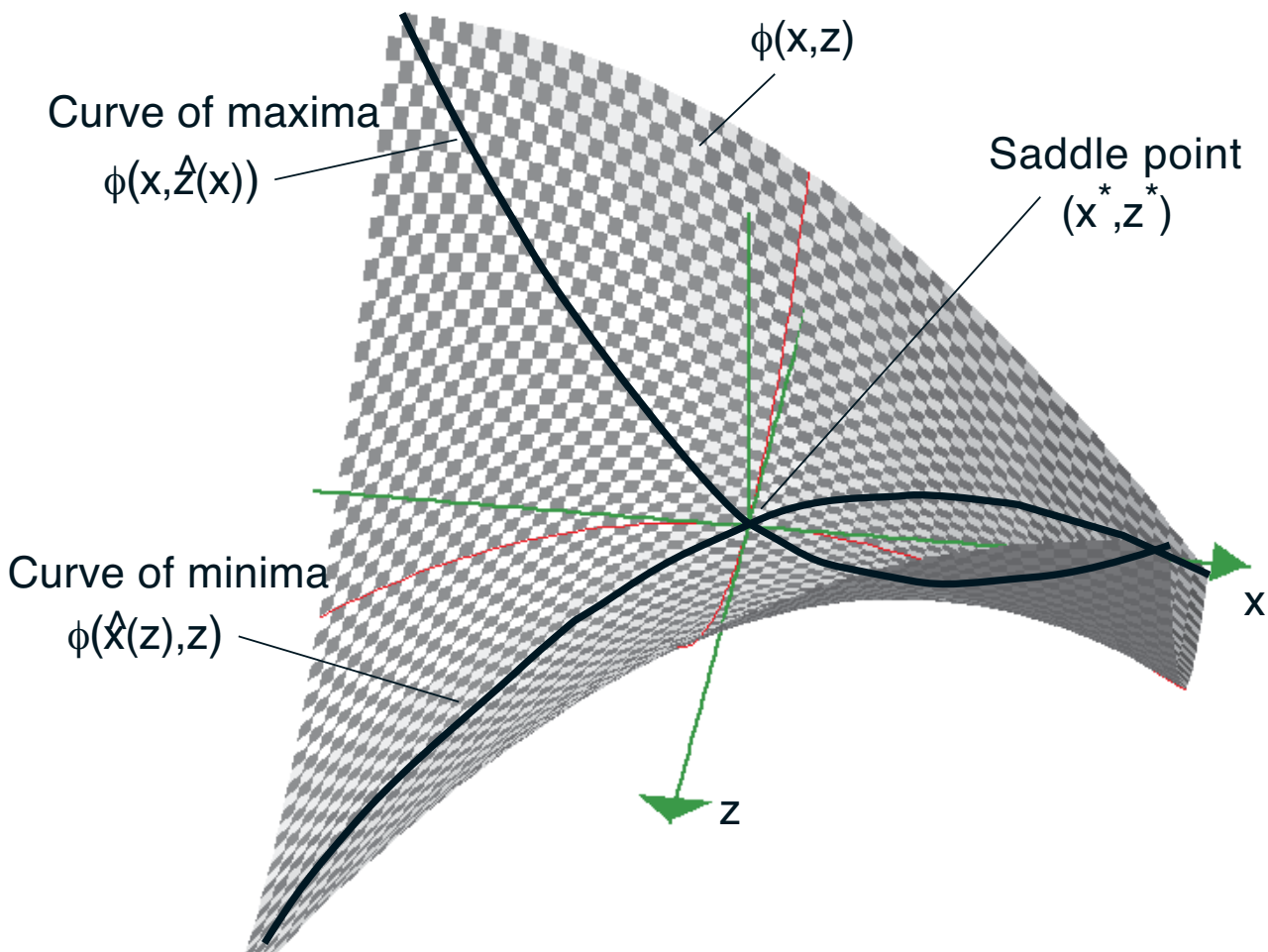
By the minimax inequality, the above holds as an equality holds throughout, so the minimax equality and Eq. (*) hold.

Conversely, if Eq. (*) holds, then

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

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

VISUALIZATION



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

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

Saddle points correspond to points where these two curves meet.

MIN COMMON/MAX CROSSING FRAMEWORK

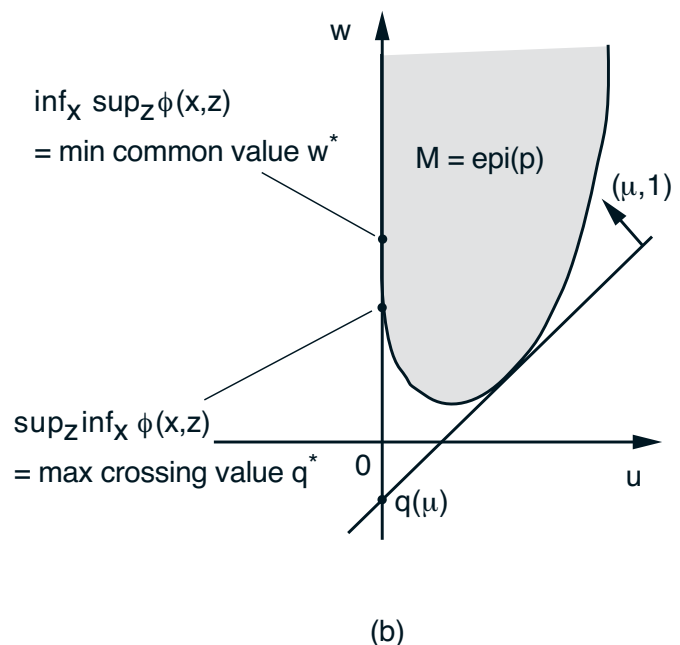
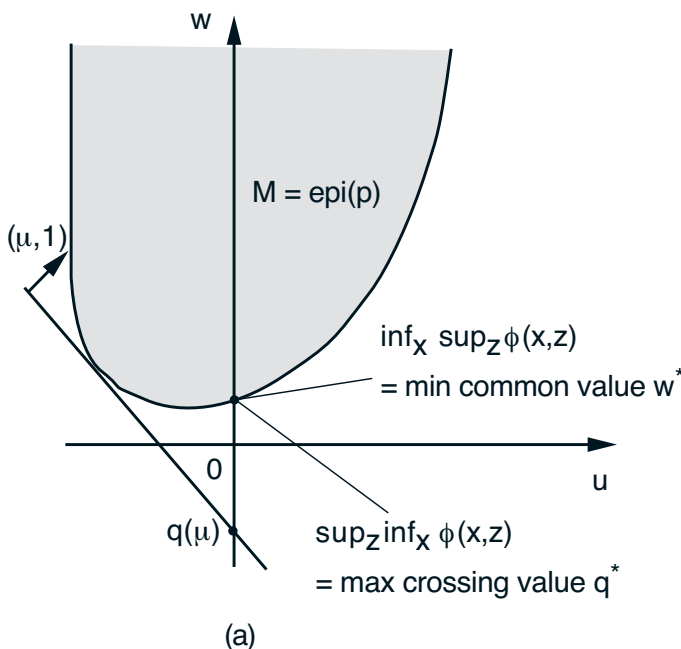
- Introduce perturbation function $p : \mathbb{R}^m \mapsto [-\infty, \infty]$

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

- 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 $\phi(\cdot, z) : X \mapsto \mathfrak{R}$ 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

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

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 $\mu \in \mathbb{R}^n$, where

$$\begin{aligned} q(\mu) &= \inf_{(u,w) \in \text{epi}(p)} \{w + \mu' u\} = \inf_{\{(u,w) | p(u) \leq w\}} \{w + \mu' u\} \\ &= \inf_{u \in \mathbb{R}^m} \{p(u) + \mu' u\} \end{aligned}$$

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

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

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

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

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

$$\begin{aligned} \sup_{z \in Z} \inf_{x \in X} \phi(x, z) &\leq \sup_{\mu \in \mathbb{R}^m} q(\mu) = q^* \\ &\leq w^* = p(0) = \inf_{x \in X} \sup_{z \in Z} \phi(x, z) \end{aligned}$$

IMPLICATIONS OF CONCAVITY IN Z

Lemma 2: Assume that for each $x \in X$, the function $r_x : \mathfrak{R}^m \mapsto (-\infty, \infty]$ 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,

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

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 $(\mu, r_x(\mu) - \epsilon)$, which does not belong to $\text{epi}(r_x)$. Since $\text{epi}(r_x)$ 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 $-\phi(x, \cdot) : Z \mapsto \mathfrak{R}$ is closed and convex.

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

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

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

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

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 $-\phi(x, \cdot) : Z \mapsto \mathfrak{R}$ is closed and convex.
- (5) 0 lies in the relative interior of $\text{dom}(p)$.

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 $\text{dom}(f)$.]

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

EXAMPLE I

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

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

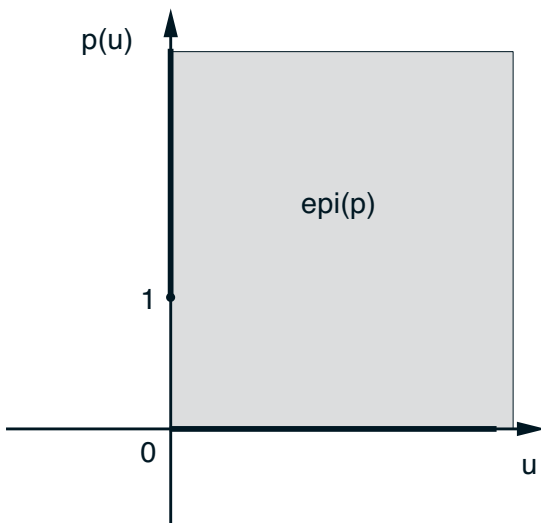
which satisfy the convexity and closedness assumptions. For all $z \geq 0$,

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

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

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

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



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

EXAMPLE II

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

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

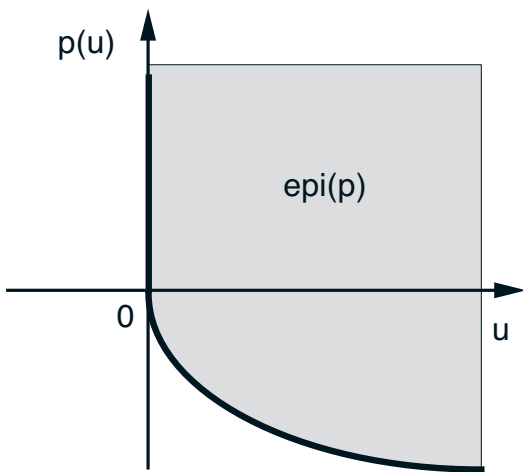
which satisfy the convexity and closedness assumptions. For all $z \geq 0$,

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

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

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

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



$$\begin{aligned} p(u) &= \inf_{x \in \mathfrak{R}} \sup_{z \geq 0} \{x + zx^2 - uz\} \\ &= \begin{cases} -\sqrt{u} & \text{if } u \geq 0, \\ \infty & \text{if } u < 0. \end{cases} \end{aligned}$$

SADDLE POINT ANALYSIS

- The preceding analysis has underscored the importance of the perturbation function

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

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

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

- 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 $\bar{z} \in Z$ and a scalar γ such that the level set $\{x \in X \mid \phi(x, \bar{z}) \leq \gamma\}$ is nonempty and compact.
- (3) X is compact and there exists a vector $\bar{x} \in X$ and a scalar γ such that the level set $\{z \in Z \mid \phi(\bar{x}, z) \geq \gamma\}$ is nonempty and compact.
- (4) There exist vectors $\bar{x} \in X$ and $\bar{z} \in Z$, and a scalar γ such that the level sets
$$\{x \in X \mid \phi(x, \bar{z}) \leq \gamma\}, \quad \{z \in Z \mid \phi(\bar{x}, z) \geq \gamma\},$$
are nonempty and compact.

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

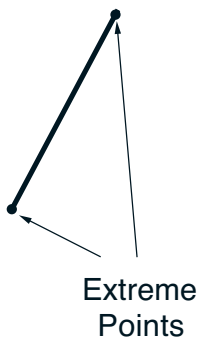
LECTURE 10

LECTURE OUTLINE

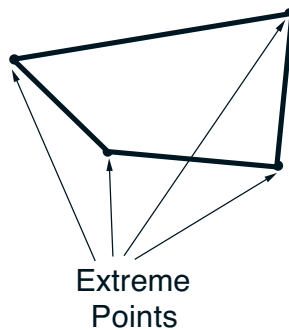
- Extreme points
 - Polar cones and polar cone theorem
 - Polyhedral and finitely generated cones
 - Farkas Lemma, Minkowski-Weyl Theorem
-
- The main convexity concepts so far have been:
 - Closure, convex hull, affine hull, rel. interior
 - Directions of recession and set intersection theorems
 - Preservation of closure under linear transformation and partial minimization
 - Existence of optimal solutions
 - Hyperplanes, min common/max crossing duality, and application in minimax
 - We now introduce new concepts with important theoretical and algorithmic implications: extreme points, polyhedral convexity, and related issues.

EXTREME POINTS

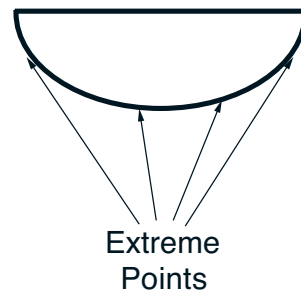
- A vector x is an *extreme point* of a convex set C if $x \in C$ and x does not lie strictly within a line segment contained in C .



(a)

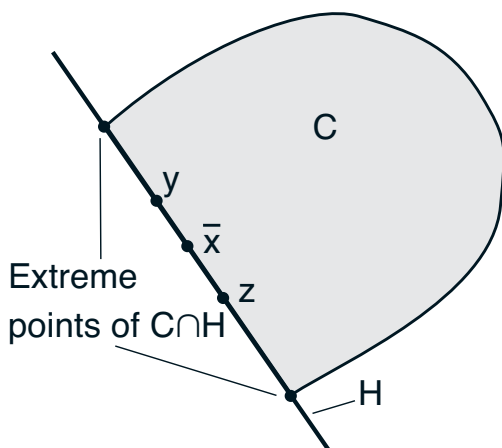


(b)



(c)

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



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

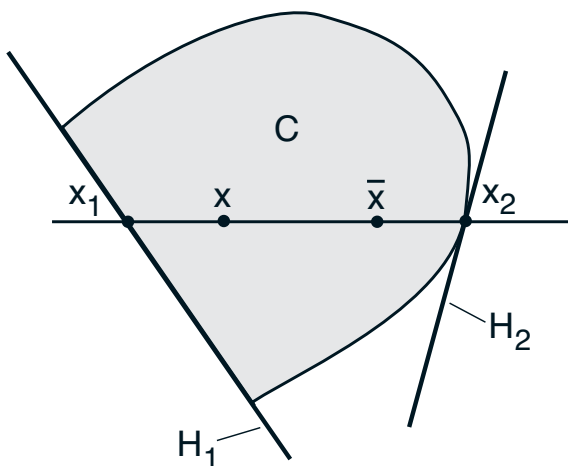
PROPERTIES OF EXTREME POINTS I

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

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

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

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



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

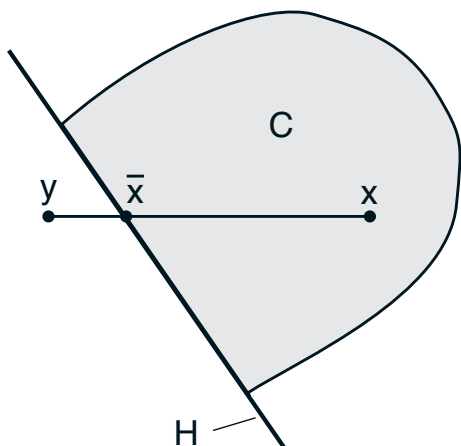
PROPERTIES OF EXTREME POINTS II

Proposition: A closed 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 (use the Recession Cone Theorem) - impossible.

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

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



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

CHARACTERIZATION OF EXTREME POINTS

Proposition: Consider a polyhedral set

$$P = \{x \mid a'_j x \leq b_j, j = 1, \dots, r\},$$

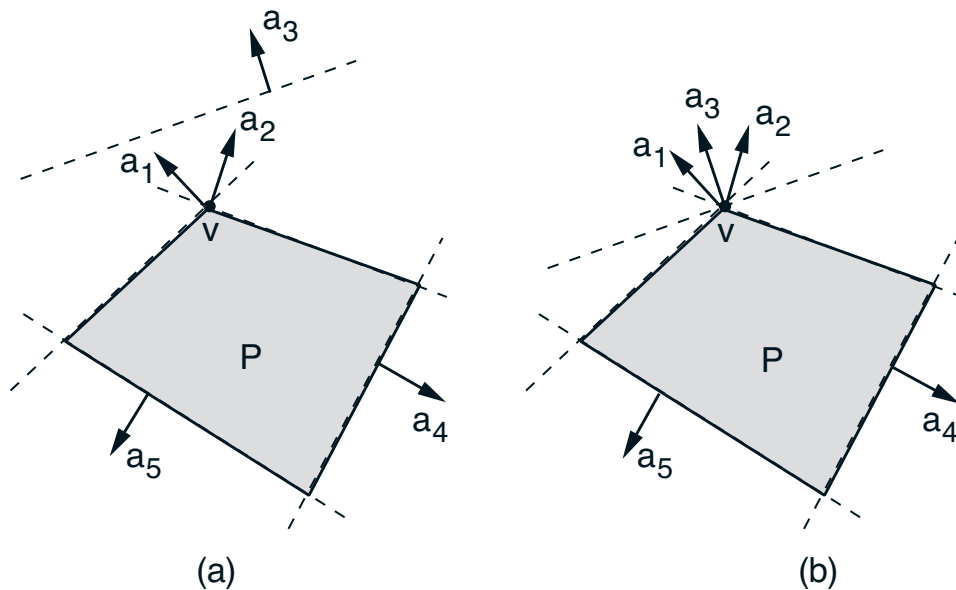
where a_j and b_j are given vectors and scalars.

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

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

contains n linearly independent vectors.

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



PROOF OUTLINE

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

$$a'_j w = 0, \quad \forall a_j \in A_v$$

has a nonzero solution \bar{w} . For small $\gamma > 0$, we have $v + \gamma\bar{w} \in P$ and $v - \gamma\bar{w} \in P$, 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 $a_j \in \bar{A}_v$,

$$b_j = a'_j v = \alpha a'_j y + (1 - \alpha) a'_j z \leq \alpha b_j + (1 - \alpha) b_j = b_j$$

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

$$a'_j w = b_j, \quad \forall a_j \in \bar{A}_v$$

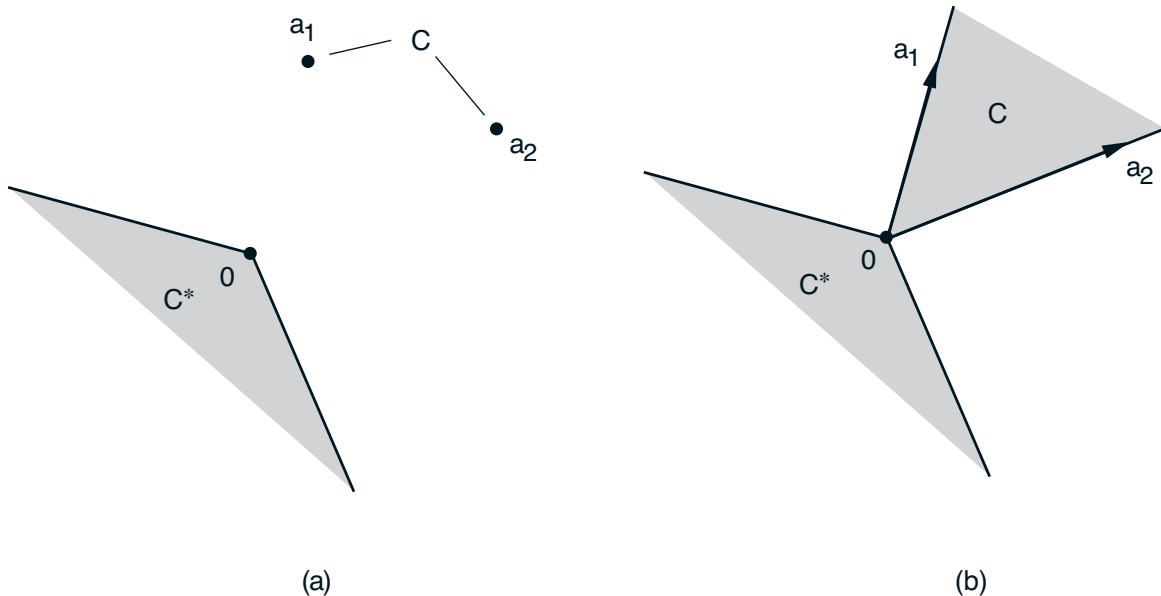
Hence, $v = y = z$, implying that v is an extreme point of P .

POLAR CONES

- Given a set C , the cone given by

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

is called the *polar cone* of C .



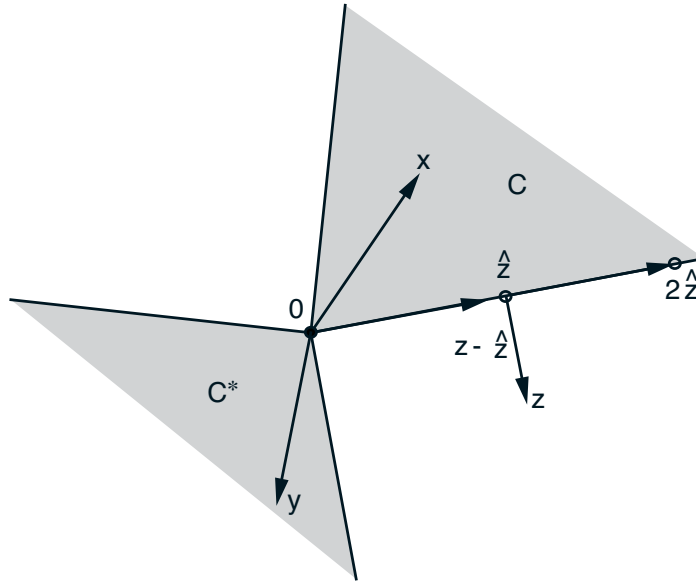
- C^* is a closed convex cone, since it is the intersection of closed halfspaces.
- Note that

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

- 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 $(C^*)^* = \text{cl}(\text{conv}(C))$.
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 $(C^*)^* \subset C$.

POLARS OF POLYHEDRAL CONES

- A cone $C \subset \mathbb{R}^n$ is *polyhedral*, if

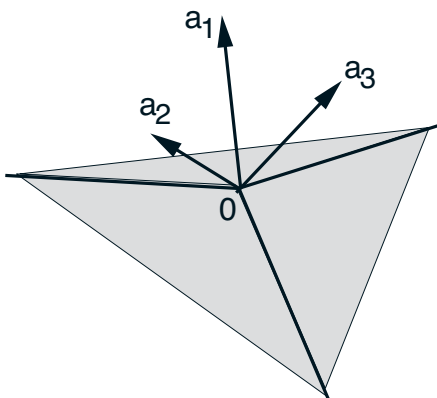
$$C = \{x \mid a'_j x \leq 0, j = 1, \dots, r\},$$

where a_1, \dots, a_r are some vectors in \mathbb{R}^n .

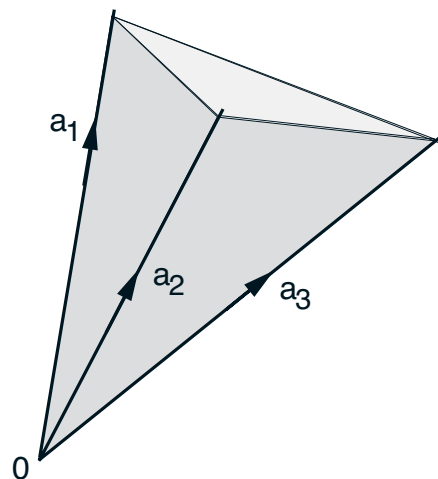
- A cone $C \subset \mathbb{R}^n$ is *finitely generated*, if

$$C = \left\{ x \mid x = \sum_{j=1}^r \mu_j a_j, \mu_j \geq 0, j = 1, \dots, r \right\}$$
$$= \text{cone}(\{a_1, \dots, a_r\}),$$

where a_1, \dots, a_r are some vectors in \mathbb{R}^n .



(a)



(b)

FARKAS-MINKOWSKI-WEYL THEOREMS

Let a_1, \dots, a_r be given vectors in \mathbb{R}^n , and let

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

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

$$(\{y \mid a'_j y \leq 0, j = 1, \dots, r\})^* = C$$

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

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

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

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

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

PROOF OUTLINE

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

$$C^* = (\text{cone}(\{a_1, \dots, a_r\}))^* = \{y \mid a'_j y \leq 0, j = 1, \dots, r\}.$$

If $y'a_j \leq 0$ for all j , then $y'x \leq 0$ for all $x \in C$,

so

$$C^* \supset \{y \mid a'_j y \leq 0, j = 1, \dots, r\}$$

Conversely, if $y \in C^*$, i.e., if $y'x \leq 0$ for all $x \in C$, then, since $a_j \in C$, we have $y'a_j \leq 0$, for all j . Thus,

$$C^* \subset \{y \mid a'_j y \leq 0, j = 1, \dots, r\}$$

- $\text{cone}(\{a_1, \dots, a_r\})$ is closed because closedness of polyhedral sets is preserved by linear transformations. (Shown in Section 1.5 using set intersection theorems.)

- Use the Polar Cone Theorem.

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

LECTURE 11

LECTURE OUTLINE

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

Recall some of the facts of polyhedral convexity:

- Polarity relation between polyhedral and finitely generated cones

$$\{x \mid a'_j x \leq 0, j = 1, \dots, r\} = \text{cone}(\{a_1, \dots, a_r\})^*$$

- Farkas' Lemma

$$\{x \mid a'_j x \leq 0, j = 1, \dots, r\}^* = \text{cone}(\{a_1, \dots, a_r\})$$

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

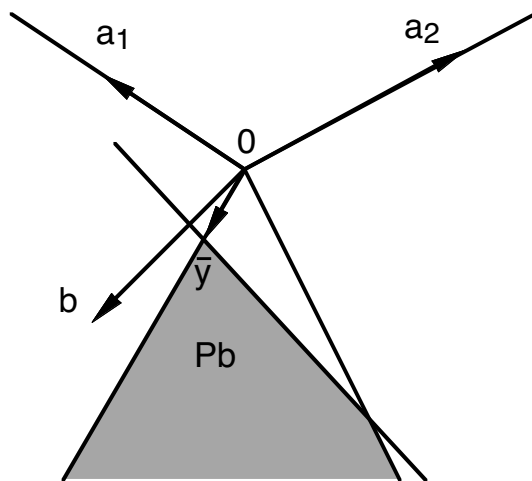
$$\text{Polyhedral set } P = \text{conv}(\{v_1, \dots, v_m\}) + R_P$$

MINKOWSKI-WEYL PROOF OUTLINE

- **Step 1:** Show $\text{cone}(\{a_1, \dots, a_r\})$ is polyhedral.
- **Step 2:** Use Step 1 and Farkas to show that $\{x \mid a'_j x \leq 0, j = 1, \dots, r\}$ is finitely generated.
- **Proof of Step 1:** Assume first that a_1, \dots, a_r span \mathbb{R}^n . Given $b \notin \text{cone}(\{a_1, \dots, a_r\})$,

$$P_b = \{y \mid b'y \geq 1, a'_j y \leq 0, j = 1, \dots, r\}$$

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



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

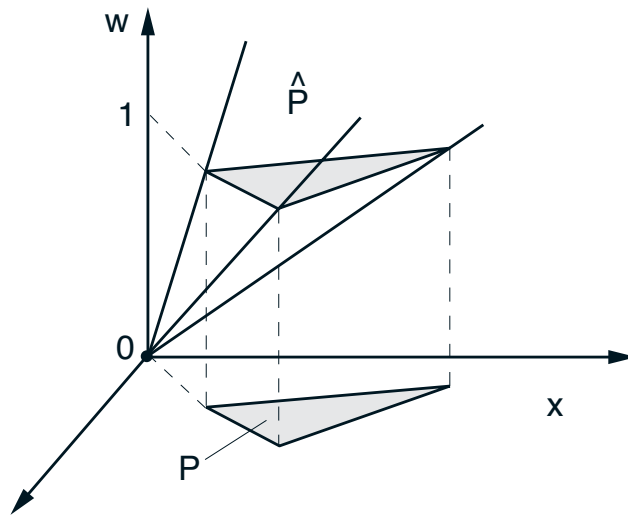
POLYHEDRAL REPRESENTATION PROOF OUTLINE

- We “lift the polyhedral set into a cone”. Let

$$P = \{x \mid a'_j x \leq b_j, j = 1, \dots, r\},$$

$$\hat{P} = \{(x, w) \mid 0 \leq w, a'_j x \leq b_j w, j = 1, \dots, r\}$$

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



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

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

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

PROOF CONTINUED

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

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

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

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

Thus,

$$P = \text{conv}(\{v_j \mid j \in J^+\}) + \left\{ \sum_{j \in J^0} \mu_j v_j \mid \mu_j \geq 0, j \in J^0 \right\}$$

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

POLYHEDRAL FUNCTIONS

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

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

$$f(x) = \max_{j=1, \dots, m} \{a'_j x + b_j\}, \quad \forall x \in \text{dom}(f),$$

for some $a_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$.

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

$$\begin{aligned} \text{epi}(f) &= \{(x, w) \mid x \in \text{dom}(f)\} \\ &\quad \cap \{(x, w) \mid a'_j x + b_j \leq w, j = 1, \dots, m\}. \end{aligned}$$

Since the two sets on the right are polyhedral, $\text{epi}(f)$ 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 $\{(x, w) \mid a'_j x + b_j \leq c_j w\}$, $j = 1, \dots, r$, where $a_j \in \mathbb{R}^n$, and $b_j, c_j \in \mathbb{R}$.
- Since for any $(x, w) \in \text{epi}(f)$, we have $(x, w + \gamma) \in \text{epi}(f)$ for all $\gamma \geq 0$, it follows that $c_j \geq 0$, so by normalizing if necessary, we may assume without loss of generality that either $c_j = 0$ or $c_j = 1$. Letting $c_j = 1$ for $j = 1, \dots, m$, and $c_j = 0$ for $j = m + 1, \dots, r$, where m is some integer,

$$\text{epi}(f) = \{(x, w) \mid a'_j x + b_j \leq w, j = 1, \dots, m, \\ a'_j x + b_j \leq 0, j = m + 1, \dots, r\}.$$

Thus

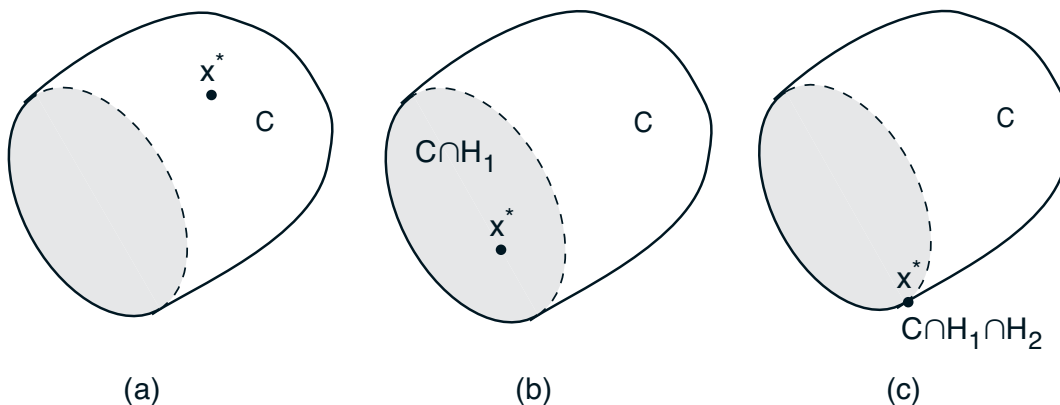
$$\text{dom}(f) = \{x \mid a'_j x + b_j \leq 0, j = m + 1, \dots, r\},$$

$$f(x) = \max_{j=1, \dots, m} \{a'_j x + b_j\}, \quad \forall x \in \text{dom}(f)$$

Q.E.D.

EXTREME POINTS AND CONCAVE MINIMIZATION

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



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

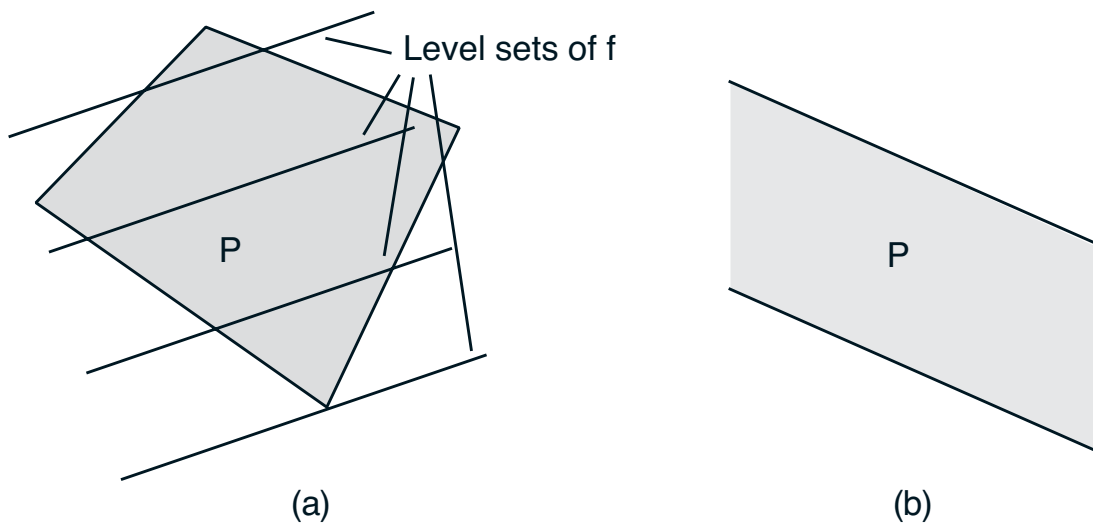
If $x^* \in \text{ri}(C \cap H_1)$ [see (b)], then f must be constant over $C \cap H_1$, so it attains a minimum at an extreme point $C \cap H_1$. This optimal extreme point is also an extreme point of C . If $x^* \notin \text{ri}(C \cap H_1)$, there is a hyperplane H_2 supporting $C \cap H_1$ through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

FUNDAMENTAL THEOREM OF LP

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

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

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



LINEAR PROGRAMMING DUALITY

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

$$\text{minimize } c'x$$

$$\text{subject to } a'_j x \geq b_j, \quad j = 1, \dots, r,$$

where c and a_1, \dots, a_r are vectors in \mathbb{R}^n .

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

$$\text{maximize } b'\mu$$

$$\text{subject to } \sum_{j=1}^r a_j \mu_j = c, \quad \mu_j \geq 0, \quad j = 1, \dots, r.$$

- **Duality Theorem:**

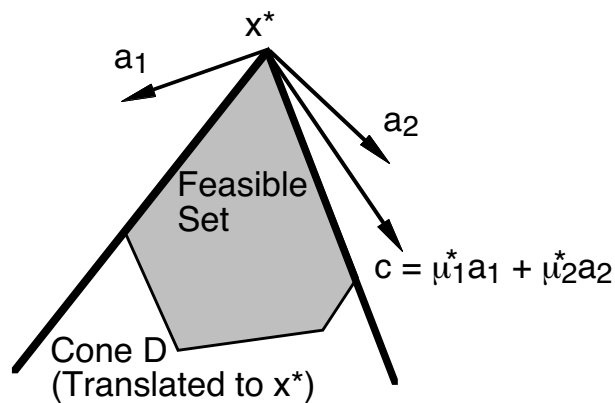
(a) If either f^* or q^* is finite, then $f^* = q^*$ and both problems have optimal solutions.

(b) If $f^* = -\infty$, then $q^* = -\infty$.

(c) If $q^* = \infty$, then $f^* = \infty$.

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

LINEAR PROGRAMMING DUALITY PROOF



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

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

Take inner product with x^* :

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

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

EXTREME POINTS AND INTEGER PROGRAMMING

- Consider a polyhedral set

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

where A is $m \times n$, $b \in \mathbb{R}^m$, and $c, d \in \mathbb{R}^n$. 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

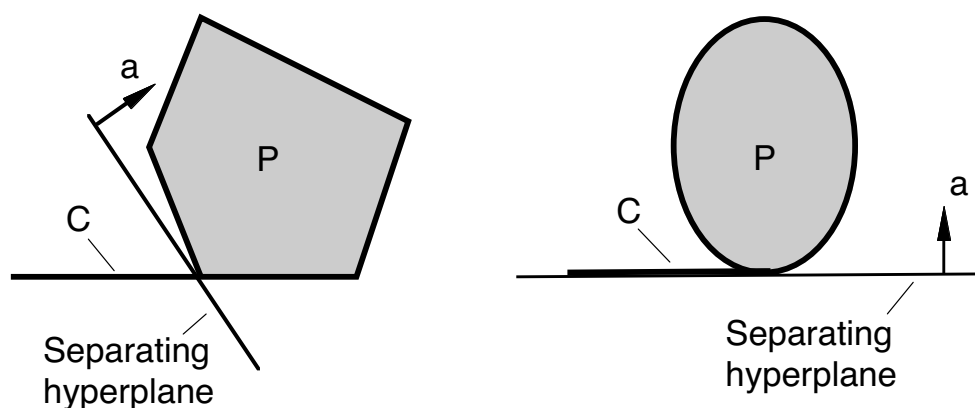
$$\text{ri}(C) \cap \text{ri}(P) = \emptyset$$

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

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

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 C/MAX C TH. - POLYHEDRAL VERSION

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

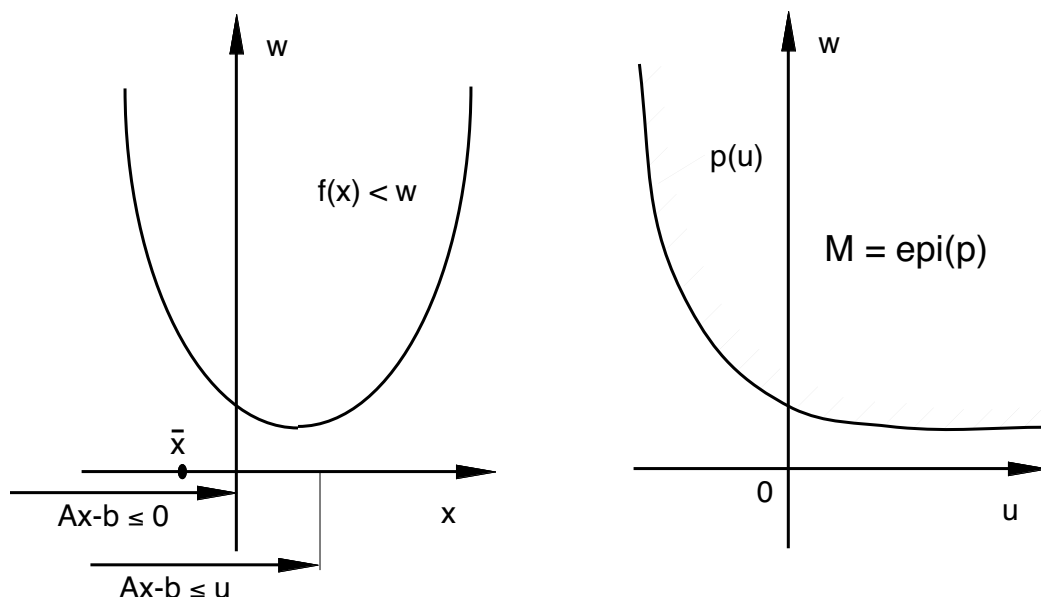
(1) The set M is defined in terms of a convex function $f : \mathbb{R}^m \mapsto (-\infty, \infty]$, an $r \times m$ matrix A , and a vector $b \in \mathbb{R}^r$:

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

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

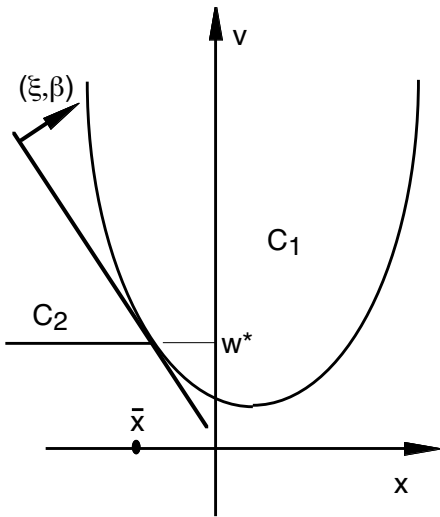
Then $q^* = w^*$ and there is a $\mu \geq 0$ with $q(\mu) = q^*$.

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



PROOF

- Consider the disjoint convex sets



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

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

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

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

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

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

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

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

PROOF (CONTINUED)

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

$$\xi' y \leq 0, \quad \forall y \text{ with } a'_j y \leq 0, \quad \forall j \in \bar{J},$$

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

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

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

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

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

NONLINEAR FARKAS' LEMMA

- Let $C \subset \mathbb{R}^n$ be convex, and $f : C \mapsto \mathbb{R}$ and $g_j : C \mapsto \mathbb{R}$, $j = 1, \dots, r$, be convex functions. Assume that

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

and one of the following two conditions holds:

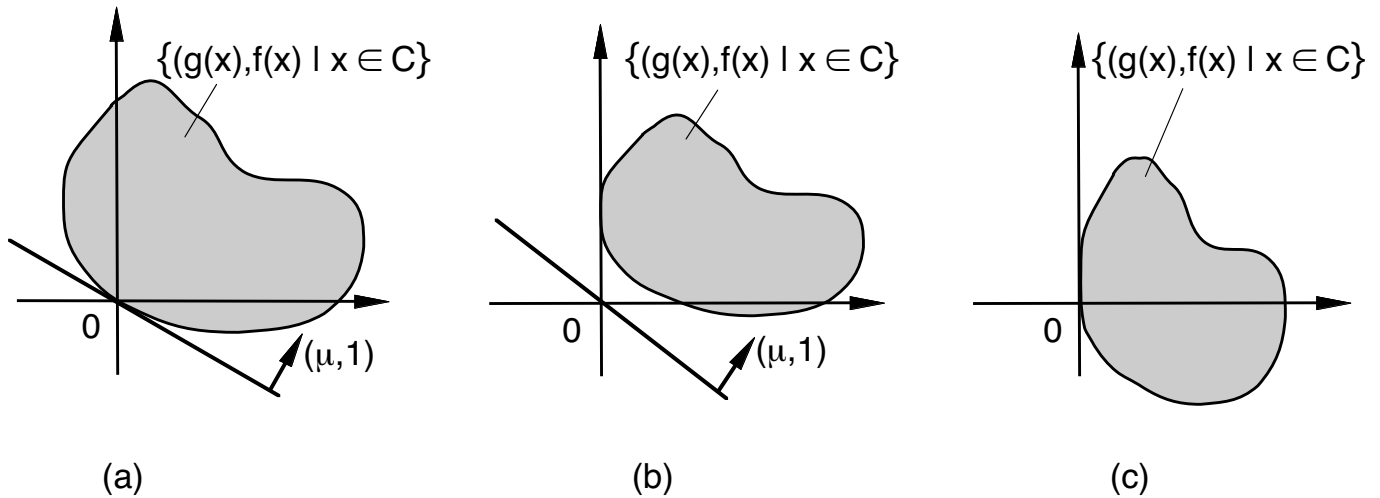
- (1) 0 is in the relative interior of the set $D = \{u \mid g(x) \leq u \text{ for some } x \in C\}$.
- (2) The functions g_j , $j = 1, \dots, r$, are affine, and F contains a relative interior point of C .

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

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

- Reduces to Farkas' Lemma if $C = \mathbb{R}^n$, 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 \mathbb{R}^{r+1} , with normal $(\mu, 1)$, that passes through the origin and contains the set

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

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

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

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

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

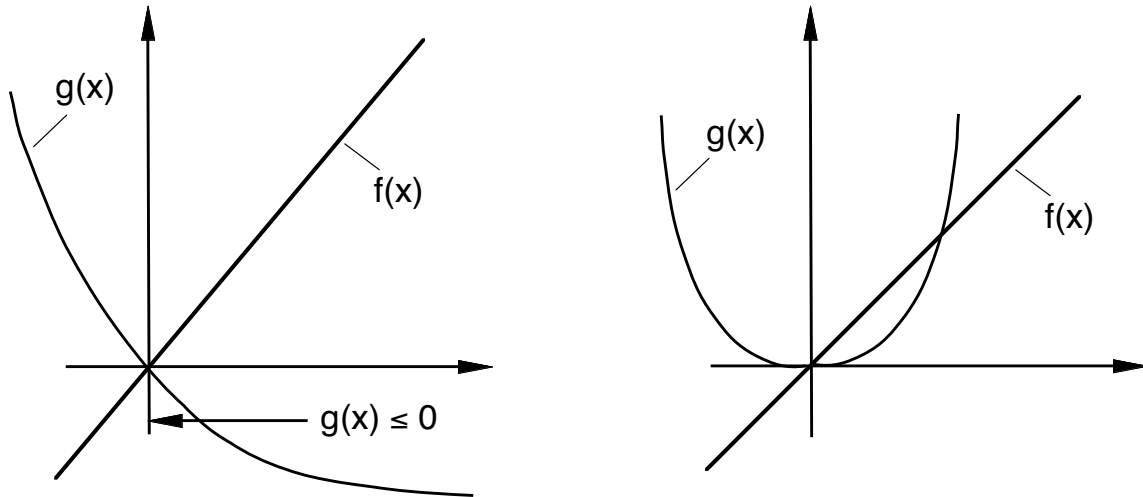
- Under condition (2), Min Common/Max Crossing Theorem III applies: $g(x) \leq 0$ can be written as $Ax - b \leq 0$.

- 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 ,

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

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

EXAMPLE



- Here $C = \mathfrak{R}$, $f(x) = x$. 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

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

- 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 $x \in F = \{x \in C \mid g_j(x) \leq 0, j = 1, \dots, r\}$

where $C \subset \mathbb{R}^n$ is convex, and $f : C \mapsto \mathbb{R}$ and $g_j : C \mapsto \mathbb{R}$ 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

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

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 $\bar{x} \in C$ with $g_j(\bar{x}) < 0$ for all j , or (2) the constraint functions g_j are affine and there is a feasible point in $\text{ri}(C)$.

LECTURE 13

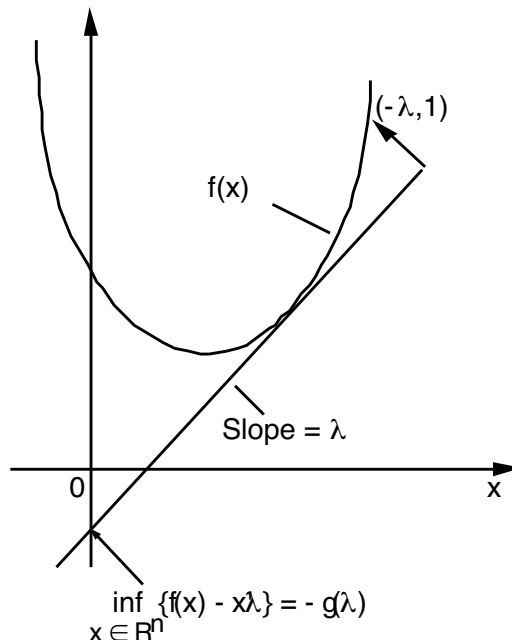
LECTURE OUTLINE

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

- Given f and its epigraph consider the function

Nonvertical hyperplanes supporting $\text{epi}(f)$

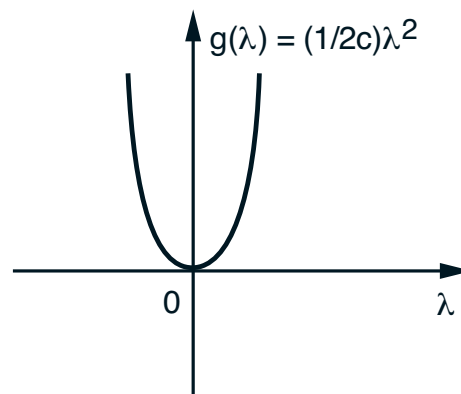
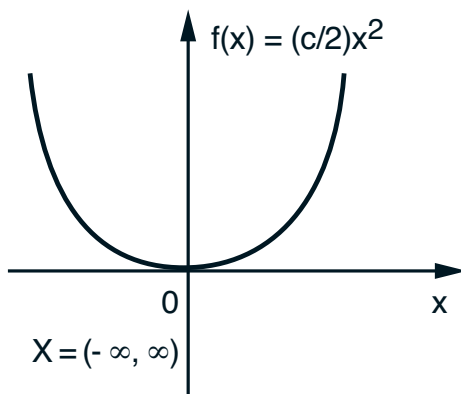
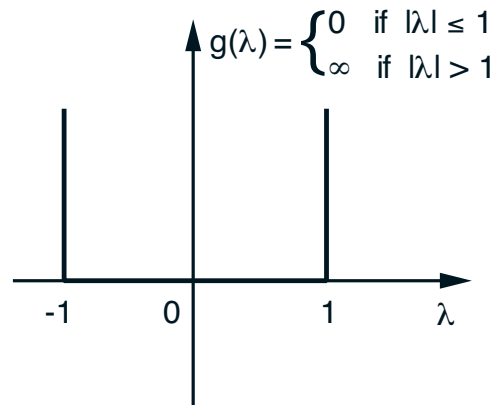
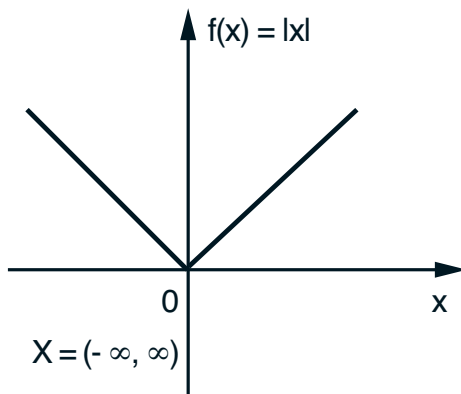
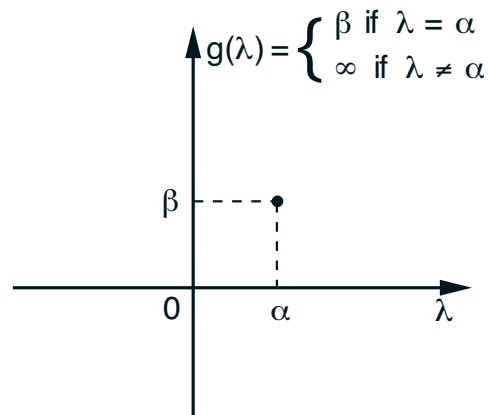
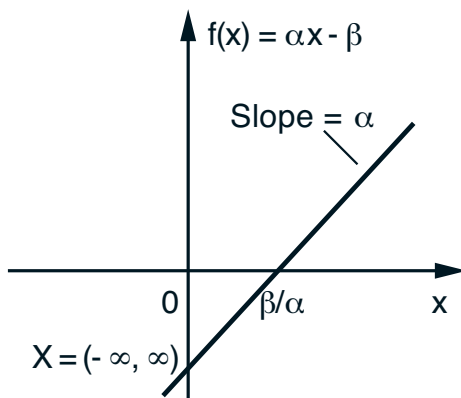
\mapsto Crossing points of vertical axis



CONJUGATE FUNCTIONS

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

$$g(\lambda) = \sup_{x \in \mathbb{R}^n} \{x' \lambda - f(x)\}, \quad \lambda \in \mathbb{R}^n.$$



CONJUGATE OF CONJUGATE

- From the definition

$$g(\lambda) = \sup_{x \in \mathfrak{R}^n} \{x' \lambda - f(x)\}, \quad \lambda \in \mathfrak{R}^n,$$

note that g is convex and closed.

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

$$g_x(\lambda) = x' \lambda - f(x)$$

as x ranges over \mathfrak{R}^n .

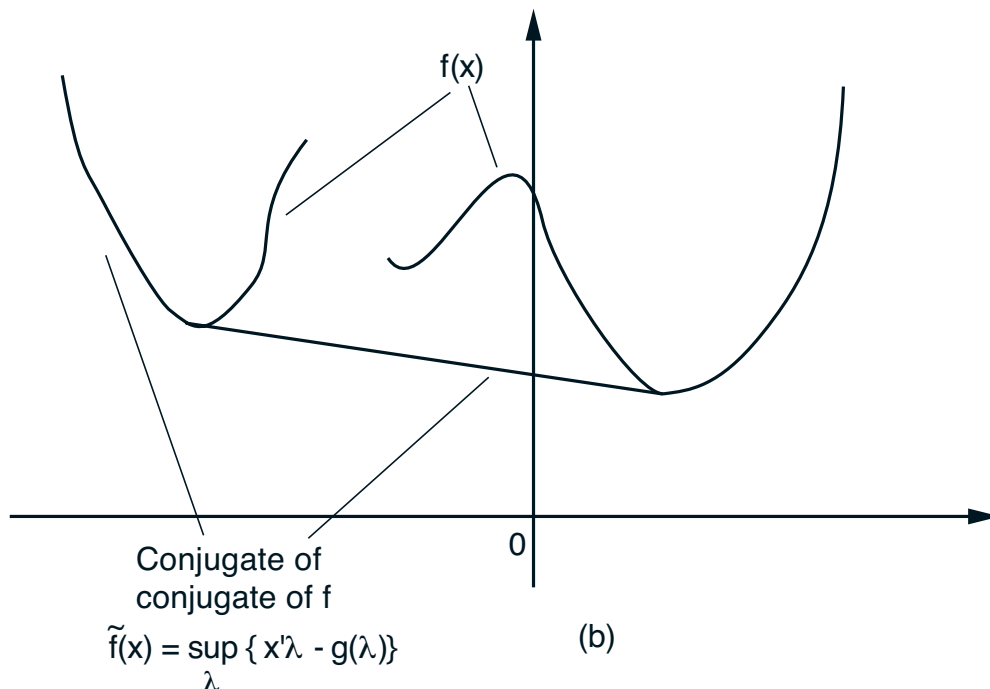
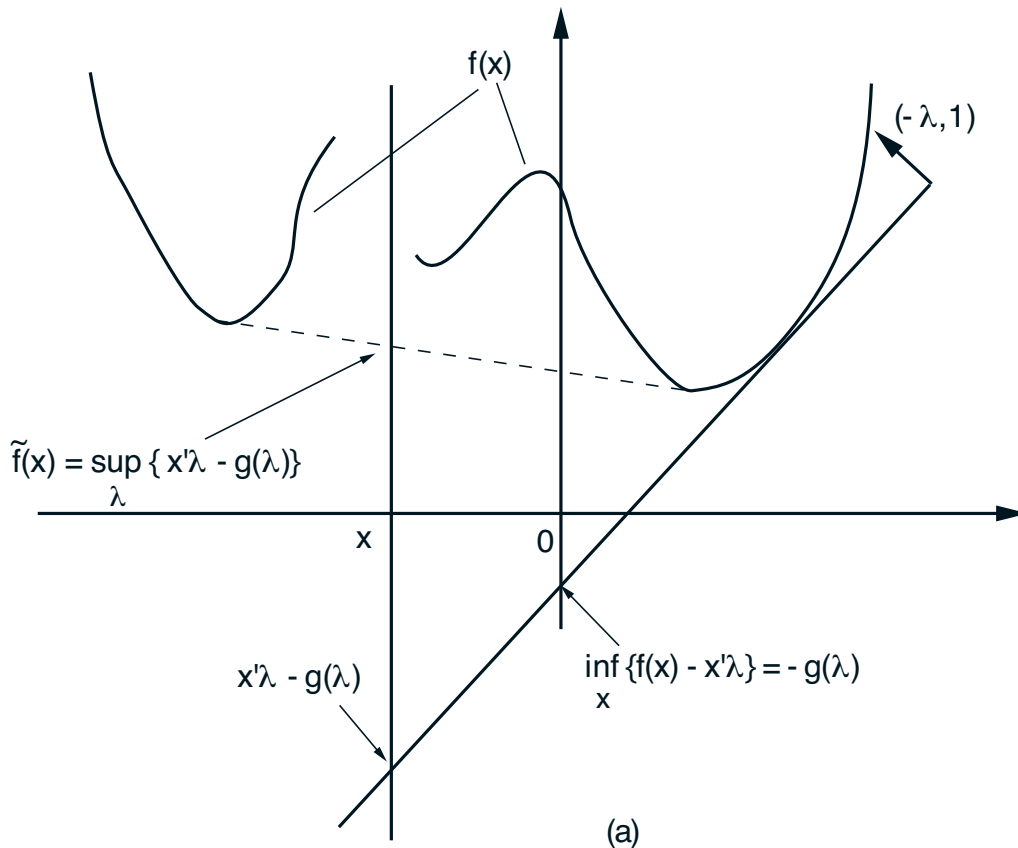
- Consider the conjugate of the conjugate:

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

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

VISUALIZATION

$$g(\lambda) = \sup_{x \in \mathbb{R}^n} \{x' \lambda - f(x)\}, \quad \lambda \in \mathbb{R}^n$$



EXTENSION

- Let $f : \mathbb{R}^n \mapsto [-\infty, \infty]$ be any function.
- Define $\hat{f} : \mathbb{R}^n \mapsto [-\infty, \infty]$, the *convex closure of f* , as the function that has as epigraph the closure of the convex hull of $\text{epi}(f)$ [also the smallest closed and convex set containing $\text{epi}(f)$].
- The conjugate of the conjugate of f is \hat{f} , assuming $\hat{f}(x) > -\infty$ for all x .
- A counterexample: Consider

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

We have

$$g(\lambda) = \infty, \quad \forall \lambda \in \mathbb{R}^n,$$

$$\tilde{f}(x) = -\infty, \quad \forall x \in \mathbb{R}^n.$$

But the convex closure of f is

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

CONJUGATE FUNCTION THEOREM

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

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

(a) We have

$$f(x) \geq \tilde{f}(x), \quad \forall x \in \mathbb{R}^n$$

(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

$$f(x) = \tilde{f}(x), \quad \forall x \in \mathbb{R}^n$$

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

$$\hat{f}(x) = \tilde{f}(x), \quad \forall x \in \mathbb{R}^n$$

A FEW EXAMPLES

- Logarithmic/Exponential Conjugacy
- l_p norm and l_q norm conjugacy, where

$$\frac{1}{p} + \frac{1}{q} = 1$$

- Conjugate of a strictly convex quadratic

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

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

- Conjugate of function obtained by invertible linear transformation and translation

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

$$g(\lambda) = q((A')^{-1}(\lambda - a)) + c'\lambda + d,$$

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

SUPPORT FUNCTIONS

- Consider a nonempty set X and its indicator function

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

- The conjugate of δ_X ,

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

is called the *support function of X* .

- Support functions provide alternative descriptions of closed convex sets.
- A lot of interesting results about sets, relative interiors, generated cones, etc, can be expressed in terms of support functions.
- **Example:** The support function of a cone is equal to the indicator function of its polar. Conjugacy theorem yields the polar cone theorem.

SUPPORT FUNCTIONS AND POLYHEDRAL SETS I

- Consider the Minkowski-Weyl representation of a polyhedral set

$$X = \text{conv}(\{v_1, \dots, v_m\}) + \text{cone}(\{y_1, \dots, y_r\})$$

- The support function is

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

Thus the support function of a polyhedral set is a polyhedral function.

SUPPORT FUNCTIONS AND POLYHEDRAL SETS II

- Consider f , g , and $\text{epi}(f)$. We have

$$\begin{aligned} g(\lambda) &= \sup_{x \in \mathbb{R}^n} \{x' \lambda - f(x)\} \\ &= \sup_{(x, w) \in \text{epi}(f)} \{x' \lambda - w\} \\ &= \sigma_{\text{epi}(f)}(\lambda, -1) \end{aligned}$$

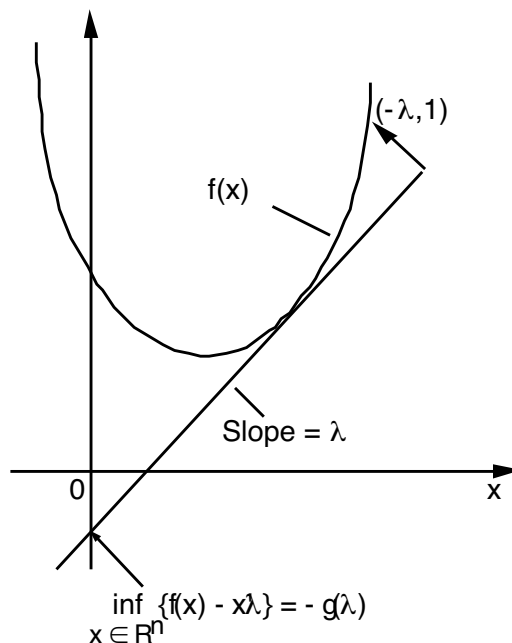
- If f is polyhedral, $\text{epi}(f)$ is a polyhedral set, so $\sigma_{\text{epi}(f)}$ is a polyhedral function, so g is a polyhedral function.
- Conclusion: Conjugates of polyhedral functions are polyhedral.

LECTURE 14

LECTURE OUTLINE

- Properties of convex conjugates and support functions

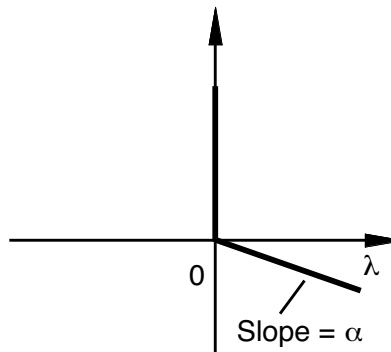
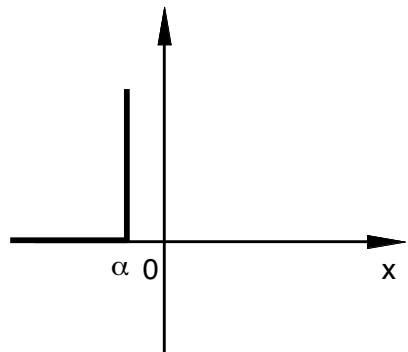
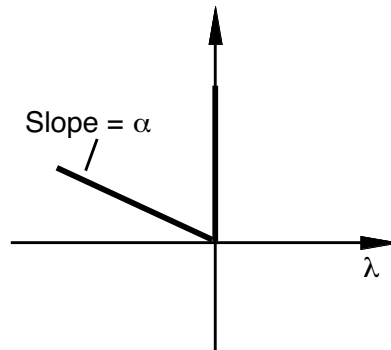
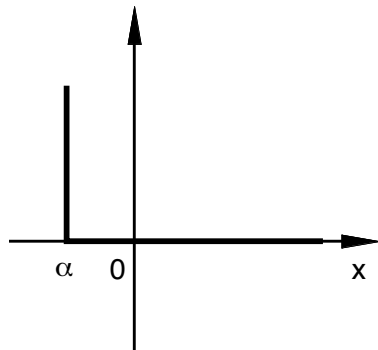
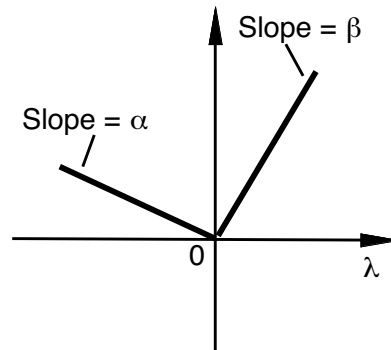
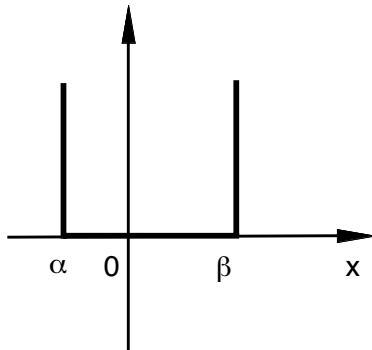
- Conjugate of f : $g(\lambda) = \sup_{x \in \mathbb{R}^n} \{x' \lambda - f(x)\}$



- Support function of set C = Conjugate of its indicator function
- **Conjugacy Theorem:** The conjugate of the conjugate of a proper convex function f is the closure of f

SUPPORT FUNCTIONS

- $\sigma_C(\lambda) = \sup_{x \in C} \lambda'x$



Indicator Functions

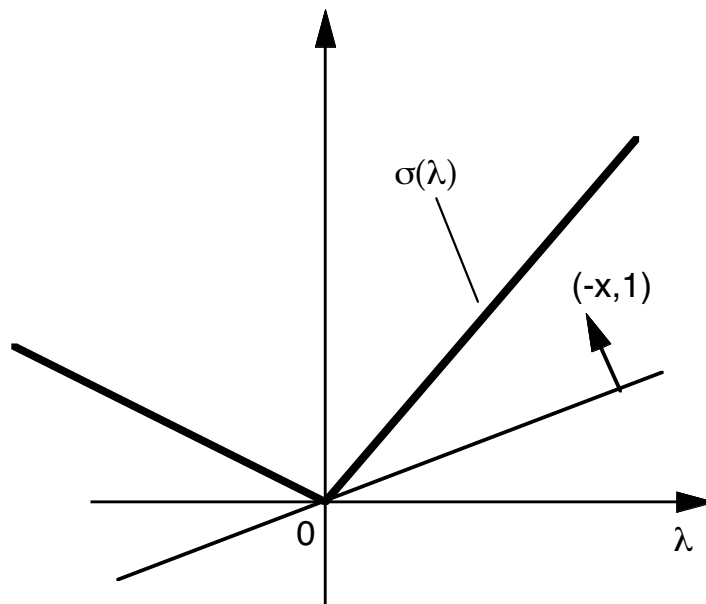
Support Functions

- The epigraph of σ_C is a closed cone

POSITIVELY HOMOGENEOUS FUNCTIONS

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

$$f(\gamma x) = \gamma f(x), \quad \forall \gamma > 0, \forall x \in \mathfrak{R}^n$$

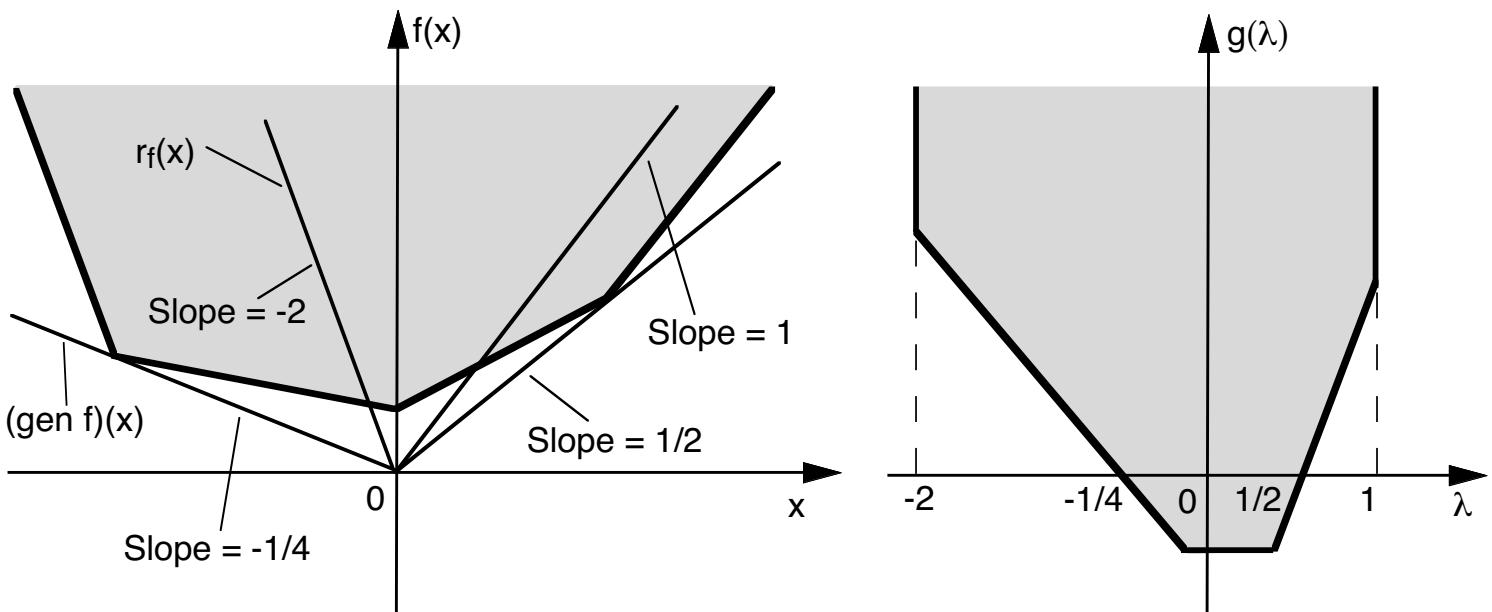


- A support function is closed, proper, convex, and positively homogeneous.
- *Converse Result:* The closure of a proper, convex, and positively homogeneous function σ is the support function of the closed convex set

$$X = \{x \mid \lambda'x \leq (\text{cl } \sigma)(\lambda), \forall \lambda \in \mathfrak{R}^n\}$$

CONES ASSOCIATED WITH SETS AND FUNCTIONS

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



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

FORMULAS FOR DOMAIN, LEVEL SETS, ETC I

• **Support Function of Domain:** Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a proper convex function, and let g be its conjugate.

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

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

• **Polar and Recession Cones, and Support Function:** Let C be a nonempty convex set in \mathbb{R}^n .

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

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

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

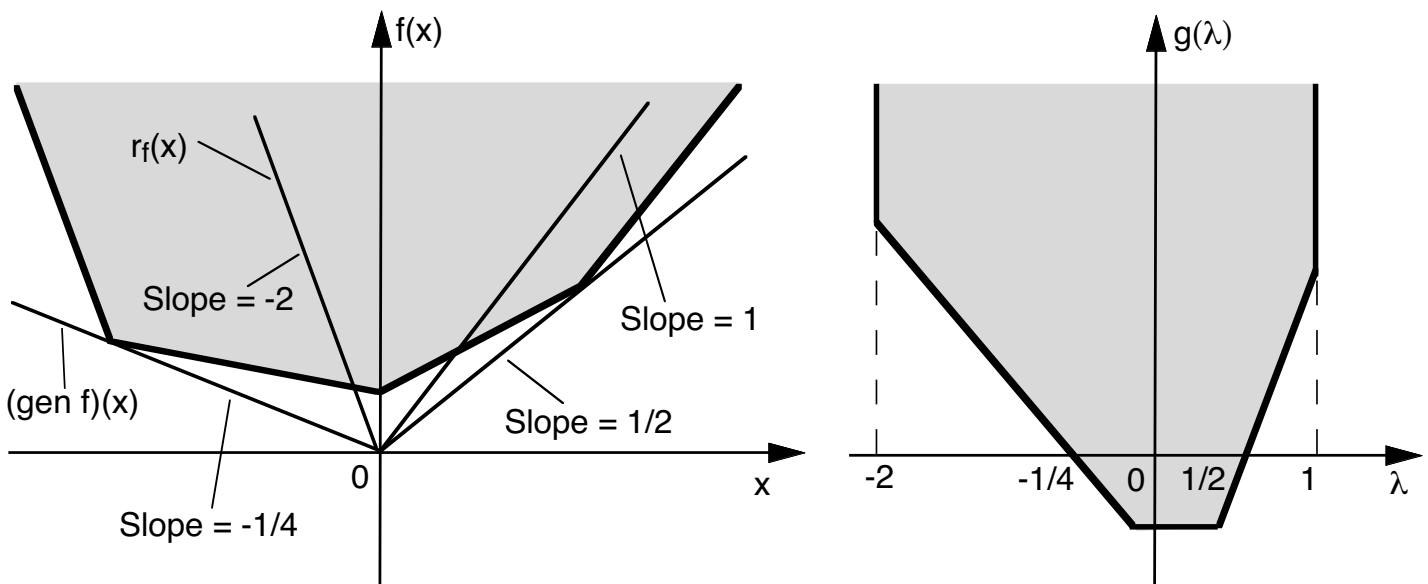
$$R_C = (\text{dom}(\sigma_C))^*.$$

FORMULAS FOR DOMAIN, LEVEL SETS, ETC II

- **Support Function of 0-Level Set:** Let $f : \mathcal{R}^n \mapsto (-\infty, \infty]$ be a closed proper convex function and let g be its conjugate.

(a) If the level set $\{\lambda \mid g(\lambda) \leq 0\}$ is nonempty, its support function is the closed function generated by f .

(b) If the level set $\{x \mid f(x) \leq 0\}$ is nonempty, its support function is the closed function generated by g .



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

CALCULUS OF CONJUGATE FUNCTIONS

- It is useful to have formulas for the conjugate of the sum or linear composition of convex functions.
- **Example:** Consider $F(x) = f(Ax)$, where f is closed proper convex, and A is a matrix.
- If g is the conjugate of f , we have

$$\begin{aligned} f(Ax) &= \sup_{\mu} \{x' A' \mu - g(\mu)\} \\ &= \sup_{\lambda} \sup_{A' \mu = \lambda} \{x' A \mu - g(\mu)\} \\ &= \sup_{\lambda} \left\{ x' \lambda - \inf_{A' \mu = \lambda} g(\mu) \right\} \end{aligned}$$

so F is the conjugate of G given by

$$G(\lambda) = \inf_{A' \mu = \lambda} g(\mu)$$

called the *image function* of g under A' . Hence the conjugate of F is the *closure* of G .

- Issues of preservation of closedness under partial minimization.

CLOSEDNESS OF IMAGE FUNCTION

- We view the image function

$$G(\lambda) = \inf_{A' \mu = \lambda} g(\mu)$$

as the result of partial minimization with respect to μ of a function of (μ, λ)

- We use the results on preservation of closedness under partial minimization
 - The image function is closed and the infimum is attained for all $\lambda \in \text{dom}(G)$ if g is closed and every direction of recession of g that belongs to $N(A')$ is a direction along which g is constant.
- This condition can be translated to an alternative and more useful condition involving the relative interior of the domain of the conjugate of g . In particular, we will show that the condition is true if and only if

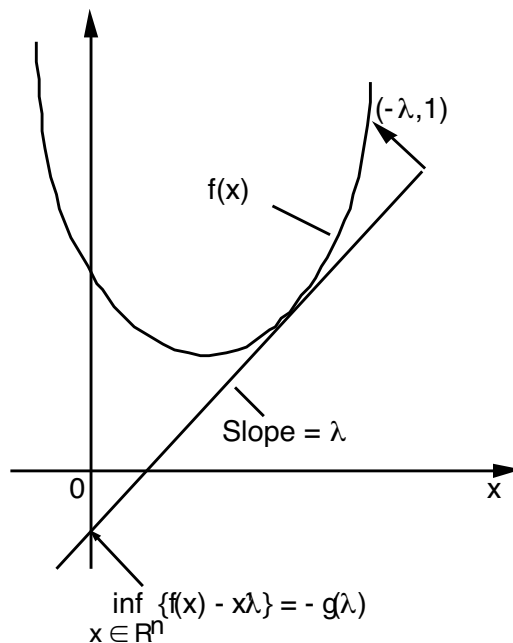
$$R(A) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$$

LECTURE 15

LECTURE OUTLINE

- Calculus of convex conjugate functions
- Subgradients

- Conjugate of f : $g(\lambda) = \sup_{x \in \mathbb{R}^n} \{x' \lambda - f(x)\}$
- Support function of set C = Conjugate of its indicator function



- We want to develop formulas that allow us to calculate convex conjugate functions

IMAGE AND INFIMAL CONVOLUTION I

- Let $f, f_i : \mathbb{R}^n \mapsto (-\infty, \infty]$ be proper convex functions, and A be a matrix. Consider the following conjugate pair (modulo a closure operation)

- Linear composition: $f(Ax)$

- Image: $\inf_{Ay=x} f(y)$

- Another conjugate pair – the special case where

$$x = (x_1, \dots, x_m) \in \mathbb{R}^{mn}, \quad f(x) = f_1(x_1) + \dots + f_m(x_m),$$

$$A(x_1, \dots, x_m) = x_1 + \dots + x_m$$

- Sum (special case of linear composition):

$$f_1(x) + \dots + f_m(x)$$

- Infimal convolution (special case of image):

$$\inf_{x_1 + \dots + x_m = x} \{ f_1(x_1) + \dots + f_m(x_m) \}$$

- Note that image and infimal convolution are special cases of partial minimization, and their closure is not guaranteed even if f, f_i are closed

IMAGE AND INFIMAL CONVOLUTION II

- Note also that for $F(x) = \inf_{Ay=x} f(y)$, we have

$$\overline{A} \text{epi}(f) \subset \text{epi}(F) \subset \text{cl}(\overline{A} \text{epi}(f)),$$

where \overline{A} is

$$\overline{A}(x, w) = (Ax, w)$$

- From preservation of closure results:
 - F is closed and proper, and the infimum in the definition of F is attained for all $x \in \text{dom}(F)$ if every direction of recession of f that belongs to $N(A)$, is a direction along which f is constant.
- For $F = f_1 + \cdots + f_m$, we have

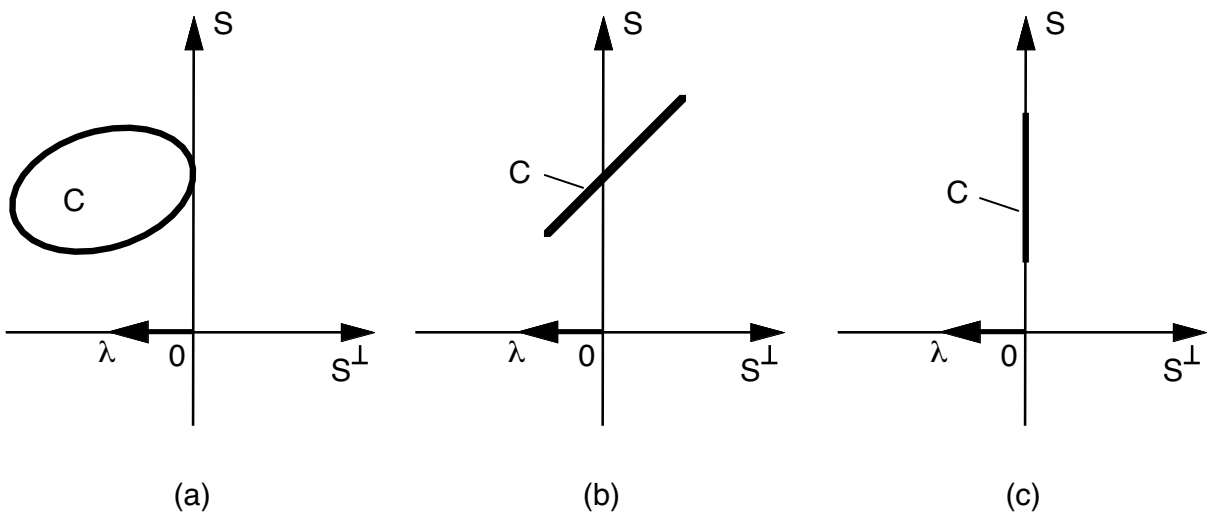
$$\text{epi}(f_1) + \cdots + \text{epi}(f_m) \subset \text{epi}(F) \subset \text{cl}(\text{epi}(f_1) + \cdots + \text{epi}(f_m))$$

- Also F is closed and proper, and the infimum is attained for all $x \in \text{dom}(F)$ under a similar condition on directions of recession of f_i

DUAL VIEW OF PRESERVATION OF CLOSURE

- **Proper Separation Lemma:** Let S and C be a subspace and a convex set. Then $S \cap \text{ri}(C) \neq \emptyset$ if and only if the support function σ_C satisfies

$$\sigma_C(\lambda) = \sigma_C(-\lambda) = 0, \quad \forall \lambda \in S^\perp \text{ with } \sigma_C(\lambda) \leq 0.$$



- Let S be a subspace of \mathbb{R}^n , let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a proper convex function, and let g be its conjugate. We have

$$S \cap \text{ri}(\text{dom}(f)) \neq \emptyset$$

if and only if every direction of recession of g that belongs to S^\perp is a direction along which g is constant.

CONJUGATE OF IMAGE THEOREM

- Let $f : \Re^n \mapsto (-\infty, \infty]$ be a proper convex function, and let g be its conjugate. Consider the function F defined by

$$F(x) = \inf_{Ay=x} f(y),$$

where A is an $r \times n$ matrix.

- (a) F is convex and its conjugate is

$$G(\lambda) = g(A'\lambda).$$

- (b) If f is closed and the range of A' contains a point in $\text{ri}(\text{dom}(g))$, then F is closed and proper, and the infimum in the definition of F is attained for all $x \in \text{dom}(F)$.
- (c) If f is polyhedral, and the range of A' contains a point in $\text{dom}(g)$, then F is polyhedral, and the infimum in the definition of F is attained for all $x \in \text{dom}(F)$.

CONJUGATE OF INFIMAL CONVOLUTION THEOREM

- Let $f_i : \mathfrak{R}^n \mapsto (-\infty, \infty]$, $i = 1, \dots, m$, be proper convex functions, and let g_i be their corresponding conjugates. Consider the function F defined by

$$F(x) = \inf_{x_1 + \dots + x_m = x} \{f_1(x_1) + \dots + f_m(x_m)\}.$$

- (a) F is convex and its conjugate is

$$G(\lambda) = g_1(\lambda) + \dots + g_m(\lambda).$$

- (b) If for some k , the functions f_1, \dots, f_k are polyhedral, the functions f_{k+1}, \dots, f_m are closed, and

$$\left(\bigcap_{i=1}^k \text{dom}(g_i) \right) \cap \left(\bigcap_{i=1}^m \text{ri}(\text{dom}(g_i)) \right) \neq \emptyset,$$

then F is closed and proper, and the infimum in the definition of F is attained for all $x \in \text{dom}(F)$.

LINEAR TRANSFORMATIONS – VECTOR SUMS

- Let $C, C_i, i = 1, \dots, m$, be nonempty closed convex subsets of \mathbb{R}^n , and let

$$D = \text{dom}(\sigma_C), \quad D_i = \text{dom}(\sigma_{C_i}), \quad i = 1, \dots, m.$$

- (a) Let A be an $m \times n$ matrix. Then the set AC is closed if

$$R(A') \cap \text{ri}(D) \neq \emptyset,$$

where $R(A')$ is the range of A' .

- (b) The vector sum $C_1 + \dots + C_m$ is closed if

$$\bigcap_{i=1}^m \text{ri}(D_i) \neq \emptyset.$$

- (c) The vector sum $C_1 + \dots + C_m$ is closed if for some $k \geq 1$, C_1, \dots, C_k is polyhedral, and

$$\left(\bigcap_{i=1}^k D_i\right) \cap \left(\bigcap_{i=k+1}^m \text{ri}(D_i)\right) \neq \emptyset.$$

SUBGRADIENTS

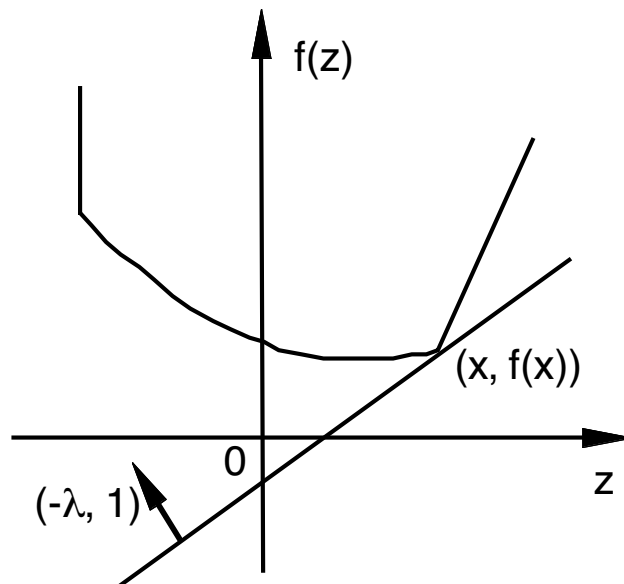
- Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a convex function. A vector $\lambda \in \mathbb{R}^n$ is a *subgradient* of f at a point $x \in \text{dom}(f)$ if

$$f(z) \geq f(x) + (z - x)' \lambda, \quad \forall z \in \mathbb{R}^n$$

- λ is a subgradient if and only if

$$f(z) - z' \lambda \geq f(x) - x' \lambda, \quad \forall z \in \mathbb{R}^n$$

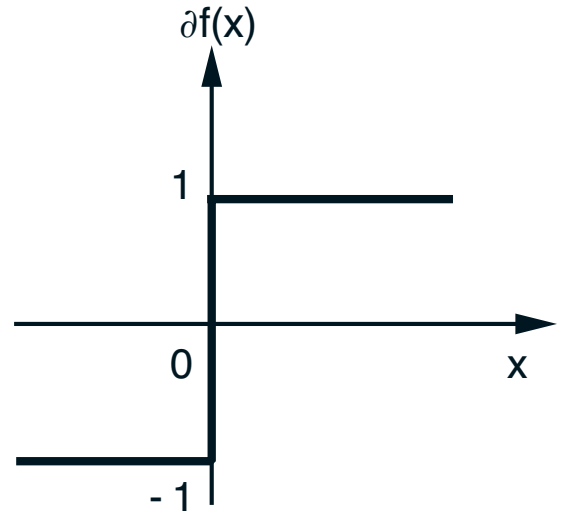
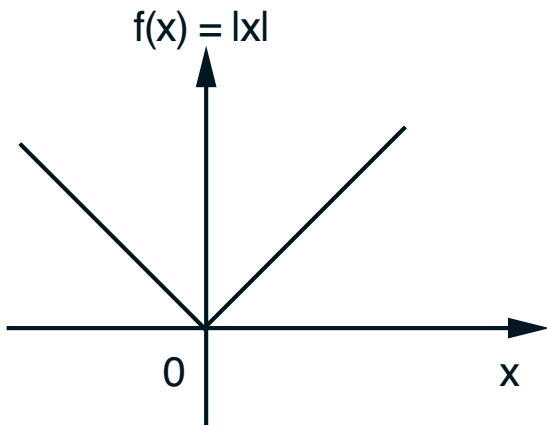
so λ is a subgradient at x if and only if the hyperplane in \mathbb{R}^{n+1} that has normal $(-\lambda, 1)$ and passes through $(x, f(x))$ supports the epigraph of f .



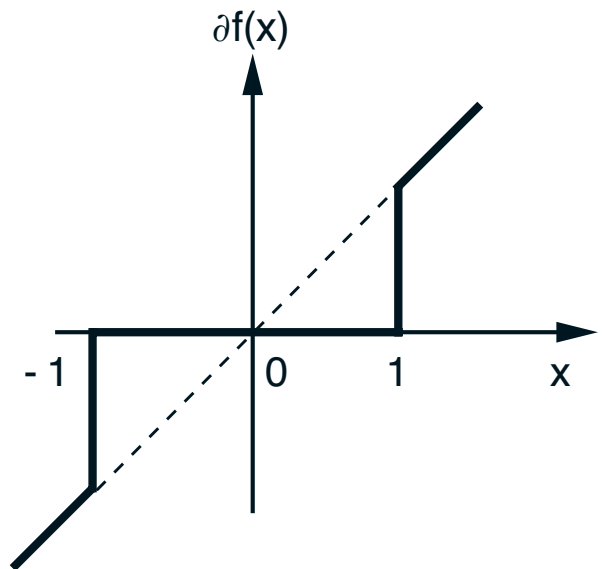
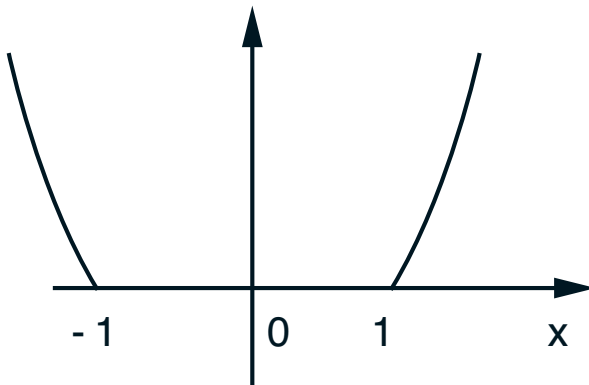
- The set of all subgradients at x is the *subdifferential* of f at x , denoted by $\partial f(x)$.

EXAMPLES OF SUBDIFFERENTIALS

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

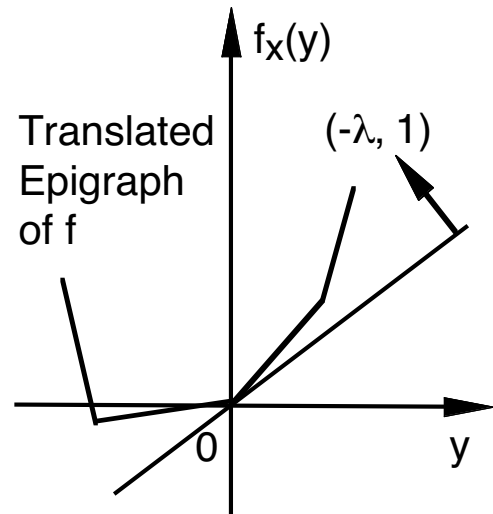
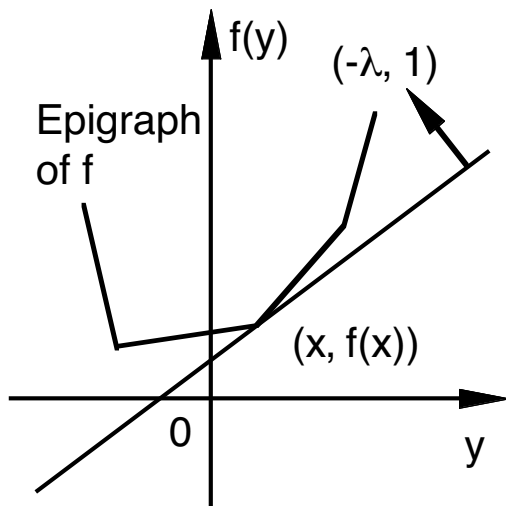


$f(x) = \max\{0, (1/2)(x^2 - 1)\}$



EXISTENCE OF SUBGRADIENTS

- Note the connection with the min common/max crossing construction.



- Let $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ be a proper convex function. For every $x \in \text{ri}(\text{dom}(f))$,

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

where:

- S is the subspace that is parallel to the affine hull of $\text{dom}(f)$
- G is a nonempty and compact set.
- Furthermore, $\partial f(x)$ is nonempty and compact if and only if x is in the interior of $\text{dom}(f)$.

LECTURE 16

LECTURE OUTLINE

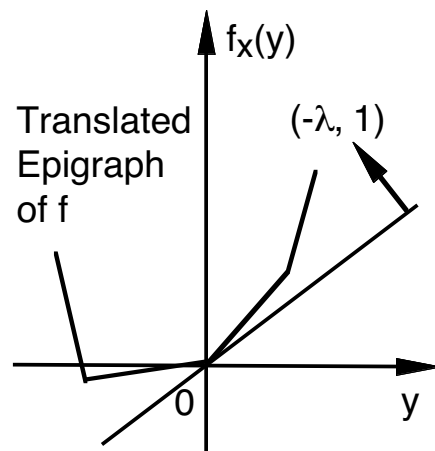
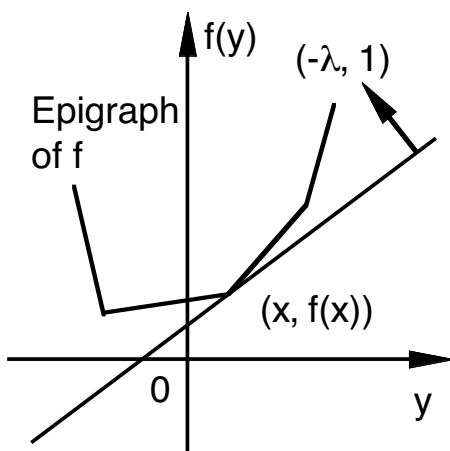
- Calculus of subgradients
- Constrained optimality conditions
- Directional derivatives

- Subgradient inequality: $\lambda \in \partial f(x)$ if

$$f(z) \geq f(x) + (z - x)' \lambda, \quad \forall z \in \mathbb{R}^n$$

- Connection with min common/max crossing and conjugacy, using x -translation

$$f_x(y) = f(x + y) - f(x)$$



EXAMPLE: SUBDIFFERENTIAL OF INDICATOR

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

$$\delta_C(z) \geq \delta_C(x) + \lambda'(z - x), \quad \forall z \in C,$$

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

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

- **Example:** For the case of a polyhedral set

$$P = \{x \mid a'_i x \leq b_i, i = 1, \dots, m\},$$

we have

$$N_P(x) = \begin{cases} \mathbb{R}^n & \text{if } x \in \text{int}(P), \\ \text{cone}(\{a_i \mid a'_i x = b_i\}) & \text{if } x \notin \text{int}(P). \end{cases}$$

FENCHEL INEQUALITY

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

$$x' \lambda \leq f(x) + g(\lambda), \quad \forall x \in \mathfrak{R}^n, \lambda \in \mathfrak{R}^n.$$

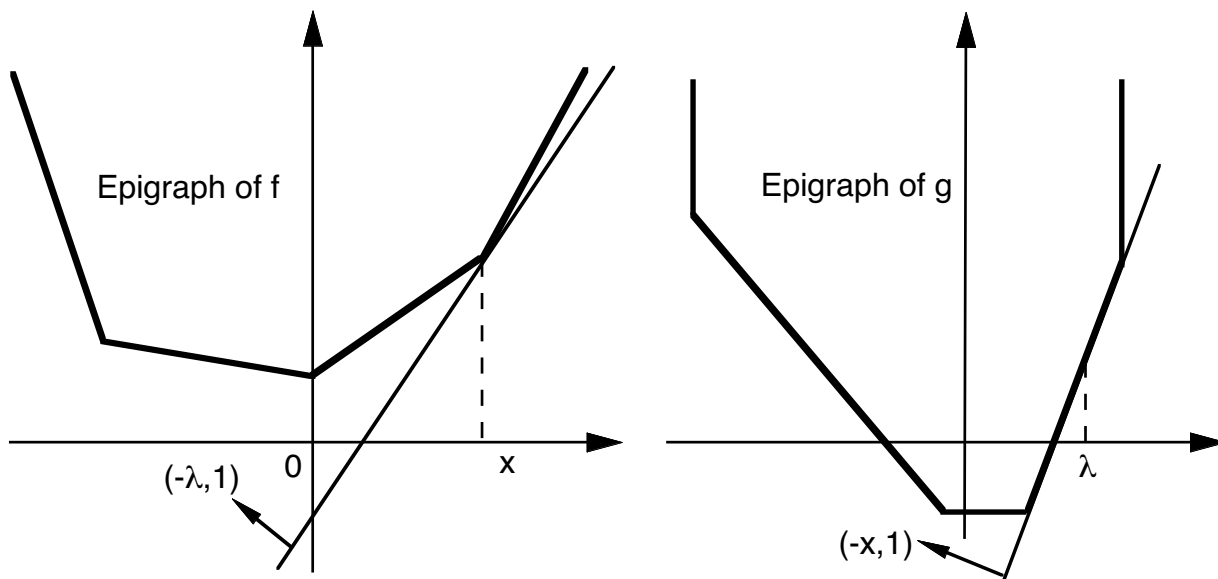
- Proposition:** The following two relations are equivalent for a pair of vectors (x, λ) :

(i) $x' \lambda = f(x) + g(\lambda)$.

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

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

(iii) $x \in \partial g(\lambda)$.



MINIMA OF CONVEX FUNCTIONS

• **Application:** Let f be closed convex and let X^* be the set of minima of f over \mathbb{R}^n . Then:

- (a) $X^* = \partial g(0)$. [Since $0 \in \partial f(x)$ iff $x \in \partial g(0)$.]
- (b) X^* is nonempty if $0 \in \text{ri}(\text{dom}(g))$.
- (c) X^* is nonempty and compact if and only if $0 \in \text{int}(\text{dom}(g))$.

• **Proof:** (a) From the subgradient inequality,

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

which is true if and only if

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

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

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

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

EXAMPLE: SUBDIFF. OF SUPPORT FUNCTION

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

$$h(\lambda) = \sigma_C(\lambda + \bar{\lambda}), \quad \lambda \in \mathfrak{R}^n.$$

- We have $\partial\sigma_C(\bar{\lambda}) = \partial h(0)$, so $\partial\sigma_C(\bar{\lambda})$ is equal to the set of minima over \mathfrak{R}^n of the conjugate of h .
- The conjugate of h is $\sup_{\lambda \in \mathfrak{R}^n} \{\lambda'x - h(\lambda)\}$, or

$$\sup_{\lambda \in \mathfrak{R}^n} \{\lambda'x - \sigma_C(\lambda + \bar{\lambda})\} = \delta(x) - \bar{\lambda}'x,$$

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

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

EXAMPLE: SUBDIFF. OF POLYHEDRAL FUNCTION

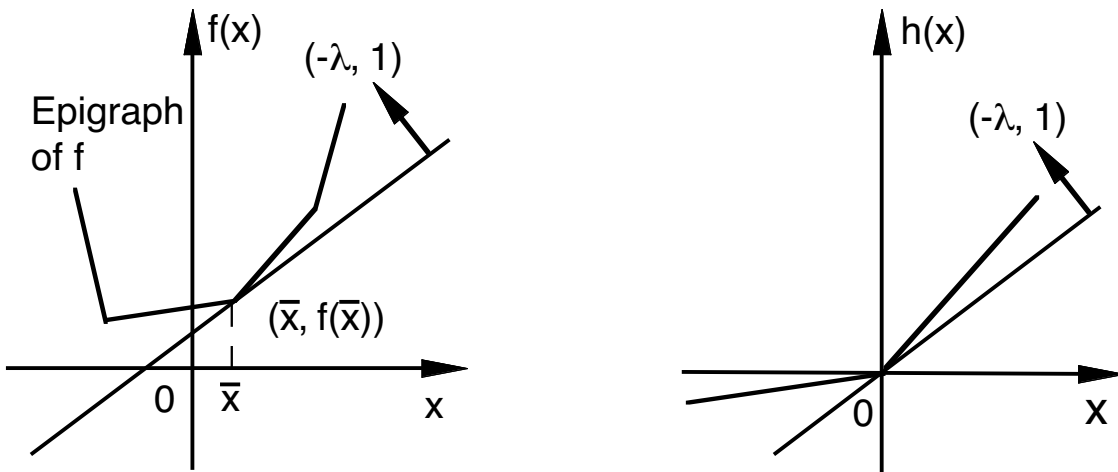
- Let

$$f(x) = \max\{a'_1 x + b_1, \dots, a'_r x + b_r\}.$$

- For a fixed $\bar{x} \in \mathbb{R}^n$, consider

$$A_{\bar{x}} = \{j \mid a'_j \bar{x} + b_j = f(\bar{x})\}$$

and the function $h(x) = \max\{a'_j x \mid j \in A_{\bar{x}}\}$.



- We have $\partial f(\bar{x}) = \partial h(0)$.
- Since h is the support function of the finite set $\{a_j \mid j \in A_{\bar{x}}\}$, we see that $\partial h(0)$ is

$$\partial f(\bar{x}) = \partial h(0) = \text{conv}(\{a_j \mid j \in A_{\bar{x}}\})$$

SUBDIFFERENTIAL CALCULUS: CHAIN RULE

- Let $f : \mathbb{R}^m \mapsto (-\infty, \infty]$ be a proper convex, and A be a matrix. We want to calculate $\partial F(x)$, where $F(x) = f(Ax)$.

- **Claim:** If $R(A) \cap \text{ri}(\text{dom}(f))$, then

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

- **Proof:** The conjugate of F is

$$G(\lambda) = \inf_{A'\mu = \lambda} g(\mu)$$

where g is the conjugate of f , and the infimum is attained for all $\lambda \in \text{dom}(G)$.

- We have $\lambda \in \partial F(x)$ iff $F(x) + G(\lambda) = x'\lambda$, or iff there exists a vector μ such that $A'\mu = \lambda$ and $F(x) + g(\mu) = x'A'\mu$, or

$$f(Ax) + g(\mu) = x'A'\mu.$$

Therefore, $\lambda \in \partial F(x)$ iff for some μ such that $A'\mu = \lambda$, we have $\mu \in \partial f(Ax)$. **Q.E.D.**

SUM OF FUNCTIONS

- Let $f_i : \mathbb{R}^n \mapsto (-\infty, \infty]$, $i = 1, \dots, m$, be proper convex functions, and let

$$f = f_1 + \dots + f_m.$$

- Assume that

$$\bigcap_{i=1}^m \text{ri}(\text{dom}(f_i)) \neq \emptyset.$$

- Then

$$\partial f(x) = \partial f_1(x) + \dots + \partial f_m(x), \quad \forall x \in \mathbb{R}^n.$$

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

$$\left(\bigcap_{i=1}^k \text{dom}(f_i) \right) \cap \left(\bigcap_{i=k+1}^m \text{ri}(\text{dom}(f_i)) \right) \neq \emptyset.$$

CONSTRAINED OPTIMALITY CONDITION

- Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be proper convex, let X be a convex subset of \mathbb{R}^n , and assume that one of the following four conditions holds:
 - (i) $\text{ri}(\text{dom}(f)) \cap \text{ri}(X) \neq \emptyset$.
 - (ii) f is polyhedral and $\text{dom}(f) \cap \text{ri}(X) \neq \emptyset$.
 - (iii) X is polyhedral and $\text{ri}(\text{dom}(f)) \cap X \neq \emptyset$.
 - (iv) f and X are polyhedral, and $\text{dom}(f) \cap X \neq \emptyset$.

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

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

Proof: x^* minimizes

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

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

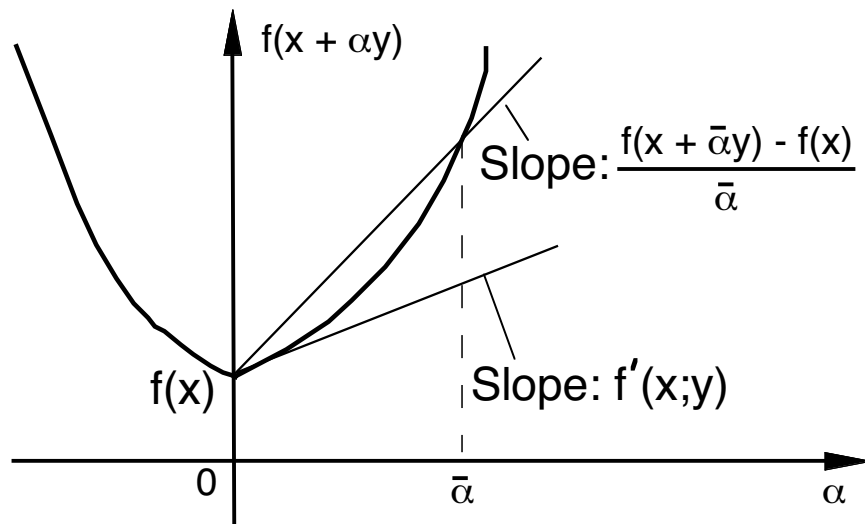
DIRECTIONAL DERIVATIVES

- Consider a proper convex $f : \mathbb{R}^n \mapsto (-\infty, \infty]$. For $x \in \text{dom}(f)$, the *directional derivative of f at x in the direction y* is

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

- **Important fact:** By convexity, the ratio is monotonically nondecreasing with α , so

$$f'(x; y) = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x)}{\alpha}$$



DIRECTIONAL DERIVATIVES AND SUBGRADIENTS

- Since $f'(x; \cdot)$ is positively homogeneous, its closure is the support function of the set whose indicator function is the conjugate of $f'(x; \cdot)$, i.e., the set

$$\{\lambda \mid \lambda'y - f'(x; y) \leq 0, \forall y \in \mathbb{R}^n\}$$

- From the equation

$$f'(x; y) = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x)}{\alpha}$$

this set is $\partial f(x)$.

- **Conclusion:** The support function of $\partial f(x)$ is $(\text{cl } f')(x, \cdot)$.
- Also, for $x \in \text{ri}(\text{dom}(f))$,

$$f'(x; \cdot) = (\text{cl } f')(x, \cdot)$$

- **Special case:** If f is real-valued,

$$f'(x; y) = \sigma_{\partial f(x)}(y), \quad \forall x, y \in \mathbb{R}^n$$

DIRECTIONAL DERIVATIVE OF THE MAX FUNCTION

- Consider the directional derivative of the function

$$f(x) = \max\{f_1(x), \dots, f_r(x)\},$$

where $f_j : \mathbb{R}^n \mapsto \mathbb{R}$, $j = 1, \dots, r$, are convex functions.

- Denote

$$A_x = \{j \mid f_j(x) = f(x)\}.$$

Then it can be shown that

$$f'(x; y) = \max\{f'_j(x; y) \mid j \in A_x\}, \quad \forall x, y \in \mathbb{R}^n.$$

- Thus the epigraph of $f'(x; \cdot)$ is the intersection of the epigraphs of $f'_j(x; \cdot)$, $j \in A_x$. From this, it follows that

$$\partial f(x) = \text{conv}(\cup_{j \in A_x} \partial f_j(x)).$$

LECTURE 17

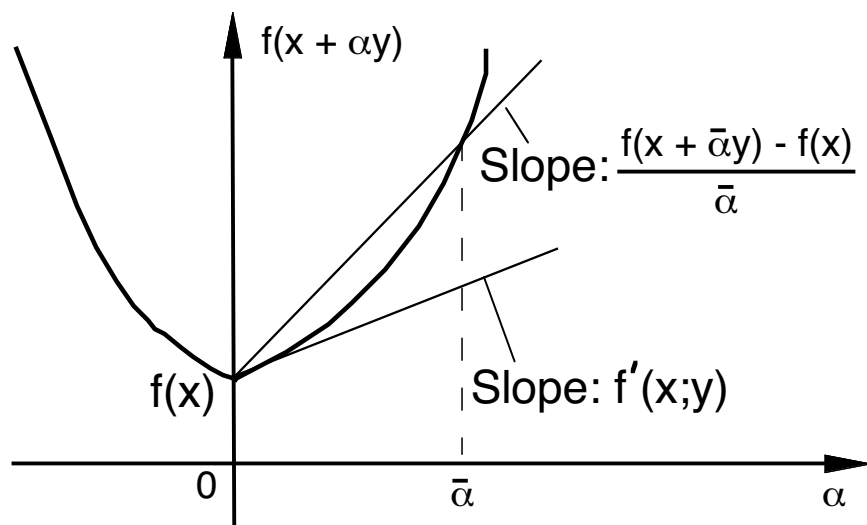
LECTURE OUTLINE

- Descent algorithms for minimization
- Steepest descent
- ϵ -subgradients and ϵ -descent

- Recall that the directional derivative of f at x in the direction y is

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

- It is the support function of $\partial f(x)$



STEEPEST DESCENT

- Given convex $f : \mathbb{R}^n \mapsto \mathbb{R}$, the *direction of steepest descent* at x solves

$$\begin{aligned} & \text{minimize} && f'(x; y) \\ & \text{subject to} && \|y\| \leq 1 \end{aligned}$$

- Equivalently, it solves

$$\begin{aligned} & \text{minimize} && \max_{\lambda \in \partial f(x)} y' \lambda \\ & \text{subject to} && \|y\| \leq 1 \end{aligned}$$

- Using minimax equality, the direction of steepest descent is $y^* = -\lambda^* / \|\lambda^*\|$, where λ^* is the vector of min norm on $\partial f(x)$.

- Steepest descent method:

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

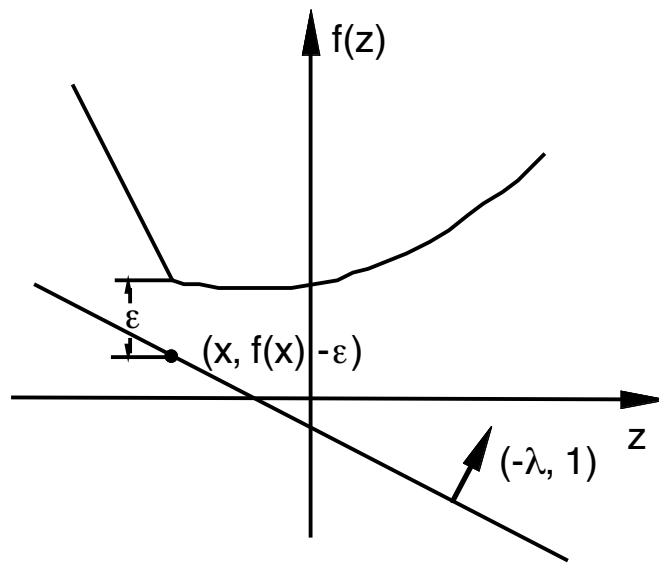
where λ_k is the vector of minimum norm in $\partial f(x_k)$, and $\alpha_k > 0$ is such that $f(x_{k+1}) < f(x_k)$.

- Difficulties due to discontinuity of $\partial f(x)$.

ϵ -SUBGRADIENTS

- Given a proper convex $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ and $\epsilon > 0$, we say that λ is an ϵ -subgradient of f at a point $x \in \text{dom}(f)$ if

$$f(z) \geq f(x) + (z - x)' \lambda - \epsilon, \quad \forall z \in \mathfrak{R}^n$$



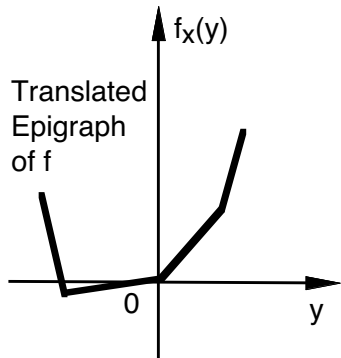
- The ϵ -subdifferential $\partial_\epsilon f(x)$ is the set of all ϵ -subgradients of f at x .
- By convention, $\partial_\epsilon f(x) = \emptyset$ for $x \notin \text{dom}(f)$.
- We will show that for f closed, $\partial_\epsilon f(x) \neq \emptyset$ for $x \in \text{dom}(f)$.

ϵ -SUBGRADIENTS AND CONJUGACY I

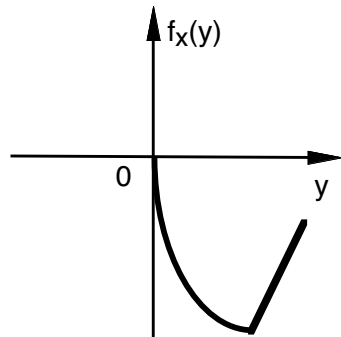
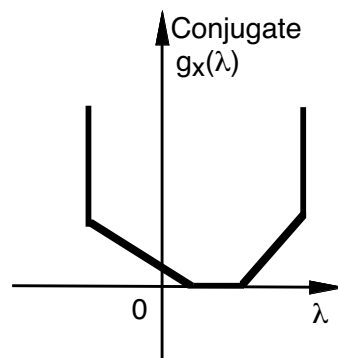
- Consider x -translation: $f_x(y) = f(x + y) - f(x)$
- The conjugate of f_x , denoted g_x is

$$g_x(\lambda) = \sup_{y \in \mathbb{R}^n} \{y' \lambda - f(x + y) + f(x)\} = g(\lambda) + f(x) - \lambda' x,$$

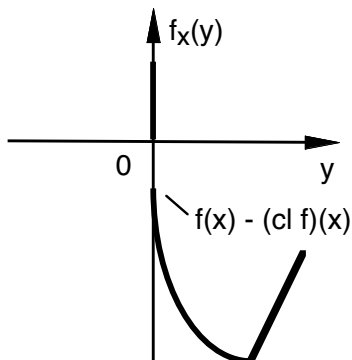
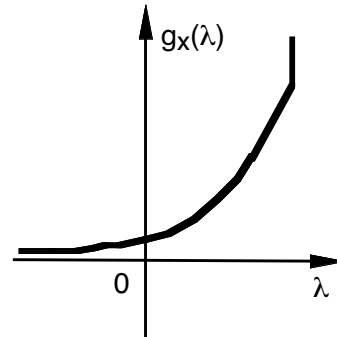
where g is the conjugate of f .



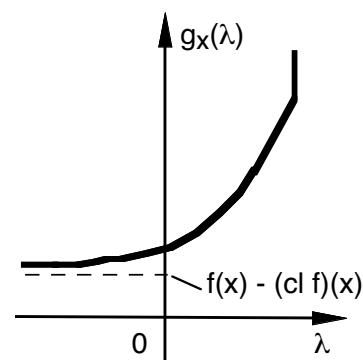
(a)



(b)



(c)



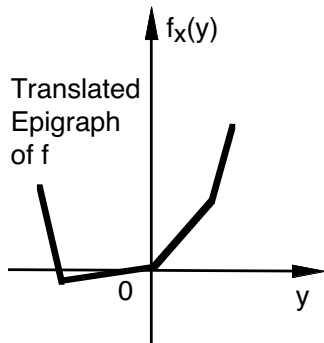
ϵ -SUBGRADIENTS AND CONJUGACY II

- Since

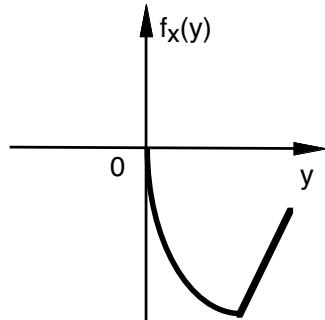
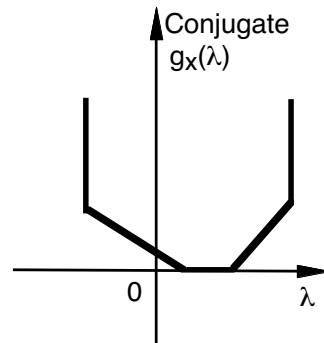
$$\lambda \in \partial f(x) \text{ iff } \sup_{y \in \mathbb{R}^n} \{ \lambda' y - f(x + y) + f(x) \} \leq 0,$$

we have

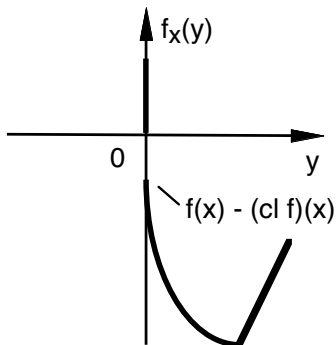
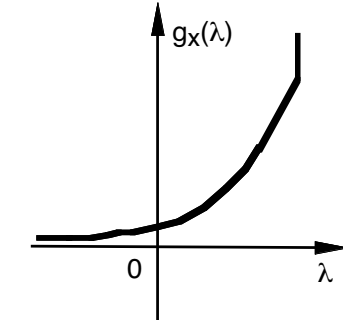
$$\partial f(x) = \{ \lambda \mid g_x(\lambda) \leq 0 \}, \quad \partial_\epsilon f(x) = \{ \lambda \mid g_x(\lambda) \leq \epsilon \}$$



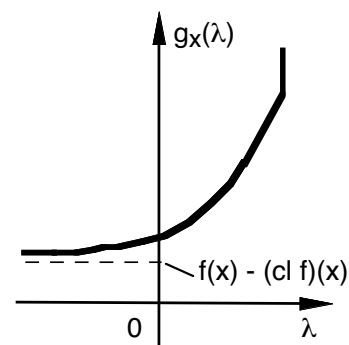
(a)



(b)



(c)



ϵ -SUBGRADIENT PROPERTIES

• Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a proper convex and let $\epsilon > 0$. For every $x \in \text{dom}(f)$, the following hold:

(a) $\partial_\epsilon f(x)$ is a closed convex set.

(b) If $(\text{cl } f)(x) = f(x)$, then $\partial_\epsilon f(x) \neq \emptyset$ and its support function is

$$\sigma_{\partial_\epsilon f(x)}(y) = \sup_{\lambda \in \partial_\epsilon f(x)} y' \lambda = \inf_{\alpha > 0} \frac{f(x + \alpha y) - f(x) + \epsilon}{\alpha}$$

(c) $\partial_\epsilon f(x)$ is compact iff $x \in \text{int}(\text{dom}(f))$.

(d) If f is real-valued, $\partial_\epsilon f(x)$ is compact.

ϵ -DESCENT DIRECTIONS

- We say that y is an ϵ -descent direction at $x \in \text{dom}(f)$ if

$$\inf_{\alpha > 0} f(x + \alpha y) < f(x) - \epsilon.$$

- Let $f : \mathfrak{R}^n \mapsto (-\infty, \infty]$ be a proper convex, let ϵ be a positive scalar, and let $x \in \text{dom}(f)$ be such that $(\text{cl } f)(x) = f(x)$. Then:

- (a) We have $0 \in \partial_\epsilon f(x)$ if and only if

$$f(x) \leq \inf_{z \in \mathfrak{R}^n} f(z) + \epsilon.$$

- (b) We have $0 \notin \partial_\epsilon f(x)$ if and only if there exists an ϵ -descent direction.

- (c) If $0 \notin \partial_\epsilon f(x)$, the vector $-\bar{\lambda}$ where

$$\bar{\lambda} = \arg \min_{\lambda \in \partial_\epsilon f(x)} \|\lambda\|,$$

is an ϵ -descent direction.

ϵ -DESCENT METHOD

- Let $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ be a proper convex function to be minimized, and assume that $(\text{cl } f)(x) = f(x)$ for all $x \in \text{dom}(f)$.
- ϵ -Descent method: It stops at an ϵ -optimal solution, and otherwise

$$x_{k+1} = x_k + \alpha_k y_k,$$

where y_k is an ϵ -descent direction and $\alpha_k > 0$ is such that

$$f(x_k + \alpha_k y_k) \leq f(x_k) - \epsilon$$

- **Implementation:** Find the projection

$$\lambda_k = \arg \min_{\lambda \in \partial_\epsilon f(x_k)} \|\lambda\|,$$

to determine whether $\lambda_k = 0$.

- If $\lambda_k \neq 0$, then $y_k = -\lambda_k$ is an ϵ -descent direction.

OUTER APPROXIMATION VARIANT

- Approximate $\partial_\epsilon f(x_k)$ by a set $A(x_k)$ such that

$$\partial_\epsilon f(x_k) \subset A(x_k) \subset \partial_{\gamma\epsilon} f(x_k),$$

where γ is a scalar with $\gamma > 1$.

- Use $y_k = -\lambda_k$ with $\lambda_k = \arg \min_{\lambda \in A(x_k)} \|\lambda\|$
- If $\lambda_k = 0$, the method stops, and x_k is within $\gamma\epsilon$ of being optimal.
- If $\lambda_k \neq 0$, by suitable choice of the stepsize α_k , we can move along the direction $y_k = -\lambda_k$ to decrease the cost function by at least ϵ .
- Important special case:

$$f(x) = f_1(x) + \cdots + f_m(x),$$

$$\partial_\epsilon f(x) \subset \text{cl}(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x)) \subset \partial_{m\epsilon} f(x)$$

At each iteration project on

$$A(x_k) = \text{cl}(\partial_\epsilon f_1(x) + \cdots + \partial_\epsilon f_m(x))$$

LECTURE 18

LECTURE OUTLINE

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

GEOMETRICAL FRAMEWORK FOR MULTIPLIERS

- We consider the problem

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

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

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

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

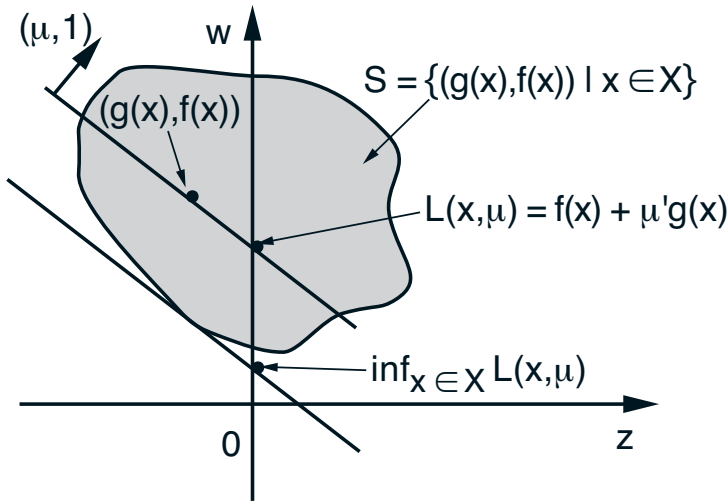
$$f^* = \inf_{x \in X} L(x, \mu^*),$$

where

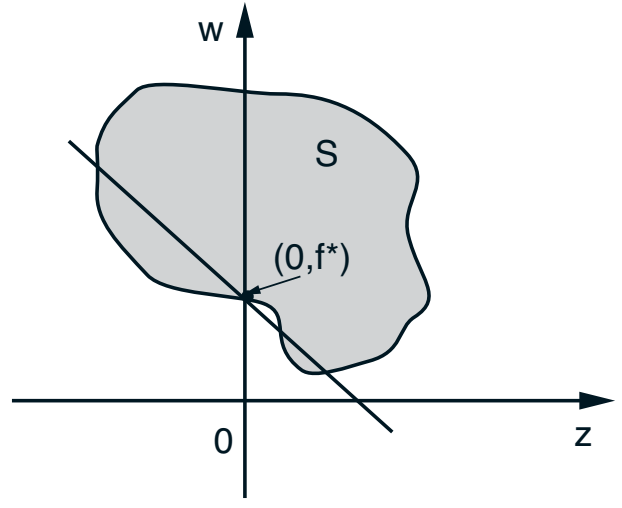
$$L(x, \mu) = f(x) + \mu'g(x)$$

- Note that a G-multiplier resembles a Lagrange multiplier, but is associated with the problem and not with a specific minimum.

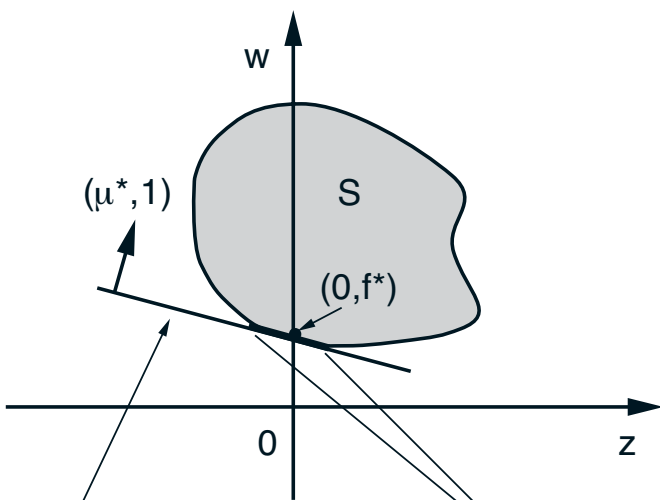
VISUALIZATION



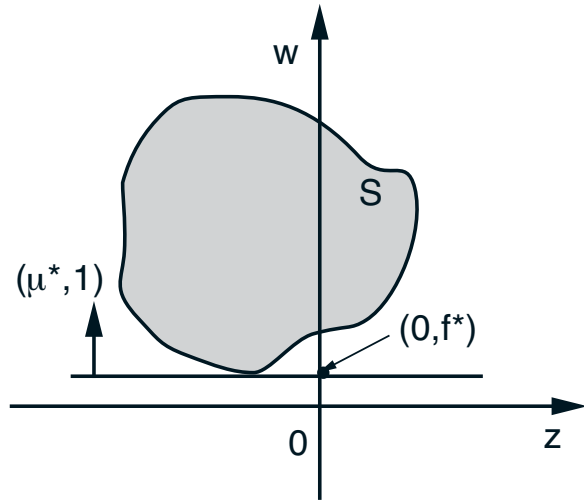
(a)



(b)



(c)



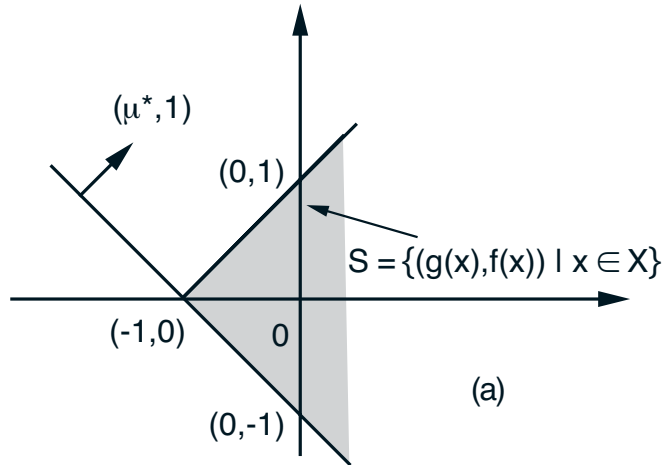
(d)

$$H = \{(z, w) \mid f^* = w + \sum_j \mu_j^* z_j\}$$

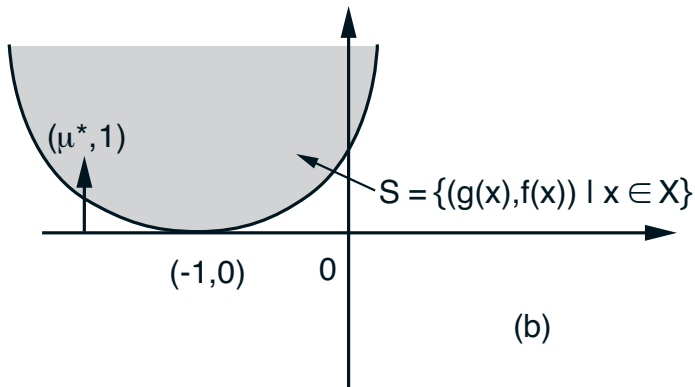
Set of pairs $(g(\bar{x}), f(\bar{x}))$ corresponding to \bar{x} that minimize $L(x, \mu^*)$ over X

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

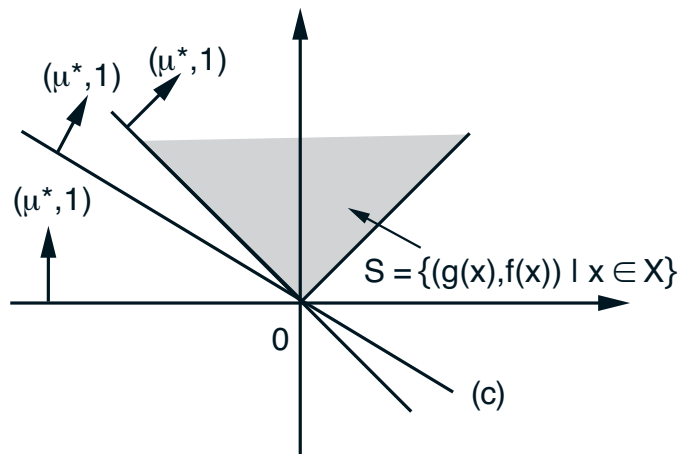
EXAMPLES: A G-MULTIPLIER EXISTS



$$\begin{aligned} \min f(x) &= x_1 - x_2 \\ \text{s.t. } g(x) &= x_1 + x_2 - 1 \leq 0 \\ x \in X &= \{(x_1, x_2) \mid x_1 \geq 0, x_2 \geq 0\} \end{aligned}$$

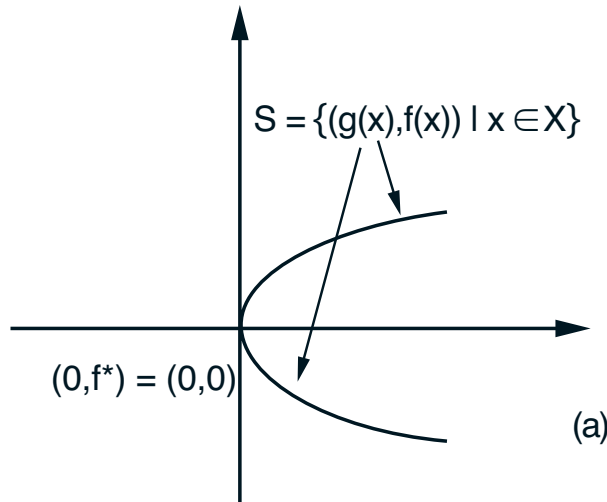


$$\begin{aligned} \min f(x) &= \frac{1}{2}(x_1^2 + x_2^2) \\ \text{s.t. } g(x) &= x_1 - 1 \leq 0 \\ x \in X &= \mathbb{R}^2 \end{aligned}$$

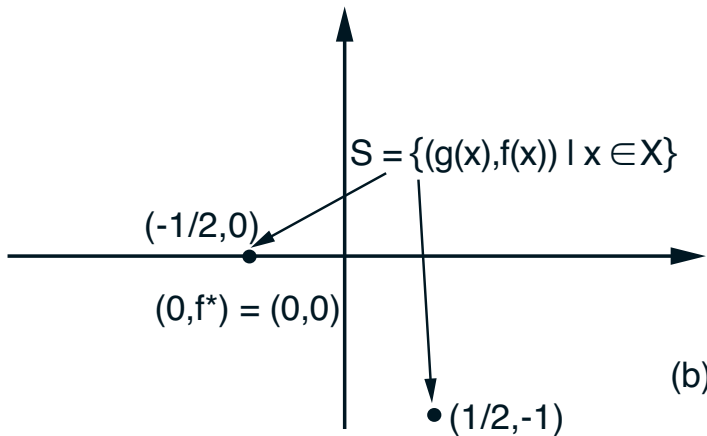


$$\begin{aligned} \min f(x) &= |x_1| + x_2 \\ \text{s.t. } g(x) &= x_1 \leq 0 \\ x \in X &= \{(x_1, x_2) \mid x_2 \geq 0\} \end{aligned}$$

EXAMPLES: A G-MULTIPLIER DOESN'T EXIST



$$\begin{aligned} \min f(x) &= x \\ \text{s.t. } g(x) &= x^2 \leq 0 \\ x \in X &= \mathbb{R} \end{aligned}$$



$$\begin{aligned} \min f(x) &= -x \\ \text{s.t. } g(x) &= x - 1/2 \leq 0 \\ x \in X &= \{0, 1\} \end{aligned}$$

- **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

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

THE DUAL FUNCTION AND THE DUAL PROBLEM

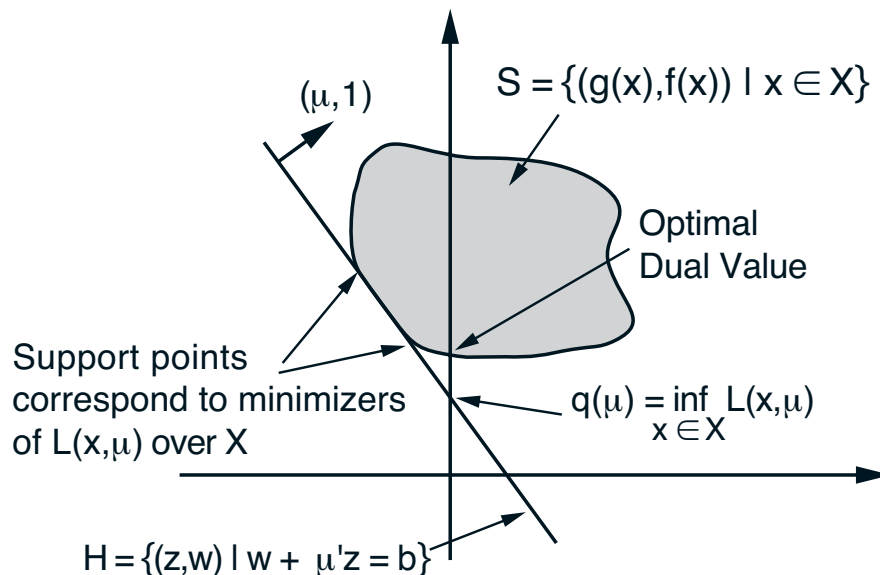
- The *dual problem* is

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

where q is the dual function

$$q(\mu) = \inf_{x \in X} L(x, \mu), \quad \forall \mu \in \mathbb{R}^r$$

- Note: The dual problem is equivalent to a max-crossing problem.



THE DUAL OF A LINEAR PROGRAM

- Consider the linear program

minimize $c'x$

subject to $e'_i x = d_i, \quad i = 1, \dots, m, \quad x \geq 0$

- 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

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

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

THE DUAL OF A QUADRATIC PROGRAM

- Consider the quadratic program

$$\text{minimize } \frac{1}{2}x'Qx + c'x$$

$$\text{subject to } Ax \leq b,$$

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

- Dual function:

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

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

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

- The dual problem, after a sign change, is

$$\text{minimize } \frac{1}{2}\mu'P\mu + t'\mu$$

$$\text{subject to } \mu \geq 0,$$

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

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

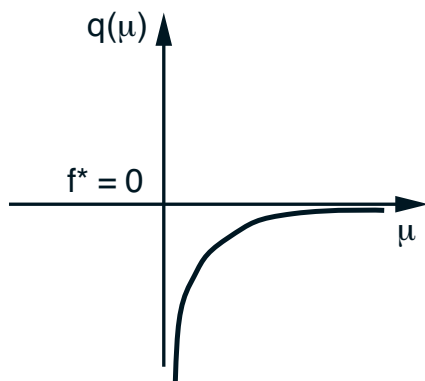
Proposition: (a) If $q^* = f^*$, the set of G-multipliers is equal to the set of optimal dual solutions.

(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 $q^* \leq f^*$,

$$\mu^* \geq 0 \text{ is a G-multiplier} \quad \text{iff} \quad q(\mu^*) = q^* = f^*$$

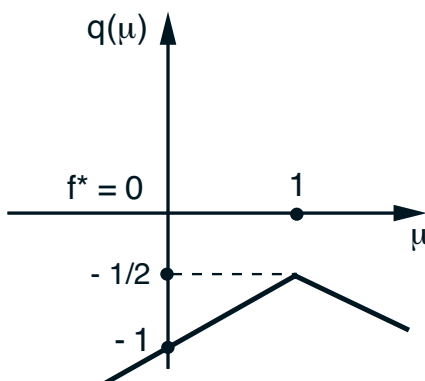
- Examples (dual functions for the two problems with no G-multipliers, given earlier):



$$\begin{aligned} \min f(x) &= x \\ \text{s.t. } g(x) &= x^2 \leq 0 \\ x \in X &= \mathbb{R} \end{aligned}$$

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

(a)



$$\begin{aligned} \min f(x) &= -x \\ \text{s.t. } g(x) &= x - 1/2 \leq 0 \\ x \in X &= \{0, 1\} \end{aligned}$$

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

(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

$$x^* \in X, \quad g(x^*) \leq 0, \quad (\text{Primal Feasibility}),$$

$$\mu^* \geq 0, \quad (\text{Dual Feasibility}),$$

$$x^* = \arg \min_{x \in X} L(x, \mu^*), \quad (\text{Lagrangian Optimality}),$$

$$\mu_j^* g_j(x^*) = 0, \quad j = 1, \dots, r, \quad (\text{Compl. Slackness}).$$

- **Saddle Point Theorem:** (x^*, μ^*) is an optimal solution/G-multiplier pair if and only if $x^* \in X$, $\mu^* \geq 0$, and (x^*, μ^*) is a saddle point of the Lagrangian, in the sense that

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

A CONVEX PROBLEM WITH A DUALITY GAP

- Consider the two-dimensional problem

minimize $f(x)$

subject to $x_1 \leq 0$, $x \in X = \{x \mid x \geq 0\}$,

where

$$f(x) = e^{-\sqrt{x_1 x_2}}, \quad \forall x \in X,$$

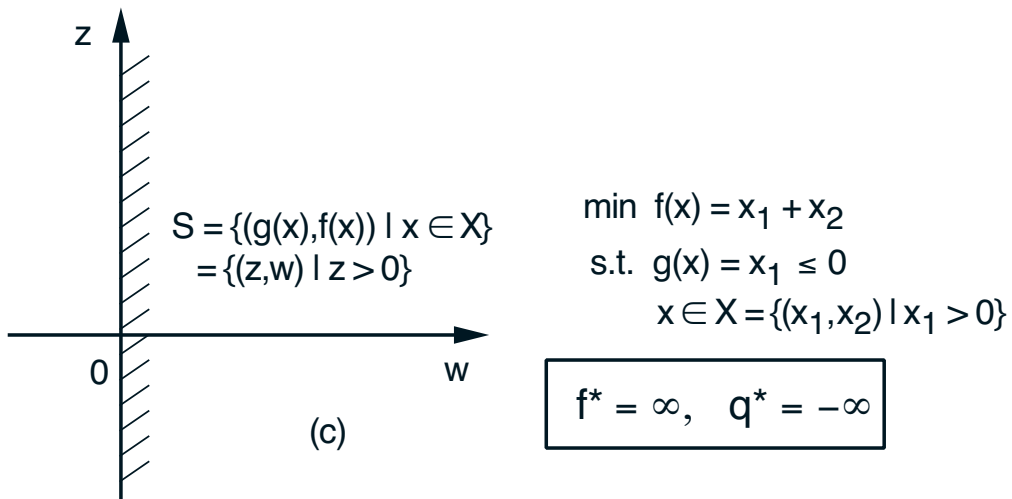
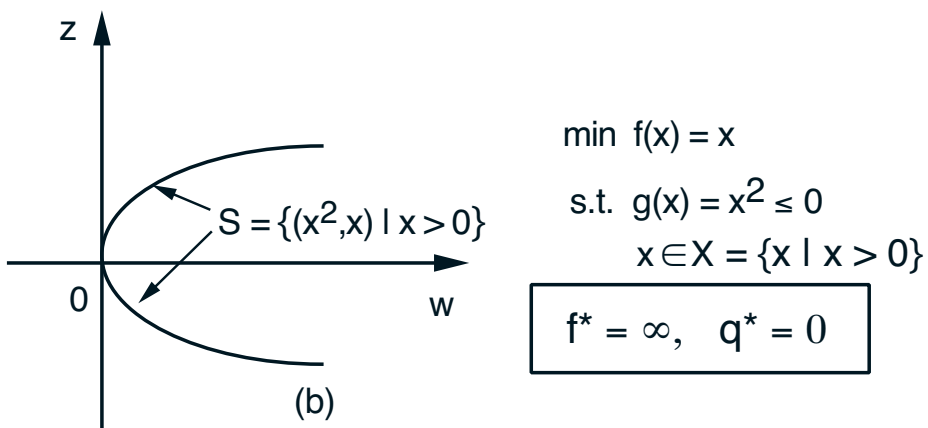
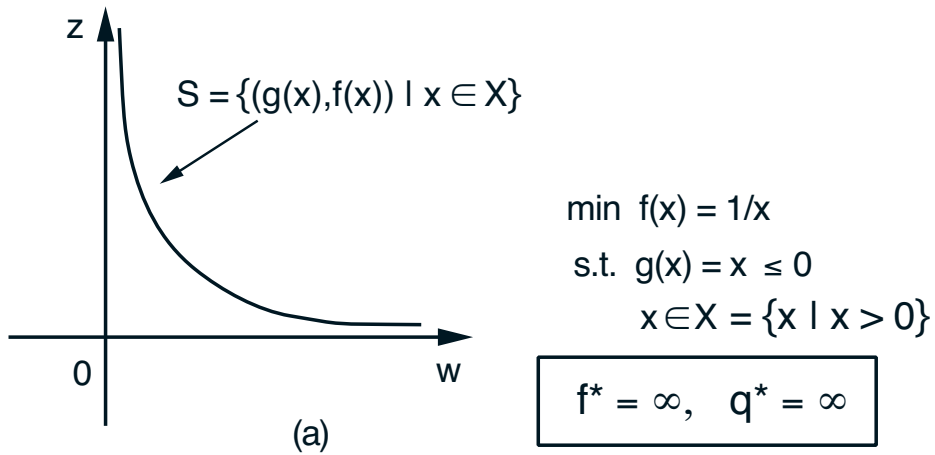
and $f(x)$ is arbitrarily defined for $x \notin X$.

- f is convex over X (its Hessian is positive definite in the interior of X), and $f^* = 1$.
- Also, for all $\mu \geq 0$ we have

$$q(\mu) = \inf_{x \geq 0} \{e^{-\sqrt{x_1 x_2}} + \mu x_1\} = 0,$$

since the expression in braces is nonnegative for $x \geq 0$ and can approach zero by taking $x_1 \rightarrow 0$ and $x_1 x_2 \rightarrow \infty$. It follows that $q^* = 0$.

INFEASIBLE AND UNBOUNDED PROBLEMS



LECTURE 19

LECTURE OUTLINE

- Conditions for existence of geometric multipliers
 - Conditions for strong duality
 - Fritz John optimality conditions
-

- Primal problem: Minimize $f(x)$ subject to $x \in X$, and $g_1(x) \leq 0, \dots, g_r(x) \leq 0$ (assuming $-\infty < f^* < \infty$). It is equivalent to $\inf_{x \in X} \sup_{\mu \geq 0} L(x, \mu)$.
- Dual problem: Maximize $q(\mu)$ subject to $\mu \geq 0$, where $q(\mu) = \inf_{x \in X} L(x, \mu)$. It is equivalent to $\sup_{\mu \geq 0} \inf_{x \in X} L(x, \mu)$.
- μ^* is a geometric multiplier if and only if $f^* = q^*$, and μ^* is an optimal solution of the dual problem.
- Question: Under what conditions $f^* = q^*$ and there exists a geometric multiplier?

RECALL NONLINEAR FARKAS' LEMMA

Let $X \subset \mathfrak{R}^n$ be convex, and $f : X \mapsto \mathfrak{R}$ and $g_j : X \mapsto \mathfrak{R}$, $j = 1, \dots, r$, be convex functions. Assume that

$$f(x) \geq 0, \quad \forall x \in F = \{x \in X \mid g_j(x) \leq 0\},$$

and one of the following two conditions holds:

- (1) There exists $\bar{x} \in X$ such that $g(\bar{x}) < 0$.
- (2) The functions g_j , $j = 1, \dots, r$, are affine, and F contains a relative interior point of X .

Then, there exists a vector $\mu^* = (\mu_1^*, \dots, \mu_r^*) \geq 0$, such that

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

In case (1) the set of such μ^* is also compact.

APPLICATION TO CONVEX PROGRAMMING

Consider the problem

minimize $f(x)$

subject to $x \in X$, $g_j(x) \leq 0$, $j = 1, \dots, r$,

where X , $f : X \mapsto \mathfrak{R}$, and $g_j : X \mapsto \mathfrak{R}$ 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

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

Since $F \subset X$ 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

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

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

$$P = \{x \mid a'_j x - b_j \leq 0, j = r + 1, \dots, p\}$$

Apply Farkas to the extended representation, with

$$F = \{x \in C \mid a'_j x - b_j \leq 0, j = 1, \dots, p\}$$

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 , $j = 1, \dots, \bar{r}$, are convex over C , and the functions g_j , $j = \bar{r} + 1, \dots, r$, are affine.
- (3) There exists a feasible vector \bar{x} such that $g_j(\bar{x}) < 0$ for all $j = 1, \dots, \bar{r}$.
- (4) There exists a vector that satisfies the linear constraints [but not necessarily the constraints $g_j(x) \leq 0$, $j = 1, \dots, \bar{r}$] 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

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

- Properties of p around $u = 0$ are critical in analyzing the presence of a duality gap and the existence of primal and dual optimal solutions.
- p is known as the *primal function* of the constrained optimization problem.
- We have

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

- So

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

and $p(u)$ can be viewed as a *perturbed optimal value* [note that $p(0) = f^*$].

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.

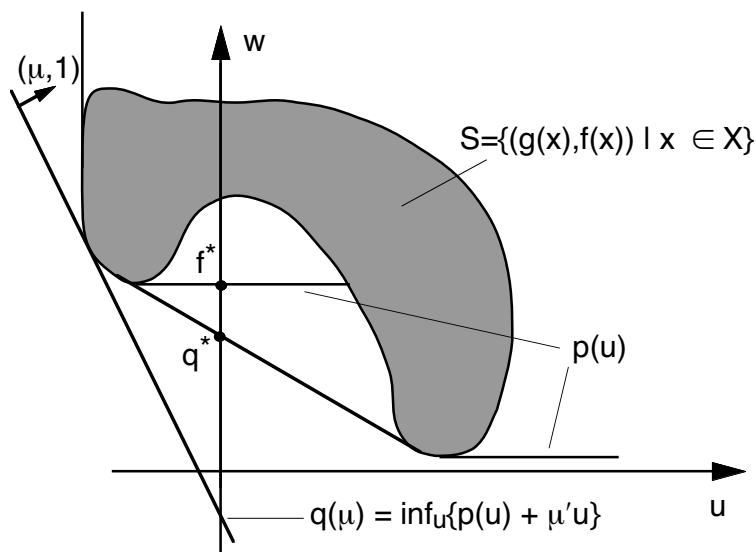
RELATION OF PRIMAL AND DUAL FUNCTIONS

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

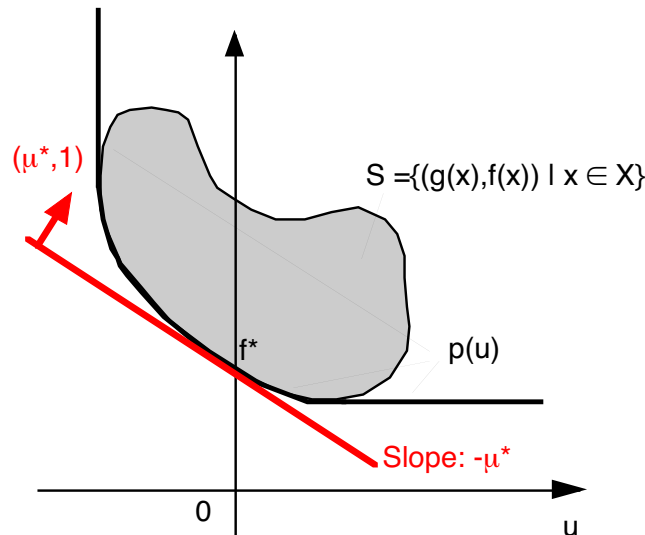
$$\begin{aligned}
 q(\mu) &= \inf_{x \in X} \{f(x) + \mu'g(x)\} \\
 &= \inf_{\{(u,x) | x \in X, g(x) \leq u, j=1, \dots, r\}} \{f(x) + \mu'g(x)\} \\
 &= \inf_{\{(u,x) | x \in X, g(x) \leq u\}} \{f(x) + \mu'u\} \\
 &= \inf_{u \in \mathcal{R}^r} \inf_{x \in X, g(x) \leq u} \{f(x) + \mu'u\}.
 \end{aligned}$$

- Thus

$$q(\mu) = \inf_{u \in \mathcal{R}^r} \{p(u) + \mu'u\}, \quad \forall \mu \geq 0$$



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 $-\partial p(0)$. This follows from the relation

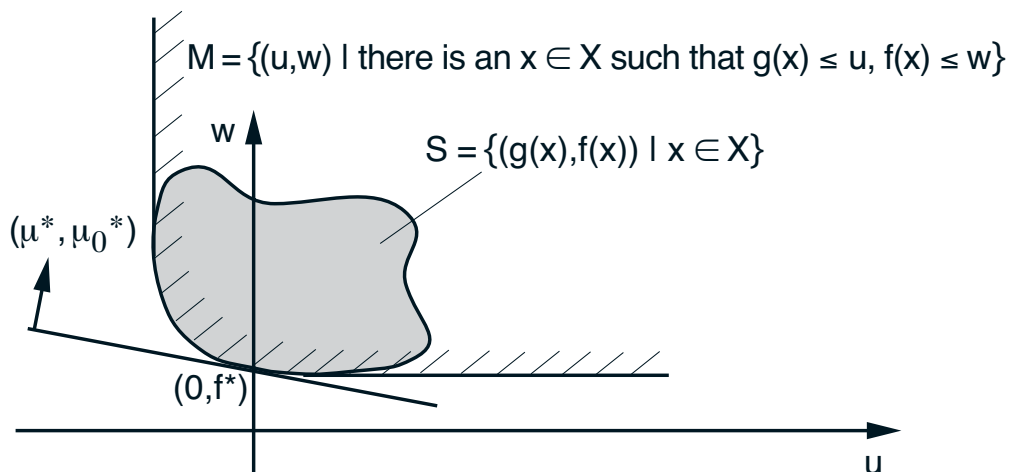
$$q(\mu) = \inf_{u \in \mathbb{R}^r} \{p(u) + \mu'u\}$$

- If p is differentiable at 0, there is a unique G-multiplier: $\mu^* = -\nabla p(0)$.
- If the origin lies in the interior of $\text{dom}(p)$, the set of G-multipliers is nonempty and compact. (This is true iff the Slater condition holds.)

FRITZ JOHN THEORY 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 $\mu^* = (\mu_1^*, \dots, \mu_r^*)$ satisfying the following conditions:

- (i) $\mu_0^* f^* = \inf_{x \in X} \{ \mu_0^* f(x) + \mu^{*'} g(x) \}$.
- (ii) $\mu_j^* \geq 0$ for all $j = 0, 1, \dots, r$.
- (iii) $\mu_0^*, \mu_1^*, \dots, \mu_r^*$ 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 $\bar{x} \in X$ s.t. $g(\bar{x}) < 0$), μ_0^* cannot be 0; if it were, then $0 = \inf_{x \in X} \mu^{*'} g(x)$ for some $\mu^* \geq 0$ with $\mu^* \neq 0$, while we would also have $\mu^{*'} g(\bar{x}) < 0$.

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 $\mu^* = (\mu_1^*, \dots, \mu_r^*)$, satisfying the following conditions:

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

- (ii) $\mu_j^* \geq 0$ for all $j = 0, 1, \dots, r.$

- (iii) $\mu_0^*, \mu_1^*, \dots, \mu_r^*$ 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 $\mu^{*'} g(\tilde{x}) > 0.$

- 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 $\text{ri}(C)$ contains a feasible solution.

- **Conclusion:** The Fritz John theory is sufficiently powerful to show the major constraint qualification theorems for convex programming.

LECTURE 20

LECTURE OUTLINE

- Fenchel Duality
 - Monotropic Programming
-

- We introduce another “standard” framework:

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

where $f_1, f_2 : \mathcal{R}^n \mapsto \mathcal{R}$, and X_1, X_2 are subsets of \mathcal{R}^n .

- While it can be shown to be equivalent to the Lagrangian framework

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

it is more convenient for some types of applications, including network flow, and conic/semidefinite programming.

FENCHEL DUALITY FRAMEWORK

- Consider the problem

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

where $f_1, f_2 : \mathfrak{R}^n \mapsto \mathfrak{R}$, and X_1, X_2 are subsets of \mathfrak{R}^n .

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

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

and dualize the constraint $z = y$:

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

PRIMAL FENCHEL DUALITY THEOREM

- We view f_1 and $-f_2$ as extended real-valued with domains X_1 and X_2 , and write the primal and dual problems as

$$\min_{x \in \mathbb{R}^n} \{ f_1(x) - f_2(x) \}, \quad \max_{\lambda \in \mathbb{R}^n} \{ g_2(\lambda) - g_1(\lambda) \}$$

- Use strong duality theorems for the problem

$$\min_{z=y, y \in X_1, z \in X_2} \{ f_1(y) - f_2(z) \}$$

- **Primal Fenchel Duality Theorem:** The dual problem has an optimal solution and we have

$$\inf_{x \in \mathbb{R}^n} \{ f_1(x) - f_2(x) \} = \max_{\lambda \in \mathbb{R}^n} \{ g_2(\lambda) - g_1(\lambda) \},$$

if f_1 , $-f_2$, X_1 , X_2 are convex, and *one* of the following two conditions holds:

- The relative interiors of X_1 and X_2 intersect
- X_1 and X_2 are polyhedral, and f_1 and f_2 can be extended to real-valued convex and concave functions over \mathbb{R}^n .

OPTIMALITY CONDITIONS

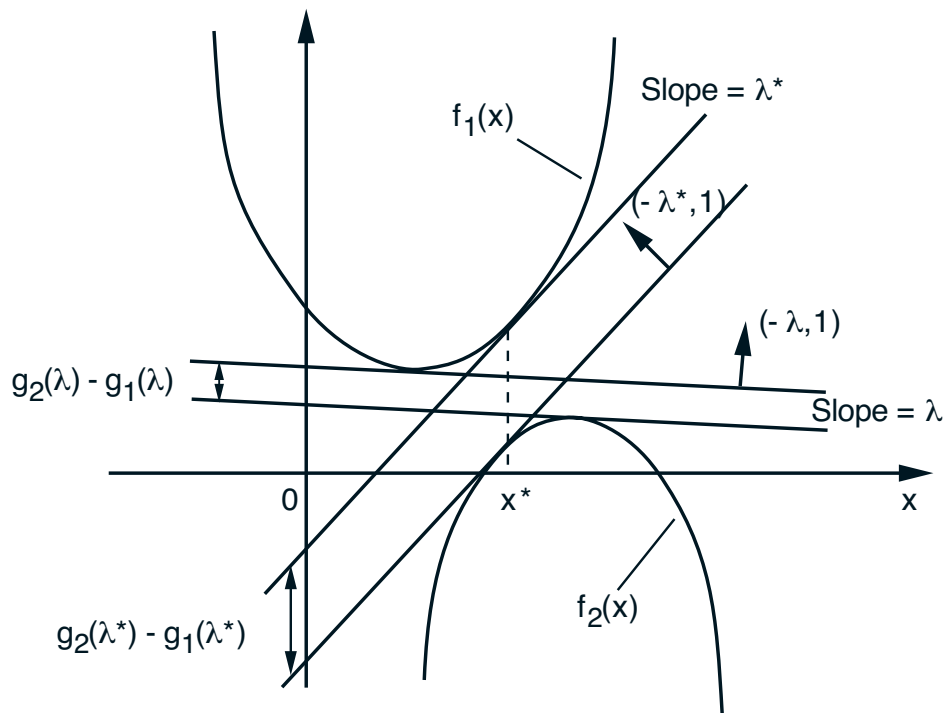
- There is no duality gap, while (x^*, λ^*) is an optimal primal and dual solution pair, if and only if

$$x^* \in \text{dom}(f_1) \cap \text{dom}(-f_2), \quad (\text{primal feasibility}),$$

$$\lambda^* \in \text{dom}(g_1) \cap \text{dom}(-g_2), \quad (\text{dual feasibility}),$$

$$x^* \in \arg \max_{y \in \mathbb{R}^n} \{y' \lambda^* - f_1(y)\}$$

$$x^* \in \arg \min_{z \in \mathbb{R}^n} \{z' \lambda^* - f_2(z)\}, \quad (\text{Lagr. optimality}).$$



- Note: The Lagrangian optimality condition is equivalent to $\lambda^* \in \partial f_1(x^*) \cap \partial f_2(x^*)$.

DUAL FENCHEL DUALITY THEOREM

- The dual problem

$$\max_{\lambda \in \mathfrak{R}^n} \{g_2(\lambda) - g_1(\lambda)\}$$

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 and we have

$$\min_{x \in \mathfrak{R}^n} \{f_1(x) - f_2(x)\} = \sup_{\lambda \in \mathfrak{R}^n} \{g_2(\lambda) - g_1(\lambda)\},$$

if *one* of the following two conditions holds

- The relative interiors of $\text{dom}(g_1)$ and $\text{dom}(-g_2)$ intersect
- $\text{dom}(g_1)$ and $\text{dom}(-g_2)$ are polyhedral, and g_1 and g_2 can be extended to real-valued convex and concave functions over \mathfrak{R}^n .

MONOTROPIC PROGRAMMING

$$\min_{x \in S} \sum_{i=1}^n f_i(x_i)$$

where $x = (x_1, \dots, x_n) \in \mathfrak{R}^n$ and

$f_i : \mathfrak{R} \mapsto (-\infty, \infty]$ are proper convex

S is a subspace

- Special cases: Linear and quadratic programs
- Duality framework:

$$\min_{x \in S, z=x} \sum_{i=1}^n f_i(z_i)$$

Assign a multiplier $\lambda \in \mathfrak{R}^n$ to $z = x$, and obtain the dual function

$$\begin{aligned} q(\lambda) &= \inf_{x \in S} \lambda'x + \sum_{i=1}^n \inf_{z_i \in \mathfrak{R}} \{ f_i(z_i) - \lambda_i z_i \} \\ &= \begin{cases} \sum_{i=1}^n q_i(\lambda_i) & \text{if } \lambda \in S^\perp, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where $q_i(\lambda_i) = \inf_{z_i \in \mathfrak{R}} \{ f_i(z_i) - \lambda_i z_i \}$.

OPTIMALITY CONDITIONS

- A pair (x, λ) satisfies the Lagrangian optimality condition, if and only if x_i attains the infimum in the equation

$$q_i(\lambda_i) = \inf_{z_i \in \mathfrak{R}} \{ f_i(z_i) - \lambda_i z_i \}, \quad i = 1, \dots, n,$$

- Equivalently, if and only if (x, λ) lies in the set

$$\Gamma = \{ (x, \lambda) \mid x_i \in \text{dom}(f_i), f_i^-(x_i) \leq \lambda_i \leq f_i^+(x_i), \forall i \}$$

- Assume that $-\infty < f^* < \infty$. The vectors x^* and λ^* are optimal primal and dual solutions, and there is no duality gap if and only if

$$x^* \in S, \quad \lambda^* \in S^\perp, \quad (x^*, \lambda^*) \in \Gamma$$

- In specific contexts (e.g., electric networks), these conditions can be interpreted as physical laws (e.g., Kirchhoff's and Ohm's laws).

EXTENDED MONOTROPIC PROGRAMMING

- Additive but nonseparable version:

$$\min_{x \in S} \sum_{i=1}^m f_i(x)$$

where $f_i : \mathfrak{R} \mapsto (-\infty, \infty]$ are proper convex and S is a subspace.

- Duality framework:

$$\min_{x \in S, z_i = x, i=1, \dots, m} \sum_{i=1}^m f_i(z_i)$$

Assign a multiplier $\lambda_i \in \mathfrak{R}^n$ to each constraint $z_i = x$, and obtain the dual function

$$\begin{aligned} q(\lambda) &= \inf_{x \in S} \left(\sum_{i=1}^m \lambda_i \right)' x + \sum_{i=1}^m \inf_{z_i \in \mathfrak{R}^n} \{ f_i(z_i) - \lambda_i' z_i \} \\ &= \begin{cases} \sum_{i=1}^m q_i(\lambda_i) & \text{if } \sum_{i=1}^m \lambda_i \in S^\perp, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where $q_i(\lambda_i) = \inf_{z_i \in \mathfrak{R}^n} \{ f_i(z_i) - \lambda_i' z_i \}$.

DUALITY THEORY FOR EXTENDED M.P.

- **Main Duality Theorem:** Assume that the problem is feasible, and that for all feasible x ,

$$f_i(x) = (\text{cl } f_i)(x), \quad i = 1, \dots, m,$$

and the set

$$T(x, \epsilon) = S^\perp + \partial_\epsilon f_1(x) + \dots + \partial_\epsilon f_m(x)$$

is closed for all $\epsilon > 0$. Then $q^* = f^*$.

- The proof is unusual. We apply the ϵ -descent method with outer approximation of the ϵ -subdifferential. Assuming $f^* > -\infty$, the method terminates with x such that $0 \in T(x, \epsilon)$, and $\lambda_i \in \partial_\epsilon f_i(x)$ such that

$$\lambda_1 + \dots + \lambda_m \in S^\perp.$$

We argue that

$$f(x) \leq q(\lambda_1, \dots, \lambda_m) + m\epsilon$$

from which we obtain $f^* \leq q^* + m\epsilon$.

DUALITY THEOREMS

- Assume the functions f_i are closed proper convex.
- **Primal Theorem:** Assume that the primal problem is feasible, and that each f_i is real-valued, or is polyhedral, or is one-dimensional. Then $q^* = f^*$.
- **Monotropic Programming Duality Theorem:** Consider the separable special case (i.e., the monotropic programming problem). Assume that either the primal or the dual problem is feasible. Then $q^* = f^*$.
- Slight extensions are possible, where each f_i is assumed to satisfy

$$f_i(x) = (\text{cl } f_i)(x),$$

for all feasible x , instead of being closed, etc.

LECTURE 21

LECTURE OUTLINE

- Conic Programming
 - Quadratic-Conic Programming
 - Semidefinite Programming
-

- A conic problem is to minimize a convex function subject to a cone constraint.
- Can be analyzed as a special case of Fenchel duality.
- There are many interesting applications of conic problems, including in discrete optimization.

CONIC DUALITY I

- Consider the problem

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

where C is a convex cone, and $f : \mathbb{R}^n \mapsto (-\infty, \infty]$ is convex.

- Apply Fenchel duality with the definitions

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

We have

$$g_1(\lambda) = \sup_{x \in \mathbb{R}^n} \{ \lambda'x - f(x) \}, \quad g_2(\lambda) = \inf_{x \in C} x'\lambda = \begin{cases} 0 & \text{if } \lambda \in \hat{C}, \\ -\infty & \text{if } \lambda \notin \hat{C}, \end{cases}$$

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

(b) f can be extended to a real-valued convex function over \mathfrak{R}^n , and $\text{dom}(f)$ 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) $\text{ri}(\text{dom}(g)) \cap \text{ri}(\hat{C}) \neq \emptyset$.

(b) g can be extended to a real-valued convex function over \mathfrak{R}^n , and $\text{dom}(g)$ and \hat{C} are polyhedral.

LINEAR-CONIC PROBLEMS

- Let f be affine, $f(x) = c'x$, with $\text{dom}(f)$ being an affine set, $\text{dom}(f) = b + S$, where S is a subspace.
- The primal problem is

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

- 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

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

- The primal and dual have the same form.
- If C is closed, the dual of the dual yields the primal.

CONIC-QUADRATIC PROGRAMMING

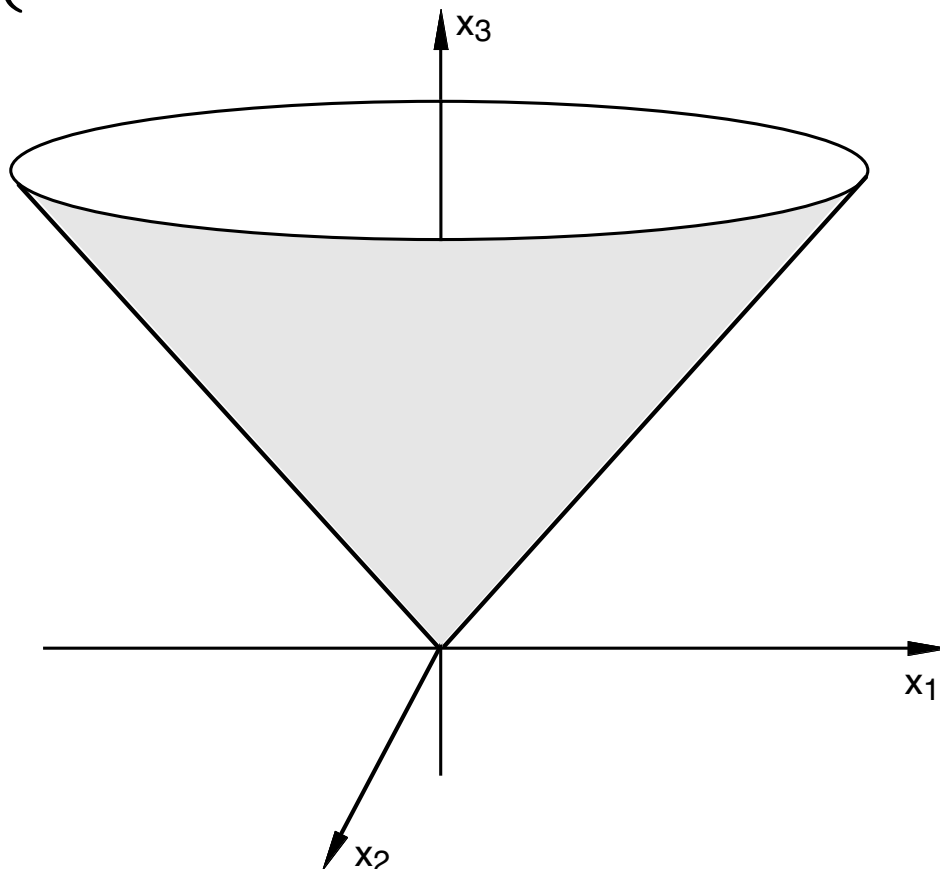
- Conic-quadratic programming is the linear-conic problem

$$\text{minimize } c'x$$

$$\text{subject to } A_i x - b_i \in C_i, \quad i = 1, \dots, m,$$

where c, b_i are vectors, A_i are matrices, b_i is a vector in \mathbb{R}^{n_i} , and C_i is the second order cone

$$C = \left\{ (x_1, \dots, x_n) \mid x_n \geq \sqrt{x_1^2 + \dots + x_{n-1}^2} \right\}$$



CONIC-QUADRATIC DUALITY

- The dual of the quadratic-conic problem (viewed as a special case of a linear-conic problem) is (after some manipulation)

$$\begin{aligned} &\text{maximize} && \sum_{i=1}^m b'_i \lambda_i \\ &\text{subject to} && \sum_{i=1}^m A'_i \lambda_i = c, \quad \lambda_i \in C_i, \quad i = 1, \dots, m, \end{aligned}$$

where $\lambda = (\lambda_1, \dots, \lambda_m)$.

- The duality theory is derived from (and is no more favorable than) the one for linear-conic problems.
- There is no duality gap if there exists a feasible solution in the interior of the 2nd order cones C_i .
- Generally, conic-quadratic problems can be recognized from the presence of norm or convex quadratic functions in the cost or the constraint functions.
- There are many applications.

SEMIDEFINITE PROGRAMMING

- Consider the symmetric $n \times n$ matrices. Inner product $\langle X, Y \rangle = \text{trace}(XY) = \sum_{i,j=1}^n x_{ij}y_{ij}$.
- Let D be the cone of pos. semidefinite matrices. Note that D is self-dual [$D = \hat{D}$, i.e., $\langle X, Y \rangle \geq 0$ 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, \dots, A_m , and vectors b_1, \dots, b_m , and consider

minimize $\langle C, X \rangle$

subject to $\langle A_i, X \rangle = b_i, \quad i = 1, \dots, m, \quad X \in D$

- Viewing this as an affine cost conic problem, the dual problem (after some manipulation) is

maximize $\sum_{i=1}^m b_i \lambda_i$

subject to $C - (\lambda_1 A_1 + \dots + \lambda_m A_m) \in D$

- There is no duality gap if there exists $\bar{\lambda}$ such that $C - (\bar{\lambda}_1 A_1 + \dots + \bar{\lambda}_m A_m)$ is pos. definite.

LECTURE 22

LECTURE OUTLINE

- Overview of Dual Methods
- Nondifferentiable Optimization

- Consider the primal problem

minimize $f(x)$

subject to $x \in X$, $g_j(x) \leq 0$, $j = 1, \dots, r$,

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.

DUAL DERIVATIVES

- Let

$$x_\mu \in \arg \min_{x \in X} L(x, \mu) = \arg \min_{x \in X} \{f(x) + \mu'g(x)\}$$

Then for all $\nu \in \mathbb{R}^r$,

$$\begin{aligned} q(\nu) &= \inf_{x \in X} \{f(x) + \nu'g(x)\} \\ &\leq f(x_\mu) + \nu'g(x_\mu) \\ &= f(x_\mu) + \mu'g(x_\mu) + (\nu - \mu)'g(x_\mu) \\ &= q(\mu) + (\nu - \mu)'g(x_\mu). \end{aligned}$$

- 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

$$\nabla q(\mu) = g(x_\mu), \quad \forall \mu \in \mathbb{R}^r$$

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,

$$q(\mu) = \min_{i \in I} \{a'_i \mu + b_i\},$$

where I is a finite index set, and $a_i \in \mathbb{R}^r$ 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

$$I_\mu = \{i \in I \mid a'_i \mu + b_i = q(\mu)\}$$

The set of all subgradients of q at μ is

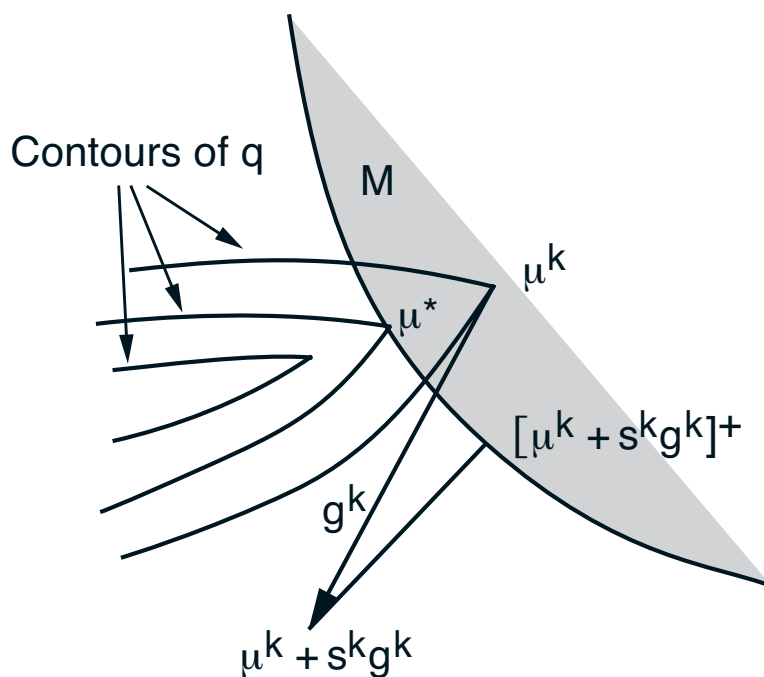
$$\partial q(\mu) = \text{conv}(\{a_i \mid i \in I_\mu\})$$

NONDIFFERENTIABLE OPTIMIZATION

- Consider maximization of $q(\mu)$ over $M = \{\mu \geq 0 \mid q(\mu) > -\infty\}$
- Subgradient method:

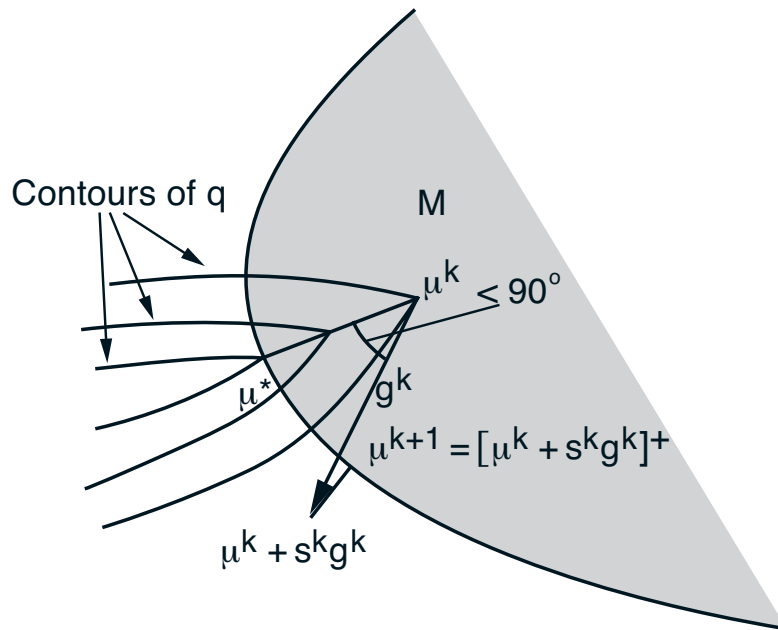
$$\mu^{k+1} = [\mu^k + s^k g^k]^+,$$

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

$$0 < s^k < \frac{2(q(\mu^*) - q(\mu^k))}{\|g^k\|^2}$$

CONVERGENCE MECHANISM

- Assume constant stepsize: $s^k \equiv s$
- If $\|g^k\| \leq C$ for some constant C and all k ,

$$\|\mu^{k+1} - \mu^*\|^2 \leq \|\mu^k - \mu^*\|^2 - 2s(q(\mu^*) - q(\mu^k)) + s^2 C^2,$$

so the distance to the optimum decreases if

$$0 < s < \frac{2(q(\mu^*) - q(\mu^k))}{C^2}$$

or equivalently, if μ^k belongs to the level set

$$\left\{ \mu \mid q(\mu) < q(\mu^*) - \frac{sC^2}{2} \right\}$$

- With a little further analysis, it can be shown that the method, at least asymptotically, reaches this level set, i.e.

$$\limsup_{k \rightarrow \infty} q(\mu^k) \geq q(\mu^*) - \frac{sC^2}{2}$$

CONVERGENCE ANALYSIS

- Consider a generic convex problem

$$\min_{x \in X} f(x)$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ 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 .

- **Assumption:** (Subgradient Boundedness)

$$\|g\| \leq C, \quad \forall g \in \partial f(x_k), \quad \forall i, k,$$

for some scalar C . (Satisfied when f is polyhedral as in integer programming.)

- **Key Lemma:** For all $y \in X$ and k ,

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

STEP SIZE RULES

- Constant Stepsize $\alpha_k \equiv \alpha$:
 - By key lemma with $f(y) \approx f^*$, it makes progress to the optimal if $0 < \alpha < \frac{2(f(x_k) - f^*)}{C^2}$, i.e., if

$$f(x_k) > f^* + \frac{\alpha C^2}{2}$$

- Diminishing Stepsize $\alpha_k \rightarrow 0, \sum_k \alpha_k = \infty$:
 - Eventually makes progress (once α_k becomes small enough). Can show that

$$\liminf_{k \rightarrow \infty} f(x_k) = f^*$$

- 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 $\{x \mid f(x) \leq f_k\}$.
- f_k can be adjusted based on the progress of the method.

LECTURE 23

LECTURE OUTLINE

- Additional Dual Methods
- Cutting Plane Methods
- Decomposition

- Consider the primal problem

minimize $f(x)$

subject to $x \in X$, $g_j(x) \leq 0$, $j = 1, \dots, r$,

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 $\mu \in M = \{\mu \mid \mu \geq 0, q(\mu) > -\infty\}$.

CUTTING PLANE METHOD

- k th iteration, after μ^i and $g^i = g(x_{\mu^i})$ have been generated for $i = 0, \dots, k - 1$: Solve

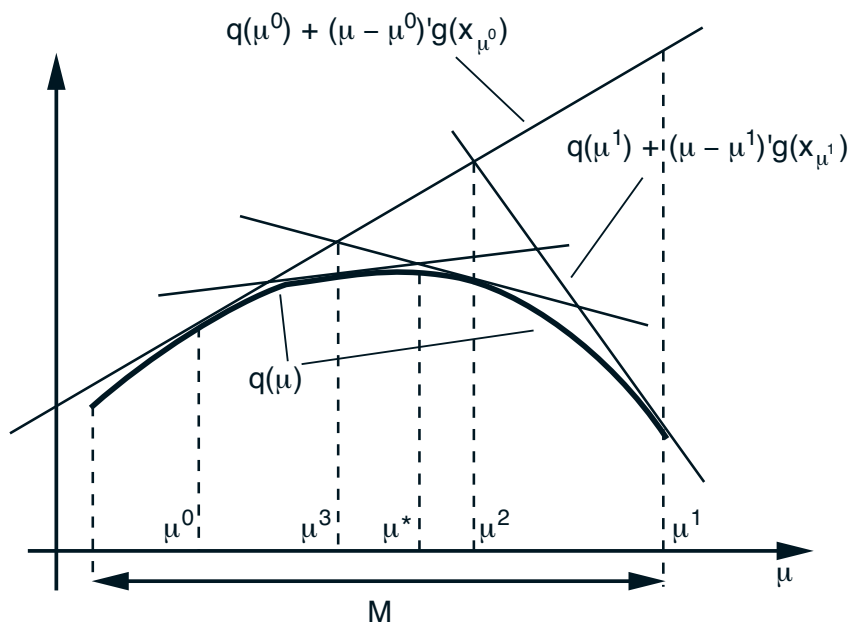
$$\max_{\mu \in M} Q^k(\mu)$$

where

$$Q^k(\mu) = \min_{i=0, \dots, k-1} \{q(\mu^i) + (\mu - \mu^i)'g^i\}$$

Set

$$\mu^k = \arg \max_{\mu \in M} Q^k(\mu)$$

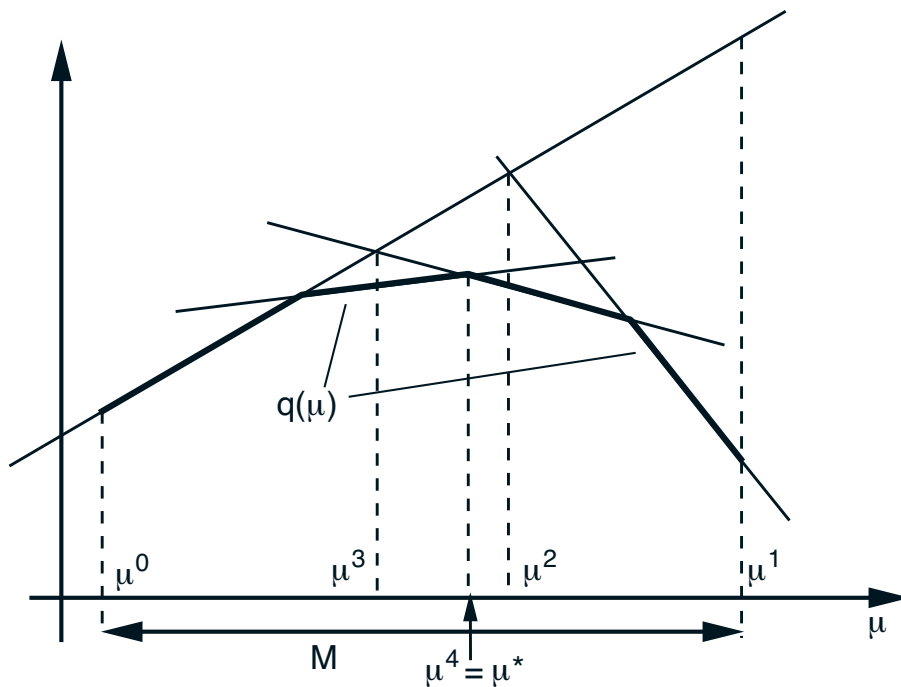


POLYHEDRAL CASE

$$q(\mu) = \min_{i \in I} \{a'_i \mu + b_i\}$$

where I is a finite index set, and $a_i \in \mathbb{R}^r$ 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 max 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

$$q(\mu^i) + (\mu - \mu^i)'g^i \geq q(\mu), \quad \forall \mu \in M,$$

$$Q^k(\mu^k) \geq Q^k(\mu) \geq q(\mu), \quad \forall \mu \in M. \quad (1)$$

• 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 $i \rightarrow \infty$, $k \rightarrow \infty$, $i \in K$, $k \in K$,

$$\lim_{k \rightarrow \infty, k \in K} Q^k(\mu^k) = q(\bar{\mu})$$

Combining with (1), $q(\bar{\mu}) = \max_{\mu \in M} q(\mu)$.

LAGRANGIAN RELAXATION

- Solving the dual of the separable problem

$$\text{minimize } \sum_{j=1}^J f_j(x_j)$$

$$\text{subject to } x_j \in X_j, \quad j = 1, \dots, J, \quad \sum_{j=1}^J A_j x_j = b.$$

- Dual function is

$$\begin{aligned} q(\lambda) &= \sum_{j=1}^J \min_{x_j \in X_j} \{ f_j(x_j) + \lambda' A_j x_j \} - \lambda' b \\ &= \sum_{j=1}^J \{ f_j(x_j(\lambda)) + \lambda' A_j x_j(\lambda) \} - \lambda' b \end{aligned}$$

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 cutting plane applied to the dual problem $\max_{\lambda} q(\lambda)$.
- At the k th iteration, we solve the “approximate dual”

$$\lambda^k = \arg \max_{\lambda \in \mathbb{R}^r} Q^k(\lambda) \equiv \min_{i=0, \dots, k-1} \{q(\lambda^i) + (\lambda - \lambda^i)' g^i\}$$

- Equivalent linear program in v and λ

maximize v

subject to $v \leq q(\lambda^i) + (\lambda - \lambda^i)' g^i, \quad i = 0, \dots, k - 1$

The dual of this (called *master problem*) is

$$\text{minimize} \quad \sum_{i=0}^{k-1} \xi^i (q(\lambda^i) - \lambda^{i'} g^i)$$

$$\text{subject to} \quad \sum_{i=0}^{k-1} \xi^i = 1, \quad \sum_{i=0}^{k-1} \xi^i g^i = 0,$$

$$\xi^i \geq 0, \quad i = 0, \dots, k - 1,$$

DANTSIG-WOLFE DECOMPOSITION (CONT.)

- The master problem is written as

$$\text{minimize } \sum_{j=1}^J \left(\sum_{i=0}^{k-1} \xi^i f_j(x_j(\lambda^i)) \right)$$

$$\text{subject to } \sum_{i=0}^{k-1} \xi^i = 1, \quad \sum_{j=1}^J A_j \left(\sum_{i=0}^{k-1} \xi^i x_j(\lambda^i) \right) = b,$$

$$\xi^i \geq 0, \quad i = 0, \dots, k-1.$$

- The primal cost function terms $f_j(x_j)$ are approximated by

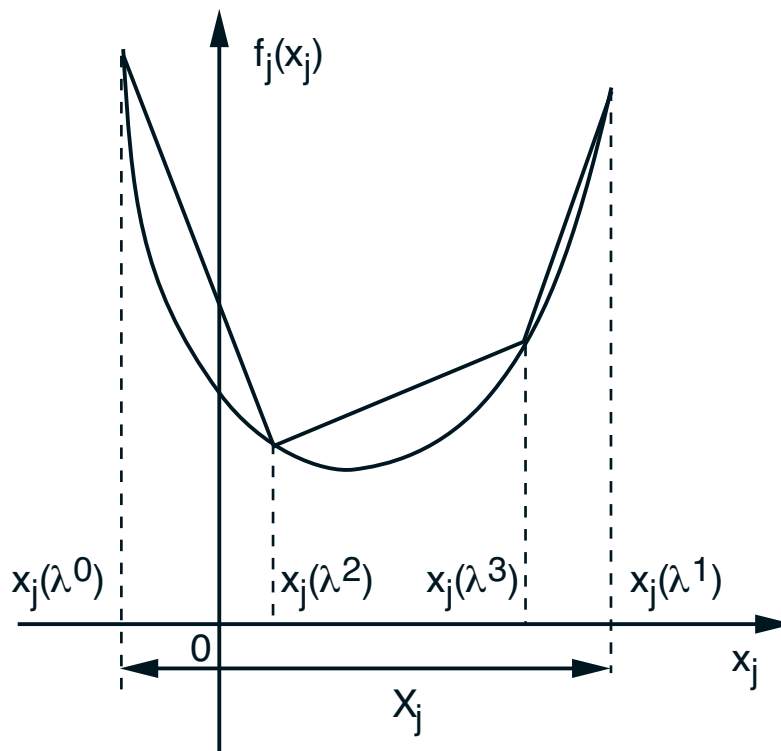
$$\sum_{i=0}^{k-1} \xi^i f_j(x_j(\lambda^i))$$

- 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*.

LECTURE 24

LECTURE OUTLINE

- We start a 2-lecture series on nonconvex (but smooth) optimization
 - Conical approximations
 - Conditions for optimality
-

- A basic necessary condition:
 - If x^* minimizes a function $f(x)$ over $x \in X$, then for every $y \in \mathfrak{R}^n$, $\alpha^* = 0$ 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):
 - $X = \mathfrak{R}^n$: $\nabla f(x^*) = 0$.
 - 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 \mathbb{R}^n and a vector $x \in X$.
- A vector $y \in \mathbb{R}^n$ is a *feasible direction* of X at x if there exists an $\bar{\alpha} > 0$ such that $x + \alpha y \in X$ for all $\alpha \in [0, \bar{\alpha}]$.
- 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 $\alpha(\bar{x} - x)$ with $\alpha > 0$ and $\bar{x} \in X$.
- **Easy optimality condition:** If x^* minimizes a differentiable function $f(x)$ over $x \in X$, then

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

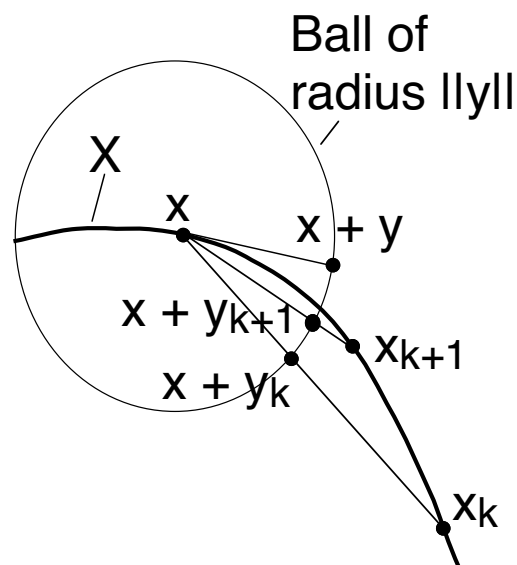
- **Difficulty:** The condition may be vacuous because there may be no feasible directions (other than 0), e.g., take X to be the boundary of a circle.

TANGENT CONE

- Consider a subset X of \mathbb{R}^n and a vector $x \in X$.
- A vector $y \in \mathbb{R}^n$ 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

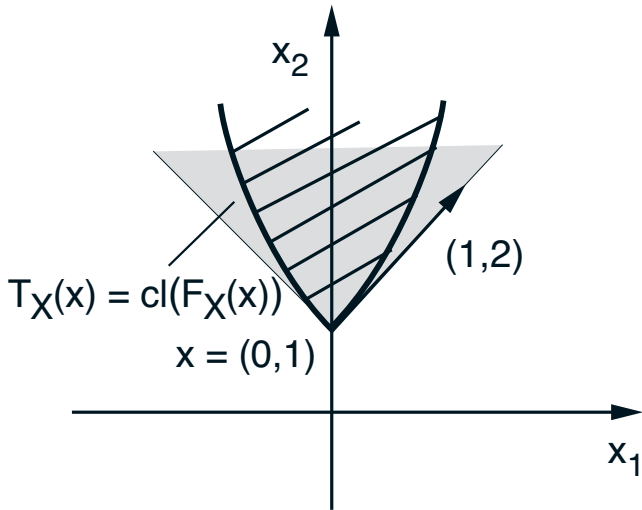
$$x_k \rightarrow x, \quad \frac{x_k - x}{\|x_k - x\|} \rightarrow \frac{y}{\|y\|}$$

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

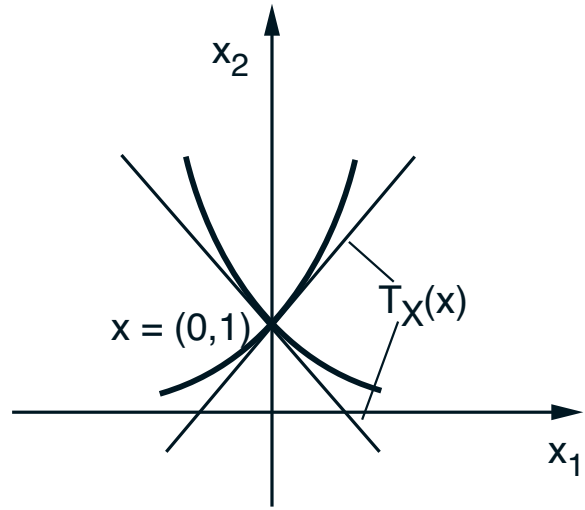


- y is a tangent of X at x iff there exists $\{x_k\} \subset X$ with $x_k \rightarrow x$, and a positive scalar sequence $\{\alpha_k\}$ such that $\alpha_k \rightarrow 0$ and $(x_k - x)/\alpha_k \rightarrow y$.

EXAMPLES



(a)



(b)

- 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) $\text{cl}(F_X(x)) \subset T_X(x)$.
 - (c) If X is convex, then $F_X(x)$ and $T_X(x)$ are convex, and we have

$$\text{cl}(F_X(x)) = T_X(x)$$

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.

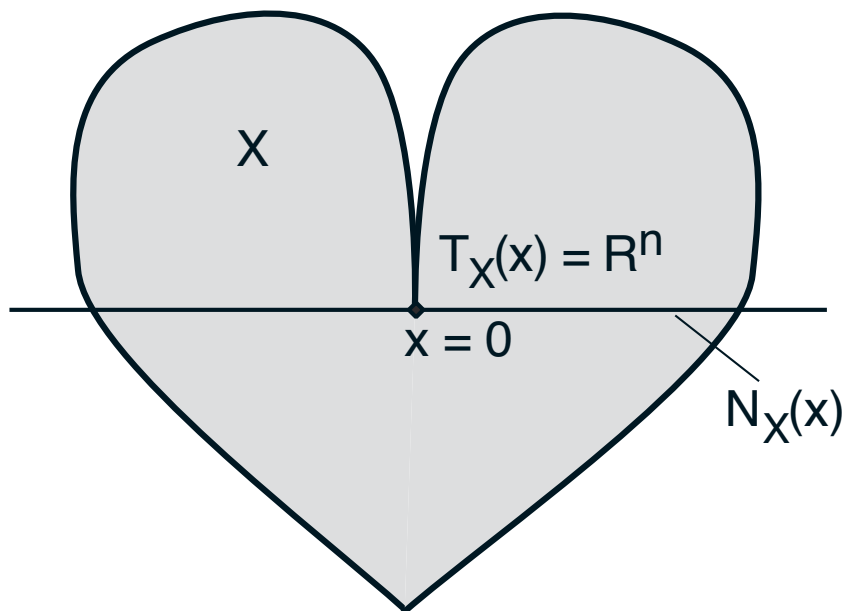
(c) Since X is convex, the set $F_X(x)$ consists of the vectors of the form $\alpha(\bar{x} - x)$ with $\alpha > 0$ and $\bar{x} \in X$. Verify definition of convexity ...

NORMAL CONE

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

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

- 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

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

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

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

- 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.

A BASIC OPTIMALITY CONDITION

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

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

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

CONSTRAINTS - LAGRANGE MULTIPLIERS

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

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

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

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

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

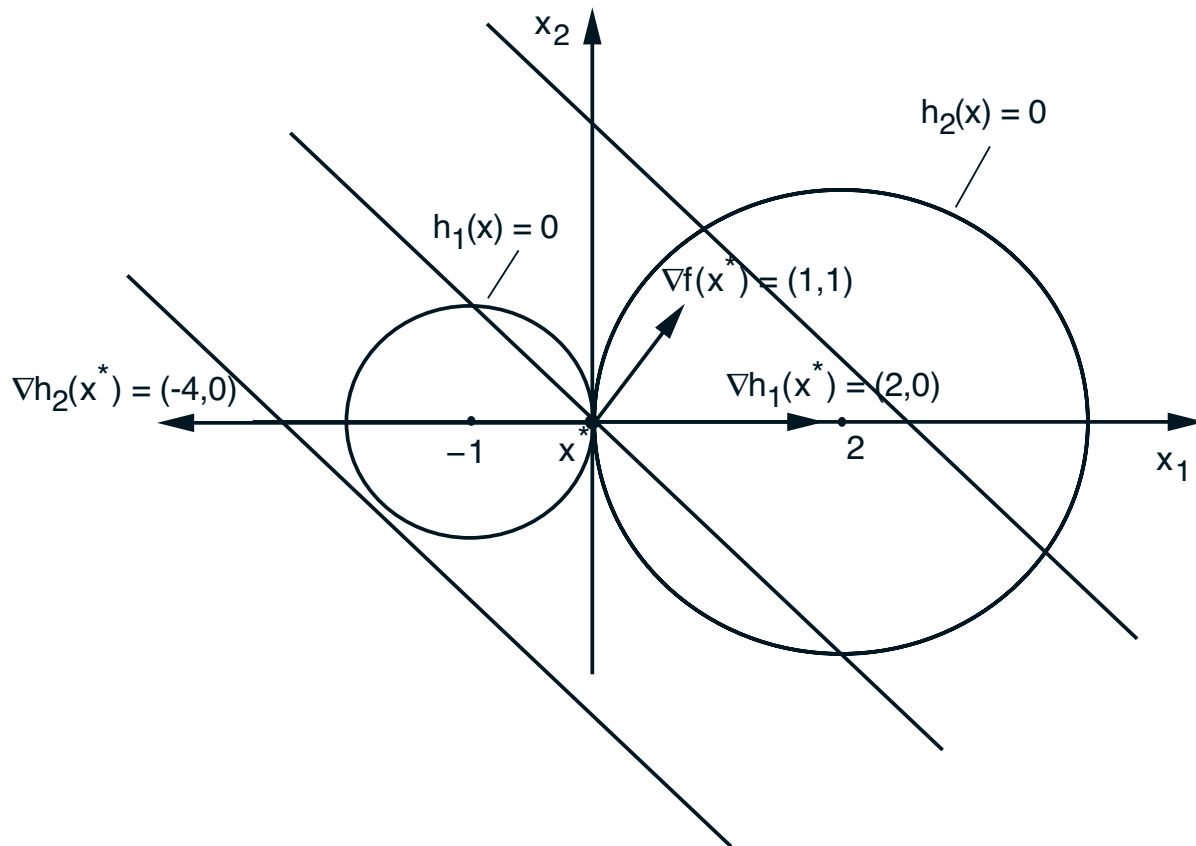
where

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

- **Note:** When $X = \mathfrak{R}^n$, then $T_X(x^*) = \mathfrak{R}^n$ and

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

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

$$h_i(x) = a_i'x - b_i, \quad i = 1, \dots, m,$$

- The tangent cone is

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

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

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

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.,

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

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

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

LECTURE 25

LECTURE OUTLINE

- Enhanced Fritz John Conditions
 - Pseudonormality
 - Constraint qualifications
-

- Problem

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

where $f, h_i, g_j : \mathbb{R}^n \mapsto \mathbb{R}$ 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 ,

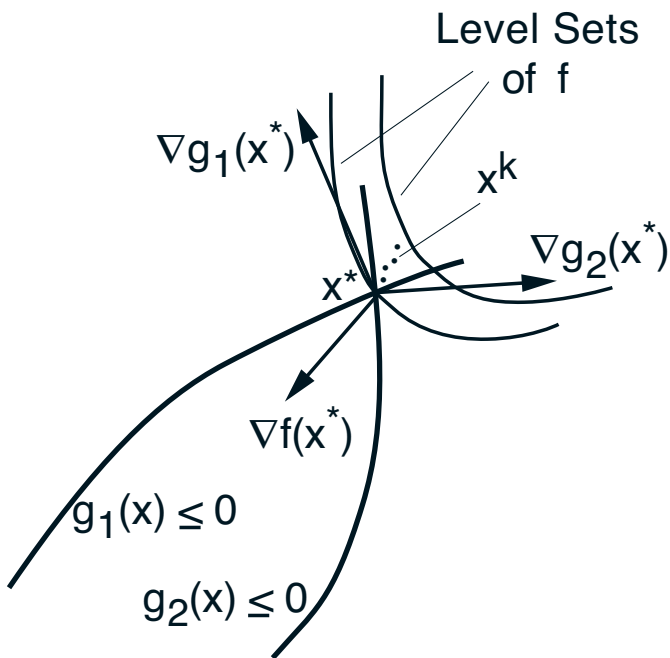
$$\mu_j^* \geq 0, \quad \mu_j^* = 0 \text{ if } g_j(x^*) < 0,$$

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):

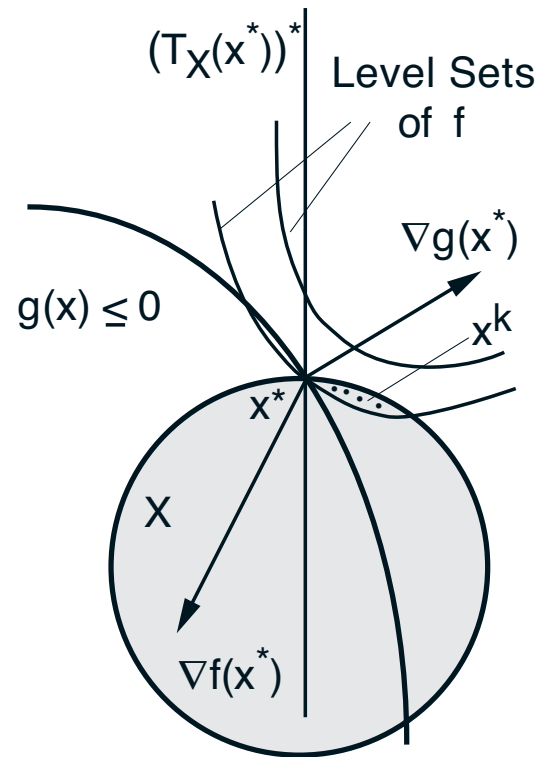
$$\nabla_x L(x^*, \lambda^*, \mu^*)' y \geq 0, \quad \forall y \in T_X(x^*)$$

- **Note 1:** If $X = \mathfrak{R}^n$, 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)



(b)

- (a) Case where $X = \mathbb{R}^n$: $-\nabla f(x^*)$ is in the cone generated by the gradients $\nabla g_j(x^*)$ of the active constraints.
- (b) Case where $X \neq \mathbb{R}^n$: $-\nabla f(x^*)$ 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 $\mu_0^*, \mu_1^*, \dots, \mu_r^*$, satisfying the following:

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

$$(ii) \quad \mu_0^*, \mu_1^*, \dots, \mu_r^* \geq 0 \text{ and not all } 0$$

(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 ,

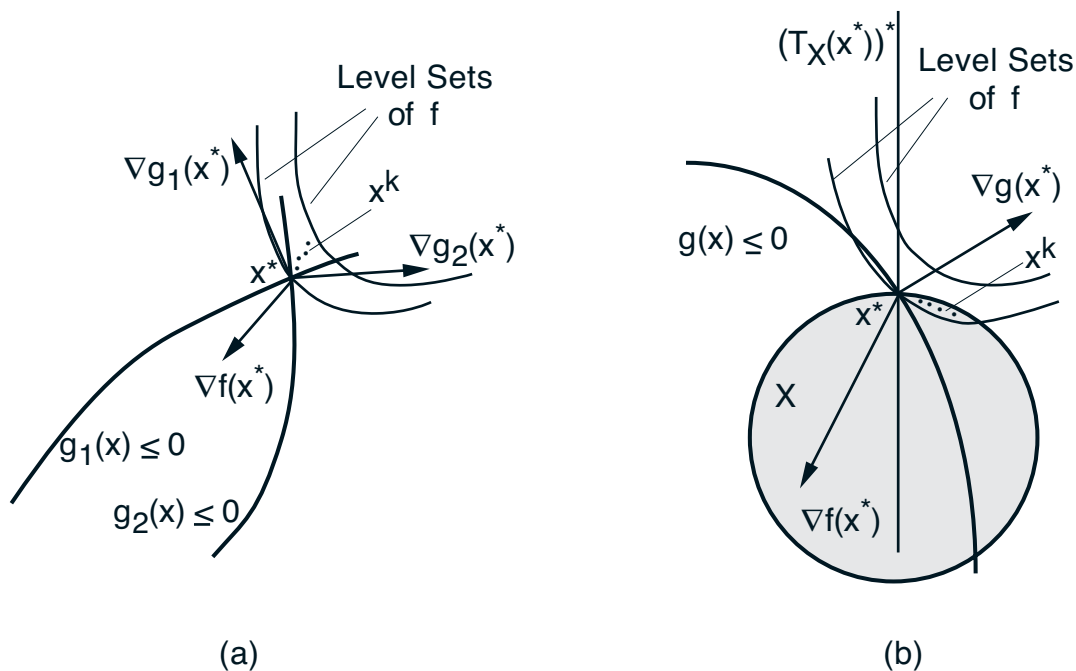
$$f(x^k) < f(x^*), \quad g_j(x^k) > 0, \quad \forall j \in J,$$

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

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

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

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 $(\mu_0^*, \mu_1^*, \mu_2^*)$ satisfying the classical FJ conditions are:

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

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

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

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

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

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

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

- Since $\bar{x} \in S$ and \bar{x} is feasible, we have $f(x^*) \leq f(\bar{x})$, so $\bar{x} = x^*$. Thus $x^k \rightarrow x^*$, 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 $\zeta_j^k = kg_j^+(x^k)$. 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 $\{\mu_0^k, \mu_1^k, \dots, \mu_r^k\}$ is bounded and must contain a subsequence that converges to some limit $\{\mu_0^*, \mu_1^*, \dots, \mu_r^*\}$. This limit has the required properties ...

CONSTRAINT QUALIFICATIONS

Suppose there do NOT exist μ_1, \dots, μ_r , satisfying:

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

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

- Then we must have $\mu_0^* > 0$ in FJ, and can take $\mu_0^* = 1$. So there exist μ_1^*, \dots, μ_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 $X = \mathbb{R}^n$ 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 μ_1, \dots, μ_r , and a sequence $\{x^k\} \subset X$ such that:

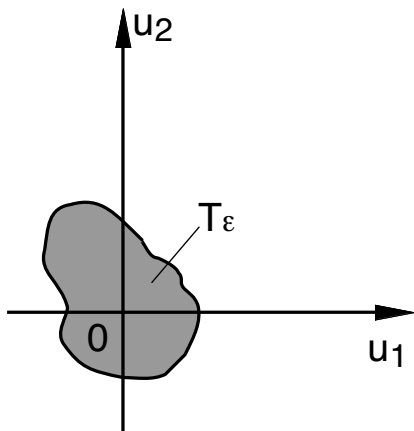
- (i) $-\left(\sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.
- (ii) $\mu_j \geq 0$, for all $j = 1, \dots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.
- (iii) $\{x^k\}$ converges to x^* and

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

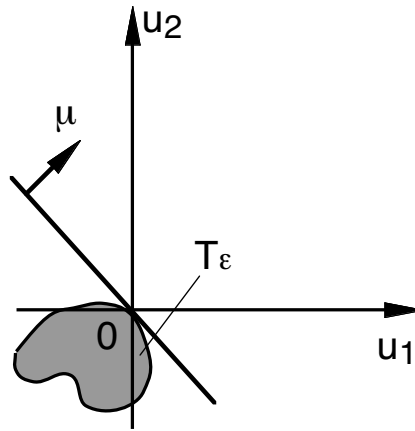
- From Enhanced FJ conditions:
 - If x^* is pseudonormal there exists an R-multiplier vector.
 - If in addition X is regular at x^* , there exists a Lagrange multiplier vector.

GEOM. INTERPRETATION OF PSEUDONORMALITY I

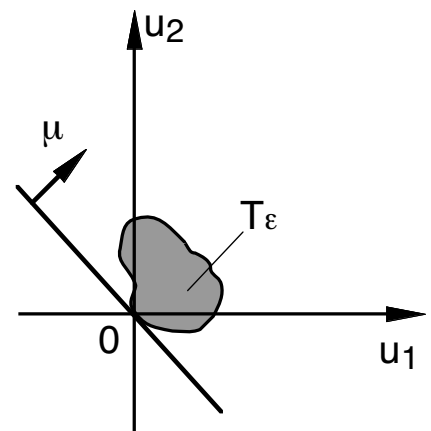
- Assume that $X = \mathbb{R}^n$



Pseudonormal
 ∇g_j : Linearly Indep.



Pseudonormal
 g_j : Concave



Not Pseudonormal

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

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

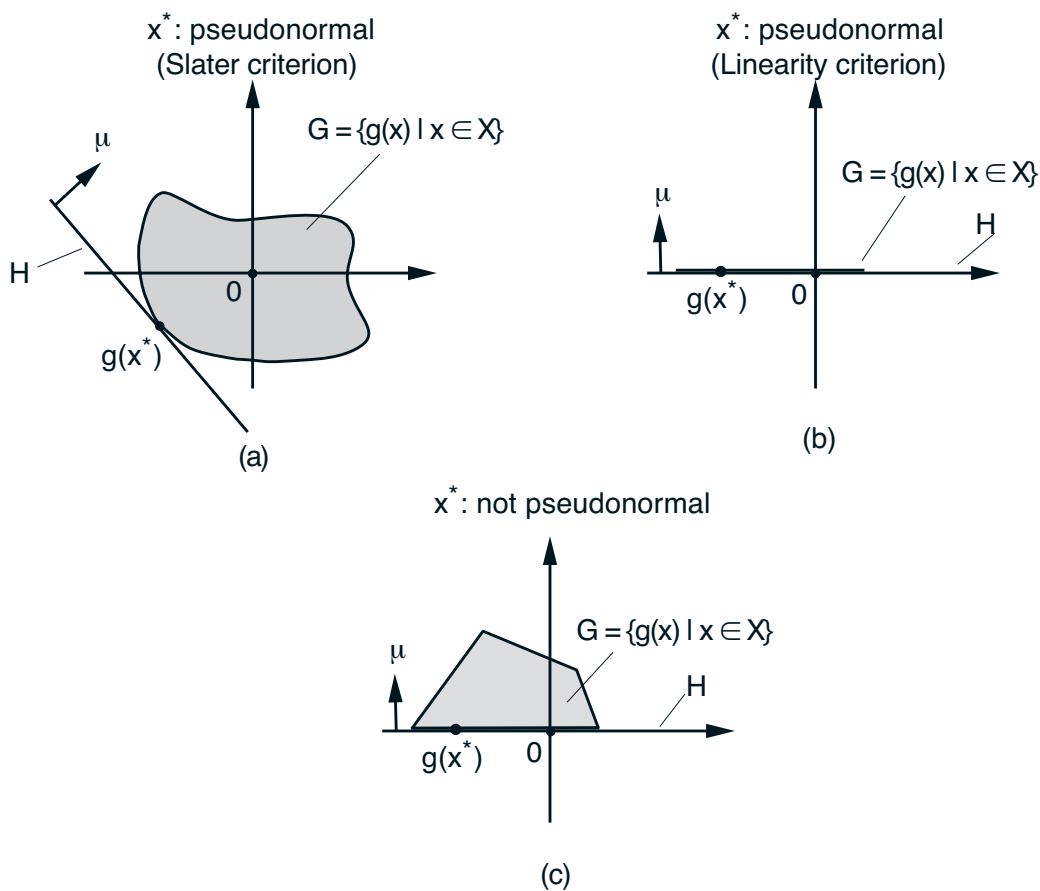
- x^* is pseudonormal if and only if either
 - (1) the gradients $\nabla g_j(x^*)$, $j = 1, \dots, r$, are linearly independent, or
 - (2) for every $\mu \geq 0$ with $\mu \neq 0$ and such that $\sum_{j=1}^r \mu_j \nabla g_j(x^*) = 0$, there is a small enough ϵ , such that the set T_ϵ does not cross 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 $x \in \mathbb{R}^n$].

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: $X = \mathbb{R}^n$, 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 $\bar{x} \in X$ s.t. $g_j(\bar{x}) < 0$ 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 $\lambda = (\lambda_1, \dots, \lambda_m)$ 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

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

CONSTRAINT QUALIFICATION THEOREM

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

Proof: Assume that there are scalars μ_j , $j = 1, \dots, r$, 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 $x \in \mathfrak{R}^n$, so

$$\mu'g(x) \leq \mu'g(x^*) = 0, \quad \forall x \in \mathfrak{R}^n$$

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,

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

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

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

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

- If $X = \mathfrak{R}^n$, 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 $X \neq \mathfrak{R}^n$, as long as X is regular at x^* . These conditions can in turn 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).