BREAKTHROUGH IN PROBLEM SOLVING


+ Karmakar (28 years old, recent UCB PhD) features twice in NYT in '84.
+ This was a poly-time interior point method. We'll study this.
+ "It has also set off a deluge of inquiries from brokerage houses, oil companies and airlines, industries with millions of dollars at stake in problems known as linear programming."
+ "This is a path-breaking result," said Dr. Ronald L. Graham, director of mathematical sciences for Bell Labs in Murray Hill, N.J. "Science has its moments of great progress, and this may well be one of them."
+ K talks with American Airlines: How much fuel to carry? Where to fuel?
+ Exon's research head says "studies underway".
+ Dantzig is cautious; he was partial to the simplex method.

A Soviet Discovery Rocks World of Mathematics


+ Khachiyan (late 20s?) features twice in NYT in '79.
+ "applicable in weather prediction, complicated industrial processes, petroleum refining, the scheduling of workers at large factories, secret codes and many other things."
+ This was the ellipsoid algorithm; also poly time. We'll study this.
Karan: Today, such press seems parallel to the coverage ML/deep learning gets.

George Dantzig's Story
+ From WWI/WWII era (distributed) logistics, productions problems. Questions around: what to do/when to do to arrive at some state/achieve some objective. Semantics: programming – planning.
+ Dantzig (USAF) 1947 formulates/recognizes the general linear programming problem as a possible compromise between solvable and interesting problem classes. Also, proposes the simplex algorithm.
+ His claim (in his text): previous work did not have an objective function, i.e. only posed feasibility problems. An example is Motzkin's 1936 thesis which cites 42 pages, none considering an objective.
+ Some LP special cases (Koopmans, Leontif, Kantorvich) would win Nobel in Econ.
+ Meets von Neumann to discuss. Von Neumann is annoyed, “get to the point!”. On seeing the problem, delivers an impromptu 1.5 hour lecture to Dantzig and describes both LP duality (including Farkas's Lemma) and an early interior point method. What triggered this?

Von Neumann’s Story
+ See [https://wwnorton.com/books/the-man-from-the-future](https://wwnorton.com/books/the-man-from-the-future). A prodigy, and reputed as a deep mathematician who interfaces with applied problems/worldly affairs, e.g., consults on Manhattan project.
+ Early contributions include a resolution to fundamental inconsistencies in mathematics (Russel's paradox: S is the set of all sets which are not members of themselves. Is S in S? Others resolved it simultaneously by better means.), and rigorous unification of wave equation and matrix mechanics in QM (earlier heuristic argument by Dirac using delta functions).
+ In 1944, a book with an Economist Morgenstern on The Theory of Games and Economic Behavior.
+ Minimax duality in 2-person zero-sum games is same as LP duality. Today, called von Neumann duality. But, it is von Neumann's?

Truer Origins of Duality
+ Monge proposes a question about the transportation problem in 1700's, used to model moving ores from mines to factories at minimum cost.
+ Kantorvich (1939) solves it, constructs the dual. Transportation is as general as LP. Kantorvich largely ignored in Russia. We will study this too.
Linear Programming — a class of optimization problems that are useful and simultaneously tractable.

**What is LINEAR?**

**Def.** A function \( f : X \rightarrow \mathbb{R} \), where \( X \subseteq \) some vector space \( V \), is LINEAR if

(a) \( f(x + y) = f(x) + f(y) \) \( \forall x, y \in X \).

(b) \( f(\alpha x) = \alpha f(x) \) \( \forall x \in X, \alpha \in \mathbb{R} \).

**Prop.** If \( V = \mathbb{R}^n \), then for any \( X \subseteq V \) and LINEAR \( f : X \rightarrow \mathbb{R} \)

\( \exists f(0) \in \mathbb{R}^n \) such that \( f(x) = c^T x \) \( \forall x \in X \).

Furthermore, if \( X = \mathbb{R}^n \), then the choice of such \( c \in \mathbb{R}^n \)

given a fixed \( f \) is unique.

**Comment:** \( x \rightarrow c^T x \) is of course linear. The interesting bit is that any linear function on \( X \subseteq \mathbb{R}^n \) can be written this way. Also, this representation may not be unique if \( X \) is a subset of a proper subspace of \( \mathbb{R}^n \), i.e., is lower-dimensional.

**Ex.** Try to see why! This does not impugn existence.

**Proof Sketch:** We will only consider the case when \( X = \mathbb{R}^n \).

First, the general case \( V = \mathbb{R}^n \) can be your own.

Any \( x \in \mathbb{R}^n \) can be written as \( x = \sum_{i=1}^{n} x_i e_i \)

where \( e_i = [0 \ldots 1 0 \ldots 0]^T \).

Hence, \( c = [f(e_1), \ldots, f(e_n)]^T \) satisfies said claim.

**Motivating Definition of LPS**

**ATTEMPT 1:** \( \max \mathbb{C}^T x \)

subject to \( x \in \mathbb{R}^n \)

(\( \text{cor such that/ s.t.} \))

\( c \) or \( + \infty \) when \( c \neq 0 \).

\( \text{Not interesting/useful.} \)

**ATTEMPT 2:** Impose constraints — will return

A few definitions

(1) A set \( S \) is CLOSED if it contains all its limit points, i.e.,

\[ \lim_{n \to \infty} x_n \in S \]

where \( x_n \in S \) where \( x_n \to \) exists for \( x_n \in S \).
(2) A set $S$ is BOUNDED if $\exists C \in \mathbb{R}$ such that $\forall x \in S$, $\|x\| \leq C$. Choice of norm (in finite-dimensional spaces) is not crucial / is immaterial to this definition.

(3) $S$ is COMPACT if it is CLOSED and BOUNDED.

(4) A function $f: X \to Y$ is CONTINUOUS if for any $x_1, x_2, \ldots \in X$ such that $\lim_{n \to \infty} x_n$ exists and is in $X$, we have $\lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n)$.

(5) Given a function $f: X \to \mathbb{R}$, its maximum is ATTAINED on $S \subseteq X$ if $\exists x^* \in S$ such that $\forall x \in S$, $f(x) \leq f(x^*)$.

A nice result; makes life easy.

**THEOREM:** A continuous function ATTAINS its maximum (and minimum) (Weierstrass) on any non-empty compact set.

RETURNING TO LP MOTIVATIONS

Take any compact set $S$.

$$\max_{x \in S} c^T x \quad \text{equivalent} \quad \max_{(t, x)} \quad \text{s.t.} \quad x \in S \quad f(x) \geq t.$$  

Optionally add $-10^{10} \leq t \leq 10^{10}$ to ensure compactness: NITPICK.

**ATTEMPT 3:** $\max_{(\text{George Dantzig}, 1947)} c^T x \quad \text{s.t.} \quad A^T_1 x \preceq b_1 \quad A^T_2 x \preceq b_2 \quad A^T_m x \preceq b_m$  

Can express many problems + efficiently scalable.

These are technically affine (We will call them linear).

In addition to linearity, having a finite number of constraints is also important to guarantee tractability.
LPs as a special case of **convex programs**

1. A set \( S \subseteq \mathbb{R}^n \) is **convex** if \( \forall x, y \in S \), \( \forall \lambda \in [0,1] \), \( \lambda x + (1-\lambda)y \in S \).

   \[ S_{x,y} = \{ \lambda x + (1-\lambda)y : \lambda \in [0,1] \} \text{ is a line segment between } x \text{ and } y. \]

   **Comments:**
   1. \( \{ x : a^T x \leq b \} \) is a convex set.
   2. The intersection of two (or any number of) convex sets is convex. (Ex)
      
      \[ \text{Then, } \left\{ x : \begin{array}{l} a_1^T x \leq b_1 \\ a_m^T x \leq b_m \end{array} \right\} \text{ is convex.} \]
      \[ A \cap B = \{ a : a \in A \text{ and } a \in B \} \]
   3. If \( A \) and \( B \) are convex, then \( A \times B = \{ (a,b) : a \in A, b \in B \} \) is convex. (Ex)
      
      \[ \text{CARTESIAN PRODUCT} \]
   4. If \( A \) and \( B \) are convex, then \( A + B = \{ a + b : a \in A, b \in B \} \) is convex. (Ex)
      
      \[ \text{MINKOWSKI SUM} \]
   5. If \( S \) is a convex set in \( \mathbb{R}^n \), \( A \subseteq \mathbb{R}^m \), then \( \{ A x : x \in S \} \) is convex. (Ex)
   6. For any convex function \( f \) on a convex set \( S \),
      
      \[ \text{argmin}_{x \in S} f(x) = \left\{ x^\star \in S : f(x^\star) \leq f(x) \text{ for all } x \in S \right\} \text{ is convex.} \]

2. A function \( f : X (\text{convex}) \rightarrow \mathbb{R} \) is said to be convex if \( \forall x, y \in X \), \( \forall \lambda \in [0,1] \),

   \[ f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \]

   In words, interpolating line lies above.
(3) How is this related to convex sets? 
\[ \text{epi}(f) = \{ (t,x) : t \geq f(x) \} \]
**Definition**

**Proposition.** $f$ is a convex function if $\text{epi}(f)$ is a convex set.

A tantalizing, but incorrect way to link convex sets and functions is: ask $f$ is a convex function if and only if $\forall t \in \mathbb{R}, S_t[f] = \{ x : f(x) \leq t \}$ is convex? FALSE. Such functions are called quasi-convex.

**Example:** $\sqrt{|x|}$ is quasi-convex, but not convex. Convex $\Rightarrow$ quasi-convex.

(4) Finally, $x \mapsto Cx$ is a convex function. Hence, $C$ are convex.

General convex program: \[ \min f(x) \quad \text{convex function} \]
\[ \text{s.t. } x \in S \quad \text{convex set} \]

More generally, for \[ \max f(x) \]
\[ \text{s.t. } x \in S \]

$S$ is called feasible region. $x^*$ maximizing $f$ on $S$ is called an optimal solution.

References:
1. History—
   1. p 209 in Schrijver
   2. Dantzig’s article
2. Basics of Convexity— chapters 2 & 3 in Boyd
**Examples and Notations Involving LPS**

* $x \leq y$ for vectors $x, y \in \mathbb{R}^n$ if $x_i \leq y_i \; \forall i \in \{1, \ldots, n\}$.

  Note: for vectors, it is not true that either $x \leq y$ or $y \geq x$.

Also, very much basis dependent (here, standard basis).

$$\begin{align*}
\max \; c^T x \\
\text{s.t. } a^T x \leq b \\
\end{align*}$$

$$\begin{align*}
\max \; c^T x \\
\text{s.t. } \begin{bmatrix}
a_1^T \\
a_2^T \\
\vdots \\
a_m^T \\
\end{bmatrix} x \leq \\
\begin{bmatrix}
b_1 \\
b_2 \\
\vdots \\
b_m \\
\end{bmatrix}
\end{align*}$$

**Example: Diet Problem**

$n$ food items; food $j$ has cost $c_j$.

$m$ nutrients; minimum acceptable level of nutrient $i$ is $b_i$.

$a_{ij}$ is amount of nutrient $i$ in food $j$.

$$\begin{align*}
A &= \begin{bmatrix} a_{ij} \end{bmatrix} \in \mathbb{R}^{m \times n} \\
\min \sum_{j=1}^n c_j x_j \\
\text{s.t. } \sum_{j=1}^n a_{ij} x_j \geq b_i
\end{align*}$$

**Standardization of LPS**

How to transform

(1) $\max c^T x = -\min (-c)^T x$

(2) $a^T x \leq b \iff (-a)^T x \geq (-b)$

(3) $a^T x = b \iff a^T x \leq b \land a^T x \geq b$

(4) $a^T x \leq b \iff a^T x + s = b \land s \geq 0$

(5) unconstrained $x \iff$ replace with $x^+ - x^- \geq 0$

where $x^+, x^- \geq 0$.

General Form

$$\begin{align*}
\max \; c^T x \\
\text{s.t. } A x \leq b \\
\end{align*}$$

Standard Form

$$\begin{align*}
\min \; c^T x \\
\text{s.t. } A x = b, \; x \geq 0
\end{align*}$$
\* Example

\[
\min_{x \in \mathbb{R}^n} \max_{i \in [m]} (a_i^T x + b_i) \iff \min_{(y, x)} t \quad \text{s.t.} \quad t \geq a_i^T x + b_i \quad \forall i \in [m]
\]

Why is this tractable?

Given convex functions \(f_1, \ldots, f_m\), \(f(x) = \max_{i \in [m]} f_i(x)\) is convex.

\* Example

\[
\max \; C^T x \\
\text{s.t.} \; \|x\|_1 = \sum_{i=1}^n |x_i| \leq 1
\]

Manually solving this.

Max value = \(\max_{i \in [n]} |c_i|\)

Let \(S = \{ i : |c_i| = \max_{i \in [n]} |c_i| \}\)

\[\arg \max = \text{sign}(c_i) e_i \text{ for } i \in S\]

Let's express this as an LP.

Attempt 1: \(\max \; C^T x \quad \text{s.t.} \; \sum_{i=1}^n |x_i| \leq 1\)

Exponentially many constraints.

Attempt 2: \(\max \; C^T x \quad \text{s.t.} \; y_i \geq x_i, y_i \geq -x_i, \sum_{i=1}^n y_i \leq 1\)

2n+1 constraints! Exponential improvement.

Extended formulations (area of study: extension complexity) aim to achieve drastic reductions in the number of constraints, by introducing a few (polynomially) more variables.

History:

* Early 1980s/late 1980s: numerous attempts at solving NP-complete problems using LPs; hence, at P=NP.

John: Typical hard problems (like TSP) can be written as LPs with exponentially many constraints. Introduce a few more variables to achieve (similar to the above example) polynomially many constraints.
* Yamazaki ‘91: Any symmetric formulation of TSP as LP has exponentially many constraints.
* FMPTW ‘12: Any formulation of TSP as LP has exponentially many constraints.
* Kidding the 1980/90’s hope.

* EXAMPLE: OPTIMAL TRANSPORT (equivalent to general LPs)

Monge 1700’s: Want to move iron ore from mines to factories. Mines produce \( u(x) \) ore at \( x \), factories at \( y \) consume \( v(y) \) ore.

Cost of transporting unit quantity from \( x \) to \( y \): \( c(x, y) \).

Minimize:

\[
\min_{T: x \rightarrow y \text{ invertible}} \int c(x, T(x)) \, u(x) \, dx
\]

Subject to:

\[
\begin{align*}
\text{S.t. } & \quad u(y) = \left| \det \nabla T(y) \right| v(T^{-1}(y)) + y \in \mathbb{R}^d \\
& \quad \text{push-forward of } y \text{ through } T \\
& \quad \text{Roughly, } \text{by conserving probability measure}
\end{align*}
\]

\[
\begin{align*}
u(x) \, dx &= \nabla T(x) \left| \det \nabla T(x) \right| \, dx \\
&= v(T(x)) \, dy
\end{align*}
\]

(\text{Assume, assume that})

\[
\int_{x} \nu(y) \, dx = \int_{y} u(y) \, dy = 1
\]

Interpretable as probability measures.

Highly non-linear problem, + 2 Evar Nakel figures

Today applications in concentration of measure
* Dynamical systems
Cedric Villani won a Fields Medal solving related problems.

Kantorovich 1939: Tractable solution by reformulating the problem. Notice that this is pre-LP/Neumann minimax duality. Also, game dual.

\[
\min_{\Gamma: X \times Y \to \mathbb{R}} \int \int \Gamma(x,y) u(x) v(y) \, dx \, dy
\]

s.t. \forall y \in Y \quad \int \Gamma(x,y) u(x) \, dx = v(y)

\forall x \in X \quad \int \Gamma(x,y) v(y) \, dy = u(x)

Notice this is linear in \( \Gamma \). \( u, v \) are probability mass functions, instead of density, same as LP below.

\[
\min_{\Gamma} \sum_{x \in X} \sum_{y \in Y} \Gamma(x,y) u(x) v(y)
\]

s.t. \forall y \in Y \quad \sum_{x \in X} \Gamma(x,y) u(x) = v(y)

\forall x \in X \quad \sum_{y \in Y} \Gamma(x,y) v(y) = u(x)

Intuitively, in Kantorovich’s Transport Plan formulation, production & supply capacities arise simultaneously while maintaining marginal rates of \( u, v \).

Also, related: “Coupling” in probability theory.

**Convention in Convex Optimization**

\[
\min_{x \in S} f(x) \text{ is equivalent to } \min_{x \in \mathbb{R}^n} f_S(x),
\]
where $f_S(x) = \begin{cases} +\infty & \text{if } x \not\in S \\ f(x) & \text{if } x \in S \end{cases}$

Similarly, $\max_{x \in S} f(x) = \max_{x \in \mathbb{R}^n} f'_S(x)$,

where $f'_S(x) = \begin{cases} -\infty & \text{if } x \not\in S \\ f(x) & \text{if } x \in S \end{cases}$

Feasibility programs ask does there exist $x \in \mathbb{R}^n$ such that $x \in S$?

Linear Feasibility: $\exists \ x \in \mathbb{R}^n$ s.t. $Ax \leq b$ (Generally represented)

Optimization & feasibility are closely related in computational terms.

Using an optimization solver to check feasibility:

\[ O(c) = \max_{x} c^T x \]

\[ \text{s.t. } x \in S \]

Obviously, $O(c) = 0$ iff $\exists x \in \mathbb{R}^n$ s.t. $x \in S$.

Using a feasibility solver for optimization:

\[ \mathcal{F}(S) = \begin{cases} \text{YES} & \text{if } \exists x : x \in S \\ \text{NO} & \text{otherwise} \end{cases} \]

Let's say we have some a-priori range for $-10^{100} \leq \max_{x \in S} f(x) \leq 10^{100}$.

Using $\mathcal{F}$, we will solve $\max_{x \in S} f(x)$ to $\varepsilon$-accuracy.

**Algorithm $A$:**

1. If $|b - u| \leq \varepsilon$, then output any value in $[l, u]$.
2. Else, $t = \frac{l+u}{2}$.

   \[ \text{If } \mathcal{F}(S \cap \{x: f(x) \geq t\}) = \text{YES}, \quad \forall t \leq \frac{b+u}{2} \]

   ...
call \( A \) on \([i,j] \).

Else, call \( A \) on \([l,t] \).

Start with \( A \) on \([-10^6, 10^6] \): INIALIZATION

Comments:
1. \( \max_{x \in S} f(x) \in [l, u] \) holds at the start of any call to the algorithm \( A \), because
   \( f(S \cap \{x : f(x) > t\}) = \text{YES} \Rightarrow f(S \cap \{x : f(x) > t'\}) = \text{YES} \)
   for all \( t' \leq t \).
2. In each successive call, the length of the argument interval \([l,u] \) is halved. After \( T \) calls, we have a \( \frac{10^{10} \times 2^{-T}}{2^T} \)-sized interval containing \( \max_{x \in S} f(x) \).

   \( T \approx \log \frac{1}{\varepsilon} \), we know \( \max_{x \in S} f(x) \) to \( \varepsilon \)-accuracy.

3. If the feasibility oracle also returns a feasible point \( \bar{x} \) on \( \text{YES} \). Then, can recover a point \( \bar{x} \in S \) s.t.
   \( f(\bar{x}) \geq \max_{x \in S} f(x) - \varepsilon \).

References:
1. Standardization— section 1.1 in Nemirovski
2. (Beyond this course) optimal transport— chapter 1 in Thorpe
3. (Beyond this course) extension complexity— Gerard’s survey
4. Feasibility-optimization reduction— 4.2.5 in Boyd
LECTURE 3: ALGEBRA

LPs as a proof system

t = \max c^T x \quad \text{can be interpreted as} \quad t x, \quad A x \leq b \Rightarrow c^T x \leq t.

One way to prove statements of the latter form is by combining existing inequalities with non-negative multipliers (these don't flip the sign.)

\((x_1 \leq 2) \times 3 \quad \Rightarrow \quad 3x_1 + 2x_2 \leq 14\)
\((x_2 \leq 4) \times 2 \quad \text{also valid}\)

The non-trivial / interesting bit for LPs is that such a proof (in the restricted language of multiplying existing inequalities with non-negative statements, then adding) always exists for any valid inequality. We will see an algorithmic proof. This remarkable fact is not true about mathematics in general, i.e., courtesy Godel, there are `true' but `unprovable' statements in mathematics.

Favati-Macutekkin Elimination

Rough Idea: Eliminate variables by adding more constraints, opposite of extended formulations.

An algorithm to solve linear feasibility problems, i.e., does there exist \(x \in \mathbb{R}^n\) such that \(A x \leq b\)?

Note, can use this for (high-accuracy) optimization via the optimization - feasibility reduction.

1-step of FM Elimination

Input: \(m\) inequalities of the form, \(a^T x \leq b\); \(\bar{H}, \bar{c}\)
1. Divide all inequalities into 3 sets

\[ Z = \{ \text{all inequalities that don't involve } x_1; a_{i1} = 0 \} \]

\[ P = \{ \text{all } x_i \text{ with } a_{i1} > 0 \} \]

Each can be rewritten as

\[ x_1 \leq \frac{b_i - \sum_{j=1}^{n} a_{ij} x_j}{a_{i1}} \]

Call this \( p(x_2, \ldots, x_n) \)

\[ N = \{ \text{all } x_i \text{ with } a_{i1} < 0 \} \]

Each can be rewritten as

\[ x_1 \geq \frac{b_i - \sum_{j=1}^{n} a_{ij} x_j}{a_{i1}} \]

Call this \( n(x_2, \ldots, x_n) \)

2. Construct a new feasibility problem by

(a) Copying all of \( Z \).

(b) \( \forall x \in P \not\in N \), introduce \( n(x_2, \ldots, x_n) \leq p(x_2, \ldots, x_n) \)

Thus we no longer contain \( x_1 \).

Rearrange (b) into \( a^T x \leq b \) form.

Claim: New LP is feasible iff old LP is feasible.

Proof. \( x_1, \ldots, n \) satisfies old LP. (IF)

\[ \Rightarrow x_2, \ldots, n \text{ satisfies } Z \text{ and } p(x_2, \ldots, n) \geq n \Rightarrow n(x_2, \ldots, n) \forall n \in N, \text{ PEP} \]

\[ \Rightarrow x_2, \ldots, n \text{ satisfies new LP.} \]

\( x_2, \ldots, n \) satisfies new LP. (ONLY IF)

By construction, \( x_2, \ldots, n \) satisfies \( Z \).

Also, \( \max_{n \in N} n(x_2, \ldots, n) \leq \min_{\text{PEP}} p(x_2, \ldots, n) \).

(Recall \( \max_{\text{PEP}} \phi = -\infty, \min_{\text{PEP}} \phi = +\infty \).

Choose \( n_1 \in [\max_{n \in N} n(x_2, \ldots, n), \min_{\text{PEP}} p(x_2, \ldots, n)] \)

arbitrarily.


Comments. 1. n-steps of FM elimination solves any feasibility problem. At termination, we either have all tautological inequalities or a contradiction.

2. If old LP has m constraints, new LP has \( m^2 \) constraints. Therefore, the transcript produced by the algorithm (over n-steps), and hence the running time is \( \propto n^{2n} \).

\[
\begin{align*}
& m \text{ ineq} \rightarrow m^2 \text{ ineq} \rightarrow m^4 \text{ ineq} \rightarrow m^8 \text{ ineq} \\
& n \text{ var} \rightarrow n-1 \text{ var} \rightarrow n-2 \text{ var} \rightarrow n-3 \text{ var}
\end{align*}
\]

So, FM is largely a conceptual algorithm.

3. If \( A, b \) only contain rationals, then
\[
Ax \leq b \text{ is feasible } \Rightarrow \exists x \text{ rational feasible.}
\]
Why? Because FM only creates inequalities with rational coefficients, given rational \( A, b \).

Observation: Any new inequality produced during FM is done by combining existing ones with non-negative coefficients.

\[
\begin{align*}
(1) \quad a_1 x_1 + \sum_{i=2}^{n} a_i x_i & \leq b \quad (a_i > 0) \\
(2) \quad a_1' x_1 + \sum_{i=2}^{n} a_i' x_i & \leq b' \quad (a_i' < 0)
\end{align*}
\]

Some as

\[
\frac{1}{a_1} x(1) + \frac{1}{a_1'} x(2)
\]

Farless' \( Ax \leq b \) is infeasible iff \( \exists u \geq 0, u^T A = 0, u^T b < 0 \).

Lemma.

Interpretation. If \( Ax \leq b \), then for any \( u \geq 0 \), \( u^T A x \leq u^T b \).

So if \( \exists u \geq 0, u^T A = 0, u^T b < 0 \), that implies \( Ax \leq b \) is infeasible. Because, otherwise \( 0 = u^T A x \leq u^T b < 0 \), a contradiction.
'w' is therefore a certificate of infeasibility; its existence guarantees infeasibility of $Ax \leq b$. The interesting but is that such a 'liberal' certificate always exists whenever $Ax \leq b$ is infeasible. In this sense, the linear inequality proof system is complete, not just sound.

Proof. Based on the correctness of FM, for any infeasible system $Ax \leq b$, FM must terminate in a contradiction of the form $0 \leq b_0$, where $b_0 < 0$. By the last observation, FM implicitly produces a vector $u \geq 0$ such that $u^TA = 0$ and $u^Tb = b_0 < 0$.

This is a central result in the theory of LPs, and only a step away from LP duality itself. We will complete this later.

3 VIEWS OF LPs

will assume $P = \{ x : Ax = b, x \geq 0 \}, \max_{x \in P} c^T x$, where

(a) $Ax = b$ has at least one solution, or $b \notin \text{COLSP}(A)$. Else, $P$ is infeasible/empty.

(b) Rows of $A$ are linearly independent. ($m \leq n$).

OPTIMIZATION

Defn. $x \in P$ is a VERTEX if $\exists c \geq c^T y \forall y \in P - \{ x \}.$

Defn. $F \subseteq P$ is a $k$-dimensional face of $P$ if

(1) $Fx_0$ and $k$-dimensional subspace $V$ such $P \subseteq x_0 + V$,
(2) $\exists c, z$ such $c^T x = z \forall x \in P$, $\forall y \in P - F$, $c^T y < z$.

$F$ is a proper $k$-dim face of $P$ if it is a $k$-dim face, but not a $(k-1)$-dimensional face.

A VERTEX is a 0-dimensional proper face.

EDGE is a 1-dimensional proper face.
GEOMETRIC

**Def.** \( x \in \mathcal{P} \) is an EXTREMAL POINT if \( \nexists u + v, \lambda \in (0,1) \) such that \( \lambda u + (1-\lambda)v = x \).

ALGEBRAIC

**Def.** \( x \in \mathcal{P} \) is a BASIC FEASIBLE SOLN if \( \exists B \subseteq [n], |B| = m \) such that \( A_B \in \mathbb{R}^{m \times m} \) is invertible and \( x_B = 0 \).

**Notation:** For any \( S \subseteq [n] \),
(a) \( \overline{S} = [n] - S \), complement of \( S \)
(b) \( A_S \) is a \( |S| \times |S| \)-sized matrix composed of columns whose indices are in \( S \).
(c) \( x_S \) is \( |S| \)-sized vector composed only of coordinates whose indices are in \( S \).

For a vector \( x \), \( \text{Supp}(x) = \{ j : x_j \neq 0 \} \).

* Notice that every \( B \) can correspond to at most one BFS.

Given \( B \), \( A_B^{-1} b \in \mathbb{R}^m \) extended to \( \mathbb{R}^n \) by padding with \( 0's \) on \( \overline{B} \) is the only possible candidate for BFS, but it's possible that \( A_B^{-1} b \geq 0 \) fails.

These 3 views are equivalent.

**Theorem.** \( x \in \mathcal{P} \) is a vertex \( \iff \) it is an extreme point \( \iff \) it is a BFS.

**Proof.** \( V \Rightarrow E \)

\( x \) is a vertex. \( \exists c \) such \( c^T x > c^T y \forall y \in \mathcal{P} - x \). Say \( \exists u + v, \lambda \in (0,1) \) such \( x = \lambda u + (1-\lambda)v \). But then \( c^T x = \lambda c^T u + (1-\lambda)c^T v \).

This is a contradiction because \( u \perp x \) and \( v \perp x \), hence \( c^T u < c^T x \) and \( c^T v < c^T x \).

\( E \Rightarrow BFS. \)

\( x \) is an extreme point. Recall \( \text{supp}(x) = \{ j : x_j > 0 \} \).

**Case A.** \( A_{\text{supp}(x)} \) has linearly independent columns.

Implies that \( |\text{supp}(x)| \leq m. \) Since \( x_{\text{supp}(x)} = 0 \), it is tempting to think \( B = \text{supp}(x) \) concludes this case. This almost
Here's a fun: some rows of $A$ are linearly independent, it's possible to construct $B$ by starting with $\text{supp}(x)$, and then expanding this set incrementally till it includes $m'$ indices by choosing column of $A_B$ that are linearly independent of that of $A_B$. At the end, we have $B \subseteq \text{supp}(x)$, $|B| = m$, $A_B$ is invertible. Finally, since $B \subseteq \text{supp}(x)$, $\pi_B = 0$. Hence, $x$ is a BFS.

**CASE B:** $A(\text{supp}(x))$ has linearly dependent columns.

$\Rightarrow \exists w, A(\text{supp}(x))w = 0, \ w \neq 0$. By padding $w$ with $0$'s we can construct $\tilde{w} \in \mathbb{R}^n$, $\tilde{w} = w$, $\tilde{w}(\text{supp}(x)) = 0$, $A\tilde{w} = 0$.

Define $y_+ = \pi + \varepsilon \tilde{w}$, $y_- = \pi - \varepsilon \tilde{w}$. Note $\pi = \frac{y_+ + y_-}{2}$; yet $y_+ \neq y_-$ for any $\varepsilon > 0$ since $\tilde{w} \neq 0$. Also $A_y = Ax + A\tilde{w} = B$ and $A\pi = B$. So, if we can ensure $y_+, y_- > 0$, then $y_+, y_- \in \mathbb{P}$ implying we have reached a contradiction. Choose $0 < \varepsilon \leq \frac{\min x_i}{\max x_i} \ |w_i|$, observe $\text{supp}(w) \subseteq \text{supp}(x)$ to conclude $y_+, y_- > 0$.

**BFS $\Rightarrow V**

$r$ is a BFS. $\exists B$, $|B| = m$ such $A_B$ is invertible and $\pi_B = 0$.

Note that $A_B \pi_B = B$. Construct $c \in \mathbb{R}^n$ such $c_j = \frac{\pi - y_j}{y_j} - 1$ if $j \in B^c$

Now, $c^T x = 0$. Notice for any $y \in \mathbb{P}$, since $y \geq 0$, $c^T y \leq 0$.

We'll prove if $c^T y = 0$ then $y = \pi$ to establish that $r$ is a vertex. Consider $y \in \mathbb{P}$ such that $c^T y = 0$. $y_B = 0$ and $A_B$ is invertible. $y$ is a BFS. But then for $B$ can correspond to at most one BFS, $\therefore y = \pi$. $

**References:**

1. Fourier-Motzkin Elimination—
   1. I like 3.1 and 3.2 in Gerard's *book*; includes proof of Farkas' Lemma
   2. Also section 6.7 in *Matousek*
   3. Alternative: Section 2.8 in *Bertsimas*
2. BFS-Vertex-Extreme Equivalence—
   1. Section 2.2 and 2.3 in *Bertsimas*
   2. Chapter 4 in *Matousek*
Since every $B \subseteq [n]$ of size $m$ corresponds to at most one BFS.
The number of BFSs is at most $\binom{n}{m}$. Recall that we are
delaying at LPs of the form:

$$\max \quad C^T x$$
$$s.t. \quad P = \begin{cases} \text{Ax=b} \Rightarrow & 0 \text{ Ax=b has at least one solution, i.e. be in \text{Col}(A)}. \\ x \geq 0 \end{cases}$$

- $\text{Ax=b}$ has at least one solution, i.e. be in $\text{Col}(A)$.
- Rows of $A$ are linearly independent.

The next result we prove states that any LP of this form
chooses between one of these three cases:

1. The LP is infeasible, i.e. $\max_{x \in P} C^T x = -\infty$.
2. The LP has unbounded optima, i.e. $\max_{x \in P} C^T x = +\infty$.
3. A (finite) optima exists, and a BFS $z$ achieves it.

Implicitly, we have the following finite-time algorithm:

**SOLVING LPs WITH BOUNDED OPTIMAL VALUE BY ENUMERATION**

1. For each $B \subseteq [n]$ of size $m$, solve $x_B = A_B^{-1} b$. Check if $x_B \geq 0$.
   - On 'YES', set $x_B = 0$ and add $z$ to the set of BFSs.

2. If no BFSs are found, output INFEASIBLE, else output
   the highest objective value achieved by any BFS.

This takes $\mathcal{O}(\binom{m}{n} \times \text{poly}(n))$ time. The simplex algorithm reuses
this idea, but searches for BFSs greedily.

**FUNDAMENTAL THEOREM OF SIMPLEX.**

In any feasible LP with bounded optima

\( \text{In any feasible LP with bounded optima} \)

- If $z$ is an extreme point

- $\exists$ a BFS $z$ that achieves the optimal value.

Let's start with a somewhat seemingly unrelated observation.

**Lemma.** Every feasible LP in the standard form has an extreme point.

**Proof.** Since $P = \{ x : Ax=b, x \geq 0 \}$ is feasible, choose $z$ to be
any feasible point with the smallest number of non-zero entries. We claim $z$ is an extreme point. If not, $\exists u \in \text{ep}\left( x \right)$

\[ \{ u \} \cup \{ (1-v)u : v \in [0,1] \} \] such that $z = u + (1-v)u = x$. Since $u, v \geq 0$, we have

\[ \text{supp}(u), \text{supp}(v) \subseteq \text{supp}(z) \], where there can be no coordinate cancellations.

We claim $\exists j \in [n]$ such that $u_j < v_j > 0$, since $u+v$

and if $u - v$ is all negative, we relabel $(u, v, z) \longleftrightarrow (z, u, v, 1-z)$. 

Now, consider \( y = x - \varepsilon (u - v) \). Choose \( \varepsilon = \min_{j : (u - v)_j > 0} \frac{x_j}{(u - v)_j} \).

Now, \( Ay = Ax - \varepsilon (Au - Av) = b \), and \( y \geq 0 \) since \( \text{supp}(u - v) \subseteq \text{supp}(x) \).

Let \( y \) have at least one fewer nonzero coordinate than \( x \). Hence, \( x \) must be an extreme point.

Because of BFS-V-E equivalence, we prove the fundamental theorem of simplex by proving the existence of an optimal extreme point.

**Proof of Fund-Thm. of Simplex.**

Since \( \max_{x \in P} c^T x = \frac{1}{2} x : Ax = b, x \geq 0, x \geq 0 \) is feasible and has bounded optima, \( \exists * \in IR \)

\( \max_{x \in P} c^T x = * \) (consider \( P = \{ x : Ax = b, c^T x = *, x \geq 0 \} \)).

We know \( * \) an extreme point \( x \) in \( P \). Note that \( c^T x = * \) and \( x \in P \) since \( Q \subseteq P \). So, it only remains to show such an \( x \) is extreme in \( P \) too. Suppose not. Then, \( \exists u \in \text{int} P \), \( \lambda (0,1) \)

such that \( x = \lambda u + (1 - \lambda) v \). Since \( x \) is extreme in \( Q \), at least one of \( u, v \) must not lie in \( Q \). Hence, \( \min \{ c^T u, c^T v \} < * \).

Also, \( \max \{ c^T u, c^T v \} \leq * \). But \( c^T x = A c u + (1 - \lambda) c v \)

implying \( c^T x < * \). Hence, \( x \) is extreme in \( P \).

**Applications of the Fundamental Theorem**

1. **Affine Hull** \( S = \{ x : \sum_{i=1}^{k} \lambda_i x_i : \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0, x_i \in S \} \)

2. **Conic Hull** \( S = \{ x : \sum_{i=1}^{k} \lambda_i x_i : \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0, x_i \in S \} \)

3. **Convex Hull** \( S = \{ x : \sum_{i=1}^{k} \lambda_i x_i : \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0, x_i \in S \} \)

The next theorem are on the sufficiency of minimal representations: sure you can represent any \( x \in \text{Conic Hull}(S) \)

as non-negative combinations of finitely many points in \( S \), but from many are needed in the worst-case?

**Caratheodory’s** \( x \in \text{Conic Hull}(S) \), where \( S \subseteq \mathbb{R}^n \), then \( \exists x_1, \ldots, x_n \in S \)

**Theorem for Cones**

Prove: Since \( x \in \text{Conic Hull}(S) \), \( \exists x_1, \ldots, x_k \in S \) such that \( x = \sum \lambda_i x_i \).

Let \( x = [x_1, \ldots, x_k] \) be the matrix whose columns are \( x_i \)'s. Then \( P = \{ x : x = \lambda x, \lambda > 0 \} \) is feasible. Consider \( \max_{x \in P} c^T x \). Then \( \exists \)
a BFS \( \lambda \) in \( P \) with at most \( n \) number of equality constraints, many non-zero elements, proving the theorem.

CARATHEODORY'S THEOREM FOR CONVEX HULLS

If \( x \in \text{CONVEX-HULL}(S) \), where \( S \subseteq \mathbb{R}^n \), then \( \exists x_1, \ldots, x_{n+1} \in S \) such that \( x \in \text{CONVEX-HULL}(\{x_1, \ldots, x_{n+1}\}) \).

**Proof:** This time we have that \( P = \{x \in \mathbb{R}^n : \lambda^T x = 1, \lambda \geq 0, \lambda^T 1 = 1\} \) is feasible. Hence, \( \exists \) a BFS \( \lambda \) in \( P \) with at most \( n+1 \) number of equality constraint many non-zeroes.

**Hints of Duality**

* \( \{x : Ax \leq b\} \) is a polyhedron.
* \( \{x : Ax \leq 0\} \) is a polyhedral cone.

These descriptions create a set by exclusion: each inequality rejects some subset of points in \( \mathbb{R}^n \); a non-empty subset is more than satisfy simultaneously all these checks.

* \( \text{CONIC-HULL}(\{x_1, \ldots, x_k\}) \) is a finitely generated cone.
* \( \text{CONVEX-HULL}(\{x_1, \ldots, x_k\}) \) is a polytope.

These descriptions create sets by inclusion: as long as a small subset (even 2 points) linearly combine to produce a point, it is in the set. A deep result in polyhedral theory is that these ways of constructing sets are equally powerful.

**Some Prelim.**

* Recall Farkas' Lemma: \( Ax \leq b \) is infeasible iff \( \exists x > 0, x^TA = 0, x^TB < 0 \).

**Farkas' Lemma**

\[
Ax = b \text{ is feasible iff } \forall \lambda \lambda^TA \leq 0 \Rightarrow \lambda^TB \leq 0.
\]

**Interpretation:** \( Ax = b \) is feasible iff \( b \in \text{CONIC-HULL}(\{a_1, \ldots, a_n\}) \).

The theorem says either this happens, or there is a hyperplane passing through the origin which separates \( b \) and \( \{a_1, \ldots, a_n\} \).

Either

\[
\begin{align*}
\text{either} & \quad \text{or} \\
\begin{aligned}
& a_1 \\
& \quad \text{or} \\
& a_2
\end{aligned}
\end{align*}
\]

The existence of such hyperplane holds for any 2 closed disjoint convex sets, at least one of which is compact. But for our purposes, Farkas' lemma suffices.
PROOF. (⇒) If $A x = b$ is feasible, then $\forall \lambda > 0$ $x^T A x = \lambda^T b$. If $x^T A x \leq 0$, then $x^T b = (x^T A)x \leq 0$ since $x^T A$ is a dot product between a non-negative $x$ and a non-positive $(x^T A)^T$.

(⇐) We will prove the contrapositive; i.e. $A \Rightarrow B \Rightarrow \neg A$.

If $A x \leq b$, $-A x \leq -b$, $-x \leq 0$ is impossible, then $\exists \lambda_1, \lambda_2, \lambda_3 > 0$ such that $(A_1 - \lambda_2) x - \lambda_3 = 0$ and $(A_1 - \lambda_2)^T b < 0$. We rewrite this as

$$(A_2 - \lambda_3) x = -\lambda_3 \leq 0,$$

and $(A - \lambda)^T b > 0$ to complete the proof.

* $(A, R)$ is a double description pair iff $\forall x$

$$A x \leq 0 \iff \exists \lambda \geq 0 \quad x = R \lambda.$$

**Lemma.** $(A, R)$ is a DDP iff $(R^T, A^T)$ is a DDP.

**Proof.** By symmetry, it is enough to prove one side. Say $(A, R)$ is a DDP. Then, we have for any $\lambda$, that

$$R^T x \leq 0$$

$$\iff \forall \lambda > 0 
   x^T R x = (R \lambda)^T x \leq 0$$

$$\iff \forall y A y \leq 0 \quad y^T x \leq 0$$

using $(A, R)$ is DDP.

$$\iff \exists \lambda \leq 0, A \lambda = x$$

my Farkas' Standard-form Lemma.

**Minkowski-Weyl Theorem for Cones.** Any polyhedral cone is a finitely generated cone, and vice versa.

**Proof.** We'll prove that $\forall R, \exists A$ such that $\text{CONIC HULL}(R) \exists x \iff A x \leq 0$. Take any $z$. Consider $\exists A R^T z = 0$, $z - R A x \leq 0$, $-\lambda \leq 0$.

Run Fourier-Motzkin to eliminate all $\lambda$'s. Since we start with homogenous inequalities, we continue at $\exists A x \leq 0$

for some $A$ such that the new system is feasible in $\forall A x \leq 0$ the former system is feasible in $(A, \lambda)$ for $\lambda \geq 0$.

This establishes that any finitely generated cone is a polyhedral cone. Finally, using the DDP lemma, we also get that $\forall R, \exists A$ CONIC HULL(A) $\exists x \iff R^T x \leq 0$, completing the proof.

**Minkowski-Weyl Theorem for Polyhedra.** Any polyhedra is expressible as the sum of a polytope and a finitely generated cone, and vice versa.

* $P = \{x : A x \leq b\}$, $C_P = \text{CONIC HULL}\{\{x\} : x \in P\} = \{A y - b \leq 0\}$, $y \in E_P$ $\iff \{x\} \in E_P$, from MW.
PROOF: $\Rightarrow$ If $Ax \leq b$ off $[m \times n]$ and $C_p = \text{cone}(\{a_1, a_2, ..., a_p\})$, then by MW for cones, for some $p_i, q_i$, $C_p \subseteq \text{cone}(\{a_1, a_2, ..., a_p\})$ if $x = \sum_{i=1}^{k} \lambda_i p_i + \sum_{i=1}^{m} \mu_i q_i$, and $\sum_{i=1}^{k} \lambda_i = 1$, therefore $P = \text{convex}(P_1 \cdots P_k) + \text{cone}(Q_1 \cdots Q_k)$.

$\Leftarrow$ Note or $\subseteq$ convex $(P_1 \cdots P_k) + \text{cone}(Q_1 \cdots Q_k)$ iff $x \in C_p \subseteq \text{cone}(\{p_i\} \cup \{q_i\})$. Then, by MW for cones, $\exists A, b$ such that $C_p = \{y \in \mathbb{R}^n : Ay - b \leq 0\}$.

Take $P = \{x : (x, C_p) \subseteq (\frac{1}{m} x : x \leq b)\}$. \square

* Note that $\frac{1}{m} \mathbb{R}^m$ is the bounded cone. Hence, we reach the following.

COROLLARY: Any bounded polyhedra is a polytope, and vice versa. In fact, we can characterize such polytopes.

COROLLARY: Any bounded polyhedra is a convex hull of its vertices.

PROOF: Any bounded polyhedron $P = \text{convex hull}(x_1, x_2, ..., x_p)$ for some vertices $x_i$. Iteratively delete any vertex $x_i \subseteq \text{convex hull}(x_1, x_2, ..., x_{i-1})$. Then $P = \text{convex hull}(x_1, x_2, ..., x_{i-1})$ where $x_i \subseteq \text{convex hull}(x_1, x_2, ..., x_{i-1})$.

Let $V$ be the set of all vertices of $P$. We claim $Q \subseteq V$. Else, generally say $v \notin V$. Then $\exists u \notin v \in P$, $\lambda \in (0, 1)$ such that $v = \lambda u + (1-\lambda)v$. Also, note $u = \sum_{i=1}^{n} a_i x_i$, $v = \sum_{i=1}^{n} b_i x_i$, where $a_i, b_i \geq 0$, $1^T a = 1^T b = 1$ and $a_1, b_1 < 1$. But then

$x_i = \frac{1}{1 - \lambda a_i - (1-\lambda) b_i} \sum_{i=1}^{n} (\lambda a_i + (1-\lambda) b_i) x_i$: a contradiction.

Hence, $P = \text{convex hull}(Q) \subseteq \text{convex hull}(V)$. Since $V \subseteq P$, $\text{convex hull}(V) \subseteq P$. \square

COMMENTS ON GENERAL FORM LPs

* No good reason why LPs $Ax \leq b$ should have sparse solutions at all.

* Modified BFS definition: $x$ is a BFS for $Ax \leq b$ ($x \in \mathbb{R}^n$) if there are at least $n$ active/tight constraints at $x$, with linearly independent $a_i$'s.

* But consider: max $c^T x$. It's feasible, has bounded optima.

Yet max $c^T x$ has no extreme points/vertices/BFS.

s.t. $CTx = \delta$ $-CTx \leq -\delta$
A set $S$ is pointed if it does not contain a line (extending infinitely in both directions), i.e. if $\mathbb{R}^n$ such that $\mathbb{R} \times S, \mathbb{R} + A \times S$.

Fundamental Theorem of Simplex: For feasible LP in the general form with a pointed feasible set and bounded optima, $\exists$ a BFS which attains the optimal value.

References:
1. Optimality of BFSs—
   1. Section 4.2 in Matousek
2. Minkowski Weyl Theorems—
   1. Section 3.5 in Gerard’s book
   2. Section 3.5 in Fukuda
3. Results on LPs in general form
   1. Section 2.2 and 2.3 in Bertsimas
LECTURE 5: DUALITY

Heuristically: $\max_x c^T x \begin{array}{c} \text{Computing} \\ \text{Primal} \end{array} \leq b \begin{array}{c} \text{Duals} \end{array}$

$= \max_x \min_{y \geq 0} \begin{cases} c^T x + y^T (b - Ax) \\ y \geq 0 \end{cases}$

$= \min_{y \geq 0} \max_x y^T b + x^T (c - A^T y)$

$= \min_{y \geq 0} b^T y \begin{array}{c} \text{Dual} \end{array}$

$\min c^T x \begin{array}{c} x \geq 0 \end{array} \begin{array}{c} \Rightarrow \text{maximize} \begin{cases} \begin{array}{c} b^T \lambda + x^T (c - A^T \lambda) \\ \lambda \geq 0 \end{cases} \end{array} \Rightarrow \lambda \geq 0 \end{array}$

$\Rightarrow \max b^T \lambda \begin{array}{c} \text{Dual} \end{array}$

We'll make the questionable steps concrete by the end of the lecture.

*Notice that dual of the dual of a LP is the LP itself.*

**Interpretation:** Recall the diet problem. $\min c^T x \begin{array}{c} A \geq b \begin{array}{c} \text{B} \end{array} \begin{array}{c} \text{Dual} \end{array} \text{y \geq c} \end{array}$

**Weak Duality** Any feasible $y$ in $\text{max } b^T y \begin{array}{c} \text{Dual} \end{array}$ provides a $\begin{array}{c} \text{Duality} \end{array}$ lower bound on value of $\min c^T x \begin{array}{c} \text{Primal} \end{array} \begin{array}{c} \Rightarrow \text{maximize} \begin{cases} \begin{array}{c} b^T \lambda + x^T (c - A^T \lambda) \\ \lambda \geq 0 \end{cases} \end{array} \Rightarrow \lambda \geq 0 \end{array}$

Proof: Take any feasible $x, y$; if primal is infeasible, any real $< \infty$.

Then, $c^T x \geq (A^T y)^T x = y^T A x = y^T b$ since $x \geq 0$, $c - A^T y \geq 0$.

Implication: If dual is unbounded, then primal is infeasible & vice-versa.

Weak duality is not an accident. Whenever we exchanged the orders of $\min / \max$ operators in the heuristic derivation, there's a consistent assignment of $\geq / \leq$ that held consistently.

Take $f : X \times Y \rightarrow R$. Clearly, $\forall x \in X \forall y \in Y$, $\max_{y \in Y} f(x, y) > f(x, y)$.

Then, $\forall y \in Y$, $\min_{x \in X} \max_{y \in Y} f(x, y) \geq \min_{x \in X} f(x, y)$, hence $\max_{x \in X} \min_{y \in Y} f(x, y)$.
Any PRIMAL/DUAL pair suffers from 1 of 4 fates:
(1) Both P & D are infeasible.
(2) P is infeasible, D has unbounded optima.
(3) D is infeasible, P has unbounded optima.
(4) P is feasible and has bounded optima. Then D is feasible and has bounded optima. Further, the optimal values of P & D match.

**Strong Duality**

**Thm.**

If \( \text{min} \ \frac{c^T x}{A x = b \quad x \geq 0} \) is feasible & has bounded optima, then \( \text{max} \ \frac{b^T y}{c^T A y} \) is feasible and

\[
\text{min} \ \frac{c^T x}{A x = b \quad x \geq 0} = \text{max} \ \frac{b^T y}{c^T A y} \quad c \geq A^T y.
\]

**Proof.** Let \( v^* = \text{min} \ \frac{c^T x}{A x = b \quad x \geq 0} \). We will prove \( y \) such that \( A^T y \leq c \)

\( b^T y \geq v^* \).

This is enough since for any feasible \( y \), \( b^T y \leq v^* \) by weak duality. Let's assume \( y \) such \( A^T y \leq c \). Then, by Farkas' Lemma,

\( \exists \lambda \geq 0 \) such \( \lambda^T A^T - \lambda_0 b^T = 0 \) and \( \lambda^T c = -\lambda_0 v^* < 0 \).

(Or \( \lambda A = \lambda_0 b \)) (Or \( \lambda^T c = \lambda_0 v^* \))

Case A: If \( \lambda_0 > 0 \), then \( \frac{\lambda^T c}{\lambda_0} < V^* \).

Consider \( \tilde{x} = x + \lambda \). Then, \( \tilde{x} \geq 0 \), \( A^T \tilde{x} = A x + A \lambda = b + D = b \)

and \( c^T \tilde{x} = c^T x + \lambda^T c < V^* \). A contradiction.

Case B: If \( \lambda_0 = 0 \), then take any feasible \( x^* \) with \( c^T x^* = V^* \).

Although we are using Farkas' Lemma here, morally, strong duality is an 'obvious' consequence of the completeness of the Fourier-Motzkin algorithm in deriving valid linear inequalities. Concretely

\[ \max \ \frac{c^T x}{A x \leq b} = v^* \]

is equivalent to \( \forall x A x \leq b \implies c^T x \leq v^* \). If the last implication is true, FM can prove it by combining some of \( A x \leq b \)

with non-negative multipliers \( y \geq 0 \). If so, \( y^T A = c \) & \( y^T b \leq v^* \).

Also, for any such \( y \), by weak duality we have \( y^T b \leq v^* \).
**Theorem on Com. Slackness**

If \( x^* \) maximizes \( c^T x \) s.t. \( Ax \leq b \), \( y^* \) minimizes \( b^T y \) s.t. \( A^T y = c \), then

\[ y_i \in \{0\}, \quad y^*_i (a^T x^* - b_i) = 0. \]

In words, \( y^*_i \) can be positive only when the corresponding inequality \( a^T x^* \leq b_i \) is tight for \( x^* \).

**Proof:** By strong duality, \( 0 = c^T x^* - b^T y^* = y^*_T (A^T x^* - b) = 0. \)

Yet, \( y^* > 0 \) and \( A^T x^* - b \leq 0 \). Each term in the dot product is non-positive, yet the sum is zero \( \Rightarrow \) each term is zero.

**Geometric Interpretation:** \( A^T y = c \) is same as \( c \in \text{cone} (a_1, \ldots, a_m) \) or \( y \geq 0 \).

In fact, Strong Duality is same as \( (c, v^*) \in \text{cone} (\text{supp}(a_i) \text{ for active } a^T x^* \leq b_i). \)

**Application One: Robust LPs**

(c, A, b) are inputs to a LP. But do we know these to absolute certainty? Often not. Say we know \( c \in C \), \( a_i \in Ua_i \), \( b_i \in Ub_i \). Can we optimize for the worst-case value?

\[
\min_{x} \max_{c \in C} \min_{x, z} c^T x + \epsilon \leq 0 \quad \text{w.r.t. } \epsilon \leq Ub_i \text{ and } a_i \leq Ua_i
\]

Thus we can assume w.l.o.g. that there's uncertainty in \( c \).

Similarly, we can get rid of uncertainty in \( b_i \), by choosing \( b_i = \min b_i \) regardless of \( v \), since \( a^T x \leq b_i \Rightarrow a^T x \leq b_1 + b_2 + \cdots + b_i \).

So, there's just \( Ua_i \) left. Let's stick to polyhedral uncertainty: \( Ua_i = a_i : Di_a \leq d_i \). A priori, this is a LP with uncountably infinite constraints. Can we solve it?

Assume for simplicity that \( Ua_i \) is bounded & feasible.
\[
\begin{align*}
\min_{x} & \quad c^T x \\
\text{s.t.} & \quad a^T x \leq b_i, \quad \forall i \in [m], \quad D_{i} a_i \leq d_i
\end{align*}
\]

By duality, \(\max a^T x = \min D_i^T p_i \quad \text{s.t.} \quad D_i^T p_i = x, \quad \pi_i \geq 0\).

But now, we have
\[
\begin{align*}
\min_{x} & \quad c^T x \\
\text{s.t.} & \quad \min_{\pi_i} \left[ \max_{a_i} a_i^T x \right] \\
\text{s.t.} & \quad D_i^T p_i = x, \quad \pi_i \geq 0
\end{align*}
\]

Prove the last equivalence.

Finally, we have an LP & can solve relaxations LPs with polyhedral uncertainty efficiently.

CAUTIONARY TALE: BILEVEL LPs

Generally, LPs embedded inside other LPs is a recipe for computational intractability. We got lucky above.

General Bilevel LP
\[
\begin{align*}
\max_{x} & \quad c^T x + d^T y \\
\text{s.t.} & \quad A x + B y \leq f \\
& \quad y^* = \arg\max_{y} d^T y \\
& \quad \text{s.t.} \quad C x + D y \leq g
\end{align*}
\]

We will embed Knapsack in a bilevel LP to demonstrate that solving general bilevel LPs is NP-hard.

Knapsack is NP-hard.
\[
\begin{align*}
\max & \quad \sum_{i=1}^{n} a_i x_i \\
\text{s.t.} & \quad \sum_{i=1}^{n} a_i x_i \leq B \\
& \quad x_i \in \{0, 1\}, \quad i \in [n]
\end{align*}
\]

We will embed Knapsack in a bilevel LP to demonstrate that solving general bilevel LPs is NP-hard.

Integral to Switching Constraints
\[
\begin{align*}
\max & \quad \sum_{i=1}^{n} a_i x_i - 10^{100} \sum_{i=1}^{n} y_i \\
\text{s.t.} & \quad \sum_{i=1}^{n} a_i x_i + 10^{100} \sum_{i=1}^{n} y_i \leq B \\
& \quad y_i = \min \{ x_i, 1 - x_i \} \\
& \quad 0 \leq x_i \leq 1
\end{align*}
\]

Ex)

Prove that if \(\max a_i \leq 10^{100}\), then optima of switching formulation and knapsack coincide.
APPLICATION TWO: TWO PLAYER ZERO SUM GAMES

<table>
<thead>
<tr>
<th>R</th>
<th>P</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,0</td>
<td>-1,1</td>
<td>1,-1</td>
</tr>
<tr>
<td>1,-1</td>
<td>0,0</td>
<td>-1,1</td>
</tr>
<tr>
<td>-1,1</td>
<td>1,-1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

Let $A$ be a matrix of payoffs for the column player.

If row player goes first, $\min_{i} \max_{j} A_{ij}$.

If column player goes first, $\max_{j} \min_{i} A_{ij}$.

Payoffs of row player, column player. These are clearly not equal, note their sum to zero, i.e. $+1 \neq -1$.

If players are permitted to choose randomized/mixed strategies, then order of play is irrelevant.

\[ \min_{x \in \Delta} \max_{y \in \Delta} z^TAy = \max_{y \in \Delta} \min_{x \in \Delta} z^TAy. \]

**Proof**

Skelet

\[ \min_{x \in \Delta} \max_{y \in \Delta} z^TAy = \max_{y \in \Delta} \min_{x \in \Delta} z^TAy = \max \min_{y \in \Delta} (A_{j})_i \]

\[ \min \max (A^T)_{ij} \]

**P = min \ z \ s.t. \ z \geq A^T x \ & \ z \geq \max_{i,j} |a_{ij}| \]

**D = max \ w \ s.t. \ w 1 \leq Ay \ & \ w \leq -\max_{i,j} |a_{ij}| \]

Therefore, by strong duality, enough the primal $P$ and dual $D$ are equal.

**Verify this via explicit computation. More subtle point: $P$ & $D$ are automatically duals since this is exactly how we derive duals.$\Box$

References:

1. Computing duals—Section 6.2 in Matousek; also see this.
2. Via minimax inequality—Sections 5.2.1 and 5.4 in Boyd.
3. Proofs of Strong Duality—Lecture 5 from Amir Ali’s course notes are the tidiest; also discusses robust LPs.
4. Section 3.3 in Gerard’s book provides a direct proof via FM.
5. (Beyond this course) More on robust programs by Nemirovski.
6. Zero sum games—Section 5.2.5 in Boyd.
7. Hardness of Bilevel LPs—this paper.
Let us recall the BFS enumeration algorithm from **Geometry**. The simplex algorithm searches for an optimal BFS in a local manner. Questions that arise:

1. What is this notion of locality?
2. Why does local search lead to optimality?
3. How to even select a feasible BFS? Recall that checking for feasibility is almost as hard as optimization itself.

**Answer 1.** Recall that each $B(x)$ of size $m$ can correspond to at most one BFS in $\{Ax=b, x \geq 0\}$. We can draw a graph over the $B$'s that yield a (feasible) BFS.

![Algebraic Graph](image)

$B - B'$ iff $|B_n B'| = m - 1$, i.e., if $B$ and $B'$ share all but one coordinate.

Roughly, simplex searches for better (or next worse) neighbors in this graph.

Note multiple $B$'s might correspond to same BFS. This creates complications later. To keep things non-degenerate: All BFSs have $m$ non-zero coordinates, or equivalently, each feasible $B$ produces a unique BFS.

![Geometric Graph](image)

$V - V'$ are connected iff they are connected by an edge, i.e., a $1$-dimensional face.

Under non-degeneracy, **algebraic graph** = **geometric graph**.

**Graphically,** the geometric graph is an edge contraction of the algebraic graph, formed by collapsing $B$'s that lead to the same BFS. By its nature, simplex performs local
search over the algebraic graph. But local search over the geometric graph is easier to analyze.

Answer 3. In short, by cheating. Construct an auxiliary LP for which a BFS is easy to guess.

\[
\begin{align*}
\text{(ORIGINAL LP)} & \quad \min c^T x \quad \text{(AUXILIARY LP)} & \quad \min \frac{1}{2} w^T S \\
x & = Ax = b \quad x & = Ax + s = b \\
& x \geq 0 \quad x \geq 0 \\
& \text{Without loss of generality, } b > 0. \quad s \geq 0 \\
& \text{Else, flip sign.}
\end{align*}
\]

Observation: Original LP is feasible \(\iff\) Auxiliary LP's optimum is 0.  
Idea: Run simplex on Aux LP starting with \(x = 0, s = b\)  
(which is a BFS). If opt > 0 or unbounded, original LP is infeasible. Else, we end up with a BFS for original LP to run simplex on (maybe after using duality expansion from LECTURE 3).

Answer 2: Okay, this is a bit unwelcome/annoying.

Assuming non-degeneracy, simplex:
1. Start with a BFS.
2. Check if there's a neighboring BFS with strictly better value. If so, then move to it & repeat.
   Else, declare current BFS is optimal.

Consider any neighbor of \(B\) with BFS \(x\), \(B'\) with BFS \(y\). Let \(B' - B = \xi i \beta\), \(d_B = y_B - x_B\). Then, \(A_B (a_B + d_B) + a_i y_B = b\)
or \(A_B d_B + a_i y_B = 0\), where \(a_i\) is \(i\)th column of \(A\). Also,
\[
c^T y = c^T x = c_i y_i + c^T_B d_B = (c_i - c^T_B A_B^{-1} a_i) y_i.
\]

Observation: \(c^T x \leq \min_{y \in \text{neighbor of } x} c^T y\) for \(B\) generating BFS \(x\).

\[
\iff (c_i - c^T_B A_B^{-1} a_i) y_i \geq 0 \quad \text{if } i \notin B.
\]

where \(y\) is a BFS for \(B'\) such that \(B' - B = \xi i \beta\).
\[ C_i - C_B^T A_B^{-1} A_i \geq 0 \quad \forall i \in [n] \text{ since } y_i \geq 0. \]

**Definition:** The reduced cost at \( B \)

\[ C_i - C_B^T A_B^{-1} A_i \geq 0 \quad \forall i \in [n] \text{ since } y_i > 0. \]

For non-degenerate LPs

Here, we are using that \( \forall i \in B, \quad C_i = C_B^T A_B^{-1} A_i = C_B^T e_i, \)
using definition of matrix inverse.

**Theorem:** Define \( \bar{C} = C^T - C_B^T A_B^{-1} A \) to be the reduced cost at \( B \).

Then, (1) \( \bar{C} \geq 0 \Rightarrow \text{BFS } \bar{r} \text{ at } B \text{ is optimal}; \)
(2) \( \text{BFS } \bar{r} \text{ at } B \text{ is optimal and LP is non-degenerate } \Rightarrow \bar{C} \geq 0. \)

**Proof:** Let’s start with (2). If \( \bar{x} \) is optimal, it must lie at least as good as its neighbors. Then for non-degenerate LPs, \( \bar{C} \geq 0 \), by the previous observation.

For (1), we’ll certify optimality by constructing a dual feasible solution. \( \bar{C} \geq 0 \iff \bar{C}^T \geq C_B^T A_B^{-1} A = y^T A \), where \( y = C_B^T A_B^{-1} b \), or \( A^T y \leq C \). Thus \( y \) is a dual feasible solution. Yet, \( b^T y = C_B^T A_B^{-1} b = C_B^T b_x = C^T x \). Hence, \( \bar{C} \) makes \( \bar{x} \) optimal.

**Corollary:** For non-degenerate LPs, simplex with strictly better neighbor will terminate in a finite number of steps and reaches an optimum.

This corollary is immediate since no zero improvement at each step implies no vertex/BFS is visited twice, and recall that there are at most \( n \) of them.

The last theorem guarantees optimality at stopping.

**Simplex for Possibly Degenerate LP.**

1. Start at some BFS \( \bar{r} \) with base \( B \).
2. Check if \( \bar{C} \geq 0 \). If so, declare optimality. Else,
choose a neighboring vertex \( B \) such that \( B' = B + \frac{1}{3} \mathbf{z} \)

such that \( \mathbf{c}_i < 0 \), using a PIVOTING RULE. Move to it; repeat.

Now, that we have given up the invariant of strict improvement every step, there’s the possibility that simplex cycles (visits the same box \( B \) twice) and never terminates. Recall that asking for strict improvement (and stopping when it is not possible) impugns on the correctness / optimality; that’s worse. Many natural pivoting rules for simplex cycle.

**Bland’s Rule:** When at a base \( B \), choose the smallest index \( i \) for which \( \mathbf{c}_i < 0 \); move to \( B' = B + \frac{1}{3} \mathbf{z} \).

**Theorem:** Simplex with Bland’s rule does not cycle, and hence, terminates at an optimum in finite steps.

We will not prove this in interest of time. Maybe in a future iteration of this course. Generally, seek lexicographic (not invariant to naming / order of indexing) rule provide a consistent way of tie-breaking.

**Comments of Running Time of Simplex**

* A reasonable algorithm for LPs arising in practice.
  Mixed evidence on if cycling is a real concern.

* Many many pivoting rules require an exponential number of steps in the worst-case.
  Klee-Minty rule & variants are a common source of such hard examples.

* Multiple decades-long push to find a pivoting rule that results in polynomial complexity.
Q: Is this even true for an "ordinary" pivoting rule?

For a polytope $P$, let $G_P$ be its BFS graph.

**Hirsch Conjectures** (1957)

$diam(G_P) \leq n-d$ where $P$ is a $d$-dimensional polyhedron with $n$ constraints.

True when $d \leq 3$

True when $n-d \leq 6$

**Counterexample** (2010 Santos)

$+$ a counterexample to the conjecture in $d=43$.

**Theorem** (Kalai-Kleitman)

$diam(G_P) \leq 2n \log_2 d + 1$

$diam(G_P) \leq \text{poly}(n,d)$?

* Concession: Maybe worst-case complexity is really bad, but perhaps average-case is good. Caution: notion of "average" in average-case is tricky. For example, an early result of this kind proved that if $(a_i,b_i) \sim D$ are sampled independently from some distribution $D$ satisfying equipodal symmetry, i.e. if $Pr_D((a,b)) = Pr_D((-a,-b))$, then with $n$ such constraints in $d$ dimensions, simplex can max $C^T x$ s.t. $Ax \leq b$ takes polynomially many steps with high probability. This is difficult to judge the significance of since for $m > 2n$, such LPs are infeasible with high probability, for most (non-degenerate) distributions.
* Smoothed analysis*: In 2002, Spielman & Teng showed that given any \((A, b)\), with
\[
\begin{align*}
\tilde{A} &= A + \text{random noise of size } \sigma \\
\tilde{b} &= b + \text{random noise of size } \sigma,
\end{align*}
\]
the simplex algorithm takes poly \((n, d, 1/\sigma)\) steps on \(\max C^T x \leq \tilde{b} \). Notice that although this is a statement about random instances, the randomness is very “localized”. Submode of analysis between worst-case and average-case is called Smoothed Analysis, and has proven useful in studying efficiency of algorithms in beyond worst-case settings more generally.

References:

1. Finding an initial BFS— page 70 in [Matousek](#)
2. Simplex algorithm
   1. Sections 3.1 and 3.5 in [Bertsimas](#)
   2. Section 11.1 in [Schrijver](#); also proves termination of Bland’s rule
3. (Beyond this course) [Survey](#) on Hirsch Conjecture
4. (Beyond this course) Smoothed analysis— Daniel Dadush’s [talk](#)
LECTURE 7: CENTER OF MASS

For any compact $K \subseteq \mathbb{R}^n$, $\text{vol}(K) = \int_K d\nu$, $\text{com}(K) = \int_K \frac{d\nu}{\text{vol}(K)}$.

Note $\text{com}(K) = \mathbb{E}_x [x]$. 

Take any $\min C^Tx$; this can represent any convex program.

ALGORITHM

1. Set $K_1 = K$. 
2. For $t = 1, \ldots, T$
   
   Compute $x_t = \frac{1}{\text{vol}(K_t)} \int_{K_t} d\nu$. 

   Take $K_{t+1} = K_t \cap \{x : C^Tx \leq C^Tx_t\}$. 

3. Output $\overline{x} = \arg\min_{x \in \{x_1, \ldots, x_T\}} C^Tx$. 

CLAIM. If \( \max_{x,y \in K} C^T(x-y) \leq F \), then \( C^T\overline{x} \leq \min_{x \in K} C^T x + F \left( \frac{1}{T} \right)^{\frac{1}{n}} \).

In particular, if $T \geq n \log \frac{F}{\varepsilon}$, then we must be $\varepsilon$-optimal.

GRUNBAUM'S LEMMA. For any convex, compact $K$, with $\text{com} \, \nu_0$, \( \forall C \in \mathbb{R}^n \), \( \frac{\text{vol}(K \cap \{x : C^T(x-x_0) \leq 0\})}{\text{vol}(K)} \geq \frac{1}{e} \).

In words, any half-space through the center of mass of a convex body rejects at least $1/e$ fraction of the volume. With this interpretation, the COM algorithm has the same flavor as binary search.

PROOF OF CLAIM. Let $x^* = \arg\min_{x \in K} C^Tx$. Then, take $x^*_k = \frac{1}{n} (1-\varepsilon) x^* + \varepsilon k$.

Now, $\text{vol}(x^*_k) = \left( 1 - \varepsilon \right)^n \text{vol}(K)$.

Also, $\max_{x \in x^*_k} C^T x \geq \min_{x \in K} C^T x + \varepsilon F$, by construction of $x^*_k$. 

\( * \)
In words, $x_*^t$ is a small set of points, all with greed objective value. We'll prove that although initially completely inside $K_t$, some of it must fall outside $K_t$ for large enough $t$. Whenever this first happens, $n_t$ must be better than some $n_t x_*^t$. To see this:

$$\nu e l(K_{t+1}) \leq (1 - \frac{1}{e^{t}}) \nu e l(K_t)$$

```
ARUNBAHM
```

$$\leq \left(1 - \frac{1}{e^{t}}\right) \nu e l(K_1).$$

By repetition.

Now, set $\varepsilon > (1 - \frac{1}{e^{t}})^{T/N}$. Then, $x_*^t \in K_t$, yet $\nu e l(K_t) < \nu e l(x_*^t)$.

Hence, $\exists t \in [T], x_*^t \in X_*, \, \forall x_\in K_t, x_\in K_{t+1}$. By construction, $c^T x_\in K_t < c^T x_*^t \leq \min_{x \in K} c^T x + \varepsilon F$.

$\square$

In the rest of this note, we will prove Gunthaus's lemma.

**Observation:** Say $(n-1)$-dimensional volume of a sphere is $V_{n-1}^{n-1}$. (radius $r$)

$$\nu e l(K) = \int_0^r \nu e l(K) = \int_0^r C_{n-1} r^{n-1} \, dr = \frac{C_{n-1} R^n}{n}.$$  

$$C_{n-1} r^{n-1} \, dr = \frac{n}{n+1} R.$$  

$$C_{n-1} r^{n-1} \, dr = \frac{n}{n+1} R.$$  

$$\nu e l(R) = \left(\frac{n}{n+1}\right)^n \geq \frac{1}{e}. \quad \text{In some sense, cone is the lowest case for Gunthaus.}$$

Although, this is an example, we will use it as a proof strategy. We will reduce general common branches to (right) cones.
For non-empty compact sets $A, B \subseteq \mathbb{R}^n$,\n\[ \text{vol}(A+B)^{\frac{1}{n}} \geq \text{vol}(A)^{\frac{1}{n}} + \text{vol}(B)^{\frac{1}{n}}. \]

**Proof.** In the 1D, you have proven this for the case when $A$ and $B$ are axis-aligned (Chebyshev) rectangles. We will extend this to when $A$ and $B$ are unions of disjoint cubes by using an inductive argument. Our induction hypothesis is that the stated inequality is true when $A$ and $B$ contain no disjoint cubes in total. Volume is translation invariant. Hence, shift the $x_1 = 0$ plane so that at least one cube lies entirely above in $A$.

Translate $B$ along $x_1$ so that \[ \frac{\text{vol}(A^+)}{\text{vol}(A)} = \frac{\text{vol}(B^+)}{\text{vol}(B)}; \]
such translation always exists due to the Inclusion-Exclusion Principle. Notice that \((A^+ + B^+) \cap (A^- + B^-) = \emptyset\) since $x_1 = 0$ separates them, yet \((A^+ + B^+) \cup (A^- + B^-) \subseteq A + B$). Hence, we have\n\[ \text{vol}(A + B) \geq \text{vol}(A^+ + B^+) + \text{vol}(A^- + B^-) \]
\[ \geq (\text{vol}(A^+)^{\frac{1}{n}} + \text{vol}(B^+)^{\frac{1}{n}})^n + (\text{vol}(A^-)^{\frac{1}{n}} + \text{vol}(B^-)^{\frac{1}{n}})^n \]
\[ = (\text{vol}(A^+)^{\frac{1}{n}} + \text{vol}(B^+)^{\frac{1}{n}})^n. \]

**Corollary:** \(\text{vol}(K \cap x_1 = \alpha 3)^{\frac{n}{n-1}}\) is concave in $x_1$ for any compact, convex set $K \subseteq \mathbb{R}^n$. This is an $(n-1)$-dimensional volume.
PROOF: Let \( K_\alpha = K \setminus \{ x_1 = \alpha^3 \} \). Note that \( \lambda K_\alpha + (1 - \lambda) K_\beta \subseteq K_{\alpha + (1 - \lambda) \beta} \)

for all \( \alpha, \beta \in \mathbb{R}, \lambda \in [0, 1] \) since \( K \) is convex. Thus,

\[
\text{vol}(K_{\alpha + (1 - \lambda) \beta})^{\frac{1}{n-1}} \geq \lambda \left( \text{vol}(K_\alpha) \right)^{\frac{1}{n-1}} + (1 - \lambda) \left( \text{vol}(K_\beta) \right)^{\frac{1}{n-1}}.
\]

Now, we are ready to complete Grunbaum's lemma.

**Proof of Grunbaum's Lemma.**

Without loss of generality, we can orient our cone so that \( x_1 = 0 \) is the cutting hyperplane. Replace every slice of \( K \) along \( x_1 = 0 \) with a \((n-1)\)-dimensional sphere of equal \((n-1)\)-dimensional volume. This step preserves volumes of both sections on either side of \( x_1 = 0 \); also \( x_1 \) coordinate of center of mass stays the same. So, it suffices to establish the claim for this new body.

\( K^+ = K \cap \{ x_1 \geq 0 \} \). First note, this new body is convex,

\( K^- = K \cap \{ x_1 \leq 0 \} \), since we didn't modify \( \text{vol}(K \cap \{ x_1 = \alpha^3 \}) \)

and \( \text{vol}(K \cap \{ x_1 = 0 \})^{\frac{1}{n-1}} \) was concave in \( \lambda \) for the old (and hence, the new) body.

Replace \( K^+ \) with a cone with the same spherical mass as \( K^+ \), so that the cone and \( K^+ \) are equivalent volume.

Extend this cone in the negative \( x_1 \)-region still this extension has volume equal to \( \text{vol}(K^-) \), again always possible by intermediate value theorem.

These operations are volume preserving, but what happens to the center of mass? The claim is that it can only move rightwards. In other words, this transformation increases the \( x_1 \) coordinate of center of mass from 0 to something non-negative. This is once again a consequence of concavity of \( \text{vol}(K \cap \{ x_1 = \alpha^3 \})^{\frac{1}{n-1}} \) in \( \alpha \). Post this transformation, we have a perfect
cone with non-negative $n_3$ coordinate of $\text{COM}$. Hence,
\[
\frac{\text{vol}(K_+)}{\text{vol}(K)} = \frac{\text{vol}(K \cap \{ x_3 > 0 \})}{\text{vol}(K)} \geq \frac{\text{vol}(K \cap \{ x_3 \geq n_{3,\text{COM}} \})}{\text{vol}(K)} \geq \frac{1}{e},
\]

where $n_{3,\text{COM}}$ is $n_3$ coordinate of $\text{COM}$.

\[ \square \]

Mass moves rightward. Hence, this transformation shifts COM rightward.

(Informal proof by constructing a transport map.)
LECTURE 8: ELLIPSOID

The ellipsoid method can be thought of as a variant of the center-of-mass method, but one implementable in polynomial time. It was the first practically poly-time algorithm for IPs.

**ASSUMPTION:** $K$ is convex, compact and $rB_2 \subseteq K \subseteq rB_2$

where $B_2 = \{ x : \| x \|_2 \leq 1 \}$.

**ALGORITHM**

1. Initialize $E_1 = rB_2$.
2. For $t = 1, \ldots, T$
   1. Let $x_t$ lie the center of $E_t$.
   2. Is $x_t \in K$? (MEMBERSHIP ORACLE)
   3. If yes, construct an ellipse $E_{t+1}$ containing $E_t \cap \{ x : C^T(x-x_t) \geq 0 \}$.
   4. If no, ask for a half space $w^T x \geq D$. Hence, all of $K$ is contained in $w^T(x-x_t) \geq 0$. Construct an ellipse $E_{t+1}$ containing $E_t \cap \{ x : w^T(x-x_t) \geq 0 \}$.

**IMPLEMENTATION**

If $K = \{ x : A x \leq b \}$, then $x_t \in K$ can be answered in linear time by checking all constraints one-by-one $a_i^T x \leq b_i$.

If $x_t \notin K$, then $i \in \text{argmin} [a_i^T x_t]$. But $x \notin K$, $a_i^T x \leq b_i$, implying $a_i^T(x-x_t) \leq 0$ $\forall x \in K$. This gives us the required separating hyperplane.

**IMPORTANT NOTE:** We've demonstrated that for IPs with polynomially many constraints, membership & separation oracles can be implemented efficiently. However, this is not the only case when this is possible. For...
certain structured LPs with exponentially/infinitely many constraints, ellipsoid is still a poly-time algorithm as long as SEPARATION/MEMBERSHIP QUERIES are efficiently answerable.

**Analysis**

**Volume Reduction Lemma.** For any ellipsoid $E_0$ with center $x_0 \in \mathbb{R}^n$, and vector $w_0$, we can efficiently construct an ellipsoid $E_1$ containing $E_0 \cap \{x : w_0^T(x-x_0) \geq 0\}$ with $\text{vol}(E_1) \leq \text{vol}(E_0) e^{-\frac{T}{2(n+1)}}$.

**Claim.** Let $x$ be a feasible point in $\mathbb{R}^n$ achieving the minimum objective value. Then

$$C^T x = \min_{x \in K} C^T x \leq \frac{F}{\xi} e^{-T/2(n+1)n},$$

where

$$F = \max_{x, y \in K} C^T (x-y).$$

Hence, as long as $T \geq 2n(n+1) \log \frac{FR}{\xi \varepsilon}$, we must be $\varepsilon$-optimal. Notice that this is slower than complexity a factor of $n$, because volume reduction is $1 - \frac{1}{2(n+1)}$ in each step, instead of a constant. This is the price ellipsoid method pays for efficient implementability.

**Proof of Claim.** We will follow the same recipe as that for con.

Let $x^* \in \arg\min_{x \in K} C^T x$, and $x_{\varepsilon}^* = (1-\varepsilon)x^* + \varepsilon K \subseteq K$.

Now, we have $\text{vol}(x_{\varepsilon}^*) \geq \text{vol}(E_1 \circ B_2) = C_n(\varepsilon e)^n$ for some $C_n$ such that $\text{vol}(B_2) = C_n$. Also, we have

$$\max_{x \in \mathbb{R}^n} C^T x = C^T x^* + \max_{x \in \mathbb{R}^n} C^T(x-x^*) \leq C^T x^* + \varepsilon F.$$

By volume reduction lemma, with expected applications, we get $\text{vol}(E_{T+1}) \leq \text{vol}(E_1) e^{-\frac{T}{2(n+1)}} = C_n R^n e^{-\frac{T}{2(n+1)}}$. 


Choose \( \varepsilon > \frac{R}{R_b} e^{-\frac{1}{2cR_bT}} \). Then since \( \text{vol}(E_{t+1}) < \text{vol}(E_{t}) \),

\( \exists t, \pi^*_t \in X^*_\varepsilon \) such that \( \pi^*_t \in E_t \), \( \pi^*_t \not\in E_{t+1} \). Further,

\( \exists t \) that this can only happen on the YES branch;

because all existing feasible points are retained on the NO branch. Hence, \( c^T \pi_t < c^T \pi^*_t \leq c^T \pi^*_t + \varepsilon F \).

Now we will finish up the proof of the volume reduction lemma. Note that this lemma was crucial for the (sub)optimality result. Also, such niceties don’t work for some simpler shapes.

**Proof of Volume Reduction Lemma.**

Consider a simpler case when we start with the unit ball \( B_2 = \frac{1}{2} \mathbb{R}^2 : \|x\|_2 \leq 1 \), \( \pi_1 \geq 0 \) is halfspace.

Any ellipse can be written as

\[ \|x - x_0\|_{H_0^{-1}}^2 = (x - x_0)^T H_0^{-1} (x - x_0) \leq 1 \]

where \( x_0 \) is the center, the eigenvalues of \( H_0 \) are its principal axes with lengths \( \sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n} \) where \( \lambda_i \)'s are the eigenvalues of \( H_0 \).

Think of it as a generalization of

\[ \frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} \leq 1 \]

Clearly, an ellipse with principal axes of lengths \( \sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n} \) has volume = \( C_n \sqrt[2n]{\lambda_1 \cdots \lambda_n} = C_n \sqrt{n \text{det}(H)} \). Stretching a body along a single axis by \( 2x \), increases volume by \( 2x \).
Each volume element decreases in volume while stretching an axis by 2x.

Now, back to $E_1$. By symmetry, we take $\eta_1 = t e_1$, as the center of $E_1$. This ellipse touches $e_1$ and $b_2 = \ell_2 = 0$. So, our ansatz is $H = a e_1 e_1^T + \frac{1}{b} (I - e_1 e_1^T)$. This is an eigen-value decomposition. $H^{-1} = \frac{1}{a} e_1 e_1^T + \frac{1}{b} (I - e_1 e_1^T)$. So,

$$
\frac{(1-t)^2}{a} = 1 \Rightarrow a = (1-t)^2 \quad \frac{\frac{1}{a} + \frac{1}{b} = 1}{1} \Rightarrow b = \frac{1}{1 - t^2/a} = \frac{(1-t)^2}{1-2t}.
$$

volume $(E_1) = C_n \sqrt{a b^{-1}} = C_n \frac{(1-t)^n}{(1-2t)^{n/2}}$.

Maximizing this for $t$, 

$$
\frac{(1-t)^n}{(1-2t)^{n/2}} \Rightarrow n = \frac{(1-t)^n}{(1-2t)^{n/2}} \Rightarrow t = \frac{1}{n+1}.
$$

$$
a = (1-t)^2 = \left(\frac{n}{n+1}\right)^2, \quad b = \frac{(\frac{n}{n+1})^2}{\frac{n-1}{n+1}} = \frac{n^2}{n^2-1}.
$$

volume $(E_1) = C_n \sqrt{a b^{-1}} = C_n \left(\frac{n}{n+1}\right) \left(1 + \frac{1}{n^2-1}\right)^{n-1/2} \leq C_n e^{2/(n+1)} e^{-1/2(n+1)} \text{ volume } (E_0 = B_2).

using $1+x \leq e^x, \forall x \in \mathbb{R}$.

Note that $\frac{\text{volume } (E_1)}{\text{volume } (E_0)}$ is invariant under rotations, (since all volumes are) but also under stretching of any axes as we have seen.

Thus $\frac{\text{volume } (E_1)}{\text{volume } (E_0)}$ is invariant under any invertible coordinate transformation, since any invertible linear map $A = UV^T$ for orthogonal $U, V$
and diagonal \( \Sigma \) with positive entries by singular value decomposition. Thus our anecdote reduction result holds starting with any ellipse and half-space.

Verify that \( \mathcal{E}_1 \equiv \mathcal{E}_0 \cap \frac{2}{3} x_1 \geq 0 \).

Finally, although unnecessary for our proof, note we have also constructed the smallest ellipse subject to the containment requirement; our upper bounds on its volume might have been a bit loose though.

For computationally explicit implementation, we extend this construction to the general case, i.e.,

\[
\mathcal{E}_0 = \{ x : \| x - x_0 \|_{H_0^{-1}}^2 \leq 1 \} \quad \text{will construct} \quad \mathcal{E}_1 \equiv \mathcal{E}_0 \cap \frac{2}{3} x_1 \geq 0.
\]

In the space, \( \mathcal{E}_0 = \{ y : \| y \|_2^2 \leq 1 \} \),

\[
\begin{align*}
\mathcal{E}_1 &= \left\{ y : \left\| y - \frac{1}{n+1} \frac{H_0 w}{\| w \|_{H_0}} \right\|_{H_0}^2 \leq \left( \frac{n+1}{n} \right)^2 \frac{H_0^{1/2} w w^T H_0^{1/2}}{\| w \|_{H_0}^2} + \frac{n^2-1}{n^2} \left( I - \frac{H_0^{1/2} w w^T H_0^{1/2}}{\| w \|_{H_0}^2} \right) \right\} \\
&= \{ x : \| x - \mathcal{x}_1 \|_{H_1^{-1}}^2 \leq 1 \}
\end{align*}
\]

where

\[
\mathcal{x}_1 = \mathcal{x}_0 + \frac{1}{n+1} \frac{H_0 w}{\| w \|_{H_0}} \quad H_1 = \left( \frac{n+1}{n} \right)^2 \frac{w w^T}{\| w \|_{H_0}^2} + \frac{n^2-1}{n^2} \left( I - \frac{H_0^{-1/2} w w^T H_0^{-1/2}}{\| w \|_{H_0}^2} \right)^{-1}.
\]

References:

1. Ellipsoid algorithm
   - Section 2.2 in Bubeck
2. Sections 3.2 and 3.3 in Lee-Vempala
3. (Beyond this course) Applying ellipsoid to large LPs— Chapter 3+ in GLS
LECTURE 9: REGRET

STORY SO FAR...

<table>
<thead>
<tr>
<th>ALGEBRAIC</th>
<th>GEOMETRIC</th>
<th>LOGICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log \frac{n}{\varepsilon} )</td>
<td>Ellipsoid</td>
<td>Fourier-Motzkin</td>
</tr>
<tr>
<td>? Interior-points</td>
<td>Oracle-based COM</td>
<td></td>
</tr>
<tr>
<td>? Gradient Descent</td>
<td>Sampling-based COM</td>
<td>Multiplication Weights</td>
</tr>
</tbody>
</table>

We will fill this quadrants now.

EXPERTS SETTING

\[ t = 1, \ldots, T \]
\( 'N' \) experts make recommendations \( \frac{\varepsilon}{\varepsilon+1} \) to the learner.
Learner chooses an expert to follow, say \( i \in [N] \).
Adversary choose losses \( \frac{\varepsilon}{\varepsilon+1} \) for each recommendation,
(More loss is hard.)

Repeat

Assumption FOR NOW: If an expert were is perfect, on all days
incurs 0 loss; learner doesn't know which one.

Q. What strategy should the learner follow to minimize
her cumulative number of mistakes?

NAIVE STRATEGY: Follow friend \( i \) till they make a mistake.
If/when they do, start following friend \( i+1 \).

Upon each of the learner's mistakes, one expert is eliminated.
Hence, \# mistakes for learner (or cumulative loss) \leq N-1
in the worst-case. But we can do much better.

SURVIVING MAJORITY: Each day take a majority vote among
all surviving experts. At the conclusion of the day,
eliminate those who made mistakes.

Every time the learner makes a mistake, \( \frac{1}{2} \) of the
expert pool is eliminated. Can only happen so many times.
Hence, \( \# \text{ mistakes} \leq \log_2 N \). This is an exponential improvement! Fantastic! But we would like to get rid of the realizability/perfect expert assumption. Natural generalization is to initially assign each expert some credibility that goes down, but doesn’t become zero like before, when the experts make mistakes. This works to an extent, but comes up short against the following barrier.

I try to produce an upper bound on number of mistakes by halving the credibility of a wrong expert in each round & taking weighted majority.

**CLAIM:** For any strategy* for the learner, there exists a worst-case assignment of losses guaranteeing the learner makes AT LEAST twice the number of mistakes for the best expert.

**PROOF:** Take 2 experts - A predicts +1 everyday, B predicts -1. The adversary assigns +1 loss to whichever expert you as the learner, pick, and 0 to the other. Hereby, on each day you make a mistake, i.e. after \( T \) days, \( T \) cumulative mistakes. However, on each day exactly one expert makes a mistake. Hence, because minimum average, an expert that at the end of \( T \) days has made at most \( T/2 \) mistakes. \( \square \)

Let \( m^* \) = minimum number of mistakes for any expert. Thus, \( 2m^* \) seems like a natural barrier. However, the above learner bound construction crucially depends on the learner’s strategy being deterministic. For a randomized strategy (where the adversary cannot inspect the learner’s coin-flips just everything else), one can plausibly...
MULTIPLICATIVE WEIGHTS / HEDGE ALGORITHM

Set \( w_i^0 = 1 \) \( \forall i \in [N] \)

For \( t = 1, \ldots, T \)

Play \( i_t \sim P_t \) where \( P_t^i = \frac{w_t^i}{\sum_{i \in [N]} w_t^i} \).

Adversary chooses loss vector \( l_t \in [-1, 1]^N \),

that can depend on past losses, weights, past &
current, all actions \( i_1 \cdots i_{t-1} \), but not \( i_t \).

(Equivalently, it can depend on \( i_t \) as long as the

learner’s payoff \( \Delta \equiv \mathbb{E}_{i \sim P_t} \left[ l_t^i \cdot P_t^i \right] \).

Update \( w_t^i 

= w_t^i \cdot e^{-\eta l_t^i} \).

THEOREM:

\[
\mathbb{E} \left[ \sum_{t=1}^T l_t^i \right] - \min_{i \in [N]} \sum_{t=1}^T l_t^i 
\leq \sum_{t=1}^T P_t^i l_t - \min_{P \in \text{Lin}[N]} \sum_{t=1}^T P^T l_t 
\leq \sqrt{T \log N}
\]

LEARNER’S.cumulative loss

Best expert’s

log in hindsight

Also called REURRENT

Let us explore the implication before diving into a proof.

Diminishing log \( T \), we get

LEARNER’S.AVERAGE LOSS \leq AVERAGE LOSS OF THE BEST EXPERT IN HINDSIGHT + \sqrt{\frac{\log N}{T}} .

EXCESS AVERAGE LOSS

Comments:

1. This guarantee holds for arbitrarily, or even adversarially,

chosen loss assignment vectors. No distributional

assumptions were made unlike stats/ML/stochastics.

2. There’s no 2x multiplier associated with the best

expert. This breaks our lower bound.

3. Excess average loss \( \rightarrow 0 \) as \( T \rightarrow \infty \).

In particular, if \( T \geq \log N / \varepsilon^2 \), excess average loss \( \leq \varepsilon \).
4. It's a relative error guarantee; no one can ensure low absolute error even for random loss functions. A relative (additive) error metric is something experts in other fields outside ML/statistical learning find hard to swallow (although situation is very rapidly changing), but it has proven to be one of the most far-reaching design choices in ML theory.

5. The cost for housing many inaccurate experts, as long as there is one good one, is small because of the log\(N\) dependence. Empirical \(N\) still yields reasonable bounds.

6. The nature of the bound in \#3 is not an accident. It closely resembles uniform convergence results from statistical learning, precisely because online learning is an algorithmic theory as opposed to an analytic theory that generalizes the former.

**Proof of Theorem.** Our basic proof strategy is to construct a potential function that decreases when a learner makes mistakes. Taking inspiration from MAJORITY/HALVING, we take \(\Phi_1 = \sum_{i \in [N]} w_i^1 = N\). Now,

\[
\Phi_{t+1} = \sum_{i \in [N]} w_i^{t+1} = \sum_{i \in [N]} w_i^t e^{-\eta l_i^t} = \Phi_t \sum_{i \in [N]} \frac{w_i^t}{\sum_{i \in [N]} w_i^t} e^{-\eta l_i^t}
\]

\[
\leq \Phi_t \sum_{i \in [N]} p_i^t (1 - \eta l_i^t + \eta^2 (l_i^t)^2)
\]

\[
\leq \Phi_t \left(1 - \eta P_t^t l_t + \eta^2 \right) \leq \Phi_t e^{-\eta P_t^t l_t + \eta^2}
\]

using \(e^x \leq 1 + x + x^2 \) \(\forall x \leq 1\), and \(1 + x \leq e^x \forall x \in \mathbb{R}\) in successive steps of the algorithm.
We are almost done. Let $i^* = \arg\min_{i \in [n]} \frac{1}{T} \sum_{t=1}^{T} L_t^i$ be the best expert in hindsight. Then using the above:

$$\hat{C}^* \leq \Phi_{T+1} \leq \Phi_1 C$$

Taking log on both sides, we get:

$$-\eta \sum_{t=1}^{T} L_t^{i^*} \leq \log N - \eta \sum_{t=1}^{T} P_t^i L_t + \eta^2 T.$$  

$$\Rightarrow \sum_{t=1}^{T} P_t^i L_t - \sum_{t=1}^{T} L_t^{i^*} \leq \frac{\log N}{\eta} + \eta T \leq 2\sqrt{T \log N}. \quad \Box$$

**APPLICATIONS OF MULTI WEIGHTS**

**EXAMPLE ONE: SOLVING LPs**

We'll solve the feasibility problem: \( \exists x \in K : Ax \leq b \), where \( K \) is a "simple" convex set. Why restrict to simple sets? Because we'll use a subroutine/oracle that answers \( \exists x \in K : Cx \leq d \). Note that this only has one inequality constraint, instead of \( m \).

**Example 1:** \( K = B_2^2 \leq 1 \). \( \exists x \in K, Cx \leq d \) is YES iff (a) \( d \geq 0 \) (or \( 0^T c \leq d \)), OR (b) \( \text{dist}(0,K) = \frac{\|d\|_2}{\|C\|_2} \leq 1 \).

**Example 2:** \( K = \{ x \geq 0 \} \). \( \exists x \in K, Cx \leq d \) is YES iff (a) \( d \geq 0 \) (or \( 0^T c \leq d \)), OR (b) \( \exists i, c_i < 0 \).

**Prove the correctness of the procedures above for \( K = B_2^2 \) and \( K = \{ x \geq 0 \} \). Also, describe a procedure to efficiently solve \( \exists x \in \{ \|x\|_\infty \leq 1 \} \) \( C^T x \leq d \).

We describe the algorithm next. Think of it as a game where the learner tries to prove the LP is infeasible assisted by the constraints as experts, by asking "gotcha" questions. The \( \text{ORACLE} \) assigns the learner's concern.
ALGORITHM.

1. Each constraint is an expert, with $w_i^0 = 1 \forall i \in [m].$
2. For $t = 1, \ldots, T$
   
   Learner chooses $p_t^i = \frac{w_t^i}{\sum_{i \in [m]} w_t^i}$
   
   Asks the oracle $\exists x \in K$, $p_t^i A x \leq p_t^b b.$
   
   If NO, output that the original LP is infeasible.
   
   If YES, ORACLE returns $x_t \in K$ such that $p_t^i A x_t \leq p_t^b b.$

   Each constraint $i$ receives loss $\frac{1}{T} (b_i - a_i^T x_t)$.
   
   $w_i^{t+1} = w_i^t e^{-\eta (b_i - a_i^T x_t)/p} \in [-1,1]$ update.

Here, $p = \max_{i \in [m], x \in K} |b_i - a_i^T x| = WIDTH$ of $Ax \leq b$ against $K.$

CLAIM. For any $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m,$ this algorithm either outputs:

(A) that $\exists x \in K, Ax \leq b$ is infeasible, correctly.

(B) a point $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t \in K$ such that $Ax \leq b + \rho \sqrt{\frac{\log m}{T}} 1.$

In words, either the algorithm correctly declares that the LP is infeasible or outputs an $\varepsilon$-feasible solution when run for long enough, i.e., when $T = \frac{e^2 \log m}{\varepsilon^2}.$

PROOF. Clearly, if $Ax \leq b, x \in K$ is feasible, then $\forall p \geq 0,$ we have that $p^T A x \leq p^T b, x \in K$ is feasible. Thus, we only need the prime part B assuming ORACLE says YES on all rounds. If so, by the regret guarantee:

$$\sum_{t=1}^T p_t^i (b-Ax_t) \leq \min_{i \in [m]} \sum_{t=1}^T (b_i - a_i^T x_t) + \rho \sqrt{T \log m}.$$ 

But $\forall t, p_t^i b \geq p_t^i A x_t.$ Hence

$$0 \leq \frac{1}{T} \sum_{t=1}^T p_t (b-Ax_t) \leq \min_{i \in [m]} \left( b_i - a_i^T \left( \frac{\sum_{t=1}^T x_t}{T} \right) \right) + \rho \sqrt{\frac{\log m}{T}}.$$ 

Rearranging, $\forall i \in [m], a_i^T \bar{x} \leq b_i + \rho \sqrt{\frac{\log m}{T}}. \quad \square$
EXAMPLE 2: CONSTRUCTIVE MINIMAX THEOREM

Recall that \( \min_{x \in \Delta} \max_{y \in \Delta} x^T A y = \max_{y \in \Delta} \min_{x \in \Delta} x^T A y. \)

We achieved it using strong duality, in fact it is equivalent to strong duality. We will give a constructive algorithmically efficient proof of this statement. In fact, the previous algorithm can be seen as an efficient algorithmic "Farkas' Lemma."

Assume \( \max_{i,j} |a_{ij}| \leq 1. \) Else, we scale.

ALGORITHM

1. Row player thinks of each row as an expert.
2. \( t = 1, \ldots, T \)
   - Row player plays \( x_t \in \Delta \) as per Multi Weight
   - Column player plays \( y_t = \arg\max_{y \in \Delta} x_t^T Ay. \)
     - Row i's loss is \( e_i^T A y_t. \)

CLAIM. \( \frac{T}{\log m} \geq \frac{1}{\varepsilon^2}, \) then \( z = \frac{1}{T} \sum_{t=1}^T x_t \) satisfies

\[
\max_{y \in \Delta} x^T A y \leq \max_{y \in \Delta} \min_{x \in \Delta} x^T A y + \varepsilon.
\]

Note that this is the non-trivial direction. By weak duality or definition of \( \min/\max, \) \( \min_{x \in \Delta} \max_{y \in \Delta} x^T A y \geq \max_{y \in \Delta} \min_{x \in \Delta} x^T A y. \)

By compactness + continuity, we get \( \exists x', \max_{y \in \Delta} x'^T A y \leq \max_{y \in \Delta} \min_{x \in \Delta} x^T A y. \)

PROOF. \( \max_{y \in \Delta} z^T A y = \max_{y \in \Delta} \frac{1}{T} \sum_{t=1}^T x_t^T A y \leq \frac{1}{T} \sum_{t=1}^T \max_{y \in \Delta} x_t^T A y \)

\[
= \frac{1}{T} \sum_{t=1}^T x_t^T A y_t \leq \min_{y \in \Delta} \left( \frac{1}{T} \sum_{t=1}^T x^T A y_t + \sqrt{\frac{\log m}{T}} \right)
\]

using regret guarantee

\[
= \min_{x \in \Delta} x^T A \left( \frac{1}{T} \sum_{t=1}^T y_t \right) + \sqrt{\frac{\log m}{T}}
\]
So, low regret $\Rightarrow$ minimax theorem. But can we go back? David Blackwell in 1956 proved a close equivalence between existence of low-regret strategy & a certain generalization of the minimax theorem.

References:

1. The best reference for regret minimization & applications to LPs/minimax duality is Elad’s book—specifically chapters 1 & 8.
2. See this survey from Sanjeev, Elad and Satyen for applications of the multiplicative weights algorithm.
3. See this fantastic paper by Yoav Freund and Robert Schapire, who pioneered the Godel prize-winning boosting approach to machine learning using the regret-minimax link.
4. This NYTimes article quoting Rakesh Vohra chronicling the (independent) rediscovery of multiplicative weights in many academic fields; I think of this as convergent evolution. In 1957, for example, a statistician named James Hanna called his theorem Bayesian Regret. He had been preceded by David Blackwell, also a statistician, who called his theorem Controlled Random Walks. Other, later papers had titles like “On Pseudo Games,” “How to Play an Unknown Game,” “Universal Coding” and “Universal Portfolios,” Dr. Vohra said, adding, “It’s not obvious how you do a literature search for this result.”
Q1. PART A - 2 points.
Prove that \( f : \mathbb{R} \rightarrow \mathbb{R} \), \( f(x) = \frac{c^T x + p}{d^T x + q} \) is quasi-convex over \( \mathcal{N} = \{ x : d^T x + q \geq 0 \} \).

PART B - 8 points
Consider the following optimization problem.
\[
\begin{align*}
\max_{x \in \mathcal{N}} & \quad \frac{c^T x + p}{d^T x + q} \\
\text{s.t.} & \quad A x \leq b
\end{align*}
\]
Propose a linear programming reformulation of this optimization problem. Also, describe how would one reconstruct a solution to (0) given an optimal solution of your proposed LP.

Comment: Upto 5 points, if you do not have a formulation, but propose an algorithm that solves (0) by solving multiple LPs.

Q2. PART A - 2 points
How far in Euclidean distance is a point \( x' \in \mathbb{R}^n \) from the hyperplane \( H = \{ x : a^T x = b \} \)?

PART B - 8 points
Consider the set \( P = \{ x : A x \leq b \} \); assume it's compact and non-empty. Provide a linear program to compute the center & the radius of the largest sphere contained (entirely) inside \( P \).
0.3. PART A - 6 points.

In 2-dimensions, consider

\[ A = \{ x \in \mathbb{R}^2 : \max \{ x_1, |x_2| \} \leq 1 \} \]

\[ B = \{ x \in \mathbb{R}^2 : |x_1| + |x_2| \leq 1 \} \]

\[ C(\varepsilon) = \{ x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq \varepsilon^2 \} \]

Sketch \( A + B \) and \( A + C(1) \); `+` is the Minkowski Sum.

Compute \( \lim_{\varepsilon \to 0^+} \frac{\text{Area}(A + C(\varepsilon)) - \text{Area}(A)}{\varepsilon} \).

PART B - 4 points.

Now, consider 2 amio-aligned hyper-cubes (i.e., 2 amio-aligned \( n \) dimensional rectangles) \( A \) and \( B \), possibly of unequal sizes.

Prove \( \text{vol}(A + B)^{\frac{1}{n}} \geq \text{vol}(A)^{\frac{1}{n}} + \text{vol}(B)^{\frac{1}{n}} \).
Q.1. (10 points)
Minkowski-Weyl says that any bounded polyhedron can be written in two ways: either via inequalities defining it, or as convex hull of some set of points. But how do we know if ultimately we are talking about the same set expressed differently. Concretely consider:

\[ A = \text{CONVEX HULL} \left( \{x_1, \ldots, x_m\} \right) \quad B = \text{CONVEX HULL} \left( \{y_1, \ldots, y_m\} \right) \]

\[ C = \{x : A x \leq b\} \quad D = \{x : C x \leq d\}. \]

Assume you can solve any LP with \( n \) variables and \( m \) constraints in \( \text{poly}(m,n) = (m+n)^{10} \) time. Here \( A, C, B, D \in \mathbb{R}^{m \times n} \), \( x, y \in \mathbb{R}^n \). Give polynomial time (e.g. \( (m+n)^{100} \) time) algorithms to answer as many of these as possible:

1. \( \text{Is } A \subseteq B? \)
2. \( \text{Is } A \subseteq C? \)
3. \( \text{Is } C \subseteq D? \)
4. \( \text{Is } C \subseteq A? \)

Hint: Three of these are solvable in \( \text{poly}\)-time.

Q.2. (10 points)
Recall that:
1. Convex hull requires \( \lambda_i \geq 0 \) \( \forall i \).
2. Affine hull requires \( \sum \lambda_i = 1 \).
3. Convex hull requires both \( \lambda_i \geq 0 \) \( \forall i \), \( \sum \lambda_i = 1 \).

This suggests that for any set \( S \),

\[ \text{CONVEX HULL} (S) = \text{CONV HULL} (S) \cap \text{AFFINE HULL} (S). \]
PART A: Prove that this is false. For example, construct a set $S$ for which $(*)$ is false.

PART B: What minimal conditions must $\text{aff}(S)$ satisfy so that $(*)$ is true for $S$? Prove that $(*)$ indeed holds under your proposed conditions.

Q3. (10 points)

Let $2^{[n]}$ be the set of all subsets of $[n] = \{1, 2, \ldots, n\}$. Consider a function $f : 2^{[n]} \to \mathbb{R}_+$ such that

1. $f(\emptyset) = 0$.
2. $f(S) \leq f(T)$ for all $S \subseteq T \subseteq [n]$.
3. $f(S) + f(T) \geq f(S \cup T) + f(S \cap T)$ for all $S, T \subseteq [n]$.

Now, consider the following LP with exponentially many constraints, and some vector $C \in \mathbb{R}^n$.

$$\max C^T x$$

subject to

$$\sum_{j \in S} x_j \leq f(S) \quad \forall S \subseteq [n]$$

and $x \geq 0$.

Give a polynomial (in $n$) time algorithm to solve this LP; the algorithm must evaluate the function $f$ polynomial (in $n$) times.

Prove your algorithm is correct by constructing a dual feasible solution that attains the same objective value as the output of your algorithm.

Note: 3 points for listing the dual LP.
Assignment #3

Q1. (10 points)
Let \( x^* \) be a BFS with basis \( B \) for \( \min_{Ax=b, x \geq 0} c^T x \).

(A) Prove that if the reduced cost of every nonbasic variable in \( B \) is positive, then \( x^* \) is the UNIQUE minimum.

(B) If \( x^* \) is the UNIQUE minimum and the LP is non-degenerate, then the reduced cost of any nonbasic variable not in \( B \) is positive.

Q2. (10 points)
Assume \( \min_{Ax=b, x \geq 0} c^T x \) is non-degenerate.

Consider \( f(\lambda) = \min_{x} (c + \lambda d)^T x \), for \( \lambda \geq 0 \).

Say \( x^* \) with basis \( B \) is optimal at \( \lambda = 0 \).

(A) Prove the set of \( \lambda \)'s for which \( x^* \) is an optimum is \( [0, \lambda_1] \) for some \( \lambda_1 > 0 \). In fact, give an efficient algorithm as possible to compute the largest such \( \lambda_1 \).

(B) Prove that \( \exists 0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_k \leq \lambda_{k+1} = \infty \) where \( k \geq 0 \) and bases \( B_0, \ldots, B_k \) such that some \( x^* \) with basis \( B_i \) is optimal iff \( \lambda \in [\lambda_i, \lambda_{i+1}] \).

Q3. (10 points)

(A) Prove \( f(b) = \min_{Ax=b, x \geq 0} c^T x \) is concave.

(B) Prove \( g(c) = \min_{Ax=b, x \geq 0} c^T x \) is concave.

(C) Rewrite the set \( C(x^*) = \{ c : c^T x^* \geq \max_{y \in P} c^T y \} \) where \( P = \{ x : Ax \leq b \} \) as a polyhedron with polynomially many constraints.