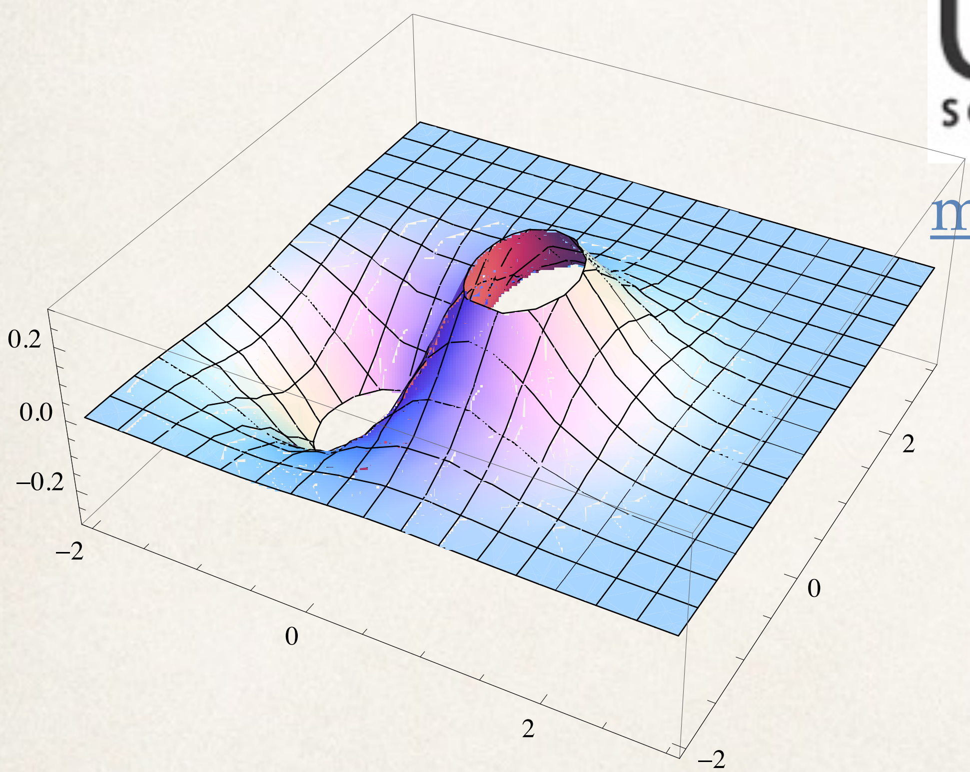
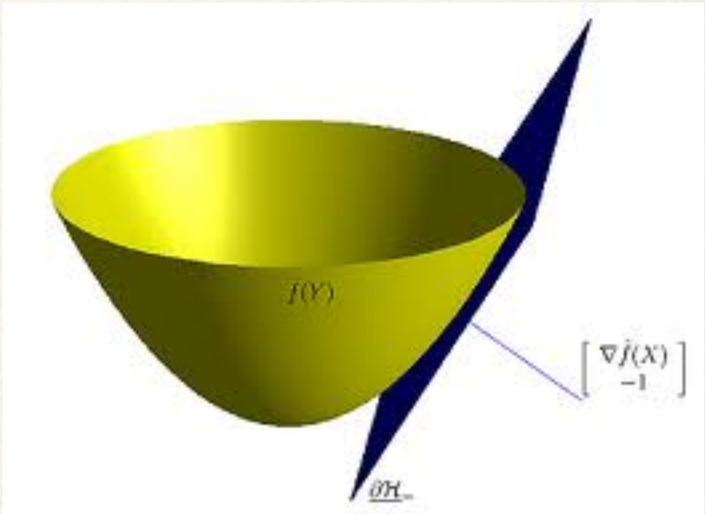
Optimization for CS: Math Intro II

Sridhar Mahadevan

UMASS CS
SCHOOL OF COMPUTER SCIENCE

mahadeva@cs.umass.edu





Outline

- Hahn-Banach theorem in general normed vector spaces
- Farkas lemma and Duality in Linear Programming
- Polar cone theorem and systems of inequalities
- Conjugate Functions and Conjugate Duality

Dual spaces

- Given a vector space X, the dual space X^* is defined as the space of all real-valued linear functionals on X
- Dual spaces play a crucial role in the theory of optimization
- Examples:
  - In Euclidean n-dimensions, $X^* = X$
  - The dual of a Hilbert space is another Hilbert space

Norm of a linear functional

- First, given f*, g* in dual space X*, it is easy to see that
- (3 f* - 2g*)(x) = 3 f*(x) - 2 g*(x)
- Given a linear functional f* on the dual space X*, its norm is defined as

$$\|f^*\| = \inf_M \{M : |f^*(x)| \leq M\|x\|, x \in X\}$$

$$\|f^*\| = \sup_{x \in X} \frac{|f^*(x)|}{\|x\|}$$

$$\|f^*\| = \sup_{\|x\|=1} |f^*(x)|$$

Decomposition using Linear Functionals

- Let X be a normed linear vector space.
- Denote the dual space by X*, elements of which are linear functionals on X
- For any f* in X*, and a point x in X for which f*(x) is non-zero, we have

$$\text{Kernel}(f^*) = \{x | f^*(x) = 0\}$$

$$X = [\text{Kernel}(f^*)] \oplus \text{Span}[x]$$

Proof of Kernel Decomposition

- Note we can write each element y in X as

$$y = \left[y - \frac{f^*(y)}{f^*(x)} x \right] + \frac{f^*(y)}{f^*(x)} x$$

- where it follows that the first element is in the kernel of f*

$$f^* \left(y - \frac{f^*(y)}{f^*(x)} x \right) = 0$$

- A linear subspace Z is said to be of **co-dimension** 1 if

$$X = Z \oplus \mathrm{span}(x_0)$$

- A hyperplane is a translate of a subspace of co-dimension 1

Co-Dimension of Subspaces

- **Lemma**: A linear subspace Z of a vector space X is of co-dimension 1 if and only if it is the kernel of some linear functional f*

- **Proof**: We have already shown that the kernel of a linear functional is of co-dimension 1. It remains to show that given a subspace of co-dimension 1, we can define a suitable linear functional f*

$$X = Z \oplus \mathrm{span}(x_0)$$

$$f^*(z + \lambda x_0) = \lambda$$

Extension of Linear Functionals

- **Lemma:** If Y is a subspace of X, any linear functional f* on Y can be extended to all of X. Also, for each x in X, there is a linear functional f* which is non-zero on x.

- **Proof**: Simply define the extension g* of f* as

$$X = Y \oplus Z \Rightarrow x = y + z$$

$$\text{Define } g^*(x) = f^*(y)$$

- For the second part, given any x in X, define the linear functional

$$f^*(\lambda x) = \lambda \|x\|$$

Sublinear functionals

- A real-valued function p on X is sublinear if it is positively homogeneous and sub-additive

$$p(\alpha x) = \alpha p(x), \alpha \geq 0$$
$$p(x_1 + x_2) \leq p(x_1) + p(x_2)$$

- For example, norms are sublinear functionals

Hahn-Banach Lemma

Hahn-Banach Lemma: Let p be a sub linear functional on the linear vector space X, and let Y be a subspace of X. Let $f^*(x)$ be a linear functional defined on Y such that

$$f^*(x) \leq p(x) \text{ for all } x \in Y$$

Let z be a vector in $X \sim Y$. Then f^* can be extended to the space Span($Y + z$) such that

$$f^*(x) \leq p(x) \text{ for all } x \in \text{Span}(Y + z)$$

Proof

Proof: Note that every vector in Span($Y + z$) can be written uniquely as $x = y + \lambda z$ where $y \in Y$. What we need to show its that it is possible to extend f^* such that

$$f^*(y + \lambda z) = f^*(y) + \lambda f^*(z) \leq p(y + \lambda z), \lambda \in R$$

All we need to do in fact is find a number $f^*(z)$ such that the above inequality holds. For any two vectors $y_1, y_2 \in Y$, since f^* is a linear functional on Y, it follows that

$$f^*(y_1 + y_2) = f^*(y_1) + f^*(y_2) \leq p(y_1 + y_2) = p(y_1 - z + z + y_2) \leq p(y_1 - z) + p(y_2 + z)$$

Therefore, it follows that

$$f^*(y_1) - p(y_1 - z) \leq -f^*(y_2) + p(y_2 + z)$$

To select a value for $f^*(z)$, simply pick the supremum value of the left-hand equation above.

Zorn's Lemma

- Let a partial ordering be defined on a set X:

$$a \leq a, \quad \text{for all} \quad a \in X$$

$$a \leq b \quad \text{and} \quad b \leq a \quad \Rightarrow a = b, \quad \text{for all } a, b \in X$$

$$a \leq b, \quad b \leq c \quad \Rightarrow a \leq c, \quad \text{for all } a, b, c \in X$$

- **Zorn's Lemma:** Let X be a partially ordered set where every totally ordered subset has an upper bound. Then, X has a maximal element.

Hahn Banach Theorem: Extension form

Hahn Banach Theorem: Let p be a sub linear functional on the linear vector space X, and let Y be a subspace of X. Let $f^*(x)$ be a linear functional defined on Y such that

$$f^*(x) \leq p(x) \text{ for all } x \in Y$$

Then f^* can be extended to the entire space X such that

$$f^*(x) \leq p(x) \text{ for all } x \in X$$

Proof: Consider the set of all extensions, which is a partially ordered set. Apply Zorn's Lemma. This set has a maximal element which is defined over all X.

$$(Y_1, f_1^*) \leq (Y_2, f_2^*) \quad \text{if} \quad Y_1 \subseteq Y_2 \quad \text{and} \quad f_2^*(x) = f_1^*(x) \quad \forall x \in Y_1$$

Alignment and Orthogonality

- Let X* be the dual space to X. Let f* be some linear functional in X*.
- f* is **aligned** with a vector x in X if $f^*(x) = \langle x, f^* \rangle = \|f^*\| \|x\|$
- The vector x and functional f* are **orthogonal** if $f^*(x) = \langle x, f^* \rangle = 0$
- The orthogonal complement of a subset S of X is the set of all f* in X* that are orthogonal to every x in S

Minimum Norm problem in dual spaces

- Let x be a vector in linear space X at a distance d from subspace M
- Then the vector in M that is closest to x can be computed as:

$$d = \inf_{m \in M} \|x - m\| = \max_{\|f^*\| \leq 1, f^* \in M^\perp} f^*(x) = \langle f^*, x \rangle$$

- If f* is a linear functional at distance d from $M^\perp$, then

$$d = \min_{m^* \in M^\perp} \|f^* - m^*\| = \sup_{x \in M, \|x\| \leq 1} f^*(x) = \langle f^*, x \rangle$$

Consequence of HB Theorem

- Let X be a normed linear space, and Y be a subspace of X where f* is a linear functional defined on Y. Then f* can be extended to a linear functional F* on X such that

$$f^*(x) = F^*(x) \text{ forall } x \in Y$$

$$\|f^*\| = \|F^*\|$$

- **Proof**: Pick p, the sublinear functional p that dominates f*, as:

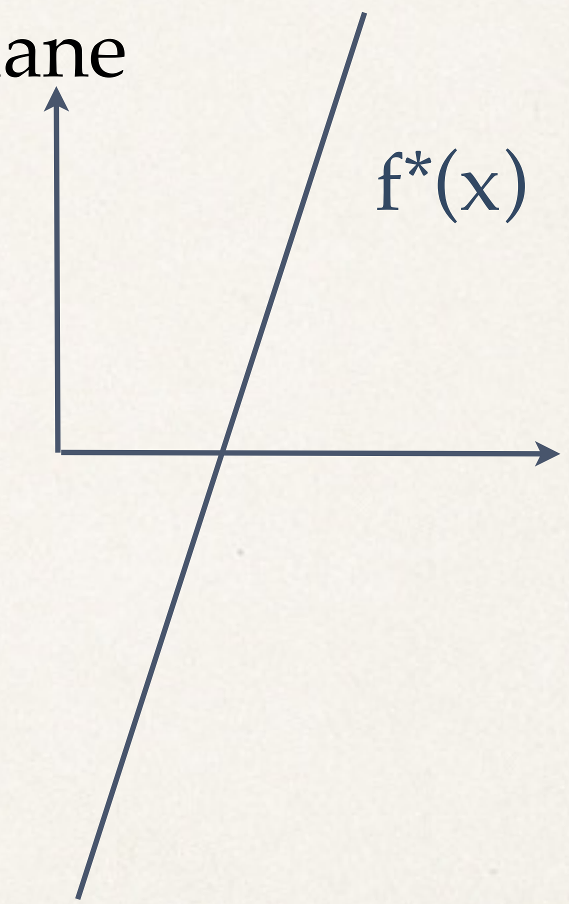
$$p(x) = \|f^*\| \|x\|$$

Application: Approximation

- The Hahn-Banach theorem is extremely valuable in answering questions regarding whether some class of functions can be approximated arbitrarily well by another class
- Example: consider the space of continuous functions C[0,1]. Can any continuous function be approximated arbitrarily well by a polynomial? (Stone Weierstrass theorem)
- Example: Can any continuous function on R^n be approximated by a feedforward neural net as a linear superposition of sigmoid functions? (Cybenko)

Hyperplanes

- A hyperplane H in a linear vector space X is a maximal proper linear variety
- If H is contained in some variety V, either H=V or V=X
- For every hyperplane H in X, there is a linear function f* and c such that H = {x : f*(x) = c}
- Conversely, the level set of any functional defines a hyperplane



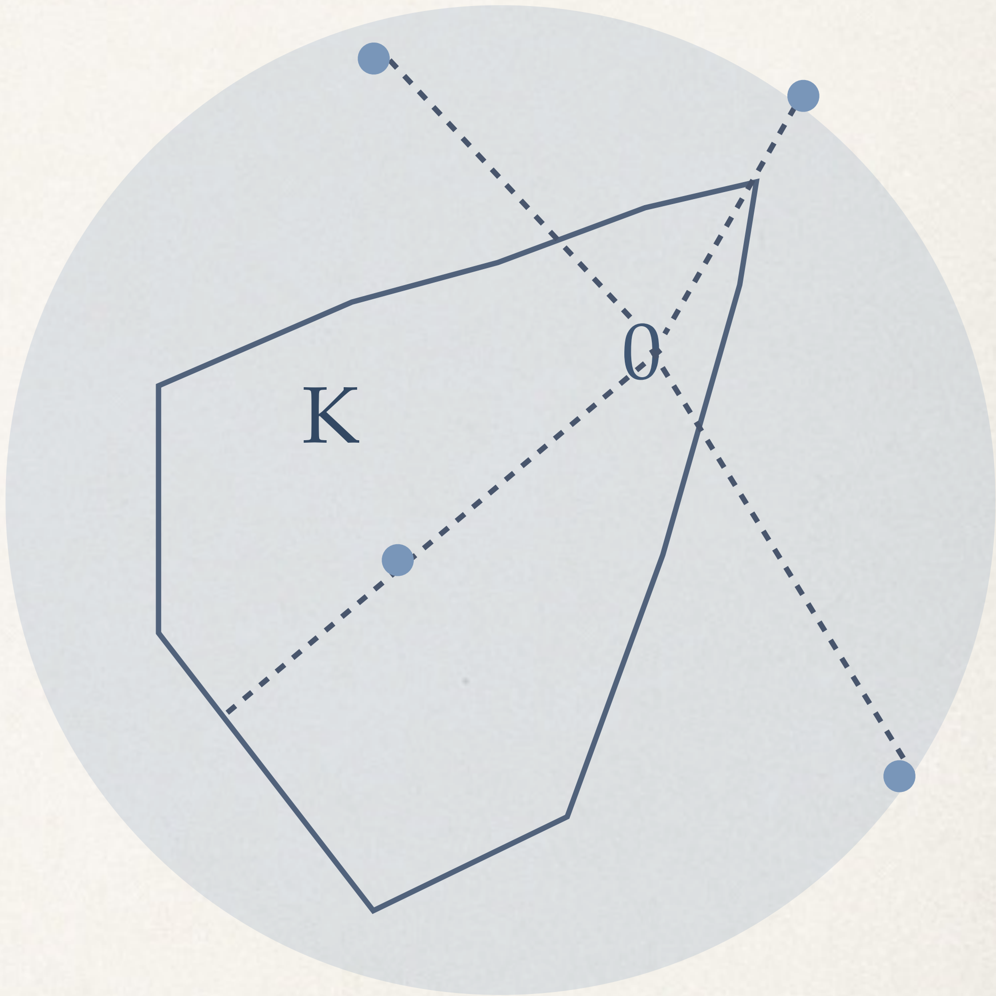
Minkowski functional of a convex set

- Let K be a convex set in a linear normed space X, such that 0 is an interior point of K

- The Minkowski functional p of K is defined on X as

$$p(x) = \inf\{r : \frac{x}{r} \in K, r > 0\}$$

- It can be shown that the Minkowski functional is a sublinear functional



Properties of Minkowski Functional

- The Minkowski functional of a convex set K is positive homogeneous and convex

$$\text{if } \frac{x}{r} \in K \text{ then } p_K(\alpha x) = \alpha r, \;\; \alpha \geq 0$$

$$\text{if } \frac{x}{a}, \frac{y}{b} \in K \text{ then } \frac{a}{a+b}\frac{x}{a} + \frac{b}{a+b}\frac{y}{b} = \frac{x+y}{a+b} \in K$$

$$\text{Hence } \;\; p_K(x+y) \leq a+b = p_K(x) + p_K(y)$$

- If x is an interior point of K, its Minkowski functional is < 1

Lagrange Duality: Example

- Consider solving the following minimization problem:
    - Minimize $x^2 + y^2$ subject to $x + y = 1$
- Introduce a new variable (``Lagrange dual”) p that intuitively measures the ``price” for violating the constraint $x+y = 1$
    - $L(x,y,p) = x^2 + y^2 + p(1 - x - y)$
- Minimize L w.r.t. x and y holding p fixed
    - This yields $x = y = p/2$ or $x = y = 1/2$

Lagrangian Duality

- We will illustrate the usefulness of Hahn-Banach theorem in the duality principle for LP
- Each LP has a corresponding dual
  - Minimizing the primal is equivalent to maximizing the dual

$$\min_x c^T x$$

$$Ax = b$$

$$x \geq 0$$

Lagrangian Duality in LP

- Define the original LP problem as a primal problem
  - Minimize $c'x$ subject to
  - $Ax = b$, $x >= 0$
- Consider the relaxed problem
  - Minimize $c'x + p'(b - Ax)$
  - such that $x >=0$
- The vector p represents a ``price'' vector for violating $Ax = b$

Duality in LP

- Note that the dual problem solution is a lower bound on the solution of the primal problem

$$min_{x \geq 0} \left(c'x + p'(b - Ax) \right) \leq c'x^* + p'(b - Ax^*) = c'x^*$$

- We can write the dual problem as

$$p'b + min_{x \geq 0} \left((c' - p'A)x \right)$$

Duality in LP

- Note that
$$\min_{x \geq 0} (c' - p'A)x = -\infty \text{ when } (c' - p'A) < 0$$

$$\text{Dual LP Problem:} \quad \max_{p \in \mathbb{R}^m} p'b$$
$$\text{such that} \quad p'A \leq c'$$

Example of Dual LP problem

- Primal problem:

Minimize $x_1 + x_2$ such that

$x_1 + 2x_2 - x_3 = 2$

$x_1 - x_4 = 1$

$x_1, x_2, x_3, x_4 >= 0$

- Dual Problem:

Maximize $2p_1 + p_2$ such that

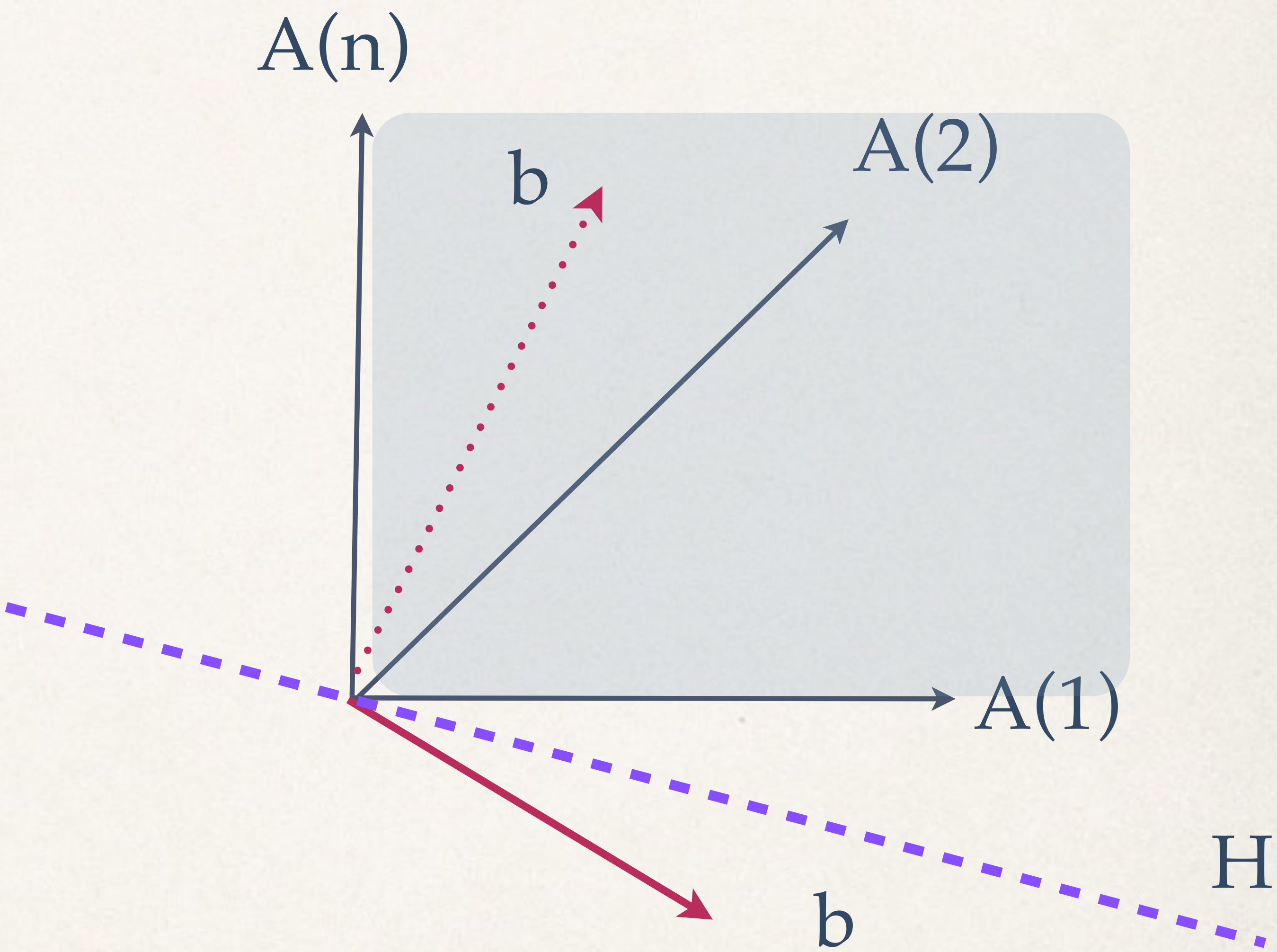
$p_1 + p_2 <= 1$

$2p_1 <= 1$

$p_1, p_2 >= 0$

Farkas Lemma and LP

- Farkas' lemma states that either a vector is in a given convex cone, or there is a hyperplane separating the vector from the cone
- Fundamental result that forms the basis for understanding duality in LP



Farkas Lemma

- Let A be a matrix of size m x n.
- Let b be a column vector of size m
- Then, only one of the following holds:
  1. $Ax = b$ has a nonnegative solution ($x >= 0$)
  2. $y' A >= 0$ and $y' b < 0$ has a solution
- Proof: Follows from Hahn-Banach theorem

Proof of Farkas Lemma

- **Proof**: Suppose both conditions are true. Then, a contradiction ensues:

$$0 <= (y' A)x = y'(Ax) = y' b < 0!$$

- Let C be the convex set of non-negative vectors Ax, x>=0.
- If b is in C, then condition 1 holds. Otherwise, b is separable from C by a hyperplane. Specifically, there is a vector p such that p'b < p'y for all y in C
- Since $A_i\delta$ is in C for $\delta >= 0$, it follows $p'b < \delta p' A_i$.
- This implies $(1/\delta)p'b < p'A_i$. As $\delta \to$ infinity, $p'A_i >= 0$, so $p'A >= 0$

Duality Theorem for LP

- **Lemma**: if there exists x and p such that $c'x = p'b$, then $x = x^*$ is the optimal primal solution and $p = p^*$ is the optimal dual solution
- **Proof**: Note that since any dual solution forms a lower bound on the primal solution, it follows that:
  - $c'x = p'b <= c'y$ (for any y)
  - Consequently, x must be optimal

Duality in LP: Alternate Form

- Primal:
  - Minimize $c'x$ such that $Ax >= b$
- Dual form:
  - Maximize $p'b$ such that
  - $p'A = c'$
  - $p >= 0$

LP Duality Theorem Proof

- Let $I = \{i \mid a'_i x^* = b_i\}$ where x^* is the optimal primal solution
- Any vector d such that $a_i'd >= 0$ for any i in I also satisfies $c'd>=0$
- This holds since $a'_i(x^* + \varepsilon d) >= b_i$ for all i. Hence $c'(x^* + \varepsilon d) >= c^*$.
- By Farkas lemma, c can be written as $c = \Sigma_{i \varepsilon I}\, p_i\, a_i$. Define $p_i=0$ for i not in I. This implies $p'A = c'$.
- Also, $p'b = \Sigma_{i \varepsilon I}\, p_i\, b_i = \Sigma_{i \varepsilon I}\, p_i\, a'_i\, x^* = c'x^*$.
- This means dual cost = primal cost, and hence both are optimal!

Polar Cone Theorem

- Recall the definition of a cone:

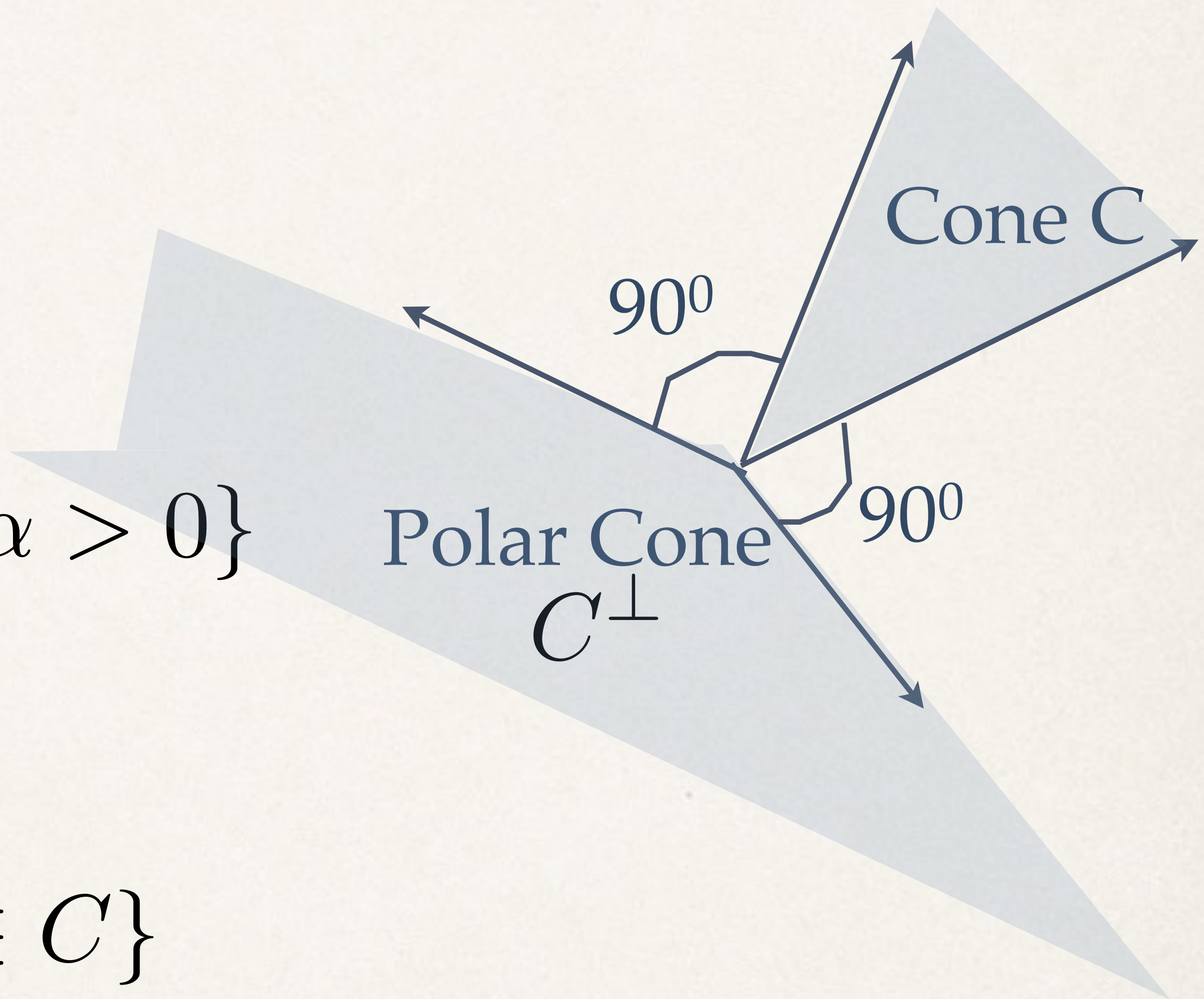
$$C = \{y \in X : y = \alpha x, x \in X, \alpha > 0\}$$

- The **polar cone** is defined as:

$$C^{\perp} = \{y \in X : \langle y, x \rangle \leq 0, x \in C\}$$

- The polar cone theorem states:

$$(C^{\perp})^{\perp} = C$$



System of Inequalities

Let C be a convex set and $f_1, \ldots, f_k$ be convex functions such that dom $f_i \supset$ ri C. Let $g_1, \ldots, g_l$ be affine functions such that the system

$$g_1(x) \leq 0, \ldots, g_l(x) \leq 0$$

has at least one solution in the ri C. Then, only one of the following alternatives holds:

- There exists some $x \in C$ such that

$$f_1(x) < 0, \ldots, f_k(x) < 0 \qquad g_1(x) \leq 0, \ldots g_l(x) \leq 0$$

- There exists non-negative real numbers $\lambda_1, \ldots, \lambda_k, \xi_1, \ldots, \xi_l$ such that at least one of the λ_i are non-zero and

$$\lambda_1 f_1(x) + \ldots + \lambda_k f_k(x) + \xi_1 g_1(x) + \ldots + \xi_l g_l(x) \geq 0, \quad \forall x \in C$$

Farkas Lemma

- Let A be a matrix of size m x n.
- Let b be a column vector of size m
- Then, only one of the following holds:

1. $Ax = b$ has a nonnegative solution ($x >= 0$)
2. $y' A >= 0$ and $y' b < 0$ has a solution

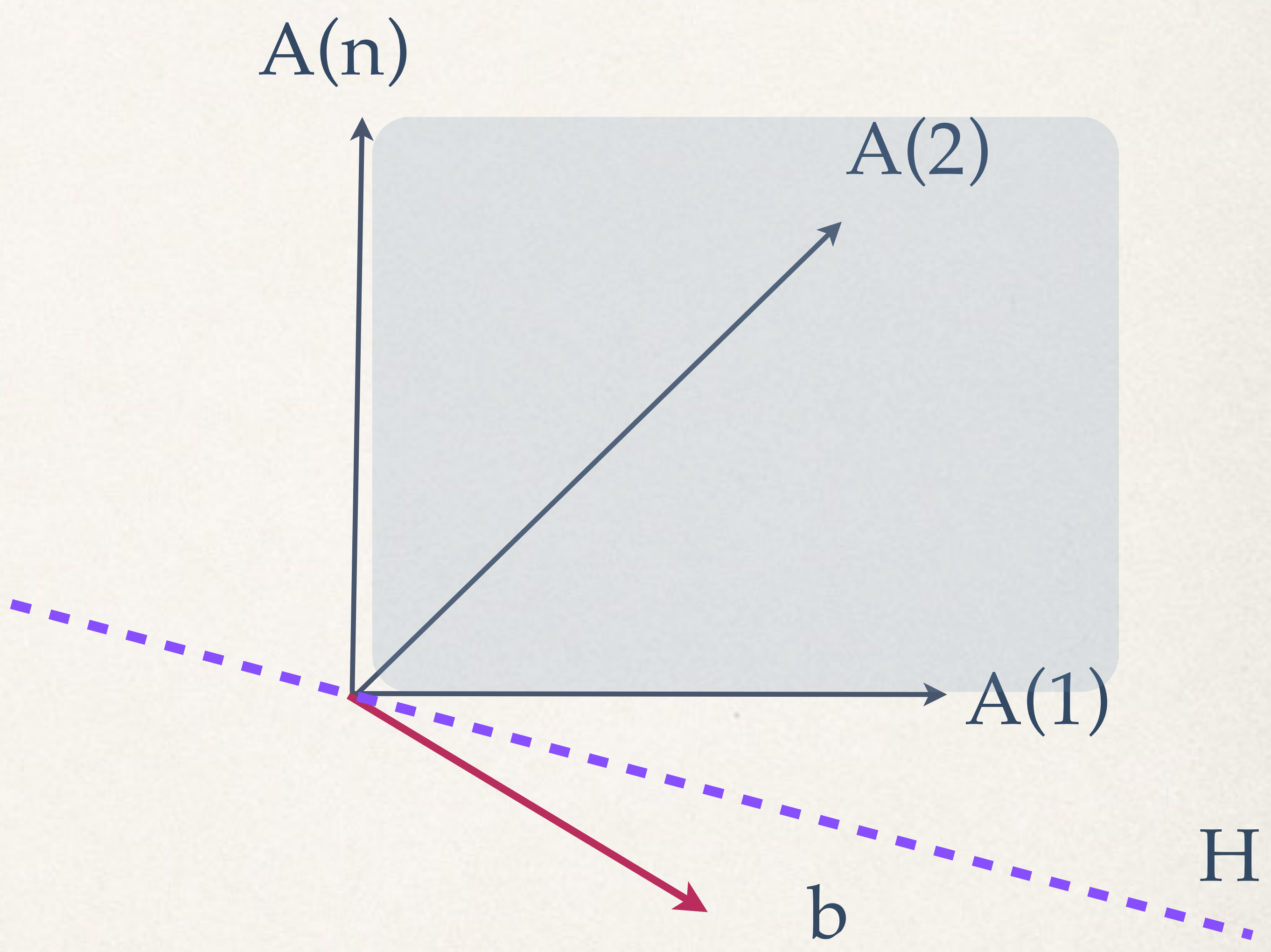
- **Proof**: Suppose both conditions are true. Then, a contradiction ensues:

  $0 <= (y' A)x = y'(Ax) = y' b < 0$!

- Let C be the convex set of non-negative vectors Ax, x>=0.
- If b is in C, then condition 1 holds. Otherwise, b is separable from C by a hyperplane.

Farkas Lemma and LP

- Farkas' lemma states that either a vector is in a given convex cone, or there is a hyperplane separating the vector from the cone
- Fundamental result that forms the basis for understanding duality in LP



Reading Assignment

- Conjugate Duality is covered in Chapter 3 of Boyd and Vandenberghe
- Lagrange Duality is covered in Chapter 5 of Boyd and Vandenberghe
- Read article on Hahn-Banach theorem on Moodle
- Work out all the examples in this lecture