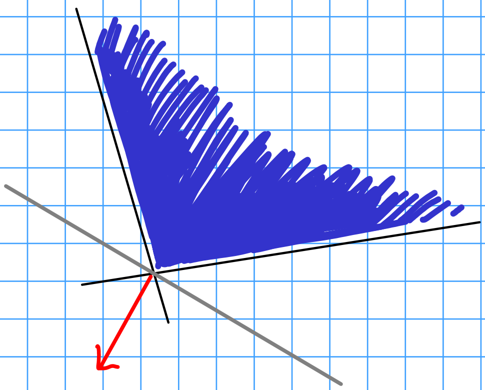
Geometry of Duality

AN ILLUSTRATED GUIDE

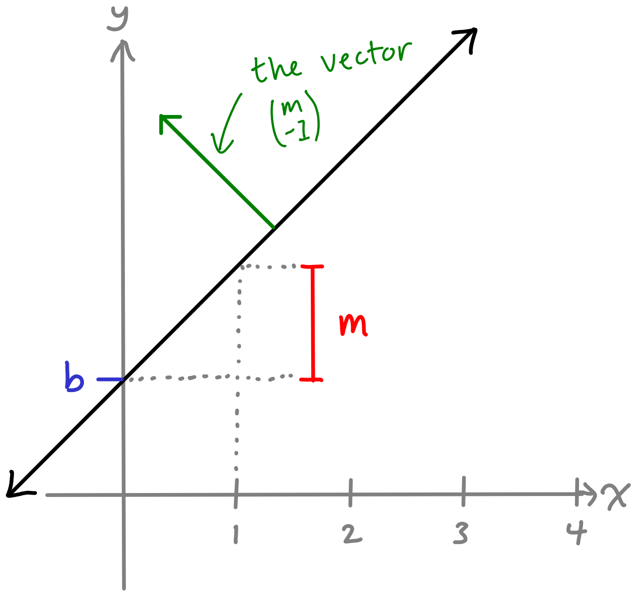


by Jenny Wong

SASMS Fall 2012 - University of Waterloo

Lines in $\mathbb{R}^2$

You've probably seen lines described in the form



slope (rise over run)

$$y = mx + b$$

y-intercept

We can rearrange the equation:

$$0 = mx - y + b$$

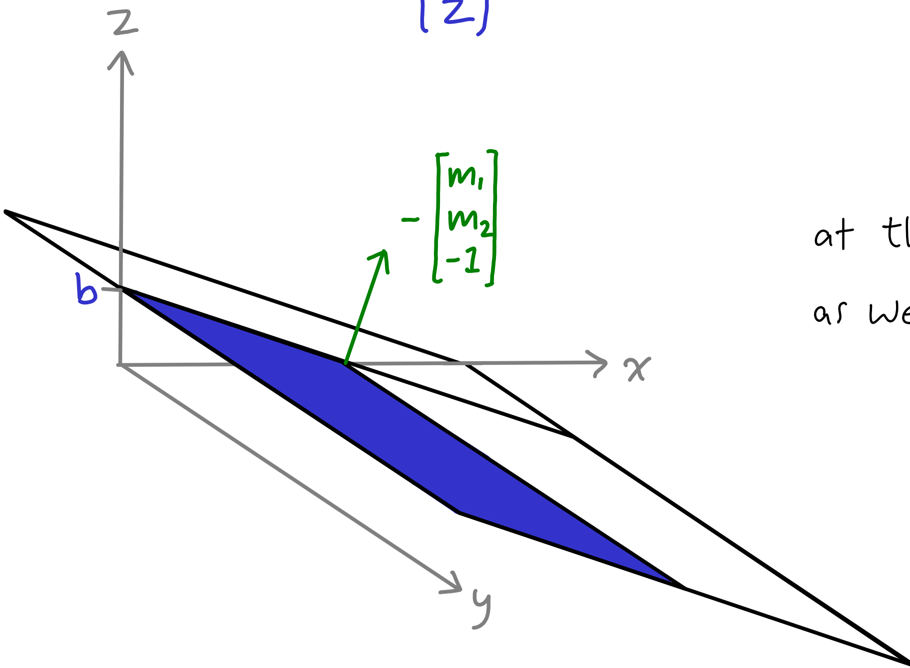
And even write it in terms of the vector $\binom{x}{y}$:

$$0 = (m \ -1)\begin{pmatrix} x \\ y \end{pmatrix} + b$$

Planes in $\mathbb{R}^3$

We take the idea of writing lines in terms of vectors and generalize.

$0 = (m_1 \;\; m_2 \;\; -1)\begin{bmatrix} x \\ y \\ z \end{bmatrix} + b$ corresponds to the plane

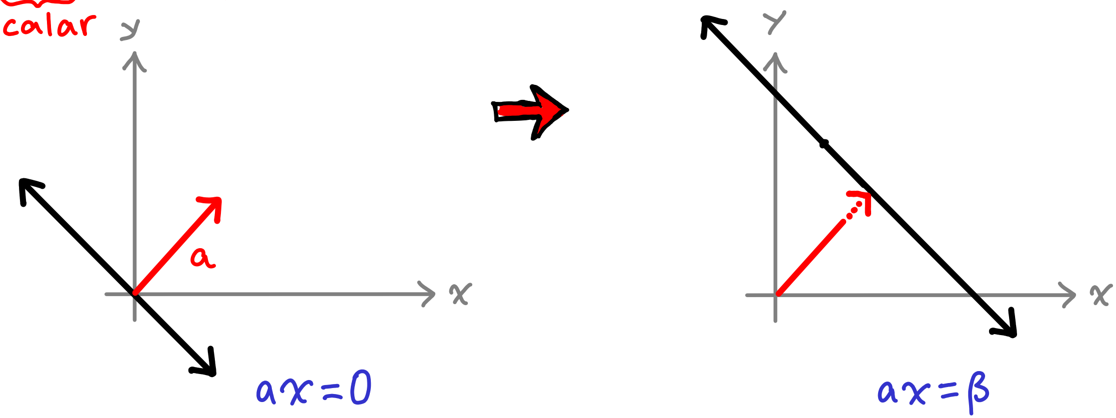


at this point, we might as well keep going past $\mathbb{R}^3$.

$(n-1)$-dimensional Hyperplanes in $\mathbb{R}^n$

Let $a \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$. We say that the set of points $x \in \mathbb{R}^n$ that satisfy $0 = ax - \beta$ (or equivalently, that $ax = \beta$) form a **hyperplane in** $\mathbb{R}^n$.

β behaves as an **offset** of sorts. It tells us that the hyperplane given by a and β is "shifted" from the origin by $\underbrace{\frac{\beta}{\|a\|_2}}_{\text{scalar}} \cdot \underbrace{a}_{\text{vector}}$.



Duality

Every hyperplane H corresponds to some pair $(a, \beta) \in \mathbb{R}^n \times \mathbb{R}$, and every pair $(a, \beta) \in \mathbb{R}^n \times \mathbb{R}$ corresponds to a hyperplane in that you can find the slope and offset of any hyperplane, and that given a slope and offset, you can construct a hyperplane with that slope and offset.

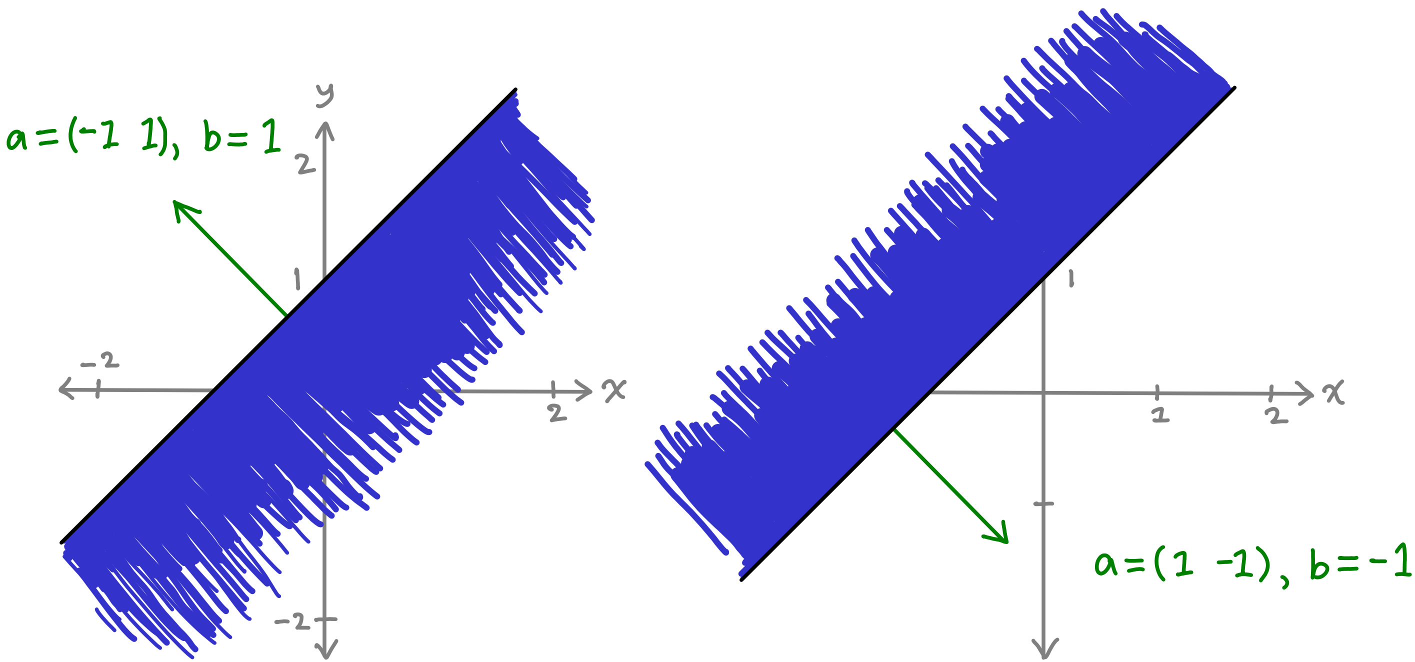
In some sense, sets of points and (slope, offset) pairs are **equally good ways to represent a hyperplane.**

$$\boxed{\{x \in \mathbb{R}^n : ax = \beta\} \Leftrightarrow (a, \beta)}$$

Half-spaces

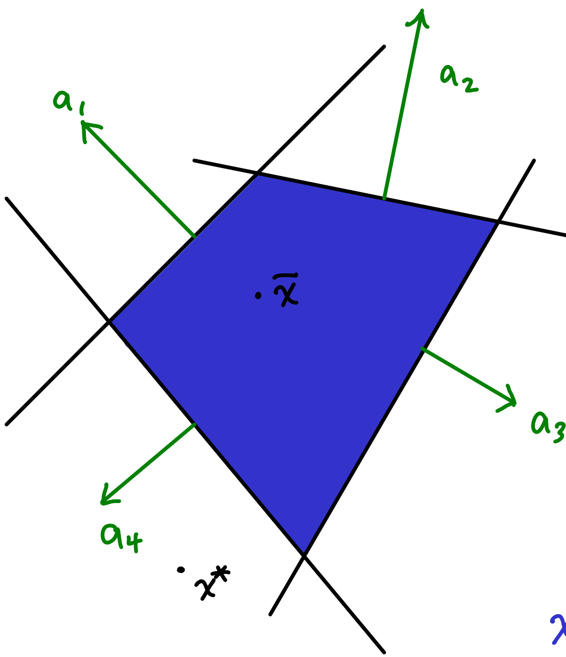
Hyperplanes in $\mathbb{R}^n$ split the vector space $\mathbb{R}^n$ in half.

Defn. X is a **closed half-space** if $X = \{x \in \mathbb{R}^n : ax \leq \beta\}$
for some $a \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$.



Polyhedra

Def'n. A **polyhedron** (plural **polyhedra**) is the intersection of a finite number of closed half-planes.



$\bar{x}$ is in this polyhedron since

$a_1\bar{x} \leq b_1$,

$a_2\bar{x} \leq b_2$,

$a_3\bar{x} \leq b_3$ and

$a_4\bar{x} \leq b_4$.

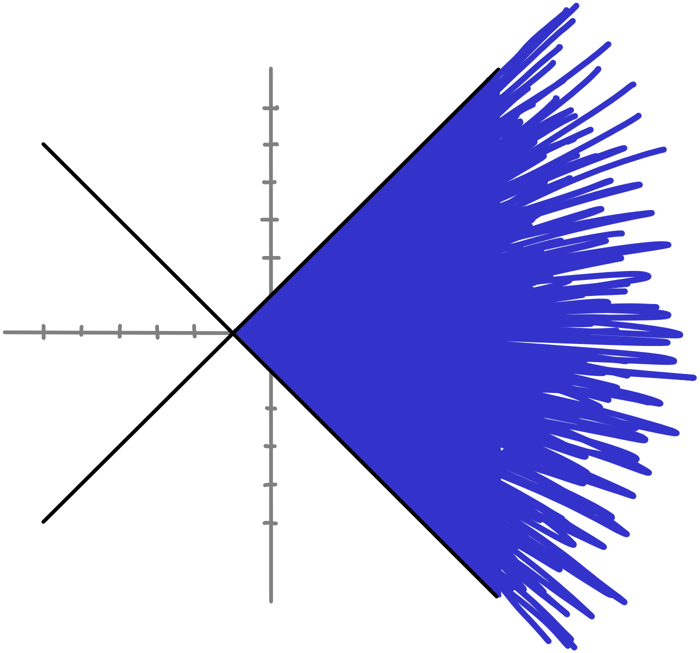
x^* is not because $a_4x^* \not\leq b_4$.

Polyhedra

Polyhedra don't need to be bounded.

$$-x + y \leq 1$$
$$-x - y \leq 1$$

Contains a half-line!



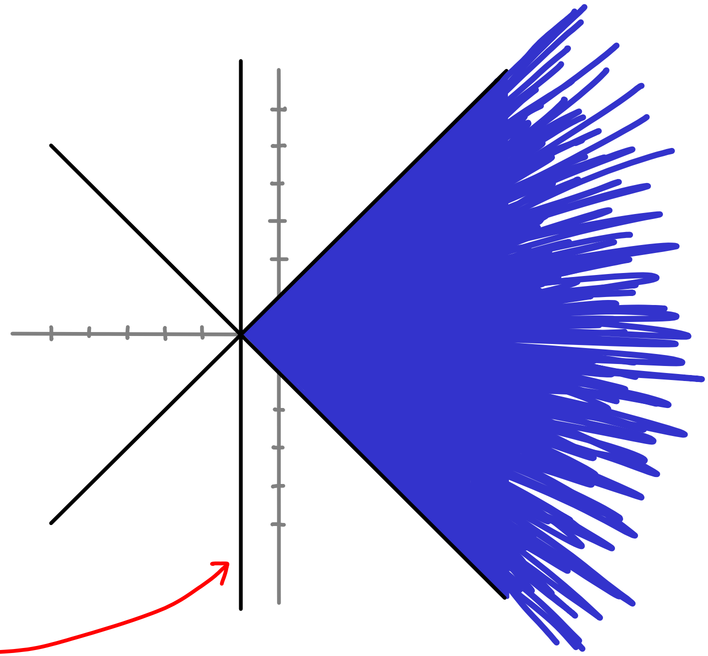
Polyhedra

Sometimes, some of the enclosing half-planes are **redundant**.

$$-x + y \leq 1$$

$$-x - y \leq 1$$

$$-x + y + (-x - y) = -2x \leq 2$$

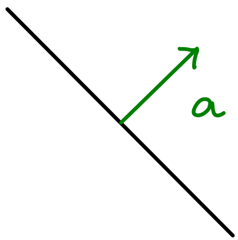


doesn't change set of points in the polyhedron

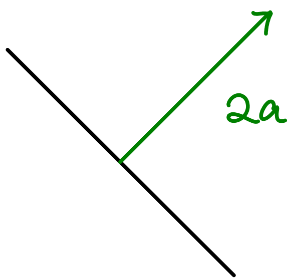
Polyhedra

It's actually really easy to find new redundant half-planes.

Let $ax \le b$ is satisfied by every x in some polyhedron P.



offset is $\dfrac{\beta}{\|a\|_2}$.



offset is $\dfrac{\beta}{\|2a\|_2} = \dfrac{1}{2} \cdot \dfrac{\beta}{\|a\|_2}$.

β gives offset in terms of the length of a.

The closed half-plane given by $(2a, 2\beta)$ is the same as the one given by (a, β).

Conical Combinations

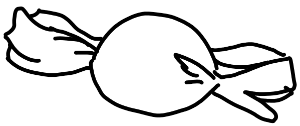
Def'n. Let $x_1, x_2, \cdots, x_k$ be vectors in $\mathbb{R}^n$. A **conical combination** is a vector of the form

$$\underbrace{\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n}_{\text{non-negative linear combination}} \text{ where } \alpha_i \geqslant 0 \text{ for } i = 1, \cdots, n.$$

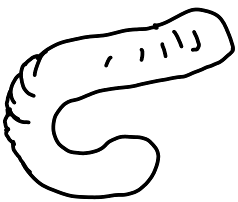
! If $a_i x \leq b_i$ for each i from 1 to m, then **every conical combination of** $(a_1, b_1), (a_2, b_2), \cdots, (a_m, b_m)$ corresponds to a "redundant" closed half-plane.

A Quick Detour to Reality™

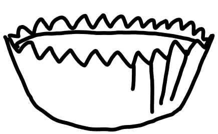
Suppose you own a ~~kitchen~~ candy factory. At the moment, you have 100 units of sugar and 210 units of food dye.



Hard candies require 3 units of sugar and 2 units of food dye. They sell for 8 cents each.



Gummy worms require 2 units of sugar and 3 units of food dye. They sell for 5 cents each.



Chocolatey cups require 4 units of sugar and 1 unit of food dye. They sell for 12 cents each.

A Quick Detour to Reality™

Suppose that you won't get another shipment of sugar and food dye for a while. How much of each type of candy should you make in the factory?

Let
x_1 = number of hard candies
x_2 = number of gummy worms
x_3 = number of chocolatey cups
.

sugar per unit

total sugar

① $3 \cdot x_1 + 2 \cdot x_2 + 4 \cdot x_3 \leq 100$

② $2 \cdot x_1 + 3 \cdot x_2 + 1 \cdot x_3 \leq 210$

food dye per unit

total food dye

③ $x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$

A Quick Detour to Reality™

Suppose that you won't get another shipment of sugar and food dye for a while. How much of each type of candy should you make in the factory?

$$① \quad 3\cdot x_1 + 2\cdot x_2 + 4\cdot x_3 \leq 100$$

$$② \quad 2\cdot x_1 + 3\cdot x_2 + 1\cdot x_3 \leq 210$$

$$③ \quad x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$$

Any $x \in \mathbb{R}^3$ that satisfies ①, ② and ③ is in the polyhedron corresponding to $((3,2,4), 100)$ and $((2,3,1), 210)$. We call x a **feasible solution**.

A Quick Detour to Reality™

Suppose that you won't get another shipment of sugar and food dye for a while. How much of each type of candy should you make in the factory?

Of course, you are only interested in the feasible solutions that $\underbrace{\textbf{maximize your profit}}_{\text{objective}}$.

We can express the desirability of a given feasible candy order (solution) by defining an **objective function**.

assigns "goodness" values

A Quick Detour to Reality™

So, expressed mathematically, the problem is to:

a linear program

maximize $8x_1 + 5x_2 + 12x_3$

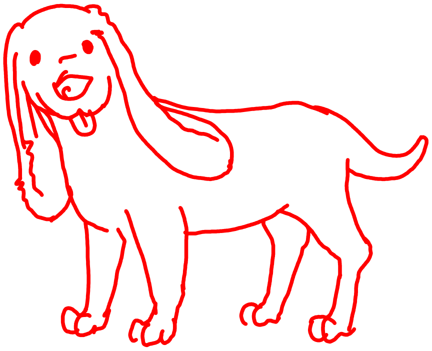
subject to the constraints

1. $3 \cdot x_1 + 2 \cdot x_2 + 4 \cdot x_3 \leq 100$
2. $2 \cdot x_1 + 3 \cdot x_2 + 1 \cdot x_3 \leq 210$
3. $x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$

Aside: There's an entire discipline dedicated to formulating real life problems as optimization/stats problems called **operations research**.

Diagrams in $\mathbb{R}^3$ are hard.

As an apology, have a puppy.



Finding a Solution

maximize $8x_1 + 5x_2 + 12x_3$

subject to the constraints

① $3 \cdot x_1 + 2 \cdot x_2 + 4 \cdot x_3 \leq 100$

② $2 \cdot x_1 + 3 \cdot x_2 + 1 \cdot x_3 \leq 210$

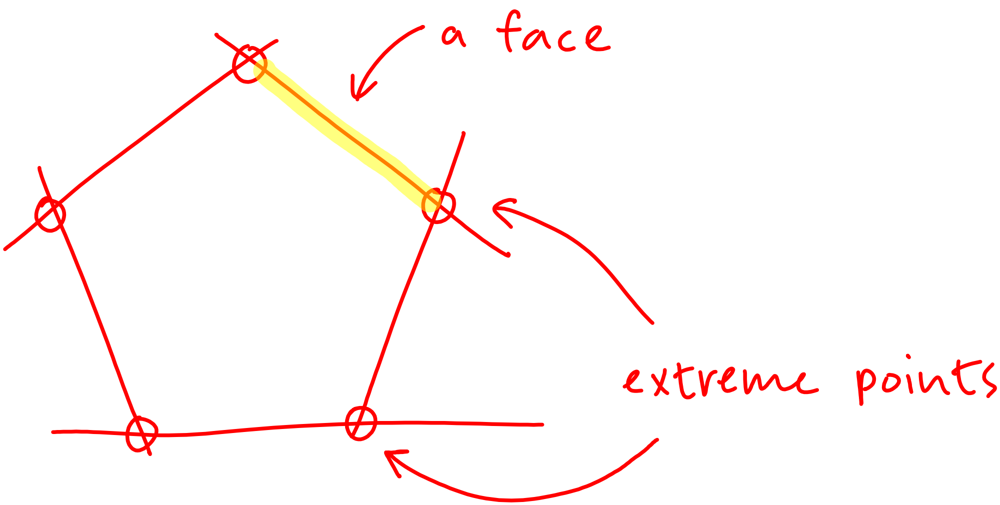
③ $x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$

So we have a well-defined mathematical formulation and a well-defined way to check whether a solution is feasible.

$(2, 1, -1)$	cannot make −1 chocolatey cup.
$(5, 10, 20)$	don't have enough sugar for this.
$(4, 8, 15)$	doable. Will earn \$2.52.
$(5, 8, 15)$	doable. Will earn \$2.60.

Polyhedron Terminology

First, I'll introduce terminology for some generalizations of concepts you are already familiar with in $\mathbb{R}^2$ and $\mathbb{R}^3$.

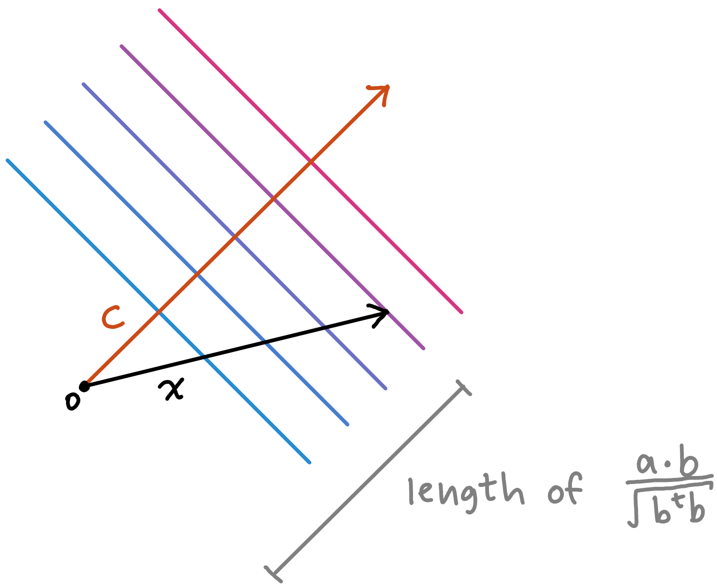


What does $c^t x$ mean?

Let's just consider the objective function for now.

How do we interpret $c^t x$?

We are **projecting** x **onto** c.



Look at the set of vectors that give the same "value".

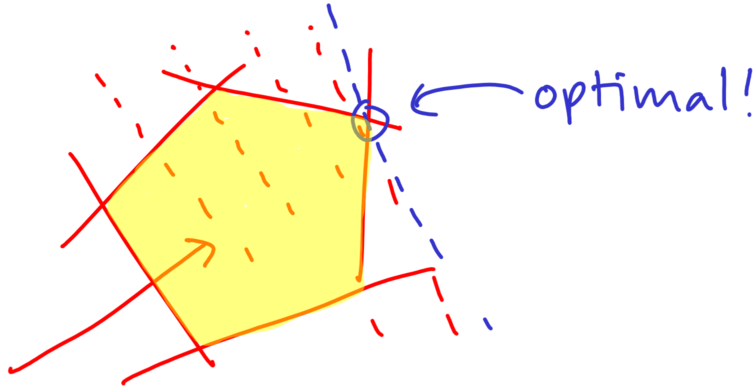
They form a hyperplane corresponding to "slope" c.

Maximizing the Objective Function

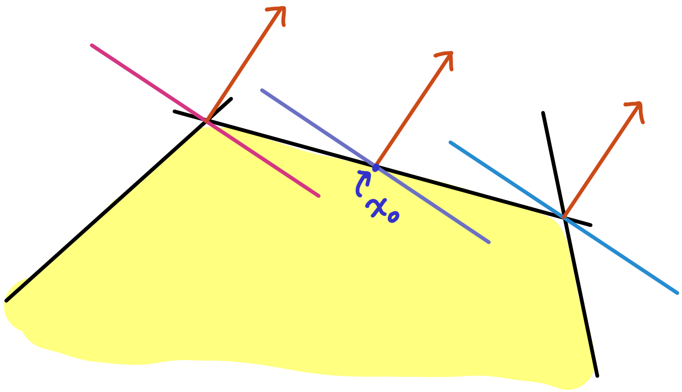
When the objective function is linear, maximizing $c^t x$ subject to constraints just means that we are trying to find the hyperplane with "slope" c that **has the biggest "offset"** out of all the hyperplanes that

a) have slope c, and

b) intersect with the feasible region.

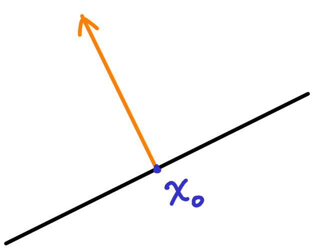


Linear Cost Functions



Let x_0 not be an extreme point. Find the hyperplane corresponding to "slope" c that goes through the point x_0.

There is always a "worse" direction, unless...



The **entire face** is contained in the hyperplane with slope given by c that goes through x_0.

wow

the fundamental theorem of linear programming

cool

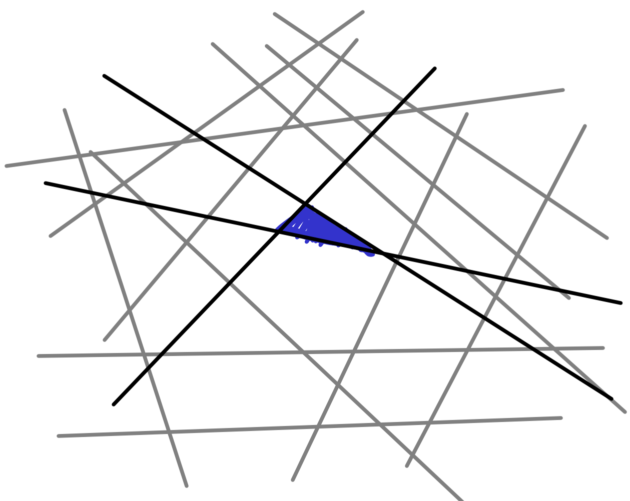
Fundamental Theorem of Linear Programming

Every optimal solution to a given linear program lies either at an extreme point of the feasible region or on a face where every point on the face is optimal.

means we can be a bit lazy. Yay!

Still not quite there yet.

- successfully narrowed down our searchspace.
- only need to look at the "outside" (**boundary**) of the feasible region (polyhedron formed by the constraints).
- how do we find **every single extreme point**?



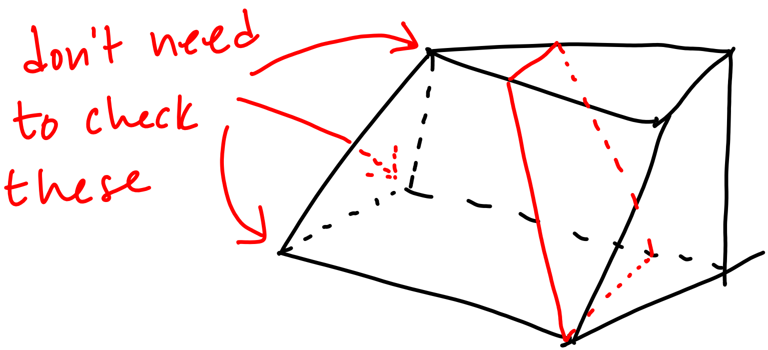
Even if we can find all extreme points,* can we do so efficiently?

This is unclear.

* look up Fourier-Motzkin elimination if you're curious

Still not quite there yet.

It's also unclear that it's necessary to look at every single extreme point, even if we could find all of them in a reasonable amount of time.



An Observation

Hey, do you remember conical combinations?

Def'n. Let $x_1, x_2, \cdots, x_k$ be vectors in $\mathbb{R}^n$. A **conical combination** is a vector of the form

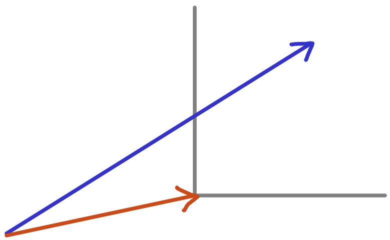
$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$ where $\alpha_i \geq 0$ for $i = 1, \cdots, n$.

Can we use conical combinations of constraint halfplanes to help solve our linear programming problem?

Spoiler: yes

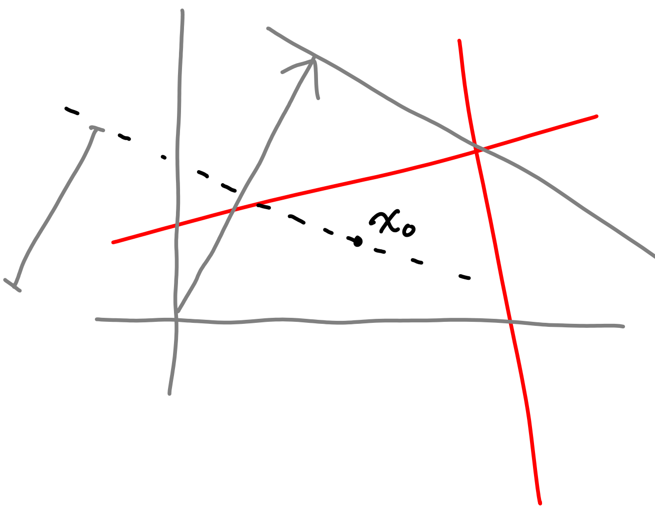
Conical Combinations of Hyperplanes

What if we found a conical combination $y_1 a_1 + \cdots + y_m a_m = y^t A$ where $(y^t A)_i \geq c_i$ for $i \in \{1, 2, \cdots, m\}$?

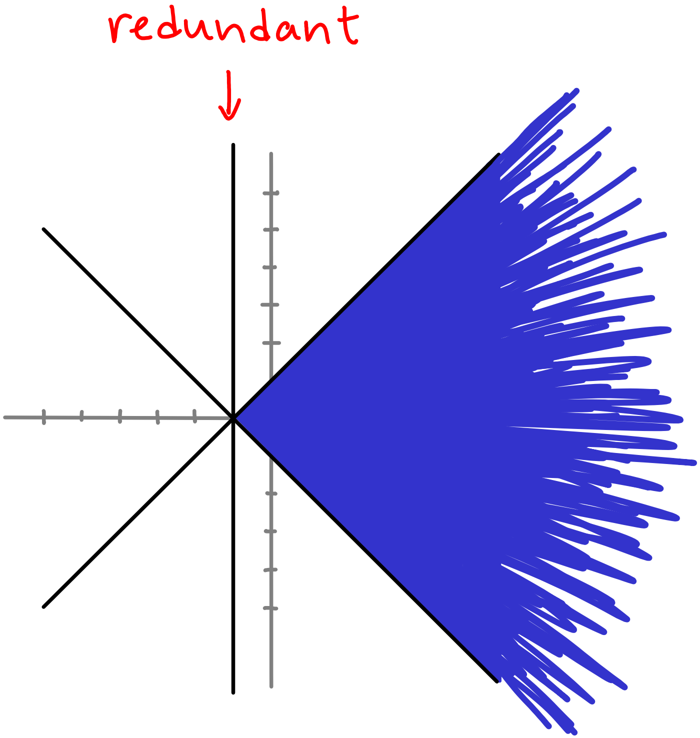


Graphically, $y^t A$ is a **longer** vector than c.

We also know that $y^t A x \leq y^t b$ for all x feasible to the original problem since $(y^t A, y^t b)$ corresponds to a redundant hyperplane.



Redundancy

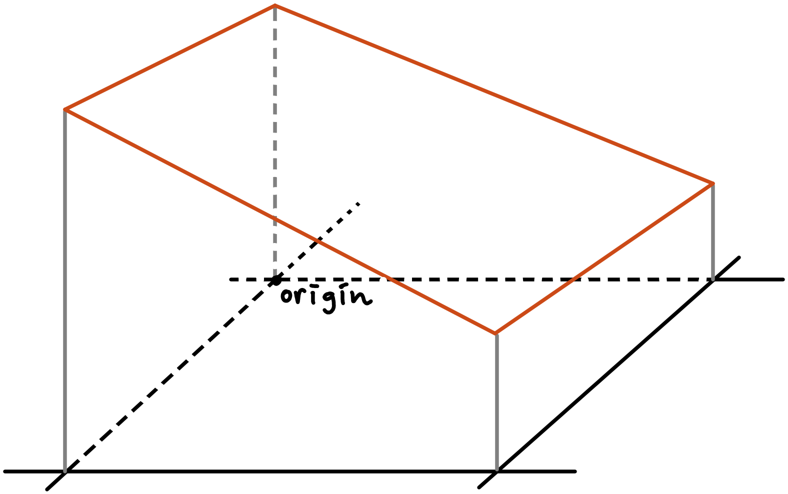


Recall that "redundant" means that the closed half-plane in question is not necessary to describe the feasible region.

This implies that **every feasible solution is in the closed half-plane given by the redundant hyperplane.**

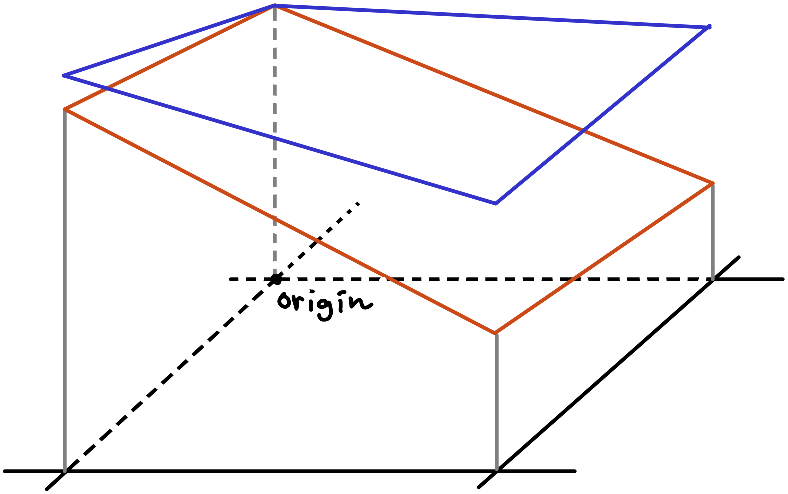
Graphs

Let's **graph** the function $f: \mathbb{R}^n \to \mathbb{R},\ f(x) = c^t x$ over the feasible region.



Graphs

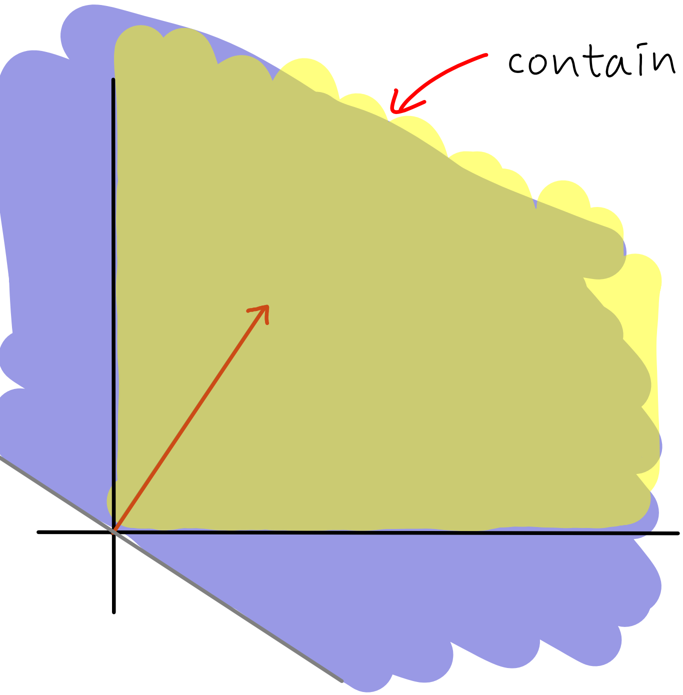
Let's **graph** the function $f: \mathbb{R}^n \to \mathbb{R}$, $f(x) = c^t x$ over the feasible region. $g: \mathbb{R}^n \to \mathbb{R}$, $g(x) = y^t A x$, too.



← this is a picture in $\mathbb{R}^{n+1}$

The Non-negative Orthant

Look at the set of points for which $(y^tA-c)^t x$ is non-negative.



contains the non-negative orthant

y^tAx will **overestimate** c^tx for all x in the non-negative orthant since

$y^tAx = c^tx + (y^tA-c)x$ and

* $(y^tA-c)x \geq 0$ for all x in the non-negative orthant.

*This is our motivation for restricting feasible solutions to be non-negative in all components.

Everything So Far

consequence of y^tA being a conical combination

If we know that $y^tA \geq c$ and $y \geq 0$,

y^tAx overestimates c^tx in the non-negative orthant

y^tb overestimates y^tAx for all x in the feasible region

inside non-negative orthant!

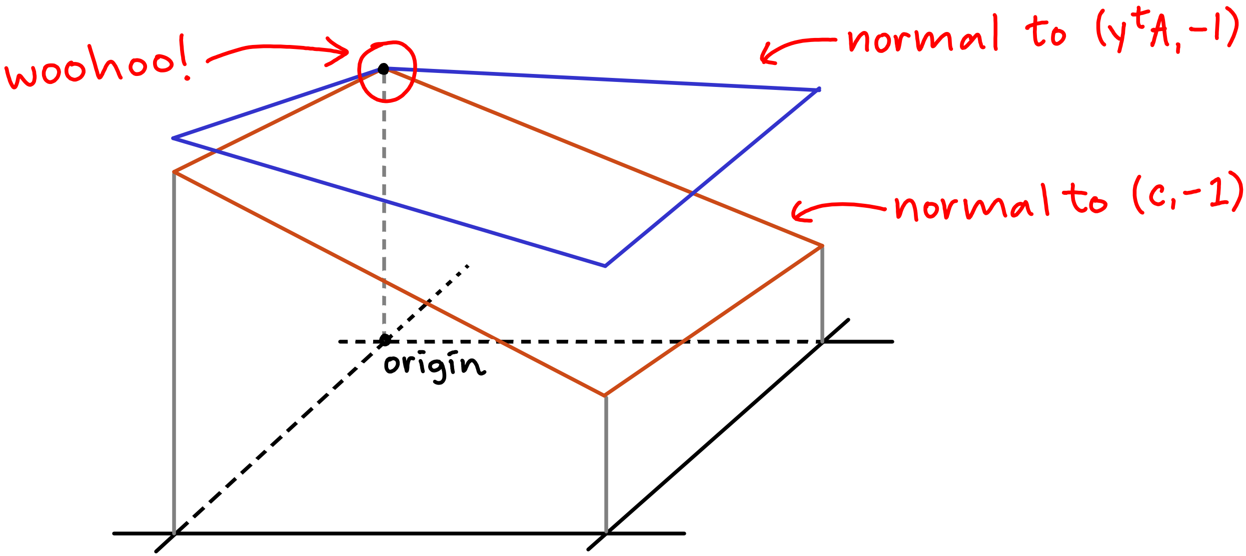
So y^tb **overestimates** c^tx for all feasible x!

As you can see, finding conical combinations of constraints that **overestimate the objective function** allows us to put **upper bounds** on the maximum value attainable by the objective function.

we have a stopping condition!

Upper Bounds For Fun and Profit

If we can find some feasible x for our linear program and some conical combination $y^t A$ where $y^t b = c^t x$, then we know that we **cannot find a better solution**!



A Pair of Problems

PRIMAL PROBLEM

original goal

maximize $c^t x$

subject to

$a_1 x \leq b_1$

$a_2 x \leq b_2$

$\vdots$

$a_m x \leq b_m$

$x_i \geq 0 \quad i = 1, 2, \cdots, m$

stay in the non-neg. orthant.

DUAL PROBLEM

upper bound on $c^t x$

minimize $b^t y$

subject to

$(y^t A)_1 \geq c_1$

$(y^t A)_2 \geq c_2$

$\vdots$

$(y^t A)_n \geq c_n$

so $y^t A x \geq c^t x$ in the non-neg orthant

$y_j \geq 0 \quad j = 1, 2, \cdots, n$

conical combinations

The Dual Problem

The dual problem is itself a linear program.

minimize $b^t y$ $\longrightarrow$ maximize $-b^t y$

subject to

$$\left.\begin{array}{l} (y^tA)_1 \geq c_1 \\ (y^tA)_2 \geq c_2 \\ \vdots \\ (y^tA)_n \geq c_n \end{array}\right\} \longrightarrow [(-A^t)y]_j \leq -c_j \quad \text{for } j=1,\cdots,n$$

Aside: oddly enough, the dual of the dual is the original primal problem. They are both equally good representations of a linear programming problem.

Interpreting the Dual Problem

While it is possible to treat the set of dual feasible solutions as a separate polyhedron in $\mathbb{R}^m$ ← # of primal constraints it **doesn't really make sense to**.

Rather, it encodes the set of redundant hyperplanes that sit on top of the graph of $c^t x$ over the set of feasible primal solutions.

In Case You Were Sleeping...

Here's a recap.

- for every linear programming problem, we can construct a **dual problem**.

"weak duality" →

- Every feasible solution for the dual **gives an upper bound** for the max. value attainable by the objective function over the feasible region of the original problem.

Tying Up Loose Ends

maximize $8x_1 + 5x_2 + 12x_3$

subject to the constraints

① $3 \cdot x_1 + 2 \cdot x_2 + 4 \cdot x_3 \leq 100$

② $2 \cdot x_1 + 3 \cdot x_2 + 1 \cdot x_3 \leq 210$

③ $x_1 \geq 0, \quad x_2 \geq 0, \quad x_3 \geq 0$

minimize $100y_1 + 210y_2$

subject to the constraints

- $3y_1 + 2y_2 \geq 8$
- $4y_1 + 1y_2 \geq 12$
- $2y_1 + 3y_2 \geq 5$
- $y_1 \geq 0, \; y_2 \geq 0$

$(x_1, x_2, x_3) = (0, 0, 25)$

$8 \cdot 0 + 5 \cdot 0 + 12 \cdot 25$

$=$

$0 + 0 + 300$

$=$

300

$=$

$300 + 0$

$100 \cdot 3 + 210 \cdot 0$

$(y_1, y_2) = (3, 0)$

More Questions

How do you find feasible solutions?

How do you find extreme points?

Is there a systematic way to find optimal solutions?

Well...

your CO255 or CO250 instructor is paid to tell you these things.

Other Interesting Things

Sometimes, your objective function isn't linear.

Sometimes your constraints aren't even linear.

These problems are outside the scope of this talk.

However, if you're interested in them, consider the following courses:

- CO 255 - Introduction to Optimization (adv level)
- CO 367 - Nonlinear Optimization
- CO 450 - Combinatorial Optimization
- CO 452 - Integer Programming
- CO 463 - Convex Optimization and Analysis
- CO 466 - Continuous Optimization
- CO 471 - Semidefinite Optimization

Thanks for Coming!

And also thanks to everyone who gave feedback on these slides.
You guys are awesome.